Mathematics Interdisciplinary Research **11** (3) (2026) 299 − 319

Original Scientific Paper

Some Characteristics of Tensor-Variate t and Skew-t Distributions

Seyed Ali Asghar Tajadod, Fatemeh Yousefzadeh⋆ and Raynaldo Boris Arellano-Valle

Abstract

In recent years, rapid technological advances have led to the emergence of increasingly complex data structures, among which tensor data play a central role. To address the challenges of analyzing such data, it is essential to extend statistical concepts and methodologies to the tensor domain. In this paper, we introduce the tensor-variate t and skew-t distributions and derive several of their statistical properties. Parameter estimation is carried out using the Expectation–Maximization (EM) algorithm, and the performance of the proposed estimators is assessed through extensive simulation studies. Finally, the effectiveness and practical applicability of the estimators are demonstrated using real world data.

Keywords: Kronecker-separable covariance, Multidimensional array, Tensor, EM algorithm, Image learning.

2020 Mathematics Subject Classification: 62H10; 70G45; 62H12.

How to cite this article
S. A. A. Tajadod, F. Yousefzadeh, R. B. Arellano-Valle, Some characteristics of tensor-variate t and skew-t distributions, *Math. Interdisc. Res.* **11** (3) (2026) 299-319.

1. Introduction

Over the past decade, the structure of collected data has become increasingly complex across many scientific domains. Among these complex forms, one particularly

⋆Corresponding author (E-mail: fyousefadeh@birjand.ac.ir)
Academic Editor: Mohammad Amini
Received 1 November 2025, Accepted 16 January 2026
DOI: 10.22052/MIR.2026.257883.1553

notable structure is tensor-valued or multi-dimensional data. Tensors generalize vectors and matrices to higher dimensions; formally, an Nth-order tensor is an element of the tensor product of vector spaces, each associated with its own dimension or coordinate system. It is important to distinguish this mathematical notion of tensors from the classical usage in physics and engineering (e.g., stress or strain tensors), which are more accurately described as tensor fields in mathematical terminology [1, 2]. Extensive work on elliptical models, tensor matrix variate distributions and robust estimation aligns perfectly with [3]. As high-dimensional and structured data becomes prevalent in applied research such as chemometrics, computer vision, image analysis, and psychometrics, there is a growing need to develop statistical theory and methodology that directly accommodates tensorial structures. While substantial advances have been made in extending classical statistical tools from random vectors (multivariate analysis) to random matrices, a further generalization to random tensors is essential for preserving intrinsic data structures and ensuring analytical accuracy. The importance of this extension has been emphasized in foundational works such as [4, 5]. These studies argue that preserving the tensor structure during statistical analysis is critical because flattening tensors into vectors or matrices often leads to the loss of multidimensional dependencies and valuable information. A compelling argument for this need is presented in [6], where the authors demonstrate how employing vector-based methods on inherently tensorial data can result in misleading or suboptimal inferences. Despite the widespread application of tensor data in numerous disciplines, the development of statistical distributions and inference methods tailored to tensors remains limited. While some research has addressed tensor normal distributions [5], elliptical distributions [7], and risk estimation under tensor-elliptical errors [8], these works predominantly assume symmetry in the underlying distributions -a condition often violated in real-world data. Moreover, outliers and skewed data structures are common, further highlighting the need for flexible and robust tensor-based models. Recent contributions such as [9] have introduced skewed tensor-variate distributions that generalize their matrix counterparts and are capable of capturing skewness and kurtosis. Similarly, [10] proposed a regression model for skewed tensor responses, marking an important step toward accommodating real-world data complexities. Nonetheless, the literature on skewed and heavy-tailed tensor distributions remains scarce. To address this gap, the present study investigates the Tensor-Variate t and skew-t (TVT, TVST) distributions, motivated by the well-established role of the classical and skew-t distributions in modeling heavy-tailed data and robust inference. This work aims to extend the applicability of the t and skew-t distributions to the tensor domain, providing a rigorous framework for modeling and inference.

The remainder of this paper is organized as follows. Section 2 introduces the TVT and TVST distributions, along with notational conventions, and derives key properties such as the density function, expectation, and characteristic function. 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 illustrated and using the results obtained in Section 3, we discuss the efficiency of the model presented in the paper.

2. Model

In this section, we introduce the basic concepts of tensors, the TVT and TVST distributions and establish some of their properties.

2.1 Background and preliminaries

A tensor is formally defined as a multidimensional array. Specifically, 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, also referred to as its modes or ways, is given by k. Here, p_j represents the marginal dimension of the jth mode ($j = 1, 2, \ldots, k$). The $(i_1, i_2, \ldots, i_k)$th entry of $\mathcal{X}$ is denoted by $x_{i_1,i_2,\ldots,i_k}$, where $i_j = 1, 2, \ldots, p_j$ for each $j = 1, 2, \ldots, k$. Moreover, by fixing certain indices of the tensor, one can obtain a tensor of lower order. In particular, a tensor can be unfolded or flattened into a matrix or a vector, denoted by $\boldsymbol{X}$ and $\boldsymbol{x}$, respectively. This process facilitates explicit computations and provides a convenient link between tensor-based and matrix/vector-based methods. For further details on these concepts, see [4] and [11].

The vectorial representation of a tensor provides a convenient way to simplify related inferences. Let $\text{vec}(\mathcal{X})$ denote the vectorization of a tensor

$$\mathcal{X} = (x_{i_1,i_2,\ldots,i_k}),$$

which arranges all tensor entries into a single column vector. Following the definition given in [4], this operation allows tensor-valued objects to be expressed in a linear algebraic form, thereby facilitating theoretical analysis and computational implementation.

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

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 $\boldsymbol{I_p} = \{i_1, \ldots, i_k : 1 \leq i_j \leq p_j, 1 \leq \text{j} \leq k\}$.

In [5], the authors focused on the estimation of a Kronecker-structured covariance matrix of order three ($k = 3$), known as the double separable covariance matrix. This work generalizes the earlier results of [12] for the 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. Whenever no ambiguity arises, the dimension of the basis vectors will be omitted and we shall simply write $\boldsymbol{e}^{p_1}_{i_1}$ as $\boldsymbol{e}_{i_1}$, and $\boldsymbol{e}^{p}_{i_1:i_k}$ as $\boldsymbol{e}_{i_1:i_k}$, etc.

Let $\mathcal{T}^p$ denote the space of all p dimensional vectors $\boldsymbol{x} = vec(\mathcal{X})$, where $\mathcal{X}$ is a tensor of order k, i.e., $\mathcal{T}^p = \{\boldsymbol{x} : \boldsymbol{x} = \sum_{\boldsymbol{I_p}} x_{i_1 i_2 \ldots i_k} \boldsymbol{e}^{\boldsymbol{p}}_{1:k}\}$. It is worth noting that this tensor space is described using vectors. Alternatively, tensor spaces may be defined in terms of matrices, as presented in the following definition.

Definition 2.1. Let

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

The derivation of the TVT and TVST distributions relies on the Tensor-variate normal (TVN) distribution. The probability density function (pdf) of the TVN distribution is given as follows.

Definition 2.2. If $\mathcal{X}$ is a random order-k tensor, with dimensional lengths $p_1 = p_2 = \ldots = p_k$, then it follows that the density function of a TVN distribution, $TVT\,(\mathcal{M}, \boldsymbol{\Sigma}_1, \ldots, \boldsymbol{\Sigma}_k)$ can be written as

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

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

2.2 Characterizing the TVT and TVST distributions

In this subsection, we introduce the TVT and TVST distributions and establish some of their fundamental properties. In the multivariate and matrix variate settings, the normal variance mean mixture model is a computationally efficient way to introduce t and skew-t distributions. In the multivariate case, this formulation assumes that the random vector $\boldsymbol{x}$ can be expressed in the form $\boldsymbol{x} = \boldsymbol{\mu} + W\boldsymbol{\delta} + \sqrt{W}\boldsymbol{V}$; where $\boldsymbol{\mu}$ is a location vector, $\boldsymbol{\delta}$ is a skewness vector, $\boldsymbol{V} \sim N(\boldsymbol{0}; \boldsymbol{\Sigma})$, and $W > 0$ is a positive random variable. This approach was extended to the matrix variate case by [13] and this is now generated to the tensor case to derive t and skew-t tensor-variate distributions. We now provide a derivation of a TVT and TVST distributions. We say that an k order random tensor $\mathcal{X}$ has a TVST distribution, $TVST(\mathcal{M}, \mathcal{S}, \boldsymbol{\Sigma}_{1:k}, \nu)$ if $\mathcal{X}$ admits the representation $\mathcal{X} = \mathcal{M} + W\mathcal{S} + \sqrt{W}\mathcal{V}$ and reduces to the TVT distribution when $\mathcal{S} \in \mathcal{T}^p_{\otimes} = \boldsymbol{0}$ where $\mathcal{M} \in \mathcal{T}^p_{\otimes}$ is k dimensional order tensors, $\mathcal{V} \sim N(\mathcal{M}, \boldsymbol{\Sigma}_{1:k})$ and W is Inv-Gamma$(\nu/2; \nu/2)$ where Inv-Gamma(.) denotes the inverse-gamma distribution. Similar to its multivariate [14] and matrix variate [13] counterparts, $\boldsymbol{\mu} = vec(\mathcal{M})$ acts as a location tensor, $\boldsymbol{\delta} = vec(\mathcal{S})$ is a skewness parameter, $\boldsymbol{\Sigma}_1, ..., \boldsymbol{\Sigma}_k$ are scale matrices, and ν is the degrees of freedom. It then follows that $\mathcal{X}|W = w \sim N(\boldsymbol{\mu} + w\boldsymbol{\delta}, w\boldsymbol{\Sigma})$ and thus the joint density of $\mathcal{X}$ and W is

$$\begin{aligned} f_{\mathcal{X}}(\boldsymbol{x}; w) &= f_{(\mathcal{X}|W=w)}(\boldsymbol{x}|w) f_w(w) \\ &= \frac{(\nu/2)^{(\nu/2)} w^{-\frac{\nu+p^*}{2}-1}}{(2\pi)^{(-p^*/2)} \left(\prod_{i=1}^k \left|\boldsymbol{\Sigma}_i\right|^{-p^*/(2p_i)}\right) \Gamma(\nu/2)} \\ &\times \; exp\Big\{ -\frac{1}{2w}\left((\boldsymbol{x} - \boldsymbol{\mu} - w\boldsymbol{\delta})' \boldsymbol{\Sigma}^{-1}_{1:k} (\boldsymbol{x} - \boldsymbol{\mu} - w\boldsymbol{\delta}) + \nu\right) \Big\}. \end{aligned}$$

Therefore, the marginal density of $\mathcal{X}$ is

$$\begin{aligned} f_{\mathcal{X}}(\boldsymbol{x}) &= \int_0^\infty f_{(\mathcal{X}|W=w)}(\boldsymbol{x}|w) f_w(w) dw \\ &= \frac{\nu^{\frac{\nu}{2}}}{(2\pi)^{\frac{p^*}{2}} \prod_{i=1}^k |\Sigma_i|^{p^*/2p_i} \Gamma(\nu/2)} \exp\left\{ -\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}_{1:k} \boldsymbol{\delta} \right\} \\ &\times \int_0^\infty w^{-\frac{\nu+p^*}{2}-1} \exp\left\{ -\frac{1}{2}\left[wA(\boldsymbol{\delta}, \boldsymbol{\Sigma}^{-1}_{1:k}) + \frac{B(\boldsymbol{x}; \mu, \boldsymbol{\Sigma}^{-1}_{1:k}) + \nu}{w} \right] \right\} dw, \end{aligned}$$

where $A(\boldsymbol{\delta}, \boldsymbol{\Sigma}^{-1}_{1:k}) = \boldsymbol{\delta}^T \boldsymbol{\Sigma}^{-1}_{1:k} \boldsymbol{\delta}$ and $B(\boldsymbol{x}; \mu, \boldsymbol{\Sigma}^{-1}_{1:k}) = (\boldsymbol{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}_{1:k} (\boldsymbol{x} - \boldsymbol{\mu})$. Now, we can derive the density of TVST and TVT distributions.

Theorem 2.3. *The density function of the TVST and TVT distribution are given as*

$$f_{TVST}(\boldsymbol{x}) = \frac{2\nu^{\frac{\nu}{2}} \exp\left\{-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}_{1:k} \boldsymbol{\delta}\right\}}{(2\pi)^{\frac{p^*}{2}} \prod_{i=1}^k |\Sigma_i|^{p^*/2p_i} \Gamma(\nu/2)} \left(\frac{B(\boldsymbol{x}; \mu, \boldsymbol{\Sigma}^{-1}_{1:k}) + \nu}{A(\boldsymbol{\delta}, \boldsymbol{\Sigma}^{-1}_{1:k})} \right)^{-\frac{\nu+p*}{2}}$$

$$\begin{aligned} & \times \quad K_{\frac{\nu+p^*}{2}}\left(\sqrt{A(\boldsymbol{\delta},\boldsymbol{\Sigma}_{1:k}^{-1})(B(\boldsymbol{x};\mu,\boldsymbol{\Sigma}_{1:k}^{-1})+\nu)}\right), \\ f_{TVT}(\boldsymbol{x}) &= \frac{2^{\frac{\nu+p^*}{2}}(\nu/2)^{(\nu/2)}}{(2\pi)^{(-p^*/2)}\left(\prod_{i=1}^{k}\left|\boldsymbol{\Sigma}_i\right|^{-p^*/(2p_i)}\right)\Gamma(\nu/2)}\frac{\Gamma\left(\frac{\nu+p^*}{2}\right)}{\left(B(\boldsymbol{x};\mu,\boldsymbol{\Sigma}_{1:k}^{-1})+\nu\right)^{\frac{\nu+p^*}{2}}}. \end{aligned}$$

where $\nu > 0$.

To simplify the notation, these distributions will hereafter be denoted by

$$TVST_p(\boldsymbol{\mu},\boldsymbol{\delta},\boldsymbol{\Sigma}_{1:k},\boldsymbol{\nu}), T_p(\boldsymbol{\mu},\boldsymbol{\Sigma}_{1:k},\boldsymbol{\nu}).$$

Theorem 2.4. *Characteristic functions of TVST and TVT distributions are given as*

$$\varphi_{TVST}(\boldsymbol{t}) = 2\exp\left(i\boldsymbol{t}'\boldsymbol{\mu}\right)\frac{(\nu/2)^{\nu/4}K_{-\nu/2}\left(\sqrt{(\boldsymbol{t}'\boldsymbol{\Sigma}_{1:k}\boldsymbol{t}-2i\boldsymbol{t}'\boldsymbol{\delta})\,\nu}\right)}{\Gamma(\nu/2)\left(\frac{\boldsymbol{t}'\boldsymbol{\Sigma}_{1:k}\boldsymbol{t}}{2}-i\boldsymbol{t}'\boldsymbol{\delta}\right)^{-\nu/4}}, \tag{5}$$

$$\varphi_{TVT}(\boldsymbol{t}) = 2\exp\left(i\boldsymbol{t}'\boldsymbol{\mu}\right)\frac{K_{-\nu/2}\left(\sqrt{(\boldsymbol{t}'\boldsymbol{\Sigma}_{1:k}\boldsymbol{t})\,\nu}\right)}{\Gamma(\nu/2)\left(\frac{\boldsymbol{t}'\boldsymbol{\Sigma}_{1:k}\boldsymbol{t}}{\nu}\right)^{-\nu/4}}, \tag{6}$$

where $K_\lambda(x) = \frac{1}{2}\int_0^\infty y^{\lambda-1}exp\left\{\frac{x}{2}\left(y+\frac{1}{y}\right)\right\}dy$ *is the modified Bessel function of the third kind with index* λ.

Proof. Without loss of generality, suppose that $\boldsymbol{x} \sim TVST_{\boldsymbol{p}}(\boldsymbol{\mu},\boldsymbol{\delta},\boldsymbol{\Sigma}_{1:k},\boldsymbol{\nu})$. So,

$$\begin{aligned} \varphi_{\mathcal{X}}(\boldsymbol{t}) &= E\left\{E\left(\exp\left(i\boldsymbol{t}'\boldsymbol{x}\right)\mid W=w\right)\right\} \\ &= \exp\left(i\boldsymbol{t}'\boldsymbol{\mu}\right)\int_0^\infty \exp\left\{i\boldsymbol{t}'\boldsymbol{\delta}w-\frac{1}{2}\left(\boldsymbol{t}'\boldsymbol{\Sigma}_{1:k}\boldsymbol{t}\right)w\right\}f_W(w)\,dw \\ &= 2\exp\left(i\boldsymbol{t}'\boldsymbol{\mu}\right)\frac{(\nu/2)^{\nu/4}\,K_{-\nu/2}\left(\sqrt{(\boldsymbol{t}'\boldsymbol{\Sigma}_{1:k}\boldsymbol{t}-2i\boldsymbol{t}'\boldsymbol{\delta})\,\nu}\right)}{\Gamma(\nu/2)\left(\frac{\boldsymbol{t}'\boldsymbol{\Sigma}_{1:k}\boldsymbol{t}}{2}-i\boldsymbol{t}'\boldsymbol{\delta}\right)^{-\nu/4}}. \end{aligned}$$

Therefore, the characteristic function of the TVT distribution can be readily derived. □

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,

$$\frac{d\boldsymbol{Y}}{d\boldsymbol{X}}=\frac{d\boldsymbol{y}}{d\boldsymbol{x}}=\sum_{i_1:i_{k_1}}\sum_{j_1:j_{k_2}}\sum_{m_1:m_{k_3}}\sum_{n_1:n_{k_4}}\frac{\partial y_{i_1:i_{k_1}j_1:j_{k_2}}}{\partial x_{m_1:m_{k_3}n_1:n_{k_4}}}\left(\mathbf{e}^{\mathbf{s}(1:k_4)}_{n_1:n_{k_4}}\otimes\mathbf{e}^{\mathbf{r}(1:k_3)}_{m_1:m_{k_3}}\right)\left(\mathbf{e}^{\mathbf{q}(1:k_2)}_{j_1:j_{k_2}}\otimes\mathbf{e}^{\mathbf{p}(1:k_1)}_{i_1:i_{k_1}}\right)'. \tag{7}$$

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

$$\frac{d^k\boldsymbol{Y}}{d\boldsymbol{X}^k}=\frac{d}{d\boldsymbol{X}}\frac{d^{k-1}\boldsymbol{Y}}{d\boldsymbol{X}^{k-1}}. \tag{8}$$

Theorem 2.5. *The mean of random tensor of TVST and TVT distributions are given by*

$$\begin{aligned}E_{TVST}(\boldsymbol{x}) &= \boldsymbol{\mu}+\frac{\nu}{\nu-2}\boldsymbol{\delta},\quad \nu>2,\\ E_{TVT}(\boldsymbol{x}) &= \boldsymbol{\mu}.\end{aligned}$$

Proof. Applying (7) to $\varphi(\mathbf{t})$, and evaluating the derivatives at $\mathbf{t}=\mathbf{0}$, the desired result is obtained. □

If we aim to express a multivariate density or distribution function using the multivariate Student's t distribution, certain expansions arise in which multivariate densities, distribution functions, cumulants, and Hermite polynomials naturally appear. The derivation of explicit expressions for low-order Hermite polynomials constitutes the concluding focus of this subsection.

Definition 2.6. 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 (8), and $f_{\mathcal{X}}(\boldsymbol{x})$ is the pdf (5).

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

Theorem 2.7. *Let $\boldsymbol{x}\sim TVT_{\boldsymbol{p}}(\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\delta})$, 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:*

$$\begin{aligned}(i)\quad H_0(\boldsymbol{x},\boldsymbol{\mu},\boldsymbol{\Sigma}_{1:k}) &= 1,\\ (ii)\quad \boldsymbol{H}_1(\boldsymbol{x},\boldsymbol{\mu},\boldsymbol{\Sigma}_{1:k}) &= \boldsymbol{\Sigma}^{-1}_{1:k}(\boldsymbol{x}-\boldsymbol{\mu})\frac{\nu+p*}{\zeta(\boldsymbol{x}-\boldsymbol{\mu})+\nu},\\ (iii)\quad \boldsymbol{H}_2(\boldsymbol{x},\boldsymbol{\mu},\boldsymbol{\Sigma}_{1:k}) &= \boldsymbol{\Sigma}^{-1}_{1:k}(\boldsymbol{x}-\boldsymbol{\mu})(\boldsymbol{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}_{1:k}\frac{(\nu+p*+2)(\nu+p*)}{(\zeta(\boldsymbol{x}-\boldsymbol{\mu})+\nu)^2}-\boldsymbol{\Sigma}^{-1}_{1:k}\frac{\nu+p*}{\zeta(\boldsymbol{x}-\boldsymbol{\mu})+\nu},\end{aligned}$$

where $\zeta(\boldsymbol{x})=\boldsymbol{x}'\boldsymbol{\Sigma}^{-1}_{1:k}\boldsymbol{x}$.

Proof. For $k = 0$, the relation in (Definition 2.6) turns into a trivial identity and we obtain that

$$H_0\left(\boldsymbol{x}, \boldsymbol{\mu}, \boldsymbol{\Sigma}\right) = 1,$$

For $k = 1$ the first order derivative equals

$$\begin{aligned}
\frac{\mathrm{d}f_{\mathcal{X}}(\boldsymbol{x})}{\mathrm{d}\boldsymbol{x}} &= \int \frac{\mathrm{d}f_{\mathcal{X}|W}(\boldsymbol{x}|w)f_W(w)}{\mathrm{d}\boldsymbol{x}}dw \\
&= \boldsymbol{\Sigma}_{1:k}^{-1}\left(\boldsymbol{x}-\boldsymbol{\mu}\right)E\left(W^{-1}|\mathcal{X}\right)f_{\mathcal{X}}(\boldsymbol{x}) \\
&= \boldsymbol{\Sigma}_{1:k}^{-1}\left(\boldsymbol{x}-\boldsymbol{\mu}\right)\frac{\nu+p*}{\zeta(\boldsymbol{x}-\boldsymbol{\mu})+\nu}f_{\mathcal{X}}(\boldsymbol{x}),
\end{aligned}$$

where $W|\mathcal{X}$ is Inv-Gamma $((\nu + p*)/2, (\zeta(\boldsymbol{x} - \boldsymbol{\mu}) + \nu)/2)$. Thus, (Definition 2.6) yields (ii). To prove (iii) we have to find the second order derivative, i.e.

$$\begin{aligned}
\frac{\mathrm{d}^2 f_{\mathcal{X}}(\boldsymbol{x})}{\mathrm{d}\boldsymbol{x}^2} &= \int \frac{\mathrm{d}^2 f_{\mathcal{X}|W}(\boldsymbol{x}|w)f_W(w)}{\mathrm{d}\boldsymbol{x}^2}dw \\
&= \boldsymbol{\Sigma}_{1:k}^{-1}\left(\boldsymbol{x}-\boldsymbol{\mu}\right)\left(\boldsymbol{x}-\boldsymbol{\mu}\right)'\boldsymbol{\Sigma}_{1:k}^{-1}E\left(W^{-2}|\mathcal{X}\right)f_{\mathcal{X}}(\boldsymbol{x}) - \boldsymbol{\Sigma}_{1:k}^{-1}E\left(W^{-1}|\mathcal{X}\right)f_{\mathcal{X}}(\boldsymbol{x}) \\
&= \boldsymbol{\Sigma}_{1:k}^{-1}\left(\boldsymbol{x}-\boldsymbol{\mu}\right)\left(\boldsymbol{x}-\boldsymbol{\mu}\right)'\boldsymbol{\Sigma}_{1:k}^{-1}\frac{(\nu+p*+2)(\nu+p*)}{(\zeta(\boldsymbol{x}-\boldsymbol{\mu})+\nu)^2}f_{\mathcal{X}}(\boldsymbol{x}) - \boldsymbol{\Sigma}_{1:k}^{-1}\frac{\nu+p*}{\zeta(\boldsymbol{x}-\boldsymbol{\mu})+\nu}f_{\mathcal{X}}(\boldsymbol{x}).
\end{aligned}$$

□

Hermite polynomials have extensive applications across mathematics, statistics, physics, and numerical analysis. In probability theory, they play a central role in Edgeworth and Gram–Charlier expansions, where they refine normal approximations and model departures from normality. They are fundamental in Gaussian quadrature for numerical integration and arise naturally in quantum mechanics as representations of wave functions for harmonic oscillators. In signal processing, Hermite polynomials are applied in noise reduction and signal approximation, while in machine learning, they support feature engineering and the modeling of nonlinear relationships. Furthermore, in financial mathematics, they are utilized in option pricing and risk analysis. Collectively, these diverse applications highlight their indispensable role in both theoretical developments and practical problem-solving.

In the following, we derive the marginal 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

$$\boldsymbol{x}_{r_1} = vec(\mathcal{X}_{r_1}) = \boldsymbol{M}_{r_1}\boldsymbol{x}, \tag{9}$$

$$\boldsymbol{x}_{r_2} = vec(\mathcal{X}_{r_2}) = \boldsymbol{M}_{r_2}\boldsymbol{x}, \tag{10}$$

where

$$\boldsymbol{M}_{r_1} = \sum_{I_{r_1}} \boldsymbol{e}_{i_1:i_{r-1}}^{\boldsymbol{p}(1:r-1)} \otimes \boldsymbol{e}_{i_1}^{m} \otimes \boldsymbol{e}_{i_{r+1}:k}^{\boldsymbol{p}(r+1:k)} \left(\boldsymbol{e}_{i_1:i_k}^{\boldsymbol{p}}\right)', \tag{11}$$

$$\boldsymbol{M}_{r_2} = \sum_{I_{r_1}} \boldsymbol{e}_{i_1:i_{r-1}}^{\boldsymbol{p}(1:r-1)} \otimes \boldsymbol{e}_{i_r-m}^{p_r-m} \otimes \boldsymbol{e}_{i_{r+1}:k}^{\boldsymbol{p}(r+1:k)} \left(\boldsymbol{e}_{i_1:i_k}^{\boldsymbol{p}}\right)', \tag{12}$$

with their respective index sets

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

Theorem 2.8. *Let $\boldsymbol{x}_{r_1}$, be as defined in (9), then*

$$\begin{aligned} \boldsymbol{x}_{r_1} &\sim TVST_{\boldsymbol{r}_1}(\boldsymbol{M}_{r_1}\boldsymbol{\mu}, \boldsymbol{M}_{r_1}\boldsymbol{\delta}, \boldsymbol{M}_{r_1}\boldsymbol{\Sigma}_{1:k}\boldsymbol{M}'_{r_1}, \nu), \\ \boldsymbol{x}_{r_1} &\sim TVT_{\boldsymbol{r}_1}(\boldsymbol{M}_{r_1}\boldsymbol{\mu}, \boldsymbol{M}_{r_1}\boldsymbol{\Sigma}_{1:k}\boldsymbol{M}'_{r_1}, \nu). \end{aligned}$$

Proof. We start from the following representation:

$$\boldsymbol{x}_{r_1} = \boldsymbol{M}_{r_1}\boldsymbol{\mu} + W\boldsymbol{M}_{r_1}\boldsymbol{\delta} + \sqrt{W}\,\boldsymbol{M}_{r_1}\boldsymbol{V}.$$

According to [5],

$$\boldsymbol{V}_{r_1} = \boldsymbol{M}_{r_1}\boldsymbol{V},$$

follows a tensor-variate normal distribution

$$\boldsymbol{V}_{r_1} \sim N_{r_1}\left(\boldsymbol{0},\, \boldsymbol{M}_{r_1}\boldsymbol{\Sigma}_{1:k}\boldsymbol{M}_{r_1}^{\top}\right).$$

By invoking Theorem 2.3, the desired result immediately follows. □

At the end of this section, we present the following theorem, which allows for the examination of several noteworthy special cases through appropriate choices of $\boldsymbol{A}$.

Theorem 2.9. *Let*

$$\boldsymbol{x} \sim TVST_{\boldsymbol{p}}(\boldsymbol{\mu}, \boldsymbol{\delta}, \boldsymbol{\Sigma}_{1:k}, \nu),$$

where $\boldsymbol{x} \in \mathcal{T}^{\boldsymbol{p}}$ and $\boldsymbol{\Sigma}_{1:k} = \bigotimes_{j=1}^{k} \boldsymbol{\Sigma}_j$ denotes the Kronecker-structured scale matrix. Let $\boldsymbol{A} \in \mathcal{T}_{\otimes}^{\boldsymbol{q}\times\boldsymbol{p}}$ be a nonsingular (invertible) linear operator acting on the tensor space (equivalently, an invertible $q \times p$ matrix on the vectorized representation). Then the linear transformation $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$ has the following distributions:

$$\boldsymbol{A}\boldsymbol{x} \sim TVST_{\boldsymbol{q}}\big(\boldsymbol{A}\boldsymbol{\mu},\, \boldsymbol{A}\boldsymbol{\delta},\, \boldsymbol{A}\boldsymbol{\Sigma}_{1:k}\boldsymbol{A}',\, \nu\big),$$

and in the special case $\boldsymbol{\delta} = \boldsymbol{0}$ $\boldsymbol{A}\boldsymbol{x} \sim TVT_{\boldsymbol{q}}\big(\boldsymbol{A}\boldsymbol{\mu},\, \boldsymbol{A}\boldsymbol{\Sigma}_{1:k}\boldsymbol{A}',\, \nu\big)$. In other words, under an invertible linear mapping, the TVST family is closed and the parameters transform by

$$\boldsymbol{\mu}_A = \boldsymbol{A}\boldsymbol{\mu}, \qquad \boldsymbol{\delta}_A = \boldsymbol{A}\boldsymbol{\delta}, \qquad \boldsymbol{\Sigma}_A = \boldsymbol{A}\boldsymbol{\Sigma}_{1:k}\boldsymbol{A}'.$$

Proof. We use the scale–mean mixture representation of the TVST distribution. By assumption, there exists a positive scalar mixing variable W and applying the linear mapping $\boldsymbol{A}$ to both sides

$$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{A}\boldsymbol{\mu} + W(\boldsymbol{A}\boldsymbol{\delta}) + \sqrt{W}\,(\boldsymbol{A}\boldsymbol{V}).$$

Since $\boldsymbol{V} \sim N_p(\boldsymbol{0}, \boldsymbol{\Sigma}_{1:k})$, linearity of the multivariate normal implies

$$\boldsymbol{A}\boldsymbol{V} \sim N_q\big(\boldsymbol{0},\ \boldsymbol{A}\boldsymbol{\Sigma}_{1:k}\boldsymbol{A}'\big).$$

Therefore, conditional on $W = w$,

$$\boldsymbol{A}\boldsymbol{x} \mid W = w \sim N_q\big(\boldsymbol{A}\boldsymbol{\mu} + w\boldsymbol{A}\boldsymbol{\delta},\ w\,\boldsymbol{A}\boldsymbol{\Sigma}_{1:k}\boldsymbol{A}'\big).$$

Marginalizing over the same mixing law W (with unchanged distribution) yields exactly the scale–mean mixture form that defines a q-dimensional tensor-variate skew-t with location $\boldsymbol{A}\boldsymbol{\mu}$, skewness $\boldsymbol{A}\boldsymbol{\delta}$, scale $\boldsymbol{A}\boldsymbol{\Sigma}_{1:k}\boldsymbol{A}'$ and degrees of freedom ν. This proves the first claim. For the second statement, set $\boldsymbol{\delta} = \boldsymbol{0}$. Then the mixture reduces to a (symmetric) scale mixture of normals that defines the tensor-variate t distribution. Applying the same linear mapping shows closure of the TVT family under nonsingular linear transformations with transformed parameters $\boldsymbol{A}\boldsymbol{\mu}$ and $\boldsymbol{A}\boldsymbol{\Sigma}_{1:k}\boldsymbol{A}'$ (and the same degrees of freedom ν). This completes the proof. □

3. Parameter estimation via EM algorithm

Assume that there are n independent tensor observations $\boldsymbol{\mathcal{X}} = (\mathcal{X}_1, \mathcal{X}_2, \ldots, \mathcal{X}_n)$ drawn from the density function $f_{\mathcal{X}}(\boldsymbol{x})$, with their corresponding vector representations $\boldsymbol{x} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n)$. Let $\boldsymbol{\theta}$ denote the parameter vector comprising $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}_1, \ldots, \boldsymbol{\Sigma}_k$, and ν. In this section, we address the maximum likelihood estimation (MLE) of $\boldsymbol{\theta}$ based on the observed samples $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$, and we derive Expectation Maximization (EM) algorithm for its computation. We treat the observed data as incomplete by introducing latent variables w_i associated with each observation. Accordingly, the complete log-likelihood function for the TVST and TVT distributions (the latter obtained when $\boldsymbol{\delta} = \boldsymbol{0}$) is given by

$$\begin{aligned} l_c(\boldsymbol{\theta}) = \frac{n\nu}{2}\log\Big(\frac{\nu}{2}\Big) - \sum_{i=1}^{n}\left(\frac{\nu + p^*}{2} + 1\right)\log w_i - n\log\Gamma(\nu/2) - n\sum_{i=1}^{k}\frac{p^*}{2p_i}\log|\boldsymbol{\Sigma}_i| \\ - \sum_{i=1}^{n}\frac{1}{2w_i}\left[(\boldsymbol{x}_i - \boldsymbol{\mu} - w_i\boldsymbol{\delta})'\boldsymbol{\Sigma}_{1:k}^{-1}(\boldsymbol{x}_i - \boldsymbol{\mu} - w_i\boldsymbol{\delta}) + \nu\right], \end{aligned} \tag{13}$$

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 w_i given the observed data $\boldsymbol{x}_i$ and the current estimate of the parameter $\hat{\boldsymbol{\theta}}^{(m)}$ at the iteration mth is

$$\begin{aligned}
Q(\boldsymbol{\theta} \mid \hat{\boldsymbol{\theta}}^{(m)}) &= E\Big(l_c(\boldsymbol{\theta}) \mid \boldsymbol{x}, \hat{\boldsymbol{\theta}}^{(m)}\Big) \\
&= \frac{n\nu}{2}\log\Big(\frac{\nu}{2}\Big) - \sum_{i=1}^{n}\left(\frac{\nu+p^*}{2}+1\right) E\Big(\log W_i \mid \boldsymbol{x}_i, \hat{\boldsymbol{\theta}}^{(m)}\Big) \\
&\quad - n\log\Gamma(\nu/2) - n\sum_{i=1}^{k}\frac{p^*}{2p_i}\log|\boldsymbol{\Sigma}_i| \\
&\quad - \sum_{i=1}^{n} E\left(\frac{1}{2W_i} \mid \boldsymbol{x}_i, \hat{\boldsymbol{\theta}}^{(m)}\right)\left[(\boldsymbol{x}_i-\boldsymbol{\mu})'\boldsymbol{\Sigma}_{1:k}^{-1}(\boldsymbol{x}_i-\boldsymbol{\mu})+\nu\right] \\
&\quad + \frac{1}{2}\sum_{i=1}^{n}(\boldsymbol{x}_i-\boldsymbol{\mu})'\boldsymbol{\Sigma}_{1:k}^{-1}\boldsymbol{\delta} + \frac{1}{2}\sum_{i=1}^{n}\boldsymbol{\delta}'\boldsymbol{\Sigma}_{1:k}^{-1}(\boldsymbol{x}_i-\boldsymbol{\mu}) \\
&\quad - \frac{1}{2}\sum_{i=1}^{n} E\Big(W_i \mid \boldsymbol{x}_i, \hat{\boldsymbol{\theta}}^{(m)}\Big)\,\boldsymbol{\delta}'\boldsymbol{\Sigma}_{1:k}^{-1}\boldsymbol{\delta}.
\end{aligned}$$

By definition, the conditional distribution of W given $\boldsymbol{X}=\boldsymbol{x}$ is a Generalized inverse Gaussian (GIG) distribution with density function as

$$\begin{aligned}
f_W(w \mid \boldsymbol{x}) &= \frac{\left(\frac{A\left(\boldsymbol{\delta},\boldsymbol{\Sigma}_{1:k}^{-1}\right)}{B\left(\boldsymbol{x};\boldsymbol{\mu},\boldsymbol{\Sigma}_{1:k}^{-1}\right)}\right)^{-\frac{\nu+p^*}{4}}}{2\,K_{-\frac{\nu+p^*}{2}}\left(\sqrt{A\left(\boldsymbol{\delta},\boldsymbol{\Sigma}_{1:k}^{-1}\right)B\left(\boldsymbol{x};\boldsymbol{\mu},\boldsymbol{\Sigma}_{1:k}^{-1}\right)+\nu}\right)} w^{-\frac{\nu+p^*}{2}} \\
&\quad \times \exp\left\{-\frac{1}{2}\left[A\left(\boldsymbol{\delta},\boldsymbol{\Sigma}_{1:k}^{-1}\right)w + \frac{B\left(\boldsymbol{x};\boldsymbol{\mu},\boldsymbol{\Sigma}_{1:k}^{-1}\right)+\nu}{w}\right]\right\},
\end{aligned}$$

where $A(\boldsymbol{\delta},\boldsymbol{\Sigma}_{1:k}^{-1}) = \boldsymbol{\delta}^T\boldsymbol{\Sigma}_{1:k}^{-1}\boldsymbol{\delta}$ and $B(\boldsymbol{x};\mu,\boldsymbol{\Sigma}_{1:k}^{-1}) = (\boldsymbol{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}_{1:k}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})$. So, we can write [9]:

$$a(\boldsymbol{\theta}|\boldsymbol{x}) = E(W_i|\boldsymbol{x},\boldsymbol{\theta}) = \frac{\sqrt{B\left(\boldsymbol{x};\boldsymbol{\mu},\boldsymbol{\Sigma}_{1:k}^{-1}\right)+\nu}}{\sqrt{A\left(\boldsymbol{\delta},\boldsymbol{\Sigma}_{1:k}^{-1}\right)}} \frac{K_{-\frac{\nu+p^*}{2}+1}\left(\sqrt{A\left(\boldsymbol{\delta},\boldsymbol{\Sigma}_{1:k}^{-1}\right)\left(B\left(\boldsymbol{x};\boldsymbol{\mu},\boldsymbol{\Sigma}_{1:k}^{-1}\right)+\nu\right)}\right)}{K_{-\frac{\nu+p^*}{2}}\left(\sqrt{A\left(\boldsymbol{\delta},\boldsymbol{\Sigma}_{1:k}^{-1}\right)\left(B\left(\boldsymbol{x};\boldsymbol{\mu},\boldsymbol{\Sigma}_{1:k}^{-1}\right)+\nu\right)}\right)},$$

$$b(\boldsymbol{\theta}|\boldsymbol{x}) = E(\frac{1}{W}|\boldsymbol{x},\boldsymbol{\theta}) = \frac{\sqrt{A\left(\boldsymbol{\delta},\boldsymbol{\Sigma}_{1:k}^{-1}\right)}K_{-\frac{\nu+p^*}{2}+1}\left(\sqrt{A\left(\boldsymbol{\delta},\boldsymbol{\Sigma}_{1:k}^{-1}\right)\left(B\left(\boldsymbol{x};\boldsymbol{\mu},\boldsymbol{\Sigma}_{1:k}^{-1}\right)+\nu\right)}\right)}{\sqrt{B\left(\boldsymbol{x};\boldsymbol{\mu},\boldsymbol{\Sigma}_{1:k}^{-1}\right)+\nu}K_{-\frac{\nu+p^*}{2}}\left(\sqrt{A\left(\boldsymbol{\delta},\boldsymbol{\Sigma}_{1:k}^{-1}\right)\left(B\left(\boldsymbol{x};\boldsymbol{\mu},\boldsymbol{\Sigma}_{1:k}^{-1}\right)+\nu\right)}\right)}$$

$$+\frac{(\nu+p^*)}{B\left(\boldsymbol{x};\boldsymbol{\mu},\boldsymbol{\Sigma}_{1:k}^{-1}\right)+\nu},$$

$$c(\boldsymbol{\theta}|\boldsymbol{x}) = E(\log(W)|\boldsymbol{x},\boldsymbol{\theta}) = \log\frac{\sqrt{B\left(\boldsymbol{x};\boldsymbol{\mu},\boldsymbol{\Sigma}_{1:k}^{-1}\right)+\nu}}{\sqrt{A\left(\boldsymbol{\delta},\boldsymbol{\Sigma}_{1:k}^{-1}\right)}}$$
$$+\frac{\partial}{\partial p}\log K_p\left(\sqrt{A\left(\boldsymbol{\delta},\boldsymbol{\Sigma}_{1:k}^{-1}\right)\left(B\left(\boldsymbol{x};\boldsymbol{\mu},\boldsymbol{\Sigma}_{1:k}^{-1}\right)+\nu\right)}\right).$$

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

  1. Update $\boldsymbol{\mu}$

$$\hat{\boldsymbol{\mu}}_{TVST}^{(m+1)} = \frac{\sum_{i=1}^n \boldsymbol{x}_i\left(\overline{a}(\hat{\boldsymbol{\theta}}^{(m)}|\boldsymbol{x})b(\hat{\boldsymbol{\theta}}^{(m)}|\boldsymbol{x}_i)-1\right)}{\overline{a}(\hat{\boldsymbol{\theta}}^{(m)}|\boldsymbol{x})\sum_{i=1}^n b(\hat{\boldsymbol{\theta}}^{(m)}|\boldsymbol{x}_i)-n},$$
$$\hat{\boldsymbol{\mu}}_{TVT}^{(m+1)} = \frac{\sum_{i=1}^n b(\hat{\boldsymbol{\theta}}^{(m)}|\boldsymbol{x}_i)\boldsymbol{x}_i}{\sum_{i=1}^n b(\hat{\boldsymbol{\theta}}^{(m)}|\boldsymbol{x}_i)}.$$

  2. Update $\boldsymbol{\delta}$

$$\hat{\boldsymbol{\delta}}_{TVST}^{(m+1)} = \frac{\sum_{i=1}^n \boldsymbol{x}_i\left(\overline{b}(\hat{\boldsymbol{\theta}}^{(m)}|\boldsymbol{x})-b(\hat{\boldsymbol{\theta}}^{(m)}|\boldsymbol{x}_i)\right)}{\overline{b}(\hat{\boldsymbol{\theta}}^{(m)}|\boldsymbol{x})\sum_{i=1}^n a(\hat{\boldsymbol{\theta}}^{(m)}|\boldsymbol{x}_i)-n}.$$

  3. Update $\boldsymbol{\Sigma}_j, j=1,...,k$,

$$\begin{aligned}\hat{\boldsymbol{\Sigma}}_{j(TVST)}^{(m+1)} &= \frac{p_j}{np^*}\sum_{i=1}^n b(\hat{\boldsymbol{\theta}}^{(m)}|\boldsymbol{x}_i)\left(\boldsymbol{x}_{i(j)}-\hat{\boldsymbol{\mu}}_{(j)}^{(m+1)}\right)\bigotimes_{d\neq j}\hat{\boldsymbol{\Sigma}}_d^{-1}\left(\boldsymbol{x}_{i(j)}-\hat{\boldsymbol{\mu}}_{(j)}^{(m+1)}\right)'\\ &- \hat{\boldsymbol{\delta}}_{(j)}^{(m+1)}\bigotimes_{d\neq j}\hat{\boldsymbol{\Sigma}}_d^{-1}\left(\boldsymbol{x}_{i(j)}-\hat{\boldsymbol{\mu}}_{(j)}^{(m+1)}\right)'\\ &- \left(\boldsymbol{x}_{i(j)}-\hat{\boldsymbol{\mu}}_{(j)}^{(m+1)}\right)\bigotimes_{d\neq j}\hat{\boldsymbol{\Sigma}}_d^{-1}\left(\hat{\boldsymbol{\delta}}_{(j)}^{(m+1)}\right)'\\ &+ a(\hat{\boldsymbol{\theta}}^{(m)}|\boldsymbol{x}_i)\hat{\boldsymbol{\delta}}_{(j)}^{(m+1)}\bigotimes_{d\neq j}\hat{\boldsymbol{\Sigma}}_d^{-1}\left(\hat{\boldsymbol{\delta}}_{(j)}^{(m+1)}\right)',\end{aligned}$$
$$\hat{\boldsymbol{\Sigma}}_{j(TVT)}^{(m+1)} = \frac{p_j}{np^*}\sum_{i=1}^n b(\hat{\boldsymbol{\theta}}^{(m)}|\boldsymbol{x}_i)(\boldsymbol{x}_{i(j)}-\hat{\boldsymbol{\mu}}_j^{(m+1)})\bigotimes_{d\neq j}\hat{\boldsymbol{\Sigma}}_d^{-1}(\boldsymbol{x}_{i(j)}-\hat{\boldsymbol{\mu}}_j^{(m+1)})'.$$

  As was discussed in [15] and [9] for parameter estimation in the matrix variate case and tensor-variate case, and in [16] for the order k case,

the estimates of $\mathbf{\Sigma}_k$ are unique only up to a multiplicative constant. Indeed, if we let

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

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

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

then

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

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

4. The update for the degrees of freedom cannot be obtained in closed form. Instead, we solve Equation (14) for ν to obtain $\hat{\nu}^{(m+1)}$.

$$log(\frac{\nu}{2}) + 1 - \phi(\frac{\nu}{2}) - \frac{1}{n}\sum_{i=1}^{n}\left(b(\hat{\boldsymbol{\theta}}^{(m)}|\boldsymbol{x}_i) + c(\hat{\boldsymbol{\theta}}^{(m)}|\boldsymbol{x}_i)\right), \qquad (14)$$

where ϕ is the digamma function and $c(\boldsymbol{\theta}|\boldsymbol{x}_i) = E(\log W_i|\boldsymbol{x}_i, \boldsymbol{\theta})$.
The iterations of the EM algorithm are repeated until the desired convergence criteria are achieved.

4. A simulation study

To assess the performance of the estimators proposed in the previous section under different sample sizes, we consider both the TVT and TVST distributions. Random samples of sizes $n = 100, 150$, and 200 are generated, each with 1000 replications, from the distributions $TVT_{\boldsymbol{p}}(\boldsymbol{\mu}, \mathbf{\Sigma}_{1:k}, \nu)$ and $TVST_{\boldsymbol{p}}(\boldsymbol{\mu}, \boldsymbol{\delta}, \mathbf{\Sigma}_{1:k}, \nu)$, where $\boldsymbol{p} = (2, 2, 2)$ and $\mathbf{\Sigma}_i = \boldsymbol{I}_p$ for $i = 1, 2, 3$. The tensor mean parameters are specified as follows:

$$\boldsymbol{\mu}_{::1} = \begin{bmatrix} 0.5 & 1.0 \\ -0.5 & 0.5 \end{bmatrix}, \quad \boldsymbol{\mu}_{::2} = \begin{bmatrix} 1.0 & -0.5 \\ 0.5 & 1.0 \end{bmatrix}, \quad \nu = 4.$$

Note that to generate random tensor samples from the aforementioned distributions, the `rTensor` package in R software can be employed.

As for the initialization of the EM algorithm, the sample mean tensor is considered as

$$\boldsymbol{\mu}_0 = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i,$$

and the initial estimates of the covariance matrices are obtained as

$$\boldsymbol{\Sigma}_j^{(0)} = \frac{p_j}{np^*} \sum_{i=1}^{n} (\boldsymbol{x}_{i(j)} - \boldsymbol{\mu}_j^{(0)})'(\boldsymbol{x}_{i(j)} - \boldsymbol{\mu}_j^{(0)}), \quad j = 1, 2, \ldots, k,$$

where $\boldsymbol{x}_{i(j)}$ denotes the mode-j matricization of the tensor $\boldsymbol{x}_i$ and $p^* = \prod_{j=1}^{k} p_j$. Furthermore, the convergence of the EM algorithm is assessed using the criterion

$$\max_{r \in \{1,\ldots,d\}} \left| \boldsymbol{\theta}_r^{(m+1)} - \boldsymbol{\theta}_r^{(m)} \right| < 10^{-2},$$

where d denotes the dimension of the parameter vector $\boldsymbol{\theta}$.

The simulation results, including the average bias and mean squared error (MSE) of the estimates, are summarized in Tables 1 to 4, corresponding to $\boldsymbol{\mu}$, $\boldsymbol{\delta}$, ν, and $\boldsymbol{\Sigma}_j$ $(j = 1, \ldots, p)$, respectively.

As observed from these tables, increasing the sample size (n) leads to a reduction in both the absolute value of the average bias and the mean squared error of the maximum likelihood estimates for each component of $\boldsymbol{\theta}$. This behavior confirms the consistency property of the MLEs.

To assess the accuracy of the estimated model parameters, we employ the relative error measure, as reported in Table 5. This metric quantifies the closeness between the estimated parameters and their true values, and is defined as

$$\frac{\|\hat{\boldsymbol{V}} - \boldsymbol{V}\|_F}{\|\boldsymbol{V}\|_F},$$

where $\|\cdot\|_F$ denotes the Frobenius matrix norm, $\hat{\boldsymbol{V}}$ represents the estimated parameter value, and $\boldsymbol{V}$ denotes the true parameter value used to generate the simulated data. A smaller value of this ratio indicates a more accurate estimation of the model parameters.

The results presented in Table 5 demonstrate that the estimation accuracy improves as the sample size increases. Furthermore, it can be observed that the proposed estimators perform better under the TVT distribution compared to the TVST distribution. Figure 1 illustrates the relative estimation errors under the TVT (blue) and TVST (red) models for $n = 100, 150, 200$. The TVT model yields smaller errors for the mean $(\boldsymbol{\mu})$, while the TVST achieves better accuracy for degrees of freedom (ν) and covariance components $(\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2, \boldsymbol{\Sigma}_3)$. Skewness $(\boldsymbol{\delta})$ is difficult to estimate, improving only slightly with larger n. Overall, the TVT model is preferable for location, whereas the TVST is more effective for dispersion and shape.

5. Image analysis

We now consider an analysis of red–green–blue (RGB) images, which can naturally be represented as third-order tensors. Each image consists of three color intensity

Table 1: Average bias and MSE for $\boldsymbol{\mu}$ for skew-t distribution.

Distributions	Component	Bias			MSE		
		$n = 100$	$n = 150$	$n = 200$	$n = 100$	$n = 150$	$n = 200$
ST	μ_{111}	0.0592	0.0394	0.0398	0.0885	0.0613	0.0435
	μ_{211}	0.1839	0.1654	0.1693	0.0827	0.0691	0.0541
	μ_{121}	0.0609	0.0319	0.0333	0.0728	0.0501	0.0331
	μ_{221}	0.1758	0.1597	0.1632	0.0699	0.0607	0.0455
	μ_{112}	0.0482	-0.0058	-0.0123	0.0944	0.0603	0.0421
	μ_{212}	0.1481	0.1150	0.1028	0.0707	0.0540	0.0351
	μ_{122}	0.0083	-0.0109	-0.0194	0.0706	0.0462	0.0330
	μ_{222}	0.1199	0.0982	0.0981	0.0540	0.0393	0.0287
T	μ_{111}	0.0042	-0.0026	-0.0028	0.0304	0.0203	0.0151
	μ_{211}	0.0007	-0.0045	-0.0024	0.0193	0.0119	0.0089
	μ_{121}	0.0082	0.0024	0.0003	0.0237	0.0168	0.0119
	μ_{221}	0.0093	-0.0010	-0.0018	0.0146	0.0093	0.0072
	μ_{112}	0.0123	-0.0030	0.0035	0.0272	0.0178	0.0122
	μ_{212}	0.0004	-0.0036	0.0037	0.0163	0.0104	0.0076
	μ_{122}	-0.0005	0.0051	0.0039	0.0231	0.0146	0.0112
	μ_{222}	-0.0019	0.0021	0.0022	0.0144	0.0082	0.0064

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

Distribution	Component	Bias			MSE		
		$n = 100$	$n = 150$	$n = 200$	$n = 100$	$n = 150$	$n = 200$
ST	δ_{111}	-0.2833	-0.2737	-0.2718	0.1143	0.0943	0.0910
	δ_{211}	-0.5956	-0.5859	-0.5887	0.3807	0.3577	0.3606
	δ_{121}	-0.2750	-0.2630	-0.2721	0.1005	0.0880	0.0860
	δ_{221}	-0.5886	-0.5779	-0.5848	0.3678	0.3496	0.3527
	δ_{112}	-0.2799	-0.2502	-0.2471	0.1172	0.0887	0.0776
	δ_{212}	-0.5814	-0.5623	-0.5565	0.3698	0.3352	0.3214
	δ_{122}	-0.2536	-0.2409	-0.2450	0.0936	0.0775	0.0726
	δ_{222}	-0.5637	-0.5489	-0.5545	0.3415	0.3182	0.3176

matrices corresponding to the red, green, and blue channels stacked along the third mode, thereby forming an order-3 tensor representation. The images used in this study were obtained from the CIFAR-100 dataset [17]. Specifically, we selected images of maple trees exhibiting green or yellow leaves, which belong to the CIFAR-100 class hierarchy under the superclass *trees* and subclass *maple*. To refine the selection, RGB arrays were converted into the HSL (Hue, Saturation, Lightness) color space, enabling the removal of images that did not correspond to the desired leaf colors. Each image tensor has dimensions $\boldsymbol{p} = (32, 32, 3)$. An example of one such image from our sample of 212 tensors is shown in Figure 2 (left).

Using the EM algorithm, we estimate the parameters of the TVT and TVST distributions. Figure 2 (center and right) illustrates the reconstructed color images obtained from the estimated mean tensors based on the TVT and TVST

Table 3: Average bias and MSE for $\boldsymbol{\nu}$.

Distribution	Bias			MSE		
	$n = 100$	$n = 150$	$n = 200$	$n = 100$	$n = 150$	$n = 200$
ST	0.1145	-0.0099	0.0295	0.3031	0.2770	0.2335
T	0.4384	0.4283	0.4254	0.9131	0.5215	0.4540

Table 4: Average bias and MSE for $\boldsymbol{\Sigma_r}, r = 1, 2, 3$.

(p_1, p_2, p_3)	Component		Bias			MSE		
			$n = 100$	$n = 150$	$n = 200$	$n = 100$	$n = 150$	$n = 200$
ST	Σ_1	σ_{11}	1.3367	1.6709	6.2678	2.0496	1.8482	1.1942
		$\sigma_{12} = \sigma_{21}$	0.2925	0.2186	0.2108	0.1795	0.1528	0.1251
		σ_{22}	0.4962	0.4553	0.4467	0.3330	0.3177	0.2571
	Σ_2	σ_{11}	-0.5110	-0.5097	-0.5057	0.2551	0.2523	0.2515
		$\sigma_{12} = \sigma_{21}$	-0.2142	-0.2100	-0.2105	0.0484	0.0458	0.0456
		σ_{22}	-0.6941	-0.6937	-0.6925	0.4907	0.4867	0.4818
	Σ_3	σ_{11}	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
		$\sigma_{12} = \sigma_{21}$	-0.2479	-0.2469	-0.2372	0.0634	0.0622	0.0621
		σ_{22}	-1.1554	-1.1534	-1.1412	1.3417	1.3414	1.3407
T	Σ_1	σ_{11}	1.5975	1.5665	1.5684	2.7574	2.5861	2.5617
		$\sigma_{12} = \sigma_{21}$	0.3670	0.3609	0.3576	0.1657	0.1501	0.1432
		σ_{22}	0.6235	0.6182	0.6140	0.4604	0.4191	0.4154
	Σ_2	σ_{11}	-0.5112	-0.5101	-0.5097	0.2563	0.2550	0.2543
		$\sigma_{12} = \sigma_{21}$	-0.2155	-0.2161	-0.2157	0.0490	0.0483	0.0477
		σ_{22}	-0.6972	-0.6955	-0.6946	0.4936	0.4890	0.4866
	Σ_2	σ_{11}	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
		$\sigma_{12} = \sigma_{21}$	-0.2521	-0.2531	-0.2517	0.0656	0.0654	0.0643
		σ_{22}	-1.1620	-1.1614	-1.1644	1.3563	1.3532	1.3590

distributions, respectively. As observed, key structural elements such as the sky, tree trunk, and branches are clearly distinguishable in both reconstructions. This demonstrates the capability of the proposed tensor-based models to effectively preserve essential spatial and color information. Such properties make these models particularly useful for applications including automatic vegetation classification, object detection in natural scenes, and image data compression.

The histogram in Figure 3 illustrates the pixel intensity distributions of the original image and the estimated mean tensors obtained from the TVT and TVST distributions. The original image (red) exhibits a clear bimodal pattern, with one peak located in the lower intensity range (approximately 0.2–0.4) and another near intensity 1. The TVT model (orange) successfully captures the main mode of the distribution but fails to adequately represent the secondary peak and the variability observed at higher intensity levels, indicating its limited ability to model heavy-tailed or multimodal structures in image data. The TVT model demonstrates a better fit to the real data compared to the TVST model. It more accurately captures the central peak around 0.4–0.5 and also provides an improved representation of the tails and the spread of pixel intensities. Overall, the results indicate that the TVT distribution provides a more accurate estimation of the

Table 5: Relative Error for $\boldsymbol{\mu}, \boldsymbol{\nu}, \boldsymbol{\delta}, \boldsymbol{\Sigma}_{1:3}$.

distribution	n	$\boldsymbol{\mu}$	$\boldsymbol{\nu}$	$\boldsymbol{\delta}$	$\boldsymbol{\Sigma}_1$	$\boldsymbol{\Sigma}_2$	$\boldsymbol{\Sigma}_3$
ST	100	0.3289	0.1168	0.6051	0.6294	0.6201	0.6417
	150	0.2736	0.1141	0.5788	0.6201	0.6478	0.8806
	200	0.2396	0.1070	0.5752	0.5799	0.6277	0.6327
T	100	0.1755	0.1815	-	0.6937	0.6885	0.7073
	150	0.1421	0.1422	-	0.6817	0.6894	0.6922
	200	0.1213	0.1301	-	0.6880	0.6840	0.6917

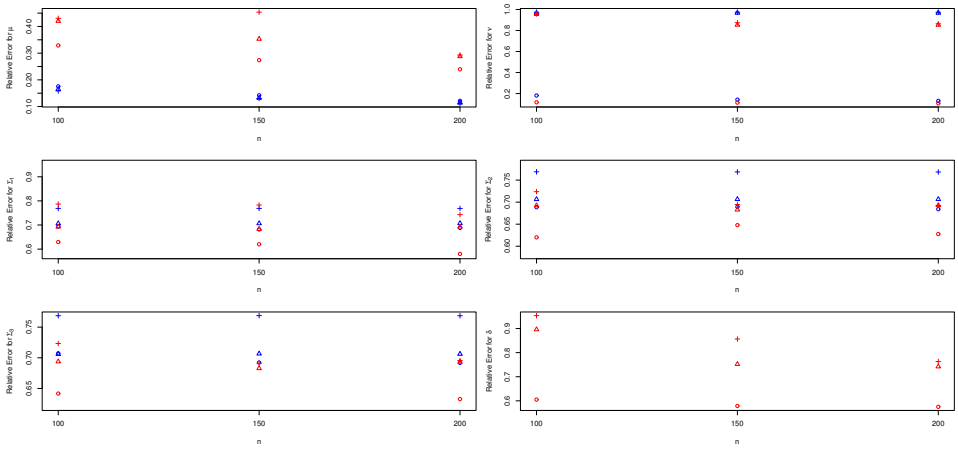

Figure 1: Relative Error for $\boldsymbol{\mu}, \nu, \boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2, \boldsymbol{\Sigma}_3$ and $\boldsymbol{\delta}$ for TVT (blue) and TVST (red) distributions.

mean and better reflects the statistical characteristics of the image data compared to the TVST model.

As the degrees of freedom increases, the proposed models reduce to the tensor-variate normal and tensor-variate skew-normal distributions. Accordingly, Table 6 presents a comparison between these two sub models ([18]) and the proposed models, via the AIC, BIC, and log-likelihood values. According to the results presented in Table 6 , the Normal and Skew-Normal models yield substantially lower AIC and BIC values compared with the tensor-variate t based models, indicating a better trade-off between model fit and complexity under these criteria. Among them, the Skew-Normal model achieves the smallest AIC and BIC, together with the highest log-likelihood value, suggesting a slightly improved fit over the Normal model. In contrast, the TVT and TVST models exhibit larger AIC and BIC values, reflecting increased model complexity. However, the TVST model attains the lowest log-likelihood, followed by the TVT model, highlighting their greater flexibility in modeling heavy-tailed and asymmetric behavior. Overall, these results

Figure 2: An image of a maple tree from CIFAR-100 data set (left) and an image of the mean tensor based on TVST and TVT distributions, respectively.

suggest that while normal-based models are favored by information criteria in this setting, tensor-variate t and skew-t models offer enhanced modeling capability for complex data structures.

Figure 2 (right) displays the color image reconstructed from the estimated mean tensor under the TVT and TVST distributions. The resulting image clearly delineates key visual elements such as the sky, tree trunk, and branches, demonstrating the ability of these tensor-based models to preserve structural details in image representation.

Table 6: Model Comparison Based on AIC, BIC, and Lnl.

	AIC	BIC	Lnl
TVT	30640404	30662099	-15312996
TVST	34642382	34664077	-17313985
Normal	22582000	22607000	-11284000
SkewNormal	22562000	22587000	-11274000

6. Conclusion

In this paper, we introduced the Tensor-Variate t (TVT) and Tensor-Variate Skew-t (TVST) distributions and investigated several of their theoretical properties, including the probability density function, expectation and characteristic function. Parameter estimation was performed using the Expectation–Maximization (EM) algorithm, and the performance of the estimators was assessed through extensive simulation studies. The results, summarized in terms of average bias and mean squared error (MSE), confirmed the consistency and accuracy of the proposed estimators, with improved performance as the sample size increased. We further applied the proposed models to real data consisting of color images represented as third-order tensors. The comparison of TVT and TVST models with alternative tensor distributions using information criteria such as AIC, BIC, and log-likelihood

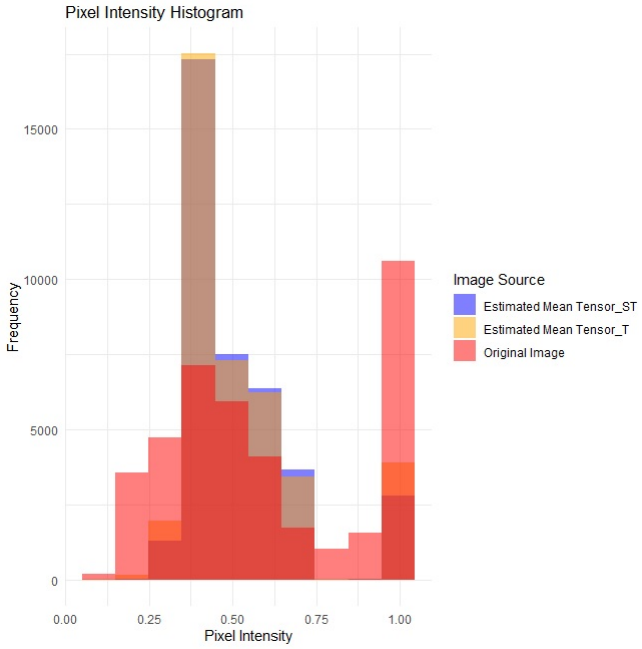

Figure 3: The distribution of pixel intensities for both the original maple tree image and the estimated mean tensors.

values demonstrated that the TVT model achieved the best overall fit, while the TVST distribution provided greater flexibility in capturing asymmetry and heavy-tailed behavior present in the data. Both models effectively reconstructed key image structures, such as the sky, branches, and leaves, showing their potential for real-world applications in image analysis and pattern recognition.

For future research, we suggest extending these models to higher-order tensors and exploring regularized or Bayesian approaches for parameter estimation to improve computational efficiency in high dimensional settings. Moreover, incorporating tensor decomposition methods and mixture extensions may further enhance clustering, classification, and feature extraction capabilities. Overall, the TVT and TVST distributions offer a robust statistical framework for modeling complex, structured, and high dimensional data.

Conflicts of Interest. The authors declare that they have no conflicts of interest regarding the publication of this article.

References

[1] H. Drygas, Linear sufficiency and some applications in multilinear estimation, *J. Multivariate Anal.* **16** (1985) 71 – 84.

[2] D. G. Northcott, *Multilinear Algebra*, Cambridge University Press, Cambridge, 1984.

[3] M. Naderi, A. Bekker, M. Arashi and A. Jamalizadeh, A theoretical framework for Landsat data modeling based on the matrix variate mean-mixture of normal model, *PLoS ONE* **15** (2020) #e0230773, https://doi.org/10.1371/journal.pone.0230773.

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

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

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

[7] M. Arashi, Some theoretical results on the tensor elliptical distribution, *J. data sci. model.* **1** (2023) 29 – 41.

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

[9] M. P. B. Gallaugher, P. A. Tait and P. D. McNicholas, Four skewed tensor distributions, arXiv:2106.08984v1, 2021.

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

[11] X. Bi, X. Tang, Y. Yuan, Y. Zhang and A. Qu, Tensors in statistics, *Annu. Rev. Stat. Appl.* **8** (2021) 345 – 368, https://doi.org/10.1146/annurev-statistics-042720-020816.

[12] M. Srivastava, T. von Rosen and D. von Rosen, Models with a Kronecker product covariance structure: estimation and testing, *Math. Methods Statist.* **17** (2008) 357 – 370, https://doi.org/10.3103/S1066530708040066.

[13] M. P. B. Gallaugher and P. D. McNicholas, Three skewed matrix variate distributions, *Statist. Probab. Lett.* **145** (2019) 103 – 109, https://doi.org/10.1016/j.spl.2018.08.012.

[14] P. M. Murray, R. P. Browne and P. D. McNicholas, Mixtures of skew-t factor analyzers, *Comput. Stat. Data Anal.* **77** (2014) 326 − 335, https://doi.org/10.1016/j.csda.2014.03.012.

[15] L. Anderlucci and C. Viroli, Covariance pattern mixture models for the analysis of multivariate heterogeneous longitudinal data, *Ann. Appl. Stat.* **9** (2015) 777 − 800, https://doi.org/10.1214/15-AOAS816.

[16] P. A. Tait, P. D. McNicholas and J. Obeid, Clustering higher order data: an application to pediatric multi-variable Longitudinal data, Available from: arXiv:1907.08566v3, 2020.

[17] A. Krizhevsky and G. Hinton, Learning multiple layers of features from tiny images, *Tech. Rep.* University of Toronto, Toronto, Ontario, 2009.

[18] S. A. A. Tajadod, F. Yousefzadeh, R. B. Arellano-Valle and S. Jomhoori, Some characteristics of tensor-variate skew-normal distribution and its application in image analysis, *J. Mahani Math. Res.* **15** (2026) 45 − 64.

S. A. A. Tajadod
Department of Statistics,
University of Birjand,
Birjand, I. R. Iran
e-mail: tajadod.aat@gmail.com

F. Yousefzadeh
Department of Statistics,
University of Birjand,
Birjand, I. R. Iran
e-mail: fyousefzadeh@birjand.ac.ir

R. B. Arellano-Valle
Department of Estadística,
Pontificia Universidad Católica de Chile,
Santiago, Chile
e-mail: reivalle@mat.uc.cl