Geometric Scale Mixtures of Normal Distributions

Deepak Prajapati[1], Sobhan Shafiei[2], Debasis Kundu*[3], and Ahad Jamalizadeh[2]

[1]Decision Sciences Area, Indian Institute of Management Lucknow, Lucknow, 226013, India
[2]Department of Statistics, Faculty of Mathematics and Computer, Shahid Bahonar University of Kerman, Kerman, Iran
[3]Department of Mathematics and Statistics, Indian Institute of Technology Kanpur, Kanpur, 208016, India

Abstract

Recently, Kundu [17] proposed a multivariate skewed distribution, termed the Geometric-Normal (GN) distribution, by compounding the multivariate normal distribution with the geometric distribution. This distribution is a viable alternative to Azzalini's multivariate skew-normal distribution and possesses several desirable properties. This paper introduces a novel class of asymmetric distributions by compounding the geometric distribution with scale mixtures of normal distributions. This class constitutes a special case of the continuous mixtures of multivariate normal distributions introduced by Arellano-Valle and Azzalini [3]. The proposed multivariate distributions exhibit high flexibility, featuring heavy tails, multi-modality, and the ability to model skewness. We have also derived several properties of this class and discussed specific examples to illustrate its applications. The expectation-maximization algorithm was employed to calculate the maximum likelihood estimates of the unknown parameters. Simulation experiments have been performed to show the effectiveness of the proposed algorithm. For illustrative purposes, we have provided one multivariate data set where it has been observed that there exist members in the proposed class of models that can provide better fit compared to skew-normal, skew-t, and generalized hyperbolic distribution. In another example, it was demonstrated that when data generated from a heavy-tailed skew-t distribution is contaminated with noise, the proposed distributions offer a better fit compared to the skew-t distribution.

*Corresponding author; e-mail: kundu@iitk.ac.in

KEYWORDS: Skew-normal distribution; heavy tails; skewness; Birnbum-Saunders distribution; maximum likelihood estimates; outliers detection.

1 INTRODUCTION

The assumption of a normal distribution is commonly used in analyzing real-life data. However, if the data are skewed or exhibit heavy tails, this assumption may not hold. As a result, numerous alternative skewed models have been proposed in the literature. Among various alternative models, the univariate skew-normal (SN) distribution introduced by Azzalini [4] has garnered significant attention. Azzalini and Dalla Valle [7] extended the univariate SN distribution to the multivariate case. The SN distribution is unimodal, with the normal distribution being a special case. Although the SN distribution is highly flexible and can assume various shapes, it cannot model heavy tails, as its tail behavior is identical to that of the normal distribution. Additionally, the Fisher information matrix can become singular when the asymmetry parameter equals zero, making it infeasible to construct confidence intervals based on the Fisher information matrix in such cases.

Since the introduction of the univariate and multivariate SN distributions, a significant amount of work has been dedicated to extending the concept to other related distributions. For instance, Azzalini and Capitanio [5] introduced the skew-t (ST) distribution, which is capable of modeling heavy tails. The issue of singularity in the Fisher information matrix, which occurs for the SN distribution, does not arise for the ST distribution. A comprehensive overview of the fundamental theory and applications of the SN distribution and its generalizations is provided in the book by Azzalini and Capitanio [6].

Kundu [16, 17] introduced a skewed univariate (multivariate) distribution by compounding the normal distribution with the geometric distribution. This distribution, referred to as

the Geometric-Normal (GN) distribution, exhibits several desirable properties. Depending on the parameter associated with the geometric distribution, the GN distribution can exhibit multiple modes when the mean of the normal distribution is non-zero, model heavy tails, and include the normal distribution as a special case. The p-variate GN can be defined as follows. Suppose N is a geometric random variable with parameter $0 < \alpha < 1$, and it has the probability mass function: $\alpha(1-\alpha)^{n-1}$ for $n = 1, 2, \ldots$. From now on it will be denoted by GE(α). Suppose the sequence of random vectors $\{\boldsymbol{X}_i; i = 1, 2, \ldots\}$ are independent and identically distributed (i.i.d.) having p-variate normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix Σ, $\mathrm{N}_p(\boldsymbol{\mu}, \Sigma)$, and they are independent of N, then

$$\boldsymbol{Y} \stackrel{d}{=} \sum_{i=1}^{N} \boldsymbol{X}_i$$

is said to have a GN distribution with parameters α, $\boldsymbol{\mu}$ and Σ, and it will be denoted by $\mathrm{GN}(\alpha, \boldsymbol{\mu}, \Sigma)$. If $\boldsymbol{Y} \sim \mathrm{GN}(\alpha, \boldsymbol{\mu}, \Sigma)$, then the probability density function (PDF) of $\boldsymbol{Y}$ becomes

$$f_{\mathrm{GN}}(\boldsymbol{y}; \alpha, \boldsymbol{\mu}, \Sigma) = \sum_{n=1}^{\infty} \alpha(1-\alpha)^{n-1} \phi_p(\boldsymbol{y}; n\boldsymbol{\mu}, n\Sigma), \tag{1}$$

where $\phi_p(\cdot; \boldsymbol{\mu}, \Sigma)$ denotes the PDF of a multivariate normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix Σ. It may be observed that if $\boldsymbol{Y} \sim \mathrm{GN}(\alpha, \boldsymbol{\mu}, \Sigma)$, then it has the following stochastic representation

$$\boldsymbol{Y} \stackrel{d}{=} N\boldsymbol{\mu} + \sqrt{N}\boldsymbol{V},$$

here N and $\boldsymbol{V} = \Sigma^{\frac{1}{2}}\boldsymbol{Z}$ are independently distributed and $\boldsymbol{Z} \sim N_p(\boldsymbol{0}, \mathbf{I}_p)$. Several properties and classical inference procedures have been developed by Kundu [17].

Recently, Arellano-Valle and Azzalini [3] introduced a generalized mixture of normal (GMN) variables. Consider a p-dimensional random variable $\boldsymbol{V} \sim N_p(\boldsymbol{0}, \Sigma)$ and univariate random variables U and W with joint distribution function $G(u, w)$, such that (V, U, W) are mutually independent. The p-variate random vector $\boldsymbol{Y}$ follows a GMN distribution if

$$\boldsymbol{Y} \stackrel{d}{=} \boldsymbol{\xi} + r(U, W)\boldsymbol{\mu} + s(U, W)\boldsymbol{V}, \tag{2}$$

where vectors $\boldsymbol{\mu}$ and $\boldsymbol{\xi}$ in $\mathbb{R}^p$, and $r(u, w)$ and $s(u, w)$ are a real-value function and a positive-valued function, respectively.

The scale mixtures of normal (SMN) distributions played an important role in the statistical literature and falls within the category of GMN distributions. Also a considerable number of asymmetric distributions proposed in the literature, including SN, ST, and generalized hyperbolic (GH) distributions, are special cases of the GMN class. Although the theory developed in [3] is more general, the specific constructions presented in [3] focused on continuous U and W. The GN distribution can also be derived as a specific instance of the GMN class by setting $r(U, W) = N \sim \text{GE}(\alpha)$ and $s(U, W) = \sqrt{N}$ in (2) and $\boldsymbol{\xi} = \mathbf{0}$. Indeed, the GN model has been developed assuming that the variable U is a discrete random variable and W is a degenerate random variable at one. In this manuscript, to enhance the flexibility of the GN distribution, particularly for modeling heavy-tailed data, a new family within the GMN class is introduced. This family is defined by assuming $r(U, W) = N \sim \text{GE}(\alpha)$ and $s(U, W) = \sqrt{N}W^{-1/2}$ in (2), where W is a positive random variable. It will be called the geometric scale mixtures of normal (GSMN) distributions. The proposed GSMN class can be defined using the simple stochastic representation as in (2). Several special cases can be obtained by using specific mixing distribution. The multivariate GN distribution can also be obtained as a special case for the proposed GSMN family. Several properties have been developed. Special attention has been paid to some of the specific examples. We also derived the maximum likelihood estimates (MLEs) of the unknown parameters using an EM-type algorithm. Simulation experiments were conducted to demonstrate the effectiveness of the proposed EM algorithm. For illustration, we analyzed a dataset and shown that there exists a GSMN distribution that provides a better fit compared to the SN, ST, GH, and contaminated normal (CN) distribution in terms of Bayes Information Criterion (BIC) and Approximate Weight of Evidence (AWE) [9]. Further, we have generated two data sets from multivariate ST distribution one with a moderate tail and the other one with a heavy tail,

and fitted six members of the GSMN class. It observed that when data generated from a heavy-tailed distribution is contaminated with noise, the proposed GSMN class provides a better fit in terms of the BIC compared to ST and CN distributions.

The rest of the paper is organized as follows. In Section 2 we define the GSMN family, establish different properties, and provide various specific examples. In Section 3, the development of the EM algorithm has been indicated. Some simulation results have been presented in Section 4 mainly to show the effectiveness of the EM algorithm, evaluate the ability of introduced distributions to deal with noisy data using simulated data sets, and also provide the analysis of a real data set. Finally, we conclude the paper in Section 5.

2 Geometric scale mixtures of normal distributions

2.1 Preliminaries

Before defining the GSMN distributions, it is helpful to revisit the definition of SMN distributions introduced by Andrews and Mallows [2].

Definition 2.1 *Suppose that the random vector $\boldsymbol{Z}$ follows a p-variate normal distribution with mean vector $\mathbf{0}$ and covariance matrix $\mathbf{I}_p$, and W is a positive-valued random variable characterized by a parameter vector $\boldsymbol{\nu}$, and have the cumulative distribution function $H(\cdot;\boldsymbol{\nu})$. Further, $\boldsymbol{Z}$ and W are independently distributed. Suppose $\boldsymbol{\mu} \in \mathbb{R}^p$, and Σ is a $p \times p$ symmetric positive definite matrix, then a p-variate random vector $\boldsymbol{U}$ is said to be a scale mixture of normal distribution if*

$$\boldsymbol{U} \stackrel{d}{=} \boldsymbol{\mu} + \Sigma^{1/2} W^{-1/2} \boldsymbol{Z}. \tag{3}$$

It will be denoted by $SMN_p(\boldsymbol{\mu}, \Sigma, H_W)$, in which $\boldsymbol{\mu}$ and Σ are location vector and dispersion matrix, respectively.

Definition 2.2 *Suppose* $\boldsymbol{Z}$ *and* W *are same as defined in Definition 2.1 and* $N \sim GE(\alpha)$. *Further,* N, $\boldsymbol{Z}$ *and* W *are independently distributed, then a p-variate random vector* $\boldsymbol{Y}$ *is said to have a GSMN distribution if it has the following stochastic representation*

$$\boldsymbol{Y} \stackrel{d}{=} \boldsymbol{\xi} + N\boldsymbol{\mu} + \Sigma^{1/2}\sqrt{N}W^{-1/2}\boldsymbol{Z}. \tag{4}$$

From now on it will be denoted by $GSMN_p(\alpha, \boldsymbol{\xi}, \boldsymbol{\mu}, \Sigma, H_W)$. *Here* $\boldsymbol{\xi} \in \mathbb{R}^p$ *is the location parameter.*

Generating a random sample from a GSMN distribution is quite simple using (4). In the following remark, we provide another representation for the GSMN class based on the sum of SMN random vectors.

Remark 2.1 *Let* $\{\boldsymbol{X}_i;\ i = 1, 2, \dots\}$ *be a sequence of random vectors, such that*

$$\begin{pmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \\ \vdots \\ \boldsymbol{X}_n \end{pmatrix} \sim SMN_{np}\left(\mathbf{1}_n \otimes \boldsymbol{\mu}, \mathbf{I}_n \otimes \Sigma, H_W\right),$$

where $\mathbf{1}_n = (1, 1, \dots, 1)^\top$ *is a 1-vector of size* n, $\mathbf{I}_n$ *is an identity matrix of order* n, *and* $\otimes$ *denotes the Kronecker product of two matrices. Then if* $N \sim GE(\alpha)$ *and independent of* $\boldsymbol{X}_i$ *for* $i = 1, 2, \dots,$ *we have*

$$\boldsymbol{Y} \stackrel{d}{=} \boldsymbol{\xi} + \sum_{i=1}^{N} \boldsymbol{X}_i \sim GSMN_p(\alpha, \boldsymbol{\xi}, \boldsymbol{\mu}, \Sigma, H_W).$$

The random vector $\boldsymbol{Y}$ will be symmetric if and only if $\boldsymbol{\mu} = \mathbf{0}$. It should be mentioned that here symmetric means central symmetric; see for example Serfling [26]. Note that $\boldsymbol{\mu}$ also plays a role in the multi-modality of the PDF of $\boldsymbol{Y}$. If $\boldsymbol{\mu} = \mathbf{0}$ it is always uni-modal, but if $\boldsymbol{\mu} \neq \mathbf{0}$, it may be multi-modal depending on the values of α and $\boldsymbol{\mu}$. Although we could not exactly characterize it, it is observed that if each component of the vector $\boldsymbol{\mu}$ is significantly away from zero and α is small then the PDF of $\boldsymbol{Y}$ becomes multi-modal.

If $\boldsymbol{Y} \sim \text{GSMN}_p(\alpha, \boldsymbol{\xi}, \boldsymbol{\mu}, \Sigma, H_W)$, then the PDF of $\boldsymbol{Y}$ is given by

$$\begin{aligned} f(\boldsymbol{y}; \alpha, \boldsymbol{\xi}, \boldsymbol{\mu}, \Sigma, H_W) &= \sum_{n=1}^{\infty} \alpha(1-\alpha)^{n-1} \int_0^{\infty} \phi_p(\boldsymbol{y}; \boldsymbol{\xi} + n\boldsymbol{\mu}, nw^{-1}\Sigma) dH(w; \boldsymbol{\nu}) \\ &= \sum_{n=1}^{\infty} \alpha(1-\alpha)^{n-1} f_{\text{SMN}}(\boldsymbol{y}; \boldsymbol{\xi} + n\boldsymbol{\mu}, n\Sigma, H_W), \qquad \boldsymbol{y} \in \mathbb{R}^p, \end{aligned} \tag{5}$$

where $f_{\text{SMN}}(\cdot; \boldsymbol{\lambda}, \Omega, H_W)$ represents the PDF of a scale mixture of normal distribution with scaling variable $W \sim H_W(\cdot; \boldsymbol{\nu})$ and location vector and dispersion matrix $\boldsymbol{\lambda}$ and Ω, respectively. Note that if W is a degenerate random variable at one, then a GSMN distribution becomes a GN distribution. Suppose W is a discrete random variable with finite (countable) support. In that case, a GSMN distribution is a finite (countable) mixture of GN distributions with the same location parameters, but different dispersion parameters. Different specific cases can be obtained by choosing different mixture distributions, either discrete or continuous. Using the representation (4), the characteristic function (CF) of a multivariate GSMN distribution with PDF (5) is defined as follows

$$C_{\boldsymbol{Y}}(\boldsymbol{t}) = \alpha \exp\left\{i\boldsymbol{t}^\top \boldsymbol{\xi}\right\} \sum_{n=1}^{\infty} (1-\alpha)^{n-1} \exp\left\{i\boldsymbol{t}^\top \boldsymbol{\mu} n\right\} E\left[\exp\left\{-\frac{n\boldsymbol{t}^\top \Sigma \boldsymbol{t}}{2} W^{-1}\right\}\right], \ \boldsymbol{t} \in \mathbb{R}^p. \tag{6}$$

2.2 Properties

Let us partition $\boldsymbol{Y}$ into two sub-vectors of sizes q and $p-q$, with corresponding partitions of the parameters into blocks of matching sizes, as follows:

$$\boldsymbol{Y} = \begin{pmatrix} \boldsymbol{Y}_1 \\ \boldsymbol{Y}_2 \end{pmatrix}, \quad \boldsymbol{\xi} = \begin{pmatrix} \boldsymbol{\xi}_1 \\ \boldsymbol{\xi}_2 \end{pmatrix}, \quad \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}. \tag{7}$$

Using the representation (4) and results of Arellano-Valle and Azzalini [3], we present the following results for the GSMN distributions.

Result 2.1 *Let* $\boldsymbol{Y} \sim$ *GSMN*$_p(\alpha, \boldsymbol{\xi}, \boldsymbol{\mu}, \Sigma, H_W)$.

(a) If $E\left[W^{-\frac{1}{2}}\right] < \infty$, *then* $E[\boldsymbol{Y}] = \boldsymbol{\xi} + \frac{\boldsymbol{\mu}}{\alpha}$;

(b) If $E[W^{-1}] < \infty$, *then* $Cov(\boldsymbol{Y}) = (1-\alpha)\alpha^{-2}\boldsymbol{\mu}\boldsymbol{\mu}^\top + \alpha^{-1}E[W^{-1}]\,\Sigma$;

(c) By considering partition (7), *we have*

$$\boldsymbol{Y}_1 \sim GSMN_q(\alpha, \boldsymbol{\xi}_1, \boldsymbol{\mu}_1, \Sigma_{11}, H_W), \quad \textit{and} \quad \boldsymbol{Y}_2 \sim GSMN_{p-q}(\alpha, \boldsymbol{\xi}_2, \boldsymbol{\mu}_2, \Sigma_{22}, H_W);$$

(d) For a q*-dimensional vector* b *and a full-rank matrix* $\boldsymbol{B}$ *of dimension* $q \times p$, *with* $q \leq p$, *we have* $b + \boldsymbol{B}\boldsymbol{Y} \sim GSMN_q(\alpha, b + \boldsymbol{B}\boldsymbol{\xi}, \boldsymbol{B}\boldsymbol{\mu}, \boldsymbol{B}\Sigma\boldsymbol{B}^\top, H_W)$;

(e) If $\boldsymbol{\xi} = \boldsymbol{0}$, *then* $WN^{-1}(\boldsymbol{Y} - N\boldsymbol{\mu})^\top\Sigma^{-1}(\boldsymbol{Y} - N\boldsymbol{\mu}) \sim \chi^2_p$.

Here we provide Mardia's measures of multivariate skewness ($\beta_{1,p}$) and kurtosis ($\beta_{2,p}$). Using equations (30) and (31) of Arellano-Valle and Azzalini [3], we can write the expressions of $\beta_{1,p}$ and $\beta_{2,p}$ as follow:

$$\beta_{1,p} = \theta C_1\left\{3(p-1)(1-\alpha)^2 + C_1^2\left[\theta C_2 + 3a_1\alpha(1-\alpha)\right]^2\right\}, \quad \text{if } a_1 < \infty, \tag{8}$$

$$\begin{aligned}\beta_{2,p} &= \theta^2 C_1^2 C_4 + 2\theta C_1 C_3\left[(p-1) + 3\alpha a_1 C_1\right] \\ &\quad + a_2\alpha(2-\alpha)\left[\frac{(p^2-1)}{a_1^2\alpha} + \frac{2(p-1)C_1}{a_1} + 3\alpha C_1^2\right], \quad \text{if } a_2 < \infty,\end{aligned} \tag{9}$$

where $C_1 = [\alpha a_1 + (1-\alpha)\theta]^{-1}, C_2 = \alpha^2 - 3\alpha + 2, C_3 = \alpha^2 - 4\alpha + 3$, $C_4 = 9 - 18\alpha + 10\alpha^2 - \alpha^3, \theta = \boldsymbol{\mu}^\top\Sigma^{-1}\boldsymbol{\mu}$, and $a_r = E[W^{-r}]$ for $r = 1, 2$

It is observed that $\boldsymbol{\mu}$, Σ, α, and $\boldsymbol{\nu}$ affect the skewness and kurtosis of the random vector $\boldsymbol{Y}$. For fixed α and $\boldsymbol{\nu}$, $\beta_{1,p}$ and $\beta_{2,p}$ are increasing functions of p.

Result 2.2 *A* p*-variate random vector* $\boldsymbol{Y} \sim GSMN_p(\alpha, \boldsymbol{\xi}, \boldsymbol{\mu}, \Sigma, H_W)$ *if and only if for all* $\boldsymbol{c} \in \mathbb{R}^p$, $\boldsymbol{c} \neq \boldsymbol{0}$, $\boldsymbol{c}^\top\boldsymbol{Y} \sim GSMN_1(\alpha, \boldsymbol{c}^\top\boldsymbol{\xi}, \boldsymbol{c}^\top\boldsymbol{\mu}, \boldsymbol{c}^\top\Sigma\boldsymbol{c}, H_W)$.

Proof: *'If' part follows from item (d) of Result 2.2. For 'Only if' part using the CF of* $\boldsymbol{c}^\top \boldsymbol{Y}$ *at point* $t = 1$*, we have*

$$C_{\boldsymbol{c}^\top \boldsymbol{Y}}(1) = \alpha \exp\left\{i\boldsymbol{c}^\top \boldsymbol{\xi}\right\} \sum_{n=1}^{\infty} (1-\alpha)^{n-1} \exp\left\{i\boldsymbol{c}^\top \boldsymbol{\mu} n\right\} E\left[\exp\left\{-\frac{n\boldsymbol{c}^\top \Sigma \boldsymbol{c}}{2} W^{-1}\right\}\right]. \quad (10)$$

Since equation (10) *holds for every* $\boldsymbol{c} \in \mathbb{R}^p$*, the right-hand side of* (10) *equals the CF of a* p*-variate GSMN distribution with PDF* (5)*, which provides the desired result.*

2.3 Examples of GSMN distributions

Now we will be providing some examples of GSMN distributions. We will be using the following notation: $d(\boldsymbol{y}; \boldsymbol{\lambda}, \Omega) = (\boldsymbol{y} - \boldsymbol{\lambda})^\top \Omega^{-1} (\boldsymbol{y} - \boldsymbol{\lambda})$, for convenience.

Example 1: Geometric multivariate t-distribution (GT): In this example the random variable W in Definition 2.2 follows a Gamma distribution with shape and scale parameters $\nu/2$ and $\nu/2$, respectively. The GT distribution has the following PDF

$$f(\boldsymbol{y}; \alpha, \boldsymbol{\xi}, \boldsymbol{\mu}, \Sigma, \nu) = \sum_{n=1}^{\infty} \alpha(1-\alpha)^{n-1} t_p(\boldsymbol{y}; \boldsymbol{\xi} + n\boldsymbol{\mu}, n\Sigma, \nu), \quad (11)$$

where

$$t_p(\boldsymbol{y}; \boldsymbol{\mu}, \Sigma, \nu) = \frac{\Gamma\left(\frac{p+\nu}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right) \pi^{p/2}} \nu^{-p/2} |\Sigma|^{-1/2} \left(1 + \frac{d(\boldsymbol{y}; \boldsymbol{\mu}, \Sigma)}{\nu}\right)^{-\frac{p+\nu}{2}}$$

denotes the PDF of a p-variate t-distribution with $\nu > 0$ degrees of freedom, and with the location vector and scale matrix $\boldsymbol{\mu}$ and Σ, respectively.

We call (11) as the PDF of a p-variate geometric t-distribution with ν degrees of freedom, and it will be denoted by $\mathrm{GT}(\alpha, \boldsymbol{\mu}, \Sigma, \nu)$. Immediately, GT has a thicker tail than GN, and as $\nu \to \infty$, GT converges to GN. We have provided surface plots of bivariate GT distribution for different parameter values in Figure 1. It can be multi-modal and it can have heavy tails. In this case using the PDF of the generalized inverse Gaussian distribution, it can easily be

concluded that

$$E\left[\exp\left\{-\frac{\chi}{2}W^{-1}\right\}\right] = \frac{K_{\frac{\nu}{2}}\left(\sqrt{\nu\chi}\right)2^{1-\frac{\nu}{2}}}{\Gamma\left(\frac{\nu}{2}\right)\left(\sqrt{\nu\chi}\right)^{-\frac{\nu}{2}}}, \quad \chi > 0, \tag{12}$$

where $K_m(\cdot)$ is the modified Bessel function of the third kind and order m. Hence, the CF of the GT distribution can be derived using (6). For $\nu > 2$, $E[W^{-1}] = \dfrac{\nu}{\nu-2}$ and $E[W^{-2}] = \dfrac{\nu^2}{(\nu-2)(\nu-4)}$, for $\nu > 4$. Thus, $\beta_{1,p}$ and $\beta_{2,p}$ as defined in (8) and (9) can be obtained for $\nu > 2$ and $\nu > 4$, respectively.

Example 2: Geometric multivariate Laplace distribution (GL): The GL distribution can be obtained by assuming the random variable W in Definition 2.2 follows an inverse Weibull distribution with shape and scale parameters 1. The PDF of GL distribution is given by

$$f(\boldsymbol{y};\alpha,\boldsymbol{\xi},\boldsymbol{\mu},\Sigma) = \sum_{n=1}^{\infty}\alpha(1-\alpha)^{n-1}g_p(\boldsymbol{y};\boldsymbol{\xi}+n\boldsymbol{\mu},n\Sigma), \tag{13}$$

where

$$g_p(\boldsymbol{y};\boldsymbol{\mu},\Sigma) = \frac{2}{(2\pi)^{p/2}|\Sigma|^{1/2}}\left(\frac{d(\boldsymbol{y};\boldsymbol{\mu},\Sigma)}{2}\right)^{\frac{2-p}{4}}K_{1-p/2}\left(\sqrt{2d(\boldsymbol{y};\boldsymbol{\mu},\Sigma)}\right)$$

denotes the PDF of a p-variate Laplace distribution with the location vector and scale matrix $\boldsymbol{\mu}$ and Σ, respectively [15].

Using (6), the CF of GL distribution with PDF (13) can be expressed as follows

$$C_{\mathrm{GL}}(\boldsymbol{t}) = 2\alpha\exp\left\{i\boldsymbol{t}^{\top}\boldsymbol{\xi}\right\}\sum_{n=1}^{\infty}(1-\alpha)^{n-1}\frac{\exp\left\{i\boldsymbol{t}^{\top}\boldsymbol{\mu}n\right\}}{2+n\boldsymbol{t}^{\top}\Sigma\boldsymbol{t}}, \ \boldsymbol{t}\in\mathbb{R}^p. \tag{14}$$

In this case $E[W^{-1}] = 1$ and $E[W^{-2}] = 2$. Using these $\beta_{1,p}$ and $\beta_{2,p}$ can be obtained for all possible values of parameters.

Example 3: Geometric multivariate Slash distribution (GS): In this case, we have $W \sim Beta(\nu,1)$, where $Beta(\alpha,\beta)$ denotes the beta distribution with parameters α and β. The PDF of geometric p-variate slash distribution can be written as

$$f(\boldsymbol{y};\alpha,\boldsymbol{\xi},\boldsymbol{\mu},\Sigma,\nu) = \sum_{n=1}^{\infty}\alpha(1-\alpha)^{n-1}\psi_p(\boldsymbol{y};\boldsymbol{\xi}+n\boldsymbol{\mu},n\Sigma,\nu), \tag{15}$$

where $\psi_p(\cdot;\boldsymbol{\mu},\Sigma,\nu)$ is the PDF of multivariate Slash distribution as introduced by [28], given as follows

$$\psi_p(\boldsymbol{y};\boldsymbol{\mu},\Sigma,\nu)=\begin{cases}\frac{\nu 2^{(2\nu+p)/2}|\Sigma|^{-1/2}\gamma((2\nu+p)/2;d(\boldsymbol{y};\boldsymbol{\mu},\Sigma)/2)}{(2\pi)^{p/2}d(\boldsymbol{y};\boldsymbol{\mu},\Sigma)^{(2\nu+p)/2}} & \boldsymbol{y}\neq\boldsymbol{\mu},\\ \frac{2\nu}{(2\pi)^{p/2}(2\nu+p)}|\Sigma|^{-1/2} & \boldsymbol{y}=\boldsymbol{\mu},\end{cases}$$

in which $\nu>0$ is the shape parameter, $\boldsymbol{\mu}\in\mathbb{R}^d$ is the location parameter, Σ is the positive definite scatter matrix, and $\gamma(a;z)$ is the incomplete gamma function.

For $\nu>1$, $E[W^{-1}]=\dfrac{\nu}{\nu-1}$ and $E[W^{-2}]=\dfrac{\nu}{\nu-2}$, for $\nu>2$. In this case, $\beta_{1,p}$ and $\beta_{2,p}$ can be obtained for $\nu>1$ and $\nu>2$, respectively.

Example 4: Geometric Normal Birnbaum-Saunders distribution (GNBS): Suppose that W has a Birnbaum-Saunders (BS) distribution with shape parameter $\nu>0$ and scale parameter 1, with the following PDF,

$$f_W(w;\nu)=\frac{1}{2\sqrt{2\pi}\nu}\left[\left(\frac{1}{w}\right)^{1/2}+\left(\frac{1}{w}\right)^{3/2}\right]\exp\left[-\frac{1}{2\nu^2}\left(w+w^{-1}-2\right)\right],\quad w>0,$$

then the PDF of $\boldsymbol{U}$ as in Definition 2.1 becomes

$$\varphi_p(\boldsymbol{y};\boldsymbol{\mu},\Sigma,\nu)=\frac{e^{\frac{1}{\nu^2}}(2\pi)^{-(p+1)/2}|\Sigma|^{-1/2}}{\nu M_\nu(\boldsymbol{y})^{(p+1)/2}}\left[K_{\frac{p-1}{2}}\left(\frac{M_\nu(\boldsymbol{y})}{\nu^2}\right)M_\nu(\boldsymbol{y})+K_{\frac{p+1}{2}}\left(\frac{M_\nu(\boldsymbol{y})}{\nu^2}\right)\right],$$

where $M_\nu(\boldsymbol{y})=\sqrt{1+\nu^2 d(\boldsymbol{y};\boldsymbol{\mu},\Sigma)}$. Therefore, the PDF of random vector $\boldsymbol{Y}$ in Definition 2.2 takes the form

$$f(\boldsymbol{y};\alpha,\boldsymbol{\xi},\boldsymbol{\mu},\Sigma,\nu)=\sum_{n=1}^{\infty}\alpha(1-\alpha)^{n-1}\varphi_p(\boldsymbol{y};\boldsymbol{\xi}+n\boldsymbol{\mu},n\Sigma,\nu).$$

According to the results of Balakrishnan and Kundu [8], if the random variable W follows a BS distribution with shape parameter $\nu>0$ and scale parameter 1, we have

$$E\left[\exp\left\{bW^{-1}\right\}\right]=\frac{1}{2}\exp\left[\frac{1}{\nu^2}\left(1-(1-2b\nu^2)^{1/2}\right)\right]\left(1+(1-2b\nu^2)^{-1/2}\right),\quad b\in\mathbb{R}.\qquad(16)$$

Therefore, the CF of the GNBS distribution can be readily derived by substituting (16) into equation (6). In this case since $E[W^{-1}]=1+\dfrac{1}{2}\nu^2$ and $E[W^{-2}]=\dfrac{3}{2}\nu^4+2\nu^2+1$, $\beta_{1,p}$ and $\beta_{2,p}$ can be defined for all $\nu>0$.

Example 5: Geometric contaminated normal distribution (GCN): Here, the random variable W is a discrete random variable taking one of two states. The probability function of W is given by

$$h(w;\boldsymbol{\nu}) = \nu_1 \mathbb{I}_{(w=1)} + (1-\nu_1)\mathbb{I}_{(w=\nu_2^{-1})}, \qquad 0 < \nu_1 < 1, \nu_2 \geq 1.$$

As a result, the associated PDF of GCN distribution, denoted by $\boldsymbol{Y} \sim \text{GCN}_p(\boldsymbol{\alpha}, \boldsymbol{\mu}, \Sigma, \boldsymbol{\nu})$, is given by

$$\begin{aligned} f(\boldsymbol{y};\alpha,\boldsymbol{\xi},\boldsymbol{\mu},\Sigma,\boldsymbol{\nu}) &= \sum_{n=1}^{\infty} \alpha(1-\alpha)^{n-1}\left[\nu_1\phi_p(\boldsymbol{y};\boldsymbol{\xi}+n\boldsymbol{\mu},n\Sigma) + (1-\nu_1)\phi_p(\boldsymbol{y};\boldsymbol{\xi}+n\boldsymbol{\mu},n\nu_2\Sigma)\right] \\ &= \nu_1 f_{\text{GN}}(\boldsymbol{y}-\boldsymbol{\xi};\alpha,\boldsymbol{\mu},\Sigma) + (1-\nu_1) f_{\text{GN}}(\boldsymbol{y}-\boldsymbol{\xi};\alpha,\boldsymbol{\mu},\nu_2\Sigma), \end{aligned} \tag{17}$$

where $f_{\text{GN}}(\cdot;\alpha,\boldsymbol{\mu},\Sigma)$ denotes the PDF of GN distribution as defined in (1). By considering (6), the CF of the GCN distribution can be expressed as follows:

$$C_{\text{GCN}}(\boldsymbol{t}) = \alpha \exp\left\{i\boldsymbol{t}^\top\boldsymbol{\xi}\right\}\left[\nu_1\frac{C_N(\boldsymbol{t};\boldsymbol{\mu},\Sigma)}{1-(1-\alpha)C_N(\boldsymbol{t};\boldsymbol{\mu},\Sigma)} + (1-\nu_1)\frac{C_N(\boldsymbol{t};\boldsymbol{\mu},\nu_2\Sigma)}{1-(1-\alpha)C_N(\boldsymbol{t};\boldsymbol{\mu},\nu_2\Sigma)}\right],\ \boldsymbol{t}\in\mathbb{R}^p,$$

where $C_N(\boldsymbol{t};\boldsymbol{\lambda},\Omega)$ is the CF of the multivariate normal distribution with mean vector $\boldsymbol{\lambda}$ and covariance matrix Ω.

According to (17), GCN distribution is a two-component GN mixture in which one of the components, with a large prior probability, represents the good observations (reference cluster distribution), and the other, with a small prior probability, the same mean, and an inflated covariance matrix represents the bad observations (or outliers). An advantage of the GCN is that once the parameters are estimated, a generic point $\boldsymbol{y}^*$ can be classified as good or bad via the posterior probability.

3 Parameter estimation via an EM-type algorithm

In this section, we discuss the MLEs of the unknown parameters of a GSMN model. Let $(\boldsymbol{y}_1,\ldots,\boldsymbol{y}_m)$ be an i.i.d. random sample arising from a distribution with PDF as in (5).

Then, the observed data log-likelihood function for the parameter vector Θ is given by

$$\ell\left(\Theta;\boldsymbol{y}\right)=\sum_{i=1}^{m}\log\left(\sum_{n=1}^{\infty}\alpha(1-\alpha)^{n-1}f_{\text{SMN}}(\boldsymbol{y}_i;\boldsymbol{\xi}+n\boldsymbol{\mu},n\Sigma,H_W)\right), \tag{18}$$

where $\Theta=(\alpha,\boldsymbol{\xi}^\top,\boldsymbol{\mu}^\top,\boldsymbol{\sigma}^\top,\boldsymbol{\nu}^\top)^\top$ is the vector of model parameters. Accordingly, vector $\boldsymbol{\sigma}$ gives the elements of the upper triangular part of matrix Σ. Due to the complexity of the observed data log-likelihood function in (18), the MLEs of the model parameters cannot be obtained directly in closed form by maximizing the log-likelihood function. So, we utilize an EM-type algorithm, called the expectation conditional maximization either (ECME) algorithm [19].

Using the convolutional representation of the GSMN distribution in Definition 2.2, a three-level hierarchical representation for the GSMN class can be given as follows:

$$\begin{aligned}\boldsymbol{Y}_i\big|\,(W=w_i,N=n_i) &\sim \text{N}_d\left(\boldsymbol{\xi}+n_i\boldsymbol{\mu},n_i\Sigma/w_i\right),\\ W_i|N_i=n_i &\sim H_W(w_i;\boldsymbol{\nu}),\\ N_i &\sim GE(\alpha),\quad i=1,\ldots,m.\end{aligned} \tag{19}$$

Suppose we have the complete observations of the form $\mathcal{D}_c=\{(\boldsymbol{y}_i,n_i,w_i)\}_{i=1}^m$. Considering the hierarchical representation in (19), the log-likelihood function based on the complete observation becomes (without the additive constant)

$$\begin{aligned}\ell_c(\Theta;\mathcal{D}_c) \;=\; & \text{constant}-\frac{m}{2}\log|\Sigma|+m\log\alpha+\left(\sum_{i=1}^{m}n_i-m\right)\log(1-\alpha)+\sum_{i=1}^{m}\log h_W(w_i;\boldsymbol{\nu})\\ & -\frac{1}{2}\texttt{trace}\left\{\Sigma^{-1}\sum_{i=1}^{m}\left[\frac{w_i}{n_i}(\boldsymbol{y}_i-\boldsymbol{\xi})(\boldsymbol{y}_i-\boldsymbol{\xi})^\top-w_i(\boldsymbol{y}_i-\boldsymbol{\xi})\boldsymbol{\mu}^\top\right.\right.\\ & \left.\left.-w_i\boldsymbol{\mu}(\boldsymbol{y}_i-\boldsymbol{\xi})^\top+w_in_i\boldsymbol{\mu}\boldsymbol{\mu}^\top\right]\right\}.\end{aligned} \tag{20}$$

Based on the EM algorithm principle, in the E-step, we have to calculate the conditional expectation of the complete-data log-likelihood function in (20) at the k-th iteration with Θ

evaluated at $\hat{\Theta}^{(k)}$, which is

$$\begin{aligned} Q\left(\Theta;\boldsymbol{x},\hat{\Theta}^{(k)}\right) &= \text{constant} - \frac{m}{2}\log|\Sigma| + m\log\alpha + \left(\sum_{i=1}^{m} E_1^0(\boldsymbol{y}_i;\hat{\Theta}^{(k)}) - m\right)\log(1-\alpha) \\ &\quad -\frac{1}{2}\texttt{trace}\Bigg\{\Sigma^{-1}\sum_{i=1}^{m}\Bigg[E_{-1}^{+1}(\boldsymbol{y}_i;\hat{\Theta}^{(k)})(\boldsymbol{y}_i-\boldsymbol{\xi})(\boldsymbol{y}_i-\boldsymbol{\xi})^\top + E_1^1(\boldsymbol{y}_i;\hat{\Theta}^{(k)})\boldsymbol{\mu}\boldsymbol{\mu}^\top \\ &\quad -E_0^1(\boldsymbol{y}_i;\hat{\Theta}^{(k)})(\boldsymbol{y}_i-\boldsymbol{\xi})\boldsymbol{\mu}^\top - E_0^1(\boldsymbol{y}_i;\hat{\Theta}^{(k)})\boldsymbol{\mu}(\boldsymbol{y}_i-\boldsymbol{\xi})^\top\Bigg]\Bigg\} \\ &\quad +\sum_{i=1}^{m} H_i(\boldsymbol{y}_i;\hat{\Theta}^{(k)}), \end{aligned} \tag{21}$$

with

$$\begin{aligned} E_q^s &\equiv E_q^s(\boldsymbol{y}_i;\hat{\Theta}^{(k)}) = \mathrm{E}\left[N^q W^s \mid \boldsymbol{Y}=\boldsymbol{y}_i;\hat{\Theta}^{(k)}\right] \\ &= \frac{\sum_{n=1}^{\infty}(1-\alpha)^n n^{q-\frac{p}{2}}\int_0^\infty w^{s+\frac{p}{2}}\exp\left\{-\frac{w}{2n}d(\boldsymbol{y}_i;\boldsymbol{\xi}+n\boldsymbol{\mu},\Sigma)\right\}dH_W(w;\boldsymbol{\nu})}{\sum_{t=1}^{\infty}(1-\alpha)^t t^{-\frac{p}{2}}\int_0^\infty w^{\frac{p}{2}}\exp\left\{-\frac{w}{2t}d(\boldsymbol{y}_i;\boldsymbol{\xi}+t\boldsymbol{\mu},\Sigma)\right\}dH_W(w;\boldsymbol{\nu})}, \quad s,q\in\mathbb{Z}, \\ H_i(\boldsymbol{y}_i;\hat{\Theta}^{(k)}) &= \mathrm{E}\left[\log h(W;\boldsymbol{\nu})|\boldsymbol{Y}=\boldsymbol{y}_i;\hat{\Theta}^{(k)}\right]. \end{aligned}$$

The closed-form formulas for calculating conditional expectations of the special cases of the GSMN class used in this study is presented in Appendix A.

The M-step is to maximize $Q\left(\Theta;\boldsymbol{x},\hat{\Theta}^{(k)}\right)$ with respect to Θ. Thus, the iterative procedure of the ECME algorithm can be summarized as follows:

- E-step: Given $\Theta=\hat{\Theta}^{(k)}$, calculate $\hat{E}_q^{s(k)}$ and $\hat{H}_i^{(k)}$, for $i=1,\ldots,m$, and $q,s\in\{-1,0,1\}$;
- CM-step 1: Updating $\hat{\boldsymbol{\xi}}$ by maximization of (21) over $\boldsymbol{\xi}$ gives
$$\hat{\boldsymbol{\xi}}^{(k+1)} = \frac{\sum_{i=1}^{m}\left(\hat{E}_{-1}^{+1(k)}\boldsymbol{y}_i - \hat{E}_0^{1(k)}\hat{\boldsymbol{\mu}}_j^{(k)}\right)}{\sum_{i=1}^{m}\hat{E}_{-1}^{+1(k)}};$$
- CM-step 2: Fix $\boldsymbol{\xi}=\hat{\boldsymbol{\xi}}^{(k+1)}$ and update $\hat{\boldsymbol{\mu}}^{(k)}$ by maximizing (21) over $\boldsymbol{\mu}$, which leads to
$$\hat{\boldsymbol{\mu}}^{(k+1)} = \frac{\sum_{i=1}^{m}\hat{E}_0^{1(k)}\left(\boldsymbol{y}_i - \hat{\boldsymbol{\xi}}^{(k+1)}\right)}{\sum_{i=1}^{m}\hat{E}_1^{1(k)}};$$

- CM-step 3: Fix $\boldsymbol{\xi} = \hat{\boldsymbol{\xi}}^{(k+1)}$ and $\boldsymbol{\mu} = \hat{\boldsymbol{\mu}}^{(k+1)}$, and update $\hat{\Sigma}^{(k)}$ by maximizing (21) over Σ, which gives

$$\hat{\Sigma}^{(k+1)} = \sum_{i=1}^{m} \left[\hat{E}_{-1}^{+1(k)} \hat{\boldsymbol{e}}_i \hat{\boldsymbol{e}}_i^{\top} - \hat{E}_0^{1(k)} \hat{\boldsymbol{e}}_i \hat{\boldsymbol{\mu}}^{(k+1)\top} - \hat{E}_0^{1(k)} \hat{\boldsymbol{\mu}}^{(k+1)} \hat{\boldsymbol{e}}_i^{\top} + \hat{E}_1^{1(k)} \hat{\boldsymbol{\mu}}^{(k+1)} \hat{\boldsymbol{\mu}}^{(k+1)\top} \right] / m,$$

where $\hat{\boldsymbol{e}}_i = \left(\boldsymbol{y}_i - \hat{\boldsymbol{\xi}}^{(k+1)} \right)$;

- CM-step 4: The MLE of α at the $(k+1)$-th iteration, say $\hat{\alpha}^{(k+1)}$, is updated by

$$\hat{\alpha}^{(k+1)} = \frac{m}{\sum_{i=1}^{m} \hat{E}_1^{0(k)}}.$$

- CML-step 5: Updating $\boldsymbol{\nu}$ is dependent on the distribution of the latent variable W. If the conditional expectation $\hat{\mathcal{H}}_i^{(k)}$ does not have a simple closed-form expression, one can estimate $\boldsymbol{\nu}$ by maximizing the restricted actual log-likelihood function as

$$\hat{\boldsymbol{\nu}}^{(k+1)} = \arg\max_{\boldsymbol{\nu}} \sum_{i=1}^{m} \log f(\boldsymbol{y}_i; \hat{\alpha}^{(k+1)}, \hat{\boldsymbol{\xi}}^{(k+1)}, \hat{\Sigma}^{(k+1)}, \hat{\boldsymbol{\mu}}^{(k+1)}, H_W).$$

For the special case GT distribution, the update of $\hat{\nu}^{(k)}$ can also be found numerically by solving the following nonlinear equation

$$\left[\frac{\nu}{2} \log \frac{\nu}{2} - \log \Gamma \left(\frac{\nu}{2} \right) \right] m + (\frac{\nu}{2} - 1) \sum_{i=1}^{m} \mathrm{E} \left[\log W_i \mid \boldsymbol{y}_i; \hat{\alpha}^{(k)} \right] - \frac{\nu}{2} \sum_{i=1}^{m} \hat{E}_0^{1(k)} = 0,$$

where

$$\begin{aligned} \mathrm{E} \left[\log W_i \mid \boldsymbol{y}_i; \hat{\alpha} \right] &= \psi \left(\frac{\nu + d}{2} \right) \\ &\quad - \frac{\sum_{n=1}^{\infty} \frac{(1-\alpha)^n}{n^{d/2}} \log \left(\frac{\nu}{2} + \frac{d(\boldsymbol{y}_i; \boldsymbol{\xi} + n\boldsymbol{\mu}, \Sigma)}{2n} \right) \left(1 + \frac{d(\boldsymbol{y}_i; \boldsymbol{\xi} + n\boldsymbol{\mu}, \Sigma)}{n\nu} \right)^{-\frac{\nu+d}{2}}}{\sum_{t=1}^{\infty} (1-\alpha)^t t^{-d/2} \left(1 + \frac{d(\boldsymbol{y}_i; \boldsymbol{\xi} + t\boldsymbol{\mu}, \Sigma)}{t\nu} \right)^{-\frac{\nu+d}{2}}}. \end{aligned}$$

Also, for the special case GNBS distribution, the update of $\hat{\nu}^{(k)}$ can be made as

$$\hat{\nu}^{(k+1)} = \sqrt{\frac{\sum_{i=1}^{m} \left(\hat{E}_0^{1(k)} + \hat{E}_0^{-1(k)} - 2 \right)}{m}}.$$

In the GCN distribution, there is another source of missing data arising from the fact that we do not know whether an observation is good or bad. To denote this source of missing data, we use $\{b_i\}_{i=1}^m$, so that $b_i = 1$ if observation i is good and $b_i = 0$ if observation i is bad. Therefore, likelihood function based on the complete data $\mathcal{S} = \{(\boldsymbol{y}_i, n_i, b_i)\}_{i=1}^m$ is given by

$$L(\Theta; \mathcal{S}) = \prod_{i=1}^{m} \alpha(1-\alpha)^{n_i-1} \left[\nu_1 \phi_p(\boldsymbol{y}_i; \boldsymbol{\xi} + n_i\boldsymbol{\mu}, n_i\Sigma)\right]^{b_i} \left[(1-\nu_1)\phi_p(\boldsymbol{y}_i; \boldsymbol{\xi} + n_i\boldsymbol{\mu}, \nu_2 n_i\Sigma)\right]^{1-b_i}.$$

By considering, $\boldsymbol{\xi} = \hat{\boldsymbol{\xi}}^{(k+1)}$, $\boldsymbol{\mu} = \hat{\boldsymbol{\mu}}^{(k+1)}$, and $\Sigma = \hat{\Sigma}^{(k+1)}$, the conditional expectation of complete-data log-likelihood function given $\boldsymbol{y}_i$ and $\Theta = \hat{\Theta}^{(k)}$, can be written as

$$\begin{aligned} Q(\boldsymbol{\nu}; \mathcal{S}) \;=\; & \text{constant} + \log\nu_1 \sum_{i=1}^{m} \hat{\mathcal{B}}_i^{(k)} + \log(1-\nu_1)\left[m - \sum_{i=1}^{m} \hat{\mathcal{B}}_i^{(k)}\right] \\ & -\frac{1}{2\nu_2} trace\Bigg(\hat{\Sigma}^{-1(k+1)} \sum_{i=1}^{m} (1-\hat{\mathcal{B}}_i^{(k)}) \Bigg[\hat{E}_{-1}^{0(k)} (\boldsymbol{x}_i - \hat{\boldsymbol{\xi}}^{(k+1)})(\boldsymbol{x}_i - \hat{\boldsymbol{\xi}}^{(k+1)})^\top \\ & -(\boldsymbol{x}_i - \hat{\boldsymbol{\xi}}^{(k+1)})\hat{\boldsymbol{\mu}}^{(k+1)\top} - \hat{\boldsymbol{\mu}}^{(k+1)}(\boldsymbol{x}_i - \hat{\boldsymbol{\xi}}^{(k+1)})^\top \\ & +\hat{E}_1^{0(k)} \hat{\boldsymbol{\mu}}^{(k+1)} \hat{\boldsymbol{\mu}}^{(k+1)\top}\Bigg]\Bigg) - \frac{p\log\nu_2}{2} \sum_{i=1}^{m} (1-\hat{\mathcal{B}}_i^{(k)}), \end{aligned} \tag{22}$$

where

$$\begin{aligned} \hat{\mathcal{B}}_i^{(k)} \;&\equiv\; \mathcal{B}(\boldsymbol{y}_i; \hat{\Theta}^{(k)}) = \mathrm{E}\left[b_i \mid \boldsymbol{Y} = \boldsymbol{y}_i; \hat{\Theta}^{(k)}\right] \\ &=\; \frac{\hat{\nu}_1^{(k)} f_{\mathrm{GN}}(\boldsymbol{y}_i; \hat{\alpha}^{(k)}, \hat{\boldsymbol{\xi}}^{(k)}, \hat{\boldsymbol{\mu}}^{(k)}, \hat{\Sigma}^{(k)})}{f_{\mathrm{GCN}}(\boldsymbol{y}_i; \hat{\alpha}^{(k)}, \hat{\boldsymbol{\xi}}^{(k)}, \hat{\boldsymbol{\mu}}^{(k)}, \hat{\Sigma}^{(k)}, \hat{\boldsymbol{\nu}}^{(k)})}. \end{aligned}$$

Thus, $\hat{\boldsymbol{\nu}}^{(k)}$ can be updated by maximizing (22) over $\boldsymbol{\nu}$, as in the following equation

$$\hat{\nu}_1^{(k+1)} = \frac{\sum_{i=1}^{m} \hat{\mathcal{B}}_i^{(k)}}{m}, \qquad \text{and} \qquad \hat{\nu}_2^{(k+1)} = \max\left\{\nu_2^0, \hat{\nu}_2^{*(k+1)}\right\},$$

with

$$\begin{aligned} \hat{\nu}_2^{*(k+1)} \;=\; & trace\Bigg(\hat{\Sigma}^{-1(k+1)} \sum_{i=1}^{m} (1-\hat{\mathcal{B}}_i^{(k)}) \Bigg[\hat{E}_{-1}^{0(k)} (\boldsymbol{x}_i - \hat{\boldsymbol{\xi}}^{(k+1)})(\boldsymbol{x}_i - \hat{\boldsymbol{\xi}}^{(k+1)})^\top \\ & -(\boldsymbol{x}_i - \hat{\boldsymbol{\xi}}^{(k+1)})\hat{\boldsymbol{\mu}}^{(k+1)\top} - \hat{\boldsymbol{\mu}}^{(k+1)}(\boldsymbol{x}_i - \hat{\boldsymbol{\xi}}^{(k+1)})^\top \\ & +\hat{E}_1^{0(k)} \hat{\boldsymbol{\mu}}^{(k+1)} \hat{\boldsymbol{\mu}}^{(k+1)\top}\Bigg]\Bigg) \Big/ \left(pm - p\sum_{i=1}^{m} \hat{\mathcal{B}}_i^{(k)}\right), \end{aligned}$$

and ν_2^0 is a number close to 1 from the right, we use 1.001 in simulated and real data analyses. After estimating the unknown parameters of the GCN distribution, one can establish whether an observation is good or bad (outlier) via the posterior probabilities $\hat{\mathcal{B}}_i$. More precisely, observation i is considered good, if $\hat{\mathcal{B}}_i > 0.5$; otherwise, it is bad.

The E- and CM-steps are repeatedly iterated until an acceptable convergence criterion is satisfied. For example, the difference in successive log-likelihood values is less than a pre-specified tolerance value.

3.1 Algorithm implementation

EM-based algorithms are deterministic, so the initialization is important [10]. We used the following simple method to obtain suitable initial values for the proposed ECME algorithm. The sample median and the sample covariance are used to initialize the $\hat{\boldsymbol{\xi}}^{(0)}$ and $\hat{\Sigma}^{(0)}$, respectively. We started the algorithm with $\hat{\boldsymbol{\mu}}^{(0)} = \mathbf{0}$ and $\hat{\alpha}^{(0)} = 0.5$. The initial value for $\boldsymbol{\nu}$ depends on the distribution of the variable W. We set $\hat{\nu}^{(0)} = 1$ for the GT, GNBS, and GS distributions in our numerical work, and for GCN distribution, $\hat{\boldsymbol{\nu}}^{(0)}$ is initialized as $(0.99, 1.001)^\top$. As a simple alternative way for choosing the starting point of $\boldsymbol{\nu}$ and α, we can randomly select several values in the range of permissible values of the parameters and run the algorithm ten times for each initial value separately. The best initial value will then correspond to the highest log-likelihood value.

A common criterion to monitor the convergence based on the likelihood-increasing property of the ECME algorithm is the Aitken acceleration criterion [1]. The Aitken acceleration at iteration $(k+1)$ is given by

$$a^{k+1} = \frac{\ell\left(\hat{\boldsymbol{\Theta}}^{(k+2)}; \boldsymbol{y}\right) - \ell\left(\hat{\boldsymbol{\Theta}}^{(k+1)}; \boldsymbol{y}\right)}{\ell\left(\hat{\boldsymbol{\Theta}}^{(k+1)}; \boldsymbol{y}\right) - \ell\left(\hat{\boldsymbol{\Theta}}^{(k)}; \boldsymbol{y}\right)},$$

where $\ell\left(\hat{\boldsymbol{\Theta}}^{(k)};\boldsymbol{y}\right)$ is the log-likelihood value based on the observed data from iteration k. Then, the asymptotic estimate of the log-likelihood at iteration $(k+2)$ is

$$\ell_{\infty}^{(k+2)} = \ell\left(\hat{\boldsymbol{\Theta}}^{(k+1)};\boldsymbol{y}\right) + \frac{1}{1-a^{k+1}}\left[\ell\left(\hat{\boldsymbol{\Theta}}^{(k+2)};\boldsymbol{y}\right) - \ell\left(\hat{\boldsymbol{\Theta}}^{(k+1)};\boldsymbol{y}\right)\right].$$

The ECME algorithm can be considered to have converged when $\ell_{\infty}^{(k+2)} - \ell\left(\hat{\boldsymbol{\Theta}}^{(k+1)};\boldsymbol{y}\right) < \epsilon$, where ϵ is a user-specified tolerance value [18]. In our analysis, the algorithm was terminated when either the maximum number of iterations $k.max = 2000$ is reached or $\ell_{\infty}^{(k+2)} - \ell\left(\hat{\boldsymbol{\Theta}}^{(k+1)};\boldsymbol{y}\right)$ is less than 10^{-5}.

4 Simulation Results and Data Analysis

4.1 Evaluation of the ECME Algorithm for Parameter Estimation

In this section, we have performed some simulation experiments for different p and different sample sizes mainly to see how the proposed ECME algorithm works in practice. For simulation purposes, we have used GT distribution and have implemented the ECME algorithm as described in detail in Section 3, although any other cases also can be implemented along the same line. Further, we have used the first 100 terms in approximating all the infinite series. We have used (a) Case 1: $p = 2$ and (b) Case 2: $p = 3$.

Case 1: In this case, the observations are simulated from a bivariate GT distribution using representation (4) by considering the following parameter values:

$$\nu = 4,\ \boldsymbol{\xi} = (0,0)^{\top},\ \boldsymbol{\mu} = (7,5)^{\top},\ \Sigma = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}.$$

Case 2: Here the observations are simulated from a trivariate GT distribution using repre-

Table 1: Average MLEs and MSEs (within parentheses) of $\hat{\alpha}$, $\hat{\boldsymbol{\xi}}$, $\hat{\boldsymbol{\mu}}$ and $\hat{\Sigma}$ for case 1 in the simulation study 4.1.

α	n	$\hat{\alpha}$	$\hat{\nu}$	$\hat{\boldsymbol{\xi}}$	$\hat{\boldsymbol{\mu}}$	$\hat{\Sigma}$	
0.2	64	0.2015(0.0226)	5.1963(1.0334)	−0.0076(0.4722)	7.0062(0.2421)	0.9787(0.5211)	0.4869(0.3324)
				−0.0048(0.4002)	4.9998(0.1871)		1.0107(0.4122)
	128	0.2015(0.0160)	4.8156(0.8608)	0.0040(0.2879)	6.9990(0.1416)	0.9770(0.2395)	0.4897(0.1821)
				0.0044(0.2596)	4.9981(0.1122)		0.9995(0.2512)
	256	0.2010(0.0112)	4.1768(0.6588)	0.0072(0.1735)	6.9960(0.0794)	0.9957(0.1619)	0.4900(0.1213)
				0.0047(0.1661)	4.9975(0.0667)		0.9941(0.1651)
0.5	64	0.4964(0.0476)	5.1089(1.0198)	0.0042(0.3959)	6.9969(0.2940)	0.9948(0.3006)	0.4950(0.2191)
				0.0061(0.3743)	4.9974(0.2563)		1.0159(0.3227)
	128	0.4994 (0.0338)	4.6121(0.7817)	−0.0084(0.2550)	7.0001(0.1806)	1.0051(0.1980)	0.5029(0.1429)
				−0.0011(0.2395)	4.9998(0.1582)		1.0152(0.2123)
	256	0.4983(0.0227)	4.1178(0.5247)	0.0052(0.1523)	6.9958(0.1057)	0.9898(0.1384)	0.4942(0.0998)
				0.0029(0.1496)	4.9957(0.0985)		0.9939(0.1449)
0.8	64	0.8021(0.0478)	5.0834(1.1770)	0.0111(0.5213)	6.9907(0.4751)	1.0034(0.2683)	0.5005(0.1893)
				0.0182(0.4656)	4.9797(0.4156)		1.0022(0.2792)
	128	0.8018(0.0332)	4.7212(0.7374)	−0.0011(0.3168)	7.0012(0.2812)	1.0121(0.1972)	0.5071(0.1459)
				0.0001(0.3036)	5.0041(0.2680)		1.0092(0.2044)
	256	0.8012(0.0238)	4.2751(0.4823)	0.0056(0.1829)	6.9972(0.1594)	1.0069(0.1306)	0.4996(0.0896)
				0.0031(0.1802)	5.0028(0.1586)		0.9995(0.1341)

sentation (4) by considering the following parameter values:

$$\nu = 4,\ \boldsymbol{\xi} = (0,0,0)^\top,\ \boldsymbol{\mu} = (7,5,5)^\top,\ \Sigma = \begin{bmatrix} 1 & 0.5 & 0.5 \\ 0.5 & 1 & 0.5 \\ 0.5 & 0.5 & 1 \end{bmatrix}.$$

We have considered various α values and different sample sizes in both cases. The average estimates (AEs) and the associated mean squared errors (MSEs) based on 1000 replications have been reported in Tables 1 and 2 for cases 1 and 2, respectively. Some of the points are quite clear from this simulation. The proposed ECME algorithm is working in this case for different p values. In all the cases the EM algorithm stops on average in less than 30 steps. In all the cases considered here, it is observed as the sample size increases the biases and the MSEs decrease for all the parameters. From the simulation experiment, it is clear that GT distribution can be used in practice quite conveniently.

Table 2: Average MLEs and MSEs (within parentheses) of $\hat{\alpha}$, $\hat{\boldsymbol{\xi}}$, $\hat{\boldsymbol{\mu}}$ and $\hat{\Sigma}$ for case 2 in the simulation study 4.1.

α	n	$\hat{\alpha}$	$\hat{\nu}$	$\hat{\boldsymbol{\xi}}$	$\hat{\boldsymbol{\mu}}$	$\hat{\Sigma}$		
				0.0109(0.4250)	6.9992(0.2138)	1.0301(0.4736)	0.5095(0.3560)	0.5076(0.3589)
0.2	64	0.2022(0.0232)	4.9013(1.1104)	0.0084(0.3776)	4.9954(0.1687)		1.0122(0.4237)	0.5049(0.3245)
				0.0087(0.3865)	4.9969(0.1690)			1.0083(0.4135)
				−0.0143(0.2687)	6.9998(0.1275)	0.9944(0.2321)	0.4898(0.1798)	0.4853(0.1897)
	128	0.2015(0.0164)	4.5962(0.7058)	−0.0142(0.2488)	5.0012(0.1004)		0.9922(0.2317)	0.4843(0.1796)
				−0.0070(0.2418)	4.9996(0.1025)			0.9891(0.2549)
				0.0052(0.1675)	6.9982(0.0732)	0.9989(0.1565)	0.4972(0.1182)	0.4971(0.1214)
	256	0.2004(0.0115)	4.2893(0.4509)	0.0023(0.1580)	4.9982(0.0598)		1.0001(0.1630)	0.4969(0.1183)
				0.0044(0.1551)	4.9977(0.0599)			1.0046(0.1612)
				0.0063(0.3938)	6.9872(0.2805)	0.9896(0.2777)	0.4920(0.2051)	0.4845(0.1988)
0.5	64	0.4967(0.0471)	5.0096(1.0072)	0.0114(0.3580)	4.9873(0.2408)		0.9974(0.2910)	0.4880(0.2008)
				0.0208(0.3741)	4.9807(0.2481)			0.9878(0.2873)
				0.0043(0.2412)	6.9977(0.1705)	0.9983(0.1863)	0.4978(0.1384)	0.4976(0.1430)
	128	0.4983(0.0320)	4.4350(0.7463)	0.0053(0.2237)	4.9995(0.1558)		1.0060(0.1983)	0.5031(0.1407)
				−0.0053(0.2272)	5.0060(0.1549)			1.0032(0.1998)
				−0.0016(0.1449)	6.9983(0.1016)	1.0004(0.1359)	0.4979(0.0971)	0.4974(0.0991)
	256	0.4994(0.0227)	4.1985(0.5209)	0.0020(0.1507)	4.9984(0.0964)		0.9983(0.1377)	0.4962(0.1007)
				−0.0033(0.1442)	5.0023(0.0919)			0.9984(0.1393)
				−0.0140(0.5119)	7.0122(0.4653)	1.0003(0.2687)	0.5016(0.1921)	0.4995(0.1905)
0.8	64	0.8018(0.0458)	5.0292(0.9053)	−0.0027(0.4727)	5.0019(0.4226)		1.0008(0.2681)	0.4977(0.1887)
				−0.0084(0.4612)	5.0004(0.4063)			0.9889(0.2634)
				0.0049(0.2853)	6.9940(0.2614)	1.0003(0.1764)	0.4973(0.1265)	0.5016(0.1246)
	128	0.8014(0.0327)	4.3979(0.7509)	0.0050(0.2908)	4.9973(0.2536)		0.9906(0.1735)	0.4935(0.1237)
				0.0134(0.2905)	4.9884(0.2543)			0.9964(0.1788)
				−0.0002(0.1799)	6.9994(0.1606)	1.0011(0.1235)	0.4982(0.0875)	0.5025(0.0916)
	256	0.7995(0.0239)	4.1875(0.3893)	−0.0073(0.1668)	5.0010(0.1482)		0.9968(0.1259)	0.5006(0.0903)
				−0.0028(0.1782)	4.9995(0.1547)			1.0057(0.1276)

4.2 Assessing the outliers detection

In this section, we perform a simulation study to investigate the performance of the GCN model in detecting bad observations. Similar to the one conducted by Tong and Tortora [27], we focus on two-dimensional data sets ($p = 2$) generated from a bivariate normal distribution with mean vector $\boldsymbol{\mu} = (0, 0)^\top$ and covariance matrix

$$\Sigma = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}.$$

To check the performance of the GCN model in outlier detection, for a specified noise level (q), we consider two scenarios:

(I) In this scenario, $q\%$ of the observations of the sample are randomly selected and replaced by $(0, x^*_{i2})$, where x^*_{i2} is a realization of a uniform random variable over the interval $(10, 15)$. These highly atypical points are mild outliers.

(II In the second scenario, $q\%$ of the observations of the sample are randomly selected and replaced by noise points for which each dimension is a realization of a uniform random variable over the interval $(-10, 10)$.

For each scenario, we simulate 500 data sets for two sample sizes of 128 and 512. In each scenario, we record their true positive rates (TPR) and false positive rates (FPR) in outlier detection, where TPR is computed as the number of outliers correctly detected over the total number of outliers, and FPR is computed as the number of points wrongly classified as outliers over the total number of good points. The average and standard deviation (in parentheses) of the TPR and FPR scores over 500 trials are reported in Table 3 for different noise levels.

According to Table 3, the results indicate a good performance of the GCN distribution in detecting outliers in terms of TPR and FPR in both scenarios. As reported in Table 3,

Table 3: The mean of the TPR and FPR values (with standard deviation in parentheses) over 500 trials, for different noise levels.

		Scenario I				Scenario II			
n		1	5	10	20	1	5	10	30
128	TPR	1	1	1	0.9980	1	0.9201	0.9178	0.9260
		(0)	(0)	(0)	(0.0447)	(0)	(0.0749)	(0.0720)	(0.0500)
	FPR	0	0.0002	0.0011	0.0084	0.0003	0.0010	0.0021	0.0129
		(0)	(0.0011)	(0.0032)	(0.0121)	(0.0017)	(0.0030)	(0.0046)	(0.0128)
512	TPR	1	1	1	1	0.9	0.9	0.9161	0.9365
		(0)	(0)	(0)	(0)	(0.1)	(0.0447)	(0.0406)	(0.0466)
	FPR	0.0000	0.0002	0.0017	0.0076	0.0001	0.0006	0.0023	0.0145
		(0.0002)	(0.0007)	(0.0018)	(0.0050)	(0.0004)	(0.0011)	(0.0024)	(0.0070)

for all noise levels and sample sizes, TPR and FPR values are always close to the optimal values 1 and 0, respectively.

4.3 Model comparison when data came from a heavy-tailed skew distribution

In the last simulation study (as suggested by a referee), the performance of the proposed models is examined when data are generated from a skewed and heavy-tailed model. We use the ST distribution for data generation. Specifically, in each replication of 100 trials, 256 observations are simulated from trivariate ST distribution using *rmmix* function in *mixsmsn* package [24] with the following parameters

$$\boldsymbol{\mu} = (0,0,0)^\top, \quad \boldsymbol{\lambda} = (1,-1,2)^\top, \quad \Sigma = \begin{bmatrix} 1 & 0.5 & 0.5 \\ 0.5 & 1 & 0.5 \\ 0.5 & 0.5 & 1 \end{bmatrix}.$$

We have generated samples from a heavy tail with $\nu = 2.05$ and a not-so-heavy tail with $\nu = 10.0$. We consider four different noise levels (q), for investigating the effect of noise level. For each level, $q\%$ of the observations of the sample are randomly selected and replaced by $(0, x^*_{i2})$, where x^*_{i2} is a realization of a uniform random variable over the interval $(10, 15)$.

We have fitted ST, CN, and all six special cases of the GSMN class to see how the eight

different distributions fit the generated data set when the data is contaminated with noise. Table 4 lists the average of BIC values based on 100 trials. Also for choosing the most plausible model, the number of times (freq.) at which each model was selected as the best model supported by BIC is reported.

Table 4: The average of the BIC values and the number of times the distribution was selected as the best model, for all considered models over 100 samples generated in the simulation study 4.3.

		$\nu = 2.05$				$\nu = 10$			
Dist.		1%	5%	10%	30%	1%	5%	10%	30%
ST	BIC	2613.371	2809.568	3017.284	3376.822	2084.471	2326.906	2536.561	2843.902
	freq.	71	4	12	7	1			1
GN	BIC	2746.144	2970.088	3151.996	3495.748	2120.533	2292.312	2371.580	2726.553
	freq.					1	45	21	61
GT	BIC	2627.934	2830.065	3044.533	3244.439	2093.964	2346.945	2602.688	2758.395
	freq.		2		65				36
GL	BIC	2686.010	2894.390	3107.448	3466.938	2146.681	2404.640	2644.309	3004.834
	freq.								
GS	BIC	2622.596	2810.632	3021.764	3350.018	2083.329	2328.042	2384.306	2973.527
	freq.	11	4	4	14			63	
GNBS	BIC	2653.529	2839.239	3049.155	3292.160	2110.528	2374.559	2509.805	2903.309
	freq.			1	3		3	14	2
GCN	BIC	2631.851	2801.285	3014.822	3373.473	2071.479	2281.891	2386.893	2737.602
	freq.		65	61	10	52	18	1	
CN	BIC	2739.698	2894.676	3111.356	3514.784	2078.998	2282.834	2516.968	3008.048
	freq.	18	25	22	1	46	34	1	

According to Table 4, some members of the proposed class consistently achieved the smallest BIC compared to ST and CN distributions. The ST provides a slightly better fit for heavy tail cases when the noise percentage is 1 and in other cases, its performance is poor compared to the GN, GT, GS, and GCN distributions.

4.4 Real Data Example

Here, we consider a dataset of the U.S. electricity companies from years 1955, reported by Nerlove [22]. The data set includes a total of eight variables: cost, output, labor, laborshare, capital, capitalshare, fuel, fuelshare. The sample size is 145. This dataset is widely used as

Table 5: The results of various distributions for the U.S. electricity companies data set.

p		CN	VG	SN	ST	GH	GN	GT	GL	GS	GNBS	GCN
5	BIC	4280.42	4231.65	4336.76	4185.52	4181.37	4199.86	4163.77	4194.26	**4154.20**	4201.89	4191.48
	AWE	4455.91	4439.05	4536.18	4392.92	4396.74	4407.26	4379.14	4426.39	**4369.58**	4417.26	4401.65
8	BIC	3318.40	3632.77	4315.24	4230.93	3463.84	3645.06	3052.48	3100.18	3004.60	3072.00	**2975.12**
	AWE	3685.33	4055.53	4730.03	4653.70	3894.58	4067.83	3483.23	3522.94	3435.34	3502.74	**3413.84**

a textbook example case (Greene [13]) for fitting multivariate distribution, and they have been re-examined by frontier methods; see e.g. Greene [12]. We first illustrate for $p = 5$ with the subset of variables: cost, output, labor, laborshare, capital and then we will consider all the eight variables with $p = 8$. In both cases, we have fitted all six special cases of the GSMN family mentioned in Section 2.3. Further, to compare the proposed models with the other members of the GMN class, we also fitted SN, ST, GH, variance gamma (VG), and CN distributions. We use *ContaminatedMixt* package [23] in *R* for fitting CN distribution. We utilize *smsn.mmix* function in *mixsmsn* package [24] to fit SN and ST distributions, and other distributions are fitted using *mixture* package [11]. The results have been reported in Table 5.

According to Table 5, some members of the proposed class, namely GS and GCN, have lower BIC and AWE in both cases compared to the other members of the GMN class, including SN, ST, VG, and GH distributions. Hence, it gives the practitioner other possible models for data analysis than the most popular skew distributions. Estimated values of the unknown parameters of the best-fitting distribution, namely GS and GCN for $p = 5$ and $p = 8$, respectively, are reported in Table 6.

Another important point we want to mention is about the computational time. Note that to compute MLEs of the proposed distributions, one needs to compute some infinite series. The performance of the MLEs depends on the number of terms considered in the infinite series computation. Table 7 provides the performance of the MLEs in terms of the log-likelihood function value (LLF), the number of iterations required to converge (Iter.),

Table 6: Estimated values of the parameters of the best-fitting distribution on the U.S. electricity companies data set for different values of p.

p	θ	$\boldsymbol{\nu}$	$\boldsymbol{\xi}$	$\boldsymbol{\mu}$	Σ
5	0.4290	1.5028	$\begin{pmatrix} -3.5838 \\ -1077.7894 \\ 2.0071 \\ 0.1279 \\ 171.9208 \end{pmatrix}$	$\begin{pmatrix} 8.0842 \\ 1685.5119 \\ -0.0247 \\ -0.0130 \\ 0.3842 \end{pmatrix}$	$\begin{pmatrix} 14.4282 & 1790.141 & 0.1909 & -0.0248 & -3.3128 \\ & 236862.899 & 24.1213 & -4.4215 & -414.4208 \\ & & 0.0229 & 0.00004 & -0.3736 \\ & & & 0.0008 & -0.0294 \\ & & & & 113.8928 \end{pmatrix}$
8	0.3432	$\begin{pmatrix} 0.9793 \\ 124942.5 \end{pmatrix}$	$\begin{pmatrix} -10.1959 \\ -1343.7653 \\ 1.8693 \\ 0.1237 \\ 172.7389 \\ 0.4326 \\ 24.6515 \\ 0.4437 \end{pmatrix}$	$\begin{pmatrix} 15.3856 \\ 2309.0727 \\ 0.0669 \\ -0.0110 \\ 1.0130 \\ -0.0032 \\ 1.0249 \\ 0.0142 \end{pmatrix}$	$\begin{pmatrix} 17.173 & 4094.427 & -0.1286 & -0.0767 & -12.4183 & 0.0321 & -9.5452 & 0.0445 \\ & 1171086.894 & -50.8140 & -19.1026 & -2115.4826 & 19.3515 & -3411.8807 & -0.2454 \\ & & 0.0427 & 0.0011 & -0.5968 & -0.0028 & 0.4743 & 0.0018 \\ & & & 0.0019 & -0.0178 & -0.0001 & 0.0189 & -0.0017 \\ & & & & 258.5941 & -0.0219 & 13.1150 & 0.0398 \\ & & & & & 0.0110 & -0.3258 & -0.0109 \\ & & & & & & 56.9615 & 0.3069 \\ & & & & & & & 0.0126 \end{pmatrix}$

and the computational time for different numbers of terms considered in the infinite series.

Table 7: The LLF values, number of iterations, and computing time of the MLE (in seconds) for different numbers of terms considered in the infinite series computation in analyzing the U.S. electricity companies dataset.

		p=5						p=8					
	dist.	5	10	25	50	75	100	5	10	25	50	75	100
LLF	GN	-2053.29	-2033.99	-2035.68	-2035.34	-2035.34	-2035.34	-1876.71	-1815.40	-1750.36	-1712.47	-1697.25	-1690.65
	GT	-2022.30	-2014.70	-2014.70	-2014.70	-2014.70	-2014.70	-1412.82	-1391.82	-1391.87	-1391.87	-1391.87	-1391.87
	GL	-2065.86	-2055.08	-2031.69	-2032.43	-2032.43	-2032.43	-1708.23	-1687.05	-1631.10	-1434.34	-1418.20	-1418.20
	GS	-2016.35	-2009.91	-2009.93	-2009.93	-2009.93	-2009.93	-1396.11	-1368.42	-1367.93	-1367.93	-1367.93	-1367.93
	GNBS	-2033.14	-2030.24	-2033.82	-2033.73	-2033.73	-2033.73	-1515.85	-1495.89	-1516.35	-1401.63	-1401.63	-1401.63
	GCN	-2019.69	-2026.89	-2025.98	-2025.98	-2025.98	-2025.98	-1369.31	-1346.87	-1350.71	-1350.71	-1350.71	-1350.71
Iter.	GN	137	116	28	28	28	28	96	72	28	155	319	510
	GT	95	205	287	287	287	287	43	55	97	97	97	97
	GL	133	121	147	157	157	157	319	154	165	293	210	210
	GS	86	161	215	215	215	215	34	290	1010	1010	1010	1010
	GNBS	82	79	44	44	44	44	203	242	96	48	48	48
	GCN	75	45	37	37	37	37	187	163	235	235	235	235
Time	GN	1.52	1.61	0.70	1.24	1.70	2.25	1.56	1.14	0.71	6.67	19.74	40.29
	GT	1.13	3.17	7.81	13.43	18.95	24.61	0.51	0.93	2.67	4.69	6.51	8.65
	GL	1.52	1.90	4.40	8.57	12.36	15.74	4.30	2.83	5.75	18.11	18.25	23.94
	GS	3.27	10.25	29.56	56.23	82.73	108.49	1.73	21.92	303.71	586.34	854.11	1130.48
	GNBS	1.13	1.50	1.61	2.86	3.97	5.20	2.78	4.89	3.34	2.87	4.10	5.40
	GCN	1.59	1.36	2.15	4.19	5.66	7.34	4.50	5.50	15.07	25.93	37.24	48.58

According to Table 7, the computational time to compute the MLEs is satisfactory in all cases of the GSMN class except for the GS case. In the GS case, to compute the PDF and conditional expectations in the E-step of the EM algorithm, we need to calculate the incomplete gamma function which is a time-consuming process. This problem exists in all statistical models developed based on the Slash distribution. Further, it is observed that even if one takes the number of terms to be 50, the performance of the MLEs is quite satisfactory. It is worth noting that our code is written using the *map* function and its variants from

the *purrr* package in the statistical software R to reduce the computational time. However, the performance and computational efficiency can be enhanced by rewriting key functions in $C++$. All experiments were conducted on a PC equipped with an Intel Core i7 processor running at 3.60 GHz and 16 GB of RAM.

5 Conclusion

A new class of statistical distributions called GSMN distributions was introduced in this paper. It provides a very flexible class of skewed distributions with heavy tails, and the multivariate geometric SN distribution can be obtained as a special case of this class. We have derived different properties of this class of distributions and provided several examples. An EM-type algorithm has been developed to compute the MLEs of the unknown parameters. Extensive simulation results suggest that the proposed ECME algorithm works quite well. We have presented one real data example where it has been observed that some members of the proposed class provide a better fit than the best-fitted SN, ST, VG, CN, and GH distributions in terms of AWE and BIC. Further, we investigated the performance of GCN distribution in detecting outliers using simulated data sets. Our results showed that our proposed method can satisfactorily distinguish good from bad observations. In another example, we have observed that when the data set has been generated from the ST distribution and contaminated by noise, then some members of the introduced class provide a better fit than ST and CN distributions. It indicates that the proposed GSMN class can be used in place of SN, ST, and mean and/or scale mixtures of normal distribution for certain data analysis purposes. Notably, three of our special cases, i.e. GT, GS, and GNBS, have only one more parameter than the ST distribution. Further, the number of unknown parameters of all considered special cases except GCN is less or equal to GH distribution which is a popular distribution in the literature of asymmetric distributions. According to Table 7,

the estimating process of our models is not too time-consuming except for GS distribution. Similarly to the *mixture* package in R for the GH distribution, computational efficiency can be improved by rewriting key functions in the C programming language. All computations have been performed using the statistical software R, and all source codes used in the current study are available from the corresponding author upon reasonable request.

There are some open avenues for future research development. It would be of interest to develop the current approach to formulate new finite mixture models for clustering and semi-supervised clustering. In the definition of the GSMN class, we assumed the random variable N follows a geometric distribution. Since the geometric distribution is a member of the power series class, the present work can be developed using any distribution of this class such as the Poisson distribution. As a further development, a novel class of multivariate mixture-of-experts models can also be extended to analyze multi-target regression data similarly to Sepahdar et al. [25].

Acknowledgements

The authors would like to thank two unknown reviewers for their many constructive comments which has helped us significantly improve the earlier draft of the manuscript.

Appendix

To compute the $E_q^s(\boldsymbol{y}_i;\Theta) = \mathrm{E}\left[N^q W^s \mid \boldsymbol{Y} = \boldsymbol{y}_i;\Theta\right]$, we need to solve the following integral

$$I_n = \int_0^\infty w^{s+\frac{p}{2}} \exp\left\{-\frac{w}{2n} d(\boldsymbol{y}_i;\boldsymbol{\xi}+n\boldsymbol{\mu},\Sigma)\right\} dH_W(w;\boldsymbol{\nu}), \quad \text{for } s=-1,0,+1.$$

For the considered special cases of the GSMN class in Section 2.3, I_n can be calculated as follows:

- GT
$$I_n = \frac{\Gamma\left(\frac{\nu+p}{2}+s\right)}{\Gamma\left(\frac{\nu}{2}\right)}\left(\frac{\nu}{2}\right)^{-\frac{p}{2}-s}\left[1+\frac{d(\boldsymbol{y}_i;\boldsymbol{\xi}+n\boldsymbol{\mu},n\Sigma)}{\nu}\right]^{-\frac{p+\nu}{2}-s}.$$

- GL
$$I_n = 2K_{s+\frac{p}{2}-1}\left(\sqrt{2d(\boldsymbol{y}_i;\boldsymbol{\xi}+n\boldsymbol{\mu},n\Sigma)}\right)\left(\frac{2}{d(\boldsymbol{y}_i;\boldsymbol{\xi}+n\boldsymbol{\mu},n\Sigma)}\right)^{\frac{s+p/2-1}{2}}.$$

- GS
$$I_n = \begin{cases} \nu\gamma\left(d(\boldsymbol{y}_i;\boldsymbol{\xi}+n\boldsymbol{\mu},n\Sigma)/2;\nu+s+p/2\right)\left(\frac{2}{d(\boldsymbol{y}_i;\boldsymbol{\xi}+n\boldsymbol{\mu},n\Sigma)}\right)^{\nu+s+\frac{p}{2}} & d(\boldsymbol{y}_i;\boldsymbol{\xi}+n\boldsymbol{\mu},n\Sigma)\neq 0, \\ \frac{\nu}{\nu+s+p/2} & d(\boldsymbol{y}_i;\boldsymbol{\xi}+n\boldsymbol{\mu},n\Sigma)=0. \end{cases}$$

- GNBS
$$I_n = \frac{e^{\frac{1}{\nu^2}}(2\pi)^{-1/2}}{\nu\left[M_n\right]^{s+(p+1)/2}}\left[K_{s+\frac{p-1}{2}}\left(\frac{M_n}{\nu^2}\right)M_n + K_{s+\frac{p+1}{2}}\left(\frac{M_n}{\nu^2}\right)\right],$$
where $M_n = \sqrt{1+\nu^2 d(\boldsymbol{y}_i;\boldsymbol{\xi}+n\boldsymbol{\mu},n\Sigma)}$.

- GCN
$$I_n = \nu_1 \exp\left\{-\frac{d(\boldsymbol{y}_i;\boldsymbol{\xi}+n\boldsymbol{\mu},n\Sigma)}{2}\right\} + (1-\nu_1)\nu_2^{-s-p/2}\exp\left\{-\frac{d(\boldsymbol{y}_i;\boldsymbol{\xi}+n\boldsymbol{\mu},n\Sigma)}{2\nu_2}\right\}.$$

References

[1] Aitken, A.C. (1926), Iii.—a series formula for the roots of algebraic and transcendental equations, *Proceedings of the Royal Society of Edinburgh*, vol. 45(1), 14–22.

[2] Andrews, D.F. and Mallows, C.L. (1974), Scale mixtures of normal distributions, *Journal of the Royal Statistical Society, Series B*, vol. 36, 99–102.

[3] Arellano-Valle, R.B. and Azzalini, A. (2021), A formulation for continuous mixtures of multivariate normal distributions, *Journal of Multivariate Analysis*, vol. 185, 104780.

[4] Azzalini, A.A. (1985), A class of distributions which include the normal, *Scandinavian Journal of Statistics*, vol. 12, 171–178.

[5] Azzalini, A. and Capitanio, A. (2003), Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution, *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 65(2), 367-389.

[6] Azzalini, A.A. and Capitanio, A. (2014), *The skew-normal and related families*, Cambridge University Press, Cambridge, United Kingdom.

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

[8] Balakrishnan, N. and Kundu, D. (2019), Birnbaum-Saunders Distribution: A Review of Models, Analysis and Applications (with discussions), *Applied Stochastic Models in Business and Industry*, vol. 35, no. 1, 4-132.

[9] Banfield, J.D. and Raftery, A.E. (1993), Model-based Gaussian and non-Gaussian clustering, *Biometrics*, 803–821.

[10] Biernacki, C., Celeux, G. and Govaert, G. (2003), Choosing starting values for the EM algorithm for getting the highest likelihood in multivariate Gaussian mixture models, *Computational Statistics & Data Analysis*, vol. 41(3-4), 561–575.

[11] Browne, R.P., ElSherbiny, A. and McNicholas, P.D. (2024), mixture: mixture models for clustering and classification, R package version, 2.1.1.

[12] Greene, W.H. (2008a), The Econometric approach to efficiency analysis, Chapter 2 in H.Friedet al. (eds.), *The Measurement of Productive Efficiency and Productivity Change*, Oxford University Press, New York.

[13] Greene, W.H. (2008b), *Econometric Analysis*, 6th edn., Pearson Education, Upper SaddleRiver, NJ.

[14] Joarder, A.H. and Ali, M.M. (1996), On the characteristic function of the multivariate t-distribution, *Pakisthan Journal of Statistics*, vol. 12, 55–62.

[15] Kotz, S., Kozubowski, T. and Podgorski, K. (2012), *The Laplace distribution and generalizations: a revisit with applications to communications, economics, engineering, and finance*. Springer Science & Business Media.

[16] Kundu, D. (2014), Geometric skewed normal distribution, *Sankhya, Ser. B.*, vol. 76-B, Part 2, 167–189.

[17] Kundu, D. (2017), Multivariate geometric skew-normal distribution, *Statistics*, vol. 51, no. 6, 1377–1397.

[18] Lindsay, B.G. (1995), Mixture models: theory, geometry and applications, *NSF-CBMS Regional Conference Series in Probability and Statistics*, vol. 5, i–163.

[19] Liu, C. and Rubin, D.B. (1994), The ECME algorithm: a simple extension of EM and ECM with faster monotone convergence. *Biometrika*, 81(4), 633–648.

[20] Mardia, K.V. (1970), Measures of multivariate skewness and kurtosis with applications, *Biometrika*, vol. 57, 519–530.

[21] Moradia, Z.G., Arashib, M., Arslanc, O., Iranmanesha, A. (2017), A new family of multivariate slash distributions, *Communications in Statistics - Theory and Methods*, vol. 46, 3264–3275.

[22] Nerlove, M. (1963), Returns to Scale in Electricity Supply. In C. Christ (ed.), *Measurement in Economics: Studies in Mathematical Economics and Econometrics in Memory of Yehuda Grunfeld*. Stanford University Press.

[23] Punzo, A., Mazza, A. and McNicholas, P.D. (2023), ContaminatedMixt: Clustering and Classification with the Contaminated Normal. R package version 1.3.8.

[24] Prates, M.O., Lachos, V.H. and Cabral, C.R.B. (2013), mixsmsn: Fitting finite mixture of scale mixture of skew-normal distributions, *Journal of Statistical Software*, vol. 54, 1–20.

[25] Sepahdar, A., Madadi, M., Balakrishnan, N. and Jamalizadeh, A. (2022), Parsimonious mixture-of-experts based on mean mixture of multivariate normal distributions, *Stat*, 11(1), e421.

[26] Serfling, R.J. (2006), Multivariate symmetry and asymmetry, *Encylopedia of Statistical Sciences*, II edn, vol. 8, Eds. Kotz, S., Balakrishnan, N., Read, C.B., and Vidakovic, B., John Wiley & Sons, New York, 5338 - 5345.

[27] Tong, H. and Tortora, C. (2022), Model-based clustering and outlier detection with missing data, *Advances in Data Analysis and Classification*, vol. 16(1), 5–30.

[28] Wang, J. and Genton, M.G. (2006), The multivariate skew-slash distribution, *Journal of Statistical Planning and Inference*, vol. 136, 209–220.

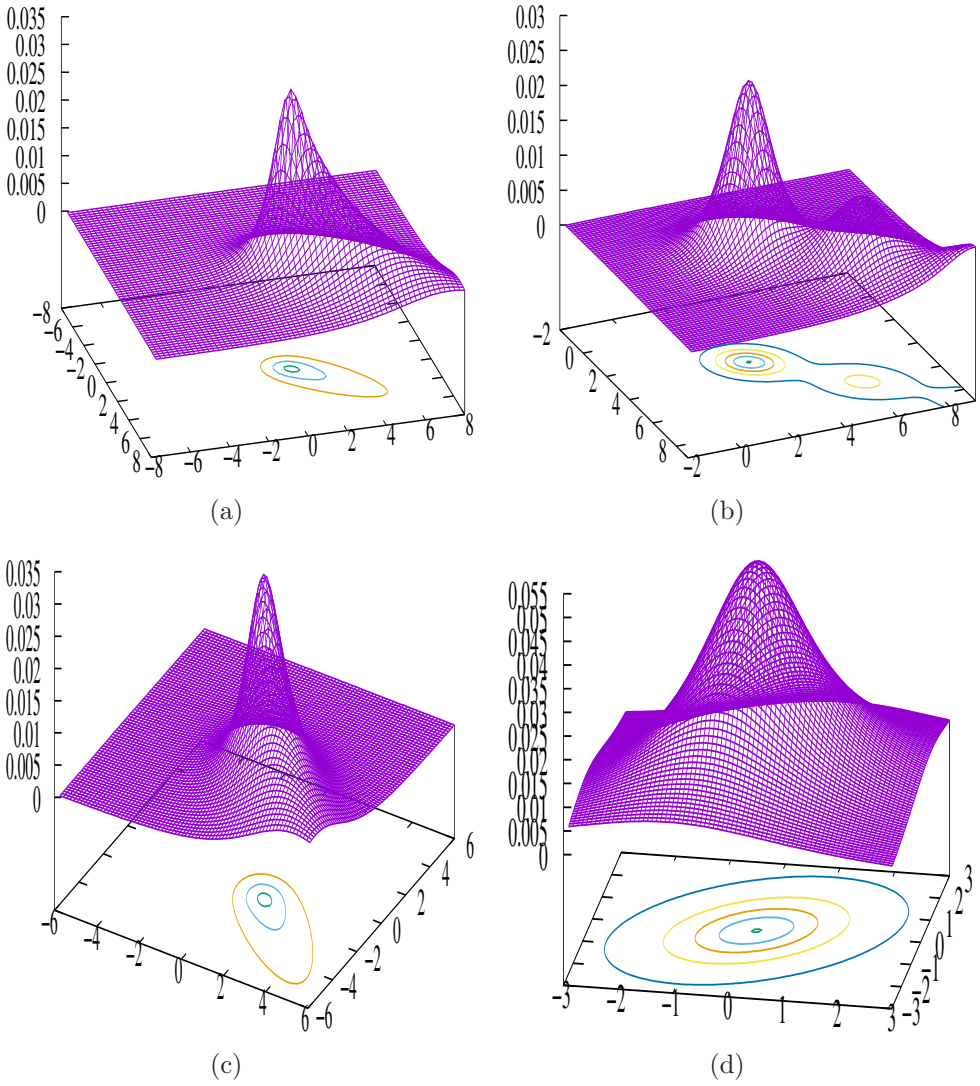

Figure 1: The contour plot of the bivariate GT distribution having the joint PDF of (11) when $\nu = 2$, and (a) $\mu_1 = \mu_2 = 1$, $\sigma_1^2 = \sigma_2^2 = 1.0$, $\sigma_{12} = \sigma_{21} = 0.25$, $\alpha = 0.15$, (b) $\mu_1 = \mu_2 = 3$, $\sigma_1^2 = \sigma_2^2 = 1.0$, $\sigma_{12} = \sigma_{21} = 0.25$, $\alpha = 0.15$, (c) $\mu_1 = 1.0$, $\mu_2 = -1.0$, $\sigma_1^2 = \sigma_2^2 = 1.0$, $\sigma_{12} = \sigma_{21} = 0.25$, $\alpha = 0.15$, (d) $\mu_1 = \mu_2 = 0.0$, $\sigma_1^2 = \sigma_2^2 = 1.0$, $\sigma_{12} = \sigma_{21} = -0.5$, $\alpha = 0.10$,