Mathematics

http://www.aimspress.com/journal/Math

AIMS Mathematics, 7(7): 12390–12414.
DOI: 10.3934/math.2022688
Received: 07 March 2022
Revised: 20 April 2022
Accepted: 22 April 2022
Published: 26 April 2022

Research article

Family of extended mean mixtures of multivariate normal distributions: Properties, inference and applications

Guangshuai Zhou and Chuancun Yin*

School of Statistics and Data Science, Qufu Normal University, Qufu 273165, China

* **Correspondence:** Email: ccyin@qfnu.edu.cn; Tel: +(086)5374453221.

Abstract: A new class of skewed distributions, with a matrix skewness parameter, called extended mean mixtures of multivariate normal (EMMN) distributions, is constructed. The family of EMMN distributions includes the SN and MMN distributions as special cases. Some basic properties of this family, such as characteristic function, moment generating function, affine transformation and canonical forms of the distributions are derived. An EM-type algorithm is developed to carry out the maximum likelihood estimation of the parameters. Two special cases of this family are studied in detail. A simulation is carried out to examine the performance of the estimation method, and the flexibility is illustrated by fitting a special case of this family to a real data. Finally, the theoretical formula of the multivariate tail conditional expectation of the EMMN distribution is derived.

Keywords: EM algorithm; matrix skewness parameter; mean mixtures of multivariate normal distribution; multivariate tail conditional expectation; risk measure
Mathematics Subject Classification: 60E07, 60E10, 62H05, 62H10, 62H12

1. Introduction

The multivariate normal distribution has long occupied a central position in statistical theoretical research and has a wide range of applications in practice. One reason is that the approximate distribution of many variables is normal distribution, as known from the central limit theorem, and the other is that a significant feature of normal distribution is its symmetry, which makes the study of related problem mathematically tractable. However, in the fields of physics, biomedicine, finance and insurance, real data often exhibit heavy-tailed and skewed characteristics. For example, the distribution of life expectancy, long-tailed insurance claims, etc. Hence, many researchers search for more flexible distributions for modeling skewed data.

One of the well-known classes of distributions is the class of Normal Mean-Variance Mixtures (NMVM) distributions introduced by Barndorff-Nielsen et al. [11]. According to that, a

p-dimensional random vector $\boldsymbol{Y}$ is said to have a multivariate NMVM distributions if it has the stochastic representation

$$\boldsymbol{Y} \stackrel{d}{=} \boldsymbol{\xi} + \boldsymbol{\delta} U + \sqrt{U}\boldsymbol{Z}, \tag{1.1}$$

where $\stackrel{d}{=}$ stands for equality in distribution, $\boldsymbol{Z} \sim N_p(\boldsymbol{0}, \boldsymbol{\Sigma})$, independently of U. And U is a positive random variable with cumulative distribution function (CDF) $H(\cdot; \boldsymbol{\nu})$, indexed by parameter $\boldsymbol{\nu}$. Specifically, NMVM distributions have the following hierarchial representation:

$$\boldsymbol{Y}|(U = u) \sim N_p(\boldsymbol{\xi} + \boldsymbol{\delta} u, u\boldsymbol{\Sigma}). \tag{1.2}$$

There is a large literature of studies based on the above stochastic representations (1.1). For example, the variance-mean mixture of kotz-type distributions were studied by Arslan [4], and the variance-mean mixture of the multivariate skew normal distributions were studied by Arslan [5]. Jamalizadeh and Balakrishnan [15] explored the the conditional distribution of a multivariate normal mean-variance mixtures (MNMVM) distributions. Naderi et al. [21] introduced a new asymmetric distribution-multivariate normal mean-variance mixture distribution based on Lindley distribution. Pourmousa et al. [24] introduced the multivariate normal mean-variance mixture distribution based on Birnbaum-Saunders (NMVMBS) distribution.

Besides the NMVM distributions, another important class of distributions is the skewed distributions. Azzalini [6] formulated the univariate skew normal (SN) distribution, and Azzalini and Valle [10] extended to the multivariate case. The statistical applications of the multivariate normal distribution were explored by Azzalini and Capitanio [8]. Azzalini and Capitanio [9] introduced a comprehensive survey of this skewed distribution classes including the skew-t (ST) distribution and the skew-elliptical distributions. To establish notation, the probability density function (PDF) and the CDF of the p-dimensional normal distribution, with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$, will be denoted by $\phi_p(\cdot; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ and $\Phi_p(\cdot; \boldsymbol{\mu}, \boldsymbol{\Sigma})$, respectively. Specially, denote $\Phi_p(\cdot)$ as the CDF of the multivariate standard normal distribution, and denote $\phi(\cdot)$, $\Phi(\cdot)$ as the PDF and CDF of the univariate standard normal distribution, respectively. We denote the doubly truncated p-dimensional normal vector $\boldsymbol{X}|(\boldsymbol{a} < \boldsymbol{X} \leq \boldsymbol{b})$ by $\boldsymbol{X} \sim TN_p(\boldsymbol{a}, \boldsymbol{b}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$, and denote

$$Pr(\boldsymbol{a} < \boldsymbol{X} \leq \boldsymbol{b}) = Pr(a_1 < X_1 \leq b_1, \ldots, a_p < X_p \leq b_p)$$

by $\bar{\Phi}_p(\boldsymbol{a}, \boldsymbol{b}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$, for any $\boldsymbol{a} < \boldsymbol{b}$ ($\boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^p$).

From [9, 10], a p -dimensional random vector $\boldsymbol{Y}$ is said to have a multivariate skewed normal (SN) distribution, if it has the PDF

$$f(\boldsymbol{y}) = 2\phi_p(\boldsymbol{y}; \boldsymbol{\xi}, \boldsymbol{\Omega})\Phi(\boldsymbol{\alpha}^\top \boldsymbol{\omega}^{-1}(\boldsymbol{y} - \boldsymbol{\xi})), \quad \boldsymbol{y} \in \mathbb{R}^p, \tag{1.3}$$

and it has equivalent form

$$f(\boldsymbol{y}) = 2\phi_p(\boldsymbol{y}; \boldsymbol{\xi}, \boldsymbol{\Omega})\Phi\left(\frac{\boldsymbol{\delta}^\top \bar{\boldsymbol{\Omega}} \omega^{-1}(\boldsymbol{y} - \boldsymbol{\xi})}{\sqrt{1 - \boldsymbol{\delta}^\top \bar{\boldsymbol{\Omega}}^{-1} \boldsymbol{\delta}}}\right), \quad \boldsymbol{y} \in \mathbb{R}^p, \tag{1.4}$$

where $\boldsymbol{\xi} \in \mathbb{R}^p$ is the location parameter, $\boldsymbol{\alpha} \in \mathbb{R}^p$ is the slant parameter and $\boldsymbol{\delta} = (1 + \boldsymbol{\alpha}^\top \bar{\boldsymbol{\Omega}} \boldsymbol{\alpha})^{-1/2} \bar{\boldsymbol{\Omega}} \boldsymbol{\alpha}$ is the skewness parameter vector, with $-\boldsymbol{1}_p < \boldsymbol{\delta} = (\delta_1, \ldots, \delta_p)^\top < \boldsymbol{1}_p$. Here, $\boldsymbol{1}_p$ is p-dimensional vector with all elements 1. The matrix $\boldsymbol{\omega} = \operatorname{diag}(\omega_1, \ldots, \omega_p) = (\boldsymbol{\Omega} \odot \boldsymbol{I}_p)^{1/2} > 0$ is a diagonal matrix, where $\boldsymbol{I}_p$ is identity matrix of size p and $\odot$ denotes the Hadamard product. Scale matrix $\boldsymbol{\Omega}$ is a positive definite matrix and $\bar{\boldsymbol{\Omega}}$ is a positive definite $p \times p$ correlation matrix, with $\boldsymbol{\Omega} = \boldsymbol{\omega}\bar{\boldsymbol{\Omega}}\boldsymbol{\omega}$. The random variable $\boldsymbol{Y}$ is usually denoted as $\boldsymbol{Y} \sim SN_p(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\alpha})$ and $\boldsymbol{Y} \sim SN_p(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\delta})$ if $\boldsymbol{Y}$ has PDF (1.3) and (1.4), respectively. It is neccessary to mention that the SN distribution has the stochastic representation:

$$\boldsymbol{Y} \stackrel{d}{=} \boldsymbol{\xi} + \boldsymbol{\omega}(\boldsymbol{\delta} U + \boldsymbol{Z}), \tag{1.5}$$

where $\boldsymbol{\xi} \in \mathbb{R}^p$, $\boldsymbol{Z} \sim N_p(\boldsymbol{0}, \bar{\boldsymbol{\Omega}} - \boldsymbol{\delta}\boldsymbol{\delta}^\top)$, univariate random variable $U \sim TN_1(0, +\infty; 0, 1)$, independently of $\boldsymbol{Z}$. Specifically, the SN distribution has the following hierarchial representation:

$$\boldsymbol{Y}|(U = u) \sim N_p(\boldsymbol{\xi} + \boldsymbol{\omega}\boldsymbol{\delta} u, \boldsymbol{\Omega} - \boldsymbol{\omega}\boldsymbol{\delta}\boldsymbol{\delta}^\top\boldsymbol{\omega}). \tag{1.6}$$

From (1.6) the SN distribution is a mean mixtures of the normal distribution, as stated in Negarestani et al. [22], the SN distribution cannot be obtained through the NMVM distributions.

For the last three decades, a growing body of literature related to SN distribution has been widely explored, as witnessed by the numerous applications in financial markets, such as financial risk measurement and portfolio optimization. For example in Bernardi et al. [12] and Mousavi et al. [20]. Also there is considerable amount of literatures of extensions and alternative formulations related to skewed distributions, such as the hierarchical skew-normal (HSN) distribution in Liseo and Loperfido [19], the closed skew-normal (CSN) distribution in González-Farías et al. [14], the fundamental skew-normal (FUSN) distribution in Arellano-Valle and Genton [3], the pseudo normal (PN) family of distributions in Ogasawara [23]. Azzalini [7] provided an overview on the progeny of the skew-normal family.

According to Arellano-Valle and Azzalini [2], these distributions (HSN, CSN, FUSN) can be seen as the special cases of unified skew normal distribution (SUN). From [2, 9], a d-dimensional random vector $\boldsymbol{Y}$ is said to have a SUN distribution, denoted by $SUN_{d,m}(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\Delta}, \boldsymbol{\gamma}, \boldsymbol{\Gamma})$, if its PDF is

$$f(\boldsymbol{y}) = \phi_d(\boldsymbol{y}; \boldsymbol{\xi}, \boldsymbol{\Omega}) \frac{\Phi_m(\boldsymbol{\gamma} + \boldsymbol{\Delta}^\top\bar{\boldsymbol{\Omega}}^{-1}\boldsymbol{\omega}^{-1}(\boldsymbol{y} - \boldsymbol{\xi}); \boldsymbol{0}, \boldsymbol{\Gamma} - \boldsymbol{\Delta}^\top\bar{\boldsymbol{\Omega}}^{-1}\boldsymbol{\Delta})}{\Phi_m(\boldsymbol{\gamma}; \boldsymbol{0}, \boldsymbol{\Gamma})}, \quad \boldsymbol{y} \in \mathbb{R}^d, \quad \boldsymbol{\gamma} \in \mathbb{R}^m, \tag{1.7}$$

which has the stochastic representation

$$\boldsymbol{Y} \stackrel{d}{=} \boldsymbol{\xi} + \boldsymbol{\omega}\left(\boldsymbol{\Delta}\boldsymbol{\Gamma}^{-1}\boldsymbol{U} + \boldsymbol{Z}\right), \tag{1.8}$$

where $\boldsymbol{\xi}$, $\boldsymbol{\omega}$ are as in (1.4), $\boldsymbol{Z} \sim N_d(\boldsymbol{0}, \bar{\boldsymbol{\Omega}} - \boldsymbol{\Delta}\boldsymbol{\Gamma}^{-1}\boldsymbol{\Delta}^\top)$, $\boldsymbol{\Omega} = \boldsymbol{\omega}\bar{\boldsymbol{\Omega}}\boldsymbol{\omega} > 0$ is a positive definite matrix, $\bar{\boldsymbol{\Omega}}$ and $\boldsymbol{\Gamma}$ are $d \times d$ and $m \times m$ full-rank correlation matrices, respectively. $\boldsymbol{\Delta} = (\delta_{ij})$ is $d \times m$ matrix, with $-1 < \delta_{ij} < 1$, for $i \in \{1, \ldots, d\}$, $j \in \{1, \ldots, m\}$. m-dimensional random vector $\boldsymbol{U} \sim TN_m(-\boldsymbol{\gamma}, +\infty; \boldsymbol{0}, \boldsymbol{\Gamma})$, independently of $\boldsymbol{Z}$. When $\boldsymbol{\gamma} = \boldsymbol{0}$, $m = 1$, the PDF in (1.7) reduces to (1.4), then the SUN distribution degenerates to the SN distribution. When $m = 1$ the SUN distribution degenerates to the extended skewed normal (ESN) distribution.

Recently, Negarestani et al. [22] proposed a new class of mean mixtures of univariate normal distributions by replacing the standard half normal random variable U in the stochastic representation (1.5) of the SN distribution with a general random variable, and analyzed its basic properties and its flexibility in fitting skewed data. Abdi et al. [1] extended the mean mixtures of univariate normal distributions in [22] to multivariate cases, and proposed the mean mixtures of multivariate normal (MMN) distributions, and studied their flexibility and calculated a variety of different skewness parameters.

Following this way, in this paper, we prposed the extended mean mixtures of multivariate normal (EMMN) distributions by considering the stochastic representation of the SUN distribution (when $\boldsymbol{\gamma} = \mathbf{0}, \boldsymbol{\Gamma} = \boldsymbol{I}_m$) and studying the case when the multivariate random variable $\boldsymbol{U}$ in (1.8) takes a general multivariate random variable.

The remainder of this paper is laid out as follows. In Section 2, we give the definition of the EMMN distributions. Some basic properties are studied in Section 3. In Section 4, an Expectation Maximization (EM) algorithm is developed for computing the maximum likelihood estimator of the EMMN distributions. Two sepecial cases, the EMMNG and EMMNE distributions, are studied in detail in Section 5. In Section 6, a simulation study is carried out for evaluating the performance of the EM-based estimators. In Section 7, to illustrate the applicability of the proposed distributions, a real data sets are analyzed. The multivariate tail conditional expectation (MTCE) is derived for the EMMN family in Section 8. Finally, in Section 9, some concluding remarks are made.

2. EMMN distribution

Definition 1. *A d-dimensional random vector $\boldsymbol{Y}$ has an extended mean mixtures of multivariate normal (EMMN) distribution, denoted by $\boldsymbol{Y} \sim EMMN_{d,m}(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\Delta}; H)$, if and only if it has the stochastic representation*

$$\boldsymbol{Y} \overset{d}{=} \boldsymbol{\xi} + \omega(\boldsymbol{\Delta}\boldsymbol{U} + \boldsymbol{Z}), \tag{2.1}$$

where $\boldsymbol{\Omega} = \omega\bar{\boldsymbol{\Omega}}\omega > 0$ is a positive definite matrix, and $\boldsymbol{\xi}$, ω, $\boldsymbol{\Delta}$ are as in stochastic representation (1.8). $\boldsymbol{Z} \sim N_d(\mathbf{0}, \bar{\boldsymbol{\Omega}} - \boldsymbol{\Delta}\boldsymbol{\Delta}^{\mathsf{T}})$, $\boldsymbol{U}$ is an arbitrary m-dimensional random variable whose components are independent of each other and whose CDF is denoted as $H(\cdot; \boldsymbol{\nu})$, indexed by the parameter $\boldsymbol{\nu} = (\nu_1, \dots, \nu_k)$, and $\boldsymbol{U}$ is independent of $\boldsymbol{Z}$.

If $\boldsymbol{U}$ has a PDF $h(\cdot; \boldsymbol{\nu})$, an integral form of the PDF of $\boldsymbol{Y} \sim EMMN_{d,m}(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\Delta}; H)$ can be obtained as

$$\begin{aligned} f_{EMMN_{d,m}}(\boldsymbol{y}; \boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\Delta}, \boldsymbol{\nu}) &= \int_{\mathbb{R}^m} \phi_d\left(\boldsymbol{y}; \boldsymbol{\xi} + \omega\boldsymbol{\Delta}\boldsymbol{u}, \boldsymbol{\Omega} - \omega\boldsymbol{\Delta}\boldsymbol{\Delta}^{\mathsf{T}}\omega\right) \mathrm{d}H(\boldsymbol{u}; \boldsymbol{\nu}) \\ &= \int_{\mathbb{R}^m} \phi_d\left(\boldsymbol{y}; \boldsymbol{\xi} + \omega\boldsymbol{\Delta}\boldsymbol{u}, \boldsymbol{\Omega} - \omega\boldsymbol{\Delta}\boldsymbol{\Delta}^{\mathsf{T}}\omega\right) h(\boldsymbol{u}; \boldsymbol{\nu})\, \mathrm{d}\boldsymbol{u}, \quad \boldsymbol{y} \in \mathbb{R}^d. \end{aligned} \tag{2.2}$$

Noting that, in (2.1), when $m = 1$, the EMMN distribution has the same stochastic representation as the MMN distribution in [1], which indicates the MMN distibution is a special case of the EMMN distribution. When $m = 1$ and $\boldsymbol{U}$ follows univariate standard half normal distribution, the EMMN distribution degenerates to the SN distribution. The hierarchical representation for the EMMN distribution is

$$\boldsymbol{Y}|(\boldsymbol{U} = \boldsymbol{u}) \sim N_d(\boldsymbol{\xi} + \omega\boldsymbol{\Delta}\boldsymbol{u}, \boldsymbol{\Omega} - \omega\boldsymbol{\Delta}\boldsymbol{\Delta}^{\mathsf{T}}\omega), \qquad \boldsymbol{U} \sim H(\cdot; \boldsymbol{\nu}). \tag{2.3}$$

From (2.3), for EMMN distributions, just the mean parameter is mixed with arbitrary multivariate random variable $\boldsymbol{U}$, but just as shown in [1, 22], EMMN distributions cannot be obtained from the

Normal Mean-Variance Mixtures (NMVM) family either. Table 1 shows the EMMN, MMN, and SN distributions.

Table 1. Summary of the EMMN, MMN, and SN distributions.

Distribution	EMMN	MMN	SN
Notation	$\boldsymbol{Y} \sim EMMN_{p,m}(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\Delta}; H)$	$\boldsymbol{Y} \sim MMN_p(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\delta}; H)$	$\boldsymbol{Y} \sim SN_p(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\delta})$
Density	$\int_{\mathbb{R}^m} \phi_p\left(\boldsymbol{y}; \boldsymbol{\xi} + \omega\boldsymbol{\Delta}\boldsymbol{u}, \boldsymbol{\Omega} - \omega\boldsymbol{\Delta}\boldsymbol{\Delta}^\top\omega\right) h(\boldsymbol{u}; \boldsymbol{\nu})\, d\boldsymbol{u}$	$\int_{\mathbb{R}} \phi_d\left(\boldsymbol{y}; \boldsymbol{\xi} + \omega\boldsymbol{\delta}u, \boldsymbol{\Omega} - \omega\boldsymbol{\delta}\boldsymbol{\delta}^\top\omega\right) h(u; \boldsymbol{\nu})\, du$	$2\phi_p(\boldsymbol{y}; \boldsymbol{\xi}, \boldsymbol{\Omega})\Phi(\alpha^\top\omega^{-1}(\boldsymbol{y} - \boldsymbol{\xi}))$
Stochastic	$\boldsymbol{Y} \overset{d}{=} \boldsymbol{\xi} + \omega(\boldsymbol{\Delta}\boldsymbol{U} + \boldsymbol{Z})$	$\boldsymbol{Y} \overset{d}{=} \boldsymbol{\xi} + \omega(\boldsymbol{\delta}U + \boldsymbol{Z})$	$\boldsymbol{Y} \overset{d}{=} \boldsymbol{\xi} + \omega(\boldsymbol{\delta}U + \boldsymbol{Z})$
representation	$\boldsymbol{U} : m \times 1 \sim h(\boldsymbol{u}; \boldsymbol{\nu}), \boldsymbol{\nu} \in \mathbb{R}^k$	$U \sim h(u; \boldsymbol{\nu}), \boldsymbol{\nu} \in \mathbb{R}^k$	$U \overset{d}{=} \lvert U_0 \rvert, U_0 \sim N(0, 1)$
	$\boldsymbol{Z} \sim N_p(\boldsymbol{0}, \bar{\boldsymbol{\Omega}} - \boldsymbol{\Delta}\boldsymbol{\Delta}^\top)$	$\boldsymbol{Z} \sim N_p(\boldsymbol{0}, \bar{\boldsymbol{\Omega}} - \boldsymbol{\delta}\boldsymbol{\delta}^\top)$	$\boldsymbol{Z} \sim N_p(\boldsymbol{0}, \bar{\boldsymbol{\Omega}} - \boldsymbol{\delta}\boldsymbol{\delta}^\top)$
Mean	$\boldsymbol{\xi} + \omega\boldsymbol{\Delta}E(\boldsymbol{U})$	$\boldsymbol{\xi} + \omega\boldsymbol{\delta}E(U)$	$\boldsymbol{\xi} + \sqrt{\frac{2}{\pi}}\omega\boldsymbol{\delta}$
Covariance matrix	$\boldsymbol{\Omega} + \omega\boldsymbol{\Delta}\left[var(\boldsymbol{U}) - \boldsymbol{I}_m\right]\boldsymbol{\Delta}^\top\omega$	$\boldsymbol{\Omega} + \omega\boldsymbol{\delta}\left[var(U) - 1\right]\boldsymbol{\delta}^\top\omega$	$\boldsymbol{\Omega} - \frac{2}{\pi}\omega\boldsymbol{\delta}\boldsymbol{\delta}^\top\omega$
Skewness Parameter	$\boldsymbol{\Delta} \in \mathbb{R}^{p\times m}$	$\boldsymbol{\delta} \in \mathbb{R}^p$	$\boldsymbol{\delta} \in \mathbb{R}^p$
Free Parameters	$p + \frac{1}{2}(1 + p)p + pm + k$	$2p + \frac{1}{2}(1 + p)p + k$	$2p + \frac{1}{2}(1 + p)p$

3. Properties

In this section, we explore some basic properties of the EMMN distribution.

Remark 1. *The normalized EMMN distribution can be constructed through the transformation $\boldsymbol{X} = \boldsymbol{\omega}^{-1}(\boldsymbol{Y} - \boldsymbol{\xi})$, which has the stochastic representation of $\boldsymbol{X} = \boldsymbol{\Delta}\boldsymbol{U} + \boldsymbol{Z}$ with the hierarchial representation $\boldsymbol{X}|(\boldsymbol{U} = \boldsymbol{u}) \sim N_d(\boldsymbol{\Delta}\boldsymbol{u}, \bar{\boldsymbol{\Omega}} - \boldsymbol{\Delta}\boldsymbol{\Delta}^\top)$, $\boldsymbol{U} \sim H(\cdot; \boldsymbol{\nu})$. Denoted by $\boldsymbol{X} \sim EMMN_{d,m}(\boldsymbol{0}, \bar{\boldsymbol{\Omega}}, \boldsymbol{\Delta}; H)$.*

Lemma 1. *If $\boldsymbol{Y} \sim EMMN_{d,m}(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\Delta}; H)$, the characteristic function (CF) and the moment generating function (MGF) of $\boldsymbol{Y}$ are as follows:*

$$C_{\boldsymbol{Y}}(\boldsymbol{t}) = \mathrm{e}^{\mathrm{i}\boldsymbol{t}^\top\boldsymbol{\xi} - \frac{1}{2}\boldsymbol{t}^\top\boldsymbol{\Sigma}_{\boldsymbol{Y}}\boldsymbol{t}} C_{\boldsymbol{U}}(\mathrm{i}\boldsymbol{\Delta}^\top\boldsymbol{\omega}\boldsymbol{t}; \boldsymbol{\nu}), \qquad M_{\boldsymbol{Y}}(\boldsymbol{t}) = \mathrm{e}^{\boldsymbol{t}^\top\boldsymbol{\xi} + \frac{1}{2}\boldsymbol{t}^\top\boldsymbol{\Sigma}_{\boldsymbol{Y}}\boldsymbol{t}} M_{\boldsymbol{U}}(\boldsymbol{\Delta}^\top\boldsymbol{\omega}\boldsymbol{t}; \boldsymbol{\nu}), \tag{3.1}$$

respectively, where $\mathrm{i} = \sqrt{-1}$, $\boldsymbol{\Sigma}_{\boldsymbol{Y}} = \boldsymbol{\Omega} - \omega\boldsymbol{\Delta}\boldsymbol{\Delta}^\top\omega$, and $C_{\boldsymbol{U}}(\cdot; \boldsymbol{\nu}) = C_{\boldsymbol{U}}(\cdot)$, $M_{\boldsymbol{U}}(\cdot; \boldsymbol{\nu}) = M_{\boldsymbol{U}}(\cdot)$ are the CF and MGF of $\boldsymbol{U}$, respectively.

Proof.

$$\begin{aligned} C_{\boldsymbol{Y}}(\boldsymbol{t}) &= E\left[\mathrm{e}^{\mathrm{i}\boldsymbol{t}^\top\boldsymbol{Y}}\right] = E\left[\mathrm{e}^{\mathrm{i}\boldsymbol{t}^\top(\boldsymbol{\xi} + \omega(\boldsymbol{\Delta}\boldsymbol{U} + \boldsymbol{Z}))}\right] \\ &= \mathrm{e}^{\mathrm{i}\boldsymbol{t}^\top\boldsymbol{\xi}} E\left[\mathrm{e}^{\mathrm{i}\boldsymbol{t}^\top\omega\boldsymbol{\Delta}\boldsymbol{U}}\right] E\left[\mathrm{e}^{\mathrm{i}\boldsymbol{t}^\top\omega\boldsymbol{Z}}\right] = \mathrm{e}^{\mathrm{i}\boldsymbol{t}^\top\boldsymbol{\xi} - \frac{1}{2}\boldsymbol{t}^\top\boldsymbol{\Sigma}_{\boldsymbol{Y}}\boldsymbol{t}} C_{\boldsymbol{U}}(\mathrm{i}\boldsymbol{\Delta}^\top\boldsymbol{\omega}\boldsymbol{t}; \boldsymbol{\nu}), \end{aligned}$$

and

$$\begin{aligned} M_{\boldsymbol{Y}}(\boldsymbol{t}) &= E\left[\mathrm{e}^{\boldsymbol{t}^\top\boldsymbol{Y}}\right] = E\left[\mathrm{e}^{\boldsymbol{t}^\top(\boldsymbol{\xi} + \omega(\boldsymbol{\Delta}\boldsymbol{U} + \boldsymbol{Z}))}\right] \\ &= \mathrm{e}^{\boldsymbol{t}^\top\boldsymbol{\xi}} E\left[\mathrm{e}^{\boldsymbol{t}^\top\omega\boldsymbol{\Delta}\boldsymbol{U}}\right] E\left[\mathrm{e}^{\boldsymbol{t}^\top\omega\boldsymbol{Z}}\right] = \mathrm{e}^{\boldsymbol{t}^\top\boldsymbol{\xi} + \frac{1}{2}\boldsymbol{t}^\top\boldsymbol{\Sigma}_{\boldsymbol{Y}}\boldsymbol{t}} M_{\boldsymbol{U}}(\boldsymbol{\Delta}^\top\boldsymbol{\omega}\boldsymbol{t}; \boldsymbol{\nu}). \end{aligned}$$

□

If $\boldsymbol{X} \sim EMMN_{d,m}(\boldsymbol{0}, \bar{\boldsymbol{\Omega}}, \boldsymbol{\Delta}; H)$, then the CF and MGF of $\boldsymbol{X}$ are

$$C_{\boldsymbol{X}}(\boldsymbol{t}) = \mathrm{e}^{\frac{1}{2}\boldsymbol{t}^\top\boldsymbol{\Sigma}_{\boldsymbol{X}}\boldsymbol{t}} C_{\boldsymbol{U}}(\mathrm{i}\boldsymbol{\Delta}^\top\boldsymbol{t}; \boldsymbol{\nu}), \qquad M_{\boldsymbol{X}}(\boldsymbol{t}) = \mathrm{e}^{\frac{1}{2}\boldsymbol{t}^\top\boldsymbol{\Sigma}_{\boldsymbol{X}}\boldsymbol{t}} M_{\boldsymbol{U}}(\boldsymbol{\Delta}^\top\boldsymbol{t}; \boldsymbol{\nu}),$$

respectively, where $\boldsymbol{\Sigma}_{\boldsymbol{X}} = \bar{\boldsymbol{\Omega}} - \boldsymbol{\Delta}\boldsymbol{\Delta}^\top$.

It can be easily concluded that the mean vector and covariance matrix of $EMMN_{d,m}(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\Delta}; H)$ are $E(\boldsymbol{Y}) = \boldsymbol{\xi} + \omega\boldsymbol{\Delta}E(\boldsymbol{U})$ and $Var(\boldsymbol{Y}) = \boldsymbol{\Omega} + \omega\boldsymbol{\Delta}\left[Var(\boldsymbol{U}) - \boldsymbol{I}_m\right]\boldsymbol{\Delta}^\top\omega$. It should be noted that theoretically it is possible to calculate the moments of random variables of any order by MGF (3.1), but in practice it is complicated to derive the explicit formula for moments of order three or higher.

Theorem 1. *If $\boldsymbol{Y}_1 \sim EMMN_{d,m}(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\Delta}; H)$ and $\boldsymbol{Y}_2 \sim N_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ are independent variables, then $\boldsymbol{Y} = \boldsymbol{Y}_1 + \boldsymbol{Y}_2 \sim EMMN_{d,m}(\boldsymbol{\xi}_{\boldsymbol{Y}}, \boldsymbol{\Omega}_{\boldsymbol{Y}}, \boldsymbol{\Delta}_{\boldsymbol{Y}}; H)$, where $\boldsymbol{\xi}_{\boldsymbol{Y}} = \boldsymbol{\xi} + \boldsymbol{\mu}$, $\boldsymbol{\Omega}_{\boldsymbol{Y}} = \boldsymbol{\Omega} + \boldsymbol{\Sigma}$, and $\boldsymbol{\Delta}_{\boldsymbol{Y}} = \boldsymbol{\omega}_{\boldsymbol{Y}}^{-1}\omega\boldsymbol{\Delta}$, with $\boldsymbol{\omega}_{\boldsymbol{Y}} = (\boldsymbol{\Omega}_{\boldsymbol{Y}} \odot \boldsymbol{I}_p)^{1/2}$.*

Proof. From (2.1) and the MGF of $N_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is $\exp\left(\boldsymbol{t}^\top\boldsymbol{\mu} + \frac{1}{2}\boldsymbol{t}^\top\boldsymbol{\Sigma}\boldsymbol{t}\right)$. It can be immediately obtained that

$$
\begin{aligned}
M_{\boldsymbol{Y}}(\boldsymbol{t}) &= M_{\boldsymbol{Y}_1+\boldsymbol{Y}_2}(\boldsymbol{t}) = E\left[\mathrm{e}^{\boldsymbol{t}^\top(\boldsymbol{Y}_1+\boldsymbol{Y}_2)}\right] = E\left[\mathrm{e}^{\boldsymbol{t}^\top\boldsymbol{Y}_1}\right]E\left[\mathrm{e}^{\boldsymbol{t}^\top\boldsymbol{Y}_2}\right] \\
&= \mathrm{e}^{\boldsymbol{t}^\top(\boldsymbol{\xi}+\boldsymbol{\mu})}\mathrm{e}^{\frac{1}{2}\boldsymbol{t}^\top[(\boldsymbol{\Omega}+\boldsymbol{\Sigma})-\omega\boldsymbol{\Delta}\boldsymbol{\Delta}^\top\omega]\boldsymbol{t}}M_{\boldsymbol{U}}(\boldsymbol{\Delta}^\top\boldsymbol{\omega}\boldsymbol{t};\boldsymbol{\nu}) \\
&= \mathrm{e}^{\boldsymbol{t}^\top(\boldsymbol{\xi}+\boldsymbol{\mu})}\mathrm{e}^{\frac{1}{2}\boldsymbol{t}^\top[(\boldsymbol{\Omega}+\boldsymbol{\Sigma})-\omega_{\boldsymbol{Y}}\boldsymbol{\Delta}_{\boldsymbol{Y}}\boldsymbol{\Delta}_{\boldsymbol{Y}}^\top\omega_{\boldsymbol{Y}}]\boldsymbol{t}}M_{\boldsymbol{U}}(\boldsymbol{\Delta}_{\boldsymbol{Y}}^\top\boldsymbol{\omega}_{\boldsymbol{Y}}\boldsymbol{t};\boldsymbol{\nu}) \\
&= \mathrm{e}^{\boldsymbol{t}^\top\boldsymbol{\xi}_{\boldsymbol{Y}}}\mathrm{e}^{\frac{1}{2}\boldsymbol{t}^\top[\boldsymbol{\Omega}_{\boldsymbol{Y}}-\omega_{\boldsymbol{Y}}\boldsymbol{\Delta}_{\boldsymbol{Y}}\boldsymbol{\Delta}_{\boldsymbol{Y}}^\top\omega_{\boldsymbol{Y}}]\boldsymbol{t}}M_{\boldsymbol{U}}(\boldsymbol{\Delta}_{\boldsymbol{Y}}^\top\boldsymbol{\omega}_{\boldsymbol{Y}}\boldsymbol{t};\boldsymbol{\nu}),
\end{aligned}
$$

where $\boldsymbol{\omega}_{\boldsymbol{Y}} = (\boldsymbol{\Omega}_{\boldsymbol{Y}} \odot \boldsymbol{I}_p)^{1/2} = [(\boldsymbol{\Omega} + \boldsymbol{\Sigma}) \odot \boldsymbol{I}_d]^{1/2}$, and by $\boldsymbol{\Delta}_{\boldsymbol{Y}}^\top\boldsymbol{\omega}_{\boldsymbol{Y}} = \boldsymbol{\Delta}^\top\omega$, then $\boldsymbol{\Delta}_{\boldsymbol{Y}} = \boldsymbol{\omega}_{\boldsymbol{Y}}^{-1}\omega\boldsymbol{\Delta}$, hence the results. □

Theorem 2. *If $\boldsymbol{X} \sim EMMN_{d,m}(\boldsymbol{0}, \bar{\boldsymbol{\Omega}}, \boldsymbol{\Delta}; H)$ and $\boldsymbol{A}$ is a non-singular $d \times d$ matrix such that $\mathrm{diag}(\boldsymbol{A}^\top\bar{\boldsymbol{\Omega}}\boldsymbol{A}) = \boldsymbol{I}_d$, that is, $\boldsymbol{A}^\top\bar{\boldsymbol{\Omega}}\boldsymbol{A}$ is a correlation matrix, then $\boldsymbol{A}^\top\boldsymbol{X} \sim EMMN_{d,m}(\boldsymbol{0}, \boldsymbol{A}^\top\bar{\boldsymbol{\Omega}}\boldsymbol{A}, \boldsymbol{A}^\top\boldsymbol{\Delta}; H)$.*

Proof. From (2.1), we can easily obtain

$$
\begin{aligned}
M_{\boldsymbol{A}^\top\boldsymbol{X}}(\boldsymbol{t}) &= E\left[\mathrm{e}^{\boldsymbol{t}^\top\boldsymbol{A}^\top\boldsymbol{X}}\right] \\
&= E\left[\mathrm{e}^{\boldsymbol{t}^\top\boldsymbol{A}^\top(\boldsymbol{\Delta}\boldsymbol{U}+\boldsymbol{Z})}\right] \\
&= E\left[\mathrm{e}^{\boldsymbol{t}^\top\boldsymbol{A}^\top\boldsymbol{\Delta}\boldsymbol{U}}\right]E\left[\mathrm{e}^{\boldsymbol{t}^\top\boldsymbol{A}^\top\boldsymbol{Z}}\right] \\
&= \mathrm{e}^{\frac{1}{2}\left\{\boldsymbol{t}^\top(\boldsymbol{A}^\top(\bar{\boldsymbol{\Omega}}-\boldsymbol{\Delta}\boldsymbol{\Delta}^\top)\boldsymbol{A})\boldsymbol{t}\right\}}M_{\boldsymbol{U}}(\boldsymbol{\Delta}^\top\boldsymbol{A}\boldsymbol{t}) \\
&= \mathrm{e}^{\frac{1}{2}\left\{\boldsymbol{t}^\top(\boldsymbol{A}^\top\bar{\boldsymbol{\Omega}}\boldsymbol{A}-\boldsymbol{A}^\top\boldsymbol{\Delta}\boldsymbol{\Delta}^\top\boldsymbol{A})\boldsymbol{t}\right\}}M_{\boldsymbol{U}}(\boldsymbol{\Delta}^\top\boldsymbol{A}\boldsymbol{t}).
\end{aligned}
$$

Upon using the uniqueness property of the moment generating function, the required result is obtained. □

Theorem 2 presents that the EMMN distribution is closed under affine transformations.

Theorem 3. *If $\boldsymbol{Y} \sim EMMN_{d,m}(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\Delta}; H)$, $\boldsymbol{A}$ is a full-rank $d \times h$ matrix, with $h \leq d$, and $\boldsymbol{c} \in \mathbb{R}^h$, then $\boldsymbol{T} = \boldsymbol{c} + \boldsymbol{A}^\top\boldsymbol{Y} \sim EMMN_{h,m}(\boldsymbol{\xi}_{\boldsymbol{T}}, \boldsymbol{\Omega}_{\boldsymbol{T}}, \boldsymbol{\Delta}_{\boldsymbol{T}}; H)$, where $\boldsymbol{\xi}_{\boldsymbol{T}} = \boldsymbol{c} + \boldsymbol{A}^\top\boldsymbol{\xi}$, $\boldsymbol{\Omega}_{\boldsymbol{T}} = \boldsymbol{A}^\top\boldsymbol{\Omega}\boldsymbol{A}$, and $\boldsymbol{\Delta}_{\boldsymbol{T}} = \boldsymbol{\omega}_{\boldsymbol{T}}^{-1}\boldsymbol{A}^\top\omega\boldsymbol{\Delta}$, with $\boldsymbol{\omega}_{\boldsymbol{T}} = (\boldsymbol{\Omega}_{\boldsymbol{T}} \odot I_h)^{1/2}$.*

Proof. From (2.1),

$$
\begin{aligned}
M_{\boldsymbol{T}} &= E\left[\mathrm{e}^{\boldsymbol{t}^\top\boldsymbol{T}}\right] \\
&= E\left[\mathrm{e}^{\boldsymbol{t}^\top(\boldsymbol{c}+\boldsymbol{A}^\top(\boldsymbol{\xi}+\omega(\boldsymbol{\Delta}\boldsymbol{U}+\boldsymbol{Z})))}\right] \\
&= \mathrm{e}^{\boldsymbol{t}^\top(\boldsymbol{c}+\boldsymbol{A}^\top\boldsymbol{\xi})}E\left[\mathrm{e}^{\boldsymbol{t}^\top\boldsymbol{A}^\top\omega\boldsymbol{\Delta}\boldsymbol{U}}\right]E\left[\mathrm{e}^{\boldsymbol{A}^\top\omega\boldsymbol{Z}}\right]
\end{aligned}
$$

$$= e^{t^{\top}(c+A^{\top}\xi)} e^{\frac{1}{2}\{t^{\top}(A^{\top}\omega\bar{\Omega}\omega A - A^{\top}\omega\Delta\Delta^{\top}\omega A)t\}} M_U(\Delta^{\top}\omega A t)$$
$$= e^{t^{\top}\xi_T} e^{\frac{1}{2}\{t^{\top}(\Omega_T - \omega_T\Delta_T\Delta_T^{\top}\omega_T)t\}} M_U(\Delta_T^{\top}\omega_T t),$$

it can be obtained $\boldsymbol{\xi_T} = \boldsymbol{c} + \boldsymbol{A}^{\top}\boldsymbol{\xi}$, $\boldsymbol{\omega_T} = (\boldsymbol{\Omega_T} \odot I_h)^{1/2} = ((\boldsymbol{A}^{\top}\boldsymbol{\Omega}\boldsymbol{A}) \odot I_h)^{1/2}$, and with $\boldsymbol{\Delta}_T^{\top}\boldsymbol{\omega_T} = \boldsymbol{A}^{\top}\boldsymbol{\omega}\boldsymbol{\Delta}$, then $\boldsymbol{\Delta_T} = \boldsymbol{\omega}_T^{-1}\boldsymbol{A}^{\top}\boldsymbol{\omega}\boldsymbol{\Delta}$, which completes the proof. □

Theorem 4. *If $\boldsymbol{Y} \sim EMMN_{d,m}(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\Delta}; H)$, $\boldsymbol{\Delta}$ is a full-rank $d \times m$ matrix, with $m \le d$, then there exists a linear transformation $\boldsymbol{Z}^* = \boldsymbol{A}_*(\boldsymbol{Y} - \boldsymbol{\xi})$ such that $\boldsymbol{Z}^* \sim EMMN_{d,m}(\boldsymbol{0}, \boldsymbol{I}_d, \boldsymbol{\Delta}_*; H)$ where $\boldsymbol{\Delta}_* = (\delta^*_{ij})$ is $d \times m$ matrix, with $\delta^*_{ii} = \delta^*_i = (\boldsymbol{\delta}_i^{\top}\bar{\boldsymbol{\Omega}}^{-1}\boldsymbol{\delta}_i)^{1/2}$ $(i = j)$, $\delta^*_{ij} = 0$ $(i \neq j)$, for $i \in \{1, \cdots, d\}$, $j \in \{1, \cdots, m\}$. $\boldsymbol{\delta}_i$ is the ith column of $\boldsymbol{\Delta}$.*

Proof. Note that the matrix $\boldsymbol{\Omega} > 0$ is assumed in the EMMN distribution, through the factorization $\boldsymbol{\Omega} = \omega\bar{\boldsymbol{\Omega}}\omega$. The matrix $\bar{\boldsymbol{\Omega}}$ is a positive definite matrix if and only if there exists some invertible matrix $\boldsymbol{C}$ such that $\bar{\boldsymbol{\Omega}} = \boldsymbol{C}^{\top}\boldsymbol{C}$. If $\boldsymbol{\Delta} \neq \boldsymbol{0}$, there exists an orthogonal matrix $\boldsymbol{P}$ with the first m column being proportional to $\boldsymbol{C}\bar{\boldsymbol{\Omega}}^{-1}\boldsymbol{\Delta}$, while for $\boldsymbol{\Delta} = \boldsymbol{0}$ we set $\boldsymbol{P} = \boldsymbol{I}_p$. Finally, define $\boldsymbol{A}_* = (\boldsymbol{C}^{-1}\boldsymbol{P})^{\top}\omega^{-1}$. By using Theorem 3, we have

$$\boldsymbol{Z}^* \sim EMMNE_{d,m}(\boldsymbol{\xi}_{\boldsymbol{Z}^*}, \boldsymbol{\Omega}_{\boldsymbol{Z}^*}, \boldsymbol{\Delta}_{\boldsymbol{Z}^*}; H),$$

where $\boldsymbol{\xi}_{\boldsymbol{Z}^*} = \boldsymbol{0}$, $\boldsymbol{\Omega}_{\boldsymbol{Z}^*} = \boldsymbol{A}_*\boldsymbol{\Omega}\boldsymbol{A}_*^{\top} = \boldsymbol{I}_d$, $\omega_{\boldsymbol{Z}^*} = \boldsymbol{I}_p$, $\boldsymbol{\Delta}_{\boldsymbol{Z}^*} = \boldsymbol{A}_*\omega\boldsymbol{\Delta} = (\boldsymbol{C}^{-1}\boldsymbol{P})^{\top}\boldsymbol{\Delta}$. Let $\boldsymbol{\Delta} = (\boldsymbol{\delta}_1, \boldsymbol{\delta}_2, \ldots, \boldsymbol{\delta}_m)$ and rewrite orthogonal matrix $\boldsymbol{P}$ as partitioned matrix

$$\boldsymbol{P} = \left(\boldsymbol{C}\bar{\boldsymbol{\Omega}}^{-1}\boldsymbol{\Delta}\boldsymbol{K} \mid \boldsymbol{Q}\right),$$

where $\boldsymbol{K} = \operatorname{diag}(k_1, k_2, \ldots, k_m)$, $(k_1, k_2, \ldots, k_m)^{\top} \in \mathbb{R}^m$, from the properties of orthogonal matrix, we can conclude that

$$k_i = \frac{1}{\sqrt{\boldsymbol{\delta}_i^{\top}\bar{\boldsymbol{\Omega}}^{-1}\boldsymbol{\delta}_i}}, \quad i \in \{1, \ldots, m\},$$

and $\boldsymbol{Q}^{\top}\boldsymbol{C}^{-1}\boldsymbol{\Delta} = \boldsymbol{0}$. Then,

$$\boldsymbol{\Delta}_{\boldsymbol{Z}^*} = \left(\boldsymbol{C}^{-1}\boldsymbol{P}\right)^{\top}\boldsymbol{\Delta} = \begin{pmatrix} \boldsymbol{K}^{\top}\boldsymbol{\Delta}^{\top}\bar{\boldsymbol{\Omega}}^{-1}\boldsymbol{C} \\ \boldsymbol{Q}^{\top} \end{pmatrix} \boldsymbol{C}^{-1}\boldsymbol{\Delta} = \begin{pmatrix} \boldsymbol{K}^{\top}\boldsymbol{\Delta}^{\top}\bar{\boldsymbol{\Omega}}^{-1}\boldsymbol{\Delta} \\ \boldsymbol{Q}^{\top}\boldsymbol{C}^{-1}\boldsymbol{\Delta} \end{pmatrix} = \begin{pmatrix} \left(\operatorname{diag}(\boldsymbol{\Delta}^{\top}\bar{\boldsymbol{\Omega}}^{-1}\boldsymbol{\Delta})\right)^{\frac{1}{2}} \\ \boldsymbol{O} \end{pmatrix},$$

where $\operatorname{diag}(\boldsymbol{\Delta}^{\top}\bar{\boldsymbol{\Omega}}^{-1}\boldsymbol{\Delta})$ is a diagonal matrix whose ith diagonal element coincide with the ith diagonal element of the matrix $\boldsymbol{\Delta}^{\top}\bar{\boldsymbol{\Omega}}^{-1}\boldsymbol{\Delta}$, for $i \in \{1, \ldots, m\}$, and $\boldsymbol{O}$ is a matrix with all elements 0. Then, the required result is obtained. □

The variable $\boldsymbol{Z}^*$, which we shall somestimes refer to as, like the MMN distribution in [1], a canonical variate. It is assumed the components of $\boldsymbol{U}$ are mutually independent in (2.1), then the joint density of $\boldsymbol{Z}^*$ can be derived from (2.1), and be given by the product of $(d - m)$ standard normal densities and at most m non-Gaussian components $MMN_1(\boldsymbol{0}, 1, \delta^*_i; H_i)$, $i \in \{1, \ldots, m\}$, where H_i is CDF of the ith component of $\boldsymbol{U}$. Hence, the density of $\boldsymbol{Z}^*$ is

$$f_{\boldsymbol{Z}^*}(\boldsymbol{z}) = \prod_{i=1}^{m} f_{Z^*_i}(z_i) \prod_{i=m+1}^{d} \phi(z_i), \tag{3.2}$$

where $\boldsymbol{Z}^*_i \sim MMN_1(0, 1, \delta^*_i; H_i)$. For more details about this distribution, one may refer to [22].

Remark 2. *When $m = 1$, by (3.2), $f_{Z^*}(z) = f_{Z_1^*}(z_1) \prod_{i=2}^{d} \phi(z_i)$, which is the result of the canonical form of MMN distribution in [1].*

Theorem 5. *If $\boldsymbol{Y} \sim EMMN_{d,m}(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\Delta}; H)$, then, the mode of $\boldsymbol{Y}$ is*

$$\boldsymbol{M}_0 = \boldsymbol{\xi} + \boldsymbol{\omega}\boldsymbol{\Delta}\left(\frac{m_{01}^*}{\delta_1^*}\boldsymbol{e}_1 + \cdots + \frac{m_{0m}^*}{\delta_m^*}\boldsymbol{e}_m\right),$$

where m_{0i}^* is the mode of $MMN_1(0, 1, \delta_i^*; H_i)$ distribution, $i \in \{1, \ldots, m\}$, δ_i^* is the same as in Theorem 4 and $\boldsymbol{e}_i$ is m-dimensional unit normal vector whose ith element is 1.

Proof. First, it can be obtained that the mode of the corresponding canonical variable $\boldsymbol{Z}^* \sim EMMN_{d,m}(\boldsymbol{0}, \boldsymbol{I}_d, \boldsymbol{\Delta}_*; H)$ by solving the following equations with respect to $z_1, \ldots, z_d$:

$$\frac{\partial f_{z_1^*}(z_1)}{\partial z_1} = 0, \quad \frac{\partial f_{z_2^*}(z_2)}{\partial z_2} = 0, \ldots, \frac{\partial f_{z_m^*}(z_m)}{\partial z_m} = 0, \quad z_j \prod_{i=1}^{m} f_{z_i^*}(z_i) = 0, \quad j \in \{m+1, \ldots, d\}.$$

When $z_j = 0$, the last $d - m$ equations are fullfilled, and the root of the first m equations correspond to the mode, $m_{01}^*, m_{02}^*, \ldots, m_{0m}^*$ say, of the $MMN_1(0, 1, \delta_i^*; H_i), i \in \{1, \ldots, m\}$ distribution, respectively. Therefore, the mode of $\boldsymbol{Z}^*$ is

$$\boldsymbol{M}_0^* = (m_{01}^*, m_{02}^*, \ldots, m_{0m}^*, 0, \ldots, 0)^\top = \boldsymbol{\Delta}_*\left(\frac{m_{01}^*}{\delta_1^*}\boldsymbol{e}_1 + \cdots + \frac{m_{0m}^*}{\delta_m^*}\boldsymbol{e}_m\right).$$

By using Theorem 4, $\boldsymbol{Y} = \boldsymbol{\xi} + \boldsymbol{\omega}\boldsymbol{C}^\top\boldsymbol{P}\boldsymbol{Z}^*$ and $\boldsymbol{\Delta}_* = \boldsymbol{P}^\top\boldsymbol{C}\bar{\boldsymbol{\Omega}}^{-1}\boldsymbol{\Delta}$. Since the mode is equivariant with respect to affine transformations, the mode of $\boldsymbol{Y}$ is

$$\begin{aligned}
\boldsymbol{M}_0 = \boldsymbol{\xi} + \boldsymbol{\omega}\boldsymbol{C}^\top\boldsymbol{P}\boldsymbol{M}_0^* &= \boldsymbol{\xi} + \boldsymbol{\omega}\boldsymbol{C}^\top\boldsymbol{P}\boldsymbol{\Delta}_*\left(\frac{m_{01}^*}{\delta_1^*}\boldsymbol{e}_1 + \cdots + \frac{m_{0m}^*}{\delta_m^*}\boldsymbol{e}_m\right) \\
&= \boldsymbol{\xi} + \boldsymbol{\omega}\boldsymbol{C}^\top\boldsymbol{P}\boldsymbol{P}^\top\boldsymbol{C}\bar{\boldsymbol{\Omega}}^{-1}\boldsymbol{\Delta}\left(\frac{m_{01}^*}{\delta_1^*}\boldsymbol{e}_1 + \cdots + \frac{m_{0m}^*}{\delta_m^*}\boldsymbol{e}_m\right) \\
&= \boldsymbol{\xi} + \boldsymbol{\omega}\boldsymbol{\Delta}\left(\frac{m_{01}^*}{\delta_1^*}\boldsymbol{e}_1 + \cdots + \frac{m_{0m}^*}{\delta_m^*}\boldsymbol{e}_m\right),
\end{aligned}$$

which complete the proof of the result. □

Remark 3. *When $m = 1$, $\boldsymbol{M}_0 = \boldsymbol{\xi} + \frac{m_{01}^*}{\delta_1^*}\boldsymbol{\omega}\boldsymbol{\delta}_1$, which is consistent with the results of the MMN distribution in [1].*

4. Likelihood estimation through EM algorithm

In this section, we propose an EM-type algorithm, as taken in [1], to obtain the maximum likelihood estimates (MLE) of all parameters of $EMMN_{d,m}(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\Delta}; H)$. Let $\boldsymbol{Y}_1, \boldsymbol{Y}_2, \ldots, \boldsymbol{Y}_n$ be a random sample of size n from a $EMMN_{d,m}(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\Delta}; H)$ distribution. To implement the EM-type algorithm, the random vector $\boldsymbol{Y}$ and $\boldsymbol{U}$ in stochastic representation (2.1) are considered as observable variable and latent variable, respectively. For $\boldsymbol{Y}_i, i \in \{1, \ldots, n\}$ in (2.1), let $(\boldsymbol{Y}_i, \boldsymbol{U}_i), i \in \{1, \ldots, n\}$, be the complete-data, where $\boldsymbol{Y}_i$ is the observed data or incomplete-data, and $\boldsymbol{U}_i$ is considered as missing data. Let $\boldsymbol{\theta} = (\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\Delta}, \boldsymbol{\nu})$. By (2.3), for $\boldsymbol{Y}_i$, $i \in \{1, \ldots, n\}$,

$$\boldsymbol{Y}_i|(\boldsymbol{U}_i = \boldsymbol{u}_i) \sim N_d(\boldsymbol{\xi} + \omega\boldsymbol{\Delta}\boldsymbol{u}_i, \boldsymbol{\Sigma}_{\boldsymbol{Y}}), \qquad \boldsymbol{U}_i \overset{i.i.d.}{\sim} H(\cdot; \boldsymbol{\nu}),$$

where $\boldsymbol{\Sigma}_{\boldsymbol{Y}} = \boldsymbol{\Omega} - \omega\boldsymbol{\Delta}\boldsymbol{\Delta}^\top\omega$. Let $\boldsymbol{y} = (\boldsymbol{y}_1^\top, \ldots, \boldsymbol{y}_n^\top)^\top$, and $\boldsymbol{y}_i$ is a realization of $EMMN_{d,m}(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\Delta}; H)$. Note that

$$\begin{aligned} f(\boldsymbol{y}_i, \boldsymbol{u}_i) &= f(\boldsymbol{y}_i|\boldsymbol{u}_i)h(\boldsymbol{u}_i; \boldsymbol{\nu}) \\ &= (2\pi)^{-\frac{d}{2}}|\boldsymbol{\Sigma}_{\boldsymbol{Y}}|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(\boldsymbol{y}_i - \boldsymbol{\xi} - \omega\boldsymbol{\Delta}\boldsymbol{u}_i)^\top\boldsymbol{\Sigma}_{\boldsymbol{Y}}^{-1}(\boldsymbol{y}_i - \boldsymbol{\xi} - \omega\boldsymbol{\Delta}\boldsymbol{u}_i)\right\} \cdot h(\boldsymbol{u}_i; \boldsymbol{\nu}), \end{aligned} \tag{4.1}$$

the complete-data likelihood function is as follows

$$L_c(\boldsymbol{\theta}) = (2\pi)^{-\frac{nd}{2}}|\boldsymbol{\Sigma}_{\boldsymbol{Y}}|^{-\frac{n}{2}} \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}(\boldsymbol{y}_i - \boldsymbol{\xi} - \omega\boldsymbol{\Delta}\boldsymbol{u}_i)^\top\boldsymbol{\Sigma}_{\boldsymbol{Y}}^{-1}(\boldsymbol{y}_i - \boldsymbol{\xi} - \omega\boldsymbol{\Delta}\boldsymbol{u}_i)\right\} \cdot \prod_{i=1}^{n} h(\boldsymbol{u}_i; \boldsymbol{\nu}),$$

and the complete-data log-likelihood function is given by

$$\begin{aligned} l_c(\boldsymbol{\theta}) = &-\frac{n}{2}\ln|\boldsymbol{\Sigma}_{\boldsymbol{Y}}| - \frac{1}{2}\sum_{i=1}^{n}(\boldsymbol{y}_i - \boldsymbol{\xi})^\top\boldsymbol{\Sigma}_{\boldsymbol{Y}}^{-1}(\boldsymbol{y}_i - \boldsymbol{\xi}) + \sum_{i=1}^{n}(\boldsymbol{y}_i - \boldsymbol{\xi})^\top\boldsymbol{\Sigma}_{\boldsymbol{Y}}^{-1}\omega\boldsymbol{\Delta}\boldsymbol{u}_i \\ &- \frac{1}{2}\sum_{i=1}^{n}\boldsymbol{u}_i^\top\boldsymbol{\Delta}^\top\omega\boldsymbol{\Sigma}_{\boldsymbol{Y}}^{-1}\omega\boldsymbol{\Delta}\boldsymbol{u}_i + \sum_{i=1}^{n}\ln h(\boldsymbol{u}_i; \boldsymbol{\nu}), \end{aligned}$$

where the constant and parameter-independent terms in the above equation are omitted. Let $\widehat{\boldsymbol{\theta}}^{(k)} = \left(\widehat{\boldsymbol{\xi}}^{(k)}, \widehat{\boldsymbol{\Omega}}^{(k)}, \widehat{\boldsymbol{\Delta}}^{(k)}, \widehat{\boldsymbol{\nu}}^{(k)}\right)$ be the updated estimates after the kth iteration of the EM algorithm.

E-step: To compute the so-called Q-function, denoted by $Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)})$,

$$\begin{aligned} Q\left(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)}\right) &= E\left[l_c(\boldsymbol{\theta})|\boldsymbol{Y}_i = \boldsymbol{y}_i, \widehat{\boldsymbol{\theta}}^{(k)}\right] \\ &= -\frac{n}{2}\ln|\boldsymbol{\Sigma}_{\boldsymbol{Y}}| - \frac{1}{2}\sum_{i=1}^{n}(\boldsymbol{y}_i - \boldsymbol{\xi})^\top\boldsymbol{\Sigma}_{\boldsymbol{Y}}^{-1}(\boldsymbol{y}_i - \boldsymbol{\xi}) + \sum_{i=1}^{n}(\boldsymbol{y}_i - \boldsymbol{\xi})^\top\boldsymbol{\Sigma}_{\boldsymbol{Y}}^{-1}\omega\boldsymbol{\Delta}E\left[\boldsymbol{u}_i|\boldsymbol{Y}_i = \boldsymbol{y}_i, \widehat{\boldsymbol{\theta}}^{(k)}\right] \\ &\quad - \frac{1}{2}\sum_{i=1}^{n}E\left[\boldsymbol{u}_i^\top\boldsymbol{\Delta}^\top\omega\boldsymbol{\Sigma}_{\boldsymbol{Y}}^{-1}\omega\boldsymbol{\Delta}\boldsymbol{u}_i|\boldsymbol{Y}_i = \boldsymbol{y}_i, \widehat{\boldsymbol{\theta}}^{(k)}\right] + \sum_{i=1}^{n}E\left[\ln h(\boldsymbol{u}_i; \boldsymbol{\nu})|\boldsymbol{Y}_i = \boldsymbol{y}_i, \widehat{\boldsymbol{\theta}}^{(k)}\right], \end{aligned}$$

and set $\boldsymbol{A} = \omega\boldsymbol{\Delta}$, then

$$\begin{aligned} Q\left(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)}\right) = &-\frac{n}{2}\ln|\boldsymbol{\Sigma}_{\boldsymbol{Y}}| - \frac{1}{2}\sum_{i=1}^{n}(\boldsymbol{y}_i - \boldsymbol{\xi})^\top\boldsymbol{\Sigma}_{\boldsymbol{Y}}^{-1}(\boldsymbol{y}_i - \boldsymbol{\xi}) + \sum_{i=1}^{n}(\boldsymbol{y}_i - \boldsymbol{\xi})^\top\boldsymbol{\Sigma}_{\boldsymbol{Y}}^{-1}\boldsymbol{A}E\left[\boldsymbol{u}_i|\boldsymbol{Y}_i = \boldsymbol{y}_i, \widehat{\boldsymbol{\theta}}^{(k)}\right] \\ &- \frac{1}{2}\sum_{i=1}^{n}E\left[\boldsymbol{u}_i^\top\boldsymbol{A}^\top\boldsymbol{\Sigma}_{\boldsymbol{Y}}^{-1}\boldsymbol{A}\boldsymbol{u}_i|\boldsymbol{Y}_i = \boldsymbol{y}_i, \widehat{\boldsymbol{\theta}}^{(k)}\right] + \sum_{i=1}^{n}E\left[\ln h(\boldsymbol{u}_i; \boldsymbol{\nu})|\boldsymbol{Y}_i = \boldsymbol{y}_i, \widehat{\boldsymbol{\theta}}^{(k)}\right]. \end{aligned} \tag{4.2}$$

M-step: Maximize $Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)})$ of (4.2) with respect to $\boldsymbol{\theta}$,

$$\frac{\partial Q\left(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)}\right)}{\partial\boldsymbol{\xi}} = 0, \quad \frac{\partial Q\left(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)}\right)}{\partial\boldsymbol{A}} = 0, \quad \frac{\partial Q\left(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)}\right)}{\partial\boldsymbol{\Sigma}_{\boldsymbol{Y}}} = 0.$$

After some algebraic manipulation, we can get the following closed-form expressions:

$$\widehat{\boldsymbol{A}}^{(k+1)} = \left\{\sum_{i=1}^{n} \boldsymbol{y}_i E\left[\boldsymbol{u}_i|\boldsymbol{Y}_i = \boldsymbol{y}_i, \widehat{\boldsymbol{\theta}}^{(k)}\right]^\top - \bar{\boldsymbol{y}} \sum_{i=1}^{n} E\left[\boldsymbol{u}_i|\boldsymbol{Y}_i = \boldsymbol{y}_i, \widehat{\boldsymbol{\theta}}^{(k)}\right]^\top\right\} \times$$

$$\left\{\sum_{i=1}^{n} E\left[\boldsymbol{u}_i\boldsymbol{u}_i^\top|\boldsymbol{Y}_i = \boldsymbol{y}_i, \widehat{\boldsymbol{\theta}}^{(k)}\right] - \frac{1}{n}\left(\sum_{i=1}^{n} E[\boldsymbol{u}_i|\boldsymbol{Y}_i = \boldsymbol{y}_i, \widehat{\boldsymbol{\theta}}^{(k)}]\right)\left(\sum_{i=1}^{n} E\left[\boldsymbol{u}_i|\boldsymbol{Y}_i = \boldsymbol{y}_i, \widehat{\boldsymbol{\theta}}^{(k)}\right]^\top\right)\right\}^{-1},$$

$$\widehat{\boldsymbol{\xi}}^{(k+1)} = \bar{\boldsymbol{y}} - \frac{1}{n}\widehat{\boldsymbol{A}}^{(k+1)} \sum_{i=1}^{n} E\left[\boldsymbol{u}_i|\boldsymbol{Y}_i = \boldsymbol{y}_i, \widehat{\boldsymbol{\theta}}^{(k)}\right],$$

$$\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{Y}}^{(k+1)} = \frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{y}_i - \widehat{\boldsymbol{\xi}}^{(k+1)})(\boldsymbol{y}_i - \widehat{\boldsymbol{\xi}}^{(k+1)})^\top - \frac{2}{n}\sum_{i=1}^{n}(\boldsymbol{y} - \widehat{\boldsymbol{\xi}}^{(k+1)})E\left[\boldsymbol{u}_i|\boldsymbol{Y}_i = \boldsymbol{y}_i, \widehat{\boldsymbol{\theta}}^{(k)}\right]^\top \left(\widehat{\boldsymbol{A}}^{(k+1)}\right)^\top$$

$$+ \frac{1}{n}\widehat{\boldsymbol{A}}^{(k+1)} \sum_{i=1}^{n} E\left[\boldsymbol{u}_i\boldsymbol{u}_i^\top|\boldsymbol{Y}_i = \boldsymbol{y}_i, \widehat{\boldsymbol{\theta}}^{(k)}\right]\left(\widehat{\boldsymbol{A}}^{(k+1)}\right)^\top,$$

where $\bar{\boldsymbol{y}} = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{y}_i$ is the sample mean vector. Then,

$$\widehat{\boldsymbol{\Omega}}^{(k+1)} = \widehat{\boldsymbol{\Sigma}}_{\boldsymbol{Y}}^{(k+1)} + \widehat{\boldsymbol{A}}^{(k+1)}\left(\widehat{\boldsymbol{A}}^{(k+1)}\right)^\top,$$

$$\widehat{\boldsymbol{\Delta}} = \left(\widehat{\boldsymbol{\omega}}^{(k+1)}\right)^{-1}\widehat{\boldsymbol{A}}^{(k+1)}, \widehat{\boldsymbol{\omega}} = \left(\widehat{\boldsymbol{\Omega}} \odot \boldsymbol{I}_p\right)^{1/2}.$$

From [1], two strategies for update of $\widehat{\boldsymbol{\nu}}^{(k)}$ were proposed as:
M-step 2:

$$\widehat{\boldsymbol{\nu}}^{(k+1)} = \arg\max_{\boldsymbol{\nu}} \sum_{i=1}^{n} E\left[\ln h(\boldsymbol{u}_i; \boldsymbol{\nu})|\boldsymbol{Y}_i = \boldsymbol{y}_i, \widehat{\boldsymbol{\theta}}^{(k)}\right].$$

Modified M-step 2:

$$\widehat{\boldsymbol{\nu}}^{(k+1)} = \arg\max_{\boldsymbol{\nu}} \sum_{i=1}^{n} \ln f_{EMMN_{d,m}}\left(\boldsymbol{y}_i; \widehat{\boldsymbol{\xi}}^{(k+1)}, \widehat{\boldsymbol{\Omega}}^{(k+1)}, \widehat{\boldsymbol{\Delta}}^{(k+1)}, \boldsymbol{\nu}\right).$$

Stopping criterion: Relative likelihood-based approach is the common way for stopping criterion, that is

$$\frac{\left|l\left(\widehat{\boldsymbol{\theta}}^{(k+1)}|\boldsymbol{y}\right) - l\left(\widehat{\boldsymbol{\theta}}^{(k)}|\boldsymbol{y}\right)\right|}{\left|l\left(\widehat{\boldsymbol{\theta}}^{(k)}|\boldsymbol{y}\right)\right|} < \epsilon,$$

where $l(\widehat{\boldsymbol{\theta}}^{(k)}|\boldsymbol{y}) = \sum_{i=1}^{n} \ln f_{EMMN_{d,m}}\left(\boldsymbol{y}_i; \widehat{\boldsymbol{\xi}}^{(k)}, \widehat{\boldsymbol{\Omega}}^{(k)}, \widehat{\boldsymbol{\Delta}}^{(k)}, \widehat{\boldsymbol{\nu}}^{(k)}\right)$, and the threshold ϵ usually be set to 10^{-5}. For other stopping criterions, one may refer to Lee and McLachlan [17].

Remark 4. *The initial value of local and scale parameter $\boldsymbol{\xi}$, $\boldsymbol{\Omega}$ could be obtained by sample mean and sample covariance, respectively, and for the skewness matrix $\boldsymbol{\Delta}$, whose initial value could be set according to the sample skewness, and other strategies related to this could be found in Lee et al. [18].*

5. Special case of EMMN distribution

In this section, we explore a special case of the EMMN family when the m-dimensional mixture random vector $\boldsymbol{U}$ in stochastic representation (2.1) follows the multi-standard gamma distribution with correponding PDF $h(\boldsymbol{u};\boldsymbol{v}) = \prod_{i=1}^{m} u_i^{v_i-1}\mathrm{e}^{-u_i}/\Gamma(v_i), v_i > 0, u_i > 0, i \in \{1,\ldots,m\}$. It means that each component U_i of random vector $\boldsymbol{U}$ is independent of each other and $U_i \sim Ga(v_i, 1)$, following univariate standard gamma distribution with parameter v_i. We denote this special distribution by $\boldsymbol{Y} \sim EMMNG_{d,m}(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\Delta}, \boldsymbol{v})$, $\boldsymbol{v} = (v_1,\ldots,v_m)$. We obtain the PDF of $\boldsymbol{Y}$ by (2.2) as follows:

$$
\begin{aligned}
&f_{EMMNG_{d,m}}(\boldsymbol{y};\boldsymbol{\xi},\boldsymbol{\Omega},\boldsymbol{\Delta},\boldsymbol{v})\\
=&\phi_d(\boldsymbol{y};\boldsymbol{\xi},\boldsymbol{\Sigma_Y})\int_{\mathbb{R}_+^m}\exp\left\{-\frac{1}{2}[-2(\boldsymbol{y}-\boldsymbol{\xi})^\top\boldsymbol{\Sigma_Y}^{-1}\boldsymbol{\omega\Delta u}+(\boldsymbol{\omega\Delta u})^\top\boldsymbol{\Sigma_Y}^{-1}(\boldsymbol{\omega\Delta u})]\right\}\prod_{i=1}^{m}\frac{1}{\Gamma(v_i)}u_i^{v_i-1}\mathrm{e}^{-u_i}\,\mathrm{d}\boldsymbol{\mu},\ \boldsymbol{y}\in\mathbb{R}^d.
\end{aligned}
$$

By (3.1), the MGF of $\boldsymbol{Y} \sim EMMNG_{d,m}(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\Delta}, \boldsymbol{v})$ is

$$
M_{\boldsymbol{Y}}(\boldsymbol{t}) = \mathrm{e}^{\boldsymbol{t}^\top\boldsymbol{\xi}+\frac{1}{2}\boldsymbol{t}^\top\boldsymbol{\Sigma_Y}\boldsymbol{t}}\prod_{i=1}^{m}(1-\boldsymbol{t}^\top\boldsymbol{\omega\Delta e}_i)^{-v_i}. \tag{5.1}
$$

For $U \sim Ga(v, 1)$, $E(U^r) = \frac{\Gamma(v+r)}{\Gamma(v)}$, $r > 0$. We can conclude that $E(\boldsymbol{U}) = \boldsymbol{v}$, $E(\boldsymbol{U}\otimes\boldsymbol{U}^\top) = \boldsymbol{v}\otimes\boldsymbol{v}^\top + \mathrm{diag}(\boldsymbol{v})$. Specifically, $E(\boldsymbol{Y}) = \boldsymbol{\xi} + \boldsymbol{\omega\Delta v}$, $var(\boldsymbol{Y}) = \boldsymbol{\Omega} + \boldsymbol{\omega\Delta}\mathrm{diag}(\boldsymbol{v} - \boldsymbol{1}_d)$.

Definition 2. *(Abdi et al. [1]) A random vector $\boldsymbol{Y}$ (or its distribution) is said to be infinitely divisible if, for each $n \geq 1$, there exist independent and identically distributed (i.i.d.) random vectors $\boldsymbol{Y}_1,\ldots,\boldsymbol{Y}_n$ such that $\boldsymbol{Y} \overset{d}{=} \boldsymbol{Y}_1 + \cdots + \boldsymbol{Y}_n$.*

Theorem 6. *The EMMNG distribution is infinitely divisible.*

Proof. Considering $\boldsymbol{X} \sim EMMNG_{d,m}(\boldsymbol{0}, \bar{\boldsymbol{\Omega}}, \boldsymbol{\Delta}, \boldsymbol{v})$, $\boldsymbol{v} = (v_1,\ldots,v_m)$ and $\boldsymbol{X}_i \overset{d}{=} \boldsymbol{\Delta U}_i + \boldsymbol{Z}_i$, where $\boldsymbol{U}_i = (U_{i1},\ldots,U_{im})^\top$, $U_{i1},\ldots,U_{im}$ are independent and $U_{ij} \sim Ga\left(\alpha = \frac{v_j}{n}, \beta = 1\right)$ for $j \in \{1,\ldots,m\}$. $\boldsymbol{Z}_i \sim N_d\left(\boldsymbol{0}, \frac{1}{n}(\bar{\boldsymbol{\Omega}} - \boldsymbol{\Delta\Delta}^\top)\right)$. Note that $\sum_{i=1}^{n}\boldsymbol{U}_{ij} \sim Ga(v_j, 1)$ and $\sum_{i=1}^{m}\boldsymbol{Z}_i \sim N_d(\boldsymbol{0}, \bar{\boldsymbol{\Omega}} - \boldsymbol{\Delta\Delta}^\top)$, then it can be concluded that $\boldsymbol{X} \overset{d}{=} \boldsymbol{X}_1 + \cdots + \boldsymbol{X}_n$, which completes the proof of Theorem 6. □

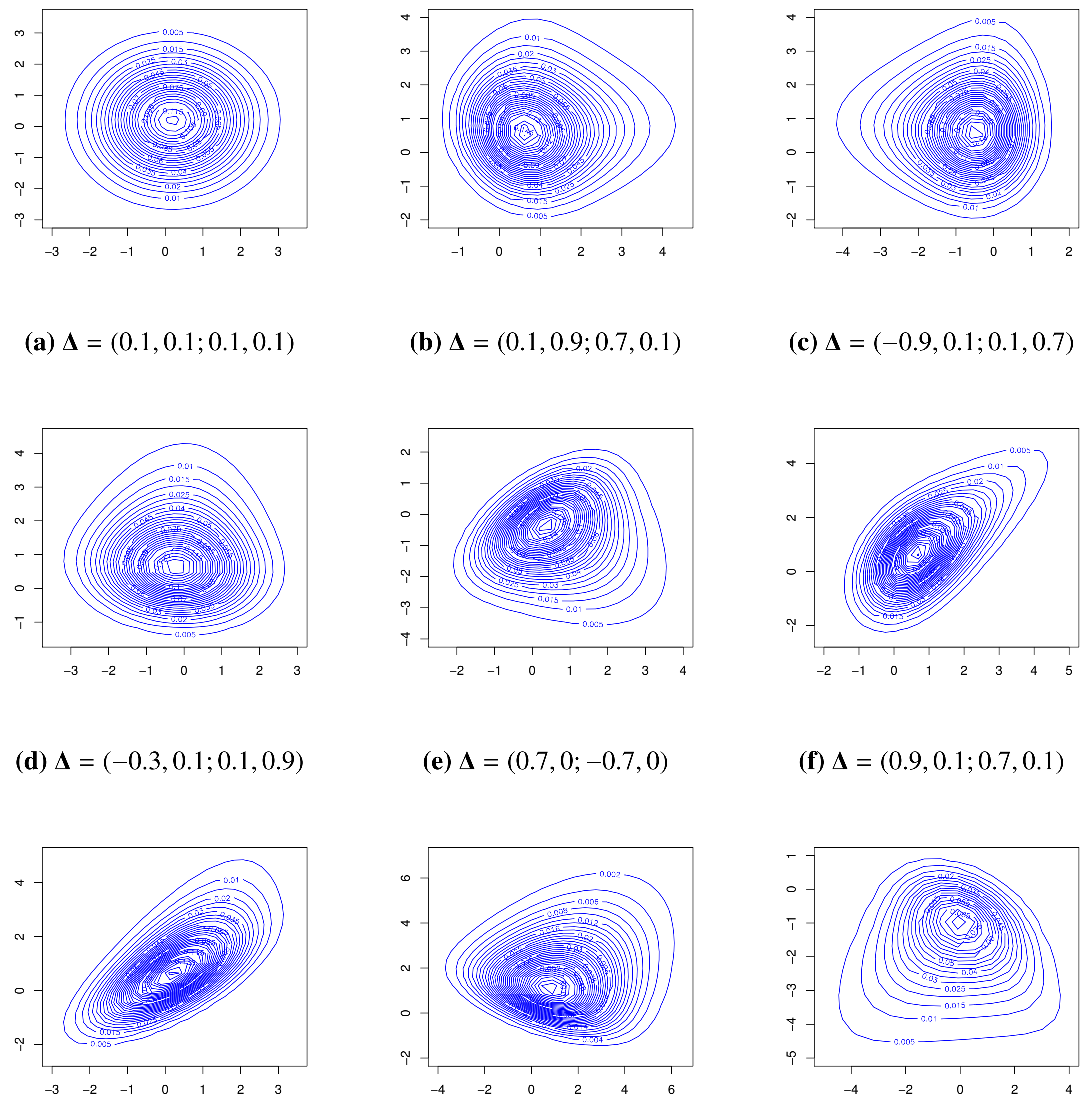
When $\boldsymbol{v} = \boldsymbol{1}_m$ in the EMMNG distribution, i.e., $\boldsymbol{U}$ in stochastic representation (2.1) follows the multi-standard exponential distribution, and we write the d-dimensional random vector $\boldsymbol{Y} \sim EMMNE_{d,m}(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\Delta})$. By (2.2), the PDF of $\boldsymbol{Y}$ is obtained as

$$
\begin{aligned}
f_{\boldsymbol{Y}}(\boldsymbol{y}) &= \phi_d(\boldsymbol{y};\boldsymbol{\xi},\boldsymbol{\Sigma_Y})\int_{\mathbb{R}_+^m}\exp\left\{-\frac{1}{2}\left[-2(\boldsymbol{y}-\boldsymbol{\xi})^\top\boldsymbol{\Sigma_Y}^{-1}\boldsymbol{\omega\Delta u}+(\boldsymbol{\omega\Delta u})^\top\boldsymbol{\Sigma_Y}^{-1}(\boldsymbol{\omega\Delta u})\right]\right\}\mathrm{e}^{\boldsymbol{1}_m^\top\boldsymbol{u}}\,\mathrm{d}\boldsymbol{u}\\
&= (2\pi)^{\frac{m}{2}}\left|\boldsymbol{\Sigma}_{\mathcal{E}}\right|^{-\frac{1}{2}}\phi_d\left(\boldsymbol{y};\boldsymbol{\xi},\boldsymbol{\Sigma_Y}\right)\exp\left\{\frac{1}{2}\boldsymbol{\eta}^\top\boldsymbol{\Sigma}_{\mathcal{E}}\boldsymbol{\eta}\right\}\int_{\mathbb{R}_+^m}\phi_m\left(\boldsymbol{u};\boldsymbol{\mu}_{\mathcal{E}},\boldsymbol{\Sigma}_{\mathcal{E}}\right)\mathrm{d}\boldsymbol{u},\quad \boldsymbol{y}\in\mathbb{R}^d,
\end{aligned} \tag{5.2}
$$

where $\boldsymbol{\mu}_{\mathcal{E}} = \boldsymbol{\Sigma}_{\mathcal{E}}\boldsymbol{\eta}$, $\boldsymbol{\eta} = \boldsymbol{\Delta}^\top\boldsymbol{\omega\Sigma_Y}^{-1}(\boldsymbol{y}-\boldsymbol{\xi}) - \boldsymbol{1}_m$, $\boldsymbol{\Sigma}_{\mathcal{E}} = \left(\boldsymbol{\Delta}^\top\boldsymbol{\omega\Sigma_Y}^{-1}\boldsymbol{\omega\Delta}\right)^{-1}$.

Figure 1 illustrates the contour plots of the bivariate EMMNE distributions, where (a)–(e) correspond to $\boldsymbol{\Omega} = (1, 0; 0, 1)$, (f) and (g) correspond to $\boldsymbol{\Omega} = (1, 1.5; 1, 1)$, and (h) and (i) correspond to

$\boldsymbol{\Omega} = (3, 0; 0, 1.5)$, and different choice of $\boldsymbol{\Delta}$ for $\boldsymbol{\xi} = (0, 0)^\top$, we can see that the EMMNE distribution has flexible shapes which clearly depends on $\boldsymbol{\Omega}$ and $\boldsymbol{\Delta}$.



(a) $\boldsymbol{\Delta} = (0.1, 0.1; 0.1, 0.1)$ **(b)** $\boldsymbol{\Delta} = (0.1, 0.9; 0.7, 0.1)$ **(c)** $\boldsymbol{\Delta} = (-0.9, 0.1; 0.1, 0.7)$

(d) $\boldsymbol{\Delta} = (-0.3, 0.1; 0.1, 0.9)$ **(e)** $\boldsymbol{\Delta} = (0.7, 0; -0.7, 0)$ **(f)** $\boldsymbol{\Delta} = (0.9, 0.1; 0.7, 0.1)$

(g) $\boldsymbol{\Delta} = (0.315, 0; 0.795, 0)$ **(h)** $\boldsymbol{\Delta} = (0.5, 0.3; 0.7, 0.8)$ **(i)** $\boldsymbol{\Delta} = (0.5, -0.7; -0.7, -0.8)$

Figure 1. Contour plots of EMMNE distributions for different choices of $\boldsymbol{\Delta}$.

Lemma 2. *Let* $\boldsymbol{X} \sim TN_p(\boldsymbol{a}, \boldsymbol{b}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$. *Then*

$$E[\boldsymbol{X}] = \boldsymbol{\mu} + \bar{\Phi}_p^{-1}(\boldsymbol{a}, \boldsymbol{b}; \boldsymbol{\mu}, \boldsymbol{\Sigma})\boldsymbol{\Sigma}\boldsymbol{q}^N(\boldsymbol{a}, \boldsymbol{b}; \boldsymbol{\mu}, \boldsymbol{\Sigma}),$$

$$E\left[\boldsymbol{X}\boldsymbol{X}^\top\right] = \boldsymbol{\Sigma} + \boldsymbol{\mu}\boldsymbol{\mu}^\top + \bar{\Phi}_p^{-1}(\boldsymbol{a}, \boldsymbol{b}; \boldsymbol{\mu}, \boldsymbol{\Sigma})\left\{\boldsymbol{\mu}\left(\boldsymbol{\Sigma}\boldsymbol{q}^N(\boldsymbol{a}, \boldsymbol{b}; \boldsymbol{\mu}, \boldsymbol{\Sigma})\right)^\top + \boldsymbol{\Sigma}\boldsymbol{q}^N(\boldsymbol{a}, \boldsymbol{b}; \boldsymbol{\mu}, \boldsymbol{\Sigma})\boldsymbol{\mu}^\top \right.$$
$$\left. + \boldsymbol{\Sigma}\left(\boldsymbol{H}^N(\boldsymbol{a}, \boldsymbol{b}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) + \boldsymbol{D}^N(\boldsymbol{a}, \boldsymbol{b}; \boldsymbol{\mu}, \boldsymbol{\Sigma})\right)\boldsymbol{\Sigma}\right\}.$$

The expression of $\boldsymbol{q}^N(\boldsymbol{a},\boldsymbol{b};\boldsymbol{\mu},\boldsymbol{\Sigma})$, $\boldsymbol{H}^N(\boldsymbol{a},\boldsymbol{b};\boldsymbol{\mu},\boldsymbol{\Sigma})$, $\boldsymbol{D}^N(\boldsymbol{a},\boldsymbol{b};\boldsymbol{\mu},\boldsymbol{\Sigma})$ and for more details, one may refer to Roozegar et al. [25].

The following theorem plays a vital role in implementing the EM algorithm for parameter estimation of the EMMNE distribution.

Theorem 7. *If $\boldsymbol{Y} \sim EMMNE_{d,m}(\boldsymbol{\xi},\boldsymbol{\Omega},\boldsymbol{\Delta})$ with $\boldsymbol{U}$ in stochastic representation (2.1) follows the multi-standard exponential distribution, then $\boldsymbol{U}|(\boldsymbol{Y}=\boldsymbol{y}) \sim TN_m(\boldsymbol{0},+\boldsymbol{\infty};\boldsymbol{\mu}_{\mathcal{E}},\boldsymbol{\Sigma}_{\mathcal{E}})$, where $\boldsymbol{\mu}_{\mathcal{E}}=\boldsymbol{\Sigma}_{\mathcal{E}}\boldsymbol{\eta}$, $\boldsymbol{\eta}=\boldsymbol{\Delta}^{\top}\omega\boldsymbol{\Sigma}_{\boldsymbol{Y}}^{-1}(\boldsymbol{y}-\boldsymbol{\xi})-\boldsymbol{1}_m$, $\boldsymbol{\Sigma}_{\mathcal{E}}=\left(\boldsymbol{\Delta}^{\top}\omega\boldsymbol{\Sigma}_{\boldsymbol{Y}}^{-1}\omega\boldsymbol{\Delta}\right)^{-1}=\left(\boldsymbol{\Delta}^{\top}(\bar{\boldsymbol{\Omega}}-\boldsymbol{\Delta}\boldsymbol{\Delta}^{\top})^{-1}\boldsymbol{\Delta}\right)^{-1}$. Furthermore,*

$$\begin{aligned}
E\left[\boldsymbol{U}|\boldsymbol{Y}=\boldsymbol{y}\right] &= \boldsymbol{\mu}_{\mathcal{E}}+\bar{\Phi}_m^{-1}(\boldsymbol{0},+\boldsymbol{\infty};\boldsymbol{\mu}_{\mathcal{E}},\boldsymbol{\Sigma}_{\mathcal{E}})\boldsymbol{\Sigma}_{\mathcal{E}}\boldsymbol{q}^N(\boldsymbol{0},+\boldsymbol{\infty};\boldsymbol{\mu}_{\mathcal{E}},\boldsymbol{\Sigma}_{\mathcal{E}}),\\
E\left[\boldsymbol{U}\boldsymbol{U}^{\top}|\boldsymbol{Y}=\boldsymbol{y}\right] &= \boldsymbol{\Sigma}_{\mathcal{E}}+\boldsymbol{\mu}_{\mathcal{E}}\boldsymbol{\mu}_{\mathcal{E}}^{\top}+\bar{\Phi}_m^{-1}(\boldsymbol{0},+\boldsymbol{\infty};\boldsymbol{\mu}_{\mathcal{E}},\boldsymbol{\Sigma}_{\mathcal{E}})\left\{\boldsymbol{\mu}_{\mathcal{E}}\left(\boldsymbol{\Sigma}_{\mathcal{E}}\boldsymbol{q}^N(\boldsymbol{0},+\boldsymbol{\infty};\boldsymbol{\mu}_{\mathcal{E}},\boldsymbol{\Sigma}_{\mathcal{E}})\right)^{\top}\right.\\
&\quad\left.+\boldsymbol{\Sigma}_{\mathcal{E}}\left(\boldsymbol{H}^N(\boldsymbol{0},+\boldsymbol{\infty};\boldsymbol{\mu}_{\mathcal{E}},\boldsymbol{\Sigma}_{\mathcal{E}})+\boldsymbol{D}^N(\boldsymbol{0},+\boldsymbol{\infty};\boldsymbol{\mu}_{\mathcal{E}},\boldsymbol{\Sigma}_{\mathcal{E}})\right)\boldsymbol{\Sigma}_{\mathcal{E}}\right\}.
\end{aligned}$$

The completed proof can be found in Appendix.

Remark 5. *The MLE of the parameters of the EMMNE distribution can be obtained by Theorem 7 with*

$$\begin{aligned}
E\left[\boldsymbol{U}_i|\boldsymbol{Y}_i=\boldsymbol{y}_i,\widehat{\boldsymbol{\theta}}^{(k)}\right] &= \widehat{\boldsymbol{\mu}}_{\mathcal{E}}^{(k)}+\bar{\Phi}_p^{-1}(\boldsymbol{0},+\boldsymbol{\infty};\widehat{\boldsymbol{\mu}}_{\mathcal{E}}^{(k)},\widehat{\boldsymbol{\Sigma}}_{\mathcal{E}}^{(k)})\widehat{\boldsymbol{\Sigma}}_{\mathcal{E}}^{(k)}\boldsymbol{q}^N(\boldsymbol{0},+\boldsymbol{\infty};\widehat{\boldsymbol{\mu}}_{\mathcal{E}}^{(k)},\widehat{\boldsymbol{\Sigma}}_{\mathcal{E}}^{(k)}),\\
E\left[\boldsymbol{U}_i\boldsymbol{U}_i^{\top}|\boldsymbol{Y}_i=\boldsymbol{y}_i,\widehat{\boldsymbol{\theta}}^{(k)}\right] &= \widehat{\boldsymbol{\Sigma}}_{\mathcal{E}}^{(k)}+\widehat{\boldsymbol{\mu}}_{\mathcal{E}}^{(k)}\widehat{\boldsymbol{\mu}}_{\mathcal{E}}^{(k)\top}+\bar{\Phi}_p^{-1}(\boldsymbol{0},+\boldsymbol{\infty};\widehat{\boldsymbol{\mu}}_{\mathcal{E}}^{(k)},\widehat{\boldsymbol{\Sigma}}_{\mathcal{E}}^{(k)})\\
&\quad\times\left\{\widehat{\boldsymbol{\mu}}_{\mathcal{E}}^{(k)}\left(\widehat{\boldsymbol{\Sigma}}_{\mathcal{E}}^{(k)}\boldsymbol{q}^N(\boldsymbol{0},+\boldsymbol{\infty};\widehat{\boldsymbol{\mu}}_{\mathcal{E}}^{(k)},\widehat{\boldsymbol{\Sigma}}_{\mathcal{E}}^{(k)})\right)^{\top}+\widehat{\boldsymbol{\Sigma}}_{\mathcal{E}}^{(k)}\boldsymbol{q}^N(\boldsymbol{0},+\boldsymbol{\infty};\widehat{\boldsymbol{\mu}}_{\mathcal{E}}^{(k)},\widehat{\boldsymbol{\Sigma}}_{\mathcal{E}}^{(k)})\widehat{\boldsymbol{\mu}}_{\mathcal{E}}^{(k)\top}\right.\\
&\quad\left.+\widehat{\boldsymbol{\Sigma}}_{\mathcal{E}^{(k)}}\left(\boldsymbol{H}^N(\boldsymbol{0},+\boldsymbol{\infty};\widehat{\boldsymbol{\mu}}_{\mathcal{E}}^{(k)},\widehat{\boldsymbol{\Sigma}}_{\mathcal{E}}^{(k)})+\boldsymbol{D}^N(\boldsymbol{0},+\boldsymbol{\infty};\widehat{\boldsymbol{\mu}}_{\mathcal{E}}^{(k)},\widehat{\boldsymbol{\Sigma}}_{\mathcal{E}}^{(k)})\right)\widehat{\boldsymbol{\Sigma}}_{\mathcal{E}}^{(k)}\right\}.
\end{aligned}$$

Since $\boldsymbol{\nu}=\boldsymbol{1}_m$, for EMMNE distribution, the M-step 2 in EM algorithm will be skipped.

Theorem 8. *If $\boldsymbol{Y} \sim EMMNE$ distribution, then $\boldsymbol{Y}$ is log-concave.*

Proof. The proof is similar to that of Theorem 9 in [1], considering the PDF of the canonical form of $\boldsymbol{Y}$, and by (3.2), we have

$$f_{\boldsymbol{Z}^*}(\boldsymbol{z})=\prod_{i=1}^{m}f_{Z_i^*}(z_i)\prod_{i=m+1}^{d}\phi(z_i), \tag{5.3}$$

where $Z_i^* \sim EMMNE_1(0,1,\delta_i^*)$, and was proved to be log-concave in Negarestani et al. [22]. The univariate normal distribution is also log-concave, hence the canoical form $\boldsymbol{Z}^*=(Z_1^*,\ldots,Z_d^*)$ is log-concave, considering the property that log-concavity is preserved by affine transformation, then we complete the proof. □

6. Simulation study

In this section, we examine the performance of the EM-type algorithm in parameter estimation of the EMMNE distribution. Four terms are considered, including average values (Mean), standard deviations (Std.), bias (Bias) and mean squared error (MSE) which are calculated with sample sizes

$n \in \{50, 100, 300, 500\}$ based on 300 replications from the $EMMNE_{d,m}(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\Delta})$, where $d = m = 2$, and parameters are considered as follows:

$$\boldsymbol{\xi} = (5, 10), \quad \boldsymbol{\Omega} = \begin{pmatrix} 0.4 & 0 \\ 0 & 0.6 \end{pmatrix}, \quad \boldsymbol{\Delta} = \begin{pmatrix} 0.3 & 0.1 \\ 0.2 & 0.7 \end{pmatrix}.$$

The formulations of Mean, Std. Bias and MSE are

$$\overline{\widehat{\theta}} = \frac{1}{n}\sum_{i=1}^{n}\widehat{\theta_i}, \quad \sqrt{\frac{\sum_{i=1}^{n}(\widehat{\theta_i} - \overline{\widehat{\theta}})^2}{n}}, \quad \frac{\sum_{i=1}^{n}(\widehat{\theta_i} - \theta)}{n}, \quad \frac{\sum_{i=1}^{n}(\widehat{\theta_i} - \theta)^2}{n},$$

respectively, where θ is the true parameter (each of $\boldsymbol{\xi} = (\xi_1, \xi_2)^\top$, $\boldsymbol{\Delta} = (\delta_{11}, \delta_{12}; \delta_{21}, \delta_{22})^\top$ and $\boldsymbol{\Omega} = \mathrm{diag}(\sigma_{11}, \sigma_{22})$) and $\widehat{\theta_i}$ is the estimate from the ith simulated sample.

Table 2 presents the results of the simulation, containing the average value (Mean), the corresponding standard deviations (Std.), Bias and MSE of the EM estimates of all the parameters of the EMMNE distribution in 300 simulated samples for each sample size. It can be seen that the Bias and MSE decrease as sample size n increases, thus varifying the asymptotic unbiasedness and consistency of the MLE.

Table 2. Mean, Std., Bias and MSE of the EM estimates over 300 samples from the MMNE model.

Sample size	Measure	$\boldsymbol{\xi} = (\xi_1, \xi_2)$		$\boldsymbol{\Sigma} = (\sigma_{11}, \sigma_{12}; \sigma_{21}, \sigma_{22})$				$\boldsymbol{\Delta} = (\delta_{11}, \delta_{12}; \delta_{21}, \delta_{22})$			
		ξ_1	ξ_2	σ_{11}	σ_{12}	σ_{21}	σ_{22}	δ_{11}	δ_{12}	δ_{21}	δ_{22}
50	Mean	5.0430	10.0164	0.3991	-0.0016	-0.0016	0.5838	0.2968	0.1000	0.2011	0.6986
	Std	0.0942	0.1112	0.0811	0.0671	0.0671	0.1279	0.0250	0.0367	0.0089	0.0200
	Bias	0.0043	0.0164	-0.0009	-0.0016	-0.0016	-0.0162	-0.0032	< 0.0001	0.0011	-0.0014
	MSE	0.0089	0.0126	0.0066	0.0045	0.0045	0.0166	0.0006	0.0013	0.0001	0.0004
100	Mean	5.0110	10.0064	0.3973	0.0022	0.0022	0.6050	0.2975	0.1031	0.2011	0.6983
	Std	0.0680	0.0810	0.0622	0.0461	0.0461	0.0946	0.0160	0.0251	0.0062	0.0135
	Bias	0.0110	0.0064	-0.0027	0.0022	0.0022	0.0050	-0.0025	0.0031	0.0011	-0.0017
	MSE	0.0047	0.0066	0.0039	0.0021	0.0021	0.0090	0.0003	0.0006	< 0.0001	0.0002
300	Mean	5.0077	9.9976	0.4005	-0.0004	-0.0004	0.6068	0.3018	0.0988	0.1998	0.7005
	Std	0.0392	0.0462	0.0309	0.0262	0.0262	0.0536	0.0107	0.01600	0.0039	0.0087
	Bias	0.0077	-0.0024	0.0005	-0.0004	-0.0004	0.0068	0.0018	-0.0012	-0.0002	0.0005
	MSE	0.0016	0.0021	0.0010	0.0007	0.0007	0.0029	0.0001	0.0003	< 0.0001	0.0001
500	Mean	5.0030	10.0027	0.4023	0.0047	0.0047	0.6009	0.3000	0.1013	0.2004	0.6994
	Std	0.0311	0.0350	0.0252	0.0216	0.0216	0.0428	0.0075	0.0123	0.0028	0.0061
	Bias	0.0030	0.0027	0.0023	0.0047	0.0047	0.0009	< 0.0001	0.0013	0.0004	-0.0006
	MSE	0.0010	0.0012	0.0006	0.0005	0.0005	0.0018	0.0001	0.0002	< 0.0001	< 0.0001

Table 3 shows the average computational time of the EM algorithm for each sample size, noting that the value would be influenced by the CPU of the computer. As the dimension of skewness parameter increasing, comparing the reult with the MMNE distribution in [1], the average computational time spent for the EMMNE distribution is quite long.

Table 3. Average run times (in seconds) for the EM algorithm.

Sample size (n)	Time
50	25.4162
100	50.9620
300	151.1491
500	242.2890

7. Real data example

In this section, we fit the EMMNE distribution for the Italian olive oil data which can be avilable in the R software, **pgmm** package. The data set consists of 572 observations of 10 columns. To support our illustration, we consider the first 323 observations of the 8th and 9th columns, the Linolenic and Arachidic fatty acids, respectively, and present the histograms in Figure 2.

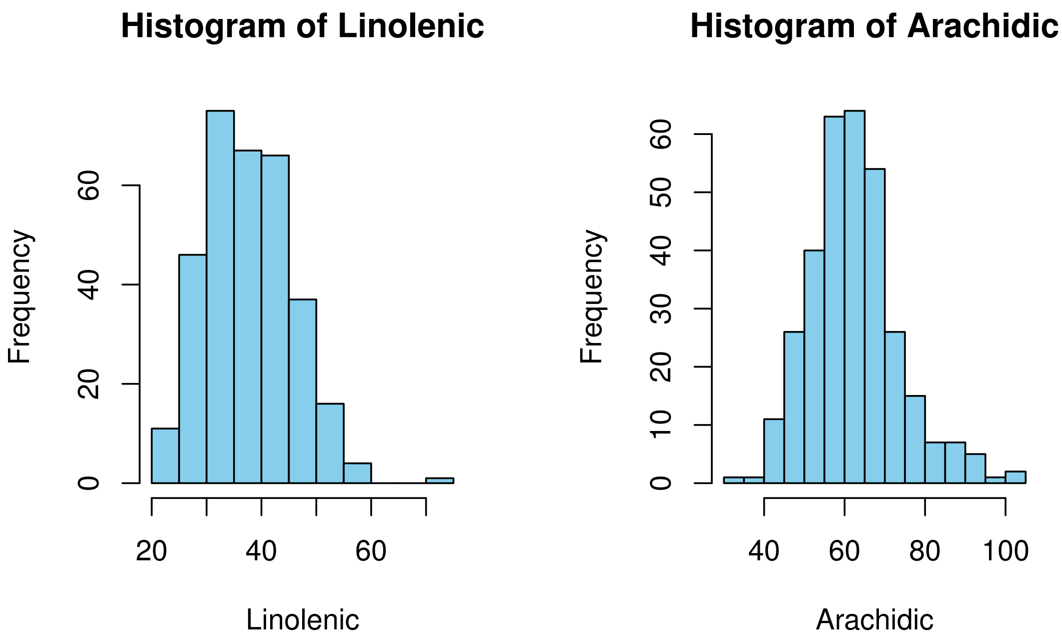


Figure 2. The histograms for the Linolenic and Arachidic fatty acids of olive oil data set.

To fit the EMMNE distribution, we consider the sample mean, sample covariance and take $\boldsymbol{\Delta} = (0.6546, -0.3750; 0.2000, 0.4977)$, which are given in Table 4, as the initial values in the EM algorithms for the parameters $\boldsymbol{\xi}, \boldsymbol{\Omega}$ and $\boldsymbol{\Delta}$, respectively. We compared the EMMNE distribution with the MMNE, SN and ST distributions in terms of the log likelihood values, AICs and BICs, and the results were presented in Table 5. As shown in Table 5, the EMMNE distribution produces the highest log-likelihood value, and smallest AIC and BIC values, hence, it provides a better fit than other distributions. Figure 3 shows the scatter plot of the real data and the contour plots of the fitted EMMNE, MMNE, SN and ST distributions.

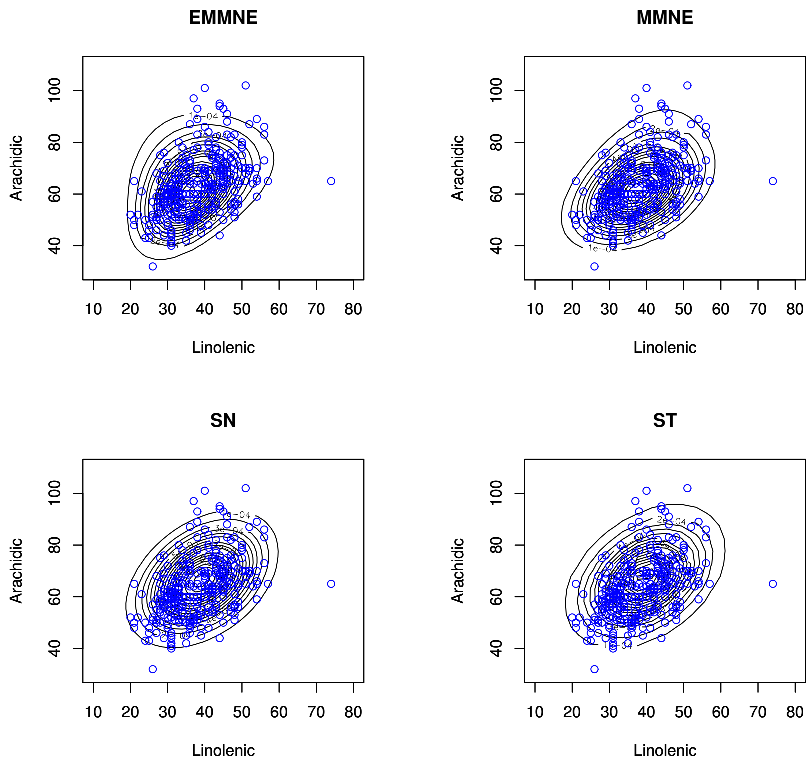


Figure 3. Scatter plots of the olive oil data, and the contour plots of the fitted EMMNE, MMNE, skew-normal (SN) and skew-t (ST) distributions.

Table 4. Initial values in the EM algorithm.

$\boldsymbol{\xi}$	$\boldsymbol{\Omega}$	$\boldsymbol{\Delta}$
$\begin{pmatrix} 38.0650 \\ 63.1177 \end{pmatrix}$	$\begin{pmatrix} 63.3673 & 40.9800 \\ 40.9800 & 124.2586 \end{pmatrix}$	$\begin{pmatrix} 0.6546 & -0.3750 \\ 0.2000 & 0.4977 \end{pmatrix}$

Table 5. Results of fitting the distributions. $l(\widehat{\boldsymbol{\theta}}|\boldsymbol{y})$: Log-likelihood value, AIC: Akaike information criterion, BIC: Bayesian information criterion.

Distribution	$l(\widehat{\boldsymbol{\theta}}|\boldsymbol{y})$	AIC	BIC
SN	-2320.039	4654.079	4680.522
ST	-2316.320	4648.640	4678.861
MMNE	-2314.604	4643.207	4669.651
EMMNE	-2291.114	4600.228	4634.227

Table 6. Parameter estimates of the EMMNE distribution by using EM Algorithm.

$\widehat{\boldsymbol{\xi}}$	$\widehat{\boldsymbol{\Omega}}$	$\widehat{\boldsymbol{\Delta}}$
$\begin{pmatrix} 34.5675 \\ 55.6717 \end{pmatrix}$	$\begin{pmatrix} 61.4410 & 32.2205 \\ 32.2205 & 136.2856 \end{pmatrix}$	$\begin{pmatrix} 0.6529 & -0.3444 \\ 0.1736 & 0.5144 \end{pmatrix}$

8. MTCE for EMMN distribution

In this section, we introduce the important risk measure: The multivariate tail conditional expectation (MTCE) and calculate its expression for EMMN distributions. Briefly, a risk measure could be seen as a mapping from a set of random variables associating the risks to the real line. A comprehensive study of risk theorey can be found in the book of Denult et al. [13].

Definition 3. *For a univariate random variable X, the tail conditional expectation (TCE) is defined by*

$$TCE_q(X) = E\left[X|X > VaR_q(X)\right], \quad q \in (0,1),$$

where $VaR_q(X) = F_X^{-1}(q) = \inf\{x \in \mathbb{R}|F_X(x) \geq q\} = \sup\{x \in \mathbb{R}|F_X(x) < q\}$*, which is known as the value at risk (VaR) measure.*

Definition 4. *For an n-dimensional random vector* $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)^\top$*, the multivariate tail conditional expectation (MTCE) introduced by Landsman et al. [16] is as follows,*

$$\begin{aligned} MTCE_{\boldsymbol{q}}(\boldsymbol{X}) &= E\left[\boldsymbol{X}|\boldsymbol{X} > VaR_{\boldsymbol{q}}(\boldsymbol{X})\right] \\ &= E\left[\boldsymbol{X}|X_1 > VaR_{q_1}(X_1), \ldots, X_n > VaR_{q_n}(X_n)\right], \quad \boldsymbol{q} = (q_1, \ldots, q_n) \in (0,1)^n. \end{aligned}$$

Here $VaR_{\boldsymbol{q}}(\boldsymbol{X})$ *is the* $n \times 1$ *vector* $(VaR_{q_1}(X_1), VaR_{q_2}(X_2), \ldots, VaR_{q_n}(X_n))^\top$.

Theorem 9. *Let* $\boldsymbol{Y} \sim EMMN_{d,m}(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\Delta}; H)$*. Then, the MTCE of* $\boldsymbol{Y}$ *is given by*

$$MTCE_{\boldsymbol{q}}(\boldsymbol{Y}) = \frac{1}{\bar{F}_{\boldsymbol{Y}}(\boldsymbol{y_q})}\left\{\left(\boldsymbol{\Omega} - \omega\boldsymbol{\Delta}\boldsymbol{\Delta}^\top\omega\right)^{\frac{1}{2}}\Lambda_{\boldsymbol{q}}(\boldsymbol{u}) + \boldsymbol{\xi}E_{\boldsymbol{U}}\left[\Phi_d(-\boldsymbol{\kappa_q}(\boldsymbol{U}))\right] + \omega\boldsymbol{\Delta}E_{\boldsymbol{U}}\left[\boldsymbol{U}\Phi_d(-\boldsymbol{\kappa_q}(\boldsymbol{U}))\right]\right\},$$

where

$$\begin{aligned} \Lambda_{\boldsymbol{q}}(\boldsymbol{u}) &= (\lambda_{1,q_1}, \lambda_{2,q_2}, \ldots, \lambda_{d,q_d})^\top, \\ \lambda_{i,\boldsymbol{q}} &= E_{\boldsymbol{U}}\left[\phi\left(\boldsymbol{\kappa}_{\boldsymbol{q},i}(\boldsymbol{u})\right)\Phi_{d-1}\left(-\boldsymbol{\kappa}_{\boldsymbol{q},-i}(\boldsymbol{u})\right)\right], \quad i \in \{1, \ldots, d\}, \\ \boldsymbol{\kappa_q}(\boldsymbol{u}) &= \left(\kappa_{q_1,1}(\boldsymbol{u}), \kappa_{q_2,2}(\boldsymbol{u}), \ldots, \kappa_{q_d,d}(\boldsymbol{u})\right)^\top \\ &= \left(\boldsymbol{\Omega} - \omega\boldsymbol{\Delta}\boldsymbol{\Delta}^\top\omega\right)^{-\frac{1}{2}}\left(VaR_{\boldsymbol{q}}(\boldsymbol{Y}) - \boldsymbol{\xi} - \omega\boldsymbol{\Delta}\boldsymbol{u}\right), \\ \boldsymbol{\kappa}_{\boldsymbol{q},-i}(\boldsymbol{u}) &= \left(\kappa_{q_1,1}(\boldsymbol{u}), \ldots, \kappa_{q_{i-1},i-1}(\boldsymbol{u}), \kappa_{q_{i+1},i+1}(\boldsymbol{u}), \ldots, \kappa_{q_d,d}(\boldsymbol{u})\right)^\top. \end{aligned}$$

Proof. By (2.2), note that

$$
\begin{aligned}
MTCE_{\boldsymbol{q}}(\boldsymbol{Y}) &= E\left[\boldsymbol{Y}|Y_1 > VaR_{q_1}(Y_1), \dots, Y_d > VaR_{q_d}(Y_d)\right] \\
&= \frac{1}{\bar{F}_{\boldsymbol{Y}}(\boldsymbol{y_q})} \int_{VaR_{\boldsymbol{q}(\boldsymbol{Y})}}^{+\infty} \boldsymbol{y} \int_{\mathbb{R}^m} \phi_d(\boldsymbol{y}; \boldsymbol{\xi} + \boldsymbol{\omega\Delta u}, \boldsymbol{\Omega} - \boldsymbol{\omega\Delta\Delta}^\top \boldsymbol{\omega}) h(\boldsymbol{u}; \boldsymbol{v})\, \mathrm{d}\boldsymbol{u}\, \mathrm{d}\boldsymbol{y},
\end{aligned}
$$

where $\bar{F}_{\boldsymbol{Y}}(\boldsymbol{y_q}) = P(\boldsymbol{Y} > \boldsymbol{y_q})$, and the $\int_{VaR_{\boldsymbol{q}(\boldsymbol{Y})}}^{+\infty} \mathrm{d}\boldsymbol{y}$ is multi-dimensional integrals.

Let $\boldsymbol{t} = \left(\boldsymbol{\Omega} - \boldsymbol{\omega\Delta\Delta}^\top \boldsymbol{\omega}\right)^{-\frac{1}{2}} (\boldsymbol{y} - \boldsymbol{\xi} - \boldsymbol{\omega\Delta u})$, then

$$
\begin{aligned}
MTCE_{\boldsymbol{q}}(\boldsymbol{Y}) &= \frac{1}{\bar{F}_{\boldsymbol{Y}}(\boldsymbol{y_q})} \int_{\mathbb{R}^m} \int_{\boldsymbol{\kappa_q}(\boldsymbol{u})}^{+\infty} \left[\left(\boldsymbol{\Omega} - \boldsymbol{\omega\Delta\Delta}^\top \boldsymbol{\omega}\right)^{\frac{1}{2}} \boldsymbol{t} + (\boldsymbol{\xi} + \boldsymbol{\omega\Delta u})\right] \phi_d(\boldsymbol{t}; \boldsymbol{0}, \boldsymbol{I}_d) h(\boldsymbol{u}; \boldsymbol{v})\, \mathrm{d}\boldsymbol{t}\, \mathrm{d}\boldsymbol{u} \\
&= \frac{1}{\bar{F}_{\boldsymbol{Y}}(\boldsymbol{y_q})} \left\{\left(\boldsymbol{\Omega} - \boldsymbol{\omega\Delta\Delta}^\top \boldsymbol{\omega}\right)^{\frac{1}{2}} \int_{\mathbb{R}^m} \int_{\boldsymbol{\kappa_q}(\boldsymbol{u})}^{+\infty} \boldsymbol{t} \phi_d(\boldsymbol{t}; \boldsymbol{0}, \boldsymbol{I}_d) h(\boldsymbol{u}; \boldsymbol{v})\, \mathrm{d}\boldsymbol{t}\, \mathrm{d}\boldsymbol{u} \right. \\
&\quad \left. + \boldsymbol{\xi} E_{\boldsymbol{U}}\left[\Phi_d(-\boldsymbol{\kappa_q}(\boldsymbol{U}))\right] + \boldsymbol{\omega\Delta} E_{\boldsymbol{U}}\left[\boldsymbol{U}\Phi_d(-\boldsymbol{\kappa_q}(\boldsymbol{U}))\right]\right\},
\end{aligned}
$$

where $\int_{\boldsymbol{\kappa_q}(\boldsymbol{u})}^{+\infty} \mathrm{d}\boldsymbol{t}$ is multi-dimensional integrals. And for

$$
\Lambda_{\boldsymbol{q}}(\boldsymbol{u}) = \int_{\mathbb{R}^m} \int_{\boldsymbol{\kappa_q}(\boldsymbol{u})}^{+\infty} \boldsymbol{t} \phi_d(\boldsymbol{t}; \boldsymbol{0}, \boldsymbol{I}_d) h(\boldsymbol{u}; \boldsymbol{v})\, \mathrm{d}\boldsymbol{t}\, \mathrm{d}\boldsymbol{u},
$$

by

$$
\int_{\boldsymbol{\kappa_q}(\boldsymbol{u})}^{+\infty} \phi_d(\boldsymbol{t}; \boldsymbol{0}, \boldsymbol{I}_d)\, \mathrm{d}\boldsymbol{t} = \Phi_d(-\boldsymbol{\kappa_q}(\boldsymbol{u})),
$$

we conclude that

$$
\begin{aligned}
\lambda_{i,\boldsymbol{q}} &= \int_{\mathbb{R}^m} \int_{\boldsymbol{\kappa_q}(\boldsymbol{u})}^{+\infty} t_i \phi_d(\boldsymbol{t}; \boldsymbol{0}, \boldsymbol{I}_d) h(\boldsymbol{u}; \boldsymbol{v})\, \mathrm{d}\boldsymbol{t}\, \mathrm{d}\boldsymbol{u} \\
&= \int_{\mathbb{R}^m} \int_{\boldsymbol{\kappa}_{\boldsymbol{q},i}(\boldsymbol{u})}^{\infty} t_i \frac{e^{-\frac{1}{2}t_i^2}}{\sqrt{2\pi}}\, \mathrm{d}t_i \int_{\boldsymbol{\kappa}_{\boldsymbol{q},-i}(\boldsymbol{u})}^{+\infty} \frac{e^{-\frac{1}{2}\boldsymbol{t}_{d-1,-i}^\top \boldsymbol{t}_{d-1,-i}}}{(2\pi)^{\frac{d-1}{2}}}\, \mathrm{d}\boldsymbol{t}_{d-1,-i}\, h(\boldsymbol{u}; \boldsymbol{v})\, \mathrm{d}\boldsymbol{u} \\
&= E_{\boldsymbol{U}}\left[\phi\left(\boldsymbol{\kappa}_{\boldsymbol{q},i}(\boldsymbol{u})\right) \Phi_{d-1}\left(-\boldsymbol{\kappa}_{\boldsymbol{q},-i}(\boldsymbol{u})\right)\right],
\end{aligned}
$$

where $\boldsymbol{t} = (t_1, \dots, t_d)^\top$, $\boldsymbol{t}_{d-1,-i} = (t_1, \dots, t_{i-1}, t_{i+1}, \dots t_d)^\top$, thus completes the proof of the result. □

Corollary 1. *When $m = 1$, $U \sim Exp(1)$, $\boldsymbol{Y}$ degenerates to the MMNE distribution, and denoted by $\boldsymbol{Y} \sim MMNE_d(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\Delta})$ (see [1]), then*

$$
MTCE_{\boldsymbol{q}}(\boldsymbol{Y}) = \frac{1}{\bar{F}_Y(\boldsymbol{y_q})} \left\{\left(\boldsymbol{\Omega} - \boldsymbol{\omega\Delta\Delta}^\top \boldsymbol{\omega}\right)^{\frac{1}{2}} \Lambda_{\boldsymbol{q}}(u) + \boldsymbol{\xi} E_U\left[\Phi_d(-\boldsymbol{\kappa_q}(U))\right] + \boldsymbol{\omega\Delta} E_U\left[U\Phi_d(-\boldsymbol{\kappa_q}(U))\right]\right\},
$$

where,

$$
\lambda_{i,\boldsymbol{q}} = \frac{\exp\left\{\frac{1-2\zeta_i\beta_i}{2\beta_i^2}\right\}}{\beta_i} \Phi\left(\frac{\zeta_i\beta_i - 1}{\beta_i}\right) E_T\left[\Phi_{d-1}\left(-\boldsymbol{\kappa}_{\boldsymbol{q},-i}\left(\frac{T}{\beta_i} + \frac{\zeta_i\beta_i - 1}{\beta_i^2}\right)\right)\right], \quad i \in \{1, \dots, d\},
$$

$T \sim TN_1(-\frac{\zeta_i\beta_i-1}{\beta_i}, +\infty; 0, 1)$,

$$\boldsymbol{\zeta} = (\zeta_1, \ldots, \zeta_d)^\top = \left(\boldsymbol{\Omega} - \boldsymbol{\omega}\boldsymbol{\Delta}\boldsymbol{\Delta}^\top\boldsymbol{\omega}\right)^{-\frac{1}{2}} \left[VaR_{\boldsymbol{q}}(\boldsymbol{Y}) - \boldsymbol{\xi}\right],$$

$$\boldsymbol{\beta} = (\beta_1, \ldots, \beta_d)^\top = \left(\boldsymbol{\Omega} - \boldsymbol{\omega}\boldsymbol{\Delta}\boldsymbol{\Delta}^\top\boldsymbol{\omega}\right)^{-\frac{1}{2}} \boldsymbol{\omega}\boldsymbol{\Delta}.$$

Proof. By using Theorem 9,

$$\begin{aligned}
\lambda_{i,\boldsymbol{q}} &= E_U\left[\phi\left(\boldsymbol{\kappa}_{\boldsymbol{q},i}(U)\right)\Phi_{d-1}\left(-\boldsymbol{\kappa}_{\boldsymbol{q},-i}(U)\right)\right] \\
&= \int_0^{+\infty} \phi\left(\boldsymbol{\kappa}_{\boldsymbol{q},i}(u)\right)\Phi_{d-1}\left(-\boldsymbol{\kappa}_{\boldsymbol{q},-i}(u)\right)e^{-u}\,\mathrm{d}u \\
&= \int_0^{+\infty} \phi\left(\zeta_i - \beta_i u\right)\Phi_{d-1}\left(-\boldsymbol{\kappa}_{\boldsymbol{q},-i}(u)\right)e^{-u}\,\mathrm{d}u \\
&= \int_0^{+\infty} \frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{1}{2}(\zeta_i - \beta_i u)^2 - u\right\}\Phi_{d-1}\left(-\boldsymbol{\kappa}_{\boldsymbol{q},-i}(u)\right)\mathrm{d}u,
\end{aligned}$$

writing $\beta_i\left(u - \frac{\zeta_i\beta_i-1}{\beta_i^2}\right) = t$, then

$$\begin{aligned}
\lambda_{i,\boldsymbol{q}} &= \int_{-\frac{\zeta_i\beta_i-1}{\beta_i}}^{+\infty} \frac{1}{\beta_i}\exp\left\{\frac{1-2\zeta_i\beta_i}{2\beta_i^2}\right\}\phi(t)\Phi_{d-1}\left(-\boldsymbol{\kappa}_{\boldsymbol{q},-i}\left(\frac{t}{\beta_i} + \frac{\zeta_i\beta_i - 1}{\beta_i^2}\right)\right)\mathrm{d}t \\
&= \frac{\exp\left\{\frac{1-2\zeta_i\beta_i}{2\beta_i^2}\right\}}{\beta_i}\Phi\left(\frac{\zeta_i\beta_i-1}{\beta_i}\right)E_T\left[\Phi_{d-1}\left(-\boldsymbol{\kappa}_{\boldsymbol{q},-i}\left(\frac{T}{\beta_i} + \frac{\zeta_i\beta_i-1}{\beta_i^2}\right)\right)\right].
\end{aligned}$$

□

Corollary 2. *When $m = 1$, $U \sim Ga(\nu, 1)$, $\boldsymbol{Y}$ degenerates to the MMNG distribution, and denoted by $\boldsymbol{Y} \sim MMNG_d(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\Delta}, \nu)$ (see [1]), then*

$$MTCE_{\boldsymbol{q}}(\boldsymbol{Y}) = \frac{1}{\bar{F}_{\boldsymbol{Y}}(\boldsymbol{y}_{\boldsymbol{q}})}\left\{\left(\boldsymbol{\Omega} - \boldsymbol{\omega}\boldsymbol{\Delta}\boldsymbol{\Delta}^\top\boldsymbol{\omega}\right)^{\frac{1}{2}}\Lambda_{\boldsymbol{q}}(u) + \boldsymbol{\xi}E_U\left[\Phi_d(-\boldsymbol{\kappa}_{\boldsymbol{q}}(U))\right] + \boldsymbol{\omega}\boldsymbol{\Delta}E_U\left[U\Phi_d(-\boldsymbol{\kappa}_{\boldsymbol{q}}(U))\right]\right\},$$

where

$$\lambda_{i,\boldsymbol{q}} = c_i E_T\left[\left(\frac{T}{\beta_i} + \frac{\zeta_i\beta_i-1}{\beta_i^2}\right)^{\nu-1}\Phi_{d-1}\left(-\boldsymbol{\kappa}_{\boldsymbol{q},-i}\left(\frac{T}{\beta_i} + \frac{\zeta_i\beta_i-1}{\beta_i^2}\right)\right)\right],\ \ i \in \{1, \ldots, d\},$$

$$c_i = \frac{\exp\left\{\frac{1-2\zeta_i\beta_i}{2\beta_i^2}\right\}}{\Gamma(\nu)\beta_i}\Phi\left(\frac{\zeta_i\beta_i-1}{\beta_i}\right),$$

$T \sim TN_1(-\frac{\zeta_i\beta_i-1}{\beta_i}, +\infty; 0, 1)$*, $\boldsymbol{\zeta} = (\zeta_1, \ldots, \zeta_d)^\top$, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_d)^\top$ are the same as those in Corllary 1.*

Proof. By using Theorem 9,

$$
\begin{aligned}
\lambda_{i,\boldsymbol{q}} &= E_U\left[\phi\left(\boldsymbol{\kappa}_{\boldsymbol{q},i}(U)\right)\Phi_{d-1}\left(-\boldsymbol{\kappa}_{\boldsymbol{q},-i}(U)\right)\right]\\
&= \int_0^{+\infty}\phi\left(\boldsymbol{\kappa}_{\boldsymbol{q},i}(u)\right)\Phi_{d-1}\left(-\boldsymbol{\kappa}_{\boldsymbol{q},-i}(u)\right)\frac{1}{\Gamma(\nu)}u^{\nu-1}e^{-u}\,\mathrm{d}u\\
&= \int_0^{+\infty}\phi\left(\zeta_i-\beta_i u\right)\Phi_{d-1}\left(-\boldsymbol{\kappa}_{\boldsymbol{q},-i}(u)\right)\frac{1}{\Gamma(\nu)}u^{\nu-1}e^{-u}\,\mathrm{d}u\\
&= \int_0^{+\infty}\frac{1}{\Gamma(\nu)}\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{1}{2}(\zeta_i-\beta_i u)^2-u\right\}u^{\nu-1}\Phi_{d-1}\left(-\boldsymbol{\kappa}_{\boldsymbol{q},-i}(u)\right)\mathrm{d}u,
\end{aligned}
$$

writing $\beta_i\left(u-\frac{\zeta_i\beta_i-1}{\beta_i^2}\right)=t$, then

$$
\begin{aligned}
\lambda_{i,\boldsymbol{q}} &= \int_{-\frac{\zeta_i\beta_i-1}{\beta_i}}^{+\infty}\frac{\exp\left\{\frac{1-2\zeta_i\beta_i}{2\beta_i^2}\right\}}{\Gamma(\nu)\beta_i}\Phi\left(\frac{\zeta_i\beta_i-1}{\beta_i}\right)\phi(t)\left(\frac{t}{\beta_i}+\frac{\zeta_i\beta_i-1}{\beta_i^2}\right)^{\nu-1}\\
&\quad\times\Phi_{d-1}\left(-\boldsymbol{\kappa}_{\boldsymbol{q},-i}\left(\frac{t}{\beta_i}+\frac{\zeta_i\beta_i-1}{\beta_i^2}\right)\right)\mathrm{d}t\\
&= c_iE_T\left[\left(\frac{T}{\beta_i}+\frac{\zeta_i\beta_i-1}{\beta_i^2}\right)^{\nu-1}\Phi_{d-1}\left(-\boldsymbol{\kappa}_{\boldsymbol{q},-i}\left(\frac{T}{\beta_i}+\frac{\zeta_i\beta_i-1}{\beta_i^2}\right)\right)\right].
\end{aligned}
$$

□

Corollary 3. *When $m=1$, $U\sim TN_1(0,+\infty;0,1)$, then $\boldsymbol{Y}\sim SN_d(\boldsymbol{\xi},\boldsymbol{\Omega},\boldsymbol{\Delta})$, and we have*

$$
MTCE_{\boldsymbol{q}}(\boldsymbol{Y})=\frac{1}{\bar{F}_{\boldsymbol{Y}}(\boldsymbol{y_q})}\left\{\left(\boldsymbol{\Omega}-\boldsymbol{\omega}\boldsymbol{\Delta}\boldsymbol{\Delta}^{\top}\boldsymbol{\omega}\right)^{\frac{1}{2}}\Lambda_{\boldsymbol{q}}(u)+\boldsymbol{\xi}E_U\left[\Phi_d(-\boldsymbol{\kappa}_{\boldsymbol{q}}(U))\right]+\boldsymbol{\omega}\boldsymbol{\Delta}E_U\left[U\Phi_d(-\boldsymbol{\kappa}_{\boldsymbol{q}}(U))\right]\right\},
$$

where

$$
\lambda_{i,\boldsymbol{q}}=\frac{\sqrt{2}\exp\left\{-\frac{\zeta_i^2}{2(1+\beta_i^2)}\right\}}{\sqrt{\pi(1+\beta_i^2)}}\Phi\left(\frac{\zeta_i\beta_i}{1+\beta_i^2}\right)E_T\left[\Phi_{d-1}\left(-\boldsymbol{\kappa}_{\boldsymbol{q},-i}\left(\frac{T}{\sqrt{1+\beta_i^2}}+\frac{\zeta_i\beta_i}{1+\beta_i^2}\right)\right)\right],\quad i\in\{1,\ldots,d\},
$$

$T\sim TN_1(-\frac{\zeta_i\beta_i}{1+\beta_i^2},+\infty;0,1)$, $\boldsymbol{\zeta}=(\zeta_1,\ldots,\zeta_d)^{\top}$, $\boldsymbol{\beta}=(\beta_1,\ldots,\beta_d)^{\top}$ are the same as those in Corollary 1.

Proof. By using Theorem 9,

$$
\begin{aligned}
\lambda_{i,\boldsymbol{q}} &= E_U\left[\phi\left(\boldsymbol{\kappa}_{\boldsymbol{q},i}(U)\right)\Phi_{d-1}\left(-\boldsymbol{\kappa}_{\boldsymbol{q},-i}(U)\right)\right]\\
&= 2\int_0^{+\infty}\phi\left(\boldsymbol{\kappa}_{\boldsymbol{q},i}(u)\right)\Phi_{d-1}\left(-\boldsymbol{\kappa}_{\boldsymbol{q},-i}(u)\right)\phi(u)\,\mathrm{d}u\\
&= 2\int_0^{+\infty}\phi\left(\zeta_i-\beta_i u\right)\Phi_{d-1}\left(-\boldsymbol{\kappa}_{\boldsymbol{q},-i}(u)\right)\phi(u)\,\mathrm{d}u\\
&= \int_0^{+\infty}\frac{1}{\pi}\exp\left\{-\frac{1}{2}(\zeta_i-\beta_i u)^2-\frac{1}{2}u^2\right\}\Phi_{d-1}\left(-\boldsymbol{\kappa}_{\boldsymbol{q},-i}(u)\right)\mathrm{d}u,
\end{aligned}
$$

writing $\frac{\left(u-\frac{\zeta_i\beta_i}{1+\beta_i^2}\right)}{\frac{1}{\sqrt{1+\beta_i^2}}}=t$, then

$$\begin{aligned}
\lambda_{i,\boldsymbol{q}} &= \int_{-\frac{\zeta_i\beta_i}{1+\beta_i^2}}^{+\infty} \sqrt{\frac{2}{\pi(1+\beta_i^2)}} \exp\left\{-\frac{\zeta_i^2}{2(1+\beta_i^2)}\right\} \phi(t)\Phi_{d-1}\left(-\boldsymbol{\kappa}_{\boldsymbol{q},-i}\left(\frac{t}{\sqrt{1+\beta_i^2}}+\frac{\zeta_i\beta_i}{1+\beta_i^2}\right)\right)\mathrm{d}t \\
&= \frac{\sqrt{2}\exp\left\{-\frac{\zeta_i^2}{2(1+\beta_i^2)}\right\}}{\sqrt{\pi(1+\beta_i^2)}}\Phi\left(\frac{\zeta_i\beta_i}{1+\beta_i^2}\right)E_T\left[\Phi_{d-1}\left(-\boldsymbol{\kappa}_{\boldsymbol{q},-i}\left(\frac{T}{\sqrt{1+\beta_i^2}}+\frac{\zeta_i\beta_i}{1+\beta_i^2}\right)\right)\right].
\end{aligned}$$

□

Remark 6. *When $\boldsymbol{\xi}=\mathbf{0}$, $\boldsymbol{\Omega}=\bar{\boldsymbol{\Omega}}$, the results of Corollary 3 coincide with the MTCE of multivariate skew-normal distribution in Mousavi et al. [20].*

Corollary 4. *When $\boldsymbol{\Delta}=\boldsymbol{O}$, is a zero matrix, $\boldsymbol{Y}$ degenerates to the multivariate normal distribution, and denoted by $\boldsymbol{Y}\sim N_d(\boldsymbol{\xi},\boldsymbol{\Omega})$, then*

$$MTCE_{\boldsymbol{q}}(\boldsymbol{Y})=\boldsymbol{\xi}+\boldsymbol{\Omega}^{\frac{1}{2}}\Upsilon_{\boldsymbol{q}},$$

where, $\Upsilon_{\boldsymbol{q}}=(\upsilon_1,\ldots,\upsilon_d)^\top$,

$$\upsilon_i=\frac{\phi\left(\boldsymbol{\kappa}_{\boldsymbol{q},i}\right)\Phi_{d-1}\left(-\boldsymbol{\kappa}_{\boldsymbol{q},-i}\right)}{\bar{\Phi}_d(\boldsymbol{\kappa}_{\boldsymbol{q}})}=\frac{\phi\left(\boldsymbol{\kappa}_{\boldsymbol{q},i}\right)\bar{\Phi}_{d-1}\left(\boldsymbol{\kappa}_{\boldsymbol{q},-i}\right)}{\bar{\Phi}_d(\boldsymbol{\kappa}_{\boldsymbol{q}})},\quad i\in\{1,\ldots,d\}.$$

Proof. By using Theorem 9, and when $\boldsymbol{\Delta}=\boldsymbol{O}$,

$$\begin{aligned}
MTCE_{\boldsymbol{q}}(\boldsymbol{Y}) &= \frac{1}{\bar{F}_{\boldsymbol{Y}}(\boldsymbol{y}_{\boldsymbol{q}})}\left\{\boldsymbol{\Omega}^{\frac{1}{2}}\Lambda_{\boldsymbol{q}}(\boldsymbol{u})+\boldsymbol{\xi}E_{\boldsymbol{U}}\left[\Phi_d(-\boldsymbol{\kappa}_{\boldsymbol{q}}(\boldsymbol{u}))\right]\right\} \\
\Lambda_{\boldsymbol{q}}(\boldsymbol{u}) &= \Lambda_{\boldsymbol{q}}=(\lambda_{1,q_1},\lambda_{2,q_2},\ldots,\lambda_{d,q_d})^\top, \\
\lambda_{i,\boldsymbol{q}} &= \phi\left(\boldsymbol{\kappa}_{\boldsymbol{q},i}(\boldsymbol{u})\right)\Phi_{d-1}\left(-\boldsymbol{\kappa}_{\boldsymbol{q},-i}(\boldsymbol{u})\right)=\phi\left(\boldsymbol{\kappa}_{\boldsymbol{q},i}\right)\Phi_{d-1}\left(-\boldsymbol{\kappa}_{\boldsymbol{q},-i}\right),\quad i\in\{1,\ldots,d\}, \\
\boldsymbol{\kappa}_{\boldsymbol{q}}(\boldsymbol{u}) &= \boldsymbol{\kappa}_{\boldsymbol{q}}=\boldsymbol{\Omega}^{-\frac{1}{2}}\left(VaR_{\boldsymbol{q}}(\boldsymbol{Y})-\boldsymbol{\xi}\right), \\
\boldsymbol{\kappa}_{\boldsymbol{q},-i} &= \left(\kappa_{q_1,1},\ldots,\kappa_{q_{i-1},i-1},\kappa_{q_{i+1},i+1},\ldots,\kappa_{q_d,d}\right)^\top.
\end{aligned}$$

By replacing $\bar{F}_{\boldsymbol{Y}}(\boldsymbol{y}_{\boldsymbol{q}})=\bar{\Phi}_d(\boldsymbol{\kappa}_{\boldsymbol{q}})$, $E_{\boldsymbol{U}}\left[\Phi_d(-\boldsymbol{\kappa}_{\boldsymbol{q}}(\boldsymbol{u}))\right]=\Phi_d(-\boldsymbol{\kappa}_{\boldsymbol{q}})=\bar{\Phi}_d(\boldsymbol{\kappa}_{\boldsymbol{q}})$, we complete the proof of the result. □

Remark 7. *The results of Corollary 4 coincide with the MTCE of the multivariate normal distribution in Landsman et al. [16].*

9. Conclusions

In this paper, we have presented the extended mean mixtures of multivariate normal (EMMN) distributions, which include the MMN, SN and the multivariate normal distribution as special cases. We have explored some basic properties of this family of distributions, including characteristic function, moment generating function, the distributions of affine transformations and canonical forms. In addition, we have developed a general EM-type algorithm for estimating parameters in the EMMN distribution. Two special cases of the EMMN family corresponding to the mixtures of multi-standard gamma and multi-standard exponential distributions have been studied in detail. Numerical results from simulation study have been shown the performance of the MLE of parameters in the EMMNE distribution. The results of the fitting of real data sets revealed that the EMMNE distribution perform better in fitting compared to the MMNE, SN and ST distributions. Finally, we introduced the risk measure theory of tail conditional expectation, and derived the theoretical formula of the multivariate tail conditional expectation (MTCE) of the EMMN distribution.

Acknowledgments

Thanks to the anonymous reviewers for their careful reading of the manuscript, correction of errors and many insightful comments and suggestions on the earlier version of this paper. The research was supported by the National Natural Science Foundation of China (Nos.12071251, 11571198).

Conflict of interest

No potential competing interest was reported by the authors.

Appendix

Proof of Theorem 7. Note that

$$f_{U|Y}(\boldsymbol{u}) = \frac{f_U(\boldsymbol{u}) f(\boldsymbol{y}|\boldsymbol{u})}{\int_{\mathbb{R}^m_+} f_U(\boldsymbol{u}) f(\boldsymbol{y}|\boldsymbol{u})\, \mathrm{d}\boldsymbol{u}},$$

where

$$f_U(\boldsymbol{u}) f(\boldsymbol{y}|\boldsymbol{u}) = (2\pi)^{-\frac{d}{2}} |\boldsymbol{\Sigma}_Y|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(\boldsymbol{y} - \boldsymbol{\xi} - \omega\boldsymbol{\Delta u})^\top \boldsymbol{\Sigma}_Y^{-1} (\boldsymbol{y} - \boldsymbol{\xi} - \omega\boldsymbol{\Delta u})\right\} \exp\left\{-\boldsymbol{u}^\top \mathbf{1}_m\right\}, \tag{a.1}$$

and

$$\begin{aligned} &-\frac{1}{2}(\boldsymbol{y} - \boldsymbol{\xi} - \omega\boldsymbol{\Delta u})^\top \boldsymbol{\Sigma}_Y^{-1} (\boldsymbol{y} - \boldsymbol{\xi} - \omega\boldsymbol{\Delta u}) \\ =& -\frac{1}{2}\left[(\boldsymbol{y} - \boldsymbol{\xi})^\top \boldsymbol{\Sigma}_Y^{-1} (\boldsymbol{y} - \boldsymbol{\xi}) - \boldsymbol{u}^\top \boldsymbol{\Delta}^\top \omega \boldsymbol{\Sigma}_Y^{-1} (\boldsymbol{y} - \boldsymbol{\xi}) - (\boldsymbol{y} - \boldsymbol{\xi})^\top \boldsymbol{\Sigma}_Y^{-1} \omega \boldsymbol{\Delta u} + \boldsymbol{u}^\top \boldsymbol{\Delta}^\top \omega \boldsymbol{\Sigma}_Y^{-1} \omega \boldsymbol{\Delta u}\right], \end{aligned}$$

then, (a.1) turns to

$$\phi_d(\boldsymbol{y}; \boldsymbol{\xi}, \boldsymbol{\Sigma}_Y) \exp\left\{-\frac{1}{2}\left[-\boldsymbol{u}^\top \boldsymbol{\Delta}^\top \omega \boldsymbol{\Sigma}_Y^{-1} (\boldsymbol{y} - \boldsymbol{\xi}) - (\boldsymbol{y} - \boldsymbol{\xi})^\top \boldsymbol{\Sigma}_Y^{-1} \omega \boldsymbol{\Delta u} + \boldsymbol{u}^\top \boldsymbol{\Delta}^\top \omega \boldsymbol{\Sigma}_Y^{-1} \omega \boldsymbol{\Delta u} + \boldsymbol{u}^\top \mathbf{1}_m + \mathbf{1}_m^\top \boldsymbol{u}\right]\right\}, \tag{a.2}$$

where

$$
\begin{aligned}
& -\boldsymbol{u}^\top\boldsymbol{\Delta}^\top\omega\boldsymbol{\Sigma}_{\boldsymbol{Y}}^{-1}(\boldsymbol{y}-\boldsymbol{\xi})-(\boldsymbol{y}-\boldsymbol{\xi})^\top\boldsymbol{\Sigma}_{\boldsymbol{Y}}^{-1}\omega\boldsymbol{\Delta}\boldsymbol{u}+\boldsymbol{u}^\top\boldsymbol{\Delta}^\top\omega\boldsymbol{\Sigma}_{\boldsymbol{Y}}^{-1}\omega\boldsymbol{\Delta}\boldsymbol{u}+\boldsymbol{u}^\top\mathbf{1}_m+\mathbf{1}_m^\top\boldsymbol{u} \\
&= -\boldsymbol{u}^\top\left[\boldsymbol{\Delta}^\top\omega\boldsymbol{\Sigma}_{\boldsymbol{Y}}^{-1}(\boldsymbol{y}-\boldsymbol{\xi})-\mathbf{1}_m\right]-\left[(\boldsymbol{y}-\boldsymbol{\xi})^\top\boldsymbol{\Sigma}_{\boldsymbol{Y}}^{-1}\omega\boldsymbol{\Delta}-\mathbf{1}_m^\top\right]\boldsymbol{u}+\boldsymbol{u}^\top\boldsymbol{\Delta}^\top\omega\boldsymbol{\Sigma}_{\boldsymbol{Y}}^{-1}\omega\boldsymbol{\Delta}\boldsymbol{u} \\
&= \left\{\boldsymbol{u}-\left(\boldsymbol{\Delta}^\top\omega\boldsymbol{\Sigma}_{\boldsymbol{Y}}^{-1}\omega\boldsymbol{\Delta}\right)^{-1}\left[\boldsymbol{\Delta}^\top\omega\boldsymbol{\Sigma}_{\boldsymbol{Y}}^{-1}(\boldsymbol{y}-\boldsymbol{\xi})-\mathbf{1}_m\right]\right\}^\top\boldsymbol{\Delta}^\top\omega\boldsymbol{\Sigma}_{\boldsymbol{Y}}^{-1}\omega\boldsymbol{\Delta} \\
&\times\left\{\boldsymbol{u}-\left(\boldsymbol{\Delta}^\top\omega\boldsymbol{\Sigma}_{\boldsymbol{Y}}^{-1}\omega\boldsymbol{\Delta}\right)^{-1}\left[\boldsymbol{\Delta}^\top\omega\boldsymbol{\Sigma}_{\boldsymbol{Y}}^{-1}(\boldsymbol{y}-\boldsymbol{\xi})-\mathbf{1}_m\right]\right\} \\
&-\left[\boldsymbol{\Delta}^\top\omega\boldsymbol{\Sigma}_{\boldsymbol{Y}}^{-1}(\boldsymbol{y}-\boldsymbol{\xi})-\mathbf{1}_m\right]^\top\left(\boldsymbol{\Delta}^\top\omega\boldsymbol{\Sigma}_{\boldsymbol{Y}}^{-1}\omega\boldsymbol{\Delta}\right)^{-1}\left[\boldsymbol{\Delta}^\top\omega\boldsymbol{\Sigma}_{\boldsymbol{Y}}^{-1}(\boldsymbol{y}-\boldsymbol{\xi})-\mathbf{1}_m\right],
\end{aligned}
$$

then, (a.2) turns to

$$
\begin{aligned}
& (2\pi)^{\frac{m}{2}}|\boldsymbol{\Delta}^\top\omega\boldsymbol{\Sigma}_{\boldsymbol{Y}}^{-1}\omega\boldsymbol{\Delta}|^{-\frac{1}{2}}\phi_d\left(\boldsymbol{y};\boldsymbol{\xi},\boldsymbol{\Sigma}_{\boldsymbol{Y}}\right)\phi_m\left(\boldsymbol{u};\left(\boldsymbol{\Delta}^\top\omega\boldsymbol{\Sigma}_{\boldsymbol{Y}}^{-1}\omega\boldsymbol{\Delta}\right)^{-1}\left[\boldsymbol{\Delta}^\top\omega\boldsymbol{\Sigma}_{\boldsymbol{Y}}^{-1}(\boldsymbol{y}-\boldsymbol{\xi})-\mathbf{1}_m\right],\left(\boldsymbol{\Delta}^\top\omega\boldsymbol{\Sigma}_{\boldsymbol{Y}}^{-1}\omega\boldsymbol{\Delta}\right)^{-1}\right) \\
&\times\exp\left\{\frac{1}{2}\left[\boldsymbol{\Delta}^\top\omega\boldsymbol{\Sigma}_{\boldsymbol{Y}}^{-1}(\boldsymbol{y}-\boldsymbol{\xi})-\mathbf{1}_m\right]^\top\left(\boldsymbol{\Delta}^\top\omega\boldsymbol{\Sigma}_{\boldsymbol{Y}}^{-1}\omega\boldsymbol{\Delta}\right)^{-1}\left[\boldsymbol{\Delta}^\top\omega\boldsymbol{\Sigma}_{\boldsymbol{Y}}^{-1}(\boldsymbol{y}-\boldsymbol{\xi})-\mathbf{1}_m\right]\right\} \\
&= (2\pi)^{\frac{m}{2}}|\boldsymbol{\Sigma}_{\mathcal{E}}|^{-\frac{1}{2}}\phi_d(\boldsymbol{y};\boldsymbol{\xi},\boldsymbol{\Sigma}_{\boldsymbol{Y}})\phi_m(\boldsymbol{u};\boldsymbol{\mu}_{\mathcal{E}},\boldsymbol{\Sigma}_{\mathcal{E}})\exp\left\{\frac{1}{2}\boldsymbol{\eta}^\top\boldsymbol{\mu}_{\mathcal{E}}\right\} \\
&= (2\pi)^{\frac{m}{2}}|\boldsymbol{\Sigma}_{\mathcal{E}}|^{-\frac{1}{2}}\phi_d(\boldsymbol{y};\boldsymbol{\xi},\boldsymbol{\Sigma}_{\boldsymbol{Y}})\phi_m(\boldsymbol{u};\boldsymbol{\mu}_{\mathcal{E}},\boldsymbol{\Sigma}_{\mathcal{E}})\exp\left\{\frac{1}{2}\boldsymbol{\eta}^\top\boldsymbol{\Sigma}_{\mathcal{E}}\boldsymbol{\eta}\right\},
\end{aligned}
$$

where $\boldsymbol{\mu}_{\mathcal{E}}=\boldsymbol{\Sigma}_{\mathcal{E}}\boldsymbol{\eta}$, $\boldsymbol{\eta}=\left[\boldsymbol{\Delta}^\top\omega\boldsymbol{\Sigma}_{\boldsymbol{Y}}^{-1}(\boldsymbol{y}-\boldsymbol{\xi})-\mathbf{1}_m\right]$, $\boldsymbol{\Sigma}_{\mathcal{E}}=\left(\boldsymbol{\Delta}^\top\omega\boldsymbol{\Sigma}_{\boldsymbol{Y}}^{-1}\omega\boldsymbol{\Delta}\right)^{-1}$. Hence,

$$
f_{U|\boldsymbol{Y}}(\boldsymbol{u})=\frac{(2\pi)^{\frac{m}{2}}|\boldsymbol{\Sigma}_{\mathcal{E}}|^{-\frac{1}{2}}\phi_d(\boldsymbol{y};\boldsymbol{\xi},\boldsymbol{\Sigma}_{\boldsymbol{Y}})\phi_m(\boldsymbol{u};\boldsymbol{\mu}_{\mathcal{E}},\boldsymbol{\Sigma}_{\mathcal{E}})\exp\left\{\frac{1}{2}\boldsymbol{\eta}^\top\boldsymbol{\mu}_{\mathcal{E}}\right\}}{\int_{\mathbb{R}_+^m}(2\pi)^{\frac{m}{2}}|\boldsymbol{\Sigma}_{\mathcal{E}}|^{-\frac{1}{2}}\phi_d(\boldsymbol{y};\boldsymbol{\xi},\boldsymbol{\Sigma}_{\boldsymbol{Y}})\phi_m(\boldsymbol{u};\boldsymbol{\mu}_{\mathcal{E}},\boldsymbol{\Sigma}_{\mathcal{E}})\exp\left\{\frac{1}{2}\boldsymbol{\eta}^\top\boldsymbol{\mu}_{\mathcal{E}}\right\}\mathrm{d}\boldsymbol{u}}=\frac{\phi_m(\boldsymbol{u};\boldsymbol{\mu}_{\mathcal{E}},\boldsymbol{\Sigma}_{\mathcal{E}})}{\int_{\mathbb{R}_+^m}\phi_m(\boldsymbol{u};\boldsymbol{\mu}_{\mathcal{E}},\boldsymbol{\Sigma}_{\mathcal{E}})\,\mathrm{d}\boldsymbol{u}},
$$

which completes the proof of the result. □

References

1. M. Abdi, M. Madadi, N. Balakrishnan, A. Jamalizadeh, Family of mean-mixtures of multivariate normal distributions: Properties, inference and assessment of multivariate skewness, *J. Multivar. Anal.*, **181** (2021), 104679. https://doi.org/10.1016/j.jmva.2020.104679
2. R. B. Arellano-Valle, A. Azzalini, On the unification of families of skew-normal distributions, *Scand. J. Stat.*, **33** (2006), 561–574. https://doi.org/10.1111/j.1467-9469.2006.00503.x
3. R. B. Arellano-Valle, M. G. Genton, On fundamental skew distributions, *J. Multivar. Anal.*, **96** (2005), 93–116. https://doi.org/10.1016/j.jmva.2004.10.002
4. O. Arslan, Variance-mean mixture of Kotz-type distributions, *Commun. Stat.-Theory Methods*, **38** (2009), 272–284. https://doi.org/10.1080/03610920802192497
5. O. Arslan, Variance-mean mixture of the multivariate skew normal distribution, *Stat. Papers*, **56** (2015), 353–378. https://doi.org/10.1007/s00362-014-0585-7

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

7. A. Azzalini, An overview on the progeny of the skew-normal family–A personal perspective, *J. Multivar. Anal.*, **188** (2022), 104851. https://doi.org/10.1016/j.jmva.2021.104851

8. A. Azzalini, A. Capitanio, Statistical applications of the multivariate skew-normal distribution, *J. R. Stat. Soc. Ser. B-Stat. Methodol.*, **61** (1999), 579–602. https://doi.org/10.1111/1467-9868.00194

9. A. Azzalini, A. Capitanio, *The skew-normal and related families*, New York: Cambridge University Press, 2014. https://doi.org/10.1017/CBO9781139248891

10. A. Azzalini, A. D. Valle, The multivariate skew-normal distribution, *Biometrika*, **83** (1996), 715–726. https://doi.org/10.1093/biomet/83.4.715

11. O. Barndorff-Nielsen, J. Kent, M. Sørensen, Normal variance-mean mixtures and z distributions, *Int. Stat. Rev.*, **50** (1982), 145–159. https://doi.org/10.2307/1402598

12. M. Bernardi, R. Cerqueti, A. Palestini, The skew normal multivariate risk measurement framework, *Comput. Manag. Sci.*, **17** (2020), 105–119. https://doi.org/10.1007/s10287-019-00350-8

13. M. Denuit, J. Dhaene, M. Goovaerts, R. Kaas, *Actuarial theory for dependent risks: Measures, orders and models*, England: John Wiley & Sons, Ltd, 2005. https://doi.org/10.1002/0470016450

14. G. González-Farías, A. Domínguez-Molina, A. K. Gupta, Additive properties of skew normal random vectors, *J. Stat. Plan. Infer.*, **126** (2004), 521–534. https://doi.org/10.1016/j.jspi.2003.09.008

15. A. Jamalizadeh, N. Balakrishnan, Conditional distributions of multivariate normal mean–variance mixtures, *Stat. Probab. Lett.*, **145** (2019), 312–316. https://doi.org/10.1016/j.spl.2018.10.005

16. Z. Landsman, U. Makov, T. Shushi, A multivariate tail covariance measure for elliptical distributions, *Insur. Math. Econ.*, **81** (2018), 27–35. https://doi.org/10.1016/j.insmatheco.2018.04.002

17. S. X. Lee, G. J. McLachlan, EMMIXcskew: An R package for the fitting of a mixture of canonical fundamental skew t-distributions, *J. Stat. Softw.*, **83** (2018), 1–32. https://doi.org/10.18637/jss.v083.i03

18. S. X. Lee, T. Lin, G. J. McLachlan, Mixtures of factor analyzers with scale mixtures of fundamental skew normal distributions, *Adv. Data Anal. Classif.*, **15** (2021), 481–512. https://doi.org/10.1007/s11634-020-00420-9

19. B. Liseo, N. Loperfido, A Bayesian interpretation of the multivariate skew-normal distribution, *Stat. Probab. Lett.*, **61** (2003), 395–401. https://doi.org/10.1016/S0167-7152(02)00398-X

20. S. A. Mousavi, V. Amirzadeh, M. Rezapour, A. Sheikhy, Multivariate tail conditional expectation for scale mixtures of skew-normal distribution, *J. Stat. Comput. Simul.*, **89** (2019), 3167–3181. https://doi.org/10.1080/00949655.2019.1657864

21. M. Naderi, A. Arabpour, A. Jamalizadeh, Multivariate normal mean-variance mixture distribution based on Lindley distribution, *Commun. Stat.-Simul. Comput.*, **47** (2018), 1179–1192. https://doi.org/10.1080/03610918.2017.1307400

22. H. Negarestani, A. Jamalizadeh, S. Shafiei, N. Balakrishnan, Mean mixtures of normal distributions: Properties, inference and application, *Metrika*, **82** (2019), 501–528. https://doi.org/10.1007/s00184-018-0692-x

23. H. Ogasawara, A non-recursive formula for various moments of the multivariate normal distribution with sectional truncation, *J. Multivar. Anal.*, **183** (2021), 104729. https://doi.org/10.1016/j.jmva.2021.104729

24. R. Pourmousa, A. Jamalizadeh, M. Rezapour, Multivariate normal mean–variance mixture distribution based on Birnbaum–Saunders distribution, *J. Stat. Comput. Simul.*, **85** (2015), 2736–2749. https://doi.org/10.1080/00949655.2014.937435

25. R. Roozegar, N. Balakrishnan, A. Jamalizadeh, On moments of doubly truncated multivariate normal mean–variance mixture distributions with application to multivariate tail conditional expectation, *J. Multivar. Anal.*, **177** (2020), 104586. https://doi.org/10.1016/j.jmva.2019.104586