Electronic Journal of Statistics
Vol. 11 (2017) 2424–2460
ISSN: 1935-7524
DOI: 10.1214/17-EJS1289

Estimation of mean form and mean form difference under elliptical laws

José A. Díaz-García

Universidad Autónoma Agraria Antonio Narro
Calzada Antonio Narro 1923, Col. Buenavista
25315 Saltillo, Coahuila, México
e-mail: jadiaz@uaaan.mx

and

Francisco J. Caro-Lopera

Faculty of Basic Sciences
Universidad de Medellín
Medellín, Colombia
e-mail: fjcaro@udem.edu.co

Abstract: The matrix variate elliptical generalization of [30] is presented in this work. The published Gaussian case is revised and modified. Then, new aspects of identifiability and consistent estimation of mean form and mean form difference are considered under elliptical laws. For example, instead of using the Euclidean distance matrix for the consistent estimates, exact formulae are derived for the moments of the matrix $\mathbf{B} = \mathbf{X}^c \left(\mathbf{X}^c\right)^T$; where $\mathbf{X}^c$ is the centered landmark matrix. Finally, a complete application in Biology is provided; it includes estimation, model selection and hypothesis testing.

MSC 2010 subject classifications: Primary 62E15, 62E05, 62H12; secondary 62H30, 62H35.
Keywords and phrases: Coordinate free approach, non-central singular Pseudo-Wishart distribution, matrix variate elliptical distribution, matrix variate Gaussian distribution, statistical shape theory.

Received August 2015.

1. Introduction

Statistical theory of shape is a versatile technique of classification and comparison of "objects" in several disciplines. It can be set in terms of matrix variate theory, then a plenty of distributional results are available for applications. However, some strong problems appear: the use of asymptotic distributions, tangent plane inference, isotropic models, Gaussian assumptions, etc.; see [16]. In particular, the implementation of procrustes theory has received critics from experts on morphometrics and related fields, see [30].

Recently, the so called generalized shape theory has emerged as a robust alternative for the addressed disadvantages. It is based on exact shape densities indexed by families of distributions with elliptical contours. The shape densities are given in terms of series of Jack polynomials, which are eigenfunctions of the Laplace-Beltrami operator, [40]. Single polynomial can be calculated by the algorithm of [27], but the series involve difficult problems. The generalized shape theory has reached a moderate computational success in the following contexts: via QR decomposition ([22], [10]), singular value decompositions ([28], [21], [14], [7], [8]), affine ([22], [14], [4], [3], [5]), and Pseudo-Wishart ([9]).

Each approach models the shape of the objects under geometrical transformations. The methods give some invariants of interest for applications. For example: 1. affine transformation removes any geometrical information of rotation, translation, scaling and uniform shear; 2. similarity transformations via QR, SVD or Pseudo-Wishart remove rotation, translation and scaling. 3. projective shape removes affine effect and projection.

About the computation in the likelihood estimation, some affine densities can be reduced to polynomials of low degree; then the associated inference gives robust estimation of location and scale population parameters.

However, likelihood estimation with similarity shape distributions is a difficult task, because it demands the computation of infinite series of Jack polynomials. A common computational practice considers isotropic models, but this assumption is unrealistic in applications. Users of Euclidean shape theory expect estimation of the correlation structure of the landmarks, in order to provide a full description of the objects, see [32].

Instead of the likelihood approach, some authors have proposed the method-of-moments estimators under Gaussian models. The technique provides a computable and consistent estimation of mean form and variance-covariance structure, see [30], [31], [39] and the references therein. In fact, [30] showed that the addressed procrustes analysis yield inconsistent estimators of mean form as well as shape. He also proved that variance-covariance parameters are nonidentifiable under this analysis. [41] also reported the inability of procrustes methods to estimate the correct variance-covariance structure. It is important to note that procrustes analysis is a widely used method for shape estimation in several fields, see [16].

Thus the method-of-moments estimators can be considered as a promising technique in shape theory. Some studies can include: a revision of the Gaussian case given in [30]; a generalization in the context of elliptical laws, model selection criteria and shape hypothesis testing.

With the modified Gaussian case of [30], we can establish a connection with the theoretical studies of [33], [36] and [11]. And then we can provide a unified approach in the general framework of the matrix variate elliptical shape theory.

The above discussion is placed in this work as follows: Section 2 clarifies some results of the published Gaussian case and proposes a generalization in the context of matrix variate elliptical distributions. The section includes: the identifiability and estimability of the parameters; the perturbation model under a matrix variate elliptical distribution; the invariance and the nuisance parameters. Section 3 studies the consistent estimation of the population parameters under dependence and independence. It also gives exact formulae for the moments estimators. Section 4 provides a consistent estimation for a general non-negative definite correlation matrix. Section 5 gives extensions of form difference under the Euclidean Distance Matrix and elliptical models. Finally, Section 6 gives a complete example with the main results of the paper. It also proposes a model selection criteria.

2. Preliminary results

In this section we review some Gaussian distributional results of [30]. Then the statistical model of the paper is proposed under a general matrix variate elliptical distribution.

2.1. Matrix variate elliptical distribution

A detailed discussion of matrix variate elliptical distributions can be found in [19] and [23].

Remark 2.1. *There are two definitions for a matrix variate Gaussian distribution. The first one is written as*

$$\mathbf{Y} \sim \mathcal{N}_{K\times D}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\Theta}),$$

see [1], [18] and [19]. In general, the estimation of $\boldsymbol{\Sigma}$ *or* $\boldsymbol{\Theta}$ *is not possible, but the Kronecker product* $\boldsymbol{\Sigma}\otimes\boldsymbol{\Theta}$ *(or* $\boldsymbol{\Theta}\otimes\boldsymbol{\Sigma}$*) is identifiable, see [30], [18]. Then, given that* $\operatorname{Cov}(\operatorname{vec}\mathbf{Y}) = \boldsymbol{\Theta}\otimes\boldsymbol{\Sigma}$*, and* $\operatorname{Cov}(\operatorname{vec}\mathbf{Y}^T) = \boldsymbol{\Sigma}\otimes\boldsymbol{\Theta}$*, a number of authors use the alternative notation,*

$$\mathbf{Y} \sim \mathcal{N}_{K\times D}(\boldsymbol{\mu}, \boldsymbol{\Sigma}\otimes\boldsymbol{\Theta}),$$

where "vec" denotes the vectorization operator, see [36] and [23]. The same situation appears in the matrix variate elliptical case; in this paper we will follow the second notation.

Definition 2.1. *The* $K\times D$ *random matrix* $\mathbf{Y}$ *is said to have a matrix variate elliptical distribution, with location parameter* $\boldsymbol{\mu}\in\Re^{K\times D}$ *and scale parameter* $\boldsymbol{\Sigma}\otimes\boldsymbol{\Theta}\in\Re^{KD\times KD}$*, if its density function with respect to the Lebesgue measure* $(d\mathbf{Y})$ *is given by*

$$dF_{\mathbf{Y}}(\mathbf{Y}) = |\boldsymbol{\Sigma}|^{-D/2}|\boldsymbol{\Theta}|^{-K/2}h[\operatorname{tr}\boldsymbol{\Theta}^{-1}(\mathbf{Y}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\mathbf{Y}-\boldsymbol{\mu})](d\mathbf{Y}). \tag{2.1}$$

Here $\boldsymbol{\Sigma}\in\Re^{K\times K}$ *and* $\boldsymbol{\Theta}\in\Re^{D\times D}$ *are positive definite matrices, denoted by* $\boldsymbol{\Sigma}>0$ *and* $\boldsymbol{\Theta}>0$*. The function* $h:\Re\to[0,\infty)$ *is termed the density generator and it is required that* $\int_0^\infty u^{KD/2-1}h(u)du<\infty$*. We collect the above definition in the notation* $\mathbf{Y}\sim\mathcal{E}_{K\times D}(\boldsymbol{\mu},\boldsymbol{\Sigma}\otimes\boldsymbol{\Theta},h)$*. Finally, the characteristic function* $\psi_{\mathbf{Y}}(\mathbf{T})$ *of* $\mathbf{Y}$ *is given by*

$$\psi_{\mathbf{Y}}(\mathbf{T}) = \operatorname{etr}(i\boldsymbol{\mu}^T\mathbf{T})\phi(\operatorname{tr}\mathbf{T}\boldsymbol{\Theta}\mathbf{T}^T\boldsymbol{\Sigma}), \tag{2.2}$$

with $i = \sqrt{-1}$*,* $\phi : [0, \infty) \to \Re$ *and* $\operatorname{etr}(\cdot) = \exp(\operatorname{tr}(\cdot))$*. Note that* $\psi_{\mathbf{Y}}$ *also exists when* $\mathbf{\Sigma}$ *and/or* $\mathbf{\Theta}$ *are semidefinite positive matrices; in such case* $\mathbf{Y}$ *is said to have a singular matrix multivariate elliptical distribution, see Remark 2.2.*

Moreover, observe that $\operatorname{Cov}(\operatorname{vec}\mathbf{Y}) = c_0 \mathbf{\Theta} \otimes \mathbf{\Sigma}$*, and* $\operatorname{Cov}(\operatorname{vec}\mathbf{Y}^T) = c_0 \mathbf{\Sigma} \otimes \mathbf{\Theta}$ *where* $c_0 = -2\phi'(0)$ *and*

$$\phi'(0) = \left.\frac{d\phi(t^2)}{dt}\right|_{t=0};$$

see [19, Theorm 2.6.5, p. 62] and [23, Corollary 3.2.1.1, p. 94 and Theorem 2.4.1, p. 33].

Now, if

$$\mathbf{Y} = \begin{pmatrix} \mathbf{Y}_{(1)}^T \\ \mathbf{Y}_{(2)}^T \\ \vdots \\ \mathbf{Y}_{(k)}^T \end{pmatrix} = (\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_D), \quad \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_{(1)}^T \\ \boldsymbol{\mu}_{(2)}^T \\ \vdots \\ \boldsymbol{\mu}_{(k)}^T \end{pmatrix} = (\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \dots, \boldsymbol{\mu}_D),$$

$$\mathbf{\Sigma} = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1K} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{K1} & \sigma_{K2} & \cdots & \sigma_{KK} \end{pmatrix} \text{ and } \mathbf{\Theta} = \begin{pmatrix} \theta_{11} & \theta_{12} & \cdots & \theta_{1D} \\ \theta_{21} & \theta_{22} & \cdots & \theta_{2D} \\ \vdots & \vdots & \ddots & \vdots \\ \theta_{D1} & \theta_{D2} & \cdots & \theta_{kD} \end{pmatrix},$$

then it is easy to check that

1. $\mathbf{Y}_{(i)} \sim \mathcal{E}_D(\boldsymbol{\mu}_{(i)}, \sigma_{ii}\mathbf{\Theta}, h), \quad i = 1, 2, \dots, K,$
2. $\mathbf{Y}_j \sim \mathcal{E}_K(\boldsymbol{\mu}_j, \theta_{jj}\mathbf{\Sigma}, h), \quad j = 1, 2, \dots, D,$

see [19]. Thus,

1. $\operatorname{Cov}(\mathbf{Y}_{(i)}) = c_0 \sigma_{ii}\mathbf{\Theta}, \quad i = 1, 2, \dots, K,$
2. $\operatorname{Cov}(\mathbf{Y}_j) = c_0 \theta_{jj}\mathbf{\Sigma}, \quad j = 1, 2, \dots, D.$

We must point out that the last two asseverations are incorrectly stated in [30] under the context of a perturbation model.

Finally, note that the matrix variate elliptical distributions generalize the Gaussian case. They include the contaminated Gaussian, Pearson type II and VII, Kotz, Jensen-Logistic, power exponential and Bessel distributions, among others; and these distributions have tails that are more or less weighted, and/or present a greater or smaller degree of kurtosis than the Gaussian distribution.

2.2. Identifiability and estimability of the parameters

In this section we study the identifiability and estimability of the parameters $\boldsymbol{\mu}$ and $\mathbf{\Sigma} \otimes \mathbf{\Theta}$.

First observe that the density (2.1) can be written as

$$dF_{\operatorname{vec}\mathbf{Y}^T}(\operatorname{vec}\mathbf{Y}^T) = |\mathbf{\Sigma} \otimes \mathbf{\Theta}|^{-1/2} h[\operatorname{vec}^T(\mathbf{Y} - \boldsymbol{\mu})(\mathbf{\Sigma} \otimes \mathbf{\Theta})^{-1}\operatorname{vec}(\mathbf{Y} - \boldsymbol{\mu})](d\operatorname{vec}\mathbf{Y}^T), \tag{2.3}$$

see [36, p. 79] and [23, Theorem 2.1.1, p. 20]. Here we have used the fact that $\operatorname{vec}^T \mathbf{X}(\mathbf{DB} \otimes \mathbf{C}^T) \operatorname{vec} \mathbf{X} = \operatorname{tr}(\mathbf{BX}^T\mathbf{CXD})$ and $|\mathbf{A}|^m|\mathbf{B}|^n = |\mathbf{A} \otimes \mathbf{B}|$, with $\operatorname{vec}^T \mathbf{X} \equiv (\operatorname{vec} \mathbf{X})^T$, $\mathbf{A} \in \Re^{n\times n}$ and $\mathbf{B} \in \Re^{m\times m}$, see [36, Section 2.2, pp. 72–76] and [19, Section 1.4, pp. 11–13]. Denoting $\operatorname{vec} \mathbf{Y}^T = \mathbf{y} \in \Re^{KD}$ and $\boldsymbol{\Sigma}\otimes\boldsymbol{\Theta} = \boldsymbol{\Xi}$, the density (2.3) provides the distribution of the vector $\mathbf{y}$; moreover, $\mathbf{y} \sim \mathcal{E}_{KD}(\operatorname{vec} \boldsymbol{\mu}, \boldsymbol{\Xi}, h)$.

Now, assume that our data consist of a sample of matrices $\mathbf{Y}_1, \mathbf{Y}_2, \ldots, \mathbf{Y}_n$ from a given population. Define the random matrix

$$\mathbb{Y} = (\operatorname{vec} \mathbf{Y}_1^T, \operatorname{vec} \mathbf{Y}_2^T, \ldots, \operatorname{vec} \mathbf{Y}_n^T)^T \in \Re^{n\times KD},$$

and suppose that $\mathbf{Y}_1, \mathbf{Y}_2, \ldots, \mathbf{Y}_n$ are independent. Then, by [11], the density function of $\mathbb{Y}$ is

$$dF_{\mathbb{Y}}(\mathbb{Y}) = |\boldsymbol{\Sigma} \otimes \boldsymbol{\Theta}|^{-n/2} h[\operatorname{tr}(\mathbb{Y} - \mathbb{M})(\boldsymbol{\Sigma} \otimes \boldsymbol{\Theta})^{-1}(\mathbb{Y} - \mathbb{M})^T](d\mathbb{Y}), \tag{2.4}$$

where

$$\mathbb{M} = \mathbf{1}_n \operatorname{vec}^T \boldsymbol{\mu} \in \Re^{n\times KD},$$

and $\mathbf{1}_n = (1, 1, \ldots, 1)^T \in \Re^n$; i.e. $\mathbb{Y} \sim \mathcal{E}_{n\times KD}(\mathbb{M}, \boldsymbol{\Sigma} \otimes \boldsymbol{\Theta} \otimes \mathbf{I}_n, h)$. Now, consider $p = KD$ in [19, Theorem 4.1.1, p.129], where $KD < n$. Given that $h(\cdot)$ is nonincreasing and continuous, then the *maximum likelihood estimate* of $(\operatorname{vec} \boldsymbol{\mu}, \boldsymbol{\Sigma}\otimes\boldsymbol{\Theta})$ is

$$\left(\widetilde{\operatorname{vec} \boldsymbol{\mu}}, \widetilde{\boldsymbol{\Sigma} \otimes \boldsymbol{\Theta}}\right) = (\bar{\mathbf{y}}, \lambda_{\max}\mathbf{S}).$$

In this case, $\lambda_{\max}$ is the critical point where $h^*(\lambda)$ attains its maximum, and

$$h^*(\lambda) = \lambda^{-KDn/2} h(KD/\lambda).$$

Note that

$$\bar{\mathbf{y}} = \frac{1}{n}\mathbb{Y}^T\mathbf{1}_n \in \Re^{KD}, \text{ and } \mathbf{S} = \mathbb{Y}^T\mathbf{H}_n\mathbb{Y} \in \Re^{KD\times KD},$$

where $\mathbf{H}_n = \mathbf{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^T$ defines an orthogonal projection, i.e., $\mathbf{H}_n = \mathbf{H}_n^T = \mathbf{H}_n^2$. Alternatively

$$\bar{\mathbf{y}} = \frac{1}{n}\sum_{i=1}^{n} \operatorname{vec} \mathbf{Y}_i^T, \text{ and } \mathbf{S} = \sum_{i=1}^{n}(\operatorname{vec} \mathbf{Y}_i^T - \bar{\mathbf{y}})(\operatorname{vec} \mathbf{Y}_i^T - \bar{\mathbf{y}})^T,$$

then the estimator of $\boldsymbol{\mu}$ is

$$\widetilde{\boldsymbol{\mu}} = \bar{\mathbf{Y}} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{Y}_i.$$

At this point, we can apply [19, Section 4.3] and derive classical properties of the maximum likelihood estimators $\widetilde{\boldsymbol{\mu}}$ and $\widetilde{\boldsymbol{\Sigma} \otimes \boldsymbol{\Theta}}$. Some characteristics are:

sufficiency, completeness, consistency and unbiasedness. For example, if $h(\cdot)$ is nonincreasing and continuous, and $\mathbb{Y}$ has a finite 2nd moment, then

$$\widehat{\boldsymbol{\mu}} = \bar{\mathbf{Y}} \quad \text{and} \quad \widehat{\boldsymbol{\Sigma} \otimes \boldsymbol{\Theta}} = \frac{1}{2(1-n)\psi'(0)}\mathbf{S},$$

are *unbiased estimators* of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma} \otimes \boldsymbol{\Theta}$.

Remark 2.2. *When the columns and/or rows of* $\mathbf{Y} \sim \mathcal{E}_{K\times D}(\boldsymbol{\mu}, \boldsymbol{\Sigma} \otimes \boldsymbol{\Theta}, h)$ *are linearly dependent, the matrix* $\mathbf{Y}$ *is said to have a singular matrix variate elliptical distribution. In this case* $\mathbf{Y}$ *has a density respect to the Hausdorff measure. Moreover, such dependence is summarized in the rank of the matrices* $\boldsymbol{\Sigma}$ *and/or* $\boldsymbol{\Theta}$. *This will be denoted by* $\mathbf{Y} \sim \mathcal{E}^{s,r}_{K\times D}(\boldsymbol{\mu}, \boldsymbol{\Sigma} \otimes \boldsymbol{\Theta}, h)$, *where* $s = \operatorname{rank}(\boldsymbol{\Sigma}) \leq K$ *and* $r = \operatorname{rank}(\boldsymbol{\Theta}) \leq D$, *see [23, Definition 2.1.1, p. 19], [13] and [15]. We also note that the maximum likelihood estimators under singular matrix variate elliptical models follow the same rules of the singular Gaussian case, see [26] and [38, Section 8a.5, pp. 528–532].*

2.3. Perturbation model under a matrix variate elliptical distribution

Assume that the random matrix $\mathbf{X} \in \Re^{K\times D}$ represents a geometrical figure comprising K landmarks in D dimensions, with $K > D$. $\mathbf{X}$ is called the *landmark coordinate matrix*, see [30].

Consider an independent sample of landmark coordinate matrices $\mathbf{X}_i \in \Re^{K\times D}$, $i = 1, 2, \dots, n$, from a given population.

The statistical model of this work is a generalization of the perturbation law used by [30]. If $\boldsymbol{\mu} \in \Re^{K\times D}$ is the corresponding mean form, then we propose the model

$$\mathbf{X}_i = (\boldsymbol{\mu} + \mathbf{E}_i)\boldsymbol{\Gamma}_i + \mathbf{t}_i, \quad , i = 1, 2, \dots, n, \tag{2.5}$$

where $\mathbf{E}_i \sim \mathcal{E}_{K\times D}(\mathbf{0}, \boldsymbol{\Sigma}_K \otimes \boldsymbol{\Sigma}_D, h)$. The orthogonal matrices $\boldsymbol{\Gamma}_i \in \Re^{D\times D}$ are rotation and/or reflection of $(\boldsymbol{\mu} + \mathbf{E}_i)$. Meanwhile, the matrices $\mathbf{t}_i \in \Re^{K\times D}$ ($\mathbf{t}_i = \mathbf{1}_k \mathbf{a}_i^T$) represent translations with some $\mathbf{a}_i \in \Re^D$. From [19, eq. (3.3.10), p. 103] or [23, Theorem 2.1.2, p. 20] we have that

$$\mathbf{X}_i \sim \mathcal{E}_{K\times D}(\boldsymbol{\mu}\boldsymbol{\Gamma}_i + \mathbf{t}_i, \boldsymbol{\Sigma}_K \otimes \boldsymbol{\Gamma}_i^T \boldsymbol{\Sigma}_D \boldsymbol{\Gamma}_i, h), \quad i = 1, 2, \dots, n. \tag{2.6}$$

The parameters of interest are $(\boldsymbol{\mu}, \boldsymbol{\Sigma}_K \otimes \boldsymbol{\Sigma}_D)$ and the nuisance parameters are $(\boldsymbol{\Gamma}_i^T, \mathbf{t}_i)$ $i = 1, 2, ..., n$. A detailed explanation of the corresponding Gaussian perturbation model can be found in [30].

Now, note that

$$\operatorname{vec} \mathbb{X}^T = \operatorname{diag}(\mathbb{G}) \operatorname{vec}(\mathbb{M} + \mathbb{E})^T + \operatorname{vec} \mathbb{T}^T,$$

with

$$\operatorname{diag}(\mathbb{G}) = \begin{pmatrix} \mathbf{I}_K \otimes \boldsymbol{\Gamma}_1^T & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_K \otimes \boldsymbol{\Gamma}_2^T & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{I}_K \otimes \boldsymbol{\Gamma}_n^T \end{pmatrix},$$

then the model (2.5) can be written in the form

$$\mathbb{X} = \operatorname{diag}(\mathbb{M} + \mathbb{E})\mathbb{G}^T + \mathbb{T}.$$

Where

$$\operatorname{diag}(\mathbb{M}+\mathbb{E}) = \begin{pmatrix} \operatorname{vec}^T(\boldsymbol{\mu}+\mathbf{E}_1)^T & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \operatorname{vec}^T(\boldsymbol{\mu}+\mathbf{E}_2)^T & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \operatorname{vec}^T(\boldsymbol{\mu}+\mathbf{E}_n)^T \end{pmatrix},$$

$$\mathbb{X} = \begin{pmatrix} \operatorname{vec}^T \mathbf{X}_1^T \\ \operatorname{vec}^T \mathbf{X}_2^T \\ \vdots \\ \operatorname{vec}^T \mathbf{X}_n^T \end{pmatrix}, \ \mathbb{M} = \mathbf{1}_n \operatorname{vec}^T \boldsymbol{\mu}^T, \ \mathbb{E} = \begin{pmatrix} \operatorname{vec}^T \mathbf{E}_1^T \\ \operatorname{vec}^T \mathbf{E}_2^T \\ \vdots \\ \operatorname{vec}^T \mathbf{E}_n^T \end{pmatrix} \ \mathbb{T} = \begin{pmatrix} \operatorname{vec}^T \mathbf{t}_1^T \\ \operatorname{vec}^T \mathbf{t}_2^T \\ \vdots \\ \operatorname{vec}^T \mathbf{t}_n^T \end{pmatrix},$$

and $\mathbb{G} = \left(\mathbf{I}_K \otimes \boldsymbol{\Gamma}_1^T | \mathbf{I}_K \otimes \boldsymbol{\Gamma}_2^T | \cdots | \mathbf{I}_K \otimes \boldsymbol{\Gamma}_n^T\right)$. Observe also that

$$\mathbb{E} \sim \mathcal{E}_{n\times KD}(\mathbf{0}, \mathbf{I}_n \otimes \boldsymbol{\Sigma}_K \otimes \boldsymbol{\Sigma}_D, h),$$

or

$$\operatorname{vec} \mathbb{E}^T \sim \mathcal{E}_{nKD}(\operatorname{vec} \mathbf{0}, \mathbf{I}_n \otimes \boldsymbol{\Sigma}_K \otimes \boldsymbol{\Sigma}_D, h).$$

Hence

$$\operatorname{vec} \mathbb{X}^T \sim \mathcal{E}_{nKD}\left(\operatorname{diag}(\mathbb{G}) \operatorname{vec} \mathbb{M}^T + \operatorname{vec} \mathbb{T}^T, \operatorname{diag}(\mathbb{G})(\mathbf{I}_n \otimes \boldsymbol{\Sigma}_K \otimes \boldsymbol{\Sigma}_D) \operatorname{diag}(\mathbb{G})^T, h\right).$$

Recall that $\operatorname{vec}(\mathbf{y}\mathbf{x}^T) = \mathbf{x} \otimes \mathbf{y}$, for vectors $\mathbf{x}$ and $\mathbf{y}$, then

$$\begin{aligned} \operatorname{diag}(\mathbb{G}) \operatorname{vec} \mathbb{M}^T &= \begin{pmatrix} \mathbf{I}_K \otimes \boldsymbol{\Gamma}_1^T & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_K \otimes \boldsymbol{\Gamma}_2^T & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{I}_K \otimes \boldsymbol{\Gamma}_n^T \end{pmatrix} \left(\mathbf{1}_n \otimes \operatorname{vec} \boldsymbol{\mu}^T\right) \\ &= \begin{pmatrix} (\mathbf{I}_K \otimes \boldsymbol{\Gamma}_1^T) \operatorname{vec} \boldsymbol{\mu}^T \\ (\mathbf{I}_K \otimes \boldsymbol{\Gamma}_2^T) \operatorname{vec} \boldsymbol{\mu}^T \\ \vdots \\ (\mathbf{I}_K \otimes \boldsymbol{\Gamma}_n^T) \operatorname{vec} \boldsymbol{\mu}^T \end{pmatrix} \\ &= \begin{pmatrix} \operatorname{vec}(\boldsymbol{\mu}\boldsymbol{\Gamma}_1)^T \\ \operatorname{vec}(\boldsymbol{\mu}\boldsymbol{\Gamma}_2)^T \\ \vdots \\ \operatorname{vec}(\boldsymbol{\mu}\boldsymbol{\Gamma}_n)^T \end{pmatrix}. \end{aligned}$$

So, $\operatorname{diag}(\mathbb{G})(\mathbf{I}_n \otimes \boldsymbol{\Sigma}_K \otimes \boldsymbol{\Sigma}_D) \operatorname{diag}(\mathbb{G})^T$ is

$$= \begin{pmatrix} \boldsymbol{\Sigma}_K \otimes \boldsymbol{\Gamma}_1^T \boldsymbol{\Sigma}_D \boldsymbol{\Gamma}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_K \otimes \boldsymbol{\Gamma}_2^T \boldsymbol{\Sigma}_D \boldsymbol{\Gamma}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \boldsymbol{\Sigma}_K \otimes \boldsymbol{\Gamma}_n^T \boldsymbol{\Sigma}_D \boldsymbol{\Gamma}_n \end{pmatrix}$$

$$= \sum_{i=1}^{n} \mathbf{E}_{ii}^n \otimes \boldsymbol{\Sigma}_K \otimes \boldsymbol{\Gamma}_i^T \boldsymbol{\Sigma}_D \boldsymbol{\Gamma}_i,$$

where $\mathbf{E}_{ii}^n = \mathbf{e}_i^n(\mathbf{e}_i^n)^T$ and $\mathbf{e}_i^n$ is the ith column unit vector of order n.

Finally, observe that

$$E(\operatorname{vec} \mathbb{X}^T) = \begin{pmatrix} \operatorname{vec}(\boldsymbol{\mu}\boldsymbol{\Gamma}_1)^T \\ \operatorname{vec}(\boldsymbol{\mu}\boldsymbol{\Gamma}_2)^T \\ \vdots \\ \operatorname{vec}(\boldsymbol{\mu}\boldsymbol{\Gamma}_n)^T \end{pmatrix} + \operatorname{vec} \mathbb{T}^T = \sum_{i=1}^{n} \mathbf{e}_i^n \otimes \left(\operatorname{vec}(\boldsymbol{\mu}\boldsymbol{\Gamma}_i)^T + \operatorname{vec} \mathbf{t}_i^T\right),$$

and

$$E(\mathbb{X}) = \sum_{i=1}^{n} \mathbf{e}_i^n \left(\operatorname{vec}(\boldsymbol{\mu}\boldsymbol{\Gamma}_i)^T + \operatorname{vec} \mathbf{t}_i^T\right)^T,$$

therefore

$$\mathbb{X} \sim \mathcal{E}_{n \times KD}\left(\sum_{i=1}^{n} \mathbf{e}_i^n \left(\operatorname{vec}(\boldsymbol{\mu}\boldsymbol{\Gamma}_i)^T + \operatorname{vec} \mathbf{t}_i^T\right)^T, \sum_{i=1}^{n} \mathbf{E}_{ii}^n \otimes \boldsymbol{\Sigma}_K \otimes \boldsymbol{\Gamma}_i^T \boldsymbol{\Sigma}_D \boldsymbol{\Gamma}_i, h\right).$$

2.4. Invariance and nuisance parameters

A first analysis of the elliptical perturbation law involves the nuisance parameters. As in the Gaussian case of [30], we can remove the nuisance parameters by using a simple transformation.

From (2.6)

$$\mathbf{X}_i \sim \mathcal{E}_{K \times D}(\boldsymbol{\mu}\boldsymbol{\Gamma}_i + \mathbf{t}_i, \boldsymbol{\Sigma}_K \otimes \boldsymbol{\Gamma}_i^T \boldsymbol{\Sigma}_D \boldsymbol{\Gamma}_i, h), \quad i = 1, 2, \ldots, n.$$

Setting $\mathbf{X}_i^c = \mathbf{H}_K \mathbf{X}_i$ and using $\mathbf{H}_K \mathbf{1}_K = \mathbf{0}_k$ and $\mathbf{1}_K^T \mathbf{H}_K = \mathbf{0}_k^T$, we obtain

$$\mathbf{X}_i^c \sim \mathcal{E}_{K \times D}^{(K-1),D}(\boldsymbol{\mu}^* \boldsymbol{\Gamma}_i, \boldsymbol{\Sigma}_K^* \otimes \boldsymbol{\Gamma}_i^T \boldsymbol{\Sigma}_D \boldsymbol{\Gamma}_i, h), \quad i = 1, 2, \ldots, n, \tag{2.7}$$

where $\boldsymbol{\mu}^* = \mathbf{H}_K \boldsymbol{\mu}$ and $\boldsymbol{\Sigma}_K^* = \mathbf{H}_K \boldsymbol{\Sigma}_K \mathbf{H}_K$. Note that $\mathbf{H}_K \mathbf{t}_i = \mathbf{H}_K \mathbf{1}_K \mathbf{a}_i^T = \mathbf{0}$ for all $i = 1, 2, \ldots, n$, then the summation of the columns of $\boldsymbol{\mu}^*$ is zero, which means that it is a centered matrix.

Recall that $K > D$ and $\operatorname{rank}(\boldsymbol{\Sigma}_K^*) = K - 1$, then by [13] and [15] we get

$$\mathbf{B}_i = \mathbf{X}_i^c (\boldsymbol{\Gamma}_i^T \boldsymbol{\Sigma}_D \boldsymbol{\Gamma}_i)^{-1} (\mathbf{X}_i^c)^T \sim \mathcal{GPW}_K^q(D, \boldsymbol{\Sigma}_K^*, \boldsymbol{\Sigma}_D, \boldsymbol{\Omega}, h), \quad i = 1, 2, \ldots, n. \tag{2.8}$$

where

$$\mathbf{\Omega} = (\mathbf{\Sigma}_K^*)^- \boldsymbol{\mu}^* \mathbf{\Gamma}_i (\mathbf{\Gamma}_i^T \mathbf{\Sigma}_D \mathbf{\Gamma}_i)^{-1} \mathbf{\Gamma}_i^T (\boldsymbol{\mu}^*)^T = (\mathbf{\Sigma}_K^*)^- \boldsymbol{\mu}^* \mathbf{\Sigma}_D^{-1} (\boldsymbol{\mu}^*)^T,$$

$q = \min((K-1), D)$ and $\mathbf{A}^-$ is any symmetric generalized inverse of $\mathbf{A}$ such that $\mathbf{A}\mathbf{A}^-\mathbf{A} = \mathbf{A} = \mathbf{A}^T$. By definition, $\mathbf{B}_i$ is said to have a *generalized singular pseudo-Wishart distribution*, which is independent of the nuisance parameters.

Remark 2.3. *Observe that* $\mathbf{B}_i$ *can be written as*

$$\mathbf{B}_i = \mathbf{X}_i^c (\mathbf{\Gamma}_i^T \mathbf{\Sigma}_D \mathbf{\Gamma}_i)^{-1} (\mathbf{X}_i^c)^T = \mathbf{X}_i^c \mathbf{\Gamma}_i^T \mathbf{\Sigma}_D^{-1} \mathbf{\Gamma}_i (\mathbf{X}_i^c)^T = \mathbf{Y}_i \mathbf{\Sigma}_D^{-1} \mathbf{Y}_i^T$$

where $\mathbf{Y}_i = \mathbf{X}_i^c \mathbf{\Gamma}_i^T$ *and*

$$\mathbf{Y}_i \sim \mathcal{E}_{K\times D}^{(K-1),D}(\boldsymbol{\mu}^*, \mathbf{\Sigma}_K^* \otimes \mathbf{\Sigma}_D, h), \quad i = 1, 2, \dots, n.$$

In particular, if $\mathbf{\Sigma}_D = \mathbf{I}_D$ and

$$\mathbf{X}_i^c = (\mathbf{X}_{1,i}^c | \mathbf{X}_{2,i}^c | \cdots | \mathbf{X}_{D,i}^c),$$

with

$$\mathbf{X}_{d,i}^c \sim \mathcal{E}_K^{(K-1)}(\boldsymbol{\mu}^* \mathbf{\Gamma}_i \mathbf{e}_d^K, \mathbf{\Sigma}_K^*, h), \quad d = 1, 2, \dots, D; \; i = 1, 2, \dots, n,$$

then we have,

$$\mathbf{B}_i = \mathbf{X}_i^c (\mathbf{X}_i^c)^T = \sum_{d=1}^{D} \mathbf{X}_{d,i}^c (\mathbf{X}_{d,i}^c)^T;$$

furthermore,

$$\mathbf{B}_i \sim \mathcal{GPW}_K^q(D, \mathbf{\Sigma}_K^*, \mathbf{I}_D, \mathbf{\Omega}, h), \quad i = 1, 2, \dots, n, \tag{2.9}$$

where $\mathbf{\Omega} = (\mathbf{\Sigma}_K^*)^- \boldsymbol{\mu}^* (\boldsymbol{\mu}^*)^T$.

Remark 2.4. *The corresponding result in [30] is obtained as a particular case of (2.9); just note that* $\boldsymbol{\mu}^* (\boldsymbol{\mu}^*)^T$ *is the noncentrality parameter of [30] and we have used* $\mathbf{\Omega} = (\mathbf{\Sigma}_K^*)^- \boldsymbol{\mu}^* (\boldsymbol{\mu}^*)^T$*, see [36, Definition 10.3.1, pp. 441–442].*

In addition, defining $\mathbb{X}^c$ as $\mathbb{X}$, we get

$$\operatorname{vec}(\mathbb{X}^c)^T = [\mathbf{I}_n \otimes (\mathbf{H}_k \otimes \mathbf{I}_D)] \operatorname{vec} \mathbb{X}^T.$$

Hence, $\mathbb{X}^c = \mathbb{X}(\mathbf{H}_k \otimes \mathbf{I}_D)$. Now, observing that $(\mathbf{H}_k \otimes \mathbf{I}_D) \operatorname{vec} \mathbf{t}_i^T = \mathbf{0}$, for all $i = 1, 2, \dots, n$, we obtain

$$\mathbb{X}^c \sim \mathcal{E}_{n\times KD}^{n,(K-1)D}\left(\sum_{i=1}^{n} \mathbf{e}_i^n \operatorname{vec}^T(\boldsymbol{\mu}^* \mathbf{\Gamma}_i)^T, \sum_{i=1}^{n} \mathbf{E}_{ii}^n \otimes \mathbf{\Sigma}_K^* \otimes \mathbf{\Gamma}_i^T \mathbf{\Sigma}_D \mathbf{\Gamma}_i, h\right), \tag{2.10}$$

where $\mathbf{\Sigma}_K^* = \mathbf{H}_k \mathbf{\Sigma}_K \mathbf{H}_k$.

In the same way of [30], assuming $\mathbf{\Sigma}_D = \mathbf{I}_D$ and using

$$\mathbf{I}_n = \sum_{i=1}^{n} \mathbf{E}_{ii}^n,$$

we have

$$\mathbb{X}^c \sim \mathcal{E}_{n\times KD}^{n,(K-1)D}\left(\sum_{i=1}^{n} \mathbf{e}_i^n \operatorname{vec}^T(\boldsymbol{\mu}^*\mathbf{\Gamma}_i)^T, \mathbf{I}_n \otimes \mathbf{\Sigma}_K^* \otimes \mathbf{I}_D, h\right).$$

3. Consistent estimation of $\boldsymbol{\mu}$ and $\mathbf{\Sigma}_K$

The Gaussian case of [30] studied the consistent estimation with the Euclidean distance matrix, in our elliptical setting we can go further and use the exact expressions for the first two moments of the matrix $\mathbf{B}$.

Behind the common assumption of independent landmarks along the D axes, we are formally considering $\mathbf{\Sigma}_D = \mathbf{I}_D$ in the context of a matrix variate Gaussian model. However, in the elliptical case the perspective is wider and we have two possible scenarios:

1. Independence and non correlated landmarks
2. Probabilistic dependence and non correlated landmarks.

In both cases $\mathbf{\Sigma}_D = \mathbf{I}_D$, but the moments of the matrix $\mathbf{B}$ are different.

Remark 3.1. *In the context of matrix variate elliptical theory, the concepts of independence and non correlation are only equivalent in the Gaussian case. Suppose that the vector* $\mathbf{Z} = (z_1, z_2)^T$ *has a bi-dimensional elliptical distribution and* $\operatorname{Cov}(\mathbf{Z}) = \mathbf{I}_2$*, then* z_1 *and* z_2 *are independent if and only if* $\mathbf{Z}$ *has a bi-dimensional Gaussian distribution. But, if* z_1 *and* z_2 *have a one-dimensional elliptical distribution with* $\operatorname{Var}(z_i) = 1$ *and* $\operatorname{Cov}(z_1, z_2) = 0$*, then they are uncorrelated and can be considered independent, see [23, Section 6.2, p. 1] and [20, Section 4.3, p. 105].*

In summary, we will find the first two moments of

$$\mathbf{B} = \mathbf{Y}\mathbf{Y}^T = \sum_{d=1}^{D} \mathbf{y}_d \mathbf{y}_d^T, \tag{3.1}$$

where $\mathbf{\Sigma}_D = \mathbf{I}_D$. This includes two cases: a) the $\mathbf{y}_d$'s are independent and uncorrelated; or b) the $\mathbf{y}_d$'s are dependent and uncorrelated.

3.1. Moments of B *under dependence*

First assume that $\mathbf{\Sigma}_D \neq \mathbf{I}_D$.

Suppose that $\mathbf{Y} \sim \mathcal{E}_{K\times D}^{(K-1),D}(\boldsymbol{\mu}, \mathbf{\Sigma} \otimes \mathbf{\Theta}, h)$, where

$$\mathbf{Y} = (\mathbf{y}_1|\mathbf{y}_2|\cdots|\mathbf{y}_D) \text{ and } \boldsymbol{\mu} = (\boldsymbol{\mu}_1|\boldsymbol{\mu}_2|\cdots|\boldsymbol{\mu}_D),$$

$\mathbf{\Sigma}_D = \mathbf{\Theta}$, $\mathbf{\Sigma}_K^* = \mathbf{\Sigma}$ and $\boldsymbol{\mu}^* = \boldsymbol{\mu}$.

Now, for $\mathbf{x}, \mathbf{y} \in \Re^n$ we know that $\operatorname{vec} \mathbf{x}\mathbf{y}^T = \mathbf{y} \otimes \mathbf{x}$ and $\mathbf{x}\mathbf{y}^T = \mathbf{x} \otimes \mathbf{y}^T = \mathbf{y}^T \otimes \mathbf{x}$, see [33]. Therefore $\operatorname{vec} \mathbf{y}\mathbf{y}^T \operatorname{vec}^T \mathbf{y}\mathbf{y}^T = \mathbf{y} \otimes \mathbf{y}^T \otimes \mathbf{y} \otimes \mathbf{y}^T$; this property will be useful in the following result.

Theorem 3.1. *Let* $\mathbf{Y} \sim \mathcal{E}_{K\times D}^{(K-1),D}(\boldsymbol{\mu}, \mathbf{\Sigma} \otimes \mathbf{\Theta}, h)$. *Then*

1. $E(\operatorname{vec} \mathbf{Y} \operatorname{vec}^T \mathbf{Y}) = c_0(\mathbf{\Theta} \otimes \mathbf{\Sigma}) + \operatorname{vec} \boldsymbol{\mu} \operatorname{vec}^T \boldsymbol{\mu}$, *and*
2. $E(\operatorname{vec} \mathbf{Y} \operatorname{vec} \mathbf{Y}^T \otimes \operatorname{vec} \mathbf{Y} \operatorname{vec} \mathbf{Y}^T)$

$$\begin{aligned}
&= \kappa_0[(\mathbf{I}_{(KD)^2} + \mathbf{K}_{KD})(\mathbf{\Theta} \otimes \mathbf{\Sigma} \otimes \mathbf{\Theta} \otimes \mathbf{\Sigma}) + \operatorname{vec}(\mathbf{\Theta} \otimes \mathbf{\Sigma}) \operatorname{vec}^T(\mathbf{\Theta} \otimes \mathbf{\Sigma})] \\
&\quad + c_0(\mathbf{I}_{K^2} + \mathbf{K}_K)[\operatorname{vec} \boldsymbol{\mu} \operatorname{vec}^T \boldsymbol{\mu} \otimes (\mathbf{\Theta} \otimes \mathbf{\Sigma}) + (\mathbf{\Theta} \otimes \mathbf{\Sigma}) \otimes \operatorname{vec} \boldsymbol{\mu} \operatorname{vec}^T \boldsymbol{\mu}] \\
&\quad + c_0[\operatorname{vec}(\mathbf{\Theta} \otimes \mathbf{\Sigma})(\operatorname{vec}^T \boldsymbol{\mu}\boldsymbol{\mu}^T) + (\operatorname{vec}^T \boldsymbol{\mu}\boldsymbol{\mu}^T) \operatorname{vec}(\mathbf{\Theta} \otimes \mathbf{\Sigma})] \\
&\quad + \operatorname{vec} \boldsymbol{\mu} \operatorname{vec}^T \boldsymbol{\mu} \otimes \operatorname{vec} \boldsymbol{\mu} \operatorname{vec}^T \boldsymbol{\mu},
\end{aligned}$$

where $\mathbf{K}_{KD}$ *is the commutation matrix defined in [33].* $\psi_U(t) = \phi(t^2)$ *is the characteristic function of the univariate elliptical distribution,* $c_0 = E(u^2)$ *and* $3\kappa_0 = E(u^4)$, *with*

$$E(u^2) = \frac{1}{i^2} \frac{d^2 \psi_U(t)}{dt^2}\bigg|_{t=0} \text{ and } E(u^4) = \frac{1}{i^4} \frac{d^4 \psi_U(t)}{dt^4}\bigg|_{t=0},$$

see [23, p. 127].

Some particular values of c_0 *and* κ_0 *are given in Table 1.*

Proof. Differentiating (2.2) and using [12], we get

$$\begin{aligned}
E(\operatorname{vec} \mathbf{Y} \otimes \operatorname{vec}^T \mathbf{Y}) &= E(\operatorname{vec} \mathbf{Y} \operatorname{vec}^T \mathbf{Y}) \\
&= \frac{1}{i^2} \frac{\partial^2 \psi_{\operatorname{vec} \mathbf{Y}}(\operatorname{vec} \mathbf{T})}{\partial \operatorname{vec} \mathbf{T} \partial \operatorname{vec} \mathbf{T}^T}\bigg|_{\operatorname{vec} \mathbf{T}=0}
\end{aligned}$$

and

$$\begin{aligned}
&E(\operatorname{vec} \mathbf{Y} \otimes \operatorname{vec} \mathbf{Y}^T \otimes \operatorname{vec} \mathbf{Y} \otimes \operatorname{vec} \mathbf{Y}^T) \\
&\quad = E(\operatorname{vec} \mathbf{Y} \operatorname{vec} \mathbf{Y}^T \otimes \operatorname{vec} \mathbf{Y} \operatorname{vec} \mathbf{Y}^T) \\
&\quad = \frac{1}{i^4} \frac{\partial^4 \psi_{\operatorname{vec} \mathbf{Y}}(\operatorname{vec} \mathbf{T})}{\partial \operatorname{vec} \mathbf{T} \partial \operatorname{vec} \mathbf{T}^T \partial \operatorname{vec} \mathbf{T} \partial \operatorname{vec} \mathbf{T}^T}\bigg|_{\operatorname{vec} \mathbf{T}=0}.
\end{aligned}$$

□

Now,

$$\mathbf{B} = \mathbf{Y}\mathbf{Y}^T = \sum_{d=1}^{D} \mathbf{y}_d \mathbf{y}_d^T,$$

thus

$$E(\mathbf{B}) = E\left(\mathbf{Y}\mathbf{Y}^T\right) = E\left(\sum_{d=1}^{D} \mathbf{y}_d \mathbf{y}_d^T\right) = \sum_{d=1}^{D} E\left(\mathbf{y}_d \mathbf{y}_d^T\right).$$

Also note that

$$\operatorname{Cov}(\operatorname{vec} \mathbf{Y}) = E(\operatorname{vec} \mathbf{Y} \operatorname{vec}^T \mathbf{Y}) - E(\operatorname{vec} \mathbf{Y}) E(\operatorname{vec}^T \mathbf{Y}),$$

TABLE 1
Particular values of c_0 and κ_0.

Distribution	c_0	κ_0
Multiuniforme[a]	1	$\frac{1}{3}$
Gaussian[b]	1	1
Kotz[c]	$\dfrac{\Gamma\left[\frac{2N+1}{2s}\right]}{r^{1/s}\Gamma\left[\frac{2N-1}{2s}\right]}$	$\dfrac{\Gamma\left[\frac{2N+3}{2s}\right]}{3r^{2/s}\Gamma\left[\frac{2N-1}{2s}\right]}$
t[d]	$\dfrac{m}{m-2}$	$\dfrac{m^2}{(m-2)(m-4)}$
Pearson Type II[e]	$\dfrac{1}{2m+3}$	$\dfrac{1}{(2m+3)(2m+5)}$
Pearson type VII[f]	$\dfrac{m}{2N-3}$	$\dfrac{m^2}{(2N-3)(2N-5)}$

[a]From [20, Theorem 3.3, p. 72].
[b]From [23, Remark 3.2.2, p. 125].
[c]From [37], where $r, s > 0$ and $2N + 1 > 2$.
[d]From [23, p. 128], or [20, p. 88], where $m > 0$.
[e]From [20, Section 3.4.2, p. 89], where $m > -1$.
[f]From [20, Section 3.3.4, p. 84], where $N > 1/2$, $m > 0$.

where $\mathbf{Y} \in \Re^{K\times D}$. Therefore

$$\begin{aligned}
\operatorname{Cov}(\operatorname{vec}\mathbf{B}) &= \operatorname{Cov}\left(\operatorname{vec}\left(\mathbf{Y}\mathbf{Y}^T\right)\right) = \operatorname{Cov}\left(\sum_{d=1}^{D}\operatorname{vec}\left(\mathbf{y}_d\mathbf{y}_d^T\right)\right)\\
&= E\left[\left(\sum_{d=1}^{D}\operatorname{vec}\left(\mathbf{y}_d\mathbf{y}_d^T\right)\right)\left(\sum_{s=1}^{D}\operatorname{vec}^T\left(\mathbf{y}_s\mathbf{y}_s^T\right)\right)\right]\\
&\quad - E\left(\sum_{d=1}^{D}\operatorname{vec}\left(\mathbf{y}_d\mathbf{y}_d^T\right)\right)E\left(\sum_{d=1}^{D}\operatorname{vec}^T\left(\mathbf{y}_d\mathbf{y}_d^T\right)\right)\\
&= \left[\sum_{d=1}^{D}\sum_{s=1}^{D}E\left(\mathbf{y}_d\mathbf{y}_s^T\otimes\mathbf{y}_d\mathbf{y}_s^T\right)\right]\\
&\quad - \operatorname{vec}\left(\sum_{d=1}^{D}E\left(\mathbf{y}_d\mathbf{y}_d^T\right)\right)\operatorname{vec}^T\left(\sum_{s=1}^{D}E\left(\mathbf{y}_s\mathbf{y}_s^T\right)\right). \qquad (3.2)
\end{aligned}$$

Then, we need to find $E\left(\mathbf{y}_d\mathbf{y}_d^T\right)$ and $E\left(\mathbf{y}_d\mathbf{y}_s^T\otimes\mathbf{y}_d\mathbf{y}_s^T\right)$. These moments are derived in the following result.

Theorem 3.2. *Assume that* $\mathbf{Y} \sim \mathcal{E}_{K\times D}^{(K-1),D}(\boldsymbol{\mu}, \boldsymbol{\Sigma}\otimes\boldsymbol{\Theta}, h)$*, with*

$$\mathbf{Y} = (\mathbf{y}_1|\mathbf{y}_2|\cdots|\mathbf{y}_D), \boldsymbol{\mu} = (\boldsymbol{\mu}_1|\boldsymbol{\mu}_2|\cdots|\boldsymbol{\mu}_D),$$

and $\boldsymbol{\Theta} = (\theta_{ds})$*. Then*

1. $E(\mathbf{y}_d\mathbf{y}_d^T) = c_0\theta_{dd}\mathbf{\Sigma} + \boldsymbol{\mu}_d\boldsymbol{\mu}_d^T$, *and*
2. $$\begin{aligned} E\left(\mathbf{y}_d\mathbf{y}_s^T \otimes \mathbf{y}_d\mathbf{y}_s^T\right) = {} & \kappa_0\theta_{ds}^2[(\mathbf{I}_{K^2} + \mathbf{K}_K)(\mathbf{\Sigma}\otimes\mathbf{\Sigma}) + \operatorname{vec}\mathbf{\Sigma}\operatorname{vec}^T\mathbf{\Sigma} \\ & + c_0\theta_{ds}[(\mathbf{I}_{K^2} + \mathbf{K}_K)(\boldsymbol{\mu}_d\boldsymbol{\mu}_s^T\otimes\mathbf{\Sigma} + \mathbf{\Sigma}\otimes\boldsymbol{\mu}_d\boldsymbol{\mu}_s^T)] \\ & + c_0\theta_{ds}[\operatorname{vec}\mathbf{\Sigma}\operatorname{vec}^T\boldsymbol{\mu}_d\boldsymbol{\mu}_s^T + \operatorname{vec}\boldsymbol{\mu}_d\boldsymbol{\mu}_s^T\operatorname{vec}^T\mathbf{\Sigma}] \\ & + \boldsymbol{\mu}_d\boldsymbol{\mu}_s^T\otimes\boldsymbol{\mu}_d\boldsymbol{\mu}_s^T. \end{aligned}$$

Proof. First note that $\mathbf{y}_d = \mathbf{Y}\mathbf{e}_d^D$, then applying Theorem 3.1 we find

$$\begin{aligned} E(\mathbf{y}_d\mathbf{y}_d^T) &= E(\operatorname{vec}\mathbf{y}_d\operatorname{vec}^T\mathbf{y}_d) = E(\operatorname{vec}\mathbf{Y}\mathbf{e}_d^D\operatorname{vec}^T\mathbf{Y}\mathbf{e}_d^D) \\ &= (\mathbf{e}_d^{D\,T}\otimes\mathbf{I}_K)E(\operatorname{vec}\mathbf{Y}\operatorname{vec}^T\mathbf{Y})(\mathbf{e}_d^D\otimes\mathbf{I}_K) \\ &= (\mathbf{e}_d^{D\,T}\otimes\mathbf{I}_K)(c_0(\mathbf{\Theta}\otimes\mathbf{\Sigma}) + \operatorname{vec}\boldsymbol{\mu}\operatorname{vec}^T\boldsymbol{\mu})(\mathbf{e}_d^D\otimes\mathbf{I}_K) \\ &= c_0\theta_{dd}\mathbf{\Sigma} + \boldsymbol{\mu}_d\boldsymbol{\mu}_d^T. \end{aligned}$$

Where we have used the fact that $\operatorname{vec}\mathbf{ABC} = (\mathbf{C}^T\otimes\mathbf{B})\operatorname{vec}\mathbf{B}$, $a\otimes\mathbf{A} = a\mathbf{A}$ and $(\mathbf{A}\otimes\mathbf{D})(\mathbf{B}\otimes\mathbf{E})(\mathbf{C}\otimes\mathbf{F}) = (\mathbf{ABC}\otimes\mathbf{DEF})$. Similarly,

$$E\left(\mathbf{y}_d\mathbf{y}_s^T\otimes\mathbf{y}_d\mathbf{y}_s^T\right) = \mathbf{R}^T E(\operatorname{vec}\mathbf{Y}\operatorname{vec}\mathbf{Y}^T\otimes\operatorname{vec}\mathbf{Y}\operatorname{vec}\mathbf{Y}^T)\mathbf{R}_1$$

with $\mathbf{R}^T = (\mathbf{e}_d^{D\,T}\otimes\mathbf{I}_K)\otimes(\mathbf{e}_d^{D\,T}\otimes\mathbf{I}_K)$ and $\mathbf{R}_1 = (\mathbf{e}_s^D\otimes\mathbf{I}_K)\otimes(\mathbf{e}_s^D\otimes\mathbf{I}_K)$. Finally, the required result is obtained by using $\mathbf{K}_{mn}(\mathbf{A}\otimes\mathbf{B}) = (\mathbf{B}\otimes\mathbf{A})\mathbf{K}_{ts}$ and $\mathbf{K}_{mm}\equiv\mathbf{K}_m$; where $\mathbf{A}\in\Re^{n\times s}$ and $\mathbf{B}\in\Re^{m\times t}$, see [33]. □

Consider the following notation for certain double summation.

Definition 3.1. *Let* $\mathbf{A}\in\Re^{p\times q}$ *such that*

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} & \cdots & \mathbf{A}_{1n} \\ \mathbf{A}_{21} & \mathbf{A}_{22} & \cdots & \mathbf{A}_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{A}_{m1} & \mathbf{A}_{m2} & \cdots & \mathbf{A}_{mm} \end{pmatrix}, \quad \mathbf{A}_{ij}\in\Re^{r\times s}$$

with, $mr = p$ *and* $ns = q$. *Then we define*

$$\mathop{\boxplus}_{i,j}^{m,n}\mathbf{A} = \sum_{i=1}^{m}\sum_{j=1}^{n}\mathbf{A}_{ij}\in\Re^{r\times s}.$$

If $m = n$ *then,* $\mathop{\boxplus}_{i,j}^{m,m}\equiv\mathop{\boxplus}_{i,j}^{m}$.

Let the partitioned matrices $\mathbf{A} = (\mathbf{A}_{ij})$ and $\mathbf{B} = (\mathbf{B}_{ij})$. If $\odot$ denotes the Khatri-Rao product, then

$$\mathbf{A}\odot\mathbf{B} = \left(\mathbf{A}_{ij}\otimes\mathbf{B}_{ij}\right)_{ij},$$

see [38, p.30]. In particular, for $\mathbf{C} = (c_{ij})$, we have

$$\mathbf{C}\odot\mathbf{A} = \left(c_{ij}\mathbf{A}_{ij}\right)_{ij}.$$

Moreover,

$$\mathop{\boxplus}_{i,j}(\mathbf{C}\odot\mathbf{A}) = \sum_i\sum_j\left(c_{ij}\mathbf{A}_{ij}\right)_{ij}.$$

Theorem 3.3. *Suppose that* $\mathbf{Y} \sim \mathcal{E}_{K\times D}^{(K-1),D}(\boldsymbol{\mu}, \boldsymbol{\Sigma}\otimes\boldsymbol{\Theta}, h)$*, with*

$$\mathbf{Y} = (\mathbf{y}_1|\mathbf{y}_2|\cdots|\mathbf{y}_D) \text{ and } \boldsymbol{\mu} = (\boldsymbol{\mu}_1|\boldsymbol{\mu}_2|\cdots|\boldsymbol{\mu}_D).$$

If we define

$$\mathbf{B} = \mathbf{Y}\mathbf{Y}^T = \sum_{d=1}^{D} \mathbf{y}_d\mathbf{y}_d^T,$$

then

$$E(\mathbf{B}) = c_0 \operatorname{tr}(\boldsymbol{\Theta})\boldsymbol{\Sigma} + \boldsymbol{\mu}\boldsymbol{\mu}^T,$$

and

$$\begin{aligned}
\operatorname{Cov}(\operatorname{vec}\mathbf{B})) = & (\mathbf{I}_{K^2} + \mathbf{K}_K)\left\{\kappa_0 \operatorname{tr}\left(\boldsymbol{\Theta}^2\right)(\boldsymbol{\Sigma}\otimes\boldsymbol{\Sigma})\right. \\
& + c_0\left[\mathop{\boxplus}_{i,j}^{D}\left(\boldsymbol{\Theta}\odot\operatorname{vec}\boldsymbol{\mu}\operatorname{vec}^T\boldsymbol{\mu}\right)\otimes\boldsymbol{\Sigma}\right. \\
& \left.\left. +\boldsymbol{\Sigma}\otimes\mathop{\boxplus}_{i,j}^{D}\left(\boldsymbol{\Theta}\odot\operatorname{vec}\boldsymbol{\mu}\operatorname{vec}^T\boldsymbol{\mu}\right)\right]\right\} \\
& + \left[\kappa_0 \operatorname{tr}\left(\boldsymbol{\Theta}^2\right) - c_0^2\operatorname{tr}^2(\boldsymbol{\Theta})\right]\operatorname{vec}\boldsymbol{\Sigma}\operatorname{vec}^T\boldsymbol{\Sigma} \\
& + c_0\left\{\operatorname{vec}\boldsymbol{\Sigma}\operatorname{vec}^T\mathop{\boxplus}_{i,j}^{D}\left(\boldsymbol{\Theta}\odot\operatorname{vec}\boldsymbol{\mu}\operatorname{vec}^T\boldsymbol{\mu}\right)\right. \\
& + \operatorname{vec}\mathop{\boxplus}_{i,j}^{D}\left(\boldsymbol{\Theta}\odot\operatorname{vec}\boldsymbol{\mu}\operatorname{vec}^T\boldsymbol{\mu}\right)\operatorname{vec}^T\boldsymbol{\Sigma} \\
& \left. +\operatorname{tr}(\boldsymbol{\Theta})\left[\operatorname{vec}\boldsymbol{\Sigma}\operatorname{vec}^T\boldsymbol{\mu}\boldsymbol{\mu}^T + \operatorname{vec}\boldsymbol{\mu}\boldsymbol{\mu}^T\operatorname{vec}\boldsymbol{\Sigma}\right]\right\}.
\end{aligned}$$

Proof. This is a consequence of (3.1), (3.2), Definition 3.1 and Theorem 2. □

Corollary 3.1. *In Theorem 3.3 assume that* $\boldsymbol{\Theta} = \mathbf{I}_D$*. Then*

$$E(\mathbf{B}) = Dc_0\boldsymbol{\Sigma} + \boldsymbol{\mu}\boldsymbol{\mu}^T,$$

and

$$\begin{aligned}
\operatorname{Cov}(\operatorname{vec}\mathbf{B})) = & (\mathbf{I}_{K^2} + \mathbf{K}_K)\left\{D\kappa_0(\boldsymbol{\Sigma}\otimes\boldsymbol{\Sigma}) + c_0\left[\boldsymbol{\mu}\boldsymbol{\mu}^T\otimes\boldsymbol{\Sigma} + \boldsymbol{\Sigma}\otimes\boldsymbol{\mu}\boldsymbol{\mu}^T\right]\right\} \\
& + D\left[\kappa_0 - Dc_0^2\right]\operatorname{vec}\boldsymbol{\Sigma}\operatorname{vec}^T\boldsymbol{\Sigma} \\
& + (1-D)c_0[\operatorname{vec}\boldsymbol{\Sigma}\operatorname{vec}^T\boldsymbol{\mu}\boldsymbol{\mu}^T + \operatorname{vec}\boldsymbol{\mu}\boldsymbol{\mu}^T\operatorname{vec}^T\boldsymbol{\Sigma}].
\end{aligned}$$

With some errors, [23, Theorem 3.2.13 and Example 3.2.1] derived general and particular examples. For instance, in the central case $\boldsymbol{\mu} = \mathbf{0}$ with $D = n-1$, they found the factor $D(\kappa_0 - c_0^2)$ instead of the right term $D\left(\kappa_0 - Dc_0^2\right)$. The univariate published version of the one-dimensional Student's t-distribution is also incorrect.

3.2. Moments of B *under independence*

Define $\mathbf{Y}$ and $\boldsymbol{\mu}$ as follows:

$$\mathbf{Y} = (\mathbf{y}_1|\mathbf{y}_2|\cdots|\mathbf{y}_D) \text{ and } \boldsymbol{\mu} = (\boldsymbol{\mu}_1|\boldsymbol{\mu}_2|\cdots|\boldsymbol{\mu}_D),$$

where $\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_D$ are independent and

$$\mathbf{y}_d \sim \mathcal{E}_K^{(K-1)}(\boldsymbol{\mu}_d, \theta_{dd}\boldsymbol{\Sigma}; h).$$

Given

$$\mathbf{B} = \mathbf{Y}\mathbf{Y}^T = \sum_{d=1}^{D} \mathbf{y}_d\mathbf{y}_d^T,$$

we have

$$E(\mathbf{B}) = E\left(\mathbf{Y}\mathbf{Y}^T\right) = E\left(\sum_{d=1}^{D} \mathbf{y}_d\mathbf{y}_d^T\right) = \sum_{d=1}^{D} E\left(\mathbf{y}_d\mathbf{y}_d^T\right).$$

For $d \neq s = 1, 2\dots, D$, the independence assumption means that $\operatorname{Cov}(\mathbf{y}_d, \mathbf{y}_s) = \mathbf{0}$.

But, the independence of $\mathbf{y}_d$ under $d = 1, 2, \dots, D$ implies that

$$\begin{aligned}\operatorname{Cov}(\operatorname{vec}\mathbf{B}) &= \operatorname{Cov}\left(\operatorname{vec}\left(\mathbf{Y}\mathbf{Y}^T\right)\right) = \operatorname{Cov}\left(\sum_{d=1}^{D} \operatorname{vec}\left(\mathbf{y}_d\mathbf{y}_d^T\right)\right) \\ &= \sum_{d=1}^{D} \operatorname{Cov}\left(\operatorname{vec}\left(\mathbf{y}_d\mathbf{y}_d^T\right)\right) = \sum_{d=1}^{D} \operatorname{Cov}\left(\mathbf{y}_d \otimes \mathbf{y}_d\right).\end{aligned}$$

Then, we need to find $E\left(\mathbf{y}_d\mathbf{y}_d^T\right)$ and

$$\begin{aligned}\operatorname{Cov}\left(\mathbf{y}_d \otimes \mathbf{y}_d\right) &= E(\left(\mathbf{y}_d \otimes \mathbf{y}_d\right)\left(\mathbf{y}_d \otimes \mathbf{y}_d\right)^T) - E\left(\mathbf{y}_d \otimes \mathbf{y}_d\right) E\left(\mathbf{y}_d \otimes \mathbf{y}_d\right)^T \\ &= E\left(\mathbf{y}_d\mathbf{y}_d^T \otimes \mathbf{y}_d\mathbf{y}_d^T\right) - E\left(\operatorname{vec}\mathbf{y}_d\mathbf{y}_d^T\right) E\left(\operatorname{vec}^T \mathbf{y}_d\mathbf{y}_d^T\right).\end{aligned}$$

These results are summarized next.

Corollary 3.2. *Let* $\mathbf{y}_d \sim \mathcal{E}_K^{(K-1)}(\boldsymbol{\mu}_d, \theta_{dd}\boldsymbol{\Sigma}; h)$, $d = 1, 2, \dots, D$, *where* $\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_D$ *are independent. Then*

1. $E(\mathbf{y}_d\mathbf{y}_d^T) = c_0\theta_{dd}\boldsymbol{\Sigma} + \boldsymbol{\mu}_d\boldsymbol{\mu}_d^T$, *and*
2. $\operatorname{Cov}(\mathbf{y}_d \otimes \mathbf{y}_d) = \operatorname{Cov}(\operatorname{vec}\mathbf{y}_d\mathbf{y}_d^T)$

$$\begin{aligned}&= (\mathbf{I}_{K^2} + \mathbf{K}_K)\left\{\kappa_0\theta_{dd}^2(\boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma}) + c_0\theta_{dd}\left[\boldsymbol{\mu}_d\boldsymbol{\mu}_d^T \otimes \boldsymbol{\Sigma} + \boldsymbol{\Sigma} \otimes \boldsymbol{\mu}_d\boldsymbol{\mu}_d^T\right]\right\} \\ &\quad + \theta_{dd}^2(\kappa_0 - c_0^2)\operatorname{vec}\boldsymbol{\Sigma}\operatorname{vec}^T\boldsymbol{\Sigma}.\end{aligned}$$

Proof. The results follow by taking $d = s$ in Theorem 3.2. □

Theorem 3.4. *Suppose that* $\mathbf{y}_d \sim \mathcal{E}_K^{(K-1)}(\boldsymbol{\mu}_d, \theta_{dd}\boldsymbol{\Sigma}; h), \quad d = 1, 2, \ldots, D$, *with*

$$\mathbf{Y} = (\mathbf{y}_1|\mathbf{y}_2|\cdots|\mathbf{y}_D) \text{ and } \boldsymbol{\mu} = (\boldsymbol{\mu}_1|\boldsymbol{\mu}_2|\cdots|\boldsymbol{\mu}_D).$$

Define $\boldsymbol{\Theta} = \operatorname{diag}(\theta_{11}, \theta_{22}, \ldots, \theta_{dd})$, *and*

$$\mathbf{B} = \mathbf{Y}\mathbf{Y}^T = \sum_{d=1}^{D} \mathbf{y}_d\mathbf{y}_d^T.$$

Then,

$$\begin{aligned}
E(\mathbf{B}) &= c_0 \operatorname{tr}(\boldsymbol{\Theta})\boldsymbol{\Sigma} + \boldsymbol{\mu}\boldsymbol{\mu}^T \\
\operatorname{Cov}(\operatorname{vec}\mathbf{B}) &= (\mathbf{I}_{K^2} + \mathbf{K}_K)\left\{\kappa_0 \operatorname{tr}\left(\boldsymbol{\Theta}^2\right)(\boldsymbol{\Sigma}\otimes\boldsymbol{\Sigma})\right. \\
&\quad \left. +\ c_0\left[\left(\sum_{d=1}^{D}\theta_{dd}\boldsymbol{\mu}_d\boldsymbol{\mu}_d^T\right)\otimes\boldsymbol{\Sigma} + \boldsymbol{\Sigma}\otimes\left(\sum_{d=1}^{D}\theta_{dd}\boldsymbol{\mu}_d\boldsymbol{\mu}_d^T\right)\right]\right\} \\
&\quad +\ (\kappa_0 - c_0^2)\operatorname{tr}\left(\boldsymbol{\Theta}^2\right)\operatorname{vec}\boldsymbol{\Sigma}\operatorname{vec}^T\boldsymbol{\Sigma}
\end{aligned}$$

Proof. By Corollary 3.2,

$$\begin{aligned}
E(\mathbf{B}) &= E\left(\mathbf{Y}\mathbf{Y}^T\right) = E\left(\sum_{d=1}^{D}\mathbf{y}_d\mathbf{y}_d^T\right) = \sum_{d=1}^{D} E\left(\mathbf{y}_d\mathbf{y}_d^T\right) \\
&= \sum_{d=1}^{D}\left(c_0\theta_{dd}\boldsymbol{\Sigma} + \boldsymbol{\mu}_d\boldsymbol{\mu}_d^T\right) \\
&= c_0 \operatorname{tr}(\boldsymbol{\Theta})\boldsymbol{\Sigma} + \sum_{d=1}^{D}\boldsymbol{\mu}_d\boldsymbol{\mu}_d^T = c_0 \operatorname{tr}(\boldsymbol{\Theta})\boldsymbol{\Sigma} + \boldsymbol{\mu}\boldsymbol{\mu}^T.
\end{aligned}$$

Similarly,

$$\begin{aligned}
\operatorname{Cov}(\operatorname{vec}\mathbf{B}) &= \operatorname{Cov}\left(\operatorname{vec}\left(\mathbf{Y}\mathbf{Y}^T\right)\right) = \operatorname{Cov}\left(\sum_{d=1}^{D}\operatorname{vec}\left(\mathbf{y}_d\mathbf{y}_d^T\right)\right) \\
&= \sum_{d=1}^{D}\operatorname{Cov}\left(\operatorname{vec}\left(\mathbf{y}_d\mathbf{y}_d^T\right)\right),
\end{aligned}$$

and the required result is derived. □

Now, if $\boldsymbol{\Theta} = \mathbf{I}_D$ we can get the following statement:

Corollary 3.3. *Let* $\mathbf{y}_d \sim \mathcal{E}_K^{(K-1)}(\boldsymbol{\mu}_d, \boldsymbol{\Sigma}; h), \quad d = 1, 2, \ldots, D$, *where* $\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_D$ *are independent. Then*

1. $E(\mathbf{y}_d\mathbf{y}_d^T) = c_0\boldsymbol{\Sigma} + \boldsymbol{\mu}_d\boldsymbol{\mu}_d^T$, *and*
2. $\operatorname{Cov}(\mathbf{y}_d\otimes\mathbf{y}_d) = \operatorname{Cov}(\operatorname{vec}\mathbf{y}_d\mathbf{y}_d^T)$

$$\begin{aligned}
&= (\mathbf{I}_{K^2} + \mathbf{K}_K)[\kappa_0(\boldsymbol{\Sigma}\otimes\boldsymbol{\Sigma}) + c_0(\boldsymbol{\mu}_d\boldsymbol{\mu}_d^T\otimes\boldsymbol{\Sigma} + \boldsymbol{\Sigma}\otimes\boldsymbol{\mu}_d\boldsymbol{\mu}_d^T)] \\
&\quad +(\kappa_0 - c_0^2)\operatorname{vec}\boldsymbol{\Sigma}\operatorname{vec}^T\boldsymbol{\Sigma},
\end{aligned}$$

Theorem 3.5. *Suppose that* $\mathbf{y}_d \sim \mathcal{E}_K^{(K-1)}(\boldsymbol{\mu}_d, \boldsymbol{\Sigma}; h), \quad d = 1, 2, \ldots, D,$ *where*

$$\mathbf{Y} = (\mathbf{y}_1|\mathbf{y}_2|\cdots|\mathbf{y}_D), \qquad \boldsymbol{\mu} = (\boldsymbol{\mu}_1|\boldsymbol{\mu}_2|\cdots|\boldsymbol{\mu}_D)$$

and

$$\mathbf{B} = \mathbf{Y}\mathbf{Y}^T = \sum_{d=1}^{D} \mathbf{y}_d\mathbf{y}_d^T.$$

Then,

$$\begin{aligned} E(\mathbf{B}) &= Dc_0\boldsymbol{\Sigma} + \boldsymbol{\mu}\boldsymbol{\mu}^T \\ \operatorname{Cov}(\operatorname{vec}\mathbf{B}) &= (\mathbf{I}_{K^2} + \mathbf{K}_K)[D\kappa_0(\boldsymbol{\Sigma}\otimes\boldsymbol{\Sigma}) + c_0(\boldsymbol{\mu}\boldsymbol{\mu}^T\otimes\boldsymbol{\Sigma} + \boldsymbol{\Sigma}\otimes\boldsymbol{\mu}\boldsymbol{\mu}^T)] \\ &\quad + D(\kappa_0 - c_0^2)\operatorname{vec}\boldsymbol{\Sigma}\operatorname{vec}^T\boldsymbol{\Sigma} \end{aligned}$$

Proof. This is a consequence of Theorem 3.4. □

The Gaussian case follows easily:

Corollary 3.4. *If* $\mathbf{Y} \sim \mathcal{N}_{K\times D}^{(K-1),D}(\boldsymbol{\mu}, \boldsymbol{\Sigma}\otimes\mathbf{I}_D)$. *Then,* $c_0 = \kappa_0 = 1$, *and*

$$\begin{aligned} E(\mathbf{B}) &= D\boldsymbol{\Sigma} + \boldsymbol{\mu}\boldsymbol{\mu}^T \\ \operatorname{Cov}(\operatorname{vec}\mathbf{B}) &= (\mathbf{I}_{K^2} + \mathbf{K}_K)[D(\boldsymbol{\Sigma}\otimes\boldsymbol{\Sigma}) + \boldsymbol{\mu}\boldsymbol{\mu}_d^T\otimes\boldsymbol{\Sigma} + \boldsymbol{\Sigma}\otimes\boldsymbol{\mu}\boldsymbol{\mu}^T]. \end{aligned}$$

3.3. Method-of-moments estimators

We return to the original notation $\boldsymbol{\Theta} = \boldsymbol{\Sigma}_D$, $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_K^*$ and $\boldsymbol{\mu} = \boldsymbol{\mu}^*$.

In this section we will derive the method-of-moments estimators of $\boldsymbol{\Sigma}_K^*$ and $\boldsymbol{\mu}^*$.

The first two sample moment estimators of $\mathbf{B}$ are given by

$$\widetilde{E(\mathbf{B})} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{B}_i = \bar{\mathbf{B}} = (\bar{b}_{ij}), \quad i, j = 1, \ldots, K,$$

and

$$\widetilde{\operatorname{Cov}(\operatorname{vec}\mathbf{B})} = \frac{1}{n}\sum_{i=1}^{n}(\operatorname{vec}\mathbf{B}_i^T - \operatorname{vec}\widetilde{E(\mathbf{B})})(\operatorname{vec}\mathbf{B}_i^T - \operatorname{vec}\widetilde{E(\mathbf{B})})^T = \mathbf{S},$$

where $\mathbf{S} = (s_{_{tr}})$, $t, r = 1, 2, \ldots, K^2$. Moreover, for $i \leq j$, $\mathbf{M} = \boldsymbol{\mu}^*\boldsymbol{\mu}^{*T} = (m_{ij}) = \mathbf{M}^T$ and $\boldsymbol{\Sigma}_K^* = (\sigma_{ij})$, we have

$$\begin{aligned} E(b_{ij}) &= E(\mathbf{e}_i^T\mathbf{B}\mathbf{e}_j) = \mathbf{e}_i^T E(\mathbf{B})\mathbf{e}_j \\ &= \mathbf{e}_i^T(Dc_0\boldsymbol{\Sigma}_K^* + \boldsymbol{\mu}^*\boldsymbol{\mu}^{*T})\mathbf{e}_j = Dc_0\sigma_{ij} + m_{ij}. \end{aligned} \tag{3.3}$$

Note that the above expectation holds for independent and dependent cases.

3.3.1. Dependent case

Start with:

$$\begin{aligned}
\mathrm{Cov}(b_{ij}) &= \mathrm{Cov}(\mathbf{e}_i^T\mathbf{B}\mathbf{e}_j) = \mathrm{Cov}(\mathrm{vec}\,\mathbf{e}_i^T\mathbf{B}\mathbf{e}_j) = \mathrm{Cov}((\mathbf{e}_j\otimes\mathbf{e}_i)^T\,\mathrm{vec}\,\mathbf{B}) \\
&= (\mathbf{e}_j\otimes\mathbf{e}_i)^T\,\mathrm{Cov}(\mathrm{vec}\,\mathbf{B})(\mathbf{e}_j\otimes\mathbf{e}_i) \\
&= (\mathbf{e}_j\otimes\mathbf{e}_i)^T\{(\mathbf{I}_{K^2}+\mathbf{K}_K)\,\{D\kappa_0(\mathbf{\Sigma}_K^*\otimes\mathbf{\Sigma}_K^*) \\
&\quad + c_0\left[\boldsymbol{\mu}^*\boldsymbol{\mu}^{*T}\otimes\mathbf{\Sigma}_K^*+\mathbf{\Sigma}_K^*\otimes\boldsymbol{\mu}^*\boldsymbol{\mu}^{*T}\right]\} \\
&\quad + D\left[\kappa_0 - Dc_0^2\right]\mathrm{vec}\,\mathbf{\Sigma}_K^*\,\mathrm{vec}^T\,\mathbf{\Sigma}_K^* \\
&\quad + (1-D)c_0[\mathrm{vec}\,\mathbf{\Sigma}_K^*\,\mathrm{vec}^T\,\boldsymbol{\mu}^*\boldsymbol{\mu}^{T*} \\
&\quad + \,\mathrm{vec}\,\boldsymbol{\mu}^*\boldsymbol{\mu}^{*T}\,\mathrm{vec}^T\,\mathbf{\Sigma}_K^*]\}\,(\mathbf{e}_j\otimes\mathbf{e}_i).
\end{aligned}$$

Now, for $i \le j$, $i,j = 1,2,\dots,K$ we obtain

$$\begin{aligned}
\mathrm{Cov}(b_{ij}) &= D\left[\kappa_0\sigma_{ii}\sigma_{jj} + (2\kappa_0 - c_0^2)\sigma_{ij}^2\right] \\
&\quad + c_0\left[m_{jj}\sigma_{ii} + m_{ii}\sigma_{jj} + 2(2-D)m_{ij}\sigma_{ij}\right]; \qquad (3.4)
\end{aligned}$$

where we have used the equalities: $(\mathbf{e}_j\otimes\mathbf{e}_i)^T\mathbf{K}_K = (\mathbf{e}_i\otimes\mathbf{e}_j)^T$ and $(\mathbf{A}\otimes\mathbf{B})(\mathbf{C}\otimes\mathbf{D}) = (\mathbf{AC}\otimes\mathbf{BD})$. By (3.3) and taking $m_{ij} = \bar{b}_{ij} - Dc_0\sigma_{ij}$ in (3.4), we get

$$\begin{aligned}
\mathrm{Cov}(b_{ij}) &= D(\kappa_0 - 2c_0^2)\sigma_{ii}\sigma_{jj} + D(2\kappa_0 - (1+2(2-D))c_0^2)\sigma_{ij}^2 \\
&\quad + c_0\left[\bar{b}_{jj}\sigma_{ii} + \bar{b}_{ii}\sigma_{jj} + 2(2-D)\bar{b}_{ij}\sigma_{ij}\right]. \qquad (3.5)
\end{aligned}$$

Finally, by (3.5) and $s_{ij} = \widetilde{\mathrm{Cov}(b_{ij})}$ we have:

Theorem 3.6. *Assume that* $\mathbf{B} \sim \mathcal{GPW}_K^q(D, \mathbf{\Sigma}_K^*, \mathbf{I}_D, \mathbf{\Omega}, h)$*. The method-of-moments estimators of* $\mathbf{\Sigma}_K^*$ *and* $\mathbf{M} = \boldsymbol{\mu}^*\boldsymbol{\mu}^{*T}$ *are given by the following exact expressions:*

For $i = 1,2,\dots,K$***:***

$$\widetilde{\sigma}_{ii} = \frac{\sqrt{Q_{ii}^2 + 4Ps_{ii}} - Q_{ii}}{2P}, \qquad (3.6)$$

with $Q_{ii}^2 + 4Ps_{ii} \ge 0$*,* $P = D(\kappa_0 - 2c_0^2) + D(2\kappa_0 - (1+2(2-D))c_0^2)$*, and* $Q_{ii} = 2c_0(3-D)\bar{b}_{ii}$*.*

$$\widetilde{m}_{ii} = \bar{b}_{ii} - Dc_0\widetilde{\sigma}_{ii}, \qquad (3.7)$$

where $\widetilde{\sigma}_{ii}$ *has been previously found in (3.6).*

If $P = 0$*, then* $\widetilde{\sigma}_{ii} = s_{ii}/Q_{ii}$*.*

For $i < j$***,*** $i = 1,\dots,(K-1), j = 2,\dots,K$***:***

$$\widetilde{\sigma}_{ij} = \frac{\sqrt{(2-D)^2c_0^2\bar{b}_{ij}^2 - R(T_{ij} - s_{ij})} - (2-D)c_0\bar{b}_{ij}}{R}, \qquad (3.8)$$

where $(2-D)^2c_0^2\bar{b}_{ij}^2 - R(T_{ij} - s_{ij}) \ge 0$*,* $R = D(2\kappa_0 - (1+2(2-D))c_0^2)$*, and*

$$T_{ij} = D(\kappa_0 - 2c_0^2)\widetilde{\sigma}_{ii}\widetilde{\sigma}_{jj} + c_0(\bar{b}_{jj}\widetilde{\sigma}_{ii} + \bar{b}_{ii}\widetilde{\sigma}_{jj}).$$

Here $\widetilde{\sigma}_{ii}$ and $\widetilde{\sigma}_{jj}$ were previously computed in (3.6).

$$\widetilde{m}_{ij} = \overline{b}_{ij} - Dc_0\widetilde{\sigma}_{ij}, \tag{3.9}$$

Denote the solution as

$$(\widetilde{\mathbf{M}}, \widetilde{\mathbf{\Sigma}}_K^*).$$

Note that s_{ij} comes from the diagonal of $\mathbf{S} \in \Re^{K^2\times K^2}$, because $s_{ij} = \widetilde{Cov(b_{ij})}$.
Finally, if $R = 0$, then $\widetilde{\sigma}_{ij} = (s_{ij} - T_{ij}) \,/\, \big(2(2-D)c_0\overline{b}_{ij}\big)$.

Remark 3.2. *Special attention must be payed on P, Q_{ii}, R, T_{ij}, and the sign of the square root. They depend on the selected model, the sample statistics s_{ij} and $\overline{b}_{ij}$.*

3.3.2. Independent case

For this case,

$$\begin{aligned}
\operatorname{Cov}(b_{ij}) &= \operatorname{Cov}(\mathbf{e}_i^T\mathbf{B}\mathbf{e}_j) = \operatorname{Cov}(\operatorname{vec}\mathbf{e}_i^T\mathbf{B}\mathbf{e}_j) = \operatorname{Cov}((\mathbf{e}_j\otimes\mathbf{e}_i)^T\operatorname{vec}\mathbf{B})\\
&= (\mathbf{e}_j\otimes\mathbf{e}_i)^T\operatorname{Cov}(\operatorname{vec}\mathbf{B})(\mathbf{e}_j\otimes\mathbf{e}_i)\\
&= (\mathbf{e}_j\otimes\mathbf{e}_i)^T\{(\mathbf{I}_{K^2}+\mathbf{K}_K)[D\kappa_0(\mathbf{\Sigma}_K^*\otimes\mathbf{\Sigma}_K^*)\\
&\quad + c_0(\boldsymbol{\mu}^*\boldsymbol{\mu}^{*T}\otimes\mathbf{\Sigma}_K^* + \mathbf{\Sigma}_K^*\otimes\boldsymbol{\mu}^*\boldsymbol{\mu}^{*T})]\\
&\quad + D(\kappa_0 - c_0^2)\operatorname{vec}\mathbf{\Sigma}_K^*\operatorname{vec}^T\mathbf{\Sigma}_K^*\}(\mathbf{e}_j\otimes\mathbf{e}_i).
\end{aligned}$$

Hence

$$\operatorname{Cov}(b_{ij}) = D\left[\kappa_0\sigma_{ii}\sigma_{jj} + (2\kappa_0 - c_0^2)\sigma_{ij}^2\right] + c_0\left(m_{jj}\sigma_{ii} + m_{ii}\sigma_{jj} + 2m_{ij}\sigma_{ij}\right). \tag{3.10}$$

By (3.3) and substitution of $m_{ij} = s_{ij} - Dc_0\sigma_{ij}$ in (3.10), we obtain

$$\operatorname{Cov}(b_{ij}) = D(\kappa_0 - 2c_0^2)\sigma_{ii}\sigma_{jj} + D(2\kappa_0 - 3c_0^2)\sigma_{ij}^2 + c_0\left[\overline{b}_{jj}\sigma_{ii} + \overline{b}_{ii}\sigma_{jj} + 2\overline{b}_{ij}\sigma_{ij}\right].$$

Summarizing:

Theorem 3.7. *Assume independent $\mathbf{y}_d \sim \mathcal{E}_K^{(K-1)}(\boldsymbol{\mu}_d, \mathbf{\Sigma}; h)$, for $d = 1, 2, \dots, D$. Define*

$$\mathbf{Y} = (\mathbf{y}_1|\mathbf{y}_2|\cdots|\mathbf{y}_D) \qquad \boldsymbol{\mu} = (\boldsymbol{\mu}_1|\boldsymbol{\mu}_2|\cdots|\boldsymbol{\mu}_D),$$

and

$$\mathbf{B} = \mathbf{Y}\mathbf{Y}^T = \sum_{d=1}^{D}\mathbf{y}_d\mathbf{y}_d^T.$$

Then, the method-of-moments estimators of $\mathbf{\Sigma}_K^*$ and $\mathbf{M} = \boldsymbol{\mu}^*\boldsymbol{\mu}^{*T}$ are given by the following exact expressions:

For** $i = 1, 2, \dots, K$**:

$$\widetilde{\sigma}_{ii} = \frac{\sqrt{Q_{ii}^2 + 4Ps_{ii}} - Q_{ii}}{2P}, \tag{3.11}$$

with $Q_{ii}^2 + 4Ps_{ii} \geq 0$, $P = D(3\kappa_0 - 5c_0^2)$, *and* $Q_{ii} = 4c_0\overline{b}_{ii}$.

$$\widetilde{m}_{ii} = \overline{b}_{ii} - Dc_0\widetilde{\sigma}_{ii}, \tag{3.12}$$

where $\widetilde{\sigma}_{ii}$ *has been previously found in (3.11).*

If $P = 0$, *then* $\widetilde{\sigma}_{ii} = s_{ii}/Q_{ii}$.

For $i < j$, $i = 1, \ldots, (K-1), j = 2, \ldots, K$***:***

$$\widetilde{\sigma}_{ij} = \frac{\sqrt{c_0^2\overline{b}_{ij}^2 - R(T_{ij} - s_{ij})} - c_0\overline{b}_{ij}}{R}, \tag{3.13}$$

where $\overline{b}_{ij}^2 - R(T_{ij} - s_{ij}) \geq 0$, $R = D(2\kappa_0 - 3c_0^2)$ *and*

$$T_{ij} = D(\kappa_0 - 2c_0^2)\widetilde{\sigma}_{ii}\widetilde{\sigma}_{jj} + c_0\overline{b}_{jj}\widetilde{\sigma}_{ii} + c_0\overline{b}_{ii}\widetilde{\sigma}_{jj}.$$

Here $\widetilde{\sigma}_{ii}$ *and* $\widetilde{\sigma}_{jj}$ *were previously computed in (3.11).*

$$\widetilde{m}_{ij} = \overline{b}_{ij} - Dc_0\widetilde{\sigma}_{ij}, \tag{3.14}$$

Denote the solution as

$$(\widetilde{\mathbf{M}}, \widetilde{\mathbf{\Sigma}}_K^*).$$

As before, given that $s_{ij} = \widetilde{Cov(b_{ij})}$, *then the* s_{ij}*'s are in the diagonal of matrix* $\mathbf{S} \in \Re^{K^2 \times K^2}$.

Finally, if $R = 0$, *then* $\widetilde{\sigma}_{ij} = (s_{ij} - T_{ij}) / \left(2c_0\overline{b}_{ij}\right)$.

Remark 3.3. *Recall that the method-of-moments estimators are not uniquely defined. For example, assume that the method-of-moments estimator of* $g(\theta)$ *is required, instead of the estimation of* θ. *We have a number of different ways. We can obtain the method-of-moments estimator* $\widetilde{\theta}$ *of* θ *and then use it for finding* $g(\widetilde{\theta})$ *as an estimator of* $g(\theta)$. *Alternatively, we can find the moments of the function* $g(\theta)$ *and then obtain the method-of-moments estimator* $\widetilde{g(\theta)}$ *of* $g(\theta)$. *The resulting method-of-moments estimators are usually different in both cases, see [35, Section 7.2.1, p.276]. Furthermore, by the law of large numbers the method-of-moments estimators provides (under suitable conditions) consistent estimators, see [38, section 5d.1, p. 351].*

Principal Coordinate Analysis, summarized by [30], provides the following algorithm for $\boldsymbol{\mu}^*$. The procedure finds the estimated coordinates of the mean form by using the method-of-moments estimator $\widetilde{\mathbf{M}}$. Note that the mean form is invariant under translation, rotation, and reflection transformations.

Theorem 3.8. *Let* $\widetilde{\mathbf{M}}$ *be the method-of-moments estimator of* $\mathbf{M} = \boldsymbol{\mu}^*\boldsymbol{\mu}^{*T}$ *(for dependent or independent cases). Let* $\widetilde{\mathbf{M}} = \mathbf{V}_1\mathbf{L}\mathbf{V}_1^T$ *the nonsingular part of the corresponding spectral decomposition, where* $\mathbf{V}_1$ *is a semiorthogonal matrix. Here* $\mathbf{V}_1 \in \Re^{K \times D}$, $\mathbf{V}_1^T\mathbf{V}_1 = \mathbf{I}_D$ *and* $\mathbf{L} = \mathrm{diag}(\lambda_1, \ldots, \lambda_D)$, *where* D *is the rank of* $\widetilde{\mathbf{M}}$. *Then the method-of-moments estimator of* $\boldsymbol{\mu}^*$ *is*

$$\widetilde{\boldsymbol{\mu}}^* = \mathbf{V}_1\mathbf{W},$$

with $\mathbf{W} = \mathrm{diag}(\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_D})$.

Proof. The required result follows from Remark 3.3. □

Theorem 3.9. *Let* $(\widetilde{\boldsymbol{\mu}}^*, \widetilde{\boldsymbol{\Sigma}}_K^*)$ *be the method-of-moments estimators of*

$$(\boldsymbol{\mu}^*, \boldsymbol{\Sigma}_K^*).$$

Then as $n \to \infty$

$$(\widetilde{\boldsymbol{\mu}}^*, \widetilde{\boldsymbol{\Sigma}}_K^*) \to (\boldsymbol{\mu}^*, \boldsymbol{\Sigma}_K^*) \qquad \textit{in probability.}$$

Proof. The required result is derived by applying the consistency of the sample moments and the continuity of the function $(\boldsymbol{\mu}^*, \boldsymbol{\Sigma}_K^*)$ in $(E(\mathbf{B}), \mathrm{Cov}(\mathrm{vec}\,\mathbf{B}))$, see [38, Section 5d.1, p. 351]. □

4. Consistent estimation when $\boldsymbol{\Sigma}_D$ is a general non-negative definite matrix

This section is motivated by the Gaussian version of the maximum likelihood estimation derived by [18]. We provide a heuristic evaluation of the elliptical case under our approach based in the method-of-moments estimation. Our algorithm is based in the following modified expressions:

$$\widetilde{\boldsymbol{\Sigma}}_D = \frac{1}{nK}\sum_{i=1}^{n}\left(\mathbf{X}_i^c - \widetilde{\boldsymbol{\mu}}^*\right)^T (\widetilde{\boldsymbol{\Sigma}}_K^*)^- \left(\mathbf{X}_i^c - \widetilde{\boldsymbol{\mu}}^*\right), \tag{4.1}$$

$$\widetilde{\boldsymbol{\Sigma}}_K^* = \frac{1}{nD}\sum_{i=1}^{n}\left(\mathbf{X}_i^c - \widetilde{\boldsymbol{\mu}}^*\right) \widetilde{\boldsymbol{\Sigma}}_D^{-1} \left(\mathbf{X}_i^c - \widetilde{\boldsymbol{\mu}}^*\right)^T. \tag{4.2}$$

Algorithm
Initialization:
$r = 0$; $\boldsymbol{\Sigma}_K^{*r} = \widetilde{\boldsymbol{\Sigma}}_K^*$; $\boldsymbol{\Sigma}_D^r = \dfrac{1}{nK}\displaystyle\sum_{i=1}^{n}\left(\mathbf{X}_i^c - \widetilde{\boldsymbol{\mu}}^*\right)^T (\boldsymbol{\Sigma}_K^{*r})^- \left(\mathbf{X}_i^c - \widetilde{\boldsymbol{\mu}}^*\right)$;
$r = r + 1$

$$\boldsymbol{\Sigma}_K^{*r+1} = \frac{1}{nD}\sum_{i=1}^{n}\left(\mathbf{X}_i^c - \widetilde{\boldsymbol{\mu}}^*\right)(\boldsymbol{\Sigma}_D^r)^{-1}\left(\mathbf{X}_i^c - \widetilde{\boldsymbol{\mu}}^*\right)^T;$$

$$\boldsymbol{\Sigma}_D^{r+1} = \frac{1}{nK}\sum_{i=1}^{n}\left(\mathbf{X}_i^c - \widetilde{\boldsymbol{\mu}}^*\right)^T (\boldsymbol{\Sigma}_K^{*r})^- \left(\mathbf{X}_i^c - \widetilde{\boldsymbol{\mu}}^*\right);$$

While
$||\boldsymbol{\Sigma}_D^{r+1} - \boldsymbol{\Sigma}_D^r||_2 > \varepsilon_1$ or $||\boldsymbol{\Sigma}_K^{*r+1} - \boldsymbol{\Sigma}_K^{*r}||_2 > \varepsilon_2$,
Repeat:
  $r = r + 1$;
  $\boldsymbol{\Sigma}_K^{*r} = \boldsymbol{\Sigma}_K^{*r+1}$;
  $\boldsymbol{\Sigma}_D^r = \boldsymbol{\Sigma}_D^{r+1}$;
  Recompute $\boldsymbol{\Sigma}_K^{*r+1}$ and $\boldsymbol{\Sigma}_D^{r+1}$.
The solutions are:
$\widetilde{\boldsymbol{\Sigma}}_K^* = \boldsymbol{\Sigma}_K^{*r}$; $\widetilde{\boldsymbol{\Sigma}}_D = \boldsymbol{\Sigma}_D^r$.

Where ε_1 and ε_2 define two infinitesimal positive quantities and $||\cdot||_2$ is the Euclidean norm, $\left(||\mathbf{A}||_2 = \sqrt{\text{tr}\left(\mathbf{A}\mathbf{A}^T\right)}\right)$.

Theorem 4.1. *Let* $(\widetilde{\boldsymbol{\mu}}^*, \widetilde{\boldsymbol{\Sigma}}_K^* \otimes \widetilde{\boldsymbol{\Sigma}}_D)$ *be the method-of-moments estimators of* $(\boldsymbol{\Sigma}_K^*, \boldsymbol{\mu}^* \otimes \boldsymbol{\Sigma}_D)$. *Then as* $n \to \infty$

$$(\widetilde{\boldsymbol{\mu}}^*, \widetilde{\boldsymbol{\Sigma}}_K^* \otimes \widetilde{\boldsymbol{\Sigma}}_D) \to (\boldsymbol{\mu}^*, \boldsymbol{\Sigma}_K^* \otimes \boldsymbol{\Sigma}_D) \qquad \text{in probability.}$$

Proof. This is a consequence of Remark 3.3. □

5. Estimation of the form difference

A detailed discussion of Euclidean Distance Matrix, matrix form, form difference and their probabilistic and geometrical properties can be found in [29, 30].

Consider the following square symmetric matrix, also known as the Euclidean Distance Matrix:

$$\mathbf{F}(\mathbf{X}) = \begin{pmatrix} 0 & d(1,2) & \dots & d(1,K-1) & d(1,K) \\ d(2,1) & 0 & \dots & d(2,K-1) & d(2,K) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ d(K,1) & d(K,2) & \dots & d(K,K-1) & 0 \end{pmatrix},$$

where $d(i,j)$ denotes the Euclidean distance between landmarks i and j. In shape theory this matrix is termed *form matrix*. Some interesting properties of form matrix are given in [29]. For example, $\mathbf{F}(\mathbf{X})$ is a maximal invariant under the group of transformations consisting of translation, rotation, and reflection. Therefore, $\mathbf{F}(\mathbf{X})$ retains all the geometrical relevant information about the form of the object.

Let $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n$ be n independent observations from a population I and let $\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_m$ be m independent observations from a population II. Let $\boldsymbol{\mu}^{\mathbf{X}}$ be the mean form of population I with corresponding form matrix $\mathbf{F}\left(\boldsymbol{\mu}^{\mathbf{X}}\right)$. In a similar way denote $\boldsymbol{\mu}^{\mathbf{Y}}$ and $\mathbf{F}\left(\boldsymbol{\mu}^{\mathbf{Y}}\right)$ for the population II. [30] provides the the following concept:

Definition 5.1. *The form difference between population I and II is defined as*

$$\mathbf{FDM}\left(\boldsymbol{\mu}^{\mathbf{X}}, \boldsymbol{\mu}^{\mathbf{Y}}\right) = \mathbf{F}\left(\boldsymbol{\mu}^{\mathbf{X}}\right) * \mathbf{F}\left(\boldsymbol{\mu}^{\mathbf{Y}}\right)^{-H},$$

where $*$ *denotes the Hadamard product and* $0/0 = 0$. *Here* $\mathbf{A}^{-H}$ *denotes the inverse of* $\mathbf{A}$ *with respect to the Hadamard product. [6] give a formula for this inverse in terms of the usual product.*

The next result applies the Remark 3.3 in the consistent estimation of the form difference between two populations. It assumes non-isotropic perturbed landmarks within the axes but isotropic relation between them.

Theorem 5.1. *Let* $\left(\boldsymbol{\mu}^{\mathbf{X}}, \boldsymbol{\Sigma}^*_{K\mathbf{X}} \otimes \boldsymbol{\Sigma}_{D\mathbf{X}}\right)$ *and* $\left(\boldsymbol{\mu}^{\mathbf{Y}}, \boldsymbol{\Sigma}^*_{K\mathbf{Y}} \otimes \boldsymbol{\Sigma}_{D\mathbf{Y}}\right)$ *be the corresponding parameters of populations I and II. If* $\boldsymbol{\Sigma}_{D\mathbf{X}} = \boldsymbol{\Sigma}_{D\mathbf{Y}} = \mathbf{I}_D$*, then*

$$\widetilde{\mathbf{FDM}}\left(\widetilde{\boldsymbol{\mu}}^{\mathbf{X}}, \widetilde{\boldsymbol{\mu}}^{\mathbf{Y}}\right) = \widetilde{\mathbf{F}}\left(\widetilde{\boldsymbol{\mu}}^{\mathbf{X}}\right) * \widetilde{\mathbf{F}}\left(\widetilde{\boldsymbol{\mu}}^{\mathbf{Y}}\right)^{-H} \to \mathbf{FDM}\left(\boldsymbol{\mu}^{\mathbf{X}}, \boldsymbol{\mu}^{\mathbf{Y}}\right) \quad \textit{in probability.}$$

Theorem 5.2. *Let* $\left(\boldsymbol{\mu}^{\mathbf{X}}, \boldsymbol{\Sigma}^*_{K\mathbf{X}} \otimes \boldsymbol{\Sigma}_{D\mathbf{X}}\right)$ *and* $\left(\boldsymbol{\mu}^{\mathbf{Y}}, \boldsymbol{\Sigma}^*_{K\mathbf{Y}} \otimes \boldsymbol{\Sigma}_{D\mathbf{Y}}\right)$ *be the corresponding parameters of populations I and II. Then*

$$\widetilde{\mathbf{FDM}}\left(\widetilde{\boldsymbol{\mu}}^{\mathbf{X}}, \widetilde{\boldsymbol{\mu}}^{\mathbf{Y}}\right) = \widetilde{\mathbf{F}}\left(\widetilde{\boldsymbol{\mu}}^{\mathbf{X}}\right) * \widetilde{\mathbf{F}}\left(\widetilde{\boldsymbol{\mu}}^{\mathbf{Y}}\right)^{-H} \to \mathbf{FDM}\left(\boldsymbol{\mu}^{\mathbf{X}}, \boldsymbol{\mu}^{\mathbf{Y}}\right) \quad \textit{in probability.}$$

6. Example

The mouse vertebra problem was originally studied in the Gaussian case by [16] (see also [34]). A further analysis under elliptical models was implemented by [8]. The experiment considers the second thoracic vertebra T2 of two groups of mice: large and small. The mice are selected and classified according to large or small body weight; in this case, the sample consists of 23, 23 and 30 large, small and control bones, respectively. The vertebras are digitized and summarized in six mathematical landmarks which are placed at points of high curvature; they are symmetrically selected by measuring the extreme positive and negative curvature of the bone. Figure 1 shows the sample for the three groups with the corresponding (x, y) cartesian coordinates of the six landmarks in the bones. Note that the clusters of the landmarks have different shapes, then the usual normal isotropic model considered in the Gaussian literature is not appropriate, see [16] and the references within. The shape difference analysis among the three groups is quite solved by different approaches. However the correlation structure among landmarks requires more research. Perform inference with such non isotropic shape distributions is very difficult. It has forced the use of strong assumptions about correlation.

More than an example, this landmark data is highly valuable for a correlation structure analysis. The symmetry of the vertebra, certainly suggests a priori non isotropic model. The control group is also useful for comparisons and correctness.

Theorems 3.6 and 3.8 can be easily implemented for a number of models. We focus on the main novelty (Theorem 3.6) and a Kotz type model (including Gaussian). This model is very flexible and meaningful for various values of the parameters r, s and N, see appendix.

First we illustrate Theorem 3.7 under six different models with independent landmarks. Moment-method estimates of mean shape by using the classical Gaussian model is shown in figure 2. For standard plots, we show the Bookstein's coordinates of the cartesian coordinates computed in the theorems of the paper, see [16]. Then we can compare directly the estimations. Bookstein's coordinates are easily obtained by fixing a reference line with landmarks 1 and 2. Both landmarks are sent to $(-0.5, 0)$ and $(0.5, 0)$, respectively, then we plot the coordinates of the remaining four landmarks of the mean shape. This display is also useful because the mean shape is collected in four landmarks instead of six.

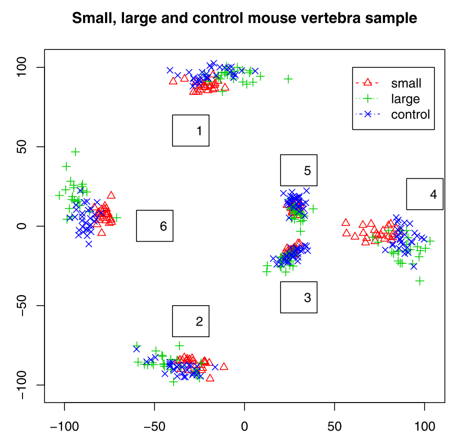

FIG 1. *Mouse vertebra sample*

In the Gaussian case the estimate is unrealistic, we expect a mean form similar to the ideal vertebra. Note that each point of the polygon represents a landmark of the bone. In the Gaussian case, the assumption of landmark independence explains the broken symmetry in the estimate. However, if we consider more robust isotropic models than Gaussian, the estimation tends to recover the symmetry of the vertebra. Indeed, this is a surprising and interesting aspect to research, because the independence of the landmarks is neglected by the complexity of the model and the mean form gets closer to the ideal vertebra. The addressed evolution from Kotz 1 to Kotz 5 is depicted in figures 3 to 7. The Bookstein's coordinates of the estimated mean form are colored in order to perform comparisons. Here S, L and C, denote the small, large and control groups, respectively.

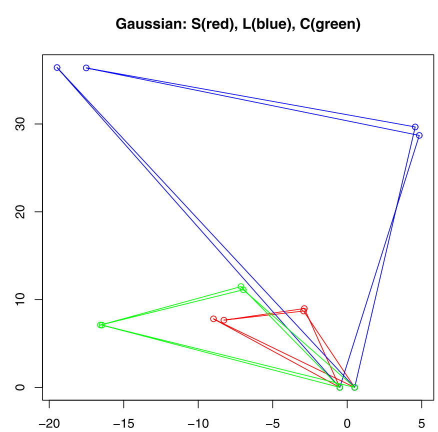

FIG 2. *Moment method estimates under independence: Gaussian model*

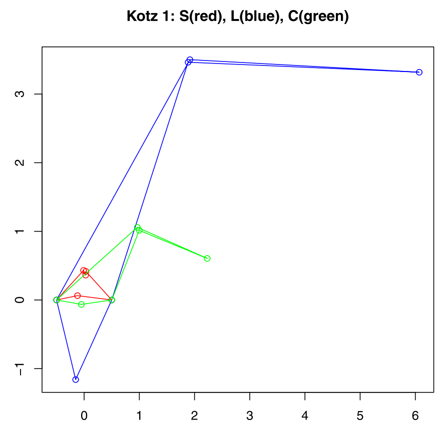

FIG 3. *Moment method estimates under independence: Kotz 1 model*

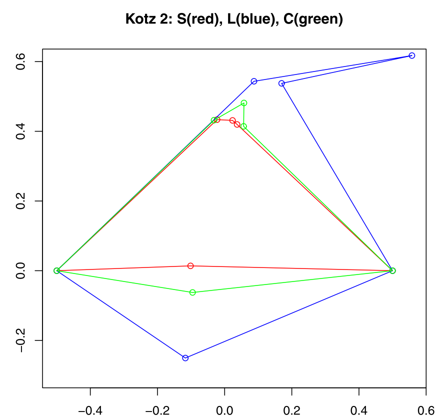

FIG 4. *Moment method estimates under independence: Kotz 2 model*

The addressed heuristic evolution suggests also the role of the analysis in shape data. First note that the mouse vertebra data is based on non anatomical landmarks, then a number of models have equal chance to estimate the mean form and the correlation structure. In this case, we have noticed that the classical isotropic Gaussian model of the literature is not appropriate. Then, we can propose robust laws and a selection criteria. However, for anatomical landmark data modeled by an expert in morphometrics, we must follow the proposed law and robust models cannot be implemented.

We now focus on the dependent case and the moment method estimators of Theorem 3.6. Given that the Gaussian case is out of any consideration, then we study other Kotz models. In order to illustrate the important effect of landmark dependence, we study a simple non Gaussian Kotz model. Consider the Kotz 1 function with $N = 2$, $r = 1/2$ and $s = 1$. We will compare its performance

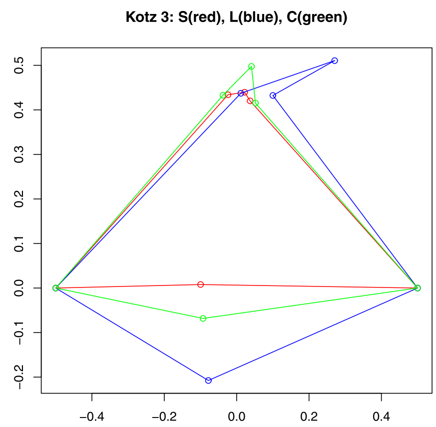

FIG 5. *Moment method estimates under independence: Kotz 3 model*

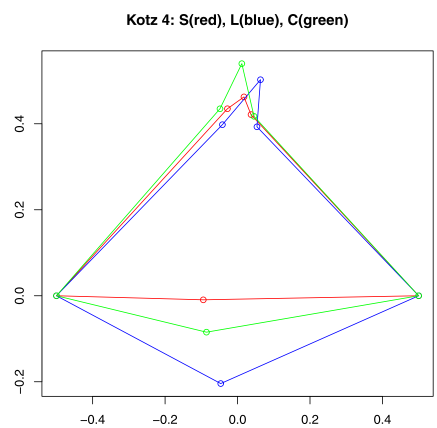

FIG 6. *Moment method estimates under independence: Kotz 4 model*

via Theorem 3.6 with another mean shape estimations. Table 2 provides comparisons among mean shape estimates of the small group. We include the mean shape by moments of Theorem 3.6, the mean shape by the Fréchet method (see [25]), and the mean shape by Bookstein method (see [2]). We can compare directly the estimations if we use Bookstein's coordinates, i.e. the reference line is given by landmarks 1 and 2. Both landmarks are sent to $(-0.5, 0)$ and $(0.5, 0)$, respectively, then the mean shape is comprised in the remaining four landmarks. Certainly, the estimations are truly similar. Recall that Kotz 1 law with independent landmarks gave a low moment method estimator of mean shape. However, under the expected and realistic dependence, the same model equals the mean shape estimators derived by standard shape theories, see figures 3 and 8, respectively.

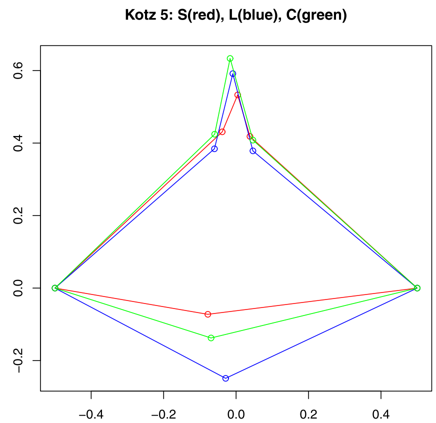

FIG 7. *Moment method estimates under independence: Kotz 5 model*

TABLE 2
Mean shape estimation in the small group by Theorem 3.6 (Kotz 1), Fréchet (F), and Bookstein (B).

Th. 6, $\widetilde{\mu}_1$	Th. 6, $\widetilde{\mu}_2$	B. $\widetilde{\mu}_1$	B. $\widetilde{\mu}_2$	F. $\widetilde{\mu}_1$	F. $\widetilde{\mu}_2$
-0.5	0	-0.5	0	-0.5	0
0.5	0	0.5	0	0.5	0
0.084507028	0.3301634	0.08469746	0.2933430	0.08490820	0.2924684
0.014836162	0.6957339	0.01215768	0.5613175	0.01245608	0.5589496
-0.073397569	0.3394693	-0.06874750	0.2991278	-0.06869796	0.2982314
-0.005026754	-0.2184060	-0.02502185	-0.3041418	-0.02512807	-0.3044915

The exact estimation of Theorem 3.6 also agrees with literature about strong difference in Gaussian mean shape between the small (S) and large (L) groups. Figure 8 also shows the mean shape estimation of the control (C) group. As we expect, the control group must tend to show strong symmetry among landmarks, by "averaging" in some sense the small and large estimates.

For a complete analysis we have included the so called Kotz 1, Kotz 2, Kotz 3, Kotz 4 and Kotz 5 models, with parameters $(N = 2, s = 1, r = 1/2)$; $(N = 3, s = 1, r = 1/2)$; $(N = 2, s = 2, r = 1/2)$; $(N = 2, s = 3, r = 1/2)$ and $(N = 20, s = 20, r = 1/2)$, respectively. The appendix gives the technical details about the generalized singular Pseudo-Wishart distributions and the particular Kotz Pseudo-Wishart distributions referred in this example. The corresponding mean shapes estimates were computed for the 5 models, but for reasons of space, we only show the results of the Kotz 5 function. This model was suggested by its performance with Theorem 3.7 and certain selection criteria that we will propose later, see figure 9. The figure shows the sample of the three groups, Small:$\triangle$, Large: $+$, Control: $\times$. It also plots the Kotz 5 mean shape joined with colored dash lines. For a simple plot, we have used again the addressed Bookstein's coordinates with landmarks 1 and 2 as the reference base-line.

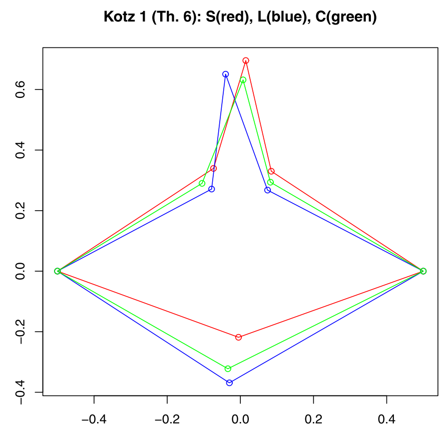

FIG 8. *Moment method estimates under dependence: Kotz 1 model*

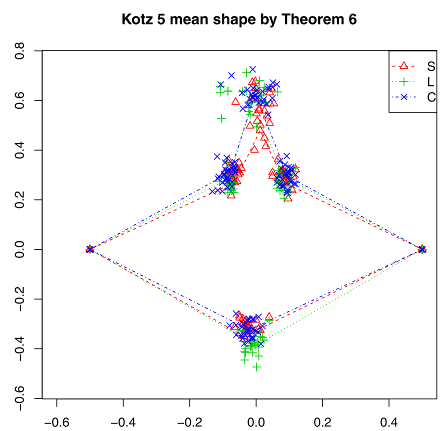

FIG 9. *Mean shape estimates for large, small and control groups under the Kotz 5 model*

Now we apply the routine of Section 4 for a consistent estimation of the general non-negative definite matrix $\mathbf{\Sigma}_D$. We show the results for the Kotz 5 model. First, we set the tolerance $\varepsilon_1 = \varepsilon_2 = 0.000005$ in the three groups (small, large and control). The number of iterations to reach that tolerance in the three groups was 57, 53 and 61, respectively.

The estimated covariance matrices for the small group are given next. For a simple interpretation we provide the correlation matrix $\boldsymbol{\rho}$ instead of $\mathbf{\Sigma}$. Recall that $\boldsymbol{\rho} = (\mathrm{diag}(\mathbf{\Sigma}))^{-\frac{1}{2}}\ \mathbf{\Sigma}\ (\mathrm{diag}(\mathbf{\Sigma}))^{-\frac{1}{2}}$):

$$\widetilde{\boldsymbol{\rho}}_K^* = \begin{pmatrix} 1.0000000 & -0.88123087 & -0.45286210 & -0.0947470 & 0.3167573 & 0.1049686 \\ -0.8812309 & 1.00000000 & -0.01931031 & -0.3859582 & -0.7246992 & 0.3741399 \\ -0.4528621 & -0.01931031 & 1.00000000 & 0.9267571 & 0.6990041 & -0.9323209 \\ -0.0947470 & -0.38595825 & 0.92675709 & 1.0000000 & 0.9113738 & -0.9979133 \\ 0.3167573 & -0.72469917 & 0.69900414 & 0.9113738 & 1.0000000 & -0.9087033 \\ 0.1049686 & 0.37413987 & -0.93232089 & -0.9979133 & -0.9087033 & 1.0000000 \end{pmatrix};$$

and

$$\widetilde{\boldsymbol{\rho}}_D = \begin{pmatrix} 1.00000000 & -0.1305434 \\ -0.1305434 & 1.00000000 \end{pmatrix}.$$

For the large group the estimated correlation matrices are:

$$\widetilde{\boldsymbol{\rho}}_K^* = \begin{pmatrix} 1.00000000 & -0.65790585 & -0.49909037 & -0.05610956 & -0.07744806 & 0.42376772 \\ -0.65790585 & 1.00000000 & 0.06499537 & -0.44441269 & -0.29572585 & 0.03327717 \\ -0.49909037 & 0.06499537 & 1.00000000 & 0.42647926 & 0.65775361 & -0.76574318 \\ -0.05610956 & -0.44441269 & 0.42647926 & 1.00000000 & 0.65493138 & -0.73668655 \\ -0.07744806 & -0.29572585 & 0.65775361 & 0.65493138 & 1.00000000 & -0.79064351 \\ 0.42376772 & 0.03327717 & -0.76574318 & -0.73668655 & -0.79064351 & 1.00000000 \end{pmatrix};$$

and

$$\widetilde{\boldsymbol{\rho}}_D = \begin{pmatrix} 1.00000000 & -0.2080039 \\ -0.2080039 & 1.00000000 \end{pmatrix}.$$

Meanwhile in the control group the estimated correlation matrices are:

$$\widetilde{\boldsymbol{\rho}}_K^* = \begin{pmatrix} 1.00000000 & -0.6179802 & -0.6137305 & -0.3968700 & 0.05198526 & 0.5036286 \\ -0.61798019 & 1.0000000 & -0.1036402 & -0.4037052 & -0.70811391 & 0.2660900 \\ -0.61373047 & -0.1036402 & 1.0000000 & 0.7604036 & 0.50488092 & -0.8386367 \\ -0.39687003 & -0.4037052 & 0.7604036 & 1.0000000 & 0.64810504 & -0.9587295 \\ 0.05198526 & -0.7081139 & 0.5048809 & 0.6481050 & 1.00000000 & -0.6427126 \\ 0.50362856 & 0.2660900 & -0.8386367 & -0.9587295 & -0.64271257 & 1.0000000 \end{pmatrix};$$

and

$$\widetilde{\boldsymbol{\rho}}_D = \begin{pmatrix} 1.00000000 & 0.1048453 \\ 0.1048453 & 1.00000000 \end{pmatrix}.$$

The three groups reveal null correlation between the axes. However, high correlation among landmarks is found, as we expected from the symmetry of the bones. The estimates in the experimental groups (S, L) detect the differential landmarks for both mean shapes. In the control group, the estimates tends to follow the main contribution of those differential landmarks, as we expect.

We have also run the routines for the models Kotz 1, to Kotz 4 with the same tolerance. They reached the stability between 50 to 70 iterations in the three groups. Similar conclusions about the almost null correlation among axes and strong correlation among landmarks were found in the models. We omit the results for models 1 to 4 and focus on the Kotz 5 model, which is the "best" of them under the following selection criteria.

For a model selection criteria, the control group plays a fundamental role. In this case we look for the minimum coefficient of variation with the small and large groups. We also consider the distance between the small and the large group relative to the mean with controls. We apply a non-Euclidian distance between covariance matrices, a technique due to [17]. The method is appropriate for meaningful correlation matrices, in this case it is performed only for $\widetilde{\Sigma}_K^*$. In Tables 3 and 4, the notation $K1$,..., $K5$, s, l, c, stand for Kotz 1,..., Kotz 5, small, large and control, respectively.

Tables 3 and 4 show the pairwise covariance distances and the percentage of the variation coefficient is presented in parenthesis. We are searching for models which reflect the role of the control group and separate the classes properly. The analysis must be complemented with the mean shape estimates. Moreover, a third criterion considers the distance with another accepted estimate; in this

Table 3
Model selection criteria.

	K1l	K1c	K2s	K2l	K2c	K3s	K3l	K3c
K1s	12.9	8.7(37)	12.8	15.9	14.1	11.8	15.7	13.7
K1l		5.1(37)	10.6	6.1	8.4	10.8	5.1	8.5
K1c			9.6	9.1	8.9	9.2	8.6	8.7
K2s				11.0	11.0(14)	6.0	11.1	10.7
K2l					9.0(14)	12.0	1.5	9.1
K2c						11.6	8.8	0.9
K3s							12.1	11.2(15)
K3l								9.0(15)

Table 4
Model selection criteria.

	K4s	K4l	K4c	K5s	K5l	K5c
K1s	11.1	15.0	13.1	11.1	13.9	12.1
K1l	11.3	3.9	8.6	12.2	2.4	2.8
K1c	9.2	7.7	8.4	9.7	6.4	4.8
K2s	8.9	10.8	10.3	10.9	10.1	9.7
K2l	13.3	3.0	9.5	14.7	4.4	7.0
K2c	12.4	8.5	2.1	13.4	8.2	8.4
K3s	6.4	11.6	10.7	9.2	10.7	9.8
K3l	13.2	1.6	9.5	14.6	3.3	6.2
K3c	11.9	8.7	1.2	12.9	8.4	8.3
K4s		12.6	11.4(16)	6.8	11.6	10.6
K4l			9.1(16)	13.9	1.8	5.1
K4c				12.3	8.7	8.3
K5s					12.8	11.6(74)
K5l						3.6(74)

case we use the Fréchet mean shape. The addressed mean shape distance can be achieved by a number of approaches, see for example [24].

Kotz 3 and Kotz 4 models behave well with low variation coefficient, but they are far from the control group and the sample. If we find the so called Riemannian distance among the moment method estimates and the Fréchet and Bookstein mean shapes, we obtain the results of table 5:

Table 5
Model selection criteria.

	K2	K3	K4	K5	F.	B.
K1	0.274	0.236	0.211	0.180	0.113	0.112
K2		0.082	0.128	0.153	0.189	0.191
K3			0.048	0.078	0.131	0.133
K4				0.035	0.099	0.100
K5					0.067	0.068
F						0.002

The mean shape of the Kotz 5 model is very near to the estimates computed by Fréchet and Bookstein (which are also similar). It also reflects good difference between the small and large groups. Then collecting the results, we can propose Kotz type 5 model as a suitable law for modeling this particular example. Note that this selection agrees with the conclusion proposed in the independent case.

It is important to quote, that a mathematical or statistical selection is just a suggestion. In this case we proceed because the experiment lacks of a priori assumption for anatomical landmark distribution. In our case, literature shows no expert assumption about normality, see [16] and the references therein. It was traditionally set in the Gaussian theory in order to simplify computations.

However, if an expert in morphometrics has set the Gaussian model for this experiment, then the above selection criteria is out of significance. Because, as we have shown in this landmark data, the moments-method estimates does not work properly under Gaussian models.

Now, at this stage, the conclusion about Kotz 5 model ratifies that non-Gaussian models explain better the three samples. An elliptical isotropic approach also verified this conclusion, see for example [8].

Once the model is selected, we are interested in application of Section 5, about estimation of mean form difference. In fact, we can go further with hypothesis testing for equality of the associated Euclidean Distance Matrices of two populations.

The methodology can be found in [31] and the references therein. We want to test $H_0 : \mathbf{F}(\boldsymbol{\mu}^{\mathbf{X}}) = c\mathbf{F}(\boldsymbol{\mu}^{\mathbf{Y}})$, for some $c > 0$, where $\boldsymbol{\mu}^{\mathbf{X}}$ and $\boldsymbol{\mu}^{\mathbf{Y}}$ are the population mean shape. Consider a sample of objects $\mathbf{X}$'s and $\mathbf{Y}$'s. The exact formulae of Theorem 3.6 give the estimated mean shapes $\widetilde{\boldsymbol{\mu}}^{X}$ and $\widetilde{\boldsymbol{\mu}}^{Y}$. Then we can derive the form difference matrix $\mathbf{FDM}\left(\widetilde{\boldsymbol{\mu}}^{\mathbf{X}}, \widetilde{\boldsymbol{\mu}}^{\mathbf{Y}}\right)$. This matrix can define a number of statistics for testing H_0; however, [31] recommend the following:

$$T = \max_{i,j} FDM_{ij} / \min_{i,j} FDM_{ij},$$

where FDM_{ij} is the $i,j-$element of matrix $\mathbf{FDM}$. Note that if H_0 is true T is close to 1. Moreover, T satisfies the desirable property of invariance under scaling, see [31] for more details.

The null distribution is difficult to obtain even in the Gaussian case. But, we can obtain an empirical null distribution by using a bootstrap procedure, see [31] and the references therein. For similar samples of the current example, those authors propose a bootstrap of size 100.

Once the empirical distribution is obtained, a p-value based on the upper tail of the observed statistics can be computed. It rejects H_0 for small values near to 0.1.

Table 6 reports the tests for the Gaussian case and the Kotz 5 model. It shows the p-values for the three possible pairs of groups in this experiment. Under the expected dependent condition of Theorem 3.6, the Gaussian model cannot detect the role of the control test, giving a wrong conclusion. The selected model with covariance distances, separates the control group but does not provide strong evidence of shape difference between small and large bones. This opens a discussion about the coordinate free approach of [31], because the pairwise-element quotient in the matrix form difference is neglecting some important information of this matrix. Then two challenges for a future work can be proposed: 1) Robust definitions of $\mathbf{FDM}$ and 2) exact distributions of $\mathbf{FDM}(\mathbf{X}, \mathbf{Y})$.

TABLE 6
p-values of testing the equality of mean shape under different models and pairs of populations.

	Small-Large	Small-Control	Large-Control
Gaussian	0.00	0.00	0.00
Kotz 5	0.12	0.51	0.74

7. Conclusions

1. This work proposes a theory for the estimation of mean form and mean form difference under elliptical laws. This approach models a wide range of real situations, with more or less heavy tails and more or less kurtosis than the Gaussian model.
2. The method-of-moments provides consistent estimators under the elliptical models.
3. Exact formulae, easy to compute, are given for the estimators.
4. Some research alternatives can be proposed. Instead of the assumptions in subsections 2.3 and 2.4, we can consider: Assume that the joint distribution of $\mathbf{E}_1, \mathbf{E}_2, \ldots, \mathbf{E}_n$ is

$$\mathbb{E} = (\mathbf{E}_1, \mathbf{E}_2, \ldots, \mathbf{E}_n) \sim \mathcal{E}_{K\times nD}(\mathbf{0}, \boldsymbol{\Sigma}_K \otimes \boldsymbol{\Sigma}_D \otimes \mathbf{I}_n, h),$$

   where $\operatorname{Cov}(\operatorname{vec} \mathbb{E}^T) = \boldsymbol{\Sigma}_K \otimes \boldsymbol{\Sigma}_D \otimes \mathbf{I}_n$. Then non central generalization of [19, Eq. 3.4.14, p. 109] and [23, Theorem 5.1.6, p. 170], provides the joint distribution of $\mathbf{B}_1, \mathbf{B}_2, \ldots, \mathbf{B}_n$ as follows:

$$\mathbb{B} = (\mathbf{B}_1, \mathbf{B}_2, \ldots, \mathbf{B}_n) \sim \mathcal{GPW}_{K,n}\left(\boldsymbol{\Sigma}_K^*, \frac{D}{2}, \frac{D}{2}, \ldots, \frac{D}{2}, \boldsymbol{\Omega}, h\right),$$

   where $\boldsymbol{\Omega} = \left(\boldsymbol{\Sigma}_K^*\right)^- \boldsymbol{\mu}^* \boldsymbol{\Sigma}_D^{-1} \boldsymbol{\mu}^{*T}$. Note that this case assumes a dependent sample $\mathbf{X}_1, \ldots, \mathbf{X}_n$.
5. Recall that the method-of-moments estimators are not uniquely defined, see Remark 3.3. Then the method-of-moments estimator of $\boldsymbol{\Sigma}_D$ can be obtained from the first two moments of $\mathbf{B}$.

Appendix A: Particular generalized Pseudo-Wishart singular distributions

The following result is a particular case of [13] or [15], when $\boldsymbol{\Theta}$ is non singular matrix.

Theorem A.1 (Generalized singular Pseudo-Wishart distributions)**.** *Assume that* $\mathbf{Y} \sim \mathcal{E}_{K\times D}^{K-1,D}(\boldsymbol{\mu}, \boldsymbol{\Sigma} \otimes \boldsymbol{\Theta}, h)$*, where* h *admits a power series expansion*

$$h(v + a) = \sum_{t=0}^{\infty} \frac{h^{(t)}(a) v^t}{t!}.$$

in $\Re$. Let $q=\min(K-1,D)$; then the density of $\mathbf{B}=\mathbf{Y}\mathbf{\Theta}^{-1}\mathbf{Y}^T$ is given by

$$= \frac{\pi^{qD/2}|\mathbf{L}|^{(D-K-1)/2}}{\Gamma_q[D/2]\left(\prod_{i=1}^{(K-1)}\lambda_i^{D/2}\right)}\sum_{t=0}^{\infty}\sum_{\kappa}\frac{h^{(2t)}(\mathrm{tr}(\mathbf{\Sigma}^{-}\mathbf{B}+\mathbf{\Omega}))}{t!}\frac{C_\kappa(\mathbf{\Omega}\mathbf{\Sigma}^{-}\mathbf{B})}{\left(\frac{1}{2}D\right)_\kappa}(d\mathbf{B}) \tag{A.1}$$

where $\mathbf{B}=\mathbf{W}_1\mathbf{L}\mathbf{W}_1^T$, is the nonsingular spectral decomposition of $\mathbf{B}$ with $\mathbf{W}_1$ a semiorthogonal matrix, i.e. $\mathbf{W}_1^T\mathbf{W}_1=\mathbf{I}_q$, and $\mathbf{L}=\mathrm{diag}(l_1,\dots,l_q)$; $\mathbf{\Omega}=\mathbf{\Sigma}^{-}\boldsymbol{\mu}\mathbf{\Theta}^{-1}\boldsymbol{\mu}^T$. $(d\mathbf{B})$ is the Hausdorff measure defined in [13, Section 5]. λ_i, $i=1,\dots,(K-1)$, are the non null eigenvalues of $\mathbf{\Sigma}$. $C_\kappa(\mathbf{A})$ are the zonal polynomials of $\mathbf{A}$ indexed by the partition $\kappa=(t_1,\dots,t_\alpha)$ of t, with $\sum_1^\alpha t_i=t$. $(a)_\kappa=\prod_{j=1}^{\alpha}(a-(j-1)/2)_{t_j}$, $(a)_t=a(a+1)\cdots(a+t-1)$ is the generalized hypergeometric coefficient and $\Gamma_s(a)=\pi^{s(s-1)/4}\prod_{j=1}^{s}\Gamma(a-(j-1)/2)$ is the multivariate gamma function, see [36];

Corollary A.1 (Singular Pseudo-Wishart Gaussian distribution)**.** *Assume that $\mathbf{Y}\sim\mathcal{N}_{K\times D}^{K-1,D}(\boldsymbol{\mu},\mathbf{\Sigma}\otimes\mathbf{\Theta})$, and let $q=\min(K-1,D)$; then the density of $\mathbf{B}=\mathbf{Y}\mathbf{\Theta}^{-1}\mathbf{Y}^T$ is given by*

$$= C\,\mathrm{etr}\left(-\frac{1}{2}(\mathbf{\Sigma}^{-}\mathbf{B}-\mathbf{\Omega})\right){}_0F_1\left(\frac{1}{2}D;\frac{1}{4}\mathbf{\Omega}\mathbf{\Sigma}^{-}\mathbf{B}\right)(d\mathbf{B}). \tag{A.2}$$

Here

$$C=\frac{\pi^{D(q-(K-1))/2}|\mathbf{L}|^{(D-K-1)/2}}{2^{D(K-1)/2}\Gamma_q[D/2]\left(\prod_{i=1}^{K-1}\lambda_i^{D/2}\right)},$$

where ${}_0F_1(\cdot)$ is a hypergeometric function with a matrix argument, see [36, p. 258].

Appendix B: Singular Pseudo-Wishart Kotz distribution

Firs recall that the $K\times D$ random matrix $\mathbf{X}$ is said to have a *singular matrix variate symmetric Kotz type distribution* with parameters $N,r,s\in\Re$, $\boldsymbol{\mu}:K\times D$, $\mathbf{\Sigma}:K\times K$, of rank $K-1$, $\mathbf{\Theta}:D\times D$ with $r>0$, $s>0$, $2N+(K-1)D>2$, $\mathbf{\Sigma}>\mathbf{0}$, and $\mathbf{\Theta}>\mathbf{0}$ if its density is

$$\frac{sr^{(2N+(K-1)D-2)/2s}\Gamma\left[(K-1)D/2\right]}{\pi^{(K-1)D/2}\Gamma\left[(2N+(K-1)D-2)/2s\right]\left(\prod_{i=1}^{K-1}\lambda_i^{D/2}\right)|\mathbf{\Theta}|^{(K-1)/2}}$$

$$\times\ \left[\mathrm{tr}\,\mathbf{\Theta}^{-1}(\mathbf{Y}-\boldsymbol{\mu})^T\mathbf{\Sigma}^{-}(\mathbf{Y}-\boldsymbol{\mu})\right]^{N-1}\exp\left\{-r\,\mathrm{tr}^s\,\mathbf{\Theta}^{-1}(\mathbf{Y}-\boldsymbol{\mu})^T\mathbf{\Sigma}^{-}(\mathbf{Y}-\boldsymbol{\mu})\right\}.$$

When $T=s=1$, and $R=1/2$ we get the singular matrix variate gaussian distribution.

Note that particular singular Pseudo-Wishart distributions just depend on the general derivative $h^{(2t)}(\cdot)$ of the elliptical generator function; it seems a trivial fact, but the general formulae involves cumbersome expressions indexed by partitions, see [4]. In the Kotz type distributions they derived the following expressions.

When $s = 1$, the Kotz type models and their general derivative simplify substantially. Thus, the following expressions applies for Gaussian, Kotz 1, and Kotz 2 models, with parameters $N = 1, s = 1, r = 1/2$; $N = 2, s = 1, r = 1/2$; $N = 3, s = 1, r = 1/2$; respectively. The generator model is given by

$$h(y) = \frac{r^{N-1+(K-1)D/2}\Gamma\left[(K-1)D/2\right]}{\pi^{(K-1)D/2}\Gamma\left[N-1+(K-1)D/2\right]} y^{N-1}\exp\{-ry\},$$

And, the corresponding k-th derivative of h, follows from

$$\frac{d^k}{dy^k} y^{N-1}\exp[-ry],$$

which is given by

$$(-r)^k y^{N-1}\exp[-ry]\left\{1+\sum_{v=1}^{k}\binom{k}{v}\left[\prod_{i=0}^{v-1}(N-1-i)\right](-ry)^{-v}\right\},$$

where $k = 2t$.

For the remaining models of the example, the so termed Kotz 3, Kotz 4 and Kotz 5, have parameters $N = 2, s = 2, r = 1/2$; $N = 2, s = 3, r = 1/2$ and $N = 20, s = 20, r = 1/2$, respectively. The generator function is given by:

$$h(y) = \frac{sr^{(2N+(K-1)D-2)/2s}\ \Gamma[(K-1)D/2]}{\pi^{(K-1)D/2}\Gamma\left[(2N+(K-1)D-2)/2s\right]}\ y^{N-1}\exp\left(-ry^s\right).$$

The required k-th derivative of h, follows from $\frac{d^k}{dy^k}\exp\left(-ry^s\right)$, which is given by

$$y^{T-1}e^{-Ry^s}\left\{\sum_{\kappa\in P_k}\frac{k!(-R)^{\sum_{i=1}^{k}v_i}\prod_{j=0}^{k-1}(s-j)^{\sum_{i=j+1}^{k}v_i}}{\prod_{i=1}^{k}v_i!(i!)^{v_i}}y^{\sum_{i=1}^{k}(s-i)v_i}\right.$$
$$+\sum_{m=1}^{k}\binom{k}{m}\left[\prod_{i=0}^{m-1}(T-1-i)\right]$$
$$\left.\times\sum_{\kappa\in P_{k-m}}\frac{(k-m)!(-R)^{\sum_{i=1}^{k-m}v_i}\prod_{j=0}^{k-m-1}(s-j)^{\sum_{i=j+1}^{k-m}v_i}}{\prod_{i=1}^{k-m}v_i!(i!)^{v_i}}y^{\sum_{i=1}^{k-m}(s-i)v_i-m}\right\},$$

where $\sum_{\kappa\in P_k}$ denotes the summation over all the partitions

$$\kappa = \left(k^{v_k}, (k-1)^{v_k-1}, \ldots, 3^{v_3, 2^{v_2}, 1^{v_1}}\right)$$

of k, with $\sum_{i=1}^{k} iv_i = k$, i.e. κ is a partition of k consisting of v_1 ones, v_2 twos, v_3 threes, etc. It is important to quote that all the singular Pseudo-Wishart distributions based on Kotz type kernels can be computed by some modifications of the Gaussian version in [27]. See for example [9] and similar works of the authors on shape theory.

Acknowledgments

The authors wish to thank the Editor and the anonymous reviewers for their constructive comments on the preliminary version of this paper. This article was written under the existing research agreement between the first author and the Universidad Autónoma Agraria Antonio Narro, Saltillo, México. F. Caro was supported by Universidad de Medellín, Colombia.

References

[1] Arnold, S. F. (1981). Theory of linear models and multivariate analysis, John Wiley & Sons, Inc., New York. MR0606011

[2] Bookstein, F. L. (1986). Size and shape spaces for landmark data in two dimensions (with discussion), *Stat. Sci.* **1** 181–242.

[3] Caro-Lopera, F. C. and Díaz-García, J. A. (2012). Matrix Kummer-Pearson VII relation and polynomial pearson VII configuration density, *J. Iranian Stat. Soc.* **11**(2) 217–230. MR3010346

[4] Caro-Lopera, F. J., Díaz-García, J. A., and González-Farías, G. (2009). Noncentral elliptical configuration density, *J. Multivariate Anal.* **101**(1) 32–43. MR2557616

[5] Caro-Lopera, F. J., Díaz-García, J. A., and González-Farías, G. (2014). Inference in affine shape theory under elliptical models, *J. Korean Stat. Soc.* **43**(1) 67–77. MR3173236

[6] Caro-Lopera, F. J., Leiva, V. and Balakrishnan. N. (2012). Connection between the Hadamard and matrix products with an application to matrix-variate Birnbaum-Saunders distributions, *J. Multivariate Anal.* **104** 126–139. MR2832191

[7] Díaz-García, J. A., and Caro-Lopera, F. J. (2012a). Generalised shape theory via SV decomposition I, *Metrika* **75**(4) 541–565. MR2922161

[8] Díaz-García, J. A., and Caro-Lopera, F. J. (2012b). Statistical theory of shape under elliptical models and singular value decompositions, *J. Multivariate Anal.* **103**(1) 77–92. MR2823710

[9] Díaz-García, J. A., and Caro-Lopera, F. J. (2013). Generalised shape theory via pseudo Wishart distribution, *Sankhyā A* **75**(2) 253–276. MR3092821

[10] Díaz-García, J. A., and Caro-Lopera, F. J. (2014). Statistical theory of shape under elliptical models via QR decomposition, *Statistics* **48**(2) 456–472. MR3175782

[11] Díaz-García, J. A. (1994). *Contributions to the theory of Wishart and multivariate elliptical distributions*, Ph.D. disertation, Universidad de Granada, Spain, (in Spanish).

[12] Díaz-García, J. A., and R. Gutiérrez Jáimez, R. (1996). *Matrix differential calculus and moments of a random matrix elliptical*, Serie Colección "Estadística Multivariable y Procesos Estocásticos". Universidad de Granada, España, (in Spanish).

[13] Díaz-García, J. A., and González-Farías, G. (2005). Singular random matrix decompositions: Distributions, *J. Multivariate Anal.* **94**(1) 109–122. MR2161213

[14] Díaz-García, J. A., Gutiérrez, R. J. and Ramos-Quiroga, R. (2003). Size-and-shape cone, shape disk and configuration densities for the elliptical models, *Brazilian J. Prob. Stati.* **17** 135–146. MR2099664

[15] Díaz-García, J. A., and Gutiérrez-Jáimez, R. (2006). Wishart and Pseudo-Wishart distributions under elliptical laws and related distributions in the shape theory context, *J. Statist. Plann. Inference* **136**(12) 4176–4193. MR2323409

[16] Dryden, I. L. and and Mardia, K. V. (1998). *Statistical shape analysis*, John Wiley and Sons, Chichester. MR1646114

[17] Dryden, I. L., Koloydenko, A., and Zhou D. (2009). Non-Euclidean statistics for covariance matrices, with applications to diffusion tensor imaging, *An. App. Statist.* **3** 1102–1123. MR2750388

[18] Dutilleul, P. (1999). The mle algorithm for the matrix normal distribution, *J. Statist. Comput. Simul.* **64** 105–123.

[19] Fang, K. T. and Zhang, Y. T. (1990). *Generalized multivariate analysis*, Science Press, Beijing, Springer-Verlang. MR1079542

[20] Fang, K. T., Kotz, S., and Ng, K. W. (1990). *Symmetric multivariate and related distributions*, Chapman Hall, London. MR1071174

[21] Goodall, C. R. (1991). Procustes methods in the statistical analysis of shape (with discussion), *J. Royal Stat. Soc. B* **53** 285–339. MR1108330

[22] Goodall, C. R., and Mardia, K. V. (1993). Multivariate aspects of shape theory, *Ann. Statist.* **21** 848–866. MR1232522

[23] Gupta, A. K., and Varga, T. (1993). *Elliptically contoured models in statistics*, Kluwer Academic Publishers, Dordrecht MR1256485

[24] Kent D. G. (1984). Shape manifolds, Procrustean metrics and complex projective spaces, *Bulletin of the London Mathematical Society* **16** 81–121. MR0737237

[25] Kent J. T. (1992). *New directions in shape analysis*, In Mardia, K. V., editor, The Art of Statistical Science, 115–127. Wiley, Chichester. MR1175661

[26] Khatri, C. G. (1968). Some results for the singular normal multivariate regression nodels, *Sankhyā A* **30** 267–280. MR0243692

[27] Koev, P., and Edelman, A. (2006). The efficient evaluation of the hypergeometric function of a matrix argument, *Math. Comp.* **75** 833–846. MR2196994

[28] Le, H. L., and Kendall, D. G. (1993). The Riemannian structure of

Euclidean spaces: A novel environment for statistics, *Ann. Statist.* **21** 1225–1271. MR1241266

[29] Lele, S. (1991). Some comments on coordinate free and scale invariant method in morphometrics, *Am. J. Phys. Anthropol.* **85** 407–418.

[30] Lele, S. (1993). Euclidean distance matrix analysis (EDMA): Estimation of mean form and mean form difference, *Math. Geol.* **25**(5) 573–602. MR1225013

[31] Lele, S., and Richtsmeier, J. (1991). Euclidean distance matrix analysis: A coordinate free approach for comparing biological shapes using landmark data, *Amer. J. Phys. Anthropol.* **86** 415–427.

[32] Lele, S., and Richtsmeier, J. (1990). Statistical models in morphometrics: Are they realistic? *Syst. Zool.* **39**(l) 60–69.

[33] Magnus, J. R., and Neudecker, H. (1979). The commutation matrix: Some properties and applications, *Ann. Statist.* **7**(2) 381–394. MR0520247

[34] Mardia, K. V., and Dryden, I. L. (1989) The statistical analysis of shape data, *Biometrika* **76** 271–281. MR1016017

[35] Mood, A. M., Graybill, F. A., and Boes, D. C. (1974). *Introduction to the theory statistics*, Third edition, McGraw-Hill Series in Probability and Statistics New York.

[36] Muirhead, R. J. (1982). *Aspects of multivariate statistical theory*, Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York. MR0652932

[37] Nadarajah, S. (2003). The Kotz-type distribution with applications, *Statistics* **37**(4) 341–358. MR1997185

[38] Rao, C. R. (1973). *Linear statistical inference and its applications*, John Wiley & Sons, New York, 1973. MR0346957

[39] Richtsmeier, J. T., Deleon, V. B., and Lele, S. R. (2002). The promise of geometric morphometrics, *Yearbook of Phis. Anthropol.* **45** 63–91.

[40] P. Sawyer. (1997). Spherical functions on symmetric cones, *Transactions of the American Mathematical Society* **349**, 3569–3584. MR1325919

[41] Walker, J. (2001). Ability of geometric morphometric methods to estimate a known covariance matrix, *Syst. Biol.* **49** 686–696.