Shahid Bahonar University of Kerman
Journal of Mahani Mathematical Research
Print ISSN: 2251-7952 Online ISSN: 2645-4505

SOME CHARACTERISTICS OF TENSOR-VARIATE SKEW-NORMAL DISTRIBUTION AND ITS APPLICATION IN IMAGE ANALYSIS

S.A.A. Tajadod, F. Yousefzadeh ✉, R.B. Arellano-Valle, and S. Jomhoori

Article type: Research Article
(Received: 11 February 2025, Received in revised form 14 May 2025)
(Accepted: 08 July 2025, Published Online: 09 July 2025)

Abstract. In recent years, due to the increasing growth of technology and new technologies, data is obtained in more complex structures as the main component in analysis. One of these complex structures is tensors. Therefore, in order to answer this need (analysis of data with tensor structure), it is necessary to expand statistical concepts and methods in the field of data with tensor structure. On the other hand, in reality, we may also encounter skew data. Therefore, in this article, we have introduced the skew normal tensor distribution and obtained some of its important statistical properties. Subsequently, we employed the EM algorithm to obtain maximum likelihood estimates of the parameters and assessed their accuracy through simulation studies. Finally, we have shown the effectiveness of the obtained estimators with real data.

Keywords: Kronecker-separable covariance, Multidimensional array, Tensor, Skew-distributions, EM algorithm, Image learning.
2020 MSC: 62H10, 70G45, 62H12.

1. Introduction

In the last decade, data has been obtained in increasingly complex structures. One of these complex structures is multidimensional or tensor data. In other words, a tensor is a multi-dimensional (multi-dimensional or multi-component) array, where each component represents a vector of observations. More formally, the Nth order tensor is an element of the multiplication of the N vector space of tensors, each of which has its own coordinates. Of course, this concept of tensors should not be confused with tensors in physics and engineering such as stress tensors, [13], which are generally called tensor fields in mathematics [6].

Given that complex data structures have become commonplace across various applied research fields, there is a growing need to develop and expand theoretical concepts for analyzing high-dimensional data. In order to answer

✉ fyousefzadeh@birjand.ac.ir, ORCID: 0000-0003-2648-5897
https://doi.org/10.22103/jmmr.2025.24813.1763
Publisher: Shahid Bahonar University of Kerman
How to cite: S.A.A. Tajadod, F. Yousefzadeh R.B. Arellano-Valle, S. Jomhoori, *Some characteristics of tensor-variate skew-normal distribution and its application in image analysis*, J. Mahani Math. Res. 2026; 15(2): 45-64.

this need, we extended statistical concepts from the field of random vectors (multivariate) to random matrices, now they should be extended to random tensors, which are actually an extension of matrix data.

The expansion of this concept has been used and investigated by many researchers in the fields of chemometrics, image measurement, face recognition and psychometrics. In this context, we refer the reader to [9] and [14] and the references mentioned in them.

What can be inferred from these articles is that statistical methods that preserve the structure of tensors are very important because converting tensor data into a vector or matrix leads to the loss of some information. A good work, especially focusing on the need to go from a matrix distribution to a tensor distribution, is presented in [4]. They argue why a vector analysis of a complex dataset that actually requires a tensor analysis and the use of a tensor distribution can lead to wrong or inefficient conclusions. In proportion to the widespread use of the concept of tensors in various fields, but limited statistical research related to tensor distributions, probabilistic properties and the problem of estimation have not grown in the same proportion. Of course, in recent years, researchers have conducted research in this regard, including [14] which is related to the analysis of tensor type data to multilinear/tensor normal distribution and also [1], using [14] concepts, was able to generalize it to tensor elliptical distributions, or [8] calculated the Stein-type risk estimator when the errors follow a tensor elliptic distribution. Of course, these researches are also in the field of symmetric distributions, although it is relatively good mathematically, but the assumption of symmetry is often violated. In addition, there may be outliers in the data, which can be problematic, and more importantly, the data structure in the real world can be skewed. Therefore, the study of the tensor distributions in this field can be useful, but currently, according to our knowledge, there has been limited research related to these distributions, we can refer to [7] in which, they presented four tensor distributions that can be considered as generalizations of their variable matrix counterparts and are able to model skewness and kurtosis, as well as [11] who studied a regression model in which the response are of the type of skew tensor data. In recent years, the development of tensor-based statistical models has attracted increasing attention. However, the extension of well-known distributions to the tensor domain—particularly the skew normal distribution—has not been thoroughly investigated. To address this gap, in this paper, we focus on the tensor variate skew normal distribution (TVSN) in order to expand the concept of tensor distributions, given the central role of the normal distribution in statistical modeling and inference. The rest of the article is organized as follows. Section 2 introduces the TVSN distribution along with some symbols that help in simplifying the calculations, also some features of the TVSN distribution, such as the density function, expectation and variance and the characteristic function are given. In section 3, parameter estimation using EM algorithm is discussed. Simulation and analysis of observed data is presented in section 4. In section

5, a real example is presented and using the results obtained in section 3, we discuss the efficiency of the model presented in the paper.

2. Model

In this section, while recalling the basic concepts of tensors and the tensor variate normal (TVN) distribution, we introduce the TVSN distribution and establish some of its properties.

2.1. **Background and preliminaries.** A Tensor is introduced as a multidimensional array; a kth order tensor is an array with k dimensions, denoted by $\mathcal{X} \in \mathbb{R}^{\times_{j=1}^{k} p_j}$. The order of a tensor is the number of dimensions, also known as ways or modes; which we have shown here with k. The p_j is the marginal dimension of the jth mode ($j = 1, 2, \ldots, k$). The $(i_1, i_2, \ldots, i_k)$th element of the tensor $\mathcal{X}$ is denoted by $x_{i_1,i_2,\ldots,i_k}$ where $i_j = 1, 2, \ldots, p_j$ and $j = 1, 2, \ldots, k$. In addition, by fixing some indices of the tensor, it is possible to reach the tensor of lower order. In particular, with this work, a tensor can be converted into a matrix or vector, making explicit calculations more facile, denoted by $\boldsymbol{X}$, $\boldsymbol{x}$ respectively. See [9] and [5] for further discussion of these concepts.

The vectorial representation of a tensor, makes the related inference much simpler. Let $vec(\mathcal{X})$ denote the vectorization of tensor $\mathcal{X} = (x_{i_1 i_2 \ldots i_k})$, according to the definition of [9] given by

$$\begin{aligned} \boldsymbol{x} = vec(\mathcal{X}) &= \sum_{i_1=1}^{p_1} \cdots \sum_{i_k=1}^{p_k} x_{i_1 i_2 \ldots i_k} \boldsymbol{e}_{i_1}^{1} \otimes \cdots \otimes \boldsymbol{e}_{i_k}^{k}, \\ &= \sum_{I_p} x_{i_1 i_2 \ldots i_k} \boldsymbol{e}_{1:k}^{\boldsymbol{p}}, \end{aligned} \tag{1}$$

where $\boldsymbol{e}_{i_k}^{k}, \boldsymbol{e}_{i_k-1}^{k-1}, \ldots, \boldsymbol{e}_{i_1}^{1}$ are the unit basis vectors (a p-vector with 1 in the jth position, and 0 elsewhere) of size p_k, p_{k-1}, ..., p_1, respectively, $\boldsymbol{e}_{1:k}^{\boldsymbol{p}} = \boldsymbol{e}_{i_1}^{1} \otimes \ldots \otimes \boldsymbol{e}_{i_k}^{k}$, where $\otimes$ denotes the Kronecker product, $\boldsymbol{I_p}$ is the index set defined as $\boldsymbol{I_p} = \{i_1, \ldots, i_k : 1 \leq i_j \leq p_j, 1 \leq \mathrm{j} \leq k\}$.

In [14], the authors concentrated on the estimation of a Kronecker structured covariance matrix of order three ($k = 3$), the so called double separable covariance matrix, generalizing the work of [15], for multilinear normal (MLN) distribution. Further, let

$$p_{j:l}^{*} = \prod_{i=j}^{l} p_i \qquad and \qquad p_{j:l}^{+} = \sum_{i=j}^{l} p_i, \tag{2}$$

with the special cases

$$p^{*} = p_{1:k}^{*} \qquad and \qquad p^{+} = p_{1:k}^{+} \tag{3}$$

respectively. When there is no ambiguity, we shall drop the dimension from the basis vectors and write $\boldsymbol{e}_{i_1}^{p_1}$ as $\boldsymbol{e}_{i_1}$, and $\boldsymbol{e}_{i_1:i_k}^{p}$ as $\boldsymbol{e}_{i_1:i_k}$, etc.

Let $\mathcal{T}^p$ denote the space of all vectors $\boldsymbol{x} = vec(\mathcal{X})$, where $\mathcal{X}$ is a tensor of order k, i.e., $\mathcal{T}^p = \{\boldsymbol{x} : \boldsymbol{x} = \sum_{\boldsymbol{I_p}} x_{i_1 i_2 \ldots i_k} \boldsymbol{e}^{\boldsymbol{p}}_{1:k}\}$. Note that this tensor space is described using vectors. However, we can define tensor spaces using matrices. This is given in the following definition.

Definition 2.1. Let

$$
\begin{aligned}
(i) \qquad \mathcal{T}^{pq} &= \{\boldsymbol{X} : \boldsymbol{X} = \sum_{I_p \cup \mathrm{I}_q} x_{1,\ldots,i_k,j_1,\ldots,j_l} \boldsymbol{e}^{\boldsymbol{p}}_{1:k} (\boldsymbol{e}^{\boldsymbol{q}}_{1:l})'\}, \text{ where} \\
&\quad \boldsymbol{I}_q = \{j_1, \ldots, j_l : 1 \le j_i \le p_i, 1 \le \mathrm{i} \le l\} \\
(ii) \qquad \mathcal{T}^{pq}_{\otimes} &= \{\boldsymbol{X} \in \mathcal{T}^{pq} : \boldsymbol{X} = \boldsymbol{X}_1 \otimes \ldots \otimes \boldsymbol{X}_k,\ \boldsymbol{X}_i : p_i \times q_i\} \\
(iii) \qquad \mathcal{T}^{p}_{\otimes} &= \{\boldsymbol{X} \in \mathcal{T}^{pp}_{\otimes} : \boldsymbol{X} = \boldsymbol{X}_1 \otimes \ldots \otimes \boldsymbol{X}_k,\ \boldsymbol{X}_i : p_i \times p_i\}
\end{aligned}
$$

In the following, we will introduce the normal tensor (TVN) distribution and present the probability density function (pdf) and its characteristic function (cf), which we will use in the next topic. For more information about this distribution, we refer the reader to [14] and [12].

Definition 2.2. A tensor random variable $\mathcal{U} \in \mathbb{R}^{p_1 \times \ldots \times p_k}$ has standard tensor normal distribution when all the elements of $\mathcal{U}$ are independent standard normal distribution. Thus, $\mathcal{X} = \mathcal{M} + [\![\mathcal{U}; \boldsymbol{\Sigma}_1^{\frac{1}{2}}, \ldots, \boldsymbol{\Sigma}_k^{\frac{1}{2}}]\!]$ has a TVN distribution denoted by $\mathcal{X} \sim TVN(\mathcal{M}, \boldsymbol{\Sigma}_1, \ldots, \boldsymbol{\Sigma}_k)$, where the positive definite matrix $\boldsymbol{\Sigma}_k = \boldsymbol{\Sigma}_k^{\frac{1}{2}} \boldsymbol{\Sigma}_k^{\frac{1}{2}}$ models the dependence structure on the kth-mode and $[\![\mathcal{U}; \boldsymbol{\Sigma}_1^{\frac{1}{2}}, \ldots, \boldsymbol{\Sigma}_k^{\frac{1}{2}}]\!]$ is the Tucker product between $\mathcal{U}$ and $\boldsymbol{\Sigma}_1^{\frac{1}{2}}, \ldots, \boldsymbol{\Sigma}_k^{\frac{1}{2}}$ (see [5]). In this case, $vec(\mathcal{X}) = vec(\mathcal{M}) + \boldsymbol{\Sigma}_{1:k}^{\frac{1}{2}} vec(\mathcal{U})$ has the multivariate normal distribution, with location vector $\boldsymbol{\mu} = vec(\mathcal{M})$ and positive definite dispersion matrix $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_{1:k} = \boldsymbol{\Sigma}_1 \otimes \ldots \otimes \boldsymbol{\Sigma}_k$.

Under the assumptions of definition 2.2 the pdf and the cf of the TVN distribution are given as

$$
(4) \qquad f_{\mathcal{X}}(\boldsymbol{x}) = (2\pi)^{\frac{-p^*}{2}} \left(\prod_{i=1}^{k} \left| \boldsymbol{\Sigma}_i \right|^{\frac{-p^*}{(2p_i)}} \right) exp\Big\{ -\frac{1}{2} (\boldsymbol{x} - \boldsymbol{\mu})' \boldsymbol{\Sigma}_{1:k}^{-1} (\boldsymbol{x} - \boldsymbol{\mu}) \Big\},
$$

where $\boldsymbol{x}, \boldsymbol{\mu} \in \mathcal{T}^p$, $\boldsymbol{\Sigma}_{1:k}^{-1} = (\boldsymbol{\Sigma}_1 \otimes \ldots \otimes \boldsymbol{\Sigma}_k)^{-1} = \boldsymbol{\Sigma}_1^{-1} \otimes \ldots \otimes \boldsymbol{\Sigma}_k^{-1}$, $\prod_{i=1}^{k} \left| \boldsymbol{\Sigma}_i \right|^{\frac{-p^*}{(2p_i)}} = \left(\left| \boldsymbol{\Sigma}_1 \right|^{\frac{-p^*}{p_1}} \right)^{-\frac{1}{2}} \times \left(\left| \boldsymbol{\Sigma}_2 \right|^{\frac{-p^*}{p_2}} \right)^{-\frac{1}{2}} \times \ldots \times \left(\left| \boldsymbol{\Sigma}_k \right|^{\frac{-p^*}{p_k}} \right)^{-\frac{1}{2}} = \left| \boldsymbol{\Sigma}_1 \otimes \boldsymbol{\Sigma}_2 \otimes \ldots \otimes \boldsymbol{\Sigma}_k \right|^{-\frac{1}{2}} = \left| \boldsymbol{\Sigma} \right|^{-\frac{1}{2}}$ and p^* is defined in (3).

$$
(5) \qquad \varphi_{\mathcal{X}}(\boldsymbol{t}) = E[exp\{i \boldsymbol{t}' \boldsymbol{x}\}] = exp[i \boldsymbol{t}' \boldsymbol{\mu} - \frac{1}{2} \boldsymbol{t}' \boldsymbol{\Sigma}_{1:k} \boldsymbol{t}], \quad \boldsymbol{t} \in \mathcal{T}^p.
$$

2.2. **Characterizing the TVSN distribution.** In this subsection, we introduce the tensor-variate skew-normal distribution (TVSN) and establish some of its properties.

The methodology behind our definition of TVSN distribution stems from the fact that the difference between vector-variate skew-normal and TVSN lies in the structure of the parameter space generated by the location vector $\boldsymbol{\mu}$, dispersion (scatter) matrix $\boldsymbol{\Sigma}_{1:k}$ and skewness vector $\boldsymbol{\delta}$.

Definition 2.3. A random variable $\mathcal{X}$ is said to have a TVSN distribution of order k, with location vector $\boldsymbol{\mu} \in \mathcal{T}^p$, positive definite dispersion matrix $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_{1:k} \in \mathcal{T}^p_{\otimes}$ and skewness vector $\boldsymbol{\delta} \in \mathcal{T}^p$, denoted by $\mathcal{X} \sim SN_{\boldsymbol{p}}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\delta})$, if

$$\boldsymbol{x} = \boldsymbol{\mu} + \boldsymbol{\Sigma}_{1:k}^{\frac{1}{2}} \boldsymbol{u}, \tag{6}$$

Where $\boldsymbol{p} = (p_1, \ldots, p_k)$ and the elements of $\boldsymbol{u} \in \mathcal{T}^p$ are independent skew-normal distributed.

Proposition 2.4. *Suppose that $\mathcal{X}$ has the tensor variate normal standard distribution and w has the univariate half normal distribution, independently of $\mathcal{X}$, then the density function of the TVSN distribution of order k, denoted by $\mathcal{U} \sim SN_{\boldsymbol{p}}(\boldsymbol{0}, \boldsymbol{I}_{1:k}, \boldsymbol{\delta})$ with the following characterization*

$$\boldsymbol{u} = \frac{\boldsymbol{x} + \boldsymbol{\delta} w}{\sqrt{1 + \boldsymbol{\delta}'\boldsymbol{\delta}}},$$

is given as

$$f_{\mathcal{U}}(\boldsymbol{u}) = 2(2\pi)^{-\frac{p^*}{2}} exp\{-\frac{1}{2}\boldsymbol{u}'\boldsymbol{u}\}\Phi(\boldsymbol{\delta}'\boldsymbol{u}).$$

or in short

$$f_{\mathcal{U}}(\boldsymbol{u}) = 2\phi_{\boldsymbol{p}}(\boldsymbol{u}; \boldsymbol{0}, \boldsymbol{I_{1:k}})\Phi(\boldsymbol{\delta}'\boldsymbol{u}). \tag{7}$$

Where $p^* = p^*_{1:k} = \prod_{i=1}^k p_i$, $\boldsymbol{u}, \boldsymbol{x}, \boldsymbol{\delta} \in \mathcal{T}^p$, $\boldsymbol{I}_{1:k} \in \mathcal{T}^p_{\otimes}$, $\phi(\cdots)$ represents the pdf of the TVSN distribution and $\Phi(\cdots)$ represents the distribution function of a univariate standard normal distribution.

Proof. Since $\mathcal{X}$ has the tensor variate normal standard distribution and w has the univariate half normal distribution we have

$$f_{\mathcal{X},W}(\boldsymbol{x}, w) = \frac{2}{2\pi}(2\pi)^{(-p^*/2)} exp\left\{ . - \frac{1}{2}\boldsymbol{x}'\boldsymbol{x} - \frac{w}{2} . \right\}.$$

So, we can write: $f_{\mathcal{U},W}(\boldsymbol{u}, w) =$

$$\begin{aligned} &= \tfrac{2}{2\pi}(2\pi)^{(-p^*/2)}\sqrt{1+\boldsymbol{\delta}'\boldsymbol{\delta}}exp\{-\tfrac{1}{2}(\sqrt{1+\boldsymbol{\delta}'\boldsymbol{\delta}}u - \boldsymbol{\delta}w)'(\sqrt{1+\boldsymbol{\delta}'\boldsymbol{\delta}}u - \boldsymbol{\delta}w) - \tfrac{w}{2}\}, \\ &= \tfrac{2}{2\pi}(2\pi)^{(-p^*/2)}\sqrt{1+\boldsymbol{\delta}'\boldsymbol{\delta}}exp\{-\tfrac{1}{2}(1+\boldsymbol{\delta}'\boldsymbol{\delta})(w - \tfrac{\boldsymbol{\delta}'\boldsymbol{u}}{\sqrt{1+\boldsymbol{\delta}'\boldsymbol{\delta}}})^2\}exp\{-\tfrac{1}{2}\boldsymbol{u}'\boldsymbol{u}\}, \end{aligned}$$

Therefore, the marginal distribution of $\boldsymbol{u}$ is

$$\begin{aligned} f_{\mathcal{U}}(\boldsymbol{u}) &= \int_0^\infty f(\boldsymbol{u}, w)dw, \\ &= 2\phi_{\boldsymbol{p}}(\boldsymbol{u}; \boldsymbol{0}, \boldsymbol{I_{1:k}})\Phi(\boldsymbol{\delta}'\boldsymbol{u}), \end{aligned}$$

where $p^* = p^*_{1:k} = \prod_{i=1}^{k} p_i$, $\boldsymbol{u}, \boldsymbol{x}, \boldsymbol{\delta} \in \mathcal{T}^p$, $\boldsymbol{I}_{1:k} \in \mathcal{T}^p_{\otimes}$, $\phi(\cdots)$ represents the pdf of the TVN distribution and $\Phi(\cdots)$ represents the distribution function of a univariate normal distribution. □

There are many other types of TVSN distribution. (7) represents what arguably is the simplest option involving a modulation factor of Gaussian type operating on a tensor variate normal base density.

Theorem 2.5. *The density function of the TVSN distribution (Definition 2.3) is given as*

$$f_{\mathcal{X}}(\boldsymbol{x}) = 2\phi_{\boldsymbol{p}}(\boldsymbol{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma_{1:k}})\Phi\left(\boldsymbol{\delta}'\left(\boldsymbol{\Sigma}^{-\frac{1}{2}}_{1:k}(\boldsymbol{x}-\boldsymbol{\mu})\right)\right). \tag{8}$$

where $\phi_{\boldsymbol{p}}(\boldsymbol{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}_{1:k})$ represents the pdf of the tensor variate normal distribution and $\Phi(\cdots)$ represents the distribution function of a univariate normal distribution.

Proof. Let $\mathcal{U} \sim \text{TVSN}_{\boldsymbol{p}}(\boldsymbol{0}, \boldsymbol{I}_{1:k}, \boldsymbol{\delta})$ be a tensor variate standard skew normal distribution. According to Proposition 2.4, the probability density function (pdf) of $\mathcal{U}$ is given by

$$f_{\mathcal{U}}(\boldsymbol{u}) = 2\,\phi_{\boldsymbol{p}}(\boldsymbol{u}; \boldsymbol{0}, \boldsymbol{I}_{1:k})\,\Phi(\boldsymbol{\delta}^\top \boldsymbol{u}),$$

where $\phi_{\boldsymbol{p}}(\cdot)$ denotes the pdf of the tensor variate normal distribution with identity covariance structure, and $\Phi(\cdot)$ is the cumulative distribution function of the univariate standard normal distribution.

Now consider a general tensor variate skew normal random variable $\mathcal{X} \sim \text{TVSN}_{\boldsymbol{p}}(\boldsymbol{\mu}, \boldsymbol{\Sigma}_{1:k}, \boldsymbol{\delta})$. Define the linear transformation

$$\boldsymbol{u} = \boldsymbol{\Sigma}^{-1/2}_{1:k}(\boldsymbol{x}-\boldsymbol{\mu}), \quad \text{equivalently,} \quad \boldsymbol{x} = \boldsymbol{\Sigma}^{1/2}_{1:k}\boldsymbol{u} + \boldsymbol{\mu}.$$

Since this is a linear transformation and $\boldsymbol{x}, \boldsymbol{\mu} \in \mathcal{T}^p$, the Jacobian determinant of the transformation is

$$\left|\frac{\partial \boldsymbol{x}}{\partial \boldsymbol{u}}\right| = \left|\boldsymbol{\Sigma}^{1/2}_{1:k}\right| = |\boldsymbol{\Sigma}_{1:k}|^{1/2}.$$

Therefore, using the change of variables formula, the pdf of $\boldsymbol{x}$ becomes

$$f_{\mathcal{X}}(\boldsymbol{x}) = f_{\mathcal{U}}(\boldsymbol{\Sigma}^{-1/2}_{1:k}(\boldsymbol{x}-\boldsymbol{\mu})) \cdot |\boldsymbol{\Sigma}^{-1/2}_{1:k}|.$$

Substituting the known expression for $f_{\mathcal{U}}$ gives

$$f_{\mathcal{X}}(\boldsymbol{x}) = 2\,\phi_{\boldsymbol{p}}(\boldsymbol{\Sigma}^{-1/2}_{1:k}(\boldsymbol{x}-\boldsymbol{\mu}); \boldsymbol{0}, \boldsymbol{I}_{1:k})\,\Phi(\boldsymbol{\delta}^\top \boldsymbol{\Sigma}^{-1/2}_{1:k}(\boldsymbol{x}-\boldsymbol{\mu})) \cdot |\boldsymbol{\Sigma}^{-1/2}_{1:k}|.$$

Using the identity

$$\phi_{\boldsymbol{p}}(\boldsymbol{\Sigma}^{-1/2}_{1:k}(\boldsymbol{x}-\boldsymbol{\mu}); \boldsymbol{0}, \boldsymbol{I}_{1:k}) \cdot |\boldsymbol{\Sigma}^{-1/2}_{1:k}| = \phi_{\boldsymbol{p}}(\boldsymbol{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}_{1:k}),$$

we obtain

$$f_{\mathcal{X}}(\boldsymbol{x}) = 2\,\phi_{\boldsymbol{p}}(\boldsymbol{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}_{1:k})\,\Phi(\boldsymbol{\delta}^\top \boldsymbol{\Sigma}^{-1/2}_{1:k}(\boldsymbol{x}-\boldsymbol{\mu})).$$

This completes the proof. □

2.2.1. *Moments, characteristic function and cumulants.*

Theorem 2.6. *Let* $\boldsymbol{x} \sim SN_{\boldsymbol{p}}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\delta})$, *where* $\boldsymbol{x} \in \mathcal{T}^{p}$. *The characteristic function of* $\boldsymbol{x}$ *is*

$$\varphi_{\mathcal{X}}(\boldsymbol{t}) = 2[exp(i\boldsymbol{t}'\boldsymbol{\mu} + \frac{1}{2}\boldsymbol{t}'\boldsymbol{\Sigma}_{1:k}\boldsymbol{t})]\Phi(\boldsymbol{\lambda}'\boldsymbol{\Sigma}_{1:k}^{\frac{1}{2}}i\boldsymbol{t}), \qquad \boldsymbol{t} \in \mathcal{T}^{\boldsymbol{p}} \tag{9}$$

and the cumulant generating function is

$$\begin{aligned} \mathsf{k}_{\mathcal{X}}(\boldsymbol{t}) &= Ln\left[\varphi_{\mathcal{X}}(\boldsymbol{t})\right] \\ &= Ln(2) + i\boldsymbol{t}'\boldsymbol{\mu} + \frac{1}{2}\boldsymbol{t}'\boldsymbol{\Sigma}_{1:k}\boldsymbol{t} + Ln[\Phi(\boldsymbol{\lambda}'\boldsymbol{\Sigma}_{1:k}^{\frac{1}{2}}i\boldsymbol{t})], \qquad \boldsymbol{t} \in \mathcal{T}^{\boldsymbol{p}}, \end{aligned} \tag{10}$$

where $\boldsymbol{\lambda} = (1 + \boldsymbol{\delta}'\boldsymbol{\delta})^{-\frac{1}{2}}\boldsymbol{\delta}, \quad \boldsymbol{\delta} \in \mathcal{T}^{p}$.

Proof. Let $\boldsymbol{u} = vec(\mathcal{U}) \sim SN_{\boldsymbol{p}}(\boldsymbol{0}, \boldsymbol{I}_{1:k}, \boldsymbol{\delta})$. So,

$$\begin{aligned} \varphi_{\mathcal{U}}(\boldsymbol{t}) &= E[exp(i\boldsymbol{t}'\boldsymbol{u})], \\ &= \int_{\mathbb{R}^{\boldsymbol{p}}} exp(i\boldsymbol{t}'\boldsymbol{u})f(\boldsymbol{u})d\boldsymbol{u}, \\ &= 2\int (2\pi)^{-\frac{p^*}{2}} exp\{-\frac{1}{2}(\boldsymbol{u} - i\boldsymbol{t})'(\boldsymbol{u} - i\boldsymbol{t})\}exp\{\frac{1}{2}\boldsymbol{t}'\boldsymbol{t}\}\Phi(\boldsymbol{\delta}'\boldsymbol{u})d\boldsymbol{u}. \end{aligned}$$

By making the transformation $\boldsymbol{z} = \boldsymbol{u} - i\boldsymbol{t}$, we get

$$\begin{aligned} \varphi_{\mathcal{U}}(\boldsymbol{t}) &= 2exp\Big\{\frac{\boldsymbol{t}'\boldsymbol{t}}{2}\Big\}\int (2\pi)^{\frac{-p^*}{2}} exp\{-\frac{1}{2}\boldsymbol{z}'\boldsymbol{z}\}\Phi\big(\boldsymbol{\delta}'(\boldsymbol{z} + i\boldsymbol{t})\big)d\boldsymbol{z}, \\ &= 2exp\Big\{\frac{\boldsymbol{t}'\boldsymbol{t}}{2}\Big\}E\Big\{\Phi\big(\boldsymbol{\delta}'(\boldsymbol{z} + i\boldsymbol{t})\big)\Big\}, \\ &= 2exp\Big\{\frac{\boldsymbol{t}'\boldsymbol{t}}{2}\Big\}\Phi\Big[\frac{\boldsymbol{\delta}'i\boldsymbol{t}}{(1 + \boldsymbol{\delta}'\boldsymbol{\delta})^{\frac{1}{2}}}\Big]. \end{aligned} \tag{11}$$

The last equality is obtained using Lemma 5.2 in [3].
Now, using equations (6) and (11), we get the desired result. □

To compute moments, we need a suitable differential operator (matrix derivative). Let $\boldsymbol{Y} \in \mathcal{T}^{\mathbf{pq}}$ be a function of $\boldsymbol{X} \in \mathcal{T}^{\mathbf{rs}}$, with their vectorized versions $\boldsymbol{y}$ and $\boldsymbol{x}$, defined as

$$\begin{aligned} \boldsymbol{y} &= \sum_{i_1:i_{k_1}} \sum_{j_1:j_{k_2}} \boldsymbol{y}_{i_1:i_{k_1}j_1:j_{k_2}} \mathbf{e}^{\mathbf{q}(1:k_2)}_{j_1:j_{k_2}} \otimes \mathbf{e}^{\mathbf{p}(1:k_1)}_{i_1:i_{k_1}}, \\ \boldsymbol{x} &= \sum_{m_1:m_{k_3}} \sum_{n_1:n_{k_4}} \boldsymbol{x}_{m_1:m_{k_3}n_1:n_{k_4}} \mathbf{e}^{\mathbf{s}(1:k_4)}_{n_1:n_{k_4}} \otimes \mathbf{e}^{\mathbf{r}(1:k_3)}_{m_1:m_{k_3}}, \end{aligned}$$

respectively. Then, $\frac{d\boldsymbol{Y}}{d\boldsymbol{X}} = \frac{d\boldsymbol{y}}{d\boldsymbol{x}} =$

$$= \sum_{i_1:i_{k_1}} \sum_{j_1:j_{k_2}} \sum_{m_1:m_{k_3}} \sum_{n_1:n_{k_4}} \frac{\partial y_{i_1:i_{k_1}j_1:j_{k_2}}}{\partial x_{m_1:m_{k_3}n_1:n_{k_4}}} \left(\mathbf{e}^{\mathbf{s}(1:k_4)}_{n_1:n_{k_4}} \otimes \mathbf{e}^{\mathbf{r}(1:k_3)}_{m_1:m_{k_3}}\right)\left(\mathbf{e}^{\mathbf{q}(1:k_2)}_{j_1:j_{k_2}} \otimes \mathbf{e}^{\mathbf{p}(1:k_1)}_{i_1:i_{k_1}}\right)' . \tag{12}$$

Higher order derivatives may be defined recursively, i.e.,

$$\frac{d^k \boldsymbol{Y}}{d\boldsymbol{X}^k} = \frac{d}{d\boldsymbol{X}} \frac{d^{k-1}\boldsymbol{Y}}{d\boldsymbol{X}^{k-1}}. \tag{13}$$

Theorem 2.7. *The mean tensor and covariance matrix of random tensor* $\boldsymbol{x} \sim SN_{\boldsymbol{p}}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\delta})$*, where* $\boldsymbol{x} \in \mathcal{T}^p$ *are given by*

$$\begin{aligned} E(\boldsymbol{x}) &= \boldsymbol{\mu} + \frac{2}{\sqrt{2\pi}}\boldsymbol{\gamma}. \\ var(\boldsymbol{x}) &= \boldsymbol{\Sigma}_{\mathbf{1:k}} - \frac{2}{\pi}\boldsymbol{\gamma}'\boldsymbol{\gamma}. \end{aligned}$$

where $\boldsymbol{\gamma} = \boldsymbol{\Sigma}_{1:k}^{\frac{1}{2}} \left(1 + \boldsymbol{\delta}'\boldsymbol{\delta}\right)^{-\frac{1}{2}} \boldsymbol{\delta}$.

Proof. Applying (12) to $\varphi(\mathbf{t})$, and evaluating the derivatives at $\mathbf{t} = \mathbf{0}$, the desired result is obtained. □

If we intend to present a multivariate density or distribution function through a multivariate skew normal distribution, there exist expansions where multivariate densities or distribution functions, multivariate cumulants and multivariate Hermite polynomials will appear in the formulas. Finding explicit expressions for the Hermite polynomials of low order will be the final topic of this subsection.

Definition 2.8. The matrix $\boldsymbol{H}(\boldsymbol{x}, \boldsymbol{\mu}, \boldsymbol{\Sigma})$ is called multivariate Hermite polynomial of order k for the vector $\boldsymbol{\mu}$ and the matrix $\boldsymbol{\Sigma} > 0$, if it satisfies the equality:

$$\frac{d^k f_{\mathcal{X}}(\boldsymbol{x})}{d\boldsymbol{x}^k} = (-1)^k \boldsymbol{H}(\boldsymbol{x}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) f_{\mathcal{X}}(\boldsymbol{x}), \quad k = 0, 1, \cdots,$$

where $\frac{d^k}{d\boldsymbol{x}^k}$ is given by (13), and $f_{\mathcal{X}}(\boldsymbol{x})$ is the pdf (8).

The explicit formulas for the first two Hermite polynomials will be given in the next theorem.

Theorem 2.9. *Let* $\boldsymbol{x} \sim SN_{\boldsymbol{p}}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\delta})$*, where* $\boldsymbol{x} \in \mathcal{T}^{\mathbf{p}}$*. The Hermite polynomials,* $\boldsymbol{H}(\boldsymbol{x}, \boldsymbol{\mu}, \boldsymbol{\Sigma}), k = 0, 1, 2$*, are of the form:*

$$\begin{aligned} (i) \quad & H_0(\boldsymbol{x}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = 1, \\ (ii) \quad & \boldsymbol{H}_1(\boldsymbol{x}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = (\boldsymbol{x} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} - \boldsymbol{\delta}' \boldsymbol{\Sigma}^{-\frac{1}{2}} \zeta\left(\boldsymbol{\delta}' \boldsymbol{\Sigma}^{-\frac{1}{2}} (\boldsymbol{x} - \boldsymbol{\mu})\right), \\ (iii) \quad & \boldsymbol{H}_2(\boldsymbol{x}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \boldsymbol{\Sigma}^{-1}(\boldsymbol{x} - \boldsymbol{\mu})(\boldsymbol{x} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} \\ & \qquad - \boldsymbol{\Sigma}^{-1} - 2\boldsymbol{\Sigma}^{-1}(\boldsymbol{x} - \boldsymbol{\mu}) \boldsymbol{\delta}' \boldsymbol{\Sigma}^{-\frac{1}{2}} \zeta\left(\boldsymbol{\delta}' \boldsymbol{\Sigma}^{-\frac{1}{2}} (\boldsymbol{x} - \boldsymbol{\mu})\right) \\ & \qquad - \boldsymbol{\Sigma}^{-1} \boldsymbol{\delta}\boldsymbol{\delta}' (\boldsymbol{x} - \boldsymbol{\mu}) \boldsymbol{\delta}' \boldsymbol{\Sigma}^{-\frac{1}{2}} \zeta\left(\boldsymbol{\delta}' \boldsymbol{\Sigma}^{-\frac{1}{2}} (\boldsymbol{x} - \boldsymbol{\mu})\right). \end{aligned}$$

where $\zeta(\boldsymbol{x}) = \frac{\phi(\boldsymbol{x})}{\Phi(\boldsymbol{x})}$.

Hermite polynomials have wide-ranging applications in mathematics, statistics, physics, and numerical analysis. In probability, they are used in Edgeworth and Gram-Charlier expansions to refine normal approximations and model deviations from normality. They are essential in Gaussian quadrature for numerical integration and appear in quantum mechanics, describing wavefunctions of harmonic oscillators. In signal processing, they assist in noise reduction and signal approximation, while in machine learning, they aid in feature engineering and capturing non-linear relationships. Additionally, Hermite polynomials are employed in financial mathematics for option pricing and risk analysis, making them indispensable tools in both theoretical and practical domains.

2.2.2. *Marginal and conditional distributions.* In the sequel, we find the marginal and conditional distributions. For this purpose, suppose that the tensor $\mathcal{X}$ is partitioned as $\mathcal{X} = \begin{pmatrix} \mathcal{X}_{r_1} \\ \mathcal{X}_{r_2} \end{pmatrix}$, over the rth mode so that

$$\boldsymbol{x}_{r_1} = vec(\mathcal{X}_{r_1}) = \boldsymbol{M}_{r_1}\boldsymbol{x}, \tag{14}$$

$$\boldsymbol{x}_{r_2} = vec(\mathcal{X}_{r_2}) = \boldsymbol{M}_{r_2}\boldsymbol{x}, \tag{15}$$

where

$$\boldsymbol{M}_{r_1} = \sum_{I_{r_1}} \boldsymbol{e}^{\boldsymbol{p}(1:r-1)}_{i_1:i_{r-1}} \otimes \boldsymbol{e}^{m}_{i_1} \otimes \boldsymbol{e}^{\boldsymbol{p}(r+1:k)}_{i_{r+1}:k} \left(\boldsymbol{e}^{\boldsymbol{p}}_{i_1:i_k}\right)', \tag{16}$$

$$\boldsymbol{M}_{r_2} = \sum_{I_{r_1}} \boldsymbol{e}^{\boldsymbol{p}(1:r-1)}_{i_1:i_{r-1}} \otimes \boldsymbol{e}^{p_r-m}_{i_r-m} \otimes \boldsymbol{e}^{\boldsymbol{p}(r+1:k)}_{i_{r+1}:k} \left(\boldsymbol{e}^{\boldsymbol{p}}_{i_1:i_k}\right)', \tag{17}$$

with their respective index sets

$$I_{r_1} = \{i_1, \cdots, i_k : 1 \le i_t \le p_t, t = 1, \cdots, r-1, r+1, \cdots, k, 1 \le i_r \le m\},$$
$$I_{r_2} = \{i_1, \cdots, i_k : 1 \le i_t \le p_t, t = 1, \cdots, r-1, r+1, \cdots, k, m+1 \le i_r \le p_r\}.$$

Theorem 2.10. *Suppose that* $\boldsymbol{x} \sim SN_{\boldsymbol{p}}(\boldsymbol{0}, \boldsymbol{\Sigma}, \boldsymbol{\delta})$. *Let* $\boldsymbol{x}_{r_1}$, $\boldsymbol{x}_{r_2}$ *be as defined in (14) and (15), respectively,* $\boldsymbol{\delta} = \begin{pmatrix} \boldsymbol{\delta}_1 \\ \boldsymbol{\delta}_2 \end{pmatrix}$ *and* $\boldsymbol{\Sigma}_r = \begin{pmatrix} \boldsymbol{\Sigma}^r_{11} & \boldsymbol{\Sigma}^r_{12} \\ \boldsymbol{\Sigma}^r_{21} & \boldsymbol{\Sigma}^r_{22} \end{pmatrix}$. *Then*

$$\boldsymbol{x}_1 = vec(\mathcal{X}_1) \sim SN_{\boldsymbol{r}_1}(\boldsymbol{0}, \boldsymbol{\Sigma}^r_{11}, \boldsymbol{\delta}_{1(2)}),$$

where

$$\boldsymbol{\delta}_{1(2)} = \left(1 + \boldsymbol{\delta}'_{r_2}\boldsymbol{\Sigma}^r_{2.1}\boldsymbol{\delta}_{r_2}\right)^{-\frac{1}{2}} \left(\boldsymbol{\delta}_{r_1} + {\boldsymbol{\Sigma}^r_{11}}^{-1}\boldsymbol{\Sigma}^r_{12}\boldsymbol{\delta}_{r_2}\right),$$

where

$${\boldsymbol{\Sigma}^r_{11}}^{-1} = \left(\boldsymbol{\Sigma}^r_{11}\right)^{-1}, \qquad \boldsymbol{\Sigma}^r_{2.1} = \boldsymbol{\Sigma}^r_{22} - \boldsymbol{\Sigma}^r_{21}\left(\boldsymbol{\Sigma}^r_{11}\right)^{-1}\boldsymbol{\Sigma}^r_{12}. \tag{18}$$

Proof. First, we find the joint density of $\boldsymbol{x} = (\boldsymbol{x'}_{r_1}, \boldsymbol{x'}_{r_2})'$. since $\begin{pmatrix} \boldsymbol{\Sigma}^r_{11} & \boldsymbol{\Sigma}^r_{12} \\ \boldsymbol{\Sigma}^r_{21} & \boldsymbol{\Sigma}^r_{22} \end{pmatrix}^{-1}$

$$= \begin{pmatrix} (\boldsymbol{\Sigma}^r_{11})^{-1} + (\boldsymbol{\Sigma}^r_{11})^{-1}\boldsymbol{\Sigma}^r_{12}(\boldsymbol{\Sigma}^r_{2.1})^{-1}\boldsymbol{\Sigma}^r_{21}(\boldsymbol{\Sigma}^r_{11})^{-1} & -(\boldsymbol{\Sigma}^r_{11})^{-1}\boldsymbol{\Sigma}^r_{12}(\boldsymbol{\Sigma}^r_{2.1})^{-1} \\ -(\boldsymbol{\Sigma}^r_{2.1})^{-1}\boldsymbol{\Sigma}^r_{21}(\boldsymbol{\Sigma}^r_{11})^{-1} & (\boldsymbol{\Sigma}^r_{2.1})^{-1} \end{pmatrix},$$

where $\boldsymbol{\Sigma}^r_{2.1}$ is given in equation (18). Also $|\boldsymbol{\Sigma}^r_{11}||\boldsymbol{\Sigma}^r_{2.1}| = |\boldsymbol{\Sigma}^r_{22}||\boldsymbol{\Sigma}^r_{1.2}|$ hence

$$f_{\mathcal{X}_{r_1},\mathcal{X}_{r_2}}(\boldsymbol{x}_{r_1},\boldsymbol{x}_{r_2}) = 2\phi_{r_1}\left(\boldsymbol{x}_{r_1},\boldsymbol{\mu}_{r_1},\boldsymbol{\Sigma}^r_{11}\right)\phi_{r_2}\left(\boldsymbol{x}_{r_2},\boldsymbol{\mu}_{r_2},\boldsymbol{\Sigma}^r_{2.1}\right)\Phi\left(\boldsymbol{\delta'}_{r_1}\boldsymbol{x}_{r_1}+\boldsymbol{\delta'}_{r_2}\boldsymbol{x}_{r_2}\right)$$

$$f_{\mathcal{X}_{r_1},\mathcal{X}_{r_2}}(\boldsymbol{x}_{r_1},\boldsymbol{x}_{r_2}) = 2\phi_{r_1}\left(\boldsymbol{x}_{r_1},\boldsymbol{0},\boldsymbol{\Sigma}^r_{11}\right)\phi_{r_2}\left(\boldsymbol{x}_{r_2},\boldsymbol{\mu}_{2.1},\boldsymbol{\Sigma}^r_{2.1}\right)\Phi\left(\boldsymbol{\delta}^{r'}_{2.1}\left(\boldsymbol{x}_{r_2}-\boldsymbol{\mu}^r_{2.1}\right)+\boldsymbol{\delta}^r_{1.1}\boldsymbol{x}_{r_1}\right)$$

where $\boldsymbol{\delta}^r_{2.1} = \boldsymbol{\delta}_{r_2}$, $\boldsymbol{\delta}^r_{1.1} = \boldsymbol{\delta}_{r_1} + \boldsymbol{\Sigma}^{r^{-1}}_{11}\boldsymbol{\Sigma}^r_{12}\boldsymbol{\delta}_{r_2}$, $\boldsymbol{\mu}^r_{2.1} = \boldsymbol{\Sigma}^r_{21}\boldsymbol{\Sigma}^{r^{-1}}_{11}\boldsymbol{x}_{r_1}$ and $\boldsymbol{\delta}^{r'}_{2.1}\left(\boldsymbol{x}_{r_2}-\boldsymbol{\mu}^r_{2.1}\right)+\boldsymbol{\delta}^r_{1.1}\boldsymbol{x}_{r_1} = \boldsymbol{\delta'}_{r_1}\boldsymbol{x}_{r_1}+\boldsymbol{\delta'}_{r_2}\boldsymbol{x}_{r_2}$,

Now by integrating the joint density over $\boldsymbol{x}_{r_2}$ and using Lemma 5.2 in [3], we get the marginal density, and the proof is complete.

$$\begin{aligned} f_{\mathcal{X}_{r_1}}(\boldsymbol{x}_{r_1}) &= \int_{\mathbb{R}^{r_2}} f_{\mathcal{X}_{r_1},\mathcal{X}_{r_2}}(\boldsymbol{x}_{r_1},\boldsymbol{x}_{r_2})\,d\boldsymbol{x}_2 \\ &= 2\phi_{r_1}\left(\boldsymbol{x}_{r_1},\boldsymbol{\mu}_{r_1},\boldsymbol{\Sigma}^r_{11}\right)E\left[\Phi\left(\boldsymbol{\delta'}^r_{2.1}\left(\boldsymbol{x}_{r_2}-\boldsymbol{\mu}^r_{2.1}\right)+\boldsymbol{\delta}^r_{1.1}\boldsymbol{x}_{r_1}\right)\right] \\ &= 2\phi_{r_1}\left(\boldsymbol{x}_{r_1},\boldsymbol{\mu}_{r_1},\boldsymbol{\Sigma}^r_{11}\right)\Phi\left[\left(1+\boldsymbol{\delta}'^r_{2.1}\boldsymbol{\Sigma}^r_{2.1}\boldsymbol{\delta}^r_{2.1}\right)^{-\frac{1}{2}}\boldsymbol{\delta}^r_{1.1}\boldsymbol{x}_{r_1}\right]. \end{aligned}$$

□

Corollary 2.11. *Under the assumptions of theorem 2.10 the conditional density of $\boldsymbol{x}_2|\boldsymbol{x}_1$ is as follows*

$$f_{\mathcal{X}_2|\mathcal{X}_1}(\boldsymbol{x}_2|\boldsymbol{x}_1) = \frac{1}{\Phi^*(\tau_{2.1})}\phi_{\boldsymbol{r}_2}\left(\boldsymbol{x}_{r_2},\boldsymbol{\mu}^r_{2.1},\boldsymbol{\Sigma}^r_{2.1}\right)\Phi\left[\boldsymbol{\delta}^{r'}_{2.1}\left(\boldsymbol{x}_{r_2}-\boldsymbol{\mu}^r_{2.1}\right)+\tau^r_{2.1}\right],$$

where

$$\begin{aligned} \boldsymbol{\mu}^r_{2.1} &= \boldsymbol{\Sigma}^r_{21}\boldsymbol{\Sigma}^{-1^r}_{11}\boldsymbol{x}_{r_1}, \\ \boldsymbol{\Sigma}^r_{2.1} &= \boldsymbol{\Sigma}^r_{22}-\boldsymbol{\Sigma}^r_{21}(\boldsymbol{\Sigma}^r_{11})^{-1}\boldsymbol{\Sigma}^r_{12}, \\ \boldsymbol{\delta}^r_{1.1} &= \boldsymbol{\delta}_{r_1}+\boldsymbol{\Sigma}^{-1^r}_{11}\boldsymbol{\Sigma}^r_{12}\boldsymbol{\delta}_{r_2}, \\ \boldsymbol{\delta}^r_{2.1} &= \boldsymbol{\delta}_{r_2}, \\ \tau^r_{2.1} &= \boldsymbol{\delta}^{r'}_{1.1}\boldsymbol{x}_{r_1}, \\ \Phi^*(\tau_{2.1}) &= \Phi\left[\left(1+\boldsymbol{\delta}'^r_{2.1}\boldsymbol{\Sigma}^r_{2.1}\boldsymbol{\delta}^r_{2.1}\right)^{-\frac{1}{2}}\boldsymbol{\tau}^r_{2.1}\right]. \end{aligned}$$

Proof. Using the law of obtaining conditional density and the joint and marginal densities obtained from theorem 2.10, the proof is complete. □

At the end of this section, we state the following theorem in which several interesting special cases can be investigated by choosing $\boldsymbol{A}$ appropriately.

Theorem 2.12. *Let $\boldsymbol{x} \sim SN_{\boldsymbol{p}}(\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\delta})$, where $\boldsymbol{x} \in \mathcal{T}^{\boldsymbol{p}}$, and let $\boldsymbol{A} \in \mathcal{T}^{\boldsymbol{qp}}_{\otimes}$ be nonsingular. Then,*

$$\boldsymbol{Ax} \sim SN_{\boldsymbol{q}}\left(\boldsymbol{\mu_A},\boldsymbol{\Sigma_A},\boldsymbol{\delta_A}\right),$$

Where $\boldsymbol{\mu_A} = \boldsymbol{A\mu}$, $\boldsymbol{\Sigma_A} \in \mathcal{T}^{\boldsymbol{q}}_{\otimes} = \boldsymbol{A\Sigma A'}$ and $\boldsymbol{\delta_A} = \boldsymbol{\Sigma}^{\frac{1}{2}}_{\boldsymbol{A}}\left(\boldsymbol{A}^{-1}\right)'\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{\delta}$.

Proof. Put $\boldsymbol{y} = vec(\mathcal{Y}) = \boldsymbol{Ax} \in \mathcal{T}^{\boldsymbol{q}}$. Using of transformation technique,

$$\begin{aligned} f_{\mathcal{Y}}(\boldsymbol{y}) &= |\boldsymbol{A}|^{-1} f_{\mathcal{X}}\left(\boldsymbol{A}^{-1}\boldsymbol{y}\right) \\ &= 2(2\pi)^{-\frac{q^*}{2}} |\boldsymbol{A\Sigma A'}|^{-\frac{1}{2}} exp\left[-\frac{1}{2}(\boldsymbol{y}-\boldsymbol{A\mu})'(\boldsymbol{A\Sigma A'})^{-1}(\boldsymbol{y}-\boldsymbol{A\mu})\right] \\ &\quad \times \Phi\left[\boldsymbol{\delta}'\boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{A}^{-1}(\boldsymbol{y}-\boldsymbol{A\mu})\right]. \end{aligned}$$

And the proof is complete. □

3. **Parameter Estimation via EM Algoritheorem**

Assume that there are n independent tensor observations $\boldsymbol{\mathcal{X}} = (\mathcal{X}_1, \cdots, \mathcal{X}_n)$, from $f_{\mathcal{X}}(\boldsymbol{x})$, with vector representation $\boldsymbol{x} = (\mathbf{x_1}, \mathbf{x_2}, \cdots, \mathbf{x_n})$. Let $\boldsymbol{\theta}$ be the parameter vector that consists of $\boldsymbol{\mu}, \boldsymbol{\Sigma_1}, ..., \boldsymbol{\Sigma_k}$ and $\boldsymbol{\delta}$, (or $\boldsymbol{\lambda}$). In this section we consider the maximum likelihood estimation of $\boldsymbol{\theta}$ based on $\mathbf{x_1}, \mathbf{x_2}, ..., \mathbf{x_n}$ and provide an EM algorithm in a closed form.
Writing the cdf Φ in (7) as the integral of its pdf ϕ leads to the following representation of the Skew Normal distribution
(19)

$$f(\boldsymbol{x}, \boldsymbol{\mu}, \boldsymbol{\Sigma}_1, ..., \boldsymbol{\Sigma}_k, \boldsymbol{\delta}) = 2\int_0^{\infty} \phi_{\mathbf{p}}(\boldsymbol{x}, \boldsymbol{\mu}, \boldsymbol{\Sigma}_1, ..., \boldsymbol{\Sigma}_k, \boldsymbol{\delta})\phi(u - \boldsymbol{\delta}'\boldsymbol{\Sigma}_{1:k}^{-\frac{1}{2}}(\boldsymbol{x}-\boldsymbol{\mu}))du,$$

where $\mathbf{p} = (p_1, ..., p_k), \boldsymbol{\Sigma}_{1:k} = \boldsymbol{\Sigma}_1 \otimes ... \otimes \boldsymbol{\Sigma}_k$. This suggests introducing a non negative random variable or a latent variable U such that the joint density function is just the integrand in (19), i.e.,

$$f(\boldsymbol{x}, u) = 2\phi_{\mathbf{p}}(\boldsymbol{x}, \boldsymbol{\mu}, \boldsymbol{\Sigma}_1, ..., \boldsymbol{\Sigma}_k, \boldsymbol{\delta})\phi(u - \boldsymbol{\delta}'\boldsymbol{\Sigma}_{1:k}^{-\frac{1}{2}}(\boldsymbol{x}-\boldsymbol{\mu}))I(u > 0).$$

By definition, the conditional distribuion of U given $\mathbf{X} = \boldsymbol{x}$ is

$$f(u|\boldsymbol{x}) = \frac{\phi(u - \boldsymbol{\lambda}'(\boldsymbol{x}-\boldsymbol{\mu}))}{\Phi(\boldsymbol{\lambda}'(\boldsymbol{x}-\boldsymbol{\mu}))} I(u > 0),$$

where $\boldsymbol{\lambda} = \boldsymbol{\Sigma}_{1:k}^{-\frac{1}{2}}\boldsymbol{\delta}$. So, we can write the conditional mean as

$$E(U|\boldsymbol{x}, \boldsymbol{\theta}) = \boldsymbol{\lambda}'(\boldsymbol{x}-\boldsymbol{\mu}) + \frac{\phi(\boldsymbol{\lambda}'(\boldsymbol{x}-\boldsymbol{\mu}))}{\Phi(\boldsymbol{\lambda}'(\boldsymbol{x}-\boldsymbol{\mu}))}. \tag{20}$$

Let $(\boldsymbol{x}_i, u_i), i = 1, ..., n$ be the complete data, where $\boldsymbol{x}_i$ and u_i are considered as incomplete observed and missing data, respectively. The complete data log-likelihood function for $\boldsymbol{\theta} = (\boldsymbol{\mu}, \boldsymbol{\Sigma_1}, ..., \boldsymbol{\Sigma_k}, \boldsymbol{\delta}(\boldsymbol{\lambda}))$ is given by
(21)

$$l_c(\boldsymbol{\theta}) = C - n\sum_{i=1}^{k} \frac{p^*}{2p_i} log|\boldsymbol{\Sigma_i}| - \frac{1}{2}\sum_{i=1}^{n}(\boldsymbol{x}_i-\boldsymbol{\mu})'\boldsymbol{\Sigma}_{1:k}^{-1}(\boldsymbol{x}_i-\boldsymbol{\mu}) - \frac{1}{2}\sum_{i=1}^{n}(u_i - \boldsymbol{\lambda}'(\boldsymbol{x}_i-\boldsymbol{\mu})),$$

where C is a constant that does not depend on the parameters. We proceed by using an ECM algorithm described below.

- Initialization: Initialize the parameters $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}_j$ for $j = 1, \ldots, p$ and $\boldsymbol{\delta}$ (or $\boldsymbol{\lambda}$).
- E Step: The expected value of the complete data log-likelihood $l_c(\boldsymbol{\theta})$with respect to the conditional distribution of the missing data u_i given the observed data $\boldsymbol{x}_i$ and the current estimate of the parameter $\hat{\boldsymbol{\theta}}^{(m)}$ at the iteration mth is

$$\begin{aligned} Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(m)}) &= E(l_c(\boldsymbol{\theta})|\boldsymbol{x}, \hat{\boldsymbol{\theta}}^{(m)}) \\ &= C - n\sum_{i=1}^{k} \frac{p^*}{2p_i} log|\boldsymbol{\Sigma_i}| - \frac{1}{2}\sum_{i=1}^{n}(\boldsymbol{x}_i - \boldsymbol{\mu})'\boldsymbol{\Sigma}_{1:k}^{-1}(\boldsymbol{x}_i - \boldsymbol{\mu}) \\ &- \frac{1}{2}\sum_{i=1}^{n}(E(U_i|\boldsymbol{x}_i, \hat{\boldsymbol{\theta}}^{(m)}) - \boldsymbol{\lambda}'(\boldsymbol{x}_i - \boldsymbol{\mu}))^2. \end{aligned}$$

- CM Step: By maximizing $Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(m)})$ over $\boldsymbol{\theta}$, drive $\hat{\boldsymbol{\theta}}^{(m+1)}$ which are given by
  (1) Update $\boldsymbol{\mu}$

$$\hat{\boldsymbol{\mu}}^{(m+1)} = \bar{\boldsymbol{x}} - (\hat{\boldsymbol{\Sigma}}_{1:k}^{(m)})^{\frac{1}{2}}\hat{\boldsymbol{\delta}}^{(m)} c(\hat{\boldsymbol{\delta}}^{(m)})\frac{1}{n}\sum_{i=1}^{n} a(\hat{\boldsymbol{\theta}}^{(m)}|\boldsymbol{x}_i),$$

  where $c(\boldsymbol{\delta}) = (1 + \boldsymbol{\delta}'\boldsymbol{\delta})^{-1}$ and $a(\boldsymbol{\theta}|\boldsymbol{x}_i) = E(U_i|\boldsymbol{x}_i, \boldsymbol{\theta})$ in (20).

  (2) Update $\boldsymbol{\Sigma}_j, j = 1, ..., k$,

$$\hat{\boldsymbol{\Sigma}}_j^{(m+1)} = \frac{p_j}{np^*}\sum_{i=1}^{n}(\boldsymbol{x}_{i(j)} - \hat{\boldsymbol{\mu}}_j^{(m+1)})' \otimes_{d\neq j} (\hat{\boldsymbol{\Sigma}}_d^{(m)})^{-1}(\boldsymbol{x}_{i(j)} - \hat{\boldsymbol{\mu}}_j^{(m+1)}),$$

  also $\hat{\boldsymbol{\Sigma}}^{(m+1)} = \hat{\boldsymbol{\Sigma}}_{1:k}^{(m+1)} = \frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{x}_i - \hat{\boldsymbol{\mu}}^{(m+1)})'(\boldsymbol{x}_i - \hat{\boldsymbol{\mu}}^{(m+1)})$. As was discussed in [2] and [7] for parameter estimation in the matrix variate case, and in [16] for the order k case, the estimates of $\boldsymbol{\Sigma}_k$ are unique only up to a multiplicative constant. Indeed, if we let

$$\tilde{\boldsymbol{\Sigma}}_{(k)} = \frac{1}{\sigma_{(k),11}}\boldsymbol{\Sigma}_{(k)},\ k = 2, ..., p,$$

  where $\sigma_{(k),11}$ is the first entry in $\boldsymbol{\Sigma}_k$ and

$$\tilde{\boldsymbol{\Sigma}}_{(1)} = \prod_{k=2}^{p}\sigma_{(k),11}\boldsymbol{\Sigma}_{(1)}, p$$

  then

$$\bigotimes_{k=1}^{p}\tilde{\boldsymbol{\Sigma}}_{(k)} = \bigotimes_{k=1}^{p}\boldsymbol{\Sigma}_{(k)},$$

and therefore the likelihood is unchanged. So the estimate of the Kronecker product would be unique.

(3) Update $\boldsymbol{\delta} = \boldsymbol{\Sigma}_{1:k}^{-\frac{1}{2}}\boldsymbol{\lambda}$,

$$\hat{\boldsymbol{\delta}}^{(m+1)} = (\hat{\boldsymbol{\Sigma}}_{1:k}^{(m+1)})^{-\frac{1}{2}}\frac{1}{n}\sum_{i=1}^{n} a(\hat{\boldsymbol{\theta}}^{(m)}|\boldsymbol{x}_i)(\boldsymbol{x}_i - \hat{\boldsymbol{\mu}}^{(m+1)}).$$

The iterations of the EM algorithm is repeated until the desired convergence criteria is achieved. A discussion of the computational complexity of the proposed EM algorithm is particularly pertinent in high-dimensional settings, where several challenges may arise. As the dimensionality of the data increases, both the E-step and M-step of the EM algorithm become computationally expensive. This can lead to excessive memory usage and significantly longer execution times. Moreover, the algorithm may converge slowly or become trapped in local optima, which affects the quality and efficiency of the results. To address these and other computational challenges, several optimization strategies can be considered. Regularization techniques such as penalties can help prevent overfitting and improve numerical stability. Also, parallelization techniques leveraging multi-core processors or GPUs can substantially improve the scalability of the algorithm in large-scale applications.

4. **A simulation study**

To evaluate the performance of the estimators considered in the previous section for varying samples sizes, we consider the tensor variate skew normal distribution and generate random samples of size $n = 100, 150, 200$ with 1000 replications from the $SN_{\boldsymbol{p}}(\boldsymbol{\mu}, \boldsymbol{\Sigma}_{1:k}, \boldsymbol{\delta})$ where $\boldsymbol{p} = (2,2,2), \boldsymbol{p} = (3,3,3), \boldsymbol{p} = (8,8,8), \boldsymbol{p} = (10,10,10)$ and $\boldsymbol{\Sigma}_i = \boldsymbol{I}_p, i = 1,2,3$, $\boldsymbol{p} = (2,2,2)$;
Also
$\boldsymbol{\mu}_{::1} = \begin{bmatrix} 0.5 & 1.0 \\ -0.5 & 0.5 \end{bmatrix}, \boldsymbol{\mu}_{::2} = \begin{bmatrix} 1.0 & -0.5 \\ 0.5 & 1.0 \end{bmatrix}$; $\boldsymbol{\delta}_{::1} = \begin{bmatrix} 0.5 & 0.7 \\ 1.0 & 0.5 \end{bmatrix}, \boldsymbol{\delta}_{::2} = \begin{bmatrix} 0.7 & 1.0 \\ 0.5 & 0.7 \end{bmatrix}$.
Note that for generating a sample tensor from the tensor variate skew normal distribution, we can use package rTensor in R software.
Finally, as initial values for the EM algorithm, we use the sample mean tensor $\boldsymbol{\mu}_0 = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i$, the sample covariance matrix and the moment estimator of $\boldsymbol{\delta}$ in [**?**] as $\boldsymbol{\Sigma}_j^0 = \frac{p_j}{np^*}\sum_{i=1}^{n}(\boldsymbol{x}_{i(j)} - \boldsymbol{\mu}_j^0)'(\boldsymbol{x}_{i(j)} - \boldsymbol{\mu}_j^0)$, $\boldsymbol{\delta}^0 = \left(\frac{\frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{x}_i - \boldsymbol{\mu}^0)^3}{(\frac{4}{\pi}-1)\sqrt{\frac{2}{\pi}}}\right)^{\frac{1}{3}}$.
Also, we consider the convergence criterion $max_{r\in 1,\ldots,d}|\boldsymbol{\theta}_r^{(m+1)} - \boldsymbol{\theta}_r^{(m)}| < 10^{-2}$, where d is the size of $\boldsymbol{\theta}$.
The results, contain the average bias and the mean squared error of the estimates, are reported in Tables 1,2 and 3 for $\boldsymbol{\mu}, \boldsymbol{\delta}$ and $\boldsymbol{\Sigma}_j, j = 1, \cdots, p$ respectively. It can be seen from these Tables that, by increasing the sample size (n),

the absolute value of the average bias and the mean squared error of the MLE decrease for each component of $\boldsymbol{\theta}$, thus verifying the consistency property of the MLE, which states that as the sample size grows, the estimates produced by MLE converge in probability to the true parameter values. Moreover, the performance is better for lower values of p^*. This suggests that when the number of parameters being estimated (or the dimensionality of the data) is lower, the estimates are more accurate, resulting in lower bias and MSE. This could be due to reduced complexity and less chance of overfitting when fewer parameters are involved.

Relative error is a measure of the uncertainty or inaccuracy of a measurement compared to the true value. It is expressed as a fraction or percentage of the true value, providing a normalized way to assess the size of the error in relation to the actual quantity being measured. The formula for calculating relative error is $\frac{\|\hat{\boldsymbol{V}}-\boldsymbol{V}\|_F}{\|\boldsymbol{V}\|_F}$, where $\|.\|_F$ is the Frobenius matrix norm,$\hat{\boldsymbol{V}}$ is the estimated parameter value, and $\boldsymbol{V}$ is the true parameter value used to generate the simulated data. Relative error is particularly useful in scientific and engineering contexts, as it allows for the comparison of errors across different scales and units. A smaller relative error indicates a more precise measurement, while a larger relative error suggests greater uncertainty. We use the relative error in Figures 1 and 2 to determine how close the estimated model parameters are to the true parameters and visualize that the performance improves as the sample size increases. Moreover, the performance is better for lower values of p^*. In other words larger tensors result in elevated relative errors.

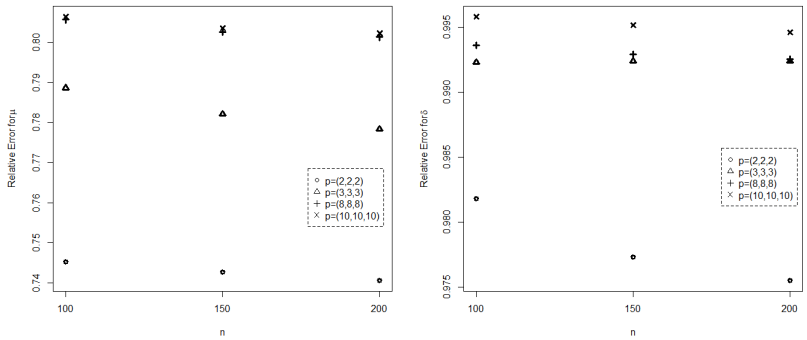

FIGURE 1. Relative Error for $\boldsymbol{\mu}$ (left) and $\boldsymbol{\delta}$ (right)

5. **Image Analysis**

We now observe an analysis of red-green-blue (RGB) images. These images come in the form three color intensity matrices (red, green, and blue), on

TABLE 1. Average bias and MSE for $\boldsymbol{\mu}$

(p_1, p_2, p_3)	Component	Bias			MSE		
		$n = 100$	$n = 150$	$n = 200$	$n = 100$	$n = 150$	$n = 200$
(2,2,2)	μ_{111}	0.3574	0.3561	0.3546	0.1374	0.1325	0.1304
	μ_{211}	0.5555	0.5581	0.5584	0.3161	0.3163	0.3153
	μ_{121}	0.7027	0.7010	0.6999	0.4991	0.4955	0.4927
	μ_{221}	0.3561	0.3554	0.3578	0.1358	0.1330	0.1326
	μ_{112}	0.5588	0.5592	0.5622	0.3198	0.3176	0.3196
	μ_{212}	0.7033	0.7048	0.7024	0.4996	0.5007	0.4959
	μ_{122}	0.3564	0.3576	0.3585	0.1363	0.1347	0.1332
	μ_{222}	0.5626	0.5607	0.5589	0.3240	0.3194	0.3157
(3,3,3)	μ_{111}	0.3530	0.3509	0.3540	0.1348	0.1301	0.1312
	μ_{211}	0.5487	0.5475	0.5458	0.3088	0.3051	0.3027
	μ_{311}	0.6867	0.6808	0.6779	0.4770	0.4674	0.4629
	μ_{121}	0.3552	0.3542	0.3492	0.1367	0.1325	0.1276
	μ_{221}	0.5499	0.5447	0.5462	0.3097	0.3023	0.3028
	μ_{321}	0.6828	0.6820	0.6736	0.4721	0.4693	0.4569
	μ_{131}	0.3583	0.3534	0.3518	0.1383	0.1322	0.1297
	μ_{231}	0.5525	0.5548	0.5458	0.3129	0.3131	0.3028
	μ_{331}	0.6833	0.6811	0.6741	0.4722	0.4679	0.4575
	μ_{112}	0.3541	0.3523	0.3527	0.1349	0.1306	0.1296
	μ_{212}	0.5548	0.5478	0.5497	0.3157	0.3052	0.3063
	μ_{312}	0.6913	0.6894	0.6839	0.4832	0.4790	0.4706
	μ_{122}	0.3539	0.3526	0.3530	0.1349	0.1305	0.1296
	μ_{222}	0.5517	0.5503	0.5515	0.3121	0.3077	0.3082
	μ_{322}	0.6935	0.6901	0.6868	0.4858	0.4798	0.4746
	μ_{132}	0.3511	0.3566	0.3577	0.1323	0.1341	0.1333
	μ_{232}	0.5538	0.5510	0.5515	0.3140	0.3085	0.3082
	μ_{332}	0.6948	0.6885	0.6864	0.4881	0.4777	0.4740
	μ_{113}	0.3603	0.3541	0.3568	0.1390	0.1314	0.1325
	μ_{213}	0.5554	0.5557	0.5544	0.3162	0.3134	0.3107
	μ_{313}	0.7001	0.6976	0.6938	0.4948	0.4902	0.4840
	μ_{123}	0.3575	0.3526	0.3565	0.1367	0.1307	0.1319
	μ_{223}	0.5564	0.5551	0.5558	0.3168	0.3135	0.3126
	μ_{323}	0.6987	0.6935	0.6948	0.4930	0.4843	0.4854
	μ_{133}	0.3521	0.3543	0.3528	0.1331	0.1321	0.1293
	μ_{233}	0.5561	0.5548	0.5512	0.3174	0.3126	0.3078
	μ_{333}	0.6993	0.6928	0.6921	0.4939	0.4833	0.4815

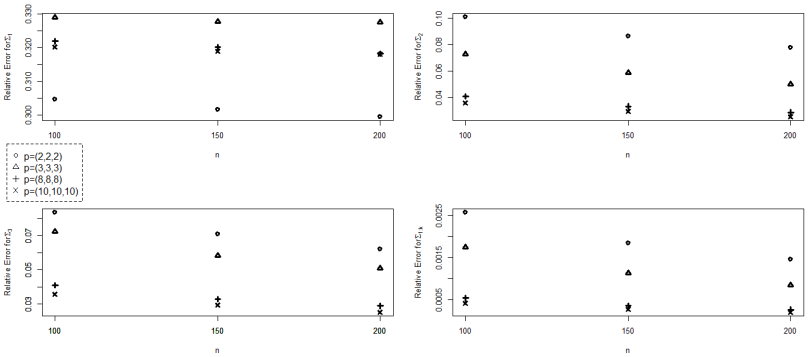

FIGURE 2. Relative Error for $\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2, \boldsymbol{\Sigma}_3$ and $\boldsymbol{\Sigma}_{1:3}$

TABLE 2. Average bias and MSE for $\boldsymbol{\delta}$

(p_1, p_2, p_3)	Component	Bias			MSE		
		$n = 100$	$n = 150$	$n = 200$	$n = 100$	$n = 150$	$n = 200$
(2,2,2)	δ_{111}	-0.4980	-0.4914	-0.4950	0.2508	0.2444	0.2480
	δ_{211}	-0.9830	-0.9773	-0.9731	0.9682	0.9572	0.9489
	δ_{121}	-1.9580	-1.9495	-1.9463	3.8348	3.8017	3.7889
	δ_{221}	-0.4972	-0.4947	-0.4946	0.2498	0.2475	0.2475
	δ_{112}	-0.9802	-0.9774	-0.9745	0.9627	0.9572	0.9516
	δ_{212}	-1.9613	-1.9536	-1.9503	3.8479	3.8174	3.8044
	δ_{122}	-0.4956	-0.4917	-0.4939	0.2482	0.2446	0.2467
	δ_{222}	-0.9822	-0.9769	-0.9731	0.9666	0.9563	0.9486
(3,3,3)	δ_{111}	-0.4971	-0.4983	-0.4985	0.2481	0.2492	0.2494
	δ_{211}	-0.9917	-0.9909	-0.9918	0.9843	0.9826	0.9845
	δ_{311}	-1.9826	-1.9820	-1.9823	3.9314	3.9289	3.9304
	δ_{121}	-0.4971	-0.4986	-0.4989	0.2482	0.2496	0.2498
	δ_{221}	-0.9932	-0.9899	-0.9916	0.9873	0.9807	0.9839
	δ_{321}	-1.9808	-1.9824	-1.9825	3.9243	3.9308	3.9312
	δ_{131}	-0.4973	-0.4976	-0.4985	0.2483	0.2485	0.2494
	δ_{231}	-0.9899	-0.9926	-0.9908	0.9806	0.9859	0.9825
	δ_{331}	-1.9818	-1.9825	-1.9826	3.9282	3.9308	3.9316
	δ_{112}	-0.4981	-0.4977	-0.4992	0.2489	0.2484	0.2500
	δ_{212}	-0.9913	-0.9918	-0.9918	0.9833	0.9843	0.9842
	δ_{312}	-1.9835	-1.9838	-1.9842	3.9347	3.9360	3.9379
	δ_{122}	-0.4978	-0.4990	-0.4983	0.2487	0.2497	0.2490
	δ_{222}	-0.9920	-0.9924	-0.9906	0.9847	0.9855	0.9819
	δ_{322}	-1.9835	-1.9842	-1.9839	3.9349	3.9375	3.9364
	δ_{132}	-0.4978	-0.4981	-0.4976	0.2486	0.2488	0.2484
	δ_{232}	-0.9920	-0.9912	-0.9910	0.9847	0.9831	0.9826
	δ_{332}	-1.9832	-1.9843	-1.9838	3.9335	3.9380	3.9362
	δ_{113}	-0.4997	-0.4986	-0.4992	0.2502	0.2492	0.2497
	δ_{213}	-0.9941	-0.9927	-0.9926	0.9886	0.9859	0.9858
	δ_{313}	-1.9859	-1.9864	-1.9862	3.9443	3.9462	3.9455
	δ_{123}	-0.4986	-0.4982	-0.4985	0.2492	0.2487	0.2489
	δ_{223}	-0.9939	-0.9931	-0.9921	0.9883	0.9867	0.9848
	δ_{323}	-1.9862	-1.9862	-1.9862	3.9452	3.9456	3.9457
	δ_{133}	-0.4987	-0.4989	-0.4988	0.2493	0.2494	0.2493
	δ_{233}	-0.9942	-0.9927	-0.9924	0.9890	0.9859	0.9853
	δ_{333}	-1.9859	-1.9861	-1.9862	3.9441	3.9452	3.9453

top of each other, thus creating an order-3 tensor. The images come from the CIFAR-100 dataset [10]. For more information about these images please see the following address: https://www.cs.toronto.edu/ kriz/cifar.html . Here images of maple trees that had green or yellow leaves are chosen and came from the following CIFAR-100 class hierarchy: super class trees and class maple. The maple tree images were converted from RGB arrays to an HSL format to filter out trees that did not have green or yellow leaves. These tensors had an $p = (32, 32, 3)$.

Figure 3 (left) is an example of one of the images in our sample of 212 tensors Using EM algorithm, we estimate the parameter of the TVSN distribution and Figure 3 (right) shows the color image that results from the estimated $E(\mathcal{X})$ tensor. The sky, tree trunk and branches are clearly visible. This feature can be useful in areas such as automatic vegetation classification, object detection in natural images, and image data compression. Also, Figure 4 displays the distribution of pixel intensities for both the original maple tree image and the

TABLE 3. Average bias and MSE for $\mathbf{\Sigma_r}, r = 1, 2, 3$

(p_1, p_2, p_3)	Component		Bias			MSE		
			$n = 100$	$n = 150$	$n = 200$	$n = 100$	$n = 150$	$n = 200$
(2,2,2)	Σ_1	σ_{11}	-0.2799	-0.2794	-0.2763	0.0839	0.0817	0.0790
		$\sigma_{12} = \sigma_{21}$	0.0001	-0.0003	0.0001	0.0012	0.0009	0.0006
		σ_{22}	-0.3186	-0.3159	-0.3168	0.1070	0.1034	0.1029
	Σ_2	σ_{11}	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
		$\sigma_{12} = \sigma_{21}$	-0.0004	0.0002	0.0015	0.0027	0.0020	0.0014
		σ_{22}	0.0925	0.0894	0.0848	0.0222	0.0159	0.0131
	Σ_3	σ_{11}	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
		$\sigma_{12} = \sigma_{21}$	0.0020	0.0010	0.0024	0.0024	0.0016	0.0012
		σ_{22}	-0.0417	-0.0487	-0.0497	0.0128	0.0097	0.0075
(3,3,3)	Σ_1	σ_{11}	-0.0750	-0.0786	-0.0777	0.0106	0.0095	0.0085
		$\sigma_{12} = \sigma_{21}$	0.0007	0.0009	0.0015	0.0008	0.0005	0.0004
		$\sigma_{13} = \sigma_{31}$	-0.0004	-0.0004	-0.0002	0.0005	0.0004	0.0003
		σ_{22}	-0.2784	-0.2802	-0.2808	0.0803	0.0803	0.0801
		$\sigma_{23} = \sigma_{32}$	0.0018	-0.0002	0.0016	0.0004	0.0003	0.0002
		σ_{33}	-0.4817	-0.4810	-0.4819	0.2333	0.2322	0.2328
	Σ_2	σ_{11}	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
		$\sigma_{12} = \sigma_{21}$	0.0005	0.0012	0.0004	0.0013	0.0009	0.0006
		$\sigma_{13} = \sigma_{31}$	0.0003	0.0004	0.0004	0.0013	0.0008	0.0006
		σ_{22}	0.0076	0.0026	0.0014	0.0055	0.0035	0.0024
		$\sigma_{23} = \sigma_{32}$	0.0008	0.0006	0.0004	0.0013	0.0008	0.0007
		σ_{33}	0.0037	0.0008	0.0010	0.0052	0.0032	0.0025
	Σ_3	σ_{11}	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
		$\sigma_{12} = \sigma_{21}$	0.0013	-0.0008	0.0008	0.0012	0.0008	0.0006
		$\sigma_{13} = \sigma_{31}$	0.0015	0.0005	-0.0004	0.0013	0.0008	0.0007
		σ_{22}	-0.0026	0.0026	-0.0003	0.0051	0.0032	0.0025
		$\sigma_{23} = \sigma_{32}$	0.0019	0.0011	0.0017	0.0013	0.0008	0.0007
		σ_{33}	0.0004	-0.0002	0.0008	0.0053	0.0035	0.0025

estimated mean tensor. Table 4 shows the AIC, BIC and the Logarithm of the

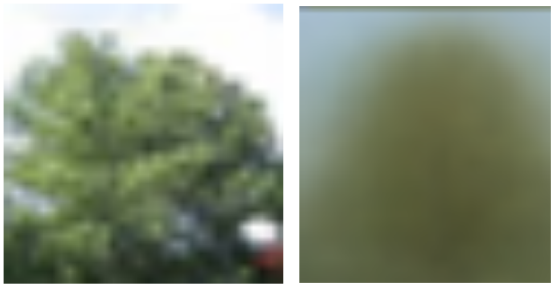

FIGURE 3. An image of a maple tree from CIFAR-100 dataset (left) and An image of the estimated mean tensor (right)

likelihood function criteria to select the better model for this data, Considering that the skew normal distribution has the highest value of the logarithm of the likelihood and the lowest values of AIC and BIC,therefore Skew Normal distribution obtained the best performance, which can serve as a useful guide

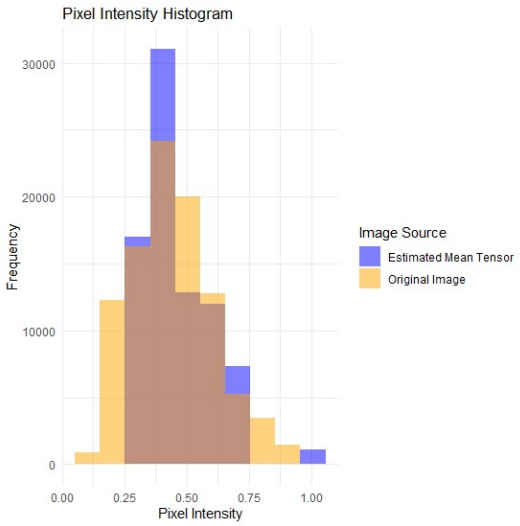

FIGURE 4. The distribution of pixel intensities for both the original maple tree image and the estimated mean tensor

for model selection in real-world image processing projects—especially when the data exhibit asymmetry.

	AIC	BIC	LNL
Normal	22582000	22607000	-11284000
SkewNormal	22562000	22587000	-11274000

TABLE 4. AIC,BIC,LNL results for the image analysis

6. Conclusion

In this paper, we introduced the Tensor Variate Skew-Normal (TVSN) distribution and presented its key properties, such as moments, characteristic functions. We also derived multivariate Hermite polynomials, which are useful for Edgeworth expansions and numerical integration such as Gaussian Quadrature. Through simulation studies, we evaluated the parameter estimation process using the EM algorithm, reporting the average bias and mean squared error (MSE) of the estimates. The model's relative error was also calculated, showing how closely the estimated parameters matched the true values. Although the TVSN model performed well in low-dimensional cases, we noted that parameter estimation can become computationally infeasible in higher tensor dimensions due to the increased complexity. The TVSN distribution was fitted to a dataset of colored images of maple trees, and its performance

was compared to the standard tensor variate normal and skew-normal distributions using model selection criteria such as AIC, BIC, and LNL. The results demonstrated that the TVSN model provided the best fit to the data, accurately representing the trunk, branches, leaves, and sky in the images.

In future work, we propose exploring dimensionality reduction techniques, such as PCA or autoencoders, to handle high-dimensional image data more efficiently. We also suggest developing more scalable algorithms for parameter estimation and considering the use of finite mixture models for clustering and classification tasks. Overall, the TVSN model provides a powerful tool for image analysis and opens up new possibilities for modeling complex, high-dimensional data.

7. Aknowledgement

The authors are very thankful to the anonymous referees for their valuable comments and suggestions, which have improved the manuscript immensely.

References

[1] Arashi, M. (2021). Some theoretical results on tensor elliptical distribution. *Computational Statistics and Modeling,* **1**(2), 29–41. https://doi.org/10.48550/arXiv.1709.00801

[2] Anderlucci, L. and Viroli, C. (2015). Covariance pattern mixture models for the analysis of multivariate heterogeneous longitudinal data. *The Annals of Applied Statistics,* **9**(2), 777–800. https://doi.org/10.1214/15-AOAS816

[3] Azzalini, A. (2013). *The Skew-Normal and Related Families*. Cambridge, Cambridge University Press.

[4] Basser, P.J. and Pajevic, S. (2003). A normal distribution for tensor-valued random variables: applications to diffusion tensor MRI. *IEEE Transactions on Medical Imaging,* **22**(7), 785–794. https://doi.org/10.1109/TMI.2003.815059

[5] Bi, X., Tang, X., Yuan, Y., Zhang, Y. and Qu, Annie. (2021). Tensors in Statistics. *Annual Review of Statistics and Its Application,* **8**(1) , 345–368. http://dx.doi.org/10.1146/annurev-statistics-042720-020816

[6] Drygas, H. (1984). Linear sufficiency and some applications in multilinear estimation. *Journal of Multivariate Analysis,* **16**(1), 71–84. https://doi.org/10.1016/0047-259X(85)90052-1

[7] Gallaugher, Michael P. B., Tait, Peter A. and McNicholas Paul D. (2021). Four skewed tensor distributions. Available from:arXiv:2106.08984. https://doi.org/10.48550/arXiv.2106.08984

[8] Ghannam, M. and Nkurunziza, S. (2022). The risk of tensor Stein-rules in elliptically contoured distributions. *Statistics,* **56**(2), 421–454. https://doi.org/10.1080/02331888.2022.2051508

[9] Kolda, T.G. and Bader, B.W. (2009). Tensor decompositions and applications. *SIAM Review,* **51**(3), 455–500. https://doi.org/10.1137/07070111X

[10] Krizhevsky, A. and Hinton, Geoffrey. (2009). *Learning Multiple Layers of Features from Tiny Images*. Toronto, ON, Canada. https://www.cs.utoronto.ca/ kriz/learning-features-2009-TR.pdf

[11] Lee, I., Sinha, D., Mai, Q., Zhang, X. and Bandyopadhyay, D. (2023). Bayesian regression analysis of skewed tensor responses. *Biometrics,* **79** , 1814–1825. https://doi.org/10.1111 /biom.13743

[12] Llosa-Vite, C. and Maitra, R. (2022). Elliptically-Contoured Tensor-variate Distributions with Application to Improved Image Learning. Available from:arXiv:2211.06940. https://doi.org/10.48550/arXiv.2211.06940

[13] Northcott, D.G. (1984). *Multilinear Algebra.* Cambridge, Cambridge University Press. https://doi.org/10.1017/cbo9780511565939

[14] Ohlson, M., Rauf Ahmad, M. and Von Rosen, D. (2013). The multilinear normal distribution: Introduction and some basic properties. *Journal of Multivariate Analysis,* **113** , 37–47. https://doi.org/10.1016/j.jmva.2011.05.015

[15] Srivastava, M., von Rosen, T. and von Rosen, D. (2008). Models with a Kronecker product covariance structure: Estimation and testing. *Mathematical Methods of Statistics,* **17**, 357–370. https://doi.org/10.3103/S1066530708040066

[16] Tait, Peter A., McNicholas, Paul D. and Obeid, J. (2020). Clustering Higher Order Data: An Application to Pediatric Multi-variable Longitudinal Data. Available from:arXiv:1907.08566. https://doi.org/10.48550/arXiv.1907.08566

S.A.A. Tajadod
Orcid number: 0009-0008-7447-5435
Department of Statistics
University of Birjand
Birjand, Iran
Email address: tajadod.aat@gmail.com

F. Yousefzadeh
Orcid number: 0000-0003-2648-5897
Department of Statistics
University of Birjand
Birjand, Iran
Email address: fyousefzadeh@birjand.ac.ir

R.B. Arellano-Valle
Orcid number: 0000-0002-5121-9702
Department of Estadística
Pontificia Universidad Católica de Chile
Santiago, Chile
Email address: reivalle@mat.uc.cl

S. Jomhoori
Orcid number: 0000-0003-2413-0371
Department of Statistics
University of Birjand
Birjand, Iran
Email address: sjomhoori@birjand.ac.ir