ECORE

INTERNATIONAL ASSOCIATION FOR RESEARCH AND TEACHING
Economics, Finance, Operations Research, Econometrics and Statistics

++ research ++ teaching ++

ECORE DISCUSSION PAPER

2010/40

# Skew-Symmetric Distributions and Fisher Information – A Tale of Two Densities

Marc HALLIN
Christoophe LEY

CORE
Université catholique de Louvain

ECARES
Université Libre de Bruxelles

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

Marc Hallin

Christophe Ley

*E.C.A.R.E.S., CP 114, and Département de Mathématique, CP 210, Université Libre de Bruxelles, 50 Avenue F.D. Roosevelt, 1050 Brussels, Belgium*

**Summary**. 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 major 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 very special and somewhat mysterious role in that context. We totally dispel that widespread opinion by providing a full characterization 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*: Skewing function, Skew-normal distributions, Skew-symmetric distributions, Singular Fisher information, 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 these skewed distributions started with the seminal papers by Azzalini (1985, 1986) 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 pdf

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

where $\phi$ and $\Phi$ respectively denote the probability density function (pdf) and cumulative distribution function (cdf) of a standard normal distribution. Besides their many appealing features, however, skew-normal densities unfortunately also suffer from a major inferential weakness: 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 hypothesis of symmetry. Such a situation has been studied by Rotnitzky et al. (2000), 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 (1996) and Azzalini and Capitanio (1999) consider multivariate skew-normal distributions resulting from replacing in (1.1) the univariate normal kernel $\phi$ with its $k$-variate version $\phi_k$. In the same paper, Azzalini and Capitanio also propose substituting an elliptical kernel $f_k$ for the normal one $\phi_k$, and replacing the skewing factor $\Phi$ in (1.1) with the cdf $F_1$ corresponding to the univariate counterpart $f_1$ of $f_k$. The resulting distributions are called *skew-elliptical*. The class of skew-elliptical distributions is also studied in detail by Branko and Dey (2001), based, however, on a slightly different definition. Azzalini and Capitanio (2003) introduce the class of multivariate skew-$t$ distributions, and propose a general model encompassing all previous ones. In a similar spirit, Genton and Loperfido (2005) introduce a concept of generalized skew-elliptical distributions where arbitrary skewing functions (satisfying cdf-type conditions, but not necessarily of the form $F_1$) can be used in conjunction with the elliptical kernel $f_k$. Finally, Wang et al. (2004) are relaxing the assumption of elliptically symmetric kernels into a weaker assumption of central symmetry, with multivariate *skew-symmetric* densities of the form

$$\mathbf{x} \mapsto f^{\Pi}_{\boldsymbol{\vartheta}}(\mathbf{x}) = f^{\Pi}_{\boldsymbol{\mu},\Sigma,\boldsymbol{\delta}}(\mathbf{x}) := 2\,|\Sigma|^{-1/2} f(\Sigma^{-1/2}(\mathbf{x}-\boldsymbol{\mu}))\Pi(\Sigma^{-1/2}(\mathbf{x}-\boldsymbol{\mu}),\boldsymbol{\delta}), \quad \mathbf{x} \in \mathbb{R}^k, \quad (1.2)$$

where

(a) $\boldsymbol{\mu} \in \mathbb{R}^k$ is a location parameter, $\Sigma \in \mathcal{S}^k$ (throughout, $|M|$ denotes the determinant and $M^{1/2}$ the symmetric square-root of any $M$ 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 pdf, 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. (2004) actually do not consider any specific $\boldsymbol{\delta}$-parametrization. Our parametric approach (with the regularity assumptions (A2) and (B2) of Sections 2.1 and 3.1) 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 2000). For further information about skew-symmetric models and related topics, we refer the reader to the recent monograph by Genton (2004) and to the review papers by Arnold and Beaver (2002) and Azzalini (2005).

The issue of singular Fisher information runs like a red thread through all those developments. Mentioned in Azzalini (1985) itself, it is discussed, in the univariate and multivariate skew-normal context, by Azzalini and Capitanio (1999), Pewsey (2000), Chiogna (2005) and Arellano-Valle and Azzalini (2008). The same issue has been considered in various subclasses of skew-symmetric distributions. Pewsey (2006) and Azzalini and Genton (2008) establish that the singularity problem remains after replacement of the cdf $\Phi$ in (1.1) with any cdf $H$ satisfying mild regularity assumptions. DiCiccio and Monti (2004) prove that, within the class of univariate skew-exponential power distributions of Azzalini (1986), 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 Gòmez et al. (2007) and DiCiccio and Monti (2010).

The multivariate counterparts of these statements are provided in Ley and Paindaveine (2010a and 2010b), respectively.

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

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

encompassing all previous cases, is considered in Ley and Paindaveine (2010a), 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.

All this tends to indicate that, for some obscure reasons, normal kernels and, consequently, skew-normal distributions play a very special role in skew-symmetric families, at the boundary between symmetry and asymmetry.

In this paper, we completely dispel that feeling of a mythical role of Gaussian kernels by 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 $\Pi$—more specifically, between two densities, the kernel $f$ and a density $g_\Pi$ associated with the skewing function $\Pi$.

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. Some 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 univariate skew-symmetric distributions with pdfs of the form

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

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* pdf, 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 $\sigma$ that does not imply any loss of generality. Classical standardization, with a constraint of the form $\int_{-\infty}^{\infty} z^2 f(z)dz = 1$, involves the variance of $Z$ with pdf $f$; the scale parameter $\sigma^2$ then is the mean squared deviation $\mathrm{E}[(X-\mu)^2]$ with respect to $\mu$ of $X$ with pdf $f^{\Pi}_{\mu,\sigma,0}$. If moment assumptions are to be avoided, one may rather consider, for instance, medians of squares, with an identification constraint of the form $\int_{-\infty}^{1} f(z)dz = 0.75$: if $X$ has pdf $f^{\Pi}_{\mu,\sigma,0}$, $\sigma$ 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. Once that choice has been made, it has no impact on the results of this paper, and hence does not require further specification.

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 1985), $\Pi(z, \delta) = \Phi(\delta\,\mathrm{sign}(z)|z|^{\alpha/2}(2/\alpha)^{1/2})$ (skew-exponential power distributions, Azzalini 1986) or $\Pi(z, \delta) = G(\delta z)$ for any symmetric univariate cdf $G$ (skew-symmetric distributions, Azzalini and Capitanio 1999). The class of skewing functions considered here is much broader.

The regularity assumptions we are making on $f$ and $\Pi$ are as follows.

ASSUMPTION (A1). $z \mapsto f(z)$ is differentiable, with derivative $\dot{f}$ such that, letting

$\varphi_f := -\dot{f}/f$, the information quantities (for location and scale, respectively)

$$\mathcal{I}_f := \int_{-\infty}^{\infty} \varphi_f^2(z)f(z)dz \quad \text{and} \quad \mathcal{J}_f := \int_{-\infty}^{\infty} (z\varphi_f(z)-1)^2 f(z)dz$$

are finite.

Assumption (A2). $\delta \mapsto \Pi(z,\delta)$ 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 $\int_{-\infty}^{\infty} \psi^2(z)f(z)dz$ is finite and strictly positive; $z \mapsto \psi(z)$ moreover admits a primitive, denoted as $\Psi$.

Assumptions (A1) and (A2) 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 classical definition of Fisher information.

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

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

where the factor 2 in $\ell^3_{f;\boldsymbol{\vartheta}_0}$ follows from the fact that $\Pi(z,0) = 1/2$ for all $z \in \mathbb{R}$. The resulting Fisher information matrix is

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

with

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

and

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

The zeroes in $\mathit{\Gamma}_{f;\boldsymbol{\vartheta}_0}$ 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 $\mathit{\Gamma}_{f;\boldsymbol{\vartheta}_0}$ only can be due to the singularity of the $2 \times 2$ submatrix

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

which, clearly, either is full-rank or, in case $\gamma^{11}_{f;\boldsymbol{\vartheta}_0}\gamma^{33}_{f;\boldsymbol{\vartheta}_0} = (\gamma^{13}_{f;\boldsymbol{\vartheta}_0})^2$, has rank 1.

Now, the Cauchy-Schwarz inequality implies that $(\gamma^{13}_{f;\boldsymbol{\vartheta}_0})^2 \leq \gamma^{11}_{f;\boldsymbol{\vartheta}_0}\gamma^{33}_{f;\boldsymbol{\vartheta}_0}$ with equality iff $\varphi_f = a\psi$ $f$-a.s. (equivalently, Lebesgue-a.e.) for some constant $a \in \mathbb{R}$. It thus follows that

$$\Gamma_{f;\boldsymbol{\vartheta}_0} \text{ is singular iff } \varphi_f = a\psi \text{ Lebesgue-a.e. for some } a \in \mathbb{R}. \tag{2.2}$$

Replacing $\varphi_f$ with its definition, (2.2) yields a first-order differential equation whose solutions are of the form $f(x) = c\exp(-a\Psi(x))$, where $\Psi$ is a primitive of $\psi$ and $c \in \mathbb{R}$ an integration constant.

Summing up, $\Gamma_{f;\boldsymbol{\vartheta}_0}$ is singular if and only if the symmetric kernel $f$ belongs to the *natural exponential family*

$$\left\{ g_a := \exp(-a\Psi)/\int_{-\infty}^{\infty}\exp(-a\Psi(z))dz \,\middle|\, a \text{ such that } \int_{-\infty}^{\infty}\exp(-a\Psi(z))dz < \infty \right\} \tag{2.3}$$

with *privileged statistic* $\Psi$ and *natural parameter* $-a$.

Note that $\mathcal{A} := \{a \text{ such that } \int_{-\infty}^{\infty}\exp(-a\Psi(z))dz < \infty\}$, as the *natural parameter space* of an exponential family, is an open interval of $\mathbb{R}$. The unique value $a_\Pi$ of $a \in \mathcal{A}$ such that $f$ and $g_{a_\Pi}$ coincide is entirely determined by the standardization constraint on $f$. If the classical variance-based standardization is adopted, classical results on exponential families imply that $a_\Pi$ is solution of the equation $\partial^2_a(\log\int_{-\infty}^{\infty}\exp(-a\Psi(z))dz) = 1$. If standardization is imposed via medians of squares, $a_\Pi$ is the unique solution of

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

It follows that, for any symmetric density $f$ satisfying Assumption (A1), there exists a skewing function $\Pi_f$ (infinitely many of them, actually) such that $\Gamma_{f;\boldsymbol{\vartheta}_0}$ is singular; among them, $\Pi_f(z,\delta) := \Phi(\delta\varphi_f(z))$. Conversely, for any skewing function $\Pi$ satisfying Assumption (A2) with $\Psi$ such that, for some $a_\Pi$, $z \mapsto \frac{1}{c}\exp(-a_\Pi\Psi(z))$ is a pdf with scale 1, there exists a symmetric $f_\Pi$ (namely, $f_\Pi(z) := \frac{1}{c}\exp(-a_\Pi\Psi(z))$) such that $\Gamma_{f_\Pi;\boldsymbol{\vartheta}_0}$ similarly is singular. A tale of two densities, $f$ and $g_{a_\Pi}$, demythifying the seemingly singular role of the Gaussian, is born.

This treatment of the univariate case provides a good intuition for the more subtle $k$-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 privileged 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 $\dot{\Pi}(0) := d\Pi(y)/dy|_{y=0}$ exists and differs from 0. Clearly, any univariate cdf could be used, in which case we retrieve the skew-symmetric distributions of Azzalini and Capitanio (1999), and, for $f = \phi$ and $\Pi = \Phi$, the skew-normal distributions of Azzalini (1985). For more examples of skewed distributions of this type, we refer the reader to Gòmez et al. (2007). Straightforward calculation shows that $\psi_1(z) = \dot{\Pi}(0)z$, and hence the privileged statistic characterizing the exponential family (2.3) is $\Psi_1(z) = \dot{\Pi}(0)z^2/2$. The resulting exponential family thus is nothing but the family of centered normal densities of the form

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

a scale family, with natural parameter space $\mathcal{A}_1 := \text{sign}(\dot{\Pi}(0))\mathbb{R}_0^+$. In other words, whenever a traditional skewing function of the type $\Pi_1$ is used, Gaussian kernels are the only problematic ones regarding singular Fisher information at $\delta = 0$. This result, combined with the popularity of those skewing functions, explains the long-standing belief in a particular role of the Gaussian distribution. Note that our findings are in line with earlier ones by Gòmez et al. (2007), who show that, by combining a Student kernel with $\nu$ degrees of freedom and a skewing function of the form $\Pi_1$, Fisher information at $\delta = 0$ remains strictly positive but tends to zero as $\nu \to \infty$. And, more generally, our results are in total accordance with those of Ley and Paindaveine (2010a) for the total class of skew-symmetric distributions of this kind.

Next consider the class of skewing functions $\Pi_2(z, \delta) := \Pi(\delta\, \text{sign}(z)|z|^{\alpha/2}(2/\alpha)^{1/2})$ with $\alpha > 1$ and $y \mapsto \Pi(y)$ satisfying the usual conditions. Clearly, for $\alpha = 2$, $\Pi_2$ coincides with $\Pi_1$. This second type of skewing function was used, with $\Pi = \Phi$, by Azzalini (1986) to define skew-exponential power distributions. One immediately obtains that $\psi_2(z) = \dot{\Pi}(0)\text{sign}(z)|z|^{\alpha/2}(2/\alpha)^{1/2}$, and, consequently,

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

The corresponding exponential family contains all densities of the form

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

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 $\dot{\Pi}(0)$. DiCiccio and Monti (2004) prove that, for $\alpha \neq 2$, skew-exponential power distributions do not suffer from singular Fisher information matrices. 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

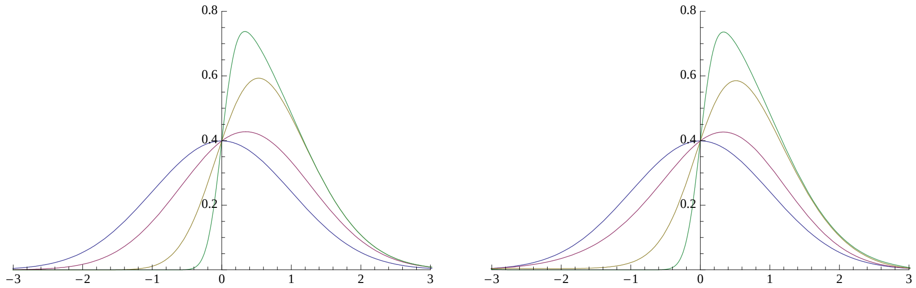


**Fig. 1.** Plots of the original Azzalini (1985) skew-normal density $2\phi(x)\Phi(\delta x)$ (left) and the $\Pi_3$-based version $2\phi(x)\Phi(\delta \sin(x))$ (right), for $\delta$ =0 (blue), 0.5 (red), 2 (yellow), and 6 (green).

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 (1986) 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 $\Pi$ belonging to the same class of functions as in the two preceding examples. It is easy to check that $\Pi_3$ 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 exponential family corresponding to the privileged statistic $\Psi_3$ is empty. In other words, no symmetric kernel $f$ yields a reduced Fisher information matrix when the skewing function $\Pi_3$ is adopted. Figure 1 shows some of the skewed densities obtained by combining $\Pi_3$ (for $\Pi = \Phi$) with a standard normal kernel. Comparison with the original skew-normal distributions of Azzalini (1985) 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 $M$, we denote by $\mathrm{vec}(M)$ the $k^2$-vector obtained by stacking the columns of $M$ on top of each other, and by $\mathrm{vech}(M)$ the $k(k+1)/2$-subvector of $\mathrm{vec}(M)$ for which only upper diagonal entries in $M$ are considered. We write $P_k$ for the $k(k+1)/2\times k^2$ matrix such that $P_k'(\mathrm{vech}\, M) = \mathrm{vec}(M)$ for any symmetric $M$ and $I_k$ 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 $\Pi$ satisfying the general conditions (a)-(c). The symmetric kernel $f$ moreover is supposed to have identity scatter matrix $I_k$, which provides the required identification constraint for $\Sigma$.

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

ASSUMPTION (B1). $\mathbf{z} \mapsto f(\mathbf{z})$ is differentiable, with gradient $\dot{f}$ such that, letting $\boldsymbol{\varphi}_f := -\dot{f}/f$, the $k \times k$ information matrices (for location and scatter, respectively)

$$\mathcal{I}_f := \int_{\mathbb{R}^k} \boldsymbol{\varphi}_f(\mathbf{z})\boldsymbol{\varphi}'_f(\mathbf{z})f(\mathbf{z})\,d\mathbf{z} \quad \text{and} \quad \mathcal{J}_f := \int_{\mathbb{R}^k} \text{vec}(\mathbf{z}\,\boldsymbol{\varphi}'_f(\mathbf{z}) - I_k)(\text{vec}(\mathbf{z}\,\boldsymbol{\varphi}'_f(\mathbf{z}) - I_k))' f(\mathbf{z})\,d\mathbf{z}$$

are finite and invertible.

ASSUMPTION (B2). $\boldsymbol{\delta} \mapsto \Pi(\mathbf{z}, \boldsymbol{\delta})$ is differentiable at $\boldsymbol{\delta} = \mathbf{0}$ for all $\mathbf{z} \in \mathbb{R}^k$, with gradient (at $\boldsymbol{\delta} = \mathbf{0}$) $\text{grad}_{\boldsymbol{\delta}} \Pi(\mathbf{z}, \boldsymbol{\delta})|_{\boldsymbol{\delta}=\mathbf{0}} =: \boldsymbol{\psi}(\mathbf{z})$ such that $\int_{\mathbb{R}^k} \boldsymbol{\psi}(\mathbf{z})\boldsymbol{\psi}'(\mathbf{z})f(\mathbf{z})\,d\mathbf{z}$ is finite and invertible; $\mathbf{z} \mapsto \boldsymbol{\psi}(\mathbf{z})$ moreover admits a primitive, that is, there exists a real-valued function $\mathbf{z} \mapsto \Psi(\mathbf{z})$ such that $\text{grad}_{\mathbf{z}} \Psi(\mathbf{z}) = \psi(\mathbf{z})$.

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. The interpretation of the scatter matrix $\Sigma$ is related to the choice of a standardization constraint on $f$. If we impose that $\mathbf{Z}$ with pdf $f$ has unit covariance matrix, then $\Sigma = \int_{\mathbb{R}^k} (\mathbf{x}-\boldsymbol{\mu})(\mathbf{x}-\boldsymbol{\mu})' f^{\Pi}_{\boldsymbol{\mu},\Sigma,\mathbf{0}}(\mathbf{x})\,d\mathbf{x}$. However, concepts of scatter that exist irrespective of the underlying density can also be used in this multivariate setup, such as the celebrated Tyler matrix (Tyler 1987), defined as the unique symmetric positive definite matrix $V$ with $\text{tr}\,V = k$ satisfying $\text{E}[(\mathbf{X}-\boldsymbol{\mu})(\mathbf{X}-\boldsymbol{\mu})'/(\mathbf{X}-\boldsymbol{\mu})' V^{-1}(\mathbf{X}-\boldsymbol{\mu})] = k^{-1} V$. As in the univariate case, we assume that an identification constraint, hence a definition of $\Sigma$, has been adopted; that choice has no impact on the results, and will not be specified any further.

Here again, 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 $\boldsymbol{\ell}^{\Pi}_{f;\boldsymbol{\vartheta}}$, at $\boldsymbol{\vartheta}_0 := (\boldsymbol{\mu}', \text{vech}(\Sigma^{1/2})', \mathbf{0}')'$, takes the form

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

where $\otimes$ stands for the standard Kronecker product. Note that, for $k = 1$, this score vector coincides with that of Section 2.1. The corresponding Fisher information matrix

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

naturally partitions into

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

with

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

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

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

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

Contrary to the univariate case, where the $2\times 2$ matrix $\mathit{\Gamma}^{0}_{f;\boldsymbol{\vartheta}_0}$ was either full-rank or singular with rank 1, the $2k\times 2k$ matrix $\mathit{\Gamma}^{0}_{f;\boldsymbol{\vartheta}_0}$ 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 $2k-m$, $m\in\{1,\ldots,k\}$, the relation between the kernel $f$ and the skewing function $\Pi$ causing such degeneracy.

Proposition 3.1. *Let Assumptions (B1) and (B2) hold. The following statements are equivalent.*

*(i)* $\mathit{\Gamma}_{f;\boldsymbol{\vartheta}_0}$ *is singular with rank* $2k-m+k(k+1)/2$, $1\le m\le k$;

*(ii)* $\mathit{\Gamma}^{0}_{f;\boldsymbol{\vartheta}_0}$ *is singular with rank* $2k-m$, $1\le m\le k$;

*(iii) there exists an orthogonal matrix* $O' = (O_1', O_2')$, *where* $O_1'$ *and* $O_2'$ *are* $k\times m$- *and* $k\times(k-m)$-*dimensional, respectively, such that, letting* $\mathbf{Y} := O\mathbf{Z}$ *and* $\mathbf{y} := O\mathbf{z}$, *for Lebesgue-almost all* $O_2\mathbf{z} = (y_{m+1},\ldots,y_k)'\in\mathbb{R}^{k-m}$, *the density of* $O_1\mathbf{Z} = (Y_1,\ldots,Y_m)'$ *conditional on* $O_2\mathbf{Z} = (Y_{m+1},\ldots,Y_k)' = (y_{m+1},\ldots,y_k)'$ *belongs to the natural exponential family*

$$\Big\{(y_1,\ldots,y_m)\mapsto g_a(y_1,\ldots,y_m) := C^{-1}\exp(-a\Psi(O'\mathbf{y}))\Big|\, a \text{ such that } C = C(y_{m+1},\ldots,y_k) := \int_{\mathbb{R}^m}\exp(-a\Psi(O'\mathbf{y}))dy_1\ldots dy_m < \infty\Big\} \tag{3.1}$$

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

Note that the natural parameter space

$$\mathcal{A} = \mathcal{A}(y_{m+1}, \ldots, y_k) := \{a \text{ such that } \int_{\mathbb{R}^m} \exp(-a\Psi(O'\mathbf{y}))dy_1 \ldots dy_m < \infty\}$$

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 (iii) 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) For any symmetric kernel $f$ satisfying Assumption (B1), there exists a skewing function $\Pi_f$ such that the rank of $\Gamma^0_{f;\boldsymbol{\vartheta}_0}$ reaches its minimal value $k$;*

*(ii) for any skewing function $\Pi$ satisfying Assumption (B2) with $\Psi$ such that, for some $a_\Pi$, $\mathbf{z} \mapsto \frac{1}{C}\exp(-a_\Pi\Psi(\mathbf{z}))$ is a pdf with identity scatter matrix, there exists a symmetric kernel $f_\Pi$ such that the rank of $\Gamma^0_{f_\Pi;\boldsymbol{\vartheta}_0}$ reaches its minimal value $k$.*

PROOF OF PROPOSITION 3.1. The equivalence between (i) and (ii) trivially follows from the fact that the cross-information blocks between scatter and location and between scatter and asymmetry are zero. We thus only have to prove the equivalence between (ii) and (iii).

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

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

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

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

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

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

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

This, combined with (3.3), yields

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

or, more explicitly,

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

As a function of $(y_1, \dots, y_m)$, the left-hand side in (3.4) has primitive $\log f^{\mathbf{Y}}(y_1, \dots, y_m, y_{m+1}, \dots, y_k) + c(y_{m+1}, \dots, y_k)$, where the "integration constant" $c$ is an arbitrary function of $(y_{m+1}, \dots, y_k)$. The right-hand side therefore has the same primitive, still up to $c(y_{m+1}, \dots, y_k)$. Now, partitioning $O'$ into $(O_1', O_2')$ where $O_1'$ and $O_2'$ are $k \times m$ and $k \times (k-m)$, respectively, a necessary condition for $(y_1, \dots, y_m) \mapsto W'\boldsymbol{\psi}(O_1'(y_1, \dots, y_m)' + O_2'(y_{m+1}, \dots, y_k)')$ to be the gradient of a scalar function is $W' = aO_1$ for some $a = a(y_{m+1}, \dots y_k) \in \mathbb{R}$: in view of Assumption (B2), a primitive of

$$(y_1, \dots, y_m) \mapsto a\,O_1\boldsymbol{\psi}(O_1'(y_1, \dots, y_m)' + O_2'(y_{m+1}, \dots, y_k)')$$

is then $a\Psi(O_1'(y_1, \dots, y_m)' + O_2'(y_{m+1}, \dots, y_k)')$, up to the usual additive constant—here, an arbitrary function of $(y_{m+1}, \dots, y_k)$. The necessary and sufficient condition (3.4) thus takes the further form

$$f^{\mathbf{Y}}(\mathbf{y}) = \exp(-c(y_{m+1}, \dots, y_k)) \exp(-a\Psi(O_1'(y_1, \dots, y_m)' + O_2'(y_{m+1}, \dots, y_k)'))$$

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

$$\begin{aligned} & f^{(Y_1,\dots,Y_m)' \,|\, (Y_{m+1},\dots,Y_k)'=(y_{m+1},\dots,y_k)'}(y_1, \dots, y_m) \\ &= f^{\mathbf{Y}}(y_1, \dots, y_m, y_{m+1}, \dots, y_k) \Big/ \int_{\mathbb{R}^m} f^{\mathbf{Y}}(y_1, \dots, y_m, y_{m+1}, \dots, y_k) dy_1 \dots dy_m \\ &= C(y_{m+1}, \dots, y_k) \exp(-a\Psi(O_1'(y_1, \dots, y_m)' + O_2'(y_{m+1}, \dots, y_k)')), \end{aligned} \tag{3.5}$$

where $C^{-1}(y_{m+1}, \dots, y_k) := \int_{\mathbb{R}^m} \exp(-a\Psi(O_1'(y_1, \dots, y_m)' + O_2'(y_{m+1}, \dots, y_k)'))dy_1 \dots dy_m$, for some $a = a(y_{m+1}, \dots, y_k) \in \mathbb{R}$.

Summing up, there exists an orthogonal matrix $O' = (O_1', O_2')$ such that, for any $(y_{m+1}, \dots, y_k)' \in \mathbb{R}^{k-m}$, the density of $O_1\mathbf{Z} =: (Y_1, \dots, Y_m)'$ conditional on $O_2\mathbf{Z} = (y_{m+1}, \dots, y_k)'$ belongs to the natural exponential family with privileged statistic $\Psi(O_1'(Y_1, \dots, Y_m)' + O_2'(y_{m+1}, \dots, y_k)')$, as was to be proved. □

The tale of two densities has turned into a more elaborate one involving infinitely many of them.

### *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 $\Pi_1$ 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 (1996). When $f$ is only required to be spherically symmetric, with radial density $\pi$, and the skewing function $\Pi$ is the cdf with density $\pi$, we obtain the class of skew-elliptical distributions as defined by Azzalini and Capitanio (1999), which is itself a subclass of the generalized skew-elliptical distributions of Genton and Loperfido (2005) where $\Pi$ is left unspecified. Finally, relaxing the assumption of spherical symmetry into the weaker assumption of central symmetry, we retrieve the popular class of skew-symmetric distributions analyzed in Ley and Paindaveine (2010a).

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 $\mathbf{z}_2 \in \mathbb{R}^{k-m}$, $m = 1, \ldots, k$, privileged statistics of the form $\Psi_1^{(k)}(O'(\mathbf{Z}_1', \mathbf{z}_2')') = \dot{\Pi}(0)(\mathbf{Z}_1'\mathbf{Z}_1/2 + \mathbf{z}_2'\mathbf{z}_2/2)$ for a $k \times k$ orthogonal matrix decomposing into $O' = (O_1', O_2')$. Quite nicely, the possibility of separating the vectors $\mathbf{Z}_1$ and $\mathbf{z}_2$ in $\Psi_1^{(k)}(O'(\mathbf{Z}_1', \mathbf{z}_2')')$ allows us to express the corresponding exponential densities in terms of $\mathbf{z}_1$ only, yielding the $m$-dimensional Gaussian densities

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

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 $O\mathbf{Z}$ of $\mathbf{Z} \sim f$, the $m$-dimensional marginal distribution of $[O\mathbf{Z}]_{1\ldots m}$ 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 of Ley and Paindaveine (2010a), 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 $\Pi_2^{(k)}(\mathbf{z},\boldsymbol{\delta}) := \Pi(\boldsymbol{\delta}'\mathbf{z}(\nu+k)^{1/2}(\mathbf{z}'\mathbf{z}+\nu)^{-1/2})$, where $\Pi$ satisfies the same properties as above, and $\nu > 0$. Clearly, as $\nu \to \infty$, $\Pi_2^{(k)}$ tends to skewing functions of the $\Pi_1^{(k)}$ type just considered. When $\Pi$ corresponds to the cdf $T_1(\,\cdot\,,\nu+k)$ of a Student variable with $\nu+k$ degrees of freedom, and the symmetric kernel used is a $k$-dimensional $t$ variate with $\nu$ degrees of freedom, then we obtain the celebrated multivariate skew-$t$ distributions of Azzalini and Capitanio (2003) (up to some minor details, since their non-standardized skewing functions are of the form $T_1(\boldsymbol{\delta}'\omega^{-1}(\mathbf{x}-\boldsymbol{\mu})(\nu+k)^{1/2}((\mathbf{x}-\boldsymbol{\mu})'\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})+\nu)^{-1/2};\nu+k)$, with $\omega = \mathrm{diag}(\Sigma_{11},\ldots,\Sigma_{kk})^{1/2}$). 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 privileged statistics of the form $\Psi_2^{(k)} = \dot{\Pi}(0)(\nu+k)^{1/2}(\mathbf{z}'\mathbf{z}+\nu)^{1/2}$ and exponential densities

$$\exp(-a\dot{\Pi}(0)(\nu+k)^{1/2}(\mathbf{z}'\mathbf{z}+\nu)^{1/2}) \Big/ \int_{\mathbb{R}^m} \exp(-a\dot{\Pi}(0)(\nu+k)^{1/2}(\mathbf{z}'\mathbf{z}+\nu)^{1/2}) dz_1 \ldots dz_m. \quad (3.6)$$

Here again, the sign of $a$ is determined by the sign of $\dot{\Pi}(0)$. Azzalini and Genton (2008) conjecture that, as long as $\nu$ is finite, multivariate skew-$t$ distributions should be free of singularity problems. DiCiccio and Monti (2010) prove the conjecture in the univariate case, Ley and Paindaveine (2010b) in higher dimension; Proposition 3.1 confirms those earlier results, as (3.6), whatever the value of $a$, cannot be derived from a $k$-dimensional $t$ distribution with $\nu$ degrees of freedom. Actually, letting $\mathbf{X} = (\mathbf{X}_1', \mathbf{X}_2')'$ follow a $k$-variate $t$ distribution where $\mathbf{X}_1$ and $\mathbf{X}_2$ are, respectively, random $m$- and $(k-m)$-vectors, it can be shown that the density of $\mathbf{X}_1 \,|\, \mathbf{X}_2 = \mathbf{x}_2$ cannot be of the form (3.6).

We conclude this section with a possible extension of the singularity-free univariate skewing function $\Pi_3$ of Section 2.2. Consider $\Pi_3^{(k)}(\mathbf{z},\boldsymbol{\delta}) := \Pi(\boldsymbol{\delta}'\mathrm{Sin}(\mathbf{z}))$, with $\Pi$ defined as above and $\mathrm{Sin}(\mathbf{z}) := (\sin(z_1),\ldots,\sin(z_k))'$. Checking the validity of Assumption (B2) is immediate, and one also directly obtains that $\boldsymbol{\psi}_3^{(k)}(\mathbf{z}) = \dot{\Pi}(0)\mathrm{Sin}(\mathbf{z})$ and $\Psi_3^{(k)} = -\dot{\Pi}(0)(\cos(z_1) + \ldots + \cos(z_k))$. The same reasoning as for $\Pi_3$ readily yields that the natural parameter space related to the exponential family with privileged 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-symmeric distributions, constitute a mysterious 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 information singularity, 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.

**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.

## 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.

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

[3] Azzalini, A. (1985) A class of distributions which includes the normal ones. *Scand. J. Statist.*, 12, 171–178.

[4] Azzalini, A. (1986) Further results on a class of distributions which includes the normal ones. *Statistica*, 46, 199–208.

[5] Azzalini, A. (2005) The skew-normal distribution and related multivariate families (with discussion). *Scand. J. Statist.*, 32, 159–188.

[6] Azzalini, A. and Capitanio, A. (1999) Statistical applications of the multivariate skew-normal distributions. *J. R. Stat. Soc. B*, 61, 579–602.

[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. B*, 65, 367–389.

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

[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.

[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.

[12] DiCiccio, T. J. and Monti, A. C. (2004) Inferential aspects of the skew-exponential power distribution. *J. Amer. Statist. Assoc.*, 99, 439–450.

[13] DiCiccio, T. J. and Monti, A. C. (2010) Inferential aspects of the skew-$t$ distribution. Unpublished manuscript.

[14] Genton, M. G. (2004) *Skew-elliptical Distributions and Their Applications: A Journey Beyond Normality*, edited volume, Boca Raton, FL: Chapman and Hall/CRC.

[15] Genton, M. G. and Loperfido, N. (2005) Generalized skew-elliptical distributions and their quadratic forms. *Ann. Inst. Statist. Math.*, 57, 389–401.

[16] Gòmez, H. W., Torres, F. J. and Bolfarine, H. (2007) Large-sample inference for the epsilon-skew-$t$ distribution. *Commun. Statist. - Theory* & *Methods*, **36**, 73–81.

[17] Ley, C. and Paindaveine, D. (2010a) On the singularity of multivariate skew-symmetric models. *J. Multivariate Anal.*, to appear.

[18] Ley, C. and Paindaveine, D. (2010b) On Fisher information matrices and profile log-likelihood functions in generalized skew-elliptical models. *Metron*, special issue on “Skew-symmetric and flexible distributions”, to appear.

[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* (eds N. Balakrishnan, E. Castillo and J. M. Sarabia), pp. 75-84. Boston: Statistics for Industry and Technology, Birkhäuser.

[21] Rotnitzky, A., Cox, D. R., Bottai, M. and Roberts, J. (2000) Likelihood-based inference with singular information matrix. *Bernoulli*, 6, 243–284.

[22] Tyler, D. E. (1987) A distribution-free $M$-estimator of multivariate scatter. *Ann. Statist.*, **15**, 234–251.

[23] van der Vaart, A. W. (2000) *Asymptotic Statistics*. Cambridge : Cambridge University Press.

[24] Wang, J., Boyer, J. and Genton, M. G. (2004) A skew-symmetric representation of multivariate distribution. *Statist. Sinica*, 14, 1259–1270.