Bernoulli **20**(3), 2014, 1432–1453
DOI: 10.3150/13-BEJ528

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

MARC HALLIN[1] and CHRISTOPHE LEY[2]

[1]*ECARES, Université Libre de Bruxelles, CP114, 50 Ave. F.D. Roosevelt, 1050 Brussels, Belgium, and ORFE, Princeton University, Sherrerd Hall, Princeton, NJ 08544, USA. E-mail: mhallin@ulb.ac.be*
[2]*ECARES and Département de Mathématique, Université Libre de Bruxelles, CP210, Boulevard du Triomphe, 1050 Brussels, Belgium. E-mail: chrisley@ulb.ac.be*

Hallin and Ley [*Bernoulli* **18** (2012) 747–763] 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) problem, showing that singularity 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. Depending on the degree of the singularity, our analysis also suggests a simple reparametrization that offers an alternative to the so-called *centred parametrization* proposed, in the particular case of skew-normal and skew-t families, by Azzalini [*Scand. J. Stat.* **12** (1985) 171–178], Arellano-Valle and Azzalini [*J. Multivariate Anal.* **113** (2013) 73–90], and DiCiccio and Monti [*Quaderni di Statistica* **13** (2011) 1–21], respectively.

Keywords: centred parametrization; consistency rates; skewing function; skew-normal distributions; skew-symmetric distributions; singular Fisher information

1. Introduction

The *skew-symmetric* families, originally proposed in Azzalini and Capitanio [8] and Wang, Boyer and Genton [29], are, in their univariate version, parametric families of probability density functions (p.d.f.s) of the form

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

(a) $\boldsymbol{\vartheta} = (\mu, \sigma, \delta)'$, with $\mu \in \mathbb{R}$ a *location parameter* and $\sigma \in \mathbb{R}_0^+$ 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 nonvanishing symmetric p.d.f. (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

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

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

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

The assumption of a nonvanishing kernel f is not essential, and has been added, as in Hallin and Ley [17], in order to ease calculations and avoid trivial complications. Condition (1.2) is classical; the less classical condition (1.3) is justified by 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 [17], satisfy (1.3). Further comments on skewing functions of the form $\Pi(z, \delta) = \Pi(\delta z)$ can be found in Section 6.2.

The *skew-normal* family of Azzalini [4], for which the symmetric kernel f is the standard Gaussian p.d.f. ϕ and the skewing function $\Pi(z, \delta) = \Phi(\delta z)$ for Φ the standard Gaussian cumulative distribution function (c.d.f.), is by far 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 [5], the *skew-Cauchy distributions* of Arnold and Beaver [3], the *skew-t densities* of Azzalini and Capitanio [8], or the *generalized skew-normal distributions* of Loperfido [22]; the latter result (along with arbitrary Π) from letting $f = \phi$ in (1.1), and play an important role in this paper. We refer to Genton [14], Azzalini [6] or Ley [20] for background reading, details and examples.

Since the pioneering paper by Azzalini [4], 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 may be difficult from an inferential point of view; in particular, traditional tests of the null hypothesis of symmetry, at first sight, seem problematic.

That singularity problem has been discussed at length in a number of papers, among which Azzalini and Capitanio [7], Pewsey [25], DiCiccio and Monti [12], Chiogna [10], Azzalini and Genton [9] or Ley and Paindaveine [21]; see Hallin and Ley [17] for a detailed account. While all authors were pointing at some special status for normal kernels, hence skew-normal distributions, Hallin and Ley [17] have shown that this information singularity 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).

Singularity of Fisher information 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 singularity of order one, by Rotnitzky *et al.* [27], who unify and reinforce earlier proposals by, for example, Cox and Hinkley ([11], pages 117 and 118) or Lee and Chesher [19].

The typical rate, corresponding to a "simple singularity", would be $n^{1/4}$. However, it is well known (see, e.g., Chiogna [10]) that, for skew-normal distributions, that $n^{1/4}$ 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 [17]. 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 excluded.

Our aim is to characterize, in the spirit of Hallin and Ley [17], among all families of univariate skew-symmetric distributions suffering from Fisher singularity, those exhibiting that simple/double/triple singularity phenomenon, and to show that there exist no higher-order ones. It turns out that only Gaussian kernels can exhibit double (a fortiori, triple) singularity. The skew-normal family is but one example; other cases are found in the aforementioned class of generalized skew-normal distributions (Loperfido [22]). We also provide (in the spirit of Rotnitzky *et al.* [27]) 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, for δ at $\delta = 0$.

With the same objective of overcoming the Fisher singularity problem, Azzalini [4], Arellano-Valle and Azzalini [1], Arellano-Valle and Azzalini [2] and DiCiccio and Monti [13], in the univariate skew-normal, multivariate skew-normal, and univariate skew-t cases, respectively, also propose a nonlinear reparametrization, called the *centred parametrization*, which we discuss in some detail in Section 6.1. Based on Gram–Schmidt orthogonalization (in the space of *scores*) ideas, our reparametrizations are analytically simple, and lead to closed-form expressions of the scores at $\delta = 0$. As far as the asymmetry parameter is concerned, and depending on the degree of Fisher singularity, they involve $\delta^{(1)} := \mathrm{sign}(\delta)\delta^2$, $\delta^{(2)} := \delta^3$, or $\delta^{(3)} := \mathrm{sign}(\delta)\delta^4$, thus preserving the interpretation of skewness as a tuning parameter of the skewing mechanism and the characterization of the null hypothesis of symmetry as $\delta = 0$. Finally, our reparametrizations apply to arbitrary skew-symmetric families with singular Fisher information at $\delta = 0$.

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 are provided in Section 5. We conclude the paper with a discussion of Azzalini's centred parametrization and its relation to ours (Section 6.1), and some warnings (Section 6.2) about the potential dangers of the standard skewing functions $\Pi(z, \delta) = \Pi(\delta z)$.

2. Simple singularity

In this section, we first briefly revisit the main result of Hallin and Ley [17] in order, mainly, to settle the notation. We then show how to resolve the singularity problem via an adequate reparametrization leading, in general, to $n^{1/4}$ consistency rates for δ in the vicinity of symmetry – equivalently, to a reparametrization of skewness, of the form $\delta^{(1)} := \operatorname{sign}(\delta)\delta^2$, recovering $n^{1/2}$ rates all over the parameter space.

2.1. Simple singularity: A mismatch between f and Π

Throughout, we consider the skew-symmetric distributions with p.d.f. (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 [17].

Assumption (A1). (i) *The symmetric kernel f is a* standardized *symmetric p.d.f.* (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*

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

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 $\int_{-\infty}^{\infty} \psi^2(z) f(z)\,\mathrm{d}z$ 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 [17] for a discussion. 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

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

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

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

with

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

and

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

The zeroes in $\boldsymbol{\Gamma}_{f;\boldsymbol{\vartheta}_0}$ are easily obtained by noting that $\ell^1_{f;\boldsymbol{\vartheta}_0}$ and $\ell^3_{f;\boldsymbol{\vartheta}_0}$ are 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 [17] 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*

$$\mathcal{E}_\Psi := \left\{ g_a := \exp(-a\Psi) \Big/ \int_{-\infty}^{\infty} \exp\bigl(-a\Psi(z)\bigr)\,\mathrm{d}z \,\Big|\, a \in \mathcal{A} \right\} \tag{2.1}$$

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

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

yielding

$$\gamma^{11}_{f;\boldsymbol{\vartheta}_0} = \sigma^{-2}a^2\int_{-\infty}^{\infty} \psi^2(z) f(z)\,\mathrm{d}z \quad \text{and} \quad \gamma^{13}_{f;\boldsymbol{\vartheta}_0} = 2\sigma^{-1}a\int_{-\infty}^{\infty} \psi^2(z) f(z)\,\mathrm{d}z. \tag{2.2}$$

We refer the reader to the end of Section 2.1 in Hallin and Ley [17] 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.* [27]. Assume that f and Π are such that $f \in \mathcal{E}_\Psi$. The collinearity at $\boldsymbol{\vartheta}_0$ 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

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

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 (2.2),

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

This (orthogonal at $\boldsymbol{\vartheta}_0$) system of scores is associated with the reparametrization $(\mu^{(1)}, \sigma^{(1)}, \delta)'$, where $\mu^{(1)} = \mu + 2\delta\sigma/a$ and $\sigma^{(1)} = \sigma$, hence

$$\begin{aligned} & f^{\Pi}_{\mu^{(1)},\sigma^{(1)},\delta}(x) \\ & \quad := 2\big(\sigma^{(1)}\big)^{-1} f\big(\big(x-\mu^{(1)}+2\delta\sigma^{(1)}/a\big)/\sigma^{(1)}\big)\Pi\big(\big(x-\mu^{(1)}+2\delta\sigma^{(1)}/a\big)/\sigma^{(1)},\delta\big). \end{aligned} \tag{2.3}$$

Note that this reparametrization, which only affects the location parameter, is adopted for the family as a whole, although the orthogonalization argument it is based on only holds at $(\mu^{(1)}, \sigma^{(1)}, 0)' = (\mu, \sigma, 0)' = \boldsymbol{\vartheta}_0$.

Since this reparametrization cancels, at $\delta = 0$, the score for skewness, hence the linear term in the Taylor expansion of the log-likelihood, second derivatives with respect to δ naturally come into the picture. To be precise, since the linear term $\tau_3\partial_\delta \log f^{\Pi}_{\mu^{(1)},\sigma^{(1)},\delta}(x)|_{\boldsymbol{\vartheta}_0}$ in the Taylor expansion of $\log f^{\Pi}_{\boldsymbol{\vartheta}_0+(0,0,\tau_3)'}(x)$ about $\log f^{\Pi}_{\boldsymbol{\vartheta}_0}(x)$ happens to be zero, the first local approximation is provided by the quadratic term $(\tau_3^2/2)\partial^2_\delta \log f^{\Pi}_{\mu^{(1)},\sigma^{(1)},\delta}(x)|_{\boldsymbol{\vartheta}_0}$ – provided that second derivatives exist. 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.* [27]). This is the structural reason for slower-than-root-n consistency rates (at $\boldsymbol{\vartheta}_0$) for the skewness parameter δ in the singular case.

The existence of second-order derivatives requires reinforcing regularity assumptions; at the same time it suggests reparametrizing skewness in terms of $\delta^{(1)} = \mathrm{sign}(\delta)\delta^2$ instead of δ.

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

Letting $\delta^{(1)} = \mathrm{sign}(\delta)\delta^2$, consider the reparametrization $\boldsymbol{\vartheta}^{(1)} := (\mu^{(1)}, \sigma^{(1)}, \delta^{(1)})'$. The reinforced regularity assumptions we need (at $\boldsymbol{\vartheta}^{(1)}_0 = (\mu^{(1)}, \sigma^{(1)}, 0)' = \boldsymbol{\vartheta}_0$) are as follows – recall that here we only address the singular case under which f and Π are such that $f = g_a \in \mathcal{E}_\Psi$ for some $a \in \mathcal{A}$ (see (2.1)), so that f is entirely determined by Π and the constant a, and we only need strengthening Assumption (A2).

Assumption (A2$^+$). *Same as Assumption (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*

$$\int_{-\infty}^{\infty}\psi^2(z)z^2 f(z)\,\mathrm{d}z \quad \textit{and} \quad \int_{-\infty}^{\infty}\big(2a^{-1}\dot{\psi}(z) - 2\psi^2(z)\big)^2 f(z)\,\mathrm{d}z$$

are finite.

Assumption (A2$^+$)(i) ensures the existence of the second derivative $\partial^2_\delta f^\Pi_{\mu^{(1)},\sigma^{(1)},\delta}(x)|_{\boldsymbol{\vartheta}_0}$ (hence, via l'Hospital's rule, that of the simple derivative $\partial_{\delta^{(1)}} f^\Pi_{\boldsymbol{\vartheta}^{(1)}}(x)|_{\boldsymbol{\vartheta}_0}$), while Assumption (A2$^+$)(ii) guarantees the 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 Assumption (A2$^+$)(ii)). As already pointed out, Assumption (A2$^+$) not only reinforces Assumption (A2) but also, via the requirement that $f = g_a \in \mathcal{E}_\Psi$ for some $a\in\mathcal{A}$, entails Assumption (A1), which is no longer needed.

Under Assumption (A2$^+$), differentiating $\log f^\Pi_{\mu^{(1)},\sigma^{(1)},\delta^{(1)}}$ with respect to $\delta^{(1)}$ and, at $\delta^{(1)} = \delta = 0$, applying l'Hospital's rule once leads to (with $f^\Pi_{\mu^{(1)},\sigma^{(1)},\delta}(x)$ as in (2.3))

$$\partial_{\delta^{(1)}}\log f^\Pi_{\boldsymbol{\vartheta}^{(1)}}(x) = \begin{cases} \dfrac{1}{2\sqrt{|\delta^{(1)}|}}\partial_\delta \log f^\Pi_{\mu^{(1)},\sigma^{(1)},\delta}(x)\Big|_{\delta=\mathrm{sign}(\delta^{(1)})(\delta^{(1)})^{1/2}}, & \text{if } \delta^{(1)}\neq 0,\\ \pm\dfrac{1}{2}\partial^2_\delta\log f^\Pi_{\mu^{(1)},\sigma^{(1)},\delta}(x)|_{\delta=0}, & \text{if } \delta^{(1)} = 0,\end{cases} \tag{2.4}$$

where the undetermined sign at $\delta = 0$ is due to the fact that the left derivative (minus sign) and the right derivative (plus sign) do not coincide. It follows that, with the same sign indeterminacy,

$$\partial_{\delta^{(1)}}\log f^\Pi_{\boldsymbol{\vartheta}^{(1)}}(x)|_{\boldsymbol{\vartheta}_0^{(1)}} = \pm 2\big[a^{-1}\dot\psi\big(\sigma^{-1}(x-\mu)\big) - \psi^2\big(\sigma^{-1}(x-\mu)\big)\big], \tag{2.5}$$

hence, in line with Section 2.1,

$$\begin{aligned}\boldsymbol{\ell}_{f;\boldsymbol{\vartheta}_0^{(1)}}(x) &:= \big(\ell^1_{f;\boldsymbol{\vartheta}_0^{(1)}}(x), \ell^2_{f;\boldsymbol{\vartheta}_0^{(1)}}(x), \ell^3_{f;\boldsymbol{\vartheta}_0^{(1)}}(x)\big)'\\ &:= \begin{pmatrix}\partial_{\mu^{(1)}}\log f^\Pi_{\boldsymbol{\vartheta}^{(1)}}(x)|_{\boldsymbol{\vartheta}_0}\\ \partial_{\sigma^{(1)}}\log f^\Pi_{\boldsymbol{\vartheta}^{(1)}}(x)|_{\boldsymbol{\vartheta}_0}\\ \partial_{\delta^{(1)}}\log f^\Pi_{\boldsymbol{\vartheta}^{(1)}}(x)|_{\boldsymbol{\vartheta}_0}\end{pmatrix}\\ &= \begin{pmatrix}\sigma^{-1}a\psi\big(\sigma^{-1}(x-\mu)\big)\\ \sigma^{-1}\big(\sigma^{-1}(x-\mu)a\psi\big(\sigma^{-1}(x-\mu)\big)-1\big)\\ \pm2\big[a^{-1}\dot\psi\big(\sigma^{-1}(x-\mu)\big)-\psi^2\big(\sigma^{-1}(x-\mu)\big)\big]\end{pmatrix}\end{aligned}$$

with covariance

$$\begin{aligned}\boldsymbol{\Gamma}_{f;\boldsymbol{\vartheta}_0^{(1)}} &:= \sigma^{-1}\int_{-\infty}^{\infty}\boldsymbol{\ell}_{f;\boldsymbol{\vartheta}_0^{(1)}}(x)\boldsymbol{\ell}'_{f;\boldsymbol{\vartheta}_0^{(1)}}(x) f\big(\sigma^{-1}(x-\mu)\big)\,\mathrm{d}x\\ &=: \begin{pmatrix}\gamma^{11}_{f;\boldsymbol{\vartheta}_0^{(1)}} & 0 & 0\\ 0 & \gamma^{22}_{f;\boldsymbol{\vartheta}_0^{(1)}} & \pm\gamma^{23}_{f;\boldsymbol{\vartheta}_0^{(1)}}\\ 0 & \pm\gamma^{23}_{f;\boldsymbol{\vartheta}_0^{(1)}} & \gamma^{33}_{f;\boldsymbol{\vartheta}_0^{(1)}}\end{pmatrix},\end{aligned} \tag{2.6}$$

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

$$\gamma^{11}_{f;\boldsymbol{\vartheta}_0^{(1)}} = a^2\sigma^{-2}\int_{-\infty}^{\infty}\psi^2(z)f(z)\,\mathrm{d}z,$$
$$\gamma^{22}_{f;\boldsymbol{\vartheta}_0^{(1)}} = \sigma^{-2}\int_{-\infty}^{\infty}\bigl(a\psi(z)z-1\bigr)^2 f(z)\,\mathrm{d}z,$$
$$\gamma^{33}_{f;\boldsymbol{\vartheta}_0^{(1)}} = 4\int_{-\infty}^{\infty}\bigl(a^{-1}\dot{\psi}(z)-\psi^2(z)\bigr)^2 f(z)\,\mathrm{d}z$$

and

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

The existence of a left and a right score for $\delta^{(1)}$ at $\delta^{(1)}=0$, a fact which does not occur with $\frac{1}{2}\partial_\delta^2\log f^\Pi_{\mu^{(1)},\sigma^{(1)},\delta}(x)|_{\boldsymbol{\vartheta}_0}$, is not a problem, as the linear term in the Taylor expansion of $\log f^\Pi_{\boldsymbol{\vartheta}_0^{(1)}+(0,0,\tau_3)'}(x)$ about $\log f^\Pi_{\boldsymbol{\vartheta}_0^{(1)}}(x)$ now is of the form $\frac{|\tau_3|}{2}\partial_\delta^2\log f^\Pi_{\mu^{(1)},\sigma^{(1)},\delta}(x)|_{\boldsymbol{\vartheta}_0}$, so that only the sign of τ_3 gets lost, as already mentioned.

Now, let us first assume that $\boldsymbol{\Gamma}_{f;\boldsymbol{\vartheta}_0^{(1)}}$ has full rank. Under Assumption (A2$^+$) and the new $\boldsymbol{\vartheta}^{(1)}$-parametrization, the model enjoys all the properties required for traditional root-n maximum likelihood and Lagrange Multiplier or Rao score tests. For instance, denoting by $X_1,\dots,X_n$ an i.i.d. sample of size n from $f^\Pi_{\boldsymbol{\vartheta}_0^{(1)}}$, the Lagrange Multiplier test rejects the null hypothesis of symmetry (in favor of an asymmetry of unspecified sign) whenever the quadratic statistic

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

($\hat{\boldsymbol{\vartheta}}_0^{(1)}=(\hat{\mu},\hat{\sigma},0)$ stands for a root-n consistent, under $\delta=0$, estimator of $\boldsymbol{\vartheta}_0^{(1)}=\boldsymbol{\vartheta}_0$) exceeds the chi-square quantile (one degree of freedom) of order $(1-\alpha)$.

Summing up, provided that $\boldsymbol{\Gamma}_{f;\boldsymbol{\vartheta}_0^{(1)}}$ has full rank, root-n consistency/contiguity rates are achieved for $\delta^{(1)}=\mathrm{sign}(\delta)\delta^2$. This implies the same root-n rates at any $\delta\neq 0$; at $\delta=0$, however, an $n^{1/2}$ rate for $\delta^{(1)}$ means an $n^{1/4}$ rate for $\delta=\mathrm{sign}(\delta^{(1)})\sqrt{|\delta^{(1)}|}$. Note, however, that, despite the fact that a Gram–Schmidt argument was used in the construction, $\boldsymbol{\Gamma}_{f;\boldsymbol{\vartheta}_0^{(1)}}$ in general is not diagonal: there is no reason, indeed, for the new score (2.5) to be orthogonal to those for $\mu^{(1)}$ and $\sigma^{(1)}$.

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 nonsingular; our objective of a singularity-free parametrization thus is achieved, with consistency rate, in the vicinity of symmetry, of $n^{1/4}$ for δ. But this is not a general rule: in the case of the skew-normal family, for instance, Chiogna [10] showed that the correct rate for δ is only $n^{1/6}$. The explanation, as we shall see in the next section, 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 $\ell^1_{f;\boldsymbol{\vartheta}_0^{(1)}}$, at $\boldsymbol{\vartheta}_0^{(1)}$, is orthogonal to the other two scores).

3. Double singularity

3.1. Double singularity: A special role for Gaussian kernels

The double singularity phenomenon ($\ell^2_{f;\boldsymbol{\vartheta}_0^{(1)}}$ and $\ell^3_{f;\boldsymbol{\vartheta}_0^{(1)}}$ collinear) takes place if and only if

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

(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 (2.1)). Rewriting this equation as

$$\dot{\psi}(z) = -\frac{ab}{2\sigma} + \frac{a^2 b}{2\sigma} z\psi(z) + a\psi^2(z) \qquad \text{a.e.} \tag{3.1}$$

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

$$\psi(z) = \frac{-ab}{2\sigma} z \tag{3.2}$$

or

$$\psi(z) = \frac{-ab}{2\sigma} z + \exp\left(-\frac{a^2 b z^2}{4\sigma}\right) \Big/ \left(c - a\int_0^z \exp\left(-\frac{a^2 b y^2}{4\sigma}\right) \mathrm{d}y\right), \qquad b, c \in \mathbb{R}. \tag{3.3}$$

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 $z \mapsto a\int_0^z \exp(-a^2 b y^2/4\sigma)\,\mathrm{d}y$ and ψ are odd, the constant c in (3.3) has to be zero. By (2.1), the natural parameter space $\mathcal{A}$ for the exponential family $\mathcal{E}_\Psi$ associated with the mapping ψ of (3.3) then consists of the set of values of a for which the integral

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

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

$$\int_{-\infty}^{\infty} \exp\bigl(-z^2\bigr)\left|\int_0^z \exp\bigl(y^2\bigr)\,\mathrm{d}y\right|\mathrm{d}z < \infty. \tag{3.4}$$

However, one easily can check that $\lim_{z\to\infty} z\exp(-z^2)|\int_0^z \exp(y^2)\,\mathrm{d}y| = 1/2$, which means that $\exp(-z^2)|\int_0^z \exp(y^2)\,\mathrm{d}y|$ behaves as $1/z$ for large values of z. It follows that (3.4) is impossible. Hence, the natural parameter space $\mathcal{A}$ is empty, meaning that no symmetric kernel f associated to the mapping ψ of (3.3) can yield singular Fisher information. Therefore, the only admissible solution to (3.1) is (3.2).

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 thus is specific to the Gaussian kernel, hence to a well-determined subclass of *generalized skew-normal distributions* (in the sense of Loperfido [22], see the Introduction). This also implies that, under the assumptions made, $n^{1/4}$ consistency rates for δ 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* (2.1);

(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* [22]).

This theorem completely characterizes the double singularity problem, hence complements the simple singularity characterization of Hallin and Ley [17].

3.2. A singularity-free reparametrization

Still inspired by Rotnitzky *et al.* [27], 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 limit ourselves to $f = \phi$. Moreover, we know from the previous section that ψ has to be of the form $\psi(z) = c_2 z$; 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 idea as in Section 2.2, but with the score for scale $\ell^2_{\phi;\boldsymbol{\vartheta}^{(1)}}$ substituted for the score for location, we project $\ell^3_{\phi;\boldsymbol{\vartheta}^{(1)}}$ onto the subspace orthogonal to $\ell^1_{\phi;\boldsymbol{\vartheta}^{(1)}}$ and $\ell^2_{\phi;\boldsymbol{\vartheta}^{(1)}}$ in the L_2 geometry of the information matrix (2.6). The resulting residual score for skewness then, as expected, is zero at $\boldsymbol{\vartheta}_0^{(1)}$; indeed

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

Translating, as in Section 2.2, this projection in terms of parameters leads to a reparametrization of the form $(\mu^{(2)}, \sigma^{(2)}, \delta)'$, with

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

and

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

Note that this reparametrization again is global, although the orthogonality argument above only holds at $\delta = 0$. Also note that $\sigma^{(2)}$ is not necessarily positive; although $\mu^{(2)}$ and $\sigma^{(2)}$ jointly characterize location and scale, they cannot be interpreted separately as a location and a scale parameter. In line with previous notation, we denote by $f^{\Pi}_{\mu^{(2)},\sigma^{(2)},\delta}$ the resulting skew-symmetric density, keeping in mind the fact that the symmetric kernel f is ϕ. The same remarks as for the first reparametrization are in order: keeping δ as the skewness parameter yields $n^{1/6}$ consistency/contiguity rates. Indeed, the first two derivatives with respect to δ now cancel at $\delta = 0$, so that derivatives of order three play the dominant role in local approximations of log-likelihoods.

In the particular case of the skew-normal family, similar ideas have been exploited by Chiogna [10], where the $n^{1/6}$ rates also are established. The reparametrization developed there, however, does not coincide with ours, as the orthogonalization it is based on holds at one specific value $\boldsymbol{\vartheta}_0^* = (\mu^*, \sigma^*, 0)$ of $\boldsymbol{\vartheta}_0$ only ((μ^*, σ^*) arbitrary but fixed). The resulting score for skewness accordingly vanishes at $\boldsymbol{\vartheta}_0^*$, while ours vanishes at all $\boldsymbol{\vartheta}_0$.

Appearance of third derivatives, in turn, suggests using $\delta^{(2)} = \delta^3$ as a new parameter of skewness, yielding the reparametrization $\boldsymbol{\vartheta}^{(2)} := (\mu^{(2)}, \sigma^{(2)}, \delta^{(2)})'$, with $\boldsymbol{\vartheta}_0^{(2)} := (\mu, \sigma, 0)' = \boldsymbol{\vartheta}_0$. The new score for skewness then will be calculated according to

$$\partial_{\delta^{(2)}} \log f^{\Pi}_{\boldsymbol{\vartheta}^{(2)}}(x) = \begin{cases} \dfrac{1}{3(\delta^{(2)})^{2/3}}\partial_\delta \log f^{\Pi}_{\mu^{(2)},\sigma^{(2)},\delta}(x)\Big|_{\delta=(\delta^{(2)})^{1/3}}, & \text{if } \delta^{(2)} \neq 0,\\ \dfrac{1}{6}\partial_\delta^3 \log f^{\Pi}_{\mu^{(2)},\sigma^{(2)},\delta}(x)|_{\delta=0}, & \text{if } \delta^{(2)} = 0, \end{cases} \tag{3.5}$$

and follows on applying l'Hospital's rule twice. This, however, requires the following reinforcement of Assumption (A2$^+$).

Assumption (A2^{++}). *Same as Assumption* (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* $\Upsilon(z):=\partial_\delta^3\Pi(z,\delta)|_{\delta=0}$, *the integral* $\int_{-\infty}^{\infty}(\frac{8}{3a^3}z^3-\frac{8}{a^3}z+\frac{1}{3}\Upsilon(z))^2\phi(z)\,\mathrm{d}z$ *is finite.*

Assumption (A2^{++})(i) ensures the existence of the third-order derivative $\partial_\delta^3 f^{\Pi}_{\mu^{(2)},\sigma^{(2)},\delta}$ at $\boldsymbol{\vartheta}_0^{(2)}=\boldsymbol{\vartheta}_0$, while Assumption (A2^{++})(ii) guarantees finiteness of the corresponding covariance matrix. Also note that the mixed derivative $\partial_z\partial_\delta^2\Pi(z,\delta)|_{\delta=0}$ vanishes by the definition of skewing functions; so does $\partial_z^2\partial_\delta\Pi(z,\delta)|_{\delta=0}=\partial_z^2\psi(z)$ for all z, since we are dealing (Theorem 3.1(ii)) with skewing functions such that $\psi(z)=z/a$ is linear. Finally note that $\Upsilon(z)$, by (1.2), is an odd function. These facts greatly simplify calculations.

Assumption (A2^{++}) implies, for this second reparametrization, the existence, at $\boldsymbol{\vartheta}_0$, 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 scores described in Section 2.3. Elementary algebra yields

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

and

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

with

$$
\begin{aligned}
&\gamma^{11}_{\phi;\boldsymbol{\vartheta}_0^{(2)}}=\sigma^{-2}\int_{-\infty}^{\infty}z^2\phi(z)\,\mathrm{d}z=\sigma^{-2},\qquad \gamma^{22}_{\phi;\boldsymbol{\vartheta}_0^{(2)}}=\sigma^{-2}\int_{-\infty}^{\infty}\big(z^2-1\big)^2\phi(z)\,\mathrm{d}z=2\sigma^{-2},\\
&\gamma^{33}_{\phi;\boldsymbol{\vartheta}_0^{(2)}}=\int_{-\infty}^{\infty}\bigg(\frac{8}{3a^3}z^3-\frac{8}{a^3}z+\frac{1}{3}\Upsilon(z)\bigg)^2\phi(z)\,\mathrm{d}z
\end{aligned}
$$

and

$$
\gamma^{13}_{\phi;\boldsymbol{\vartheta}_0^{(2)}}=\frac{1}{3}\sigma^{-1}\int_{-\infty}^{\infty}z\Upsilon(z)\phi(z)\,\mathrm{d}z.
$$

Here again, $\gamma^{13}_{\phi;\boldsymbol{\vartheta}_0^{(2)}}$, in general, is not zero, and $\boldsymbol{\Gamma}_{\phi;\boldsymbol{\vartheta}_0^{(2)}}$ is not diagonal.

If we assume, as in 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

$$\frac{1}{n}\sum_{i=1}^{n}\bigl(\ell^3_{\phi;\hat{\boldsymbol{\vartheta}}_0^{(2)}}(X_i) - \sigma^2\gamma^{13}_{\phi;\hat{\boldsymbol{\vartheta}}_0^{(2)}}\ell^1_{\phi;\hat{\boldsymbol{\vartheta}}_0^{(2)}}(X_i)\bigr)^2 / \bigl(\gamma^{33}_{\phi;\hat{\boldsymbol{\vartheta}}_0^{(2)}} - \sigma^2\bigl(\gamma^{13}_{\phi;\hat{\boldsymbol{\vartheta}}_0^{(2)}}\bigr)^2\bigr),$$

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 $n^{1/6}$ while that for $\delta^{(2)}$ uniformly remains $n^{1/2}$, and the same comments can be made as in Section 2.3. The particular case of the skew-normal family, from the point of view of Le Cam's asymptotic theory of statistical experiments, is studied in full detail by Hallin, Ley and Monti [18], thereby generalizing and extending previous work by Salvan [28] where, despite the singularities, a locally optimal test for normality against skew-normal alternatives is derived.

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 $\boldsymbol{\vartheta}_0^{(2)}$) collinear to the score for location $\ell^1_{\phi;\boldsymbol{\vartheta}_0^{(2)}}$ (note that, very clearly, it is orthogonal to the score for 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 for δ (but keeping $n^{1/2}$ rates for a reparametrization $\delta^{(3)} = \operatorname{sign}(\delta)\delta^4$ of skewness). That $n^{1/8}$ rate, however, as we shall see, is the worst possible one.

We are skipping details, as they are very much the same as in Section 3. 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 linearly dependent of $\ell^1_{\phi;\boldsymbol{\vartheta}_0^{(2)}} = z/\sigma$, $\Upsilon(z)$ necessarily has to be of the form $\alpha_1 z + \alpha_2 z^3$, with $\alpha_1 \in \mathbb{R}$ and $\alpha_2 = -8/a^3$ in order to cancel the term in z^3. This condition on the third derivative with respect to δ thus characterizes what we would call a *triple singularity* case (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 weird peculiarity; see Section 5.4.

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

$$\mu^{(3)} = \mu^{(2)} + \left(-\frac{8}{a^3} + \frac{\alpha_1}{3}\right)\sigma\delta^3 = \mu + \frac{2}{a}\sigma\delta + \left(-\frac{8}{a^3} + \frac{\alpha_1}{3}\right)\sigma\delta^3,$$

$$\sigma^{(3)} = \sigma^{(2)} = \sigma\bigl(1 - 2\delta^2/a^2\bigr)$$

and

$$\delta^{(3)} = \operatorname{sign}(\delta)\delta^4.$$

This reparametrization of skewness entails, with the same left and right derivative interpretation for $\pm$ as in (2.4),

$$\partial_{\delta^{(3)}} \log f^{\Pi}_{\boldsymbol{\vartheta}^{(3)}}(x) = \begin{cases} \dfrac{1}{4|\delta^{(3)}|^{3/4}} \partial_\delta \log f^{\Pi}_{\mu^{(3)},\sigma^{(3)},\delta}(x)\Big|_{\delta=\operatorname{sign}(\delta^{(3)})(\delta^{(3)})^{1/4}}, & \text{if } \delta^{(3)} \neq 0, \\ \pm\dfrac{1}{24}\partial^4_\delta \log f^{\Pi}_{\mu^{(3)},\sigma^{(3)},\delta}(x)\Big|_{\delta=0}, & \text{if } \delta^{(3)} = 0, \end{cases}$$

by means of a triple use of l'Hospital's rule. This, however, requires fourth-order derivatives, hence further strengthening of Assumption (A2^{++}).

Assumption (A2^{+++}). *Same as Assumption* (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 we do not need to assume finiteness of Fisher information for skewness, as this, as we shall see, will always be the case after this third reparametrization. Clearly, as in all previous cases, both the score for location $\ell^1_{\phi;\boldsymbol{\vartheta}^{(3)}_0}$ and the score for scale $\ell^2_{\phi;\boldsymbol{\vartheta}^{(3)}_0}$ remain the same as in the original parametrization, and the new fourth-order score for skewness, for skewing functions satisfying $\partial^3_\delta \Pi(z,\delta)|_{\delta=0} = \alpha_1 z - \frac{8}{a^3}z^3$, becomes (after lengthy but quite elementary calculation)

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

One again easily can check that this quantity is centred under $\boldsymbol{\vartheta}^{(3)}_0$. The interesting feature, however, is that the term $\frac{4}{3a^4}(\frac{x-\mu}{\sigma})^4$ by no means can cancel out, and hence prevents $\ell^3_{\phi;\boldsymbol{\vartheta}^{(3)}_0}$ from being any linear combination of the location and scale scores. Thus, the resulting (at $\boldsymbol{\vartheta}^{(3)}_0$) Fisher information matrix

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

(the finiteness of which is obvious) *cannot* be singular, which in turn implies that $n^{1/8}$ rates of convergence for δ are the worst possible! The structural reason behind this result lies in the fact that, by the definition of skewing functions, $\partial_\delta^4 \Pi(z,\delta)|_{\delta=0}$ equals zero, hence cannot interfere in the fourth derivative, contrary to $\partial_\delta^3 \Pi(z,\delta)|_{\delta=0}$ which plays the crucial role in canceling 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 triple singularity phenomenon if and only if the symmetric kernel* f *is the normal kernel* ϕ *and the skewing function* Π *moreover is such that* $\psi(z) := \partial_\delta \Pi(z,\delta)|_{\delta=0} = z/a$ *for some nonzero real constant* a, *and has third-order derivative* $\Upsilon(z) := \partial_\delta^3 \Pi(z,\delta)|_{\delta=0} = \alpha_1 z - \frac{8}{a^3} z^3$ *for some real constant* α_1 *possibly zero.*

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

5. Examples

In this section, we illustrate our findings on the 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 [5] with p.d.f. $2c^{-1}\exp(-|z|^\alpha/\alpha)\Phi(\delta\,\mathrm{sign}(z)|z|^{\alpha/2}(2/\alpha)^{1/2})$ for $\alpha > 1$ and $c = 2\alpha^{1/\alpha-1}\Gamma(1/\alpha)$, and the skew-$t$ distributions of Azzalini and Capitanio [8] with p.d.f. $2t_\nu(z)T_{\nu+1}(\delta z(\nu+1)^{1/2}(z^2+\nu)^{-1/2})$ where t_η and T_η, respectively, stand for the p.d.f. and c.d.f. of a standard Student distribution with η degrees of freedom. These examples are discussed at length in Hallin and Ley [17], 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 $\dot{\Pi}(0) = \mathrm{d}\Pi(y)/\mathrm{d}y|_{y=0}$ exists and differs from zero.

5.2. Simple singularities

As shown in Hallin and Ley [17], 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 $\ddot{\Pi}(0) := d^2\Pi(y)/(\mathrm{d}y)^2|_{y=0}$ exists and, by construction, equals zero. Theorem 3.1 then tells us that among the p.d.f.s $2f(z)\Pi(\delta\varphi_f(z))$ 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 double singularity, a prominent example is of course Azzalini's skew-normal family, with p.d.f. $2\phi(z)\Phi(\delta z)$. Let us briefly show that higher-order singularities are excluded in that family. Straightforward calculation yields $a = \sqrt{2\pi}$ and $\Upsilon(z) = -(2\pi)^{-1/2}z^3$, which is different from $-\frac{8}{a^3}z^3 = -(2/\pi)^{3/2}z^3$. Hence, Theorem 4.1 readily yields the well-known result of $n^{1/6}$ consistency rates for δ in 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 [24] propose another family of skew densities generated by the normal kernel, with p.d.f.s of the form $2\phi(z)G(\delta z)$ where G is some univariate symmetric c.d.f. 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 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 6.2 for further discussion of such skewing functions.

5.4. Higher-order singularities

Let us further analyze the families of Nadarajah and Kotz [24]. Assume that G is three times continuously differentiable. Elementary calculations show that $a = 1/g(0)$, where $g(z) := \mathrm{d}G(z)/\mathrm{d}z$, 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 [24], 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 yield $n^{1/6}$ consistency rates for δ, the skew normal-logistic one requires the worst possible rates, namely $n^{1/8}$ rates.

Finally, consider the "lifted" skew-normal distribution, with p.d.f.

$$2\phi(z)\Phi\big(\delta z - (4-\pi)(6\pi)^{-1}\delta^3 z^3\big). \tag{5.1}$$

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 $n^{1/8}$ consistency rates for δ. Note that this distribution is part of the so-called class of *flexible generalized skew-normal distributions* defined in Ma and Genton [23]. 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 power 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 any odd polynomial H_ℓ with the terms in δz and $(\delta z)^3$ as in (5.1), for instance,

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

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

6. Some concluding remarks

We conclude this paper by a short discussion of two structural issues: the centred parametrization (Section 6.1) and the type of skewing function that causes most of the trouble when using a Gaussian kernel (Section 6.2).

6.1. The centred parametrization

In order to remedy Fisher singularity problems, Azzalini [4], in the very same paper where he first introduces the skew-normal densities, and for the specific case of the skew-normal family, suggested an alternative parametrization, the so-called *centred parametrization*. Denoting by Z a random variable with skew-normal density $2\phi(z)\Phi(\delta z)$, let $Y:=\mu+\sigma Z$ a.s. ($\sigma>0$): Y then has skew-normal density $2\sigma^{-1}\phi(\sigma^{-1}(z-\mu))\Phi(\delta\sigma^{-1}(z-\mu))$. That density has finite third-order moments: letting $\theta_1:=\mathrm{E}[Y]$ and $\theta_2:=\mathrm{Var}^{1/2}[Y]$, define $\gamma_1:=\mathrm{E}[(Y-\theta_1)^3]/\theta_2^3$ as Y's (hence also Z's) third standardized cumulant. The triple $\boldsymbol{\theta}:=(\theta_1,\theta_2,\gamma_1)'$ provides a parametrization of the skew-normal family, the *centred parametrization* (hereafter CP). The terminology "centred" refers to the fact that the new location and scale parameters θ_1 and θ_2 are such that $(Y-\theta_1)/\theta_2=(Z-\mathrm{E}[Z])/\mathrm{Var}^{1/2}[Z]$ has mean zero and variance one, whereas the original location and scale μ and σ values lead to $(Y-\mu)/\sigma=Z$, which is not centred about its mean. Azzalini calls μ, σ and δ *direct parameters*, since they can be directly read from the density of Y.

Besides Azzalini [4], the centred parametrization has been discussed in Azzalini and Capitanio [7], Pewsey [25] and Chiogna [10], to cite but these, for the skew-normal family, extended to the multinormal setup in Arellano-Valle and Azzalini [1], and to the skew-t distributions in DiCiccio and Monti [13] and Arellano-Valle and Azzalini [2].

The CP does not suffer from the Fisher singularity problem, and provides parameters that can all be estimated at the usual $n^{1/2}$ rate. It is well-suited for inferential purposes (see, e.g., Pewsey [25]), and enjoys a simple traditional moment-based interpretation (which is the main motivation for Arellano-Valle and Azzalini [2] to extend it to the skew-t context although skew-t families do not exhibit any Fisher singularity).

The main drawback of the CP lies in its complicated analytical form. Expressing the centred parameters $\boldsymbol{\theta}$ in terms of the original ones $\boldsymbol{\vartheta}$ yields (for the skew-normal family)

$$\theta_1=\mu+\sigma\sqrt{2/\pi}\,\delta\bigl(1+\delta^2\bigr)^{-1/2},\qquad\theta_2=\sigma\bigl(1+\delta^2(1-2/\pi)\bigr)^{1/2}\bigl(1+\delta^2\bigr)^{-1/2}$$

and

$$\gamma_1=\frac{4-\pi}{2}\left(\frac{2}{\pi}\right)^{3/2}\delta^3\bigl(1+\delta^2(1-2/\pi)\bigr)^{-3/2}.$$

The success of skew-symmetric families is largely due to the analytical simplicity of the decomposition (1.1) of a skew density into the product of a symmetric kernel and a skewing function. That simplicity is closely related to the original $\boldsymbol{\vartheta}$ parametrization, and gets lost in the CP. So is the flexibility that allows for combining various symmetric kernels and skewing functions in (1.1) without altering its structure: contrary to the mapping $\boldsymbol{\vartheta} \mapsto f_{\boldsymbol{\vartheta}}^{\Pi}(x)$, which does not depend on (f, Π), the mapping $\boldsymbol{\theta} \mapsto f_{\boldsymbol{\theta}}^{\Pi}(x)$ very much does. Although simple from the point of view of interpretation, the CP thus does not avoid analytical complexity, which results into lengthy and very tedious calculations – see Chiogna [10].[1]

Both the centred parametrization and those we are proposing in this paper are motivated by Fisher singularity problems at $\delta = 0$, and both are losing some of the simplicity of the original parametrization. The CP, however, is guided by interpretability considerations (the centred parameters *always* are the mean, the standard error, and the third-order cumulant); although solving the Fisher singularity problem in the skew-normal family, there is no guarantee it does so in other skew-symmetric families. Our reparametrizations, on the contrary, are guided by Fisher information considerations, and are specifically designed to solve the Fisher singularity problem, irrespective of the skew-symmetric family under study – with distinctive forms for simple, double, and triple singularity. The way they deal with δ (exponentiating it into $\text{sign}(\delta)\delta^2$, δ^3, or $\text{sign}(\delta)\delta^4$) preserves its interpretation, regardless of the chosen symmetric kernel, as a tuning quantity in the skewing mechanism that generates the family. Finally, due to their Gram–Schmidt nature, our reparametrizations are tailored for the construction of optimal tests of symmetry of the Lagrange Multiplier or Rao score type.

For the sake of comparison, we provide, in the Appendix, the scores for skewness associated with each of the two reparametrizations (the CP and ours) in the skew-normal family.

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

We conclude this section 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 the required differentiability conditions). Such functions are the most natural examples of a skewing function such that $\psi(z)$ is linear, yielding a risky combination with the Gaussian kernel ϕ.

The original skew-normal family of Azzalini [4] is based on $\Pi(z, \delta) = \Phi(\delta z)$; the same type of skewing function has been used, inter alia, by:

- Azzalini and Capitanio [7] for skew-symmetric densities of the form $2f(z)G(\delta z)$, with G some univariate symmetric distribution function (in fact, Azzalini and Capitanio proposed multivariate skew-elliptical distributions, but elliptical symmetry here boils down to plain univariate symmetry);

[1] We take this opportunity to indicate two unfortunate typos in that paper, namely (i) an exponent $^{-1/2}$ is missing in the expression of δ (λ in the notation of the paper) in terms of γ_1 on page 338, and (ii) the score function $u^{\Theta}(\phi^*, \gamma_1^*)$ on page 339, instead of three linear terms, should involve a quadratic term for the θ_2-score and a term of the form $(x^3 - 3x)$ for the γ_1-score.

- Gupta, Chang and Huang [16] for their skew-uniform, skew-t, skew-Cauchy, skew-Laplace and skew-logistic distributions, which all are special cases of Azzalini and Capitanio's [7] construction;
- Nadarajah and Kotz [24] for their skew normal-G distributions, as described in the previous sections; and by
- Gómez, Venegas and Bolfarine [15] 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 exhibit at least double singularity). An early important warning on the combination of a Gaussian kernel with such skewing functions has been given by Pewsey [26], who has shown that all densities of the form $2\phi(z)G(\delta z)$ (G some symmetric univariate c.d.f.) suffer from the singularity problem. His results are thus in total agreement with our general findings. The peculiarities of the skew-normal distribution, which belongs to all of the above-cited classes of distributions, have been discussed at length in the literature. We hope that this paper sheds some more light on the structural reasons behind those peculiarities, and provides further warning about the dangers of Gaussian kernels in combination with skewing functions of the form $\Pi(\delta z)$.

Appendix: Expressions of the score functions for the skew-normal distribution

In this Appendix, we provide the explicit expressions of the score functions for skewness in the skew-normal case (at any value of the skewness parameter, not only in the vicinity of symmetry) for both the centred parametrization and ours (as described in Section 3.2).

In our reparametrization $\boldsymbol{\vartheta}^{(2)}$, the score for skewness $\partial_{\delta^{(2)}} \log f^{\Pi}_{\boldsymbol{\vartheta}^{(2)}}(x)$ takes on the guise of a ratio $h_1(\mu^{(2)}, \sigma^{(2)}, \delta^{(2)})/h_2(\mu^{(2)}, \sigma^{(2)}, \delta^{(2)})$, with

$$
\begin{aligned}
&h_1(\mu^{(2)}, \sigma^{(2)}, \delta^{(2)}) \\
&\quad := \exp\left(-\frac{(\delta^{(2)})^{2/3}(\sqrt{2\pi}(\delta^{(2)})^{1/3}\sigma^{(2)} - (\delta^{(2)})^{2/3}(x-\mu^{(2)}) + \pi(x-\mu^{(2)}))^2}{2\pi^2(\sigma^{(2)})^2}\right)(\sigma^{(2)})^{-1} \\
&\qquad \times \left(\pi - \left(\delta^{(2)}\right)^{2/3}\right)\left(4\pi\left(\delta^{(2)}\right)^{1/3}\sigma^{(2)} + \sqrt{2}\pi^{3/2}\left(x-\mu^{(2)}\right) - 3\sqrt{2\pi}\left(\delta^{(2)}\right)^{2/3}\left(x-\mu^{(2)}\right)\right) \\
&\qquad - 4\pi^2\left(\delta^{(2)}\right)^{1/3}\Phi\left(\frac{\sqrt{2\pi}(\delta^{(2)})^{2/3}\sigma^{(2)} - \delta^{(2)}(x-\mu^{(2)}) + \pi(\delta^{(2)})^{1/3}(x-\mu^{(2)})}{\pi\sigma^{(2)}}\right) \\
&\qquad + \frac{2}{(\sigma^{(2)})^2}\left(\pi - \left(\delta^{(2)}\right)^{2/3}\right)\Phi\left(\frac{\sqrt{2\pi}(\delta^{(2)})^{2/3}\sigma^{(2)} - \delta^{(2)}(x-\mu^{(2)}) + \pi(\delta^{(2)})^{1/3}(x-\mu^{(2)})}{\pi\sigma^{(2)}}\right) \\
&\qquad \times \left(-\sqrt{2}\pi^{3/2}\sigma^{(2)}\left(x-\mu^{(2)}\right) - 2\delta^{(2)}\left(x-\mu^{(2)}\right)^2 + 3\sqrt{2\pi}\left(\delta^{(2)}\right)^{2/3}\sigma^{(2)}\left(x-\mu^{(2)}\right)\right. \\
&\qquad \left. + 2\pi\left(\delta^{(2)}\right)^{1/3}\left(\left(x-\mu^{(2)}\right)^2 - \left(\sigma^{(2)}\right)^2\right)\right)
\end{aligned}
$$

and $h_2(\mu^{(2)}, \sigma^{(2)}, \delta^{(2)}) := 6\pi^2(\pi - (\delta^{(2)})^{2/3})(\delta^{(2)})^{2/3}$

$$\times \Phi\left(\frac{\sqrt{2\pi}(\delta^{(2)})^{2/3}\sigma^{(2)} - \delta^{(2)}(x-\mu^{(2)}) + \pi(\delta^{(2)})^{1/3}(x-\mu^{(2)})}{\pi\sigma^{(2)}}\right).$$

In the CP reparametrization $\boldsymbol{\theta} = (\theta_1, \theta_2, \gamma_1)'$, the score for skewness $\partial_{\gamma_1} \log f^{\Pi}_{(\theta_1,\theta_2,\gamma_1)}(x)$ takes on the guise $h_1^{\mathrm{CP}}(\theta_1, \theta_2, \gamma_1) + h_2^{\mathrm{CP}}(\theta_1, \theta_2, \gamma_1)/h_3^{\mathrm{CP}}(\theta_1, \theta_2, \gamma_1)$, with

$$\begin{aligned}
h_1^{\mathrm{CP}}(\theta_1, \theta_2, \gamma_1) = &-\frac{1}{3\gamma_1^{1/3}(\gamma_1^{2/3} + ((4-\pi)/2)^{2/3})} \\
&+ \frac{1}{3}\left(\frac{2}{4-\pi}\right)^{2/3} \frac{(x - \theta_1 + (2/(4-\pi))^{1/3}\gamma_1^{1/3}\theta_2)^2}{\gamma_1^{1/3}\theta_2^2(1+\gamma_1^{2/3}(2/(4-\pi))^{2/3})^2} \\
&- \frac{1}{3}\left(\frac{2}{4-\pi}\right)^{1/3} \frac{x - \theta_1 + (2/(4-\pi))^{1/3}\gamma_1^{1/3}\theta_2}{\gamma_1^{2/3}\theta_2(1+\gamma_1^{2/3}(2/(4-\pi))^{2/3})},
\end{aligned}$$

$$\begin{aligned}
&h_2^{\mathrm{CP}}(\theta_1, \theta_2, \gamma_1) \\
&\quad = 2^{1/6}\exp\left(\frac{\pi(4-\pi)^{-2/3}\gamma_1^{2/3}(x-\theta_1+(2/(4-\pi))^{1/3}\gamma_1^{1/3}\theta_2)^2}{2^{4/3}(1+(2/(4-\pi))^{2/3}\gamma_1^{2/3})(-1+2^{-1/3}(\pi-2)(4-\pi)^{-2/3}\gamma_1^{2/3})\theta_2^2}\right) \\
&\qquad \times \big((x-\theta_1)\big(2(4-\pi)^{2/3}(\pi-2)\gamma_1^{4/3} + 2^{2/3}(\pi-4)^2\big) \\
&\qquad\quad + \theta_2\big(2^{2/3}(4-\pi)^2\gamma_1 + 4(4-\pi)^{5/3}\gamma_1^{1/3}\big)\big)
\end{aligned}$$

and

$$\begin{aligned}
&h_3^{\mathrm{CP}}(\theta_1, \theta_2, \gamma_1) \\
&\quad = 3(4-\pi)^{7/3}\theta_2\gamma_1^{2/3}\left(1+\gamma_1^{2/3}\left(\frac{2}{4-\pi}\right)^{2/3}\right)^{3/2}\big(2-2^{2/3}(\pi-2)(4-\pi)^{-2/3}\gamma_1^{2/3}\big)^{3/2} \\
&\qquad \times \Phi\left(\frac{2^{1/3}\sqrt{\pi}(4-\pi)^{-1/3}\gamma_1^{1/3}(x-\theta_1+\gamma_1^{1/3}\theta_2(2/(4-\pi))^{1/3})}{\theta_2(1+2^{2/3}(4-\pi)^{-2/3}\gamma_1^{2/3})^{1/2}(2-2^{2/3}(\pi-2)(4-\pi)^{-2/3}\gamma_1^{2/3})^{1/2}}\right).
\end{aligned}$$

Both score functions look equally complex. Under symmetry (either $\delta^{(2)} = 0$ or $\gamma_1 = 0$), they both yield an indetermination of the form $0/0$, apparently calling for tedious applications of l'Hospital's rule. For the $\boldsymbol{\vartheta}^{(2)}$ reparametrization we are proposing, such algebra is not required, though, as explicit expressions involving higher order derivatives already have been derived analytically (and follow quite directly from the Gram–Schmidt orthogonalization structure): see (3.5). In the CP reparametrization $\boldsymbol{\theta}$, the required algebra is so tedious as to defeat our version of MATHEMATICA.

Acknowledgements

Marc Hallin is 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), the Belgian Science Policy Office (2012–2017) Interuniversity Attraction Poles, 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.

Both authors would like to thank two anonymous referees for their helpful comments that led to an improvement of the present paper, as well as Adelchi Azzalini and Reinaldo Arellano-Valle for interesting discussions on the centred parametrization.

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. Corrigendum: **100** (2009), 816. MR2424355

[2] Arellano-Valle, R.B. and Azzalini, A. (2013). The centred parameterization and related quantities of the skew-*t* distribution. *J. Multivariate Anal.* **113** 73–90. MR2984357

[3] Arnold, B.C. and Beaver, R.J. (2000). The skew-Cauchy distribution. *Statist. Probab. Lett.* **49** 285–290. MR1794746

[4] Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scand. J. Stat.* **12** 171–178. MR0808153

[5] Azzalini, A. (1986). Further results on a class of distributions which includes the normal ones. *Statistica (Bologna)* **46** 199–208. MR0877720

[6] Azzalini, A. (2005). The skew-normal distribution and related multivariate families (with discussion). *Scand. J. Stat.* **32** 159–188. MR2188669

[7] Azzalini, A. and Capitanio, A. (1999). Statistical applications of the multivariate skew normal distribution. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **61** 579–602. MR1707862

[8] Azzalini, A. and Capitanio, A. (2003). Distributions generated by perturbation of symmetry with emphasis on a multivariate skew *t*-distribution. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **65** 367–389. MR1983753

[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] Chiogna, M. (2005). A note on the asymptotic distribution of the maximum likelihood estimator for the scalar skew-normal distribution. *Stat. Methods Appl.* **14** 331–341. MR2211338

[11] Cox, D.R. and Hinkley, D.V. (1974). *Theoretical Statistics*. London: Chapman & Hall. MR0370837

[12] DiCiccio, T.J. and Monti, A.C. (2004). Inferential aspects of the skew exponential power distribution. *J. Amer. Statist. Assoc.* **99** 439–450. MR2062829

[13] DiCiccio, T.J. and Monti, A.C. (2011). Inferential aspects of the skew *t*-distribution. *Quaderni di Statistica* **13** 1–21.

[14] Genton, M.G., ed. (2004). *Skew-Elliptical Distributions and Their Applications: A Journey Beyond Normality*. Boca Raton, FL: Chapman & Hall/CRC. MR2156754

[15] Gómez, H.W., Venegas, O. and Bolfarine, H. (2007). Skew-symmetric distributions generated by the distribution function of the normal distribution. *Environmetrics* **18** 395–407. MR2370731

[16] Gupta, A.K., Chang, F.C. and Huang, W.J. (2002). Some skew-symmetric models. *Random Oper. Stoch. Equ.* **10** 133–140. MR1912936

[17] Hallin, M. and Ley, C. (2012). Skew-symmetric distributions and Fisher information—A tale of two densities. *Bernoulli* **18** 747–763. MR2948899

[18] Hallin, M., Ley, C. and Monti, A.C. (2013). Optimal tests for symmetry against skew-normal alternatives. Unpublished manuscript.

[19] Lee, L.F. and Chesher, A. (1986). Specification testing when score test statistics are identically zero. *J. Econometrics* **31** 121–149. MR0846287

[20] Ley, C. (2012). Skew distributions. In *Statistical Theory and Methods*, *Encyclopedia of Environmetrics*, 2*nd Edition* (A. El-Shaarawi and W. Piegorsch, eds.) 1944–1949. New York: Wiley.

[21] Ley, C. and Paindaveine, D. (2010). On the singularity of multivariate skew-symmetric models. *J. Multivariate Anal.* **101** 1434–1444. MR2609504

[22] Loperfido, N.M.R. (2004). Generalized skew-normal distributions. In *Skew-Elliptical Distributions and Their Applications* 65–80. Boca Raton, FL: Chapman & Hall/CRC. MR2155323

[23] Ma, Y. and Genton, M.G. (2004). Flexible class of skew-symmetric distributions. *Scand. J. Stat.* **31** 459–468. MR2087837

[24] Nadarajah, S. and Kotz, S. (2003). Skewed distributions generated by the normal kernel. *Statist. Probab. Lett.* **65** 269–277. MR2018040

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

[26] Pewsey, A. (2006). Some observations on a simple means of generating skew distributions. In *Advances in Distribution Theory*, *Order Statistics*, *and Inference* (N. Balakrishnan, E. Castillo and J. M. Sarabia, eds.). *Stat. Ind. Technol.* 75–84. Boston, MA: Birkhäuser. MR2226216

[27] Rotnitzky, A., Cox, D.R., Bottai, M. and Robins, J. (2000). Likelihood-based inference with singular information matrix. *Bernoulli* **6** 243–284. MR1748721

[28] Salvan, A. (1986). Locally most powerful invariant test of normality (in Italian). *Atti XXXIII Riunione Società Italiana di Statistica* **2** 173–179.

[29] Wang, J., Boyer, J. and Genton, M.G. (2004). A skew-symmetric representation of multivariate distributions. *Statist. Sinica* **14** 1259–1270. MR2126352

Received September 2012 and revised April 2013