Multivariate Geometric Skew-Normal Distribution

Debasis Kundu[1]

Abstract

Azzalini [3] introduced a skew-normal distribution of which normal distribution is a special case. Recently Kundu [9] introduced a geometric skew-normal distribution and showed that it has certain advantages over Azzalini's skew-normal distribution. In this paper we discuss about the multivariate geometric skew-normal distribution. It can be used as an alternative to Azzalini's skew normal distribution. We discuss different properties of the proposed distribution. It is observed that the joint probability density function of the multivariate geometric skew normal distribution can take variety of shapes. Several characterization results have been established. Generation from a multivariate geometric skew normal distribution is quite simple, hence the simulation experiments can be performed quite easily. The maximum likelihood estimators of the unknown parameters can be obtained quite conveniently using expectation maximization (EM) algorithm. We perform some simulation experiments and it is observed that the performances of the proposed EM algorithm are quite satisfactory. Further, the analyses of two data sets have been performed, and it is observed that the proposed methods and the model work very well.

Key Words and Phrases: Skew-normal distribution; moment generating function; infinite divisible distribution; maximum likelihood estimators; EM algorithm; Fisher information matrix.

AMS 2000 Subject Classification: Primary 62F10; Secondary: 62H10

[1]Department of Mathematics and Statistics, Indian Institute of Technology Kanpur, Kanpur, Pin 208016, India. e-mail: kundu@iitk.ac.in.

1 Introduction

Azzalini [3] proposed a class of three-parameter skew-normal distributions which includes the normal one. Azzalini's skew normal (ASN) distribution has received a considerable attention in the last two decades due to its flexibility and its applications in different fields. The probability density function (PDF) of ASN takes the following form:

$$f(x;\mu,\sigma,\lambda) = \frac{2}{\sigma}\phi\left(\frac{x-\mu}{\sigma}\right)\Phi\left(\frac{\lambda(x-\mu)}{\sigma}\right), \qquad -\infty < x, \mu, \lambda < \infty, \quad \sigma > 0,$$

where $\phi(x)$ and $\Phi(x)$ denote the standard normal PDF and standard normal cumulative distribution function (CDF), respectively, at the point x. Here μ, σ and λ are known as the location, scale and skewness or tilt parameters, respectively. ASN distribution has an unimodal PDF, and it can be both positively or negatively skewed depending on the skewness parameter. Arnold and Beaver [2] provided an interesting interpretation of this model in terms of hidden truncation. This model has been used quite effectively to analyze skewed data in different fields due to its flexibility.

Later Azzalini and Dalla Valle [5] constructed a multivariate distribution with skew normal marginals. From now on we call it as Azzalini's multivariate skew-normal (AMSN) distribution, and it can be defined as follows. A random vector $\boldsymbol{Z} = (Z_1, \ldots, Z_d)^T$ is a d-dimensional AMSN distribution, if it has the following PDF

$$g(\boldsymbol{z}) = 2\phi_d(\boldsymbol{z};\boldsymbol{\Omega})\Phi(\boldsymbol{\alpha}^T\boldsymbol{z}), \quad \boldsymbol{z} \in \mathbb{R}^d,$$

where $\phi_d(\boldsymbol{z},\boldsymbol{\Omega})$ denotes the PDF of the d-dimensional multivariate normal distribution with standardized marginals, and correlation matrix $\boldsymbol{\Omega}$. We denote such a random vector as $\boldsymbol{Z} \sim \mathrm{SN}_d(\boldsymbol{\Omega},\boldsymbol{\alpha})$. Here the vector $\boldsymbol{\alpha}$ is known as the shape vector, and it can be easily seen that the PDF of AMSN distribution is unimodal and can take different shapes. It has several interesting properties, and it has been used quite successfully to analyze several multivariate data sets in different areas because of its flexibility.

Although ASN distribution is a very flexible distribution, it cannot be used to model moderate or heavy tail data; see for example Azzalini and Capitanio [4]. It is well known to be a thin tail distribution. Since the marginals of AMSN are ASN, multivariate heavy tail data cannot be modeled by using AMSN. Due to this reason several other skewed distributions, often called skew-symmetric distributions, have been suggested in the literature using different kernel functions other than the normal kernel function and using the same technique as Azzalini [3]. Depending on the kernel function the resulting distribution can have moderate or heavy tail behavior. Among different such distributions, skew-t distribution is quite commonly used in practice, which can produce heavy tail distribution depending on the degrees of freedom of the associated t-distribution. It has a multivariate extension also. For a detailed discussions on different skew-symmetric distribution, the readers are referred to the excellent monograph by Azzalini and Capitanio [4].

Although ASN model is a very flexible one dimensional model, and it has several interesting properties, it is well known that computing the maximum likelihood estimators (MLEs) of the unknown parameters of an ASN model is a challenging issue. Azzalini [3] has shown that there is a positive probability that the MLEs of the unknown parameters of a ASN model do not exist. If all the data points have same sign, then the MLEs of unknown parameters of the ASN model may not exist. The problem becomes more severe for AMSN model, and the problem exists for other kernels also.

Recently, the author [9] proposed a new three-parameter skewed distribution, of which normal distribution is a special case. The proposed distribution can be obtained as a geometric sum of independent identically distributed (i.i.d.) normal random variables, and it is called as the geometric skew normal (GSN) distribution. It can be used quite effectively as an alternative to an ASN distribution. It is observed that the GSN distribution is a very flexible distribution, as its PDF can take different shapes depending on the parameter values.

Moreover, the MLEs of the unknown parameters can be computed quite conveniently using the EM algorithm. It can be easily shown that the 'pseudo-log-likelihood' function has a unique maximum, and it can be obtained in explicit forms. Several interesting properties of the GSN distribution have also been developed by Kundu [9].

The main aim of this paper is to consider the multivariate geometric skew-normal (MGSN) distribution, develop its various properties and discuss different inferential issues. Several characterization results and dependence properties have also been established. It is observed that the generation from a MGSN distribution is quite simple, hence simulation experiments can be performed quite conveniently. Note that the d-dimensional MGSN model has $d + 1 + d(d + 1)/2$ unknown parameters. The MLEs of the unknown parameters can be obtained by solving $d + 1 + d(d + 1)/2$ non-linear equations. We propose to use EM algorithm, and it is observed that the 'pseudo-log-likelihood' function has a unique maximum, and it can be obtained in explicit forms. Hence, the implementation of the EM algorithm is quite simple, and the algorithm is very efficient. We perform some simulation experiments to see the performances of the proposed EM algorithm and the performances are quite satisfactory. We also perform the analyses of two data sets to illustrate how the proposed methods can be used in practice. It is observed that the proposed methods and the model work quite satisfactorily.

The main motivation to introduce the MGSN distribution can be stated as follows. Although there are several skewed distributions available in one-dimension, the same is not true in $\mathbb{R}^d$. The proposed MGSN distribution is a very flexible multivariate distribution which can produce variety of shapes. The joint PDF can be unimodal or multimodal and the marginals can have heavy tails depending on the parameters. It has several interesting statistical properties. Computation of the MLEs can be performed in a very simple manner even in high dimension. Hence, if it is known that the data are obtained from a multivariate skewed distribution, the proposed model can be used for analysis purposes. Generating random samples from a MGSN distribution is quite simple, hence any simulation experiment related to this distribution can be performed quite conveniently. Further, it is observed that in one of our data example the MLEs of AMSN do not exist, whereas the MLEs of MGSN distribution exist. Hence, in certain cases the implementation of MGSN distribution becomes easier than the AMSN distribution. The proposed MGSN distribution provides a choice to a practitioner of a new multivariate skewed distribution to analyze multivariate skewed data.

Rest of the paper is organized as follows. In Section 2, first we briefly describe the univariate GSN model, and discuss some of its properties, and then we describe MGSN model. Different properties are discussed in Section 3. In Section 4, we discuss the implementation of the EM algorithm, and some testing of hypotheses problems. Simulation results are presented in Section 5. The analysis of two data sets are presented in Section 6, and finally we conclude the paper in Section 7.

2 GSN and MGSN Distributions

We use the following notations in this paper. A normal random variable with mean μ and variance σ^2 will be denoted by $\mathrm{N}(\mu, \sigma^2)$. A d-variate normal random variable with mean vector $\boldsymbol{\mu}$ and dispersion matrix $\boldsymbol{\Sigma}$ will be denoted by $\mathrm{N}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. The corresponding PDF and CDF at the point $\boldsymbol{x}$ will be denoted by $\phi_d(\boldsymbol{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ and $\Phi_d(\boldsymbol{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$, respectively. A geometric random variable with parameter p will be denoted by $\mathrm{GE}(p)$, and it has the probability mass function (PMF): $p(1-p)^{n-1}$ for $n = 1, 2, \ldots$.

2.1 GSN Distribution

Suppose $N \sim \text{GE}(p)$ and $\{X_i; i = 1, 2, \ldots,\}$ are i.i.d. Gaussian random variables. It is assumed that N and X_i's are independently distributed. Then the random variable

$$X \stackrel{dist}{=} \sum_{i=1}^{N} X_i$$

is known as GSN random variable and its distribution will be denoted by $\text{GSN}(\mu, \sigma, p)$. Here, '$\stackrel{dist}{=}$' means equal in distribution. The GSN distribution can be seen as one of the compound geometric distributions. The PDF of X takes the following form:

$$f_X(x; \mu, \sigma, p) = \sum_{k=1}^{\infty} \frac{p}{\sigma\sqrt{k}} \phi\left(\frac{x - k\mu}{\sigma\sqrt{k}}\right) (1-p)^{k-1}.$$

When $\mu = 0$ and $\sigma = 1$, we say that X has a standard GSN distribution, and it will be denoted by $\text{GSN}(p)$.

The standard GSN is symmetric about 0, and unimodal, but the PDF of $\text{GSN}(\mu, \sigma, p)$ can take different shapes. It can be unimodal or multimodal depending on μ, σ and p values. The hazard function is always an increasing function. If $X \sim \text{GSN}(\mu, \sigma, p)$, then the moment generating function (MGF) of X becomes

$$M_X(t) = \frac{p e^{\mu t + \frac{\sigma^2 t^2}{2}}}{1 - (1-p) e^{\mu t + \frac{\sigma^2 t^2}{2}}}, \quad t \in A_1(\mu, \sigma, p), \tag{1}$$

where

$$\begin{aligned} A_1(\mu, \sigma, p) &= \left\{t; t \in \mathbb{R}, (1-p) e^{\mu t + \frac{\sigma^2 t^2}{2}} < 1\right\} \\ &= \left\{t; t \in \mathbb{R}, 2\mu t + \sigma^2 t^2 + 2\ln(1-p) < 0\right\}. \end{aligned}$$

The corresponding cumulant generating (CGF) function of X is

$$K_X(t) = \ln M_X(t) = \ln p + \mu t + \frac{\sigma^2 t^2}{2} - \ln\left(1 - (1-p) e^{\mu t + \frac{\sigma^2 t^2}{2}}\right). \tag{2}$$

From (2), the mean, variance, skewness and kurtosis can be easily obtained as

$$\begin{aligned} E(X) &= \frac{\mu}{p}, &(3)\\ V(X) &= \frac{\sigma^2 p + \mu^2(1-p)}{p^2}, &(4)\\ \gamma_1 &= \frac{(1-p)\left(\mu^3(2-p) + 3\mu\sigma^2 p\right)}{(p\sigma^2 + \mu^2(1-p))^{3/2}}, \\ \gamma_2 &= \frac{\mu^4(1-p)(p^2-6p+6) - 2\mu^2\sigma^2 p(1-p)(p^2+3p-6) + 3\sigma^4 p^2}{(p\sigma^2 + \mu^2(1-p))^2}, \end{aligned}$$

respectively. It is clear from the expressions of (3) and (4) that as $p \to 0$, $|E(X)|$ and $V(X)$ diverge to ∞. It indicates that GSN model can be used to model heavy tail data. It has been shown that the GSN law is infinitely divisible, and an efficient EM algorithm has been suggested to compute the MLEs of the unknown parameters.

2.2 MGSN Distribution

A d-variate MGSN distribution can be defined as follows. Suppose $N \sim \text{GE}(p)$, $\{\boldsymbol{X}_i; i = 1, 2 \ldots\}$ are i.i.d. $\text{N}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ random vectors and all the random variables are independently distributed. Define

$$\boldsymbol{X} \stackrel{dist}{=} \sum_{i=1}^{N} \boldsymbol{X}_i, \tag{5}$$

then $\boldsymbol{X}$ is said to have a d-variate geometric skew-normal distribution with parameters p, $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, and its distribution will be denoted by $\text{MGSN}_d(p, \boldsymbol{\mu}, \boldsymbol{\Sigma})$. If $\boldsymbol{X} \sim \text{MGSN}_d(p, \boldsymbol{\mu}, \boldsymbol{\Sigma})$, then the CDF and PDF of $\boldsymbol{X}$ become

$$F_{\boldsymbol{X}}(\boldsymbol{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, p) = \sum_{k=1}^{\infty} p(1-p)^{k-1} \Phi_d(\boldsymbol{x}; k\boldsymbol{\mu}, k\boldsymbol{\Sigma})$$

and

$$\begin{aligned} f_{\boldsymbol{X}}(\boldsymbol{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, p) &= \sum_{k=1}^{\infty} p(1-p)^{k-1} \phi_d(\boldsymbol{x}; k\boldsymbol{\mu}, k\boldsymbol{\Sigma}) \\ &= \sum_{k=1}^{\infty} \frac{p(1-p)^{k-1}}{(2\pi)^{d/2} |\boldsymbol{\Sigma}|^{1/2} k^{d/2}} e^{-\frac{1}{2k}(\boldsymbol{x} - k\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{x} - k\boldsymbol{\mu})}, \end{aligned}$$

respectively. Here $\Phi_d(\boldsymbol{x}; k\boldsymbol{\mu}, k\boldsymbol{\Sigma})$ and $\phi_d(\boldsymbol{x}; k\boldsymbol{\mu}, k\boldsymbol{\Sigma})$ denote the CDF and PDF of a d-variate normal distribution, respectively, with the mean vector $k\boldsymbol{\mu}$ and dispersion matrix $k\boldsymbol{\Sigma}$.

If $\boldsymbol{\mu} = \boldsymbol{0}$ and $\boldsymbol{\Sigma} = \boldsymbol{I}$, we say that $\boldsymbol{X}$ is a standard d-variate MGSN random variable, and its distribution will be denoted by $\text{MGSN}_d(p)$. The PDF of $\text{MGSN}_d(p)$ is symmetric and unimodal, for all values of d and p, whereas the PDF of $\text{MGSN}_d(p, \boldsymbol{\mu}, \boldsymbol{\Sigma})$ may not be symmetric, and it can be unimodal or multimodal depending on parameter values. The MGF of MGSN can be obtained in explicit form. If $\boldsymbol{X} \sim \text{MGSN}_d(p, \boldsymbol{\mu}, \boldsymbol{\Sigma})$, then the MGF of $\boldsymbol{X}$ is

$$M_{\boldsymbol{X}}(\boldsymbol{t}) = \frac{p e^{\boldsymbol{\mu}^T \boldsymbol{t} + \frac{1}{2}\boldsymbol{t}^T \boldsymbol{\Sigma} \boldsymbol{t}}}{1 - (1-p) e^{\boldsymbol{\mu}^T \boldsymbol{t} + \frac{1}{2}\boldsymbol{t}^T \boldsymbol{\Sigma} \boldsymbol{t}}}, \qquad \boldsymbol{t} \in A_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, p), \tag{6}$$

where

$$\begin{aligned} A_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, p) &= \left\{\boldsymbol{t}; \boldsymbol{t} \in \mathbb{R}^d, (1-p) e^{\boldsymbol{\mu}^T \boldsymbol{t} + \frac{1}{2}\boldsymbol{t}^T \boldsymbol{\Sigma} \boldsymbol{t}} < 1\right\} \\ &= \left\{\boldsymbol{t}; \boldsymbol{t} \in \mathbb{R}^d, \boldsymbol{\mu}^T \boldsymbol{t} + \frac{1}{2}\boldsymbol{t}^T \boldsymbol{\Sigma} \boldsymbol{t} + \ln(1-p) < 0\right\}. \end{aligned}$$

Further the generation of MGSN distribution is very simple. The following algorithm can be used to generate samples from a MGSN random variable.

Algorithm 1:

- Step 1: Generate n from a GE(p)
- Step 2: Generate $\boldsymbol{X} \sim \text{N}_d(n\boldsymbol{\mu}, n\boldsymbol{\Sigma})$.

In Figure 1 we provide the joint PDF of a bivariate geometric skew normal distribution for different parameter values: (a) $p = 0.75$, $\mu_1 = \mu_2 = 0$, $\sigma_1^2 = \sigma_2^2 = 2$, $\sigma_{12} = \sigma_{21} = 0$, (b) $p = 0.5$, $\mu_1 = \mu_2 = 2.0$, $\sigma_1^2 = \sigma_2^2 = 1$, $\sigma_{12} = \sigma_{21} = -0.5$, (c) $p = 0.15$, $\mu_1 = 2.0$, $\mu_2 = 1.0$, $\sigma_1^2 = \sigma_2^2 = 1$, $\sigma_{12} = \sigma_{21} = -0.5$, (d) $p = 0.15$, $\mu_1 = 0.5$, $\mu_2 =$ -2.5, $\sigma_1^2 = \sigma_2^2 = 1.0$, $\sigma_{12} = \sigma_{21} = 0.5$.

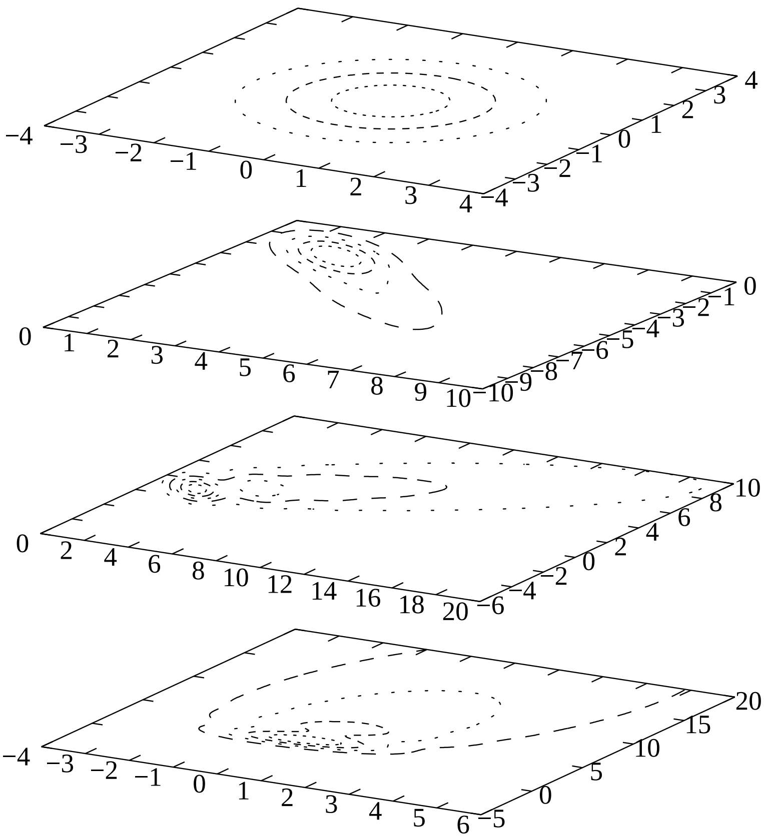

Figure 1: The joint PDF of a bivariate geometric skew normal distribution for different parameter values.

3 Properties

In this section we discuss different properties of a MGSN distribution. We use the following notations:

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

Here the vectors $\boldsymbol{X}$ and $\boldsymbol{\mu}$ are of the order d each, and the matrix $\boldsymbol{\Sigma}_{11}$ is of the order $h \times h$. Rest of the quantities are defined, so that they are compatible. The following result provides the marginals of a MGSN distribution.

Result 1: If $\boldsymbol{X} \sim \text{MGSN}_d(p, \boldsymbol{\mu}, \boldsymbol{\Sigma})$ and $\boldsymbol{X}_1 \sim \text{MGSN}_h(p, \boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11})$ then

$$\boldsymbol{X}_2 \sim MGSN_{d-h}(p, \boldsymbol{\mu}_2, \boldsymbol{\Sigma}_{22}).$$

Proof: The result easily follows from the MGF of MGSN as provided in (6). ∎

We further have the following results similar to the multivariate normal distribution. The result may be used for testing simultaneously a set of linear hypothesis on the parameter vector $\boldsymbol{\mu}$ or it may have some independent interest also; see for example Rao [12].

Theorem 1: If $\boldsymbol{X} \sim \text{MGSN}_d(p, \boldsymbol{\mu}, \boldsymbol{\Sigma})$, then $\boldsymbol{Z} = \boldsymbol{D}\boldsymbol{X} \sim \text{MGSN}_s(p, \boldsymbol{D}\boldsymbol{\mu}, \boldsymbol{D}\boldsymbol{\Sigma}\boldsymbol{D}^T)$, where $\boldsymbol{D}$ is a $s \times d$ matrix of rank $s \leq d$.

Proof: The MGF of the random vector $\boldsymbol{Z}$ is

$$\begin{aligned} M_{\boldsymbol{Z}}(\boldsymbol{t}) &= E\left(e^{\boldsymbol{t}^T\boldsymbol{Z}}\right) = E\left(e^{\boldsymbol{t}^T\boldsymbol{D}\boldsymbol{X}}\right) = E\left(e^{\left(\boldsymbol{D}^T\boldsymbol{t}\right)^T\boldsymbol{X}}\right) \\ &= \frac{pe^{(\boldsymbol{D}\boldsymbol{\mu})^T\boldsymbol{t}+\frac{1}{2}\boldsymbol{t}^T\boldsymbol{D}\boldsymbol{\Sigma}\boldsymbol{D}^T\boldsymbol{t}}}{1-(1-p)e^{(\boldsymbol{D}\boldsymbol{\mu})^T\boldsymbol{t}+\frac{1}{2}\boldsymbol{t}^T\boldsymbol{D}\boldsymbol{\Sigma}\boldsymbol{D}^T\boldsymbol{t}}}, \quad \text{for} \quad \boldsymbol{t} \in A_s^D, \end{aligned}$$

where

$$\begin{aligned} A_s^D &= \left\{\boldsymbol{t}; \boldsymbol{t} \in \mathbb{R}^s, (1-p)e^{(\boldsymbol{D}\boldsymbol{\mu})^T\boldsymbol{t}+\frac{1}{2}\boldsymbol{t}^T\boldsymbol{D}\boldsymbol{\Sigma}\boldsymbol{D}^T\boldsymbol{t}} < 1\right\} \\ &= \left\{\boldsymbol{t}; \boldsymbol{t} \in \mathbb{R}^s, \ln(1-p) + (\boldsymbol{D}\boldsymbol{\mu})^T\boldsymbol{t} + \frac{1}{2}\boldsymbol{t}^T\boldsymbol{D}\boldsymbol{\Sigma}\boldsymbol{D}^T\boldsymbol{t} < 0\right\}. \end{aligned}$$

Hence the result follows. ∎

If $\boldsymbol{X} = (X_1, \ldots, X_d)^T \sim \text{MGSN}_d(p, \boldsymbol{\mu}, \boldsymbol{\Sigma})$, and if we denote $\boldsymbol{\mu}^T = (\mu_1, \ldots, \mu_d)$, $\boldsymbol{\Sigma} = ((\sigma_{ij}))$, then the moments and cumulants of $\boldsymbol{X}$, for $i, j = 1, 2, \ldots, d$, can be obtained from the MGF as follows:

$$E(X_i) = \frac{\partial}{\partial t_i} M_{\boldsymbol{X}}(\boldsymbol{t})\Big|_{\boldsymbol{t}=\boldsymbol{0}} = \frac{\mu_i}{p} \tag{8}$$

$$E(X_iX_j) = \frac{\partial^2}{\partial t_i \partial t_j} M_{\boldsymbol{X}}(\boldsymbol{t})\Big|_{\boldsymbol{t}=\boldsymbol{0}} = \frac{p\sigma_{ij} + \mu_i\mu_j(2-p)}{p^2}.$$

Hence,

$$Var(X_i) = \frac{p\sigma_{ii} + \mu_i^2(1-p)}{p^2}, \tag{9}$$

$$Cov(X_i, X_j) = \frac{p\sigma_{ij} + \mu_i\mu_j(1-p)}{p^2}, \tag{10}$$

and

$$Corr(X_i, X_j) = \frac{p\sigma_{ij} + \mu_i\mu_j(1-p)}{\sqrt{p\sigma_{ii} + \mu_i^2(1-p)}\sqrt{p\sigma_{jj} + \mu_j^2(1-p)}}. \tag{11}$$

It is clear from (11) that the correlation between X_i and X_j for $i \neq j$, not only depends on σ_{ij}, but it also depends on μ_i and μ_j. For fixed p, σ_{ij}, if $\mu_j \to \infty$ and $\mu_i \to \infty$, then $Corr(X_i, X_j) \to 1$, and if $\mu_j \to \infty$ and $\mu_i \to -\infty$, then $Corr(X_i, X_j) \to$ -1. From (11) it also follows that if $\boldsymbol{X}$ is a standard d-variate MGSN random variable, i.e. for $i \neq j$, $\mu_i = \mu_j = \sigma_{ij}$ $= 0$, hence $Corr(X_i, X_j) = 0$. Therefore, in this case although X_i and X_j are uncorrelated, they are not independent.

Now we would like to compute the multivariate skewness indices of the MGSN distribution. Different multivariate skewness measures have been introduced in the literature. Among them the skewness index of Mardia [10, 11] is the most popular one. To define Mardia's multivariate skewness index let us introduce the following notations of a random vector $\boldsymbol{X} = (X_1, \ldots, X_d)$.

$$\mu_{i_1,\ldots,i_s}^{(r_1,\ldots,r_s)} = E\left[\prod_{k=1}^{s}(X_{r_k} - \mu_{r_k})^{i_k}\right],$$

where $\mu_{r_k} = E(X_{r_k}), k = 1, \ldots, s$. Mardia [10] defined the multivariate skewness index as

$$\beta_1 = \sum_{r,s,t=1}^{d} \sum_{r',s',t'=1}^{d} \sigma^{rr'}\sigma^{ss'}\sigma^{tt'}\ \mu_{111}^{rst}\ \mu_{111}^{r's't'},$$

here σ^{jk} for $j, k = 1, \ldots, d$ denotes the (j,k)-th element of the inverse of the dispersion matrix of the random vector $\boldsymbol{X}$. In case of MGSN distribution

$$\mu_{111}^{lhm} = \frac{1}{p^4}\left\{p(1-p)(2-p)\mu_h\mu_l\mu_m + p^2(1-p)(\mu_m\sigma^{hl} + \mu_l\sigma^{hm} + \mu_h\sigma^{lm})\right\}. \tag{12}$$

It is clear from (12) that if $p = 1$ then $\beta_1 = 0$. Also if $\mu_j = 0$ for all $j = 1, \ldots, d$, then $\beta_1 = 0$. Moreover, if $\mu_j \neq 0$ for some $j = 1, \ldots, d$, then the skewness index β_1 may diverge to ∞ or $-\infty$ as $p \to 0$. Therefore, for MGSN distribution Mardia's multivariate skewness index varies from $-\infty$ to ∞.

If $\boldsymbol{X} \sim \text{MGSN}_d(p, \boldsymbol{\mu}, \boldsymbol{\Sigma})$ and if we denote the mean vector and dispersion matrix of $\boldsymbol{X}$, as $\boldsymbol{\mu}_{\boldsymbol{X}}$ and $\boldsymbol{\Sigma}_{\boldsymbol{X}}$, respectively, then from (8), (9) and (10), we have the following relation:

$$p\ \boldsymbol{\mu}_{\boldsymbol{X}} = \boldsymbol{\mu} \quad \text{and} \quad p^2\ \boldsymbol{\Sigma}_{\boldsymbol{X}} = p\ \boldsymbol{\Sigma} + (1-p)\ \boldsymbol{\mu}\boldsymbol{\mu}^T.$$

The following result provides the canonical correlation between $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$. It may be mentioned that canonical correlation is very useful in multivariate data analysis. In an experimental context suppose we take two sets of variables, then the canonical correlation can be used to see what is common among these two sets of variables; see for example Rao [12].

Theorem 2: Suppose $\boldsymbol{X} \sim \text{MGSN}_d(p, \mathbf{0}, \boldsymbol{\Sigma})$. Further $\boldsymbol{X}$ and $\boldsymbol{\Sigma}$ are partitioned as in (7). Then for $\boldsymbol{\alpha} \in \mathbb{R}^h$ and $\boldsymbol{\beta} \in \mathbb{R}^{d-h}$ such that $\boldsymbol{\alpha}^T \boldsymbol{\Sigma}_{11} \boldsymbol{\alpha} = 1$ and $\boldsymbol{\beta}^T \boldsymbol{\Sigma}_{22} \boldsymbol{\beta} = 1$, the maximum $\text{corr}(\boldsymbol{\alpha}^T \boldsymbol{X}_1, \boldsymbol{\beta}^T \boldsymbol{X}_2) = \lambda_1$, where λ_1 is the maximum root of the d-degree polynomial equation

$$\begin{vmatrix} -\lambda \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & -\lambda \Sigma_{22} \end{vmatrix} = 0.$$

Proof: From Theorem 1, we obtain

$$\begin{pmatrix} \boldsymbol{\alpha}^T \boldsymbol{X}_1 \\ \boldsymbol{\beta}^T \boldsymbol{X}_2 \end{pmatrix} \sim \text{MGSN}\left(2, p, \begin{pmatrix} \boldsymbol{\alpha}^T \boldsymbol{\mu}_1 \\ \boldsymbol{\beta}^T \boldsymbol{\mu}_2 \end{pmatrix}, \begin{pmatrix} \boldsymbol{\alpha}^T \boldsymbol{\Sigma}_{11} \boldsymbol{\alpha} & \boldsymbol{\alpha}^T \boldsymbol{\Sigma}_{12} \boldsymbol{\beta} \\ \boldsymbol{\beta}^T \boldsymbol{\Sigma}_{21} \boldsymbol{\alpha} & \boldsymbol{\beta}^T \boldsymbol{\Sigma}_{22} \boldsymbol{\beta} \end{pmatrix}\right).$$

Therefore, using (11), it follows that the problem is to find $\boldsymbol{\alpha} \in \mathbb{R}^h$ and $\boldsymbol{\beta} \in \mathbb{R}^{d-h}$ such that it maximizes

$$\text{corr}(\boldsymbol{\alpha}^T \boldsymbol{X}_1, \boldsymbol{\beta}^T \boldsymbol{X}_2) = \frac{p \boldsymbol{\alpha}^T \boldsymbol{\Sigma}_{12} \boldsymbol{\beta}^T}{\sqrt{p \boldsymbol{\alpha}^T \boldsymbol{\Sigma}_{11} \boldsymbol{\alpha}} \sqrt{p \boldsymbol{\beta}^T \boldsymbol{\Sigma}_{22} \boldsymbol{\beta}}} = \boldsymbol{\alpha}^T \boldsymbol{\Sigma}_{12} \boldsymbol{\beta}^T,$$

subject to the restrictions $\boldsymbol{\alpha}^T\boldsymbol{\Sigma}_{11}\boldsymbol{\alpha} = 1$ and $\boldsymbol{\beta}^T\boldsymbol{\Sigma}_{22}\boldsymbol{\beta} = 1$. Now following the same steps as in the multivariate normal cases, Anderson [1], the result follows. ∎

The following result provides the characteristic function of the Wishart type matrix based on MGSN random variables.

Theorem 3: Suppose $Z_1, \ldots, Z_n$ are n i.i.d. random variables, and $Z_1 \sim \text{MGSN}_d(p, \mathbf{0}, \boldsymbol{\Sigma})$. Let us consider the Wishart type matrix

$$\boldsymbol{A} = \sum_{m=1}^{n} \boldsymbol{Z}_m \boldsymbol{Z}_m^T = ((A_{ij})), \quad i, j = 1, \ldots, d.$$

If $\boldsymbol{\Theta} = ((\theta_{ij}))$ with $\theta_{ij} = \theta_{ji}$ is a $d \times d$ matrix, then the characteristic function of $(A_{11}, \ldots, A_{pp}, 2A_{12}, 2A_{13}, \ldots, 2A_{p-1,p})$ is

$$E\left(e^{i\text{tr}(\boldsymbol{A\Theta})}\right) = p^n \left[\sum_{k=1}^{\infty} |\boldsymbol{I} - 2ik\boldsymbol{\Theta\Sigma}|^{-1/2} (1-p)^{k-1}\right]^n.$$

Proof:

$$\begin{aligned} E\left[e^{i\text{tr}(\boldsymbol{A\Theta})}\right] &= E\left[e^{i\text{tr}\left(\sum_{m=1}^{n} \boldsymbol{Z}_m \boldsymbol{Z}_m^T \boldsymbol{\Theta}\right)}\right] = E\left[e^{i\text{tr}\left(\sum_{m=1}^{n} \boldsymbol{Z}_m^T \boldsymbol{\Theta} \boldsymbol{Z}_m\right)}\right] \\ &= E\left[e^{i\left(\sum_{m=1}^{n} \boldsymbol{Z}_m^T \boldsymbol{\Theta} \boldsymbol{Z}_m\right)}\right] = \left(E\left[e^{i\left(\boldsymbol{Z}_1^T \boldsymbol{\Theta} \boldsymbol{Z}_1\right)}\right]\right)^n. \end{aligned}$$

Now we would like to compute $E\left[e^{i\left(\boldsymbol{Z}_1^T \boldsymbol{\Theta} \boldsymbol{Z}_1\right)}\right]$. For a $d \times d$ real symmetric matrix $\boldsymbol{\Theta}$, there is a real $d \times d$ matrix $\boldsymbol{B}$ such that

$$\boldsymbol{B}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{B} = \boldsymbol{I} \quad \text{and} \quad \boldsymbol{B}^T\boldsymbol{\Theta}\boldsymbol{B} = \boldsymbol{D} = \text{diag}\{\delta_1, \ldots, \delta_d\}.$$

Here, $\text{diag}\{\delta_1, \ldots, \delta_d\}$ means a $d \times d$ diagonal matrix with diagonal entries as $\delta_1, \ldots, \delta_d$. If we make the transformation $\boldsymbol{Z}_1 = \boldsymbol{BY}$, then using Theorem 1, $\boldsymbol{Y} \sim \text{MGSN}_d(p, \mathbf{0}, \boldsymbol{I})$. Using the definition MGSN distribution it follows that

$$\boldsymbol{Y} \stackrel{d}{=} \sum_{m=1}^{N} \boldsymbol{Y}_m,$$

here $N \sim \mathrm{GE}(p)$, and $\boldsymbol{Y}_m$'s are i.i.d. random vectors, and $\boldsymbol{Y}_1 \sim \mathrm{N}_d(\boldsymbol{0}, \boldsymbol{I})$. We use the following notation

$$\boldsymbol{Y} = (Y_1, \ldots, Y_d)^T \quad \text{and} \quad \boldsymbol{Y}_m = (Y_{m1}, \ldots, Y_{md})^T.$$

Hence, $Y_j = \sum_{m=1}^{N} Y_{mj}$, for $j = 1, \ldots, d$. Therefore,

$$\begin{aligned}
E\left[e^{i\left(\boldsymbol{Z}_1^T \boldsymbol{\Theta} \boldsymbol{Z}_1\right)}\right] &= E\left[e^{i\left(\boldsymbol{Y}^T \boldsymbol{D} \boldsymbol{Y}\right)}\right] = E\left[e^{i\left(\sum_{j=1}^d \delta_j Y_j^2\right)}\right] = E\left[e^{i\left(\sum_{j=1}^d \delta_j \left(\sum_{m=1}^N Y_{mj}\right)^2\right)}\right] \\
&= E_N E\left[e^{i\left(\sum_{j=1}^d \delta_j \left(\sum_{m=1}^N Y_{mj}\right)^2\right)} \middle| N\right] = E_N E\left[e^{i\left(\sum_{j=1}^d \delta_j N \left(\sum_{m=1}^N Y_{mj}/\sqrt{N}\right)^2\right)} \middle| N\right] \\
&= E_N \prod_{j=1}^d E\left[e^{i\left(\delta_j N \left(\sum_{m=1}^N Y_{mj}/\sqrt{N}\right)^2\right)} \middle| N\right] = E_N \prod_{j=1}^d (1 - 2i\delta_j N)^{-1/2} \\
&= E_N \left|\boldsymbol{I} - 2iN\boldsymbol{D}\right|^{-1/2} = E_N \left|\boldsymbol{I} - 2iN\boldsymbol{\Theta}\boldsymbol{\Sigma}\right|^{-1/2} \\
&= p \sum_{k=1}^{\infty} \left|\boldsymbol{I} - 2ik\boldsymbol{\Theta}\boldsymbol{\Sigma}\right|^{-1/2} (1-p)^{k-1}.
\end{aligned}$$

■

THEOREM 4: If for any $\boldsymbol{c} \neq \boldsymbol{0}, \boldsymbol{c} \in \mathbb{R}^d$, $\boldsymbol{c}^T \boldsymbol{X} \sim \mathrm{GSN}(\mu(\boldsymbol{c}), \sigma(\boldsymbol{c}), p)$ for a d dimensional random vector $\boldsymbol{X}$, then there exists a d-dimensional vector $\boldsymbol{\mu}$ and a $d \times d$ symmetric matrix $\boldsymbol{\Sigma}$ such that $\boldsymbol{X} \sim \mathrm{MGSN}_d(p, \boldsymbol{\mu}, \boldsymbol{\Sigma})$. Here $-\infty < \mu(\boldsymbol{c}) < \infty, 0 < \sigma(\boldsymbol{c}) < \infty$ are functions of $\boldsymbol{c}$.

PROOF: If we denote the mean vector and dispersion matrix of the random vector $\boldsymbol{X}$, as $\boldsymbol{\mu}_{\boldsymbol{X}}$ and $\boldsymbol{\Sigma}_{\boldsymbol{X}}$, respectively, then we have $E(\boldsymbol{c}^T \boldsymbol{X}) = \boldsymbol{c}^T \boldsymbol{\mu}_{\boldsymbol{X}}$ and $V(\boldsymbol{c}^T \boldsymbol{X}) = \boldsymbol{c}^T \boldsymbol{\Sigma}_{\boldsymbol{X}} \boldsymbol{c}$. Hence from (3) and (4), we have

$$\mu(\boldsymbol{c}) = p\boldsymbol{c}^T \boldsymbol{\mu}_{\boldsymbol{X}} \quad \text{and} \quad \sigma^2(\boldsymbol{c}) = p\boldsymbol{c}^T \boldsymbol{\Sigma}_{\boldsymbol{X}} \boldsymbol{c} - p(1-p)\left(\boldsymbol{c}^T \boldsymbol{\mu}_{\boldsymbol{X}}\right)^2. \tag{13}$$

Therefore, from (1), using $t = 1$, it follows that

$$E\left(\exp(\boldsymbol{c}^T \boldsymbol{X})\right) = \frac{p \exp(\mu(\boldsymbol{c}) + \sigma^2(\boldsymbol{c})/2)}{1 - (1-p)\exp(\mu(\boldsymbol{c}) + \sigma^2(\boldsymbol{c})/2)}. \tag{14}$$

Let us define a d-dimensional vector $\boldsymbol{\mu}$ and a $d \times d$ symmetric matrix $\boldsymbol{\Sigma}$ as given below:

$$\boldsymbol{\mu} = p\boldsymbol{\mu}_{\boldsymbol{X}} \quad \text{and} \quad \boldsymbol{\Sigma} = p\boldsymbol{\Sigma}_{\boldsymbol{X}} - p(1-p)\boldsymbol{\mu}_{\boldsymbol{X}}\boldsymbol{\mu}_{\boldsymbol{X}}^T. \tag{15}$$

Therefore,

$$\mu(\boldsymbol{c}) + \frac{\sigma^2(\boldsymbol{c})}{2} = \boldsymbol{c}^T\boldsymbol{\mu} + \frac{1}{2}\boldsymbol{c}^T\boldsymbol{\Sigma}\boldsymbol{c},$$

and (14) can be written as

$$E\left(e^{\boldsymbol{c}^T\boldsymbol{X}}\right) = \frac{p\exp\left(\boldsymbol{c}^T\boldsymbol{\mu} + \frac{1}{2}\boldsymbol{c}^T\boldsymbol{\Sigma}\boldsymbol{c}\right)}{1-(1-p)\exp\left(\boldsymbol{c}^T\boldsymbol{\mu} + \frac{1}{2}\boldsymbol{c}^T\boldsymbol{\Sigma}\boldsymbol{c}\right)} = M_{\boldsymbol{X}}(\boldsymbol{c}).$$

Hence the result follows. ∎

Therefore, combining Theorems 1 and 2, we have the following characterization results for a d-variate MGSN distribution.

THEOREM 5: If a d-dimensional random vector $\boldsymbol{X}$ has a mean vector $\mu_{\boldsymbol{X}}$ and a dispersion matrix $\boldsymbol{\Sigma}_{\boldsymbol{X}}$, then $\boldsymbol{X} \sim \text{MGSN}_d(p, \boldsymbol{\mu}, \boldsymbol{\Sigma})$, here $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are as defined in (15), if and only if for any $\boldsymbol{c} \neq \boldsymbol{0}, \boldsymbol{c} \in \mathbb{R}^d$, $\boldsymbol{c}^T\boldsymbol{X} \sim \text{GSN}(\mu(\boldsymbol{c}), \sigma(\boldsymbol{c}), p)$, where $\mu(\boldsymbol{c})$ and $\sigma(\boldsymbol{c})$ are as in (13).

Now we provide another characterization of the MGSN distribution.

THEOREM 6: Suppose $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots$ is a sequence of i.i.d. d-dimensional random vectors, and $M \sim \text{GE}(\alpha)$, for $0 < \alpha \leq 1$. Consider a new d-dimensional random vector

$$\boldsymbol{Y} = \sum_{i=1}^{M} \boldsymbol{X}_i.$$

Then $\boldsymbol{Y} \sim \text{MGSN}_d(\beta, \boldsymbol{\mu}, \boldsymbol{\Sigma})$ for $\beta \leq \alpha$, if and only if $\boldsymbol{X}_1$ has a MGSN distribution.

PROOF: If part. Suppose $\boldsymbol{X}_1 \sim \text{MGSN}_d(p, \boldsymbol{\mu}, \boldsymbol{\Sigma})$, then the MGF of $\boldsymbol{Y}$ for $\boldsymbol{t} \in \mathbb{R}^d$, can be written as

$$M_{\boldsymbol{Y}}(\boldsymbol{t}) = E\left(e^{\boldsymbol{t}^T\boldsymbol{Y}}\right) = \sum_{m=1}^{\infty} E\left(e^{\sum_{i=1}^{M}\boldsymbol{t}^T\boldsymbol{X}_i}|M=m\right) P(M=m)$$

$$= \sum_{m=1}^{\infty} \alpha(1-\alpha)^{m-1} \left(\frac{p e^{\boldsymbol{\mu}^T \boldsymbol{t} + \frac{1}{2}\boldsymbol{t}^T \boldsymbol{\Sigma} \boldsymbol{t}}}{1-(1-p)e^{\boldsymbol{\mu}^T \boldsymbol{t} + \frac{1}{2}\boldsymbol{t}^T \boldsymbol{\Sigma} \boldsymbol{t}}} \right)^m$$

$$= \frac{\alpha p e^{\boldsymbol{\mu}^T \boldsymbol{t} + \frac{1}{2}\boldsymbol{t}^T \boldsymbol{\Sigma} \boldsymbol{t}}}{1-(1-\alpha p)e^{\boldsymbol{\mu}^T \boldsymbol{t} + \frac{1}{2}\boldsymbol{t}^T \boldsymbol{\Sigma} \boldsymbol{t}}} = \frac{\beta e^{\boldsymbol{\mu}^T \boldsymbol{t} + \frac{1}{2}\boldsymbol{t}^T \boldsymbol{\Sigma} \boldsymbol{t}}}{1-(1-\beta)e^{\boldsymbol{\mu}^T \boldsymbol{t} + \frac{1}{2}\boldsymbol{t}^T \boldsymbol{\Sigma} \boldsymbol{t}}},$$

here $\beta = \alpha p \le \alpha$.

Only if part. Suppose $\boldsymbol{Y} \sim \text{MGSN}_d(\beta, \boldsymbol{\mu}, \boldsymbol{\Sigma})$ for some $0 < \beta \le \alpha$, and the MGF of $\boldsymbol{X}_1$ is $M_{\boldsymbol{X}_1}(\boldsymbol{t})$. We have the following relation:

$$\frac{\beta e^{\boldsymbol{\mu}^T \boldsymbol{t} + \frac{1}{2}\boldsymbol{t}^T \boldsymbol{\Sigma} \boldsymbol{t}}}{1-(1-\beta)e^{\boldsymbol{\mu}^T \boldsymbol{t} + \frac{1}{2}\boldsymbol{t}^T \boldsymbol{\Sigma} \boldsymbol{t}}} = \frac{\alpha M_{\boldsymbol{X}_1}(\boldsymbol{t})}{1-(1-\alpha)M_{\boldsymbol{X}_1}(\boldsymbol{t})}. \tag{16}$$

From (16), we obtain

$$M_{\boldsymbol{X}_1}(\boldsymbol{t}) = \frac{\gamma e^{\boldsymbol{\mu}^T \boldsymbol{t} + \frac{1}{2}\boldsymbol{t}^T \boldsymbol{\Sigma} \boldsymbol{t}}}{1-(1-\gamma)e^{\boldsymbol{\mu}^T \boldsymbol{t} + \frac{1}{2}\boldsymbol{t}^T \boldsymbol{\Sigma} \boldsymbol{t}}},$$

for $\gamma = \beta/\alpha \le 1$. Therefore, $\boldsymbol{X}_1 \sim \text{MGSN}_d(\gamma, \boldsymbol{\mu}, \boldsymbol{\Sigma})$. ∎

Stochastic ordering plays a very important role in the distribution theory. It has been studied quite extensively in the statistical literature. For its importance and different applications, interested readers are referred to Shaked and Shantikumar [14]. Now we will discuss the multivariate total positivity of order two (MTP_2) property, in the sense of Karlin and Rinott [8], of the joint PDF of MGSN distribution. We shall be using the following notation here. For any two real numbers a and b, let $a \wedge b = \min\{a, b\}$, and $a \vee b = \max\{a, b\}$. For any vector $\boldsymbol{x} = (x_1, \ldots, x_d)^T$ and $\boldsymbol{y} = (y_1, \ldots, y_d)^T$, let $\boldsymbol{x} \vee \boldsymbol{y} = (x_1 \vee y_1, \ldots, x_d \vee y_d)^T$ and $\boldsymbol{x} \wedge \boldsymbol{y} = (x_1 \wedge y_1, \ldots, x_d \wedge y_d)^T$. Let us recall the definition of MTP_2 property. A function $g : \mathbb{R}^d \to \mathbb{R}^+$ is said to have MTP_2 property, in the sense of Karlin and Rinott [8], if $g(\boldsymbol{x})g(\boldsymbol{y}) \le g(\boldsymbol{x} \wedge \boldsymbol{y})g(\boldsymbol{x} \vee \boldsymbol{y})$, for all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d$. We then have the following result for MGSN distribution.

Theorem 7: Let $\boldsymbol{X} \sim \text{MGSN}_d(p, \boldsymbol{0}, \boldsymbol{\Sigma})$, and all the off-diagonal elements of $\boldsymbol{\Sigma}^{-1}$ are less than or equal to zero, then the PDF of $\boldsymbol{X}$ has MTP_2 property.

Proof: To prove that the PDF of $\boldsymbol{X}$ has MTP_2 property, it is enough to show that

$$\boldsymbol{x}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{x} + \boldsymbol{y}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{y} \geq (\boldsymbol{x} \vee \boldsymbol{y})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x} \vee \boldsymbol{y}) + (\boldsymbol{x} \wedge \boldsymbol{y})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x} \wedge \boldsymbol{y}) \tag{17}$$

for any $\boldsymbol{x} = (x_1, \ldots, x_d)^T \in \mathbb{R}^d$ and $\boldsymbol{y} = (y_1, \ldots, y_d)^T \in \mathbb{R}^d$. If the elements of $\boldsymbol{\Sigma}^{-1}$ are denoted by $((\sigma^{kj}))$, for $k, j = 1, \ldots, d$, then proving (17) is equivalent to showing

$$\sum_{\substack{k,j=1 \\ k \neq j}}^{d} (x_j x_k + y_j y_k)\sigma^{jk} \geq \sum_{\substack{k,j=1 \\ k \neq j}}^{d} ((x_j \wedge y_j)(x_k \wedge y_k) + (x_j \vee y_j)(x_k \vee y_k))\sigma^{jk}.$$

For all $k, j = 1, \ldots, d$,

$$(x_j x_k + y_j y_k) \leq (x_j \wedge y_j)(x_k \wedge y_k) + (x_j \vee y_j)(x_k \vee y_k),$$

which can be easily shown by taking any ordering of x_k, x_j, y_k, y_j. Now the result follows since $\sigma^{jk} \leq 0$. ∎

The following two decompositions of a MGSN distribution are possible. We use the following notations. The distribution of a negative binomial random variable with parameters r and p, where r is a non-negative integer and $0 < p < 1$, will be denoted by $\text{NB}(r, p)$. If $T \sim \text{NB}(r, p)$, then the MGF of T is

$$M_T(t) = \left(\frac{1-p}{1-pe^t}\right)^r \quad \text{for} \quad t < -\ln p.$$

A discrete random variable Z is said to have a logarithmic distribution with parameter p, for $0 < p < 1$, if the PMF of Z is

$$P(Z = k) = \frac{(1-p)^k}{\lambda k} \quad \text{for} \quad k = 1, 2, \ldots, \quad \text{where} \quad \lambda = -\ln p,$$

and it will be denoted by $\text{LD}(p)$. Now we provide two decompositions of MGSN distribution.

Decomposition 1: Suppose $\boldsymbol{X} \sim \text{MGSN}_d(p, \boldsymbol{\mu}, \boldsymbol{\Sigma})$. Further, for any positive integer n and for any $1 \leq k \leq n$, suppose

$$\boldsymbol{Z}_{kn} \stackrel{disp}{=} \sum_{j=1}^{1+nT} \boldsymbol{Y}_j,$$

where $T \sim \text{NB}(r, p)$ and $r = 1/n$, $\boldsymbol{Y}_j$'s are i.i.d. $\text{N}_d(r\boldsymbol{\mu}, r\boldsymbol{\Sigma})$, T and $\boldsymbol{Y}_j$'s are independently distributed, then

$$\boldsymbol{X} \overset{disp}{=} \boldsymbol{Z}_{1n} + \ldots + \boldsymbol{Z}_{nn}.$$

Proof: The MGF of $\boldsymbol{Z}_{kn}$ can be written as

$$\begin{aligned} M_{\boldsymbol{Z}_{kn}}(\boldsymbol{t}) &= E\left(e^{\boldsymbol{t}^T\boldsymbol{Z}_{kn}}\right) = \sum_{j=0}^{\infty} E\left(e^{\boldsymbol{t}^T\boldsymbol{Z}_{kn}}|T=j\right) P(T=j) \\ &= \left[\frac{pe^{\boldsymbol{\mu}^T\boldsymbol{t}+\frac{1}{2}\boldsymbol{t}^T\boldsymbol{\Sigma}\boldsymbol{t}}}{1-(1-p)e^{\boldsymbol{\mu}^T\boldsymbol{t}+\frac{1}{2}\boldsymbol{t}^T\boldsymbol{\Sigma}\boldsymbol{t}}}\right]^r = \left[M_{\boldsymbol{X}}(\boldsymbol{t})\right]^r. \end{aligned}$$

■

It implies that MGSN law is infinitely divisible. The following decomposition is also possible.

Decomposition 2: Suppose Q is a Poisson random variable with parameter λ, and $\{Z_i, i = 1, 2, \ldots\}$ is a sequence of i.i.d. random variables having logarithmic distribution with the following probability mass function for $\lambda = -\ln p$;

$$P(Z_1 = k) = \frac{(1-p)^k}{\lambda k}; \quad k = 1, 2, \ldots,$$

and all the random variables are independently distributed. If $\boldsymbol{X} \sim \text{MGSN}_d(p, \boldsymbol{\mu}, \boldsymbol{\Sigma})$, then the following decomposition is possible

$$\boldsymbol{X} \overset{disp}{=} \boldsymbol{Y} + \sum_{i=1}^{Q} \boldsymbol{Y}_i, \tag{18}$$

here $\{\boldsymbol{Y}_i | Z_i = k\} \sim \text{N}_d(k\boldsymbol{\mu}, k\boldsymbol{\Sigma})$ for $i = 1, 2, \ldots$, and they are independently distributed, $\boldsymbol{Y} \sim \text{N}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, and it is independent of Q, and $(\boldsymbol{Y}_i, Z_i)$ for all $i = 1, 2, \ldots$.

Proof: First note that the probability generating function of Q and Z_1 are as follows:

$$E(t^Q) = e^{\lambda(t-1)} \quad \text{and} \quad E(t^{Z_1}) = \frac{\ln(1-(1-p)t)}{\ln p}; \quad t < (1-p)^{-1}.$$

The MGF of $\boldsymbol{Y}_i$ for $\boldsymbol{t} \in \mathbb{R}^d$, such that $(1-p)e^{\boldsymbol{\mu}^T\boldsymbol{t}+\frac{1}{2}\boldsymbol{t}^T\boldsymbol{\Sigma}\boldsymbol{t}} < 1$, can be obtained as

$$M_{\boldsymbol{Y}_i}(\boldsymbol{t}) = E\left(e^{\boldsymbol{t}^T\boldsymbol{Y}_i}\right) = E_{Z_i}E_{\boldsymbol{Y}_i|Z_i}\left(e^{\boldsymbol{t}^T\boldsymbol{Y}_i}\right) = \frac{\ln\left(1-(1-p)e^{\boldsymbol{\mu}^T\boldsymbol{t}+\frac{1}{2}\boldsymbol{t}^T\boldsymbol{\Sigma}\boldsymbol{t}}\right)}{\ln p}.$$

Therefore, the MGF of the right hand side of (18) can be written as

$$\begin{aligned} E\left[e^{\boldsymbol{t}^T(\boldsymbol{Y}+\sum_{i=1}^{Q}\boldsymbol{Y}_i)}\right] &= e^{\boldsymbol{\mu}^T\boldsymbol{t}+\frac{1}{2}\boldsymbol{t}^T\boldsymbol{\Sigma}\boldsymbol{t}} \times E\left[\frac{\ln\left(1-(1-p)e^{\boldsymbol{\mu}^T\boldsymbol{t}+\frac{1}{2}\boldsymbol{t}^T\boldsymbol{\Sigma}\boldsymbol{t}}\right)}{\ln p}\right]^Q \\ &= \frac{pe^{\boldsymbol{\mu}^T\boldsymbol{t}+\frac{1}{2}\boldsymbol{t}^T\boldsymbol{\Sigma}\boldsymbol{t}}}{1-(1-p)e^{\boldsymbol{\mu}^T\boldsymbol{t}+\frac{1}{2}\boldsymbol{t}^T\boldsymbol{\Sigma}\boldsymbol{t}}} = M_{\boldsymbol{X}}(\boldsymbol{t}). \end{aligned}$$

∎

The following results will be useful for further development. Let us consider the random vector $(\boldsymbol{X}, N)$, where $\boldsymbol{X}$ and N are same as defined in (5). The joint PDF of $(\boldsymbol{X}, N)$ can be written as

$$f_{\boldsymbol{X},N}(\boldsymbol{x}, n) = \begin{cases} \frac{p(1-p)^{n-1}}{(2\pi)^{d/2}|\boldsymbol{\Sigma}|^{1/2}n^{d/2}} e^{-\frac{1}{2n}(\boldsymbol{x}-n\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-n\boldsymbol{\mu})} & \text{if} \quad 0 < p < 1 \\ \frac{1}{(2\pi)^{d/2}|\boldsymbol{\Sigma}|^{1/2}} e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})} & \text{if} \quad p = 1, \end{cases}$$

for $\boldsymbol{x} \in \mathbb{R}^d$ and for any positive integer n. Therefore, the conditional probability mass function of N given $\boldsymbol{X} = \boldsymbol{x}$ becomes

$$P(N = n|\boldsymbol{X} = \boldsymbol{x}) = \frac{(1-p)^{n-1}e^{-\frac{1}{2n}(\boldsymbol{x}-n\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-n\boldsymbol{\mu})}n^{-d/2}}{\sum_{k=1}^{\infty}(1-p)^{k-1}e^{-\frac{1}{2k}(\boldsymbol{x}-k\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-k\boldsymbol{\mu})}k^{-d/2}}.$$

Therefore,

$$E(N|\boldsymbol{X} = \boldsymbol{x}) = \frac{\sum_{n=1}^{\infty}(1-p)^{n-1}e^{-\frac{1}{2n}(\boldsymbol{x}-n\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-n\boldsymbol{\mu})}n^{-d/2+1}}{\sum_{k=1}^{\infty}(1-p)^{k-1}e^{-\frac{1}{2k}(\boldsymbol{x}-k\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-k\boldsymbol{\mu})}k^{-d/2}}, \tag{19}$$

and

$$E(N^{-1}|\boldsymbol{X} = \boldsymbol{x}) = \frac{\sum_{n=1}^{\infty}(1-p)^{n-1}e^{-\frac{1}{2n}(\boldsymbol{x}-n\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-n\boldsymbol{\mu})}n^{-d/2-1}}{\sum_{k=1}^{\infty}(1-p)^{k-1}e^{-\frac{1}{2k}(\boldsymbol{x}-k\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-k\boldsymbol{\mu})}k^{-d/2}}. \tag{20}$$

4 Statistical Inference

4.1 Estimation

In this section we discuss the maximum likelihood estimators (MLEs) of the unknown parameters, when $0 < p < 1$. When $p = 1$, the MLEs of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ can be easily obtained as the sample mean and the sample variance covariance matrix, respectively. Suppose $\mathcal{D} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$ is a random sample of size n from $\text{MGSN}_d(p, \boldsymbol{\mu}, \boldsymbol{\Sigma})$, then the log-likelihood function becomes

$$\begin{aligned} l(p, \boldsymbol{\mu}, \boldsymbol{\Sigma}) &= \sum_{i=1}^{n} \ln f_{\boldsymbol{X}}(\boldsymbol{x}_i; \boldsymbol{\mu}, \boldsymbol{\Sigma}, p) \\ &= \sum_{i=1}^{n} \ln \left[\sum_{k=1}^{\infty} \frac{p(1-p)^{k-1}}{(2\pi)^{d/2} |\boldsymbol{\Sigma}|^{1/2} k^{d/2}} e^{-\frac{1}{2k}(\boldsymbol{x}_i - k\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{x}_i - k\boldsymbol{\mu})} \right]. \end{aligned} \tag{21}$$

The maximum likelihood estimators (MLEs) of the unknown parameters can be obtained by maximizing (21) with respect to the unknown parameters. It involves solving a $(d+1+d(d+1)/2)$ dimensional optimization problem. Therefore, for large d, it is a challenging issue.

To avoid that problem, first it is assumed that p is known. For a known p, we estimate the MLEs of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ by using EM algorithm, say $\widehat{\boldsymbol{\mu}}(p)$ and $\widehat{\boldsymbol{\Sigma}}(p)$, respectively. We maximize $l(p, \widehat{\boldsymbol{\mu}}(p), \widehat{\boldsymbol{\Sigma}}(p))$ to compute $\widehat{p}$, the MLE of p. Finally we obtain the MLE of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ as $\widehat{\boldsymbol{\mu}} = \widehat{\boldsymbol{\mu}}(\widehat{p})$ and $\widehat{\boldsymbol{\Sigma}} = \widehat{\boldsymbol{\Sigma}}(\widehat{p})$, respectively. Now we will show how to compute $\widehat{\boldsymbol{\mu}}(p)$ and $\widehat{\boldsymbol{\Sigma}}(p)$, for a given p using EM algorithm. We treat the problem as a missing value problem, and the main idea is as follows.

It is assumed that p is known. Suppose we have the complete observations of the form $\{(\boldsymbol{x}_1, m_1), \ldots, (\boldsymbol{x}_n, m_n)\}$ from $(\boldsymbol{X}, N)$. Then the log-likelihood function based on the complete observation becomes (without the additive constant)

$$l_c(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = -\frac{n}{2} \ln |\boldsymbol{\Sigma}| - \frac{1}{2} \sum_{i=1}^{n} \frac{1}{m_i} (\boldsymbol{x}_i - m_i \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{x}_i - m_i \boldsymbol{\mu}).$$

Therefore, if we define the MLEs of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ based on the complete observations as $\widehat{\boldsymbol{\mu}}_c(p)$ and $\widehat{\Sigma}_c(p)$, respectively, then for $K = \sum_{i=1}^{n} m_i$,

$$\widehat{\boldsymbol{\mu}}_c(p) = \frac{1}{K} \sum_{i=1}^{n} \boldsymbol{x}_i, \tag{22}$$

and

$$\widehat{\boldsymbol{\Sigma}}_c(p) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{m_i} (\boldsymbol{x}_i - m_i \widehat{\boldsymbol{\mu}}_c)(\boldsymbol{x}_i - m_i \widehat{\boldsymbol{\mu}}_c)^T$$

$$= \frac{1}{n}\left[\sum_{i=1}^{n}\frac{1}{m_i}\boldsymbol{x}_i\boldsymbol{x}_i^T - \sum_{i=1}^{n}(\widehat{\boldsymbol{\mu}}_c\boldsymbol{x}_i^T + \boldsymbol{x}_i\widehat{\boldsymbol{\mu}}_c^T) + K\widehat{\boldsymbol{\mu}}_c\widehat{\boldsymbol{\mu}}_c^T\right]. \quad (23)$$

Note that $\widehat{\boldsymbol{\mu}}_c(p)$ is obtained by taking derivative of $\sum_{i=1}^{n}\frac{1}{m_i}(\boldsymbol{x}_i - m_i\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_i - m_i\boldsymbol{\mu})$ with respect to $\boldsymbol{\mu}$, and equate it to zero. Similarly, $\widehat{\boldsymbol{\Sigma}}_c(p)$ is obtained by using Lemma 3.2.2 of Anderson [1].

Now we are ready to provide the EM algorithm for a given p. The EM algorithm consists of maximizing the conditional expectation of the complete log-likelihood function, based on the observed data and the current value of $\boldsymbol{\theta} = (\boldsymbol{\mu}, \boldsymbol{\Sigma})$, say $\widetilde{\boldsymbol{\theta}}$, in an iterative two-step algorithm process, see for example Dempster et al. [6]. The E-step is to compute the conditional expectation denoted by $Q(\boldsymbol{\theta}|\widetilde{\boldsymbol{\theta}})$, and the M-step is maximizing $Q(\boldsymbol{\theta}|\widetilde{\boldsymbol{\theta}})$, with respect to $\boldsymbol{\theta}$. We use the following notations:

$$a_i = E(N|\boldsymbol{X} = \boldsymbol{x}_i, \widetilde{\boldsymbol{\theta}}) \quad \text{and} \quad b_i = E(N^{-1}|\boldsymbol{X} = \boldsymbol{x}_i, \widetilde{\boldsymbol{\theta}}),$$

where a_i and b_i are obtained using (19) and (20), respectively.

E-Step: It consists of calculating $Q(\boldsymbol{\theta}|\widetilde{\boldsymbol{\theta}})$, $\widetilde{\boldsymbol{\theta}}$ being the current parameter value.

$$\begin{aligned} Q(\boldsymbol{\theta}|\widetilde{\boldsymbol{\theta}}) &= E(l_c(\boldsymbol{\theta}|\mathcal{D}, \widetilde{\boldsymbol{\theta}})) \\ &= -\frac{n}{2}\ln|\boldsymbol{\Sigma}| - \frac{1}{2}\text{trace}\left\{\boldsymbol{\Sigma}^{-1}\left(\sum_{i=1}^{n} b_i\boldsymbol{x}_i\boldsymbol{x}_i^T - \sum_{i=1}^{n}(\boldsymbol{x}_i\boldsymbol{\mu}^T + \boldsymbol{\mu}\boldsymbol{x}_i^T) + \boldsymbol{\mu}\boldsymbol{\mu}^T\sum_{i=1}^{n} a_i\right)\right\}. \end{aligned}$$

M-Step: It involves maximizing $Q(\boldsymbol{\theta}|\widetilde{\boldsymbol{\theta}})$ with respect to $\boldsymbol{\theta}$, to obtain $\overline{\boldsymbol{\theta}}$, where

$$\overline{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} Q(\boldsymbol{\theta}|\widetilde{\boldsymbol{\theta}}).$$

Here, $\arg\max_{\boldsymbol{\theta}} Q(\boldsymbol{\theta}|\widetilde{\boldsymbol{\theta}})$ means the value of $\boldsymbol{\theta}$ for which the function $Q(\boldsymbol{\theta}|\widetilde{\boldsymbol{\theta}})$ takes the maximum value. From (22) and (23), we obtain

$$\overline{\boldsymbol{\mu}} = \frac{1}{\sum_{j=1}^{n} a_j}\sum_{i=1}^{n}\boldsymbol{x}_i$$

and

$$\overline{\boldsymbol{\Sigma}} = \frac{1}{n}\left[\sum_{i=1}^{n} b_i \boldsymbol{x}_i \boldsymbol{x}_i^T - \sum_{i=1}^{n}(\boldsymbol{x}_i \overline{\boldsymbol{\mu}}^T + \overline{\boldsymbol{\mu}} \boldsymbol{x}_i^T) + \overline{\boldsymbol{\mu}}\ \overline{\boldsymbol{\mu}}^T \sum_{i=1}^{n} a_i\right]. \tag{24}$$

We propose the following algorithm to compute the MLEs of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ for a known p.

Algorithm 2:

- Step 1: Choose an initial guess of $\boldsymbol{\theta}$, say $\boldsymbol{\theta}^{(0)}$.
- Step 2: Obtain

$$\boldsymbol{\theta}^{(1)} = \arg\max_{\boldsymbol{\theta}} Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(0)}).$$

- Step 3: Continue the process until convergence takes place.

Once for a given p, the MLEs of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are obtained, say $\widehat{\boldsymbol{\mu}}(p)$ and $\widehat{\boldsymbol{\Sigma}}(p)$, respectively, then the MLE of p can be obtained by maximizing the profile log-likelihood function of p, i.e. $l(p, \widehat{\boldsymbol{\mu}}(p), \widehat{\boldsymbol{\Sigma}}(p))$, with respect to p. If it is denoted by $\widehat{p}$, then the MLEs of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ become $\widehat{\boldsymbol{\mu}} = \widehat{\boldsymbol{\mu}}(\widehat{p})$ and $\widehat{\boldsymbol{\Sigma}} = \widehat{\boldsymbol{\Sigma}}(\widehat{p})$, respectively. The details will be explained in Section 5. We have used the sample mean vector and the sample variance covariance matrix as the initial guess of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, respectively, of the proposed EM algorithm for all p.

4.2 Testing of Hypotheses

In this section we discuss three different testing of hypotheses problems which can be useful in practice. We propose to use the likelihood ratio test (LRT) in all the cases, and we indicate the asymptotic distribution of the LRT tests under the null hypothesis in each case. With the abuse of notations, in each case if δ is any unknown parameter, the MLE of δ under the null hypothesis will be denoted by $\widehat{\delta}_H$.

Test 1:

$$H_0 : p = 1 \quad \text{vs.} \quad H_1 : p < 1. \tag{25}$$

The above testing problem (25) is important in practice as it tests the normality of the distribution. In this $\widehat{\boldsymbol{\mu}}_H$ and $\widehat{\boldsymbol{\Sigma}}_H$, respectively, can be obtained as the sample mean and the sample variance covariance matrix. Since in this case p is in the boundary, the standard results do not work. But using result 3 of Self and Liang [13] it follows that under the null hypothesis

$$T_1 = 2(l(\widehat{p}, \widehat{\boldsymbol{\mu}}, \widehat{\boldsymbol{\Sigma}}) - l(1, \widehat{\boldsymbol{\mu}}_H, \widehat{\boldsymbol{\Sigma}}_H)) \longrightarrow \frac{1}{2} + \frac{1}{2}\chi^2_1.$$

Test 2:

$$H_0 : \boldsymbol{\mu} = \mathbf{0} \quad \text{vs.} \quad H_1 : \boldsymbol{\mu} \neq \mathbf{0}. \tag{26}$$

The above testing problem (26) is important as it tests the symmetry of the distribution. In this case under the null hypothesis the MLEs of p and $\boldsymbol{\Sigma}$ can be obtained as follows. For a given p, the MLE of $\boldsymbol{\Sigma}$ can be obtained using the EM algorithm as before, and then the MLE of p can be obtained by maximizing the profile likelihood function. In this case the 'E-step' and 'M-Step' can be obtained from (24) and (24), respectively, by replacing $\boldsymbol{\mu} = \mathbf{0}$. Under H_0, then

$$T_2 = 2(l(\widehat{p}, \widehat{\boldsymbol{\mu}}, \widehat{\boldsymbol{\Sigma}}) - l(\widehat{p}_H, \mathbf{0}, \widehat{\Sigma}_H)) \longrightarrow \chi^2_d.$$

Test 3:

$$H_0 : \boldsymbol{\Sigma} \text{ is a diagonal matrix} \quad \text{vs.} \quad \boldsymbol{\Sigma} \text{ is arbitrary.} \tag{27}$$

The above testing problem (27) is important as it tests the uncorrelatedness of the components. In this case the diagonal elements of the matrix $\boldsymbol{\Sigma}$ will be denoted by $\sigma_1^2, \ldots, \sigma_d^2$, i.e. $\boldsymbol{\Sigma} = \text{diag}\left\{\sigma_1^2, \ldots, \sigma_d^2\right\}$. Now we will mention how to compute the MLEs of the unknown parameters p, $\boldsymbol{\mu}$ and $\sigma_1^2, \ldots, \sigma_d^2$, under the null hypothesis. In this case also as before for a given p, we use EM algorithm to compute the MLEs of $\boldsymbol{\mu}$ and $\sigma_1^2, \ldots, \sigma_d^2$, and finally the MLE of p can be obtained by maximizing the profile likelihood function. Now we will describe how to compute the MLEs of $\boldsymbol{\mu}$ and $\sigma_1^2, \ldots, \sigma_d^2$, for a given p, by using the EM algorithm.

We use the following notation for further development. The matrix $\boldsymbol{\Delta}_k$ is a $d \times d$ matrix with all the entries 0, except the (k,k)-th element which is 1. Now under H_0, the 'E-Step' of the EM algorithm can be written as follows:

$$
\begin{aligned}
Q(\boldsymbol{\theta}|\widetilde{\boldsymbol{\theta}}) \quad = \quad & -\frac{n}{2}\left(\sum_{k=1}^{d} \ln \sigma_k^2\right) \\
& -\frac{1}{2}\sum_{k=1}^{d}\frac{1}{\sigma_k^2}\left\{\sum_{i=1}^{n} b_i \boldsymbol{x}_i^T \boldsymbol{\Delta}_k \boldsymbol{x}_i - \sum_{i=1}^{n}\left(\boldsymbol{x}_i^T \boldsymbol{\Delta}_k \boldsymbol{\mu} + \boldsymbol{\mu}^T \boldsymbol{\Delta}_k \boldsymbol{x}_i\right) + \boldsymbol{\mu}^T \boldsymbol{\Delta}_k \boldsymbol{\mu} \sum_{i=1}^{n} a_i\right\} \quad (28)
\end{aligned}
$$

The 'M-Step' involves maximizing (28) with respect to $\boldsymbol{\mu}$, $\sigma_1^2, \ldots, \sigma_d^2$ to obtain updated $\boldsymbol{\mu}$, $\sigma_1^2, \ldots, \sigma_d^2$, say $\overline{\boldsymbol{\mu}}$, $\overline{\boldsymbol{\sigma}_1^2}, \ldots, \overline{\boldsymbol{\sigma}_d^2}$, respectively. From (28), we obtain

$$
\overline{\boldsymbol{\mu}} = \frac{1}{\sum_{j=1}^{n} a_j} \sum_{i=1}^{n} \boldsymbol{x}_i
$$

and

$$
\overline{\boldsymbol{\sigma}_k^2} = \frac{1}{n}\left[\sum_{i=1}^{n} b_i \boldsymbol{x}_i \boldsymbol{x}_i^T - \sum_{i=1}^{n}\left(\overline{\boldsymbol{\mu}} \boldsymbol{x}_i^T + \boldsymbol{x}_i \overline{\boldsymbol{\mu}}^T\right) + \left(\sum_{i=1}^{n} a_i\right) \overline{\boldsymbol{\mu}}\, \overline{\boldsymbol{\mu}}^T\right]_k .
$$

Here for a square matrix $\boldsymbol{A}$, $\boldsymbol{A}_k$ denotes the k-th diagonal element of the matrix $\boldsymbol{A}$. Under the null hypothesis

$$
T_3 = 2(l(\widehat{p}, \widehat{\boldsymbol{\mu}}, \widehat{\boldsymbol{\Sigma}}) - l(\widehat{p}_H, \widehat{\boldsymbol{\mu}}_H, \text{diag}\{\widehat{\sigma}_{1H}^2, \ldots, \widehat{\sigma}_{dH}^2\})) \longrightarrow \chi^2_{d(d+1)/2}.
$$

5 Simulations and Data Analysis

In this section we perform some Monte Carlo simulations to show how the proposed EM algorithm performs and we perform the analyses of two data sets analysis to show how the proposed model and the methods can be used in practice.

5.1 Simulation Results

For simulation purposes we have used the following sample size and the parameter values;

$$n = 100, \quad d = 4, \quad p = 0.50, \quad p = 0.75, \quad \boldsymbol{\mu} = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}, \quad \boldsymbol{\Sigma} = \begin{bmatrix} 2 & 2 & 1 & 0 \\ 2 & 3 & 2 & 0 \\ 1 & 2 & 3 & 2 \\ 0 & 1 & 2 & 2 \end{bmatrix}.$$

Now to show the effectiveness of the EM algorithm we have considered both the cases namely when (a) p is known and (b) p is unknown. We have generated samples from the above configuration and computed the MLEs of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ using EM algorithm. In all the cases we have used the sample mean and the sample variance covariance matrix as the initial guesses of the EM algorithm. We replicate the process 1000 times and report the average estimates and the associated mean squared errors (MSEs). For known p, the results are reported in Tables 1 and 3 and for unknown p, the results are reported in Tables 2 and 4. In each box of a table, the first figure, second figure and the third figure represent the true value, the average estimate and the corresponding MSE, respectively.

It is clear that the performances of the proposed EM algorithm are quite satisfactory. It is observed that the sample mean and the sample variance covariance matrix can be used as good initial guesses of the EM algorithm. In all the cases considered it is observed that the EM algorithm converges within 30 iterations, hence it can be used in practice quite conveniently. Further, it is observed that the profile likelihood method is also quite effective in estimating p, when it is unknown.

In this section we present the analysis of two data sets namely (i) one simulated data set and (ii) one real data set mainly to illustrate how the proposed EM algorithm and the other testing procedures can be used in practice.

Table 1: Average estimates and MSEs of $\widehat{\boldsymbol{\mu}}$ and $\widehat{\boldsymbol{\Sigma}}$, when $p = 0.5$ and it is known

$\boldsymbol{\mu}$	0.0000 0.0053 (0.0984)	0.0000 0.0047 (0.1213)	1.0000 1.0097 (0.1430)	1.0000 1.0055 (0.1237)
σ_{ij}	2.0000 2.0024 (0.3299)	2.0000 1.9984 (0.3602)	1.0000 0.9942 (0.2879)	0.0000 -0.0060 (0.2262)
	2.0000 1.9984 (0.3602)	3.0000 2.9922 (0.4932)	2.0000 1.9907 (0.4130)	1.0000 0.9907 (0.3092)
	1.0000 0.9942 (0.2879)	2.0000 1.9907 (0.4130)	3.0000 2.9757 (0.5350)	2.0000 1.9823 (0.4054)
	0.0000 -0.0060 (0.2262)	1.0000 0.9907 (0.3092)	2.0000 1.9823 (0.4054)	2.0000 1.9859 (0.3675)

Table 2: Average estimates and MSEs of $\widehat{\boldsymbol{\mu}}$, $\widehat{\boldsymbol{\Sigma}}$ and $\widehat{p}$, when $p = 0.5$ and it is unknown

$\boldsymbol{\mu}$	0.0000 -0.0067 (0.1029)	0.0000 -0.0079 (0.1255)	1.0000 1.0094 (0.1553)	1.0000 1.0122 (0.1390)
σ_{ij}	2.0000 2.0105 (0.3579)	2.0000 2.0103 (0.3964)	1.0000 1.0055 (0.3165)	0.0000 0.0011 (0.2377)
	2.0000 2.0103 (0.3964)	3.0000 3.0194 (0.5420)	2.0000 2.0095 (0.4443)	1.0000 1.0091 (0.3233)
	1.0000 1.0055 (0.3165)	2.0000 2.0095 (0.4443)	3.0000 2.9987 (0.5577)	2.0000 2.0006 (0.4161)
	0.0000 0.0011 (0.2377)	1.0000 1.0091 (0.3233)	2.0000 2.0006 (0.4161)	2.0000 2.0058 (0.3827)
2*p	0.5000 0.5068 (0.0433)			

Table 3: Average estimates and MSEs of $\widehat{\boldsymbol{\mu}}$ and $\widehat{\boldsymbol{\Sigma}}$, when $p = 0.75$ and it is known

$\boldsymbol{\mu}$	0.0000 0.0057 (0.1205)	0.0000 0.0046 (0.1481)	1.0000 1.0118 (0.1605)	1.0000 1.0068 (0.1350)
σ_{ij}	2.0000 2.0015 (0.3126)	2.0000 1.9969 (0.3416)	1.0000 0.9923 (0.2720)	0.0000 -0.0065 (0.2099)
	2.0000 1.9969 (0.3416)	3.0000 2.9906 (0.4645)	2.0000 1.9887 (0.3875)	1.0000 0.9898 (0.2907)
	1.0000 0.9923 (0.2720)	2.0000 1.9897 (0.3875)	3.0000 2.9778 (0.4848)	2.0000 1.9864 (0.3864)
	0.0000 -0.0065 (0.2099)	1.0000 0.9898 (0.2907)	2.0000 1.9864 (0.3684)	2.0000 1.9903 (0.3320)

Table 4: Average estimates and MSEs of $\widehat{\boldsymbol{\mu}}$, $\widehat{\boldsymbol{\Sigma}}$ and $\widehat{p}$, when $p = 0.75$ and it is unknown

$\boldsymbol{\mu}$	0.0000 0.0047 (0.1392)	0.0000 0.0035 (0.1704)	1.0000 1.0141 (0.1938)	1.0000 1.0093 (0.1615)
σ_{ij}	2.0000 2.0183 (0.3793)	2.0000 2.0258 (0.4260)	1.0000 1.0263 (0.3453)	0.0000 0.0121 (0.2620)
	2.0000 2.0258 (0.4260)	3.0000 3.0351 (0.5807)	2.0000 2.0272 (0.4842)	1.0000 1.0088 (0.3448)
	1.0000 1.0263 (0.3453)	2.0000 2.0272 (0.4842)	3.0000 3.0120 (0.5897)	2.0000 1.9961 (0.4256)
	0.0000 0.0121 (0.2620)	1.0000 1.0088 (0.3448)	2.0000 1.9961 (0.4256)	2.0000 1.9892 (0.3726)
2*p	0.7500 0.7575 (0.0446)			

5.2 Simulated Data Set

We have generated a data set using the Algorithm 1 as suggested in Section 2, with the following specification:

$$n = 100, \quad d = 4, \quad p = 0.50, \quad \boldsymbol{\mu} = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}, \quad \boldsymbol{\Sigma} = \begin{bmatrix} 2 & 2 & 1 & 0 \\ 2 & 3 & 2 & 0 \\ 1 & 2 & 3 & 2 \\ 0 & 1 & 2 & 2 \end{bmatrix}.$$

It is available in http://home.iitk.ac.in/∼kundu/fort.76. We present some basic statistics of the data set. The sample mean vector, and the sample variance covariance matrix are as follows:

$$\bar{\boldsymbol{x}} = \begin{bmatrix} 0.1489 \\ 0.1323 \\ 1.9803 \\ 1.9246 \end{bmatrix} \quad \text{and} \quad \boldsymbol{S} = \begin{bmatrix} 3.4240 & 3.1869 & 1.6040 & -0.2736 \\ 3.1869 & 5.3792 & 4.1521 & 2.3513 \\ 1.6040 & 4.1521 & 7.0360 & 5.5817 \\ -0.2736 & 2.3513 & 5.5817 & 6.1312 \end{bmatrix}. \tag{29}$$

We start the EM algorithm for each p with the above initial guesses. The profile log-likelihood function is plotted in Figure 2. Finally, the MLEs of the unknown parameters are obtained

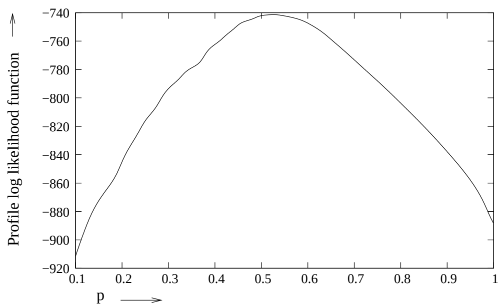

Figure 2: The profile log-likelihood function.

as follows:

$$\widehat{p} = 0.5260, \quad \widehat{\boldsymbol{\mu}} = \begin{bmatrix} 0.0674 \\ 0.0636 \\ 0.9978 \\ 0.9674 \end{bmatrix}, \quad \widehat{\boldsymbol{\Sigma}} = \begin{bmatrix} 1.6511 & 1.5361 & 0.6973 & -0.2077 \\ 1.5361 & 2.5957 & 1.9258 & 1.0580 \\ 0.6973 & 1.9258 & 2.0314 & 1.3745 \\ -0.2077 & 1.0580 & 1.3745 & 1.6847 \end{bmatrix},$$

and the associated log-likelihood value is -741.347. It may be mentioned for each p, the EM algorithm is continued for 20 iterations, and the log-likelihood value (21) is calculated based on the first 50 terms of the infinite series. The program is written in FORTRAN-77, and it is available in http://home.iitk.ac.in/∼kundu/mv-geo-sn-em-punknown-data.for.

For illustrative purposes, we would like to perform the test:

$$H_0 : p = 1, \quad \text{vs.} \quad H_1 : p < 1.$$

Under H_0, the MLEs of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ become $\bar{\boldsymbol{x}}$ and $\boldsymbol{S}$, respectively, as given in (29), and the associated log-likelihood value is -887.852. Therefore, the value of the test statistic $T_1 = 293.01$, and the associated p value is less than 0.00001. Hence, we reject H_0. Next we consider the following testing problem

$$H_0 : \boldsymbol{\mu} = \mathbf{0}, \quad \text{vs.} \quad H_1 : \boldsymbol{\mu} \neq \mathbf{0}.$$

In this case under H_0, the MLEs of p and $\boldsymbol{\Sigma}$ are as follows

$$\widehat{p}_H = 0.459, \qquad \widehat{\boldsymbol{\Sigma}}_H = \begin{bmatrix} 1.9295 & 1.7953 & 1.0631 & 0.0073 \\ 1.7953 & 3.0215 & 2.4714 & 1.4589 \\ 1.0631 & 2.4713 & 6.1348 & 5.2589 \\ 0.0073 & 1.4589 & 5.2589 & 5.5066 \end{bmatrix},$$

and the associated log-likelihood value is -917.674. In this case the value of the test statistics $T_2 = 352.654$. Since the associated p value is less than 0.00001, we reject the null hypothesis. Finally we consider the testing problem:

$$H_0 : \boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & 0 & 0 & 0 \\ 0 & \sigma_2^2 & 0 & 0 \\ 0 & 0 & \sigma_3^2 & 0 \\ 0 & 0 & 0 & \sigma_4^2 \end{bmatrix} \quad \text{vs.} \quad H_1 : \boldsymbol{\Sigma} \text{ is arbitrary.}$$

In this case under H_0, the MLEs of the unknown parameters are as follows:

$$\widehat{p}_H = 0.585, \quad \widehat{\boldsymbol{\mu}}_H = (0.0479, 0.2502, 1.0202, 0.9967)^T$$

and

$$\widehat{\sigma}^2_{1H} = 1.5695, \quad \widehat{\sigma}^2_{2H} = 2.2356, \quad \widehat{\sigma}^2_{3H} = 1.1014, \quad \widehat{\sigma}^2_{4H} = 0.7652.$$

The associated log-likelihood value is -1036.80. The value of $T_3 = 590.91$. In this case also we reject H_0, as the associated p values is less than 0.00001.

5.3 Stiffness Data Set

In this section we present the analysis of a real data set to show how the proposed model and the methodologies work in practice. The data set represents the four different measurements of stiffness, x_1, x_2, x_3, x_4 of 'Shock' and 'Vibration' of each of 30 boards. The first measurement (Shock) involves sending a shock wave down the board and the second measurement (Vibration) is determined while vibrating the board. The last two measurements are obtained from static tests. The data set is available in Johnson and Wichern [7]. For easy reference it is presented in Table 5. Since all the entries of the data set are non-negative, if we want to the fit the multivariate skew normal distribution to this data set, the MLEs of the unknown parameters may not exist. In fact we have tried to fit univariate skew-normal distribution to x_1 and it is observed that the likelihood function is an increasing function of the 'tilt' parameter for fixed location and scale parameters. Therefore, the MLEs do not exist in this case. It is expected the same phenomenon even for skew-t distribution for large values of the degrees of freedom.

Before progressing further we have divided all the measurements by 100, and it is not going to make any difference in the inferential procedure. The sample mean vector and the sample variance covariance matrix of the transformed data are

$$\bar{\boldsymbol{x}} = \begin{bmatrix} 19.0610 \\ 17.4953 \\ 15.0790 \\ 17.2497 \end{bmatrix} \quad \text{and} \quad \boldsymbol{S} = \begin{bmatrix} 10.2096 & 9.1460 & 8.4590 & 9.1090 \\ 9.1460 & 9.8126 & 7.3791 & 7.8362 \\ 8.4590 & 7.3791 & 8.9212 & 8.7615 \\ 9.1090 & 7.8362 & 8.7615 & 10.0754. \end{bmatrix} \tag{30}$$

Based on the EM algorithm and using the profile likelihood method, we obtain the MLEs of the unknown parameters as follows:

$$\widehat{p} = 0.9640, \quad \widehat{\boldsymbol{\mu}} = \begin{bmatrix} 18.2409 \\ 16.7594 \\ 14.4459 \\ 16.4719 \end{bmatrix}, \quad \widehat{\boldsymbol{\Sigma}} = \begin{bmatrix} 7.6625 & 6.4895 & 6.1284 & 7.4978 \\ 6.4895 & 7.0296 & 4.9941 & 6.0798 \\ 6.1284 & 4.9941 & 6.7659 & 7.1631 \\ 7.4978 & 6.0798 & 7.1631 & 9.1815. \end{bmatrix},$$

and the associated log-likelihood value is -271.969. Now to check whether the proposed

Table 5: Four different stiffness measurements of 30 boards

No.	x_1	x_2	x_3	x_4	No.	x_1	x_2	x_3	x_4
1	1889	1651	1561	1778	2	2403	2048	2087	2197
3	2119	1700	1815	2222	4	1645	1627	1110	1533
5	1976	1916	1614	1883	6	1712	1712	1439	1546
7	1943	1685	1271	1671	8	2104	1820	1717	1874
9	2983	2794	2412	2581	10	1745	1600	1348	1508
11	1710	1591	1518	1667	12	2046	1907	1627	1898
13	1840	1841	1595	1741	14	1867	1685	1493	1678
15	1859	1649	1389	1714	16	1954	2149	1180	1281
17	1325	1170	1002	1176	18	1419	1371	1251	1308
19	1828	1634	1602	1755	20	1725	1594	1313	1646
21	2276	2189	1547	2111	22	1899	1614	1422	1477
23	1633	1513	1290	1516	24	2061	1867	1646	2037
25	1856	1493	1356	1533	26	1727	1412	1238	1469
27	2168	1896	1701	1834	28	1655	1675	1414	1597
29	2326	2301	2065	2234	30	1490	1382	1214	1284

MGSN distribution provides a better fit than the multivariate normal distribution or not, we perform the following test:

$$H_0 : p = 1, \quad \text{vs.} \quad H_1 : p < 1.$$

Under H_0, the MLEs of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are provided in (30), and the associated log-likelihood value is -277.761. Therefore, the value of the test statistic $T_1 = 11.584$, and the associated p value is 0.0000025. Hence, we reject the null hypothesis, and it indicates that the proposed MGSN distribution provides a better fit than the multivariate normal distribution to the given stiffness data set. AIC also prefers MGSN distribution than the multivariate normal distribution for this data set.

6 Conclusion

In this paper we have discusses different properties of the MGSN distribution in details. Different characterization results and dependence properties have been established. The d-dimensional MGSN distribution has $d + 1 + d(d + 1)/2$ unknown parameters. We have proposed to use EM algorithm and the profile likelihood method to compute the MLEs of the unknown parameters, and it is observed that the proposed algorithm can be implemented very easily. We have discussed some testing of hypothesis problems also. Two data sets have been analyzed to show the effectiveness of the proposed methods, and it is observed that for the real 'stiffness' data set MGSN provides a better fit than the multivariate normal distribution. Hence, this model can be used as an alternative to Azzalini's multivariate skew normal distribution.

Acknowledgements

The author would like to thank Prof. Dimitris Karlis from the Athens University of Economics and Business, Greece, for his helpful comments. The author is really thankful to two reviewers and the associate editor for their careful reading and providing constructive comments.

References

[1] Anderson, T.W. (1984), *An introduction to multivariate statistical analysis*, 2nd-edition, John-Wiley & Sons, New York.

[2] Arnold, B.C. and Beaver, R.J. (2000), "Hidden truncation models", *Sankhya*, Ser. A, vol. 62, 23 - 35.

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

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

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

[6] Dempster, A.P., Laird, N.M. and Rubin, D.B. (1977), "Maximum likelihood from incomplete data via EM algorithm", *Journal of the Royal Statistical Society*, Ser. B., vol. 39, 1 - 38.

[7] Johnson, R. A. and Wichern, D.W. (2013), *Applied multivariate Statistical Analysis*, 6-th edition, Pearson New International Edition, New York, USA.

[8] Karlin, S. and Rinott, Y. (1980), "Classes of orderings of measures and related correlation inequalities, I, Multivariate totally positive distributions", *Journal of Multivariate Analysis*, vol. 10, 467 - 498.

[9] Kundu, D. (2014), "Geometric skew normal distribution", *Sankhya*, Ser. B, vol. 76, 167 - 189.

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

[11] Mardia, K. V. (1974), "Applications of some measures of multivariate skewness and kurtosis in testing normality and robustness studies", *Sankhya*, Ser. B, vol. 36, 115 – 128.

[12] Rao, C.R. (1973), *Linear Statistical Inference and its Applications*, 2nd. edition, Wiley and Sons, New York.

[13] Self, S.G. and Liang, K-L. (1987), "Asymptotic properties of the maximum likelihood estimators and likelihood ratio test under non-standard conditions", *Journal of the American Statistical Association*, vol. 82, 605 - 610.

[14] Shaked, M. and Shantikumar, J.G. (1994), *Stochastic Orders and their Applications*, Acadmic Press, London, UK.