Bernoulli **18**(3), 2012, 747–763
DOI: 10.3150/12-BEJ346

Skew-symmetric distributions and Fisher information – a tale of two densities

MARC HALLIN[1] and CHRISTOPHE LEY[1,2]

[1]*E.C.A.R.E.S., CP 114, Université Libre de Bruxelles, 50 Avenue F.D. Roosevelt, 1050 Brussels, Belgium. E-mail: mhallin@ulb.ac.be*

[2]*Département de Mathématique, CP 210, Université Libre de Bruxelles, Boulevard du Triomphe, 1050 Brussels, Belgium. E-mail: chrisley@ulb.ac.be*

Skew-symmetric densities recently received much attention in the literature, giving rise to increasingly general families of univariate and multivariate skewed densities. Most of those families, however, suffer from the inferential drawback of a potentially singular Fisher information in the vicinity of symmetry. All existing results indicate that Gaussian densities (possibly after restriction to some linear subspace) play a special and somewhat intriguing role in that context. We dispel that widespread opinion by providing a full characterization, in a general multivariate context, of the information singularity phenomenon, highlighting its relation to a possible link between symmetric kernels and skewing functions – a link that can be interpreted as the mismatch of two densities.

Keywords: singular Fisher information; skew-normal distributions; skew-symmetric distributions; skewing function; symmetric kernel

1. Introduction

Models for skewed distributions have become increasingly popular in recent years, as they provide a much better fit for data presenting some departure from normality, and from symmetry in general. Many of the proposed models in the literature allow for a continuous variation from symmetry to asymmetry, regulated by some finite-dimensional parameter.

The success of those skewed distributions started with the seminal papers by Azzalini [3,4] introducing the scalar *skew-normal* model, which embeds the univariate normal distributions into a flexible parametric class of (possibly) skewed distributions. More 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 $\delta \in \mathbb{R}$ if it admits the probability density function (p.d.f.)

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

where ϕ and Φ respectively denote the p.d.f. and cumulative distribution function (c.d.f.) of a standard normal distribution. Besides their many appealing features, however, skew-normal densities unfortunately also suffer from an unpleasant inferential drawback: in the vicinity of symmetry, that is, at $\delta = 0$, the Fisher information matrix for the three-parameter density (1.1) is singular – typically, with rank 2 instead of 3. Consequently, skew-normal distributions happen to be problematic from an inferential point of view, since that singularity violates the assumptions for standard Gaussian asymptotics and precludes, at first sight, any nontrivial test of the null

1350-7265

hypothesis of symmetry. Such a situation has been studied by Rotnitzky *et al.* [21], who show that one of the parameters then cannot be estimated at the usual root-n rate, while the limit distribution of maximum likelihood estimators might be bimodal.

This Fisher singularity problem, however, did not hamper the success of skew-normal densities among practitioners, while theoretical extensions were developing into various directions. Azzalini and Dalla Valle [8] and Azzalini and Capitanio [6] consider multivariate skew-normal distributions resulting from replacing in (1.1) the univariate normal kernel ϕ with its k-variate version ϕ_k. In the same paper, Azzalini and Capitanio also propose substituting an elliptical kernel f_k for the normal one ϕ_k, and replacing the skewing factor Φ in (1.1) with an arbitrary, possibly non-Gaussian, univariate symmetric c.d.f. G_1. The resulting distributions are called *skew-elliptical*. The class of skew-elliptical distributions is also studied in detail by Branco and Dey [10], based, however, on a slightly different definition. Genton and Loperfido [15] introduce a concept of *generalized skew-elliptical* distributions encompassing all previous ones, where arbitrary skewing functions (not necessarily c.d.f.s, but satisfying c.d.f.-type conditions) can be used in conjunction with the elliptical kernel f_k. Finally, Azzalini and Capitanio [7] (who also propose the nowadays commonly adopted definition of multivariate skew-t distributions), then Wang *et al.* [24] are relaxing the assumption of elliptically symmetric kernels into a weaker assumption of central symmetry, leading to multivariate *skew-symmetric* densities of the form

$$\begin{aligned}\mathbf{x} \mapsto f_{\boldsymbol{\vartheta}}^{\Pi}(\mathbf{x}) &= f_{\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\delta}}^{\Pi}(\mathbf{x}) \\ &:= 2|\boldsymbol{\Sigma}|^{-1/2} f\big(\boldsymbol{\Sigma}^{-1/2}(\mathbf{x}-\boldsymbol{\mu})\big)\Pi\big(\boldsymbol{\Sigma}^{-1/2}(\mathbf{x}-\boldsymbol{\mu}),\boldsymbol{\delta}\big), \qquad \mathbf{x}\in\mathbb{R}^k,\end{aligned} \tag{1.2}$$

where

(a) $\boldsymbol{\mu} \in \mathbb{R}^k$ is a location parameter, $\boldsymbol{\Sigma} \in \mathcal{S}_k$ (throughout, $|\mathbf{M}|$ denotes the determinant and $\mathbf{M}^{1/2}$ the symmetric square-root of any $\mathbf{M}$ in the class $\mathcal{S}_k$ of symmetric positive definite $k \times k$ matrices) a scatter matrix, while $\boldsymbol{\delta} \in \mathbb{R}^k$ plays the role of a skewness parameter;
(b) the *symmetric kernel* f is a centrally symmetric nonvanishing p.d.f., meaning that $0 \neq f(-\mathbf{z}) = f(\mathbf{z})$, $\mathbf{z} \in \mathbb{R}^k$, and
(c) the *skewing function* $\Pi : \mathbb{R}^k \times \mathbb{R}^k \to [0,1]$ satisfies $\Pi(-\mathbf{z},\boldsymbol{\delta}) + \Pi(\mathbf{z},\boldsymbol{\delta}) = 1$, $\mathbf{z},\boldsymbol{\delta} \in \mathbb{R}^k$, and $\Pi(\mathbf{z},\mathbf{0}) = 1/2$, $\mathbf{z} \in \mathbb{R}^k$.

This definition is the one we are adopting in the sequel. While $\Pi(\mathbf{z},\boldsymbol{\delta})$, in most practical situations, is of the simple form $\Pi(\boldsymbol{\delta}'\mathbf{z})$, with $\Pi : \mathbb{R} \to [0,1]$, Wang *et al.* [24] actually do not consider any specific $\boldsymbol{\delta}$-parameterization. Our parametric approach (with the regularity assumptions (A2)–(A2$^+$) and (B2)–(B2$^+$) of Sections 2.1 and 3.1, resp.) is in the spirit of – if not at the same level of mathematical generality as – the differentiable path and tangent space approach taken in the local and asymptotic treatment of semiparametric models (see, e.g., Chapter 25 of van der Vaart [23]). Also, the condition that f is a nonvanishing density is not imposed by Wang *et al.* [24]; we are adding that requirement in order to avoid inessential complications related with bounded and parameter-dependent supports. For further information about skew-symmetric models and related topics, we refer the reader to the recent monograph by Genton [14], and to the review papers Arnold and Beaver [2] and Azzalini [5].

The issue of singular Fisher information runs like a red thread through all those developments. Mentioned, from the very beginning, in Azzalini [3] itself, it is discussed, in the univariate and multivariate skew-normal context, by Azzalini and Capitanio [6], Pewsey [19], Chiogna [11] and Arellano-Valle and Azzalini [1]. The same issue has been considered in various subclasses of skew-symmetric distributions. Pewsey [20] and Azzalini and Genton [9] establish that the singularity problem remains after replacement of the c.d.f. Φ in (1.1) with any c.d.f. H satisfying mild regularity assumptions. DiCiccio and Monti [12] prove that, within the class of univariate skew-exponential power distributions of Azzalini [4], the normal kernels are the only ones suffering from singular Fisher information. The same result is shown to hold true for two classes of scalar skew-t distributions by Gómez *et al.* [16] and DiCiccio and Monti [13]. The multivariate counterparts of these statements are provided in Ley and Paindaveine [17,18], respectively.

Finally, the very general (still a special case of (1.2), though) class of multivariate skew-symmetric densities of the form

$$\mathbf{x} \mapsto 2|\mathbf{\Sigma}|^{-1/2} f\big(\mathbf{\Sigma}^{-1/2}(\mathbf{x}-\boldsymbol{\mu})\big)\Pi\big(\boldsymbol{\delta}'\mathbf{\Sigma}^{-1/2}(\mathbf{x}-\boldsymbol{\mu})\big), \qquad \mathbf{x} \in \mathbb{R}^k, \tag{1.3}$$

encompassing all previous cases, is considered in Ley and Paindaveine [17], who characterize, for each possible value $1 \le m \le k$ of the Fisher information rank deficiency, the form of the symmetric kernels giving rise to such deficiency. Here again, Gaussian kernels are playing a very special role. In the univariate setup and within the subclass of multivariate generalized skew-elliptical distributions, only the skew-normal densities are affected by the singularity problem. Although results in the fully general (for densities of the form (1.3)) multivariate case are more complex, only kernels exhibiting Gaussian restrictions on some m-dimensional linear subspaces can lead to degenerate Fisher information.

A tentative remedy to that singularity problem was suggested by Azzalini himself who, as early as 1985, in his original paper, proposes a reparametrization of skew-normal families, the so-called *centered parametrization*, under which Fisher information matrices remain full-rank. The multivariate version of that reparametrization is examined in detail by Arellano-Valle and Azzalini [1]. That solution, however, never really caught up in practice, partly because the structure of the skewing mechanism, hence of the resulting skew-normal family, under the new parametrization, loses much of its simplicity (certainly so in the multivariate context), partly because of its limitation to skew-normal families. Azzalini and Genton [9] therefore once again emphasize the need for a clarification of the Fisher singularity phenomenon in order to "*remove, or at least alleviate, the necessity of an alternative parametrization.*"

The objective of the present paper is to provide such a clarification. While all comments and existing results, in this singular Fisher information issue, seemed to be pointing at some special status for normal kernels and, consequently, skew-normal distributions, we completely dispel the idea of any particular role of Gaussian kernels. Turning to the fully general class of skew-symmetric densities described in (1.2), we show indeed that information deficiency actually originates in an unfortunate mismatch between f and Π – more specifically, between two densities, the kernel f and an exponential density g_Π associated with the skewing function Π.

A tale of two densities, thus, rather than a Gaussian mystery…

The paper is organized as follows. Section 2.1 deals with the univariate setup, where the singularity problem is simple, as the rank of the three-parameter Fisher information matrix only can be 3 or 2. The result is derived in an informal way, and some examples of skewing functions are treated in Section 2.2. A more formal statement of the general solution is provided for the multivariate setup in Section 3.1, along with some examples in Section 3.2. Final comments and conclusions are given in Section 4.

2. The univariate setup

2.1. A tale of two densities ...

We start by analyzing the information singularity problem in the univariate case. To do so, consider the class of skew-symmetric probability density families of the form

$$x \mapsto f_{\boldsymbol{\vartheta}}^{\Pi}(x) = f_{\mu,\sigma,\delta}^{\Pi}(x) := 2\sigma^{-1} f\big(\sigma^{-1}(x-\mu)\big)\Pi\big(\sigma^{-1}(x-\mu),\delta\big), \qquad x \in \mathbb{R}, \tag{2.1}$$

with $\boldsymbol{\vartheta} := (\mu, \sigma, \delta)'$, where $\mu \in \mathbb{R}$ is a location parameter, $\sigma \in \mathbb{R}_0^+$ a scale parameter and $\delta \in \mathbb{R}$ an asymmetry parameter.

The symmetric kernel $f : \mathbb{R} \to \mathbb{R}^+$ in (2.1) is a nonvanishing symmetric *standardized* p.d.f., that is, a probability density function such that $f(z) = f(-z) \neq 0$ for all $z \in \mathbb{R}$, with scale parameter one – an identification constraint for σ that does not imply any loss of generality. Classical standardization, with a constraint of the form $\int_{-\infty}^{\infty} z^2 f(z)\,\mathrm{d}z = 1$, involves the variance of Z with p.d.f. f; the scale parameter σ^2 then is the mean squared deviation $\mathrm{E}[(X-\mu)^2]$ with respect to μ of X with p.d.f. $f_{\mu,\sigma,0}^{\Pi}$. If moment assumptions are to be avoided, one may rather consider, for instance, medians of squares, with an identification constraint of the form $\int_{-\infty}^{1} f(z)\,\mathrm{d}z = 0.75$: if X has p.d.f. $f_{\mu,\sigma,0}^{\Pi}$, σ then is the median of the absolute deviation $|X-\mu|$, which exists irrespective of the density of X. Other quantiles of $|X-\mu|$ would enjoy similar properties. We throughout assume that such an identification constraint, hence a concept of scale, has been adopted. That choice, however, is completely arbitrary, and any element in the scale family of p.d.f.'s of the form (2.1) with $\mu = \delta = 0$ could be chosen as the reference density characterizing unit scale – hence could serve as a symmetric kernel for the same skew-symmetric family. As we shall see, that choice has no impact on the results of this paper.

The second factor in (2.1) is a skewing function, namely, a function $\Pi : \mathbb{R} \times \mathbb{R} \to [0,1]$ such that $\Pi(-z,\delta) + \Pi(z,\delta) = 1$ for all $z, \delta \in \mathbb{R}$, and $\Pi(z,0) = 1/2$ for all $z \in \mathbb{R}$. Traditional choices involve $\Pi(z,\delta) = \Phi(\delta z)$ (skew-normal distributions, Azzalini [3]), $\Pi(z,\delta) = \Phi(\delta\,\mathrm{sign}(z)|z|^{\alpha/2}(2/\alpha)^{1/2})$ (skew-exponential power distributions, Azzalini [4]) or $\Pi(z,\delta) = G(\delta z)$ for any symmetric univariate c.d.f. G (skew-symmetric distributions, Azzalini and Capitanio [6]). The class of skewing functions considered here is much broader.

The regularity assumptions we are making on f and Π are as follows.

Assumption (A1). *The mapping $z \mapsto f(z)$ is differentiable, with derivative $\dot{f}$ such that, letting $\varphi_f := -\dot{f}/f$, the information quantity for location $\sigma^{-2}\mathcal{I}_f$, with*

$$\mathcal{I}_f := \int_{-\infty}^{\infty} \varphi_f^2(z) f(z)\,\mathrm{d}z,$$

is finite.

Assumption (A1$^+$). *Same as* (A1), *but the information quantity for scale* $\sigma^{-2}\mathcal{J}_f$, *with*

$$\mathcal{J}_f := \int_{-\infty}^{\infty} \big(z\varphi_f(z) - 1\big)^2 f(z)\,\mathrm{d}z,$$

moreover is finite.

Assumption (A2). (i) *The mapping* $z \mapsto \Pi(z,\delta)$ *is differentiable, and its derivative equals* 0 *at* $\delta = 0$. (ii) *The mapping* $\delta \mapsto \Pi(z,\delta)$ *is differentiable at* $\delta = 0$ *for all* $z \in \mathbb{R}$, *with derivative* (*at* $\delta = 0$) $\partial_\delta \Pi(z,\delta)|_{\delta=0} =: \psi(z)$ *such that* $z \mapsto \psi(z)$ *admits a primitive, denoted as* Ψ.

Assumption (A2$^+$). *Same as* (A2), *but the quantity*

$$\int_{-\infty}^{\infty} \psi^2(z) f(z)\,\mathrm{d}z$$

moreover is finite.

These assumptions essentially guarantee the existence and finiteness of Fisher information at $\delta = 0$; the differentiability and integrability conditions could be relaxed into weaker differentiability properties such as quadratic mean differentiability. This small gain of generality, however, would require a generalized definition of information (in the Le Cam style), with non-negligible technical complications. For the sake of simplicity, we stick to a more traditional approach and the traditional definition of Fisher information. Note that this definition differs from the one, used by some authors, of an *observed Fisher information*, that is, the empirical value of the matrix of negative second-order derivatives of the log-likelihood evaluated at the maximum likelihood estimator of the parameters.

Under Assumptions (A1) and (A2), the *score vector* $\boldsymbol{\ell}_{f;\boldsymbol{\vartheta}}$, at $(\mu,\sigma,0)' =: \boldsymbol{\vartheta}_0$, takes the form

$$\boldsymbol{\ell}_{f;\boldsymbol{\vartheta}_0}(x) := \mathrm{grad}_{\boldsymbol{\vartheta}} \log f_{\boldsymbol{\vartheta}}^{\Pi}(x)|_{\boldsymbol{\vartheta}_0} =: \big(\ell^1_{f;\boldsymbol{\vartheta}_0}(x), \ell^2_{f;\boldsymbol{\vartheta}_0}(x), \ell^3_{f;\boldsymbol{\vartheta}_0}(x)\big)'$$
$$= \begin{pmatrix} \sigma^{-1}\varphi_f\big(\sigma^{-1}(x-\mu)\big) \\ \sigma^{-1}\big(\sigma^{-1}(x-\mu)\varphi_f\big(\sigma^{-1}(x-\mu)\big) - 1\big) \\ 2\psi\big(\sigma^{-1}(x-\mu)\big) \end{pmatrix},$$

where the factor 2 in $\ell^3_{f;\boldsymbol{\vartheta}_0}$ follows from the fact that $\Pi(z,0) = 1/2$ for all $z \in \mathbb{R}$. Assumption (A2)(i) is a mild requirement which, in regular models, readily follows from the fact that $\Pi(z,0) = 1/2$, and ensures that the skewing function Π plays no role in the score functions for μ and σ at $\delta = 0$.

Under Assumptions (A1$^+$) and (A2$^+$), the 3×3 Fisher information matrix for (μ,σ,δ) exists, and takes the form

$$\boldsymbol{\Gamma}_{f;\boldsymbol{\vartheta}_0} := \sigma^{-1} \int_{-\infty}^{\infty} \boldsymbol{\ell}_{f;\boldsymbol{\vartheta}_0}(x)\boldsymbol{\ell}'_{f;\boldsymbol{\vartheta}_0}(x) f\big(\sigma^{-1}(x-\mu)\big)\,\mathrm{d}x =: \begin{pmatrix} \gamma^{11}_{f;\boldsymbol{\vartheta}_0} & 0 & \gamma^{13}_{f;\boldsymbol{\vartheta}_0} \\ 0 & \gamma^{22}_{f;\boldsymbol{\vartheta}_0} & 0 \\ \gamma^{13}_{f;\boldsymbol{\vartheta}_0} & 0 & \gamma^{33}_{f;\boldsymbol{\vartheta}_0} \end{pmatrix},$$

with

$$\gamma^{11}_{f;\boldsymbol{\vartheta}_0} = \sigma^{-2}\mathcal{I}_f, \qquad \gamma^{22}_{f;\boldsymbol{\vartheta}_0} = \sigma^{-2}\mathcal{J}_f, \qquad \gamma^{33}_{f;\boldsymbol{\vartheta}_0} = 4\int_{-\infty}^{\infty} \psi^2(z) f(z)\,\mathrm{d}z$$

and

$$\gamma^{13}_{f;\boldsymbol{\vartheta}_0} = 2\sigma^{-1}\int_{-\infty}^{\infty} \varphi_f(z)\psi(z) f(z)\,\mathrm{d}z.$$

The zeroes in $\boldsymbol{\Gamma}_{f;\boldsymbol{\vartheta}_0}$ are easily obtained by noting that $\ell^1_{f;\boldsymbol{\vartheta}_0}$ and $\ell^3_{f;\boldsymbol{\vartheta}_0}$ are antisymmetric functions of $(x-\mu)$, whereas $\ell^2_{f;\boldsymbol{\vartheta}_0}$ is symmetric with respect to the same quantity.

It then trivially follows that singularity of $\boldsymbol{\Gamma}_{f;\boldsymbol{\vartheta}_0}$ only can be due to the singularity of the 2×2 submatrix

$$\boldsymbol{\Gamma}^0_{f;\boldsymbol{\vartheta}_0} := \begin{pmatrix} \gamma^{11}_{f;\boldsymbol{\vartheta}_0} & \gamma^{13}_{f;\boldsymbol{\vartheta}_0} \\ \gamma^{13}_{f;\boldsymbol{\vartheta}_0} & \gamma^{33}_{f;\boldsymbol{\vartheta}_0} \end{pmatrix},$$

the existence of which, however, only requires Assumptions (A1) and (A2$^+$). Clearly, either $\boldsymbol{\Gamma}^0_{f;\boldsymbol{\vartheta}_0}$ is full-rank or, in case $\gamma^{11}_{f;\boldsymbol{\vartheta}_0}\gamma^{33}_{f;\boldsymbol{\vartheta}_0} = (\gamma^{13}_{f;\boldsymbol{\vartheta}_0})^2$, it has rank 1.

Now, the Cauchy–Schwarz inequality implies that $(\gamma^{13}_{f;\boldsymbol{\vartheta}_0})^2 \le \gamma^{11}_{f;\boldsymbol{\vartheta}_0}\gamma^{33}_{f;\boldsymbol{\vartheta}_0}$, with equality if and only if

$$\varphi_f = a\psi \qquad f\text{-a.s. (equivalently, Lebesgue-a.e.)} \tag{2.2}$$

for some constant $a \in \mathbb{R}$. It thus follows that $\boldsymbol{\Gamma}^0_{f;\boldsymbol{\vartheta}_0}$ is singular for any $\boldsymbol{\vartheta}_0 = (\mu, \sigma, 0)'$ if and only if (2.2) is satisfied for some $a \in \mathbb{R}$. This holds under Assumptions (A1) and (A2$^+$). If Assumption (A1) is reinforced into (A1$^+$), the 2×2 singularity of $\boldsymbol{\Gamma}^0_{f;\boldsymbol{\vartheta}_0}$ in turn is equivalent to the 3×3 singularity of $\boldsymbol{\Gamma}_{f;\boldsymbol{\vartheta}_0}$. Replacing φ_f with its definition, the necessary and sufficient condition $\varphi_f = a\psi$ yields a first-order differential equation whose solutions are of the form $f(x) = c\exp(-a\Psi(x))$ for some $a \in \mathbb{R}$, where Ψ is a primitive of ψ and $c \in \mathbb{R}^+$ an integration constant.

Summing up, let the couple (f, Π) satisfy Assumptions (A1$^+$) and (A2$^+$): $\boldsymbol{\Gamma}_{f;\boldsymbol{\vartheta}_0}$ is singular for all $\boldsymbol{\vartheta}_0$ if and only if the symmetric kernel f belongs to the *exponential family*

$$\mathcal{E}_\Psi := \left\{ g_a := \exp(-a\Psi) \Big/ \int_{-\infty}^{\infty} \exp(-a\Psi(z))\,\mathrm{d}z \;\middle|\; a \in \mathcal{A} \right\} \tag{2.3}$$

with *minimal sufficient statistic* Ψ, *natural parameter* $-a$, and *natural parameter space*

$$\mathcal{A} := \left\{ a \in \mathbb{R} \text{ such that } \int_{-\infty}^{\infty} \exp(-a\Psi(z))\,\mathrm{d}z < \infty \right\}.$$

The same statement can be made under Assumptions (A1) and (A2$^+$) about the singularity of $\boldsymbol{\Gamma}^0_{f;\boldsymbol{\vartheta}_0}$.

Note that $\mathcal{A}$, as the natural parameter space of an exponential family, is an open interval of $\mathbb{R}$. The unique value a_Π of $a \in \mathcal{A}$ such that f and g_{a_Π} coincide, if any, is entirely determined by the standardization constraint on f. If the classical variance-based standardization is adopted, then a_Π is solution of the equation

$$\int_{-\infty}^{\infty} z^2 \exp(-a\Psi(z))\,\mathrm{d}z = \int_{-\infty}^{\infty} \exp(-a\Psi(z))\,\mathrm{d}z.$$

If standardization is imposed via medians of squares, a_Π is solution of

$$\int_{-\infty}^{1} \exp(-a\Psi(z))\,\mathrm{d}z = 3\int_{1}^{\infty} \exp(-a\Psi(z))\,\mathrm{d}z.$$

Letting $f_\sigma(x) := \sigma^{-1} f(x/\sigma)$, $\sigma \in \mathbb{R}_0^+$, also note that $f \in \mathcal{E}_\Psi$ if and only if $f_\sigma \in \mathcal{E}_{\Psi\circ\sigma^{-1}}$, where $\mathcal{E}_{\Psi\circ\sigma^{-1}}$ stands for the exponential family with minimal sufficient statistic $\Psi\circ\sigma^{-1} : z \mapsto \Psi(\sigma^{-1}z)$. It is easy to see that both conditions moreover determine the same a_Π, which confirms that the arbitrary choice of a scale parameter has no impact on the result.

As a consequence of those results, it follows that, for any symmetric density f satisfying Assumption (A1$^+$) (resp., Assumption (A1)), there exists a skewing function Π_f (infinitely many of them, actually) such that $\mathbf{\Gamma}_{f;\boldsymbol{\vartheta}_0}$ (resp., $\mathbf{\Gamma}^0_{f;\boldsymbol{\vartheta}_0}$) exists and is singular for any $\boldsymbol{\vartheta}_0$; among them, with $a_\pi = \sqrt{2\pi}$, $\Pi_f(z,\delta) := \Phi(\delta\varphi_f(z))$, for which Assumption (A2$^+$) holds.

The converse is slightly more subtle. Let Π be a skewing function satisfying Assumption (A2); a function Ψ with derivative ψ thus exists, which automatically satisfies $\Psi(z) = \Psi(-z)$. If there exists a density g_a in the corresponding exponential family (2.3) such that $\int_{-\infty}^{\infty} \psi^2(z) g_a(z)\,\mathrm{d}z$ is finite, then the skew-symmetric family with symmetric kernel $f = g_a$ and skewing function Π is such that Assumptions (A1) and (A2$^+$) hold, and the corresponding 2×2 matrix $\mathbf{\Gamma}^0_{f;\boldsymbol{\vartheta}_0}$ exists and is singular for any $\boldsymbol{\vartheta}_0$. If moreover $f = g_a$ also satisfies Assumption (A1$^+$), then the 3×3 information matrix $\mathbf{\Gamma}_{f;\boldsymbol{\vartheta}_0}$ exists, and is singular for any $\boldsymbol{\vartheta}_0$. Note, however, that the reference density for scale – the one that, by definition, provides the unit scale – here is $f = g_a$.

A tale of two densities, f and g_{a_Π}, is emerging, which demythifies the seemingly singular role of the Gaussian distribution.

This treatment of the univariate case provides a good intuition for the more complex k-dimensional problem where, as we shall see, the rank of the Fisher information matrix can take any value between $k + k(k+1)/2 = k(k+3)/2$ and $2k + k(k+1)/2 = k(k+5)/2$. Since the univariate case follows as a particular case by letting $k = 1$ in the general result of Proposition 3.1 of the next section, we do not provide a more formal statement here.

2.2. Some examples

In order to illustrate the results of the previous section, we now apply our findings in three examples of skewing functions and determine the exponential family with corresponding minimal sufficient statistic and natural parameter space leading to singular Fisher information matrices.

As a first example, we propose the most usual class of skewing functions, namely those of the form $\Pi_1(z,\delta) := \Pi(\delta z)$, where $\Pi : \mathbb{R} \to [0,1]$ is a function satisfying $\Pi(-y) + \Pi(y) = 1$ for all $y \in \mathbb{R}$ (hence $\Pi(0) = 1/2$) and such that $\dot{\Pi}(0) := \mathrm{d}\Pi(y)/\mathrm{d}y|_{y=0}$ exists and differs from 0.

Clearly, any univariate c.d.f. could be used, in which case we retrieve the skew-symmetric distributions of Azzalini and Capitanio [6], and, for $f = \phi$ and $\Pi = \Phi$, the skew-normal distributions of Azzalini [3]. For more examples of skewed distributions of this type, we refer the reader to Gómez *et al.* [16]. Straightforward calculations show that $\psi_1(z) = \dot{\Pi}(0)z$, and hence the minimal sufficient statistic characterizing the exponential family (2.3) is $\Psi_1(z) = \dot{\Pi}(0)z^2/2$. The resulting exponential family $\mathcal{E}_{\Psi_1}$ thus is nothing but the family of centered normal densities of the form

$$g_a^{(1)}(z) = \exp\bigl(-a\dot{\Pi}(0)z^2/2\bigr)\bigl(2\pi/(a\dot{\Pi}(0))\bigr)^{-1/2},$$

with natural parameter space $\mathcal{A}_1 := \mathrm{sign}(\dot{\Pi}(0))\mathbb{R}_0^+$. Assumptions (A1$^+$) and (A2$^+$) are satisfied, hence the 3×3 matrix $\mathbf{\Gamma}_{f;\boldsymbol{\vartheta}_0}$ exists. Thus, whenever the traditional skewing function Π_1 is used, Gaussian kernels are the only problematic ones regarding singular Fisher information at $\delta = 0$. This result, combined with the popularity of Π_1 as a skewing function, explains the long-standing belief in a particular role of the Gaussian distribution. Note that our findings are in line with earlier ones by Gómez *et al.* [16], who show that, by combining a Student kernel with ν degrees of freedom and a skewing function of the form Π_1, Fisher information at $\delta = 0$ is non-singular in general but becomes singular as $\nu \to \infty$. And, more generally, our results are in total accordance with those of Ley and Paindaveine [17] for the total class of skew-symmetric distributions of this kind.

Next consider the class of skewing functions $\Pi_2(z, \delta) := \Pi(\delta\,\mathrm{sign}(z)|z|^{\alpha/2}(2/\alpha)^{1/2})$ with $\alpha > 1$ and $y \mapsto \Pi(y)$ satisfying the usual conditions. Clearly, for $\alpha = 2$, Π_2 coincides with Π_1. This second type of skewing function was used, with $\Pi = \Phi$, by Azzalini [4] to define skew-exponential power distributions. One immediately obtains $\psi_2(z) = \dot{\Pi}(0)\,\mathrm{sign}(z)|z|^{\alpha/2}(2/\alpha)^{1/2}$, and, consequently,

$$\Psi_2(z) = \dot{\Pi}(0)|z|^{\alpha/2+1}(2/\alpha)^{1/2}(\alpha/2+1)^{-1}.$$

The corresponding exponential family $\mathcal{E}_{\Psi_2}$ contains all densities of the form

$$g_a^{(2)}(z) = c\exp\bigl(-a\dot{\Pi}(0)(2/\alpha)^{1/2}(\alpha/2+1)^{-1}|z|^{\alpha/2+1}\bigr),$$

where c is a normalization constant and a again ranges over either the positive or the negative real half line, depending on the sign of $\dot{\Pi}(0)$. One easily can check that the complete Fisher information matrix is well-defined in this case. DiCiccio and Monti [12] prove that, for $\alpha \neq 2$, skew-exponential power distributions do not suffer from singular Fisher information matrices in the vicinity of symmetry. Our findings do not only confirm that result, but also provide some further insight into the reasons for that absence of singularity. Actually, the exponent of $|z|$ in $g_a^{(2)}$ has to be $\alpha/2+1$, while the symmetric kernels in skew-exponential power distributions as defined in Azzalini [4] are of the form $c\exp(-|z|^\alpha/\alpha)$. Thus, while skew-normal distributions involve a symmetric kernel and a skewing function which are in a problematic relationship, this is avoided with the class of skew-exponential power distributions.

As a final example, consider skewing functions of the form $\Pi_3(z, \delta) := \Pi(\delta\sin(z))$, with Π belonging to the same class of functions as in the two preceding examples. It is easy to check that Π_3 then actually is a skewing function satisfying Assumption (A2$^+$). Direct manipulations yield $\psi_3(z) = \dot{\Pi}(0)\sin(z)$ and $\Psi_3(z) = -\dot{\Pi}(0)\cos(z)$. The natural parameter space $\mathcal{A}_3$ of the

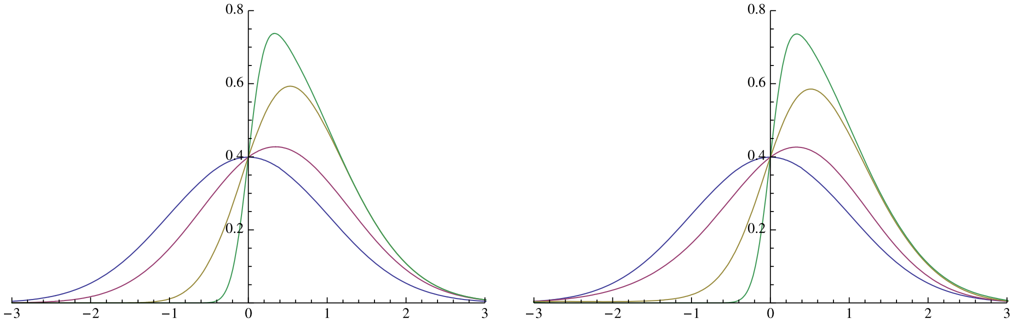

Figure 1. Plots of the original Azzalini [3] skew-normal density $2\phi(x)\Phi(\delta x)$ (left) and the Π_3-based version $2\phi(x)\Phi(\delta\sin(x))$ (right), for $\delta = 0$ (darker), 0.5, 2, and 6 (lighter).

exponential family $\mathcal{E}_{\Psi_3}$ corresponding to the minimal sufficient statistic Ψ_3 is empty. In other words, no symmetric kernel f yields a reduced Fisher information matrix when the skewing function Π_3 is adopted. Figure 1 shows some of the skewed densities obtained by combining Π_3 (for $\Pi = \Phi$) with a standard normal kernel. Comparison with the original skew-normal distributions of Azzalini [3] indicates that the new family, which is immune from degenerate Fisher information problems, is nevertheless extremely close to Azzalini's classical one.

3. The multivariate setup

3.1. A further tale . . .

Before starting our investigation of the multivariate case, let us introduce some further notations required when passing from dimension 1 to $k > 1$. For any given $k \times k$ matrix $\mathbf{M}$, we denote by $\text{vec}(\mathbf{M})$ the k^2-vector obtained by stacking the columns of $\mathbf{M}$ on top of each other, and by $\text{vech}(\mathbf{M})$ the $k(k+1)/2$-subvector of $\text{vec}(\mathbf{M})$ for which only upper diagonal entries in $\mathbf{M}$ are considered. We write $\mathbf{P}_k$ for the $k(k+1)/2 \times k^2$ matrix such that $\mathbf{P}'_k(\text{vech}\,\mathbf{M}) = \text{vec}(\mathbf{M})$ for any symmetric $\mathbf{M}$ and $\mathbf{I}_k$ for the $k \times k$ identity matrix.

The general multivariate skew-symmetric densities (generalizing (2.1)) we are considering are of the form (1.2), with $\boldsymbol{\vartheta}$, f and Π satisfying the general conditions (a)–(c). The symmetric kernel f moreover is supposed to have identity scatter matrix $\mathbf{I}_k$, which provides the required identification constraint for $\boldsymbol{\Sigma}$.

As in the univariate setup, we need to impose some mild regularity assumptions on f and Π.

Assumption (B1). *The mapping $\mathbf{z} \mapsto f(\mathbf{z})$ is differentiable, with gradient $\dot{f}$ such that, letting $\boldsymbol{\varphi}_f := -\dot{f}/f$, the $k \times k$ information matrix for location $\boldsymbol{\Sigma}^{-1/2}\mathcal{I}_f\boldsymbol{\Sigma}^{-1/2}$, with*

$$\mathcal{I}_f := \int_{\mathbb{R}^k} \boldsymbol{\varphi}_f(\mathbf{z})\boldsymbol{\varphi}'_f(\mathbf{z}) f(\mathbf{z})\,\mathrm{d}\mathbf{z},$$

is finite and invertible.

Assumption (B1$^+$). *Same as* (B1), *but the* $k(k+1)/2 \times k(k+1)/2$ *information matrix for scatter (actually, for* $\mathbf{\Sigma}^{1/2}$, *or more precisely, for* $\mathrm{vech}(\mathbf{\Sigma}^{1/2})$, *as* $\mathbf{\Sigma}^{1/2}$ *is symmetric)* $\mathbf{P}_k(\mathbf{\Sigma}^{-1/2} \otimes \mathbf{I}_k)\mathcal{J}_f(\mathbf{\Sigma}^{-1/2} \otimes \mathbf{I}_k)\mathbf{P}_k'$, *with*

$$\mathcal{J}_f := \int_{\mathbb{R}^k} \mathrm{vec}\big(\mathbf{z}\boldsymbol{\varphi}'_f(\mathbf{z}) - \mathbf{I}_k\big)\big(\mathrm{vec}\big(\mathbf{z}\boldsymbol{\varphi}'_f(\mathbf{z}) - \mathbf{I}_k\big)\big)' f(\mathbf{z})\,\mathrm{d}\mathbf{z},$$

moreover is finite and invertible.

Assumption (B2). (i) *The mapping* $\mathbf{z} \mapsto \Pi(\mathbf{z}, \boldsymbol{\delta})$ *is differentiable, and has gradient* $\mathbf{0}$ *at* $\boldsymbol{\delta} = \mathbf{0}$. (ii) *The mapping* $\boldsymbol{\delta} \mapsto \Pi(\mathbf{z}, \boldsymbol{\delta})$ *is differentiable at* $\boldsymbol{\delta} = \mathbf{0}$ *for all* $\mathbf{z} \in \mathbb{R}^k$, *with gradient (at* $\boldsymbol{\delta} = \mathbf{0}$) $\mathrm{grad}_{\boldsymbol{\delta}}\,\Pi(\mathbf{z}, \boldsymbol{\delta})|_{\boldsymbol{\delta}=\mathbf{0}} =: \boldsymbol{\psi}(\mathbf{z})$ *such that* $\boldsymbol{\psi}$ *admits a primitive* Ψ, *that is, a real-valued function* $\mathbf{z} \mapsto \Psi(\mathbf{z})$ *such that* $\mathrm{grad}_{\mathbf{z}}\Psi(\mathbf{z}) = \boldsymbol{\psi}(\mathbf{z})$.

Assumption (B2$^+$). *Same as* (B2), *but the* $k \times k$ *matrix*

$$\int_{\mathbb{R}^k} \boldsymbol{\psi}(\mathbf{z})\boldsymbol{\psi}'(\mathbf{z}) f(\mathbf{z})\,\mathrm{d}\mathbf{z}$$

moreover is finite and invertible.

These assumptions admit the same interpretation as in the univariate case, and basically ensure the existence of a finite Fisher information matrix. The standardization issue also calls for the same comments as in Section 2.1. The interpretation of the scatter matrix $\mathbf{\Sigma}$ is related to the choice of a standardization constraint on f. If we impose that $\mathbf{Z}$ with p.d.f. f has unit covariance matrix, then $\mathbf{\Sigma} = \int_{\mathbb{R}^k} (\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})' f^{\Pi}_{\boldsymbol{\mu},\mathbf{\Sigma},\mathbf{0}}(\mathbf{x})\,\mathrm{d}\mathbf{x}$. However, concepts of scatter that make sense irrespective of the underlying density also can be used in this multivariate setup, such as the celebrated *Tyler matrix* $\mathbf{V}_{\mathrm{Tyler}}$ (Tyler [22]), defined as the unique symmetric positive definite matrix $\mathbf{V}$ with $\mathrm{tr}\,\mathbf{V} = k$ satisfying

$$\mathrm{E}\big[(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})'/\big((\mathbf{X} - \boldsymbol{\mu})'\mathbf{V}^{-1}(\mathbf{X} - \boldsymbol{\mu})\big)\big] = k^{-1}\mathbf{V}.$$

Note however that the Tyler matrix $\mathbf{V}_{\mathrm{Tyler}}$ in fact is a *shape matrix*, not a scatter matrix: the corresponding scatter is $\mathbf{\Sigma} = \sigma\mathbf{V}_{\mathrm{Tyler}}$, with $\sigma = k^{-1}\,\mathrm{tr}(\mathbf{\Sigma})$. As in the univariate case, the scatter $\mathbf{\Sigma}$, for the kernel f, safely and without any loss of generality, can be fixed to identity for identification purposes, implying that, for f, σ takes value 1, while $\mathbf{V}_{\mathrm{Tyler}}$ is an identity matrix. As in the univariate case, this choice has no impact on the final results.

Here also, we could relax classical differentiability conditions by considering weaker differentiability and generalized Fisher information concepts, at the expense, however, of non-negligible technical complications.

Under Assumptions (B1) and (B2), the score vector $\boldsymbol{\ell}_{f;\boldsymbol{\vartheta}}$, at $\boldsymbol{\vartheta}_0 := (\boldsymbol{\mu}', \mathrm{vech}(\mathbf{\Sigma}^{1/2})', \mathbf{0}')'$, takes the form

$$\begin{aligned}\boldsymbol{\ell}_{f;\boldsymbol{\vartheta}_0}(\mathbf{x}) &:= \mathrm{grad}_{\boldsymbol{\vartheta}} \log f^{\Pi}_{\boldsymbol{\vartheta}}(\mathbf{x})|_{\boldsymbol{\vartheta}_0} =: \big(\boldsymbol{\ell}^{1\prime}_{f;\boldsymbol{\vartheta}_0}(\mathbf{x}) \quad \boldsymbol{\ell}^{2\prime}_{f;\boldsymbol{\vartheta}_0}(\mathbf{x}) \quad \boldsymbol{\ell}^{3\prime}_{f;\boldsymbol{\vartheta}_0}(\mathbf{x})\big)' \\ &= \begin{pmatrix} \mathbf{\Sigma}^{-1/2}\boldsymbol{\varphi}_f\big(\mathbf{\Sigma}^{-1/2}(\mathbf{x} - \boldsymbol{\mu})\big) \\ \mathbf{P}_k(\mathbf{\Sigma}^{-1/2} \otimes \mathbf{I}_k)\,\mathrm{vec}\big(\mathbf{\Sigma}^{-1/2}(\mathbf{x} - \boldsymbol{\mu})\boldsymbol{\varphi}'_f\big(\mathbf{\Sigma}^{-1/2}(\mathbf{x} - \boldsymbol{\mu})\big) - \mathbf{I}_k\big) \\ 2\boldsymbol{\psi}\big(\mathbf{\Sigma}^{-1/2}(\mathbf{x} - \boldsymbol{\mu})\big) \end{pmatrix},\end{aligned}$$

where $\otimes$ stands for the standard Kronecker product. Note that, for $k = 1$, this score vector coincides with the one we obtained in Section 2.1. Under Assumptions (B1$^+$) and (B2$^+$), the corresponding Fisher information matrix

$$\mathbf{\Gamma}_{f;\boldsymbol{\vartheta}_0} := |\mathbf{\Sigma}|^{-1/2} \int_{\mathbb{R}^k} \boldsymbol{\ell}_{f;\boldsymbol{\vartheta}_0}(\mathbf{x}) \boldsymbol{\ell}'_{f;\boldsymbol{\vartheta}_0}(\mathbf{x}) f\big(\mathbf{\Sigma}^{-1/2}(\mathbf{x}-\boldsymbol{\mu})\big)\, \mathrm{d}\mathbf{x}$$

exists and is finite, and naturally partitions into

$$\mathbf{\Gamma}_{f;\boldsymbol{\vartheta}_0} = \begin{pmatrix} \mathbf{\Gamma}^{11}_{f;\boldsymbol{\vartheta}_0} & \mathbf{0} & \mathbf{\Gamma}^{13}_{f;\boldsymbol{\vartheta}_0} \\ \mathbf{0} & \mathbf{\Gamma}^{22}_{f;\boldsymbol{\vartheta}_0} & \mathbf{0} \\ \mathbf{\Gamma}^{13\prime}_{f;\boldsymbol{\vartheta}_0} & \mathbf{0} & \mathbf{\Gamma}^{33}_{f;\boldsymbol{\vartheta}_0} \end{pmatrix},$$

with

$$\mathbf{\Gamma}^{11}_{f;\boldsymbol{\vartheta}_0} = \mathbf{\Sigma}^{-1/2} \mathcal{I}_f \mathbf{\Sigma}^{-1/2}, \qquad \mathbf{\Gamma}^{22}_{f;\boldsymbol{\vartheta}_0} = \mathbf{P}_k(\mathbf{\Sigma}^{-1/2} \otimes \mathbf{I}_k) \mathcal{J}_f (\mathbf{\Sigma}^{-1/2} \otimes \mathbf{I}_k) \mathbf{P}'_k,$$

$$\mathbf{\Gamma}^{33}_{f;\boldsymbol{\vartheta}_0} = 4 \int_{\mathbb{R}^k} \boldsymbol{\psi}(\mathbf{z}) \boldsymbol{\psi}'(\mathbf{z}) f(\mathbf{z})\, \mathrm{d}\mathbf{z} \quad \text{and} \quad \mathbf{\Gamma}^{13}_{f;\boldsymbol{\vartheta}_0} = 2\mathbf{\Sigma}^{-1/2} \int_{\mathbb{R}^k} \boldsymbol{\varphi}_f(\mathbf{z}) \boldsymbol{\psi}'(\mathbf{z}) f(\mathbf{z})\, \mathrm{d}\mathbf{z}.$$

As in the univariate case, the blocks of zeroes in $\mathbf{\Gamma}_{f;\boldsymbol{\vartheta}_0}$ readily follow from symmetry arguments and, without loss of generality, we can focus our attention on the submatrix

$$\mathbf{\Gamma}^{0}_{f;\boldsymbol{\vartheta}_0} := \begin{pmatrix} \mathbf{\Gamma}^{11}_{f;\boldsymbol{\vartheta}_0} & \mathbf{\Gamma}^{13}_{f;\boldsymbol{\vartheta}_0} \\ \mathbf{\Gamma}^{13\prime}_{f;\boldsymbol{\vartheta}_0} & \mathbf{\Gamma}^{33}_{f;\boldsymbol{\vartheta}_0} \end{pmatrix},$$

the existence of which only requires Assumptions (B1) and (B2$^+$). In the univariate case, the 2×2 matrix $\mathbf{\Gamma}^{0}_{f;\boldsymbol{\vartheta}_0}$ was either full-rank or singular with rank 1; here, the $2k \times 2k$ matrix $\mathbf{\Gamma}^{0}_{f;\boldsymbol{\vartheta}_0}$ can be singular with any rank ranging from k to $2k-1$ (note that the lower bound k is a direct consequence of either Assumption (B1) or (B2$^+$)).

The following proposition fully characterizes, for each possible rank $2k - m$, $m \in \{1, \dots, k\}$, the relation between the kernel f and the skewing function Π causing such degeneracy (for simplicity, we restrict to a characterization of the singularity of $\mathbf{\Gamma}^{0}_{f;\boldsymbol{\vartheta}_0}$).

Proposition 3.1. *Let the symmetric kernel f and the skewing function Π satisfy Assumptions* (B1) *and* (B2$^+$). *The following statements are equivalent*:

(i) *the $2k \times 2k$ matrix $\mathbf{\Gamma}^{0}_{f;\boldsymbol{\vartheta}_0}$ is singular with rank $2k - m$, $1 \le m \le k$, for any $\boldsymbol{\vartheta}_0$*;

(ii) *denoting by $\mathbf{Z}$ a random k-vector with p.d.f. f, there exists a $k \times k$ orthogonal matrix $\mathbf{O}' = (\mathbf{O}'_1, \mathbf{O}'_2)$, where $\mathbf{O}'_1$ and $\mathbf{O}'_2$ are $k \times m$- and $k \times (k-m)$-dimensional, respectively, such that, letting $\mathbf{Y} := \mathbf{O}\mathbf{Z}$ and $\mathbf{y} := \mathbf{O}\mathbf{z}$, for Lebesgue-almost all $\mathbf{O}_2\mathbf{z} = (y_{m+1}, \dots, y_k)' \in \mathbb{R}^{k-m}$, the density of $\mathbf{O}_1\mathbf{Z} = (Y_1, \dots, Y_m)'$ conditional on $\mathbf{O}_2\mathbf{Z} = (Y_{m+1}, \dots, Y_k)' = (y_{m+1}, \dots, y_k)'$ belongs to the exponential family*

$$\Big\{ (y_1, \dots, y_m) \mapsto g_a(y_1, \dots, y_m) := C^{-1} \exp(-a\Psi(\mathbf{O}'\mathbf{y})) \ \Big| \ a \text{ such that } C = C(y_{m+1}, \dots, y_k) := \int_{\mathbb{R}^m} \exp(-a\Psi(\mathbf{O}'\mathbf{y}))\, \mathrm{d}y_1 \cdots \mathrm{d}y_m < \infty \Big\} \tag{3.1}$$

with parameter a and minimal sufficient statistic $\Psi(\mathbf{O}'(Y_1, \ldots, Y_m, y_{m+1}, \ldots, y_k)')$.

Note that the natural parameter space

$$\mathcal{A} = \mathcal{A}(y_{m+1}, \ldots, y_k) := \left\{ a \in \mathbb{R} \text{ such that } \int_{\mathbb{R}^m} \exp(-a\Psi(\mathbf{O}'\mathbf{y}))\, \mathrm{d}y_1 \cdots \mathrm{d}y_m < \infty \right\}$$

of the exponential family (3.1) in principle also depends on $(y_{m+1}, \ldots, y_k)$. Natural parameters in exponential families being well identified, the values $a_\Pi(y_{m+1}, \ldots, y_k)$ of the natural parameter a achieving, whenever condition (ii) of Proposition 3.1 holds, the matchings $f = g_a$, are uniquely defined for Lebesgue-almost all $(k - m)$-tuple $(y_{m+1}, \ldots, y_k)$, yielding exponential densities $g_{a_\Pi(y_{m+1},\ldots,y_k)}$.

Proposition 3.1 has the following straightforward corollary.

Corollary 3.1. (i) *Let f be a symmetric kernel satisfying Assumption* (B1): *there exists a skewing function Π_f such that the rank of $\boldsymbol{\Gamma}^0_{f;\boldsymbol{\vartheta}_0}$ reaches its minimal value k for any $\boldsymbol{\vartheta}_0$.*

(ii) *Let Π be a skewing function satisfying Assumption* (*B*2) *with Ψ such that, for some a_Π,*

(iia) $\mathbf{z} \mapsto g_{a_\Pi}(\mathbf{z}) := C^{-1} \exp(-a_\Pi \Psi(\mathbf{z}))$ *is a p.d.f. with identity scatter matrix, and*
(iib) $\int_{\mathbb{R}^k} \boldsymbol{\psi}(\mathbf{z})\boldsymbol{\psi}'(\mathbf{z}) f(\mathbf{z})\, \mathrm{d}\mathbf{z}$ *is finite and invertible* (*meaning that* ($B2^+$) *is satisfied*).

Then, there exists a symmetric kernel f_Π such that the rank of $\boldsymbol{\Gamma}^0_{f_\Pi;\boldsymbol{\vartheta}_0}$ reaches its minimal value k for any $\boldsymbol{\vartheta}_0$.

Proof of Proposition 3.1. Clearly, $\boldsymbol{\Gamma}^0_{f;\boldsymbol{\vartheta}_0}$ has rank $2k - m$, $1 \le m \le k$, if and only if m is the largest integer such that there exist $(k \times m)$ matrices $\mathbf{V}$ and $\mathbf{W}$ with $(\mathbf{V}', -\mathbf{W}')$ of rank m such that

$$\mathbf{V}'\boldsymbol{\varphi}_f = \mathbf{W}'\boldsymbol{\psi} \qquad \text{Lebesgue-a.e.} \tag{3.2}$$

(note that the matrix $\boldsymbol{\Sigma}^{-1/2}$ is incorporated in $\mathbf{V}$, and hence plays no role in the characterization (3.2)). Both $\mathbf{V}$ and $\mathbf{W}$ are of maximal rank m. Suppose indeed that $\mathbf{V}$ is not: then, there exists $\mathbf{0} \neq \boldsymbol{\lambda} \in \mathbb{R}^m$ such that $\mathbf{V}\boldsymbol{\lambda} = \mathbf{0}$, so that $\boldsymbol{\lambda}'\mathbf{W}'\boldsymbol{\psi} = \boldsymbol{\lambda}'\mathbf{V}'\boldsymbol{\varphi}_f = 0$ (Lebesgue-a.e.). Then, in view of Assumption (B2), $\mathbf{W}\boldsymbol{\lambda} = \mathbf{0}$ as well, hence $\boldsymbol{\lambda}'(\mathbf{V}', -\mathbf{W}') = \mathbf{0}$, which contradicts the assumption that $(\mathbf{V}', -\mathbf{W}')$ has rank m. The same reasoning holds for $\mathbf{W}$. It follows that $\mathbf{V}$, without loss of generality, can be assumed to be orthonormal, and therefore can be extended into an orthogonal matrix $\mathbf{O}' := (\mathbf{V}, \mathbf{v})$, $\mathbf{v}$ being the $k \times (k - m)$ orthogonal complement to $\mathbf{V}$. The necessary and sufficient condition (3.2) then takes the form

$$[\mathbf{O}\boldsymbol{\varphi}_f]_{1\ldots m} = \mathbf{W}'\boldsymbol{\psi} \qquad \text{Lebesgue-a.e.} \tag{3.3}$$

where $[\mathbf{O}\boldsymbol{\varphi}_f]_{1\ldots m}$ stands for $\mathbf{O}\boldsymbol{\varphi}_f$'s m first rows.

Define $\mathbf{Y} := \mathbf{O}\mathbf{Z}$. Since $\mathbf{Z}$ has density f, $\mathbf{Y}$ has density $\mathbf{y} \mapsto f^{\mathbf{Y}}(\mathbf{y}) = f(\mathbf{O}'\mathbf{y})$. This density $f^{\mathbf{Y}}$ has gradient $\dot{f}^{\mathbf{Y}}$ and score $\boldsymbol{\varphi}_{f^{\mathbf{Y}}}$, with

$$\boldsymbol{\varphi}_{f^{\mathbf{Y}}}(\mathbf{y}) := -\dot{f}^{\mathbf{Y}}(\mathbf{y})/f^{\mathbf{Y}}(\mathbf{y}) = -\mathbf{O}\dot{f}(\mathbf{O}'\mathbf{y})/f(\mathbf{O}'\mathbf{y}) = \mathbf{O}\boldsymbol{\varphi}_f(\mathbf{O}'\mathbf{y}).$$

This, combined with (3.3), yields

$$[\boldsymbol{\varphi}_{f^{\mathbf{Y}}}(\mathbf{y})]_{1\ldots m} = \mathbf{W}'\boldsymbol{\psi}(\mathbf{O}'\mathbf{y}) \qquad \text{Lebesgue-a.e.}$$

or, more explicitly,

$$\begin{pmatrix} \partial_{y_1} \log f^{\mathbf{Y}}(\mathbf{y}) \\ \vdots \\ \partial_{y_m} \log f^{\mathbf{Y}}(\mathbf{y}) \end{pmatrix} = -\mathbf{W}'\boldsymbol{\psi}(\mathbf{O}'\mathbf{y}) \qquad \text{Lebesgue-a.e.} \tag{3.4}$$

As a function of $(y_1, \ldots, y_m)$, the left-hand side in (3.4) has primitive

$$\log f^{\mathbf{Y}}(y_1, \ldots, y_m, y_{m+1}, \ldots, y_k) + c(y_{m+1}, \ldots, y_k),$$

where the "integration constant" c is an arbitrary function of $(y_{m+1}, \ldots, y_k)$. The right-hand side therefore has the same primitive, still up to an additive $c(y_{m+1}, \ldots, y_k)$. Now, partitioning $\mathbf{O}'$ into $(\mathbf{O}'_1, \mathbf{O}'_2)$ where $\mathbf{O}'_1$ and $\mathbf{O}'_2$ are $k \times m$ and $k \times (k-m)$, respectively, a necessary condition for

$$(y_1, \ldots, y_m) \mapsto \mathbf{W}'\boldsymbol{\psi}\big(\mathbf{O}'_1(y_1, \ldots, y_m)' + \mathbf{O}'_2(y_{m+1}, \ldots, y_k)'\big)$$

to be the gradient of a scalar function is $\mathbf{W}' = a\mathbf{O}_1$ for some $a = a(y_{m+1}, \ldots, y_k) \in \mathbb{R}$: in view of Assumption (B2), a primitive of

$$(y_1, \ldots, y_m) \mapsto a\mathbf{O}_1\boldsymbol{\psi}\big(\mathbf{O}'_1(y_1, \ldots, y_m)' + \mathbf{O}'_2(y_{m+1}, \ldots, y_k)'\big)$$

is then $a\Psi(\mathbf{O}'_1(y_1, \ldots, y_m)' + \mathbf{O}'_2(y_{m+1}, \ldots, y_k)')$, up to the usual additive constant – here, an arbitrary function of $(y_{m+1}, \ldots, y_k)$. The necessary and sufficient condition (3.4) thus takes the further form

$$f^{\mathbf{Y}}(\mathbf{y}) = \exp(-c(y_{m+1}, \ldots, y_k)) \exp\big(-a\Psi\big(\mathbf{O}'_1(y_1, \ldots, y_m)' + \mathbf{O}'_2(y_{m+1}, \ldots, y_k)'\big)\big)$$

for some $a = a(y_{m+1}, \ldots, y_k) \in \mathbb{R}$; in other words, the conditional density of $(Y_1, \ldots, Y_m)'$ given $(Y_{m+1}, \ldots, Y_k)' = (y_{m+1}, \ldots, y_k)'$ is

$$\begin{aligned} & f^{(Y_1,\ldots,Y_m)'|(Y_{m+1},\ldots,Y_k)'=(y_{m+1},\ldots,y_k)'}(y_1, \ldots, y_m) \\ &\quad = f^{\mathbf{Y}}(y_1, \ldots, y_m, y_{m+1}, \ldots, y_k) \Big/ \int_{\mathbb{R}^m} f^{\mathbf{Y}}(y_1, \ldots, y_m, y_{m+1}, \ldots, y_k)\,\mathrm{d}y_1 \cdots \mathrm{d}y_m \\ &\quad = C(y_{m+1}, \ldots, y_k) \exp\big(-a\Psi\big(\mathbf{O}'_1(y_1, \ldots, y_m)' + \mathbf{O}'_2(y_{m+1}, \ldots, y_k)'\big)\big), \end{aligned} \tag{3.5}$$

where $C^{-1}(y_{m+1}, \ldots, y_k) := \int_{\mathbb{R}^m} \exp(-a\Psi(\mathbf{O}'_1(y_1, \ldots, y_m)' + \mathbf{O}'_2(y_{m+1}, \ldots, y_k)'))\,\mathrm{d}y_1 \cdots \mathrm{d}y_m$, for some $a = a(y_{m+1}, \ldots, y_k) \in \mathbb{R}$.

Summing up, there exists an orthogonal matrix $\mathbf{O}' = (\mathbf{O}'_1, \mathbf{O}'_2)$ such that, for any $(y_{m+1}, \ldots, y_k)' \in \mathbb{R}^{k-m}$, the density of $\mathbf{O}_1\mathbf{Z} =: (Y_1, \ldots, Y_m)'$ conditional on $\mathbf{O}_2\mathbf{Z} = (y_{m+1}, \ldots, y_k)'$ belongs to the exponential family with minimal sufficient statistic

$$\Psi\big(\mathbf{O}'_1(Y_1, \ldots, Y_m)' + \mathbf{O}'_2(y_{m+1}, \ldots, y_k)'\big),$$

as was to be proved. □

So far, we have formally solved the singularity problem for the $2k \times 2k$ information matrix $\mathbf{\Gamma}^0_{f;\boldsymbol{\vartheta}_0}$. As in the univariate case, the singularity problem for the full $k(k+5)/2 \times k(k+5)/2$ information matrix $\mathbf{\Gamma}_{f;\boldsymbol{\vartheta}_0}$ is slightly different. Indeed, the existence of $\mathbf{\Gamma}_{f;\boldsymbol{\vartheta}_0}$ requires the stronger Assumption (B1$^+$), as the information for scatter, which is not present in $\mathbf{\Gamma}^0_{f;\boldsymbol{\vartheta}_0}$, has to exist as well; this adds a further condition on the exponential family in Proposition 3.1. Nevertheless, there is no fundamental difference between the two setups: it only could happen that a solution to the singularity problem of $\mathbf{\Gamma}^0_{f;\boldsymbol{\vartheta}_0}$ is not a solution of the larger problem because the matrix $\mathbf{\Gamma}_{f;\boldsymbol{\vartheta}_0}$ simply does not exist, hence cannot be singular. This explains why, for the sake of simplicity, we state the results of this section in terms of $\mathbf{\Gamma}^0_{f;\boldsymbol{\vartheta}_0}$. The message is clear: the tale of two densities has turned into a more elaborate plot, starring a much larger number of actors.

3.2. Further examples

As in the univariate case, we now analyze three concrete examples of skewing functions in the light of the findings of the previous section, which provides the theoretical statement in Proposition 3.1 with some further intuition.

The first example is the natural extension of the univariate skewing function Π_1 to the multivariate context, with $\Pi_1^{(k)}(\mathbf{z}, \boldsymbol{\delta}) := \Pi(\boldsymbol{\delta}'\mathbf{z})$, where $\Pi : \mathbb{R} \to [0, 1]$ satisfies exactly the same conditions as in Section 2.2. The resulting class of skewing functions $\Pi_1^{(k)}$ is the most common one in the literature. A skewing function $\Pi = \Phi$ combined with a multinormal kernel $f = \phi_k$ yields the class of skew-multinormal densities of Azzalini and Dalla Valle [8]. When f is only required to be spherically symmetric and the skewing function Π is a univariate symmetric c.d.f., we obtain the class of skew-elliptical distributions as defined by Azzalini and Capitanio [6], itself a subclass of the generalized skew-elliptical distributions of Genton and Loperfido [15] where Π is left unspecified. Finally, relaxing the assumption of spherical symmetry into the weaker assumption of central symmetry, we retrieve the popular class of multivariate skew-symmetric distributions analyzed in Ley and Paindaveine [17].

Direct calculation yields $\boldsymbol{\psi}_1^{(k)}(\mathbf{z}) = \dot{\Pi}(0)\mathbf{z}$, hence, writing $\mathbf{z} = (\mathbf{z}_1', \mathbf{z}_2')'$ with $\mathbf{z}_1 \in \mathbb{R}^m$ and $\mathbf{z}_2 \in \mathbb{R}^{k-m}$, $m = 1, \ldots, k$, we obtain minimal sufficient statistics of the form

$$\Psi_1^{(k)}(\mathbf{O}'(\mathbf{Z}_1', \mathbf{z}_2')') = \dot{\Pi}(0)(\mathbf{Z}_1'\mathbf{Z}_1/2 + \mathbf{z}_2'\mathbf{z}_2/2)$$

for a $k \times k$ orthogonal matrix decomposing into $\mathbf{O}' = (\mathbf{O}_1', \mathbf{O}_2')$. Quite nicely, the possibility of separating the vectors $\mathbf{Z}_1$ and $\mathbf{z}_2$ in $\Psi_1^{(k)}(\mathbf{O}'(\mathbf{Z}_1', \mathbf{z}_2')')$ allows us to express the corresponding exponential densities in terms of $\mathbf{z}_1$ only, yielding the m-dimensional Gaussian densities

$$\mathbf{z}_1 \mapsto \exp\bigl(-a\dot{\Pi}(0)\mathbf{z}_1'\mathbf{z}_1/2\bigr)\bigl(2\pi/(a\dot{\Pi}(0))\bigr)^{-m/2}.$$

As in the univariate case, the sign of a is the same as that of $\dot{\Pi}(0)$. Degenerate information thus takes place iff, for some adequate rotation $\mathbf{OZ}$ of $\mathbf{Z} \sim f$, the m-dimensional marginal distribution of $[\mathbf{OZ}]_{1\ldots m}$ is standard m-variate normal. Note that this does not imply k-variate normal distributions. Consider, for example, a random k-vector whose first m components are i.i.d. standard Gaussian, and independent of the remaining $k-m$ ones, themselves i.i.d. with some other standardized univariate symmetric distribution. In such a case, the conditional distribution of the m first components given the $k-m$ last ones belongs to the exponential family of distributions just described. Thus, contrary to the univariate setup, multinormal densities are not the only symmetric kernels leading to singular Fisher information when combined with the skewing functions $\Pi_1^{(k)}$. Multinormal kernels, however, are the only ones for which Fisher information has minimal rank (corresponding to $m=k$). All this is in total accordance with earlier findings by Ley and Paindaveine [17], who examine in detail the singularity issues related to skew-symmetric distributions generated via $\Pi_1^{(k)}$. We therefore refer the reader to that reference for more details about the skewing functions $\Pi_1^{(k)}$, especially so for the special case of skew-elliptical distributions.

Our second example corresponds to another classical type of skewing functions, namely

$$\Pi_2^{(k)}(\mathbf{z},\boldsymbol{\delta}) := \Pi\big(\boldsymbol{\delta}'\mathbf{z}(\nu+k)^{1/2}(\mathbf{z}'\mathbf{z}+\nu)^{-1/2}\big), \tag{3.6}$$

where Π satisfies the same properties as above, and $\nu>0$. Clearly, as $\nu\to\infty$, the skewing functions $\Pi_2^{(k)}$ tend to skewing functions of the $\Pi_1^{(k)}$ type just considered. When Π in (3.6) corresponds to the c.d.f. $T_1(\cdot,\nu+k)$ of a Student variable with $\nu+k$ degrees of freedom, and the symmetric kernel used is a k-dimensional t variable with ν degrees of freedom, then we obtain the celebrated multivariate skew-t distributions of Azzalini and Capitanio [7] – up to some minor details, since their non-standardized skewing functions are of the form

$$T_1\big(\boldsymbol{\delta}'\boldsymbol{\omega}^{-1}(\mathbf{x}-\boldsymbol{\mu})(\nu+k)^{1/2}\big((\mathbf{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})+\nu\big)^{-1/2};\nu+k\big),$$

with $\boldsymbol{\omega}=\mathrm{diag}(\boldsymbol{\Sigma}_{11},\ldots,\boldsymbol{\Sigma}_{kk})^{1/2}$. Elementary calculation yields

$$\boldsymbol{\psi}_2^{(k)}(\mathbf{z}) = \dot{\Pi}(0)\mathbf{z}(\nu+k)^{1/2}(\mathbf{z}'\mathbf{z}+\nu)^{-1/2},$$

hence minimal sufficient statistics and exponential densities of the form

$$\Psi_2^{(k)} = \dot{\Pi}(0)(\nu+k)^{1/2}(\mathbf{z}'\mathbf{z}+\nu)^{1/2}$$

and

$$\begin{aligned}&\exp\big(-a\dot{\Pi}(0)(\nu+k)^{1/2}(\mathbf{z}'\mathbf{z}+\nu)^{1/2}\big)\\ &\quad\Big/\int_{\mathbb{R}^m}\exp\big(-a\dot{\Pi}(0)(\nu+k)^{1/2}(\mathbf{z}'\mathbf{z}+\nu)^{1/2}\big)\,\mathrm{d}z_1\cdots\mathrm{d}z_m,\end{aligned} \tag{3.7}$$

respectively. Here again, the sign of a is determined by the sign of $\dot{\Pi}(0)$. Azzalini and Genton [9] conjecture that, as long as ν is finite, multivariate skew-t distributions should be free of singularity problems. DiCiccio and Monti [13] prove the conjecture in the univariate case, Ley and Paindaveine [18] in any dimension k. Proposition 3.1 confirms those earlier results, as (3.7), whatever the value of a, cannot be derived from a k-dimensional t distribution with ν degrees of freedom. Actually, letting $\mathbf{X}=(\mathbf{X}_1',\mathbf{X}_2')'$ follow a k-variate t distribution where $\mathbf{X}_1$ and $\mathbf{X}_2$, respectively, are m- and $(k-m)$-dimensional random vectors, it can be shown that the density of $\mathbf{X}_1|\mathbf{X}_2 = \mathbf{x}_2$ cannot be of the form (3.7).

We conclude this section with a possible extension of the singularity-free univariate skewing function Π_3 of Section 2.2. Consider $\Pi_3^{(k)}(\mathbf{z}, \boldsymbol{\delta}) := \Pi(\boldsymbol{\delta}' \operatorname{Sin}(\mathbf{z}))$, with Π defined as above and $\operatorname{Sin}(\mathbf{z}) := (\sin(z_1), \ldots, \sin(z_k))'$. Checking the validity of Assumption (B2$^+$) is immediate, and one also directly obtains that $\boldsymbol{\psi}_3^{(k)}(\mathbf{z}) = \dot{\Pi}(0) \operatorname{Sin}(\mathbf{z})$ and $\Psi_3^{(k)} = -\dot{\Pi}(0)(\cos(z_1) + \cdots + \cos(z_k))$. The same reasoning as for Π_3 readily yields that the natural parameter space related to the exponential family with minimal sufficient statistic $\Psi_3^{(k)}$ is empty, hence skewing functions of the type $\Pi_3^{(k)}$ can be used without worrying about possibly singular Fisher information.

4. Final comments

In this paper, we fully dispel the widespread opinion that Gaussian densities, in the context of skew-symmetric distributions, constitute an intriguing worst-case situation, being the only ones (possibly, after restriction to linear subspaces) leading to degenerate Fisher information matrices in the vicinity of symmetry. Our main result provides a complete characterization of that information degeneracy phenomenon, which generalizes and extends all previous results of that type, and highlights the link between the symmetric kernel and the skewing function causing singularity. We also show how that link, in the univariate as well as in the multivariate case, can be described as a mismatch between two densities, in which the Gaussian distribution plays no particular role. By avoiding such mismatch, one can deal with skew-symmetric distributions without worrying about singular Fisher information and its consequences.

Acknowledgements

Marc Hallin is also member of the Académie Royale de Belgique and ECORE, and an extra-muros Fellow of CentER, Tilburg University. His research is supported by the Sonderforschungsbereich "Statistical modeling of nonlinear dynamic processes" (SFB 823) of the German Research Foundation (Deutsche Forschungsgemeinschaft) and a Discovery Grant of the Australian Research Council. The financial support and hospitality of ORFE and the Bendheim Center at Princeton University, where part of this work was completed, is gratefully acknowledged. Christophe Ley thanks the Fonds National de la Recherche Scientifique, Communauté française de Belgique, for support via a Mandat d'Aspirant FNRS. Both authors would like to thank two anonymous referees for helpful comments that led to an improvement of the paper.

References

[1] Arellano-Valle, R.B. and Azzalini, A. (2008). The centred parametrization for the multivariate skew-normal distribution. *J. Multivariate Anal.* **99** 1362–1382. MR2424355

[2] Arnold, B.C. and Beaver, R.J. (2002). Skewed multivariate models related to hidden truncation and/or selective reporting. *Test* **11** 7–54. MR1915776

[3] Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scand. J. Statist.* **12** 171–178. MR0808153
[4] Azzalini, A. (1986). Further results on a class of distributions which includes the normal ones. *Statistica (Bologna)* **46** 199–208. MR0877720
[5] Azzalini, A. (2005). The skew-normal distribution and related multivariate families. *Scand. J. Statist.* **32** 159–188. MR2188669
[6] Azzalini, A. and Capitanio, A. (1999). Statistical applications of the multivariate skew normal distribution. *J. R. Stat. Soc. Ser. B* **61** 579–602. MR1707862
[7] Azzalini, A. and Capitanio, A. (2003). Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution. *J. R. Stat. Soc. Ser. B* **65** 367–389. MR1983753
[8] Azzalini, A. and Dalla Valle, A. (1996). The multivariate skew-normal distribution. *Biometrika* **83** 715–726. MR1440039
[9] Azzalini, A. and Genton, M.G. (2008). Robust likelihood methods based on the skew-t and related distributions. *International Statistical Review* **76** 106–129.
[10] Branco, M.D. and Dey, D.K. (2001). A general class of multivariate skew-elliptical distributions. *J. Multivariate Anal.* **79** 99–113. MR1867257
[11] Chiogna, M. (2005). A note on the asymptotic distribution of the maximum likelihood estimator for the scalar skew-normal distribution. *Stat. Methods Appl.* **14** 331–341. MR2211338
[12] DiCiccio, T.J. and Monti, A.C. (2004). Inferential aspects of the skew-exponential power distribution. *J. Amer. Statist. Assoc.* **99** 439–450. MR2062829
[13] DiCiccio, T.J. and Monti, A.C. (2009). Inferential aspects of the skew-t distribution. Unpublished manuscript.
[14] Genton, M.G., ed. (2004). *Skew-Elliptical Distributions and Their Applications: A Journey Beyond Normality*. Boca Raton, FL: Chapman & Hall/CRC. MR2156754
[15] Genton, M.G. and Loperfido, N. (2005). Generalized skew-elliptical distributions and their quadratic forms. *Ann. Inst. Statist. Math.* **57** 389–401. MR2160656
[16] Gómez, H.W., Venegas, O. and Bolfarine, H. (2007). Skew-symmetric distributions generated by the distribution function of the normal distribution. *Environmetrics* **18** 395–407. MR2370731
[17] Ley, C. and Paindaveine, D. (2010). On the singularity of multivariate skew-symmetric models. *J. Multivariate Anal.* **101** 1434–1444. MR2609504
[18] Ley, C. and Paindaveine, D. (2010). On Fisher information matrices and profile log-likelihood functions in generalized skew-elliptical models. *Metron* **68**, special issue on "Skew-symmetric and flexible distributions," 235–250.
[19] Pewsey, A. (2000). Problems of inference for Azzalini's skew-normal distribution. *J. Appl. Statist.* **27** 859–870.
[20] Pewsey, A. (2006). Some observations on a simple means of generating skew distributions. In *Advances in Distribution Theory, Order Statistics, and Inference* (N. Balakrishman, E. Castillo and J.M. Sarabia, eds.) 75–84. Boston, MA: Birkhäuser. MR2226216
[21] Rotnitzky, A., Cox, D.R., Bottai, M. and Robins, J. (2000). Likelihood-based inference with singular information matrix. *Bernoulli* **6** 243–284. MR1748721
[22] Tyler, D.E. (1987). A distribution-free M-estimator of multivariate scatter. *Ann. Statist.* **15** 234–251. MR0885734
[23] van der Vaart, A.W. (2000). *Asymptotic Statistics*. Cambridge: Cambridge Univ. Press.
[24] Wang, J., Boyer, J. and Genton, M.G. (2004). A skew-symmetric representation of multivariate distributions. *Statist. Sinica* **14** 1259–1270. MR2126352

Received July 2010 and revised November 2010