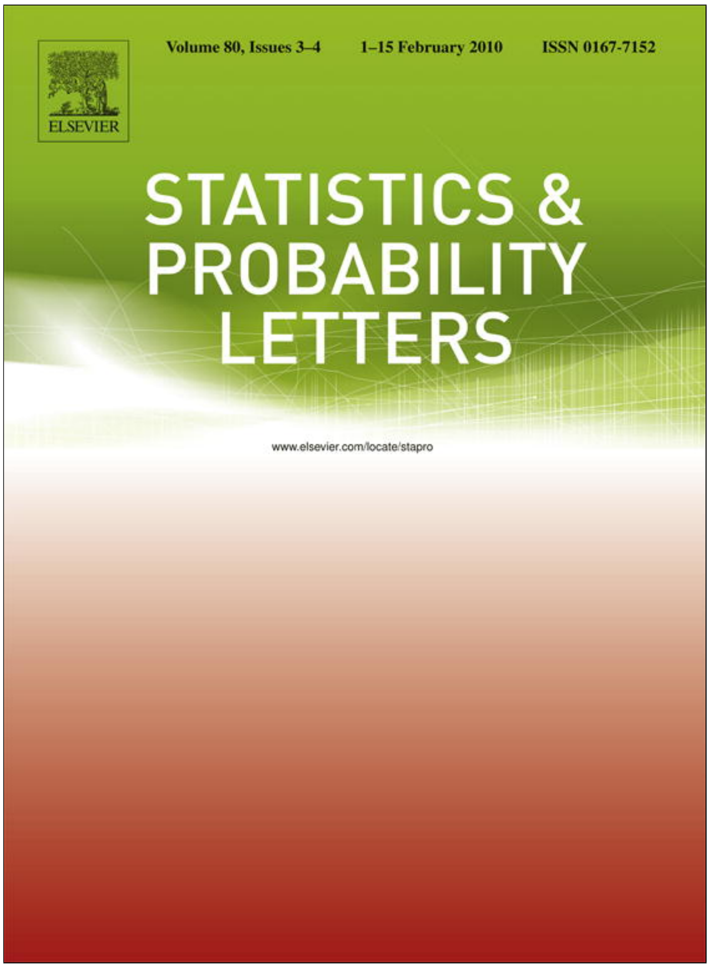
Volume 80, Issues 3–4
1–15 February 2010
ISSN 0167-7152
ELSEVIER
STATISTICS &
PROBABILITY
LETTERS
www.elsevier.com/locate/stapro

Statistics and Probability Letters 80 (2010) 228–235

Contents lists available at ScienceDirect

Statistics and Probability Letters

journal homepage: www.elsevier.com/locate/stapro

On the asymptotic behavior of general projection-pursuit estimators under the common principal components model

Graciela Boente *, Julieta Molina, Mariela Sued

Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Argentina
CONICET, Argentina

ARTICLE INFO

Article history:
Received 10 June 2009
Received in revised form 14 October 2009
Accepted 14 October 2009
Available online 20 October 2009

ABSTRACT

The common principal components model for several groups of multivariate observations assumes equal principal axes among the groups. Robust estimators can be defined replacing the sample variance by a robust dispersion measure. This paper studies the asymptotic distribution of robust projection-pursuit estimators under a common principal components model.

1. Introduction

In many situations, when dealing with several populations, in multivariate analysis, models for common structure dispersion need to be considered to overcome the problem of an excessive number of parameters. Flury (1984) introduced the so-called *Common Principal Components* (CPC) model, in which the common structure assumes that the k covariance matrices have possibly different eigenvalues but identical eigenvectors, i.e.,

$$\boldsymbol{\Sigma}_i = \boldsymbol{\beta}\boldsymbol{\Lambda}_i\boldsymbol{\beta}^{\mathrm{T}}, \quad 1 \le i \le k, \tag{1.1}$$

where $\boldsymbol{\Lambda}_i$ are diagonal matrices, $\boldsymbol{\beta}$ is the orthogonal matrix of the common eigenvectors $\boldsymbol{\beta}_j$ and $\boldsymbol{\Sigma}_i$ is the covariance matrix of the ith population. In the one-group principal component analysis, the eigenvectors $\boldsymbol{\beta}_j$ are usually arranged according to decreasing values of the associated eigenvalues. In the CPC model, no obvious fixed order of the columns of $\boldsymbol{\beta}$ is available, since the order among the diagonal elements of $\boldsymbol{\Lambda}_i$ need not be the same for all the populations. Therefore, the common directions can be ordered as $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_p$ according to several criteria some of which will be defined below. The maximum likelihood estimators of $\boldsymbol{\beta}$ and $\boldsymbol{\Lambda}_i$ are derived in Flury (1984), assuming multivariate normality of the original variables while Flury (1988) considered a unified study of the maximum likelihood estimators under different hierarchical models.

Let $(\mathbf{x}_{ij})_{1\le j\le n_i, 1\le i\le k}$ be independent observations from k independent samples in $\mathbb{R}^p$ with location parameter $\boldsymbol{\mu}_i$ and scatter matrix $\boldsymbol{\Sigma}_i$. Let $N = \sum_{i=1}^k n_i$, $\tau_{iN} = n_i/N$, where $\tau_{iN} \to \tau_i \in (0, 1)$ as $N \to \infty$, and $\mathbf{X}_i = (\mathbf{x}_{i1}, \ldots, \mathbf{x}_{in_i})$. For the CPC model, the common decomposition given in (1.1) implies that for any $\mathbf{a} \in \mathbb{R}^p$, and $1 \le i \le k$, $\mathrm{VAR}(\mathbf{a}^{\mathrm{T}}\mathbf{x}_{i1}) = \mathbf{a}^{\mathrm{T}}\boldsymbol{\beta}\boldsymbol{\Lambda}_i\boldsymbol{\beta}^{\mathrm{T}}\mathbf{a}$. Therefore, the first axis $\boldsymbol{\beta}_1$ could be defined through a projection approach by maximizing $\sum_{i=1}^k \tau_i \mathrm{VAR}(\mathbf{a}^{\mathrm{T}}\mathbf{x}_{i1})$ over $\mathbf{a} \in \mathbb{R}^p$ with $\|\mathbf{a}\| = 1$. By considering orthogonal directions to $\boldsymbol{\beta}_1$, the second axis is defined and so on. If the order of the eigenvalues, λ_{ij}, is preserved across populations, i.e., if $\lambda_{\ell 1} > \cdots > \lambda_{\ell p}$, for all ℓ, $\boldsymbol{\beta}_j$ is also the direction related to $\lambda_{\ell j}$, regardless of the index ℓ of the selected population. Otherwise, we are ordering the axis according to decreasing values of the pooled matrix eigenvalues, $\sum_{i=1}^k \tau_i\lambda_{ij}$, $1 \le j \le p$, as far as they have multiplicity one. It is well known that, as in the one-population setting, the classical CPC analysis can be affected by the existence of outliers in a sample. In a one-population setting,

* Corresponding address: Instituto de Cálculo, FCEyN, UBA Ciudad Universitaria, Pabellón 2 Buenos Aires, C1428EHA, Argentina. Fax: +54 11 457 63375.
E-mail addresses: gboente@dm.uba.ar (G. Boente), msued@dm.uba.ar (M. Sued).

0167-7152/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.spl.2009.10.011

robust estimators for the principal directions using alternative measures of variability, were first considered in Li and Chen (1985) who proposed projection-pursuit estimators maximizing (or minimizing) a robust scale. Later on, Croux and Ruiz-Gazen (2005) provided the influence functions of the resulting principal components while their asymptotic distribution was studied in Cui et al. (2003). The above described projection approach allows to define, for several populations, robust projection-pursuit estimators by considering a robust measure of dispersion s instead of the standard deviation (see Boente and Orellana, 2001) and (Boente et al., 2002) and provides clear interpretations of the resulting common directions as those maximizing the overall variability of the projected data $\sum_{i=1}^{k}\tau_{iN}s^2\left(\mathbf{a}^{\mathrm{T}}\mathbf{x}_{i1},\ldots,\mathbf{a}^{\mathrm{T}}\mathbf{x}_{in_i}\right)$.

One main disadvantage of the above described projection-pursuit approach, is that, when considering the sample standard deviation as dispersion measure, the maximum likelihood estimators are not obtained. This suggests that the criterion considered may produce an extra loss of efficiency that is not only related to the robust scale. For that reason, Boente et al. (2006) considered a general approach which consists of applying a score function to the scale estimator. To motivate this approach, remind that the maximum likelihood estimator of $\boldsymbol{\beta}$ for gaussian populations minimizes $\sum_{j=1}^{p}\sum_{i=1}^{k}n_i\ln\left(\ell_{ij}\right)$, where ℓ_{ij} are the diagonal elements of $\mathbf{F}_i=\mathbf{A}^{\mathrm{T}}\mathbf{S}_i\mathbf{A}$, i.e., ℓ_{ij} equals the sample variance of the projected vectors $\mathbf{a}_j^{\mathrm{T}}\mathbf{x}_{i1},\ldots,\mathbf{a}_j^{\mathrm{T}}\mathbf{x}_{in_i}$ (see Flury (1988)). Therefore, a natural way to robustify the maximum likelihood estimators could be to maximize iteratively $\sum_{i=1}^{k}n_i\ln\left(s^2\left(\mathbf{a}^{\mathrm{T}}\mathbf{x}_{i1},\ldots,\mathbf{a}^{\mathrm{T}}\mathbf{x}_{in_i}\right)\right)$, for a robust dispersion s. More generally, Boente et al. (2006) considered an increasing score function $f:\mathbb{R}^+\to\mathbb{R}$ and a univariate scale estimator s, and they propose to estimate the common directions as

$$\begin{cases}\widehat{\boldsymbol{\beta}}_1=\operatorname*{argmax}_{\|\mathbf{a}\|=1}\sum_{i=1}^{k}\tau_{iN}f\left(s^2(\mathbf{a}^{\mathrm{T}}\mathbf{x}_{i1},\ldots,\mathbf{a}^{\mathrm{T}}\mathbf{x}_{in_i})\right)\\ \widehat{\boldsymbol{\beta}}_m=\operatorname*{argmax}_{\mathbf{a}\in\widehat{\mathcal{B}}_m}\sum_{i=1}^{k}\tau_{iN}f\left(s^2(\mathbf{a}^{\mathrm{T}}\mathbf{x}_{i1},\ldots,\mathbf{a}^{\mathrm{T}}\mathbf{x}_{in_i})\right)\quad 2\le m\le p;\end{cases}\tag{1.2}$$

where $\widehat{\mathcal{B}}_m=\{\mathbf{a}\in\mathbb{R}^p:\|\mathbf{a}\|=1,\mathbf{a}^{\mathrm{T}}\widehat{\boldsymbol{\beta}}_j=0\text{ for }1\le j\le m-1\}$. The estimators of the eigenvalues of the i-th population are then computed as $\widehat{\lambda}_{im}=s^2(\widehat{\boldsymbol{\beta}}_m^{\mathrm{T}}\mathbf{x}_{i1},\ldots,\widehat{\boldsymbol{\beta}}_m^{\mathrm{T}}\mathbf{x}_{in_i})$ for $1\le m\le p$. A different definition arises by minimizing instead of maximizing, which lead to different solutions (beyond the order) due to the use of a robust scale (see Li and Chen, 1985). However, both proposals will have the same asymptotic behavior. It is worth noticing that the robustness properties of the robust scale considered are inherited by the projection-pursuit estimators. When $f(t)=t$, Li and Chen (1985), for $k=1$, and Boente and Orellana (2001), for several populations, obtained some results in that direction which can be extended to the general projection-pursuit estimators defined by (1.2) for any increasing function f. Therefore, in Boente et al. (2006), the aim of introducing the score function f was not to achieve better robustness properties but to obtain, for a given scale functional, the optimal function f in the sense of minimizing the asymptotic variance of the proposed estimators. It is well known that when considering the variance, the optimal choice corresponds to $f(t)=\ln(t)$. When the populations share the same elliptical distribution up to location and scatter parameters, $\ln(t)$ is still optimal for any robust scale functional, under a proportionality model, i.e., it minimizes the asymptotic variance of the common direction estimators in the class of all increasing differentiable score functions f. Moreover, the same conclusion holds under a CPC model when the eigenvalues preserve their order among populations and if we restrict the class of possible score functions to the well-known Box and Cox class (see Propositions 2 and 3 in Boente et al., 2006).

Partial influence functions of the described projection-pursuit estimators were derived in Boente et al. (2006). The aim of this paper is to obtain under mild conditions their consistency and asymptotic normality. Asymptotic normality will be derived through Bahadur representations that are applicable to some common choices of robust dispersions. In this sense, our results extend those given by Cui et al. (2003), from one to several populations. Under elliptically symmetric models, our results simplify to provide the same asymptotic variances computed by Boente et al. (2006) using partial influence functions.

In Section 2, we describe the general projection-index estimators and the assumptions needed to derive the asymptotic behavior. Our main results are stated in Section 3 where the situation in which all the populations have elliptical distribution except for changes in location and scatter is also discussed. In Boente et al. (2006), it was assumed that $n_i=\tau_i N$, where $0<\tau_i<1$, are fixed numbers such that $\sum_{i=1}^{k}\tau_i=1$, i.e., that $\tau_{iN}=n_i/N=\tau_i$. To consider a more general framework, the asymptotic results will be stated by assuming that the sample sizes, n_i, go to infinity in such a way that $\tau_{iN}\to\tau_i\in(0,1)$ and $N^{\frac{1}{2}}(\tau_{iN}-\tau_i)\to 0$. This includes the situation in which n_i is the integer part of $\tau_i N$. Proofs are left to the Appendix.

2. Projection-index estimators: Notation and assumptions

As mentioned in the Introduction, under a CPC model, robust projection-pursuit estimators were introduced by Boente and Orellana (2001) who considered as score function f the identity function in (1.2) while in Boente et al. (2002) their partial influence function was obtained. Boente et al. (2006) proposed the general projection-pursuit estimators defined through (1.2) to estimate the common directions.

From now on $\mathbf{X}_i=(\mathbf{x}_{i1},\ldots,\mathbf{x}_{in_i})$ will denote independent vectors from k independent samples in $\mathbb{R}^p$ such that, for $1\le j\le n_i$, $\mathbf{x}_{ij}\sim F_i$, where F_i is a p-dimensional distribution with location parameter $\boldsymbol{\mu}_i$ and scatter matrix $\boldsymbol{\Sigma}_i$ satisfying (1.1). Note that for any unidimensional scale estimator $s^2(\mathbf{a}^{\mathrm{T}}\mathbf{x}_{i1},\ldots,\mathbf{a}^{\mathrm{T}}\mathbf{x}_{in_i})=s^2(\mathbf{a}^{\mathrm{T}}(\mathbf{x}_{i1}-\mathbf{b}),\ldots,\mathbf{a}^{\mathrm{T}}(\mathbf{x}_{in_i}-\mathbf{b}))$ for any $\mathbf{b}\in\mathbb{R}^p$, thus, as in Boente et al. (2006), without loss of generality, we may assume that $\boldsymbol{\mu}_i = 0$. Denote by $F_i[\mathbf{a}]$ the distribution of $\mathbf{a}^{\mathrm{T}}\mathbf{x}_{i1}$, and by F the product measure $F = F_1 \times F_2 \cdots \times F_k$. Let $\mathcal{F}_1$ be the one dimensional distribution space, $\mathcal{S}_p$ the p-dimensional unit sphere and $\mathbf{I}_p$ the identity matrix in $\mathbb{R}^{p\times p}$.

Moreover, let ς be a projection index, i.e., a functional $\varsigma : \mathcal{F}_1 \to \mathbb{R}_{\geq 0}$ and $\sigma(\cdot)$ a univariate scale functional. Denote by $s^2_{i,n_i} : \mathbb{R}^p \to \mathbb{R}$ and $\varsigma_{i,n_i} : \mathbb{R}^p \to \mathbb{R}$ the functions $s^2_{i,n_i}(\mathbf{a}) = \sigma^2(\mathbf{a}^{\mathrm{T}}\mathbf{X}_i)$ and $\varsigma_{i,n_i}(\mathbf{a}) = \varsigma(\mathbf{a}^{\mathrm{T}}\mathbf{X}_i)$, respectively, where $\varsigma(\mathbf{a}^{\mathrm{T}}\mathbf{X}_i)$ and $\sigma^2(\mathbf{a}^{\mathrm{T}}\mathbf{X}_i)$ stand for the functionals ς and σ computed at the empirical distribution of $\mathbf{a}^{\mathrm{T}}\mathbf{x}_{i1}, \ldots, \mathbf{a}^{\mathrm{T}}\mathbf{x}_{in_i}$, respectively. Analogously, $\sigma_i : \mathbb{R}^p \to \mathbb{R}$ and $\varsigma_i : \mathbb{R}^p \to \mathbb{R}$ will stand for $\sigma_i(\mathbf{a}) = \sigma(F_i[\mathbf{a}])$ and $\varsigma_i(\mathbf{a}) = \varsigma(F_i[\mathbf{a}])$, respectively. The estimators defined in Boente et al. (2006) correspond to the choice $\varsigma(F) = f(\sigma^2(F))$. We will assume that $\varsigma_i(\mathbf{a}) = \varsigma_i(-\mathbf{a})$ and $\varsigma_{i,n_i}(\mathbf{a}) = \varsigma_{i,n_i}(-\mathbf{a})$, that holds if $\varsigma(F) = f(\sigma^2(F))$. Denote by $\rho_N(\mathbf{a}) = \sum_{i=1}^k \tau_{iN}\, \varsigma_{i,n_i}(\mathbf{a})$ and $\rho(\mathbf{a}) = \sum_{i=1}^k \tau_i \varsigma_i(\mathbf{a})$. Then, a more general framework than (1.2) defines the estimators of the common directions as

$$\widehat{\boldsymbol{\beta}}_1 = \underset{\|\mathbf{a}\|=1}{\operatorname{argmax}}\, \rho_N(\mathbf{a}) \qquad \widehat{\boldsymbol{\beta}}_m = \underset{\mathbf{a}\in\widehat{\mathcal{B}}_m}{\operatorname{argmax}}\, \rho_N(\mathbf{a}) \quad 2 \leq m \leq p. \tag{2.1}$$

where $\widehat{\mathcal{B}}_m = \{\mathbf{a} \in \mathbb{R}^p : \|\mathbf{a}\| = 1, \mathbf{a}^{\mathrm{T}}\widehat{\boldsymbol{\beta}}_j = 0,\ \forall\ 1 \leq j \leq m-1\}$. The estimators of the eigenvalues of the i-th population are then, computed as

$$\widehat{\lambda}_{im} = \sigma^2(\widehat{\boldsymbol{\beta}}_m^{\mathrm{T}}\mathbf{X}_i) = s^2_{i,n_i}(\widehat{\boldsymbol{\beta}}_m), \quad 1 \leq m \leq p. \tag{2.2}$$

We will now introduce the statistical functional related to (2.1). The projection-index common directions functional $\boldsymbol{\beta}_\varsigma(F) = (\boldsymbol{\beta}_{1,\varsigma}(F), \ldots, \boldsymbol{\beta}_{p,\varsigma}(F))$ is defined as the solution of

$$\boldsymbol{\beta}_{1,\varsigma}(F) = \underset{\|\mathbf{a}\|=1}{\operatorname{argmax}}\, \rho(\mathbf{a}) \qquad \boldsymbol{\beta}_{m,\varsigma}(F) = \underset{\mathbf{a}\in\mathcal{B}_m}{\operatorname{argmax}}\, \rho(\mathbf{a}) \quad 2 \leq m \leq p, \tag{2.3}$$

where $\mathcal{B}_m = \{\mathbf{a} \in \mathbb{R}^p : \|\mathbf{a}\| = 1, \mathbf{a}^{\mathrm{T}}\boldsymbol{\beta}_{\ell,\varsigma}(F) = 0,\ \forall\ 1 \leq \ell \leq m-1\}$. It is clear that both $\widehat{\boldsymbol{\beta}}_m$ and $\boldsymbol{\beta}_{m,\varsigma}(F)$ are defined except for a multiplicative factor -1. The eigenvalue functional is defined as

$$\lambda_{im,\varsigma,\sigma}(F) = \sigma^2(F_i[\boldsymbol{\beta}_{m,\varsigma}(F)]) \quad 1 \leq m \leq p,\ 1 \leq i \leq k. \tag{2.4}$$

Remark 2.1. When $\varsigma(F) = f(\sigma^2(F))$, conditions under which the functional defined through (2.3) will be Fisher-consistent for elliptical distributions were obtained in Boente et al. (2006), while the particular case in which $f(t) = t$ was studied in Boente and Orellana (2001). In particular, if the order between eigenvalues is preserved across populations the functionals related to $\varsigma(F) = f(\sigma^2(F))$ are consistent for any increasing function f. For more general CPC models, the populations need to verify, for instance, that $\nu_{f,1} > \cdots > \nu_{f,p}$ where $\nu_{f,j} = \sum_{i=1}^k \tau_i f(\lambda_{ij})$. On the other hand, if $\nu_{f,1} \geq \cdots \geq \nu_{f,p}$, Fisher-consistency can be obtained if for each $m \neq \ell$ there exists i_0 such that $\lambda_{i_0 m} \neq \lambda_{i_0 \ell}$ and some convexity conditions are required to the score function (see Boente et al. (2006), for details).

To simplify the notation, we will avoid the subscript ς and/or σ and so, we will indicate $\boldsymbol{\beta}_m(F) = \boldsymbol{\beta}_{m,\varsigma}(F)$ and $\lambda_{im}(F) = \lambda_{im,\varsigma,\sigma}(F)$. For $1 \leq m \leq p$, consider

$$\nu_m(F) = \max_{\mathbf{a}\in\mathcal{B}_m} \rho(\mathbf{a}) \quad \text{and} \quad \widehat{\nu}_m = \max_{\mathbf{a}\in\widehat{\mathcal{B}}_m} \rho_N(\mathbf{a}). \tag{2.5}$$

From now on, the notation $\dot{h}(\mathbf{x}, \mathbf{a})$ will be used for the derivative of the function $h(\mathbf{x}, \mathbf{a})$ with respect to $\mathbf{a}$. Throughout this paper we will consider the following set of assumptions

S0. For some $q \leq p$, we have that $\nu_1(F) > \nu_2(F) \cdots > \nu_q(F)$. Moreover, for $1 \leq m \leq q$, $\boldsymbol{\beta}_m(F)$ are unique except for changes in their sign.

S1. $\varsigma_{i,n_i}(\mathbf{a}) - \varsigma_i(\mathbf{a}) = n_i^{-1}\sum_{j=1}^{n_i} h_i(\mathbf{x}_{ij}, \mathbf{a}) + R_{i,n_i}$, where
- (a) $\varsigma_i(\mathbf{a})$ is a continuous function of $\mathbf{a}$.
- (b) $h_i(\mathbf{x}, \mathbf{a})$ is continuous in both variables.
- (c) $\varsigma_{i,n_i}(\mathbf{a})$ is a continuous function of $\mathbf{a}$ a.e.
- (d) $\mathbb{E}h_i(\mathbf{x}_{i1}, \mathbf{a}) = 0$ and $\mathbb{E}\left(\sup_{\mathbf{a}\in\mathcal{S}_p} |h_i(\mathbf{x}_{i1}, \mathbf{a})|\right) < \infty$.
- (e) $\sup_{\mathbf{a}\in\mathcal{S}_p} |R_{i,n_i}| \xrightarrow{p} 0$, i.e., $R_{i,n_i} = o_p(1)$ uniformly in $\mathbf{a} \in \mathcal{S}_p$.
- (f) $\mathbb{E}\left(\sup_{\mathbf{a}\in\mathcal{S}_p} h_i^2(\mathbf{x}_{i1}, \mathbf{a})\right) < \infty$ and $R_{i,n_i} = o_p\left(n_i^{-1/2}\right)$ uniformly in $\mathbf{a} \in \mathcal{S}_p$.

S2. $\varsigma_i(\mathbf{a})$ is twice continuously differentiable with respect to $\mathbf{a}$. $\dot{\varsigma}_i(\mathbf{a})$ and $\ddot{\varsigma}_i(\mathbf{a})$ will stand for its first and second derivatives, respectively.

S3. The function $\varsigma_{i,n_i}(\mathbf{a})$ is differentiable with respect to $\mathbf{a}$ for any $\mathbf{a} \in \mathcal{S}_p$, almost everywhere. Moreover, $\dot{\varsigma}_{i,n_i}(\mathbf{a}) - \dot{\varsigma}_i(\mathbf{a}) = n_i^{-1}\sum_{j=1}^{n_i} h_i^\star(\mathbf{x}_{ij}, \mathbf{a}) + \mathbf{o}_p(n_i^{-1/2})$, uniformly in $\mathbf{a} \in \mathcal{S}_p$, with $h_i^\star : \mathbb{R}^p \times \mathbb{R}^p \to \mathbb{R}^p$ such that
- (a) For any given $\mathbf{x}$, $h_i^\star(\mathbf{x}, \mathbf{a})$ is continuous in $\mathbf{a}$.
- (b) $\mathbb{E}\, h_i^\star(\mathbf{x}_{i1}, \mathbf{a}) = \mathbf{0}$ for all $\mathbf{a} \in \mathbb{R}^p$ and $\mathbb{E}\left(\sup_{\mathbf{a}\in\mathcal{S}_p} \|h_i^\star(\mathbf{x}_{i1}, \mathbf{a})\|^2\right) < \infty$.

S4. $s_{i,n_i}(\mathbf{a}) - \sigma_i(\mathbf{a}) = n_i^{-1} \sum_{j=1}^{n_i} h_{i,\sigma}(\mathbf{x}_{ij}, \mathbf{a}) + R_{i,n_i,\sigma}$, where
   (a) $\sigma_i(\mathbf{a})$ is a continuous function of $\mathbf{a}$.
   (b) $\mathbb{E}h_{i,\sigma}(\mathbf{x}_{i1}, \mathbf{a}) = 0$ and $\mathbb{E}\left(\sup_{\mathbf{a}\in\mathcal{S}_p} |h_{i,\sigma}(\mathbf{x}_{i1}, \mathbf{a})|\right) < \infty$.
   (c) $h_{i,\sigma}(\mathbf{x}, \mathbf{a})$ is continuous in both variables.
   (d) $R_{i,n_i,\sigma} = o_p(1)$ uniformly in $\mathbf{a} \in \mathcal{S}_p$.
   (e) $\mathbb{E}\left(\sup_{\mathbf{a}\in\mathcal{S}_p} h^2_{i,\sigma}(\mathbf{x}_{i1}, \mathbf{a})\right) < \infty$ and $R_{i,n_i,\sigma} = o_p\left(n_i^{-1/2}\right)$, uniformly in $\mathbf{a} \in \mathcal{S}_p$.

S5. The families of functions $\mathcal{H}_i = \{f(\mathbf{x}) = h_i(\mathbf{x}, \mathbf{a})\ \mathbf{a} \in \mathcal{S}_p\}, \mathcal{H}_{i,\sigma} = \{f(\mathbf{x}) = h_{i,\sigma}(\mathbf{x}, \mathbf{a})\ \mathbf{a} \in \mathcal{S}_p\}$ and $\mathcal{H}^\star_{i,\ell} = \{f(\mathbf{x}) = h^\star_{i,\ell}(\mathbf{x}, \mathbf{a})\ \mathbf{a} \in \mathcal{S}_p\}$, for $1 \le i \le k$, $1 \le \ell \le p$, with envelopes $H_i(\mathbf{x}) = \sup_{\mathbf{a}\in\mathcal{S}_p} |h_i(\mathbf{x}, \mathbf{a})|$, $H_{i,\sigma}(\mathbf{x}) = \sup_{\mathbf{a}\in\mathcal{S}_p} |h_{i,\sigma}(\mathbf{x}, \mathbf{a})|$ and $H^\star_{i,\ell}(\mathbf{x}) = \sup_{\mathbf{a}\in\mathcal{S}_p} |h^\star_{i,\ell}(\mathbf{x}, \mathbf{a})|$, respectively, have finite uniform-entropy, where $h^\star_{i,\ell}(\mathbf{x}, \mathbf{a})$ stands for the ℓ-th component of $h^\star_i(\mathbf{x}, \mathbf{a})$.

Remark 2.2. As in Cui et al. (2003), **S0** states that $\boldsymbol{\beta}_m(F)$ and $-\boldsymbol{\beta}_m(F)$ are considered equivalent and it does not matter which one we take. In this sense, the convergence $\widehat{\boldsymbol{\beta}}_m \xrightarrow{p} \boldsymbol{\beta}_m(F)$ mean convergence in axis, not in the signed vector. In order to identify the vectors (functional and estimators), one can choose them such that the component with its largest absolute value will be positive, for instance. It is worth noticing that conditions **S1** to **S4** are analogous to Conditions 1 to 5 in Cui et al. (2003). On the other hand, **S5** is fulfilled if, for instance, $|h_i(\mathbf{x}, \mathbf{a}_1) - h_i(\mathbf{x}, \mathbf{a}_2)| \le G_i(\mathbf{x})\|\mathbf{a}_1 - \mathbf{a}_2\|$ with $\mathbb{E}G_i^2(\mathbf{x}_{i1}) < \infty$, see for instance, van der Vaart and Wellner (1996). On the other hand, if $h_i(\mathbf{x}, \mathbf{a}) = \chi_i(\mathbf{a}^{\mathrm{T}}\mathbf{x}/g(\mathbf{a}))$ where $\chi_i : \mathbb{R} \to \mathbb{R}$ is a bounded function with bounded variation and $g : \mathbb{R}^p \to \mathbb{R}$, then, **S5** holds. This result follows easily using the permanence properties stated in van der Vaart and Wellner (1996) and that the fact that, given $\epsilon > 0$, for any classes of functions $\mathcal{G}_1$ and $\mathcal{G}_2$, if $\mathcal{G} = \{g = g_1 + g_2 : g_i \in \mathcal{G}_i,\ i = 1, 2\}$, then $N\left(\epsilon, \mathcal{G}, L^2(Q)\right) \le N\left(\epsilon/2, \mathcal{G}_1, L^2(Q)\right) .N\left(\epsilon/2, \mathcal{G}_2, L^2(Q)\right)$.

Let us define

- $\mathbf{u}_m = -N^{-1/2} \sum_{i=1}^k \sum_{j=1}^{n_i} h^\star_i(\mathbf{x}_{ij}, \boldsymbol{\beta}_m(F))$
- $\mathbf{P}_{m+1} = \mathbf{I}_p - \sum_{j=1}^m \boldsymbol{\beta}_j(F)\boldsymbol{\beta}_j(F)^{\mathrm{T}}$ the projection matrix over the linear space orthogonal to that spanned by $\boldsymbol{\beta}_1(F), \ldots, \boldsymbol{\beta}_m(F)$,
- $\mathbf{B}_{jm} = \boldsymbol{\beta}_j(F)^{\mathrm{T}}\ \dot\rho(\boldsymbol{\beta}_m(F))\ \mathbf{I}_p + \boldsymbol{\beta}_j(F)\ \dot\rho(\boldsymbol{\beta}_m(F))^{\mathrm{T}}$,
- $\mathbf{A}_m = \mathbf{P}_{m+1}\ \ddot\rho(\boldsymbol{\beta}_m(F)) - \boldsymbol{\beta}_m(F)^{\mathrm{T}}\ \dot\rho(\boldsymbol{\beta}_m(F))\mathbf{I}_p - \sum_{j=1}^{m-1} \boldsymbol{\beta}_j(F)^{\mathrm{T}}\ \dot\rho(\boldsymbol{\beta}_m(F))\boldsymbol{\beta}_m(F)\ \boldsymbol{\beta}_j(F)^{\mathrm{T}}$.
- $\mathbf{Z}_0 = 0$ and define $\mathbf{Z}_m$ recursively as $\mathbf{Z}_m = \sum_{j=0}^{m-1} \mathbf{A}_m^{-1}\mathbf{B}_{jm}\mathbf{Z}_j + \mathbf{A}_m^{-1}\mathbf{P}_{m+1}\mathbf{u}_m$, for $1 \le m \le q$, provided that $\mathbf{A}_j^{-1}$ exists for $1 \le j \le m$.

  It is clear that the process $\mathbf{Z}_m$ can by represented by $\mathbf{Z}_m = \sum_{j=0}^{m-1} \mathbf{C}_{jm}\mathbf{u}_j$, for some sequence of matrices $\mathbf{C}_{jm}$ depending on $\mathbf{A}_m$, $\mathbf{B}_{jm}$ and $\mathbf{P}_{m+1}$.
- $\boldsymbol{\xi}_{i,m}(\mathbf{x}) = \sum_{\ell=1}^m \mathbf{C}_{\ell m}\ h^\star_i(\mathbf{x}, \boldsymbol{\beta}_\ell(F))$, for $\mathbf{x} \in \mathbb{R}^p$.
- $\boldsymbol{\xi}_m(\overrightarrow{\mathbf{x}}) = \sum_{i=1}^k \tau_i^{1/2}\boldsymbol{\xi}_{i,m}(\mathbf{x}_i)$, where $\overrightarrow{\mathbf{x}} = (\mathbf{x}_1, \ldots, \mathbf{x}_k)$ and $\mathbf{x}_i \in \mathbb{R}^p$.

3. Main results

3.1. Consistency and asymptotic distribution

The following theorem establishes the consistency of the estimators of the common directions defined through (2.1), under mild conditions. Its proof can be found in Boente et al. (2009). From their consistency, it is easy to derive that of the eigenvalue estimators (2.2) and also that of the estimators of the i-scatter matrix defined as $\widehat{\mathbf{V}}_i = \sum_{j=1}^p \widehat\lambda_{im}\widehat{\boldsymbol{\beta}}_m\widehat{\boldsymbol{\beta}}_m^{\mathrm{T}}$.

Theorem 3.1. *Let $\mathbf{X}_i = (\mathbf{x}_{i1}, \ldots, \mathbf{x}_{in_i})$ denote independent vectors from k independent samples in $\mathbb{R}^p$ such that, for $1 \le j \le n_i$, $\mathbf{x}_{ij} \sim F_i$, where F_i is a p-dimensional distribution. Moreover, assume that $n_i = \tau_{iN}N$, with $0 < \tau_{iN} < 1$ such that $\sum_{i=1}^k \tau_{iN} = 1$ and $\tau_{iN} \to \tau_i \in (0, 1)$. Let $\boldsymbol{\beta}_m$, λ_{im} and ν_m be the functionals defined through (2.3)–(2.5), respectively. Let $\widehat{\boldsymbol{\beta}}_m$ and $\widehat\lambda_{im}$ be the estimators defined in (2.1) and (2.2), respectively. Under* S0, S1 (a) *to* (e) *and* S4 (a) *to* (d), *we have that, for $1 \le m \le q$, $\widehat{\boldsymbol{\beta}}_m \xrightarrow{p} \boldsymbol{\beta}_m(F)$ and $\widehat\lambda_{im} \xrightarrow{p} \lambda_{im}(F)$, for $1 \le i \le k$ as $N \to \infty$.*

The following theorem gives a Bahadur representation for the estimators $\widehat{\boldsymbol{\beta}}_m$ and $\widehat\lambda_{im}$ which allows to derive easily their asymptotic distribution. A sketch of its proof is given in the Appendix.

Theorem 3.2. *Under the conditions of Theorem 3.1, if, in addition, $N^{\frac{1}{2}}(\tau_{iN} - \tau_i) \to 0$,* **S1** *to* **S5** *hold and the matrices $\mathbf{A}_m$, $1 \le m \le q$, are non-singular, we have that, for $1 \le m \le q$,*

$$\widehat{\boldsymbol{\beta}}_m - \boldsymbol{\beta}_m(F) = \frac{1}{N}\sum_{i=1}^{k}\sum_{j=1}^{n_i} \boldsymbol{\xi}_{i,m}(\mathbf{x}_{ij}) + \mathbf{o}_p(N^{-1/2}) \tag{3.1}$$

$$\widehat{\lambda}_{im} - \lambda_{im}(F) = \frac{1}{n_i}\sum_{j=1}^{n_i} h_{i,\sigma}(\mathbf{x}_{ij}, \boldsymbol{\beta}_m(F)) + o_p(n_i^{-1/2}). \tag{3.2}$$

Theorem 3.2 entails that, for $1 \le i \le k$, the joint distribution of $N^{-1/2}(\widehat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1(F), \ldots, \widehat{\boldsymbol{\beta}}_q - \boldsymbol{\beta}_q(F), \widehat{\lambda}_{i1} - \lambda_{i1}(F), \ldots, \widehat{\lambda}_{iq} - \lambda_{iq}(F))$ converges to a multivariate normal distribution with mean $\mathbf{0}$ and covariance matrix $\mathrm{cov}_F\left(\xi_1(\overrightarrow{\mathbf{x}_1}), \ldots, \xi_q(\overrightarrow{\mathbf{x}_1}), \ldots, h_{i,\sigma}(\mathbf{x}_{i1}, \boldsymbol{\beta}_1(F)), \ldots, h_{i,\sigma}(\mathbf{x}_{i1}, \boldsymbol{\beta}_q(F))\right)$, where $\overrightarrow{\mathbf{x}_1} = (\mathbf{x}_{11}, \ldots, \mathbf{x}_{k1})$.

It is worth noticing that when dealing with only one population, i.e., when $k = 1$, Theorem 3.2 provides the Bahadur expansion given in Cui et al. (2003).

3.2. Example: General projection-pursuit estimates under the CPC model

Let σ be a univariate robust scale functional and $f : \mathbb{R}^+ \to \mathbb{R}$ an increasing score function. Considering the functional $\varsigma(\cdot) = f\{\sigma^2(\cdot)\}$ in (2.1), we obtain the estimators defined through (1.2) in Boente et al. (2006). As mentioned above, these authors studied conditions for the Fisher-consistency of the common direction functional defined through ς and they also provide an expression for their partial influence functions. From the general results in Pires and Branco (2002), the asymptotic variance of the common direction and eigenvalue estimates, i.e., the variance of the approximating normal distribution, was also computed in Boente et al. (2006).

The aim of this section is to show that the expansion obtained Theorem 3.2 allows to obtain under mild conditions the expressions obtained by these authors. Under regularity conditions on σ, we get that $h_{i,\sigma}(\mathbf{x}, \mathbf{a}) = 2\sigma(F_i[\mathbf{a}])\psi_i(\mathbf{x}, \mathbf{a})$, $h_i(\mathbf{x}, \mathbf{a}) = f'\left(\sigma^2(F_i[\mathbf{a}])\right) 2\sigma(F_i[\mathbf{a}])\psi_i(\mathbf{x}, \mathbf{a})$ and

$$\begin{aligned} h_i^\star(\mathbf{x}, \mathbf{a}) &= \dot{h}_i(\mathbf{x}, \mathbf{a}) = f'\left(\sigma^2(F_i[\mathbf{a}])\right)\left(2\sigma(F_i[\mathbf{a}])\,\psi_i^\star(\mathbf{x}, \mathbf{a}) + 2\dot{\sigma}(F_i[\mathbf{a}])\,\psi_i(\mathbf{x}, \mathbf{a})\right) \\ &\quad + 4 f''\left(\sigma^2(F_i[\mathbf{a}])\right)\dot{\sigma}(F_i[\mathbf{a}])\sigma^2(F_i[\mathbf{a}])\psi_i(\mathbf{x}, \mathbf{a}) \qquad 1 \le i \le k, \end{aligned}$$

where $\psi_i(\mathbf{x}, \mathbf{a}) = \mathrm{IF}(\mathbf{x}, \sigma_{\mathbf{a}}; F_i)$, $\psi_i^\star(\mathbf{x}, \mathbf{a}) = \mathrm{IF}(\mathbf{x}, \dot{\sigma}_{\mathbf{a}}; F_i) = \dot{\psi}_i(\mathbf{x}, \mathbf{a})$, $\sigma_{\mathbf{a}} : \mathcal{F}_1 \to \mathbb{R}^+$ is such that $\sigma_{\mathbf{a}}(F) = \sigma(F[\mathbf{a}])$ and $\dot{\sigma}_{\mathbf{a}}(F)$ is the derivative of $\sigma_{\mathbf{a}}(F)$ respect to $\mathbf{a}$.

Let us consider the following assumptions.

A1. $\sigma(\cdot)$ is a robust scale functional, equivariant under scale transformations.

A2. F_i is an ellipsoidal distribution with location parameter $\boldsymbol{\mu}_i = 0$ and scatter matrix $\boldsymbol{\Sigma}_i = \mathbf{C}_i\mathbf{C}_i^{\mathrm{T}}$ satisfying (1.1). Moreover, the scatter matrices $\boldsymbol{\Sigma}_i$ and the scale functional are such that $\sigma(G_{0,i}) = 1$, with $G_{0,i}$ the distribution of z_{i1}, $1 \le i \le k$, where $\mathbf{z}_i = \mathbf{C}_i^{-1}\mathbf{x}_{i1}$ has spherical distribution G_i, for all $1 \le i \le k$.

A3. For $G = G_{0,i}$, $1 \le i \le k$, the function $(\varepsilon, y) \to \sigma((1-\varepsilon)G + \varepsilon\Delta_y)$ is twice continuously differentiable in $(0, y)$, $y \in \mathbb{R}$ where Δ_y denotes the point mass at y.

A4. f is a twice continuously differentiable function.

A5. For any $1 \le m \le p$, the eigenvalues $\eta_{m\ell} = \sum_{i=1}^k \tau_i f'(\lambda_{im})\lambda_{i\ell}$ of $\widetilde{\boldsymbol{\Sigma}}_m = \sum_{i=1}^k \tau_i f'(\lambda_{im})\boldsymbol{\Sigma}_i$ are such that $\eta_{m\ell} \ne \eta_m = \eta_{mm} = \sum_{i=1}^k \tau_i f'(\lambda_{im})\lambda_{im}$ for $\ell \ne m$.

Remark 3.1. Note that if all populations share the same elliptical distribution up to location and scatter, i.e., if $G_i = G$ for all i, then the scale functional can be calibrated so that $\sigma(G_0) = 1$, where G_0 is the distribution of z_1 when $\mathbf{z} \sim G$. Otherwise, assumption **A2** provides a way to choose the scatter matrices related to the scale functional. If $E\|\mathbf{x}_{i1}\|^2 < \infty$, it is well known that $\boldsymbol{\Sigma}_i$ is up to a constant α_i the covariance matrix of $\mathbf{x}_{i1}$. Assumption **A2**, states that in this situation, the constant does not need to be equal to 1 but may depend on the population when the related spherical distributions are not equal. This parametrization does not change the percentage of total variation explained in each population by the common directions. It is worth noticing that under **A1** to **A4**, the Bahadur expansions required in **S1**, **S3** and **S4** can be obtained under mild conditions on the scale functional, see Cui et al. (2003) for a discussion. On the other hand, as mentioned in Boente et al. (2006), under a proportional model $\eta_{m\ell} = \lambda_\ell \sum_{i=1}^k \tau_i f'(\lambda_{im})\,\rho_i$ and so, $\eta_{m\ell} \ne \eta_{mm}$ for $\ell \ne m$ if and only if the eigenvalues of the first population are different, which is a usual assumption in order to identify the common directions. More generally, **A5** holds if the eigenvalues preserve the order among populations, i.e., $\lambda_{i1} \ge \lambda_{i2} \ge \ldots \lambda_{ip}$ and if for each $\ell \ne m$ there exists $1 \le i \le k$ such that $\lambda_{i\ell} \ne \lambda_{im}$.

Typically, the influence function of a robust scale functional is bounded. Therefore, using that if **A1** to **A4** hold, $\psi_i(\mathbf{x}, \mathbf{a}) = \sigma_{\mathbf{a}}(F_i)\mathrm{IF}(\mathbf{a}^{\mathrm{T}}\mathbf{x}/\sigma_{\mathbf{a}}(F_i), \sigma; G_{0,i})$ and $\sigma_{\mathbf{a}}^2(F_i) = \mathbf{a}^{\mathrm{T}}\boldsymbol{\Sigma}_i\mathbf{a}$, we get easily that $h_{i,\sigma}(\mathbf{x}, \mathbf{a})$ and $h_i(\mathbf{x}, \mathbf{a})$ are bounded. Moreover, if $\mathrm{IF}(y, \sigma; G_{0,i}) = \chi_i(y)$ for some function χ_i of bounded variation, as is, for instance, the case of an M-scale function, then $\mathcal{H}_{i,\sigma}$ and $\mathcal{H}_i$ will have finite uniform-entropy. On the other hand, if $\boldsymbol{\Sigma}_i$ is non-singular, using that

$$\dot{\psi}_i(\mathbf{x}, \mathbf{a}) = \sigma_{\mathbf{a}}^{-1}(F_i)\mathrm{IF}\left(\frac{\mathbf{a}^{\mathrm{T}}\mathbf{x}}{\sigma_{\mathbf{a}}(F_i)}, \sigma; G_{0,i}\right)\boldsymbol{\Sigma}_i\mathbf{a} + \mathrm{DIF}\left(\frac{\mathbf{a}^{\mathrm{T}}\mathbf{x}}{\sigma_{\mathbf{a}}(F_i)}, \sigma; G_{0,i}\right)\left(\mathbf{I}_p - \frac{1}{\sigma_{\mathbf{a}}^2(F_i)}\boldsymbol{\Sigma}_i\mathbf{a}\mathbf{a}^{\mathrm{T}}\right)\mathbf{x},$$

where $\text{DIF}(y, \sigma; \mathbf{G}_0)$ denotes the derivative of the influence function $\text{IF}(y, \sigma; \mathbf{G}_0)$ with respect to y, it is easy to see that the first term on the right hand side will be bounded while the second one, can be unbounded for some values of $\mathbf{a}$, for instance, when $\mathbf{a} = \boldsymbol{\beta}_m$. Hence, to ensure that the envelope $H^{\star}_{i,\ell}$ has second finite moment it is enough to require that $\mathbb{E}\|\mathbf{x}_{i1}\|^2 < \infty$. Therefore, if f' and f'' are functions of bounded variation, $\mathcal{H}^{\star}_{i,\ell}$ will have finite entropy, if $\text{IF}(y, \sigma; G_{0,i}) = \chi_i(y)$ for some continuously differentiable function χ_i such that $\chi_{i,1}(y) = \chi_i'(y)$ and $\chi_{i,2}(y) = y\chi_i'(y)$ have bounded variation (see Boente et al. (2009), for details).

Under **A1** to **A4**, we get that $\boldsymbol{\beta}_m(F) = \boldsymbol{\beta}_m$, $\sigma_i^2(\boldsymbol{\beta}_m) = \lambda_{im}$, $\varsigma_i(\boldsymbol{\beta}_m) = f(\lambda_{im})$, $\dot{\varsigma}_i(\boldsymbol{\beta}_m) = f'(\lambda_{im})\,\lambda_{im}\,\boldsymbol{\beta}_m$ and so,

$$\psi_i(\mathbf{x}, \boldsymbol{\beta}_m) = \sqrt{\lambda_{im}}\,\text{IF}\left(\frac{\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}_m}{\sqrt{\lambda_{im}}}, \sigma; G_{0,i}\right),$$

$$\dot{\psi}_i(\mathbf{x}, \boldsymbol{\beta}_m) = \sqrt{\lambda_{im}}\boldsymbol{\beta}_m\text{IF}\left(\frac{\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}_m}{\sqrt{\lambda_{im}}}, \sigma; G_{0,i}\right) + \text{DIF}\left(\frac{\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}_m}{\sqrt{\lambda_{im}}}, \sigma; G_{0,i}\right)(\mathbf{I}_p - \boldsymbol{\beta}_m\boldsymbol{\beta}_m^{\mathrm{T}})\mathbf{x}.$$

Moreover, we have that $\rho(\boldsymbol{\beta}_m) = \sum_{i=1}^k \tau_i f(\lambda_{im}) = \nu_m$, $\dot{\rho}(\boldsymbol{\beta}_m) = 2\eta_m\boldsymbol{\beta}_m$, $\ddot{\rho}(\boldsymbol{\beta}_m) = 4\sum_{i=1}^k \tau_i f''(\lambda_{im})\lambda_{im}^2\boldsymbol{\beta}_m\boldsymbol{\beta}_m^{\mathrm{T}} + 2\widetilde{\boldsymbol{\Sigma}}_m$ and

$$\dot{h}_i(\mathbf{x}, \boldsymbol{\beta}_m) = 2\sqrt{\lambda_{im}}f'(\lambda_{im})\left[\dot{\psi}_i(\mathbf{x}, \boldsymbol{\beta}_m) + \psi_i(\mathbf{x}, \boldsymbol{\beta}_m)\boldsymbol{\beta}_m\right] + 4\lambda_{im}^{\frac{3}{2}} f''(\lambda_{im})\,\psi_i(\mathbf{x}, \boldsymbol{\beta}_m)\boldsymbol{\beta}_m.$$

Therefore, using that $\boldsymbol{\beta}_j^{\mathrm{T}}\boldsymbol{\beta}_m = 0$ for $1 \le j \le m-1$, we get that $\mathbf{A}_m = 2\sum_{j=m+1}^p \eta_{mj}\boldsymbol{\beta}_j\boldsymbol{\beta}_j^{\mathrm{T}} - 2\eta_m\mathbf{I}_p$, which after straightforward calculations, using (3.1), lead to

$$N^{1/2}(\widehat{\boldsymbol{\beta}}_m - \boldsymbol{\beta}_m) = N^{-1/2}\sum_{i=1}^k\sum_{j=1}^{n_i}\sum_{\ell=1}^{m-1}\sqrt{\lambda_{i\ell}}f'(\lambda_{i\ell})\,\text{DIF}\left(\frac{\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta}_\ell}{\sqrt{\lambda_{i\ell}}}, \sigma; G_{0,i}\right)\frac{1}{(\eta_{\ell m} - \eta_\ell)}(\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta}_m)\,\boldsymbol{\beta}_\ell$$
$$+ N^{-1/2}\sum_{i=1}^k\sum_{j=1}^{n_i}\sum_{l=m+1}^{p}\sqrt{\lambda_{im}}f'(\lambda_{im})\,\text{DIF}\left(\frac{\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta}_m}{\sqrt{\lambda_{im}}}, \sigma; G_{0,i}\right)\frac{1}{(\eta_m - \eta_{m\ell})}(\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta}_\ell)\,\boldsymbol{\beta}_\ell + \mathbf{o}_p(1).$$

When $G_i = G$, for all i, i.e., when all the populations share the same elliptical distribution up to location and scatter, the above expansion is that suggested by the partial influence functions obtained in Boente et al. (2006) and the expansion given in Pires and Branco (2002), see Boente et al. (2009) for a detailed derivation. In particular, when $f(t) = t$, we obtain the Bahadur representation of the estimators defined in Boente and Orellana (2001), suggested by the partial influence functions derived in Boente et al. (2002).

As in Cui et al. (2003) our results demonstrate the need of using a robust dispersion measure with a continuous influence function. As mentioned by these authors, a lower order of convergence may be attained when using a robust dispersion with a non-differentiable influence function. This fact provides evidence against using as dispersion measure the median absolute deviation when estimating robustly the common principal components. The simulation study performed in Boente et al. (2006) do not allow us to illustrate clearly this phenomenon since only plots of the density estimators of the cosines of the angle between the true and the estimated direction were given. On the other hand, the tables reported in Rodrigues (2003) show that the robust estimators of the common directions based on the median absolute deviation have a very low efficiency, in particular, in the simulation study involving $k = 3$ populations and when estimating the eigenvalues.

Acknowledgements

The authors wish to thank two anonymous referees for their valuable comments which lead to an improved version of the original paper. This research was partially supported by Grants X-018 and X-447 from the Universidad de Buenos Aires, PIP 5505, 0592 and 0216 from CONICET and PICT 21407, 00883 and 00821 from ANPCYT, Argentina.

Appendix

From now on, to simplify the notation $\boldsymbol{\beta}_j$ and λ_{ij} will stand for $\boldsymbol{\beta}_j(F)$ and $\lambda_{ij}(F)$, respectively. The proof of Theorem 3.2 follows the same steps as those considered in Cui et al. (2003), we skip most of the details that can be found in Boente et al. (2009). We state the lemmas needed to derive Theorem 3.2, details of their proof can be found in Boente et al. (2009).

Lemma A.1. *Under the conditions of Theorem 3.2, we have that*

(i) $\rho_N(\widehat{\boldsymbol{\beta}}_m) - \rho_N(\boldsymbol{\beta}_m) - \rho(\widehat{\boldsymbol{\beta}}_m) + \rho(\boldsymbol{\beta}_m) = o_p(N^{-1/2})$

(ii) $\dot{\rho}(\widehat{\boldsymbol{\beta}}_m) + N^{-1}\sum_{i=1}^k\sum_{j=1}^{n_i} h_i^{\star}(\mathbf{x}_{ij}, \widehat{\boldsymbol{\beta}}_m) - \left\{\dot{\rho}(\boldsymbol{\beta}_m) + N^{-1}\sum_{i=1}^k\sum_{j=1}^{n_i} h_i^{\star}(\mathbf{x}_{ij}, \boldsymbol{\beta}_m)\right\} = \ddot{\rho}(\boldsymbol{\beta}_m)(\widehat{\boldsymbol{\beta}}_m - \boldsymbol{\beta}_m) + \mathbf{o}_p(\|\widehat{\boldsymbol{\beta}}_m - \boldsymbol{\beta}_m\|) + \mathbf{o}_P(N^{-1/2})$.

To derive the Bahadur expansions given in (3.1), we need to obtain, as in Cui et al. (2003) some identities satisfied by the common direction estimators.Using the Lagrange multiplier method, we have that $\widehat{\boldsymbol{\beta}}_1$ maximizes $G_1(\mathbf{a}, \mu_1) =$

$\rho_N(\mathbf{a}) - \mu_1(\mathbf{a}^{\mathrm{T}}\mathbf{a} - 1)$, where $\mu_1 \in \mathbb{R}$. Hence, differentiating G_1 respect to $\mathbf{a}$, we get that $\dot{\rho}_N(\widehat{\boldsymbol{\beta}}_1) = 2\mu_1\widehat{\boldsymbol{\beta}}_1$. Note that $N^{1/2}(\tau_{iN} - \tau_i) \to 0$ implies that $N^{1/2}\sup_{\|\mathbf{a}\|=1} |\sum_{i=1}^{k}(\tau_{iN} - \tau_i)\dot{\varsigma}_i(\mathbf{a})| \to 0$ and so, using **S3**, we have

$$\dot{\rho}_N(\mathbf{a}) = \dot{\rho}(\mathbf{a}) + \frac{1}{N}\sum_{i=1}^{k}\sum_{j=1}^{n_i} h_i^{\star}(\mathbf{x}_{ij}, \mathbf{a}) + \mathbf{o}_p(N^{-1/2}) \tag{A.1}$$

which entails that $\dot{\rho}(\widehat{\boldsymbol{\beta}}_1) + N^{-1}\sum_{i=1}^{k}\sum_{j=1}^{n_i} h_i^{\star}(\mathbf{x}_{ij}, \widehat{\boldsymbol{\beta}}_1) = 2\mu_1\widehat{\boldsymbol{\beta}}_1 + \mathbf{o}_p(N^{-1/2})$. Let, for $1 \le m \le p$, $\widehat{\mathbf{P}}_{m+1} = \mathbf{I}_p - \sum_{j=1}^{m}\widehat{\boldsymbol{\beta}}_j\widehat{\boldsymbol{\beta}}_j^{\mathrm{T}}$ be the projection matrix over the linear space orthogonal to that spanned by $\widehat{\boldsymbol{\beta}}_1, \dots, \widehat{\boldsymbol{\beta}}_m$. Then, we have that $\widehat{\mathbf{P}}_2\widehat{\boldsymbol{\beta}}_1 = 0$ and so, we get $\widehat{\mathbf{P}}_2\left(\dot{\rho}(\widehat{\boldsymbol{\beta}}_1) + N^{-1}\sum_{i=1}^{k}\sum_{j=1}^{n_i} h_i^{\star}(\mathbf{x}_{ij}, \widehat{\boldsymbol{\beta}}_1)\right) = \mathbf{o}_p(N^{-1/2})$. Similarly, we have that $\widehat{\boldsymbol{\beta}}_m$ maximizes $G_m(\mathbf{a}, \mu_1, \dots, \mu_m) = \rho_N(\mathbf{a}) - \sum_{j=1}^{m-1}\mu_j\widehat{\boldsymbol{\beta}}_j^{\mathrm{T}}\mathbf{a} - \mu_m(\mathbf{a}^{\mathrm{T}}\mathbf{a} - 1)$, for $1 \le m \le q$, which implies that $\dot{\rho}_N(\widehat{\boldsymbol{\beta}}_m) = \sum_{j=1}^{m-1}\mu_j\widehat{\boldsymbol{\beta}}_j + 2\mu_m\widehat{\boldsymbol{\beta}}_m$. Therefore, using again **S3**, the fact that $N^{1/2}(\tau_{iN} - \tau_i) \to 0$, (A.1) and that $\widehat{\mathbf{P}}_{m+1}\widehat{\boldsymbol{\beta}}_j = 0$, $1 \le j \le m$, we obtain

$$\widehat{\mathbf{P}}_{m+1}\left(\dot{\rho}(\widehat{\boldsymbol{\beta}}_m) + N^{-1}\sum_{i=1}^{k}\sum_{j=1}^{n_i} h_i^{\star}(\mathbf{x}_{ij}, \widehat{\boldsymbol{\beta}}_m)\right) = \mathbf{o}_p(N^{-1/2}). \tag{A.2}$$

The Eq. (A.2) has its asymptotic version given in the next lemma whose proof we omit since it follows as in Cui et al. (2003).

Lemma A.2. *Under the conditions of Theorem 3.2. the following equation holds* $\mathbf{P}_{m+1}\dot{\rho}(\boldsymbol{\beta}_m) = 0$.

Moreover, we have the following relation between the estimators $\widehat{\boldsymbol{\beta}}_m$ and the projection matrix, which follows using standard arguments.

Lemma A.3. *Under the conditions of Theorem 3.2. we have that, for all* $\mathbf{b} \in \mathbb{R}^p$

$$(\widehat{\mathbf{P}}_{m+1} - \mathbf{P}_{m+1})\mathbf{b} = -\sum_{j=1}^{m}(\boldsymbol{\beta}_j^{\mathrm{T}}\mathbf{b}\mathbf{I}_p + \boldsymbol{\beta}_j\mathbf{b}^{\mathrm{T}})(\widehat{\boldsymbol{\beta}}_j - \boldsymbol{\beta}_j) + \mathbf{O}_p\left(\|\mathbf{b}\|\sum_{i=1}^{m}\|\widehat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_i\|^2\right).$$

Lemma A.4 gives the key point to obtain the equality given in (3.1). We omit its proof since it follows using analogous arguments as those considered in Cui et al. (2003) and tedious calculations.

Lemma A.4. *Under conditions of Theorem 3.2, we have that*

$$\begin{aligned}&\left\{\mathbf{P}_{m+1}\ddot{\rho}(\boldsymbol{\beta}_m) - \boldsymbol{\beta}_m^{\mathrm{T}}\dot{\rho}(\boldsymbol{\beta}_m)\mathbf{I}_p - \boldsymbol{\beta}_m\dot{\rho}(\boldsymbol{\beta}_m)^{\mathrm{T}}\right\}(\widehat{\boldsymbol{\beta}}_m - \boldsymbol{\beta}_m) + \mathbf{o}_p(\|\widehat{\boldsymbol{\beta}}_m - \boldsymbol{\beta}_m\|)\\ &\quad = \sum_{j=1}^{m-1}\left\{\boldsymbol{\beta}_j^{\mathrm{T}}\dot{\rho}(\boldsymbol{\beta}_m)\mathbf{I}_p + \boldsymbol{\beta}_j\dot{\rho}(\boldsymbol{\beta}_m)^{\mathrm{T}}\right\}(\widehat{\boldsymbol{\beta}}_j - \boldsymbol{\beta}_j) + N^{-1/2}\mathbf{P}_{m+1}\mathbf{u}_m + \mathbf{o}_p\left(\sum_{i=1}^{m-1}\|\widehat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_i\|\right) + \mathbf{o}_p(N^{-1/2}).\end{aligned}$$

From Lemma A.4, we easily obtain the following result.

Lemma A.5. *Under the conditions of Theorem 3.2, we have* $\mathbf{A}_m(\widehat{\boldsymbol{\beta}}_m - \boldsymbol{\beta}_m) + \mathbf{o}_p(\|\widehat{\boldsymbol{\beta}}_m - \boldsymbol{\beta}_m\|) = \sum_{j=1}^{m-1}\mathbf{B}_{jm}(\widehat{\boldsymbol{\beta}}_j - \boldsymbol{\beta}_j) + N^{-1/2}\mathbf{P}_{m+1}\mathbf{u}_m + \mathbf{o}_p(\sum_{j=1}^{m-1}\|\widehat{\boldsymbol{\beta}}_j - \boldsymbol{\beta}_j\|) + \mathbf{o}_p(N^{-1/2})$*, where* $\mathbf{A}_m = \mathbf{P}_{m+1}\ddot{\rho}(\boldsymbol{\beta}_m) - \boldsymbol{\beta}_m^{\mathrm{T}}\dot{\rho}(\boldsymbol{\beta}_m)\mathbf{I}_p - \sum_{i=1}^{m-1}\boldsymbol{\beta}_i^{\mathrm{T}}\dot{\rho}(\boldsymbol{\beta}_m)\boldsymbol{\beta}_m\boldsymbol{\beta}_i^{\mathrm{T}}$ *and* $\mathbf{B}_{jm} = \boldsymbol{\beta}_j^{\mathrm{T}}\dot{\rho}(\boldsymbol{\beta}_m)\mathbf{I}_p + \boldsymbol{\beta}_j\dot{\rho}(\boldsymbol{\beta}_m)^{\mathrm{T}}$.

Lemma A.6. *Under the conditions of Theorem 3.2. we have* $\widehat{\boldsymbol{\beta}}_m - \boldsymbol{\beta}_m = \mathbf{O}_p(N^{-1/2})$.

Proof. We begin by proving that the result holds for $m = 1$. Using Lemma A.5 with $m = 1$, we get easily that $(\widehat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1) = \mathbf{o}_p(\|\widehat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1\|) + \mathbf{A}_1^{-1}N^{-1/2}\mathbf{P}_2\mathbf{u}_1 + \mathbf{o}_p(N^{-1/2})$. Since $\mathbf{P}_2\mathbf{u}_1$ converges in distribution to a normal random variable, we have that $\mathbf{A}_1^{-1}N^{-1/2}\mathbf{P}_2\mathbf{u}_1 = \mathbf{O}_p(N^{-1/2})$, and so, $N^{1/2}(\widehat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1) = N^{1/2}\mathbf{o}_p(\|\widehat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1\|) + \mathbf{O}_p(1) + \mathbf{o}_p(1)$. Thus,

$$N^{1/2}(\widehat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1) = N^{1/2}\mathbf{o}_p(\|\widehat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1\|) + \mathbf{O}_p(1). \tag{A.3}$$

Hence, $N^{1/2}\|\widehat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1\|\left[(\widehat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1)/\|\widehat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1\| - \mathbf{o}_p(\|\widehat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1\|)/\|\widehat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1\|\right] = \mathbf{O}_p(1)$ which implies that $N^{1/2}\|\widehat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1\| = O_p(1)$, since the norm of the term between brackets converges in probability to 1.

Let us show if $N^{1/2}\|\widehat{\boldsymbol{\beta}}_j - \boldsymbol{\beta}_j\| = O_p(1)$, for all $2 \le j \le m-1$, then $N^{1/2}\|\widehat{\boldsymbol{\beta}}_m - \boldsymbol{\beta}_m\| = O_p(1)$. Indeed, from Lemma A.5, we get $N^{1/2}(\widehat{\boldsymbol{\beta}}_m - \boldsymbol{\beta}_m) = N^{1/2}\mathbf{o}_p(\|\widehat{\boldsymbol{\beta}}_m - \boldsymbol{\beta}_m\|) + \sum_{j=1}^{m-1}\mathbf{A}_m^{-1}\mathbf{B}_{jm}N^{1/2}(\widehat{\boldsymbol{\beta}}_j - \boldsymbol{\beta}_j) + \mathbf{A}_m^{-1}\mathbf{P}_{m+1}\mathbf{u}_m + N^{1/2}\mathbf{o}_p(\sum_{j=1}^{m-1}\|\widehat{\boldsymbol{\beta}}_j - \boldsymbol{\beta}_j\|) + \mathbf{o}_p(1)$. Since $\mathbf{P}_{m+1}\mathbf{u}_m$ converges in distribution and using the inductive assumption, we get $N^{1/2}(\widehat{\boldsymbol{\beta}}_m - \boldsymbol{\beta}_m) = N^{1/2}\mathbf{o}_p(\|\widehat{\boldsymbol{\beta}}_m - \boldsymbol{\beta}_m\|) + \mathbf{O}_p(1)$, which is analogous to (A.3), and so, we get easily that $\widehat{\boldsymbol{\beta}}_m - \boldsymbol{\beta}_m = \mathbf{O}_p(N^{-1/2})$ concluding the proof. □

The Bahadur representation for $\widehat{\boldsymbol{\beta}}_m$ given in (3.1) follows now easily using Lemmas A.5 and A.6.

Let us obtain the expansion given in (3.2). Similar arguments to those considered in Lemma A.3, allow us to show that $n_i^{-1}\sum_{j=1}^{n_i} h_{i,\sigma}(\mathbf{x}_{ij}, \widehat{\boldsymbol{\beta}}_m) = n_i^{-1}\sum_{j=1}^{n_i} h_{i,\sigma}(\mathbf{x}_{ij}, \boldsymbol{\beta}_m) + o_p(n_i^{-1/2})$, which implies that $\widehat{\lambda}_{im} - \lambda_{im} = \varsigma_{i,n_i}(\widehat{\boldsymbol{\beta}}_m) - \varsigma_i(\widehat{\boldsymbol{\beta}}_m) + \varsigma_i(\widehat{\boldsymbol{\beta}}_m) - \varsigma_i(\boldsymbol{\beta}_m) = n_i^{-1}\sum_{j=1}^{n_i} h_{i,\sigma}(\mathbf{x}_{ij}, \boldsymbol{\beta}_m) + o_p(n_i^{-1/2}) + \dot{\varsigma}_i(\boldsymbol{\beta}_m)^{\mathrm{T}}(\widehat{\boldsymbol{\beta}}_m - \boldsymbol{\beta}_m) + o_p(\|\widehat{\boldsymbol{\beta}}_m - \boldsymbol{\beta}_m\|) = n_i^{-1}\sum_{j=1}^{n_i} h_{i,\sigma}(\mathbf{x}_{ij}, \boldsymbol{\beta}_m) + o_p(n_i^{-1/2})$ concluding the proof. □

References

Boente, G., Molina, J., Sued, M., 2009. On the asymptotic behavior of general projection-pursuit estimators under the common principal components model. http://www.ic.fcen.uba.ar/preprints/Boente_Molina_Sued.pdf.

Boente, G., Orellana, L., 2001. A robust approach to common principal components. In: Fernholz, L.T., Morgenthaler, S., Stahel, W. (Eds.), Statistics in Genetics and in the Environmental Sciences. Birkhauser, Basel, pp. 117–147.

Boente, G., Pires, A.M., Rodrigues, I.M., 2002. Influence functions and outlier detection under the common principal components model: A robust approach. Biometrika 89, 861–875.

Boente, G., Pires, A.M., Rodrigues, I.M., 2006. General projection-pursuit estimators for the common principal components model: Influence funtions and Monte Carlo study. Journal of Multivariate Analysis 97, 124–147.

Croux, C., Ruiz-Gazen, A., 2005. High breakdown estimators for principal components: The projection-pursuit approach revisited. Journal of Multivariate Analysis 95, 206–226.

Cui, H., He, X., Ng, K.W., 2003. Asymptotic distribution of principal components based on robust dispersions. Biometrika 90, 953–966.

Flury, B.K., 1984. Common principal components in *k* groups. Journal of the American Statistical Association 79, 892–898.

Flury, B.K., 1988. Common Principal Components and Related Multivariate Models. John Wiley, New York.

Li, G., Chen, Z., 1985. Projection–pursuit approach to robust dispersion matrices and principal components: Primary theory and Monte Carlo. Journal of the American Statistical Association 80, 759–766.

Pires, A.M., Branco, J., 2002. Partial influence functions. Journal of Multivariate Analysis 83, 458–468.

Rodrigues, I.M., 2003. Métodos Robustos em Análise de Componentes Principais Comuns. Universidade Técnica de Lisboa. Unpublished Ph.D. Thesis (in Portuguese). Available on: http://www.math.ist.utl.pt/~apires/phd.html.

van der Vaart, A., Wellner, J., 1996. Weak Convergence and Empirical Processes. With Applications to Statistics. Springer, New York.