On Fisher Information Matrices and Profile Log-Likelihood Functions in Generalized Skew-Elliptical Models

Christophe Ley* and Davy Paindaveine†

E.C.A.R.E.S., Institut de Recherche en Statistique, and Département de Mathématique
Université Libre de Bruxelles, Brussels, Belgium

Abstract

In recent years, the skew-normal models introduced in Azzalini (1985) have enjoyed an amazing success, although an important literature has reported that they exhibit, in the vicinity of symmetry, singular Fisher information matrices and stationary points in the profile log-likelihood function for skewness, with the usual unpleasant consequences for inference. For general multivariate skew-symmetric and skew-elliptical models, the open problem of determining which symmetric kernels lead to each such singularity has been solved in Ley and Paindaveine (2009). In the present paper, we provide a simple proof that, in generalized skew-elliptical models involving the same skewing scheme as in the skew-normal distributions, Fisher information matrices, in the vicinity of symmetry, are singular for Gaussian kernels only. Then we show that if the profile log-likelihood function for skewness always has a point of inflection in the vicinity of symmetry, the generalized skew-elliptical distribution considered is actually skew-(multi)normal. In addition, we show that the class of multivariate skew-t distributions (as defined in Azzalini and Capitanio 2003), which was not covered by Ley and Paindaveine (2009), does not suffer from singular Fisher information matrices in the vicinity of symmetry. Finally, we briefly discuss the implications of our results on inference.

AMS 1980 subject classification : 62M15, 62G35.

Key words and phrases : Generalized skew-elliptical distributions, Profile likelihood, Singular Fisher information matrix, Skewness, Skew-normal distributions, Skew-t distributions.

Running title: Singularities in skew-elliptical models.

1 Introduction

Azzalini (1985) introduced the so-called skew-normal model, which embeds the univariate normal distributions into a flexible parametric class of (possibly) skewed distributions. More

*Research supported by a Mandat d'Aspirant of the Fonds National de la Recherche Scientifique, Communauté française de Belgique.

†Research supported by a Mandat d'Impulsion Scientifique of the Fonds National de la Recherche Scientifique, Communauté française de Belgique. Davy Paindaveine is member of ECORE, the association between CORE and ECARES.

formally, a random variable X is said to be skew-normal with location parameter $\mu \in \mathbb{R}$, scale parameter $\sigma \in \mathbb{R}_0^+$ and skewness parameter $\lambda \in \mathbb{R}$ if it admits the pdf

$$x \mapsto 2\sigma^{-1}\phi\Big(\frac{x-\mu}{\sigma}\Big)\Phi\Big(\lambda\Big(\frac{x-\mu}{\sigma}\Big)\Big), \qquad x \in \mathbb{R}, \tag{1.1}$$

where ϕ and Φ respectively denote the probability density function (pdf) and cumulative distribution function (cdf) of the standard normal distribution. An intensive study of these distributions has revealed some nice stochastic properties, which led to numerous generalizations of that univariate model. To cite a few, Azzalini and Dalla Valle (1996) introduced the multivariate skew-normal distribution by replacing the normal kernel ϕ with its k-variate extension, Branco and Dey (2001) and Azzalini and Capitanio (2003) defined multivariate skew-t distributions, while Azzalini and Capitanio (1999) and Branco and Dey (2001) proposed skew-elliptical distributions based on elliptically symmetric kernels. Azzalini and Capitanio (2003) also established a link between the distinct constructions of skew-elliptical distributions, and extended the latter into a more general class of skewed distributions containing all the pre-cited examples. This generalization is very much in the spirit of the so-called generalized skew-elliptical distributions analyzed in Genton and Loperfido (2005), associated with pdfs of the form

$$x \mapsto 2|\Sigma|^{-1/2} f(\Sigma^{-1/2}(x-\mu))\,\Pi(\Sigma^{-1/2}(x-\mu)), \qquad x \in \mathbb{R}^k, \tag{1.2}$$

where $\mu \in \mathbb{R}^k$ is a location parameter, Σ is a scatter parameter belonging to the class $\mathcal{S}_k$ of symmetric and positive definite $k \times k$ matrices (throughout, $|A|$ stands for the determinant of A, and $A^{1/2}$, for $A \in \mathcal{S}_k$, denotes the symmetric square-root of A, although any other square-root could be used), f (the *symmetric kernel*) is a spherically symmetric pdf, and where the mapping $\Pi : \mathbb{R}^k \to [0,1]$ satisfies $\Pi(-x) = 1 - \Pi(x)\ \forall x \in \mathbb{R}^k$. The densities in (1.2) in turn can be generalized by relaxing the spherical symmetry assumption on f into a weaker central symmetry one (under which $f(-x) = f(x)\ \forall x \in \mathbb{R}^k$); see Wang, Boyer and Genton (2004). For further information about models of skewed distributions and related topics, we refer the reader to the review paper Azzalini (2005).

Besides the quite appealing stochastic properties, the skew-normal model, as already noticed in Azzalini (1985), suffers from two closely related inferential problems: at $\lambda = 0$, corresponding to the symmetric situation, (i) the profile log-likelihood function for λ always admits a stationary point, and consequently, (ii) the Fisher information matrix for the three parameters in (1.1) is singular. These two unpleasant features are treated in, e.g., Pewsey (2000) or Azzalini and Genton (2008). Most authors show that, if a normal kernel is used, in some specific class of skewed densities containing those of (1.1), the problems in (i) and (ii) occur. However, from an inferential point of view, it is more important to derive an "iff" result, stating the exact conditions under which such singularities happen. Some "iff" results have been provided in parametric subclasses such as the skew-exponential one (DiCiccio and Monti 2004) or skew-t one (Gómez et al. 2007, DiCiccio and Monti 2009), until Ley and Paindaveine (2009) determined the collection of symmetric kernels bringing such singularities in broad semiparametric classes of multivariate skew-symmetric distributions.

In this paper, we first provide a simple and direct proof that, in generalized skew-elliptical models involving the same skewing scheme as in the skew-normal model, Fisher information matrices, in the vicinity of symmetry, are singular for Gaussian kernels only (Section 2). Then, after discussing the behavior of the first derivative of the profile log-likelihood function for skewness, we investigate the second derivative of that function; more precisely, we show that, if the profile log-likelihood function has a systematic point of inflection in the vicinity of symmetry ("systematic" here meaning for any sample of any sample size), then the generalized skew-elliptical distribution considered is actually skew-(multi)normal (Section 3). Still within the class of generalized skew-elliptical distributions, we then turn our attention towards the multivariate skew-t distributions as defined in Azzalini and Capitanio (2003); these distributions do not belong to the class of densities considered in Ley and Paindaveine (2009), but are the commonly used form of skew-t distributions in the literature. We show (Section 4) that the "iff" result from DiCiccio and Monti (2009) concerning the aforementioned singularities actually extends from the univariate to the multivariate setup. Finally, we briefly discuss the impact of our results on inference (Section 5).

2 On the singularity of Fisher information matrices

As mentioned in the Introduction, there exist many distinct generalizations of the skew-normal distributions described in Azzalini (1985). We now focus on generalized skew-elliptical distributions involving the same skewing scheme as in the skew-normal model. More precisely, within the class of densities of type (1.2), we consider densities of the form

$$x \mapsto f^{\Pi}_{g;\mu,\Sigma,\lambda}(x) := 2|\Sigma|^{-1/2} g(\|\Sigma^{-1/2}(x-\mu)\|)\, \Pi(\lambda'\Sigma^{-1/2}(x-\mu)), \qquad x \in \mathbb{R}^k, \qquad (2.3)$$

where $\mu \in \mathbb{R}^k$ is a location parameter, $\Sigma \in \mathcal{S}_k$ a scatter parameter, $\lambda \in \mathbb{R}^k$ a skewness parameter, where the *radial density* $g : \mathbb{R}_0^+ \to \mathbb{R}^+$ satisfies (a) $\int_{\mathbb{R}^k} g(\|x\|)\, dx = 1$ and (b) $\int_{\mathbb{R}^k} \|x\|^2 g(\|x\|)\, dx = k$, and where the skewing function Π is described in Assumption (A) below. Condition (b) allows to identify the elliptically symmetric part (obtained for $\lambda = 0$) of (2.3); on one hand, it guarantees that Σ and g are identifiable, and on the other hand it is so that the (multi)normal case is associated with the *Gaussian kernel* $g(r) := c_k \exp(-r^2/2)$, where $c_k := (2\pi)^{-k/2}$ is the normalizing constant determined by Condition (a). The functions g and Π further need to satisfy

Assumption (A). *(i) The radial density $g : \mathbb{R}_0^+ \to \mathbb{R}^+$ belongs to the collection $\mathcal{G}$ of a.e. positive, continuously differentiable functions for which Conditions (a)-(b) above hold and for which, letting $\varphi_g := -g'/g$, $\int_{\mathbb{R}^k} \varphi_g^2(\|x\|) g(\|x\|)\, dx$ and $\int_{\mathbb{R}^k} \|x\|^2 \varphi_g^2(\|x\|) g(\|x\|)\, dx$ are finite. (ii) The skewing function $\Pi : \mathbb{R} \to [0,1]$ is a continuously differentiable function that satisfies $\Pi(-x) = 1 - \Pi(x)$ for all $x \in \mathbb{R}$, and $\Pi'(0) \neq 0$.*

We adopt the following notation. Let $\mathrm{vec}\,\Sigma$ denote the k^2-vector obtained by stacking the columns of Σ on top of each other, and $\mathrm{vech}\,\Sigma$ be the $k(k+1)/2$-subvector of $\mathrm{vec}\,\Sigma$ where only upper diagonal entries in Σ are considered (in order to avoid heavy notations, we will write ∇_Σ instead of $\nabla_{\mathrm{vech}\,\Sigma}$ in what follows). Define $\Sigma^{\otimes 2} := \Sigma \otimes \Sigma$, where $\otimes$ stands for the usual Kronecker product, and let I_ℓ and K_k be the ℓ-dimensional identity matrix and the $k^2 \times k^2$ commutation matrix, respectively. Finally, put $J_k := (\mathrm{vec}\, I_k)(\mathrm{vec}\, I_k)'$ and let P_k be the matrix such that $P_k'(\mathrm{vech}\, A) = \mathrm{vec}\, A$ for any $k \times k$ symmetric matrix A.

Under Assumption (A), the scores for $\mu, \mathrm{vech}\,\Sigma$ and λ, in the vicinity of symmetry (that is, at $\lambda = 0$), are the quantities $d_{g;\mu}(x), d_{g;\Sigma}(x)$ and $d_{g;\lambda}(x)$, respectively, in

$$\ell_{g;\mu,\Sigma,\lambda}(x) := \begin{pmatrix} d_{g;\mu}(x) \\ d_{g;\Sigma}(x) \\ d_{g;\lambda}(x) \end{pmatrix}$$
$$:= \begin{pmatrix} \frac{\varphi_g(||\Sigma^{-1/2}(x-\mu)||)}{||\Sigma^{-1/2}(x-\mu)||}\Sigma^{-1}(x-\mu) \\ \frac{1}{2}P_k(\Sigma^{\otimes 2})^{-1/2}\mathrm{vec}\left(\frac{\varphi_g(||\Sigma^{-1/2}(x-\mu)||)}{||\Sigma^{-1/2}(x-\mu)||}\Sigma^{-1/2}(x-\mu)(x-\mu)'\Sigma^{-1/2} - I_k\right) \\ 2\Pi'(0)\Sigma^{-1/2}(x-\mu) \end{pmatrix},$$

where the factor 2 in the λ-score follows from the fact that $\Pi(0) = 1/2$. The corresponding Fisher information matrix is then given by $\Gamma_g := \int_{\mathbb{R}^k} \ell_{g;\mu,\Sigma,\lambda}(x)\ell'_{g;\mu,\Sigma,\lambda}(x)f^{\Pi}_{g;\mu,\Sigma,0}(x)dx$, a matrix that naturally partitions into

$$\Gamma_g = \begin{pmatrix} \Gamma_{g;\mu\mu} & 0 & \Gamma_{g;\mu\lambda} \\ 0 & \Gamma_{g;\Sigma\Sigma} & 0 \\ \Gamma_{g;\lambda\mu} & 0 & \Gamma_{g;\lambda\lambda} \end{pmatrix},$$

with

$$\Gamma_{g;\mu\mu} := \frac{1}{k}\mathrm{E}[\varphi_g^2(\|Z\|)]\Sigma^{-1}, \quad \Gamma_{g;\mu\lambda} := 2\Pi'(0)\Sigma^{-1/2} =: \Gamma_{g;\lambda\mu}, \quad \Gamma_{g;\lambda\lambda} := \frac{4}{k}(\Pi'(0))^2\mathrm{E}[\|Z\|^2]I_k,$$

and

$$\Gamma_{g;\Sigma\Sigma} := \frac{1}{4}P_k(\Sigma^{\otimes 2})^{-1/2}\{(k(k+2))^{-1}\mathrm{E}[\|Z\|^2\varphi_g^2(\|Z\|)](I_{k^2} + K_k + J_k) - J_k\}(\Sigma^{\otimes 2})^{-1/2}P_k';$$

here, Z stands for a random vector with the same distribution as $\Sigma^{-1/2}(X - \mu)$, where X has pdf $f^{\Pi}_{g;\mu,\Sigma,0}$. The zero blocks in Γ_g can easily be obtained by noticing that the score in $\mathrm{vech}\,\Sigma$ is symmetric with respect to $x - \mu$, while the scores in μ and λ are anti-symmetric with respect to the same quantity.

The expression of Γ_g enables us to have

Theorem 2.1 *Let Assumption (A) hold. Then, Γ_g is singular iff g is the Gaussian kernel (i.e., $g(r) = c_k \exp(-r^2/2)$ for all $r > 0$).*

We here propose the following simple and direct proof of this result.

Proof of Theorem 2.1. Since $\Gamma_{g;\Sigma\Sigma}$ is invertible (see, e.g., Hallin and Paindaveine 2006), Γ_g is singular iff

$$\Gamma_g^{\mathrm{sub}} := \begin{pmatrix} \Gamma_{g;\mu\mu} & \Gamma_{g;\mu\lambda} \\ \Gamma_{g;\lambda\mu} & \Gamma_{g;\lambda\lambda} \end{pmatrix}$$

is singular. Now, since $\Gamma_{g;\mu\mu} = \frac{1}{k}\mathrm{E}[\varphi_g^2(\|Z\|)]\Sigma^{-1}$ is invertible, we have that $|\Gamma_g^{\mathrm{sub}}| = |\Gamma_{g;\mu\mu}||\Gamma_{g;\lambda\lambda} - \Gamma_{g;\lambda\mu}\Gamma^{-1}_{g;\mu\mu}\Gamma_{g;\mu\lambda}|$. Hence, Γ_g is singular iff $\Gamma_{g;\lambda\lambda} - \Gamma_{g;\lambda\mu}\Gamma^{-1}_{g;\mu\mu}\Gamma_{g;\mu\lambda} = 4(\Pi'(0))^2 (\mathrm{E}[\|Z\|^2]\mathrm{E}[\varphi_g^2(\|Z\|)] - k^2)I_k/(k\mathrm{E}[\varphi_g^2(\|Z\|)])$ is, which, under Assumption (A), is the case iff

$$\mathrm{E}[\|Z\|^2]\mathrm{E}[\varphi_g^2(\|Z\|)] = k^2. \tag{2.4}$$

Now, by applying the Cauchy-Schwarz inequality and by integrating by parts, we obtain

$$\sqrt{\mathrm{E}[\|Z\|^2]\mathrm{E}[\varphi_g^2(\|Z\|)]} \geq \mathrm{E}[\|Z\|\varphi_g(\|Z\|)] = \Big(\int_0^\infty r^{k-1} g(r)\, dr\Big)^{-1}\Big(\int_0^\infty r\, \varphi_g(r) r^{k-1}\, g(r)\, dr\Big) = k.$$

Equation (2.4) indicates that the Cauchy-Schwarz inequality here is an equality, which implies that there exists some real number η such that $\varphi_g(r) = \eta r$ for all $r > 0$. Solving the latter differential equation yields $g(r) = \gamma \exp[-\eta r^2/2]$. Since $\int_{\mathbb{R}^k} g(\|x\|)\, dx = 1$ and $\int_{\mathbb{R}^k} \|x\|^2 g(\|x\|)\, dx = k$, we must have $\gamma = c_k$ and $\eta = 1$, which establishes the result. □

Theorem 2.1 states that, in the semiparametric class of generalized skew-elliptical densities of type (2.3), only the skew-(multi)normal parametric submodel may suffer from the inferential problems usually associated with singular Fisher information matrices. We refer to Ley and Paindaveine (2009) for a discussion of such issues (see also Section 5 below), as well as for possible extensions of Theorem 2.1 to the skew-symmetric setup. Here, we only stress that the simple proof above can be straightforwardly adapted to the setup where Σ, in the argument of the mapping Π in (2.3), is replaced with the identity matrix I_k (Genton and Loperfido 2005 also consider generalized skew-elliptical distributions of this form).

3 On the profile log-likelihood function for skewness

For a given random sample $X^{(n)} := (X_1, \ldots, X_n)$ from (2.3), we define the profile log-likelihood function for the skewness parameter λ as

$$\lambda \mapsto \tilde{L}^{\Pi}_{g;\lambda}(X^{(n)}) := \sup_{\mu\in\mathbb{R}^k,\, \Sigma\in\mathcal{S}_k} L^{\Pi}_{g;\mu,\Sigma,\lambda}(X^{(n)}), \qquad \lambda \in \mathbb{R}^k, \tag{3.5}$$

where $L^{\Pi}_{g;\mu,\Sigma,\lambda}(X^{(n)}) := \sum_{i=1}^n \log f^{\Pi}_{g;\mu,\Sigma,\lambda}(X_i)$ is the standard log-likelihood function associated with $X^{(n)}$. This expression can be rewritten under the more tractable form $\tilde{L}^{\Pi}_{g;\lambda}(X^{(n)}) = L^{\Pi}_{g;\hat{\mu}_g(\lambda),\hat{\Sigma}_g(\lambda),\lambda}(X^{(n)})$, where $\hat{\mu}_g(\lambda)$ and $\hat{\Sigma}_g(\lambda)$ stand for the MLEs of μ and Σ at fixed λ. This equality implies that $\nabla_\lambda \tilde{L}^{\Pi}_{g;\lambda}(X^{(n)}) = \nabla_\lambda L^{\Pi}_{g;\mu,\Sigma,\lambda}(X^{(n)})|_{(\mu,\Sigma,\lambda)=(\hat{\mu}_g(\lambda),\hat{\Sigma}_g(\lambda),\lambda)}$ (see Ley and Paindaveine 2009 for explicit calculations, or Barndorff-Nielsen and Cox 1994 for general results on profile log-likelihood functions). Consequently, a necessary and sufficient condition for the profile log-likelihood function to always admit a stationary point at $\lambda = 0$ is that

$$\nabla_\lambda \tilde{L}^{\Pi}_{g;\lambda}(X^{(n)})|_{\lambda=0} = 2\Pi'(0)(\hat{\Sigma}_g(0))^{-1/2} \sum_{i=1}^n (X_i - \hat{\mu}_g(0)) = 0$$

for any sample $X^{(n)}$. Since $\hat{\Sigma}_g(0) \in \mathcal{S}_k$, hence is invertible, this means that $\hat{\mu}_g(0)$, the MLE for the location parameter μ at $\lambda = 0$, must coincide, for any $X^{(n)}$, with $\bar{X}^{(n)} := \frac{1}{n}\sum_{i=1}^n X_i$. The result in Theorem 3.1 below—which is in line with the statement of Theorem 2.1—thus directly follows from a well-known characterization property of (multi)normal distributions which can be traced back to Gauss (see Azzalini and Genton 2007 for a recent account).

Theorem 3.1 *Let Assumption (A) hold. Then, the profile log-likelihood function for skewness in (3.5) admits, for any sample $X^{(n)}$ of size $n \geq 3$, a stationary point at $\lambda = 0$ iff g is the Gaussian kernel.*

Although Theorem 3.1 only considers the first derivative of the profile log-likelihood function, it also shows that, in the class of generalized skew-elliptical densities of type (2.3), only the skew-(multi)normal ones may exhibit a systematic saddle point at $\lambda = 0$ in the profile log-likelihood function for skewness (since a saddle point is in particular a stationary point). In contrast with this, we now investigate the conditions under which the profile log-likelihood function for skewness has a systematic *point of inflection* at $\lambda = 0$, without requiring anything on the first derivative (hence, in particular, without requiring that this point of inflection is a saddle point).

To avoid any confusion, we will say that $x_0 \in \mathbb{R}^k$ is a point of inflection for $h : \mathbb{R}^k \to \mathbb{R}$ (of class C^1) if there is no neighborhood of $(x_0', h(x_0))' \in \mathbb{R}^{k+1}$ in which the graph G_h of h is fully contained in one of the two (closed) halfspaces determined by the hyperplane tangent to G_h at $(x_0', h(x_0))'$. Defining Assumption (A′) as the reinforcement of Assumption (A) obtained by further requiring that both g and Π are of class C^2 (which clearly implies that $\Pi''(0) = 0$), we have the following result.

Theorem 3.2 *Let Assumption (A′) hold. Then, if the profile log-likelihood function for skewness in (3.5) admits a point of inflection at* $\lambda = 0$ *for any sample* $X^{(n)}$, g *is the Gaussian kernel.*

Proof of Theorem 3.2. First note that, under Assumption (A′), the profile log-likelihood function for skewness is of class C^2. Hence, if the latter has a systematic point of inflection at $\lambda = 0$, there exists a non-zero k-vector v such that

$$v'[(\nabla_\lambda \nabla_\lambda')\tilde{L}^{\Pi}_{g;\lambda}(X^{(n)})|_{\lambda=0}]v = 0 \tag{3.6}$$

for any sample $X^{(n)}$; throughout, we denote by $(\nabla_a \nabla_b')h_{a,b}(.)$ the matrix whose entry (i,j) is given by $\frac{\partial^2 h_{a,b}(.)}{\partial a_i \partial b_j}$. By using the main result in Patefield (1977), we can write

$$(\nabla_\lambda \nabla_\lambda')\tilde{L}^{\Pi}_{g;\lambda}(X^{(n)}) = \Big[(\nabla_\lambda \nabla_\lambda')L^{\Pi}_{g;\alpha,\lambda}(X^{(n)}) \\ -(\nabla_\lambda \nabla_\alpha')L^{\Pi}_{g;\alpha,\lambda}(X^{(n)})[(\nabla_\alpha \nabla_\alpha')L^{\Pi}_{g;\alpha,\lambda}(X^{(n)})]^{-1}(\nabla_\alpha \nabla_\lambda')L^{\Pi}_{g;\alpha,\lambda}(X^{(n)})\Big]|_{(\alpha,\lambda)=(\hat{\alpha}_g(\lambda),\lambda)},$$

where $\alpha = (\mu', (\mathrm{vech}\,\Sigma)')'$ and $\hat{\alpha}_g(\lambda)$ stands for the MLE of α at fixed λ. By (3.6), we then have

$$v'\Big\{\Big[\tfrac{1}{n}(\nabla_\lambda \nabla_\lambda')L^{\Pi}_{g;\alpha,\lambda}(X^{(n)}) \\ -\tfrac{1}{n}(\nabla_\lambda \nabla_\alpha')L^{\Pi}_{g;\alpha,\lambda}(X^{(n)})[\tfrac{1}{n}(\nabla_\alpha \nabla_\alpha')L^{\Pi}_{g;\alpha,\lambda}(X^{(n)})]^{-1}\tfrac{1}{n}(\nabla_\alpha \nabla_\lambda')L^{\Pi}_{g;\alpha,\lambda}(X^{(n)})\Big]|_{(\alpha,\lambda)=(\hat{\alpha}_g(0),0)}\Big\}v = 0 \tag{3.7}$$

for any sample $X^{(n)}$. Now, writing $M_g(x) := (\nabla_x \nabla_x') \log g(\|x\|)$, direct computations show that

$$(\nabla_\lambda \nabla_\mu')L^{\Pi}_{g;\mu,\Sigma,\lambda}(X^{(n)})|_{\lambda=0} = -2n\Pi'(0)\Sigma^{-1/2},$$

$$(\nabla_\mu \nabla_\mu')L^{\Pi}_{g;\mu,\Sigma,\lambda}(X^{(n)})|_{\lambda=0} = \sum_{i=1}^n \Sigma^{-1/2} M_g(\Sigma^{-1/2}(X_i - \mu))\Sigma^{-1/2},$$

and

$$(\nabla_\lambda \nabla_\lambda')L^{\Pi}_{g;\mu,\Sigma,\lambda}(X^{(n)})|_{\lambda=0} = -4(\Pi'(0))^2\Sigma^{-1/2}\left[\sum_{i=1}^n (X_i - \mu)(X_i - \mu)'\right]\Sigma^{-1/2},$$

whereas it is easy to show (by using the expression for $\bar{\Delta}^n_{\xi,f_1}$ on page 2246 of Paindaveine 2008) that, for some mappings $N_{g;\Sigma}$ and $P_{g;\Sigma}$ satisfying $\mathrm{E}[N_{g;\Sigma}(X-\mu)]=0$ and $\mathrm{E}[P_{g;\Sigma}(X-\mu)]=0$ if X admits the pdf $f^{\Pi}_{g;\mu,\Sigma,0}$,

$$(\nabla_\mu \nabla'_\Sigma) L^{\Pi}_{g;\mu,\Sigma,\lambda}(X^{(n)})|_{\lambda=0} = \sum_{i=1}^n N_{g;\Sigma}(X_i-\mu),$$

and

$$(\nabla_\lambda \nabla'_\Sigma) L^{\Pi}_{g;\mu,\Sigma,\lambda}(X^{(n)})|_{\lambda=0} = \sum_{i=1}^n P_{g;\Sigma}(X_i-\mu).$$

Since $\frac{1}{n}(\nabla_\mu \nabla'_\Sigma) L^{\Pi}_{g;\mu,\Sigma,\lambda}(X^{(n)})|_{\lambda=0}$ and $\frac{1}{n}(\nabla_\lambda \nabla'_\Sigma) L^{\Pi}_{g;\mu,\Sigma,\lambda}(X^{(n)})|_{\lambda=0}$ converge to zero in probability as $n\to\infty$, substituting in (3.7) and taking the limit in probability as $n\to\infty$ yields

$$-4(\Pi'(0))^2\, v'(\Sigma^{-1/2}\mathrm{E}[(X-\mu)(X-\mu)']\Sigma^{-1/2} + (\mathrm{E}[M_g(\Sigma^{-1/2}(X-\mu))])^{-1})v = 0.$$

Dividing by $-4(\Pi'(0))^2$, letting, as before, Z stand for a random vector with the same distribution as $\Sigma^{-1/2}(X-\mu)$, where X has pdf $f^{\Pi}_{g;\mu,\Sigma,0}$, and integrating by parts in the second term, yields

$$v'(\tfrac{1}{k}\mathrm{E}[\|Z\|^2]I_k + (-\tfrac{1}{k}\mathrm{E}[\varphi_g^2(\|Z\|)]I_k)^{-1})v = \frac{\mathrm{E}[\|Z\|^2]\mathrm{E}[\varphi_g^2(\|Z\|)]-k^2}{k\mathrm{E}[\varphi_g^2(\|Z\|)]}\|v\|^2 = 0.$$

Since v is a non-zero vector, we must have that $\mathrm{E}[\|Z\|^2]\mathrm{E}[\varphi_g^2(\|Z\|)] = k^2$, which coincides with (2.4). We therefore obtain, in the same way as in the proof of Theorem 2.1, that g must be the Gaussian kernel, which concludes the proof. □

This result shows that, in the class of generalized skew-elliptical densities of type (2.3), a systematic point of inflection (at $\lambda=0$) again can happen at skew-(multi)normal densities only. Similarly to the previous section, the theorem remains valid if Σ is replaced with I_k in the argument of Π in (2.3). The proof is actually easier in this setup since $(\nabla_\lambda \nabla'_\Sigma) L^{\Pi}_{g;\mu,\Sigma,\lambda}(X^{(n)})|_{\lambda=0}$ is then exactly zero—and not only a $o_P(n)$ quantity, as in the proof of Theorem 3.2.

Strengthening Theorem 3.2 into an "iff" result would require an investigation of higher-order derivatives of the profile log-likelihood function for skewness. This has been achieved for the univariate case in Chiogna (2005), where it is shown that the profile function, under skew-normal distributions, always admits a point of inflection in the vicinity of symmetry. Hence, the desired "iff" statement is formally established in the scalar case, and should extend to the general multinormal setup. Finally, whereas Theorem 3.1 can be shown to hold in broader multivariate skew-symmetric models (see Ley and Paindaveine 2009 for a precise statement), Theorem 3.2 seems to be limited to generalized skew-elliptical distributions, mainly due to the very subtle interplay, in the multivariate skew-symmetric setup, between the nature of points of inflection and the rank of the corresponding Hessian matrices.

4 Investigating the case of multivariate skew-t distributions.

In this section, we study the possible singularity of Fisher information matrices in multivariate skew-t models. It should be noted that the commonly adopted form of those distributions (namely the one from Azzalini and Capitanio 2003) is based on a skewing function that *is not* of the form $\Pi(\lambda' \cdot)$ (see the pdf (4.8) below), which explains that the important case of multivariate skew-t densities does not enter the framework of the previous sections and has to be treated in this separate section.

For any $\nu > 0$ and any positive integer k, consider the mapping

$$r \mapsto g_{k,\nu}(r) = \frac{\Gamma((\nu+k)/2)}{(\pi\nu)^{k/2}\Gamma(\nu/2)}(1 + r^2/\nu)^{-(\nu+k)/2}, \quad r > 0,$$

which is such that $x \mapsto g_{k,\nu}(\|x\|)$, $x \in \mathbb{R}^k$, is the pdf of the k-dimensional standard t distribution with ν degrees of freedom. Also denote by $y \mapsto T_\eta(y)$, $y \in \mathbb{R}$, the cdf of the scalar t distribution with η degrees of freedom. Multivariate skew-t densities—as defined by Azzalini and Capitanio (2003)—then have densities of the form

$$\begin{aligned} x \mapsto f^{\mathrm{ST}}_{\mu,\Sigma,\lambda,\nu}(x) \;&:=\; 2|\Sigma|^{-1/2} g_{k,\nu}(||\Sigma^{-1/2}(x-\mu)||) \\ &\times T_{\nu+k}\left(\lambda'\sigma^{-1}(x-\mu)\left(\frac{\nu+k}{||\Sigma^{-1/2}(x-\mu)||^2+\nu}\right)^{1/2}\right), \quad x \in \mathbb{R}^k, \end{aligned} \tag{4.8}$$

where $\mu \in \mathbb{R}^k$ is the location parameter, $\Sigma \in \mathcal{S}^k$ the scatter parameter, $\lambda \in \mathbb{R}^k$ the skewness parameter, $\nu > 0$ the tail weight parameter, and where σ is the $k \times k$ diagonal matrix with diagonal entries $\sigma_{ii} = \Sigma_{ii}^{1/2}$, $i = 1, \ldots, k$. When $\nu \to \infty$, (4.8) becomes the density of a k-dimensional skew-normal distribution.

Azzalini and Genton (2008) provided evidence that, for any finite value of ν, the univariate skew-t distributions do not suffer, in the vicinity of symmetry, from a singular Fisher information matrix—and hence not from a systematic stationary point in the profile log-likelihood function for skewness. DiCiccio and Monti (2009) formally proved that statement. As for the multivariate setup, Azzalini and Genton (2008) showed that those singularities in the vicinity of symmetry also appear in multivariate skew-normal distributions, and conjectured that, as in the univariate case, no member from the multivariate skew-t class should suffer from those problems. We now prove that their conjecture actually holds.

For multivariate skew-t distributions with fixed ν value, the score for $(\mu', (\mathrm{vech}\,\Sigma)', \lambda')'$, in the vicinity of symmetry, is given by

$$\begin{aligned} \ell^{\mathrm{ST}}_{g_{k,\nu};\mu,\Sigma,\lambda}(x) &:= \begin{pmatrix} d^{\mathrm{ST}}_{g_{k,\nu};\mu}(x) \\ d^{\mathrm{ST}}_{g_{k,\nu};\Sigma}(x) \\ d^{\mathrm{ST}}_{g_{k,\nu};\lambda}(x) \end{pmatrix} = \begin{pmatrix} d_{g_{k,\nu};\mu}(x) \\ d_{g_{k,\nu};\Sigma}(x) \\ d^{\mathrm{ST}}_{g_{k,\nu};\lambda}(x) \end{pmatrix} \\ &= \begin{pmatrix} \frac{(1+k/\nu)}{(1+||\Sigma^{-1/2}(x-\mu)||^2/\nu)}\Sigma^{-1}(x-\mu) \\ \frac{1}{2}P_k(\Sigma^{\otimes 2})^{-1/2}\mathrm{vec}\left(\frac{(1+k/\nu)}{(1+||\Sigma^{-1/2}(x-\mu)||^2/\nu)}\Sigma^{-1/2}(x-\mu)(x-\mu)'\Sigma^{-1/2} - I_k\right) \\ 2\,t_{\nu+k}(0)\sigma^{-1}(x-\mu)\left(\frac{1+k/\nu}{1+||\Sigma^{-1/2}(x-\mu)||^2/\nu}\right)^{1/2} \end{pmatrix}, \end{aligned} \tag{4.9}$$

where $t_{\nu+k}(0)$ stands for the derivative of $y \mapsto T_{\nu+k}(y)$ at 0; the exponent ST makes the notation somewhat heavy, but actually stresses the fact that $\ell_{g_{k,\nu};\mu,\Sigma,\lambda}(x)$ (that is, the score associated with the $g = g_{k,\nu}$ version of the $\ell_{g;\mu,\Sigma,\lambda}(x)$ score from Section 2) and $\ell^{\mathrm{ST}}_{g_{k,\nu};\mu,\Sigma,\lambda}(x)$ do not coincide, due to the fact that both skewing mechanisms are of a different nature. As suggested in (4.9), they differ only through the λ-part of the scores, though.

Now, note that the new λ-score $d^{\rm ST}_{g_{k,\nu};\lambda}(x)$ remains an anti-symmetric function of $x-\mu$. Hence, the symmetry properties, still with respect to $x-\mu$, of $d^{\rm ST}_{g_{k,\nu};\mu}(x)=d_{g_{k,\nu};\mu}(x)$ and $d^{\rm ST}_{g_{k,\nu};\Sigma}(x)=d_{g_{k,\nu};\Sigma}(x)$ entail that the resulting Fisher information matrix takes the form

$$\Gamma^{\rm ST}_{g_{k,\nu}} = \begin{pmatrix} \Gamma_{g_{k,\nu};\mu\mu} & 0 & \Gamma^{\rm ST}_{g_{k,\nu};\mu\lambda} \\ 0 & \Gamma_{g_{k,\nu};\Sigma\Sigma} & 0 \\ \Gamma^{\rm ST}_{g_{k,\nu};\lambda\mu} & 0 & \Gamma^{\rm ST}_{g_{k,\nu};\lambda\lambda} \end{pmatrix},$$

with the same quantities $\Gamma_{g_{k,\nu};\mu\mu}$ and $\Gamma_{g_{k,\nu};\Sigma\Sigma}$ as in Section 2 (here evaluated at the radial density $g_{k,\nu}$). Note that the finiteness of $\Gamma^{\rm ST}_{g_{k,\nu}}$ requires that $\nu>2$, a condition that guarantees that the parent elliptically symmetric t distribution has finite second-order moments; see Assumption (A).

We can now state the following result.

Theorem 4.1 *Fix $\nu\in(2,\infty)$ and an arbitrary positive integer k. Then, at any parameter value $(\mu',(\mathrm{vech}\,\Sigma)',0')'\in\mathbb{R}^k\times\mathcal{S}^k\times\{0\}$, the Fisher information matrix associated with the fixed-ν subclass of multivariate skew-t densities in (4.8), namely $\Gamma^{\rm ST}_{g_{k,\nu}}$, is non-singular.*

Proof of Theorem 4.1. First note that, for $\nu>2$, the radial density $g_{k,\nu}$ satisfies the g-part of Assumption (A), which entails that (see Hallin and Paindaveine 2006) $\Gamma_{g_{k,\nu};\mu\mu}$ and $\Gamma_{g_{k,\nu};\Sigma\Sigma}$ are invertible. It is therefore sufficient to show that the (μ,λ)-submatrix of $\Gamma^{\rm ST}_{g_{k,\nu}}$ is non-singular. Denoting by Z a k-variate random vector with the same distribution as $\Sigma^{-1/2}(X-\mu)$, where X admits the pdf $f^{\rm ST}_{\mu,\Sigma,0,\nu}$, we easily obtain that

$$\Gamma_{g_{k,\nu};\mu\mu}=(1+k/\nu)^2\Sigma^{-1/2}\mathrm{E}\Big[\frac{ZZ'}{(1+||Z||^2/\nu)^2}\Big]\Sigma^{-1/2},$$

$$\Gamma^{\rm ST}_{g_{k,\nu};\lambda\lambda}=4(t_{\nu+k}(0))^2(1+k/\nu)\sigma^{-1}\Sigma^{1/2}\mathrm{E}\Big[\frac{ZZ'}{(1+||Z||^2/\nu)}\Big]\Sigma^{1/2}\sigma^{-1},$$

and

$$\Gamma^{\rm ST}_{g_{k,\nu};\mu\lambda}=2t_{\nu+k}(0)(1+k/\nu)^{3/2}\Sigma^{-1/2}\mathrm{E}\Big[\frac{ZZ'}{(1+||Z||^2/\nu)^{3/2}}\Big]\Sigma^{1/2}\sigma^{-1}=(\Gamma^{\rm ST}_{g_{k,\nu};\lambda\mu})'.$$

Since $\Gamma_{g_{k,\nu};\mu\mu}$ is invertible, we have (as in the proof of Theorem 2.1) that the (μ,λ)-submatrix of $\Gamma^{\rm ST}_{g_{k,\nu}}$ is non-singular iff $\Gamma^{\rm ST}_{g_{k,\nu};\lambda\lambda.\mu}:=\Gamma^{\rm ST}_{g_{k,\nu};\lambda\lambda}-\Gamma^{\rm ST}_{g_{k,\nu};\lambda\mu}\Gamma^{-1}_{g_{k,\nu};\mu\mu}\Gamma^{\rm ST}_{g_{k,\nu};\mu\lambda}$ is non-singular.

Now, letting

$$A_{k,\nu}:=\mathrm{E}\Big[\frac{ZZ'}{(1+||Z||^2/\nu)}\Big]-\mathrm{E}\Big[\frac{ZZ'}{(1+||Z||^2/\nu)^{3/2}}\Big]\Big(\mathrm{E}\Big[\frac{ZZ'}{(1+||Z||^2/\nu)^2}\Big]\Big)^{-1}\mathrm{E}\Big[\frac{ZZ'}{(1+||Z||^2/\nu)^{3/2}}\Big],$$

and

$$b^{(r)}_{k,\nu}:=\mathrm{E}\Big[\frac{Z_1^2}{(1+||Z||^2/\nu)^r}\Big],$$

simple algebra and the fact that Z has a spherically symmetric distribution yield

$$\begin{aligned}\Gamma^{\rm ST}_{g_{k,\nu};\lambda\lambda.\mu} &= 4(t_{\nu+k}(0))^2(1+k/\nu)\sigma^{-1}\Sigma^{1/2}A_{k,\nu}\Sigma^{1/2}\sigma^{-1}\\ &= (4/b^{(2)}_{k,\nu})\Big(b^{(1)}_{k,\nu}b^{(2)}_{k,\nu}-(b^{(3/2)}_{k,\nu})^2\Big)(t_{\nu+k}(0))^2(1+k/\nu)\sigma^{-1}\Sigma\sigma^{-1}.\end{aligned}$$

The Cauchy-Schwarz inequality and the fact that $\|Z\|$ is an absolutely continuous random variable entail that $b^{(1)}_{k,\nu}b^{(2)}_{k,\nu} - (b^{(3/2)}_{k,\nu})^2 > 0$. This allows to conclude since the invertibility of σ and Σ then guarantees that $\Gamma^{\rm ST}_{g_{k,\nu};\lambda\lambda.\mu}$ is non-singular. □

Since, as $\nu \to \infty$, the pdf $f^{\rm ST}_{\mu,\Sigma,\lambda,\nu}(x)$ in (4.8) tends to the pdf of a skew-multinormal density, the resulting Fisher information matrix at $\lambda = 0$ becomes singular. In this sense, Theorem 4.1 requires the fixed value of ν to be finite.

Now, one might argue that, nice as it is, the non-singularity result of Theorem 4.1 is obtained only for fixed values of ν, and that one should also investigate a possible singularity of Fisher information matrices (still, in the vicinity of symmetry) in the class of multivariate skew-t densities indexed by μ, Σ, λ, *and* ν. It is indeed unrealistic to assume that the tail weight parameter ν is known in practice, so that ν should enter score functions and Fisher information matrices when performing inference for such distributions. The rest of this section therefore deals with the ν-unspecified case.

The parameter ν enters the score function $\ell^{\rm ST}_{g_{k,\nu};\mu,\Sigma,\lambda}(x)$ through a further component given by (in the vicinity of symmetry)

$$d^{\rm ST}_{g_{k,\nu};\nu}(x) := c_v - \frac{\log(1 + ||\Sigma^{-1/2}(x-\mu)||^2/\nu)}{2} + \frac{(\nu+k)}{2\nu^2}\frac{||\Sigma^{-1/2}(x-\mu)||^2}{(1 + ||\Sigma^{-1/2}(x-\mu)||^2/\nu)}, \quad (4.10)$$

where $c_v := \frac{d}{d\nu}\log\left(\frac{\Gamma((\nu+k)/2)}{(\pi\nu)^{k/2}\Gamma(\nu/2)}\right)$, which gives rise to a Fisher information matrix of the form

$$\Gamma^{\rm ST} = \begin{pmatrix} \Gamma_{g_{k,\nu};\mu\mu} & 0 & \Gamma^{\rm ST}_{g_{k,\nu};\mu\lambda} & 0 \\ 0 & \Gamma_{g_{k,\nu};\Sigma\Sigma} & 0 & \Gamma^{\rm ST}_{g_{k,\nu};\Sigma\nu} \\ \Gamma^{\rm ST}_{g_{k,\nu};\lambda\mu} & 0 & \Gamma^{\rm ST}_{g_{k,\nu};\lambda\lambda} & 0 \\ 0 & \Gamma^{\rm ST}_{g_{k,\nu};\nu\Sigma} & 0 & \Gamma^{\rm ST}_{g_{k,\nu};\nu\nu} \end{pmatrix};$$

again, the extra zero blocks in $\Gamma^{\rm ST}$ follow from the fact that the ν-score in (4.10) is symmetric with respect to $x - \mu$.

Theorem 4.2 *Fix an arbitrary positive integer k. Then, at any parameter value $(\mu', (\mathrm{vech}\,\Sigma)', 0', \nu)' \in \mathbb{R}^k \times \mathcal{S}^k \times \{0\} \times (2, \infty)$, the Fisher information matrix associated with the class of multivariate skew-t densities in (4.8), namely $\Gamma^{\rm ST}$, is non-singular.*

Proof of Theorem 4.2. Clearly, it is sufficient to show that the (μ, λ)- and (Σ, ν)-submatrices of $\Gamma^{\rm ST}$ are non-singular. Since the (μ, λ)-submatrix was shown to be non-singular in the proof of Theorem 4.1, we may focus on the (Σ, ν)-submatrix. As already stated in the proof of Theorem 4.1, $\Gamma_{g_{k,\nu};\Sigma\Sigma}$ is invertible, so that the (Σ, ν)-submatrix is non-singular iff $\Gamma^{\rm ST}_{g_{k,\nu};\nu\nu.\Sigma} := \Gamma^{\rm ST}_{g_{k,\nu};\nu\nu} - \Gamma^{\rm ST}_{g_{k,\nu};\nu\Sigma}\Gamma^{-1}_{g_{k,\nu};\Sigma\Sigma}\Gamma^{\rm ST}_{g_{k,\nu};\Sigma\nu}$ is non-singular, that is, iff the latter is non-zero.

Consider the random variable $Y := d^{\rm ST}_{g_{k,\nu};\nu}(X) - \Gamma^{\rm ST}_{g_{k,\nu};\nu\Sigma}\Gamma^{-1}_{g_{k,\nu};\Sigma\Sigma}d_{g_{k,\nu};\Sigma}(X)$, where X has pdf $f^{\rm ST}_{\mu,\Sigma,0,\nu}$. Clearly, the expectation and variance of Y are equal to 0 and $\Gamma^{\rm ST}_{g_{k,\nu};\nu\nu.\Sigma}$, respectively. If $\Gamma^{\rm ST}_{g_{k,\nu};\nu\nu.\Sigma} = 0$, we must therefore have that

$$d^{\rm ST}_{g_{k,\nu};\nu}(x) - \Gamma^{\rm ST}_{g_{k,\nu};\nu\Sigma}\Gamma^{-1}_{g_{k,\nu};\Sigma\Sigma}d_{g_{k,\nu};\Sigma}(x) = 0 \quad \text{a.e. over } \mathbb{R}^k.$$

This, however, cannot hold, since substituting, in $d_{g_{k,\nu};\Sigma}(x)$ and $d^{\mathrm{ST}}_{g_{k,\nu};\nu}(x)$, $\Sigma^{-1/2}(x-\mu)$ with $s\Sigma^{-1/2}(x-\mu)$ ($s \in \mathbb{R}$) and letting s go to infinity indeed reveals that $d_{g_{k,\nu};\Sigma}(x)$, unlike $d^{\mathrm{ST}}_{g_{k,\nu};\nu}(x)$, is a bounded function of s. Therefore, we must have that $\Gamma^{\mathrm{ST}}_{g_{k,\nu};\nu\nu.\Sigma} > 0$, which establishes the result. □

Theorems 4.1 and 4.2 prove that the conjecture of Azzalini and Genton (2008) is correct. Finally, note that the proof of Theorem 4.2 did not require an exact determination of the Fisher information matrix Γ^{ST}, which is known to be difficult (see Azzalini and Capitanio 2003).

5 Implications for inference.

We have shown in the previous sections that, for the class of generalized skew-elliptical densities of type (2.3) and for the class of multivariate skew-t distributions, singular Fisher information matrices and systematic stationary points in the profile log-likelihood function for skewness, both in the vicinity of symmetry, only occur for skew-(multi)normal distributions (here, we of course consider that the skew-normal distributions are embedded in the class of skew-t distributions, that is, we allow the number of degrees of freedom to be infinite in that class). We now provide a short overview of the implications of our findings on inference.

Azzalini (1985) proposed an alternative parameterization, the so-called centred parameterization, in order to get rid of the singularities of the univariate skew-normal model. Arellano-Valle and Azzalini (2008) adopted the same approach for the multivariate skew-normal densities. Our results strongly alleviate the necessity of such a rather cumbersome reparameterization for broad classes of skewed densities which do not contain the skew-normal ones, which appears to be a welcomed feature (see Azzalini and Genton 2008, Section 1.2). Moreover, for distributions other than the skew-normal, the more regular behavior of the profile log-likelihood function for skewness should remove estimation problems encountered in the neighborhood of $\lambda = 0$. Non-singular Fisher information matrices also allow for using the standard asymptotic theory of MLEs, hence avoid further mathematical complications.

As for hypothesis testing, the singularities related to the skew-normal densities imply that the Le Cam optimal tests for symmetry about an unspecified centre against skew-normal alternatives coincide with the trivial test, that is, the test discarding the observations and rejecting the null of symmetry at level α whenever an auxiliary Bernoulli variable with parameter α takes value one. It can be shown that non-singular information matrices protect from such "worst" Le Cam optimal tests for symmetry (see Ley and Paindaveine 2009 for more details); our findings therefore encourage the construction of Le Cam optimal tests for symmetry against *non-Gaussian* generalized skew-elliptical alternatives.

Further inferential aspects related to singular information matrices can be found in Rotnitzky et al. (2000) or Bottai (2003).

References

ARELLANO-VALLE, R.B. and AZZALINI, A. (2008) The centred parametrization for the multivariate skew-normal distribution, *Journal of Multivariate Analysis*, 99, 1362–1382.

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

AZZALINI, A. (2005) The skew-normal distribution and related multivariate families (with discussion), *Scandinavian Journal of Statistics*, 32, 159–188.

AZZALINI, A. and CAPITANIO, A. (1999) Statistical applications of the multivariate skew-normal distributions, *Journal of the Royal Statistical Society B*, 61, 579–602.

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

AZZALINI, A. and DALLA VALLE, A. (1996) The multivariate skew-normal distribution, *Biometrika*, 83, 715–726.

AZZALINI, A. and GENTON, M.G. (2007) On Gauss's characterization of the normal distribution, *Bernoulli*, 13, 169–174.

AZZALINI, A. and GENTON, M.G. (2008) Robust likelihood methods based on the skew-t and related distributions, *International Statistical Review*, 76, 106–129.

BARNDORFF-NIELSEN, O.E. and COX, D.R. (1994) *Inference and asymptotics*, Chapman and Hall. London.

BOTTAI, M. (2003) Confidence regions when the Fisher information is zero, *Biometrika*, 90, 73–84.

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

CHIOGNA, M. (2005) A note on the asymptotic distribution of the maximum likelihood estimator for the scalar skew-normal distribution, *Statistical Methods and Applications*, 14, 331–341.

DiCICCIO, T.J. and MONTI, A.C. (2004) Inferential aspects of the skew exponential power distribution, *Journal of the American Statistical Association*, 99, 439–450.

DiCICCIO, T.J. and MONTI, A.C. (2009) Inferential aspects of the skew t-distribution, *Manuscript in preparation.*

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

HALLIN, M. and PAINDAVEINE, D. (2006) Semiparametrically efficient rank-based inference for shape. I. Optimal rank-based tests for sphericity, *Annals of Statistics*, 34, 2707–2756.

GOMEZ, H.W., VENEGAS, O. and BOLFARINE, H. (2007) Skew-symmetric distributions generated by the distribution function of the normal distribution, *Environmetrics*, 18, 395–407.

LEY, C. and PAINDAVEINE, D. (2009) On the Singularity of Multivariate Skew-Symmetric Models, *Journal of Multivariate Analysis*, to appear.

PAINDAVEINE, D. (2008) A canonical definition of shape, *Statistics and Probability Letters*, 78, 2240–2247.

PATEFIELD, W.M. (1977) On the maximized likelihood function, *Sankhyā Series B*, 39, 92–96.

PEWSEY, A. (2000) Problems of inference for Azzalini's skew-normal distribution, *Journal of Applied Statistics*, 27, 859–870.

ROTNITZKY, A., COX, D.R., BOTTAI, M., and ROBERTS, J. (2000) Likelihood-based inference with singular information matrix, *Bernoulli*, 6, 243–284.

WANG, J., BOYER, J. and GENTON, M.G. (2004) A skew-symmetric representation of multivariate distribution, *Statistica Sinica*, 14, 1259–1270.