Received: 19 May 2018 | Revised: 11 December 2018 | Accepted: 19 December 2018

DOI: 10.1111/sjos.12377

ORIGINAL ARTICLE

Scandinavian Journal of Statistics

Identifiability of two-component skew normal mixtures with one known component

Shantan Jain[1] | Michael Levine[2] | Predrag Radivojac[1] | Michael W. Trosset[3]

[1]Khoury College of Computer and Information Sciences, Northeastern University, Boston, Massachusetts

[2]Department of Statistics, Purdue University, West Lafayette, Indiana

[3]Department of Statistics, Indiana University Bloomington, Indiana

Correspondence
Michael Levine, Department of Statistics, Purdue University, 250 N. University St., West Lafayette, IN 47907.
Email: mlevins@purdue.edu

Abstract

We give sufficient identifiability conditions for estimating mixing proportions in two-component mixtures of skew normal distributions with one known component. We consider the univariate case and two multivariate extensions: a multivariate skew normal distribution (MSN) and the canonical fundamental skew normal distribution (CFUSN). The characteristic function of the CFUSN distribution is additionally derived.

KEYWORDS

characteristic function, identifiability, one known component, skew normal, two-component mixture

1 | INTRODUCTION

Many real-world studies consider a population that can be divided into two subpopulations, based on the presence and absence of a certain property or trait of interest. Such a population is usually modeled using mixtures with two-component distributions, one for each subpopulation. Among different variants of parameter estimation in this setting, we are motivated by the problem of inferring mixing proportions (proportions of the subpopulations in the subsuming population) given a sample from the mixture distribution (a sample from the population) and a sample from one of the components (a sample from the subpopulation that satisfies the property). This setting is common in domains where an absence of the property cannot be easily verified due to practical or systemic constraints that are typically seen in social networks, molecular biology, etc. In social networks, for example, users may only be allowed to click "like" for a particular product and, thus, the data can be collected only for one of the component samples (a sample from the users who clicked "like") and the mixture (a sample from all users). Accurate estimation of mixing

Abbreviations: CF, characteristic function; MGF, moment-generating function; pdf, probability density function; SN, univariate skew normal family; MSN, the first multivariate skew normal family; CFUSN, canonical fundamental skew normal family.

proportions in this setting is important, with fundamental implications for false discovery rate estimation (Storey, 2002, 2003; Storey & Tibshirani, 2003) and, in the context of classification, more precisely, positive-unlabeled learning (Denis, Gilleron, & Letouzey, 2005), for estimating posterior distributions (Jain, White, & Radivojac, 2016; Jain, White, Trosset, & Radivojac, 2016; Ward, Hastie, Barry, Elith, & Leathwick, 2009) and recovering true classifier performance (Jain, White, & Radivojac, 2017; Menon, van Rooyen, Ong, & Williamson, 2015).

In this work, we consider two-component skew normal mixture families in univariate and multivariate settings and establish sufficient conditions that ensure identifiability of mixing proportions when the parameters of one of the components are known. Identifiability and estimation of mixtures have been extensively studied and well understood (Allman, Matias, & Rhodes, 2009; Dempster, Laird, & Rubin, 1977; McLachlan & Peel, 2000; Tallis & Chesson, 1982; Yakowitz & Spragins, 1968). More recently, however, the case with one known component has been considered in the nonparametric setting because of its connections with novelty detection and classification (Blanchard, Lee, & Scott, 2010; Bordes, Delmas, & Vandekerkhove, 2006; Jain et al., 2016; Patra & Sen, 2016; Ward et al., 2009). Though the nonparametric formulation is highly flexible, it can also be problematic due to the curse-of-dimensionality issues or when the irreducibility assumption is violated (Blanchard et al., 2010; Jain et al., 2016; Patra & Sen, 2016).[1] Such a formulation also leads to difficulties in ensuring unimodality of density components, which is a reasonable practical requirement. To guarantee unimodality of components and allow for the skewness, we model the components with a skew normal family, a generalization of the Gaussian family with good theoretical properties and tractable inference (Genton, 2004). Despite being relatively young (see, e.g., Azzalini, 1985, 1986; Azzalini & Dalla Valle, 1996), the skew normal family also has many practical applications (Abanto-Valle, Lachos, & Dey, 2015; Asparouhov & Muthén, 2016; Bernardi, 2013; Genton, 2004; Hu et al., 2013; Lee & McLachlan, 2013a, 2013b, 2018; Lee, McLachlan, & Pyne, 2014; Lee, McLachlan, & Pyne, 2016; Lin, McLachlan, & Lee, 2016; Lin, Wu, McLachlan, & Lee, 2015; Muthén & Asparouhov, 2015; Pyne et al., 2009; Pyne, Lee, & McLachlan, 2015; Pyne et al., 2014; Riggi & Ingrassia, 2013; Schaarschmidt, Hofmann, Jaki, Grün, & Hothorn, 2015).

Genton (2004) provides several applications of the skew elliptical families, obtained as extensions of skew normal families, in economics, finance, oceanography, climatology, environmetrics, engineering, image processing, astronomy, and biomedical sciences. Lee and McLachlan (2013a) demonstrated the usefulness of skew normal and t-component distributions on six real data sets from different domains such as finance, biology, sports, and image processing on clustering and classification tasks. Muthén and Asparouhov (2015) used skewed t-component mixtures for growth mixture modeling. Bernardi (2013) demonstrated the usefulness of multivariate skew normal mixtures to model assets distribution, by showing that the resulting distribution of the portfolio returns (a linear combination of the assets) is a univariate skew normal. Hu et al. (2013) and Pyne et al. (2014) used multivariate skew normal and t-distributions for cytometric data analysis.

Until recently, the literature on identifiability of parametric mixture models has emphasized identifiability with respect to a subset of parameters; for example, when only a single location parameter, or location and scale parameters, can change, and few studies have considered identifiability of mixtures of general multivariate densities with respect to all of their parameters (Browne & McNicholas, 2015; Holzmann, Munk, & Gneiting, 2006). We take the latter approach and investigate identifiability of two-component mixtures of skew normal distributions with one

[1]The irreducibility assumption constrains the unknown component so that it cannot be expressed as a nontrivial mixture containing the known component.

known component. Though we only focus on identifiability of mixing proportions, the derived conditions imply identifiability of all parameters in a typical well-behaved skew normal family. We begin with a univariate skew normal family (SN) introduced by Azzalini (1985) and, then, extend our results to two forms of multivariate skew normal families (MSN and CFUSN) introduced by Azzalini and Dalla Valle (1996) and Arellano-Valle and Genton (2005), respectively. Our main contributions are presented in Theorems 2–4 that state sufficient conditions for identifiability of mixing proportions in the mixtures of SN, MSN, and CFUSN components, respectively. We also derive a concise formula for the characteristic function of CFUSN in Appendix B.

2 | BACKGROUND AND IDENTIFIABILITY

Let $\mathcal{P}_0$ and $\mathcal{P}_1$ be families of probability density functions (pdf) on $\mathbb{R}^K$. Let $\mathcal{F}(\mathcal{P}_0, \mathcal{P}_1)$ be a family of pdfs having the form

$$f = \alpha f_1 + (1-\alpha) f_0, \tag{1}$$

where $f_0 \in \mathcal{P}_0$, $f_1 \in \mathcal{P}_1$, and $\alpha \in (0, 1)$. Densities f_1 and f_0 will be referred to as component pdfs, f will be referred to as the mixture pdf, and α will be referred to as the mixing proportion. $\mathcal{F}(\mathcal{P}_0, \mathcal{P}_1)$, therefore, is a family of two-component mixtures. We will later restrict $\mathcal{P}_0$ and $\mathcal{P}_1$ to three different skew normal families, one univariate and two multivariate, as defined in (Genton, 2004). An example of this situation is shown in Figure 1, where both f_1 and f_0 belong to the SN (Section 3).

Our goal is to identify the conditions under which estimation of α using samples from f and f_1 is well posed, that is, conditions under which α can be uniquely identified from f and f_1. As a reasonable simplification of having a sample from the component, we assume f_1 to be fixed to discuss identifiability. This is equivalent to restricting $\mathcal{P}_1$ to a singleton set, that is, $\mathcal{P}_1 = \{f_1\}$. With a minor abuse of notation, we denote the family of mixtures $\mathcal{F}(\mathcal{P}_0, \mathcal{P}_1)$ as $\mathcal{F}(\mathcal{P}_0, f_1)$.

Next, we formalize identifiability and derive Theorem 1 that gives a useful technique for proving identifiability, which we will later apply to skew normal mixtures. Once f_1 is fixed, f in Equation (1) can be treated as a pdf parametrized by α and f_0. To reflect this parameterization, we rewrite f as a function of α and f_0; that is, $f : (0, 1) \times \mathcal{P}_0 \to \mathcal{F}(\mathcal{P}_0, f_1)$ given by

$$f(\alpha, f_0) = \alpha f_1 + (1-\alpha) f_0.$$

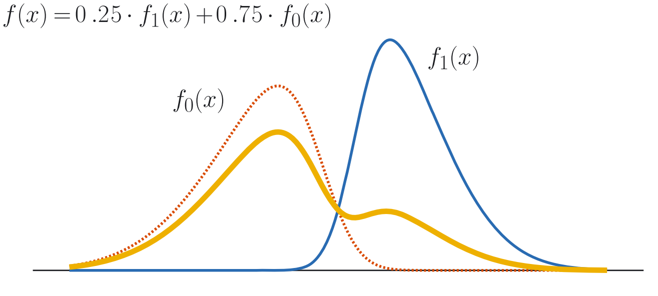

FIGURE 1 Problem illustration: a mixture of two univariate skew normal distributions, as defined later in Section 3. Density functions were drawn as $f_0(x) = \text{SN}(0, 2.5, -3)$ (red curve, left skewed), $f_1(x) = \text{SN}(1, 2, 3)$ (blue curve, right skewed), and $f(x) = \alpha f_1(x) + (1-\alpha) f_0(x)$ with $\alpha = 0.25$ (yellow curve). This work considers a setting when samples are available from $f(x)$ and $f_1(x)$, as indicated by the solid curves (ideally, $f_1(x)$ is known). On the other hand, no sample is available from $f_0(x)$, as shown by the dotted curve. The objective of our work is to derive sufficient identifiability conditions for estimating α [Colour figure can be viewed at wileyonlinelibrary.com]

A family of distributions $\mathcal{G} = \{g_\theta : \theta \in \Theta\}$ is said to be identifiable[2] if the mapping from θ to g_θ is one to one.[3] Therefore, $\mathcal{F}(\mathcal{P}_0, f_1)$ is identifiable if $\forall a, b \in (0, 1)$ and $\forall h_0, g_0 \in \mathcal{P}_0$,

$$f(a, h_0) = f(b, g_0) \Rightarrow (a, h_0) = (b, g_0). \tag{2}$$

The lack of identifiability means that even if (a, h_0) and (b, g_0) are different, the target density f contains no information to tell them apart. If we are only interested in estimating α, we need $\mathcal{F}(\mathcal{P}_0, f_1)$ to be identifiable in α. That is, $\forall a, b \in (0, 1)$ and $\forall h_0, g_0 \in \mathcal{P}_0$,

$$f(a, h_0) = f(b, g_0) \Rightarrow a = b. \tag{3}$$

Identifiability of $\mathcal{F}(\mathcal{P}_0, f_1)$ in α might seem to be a weaker requirement as compared to the identifiability of $\mathcal{F}(\mathcal{P}_0, f_1)$ in (α, f_0). However, Jain et al. (2016) (Lemma 2) showed that the two notions of identifiability are equivalent; that is,

$$\mathcal{F}(\mathcal{P}_0, f_1) \text{ is identifiable if and only if } \mathcal{F}(\mathcal{P}_0, f_1) \text{ is identifiable in } \alpha.$$

Consider now the largest possible $\mathcal{P}_0$, that is, $\mathcal{P}_0$ that contains all pdfs in $\mathbb{R}^K$, except f_1 (or any pdf equal to f_1 almost everywhere). Then, $\mathcal{F}(\mathcal{P}_0, f_1)$ contains all nontrivial two-component mixtures on $\mathbb{R}^K$ with f_1 as one of the components. Lemma C.1 (Appendix) shows that this family is not identifiable. Jain et al. (2016, Lemma 5) gave the following necessary and sufficient condition for identifiability of $\mathcal{F}(\mathcal{P}_0, f_1)$:

$$\mathcal{F}(\mathcal{P}_0, f_1) \text{ is identifiable if and only if } \mathcal{F}(\mathcal{P}_0, f_1) \cap \mathcal{P}_0 = \emptyset.$$

The next lemma gives a sufficient condition for identifiability that is mathematically convenient. Let Span($\mathcal{P}$) denote the span (set of all finite linear combinations) of a set of functions $\mathcal{P}$; that is, $\text{Span}(\mathcal{P}) = \left\{ \sum_{i=1}^{k} a_i g_i : k \in \mathbb{N}, a_i \in \mathbb{R}, g_i \in \mathcal{P} \right\}$. For a given $g \in \mathcal{P}$, let $\Psi_g(t)$ denote a mapping from a $t \in \mathbb{R}^K$ to a real or complex number. We assume that Ψ is linear in g; that is, $\Psi_{ag+bh}(\cdot) = a\Psi_g(\cdot) + b\Psi_h(\cdot)$, for any complex (or real) numbers a, b and any $g, h \in \text{Span}(\mathcal{P})$. Let $\text{Supp}(\Psi_g)$ denote the support of $\Psi_g(\cdot)$; that is, $\text{Supp}(\Psi_g) = \{s \in \mathbb{R}^K : \Psi_g(s) \neq 0\}$.

Theorem 1. *Consider the family of pdfs $\mathcal{F}(\mathcal{P}_0, f_1)$. If for all pairs of pdfs $f_0, g_0 \in \mathcal{P}_0$, there exists (1) a mapping $\Psi_g(t)$ (as defined above) with $g \in \text{Span}(\{f_0, g_0, f_1\})$ and (2) a sequence $\{t_n\}$ in $\text{Supp}(\Psi_{f_1})$ such that*

$$\lim_{n\to\infty} \frac{\Psi_{f_0}(t_n)}{\Psi_{f_1}(t_n)} = 0 \textit{ and } \lim_{n\to\infty} \frac{\Psi_{g_0}(t_n)}{\Psi_{f_1}(t_n)} \notin (-\infty, 0), \tag{4}$$

then $\mathcal{F}(\mathcal{P}_0, f_1)$ is identifiable.

Proof. We give a proof by contradiction. Suppose conditions of the theorem are satisfied but $\mathcal{F}(\mathcal{P}_0, f_1)$ is not identifiable. Thus, $\mathcal{F}(\mathcal{P}_0, f_1) \cap \mathcal{P}_0 \neq \emptyset$, that is, there exists a common element in $\mathcal{F}(\mathcal{P}_0, f_1)$ and $\mathcal{P}_0$, say f_0. Because f_0 is in $\mathcal{F}(\mathcal{P}_0, f_1)$, there exists $g_0 \in \mathcal{P}_0$ such that $f_0 = f(a, g_0)$ for some $a \in (0, 1)$. Because f_0 and g_0 are in $\mathcal{P}_0$, there exists a linear transform Ψ and a sequence $\{t_n\}$ satisfying condition (4). It follows that $f_0 = f(a, g_0) = af_1 + (1 - a)g_0$ and so $\Psi_{f_0}(t) = a\Psi_{f_1}(t) + (1 - a)\Psi_{g_0}(t)$. Now, for all $t \in \text{Supp}(\Psi_{f_1})$, we have $\frac{\Psi_{f_0}(t)}{\Psi_{f_1}(t)} = a + (1 - a)\frac{\Psi_{g_0}(t)}{\Psi_{f_1}(t)}$

[2]In statistics parlance, this is same as the more descriptive phrasing: The parameter θ from the model $\mathcal{G} = \{g_\theta : \theta \in \Theta\}$ is identifiable.
[3]Technically, we require bijection but ignore the obvious "onto" requirement for simplicity.

and, consequently, $\lim_{n\to\infty}\frac{\Psi_{g_0}(t_n)}{\Psi_{f_1}(t_n)} = -\frac{a}{1-a} \in (-\infty, 0)$ (contradiction) because $\{t_n\}$ satisfies $\lim_{n\to\infty}\frac{\Psi_{f_0}(t_n)}{\Psi_{f_1}(t_n)} = 0$ from condition (4). □

We will invoke this lemma later in this paper with two linear transforms, namely, the moment-generating function (MGF) transform and the characteristic function (CF) transform. The main ideas in this lemma (linear transforms and limits) come from Theorem 2 in Teicher (1963) on identifiability of finite mixtures.

3 | TWO-COMPONENT SKEW NORMAL MIXTURES

Gaussian mixtures are widely used in many applications to model real-world data. A straightforward use of Gaussian mixtures in the context of our problem is to assume the components to be Gaussian. However, such an assumption would not be reasonable for skewed components. A possible approach to account for the skewness is to model the components themselves as mixtures of Gaussians; however, finite Gaussian mixtures are still ill-equipped to model the skewness, especially when the component distributions are expected to be unimodal (see Figures D1 and D2 in the Appendix).

In this paper, we derive identifiability results for $\mathcal{F}(\mathcal{P}_0, f_1)$, where f_1 is a fixed skew normal density and $\mathcal{P}_0$ is a subset of a skew normal family in univariate and multivariate cases. Our contributions here are Theorem 2, Theorem 3, and Theorem 4, which give a rather large identifiable family of two-component skew normal mixtures by restricting $\mathcal{P}_0$. A similar approach has been reported by Ghosal and Roy (2011) for mixtures of normal and skew normal distributions. Our result, however, results in a much more extensive family $\mathcal{F}(\mathcal{P}_0, f_1)$. Before giving these results, we first introduce the SN as well as its two most common multivariate generalizations.

Univariate skew normal family: Azzalini (1985) introduced the skew normal (SN) family of distributions as a generalization of the normal family that allows for skewness. It has a location (μ), a scale (ω), and a shape (λ) parameter, where λ controls for skewness. The distribution is right skewed when $\lambda > 0$ and left skewed when $\lambda < 0$, and reduces to a normal distribution when $\lambda = 0$. The pdf of a random variable $X \sim \text{SN}(\mu, \omega, \lambda)$ is given by

$$f(x; \mu, \omega, \lambda) = \frac{2}{\omega}\phi\left(\frac{x-\mu}{\omega}\right)\Phi\left(\frac{\lambda(x-\mu)}{\omega}\right),\ x \in \mathbb{R},$$

where $\mu, \lambda \in \mathbb{R}$, $\omega \in \mathbb{R}^+$, and ϕ and Φ are the probability density function (pdf) and the cumulative distribution function (cdf) of $N(0,1)$ (the standard normal distribution), respectively. Alternatively, the SN family can be parameterized by Δ and Γ (defined in Table 1), instead of λ and ω. The alternate parametrization naturally arises in the stochastic representation of an SN random variable: $X \stackrel{d}{=} \mu + \Delta T + \Gamma^{1/2}U$, where $T \sim TN(0, 1, \mathbb{R}_+)$, the standard normal distribution truncated below 0; $U \sim N(0,1)$, the standard normal distribution; $\stackrel{d}{=}$ reads as "equal in distribution." The expressions for MGF and CF are given in Table 2 (Genton, 2004; Kim & Genton, 2011; Pewsey, 2003).

Multivariate skew normal families: Azzalini and Dalla Valle (1996) proposed an extension of the skew normal family to the multivariate case. The family has a useful property that it is closed under marginalization (Azzalini & Capitanio, 1999). More recently, many other multivariate skew normal families have been proposed (Lee & McLachlan, 2013c). We consider a reparameterization

TABLE 1 Alternate parametrization: the identifiability results are better formulated in terms of the alternate parameters

Family	Alternate parametrization		Related quantities
	canonical → alternate	alternate → canonical	
$\mathrm{SN}(\mu, \omega, \lambda)$	$\Delta = \omega\delta$	$\lambda = \mathrm{sign}(\Delta)\sqrt{\Delta^2/\Gamma}$	$\delta = \frac{\lambda}{\sqrt{1+\lambda^2}}$
	$\Gamma = \omega^2 - \Delta^2$	$\omega = \sqrt{\Gamma + \Delta^2}$	
$\mathrm{MSN}_K(\mu, \Omega, \lambda)$	$\Delta = \Omega^{1/2}\delta$	$\lambda = \frac{\Omega^{-1/2}\Delta}{\sqrt{1-\Delta'\Omega^{-1}\Delta}}$	$\delta = \frac{\lambda}{1+\lambda'\lambda}$
	$\Gamma = \Omega - \Delta\Delta'$	$\Omega = \Gamma + \Delta\Delta'$	
$\mathrm{CFUSN}_{K,M}(\mu, \Omega, \Lambda)$	$\Delta = \Omega^{1/2}\Lambda$	$\Lambda = \Omega^{-1/2}\Delta$	
	$\Gamma = \Omega - \Delta\Delta'$	$\Omega = \Gamma + \Delta\Delta'$	

Note. This table gives the relationship between the alternate and the canonical parameters, as well as some related quantities. SN = univariate skew normal family; MSN = the first multivariate skew normal family; CFUSN = canonical fundamental skew normal family.

TABLE 2 Skew normal families: expression for the characteristic function and moment-generating function

Family	MGF(t)	CF(t)
$\mathrm{SN}(\mu, \omega, \lambda)$	$2\exp\left(t\mu + \frac{1}{2}t^2\omega^2\right)\Phi(\Delta t)$	$\exp\left(\imath t\mu - \frac{1}{2}t^2\omega^2\right)(1 + \imath\mathfrak{I}(\Delta t))$
$\mathrm{MSN}_K(\mu, \Omega, \lambda)$	$2\exp\left(t'\mu + \frac{1}{2}t'\Omega t\right)\Phi(\Delta' t)$	$\exp\left(\imath t'\mu - \frac{1}{2}t'\Omega t\right)(1 + \imath\mathfrak{I}(\Delta' t))$
$\mathrm{CFUSN}_{K,M}(\mu, \Omega, \Lambda)$	$2^M\exp\left\{t'\mu + \frac{1}{2}t'\Omega t\right\}\Phi_M(\Delta' t)$	$\exp\left\{\imath t'\mu - \frac{1}{2}t'\Omega t\right\}\prod_{\delta\in\Delta}(1 + \imath\mathfrak{I}(\delta' t))$

Note. The noncanonical parameters are defined in Table 1. Here, $\imath$ denotes the imaginary number and $\mathfrak{I}(x) = \int_0^x \sqrt{2/\pi}\exp(u^2/2)du$. Φ and Φ_M denote the cdfs of the standard normal and $\boldsymbol{M}$-dimensional multivariate normal with 0 mean and $\mathcal{I}_M$ covariance, respectively. In addition to the $K \times M$ matrix, Δ in the expression for CFUSN CF also represents the multiset containing its column vectors. $t \in \mathbb{R}$ for SN and $t \in \mathbb{R}^K$ for MSN and CFUSN. MGF = moment-generating function; CF = characteristic function; SN = univariate skew normal family; MSN = the first multivariate skew normal family; CFUSN = canonical fundamental skew normal family; cdf = cumulative distribution function.

of the original multivariate skew normal family, MSN, and its generalization referred to as the Canonical fundamental skew normal distribution CFUSN.

MSN: Using the location scale parametrization espoused by Lachos, Ghosh, and Arellano-Valle (2010), the pdf of a K-dimensional random variable $X \sim \mathrm{MSN}_K(\mu, \Omega, \lambda)$ is given by

$$f(x; \mu, \Omega, \lambda) = 2\phi_K(x - \mu \mid \Omega)\Phi(\lambda'\Omega^{-1/2}(x - \mu)),\ x \in \mathbb{R}^K,$$

where Ω is a $K \times K$ covariance matrix; $\mu \in \mathbb{R}^K$ is a column vector giving the location parameter; $\lambda \in \mathbb{R}^K$ is a column vector giving the shape/skewness parameter; $\phi_K(\cdot \mid \Omega)$ is the density function of $N_K(0, \Omega)$, the multivariate normal distribution with Ω covariance; Φ is as defined earlier; and v' is used to denote the transpose of a vector v. As for SN, the stochastic representation of an MSN random variable can be expressed with Δ and Γ (defined in Table 1) more conveniently: $X \overset{d}{=} \mu + \Delta T + \Gamma^{1/2}U$, where $T \sim \mathrm{TN}(0, 1, \mathbb{R}_+)$ and $U \sim N_K(0, \mathcal{I}_K)$. The expressions for MGF and CF are given in Table 2. Since Azzalini and Capitanio (1999) and Kim and Genton (2011) use a different parameterization in the derivation of MGF and CF, we verify the correctness of our formula in Appendix A.

CFUSN: The pdf of a K-dimensional random variable $X \sim \text{CFUSN}_{K,M}(\mu, \Omega, \Lambda)$ is given by

$$f(x; \mu, \Omega, \Lambda) = 2^M \phi_K(x - \mu \mid \Omega)\Phi_M\left(\Lambda'\Omega^{-1/2}(x - \mu) \mid D\right), x \in \mathbb{R}^K,$$

where Ω is a $K \times K$ covariance matrix; $\mu \in \mathbb{R}^K$ is a column vector giving the location parameter; Λ is a $K \times M$ shape/skewness matrix such that $D = \mathcal{I}_M - \Lambda'\Lambda$ is a positive definite matrix, that is, $\|\Lambda a\| < 1$ for any unitary vector $a \in \mathbb{R}^M$; ϕ_K is as defined earlier; $\Phi_M(\cdot \mid D)$ is the cdf of $N_M(0, D)$; and A' denotes the transpose of matrix A. As for SN and MSN, the stochastic representation of a CFUSN random variable can be expressed with Δ and Γ (defined in Table 1) more conveniently: $X \overset{d}{=} \mu + \Delta T + \Gamma^{1/2}U$, where $T \sim \text{TN}_M(0, \mathcal{I}_M, \mathbb{R}^M_+)$, the multivariate normal distribution truncated outside the positive orthant and $U \sim \text{N}_K(0, \mathcal{I}_K)$. The expressions for MGF and CF are given in Table 2. The MGF was obtained from Arellano-Valle and Genton (2005). To the best of our knowledge, the expression for the CF was not available in the literature; we derived it in Theorem B.1 (Appendix) for the purpose of this study.

Lee and McLachlan (2013c) grouped many multivariate skew normal families, including MSN and CFUSN into four groups: restricted, unrestricted, extended, and generalized (in increasing order of generality). MSN belongs to the restricted group, and all the other families in the restricted group are mere reparameterizations. CFUSN was introduced by Arellano-Valle and Genton (2005) as a special case of the fundamental skew normal family (FUSN), the most prominent member of the generalized group. When $K = M$, $\text{CFUSN}_{K,M}$ is identical to Sahu's MSN, the most prominent example of the unrestricted group (Sahu, Dey, & Branco, 2003). Compared to the other multivariate skew normal families, MSN and CFUSN are the most natural and well-known generalizations of SN. Looking at their stochastic representation, in all the three families, skewness is achieved via a truncated normal variable T. T is univariate in the case of SN and MSN, whereas it is multivariate in the case of CFUSN, but in all the three families, 0 is used as a threshold for truncation; that is, only points greater than or equal to 0 have a positive density. The other families in the extended and generalized group either allow T to be truncated with an arbitrary threshold parameter and/or define T as a truncation of a non-Gaussian distribution.

3.1 | Identifiability

In this section, we discuss identifiability of $\mathcal{F}(\mathcal{P}_0, f_1)$ for skew normal families. We first observe that $\mathcal{P}_0$ is the set of distributions and not the set of underlying parameters. Thus, by saying that $\mathcal{F}(\mathcal{P}_0, f_1)$ is identifiable, we mean identifiability in (α, f_0). However, it does not necessarily mean that the parameters corresponding to f_0 are identifiable. In the typical case, when a skew normal family itself is identifiable, that is, there exists a bijection between the set of all skew normal densities in the family and its parameter space, the underlying skew normal parameters are identifiable. It is easy to show that SN and MSN are indeed well-behaved identifiable families; however, CFUSN is not identifiable. This can be easily seen by noting that the characteristic function of CFUSN does not change by permuting the columns of Δ.

Also note that there is a bijection between the alternate and canonical parameterization for all the three families (see Table 1). Thus, identifiability w.r.t. one parameterization implies identifiability w.r.t. the other parameterization as well.

We next give a nontrivial example that shows unidentifiability for two-component SN mixtures in the general case when $\mathcal{P}_0$ is allowed to be the entire SN family, except $f_1 = \text{SN}(\mu_1, \omega_1, \lambda_1)$, where $\lambda_1 \neq 0$. Notice that $\text{SN}(\mu_1, \omega_1, 0) \in \mathcal{P}_0$. Next, we show that $\mathcal{F}(\mathcal{P}_0, f_1)$ also contains $\text{SN}(\mu_1, \omega_1, 0)$, which proves unidentifiability because $\mathcal{F}(\mathcal{P}_0, f_1) \cap \mathcal{P}_0 \neq \emptyset$. For $f_0 = \text{SN}(\mu_1, \omega_1, -\lambda_1) \in \mathcal{P}_0$,

$$\begin{aligned} {}^1\!/_2 f_1(x) + {}^1\!/_2 f_0(x) &= \frac{1}{\omega_1}\phi\left(\frac{x-\mu_1}{\omega_1}\right)\left(\Phi\left(\frac{\lambda_1(x-\mu_1)}{\omega_1}\right) + \Phi\left(\frac{-\lambda_1(x-\mu_1)}{\omega_1}\right)\right) \\ &= \frac{1}{\omega_1}\phi\left(\frac{x-\mu_1}{\omega_1}\right) \qquad \text{(because } \forall x \in \mathbb{R}, \Phi(x)+\Phi(-x)=1\text{)} \\ &= \text{SN}(\mu_1, \omega_1, 0). \end{aligned}$$

Thus, $\text{SN}(\mu_1, \omega_1, 0) \in \mathcal{F}(\mathcal{P}_0, f_1)$.

In Theorems 2,[4] 3, and 4, we show that restricting f_0 to a certain subset, $\mathcal{P}_0$, of SN, MSN, and CFUSN families, respectively, gives a sufficient (but not a necessary) condition for the identifiability of the mixture family $\mathcal{F}(\mathcal{P}_0, f_1)$. The restricted skew normal family, $\mathcal{P}_0$, is defined in terms of the Γ parameter defined in Table 1. It is satisfying that the Γ parameter arises naturally in the stochastic representation of all three families. Note that, in the theorem statements, we define the family and state that it is identifiable, instead of specifying it as a sufficient condition, because the former is more natural stylistically. Further note that any subset of an identifiable family is also identifiable by definition of identifiability. In particular, if $\mathcal{F}(\mathcal{P}_0, f_1)$ is identifiable, a family $\mathcal{F}(\mathcal{P}, f_1)$, where $\mathcal{P} \subseteq \mathcal{P}_0$, is also identifiable.

Theorem 2 (Using the alternate parameterization for SN in Table 1).
The family of pdfs $\mathcal{F}(\mathcal{P}_0, f_1)$ with $f_1 = SN(\mu_1, \omega_1, \lambda_1)$ and

$$\mathcal{P}_0 = \{\text{SN}(\mu, \omega, \lambda) : \Gamma \neq \Gamma_1\} \tag{5}$$

is identifiable.

Proof. Consider a partition of $\mathcal{P}_0$ by sets $\mathcal{P}_0^1, \mathcal{P}_0^2$, defined as follows:

$$\begin{aligned} \mathcal{P}_0^1 &= \{\text{SN}(\mu, \omega, \lambda) : \Gamma > \Gamma_1\} \\ \mathcal{P}_0^2 &= \{\text{SN}(\mu, \omega, \lambda) : \Gamma < \Gamma_1\}. \end{aligned}$$

We now show that, for any given pair of pdfs $f_0, \ddot{f}_0$ from $\mathcal{P}_0$, the conditions of Theorem 1 are satisfied. Let Γ_0, Δ_0 ($\ddot{\Gamma}_0, \ddot{\Delta}_0$) be the parameters corresponding to f_0 ($\ddot{f}_0$), as defined in Table 1.

- If f_0 is from $\mathcal{P}_0^1$ ($\Gamma_0 > \Gamma_1$), we choose the CF transform as Ψ. First, select some $t \neq 0$ in $\mathbb{R}$. Applying Lemma C.4 (Statements 1.a. and 1.b.), we obtain $\lim_{c\to\infty} \frac{\text{CF}(ct;f_0)}{\text{CF}(ct;f_1)} = 0$ and $\lim_{c\to\infty} \frac{\text{CF}(ct;\ddot{f}_0)}{\text{CF}(ct;f_1)} \notin (-\infty, 0)$. Therefore, the sequence $T = \{t_n\}, t_n = nt$ satisfies the conditions of Theorem 1.
- If f_0 is from $\mathcal{P}_0^2$ ($\Gamma_1 > \Gamma_0$), we choose the MGF transform as Ψ. First, we select some $t \neq 0$ in $\mathbb{R}$ with $\Delta_0 t \leq 0$. Applying Lemma C.4 (Statement 2), we obtain $\lim_{c\to\infty} \frac{\text{MGF}(ct;f_0)}{\text{MGF}(ct;f_1)} = 0$. Moreover, owing to the fact that an MGF is always positive, we know

[4]We were able to prove identifiability for a larger SN mixture family, obtained by using $\mathcal{P}_0^b = \{\text{SN}(\mu, \omega, \lambda) : (\mu, \Gamma, |\Delta|) \neq (\mu_1, \Gamma_1, |\Delta_1|)\}$, which subsumes $\{\text{SN}(\mu, \omega, \lambda) : \Gamma \neq \Gamma_1\}$ of Theorem 2 (see supplementary material). Because the stronger result could not be extended to the multivariate families, we did not include it in the main text.

that $\lim_{c\to\infty} \frac{\mathrm{MGF}(ct;\ddot{f}_0)}{\mathrm{MGF}(ct;f_1)} \notin (-\infty, 0)$. The sequence $T = \{t_n\}, t_n = nt$ satisfies the conditions of Theorem 1.

Thus, all the conditions of Theorem 1 are satisfied, and consequently, $\mathcal{F}(\mathcal{P}_0, f_1)$ is identifiable. □

Looking at the expression for the variance of SN distribution, $\omega^2\left(1 - \frac{2}{\pi}\delta^2\right)$ (Lin, Lee, & Yen, 2007) and observing its similarity with $\Gamma = \omega^2(1 - \delta^2)$, we argue that Γ quantifies the dispersion of the distribution—in fact, when $\lambda = 0$, the variance is equal to Γ. Moreover, increasing ω and/or decreasing λ has the same effect on Γ and variance, that is, both increase. Thus, loosely speaking, the condition for identifiability in Theorem 2 is about the difference in the dispersion of the components. In terms of ω and λ, conditions (1) $\omega_1 > \omega_0$ with $\lambda_1 < \lambda_0$ or (2) $\omega_1 < \omega_0$ with $\lambda_1 > \lambda_0$ are both subsumed by $\Gamma_1 \neq \Gamma_0$.

Theorem 3 (Using the alternate parameterization for MSN in Table 1).
The family of pdfs $\mathcal{F}(\mathcal{P}_0, f_1)$ *with* $f_1 = MSN_K(\mu_1, \Omega_1, \lambda_1)$ *and*

$$\mathcal{P}_0 = \{\mathrm{MSN}_K(\mu, \Omega, \lambda) : \Gamma \neq \Gamma_1\}$$

is identifiable.

Proof. Consider a partition of $\mathcal{P}_0$ by sets $\mathcal{P}_0^1$, $\mathcal{P}_0^2$, defined as follows:

$$\mathcal{P}_0^1 = \{\mathrm{MSN}_K(\mu, \Omega, \lambda) : \Gamma \succeq \Gamma_1, \Gamma \neq \Gamma_1\},$$
$$\mathcal{P}_0^2 = \mathcal{P}_0 \setminus \mathcal{P}_0^1,$$

where $\succ$ is the standard partial order relationship on the space of matrices. More specifically, $A \succ B$ implies that $A - B$ is positive definite. Note that $\mathcal{P}_0^2$ also contains pdfs whose Γ matrix is unrelated to Γ_1 by the partial ordering.

We now show that, for any given pair of pdfs $f_0, \ddot{f}_0$ from $\mathcal{P}_0$, the conditions of Theorem 1 are satisfied. Let Γ_0, Δ_0 $(\ddot{\Gamma}_0, \ddot{\Delta}_0)$ be the parameters corresponding to f_0 $(\ddot{f}_0)$, as defined in Table 1.

- If f_0 is from $\mathcal{P}_0^1$, we choose the CF transform as Ψ. We pick some $t \in \mathbb{R}^K$ with $t'(\ddot{\Gamma}_0 - \Gamma_1)t \neq 0$ and $t'(\Gamma_0 - \Gamma_1)t > 0$; the existence of such a t is guaranteed by Lemma C.2. Applying Lemma C.5 (Statements 1.a. and 1.b.), we obtain $\lim_{c\to\infty} \frac{\mathrm{CF}(ct;f_0)}{\mathrm{CF}(ct;f_1)} = 0$ and $\lim_{c\to\infty} \frac{\mathrm{CF}(ct;\ddot{f}_0)}{\mathrm{CF}(ct;f_1)} \notin (-\infty, 0)$. Notice that the sequence $T = \{t_n\}, t_n = nt$ satisfies the conditions of Theorem 1.
- If f_0 is from $\mathcal{P}_0^2$, we choose the MGF transform as Ψ. We pick some $l \neq 0$ in $\mathbb{R}^k$ such that $l'(\Gamma_1 - \Gamma_0)l > 0$; existence of such an l is guaranteed by $\Gamma_0 \nsucceq \Gamma_1$. If the scalar value $\Delta_0' l \leq 0$, we choose $t = l$; otherwise, we choose $t = -l$. It is easy to see that $t'(\Gamma_1 - \Gamma_0)t > 0$ and $\Delta_0' t \leq 0$. Applying Lemma C.5 (Statement 2), we obtain $\lim_{c\to\infty} \frac{\mathrm{MGF}(ct;f_0)}{\mathrm{MGF}(ct;f_1)} = 0$. Moreover, owing to the fact that an MGF is always positive, we know that $\lim_{c\to\infty} \frac{\mathrm{MGF}(ct;\ddot{f}_0)}{\mathrm{MGF}(ct;f_1)} \notin (-\infty, 0)$. The sequence $T = \{t_n\}, t_n = nt$ satisfies the conditions of Theorem 1.

Thus, all the conditions of Theorem 1 are satisfied and consequently $\mathcal{F}(\mathcal{P}_0, f_1)$ is identifiable. □

Here again, Γ quantifies the dispersion of MSN distribution, observing its similarity with the expression for variance: $\Omega - \frac{2}{\pi}\Delta\Delta'$ (Azzalini & Capitanio, 1999). Thus, the difference in the dispersions of the components is beneficial for identifiability.

Notation 1 (Notation for the CFUSN identifiability result used in Theorem 4, Lemma C.3, Lemma C.6).

- θ is used as a placeholder for CFUSN parameters.
- Matrix as a multiset: a matrix U is also used to denote the multiset (to allow duplicates) containing its column vectors. $|U|$ represents its cardinality.
- Null(S): Given a matrix S, Null(S) denotes its null space.
- $\complement A$: Given a set A, $\complement A$ denote its complement.
- $U(t) : U(t) = \{u \in U : u't \neq 0\}$, where $t \in \mathbb{R}^K$ and U is a multiset of vectors in $\mathbb{R}^K$; that is, the multiset containing those vectors in U that are not orthogonal to t.
- $V(c; \theta, \underline{\theta}, t) := \frac{\exp(\imath(\mu-\underline{\mu})'t)\exp(-\frac{1}{2}c^2t'(\Gamma-\underline{\Gamma})t)}{c^{(|\Delta(t)|-|\underline{\Delta}(t)|)}}$, where θ and $\underline{\theta}$ are place holders for CFUSN parameters, $c \in \mathbb{R}$ and $t \in \mathbb{R}^K$.
- $\Xi(U, V, t) := \left(\imath\sqrt{\frac{2}{\pi}}\right)^{(|U(t)|-|V(t)|)} \frac{\prod_{v\in V(t)} v't}{\prod_{u\in U(t)} u't}$, where $\imath$ is the imaginary number.
- $R_N(c, x) := \sum_{n=1}^{N} \frac{(2n-1)!!}{c^{2n}x^{2n}} + \mathcal{O}\left(c^{-2(N+1)}\right) + \mathcal{O}(\exp(-c^2x^2/4))$ as $c \to \infty$, where !! is the standard double factorial notation.

Theorem 4 (Using Notation 1 and the alternate parameterization for CFUSN in Table 1). *The family of pdfs $\mathcal{F}(\mathcal{P}_0, f_1)$ with $f_1 = CFUSN_{K,M}(\mu_1, \Omega_1, \Lambda_1)$ and*

$$\mathcal{P}_0 = \left\{\text{CFUSN}_{K,M}(\mu, \Omega, \Lambda) : \Gamma \neq \Gamma_1, \Gamma_1 - \Gamma \neq kvv',\ \textit{for any } v \in \Delta \textit{ and any } k \in \mathbb{R}^+\right\},$$

is identifiable.[5]

Proof. (Notation (1) is used throughout the proof).
First, we describe a few properties of the V function used at multiple places in the proof. Note that

$$V(c; \theta_0, \theta_1, t) = V(c; \theta_0, \theta, t)V(c; \theta, \theta_1, t),$$

where θ is an arbitrary CFUSN parameter. Reasoning about the asymptotic behaviour of $V(c; \theta_0, \theta_1, t)$ as $c \to \infty$, note that the limit is primarily determined by the sign of the quadratic form $t'(\Gamma_0 - \Gamma_1)t$ and is either 0 or ∞. However, if $t'(\Gamma_0 - \Gamma_1)t = 0$, then the limit is determined by the sign of $|\Delta_0(t)| - |\Delta_1(t)|$ and is still 0 or ∞; if $|\Delta_0(t)| - |\Delta_1(t)| = 0$ as well, then $V(c; \theta_0, \theta_1, t)$ oscillates between -1 and 1 (undefined limit), unless $(\mu_0 - \mu_1)'t = 0$, in which case the limit is 1.

We give a proof by contradiction supposing that the family is not identifiable. Thus, $\mathcal{F}(\mathcal{P}_0, f_1) \cap \mathcal{P}_0 \neq \emptyset$; that is, there exist f_0 and $\ddot{f}_0$ in $\mathcal{P}_0$, such that, with the characteristic function as the linear transform,

$$\ddot{\text{CF}}_0(ct) = a\text{CF}_1(ct) + (1-a)\text{CF}_0(ct), \quad \forall t \in \mathbb{R}^K, \forall c \in \mathbb{R} \text{ and } 0 < a < 1. \tag{6}$$

We will show that Equation (6) leads to a contradiction for all possible choices of f_0 and $\ddot{f}_0$ from $\mathcal{P}_0$. Consider a partition of $\mathcal{P}_0$ by sets $\mathcal{P}_0^1, \mathcal{P}_0^2$, defined as follows:

$$\mathcal{P}_0^1 = \{\text{CFUSN}_{K,M}(\mu, \Omega, \Lambda) : \Gamma_1 \not\succeq \Gamma\},$$
$$\mathcal{P}_0^2 = \mathcal{P}_0 \setminus \mathcal{P}_0^1,$$

[5] Though CFUSN is not an identifiable family itself, a given CFUSN density uniquely determines its Γ parameter (Lemma C.8). Because $\mathcal{P}_0$ is defined in relationship to only the Γ_1 parameter of f_1, the theorem statement is meaningful, even though f_1 might admit multiple representations in the other parameters.

where $\succeq$ is the standard partial order relationship on the space of matrices. Precisely, $A \succeq B$ implies that $A - B$ is positive semidefinite.

Now, consider the following cases that cover all the contingencies.

- If f_0 is from $\mathcal{P}_0^1$ $(\Gamma_1 \nsucceq \Gamma_0)$, we proceed as follows. Equation (6) implies that, for $\mathrm{CF}_0(ct) \neq 0$,

$$\frac{\ddot{\mathrm{CF}}_0(ct)}{\mathrm{CF}_0(ct)} = a\frac{\mathrm{CF}_1(ct)}{\mathrm{CF}_0(ct)} + (1-a)$$
$$\Rightarrow \frac{\frac{\ddot{\mathrm{CF}}_0(ct)}{\mathrm{CF}_0(ct)}}{V(c;\theta_1,\theta_0,t)} = a\frac{\frac{\mathrm{CF}_1(ct)}{\mathrm{CF}_0(ct)}}{V(c;\theta_1,\theta_0,t)} + \frac{1-a}{V(c;\theta_1,\theta_0,t)}$$
$$\Rightarrow \frac{1}{V(c;\theta_1,\ddot{\theta}_0,t)}\underbrace{\frac{\frac{\ddot{\mathrm{CF}}_0(ct)}{\mathrm{CF}_0(ct)}}{V(c;\ddot{\theta}_0,\theta_0,t)}}_{A} = a\underbrace{\frac{\frac{\mathrm{CF}_1(ct)}{\mathrm{CF}_0(ct)}}{V(c;\theta_1,\theta_0,t)}}_{B} + \underbrace{\frac{1-a}{V(c;\theta_1,\theta_0,t)}}_{C}.$$

If $\lim_{c\to\infty}V(c;\theta_1,\theta_0,t) = \infty$, term (C) goes to 0 as $c \to \infty$. Applying Lemma C.6 (Statement 1.a.), the limit of term (B) as $c \to \infty$ exists in $\mathbb{C} \setminus \{0\}$ (the set of nonzero complex numbers), as well as the limit of the entire right-hand side (RHS) and, consequently, the left-hand side (LHS). It follows that, because the limit of term (A) as $c \to \infty$ exists in $\mathbb{C} \setminus \{0\}$, $\lim_{c\to\infty}\frac{1}{V(c;\theta_1,\ddot{\theta}_0,t)}$ should also exist in $\mathbb{C} \setminus \{0\}$ (so that the limit of entire LHS can exist in $\mathbb{C} \setminus \{0\}$). To summarize,

$$\lim_{c\to\infty} V(c;\theta_1,\theta_0,t) = \infty \Rightarrow \lim_{c\to\infty}\frac{1}{V(c;\theta_1,\ddot{\theta}_0,t)} \in \mathbb{C} \setminus \{0\}. \tag{7}$$

Now, we pick some $t \in \mathbb{R}^K$ with $t'(\Gamma_0 - \Gamma_1)t > 0$ and $t'(\ddot{\Gamma}_0 - \Gamma_1)t \neq 0$; existence of such a t is guaranteed by $\ddot{\Gamma}_0 \neq \Gamma_1$ and $\Gamma_1 \nsucceq \Gamma_0$, as shown in Lemma C.2. Because $t'(\Gamma_0 - \Gamma_1)t > 0$, $\lim_{c\to\infty}V(c;\theta_1,\theta_0,t) = \infty$ but $\lim_{c\to\infty}\frac{1}{V(c;\theta_1,\ddot{\theta}_0,t)}$ is either 0 or ∞ as $t'(\ddot{\Gamma}_0 - \Gamma_1)t \neq 0$, which contradicts Equation (7).

- If f_0 is from $\mathcal{P}_0^2$, we proceed as follows.
  - If $(\Gamma_0 = \ddot{\Gamma}_0)$, we use Equation (6) to get

$$\frac{\ddot{\mathrm{CF}}_0(ct)}{\mathrm{CF}_0(ct)} = a\frac{\mathrm{CF}_1(ct)}{\mathrm{CF}_0(ct)} + (1-a)$$
$$\Rightarrow \frac{\frac{\ddot{\mathrm{CF}}_0(ct)}{\mathrm{CF}_0(ct)}}{V(c;\ddot{\theta}_0,\theta_0,t)} = a\frac{\frac{\mathrm{CF}_1(ct)}{\mathrm{CF}_0(ct)}}{V(c;\ddot{\theta}_0,\theta_0,t)} + \frac{1-a}{V(c;\ddot{\theta}_0,\theta_0,t)}$$
$$\Rightarrow \frac{\frac{\ddot{\mathrm{CF}}_0(ct)}{\mathrm{CF}_0(ct)}}{V(c;\ddot{\theta}_0,\theta_0,t)} = \underbrace{\frac{1-a}{V(c;\ddot{\theta}_0,\theta_0,t)}}_{A} + \underbrace{\frac{1}{V(c;\ddot{\theta}_0,\theta_1,t)}a\underbrace{\frac{\frac{\mathrm{CF}_1(ct)}{\mathrm{CF}_0(ct)}}{V(c;\theta_1,\theta_0,t)}}_{B}}_{C}. \tag{8}$$

If $\lim_{c\to\infty}V(c;\ddot{\theta}_0,\theta_1,t) = \infty$, term (C) goes to 0 as $c \to \infty$ because the limit of term (B) exists in $\mathbb{C} \setminus \{0\}$ by Lemma C.6 (Statement 1.a.). Applying Lemma C.6 (Statement 1.a.), the limit of RHS as $c \to \infty$ exists in $\mathbb{C} \setminus \{0\}$, as well as the limit of the entire LHS and consequently term (A); that is, $\lim_{c\to\infty} \frac{1}{V(c;\ddot{\theta}_0,\theta_0,t)} \in \mathbb{C} \setminus \{0\}$. To summarize,

$$\lim_{c\to\infty} V(c;\ddot{\theta}_0,\theta_1,t) = \infty \Rightarrow \lim_{c\to\infty} \frac{1}{V(c;\ddot{\theta}_0,\theta_0,t)} \in \mathbb{C} \setminus \{0\} \tag{9}$$

and

$$\begin{aligned}\lim_{c\to\infty} V(c;\ddot{\theta}_0,\theta_1,t) = \infty \Rightarrow \lim_{c\to\infty} \frac{\frac{\ddot{\mathrm{CF}}_0(ct)}{\mathrm{CF}_0(ct)}}{V(c;\ddot{\theta}_0,\theta_0,t)} &= \lim_{c\to\infty} \frac{1-a}{V(c;\ddot{\theta}_0,\theta_0,t)}\\ \Rightarrow \Xi(\ddot{\Delta}_0,\Delta_0,t) &= \lim_{c\to\infty} \frac{1-a}{V(c;\ddot{\theta}_0,\theta_0,t)} \qquad (10)\\ &\text{(from Lemma C.6, Statement 1.a.).}\end{aligned}$$

Now,

$$\begin{aligned} t \in \complement\mathrm{Null}(\Gamma_1-\Gamma_0) &\Rightarrow t'(\Gamma_1-\Gamma_0)t > 0\\ &\Rightarrow \lim_{c\to\infty} V(c;\ddot{\theta}_0,\theta_1,t) = \infty\\ &\Rightarrow \lim_{c\to\infty} \frac{1}{V(c;\ddot{\theta}_0,\theta_0,t)} \in \mathbb{C}\setminus\{0\} \qquad \text{(from Equation 9)}\\ &\Rightarrow |\Delta_0(t)| - |\ddot{\Delta}_0(t)| = 0 \ \text{ and } \ (\ddot{\mu}_0-\mu_0)'t = 0,\end{aligned}$$

where the last step follows because $V(c;\ddot{\theta}_0,\theta_0,t) = \frac{\exp(\imath c(\ddot{\mu}_0-\mu_0)'t)}{c^{(|\ddot{\Delta}_0(t)|-|\Delta_0(t)|)}}$ when $\Gamma_0 = \ddot{\Gamma}_0$. Consequently,

$$\begin{aligned} t \in \complement\mathrm{Null}(\Gamma_1-\Gamma_0) &\Rightarrow V(c;\ddot{\theta}_0,\theta_0,t) = 1 \qquad (11)\\ &\Rightarrow \Xi(\ddot{\Delta}_0,\Delta_0,t) = 1-\alpha \quad \text{(from Equation 10).}\end{aligned}$$

To summarize, $\forall t \in \complement\mathrm{Null}(\Gamma_1-\Gamma_0)$,

$$|\Delta_0(t)| - |\ddot{\Delta}_0(t)| = 0$$
$$\Xi(\ddot{\Delta}_0,\Delta_0,t) = 1-a.$$

Because $1-\alpha \in (0,1) \subset \mathbb{R}\setminus\{-1,1\}$, from Lemma C.3, it follows that
$\Gamma_1 - \Gamma_0 = kvv'$, for some $v \in \Delta_0$ and some $k \in \mathbb{R}^+$.

Thus, $f_0 \notin \mathcal{P}_0$ and, hence, the contradiction.

– If $\Gamma_0 \neq \ddot{\Gamma}_0$,
Equation (6) implies that, for $\mathrm{CF}_1(ct) \neq 0$,

$$\begin{aligned}\frac{\ddot{\mathrm{CF}}_0(ct)}{\mathrm{CF}_1(ct)} &= a + (1-a)\frac{\mathrm{CF}_0(ct)}{\mathrm{CF}_1(ct)}\\ \Rightarrow \frac{\frac{\ddot{\mathrm{CF}}_0(ct)}{\mathrm{CF}_1(ct)}}{V(c;\theta_0,\theta_1,t)} &= \frac{a}{V(c;\theta_0,\theta_1,t)} + (1-a)\frac{\frac{\mathrm{CF}_0(ct)}{\mathrm{CF}_1(ct)}}{V(c;\theta_0,\theta_1,t)}\\ \Rightarrow \frac{1}{V(c;\theta_0,\ddot{\theta}_0,t)} \underbrace{\frac{\frac{\ddot{\mathrm{CF}}_0(ct)}{\mathrm{CF}_1(ct)}}{V(c;\ddot{\theta}_0,\theta_1,t)}}_{A} &= (1-a)\underbrace{\frac{\frac{\mathrm{CF}_0(ct)}{\mathrm{CF}_1(ct)}}{V(c;\theta_0,\theta_1,t)}}_{B} + \underbrace{\frac{a}{V(c;\theta_0,\theta_1,t)}}_{C}.\end{aligned}$$

Notice that if $\lim_{c\to\infty} V(c;\theta_0,\theta_1,t) = \infty$, then term (C) goes to 0. Applying Lemma C.6 (Statement 1.a.), the limit of term (B) as $c \to \infty$ exists in $\mathbb{C}\setminus\{0\}$, as well as the limit of the entire RHS and, consequently, the LHS. It follows that, because the limit of term (A) as $c \to \infty$ exists in $\mathbb{C} \setminus \{0\}$, $\lim_{c\to\infty} \frac{1}{V(c;\theta_0,\ddot{\theta}_0,t)}$ should also exist in $\mathbb{C} \setminus \{0\}$ (so that the limit of entire LHS exists in $\mathbb{C} \setminus \{0\}$). To summarize,

$$\lim_{c\to\infty} V(c;\theta_0,\theta_1,t) = \infty \Rightarrow \lim_{c\to\infty} \frac{1}{V(c;\theta_0,\ddot{\theta}_0,t)} \in \mathbb{C} \setminus \{0\}. \tag{12}$$

Now then, we pick some $t \in \mathbb{R}^K$ with $t'(\Gamma_1 - \Gamma_0)t > 0$ and $t'(\Gamma_0 - \ddot{\Gamma}_0)t \neq 0$; the existence of such a t is guaranteed by Lemma C.2. $t'(\Gamma_1 - \Gamma_0)t > 0$ ensures that $\lim_{c\to\infty} V(c;\theta_0,\theta_1,t) = \infty$, but $\lim_{c\to\infty} V(c;\theta_0,\ddot{\theta}_0,t)$ is either 0 or ∞ as $t'(\Gamma_0 - \ddot{\Gamma}_0)t \neq 0$, which contradicts Equation (12).

□

As before, Γ quantifies the dispersion of CFUSN distribution, observing its similarity with the expression for variance: $\Omega - \frac{2}{\pi}\Delta\Delta'$ (Arellano-Valle & Genton, 2005). Thus, the difference in the dispersions of the components is beneficial for identifiability.

Comment 1 (Extension of Theorem 4).
We speculate that Theorem 4 can be further strengthened by removing the condition $\Gamma_1 - \Gamma \neq kvv'$ (for any $v \in \Delta$ and any $k \in \mathbb{R}^+$) from the definition of $\mathcal{P}_0$. Removal of this condition breaks the current proof only in the case when $\Gamma_1 \succeq \Gamma_0$ and $\Gamma_0 = \ddot{\Gamma}_0$. Notice that this case implies that, for any $t \in \mathbb{R}^K$ such that $t \in \complement\text{Null}(\Gamma_1 - \Gamma_0)$ satisfies $V(c;\ddot{\theta}_0,\theta_1,t) = \Omega(c^{k_1}\exp(1/2 c^2 t'(\Gamma_1 - \Gamma_0)t))$, for some integer k_1 (from the definition of V), $V(c;\ddot{\theta}_0,\theta_0,t) = 1$ and $\Xi(\ddot{\Delta}_0,\Delta_0,t) = 1 - \alpha$ (as shown in Equation (11)). These implications reduce Equation (8) to

$$a\frac{\frac{\text{CF}_1(ct)}{\text{CF}_0(ct)}}{V(c;\theta_1,\theta_0,t)} = V(c;\ddot{\theta}_0,\theta_1,t)\left(\frac{\frac{\ddot{\text{CF}}_0(ct)}{\text{CF}_0(ct)}}{V(c;\ddot{\theta}_0,\theta_0,t)} - \Xi(\ddot{\Delta}_0,\Delta_0,t)\right)$$
$$= \Omega\left(c^{k_1}\exp\left(\frac{1}{2}c^2 t'(\Gamma_1 - \Gamma_0)t\right)\right)\Xi(\ddot{\Delta}_0,\Delta_0,t)$$
$$\cdot\underbrace{\left(\frac{\prod_{\ddot{\delta}_0\in\ddot{\Delta}_0(t)}\left(1 + R_N\left(c,\ddot{\delta}_0't\right) - \iota\mathcal{O}\left(\frac{c}{\exp\left(\frac{1}{2}c^2(\ddot{\delta}_0't)^2\right)}\right)\right)}{\prod_{\delta_0\in\Delta_0(t)}\left(1 + R_N\left(c,\delta_0't\right) - \iota\mathcal{O}\left(\frac{c}{\exp\left(\frac{1}{2}c^2(\delta_0't)^2\right)}\right)\right)} - 1\right)}_{A},$$

(using Lemma C.6, Statement 1.b.)

as $c \to \infty$ for any positive integer N. Looking at the definition of R_N (Notation 1), it seems that the term (A) should be $\Omega(c^{k_2})$ for some negative integer k_2, except in a difficult-to-characterize special case (when all the polynomial terms in the numerator and denominator of (A) cancel out in the expression given by numerator − denominator). This would imply that the RHS is $\Omega(c^{k_1+k_2}\exp(\frac{1}{2}c^2 t'(\Gamma_1 - \Gamma_0)t))$, which still goes to ∞ as $c \to \infty$, yet the LHS is in $\mathbb{C} \setminus \{0\}$, which leads to a contradiction.

4 | CONCLUSIONS

We give meaningful sufficient conditions that ensure identifiability of two-component mixtures with SN, MSN, and CFUSN components. We proved identifiability in terms of the Γ parameter that contains both the scale and the skewness information and has a consistent interpretation across the three skew normal families as natural parameter in the stochastic representation and as a measure of dispersion. Our results are strong in the sense that the set of parameter values not covered by the sufficient condition is a Lebesgue measure 0 set in the parameter space. Ghosal and Roy (2011) study the identifiability of a two-component mixture with the standard normal as one of the components and the second component itself given by an uncountable mixture of skew normals. Treating G from their work as a point distribution, we can make a valid comparison between our identifiability result and theirs, concluding the superiority of our results, owing to a larger coverage of the parameter space by our conditions. The ability of skew normal families to capture asymmetry, the work-around for high-dimensional data using specialized transforms, and our identifiability results position skew normal mixtures favorably as a parametric model for mixing proportion estimation. Previous work on finite skew normal mixture estimation (without a sample from the component) additionally provides a promising direction for developing practical algorithms (Lachos et al., 2010; Lin, 2009; Lin et al., 2007).

ORCID

Michael Levine https://orcid.org/0000-0002-1060-5711

REFERENCES

Abanto-Valle, C. A., Lachos, V. H., & Dey, D. K. (2015). Bayesian estimation of a skew-student-t stochastic volatility model. *Methodology and Computing in Applied Probability*, *17*, 721–738.

Allman, E. S., Matias, C., & Rhodes, J. A. (2009). Identifiability of parameters in latent structure models with many observed variables. *The Annals of Statistics*, *37*, 3099–3132.

Arellano-Valle, R.B., & Genton, M. G. (2005). On fundamental skew distributions. *Journal of Multivariate Analysis*, *96*, 93–116.

Asparouhov, T., & Muthén, B. (2016). Structural equation models and mixture models with continuous nonnormal skewed distributions. *Structural Equation Modeling: A Multidisciplinary Journal*, *23*, 1–19.

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

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

Azzalini, A., & Capitanio, A. (1999). Statistical applications of the multivariate skew normal distribution. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *61*, 579–602.

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

Bernardi, M. (2013). Risk measures for skew normal mixtures. *Statistics & Probability Letters*, *83*, 1819–1824.

Blanchard, G., Lee, G., & Scott, C. (2010). Semi-supervised novelty detection. *Journal of Machine Learning Research*, *11*, 2973–3009.

Bordes, L., Delmas, C., & Vandekerkhove, P. (2006). Semiparametric estimation of a two-component mixture model where one component is known. *Scandinavian Journal of Statistics*, *33*, 733–752.

Browne, R. P., & McNicholas, P. D. (2015). A mixture of generalized hyperbolic distributions. *The Canadian Journal of Statistics*, *43*, 176–198.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *J R Stat Soc B*, *39*, 1–38.

Denis, F., Gilleron, R., & Letouzey, F. (2005). Learning from positive and unlabeled examples. *Theoretical Computer Science*, *348*, 70–83.

Feller, W. (1968). *An introduction to probability theory and its applications* (Vol. 1). Hoboken, NJ: John Wiley & Sons, Inc.

Genton, M. G. (2004). *Skew-elliptical distributions and their applications: A journey beyond normality*. Boca Raton, FL: CRC Press.

Ghosal, S., & Roy, A. (2011). Identifiability of the proportion of null hypotheses in skew-mixture models for the p-value distribution. *Electronic Journal of Statistics*, *5*, 329–341.

Holzmann, H., Munk, A., & Gneiting, T. (2006). Identifiability of finite mixtures of elliptical distributions. *Scandinavian Journal of Statistics*, *33*, 753–763.

Hu, X., Kim, H., Brennan, P. J., Han, B., Baecher-Allan, C. M., De Jager, P. L., … Raychaudhuri, S. (2013). Application of user-guided automated cytometric data analysis to large-scale immunoprofiling of invariant natural killer T cells. *Proceedings of the National Academy of Sciences of the United States of America*, *110*, 19030–19035.

Jain, S., White, M., & Radivojac, P. (2016). Estimating the class prior and posterior from noisy positives and unlabeled data. In *Advances in Neural Information Processing Systems*, 2693–2701.

Jain, S., White, M., & Radivojac, P. (2017). Recovering true classifier performance in positive-unlabeled learning. *Proceedings of the 31st AAAI Conference on Artificial Intelligence, AAAI 2017, USA*, 2066–2072.

Jain, S., White, M., Trosset, M. W., & Radivojac, P. (2016). Nonparametric semi-supervised learning of class proportions. arXiv preprint arXiv:1601.01944. http://arxiv.org/abs/1601.01944

Kim, H.-M., & Genton, M. G. (2011). Characteristic functions of scale mixtures of multivariate skew-normal distributions. *Journal of Multivariate Analysis*, *102*, 1105–1117.

Lachos, V. H., Ghosh, P., & Arellano-Valle, R. B. (2010). Likelihood based inference for skew-normal independent linear mixed models. *Statistica Sinica*, *20*, 303–322.

Lee, S. X., & McLachlan, G. J. (2013a). Model-based clustering and classification with non-normal mixture distributions. *Statistical Methods & Applications*, *22*, 427–454.

Lee, S. X., & McLachlan, G. J. (2013b). Modelling asset return using multivariate asymmetric mixture models with applications to estimation of value-at-risk. *20th International Congress on Modelling and Simulation (MODSIM2013)*, 1128–1234.

Lee, S. X., & McLachlan, G. J. (2013c). On mixtures of skew normal and skew *t*-distributions. *Advances in Data Analysis and Classification*, *7*, 241–266.

Lee, S. X., & McLachlan, G. J. (2018). Risk measures based on multivariate skew normal and skew *t*-mixture models. In J. Alcock & S. Satchell (Eds.), *Asymmetric Dependence in Finance: diversification, correlation and portfolio management in market downturns* (pp. 152–168). Hoboken, NJ: Wiley.

Lee, S. X., McLachlan, G. J., & Pyne, S. (2014). Supervised classification of flow cytometric samples via the Joint Clustering and Matching (JCM) procedure. arXiv preprint arXiv:1411.2820. https://arxiv.org/abs/1411.2820

Lee, S. X., McLachlan, G. J., & Pyne, S. (2016). Modeling of inter-sample variation in flow cytometric data with the joint clustering and matching procedure. *Cytometry Part A*, *89*, 30–43.

Lin, T. I. (2009). Maximum likelihood estimation for multivariate skew normal mixture models. *Journal of Multivariate Analysis*, *100*, 257–265.

Lin, T. I., Lee, J. C., & Yen, S. Y. (2007). Finite mixture modelling using the skew normal distribution. *Statistica Sinica*, *17*, 909–927.

Lin, T.-I., McLachlan, G. J., & Lee, S. X. (2016). Extending mixtures of factor models using the restricted multivariate skew-normal distribution. *Journal of Multivariate Analysis*, *143*, 398–413.

Lin, T.-I., Wu, P. H., McLachlan, G. J., & Lee, S. X. (2015). A robust factor analysis model using the restricted skew-*t* distribution. *Test*, *24*, 510–531.

McLachlan, G. J., & Peel, D. (2000). *Finite mixture models*. New York, NY: John Wiley & Sons, Inc.

Menon, A. K., van Rooyen, B., Ong, C. S., & Williamson, R. C. (2015). Learning from corrupted binary labels via class-probability estimation. *Proceedings of the 32nd International Conference on Machine Learning, ICML 2015*, 125–134.

Muthén, B., & Asparouhov, T. (2015). Growth mixture modeling with non-normal distributions. *Statistics in Medicine*, *34*, 1041–1058.

Patra, R. K., & Sen, B. (2016). Estimation of a two-component mixture model with applications to multiple testing. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *78*, 869–893.

Pewsey, A. (2003). The characteristic functions of the skew-normal and wrapped skew-normal distributions. *27 Congreso Nacional de Estadistica e Investigación Operativa*.

Pyne, S., Hu, X., Wang, K., Rossin, E., Lin, T.-I., Maier, L. M. … Mesirov, J. P. (2009). Automated high-dimensional flow cytometric data analysis. *Proceedings of the National Academy of Science of the United States of America*, *106*, 8519–8524.

Pyne, S., Lee, S. X., & McLachlan, G. J. (2015). Nature and man: The goal of bio-security in the course of rapid and inevitable human development. *Journal of the Indian Society of Agricultural Statistics*, *69*, 117–125.

Pyne, S., Lee, S. X., Wang, K., Irish, J., Tamayo, P., Nazaire, M.-D., … McLachlan, G. J. (2014). Joint modeling and registration of cell populations in cohorts of high-dimensional flow cytometric data. *PLOS ONE*, *9*, e100334.

Riggi, S., & Ingrassia, S. (2013). A model-based clustering approach for mass composition analysis of high energy cosmic rays. *Astroparticle Physics*, *48*, 86–96.

Sahu, S. K., Dey, D. K., & Branco, M. D. (2003). A new class of multivariate skew distributions with applications to Bayesian regression models. *The Canadian Journal of Statistics*, *31*, 129–150.

Schaarschmidt, F., Hofmann, M., Jaki, T., Grün, B., & Hothorn, L. A. (2015). Statistical approaches for the determination of cut points in anti-drug antibody bioassays. *Journal of Immunological Methods*, *418*, 84–100.

Storey, J. D. (2002). A direct approach to false discovery rates. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *64*, 479–498.

Storey, J. D. (2003). The positive false discovery rate: A Bayesian interpretation and the q-value. *The Annals of Statistics*, *31*, 2013–2035.

Storey, J. D., & Tibshirani, R. (2003). Statistical significance for genomewide studies. *Proceedings of the National Academy of Sciences of the United States of America*, *100*, 9440–9445.

Tallis, G. M. (1961). The moment-generating function of the truncated multi-normal distribution. *Journal of the Royal Statistical Society: Series B (Methodological)*, 223–229.

Tallis, G. M., & Chesson, P. (1982). Identifiability of mixtures. *Journal of the Australian Mathematical Society*, *32*, 339–348.

Teicher, H. (1963). Identifiability of finite mixtures. *The Annals of Mathematical Statistics*, *34*, 1265–1269.

Ward, G., Hastie, T., Barry, S., Elith, J., & Leathwick, J. R. (2009). Presence-only data and the EM algorithm. *Biometrics*, *65*, 554–563.

Yakowitz, S. J., & Spragins, J. D. (1968). On the identifiability of finite mixtures. *The Annals of Mathematical Statistics*, *39*, 209–214.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Additional information for this article is available online, including Appendices A (gives correspondence between different MSN parametrizations to verify the correctness of the CF and MGF formulas), B (derives the formula for CFUSN characteristic function), C (gives the supporting lemmas for the identifiability results), and D (demonstrates the inefficacy of Gaussian mixtures to model unimodal skewed data). A supplementary material extending the identifiability results for the SN case to a larger family is also provided.

How to cite this article: Jain S, Levine M, Radivojac P, Trosset MW. Identifiability of two-component skew normal mixtures with one known component. *Scand J Statist*. 2019;46:955–986. https://doi.org/10.1111/sjos.12377

APPENDIX A

MSN MGF AND CF

Azzalini and Capitanio (1999) and Kim and Genton (2011) define the skewness parameter differently in their definition of MSN pdf. Denoting their skewness parameter as λ_A, the pdf is given by

$$f(x \mid \mu, \Omega, \lambda_A) = 2\phi_K(x-\mu \mid \Omega)\Phi\left(\lambda_A'\omega^{-1}(x-\mu)\right), \quad x \in \mathbb{R}^K,$$

where $\omega = \text{diag}(\sqrt{\omega_{11}}, \sqrt{\omega_{22}}, \dots, \sqrt{\omega_{KK}})$ and ω_{ij} is the entry in the ith row and jth column of Ω. The relationship between the two parameterization can be obtained by observing that the corresponding pdfs are equal when $\lambda'\Omega^{-1/2} = \lambda_A'\omega^{-1}$. It follows that $\lambda = \Omega^{1/2}\omega^{-1}\lambda_A$, and consequently, $\delta = \frac{\Omega^{1/2}\omega^{-1}\lambda_A}{1+\lambda_A'\omega^{-1}\Omega\omega^{-1}\lambda_A} = \frac{\Omega^{1/2}\omega^{-1}\lambda_A}{1+\lambda_A'\Omega_z\lambda_A}$, where $\Omega_z = \omega^{-1}\Omega\omega^{-1}$. The CF and MGF in Azzalini and Dalla Valle (1996) and Kim and Genton (2011) are expressed as $\text{MG}(t) = 2\exp(t'\mu + {}^{t'\Omega t}/_2)\Phi(\delta_A'\omega t)$ and $\text{CF}(t) = \exp(\imath t'\mu - {}^{t'\Omega t}/_2)(1+\imath\mathfrak{I}(\delta_A'\omega t))$, where $\delta_A = \frac{\Omega_z\lambda_A}{1+\lambda_A'\Omega_z\lambda_A}$. Our formula for the CF and MGF in Table 2 can be verified by observing that $\Delta = \Omega^{1/2}\delta = \frac{\Omega\omega^{-1}\lambda_A}{1+\lambda_A'\Omega_z\lambda_A} = \frac{\omega\Omega_z\lambda_A}{1+\lambda_A'\Omega_z\lambda_A} = \omega\delta_A$.

APPENDIX B

CFUSN CHARACTERISTIC FUNCTION

Theorem B.1. *The characteristic function of a K-dimensional* $\text{CFUSN}_{K,M}(\mu, \Omega, \Delta)$ *with a* $K \times M$ Δ *matrix is given by*

$$\text{CF}(t) = \exp\left\{\imath t'\mu - \frac{1}{2}t'\Omega t\right\}\prod_{\delta\in\Delta}(1+\imath\mathfrak{I}(\delta' t)),$$

where Δ *is also used to represent the multiset containing its column vectors,* $\imath$ *is the imaginary number, and* $\mathfrak{I}(x) = \int_0^x \sqrt{2/\pi}\exp(u^2/2)du$.

Proof. We use the stochastic representation of $X \sim \text{CFUSN}(\mu, \Omega, \Delta)$ obtained from Lin (2009), given by $X = \Delta H + G$, where $H \sim TN(0, \mathcal{I}_M, \mathbb{R}_+^M)$, the standard multivariate Normal distribution truncated below 0 in all the dimensions and $G \sim N_K(\mu, \Gamma)$ (multivariate normal distribution) for $\Gamma = \Omega - \Delta\Delta'$—a symmetric positive semidefinite matrix. It follows that the CF of X can be expressed in terms of CFs of Normal distribution and truncated Normal distribution precisely for $t \in \mathbb{R}^K$ $\text{CF}_X(t) = \text{CF}_G(t)\cdot\text{CF}_{\Delta H}(t)$. Using the expression for the CF of multivariate Normal,

$$\text{CF}_X(t) = \exp\left\{\imath t'\mu - \frac{1}{2}t'\Gamma t\right\}\text{CF}_{\Delta H}(t). \tag{B1}$$

The basic properties of a CF and its connection with the corresponding MGF give $\text{CF}_{\Delta H}(t) = \text{CF}_H(\Delta' t) = \text{MGF}_H(\imath\Delta' t)$. Using the expression for the MGF of a Truncated Normal derived in

Tallis (1961, pp. 225) and replacing R (the covariance matrix in Tallis, 1961) by $\mathcal{I}_M$, we get

$$\begin{aligned}
\mathrm{CF}_{\Delta H}(t) &= \mathrm{MGF}_H(\imath\Delta' t) \\
&= \exp\left\{-\frac{1}{2}t'\Delta\Delta' t\right\} \frac{(2\pi)^{-M/2}\int_{\mathbb{R}_+^M}\exp\left\{-\frac{1}{2}(w-\imath\Delta' t)'(w-\imath\Delta' t)\right\}dw}{\int_{\mathbb{R}_+^M}\phi_M(u)du} \\
&\quad \text{(where } \phi_M \text{ is the pdf of } N(0,\mathcal{I}_M)) \\
&= 2^M \exp\left\{-\frac{1}{2}t'\Delta\Delta' t\right\}\int_{\mathbb{R}_+^M}(2\pi)^{-M/2}\exp\left\{-\frac{1}{2}\sum_{i=1}^{M}\left(w_i-\imath\delta_i' t\right)^2\right\}dw \\
&\quad \text{(where } w=[w_i]_{i=1}^M \text{ and } \delta_i \text{ is the } i\text{th column of } \Delta) \\
&= 2^M \exp\left\{-\frac{1}{2}t'\Delta\Delta' t\right\}\int_{\mathbb{R}_+^M}(2\pi)^{-M/2}\prod_{i=1}^{M}\exp\left\{-\frac{1}{2}\left(w_i-\imath\delta_i' t\right)^2\right\}dw \\
&= 2^M \exp\left\{-\frac{1}{2}t'\Delta\Delta' t\right\}\prod_{i=1}^{M}\int_0^\infty(2\pi)^{-1/2}\exp\left\{-\frac{1}{2}\left(w_i-\imath\delta_i' t\right)^2\right\}dw_i.
\end{aligned}$$

Applying the substitution $u_i=-w_i+\imath\delta_i' t$ for the integral in the numerator changes the domain of the integration from the real line to the complex plane. To define such an integral correctly, one needs to specify the path in the complex plane across which the integration is performed. Using the path from $-\infty+\imath\delta_i' t$ to $\imath\delta_i' t$, parallel to the real line, we get

$$\begin{aligned}
\mathrm{CF}_{\Delta H}(t) &= 2^M \exp\left\{-\frac{1}{2}t'\Delta\Delta' t\right\}\prod_{i=1}^{M}\int_{-\infty+\imath\delta_i' t}^{\imath\delta_i' t}(2\pi)^{-1/2}\exp\left\{-\frac{1}{2}u_i^2\right\}du_i \\
&= 2^M \exp\left\{-\frac{1}{2}t'\Delta\Delta' t\right\}\prod_{i=1}^{M}\int_{-\infty+\imath\delta_i' t}^{\imath\delta_i' t}\phi(u_i)du_i.
\end{aligned}$$

Using Lemma 1 from Kim and Genton (2011) to simplify the integral term, we get

$$\begin{aligned}
\mathrm{CF}_{\Delta H}(t) &= 2^M \exp\left\{-\frac{1}{2}t'\Delta\Delta' t\right\}\prod_{i=1}^{M}\left(\frac{1}{2}+\imath\frac{1}{\sqrt{\pi}}\int_0^{\frac{\delta_i' t}{\sqrt{2}}}\exp\left\{u_i^2\right\}du_i\right) \\
&= 2^M \exp\left\{-\frac{1}{2}t'\Delta\Delta' t\right\}\prod_{i=1}^{M}\left(\frac{1}{2}+\imath\frac{1}{2}\int_0^{\delta_i' t}\sqrt{\frac{2}{\pi}}\exp\left\{\frac{v_i^2}{2}\right\}dv_i\right) \\
&\quad \left(\text{substituting } v_i=u_i/\sqrt{2}\right) \\
&= \exp\left\{-\frac{1}{2}t'\Delta\Delta' t\right\}\prod_{i=1}^{M}\left(1+\imath\mathfrak{J}\left(\delta_i' t\right)\right) \\
&= \exp\left\{-\frac{1}{2}t'\Delta\Delta' t\right\}\prod_{\delta\in\Delta}(1+\imath\mathfrak{J}(\delta' t)).
\end{aligned}$$

Substituting the expression for $\mathrm{CF}_{\Delta H}(t)$ in Equation (B1) completes the proof. □

APPENDIX C

SUPPORTING LEMMAS FOR IDENTIFIABILITY RESULTS

Lemma C.1. *If $\mathcal{P}_0$ contains all pdfs on $\mathbb{R}^K$ except f_1, then $\mathcal{F} = \mathcal{F}(\mathcal{P}_0, f_1)$ is not identifiable.*

Proof. Because $\mathcal{P}_0$ contains all pdfs on $\mathbb{R}^K$ except f_1, we have $\mathcal{F} \subseteq \mathcal{P}_0$ (note that $f_1 \notin \mathcal{F}$ either, because α cannot be 1). Let $a \in (0,1)$ and $b \in (0, a)$, $h_0 \in \mathcal{P}_0$ and $g_0 = f\left({}^{(a-b)}/_{(1-b)}, h_0\right)$. As g_0 is a mixture in $\mathcal{F}$ and $\mathcal{F} \subseteq \mathcal{P}_0$, it follows that g_0 is also in $\mathcal{P}_0$. Consequently, the mixture $f(b, g_0)$ is in $\mathcal{F}$. Therefore, $f(b, g_0) = bf_1 + (1-b)g_0 = bf_1 + (1-b)f\left({}^{(a-b)}/_{(1-b)}, h_0\right)$; the last expression is equivalent to $f(a, h_0)$. Thus, we have $f(a, h_0) = f(b, g_0)$. However, $b \neq a$, and hence, $\mathcal{F}$ is not identifiable. □

Lemma C.2. *For $K \times K$ symmetric matrices $A \neq 0$ and $B \neq 0$, if either $A \geq 0$ or $A \nleq 0$, then there exists a vector $t \in \mathbb{R}^K$ such that $t'Bt \neq 0$ and $t'At > 0$.*

Proof. Suppose there does not exist any vector $l \in \mathbb{R}^K$ such that $l'Al > 0$. Thus, for all $l \in \mathbb{R}^K$, $l'Al \leq 0$. This immediately contradicts $A \nleq 0$. Hence, $A \nleq 0$ implies that there exists $l \in \mathbb{R}^K$ such that $l'Al > 0$. On the other hand, $A \geq 0$ implies $l'Al \geq 0$ for all $l \in \mathbb{R}^K$. This, in combination with $l'Al \leq 0$ for all $l \in \mathbb{R}^K$, implies that $l'Al = 0$ for all $l \in \mathbb{R}^K$. This, however, is impossible because $A \neq 0$. To summarize, there exists $l \in \mathbb{R}^K$ such that $l'Al > 0$ when $A \neq 0$ and either of $A \geq 0$ or $A \nleq 0$ is true. Now, we give a recipe to find $t \in \mathbb{R}^K$ with $t'Bt \neq 0$ and $t'At > 0$. Let l be some vector in $\mathbb{R}^K$ with $l'Al > 0$ (existence of l already proved)

- If $l'Bl \neq 0$, then choose $t = l$
- else ($l'Bl = 0$), let $l_1 \in \mathbb{R}^K$ be such that $l_1'Bl_1 \neq 0$. Existence of such l_1 is guaranteed because $B \neq 0$. We choose $t = l + \epsilon l_1$, where $\epsilon > 0$ is picked so that $t'Bt \neq 0$ and $t'At > 0$. To see that such an ϵ exists, notice first that $t'Bt = (l + \epsilon l_1)'B(l + \epsilon l_1) = lBl' + 2\epsilon l_1'Bl + \epsilon^2 l_1'Bl_1 = 2\epsilon l_1'Bl + \epsilon^2 l_1'Bl_1 \neq 0$ for any $\epsilon \neq \frac{-2l_1'Bl}{l_1'Bl_1}$. Second, $t'At = (l + \epsilon l_1)'A(l + \epsilon l_1) = lAl' + 2\epsilon l_1'Al + \epsilon^2 l_1'Al_1 > 0$ for a small-enough $\epsilon > 0$. Thus, picking a small-enough $\epsilon \neq \frac{-2l_1'Bl}{l_1'Bl_1}$ ensures $t'Bt \neq 0$ and $t'At > 0$. □

Notation C.1 (Notation for Lemma C.3).

- $t \equiv l$: Given vectors $t, l \in \mathbb{R}^K$, $t \equiv l$ if $ct = l$ for some $c \neq 0$ in $\mathbb{R}$; that is, t and l have the same direction or opposite direction.
- $\mathbb{P}(U)$: Given a multiset of vectors U, $\mathbb{P}(U)$ is a partition of U defined by the equivalence relationship defined above; that is, vectors having the same or opposite direction are in the same block of the partition.
- P_C: Given $P \in \mathbb{P}(U)$, P_C denotes the canonical vector direction of P. Because P can potentially contain vectors having an opposite direction, strictly speaking, there are two canonical directions. However, for simplicity, we pick an arbitrary vector $t \in P$ and define P_C as $P_C = t/||t||$ when a $t \neq 0$ and $P_C = 0$ otherwise. This abuse of notation does not affect the result of Lemma C.3.
- $t^{\perp}$: Given a vector $t \in \mathbb{R}^K$, $t^{\perp}$ denotes the $K - 1$, the vector space orthogonal to t.

Lemma C.3 (Using Notation 1 and C.1).
Let U, V be $K \times M$ matrices and $S \neq 0$ be a $K \times K$ symmetric positive semidefinite matrix. Let $\mathbb{P} = \mathbb{P}(U \cup V)$. Suppose $|U(t)| = |V(t)$ and $\Xi(U, V, t) = r, \forall t \in \complement\text{Null}(S)$ and some $r \in \mathbb{R} \setminus \{-1, 1\}$. It follows that $S = kvv'$ for some $v \in V$ and some constant $k > 0$; that is, S is a Rank 1 matrix with all column (and row) vectors in the same or opposite direction as some column vector of V.

Proof.
Part 1 (Partitioning $\mathbb{P}$): First, we partition the elements of $\mathbb{P}$ into three sets $\mathbb{P}_0$, $\mathbb{P}_1$, and $\mathbb{P}_2$ defined below, showing that $\mathbb{P}_0$ and $\mathbb{P}_2$ cannot have more than one element and connect the nonemptiness of $\mathbb{P}_2$ with the desired result:

$$\mathbb{P}_0 = \{P \in \mathbb{P} : P_C \equiv 0\},$$
$$\mathbb{P}_1 = \{P \in \mathbb{P} : P_C \not\equiv s \text{ for some} s \neq 0 \text{ in } S\},$$
$$\mathbb{P}_2 = \mathbb{P} \setminus (\mathbb{P}_0 \cup \mathbb{P}_1).$$

Notice that $\mathbb{P}_0$ is either a singleton or empty because all the 0 vectors in $U \cup V$ are collected in a single component set in $\mathbb{P}$. If $\mathbb{P}_2 \neq \emptyset$, then any vector w in any $\tilde{P} \in \mathbb{P}_2$ is equivalent to all nonzero column vectors in S, which implicitly means that all nonzero column vectors in S are in the same or opposite direction (equivalent) and, consequently, S is Rank 1 matrix having column vectors (and row vectors as S is symmetric) equivalent to w. In other words, S can be expressed as $S = k_1 ww'$ for some constant $k_1 > 0$ ($k_1 > 0$ ensures S is positive semidefinite). To summarize,

$$\mathbb{P}_2 \neq \emptyset \Rightarrow S = k_1 ww', \text{ for a } w \in \tilde{P} \text{ from } \mathbb{P}_2 \text{ and } k_1 > 0. \tag{C1}$$

Moreover, any other vector that can appear inside $\mathbb{P}_2$ is equivalent to w, and consequently, $\mathbb{P}_2$ is also singleton set (if not empty).

Part 2 (t_0,$t_{\bar{P}}$): We define the following vectors and prove their existence.

- $t_0 : t_0 \in \complement\text{Null}(S)$ such that $P'_C t_0 \neq 0, \forall P \in \mathbb{P} \setminus \mathbb{P}_0$. Existence of t_0 is shown by using result (A) (given below) with $e = 0$ and $H = \{P_C : P \in \mathbb{P} \setminus \mathbb{P}_0\} \cup \{s\}$ for some $s \neq 0$ in S.
- $t_{\bar{P}}$: For a given $\bar{P} \in \mathbb{P}_1$, $t_{\bar{P}} \in \complement\text{Null}(S)$ such that $\bar{P}'_C t_{\bar{P}} = 0$, $P'_C t_{\bar{P}} \neq 0, \forall P \in \mathbb{P} \setminus (\{\bar{P}\} \cup \mathbb{P}_0)$. Existence of $t_{\bar{P}}$ is shown by using result (A) (given below) with $e = \bar{P}_C$ and $H = \{P_C : P \in \mathbb{P} \setminus (\{\bar{P}\} \cup \mathbb{P}_0)\} \bigcup \{s\}$, where $s \neq 0$ in S be such that $s \not\equiv \bar{P}_C$ (such an s exists by definition of $\mathbb{P}_1$).

Result (A): For a given vector e and a finite multiset of nonzero vectors H in $\mathbb{R}^K$,

$$e \not\equiv h, \forall h \in H \Rightarrow \exists t \in \mathbb{R}^K \text{ such that } e't = 0 \text{ and } h't \neq 0, \forall h \in H.$$

To prove (A), notice that choosing t from $e^\perp$ guarantees $e't = 0$. Choosing t from $\complement h^\perp$ ensures $h't \neq 0$. It follows that if the set, G, obtained by removing $h^\perp$, for all $h \in H$, from $e^\perp$ is nonempty, then any $t \in G$ satisfies both $h't \neq 0$ and $e't = 0$. To see that G is indeed nonempty, notice that removing $h^\perp$'s (finite number of $K-1$ dimensional linear spaces) from $e^\perp$ (either K dimensional when $e = 0$ or $K-1$ dimensional when $e \neq 0$) reduces it only by Lebesgue measure 0 set, provided $e^\perp$ does not coincide with any of the $h^\perp$'s, guaranteed by $e \not\equiv h$ for all $h \in H$.

Part 3 (Equal contribution from U,V): Next, we show that any $P \in \mathbb{P}$ has even an number of elements with equal contribution from U and V; that is,

$$|U \cap P| - |V \cap P| = 0, \forall P \in \mathbb{P}. \tag{C2}$$

We break the argument into three exhaustive cases, picking P from $\mathbb{P}_0$ or $\mathbb{P}_1$ or $\mathbb{P}_2$ as follows.

1. $\check{P} \in \mathbb{P}_0$: Because $t_0 \in \complement\text{Null}(S)$, $|U(t_0)| = |V(t_0)|$. Thus, the number of nonzero and, consequently, zero entries in $U' t_0$ and $V' t_0$ are equal (U and V have the same number of columns). The only source of 0's in $U' t_0$ and $V' t_0$ are column vectors in $\check{P}$ and consequently, $|U \cap \check{P}| - |V \cap \check{P}| = 0$ follows.
2. $\bar{P} \in \mathbb{P}_1$: Because $t_{\bar{P}} \in \complement\text{Null}(S)$, $|U(t_{\bar{P}})| = |V(t_{\bar{P}})|$. Thus, the number of nonzero and, consequently, zero entries in $U' t_{\bar{P}}$ and $V' t_{\bar{P}}$ are equal (U and V have the same number of columns). There are two possibilities for the source of 0s in $U' t_{\bar{P}}$ and $V' t_{\bar{P}}$:
   a. column vectors in $\bar{P}$ only (when $\mathbb{P}_0 = \emptyset$). Thus, to satisfy $|U(t_{\bar{P}})| = |V(t_{\bar{P}})|$, $|U \cap \bar{P}| - |V \cap \bar{P}| = 0$ must be true.
   b. column vectors in $\bar{P}$ and the only element in $\mathbb{P}_0$, $\check{P}$ (when $\mathbb{P}_0$ is singleton). We already know from case (1) that $|U \cap \check{P}| - |V \cap \check{P}| = 0$ is true, and consequently, to satisfy $|U(t_{\bar{P}})| = |V(t_{\bar{P}})|$, $|U \cap \bar{P}| - |V \cap \bar{P}| = 0$ must be true as well.
3. $\tilde{P} \in \mathbb{P}_2$: Because $\tilde{P}$ is the only element in $\mathbb{P}_2$, all other sets $P \in \mathbb{P} \setminus \{\tilde{P}\}$ belong to either $\mathbb{P}_0$ or $\mathbb{P}_1$ and are covered by cases (1) and (2); that is, $|U \cap P| - |V \cap P| = 0$. As a consequence, $\tilde{P}$, being the only remaining set, $|U \cap \tilde{P}| - |V \cap \tilde{P}| = 0$ must be true because both U and V have the equal number of column vectors.

Part 4: We rewrite the formula for $\Xi(U, V, l)$ as follows:

$$\Xi(U, V, l) = \prod_{P \in \mathbb{P} \setminus \mathbb{P}_0, P_C' l \neq 0} \frac{\prod_{v \in V \cap P} \frac{1}{\iota} \sqrt{\frac{\pi}{2}} v' l}{\prod_{u \in U \cap P} \frac{1}{\iota} \sqrt{\frac{\pi}{2}} u' l}.$$

Taking the absolute value squared,

$$\begin{aligned}
|\Xi(U, V, l)|^2 &= \prod_{P \in \mathbb{P} \setminus \mathbb{P}_0, P_C' l \neq 0} \frac{\prod_{v \in V \cap P} \frac{\pi}{2} (v' l)^2}{\prod_{u \in U \cap P} \frac{\pi}{2} (u' l)^2} \\
&= \prod_{P \in \mathbb{P} \setminus \mathbb{P}_0, P_C' l \neq 0} \frac{\prod_{v \in V \cap P} \frac{\pi}{2} \|v\|^2 \left(P_C' l\right)^2}{\prod_{u \in U \cap P} \frac{\pi}{2} \|u\|^2 \left(P_C' l\right)^2} \\
&= \prod_{P \in \mathbb{P} \setminus \mathbb{P}_0, P_C' l \neq 0} \left(\frac{\pi}{2} \left(P_C' l\right)^2 \right)^{|V \cap P| - |U \cap P|} \frac{\prod_{v \in V \cap P} \|v\|^2}{\prod_{u \in U \cap P} \|u\|^2} \\
&= \prod_{P \in \mathbb{P} \setminus \mathbb{P}_0, P_C' l \neq 0} \frac{\prod_{v \in V \cap P} \|v\|^2}{\prod_{u \in U \cap P} \|u\|^2} \quad \text{(Using Equation (C2))} \qquad \text{(C3)}
\end{aligned}$$

$$\begin{aligned}
&= \prod_{P \in \mathbb{P}_1, P_C' l \neq 0} \frac{\prod_{v \in V \cap P} \|v\|^2}{\prod_{u \in U \cap P} \|u\|^2} \cdot \prod_{P \in \mathbb{P}_2, P_C' l \neq 0} \frac{\prod_{v \in V \cap P} \|v\|^2}{\prod_{u \in U \cap P} \|u\|^2} \\
&= \prod_{\bar{P} \in \mathbb{P}_1, \bar{P}_C' l \neq 0} \frac{|\Xi(U, V, t_0)|^2}{|\Xi(U, V, t_{\bar{P}})|^2} \cdot \prod_{\tilde{P} \in \mathbb{P}_2, \tilde{P}_C' l \neq 0} \frac{\prod_{v \in V \cap \tilde{P}} \|v\|^2}{\prod_{u \in U \cap \tilde{P}} \|u\|^2} \quad \text{(Using Equation (C3))} \\
&= \prod_{\tilde{P} \in \mathbb{P}_2, \tilde{P}_C' l \neq 0} \frac{\prod_{v \in V \cap \tilde{P}} \|v\|^2}{\prod_{u \in U \cap \tilde{P}} \|u\|^2} \quad \text{(because } \Xi(U, V, t_0) = r \text{ and } \Xi(U, V, t_{\bar{P}}) = r\text{).}
\end{aligned}$$

Thus, $\mathbb{P}_2 = \emptyset$ is a sufficient condition for $|\Xi(U,V,l)|^2 = 1, \forall l \in \mathbb{R}^K$, but the condition $r \in \mathbb{R} \setminus \{-1,1\}$ leads to a contradiction. Thus,

$$\begin{aligned} r \in \mathbb{R} \setminus \{-1,1\} \Rightarrow \mathbb{P}_2 \neq \emptyset & \\ \Rightarrow S = k_1 ww', \text{ for a } w \in \tilde{P} \text{ from } \mathbb{P}_2 \text{ and some } k_1 > 0 & \quad \text{(from Equation C1)} \\ \Rightarrow S = kvv', \text{ for some } v \in V \cap \tilde{P} \text{ and some } k > 0; & \end{aligned}$$

existence of v is justified by Equation (C2) and the fact that $\tilde{P}$ is nonempty. □

Notation C.2 (Landau's notation).
We use Landau's asymptotic notation in the next few lemmas, defined as follows. For real-valued functions g and h defined on some subset of $\mathbb{R}$, $g(c) = \mathcal{O}(h(c))$ as $c \to \infty$ if $\limsup_{c\to\infty} |\frac{g(c)}{h(c)}| < \infty$ and $g(c) = \Omega(h(c))$ as $c \to \infty$ if $\limsup_{c\to\infty} |\frac{g(c)}{h(c)}| > 0$.

Lemma C.4 (Using the alternate parameterization for SN in Table 1)**.**
Consider two univariate skew normal distributions, $SN(\mu, \omega, \lambda)$ and $SN(\underline{\mu}, \underline{\omega}, \underline{\lambda})$. Let $c \in \mathbb{R}$ and $t \in \mathbb{R} \setminus \{0\}$.

1. *Let* CF *and* $\underline{\text{CF}}$ *be the characteristic functions corresponding to the two distributions (refer to Table 2).*
   1.a.
$$\Gamma - \underline{\Gamma} > 0 \Rightarrow \lim_{c\to\infty} \frac{\text{CF}(ct)}{\underline{\text{CF}}(ct)} = 0$$
   1.b.
$$\Gamma - \underline{\Gamma} \neq 0 \Rightarrow \lim_{c\to\infty} \frac{\text{CF}(ct)}{\underline{\text{CF}}(ct)} \in \{-\infty, 0, \infty\},$$
   provided the limit exists in $\overline{\mathbb{R}}$ (the extended real number line).
2. *Let* MGF *and* $\underline{\text{MGF}}$ *be the moment-generating functions corresponding to the two distributions (refer to Table 2). For $\Delta t \leq 0$,*
$$\underline{\Gamma} - \Gamma > 0 \Rightarrow \lim_{c\to\infty} \frac{\text{MGF}(ct)}{\underline{\text{MGF}}(ct)} = 0.$$

Proof. Here, we use Landau's $\mathcal{O}(\cdot)$ and $\Omega(\cdot)$ notation, defined in Notation (C.2).

Statement 1.a.: Instead of working directly with $\frac{\text{CF}(ct)}{\underline{\text{CF}}(ct)}$, which can be complex, we circumvent the complication by working with the ratio's absolute value squared, which is always real. Multiplying the ratio with its conjugate, we obtain an expression of its absolute value squared as follows:

$$\begin{aligned} \left|\frac{\text{CF}(ct)}{\underline{\text{CF}}(ct)}\right|^2 &= \frac{\text{CF}(ct)}{\underline{\text{CF}}(ct)} \overline{\left(\frac{\text{CF}(ct)}{\underline{\text{CF}}(ct)}\right)} \\ &= \frac{\text{CF}(ct)\overline{(\text{CF}(ct))}}{\underline{\text{CF}}(ct)\overline{(\underline{\text{CF}}(ct))}} \qquad \text{(property of complex conjugate of a fraction)} \\ &= \frac{\exp\left(-c^2\omega_0^2 t^2\right)(1 + (\Im(c\Delta t))^2)}{\exp\left(-c^2 t^2 \omega_1^2 t\right)(1 + (\Im(c\underline{\Delta} t))^2)}. \end{aligned}$$

Consider the ratio $\frac{1+(\Im(c\Delta t))^2}{1+(\Im(c\underline{\Delta} t))^2}$ from the previous expression. Using the asymptotic upper bound (for the numerator) and lower bound (for the denominator), obtained in Lemma C.7

(Statement 2c and 2d), we get

$$\frac{1+(\Im(c\Delta t)^2}{1+(\Im(c\underline{\Delta}t))^2} = \mathcal{O}\left(c^2 \exp\left(c^2(\Delta t)^2 - c^2(\underline{\Delta}t)^2\right)\right)$$
$$= \mathcal{O}\left(c^2 \exp\left(c^2 t^2\left(\Delta^2 - \underline{\Delta}^2\right)\right)\right).$$

Thus,

$$\left|\frac{\mathrm{CF}(ct)}{\underline{\mathrm{CF}}(ct)}\right|^2 = \exp\left(-c^2 t^2\left(\omega^2 - \underline{\omega}^2\right)\right)\mathcal{O}\left(c^2 \exp\left(c^2 t^2\left(\Delta^2 - \underline{\Delta}^2\right)\right)\right)$$
$$= \mathcal{O}\left(c^2 \exp\left(-c^2 t^2\left(\left(\omega^2 - \Delta^2\right) - \left(\underline{\omega}^2 - \underline{\Delta}^2\right)\right)\right)\right)$$
$$= \mathcal{O}\left(c^2 \exp\left(-c^2 t^2(\Gamma - \underline{\Gamma})\right)\right). \tag{C4}$$

Consequently,

$$\lim_{c\to\infty}\left|\frac{\mathrm{CF}(ct)}{\underline{\mathrm{CF}}(ct)}\right|^2 = 0, \quad \text{when } \Gamma - \underline{\Gamma} > 0$$

and

$$\lim_{c\to\infty}\frac{\mathrm{CF}(ct)}{\underline{\mathrm{CF}}(ct)} = 0, \quad \text{when } \Gamma - \underline{\Gamma} > 0$$

follows.

Statement 1.b. Similar to the derivation of the asymptotic upper bound for the ratio in Equation (C4), we derive the asymptotic lower bound by using Lemma C.7 (Statement 2c and 2d);

$$\left|\frac{\mathrm{CF}(ct)}{\underline{\mathrm{CF}}(ct)}\right|^2 = \Omega\left(\frac{1}{c^2}\exp\left(c^2 t^2(\underline{\Gamma} - \Gamma)\right)\right).$$

Consequently,

$$\lim_{c\to\infty}\left|\frac{\mathrm{CF}(ct)}{\underline{\mathrm{CF}}(ct)}\right|^2 = \infty, \quad \text{when } \underline{\Gamma} - \Gamma > 0$$

and

$$\lim_{c\to\infty}\frac{\mathrm{CF}(ct)}{\underline{\mathrm{CF}}(ct)} \in \{-\infty, \infty\}, \quad \text{when } \underline{\Gamma} - \Gamma > 0$$

follows, provided the limit exists in $\overline{\mathbb{R}}$. Combining the result with Statement 1.a. proves Statement 1.b.

Statement 2 From the definition of SN MGF (Table 2), we get

$$\frac{\mathrm{MGF}(ct)}{\underline{\mathrm{MGF}}(ct)} = \exp\left(c(\mu - \underline{\mu})t - \frac{c^2}{2}t^2\left(\underline{\omega}^2 - \omega^2\right)\right)\frac{\Phi(c\Delta t)}{\Phi(c\underline{\Delta}t)}.$$

Consider the ratio $\frac{\Phi(c\Delta t)}{\Phi(c\underline{\Delta}t)}$ from the previous expression. We apply the asymptotic upper bound (for the numerator) and lower bound (for the denominator), obtained in Lemma C.7 (Statement 1.a. and 1.b.). Because $\Delta t \leq 0$, the asymptotic upper bound is applicable.

$$\frac{\Phi(c\Delta t)}{\Phi(c\underline{\Delta}t)} = \mathcal{O}\left(c\exp\left(-\frac{c^2}{2}\left((\Delta t)^2 - (\underline{\Delta}t)^2\right)\right)\right)$$
$$= \mathcal{O}\left(c\exp\left(-\frac{c^2}{2}t^2\left(\Delta^2 - \underline{\Delta}^2\right)\right)\right)$$

Thus,

$$\begin{aligned}\frac{\mathrm{MGF}(ct)}{\underline{\mathrm{MGF}}(ct)} &= \exp\left(c(\mu-\underline{\mu})t-\frac{c^2}{2}t^2\left(\underline{\omega}^2-\omega^2\right)\right)\mathcal{O}\left(c\exp\left(-\frac{c^2}{2}t^2\left(\Delta^2-\underline{\Delta}^2\right)\right)\right)\\ &= \mathcal{O}\left(c\exp\left(c(\mu-\underline{\mu})t-\frac{c^2}{2}t^2\left(\left(\underline{\omega}^2-\underline{\Delta}^2\right)-\left(\omega^2-\Delta^2\right)\right)\right)\right)\\ &= \mathcal{O}\left(c\exp\left(c(\mu-\underline{\mu})t-\frac{c^2}{2}t^2(\underline{\Gamma}-\Gamma)\right)\right).\end{aligned}$$

Because c^2 term dominates the c term in the exponential above, the asymptotic upper bound goes to 0 when $\underline{\Gamma}-\Gamma>0$, irrespective of the relation between μ and $\underline{\mu}$. Consequently,

$$\lim_{c\to\infty}\frac{\mathrm{MGF}(ct)}{\underline{\mathrm{MGF}}(ct)}=0,\quad \text{when }\ \underline{\Gamma}-\Gamma>0.$$

□

Lemma C.5 (Using the alternate parameterization for MSN in Table 1). *Consider two K-dimensional skew Normal distributions,* $\mathrm{MSN}_K(\mu,\Omega,\lambda)$, *and* $\mathrm{MSN}_K(\underline{\mu},\underline{\Omega},\underline{\lambda})$. *Let* $c\in\mathbb{R}$ *and* $t\in\mathbb{R}^K$.

1. *Let* CF *and* $\underline{\mathrm{CF}}$ *be the characteristic functions corresponding to the two distributions (refer to Table 2).*

   1.a.

   $$t'(\Gamma-\underline{\Gamma})t>0\Rightarrow\lim_{c\to\infty}\frac{\mathrm{CF}(ct)}{\underline{\mathrm{CF}}(ct)}=0$$

   1.b.

   $$t'(\Gamma-\underline{\Gamma})t\neq0\Rightarrow\lim_{c\to\infty}\frac{\mathrm{CF}(ct)}{\underline{\mathrm{CF}}(ct)}\in\{-\infty,0,\infty\},$$

   provided the limit exists in $\overline{\mathbb{R}}$ *(the extended real number line).*

2. *Let* MGF *and* $\underline{\mathrm{MGF}}$ *be the moment-generating functions corresponding to the two distributions (refer to Table 2). For* $\Delta' t\le 0$,

   $$t'(\underline{\Gamma}-\Gamma)t>0\Rightarrow\lim_{c\to\infty}\frac{\mathrm{MGF}(ct)}{\underline{\mathrm{MGF}}(ct)}=0.$$

Proof. Here, we use Landau's $\mathcal{O}(\cdot)$ and $\Omega(\cdot)$ notation, defined in Notation (C.2).
Statement 1.a.: We use the approach in Lemma C.4. The expression for the squared absolute value of the characteristic function ratio, obtained by multiplying the ratio with its conjugate, is given by

$$\left|\frac{\mathrm{CF}(ct)}{\underline{\mathrm{CF}}(ct)}\right|^2=\frac{\exp\left(-c^2t'\Omega t\right)\left(1+(\Im(c\Delta' t))^2\right)}{\exp\left(-c^2t'\underline{\Omega}t\right)\left(1+(\Im(c\underline{\Delta}' t))^2\right)}.$$

Consider the ratio $\frac{1+(\Im(c\Delta' t))^2}{1+(\Im(c\underline{\Delta}' t))^2}$ from the previous expression. Using the asymptotic upper bound (for the numerator) and lower bound (for the denominator), obtained in Lemma C.7 (Statements 2.c. and 2.d.), we get

$$\begin{aligned}\frac{1+(\Im(c\Delta' t))^2}{1+(\Im(c\underline{\Delta}' t))^2}&=\mathcal{O}\left(c^2\exp\left(c^2(\Delta' t)^2-c^2(\underline{\Delta}' t)^2\right)\right)\\&=\mathcal{O}\left(c^2\exp\left(c^2t'(\Delta\Delta'-\underline{\Delta}\underline{\Delta}')t\right)\right).\end{aligned}$$

Thus,

$$\left|\frac{\mathrm{CF}(ct)}{\underline{\mathrm{CF}}(ct)}\right|^2 = \exp\left(-c^2t'(\Omega-\underline{\Omega})t\right)\mathcal{O}\left(c^2\exp\left(c^2t'(\Delta\Delta'-\underline{\Delta\Delta'})t\right)\right)$$
$$= \mathcal{O}\left(c^2\exp\left(-c^2t'\left((\Omega-\Delta\Delta')-(\underline{\Omega}-\underline{\Delta\Delta'})\right)t\right)\right)$$
$$= \mathcal{O}\left(c^2\exp\left(-c^2\left(t'(\Gamma-\underline{\Gamma})t\right)\right)\right). \tag{C5}$$

Consequently,

$$\lim_{c\to\infty}\left|\frac{\mathrm{CF}(ct)}{\underline{\mathrm{CF}}(ct)}\right|^2 = 0, \quad \text{when} \quad t'(\Gamma-\underline{\Gamma})t>0,$$

and

$$\lim_{c\to\infty}\frac{\mathrm{CF}(ct)}{\underline{\mathrm{CF}}(ct)} = 0, \quad \text{when} \quad t'(\Gamma-\underline{\Gamma})t>0$$

follows.

Statement 1.b. Similar to the derivation of the asymptotic upper bound for the ratio in Equation (C5), we derive the asymptotic lower bound by using Lemma C.7 (Statements 2.c. and 2.d.);

$$\left|\frac{\mathrm{CF}(ct)}{\underline{\mathrm{CF}}(ct)}\right|^2 = \Omega\left(\frac{1}{c^2}\exp\left(c^2\left(t'(\underline{\Gamma}-\Gamma)t\right)\right)\right).$$

Consequently,

$$\lim_{c\to\infty}\left|\frac{\mathrm{CF}(ct)}{\underline{\mathrm{CF}}(ct)}\right|^2 = \infty, \quad \text{when} \quad t'(\underline{\Gamma}-\Gamma)t>0,$$

and

$$\lim_{c\to\infty}\frac{\mathrm{CF}(ct)}{\underline{\mathrm{CF}}(ct)} \in \{-\infty,\infty\}, \quad \text{when} \quad t'(\underline{\Gamma}-\Gamma)t>0$$

follows, provided the limit exists in $\overline{\mathbb{R}}$. Combining the result with Statement 1.a. proves Statement 1.b.

Statement 2 From the definition of MSN MGF (Table 2), we get

$$\frac{\mathrm{MGF}(ct)}{\underline{\mathrm{MGF}}(ct)} = \exp\left(c(\mu'-\underline{\mu}')t-\frac{c^2}{2}t'(\underline{\Omega}-\Omega)t\right)\frac{\Phi(c\Delta't)}{\Phi(c\underline{\Delta}'t).}$$

Consider the ratio $\frac{\Phi(c\Delta't)}{\Phi(c\underline{\Delta}'t)}$ from the previous expression. We apply the asymptotic upper bound (for the numerator) and lower bound (for the denominator), obtained in Lemma C.7 (Statement 1). Because $\Delta't \le 0$, the asymptotic upper bound is applicable.

$$\frac{\Phi(c\Delta't)}{\Phi(c\underline{\Delta}'t)} = \mathcal{O}\left(c\exp\left(-\frac{c^2}{2}\left((\Delta't)^2-(\underline{\Delta}'t)^2\right)\right)\right)$$
$$= \mathcal{O}\left(c\exp\left(-\frac{c^2}{2}t'(\Delta\Delta'-\underline{\Delta\Delta'})t\right)\right).$$

Thus,

$$\frac{\mathrm{MGF}(ct)}{\underline{\mathrm{MGF}}(ct)} = \exp\left(c(\mu'-\underline{\mu}')t-\frac{c^2}{2}t'(\underline{\Omega}-\Omega)t\right)\mathcal{O}\left(c\exp\left(-\frac{c^2}{2}t'(\Delta\Delta'-\underline{\Delta\Delta'})t\right)\right)$$
$$= \mathcal{O}\left(c\exp\left(c(\mu'-\underline{\mu}')t-\frac{c^2}{2}t'\left((\underline{\Omega}-\underline{\Delta\Delta'})-(\Omega-\Delta\Delta')\right)t\right)\right)$$
$$= \mathcal{O}\left(c\exp\left(c(\mu'-\underline{\mu}')t-\frac{c^2}{2}(t'(\underline{\Gamma}-\Gamma)t)\right)\right).$$

Because c^2 term dominates the c term in the exponential above, the asymptotic upper bound goes to 0, irrespective of the relation between μ and $\underline{\mu}$. Consequently,

$$\lim_{c\to\infty} \frac{\text{MGF}(ct)}{\underline{\text{MGF}}(ct)} = 0, \quad \text{when} \quad t'(\underline{\Gamma} - \Gamma)t > 0.$$

□

Lemma C.6 (Using Notation 1 and the alternate parameterization for CFUSN in Table 1)**.** *Consider two K-dimensional skew Normal distributions,* $\text{CFUSN}_{K,M}(\mu, \Omega, \Lambda)$ *and* $\text{CFUSN}_{K,M}(\underline{\mu}, \underline{\Omega}, \underline{\Lambda})$*. Let* $c \in \mathbb{R}$ *and* $t \in \mathbb{R}^K$*.*

1. *Let* CF *and* $\underline{\text{CF}}$ *be the characteristic functions corresponding to the two distributions (refer to Table 2). Then,*

   1.a.

$$\lim_{c\to\infty} \frac{\frac{\text{CF}(ct)}{\underline{\text{CF}}(ct)}}{V(c;\theta,\underline{\theta},t)} = \Xi(\Delta, \underline{\Delta}, t).$$

   1.b. *Using Landau's* $\mathcal{O}(\cdot)$ *notation, defined in Notation (C.2), and for any positive integer N,*

$$\frac{\frac{\text{CF}(ct)}{\underline{\text{CF}}(ct)}}{V(c;\theta,\underline{\theta},t)} - \Xi(\Delta, \underline{\Delta}, t) = \Xi(\Delta, \underline{\Delta}, t)\left(\frac{\prod_{\delta\in\Delta(t)}\left(1 + R_N(c,\delta' t) - \imath\mathcal{O}\left(\frac{c}{\exp\left(\frac{1}{2}c^2(\delta' t)^2\right)}\right)\right)}{\prod_{\underline{\delta}\in\underline{\Delta}(t)}\left(1 + R_N(c,\underline{\delta}' t) - \imath\mathcal{O}\left(\frac{c}{\exp\left(\frac{1}{2}c^2(\underline{\delta}' t)^2\right)}\right)\right)} - 1\right).$$

Proof.

$$\begin{aligned}
\frac{\frac{\text{CF}(ct)}{\underline{\text{CF}}(ct)}}{V(c;\theta,\underline{\theta},t)} &= c^{(|\Delta(t)| - |\underline{\Delta}(t)|)} \frac{\exp\left(-\frac{1}{2}c^2 t'(\Omega - \underline{\Omega})t\right) \prod_{\delta\in\Delta}\left(1 + \imath\mathfrak{I}(c\delta' t)\right)}{\exp\left(-\frac{1}{2}c^2 t'(\Gamma - \underline{\Gamma})t\right) \prod_{\underline{\delta}\in\underline{\Delta}}\left(1 + \imath\mathfrak{I}(c\underline{\delta}' t)\right)} \\
&= \frac{c^{|\Delta(t)|}}{c^{|\underline{\Delta}(t)|}} \frac{\exp\left(\frac{1}{2}c^2 t'\underline{\Delta}\underline{\Delta}' t\right) \prod_{\delta\in\Delta}\left(1 + \imath\mathfrak{I}(c\delta' t)\right)}{\exp\left(\frac{1}{2}c^2 t'\Delta\Delta' t\right) \prod_{\underline{\delta}\in\underline{\Delta}}\left(1 + \imath\mathfrak{I}(c\underline{\delta}' t)\right)} \\
&= \frac{c^{|\Delta(t)|}}{c^{|\underline{\Delta}(t)|}} \frac{\exp\left(\frac{1}{2}c^2 \sum_{\underline{\delta}\in\underline{\Delta}(t)}(\underline{\delta}' t)^2\right) \prod_{\delta\in\Delta(t)}\left(1 + \imath\mathfrak{I}(c\delta' t)\right)}{\exp\left(\frac{1}{2}c^2 \sum_{\delta\in\Delta(t)}(\delta' t)^2\right) \prod_{\underline{\delta}\in\underline{\Delta}(t)}\left(1 + \imath\mathfrak{I}(c\underline{\delta}' t)\right)} \qquad (\mathfrak{I}(x) = 0 \text{ when } x = 0) \\
&= \frac{\prod_{\delta\in\Delta(t)}\left(\frac{c}{\exp\left(\frac{1}{2}c^2(\delta' t)^2\right)} + \imath\frac{c}{\exp\left(\frac{1}{2}c^2(\delta' t)^2\right)}\mathfrak{I}(c\delta' t)\right)}{\prod_{\underline{\delta}\in\underline{\Delta}(t)}\left(\frac{c}{\exp\left(\frac{1}{2}c^2(\underline{\delta}' t)^2\right)} + \imath\frac{c}{\exp\left(\frac{1}{2}c^2(\underline{\delta}' t)^2\right)}\mathfrak{I}(c\underline{\delta}' t)\right)}.
\end{aligned} \tag{C6}$$

Using Lemma C.7 (Statement 2.a.), we get

$$\lim_{c\to\infty}\frac{\frac{\mathrm{CF}(ct)}{\underline{\mathrm{CF}}(ct)}}{V(c;\theta,\underline{\theta},t)}=\frac{\prod_{\delta\in\Delta(t)}\left(\imath\sqrt{\frac{2}{\pi}}\frac{1}{\delta' t}\right)}{\prod_{\underline{\delta}\in\underline{\Delta}(t)}\left(\imath\sqrt{\frac{2}{\pi}}\frac{1}{\underline{\delta}' t}\right)}=\Xi(\Delta,\underline{\Delta},t).$$

This proves Statement (1.a).

Using Equation (C6),

$$\frac{\frac{\mathrm{CF}(ct)}{\underline{\mathrm{CF}}(ct)}}{V(c;\theta,\underline{\theta},t)}-\Xi(\Delta,\underline{\Delta},t)=\frac{\prod_{\delta\in\Delta(t)}\left(\frac{c}{\exp\left(\frac{1}{2}c^2(\delta' t)^2\right)}+\imath\frac{c}{\exp\left(\frac{1}{2}c^2(\delta' t)^2\right)}\mathfrak{I}(c\delta' t)\right)}{\prod_{\underline{\delta}\in\underline{\Delta}(t)}\left(\frac{c}{\exp\left(\frac{1}{2}c^2(\underline{\delta}' t)^2\right)}+\imath\frac{c}{\exp\left(\frac{1}{2}c^2(\underline{\delta}' t)^2\right)}\mathfrak{I}(c\underline{\delta}' t)\right)}-\Xi(\Delta,\underline{\Delta},t)$$

$$=\frac{\prod_{\delta\in\Delta(t)}\left(\frac{c}{\exp\left(\frac{1}{2}c^2(\delta' t)^2\right)}+\imath\sqrt{2/\pi}\frac{1}{\delta' t}\left(1+R_N(c,\delta' t)\right)\right)}{\prod_{\underline{\delta}\in\underline{\Delta}(t)}\left(\frac{c}{\exp\left(\frac{1}{2}c^2(\underline{\delta}' t)^2\right)}+\imath\sqrt{2/\pi}\frac{1}{\underline{\delta}' t}\left(1+R_N\left(c,\underline{\delta}' t\right)\right)\right)}-\Xi(\Delta,\underline{\Delta},t)$$

(Using Lemma (C.7), Statement 2.b.)

$$=\Xi(\Delta,\underline{\Delta},t)\frac{\prod_{\delta\in\Delta(t)}\left(1+R_N(c,\delta' t)-\imath\mathcal{O}\left(\frac{c}{\exp\left(\frac{1}{2}c^2(\delta' t)^2\right)}\right)\right)}{\prod_{\underline{\delta}\in\underline{\Delta}(t)}\left(1+R_N(c,\underline{\delta}' t)-\imath\mathcal{O}\left(\frac{c}{\exp\left(\frac{1}{2}c^2(\underline{\delta}' t)^2\right)}\right)\right)}-\Xi(\Delta,\underline{\Delta},t)$$

$$=\Xi(\Delta,\underline{\Delta},t)\left|\frac{\prod_{\delta\in\Delta(t)}\left(1+R_N(c,\delta' t)-\imath\mathcal{O}\left(\frac{c}{\exp\left(\frac{1}{2}c^2(\delta' t)^2\right)}\right)\right)}{\prod_{\underline{\delta}\in\underline{\Delta}(t)}\left(1+R_N(c,\underline{\delta}' t)-\imath\mathcal{O}\left(\frac{c}{\exp\left(\frac{1}{2}c^2(\underline{\delta}' t)^2\right)}\right)\right)}-1\right|.$$

This proves Statement (1.b.). □

Lemma C.7. *Let Φ be the standard normal cdf and $\mathfrak{I}(x)=\int_0^x\sqrt{2/\pi}\exp(u^2/2)du$. Let x be finite. Then, using Landau's $\mathcal{O}(\cdot)$ and $\Omega(\cdot)$ notation, defined in Notation (C.2), as $c\to\infty$,*

1.

1.a. *For all $x\in\mathbb{R}$,*

$$\Phi(cx)=\Omega\left(\frac{\exp\left(-\frac{1}{2}c^2x^2\right)}{c}\right).$$

1.b. *When $x\le 0$,*

$$\Phi(cx)=\mathcal{O}\left(\exp\left(-\frac{1}{2}c^2x^2\right)\right).$$

2.

2.a. *For all* $x \neq 0$,

$$\lim_{c\to\infty} \frac{c\mathfrak{I}(cx)}{\exp\left(\frac{c^2x^2}{2}\right)} = \sqrt{2/\pi}\,\frac{1}{x}.$$

2.b. *For all* $x \neq 0$,

$$\frac{c\mathfrak{I}(cx)}{\exp\left(\frac{c^2x^2}{2}\right)} = \sqrt{\frac{2}{\pi}}\frac{1}{x}\left[1+\sum_{n=1}^{N}\frac{(2n-1)!!}{c^{2n}x^{2n}} + \mathcal{O}\left(c^{-2(N+1)}\right) + \mathcal{O}\left(\exp\left(-c^2x^2/4\right)\right)\right],$$

where !! *is the standard double factorial notation.*

2.c. *For all* $x \in \mathbb{R}$,

$$1+(\mathfrak{I}(cx))^2 = \mathcal{O}\left(\exp(c^2x^2)\right).$$

2.d. *For all* $x \in \mathbb{R}$,

$$1+(\mathfrak{I}(cx))^2 = \Omega\left(\frac{\exp(c^2x^2)}{c}\right).$$

Proof.

Statement 1: From Feller (1968), $1-\Phi(z) \sim z^{-1}\phi(z)$ as $z \to \infty$, where ϕ is the standard normal density function. Because $\Phi(-z) = 1-\Phi(z)$ for any $z \in \mathbb{R}$, it follows that, for $x < 0$, $\Phi(cx) = 1-\Phi(-cx) \sim c^{-1}\phi(-cx)$ as $c \to \infty$ (treating x as a constant). Thus,

$$\Phi(cx) \sim c^{-1}\exp(-\frac{c^2x^2}{2}) \text{ as } c \to \infty, \text{ for } x<0. \tag{C7}$$

Thus, for $x < 0$, $\Phi(cx) = \mathcal{O}\left(\frac{\exp(-c^2x^2/2)}{c}\right)$ and, consequently, $\Phi(cx) = \mathcal{O}(\exp(-c^2x^2/2))$, which also holds true when $x = 0$. Thus, $\Phi(cx) = \mathcal{O}(\exp(-c^2x^2/2))$ when $x \leq 0$, which proves Statement (1.b.). Moreover, it follows from Equation (C7) that, for $x < 0$, $\Phi(cx) = \Omega\left(\frac{\exp(-c^2x^2/2)}{c}\right)$ and because it is true for $x \geq 0$ as well (because $\Phi(0) = 1/2$ and $\Phi(cx)$ approaches 1 when $x > 0$), $\Phi(cx) = \Omega\left(\frac{\exp(-c^2x^2/2)}{c}\right)$ for all $x \in \mathbb{R}$. This proves Statement (1.a.).

Statement 2: Performing integration by parts on $\mathfrak{I}(cx)$ for $x \neq 0$ gives

$$\begin{aligned}
\mathfrak{I}(x) &= \int_0^x \sqrt{\frac{2}{\pi}}\exp\left(u^2/2\right)du \\
&= \int_0^{x/\sqrt{2}} \sqrt{\frac{2}{\pi}}\exp\left(u^2/2\right)du + \int_{x/\sqrt{2}}^{x} \sqrt{\frac{2}{\pi}}\exp\left(u^2/2\right)du \\
&= \sqrt{\frac{2}{\pi}}\left[\int_0^{x/\sqrt{2}} \exp\left(u^2/2\right)du + \int_{x/\sqrt{2}}^{x}\frac{1}{u}\frac{d}{du}\left(\exp\left(u^2/2\right)\right)du\right] \\
&= \sqrt{\frac{2}{\pi}}\left[\int_0^{x/\sqrt{2}} \exp\left(u^2/2\right)du - \frac{\exp\left(x^2/4\right)}{2^{-1/2}x} + \frac{\exp\left(x^2/2\right)}{x} + \int_{x/\sqrt{2}}^{x}\frac{\exp\left(u^2/2\right)}{u^2}du\right] \\
&= \sqrt{\frac{2}{\pi}}\left[\int_0^{x/\sqrt{2}} \exp\left(u^2/2\right)du - \frac{\exp\left(x^2/4\right)}{2^{-1/2}x} + \frac{\exp\left(x^2/2\right)}{x} + \int_{x/\sqrt{2}}^{x}\frac{1}{u^3}\frac{d}{du}\exp\left(u^2/2\right)du\right]
\end{aligned}$$

$$= \sqrt{\frac{2}{\pi}}\left[\int_0^{x/\sqrt{2}} \exp\left({}^{u^2}\!/_2\right) du - \frac{\exp\left({}^{x^2}\!/_4\right)}{2^{-1/2}x} - \frac{\exp\left({}^{x^2}\!/_4\right)}{2^{-3/2}x^3} + \frac{\exp\left({}^{x^2}\!/_2\right)}{x} + \frac{\exp\left({}^{x^2}\!/_2\right)}{x^3} + \int_{x/\sqrt{2}}^{x} 3\frac{\exp\left({}^{u^2}\!/_2\right)}{u^4} du\right]$$

$$= \sqrt{\frac{2}{\pi}}\left[\frac{\exp\left({}^{x^2}\!/_2\right)}{x} + \sum_{n=1}^{N}\frac{(2n-1)!!\exp\left({}^{x^2}\!/_2\right)}{x^{2n+1}} + (2N+1)!!\int_{x/\sqrt{2}}^{x}\frac{\exp\left({}^{u^2}\!/_2\right)}{u^{2(N+1)}}du - \sum_{n=0}^{N}\frac{(2n-1)!!\exp\left({}^{x^2}\!/_4\right)}{\sqrt{2^{-(2n+1)}}x^{2n+1}} + \int_0^{x/\sqrt{2}}\exp\left({}^{u^2}\!/_2\right)du\right]$$

$$= \sqrt{\frac{2}{\pi}}\frac{\exp\left({}^{x^2}\!/_2\right)}{x}\left[1 + \sum_{n=1}^{N}\frac{(2n-1)!!}{x^{2n}} + (2N+1)!!x\frac{\int_{x/\sqrt{2}}^{x}\frac{\exp\left({}^{u^2}\!/_2\right)}{u^{2(N+1)}}du}{\exp\left({}^{x^2}\!/_2\right)} - \sum_{n=0}^{N}\frac{(2n-1)!!\exp\left(-{}^{x^2}\!/_4\right)}{\sqrt{2^{-(2n+1)}}x^{2n}} + x\frac{\int_0^{x/\sqrt{2}}\exp\left({}^{u^2}\!/_2\right)du}{\exp\left({}^{x^2}\!/_2\right)}\right].$$

Thus,

$$\mathfrak{I}(cx) = \sqrt{\frac{2}{\pi}}\frac{\exp\left({}^{c^2x^2}\!/_2\right)}{cx}\left[1 + \sum_{n=1}^{N}\frac{(2n-1)!!}{c^{2n}x^{2n}} + \overbrace{(2N+1)!!cx\frac{\int_{cx/\sqrt{2}}^{cx}\frac{\exp\left({}^{u^2}\!/_2\right)}{u^{2(N+1)}}du}{\exp\left({}^{c^2x^2}\!/_2\right)}}^{A} \overbrace{- \sum_{n=0}^{N}\frac{(2n-1)!!\exp\left(-{}^{c^2x^2}\!/_4\right)}{\sqrt{2^{-(2n+1)}}c^{2n}x^{2n}}}^{B} + \overbrace{cx\frac{\int_0^{cx/\sqrt{2}}\exp\left({}^{u^2}\!/_2\right)du}{\exp\left({}^{c^2x^2}\!/_2\right)}}^{C}+\right].$$

Notice that term (A) is of order $\mathcal{O}(c^{-2(N+1)})$ because

$$\lim_{c\to\infty}\frac{A}{c^{-2(N+1)}} = \lim_{c\to\infty}(2N+1)!!x\frac{\int_{cx/\sqrt{2}}^{cx}\frac{\exp\left({}^{u^2}\!/_2\right)}{u^{2(N+1)}}du}{\frac{\exp\left({}^{c^2x^2}\!/_2\right)}{c^{2N+3}}}$$

$$= \lim_{c\to\infty}(2N+1)!!x\frac{\frac{d}{dc}\int_{cx/\sqrt{2}}^{cx}\frac{\exp\left({}^{u^2}\!/_2\right)}{u^{2(N+1)}}du}{\frac{d}{dc}\frac{\exp\left({}^{c^2x^2}\!/_2\right)}{c^{2N+3}}} \qquad \text{(applying L'Hôpital's rule)}$$

$$= \lim_{c\to\infty}(2N+1)!!\frac{\frac{\exp\left(c^2x^2/2\right)}{c^{2(N+1)}x^{2(N+1)}}\cdot x - \frac{\exp\left(c^2x^2/4\right)2^{N+1}}{c^{2(N+1)}x^{2(N+1)}}\cdot x/\sqrt{2}}{\frac{\exp\left(c^2x^2/2\right)((cx^2)c^{2N+3}-(2N+3)c^{2(N+1)})}{c^{4N+6}}} \quad \text{(applying Leibniz integral rule)}$$

$$= \lim_{c\to\infty}(2N+1)!!\frac{c^{-2(N+1)}x^{-(2N+1)}\left(1-\exp\left(-c^2x^2/4\right)\sqrt{2^{2N+1}}\right)}{c^{-2(N+1)}(x^2-(2N+3)c^{-2})}$$

$$= \lim_{c\to\infty}(2N+1)!!\frac{1}{x^{2N+1}}\frac{1-\exp\left(-c^2x^2/4\right)\sqrt{2^{2N+1}}}{x^2-(2N+3)c^{-2}} = (2N+1)!!\frac{1}{x^{2N+3}};$$

term (B) is $\mathcal{O}(\exp(-c^2x^2/4))$ and so is term (C) because

$$\lim_{c\to\infty}\frac{C}{\exp\left(-c^2x^2/4\right)} = \lim_{c\to\infty} x\frac{\int_0^{cx/\sqrt{2}}\exp\left(u^2/2\right)du}{\frac{\exp\left(c^2x^2/4\right)}{c}}$$

$$= \lim_{c\to\infty} x\frac{\frac{d}{dc}\int_0^{cx/\sqrt{2}}\exp\left(u^2/2\right)du}{\frac{d}{dc}\frac{\exp\left(c^2x^2/4\right)}{c}} \quad \text{(applying L'Hôpital's rule)}$$

$$= \lim_{c\to\infty} x\frac{\exp\left(c^2x^2/4\right)\cdot\frac{x}{\sqrt{2}}}{\frac{\exp\left(c^2x^2/4\right)\left(\left(cx^2/2\right)c-1\right)}{c^2}} \quad \text{(applying Leibniz integral rule)}$$

$$= \lim_{c\to\infty}\frac{\frac{x^2}{\sqrt{2}}}{\left(x^2/2-1/c^2\right)}$$

$$= \sqrt{2}.$$

Consequently,

$$\frac{c\mathfrak{J}(cx)}{\exp\left(c^2x^2/2\right)} = \sqrt{\frac{2}{\pi}}\frac{1}{x}\left[1+\sum_{n=1}^{N}\frac{(2n-1)!!}{c^{2n}x^{2n}} + \mathcal{O}\left(c^{-2(N+1)}\right) + \mathcal{O}\left(\exp\left(-c^2x^2/4\right)\right)\right],$$

which proves Statement (2.b) and, consequently, Statement (2.a.).

Statement (2.a) implies that $\mathfrak{J}(cx)$ is $\mathcal{O}(\frac{\exp(c^2x^2/2)}{c})$ when $x \neq 0$. Thus, $1+(\mathfrak{J}(cx))^2$ is $\mathcal{O}(1)+\mathcal{O}(\frac{\exp(c^2x^2)}{c^2})$ and, consequently, $\mathcal{O}(\exp(c^2x^2))$ when $x \neq 0$. Notice that the $1+(\mathfrak{J}(cx))^2$ is trivially $\mathcal{O}(\exp(c^2x^2))$ when $x = 0$ as well, which completes the proof of Statement (2.c).

Statement (2.a) also implies that $\mathfrak{J}(cx)$ is $\Omega(\frac{\exp(c^2x^2/2)}{c})$, when $x \neq 0$. Thus, $1+(\mathfrak{J}(cx))^2$ is $\Omega(1)+\Omega(\frac{\exp(c^2x^2)}{c^2})$ and, consequently, $\Omega(\frac{\exp(c^2x^2)}{c^2})$, when $x \neq 0$. Notice that $1+(\mathfrak{J}(cx))^2$ is trivially $\Omega(\frac{\exp(c^2x^2)}{c^2})$ when $x = 0$ as well, which completes the proof of Statement (2.d). □

Lemma C.8 (Using Notation 1 and the alternate parameterization for CFUSN in Table 1). *The* $\text{CFUSN}_{K,M}(\mu,\Omega,\Lambda)$ *family is identifiable in* Γ*; that is,* $\text{CFUSN}_{K,M}(\mu,\Omega,\Lambda) = \text{CFUSN}_{K,M}(\underline{\mu},\underline{\Omega},\underline{\Lambda}) \Rightarrow \Gamma = \underline{\Gamma}$.

Proof. The antecedent is equivalent to $\forall t \in \mathbb{R}^K$

$$\text{CF}(t) = \underline{\text{CF}}(t) \Rightarrow \exp\left\{\imath t'\mu - \frac{1}{2}t'\Omega t\right\}\prod_{\delta\in\Delta}\left(1+\imath\mathfrak{J}(\delta' t)\right) = \exp\left\{\imath t'\underline{\mu} - \frac{1}{2}t'\underline{\Omega} t\right\}\prod_{\underline{\delta}\in\underline{\Delta}}\left(1+\imath\mathfrak{J}(\underline{\delta}' t)\right)$$

$$\Rightarrow \exp\left\{-\frac{1}{2}t'(\Gamma-\underline{\Gamma})t\right\} = \frac{c^{|\Delta(t)|}}{c^{|\underline{\Delta}(t)|}}\exp\{\imath t'(\underline{\mu}-\mu)\}$$

$$\times\frac{\prod_{\underline{\delta}\in\underline{\Delta}(t)}\left(\frac{c}{\exp\left(\frac{1}{2}c^2(\underline{\delta}' t)^2\right)}+\imath\frac{c}{\exp\left(\frac{1}{2}c^2(\underline{\delta}' t)^2\right)}\mathfrak{J}(c\underline{\delta}' t)\right)}{\prod_{\delta\in\Delta(t)}\left(\frac{c}{\exp\left(\frac{1}{2}c^2(\delta' t)^2\right)}+\imath\frac{c}{\exp\left(\frac{1}{2}c^2(\delta' t)^2\right)}\mathfrak{J}(c\delta' t)\right)}$$

(where $\Delta(t)$ is defined in Notation 1).

From Lemma C.7, the fraction with product terms on the RHS would converge to a nonzero complex number in the limit as $t \to \infty$. Looking at the other terms on the RHS, one can see that the RHS can never grow/shrink exponentially. In contrast, the LHS grows/shrinks exponentially, unless $t'(\Gamma-\underline{\Gamma})t = 0$. Because this is true for all $t \in \mathbb{R}^K$, the only way the growth rate of RHS and LHS can be consistent is when $\Gamma = \underline{\Gamma}$. □

APPENDIX D

INEFFICACY OF GAUSSIAN MIXTURES IN MODELING UNIMODAL SKEWED DATA

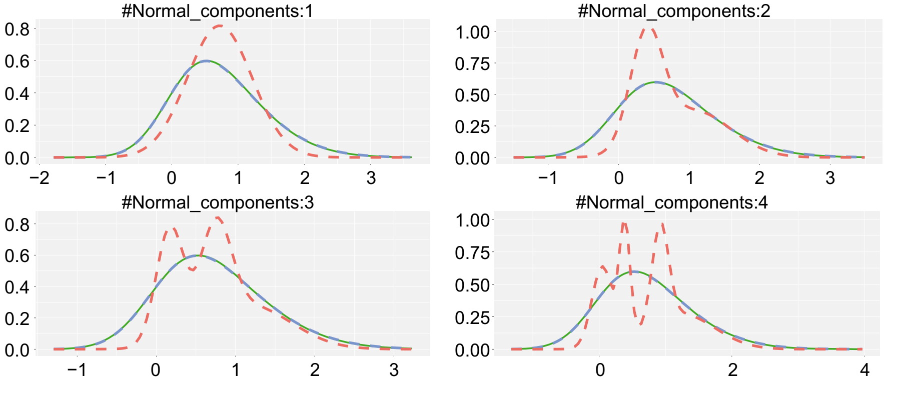

FIGURE D1 Modeling skewed density with Gaussian mixtures: 10000 points were generated from an SN density (green) with $\mu = 1, \omega = 1$, and $\delta = 0.9$. Gaussian mixtures were fitted to the data (red) using the `mclust` package in `R`. The number of Gaussian components was fixed to 1, 2, 3, and 4. An SN curve (blue) was also fitted to the data for comparison. SN = univariate skew normal family [Colour figure can be viewed at wileyonlinelibrary.com]

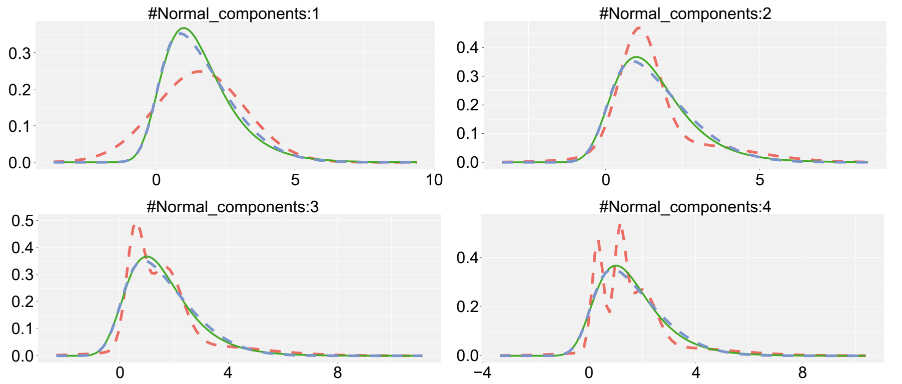

FIGURE D2 Modeling skewed density with Gaussian mixtures: 10000 points were generated from a Gumbel density (green) with location and scale parameters both set to 1. Gaussian mixtures were fitted to the data (red) using the `mclust` package in `R`. The number of Gaussian components was fixed to 1, 2, 3, and 4. An SN curve (blue) was also fitted to the data for comparison. SN = univariate skew normal family [Colour figure can be viewed at wileyonlinelibrary.com]