SFB 823

Skew-symmetric distributions and Fisher information The double sin of the skew-normal

Marc Hallin, Christophe Ley

Discussion Paper

Nr. 37/2012

Skew-Symmetric Distributions and Fisher Information
the double sin of the skew-normal

Marc Hallin[a], Christophe Ley[b]

[a]*ECARES, Université Libre de Bruxelles, CP114, 50 Ave. F.D. Roosevelt, 1050 Brussels, Belgium and ORFE, Princeton University, Sherrerd Hall, Princeton, NJ 08544, USA*
[b]*ECARES and Département de Mathématique, Université Libre de Bruxelles, CP210, Boulevard du Triomphe, 1050 Brussels, Belgium*

Abstract

Hallin and Ley (2012) investigate and fully characterize the Fisher singularity phenomenon in univariate and multivariate families of *skew-symmetric distributions.* This paper proposes a refined analysis of the (univariate) Fisher degeneracy problem, showing that it can be more or less severe, inducing $n^{1/4}$ ("simple singularity"), $n^{1/6}$ ("double singularity"), or $n^{1/8}$ ("triple singularity") consistency rates for the skewness parameter. We show, however, that simple singularity (yielding $n^{1/4}$ consistency rates), if any singularity at all, is the rule, in the sense that double and triple singularities are possible for generalized skew-normal families only. We also show that higher-order singularities, leading to worse-than-$n^{1/8}$ rates, cannot occur.

Key words: Consistency rates, Skewing function, Skew-normal distributions, Skew-symmetric distributions, Singular Fisher information, Symmetric kernel
2000 MSC: 62E10, 62F12

1. Introduction.

The *skew-symmetric* families, originally proposed in Azzalini and Capitanio (2003) and Wang *et al.* (2004), are, in their univariate version, parametric families of probability density functions (pdfs) of the form

$$x \mapsto f_{\boldsymbol{\vartheta}}^{\Pi}(x) := 2\,\sigma^{-1} f(\sigma^{-1}(x-\mu))\Pi(\sigma^{-1}(x-\mu), \delta), \quad x \in \mathbb{R}, \tag{1.1}$$

where

Email addresses: mhallin@ulb.ac.be (Marc Hallin), chrisley@ulb.ac.be (Christophe Ley)
URL: http://homepages.ulb.ac.be/~chrisley (Christophe Ley)

(a) $\boldsymbol{\vartheta} = (\mu, \sigma, \delta)'$, with $\mu \in \mathbb{R}$ a *location parameter*, $\sigma \in \mathbb{R}_0^+$ a *scale parameter*, while $\delta \in \mathbb{R}$ plays the role of a *skewness parameter*;

(b) $f : \mathbb{R} \to \mathbb{R}_0^+$, the *symmetric kernel*, is a symmetric nonvanishing pdf (such that, for any $z \in \mathbb{R}$, $0 \neq f(-z) = f(z)$), and

(c) $\Pi : \mathbb{R} \times \mathbb{R} \to [0, 1]$ is a *skewing function*, that is, satisfies

$$\Pi(-z, \delta) + \Pi(z, \delta) = 1, \quad z, \delta \in \mathbb{R}, \quad \text{and} \quad \Pi(z, 0) = 1/2, \quad z \in \mathbb{R}, \tag{1.2}$$

and, in case $(z, \delta) \mapsto \Pi(z, \delta)$ admits a derivative of order s at $\delta = 0$ for all $z \in \mathbb{R}$,

$$\partial_z^s \Pi(z, \delta)|_{\delta=0} = 0, \quad z \in \mathbb{R} \quad \text{and, for } s \text{ even,} \quad \partial_\delta^s \Pi(z, \delta)|_{\delta=0} = 0, \quad z \in \mathbb{R}. \tag{1.3}$$

While condition (1.2) is classical, (1.3), which involves the derivatives of Π, is less usual. The main justification for it lies in the analogy with skewing functions of the form $\Pi(z, \delta) = \Pi(\delta z)$, by far the most common ones. If Π is s times continuously differentiable, $\partial_z^s \Pi(\delta z) = \delta^s (\partial^s \Pi)(\delta z)$ obviously vanishes at $\delta = 0$. Similarly, the fact that $\Pi(-y) + \Pi(y) = 1$ implies that $\partial^s \Pi(\delta z)$ cancels at $\delta = 0$ for even values of s. All skewing functions considered in the literature, as well as those appearing in the examples developed in this paper and in Hallin and Ley (2012), satisfy (1.3). Further comments on the skewing functions of the form $\Pi(z, \delta) = \Pi(\delta z)$ can be found in Section 5.5.

The *skew-normal* family of Azzalini (1985), for which the symmetric kernel f is the standard Gaussian pdf ϕ and the skewing function $\Pi(z, \delta) = \Phi(\delta z)$ for Φ the standard Gaussian cumulative distribution function (cdf), is the oldest and most popular example of such a skew-symmetric family; varying f and Π, however, yields a virtually infinite number of them. Traditional examples include the skew-exponential power distributions of Azzalini (1986), the skew-Cauchy distributions of Arnold and Beaver (2000), the skew-t densities of Azzalini and Capitanio (2003), or the generalized skew-normal distributions of Loperfido (2004). We refer to Genton (2004), Azzalini (2005) or Ley (2012) for background reading, details and examples.

Since the pioneering paper by Azzalini (1985), it is well known that the scalar skew-normal distribution suffers from a Fisher information singularity problem at $\delta = 0$. More precisely, the Fisher information matrix for the three-parameter density (1.1) in the scalar skew-normal case is singular—typically, with rank 2 instead of 3—in the vicinity of symmetry, that is, at $\delta = 0$. Such a singularity violates the standard assumptions for root-n asymptotic inference, and skew-normal distributions therefore are problematic from an inferential point of view; in particular, any nontrivial traditional test of the null hypothesis of symmetry, at first sight, seems impossible.

That degeneracy problem has been discussed at length in a number of papers, among which Azzalini and Capitanio (1999), Pewsey (2000), DiCiccio and Monti (2004), Chiogna (2005), Azzalini and Genton (2008) or Ley and Paindaveine (2010); see Hallin and Ley (2012) for a detailed account. While all authors were pointing at some special status for normal kernels, hence skew-normal distributions, Hallin and Ley (2012) have shown that this information deficiency has no special relation to the skew-normal case, but actually originates in an unfortunate mismatch between f and Π—more precisely, between two densities, the kernel f and an exponential density g_Π associated with the skewing function Π (see Section 2.1).

The deficiency of the Fisher information matrix results in slower consistency rates in the estimation of the skewness parameter (at $\delta = 0$)—equivalently, it yields slower local alternative rates (*contiguity* rates) in tests of the null hypothesis of symmetry ($\delta = 0$). That impact of singular Fisher information on consistency/contiguity rates has been studied, in a general context, for the particular case of a deficiency of order one, by Rotnitzky *et al.* (2000), who unify and reinforce earlier proposals by, e.g., Cox and Hinkley (1974, pp. 117–118) or Lee and Chesher (1986).

The typical rate, corresponding to a "simple singularity", would be $n^{1/4}$. However, it is well-known from e.g. Chiogna (2005) that, for skew-normal distributions, that rate (for the estimation of δ at $\delta = 0$) drops down to $n^{1/6}$. In order to understand and explain this intriguing phenomenon, we pursue and refine, in the present paper, the analysis of Fisher singularity initiated in Hallin and Ley (2012). We show that this deterioration from $n^{1/4}$ to $n^{1/6}$ is explained by a "double singularity" property (a terminology that will become clear in the course of this paper)—the *double sin* of the skew-normal. That $n^{1/6}$ rate in turn possibly can drop further down to $n^{1/8}$, a case of "triple singularity". This, however, as we show in Theorem 4.1, is the worst case: "fourfold singularities"—*quadruple sins*—yielding $n^{1/10}$ rates or worse, are impossible.

Our aim is to characterize, in the spirit of Hallin and Ley (2012), among all families of univariate skew-symmetric distributions suffering from Fisher singularity, those exhibiting that double/triple singularity phenomenon, and to show that there exist no higher-order ones. It turns out that only Gaussian kernels can exhibit double (hence, also triple) degeneracy. The skew-normal family is one example; other ones are found in the class of generalized skew-normal distributions (Loperfido 2004). We also provide (in the spirit of Rotnitzky *et al.* 2000) the reparametrizations and the scores taking care of simple, double and triple singularities and achieving the $n^{1/4}$, $n^{1/6}$ and $n^{1/8}$ consistency/contiguity rates, respectively.

The paper is organized as follows. Section 2 deals with the simple singularity case, Section 3 with double singularity. Section 4 analyzes the triple singularity case and shows that higher-order ones are excluded. Examples for each type of singularity, and a discussion of the most standard type of skewing function are provided in Section 5.

2. The simple singularity case.

In this section, we first briefly revisit the main result of Hallin and Ley (2012); we then show how to remove the singularity problem via an adequate reparametrization leading, in general, to $n^{1/4}$ consistency rates for the skewness parameter in the vicinity of symmetry.

2.1. Simple singularity: a mismatch between f and Π.

Throughout, we consider the skew-symmetric distributions with pdf (1.1), along with regularity assumptions on f and Π that will be tightened from section to section. The minimal regularity assumptions we need are those of Hallin and Ley (2012).

Assumption (A1). (i) The symmetric kernel f is a *standardized* symmetric pdf. (ii) The mapping $z \mapsto f(z)$ is continuously differentiable, with derivative $\dot{f}$, at all $z \in \mathbb{R}$. (iii) Letting $\varphi_f := -\dot{f}/f$, the information quantities $\sigma^{-2}\mathcal{I}_f$ for location and $\sigma^{-2}\mathcal{J}_f$ for scale, with

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

are finite.

Assumption (A2). (i) The mapping $(z, \delta) \mapsto \Pi(z, \delta)$ is continuously differentiable at $\delta = 0$ for all $z \in \mathbb{R}$; (ii) the derivative $\partial_\delta \Pi(z, \delta)|_{\delta=0} =: \psi(z)$ admits a primitive Ψ; (iii) the quantity $\int_{-\infty}^{\infty} \psi^2(z) f(z) dz$ is finite.

Regarding Assumption (A1)(i), the term "standardized" means that the scale parameter (not necessarily a standard error, so that finite second-order moments are not required) of the symmetric kernel equals one—an identification constraint for σ that does not imply any loss of generality; see Hallin and Ley (2012) for a discussion of possible choices of scale parameters. All other assumptions ensure the existence and finiteness of Fisher information for the original parametrization.

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

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

where the factor 2 in $\ell^3_{f;\boldsymbol{\vartheta}_0}$ follows from the fact that $\Pi(z,0)=1/2$ for all $z\in\mathbb{R}$. We attract the reader's attention to the fact that the skewing function Π plays no role in the score functions for μ and σ at $\delta=0$. The resulting 3×3 Fisher information matrix then exists, is finite, and takes the form

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

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 $\boldsymbol{\Gamma}_{f;\boldsymbol{\vartheta}_0}$ are easily obtained by noting that $\ell^1_{f;\boldsymbol{\vartheta}_0}$ and $\ell^3_{f;\boldsymbol{\vartheta}_0}$ are odd functions of $(x-\mu)$, whereas $\ell^2_{f;\boldsymbol{\vartheta}_0}$ is even with respect to the same quantity. Consequently, Fisher singularity only can be caused by the collinearity of $\ell^1_{f;\boldsymbol{\vartheta}_0}$ and $\ell^3_{f;\boldsymbol{\vartheta}_0}$. Starting from that elementary observation, Hallin and Ley (2012) show that the family of densities (1.1) characterized by a couple (f,Π) suffers from Fisher singularity at $\delta=0$ if and only if the symmetric kernel f belongs to the *exponential family*

$$\mathcal{E}_\Psi := \Big\{ g_a := \exp(-a\Psi)\Big/\int_{-\infty}^{\infty}\exp(-a\Psi(z))dz \,\Big|\, a\in\mathcal{A}\Big\} \tag{2.4}$$

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

$$\mathcal{A} := \Big\{a\in\mathbb{R} \text{ such that } \int_{-\infty}^{\infty}\exp(-a\Psi(z))dz<\infty\Big\},$$

yielding

$$\gamma^{11}_{f;\boldsymbol{\vartheta}_0} = \sigma^{-2}a^2\int_{-\infty}^{\infty}\psi^2(z)f(z)dz \quad\text{and}\quad \gamma^{13}_{f;\boldsymbol{\vartheta}_0} = 2\sigma^{-1}a\int_{-\infty}^{\infty}\psi^2(z)f(z)dz. \tag{2.5}$$

We refer the reader to the end of Section 2.1 in Hallin and Ley (2012) for comments and a discussion on the existence of couples (f,Π) such that $f\in\mathcal{E}_\Psi$ for given f and for given Π, respectively.

2.2. Towards a singularity-free reparametrization: orthogonalization.

A natural way to handle this singularity problem consists in reparametrizing (1.1) in the spirit of Rotnitzky *et al.* (2000). Assume that f and Π are such that $f \in \mathcal{E}_\Psi$. The collinearity between the score for location and the score for skewness can be taken care of by a Gram-Schmidt orthogonalization process applied to the three components of $\boldsymbol{\ell}_{f;\boldsymbol{\vartheta}_0}$. This process projects, in the L_2 geometry of the information matrix, the score for skewness $\ell^3_{f;\boldsymbol{\vartheta}_0}$ onto the subspace orthogonal (at $\boldsymbol{\vartheta}_0$) to the scores for location and scale $\ell^1_{f;\boldsymbol{\vartheta}_0}$ and $\ell^2_{f;\boldsymbol{\vartheta}_0}$, so that the score for skewness becomes orthogonal to the score for location (since it is already orthogonal to $\ell^2_{f;\boldsymbol{\vartheta}_0}$). The resulting score for skewness is

$$\ell^{3(1)}_{f;\boldsymbol{\vartheta}_0} = \ell^3_{f;\boldsymbol{\vartheta}_0} - \ell^1_{f;\boldsymbol{\vartheta}_0}\mathrm{Cov}(\ell^1_{f;\boldsymbol{\vartheta}_0}, \ell^3_{f;\boldsymbol{\vartheta}_0})/\mathrm{Var}(\ell^1_{f;\boldsymbol{\vartheta}_0}),$$

while the other two scores remain unchanged: $\ell^{1(1)}_{f;\boldsymbol{\vartheta}_0} = \ell^1_{f;\boldsymbol{\vartheta}_0}$, $\ell^{2(1)}_{f;\boldsymbol{\vartheta}_0} = \ell^2_{f;\boldsymbol{\vartheta}_0}$. As expected, in view of (2.5),

$$\ell^{3(1)}_{f;\boldsymbol{\vartheta}_0}(x) = 2\psi(\sigma^{-1}(x-\mu)) - \sigma^{-1}a\psi(\sigma^{-1}(x-\mu))\frac{2\sigma^{-1}a\int_{-\infty}^{\infty}\psi^2(z)f(z)dz}{\sigma^{-2}a^2\int_{-\infty}^{\infty}\psi^2(z)f(z)dz} = 0. \quad (2.6)$$

This orthogonal system of scores is associated (at $\boldsymbol{\vartheta}_0$) with the reparametrization $\boldsymbol{\vartheta}^{(1)} := (\mu^{(1)}, \sigma^{(1)}, \delta^{(1)})'$, with

$$\mu^{(1)} = \mu + 2\delta\sigma/a, \quad \sigma^{(1)} = \sigma, \quad \text{and} \quad \delta^{(1)} = \delta,$$

hence with

$$f^{\Pi}_{\boldsymbol{\vartheta}^{(1)}}(x) := 2\,(\sigma^{(1)})^{-1}f((x-\mu^{(1)}+2\delta^{(1)}\sigma^{(1)}/a)/\sigma^{(1)})\Pi((x-\mu^{(1)}+2\delta^{(1)}\sigma^{(1)}/a)/\sigma^{(1)}, \delta^{(1)});$$

it is easily checked, indeed, that $\partial_{\delta^{(1)}} f^{\Pi}_{\boldsymbol{\vartheta}^{(1)}}(x)|_{\delta=\delta^{(1)}=0} = \ell^{3(1)}_{f;\boldsymbol{\vartheta}_0}(x)$. Note that, under $\delta = \delta^{(1)} = 0$ (but not in a neighborhood thereof) $\boldsymbol{\vartheta}^{(1)}$ and $\boldsymbol{\vartheta} = \boldsymbol{\vartheta}_0$ coincide.

Since this reparametrization, which only affects the location parameter, cancels (at $\boldsymbol{\vartheta}^{(1)}_0 := (\mu^{(1)}, \sigma^{(1)}, 0)' = (\mu, \sigma, 0)' = \boldsymbol{\vartheta}_0$) the score for skewness, second derivatives with respect to $\delta^{(1)} = \delta$ naturally come into the picture in the Taylor expansion of the log-likelihood. To be precise, the score $\ell^3_{f;\boldsymbol{\vartheta}^{(1)}_0}(x) = \ell^{3(1)}_{f;\boldsymbol{\vartheta}_0}(x) = \partial_\delta \log f^{\Pi}_{\boldsymbol{\vartheta}^{(1)}}(x)|_{\boldsymbol{\vartheta}^{(1)}_0}$ is supposed to provide a linear term $\tau_3\ell^3_{f;\boldsymbol{\vartheta}^{(1)}_0}(x)$ in the Taylor expansion of $\log f^{\Pi}_{\boldsymbol{\vartheta}^{(1)}_0+(0,0,\tau_3)'}(x)$ about $\log f^{\Pi}_{\boldsymbol{\vartheta}^{(1)}_0}(x)$. Since that linear term happens to be zero, the best approximation is provided by the quadratic term $\frac{\tau_3^2}{2}\partial^2_\delta \log f^{\Pi}_{\boldsymbol{\vartheta}^{(1)}}(x)|_{\boldsymbol{\vartheta}^{(1)}_0}$. The quantity $\frac{1}{2}\partial^2_\delta \log f^{\Pi}_{\boldsymbol{\vartheta}^{(1)}}(x)|_{\boldsymbol{\vartheta}^{(1)}_0}$ thus plays the role of a score function in that approximation, at $\boldsymbol{\vartheta}^{(1)}_0$—not for $\delta^{(1)}$, though, but for $(\delta^{(1)})^2$. Note indeed that, in view of (2.6),

$$\mathrm{E}_{\boldsymbol{\vartheta}^{(1)}_0}\left[\partial^2_\delta \log f^{\Pi}_{\boldsymbol{\vartheta}^{(1)}}(X)|_{\boldsymbol{\vartheta}^{(1)}_0}\right] = -\mathrm{E}_{\boldsymbol{\vartheta}^{(1)}_0}\left[\left(\partial_\delta \log f^{\Pi}_{\boldsymbol{\vartheta}^{(1)}}(X)|_{\boldsymbol{\vartheta}^{(1)}_0}\right)^2\right] = 0,$$

which is an essential property of score functions. As a result, if the impact, on the log-likelihood of an i.i.d. sample of size n, of a perturbation τ_3 of $\delta = 0$ is to exhibit the central-limit magnitude of $n^{-1/2}$, τ_3 itself has to be of magnitude $n^{-1/4}$ only; moreover, information about its sign is lost (a phenomenon which is also stressed by Rotnitzky *et al.* 2000). This is the structural reason for slower-than-$n^{1/2}$ consistency rates (at $\boldsymbol{\vartheta}_0^{(1)} = \boldsymbol{\vartheta}_0$) for the skewness parameter δ in the singular case: see the next section for details.

2.3. Towards a singularity-free reparametrization: second-order scores.

Second-order derivatives thus quite naturally enter the scene in case of degenerate Fisher information. The existence of derivatives of order two, however, requires reinforcing the regularity assumptions (A1) and (A2) on f and Π.

The reinforced regularity assumptions we need to reparametrize (at $\boldsymbol{\vartheta}_0^{(1)} = \boldsymbol{\vartheta}_0$) the family (1.1) are as follows—recall that we only address the case under which f and Π are such that $f = g_a \in \mathcal{E}_\Psi$ for some $a \in \mathcal{A}$ (see (2.4)): f thus is now entirely determined by Π and the constant a, and we only need strengthening (A2).

Assumption (A2$^+$). Same as (A2) but moreover (i) the mapping $(z, \delta) \mapsto \Pi(z, \delta)$ is twice continuously differentiable at $(z, 0)$, $z \in \mathbb{R}$; (ii) denoting by $z \mapsto \dot{\psi}(z) = \partial_\delta \partial_z \Pi(z, \delta)|_{\delta=0}$ the derivative of ψ, the quantities $\int_{-\infty}^{\infty} \psi^2(z) z^2 f(z) dz$ and $\int_{-\infty}^{\infty} (2a^{-1}\dot{\psi}(z) - 2\psi^2(z))^2 f(z) dz$ are finite.

Assumption (A2$^+$)(i) ensures the existence of the second derivative $\partial_\delta^2 f^{\Pi}_{\boldsymbol{\vartheta}^{(1)}}(x)|_{\boldsymbol{\vartheta}_0^{(1)}}$, while Assumption (A2$^+$)(ii) guarantees finiteness of the corresponding covariance matrix. Assumption (A2$^+$)(i) also entails $\partial_\delta \partial_z \Pi(z, \delta)|_{\delta=0} = \partial_z \partial_\delta \Pi(z, \delta)|_{\delta=0}$ for all $z \in \mathbb{R}$, so that this mixed derivative indeed coincides with $\dot{\psi}(z)$ (see (A2$^+$)(ii)). As already pointed out, Assumption (A2$^+$) not only reinforces (A2) but also, via the requirement that $f = g_a \in \mathcal{E}_\Psi$ for some $a \in \mathcal{A}$, entails (A1), which is no longer needed.

Now, in line with Section 2.1, and under Assumption (A2$^+$), let

$$
\begin{aligned}
\boldsymbol{\ell}_{f;\boldsymbol{\vartheta}_0^{(1)}}(x) &:= \left(\ell^1_{f;\boldsymbol{\vartheta}_0^{(1)}}(x), \ell^2_{f;\boldsymbol{\vartheta}_0^{(1)}}(x), \ell^3_{f;\boldsymbol{\vartheta}_0^{(1)}}(x)\right)' \qquad (2.7)\\
&:= \begin{pmatrix} \partial_{\mu^{(1)}} \log f^{\Pi}_{\boldsymbol{\vartheta}^{(1)}}(x)|_{\boldsymbol{\vartheta}_0^{(1)}} \\ \partial_{\sigma^{(1)}} \log f^{\Pi}_{\boldsymbol{\vartheta}^{(1)}}(x)|_{\boldsymbol{\vartheta}_0^{(1)}} \\ \frac{1}{2}\partial^2_{\delta^{(1)}} \log f^{\Pi}_{\boldsymbol{\vartheta}^{(1)}}(x)|_{\boldsymbol{\vartheta}_0^{(1)}} \end{pmatrix} = \begin{pmatrix} \sigma^{-1} a \psi(\sigma^{-1}(x-\mu)) \\ \sigma^{-1}\big(\sigma^{-1}(x-\mu) a \psi(\sigma^{-1}(x-\mu)) - 1\big) \\ \frac{2}{a}\dot{\psi}(\sigma^{-1}(x-\mu)) - 2\psi^2(\sigma^{-1}(x-\mu)) \end{pmatrix}
\end{aligned}
$$

with covariance

$$\boldsymbol{\Gamma}_{f;\boldsymbol{\vartheta}_0^{(1)}} := \sigma^{-1}\int_{-\infty}^{\infty} \boldsymbol{\ell}_{f;\boldsymbol{\vartheta}_0^{(1)}}(x)\boldsymbol{\ell}'_{f;\boldsymbol{\vartheta}_0^{(1)}}(x) f(\sigma^{-1}(x-\mu))dx =: \begin{pmatrix} \gamma^{11}_{f;\boldsymbol{\vartheta}_0^{(1)}} & 0 & 0 \\ 0 & \gamma^{22}_{f;\boldsymbol{\vartheta}_0^{(1)}} & \gamma^{23}_{f;\boldsymbol{\vartheta}_0^{(1)}} \\ 0 & \gamma^{23}_{f;\boldsymbol{\vartheta}_0^{(1)}} & \gamma^{33}_{f;\boldsymbol{\vartheta}_0^{(1)}} \end{pmatrix}$$

where (finiteness of the integrals below follows from (A2$^+$)(ii))

$$\gamma^{11}_{f;\boldsymbol{\vartheta}_0^{(1)}} = a^2\sigma^{-2}\int_{-\infty}^{\infty}\psi^2(z)f(z)dz, \qquad \gamma^{22}_{f;\boldsymbol{\vartheta}_0^{(1)}} = \sigma^{-2}\int_{-\infty}^{\infty}(a\psi(z)z-1)^2 f(z)dz,$$

$$\gamma^{33}_{f;\boldsymbol{\vartheta}_0^{(1)}} = 4\int_{-\infty}^{\infty}(a^{-1}\dot{\psi}(z)-\psi^2(z))^2 f(z)dz,$$

and

$$\gamma^{23}_{f;\boldsymbol{\vartheta}_0^{(1)}} = 2\sigma^{-1}\int_{-\infty}^{\infty}(a\psi(z)z-1)(a^{-1}\dot{\psi}(z)-\psi^2(z))f(z)dz.$$

First, let us assume that $\boldsymbol{\Gamma}_{f;\boldsymbol{\vartheta}_0^{(1)}}$ has full rank. Denoting by $X_1,\ldots,X_n$ an i.i.d. sample of size n from $f^{\Pi}_{\boldsymbol{\vartheta}_0^{(1)}}$, the vector $\boldsymbol{\ell}_{f;\boldsymbol{\vartheta}_0^{(1)}}$ defined in (2.7) provides a linear term, of the form $(\tau_1,\tau_2,\tau_3^2)\sum_{i=1}^n \boldsymbol{\ell}_{f;\boldsymbol{\vartheta}_0^{(1)}}(X_i)$, to the Taylor expansion of the log-likelihood $\sum_{i=1}^n \log f^{\Pi}_{\boldsymbol{\vartheta}_0^{(1)}+(\tau_1,\tau_2,\tau_3)'}(X_i)$ with respect to $\sum_{i=1}^n \log f^{\Pi}_{\boldsymbol{\vartheta}_0^{(1)}}(X_i)$. In order for that linear term to exhibit the required traditional central-limit behavior, the perturbation $\boldsymbol{\tau} := (\tau_1,\tau_2,\tau_3)'$ has to be of the order $(n^{-1/2},n^{-1/2},n^{-1/4})'$, hence must be of the form $\boldsymbol{\tau} = (n^{-1/2}t_1, n^{-1/2}t_2, n^{-1/4}t_3)'$, yielding $(t_1,t_2,t_3^2)n^{-1/2}\sum_{i=1}^n \boldsymbol{\ell}_{f;\boldsymbol{\vartheta}_0^{(1)}}(X_i)$ which, in view of the fact that $\boldsymbol{\ell}_{f;\boldsymbol{\vartheta}_0^{(1)}}(X_i)$ has expectation zero and finite full-rank variance $\boldsymbol{\Gamma}_{f;\boldsymbol{\vartheta}_0^{(1)}}$, is asymptotically normal under $\boldsymbol{\vartheta}_0^{(1)}$, as should be for the linear term of local log-likelihood expansions under the assumptions of the classical MLE theory.

This also naturally suggests a test rejecting the null hypothesis of symmetry (in favor of an asymmetry of unspecified sign) whenever the quadratic statistic (of the Lagrange Multiplier type; $\hat{\boldsymbol{\vartheta}}_0^{(1)} = (\hat{\mu},\hat{\sigma},0)$ stands for a root-n consistent estimator of $\boldsymbol{\vartheta}_0^{(1)}$ under $\delta = 0$)

$$n^{-1}\sum_{i=1}^{n}\left(\ell^3_{f;\hat{\boldsymbol{\vartheta}}_0^{(1)}}(X_i) - (\gamma^{23}_{f;\hat{\boldsymbol{\vartheta}}_0^{(1)}}/\gamma^{22}_{f;\hat{\boldsymbol{\vartheta}}_0^{(1)}})\ell^2_{f;\hat{\boldsymbol{\vartheta}}_0^{(1)}}(X_i)\right)^2\left(\gamma^{33}_{f;\hat{\boldsymbol{\vartheta}}_0^{(1)}} - (\gamma^{23}_{f;\hat{\boldsymbol{\vartheta}}_0^{(1)}})^2/\gamma^{22}_{f;\hat{\boldsymbol{\vartheta}}_0^{(1)}}\right)^{-1}$$

exceeds the chi-square quantile (one degree of freedom) of order $(1-\alpha)$. For all those reasons, the terminology "score vector" adequately can be used for $\boldsymbol{\ell}_{f;\boldsymbol{\vartheta}_0^{(1)}}$.

However, score vectors, in the classical MLE theory as well as in Le Cam's theory of locally asymptotically normal experiments, enjoy stronger properties, ensuring, in particular, the optimal nature of the test just described. Those properties rely on the quadratic approximation (as $n \to \infty$, under $\boldsymbol{\vartheta}_0^{(1)}$) of local log-likelihood ratios which, in the present case, should take the form

$$\sum_{i=1}^{n} \log f^{\Pi}_{\boldsymbol{\vartheta}^{(1)}+(n^{-1/2}t_1,n^{-1/2}t_2,n^{-1/4}t_3)'}(X_i)$$
$$= \sum_{i=1}^{n} \log f^{\Pi}_{\boldsymbol{\vartheta}_0^{(1)}}(X_i) + (t_1,t_2,t_3^2)n^{-1/2}\sum_{i=1}^{n} \boldsymbol{\ell}_{f;\boldsymbol{\vartheta}_0^{(1)}}(X_i) - \frac{1}{2}(t_1,t_2,t_3^2)\boldsymbol{\Gamma}_{f;\boldsymbol{\vartheta}_0^{(1)}}(t_1,t_2,t_3^2)' + o_{\mathrm{P}}(1)$$

where $\boldsymbol{\Gamma}_{f;\boldsymbol{\vartheta}_0^{(1)}}$ is the covariance matrix of $\boldsymbol{\ell}_{f;\boldsymbol{\vartheta}_0^{(1)}}$. This quadratic approximation does not hold here without additional assumptions on higher-order log-likelihood derivatives of orders three and four. This point is investigated in detail in Hallin, Ley and Monti (2012), for the particular case of the skew-normal, and we will not pursue it any further here.

We have assumed, so far, that $\boldsymbol{\Gamma}_{f;\boldsymbol{\vartheta}_0^{(1)}}$ has full rank. In most cases, the components of the new score vector $(\ell^1_{f;\boldsymbol{\vartheta}_0^{(1)}}, \ell^2_{f;\boldsymbol{\vartheta}_0^{(1)}}, \ell^3_{f;\boldsymbol{\vartheta}_0^{(1)}})'$ are not collinear anymore, so that $\boldsymbol{\Gamma}_{f;\boldsymbol{\vartheta}_0^{(1)}}$ indeed is non-singular; our objective of a singularity-free parametrization then is achieved, with consistency rate (for δ, at $\boldsymbol{\vartheta}_0$) $n^{1/4} = (n^{1/2})^{1/2}$. But this is not a general rule: in the case of the skew-normal family, for instance, Chiogna (2005) showed that the correct rate is only $n^{1/6}$. The explanation, as we shall see, lies in a *double singularity* phenomenon, which occurs when $\ell^2_{f;\boldsymbol{\vartheta}_0^{(1)}}$ and $\ell^3_{f;\boldsymbol{\vartheta}_0^{(1)}}$ in turn are collinear (by construction, the location score $\ell^1_{f;\boldsymbol{\vartheta}_0^{(1)}}$ is orthogonal to the other two).

3. The double singularity case.

3.1. Double singularity: a special role for Gaussian kernels.

The double singularity phenomenon thus takes place if and only if

$$b(az\psi(z)-1)/\sigma = (2/a)\dot{\psi}(z) - 2\psi^2(z) \quad \text{a.e.}$$

(a.e. here and in the sequel means Lebesgue-a.e.) for some constant $b \in \mathbb{R}$ and a couple (f, Π) such that $f \in \mathcal{E}_\Psi$ (see (2.4)). Rewriting this equation under the form

$$\dot{\psi}(z) = -\frac{ab}{2\sigma} + \frac{a^2 b}{2\sigma} z\psi(z) + a\psi^2(z) \quad \text{a.e.} \tag{3.8}$$

yields a classical Ricatti equation, whose solutions are of the form

$$\psi(z) = \frac{-ab}{2\sigma} z \tag{3.9}$$

or

$$\psi(z) = \frac{-ab}{2\sigma} z + \exp\left(-\frac{a^2bz^2}{4\sigma}\right) \Big/ \left(c - a\int_0^z \exp\left(-\frac{a^2by^2}{4\sigma}\right) dy\right) \quad b, c \in \mathbb{R}. \tag{3.10}$$

First, note that b has to be negative, as otherwise $\varphi_f(z) = a\psi(z)$ would tend to $-\infty$ irrespective of the sign of a when $z \to \infty$, implying positive values of $\dot{f}$ in the right tail of f, which is of course impossible for a density function. Furthermore, since both $z \mapsto a \int_0^z \exp\left(-\frac{a^2by^2}{4\sigma}\right) dy$ and ψ are odd, the constant c in (3.10) has to be zero. By (2.4), the natural parameter space $\mathcal{A}$ for the exponential family $\mathcal{E}_\Psi$ associated with the mapping ψ of (3.10) then consists of the set of values of a for which the integral

$$\begin{aligned}\int_{-\infty}^{\infty} \exp(-a\Psi(z))dz &= \int_{-\infty}^{\infty} \exp\left(\frac{a^2b}{4\sigma}z^2 + \log\left|\int_0^z \exp\left(-\frac{a^2by^2}{4\sigma}\right) dy\right|\right) dz \\ &= \int_{-\infty}^{\infty} \exp\left(\frac{a^2b}{4\sigma}z^2\right)\left|\int_0^z \exp\left(-\frac{a^2b}{4\sigma}y^2\right) dy\right| dz\end{aligned}$$

is finite. After a change of variable involving the quantity $\sqrt{a^2|b|/(4\sigma)}$, this appears to be equivalent to the requirement

$$\int_{-\infty}^{\infty} \exp(-z^2)\left|\int_0^z \exp(y^2)dy\right| dz < \infty. \tag{3.11}$$

However, one easily can check that $\lim_{z\to\infty} z\exp(-z^2)\left|\int_0^z \exp(y^2)dy\right| = 1/2$, meaning that $\exp(-z^2)\left|\int_0^z \exp(y^2)dy\right|$ behaves as $1/z$ for large values of z. It follows that (3.11) is impossible. Hence, the natural parameter space $\mathcal{A}$ is empty, meaning that no symmetric kernel f associated to the mapping ψ of (3.10) can yield singular Fisher information. Therefore, the only admissible solution to (3.8) is (3.9).

This finding is quite remarkable: combined with the fact that $f \in \mathcal{E}_\Psi$ (which is equivalent to $\varphi_f = a\psi$), it implies that double singularity only can occur for symmetric kernels f such that $\varphi_f(z) = c_1 z$ for some constant c_1—namely, for Gaussian kernels; those Gaussian kernels moreover should be combined with a skewing function Π such that $\psi(z) = c_2 z$ for some constant c_2.

While Fisher singularity arises as a mismatch between the symmetric kernel and the skewing function, and hence can occur with all possible symmetric kernels, the double singularity phenomenon is specific to the Gaussian kernel, hence to a well-determined subclass of *generalized skew-normal distributions* (in the sense of Loperfido 2004). This also implies that, under the assumptions made, $n^{1/4}$ consistency rates are achieved for all other skew-symmetric families subject to Fisher singularity.

We formalize that result in the following theorem.

Theorem 3.1. *Consider the skew-symmetric family defined in (1.1). Then,*

(i) under Assumptions (A1) and (A2), the couple (f, Π) leads to a skew-symmetric family subject to Fisher singularity at $\delta = 0$ if and only if the symmetric kernel f is related to the skewing function Π via the fact that $f \in \mathcal{E}_\Psi$, see (2.4);

(ii) under Assumption (A2$^+$), the couple (f, Π) leads to a skew-symmetric family subject to the double singularity phenomenon if and only if the symmetric kernel f is the normal kernel ϕ and the skewing function Π moreover satisfies $\psi(z) := \partial_\delta \Pi(z,\delta)|_{\delta=0} = cz$ for some real constant c; the family then is a particular case of the generalized skew-normal family (Loperfido 2004).

This theorem completely characterizes the double singularity problem, hence complements the simple singularity characterization of Hallin and Ley (2012).

3.2. A singularity-free reparametrization.

Still inspired by Rotnitzky *et al.* (2000), let us now proceed with this second singularity the way we did with the first one, producing a second, hopefully singularity-free, reparametrization. Since the symmetric kernel ϕ is the only candidate for this double singularity phenomenon, we can limit ourselves to $f = \phi$. Moreover, we know from the previous section that $\psi(z) = c_2 z$; hence, in view of the fact that $z = \varphi_\phi(z) = a\psi(z)$, we have $c_2 = 1/a$. Applying the same Gram-Schmidt process as in Section 2.2, but with the score for scale $\ell^2_{\phi;\boldsymbol{\vartheta}_0^{(1)}}$ substituted for the score for location, we project $\ell^3_{\phi;\boldsymbol{\vartheta}_0^{(1)}}$ onto the subspace orthogonal to $\ell^1_{\phi;\boldsymbol{\vartheta}_0^{(1)}}$ and $\ell^2_{\phi;\boldsymbol{\vartheta}_0^{(1)}}$. The resulting residual score for skewness then, as expected, is zero:

$$
\begin{aligned}
&\ell^3_{\phi;\boldsymbol{\vartheta}_0^{(1)}}(x) - \ell^2_{\phi;\boldsymbol{\vartheta}_0^{(1)}}(x)\mathrm{Cov}(\ell^2_{\phi;\boldsymbol{\vartheta}_0^{(1)}}, \ell^3_{\phi;\boldsymbol{\vartheta}_0^{(1)}})/\mathrm{Var}(\ell^2_{\phi;\boldsymbol{\vartheta}_0^{(1)}}) \\
&= \frac{2}{a^2} - \frac{2}{a^2}\big((x-\mu)/\sigma\big)^2 - \sigma^{-1}\big(((x-\mu)/\sigma)^2 - 1\big)\frac{2\sigma^{-1}\int_{-\infty}^{\infty}(z^2-1)(a^{-2} - a^{-2}z^2)\phi(z)dz}{\sigma^{-2}\int_{-\infty}^{\infty}(z^2-1)^2\phi(z)dz} \\
&= 0.
\end{aligned}
$$

Transposing, as in Section 2.2, this projection in terms of parameters leads to the reparametrization $\boldsymbol{\vartheta}^{(2)} := (\mu^{(2)}, \sigma^{(2)}, \delta^{(2)})'$, where

$$
\mu^{(2)} = \mu^{(1)} = \mu + 2\delta\sigma/a, \quad \sigma^{(2)} = \sigma^{(1)} + \delta^2 \frac{\mathrm{Cov}(\ell^2_{\phi;\boldsymbol{\vartheta}_0^{(1)}}, \ell^3_{\phi;\boldsymbol{\vartheta}_0^{(1)}})}{\mathrm{Var}(\ell^2_{\phi;\boldsymbol{\vartheta}_0^{(1)}})} = \sigma^{(1)}(1 - 2\delta^2/a^2),
$$

and

$$
\delta^{(2)} = \delta^{(1)} = \delta.
$$

In line with previous notations, we denote by $f^{\Pi}_{\boldsymbol{\vartheta}^{(2)}}$ the resulting skew-symmetric density despite the fact that the symmetric kernel is ϕ. It is easy to check that our reparametrization, in the skew-normal case, coincides with that of Chiogna (2005).

This second reparametrization solely affects the scale parameter, but again cancels the score for skewness. Thus, derivatives of order three with respect to $\delta^{(2)} = \delta$ come into the picture, which eventually will lead to $n^{1/6}$ consistency rates. This, however, requires a reinforcement of Assumption (A2$^+$).

ASSUMPTION (A2^{++}). Same as (A2$^+$), but now (i) the mapping $(z, \delta) \mapsto \Pi(z, \delta)$ is three times continuously differentiable at $(z, 0)$ for all $z \in \mathbb{R}$; (ii) letting $\Upsilon(z) := \partial_\delta^3 \Pi(z, \delta)|_{\delta=0}$, $\int_{-\infty}^{\infty} \left(\frac{8}{3a^3} z^3 - \frac{8}{a^3} z + \frac{1}{3}\Upsilon(z)\right)^2 \phi(z) dz$ is finite.

Assumption (A2^{++})(i) ensures the existence of the third-order derivative $\partial_\delta^3 f^\Pi_{\boldsymbol{\vartheta}^{(2)}}$ at $\boldsymbol{\vartheta}_0^{(2)} = (\mu^{(2)}, \sigma^{(2)}, 0)' = (\mu, \sigma, 0)' = \boldsymbol{\vartheta}_0$, while Assumption (A2^{++})(ii) guarantees finiteness of the corresponding covariance matrix. Also note that the mixed derivative $\partial_z \partial_\delta^2 \Pi(z, \delta)|_{\delta=0} = 0$ by definition of skewing functions, and that $\partial_z^2 \partial_\delta \Pi(z, \delta)|_{\delta=0} = \partial_z^2 \psi(z)$ vanishes for all z, since we are dealing (Theorem 3.1(ii)) with skewing functions such that $\psi(z) = z/a$ is linear. These facts greatly simplify calculations.

Assumption (A2^{++}) thus implies, for this second reparametrization, the existence, at $\boldsymbol{\vartheta}_0^{(2)}$, of a third-order score vector $\boldsymbol{\ell}_{\phi;\boldsymbol{\vartheta}_0^{(2)}}$ with finite covariance matrix $\boldsymbol{\Gamma}_{\phi;\boldsymbol{\vartheta}_0^{(2)}}$, enjoying the same properties as the second-order score described in Section 2.3, now with rates $n^{1/6}$. Elementary algebra yields

$$\boldsymbol{\ell}_{\phi;\boldsymbol{\vartheta}_0^{(2)}}(x) := \begin{pmatrix} \ell^1_{\phi;\boldsymbol{\vartheta}_0^{(2)}} \\ \ell^2_{\phi;\boldsymbol{\vartheta}_0^{(2)}} \\ \ell^3_{\phi;\boldsymbol{\vartheta}_0^{(2)}} \end{pmatrix} := \begin{pmatrix} \partial_{\mu^{(2)}} \log f^\Pi_{\boldsymbol{\vartheta}^{(2)}}(x)|_{\boldsymbol{\vartheta}_0^{(2)}} \\ \partial_{\sigma^{(2)}} \log f^\Pi_{\boldsymbol{\vartheta}^{(2)}}(x)|_{\boldsymbol{\vartheta}_0^{(2)}} \\ \frac{1}{6}\partial^3_{\delta^{(2)}} \log f^\Pi_{\boldsymbol{\vartheta}^{(2)}}(x)|_{\boldsymbol{\vartheta}_0^{(2)}} \end{pmatrix}$$
$$= \begin{pmatrix} \sigma^{-1}\left(\sigma^{-1}(x-\mu)\right) \\ \sigma^{-1}\left((\sigma^{-1}(x-\mu))^2 - 1\right) \\ \frac{8}{3a^3}\left(\sigma^{-1}(x-\mu)\right)^3 - \frac{8}{a^3}\sigma^{-1}(x-\mu) + \frac{1}{3}\Upsilon\left(\sigma^{-1}(x-\mu)\right) \end{pmatrix}$$

and

$$\boldsymbol{\Gamma}_{\phi;\boldsymbol{\vartheta}_0^{(2)}} := \sigma^{-1}\int_{-\infty}^{\infty} \boldsymbol{\ell}_{\phi;\boldsymbol{\vartheta}_0^{(2)}}(x)\boldsymbol{\ell}'_{\phi;\boldsymbol{\vartheta}_0^{(2)}}(x)\phi(\sigma^{-1}(x-\mu))dx =: \begin{pmatrix} \gamma^{11}_{\phi;\boldsymbol{\vartheta}_0^{(2)}} & 0 & \gamma^{13}_{\phi;\boldsymbol{\vartheta}_0^{(2)}} \\ 0 & \gamma^{22}_{\phi;\boldsymbol{\vartheta}_0^{(2)}} & \gamma^{23}_{\phi;\boldsymbol{\vartheta}_0^{(2)}} \\ \gamma^{13}_{\phi;\boldsymbol{\vartheta}_0^{(2)}} & \gamma^{23}_{\phi;\boldsymbol{\vartheta}_0^{(2)}} & \gamma^{33}_{\phi;\boldsymbol{\vartheta}_0^{(2)}} \end{pmatrix},$$

with

$$\gamma^{11}_{\phi;\boldsymbol{\vartheta}_0^{(2)}} = \sigma^{-2}\int_{-\infty}^{\infty} z^2\phi(z)dz = \sigma^{-2}, \qquad \gamma^{22}_{\phi;\boldsymbol{\vartheta}_0^{(2)}} = \sigma^{-2}\int_{-\infty}^{\infty} (z^2-1)^2\phi(z)dz = 2\sigma^{-2},$$

$$\gamma^{33}_{\phi;\boldsymbol{\vartheta}_0^{(2)}} = \int_{-\infty}^{\infty}\left(\frac{8}{3a^3}z^3 - \frac{8}{a^3}z + \frac{1}{3}\Upsilon(z)\right)^2\phi(z)dz,$$

$$\gamma^{13}_{\phi;\boldsymbol{\vartheta}_0^{(2)}} = \sigma^{-1}\int_{-\infty}^{\infty} z\left(\frac{8}{3a^3}z^3 - \frac{8}{a^3}z + \frac{1}{3}\Upsilon(z)\right)\phi(z)dz,$$

and

$$\gamma^{23}_{\phi;\boldsymbol{\vartheta}_0^{(2)}} = \sigma^{-1}\int_{-\infty}^{\infty} (z^2-1)\left(\frac{8}{3a^3}z^3 - \frac{8}{a^3}z + \frac{1}{3}\Upsilon(z)\right)\phi(z)dz.$$

If we assume, as is Section 2.3, that $\boldsymbol{\Gamma}_{\phi;\boldsymbol{\vartheta}_0^{(2)}}$ has full rank, denoting by $X_1, \ldots, X_n$ an i.i.d. sample of size n from $f^{\Pi}_{\boldsymbol{\vartheta}_0^{(2)}}$, the score vector $\boldsymbol{\ell}_{\phi;\boldsymbol{\vartheta}_0^{(2)}}$ provides a linear term to the Taylor expansion of the log-likelihood, as well as a Lagrange multiplier-type test of the null hypothesis of symmetry (in the generalized skew-normal family under study), based on the quadratic test statistic

$$n^{-1}\sum_{i=1}^{n}\left(\ell^3_{\phi;\hat{\boldsymbol{\vartheta}}_0^{(2)}}(X_i) - \sigma^2\big(\gamma^{13}_{\phi;\hat{\boldsymbol{\vartheta}}_0^{(2)}},\quad \gamma^{23}_{\phi;\hat{\boldsymbol{\vartheta}}_0^{(2)}}/2\big)\begin{pmatrix}\ell^1_{\phi;\hat{\boldsymbol{\vartheta}}_0^{(2)}}(X_i)\\ \ell^2_{\phi;\hat{\boldsymbol{\vartheta}}_0^{(2)}}(X_i)\end{pmatrix}\right)^2$$
$$\times\Big(\gamma^{33}_{\phi;\hat{\boldsymbol{\vartheta}}_0^{(2)}} - \sigma^2(\gamma^{13}_{\phi;\hat{\boldsymbol{\vartheta}}_0^{(2)}})^2 - \sigma^2(\gamma^{23}_{\phi;\hat{\boldsymbol{\vartheta}}_0^{(2)}})^2/2\Big)^{-1},$$

where $\hat{\boldsymbol{\vartheta}}_0^{(2)}$ is, under the null hypothesis of symmetry, a root-n consistent estimator of location and scale. The consistency/contiguity rate for δ (still, at $\delta = 0$) is $n^{1/6}$, and the same comments as in Section 2.3 are in order. The particular case of the skew-normal family is studied in full detail in Hallin, Ley and Monti (2012).

4. Higher-order singularities.

It may happen, however, that $\boldsymbol{\Gamma}_{\phi;\boldsymbol{\vartheta}_0^{(2)}}$ in turn is singular, the new third-order score for skewness $\ell^3_{\phi;\boldsymbol{\vartheta}_0^{(2)}}$ being (at $\boldsymbol{\vartheta}_0^{(2)}$) a linear combination of the scores for location $\ell^1_{\phi;\boldsymbol{\vartheta}_0^{(2)}}$ and scale $\ell^2_{\phi;\boldsymbol{\vartheta}_0^{(2)}}$. If this occurs, one has to go yet one step further with the approximation of log-likelihoods, assuming the existence of fourth-order derivatives and ending up with $n^{1/8}$ consistency/contiguity rates. That $n^{1/8}$ rate, however, as we shall see, is the worst possible one. Since this last derivation is not the main aim of this paper, we will voluntarily alleviate the reading and spare the reader computational details and the diverse steps which we have sufficiently described in the previous cases.

In order for $\ell^3_{\phi;\boldsymbol{\vartheta}_0^{(2)}} = \frac{8}{3a^3}z^3 - \frac{8}{a^3}z + \frac{1}{3}\Upsilon(z)$ to be a linear combination of $\ell^1_{\phi;\boldsymbol{\vartheta}_0^{(2)}} = z/\sigma$ and $\ell^2_{\phi;\boldsymbol{\vartheta}_0^{(2)}} = (z^2-1)/\sigma$, $\Upsilon(z)$ necessarily has be of the form $\alpha_1(-1+z^2) + \alpha_2 z + \alpha_3 z^3$, with $\alpha_1, \alpha_2 \in \mathbb{R}$ and $\alpha_3 = -\frac{8}{a^3}$ in order to annihilate the term in z^3. This condition on the third derivative w.r.t. δ thus characterizes what we would call a *triple singularity* (the result is formally stated in Theorem 4.1 at the end of this section). It is quite easy to construct examples suffering from this peculiarity; see Section 5.4.

At this stage, the by now familiar machinery *new singularity—Gram-Schmidt orthogonalization of scores—reparametrization—new higher-order score for* δ applies, leading after some direct manipulations to the reparametrization $\boldsymbol{\vartheta}^{(3)} := (\mu^{(3)}, \sigma^{(3)}, \delta^{(3)})'$, with

$$\mu^{(3)} = \mu^{(2)} + \left(-\frac{8}{a^3} + \frac{\alpha_2}{3}\right)\sigma\delta^3 = \mu + \frac{2}{a}\sigma\delta + \left(-\frac{8}{a^3} + \frac{\alpha_2}{3}\right)\sigma\delta^3,$$

$$\sigma^{(3)} = \sigma^{(2)} + \frac{\alpha_1}{3}\sigma\delta^3 = \sigma(1 - 2\delta^2/a^2 + \frac{\alpha_1}{3}\delta^3),$$

and

$$\delta^{(3)} = \delta^{(2)} = \delta^{(1)} = \delta.$$

Since this reparametrization annihilates the third-order score for skewness, we need to take fourth-order derivatives, which requires the following strengthening of Assumption (A2^{++}).

Assumption (A2^{+++}). Same as (A2^{++}), but now the mapping $(z, \delta) \mapsto \Pi(z, \delta)$ is four times continuously differentiable at $(z, 0)$, $z \in \mathbb{R}$.

Let us remark that, as will be seen below, we do not need to assume finiteness of Fisher information for skewness, as this will always be the case after this third reparametrization. Clearly, as in all previous cases, both the location score $\ell^1_{\phi;\boldsymbol{\vartheta}_0^{(3)}}$ and the scale score $\ell^2_{\phi;\boldsymbol{\vartheta}_0^{(3)}}$ remain the same as in the original parametrization, and the new fourth-order score for skewness, for skewing functions such that $\partial^3_\delta \Pi(z, \delta)|_{\delta=0} = \alpha_1(-1 + z^2) + \alpha_2 z - \frac{8}{a^3}z^3$, becomes (after very lengthy but elementary calculations)

$$\begin{aligned}
\ell^3_{\phi;\boldsymbol{\vartheta}_0^{(3)}} &= \frac{1}{24}\partial^4_{\delta^{(3)}} \log f^{\Pi}_{\boldsymbol{\vartheta}^{(3)}}(x)|_{\boldsymbol{\vartheta}_0^{(3)}} \\
&= -\frac{10}{a^4} + \frac{2\alpha_2}{3a} + \frac{2\alpha_1}{a}\left(\frac{x-\mu}{\sigma}\right) + \left(\frac{6}{a^4} - \frac{2\alpha_2}{3a}\right)\left(\frac{x-\mu}{\sigma}\right)^2 - \frac{2\alpha_1}{3a}\left(\frac{x-\mu}{\sigma}\right)^3 \\
&\quad + \frac{4}{3a^4}\left(\frac{x-\mu}{\sigma}\right)^4.
\end{aligned}$$

One again easily can check that this quantity is centered under $\boldsymbol{\vartheta}_0^{(3)} = \boldsymbol{\vartheta}_0 = (\mu, \sigma, 0)'$. The interesting feature here is that the term $\frac{4}{3a^4}\left(\frac{x-\mu}{\sigma}\right)^4$ can by no means be annihilated, and hence hampers any linear combination with the location and scale scores. Thus, the resulting Fisher information matrix (whose finiteness is obvious)

$$\begin{aligned}
\boldsymbol{\Gamma}_{\phi;\boldsymbol{\vartheta}_0^{(3)}} &:= \sigma^{-1}\int_{-\infty}^{\infty} \boldsymbol{\ell}_{\phi;\boldsymbol{\vartheta}_0^{(3)}}(x)\boldsymbol{\ell}'_{\phi;\boldsymbol{\vartheta}_0^{(3)}}(x)\phi(\sigma^{-1}(x-\mu))dx \\
&= \begin{pmatrix} \sigma^{-2} & 0 & -\frac{46\alpha_1}{\sigma a} \\ 0 & 2\sigma^{-2} & \sigma^{-1}\left(\frac{28}{a^4} - \frac{4\alpha_2}{3a}\right) \\ -\frac{46\alpha_1}{\sigma a} & \sigma^{-1}\left(\frac{28}{a^4} - \frac{4\alpha_2}{3a}\right) & \frac{1304}{3a^8} - \frac{112\alpha_2}{3a^5} + \frac{24\alpha_1^2 + 8\alpha_2^2}{9a^2} \end{pmatrix}
\end{aligned}$$

cannot be singular, which in turn implies that $n^{1/8}$ rates of convergence are the worst possible! The structural reason behind this result lies in the fact that, by definition of skewing functions, $\partial_\delta^4 \Pi(z,\delta)|_{\delta=0}$ equals zero, hence cannot interfere in the fourth derivative, contrarily to $\partial_\delta^3 \Pi(z,\delta)|_{\delta=0}$ which plays the crucial role in annihilating the third-order derivative.

Those results are summarized in the following theorem, which complements Theorem 3.1.

Theorem 4.1. *Consider the skew-symmetric family defined in (1.1). Then,*

(i) under Assumption (A2^{++}), the couple (f, Π) leads to a skew-symmetric family subject to the third singularity phenomenon if and only if the symmetric kernel f is the normal kernel ϕ and the skewing function Π moreover satisfies $\psi(z) := \partial_\delta \Pi(z,\delta)|_{\delta=0} = z/a$ for some non-zero real constant a and $\Upsilon(z) := \partial_\delta^3 \Pi(z,\delta)|_{\delta=0} = \alpha_1(-1+z^2) + \alpha_2 z - \frac{8}{a^3} z^3$ for some real constants α_1 and α_2, both possibly zero.

(ii) under Assumption (A2^{+++}), the couple (f, Π) leads to no skew-symmetric family subject to a fourfold/quadruple singularity phenomenon.

We conclude this section by noting that, in most cases (including all classical skewing functions described in Section 5.5 hereafter), Υ is an odd function, implying some simplifications in the above expressions (namely α_1 then equals 0), but clearly the final outcome does not alter.

5. Examples.

In this section, we illustrate our findings on basis of some well-known examples of the literature. Our presentation goes *crescendo*: starting, for the sake of completeness, with singularity-free families, we consider simple, double, and finally triple singularities.

5.1. Singularity-free families.

Famous singularity-free examples comprise, *inter alia*, the skew-exponential power distributions of Azzalini (1986) with pdf $2\,c^{-1}\exp(-|z|^\alpha/\alpha)\Phi(\delta\,\mathrm{sign}(z)|z|^{\alpha/2}(2/\alpha)^{1/2})$ for $\alpha > 1$ and $c = 2\alpha^{1/\alpha-1}\Gamma(1/\alpha)$, and the skew-$t$ distributions of Azzalini and Capitanio (2003) with pdf $2t_\nu(z)T_{\nu+1}(\delta z(\nu+1)^{1/2}(z^2+\nu)^{-1/2})$ where t_η and T_η respectively stand for the pdf and cdf of a standard Student distribution with η degrees of freedom. These examples are discussed at length in Hallin and Ley (2012), where we refer to for details. In that same paper, an example of skewing function for which no mismatching symmetric kernel exists is given, namely $\Pi(z,\delta) = \Pi(\delta \sin(z))$ with $\Pi : \mathbb{R} \to [0,1]$ a differentiable function satisfying $\Pi(-y) + \Pi(y) = 1$ for all $y \in \mathbb{R}$ and such that $\dot{\Pi}(0) = d\Pi(y)/dy|_{y=0}$ exists and differs from zero.

5.2. Simple singularities.

As shown in Hallin and Ley (2012), the easiest-to-construct mismatching skewing function for a given symmetric kernel f is of the form $\Pi(\delta\varphi_f(z))$, with Π as described above. For any symmetric kernel f, it is readily seen that the location and skewness scores then are collinear.

Under the assumptions made, double singularity requires the additional assumption that $\ddot{\Pi}(0) := d^2\Pi(y)/(dy)^2|_{y=0}$ exists and, by construction, equals zero. Theorem 3.1 then tells us that among the pdfs $2f(z)\Pi(\delta\varphi_f(z))$ only the skew-normal, obtained for $f = \phi$, suffers from the double singularity. Thus all non-Gaussian kernels f yield examples of simple singularities.

5.3. Double singularities.

Concerning the double singularity, a prominent example is of course Azzalini's skew-normal family, with pdf $2\phi(z)\Phi(\delta z)$. Let us briefly show that higher-order singularities are excluded in that family. Straightforward calculations yield $a = \sqrt{2\pi}$ and $\Upsilon(z) = -(2\pi)^{-1/2}z^3$, which is different from $-\frac{8}{a^3} = -(2/\pi)^{3/2}$, hence Theorem 4.1 readily yields the well-known result of $n^{1/6}$ rates of convergence for the skew-normal distribution. For the sake of completeness, we also provide for this famous example the corresponding score for skewness, which equals $\frac{4-\pi}{3\pi\sqrt{2\pi}}z^3 - \frac{4}{\pi\sqrt{2\pi}}z$.

Nadarajah and Kotz (2003) propose another family of skew densities generated by the normal kernel, with pdfs of the form $2\phi(z)G(\delta z)$ where G is some univariate symmetric cdf. They call *skew normal-G* the resulting families of densities. Their definition includes as particular cases the skew normal-normal model, the skew normal-t, the skew normal-Cauchy, the skew normal-Laplace, the skew normal-logistic and the skew normal-uniform families. Theorem 3.1 tells us that all skew normal-G models suffer from the double singularity, a fact that, except of course for the skew normal-normal (which, up to an additional scale parameter, coincides with the classical skew-normal), has never been noticed. Consequently, these models have to be treated with much care when used for inferential purposes. The problem with those families obviously stems from the product δz inside G; see Section 5.5 for further discussion of such skewing functions.

5.4. Higher-order singularities.

Let us further analyze the families of Nadarajah and Kotz (2003). Assume that G is three times continuously differentiable. Elementary calculations show that $a = 1/g(0)$, where $g(z) := dG(z)/dz$, and $\Upsilon(z) = \ddot{g}(0)z^3$. We know from Theorem 4.1 that a triple singularity can only occur if $\ddot{g}(0) = -\frac{8}{a^3} = -8(g(0))^3$. Among the distributions considered by Nadarajah and Kotz (2003), this equality holds for the skew normal-logistic only, for which $g(0) = 1/4$ and $\ddot{g}(0) = -1/8$. Thus, while all their other skew normal-G distributions have $n^{1/6}$ rates of convergence, the skew normal-logistic requires the worst possible rates, namely $n^{1/8}$ rates.

Finally, consider the "lifted" skew-normal distribution with pdf

$$2\phi(z)\Phi(\delta z - (4-\pi)(6\pi)^{-1}\delta^3 z^3). \tag{5.12}$$

Here, $a = \sqrt{2\pi}$ and $\Upsilon(z) = -(2/\pi)^{3/2}z^3 = -\frac{8}{(\sqrt{2\pi})^3}z^3 = -\frac{8}{a^3}z^3$, entailing, by Theorem 4.1, a triple singularity and hence $n^{1/8}$ rates of convergence. Note that this distribution is part of the so-called *flexible generalized skew-normal distributions* defined in Ma and Genton (2004). More generally, in that paper, the authors have proposed *flexible skew-symmetric distributions* with skewing functions of the form $\Pi(z,\delta) := \Pi(H_\ell(\delta z))$, with Π as defined in Section 5.1 and H_ℓ an odd polynomial of order ℓ (meaning that the polynomial only contains odd terms). Since, in the first four derivatives, all terms of the form $(\delta z)^s$ with odd $s \geq 5$ do not play any role, one can directly construct an infinity of flexible generalized skew-normal distributions suffering from triple singularity: take, for instance, an odd polynomial H_ℓ with the terms in δz and $(\delta z)^3$ as in (5.12), such as

$$2\phi(z)\Phi(\delta z - (4-\pi)(6\pi)^{-1}\delta^3 z^3 + \sum_{i=2}^{\ell} \alpha_{2i+1}(\delta z)^{2i+1})$$

with $\alpha_i \in \mathbb{R}$ and $2 \leq \ell \in \mathbb{N}$.

5.5. A brief discussion of skewing functions of the form $\Pi(z,\delta) = \Pi(\delta z)$.

As announced in the Introduction, we conclude this paper with a few comments on the most frequent type of skewing function, namely $\Pi(z,\delta) = \Pi(\delta z)$ with $\Pi : \mathbb{R} \to [0,1]$ satisfying $\Pi(-y) + \Pi(y) = 1$ for all $y \in \mathbb{R}$ (and satisfying the required differentiability conditions). Such functions are the most natural examples of a skewing function such that $\psi(z)$ is linear, yielding an extremely risky combination with the Gaussian kernel ϕ.

The original skew-normal family of Azzalini (1985) is based on $\Pi = \Phi$; in conjunction with a Gaussian kernel, the same type of skewing function has been used, *inter alia*, by

- Azzalini and Capitanio (1999) for their skew-symmetric densities of the form $2f(z)F(\delta z)$, with F the cdf corresponding to f;
- Gupta *et al.* (2002) for their skew-uniform, skew-t, skew-Cauchy, skew-Laplace and skew-logistic distributions, which all are special cases of Azzalini and Capitanio (1999)'s construction;
- Nadarajah and Kotz (2003) for their skew normal-G distributions, as described in the previous sections; and by
- Gómez *et al.* (2007) for their skew g-normal densities $2g(z)\Phi(\delta z)$ where, contrary to the skew normal-G distributions, normality is present in the skewing function and not in the symmetric kernel.

As shown in this paper, skewing functions of the form $\Pi(\delta z)$ are harmless whenever the symmetric kernel is not Gaussian. In view of this, the skew g-normal distributions (free of any singularity except for $g = \phi$) are inferentially preferable to the skew normal-G ones (which at least exhibit double singularity). The peculiarities of the skew-normal distribution, which belongs to all of the above-cited classes of distributions, have been discussed in length in the literature; we hope that this paper sheds some light on the structural reasons behind these inferential drawbacks, and warns the reader about the dangers of combining a Gaussian kernel with a skewing function of the form $\Pi(\delta z)$.

Azzalini and Capitanio (2003) clearly were aware of the dangers of using $\Pi(z, \delta)$ of the form $\Pi(\delta z)$: in reaction to a referee's remark, they write *A reviewer of this paper has remarked that, if we set* $d = 1$, *density (26) does not reduce to the form* $2t_1(y; \nu)T_1(\alpha y; \nu)$, *which seems to be the "most natural" univariate form of skew t density generated by Lemma 1 of Azzalini (1985)*, explain why the skewing functions they are proposing for their skew-t densities are not of that type, and suggest that the choice of $\Phi(\delta z)$ for the original skew-normal perhaps was not the best one. Our results amply justify their concern, and confirm the clear-sightedness of their diagnosis.

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.

Christophe Ley thanks the Fonds National de la Recherche Scientifique, Communauté française de Belgique, for support via a Mandat de Chargé de Recherche FNRS.

References

[1] Arnold, B. C. and Beaver, R. J. (2000) The skew-Cauchy distribution. *Statist. Probab. Lett.*, 49, 285–290.

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

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

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

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

[6] 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.

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

[8] 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.

[9] Cox, D. R. and Hinkley, D. V. (1974) *Theoretical statistics*. London: Chapman & Hall.

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

[11] Genton, M. G. (2004) *Skew-elliptical Distributions and their Applications: a Journey beyond Normality*. Boca Raton, FL: Chapman and Hall/CRC.

[12] 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.

[13] Gupta, A. K., Chang, F. C. and Huang, W. J. (2002) Some skew-symmetric models. *Random Oper. Stochastic Equations*, 10, 133–140.

[14] Hallin, M. and Ley, C. (2012) Skew-symmetric distributions and Fisher information – a tale of two densities. *Bernoulli*, 18, 747–763.

[15] Hallin, M., Ley, C. and Monti, A. C. (2012) Le Cam optimal tests for normality against skew-normal alternatives. *Work in progress.*

[16] Lee, L. E. and Chesher, A. (1986) Specification testing when score test statistics are identically zero. *Journal of Econometrics*, 31, 121–149.

[17] Ley, C. (2012) Skew distributions. *In A. El-Shaarawi and W. Piegorsch, Editors, Statistical Theory and Methods, Encyclopedia of Environmetrics,* 2nd edition, Wiley, New York, to appear.

[18] Ley, C. and Paindaveine, D. (2010) On the Singularity of Multivariate Skew-Symmetric Models. *J. Multivariate Anal.*, 101, 1434–1444.

[19] Loperfido, N. (2004) Generalized skew-normal distributions. *In M. G. Genton, ed., Skew-elliptical Distributions and their Applications: a Journey beyond Normality,* Boca Raton, FL: Chapman and Hall/CRC, 65–80.

[20] Ma, Y. and Genton, M. G. (2004) Flexible class of skew-symmetric distributions. *Scand. J. Statist.*, 31, 459–468.

[21] Nadarajah, S. and Kotz, S. (2003) Skewed distributions generated by the normal kernel. *Statist. Prob. Lett.*, 65, 269–277.

[22] Pewsey, A. (2000) Problems of inference for Azzalini's skew-normal distribution. *J. Appl. Statist.*, 27, 859–870.

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

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