SKEW SCALE MIXTURE OF NORMAL DISTRIBUTIONS: PROPERTIES AND ESTIMATION

Clécio da Silva Ferreira, Heleno Bolfarine[†]

Departamento de Estatística, Universidade de São Paulo, São Paulo, SP, Brazil

Víctor H. Lachos

Departamento de Estatística, Universidade Estadual de Campinas, Campinas, SP, Brazil

Abstract

Scale mixture of normal distributions are often used as a challenging family for statistical procedures of 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 informating matrix, which is derived in closed form for some distributions in the family. 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.

Key Words: *scale mixture of normal distributions; skewness; EM-algorithm.*

1 Introduction

The scale mixture of normal distributions (Andrews and Mallows, 1974) provide a group of thick-tailed distributions that are often used for robust inference of symmetrical data. The theory and application (through simulation or experimentation) often generate a great amount of data sets that are skewed or heavy-tailed as, for instance, the data on family income (Azzalini et al., 2003) or substance concentration (Galea-Rojas et al., 2003 and Lachos and Bolfarine, 2007). Thus, we need

[†] *Address for correspondence:* Departamento de Estatística, IME, Universidade de São Paulo, Caixa Postal 66281 - CEP 05315 970, São Paulo-SP-Brazil. E-mail: hbolfar@ime.usp.br.

appropriate distributions to fit and simulate these skewed or heavy-tailed data. Candidate distributions at our disposal for fitting and simulating these data are not very abundant in the literature. In this article, we propose a new family of distributions that combine skewness with heavy tails. Moreover, this distribution is attractive because it has a stochastic representation that allows easy implementation of the EM-algorithm and it also facilitates the study of many of its properties. Our proposal generalizes recent results found mainly in Lange and Sinsheimer (1993) (see also Andrews and Mallows, 1974).

A simpler departure from the normal distribution which allows defining the univariate skew-normal distribution with probability density function (pdf) given by

$$f(y) = 2\phi(y|\mu,\sigma^2)\Phi_1\left(\frac{\lambda(y-\mu)}{\sigma}\right), \quad y \in \mathbb{R}. \tag{1}$$

was proposed by Azzalini (1985), where $\phi(.|\mu,\sigma^2)$ stands for the probability density function (pdf) of the normal distribution with mean μ and variance σ^2, $\Phi_1(.)$ represents the cumulative distribution function (cdf) of the standard normal distribution. It is well known that asymmetry range for this distribution is $(-0.995, 0.995)$. An extension to the multivariate setting was proposed by Arellano–Valle, Bolfarine and Lachos (2005) (see also Azzalini and Dalla–Valle, 1996). 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 $\mathrm{SN}(\mu,\sigma^2,\lambda)$. Its marginal stochastic representation (Henze, 1986), which can be used to derive several of its properties, is given by

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

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

$$E[Y] = \mu + \sqrt{\frac{2}{\pi}}\sigma\delta, \tag{3}$$

$$Var[Y] = \sigma^2(1 - \frac{2}{\pi}\delta^2). \tag{4}$$

Reasoning as Azzalini (1985) and Azzalini and Capitanio (2003), it is natural to construct univariate and multivariate distributions that combine skewness with heavy tails. For instance, one can define skew-t distributions (Sahu, Dey and Branco, 2003), skew-Cauchy distributions (Arnold and Beaver, 2000), skew-slash distributions (Wang and Genton, 2006), skew-slash-t distributions (Tan and Peng, 2006), skew-elliptical distributions (Azzalini and Capitanio, 1999; Branco and Dey, 2001; Sahu, Dey and Branco, 2003; Genton and Loperfido, 2005). Differently from the ideas above, in this article, we define 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 mixture of normal distributions (Andrews and Mallows, 1974; Lange and Sinsheimer, 1993). 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 normal/independent (NI) family of models proposed in Lange and Sinsheimer (1993). Besides, the M-step involves closed form expressions facilitating the implementation of the algorithm. Furthermore, we also concluded that the information matrix has a common part for all elements in the family, which makes, apart from the ordinary skew normal, the family having nonsingular information matrix.

The paper is organized as follows. In Section 2, for the sake of completeness, we give a brief sketch of the scale mixture of normal (SMN) distributions . In Section 3, the skew scale mixture of normal distribution (SSMN) are defined by extending the SMN class. Properties like moments, linear transformation and stochastic representation of the proposed distributions are also discussed. In Section 4, an EM-type algorithm which presents advantages over the direct maximization approach is presented, 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. Concluding remarks are given in Section 6.

2 Scale mixture of normal distributions

The symmetrical class of SMN distributions has attracted attention in the last few years, particularly due to the fact that they include distributions such as the Student-t, the slash, the power exponential, the contaminated normal, among others. All of these distributions have heavier tails than the normal ones. We say that a random variable Y has a SMN distribution with location parameter $\mu \in \mathbb{R}$ and a positive definite scale parameter σ^2 if its density function assumes the form

$$f(y) = \int_0^\infty \phi(y|\mu, \kappa(u)\sigma^2) dH(u), \tag{5}$$

where $H(.;\boldsymbol{\nu})$ is a cdf of a positive random variable U indexed by the parameter vector $\boldsymbol{\nu}$. For a random variable with a pdf as in (5), we shall use the notation $Y \sim \text{SMN}(\mu, \sigma^2; H)$. Moreover, when $\mu = 0$ and $\sigma^2 = 1$, we denote $y \sim \text{SMN}(H)$. Its stochastic representation is given by

$$Y = \mu + \kappa^{1/2}(U)Z, \tag{6}$$

where $Z \sim \text{N}(0, \sigma^2)$ and U is a positive random variable with cdf H independent of Z. Some examples of SMN distributions are described subsequently (see Lange and Sinsheimer, 1993). For this family, the distributional properties of the Mahalanobis distance

$$d = \frac{(y-\mu)^2}{\sigma^2},$$

are described, because they are extremely useful in testing goodness of fit and detecting the presence of outliers.

2.1 Examples of SMN distributions

- *The Generalized t-Student distribution with $\nu > 0$ degrees of freedom, $Y \sim Gt(\mu, \sigma^2; \nu, \gamma)$.*
  The use of the t-distribution as an alternative to the normal distribution has frequently been suggested in the literature. For instance, Little (1988) and Lange, Little and Taylor (1989) use the Student-t distribution for robust modeling. Y has a density function given by

$$f(y) = \frac{1}{\sigma\sqrt{\gamma\pi}} \frac{\Gamma((\nu+1)/2)}{\Gamma(\frac{\nu}{2})} \left(1 + \frac{d}{\gamma}\right)^{-(\frac{\nu+1}{2})}. \tag{7}$$

  In this case, $\kappa(U) = 1/U$ and $U \sim Gamma(\nu/2, \gamma/2)$, with density

$$h(u; \nu, \gamma) = \frac{(\gamma/2)^{\nu/2}}{\Gamma(\nu/2)} u^{\nu/2-1} e^{-\gamma u/2}, \tag{8}$$

  with finite reciprocal moments $E[U^{-m}] = \dfrac{(\gamma/2)^m \Gamma(\nu/2 - m)}{\Gamma(\nu/2)}$, for $m < \nu/2$. Lange and Sinsheimer (1993) shown that, for $\nu = \gamma$, the Mahalanobis distance d has F-distribution. In this case, a simple algebraic manipulation lead to

$$\frac{\nu}{\gamma} d = \frac{\nu}{\gamma} \frac{(y-\mu)^2}{\sigma^2} \sim F_{1,\ \nu}.$$

- *The slash distribution, $Y \sim SL(\mu, \sigma^2; \nu)$, with a shape parameter $\nu > 0$.*
  This distribution presents heavier tails than those of the normal distribution and it includes the normal case when $\nu \uparrow \infty$. Its pdf is given by

$$f(y) = \frac{\nu}{\sqrt{2\pi}\sigma} \int_0^1 u^{\nu-1/2} e^{-ud/2} du, \tag{9}$$

  with $d = (y-\mu)^2/\sigma^2$. Here we have that $\kappa(U) = 1/U$ and U with density

$$h(u; \nu) = \nu u^{\nu-1} \mathbb{I}_{(0,1)}(u), \tag{10}$$

  with reciprocal moments $E[U^{-m}] = \dfrac{\nu}{\nu - m}$, for $m < \nu$, and the Mahalanobis distance has cdf

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

- *The contaminated normal distribution, $Y \sim CN(\mu, \sigma^2; \nu, \gamma)$, $0 \le \nu \le 1$, $0 < \gamma \le 1$* (Little, 1988).

  This distribution may also be applied for modeling symmetric data with outlying observations. The parameter ν represents the percentage of outliers, while γ may be interpreted as a scale factor. Its pdf is given by

$$f(y) = \nu\phi(y|\mu, \sigma^2/\gamma) + (1-\nu)\phi(y|\mu, \sigma^2). \tag{11}$$

In this case $\kappa(U) = 1/U$ and the probability density function $h(u; \boldsymbol{\nu})$ given by

$$h(u; \boldsymbol{\nu}) = \nu \mathbb{I}_{(u=\gamma)} + (1-\nu)\mathbb{I}_{(u=1)}, \quad \boldsymbol{\nu} = (\nu, \gamma)^\top, \tag{12}$$

where the notation $\mathbb{I}_{(A)}$ is the indicator function of the set A. Clearly, $E[U^{-m}] = \nu/\gamma^m + 1 - \nu$, and

$$Pr(d \leq r) = \nu Pr(\chi^2 \leq \gamma r) + (1-\nu) Pr(\chi^2 \leq r).$$

- *The power-exponential distribution, $Y \sim PE(\mu, \sigma^2; \nu)$, with a shape parameter $0 < \nu \leq 1$.*

  Its pdf is given by

  $$f(y) = \frac{\nu}{2^{\frac{1}{2\nu}} \sigma \Gamma(\frac{1}{2\nu})} e^{-d^\nu/2}, \tag{13}$$

  with $d = (y-\mu)^2/\sigma^2$. When $\nu = 1$, the density (13) collapses to the normal density. Here U has positive stable density with $S^P(u|\nu)$ (Branco and Dey, 2001).

  Form Lange and Sinsheimer (1993), the Mahalanobis distance has cdf given by

  $$Pr(d \leq r) = \frac{r^{1/2} G(\frac{1}{2\nu}, r^\nu/2)}{\Gamma(\frac{1}{2\nu}) 2^{\frac{1}{2\nu}}},$$

  where $G(\beta, s) = \int_0^s u^{\beta-1} e^{-u} du$ is the incomplete gamma function.

3 Skew scale mixture of normal distributions

In this section, we define the univariate SSMN distributions generalizing the SSM family and study some of its properties. We have also shown that these distributions are invariant under linear transformations.

Definition 1. *A random variable Y follows a distribution between the SSMN class with location parameter $\mu \in \mathbb{R}$, scale factor σ^2 and skewness parameter $\lambda \in \mathbb{R}$, if its pdf is given by*

$$f(y) = 2 \int_0^{+\infty} \phi(y|\mu, \sigma^2 \kappa(u)) \Phi_1(\lambda \frac{y-\mu}{\sigma}) dH(u), \tag{14}$$

where U is a positive random variable with cdf $H(u; \boldsymbol{\nu})$. For a random variable with pdf as in (14), we use the notation $y \sim SSMN(\mu, \sigma^2, \lambda; H)$. If $\mu = 0$ and $\sigma^2 = 1$ we refer to it as a standard SSMN distribution and we denote it by $SSMN(\lambda; H)$.

Clearly, from (14), when $\lambda = 0$ we get the corresponding SMN distribution defined in (5). For a random variable with pdf as in (14), we write the Mahalanobis distance as

$$d_\lambda = \frac{(y-\mu)^2}{\sigma^2}.$$

In Definition 1, note that the cdf $H(u;\boldsymbol{\nu})$ is indexed by the parameter vector $\boldsymbol{\nu}$. Hence, if we suppose that $\boldsymbol{\nu}_\infty$ is such that $\boldsymbol{\nu} \uparrow \boldsymbol{\nu}_\infty$ and that $H(u;\boldsymbol{\nu})$ converges weakly to the distribution function $H_\infty(u) = H(u;\boldsymbol{\nu}_\infty)$ of the unit point mass at 1, then the density function in (14) converges to the density function of a random variable having a skew-normal distribution. The proof of this result is similar to that of Lange and Sinsheimer (1993) for the SMN case.

For a SSMN random variable, the stochastic representation given below can be used to quickly simulate pseudo realizations of y and also to study many of its properties.

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

$$\begin{aligned} Y|U = u &\sim SN(\mu, \sigma^2\kappa(u), \lambda\sqrt{\kappa(u)}) \\ U &\sim H(u;\boldsymbol{\nu}). \end{aligned} \tag{15}$$

Proof. From (14), the (joint) distribution of (Y, U) is given by

$$g(y,u) = 2\phi(y|\mu, \sigma^2\kappa(u))\Phi_1(\lambda(y-\mu)/\sigma)f_U(u).$$

Provided $g(y,u) = f(y|u)f_U(u)$, then

$$\begin{aligned} g(y|u) &= 2\phi(y|\mu,\sigma^2\kappa(u))\Phi_1\left(\lambda\frac{(y-\mu)}{\sigma}\right) \\ &= 2\phi(y|\mu,\sigma^2\kappa(u))\Phi_1\left(\lambda\sqrt{\kappa(u)}\frac{(y-\mu)}{\sigma\sqrt{\kappa(u)}}\right), \end{aligned}$$

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

Remark 1. *In other words, to generate a skew-normal independent distribution, we proceed in two steps, that is, we generate first from the distribution of U and next from the conditional distribution $Y|U$ using, for instance, the stochastic representation given in (2).*

In the next proposition, we derive a general expression for the moment generating function (mgf) of a $SSMN$ random variable.

Proposition 2. *Let $Y \sim SSMN(\mu, \sigma^2, \lambda; H)$. Then*

$$M_Y(t) = E[e^{tY}] = \int_0^\infty 2e^{t\mu + \frac{t^2}{2}\kappa(u)\sigma^2}\Phi_1\left(\frac{\sigma\lambda\kappa(u)}{\sqrt{1+\lambda^2\kappa(u)}}\, t\right)dH(u), \quad t \in \mathbb{R}. \tag{16}$$

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 expectation, it follows that $M_Y(t) = E_U[E[e^{tY}|U]]$. From Gupta and Huang (2002), $M_Z(z) = 2\exp(\frac{z^2}{2})\Phi_1\left(\frac{\lambda z}{\sqrt{1+\lambda^2}}\right)$, where $Z \sim SN(\lambda)$. So, provided $M_{a+bY}(t) = e^{at}M_Y(bt)$, we obtain

$$M_{Y|U=u}(t) = 2e^{t\mu + \frac{t^2}{2}\kappa(u)\sigma^2}\Phi_1\left(\frac{\sigma\lambda\kappa(u)}{\sqrt{1+\lambda^2\kappa(u)}}\, t\right).$$

□

In the following proposition we present the mean and the variance of a SSMN random variable.

Proposition 3. *Suppose that $Y \sim SSMN(\mu, \sigma^2, \lambda; H)$. Then,*

a)

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

b)

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

where $b = \sqrt{\frac{2}{\pi}}$.

Remark 2. *If $\kappa(U) = 1$, then $E[Y] = \mu + b\sigma\delta$ and $Var[Y] = \sigma^2(1 - b^2\delta^2)$, are the same values of a skew-normal random variable, as in (4).*

Another important property of the SSMN class is presented next.

Proposition 4. *If $Y \sim SSMN(\mu, \sigma^2, \lambda; H)$, then for any even function τ, the distribution of $\tau(Y-\mu)$ does not depend on λ and has the same distribution as that of $\tau(X-\mu)$, where $X \sim SMN(\mu, \sigma^2; H)$. In a particular case, $(Y-\mu)^2$ and $(X-\mu)^2$ are identically distributed.*

Proof. Let $f_Y(y-\mu|0, \sigma^2; H)$ be the SMN density, as in (5), then

$$\begin{aligned}
M_{\tau(Y-\mu)}(t) &= \int_{\mathbb{R}} 2e^{\tau(y-\mu)t} f_Y(y-\mu|0,\sigma^2;H)\Phi_1\left(\lambda\frac{(y-\mu)}{\sigma}\right) dy \\
&= \int_{\mathbb{R}^+} 2e^{\tau(x)t} f_Y(x|0,\sigma^2;H)\Phi_1\left(\lambda\frac{x}{\sigma}\right) dx + \\
&\quad \int_{\mathbb{R}^-} 2e^{\tau(x)t} f_Y(x|0,\sigma^2;H)\left(1-\Phi_1\left(-\lambda\frac{x}{\sigma}\right)\right) dx \\
&= \int_{\mathbb{R}^+} 2e^{\tau(x)t} f_Y(x|0,\sigma^2;H)\left(\Phi_1\left(\lambda\frac{x}{\sigma}\right) + 1 - \Phi_1\left(\lambda\frac{x}{\sigma}\right)\right) dx \\
&= \int_{\mathbb{R}} e^{\tau(x)t} f_Y(x|0,\sigma^2;H) dx \\
&= M_{\tau(X)}(t),
\end{aligned}$$

where $X \sim SMN(0, \sigma^2; H)$. □

As a byproduct of Proposition 4, we have the following interesting result.

Corollary 1. *Let $Y \sim SSMN(\mu, \sigma^2, \lambda; H)$. Then the quadratic form*

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

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

The result of Corollary 1 is interesting because it allows us to do model checking in practice (see Section 5). On the other hand, Corollary 1 jointly with the result found, for instance, in Lange and Sinsheimer (1993, Section 2) allows us to obtain the *mth* moment of d_λ.

Corollary 2. *Let $Y \sim SSMN(\mu, \sigma^2, \lambda; H)$. Then for any $m > 0$*

$$E[d_\lambda^m] = \frac{2^m \Gamma(m + 1/2)}{\sqrt{\pi}} E[\kappa(U)^m].$$

Proof. The pdf of Y is given by $f(y) = \int_0^\infty \phi(y|\mu, \sigma^2\kappa(u))du$. Thus,

$$\begin{aligned} f(y|u) &= \phi(y|\mu, \sigma^2\kappa(u)) \\ &= \frac{\kappa(u)^{-1/2}}{\sqrt{2\pi}\sigma} e^{-\frac{d_\lambda}{2\kappa(u)}} \end{aligned}$$

Consider the random variable $Z = \dfrac{d_\lambda}{\kappa(U)} = \dfrac{(Y-\mu)^2}{\sigma^2\kappa(U)}$. The pdf of $Z|U$ has density χ_1^2. Provided U and d_λ are independent, then $E[d_\lambda^m] = \dfrac{2^m\Gamma(m+1/2)}{\sqrt{\pi}} E[\kappa^m(U)]$. □

In the next proposition we shall show that an SSMN random variable is invariant under linear transformations. This result is summarized in the following proposition:

Proposition 5. *Let $Y \sim SSMN(\mu, \sigma^2, \lambda; H)$. Then for any fixed value $b \in \mathbb{R}$,*

$$V = a + bY \sim SSMN(a + b\mu, b^2\sigma^2, sign(b)\lambda; H). \tag{17}$$

Proof. The proof follows directly from Proposition 2, since $M_{a+by}(\mathbf{t}) = e^{at}M_Y(bt)$. □

By using (17), when $a = 0$ and $b = -1$, we have the following additional property of an $SSMN$ random variable.

Corollary 3. *Let $Y \sim SSMN(\mu, \sigma^2, \lambda; H)$. Then, $-Y \sim SSMN(-\mu, \sigma^2, -\lambda; H)$.*

3.1 Examples of SSMN distributions

Some examples of SSMN distributions, include

- *The Skew Generalized Student-t distribution with $\nu > 0$ degrees of freedom, denoted $Y \sim SGt(\mu, \sigma^2, \lambda; \nu, \gamma)$.* Considering $U \sim Gamma(\nu/2, \gamma/2)$, $\kappa(U) = 1/U$, Y has the density function:

$$f(y) = 2\frac{1}{\sigma\sqrt{\gamma\pi}} \frac{\Gamma((\nu+1)/2)}{\Gamma(\frac{\nu}{2})} \left(1 + \frac{d}{\gamma}\right)^{-(\frac{\nu+1}{2})} \Phi_1\left(\lambda \frac{(y-\mu)}{\sigma}\right). \tag{18}$$

  When $\gamma = \nu$, we have a skew-t normal distribution (Gómez, Venegas and Bolfarine, 2007) where it has been shown that it can present a much wider asymmetry range than the one presented by the ordinary skew normal distribution (Azzalini, 1985).

Another particular case of the skew generalized t distribution is the skew-Cauchy normal distribution, that follows when $\nu = \gamma = 1$. Also, when $\nu \uparrow \infty$, we get the skew-normal distribution as the limiting case. The mean and variance of $Y \sim SGt(\mu, \sigma^2, \lambda; \nu, \gamma)$ are given by

$$E[Y] = \mu + b\sigma\lambda(\gamma/2)^{1/2}\frac{\Gamma((\nu-1)/2)}{\Gamma(\nu/2)}E_V\left[(V+\lambda^2)^{-1/2}\right], \tag{19}$$

$$Var[Y] = \sigma^2\left[\frac{\gamma}{\nu-2} - \frac{b^2\lambda^2\gamma}{2}\left(\frac{\Gamma((\nu-1)/2)}{\Gamma(\nu/2)}\right)^2 E_V^2\left[(V+\lambda^2)^{-1/2}\right]\right], \tag{20}$$

where $b = \sqrt{\frac{2}{\pi}}$ and $V \sim Gamma(\frac{\nu-1}{2}, \frac{\gamma}{2})$. The expected values are computed numerically.

- *The skew-slash distribution, with shape parameter $\nu > 0$, denoted $SSL(\mu, \sigma^2, \lambda; \nu)$.* With $h(u;\nu)$ as in (10) and $\kappa(U) = 1/U$, we have

$$f(y) = 2\nu\Phi_1\left(\lambda\frac{y-\mu}{\sigma}\right)\int_0^1 u^{\nu-1}\phi\left(y|\mu, \frac{\sigma^2}{u}\right)du, \quad y \in \mathbb{R}. \tag{21}$$

The skew-slash distribution reduces to the skew-normal distribution when $\nu \uparrow \infty$. The mean and variance are given by

$$E[Y] = \mu + \frac{b\sigma\lambda\nu}{\nu-1/2}E_V\left[(V+\lambda^2)^{-1/2}\right], \tag{22}$$

$$Var[Y] = \sigma^2\left(\frac{\nu}{\nu-1} - \frac{b^2\lambda^2\nu^2}{(\nu-1/2)^2}E_V^2\left[(V+\lambda^2)^{-1/2}\right]\right), \tag{23}$$

where $V \sim Beta(1, \nu - 1/2)$.

- *The skew-contaminated normal distribution, denoted $SCN(\mu, \sigma^2, \lambda; \nu, \gamma)$, $0 \le \nu \le 1$, $0 < \gamma \le 1$.* Taking $h(u;\nu)$ as in (12) it follows, straightforwardly, that

$$f(y) = 2\Big\{\nu\phi\left(y|\mu, \frac{\sigma^2}{\gamma}\right)\Phi_1\left(\lambda\frac{y-\mu}{\sigma}\right) + (1-\nu)\phi(y|\mu,\sigma^2)\Phi_1\left(\lambda\frac{y-\mu}{\sigma}\right)\Big\}. \tag{24}$$

The skew-contaminated normal distribution reduces to the skew-normal distribution when $\gamma = 1$. Hence, the mean and the variance are given by

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

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

- *The skew power-exponential distribution, $Y \sim SPE(\mu, \sigma^2, \lambda; \nu)$, with a shape parameter $0 < \nu \leq 1$.*
  Its pdf is given by

$$f(y) = 2\frac{\nu}{2^{\frac{1}{2\nu}}\sigma\Gamma(\frac{1}{2\nu})}e^{-d^{\nu}/2}\Phi_1\left(\lambda\frac{y-\mu}{\sigma}\right). \tag{25}$$

  with $d = \frac{(y-\mu)^2}{\sigma^2}$. The skew power-exponential distribution reduces to the skew-normal distribution when $\nu = 1$. In this case we have not explicit form to $\kappa(u)$.

In Figure 1, we plotted the density of the standard $SN(3)$ distribution together with the standard densities of the distributions $SGt(3; 2, 2)$, $SSL(3; 0.5)$, $SNC(3; 0.9, 0.1)$ and $SEP(3; 0.5)$. They are re-scaled 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 power-exponential, skew-slash, and the skew-t distributions have much heavier tails than the skew-normal distribution. Note that all they represent extremes situations, at least in the symmetric case, for instance, the Slash with $\nu = 0.5$ does not has finite first moment.

In what follows, we propose using the EM-algorithm to obtain the ML estimate of the parameter vector $\boldsymbol{\theta}$. We note that it is complicated to implement this approach without identifying a stochastic representation. The proposed methodology for the class of models we are dealing with does not exist in the literature. One special feature, however, is that the E-step is as in Lange and Sinsheimer (1993). Moreover, studies related to local influence for incomplete data (Zhu and Lee, 2001) can be easily extended from these results which will be reported elsewhere.

4 SSMN regression models and the EM-algorithm

Suppose that we have observations on m independent individuals, denoted $Y_1, \ldots, Y_m$, where $Y_i \sim SSMN(\mu_i, \sigma^2, \lambda; H)$, $i = 1, \ldots, m$. Associated with individual i we assume a known $p \times 1$ covariate vector $\mathbf{x}_i$, which we use to specify the linear predictor $\mu_i = \mathbf{x}_i^{\top}\boldsymbol{\beta}$, where $\boldsymbol{\beta}$ is a p-dimensional vector of unknown regression coefficients. Hence, relating the two sets of variables we consider

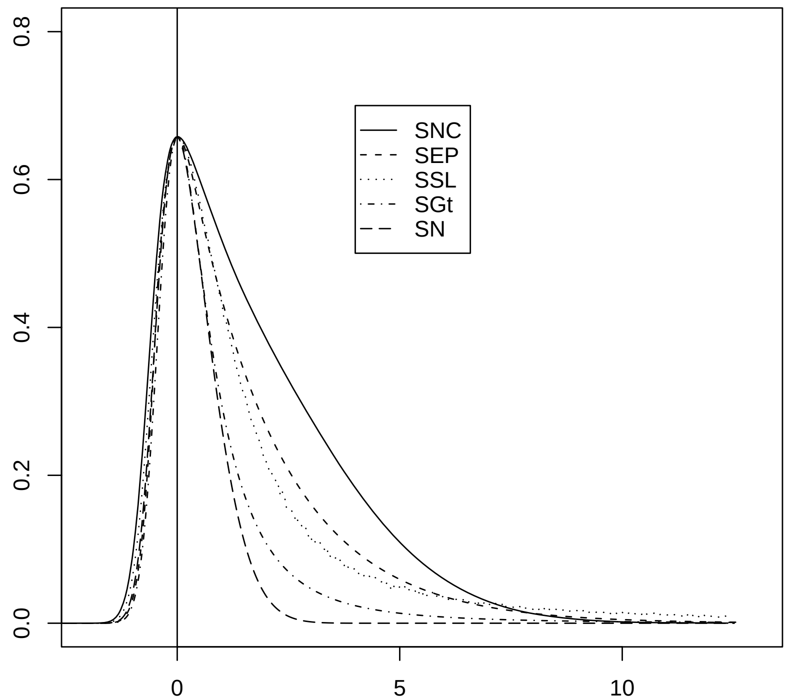
$$\begin{aligned} y_i &= \beta_0 + \textstyle\sum_{k=1}^{p} x_{ik}\beta_k + \varepsilon_i, \quad i = 1, \cdots, m, \\ &= \mathbf{x}_i^{\top}\boldsymbol{\beta} + \varepsilon_i, \quad \varepsilon_i \sim SSMN(0, \sigma^2, \lambda; H). \end{aligned} \tag{26}$$

Thus, according by expression (38) in Appendix, the observed-data log-likelihood function of $\boldsymbol{\theta} = (\boldsymbol{\beta}^{\top}, \sigma^2, \lambda)^{\top}$ is given by

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^{m} \log\left[2\int_0^{+\infty}\int_0^{+\infty} \phi(y_i|\mathbf{x}_i^{\top}\boldsymbol{\beta}, \sigma^2\kappa(u_i))\phi(t_i|\lambda(y_i - \mathbf{x}_i^{\top}\boldsymbol{\beta}), \sigma^2)h(u_i;\nu)dt_i du_i\right]. \tag{27}$$

It is not a simple task to find the ML estimate of the parameter vector $\boldsymbol{\theta}$ by directly maximizing the log-likelihood function. Thus, we prefer to implement the

Figure 1: Density curves of the univariate skew-normal $SN(3)$, skew-t $SGt(3;2,2)$, skew-slash $SSL(3;0.5)$, skew-contaminated $SNC(3;0.9,0.1)$ and skew-power exponential $SEP(3;0.5)$ normal distributions.



EM-algorithm.

The EM-algorithm is a popular iterative algorithm for ML estimation for models with incomplete data. More specifically, let $\mathbf{y}$ denote the observed data and $\mathbf{s}$ denote the missing data. The complete data $\mathbf{y}_c = (\mathbf{y}, \mathbf{s})$ is $\mathbf{y}$ augmented with $\mathbf{s}$. We denote by $\ell_c(\boldsymbol{\theta}|\mathbf{y}_c)$, $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, the complete-data log-likelihood function and by $Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}) = E[\ell_c(\boldsymbol{\theta}|\mathbf{y}_c)|\mathbf{y}, \widehat{\boldsymbol{\theta}}]$, the expected complete-data log-likelihood function. Each iteration of the EM-algorithm involves two steps; an E-step and an M-step, defined as:

- E-step: Compute $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(r)})$ as a function of $\boldsymbol{\theta}$;
- M-step: Find $\boldsymbol{\theta}^{(r+1)}$ such that $Q(\boldsymbol{\theta}^{(r+1)}|\boldsymbol{\theta}^{(r)}) = \max_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(r)})$.

Notice that, by using (2) and (15), the set-up defined above can be written as

$$
\begin{aligned}
Y_i|t_i, u_i, &\overset{\text{ind}}{\sim} N\left(\mathbf{x}_i^\top \boldsymbol{\beta} + \sigma\lambda \frac{\kappa(u_i)}{\sqrt{1+\lambda^2\kappa(u_i)}} t_i, \sigma^2\kappa(u_i)\left(1-\lambda^2\frac{\kappa(u_i)}{1+\lambda^2\kappa(u_i)}\right)\right), \\
u_i &\overset{\text{ind}}{\sim} h(u_i;\boldsymbol{\nu}) \qquad (28)\\
t_i &\overset{\text{iid}}{\sim} HN_1(0,1) \;\; i=1,\ldots,m,
\end{aligned}
$$

all independent, where $HN_1(0,1)$ denotes the univariate standard half-normal distribution (see $|T_0|$ in equation (2) or Johnson, Kotz and Balakrishnan, 1994). We assume that the parameter vector $\boldsymbol{\nu}$ is known. In applications the optimum value of $\boldsymbol{\nu}$ can be choosing by using the profile likelihood and the Schwarz Information Criterion (see Lange, Little and Taylor, 1989).

Let $\mathbf{y} = (y_1,\ldots,y_m)^\top$, $\mathbf{u} = (u_1,\ldots,u_m)^\top$ 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

$$
\begin{aligned}
\ell_c(\boldsymbol{\theta}|\mathbf{y}_c) &\propto -m\log\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^m \frac{(y_i - \mathbf{x}_i^\top\boldsymbol{\beta})^2}{\kappa(u_i)} - \frac{1}{2\sigma^2}\sum_{i=1}^m [t_i - \lambda(y_i - \mathbf{x}_i^\top\boldsymbol{\beta})]^2 \\
&= -m\log\sigma^2 - \frac{1}{2\sigma^2}\mathbf{t}^{2^\top}\mathbf{1}_m + \frac{\lambda}{\sigma^2}\mathbf{t}^\top(\mathbf{y}-\mathbf{x}\boldsymbol{\beta}) \\
&\quad -\frac{1}{2\sigma^2}(\mathbf{y}-\mathbf{x}\boldsymbol{\beta})^\top\left(\mathbf{D}(\boldsymbol{\kappa}) + \lambda^2\mathbb{I}_m\right)(\mathbf{y}-\mathbf{x}\boldsymbol{\beta}) \qquad (29)
\end{aligned}
$$

where $\mathbf{1}_m(m\times 1)$ is a vector of 1's, $\mathbb{I}_m$ is the identity matrix of order m, $\mathbf{x}^\top = (\mathbf{x}_1\ldots,\mathbf{x}_m)$ is a matrix of dimension $p\times m$ and $\mathbf{D}(\boldsymbol{\kappa}) = Diag(\kappa(u_1),\ldots,\kappa(u_m))$.

Letting $\widehat{t_i} = \mathrm{E}[t_i|\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}, \mathbf{y}_i]$, $\widehat{t^2}_i = \mathrm{E}[t_i^2|\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}, \mathbf{y}_i]$ and $\widehat{\kappa}_i = \mathrm{E}[\kappa^{-1}(u_i)|\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}, \mathbf{y}_i]$ we obtain, using the moments of the truncated normal distribution, that

$$
\widehat{t_i} = \widehat{\lambda}\widehat{\eta}_i + \widehat{\sigma}W_{\Phi_1}\left(\frac{\widehat{\lambda}\widehat{\eta}_i}{\widehat{\sigma}}\right) \text{ and } \widehat{t^2}_i = \widehat{\lambda}^2\widehat{\eta}_i^2 + \widehat{\sigma}^2 + \widehat{\lambda}\widehat{\sigma}\widehat{\eta}_i W_{\Phi_1}\left(\frac{\widehat{\lambda}\widehat{\eta}_i}{\widehat{\sigma}}\right), \qquad (30)
$$

where $W_{\Phi_1}(u) = \phi_1(u)/\Phi_1(u)$ and $\widehat{\eta}_i = (y_i - \mathbf{x}_i^\top\widehat{\boldsymbol{\beta}})$, $i = 1,\ldots,m$. For the specific distributions discussed here, $\widehat{\kappa}_i$ are given by equations (34)-(37).

It follows, after some simple algebra, that the expectation with respect to $\mathbf{t}$, $\mathbf{u}$ conditional on $\mathbf{y}$, of the complete log-likelihood function, has the form

$$
\begin{aligned}
Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}) &= \mathrm{E}[\ell_c(\boldsymbol{\theta}|\mathbf{y}_c)|\mathbf{y}, \widehat{\boldsymbol{\theta}}] \\
&= -m\log\sigma^2 - \frac{1}{2\sigma^2}\widehat{\mathbf{t}}^{2\top}\mathbf{1}_m + \frac{\lambda}{\sigma^2}\widehat{\mathbf{t}}^\top(\mathbf{y}-\mathbf{x}\boldsymbol{\beta}) \\
&\quad -\frac{1}{2\sigma^2}(\mathbf{y}-\mathbf{x}\boldsymbol{\beta})^\top\left(\mathbf{D}(\widehat{\boldsymbol{\kappa}}) + \lambda^2\mathbb{I}_m\right)(\mathbf{y}-\mathbf{x}\boldsymbol{\beta}) \qquad (31)
\end{aligned}
$$

where $\widehat{\boldsymbol{\kappa}} = [\widehat{\kappa}_1,\ldots,\widehat{\kappa}_m]^\top$. Thus, we have the following EM-algorithm:
E-step: Given $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}$, compute for $i = 1,\ldots,m$, $\widehat{t}_i$, $\widehat{t^2}_i$ using (30) and $\widehat{k}_i$ using

(34)-(37), for instance.
M-step: Update $\widehat{\boldsymbol{\theta}}$ by maximizing $Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})$ over $\boldsymbol{\theta}$, which leads to the following closed form expressions

$$\begin{aligned}
\widehat{\boldsymbol{\beta}} &= \left[\mathbf{x}^\top\left(\mathbf{D}(\widehat{\boldsymbol{\kappa}})+\lambda^2\mathbb{I}_m\right)\mathbf{x}\right]^{-1}\mathbf{x}^\top\left[\mathbf{D}(\widehat{\boldsymbol{\kappa}})\mathbf{y}-\lambda(\widehat{\mathbf{t}}-\lambda\mathbf{y})\right], \qquad (32)\\
\widehat{\sigma^2} &= \frac{1}{2m}\left[Q_{\widehat{\mathbf{k}}}(\boldsymbol{\beta})+\widehat{\mathbf{t}^2}^\top\mathbf{1}_m-2\lambda\widehat{\mathbf{t}}^\top(\mathbf{y}-\mathbf{x}\boldsymbol{\beta})+\lambda^2Q(\boldsymbol{\beta})\right],\\
\widehat{\lambda} &= \frac{\widehat{\mathbf{t}}^\top(\mathbf{y}-\mathbf{x}\boldsymbol{\beta})}{Q(\boldsymbol{\beta})},
\end{aligned}$$

where $Q(\boldsymbol{\beta})=(\mathbf{y}-\mathbf{x}\boldsymbol{\beta})^\top(\mathbf{y}-\mathbf{x}\boldsymbol{\beta})$, $Q_{\widehat{\mathbf{k}}}(\boldsymbol{\beta})=(\mathbf{y}-\mathbf{x}\boldsymbol{\beta})^\top\mathbf{D}(\widehat{\boldsymbol{\kappa}})(\mathbf{y}-\mathbf{x}\boldsymbol{\beta})$.

4.1 Conditional distributions for the EM algorithm

In this section we compute the conditional distribution $u_i|y_i$ for the distributions present in Section 3, which are used with the EM-algorithm. Before, we give an important result.

Proposition 6. *(An invariance result) If* $Y_i \sim SSMN(\mu,\sigma^2,\lambda;H)$, $i=1,\ldots,m$, *then*

$$f(u_i|Y_i=y_i) \propto h(u_i;\nu)\phi(y_i|\mu,\sigma^2\kappa(u_i)). \qquad (33)$$

Proof. In fact, from (38) we have that,

$$\begin{aligned}
f(y_i,u_i,t_i) &= 2\phi(y_i|\mu,\sigma^2\kappa(u_i))\phi(t_i|\lambda\frac{y_i-\mu}{\sigma},1)h(u_i;\nu)\mathbb{I}_{(t_i>0)}\\
&= f(u_i)f(y_i|u_i)f(t_i|y_i).
\end{aligned}$$

Then,

$$\begin{aligned}
f(u_i|y_i) &\propto f(u_i)f(y_i|u_i)\\
&= h(u_i;\nu)\phi(y_i|\mu,\sigma^2\kappa(u_i)).
\end{aligned}$$

which concludes the proof. □

From Proposition 6 it follows that, under the more general $SSMN$ distribution considered here, the conditional distribution $u_i|y_i$ reduces to considering the corresponding SMN model. This peculiarity tremendously simplify the EM algorithm implementation. Hence, for the discussed distributions, we have the following results:

- *The generalized skew-t case.* If $Y_i \sim SGt(\mu,\sigma^2,\lambda;\nu,\gamma)$, $i=1,\ldots,m$, then, from (8) jointly with Proposition 6, we have that

$$\begin{aligned}
f(u_i|y_i) &= \frac{(\gamma/2)^{\nu/2}}{\Gamma(\nu/2)}u_i^{\nu/2-1}e^{-\gamma u_i/2}\frac{u_i^{1/2}}{\sqrt{2\pi}\sigma}e^{-u_id_i/2}\\
u_i|y_i &\sim Gamma(\frac{\nu+1}{2},\frac{\gamma+d_i}{2}),
\end{aligned}$$

$d_i = d_{\lambda i} = (y_i - \mu)^2/\sigma^2$. So, $E[U_i^\alpha|y_i] = \dfrac{\Gamma\left(\frac{\nu+1}{2}+\alpha\right)}{\Gamma\left(\frac{\nu+1}{2}\right)\left(\frac{\gamma+d_i}{2}\right)^\alpha}$ and then

$$\widehat{k}_i = \frac{\nu+1}{\gamma+d_i}. \tag{34}$$

• *The skew-slash case.* If $Y_i \sim SSL(\mu,\sigma^2,\lambda;\nu)$, $i=1,\ldots,m$, we obtain

$$\begin{aligned} f(u_i|y_i) &= \nu u_i^{\nu-1}\mathbf{I}_{(0,1)}(u)\frac{u_i^{1/2}}{\sqrt{2\pi}\sigma}e^{-u_i d_i/2} \\ u_i|y_i &\sim Gamma(\nu+1/2, d_i/2)\mathbb{I}_{(0,1)}(u_i). \end{aligned}$$

Thus, $E(U_i^\alpha|y_i) = \dfrac{\Gamma\left(\nu+1/2+\alpha\right)}{\Gamma(\nu+1/2)(di/2)^\alpha}\dfrac{P_1(\nu+\alpha+1/2,di/2)}{P_1(\nu+1/2,di/2)}$ and

$$\widehat{k}_i = \frac{(2\nu+1)}{d_i}\frac{P_1(\nu+3/2,d_i/2)}{P_1(\nu+1/2,d_i/2)}, \tag{35}$$

where $P_x(a,b)$ denotes the cdf of the $Gamma(a,b)$ distribution evaluated at x.

• *The skew-contaminated normal case.* If $Y_i \sim SNC(\mu,\sigma^2,\lambda;\nu,\gamma)$, $i=1,\ldots,m$, we have that

$$f(u_i|y_i) = \nu p_i \mathbf{1}_{(u_i=\gamma)} + (1-\nu)p_i\mathbf{1}_{(u_i=1)},$$

with $p_i = \dfrac{u_i^{1/2}\exp\{-\frac{d_i u_i}{2}\}}{\nu\gamma^{1/2}\exp\{-\frac{d_i\gamma}{2}\}+(1-\nu)\exp\{-\frac{d_i}{2}\}}$.

Moreover, $E(U_i^\alpha|y_i) = \dfrac{1-\nu+\nu\gamma^{\alpha+1/2}\exp\left\{(1-\gamma)d_i/2\right\}}{1-\nu+\nu\gamma^{1/2}\exp\left\{(1-\gamma)d_i/2\right\}}$ and

$$\widehat{k}_i = \frac{1-\nu+\nu\gamma^{3/2}\exp\left\{(1-\gamma)d_i/2\right\}}{1-\nu+\nu\gamma^{1/2}\exp\left\{(1-\gamma)d_i/2\right\}}. \tag{36}$$

• *The skew power-exponential case.* In this case we have the following proposition:

Proposition 7. *If $Y_i \sim SPE(\mu,\sigma^2,\lambda;\nu)$, $i=1,\ldots,m$, then*

$$\widehat{k}_i = E[\kappa^{-1}(U_i)|y_i] = \nu d_i^{\nu-1}, \tag{37}$$

with $d_i = \frac{(y_i-\mu)^2}{\sigma^2}$.

Proof. We omitted the index i to facilitate the notation. Being $Y \sim SMN(\mu,\sigma^2;\nu)$, we have that

$$f_Y(y) = \int_0^{+\infty}\phi(y|\mu,\sigma^2\kappa(u))f_U(u)du$$

and $f_{(U,Y)}(u,y) = \phi(y|\mu,\sigma^2\kappa(u))f_U(u)$. As $V=\kappa^{-1}(U)$, $\kappa(U)=1/V$, then

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

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

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

where $d=\frac{(y-u)^2}{\sigma^2}$. Hence, $\kappa^{-1}(U)|Y$ belong to exponential family, with $\theta=-d$, $b(\theta)=-d^\nu=-(-\theta)^\nu$ and $\phi=1/2$. Thus, $E[k^{-1}(u)|y]=b'(\theta)=\nu d^{\nu-1}$. □

5 The observed information matrix

Let Y be a random variable following a SSMN distribution as in (14). Hence, for an observed sample $y_1,\ldots,y_m$, the log-likelihood function is of the form $\ell(\boldsymbol{\theta})=\sum_{i=1}^{m}\ell_i(\boldsymbol{\theta})$, with

$$\ell_i(\boldsymbol{\theta})=\log 2+\ell_{1_i}(\boldsymbol{\theta})+log[\Phi_1(\ell_{2_i}(\boldsymbol{\theta}))],\quad \boldsymbol{\theta}=(\boldsymbol{\beta}^\top,\sigma^2,\lambda)^\top,$$

where $\ell_1(\boldsymbol{\theta})$ is the log-likelihood function of the corresponding symmetric SMN distribution and $\ell_{2_i}(\boldsymbol{\theta})=\lambda\dfrac{y_i-\mathbf{x}_i^\top\boldsymbol{\beta}}{\sigma}$. Then, the first derivative of $\ell_i(\boldsymbol{\theta})$ is given by

$$\frac{\partial\ell_i(\boldsymbol{\theta})}{\partial\boldsymbol{\psi}}=\frac{\partial\ell_{1_i}(\boldsymbol{\theta})}{\partial\boldsymbol{\psi}}+W_\Phi(l_{2_i}(\boldsymbol{\theta}))\frac{\partial\ell_{2_i}(\boldsymbol{\theta})}{\partial\boldsymbol{\psi}},\quad \boldsymbol{\psi}=\boldsymbol{\beta},\sigma^2,\lambda.$$

The second derivative is given by

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

where $W_\Phi^{(1)}(x)=-W_\Phi(x)(x+W_\Phi(x))$ is the derivative of $W_\Phi(x)$.

Thus, the observed information matrix for $\boldsymbol{\theta}$ can be written as

$$I(\boldsymbol{\theta})=I_1(\boldsymbol{\theta})+I_2(\boldsymbol{\theta}),$$

$$I_\kappa(\boldsymbol{\theta})=\begin{pmatrix} I^k_{\boldsymbol{\beta}\boldsymbol{\beta}} & I^k_{\sigma^2\boldsymbol{\beta}} & I^k_{\lambda\boldsymbol{\beta}} \\ & I^k_{\sigma^2\sigma^2} & I^k_{\lambda\sigma^2} \\ & & I^k_{\lambda\lambda}\end{pmatrix},$$

for $k=1,2$, $\boldsymbol{\gamma},\boldsymbol{\psi}=\boldsymbol{\beta},\sigma^2,\lambda$, and

$$\begin{aligned}
I^1_{\boldsymbol{\gamma}\boldsymbol{\psi}^\top} &= \sum_{i=1}^{m}\frac{\partial^2\ell_{1_i}(\boldsymbol{\theta})}{\partial\boldsymbol{\gamma}\partial\boldsymbol{\psi}^\top}, \\
I^2_{\boldsymbol{\gamma}\boldsymbol{\psi}^\top} &= \sum_{i=1}^{m}\left[W_\Phi(\ell_{2_i}(\boldsymbol{\theta}))\frac{\partial^2\ell_{2_i}(\boldsymbol{\theta})}{\partial\boldsymbol{\gamma}\partial\boldsymbol{\psi}^\top}+W_\Phi^{(1)}(\ell_{2_i}(\boldsymbol{\theta}))\frac{\partial\ell_{2_i}(\boldsymbol{\theta})}{\partial\boldsymbol{\gamma}}\frac{\partial\ell_{2_i}(\boldsymbol{\theta})}{\partial\boldsymbol{\psi}^\top}\right].
\end{aligned}$$

For all distributions in the SMN class, the elements of $I_2(\boldsymbol{\theta})$ are common and are given by

$$\begin{aligned}
I^2_{\boldsymbol{\beta}\boldsymbol{\beta}} &= -\frac{\lambda^2}{\sigma^2}\mathbf{x}^\top\mathbf{D}\left(W^{(1)}_\Phi\left[\lambda\frac{(\mathbf{y}-\mathbf{x}\boldsymbol{\beta})}{\sigma}\right]\right)\mathbf{x} \\
I^2_{\sigma^2\boldsymbol{\beta}} &= -\frac{\lambda}{2\sigma^3}\mathbf{x}^\top W_\Phi\left[\lambda\frac{(\mathbf{y}-\mathbf{x}\boldsymbol{\beta})}{\sigma}\right] - \frac{\lambda^2}{2\sigma^4}\mathbf{x}^\top\mathbf{D}\left(W^{(1)}_\Phi\left[\lambda\frac{(\mathbf{y}-\mathbf{x}\boldsymbol{\beta})}{\sigma}\right]\right)(\mathbf{y}-\mathbf{x}\boldsymbol{\beta}) \\
I^2_{\lambda\boldsymbol{\beta}} &= \frac{1}{\sigma}\mathbf{x}^\top W_\Phi\left[\lambda\frac{(\mathbf{y}-\mathbf{x}\boldsymbol{\beta})}{\sigma}\right] + \frac{\lambda}{\sigma^2}\mathbf{x}^\top\mathbf{D}\left(W^{(1)}_\Phi\left[\lambda\frac{(\mathbf{y}-\mathbf{x}\boldsymbol{\beta})}{\sigma}\right]\right)(\mathbf{y}-\mathbf{x}\boldsymbol{\beta}) \\
I^2_{\sigma^2\sigma^2} &= -\frac{3\lambda}{4\sigma^5}(\mathbf{y}-\mathbf{x}\boldsymbol{\beta})^\top W_\Phi\left[\lambda\frac{(\mathbf{y}-\mathbf{x}\boldsymbol{\beta})}{\sigma}\right] - \frac{\lambda^2}{4\sigma^6}Q_w(\boldsymbol{\beta})
\end{aligned}$$

$$\begin{aligned}
I^2_{\lambda\sigma^2} &= \frac{1}{2\sigma^3}(\mathbf{y}-\mathbf{x}\boldsymbol{\beta})^\top W_\Phi\left[\lambda\frac{(\mathbf{y}-\mathbf{x}\boldsymbol{\beta})}{\sigma}\right] + \frac{\lambda}{2\sigma^4}Q_w(\boldsymbol{\beta}) \\
I^2_{\lambda\lambda} &= -\frac{1}{\sigma^2}Q_w(\boldsymbol{\beta})
\end{aligned}$$

where $Q_w(\boldsymbol{\beta}) = (\mathbf{y}-\mathbf{x}\boldsymbol{\beta})^\top\mathbf{D}\left(W^{(1)}_\Phi\left[\lambda\frac{(\mathbf{y}-\mathbf{x}\boldsymbol{\beta})}{\sigma}\right]\right)(\mathbf{y}-\mathbf{x}\boldsymbol{\beta})$. Notice that being $W_\Phi(x) = \phi(x)/\Phi(x)$, with $\phi(.)$ and $\Phi(.)$ the density and the cumulative functions of the standard normal distribution, then $-1 \leq W^{(1)}_\Phi(x) \leq 0$. The fact that this matrix is constant for all families in the class, it makes the information matrix nonsingular for nonbormnal situations. As shown in Piesey (2000) this certainly is a problem with the ordinary skew normal distribution.

The matrix $I_1(\boldsymbol{\theta})$ can be calculated for each SMN distribution considered, as follows

- *The Student-t distribution*

$$\begin{aligned}
I^1_{\boldsymbol{\beta}\boldsymbol{\beta}} &= -\frac{\nu+1}{\gamma\sigma^2}\mathbf{x}^\top\left(\frac{2}{\gamma\sigma^2}\mathbf{D}^2(\mathbf{V})\mathbf{D}^2(\mathbf{y}-\mathbf{x}\boldsymbol{\beta}) - \mathbf{D}(\mathbf{V})\right)\mathbf{x} \\
I^1_{\sigma^2\boldsymbol{\beta}} &= -\frac{\nu+1}{\gamma\sigma^4}\mathbf{x}^\top\mathbf{D}(\mathbf{y}-\mathbf{x}\boldsymbol{\beta})\left(\frac{1}{\gamma\sigma^2}\mathbf{D}^2(\mathbf{V})\mathbf{D}(\mathbf{y}-\mathbf{x}\boldsymbol{\beta})(\mathbf{y}-\mathbf{x}\boldsymbol{\beta}) - \mathbf{V}\right) \\
I^1_{\sigma^2\sigma^2} &= -\frac{m}{2\sigma^4} + \frac{\nu+1}{\gamma\sigma^6}Q_{\mathbf{V}}(\boldsymbol{\beta}) - \frac{\nu+1}{2\gamma^2\sigma^8}(\mathbf{y}-\mathbf{x}\boldsymbol{\beta})^\top\mathbf{D}^3(\mathbf{y}-\mathbf{x}\boldsymbol{\beta})\mathbf{D}(\mathbf{V})\mathbf{V} \\
I^1_{\lambda\boldsymbol{\tau}} &= \mathbf{0}, \boldsymbol{\tau} = \boldsymbol{\beta}, \sigma^2, \lambda,
\end{aligned}$$

where $Q_{\mathbf{V}}(\boldsymbol{\beta}) = (\mathbf{y}-\mathbf{x}\boldsymbol{\beta})^\top\mathbf{D}(\mathbf{V})(\mathbf{y}-\mathbf{x}\boldsymbol{\beta})$ and $\mathbf{V} = (V_1, \cdots, V_n)^\top$, $V_i = \left(1+\frac{d_i}{\gamma}\right)^{-1}$, $d_i = \frac{(y_i-\mathbf{x}_i^\top\boldsymbol{\beta})^2}{\sigma^2}$. As noted in Gómez, Venegas and Bolfarine (2007), the information matrix for the skew skew-t normal model is not singular for finite degrees of freedom when $\lambda = 0$. This certainly is a problem with the skew-normal and skew-t distributions (Pewsey, 2000).

- *The contaminated normal distribution*

$$
\begin{aligned}
I^1_{\boldsymbol{\beta}\boldsymbol{\beta}} &= \frac{\gamma+1}{\sigma^2}\mathbf{x}^\top\mathbf{x}\\
I^1_{\sigma^2\boldsymbol{\beta}} &= \frac{\gamma+1}{\sigma^4}\mathbf{x}^\top(\mathbf{y}-\mathbf{x}\boldsymbol{\beta})\\
I^1_{\sigma^2\sigma^2} &= -\frac{m}{\sigma^4}+\frac{\gamma+1}{\sigma^6}(\mathbf{y}-\mathbf{x}\boldsymbol{\beta})^\top(\mathbf{y}-\mathbf{x}\boldsymbol{\beta})\\
I^1_{\lambda\boldsymbol{\tau}} &= \mathbf{0}, \boldsymbol{\tau}=\boldsymbol{\beta},\sigma^2,\lambda.
\end{aligned}
$$

- *The power-exponential distribution*

$$
\begin{aligned}
I^1_{\boldsymbol{\beta}\boldsymbol{\beta}} &= \frac{\nu(2\nu-1)}{\sigma^{2\nu}}\mathbf{x}^\top\mathbf{D}[(\mathbf{y}-\mathbf{x}\boldsymbol{\beta})^{2\nu-2}]\mathbf{x}\\
I^1_{\sigma^2\boldsymbol{\beta}} &= \frac{\nu^2}{\sigma^{2(\nu+1)}}\mathbf{x}^\top[(\mathbf{y}-\mathbf{x}\boldsymbol{\beta})^{2\nu-1}]\\
I^1_{\sigma^2\sigma^2} &= -\frac{m}{2\sigma^4}+\frac{\nu(\nu+1)}{2\sigma^{2(\nu+2)}}[(\mathbf{y}-\mathbf{x}\boldsymbol{\beta})^{\nu}]^\top[(\mathbf{y}-\mathbf{x}\boldsymbol{\beta})^{\nu}]\\
I^1_{\lambda\boldsymbol{\tau}} &= \mathbf{0}, \boldsymbol{\tau}=\boldsymbol{\beta},\sigma^2,\lambda.
\end{aligned}
$$

- *The slash distribution*

$$
\begin{aligned}
I^1_{\boldsymbol{\beta}\boldsymbol{\beta}} &= -(2\nu+1)\mathbf{x}^\top\mathbf{D}[(\mathbf{y}-\mathbf{x}\boldsymbol{\beta})^{-2}]\mathbf{D}[P_1(\nu+1/2,\mathbf{d}/2)^{-1}]\times\\
&\quad\big[(2\nu+3)\mathbf{D}[P_1(\nu+5/2,\mathbf{d}/2)]-\mathbf{D}[P_1(\nu+3/2,\mathbf{d}/2)]\\
&\quad-(2\nu+1)\mathbf{D}[P_1(\nu+3/2,\mathbf{d}/2)^2]\mathbf{D}[P_1(\nu+1/2,\mathbf{d}/2)^{-1}]\big]\mathbf{x}\\
I^1_{\sigma^2\boldsymbol{\beta}} &= -\frac{(2\nu+1)}{2\sigma^2}\mathbf{x}^\top\mathbf{D}[(\mathbf{y}-\mathbf{x}\boldsymbol{\beta})^{-1}]\mathbf{D}[P_1(\nu+1/2,\mathbf{d}/2)^{-1}]\times\\
&\quad\big[(2\nu+3)P_1(\nu+5/2,\mathbf{d}/2)-2P_1(\nu+3/2,\mathbf{d}/2)\\
&\quad-(2\nu+1)\mathbf{D}[P_1(\nu+3/2,\mathbf{d}/2)^2]P_1(\nu+1/2,\mathbf{d}/2)^{-1}\big]\\
I^1_{\sigma^2\sigma^2} &= -\frac{m}{2\sigma^4}-\frac{(2\nu+1)}{4\sigma^4}(P_1(\nu+1/2,\mathbf{d}/2)^{-1})^\top\times\\
&\quad\big[(2\nu+3)P_1(\nu+5/2,\mathbf{d}/2)-4P_1(\nu+3/2,\mathbf{d}/2)\\
&\quad-(2\nu+1)\mathbf{D}[(P_1(\nu+3/2,\mathbf{d}/2)^2)]P_1(\nu+1/2,\mathbf{d}/2)^{-1}\big]\\
I^1_{\lambda\boldsymbol{\tau}} &= \mathbf{0}, \boldsymbol{\tau}=\boldsymbol{\beta},\sigma^2,\lambda.
\end{aligned}
$$

Asymptotic confidence intervals and test on the ML estimators can be obtained using this matrix, that is, if $\mathbf{J}=-\mathbf{I}$ denotes the observed information matrix for the marginal log-likelihood $\ell(\boldsymbol{\theta})$ of the SSMN regression model, then asymptotic confidence intervals and hypotheses tests for the parameter $\boldsymbol{\theta}$ are obtained assuming that the MLE $\boldsymbol{\theta}$ has approximately a $N_{p+3}(\boldsymbol{\theta},\mathbf{J}^{-1})$ distribution. In practice, $\mathbf{J}$ is usually unknown and has to be replaced by the MLE $\widehat{\mathbf{J}}$, that is, the matrix $\widehat{\mathbf{J}}$ evaluated at the MLE $\hat{\boldsymbol{\theta}}$.

6 Applications

In this section, we present two applications. The first one illustrates the use of the distributions SN, SGt, SSL, SNC and SEP in simulation studies, whereas the other one involves the statistical analysis of a real data set.

6.1 Simulation study

SNI distributions can be used in simulation studies as a challenging family for developing statistical procedures in asymmetric situations. As an illustration, we perform a small scale simulation study to study the behavior of two location estimators, the sample mean and the sample median, under four different standard univariate settings. We consider a standard skew-normal $SN(3)$ distribution, a skew-t $SGt(3; 2, 2)$ distribution, a skew-slash $SSL(3; 0.5)$ distribution , a skew power-exponential $SEP(3; 0.5)$ distribution and a skew-contaminated normal $SNC(3; 0.9, 0.1)$ distribution. The location mean of all the asymmetric distributions is adjusted to zero, so that all four distributions are comparable. Thus, this setting represents four distributions with the same mean, but with different tail behaviors and skewness. We simulate 500 samples of size $n = 100$ from each of these five distributions. For each sample, we compute the sample means and the sample median and depict the box-plot for each distribution in Figure 2. In the left panel, we observe that all boxplots of the estimated means are centered around zero but have larger variability for the heavy tailed distributions (skew-t and skew-slash). In the right panel, we see the boxplots of the estimated medians has a slightly larger variability than the boxplots for the estimated means for the skew-normal, skew-slash and skew-contaminated normal, but has a much smaller variability for the the skew-t and skew power-exponential distributions.

This indicates that the median is a robust estimator of location at asymmetric light tailed distributions. On the other hand, the median estimator becomes biased as soon as unexpected skewness and heavier tails arise in the underlying distribution.

6.2 Australian athletes data set revisited

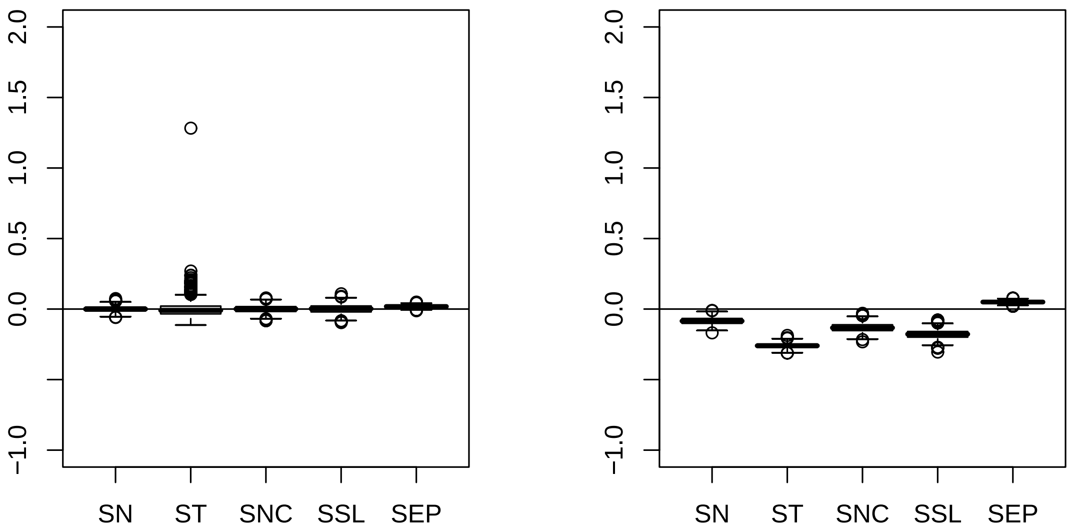
In this section we consider a likelihood analysis of a part of the data set collected on several biomedical variables by the Australian Institute of Sports (AIS data set) on a number of their athletes. A subset of the data set was previously analyzed in Azzalini and Capitanio (2003). We consider a subset of the data set considering a linear regression model relating variables SSF and $bfat$. We define the model

$$SSF_i = \alpha + \beta bfat_i + e_i,$$

$i = 1, \ldots, 202$, where $bfat_i$ is the body fat percentage of the i-th individual in the sample, SSF_i is the sum of skin folds and $e_i \overset{iid}{\sim} SSMN(0, \sigma^2, \lambda; H)$. Arellano–Valle et al. (2005) fitted a skew-normal measurement error model to these data and noted a stronger relationship between the variables and right skewness.

The univariate skew-normal, skew-t-normal, skew-slash, skew-contaminated and skew-exponential power normal distributions are applied to fit the data. Resulting parameter estimates are given in Table 1. The estimated standard errors were calculated using the observed information matrix present is Section 5. The Schwarz information criterion (or equivalently the log-likelihood) was used for choosing among some values of ν and γ as recommended by Fernández and Steel (1999). Note that

Figure 2: Boxplots of the sample mean (left panel) and sample median (right panel) based on 500 samples of sizes n=100 from the five standardized distributions: $SN(3)$; $SGt(3; 2, 2)$; $SNC(3; 0.5, 0.5)$; $SSL(3; 1)$, and $SEP(3; 0.5)$. The respective means are adjusted to zero.



using the log-likelihood values shown on the right of the Table 1 we see that the contaminated skew-normal fits the data better than the other four distributions. This came as a surprise to us because other authors found the skew-normal as presenting good fit which, obviously, is not our main conclusion. The normal skew-model is certainly (well) surpassed the the skew-contaminated model.

Table 1: MLEs of the five models fitted on the AIS data set.

Distr.	$\widehat{\beta}_0$	$\widehat{\beta}_1$	$\widehat{\sigma}^2$	$\widehat{\lambda}$	ν	γ	$l(\widehat{\boldsymbol{\theta}})$
SNC	-0.37(1.35)	4.77(0.10)	65.40(11.42)	0.69(0.30)	0.15	0.2	-720.83
SGt	-2.02(1.88)	4.78(0.14)	90.59(27.01)	1.07(0.58)	7.88	7.88	-721.97
SEP	-4.60(1.58)	4.86(0.14)	129.88(29.82)	1.55(0.31)	0.95	-	-723.18
SN	-5.08(1.57)	4.88(0.14)	143.97(32.76)	1.64(0.61)	-	-	-723.27
SSL	-5.02(1.58)	4.88(0.14)	134.25(30.72)	1.58(0.59)	15.67	-	-723.33

Replacing the ML estimates of $\boldsymbol{\theta}$ in the Mahalanobis distance $d_i = \frac{(y_i - \mathbf{x}_i^\top \boldsymbol{\beta})^2}{\sigma^2}$, we present Q-Q plots and envelopes in Figure 3 (lines represent the 5th percentile, the mean, and the 95th percentile of 100 simulated points for each observation). It seems to us that the plots in Figure (3) provide even stronger evidence (than the log-likelihood criteria), that the skew-contaminated normal distribution provides a better fit to the data set than others SSMN distributions.

7 Final Conclusion

In this work we have defined a new family of asymmetric models by extending the symmetric class of scale mixture of normal distributions . Our proposal generalized results found in Azzalini (1985) and Andrews and Mallows (1973). An important characteristic of the results obtained is that closed form expressions were derived for the iterative estimation processes. This was a consequence of the fact that the proposed distributions posses a stochastic representation that can be used to represent them hierarchically. We believe that the approaches proposed here can also be used to study other asymmetric models. The assessment of influence of data and model assumption on the result of any statistical analysis is a key aspect yet to be studied. Work is in progress addressing specifically local influence and residual analysis.

Appendix

Let Y be a distribution between the SSMN class with location parameter $\mu \in \mathbb{R}$, scale factor σ^2 and skewness parameter $\lambda \in \mathbb{R}$, if its pdf is given by

$$f(y) = 2\int_0^{+\infty} \phi(y|\mu, \sigma^2\kappa(u))\Phi_1(\lambda\frac{y-\mu}{\sigma})dH(u),$$

where U is a positive random variable with cdf $H(u;\boldsymbol{\nu})$. Thus, we have that

$$\begin{aligned} f(y) &= 2\int_0^{+\infty} \phi(y|\mu, \sigma^2\kappa(u))\Phi_1(\lambda(y-\mu)/\sigma)dH(u), \\ &= 2\int_0^{+\infty} \phi(y|\mu, \sigma^2\kappa(u))h(u;\nu)du \int_{-\infty}^{\lambda\frac{y-\mu}{\sigma}} \phi(t|0,1)dt \\ &= 2\int_0^{+\infty}\int_{-\infty}^{0} \phi(y|\mu, \sigma^2\kappa(u))\phi(t|-\lambda\frac{y-\mu}{\sigma}, 1)h(u;\nu)dtdu \\ &= 2\int_0^{+\infty}\int_0^{+\infty} \phi(y|\mu, \sigma^2\kappa(u))\phi(t|\lambda(y-\mu), \sigma^2)h(u;\nu)dtdu \qquad (38) \end{aligned}$$

since

$$\begin{aligned} \int_{-\infty}^{0} \phi(t|-\lambda\frac{y-\mu}{\sigma}, 1) &= P(Z \leq 0), \;\; Z \sim N(-\lambda\frac{y-\mu}{\sigma}, 1) \\ &= P(-Z \geq 0) \\ &= P(T \geq 0), \;\; T \sim N(\lambda\frac{y-\mu}{\sigma}, 1) \\ &= \int_0^{+\infty} \phi(t|\lambda\frac{y-\mu}{\sigma}, 1) \\ &= \int_0^{+\infty} \phi(t|\lambda(y-\mu), \sigma^2). \end{aligned}$$

References

Andrews, D. R. & Mallows, C. L. (1974). Scale mixtures of normal distributions. *J. R. Statist. Soc.*, **B, 36**, 99-102.

Arellano-Valle, R.B., Bolfarine, H. and Lachos, V.H. (2005). Skew-Normal Linear Mixed Models. *Journal of Data Science*, **3**, 415-438.

Arellano-Valle, R. B. and Genton, M.G. (2005). On fundamental skew distributions. *Journal Multivariate Analysis*, **96**, 93-116.

Arellano-Valle, R. B., Ozan, S., Bolfarine, H. and Lachos, V.H. (2005). Skew-Normal Measurement Error Models. *Journal of Multivariate Analysis*, **96**, 93-116.

Arnold, B. C. and Beaver, R. J. (2000). The skew-Cauchy distribution. *Statistics and Probability Letters*, **49**, 285-290.

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

Azzalini, A., Capitanio, A. (1999). Statistical applications of the multivariate skew normal distribution. *Journal of the Royal Statistical Society*, **61**, 579-602.

Azzalini, A. and Dalla-Valle, A. (1996). The multivariate skew-normal distribution. *Biometrika,* **83**, 715-726.

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*, B, 65, 367-389.

Azzalini, A., Dal Capello, T. and Kotz, S. (2003). Log-skew-normal and log-skew-t distributions as model for family income data. *Journal of Income Distribution*, 12-20.

Branco, M. and Dey, D. (2001). A general class of multivariate skew-elliptical distribution. *Journal of Multivariate Analysis*, **79**, 93-113.

Fernández, C. and Steel, M.F.J. (1999). Multivariate Student-t regression models. pitfall and inference. *Biometrika*, **86**, 153-167.

Galea-Rojas, M., de Castillho, M., Bolfarine, H. and de Castro, M. (2003). Detection of analytical bias. *Analyst*, **128**, 1073-1081.

Genton, M. G., Loperfido, N. (2005). Generalized skew-elliptical distributions and their quadratic forms. *Annals of the Institute of Statistical Mathematics*, **57**, 389-401.

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

Gupta, A. K. (2003). Multivariate Skew-t Distribution. *Statistics*, **37**, 359-363.

Gupta, A. K. and Huang, W.-J. (2002). Quadratic forms in skew normal variates. *Journal of Mathematical Analysis and Applications*, **273**, 558-564.

Henze, N. (1986). A probabilistic representation of the "skew-normal" distribution. *Scandinavian Journal of Statistics,* **13**, 271-275.

Johnson, N. L., Kotz, S., and Balakrishnan, N. (1994). *Continuous Univariate Distributions*, vol. 1. New York: John Wiley.

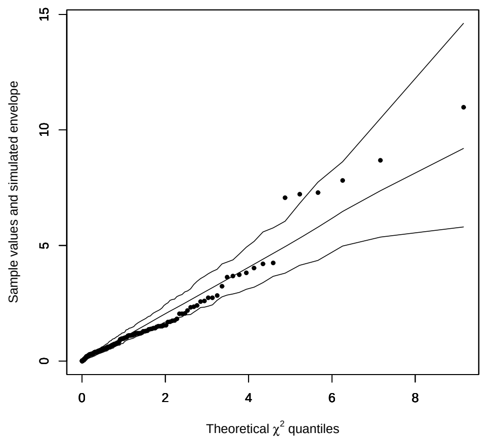
Lachos, V. H. and Bolfarine, H. (2007). Skew probit error-in-variables models. *Statistical Methodology,* **3**, 1-12.

Lange, K. and Sinsheimer. (1993). Normal/independent distributions and their applications in robust regression . *Journal of Computational and Graphical Statistics,*

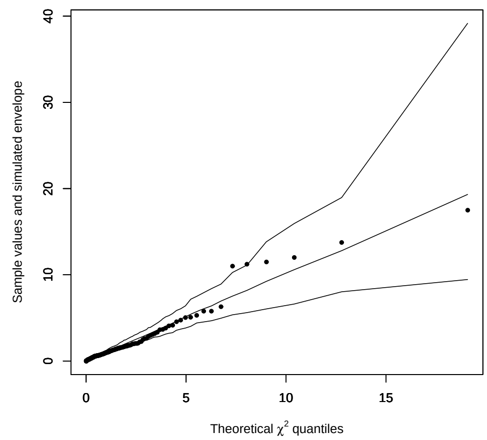
2, 175-198.
Lange, K. L., Little, J. A. and Taylor, M. G. J. (1989). Robust modeling using the t distribution . *Journal of the American Statistical Association,* **84**, 881-896.
Little, R. J. A. (1988). Robust Estimation of the mean and covariance matrix from data with missing values . *Applied Statistics,* **37**, 23-38.
Pewsey, A. (2000). Problems of inference for Azzalini's skew-normal distribution. *Journal of Applied Statistics*, **27**, 859-870.
Sahu, S. K., Dey, D. K. and Branco, M. D. (2003). A new class of multivariate skew distributions with applications to Bayesian regression models. *The Canadian Journal of Statistics*, **31**, 129-150.
Tan, F. and Peng, H. (2006). The slash and the skew-slash Student t distributions. Submitted
Wang, J. and Genton, M. (2006). The multivariate skew-slash distribution. *Journal of Statistical Planning and Inference*, **136**, 209-220.
Zhu, H. and Lee, S. (2001). Local influence for incomplete-data models. *Journal of the Royal Statistical Society,* B, **63**, 111-126.
Zacks, S. (1971). *The Theory of Statistical Inference*. Wiley. New York.

Figure 3: AIS data set. Q-Q plots and simulated envelopes: (a) Skew-normal model (b) Skew-t normal model (c) Skew-contaminated normal (d) Skew-slash model, (e) Skew-power exponential model and (f) profile likelihood for the skew-contaminated normal model.

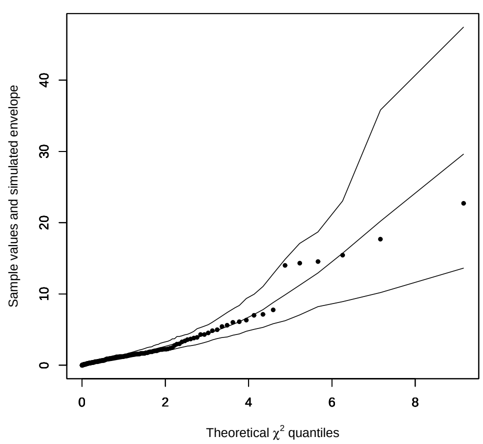
(a)



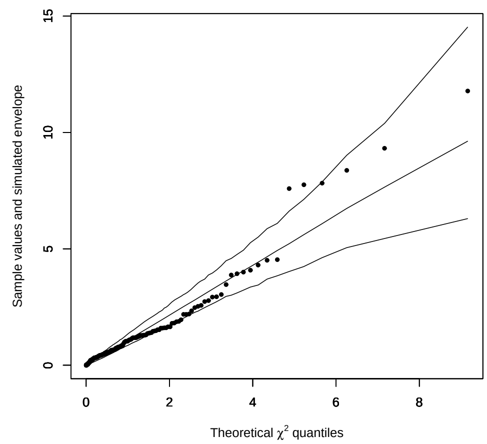
(b)



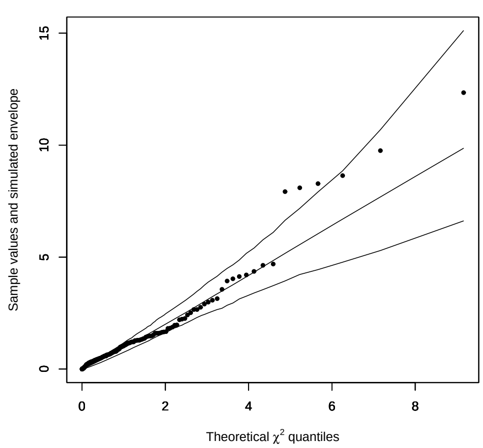
(c)



(d)



(e)



(f)