Multivariate skew-normal/independent distributions: properties and inference

Victor H. Lachos[1,2]*, Filidor V. Labra*[1]

October, 2014

Abstract

Liu (1996) discussed a class of robust normal/independent distributions which contains a group of thick-tailed cases. In this article, we develop a skewed version of these distributions in the multivariate setting, and we call them multivariate skew normal/independent distributions. We derive several useful properties for them. The main virtue of the members of this family is that they are easy to simulate and lend themselves to an EM-type algorithm for maximum likelihood estimation. For two multivariate models of practical interest, the EM-type algorithm has been discussed with emphasis on the skew-t, the skew-slash, and the contaminated skew-normal distributions. Results obtained from simulated and two real data sets are also reported.

MSC(2010): 62F03, 62F05, 62F10, 62F12.

Keywords: EM algorithm, normal/independent distributions, skewness, measurement errors models.

[1] *Departamento de Estatística, IMECC, Universidade Estadual de Campinas, Brazil.*
[2] *Supported by FAPESP.*

Victor H. Lachos, Filidor V. Labra

1. Introduction

A normal distribution is a routine assumption for analyzing real data, but it may be unrealistic, specially when strong skewness and heavy-tailed appear. In practice, we generate a great number of data that are skewed or heavy-tailed, for instance, information on family income, the CD4 cell count from AIDS studies, etc. Thus, one needs to develop a flexible class of models that can readily adapt to the non-normality behavior of certain phenomena. Flexible models that include several known distributions, including normal distribution, are of particular importance, since such models can adapt to distributions that are in the neighborhood of the normal model (DiCiccio and Monti (2004) [13]). Lange and Sinsheimer (1993) [22] developed a normal/independent distribution which contains a group of thick-tailed distributions that is often used for robust inference of symmetrical data (Liu (1996) [25]). In this article we further generalize the normal/independent (NI) distributions and combine skewness with heavy-tailed. These new classes of distributions are attractive not only because they model both cases, but because they have a stochastic representation for easy implementation of the EM-algorithm, and so facilitate the study of many useful properties. Our proposal extends some of the recent results found in Azzalini and Capitanio (2003) [6], Gupta (2003) [17], and Wang and Genton (2006) [31].

Azzalini (1985) [4] proposed a univariate skew-normal distribution that was generalized to the multivariate case by Azzalini and Dalla–Valle (1996) [7] and Arellano–Valle et al. (2005) [2]. The multivariate skew-normal density extends the multivariate normal model by allowing a shape parameter to account for skewness. The probability density function of the generic element of a multivariate skew-normal distribution is given explicitly by

$$f(\mathbf{y}) = 2\phi_p(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\Sigma})\Phi_1(\boldsymbol{\lambda}^\top \boldsymbol{\Sigma}^{-1/2}(\mathbf{y} - \boldsymbol{\mu})), \quad \mathbf{y} \in \mathbb{R}^p, \qquad (1.1)$$

where $\phi_p(.|\boldsymbol{\mu}, \boldsymbol{\Sigma})$ stands for the probability density function of the p-variate normal distribution with mean vector $\boldsymbol{\mu}$ and covariate matrix $\boldsymbol{\Sigma}$, while $\Phi_1(.)$ represents the cumulative distribution function of the

standard normal distribution, here $\boldsymbol{\Sigma}^{-1/2}$ satisfies $\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\Sigma}^{-1/2} = \boldsymbol{\Sigma}^{-1}$. When we have $\boldsymbol{\lambda} = \mathbf{0}$, the skew normal distribution reduces to the normal distribution ($\mathbf{Y} \sim \mathrm{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$). A p-dimensional random vector $\mathbf{Y}$ with probability density function as in (1.1) will be denoted by $\mathrm{SN}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda})$. Its marginal stochastic representation, which can be used to derive several of its properties, is given by

$$\mathbf{Y} \stackrel{d}{=} \boldsymbol{\mu} + \boldsymbol{\Sigma}^{1/2}(\boldsymbol{\delta}|T_0| + (\mathbf{I}_p - \boldsymbol{\delta}\boldsymbol{\delta}^\top)^{1/2}\mathbf{T}_1), \quad \text{with} \quad \boldsymbol{\delta} = \frac{\boldsymbol{\lambda}}{\sqrt{1 + \boldsymbol{\lambda}^\top\boldsymbol{\lambda}}}, \quad (1.2)$$

where $T_0 \sim \mathrm{N}_1(0, 1)$ and $\mathbf{T}_1 \sim \mathrm{N}_p(\mathbf{0}, \mathbf{I}_p)$ are independent, $|T_0|$ denotes of course the absolute value of T_0, and "$\stackrel{d}{=}$" stands for "distributed as". From (1.2) it follows that the expectation and variance of $\mathbf{Y}$ are given, respectively, by

$$E[\mathbf{Y}] = \boldsymbol{\mu} + \sqrt{2/\pi}\boldsymbol{\Sigma}^{1/2}\boldsymbol{\delta} \qquad (1.3)$$

and

$$Var[\mathbf{Y}] = \boldsymbol{\Sigma} - (2/\pi)\boldsymbol{\Sigma}^{1/2}\boldsymbol{\delta}\boldsymbol{\delta}^\top\boldsymbol{\Sigma}^{1/2}. \qquad (1.4)$$

Several extensions of the above model has been proposed. For example we have the skew-t distributions (Sahu et al., (2003) [27], Gupta, (2003) [17]), skew-Cauchy distributions (Arnold and Beaver (2000) [3]), skew-slash distributions (Wang and Genton (2006) [31]), skew-slash-t distributions (Tan and Peng (2006) [29]), and skew-elliptical distributions (Azzalini and Capitanio (1999) [5], Branco and Dey (2001) [9], Sahu et al. (2003) [27], Genton and Loperfido (2005) [16]). In this paper we define a new unified family of asymmetric distributions that offers a much needed flexibility by combining both skewness with heavy-tailed. This family contains, as a special case, the multivariate skew-normal distribution defined by Arellano-Valle et al. (2005) [2], the multivariate skew-slash distribution defined by Wang and Genton (2006) [31], the multivariate skew-t distribution defined by Azzalini and Capitanio (2003) [6], and all the distributions studied by Lange and Sinsheimer (1993) [22] in the symmetric context. Thus, our proposal is a more flexible class than the existing skewed distributions, since it allows easy implementation of inferences in any type of models. We point out that the results and methods provided here are not available elsewhere in the literature.

The plan of the article is as follows. In Section 2, the normal/independent distributions (NI) are reviewed for completeness. In Section 3, the skew-normal normal/independent distributions (SNI) are described, and the main results are presented. In Section 4, we derive the maximum likelihood estimates (MLE) for two important applications of SNI distributions. Analytical expressions for the observed information matrix are worked in Section 5. An illustrative example is presented in Section 6, depicting the usefulness of the proposed methodology. Our concluding remarks are presented in Section 7. We also include an appendix as Section 8.

2. Normal/independent distributions

The symmetric family of NI distributions has attracted much attention in the last few years, mainly because it includes distributions such as the Student-t, the slash, the power exponential, and the contaminated normal distributions. All these distributions have heavier tails than the normal.

We say that a p-dimensional vector $\mathbf{Y}$ has a **NI distribution** with location parameter $\boldsymbol{\mu} \in \mathbb{R}^p$ and positive definite scale matrix $\boldsymbol{\Sigma}$ (see for instance, Lange and Sinsheimer (1993) [22]) if its density function has the form

$$f(\mathbf{y}) = \int_0^\infty \phi_p(y|\boldsymbol{\mu}, u^{-1}\boldsymbol{\Sigma})dH(u;\boldsymbol{\nu}), \qquad (2.1)$$

where $H(u;\boldsymbol{\nu})$ is a cumulative distribution function of a unidimensional positive random variable U indexed by the parameter vector $\boldsymbol{\nu}$. For a random vector with a probability density function as in (2.1), we shall use the notation $\mathbf{Y} \sim \mathrm{NI}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}; H)$. Now, when $\boldsymbol{\mu} = \mathbf{0}$ and $\boldsymbol{\Sigma} = \mathbf{I}_p$, we simply use $\mathbf{Y} \sim \mathrm{NI}_p(H)$.

The stochastic representation of $\mathbf{Y}$ is given by

$$\mathbf{Y} = \boldsymbol{\mu} + U^{-1/2}\mathbf{Z}, \qquad (2.2)$$

with $\mathbf{Z} \sim \mathrm{N}_p(\mathbf{0}, \boldsymbol{\Sigma})$ and U a positive random variable with cumulative distribution function H independent of $\mathbf{Z}$. Examples of NI distributions

are described subsequently (see also Lange and Sinsheimer (1993) [22]). For this family, the distributional properties of the **Mahalanobis distance**

$$d = (\mathbf{y} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{y} - \boldsymbol{\mu}),$$

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

2.1   Examples of NI distributions

● *The Student-t distribution* $\mathbf{Y} \sim t_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$ *with* $\nu > 0$ *degrees of freedom.* The use of the t-distribution as an alternative to the normal distribution has frequently been suggested in the literature. For example Little (1988) [24] and Lange et al. (1989) [23] use the Student-t distribution for robust modeling. The variable $\mathbf{Y}$ has density

$$f(\mathbf{y}) = \frac{\Gamma(\frac{p+\nu}{2})}{\Gamma(\frac{\nu}{2})\pi^{p/2}} \nu^{-p/2} |\boldsymbol{\Sigma}|^{-1/2} \left(1 + \frac{d}{\nu}\right)^{-(\frac{p+\nu}{2})}. \qquad (2.3)$$

In this case, we have $U \sim Gamma(\nu/2, \nu/2)$, where the cumulative distribution function $H(u; \nu)$ has density

$$h(u; \nu) = \frac{(\nu/2)^{\nu/2} u^{\nu/2-1}}{\Gamma(\nu/2)} \exp\left(-\frac{1}{2}\nu u\right), \qquad (2.4)$$

and finite reciprocal moments $E[U^{-m}] = \frac{(\nu/2)^m \Gamma(\nu/2-m)}{\Gamma(\nu/2)}$, for $m < \nu/2$. From Lange and Sinsheimer (1993) [22] we also get

$$d = (\mathbf{y} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{y} - \boldsymbol{\mu}) \sim pF(p, \nu).$$

● *The slash distribution* $\mathbf{Y} \sim SL_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$ *with shape parameter* $\nu > 0$. This distribution presents heavier tails than the normal distribution. It also includes the limiting normal case as $\nu \uparrow \infty$. Its probability density function is given by

$$f(\mathbf{y}) = \nu \int_0^1 u^{\nu-1} \phi_p(\mathbf{y}|\boldsymbol{\mu}, u^{-1}\boldsymbol{\Sigma}) \, du. \qquad (2.5)$$

Here $H(u; \nu)$ has density

$$h(u; \nu) = \nu u^{\nu-1} \mathbb{I}_{(0,1)}, \tag{2.6}$$

with reciprocal moments $E[U^{-m}] = \dfrac{\nu}{\nu - m}$, for $m < \nu$. The Mahalanobis distance has cumulative distribution function given by

$$Pr(d \leq r) = Pr(\chi_p^2 \leq r) - \frac{2^{\nu}\Gamma(p/2 + \nu)}{r^{\nu}\Gamma(p/2)} Pr(\chi_{p+2\nu}^2 \leq r).$$

• *The contaminated normal distribution* $\mathbf{Y} \sim CN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu, \gamma)$, *with* $0 \leq \nu \leq 1$, $0 < \gamma \leq 1$ (Little (1988) [24]). 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 probability density function is

$$f(\mathbf{y}) = \nu\phi_p(\mathbf{y}|\boldsymbol{\mu}, \frac{\boldsymbol{\Sigma}}{\gamma}) + (1 - \nu)\phi_p(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\Sigma}). \tag{2.7}$$

In this case the cumulative distribution function $H(u; \boldsymbol{\nu})$ is given by

$$h(u; \boldsymbol{\nu}) = \nu\mathbb{I}_{(u=\gamma)} + (1 - \nu)\mathbb{I}_{(u=1)}, \qquad \boldsymbol{\nu} = (\nu, \gamma)^{\top}, \tag{2.8}$$

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

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

The power-exponential distribution is the type NI. However, the scale distribution $H(u; \boldsymbol{\nu})$ is not computationally attractive and will not be dealt with in this work.

3. Multivariate SNI distributions and main results

In this section, we define the multivariate SNI distributions and study some of their properties (*v.g.*, moments, kurtosis, linear transformations, and marginal and conditional distributions).

Definition 3.1. A p-dimensional random vector **Y follows a SNI distribution** with location parameter $\boldsymbol{\mu} \in \mathbb{R}^p$, scale matrix $\boldsymbol{\Sigma}$ (a $p \times p$ positive-definite matrix) and skewness parameter $\boldsymbol{\lambda} \in \mathbb{R}^p$ if its probability density function is given by

$$f(\mathbf{y}) = 2 \int_0^\infty \phi_p(\mathbf{y}|\boldsymbol{\mu}, u^{-1}\boldsymbol{\Sigma}) \Phi_1(u^{1/2}\boldsymbol{\lambda}^\top \boldsymbol{\Sigma}^{-1/2}(\mathbf{y} - \boldsymbol{\mu})) \, dH(u) \qquad (3.1)$$

$$= 2 \int_0^\infty \frac{u^{p/2}}{(2\pi)^{p/2}} |\boldsymbol{\Sigma}|^{-1/2} e^{-\frac{u}{2}d_\lambda} \Phi_1(u^{1/2}\boldsymbol{\lambda}^\top \boldsymbol{\Sigma}^{-1/2}(\mathbf{y} - \boldsymbol{\mu})) \, dH(u),$$

where U is a positive random variable with cumulative distribution function $H(u; \boldsymbol{\nu})$.

For a random vector with probability density function as in (3.1), we use the notion $\mathbf{Y} \sim SNI_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}; H)$. When $\boldsymbol{\mu} = \mathbf{0}$ and $\boldsymbol{\Sigma} = \mathbf{I}_p$, we get a standard SNI distribution and denote it by $SNI_p(\boldsymbol{\lambda}; H)$.

It is clear from (3.1) that when $\boldsymbol{\lambda} = \mathbf{0}$, we get back the NI distribution defined in (2.1). For a random vector with probability density function as in (3.1), we write the **Mahalanobis distance** as

$$d_\lambda = (\mathbf{y} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{y} - \boldsymbol{\mu}).$$

In Definition 3.1, note that the cumulative distribution function $H(u; \boldsymbol{\nu})$ is indexed by the vector $\boldsymbol{\nu}$. Thus, if we suppose that $\boldsymbol{\nu}_\infty$ is such that $\boldsymbol{\nu} \uparrow \boldsymbol{\nu}_\infty$, and $H(u; \boldsymbol{\nu})$ 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 (3.1) converges to the density function of a random vector having a skew-normal distribution. The proof of this result is similar to the one present in Lange and Sinsheimer (1993) [22] for the NI case.

For a SNI random vector, the stochastic representation given below can be used to quickly simulate pseudo-realizations of **Y**, and also to study many of their properties.

Proposition 3.2. *For* $\mathbf{Y} \sim SNI_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}; H)$ *we have*

$$\mathbf{Y} \stackrel{d}{=} \boldsymbol{\mu} + U^{-1/2}\mathbf{Z}, \qquad (3.2)$$

with $\mathbf{Z} \sim SN_p(\mathbf{0}, \boldsymbol{\Sigma}, \boldsymbol{\lambda})$ *and* U *a positive random variable with cumulative distribution function* H *independent of* **Z**. *(Compare Equation (1.1).)*

Proof. This follows from the hypothesis $\mathbf{Y}|U = u \sim \mathrm{SN}_p(\boldsymbol{\mu}, u^{-1}\boldsymbol{\Sigma}, \boldsymbol{\lambda})$. $\qquad\square$

Notice that the stochastic representation given in (2.2) for the NI case is a specialization of (3.2) for $\boldsymbol{\lambda} = \mathbf{0}$. Hence, we have extended the family of NI distributions to the skewed case. Besides, from (1.2) it follows that (3.2) can be written as

$$\mathbf{Y} \overset{d}{=} \boldsymbol{\mu} + \frac{1}{U^{1/2}}\boldsymbol{\Sigma}^{1/2}\{\boldsymbol{\delta}|X_0| + (\mathbf{I}_n - \boldsymbol{\delta}\boldsymbol{\delta}^T)^{1/2}\mathbf{X}_1\}, \tag{3.3}$$

where $\boldsymbol{\delta} = \boldsymbol{\lambda}/\sqrt{1 + \boldsymbol{\lambda}^\top\boldsymbol{\lambda}}$, and U, $X_0 \sim N_1(0,1)$ and $\mathbf{X}_1 \sim N_p(\mathbf{0}, \mathbf{I}_p)$ are independent. The marginal stochastic representation given in (3.3) is very important since it allows us to implement the EM-algorithm for a wide class of linear models similar to those of Lachos et al. (2007) [20].

In the next proposition, we derive a general expression for the moment generating function of a SNI random vector.

Proposition 3.3. *For* $\mathbf{Y} \sim SNI_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}; H)$ *and* $\mathbf{s} \in \mathbb{R}^p$, *we have*

$$M_{\mathbf{Y}}(\mathbf{s}) = E[e^{\mathbf{s}^\top\mathbf{Y}}] = \int_0^\infty 2e^{\mathbf{s}^\top\boldsymbol{\mu} + \frac{1}{2}u^{-1}\mathbf{s}^\top\boldsymbol{\Sigma}\mathbf{s}}\Phi_1(u^{-1/2}\boldsymbol{\delta}^\top\boldsymbol{\Sigma}^{1/2}\mathbf{s})dH(u). \tag{3.4}$$

Proof. From Proposition 3.2, we obtain $\mathbf{Y}|U = u \sim SN_p(\boldsymbol{\mu}, u^{-1}\boldsymbol{\Sigma}, \boldsymbol{\lambda})$. Next we get $M_{\mathbf{Y}}(\mathbf{s}) = E_U[E[e^{\mathbf{s}^\top\mathbf{Y}}|U]]$ from well known properties of conditional expectation. As U is a positive random variable with cumulative distribution function H, we derive the proof from the fact that $\mathbf{Z} \sim SN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda})$ implies $M_{\mathbf{Z}}(\mathbf{s}) = 2e^{\mathbf{s}^\top\boldsymbol{\mu} + \frac{1}{2}\mathbf{s}^\top\boldsymbol{\Sigma}\mathbf{s}}\Phi_1(\boldsymbol{\delta}^\top\boldsymbol{\Sigma}^{1/2})$. $\qquad\square$

The next proposition shows that a SNI random vector is invariant under linear transformations. This, in turn, implies that the marginal distributions of $\mathbf{Y} \sim \mathrm{SNI}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}; H)$ are still SNI.

Proposition 3.4. *Let* $\mathbf{Y} \sim SNI_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}; H)$. *Then for any fixed vector* $\mathbf{b} \in \mathbb{R}^m$ *and matrix* $\mathbf{A} \in \mathbb{R}^{m \times p}$ *of full row rank we get*

$$\mathbf{V} = \mathbf{b} + \mathbf{A}\mathbf{Y} \sim SNI_p(\mathbf{b} + \mathbf{A}\boldsymbol{\mu}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^\top, \boldsymbol{\lambda}^*; H), \tag{3.5}$$

here we have $\boldsymbol{\lambda}^* = \boldsymbol{\delta}^*/(1 - \boldsymbol{\delta}^{*\top}\boldsymbol{\delta}^*)^{1/2}$ *with* $\boldsymbol{\delta}^* = (\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^\top)^{-1/2}\mathbf{A}\boldsymbol{\Sigma}^{1/2}\boldsymbol{\delta}$. *Moreover, if* $m = p$ *and* $\mathbf{A}$ *is non-singular, then we get* $\boldsymbol{\lambda}^* = \boldsymbol{\lambda}$. *Also, for any* $\mathbf{a} \in \mathbb{R}^p$, *we obtain*

$$\mathbf{a}^\top\mathbf{Y} \sim SNI_p(\mathbf{a}^\top\boldsymbol{\mu}, \mathbf{a}^\top\boldsymbol{\Sigma}\mathbf{a}, \lambda^*; H),$$

where $\lambda^* = \alpha/(1 - \alpha^2)^{1/2}$, *with* $\alpha = \{\mathbf{a}^\top\boldsymbol{\Sigma}\mathbf{a}(1 + \boldsymbol{\lambda}^\top\boldsymbol{\lambda})\}^{-1/2}\mathbf{a}^\top\boldsymbol{\Sigma}^{1/2}\boldsymbol{\lambda}$.

Proof. The proof of this result is direct from Proposition 3.3 since we have $M_{\mathbf{b}+\mathbf{A}\mathbf{Y}}(\mathbf{s}) = e^{\mathbf{s}^\top\mathbf{b}}M_{\mathbf{Y}}(\mathbf{A}^\top\mathbf{s})$. When $\mathbf{A}$ is non-singular, it is easy to see that $\boldsymbol{\delta}^* = \boldsymbol{\delta}$ holds. $\qquad\square$

Applying Proposition 3.4 to $\mathbf{A} = [\mathbf{I}_{p_1}, \mathbf{0}_{p_2}]$, with $p_1 + p_2 = p$, we obtain the following additional properties of a SNI random vector, related to the marginal distribution this time.

Corollary 3.5. *Let* $\mathbf{Y} \sim SNI_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}; H)$ *and suppose* $\mathbf{Y}$ *is partitioned as* $\mathbf{Y}^\top = (\mathbf{Y}_1^\top, \mathbf{Y}_2^\top)^\top$ *of dimensions* p_1 *and* $p_2 = p - p_1$, *respectively. Let*

$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}, \quad \boldsymbol{\mu} = (\boldsymbol{\mu}_1^\top, \boldsymbol{\mu}_2^\top)^\top$$

be the corresponding partitions of $\boldsymbol{\Sigma}$ *and* $\boldsymbol{\mu}$. *Then, the marginal density of* $\mathbf{Y}_1$ *is* $SNI_{p_1}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11}, \boldsymbol{\Sigma}_{11}^{1/2}\widetilde{\boldsymbol{v}}; H)$, *where* $\widetilde{\boldsymbol{v}} = \dfrac{\boldsymbol{v}_1 + \boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}\boldsymbol{v}_2}{\sqrt{1 + \boldsymbol{v}_2^\top\boldsymbol{\Sigma}_{22.1}\boldsymbol{v}_2}}$, *with* $\boldsymbol{\Sigma}_{22.1} = \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}$ *and* $\boldsymbol{v} = \boldsymbol{\Sigma}^{-1/2}\boldsymbol{\lambda} = (\boldsymbol{v}_1^\top, \boldsymbol{v}_2^\top)^\top$. $\qquad\square$

Proposition 3.6. *Let* $\mathbf{Y} \sim SNI_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}; H)$. *Then the distribution of* $\mathbf{Y}_2$, *conditionally on* $\mathbf{Y}_1 = \mathbf{y}_1$ *and* $U = u$, *has density*

$$f(\mathbf{y}_2|\mathbf{y}_1, u) = \phi_{p_2}(\mathbf{y}_2|\boldsymbol{\mu}_{2.1}, u^{-1}\boldsymbol{\Sigma}_{22.1})\frac{\Phi_1(u^{1/2}\boldsymbol{v}^\top(\mathbf{y} - \boldsymbol{\mu}))}{\Phi_1(u^{1/2}\widetilde{\boldsymbol{v}}^\top(\mathbf{y}_1 - \boldsymbol{\mu}_1))}, \tag{3.6}$$

with $\boldsymbol{\mu}_{2.1} = \boldsymbol{\mu}_2 + \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}(\mathbf{y}_1 - \boldsymbol{\mu}_1)$. *Furthermore, we get*

$$E[\mathbf{Y}_2|\mathbf{y}_1, u] = \boldsymbol{\mu}_{2.1} + u^{-1/2}\frac{\phi_1(u^{1/2}\widetilde{\boldsymbol{v}}^\top(\mathbf{y}_1 - \boldsymbol{\mu}_1))}{\Phi_1(u^{1/2}\widetilde{\boldsymbol{v}}^\top(\mathbf{y}_1 - \boldsymbol{\mu}_1))}\frac{\boldsymbol{\Sigma}_{22.1}\boldsymbol{v}_2}{\sqrt{1 + \boldsymbol{v}_2^\top\boldsymbol{\Sigma}_{22.1}\boldsymbol{v}_2}}. \tag{3.7}$$

Proof. As we have $f(\mathbf{y}_2|\mathbf{y}_1, u) = f(\mathbf{y}|u)/f(\mathbf{y}_1|u)$, Formula (3.6) for the density follows after noticing $\mathbf{Y}|U = u \sim SN_p(\boldsymbol{\mu}, u^{-1}\boldsymbol{\Sigma}, \boldsymbol{\lambda})$ and $\mathbf{Y}_1|U = u \sim SN(\boldsymbol{\mu}_1, u^{-1}\,\boldsymbol{\Sigma}_{11}, \boldsymbol{\Sigma}_{11}^{1/2}\widetilde{\boldsymbol{v}})$. The expectation suggested in (3.7) is confirmed by Lemma 8.2 (in the appendix) if we take $A = \boldsymbol{v}_1^\top(\mathbf{y}_1 - \boldsymbol{\mu}_1) - \boldsymbol{v}_2^\top\boldsymbol{\mu}_2$, $\mathbf{B} = \boldsymbol{v}_2$, $\boldsymbol{\mu} = \boldsymbol{\mu}_{2.1}$, and $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_{22.1}$. This concludes the proof. $\square$

Note that given u, when we have $\boldsymbol{\Sigma}_{21} = \mathbf{0}$ and $\boldsymbol{\lambda}_2 = \mathbf{0}$, it is possible to obtain independence for the components $\mathbf{Y}_1$ and $\mathbf{Y}_2$ of a SNI random vector $\mathbf{Y}$. The following corollary is a by-product of Proposition 3.6, since we have $E[\mathbf{Y}_2|\mathbf{y}_1] = E_U[E[\mathbf{Y}_2|\mathbf{y}_1, U]|\mathbf{y}_1]$.

Proposition 3.7. *For* $\mathbf{Y} \sim SNI_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}; H)$ *the first moment of* $\mathbf{Y}_2$, *conditionally on* $\mathbf{Y}_1 = \mathbf{y}_1$, *is given by*

$$E[\mathbf{Y}_2|\mathbf{y}_1] = \boldsymbol{\mu}_{2.1} + \frac{\boldsymbol{\Sigma}_{22.1}\boldsymbol{v}_2}{\sqrt{1 + \boldsymbol{v}_2^\top\boldsymbol{\Sigma}_{22.1}\boldsymbol{v}_2}} E[U^{-1/2}\frac{\phi_1(U^{1/2}\widetilde{\boldsymbol{v}}^\top(\mathbf{y}_1 - \boldsymbol{\mu}_1))}{\Phi_1(U^{1/2}\widetilde{\boldsymbol{v}}^\top(\mathbf{y}_1 - \boldsymbol{\mu}_1))}\mathbf{y}_1],$$

with $\boldsymbol{\mu}_{2.1} = \boldsymbol{\mu}_2 + \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}(\mathbf{y}_1 - \boldsymbol{\mu}_1)$. $\square$

The next result can be useful in applications to linear models. For instance, we can use it when the linear model depends on a vector of unobservable random effects and a vector of random errors (linear mixed model) in which the random effects are assumed to have a SNI distribution and the errors are assumed to have a NI distribution.

Proposition 3.8. *Suppose we have* $\mathbf{X} \sim SNI_m(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1, \boldsymbol{\lambda}, H)$ *and* $\mathbf{Y} \sim NI_p(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2, H)$. *If there is a positive random variable* U *with cumulative distribution function* H *so that we can write* $\mathbf{X} \stackrel{d}{=} \boldsymbol{\mu}_1 + U^{-1/2}\mathbf{Z}$ *and* $\mathbf{Y} \stackrel{d}{=} \boldsymbol{\mu}_2 + U^{-1/2}\mathbf{W}$, *with* $\mathbf{Z} \sim SN_m(\mathbf{0}, \boldsymbol{\Sigma}_1, \boldsymbol{\lambda})$ *independent of* $\mathbf{W} \sim N_p(\mathbf{0}, \boldsymbol{\Sigma}_2)$, *then for any matrix* $\mathbf{A}$ *of dimension* $p \times m$ *we have*

$$\mathbf{A}\mathbf{X} + \mathbf{Y} \sim SNI_m(\mathbf{A}\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2, \mathbf{A}\boldsymbol{\Sigma}_1\mathbf{A}^\top + \boldsymbol{\Sigma}_2, \boldsymbol{\lambda}_*; H),$$

here $\boldsymbol{\lambda}_* = \boldsymbol{\delta}_*/\sqrt{1 - \boldsymbol{\delta}_*^\top\boldsymbol{\delta}_*}$, *with* $\boldsymbol{\delta}_* = (\mathbf{A}\boldsymbol{\Sigma}_1\mathbf{A}^\top + \boldsymbol{\Sigma}_2)^{-1/2}\mathbf{A}\boldsymbol{\Sigma}_1^{1/2}\boldsymbol{\delta}$.

Proof. The proof is based on Proposition 3.3. Note first that surely $\mathbf{X}$

and $\mathbf{Y}$ are independent. Next, letting $\mathbf{V} = \mathbf{AX} + \mathbf{Y}$ we obtain

$$
\begin{aligned}
M_{\mathbf{V}}(\mathbf{s}) &= E_U(E[e^{\mathbf{s}^\top \mathbf{AX}}|U]E[e^{\mathbf{s}^\top \mathbf{Y}}|U]) \\
&= \int_0^\infty 2e^{\mathbf{s}^\top \mathbf{A}\boldsymbol{\mu}_1 + \frac{1}{2u}\mathbf{s}^\top \mathbf{A}\boldsymbol{\Sigma}_1 \mathbf{A}^\top \mathbf{s}} \Phi_1\left(\frac{\boldsymbol{\delta}^\top \boldsymbol{\Sigma}_1^{1/2}\mathbf{A}^\top \mathbf{s}}{\sqrt{u}}\right) e^{\mathbf{s}^\top \boldsymbol{\mu}_2 + \frac{1}{2u}\mathbf{s}^\top \boldsymbol{\Sigma}_2 \mathbf{s}}\, dH(u) \\
&= \int_0^\infty 2e^{\mathbf{s}^\top (\mathbf{A}\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2) + \frac{1}{2u}\mathbf{s}^\top (\mathbf{A}\boldsymbol{\Sigma}_1 \mathbf{A}^\top + \boldsymbol{\Sigma}_2)\mathbf{s}} \Phi_1\left(\frac{\boldsymbol{\delta}^\top \boldsymbol{\Sigma}_1^{1/2}\mathbf{A}^\top \mathbf{s}}{\sqrt{u}}\right) dH(u) \\
&= \int_0^\infty 2e^{\mathbf{t}^\top (\mathbf{A}\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2) + \frac{1}{2u}\mathbf{s}^\top (\mathbf{A}\boldsymbol{\Sigma}_1 \mathbf{A}^\top + \boldsymbol{\Sigma}_2)\mathbf{s}} \Phi_1\left(\frac{\boldsymbol{\delta}_*^\top \boldsymbol{\Psi}^{1/2}\mathbf{s}}{\sqrt{u}}\right) dH(u),
\end{aligned}
$$

where $\boldsymbol{\Psi} = \mathbf{A}\boldsymbol{\Sigma}_1 \mathbf{A}^\top + \boldsymbol{\Sigma}_2$, $\boldsymbol{\delta}_* = \boldsymbol{\Psi}^{-1/2}\mathbf{A}\boldsymbol{\Sigma}_1^{1/2}\boldsymbol{\delta}$, and the proof follows from Proposition 3.3. $\qquad\square$

In the following proposition we derive the mean and the covariance matrix of a SNI random vector. Furthermore, we present the multidimensional kurtosis coefficient for a random vector SNI, which represent a extension of the kurtosis coefficient proposed by Azzalini and Capitanio (1999) [5].

Proposition 3.9. *Suppose we have* $\mathbf{Y} \sim SNI_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}; H)$*. Then the following conditions hold.*

a) If $E[U^{-1/2}] < \infty$*, then we have*

$$
E[\mathbf{Y}] = \boldsymbol{\mu} + \sqrt{\frac{2}{\pi}} E[U^{-1/2}]\boldsymbol{\Sigma}^{1/2}\boldsymbol{\delta}.
$$

b) If $E[U^{-1}] < \infty$*, then we have*

$$
Var[\mathbf{Y}] = \boldsymbol{\Sigma}_y = E[U^{-1}]\boldsymbol{\Sigma} - \frac{2}{\pi} E^2[U^{-1/2}]\boldsymbol{\Sigma}^{1/2}\boldsymbol{\delta}\boldsymbol{\delta}^\top \boldsymbol{\Sigma}^{1/2}.
$$

c) If $E[U^{-2}] < \infty$*, then the multidimensional kurtosis coefficient is*

$$
\gamma_2(\mathbf{Y}) = \frac{E[U^{-2}]}{E^2[U^{-1}]}a_{1y} - 4\frac{E[U^{-3/2}]}{E^2[U^{-1}]}a_{2y} + a_{3y} - p(p+2),
$$

here

$$
\begin{aligned}
a_{1y} &= p(p+2) + 2(p+2)\boldsymbol{\mu}_y^\top \boldsymbol{\Sigma}_y^{-1} \boldsymbol{\mu}_y + 3(\boldsymbol{\mu}_y^\top \boldsymbol{\Sigma}_y^{-1} \boldsymbol{\mu}_y)^2, \\
a_{2y} &= \left(p + \frac{2}{E[U^{-1/2}]} \right) \boldsymbol{\mu}_y^\top \boldsymbol{\Sigma}_y^{-1} \boldsymbol{\mu}_y \\
&\qquad + \left(1 + \frac{2}{E[U^{-1/2}]} - \frac{\pi}{2} \frac{E[U^{-1}]}{E^2[U^{-1/2}]} \right) (\boldsymbol{\mu}_y^\top \boldsymbol{\Sigma}_y^{-1} \boldsymbol{\mu}_y)^2, \\
a_{3y} &= 2(p+2)\boldsymbol{\mu}_y^\top \boldsymbol{\Sigma}_y^{-1} \boldsymbol{\mu}_y + 3(\boldsymbol{\mu}_y^\top \boldsymbol{\Sigma}_y^{-1} \boldsymbol{\mu}_y)^2,
\end{aligned}
$$

where $\boldsymbol{\mu}_y = E[\mathbf{Y} - \boldsymbol{\mu}] = \sqrt{\frac{2}{\pi}} E[U^{-1/2}] \boldsymbol{\Sigma}^{1/2} \boldsymbol{\delta}$.

Proof. The proof of a) and b) follows from Proposition 3.2. To obtain the expression in c) we use the definition of the multivariate kurtosis introduced by Mardia (1974) [26]. Without loss of generality we take $\boldsymbol{\mu} = \mathbf{0}$, so to get $\boldsymbol{\mu}_y = E[\mathbf{Y}] = \sqrt{\frac{2}{\pi}} E[U^{-1/2}] \boldsymbol{\Sigma}^{1/2} \boldsymbol{\delta}$. Note first that the kurtosis is defined by $\gamma_2(\mathbf{Y}) = E[\{(\mathbf{Y} - \boldsymbol{\mu}_y)^\top \boldsymbol{\Sigma}_y^{-1} \mathbf{Y} - \boldsymbol{\mu}_y)\}^2]$. Now, by using the stochastic representation of $\mathbf{Y}$ given in (2.2) we obtain

$$
(\mathbf{Y} - \boldsymbol{\mu}_y)^\top \boldsymbol{\Sigma}_y^{-1} \mathbf{Y} - \boldsymbol{\mu}_y) \stackrel{d}{=} U^{-1}\mathbf{Z}^\top \boldsymbol{\Sigma}_y^{-1}\mathbf{Z} - 2U^{-1/2}\mathbf{Z}^\top \boldsymbol{\Sigma}_y^{-1}\boldsymbol{\mu}_y + \boldsymbol{\mu}_y^\top \boldsymbol{\Sigma}_y^{-1}\boldsymbol{\mu}_y,
$$

where $\mathbf{Z} \sim SN_p(\mathbf{0}, \boldsymbol{\Sigma}, \boldsymbol{\lambda})$. Due to the definition of $\gamma_2(\mathbf{Y})$, the proof follows after some algebraic manipulations involving the first two moments of a quadratic form (see Genton, He and Liu, (2001) [15]) and Lemma 8.1. $\qquad\square$

Note that under the skew-normal distribution condition, i.e, when $U = 1$, the multidimensional kurtosis coefficient reduces to $\gamma_2(\mathbf{Y}) = 2(\pi - 3)(\boldsymbol{\mu}_y^\top \boldsymbol{\Sigma}_y^{-1}\boldsymbol{\mu}_y)^2$, which is the kurtosis coefficient for a skew-normal random vector (see for instance, Azzalini and Capitanio (1999) [5]).

Proposition 3.10. *If* $\mathbf{Y} \sim SNI_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}; H)$, *then for any even function g the distribution of* $g(\mathbf{Y} - \boldsymbol{\mu})$ *does not depend on* $\boldsymbol{\lambda}$ *and has the same distribution as* $g(\mathbf{X} - \boldsymbol{\mu})$, *where* $\mathbf{X} \sim NI_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}; H)$. *In particular, if* $\mathbf{A}$ *is a* $p \times p$ *symmetric matrix, then* $(\mathbf{Y} - \boldsymbol{\mu})^\top \mathbf{A}(\mathbf{Y} - \boldsymbol{\mu})$ *and* $(\mathbf{X} - \boldsymbol{\mu})^\top \mathbf{A}(\mathbf{X} - \boldsymbol{\mu})$ *are identically distributed.*

Proof. The proof follows from Proposition 3.3; a similar procedure can be found in Wang et al. (2004) [30]. □

As a by-product of Proposition 3.10 we have the following interesting result.

Corollary 3.11. *Let* $\mathbf{Y} \sim SNI_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}; H)$. *Then the quadratic form*

$$d_\lambda = (\mathbf{Y} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{Y} - \boldsymbol{\mu})$$

has the same distribution as $d = (\mathbf{X} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{X} - \boldsymbol{\mu})$, *where* $\mathbf{X} \sim NI_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}; H)$. □

Corollary 3.11 is interesting because it allows us in practice to check models (see Section 5). On the other hand, Corollary 3.11 together with a result from Lange and Sinsheimer (1993) [22, Section 2] allows us to obtain the m-th moment of d_λ.

Corollary 3.12. *Let* $\mathbf{Y} \sim SNI_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}; H)$. *Then for any* $m > 0$ *we have*

$$E[d_\lambda^m] = \frac{2^m \Gamma(m + p/2)}{\Gamma(p/2)} E[U^{-m}].$$

□

3.1   Examples of SNI distributions

We provide several examples of SNI distributions.

• *The skew-t distribution* $ST_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \nu)$ *with* ν *degrees of freedom.* Consider $U \sim Gamma(\nu/2, \nu/2)$. Similar procedures to those of Gupta (2003) [17, Section 2] lead us to the density function

$$f(\mathbf{y}) = 2t_p(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu) T_1 \left(\frac{\sqrt{v + p}\, \boldsymbol{\lambda}^\top \boldsymbol{\Sigma}^{-1/2}(\mathbf{y} - \boldsymbol{\mu})}{\sqrt{d + \nu}} \Big| 0, 1, \nu + p \right), \ \mathbf{y} \in \mathbb{R}^p, \tag{3.8}$$

where, as usual, $t_p(\cdot|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$ and $T_p(\cdot|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$ denote, respectively, the probability density function and cumulative distribution function of the

Student-t distribution $t_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$ as defined in (2.3). Absorbed by the skew-t distribution is the skew-Cauchy distribution when $\nu = 1$. Also, when $\nu \uparrow \infty$, we recover the skew-normal distribution as the limiting case; see Gupta (2003) [17] for further details. In this case, from Proposition 3.9, the mean and covariance matrix of $\mathbf{Y} \sim ST_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \nu)$ are given by

$$E[\mathbf{Y}] = \boldsymbol{\mu} + \sqrt{\nu/\pi}\,\frac{\Gamma(\frac{\nu-1}{2})}{\Gamma(\frac{\nu}{2})}\boldsymbol{\Sigma}^{1/2}\boldsymbol{\delta}, \;\; \nu > 1$$

and

$$Var[\mathbf{Y}] = \frac{\nu}{\nu - 2}\boldsymbol{\Sigma} - (\nu/\pi)\,\big(\frac{\Gamma(\frac{\nu-1}{2})}{\Gamma(\frac{\nu}{2})}\big)^2\boldsymbol{\Sigma}^{1/2}\boldsymbol{\delta}\boldsymbol{\delta}^{\top}\boldsymbol{\Sigma}^{1/2}, \;\; \nu > 2.$$

In what follows we give an important result which will be used in the implementation of the EM algorithm.

Proposition 3.13. *If $\mathbf{Y} \sim ST_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \nu)$, then we have*

$$E[U^r|\mathbf{y}] = \frac{2^{r+1}\nu^{\nu/2}\Gamma(\frac{p+\nu+2r}{2})(d+\nu)^{-\frac{p+\nu+2r}{2}}}{f(\mathbf{y})\Gamma(\nu/2)\sqrt{\pi^p}|\boldsymbol{\Sigma}|^{1/2}}T_1\big(\sqrt{\tfrac{p+\nu+2r}{d+\nu}}\,A\,|\,0, 1, p + \nu + 2r\big)$$

and

$$E[U^r W_{\Phi_1}(U^{1/2}A)] = \frac{2^{r+1/2}\nu^{\nu/2}\Gamma(\frac{p+\nu+2r}{2})(d+\nu+A^2)^{-\frac{p+\nu+2r}{2}}}{f(\mathbf{y})\Gamma(\nu/2)\sqrt{\pi^{p+1}}|\boldsymbol{\Sigma}|^{1/2}}.$$

where $A = \boldsymbol{\lambda}^{\top}\boldsymbol{\Sigma}^{-1/2}(\mathbf{y} - \boldsymbol{\mu})$ and $W_{\Phi_1}(x) = \phi_1(x)/\Phi_1(x)$, for $x \in \mathbb{R}$.

Proof. The proof follows from Lemma 1 in Azzalini and Capitanio (2003) [6, Lemma 1] as we have $f(u|\mathbf{y}) = f(\mathbf{y}, u)/f(\mathbf{y})$ plus

$$E[U^r|\mathbf{y}] = \frac{2}{f(\mathbf{y})}\int_0^{\infty} u^r \phi_p(\mathbf{y}|\boldsymbol{\mu}, u^{-1}\boldsymbol{\Sigma})\Phi_1(u^{1/2}A)G_u(\nu/2, \nu/2)du$$

and

$$E[U^r W_{\Phi_1}(U^{1/2}A)] = \frac{2}{f(\mathbf{y})}\int_0^{\infty} u^r \phi_p(\mathbf{y}|\boldsymbol{\mu}, u^{-1}\boldsymbol{\Sigma})\phi_1(u^{1/2}A)G_u(\nu/2, \nu/2)du,$$

here the probability density function of the $Gamma(\frac{\nu}{2}, \frac{\nu}{2})$ distribution is given by $G_u(\nu/2, \nu/2)$. $\square$

For a skew-t random vector $\mathbf{Y}$, partitioned as $\mathbf{Y}^\top = (\mathbf{Y}_1^\top, \mathbf{Y}_2^\top)^\top$, we have from Corollary 1 that $\mathbf{Y}_1 \sim ST_{p_1}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11}, \boldsymbol{\Sigma}_{11}^{1/2}\widetilde{\boldsymbol{v}}, \nu)$ holds. Thus, from Proposition 3.7 we have the following result.

Corollary 3.14. *For $\mathbf{Y} \sim ST_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \nu)$ we have*

$$
\begin{aligned}
E[\mathbf{Y}_2|\mathbf{y}_1] &= \boldsymbol{\mu}_{2.1} + \frac{\boldsymbol{\Sigma}_{22.1}\boldsymbol{v}_2}{\sqrt{1 + \boldsymbol{v}_2^\top \boldsymbol{\Sigma}_{22.1}\boldsymbol{v}_2}} \frac{\nu^{\nu/2}\Gamma(\frac{\nu+p_1-1}{2})}{\Gamma(\nu/2)\sqrt{\pi}^{(p_1+1)}|\boldsymbol{\Sigma}_{11}|^{1/2}} \times \\
&\quad \frac{1}{f(\mathbf{y}_1)}(\nu + d_{y_1} + (\widetilde{\boldsymbol{v}}^\top(\mathbf{y}_1 - \boldsymbol{\mu}_1))^2)^{-\frac{\nu+p_1-1}{2}},
\end{aligned}
$$

where $d_{y_1} = (\mathbf{y}_1 - \boldsymbol{\mu}_1)^\top \boldsymbol{\Sigma}_{11}^{-1}(\mathbf{y}_1 - \boldsymbol{\mu}_1)$. $\qquad\square$

- *The skew-slash distribution $SSL_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \nu)$ with the shape parameter $\nu > 0$. With $h(u;\nu)$ as in (2.6), from Proposition 3.2 can easily be derived*

$$
f(\mathbf{y}) = 2\nu \int_0^1 u^{\nu-1}\phi_p(\mathbf{y}|\boldsymbol{\mu}, \frac{\boldsymbol{\Sigma}}{u})\Phi_1(u^{1/2}\boldsymbol{\lambda}^\top\boldsymbol{\Sigma}^{-1/2}(\mathbf{y} - \boldsymbol{\mu})), \quad \mathbf{y} \in \mathbb{R}^p,
$$
(3.9)

The skew-slash distribution becomes the skew-normal distribution when $\nu \uparrow \infty$. See Wang and Genton (2006) [31] for further details. In this particular case, from Proposition 3.9 we get

$$
E[\mathbf{Y}] = \boldsymbol{\mu} + \sqrt{\frac{2}{\pi}}\frac{2\nu}{2\nu-1}\boldsymbol{\Sigma}^{1/2}\boldsymbol{\delta}, \ \nu > 1/2
$$

and

$$
Var[\mathbf{Y}] = \frac{\nu}{\nu-1}\boldsymbol{\Sigma} - \frac{2}{\pi}(\frac{2\nu}{2\nu-1})^2\boldsymbol{\Sigma}^{1/2}\boldsymbol{\delta}\boldsymbol{\delta}^\top\boldsymbol{\Sigma}^{1/2}, \ \nu > 1.
$$

As in the skew-t case we have the following results.

Proposition 3.15. *For $\mathbf{Y} \sim SSL_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \nu)$ we get*

$$
E[U^r|\mathbf{y}] = \frac{2^{\nu+r+1}\nu\Gamma(\frac{p+2\nu+2r}{2})P_1(\frac{p+2\nu+2r}{2}, \frac{d}{2})d^{-\frac{p+2\nu+2r}{2}}}{f(\mathbf{y})\sqrt{\pi^p}|\boldsymbol{\Sigma}|^{1/2}}E[\Phi(S^{1/2}A)],
$$

where $S_i \sim Gamma(\dfrac{p + 2\nu + 2r}{2}, \dfrac{d}{2})\mathbb{I}_{(0,1)}$, and

$$E[U^r W_{\Phi_1}(U^{1/2}A)] =$$
$$= \frac{2^{\nu+r+1/2}\nu\Gamma(\frac{2\nu+p+2r}{2})}{f(\mathbf{y})\sqrt{\pi}^{p+1}|\mathbf{\Sigma}|^{1/2}}(d+A^2)^{-\frac{2\nu+p+2r}{2}}P_1(\frac{2\nu+p+2r}{2}, \frac{d+A^2}{2})$$

here $P_x(a, b)$ is the cumulative distribution function of the $Gamma(a, b)$ distribution evaluated at x. □

Corollary 3.16. *If* $\mathbf{Y} \sim SSL_p(\boldsymbol{\mu}, \mathbf{\Sigma}, \boldsymbol{\lambda}, \nu)$ *then we have*

$$E[\mathbf{Y}_2|\mathbf{y}_1] = \boldsymbol{\mu}_{2.1} + \frac{\mathbf{\Sigma}_{22.1}\boldsymbol{v}_2}{\sqrt{1 + \boldsymbol{v}_2^\top \mathbf{\Sigma}_{22.1}\boldsymbol{v}_2}} \times$$
$$\frac{2^\nu \nu}{f(\mathbf{y}_1)}\frac{\Gamma^{\frac{(p_1+2\nu-1)}{2}}(d_{y1}+(\widetilde{\boldsymbol{v}}^\top(\mathbf{y}_1-\boldsymbol{\mu}_1))^2)^{-\frac{p_1+2\nu-1}{2}}}{\sqrt{\pi}^{(p_1+1)}|\mathbf{\Sigma}_{11}|^{1/2}} \times$$
$$P_1(\frac{p_1+2\nu-1}{2}, \frac{d_{y_1}+(\widetilde{\boldsymbol{v}}^\top(\mathbf{y}_1-\boldsymbol{\mu}_1))^2}{2}),$$

where $d_{y_1} = (\mathbf{y}_1 - \boldsymbol{\mu}_1)^\top \mathbf{\Sigma}_{11}^{-1}(\mathbf{y}_1 - \boldsymbol{\mu}_1)$. □

• *The contaminated skew-normal distribution* $SCN_p(\boldsymbol{\mu}, \mathbf{\Sigma}, \boldsymbol{\lambda}, \nu, \gamma)$ *with* $0 \leq \nu \leq 1$, $0 < \gamma < 1$. *Taking* $h(u; \boldsymbol{\nu})$ *as in (2.8), we get in a straightforward manner*

$$f(\mathbf{y}) = 2\{\nu\phi_p(\mathbf{y}|\boldsymbol{\mu}, \frac{\mathbf{\Sigma}}{\gamma})\Phi_1(\gamma^{1/2}\boldsymbol{\lambda}^\top\mathbf{\Sigma}^{-1/2}(\mathbf{y}-\boldsymbol{\mu}))$$
$$+ (1-\nu)\phi_p(\mathbf{y}|\boldsymbol{\mu}, \mathbf{\Sigma})\Phi_1(\boldsymbol{\lambda}^\top\mathbf{\Sigma}^{-1/2}(\mathbf{y}-\boldsymbol{\mu}))\}. \qquad (3.10)$$

In this case the contaminated skew-normal distribution reduces to the skew-normal distribution when $\gamma = 1$. Hence, the mean vector and the covariance matrix are given, respectively, by

$$E[\mathbf{Y}] = \boldsymbol{\mu} + \sqrt{\frac{2}{\pi}}(\frac{\nu}{\gamma^{1/2}} + 1 - \nu)\mathbf{\Sigma}^{1/2}\boldsymbol{\delta}$$

and

$$Var[\mathbf{Y}] = (\frac{\nu}{\gamma} + 1 - \nu)\mathbf{\Sigma} - \frac{2}{\pi}(\frac{\nu}{\gamma^{1/2}} + 1 - \nu)^2\mathbf{\Sigma}^{1/2}\boldsymbol{\delta}\boldsymbol{\delta}^\top\mathbf{\Sigma}^{1/2}.$$

From (3.10) we derive the following results.

Proposition 3.17. *For* $\mathbf{Y} \sim SCN_p(\boldsymbol{\mu}, \mathbf{\Sigma}, \boldsymbol{\lambda}, \nu, \gamma)$ *we get*

$$E[U^r|\mathbf{y}] = \frac{2}{f(\mathbf{y})}[\nu\gamma^r\phi_p(\mathbf{y}|\boldsymbol{\mu}, \gamma^{-1}\mathbf{\Sigma})\Phi_1(\gamma^{1/2}A) + (1 - \nu)\phi_p(\mathbf{y}|\boldsymbol{\mu}, \mathbf{\Sigma})\Phi_1(A)]$$

and

$$E[U^r W_{\Phi_1}(U^{1/2}A)] = $$
$$\frac{2}{f(\mathbf{y})}[\nu\gamma^r\phi_p(\mathbf{y}|\boldsymbol{\mu}, \gamma^{-1}\mathbf{\Sigma})\phi_1(\gamma^{1/2}A) + (1 - \nu)\phi_p(\mathbf{y}|\boldsymbol{\mu}, \mathbf{\Sigma})\phi_1(A)].$$

$\square$

Corollary 3.18. *For* $\mathbf{Y} \sim SCN_p(\boldsymbol{\mu}, \mathbf{\Sigma}, \boldsymbol{\lambda}, \nu, \gamma)$ *we get*

$$
\begin{aligned}
E[\mathbf{Y}_2|\mathbf{y}_1] &= \boldsymbol{\mu}_{2.1} + \frac{2\mathbf{\Sigma}_{22.1}\boldsymbol{v}_2}{f(\mathbf{y}_1)\sqrt{1 + \boldsymbol{v}_2^\top\mathbf{\Sigma}_{22.1}\boldsymbol{v}_2}} \times \\
&\quad \left[\nu\gamma^{-1/2}\phi_{p_1}(\mathbf{y}_1|\boldsymbol{\mu}_1, \gamma^{-1}\mathbf{\Sigma}_{11})\phi_1(\gamma^{1/2}\widetilde{\boldsymbol{v}}^\top(\mathbf{y}_1 - \boldsymbol{\mu}_1))+ \right. \\
&\quad \left. (1 - \nu)\phi_{p_1}(\mathbf{y}_1|\boldsymbol{\mu}_1, \mathbf{\Sigma}_{11})\phi_1(\widetilde{\boldsymbol{v}}^\top(\mathbf{y}_1 - \boldsymbol{\mu}_1))\right],
\end{aligned}
$$

where $d_{y_1} = (\mathbf{y}_1 - \boldsymbol{\mu}_1)^\top\mathbf{\Sigma}_{11}^{-1}(\mathbf{y}_1 - \boldsymbol{\mu}_1).$ $\square$

Remark 3.19. The stochastic representation given by Equation (2.2) can be used to obtain the slash Student. Let U_1 (with probability density function as in (2.6)), $U_2 \sim Gamma(\nu/2, \nu/2)$ (with $\nu > 0$), and $\mathbf{X} \sim N_p(\mathbf{0}, \mathbf{\Sigma})$ be all independently distributed. Then

$$\mathbf{Y} \overset{d}{=} \boldsymbol{\mu} + U_1^{-1/2}U_2^{-1/2}\mathbf{X} \tag{3.11}$$

has a slash student distribution (Tang and Peng (2006) [29]). The proof follows from the formula

$$\mathbf{T} = U_2^{-1/2}\mathbf{X} \sim t_p(\boldsymbol{\mu}, \mathbf{\Sigma}, \nu).$$

Remark 3.20. If $\mathbf{X} \sim \mathrm{SN}_p(\mathbf{0}, \boldsymbol{\Sigma}, \boldsymbol{\lambda})$, then $\mathbf{Y}$ in (3.11) has a skew-slash student distribution as shown by Tang and Peng (2006) [29]. Obviously, many other distributions can be constructed by choosing appropriate probability density functions (i.e, $h(.; \boldsymbol{\nu})$) for U_1 and U_2.
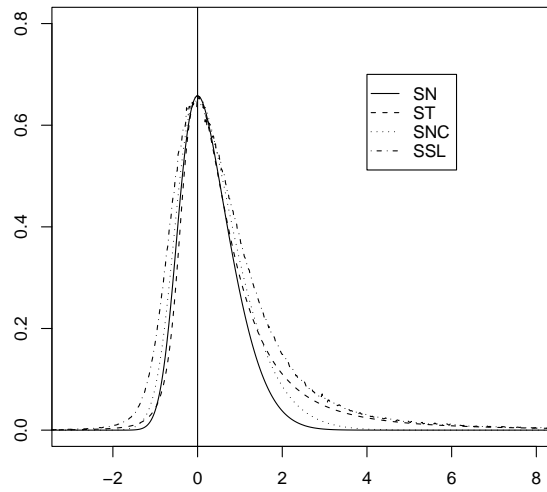

Figure 1: Density curves of the univariate skew-normal, skew-t, skew-slash and contaminated skew-normal distributions.

In Figure 1 we drew the density of the standard distribution $SN_1(3)$ together with the standard densities of the distributions $ST_1(3,2)$, SSL_1 $(3,1)$ and $SNC_1(3, 0.5, 0.5)$. They are rescaled to take the same value at the origin. The four densities are positively skewed. The skew-slash and skew-t distributions have much heavier tails than the skew-normal distribution. Actually, the skew-slash and the skew-t distributions do not have finite means nor variances. Figure 2 depicts some contours of the

densities associated with the standard bivariate skew-normal distribution $SN_2(\boldsymbol{\lambda})$, the standard bivariate skew-t distribution $ST_2(\boldsymbol{\lambda}, 2)$, the standard bivariate skew-slash distribution $SSL_2(\boldsymbol{\lambda}, 1)$, and the standard bivariate contaminated skew-normal distribution $SCN_2(\boldsymbol{\lambda}, 0.5, 0.5)$, with $\boldsymbol{\lambda} = (2, 1)^\top$ for all the distributions. Note that these contours are not elliptical and they can be strongly asymmetric depending on the choice of the parameters.
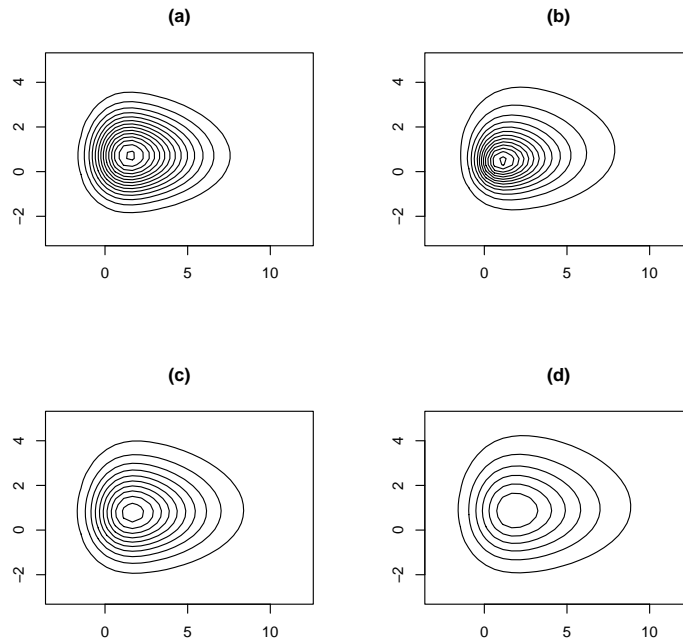

Figure 2: Contour plot of some elements of the standard bivariate SNI family. (a) $SN_2(\boldsymbol{\lambda})$, (b) $ST_2(\boldsymbol{\lambda}, 2)$, (c) $SCN_2(\boldsymbol{\lambda}, 0.5, 0.5)$, and (d) $SSL_2(\boldsymbol{\lambda}, 1)$, where $\boldsymbol{\lambda} = (2, 1)^\top$.
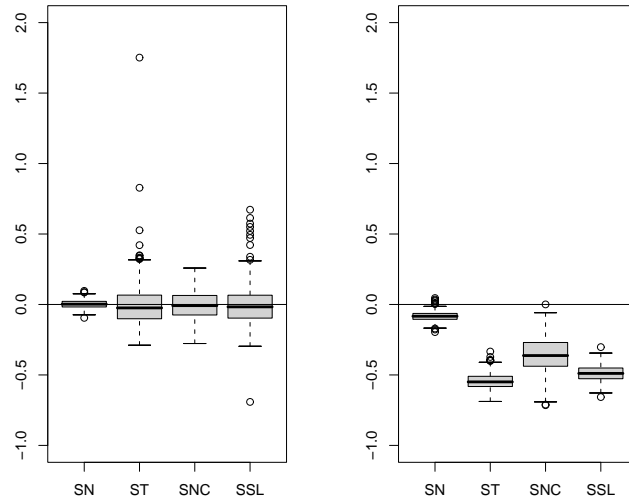
Figure 3: Box-plots of the sample mean (left panel) and sample median (right panel) on 500 samples of size n=100 from the four standardized distributions: $SN_1(3)$, $ST_1(3,2)$, $SSL_1(3,1)$, and $SNC_1(3,0.9,0.1)$. The respective means are adjusted to zero.

3.2   A Simulation study

To illustrate the usefulness of the SNI distribution, we perform a small simulation in order to study the behavior of two location estimators, the sample mean and the sample median under four different standard univariate settings. In particular, we consider a standard skew-normal $SN_1(3)$, a skew-t $ST_1(3,2)$, a skew-slash $SSL_1(3,1)$, and a contaminated

skew-normal $SCN_1(3, 0.9, 0.1)$. The 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. Note that the skew-slash and skew-t will have infinite variance when $\nu = 1$, $\nu = 2$, respectively. We simulate 500 samples of size $n = 100$ for them. For each sample, we compute the sample mean and the sample median and report the box-plot for each distribution in Figure 3. In the left panel all box-plots 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 the box-plots of the estimated medians have a slightly larger variability for the skew-normal and skew-contaminated normal, and have a much smaller variability for skew-t and skew-slash 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 heavy-tailed arise in the underlying distribution.

4. Maximum likelihood estimation

This section presents an EM-algorithm to perform maximum likelihood estimation for two multivariate SNI models of considerable practical interest.

4.1 Multivariate SNI responses

Suppose that we have observations on n independent individuals, $\wedge_1, \ldots,$ $\wedge_n$, where $\wedge_i \sim SNI_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}; H)$, $i = 1, \ldots, n$. . The parameter vector is $\boldsymbol{\theta} = (\boldsymbol{\mu}^\top, \boldsymbol{\gamma}^\top, \boldsymbol{\lambda}^\top)^\top$, where $\boldsymbol{\gamma}$ denotes a minimal set of parameters such that $\boldsymbol{\Sigma}(\boldsymbol{\gamma})$ is well defined (e.g., the upper triangular elements of $\boldsymbol{\Sigma}$ in the unstructured case).

In what follows, we illustrate the implementation of likelihood inference for the multivariate SNI via the EM-algorithm. The EM-algorithm

is a popular iterative algorithm for maximum likelihood estimation for models with incomplete data. More specifically, let $\mathbf{y}$ denote the observed data and $\mathbf{s}$ the missing one. The complete data $\mathbf{y}_c = (\mathbf{y}, \mathbf{s})$ is $\mathbf{y}$ augmented with $\mathbf{s}$. We denote by $\ell_c(\boldsymbol{\theta}|\mathbf{y}_c)$ (with $\boldsymbol{\theta} \in \boldsymbol{\Theta}$) 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 a M-step, defined as follows.

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

By using (3.3), the setup defined above can be written as

$$\mathbf{Y}_i|T_i = t_i, U_i = u_i, \quad \overset{\text{ind}}{\sim} \ N_p(\boldsymbol{\mu} + t_i\boldsymbol{\Sigma}^{1/2}\boldsymbol{\delta}, u_i^{-1}\boldsymbol{\Sigma}^{1/2}(\mathbf{I}_p - \boldsymbol{\delta}\boldsymbol{\delta}^{\top})\boldsymbol{\Sigma}^{1/2}) \quad (4.1)$$

$$T_i|U_i = u_i \quad \overset{\text{iid}}{\sim} \ HN_1(0, \frac{1}{u_i}) \quad (4.2)$$

$$U_i \quad \overset{\text{ind}}{\sim} \ h(u_i; \boldsymbol{\nu}), \quad (4.3)$$

all independent, where $HN_1(0, 1)$ denotes the univariate standard half-normal distribution (see $|X_0| = |T_0|$ in Equation (1.2) or Johnson et al. (1994) [18]). We assume that the parameter vector $\boldsymbol{\nu}$ is known. In practice, the optimum value of $\boldsymbol{\nu}$ can be determined using the profile likelihood and the Schwarz information criterion (see Lange et al. (1989) [23]).

Let $\mathbf{y} = (\mathbf{y}_1^{\top}, \ldots, \mathbf{y}_n^{\top})^{\top}$, $\mathbf{u} = (u_1, \ldots, u_n)^{\top}$, and $\mathbf{t} = (t_1, \ldots, t_n)^{\top}$. Then, under the hierarchical representation (4.1)–(4.2), with $\boldsymbol{\Delta} = \boldsymbol{\Sigma}^{1/2}\boldsymbol{\delta}$ and $\boldsymbol{\Gamma} = \boldsymbol{\Sigma} - \boldsymbol{\Delta}\boldsymbol{\Delta}^{\top}$, 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

$$\ell_c(\boldsymbol{\theta}|\mathbf{y}_c) = c - \frac{n}{2}\log|\boldsymbol{\Gamma}| - \frac{1}{2}\sum_{i=1}^{n} u_i(\mathbf{y}_i - \boldsymbol{\mu} - \boldsymbol{\Delta}t_i)^{\top}\boldsymbol{\Gamma}^{-1}(\mathbf{y}_i - \boldsymbol{\mu} - \boldsymbol{\Delta}t_i),$$

where c is a constant independent of the parameter vector $\boldsymbol{\theta}$. By letting $\widehat{u}_i = \mathrm{E}[U_i|\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}, \mathbf{y}_i]$, $\widehat{ut_i} = \mathrm{E}[U_iT_i|\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}, \mathbf{y}_i]$, $\widehat{ut_i^2} = \mathrm{E}[U_iT_i^2|\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}, \mathbf{y}_i]$, and using known properties of conditional expectation we obtain

$$\widehat{ut_i} = \widehat{u}_i\widehat{\mu}_{Ti} + \widehat{M}_{T_i}\widehat{\tau}_i \quad (4.4)$$

$$\widehat{ut_i^2} = \widehat{u}_i\widehat{\mu}_{T_i}^2 + \widehat{M}_{T_i}^2 + \widehat{M}_{T_i}\widehat{\mu}_{T_i}\widehat{\tau}_i, \quad (4.5)$$

with $\widehat{\tau}_i = E[U_i^{1/2} W_{\Phi_1}(\frac{U_i^{1/2}\widehat{\mu}_{T_i}}{\widehat{M}_{T_i}})|\widehat{\boldsymbol{\theta}}, \mathbf{y}_i]$, $W_{\Phi_1}(x) = \phi_1(x)/\Phi_1(x)$, $\widehat{M}_T^2 = 1/(1 + \widehat{\boldsymbol{\Delta}}^\top \widehat{\boldsymbol{\Gamma}}^{-1} \widehat{\boldsymbol{\Delta}})$ and $\widehat{\mu}_{T_i} = \widehat{M}_{T_i}^2 \widehat{\boldsymbol{\Delta}}^\top \widehat{\boldsymbol{\Gamma}}^{-1}(\mathbf{y}_i - \boldsymbol{\mu})$, $i = 1, \ldots, n$.

As we have $\frac{\mu_{T_i}}{M_{T_i}} = \boldsymbol{\lambda}^\top \boldsymbol{\Sigma}^{-1/2}(\mathbf{y}_i - \boldsymbol{\mu})$, the conditional expectations given in (4.4)–(4.5), specifically $\widehat{u}_i$ and $\widehat{\tau}_i$, can be easily derived from the results of Section 3.1. Thus, at least for the skew-t and skew-contaminated normal distributions of the SNI class we have closed-form expressions for the quantities $\widehat{u}_i$ and $\widehat{\tau}_i$. For the skew-slash case, Monte Carlo integration may be employed, which yield the so-called MC-EM algorithm.

It follows, after some simple algebra involving (4.4)–(4.5), that the conditional expectation of the complete log-likelihood function has the form

$$
\begin{aligned}
Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}) &= \mathrm{E}[\ell_c(\boldsymbol{\theta}|\mathbf{y}_c)|\mathbf{y}, \widehat{\boldsymbol{\theta}}] \\
&= c - \frac{n}{2}\log|\boldsymbol{\Gamma}| - \frac{1}{2}\sum_{i=1}^n \widehat{u}_i(\mathbf{y}_i - \boldsymbol{\mu})^\top \boldsymbol{\Gamma}^{-1}(\mathbf{y}_i - \boldsymbol{\mu}) \\
&\quad + \sum_{i=1}^n \widehat{ut}_i(\mathbf{y}_i - \boldsymbol{\mu})^\top \boldsymbol{\Gamma}^{-1}\boldsymbol{\Delta} - \frac{1}{2}\sum_{i=1}^n \widehat{ut^2}_i \boldsymbol{\Delta}^\top \boldsymbol{\Gamma}^{-1}\boldsymbol{\Delta}.
\end{aligned}
$$

We then have the following EM-type algorithm.

E-step: Given $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}$, compute $\widehat{ut_i^2}$, $\widehat{ut_i}$ and $\widehat{u_i}$ using (4.4)–(4.5).
M-step: Update $\widehat{\boldsymbol{\theta}}$ by maximizing $Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}})$ over $\boldsymbol{\theta}$, which leads us to the following closed-form expressions

$$
\widehat{\boldsymbol{\mu}} = \sum_{i=1}^n (\widehat{u}_i \mathbf{y}_i - \widehat{ut}_i \boldsymbol{\Delta})/(\sum_{i=1}^n \widehat{u}_i), \tag{4.6}
$$

$$
\widehat{\boldsymbol{\Gamma}} = \frac{1}{n}\sum_{i=1}^n \left[\widehat{u}_i(\mathbf{y}_i - \boldsymbol{\mu})(\mathbf{y}_i - \boldsymbol{\mu})^\top - 2\widehat{ut}_i \boldsymbol{\Delta}(\mathbf{y}_i - \boldsymbol{\mu})^\top + \widehat{ut_i^2}\boldsymbol{\Delta}\boldsymbol{\Delta}^\top \right],
$$

$$
\widehat{\boldsymbol{\Delta}} = \frac{\sum_{i=1}^n \widehat{ut}_i(\mathbf{y}_i - \boldsymbol{\mu})}{\sum_{i=1}^n \widehat{ut_i^2}}.
$$

The skewness parameter vector and the unstructured scale matrix

can be estimated by the equalities $\widehat{\boldsymbol{\Sigma}} = \widehat{\boldsymbol{\Gamma}} + \widehat{\boldsymbol{\Delta}}\widehat{\boldsymbol{\Delta}}^{T}$ and $\widehat{\boldsymbol{\lambda}} = \widehat{\boldsymbol{\Sigma}}^{-1/2}\widehat{\boldsymbol{\Delta}}/(1 - \widehat{\boldsymbol{\Delta}}^{\top}\widehat{\boldsymbol{\Sigma}}^{-1}\widehat{\boldsymbol{\Delta}})^{1/2}$. It is clear that when $\boldsymbol{\lambda} = \mathbf{0}$ (or when $\boldsymbol{\Delta} = \mathbf{0}$), the M-step equations reduce to the equations obtained assuming normal/independent distribution. This algorithm clearly generalizes results found in Lachos et al. (2007) [20, Section 2] by taking $U_i = 1$, $i = 1, \ldots, n$. Useful starting values required to implement this algorithm are those obtained under the normality assumption, with the starting values for the skewness parameter vector set equal to $\mathbf{0}$. However, in order to ensure that the true ML estimate is identified, we recommend running the EM algorithm using a range of different starting values. The log-likelihood function for $\boldsymbol{\theta} = (\boldsymbol{\mu}^{\top}, \boldsymbol{\gamma}^{\top}, \boldsymbol{\lambda}^{\top})^{\top}$, given the observed sample $\mathbf{y} = (\mathbf{y}_1^{\top}, \ldots, \mathbf{y}_n^{\top})^{\top}$, is of the form

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^{n} \ell_i(\boldsymbol{\theta}), \tag{4.7}$$

where $\ell_i(\boldsymbol{\theta}) = \log 2 - \dfrac{p}{2} log 2\pi - \dfrac{1}{2}\log|\boldsymbol{\Sigma}| + \log K_i$, with

$$K_i = K_i(\boldsymbol{\theta}) = \int_0^{\infty} u_i^{p/2}\exp\{-\frac{1}{2}u_i d_i\}\Phi_1(u_i^{1/2}A_i)dH(u_i),$$

where $d_i = (\mathbf{y}_i - \boldsymbol{\mu})^{\top}\boldsymbol{\Sigma}^{-1}(\mathbf{y}_i - \boldsymbol{\mu})$ and $A_i = \boldsymbol{\lambda}^{\top}\boldsymbol{\Sigma}^{-1}(\mathbf{y}_i - \boldsymbol{\mu})$. Explicit expressions for the observed information matrix can be derived from the results presented in Section 5.

4.2 Multivariate measurement error model

In this section we further apply the SNI distribution to a multivariate measurement error model. Let n be the sample size, X_i the observed value in unit i of the covariate, y_{ij} the j-th observed response in unit i, and x_i the unobserved (true) covariate value for unit i; here i ranges from 1 to n, and j from 1 to r. Relating these variables we postulate as working model (see also Barnett (1969) [8] and Shyr and Gleser (1986) [28]) the equations

$$X_i = x_i + u_i, \tag{4.8}$$

and

$$\mathbf{Z}_i = \boldsymbol{\alpha} + \boldsymbol{\beta} x_i + \mathbf{e}_i, \tag{4.9}$$

where $\mathbf{Z}_i = (z_{i1}, \ldots, z_{ir})^\top$ is the vector of responses for the i-th experimental unit, $\mathbf{e}_i = (e_{i1}, \ldots, e_{ir})^\top$ is a random vector of measurement errors of dimension r, and $\boldsymbol{\alpha} = (\alpha_1, ..., \alpha_r)^\top, \boldsymbol{\beta} = (\beta_1, ..., \beta_r)^\top$ are parameter vectors of dimension r.

Set $\boldsymbol{\epsilon}_i = (u_i, \mathbf{e}_i^\top)^\top$ and $\mathbf{Y}_i = (X_i, \mathbf{Z}_i^\top)^\top$. Then the model defined by Equations (4.8)–(4.9) can be rewritten as

$$\mathbf{Y}_i = \mathbf{a} + \mathbf{b} x_i + \boldsymbol{\epsilon}_i, \tag{4.10}$$

where $\mathbf{a} = (0, \boldsymbol{\alpha}^\top)^\top$ and $\mathbf{b} = (1, \boldsymbol{\beta}^\top)^\top$ are $p \times 1$ vectors, with $p = r + 1$. We assume

$$\begin{pmatrix} x_i \\ \boldsymbol{\epsilon}_i \end{pmatrix} \stackrel{\text{iid}}{\sim} SNI_{p+1} \left(\begin{pmatrix} \mu_x \\ \mathbf{0} \end{pmatrix}, D(\phi_x, \boldsymbol{\phi}), \begin{pmatrix} \lambda_x \\ \mathbf{0} \end{pmatrix}; H \right), \tag{4.11}$$

where $D(\phi_x, \boldsymbol{\phi}) = diag(\phi_x, \phi_1, \ldots, \phi_p)^\top$, with $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_p)$, called **structural SNI-MMEM**. From (2.2), this formulation implies

$$\begin{pmatrix} x_i \\ \boldsymbol{\epsilon}_i \end{pmatrix} | U_i = u_i \sim SN_{p+1} \left(\begin{pmatrix} \mu_x \\ \mathbf{0} \end{pmatrix}, u_i^{-1} D(\phi_x, \boldsymbol{\phi}), \begin{pmatrix} \lambda_x \\ \mathbf{0} \end{pmatrix} \right), \tag{4.12}$$

$$U_i \sim h(u_i; \boldsymbol{\nu}). \tag{4.13}$$

From Corollary 3.5, marginally we get

$$\boldsymbol{\epsilon}_i \stackrel{\text{ind}}{\sim} NI_{m+1}(\mathbf{0}, D(\boldsymbol{\phi}); H) \quad \text{and} \quad x_i \stackrel{\text{ind}}{\sim} SNI_1(\mu_x, \phi_x, \lambda_x; H). \tag{4.14}$$

The asymmetric parameter λ_x incorporates asymmetry in the latent variable x_i and consequently in the observed quantities $\mathbf{Y}_i$, which will be shown to have marginally multivariate SNI distributions. If $\lambda_x = 0$, then the asymmetric model reduces to the symmetric MMEM considering NI distributions. Under (4.11), it follows from (1.2) that the regression

setup defined in (4.8)–(4.11) can be written hierarchically as

$$\mathbf{Y}_i \mid x_i, U_i = u_i \overset{\text{ind}}{\sim} N_p(\mathbf{a} + \mathbf{b}x_i, u_i^{-1}D(\boldsymbol{\phi})), \tag{4.15}$$

$$x_i \mid T_i = t_i, U_i = u_i \overset{\text{ind}}{\sim} N_1(\mu_x + \phi_x^{1/2}\delta_x t_i, u_i^{-1}\phi_x(1 - \delta_x^2)), \tag{4.16}$$

$$T_i \overset{\text{iid}}{\sim} HN_1(0, \frac{1}{u_i}), \tag{4.17}$$

$$U_i \overset{\text{iid}}{\sim} h(u_i; \boldsymbol{\nu}), \tag{4.18}$$

all independent, where $\delta_x = \lambda_x/(1 + \lambda_x^2)^{1/2}$. As in Lange et al. (1989) [23], we assumed $\boldsymbol{\nu}$ to be known. Classical inference on the parameter vector $\boldsymbol{\theta} = (\boldsymbol{\alpha}^\top, \boldsymbol{\beta}^\top, \boldsymbol{\phi}^\top, \mu_x, \phi_x, \lambda_x)^\top$ in this type of model is based on the marginal distribution for $\mathbf{Y}_i$ given in the following proposition (see Bolfarine and Galea-Rojas (1995) [10]).

Proposition 4.1. *For the structural SNI-MMEM model (4.8)–(4.11), the marginal distribution of $\mathbf{Y}_i$ is given by*

$$f_{\mathbf{Y}_i}(\mathbf{y}_i|\boldsymbol{\theta}) = 2 \int_0^\infty \phi_p(\mathbf{y}_i|\boldsymbol{\mu}, u_i^{-1}\boldsymbol{\Sigma})\Phi_1(u_i^{1/2}\bar{\boldsymbol{\lambda}}_x^\top\boldsymbol{\Sigma}^{-1/2}(\mathbf{y}_i - \boldsymbol{\mu}))dH(u_i) \tag{4.19}$$

(i.e., by $\mathbf{Y}_i \overset{\text{iid}}{\sim} SNI_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \bar{\boldsymbol{\lambda}}_x; H)$, with $\boldsymbol{\mu} = \mathbf{a} + \mathbf{b}\mu_x$, $\boldsymbol{\Sigma} = \phi_x\mathbf{b}\mathbf{b}^\top + D(\boldsymbol{\phi})$, and $\bar{\boldsymbol{\lambda}}_x = \frac{\lambda_x\phi_x\boldsymbol{\Sigma}^{-1/2}\mathbf{b}}{\sqrt{\phi_x+\lambda_x^2\Lambda_x}}$; here $\Lambda_x = (\phi_x^{-1} + \mathbf{b}^\top D^{-1}(\boldsymbol{\phi})\mathbf{b})^{-1})$.

Proof. The proof is a direct consequence of Proposition 3.8 after some algebraic manipulations. □

It follows that the log-likelihood function for $\boldsymbol{\theta}$, given the observed sample $\mathbf{y} = (\mathbf{y}_1^\top, \ldots, \mathbf{y}_n^\top)^\top$, is

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^n \ell_i(\boldsymbol{\theta}), \tag{4.20}$$

where $\ell_i(\boldsymbol{\theta}) = \log 2 - \frac{p}{2}log2\pi - \frac{1}{2}\log|\boldsymbol{\Sigma}| + \log K_i$, with

$$K_i = K_i(\boldsymbol{\theta}) = \int_0^\infty u_i^{p/2} \exp\{-\frac{1}{2}u_i d_i\}\Phi_1(u_i^{1/2}A_i) \ dH(u_i),$$

and $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$, $\bar{\boldsymbol{\lambda}}_x$ as in Proposition 4.1. Here $d_i = (\mathbf{y}_i - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{y}_i - \boldsymbol{\mu})$ and $A_i = \bar{\boldsymbol{\lambda}}_x^\top \boldsymbol{\Sigma}^{-1/2}(\mathbf{y}_i - \boldsymbol{\mu}) = A_x a_i$ hold with $A_x = \frac{\lambda_x \Lambda_x}{\sqrt{\phi_x + \lambda_x^2 \Lambda_x}}$ and $a_i = (\mathbf{y}_i - \boldsymbol{\mu})^\top D^{-1}(\boldsymbol{\phi})\mathbf{b}$.

The ML estimators of the parameters in the model (4.10)–(4.11) can be found by direct maximization of the log-likelihood (4.20) which can be computed numerically using the *optim* routine in platform R or *fmincon* in Matlab. An oft-voiced complaint about these methods is that they may not converge unless good starting values are provided. The EM algorithm —which takes advantage of being insensitive to the stating values— is a tool that requires the construction of unobserved data, and has been well developed and has become a broadly applicable approach to the iterative computation of ML estimates. Thus, if we let $\mathbf{y} = (\mathbf{y}_1^\top, \ldots, \mathbf{y}_n^\top)^\top$, $\mathbf{x} = (x_1, \ldots, x_n)^\top$, $\mathbf{u} = (u_1, \ldots, u_n)^\top$, $\mathbf{t} = (t_1, \ldots, t_n)^\top$, $\nu_x^2 = \phi_x(1 - \delta_x^2)$ and $\tau_x = \phi_x^{1/2}\delta_x$, it follows that the complete log-likelihood function associated with $(\mathbf{y}, \mathbf{x}, \mathbf{t}, \mathbf{u})$ is given by

$$\ell_c(\boldsymbol{\theta}|\mathbf{y}, \mathbf{x}, \mathbf{t}, \mathbf{u}) \quad \propto \quad -\frac{n}{2}\log(|D(\boldsymbol{\phi})|) \tag{4.21}$$

$$-\frac{1}{2}\sum_{i=1}^n u_i(\mathbf{y}_i - \mathbf{a} - \mathbf{b}x_i)^\top D^{-1}(\boldsymbol{\phi})(\mathbf{y}_i - \mathbf{a} - \mathbf{b}x_i) \tag{4.22}$$

$$-\frac{n}{2}\log(\nu_x^2) - \frac{1}{2\nu_x^2}\sum_{i=1}^n u_i(x_i - \mu_x - \tau_x t_i)^2. \tag{4.23}$$

Letting $\widehat{u}_i = E[U_i|\widehat{\boldsymbol{\theta}}, \mathbf{y}_i]$, $\widehat{ut}_i = E[U_i t_i|\widehat{\boldsymbol{\theta}}, \mathbf{y}_i]$, $\widehat{t_i^2} = E[t_i^2|\widehat{\boldsymbol{\theta}}, \mathbf{y}_i]$, $\widehat{ux}_i = E[U_i x_i|\widehat{\boldsymbol{\theta}}, \mathbf{y}_i]$, $\widehat{ux_i^2} = E[U_i x_i^2|\widehat{\boldsymbol{\theta}}, \mathbf{y}_i]$, and $\widehat{utx}_i = E[U_i t_i x_i|\widehat{\boldsymbol{\theta}}, \mathbf{y}_i]$ we obtain

$$\widehat{ut}_i \quad = \quad \widehat{u}_i\widehat{\mu}_{T_i} + \widehat{M}_T E[U_i^{1/2} W_{\Phi_1}(\frac{U_i^{1/2}\widehat{\mu}_{T_i}}{\widehat{M}_T})|\widehat{\boldsymbol{\theta}}, \mathbf{y}_i], \tag{4.24}$$

$$\widehat{t_i^2} \quad = \quad \widehat{u}_i\widehat{\mu}_{T_i}^2 + \widehat{M}_T^2 + \widehat{M}_T\widehat{\mu}_{T_i} E[U_i^{1/2} W_{\Phi_1}(\frac{U_i^{1/2}\widehat{\mu}_{T_i}}{\widehat{M}_T})|\widehat{\boldsymbol{\theta}}, \mathbf{y}_i],$$

$$\widehat{ux}_i \quad = \quad \widehat{r}_i\,\widehat{u}_i + \widehat{s}\,\widehat{ut}_i, \quad \widehat{ux_i^2} = \widehat{T}_x^2 + \widehat{r}_i^2\,\widehat{u}_i + 2\widehat{r}_i\widehat{s}\,\widehat{ut}_i + \widehat{s}^2\,\widehat{t_i^2},$$

$$\widehat{utx}_i \quad = \quad \widehat{r}_i\,\widehat{ut}_i + \widehat{s}\,\widehat{t_i^2},$$

where $\widehat{M}_T^2 = [1 + \widehat{\tau}_x^2\widehat{\mathbf{b}}^\top(D(\widehat{\boldsymbol{\phi}}) + \widehat{\nu}_x^2\widehat{\mathbf{b}}\widehat{\mathbf{b}}^\top)^{-1}\widehat{\mathbf{b}}]^{-1}$, $\widehat{\mu}_{T_i} = \widehat{\tau}_x\widehat{M}_T^2\widehat{\mathbf{b}}^\top(D(\widehat{\boldsymbol{\phi}}) + \widehat{\nu}_x^2\widehat{\mathbf{b}}\widehat{\mathbf{b}}^\top)^{-1}(\mathbf{y}_i - \widehat{\mathbf{a}} - \widehat{\mathbf{b}}\widehat{\mu}_x)$, $\widehat{T}_x^2 = \widehat{\nu}_x^2[1 + \widehat{\nu}_x^2\widehat{\mathbf{b}}^\top D^{-1}(\widehat{\boldsymbol{\phi}})\widehat{\mathbf{b}}]^{-1}$,

$\widehat{r}_i = \widehat{\mu}_x + \widehat{T_x}^2 \mathbf{\widehat{b}}^\top D^{-1}(\widehat{\boldsymbol{\phi}})(\mathbf{y}_i - \widehat{\mathbf{a}} - \widehat{\mathbf{b}}\widehat{\mu}_x)$, and $\widehat{s} = \widehat{\tau}_x(1 - \widehat{T_x}^2 \mathbf{\widehat{b}}^\top D^{-1}(\widehat{\boldsymbol{\phi}})\mathbf{\widehat{b}})$.
A closed-form expression for $E[U_i^{1/2} W_{\Phi_1}(\frac{U_i^{1/2}\widehat{\mu}_{T_i}}{\widehat{M}_T})|\widehat{\boldsymbol{\theta}}, \mathbf{y}_i]$ can be found from the results given in Section 3.1.

In this way we have the following EM type algorithm.

E-step: Given $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}$, compute $\widehat{u}_i$, $\widehat{t_i^2}$, $\widehat{ut}_i$, $\widehat{ux}_i$, $\widehat{ux_i^2}$, and $\widehat{utx}_i$ using (4.24).

M-step: Update $\widehat{\boldsymbol{\theta}}$ by maximizing $E[\ell_c(\boldsymbol{\theta}|\mathbf{y}, \mathbf{x}, \mathbf{t}, \mathbf{u})|\mathbf{y}, \widehat{\boldsymbol{\theta}}]$ over $\boldsymbol{\theta}$; which leads to

$$
\begin{aligned}
\widehat{\boldsymbol{\alpha}} &= \overline{\mathbf{z}}_u - \overline{x}_u \widehat{\boldsymbol{\beta}}, \\
\widehat{\boldsymbol{\beta}} &= \frac{\sum_{i=1}^n \widehat{ux}_i(\mathbf{z}_i - \overline{\mathbf{z}}_u)}{\sum_{i=1}^n \widehat{ux_i^2} - n\overline{\widehat{u}}\overline{x}_u^2}, \\
\widehat{\phi}_1 &= \frac{1}{n}\sum_{i=1}^n (\widehat{u}_i X_i^2 - 2\widehat{ux}_i X_i + \widehat{ux_i^2}), \\
\widehat{\phi}_{j+1} &= \frac{1}{n}\sum_{i=1}^n (\widehat{u}_i z_{ij}^2 + \widehat{u}_i \alpha_j^2 + \beta_j^2 \widehat{ux_i^2} - 2\widehat{u}_i \alpha_j z_{ij} - \\
& \qquad 2y_{ij}\beta_j \widehat{ux}_i + 2\alpha_j \beta_j \widehat{ux}_i), \qquad j = 1, \ldots, r, \\
\widehat{\mu}_x &= \overline{x}_u - \widehat{\tau}_x \overline{t}_u, \quad \widehat{\nu}_x^2 = \frac{1}{n}\sum_{i=1}^n (\widehat{ux_i^2} - \widehat{\mu}_x \widehat{ux}_i) - \widehat{\tau}_x \frac{1}{n}\sum_{i=1}^n \widehat{utx}_i, \\
\widehat{\tau}_x &= \frac{\sum_{i=1}^n (\widehat{utx}_i - \overline{x}_u \widehat{ut}_i)}{\sum_{i=1}^n (\widehat{t_i^2} - \overline{t}_u \widehat{ut}_i)},
\end{aligned}
$$

where $\overline{\mathbf{z}}_u = \frac{\sum_{i=1}^n \widehat{u}_i \mathbf{z}_i}{\sum_{i=1}^n \widehat{u}_i}$, $\overline{x}_u = \frac{\sum_{i=1}^n \widehat{ux}_i}{\sum_{i=1}^n \widehat{u}_i}$, $\overline{t}_u = \frac{\sum_{i=1}^n \widehat{ut}_i}{\sum_{i=1}^n \widehat{u}_i}$, and $\overline{\widehat{u}} = \frac{1}{n}\sum_{i=1}^n \widehat{u}_i$.

When $U_i = 1$, the M-step equations reduce to the equations obtained by Lachos et al. (2005) [21] under the skew-normal distribution. When $\lambda_x = 0$ (or when $\tau_x = 0$), the M-step equations become the equations by Bolfarine and Galea-Rojas (1995) [10]. Moreover, when $U \sim Gamma(\nu/2, \nu/2)$ and $\lambda_x = 0$, the M-step reduces to equations obtained by Bolfarine and Galea-Rojas (1996) [11]. The shape and scale parameters of the latent variable x can be estimated by noting the equalities $\tau_x/\nu_x = \lambda_x$ and $\phi_x = \tau_x^2 + \nu_x^2$.

We now consider an empirical Bayes inference for the latent variable that is useful for estimating the x_i quantities. Models (4.10) and (4.14) imply $\mathbf{Y}_i|x_i \sim NI_p(\mathbf{a} + \mathbf{b}x_i, D(\boldsymbol{\phi}); H)$ and $x_i \sim SNI_1(\mu_x, \sigma_x^2, \lambda_x; H)$. The conditional density of x_i, given $\mathbf{y}_i, u_i$, is

$$f(x_i|\mathbf{y}_i, u_i) = \phi_q(x_i|\mu_x + \Lambda_x a_i, u_i^{-1}\Lambda_x)\frac{\Phi_1(u_i^{1/2}\frac{\lambda_x(x_i-\mu_x)}{\sigma_x})}{\Phi_1(u_i^{1/2}A_i)},$$

where Λ_x and a_i, A_i are as in Proposition 4.1 and Equation (4.20), respectively. It follows from Lemma 8.1 in the appendix that we have

$$E[x_i|\mathbf{y}_i, u_i] = \mu_x + \Lambda_x a_i + u_i^{-1/2}\frac{\Lambda_x\lambda_x}{\sqrt{1+\lambda_x^2\Lambda_x}}W_{\Phi_1}(u_i^{1/2}A_i),$$

and as $E[x|\mathbf{y}] = E_U[E[x|\mathbf{y}, U]|\mathbf{y}]$ holds, we conclude that the minimum mean-square error (MSE) estimator of x_i obtained by the conditional mean of x_i, given $\mathbf{y}_i$, is

$$\widehat{x}_i = E[x_i|\mathbf{y}_i] = \mu_x + \Lambda_x a_i + \frac{\Lambda_x\lambda_x}{\sqrt{1+\lambda_x^2\Lambda_x}}E[U_i^{-1/2}\frac{\phi_1(U_i^{1/2}A_i)}{\Phi_1(U_i^{1/2}A_i)}|\mathbf{y}_i], \quad (4.25)$$

If $\mathbf{Y}_i$ has distribution $ST_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \bar{\boldsymbol{\lambda}}_x, \nu)$ or $SCN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \bar{\boldsymbol{\lambda}}_x, \nu, \gamma)$, then we obtain closed-form expressions for the expected values given in (4.25) from the results exhibited in Section 3.1. In practice the Bayes estimator of x_i, namely $\widehat{x}_i$, can be obtained by substituting the ML estimate $\boldsymbol{\theta}$ into (4.25).

5. The observed information matrix

In this section we develop the observed information matrix in a general form. Suppose that we have observations on n independent individuals $\wedge_1, \ldots, \wedge_n$, where $\wedge_i \sim SNI_{n_i}(\boldsymbol{\mu}_i(\boldsymbol{\beta}), \boldsymbol{\Sigma}_i(\boldsymbol{\gamma}), \boldsymbol{\lambda}_i(\boldsymbol{\lambda}); H)$. Then the log-likelihood function for $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\gamma}^\top, \boldsymbol{\lambda}^\top)^\top \in \mathbb{R}^q$, given the observed sample $\mathbf{y} = (\mathbf{y}_1^\top, \ldots, \mathbf{y}_n^\top)^\top$, is of the form

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^{n} \ell_i(\boldsymbol{\theta}), \quad (5.1)$$

where $\ell_i(\boldsymbol{\theta}) = \log 2 - \dfrac{n_i}{2} log 2\pi - \dfrac{1}{2} \log |\boldsymbol{\Sigma}_i| + \log K_i$, with

$$K_i = K_i(\boldsymbol{\theta}) = \int_0^\infty u_i^{n_i/2} \exp\{-\frac{1}{2}u_i d_i\}\Phi_1(u_i^{1/2}A_i)dH(u_i),$$

and $d_i = (\mathbf{y}_i - \boldsymbol{\mu}_i)^\top \boldsymbol{\Sigma}_i^{-1}(\mathbf{y}_i - \boldsymbol{\mu}_i)$, $A_i = \boldsymbol{\lambda}_i^\top \boldsymbol{\Sigma}_i^{-1}(\mathbf{y}_i - \boldsymbol{\mu}_i)$. Using the notation

$$
\begin{aligned}
I_i^\Phi(w) &= \int_0^\infty u_i^w \exp\{-\frac{1}{2}u_i d_i\}\Phi_1(u_i^{1/2}A_i)dH(u_i), \\
I_i^\phi(w) &= \int_0^\infty u_i^w \exp\{-\frac{1}{2}u_i d_i\}\phi_1(u_i^{1/2}A_i|0,1)dH(u_i),
\end{aligned}
$$

(so that $K_i(\boldsymbol{\theta})$ can be expressed as $K_i(\boldsymbol{\theta}) = I_i^\Phi(\frac{n_i}{2})$), it follows that the matrix of second derivatives with respect to $\boldsymbol{\theta}$ is just

$$
\begin{aligned}
\mathbf{L} &= \sum_{i=1}^n \frac{\partial^2 \ell_i(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top} && (5.2) \\
&= -\frac{1}{2}\sum_{i=1}^n \frac{\partial^2 \log|\boldsymbol{\Sigma}_i|}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top} - \sum_{i=1}^n \frac{1}{K_i^2}\frac{\partial K_i}{\partial\boldsymbol{\theta}}\frac{\partial K_i}{\partial\boldsymbol{\theta}^\top} + \sum_{i=1}^n \frac{1}{K_i}\frac{\partial^2 K_i}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}, && (5.3)
\end{aligned}
$$

where here we have $\frac{\partial K_i}{\partial\boldsymbol{\theta}} = I_i^\phi(\frac{n_i+1}{2})\frac{\partial A_i}{\partial\boldsymbol{\theta}} - \frac{1}{2}I_i^\Phi(\frac{n_i+2}{2})\frac{\partial d_i}{\partial\boldsymbol{\theta}}$ and $\frac{\partial^2 K_i}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top} = \frac{1}{4}I_i^\Phi(\frac{n_i+4}{2})\frac{\partial d_i}{\partial\boldsymbol{\theta}}\frac{\partial d_i}{\partial\boldsymbol{\theta}^\top} - \frac{1}{2}I_i^\Phi(\frac{n_i+2}{2})\frac{\partial^2 d_i}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top} - \frac{1}{2}I_i^\phi(\frac{n_i+3}{2})(\frac{\partial A_i}{\partial\boldsymbol{\theta}}\frac{\partial d_i}{\partial\boldsymbol{\theta}^\top} + \frac{\partial d_i}{\partial\boldsymbol{\theta}}\frac{\partial A_i}{\partial\boldsymbol{\theta}^\top}) - I_i^\phi(\frac{n_i+3}{2})A_i\frac{\partial A_i}{\partial\boldsymbol{\theta}}\frac{\partial A_i}{\partial\boldsymbol{\theta}^\top} + I_i^\Phi(\frac{n_i+1}{2})\frac{\partial^2 A_i}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}$.

From Propositions 3.13, 3.15, and 3.17 we have that for each distribution considered in this work, the integrates can be written as follows.

- *Skew-t:*

$$
\begin{aligned}
I_i^\Phi(w) &= \frac{2^w \nu^{\nu/2}\Gamma(w+\nu/2)}{\Gamma(\nu/2)(\nu+d_i)^{\nu/2+w}}T_1\left(\frac{A_i}{(d_i+\nu)^{1/2}}\sqrt{2w+\nu}|0,1,2w+\nu\right), \\
I_i^\phi(w) &= \frac{2^w \nu^{\nu/2}}{\sqrt{2\pi}\Gamma(\nu/2)}\left(\frac{1}{d_i+A_i^2+\nu}\right)^{\frac{\nu+2w}{2}}\Gamma(\frac{\nu+2w}{2}).
\end{aligned}
$$

- *Skew-slash:*

$$I_i^{\Phi}(w) \;\; = \;\; \frac{2^{w+\nu}\Gamma(w+\nu)}{d_i^{w+\nu}} P_1(w+\nu, \frac{d_i}{2}) E[\Phi(S_i^{1/2}A_i)],$$

$$I_i^{\phi}(w) \;\; = \;\; \frac{\nu 2^{w+\nu}\Gamma(w+\nu)}{\sqrt{2\pi}(d_i+A_i^2)^{w+\nu}} P_1(w+\nu, \frac{d_i+A_i^2}{2}),$$

where $S_i \sim Gamma(w+\nu, \frac{d_i}{2})\mathbb{I}_{(0,1)}$.

- *Contaminated skew–normal:*

$$I_i^{\Phi}(w) \;\; = \;\; \sqrt{2\pi}\{\nu\gamma^{w-1/2}\phi_1(d_i|0, \frac{1}{\gamma})\Phi(\gamma^{1/2}A_i) + (1-\nu)\phi_1(d_i|0,1)\Phi(A_i)\}$$

$$I_i^{\phi}(w) \;\; = \;\; \nu\gamma^{w-1/2}\phi_1(d_i+A_i^2|0, \frac{1}{\gamma}) + (1-\nu)\phi_1(d_i+A_i^2).$$

In many situations the derivatives of $\log \mathbf{\Sigma}_i$, d_i, and A_i involve complicated algebraic manipulation. For SNI-MEM, the derivatives of $\log \mathbf{\Sigma}$, d_i, and A_i can be found in Lachos et al. (2007) [20]. Asymptotic confidence intervals and test on the maximun likelihood estimators can be obtained using this matrix. Thus, if $\mathbf{J} = -\mathbf{L}$ denotes the observed information matrix for the marginal log-likelihood $\ell(\boldsymbol{\theta})$, then asymptotic confidence intervals and hypotheses tests for the parameter $\boldsymbol{\theta} \in \mathbb{R}^q$ are obtained once we assume the MLE $\widehat{\boldsymbol{\theta}}$ has approximately a $N_q(\boldsymbol{\theta}, \mathbf{J}^{-1})$ distribution. In practice, $\mathbf{J}$ is usually unknown and has to be replaced by its maximun likelihood estimation $\widehat{\mathbf{J}}$, that is, the matrix $\widehat{\mathbf{J}}$ evaluated at $\hat{\boldsymbol{\theta}}$. More generally speaking, for models as those in Proposition 3.7, the observed information matrix can be derived from the results given here.

6. Some examples

We illustrate the usefulness of the proposed class of distributions by applying them to two real data sets. The first example is an application of the methodology for univariate SNI responses, while the second is an application of SNI-MEM with $p = 5$.

6.1   Fiber-glass data set

In this section we apply four specific distributions of the skew normal/independent class, specifically, the univariate skew-normal, skew-t, skew-slash, and skew-contaminated normal, to fit the data on the breaking strength of 1.5cm long glass fiber, consisting of 63 observations. Jones and Faddy (2003) [19] and Wang and Genton (2006) [31] had previously analyzed this data with a skew-t and a skew-slash distribution, respectively. They both reported a strong presence of skewness on the left as well as a heavy-tailed behavior of the data, as depicted in Figure 4. We compare in the sequel the skew-normal (SN), skew-t (ST), contaminated skew-normal (SCN), and skew-slash (SSL) fitting for this data set. The resulting parameter estimates for the four models is given in Table 1. As suggested by Lange et al. (1989) [23], for each model the Schwarz information criterion was used for choosing the value of ν. This strategy is illustrated in Figure 5. Figure 4 shows the histogram of the fiber data superimposed with the fitted curves of the densities from the four considered models. We observe that the contaminated skew-normal fits the fiber data better than the other three distributions, especially at the peak part of the histogram. This conclusion is also supported by the log-likelihoods given in Table 1. Replacing the ML estimates of $\boldsymbol{\theta}$ in the Mahalanobis distance $d_i = (y_i - \mu)^2/\sigma^2$, we present Q-Q plots and envelopes in Figure 6 (lines represent the 5th percentile, the mean, and the 95th percentile of 100 simulated points for each observation). Plots in Figure 6 again provide strong evidence that the SNI distributions provides a better fit than the skew-normal distribution.

6.2   Chipkevitch et al. (1996) [12] data set

In this application, the multivariate skew-normal, skew-t, skew-slash, and skew-contaminated normal distributions are applied to fit the data studied by Chipkevitch et al. (1996) [12], where measurements of the testicular volume of 42 adolescents were converted to certain sequences by five different techniques: ultrasound (US), a graphical method proposed
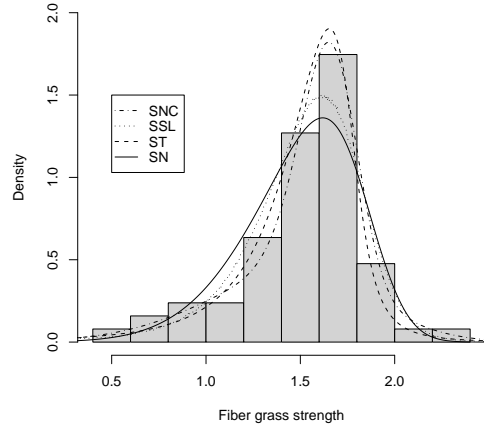
Figure 4: The histogram of the fiber grass strength superimposed with the fitted densities curves of the four distributions.

distribution	$\widehat{\mu}$	$\widehat{\sigma}$	$\widehat{\lambda}$	ν	γ	Log-likelihood
SN	1.8503	0.4705	-2.6789	-	-	
	(0.0468)	(0.0464)	(0.7513)			-13.9571
ST	1.7549	0.2725	-1.6196	3	-	
	(0.045)	(0.0353)	(0.6523)			-11.7546
SCN	1.7241	0.1615	-1.2940	0.5	0.10	
	(0.0393)	(0.0187)	(0.5080)			-9.1928
SSL	1.805591	0.2989667	-1.870298	1.7	-	
	(0.0461)	(0.0366)	(0.6320)			-12.9367

Table 1: MLE of the four fitted models on the fiber grass strength data set. Standard errors are based on the observed information matrix of Section 5.

(a)   **(b)**
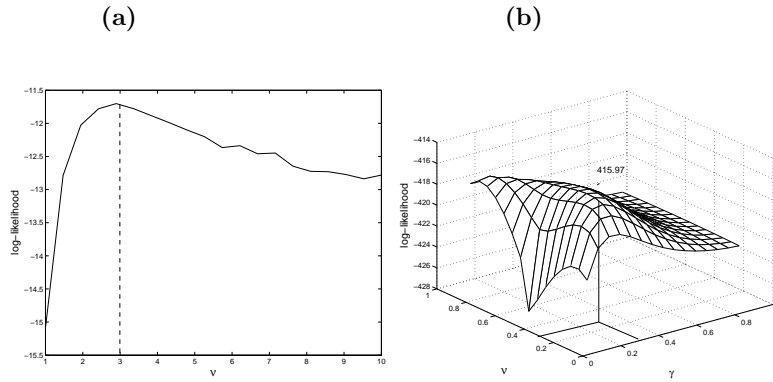
Figure 5: Plot of the profile log-likelihood for fitting (a) a skew-t model for fiber grass strength, (b) a contaminated skew-normal model for Chipkevitch data.

by the authors (I), dimensional measurement (II), Prader orchidometer (III), and ring orchidometer (IV). The ultrasound approach is assumed to be the reference measurement device. A histogram of the measurements (see Figure 7c) shows a certain asymmetry in the data set so that Galea-Rojas et al. (2002) [14] recommended a cubic root transformation to achieve better normality. Resulting parameter estimates for the four models are given in Table 2. The AIC criterion was used for choosing among some values of ν. For the ST model we found $\nu = 6$, for the SSL $\nu = 3$, and for the SCN $\nu = 0.3$, $\gamma = 0.3$. Therefore, for the three models a heavy-tailed distribution will be assumed. We can note from Table 2 that the intercept and slope estimates are similar among the four fitted models. However, the standard errors of the SNI distributions are smaller than the ones for the skew-normal model, indicating that the three models with longer than skew-normal tails seem to produce more accurate maximum likelihood estimates. The estimates for the variance components are not comparable since they are in a different scale. Note also that the log-likelihood values, shown at the bottom of Table 2, favor the SNI models. Particularly, we can see that the skew-t distribution fits the data better than the other three. The plots in Figure 7 provide even
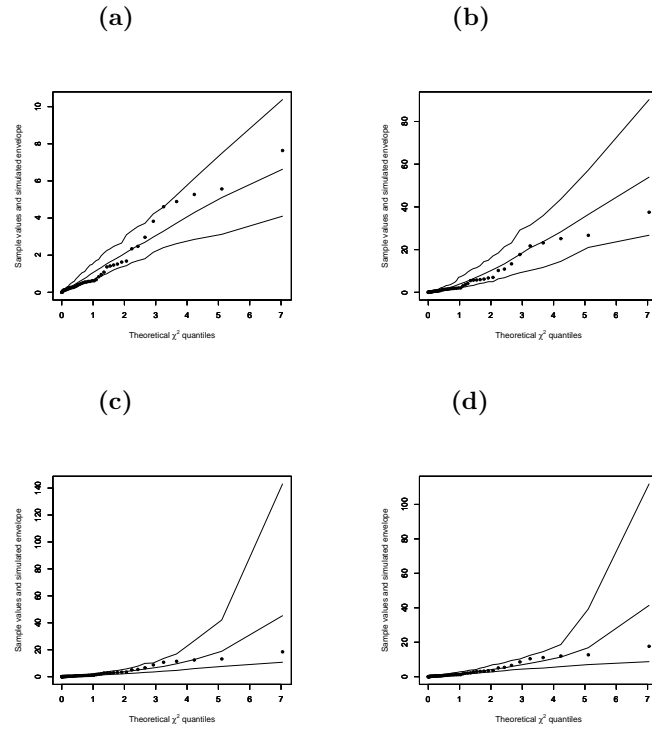
(a) **(b)**

(c) **(d)**

Figure 6: Fiber grass strength data set. Q-Q plots and simulated envelopes: (a) skew-normal model, (b) contaminated skew-normal model, (c) skew-t model, and (d) skew-slash model.

stronger evidence that the ST distribution allows a better fit to the data than the SN distribution.
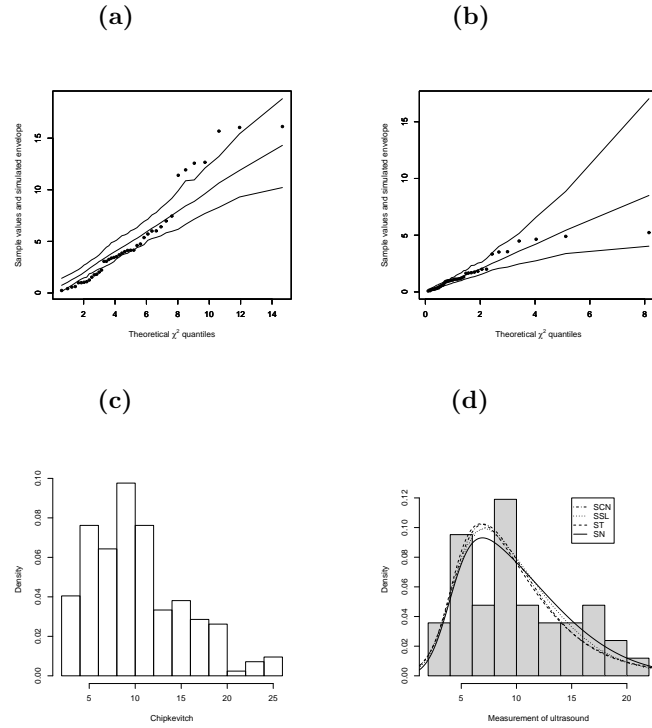
(a)

(b)

(c)

(d)

Figure 7: The Chipkevitch data set: (a) Q-Q plots and simulated envelopes for skew-normal model, (b) skew-t model, (c) histogram of the observed measurement, (d) histogram of the reference device measurement with superimposed fitted SNI densities.

Parameter	SN		ST		SCN		SSL	
	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE
α_1	0.1022	0.5655	0.0426	0.4940	0.1559	0.5017	0.1226	0.5317
α_2	-0.0096	0.6216	-0.2472	0.5261	-0.1527	0.5227	-0.1716	0.5777
α_3	0.0483	0.6277	0.1110	0.5674	0.1464	0.5643	0.1027	0.5999
α_4	1.5391	0.6337	1.5444	0.5605	1.6404	0.5729	1.5784	0.5978
β_1	0.8838	0.0509	0.8990	0.0513	0.8911	0.0511	0.8887	0.0517
β_2	0.9495	0.0559	0.9866	0.0565	0.9782	0.0547	0.9754	0.0577
β_3	1.1419	0.0565	1.1537	0.0586	1.1540	0.0574	1.1466	0.0583
β_4	1.0826	0.0570	1.0957	0.0579	1.0885	0.0584	1.0864	0.0578
ϕ_1	1.3384	0.3714	0.9291	0.2893	0.7467	0.2240	0.8345	0.2508
ϕ_2	1.3284	0.3480	0.9538	0.2841	0.7972	0.2225	0.8405	0.2369
ϕ_3	1.6736	0.4322	0.9028	0.2960	0.7294	0.2105	0.8938	0.2849
ϕ_4	1.1578	0.3710	0.9481	0.3160	0.7845	0.2537	0.7998	0.2581
ϕ_5	1.4105	0.3994	1.0196	0.3385	0.9119	0.2761	0.9080	0.2783
μ_x	3.9952	1.3958	4.1688	1.0890	4.2784	1.3844	4.1830	1.3096
σ_x^2	59.2857	21.5487	38.1512	14.3057	30.9914	13.6063	35.0036	13.3506
λ_x	4.7842	4.7925	3.4300	2.4361	3.2404	2.8759	3.7562	3.2021
ν	-	-	6	-	0.3	-	3	-
γ	-	-	-	-	0.3	-	-	-
log-likelihood	-422.1628		-416.7776		-415.9791		-419.3461	

Table 2: Results of fitting skew-normal and SNI-MEM to the Chipkevitch data. Standard errors are based on the observed information matrix of Section 5.

7. Conclusion

In this work we have defined a new family of asymmetric models by extending the symmetric normal/independent family. Our proposal generalizes results by Azzalini and Capitanio (2003) [6], Gupta (2003) [17], and Wang and Genton (2006) [31]. In addition, we have developed a general method based on the EM algorithm for estimating the parameters of the skew-normal/independent distributions. Closed-form expressions were derived for the iterative estimation processes based on the fact that the proposed distributions possess a stochastic representation that can be used to represent them hierarchically. This stochastic representation also allows us to study many of its properties easily. We believe that the approaches proposed here can be applied to other asymmetric multivariate models like those proposed by Branco and Dey (2001) [9, Section 3]. The assessment of influence of data and model assumption on the result of the statistical analysis is a key aspect of any new class of distribution. We are currently exploring the local influence and residual analysis to address this issue.

8. Appendix: some lemmas

Now we take care of some technical lemmas needed in Section 3.

Lemma 8.1. *Let* $\mathbf{Y} \sim SN_p(\boldsymbol{\lambda})$. *Then for any fixed p-dimensional vector* $\mathbf{b}$ *and a* $p \times p$ *matrix* $\mathbf{A}$ *we have*

$$E[\mathbf{Y}^\top \mathbf{A} \mathbf{Y} \mathbf{b}^\top \mathbf{Y}] = -\sqrt{\frac{2}{\pi}}[(\boldsymbol{\delta}^\top \mathbf{A} \boldsymbol{\delta} + tr(\mathbf{A}))\mathbf{b}^\top \boldsymbol{\delta} + 2\boldsymbol{\delta}^\top \mathbf{A} \mathbf{b}],$$

where $\boldsymbol{\delta}$ *is as in (3.3).*

Proof. The proof follows by the stochastic representation of $\mathbf{Y}$ given in (1.2) and the calculation of the moments $E[|X_0|]$ and $E[|X_0|^3]$, when $X_0 \sim N(0,1)$. $\qquad\qquad\square$

Lemma 8.2. *Let* $\mathbf{Y} \in \mathbb{R}^p$ *be a random vector with*

$$f(\mathbf{y}|u) = k^{-1}(u)\phi_p(\mathbf{y}|\boldsymbol{\mu}, u^{-1}\boldsymbol{\Sigma})\Phi_1(u^{1/2}A + u^{1/2}\mathbf{B}^\top\mathbf{y})$$

as probability density function, with u *a positive constant,* $A \in \mathbb{R}$, $\mathbf{B}$ *a* p-*dimensional vector, and* $k(u) = \Phi_1(u^{1/2}\dfrac{A + \mathbf{B}^\top\boldsymbol{\mu}}{\sqrt{1 + \mathbf{B}^\top\boldsymbol{\Sigma}\mathbf{B}}})$ *a standardized constant. Then we have*

$$E[\mathbf{Y}|u] = \boldsymbol{\mu} + u^{-1/2}\frac{\boldsymbol{\Sigma}\mathbf{B}}{\sqrt{1 + \mathbf{B}^\top\boldsymbol{\Sigma}\mathbf{B}}}W_{\Phi_1}(u^{1/2}\frac{A + \mathbf{B}^\top\boldsymbol{\mu}}{\sqrt{1 + \mathbf{B}^\top\boldsymbol{\Sigma}\mathbf{B}}}).$$

Proof. If we notice, by using Lemma 2 from Arellano-Valle et al. (2005) [2], that

$$
\begin{aligned}
E[\mathbf{Y}|u] &= k^{-1}(u)\int_{\mathbb{R}}\int_0^\infty \mathbf{y}\phi_1(t|u^{1/2}A + u^{1/2}\mathbf{B}^\top\mathbf{y}, 1)\phi(\mathbf{y}|\boldsymbol{\mu}, u^{-1}\boldsymbol{\Sigma})dtd\mathbf{y} \\
&= k^{-1}(u)\int_0^\infty \phi_1(t|u^{1/2}A + u^{1/2}\mathbf{B}^\top\boldsymbol{\mu}, 1 + \mathbf{B}^\top\boldsymbol{\Sigma}\mathbf{B})E_{\mathbf{Y}|t}[\mathbf{Y}]dt
\end{aligned}
$$

holds, where $\mathbf{Y}|t \sim N_p(\boldsymbol{\mu} - \boldsymbol{\Lambda}\mathbf{B}(A + \mathbf{B}^\top\boldsymbol{\mu}) + u^{-1/2}\boldsymbol{\Lambda}\mathbf{B}t, u^{-1}\boldsymbol{\Lambda})$, with $\boldsymbol{\Lambda} = (\boldsymbol{\Sigma}^{-1} + \mathbf{B}\mathbf{B}^\top)^{-1}$, then the proof follows from well known properties of the truncated normal distribution (compare Johnson et al. (1994) [18, Section 10.1]). $\square$

Lemma 8.3. *Let* $\mathbf{Y} \sim Gamma(\alpha, \beta)$. *Then for any* $a \in \mathbb{R}$ *we have*

$$E[\Phi_1(a\sqrt{(Y)})] = T_1(a\sqrt{\frac{\alpha}{\beta}}|0, 1, 2\alpha).$$

Proof. See Azallini and Capitanio (2003) [6, Lemma 1]. $\square$

References

[1] Andrews, D. F. and Mallows, C. L.; *Scale mixtures of normal distributions.* Journal of the Royal Statistical Society, Series B, **36**, 99-102, 1974.

[2] Arellano-Valle, R. B., Bolfarine, H. and Lachos, V. H.; *Skew-normal linear mixed models.* Journal of Data Science, **3**, 415-438, 2005.

[3] Arnold, B. C. and Beaver, R. J.; *Skewed multivariate models related to hidden truncation and/or selective reporting.* Test, **11**, 7-54, 2002.

[4] Azzalini, A.; *A class of distributions which includes the normal ones.* Scandinavian Journal of Statistics, **12**, 171-178, 1985.

[5] Azzalini, A., Capitanio, A.; *Statistical applications of the multivariate skew normal distribution.* Journal of the Royal Statistical Society ,Series B, **61**, 579-602, 1999.

[6] Azzalini, A., Capitanio, A.; *Distributions generated y perturbation of symmetry with emphasis on the multivariate skew t-distribution.* Journal of the Royal Statistical Society, Series B, **61**, 367-389, 2003.

[7] Azzalini, A. and Dalla-Valle, A.; *The multivariate skew-normal distribution.* Biometrika, **83**, 715-726, 1996.

[8] Barnett, V.D.; *Simultaneous pairwise linear structural relationships.* Biometrics, 25, 129-142, 1969.

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

[10] Bolfarine, H. and Galea-Rojas, M.; *Maximum likelihood estimation of simultaneous pairwise linear structural relationships.* Biometrical Journal, 37, 673-689, 1995.

[11] Bolfarine, H. and Galea-Rojas, M.; *On structural comparative calibration under a t-model.* Computational Statistics, 11, 63-85, 1996.

[12] Chipkevitch, E., Nishimura, R., Tu, D. and Galea-Rojas, M.; *Clinical measurement of testicular volume in adolescents: Comparison of the reliability of 5 methods.* Journal of Urology, 156, 2050-2053, 1996.

[13] Diciccio, T. and Monti, A. C.; *Inferential aspects of the skew exponential power distribution.* Journal of the American Statistical Association, 99, 439-450, 2004.

[14] Galea-Rojas, M., H. and de Castro, M. *Local influence in comparative calibration models.* Biometrical Journal, 44(1), 59-81, 2002.

[15] Genton, M. G., He, L. and Liu, X.; *Moments of skew-normM random vectors and their quadratic forms.* Statistics and Probability Letters, 51,319-325, 2001.

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

[17] Gupta A. K.; *Multivariate skew t-distributions.* Statistics, **37(4)**, 359-363, 2003.

[18] Johnson, N. L., Kotz, S., Balakrishnan, N.; *Continuous Univariate Distributions, vol. 1.* New York: John Wiley, 1994.

[19] Jones, M. C. and Faddy, M. J.; *A skew extension of the s distribution, with applications.* Journal of the Royal Statistical Society, Series B, **65**, 159-174, 2003.

[20] Lachos, V. H., Bolfarine, H., Arellano-Valle, R. B. and Montenegro, L. C.; *Likelihood based inference for multivariate skew-normal regression models.* Communications in Statistics: Theory and Methods, **36**, 1769-1786, 2007.

[21] Lachos, V. H., Bolfarine, H., Vilca, L. F., Galea–Rojas, M.; *Estimation and influence diagnostic for structural comparative calibration models under the skew-normal distribution.* Technical report, RT-MAE 2005-12, IME-USP, 2005.

[22] Lange, K. and Sinsheimer.; *Normal/independent distributions and their applications in robust regression.* Journal of Computational and Graphical Statistics, **2**, 175-198, 1993.

[23] Lange, K. L., Little, J. A. and Taylor, M. G. J.; *Robust statistical modeling using the t distribution.* Journal of the American Statistical Association, **84**, 881-896, 1989.

[24] Little, R. J. A.; *Robust Estimation of the mean and covariance matrix from data with missing values.* Applied Statistics, **37**, 23-38, 1988.

[25] Liu, C.; *Bayesian robust linear regression with incomplete data.* Journal of the American Statistical Association, **91**, 1219-1227, 1996.

[26] Mardia, K. V.; *Applications of some measures of multivariate skewness and kurtosis in testing normality and robustness studies.* Sankhyá B, 36(2), 115-128, 1974.

[27] Sahu, S. K., Dey, D. K. and Branco, M. D.; *A new class of multivariate skew distributions with applications to Bayesian regression models.* The Canadian Journal of Statistics, **31**, 129-150, 2003.

[28] Shyr, I. and Gleser, L.; *Inference about comparative precision in linear structural relationships.* Journal of Statistical Planning and Inference, 14, 339-358, 1986.

[29] Tan, F. and Peng, H.; *The slash and the skew-slash Student t distributions.* Preprint, 2006.

[30] Wang, J., Boyer, J. and Genton, M.; *A skew-symmetric representation of multivariate distributions.* Statistica Sinica, **14**, 1259-1270, 2004.

[31] Wang, J. and Genton, M.; *The multivariate skew-slash distribution.* Journal of Statistical Planning and Inference, **136**, 209-220, 2006.

[32] Zhu, H. and Lee S.; *Local influence for incomplete-data models.* Journal of the Royal Statistical Society, Series B, **63**, 111-126., 2001.

Resumen

Liu (1996) discute una clase de distribuciones robustas a las que apela como normal/independiente, y que contiene un grupo de distribuciones de colas pesadas. En este artículo desarrollamos una versión asimétrica de tales distribuciones en un escenario multivariado, a las que llamaremos distruciones normales asimétricas independientes multivariadas. Para tales distribuciones derivamos algunas propiedades. La principal virtud de los miembros de esta familia es que son fáciles de simular y se prestan a un algoritmo de tipo EM para realizar estimaciones de máxima verosimilitud de sus parámetros. Para dos modelos multivariados de interés práctico se discute el algoritmo EM con énfasis en las distribuciones t-asimétrica, slash asimétrica y normal asimétrica contaminada. Los resultados obtenidos a partir de simulaciones y de dos conjuntos de datos reales son reportados.

Palabras clave: Algoritmo EM, normal/independiente.

Victor Hugo Lachos
Departamento de Estatística, IMECC
Universidade Estadual de Campinas
CEP 13083-859, Campinas, São Paulo, Brazil
hlachos@ime.unicamp.br

Filidor V. Labra
Departamento de Estatística, IMECC
Universidade Estadual de Campinas
CEP 13083-859, Campinas, São Paulo, Brazil
fily@ime.unicamp.br