Statistical Methodology 8 (2011) 154–171

Contents lists available at ScienceDirect

Statistical Methodology

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

Skew scale mixtures of normal distributions: Properties and estimation

Clécio da Silva Ferreira [a], Heleno Bolfarine [b], Víctor H. Lachos [c,*]

[a] *Departamento de Estatística, Universidade Federal de Juiz de Fora, Juiz de Fora, MG, Brazil*

[b] *Departamento de Estatística, IME, Universidade de São Paulo, São Paulo, SP, Brazil*

[c] *Departamento de Estatística, IMECC, Universidade Estadual de Campinas, Campinas, SP, Brazil*

ARTICLE INFO

Article history:
Received 12 August 2009
Received in revised form
24 August 2010
Accepted 5 September 2010

Keywords:
EM algorithm
Scale mixtures of normal distributions
Skewness
Skew-t normal distribution

ABSTRACT

Scale mixtures of normal distributions are often used as a challenging class for statistical procedures for symmetrical data. In this article, we have defined a skewed version of these distributions and we have derived several of its probabilistic and inferential properties. The main virtue of the members of this family of distributions is that they are easy to simulate from and they also supply genuine EM algorithms for maximum likelihood estimation. For univariate skewed responses, the EM-type algorithm has been discussed with emphasis on the skew-t, skew-slash, skew-contaminated normal and skew-exponential power distributions. Some simplifying and unifying results are also noted with the Fisher information matrix, which is derived analytically for some members of this class. Results obtained from simulated and real data sets are reported, illustrating the usefulness of the proposed methodology. The main conclusion in reanalyzing a data set previously studied is that the models so far entertained are clearly not the most adequate ones.

1. Introduction

The scale mixtures of normal distributions [1] provide a group of thick tailed distributions that are often used for robust inference for symmetrical data. Moreover, this class includes distributions such as the Student-t, the slash, and the contaminated normal, among others. However, the theory

* Corresponding address: Departamento de Estatística, IMECC, Universidade Estadual de Campinas, Caixa Postal 6065, CEP 13083-859, Campinas, São Paulo, Brazil. Tel.: +55 1935216078.

E-mail address: hlachos@ime.unicamp.br (V.H. Lachos).

1572-3127/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.stamet.2010.09.001

and application (through simulation or experimentation) often generate a great amount of data sets that are skewed and present heavy tails—for instance, the data sets of family income [8] or substance concentration [13,9]. Thus, appropriate distributions for fitting and simulating these skewed and heavy tailed data are needed. Candidate distributions at our disposal for fitting and simulating these data are not very abundant in the literature. The skew-normal distribution is a new class of density functions dependent on an additional shape parameter, and includes the normal density as a special case. It provides a more flexible approach to the fitting of asymmetric observations.

The terminology *skew-normal distribution* was originally introduced by Azzalini [5], considering the probability density function (pdf) given by

$$f(y) = 2\phi(y|\mu, \sigma^2)\Phi\left(\frac{\lambda(y-\mu)}{\sigma}\right), \quad y \in \mathbb{R}, \tag{1}$$

where $\phi(.|\mu, \sigma^2)$ stands for the pdf of the normal distribution with mean μ and variance σ^2 and $\Phi(.)$ represents the cumulative distribution function (cdf) of the standard normal distribution. It is well known that the asymmetry range for this distribution is $(-0.995, 0.995)$. When $\lambda = 0$, the skew-normal distribution reduces to the normal distribution ($y \sim \mathrm{N}(\mu, \sigma^2)$). A random variable y with pdf as in (1) will be denoted by $SN(\mu, \sigma^2, \lambda)$. Its marginal stochastic representation [18], which can be used to derive several of its properties, is given by

$$Y \overset{d}{=} \mu + \sigma(\delta|T_0| + (1-\delta^2)^{1/2}T_1), \quad \text{with } \delta = \frac{\lambda}{\sqrt{1+\lambda^2}}, \tag{2}$$

where $T_0 \sim \mathrm{N}_1(0, 1)$ and $T_1 \sim \mathrm{N}(0, 1)$ are independent, $|T_0|$ denotes the absolute value of T_0, and "$\overset{d}{=}$" means "distributed as". From (2) it follows that the expectation and variance of Y are given, respectively, by

$$E[Y] = \mu + \sqrt{\frac{2}{\pi}}\sigma\delta, \qquad \mathrm{Var}[Y] = \sigma^2\left(1 - \frac{2}{\pi}\delta^2\right). \tag{3}$$

Reasoning as [5], it is natural to construct univariate and multivariate distributions that combine skewness with heavy tails. The main idea behind the construction above involves a density function defined as the product of the normal density with its cumulative distribution function. Hence, a more general set-up can be considered if we multiply a general density by any distribution function. This idea was put forward in [5], in which the following result can be found:

Lemma 1. *Let g be a symmetric pdf and G a distribution function such that G' exists and is symmetric (all around zero); then*

$$f_X(x|\lambda) = 2f(x)G(\lambda x), \quad -\infty < x < \infty,$$

is a density for any $\lambda \in \mathbb{R}$.

The above lemma allows generating a multitude of skew-symmetric distributions. Nadarajah and Kotz [30] consider f and G was taken as coming from the normal, Student-t, Cauchy, Laplace, logistic and uniform distributions. It can be shown for this family of distributions that the asymmetry parameter $\sqrt{\beta_1}$ is confined to the interval $-0.995 < \sqrt{\beta_1} < 0.995$, which is exactly the same as for the ordinary skew-normal distribution. On the other hand, Gupta et al. [17] consider f and G to belong to the normal, Student-t, Cauchy, Laplace, logistic or uniform family. Some other classes of asymmetric distributions are considered, for example, the skew-t distribution [33,8], skew-Cauchy distribution [3], skew-slash distribution [35], and skew-elliptical distributions [7,10,33,14].

Gómez et al. [15] consider the situation where G is fixed as the cumulative distribution function of the normal distribution, while f is taken as the density function of the normal, Student-t, logistic, Laplace and uniform distributions, calling it the skew-symmetric normal distribution. In this family of distributions, the range for asymmetry and kurtosis parameters are wider than for the family of models introduced by Nadarajah and Kotz [30] and, as a particular case, the skew-normal distribution.

By using G as in [15], an alternative family of asymmetric distributions is pursued in this article. The approach leads to a new family of asymmetric univariate distributions generated by the normal kernel (as the skewing function), using otherwise symmetric distributions of the class of scale mixtures of normal distributions [1,24]. We study some of its probabilistic and inferential properties and discuss applications to real data. One interesting and simplifying aspect of the family defined is that the implementation of the EM algorithm is facilitated by the fact that the E-step is exactly as in the scale mixtures of the normal distribution class of models proposed in [1]. Moreover, the M-step involves closed form expressions facilitating the implementation of the algorithm. Furthermore, we also found that the information matrix has a common part for all elements in the class, which makes, for instance, the skew-t-normal (StN) form have a nonsingular information matrix as shown in [15]. The SSMN class proposed here is fundamentally different from the scale mixtures of normal distributions (SMSN) developed by Branco and Dey [10] – see also [22]– because we start our construction from the SMN densities and not from the stochastic representation of a skew-normal random variable as presented in [22].

The rest of the article is organized as follows. In Section 2, the family of skew scale mixtures of normal distributions (SSMN) is defined by extending the symmetric class of scale mixtures of normal distributions (SMN). Properties like moments and a stochastic representation of the proposed distributions are also discussed. In Section 3, some examples of SSMN distributions are presented. In Section 4, the observed information matrix is derived analytically and we discuss how to compute ML estimates for the skew-normal distribution by using the Newton–Raphson method and the EM algorithm, which presents advantages over the direct maximization approach, especially in terms of robustness with respect to starting values. Section 5 reports applications to simulated and real data sets, indicating the usefulness of the proposed methodology. Finally, Section 6 concludes with some discussions, citing avenues for future research.

2. Skew scale mixtures of normal distributions

To better motivate our proposed methodology, we give a brief introduction to SMN distributions, starting with the definition. Andrews and Mallows [1] use the Laplace transform technique to prove that a standardized continuous random variable Y has a scale mixture of normal (SMN) distribution if its density function can be expressed as

$$Y = \mu + \kappa(U)^{1/2}Z, \tag{4}$$

where $\kappa(.)$ is a strictly positive function, and $Z \sim \mathrm{N}(0, \sigma^2)$ is independent of the positive random variable U with cdf $H(.; \boldsymbol{\tau})$, which is indexed by the parameter vector $\boldsymbol{\tau}$. Thus, we have the following definition:

Definition 1. A random variable Y follows a SMN distribution with location parameter $\mu \in \mathbb{R}$ and scale parameter σ^2 if its pdf assumes the form

$$f_0(y) = \int_0^\infty \phi(y|\mu, \kappa(u)\sigma^2)\mathrm{d}H(u; \boldsymbol{\tau}). \tag{5}$$

For a random variable with a pdf as in (5), we shall use the notation $Y \sim SMN(\mu, \sigma^2, H; \kappa)$. Moreover, when $\mu = 0$ and $\sigma^2 = 1$, we use the notation $Y \sim SMN(H; \kappa)$.

On the basis of Definition 1 and Lemma 1, in the following definition we introduce the new class of skew scale mixture of normal (SSMN) distributions, whose properties will be studied in this work.

Definition 2. A random variable Y follows a SSMN distribution with location parameter $\mu \in \mathbb{R}$, scale factor σ^2 and skewness parameter $\lambda \in \mathbb{R}$ if its pdf is given by

$$f(y) = 2f_0(y)\Phi\left(\lambda\frac{y-\mu}{\sigma}\right), \tag{6}$$

where $f_0(y)$ is a SMN density as defined in (5). For a random variable with pdf as in (6), we use the notation $Y \sim SSMN(\mu, \sigma^2, \lambda, H; \kappa)$. If $\mu = 0$ and $\sigma^2 = 1$ we refer to it as the standard SSMN distribution and we denote it by SSMN$(\lambda, H; \kappa)$.

Clearly, when $\lambda = 0$, we get the corresponding SMN distribution proposed by Andrews and Mallows [1]. For a SSMN random variable, a convenient stochastic representation is given next, which can be used to quickly simulate realizations of Y, to implement the EM algorithm and also to study many of its properties.

Proposition 1. *Let $Y \sim SSMN(\mu, \sigma^2, \lambda, H; \kappa)$. Then its stochastic representation is given by*

$$Y|U = u \sim SN(\mu, \sigma^2\kappa(u), \lambda\sqrt{\kappa(u)}), \quad U \sim H(u; \boldsymbol{\tau}). \tag{7}$$

Proof. From (6), the (joint) distribution of (Y, U) is given by

$$g(y, u) = 2\phi(y|\mu, \sigma^2\kappa(u))\Phi(\lambda(y-\mu)/\sigma)h(u; \boldsymbol{\tau}).$$

Then

$$g(y \mid u) = 2\phi(y|\mu, \sigma^2\kappa(u))\Phi\left(\lambda\frac{(y-\mu)}{\sigma}\right) = 2\phi(y|\mu, \sigma^2\kappa(u))\Phi\left(\lambda\sqrt{\kappa(u)}\frac{(y-\mu)}{\sigma\sqrt{\kappa(u)}}\right),$$

and hence $Y \mid U = u \sim SN(\mu, \sigma^2\kappa(u), \lambda\sqrt{\kappa(u)})$. □

Since (6) is a skew-symmetric construction, then Y allows also a conditional stochastic representation as defined in [34, Eq. (8)], which presents a familiar procedure for simulating from skew-symmetric distributions. Alternatively, from (7), to generate a SSMN random variable, we proceed in two steps, that is, we generate first from the distribution of U and next from the conditional distribution $Y \mid U$ using, for instance, the stochastic representation given in (2).

Next, we present some properties related to the SSMN class of distributions. For instance, the following proposition shows that under the more general SSMN distribution considered here, the conditional distribution $U|y$ is the same as when considering the corresponding symmetric SMN model. The proof is direct and will be omitted.

Proposition 2 (*An Invariance Result*)**.** *If $Y \sim SSMN(\mu, \sigma^2, \lambda, H; \kappa)$, then the conditional distribution of $U \mid Y$ does not depend of λ.*

The following section summarizes some distributions of the SSMN class. For each distribution, we present its pdf, mean and variance, the Mahalanobis distance and the conditional expectation $\widehat{\kappa} = E[\kappa^{-1}(U)|Y = y]$, which is exactly the same as in the SMN class and will be useful in the implementation of the EM algorithm. Other members of the SSMN class can be derived from Choy and Smith [11]. However, for some cases, the scale distribution $H(u; \boldsymbol{\tau})$ does not have a computationally attractive form and hence will not be dealt with in this work.

3. Examples of SSMN distributions

3.1. The skew-Student-t normal distribution (StN)

The use of the t distribution as an alternative to the normal distribution has been frequently suggested in the literature. For instance, Little [27] and Lange et al. [23] recommend using the Student-t distribution for robust modeling. The StN [15] distribution with ν degrees of freedom, $StN(\mu, \sigma^2, \lambda, \nu)$, can be derived from the mixture model (7), with $U \sim$ Gamma$(\nu/2, \nu/2)$, $\nu > 0$ and $\kappa(u) = 1/u$. The pdf of Y takes the form

$$f(y) = 2\frac{1}{\sigma\sqrt{\nu\pi}}\frac{\Gamma((\nu+1)/2)}{\Gamma\left(\frac{\nu}{2}\right)}\left(1+\frac{d}{\nu}\right)^{-\left(\frac{\nu+1}{2}\right)}\Phi\left(\lambda\frac{(y-\mu)}{\sigma}\right), \tag{8}$$

where $d = (y-\mu)^2/\sigma^2$. From results given in [24] in conjunction with Corollary 2 (Appendix A), it follows that the Mahalanobis distance $D = (Y-\mu)/\sigma^2 \sim F(1,\nu)$, where $F(a,b)$ denotes the F distribution with a and b degrees of freedom.

The *StN* distribution has been studied by Gómez et al. [15]. In that paper, the authors show that the *StN* distribution can present a much wider asymmetry range than the one presented by the ordinary skew-normal distribution [5]. A particular case of the *StN* distribution is the skew-Cauchy normal distribution, that follows when $\nu = 1$. Also, when $\nu \uparrow \infty$, we get the skew-normal distribution as the limiting case. In this case, from Corollary 1, the mean and variance of Y are given, respectively, by

$$E[Y] = \mu + b\sigma\lambda(\nu/2)^{1/2}\frac{\Gamma((\nu-1)/2)}{\Gamma(\nu/2)}E_V\left[(V+\lambda^2)^{-1/2}\right],$$
$$\mathrm{Var}[Y] = \sigma^2\left[\frac{\nu}{\nu-2} - \frac{b^2\lambda^2\nu}{2}\left(\frac{\Gamma((\nu-1)/2)}{\Gamma(\nu/2)}\right)^2 E_V^2\left[(V+\lambda^2)^{-1/2}\right]\right],$$

where $b = \sqrt{\frac{2}{\pi}}$ and $V \sim \text{Gamma}\left(\frac{\nu-1}{2}, \frac{\nu}{2}\right)$ and the expected values can be computed numerically. Since $U|Y=y \sim \text{Gamma}\left(\frac{\nu+1}{2}, \frac{\nu+d}{2}\right)$, it follows that $\widehat{k} = \frac{\nu+1}{\nu+d}$.

3.2. The skew-slash distribution (SSL)

Another SSMN distribution, termed the skew-slash distribution and denoted by $SSL(\mu,\sigma^2,\lambda,\nu)$, arises when $\kappa(u) = 1/u$ and the distribution of U is $\text{Beta}(\nu,1)$, $0 < u < 1$ and $\nu > 0$. Its pdf is given by

$$f(y) = 2\nu\Phi\left(\lambda\frac{y-\mu}{\sigma}\right)\int_0^1 u^{\nu-1}\phi\left(y|\mu,\frac{\sigma^2}{u}\right)\mathrm{d}u, \quad y \in \mathbb{R}. \tag{9}$$

The skew-slash distribution reduces to the skew-normal distribution when $\nu \uparrow \infty$. In this case, from Lange and Sinsheimer [24], the Mahalanobis distance has cdf

$$\Pr(D \le r) = \Pr(\chi^2 \le r) - \frac{2^\nu\Gamma(\nu+1/2)}{r^\nu\sqrt{\pi}}\Pr(\chi^2_{2\nu+1} \le r)$$

and the mean and variance are given, respectively, by

$$E[Y] = \mu + \frac{b\sigma\lambda\nu}{\nu-1/2}E_V\left[(V+\lambda^2)^{-1/2}\right],$$
$$\mathrm{Var}[Y] = \sigma^2\left(\frac{\nu}{\nu-1} - \frac{b^2\lambda^2\nu^2}{(\nu-1/2)^2}E_V^2\left[(V+\lambda^2)^{-1/2}\right]\right),$$

where $V \sim \text{Beta}(1,\nu-1/2)$. It is easy to see that $U|Y=y \sim TG(\nu+1/2, d/2, 1)$, where $TG(a,b,t)$ is the right truncated gamma distribution, with pdf $f(x|a,b,t) = \frac{b^a}{\gamma(a,bt)}x^{a-1}\exp(-bx)\mathbb{I}_{(0,t)}(x)$, and $\gamma(a,b) = \int_0^b u^{a-1}\mathrm{e}^{-u}\mathrm{d}u$ is the incomplete gamma function. Thus, $\widehat{k} = \frac{(2\nu+1)}{d}\frac{P_1(\nu+3/2,d/2)}{P_1(\nu+1/2,d/2)}$, where $P_x(a,b)$ denotes the cdf of the $\text{Gamma}(a,b)$ distribution evaluated at x.

3.3. The skew-contaminated normal distribution (SCN)

The skew-contaminated normal distribution is denoted by $SCN(\mu,\sigma^2,\lambda,\nu,\gamma)$, $0 < \nu < 1$, $0 < \gamma < 1$. Here, $\kappa(u) = 1/u$ and U is a discrete random variable taking one of two states. The probability density function of U, given the parameter vector $\boldsymbol{\tau} = (\nu,\gamma)^\top$, is denoted by $h(u;\boldsymbol{\tau}) = \nu\mathbb{I}_{(u=\gamma)} + (1-\nu)\mathbb{I}_{(u=1)}$, $\boldsymbol{\tau} = (\nu,\gamma)^\top$. It follows then that

$$f(y) = 2\left\{\nu\phi\left(y|\mu,\frac{\sigma^2}{\gamma}\right) + (1-\nu)\phi(y|\mu,\sigma^2)\right\}\Phi\left(\lambda\frac{y-\mu}{\sigma}\right).$$

The skew-contaminated normal distribution reduces to the skew-normal distribution when $\gamma \to 1$. In this case the mean and the variance are given, respectively, by

$$E[Y] = \mu + b\sigma\lambda\left(\frac{\nu}{(\gamma(\gamma+\lambda^2))^{1/2}} + \frac{1-\gamma}{(1+\lambda^2)^{1/2}}\right),$$

$$\mathrm{Var}[Y] = \sigma^2\left[\frac{\nu}{\gamma} + 1 - \nu - b^2\lambda^2\left(\frac{\nu}{(\gamma(\gamma+\lambda^2))^{1/2}} + \frac{1-\gamma}{(1+\lambda^2)^{1/2}}\right)^2\right].$$

The Mahalanobis distance has a cdf given by $\Pr(D \le r) = \nu \Pr(\chi^2 \le \gamma r) + (1-\nu)\Pr(\chi^2 \le r)$ and $\widehat{k} = \frac{1-\nu+\nu\gamma^{3/2}\exp\{(1-\gamma)d/2\}}{1-\nu+\nu\gamma^{1/2}\exp\{(1-\gamma)d/2\}}$.

3.4. *The skew-exponential power distribution (SEP)*

West [36] has shown that the univariate exponential power distribution (EP) is one of the scale mixture of normal distributions for $0.5 < \nu \le 1$, and U is distributed according to a stable distribution $U^P(u|\nu)$. However, as shown in [21], this property is not valid in a multivariate context. Recently, Gómez-Villegas et al. [16] have shown that the multivariate EP distribution is a scale mixture of normal distribution, when its kurtosis parameter belongs to the interval (0, 1]. An alternative asymmetric version of the exponential power distribution has been considered in [6,12], where some inferential aspects were studied.

The skew-exponential power distribution that will be used here, denoted by $SEP(\mu, \sigma^2, \lambda, \nu)$, has pdf given by

$$f(y) = 2\frac{\nu}{\sqrt{2^\nu}\sigma\Gamma(1/2\nu)}\mathrm{e}^{-d^\nu/2}\Phi\left(\lambda\frac{y-\mu}{\sigma}\right), \quad 0.5 < \nu \le 1, \tag{10}$$

which reduces to the skew-normal distribution when $\nu = 1$. From Lange and Sinsheimer [24], the Mahalanobis distance has a cdf given by

$$\Pr(D \le r) = \frac{r^{1/2}\gamma\left(\frac{1}{2\nu}, r^\nu/2\right)}{\Gamma\left(\frac{1}{2\nu}\right)2^{\frac{1}{2\nu}}},$$

where $\gamma(a, b)$ is the incomplete gamma function. The following result is useful for evaluating the expected value $\widehat{k}$ when $Y \sim SEP(\mu, \sigma^2, \lambda, \nu)$.

Proposition 3. *If $Y \sim SEP(\mu, \sigma^2, \lambda; \nu)$, then*

$$\widehat{\kappa} = E[\kappa^{-1}(U)|y] = \nu d^{\nu-1}, \quad \nu \neq 0.5. \tag{11}$$

Proof. $f_Y(y) = \int_0^{+\infty}\phi(y|\mu, \sigma^2\kappa(u))h(u; \boldsymbol{\tau})\mathrm{d}u$ and $f_{(U,Y)}(u, y) = \phi(y|\mu, \sigma^2\kappa(u))h(u; \boldsymbol{\tau})$. As $V = \kappa^{-1}(U)$, $\kappa(u) = 1/V$, then

$$f_{(V,Y)}(v, y) = f_{(U,Y)}(u, y)|J| \propto \phi(y|\mu, \sigma^2/v)h(\kappa^{-1}(1/v); \boldsymbol{\tau}),$$

where J is the Jacobian of the transformation $g(u, y) = (v, y)$. So,

$$\begin{aligned} f_{(V|Y=y)}(v|y) &= \frac{\phi(y|\mu, \sigma^2/v)f_U(k^{-1}(1/v))}{f(y)} \\ &= h(\kappa^{-1}(1/v); \boldsymbol{\tau})\frac{2^{\frac{1}{2\nu}}\sigma\Gamma\left(\frac{1}{2\nu}\right)}{\nu}\mathrm{e}^{d^\nu/2}\frac{v^{1/2}}{\sqrt{2\pi}\sigma}\mathrm{e}^{-\frac{(y-u)^2}{2\sigma^2}v} \\ &\propto \exp\left\{\frac{1}{2}\left[-\mathrm{d}v + d^\nu\right] + c(v, \nu)\right\}. \end{aligned}$$

Hence, $\kappa^{-1}(U)|Y$ belongs to the exponential family, with $\theta = -d$, $b(\theta) = -d^\nu = -(-\theta)^\nu$ and dispersion parameter equal to 1/2. Thus, $E_U[k^{-1}(U) \mid Y = y] = b'(\theta) = \partial b(\theta)/\partial\theta = \nu d^{\nu-1}$. □

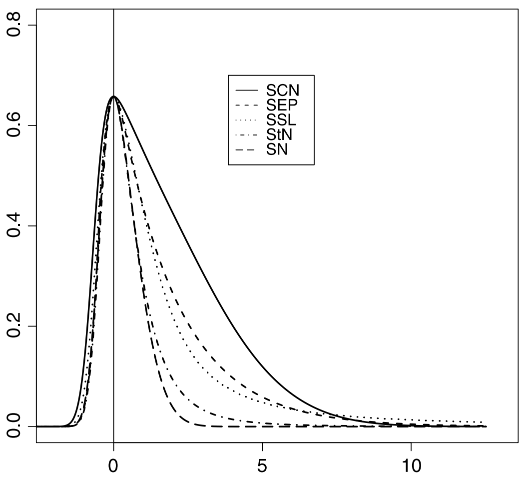

Fig. 1. Density curves of the univariate skew-normal $SN(3)$, skew-t normal $StN(3, 2)$, skew-slash $SSL(3, 0.5)$, skew-contaminated normal $SCN(3, 0.9, 0.1)$ and skew-exponential power $SEP(3, 0.5)$ distributions.

In Fig. 1, we plotted the density of the standard $SN(3)$ distribution together with the standard densities of the distributions $StN(3, 2)$, $SSL(3, 0.5)$, $SEP(3, 0.5)$ and $SCN(3, 0.9, 0.1)$. They are rescaled so that they have the same value at the origin. Note that the five densities are positively skewed, and that the skew-contaminated normal, skew-exponential power, skew-slash, and the skew-t distributions have much heavier tails than the skew-normal distribution. Note that they all represent extreme situations, even in the symmetric case where, for instance, the slash distribution with $\nu = 0.5$ does not have finite first moment.

4. Independent responses

Suppose that we have observations on m independent individuals, denoted by $Y_1, \ldots, Y_m$, where $Y_i \sim SSMN(\mu, \sigma^2, \lambda, H; \kappa)$, $i = 1, \ldots, m$. Hence, for an observed sample $\mathbf{y} = (y_1, \ldots, y_m)^\top$, the log-likelihood function of $\boldsymbol{\theta} = (\mu, \sigma^2, \lambda, \boldsymbol{\tau}^\top)^\top$ is of the form $\ell(\boldsymbol{\theta}) = \sum_{i=1}^m \ell_i(\boldsymbol{\theta})$, with $\ell_i(\boldsymbol{\theta}) = \log 2 + \ell_{1_i}(\boldsymbol{\theta}) + \log[\Phi_1(\ell_{2_i}(\boldsymbol{\theta}))]$, where $\ell_1(\boldsymbol{\theta})$ is the log-likelihood function of the corresponding symmetric SMN distribution and $\ell_{2_i}(\boldsymbol{\theta}) = \lambda \frac{y_i - \mu}{\sigma}$. Then, the vector of the first derivative and the matrix of the second derivative of $\ell_i(\boldsymbol{\theta})$ are given, respectively, by

$$\mathbf{s}_{\boldsymbol{\theta}} = \frac{\partial \ell_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \frac{\partial \ell_{1_i}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} + W_\Phi(l_{2_i}(\boldsymbol{\theta}))\frac{\partial \ell_{2_i}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}},$$

$$\frac{\partial^2 \ell_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} = \frac{\partial^2 \ell_{1_i}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} + W_\Phi(\ell_{2_i}(\boldsymbol{\theta}))\frac{\partial^2 \ell_{2_i}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} + W_\Phi^{(1)}(\ell_{2_i}(\boldsymbol{\theta}))\frac{\partial \ell_{2_i}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\frac{\partial \ell_{2_i}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^\top},$$

where $W'_\Phi(x) = -W_\Phi(x)(x + W_\Phi(x))$, with $W_\Phi(x) = \phi(x)/\Phi_1(x)$. Thus, the observed information matrix for $\boldsymbol{\theta}$ can be written as

$$\mathbf{I}_{\boldsymbol{\theta}\boldsymbol{\theta}} = -\sum_{i=1}^m \frac{\partial^2 \ell_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} = \mathbf{I}_1(\boldsymbol{\theta}) + \mathbf{I}_2(\boldsymbol{\theta}),$$

where $\mathbf{I}_1(\boldsymbol{\theta}) = -\sum_{i=1}^m \frac{\partial^2 \ell_{1_i}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top}$, and

$$\mathbf{I}_2(\boldsymbol{\theta}) = -\sum_{i=1}^m \left[W_\Phi(\ell_{2_i}(\boldsymbol{\theta}))\frac{\partial^2 \ell_{2_i}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} + W_\Phi^{(1)}(\ell_{2_i}(\boldsymbol{\theta}))\frac{\partial \ell_{2_i}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\frac{\partial \ell_{2_i}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^\top}\right].$$

For all distributions in the SMN class, the elements of $\mathbf{I}_2(\boldsymbol{\theta})$ are common and are given by

$$I^2_{\mu\mu} = -\frac{\lambda^2}{\sigma^2}\sum_{i=1}^m W'_\Phi\left(\lambda\frac{y_i-\mu}{\sigma}\right),$$

$$I^2_{\sigma^2\mu} = -\frac{\lambda}{2\sigma^3}\sum_{i=1}^m W_\Phi\left(\lambda\frac{y_i-\mu}{\sigma}\right) - \frac{\lambda^2}{2\sigma^4}\sum_{i=1}^m W'_\Phi\left(\lambda\frac{y_i-\mu}{\sigma}\right)(y_i-\mu),$$

$$I^2_{\lambda\mu} = \frac{1}{\sigma}\sum_{i=1}^m W_\Phi\left(\lambda\frac{y_i-\mu}{\sigma}\right) + \frac{\lambda}{\sigma^2}\sum_{i=1}^m W'_\Phi\left(\lambda\frac{y_i-\mu}{\sigma}\right)(y_i-\mu),$$

$$I^2_{\sigma^2\sigma^2} = -\frac{3\lambda}{4\sigma^5}\sum_{i=1}^m W_\Phi\left(\lambda\frac{y_i-\mu}{\sigma}\right)(y_i-\mu) - \frac{\lambda^2}{4\sigma^6}\sum_{i=1}^m W'_\Phi\left(\lambda\frac{y_i-\mu}{\sigma}\right)(y_i-\mu)^2,$$

$$I^2_{\lambda\sigma^2} = \frac{1}{2\sigma^3}\sum_{i=1}^m W_\Phi\left(\lambda\frac{y_i-\mu}{\sigma}\right)(y_i-\mu) + \frac{\lambda}{2\sigma^4}\sum_{i=1}^m W'_\Phi\left(\lambda\frac{y_i-\mu}{\sigma}\right)(y_i-\mu)^2,$$

$$I^2_{\lambda\lambda} = -\frac{1}{\sigma^2}\sum_{i=1}^m W'_\Phi\left(\lambda\frac{y_i-\mu}{\sigma}\right)(y_i-\mu)^2,$$

$$I^2_{\boldsymbol{\tau}\boldsymbol{\theta}} = \mathbf{0}.$$

Notice that as $W_\Phi(x) = \phi(x)/\Phi(x)$, then $-1 \le W'_\Phi(x) \le 0$. The fact that this matrix is constant for all families in the class makes the information matrix nonsingular for the StN model [15] and, we hope, for other situations. As shown in [2] (see also [31]), the singularity of the information matrix for the skew-normal model certainly is a problem with the ordinary skew-normal distribution. Expressions for the elements of $\mathbf{I}_1(\boldsymbol{\theta})$, for some elements of the SMN class, are given in the Appendix.

4.1. Maximum likelihood estimation

Since ML estimation for parameters in this class of models does not lead to closed form expressions, the popular Newton–Raphson (NR) algorithm can be applied to calculate the ML estimates iteratively. Starting from an initial point $\boldsymbol{\theta}^{(0)}$, the NR procedure proceeds according to

$$\widehat{\boldsymbol{\theta}}^{(k+1)} = \widehat{\boldsymbol{\theta}}^{(k)} + \widehat{\mathbf{I}}_{\boldsymbol{\theta}\boldsymbol{\theta}}^{-1(k)}\widehat{\mathbf{s}}_{\boldsymbol{\theta}}^{(k)}, \tag{12}$$

where $\widehat{\mathbf{s}}_{\boldsymbol{\theta}}$ and $\widehat{\mathbf{I}}_{\boldsymbol{\theta}\boldsymbol{\theta}}$ are the score vector and the observed information matrix evaluated at $\boldsymbol{\theta}^{(k)}$, respectively. Other related iterative algorithms such as steepest ascent, Davidon–Fletcher–Powell, conjugate gradient, quasi-Newton and Marquardt–Levenberg ones, which are gradient methods like the NR algorithm, can lead to equivalent calculations.

An oft-voiced complaint as regards the NR algorithm is that it may not converge unless good starting values are used. The EM algorithm, which takes advantage of being insensitive to the starting values as a powerful computational tool that requires the construction of unobserved data, has been well developed and has become a broadly applicable approach for the iterative computation of ML estimates. One of the major reasons for the popularity of the EM algorithm is that the M-step involves only complete data ML estimation, which is often computationally simple. Moreover, the EM algorithm is stable and straightforward to implement since the iterations converge monotonically and no second derivatives are required. When the M-step of EM turns out to be analytically intractable, it can be replaced with a sequence of conditional maximization steps (CM-steps). Such a modification is referred to as the ECM algorithm [29]. The ECME algorithm [28], a faster extension of EM and ECM, is obtained by maximizing the constrained Q-function (the expected complete data function) with some CM-steps that maximize the corresponding constrained actual marginal likelihood function, called the CML-steps. In the following, we demonstrate how to employ the ECME algorithms for ML estimation of a SSMN distribution.

Note first that from (2) and (7) it follows that

$$Y_i|T_i = t_i, U_i = u_i \overset{\text{ind}}{\sim} N\left(\mu + \sigma\lambda\frac{\kappa(u_i)}{\sqrt{1+\lambda^2\kappa(u_i)}}t_i, \sigma^2\kappa(u_i)\left(1-\lambda^2\frac{\kappa(u_i)}{1+\lambda^2\kappa(u_i)}\right)\right),$$
$$U_i \overset{\text{ind}}{\sim} h(u_i;\boldsymbol{\tau}), \qquad T_i \overset{\text{i.i.d.}}{\sim} HN_1(0,1), \quad i = 1,\ldots,m,$$

all independent, where $HN_1(0, 1)$ denotes the univariate standard half-normal distribution (see $|T_0|$ in Eq. (2) or [19]). Letting $\mathbf{y} = (y_1, \ldots, y_m)^\top$, $\mathbf{u} = (u_1, \ldots, u_m)^\top$ and $\mathbf{t} = (t_1, \ldots, t_m)^\top$ and treating $\mathbf{u}$ and $\mathbf{t}$ as missing data, it follows that the complete log-likelihood function associated with $\mathbf{y}_c = (\mathbf{y}^\top, \mathbf{u}^\top, \mathbf{t}^\top)^\top$ is given by

$$\ell_c(\boldsymbol{\theta}|\mathbf{y}_c) = K + \sum_{i=1}^m \left[-\log\sigma^2 - \frac{1}{2\sigma^2}t_i^2 + \frac{\lambda}{\sigma^2}t_i(y_i-\mu) - \frac{1}{2\sigma^2}(\kappa^{-1}(u_i)+\lambda^2)(y_i-\mu)^2 + \log h(u_i;\boldsymbol{\tau})\right], \tag{13}$$

where K is a constant not depending on unknown parameters. Given the current estimate $\widehat{\boldsymbol{\theta}}^{(k)} = (\widehat{\mu}^{(k)}, \widehat{\sigma^2}^{(k)}, \widehat{\lambda}^{(k)}, \widehat{\boldsymbol{\tau}}^{(k)\top})^\top$, the E-step calculates the function

$$Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}) = \mathrm{E}[\ell_c(\boldsymbol{\theta}|\mathbf{y}_c)|\mathbf{y}, \widehat{\boldsymbol{\theta}}^{(k)}] = \sum_{i=1}^m Q_{1i}(\boldsymbol{\theta}_1, \widehat{\boldsymbol{\theta}}^{(k)}) + \sum_{i=1}^m Q_{2i}(\boldsymbol{\tau}, \widehat{\boldsymbol{\theta}}^{(k)}), \tag{14}$$

with $\boldsymbol{\theta}_1 = (\mu, \sigma^2, \lambda)^\top$, $Q_{2i}(\boldsymbol{\tau}, \widehat{\boldsymbol{\theta}}^{(k)}) = E[\log h(U_i;\boldsymbol{\tau})|\mathbf{y}, \widehat{\boldsymbol{\theta}}^{(k)}]$ and

$$Q_{1i}(\boldsymbol{\theta}_1, \widehat{\boldsymbol{\theta}}^{(k)}) = -\log\widehat{\sigma^2}^{(k)} - \frac{1}{2\widehat{\sigma^2}^{(k)}}\widehat{t^2}_i^{(k)} + \frac{\lambda^{(k)}}{\widehat{\sigma^2}^{(k)}}\widehat{t}_i^{(k)}(y_i - \mu^{(k)}) - \frac{1}{2\widehat{\sigma^2}^{(k)}}[\widehat{\kappa}_i^{(k)} + (\widehat{\lambda}^{(k)})^2](y_i-\mu^{(k)})^2,$$

which requires expressions for $\widehat{t}_i^{(k)} = \mathrm{E}[T_i|y_i, \widehat{\boldsymbol{\theta}}^{(k)}]$, $\widehat{t^2}_i^{(k)} = \mathrm{E}[T_i^2|y_i, \widehat{\boldsymbol{\theta}}^{(k)}]$ and $\widehat{\kappa}_i^{(k)} = \mathrm{E}[\kappa^{-1}(U_i)|y_i, \widehat{\boldsymbol{\theta}}^{(k)}]$, which can be readily evaluated by using

$$\widehat{t}_i^{(k)} = \widehat{\lambda}^{(k)}\widehat{\eta}_i^{(k)} + \widehat{\sigma}^{(k)}W_\Phi\left(\frac{\widehat{\lambda}^{(k)}\widehat{\eta}_i^{(k)}}{\widehat{\sigma}^{(k)}}\right), \tag{15}$$

$$\widehat{t^2}_i^{(k)} = [\widehat{\lambda}^{(k)}\widehat{\eta}_i^{(k)}]^2 + \widehat{\sigma^2}^{(k)} + \widehat{\lambda}^{(k)}\widehat{\sigma}^{(k)}\widehat{\eta}_i^{(k)}W_\Phi\left(\frac{\widehat{\lambda}^{(k)}\widehat{\eta}_i^{(k)}}{\widehat{\sigma}^{(k)}}\right), \tag{16}$$

where $W_\Phi(u) = \phi_1(u)/\Phi_1(u)$ and $\widehat{\eta}_i^{(k)} = y_i - \widehat{\mu}^{(k)}$, $i = 1, \ldots, m$. For the StN, SSL, SEP and SCN distributions of the SSMN class we have computationally attractive expressions for $\widehat{\kappa}_i^{(k)}$ and thus (15)–(16) can be easily implemented. The CM-step then conditionally maximizes $Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})$ with respect to $\boldsymbol{\theta}$, obtaining a new estimate $\widehat{\boldsymbol{\theta}}^{(k+1)}$, as described below:

CM-step: Update $\widehat{\mu}^{(k+1)}$, $\widehat{\lambda}^{(k+1)}$ and $\widehat{\sigma^2}^{(k+1)}$ as

$$\widehat{\mu}^{(k+1)} = \left[m[\widehat{\lambda}^{(k)}]^2 + \sum_{i=1}^m \widehat{\kappa}_i^{(k)}\right]^{-1} \sum_{i=1}^m \left[\widehat{\kappa}_i^{(k)}y_i - \widehat{\lambda}^{(k)}(\widehat{t}_i^{(k)} - \widehat{\lambda}^{(k)}y_i)\right], \tag{17}$$

$$\widehat{\lambda}^{(k+1)} = \frac{\sum_{i=1}^{m} \widehat{t}_i^{(k)} (y_i - \widehat{\mu}^{(k)})}{\sum_{i=1}^{m} (y_i - \widehat{\mu}^{(k)})^2},$$

$$\widehat{\sigma^2}^{(k+1)} = \frac{1}{2m} \sum_{i=1}^{m} \left[[\widehat{\kappa}_i^{(k)} + (\widehat{\lambda}^{(k)})^2](y_i - \widehat{\mu}^{(k)})^2 - 2\widehat{\lambda}^{(k)} \widehat{t}_i^{(k)} (y_i - \widehat{\mu}^{(k)}) + \widehat{t2}_i^{(k)} \right].$$

CML-step: Fix $\widehat{\mu}^{(k+1)}, \widehat{\lambda}^{(k+1)}$ and $\widehat{\sigma^2}^{(k+1)}$ and update $\widehat{\boldsymbol{\tau}}^{(k)}$ by optimizing the constrained log-likelihood function, i.e.,

$$\widehat{\boldsymbol{\tau}}^{(k+1)} = argmax_{\boldsymbol{\tau}} \sum_{i=1}^{m} \log f_0(y_i | \widehat{\mu}^{(k+1)}, \widehat{\sigma^2}^{(k+1)}, \boldsymbol{\tau}), \tag{18}$$

where $f_0(y)$ is the respective symmetric pdf as defined in (5). The more efficient CML-step follows [28] (ECME), and is referred to as the conditional marginal likelihood step (CML-step), where we replace the usual M-step by a step that maximizes the restricted actual log-likelihood function. Further, this step requires one-dimensional search on StN, SSL, and SEP models and a bi-dimensional search on the SCN model, which can be easily accomplished by using, for example, the "optim" routine in R software or "fmincon" in *Matlab* ®.

The iterations of the above algorithms are repeated until a suitable convergence rule is satisfied, e.g., $\|\boldsymbol{\theta}^{(k+1)} - \boldsymbol{\theta}^{(k)}\|$ is sufficiently small. Although the EM-type algorithm tends to be robust with respect to the choice of the starting values, it may not converge when initial values are far from good ones. Thus, the choice of adequate starting values for the EM and NR algorithms plays a big role in parameter estimation. A set of reasonable initial values can be achieve by computing $\widehat{\mu}^{(0)}, \widehat{\sigma}^{2(0)}$ and $\widehat{\lambda}^{(0)}$ using the method of moments of a SN random variable given, for instance, in [25].

5. Applications

5.1. Simulation study

In order to compare the behavior of EM estimates, in this section we present a simulation study based on 5000 data samples generated from the skew-Student-*t* (ST) model proposed by Azzalini and Capitanio [8] and the skew-*t* normal (StN) model considered in this paper. The true parameter values considered were: $\mu = 0, \sigma = 1, \lambda = 1$ and degrees of freedom $\nu = 4$ for both models. We use the notation $ST(1, 4)$ for Azzalini and Capitanio's skew-*t* model and $StN(1, 4)$ for the skew-*t* model. A similar study for the skew-normal distribution is reported in [2]. As can be discerned from Figs. 2 and 3, the dispersion diagram for the skew-*t* form is a bit bent to the right indicating a behavior somewhat similar to (but not as pronounced as) that for the skew-normal model reported in [2]. The behavior of the MLE for the StN model seems quite well behaved with a moderate (and expected) degree of correlation between the estimates. This is an indication that the StN model can be proved a viable alternative to the ordinary skew-*t* model.

5.2. The fiber-glass data set

In this section we apply five specific distributions of the SSMN class, namely, the univariate skew-normal (SN), skew-*t* (StN), skew-slash (SSL), skew-contaminated normal (SCN) and skew-exponential power (SEP) distributions, to fit the data on the breaking strength of 1.5 cm long glass fiber, consisting of 63 observations. Jones and Faddy [20] and Wang and Genton [35] had previously analyzed this data set with a skew-*t* and a skew-slash distribution, respectively. They both found strong evidence of skewness to the left as well as of heavy tail behavior of the data. In this section, we revisit the "fiber-glass" data, aiming at providing additional inferences by using SSMN distributions. In order to compare, we have also fitted the skew-*t* (ST) distribution proposed by Azzalini and Capitanio [8].

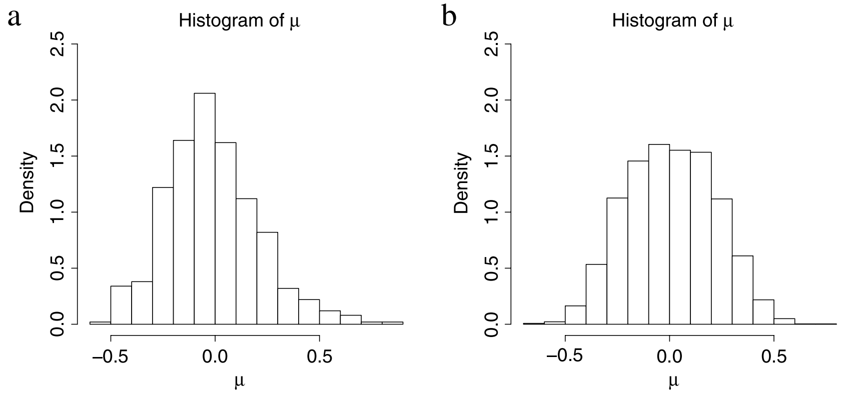

Fig. 2. Histogram of $\widehat{\mu}$ based on 5000 samples of size $n = 200$ generated from the (a) $ST(1, 4)$ distribution (Azzalini) and (b) $StN(1, 4)$ distribution.

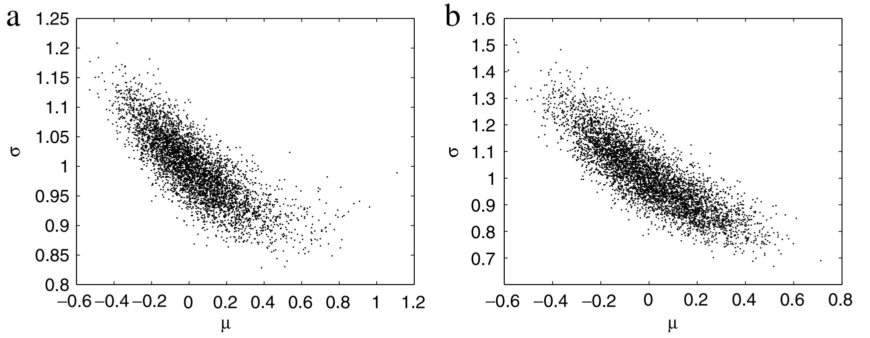

Fig. 3. Scatter plot of $(\widehat{\mu}, \widehat{\sigma})$ based on 5000 samples of size $n = 200$ generated from the (a) $ST(1, 4)$ distribution (Azzalini) and (b) $StN(1, 4)$ distribution.

We consider a location–scale model for variable fiber-glass, which is represented as

$$y_i \sim SSMN(\mu, \sigma^2, \lambda, H; \kappa), \quad i = 1, \ldots, 63.$$

Resulting parameter estimates for the six models are given in Table 1 together with their corresponding standard errors (in parentheses) calculated via the observed information matrix given in Section 3. The likelihood-ratio test $LR = -2(\ell(\widehat{\boldsymbol{\theta}}_{SSMN}) - \ell(\widehat{\boldsymbol{\theta}}_{SN}))$ indicates that the heavy tailed SSMN distributions (StN, SNC and SEP) and ST model present a better fit than the SN model, with the SCN distribution significantly better in terms of the LR test. We also note from Table 1 that the standard errors of the StN, SCN and SSL and SEP models are smaller than those under the SN model, indicating that the four models with thicker tails seem to produce more accurate maximum likelihood estimates.

Fig. 4 shows the histogram of the fiber data superimposed with the fitted curves of the SSMN densities. Note that the density curve for the SCN model appears to follow the behavior of the observed data more satisfactorily than the other models. Moreover, replacing the ML estimates of $\boldsymbol{\theta}$ in the Mahalanobis distance D we constructed the QQ-plots and envelopes [4] by simulation for the SN and SCN model as depicted in Fig. 5. The estimated envelope for the SCN model indicates no points outside the confidence band (Fig. 5(b)). Therefore, these plots enable us to claim that the SCN model provides a better fit to the data than the SN model.

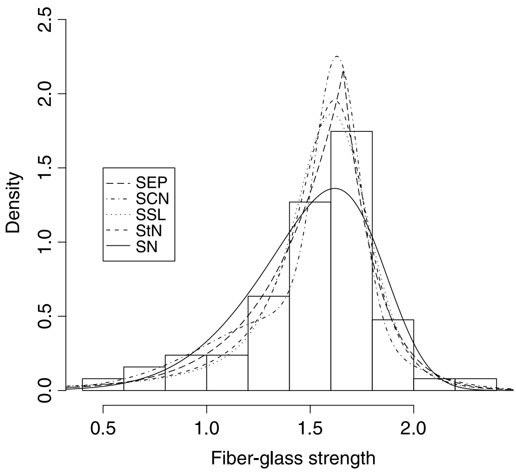

Fig. 4. Fiber-glass data set: histogram superimposed with the fitted density curves for the five SSMN distributions.

Table 1
MLEs for the six models fitted on the fiber-glass data set. The SE values, in parentheses, are estimated asymptotic standard errors. LR is the likelihood-ratio test for a SN model versus SMSN models.

Parameter	SN Estimate	ST Estimate	SSL Estimate	SEP Estimate	StN Estimate	SCN Estimate
μ	1.85(0.05)	1.75(0.06)	1.65(0.05)	1.66(2.9×10^{-6})	1.65(0.05)	1.65(0.04)
σ^2	0.22(0.05)	0.26(0.08)	0.02(0.01)	0.02(0.03)	0.03(0.02)	0.01(0.01)
λ	−2.68(0.80)	−1.55(0.57)	−0.27(0.16)	−0.31(0.22)	−0.36(0.24)	−0.21(0.17)
ν	–	0.76(0.05)	2.73(1.41)	0.55(0.17)	1.96(0.82)	0.51(0.07)
γ	–	–	–	–	–	0.05(0.01)
$\ell(\widehat{\boldsymbol{\theta}})$	−13.96	−11.70	−12.83	−11.05	−11.79	−8.86
LR	–	4.52	2.26	5.81	4.34	10.19
p-value	–	0.03	0.13	0.02	0.04	0.01

6. Discussion and future work

In this work we have defined a new family of asymmetric models by extending the symmetric class of scale mixtures of normal distributions. Our proposal generalized in some directions results found in [5,1,24]. One interesting and simplifying aspect of the defined class is that the implementation of the EM algorithm is facilitated by the fact that the E-step is exactly as in the scale mixtures of normal distribution class of models proposed in [1]. Also, the M-step involves closed form expressions facilitating the implementation of the algorithm. The observed information matrix is derived analytically, which allows direct implementation of inference on this class of models. Matlab programs are available from the third author's homepage with website address http://www.ime.unicamp.br/~hlachos/ListaPub.html.

We believe that the approaches proposed here can be extended in several directions. For example, some authors have already considered modeling nonlinear models with the StN distribution (see [26]). Thus, it is of interest to explore the SSMN class of distributions in the context of nonlinear models. Another natural extension of this work is to study the multivariate version of the SSMN class of distributions. More specifically, we say that a p-dimensional random vector $\mathbf{Y}$ follows a multivariate skew scale mixture of normal distribution with location parameter $\boldsymbol{\mu} \in \mathbb{R}^p$, a positive definite scale

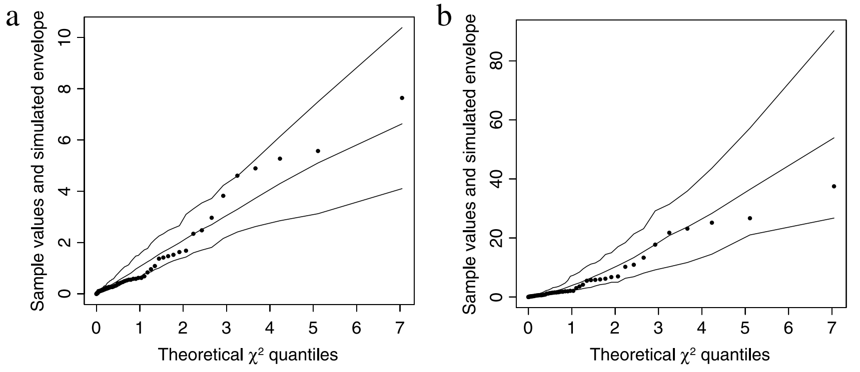

Fig. 5. Simulated envelopes based on the Mahalanobis distance for the fiber-glass data set. (a) Skew-normal model. (b) Skew-contaminated normal model.

matrix $\mathbf{\Sigma}(p \times p)$ and skewness parameter $\boldsymbol{\lambda} \in \mathbb{R}^p$ if its pdf is given by

$$f(\mathbf{y}) = 2f_0(\mathbf{y})\Phi\left(\boldsymbol{\lambda}^\top \mathbf{\Sigma}^{-1/2}(\mathbf{y}-\boldsymbol{\mu})\right), \quad \mathbf{y} \in \mathbb{R}^p,$$

where $f_0(\cdot)$ is the pdf of a multivariate SMN distribution (see [24]). Indeed, multivariate skew models have been used as a powerful tool for extending some traditional normal based models; see [22], for example.

Acknowledgements

The authors wish to thank the associate editor and one referee for constructive comments. The third author acknowledges partial financial support from Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) and CNPq. The research of Heleno Bolfarine was supported in part by CNPq.

Appendix A. Some important results related to SSMN models

Proposition 4. *Let $Y \sim SSMN(\mu, \sigma^2, \lambda, H; \kappa)$. Then the moment-generating function ($M_Y(t)$) is given by*

$$M_Y(t) = E[\mathrm{e}^{tY}] = \int_0^\infty 2\mathrm{e}^{t\mu+\frac{t^2}{2}\kappa(u)\sigma^2}\Phi\left(\frac{\sigma\lambda\kappa(u)}{\sqrt{1+\lambda^2\kappa(u)}}t\right)\mathrm{d}H(u), \quad t \in \mathbb{R}.$$

Proof. From Proposition 1, we have that $y|U = u \sim SN(\mu, \sigma^2\kappa(u), \lambda\sqrt{\kappa(u)})$. Moreover, from well known properties of conditional expectations, it follows that $M_Y(t) = E_U[E[\mathrm{e}^{tY}|U]]$. From Azzalini [6], $M_Z(z) = 2\exp\left(\frac{z^2}{2}\right) \times \Phi_1\left(\frac{\lambda z}{\sqrt{1+\lambda^2}}\right)$, where $Z \sim SN(\lambda)$. Hence, provided $M_{a+bY}(t) = \mathrm{e}^{at}M_Y(bt)$, we obtain

$$M_{Y|U=u}(t) = 2\mathrm{e}^{t\mu+\frac{t^2}{2}\kappa(u)\sigma^2}\Phi\left(\frac{\sigma\lambda\kappa(u)}{\sqrt{1+\lambda^2\kappa(u)}}t\right). \quad \square$$

From Proposition 4 we have the following corollary:

Corollary 1. *Suppose that $Y \sim SSMN(\mu, \sigma^2, \lambda, H; \kappa)$. Then, for $b = \sqrt{\frac{2}{\pi}}$:*

(a) $E[Y] = \mu + b\sigma\lambda E_U\left[\frac{\kappa(U)}{\sqrt{1+\lambda^2\kappa(U)}}\right]$,

(b) $\mathrm{Var}[Y] = \sigma^2 \left(E_U[\kappa(U)] - b^2\lambda^2 E_U^2 \left[\frac{\kappa(U)}{\sqrt{1+\lambda^2\kappa(U)}} \right] \right)$.

Proposition 5. *If $Y \sim SSMN(\mu, \sigma^2, \lambda, H; \kappa)$, then for any even function τ, the distribution of $\tau(Y-\mu)$ does not depend on λ and has the same distribution as $\tau(X-\mu)$, where $X \sim SMN(\mu, \sigma^2, H; \kappa)$. In particular, $(Y-\mu)^2$ and $(X-\mu)^2$ are identically distributed.*

Proof. This proposition is a particular case of a result found in [34].

The following corollary is interesting because it allows us to check models in practice. □

Corollary 2. *Let $Y \sim SSMN(\mu, \sigma^2, \lambda, H; \kappa)$. Then the quadratic form*

$$D_\lambda = \frac{(Y-\mu)^2}{\sigma^2}$$

has the same distribution as $D = \frac{(X-\mu)^2}{\sigma^2}$, where $X \sim SMN(\mu, \sigma^2 H; \kappa)$.

Appendix B. Derivation of the information matrix $\mathbf{I_1}(\theta)$ (symmetric portion)

B.1. The Student-t distribution

$$I^1_{\mu\mu} = -\frac{\nu+1}{\nu\sigma^2}\sum_{i=1}^m V_i\left(\frac{2}{\nu}V_i d_i - 1\right),$$

$$I^1_{\sigma^2\mu} = -\frac{\nu+1}{\nu\sigma^4}\sum_{i=1}^m V_i(y_i-\mu)\left(\frac{1}{\nu}V_i d_i - 1\right),$$

$$I^1_{\nu\mu} = -\frac{1}{2}\sum_{i=1}^m V_i(y_i-\mu)\left(1+\frac{\nu+1}{\nu^2}V_i d_i\right),$$

$$I^1_{\sigma^2\sigma^2} = -\frac{m}{2\sigma^4} + \frac{\nu+1}{\nu\sigma^4}\sum_{i=1}^m V_i d_i\left(1-\frac{1}{2\nu}V_i d_i\right),$$

$$I^1_{\nu\sigma^2} = -\frac{1}{2\nu^2\sigma^2}\sum_{i=1}^m V_i d_i\left(\frac{\nu+1}{\nu}V_i d_i - 1\right),$$

$$\begin{aligned} I^1_{\nu\nu} = & -\frac{m}{4}\left[\Psi'\left(\frac{\nu+1}{2}\right) - \Psi'\left(\frac{\nu}{2}\right) + \frac{2}{\nu^2}\right] - \frac{1}{2\nu^2}\sum_{i=1}^m V_i d_i, \\ & -\frac{1}{2\nu^4}\sum_{i=1}^m V_i d_i\left[(\nu+1)V_i d_i - \nu(\nu+2)\right], \end{aligned}$$

$$I^1_{\lambda\boldsymbol{\theta}} = I^1_{\gamma\boldsymbol{\theta}} = 0,$$

where $V_i = \left(1+\frac{d_i}{\nu}\right)^{-1}$, $d_i = \frac{(y_i-\mu)^2}{\sigma^2}$ and $\Psi(x) = \frac{\Gamma'(x)}{\Gamma(x)}$ is the digamma function. As noted in [15], the information matrix for the skew-t normal model is not singular for finite degrees of freedom when $\lambda = 0$.

B.2. The slash distribution

$$\begin{aligned}
I^1_{\mu\mu} &= -(2\nu+1)\sum_{i=1}^{m}(y_i-\mu)^{-2}P_1(\nu+1/2,d_i/2)^{-1}[(2\nu+3)P_1(\nu+5/2,d_i/2)\\
&\quad -P_1(\nu+3/2,d_i/2)-(2\nu+1)P_1(\nu+3/2,d_i/2)^2P_1(\nu+1/2,d_i/2)^{-1}],
\end{aligned}$$

$$\begin{aligned}
I^1_{\sigma^2\mu} &= -\frac{(2\nu+1)}{2\sigma^2}\sum_{i=1}^{m}(y_i-\mu)^{-1}P_1(\nu+1/2,d_i/2)^{-1}[(2\nu+3)P_1(\nu+5/2,d_i/2)\\
&\quad -2P_1(\nu+3/2,d_i/2)-(2\nu+1)P_1(\nu+3/2,d_i/2)^2P_1(\nu+1/2,d_i/2)^{-1}],
\end{aligned}$$

$$\begin{aligned}
I^1_{\nu\mu} &= -2\sum_{i=1}^{m}\frac{1}{(y_i-\mu)P_1(\nu+1/2,d_i/2)}[P_1(\nu+3/2,d_i/2)+(\nu+1/2)P_1^{(1)}(\nu+3/2,d_i/2)\\
&\quad -(\nu+1/2)P_1(\nu+3/2,d_i/2)P_1^{(1)}(\nu+1/2,d_i/2)P_1(\nu+1/2,d_i/2)^{-1}],
\end{aligned}$$

$$\begin{aligned}
I^1_{\sigma^2\sigma^2} &= -\frac{m}{2\sigma^4}-\frac{(2\nu+1)}{4\sigma^4}\sum_{i=1}^{m}P_1(\nu+1/2,d_i/2)^{-1}[(2\nu+3)P_1(\nu+5/2,d_i/2)\\
&\quad -4P_1(\nu+3/2,d_i/2)-(2\nu+1)P_1(\nu+3/2,d_i/2)^2P_1(\nu+1/2,d_i/2)^{-1}],
\end{aligned}$$

$$\begin{aligned}
I^1_{\nu\sigma^2} &= -\frac{1}{\sigma^2}\sum_{i=1}^{m}\frac{1}{P_1(\nu+1/2,d_i/2)}\Bigg[P_1(\nu+3/2,d_i/2)\\
&\quad +(\nu+1/2)P_1^{(1)}(\nu+3/2,d_i/2)-\frac{P_1(\nu+3/2,d_i/2)P_1^{(1)}(\nu+1/2,d_i/2)}{P_1(\nu+1/2,d_i/2)}\Bigg],
\end{aligned}$$

$$I^1_{\nu\nu}=\frac{m}{\nu^2}-m\Psi'(\nu+1/2)-\frac{P_1^{(2)}(\nu+1/2,d_i/2)P_1(\nu+1/2,d_i/2)-P_1^{(1)}(\nu+1/2,d_i/2)^2}{P_1(\nu+1/2,d_i/2)^2},$$

$$I^1_{\lambda\boldsymbol{\theta}}=I^1_{\gamma\boldsymbol{\theta}}=0,$$

with $P_1^{(1)}(a,b)=[\log(b)-\Psi(a)]P_1(a,b)+\frac{\gamma(a,b)}{\Gamma(a)}E_U[\log(U)]$, and $P_1^{(2)}(a,b)=-\Psi'(a)P_1(a,b)+[\log(b)-\Psi(a)]\left[P_1^{(1)}(a,b)+\frac{\gamma(a,b)}{\Gamma(a)}E_U[\log(U)]\right]+\frac{\gamma(a,b)}{\Gamma(a)}E_U[\log(U)^2]$, $U\sim TG(a,b,1)$. The expected values above can be obtained via Monte Carlo methods, with an algorithm for generation of the truncated gamma distribution given by Philippe [32].

B.3. The contaminated normal distribution

$$\begin{aligned}
I^1_{\mu\mu} &= \frac{1}{\sqrt{2\pi}\sigma^3}\sum_{i=1}^{m}\frac{1}{f_s(y_i|\boldsymbol{\theta})}\Bigg[-d_i\left(\nu\gamma^{5/2}\mathrm{e}^{-\gamma d_i/2}+(1-\nu)\mathrm{e}^{-d_i/2}\right)\\
&\quad +\frac{f_s(y_i|\boldsymbol{\theta})+(y_i-\mu)\frac{\partial f_s(y_i|\boldsymbol{\theta})}{\partial\mu}}{f_s(y_i|\boldsymbol{\theta})}\left(\nu\gamma^{3/2}\mathrm{e}^{-\gamma d_i/2}+(1-\nu)\mathrm{e}^{-d_i/2}\right)\Bigg],
\end{aligned}$$

$$\begin{aligned}
I^1_{\sigma^2\mu} &= \frac{1}{2\sqrt{2\pi}\sigma^5}\sum_{i=1}^{m}\frac{(y_i-\mu)}{f_s(y_i|\boldsymbol{\theta})}\Bigg[-d_i\left(\nu\gamma^{5/2}\mathrm{e}^{-\gamma d_i/2}+(1-\nu)\mathrm{e}^{-d_i/2}\right)\\
&\quad +\frac{1}{f_s(y_i|\boldsymbol{\theta})}\left(3f_s(y_i|\boldsymbol{\theta})+2\sigma^2\frac{\partial f_s(y_i|\boldsymbol{\theta})}{\partial\sigma^2}\right)\left(\nu\gamma^{3/2}\mathrm{e}^{-\gamma d_i/2}+(1-\nu)\mathrm{e}^{-d_i/2}\right)\Bigg],
\end{aligned}$$

$$I^1_{\nu\mu} = \frac{1}{\sqrt{2\pi}\sigma^3}\sum_{i=1}^{m}\frac{(y_i-\mu)}{f_s(y_i|\boldsymbol{\theta})}\Bigg[-\gamma^{3/2}\mathrm{e}^{-\gamma d_i/2}+\mathrm{e}^{-d_i/2} + \frac{1}{f_s(y_i|\boldsymbol{\theta})}\frac{\partial f_s(y_i|\boldsymbol{\theta})}{\partial\nu}\left(\nu\gamma^{3/2}\mathrm{e}^{-\gamma d_i/2}+(1-\nu)\mathrm{e}^{-d_i/2}\right)\Bigg],$$

$$I^1_{\gamma\mu} = \frac{1}{\sqrt{2\pi}\sigma^3}\sum_{i=1}^{m}\frac{(y_i-\mu)}{f_s(y_i|\boldsymbol{\theta})}\Bigg[-\frac{\nu\gamma^{1/2}}{2}\mathrm{e}^{-\gamma d_i/2}(3-\gamma d_i) + \frac{1}{f_s(y_i|\boldsymbol{\theta})}\frac{\partial f_s(y_i|\boldsymbol{\theta})}{\partial\gamma}\left(\nu\gamma^{3/2}\mathrm{e}^{-\gamma d_i/2}+(1-\nu)\mathrm{e}^{-d_i/2}\right)\Bigg],$$

$$I^1_{\sigma^2\sigma^2} = \frac{1}{4\sqrt{2\pi}\sigma^5}\sum_{i=1}^{m}\frac{1}{f_s(y_i|\boldsymbol{\theta})}\Bigg\{-\nu\gamma^{3/2}d_i\mathrm{e}^{-\gamma d_i/2}(\gamma d_i/2-2)-(1-\nu)d_i\mathrm{e}^{-d_i/2}(d_i/2-2) + \frac{1}{f_s(y_i|\boldsymbol{\theta})}\left[3f_s(y_i|\boldsymbol{\theta})+2\sigma^2\frac{\partial f_s(y_i|\boldsymbol{\theta})}{\partial\sigma^2}\right] \times\left[\nu\gamma^{1/2}\mathrm{e}^{-\gamma d_i/2}(\gamma d_i/2-1)+(1-\nu)\mathrm{e}^{-d_i/2}(d_i/2-1)\right]\Bigg\},$$

$$I^1_{\nu\sigma^2} = \frac{1}{2\sqrt{2\pi}\sigma^3}\sum_{i=1}^{m}\frac{1}{f_s(y_i|\boldsymbol{\theta})}\Bigg\{-\gamma^{1/2}\mathrm{e}^{-\gamma d_i/2}\,(\gamma d_i/2-1)+\mathrm{e}^{-d_i/2}\,(d_i/2-1) + \frac{1}{f_s(y_i|\boldsymbol{\theta})}\frac{\partial f_s(y_i|\boldsymbol{\theta})}{\partial\nu}\left[\nu\gamma^{1/2}\mathrm{e}^{-\gamma d_i/2}\,(\gamma d_i/2-1)+(1-\nu)\mathrm{e}^{-d_i/2}\,(d_i/2-1)\right]\Bigg\},$$

$$I^1_{\gamma\sigma^2} = \frac{1}{2\sqrt{2\pi}\sigma^3}\sum_{i=1}^{m}\frac{1}{f_s(y_i|\boldsymbol{\theta})}\Bigg\{-\frac{\nu}{4}\mathrm{e}^{-\gamma d_i/2}\left(5\gamma^{1/2}\sigma^2 d_i-\gamma^{3/2}d_i^2-2\gamma^{-1/2}\right) + \frac{1}{f_s(y_i|\boldsymbol{\theta})}\frac{\partial f_s(y_i|\boldsymbol{\theta})}{\partial\gamma}\left[\nu\gamma^{1/2}\mathrm{e}^{-\gamma d_i/2}\,(\gamma d_i/2-1)+(1-\nu)\mathrm{e}^{-d_i/2}\,(d_i/2-1)\right]\Bigg\},$$

$$I^1_{\nu\nu} = \frac{1}{\sqrt{2\pi}\sigma}\sum_{i=1}^{m}\frac{1}{f_s(y_i|\boldsymbol{\theta})^2}\frac{\partial f_s(y_i|\boldsymbol{\theta})}{\partial\nu}\left(\gamma^{1/2}\mathrm{e}^{-\gamma d_i/2}-\mathrm{e}^{-d_i/2}\right),$$

$$I^1_{\gamma\nu} = -\frac{1}{2\sqrt{2\pi}\sigma\gamma^{1/2}}\sum_{i=1}^{m}\frac{\mathrm{e}^{-\gamma d_i/2}}{f_s(y_i|\boldsymbol{\theta})}(1-\gamma d_i)\left(1-\frac{\nu}{f_s(y_i|\boldsymbol{\theta})}\frac{\partial f_s(y_i|\boldsymbol{\theta})}{\partial\nu}\right),$$

$$I^1_{\gamma\gamma} = \frac{\nu}{2\sqrt{2\pi}\sigma}\sum_{i=1}^{m}\frac{\mathrm{e}^{-\gamma d_i/2}}{f_s(y_i|\boldsymbol{\theta})}\left[d_i+\frac{(1-\gamma d_i)}{2\gamma^{1/2}f_s(y_i|\boldsymbol{\theta})}\left(d_i f_s(y_i|\boldsymbol{\theta})+\frac{f_s(y_i|\boldsymbol{\theta})}{\gamma}+2\frac{\partial f_s(y_i|\boldsymbol{\theta})}{\partial\gamma}\right)\right],$$

$$I^1_{\lambda\boldsymbol{\theta}} = 0,$$

where $f_s(y|\boldsymbol{\theta}) = \nu\phi(y|\mu,\sigma^2/\gamma)+(1-\nu)\phi(y|\mu,\sigma^2)$,
$\frac{\partial f_s(y_i|\boldsymbol{\theta})}{\partial\mu} = \frac{(y_i-\mu)}{\sqrt{2\pi}\sigma^3 f_s(y_i|\boldsymbol{\theta})}\left[\nu\gamma^{3/2}\mathrm{e}^{-\gamma d_i/2}+(1-\nu)\mathrm{e}^{-d_i/2}\right]$,
$\frac{\partial f_s(y_i|\boldsymbol{\theta})}{\partial\sigma^2} = \frac{1}{2\sqrt{2\pi}\sigma^3 f_s(y_i|\boldsymbol{\theta})}\left[\nu\gamma^{1/2}\mathrm{e}^{-\gamma d_i/2}(\gamma d_i/2-1)+(1-\nu)\mathrm{e}^{-d_i/2}(d_i/2-1)\right]$,
$\frac{\partial f_s(y_i|\boldsymbol{\theta})}{\partial\nu} = \frac{1}{\sqrt{2\pi}\sigma f_s(y_i|\boldsymbol{\theta})}(\gamma^{1/2}\mathrm{e}^{-\gamma d_i/2}-\mathrm{e}^{-d_i/2})$ and $\frac{\partial f_s(y_i|\boldsymbol{\theta})}{\partial\gamma} = \frac{\nu\mathrm{e}^{-\gamma d_i/2}}{2\sqrt{2\pi}\sigma\gamma^{1/2}f_s(y_i|\boldsymbol{\theta})}(1-\gamma d_i)$.

B.4. The exponential power distribution

$$I^1_{\mu\mu} = \frac{\nu(2\nu-1)}{\sigma^{2\nu}} \sum_{i=1}^{m} |y_i-\mu|^{2\nu-2},$$

$$I^1_{\sigma^2\mu} = \frac{\nu^2}{\sigma^{2(\nu+1)}} \sum_{i=1}^{m} \operatorname{sign}(y_i-\mu)|y_i-\mu|^{2\nu-1},$$

$$I^1_{\nu\mu} = -\frac{1}{\sigma^{2\nu}} \sum_{i=1}^{m} \operatorname{sign}(y_i-\mu)|y_i-\mu|^{2\nu-1}\left[1-\nu\log\sigma^2+2\nu\log|y_i-\mu|\right],$$

$$I^1_{\sigma^2\sigma^2} = -\frac{m}{2\sigma^4} + \frac{\nu(\nu+1)}{2\sigma^{2(\nu+2)}} \sum_{i=1}^{m} |y_i-\mu|^{2\nu},$$

$$I^1_{\nu\sigma^2} = -\frac{1}{2\sigma^{2(\nu+1)}} \sum_{i=1}^{m} |y_i-\mu|^{2\nu}\left[1-\nu\log(\sigma^2)+2\nu\log|y_i-\mu|\right],$$

$$I^1_{\nu\nu} = \frac{m}{\nu^2} + \frac{m\log 2}{\nu^3} + \frac{m[\Psi'(1/2\nu)+4\nu\Psi(1/2\nu)]}{4\nu^4} + \frac{1}{2}\sum_{i=1}^{m}(\log d_i)^2\left|\frac{y_i-\mu}{\sigma}\right|^{2\nu},$$

$$I^1_{\lambda\boldsymbol{\theta}} = I^1_{\gamma\boldsymbol{\theta}} = 0,$$

$d_i = (y_i-\mu)^2/\sigma^2$. The expressions involving $\log(y_i-\mu)$ could pose problems if $\mu = y_i$. However, $\lim_{a\to^+0} a^{2k}\log(a^2) = 0$, by L'Hospital's Rule.

References

[1] D.R. Andrews, C.L. Mallows, Scale mixtures of normal distributions, Journal of the Royal Statistical Society, Series B 36 (1974) 99–102.
[2] R.B. Arellano-Valle, A. Azzalini, The centred parametrization for the multivariate skew-normal distribution, Journal of Multivariate Analysis 99 (2008) 1362–1382.
[3] B.C. Arnold, R.J. Beaver, The skew-Cauchy distribution, Statistics and Probability Letters 49 (2000) 285–290.
[4] A.C. Atkinson, Plots, Transformations, and Regression, Oxford, Clarendon, 1985.
[5] A. Azzalini, A class of distributions which includes the normal ones, Scandinavian Journal of Statistics 12 (1985) 171–178.
[6] A. Azzalini, Further results on a class of distributions which includes the normal ones, Statistica 46 (1986) 199–208.
[7] A. Azzalini, A. Capitanio, Statistical applications of the multivariate skew normal distribution, Journal of the Royal Statistical Society, Series B 61 (1999) 579–602.
[8] A. Azzalini, A. Capitanio, Distributions generated by perturbation of symmetry with emphasis on a multivariate skew-*t* distribution, Journal of the Royal Statistical Society, Series B 65 (2003) 367–389.
[9] H. Bolfarine, V.H. Lachos, Skew probit error-in-variables models, Statistical Methodology 3 (2007) 1–12.
[10] M.D. Branco, K.D. Dey, A general class of multivariate skew-elliptical distribution, Journal of Multivariate Analysis 79 (2001) 93–113.
[11] S.T.B. Choy, A.F.M. Smith, On robust analysis of a normal location parameter, Journal of the Royal Statistical Society, B 58 (1997) 463–474.
[12] T. Di Ciccio, A.C. Monti, Inferential aspects of the exponential power distribution, Journal of the American Statistical Association 99 (2004) 439–450.
[13] M. Galea-Rojas, M. de Castillho, H. Bolfarine, M. de Castro, Detection of analytical bias, Analyst 128 (2003) 1073–1081.
[14] M.G. Genton, N. Loperfido, Generalized skew-elliptical distributions and their quadratic forms, Annals of the Institute of Statistical Mathematics 57 (2005) 389–401.
[15] H.W. Gómez, O. Venegas, H. Bolfarine, Skew-symmetric distributions generated by the normal distribution function, Environmetrics 18 (2007) 395–407.
[16] M.A. Gómez-Villegas, E. Gómez-Sánchez-Manzano, J.M. Marín, Multivariate exponential power distributions as mixtures of normal distributions with Bayesian applications, Communications in Statistics—Theory and Methods 37 (6) (2008) 972–985.
[17] A.K. Gupta, F.C. Chang, W.J. Huang, Some skew-symmetric models, Random Operators and Stochastic Equations 10 (2002) 133–140.
[18] N. Henze, A probabilistic representation of the "skew-normal" distribution, Scandinavian Journal of Statistics 13 (1986) 271–275.

[19] N.L. Johnson, S. Kotz, N. Balakrishnan, Continuous Univariate Distributions, vol. 1, John Wiley, New York, 1994.
[20] M.C. Jones, M.J. Faddy, A skew extension of the t-distribution, with applications, Journal of the Royal Statistical Society, Series B 65 (2003) 159–174.
[21] Y. Kano, Consistency property of elliptical probability density functions, Journal of Multivariate Analysis 51 (1994) 139–147.
[22] V.H. Lachos, P. Ghosh, R.B. Arellano-Valle, Likelihood based inference for skew-normal/independent linear mixed models, Statistica Sinica 20 (2010) 303–322.
[23] K.L. Lange, J.A. Little, M.G.J. Taylor, Robust modeling using the t distribution, Journal of the American Statistical Association 84 (1989) 881–896.
[24] K. Lange, J.S. Sinsheimer, Normal/independent distributions and their applications in robust regression, Journal of Computational and Graphical Statistics 2 (1993) 175–198.
[25] T.I. Lin, J.C. Lee, S.Y. Yen, Finite mixture modelling using the skew-normal distribution, Statistica Sinica 17 (2007) 909–927.
[26] J.G. Lin, F.C. Xie, B.C. Wei, Statistical diagnostics for skew-t-normal nonlinear models, Communications in Statistics—Simulation and Computation 38 (2009) 2096–2110.
[27] R.J.A. Little, Robust estimation of the mean and covariance matrix from data with missing values, Applied Statistics 37 (1988) 23–38.
[28] C.H. Liu, D.B. Rubin, The ECME algorithm: a simple extension of EM and ECM with faster monotone convergence, Biometrika 81 (1994) 633–648.
[29] X.L. Meng, D.B. Rubin, Maximum likelihood estimation via the ECM algorithm: a general framework, Biometrika 80 (1993) 267–278.
[30] S. Nadarajah, S. Kotz, Skewed distributions generated by the normal kernel, Statistics and Probability Letters 65 (2003) 269–277.
[31] A. Pewsey, Problems of inference for Azzalini's skew-normal distribution, Journal of Applied Statistics 27 (2000) 859–870.
[32] A. Philippe, Simulation of right and left truncated gamma distributions by mixtures, Statistics and Computing 7 (1997) 173–181.
[33] S.K. Sahu, D.K. Dey, M.D. Branco, A new class of multivariate skew distributions with applications to Bayesian regression models, The Canadian Journal of Statistics 31 (2003) 129–150.
[34] J. Wang, J. Boyer, M.G. Genton, A skew-symmetric representation of multivariate distributions, Statistica Sinica 14 (2004) 1259–1270.
[35] J. Wang, M. Genton, The multivariate skew-slash distribution, Journal of Statistical Planning and Inference 136 (2006) 209–220.
[36] M. West, On scale mixtures of normal distributions, Biometrika 74 (1987) 646–648.