REVSTAT – Statistical Journal
Volume 12, Number 2, June 2014, 101–118

UNIFORM APPROXIMATIONS FOR DISTRIBUTIONS OF CONTINUOUS RANDOM VARIABLES WITH APPLICATION IN DUAL STATIS METHOD

Authors: João Lita da Silva
– Department of Mathematics and CMA,
Faculty of Sciences and Technology, New University of Lisbon,
Quinta da Torre, 2829-516 Caparica, Lisbon, Portugal
jfls@fct.unl.pt

Luís Pedro Ramos
– Department of Mathematics and CMA,
Faculty of Sciences and Technology, New University of Lisbon,
Quinta da Torre, 2829-516 Caparica, Lisbon, Portugal
lpcr@fct.unl.pt

Received: December 2012 Revised: March 2013 Accepted: March 2013

Abstract:

- The matrix $\mathbf{S} = \big[\operatorname{tr}(\mathbf{W}_i\mathbf{Q}\mathbf{W}_j\mathbf{Q})\big]_{i,j=1,...,k}$ where $\mathbf{Q}$ is a symmetric positive definite matrix and $\mathbf{W}_i = \mathbf{X}_i'\mathbf{D}_i\mathbf{X}_i$, $i = 1, ..., k$ is formed by data tables $\mathbf{X}_i$ and diagonal matrices of weights $\mathbf{D}_i$, plays a central role in dual STATIS method. In this paper, we approximate the distribution function of the entries of $\mathbf{S}$, assuming data tables $\mathbf{X}_i$ given by $\mathbf{U}_i + \mathbf{E}_i$, $i = 1, ..., k$ with independent random matrices $\mathbf{E}_i$ representing errors, in order to obtain (approximately) the distribution of $\mathbf{Sv}$, where $\mathbf{v}$ is the orthonormal eigenvector of $\mathbf{S}$ associated to the largest eigenvalue. To achieve this goal, we approximate uniformly the distribution of each entry of $\mathbf{S}$. In general, our technique consists in to approximate uniformly the distribution sequence $\big\{g(\boldsymbol{V}_n + \boldsymbol{\mu}_n), n \geqslant 1\big\}$, where g is some smooth function of several variables, $\{\boldsymbol{V}_n,\, n \geqslant 1\}$ is a sequence of identically distributed random vectors of continuous type and $\{\boldsymbol{\mu}_n\}$ is a non-random vector sequence.

Key-Words:

- *dual STATIS method; uniform approximations.*

AMS Subject Classification:

- 62H25, 60E05(62E20), 60F05.

1. INTRODUCTION

The dual STATIS method is an exploratory technique of multivariate data analysis used to study simultaneously multiple data tables, each with information of groups of individuals measured on the same set of variables (see [4], [8] or [9]). The purpose of this method is to analyze the relationship between data tables and combine them into a compromise matrix corresponding to an optimal agreement between the data.

In this paper, we shall consider the n_i-by-p random data tables $\mathbf{X}_i$, $i = 1, ..., k$ consisting in the measurements of k groups of n_i individuals on the same set of p variables, the n_i-by-n_i diagonal matrices $\mathbf{D}_i$ of positive weights attached to the n_i observations of each matrix $\mathbf{X}_i$ in order to define

$$\mathbf{W}_i = \mathbf{X}_i'\mathbf{D}_i\mathbf{X}_i \; , \tag{1.1}$$

where prime denotes transpose. If the columns of $\mathbf{X}_i$ are $\mathbf{D}_i$-centered then $\mathbf{W}_i$ is the covariance matrix between the p variables of $\mathbf{X}_i$ and its elements corresponds to the scalar products between the variables in $\mathbb{R}^{n_i}$. To evaluate the closeness of two data configurations in $\mathbb{R}^{n_i}$ and $\mathbb{R}^{n_j}$, the trace $\mathrm{tr}(\mathbf{W}_i\mathbf{Q}\mathbf{W}_j\mathbf{Q})$, where $\mathbf{Q}$ is a p-by-p symmetric positive definite matrix, is commonly used as scalar product between $\mathbf{W}_i$ and $\mathbf{W}_j$, known as *Hilbert–Schmidt scalar product* between $\mathbf{W}_i$ and $\mathbf{W}_j$ (see [8], page 38).

We shall set $s_{ij} = \mathrm{tr}(\mathbf{W}_i\mathbf{Q}\mathbf{W}_j\mathbf{Q})$, $i, j = 1, ..., k$ as being the entries of the k-by-k interstructure matrix $\mathbf{S}$. The vectorial correlation coefficient RV of $\mathbf{W}_i$ and $\mathbf{W}_j$ is defined as

$$\mathrm{RV}(\mathbf{W}_i, \mathbf{W}_j) = \frac{\mathrm{tr}(\mathbf{W}_i\mathbf{Q}\mathbf{W}_j\mathbf{Q})}{\sqrt{\mathrm{tr}(\mathbf{W}_i\mathbf{Q}\mathbf{W}_i\mathbf{Q}) \; \mathrm{tr}(\mathbf{W}_j\mathbf{Q}\mathbf{W}_j\mathbf{Q})}} \; .$$

(see [4]), which appears as a measure of similarity between $\mathbf{W}_i$ and $\mathbf{W}_j$. The reader is referred to [13] for further details on the RV coefficient. Moreover, from Cholesky decomposition (see [5], page 229), there exists a unique upper triangular p-by-p matrix $\mathbf{T}$ with positive diagonal elements such that $\mathbf{Q} = \mathbf{T}'\mathbf{T}$ and putting $\mathbf{A}_i = \mathbf{D}_i^{1/2}\mathbf{X}_i\mathbf{T}'$ we get

$$\mathbf{A}_i'\mathbf{A}_i = \mathbf{T}\mathbf{X}_i'\mathbf{D}_i\mathbf{X}_i\mathbf{T}'$$

which implies

$$\mathrm{tr}(\mathbf{W}_i\mathbf{Q}\mathbf{W}_j\mathbf{Q}) = \mathrm{tr}(\mathbf{A}_i'\mathbf{A}_i\mathbf{A}_j'\mathbf{A}_j) = \left|\!\left|\!\left|\mathbf{A}_i\mathbf{A}_j'\right|\!\right|\!\right|_{\mathrm{tr}}^2 \geqslant 0$$

where $|||\mathbf{A}|||_{\mathrm{tr}} = \sqrt{\mathrm{tr}(\mathbf{A}'\mathbf{A})}$ (see [5], page 60). Denoting by $\mathbf{a}_{i\ell}$, $\ell = 1, ..., n_i$ the rows of $\mathbf{A}_i$ and $\mathbf{a}_{jm}$, $m = 1, ..., n_j$ the rows of $\mathbf{A}_j$, $\mathbf{a}_{i\ell}\mathbf{a}_{jm}'$ is the covariance between $\mathbf{a}_{i\ell}$ and $\mathbf{a}_{jm}$ so that

$$\mathrm{tr}(\mathbf{W}_i\mathbf{Q}\mathbf{W}_j\mathbf{Q}) = \sum_{\ell=1}^{n_i}\sum_{m=1}^{n_j}\left[\mathrm{cov}(\mathbf{a}_{i\ell}, \mathbf{a}_{jm})\right]^2 .$$

Consider the eigenvalues $\rho_1 \geqslant \rho_2 ... \geqslant \rho_k$ of $\mathbf{S}$ and the corresponding orthonormal eigenvectors $\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_k$. From the spectral theorem for symmetric matrices (see [7], page 104) we get

$$\mathbf{S} = \rho_1 \mathbf{v}_1 \mathbf{v}_1' + ... + \rho_k \mathbf{v}_k \mathbf{v}_k' = \mathbf{P}\mathbf{\Lambda}\mathbf{P}'$$

where $\mathbf{\Lambda}$ is the diagonal matrix whose elements are the corresponding eigenvalues and $\mathbf{P}'\mathbf{P} = \mathbf{I}$, where $\mathbf{I}$ is the k-by-k identity matrix. From the above expression, we can plot the ith stage in $\mathbb{R}^k$ as a point M_i whose coordinates are the ith row of $\mathbf{P}\mathbf{\Lambda}^{1/2}$ (see [9]). Along all text, we will assume that points M_i are all near by each other lying on around first axis, which lead us to $\rho_i \approx 0$, for each $i = 2, ..., k$ and

$$\mathbf{S} \approx \rho \mathbf{v}\mathbf{v}' \tag{1.2}$$

setting $\rho = \rho_1$ and $\mathbf{v} = \mathbf{v}_1$. In our model, we will also assume random errors on the data i.e.

$$\mathbf{X}_i = \mathbf{U}_i + \mathbf{E}_i \tag{1.3}$$

where $\mathbf{E}_i$ are independent n_i-by-p random matrices representing the errors with i.i.d. continuous entries and $\mathbf{U}_i$ are n_i-by-p non-random matrices. Moreover, we shall admit that $\mathrm{E}(\mathbf{S})$ has rank one, so that the spectral theorem for symmetric matrices allow us to write

$$\mathrm{E}(\mathbf{S}) = \lambda \boldsymbol{\alpha}\boldsymbol{\alpha}' \tag{1.4}$$

where λ is the largest eigenvalue of $\mathrm{E}(\mathbf{S})$ associated to the orthonormal eigenvector $\boldsymbol{\alpha}$. Hence, we are led to consider the model

$$\mathbf{S} = \lambda \boldsymbol{\alpha}\boldsymbol{\alpha}' + \boldsymbol{\mathcal{E}} \tag{1.5}$$

for some p-by-p random matrix $\boldsymbol{\mathcal{E}}$ satisfying $\mathrm{E}(\boldsymbol{\mathcal{E}}) = \mathbf{O}$ (null matrix).

Let us start with the following question: if the sequence of matrices $\mathbf{E}_i$, $i = 1, 2, ..., k$ are independent with i.i.d. continuous entries how can we compute the distribution function of each entry s_{ij} of the matrix $\mathbf{S}$? Generally, the distribution function of s_{ij} is hard to compute, so that our proposal answer to this question will be to approximate the distribution of s_{ij} by some computable distribution. More precisely, our results will permit us to approximate uniformly the distribution function of each entry of the random matrix $\mathbf{S}$ by its linear part. The Section 2 will describe in detail all the theoretical results required to fulfill our intentions. Once the distribution of the elements of $\mathbf{S}$ is achieved, we will be able to obtain (approximately) the distribution of $\widehat{\boldsymbol{\beta}} = \mathbf{S}\mathbf{v}$, which will be taken as an estimator of $\boldsymbol{\beta} = \lambda\boldsymbol{\alpha}$. The example exhibited in last section considering the elements of $\mathbf{E}_i$ i.i.d. normal distributed with zero mean and variance σ^2, it will illustrate our inferential purposes in a very clear way.

2. UNIFORM APPROXIMATIONS

In general way, our idea to approaching the distribution function of the entries of **S**, will consist in expanding asymptotically a sequence of r.v.'s (with unknown distribution) to obtain a random sequence with identifiable distribution. Thereafter, with the aid of a uniform bound, we will establish the uniform approximation results imposing on the remainder term the asymptotic condition $o_{\mathrm{Pr}}(1)$.[1] The driving tool in the our proof technique is to consider asymptotic Taylor expansions.

Let us introduce the following notation: $\|\boldsymbol{a}\| = \left(\sum_{i=1}^{\kappa} a_i^2\right)^{1/2}$ for the Euclidean norm of a vector $\boldsymbol{a} \in \mathbb{R}^\kappa$ (see [7], page 264) and $\|\!|\boldsymbol{A}|\!\| = \left(\sum_{i,j=1}^{\kappa} a_{ij}^2\right)^{1/2}$ for the norm of a real matrix $\boldsymbol{A} = \left[a_{ij}\right]_{i,j=1,\dots,\kappa}$ (known as *Frobenius norm*, see [7] page 291). Given a differentiable mapping $\boldsymbol{\varphi}\colon \mathbb{R}^N \to \mathbb{R}^M$ we shall denote the jacobian matrix of $\boldsymbol{\varphi}$ at the point $\boldsymbol{x}$ by $D\boldsymbol{\varphi}(\boldsymbol{x})$. For a differentiable real-valued function $\varphi\colon \mathbb{R}^N \to \mathbb{R}$ having second partial derivatives, $D\varphi(\boldsymbol{x})$ and $D^2\varphi(\boldsymbol{x})$ will denote, respectively, the gradient vector and the Hessian matrix of φ at the point $\boldsymbol{x}$.

Lemma 2.1. *If* $\boldsymbol{\varphi} \in C^1(\mathbb{R}^N, \mathbb{R}^N)$ *satisfies* $\|\!|D\boldsymbol{\varphi}(\boldsymbol{x})|\!\| = o\left(\|\boldsymbol{\varphi}(\boldsymbol{x})\|\right)$, $\|\boldsymbol{x}\| \to \infty$ *then* $\left\|\!\left|D\boldsymbol{\varphi}\big(\boldsymbol{x}+\boldsymbol{\theta}(\boldsymbol{x})\big)\right|\!\right\| = o\left(\|\boldsymbol{\varphi}(\boldsymbol{x})\|\right)$, $\|\boldsymbol{x}\| \to \infty$ *for any bounded function* $\boldsymbol{\theta}(\boldsymbol{x})$.

Proof: We have,

$$\frac{\left\|\!\left|D\boldsymbol{\varphi}\big(\boldsymbol{x}+\boldsymbol{\theta}(\boldsymbol{x})\big)\right|\!\right\|}{\|\boldsymbol{\varphi}(\boldsymbol{x})\|} = \frac{\left\|\!\left|D\boldsymbol{\varphi}\big(\boldsymbol{x}+\boldsymbol{\theta}(\boldsymbol{x})\big)\right|\!\right\|}{\left\|\boldsymbol{\varphi}\big(\boldsymbol{x}+\boldsymbol{\theta}(\boldsymbol{x})\big)\right\|}\,\frac{\left\|\boldsymbol{\varphi}\big(\boldsymbol{x}+\boldsymbol{\theta}(\boldsymbol{x})\big)\right\|}{\|\boldsymbol{\varphi}(\boldsymbol{x})\|}$$

and it is sufficient to prove that $\left\|\boldsymbol{\varphi}\big(\boldsymbol{x}+\boldsymbol{\theta}(\boldsymbol{x})\big)\right\| \sim \|\boldsymbol{\varphi}(\boldsymbol{x})\|$ as $\|\boldsymbol{x}\| \to \infty$. Considering $h\colon [0,1] \to \mathbb{R}$ defined by

$$h(t) = \log\left\|\boldsymbol{\varphi}\big(\boldsymbol{x}+t\boldsymbol{\theta}(\boldsymbol{x})\big)\right\|$$

and setting $\boldsymbol{\varphi}(\boldsymbol{x}) = \big(\varphi_1(\boldsymbol{x}), \dots, \varphi_N(\boldsymbol{x})\big)$ we get

$$\tfrac{d}{dt}h(t) = \tfrac{\varphi_1(\boldsymbol{x}+t\boldsymbol{\theta}(\boldsymbol{x}))\,\boldsymbol{\theta}(\boldsymbol{x})\cdot D\varphi_1(\boldsymbol{x}+t\boldsymbol{\theta}(\boldsymbol{x}))+\dots+\varphi_N(\boldsymbol{x}+t\boldsymbol{\theta}(\boldsymbol{x}))\,\boldsymbol{\theta}(\boldsymbol{x})\cdot D\varphi_N(\boldsymbol{x}+t\boldsymbol{\theta}(\boldsymbol{x}))}{\|\boldsymbol{\varphi}(\boldsymbol{x}+t\boldsymbol{\theta}(\boldsymbol{x}))\|^2}.$$

Hence,

$$\begin{aligned}\left|\frac{d}{dt}h(t)\right| &\leqslant \tfrac{|\varphi_1(\boldsymbol{x}+t\boldsymbol{\theta}(\boldsymbol{x}))|\,\|\boldsymbol{\theta}(\boldsymbol{x})\|\|D\varphi_1(\boldsymbol{x}+t\boldsymbol{\theta}(\boldsymbol{x}))\|+\dots+|\varphi_N(\boldsymbol{x}+t\boldsymbol{\theta}(\boldsymbol{x}))|\,\|\boldsymbol{\theta}(\boldsymbol{x})\|\|D\varphi_N(\boldsymbol{x}+t\boldsymbol{\theta}(\boldsymbol{x}))\|}{\|\boldsymbol{\varphi}(\boldsymbol{x}+t\boldsymbol{\theta}(\boldsymbol{x}))\|^2}\\ &\leqslant N\,\|\boldsymbol{\theta}(\boldsymbol{x})\|\,\tfrac{\|\!|D\boldsymbol{\varphi}(\boldsymbol{x}+t\boldsymbol{\theta}(\boldsymbol{x}))|\!\|}{\|\boldsymbol{\varphi}(\boldsymbol{x}+t\boldsymbol{\theta}(\boldsymbol{x}))\|}\end{aligned}$$

[1] $X_n = o_{\mathrm{Pr}}(1)$ means $X_n \xrightarrow{\mathrm{Pr}} 0$ as $n \to \infty$.

and mean value theorem lead us to

$$\begin{aligned}\left|\log\left\|\boldsymbol{\varphi}\big(\boldsymbol{x}+\boldsymbol{\theta}(\boldsymbol{x})\big)\right\| - \log\|\boldsymbol{\varphi}(\boldsymbol{x})\|\right| &= |h(1)-h(0)| = \\ &= \left|\frac{dh}{dt}(c)\right| \leqslant \kappa\,\|\boldsymbol{\theta}(\boldsymbol{x})\|\,\frac{\left\|\!\left| D\boldsymbol{\varphi}\big(\boldsymbol{x}+c\boldsymbol{\theta}(\boldsymbol{x})\big)\right|\!\right\|}{\left\|\boldsymbol{\varphi}\big(\boldsymbol{x}+c\boldsymbol{\theta}(\boldsymbol{x})\big)\right\|}\,, \qquad 0<c<1\end{aligned}$$

which implies $\left\|\boldsymbol{\varphi}\big(\boldsymbol{x}+\boldsymbol{\theta}(\boldsymbol{x})\big)\right\| \sim \|\boldsymbol{\varphi}(\boldsymbol{x})\|$ as $\|\boldsymbol{x}\|\to\infty$. □

Next, we present the main uniform approximation result.

Theorem 2.1. *Let $\boldsymbol{V}_n=(V_{1n},\dots,V_{Nn})$ be a sequence of random vectors of continuous type such that $\{V_{in},\, n\geqslant 1\}$ $(1\leqslant i\leqslant N)$ is identically distributed and $\sup_{n\geqslant 1}\|\boldsymbol{V}_n\|\leqslant W$ for some r.v. W. If $X_n := g(\boldsymbol{V}_n+\boldsymbol{\mu}_n)$ where $\boldsymbol{\mu}_n$ is a non-random vector sequence verifying $\|\boldsymbol{\mu}_n\|\to\infty$ and g is a $C^2(\mathbb{R}^N)$ map such that $\dfrac{Dg(\boldsymbol{t})}{\|Dg(\boldsymbol{t})\|}$ exists as $\|\boldsymbol{t}\|\to\infty$ and $\left\|\!\left|D^2g(\boldsymbol{t})\right|\!\right\| = o\big(\|Dg(\boldsymbol{t})\|\big)$, $\|\boldsymbol{t}\|\to\infty$ then, with $Y_n := g(\boldsymbol{\mu}_n)+Dg(\boldsymbol{\mu}_n)\cdot\boldsymbol{V}_n$, the law of X_n is uniformly approximate by the law of Y_n for large values of n, that is,*

$$\sup_{x\in\mathbb{R}}|F_{X_n}(x)-F_{Y_n}(x)| \to 0\ .$$

Proof: Using the Taylor formula for g we get

$$\begin{aligned}X_n := g(\boldsymbol{\mu}_n+\boldsymbol{V}_n) &= g(\boldsymbol{\mu}_n)+Dg(\boldsymbol{\mu}_n)\cdot\boldsymbol{V}_n+\frac{1}{2}\,D^2g(\boldsymbol{\mu}_n+\theta_n\boldsymbol{V}_n)\cdot\boldsymbol{V}_n^2 = \\ &= Y_n+\frac{1}{2}\,D^2g(\boldsymbol{\mu}_n+\theta_n\boldsymbol{V}_n)\cdot\boldsymbol{V}_n^2\,, \qquad 0<\theta_n<1\end{aligned}$$

where $Dg(\boldsymbol{a})\cdot\boldsymbol{V}_n=\sum_i \dfrac{\partial g}{\partial x_i}(\boldsymbol{a})V_{in}$ and $D^2g(\boldsymbol{a})\cdot\boldsymbol{V}_n^2=\sum_{i,j}\dfrac{\partial^2 g}{\partial x_i\partial x_j}(\boldsymbol{a})V_{in}V_{jn}$ (see [11], page 150). For $\varepsilon>0$ fixed we have

$$\begin{aligned}\Pr\{X_n\leqslant x\} &= \Pr\left\{X_n\leqslant x,\, \frac{|X_n-Y_n|}{\|Dg(\boldsymbol{\mu}_n)\|}\leqslant\varepsilon\right\}+\Pr\left\{X_n\leqslant x,\, \frac{|X_n-Y_n|}{\|Dg(\boldsymbol{\mu}_n)\|}>\varepsilon\right\} \\ &\leqslant \Pr\left\{Y_n\leqslant x+\varepsilon\,\|Dg(\boldsymbol{\mu}_n)\|\right\}+\Pr\left\{\frac{|X_n-Y_n|}{\|Dg(\boldsymbol{\mu}_n)\|}>\varepsilon\right\}\end{aligned}$$

and

$$\begin{aligned}\Pr\left\{Y_n\leqslant x-\varepsilon\,\|Dg(\boldsymbol{\mu}_n)\|\right\} &= \Pr\left\{Y_n\leqslant x-\varepsilon\,\|Dg(\boldsymbol{\mu}_n)\|\,, X_n\leqslant x\right\}+ \\ &\quad +\Pr\left\{Y_n\leqslant x-\varepsilon\,\|Dg(\boldsymbol{\mu}_n)\|\,, X_n>x\right\} \\ &\leqslant \Pr\{X_n\leqslant x\}+\Pr\left\{\frac{|X_n-Y_n|}{\|Dg(\boldsymbol{\mu}_n)\|}>\varepsilon\right\}\end{aligned}$$

that is,

$$\begin{aligned}-\Pr\left\{\tfrac{|X_n-Y_n|}{\|Dg(\boldsymbol{\mu}_n)\|}>\varepsilon\right\}+F_{Dg(\boldsymbol{\mu}_n)\cdot\boldsymbol{V}_1}\left(x-g(\boldsymbol{\mu}_n)-\varepsilon\,\|Dg(\boldsymbol{\mu}_n)\|\right) &\leqslant F_{X_n}(x)\leqslant \\ \leqslant F_{Dg(\boldsymbol{\mu}_n)\cdot\boldsymbol{V}_1}\left(x-g(\boldsymbol{\mu}_n)+\varepsilon\,\|Dg(\boldsymbol{\mu}_n)\|\right)+\Pr\left\{\tfrac{|X_n-Y_n|}{\|Dg(\boldsymbol{\mu}_n)\|}>\varepsilon\right\}.\end{aligned}$$

We can rewrite the above inequalities as

$$-\Pr\left\{\tfrac{|X_n-Y_n|}{\|Dg(\boldsymbol{\mu}_n)\|}>\varepsilon\right\}+F_{\frac{Dg(\boldsymbol{\mu}_n)\cdot\boldsymbol{V}_1}{\|g(\boldsymbol{\mu}_n)\|}}\left(\tfrac{x-g(\boldsymbol{\mu}_n)}{\|Dg(\boldsymbol{\mu}_n)\|}-\varepsilon\right)\leqslant F_{X_n}(x)\leqslant$$
$$\leqslant F_{\frac{Dg(\boldsymbol{\mu}_n)\cdot\boldsymbol{V}_1}{\|Dg(\boldsymbol{\mu}_n)\|}}\left(\tfrac{x-g(\boldsymbol{\mu}_n)}{\|Dg(\boldsymbol{\mu}_n)\|}+\varepsilon\right)+\Pr\left\{\tfrac{|X_n-Y_n|}{\|Dg(\boldsymbol{\mu}_n)\|}>\varepsilon\right\}.$$

Therefore,

$$|F_{X_n}(x)-F_{Y_n}(x)|\leqslant\left|F_{\frac{Dg(\boldsymbol{\mu}_n)\cdot\boldsymbol{V}_1}{\|Dg(\boldsymbol{\mu}_n)\|}}\left(\tfrac{x-g(\boldsymbol{\mu}_n)}{\|Dg(\boldsymbol{\mu}_n)\|}+\varepsilon\right)-F_{\frac{Dg(\boldsymbol{\mu}_n)\cdot\boldsymbol{V}_1}{\|Dg(\boldsymbol{\mu}_n)\|}}\left(\tfrac{x-g(\boldsymbol{\mu}_n)}{\|Dg(\boldsymbol{\mu}_n)\|}\right)\right|+$$
$$+\left|F_{\frac{Dg(\boldsymbol{\mu}_n)\cdot\boldsymbol{V}_1}{\|Dg(\boldsymbol{\mu}_n)\|}}\left(\tfrac{x-g(\boldsymbol{\mu}_n)}{\|Dg(\boldsymbol{\mu}_n)\|}\right)-F_{\frac{Dg(\boldsymbol{\mu}_n)\cdot\boldsymbol{V}_1}{\|Dg(\boldsymbol{\mu}_n)\|}}\left(\tfrac{x-g(\boldsymbol{\mu}_n)}{\|Dg(\boldsymbol{\mu}_n)\|}-\varepsilon\right)\right|+\Pr\left\{\tfrac{|X_n-Y_n|}{\|Dg(\boldsymbol{\mu}_n)\|}>\varepsilon\right\}$$

and we obtain the following uniform bound,

(2.1)

$$\sup_{x\in\mathbb{R}}|F_{X_n}(x)-F_{Y_n}(x)|\leqslant\sup_{x\in\mathbb{R}}\left|F_{\frac{Dg(\boldsymbol{\mu}_n)\cdot\boldsymbol{V}_1}{\|Dg(\boldsymbol{\mu}_n)\|}}\left(\tfrac{x-g(\boldsymbol{\mu}_n)}{\|Dg(\boldsymbol{\mu}_n)\|}+\varepsilon\right)-F_{\frac{Dg(\boldsymbol{\mu}_n)\cdot\boldsymbol{V}_1}{\|Dg(\boldsymbol{\mu}_n)\|}}\left(\tfrac{x-g(\boldsymbol{\mu}_n)}{\|Dg(\boldsymbol{\mu}_n)\|}\right)\right|+$$
$$+\sup_{x\in\mathbb{R}}\left|F_{\frac{Dg(\boldsymbol{\mu}_n)\cdot\boldsymbol{V}_1}{\|Dg(\boldsymbol{\mu}_n)\|}}\left(\tfrac{x-g(\boldsymbol{\mu}_n)}{\|Dg(\boldsymbol{\mu}_n)\|}\right)-F_{\frac{Dg(\boldsymbol{\mu}_n)\cdot\boldsymbol{V}_1}{\|Dg(\boldsymbol{\mu}_n)\|}}\left(\tfrac{x-g(\boldsymbol{\mu}_n)}{\|Dg(\boldsymbol{\mu}_n)\|}-\varepsilon\right)\right|+\Pr\left\{\tfrac{|X_n-Y_n|}{\|Dg(\boldsymbol{\mu}_n)\|}>\varepsilon\right\}$$

Since

$$\frac{Dg(\boldsymbol{\mu}_n)\cdot\boldsymbol{V}_1}{\|Dg(\boldsymbol{\mu}_n)\|}\xrightarrow{\mathrm{d}}\boldsymbol{\tau}\cdot\boldsymbol{V}_1\qquad(\boldsymbol{\tau}\neq\mathbf{0})$$

and $\boldsymbol{\tau}\cdot\boldsymbol{V}_1=\sum_i\tau_iV_{i1}$ is continuous then Polya's theorem (see [3], page 3) states

$$F_{\frac{Dg(\boldsymbol{\mu}_n)\cdot\boldsymbol{V}_1}{\|Dg(\boldsymbol{\mu}_n)\|}}\longrightarrow F_{\boldsymbol{\tau}\cdot\boldsymbol{V}_1}\ \text{ uniformly on }\mathbb{R}.$$

Hence, we can still write

$$\sup_{x\in\mathbb{R}}\left|F_{\frac{Dg(\boldsymbol{\mu}_n)\cdot\boldsymbol{V}_1}{\|Dg(\boldsymbol{\mu}_n)\|}}\left(\frac{x-g(\boldsymbol{\mu}_n)}{\|Dg(\boldsymbol{\mu}_n)\|}+\varepsilon\right)-F_{\frac{Dg(\boldsymbol{\mu}_n)\cdot\boldsymbol{V}_1}{\|Dg(\boldsymbol{\mu}_n)\|}}\left(\frac{x-g(\boldsymbol{\mu}_n)}{\|Dg(\boldsymbol{\mu}_n)\|}\right)\right|\leqslant$$
$$\leqslant\sup_{x\in\mathbb{R}}\left|F_{\frac{Dg(\boldsymbol{\mu}_n)\cdot\boldsymbol{V}_1}{\|Dg(\boldsymbol{\mu}_n)\|}}\left(\frac{x-g(\boldsymbol{\mu}_n)}{\|Dg(\boldsymbol{\mu}_n)\|}+\varepsilon\right)-F_{\boldsymbol{\tau}\cdot\boldsymbol{V}_1}\left(\frac{x-g(\boldsymbol{\mu}_n)}{\|Dg(\boldsymbol{\mu}_n)\|}+\varepsilon\right)\right|+$$
$$+\sup_{x\in\mathbb{R}}\left|F_{\boldsymbol{\tau}\cdot\boldsymbol{V}_1}\left(\frac{x-g(\boldsymbol{\mu}_n)}{\|Dg(\boldsymbol{\mu}_n)\|}+\varepsilon\right)-F_{\boldsymbol{\tau}\cdot\boldsymbol{V}_1}\left(\frac{x-g(\boldsymbol{\mu}_n)}{\|Dg(\boldsymbol{\mu}_n)\|}\right)\right|+$$
$$+\sup_{x\in\mathbb{R}}\left|F_{\boldsymbol{\tau}\cdot\boldsymbol{V}_1}\left(\frac{x-g(\boldsymbol{\mu}_n)}{\|Dg(\boldsymbol{\mu}_n)\|}\right)-F_{\frac{Dg(\boldsymbol{\mu}_n)\cdot\boldsymbol{V}_1}{\|Dg(\boldsymbol{\mu}_n)\|}}\left(\frac{x-g(\boldsymbol{\mu}_n)}{\|Dg(\boldsymbol{\mu}_n)\|}\right)\right|$$

and

$$\sup_{x\in\mathbb{R}}\left|F_{\frac{Dg(\boldsymbol{\mu}_n)\cdot\boldsymbol{V}_1}{\|Dg(\boldsymbol{\mu}_n)\|}}\left(\frac{x-g(\boldsymbol{\mu}_n)}{\|Dg(\boldsymbol{\mu}_n)\|}\right)-F_{\frac{Dg(\boldsymbol{\mu}_n)\cdot\boldsymbol{V}_1}{\|Dg(\boldsymbol{\mu}_n)\|}}\left(\frac{x-g(\boldsymbol{\mu}_n)}{\|Dg(\boldsymbol{\mu}_n)\|}-\varepsilon\right)\right|\leqslant$$
$$\leqslant\sup_{x\in\mathbb{R}}\left|F_{\frac{Dg(\boldsymbol{\mu}_n)\cdot\boldsymbol{V}_1}{\|Dg(\boldsymbol{\mu}_n)\|}}\left(\frac{x-g(\boldsymbol{\mu}_n)}{\|Dg(\boldsymbol{\mu}_n)\|}\right)-F_{\boldsymbol{\tau}\cdot\boldsymbol{V}_1}\left(\frac{x-g(\boldsymbol{\mu}_n)}{\|Dg(\boldsymbol{\mu}_n)\|}\right)\right|+$$
$$+\sup_{x\in\mathbb{R}}\left|F_{\boldsymbol{\tau}\cdot\boldsymbol{V}_1}\left(\frac{x-g(\boldsymbol{\mu}_n)}{\|Dg(\boldsymbol{\mu}_n)\|}\right)-F_{\boldsymbol{\tau}\cdot\boldsymbol{V}_1}\left(\frac{x-g(\boldsymbol{\mu}_n)}{\|Dg(\boldsymbol{\mu}_n)\|}-\varepsilon\right)\right|+$$
$$+\sup_{x\in\mathbb{R}}\left|F_{\boldsymbol{\tau}\cdot\boldsymbol{V}_1}\left(\frac{x-g(\boldsymbol{\mu}_n)}{\|Dg(\boldsymbol{\mu}_n)\|}-\varepsilon\right)-F_{\frac{Dg(\boldsymbol{\mu}_n)\cdot\boldsymbol{V}_1}{\|Dg(\boldsymbol{\mu}_n)\|}}\left(\frac{x-g(\boldsymbol{\mu}_n)}{\|Dg(\boldsymbol{\mu}_n)\|}-\varepsilon\right)\right|.$$

Choosing $\varepsilon > 0$ small enough, we get for each $\eta > 0$

$$\sup_{x\in\mathbb{R}}\left|F_{\boldsymbol{\tau}\cdot\boldsymbol{V}_1}\left(\frac{x-g(\boldsymbol{\mu}_n)}{\|Dg(\boldsymbol{\mu}_n)\|}+\varepsilon\right)-F_{\boldsymbol{\tau}\cdot\boldsymbol{V}_1}\left(\frac{x-g(\boldsymbol{\mu}_n)}{\|Dg(\boldsymbol{\mu}_n)\|}\right)\right|<\eta$$

and

$$\sup_{x\in\mathbb{R}}\left|F_{\boldsymbol{\tau}\cdot\boldsymbol{V}_1}\left(\frac{x-g(\boldsymbol{\mu}_n)}{\|Dg(\boldsymbol{\mu}_n)\|}\right)-F_{\boldsymbol{\tau}\cdot\boldsymbol{V}_1}\left(\frac{x-g(\boldsymbol{\mu}_n)}{\|Dg(\boldsymbol{\mu}_n)\|}-\varepsilon\right)\right|<\eta$$

provided that $F_{\boldsymbol{\tau}\cdot\boldsymbol{V}_1}$ is uniformly continuous. Since $\sup\limits_{n\geqslant 1}\|\theta_n\boldsymbol{V}_n\|\leqslant W$ we obtain from Lemma 2.1,

$$\frac{|X_n-Y_n|}{\|Dg(\boldsymbol{\mu}_n)\|}\leqslant\frac{\|\boldsymbol{V}_n\|^2}{2}\,\frac{\left\|D^2g(\boldsymbol{\mu}_n+\theta_n\boldsymbol{V}_n)\right\|}{\|Dg(\boldsymbol{\mu}_n)\|}\leqslant\frac{W^2}{2}\,\frac{\left\|D^2g(\boldsymbol{\mu}_n+\theta_n\boldsymbol{V}_n)\right\|}{\|Dg(\boldsymbol{\mu}_n)\|}\xrightarrow{\text{a.s.}}0$$

so that for every $\varepsilon > 0$, $\Pr\left\{\dfrac{|X_n-Y_n|}{\|Dg(\boldsymbol{\mu}_n)\|}>\varepsilon\right\}=o(1)$ as $n\to\infty$. Taking $\varepsilon > 0$ small enough, Polya's theorem permit us to conclude that

$$\sup_{x\in\mathbb{R}}|F_{X_n}(x)-F_{Y_n}(x)|\to 0. \qquad \square$$

Remark 2.1. Let us note that, in Theorem 2.1, when $N = 1$ the condition $\dfrac{Dg(\boldsymbol{t})}{\|Dg(\boldsymbol{t})\|}$ exists as $\|\boldsymbol{t}\|\to\infty$ can be dropped since in this case the uniform bound (2.1) takes the look

$$\sup_{x\in\mathbb{R}}|F_{X_n}(x)-F_{Y_n}(x)|\leqslant\sup_{x\in\mathbb{R}}\left|F_{V_1}\left(\frac{x-g(\mu_n)}{\frac{dg}{dx_1}(\mu_n)}+\varepsilon\right)-F_{V_1}\left(\frac{x-g(\mu_n)}{\frac{dg}{dx_1}(\mu_n)}\right)\right|+$$
$$+\sup_{x\in\mathbb{R}}\left|F_{V_1}\left(\frac{x-g(\mu_n)}{\frac{dg}{dx_1}(\mu_n)}\right)-F_{V_1}\left(\frac{x-g(\mu_n)}{\frac{dg}{dx_1}(\mu_n)}-\varepsilon\right)\right|+\Pr\left\{\left|\frac{X_n-Y_n}{\frac{dg}{dx_1}(\mu_n)}\right|>\varepsilon\right\}$$

and F_{V_1} is uniformly continuous.

When $N = 1$ we can consider functions $g\colon\mathbb{R}\longrightarrow\mathbb{R}$ defined by $g(x) = x^r$ (*power behavior*) or more generally any polynomial $\sum_{k=0}^{m}a_kx^k$ with real coefficients a_k. Moreover, functions g such that $\frac{d}{dx}g(x)=\exp\left(x^r\right)$, $r<1$ (*exponencial behavior*) or $\frac{d}{dx}g(x)=\log^r\left(x^2+1\right)$ (*logarithmic behavior*) can be chosen broadening the important class of polynomials. In the multidimensional case, a remarkable example occurs when $\boldsymbol{V}_1$ has multivariate normal distribution with positive-definite variance-covariance matrix and g is a polynomial function in N variables $x_1, ..., x_N$ with arbitrary real coefficients, that is,

$$g(x_1,...,x_N)=\sum_{k_1=0}^{m_1}\cdots\sum_{k_N=0}^{m_N}a_{k_1\ldots k_N}x_1^{k_1}\ldots x_N^{k_N}.$$

3. APPLICATION OF THE UNIFORM APPROXIMATIONS TO THE DUAL STATIS METHOD

Without loss of generality, we will assume that the p-by-p symmetric positive definite matrix $\mathbf{Q}$ and the n_i-by-n_i diagonal matrix $\mathbf{D}_i$ introduced in the first section are, respectively, the identity matrix and the diagonal matrix $\mathrm{diag}\left(\frac{1}{n_i}, ..., \frac{1}{n_i}\right)$. Indeed, these assumptions can be made performing the linear transformation $\mathbf{X}_i$ to $\mathbf{Y}_i = \sqrt{n_i}\,\mathbf{D}_i^{1/2}\mathbf{X}_i\mathbf{T}'$ as a preliminary step. From (1.1) and (1.3) we have

$$\begin{aligned}
\mathbf{W}_i &= \mathbf{X}_i'\mathbf{D}_i\mathbf{X}_i \\
&= (\mathbf{U}_i+\mathbf{E}_i)'\mathbf{D}_i(\mathbf{U}_i+\mathbf{E}_i) \\
&= \frac{1}{n_i}\mathbf{U}_i'\mathbf{U}_i + \frac{1}{n_i}\left[\mathbf{U}_i'\mathbf{E}_i + (\mathbf{U}_i'\mathbf{E}_i)'\right] + \frac{1}{n_i}\mathbf{E}_i'\mathbf{E}_i.
\end{aligned}$$

Thus,

$$\begin{aligned}
\mathbf{W}_i\mathbf{W}_i = {}& \frac{1}{n_i^2}(\mathbf{U}_i'\mathbf{U}_i)^2 + \frac{1}{n_i^2}(\mathbf{U}_i'\mathbf{U}_i)\big(\mathbf{U}_i'\mathbf{E}_i + (\mathbf{U}_i'\mathbf{E}_i)'\big) + \frac{1}{n_i^2}(\mathbf{U}_i'\mathbf{U}_i)(\mathbf{E}_i'\mathbf{E}_i) + \\
&+ \frac{1}{n_i^2}\big(\mathbf{U}_i'\mathbf{E}_i + (\mathbf{U}_i'\mathbf{E}_i)'\big)(\mathbf{U}_i'\mathbf{U}_i) + \frac{1}{n_i^2}\big(\mathbf{U}_i'\mathbf{E}_i + (\mathbf{U}_i'\mathbf{E}_i)'\big)^2 + \\
&+ \frac{1}{n_i^2}\big(\mathbf{U}_i'\mathbf{E}_i + (\mathbf{U}_i'\mathbf{E}_i)'\big)(\mathbf{E}_i'\mathbf{E}_i) + \frac{1}{n_i^2}(\mathbf{E}_i'\mathbf{E}_i)(\mathbf{U}_i'\mathbf{U}_i) + \\
&+ \frac{1}{n_i^2}(\mathbf{E}_i'\mathbf{E}_i)\big(\mathbf{U}_i'\mathbf{E}_i + (\mathbf{U}_i'\mathbf{E}_i)'\big) + \frac{1}{n_i^2}(\mathbf{E}_i'\mathbf{E}_i)^2.
\end{aligned}$$

and

$$\begin{aligned}
\mathrm{tr}(\mathbf{W}_i\mathbf{W}_i) = {}& \frac{1}{n_i^2}\,\mathrm{tr}\left((\mathbf{U}_i'\mathbf{U}_i)^2\right) + \frac{4}{n_i^2}\,\mathrm{tr}\left(\mathbf{U}_i'\mathbf{U}_i\mathbf{U}_i'\mathbf{E}_i\right) + \frac{2}{n_i^2}\,\mathrm{tr}\left((\mathbf{U}_i'\mathbf{U}_i)(\mathbf{E}_i'\mathbf{E}_i)\right) + \\
&+ \frac{4}{n_i^2}\,\mathrm{tr}\left(\mathbf{E}_i'\mathbf{E}_i\mathbf{U}_i'\mathbf{E}_i\right) + \frac{1}{n_i^2}\,\mathrm{tr}\left(\left(\mathbf{U}_i'\mathbf{E}_i + (\mathbf{U}_i'\mathbf{E}_i)'\right)^2\right) + \frac{1}{n_i^2}\,\mathrm{tr}\left((\mathbf{E}_i'\mathbf{E}_i)^2\right)
\end{aligned}$$

provide that $\mathrm{tr}(\mathbf{A}) = \mathrm{tr}(\mathbf{A}')$ and $\mathrm{tr}(\mathbf{AB}) = \mathrm{tr}(\mathbf{BA})$ (see [5], page 50). Moreover,

$$\begin{aligned}
\mathbf{W}_i\mathbf{W}_j = {}& \frac{1}{n_in_j}(\mathbf{U}_i'\mathbf{U}_i)(\mathbf{U}_j'\mathbf{U}_j) + \frac{1}{n_in_j}(\mathbf{U}_i'\mathbf{U}_i)\big(\mathbf{U}_j'\mathbf{E}_j + (\mathbf{U}_j'\mathbf{E}_j)'\big) + \\
&+ \frac{1}{n_in_j}(\mathbf{U}_i'\mathbf{U}_i)(\mathbf{E}_j'\mathbf{E}_j) + \frac{1}{n_in_j}\big(\mathbf{U}_i'\mathbf{E}_i + (\mathbf{U}_i'\mathbf{E}_i)'\big)(\mathbf{U}_j'\mathbf{U}_j) + \\
&+ \frac{1}{n_in_j}\big(\mathbf{U}_i'\mathbf{E}_i + (\mathbf{U}_i'\mathbf{E}_i)'\big)\big(\mathbf{U}_j'\mathbf{E}_j + (\mathbf{U}_j'\mathbf{E}_j)'\big) + \frac{1}{n_in_j}\big(\mathbf{U}_i'\mathbf{E}_i + (\mathbf{U}_i'\mathbf{E}_i)'\big)(\mathbf{E}_j'\mathbf{E}_j) + \\
&+ \frac{1}{n_in_j}(\mathbf{E}_i'\mathbf{E}_i)(\mathbf{U}_j'\mathbf{U}_j) + \frac{1}{n_in_j}(\mathbf{E}_i'\mathbf{E}_i)\big(\mathbf{U}_j'\mathbf{E}_j + (\mathbf{U}_j'\mathbf{E}_j)'\big) + \frac{1}{n_in_j}(\mathbf{E}_i'\mathbf{E}_i)(\mathbf{E}_j'\mathbf{E}_j).
\end{aligned}$$

when $i \neq j$ and so

$$\begin{aligned}\operatorname{tr}(\mathbf{W}_i\mathbf{W}_j) &= \frac{1}{n_i n_j}\operatorname{tr}\left(\mathbf{U}_i'\mathbf{U}_i\mathbf{U}_j'\mathbf{U}_j\right) + \frac{2}{n_i n_j}\operatorname{tr}\left(\mathbf{U}_i'\mathbf{U}_i\mathbf{U}_j'\mathbf{E}_j\right) + \\ &+ \frac{1}{n_i n_j}\operatorname{tr}\left(\mathbf{U}_i'\mathbf{U}_i)(\mathbf{E}_j'\mathbf{E}_j\right) + \frac{2}{n_i n_j}\operatorname{tr}\left(\mathbf{U}_i'\mathbf{E}_i\mathbf{U}_j'\mathbf{U}_j\right) + \frac{2}{n_i n_j}\operatorname{tr}\left(\mathbf{U}_i'\mathbf{E}_i\mathbf{U}_j'\mathbf{E}_j\right) + \\ &+ \frac{2}{n_i n_j}\operatorname{tr}\left(\mathbf{U}_i'\mathbf{E}_i(\mathbf{U}_j'\mathbf{E}_j)'\right) + \frac{2}{n_i n_j}\operatorname{tr}\left(\mathbf{U}_i'\mathbf{E}_i\mathbf{E}_j'\mathbf{E}_j\right) + \frac{1}{n_i n_j}\operatorname{tr}\left(\mathbf{E}_i'\mathbf{E}_i\mathbf{U}_j'\mathbf{U}_j\right) + \\ &+ \frac{2}{n_i n_j}\operatorname{tr}\left(\mathbf{E}_i'\mathbf{E}_i\mathbf{U}_j'\mathbf{E}_j\right) + \frac{1}{n_i n_j}\operatorname{tr}\left(\mathbf{E}_i'\mathbf{E}_i\mathbf{E}_j'\mathbf{E}_j\right).\end{aligned}$$

We will apply now the theoretical results of Section 2 to obtain an uniform approximation for the entries of the matrix $\mathbf{S}$ considering $g_i\colon \mathbb{R}^{n_i\times p} \longrightarrow \mathbb{R}$ defined by

$$g_i(\mathbf{X}) = \frac{1}{n_i^2}\operatorname{tr}\left((\mathbf{X}'\mathbf{X})^2\right), \quad i = 1, ..., k.$$

Representing $\mathbf{i}_j$ the jth column of an identity matrix of unspecified dimensions we have

$$\begin{aligned}\frac{\partial g_i(\mathbf{X})}{\partial x_{\ell j}} &= \frac{1}{n_i^2}\operatorname{tr}\left(\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{i}_\ell\mathbf{i}_j' + \mathbf{i}_j\mathbf{i}_\ell'\mathbf{X}) + (\mathbf{X}'\mathbf{i}_\ell\mathbf{i}_j' + \mathbf{i}_j\mathbf{i}_\ell'\mathbf{X})\mathbf{X}'\mathbf{X}\right) \\ &= \frac{4}{n_i^2}\operatorname{tr}\left(\mathbf{X}'\mathbf{X}\mathbf{X}'\mathbf{i}_\ell\mathbf{i}_j'\right)\end{aligned}$$

$\ell = 1, ..., n_i$, $j = 1, ..., p$ (see [5], pages 299 and 300). Therefore,

$$g_i(\mathbf{U}_i) + Dg_i(\mathbf{U}_i)\cdot\mathbf{E}_i = \frac{1}{n_i^2}\operatorname{tr}\left((\mathbf{U}_i'\mathbf{U}_i)^2\right) + \frac{4}{n_i^2}\operatorname{tr}\left(\mathbf{U}_i'\mathbf{U}_i\mathbf{U}_i'\mathbf{E}_i\right)$$

and from Theorem 2.1 we obtain

$$\sup_{x\in\mathbb{R}}\left|F_{g_i(\mathbf{U}_i+\mathbf{E}_i)} - F_{g_i(\mathbf{U}_i)+Dg_i(\mathbf{U}_i)\cdot\mathbf{E}_i}\right| \to 0 \tag{3.1}$$

as $\|\operatorname{vec}(\mathbf{U}_i)\| \to \infty$, where $\operatorname{vec}(\mathbf{U}_i)$ is the vectorization of $\mathbf{U}_i$ (see [5], pages 339 and 340), since the entries of $\mathbf{E}_i$ are continuous r.v.'s. For the case $i \neq j$ consider $g_{i,j}\colon \mathbb{R}^{n_i\times p}\times\mathbb{R}^{n_j\times p} \longrightarrow \mathbb{R}$ defined by

$$g_{i,j}(\mathbf{X},\mathbf{Y}) = \frac{1}{n_i n_j}\operatorname{tr}\left(\mathbf{X}'\mathbf{X}\mathbf{Y}'\mathbf{Y}\right), \quad i,j = 1,...,k, \quad i \neq j.$$

Thus,

$$\begin{aligned}g_{i,j}(\mathbf{U}_i,\mathbf{U}_j) + Dg_{i,j}(\mathbf{U}_i,\mathbf{U}_j)\cdot(\mathbf{E}_i,\mathbf{E}_j) &= \frac{1}{n_i n_j}\operatorname{tr}\left(\mathbf{U}_i'\mathbf{U}_i\mathbf{U}_j'\mathbf{U}_j\right) + \\ &+ \frac{2}{n_i n_j}\operatorname{tr}\left(\mathbf{U}_i'\mathbf{U}_i\mathbf{U}_j'\mathbf{E}_j\right) + \frac{2}{n_i n_j}\operatorname{tr}\left(\mathbf{U}_j'\mathbf{U}_j\mathbf{U}_i'\mathbf{E}_i\right)\end{aligned}$$

and again, from Theorem 2.1 we have

$$\sup_{x\in\mathbb{R}}\left|F_{g_{i,j}(\mathbf{U}_i+\mathbf{E}_i,\mathbf{U}_j+\mathbf{E}_j)} - F_{g_{i,j}(\mathbf{U}_i,\mathbf{U}_j)+Dg_{i,j}(\mathbf{U}_i,\mathbf{U}_j)\cdot(\mathbf{E}_i,\mathbf{E}_j)}\right| \to 0 \tag{3.2}$$

as $\|\text{vec}(\mathbf{U}_i)\| \to \infty$ for each i. Therefore, considering $\overline{\mathbf{S}} = [\overline{s}_{ij}]_{i,j=1,\ldots,k}$ where

$$\overline{s}_{ii} = g_i(\mathbf{U}_i) + Dg_i(\mathbf{U}_i) \cdot \mathbf{E}_i \quad \text{and} \quad \overline{s}_{ij} = g_{i,j}(\mathbf{U}_i, \mathbf{U}_j) + Dg_{i,j}(\mathbf{U}_i, \mathbf{U}_j) \cdot (\mathbf{E}_i, \mathbf{E}_j),$$

with $i \neq j$ for all $i, j = 1, ..., k$, we can state from (3.1) and (3.2) that the distribution of each entry of $\mathbf{S}$ can be uniformly approximated by the distribution of the same entry of $\overline{\mathbf{S}}$ when $\|\text{vec}(\mathbf{U}_i)\| \to \infty$ for each i.

The above exposure can be summarized in the following result

Theorem 3.1. *If $\mathbf{X}_i$, $i = 1, ..., k$ are n_i-by-p random data tables such that $\mathbf{X}_i = \mathbf{U}_i + \mathbf{E}_i$, where $\mathbf{E}_i$ are independent n_i-by-p random matrices having i.i.d. continuous entries and $\mathbf{U}_i$ are non-random n_i-by-p matrices then, for each $i, j = 1, ..., k$, the distribution of*

$$s_{ij} = \frac{1}{n_i n_j} \operatorname{tr}\left[(\mathbf{U}_i + \mathbf{E}_i)'(\mathbf{U}_i + \mathbf{E}_i)(\mathbf{U}_j + \mathbf{E}_j)'(\mathbf{U}_j + \mathbf{E}_j)\right]$$

is uniformly approximated by the distribution of

$$\overline{s}_{ij} = \frac{1}{n_i n_j} \operatorname{tr}\left(\mathbf{U}_i'\mathbf{U}_i\mathbf{U}_j'\mathbf{U}_j\right) + \frac{2}{n_i n_j} \operatorname{tr}\left(\mathbf{U}_i'\mathbf{U}_i\mathbf{U}_j'\mathbf{E}_j\right) + \frac{2}{n_i n_j} \operatorname{tr}\left(\mathbf{U}_j'\mathbf{U}_j\mathbf{U}_i'\mathbf{E}_i\right),$$

as $\|\text{vec}(\mathbf{U}_i)\| \to \infty$ for each i.

Remark 3.1. Observe that the condition $\|\text{vec}(\mathbf{U}_i)\| \to \infty$ for each i are related with the smallness coefficient of variation and it is a verifiable assumption in some scenarios.

3.1. Estimating the eigenvalue and the eigenvector of $\mathrm{E}(\mathbf{S})$

Recovering (1.5) we can write synthetically

$$\overline{\mathbf{S}} \overset{\mathrm{d}}{\approx} \mathbf{S} = \lambda \boldsymbol{\alpha}\boldsymbol{\alpha}' + \boldsymbol{\mathcal{E}} \tag{3.3}$$

when $\|\text{vec}(\mathbf{U}_i)\|$ is large enough for all i (i.e. the distribution of each entry of $\boldsymbol{\mathcal{E}}$ can be computable approximately). From (1.2) and (1.4), one estimates λ by the eigenvalue ρ and one estimates $\boldsymbol{\alpha}$ by the eigenvector $\mathbf{v}$, that is, we will consider the following estimators

$$\widehat{\lambda} = \rho \quad \text{and} \quad \widehat{\boldsymbol{\alpha}} = \mathbf{v}$$

of λ and $\boldsymbol{\alpha}$, respectively. The choice of $\widehat{\lambda}$ and $\widehat{\boldsymbol{\alpha}}$ as estimators of λ and $\boldsymbol{\alpha}$, respectively, arises in a very natural way (the same estimation method of eigenvalues and eigenvectors was already used in Anderson (1963)). On the other hand, the symmetry of $\mathbf{S}$ implies

$$\widehat{\boldsymbol{\beta}} = \mathbf{S}\mathbf{v} = \left(\mathbf{I} \otimes \mathbf{v}'\right) \text{vec}(\mathbf{S})$$

where $\otimes$ denotes the Kronecker product (see [5], page 333) and $\mathrm{vec}(\mathbf{S})$ is the vectorization of $\mathbf{S}$. Using (3.3) we can compute approximately the distribution of $\widehat{\boldsymbol{\beta}}$. Indeed, we can consider the approximated estimator $\widetilde{\boldsymbol{\beta}} = (\mathbf{I} \otimes \boldsymbol{\alpha}')\,\mathrm{vec}(\overline{\mathbf{S}})$ instead of $\widehat{\boldsymbol{\beta}}$, since its distribution is always determinable with the aid of the errors distribution.

3.2. Example with i.i.d. normal errors

Let us consider the entries of $\mathbf{E}_i$ i.i.d. normal distributed with zero mean and variance σ^2. Given an non-random q-by-n_i matrix $\mathbf{M}$, it is well-known that the trace $\mathrm{tr}(\mathbf{M}\mathbf{E}_i)$ is distributed normally with zero mean and variance $\sigma^2\,\mathrm{tr}(\mathbf{M}\mathbf{M}')$. Therefore, $\overline{s}_{ii}$ is distributed normally with mean $\frac{\mathrm{tr}\left((\mathbf{U}_i'\mathbf{U}_i)^2\right)}{n_i^2}$ and variance $\frac{16\sigma^2\,\mathrm{tr}\left((\mathbf{U}_i'\mathbf{U}_i)^3\right)}{n_i^4}$. Moreover, $\overline{s}_{ij}$ is distributed normally with mean $\frac{\mathrm{tr}(\mathbf{U}_i'\mathbf{U}_i\mathbf{U}_j'\mathbf{U}_j)}{n_in_j}$ and variance $\frac{4\sigma^2\,\mathrm{tr}\left((\mathbf{U}_i'\mathbf{U}_i)^2\mathbf{U}_j'\mathbf{U}_j+(\mathbf{U}_j'\mathbf{U}_j)^2\mathbf{U}_i'\mathbf{U}_i\right)}{n_i^2n_j^2}$ for all $i,j = 1,...,k$ with $i \neq j$. Using the covariance properties we also get

$$\mathrm{Cov}\,(\overline{s}_{ii},\overline{s}_{ii}) = \mathrm{Cov}\left(\tfrac{4}{n_i^2}\,\mathrm{tr}(\mathbf{U}_i'\mathbf{U}_i\mathbf{U}_i'\mathbf{E}_i),\tfrac{4}{n_i^2}\,\mathrm{tr}(\mathbf{U}_i'\mathbf{U}_i\mathbf{U}_i'\mathbf{E}_i)\right) = \tfrac{16\sigma^2}{n_i^4}\,\mathrm{tr}\left((\mathbf{U}_i'\mathbf{U}_i)^3\right)$$

and for $i \neq j$,

$$\begin{aligned}
&\mathrm{Cov}\,(\overline{s}_{ii},\overline{s}_{jj}) = \mathrm{Cov}\left(\tfrac{4}{n_i^2}\,\mathrm{tr}(\mathbf{U}_i'\mathbf{U}_i\mathbf{U}_i'\mathbf{E}_i),\tfrac{4}{n_j^2}\,\mathrm{tr}(\mathbf{U}_j'\mathbf{U}_j\mathbf{U}_j'\mathbf{E}_j)\right) = 0\\
&\mathrm{Cov}\,(\overline{s}_{ii},\overline{s}_{ij}) =\\
&\quad = \mathrm{Cov}\left(\tfrac{4}{n_i^2}\,\mathrm{tr}\left(\mathbf{U}_i'\mathbf{U}_i\mathbf{U}_i'\mathbf{E}_i\right),\tfrac{2}{n_in_j}\,\mathrm{tr}\left(\mathbf{U}_i'\mathbf{U}_i\mathbf{U}_j'\mathbf{E}_j\right)+\tfrac{2}{n_in_j}\,\mathrm{tr}\left(\mathbf{U}_j'\mathbf{U}_j\mathbf{U}_i'\mathbf{E}_i\right)\right)\\
&\quad = \tfrac{8\sigma^2}{n_i^3n_j}\,\mathrm{tr}\left((\mathbf{U}_i'\mathbf{U}_i)^2\,\mathbf{U}_j'\mathbf{U}_j\right)
\end{aligned}$$

$$\begin{aligned}
\mathrm{Cov}\,(\overline{s}_{ij},\overline{s}_{ij}) &= \mathrm{Cov}\left(\tfrac{2}{n_in_j}\,\mathrm{tr}\left(\mathbf{U}_i'\mathbf{U}_i\mathbf{U}_j'\mathbf{E}_j\right)+\tfrac{2}{n_in_j}\,\mathrm{tr}\left(\mathbf{U}_j'\mathbf{U}_j\mathbf{U}_i'\mathbf{E}_i\right),\right.\\
&\qquad\left.\tfrac{2}{n_in_j}\,\mathrm{tr}\left(\mathbf{U}_i'\mathbf{U}_i\mathbf{U}_j'\mathbf{E}_j\right)+\tfrac{2}{n_in_j}\,\mathrm{tr}\left(\mathbf{U}_j'\mathbf{U}_j\mathbf{U}_i'\mathbf{E}_i\right)\right)\\
&= \tfrac{4\sigma^2}{n_i^2n_j^2}\,\mathrm{tr}\left((\mathbf{U}_i'\mathbf{U}_i)^2\mathbf{U}_j'\mathbf{U}_j\right)+\tfrac{4\sigma^2}{n_i^2n_j^2}\,\mathrm{tr}\left((\mathbf{U}_j'\mathbf{U}_j)^2\mathbf{U}_i'\mathbf{U}_i\right).
\end{aligned}$$

For all different i,j,ℓ,q we still have

$$\begin{aligned}
\mathrm{Cov}\,(\overline{s}_{ii},\overline{s}_{j\ell}) &= \mathrm{Cov}\left(\tfrac{4}{n_i^2}\,\mathrm{tr}(\mathbf{U}_i'\mathbf{U}_i\mathbf{U}_i'\mathbf{E}_i),\right.\\
&\qquad\left.\tfrac{2}{n_jn_\ell}\,\mathrm{tr}(\mathbf{U}_j'\mathbf{U}_j\mathbf{U}_\ell'\mathbf{E}_\ell)+\tfrac{2}{n_jn_\ell}\,\mathrm{tr}(\mathbf{U}_\ell'\mathbf{U}_\ell\mathbf{U}_j'\mathbf{E}_j)\right)\\
&= 0
\end{aligned}$$

$$
\begin{aligned}
\mathrm{Cov}\,(\overline{s}_{ij}, \overline{s}_{i\ell}) &= \mathrm{Cov}\Big(\tfrac{2}{n_i n_j}\,\mathrm{tr}(\mathbf{U}_i'\mathbf{U}_i\mathbf{U}_j'\mathbf{E}_j) + \tfrac{2}{n_i n_j}\,\mathrm{tr}(\mathbf{U}_j'\mathbf{U}_j\mathbf{U}_i'\mathbf{E}_i),\\
&\qquad\qquad \tfrac{2}{n_i n_\ell}\,\mathrm{tr}(\mathbf{U}_i'\mathbf{U}_i\mathbf{U}_\ell'\mathbf{E}_\ell) + \tfrac{2}{n_i n_\ell}\,\mathrm{tr}(\mathbf{U}_\ell'\mathbf{U}_\ell\mathbf{U}_i'\mathbf{E}_i)\Big)\\
&= \tfrac{4\sigma^2}{n_i^2 n_j n_\ell}\,\mathrm{tr}\left(\mathbf{U}_j'\mathbf{U}_j\mathbf{U}_i'\mathbf{U}_i\mathbf{U}_\ell'\mathbf{U}_\ell\right)\\
\mathrm{Cov}\,(\overline{s}_{ij}, \overline{s}_{q\ell}) &= \mathrm{Cov}\Big(\tfrac{2}{n_i n_j}\,\mathrm{tr}(\mathbf{U}_i'\mathbf{U}_i\mathbf{U}_j'\mathbf{E}_j) + \tfrac{2}{n_i n_j}\,\mathrm{tr}(\mathbf{U}_j'\mathbf{U}_j\mathbf{U}_i'\mathbf{E}_i),\\
&\qquad\qquad \tfrac{2}{n_q n_\ell}\,\mathrm{tr}(\mathbf{U}_q'\mathbf{U}_q\mathbf{U}_\ell'\mathbf{E}_\ell) + \tfrac{2}{n_q n_\ell}\,\mathrm{tr}(\mathbf{U}_\ell'\mathbf{U}_\ell\mathbf{U}_q'\mathbf{E}_q)\Big)\\
&= 0
\end{aligned}
$$

The next table resumes all covariance computations:

Table 1: Covariance between elements of $\overline{\mathbf{S}}$.

Elements	Covariance	# of elements
$\mathrm{Cov}(\overline{s}_{ii}, \overline{s}_{ii})$	$\frac{16\,\sigma^2}{n_i^4}\,\mathrm{tr}\Big(\big(\mathbf{U}_i'\mathbf{U}_i\big)^3\Big)$	k
$\mathrm{Cov}(\overline{s}_{ii}, \overline{s}_{ij})$	$\frac{8\,\sigma^2}{n_i^3 n_j}\,\mathrm{tr}\Big(\big(\mathbf{U}_i'\mathbf{U}_i\big)^2\,\mathbf{U}_j'\mathbf{U}_j\Big)$	$k\,(k{-}1)$
$\mathrm{Cov}(\overline{s}_{ii}, \overline{s}_{ji})$	$\frac{8\,\sigma^2}{n_i^3 n_j}\,\mathrm{tr}\Big(\big(\mathbf{U}_i'\mathbf{U}_i\big)^2\,\mathbf{U}_j'\mathbf{U}_j\Big)$	$k\,(k{-}1)$
$\mathrm{Cov}(\overline{s}_{ij}, \overline{s}_{ii})$	$\frac{8\,\sigma^2}{n_i^3 n_j}\,\mathrm{tr}\Big(\big(\mathbf{U}_i'\mathbf{U}_i\big)^2\,\mathbf{U}_j'\mathbf{U}_j\Big)$	$k\,(k{-}1)$
$\mathrm{Cov}(\overline{s}_{ji}, \overline{s}_{ii})$	$\frac{8\,\sigma^2}{n_i^3 n_j}\,\mathrm{tr}\Big(\big(\mathbf{U}_i'\mathbf{U}_i\big)^2\,\mathbf{U}_j'\mathbf{U}_j\Big)$	$k\,(k{-}1)$
$\mathrm{Cov}(\overline{s}_{ii}, \overline{s}_{jj})$	0	$k\,(k{-}1)$
$\mathrm{Cov}(\overline{s}_{ii}, \overline{s}_{j\ell})$	0	$k\,(k{-}1)\,(k{-}2)$
$\mathrm{Cov}(\overline{s}_{ij}, \overline{s}_{\ell\ell})$	0	$k\,(k{-}1)\,(k{-}2)$
$\mathrm{Cov}(\overline{s}_{ij}, \overline{s}_{ij})$	$\frac{4\,\sigma^2}{n_i^2 n_j^2}\,\mathrm{tr}\Big(\big(\mathbf{U}_i'\mathbf{U}_i\big)^2\,\mathbf{U}_j'\mathbf{U}_j\Big) + \frac{4\,\sigma^2}{n_i^2 n_j^2}\,\mathrm{tr}\Big(\big(\mathbf{U}_j'\mathbf{U}_j\big)^2\,\mathbf{U}_i'\mathbf{U}_i\Big)$	$k\,(k{-}1)$
$\mathrm{Cov}(\overline{s}_{ij}, \overline{s}_{ji})$	$\frac{4\,\sigma^2}{n_i^2 n_j^2}\,\mathrm{tr}\Big(\big(\mathbf{U}_j'\mathbf{U}_j\big)^2\,\mathbf{U}_i'\mathbf{U}_i\Big) + \frac{4\,\sigma^2}{n_i^2 n_j^2}\,\mathrm{tr}\Big(\big(\mathbf{U}_j'\mathbf{U}_j\big)^2\,\mathbf{U}_i'\mathbf{U}_i\Big)$	$k\,(k{-}1)$
$\mathrm{Cov}(\overline{s}_{ij}, \overline{s}_{i\ell})$	$\frac{4\,\sigma^2}{n_i^2 n_j n_\ell}\,\mathrm{tr}\Big(\mathbf{U}_j'\mathbf{U}_j\mathbf{U}_i'\mathbf{U}_i\mathbf{U}_\ell'\mathbf{U}_\ell\Big)$	$k\,(k{-}1)\,(k{-}2)$
$\mathrm{Cov}(\overline{s}_{ij}, \overline{s}_{j\ell})$	$\frac{4\,\sigma^2}{n_j^2 n_i n_\ell}\,\mathrm{tr}\Big(\mathbf{U}_i'\mathbf{U}_i\mathbf{U}_j'\mathbf{U}_j\mathbf{U}_\ell'\mathbf{U}_\ell\Big)$	$k\,(k{-}1)\,(k{-}2)$
$\mathrm{Cov}(\overline{s}_{ij}, \overline{s}_{\ell i})$	$\frac{4\,\sigma^2}{n_i^2 n_j n_\ell}\,\mathrm{tr}\Big(\mathbf{U}_i'\mathbf{U}_i\mathbf{U}_j'\mathbf{U}_j\mathbf{U}_\ell'\mathbf{U}_\ell\Big)$	$k\,(k{-}1)\,(k{-}2)$
$\mathrm{Cov}(\overline{s}_{ij}, \overline{s}_{\ell j})$	$\frac{4\,\sigma^2}{n_j^2 n_i n_\ell}\,\mathrm{tr}\Big(\mathbf{U}_i'\mathbf{U}_i\mathbf{U}_j'\mathbf{U}_j\mathbf{U}_\ell'\mathbf{U}_\ell\Big)$	$k\,(k{-}1)\,(k{-}2)$
$\mathrm{Cov}(\overline{s}_{ij}, \overline{s}_{q\ell})$	0	$k\,(k{-}1)\,(k{-}2)\,(k{-}3)$

Generally, if $\boldsymbol{\Sigma}_i^+$ is the Moore–Penrose inverse (see [5], page 493) of the covariance matrix $\boldsymbol{\Sigma}_i$ of $\mathrm{vec}(\mathbf{U}_i + \mathbf{E}_i)$ (the vectorization of $\mathbf{U}_i + \mathbf{E}_i$) then the quadratic form $\mathrm{vec}(\mathbf{U}_i + \mathbf{E}_i)'\boldsymbol{\Sigma}_i^+\mathrm{vec}(\mathbf{U}_i + \mathbf{E}_i)$ has chi-square distribution with $r_i = \mathrm{rank}(\boldsymbol{\Sigma}_i)$ degrees of freedom and non-centrality parameter $\delta_i = \mathbf{U}_i'\boldsymbol{\Sigma}_i^+\mathbf{U}_i$ (see [12], page 182). Hence, we may use δ_i to measure the non-centrality of the sample.

In this case, the covariance matrix of $\mathrm{vec}(\mathbf{E}_i)$ is known, so that the assumption $\|\mathrm{vec}(\mathbf{U}_i)\| \to \infty$ for each i corresponds to consider observations with low variation coefficients which implies a large δ_i.

Since $\mathbf{S} = \lambda\boldsymbol{\alpha}\boldsymbol{\alpha}' + \boldsymbol{\mathcal{E}}$ with $\mathrm{E}(\boldsymbol{\mathcal{E}}) = \mathbf{O}$ we get

$$\mathrm{E}\left(\left(\mathbf{I}\otimes\boldsymbol{\alpha}'\right)\mathrm{vec}(\mathbf{S})\right) = \left(\mathbf{I}\otimes\boldsymbol{\alpha}'\right)\mathrm{vec}(\lambda\boldsymbol{\alpha}\boldsymbol{\alpha}') = \lambda\boldsymbol{\alpha} = \boldsymbol{\beta}$$

that is, $\widehat{\boldsymbol{\beta}}$ is approximately unbiased. The covariance matrix of $\widehat{\boldsymbol{\beta}}$ can also be computed (approximately) through

$$\left(\mathbf{I}\otimes\boldsymbol{\alpha}'\right)\boldsymbol{\Sigma}_{\mathrm{vec}(\overline{\mathbf{S}})}\left(\mathbf{I}\otimes\boldsymbol{\alpha}\right) \tag{3.4}$$

where $\boldsymbol{\Sigma}_{\mathrm{vec}(\overline{\mathbf{S}})}$ is the covariance matrix of $\overline{\mathbf{S}}$ (i.e. the elements of the previous table). Hence, $\widehat{\boldsymbol{\beta}}$ will have (approximately) normal distribution with mean value $\boldsymbol{\beta}$ and covariance matrix $\mathbf{C} = (\mathbf{I}\otimes\boldsymbol{\alpha}')\,\boldsymbol{\Sigma}_{\mathrm{vec}(\overline{\mathbf{S}})}\,(\mathbf{I}\otimes\boldsymbol{\alpha})$.

Remark 3.2. Relation (3.4) lead us even to consider

$$\widehat{\boldsymbol{\Sigma}} = \left(\mathbf{I}\otimes\mathbf{v}'\right)\boldsymbol{\Sigma}_{\mathrm{vec}(\overline{\mathbf{S}})}\left(\mathbf{I}\otimes\mathbf{v}\right)$$

as an estimator of the covariance matrix of $\widehat{\boldsymbol{\beta}}$.

Now, we will construct a non-random k^2-by-k^2 matrix $\mathbf{G}$ $(k > 1)$ such that the random vector $\mathbf{y} = \mathbf{G}\mathrm{vec}(\mathbf{S})$ be independent of $\widehat{\boldsymbol{\beta}}$ and

$$\mathrm{E}\big(\mathbf{G}\mathrm{vec}(\overline{\mathbf{S}})\big) = \mathbf{0}, \quad \mathbf{G}\boldsymbol{\Sigma}_{\mathrm{vec}(\overline{\mathbf{S}})}\mathbf{G}' = \begin{bmatrix} \sigma^2\mathbf{I} & \mathbf{O}_{12} \\ \mathbf{O}_{21} & \mathbf{O}_{22} \end{bmatrix}$$

where $\mathbf{I}$ is the identity matrix of some size less than or equal to k^2 and $\mathbf{O}_{12}$, $\mathbf{O}_{21}$, $\mathbf{O}_{22}$ are matrices with zero elements. First, we can obtain an k^2-by-k^2 non-random matrix $\mathbf{B}$ such that

$$\mathbf{B}\,\mathrm{E}\big(\mathrm{vec}(\overline{\mathbf{S}})\big) = \mathbf{0} \quad \text{and} \quad \mathbf{B}\boldsymbol{\Sigma}_{\mathrm{vec}(\overline{\mathbf{S}})}(\mathbf{I}\otimes\boldsymbol{\alpha}) = \mathbf{O}$$

taking, for instance,

$$\mathbf{B} = \mathbf{R}\left(\mathbf{I} - \Big[\,\boldsymbol{\Sigma}_{\mathrm{vec}(\overline{\mathbf{S}})}(\mathbf{I}\otimes\boldsymbol{\alpha})\;\; \mathrm{E}\big(\mathrm{vec}(\overline{\mathbf{S}})\big)\,\Big]\Big[\,\boldsymbol{\Sigma}_{\mathrm{vec}(\overline{\mathbf{S}})}(\mathbf{I}\otimes\boldsymbol{\alpha})\;\; \mathrm{E}\big(\mathrm{vec}(\overline{\mathbf{S}})\big)\,\Big]^{+}\right)$$

where $\mathbf{R}$ is an arbitrary (conformatable) matrix (see [1], page 295). Hence, $\mathbf{B}\mathrm{vec}(\overline{\mathbf{S}})$ will have multivariate normal distribution with mean vector $\mathbf{B}\,\mathrm{E}\big(\mathrm{vec}(\overline{\mathbf{S}})\big)$ and covariance matrix $\mathbf{B}\boldsymbol{\Sigma}_{\mathrm{vec}(\overline{\mathbf{S}})}\mathbf{B}'$ (see [14], page 32). Since $\mathbf{B}\boldsymbol{\Sigma}_{\mathrm{vec}(\overline{\mathbf{S}})}\mathbf{B}'$ is an symmetric matrix with rank r such that $\mathbf{B}\boldsymbol{\Sigma}_{\mathrm{vec}(\overline{\mathbf{S}})}\mathbf{B}'$ is either positive definite $(r = k^2)$ or positive semidefinite $(r < k^2)$ then:

(i) If $r = k^2$ then there exists a nonsingular k^2-by-k^2 matrix $\mathbf{H}$ such that $\mathbf{H}\mathbf{B}\boldsymbol{\Sigma}_{\mathrm{vec}(\overline{\mathbf{S}})}\mathbf{B}'\mathbf{H}' = \mathbf{I}$.

(ii) If $r < k^2$ then there exists a nonsingular r-by-r matrix $\mathbf{H}$ such that

$$\mathbf{H}\mathbf{B}\boldsymbol{\Sigma}_{\mathrm{vec}(\overline{\mathbf{S}})}\mathbf{B}'\mathbf{H}' = \begin{bmatrix} \mathbf{I} & \mathbf{O}_{12} \\ \mathbf{O}_{21} & \mathbf{O}_{22} \end{bmatrix}$$

where $\mathbf{O}_{12}$, $\mathbf{O}_{21}$ and $\mathbf{O}_{22}$ are r-by-$(k^2 - r)$, $(k^2 - r)$-by-r and $(k^2 - r)$-by-$(k^2 - r)$ matrices with zero elements, respectively (see [14], page 27). Therefore, we can take

$$\mathbf{G} = \sigma\mathbf{H}\mathbf{B}$$

and the components of $\mathbf{y}$ will be i.i.d. normal distributed with zero mean and variance σ^2.

For testing the assumption $\mathrm{rank}\big(\mathrm{E}(\mathbf{S})\big) = 1$ we can use the following statistical test:

$$\mathcal{F} = \frac{r\sigma^2\widehat{\boldsymbol{\beta}}'\widehat{\boldsymbol{\Sigma}}^+\widehat{\boldsymbol{\beta}}}{\nu\,\|\mathbf{y}\|^2}$$

where ν is the rank of $\widehat{\boldsymbol{\Sigma}}$.

Remark 3.3. Let us stand out that $\widehat{\boldsymbol{\Sigma}}^+ = \dfrac{1}{\sigma^2}\widetilde{\boldsymbol{\Sigma}}^+$ where $\widetilde{\boldsymbol{\Sigma}}^+$ is independent of σ and so $\mathcal{F}$ do not depend on σ.

The statistical test $\mathcal{F}$ will have (approximately) F-distribution with parameters ν, r and non-centrality parameter $\delta = \boldsymbol{\beta}'\widehat{\boldsymbol{\Sigma}}^+\boldsymbol{\beta}$ (see [6], page 609). Since $\mathrm{rank}\big(\mathrm{E}(\mathbf{S})\big) = 1$ if $\delta > 0$, we will use $\mathcal{F}$ to test the null hypothesis $H_0 : \delta = 0$ against $H_1 : \delta > 0$ and the p-value of this statistical test of hypothesis can be computed by

$$p\text{-value} = \Pr(\mathcal{F} > \mathcal{F}_{\mathrm{obs}} | \delta = 0)$$

where $\mathcal{F}_{\mathrm{obs}}$ is the observed value.

After validate the model, we are able to use statistical hypothesis tests on the components of the vector $\boldsymbol{\beta}$. Given a ℓ-by-k non-random matrix $\mathbf{Z}$ and $\boldsymbol{\psi} = \mathbf{Z}\boldsymbol{\beta}$, we can test the hypothesis $H_0 : \boldsymbol{\psi} = \mathbf{z}$, where $\mathbf{z}$ is a non-random vector. Considering the estimator $\widehat{\boldsymbol{\psi}} = \mathbf{Z}\widehat{\boldsymbol{\beta}}$ of $\boldsymbol{\psi}$ then $\widehat{\boldsymbol{\psi}}$ will have (approximately) normal distribution with mean value $\boldsymbol{\psi} = \mathbf{Z}\boldsymbol{\beta}$ and covariance matrix $\mathbf{Z}\widehat{\boldsymbol{\Sigma}}\mathbf{Z}'$ (see [14], page 32). Moreover, $\widehat{\boldsymbol{\psi}}$ will be also independent from $\mathbf{y}$ which lead us to use the following statistical test:

$$\mathcal{F} = \frac{r\sigma^2(\widehat{\boldsymbol{\psi}} - \mathbf{z})'\left(\mathbf{Z}\widehat{\boldsymbol{\Sigma}}\mathbf{Z}'\right)^+(\widehat{\boldsymbol{\psi}} - \mathbf{z})}{\nu\,\|\mathbf{y}\|^2}$$

where ν is the rank of $\mathbf{Z}\widehat{\mathbf{\Sigma}}\mathbf{Z}'$. Again, $\mathcal{F}$ do not depend on σ (see Remark 4). If H_0 is accepted then $\mathcal{F}$ will have (approximately) F-distribution with parameters ν, r and noncentrality parameter

$$\delta = (\boldsymbol{\psi} - \mathbf{z})' \left(\mathbf{Z}\widehat{\mathbf{\Sigma}}\mathbf{Z}'\right)^{+} (\boldsymbol{\psi} - \mathbf{z}).$$

If $\mathbf{Z}\widehat{\mathbf{\Sigma}}\mathbf{Z}'$ was invertible then $\delta = 0$ is equivalent to the acceptance of H_0. Hence, the p-value of this statistical test of hypothesis is given by

$$p - \text{value} = \Pr(\mathcal{F} > \mathcal{F}_{\text{obs}} | H_0 \text{ is true})$$

where $\mathcal{F}_{\text{obs}}$ is the observed value.

Remark 3.4. Observe that the above statistical test of hypothesis is a generalization of the first one, in the sense that we can use it with $\mathbf{Z} = \mathbf{I}$ and $\mathbf{z} = \mathbf{0}$ to test the assumption $\text{rank}\big(\text{E}(\mathbf{S})\big) = 1$.

Remark 3.5. Furthermore, choosing the matrix $\mathbf{Z}$ and the vector $\mathbf{z}$ appropriately we can perform statistical tests of hypothesis to compare two or more components of $\boldsymbol{\beta} = (\beta_1, ..., \beta_k)$, for instance, $H_0\colon \beta_i = \beta_j$ against $H_1\colon \beta_i \neq \beta_j$ (i.e. $\mathbf{z} = \mathbf{0}$ and $\mathbf{Z} = \big[z_{ij}\big]_{\substack{i=1,\dots,\ell \\ j=1,\dots,k}}$ defined by $z_{1i} = -z_{1j} = 1$ with all remaining entries being zero). Note also that the statistical test $H_0\colon \beta_i = \beta_j$ against $H_1\colon \beta_i \neq \beta_j$ $(i \neq j)$ is equivalent to compare two different components of $\boldsymbol{\alpha}$ provided that $\lambda \neq 0$.

4. CONCLUSIONS

The theoretical results of Section 2 arose to get the solution to the following problem: when there is no limiting distribution for a sequence of r.v.'s X_n can we approximate the limit distribution of $g(X_n)$ for large values of n and some fixed function g? The well-known *delta method* cannot be used to give an answer to this question since there is no limiting distribution for X_n. In Theorem 2.1 we partially answer to the above question considering $X_n = V_n + \mu_n$ and giving sufficient conditions on g to obtain a sequence of random variables which are of the same type of $g(X_n)$ for large values of n. Let us observe also that our result allows us to get "normalizing constants" for $g(X_n)$ when n is large enough.

Therefore, besides the uniform approximation of the distribution function sequence $F_{g(\mathbf{V}_n + \boldsymbol{\mu}_n)}$ by a computable one, this work develops inference results on the components of the vector $\lambda\boldsymbol{\alpha}$, where λ and $\boldsymbol{\alpha}$ are, respectively, the eigenvalue and the eigenvector of the rank one matrix $\text{E}(\mathbf{S})$, with $\mathbf{S}$ the interstructure matrix used in dual STATIS method admitting data tables of the form (1.3). So, our results appears in the same alignment of Lazraq (see [10]) which considered an inferencial approach for the validation of the compromise matrix obtained by the STATIS method.

In our scenario of data tables, remains an open problem the generalization of the presented inferential results to the case where the rank of E(**S**) is greater than one.

ACKNOWLEDGMENTS

This work was partially supported by the Fundação para a Ciência e a Tecnologia (Portuguese Foundation for Science and Technology) through PEst-OE/MAT/UI0297/2011 (CMA).

REFERENCES

[1] Abadir, K.M. and Magnus, J.R. (2005). *Matrix Algebra*, Cambridge University Press.

[2] Anderson, T.W. (1963). Asymptotic theory for principal component analysis, *The Annals of Mathematical Statistics*, **34**(1), 122–148.

[3] DasGupta, A. (2008). *Asymptotic Theory of Statistics and Probability*, Springer.

[4] Escoufier, Y. (1973). Le traitement des variables vectorielles, *Biometrics*, **29**(4), 751–760.

[5] Harville, D.A (1997). *Matrix Algebra From a Statistician's Perspective*, Springer-Verlag, New York.

[6] Hocking, R.R. (2003). *Methods and Applications of Linear Models: Regression and the Analysis of Variance*, 2nd edn., John Wiley & Sons.

[7] Horn, R.A. and Johnson, C.R. (1985). *Matrix Analysis*, Cambridge University Press.

[8] Lavit, C. (1988). *Analyse Conjointe de Tableaux Quantitatifs*, Masson, Paris.

[9] Lavit, C.; Escoufier, Y.; Sabatier R. and Traissac, P. (1994). The ACT (STATIS) method, *Computational Statistics & Data Analysis*, North-Holland **18**, 97–119.

[10] Lazraq, A.; Hanafi, M.; Cléroux, R.; Allaire, J. and Lepage, Y. (2008). Une approche inférentielle pour la validation du compromis de la méthode STATIS, *Journal de la Société Française de Statistique*, **149**, 97–109.

[11] Lima, E.L. (1989). *Curso de Análise, Volume 2*, 3.ª ed., Projecto Euclides.

[12] Muller, K.E. and Stewart, P.W. (2006). *Linear Model Theory: Univariate, Multivariate, and Mixed Models*, John Wiley & Sons.

[13] Robert, P. and Escoufier, Y. (1976). A unifying tool for linear multivariate statistical methods: the RV-coefficient, *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, **25**(3), 257–265.

[14] Tong, Y.L. (1990). *The Multivariate Normal Distribution*, Springer-Verlag.