Journal of Computational Statistics and Modeling, Vol. 1, No. 2, 29-41, December 2021

Some Theoretical Results on the Tensor Elliptical Distribution

M. Arashi*[1]

[1] Professor of statistics, Department of Statistics, Faculty of Mathematical Sciences
Ferdowsi University of Mashhad, Mashhad, Iran.

Recieved: 02/11/2019 Accepted: 04/11/2021

Abstract:
The multilinear normal distribution is a widely used tool in the tensor analysis of magnetic resonance imaging (MRI). Diffusion tensor MRI provides a statistical estimate of a symmetric 2^{nd}-order diffusion tensor for each voxel within an imaging volume. In this article, tensor elliptical (TE) distribution is introduced as an extension to the multilinear normal (MLN) distribution. Some properties, including the characteristic function and distribution of affine transformations, are given. An integral representation connecting densities of TE and MLN distributions is exhibited that is used in deriving the expectation of any measurable function of a TE variate.

Keywords: Characteristic generator; Inverse Laplace transform; Stochastic representation; Tensor; Vectorial operator.
Mathematics Subject Classification (2010): 62E15, 60E10, 53A45, 15A69.

*Corresponding Author: m_arashi_stat@yahoo.com

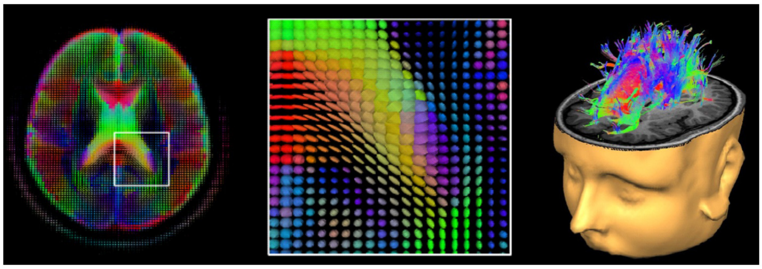

Figure 1: Visualization of tensor filed of a brain

1. Introduction

Nowadays, the analysis of matrix-valued data sets is become quite common in medical sciences, since the collected data are of multiple-way (multiple-component) arrays. For example, in medical imaging, it has become possible to collect magnetic resonance imaging (MRI) data that can be used to infer the apparent diffusivity of water in tissue in vivo. In this regard, there is a need to consider parallel extensions of bilinear forms*, namely tensor matrices. Tensor matrices have been commonly used to approximate the diffusivity profile of images. This approximation yields a diffusion tensor magnetic resonance imaging (DT-MRI) data set. The processing of DT-MRI data sets has scientific significance in clinical sciences. Figure 1 shows the tensor field in a diffusion MRI image.

In image analysis, the characteristic or precision matrix of the underlying model for tensor observations and distribution of eigenvalues play deterministic roles. Hence, the underlying tensor distribution influences the respective inference. The use of tensor and associated distributional structure in Statistics dates back to McCullagh (1987). McCullagh (1984) had already introduced tensor notation in statistics regarding the computation of polynomial cumulants. For selective papers about tensors and their applications in statistics, we refer to Sakata (2016).

In all pronounced studies in statistical tensor analysis, tensor normal (or multilinear normal) distribution is employed for the underlying distribution of observations. However, a slight change in the specification of the distribution, as pointed by Basser and Pajevic (2003), may play havoc on the resulting inferences. Hence,

*Bilinear form is a two-way (two-component) array, with each component represents a vector of observations the contribution of this study can be highlighted as follows:

- Proposing a new class of symmetric tensor distributions;
- deriving important statistical characteristics of the new class for inferential purposes;
- obtaining the maximum likelihood estimates for the location and scale parameters in the tensor field.

Therefore, this paper's plan is as follows: In section 2, we give some preliminary mathematical results for the tensor's definition. Section 3 contributes to the central part of this study and defines a new class of tensor distributions as an extension to matrix variate elliptical distribution in Statistics, along with some properties. In section 4, an underlying integral representation is given for the ease of computation, while section 5 includes some examples of the new proposed class of tensor distributions. Section 6 includes the estimation of parameters for inferential purposes. We conclude the results in section 7.

2. Preliminaries

In this section, we introduce related notation to our study and give some definitions. We adhere to the notations of Ohlson et al. (2013).

Let $\boldsymbol{\mathcal{X}}$ be a tensor of order k (k^{th}-order tensor, in tensor parlance), with the dimension $\boldsymbol{p} = (p_1, p_2, \ldots, p_k)$ in the $\boldsymbol{x} = (x_1, x_2, \ldots, x_k)$ direction. Figure 2 shows the special case when $k = 3$. Indeed 2^{nd}-order tensor is a matrix, 1^{st}-order tensor is a vector, and 0^{th}-order tensor is a scalar.

In connection with Figure 2, Figure 3 shows that the collected data can be interpreted as a tensor, where the assessment of cardiac ventricular with helical structure is done by DT-MRI.

Vectorial representation of a tensor, makes the related inference much simpler. Let $\operatorname{vec}\boldsymbol{\mathcal{X}}$ denote the vectorization of tensor $\boldsymbol{\mathcal{X}} = (x_{i_1 i_2 \cdots i_k})$, according to the definition of Kolda and Bader (2009) given by

$$\begin{aligned} \operatorname{vec}\boldsymbol{\mathcal{X}} &= \sum_{i_1=1}^{p_1} \cdots \sum_{i_k=1}^{p_k} x_{i_1 i_2 \cdots i_k} \boldsymbol{e}_{i_1}^1 \otimes \cdots \otimes \boldsymbol{e}_{i_k}^k, \\ &= \sum_{I_{\boldsymbol{p}}} x_{i_1 i_2 \cdots i_k} \boldsymbol{e}_{1:k}^{\boldsymbol{p}}, \end{aligned} \tag{2.1}$$

where $\boldsymbol{e}_{i_k}^k$, $\boldsymbol{e}_{i_{k-1}}^{k-1}$, ..., $\boldsymbol{e}_{i_1}^1$ are the unit basis vectors of size p_k, p_{k-1}, ..., p_1, respectively, $\boldsymbol{e}_{1:k}^{\boldsymbol{p}} = \boldsymbol{e}_{i_1}^1 \otimes \cdots \otimes \boldsymbol{e}_{i_k}^k$, where $\otimes$ denotes the Kronecker product, $\boldsymbol{I_p}$ is the

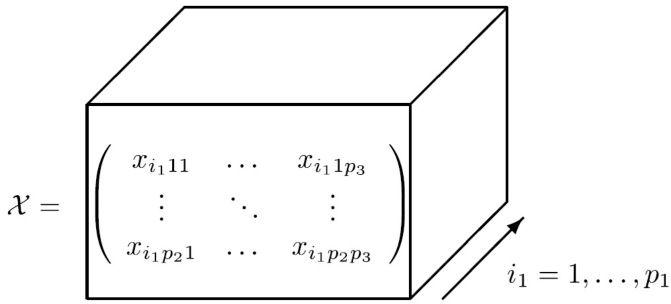

Figure 2: Visualization of a 3-dimensional data set as a 3$^{\text{rd}}$-order tensor.

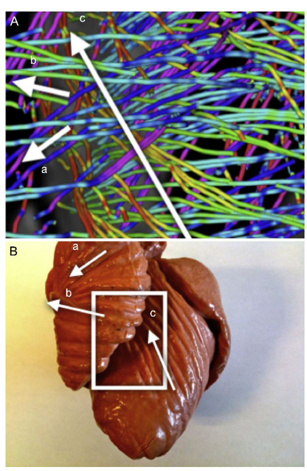

Figure 3: Helical structure of the cardiac ventricular anatomy

index set defined as $I_{\boldsymbol{p}} = \{i_1, \ldots, i_k : 1 \leq i_j \leq p_j, 1 \leq j \leq k\}$. In Ohlson et al. (2012), 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 Srivastava et al. (2008), for multilinear normal (MLN) distributions.

Let $\mathcal{T}^{\boldsymbol{p}}$ denote the space of all vectors $\boldsymbol{x} = \operatorname{vec} \boldsymbol{\mathcal{X}}$, where $\boldsymbol{\mathcal{X}}$ is a tensor of order k, i.e., $\mathcal{T}^{\boldsymbol{p}} = \{\boldsymbol{x} : \boldsymbol{x} = \sum_{I_{\boldsymbol{p}}} x_{i_1 i_2 \cdots i_k} \boldsymbol{e}^{\boldsymbol{p}}_{1:k}\}$. 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*

(i) $\boldsymbol{\mathcal{T}}^{\boldsymbol{pq}} = \left\{\boldsymbol{X} : \boldsymbol{X} = \sum_{I_{\boldsymbol{p}} \cup I_{\boldsymbol{q}}} x_{i_1, \cdots, i_k, j_1, \cdots, j_l} \boldsymbol{e}^{\boldsymbol{p}}_{1:k} (\boldsymbol{d}^{\boldsymbol{q}}_{1:l})'\right\}$,
$I_{\boldsymbol{q}} = \{j_1, \ldots, j_l : 1 \leq j_i \leq p_i,\ 1 \leq i \leq l\}$

(ii) $\boldsymbol{\mathcal{T}}^{\boldsymbol{pq}}_{\otimes} = \{\boldsymbol{X} \in \boldsymbol{\mathcal{T}}^{\boldsymbol{pq}} : \boldsymbol{X} = \boldsymbol{X}_1 \otimes \ldots \otimes \boldsymbol{X}_k, \boldsymbol{X}_i : p_i \times q_i\}$

(iii) $\boldsymbol{\mathcal{T}}^{\boldsymbol{p}}_{\otimes} = \left\{\boldsymbol{X} \in \boldsymbol{\mathcal{T}}^{\boldsymbol{pp}}_{\otimes} : \boldsymbol{X} = \boldsymbol{X}_1 \otimes \ldots \otimes \boldsymbol{X}_k, \boldsymbol{X}_i : p_i \times p_i\right\}$

Theorem 2.2. *(Ohlson et al. (2013)) A tensor $\boldsymbol{\mathcal{X}}$ is MLN of order k, denoted by $\boldsymbol{\mathcal{X}} \sim \mathcal{N}_{\boldsymbol{p}}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ if $\boldsymbol{x} = \boldsymbol{\mu} + \boldsymbol{\Sigma}^{\frac{1}{2}} \boldsymbol{u}$, where $\boldsymbol{x}, \boldsymbol{\mu} \in \boldsymbol{\mathcal{T}}^{\boldsymbol{p}}$, $\boldsymbol{\Sigma} \in \boldsymbol{\mathcal{T}}^{\boldsymbol{p}}_{\otimes}$, $\boldsymbol{p} = (p_1, \ldots, p_k)$, and the elements of $\boldsymbol{u} \in \boldsymbol{\mathcal{T}}^{\boldsymbol{p}}$ are independent standard normally distributed.*

Note that $\boldsymbol{\Sigma} \in \boldsymbol{\mathcal{T}}^{\boldsymbol{p}}_{\otimes}$ can be written as Kronecker product $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_1 \otimes \ldots \otimes \boldsymbol{\Sigma}_k$

Indeed, Theorem 2.2 configures the MLN distribution using the stochastic representation of the vector $\boldsymbol{x} \in \boldsymbol{\mathcal{T}}^{\boldsymbol{p}}$. This methodology can be mimicked to extend the above result for elliptical models. Before revealing the main result of this paper, we need to consider the definition of matrix elliptical distributions.

3. Tensor Elliptical Distributions

Let $\boldsymbol{u}^{(p^*)}$, $p^* = \prod_{i=1}^{k} p_i$, denote a random vector distributed uniformly on the unit sphere surface in $\mathbb{R}^{p^*}$, with characteristic function (cf) $\Omega_{p^*}(\cdot)$. Hereafter, using Theorem 2.2 of Fang et al. (1990), we propose a definition for tensor elliptical (TE) distribution. The methodology behind our definition of TE distribution comes from two facts: (1) a random matrix $\boldsymbol{X}$ has matrix elliptical distribution if and only if $\operatorname{vec} \boldsymbol{X}$ has a vector-variate elliptical distribution, which will be used for tensor (see Gupta et al. (2013)) (2) the difference between vector-variate elliptical and TE lies in the structure of the parameter space generated by $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$.

Definition 3.1. *A random tensor* $\boldsymbol{\mathcal{X}}$ *is TE of order* k*, denoted by* $\boldsymbol{\mathcal{X}} \sim \mathcal{E}_{\boldsymbol{p}}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \psi)$*, if*

$$\boldsymbol{x} = \mathrm{vec}(\boldsymbol{\mathcal{X}}) = \boldsymbol{\mu} + \mathcal{R}\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{u}^{(p^*)}, \tag{3.2}$$

where $\boldsymbol{x}, \boldsymbol{\mu} \in \boldsymbol{\mathcal{T}}^{\boldsymbol{p}}$*,* $\boldsymbol{\Sigma}^{\frac{1}{2}} \in \boldsymbol{\mathcal{T}}^{\boldsymbol{p}}_{\otimes}$ *is any square root,* $\boldsymbol{p} = (p_1, \ldots, p_k)$*,* $\mathcal{R} \geq 0$ *is independent of* $\boldsymbol{u}^{(p^*)}$*, and* $\mathcal{R} \sim F$*, for some cumulative distribution function (cdf)* $F(\cdot)$ *over* $[0, \infty)$*, is related to* ψ *by the following relation*

$$\psi(x) = \int_{\mathbb{R}^+} \Omega_{p^*}(xr^2)\mathrm{d}F(r). \tag{3.3}$$

The question arises whether the parameters in Definition 3.1 are uniquely defined. The answer is no. To see this, assume that a_i, $i = 1, \ldots, k$ are positive constants such that $a^* = \prod_{j=1}^{k} a_j$, $\boldsymbol{\Sigma}_j^* = a_j\boldsymbol{\Sigma}_j$, $j = 1, \ldots, k$ and $\psi^*(x) = \psi\left(\frac{1}{p^*}x\right)$. Then $\mathcal{E}_{\boldsymbol{p}}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \psi)$ and $\mathcal{E}_{\boldsymbol{p}}(\boldsymbol{\mu}, \boldsymbol{\Sigma}^*, \psi^*)$, where $\boldsymbol{\Sigma}^* = \boldsymbol{\Sigma}_1^* \otimes \ldots \otimes \boldsymbol{\Sigma}_k^*$, define the same tensor elliptical distribution.

Using the vector representation for $\boldsymbol{x}$, we can conveniently write the probability distribution function (pdf) of a TE distribution. The following result gives the pdf of a random tensor elliptical if it possesses a density, as an extension to Ohlson et al. (2013).

Theorem 3.2. *Under the assumptions of Definition 3.1, the pdf of the TE distribution is given by*

$$f_{\boldsymbol{\mathcal{X}}}(\boldsymbol{x}) = |\boldsymbol{\Sigma}|^{-\frac{1}{2}} g\left[(\boldsymbol{x} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x} - \boldsymbol{\mu})\right],$$

where $g(\cdot)$ *is a non-negative function (density generator, say) satisfying*

$$\int_{\mathbb{R}^+} y^{\frac{1}{2}p^*-1} g(y)\mathrm{d}y < \infty.$$

We designate $\boldsymbol{\mathcal{X}} \sim \mathcal{E}_{\boldsymbol{p}}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, g)$.

Similarly, we have the following result.

Theorem 3.3. *Let* $\boldsymbol{\mathcal{X}} \sim \mathcal{E}_{\boldsymbol{p}}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \psi)$*. Then, its characteristic function has the form*

$$\phi_{\boldsymbol{\mathcal{X}}}(\boldsymbol{\mathcal{S}}) = e^{i\boldsymbol{\mathcal{S}}'\boldsymbol{\mu}}\psi(\boldsymbol{\mathcal{S}}'\boldsymbol{\Sigma}\boldsymbol{\mathcal{S}}), \quad \boldsymbol{\mathcal{S}} \in \boldsymbol{\mathcal{T}}^{\boldsymbol{p}}. \tag{3.4}$$

Remark 3.4. *Since*

$$\begin{aligned} |\boldsymbol{\Sigma}|^{-\frac{1}{2}} &= |\boldsymbol{\Sigma}_1 \otimes \boldsymbol{\Sigma}_2 \otimes \ldots \otimes \boldsymbol{\Sigma}_k|^{-\frac{1}{2}} \\ &= \left(|\boldsymbol{\Sigma}_1|^{\frac{p^*}{p_1}}\right)^{-\frac{1}{2}} \times \left(|\boldsymbol{\Sigma}_2|^{\frac{p^*}{p_2}}\right)^{-\frac{1}{2}} \times \ldots \times \left(|\boldsymbol{\Sigma}_k|^{\frac{p^*}{p_k}}\right)^{-\frac{1}{2}} \end{aligned}$$

$$= \prod_{i=1}^{k} |\mathbf{\Sigma}_i|^{-\frac{p^*}{2p_i}}$$

taking $g(y) = (2\pi)^{-\frac{1}{2}p^*} \exp\left(-\frac{1}{2}y\right)$ *in Definition 3.1, gives the pdf of MLN distribution (as given in Theorem 1 of Ohlson et al. (2013)) as*

$$f_{\boldsymbol{\mathcal{X}}}(\boldsymbol{x}) = (2\pi)^{-\frac{1}{2}p^*} \left(\prod_{i=1}^{k} |\mathbf{\Sigma}_i|^{-\frac{p^*}{2p_i}}\right) \exp\left[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})'\mathbf{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right], \tag{3.5}$$

where $\mathbf{\Sigma}$ *is positive definite,* $\boldsymbol{x}, \boldsymbol{\mu} \in \boldsymbol{\mathcal{T}}^{\boldsymbol{p}}$, $\mathbf{\Sigma} \in \boldsymbol{\mathcal{T}}^{\boldsymbol{p}}_{\otimes}$, *and* $p^* = \prod_{i=1}^{k} p_i$.

The following result gives the distribution of affine transformations for TE variates.

Theorem 3.5. *Let* $\boldsymbol{\mathcal{X}} \sim \mathcal{E}_{\boldsymbol{p}}(\boldsymbol{\mu}, \mathbf{\Sigma}, \psi)$, *with* $\operatorname{vec}\boldsymbol{\mathcal{X}} \in \boldsymbol{\mathcal{T}}^{\boldsymbol{p}}$, $\boldsymbol{A} \in \boldsymbol{\mathcal{T}}^{\boldsymbol{qp}}$ *is nonsingular, and* $\boldsymbol{B} \in \boldsymbol{\mathcal{T}}^{\boldsymbol{q}}$. *Then,* $\boldsymbol{A\mathcal{X}} + \boldsymbol{B} \sim \mathcal{E}_{\boldsymbol{q}}(\boldsymbol{A\mu} + \boldsymbol{B}, \boldsymbol{A}\mathbf{\Sigma}\boldsymbol{A}')$, *where* $\boldsymbol{A\mu} + \boldsymbol{B} \in \boldsymbol{\mathcal{T}}^{\boldsymbol{q}}$ *and* $\boldsymbol{A}\mathbf{\Sigma}\boldsymbol{A}' \in \boldsymbol{\mathcal{T}}^{\boldsymbol{q}}_{\otimes}$.

Proof. Let $\boldsymbol{y} = \boldsymbol{Ax}+\boldsymbol{B}$, where $\boldsymbol{x} = \operatorname{vec}(\boldsymbol{\mathcal{X}})$. From the stochastic representation in Definition 3.1, the proof directly follows from $\boldsymbol{y} = (\boldsymbol{A\mu}+\boldsymbol{B})+\mathcal{R}(\boldsymbol{A}\mathbf{\Sigma}^{\frac{1}{2}})\boldsymbol{u}^{(p^*)}$. □

The following result is a direct consequence of Theorem 2.16 of Gupta et al. (2013) for tensor elliptical distributions.

Theorem 3.6. *Under the assumptions of Definition 3.1, the pdf of* $\mathcal{R}$ *has from*

$$h_{\mathcal{R}}(r) = \frac{2\pi^{\frac{1}{2}p^*}}{\Gamma\left(\frac{1}{2}p^*\right)} r^{p^*-1} g\left(r^2\right), \quad r \geq 0.$$

The following theorem reveals the distribution of quadratic form for a special case.

Theorem 3.7. *Let* $\boldsymbol{\mathcal{X}} \sim \mathcal{E}_{\boldsymbol{p}}(\mathbf{0}, \mathbf{\Sigma}^{(1)}, \psi)$, *where* $\mathbf{\Sigma}^{(1)} = \mathbf{\Sigma} \otimes \boldsymbol{I}_{p_2} \otimes \ldots \otimes \boldsymbol{I}_{p_k} \in \boldsymbol{\mathcal{T}}^{\boldsymbol{p}}_{\otimes}$, $\boldsymbol{p} = (p_1, \ldots, p_k)$. *Then, the pdf of* $\boldsymbol{\mathcal{A}} = \boldsymbol{\mathcal{X}}\boldsymbol{\mathcal{X}}'$ *is given by*

$$f(\boldsymbol{\mathcal{A}}) = \frac{\pi^{p^*}|\mathbf{\Sigma}|^{-\frac{1}{2}p_1}}{\Gamma_{p_1}\left(\frac{1}{2}p^{(1)}\right)} |\boldsymbol{\mathcal{A}}|^{\frac{1}{2}p^{(1)}-p_1-1} g(\operatorname{tr}\mathbf{\Sigma}^{-1}\boldsymbol{\mathcal{A}})$$

where $p^{(1)} = \prod_{j=2}^{k} p_j$.

In the forthcoming section, we provide a weighting representation of the pdf of TE variate using the Laplace operator.

4. Weighting Representation

Although the proposed theorems in the previous section are obtained conventionally, it is not easy to achieve other statistical properties of the TE distributions from Definition 3.1 straightforwardly. However, under mild conditions, one can make a connection between densities of TE and MLN pdfs and derive other properties of the TE distributions using MLN distributions. In this section, we propose a weighting representation that connects densities of the TE and MLN distributions. This result is given in the following theorem.

Theorem 4.1. *Let $\boldsymbol{\mathcal{X}} \sim \mathcal{E}_{\boldsymbol{p}}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, g)$, where $\boldsymbol{\mu} \in \boldsymbol{\mathcal{T}}^{\boldsymbol{p}}$, $\boldsymbol{\Sigma} \in \boldsymbol{\mathcal{T}}^{\boldsymbol{p}}_{\otimes}$ and $g : \mathbb{R}^+ \to \mathbb{R}^+$. Also assume that $g(s^2)$ is differentiable when s^2 is sufficiently large, and $g(s^2)$ vanishes faster than s^{-k}; $k > 1$ as $s \to \infty$. Then, the pdf of $\boldsymbol{\mathcal{X}}$ can be represented as an integral of a series of MLN pdfs given by*

$$f_{\boldsymbol{\mathcal{X}}}(\boldsymbol{x}) = \int_{\mathbb{R}^+} \mathcal{W}(t) f_{\mathcal{N}_{\boldsymbol{p}}(\boldsymbol{\mu}, t^{-1}\boldsymbol{\Sigma})}(\boldsymbol{x}) \mathrm{d}t,$$

where $f_{\mathcal{N}_{\boldsymbol{p}}(\boldsymbol{\mu}, t^{-1}\boldsymbol{\Sigma})}(\cdot)$ is the pdf of $\mathcal{N}_{\boldsymbol{p}}(\boldsymbol{\mu}, t^{-1}\boldsymbol{\Sigma})$ and $\mathcal{W}(\cdot)$ is a weighting function.

Proof. Let $s^2 = \frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x} - \boldsymbol{\mu})$ and

$$\mathcal{W}(t) = (2\pi)^{\frac{1}{2}p^*} t^{-\frac{p^*}{2}} \mathcal{L}^{-1}\left[g\left(2s^2\right)\right],$$

where $\mathcal{L}$ is the Laplace transform operator. It should be noted that under the regularity condition on $g(s^2)$, the inverse Laplace transform exists. Then, from $|\boldsymbol{\Sigma}|^{-\frac{1}{2}} = \prod_{i=1}^k |\boldsymbol{\Sigma}_i|^{-\frac{p^*}{2p_i}}$, we have

$$\begin{aligned} f(\boldsymbol{x}) = |\boldsymbol{\Sigma}|^{-\frac{1}{2}} g(2s^2) &= |\boldsymbol{\Sigma}|^{-\frac{1}{2}} \mathcal{L}\left[\mathcal{W}(t)(2\pi)^{-\frac{1}{2}p^*} t^{\frac{p^*}{2}}\right] \\ &= \mathcal{L}\left[\mathcal{W}(t)(2\pi)^{-\frac{1}{2}p^*} t^{\frac{p^*}{2}} \prod_{i=1}^k |\boldsymbol{\Sigma}_i|^{-\frac{p^*}{2p_i}}\right] \\ &= \int_{\mathbb{R}^+} \mathcal{W}(t)(2\pi)^{-\frac{1}{2}p^*} t^{\frac{p^*}{2}} \left(\prod_{i=1}^k |\boldsymbol{\Sigma}_i|^{-\frac{p^*}{2p_i}}\right) e^{-ts^2} \mathrm{d}t \\ &= \int_{\mathbb{R}^+} \mathcal{W}(t)(2\pi)^{-\frac{1}{2}p^*} \left(\prod_{i=1}^k |t^{-\frac{1}{k}}\boldsymbol{\Sigma}_i|^{-\frac{p^*}{2p_i}}\right) e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})'(t^{-1}\boldsymbol{\Sigma})^{-1}(\boldsymbol{x}-\boldsymbol{\mu})} \mathrm{d}t \\ &= \int_{\mathbb{R}^+} \mathcal{W}(t) f_{\mathcal{N}_{\boldsymbol{p}}(\boldsymbol{\mu}, t^{-1}\boldsymbol{\Sigma})}(\boldsymbol{x}) \mathrm{d}t. \end{aligned}$$

The proof is complete. □

Thus, a TE variable is an integral over all MLN variables having the same covariance subject to different scales.

Since $f_{\boldsymbol{\mathcal{X}}}(\cdot)$ is the pdf of $\boldsymbol{\mathcal{X}}$, using Fubini's theorem, we obtain

$$
\begin{aligned}
1 = \int_{\chi} f_{\boldsymbol{\mathcal{X}}}(\boldsymbol{x})\mathrm{d}\boldsymbol{x} &= \int_{\chi}\int_{\mathbb{R}^+} \mathcal{W}(t) f_{\mathcal{N}_{\boldsymbol{p}}(\boldsymbol{\mu},t^{-1}\boldsymbol{\Sigma})}(\boldsymbol{x})\mathrm{d}t\mathrm{d}\boldsymbol{x} \\
&= \int_{\mathbb{R}^+} \mathcal{W}(t) \int_{\chi} f_{\mathcal{N}_{\boldsymbol{p}}(\boldsymbol{\mu},t^{-1}\boldsymbol{\Sigma})}(\boldsymbol{x})\mathrm{d}\boldsymbol{x}\mathrm{d}t \\
&= \int_{\mathbb{R}^+} \mathcal{W}(t)\mathrm{d}t \qquad (4.6)
\end{aligned}
$$

where χ is the sample space, hence, for positive weighting functions $\mathcal{W}(\cdot)$, the weighting representation of TE distributions can be interpreted as a scale mixture of MLN distributions. However, sometimes, $\mathcal{W}(\cdot)$ can be negative. Note that a TE distribution is completely defined by the matrix $\boldsymbol{\Sigma} \in \boldsymbol{\mathcal{T}}^{\boldsymbol{p}}_{\otimes}$ and the scalar weighting function $\mathcal{W}(\cdot)$.

Theorem 4.1 enables us to describe more properties of TE distributions via MLN distributions. This can be done using the following important result.

Theorem 4.2. *Let $\boldsymbol{x} \sim \mathcal{E}_{\boldsymbol{p}}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, g)$, $\boldsymbol{\mu} \in \boldsymbol{\mathcal{T}}^{\boldsymbol{p}}$, $\boldsymbol{\Sigma} \in \boldsymbol{\mathcal{T}}^{\boldsymbol{p}}_{\otimes}$ and $g : \mathbb{R}^+ \to \mathbb{R}^+$ with weighting function $\mathcal{W}(\cdot)$, and $B(\boldsymbol{x})$ be any Borel measurable function of $\boldsymbol{x} \in \boldsymbol{\mathcal{T}}^{\boldsymbol{p}}$. Then, if $E[B(\boldsymbol{x})]$ exists, we have*

$$E[B(\boldsymbol{x})] = \int_{\mathbb{R}^+} \mathcal{W}(t) E_{\mathcal{N}_{\boldsymbol{p}}(\boldsymbol{\mu},t^{-1}\boldsymbol{\Sigma})}[B(\boldsymbol{x})]\mathrm{d}t$$

5. Examples

In this section, we provide some examples of TE distributions based on Definition 3.1 with respective weighting function, as defined in Theorem 4.1.

Firstly, we consider some examples in which the weighting function $\mathcal{W}(.)$ is always positive, resulting to scale mixture of multilinear normal distributions.

(i) Multilinear normal distribution (Ohlson et al. (2013))
The weighting function has the form

$$\mathcal{W}(t) = \delta(t-1),$$

where $\delta(\cdot)$ is the Dirac delta or impulse function having the property $\int\limits_{\mathbb{R}} f(x)\delta(x)dx = f(0)$, for every Borel-measurable function $f(\cdot)$.

(ii) Multilinear ε-contaminated normal distribution
We say the random tensor $\boldsymbol{\mathcal{X}} \in \boldsymbol{\mathcal{T}}^{\boldsymbol{p}}$ has multilinear ε-contaminated normal

distribution if it has the following density

$$f_{\mathcal{X}}(\boldsymbol{x}) = \frac{1}{(2\pi)^{\frac{1}{2}p^*}} \left(\prod_{i=1}^{k} |\boldsymbol{\Sigma}_i|^{-\frac{p^*}{2p_i}} \right) \left\{ (1-\varepsilon) \exp\left[-\frac{1}{2} (\boldsymbol{x}-\boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\boldsymbol{x}-\boldsymbol{\mu}) \right] \right.$$
$$\left. + \frac{\varepsilon}{\sigma^{p^*}} \exp\left[-\frac{1}{2\sigma^2} (\boldsymbol{x}-\boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\boldsymbol{x}-\boldsymbol{\mu}) \right] \right\}.$$

Then it can be concluded that the weighting function is given by

$$\mathcal{W}(t) = (1-\varepsilon)\delta(t-1) + \varepsilon\delta(t-\sigma^2).$$

(iii) Tensor t-distribution

We say the random matrix $\mathcal{X} \in \mathcal{T}^{\boldsymbol{p}}$ has tensor t-distribution if it has the following density

$$f_{\mathcal{X}}(\boldsymbol{x}) = \frac{\nu^{\frac{p^*}{2}} \Gamma\left(\frac{p^*+\nu}{2}\right)}{\pi^{\frac{1}{2}p^*} \Gamma(\frac{\nu}{2})} \left\{ 1 + \frac{1}{\nu} (\boldsymbol{x}-\boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\boldsymbol{x}-\boldsymbol{\mu}) \right\}^{-(p^*+\nu)}. \quad (5.7)$$

The corresponding weighting function has the form $\mathcal{W}(t) = \frac{\left(\frac{t\nu}{2}\right)^{\frac{\nu}{2}} e^{-\frac{t\nu}{2}}}{t\Gamma\left(\frac{\nu}{2}\right)}$.

The tensor Cauchy distribution is obtained by setting $\nu = 1$ in (5.7).

It is of much interest to consider cases in which the weighting function $\mathcal{W}(.)$ is not always positive. Such kinds of distributions are not scale mixture of multilinear normal distributions. The item below is not a tensor distribution, however, it is 0^{th}-order tensor distribution.

(iv) The one-dimensional distribution with the following density

$$f(x) = \frac{\sqrt{2}}{\pi\sigma} \left[1 + \left(\frac{x}{\sigma} \right)^4 \right]^{-1},$$

where the weighting function is given by $\mathcal{W}(t) = \frac{1}{\sqrt{t\pi}} \sin\left(\frac{t}{2}\right)$.

6. Inference

Theorem 6.1. *Suppose that tensor variables $\mathcal{X}_1, \ldots, \mathcal{X}_n$ are jointly distributed with the following pdf*

$$\prod_{i=1}^{k} |\boldsymbol{\Sigma}_i|^{-\frac{p^*}{2p_i}} g\left(\sum_{j=1}^{n} \boldsymbol{x}_j' \boldsymbol{\Sigma}^{-1} \boldsymbol{x}_j \right), \quad \boldsymbol{\Sigma} = \boldsymbol{\Sigma}_1 \otimes \boldsymbol{\Sigma}_2 \otimes \ldots \otimes \boldsymbol{\Sigma}_k$$

such that $\sigma^{(2)}_{p_2p_2} = \sigma^{(3)}_{p_3p_3} = \ldots = \sigma^{(k)}_{p_kp_k} = 1$, *where* $\boldsymbol{\Sigma}_r = \left(\sigma^{(r)}_{ij}\right)$. *Further, suppose* $g(\cdot)$ *is such that* $g(\boldsymbol{x}'\boldsymbol{x})$ *is a pdf in* $\mathbb{R}^{p^*}$ *and* $y^{p^*/2}g(y)$ *has a finite positive maximum* y_g. *Suppose that* $\tilde{\boldsymbol{\Sigma}}$ *is an estimator which obtains from solving the following equations*

$$\begin{aligned}
\tilde{\boldsymbol{\Sigma}}_1 &= \frac{1}{p^*_{2:k}n}\sum_{j=1}^{n}\boldsymbol{x}'_j\boldsymbol{\Sigma}^{-1}_{2:k}\boldsymbol{x}_j \\
&\quad \text{and, for } r = 2,\ldots,k \\
\tilde{\boldsymbol{\Sigma}}_r &= \frac{1}{p^*_{1:r-1}p^*_{r+1:k}n}\sum_{j=1}^{n}\boldsymbol{x}_i^{2,r(r)\prime}\left(\boldsymbol{\Sigma}^{2,r}_{1:k\backslash r}\right)^{-1}\boldsymbol{x}_i^{2,r(r)},
\end{aligned}$$

where

$$\begin{aligned}
\boldsymbol{x}_j^{2,r(r)} &= \sum_{l_{\boldsymbol{p}}} x_{i_1,\ldots i_k}\boldsymbol{e}^{2,r}_{i_1:i_k\backslash i_r}\boldsymbol{e}^{p_r\prime}_{i_r} \\
\boldsymbol{\Sigma}^{2,r}_{1:k\backslash r} &= \boldsymbol{\Sigma}_2\otimes\ldots\otimes\boldsymbol{\Sigma}_{r-1}\otimes\boldsymbol{\Sigma}_1\otimes\ldots\otimes\boldsymbol{\Sigma}_k \\
\boldsymbol{e}^{2,r}_{i_1:i_k\backslash i_r} &= \boldsymbol{e}^{p_1}_{i_1}\otimes\boldsymbol{e}^{p_3}_{i_3}\otimes\ldots\otimes\boldsymbol{e}^{p_{r-1}}_{i_{r-1}}\otimes\boldsymbol{e}^{p_{r+1}}_{i_{r+1}}\otimes\ldots\otimes\boldsymbol{e}^{p_k}_{i_k}
\end{aligned}$$

Then, the MLE of $\boldsymbol{\Sigma}$ *is given by*

$$\hat{\boldsymbol{\Sigma}} = \frac{p^*}{y_g}\tilde{\boldsymbol{\Sigma}}$$

Proof. Let $\boldsymbol{A} = |\boldsymbol{\Sigma}|^{-\frac{1}{p^*}}\boldsymbol{\Sigma}$. Also for any $j = 1,\ldots,n$ write

$$d_j = \boldsymbol{x}'_j\boldsymbol{\Sigma}^{-1}\boldsymbol{x}_j = |\boldsymbol{\Sigma}|^{-\frac{1}{p^*}}\boldsymbol{x}'_j\boldsymbol{A}^{-1}x_j. \tag{6.8}$$

Since $|\boldsymbol{\Sigma}|^{-\frac{1}{2}} = \prod_{i=1}^k|\boldsymbol{\Sigma}_i|^{-\frac{p^*}{2p_i}}$, the likelihood can be written as

$$\begin{aligned}
\mathcal{L} &= |\boldsymbol{\Sigma}|^{-\frac{1}{2}}g\left(\sum_{j=1}^n d_j\right) \\
&= \left(|\boldsymbol{\Sigma}|^{\frac{1}{p^*}}\right)^{-\frac{p^*}{2}}\left(\sum_{j=1}^n d_j\right)^{\frac{p^*}{2}} g\left(\sum_{j=1}^n d_j\right) \\
&= \left(\sum_{j=1}^n x'_j\boldsymbol{A}^{-1}x_j\right)^{-\frac{p^*}{2}} d^{\frac{p^*}{2}}g(d),
\end{aligned} \tag{6.9}$$

where $d = \sum_{j=1}^n d_j$.

The maximum of (6.9) is attained at $\hat{\boldsymbol{A}} = \tilde{\boldsymbol{A}}$ and $\hat{d} = y_g$. Then the MLE of $\boldsymbol{\Sigma}$ is given by

$$\hat{\boldsymbol{\Sigma}} = |\hat{\boldsymbol{\Sigma}}|^{\frac{1}{p^*}}\hat{\boldsymbol{A}} = \frac{|\hat{\boldsymbol{\Sigma}}|^{\frac{1}{p^*}}}{|\tilde{\boldsymbol{\Sigma}}|^{\frac{1}{p^*}}}\tilde{\boldsymbol{\Sigma}}. \tag{6.10}$$

On the other hand, from (6.8) we get

$$
\begin{aligned}
|\hat{\boldsymbol{\Sigma}}|^{\frac{1}{p^*}} &= \frac{\sum_{j=1}^{n} \boldsymbol{x}_j' \hat{\boldsymbol{A}}^{-1} \boldsymbol{x}_j}{\sum_{j=1}^{n} \hat{d}_j} = \frac{\sum_{j=1}^{n} \boldsymbol{x}_j' \tilde{\boldsymbol{A}}^{-1} \boldsymbol{x}_j}{\hat{d}} = \frac{\sum_{j=1}^{n} \boldsymbol{x}_j' \tilde{\boldsymbol{A}}^{-1} \boldsymbol{x}_j}{y_g} \\
|\tilde{\boldsymbol{\Sigma}}|^{\frac{1}{p^*}} &= \frac{\sum_{j=1}^{n} \boldsymbol{x}_j' \tilde{\boldsymbol{A}}^{-1} \boldsymbol{x}_j}{\sum_{j=1}^{n} \tilde{d}_j} = \frac{\sum_{j=1}^{n} \boldsymbol{x}_j' \tilde{\boldsymbol{A}}^{-1} \boldsymbol{x}_j}{\tilde{d}} = \frac{\sum_{j=1}^{n} \boldsymbol{x}_j' \tilde{\boldsymbol{A}}^{-1} \boldsymbol{x}_j}{p^*}
\end{aligned}
\tag{6.11}
$$

Substituting (6.11) in (6.10) and using Theorem 4.1 of Ohlson et al. (2013) gives the result. □

7. Conclusion

In this article, for robust inferring on diffusion tensor magnetic resonance imaging (DT-MRI) observations, we proposed a class of tensor elliptical (TE) distributions. This class includes many heavier tail distributions than the tensor normal or multilinear normal (MLN) distribution. Important statistical properties, including the characteristic function along with the distribution of affine transformations derived. The weighting representation is also exhibited that connects densities of TE and MLN distributions. The result of this paper can be well-used in tensor regression; see Carlos (2018) for details.

Acknowledgments

The author would like to thank the two anonymous referees for their insightful comments, which substantially improved the paper.

References

Cameron A. C. and Trivedi P. K. (2013), *Regression Analysis of Count Data*, 2nd Ed., NewYork, Cambridge University Press.

Feller W. (1972), *An Introduction to Probability Theory and Its Applications*, New York, John Wiley.

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.

Fang, K.T., Kotz, S., Ng, K.W. (1990) *Symmetric Multivariate and Related Distributions*. Chapman and Hall, London.

Gupta, A.K. and Varga, T., and Bondar, T. (2013) *Elliptically Contoured Models in Statistics and Portfolio Theory*, 2nd Ed., Springer, New York.

Kolda, T.G. and Bader, B.W. (2009). Tensor decompositions and applications. *SIAM Review*, **51**(3), 455-500.

Carlos, L. (2018). Tensor on tensor regression with tensor normal errors and tensor network states on the regression parameter. *Creative Components*, 82. https://lib.dr.iastate.edu/creativecomponents/82.

McCullagh, P. (1987). *Tensor Methods in Statistics*, Chapman & Hall, London.

McCullagh, P. (1984). Tensor notation and cumulants of polynomials. *Biometrika*, **71**(3), 461-476.

Ohlson, M., Rauf Ahmad, M., and von Rosen, D. (2012). More on the Kronecker structured covariance matrix. *Communications in Statistics Theory and Methods*, **41**, 2512-2523.

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.

Sakata, T. (2016). *Applied Matrix and Tensor Variate Data Analysis*, Springer, Japan.

Srivastava, M., von Rosen, T., and von Rosen, D. (2008). Models with a Kronecker product covariance structure: estimation and testing. *Mathematical Methods in Statistics*, **17**, 357-370.