Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t distribution

ADELCHI AZZALINI
Dipartimento di Scienze Statistiche, Università di Padova
azzalini@stat.unipd.it

ANTONELLA CAPITANIO
Dipartimento di Scienze Statistiche, Università di Bologna
capitani@stat.unibo.it

23rd September 2002
(Original version April 2001, last amendement 27th June 2003)

This is the full-length paper whose abriged version appears in
J. Roy. Statist. Soc., series B *vol. 65 (2003), pp. 367–389*

Abstract

A fairly general procedure is studied to perturbate a multivariate density satisfying a weak form of multivariate symmetry, and to generate a whole set of non-symmetric densities. The approach is general enough to encompass a number of recent proposals in the literature, variously related to the skew normal distribution. The special case of skew elliptical densities is examined in detail, establishing connections with existing similar work. The final part of the paper specializes further to a form of multivariate skew t density. Likelihood inference for this distribution is examined, and it is illustrated with numerical examples.

Key-words: asymmetry, central symmetry, elliptical distributions, Healy's plot, multivariate t distribution, quadratic forms, skewness, skew normal distribution.

1 INTRODUCTION

1.1 MOTIVATION AND AIMS

There is a growing interest in the literature on parametric families of multivariate distributions which represent a local departure from the multivariate normal family, in the sense that they exhibit a bell-shaped behaviour similar to the normal density, and they can be made arbitrarily close to the normal density by regulating a suitable parameter. The phrase 'local departure' must be interpreted appropriately, in the sense that, while these families can approach normality, they also can, under other circumstances, exhibit quite a substantial departure from normality.

The motivation of these efforts is to introduce more flexible parametric families capable of adapting as closely as possible to real data, in particular in the rather frequent case of phenomena whose empirical outcome behaves in a non-normal fashion but still retains some broad similarity with the multivariate normal distribution. Typically this departure from normality occurs in the form of a roughly bell-shaped density, but with contour levels not quite elliptically shaped and/or with contour levels not quite spaced as the normal density prescribes.

Some of this literature is connected with the so-called multivariate skew normal (SN) distribution, recently studied by Azzalini & Dalla Valle (1996) and Azzalini & Capitanio (1999); this has been further developed by other authors whose work will be referenced later in this section. The d-dimensional SN density, in the 'standard' form which does not include location and scale parameters, is

$$2\,\phi_d(y;\bar{\Omega})\,\Phi(\alpha^\top y), \qquad y \in \mathbb{R}^d, \tag{1}$$

where $\phi_d(y;\bar{\Omega})$ is the $N_d(0,\bar{\Omega})$ density at y for some correlation matrix $\bar{\Omega}$, $\Phi(\cdot)$ is the $N(0,1)$ distribution function and $\alpha \in \mathbb{R}^d$. Here α plays the role of shape parameter; when $\alpha = 0$, we recover the regular normal density.

As a further level of generalisation of the normal distribution, Azzalini & Capitanio (1999, p. 599) have presented a lemma which leads to the construction of a 'skew elliptical' density, which is an elliptical density multiplied by a suitable skewing factor, in such a way that the product is still a proper density. Branco & Dey (2001) have considered another form of skew elliptical distribution, whose connections with the one mentioned above will be discussed extensively in this paper. Other work on extensions of elliptical families has been done by Genton & Loperfido (2002), where it is shown that distributional properties of certain functions of elliptical variates extends to their skewed variants, generalizing a similar result of Branco &Dey (2001).

Arnold & Beaver (2000a) have studied a variant of (1) which replaces the argument of Φ by $\alpha_0 + \alpha^\top y$, where α_0 is an additional parameter, with consequent adjustment of the normalising constant. The same variant of the SN distribution has been considered by Capitanio *et al.* (2003) in the context of graphical models. Sahu, Dey & Branco (2001) have studied yet another form of skew elliptical distribution, where the skewing factor is a d-dimensional distribution function, rather than a scalar one like those of the previously mentioned cases. In the same spirit as (1), Arnold & Beaver (2000b) have studied a form of multivariate skew Cauchy distribution. For additional references and a recent review on the connected literature, see Arnold & Beaver (2002).

There is therefore a set of interesting developments in various directions aimed at extending (1) or adapting the underlying idea to other distributions. While all this activity is definitely promising and appealing, it also brings in the question of the inter-relationships among these contributions, which tend to appear as scattered in different directions.

One purpose of the present contribution is to propose a fairly general extension of (1); in addition, a better understanding of the connections and similarities among some of the above-described proposals is attempted. A broad formulation is presented in Section 2, and is specialised to a skew elliptical form in Section 3. This approach encompasses several of the existing proposals and it appears to provide a potentially general framework for special cases. We discuss in some detail a few of these and, from Section 4 onwards, we focus on a form of multivariate skew t distribution; since this represents a mathematically quite manageable distribution, allowing ample flexibility in skewness and kurtosis, it appears to be a promising tool for a wide range of practical problems. Associated likelihood inference for this skew t distribution and illustrative examples are presented in Section 5. Some background information on the SN distribution and the elliptical family is given in the second part of this introductory section.

1.2 Some preliminaries

The SN distribution Given a full-rank $d \times d$ covariance matrix $\Omega = (\omega_{rs})$, define

$$\omega = \mathrm{diag}(\omega_1, \ldots, \omega_d) = \mathrm{diag}(\omega_{11}, \ldots, \omega_{dd})^{1/2}$$

and let $\bar{\Omega} = \omega^{-1}\Omega\omega^{-1}$ be the associated correlation matrix; also let $\xi,\ \alpha \in \mathbb{R}^d$. A d-dimensional random variable Z is said to have a skew normal distribution if it is continuous with density function at $z \in \mathbb{R}^d$ of type

$$2\,\phi_d(z-\xi;\Omega)\,\Phi(\alpha^\top\omega^{-1}(z-\xi)). \tag{2}$$

We shall then write $Z \sim \mathrm{SN}_d(\xi, \Omega, \alpha)$, referring to ξ, Ω, α as the location, dispersion and shape or skewness parameters, respectively. Density (1) corresponds to the 'standard' distribution $\mathrm{SN}_d(0, \bar{\Omega}, \alpha)$.

By varying α, one obtains a variety of shapes; Azzalini & Dalla Valle (1996) display graphically some instances of them when $d = 2$. Clearly, when $\alpha = 0$, we are back to the $\mathrm{N}_d(\xi, \Omega)$ density. The cumulant generating function is

$$K_Z(t) = t^\top\xi + \tfrac{1}{2}t^\top\Omega t + \zeta_0(\delta^\top\omega t)$$

where

$$\delta = \frac{1}{\left(1+\alpha^\top\bar{\Omega}\alpha\right)^{1/2}}\bar{\Omega}\alpha, \qquad \zeta_0(x) = \log\{2\,\Phi(x)\}. \tag{3}$$

From the expression for δ we have

$$\alpha = \frac{1}{\left(1-\delta^\top\bar{\Omega}^{-1}\delta\right)^{1/2}}\,\bar{\Omega}^{-1}\delta. \tag{4}$$

There exists at least two stochastic representations for Z. These are useful for random number generation and for deriving in a simple way a number of formal properties.

- ⋄ *Conditioning method.* Suppose that U_0 is a scalar random variable and U is a d-dimensional variable, such that

$$\begin{pmatrix} U_0 \\ U \end{pmatrix} \sim \mathrm{N}_{d+1}\left(0, \Omega^*\right), \qquad \Omega^* = \begin{pmatrix} 1 & \delta^\top \\ \delta & \bar{\Omega} \end{pmatrix} \tag{5}$$

where Ω^* is a full-rank correlation matrix. Then the distribution of $(U|U_0 > 0)$ is $\mathrm{SN}_d(0, \bar{\Omega}, \alpha)$ where α is a function of δ and $\bar{\Omega}$; in fact, we can also set

$$Z = \begin{cases} U & \text{if } U_0 > 0, \\ -U & \text{if } U_0 < 0. \end{cases}$$

By an affine transformation of the resulting variable one obtains a distribution of type (2).

◇ *Transformation method.* Suppose now that

$$\begin{pmatrix} U_0' \\ U' \end{pmatrix} \sim \mathrm{N}_{d+1}\left(0, \begin{pmatrix} 1 & 0 \\ 0 & \Psi \end{pmatrix}\right) \tag{6}$$

where Ψ is a full-rank correlation matrix, and define

$$Z_j = \delta_j\,|U_0'| + \left(1-\delta_j^2\right)^{1/2}\,U_j', \tag{7}$$

where $-1 < \delta_j < 1$ for $j = 1, \ldots, d$. Then $(Z_1, \ldots, Z_d)$ has the d-dimensional skew normal distribution, with parameters which are suitable functions of the δ's and Ψ.

A third type of representation is known to exist in the scalar case. If (U_0, U_1) is a bivariate normal variate with standardized marginals and correlation ρ, then

$$\max(U_0, U_1) \sim \mathrm{SN}(0, 1, \alpha) \tag{8}$$

where $\alpha = ((1-\rho)/(1+\rho))^{1/2}$. This result has been given by Roberts (1966), in an early explicit occurrence of the scalar SN distribution, and later rediscovered by Loperfido (2002); the same conclusion can also be obtained as special case of a result of H. N. Nagaraja, quoted by David (1981, Exercise 5.6.4). The generalization of this type of representation to the multivariate setting to obtain (1) via a set of $\max(\cdot)$ operation on normal variates is an open question.

Among the many formal properties shared with the normal class, a noteworthy fact is that

$$(Z-\xi)^\top \Omega^{-1}(Z-\xi) \sim \chi_d^2. \tag{9}$$

Other properties of quadratic forms of SN variables are given by Azzalini & Capitanio (1999), Genton *et al.* (2001) and Loperfido (2001). Another important property of this class is closure under affine transformations of the variable Z; in particular, this implies closure under marginalization, i.e. the distribution of all sub-vectors of Z is still of type (2).

What is lacking is closure under conditioning, i.e. the conditional distribution of a set of components of Z given another set of components is not of type (2). This property is achieved by a simple extension of (2) which has been examined by Arnold & Beaver (2000a) and by Capitanio *et al.* (2003). This variant of the density takes the form

$$\Phi(\tau)^{-1}\,\phi_d(z-\xi;\Omega)\,\Phi(\alpha_0 + \alpha^\top\omega^{-1}(z-\xi)) \tag{10}$$

where τ ($\tau \in \mathbb{R}$) is an additional parameter and

$$\alpha_0 = \left(1 - \delta^\top\bar{\Omega}^{-1}\delta\right)^{-1/2}\tau\,.$$

When $\tau = 0$, $\alpha_0 = 0$ and (10) reduces to (2). Unfortunately, the χ^2 property (9) does not hold for (10), if $\tau \neq 0$. A form of genesis of (10) via conditioning using (6) is by consideration of $(U|U_0 + \tau > 0)$.

Elliptical distributions We summarize briefly a few concepts about and establish notation for elliptical distributions, confining ourselves to random variables without discrete components. For a full treatment of this topic, we refer the reader to Fang, Kotz and Ng (1990).

A d-dimensional continuous random variable Y is said to have an elliptical density if this is of the form

$$f(y;\xi,\Omega) = \frac{c_d}{|\Omega|^{1/2}}\,\tilde{f}\{(y-\xi)^\top\Omega^{-1}(y-\xi)\}, \qquad y\in\mathbb{R}^d,$$

where $\xi\in\mathbb{R}^d$, Ω is a covariance matrix, $\tilde{f}$ is a suitable function from $\mathbb{R}^+$ to $\mathbb{R}^+$, called the 'density generator', and c_d is a normalising constant. We shall then write $Y\sim \mathrm{Ell}_d(\xi,\Omega,\tilde{f})$.

The basic case is obtained by setting $\tilde{f}(x) = \exp(-x/2)$ and $c_d = (2\pi)^{-d/2}$, leading to the multivariate normal density. Two other important special cases, which will be used extensively in the sequel, are provided by the multivariate Pearson type VII distributions, whose generator and normalising constant are

$$\tilde{f}(x) = (1+x/\nu)^{-M}, \qquad c_d = \frac{\Gamma(M)}{(\pi\nu)^{d/2}\,\Gamma(M-d/2)},$$

where $\nu>0, M>d/2$, and by the multivariate Pearson type II distributions for which

$$\tilde{f}(x) = (1-x)^{\nu}, \qquad c_d = \frac{\Gamma(d/2+\nu+1)}{\pi^{d/2}\,\Gamma(\nu+1)}$$

where $0\le x\le 1, \nu>-1$. The special importance of type VII lies in the fact that it includes the multivariate t density when $M=(d+\nu)/2$, hence also the Cauchy distribution. For these distributions, we shall use the notation $\mathrm{PVII}_d(\xi,\Omega,M,\nu)$ and $\mathrm{PII}_d(\xi,\Omega,\nu)$, respectively.

A convenient stochastic representation for Y is

$$Y = \xi + RL^\top S \tag{11}$$

where $L^\top L = \Omega$, the random vector S is uniformly distributed on the unit sphere in $\mathbb{R}^d$ and R is a positive scalar random variable independent of S, called the generating variate. An immediate consequence of this representation is that $(Y-\xi)^\top\Omega^{-1}(Y-\xi) \overset{d}{=} R^2$, where $\overset{d}{=}$ means equality in distribution.

Elliptical distributions are closed under affine transformations and conditioning. In particular they are closed under marginalization, in the following sense: consider the block partition $Y^\top = (Y_1^\top, Y_2^\top)$ where $Y_1\in\mathbb{R}^h$ and a corresponding partition for ξ and Ω; then

$$Y_1 \sim \mathrm{Ell}_h(\xi_1,\Omega_{11},\tilde{f}_1)$$

Similarly, for the conditional density we have

$$(Y_1|Y_2=y_2)\sim \mathrm{Ell}_h(\xi_1+\Omega_{12}\Omega_{22}^{-1}(y_2-\xi_2), \Omega_{11}-\Omega_{12}\Omega_{22}^{-1}\Omega_{21}, \tilde{f}^{Q_y}).$$

where $Q_y = y_2^\top\Omega_{22}^{-1}y_2$. The density generators $\tilde{f}_1$ and $\tilde{f}^{Q_y}$ are not necessarily of the same form as $\tilde{f}$. Kano (1994) has shown that the form of the density generator is preserved under marginalization only in the case of elliptical distributions which can be obtained from a scale mixture of normal variates. This property is true, for instance, for multivariate Pearson type VII and II distributions. The generator $\tilde{f}^{Q_y}$ of the conditional distribution depends in general on the quantity Q_y, with the exception of the normal distribution.

2 CENTRAL SYMMETRY AND DISTRIBUTIONS OBTAINED BY ITS PERTURBATION

Our starting point is the following proposition which is closely connected to Lemma 1 of Azzalini & Capitanio (1999). Strictly speaking, the present statement is a bit more restricted than the earlier result, but it has the major advantage of requiring a set of conditions whose fulfillment is far simpler to check, and still it represents a very general formulation.

The result refers to central symmetry, a simple and wide concept of symmetry, which is commonly in use in nonparametric statistics; see Zuo & Serfling (2000). Other authors refer to the same property with alternative terms. A d-dimensional random variable Y is said to be centrally symmetric around a point ξ if $Y-\xi \stackrel{d}{=} \xi - Y$. Since we shall deal with continuous variables, the above requirement implies that the corresponding density function f satisfies $f(y-\xi) = f(\xi - y)$ for all $y \in \mathbb{R}^d$, up to a negligible set. It is immediate to see that the condition of central symmetry is satisfied by various ample families, notably the elliptical densities, but also many others; some examples are the symmetric stable laws, the Watson rotational symmetric densities, the class of distributions studied studied by Szabłowski (1998), among many others.

Proposition 1 *Denote by $f(y)$ the density function of a d-dimensional continuous random variable which is centrally symmetric around 0, and by G a scalar distribution function such that $G(-x) = 1-G(x)$ for all real x. If $w(y)$ is a function from $\mathbb{R}^d$ to $\mathbb{R}$ such that $w(-y) = -w(y)$ for all $y \in \mathbb{R}^d$, then*

$$2\, f(y)\, G\{w(y)\} \tag{12}$$

is a density function.

Proof. Denote by Y a random variable with density f, and by X a random variable with distribution function G, independent of Y. To show that $W = w(Y)$ has a distribution symmetric about 0, denote by A a Borel set of the real line and by $-A$ its mirror set obtained by reversing the sign of each element of A. Then, taking into account that Y and $-Y$ have the same distribution,

$$\mathbb{P}\{W \in -A\} = \mathbb{P}\{-W \in A\} = \mathbb{P}\{w(-Y) \in A\} = \mathbb{P}\{w(Y) \in A\}\,,$$

showing that W has the property indicated. Then, on noticing that $X - W$ has distribution symmetric about 0, write

$$\tfrac{1}{2} = \mathbb{P}\{X \le W\} = \mathbb{E}_Y\{\mathbb{P}\{X \le w(Y)|Y=y\}\} = \int_{\mathbb{R}^d} G\{w(y)\}\, f(y)\, \mathrm{d}y$$

which completes the proof.

To demonstrate graphically the ample flexibility attained by (12) for appropriate choices of f, G, and w, we present the following example in the case $d=2$. Consider the non-elliptical distribution

$$f(y) = \frac{(1-y_1^2)^{a-1}\,(1-y_2^2)^{b-1}}{4^{a+b-1}\,B(a,a)\,B(b,b)}, \qquad y = (y_1, y_2) \in (-1,1)^2,$$

obtained by multiplication of two symmetric Beta densities rescaled to the interval $(-1,1)$, with positive parameters a and b. We perturb this density by choosing

$$G(x) = \frac{e^x}{1+e^x}, \qquad w(y) = \frac{\sin(p_1 y_1 + p_2 y_2)}{1+\cos(q_1 y_1 + q_2 y_2)}$$

where p_1, p_2, q_1 and q_2 are additional parameters.

We have generated several plots of the above type of density, obtaining an extremely rich set of surfaces, as indicated by the small collection of such densities given in Figure 1. Additional regulation of the shape could be achieved, by inserting parameters in the logistic function $G(x)$, although it is doubtful that one would need the latter level of additional flexibility. The plots indicate that the effect of perturbing f via (12) is far more complex than the effect introduced, say, by the skewing factor of the normal density in (2). Clearly, the purpose of Figure 1 is purely illustrative, and it is not suggested to use the above class of density functions in practice without further investigation.

For a random variable with density (12), the stochastic representation given by Azzalini & Capitanio (1999, p. 599) for a slightly different case is still valid. In fact, the conditions required there for its validity are actually those of Proposition 1. Specifically, if Y has density function f and X is an independent variable with distribution function G, then

$$Z = \begin{cases} Y & \text{if } X < w(Y) \\ -Y & \text{if } X > w(Y) \end{cases} \tag{13}$$

has density function (12). Clearly, this provides an algorithm for generating Z and it will also turn out to be useful for theoretical purposes.

It can be shown that the conditioning method for generating skew normal random variables from (5) is a special case of (13). In fact, from consideration of the residual part of U_0 after removing the regression on U, define the variable

$$\tilde{X} = -\left(1 - \delta^\top \bar{\Omega}^{-1} \delta\right)^{-1/2} (U_0 - \delta^\top \bar{\Omega}^{-1} U) \sim N(0,1), \tag{14}$$

independent of U. After substituting symbols, the condition $\tilde{X} < \alpha^\top U$ of the top branch of (13) is equivalent to $U_0 > 0$, if α is given by (4); hence it generates a $\mathrm{SN}_d(0, \bar{\Omega}, \alpha)$ variable if we set $Z = U$. The condition of the lower branch is equivalent to $-\tilde{X} < \alpha^\top(-U)$ leading to a $\mathrm{SN}_d(0, \bar{\Omega}, -\alpha)$ variable if we set $Z = U$, hence to a $\mathrm{SN}_d(0, \bar{\Omega}, \alpha)$ variable if we set $Z = -U$.

Similarly, the stochastic representation of a variate with density (10) via $(U|U_0 + \tau > 0)$ could be reformulated in terms of the condition $\tilde{X} < \alpha_0 + \alpha^\top U$. In general, the existence of a similar correspondence would be unclear if the assumption of normality in (5) was replaced by some other distributional assumption. Luckily, a suitable transformation analogous to (14) can be obtained in a few important special cases to be discussed in Section 3.

It is immediate that, if f is an elliptical density, G corresponds to a distribution symmetric about 0 and $w(y) = \alpha^\top y$ for some $\alpha \in \mathbb{R}^d$, then the conditions required by Proposition 1 are fulfilled. We then obtain the family of densities produced by Corollary 2 of Azzalini & Capitanio (1999).

Proposition 2 *Denote by Y and Z two d-dimensional random variates having density function f and (12), respectively, satisfying the conditions of Proposition 1. If $t(\cdot)$ is a function from $\mathbb{R}^d$ to some Euclidean space, such that $t(-y) = t(y)$ for all $y \in \mathbb{R}^d$, then*

$$t(Y) \stackrel{d}{=} t(Z)\,.$$

Proof. This is immediate from representation (13).

A key example of the above result is obtained when $t(y)$ represents the distance from the origin. Since any choice of $t(\cdot)$ must satisfy the symmetry condition $t(y) = t(-y)$, then the probability

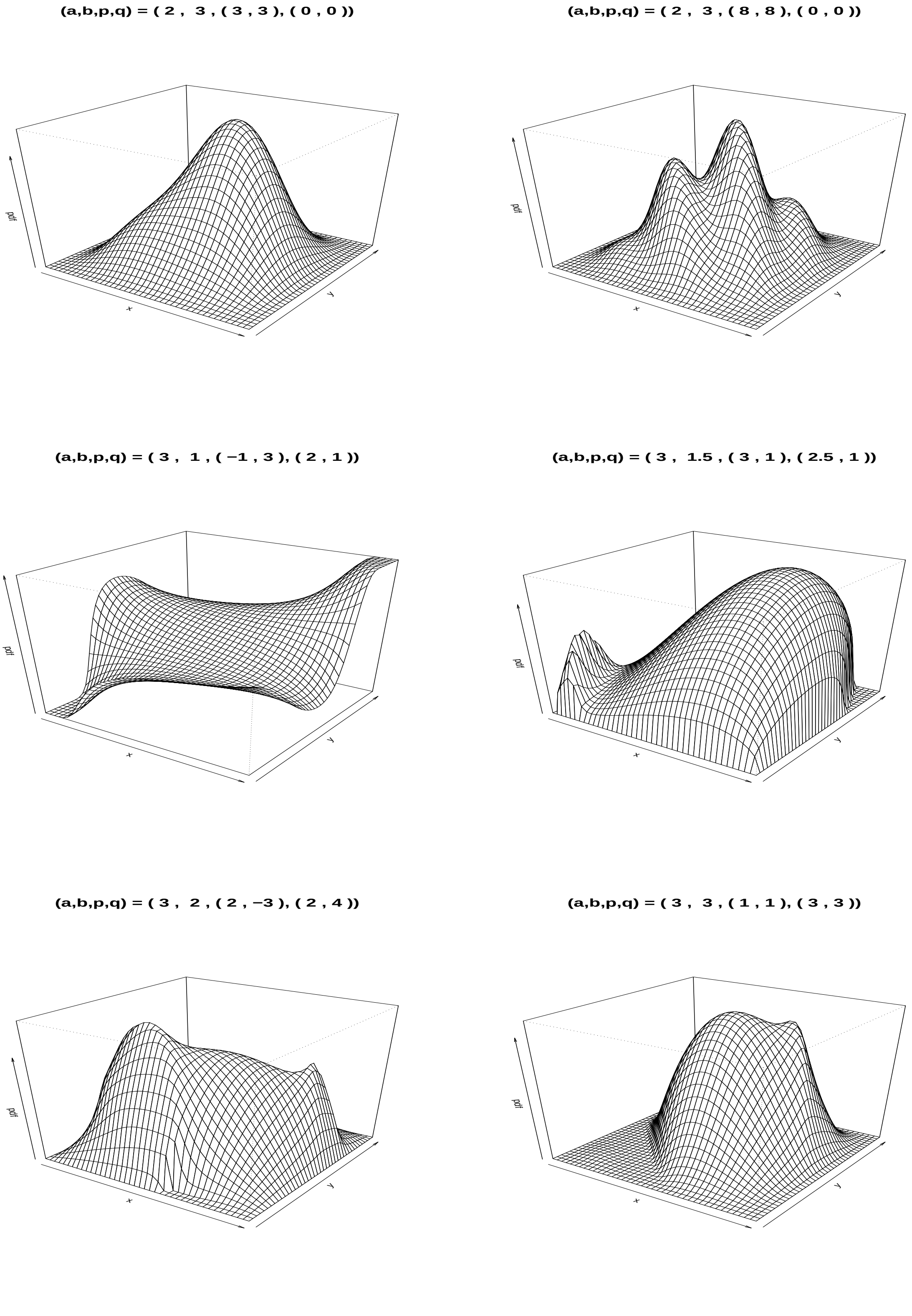

Figure 1: Examples of perturbed symmetric Beta densities. The set of parameters $(a, b, p_1, p_2, q_1, q_2)$ is shown at the top of each plot

distribution of the distance of a random point from the origin is the same for Y and for Z. In particular we can write $Y^\top BY \stackrel{d}{=} Z^\top BZ$ for any positive definite matrix B. A result similar to Proposition 2 for the case when f is an elliptical distribution has been given by Genton & Loperfido (2002).

A related set of applications of Proposition 2 is offered by various results on quadratic forms of skew normal variates, all of which lead to the conclusion that known distributional results for normal variates still hold if the variates are of skew normal type. This set of results includes Proposition 7, 8 and 9 of Azzalini & Capitanio (1999) and Proposition 1, 2 and 6 (parts 1 and 3) of Loperfido (2001). For these conclusions, one must consider functions $t(\cdot)$ in Proposition 2 taking on values in an appropriate Euclidean space, for instance $\mathbb{R}^+ \times \mathbb{R}^+$ if the independence of two quadratic forms is under consideration. Notice that Propositions 8 and 9 of Azzalini & Capitanio (1999) have added conditions on the α parameter, but these are not necessary. There is no conflict with the present conclusions since in their Proposition 8 this extra condition is part of a sufficiency requirement, and their Proposition 9 (a Fisher-Cochran type of theorem) was stated in a more restricted form than actually possible.

We conclude this section with a discussion on possible generalisations of Proposition 1. A very general form of density resembling (12) is along the following lines. Denote by $X = (X_1, \ldots, X_m)^\top$ an m-dimensional random variable with distribution function G, by Y an independent d-dimensional random variable with density function f, and by $w_1(y), \ldots, w_m(y)$ a set of functions from $\mathbb{R}^d$ to $\mathbb{R}$. For the moment, we remove any assumptions on f, G and the w_i's; there is no loss of generality in assuming $w_i(0) = 0$, since otherwise $w_i(0)$ could be absorbed into the b_i's to be introduced in a moment. Then

$$p^{-1}\, G\{w_1(y) + b_1, \ldots, w_m(y) + b_m\}\, f(y) \tag{15}$$

is a density function for any choice of the real numbers $b_1, \ldots, b_m$, if

$$p = \mathbb{P}\{X_1 - w_1(Y) \le b_1, \ldots, X_m - w_m(Y) \le b_m\}\,.$$

The statement follows immediately from the fact that

$$\begin{aligned} p &= \mathbb{E}_Y\{\mathbb{P}\{X_1 - w_1(y) \le b_1, \ldots, X_m - w_m(y) \le b_m | Y = y\}\} \\ &= \int_{\mathbb{R}^d} G\{w_1(y) + b_1, \ldots, w_m(y) + b_m\}\, f(y)\, \mathrm{d}y. \end{aligned}$$

Clearly, the difficulty is in computing the normalising constant p. This task is amenable when X and Y are multivariate normal variables. A rather simple special case of (15) is given by (10) where G is the scalar normal distribution function, and f is $\phi_d(x; \Omega)$. An instance of density (15) with multivariate G is given by Sahu *et al.* (2001); in their case, f is the d-dimensional normal density, G is the d-dimensional normal distribution function, the w_j's are d linear combinations of y and all b_j's are 0. The multivariate distribution sketched by Azzalini (1985, section 4) and the multiple constraint model outlined by Arnold and Beaver (2000a, section 6) has a G which is the product of m ($m \ge 1$) terms of type $\Phi(\alpha_i y_i)$ or $\Phi(\alpha_i^\top y + b_i)$, respectively. The 'general multivariate skew normal distribution' mentioned by Gupta, Gonzáles-Farías and Domínguez-Molina (2001, section 5) is even more general since they adopt a G which is the the m-dimensional normal distribution function.

When f or G or both, in (15), are not of Gaussian type, evaluation of p is generally much more problematic. Some form of restrictions must however be imposed, not only to make the problem tractable but also because it has little meaning to consider (15) in its full generality which is so broad as to lose nearly any structure. A reasonable setting is as follows: suppose that f and G are both centrally symmetric and $w_i(-y) = -w_i(y)$ for all $y \in \mathbb{R}^d$. Then, by using essentially the same argument as in the proof of Proposition 1, one concludes that $W = (W_1, \dots, W_m) = (w_1(Y), \dots, w_m(Y))$ is centrally symmetric; therefore so is $V = (X_1 - W_1, \dots, X_m - W_m)$, by using the properties of centrally symmetric functions. A tractable instance of this setting is offered by the skew Cauchy distribution and its variants discussed by Arnold and Beaver (2000b), using a univariate G. Exploration of other cases along the direction sketched above seems very interesting but far beyond the scope of the present paper.

3 SKEW ELLIPTICAL DENSITIES

This section focuses on an important subclass of (12) with the component f of elliptical form, aiming at three main goals. The first is to prove that the two forms of skew elliptical densities introduced by Azzalini & Capitanio (1999, p. 599) and by Branco & Dey (2001) are closely connected. The second goal is to show that the relationships among the three forms of stochastic representation of a skew normal variate recalled in Section 1.2 carry over to skew elliptical variates. Furthermore, an analogue of stochastic representation (11) for elliptical variates is obtained for skew elliptical ones.

3.1 SKEW ELLIPTICAL DENSITIES BY CONDITIONING

For simplicity of presentation, we shall work with correlation matrices, and location parameter 0. For the rest of this section, U^* denotes a $(d+1)$-dimensional variate partitioned into a scalar component U_0 and a d-dimensional vector U.

Branco & Dey (2001) have introduced a class of skew elliptical distributions generated by applying to a $(d+1)$-dimensional elliptical variate the same conditioning method described in Section 1.2 in connection with the SN distribution. The following proposition recalls their key statement, up to some inessential changes of notation.

Proposition 3 *Consider the random vector $U^* \sim \mathrm{Ell}_{d+1}(0, \Omega^*, \tilde{f})$ where Ω^* is defined in (5). Then the probability density function of $Z = (U|U_0 > 0)$ is*

$$2 f_U(z; \bar{\Omega}) \int_{-\infty}^{\alpha^\top z} c_1 \, \tilde{f}^{Q_z}(y^2) \, \mathrm{d}y \tag{16}$$

where

$$Q_z = z^\top \bar{\Omega}^{-1} z, \tag{17}$$

the vector α is defined in (4), f_U is the density of U, $\tilde{f}^{Q_z}(\cdot)$ is the density generator of $(U_0|U = z)$ and c_1 is the associated normalizing constant.

For later use, note that an alternative expression for (16) is

$$2 \int_0^\infty c_{d+1} \tilde{f}\left(u^{*\top} (\Omega^*)^{-1} u^*\right) |\Omega^*|^{-1/2} \, \mathrm{d}u_0 \, . \tag{18}$$

On defining $F^{Q_z}(\cdot)$ to be the distribution function corresponding to the density generator $\tilde{f}^{Q_z}(\cdot)$, the above result lead Branco & Dey (2001) to re-write (16) in the form

$$2 \, f_U(z; \bar{\Omega}) \, F^{Q_z}(\alpha^\top z) \tag{19}$$

where the distribution function F^{Q_z} is actually varying at each selected point z. This expression appears to be different from (12) where a fixed distribution function F is involved.

However, when the quantity Q_z can be removed from the argument of the integral in (16) by means of a suitable change in variable, the resulting density function will become

$$2\, f_U(z; \bar{\Omega})\, F\{w(z)\} \tag{20}$$

where F is a univariate distribution function and w is such that $w(z) = h(\alpha^\top z, z^\top \bar{\Omega}^{-1} z)$ for some function h from $\mathbb{R} \times \mathbb{R}^+$ to $\mathbb{R}$. It is easy to show that the property $w(-z) = -w(z)$ must hold; hence (20) is of type (12).

It is difficult to state general conditions under which a density of type (19) can actually be transformed into one of form (20), but special cases where this is indeed feasible do exist. We shall now examine in detail two important cases of this form, namely when U^* has either a PVII_{d+1} or a PII_{d+1} distribution, which are among those considered by Branco & Dey (2001).

Proposition 4 *If the random vector U^* has a $\mathrm{PVII}_{d+1}(0, \Omega^*, M, \nu)$ distribution, then the probability density function of $Z = (U|U_0 > 0)$ is*

$$2\, f_U(z; \bar{\Omega})\, F_1\left(\alpha^\top z\, (\nu + Q_z)^{-1/2}; M, 1\right), \qquad z \in \mathbb{R}^d, \tag{21}$$

where Q_z is given by (17), f_U is the density of a $\mathrm{PVII}_d(0, \bar{\Omega}, M - 1/2, \nu)$ and $F_1(\cdot; M, 1)$ is the cumulative probability function of a $\mathrm{PVII}_1(0, 1, M, 1)$.

Proof. Using results in Fang, Kotz and Ng (1990, pp. 82–83), we have

$$c_1 \tilde{f}^{Q_z}(y^2) = \frac{\Gamma(M)}{\pi^{1/2}\Gamma(M - 1/2)}\, (\nu + Q_z)^{-1/2} \left(1 + \frac{y^2}{\nu + Q_z}\right)^{-M}$$

and

$$f_U(z; \bar{\Omega}) = \frac{\Gamma(M - 1/2)}{|\bar{\Omega}|^{1/2} (\pi\nu)^{d/2} \Gamma(M - (d+1)/2)} \left(1 + \frac{Q_z}{\nu}\right)^{-M+1/2}$$

i.e. the densities of a $\mathrm{PVII}_1(0, 1, M, \nu{+}Q_z)$ and of a $\mathrm{PVII}_d(0, \bar{\Omega}, M{-}1/2, \nu)$ variate with parameters $M - 1/2$ and ν, respectively. On setting $x = y\, (\nu + Q_z)^{-1/2}$, the integral in (16) becomes

$$\int_{-\infty}^{\alpha^\top z (\nu + Q_z)^{-1/2}} \frac{\Gamma(M)}{\pi^{1/2}\Gamma(M - 1/2)} (1 + x^2)^{-M}\, \mathrm{d}x$$

which is the distribution function of a $\mathrm{PVII}_1(0, 1, M, 1)$ variate evaluated at the point $\alpha^\top z\, (\nu + Q_z)^{-1/2}$. QED

Example 1: skew t distribution. The relevance of the PVII_d class is due to the inclusion of the multivariate t family as the special case when $M = (d + \nu)/2$. The corresponding specification of Proposition 4 produces then a form of multivariate skew t density. Since Section 4 will be entirely dedicated to this distribution, we defer detailed discussion until then.

Proposition 5 *If the $(d + 1)$-dimensional elliptical random vector U^* has a $\mathrm{PII}_{d+1}(0, \Omega^*, \nu)$ distribution, then the probability density function of $Z = (U|U_0 > 0)$ is*

$$2 f_U(z; \bar{\Omega})\, F_1\left(\alpha^\top z\ (1 - Q_z)^{-1/2}; \nu\right), \qquad z \in (-1, 1)^d, \tag{22}$$

where Q_z is given by (17), f_U is the density of a $\mathrm{PII}_d(0,\bar\Omega,\nu+1/2)$ *variate, and* $F_1(\cdot;\nu)$ *is the distribution function of a* $\mathrm{PII}_1(0,1,\nu)$.

Proof. Identical to that of Proposition 4, considering the densities of marginal and conditional distributions of PII, as defined in Fang, Kotz and Ng, (1990, pp. 89-91).

The absence of Q_z in the conditional density characterizes the multivariate normal distribution among the members of the elliptical family. This fact can be used to produce an analogous characterization of the skew normal distribution within the skew elliptical family.

Proposition 6 *The function w in (20) is such that $w(z)=\alpha^\top z$ if and only if U^* is Gaussian, i.e. Z is skew normal.*

Proof. The density of $(U|U_0=z)$ does not depend on Q_z if and only if U^* is Gaussian; see Theorem 4.12 of Fang *et al.* (1990). In this case, the integral in (16) becomes $\Phi(\alpha^\top z)$, so that $Z\sim\mathrm{SN}_d(0,\bar\Omega,\alpha)$. QED

A number of parallels between the skew normal distribution and other types of skew elliptical distributions have already been shown. The next result allows us to construct a random variable $\tilde X$ playing a role analogous to the one in (14) for the skew version of a PVII_d and PII_d distribution, respectively.

Proposition 7 *Let* $U^*\sim\mathrm{PVII}_{d+1}(0,\Omega^*,M,\nu)$. *Then*

$$\tilde X=-\left(1-\delta\bar\Omega^{-1}\delta\right)^{-1/2}(U_0-\delta^\top\bar\Omega^{-1}U)\left(\nu+U^\top\bar\Omega^{-1}U\right)^{-1/2}\sim\mathrm{PVII}_1(0,1,M,1),$$

independent of U. If $U^*\sim\mathrm{PII}_{d+1}(0,\Omega^*,\nu)$ *then*

$$\tilde X=-\left(1-\delta\bar\Omega^{-1}\delta\right)^{-1/2}(U_0-\delta^\top\bar\Omega^{-1}U)\left(1-U^\top\bar\Omega^{-1}U\right)^{-1/2}\sim\mathrm{PII}_1(0,1,\nu),$$

independent of U.

Proof. By direct calculation.

Therefore, we can set

$$Z=\begin{cases}U & \text{if } \tilde X<w(U),\\ -U & \text{if } \tilde X>w(U),\end{cases}$$

where $w(z)$ is the transformation of z used in the argument of F_1 in (21) and (22), respectively; it is intended that the appropriate distribution of U^* and transformation $\tilde X$ have been selected. This formula establishes a method of type (13) to generate a skew PVII_d and skew PII_d variate, respectively.

The connections between the proposal of Azzalini & Capitanio (1999) and the one of Branco & Dey (2001) can be summarised as follows. The conditioning argument which is one of the mechanisms to generate the skew normal distribution from the normal one can be adopted to generate a form of skew elliptical distributions from the elliptical ones, leading to (19), or some similar form as obtained by Branco & Dey. This type of expression can, at least in some important special cases, be transformed into one where the skewing factor of f is a fixed distribution function, as shown by (21) and (22). These expressions are of type (12), which is essentially the form of Azzalini & Capitanio. The natural question is whether all densities of type (19) can be re-written in the form (12), but we have been unable to prove this fact in general. Notice that the converse inclusion is not true, that is, not all densities of type (12) can be written in the form (19), unless additional restrictions are imposed on the components of (12), besides the obvious condition that f is elliptical.

The next result concerns a stochastic representation of type (11) for distributions of type (12) when the density f is elliptical. For example, this representation is valid for the skew elliptical densities defined in Azzalini & Capitanio (1999, p. 599) and for the skew versions of PVII_d and PII_d examined earlier.

Proposition 8 *If Z has a density of type (12), where f is the density of $U \sim \mathrm{Ell}_d(\xi, \bar{\Omega}, \tilde{f})$, then Z admits the stochastic representation*

$$Z = \xi + RL^\top S' \tag{23}$$

where $\bar{\Omega} = L^\top L$, $R > 0$ has the same distribution as the radius of the stochastic representation (11) of U, and S' has a non-uniform distribution on the unit sphere of $\mathbb{R}^d$. Specifically, using spherical coordinates, the density of S' is equal to

$$\frac{\Gamma(d/2)}{\pi^{d/2}} \prod_{k=1}^{d-2} (\sin\theta_k)^{d-k-1} \mathbb{P}\{X \le w_L^*(\theta_1, \ldots, \theta_{d-1}, R)\},$$

where $w_L^*(\cdot)$ is a function from $\mathbb{R}^d$ to $\mathbb{R}$ defined in Appendix A, and X is an independent random variable having distribution function G. Furthermore, the conditional distribution of S' given $R = r$ is of type (12), with density

$$\frac{\Gamma(d/2)}{\pi^{d/2}} \prod_{k=1}^{d-2} (\sin\theta_k)^{d-k-1} G\{w_L^*(\theta_1, \ldots, \theta_{d-1}, r)\}.$$

Proof. In Appendix A.

Example 2: Stochastic representation (23) for skew normal distribution. If $Z \sim \mathrm{SN}_d(\xi, \bar{\Omega}, \alpha)$, then by applying Proposition 8 we obtain $R^2 \sim \chi^2_d$ and the following spherical coordinates representation of the marginal distribution of S':

$$\begin{aligned} f_\theta(\theta) &= 2\,\frac{\Gamma(d/2)}{2\pi^{d/2}} \prod_{k=1}^{d-2} (\sin\theta_k)^{d-k-1} \\ &\qquad \mathbb{P}\{X \le R\,(\alpha_1^* \cos\theta_1 + \alpha_2^* \sin\theta_1 \cos\theta_2 + \cdots + \alpha_d^* \sin\theta_1 \cdots \sin\theta_{d-1})\}, \end{aligned}$$

where $\theta = (\theta_1, \theta_2, \ldots, \theta_{d-1})^\top$, $\alpha^* = L\,\alpha$ and $X \sim \mathrm{N}(0,1)$ is independent of R. Finally, noticing that $d^{1/2} X\, R^{-1}$ has a t distribution with d degrees of freedom, we have

$$\begin{aligned} f_\theta(\theta) &= 2\,\frac{\Gamma(d/2)}{2\pi^{d/2}} \prod_{k=1}^{d-2} (\sin\theta_k)^{d-k-1} \\ &\qquad T_1\{d^{1/2}(\alpha_1^* \cos\theta_1 + \alpha_2^* \sin\theta_1 \cos\theta_2 + \ldots + \alpha_d^* \sin\theta_1 \cdots \sin\theta_{d-1}); d\} \end{aligned}$$

where $T_1(\cdot; d)$ is the distribution function of a scalar t distribution with d degrees of freedom.

3.2 Skew elliptical densities by transformation method

The next result shows how the class of skew elliptical distributions mirrors another property of the skew normal distribution. In fact the class of skew elliptical densities obtained via the conditioning method is equivalent to the one obtained by applying the transformation method recalled in Section 1.2.

Proposition 9 *Consider the random vector* $(U_0, U) \sim \mathrm{Ell}_{d+1}(0, \Psi^*, \tilde{f})$ *where* Ψ^* *is as in (6), and define*

$$Z_j = \delta_j\,|U_0| + \left(1 - \delta_j^2\right)^{1/2}\,U_j, \qquad j = 1, \ldots, d, \tag{24}$$

where $-1 < \delta_j < 1$. *Then the density of* $(Z_1, \ldots, Z_d)$ *is of type (16), where*

$$\begin{aligned}
\lambda_i &= \delta_i\left(1-\delta_i^2\right)^{-1/2}, \qquad (i = 1, \ldots, d),\\
\Delta &= \mathrm{diag}\{\left(1+\lambda_1^2\right)^{-1/2}, \ldots, \left(1+\lambda_d^2\right)^{-1/2}\},\\
\Omega &= \Delta(\Psi + \lambda\lambda^\top)\Delta,\\
\alpha &= \left(1 + \lambda^\top\Psi\lambda\right)^{-1/2}\Delta^{-1}\Psi^{-1}\lambda\,.
\end{aligned}$$

Proof. First note that the joint density function of $|U_0|$ and U takes the form $2c_{d+1}\tilde{f}(\cdot)$. Denote by B the $(d+1)\times(d+1)$ matrix implicitly defined by (24) such that $(Z_0, Z_1, \ldots, Z_d)^\top = B(|U_0|, U^\top)^\top$, and apply the usual formulae for linear transforms. Then the density function of $(Z_1, \ldots, Z_d)$ turns out to be

$$2\int_0^\infty c_{d+1}\tilde{f}\left((z_0, z^\top)A^{-1}(z_0, z^\top)^\top\right)\,|A|^{-1/2}\,\mathrm{d}x_0$$

where $A = B\Psi^* B^\top$ is a correlation matrix. Taking into account expression (18) the result follows. QED

An immediate consequence of the transformation method is a further generating method for the bivariate case. Again, this reproduces for the skew elliptical family a generation method known to hold for the skew normal distributions.

Proposition 10 *If* $(U_0, U) \sim \mathrm{Ell}_2(0, \Omega^*, \tilde{f})$, *the class generated by* $Z = \max(U_0, U)$ *is equal to the class generated by the transformation method of Proposition 9 with* $d = 2$.

Proof. First notice that $\max(U_0, U) = \frac{1}{2}|U - U_0| + \frac{1}{2}(U + U_0)$. As the joint distribution of $(U - U_0)\,(2 - 2\rho)^{-1/2}$ and $(U + U_0)\,(2 + 2\rho)^{-1/2}$ is $\mathrm{Ell}_2(0, I, \tilde{f})$, where ρ denotes the off-diagonal elements of Ω^*, the result follows by direct application of Proposition 9 on imposing $\delta = \left(\frac{1}{2}(1-\rho)\right)^{1/2}$. QED

4 A skew t distribution

For the rest of the paper we shall focus on the development of an asymmetric version of the multivariate Student's t distribution, already sketched in Section 3.1. The purpose of the present section is to provide additional support for its definition and to examine more closely its properties. Connected inferential aspects will be discussed in the subsequent section.

4.1 DEFINITION AND DENSITY

The usual construction of the t distribution is via the ratio of a normal variate and an appropriate transformation of a chi-square. If one wants to introduce an asymmetric variant of the t distribution, a quite natural option is to replace the normal variate above by a skew normal one.

A preliminary result on Gamma variates is required. We shall say that a positive random variable is distributed as $\mathrm{Gamma}(\psi, \lambda)$ if its density at x $(x > 0)$ is

$$\frac{\lambda^{\psi}}{\Gamma(\psi)}\, x^{\psi-1}\, \exp(-\lambda\, x).$$

Lemma 11 *If* $V \sim \mathrm{Gamma}(\psi, \lambda)$*, then for any* $a,\ b \in \mathbb{R}$

$$\mathbb{E}\Big\{\,\Phi(a\sqrt{V} + b)\Big\} = \mathbb{P}\Big\{T \le a\sqrt{\psi/\lambda}\Big\}$$

where T *denotes a non-central* t *variate with* 2ψ *degrees of freedom and non-centrality parameter* $-b$*.*

Proof. Let $U \sim N(0,1)$; then

$$\begin{aligned}
\mathbb{E}\Big\{\,\Phi(a\sqrt{V} + b)\Big\} &= \mathbb{E}_V\left\{\mathbb{P}\{U \le a\sqrt{v} + b | V = v\}\right\} \\
&= \mathbb{E}_V\left\{\mathbb{P}\Big\{(U-b)/\,(v\lambda/\psi)^{1/2} \le a\,(\psi/\lambda)^{1/2}\,|V = v\Big\}\right\} \\
&= \mathbb{P}\Big\{T' \le a\,(\psi/\lambda)^{1/2}\Big\}
\end{aligned}$$

where $T = (U-b)/\,(V\lambda/\psi)^{1/2}$ has the quoted t distribution. QED

As anticipated earlier, we define the skew t distribution as the one corresponding to the transformation

$$Y = \xi + V^{-1/2}\, Z \tag{25}$$

where Z has density function (2) with $\xi = 0$, and $V \sim \chi^2_{\nu}/\nu$, independent of Z. An equivalent interpretation of Y is to regard it as a scale mixture of SN variates, with mixing scale factor $V^{-1/2}$. Application of the above lemma to a $\mathrm{Gamma}(\frac{1}{2}\nu, \frac{1}{2}\nu)$ variate and some simple algebra lead to the density of Y, which is

$$f_Y(y) = 2\, t_d(y;\nu)\, T_1\left(\alpha^{\top}\omega^{-1}(y-\xi)\left(\frac{\nu+d}{Q_y+\nu}\right)^{1/2};\nu+d\right) \tag{26}$$

where ω is defined at the beginning of Section 1.2,

$$\begin{aligned}
Q_y &= (y-\xi)^{\top}\Omega^{-1}(y-\xi)\,, \\
t_d(y;\nu) &= \frac{1}{|\Omega|^{1/2}}\, g_d(Q_y;\nu) \;=\; \frac{\Gamma((\nu+d)/2)}{|\Omega|^{1/2}\,(\pi\nu)^{d/2}\,\Gamma(\nu/2)}(1+Q_y/\nu)^{-(\nu+d)/2}
\end{aligned}$$

is the density function of a d-dimensional t variate with ν degrees of freedom, and $T_1(x;\nu{+}d)$ denotes the scalar t distribution function with $\nu+d$ degrees of freedom. We shall call distribution (26) skew t, and write

$$Y \sim \mathrm{St}_d(\xi, \Omega, \alpha, \nu)\,. \tag{27}$$

It is easy to check that density (26) coincides with the one sketched in Section 3.1 using Proposition 4, which is of type (12). Moreover, for the reasons explained in that section, (26) coincides in turn with the skew t distribution of Branco & Dey (2001), although this equality is not visible from their derivation because they did not provide the above closed-form expression of the density.

Therefore, we have seen that a number of different ways to define a skew t distribution all lead to the same density (26). While additional proposals to introduce a form of a skew t density are possible, this one has the advantage of arising from various generating criteria, which in turn are linked to other portions of literature.

A reviewer of this paper has remarked that, if we set $d = 1$, density (26) does not reduce to the form $2\,t_1(y;\nu)\,T_1(\alpha y;\nu)$, which seems to be the 'most natural' univariate form of skew t density generated by Lemma 1 of Azzalini (1985), a forerunner of Proposition 1. While the latter density has the appeal of a slightly simpler mathematical expression, the arguments indicated in the previous paragraph lead us to prefer (26). In fact, one could reverse the reasoning, and claim that Lemma 1 of Azzalini (1985) 'should' had been stated in the form of Proposition 1 for $d = 1$; in other words, there is no reason to restrict $w(y)$ to the linear form αy, especially outside the normal case.

Alternative proposals of univariate skew t distributions have been made by Fernández & Steel (1998), constructed similarly to the so-called two-piece normal density, and by Jones (2001), developed by Jones & Faddy (2002), which is based on a suitable transformation of a beta density. A multivariate form of skew t distribution has been proposed by Jones (2002) but the associated inferential aspects have not been discussed. The alternative form of multivariate skew t distribution considered by Sahu *et al.* (2001) concides with (26) in the case $d = 1$; for general d, their density involves the multivariate t distribution function. The density examined in this paper allows a relatively simple mathematical treatment, and it is more naturally linked to the skew normal distribution, via mechanisms already mentioned. As a consequence, the distribution enjoys various useful formal properties, which will be examined in the remaining part of this section.

4.2 Some properties

Distribution function For simplicity of exposition, we obtain the distribution function of Y in the 'standard' case with $\xi = 0, \Omega = \bar{\Omega}$. Bearing in mind the representation of Z based on conditioning, write

$$\begin{aligned}
\mathbb{P}\{Y \le y\} &= \mathbb{P}\Big\{V^{-1/2}\,Z \le y\Big\} \\
&= \mathbb{P}\Big\{V^{-1/2}\,U \le y \mid U_0 > 0\Big\} \\
&= 2\,\mathbb{P}\left\{V^{-1/2}\begin{pmatrix}-U_0\\ U\end{pmatrix} \le \begin{pmatrix}0\\ y\end{pmatrix}\right\} \\
&= 2\,\mathbb{P}\left\{T' \le \begin{pmatrix}0\\ y\end{pmatrix}\right\}
\end{aligned}$$

where (U_0, U) has distribution (5), and the inequality signs are intended componentwise. The last expression involves the integral of a multivariate $(d+1)$-dimensional t variate T' with dispersion matrix similar to the one of (5), but with reversed sign of δ. Algorithms for computing this type of distribution function are given by Genz & Bretz (1999).

An alternative expression for the above distribution function is given by

$$\mathbb{P}\{Y \le y\} = \mathbb{P}\Big\{V^{-1/2}\, U \le y \mid U_0 > 0\Big\} = \mathbb{E}_V\left\{F_Z(yv^{1/2})|V = v\right\},$$

where F_Z denotes the distribution of Z, hence evaluating the distribution function of Y by suitably averaging the distribution of Z with respect to the distribution of V. This expression is most useful in the case $d = 1$ where a practical expression of F_Z is available; see formula (4) and subsequent remarks of Azzalini (1985).

Moments Using the representation (25), it is easy to compute the moments of Y. For algebraic convenience, we assume $\xi = 0$ throughout. If $\mathbb{E}\Big\{Y^{(m)}\Big\}$ denotes a moment of order m, write

$$\mathbb{E}\Big\{Y^{(m)}\Big\} = \mathbb{E}\Big\{V^{-m/2}\Big\}\ \mathbb{E}\Big\{Z^{(m)}\Big\} \tag{28}$$

where Z has density function (2) with $\xi = 0$. It is well-known that

$$\mathbb{E}\Big\{V^{-m/2}\Big\} = \frac{(\nu/2)^{m/2}\,\Gamma(\frac{1}{2}(\nu - m))}{\Gamma(\frac{1}{2}\nu)},$$

while, for the expressions of $\mathbb{E}\Big\{Z^{(m)}\Big\}$, we use results given by Azzalini & Capitanio (1999) and by Genton *et al.* (2001).

First, we apply (28) to the scalar case. On defining

$$\mu = \delta\ (\nu/\pi)^{1/2}\ \frac{\Gamma(\frac{1}{2}(\nu-1))}{\Gamma(\frac{1}{2}\nu)}, \qquad (\nu > 1), \tag{29}$$

one obtains, for $\xi = 0$,

$$\begin{aligned}
\mathbb{E}\{Y\} &= \omega\,\mu,\\
\mathbb{E}\{Y^2\} &= \omega^2\,\frac{\nu}{\nu-2},\\
\mathbb{E}\{Y^3\} &= \omega^3\,\mu\,(3-\delta^2)\,\frac{\nu}{\nu-3},\\
\mathbb{E}\{Y^4\} &= \omega^4\,\frac{3\,\nu^2}{(\nu-2)(\nu-4)},
\end{aligned}$$

provided that ν is larger than the corresponding order of the moment; the first two of the above expressions have been given by Branco & Dey (2001). After some algebra, the indices of skewness and kurtosis turn out to be

$$\begin{aligned}
\gamma_1 &= \mu\,\left[\frac{\nu(3-\delta^2)}{\nu-3} - \frac{3\,\nu}{\nu-2} + 2\,\mu^2\right]\left[\frac{\nu}{\nu-2} - \mu^2\right]^{-3/2} \qquad (\text{if } \nu > 3),\\
\gamma_2 &= \left[\frac{3\nu^2}{(\nu-2)(\nu-4)} - \frac{4\mu^2\nu(3-\delta^2)}{\nu-3} + \frac{6\mu^2\nu}{\nu-2} - 3\mu^4\right]\left[\frac{\nu}{\nu-2} - \mu^2\right]^{-2} - 3 \qquad (\text{if } \nu > 4).
\end{aligned}$$

In the multivariate case, we obtain from (28) that $\mathbb{E}\{Y\} = \omega\mu$ still holds, provided $\nu > 1$ and (29) and ω are intended in vector and matrix form, respectively; furthermore

$$\mathbb{E}\Big\{Y\,Y^\top\Big\} = \frac{\nu}{\nu-2}\,\Omega \qquad (\text{if } \nu > 2)\,,$$

leading to

$$\mathrm{var}\{Y\} = \frac{\nu}{\nu-2}\,\Omega - \omega\mu\mu^\top\omega\,.$$

Linear and quadratic forms Consider the affine transformation $a + AY$ where $a \in \mathbb{R}^m$ and A is a $m \times d$ constant matrix of rank m. Using (25) we can write

$$a + AY = \xi' + V^{-1/2}AZ$$

where $\xi' = a + A\xi$. Take into account that

$$AZ \sim SN_m(0, A\Omega A^\top, \alpha')$$

on the ground of results given by Azzalini & Capitanio (1999) where the explicit expression for α' is given; similar results, but in a more convenient form, are provided by Capitanio *et al.* (2003, Appendix A.2). Therefore we obtain

$$a + AY \sim \mathrm{St}_m(\xi', A\Omega A^\top, \alpha', \nu).$$

In particular for a single component, Y_r say ($r \in \{1, \ldots, d\}$), one has

$$Y_r \sim \ \mathrm{St}(\xi_r, \omega_{rr}, \alpha'_r, \nu)$$

where α'_r is given by (10) of Capitanio *et al.* (2003).

Similarly, for a quadratic form, $Q = (Y-\xi)^\top B(Y-\xi)$, where B is a symmetric $d \times d$ matrix, we can write

$$Q = Z^\top BZ/V.$$

For appropriate choices of B, the distribution of $Z^\top BZ$ is $\chi^2_{\nu'}$ for some value ν' of the degrees of freedom. One such case is (9), where $B = \Omega^{-1}$. Azzalini & Capitanio (1999, Section 3.3) consider more general forms of B; see also Genton *et al.* (2001) for additional results. In all cases when the χ^2 property holds for Z, we can state immediately

$$Q/\nu' \sim F(\nu', \nu).$$

This property allows us to produce Healy's-type plots (Healy, 1968) as a diagnostic tool in data fitting, similarly to the Normal and SN case, just using the Snedecor distribution as the reference distribution instead of the χ^2. This device will be illustrated in the subsequent numerical work.

An extended skew t distribution If the component Z in (25) is taken to have distribution (10) rather than (2), we obtain a density which parallels the role of (10) for skew t densities; this is now discussed briefly.

By using again Lemma 11, the new density turns out to be of type (26), except that T_1 refers now to a t distribution with non-centrality parameter $-\tau(1-\delta^\top\bar\Omega^{-1}\delta)^{-1/2}$ and the normalizing constant 2 is replaced by $1/\Phi(\tau)$. The distribution function is obtained with the same sort of argument of the case $\tau = 0$, namely

$$\begin{aligned}\mathbb{P}\{Y \le y\} &= \mathbb{P}\Big\{V^{-1/2}\,U \le y \,|\, U_0 + \tau > 0\Big\}\\ &= \mathbb{P}\Big\{V^{-1/2}\begin{pmatrix}-U_0-\tau\\ U\end{pmatrix} \le \begin{pmatrix}0\\ y\end{pmatrix}\Big\}\Big/\Phi(\tau)\\ &= \mathbb{P}\Big\{T'' \le \begin{pmatrix}0\\ y\end{pmatrix}\Big\}\Big/\Phi(\tau)\end{aligned}$$

where now T'' refers to a non-central multivariate t; unfortunately, the latter distribution function is appreciably harder to compute in practice than the analogous one for the central case. Moments can be computed again with the aid of (28). Those of the first and second order are, if $\xi = 0$,

$$\begin{aligned}
\mathbb{E}\{Y\} &= \mathbb{E}\Big\{V^{-1/2}\Big\}\,\zeta_1(\tau)\omega\delta, \qquad (\nu > 1),\\
\mathbb{E}\Big\{YY^\top\Big\} &= \frac{\nu}{\nu-2}\,\Big(\Omega + [\zeta_2(\tau)+\zeta_1^2(\tau)]\,\omega\,\delta\,(\omega\delta)^\top\Big), \qquad (\nu > 2),
\end{aligned}$$

where

$$\zeta_r(x) = \frac{\mathrm{d}^r}{\mathrm{d}x^r}\zeta_0(x), \qquad (r = 1, 2, \ldots),$$

and ζ_0 is defined by (3).

5 Statistical aspects of the skew t distribution

5.1 Likelihood inference

Consider n independent observations satisfying a regression model of type

$$y_i \sim \mathrm{St}_d(\xi_i, \Omega, \alpha, \nu), \quad \xi_i = \beta^\top x_i$$

for $i = 1, \ldots, n$; here x_i is a $p-$dimensional vector and β is a $p \times d$ matrix of parameters. Also let

$$X = (x_1, x_2, \ldots, x_n)^\top$$

be the $n \times p$ design matrix. Notice that we are effectively considering a multivariate regression model with error term of skew t type. It would be inappropriate to use such a distribution, and in fact even a regular elliptical distribution, for the joint modelling of the n observations, since usually these are supposed to behave independently.

It is convenient to reparametrize the problem by writing

$$\Omega^{-1} = A^\top \mathrm{diag}(e^{-2\rho})A = A^\top D A, \quad \text{and} \quad \eta = \omega^{-1}\alpha$$

where A is an upper triangular $d \times d$ matrix with diagonal terms equal to 1 and $\rho \in \mathbb{R}^d$. The loglikelihood function for the parameter $\theta = (\beta, A, \rho, \eta, \log\nu)$ is then

$$\ell(\theta) = \sum_{i=1}^{n} \ell_i(\theta) \qquad (30)$$

where $\ell_i(\theta)$ is the contribution to the loglikelihood from the i-th individual; this term is

$$\ell_i(\theta) = \log 2 + \tfrac{1}{2}\log|D| + \log g_d(Q_i;\nu) + \log T_1(t(L_i, Q_i, \nu); \nu + d)$$

where

$$u_i = y_i - \beta^\top x_i, \qquad Q_i = u_i^\top \Omega^{-1} u_i, \qquad L_i = \alpha^\top \omega^{-1} u_i, \qquad t(L, Q, \nu) = L\left(\frac{\nu+d}{Q+\nu}\right)^{1/2}.$$

Maximisation of this log-likelihood function must be accomplished numerically. To improve efficiency, the derivatives of (30) can be supplied to an optimisation algorithm; details for computing these derivatives are given in an appendix.

A suite of R routines for evaluating the above log-likelihood and its derivatives has been developed, and it is available on the WWW at `http://azzalini.stat.unipd.it/SN`.

In connection with the skew normal distribution, Azzalini (1985) and Azzalini & Capitanio (1999) have highlighted some problematic aspects of the likelihood function. A key feature is that the profile log-likelihood function for α always has a stationarity point at $\alpha = 0$, which in turn is connected to singularity of the information matrix at $\alpha = 0$. These problematic features were the motivation to introduce an alternative parametrization which overcomes most if not all of these problems.

It was a pleasant surprise to find that in the present setting the behaviour of the log-likelihood function was to be much more regular, at least for those numerical cases which we have explored. A graphical illustration of this statement is given by Figures 5 and 8 below, which show some profile log-likelihood plots. These plots refer to specific datasets, but a similar regularity was found with some other datasets which we have considered.

It would be useful to have some theoretical insight on why the log-likelihood function using the skew t distribution behaves so differently from the skew normal model, as well as to gather more numerical evidence of its behaviour. However this theme appears to be a project on its own, and cannot be pursued here.

On another front, Fernández & Steel (1999) have highlighted difficulties in regression models when the error term is assumed to have a t distribution with unspecified degrees of freedom to be estimated from the data. Specifically, their Theorem 5 states there are points of the parameter space where the likelihood function becomes unbounded, if the degrees of freedom are allowed to span over the whole range $\nu \in (0, \infty)$. To avoid this effect, one must restrict the range of ν to the interval (ν_0, ∞), where the threshold ν_0 is a function of X and y. For instance, in the case of a simple random sample with no ties in the y_i's, we obtain $\nu_0 = d/(n-1)$, which imposes a very mild limitation. For the stackloss data example discussed by Fernández & Steel (1999) with $d = 1$ and $p = 3$, the value of ν_0 is small, 8/13. In addition, they recall some numerical examples from the literature where poles have been found by various authors; in all these cases, however, these poles where found at values of ν very small, always below 0.30.

Therefore, in practice the difficulties can be circumvented by avoiding a certain portion of the parameter space which would be somewhat peculiar anyway. However, the fact that ν_0 depends on the response variable leads to a procedure which lacks complete support by the theory of likelihood inference. As advocated by Fernández & Steel, a better theoretical understanding of this sort of model and the associated log-likelihood properties is therefore called for.

It is plausible that regression models with skew t error terms behave quite similarly to analogous cases which employ a regular t distribution, as for the phenomenon discusses by Fernández & Steel (1999). In the numerical work of the next subsection, we have been driven by considerations described above, and decided to ignore poles of the log-likelihood very near $\nu = 0$. We have however searched for them, but the only case where we have successfully located one was with the stackloss data, near $\nu = 0.06$, while the maximum above the threshold $\nu_0 = 8/13$ was at $\hat{\nu} = 1.14$.

5.2 Numerical examples

AIS data It is instructive to examine the outcome of a data fitting process based on the skew t distribution in a few practical cases. Data on several biomedical variables from 202 athletes have been

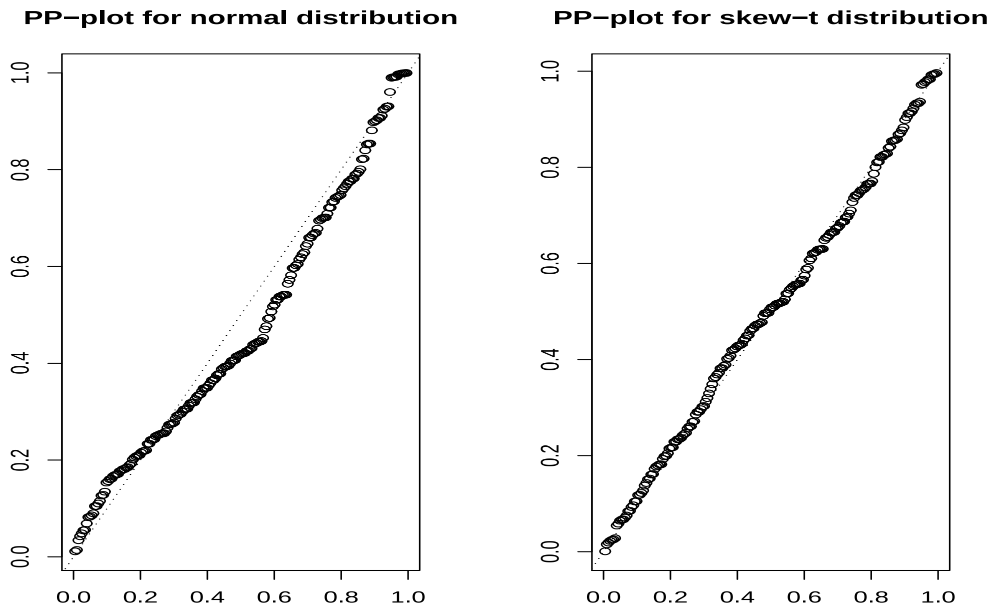

Figure 2: *AIS data: Healy's plot when either a normal distribution (left-hand side panel) or a skew t distribution (right-hand side panel) is fitted to the data*

collected at the Australian Institute of Sport; see Cook & Weisberg (1994) for their description.

We consider here four variables, $(BMI, Bfat, ssf, LBM)$, which represent represent the body mass index, the percentage of body fat, the sum of skin folds and the lean body mass, respectively. A St_4 distribution has been fitted to the 202 points, and Figure 2 shows the associated Healy's plot, using the multivariate normal and the skew t distribution, as described at the end of Section 4.2. The plots indicate a satisfactory fit to the data provided by the skew t, markedly superior to the normal one.

This figure matches with Figure 6 of Azzalini & Capitanio (1999), who fit a SN distribution to the same data. While the SN fit was definitely superior to the normal one, still there was some discrepancy from the identity line which has now vanished almost perfectly.

The full list of estimated parameters is not of particular interest, but it is noteworthy that $\hat{\nu} = 13.7$, which confirms the presence of somewhat longer tails than the normal distribution.

We do not present the analogue of Figure 5 of Azzalini & Capitanio (1999) because its graphical appearance in our case is not so markedly different from their Figure 5. These differences exist, but they become graphically evident only in a summary plot like the one reported.

Strength of fiber-glass Smith & Naylor (1987) have reported values concerning the breaking strengths of 1.5 cm long glass fibers. These data have also been considered by Jones & Faddy (2002) in association with another form of skew t distribution, and comparison with their results is the reason for including this example here.

Figure 3 shows a histogram of the data and skew t densities fitted using (26) and the Jones'

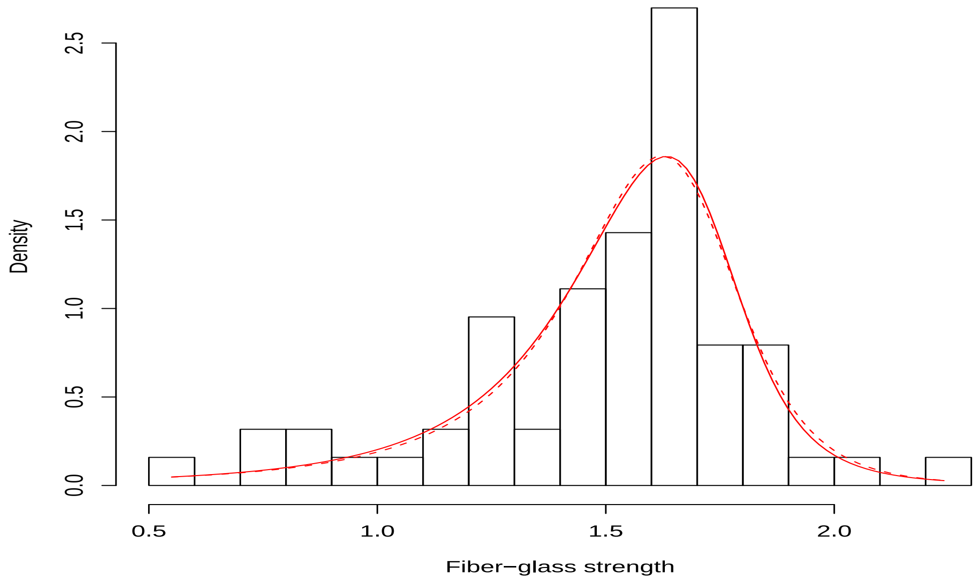

Figure 3: *Fiber–glass data: histogram and fitted skew t densities; the continuous curve refers to the density studied in this paper, the dashed curve refers to Jones' model*

distribution. The two parametric densities are graphically very close, and choice between the two distributions has to be based on other aspects, rather than empirical adequacy. The Healy plot associated to (26), in Figure 4, confirms a satisfactory fit of the parametric distribution to the data.

Other interesting features are indicated by twice the profile log-likelihood functions for the parameters α, $\log \nu$, $(\log \omega, \alpha)$ and $(\alpha, \log \nu)$ reported in panel (a) to (d) of Figure 5, respectively. The contour lines for the two parameter cases are chosen to correspond to differences from the maximum equal to the quantiles of level 0.50, 0.75, 0.90, 0.95, 0.99 of the χ^2_2 distribution; hence each contoured region can be interpreted as a confidence region for the pair of parameters, at the quoted confidence level. As anticipated earlier, these plots have a quite regular behaviour, not very far from quadratic functions.

This figure also indicates quite clearly a significant negative skewness of the distribution, since the confidence regions up to level 95% are entirely on the left of $\alpha = 0$. This conclusion is confirmed by the value of $\hat{\alpha}$ divided by its standard error, which is $-1.55/0.574 \approx -2.70$, with corresponding p-value about 0.7%. There is also an indication of a long tail of the distribution, since $\hat{\nu} = 2.73$, but rather higher values of ν are not ruled out. These conclusions are broadly similar to those of Jones & Faddy (2001); from our analysis there appears to be a slightly stronger indication of significant skewness.

Martin Marietta data Our next example considers data taken from Table 1 of Butler, McDonald, Nelson and White (1990). Based on the arguments presented in that paper, a linear regression is introduced

$$y = \beta_0 + \beta_1 \mathrm{CRSP} + \varepsilon$$

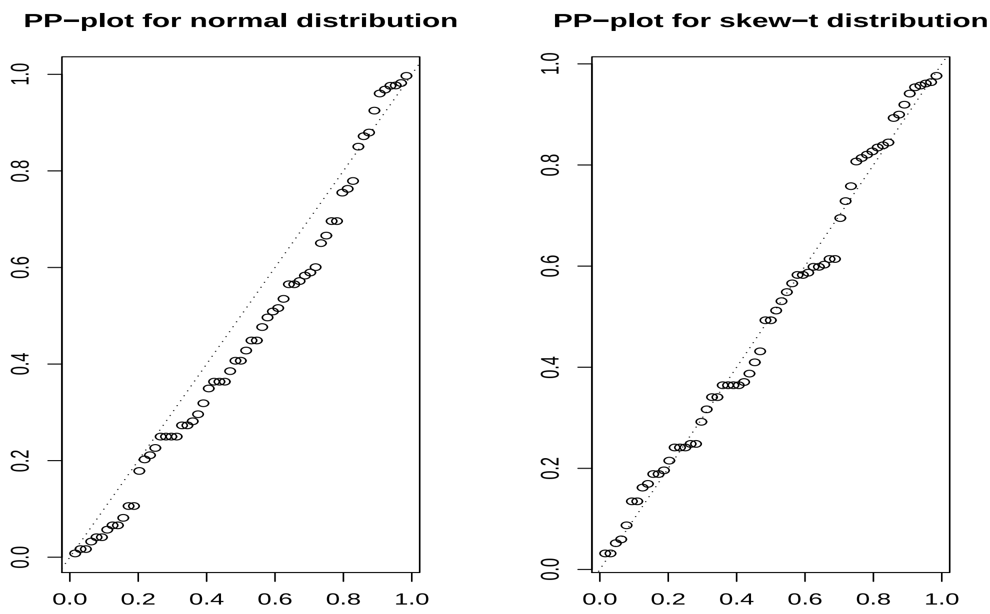

Figure 4: *Fiber–glass data: Healy's plot when either a normal distribution (left panel) or a skew t distribution (right panel) is fitted to the glass data*

where y is the excess rate of the Martin Marietta company, CRSP is an index of the excess rate of return for the New York market as a whole and ε is an error term which in our case is taken to be distributed as $\mathrm{St}(0, \omega^2, \alpha)$. Data over a period of $n = 60$ consecutive months are available.

The resulting fitted line is shown in Figure 6, which displays the scatter-plot of the data with superimposed the least squares lines and the line obtained from the above model after adjusting for $\mathbb{E}\{\varepsilon\}$, whose intercept and slope are

$$\hat{\beta}_0 + \hat{\mathbb{E}}\{\varepsilon\} = 0.0029, \qquad \hat{\beta}_1 = 1.248$$

respectively. These values are very close to those obtained using the skew t distribution of Jones (2001), and the addition of that line to Figure 6 would be barely visible, being essentially coincident with our line. The estimated skewness parameter is $\hat{\alpha} \approx 1.246$ with standardised value $1.246/0.653 \approx 1.908$ and observed significance 5.6%. The estimated degrees of freedom are $\hat{\nu} = 3.32$ $(\mathrm{s.e.} 1.43)$.

As further indication of the agreement between observed data and fitted distributions, Figure 7 shows the histogram of the residuals after removing the line $\hat{\beta}_0 + \hat{\beta}_1 \mathrm{CRSP}$, and the fitted skew t density; there appears to be a satisfactory agreement between the two. Similarly to Figure 5, the shape of the log-likelihood function displayed a nice regular behaviour, as indicated by Figure 8. Finally, Figure 9 compares the Healy's plots for the normal and a skew t fitted models. Expectedly the normal model shows obvious inadequacy, while the skew t model behaves satisfactorily.

6 DISCUSSION

A number of broadly related proposals and results have appeared in the recent literature under the connecting concept of the multivariate skew normal distribution. The present paper has examined the

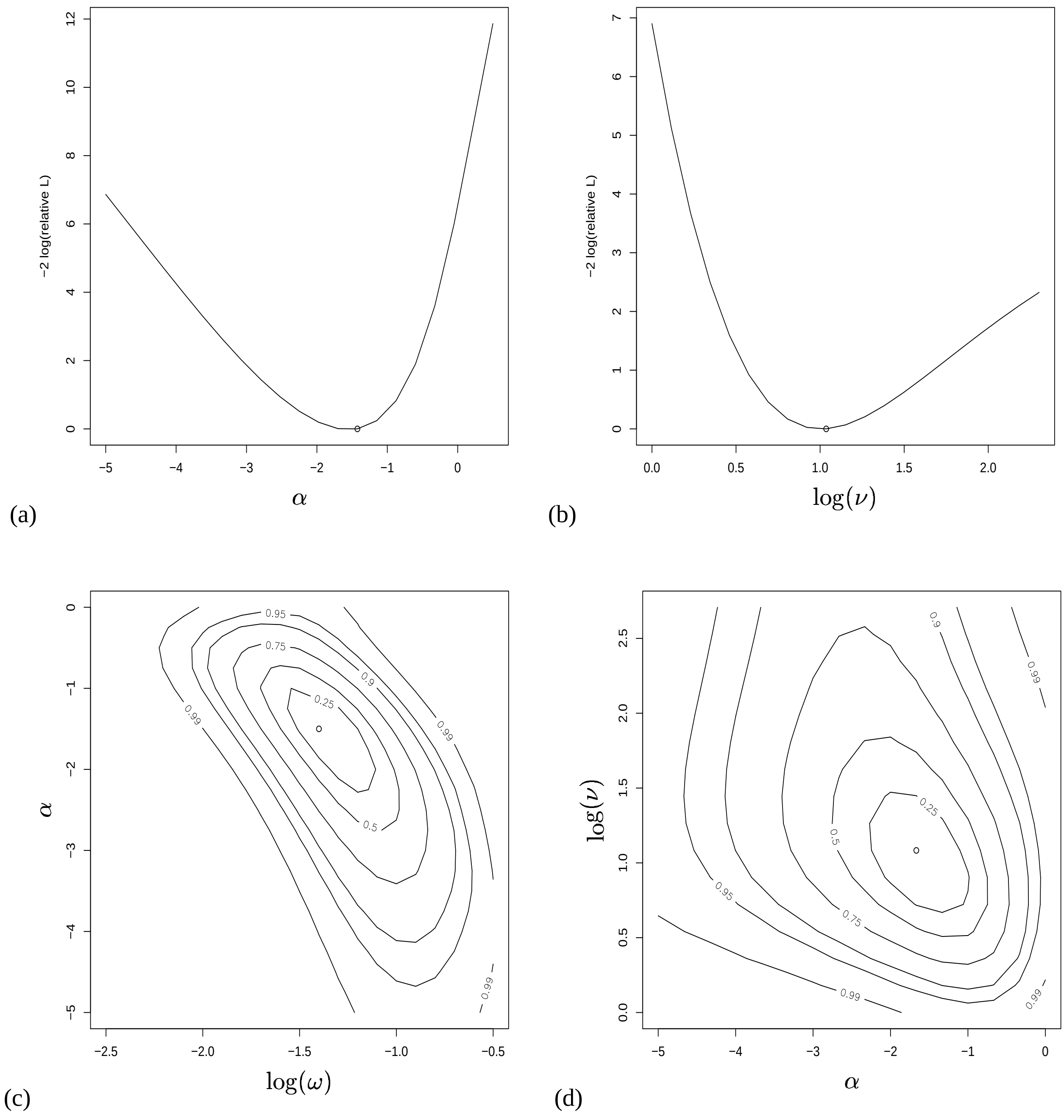

Figure 5: *Fiber–glass data: twice profile negative relative log-likelihood for parameters* α*,* $\log \nu$*,* $(\log \omega, \alpha)$ *and* $(\alpha, \log \nu)$ *are given in panel (a) to (d), respectively respectively*

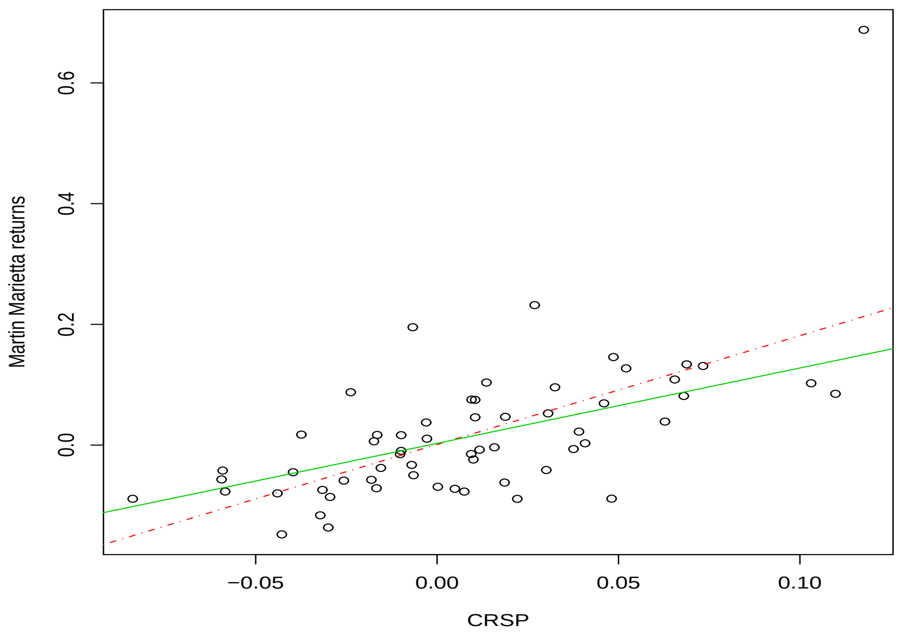

Figure 6: *Martin Marietta data: scatterplot and fitted regression lines; the dot-dashed line is the least squares fit, the continuous line is the one using a skew t error term*

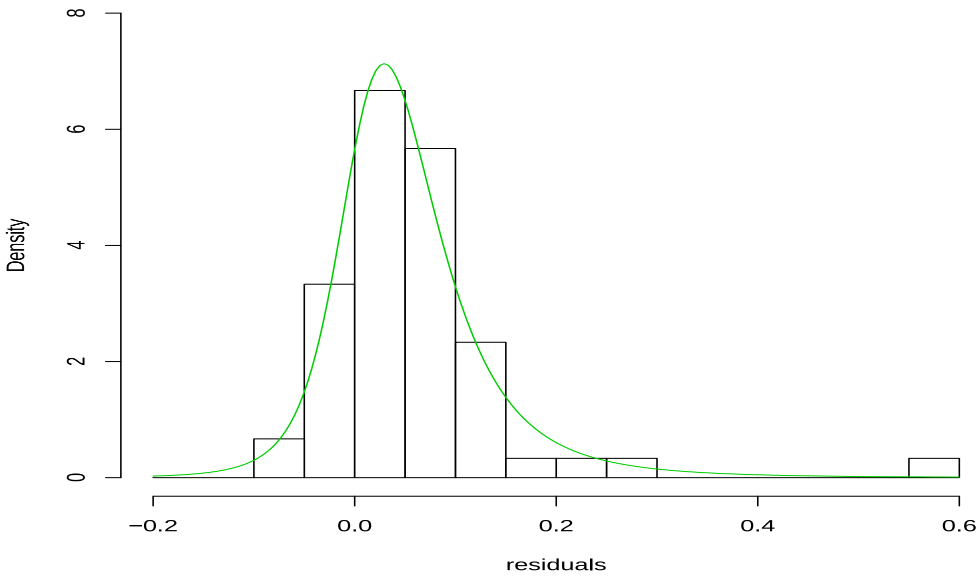

Figure 7: *Martin Marietta data: histogram of the residuals of linear regression and fitted skew t distribution*

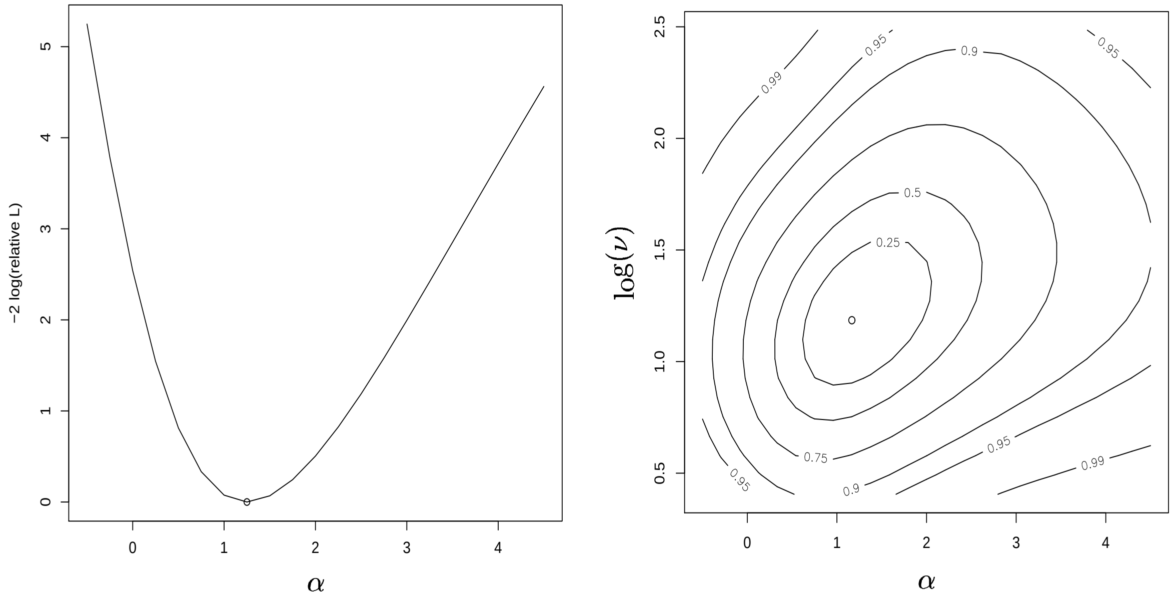

Figure 8: *Martin Marietta data: twice profile negative relative log-likelihood for parameters* α *(left panel)* $(\alpha, \log \nu)$ *(right panel)*

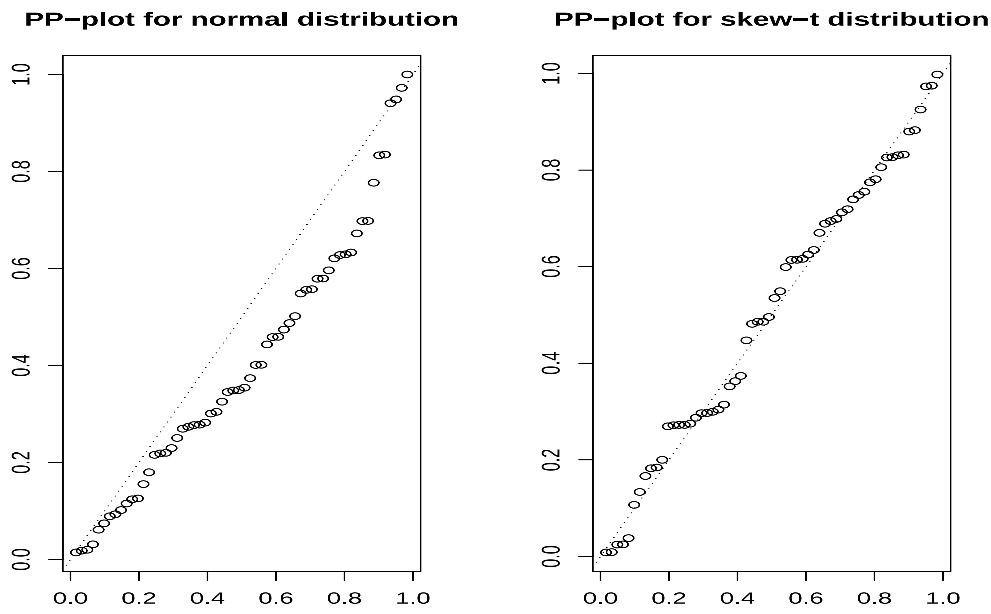

Figure 9: *Martin Marietta data: Healy's plot when either a normal distribution (left panel) or a skew t distribution (right panel)*

relationships among many of the above proposals, especially of those dealing with various formulations of skew elliptical family, by examining their connections and providing a more general approach to obtain several specific results.

Among the broad class of skew elliptical family, the multivariate skew t distribution offers ample flexibility for adapting itself to a very wide range of practical situations, and still it maintains mathematical tractability and a set of appealing formal properties. Some numerical evidence and the availability of developed software for inference provide additional support for using the distribution in practical cases. Other interesting distributions have been presented in the literature, most of which fall under the general umbrella of density (12) and its extensions discussed at the end of Section 2.

A wide and closely interconnected set of specific results is evolving towards a quite general framework. Open problems still exists, both on the probabilistic and on the inferential side of this area of work, as we have mentioned at various points in the paper, and additional, yet unexpected results will be discovered. However, what seems to us the more important direction of work, at this stage, is to make use of the available results in tackling real problems. This is the ultimate test to decide of the actual usefulness of all this work.

ACKNOWLEDGMENTS

We are grateful to Chris Jones for kindly sending us preliminary versions of his papers, to José Almer Sanqui for drawing our attention to Roberts paper, and to an anonymous referee for a number of helpful comments on an earlier versions of the paper. This research has been supported by MIUR, Italy, under grant scheme PRIN 2000.

APPENDIX

A PROOF OF PROPOSITION 8

Consider $Y = L^{-1\top}(Z - \xi)$, where the $d \times d$ matrix L is such that $\bar{\Omega} = L^\top L$. Then the density of Y is

$$2\, f(y; I) G\{w_L(y)\},$$

where $w_L(-y) = w(-L^\top y) = -w_L(y)$. Using the transformation to spherical coordinates

$$Y_j = R\left(\prod_{k=1}^{j-1} \sin\theta_k\right)\cos\theta_j, \quad 1 \le j \le d-1, \qquad Y_d = R\left(\prod_{k=1}^{d-2} \sin\theta_k\right)\sin\theta_{d-1},$$

where $R > 0$, $\theta_k \in [0, \pi)$, for $k = 1, \ldots, d-2$ and $\theta_{d-1} \in [0, 2\pi)$, and taking into account that the Jacobian is $r^{d-1}\prod_{k=1}^{d-2}(\sin\theta_k)^{d-k-1}$, we have

$$\begin{aligned} f_{\theta,R}(\theta, r) &= \\ &= 2\, c_d \tilde{f}(r^2) r^{d-1} \prod_{k=1}^{d-2} (\sin\theta_k)^{d-k-1} G\{w_L(r\cos\theta_1, r\sin\theta_1\cos\theta_2, \ldots, r\sin\theta_1\ldots\sin\theta_{d-1})\} \\ &= 2\, c_d \tilde{f}(r^2) r^{d-1} \prod_{k=1}^{d-2} (\sin\theta_k)^{d-k-1} G\{w_L^*(\theta, r)\} \end{aligned}$$

where $\theta = (\theta_1, \ldots, \theta_{d-1})^\top$, and $w_L^*(\theta, r) = w_L(r\cos\theta_1, r\sin\theta_1\cos\theta_2, \ldots, r\sin\theta_1\ldots\sin\theta_{d-1})$. Notice that $\dfrac{2\pi^{d/2}}{\Gamma(d/2)} c_d \tilde{f}(r^2) r^{d-1}$ is the density of the radius in the stochastic representation (11) of the elliptical random vector U, say, having density f, and $\dfrac{\Gamma(d/2)}{2\pi^{d/2}} \prod_{k=1}^{d-2}(\sin\theta_k)^{d-k-1}$ is the spherical coordinates representation of the uniform distribution on the unit sphere of $\mathbb{R}^d$; see Fang *et al.* (1990, Section 2.2.3). From Proposition 2 it follows that $R^2 \overset{d}{=} Y^\top Y \overset{d}{=} U^\top U$, so that the marginal density of R is given by

$$f_R(r) = \frac{2\pi^{d/2}}{\Gamma(d/2)} c_d \tilde{f}(r^2) r^{d-1}.$$

By integrating the joint density $f_{\theta,R}$ with respect to r, the marginal density of θ turns out to be

$$\begin{aligned} f_\theta(\theta) &= \frac{\Gamma(d/2)}{2\pi^{d/2}} \prod_{k=1}^{d-2} (\sin\theta_k)^{d-k-1}\, 2\int_0^\infty f_r(r) G\{w_L^*(\theta, r)\}\, \mathrm{d}r \\ &= \frac{\Gamma(d/2)}{\pi^{d/2}} \prod_{k=1}^{d-2} (\sin\theta_k)^{d-k-1} \mathbb{P}\{X \le w_L^*(\theta, R)\}\,, \end{aligned}$$

where X is a random variable with cumulative distribution function G.

The conditional density of θ given $R = r$ is equal to

$$f_{\theta|R=r}(\theta) = \frac{\Gamma(d/2)}{\pi^{d/2}} \prod_{k=1}^{d-2} (\sin\theta_k)^{d-k-1}\, G\{w_L^*(\theta, r)\},$$

which is a density of type (12) with location parameter $(\pi/2, \ldots, \pi/2, \pi)$. In fact for any $r > 0$ and any matrix L the equality

$$w_L^*(\pi - \theta_1, \pi - \theta_2, \ldots, \pi + \theta_{d-1}, r) = -w_L^*(\theta_1, \theta_2, \ldots, \theta_{d-1}, r)$$

holds true, and consequently the random variable $W_L^* = w_L^*(\theta, r)$ is symmetrically distributed around π. Then, using Lemma 1 in Azzalini & Capitanio (1999, p. 599), the result follows. QED

B DERIVATIVES OF THE SKEW t LOG-LIKELIHOOD

Write $U = (u_1, \ldots, u_n)^\top$. Then the derivatives of (30) are obtained from

$$\begin{aligned}
\frac{\partial \ell}{\partial \beta} &= -2\, X^\top \mathrm{diag}(\tilde{g}_Q + \tilde{T}_1 \odot \dot{t}_Q) U \Omega^{-1} - X^\top \mathrm{diag}(\tilde{T}_1 \odot \dot{t}_L) 1_n \eta^\top \\
\frac{\partial \ell}{\partial A} &= 2 \text{ upper triangle of } \left(D\, A\, U^\top\, \mathrm{diag}(\tilde{g}_Q + \tilde{T}_1 \odot \dot{t}_Q)\, U \right) \\
\frac{\partial \ell}{\partial D} &= I_d \odot \left(A\, U^\top \mathrm{diag}(\tilde{g}_Q + \tilde{T}_1 \odot \dot{t}_Q) U\, A^\top \right) + \tfrac{1}{2} n D^{-1} \\
\frac{\partial \ell}{\partial \eta} &= U^\top \mathrm{diag}(\tilde{T}_1 \odot \dot{t}_L) 1_n \\
\frac{\partial \ell}{\partial \nu} &= \sum \left(\frac{\partial \log g_d}{\partial \nu} + \frac{\partial \log T_1(t; \nu + d)}{\partial \nu} \right)
\end{aligned}$$

where the components of the vectors are obtained by evaluation of the quoted expressions at each of the n observations, $\odot$ denotes the Hadamard (or element-wise) product and

$$\begin{aligned}
\tilde{g}_Q &= \partial \log g_d(Q; \nu)/\partial Q = -\frac{\nu + d}{2\nu} (1 + Q/\nu)^{-1} \\
\tilde{T}_1 &= \partial \log T_1(t; \nu + d)/\partial t = T_1(t; \nu + d)^{-1} t_1(t; \nu + d) \\
\dot{t}_L &= \partial t(L, Q, \nu)/\partial L = \left(\frac{\nu + d}{Q + \nu} \right)^{1/2} \\
\dot{t}_Q &= \partial t(L, Q, \nu)/\partial Q = -\frac{L\, (\nu + d)^{1/2}}{2(Q + \nu)^{3/2}} \\
\frac{\partial \log g_d}{\partial \nu} &= \tfrac{1}{2} \left(\psi(\tfrac{1}{2}(\nu + d)) - \psi(\tfrac{1}{2}\nu) - d/\nu + \frac{(\nu + d) Q}{\nu^2 (1 + Q/\nu)} - \log(1 + Q/\nu) \right)
\end{aligned}$$

denoting by ψ the digamma function. What is not given above is an expression for

$$\frac{\partial \log T_1(t(L, Q, \nu); \nu + d)}{\partial \nu}$$

which appears intractable and must be evaluated numerically.

For transforming the above derivatives of D and ν into those of their logarithmic transform, we just use the chain rule

$$\frac{\partial \ell}{\partial \rho} = \frac{\partial \ell}{\partial D^*} (-2D^*), \qquad \frac{\partial \ell}{\partial \log \nu} = \frac{\partial \ell}{\partial \nu}\, \nu$$

where D^* denotes the diagonal of D.

The above expressions do not lend themselves to further differentiation. Therefore, in the numerical work described in Section 5, the observed information matrix has been obtained via numerical differentiation of the first derivatives.

REFERENCES

Arnold, B.C. and Beaver, R.J. (2000a). Hidden truncation models. *Sankhyā* **62**, 22–35.

Arnold, B.C. and Beaver, R.J. (2000b). The skew Cauchy distribution. *Statist. Prob. Lett.* **49**, 285–290.

Arnold, B.C. and Beaver, R.J. (2002). Skewed multivariate models related to hidden truncation and/or selective reporting. *Test* **11**, 7–54.

Azzalini, A. (1985). A class of distribution which includes the normal ones. *Scand. J. Statist.* **12**, 171–8.

Azzalini, A. and Capitanio, A. (1999). Statistical applications of the multivariate skew normal distribution. *J. Roy. Statist. Soc., B* **61** 579–602.

Azzalini, A. and Dalla Valle, A. (1996). The multivariate skew normal distribution. *Biometrika* **83**, 715–26.

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

Brownlee, K. A. (1960, 2nd ed. 1965). *Statistical Theory and Methodology in Science and Engineering*. New York: Wiley.

Butler, R. L., McDonald,J. B., Nelson, R. D. and White, S. B. (1990). Robust and partly adaptive estimation of regression models. *Rev. Econ. Statist.* **72**, 321–327.

Capitanio, A., Azzalini, A. and Stanghellini, E. (2003). Graphical models for skew normal variates. *Scand. J. Statist.* **30**, 129–144.

Cook, R. D. and Weisberg, S. (1994). *An Introduction to Regression Graphics*. Wiley, New York.

David, H. A. (1981). *Order statistics*, 2nd edition. Wiley, New York.

Fang, K.-T., Kotz, S. and Ng, K. (1990). *Symmetric multivariate and related distributions*. Chapman & Hall, London.

Fernández, C. and Steel, M. F. J. (1998). On Bayesian modelling of fat tails and skewness. *J. Am. Statist. Assoc.* **93**, 359–371.

Fernández, C. and Steel, M. F. J. (1999). Multivariate Student-t regression models: pitfalls and inference. *Biometrika* **86**, 153–168.

Genton, M. G., He, L. and Liu, X. (2001). Moments of skew normal random vectors and their quadratic forms. *Statist. & Prob. Lett.* **51**, 319–325.

Genton, M. G., & Loperfido, N. (2002). Generalized skew-elliptical distributions and their quadratic forms. Institute of Statistics Mimeo Series No. 2539, North Carolina State University. `http://www.stat.ncsu.edu/library/mimeo.html`

Genz, A. and Bretz, F. (1999). Numerical Computation of Multivariate t-Probabilities with Application to Power Calculation of Multiple Contrasts. *J. Stat. Comp. Simul.* **63**, 361–378.

Gupta, A.K., Gonzáles-Farías, G. and Domínguez-Molina, J. A. (2001). A Multivariate Skew Normal Distribution. Report I-01-19, `http://www.cimat.mx/reportes`

Healy, M. J. R. (1968). Multivariate normal plotting. *Appl. Statist.* **17**, 157–161.

Jones, M. C. (2002). Multivariate t and Beta distributions associated with the multivariate F distributions. *Metrika*, **54**, 215–231.

Jones, M.C. (2001). A skew t distribution. In *Probability and Statistical Models with Applications: a Volume in Honor of Theophilos Cacoullos*, eds: C. A. Charalambides, M. V. Koutras and N. Balakrishnan. Chapman and Hall, London, 269–278.

Jones, M.C. and Faddy, M.J.(2001). A skew extension of the t distribution, with applications. To appear.

Kano, Y. (1994). Consistency property of the elliptic probability density functions. *J. Multiv. An.* **51**, 139–147.

Loperfido, N. (2001). Quadratic forms of skew normal random vectors. *Statistics & Probability Letters* **54**, 381–387.

Loperfido, N. (2002). Statistical implications of selectively reported inferential results. *Statistics & Probability Letters* **56**, 13–22.

Roberts, C. (1966). A correlation model useful in the study of twins. *J. Am. Statist. Assoc.* **61**, 1184–1190.

Sahu, S. K., Dey, D. K. and Branco, M. (2001). A New Class of Multivariate Skew Distributions with Applications to Bayesian Regression Models. Tech. report `http://www.maths.soton.ac.uk/staff/Sahu/research/papers/skew.html`

Smith, R. L. and Naylor, J. C. (1987). A comparison of maximum likelihood and Bayesian estimators for the three-parameter Weibull distribution. *Appl. Statist.*, **36**, 358–369.

Szabłowski, P. J. (1998). Uniform distributions on spheres in finite-dimensional L_α and their generalization. *Journal of Multivariate Analysis* **64**, 103-117.

Zuo, Y., and Serfling, R. (2000). On the performance of some robust nonparametric location measures relative to a general notion of multivariate symmetry. *J. Statistical Planning and Inference* **84**, 55–79.