UNIVERSIDAD DE BUENOS AIRES
Facultad de Ciencias Exactas y Naturales
Departamento de Matemática

Tesis de Licenciatura

Sobre el comportamiento asintótico de los estimadores de ***projection–pursuit*** para el modelo de componentes principales comunes

Julieta Molina

Directoras de tesis: Dra. Graciela L. Boente Boente y Dra. Mariela Sued.

Lugar de trabajo: Instituto de Cálculo, FCEyN, UBA.

31 de Marzo de 2008

Índice general

Capítulo 1

Introducción

En muchas situaciones que involucran varias poblaciones, como análisis discriminante o análisis de la varianza multivariado, se supone igualdad de las matrices de covarianza de las distintas poblaciones. Si este supuesto no se verifica, es usual estimar las matrices de covarianza de cada grupo en forma separada lo cual suele llevar a un número excesivo de parámetros. Como menciona Flury (1987), en muchas aplicaciones, este supuesto no es satisfactorio ya que las matrices de escala de las k poblaciones pueden exhibir algún tipo de estructura común. Por otra parte, el "principio de parsimonia" sugiere que se introduzcan parámetros solamente cuando sea necesario. Varios autores han estudiado distintas formas de modelar matrices de covarianza simultaneamente bajo restricciones. Flury(1988) provee una revisión de distintos modelos para matrices de escala de varias poblaciones independientes.

Una generalización de la igualdad de matrices de escala consiste en suponer que son proporcionales

$$\mathbf{\Sigma}_i = \rho_i \mathbf{\Sigma}_1 \ , \ \text{para } 1 \leq i \leq k \text{ y } \rho_1 = 1 \ ,$$

donde $\mathbf{\Sigma}_i$ es la matriz de covarianza de la población i-ésima. El *modelo de componentes principales comunes* (CPC) generaliza este supuesto suponiendo que las matrices son conmutables, es decir, que tienen distintos autovalores pero idénticos autovectores, o sea,

$$\mathbf{\Sigma}_i = \boldsymbol{\beta} \mathbf{\Lambda}_i \boldsymbol{\beta}' \ , \quad 1 \leq i \leq k \, ,$$

donde $\mathbf{\Lambda}_i$ son matrices diagonales y $\boldsymbol{\beta}$ es una matriz ortogonal de autovectores comunes. Este modelo puede verse como una generalización del modelo de componentes principales a k grupos, ya que la transformación principal es la misma en todas las poblaciones consideradas mientras las varianzas asociadas con las direcciones principales varían entre grupos. Como en el análisis de componentes principales, el modelo CPC puede usarse para reducir la dimensión de los datos, conservando en la medida de lo posible la variabilidad presente en cada población. Aunque los ejes principales son los mismos para todas las poblaciones, el porcentaje de variabilidad explicada por cada uno de ellos varía entre poblaciones.

El modelo CPC fue introducido por Flury (1984), quien describió como obtener los estimadores de máxima verosimilitud de $\boldsymbol{\beta}$ y $\mathbf{\Lambda}_i$ y como testear la hipótesis del modelo CPC suponiendo que las variables son conjuntamente normales. La distribución asintótica del estimador de máxima verosimilitud fue estudiada por Flury (1986). En Flury (1988) se analizan y discuten varios ejemplos, incluyendo conjuntos de datos biológicos. En muchas aplicaciones biométricas, las componentes principales se interpretan como factores independientes que determinan el crecimiento, tamaño o forma de un organismo. Es natural entonces considerar un modelo en el cual los mismos factores aparezcan en especies distintas pero relacionadas. El modelo de componentes principales comunes claramente sirve para tal propósito. Aplicaciones de este modelo y otros modelos jerárquicos pueden verse en Flury (1988), Klingerberg & Neuenschwander (1996), Arnold & Phillips (1999) y Phillips & Arnold (1999).

Sin embargo, como la mayoría de los procedimientos óptimos bajo normalidad, estos estimadores no son robustos debido a la falta de robustez de la matriz de covarianza muestral. Es decir, pierden sus propiedades de optimalidad y se comportan inadecuadamente bajo pequeñas desviaciones del supuesto de normalidad, tales como la presencia de observaciones atípicas o altamente influyentes que pueden aparecer ya sea por contaminación o como resultado de un modelo de distribución con colas pesadas. Existen distintas alternativas robustas a la media y covarianza muestral. Entre otros, los $M-$estimadores (Maronna, 1976 y Huber, 1977) son una generalización de los estimadores de máxima verosimilitud y poseen una buena eficiencia para un amplio rango de modelos poblaciones. Sin embargo, tienen bajo punto de ruptura en alta dimensión y por esa razón se introdujeron estimadores afín equivariantes de alto punto de ruptura (ver, Stahel, 1981, Donoho, 1982, Rousseeuw, 1985, Davies, 1987 and Kent & Tyler, 1996). Una discusión detallada de estimadores robustos de posición y escala puede verse en Maronna, Martin & Yohai (2006).

Como en otros contextos, en muchas situaciones es deseable obtener estimadores robustos de los autovectores comunes y sus autovalores bajo un modelo CPC. Novi Inverardi & Flury (1992) propusieron estimar separadamente en forma robusta las matrices de escala de las k poblaciones y luego reemplazar estos estimadores en las ecuaciones que definen el estimador de máxima verosimilitud. Estos autores sugieren utilizar los $M-$estimadores introducidos por Maronna (1976). Un enfoque relacionado fue dado por Boente & Orellana (2001) quienes extendieron al caso de varias poblaciones el enfoque *plug–in* (**PI**) dado para componentes principales y estudiado por Croux & Haesbroeck (2000). El procedimiento *plug–in* consiste en resolver un sistema de ecuaciones semejante al obtenido para los estimadores de máxima verosimilitud con datos gaussianos reemplazando la matriz de covarianza muestral por estimadores de escala robustos que se calibran de modo a resultar asintóticamente insesgados bajo el modelo central. Boente & Orellana (2001) obtuvieron el comportamiento asintótico de los estimadores robustos *plug–in* de las direcciones comunes y sus autovalores cuando las matrices de escala robustas son asintóticamente normales y esféricamente invariantes.

Otra posibilidad para obtener estimadores robustos en el caso de una población es el método de *projection–pursuit* (**PP**) introducido por Li & Chen (1985) quienes obtienen estimadores de las direcciones principales mediante la maximización o minimización de una escala robusta de las proyecciones univariadas de las observaciones. Croux & Ruiz–Gazen (2005) estudiaron la función de influencia del funcional asociado a este procedimiento mientras que Cui, He & Ng (2003) obtuvieron un desarrollo de Bahadur de los estimadores. Boente & Orellana (2001) generalizaron al modelo CPC los estimadores de *projection–pursuit* definiendo un índice de escala adecuado. La ventaja de esta alternativa es que pueden calcularse para conjuntos de datos con más variables que observaciones. La distribución asintótica de esta clase de estimadores es el objeto de esta tesis. Las funciones de influencia parcial de ambas clases de estimadores **PI** y **PP** fueron estudiadas por Boente, Pires & Rodrigues (2002).

Los dos procedimientos mencionados producen estimadores robustos, es decir, estimadores que dan resultados confiables aunque las observaciones tengan una proporción de datos atípicos. Sin embargo, los estimadores **PI** mostraron su sensibilidad en algunas situaciones cuando se estiman las direcciones comunes aún cuando se usen estimadores de la matriz de escala de alto punto de ruptura (ver, Boente & Orellana, 2001). Por esta razón, los métodos de *projection–pursuit* se introdujeron a fin de proveer estimaciones más confiables. El problema de obtener desarrollos de Bahadur para los estimadores **PP** de las componentes principales comunes es de particular interés si uno desea extender los test robustos definidos por Boente, Pires & Rodrigues (2007) al caso de estimadores de *projection–pursuit*.

Esta tesis está organizada como sigue. En el Capítulo 2, se introducen los principales conceptos relacionados con el análisis de componentes principales para una población. En la sección 2.4 de dicho capítulo, se presenta el enfoque de *projection–pursuit* para el problema de componentes principales y se recuerdan los principales resultados obtenidos por Li & Chen (1985) y por Cui, He & Ng (2003). En el Capítulo 3 se presenta el modelo de componentes principales comunes y se describen los estimadores de máxima verosimilitud para los parámetros de dicho problema, así como el enfoque **PI** para obtener estimadores robustos y los resultados de distribución asintótica obtenidos por Boente & Orellana (2001). Finalmente, en los Capítulos 4 y 5 se presenta la contribución de esta tesis, es decir, la teoría asintótica para los estimadores de *projection–pursuit* bajo el modelo CPC y el estudio de los casos elíptico y unipoblacional respectivamante.

Capítulo 2

Componentes principales para una población

En este capítulo, vamos a definir las componentes principales de un vector aleatorio $\mathbf{x} \in I\!R^p$, para poder introducir los estimadores de *projection–pursuit* en forma natural. Dicho enfoque trata de responder a la siguiente pregunta ¿Cómo transformar $\mathbf{x}$ linealmente en un nuevo vector aleatorio $\mathbf{y}$ cuyas componentes sean no correlacionadas y de forma tal que dicha transformación sea óptima en algún sentido?

2.1. Caso poblacional

Definición 2.1.1. Sea $\mathbf{x} \in I\!R^p$ un vector aleatorio con posición $\boldsymbol{\mu}$ y matriz de covarianza $\boldsymbol{\Sigma}$. Dado que esta última es simétrica y definida positiva, tenemos su descomposición espectral, es decir, existe una matriz ortogonal $\boldsymbol{\beta}$ y una matriz diagonal $\boldsymbol{\Lambda}$ tal que $\boldsymbol{\Sigma} = \boldsymbol{\beta}\boldsymbol{\Lambda}\boldsymbol{\beta}^{\mathrm{T}}$, con $\boldsymbol{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_p)$ una matriz diagonal formada por los autovalores de $\boldsymbol{\Sigma}$ ordenados de mayor a menor, $\lambda_1 \geq \ldots \geq \lambda_p$, y $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_p)$ una matriz ortogonal con $\boldsymbol{\beta}_i$ autovector de $\boldsymbol{\Sigma}$ asociado al autovalor λ_i.

Dado $1 \leq j \leq p$, definimos la $j-$ésima componente principal del vector $\mathbf{x}$ como $y_j = \boldsymbol{\beta}_j^{\mathrm{T}}(\mathbf{x} - \boldsymbol{\mu})$. Además, llamamos vector de componentes principales de $\mathbf{x}$ a $\mathbf{y} = \boldsymbol{\beta}^{\mathrm{T}}(\mathbf{x} - \boldsymbol{\mu})$.

Observación: En la literatura, los autovectores $(\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_p)$ de la matriz $\boldsymbol{\Sigma}$ son conocidos como direcciones ó ejes principales.

Como podemos ver, y_j presenta las siguientes características:

i) Es la $j-$ésima coordenada del vector $\mathbf{x} - \boldsymbol{\mu}$ expresado en la base $(\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_p)$.

ii) Es una combinación lineal de las coordenadas del vector $\mathbf{x} - \boldsymbol{\mu}$.

La siguiente proposición nos explica, entre otras cosas, por qué esta combinación lineal de las coordenadas de $\mathbf{x} - \boldsymbol{\mu}$ se destaca por sobre otras y nos permite responder a la pregunta planteada al inicio de este capítulo.

Teorema 2.1.1. *a) Las variables $y_1, \ldots, y_p$ son no correlacionadas y $var(y_j) = \lambda_j$.*

b) $\underset{\|\mathbf{a}\|=1}{\mathrm{argmax}}\, var(\mathbf{a}^{\mathrm{T}}\mathbf{x}) = \boldsymbol{\beta}_1$ *y además,* $\max_{\|\mathbf{a}\|=1} var(\mathbf{a}^{\mathrm{T}}\mathbf{x}) = \lambda_1$.

c) La función var$(\mathbf{a}^{\mathrm{T}}\mathbf{x})$ *restringida al conjunto*

$$\mathbf{C}_j = \{\mathbf{a} \in I\!R^p / \ \|\mathbf{a}\| = 1, \ \mathbf{a}^{\mathrm{T}}\boldsymbol{\beta}_i = 0 \ \forall i = 1, 2, \ldots, j-1\}$$

toma su máximo valor cuando $\mathbf{a} = \boldsymbol{\beta}_j$*, o sea,* $\operatorname*{argmax}_{\mathbf{a}\in\mathbf{C}_j} var(\mathbf{a}^{\mathrm{T}}\mathbf{x}) = \boldsymbol{\beta}_j$ *y* $\text{máx}_{\mathbf{a}\in\mathbf{C}_j} var(\mathbf{a}^{\mathrm{T}}\mathbf{x}) = \lambda_j$*.*

Demostración: a) La matriz de covarianza $\boldsymbol{\Sigma}_{\mathbf{y}}$ del vector $\mathbf{y} = \boldsymbol{\beta}^{\mathrm{T}}(\mathbf{x}-\boldsymbol{\mu})$, está dada por $\boldsymbol{\Sigma}_{\mathbf{y}} = \boldsymbol{\Sigma}_{\boldsymbol{\beta}^{\mathrm{T}}(\mathbf{x}-\boldsymbol{\mu})} = \boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{x}}\boldsymbol{\beta} = \boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{\Sigma}\boldsymbol{\beta} = \boldsymbol{\Lambda}$, de donde obtenemos que $\mathrm{cov}(y_i, y_j) = 0$ cuando $i \neq j$ y $\mathrm{var}(y_j) = \lambda_j$.

b) Como $(\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_p)$ es una base ortonormal tenemos $\mathbf{a} = \sum_{i=1}^p (\mathbf{a}^{\mathrm{T}}\boldsymbol{\beta}_i)\boldsymbol{\beta}_i$, con lo cual $\|\mathbf{a}\| = 1$ es equivalente a $\sum_{i=1}^p (\mathbf{a}^{\mathrm{T}}\boldsymbol{\beta}_i)^2 = 1$. Por lo tanto, usando que $\lambda_i \leq \lambda_1$ para $1 \leq i \leq p$, obtenemos $\mathrm{var}(\mathbf{a}^{\mathrm{T}}\mathbf{x}) = \mathbf{a}^{\mathrm{T}}\boldsymbol{\Sigma}\mathbf{a} = \sum_{i=1}^p (\mathbf{a}^{\mathrm{T}}\boldsymbol{\beta}_i)^2\lambda_i \leq \lambda_1 \sum_{i=1}^p (\mathbf{a}^{\mathrm{T}}\boldsymbol{\beta}_i)^2 = \lambda_1 = \mathrm{var}(\boldsymbol{\beta}_1^{\mathrm{T}}\mathbf{x})$.

c) Como en b) tenemos $\mathbf{a} = \sum_{i=1}^p (\mathbf{a}^{\mathrm{T}}\boldsymbol{\beta}_i)\boldsymbol{\beta}_i$, pero usando que $\mathbf{a} \in \mathbf{C}_j$, obtenemos $\mathbf{a} = \sum_{i=j}^p (\mathbf{a}^{\mathrm{T}}\boldsymbol{\beta}_i)\boldsymbol{\beta}_i$ con $\sum_{i=j}^p (\mathbf{a}^{\mathrm{T}}\boldsymbol{\beta}_i)^2 = 1$. Por lo tanto, usando que $\lambda_i \leq \lambda_j$ para $j \leq i \leq p$, $\mathrm{var}(\mathbf{a}^{\mathrm{T}}\mathbf{x}) = \mathbf{a}^{\mathrm{T}}\boldsymbol{\Sigma}\mathbf{a} = \sum_{i=j}^p (\mathbf{a}^{\mathrm{T}}\boldsymbol{\beta}_i)^2\lambda_i \leq \lambda_j \sum_{i=j}^p (\mathbf{a}^{\mathrm{T}}\boldsymbol{\beta}_i)^2 = \lambda_j = \mathrm{var}(\boldsymbol{\beta}_j^{\mathrm{T}}\mathbf{x})$. □

La importancia de este resultado es que permite definir iterativamente las componentes principales como aquellas direcciones que maximizan la variabilidad. Esta noción será la que se utilizará para definir los estimadores de *projection–pursuit* reemplazando la variabilidad medida a través de la varianza por una medida de variabilidad o escala univariada robusta, como veremos en 2.4.

2.2. Porcentajes de variabilidad

Si $(y_1, \ldots, y_p)$ son las componentes principales del vector aleatorio $\mathbf{x}$, tenemos que

$$\sum_{j=1}^p \mathrm{var}(y_j) = \sum_{i=1}^p \lambda_j = \mathrm{tr}(\boldsymbol{\Lambda}) = \mathrm{tr}(\boldsymbol{\beta}\boldsymbol{\Lambda}\boldsymbol{\beta}^{\mathrm{T}}) = \mathrm{tr}(\boldsymbol{\Sigma}) = \sum_{j=1}^p \mathrm{var}(x_j) \ .$$

Es decir, la suma de las varianzas de las variables originales (las coordenadas del vector $\mathbf{x}$) y la suma de las varianzas de las componentes principales coinciden. Esto permite hablar del porcentaje π_j de varianza total explicado por la $j-$ésima componente principal,

$$\pi_j = \frac{\lambda_j}{\sum_{i=1}^p \lambda_i} = \frac{\lambda_j}{\sum_{i=1}^p \mathrm{var}(x_j)} \ .$$

Luego, el porcentaje de variabilidad explicado por las primeras m componentes principales estará dado por

$$\sum_{i=1}^m \pi_j = \frac{\sum_i^m \lambda_j}{\sum_{i=1}^p \mathrm{var}(x_j)} \ .$$

El objetivo del análisis de componentes principales es por lo tanto, proveer una reducción de la dimensión que conserve un porcentaje alto de la variabilidad. Es decir, partiendo de un vector aleatorio $\mathbf{x}$ de dimensión p, con p un número grande, nos interesa obtener una transformación lineal del vector original $\mathbf{A} \in I\!R^{m\times p}$ tal que el vector $\mathbf{A}\mathbf{x}$ recoja un porcentaje amplio de la variabilidad total $\sum_{i=1}^p \mathrm{var}(x_j)$.

2.3. Estimadores clásicos de las componentes principales

Cuando la posición $\boldsymbol{\mu}$ y la matriz de covarianza $\boldsymbol{\Sigma}$ del vector aleatorio $\mathbf{x}$ son desconocidas, pueden ser estimadas a partir de una muestra $\mathbf{x}_1, \ldots, \mathbf{x}_n$ de observaciones independientes e idénticamente distribuidas. La teoría clásica considera como estimadores, los estimadores de máxima verosimilitud bajo normalidad, o sea,

$$\widehat{\boldsymbol{\mu}} = \overline{\mathbf{x}} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{x}_i \ , \quad \widehat{\boldsymbol{\Sigma}} = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{x}_i - \overline{\mathbf{x}})(\mathbf{x}_i - \overline{\mathbf{x}})^{\mathrm{T}} = \frac{n-1}{n}\mathbf{S} \ .$$

Luego, para estimar las direcciones principales del vector $\mathbf{x}$, es razonable considerar la matriz ortogonal $\widehat{\boldsymbol{\beta}} = (\widehat{\boldsymbol{\beta}}_1, \ldots, \widehat{\boldsymbol{\beta}}_p)$ cuyas columnas son los autovectores de $\widehat{\boldsymbol{\Sigma}}$ (o equivalentemente de $\mathbf{S}$) asociados a los autovalores $\widehat{\lambda}_1 \geq \ldots \geq \widehat{\lambda}_p$ de $\widehat{\boldsymbol{\Sigma}}$, ordenados de mayor a menor. En Kshirsagar (1972) se muestra que estos estimadores son los estimadores de máxima verosimilitud de $\boldsymbol{\beta}$ y $\lambda_1, \ldots, \lambda_p$. Por otra parte, la distribución asintótica de estos estimadores fue obtenida por Anderson (1963) y puede verse en Muirhead (1982).

Para cada observación $\mathbf{x}_i$ se define entonces el vector de componentes principales muestrales como $\widehat{\mathbf{y}}_i = \widehat{\boldsymbol{\beta}}^{\mathrm{T}}(\mathbf{x}_i - \overline{\mathbf{x}})$.

Utilizando la distribución empírica, se deduce inmediatamente el siguiente resultado que muestra que $\widehat{\boldsymbol{\beta}}_1, \ldots, \widehat{\boldsymbol{\beta}}_p$ son las direcciones principales correspondientes a la distribución empírica F_n dada por la muestra $\mathbf{x}_1 \ldots, \mathbf{x}_n$.

Teorema 2.3.1. *Dada una muestra aleatoria $\mathbf{x}_1, \ldots, \mathbf{x}_n$, sea $\mathbf{v}$ un vector aleatorio que toma los valores $\mathbf{x}_i$ con probabilidad $1/n$. Tenemos entonces que las direcciones principales de $\mathbf{v}$ están dadas por $\widehat{\boldsymbol{\beta}}_1, \ldots, \widehat{\boldsymbol{\beta}}_p$ mientras que el vector de componentes principales de $\mathbf{v}$ asigna masa $1/n$ a cada vector de componentes principales muestrales $\widehat{\mathbf{y}}_1, \ldots, \widehat{\mathbf{y}}_n$.*

Del Teorema 2.1.1, se deduce inmediatamente que $\widehat{\boldsymbol{\beta}}_1, \ldots, \widehat{\boldsymbol{\beta}}_p$ maximizan la variabilidad empírica de las proyecciones de las observaciones, es decir,

$$\widehat{\boldsymbol{\beta}}_1 = \underset{\|\mathbf{a}\|=1}{\operatorname{argmax}}\, s_n^2(\mathbf{a}^{\mathrm{T}}\mathbf{X}) \tag{2.3.1}$$

$$\widehat{\boldsymbol{\beta}}_j = \underset{\mathbf{a}\in\mathcal{C}_j}{\operatorname{argmax}}\, s_n^2(\mathbf{a}^{\mathrm{T}}\mathbf{X}) \tag{2.3.2}$$

donde $\mathcal{C}_j = \{\mathbf{a} \in I\!R^p, \|\mathbf{a}\| = 1, \mathbf{a}^{\mathrm{T}}\widehat{\boldsymbol{\beta}}_i = 0\ ,\ 1 \leq i \leq j-1\}$, $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_n)$ y $s_n^2(\mathbf{Y}) = \sum_{i=1}^{n}(y_i - \overline{y})^2/(n-1)$ con $\mathbf{Y} = (y_1, \ldots, y_n)$ y $y_j \in I\!R$.

2.4. Estimadores robustos de las componentes principales: Método de *projection–pursuit*.

En la sección anterior hemos definido los estimadores de las direcciones principales del vector $\mathbf{x}$ como los autovectores $\widehat{\boldsymbol{\beta}}_1, \ldots, \widehat{\boldsymbol{\beta}}_p$ de la matriz de covarianza muestral $\mathbf{S}$. Como hemos mencionado en la Introducción, la matriz de covarianza muestral $\mathbf{S}$ es sensible a la presencia de datos atípicos en la muestra y, en consecuencia, los mismo ocurre con los estimadores de las direcciones principales. Varias han sido las propuestas hechas para salvar esta dificultad, pero todas se basan en dos principios, el principio *plug–in* o el de *projection–pursuit*. El primer enfoque consiste en utilizar estimadores robustos de la matriz de escala y definir las componentes principales como los autovectores asociados a dicha matriz.

Algunos resultados en esta dirección pueden verse en Devlin, Gnanadesikan & Kettenring (1981), Tyler (1983), Boente (1987) y en Croux & Haesbroeck (2000), entre otros. Un enfoque relacionado fue dado por Locantore *et al.* (1999) quienes proyectaron los datos centrados en la esfera unidad y luego definieron los ejes principales a través de una análisis de componentes principales sobre los datos proyectados. Para una descripción de dichos procedimientos puede verse Maronna, Martin & Yohai (2006).

El segundo enfoque, que es el que abordaremos en esta sección, se basa en proyecciones univariadas y en las propiedades obtenidas en el Teorema 2.1.1. Las técnicas de *projection–pursuit* son una herramienta útil en análisis multivariado, especialmente cuando se trabaja con más variables que observaciones. Estos métodos procuran disminuir la dimensión del vector original, proyectándolo sobre un subespacio de menor dimensión, de forma tal que se preserve de la mejor manera posible la riqueza de la estructura original. Diferentes formas de medir dicha riqueza dan origen a distintos criterios para reducir dimensión. Para el problema de componentes principales, el método de *projection–pursuit* (**PP**) define los ejes principales maximizando (o minimizando) una escala robusta de las observaciones proyectadas. Efectivamente, observemos que (2.3.1) y (2.3.2) sugieren que uno de los problemas del procedimiento clásicamente usado es que estamos maximizando el desvío estándar muestral de las observaciones proyectadas, siendo esta medida de escala muy sensible a observaciones atípicas. Li & Chen (1985) (ver también Huber, 1981) sugirieron reemplazar el desvío estándar por un estimador de escala robusto es decir, definir las direcciones principales como

$$\widehat{\boldsymbol{\beta}}_1 = \underset{\|\mathbf{a}\|=1}{\operatorname{argmax}}\, \mathbf{s}_n^2(\mathbf{a}^{\mathrm{T}}\mathbf{X}) \tag{2.4.1}$$

$$\widehat{\boldsymbol{\beta}}_j = \underset{\mathbf{a}\in\mathcal{C}_j}{\operatorname{argmax}}\, \mathbf{s}_n^2(\mathbf{a}^{\mathrm{T}}\mathbf{X}) \tag{2.4.2}$$

donde $\mathcal{C}_j = \{\mathbf{a} \in I\!R^p, \|\mathbf{a}\| = 1, \mathbf{a}^{\mathrm{T}}\widehat{\boldsymbol{\beta}}_i = 0\ ,\ 1 \le i \le j-1\}$, $\mathbf{a}^{\mathrm{T}}\mathbf{X} = (\mathbf{a}^{\mathrm{T}}\mathbf{x}_1, \dots, \mathbf{a}^{\mathrm{T}}\mathbf{x}_n)$ y $\mathbf{s}_n$ es un estimador de escala univariado robusto, como por ejemplo, un $M-$estimador de escala (ver Maronna, Martin & Yohai, 2006) o en particular, la mediana de los desvíos absolutos respecto de la mediana (MAD), $\mathbf{s}_n(\mathbf{Y}) = \operatorname{median}_{1\le i\le n} |y_i - \operatorname{median}_{1\le \ell\le n} y_\ell|$ con $\mathbf{Y} = (y_1, \dots, y_n)$. Los estimadores de los autovalores asociados se definen luego como $\widehat{\lambda}_j = \mathbf{s}_n^2(\widehat{\boldsymbol{\beta}}_j^{\mathrm{T}}\mathbf{X})$ y la matriz robusta de dispersión se define a partir de la noción de descomposición espectral como $\mathbf{V} = \sum_{j=1}^p \widehat{\lambda}_j\ \widehat{\boldsymbol{\beta}}_j\widehat{\boldsymbol{\beta}}_j^{\mathrm{T}}$. Este método fue aplicado por Xie *et al.* (1993) a un problema chemométrico mientras que Croux & Ruiz–Gazen (1996) dieron un algoritmo para su cálculo que fue discutido y mejorado por Croux, Filzmoser & Oliveira (2007). En esta sección, describiremos los principales resultados dados por Li & Chen (1985) y posteriormente, las propiedades asintóticas de los estimadores **PP** estudiadas por Cui, He & Ng (2003). Cabe mencionar que Patak (1991) y Berrendero (1996) estudiaron el sesgo asintótico en entornos de contaminación de las direcciones principales y de su tamaño cuando se utiliza una escala robusta. Por otra parte, Croux & Ruiz–Gazen (2000) obtuvieron una expresión para la función de influencia del funcional asociado a los estimadores de los autovectores, autovalores y la matriz de dispersión asociada.

Empezaremos introduciendo en la sección 2.4.1 el funcional asociado al procedimiento descripto y estudiando algunas de sus propiedades. En la sección 2.4.2 se presentarán los resultados dados por Cui, He & Ng (2003) quienes obtuvieron un desarrollo de Bahadur para los estimadores **PP** de las direcciones principales y su tamaño que permite fácilmente deducir la distribución asintótica de los mismos.

2.4.1. Funcional asociado

Consideremos $\mathbf{x} \in I\!R^p$ un vector aleatorio con distribución F y sin perdida de generalidad, supondremos de ahora en más que su parámetro de posición es $\mathbf{0}$. Denotemos por $F[\mathbf{a}]$ a la distribución de la variable aleatoria unidimensional $y = \mathbf{a}^{\mathrm{T}}\mathbf{x}$, donde $\mathbf{a} \in I\!R^p$. Como hemos visto en el Teorema 2.1.1, la primer componente principal asociada al vector $\mathbf{x}$ está dada por $y_1 = \boldsymbol{\beta}_1^{\mathrm{T}}\mathbf{x}$ donde $\boldsymbol{\beta}_1 \in I\!R^p$ verifica:

$$\boldsymbol{\beta}_1 = \underset{\|\mathbf{a}\|=1}{\operatorname{argmax}} \, \sigma^2(F[\mathbf{a}]) = \underset{\|\mathbf{a}\|=1}{\operatorname{argmax}} \, (\mathbf{a}^{\mathrm{T}}\boldsymbol{\Sigma}\mathbf{a}) \ ,$$

siendo $\boldsymbol{\Sigma}$ la matriz de covarianzas del vector $\mathbf{x}$ y $\sigma(G)$ el desvío standard de la distribución unidimensional G. Por otra parte, el tamaño de dicha dirección está dado por $\sigma^2(F[\boldsymbol{\beta}_1]) = \boldsymbol{\beta}_1^{\mathrm{T}}\boldsymbol{\Sigma}\boldsymbol{\beta}_1 = \lambda_1$, o sea, es el más grande de los autovalores de $\boldsymbol{\Sigma}$ asociado al autovector $\boldsymbol{\beta}_1$. El Teorema 2.1.1 permite obtener consecutivamente las componentes principales. En efecto, la $j-$ésima componente principal de $\mathbf{x}$ está dada por $y_j = \boldsymbol{\beta}_j^{\mathrm{T}}\mathbf{x}$ donde

$$\boldsymbol{\beta}_j = \underset{\|\mathbf{a}\|=1,\ \mathbf{a}^{\mathrm{T}}\beta_i=0, 1\le i\le j-1}{\operatorname{argmax}} \sigma^2(F[\mathbf{a}]) \ .$$

En este enfoque, hemos considerado el desvío estandar como medida de variabilidad ó dispersion unidimensional y encontramos que $\boldsymbol{\beta}_1$ es la dirección en la que debemos proyectar el vector $\mathbf{x}$ si pretendemos obtener la mayor variabilidad posible. Por su parte, $\boldsymbol{\beta}_2$ es (entre las direcciones ortogonales a $\boldsymbol{\beta}_1$) la que debemos utilizar para proyectar el vector $\mathbf{x}$ de forma tal de garantizar mayor desvío posible. Y así sucesivamente.

Como mencionamos anteriormente, las técnicas **PP** procuran conservar este espíritu sustituyendo el desvío estandar por diferentes escalas unidimensionales para medir variabilidad o dispersión. En el presente contexto las escalas utilizadas se denominan también índices de proyección. Intuitivamente, así como en regresión las escalas robustas dan origen a estimadores robustos del parámetro de regresión, llamados $S-$estimadores, es de esperar que escalas robustas den lugar a procedimientos robustos para estimar las direcciones principales. Vale la pena mencionar que a diferencia de lo que ocurre en el caso clásico, cuando se toma como escala el desvío estándar, el procedimiento que obtiene las direcciones por maximización consecutiva usando un estimador de escala robusto no tiene porque dar el mismo resultado que el que las obtiene por minimización consecutiva.

Antes de definir el funcional asociado al sistema (2.4.1) y (2.4.2), recordemos que es una escala.

Definición 2.4.1. Sea $\mathcal{F}_1$ el espacio de las distribuciones en $I\!R$, $\mathbf{s}(\cdot) : \mathcal{F}_1 \to I\!R_{\ge 0}$ se dice un funcional de escala si y sólo si verifica $\mathbf{s}(G_{a+bY}) = |b|\,\mathbf{s}(G_Y)$ para todo $a, b \in I\!R$, donde G_{a+bY} es la distribución de $a + bY$ cuando $Y \sim G_Y$. De ahora en más, las escalas serán también llamadas índices de proyección.

Definiremos ahora el funcional asociado al sistema (2.4.1) y (2.4.2). Diremos que el funcional de escala $\mathbf{s}(\cdot)$ está asociado al estimador de escala univariado $\mathbf{s}_n$ si dada una muestra $y_1, \ldots, y_n$ de observaciones univariadas, $\mathbf{s}_n(\mathbf{Y}) = \mathbf{s}(F_n)$ donde $\mathbf{Y} = (y_1, \ldots, y_n)$ y F_n es la distribución empírica asociada a $y_1, \ldots, y_n$.

Definición 2.4.2. Sea $\mathbf{s}(\cdot)$ un funcional de escala. Dado un vector aleatorio $\mathbf{x} \in I\!R^p$ con distribución F definimos las $\mathbf{s}$-direcciones principales $\boldsymbol{\beta}_{i,\mathbf{s}} = \boldsymbol{\beta}_{i,\mathbf{s}}(F)$ y sus correspondientes $\mathbf{s}$-dispersiones $\lambda_{i,\mathbf{s}} = \lambda_{i,\mathbf{s}}(F)$ como

$$\boldsymbol{\beta}_{1,\mathbf{s}} = \underset{\|\mathbf{a}\|=1}{\operatorname{argmax}} \, \mathbf{s}^2(F[\mathbf{a}]) \qquad \lambda_{1,\mathbf{s}} = \max_{\|\mathbf{a}\|=1} \mathbf{s}^2(F[\mathbf{a}]) = \mathbf{s}^2(F[\boldsymbol{\beta}_{1,\mathbf{s}}]) \tag{2.4.3}$$

$$\boldsymbol{\beta}_{j,\mathbf{s}} = \underset{\mathbf{a}\in\mathbf{C}_j}{\operatorname{argmax}} \, \mathbf{s}^2(F[\mathbf{a}]) \qquad \lambda_{j,\mathbf{s}} = \max_{\mathbf{a}\in\mathbf{C}_j} \mathbf{s}^2(F[\mathbf{a}]) = \mathbf{s}^2(F[\boldsymbol{\beta}_{j,\mathbf{s}}]) \tag{2.4.4}$$

siendo $\mathbf{C}_j = \{\mathbf{a} \in I\!R^p \ / \ \|\mathbf{a}\| = 1, \mathbf{a}^{\mathrm{T}}\boldsymbol{\beta}_{i,\mathbf{s}} = 0 \ \ \forall\, 1 \le i \le j-1 \ \}$.
La $\mathbf{s}$-matriz de escala $\mathbf{V}_\mathbf{s} = \mathbf{V}_\mathbf{s}(F)$ se define como $\mathbf{V}_\mathbf{s} = \sum_{i=1}^p \lambda_{i,\mathbf{s}}\boldsymbol{\beta}_{i,\mathbf{s}}\boldsymbol{\beta}_{i,\mathbf{s}}^{\mathrm{T}}$.

Li & Chen (1985) consideran funcionales de escala $\mathbf{s}(\cdot)$ robustos. No incluimos esta restricción en la definición de forma tal que, poniendo $\mathbf{s}(G) = \sigma(G)$, recuperamos las definición clásica de las componentes principales, es decir, $\boldsymbol{\beta}_{i,\sigma}(F) = \boldsymbol{\beta}_i$ siendo $\boldsymbol{\beta}_i$ el autovector de la matriz de covarianza asociado al $i-$ésimo autovalor λ_i y además, $\lambda_{i,\sigma}(F) = \lambda_i$. Más aún, $\mathbf{V}_\sigma = \boldsymbol{\Sigma}$. En principio, uno podría pensar que distintas escalas podrían dar origen a distintos parámetros y que por lo tanto, los estimadores asociados podrían no ser consistentes a los parámetros de interés en caso de existir segundo momento. El Lema 2.4.1 muestra que bajo condiciones generales, el funcional es Fisher–consistente, o sea, efectivamente estamos obteniendo los autovectores y autovalores de la matriz de dispersión.

Observemos que los estimadores definidos en (2.4.1) y (2.4.2) corresponden a tomar simplemente el funcional anterior evaluado en la distribución empírica. Es decir, tenemos la siguiente definición.

Definición 2.4.3. Sea $\mathbf{x}_1, \ldots, \mathbf{x}_n$, una muestra aleatoria de un vector $\mathbf{x} \in I\!R^p$ con distribución F. Sea F_n la distribución empírica de dicha muestra. Definimos los estimadores de *projection–pursuit* (**PP**) de las direcciones principales, su tamaño y la matriz de escala asociados al funcional de escala $\mathbf{s}$ (asociado al estimador de escala univariado $\mathbf{s}_n$) como

$$\widehat{\lambda}_{j,\mathbf{s}} = \lambda_{j,\mathbf{s}}(F_n) \ , \quad \widehat{\boldsymbol{\beta}}_{j,\mathbf{s}} = \boldsymbol{\beta}_{j,\mathbf{s}}(F_n) \ , \quad \widehat{\mathbf{V}}_{\mathbf{s}} = \mathbf{V}_{\mathbf{s}}(F_n) \ , \tag{2.4.5}$$

o sea, $\widehat{\boldsymbol{\beta}}_{j,\mathbf{s}} = \text{argmax}_{\mathbf{a}\in\mathcal{C}_j} \mathbf{s}_n^2(\mathbf{a}^{\mathrm{T}}\mathbf{X})$, con $\mathcal{C}_j = \{\mathbf{a} \in I\!R^p, \|\mathbf{a}\| = 1, \mathbf{a}^{\mathrm{T}}\widehat{\boldsymbol{\beta}}_{i,\mathbf{s}} = 0 \ , \ 1 \leq i \leq j-1\}$, $\widehat{\lambda}_{j,\mathbf{s}} = \mathbf{s}_n^2(\widehat{\boldsymbol{\beta}}_{j,\mathbf{s}}^{\mathrm{T}}\mathbf{X})$ y $\widehat{\mathbf{V}}_{\mathbf{s}} = \sum_{i=1}^p \widehat{\lambda}_{j,\mathbf{s}} \widehat{\boldsymbol{\beta}}_{j,\mathbf{s}} \widehat{\boldsymbol{\beta}}_{j,\mathbf{s}}^{\mathrm{T}}$.

Una manera de extender la noción de valor esperado y matriz de covarianza de un vector aleatorio a un contexto sin segundo momento finito la constituyen los parámetros de posición $\boldsymbol{\mu}$ y la matriz de escala $\boldsymbol{\Sigma}$ de una distribución elíptica F_0, es decir, una de la forma

$$f_0(\mathbf{x}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = |\boldsymbol{\Sigma}|^{-\frac{1}{2}} g_0\left((\mathbf{x}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right) \ ,$$

donde $|\boldsymbol{\Sigma}|$ indica el determinante de $\boldsymbol{\Sigma}$ y g_0 es una función fija no–negativa tal que

$$\int_0^\infty g_0(r) r^{\frac{p}{2}-1} dr = \frac{1}{\pi^{\frac{p}{2}}} \Gamma\left(\frac{p}{2}\right) \ .$$

Indicaremos por G_0 la distribución esférica asociada a g_0. Luego, si $\mathbf{x} \sim G_0$ se tiene que $\mathbf{Hx} \sim \mathbf{x}$ para toda matriz $\mathbf{H} \in \mathcal{O}(p)$, siendo $\mathcal{O}(p)$ el grupo de las matrices ortogonales de orden p. Es decir, $\mathbf{x}$ es invariante por transformaciones ortogonales. En nuestro contexto, como cuando se desean obtener estimadores robustos de la posición y escala $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, deseamos obtener estimadores de $\boldsymbol{\beta}$ y $\boldsymbol{\Lambda}$, con $\boldsymbol{\Sigma} = \boldsymbol{\beta}\boldsymbol{\Lambda}\boldsymbol{\beta}^{\mathrm{T}}$, suponiendo que la distribución de las observaciones es aproximadamente conocida. Para ser más precisos, se supone en muchos casos que $\mathbf{x} = \boldsymbol{\mu} + \boldsymbol{\Sigma}^{1/2}\mathbf{z}$ donde la distribución de $\mathbf{z}$ pertenece a un entorno de G_0 y usualmente los entornos que se consideran son de contaminación alrededor de la distribución normal para obtener sesgos máximos en el entorno (ver, por ejemplo, Patak 1991, Berrendero, 1996). Cuando $\mathbf{x}$ tiene una distribución elípticamente simétrica F, o sea, cuando las distribuciones del entorno de G_0 se restringen a ser esféricamente simétricas la mayoría de los funcionales robustos de escala, $\boldsymbol{\Sigma}_{\mathrm{R}}(\cdot)$ son proporcionales a $\boldsymbol{\Sigma}$ en F y pueden calibrarse de modo que $\boldsymbol{\Sigma}_{\mathrm{R}}(F_0) = \boldsymbol{\Sigma}$. De esta forma, los parámetros del problema de componentes principales quedarán unívocamente determinados. El siguiente resultado cuya demostración puede encontrarse en Li & Chen (1985) muestra que efectivamente, los funcionales asociados a las direcciones principales son Fisher–consistentes y que los asociados a los autovalores lo son si se calibra adecuadamente el funcional de escala.

Lema 2.4.1. *a) Los funcionales $\boldsymbol{\beta}_{j,\mathbf{s}}(\cdot)$ y $\mathbf{V}_{\mathbf{s}}(\cdot)$ son ortogonalmente equivariantes, mientras que $\lambda_{j,\mathbf{s}}(\cdot)$ es ortogonalmente invariante, es decir, si $\mathbf{H} \in \mathcal{O}(p)$, y si $\mathbf{x} \sim F$, entonces, para $1 \leq j \leq p$ y $\mathbf{b} \in I\!R^p$,*

$$\lambda_{j,\mathbf{s}}(F_{\mathbf{Hx+b}}) = \lambda_{j,\mathbf{s}}(F) \ , \quad \boldsymbol{\beta}_{j,\mathbf{s}}(F_{\mathbf{Hx+b}}) = \mathbf{H}\boldsymbol{\beta}_{j,\mathbf{s}}(F) \ , \quad \mathbf{V}_{\mathbf{s}}(F_{\mathbf{Hx+b}}) = \mathbf{H}\mathbf{V}_{\mathbf{s}}(F)\mathbf{H}^{\mathrm{T}} \ .$$

b) $\mathbf{V}_{\mathbf{s}}(\cdot)$ es afín equivariante dentro de la familia de distribuciones elípticas, es decir, si $\mathbf{x} \sim F$ con F una distribución elíptica entonces: $\mathbf{V}_{\mathbf{s}}(F_{\mathbf{Ax}}) = \mathbf{A}\mathbf{V}_{\mathbf{s}}(F)\mathbf{A}^{\mathrm{T}}$.

c) Sea G una distribución esférica en $I\!R^p$ y sea $G[\mathbf{e}_1]$ la distribución de $\mathbf{e}_1^{\mathrm{T}}\mathbf{x}$ cuando $\mathbf{x} \sim G$, siendo $\mathbf{e}_1$ el primer vector canónico. Sea $\mathbf{x} = \boldsymbol{\mu} + \mathbf{Cz}$ con $\mathbf{C}$ inversible, $\boldsymbol{\Sigma} = \mathbf{CC}^{\mathrm{T}}$ y $\mathbf{z} \sim G$, o sea, $\mathbf{x} \sim F$ perteneciente a la familia elíptica generada por G. Entonces, $\boldsymbol{\beta}_{j,\mathbf{s}}(F)$ es el autovector de $\boldsymbol{\Sigma}$ asociado al $j-$ésimo autovalor λ_j ($\lambda_1 \geq \cdots \geq \lambda_p$). Más aún, si $\mathbf{s}(G[\mathbf{e}_1]) = 1$,entonces $\lambda_{j,\mathbf{s}}(F) = \lambda_j$ y $\mathbf{V}_{\mathbf{s}}(F) = \boldsymbol{\Sigma}$.

El Lema 2.4.1 muestra, por lo tanto, que los estimadores definidos por (2.4.1) y (2.4.2) son equivariantes por transformaciones ortogonales, pero no tienen porqué serlo por transformaciones afines ya que la distribución empírica no tiene porqué pertenecer a una familia elíptica. La consecuencia importante de c) es que el funcional asociado a las direcciones es Fisher–consistente dentro de la familia de distribuciones elípticas mientras que el funcional asociado al tamaño de dichas direcciones puede calibrarse de modo a resultar también Fisher–consistente en el centro del entorno. De esta forma, es posible obtener estimadores consistentes a los autovectores y autovalores de la matriz de escala $\boldsymbol{\Sigma}$ para toda escala $\mathbf{s}$ convenientemente calibrada.

Li & Chen (1985) muestran que, salvo que $\mathbf{x} \sim F$ con F concentrada en un subespacio de dimensión $q < p$, el punto de ruptura del funcional asociado a los autovalores es el mismo que el del funcional de escala univariado $\mathbf{s}$, entendiendo que el funcional de autovalores se rompe si éste va a 0 o a infinito. Resultados respecto del punto de ruptura de las direcciones principales fueron obtenidos por Patak (1991) y Berrendero (1996), quienes mencionaron que la peor situación en entornos de contaminación corresponde a intercambiar dos autovalores adyacentes de $\boldsymbol{\Sigma}$. En ese caso, se espera que el sesgo de la $j-$ésima dirección dependa del cociente $r_j = \lambda_j/\lambda_{j+1}$, siendo $\lambda_1 > \ldots > \lambda_p$ los autovaores ordenados de $\boldsymbol{\Sigma}$. A medida que crece el cociente r_j, el sesgo decrece.

Cuando la distribución subyacente pertenece a una familia elíptica, para la cual como hemos mencionado la matriz de dispersión tiene una interpretación intuitiva, es posible probar que los estimadores de *projection–pursuit* son cualitativamente robustos y consistentes si la escala $\mathbf{s}$ lo es.

Teorema 2.4.1. *Supongamos que el funcional univariado de escala $\mathbf{s}$ es cualitativamente robusto y sea F una distribución elíptica.*

a) Los funcionales $\lambda_{i,\mathbf{s}}(\cdot)$ y $\mathbf{V}_{\mathbf{s}}(\cdot)$ son débilmente continuos en F. Más aún, si los autovalores de $\boldsymbol{\Sigma}$ tienen multiplicidad 1, siendo $\boldsymbol{\Sigma}$ la matriz de dispersión asociada a F, entonces, $\boldsymbol{\beta}_{i,\mathbf{s}}$ son débilmente continuos.

b) Si los autovalores de $\boldsymbol{\Sigma}$ tienen multiplicidad mayor a 1, sea m_i la multiplicidad del $i-$ésimo valor distinto, ξ_i, de los autovalores, es decir, $\xi_1 = \lambda_1 = \cdots = \lambda_{m_1} > \xi_2 = \lambda_{m_1+1} = \cdots = \lambda_{m_1+m_2} > \ldots$. Definamos $\ell_0 = 0$, $\ell_k = \sum_{j=1}^{k} m_j$ y $E_k(F) = \sum_{i=\ell_{k-1}+1}^{\ell_k} \boldsymbol{\beta}_{i,\mathbf{s}}(F)\boldsymbol{\beta}_{i,\mathbf{s}}^{\mathrm{T}}(F)$ el funcional asociado a la projección sobre el subespacio asociado a ξ_k. Entonces, $E_k(\cdot)$ es débilmente continuo en F.

c) Los funcionales $\lambda_{i,\mathbf{s}}(\cdot)$, $E_k(\cdot)$ y $\mathbf{V}_{\mathbf{s}}(\cdot)$ son cualitativamente robustos en F.

Teorema 2.4.2. *Sean $\mathbf{x}_1 \ldots \mathbf{x}_n$, $\mathbf{x}_i \in I\!R^p$, independientes e idénticamente distribuídas con distribución F perteneciente a una familia elíptica. Sea $\mathbf{s}(\cdot)$ un funcional de escala asociado a un estimador de escala univariado robusto y consistente tal que $\mathbf{s}(G[\mathbf{e}_1]) = 1$, donde $\mathbf{e}_1$ es el primer vector canónico y G es la distribución esférica asociada a F, es decir, $\mathbf{x} = \boldsymbol{\mu} + \boldsymbol{\Sigma}^{\frac{1}{2}}\mathbf{z}$ con $\mathbf{z} \sim G$. Sea como en el Teorem 2.4.1, m_i la multiplicidad del $i-$ésimo valor distinto entre los autovalores de $\boldsymbol{\Sigma}$. Entonces, los estimadores de projection–pursuit, $\widehat{\lambda}_{i,\mathbf{s}}$, $\widehat{E}_k = \sum_{i=\ell_{k-1}+1}^{\ell_k} \widehat{\boldsymbol{\beta}}_{i,\mathbf{s}}\widehat{\boldsymbol{\beta}}_{i,\mathbf{s}}^{\mathrm{T}}$ y $\widehat{\mathbf{V}}_{\mathbf{s}}$ son fuertemente consistentes para λ_i, $E_k = \sum_{i=\ell_{k-1}+1}^{\ell_k} \boldsymbol{\beta}_i\boldsymbol{\beta}_i^{\mathrm{T}}$ y $\boldsymbol{\Sigma} = \boldsymbol{\beta}\boldsymbol{\Lambda}\boldsymbol{\beta}^{\mathrm{T}}$, respectivamente.*

2.4.2. Distribución asintótica de las componentes principales basadas en una escala robusta.

En esta sección vamos a recordar los resultados obtenidos por Cui, He & Ng (2003) quienes prueban la distribución asintótica para los estimadores definidos en (2.4.5). Las funciones de influencia juegan un papel fundamental en la teoría asintótica, por lo cual comenzaremos recordando su definición.

Definición 2.4.4. La función de influencia del funcional $\mathbf{L}$ en la distribución H evaluado en el punto y, está dada por

$$\mathrm{IF}(y, \mathbf{L}; H) = \mathrm{lím}_{\epsilon \to 0} \frac{\mathbf{L}((1-\epsilon)H + \epsilon\Delta_y) - \mathbf{L}(H)}{\epsilon},$$

donde Δ_y es la masa puntual en y.

Recordemos que si H es una distribución en $I\!R^p$ y $\mathbf{a} \in I\!R^p$, $H[\mathbf{a}]$ indica la distribución de la variable aleatoria $\mathbf{a}^{\mathrm{T}}\mathbf{x}$. Dado un funcional robusto de escala univariado $\mathbf{s}(\cdot) : \mathcal{F}_1 \to I\!R_{\geq 0}$, llamaremos para cada $\mathbf{a} \in I\!R^p$ fijo, $\mathbf{s}_{\mathbf{a}} : \mathcal{F}_p \to I\!R_{\geq 0}$ al funcional $\mathbf{s}_{\mathbf{a}}(H) = \mathbf{s}(H[\mathbf{a}])$.

De ahora en más, $\mathcal{S}_p$ indicará la esfera unitaria p-dimensional.

Proposición 2.4.1. *Sea $\mathbf{s}(\cdot)$ un funcional robusto de escala, $\mathbf{s}_{\mathbf{a}} : \mathcal{F}_p \to I\!R_{\geq 0}$ el funcional $\mathbf{s}_{\mathbf{a}}(H) = \mathbf{s}(H[\mathbf{a}])$ y $H_{\mathbf{a}}$ la distribución de $\mathbf{a}^{\mathrm{T}}\mathbf{x}/\mathbf{s}_{\mathbf{a}}(H)$ cuando $\mathbf{x} \sim H$. Entonces, la función de influencia del funcional $\mathbf{s}_{\mathbf{a}}$, verifica:*

$$h(\mathbf{x}, \mathbf{a}) = IF(\mathbf{x}, \mathbf{s}_{\mathbf{a}}; H) = \mathbf{s}_{\mathbf{a}}(H)\, IF(\mathbf{x}^{\mathrm{T}}\mathbf{a}/\mathbf{s}_{\mathbf{a}}(H), \mathbf{s}; H_{\mathbf{a}})$$

La función $h(\mathbf{x}, \mathbf{a})$ depende de $\mathbf{x}$ sólo a través de $\mathbf{a}^{\mathrm{T}}\mathbf{x}$. Si bien sólo nos interesa el caso $\mathbf{a} \in \mathcal{S}_p$, consideraremos a $h(\mathbf{x}, \mathbf{a})$ como función de $\mathbf{a} \in I\!R^p$ para facilitar la diferenciación.

Enunciaremos a continuación una serie de supuestos bajo los cuales Cui, He & Ng (2003) obtienen resultados asintóticos para los estimadores **PP** definidos en (2.4.5). Conciernen al funcional de dispersión tanto en la distribución que genera los datos como en diferentes empíricas.

De ahora en más, supondremos que $\mathbf{x}_1, \ldots, \mathbf{x}_n$ son independientes e idénticamente distribuídas con distribución F fija. Indicaremos por $\mathbf{s}_n$ al estimador robusto de escala univariado asociado a $\mathbf{s}$, o sea, $\mathbf{s}_n(\mathbf{Y}) = \mathbf{s}(F_{n,\mathbf{Y}})$ donde $\mathbf{Y} = (y_1, \ldots, y_n)$ es una muestra de variables univariadas y $F_{n,\mathbf{Y}}$ es su distribución empírica. Por simplicidad de notación, $\varsigma_n(\mathbf{a})$ indicará a $\varsigma_n(\mathbf{a}) = \mathbf{s}_n(\mathbf{a}^{\mathrm{T}}\mathbf{X}) = \mathbf{s}\,(F_{n,\mathbf{a}^{\mathrm{T}}\mathbf{X}})$ donde $F_{n,\mathbf{a}^{\mathrm{T}}\mathbf{X}}$ es la distribución empírica de $\mathbf{a}^{\mathrm{T}}\mathbf{x}_1, \ldots, \mathbf{a}^{\mathrm{T}}\mathbf{x}_n$ y $\mathbf{a}^{\mathrm{T}}\mathbf{X} = (\mathbf{a}^{\mathrm{T}}\mathbf{x}_1, \ldots, \mathbf{a}^{\mathrm{T}}\mathbf{x}_n)$. Análogamente, $\varsigma(\mathbf{a})$ indicará a $\varsigma(\mathbf{a}) = \mathbf{s}(F[\mathbf{a}])$. Finalmente, $\dot{\varsigma}(\mathbf{a})$ denotará la derivada con respecto de $\mathbf{a}$ de la función $\varsigma : I\!R^p \to I\!R$ mientras que $\dot{\varsigma}_n(\mathbf{a})$ indicará la derivada de $\varsigma_n(\mathbf{a})$ con respecto a $\mathbf{a}$, en caso de existir.

> **C1:** Para cada $\mathbf{x}$ fijo, la función $h(\mathbf{x}, \mathbf{a})$ es continua en $\mathbf{a}$ y diferenciable a trozos, en el siguiente sentido: existe un entero m y una partición $\{\mathcal{R}_j(\mathbf{x}),\ j = 1, \ldots, m\}$ de $I\!R^p$ tal que $h(\mathbf{x}, \mathbf{a})$ es diferenciable en $\mathbf{a}$ en el interior de cada $\mathcal{R}_j(\mathbf{x})$.

C2: El siguiente desarrollo $\varsigma_n(\mathbf{a}) - \varsigma(\mathbf{a}) = n^{-1}\sum_{i=1}^n h(\mathbf{x}_i, \mathbf{a}) + o_p\left(n^{-\frac{1}{2}}\right)$ vale uniformemente en $\mathbf{a} \in \mathcal{S}_p$, con $\mathrm{E}h(\mathbf{x}, \mathbf{a}) = 0$ y $\mathrm{E}\left(\sup_{\mathbf{a}\in\mathcal{S}_p} |h(\mathbf{x}, \mathbf{a})|\right) < \infty$.

C3: Supondremos que $\dot{\varsigma}(\mathbf{a})$ es diferenciable con derivada continua $\ddot{\varsigma}(\mathbf{a})$.

C4: La derivada $\dot{\varsigma}_n(\mathbf{a})$ de $\varsigma_n(\mathbf{a})$ con respecto a $\mathbf{a}$ existe en casi todo punto $\mathbf{a} \in \mathcal{S}_p$. Más aún,

$$\dot{\varsigma}_n(\mathbf{a}) - \dot{\varsigma}(\mathbf{a}) = n^{-1}\sum_{i=1}^n h^\star(\mathbf{x}_i, \mathbf{a}) + o_p\left(n^{-\frac{1}{2}}\right)$$

uniformemente en $\mathbf{a} \in \mathcal{S}_p$ con $h^\star$ tal que

i) $\mathrm{E}h^\star(\mathbf{x}, \mathbf{a}) = 0$ para todo $\mathbf{a}$

ii) $h^\star(\mathbf{x}, \mathbf{a})$ continua a trozos en $\mathbf{a}$.

La condición **C2** es una representación de Bahadur uniforme para los estimadores de escala y se verifica para la mayoría de los $M-$estimadores de escala. La continuidad a trozos de $h^\star$ en **C4**ii) debe entenderse como en **C1**. En la mayoría de las situaciones $h^\star(\mathbf{x}, \mathbf{a}) = (\partial/\partial\mathbf{a})h(\mathbf{x}, \mathbf{a}) = \dot{h}(\mathbf{x}, \mathbf{a})$, pero no se impone esa restricción para facilitar la verificación de **C4**. Tampoco pedimos que la función $\dot{h}(\mathbf{x}, \mathbf{a})$ sea suave ya que de hecho el $M-$estimador de escala con función de Huber corresponde a una función $\dot{h}(\mathbf{x}, \mathbf{a})$ discontinua. La condición **C4** es la más difícil de verificar y como se menciona en Cui, He & Ng (2003) puede ser debilitada y reemplazada por la siguiente condición a fin de obtener la distribución asintótica de los estimadores de *projection–pursuit*

C4*: $\varsigma_n(\mathbf{a} + \delta\mathbf{b}) - \varsigma_n(\mathbf{a}) = n^{-1}\sum_{i=1}^n \{h(\mathbf{x}_i, \mathbf{a} + \delta\mathbf{b}) - h(\mathbf{x}_i, \mathbf{a})\} + o_p\left(\delta n^{-\frac{1}{2}}\right) + o_p\left(n^{-1}\right)$, uniformemente en $\mathbf{a} \in \mathcal{S}_p$, $\mathbf{b} \in \mathcal{S}_p$ cuando $\delta \to 0$ y $n \to \infty$.

Los siguientes Teoremas cuya demostración puede verse en Cui, He & Ng (2003) reproducen los resultados de consistencia y distribución asintótica obtenidos por dichos autores.

Teorema 2.4.3. *Sean $\mathbf{x}_1, \ldots, \mathbf{x}_n$ independientes e idénticamente distribuidas con distribución F. Sean $\boldsymbol{\beta}_{j,\mathbf{s}} = \boldsymbol{\beta}_{j,\mathbf{s}}(F)$, $\lambda_{j,\mathbf{s}} = \lambda_{j,\mathbf{s}}(F)$ las $\mathbf{s}-$funcionales de las direcciones principales y su tamaño definidos en (2.4.4) evaluados en F y sean $\widehat{\boldsymbol{\beta}}_{j,\mathbf{s}}$ y $\widehat{\lambda}_{j,\mathbf{s}}$ sus correspondientes estimadores definidos en (2.4.5). Supongamos que $\lambda_{1,\mathbf{s}} > \lambda_{2,\mathbf{s}} > \ldots > \lambda_{q,\mathbf{s}}$ para algún $q \le p$ y que $\boldsymbol{\beta}_{j,\mathbf{s}}$ son únicos salvo cambios de signo. Bajo las condiciones **C1** y **C2**, se tiene que para $1 \le j \le q$, $\widehat{\lambda}_{j,\mathbf{s}} \xrightarrow{p} \lambda_{j,\mathbf{s}}$ y $\widehat{\boldsymbol{\beta}}_{j,\mathbf{s}} \xrightarrow{p} \boldsymbol{\beta}_{j,\mathbf{s}}$, es decir, los estimadores **PP** son consistentes.*

Observemos que si la distribución F es elíptica, $\boldsymbol{\beta}_{j,\mathbf{s}} = \boldsymbol{\beta}_j$ y por lo tanto, los estimadores son consistentes a los autovectores de la matriz de dispersión de las observaciones, como se deseaba.

Para obtener una expresión para la distribución asintótica, definamos $\mathbf{I}_p$ la matriz identidad de $I\!R^{p\times p}$. En lo que sigue $\boldsymbol{\beta}_{j,\mathbf{s}} = \boldsymbol{\beta}_{j,\mathbf{s}}(F)$ y $\lambda_{j,\mathbf{s}} = \lambda_{j,\mathbf{s}}(F)$. Sean

- $H(\mathbf{x}, \mathbf{a}) = h^*(\mathbf{x}, \mathbf{a}) + \dot{\varsigma}(\mathbf{a})$,
- $\widetilde{\mathbf{u}}_m = -n^{-\frac{1}{2}}\sum_{i=1}^n h^*\left(\mathbf{x}_i, \boldsymbol{\beta}_{m,\mathbf{s}}\right)$,
- $\widetilde{\mathbf{P}}_{m+1} = \mathbf{I}_p - \sum_{i=1}^m \boldsymbol{\beta}_{i,\mathbf{s}}\boldsymbol{\beta}_{i,\mathbf{s}}^{\mathrm{T}}$ matriz de proyección sobre el subespacio $< \boldsymbol{\beta}_{1,\mathbf{s}}, \ldots, \boldsymbol{\beta}_{m,\mathbf{s}} >^\perp$,
- $\widetilde{\mathbf{B}}_{jm} = \boldsymbol{\beta}_{j,\mathbf{s}}^{\mathrm{T}}\dot{\varsigma}(\boldsymbol{\beta}_{m,\mathbf{s}})\mathbf{I}_p + \boldsymbol{\beta}_{j,\mathbf{s}}\dot{\varsigma}(\boldsymbol{\beta}_{m,\mathbf{s}})^{\mathrm{T}}$,

- $\widetilde{\mathbf{A}}_m = \widetilde{\mathbf{P}}_{m+1}\ddot{\varsigma}(\boldsymbol{\beta}_{m,\mathbf{s}}) - \boldsymbol{\beta}_{m,\mathbf{s}}^{\mathrm{T}}\dot{\varsigma}(\boldsymbol{\beta}_{m,\mathbf{s}})\mathbf{I}_p - \sum_{j=1}^{m-1} \boldsymbol{\beta}_{j,\mathbf{s}}^{\mathrm{T}}\dot{\varsigma}(\boldsymbol{\beta}_{m,\mathbf{s}})\boldsymbol{\beta}_{m,\mathbf{s}}\ \boldsymbol{\beta}_{j,\mathbf{s}}^{\mathrm{T}}$.
- En forma recurrente, podemos definir para $1 \leq m \leq q$, $\widetilde{\mathbf{Z}}_m = \sum_{j=0}^{m-1} \widetilde{\mathbf{A}}_m^{-1}\widetilde{\mathbf{B}}_{jm}\widetilde{\mathbf{Z}}_j + \widetilde{\mathbf{A}}_m^{-1}\widetilde{\mathbf{P}}_{m+1}\widetilde{\mathbf{u}}_m$, donde $\widetilde{\mathbf{Z}}_0 = 0$ y suponemos la existencia de $\widetilde{\mathbf{A}}_m^{-1}$.

  Claramente, $\widetilde{\mathbf{Z}}_m$ puede ser representado en la forma $\widetilde{\mathbf{Z}}_m = \sum_{j=0}^{m} \widetilde{\mathbf{C}}_{jm}\widetilde{\mathbf{u}}_j$, donde las matrices $\widetilde{\mathbf{C}}_{jm}$ están determinadas por $\widetilde{\mathbf{A}}_m,\ \widetilde{\mathbf{B}}_{jm}\ \ y\ \ \widetilde{\mathbf{P}}_{m+1}$.
- $\xi_m(\mathbf{x}_i) = \sum_{j=1}^{m} \widetilde{\mathbf{C}}_{jm} h^*(\mathbf{x}_i, \boldsymbol{\beta}_{j,\mathbf{s}})$
- $\eta_m(\mathbf{x}_i) = 2\sqrt{\lambda_{m,\mathbf{s}}}\left\{h(\mathbf{x}_i, \boldsymbol{\beta}_{m,\mathbf{s}}) + \sum_{j=1}^{m-1} \boldsymbol{\beta}_{j,\mathbf{s}}^{\mathrm{T}}\dot{\varsigma}(\boldsymbol{\beta}_{m,\mathbf{s}})\boldsymbol{\beta}_{j,\mathbf{s}}^{\mathrm{T}}\xi_m(\mathbf{x}_i)\right\}$

Teorema 2.4.4. *Bajo las condiciones del Teorema 2.4.3, si además también se verifican las condiciones* ***C3*** *y* ***C4*** *y si las matrices* $\widetilde{\mathbf{A}}_m$, $1 \leq m \leq q$, *son no singulares, tenemos que*

$$\begin{aligned} \widehat{\boldsymbol{\beta}}_{m,\mathbf{s}} - \boldsymbol{\beta}_{m,\mathbf{s}} &= \frac{1}{n}\sum_{i=1}^{n} \xi_m(\mathbf{x}_i) + o_p(n^{-\frac{1}{2}}) \\ \widehat{\lambda}_{m,\mathbf{s}} - \lambda_{m,\mathbf{s}} &= \frac{1}{n}\sum_{i=1}^{n} \eta_m(\mathbf{x}_i) + o_p(n^{-\frac{1}{2}})\ . \end{aligned}$$

Vale la pena mencionar que, en el caso de distribuciones elípticas, el desarrollo dado por el Teorema 2.4.4 coincide con el que se obtendría utilizando la funciones de influencia obtenidas por Croux & Ruiz–Gazen (2005) ya que el segundo sumando de $\eta_m(\mathbf{x})$ se anula pues $\varsigma(\mathbf{a}) = \mathbf{a}^{\mathrm{T}}\boldsymbol{\Sigma}\mathbf{a}$, siendo $\boldsymbol{\Sigma}$ la matriz de escala de $\boldsymbol{\Sigma}$. Como consecuencia del Teorema 2.4.4, la distribución conjunta de

$$n^{-\frac{1}{2}}(\widehat{\boldsymbol{\beta}}_{1,\mathbf{s}} - \boldsymbol{\beta}_{1,\mathbf{s}}, \ldots, \widehat{\boldsymbol{\beta}}_{q,\mathbf{s}} - \boldsymbol{\beta}_{q,\mathbf{s}}, \widehat{\lambda}_{1,\mathbf{s}} - \lambda_{1,\mathbf{s}}, \ldots, \widehat{\lambda}_{q,\mathbf{s}} - \lambda_{q,\mathbf{s}})$$

converge a una distribución normal multivariada con media 0 y matriz de covarianza dada por $\mathrm{COV}_F\,(\eta_1(\mathbf{x}), \ldots, \eta_q(\mathbf{x}), \ldots, \xi_1(\mathbf{x}), \ldots, \xi_q(\mathbf{x})\ )$.

Capítulo 3

Componentes Principales Comunes

3.1. El modelo CPC

Como mencionamos en la Introducción, el modelo de componentes principales comunes (CPC) puede verse como una generalización de las componentes principales al caso de varias poblaciones de datos multivariados. El modelo resulta parsimonioso sin llegar a exigir igualdad en las matrices de covarianza de las distintas poblaciones. Más específicamente, supongamos tener muestras independientes de k vectores aleatorios $\mathbf{x}_1 \ldots, \mathbf{x}_k$ en $I\!R^p$ independientes tales que $\mathbf{x}_i$ tiene media (parámetro de posición) $\boldsymbol{\mu}_i$ y matriz de dispersión $\boldsymbol{\Sigma}_i$. Usualmente, en análisis multivariado se tratan a las matrices $\boldsymbol{\Sigma}_i$ como no relacionadas entre sí cuando un test de igualdad entre ellas no rechaza. Como menciona Flury (1988), en contraste con la situación univariada, la desigualdad de matrices de covarianza no es simplemente desigualdad, ya que hay, de hecho, muchas formas en las que las matrices pueden diferir. El modelo CPC supone que dichas matrices son conmutables, es decir, que son simultáneamente diagonalizables y, por lo tanto, pueden escribirse de la siguiente manera

$$\boldsymbol{\Sigma}_i = \boldsymbol{\beta}\boldsymbol{\Lambda}_i\boldsymbol{\beta}^{\mathrm{T}} , \quad \text{para } i = 1, \ldots, k , \tag{3.1.1}$$

con $\boldsymbol{\Lambda}_i = \operatorname{diag}(\lambda_{i1}, \ldots, \lambda_{ip})$, la matriz diagonal formada por los autovalores de $\boldsymbol{\Sigma}_i$ y $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_p)$ una matriz ortogonal cuya $j-$ésima columna, $\boldsymbol{\beta}_j$, es el autovector asociado al autovalor λ_{ij}. En otras palabras, el modelo supone que se cuenta con k poblaciones vectores aleatorios cuyas matrices de covarianza se diagonalizan con la misma matriz ortogonal $\boldsymbol{\beta}$. Luego, ya sea para estimar las matrices $\boldsymbol{\Sigma}_i$ como sus autovalores y/o autovectores, bajo un modelo CPC, la relación (3.1.1) debe tenerse en cuenta. Cuando trabajamos con una sola población, los autovectores fueron identificados, salvo cambio de signo, ordenando en forma decreciente los correspondientes autovalores. En el presente contexto, no hay una manera natural para ordenarlos ya que, en principio, no se preserva el orden de los elementos de la diagonal de $\boldsymbol{\Lambda}_i$ entre las distintas poblaciones. Con el propósito de identificar las direcciones principales, supondremos que la matriz de covarianza de alguna de las k poblaciones tiene sus autovalores todos distintos. Sin pérdida de generalidad, podemos suponer que dicha población es la población 1 y por lo tanto, de ahora en más supondremos que $\lambda_{11} > \ldots > \lambda_{1p}$.

El modelo CPC es uno de los niveles de jerarquía considerados por Flury (1988), quien definió las siguientes relaciones entre matrices de escala

- **Nivel 1.** Las matrices $\boldsymbol{\Sigma}_i$ son arbitrarias, o sea, no exhiben estructura de relación entre ellas.

- **Nivel 2.** Las matrices cumplen el modelo (CPC), o sea, $\boldsymbol{\Sigma}_i = \boldsymbol{\beta}\boldsymbol{\Lambda}_i\boldsymbol{\beta}^{\mathrm{T}}$, donde $\boldsymbol{\beta} = \left(\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_p\right)$ es la matriz ortogonal de los autovectores comunes y $\boldsymbol{\Lambda}_i = \operatorname{diag}(\lambda_{i1}, \ldots, \lambda_{ip})$ son matrices diagonales que contienen los autovalores de cada población.

- **Nivel 3.** Las matrices son proporcionales entre sí, o sea, $\mathbf{\Sigma}_i = \rho_i \mathbf{\Sigma}_1$, con ρ_i la constante de proporcionalidad, $\rho_1 = 1$.
- **Nivel 4.** $\mathbf{\Sigma}_i = \mathbf{\Sigma}_1$, para todo i.

Sin considerar los parámetros de posición, el número de parámetros de cada nivel es $kp(p+1)/2$, $kp + p(p-1)/2$, $k-1+p(p+1)/2$ y $p(p+1)/2$, respectivamente. La diferencia entre el número de parámetros en los niveles 1 y 4 es $(k-1)p(p+1)/2$ que puede ser muy grande en la práctica, especialmente para datos de alta dimensión. Por esa razón, de ser posible debe usarse el más parsimonioso entre los niveles descriptos. En esta tesis, nos interesará estudiar el modelo CPC. En la Sección 3.2, se describirá el procedimiento de estimación por máxima verosimilitud que está ampliamente desarrollado en Flury (1988), mientras que en la Sección 3.3, se describirá el enfoque plug–in y el de *projection–pursuit* para este modelo que fue estudiado por Boente y Orellana (2001).

3.2. Estimadores clásicos bajo un modelo de componentes principales comunes

Sean $\mathbf{x}_{i1}, \dots, \mathbf{x}_{in_i}$, $1 \le i \le k$, observaciones independientes de las k poblaciones en $I\!R^p$. Estamos interesados en estimar, bajo el modelo CPC, los autovectores comunes $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_p)$ de las matrices $\mathbf{\Sigma}_i$ y los autovalores $\text{diag}(\mathbf{\Lambda}_i)$ correspondientes a la $i-$ésima población.

Flury (1988) obtuvo los estimadores de máxima verosimilitud para los parámetros del modelo CPC, suponiendo normalidad para cada población. Recordemos que $\mathcal{O}(p)$ indica el grupo de matrices ortogonales de orden p. El estimador de máxima verosimilitud $\widehat{\boldsymbol{\beta}}_{\text{MV}}$ de $\boldsymbol{\beta}$ minimiza, entre todas las matrices $\mathbf{B} \in \mathcal{O}(p)$, la función

$$\Phi(\mathbf{B}) = \prod_{i=1}^{k} \left[\frac{\det\left(\text{diag}(\mathbf{B}^{\text{T}} \widehat{\mathbf{\Sigma}}_i \mathbf{B})\right)}{\det(\mathbf{B}^{\text{T}} \widehat{\mathbf{\Sigma}}_i \mathbf{B})} \right]^{n_i}$$

donde $\widehat{\mathbf{\Sigma}}_i$ es el estimador de máxima verosimilitud de la matriz de covarianza, o sea,

$$\widehat{\mathbf{\Sigma}}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} (\mathbf{x}_{ij} - \overline{\mathbf{x}}_i)(\mathbf{x}_{ij} - \overline{\mathbf{x}}_i)^{\text{T}} = \frac{n_i - 1}{n_i} \mathbf{S}_i \ .$$

Es fácil mostrar que los estimadores de máxima verosimilitud de $\boldsymbol{\beta}$ y $\mathbf{\Lambda}_i$, denotados $\boldsymbol{\beta}_{\text{MV}} = (\boldsymbol{\beta}_{\text{MV},1}, \dots, \boldsymbol{\beta}_{\text{MV},p})$ y $\mathbf{\Lambda}_{\text{MV},i}$, respectivamente, verifican el siguiente sistema de ecuaciones:

$$\left\{ \begin{array}{rll} \widehat{\mathbf{\Lambda}}_{\text{MV},i} & = \text{diag}(\widehat{\boldsymbol{\beta}}_{\text{MV}}^{\text{T}} \widehat{\mathbf{\Sigma}}_i \widehat{\boldsymbol{\beta}}_{\text{MV}}) & \\ \widehat{\boldsymbol{\beta}}_{\text{MV},m}^{\text{T}} \displaystyle\sum_{i=1}^{k} n_i \frac{\widehat{\lambda}_{\text{MV},im} - \widehat{\lambda}_{\text{MV},ij}}{\widehat{\lambda}_{\text{MV},im} \widehat{\lambda}_{\text{MV},ij}} \widehat{\mathbf{\Sigma}}_i \, \widehat{\boldsymbol{\beta}}_{\text{MV},j} & = 0 & \text{para } \ 1 \le m \ne j \le p \\ \widehat{\boldsymbol{\beta}}_{\text{MV},j}^{\text{T}} \widehat{\boldsymbol{\beta}}_{\text{MV},m} & = \delta_{mj} & \end{array} \right. \tag{3.2.1}$$

donde $\delta_{mj} = 0$ si $m \ne j$ y $\delta_{mj} = 1$ si $m = j$. La verosimilitud $\Phi(\mathbf{B})$ es una medida de "diagonalización simultánea"de las k matrices de covarianza muestral $\widehat{\mathbf{\Sigma}}_i$ y por lo tanto, el método CPC puede verse como una rotación simultánea de los ejes que llevan a obtener variables que sean lo más no–correlacionadas sobre los k grupos. Más aún, el sistema (3.2.1) puede verse como un sistema generalizado de ecuaciones características. Si todas las matrices

$$\mathbf{A}_{mj} = \sum_{i=1}^{k} n_i \frac{\widehat{\lambda}_{\text{MV},im} - \widehat{\lambda}_{\text{MV},ij}}{\widehat{\lambda}_{\text{MV},im} \widehat{\lambda}_{\text{MV},ij}} \widehat{\mathbf{\Sigma}}_i$$

fuesen iguales, entonces (3.2.1) definiría los autovectores de dicha matriz. Por otra parte, el peso de $\widehat{\boldsymbol{\Sigma}}_i$ en $\mathbf{A}_{mj}$ será menor cuanto más parecidos sean $\widehat{\lambda}_{\text{MV},im}$ y $\widehat{\lambda}_{\text{MV},ij}$. En particular, si $\widehat{\lambda}_{im} = \widehat{\lambda}_{ij}$, o sea si las varianzas estimadas en las direcciones $m-$ésima y $j-$ésima son idénticas, entonces $\widehat{\boldsymbol{\Sigma}}_i$ desaparece de $\mathbf{A}_{mj}$. Esta condición corresponde a esfericidad de la población $i-$ésima en el plano generado por $\boldsymbol{\beta}_m$ y $\boldsymbol{\beta}_j$. En el caso extremo $\widehat{\boldsymbol{\Sigma}}_i = c\mathbf{I}_p$, la matriz $\widehat{\boldsymbol{\Sigma}}_i$ desaparece completamente de las ecuaciones (3.2.1), reflejando que $\mathbf{I}_p$ conmuta con cualquier matriz de $p \times p$. Este hecho se observa también en la expresión de la varianza asintótica de los estimadores de máxima verosimilitud de $\boldsymbol{\beta}$ cuya distribución asintótica fue obtenida por Flury (1986). Cuando las matrices $\boldsymbol{\Sigma}_i$ son diagonales, o sea, $\boldsymbol{\beta} = \mathbf{I}_p$, y si suponemos que $\lambda_{11} > \ldots > \lambda_{1p}$, la varianza asintótica de los elementos nodiagonales de $\widehat{\beta}_{\text{MV},m\ell}$, $m \neq \ell$, está dada por

$$\left[\sum_{i=1}^{k} \tau_i \frac{(\lambda_{im} - \lambda_{i\ell})^2}{\lambda_{im}\lambda_{i\ell}} \right]^{-1} ,$$

donde $\tau_i = n_i/N$, $N = \sum_{i=1}^{k} n_i$. Luego, si $\lambda_{im} = \lambda_{i\ell}$, la población $i-$ésima no contribuye a la varianza asintótica del estimador del autovector. Observemos que si en al menos una población el cociente entre dos autovalores consecutivos aumenta, la varianza asintótica decrece. Flury & Gautschi (1986) propusieron un algoritmo, llamado *algoritmo FG*, para resolver el sistema (3.2.1) buscando el mínimo de $\Phi(\mathbf{B})$ dadas las matrices $\widehat{\boldsymbol{\Sigma}}_i$ que puede verse como una generalización del método de Jacobi para diagonalizar matrices simétricas.

3.3. Estimadores robustos bajo un modelo de componentes principales comunes

En la Sección 3.2 construimos estimadores de las direcciones principales comunes bajo un modelo CPC que resultan óptimos bajo el supuesto de datos gaussianos. Dicha optimalidad se paga con la poca tolerancia de los estimadores encontrados a datos atípicos debido a la falta de robustez del estimador clásico de la matriz de covarianza. Por lo tanto, como en el Capítulo 2, estamos interesados en construir estimadores robustos de los autovectores comunes y sus autovalores asociados suponiendo que las matrices de escala $\boldsymbol{\Sigma}_i$ cumplen el modelo CPC afin de proveer estimadores robustos de las matrices de dispersión bajo el modelo CPC. En esta sección, vamos a describir dos métodos para lograr ese propósito. El primero, que es un método *plug–in* y que fue descripto en un caso particular por Novi Inverardi & Flury (1992) y estudiado por Boente & Orellana (2001), consiste en remplazar en el sistema (3.2.1) la matriz de covarianza muestral de la $i-$ésima población, $\widehat{\boldsymbol{\Sigma}}_i$, por un estimador robusto de la matriz de escala $\boldsymbol{\Sigma}_i$, para luego resolver el nuevo sistema y obtener así un estimador de $\boldsymbol{\beta}$ y $\boldsymbol{\Lambda}_i$. La segunda aproximación al problema consiste en un método de *projection–pursuit*, es decir, construye estimadores robustos para las direcciones principales comunes generalizando las técnicas estudiadas para el caso de una población en la sección 2.4. Este enfoque fue introducido por Boente & Orellana (2001) y estudiado por Boente, Pires & Rodrigues (2002).

3.3.1. Estimadores basados en una matriz de escala robusta: Enfoque plug–in

El enfoque *plug–in* (**PI**) propone estimadores de los ejes principales comunes y de sus autovalores basados en estimadores robustos de las matrices de escala. Como se mencionó anteriormente, una posible elección son los $M-$estimadores propuestos por Maronna (1976) o el estimador de elipsoide de mínimo volumen (Rousseeuw & Zomeren, 1990). Para cada elección de un estimador robusto de la matriz de covarianza, se obtiene un nuevo estimador robusto bajo CPC resolviendo el sistema (3.2.1) reemplazando la matriz de covarianza muestral $\widehat{\mathbf{\Sigma}}_i$ por un estimador robusto, $\mathbf{V}_i$, de la matriz de escala $\mathbf{\Sigma}_i$. Sea $\widehat{\boldsymbol{\beta}}$ la matriz ortogonal solución de este nuevo sistema, o sea, $\widehat{\boldsymbol{\beta}}$ es solución de

$$\widehat{\boldsymbol{\beta}}_m^{\mathrm{T}} \sum_{i=1}^{k} \frac{n_i(\widehat{\lambda}_{im} - \widehat{\lambda}_{ij})\mathbf{V}_i}{\widehat{\lambda}_{im}\widehat{\lambda}_{ij}} \widehat{\boldsymbol{\beta}}_j^{\mathrm{T}} = 0 \quad 1 \leq m \neq j \leq p \quad \widehat{\boldsymbol{\beta}}_j^{\mathrm{T}}\widehat{\boldsymbol{\beta}}_m = \delta_{mj} \tag{3.3.1}$$

y sean $\widehat{\mathbf{\Lambda}}_i = \mathrm{diag}(\widehat{\lambda}_{i1}, \ldots, \widehat{\lambda}_{ip})$, $1 \leq i \leq k$, los estimadores de las k matrices donde los estimadores de los autovalores vienen dados por

$$\widehat{\lambda}_{ij} = \widehat{\boldsymbol{\beta}}_j^{\mathrm{T}} \mathbf{V}_i \widehat{\boldsymbol{\beta}}_j \ . \tag{3.3.2}$$

Consideraremos estimadores de la matriz de escala $\mathbf{V}_i$ asintóticamente normales y esféricamente invariantes, es decir, en lo que sigue, $\mathbf{V}_i$ sera un estimador de escala robusto afín equivariante tal que

A.1. $\sqrt{n_i}(\mathbf{V}_i - \mathbf{\Sigma}_i) \longrightarrow \mathbf{Z}_i$ en distribución, donde $\mathrm{vec}(\mathbf{Z}_i)$ tiene distribución normal multivariada con media cero y matriz de covarianza $\mathbf{C}_i$

A.2. Para cualquier $\mathbf{G}$ tal que $\mathbf{G}\mathbf{G}^{\mathrm{T}} = \mathbf{\Sigma}_i^{-1}$ la distribución de $\mathbf{G}\mathbf{Z}_i\mathbf{G}^{\mathrm{T}}$ es invariante bajo transformaciones ortogonales.

Además, como $\mathbf{Z}_i$ tiene distribución normal y como el estimador $\mathbf{V}_i$ es asintóticamente insesgado, la distribución de $\mathbf{Z}_i$ se puede caracterizar por su segundo momento. Por ser $\mathbf{Z}_i$ esféricamente invariante, su segundo momento puede describirse mediante dos parámetros (ver Tyler, 1982, Teorema 1). Tenemos entonces que $\mathrm{vec}(\mathbf{Z}_i) \sim N(0, \mathbf{C}_i)$ donde:

$$\mathbf{C}_i \quad = \quad \sigma_{i1}(\mathbf{I} + \mathbf{K}_{pp})(\mathbf{\Sigma}_i \otimes \mathbf{\Sigma}_i) + \sigma_{i2}\mathrm{vec}(\mathbf{\Sigma}_i)\mathrm{vec}(\mathbf{\Sigma}_i)^{\mathrm{T}} \tag{3.3.3}$$

El siguiente teorema describe el comportamiento asintótico de los estimadores de las direcciones principales bajo el modelo CPC obtenidos con el método **PI**.

Teorema 3.3.1. *Sean $\mathbf{x}_{i1}, \ldots \mathbf{x}_{in_i}$, $1 \leq i \leq k$, observaciones independientes de k muestras independientes con parámetros de posición $\boldsymbol{\mu}_i$ y matrices de escala $\mathbf{\Sigma}_i$ que satisfacen el modelo (3.1.1) con $\boldsymbol{\beta} = \mathbf{I}_p$, o sea, $\mathbf{\Sigma}_i = \mathbf{\Lambda}_i = diag(\lambda_{i1}, \ldots, \lambda_{ip})$. Supongamos que $\lambda_{11} > \ldots > \lambda_{1p}$. Sea $n_i = \tau_i N$ con $0 < \tau_i < 1$ y $\sum_{i=1}^k \tau_i = 1$.*

*Sea $\mathbf{V}_i$ un estimador robusto de la matriz de escala $\mathbf{\Sigma}_i$, afín equivariante, que cumple **A.1** y **A.2** donde $\mathbf{C}_i$ puede descomponerse como en (3.3.3), y sean $\widehat{\lambda}_{ij}$ estimadores robustos consistentes de λ_{ij}, $1 \leq i \leq k$, $1 \leq j \leq p$.*

Sea $\widetilde{\boldsymbol{\beta}} = (\widetilde{\boldsymbol{\beta}}_1, \ldots, \widetilde{\boldsymbol{\beta}}_p) \in \mathcal{O}_p$ un estimador consistente de $\boldsymbol{\beta}$ solución de (3.3.1). Entonces, tenemos que

a) $\widetilde{f}_{jj} = \sqrt{N}(\widetilde{\boldsymbol{\beta}}_{jj} - 1) \xrightarrow{p} 0.$

b) $\widetilde{f}_{jm} = \sqrt{N}\widetilde{\boldsymbol{\beta}}_{jm}$ *para $j \leq m$ tiene distribución asintótica normal con media 0, correlaciones 0 y varianzas dadas por:*

$$\left[\sum_{i=1}^{k} \tau_i \sigma_{i1} \frac{(\lambda_{im} - \lambda_{ij})^2}{\lambda_{im}\lambda_{ij}}\right] \left[\sum_{i=1}^{k} \tau_i \frac{(\lambda_{im} - \lambda_{ij})^2}{\lambda_{im}\lambda_{ij}}\right]^{-2} .$$

Esta es también la distribución asintótica de $-\widetilde{f}_{jm}$.

El siguiente teorema describe el comportamiento asintótico de los estimadores **PI** de los autovalores de cada población.

Teorema 3.3.2. *Sean $\mathbf{x}_{i1},\ldots\mathbf{x}_{in_i}$, $1 \le i \le k$, observaciones independientes de k muestras independientes con parámetros de posición $\boldsymbol{\mu}_i$ y matrices de escala $\boldsymbol{\Sigma}_i$ que satisfacen el modelo* (3.1.1)*, o sea, $\boldsymbol{\Sigma}_i = \boldsymbol{\beta}\boldsymbol{\Lambda}_i\boldsymbol{\beta}^{\mathrm{T}}$ con $\boldsymbol{\Lambda}_i = diag(\lambda_{i1},\ldots,\lambda_{ip})$. Sea $n_i = \tau_i N$ con $0 < \tau_i < 1$ y $\sum_{i=1}^k \tau_i = 1$.*

*Sea $\mathbf{V}_i$ un estimador robusto y afín equivariante de la matriz de escala $\boldsymbol{\Sigma}_i$ que cumple **A.1**. Sea $\widehat{\boldsymbol{\Lambda}}_i$ definida por* (3.3.2) *donde $\widehat{\boldsymbol{\beta}} \in \mathcal{O}(p)$ y $\sqrt{n_i}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = O_p(1)$. Entonces, se tiene que*

$$\sqrt{n_i}\,vec\,(\widehat{\boldsymbol{\Lambda}}_i - \boldsymbol{\Lambda}_i) \xrightarrow{\mathcal{D}} N(0, \mathbf{W}_i)$$

donde $\xrightarrow{\mathcal{D}}$ indica la convergencia en distribución y $\mathbf{W}_i$ es la matriz de covarianza asintótica correspondiente a los elementos diagonales de $\sqrt{n_i}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{V}_i\boldsymbol{\beta} - \boldsymbol{\Lambda}_i)$.

En particular, si $\mathbf{C}_i$ cumple (3.3.3), tenemos que

$$\begin{aligned} \text{ASVAR}\left(\widehat{\lambda}_{ij}\right) &= (2\sigma_{i1} + \sigma_{i2})\lambda_{ij}^2\,, \quad 1 \le j \le p \\ \text{ASCOV}\left(\widehat{\lambda}_{ij}, \widehat{\lambda}_{im}\right) &= \sigma_{i2}\lambda_{ij}\lambda_{im} \quad \text{para } m \ne j\,. \end{aligned}$$

Como fue observado en Boente & Orellana (2001), si las k poblaciones tienen distribuciones elípticas que difieren sólo en sus matrices de escala y parámetros de posición y si se considera el mismo estimador de escala para cada población, la eficiencia asintótica será la misma para todas las poblaciones, o sea, $\sigma_{i1} = \sigma_1$ y $\sigma_{i2} = \sigma_2$, para todo i. Por lo tanto, la eficiencia de los estimadores de los autovalores obtenidos por (3.3.2) y de las direcciones comunes solución de (3.3.1) son las mismas que en el caso de una población.

3.3.2. Estimadores robustos bajo el modelo CPC: Enfoque de *projection–pursuit*

En la Sección 2.4 del Capítulo 2 definimos estimadores robustos de las componentes principales para el caso de una población utilizando técnicas de *projection–pursuit*. Con estas mismas ideas, Boente & Orellana (2001) definen estimadores robustos bajo el modelo CPC de los autovectores comunes y de los autovalores de cada población.

Sean $\mathbf{X}_i = (\mathbf{x}_{i1},\ldots,\mathbf{x}_{in_i})$, $1 \le i \le k$ observaciones independientes de la población i-ésima, $1 \le i \le k$ tales que $\mathbf{x}_{ij}$ tiene distribución F_i con parámetro de posición $\boldsymbol{\mu}_i$ y matriz de escala $\boldsymbol{\Sigma}_i$ que satisfacen el modelo CPC. Sin pérdida de generalidad, supondremos $\boldsymbol{\mu}_i = \mathbf{0}$.

En caso de existir segundo momento y si $\boldsymbol{\Sigma}_i$ es la matriz de covarianza de la población $i-$ésima se tiene que, para todo $\mathbf{a} \in I\!R^p$, $\text{var}(\mathbf{a}^{\mathrm{T}}\mathbf{x}_{i1}) = \mathbf{a}^{\mathrm{T}}\boldsymbol{\beta}\boldsymbol{\Lambda}_i\boldsymbol{\beta}^{\mathrm{T}}\mathbf{a}$, $1 \le i \le k$. Luego, el primer autovector común $\boldsymbol{\beta}_1$ puede obtenerse maximizando $\sum_{i=1}^k \tau_i \text{var}(\mathbf{a}^{\mathrm{T}}\mathbf{x}_{i1})$ sobre los vectores $\mathbf{a} \in I\!R^p$ con $\|\mathbf{a}\| = 1$, siendo $\boldsymbol{\beta}_1$ el autovector común asociado al máximo autovalor de $\sum_{i=1}^k \tau_i\boldsymbol{\Sigma}_i$. Si consideramos direcciones ortogonales a $\boldsymbol{\beta}_1$, el segundo autovector puede obtenerse de la misma forma y así podemos seguir con los demás. Notemos que los autovalores de la i-ésima población pueden ser escritos como $\boldsymbol{\beta}_j^{\mathrm{T}}\boldsymbol{\Sigma}_i\boldsymbol{\beta}_j$. Resulta entonces natural definir los estimadores **PP** bajo el modelo CPC de la siguiente manera.

Definición 3.3.1. Sean $\mathbf{X}_i = (\mathbf{x}_{i1},\ldots,\mathbf{x}_{in_i})$, $1 \le i \le k$ observaciones independientes de la población i-ésima, $1 \le i \le k$ tales que $\mathbf{x}_{ij}$ tiene distribución F_i con parámetro de posición $\boldsymbol{\mu}_i$ y matriz de escala $\boldsymbol{\Sigma}_i$ que satisfacen el modelo CPC. Dada $\mathbf{s}$ una escala univariada, siguiendo la notación de la sección

2.4, llamaremos $\mathbf{s}_{n_i}(\mathbf{a}^{\mathrm{T}}\mathbf{X}_i)$ al funcional $\mathbf{s}$ evaluado en la distribucion empírica de $\mathbf{a}^{\mathrm{T}}\mathbf{x}_{i1}, \ldots, \mathbf{a}^{\mathrm{T}}\mathbf{x}_{in_i}$. Sea $\tau_i = n_i/N$, $1 \leq i \leq k$, donde $N = \sum_{i=1}^{k} n_i$. Definimos los estimadores de *projection–pursuit* (**PP**) de las direcciones principales comunes, $\widehat{\boldsymbol{\beta}}_{j,\mathbf{s}}$ asociados al funcional de escala $\mathbf{s}$ (o al estimador de escala univariado $\mathbf{s}_n$) como

$$\widehat{\boldsymbol{\beta}}_{1,\mathbf{s}} = \underset{\|\mathbf{a}\|=1}{\operatorname{argmax}} \sum_{i=1}^{k} \tau_i \mathbf{s}_{n_i}^2(\mathbf{a}^{\mathrm{T}}\mathbf{X}_i) \tag{3.3.4}$$

$$\widehat{\boldsymbol{\beta}}_{j,\mathbf{s}} = \underset{\mathbf{a}\in\mathcal{C}_j}{\operatorname{argmax}} \sum_{i=1}^{k} \tau_i \mathbf{s}_{n_i}^2(\mathbf{a}^{\mathrm{T}}\mathbf{X}_i) \quad 1 \leq j \leq p \tag{3.3.5}$$

siendo $\mathcal{C}_j = \{\mathbf{a} \in I\!R^p \ / \ \|\mathbf{a}\| = 1, \mathbf{a}^{\mathrm{T}}\widehat{\boldsymbol{\beta}}_{\ell,\mathbf{s}} = 0 \quad \forall\, 1 \leq \ell \leq j-1 \ \}$. Los pesos τ_i se incluyen para adaptarse a los distintos tamaños de muestra. Los estimadores de *projection–pursuit* (**PP**) de los autovalores de la $i-$ésima población se definen como

$$\widehat{\lambda}_{ij,\mathbf{s}} = \mathbf{s}_{n_i}^2(\widehat{\boldsymbol{\beta}}_{j,\mathbf{s}}^{\mathrm{T}}\mathbf{X}_i)\ . \tag{3.3.6}$$

Se obtiene una solución diferente si, en cada paso, en lugar de maximizar se minimiza.

Los estimadores $\widehat{\boldsymbol{\beta}}_{j,\mathbf{s}}$ serán llamados estimadores robustos **PP** de las direcciones principales bajo el modelo CPC. Los ejes principales comunes pueden ser reordenados de manera tal que $\widehat{\lambda}_{11,\mathbf{s}} \geq \cdots \geq \widehat{\lambda}_{1p,\mathbf{s}}$.

Finalmente, el estimador robusto **PP** de la matriz de escala de la $i-$ésima población bajo el modelo CPC puede definirse como

$$\widehat{\mathbf{V}}_{i,\mathbf{s}} = \sum_{j=1}^{p} \widehat{\lambda}_{ij,\mathbf{s}} \widehat{\boldsymbol{\beta}}_{j,\mathbf{s}} \widehat{\boldsymbol{\beta}}_{j,\mathbf{s}}^{\mathrm{T}}\ . \tag{3.3.7}$$

3.3.3. Funcional Asociado a los estimadores de *projection–pursuit*

En esta sección definimos el funcional asociado a los estimadores definidos en la Definición 3.3.1 y damos condiciones para que sean Fisher–consistentes.

Definición 3.3.2. Sean $\mathbf{x}_1, \ldots, \mathbf{x}_k$, k vectores aleatorios independientes, tales que $\mathbf{x}_i \sim F_i$ donde F_i tiene parámetro de posición $\boldsymbol{\mu}_i$ y matriz de escala $\boldsymbol{\Sigma}_i$ que satisfacen el modelo CPC. Sea $F = F_1 \times \ldots \times F_k$ y $\mathbf{s}$ un funcional de escala robusto univariado. Denotemos por $F_i[\mathbf{a}]$ a la distribución de variable aleatoria unidimensional $\mathbf{a}^{\mathrm{T}}(\mathbf{x}_i - \boldsymbol{\mu}_i)$, donde $\mathbf{a} \in I\!R^p$. Sea

$$\rho(\mathbf{a}) = \sum_{i=1}^{k} \tau_i \mathbf{s}^2\left(F_i[\mathbf{a}]\right)\ .$$

Definimos las $\mathbf{s}-$direcciones principales comunes $\boldsymbol{\beta}_{j,\mathbf{s}} = \boldsymbol{\beta}_{j,\mathbf{s}}(F)$ y los respectivos $\mathbf{s}-$autovalores de cada población $\lambda_{ij,\mathbf{s}} = \lambda_{ij,\mathbf{s}}(F)$ como

$$\boldsymbol{\beta}_{1,\mathbf{s}} = \underset{\|\mathbf{a}\|=1}{\operatorname{argmax}} \sum_{i=1}^{k} \tau_i \mathbf{s}^2\left(F_i[\mathbf{a}]\right) = \underset{\|\mathbf{a}\|=1}{\operatorname{argmax}}\, \rho(\mathbf{a}) \tag{3.3.8}$$

$$\boldsymbol{\beta}_{j,\mathbf{s}} = \underset{\mathbf{a}\in\mathbf{C}_j}{\operatorname{argmax}} \sum_{i=1}^{k} \tau_i \mathbf{s}^2\left(F_i[\mathbf{a}]\right) = \underset{\mathbf{a}\in\mathbf{C}_j}{\operatorname{argmax}}\, \rho(\mathbf{a}) \quad 2 \leq j \leq p \tag{3.3.9}$$

$$\lambda_{ij,\mathbf{s}} = \mathbf{s}^2(F_i[\boldsymbol{\beta}_{j,\mathbf{s}}]) \quad 1 \leq j \leq p,\ \ 1 \leq i \leq k \tag{3.3.10}$$

donde $\mathbf{C}_j = \{\mathbf{a} \in I\!R^p : \ \|\mathbf{a}\| = 1, \ \ \mathbf{a}^{\mathrm{T}}\boldsymbol{\beta}_{\ell,\mathbf{s}} = 0 \ \forall \ 1 \leq \ell \leq j-1\}$, $0 < \tau_i < 1$ son números fijos tales que $\sum_{i=1}^{k} \tau_i = 1$.

También definimos la $\mathbf{s}-$matriz de escala como $\mathbf{V}_{i,\mathbf{s}} = \sum_{j=1}^{p} \lambda_{ij,\mathbf{s}} \boldsymbol{\beta}_{j,\mathbf{s}} \boldsymbol{\beta}_{j,\mathbf{s}}^{\mathrm{T}}$.

Proposición 3.3.1. *Sean $\mathbf{x}_i$, $1 \leq i \leq k$, vectores independientes con distribución elíptica F_i con parámetro de posición $\boldsymbol{\mu}_i$ y matrices escala $\boldsymbol{\Sigma}_i = \mathbf{C}_i\mathbf{C}_i^{\mathrm{T}} = \boldsymbol{\beta}\boldsymbol{\Lambda}_i\boldsymbol{\beta}^{\mathrm{T}}$, es decir, $\boldsymbol{\Sigma}_i$ cumplen el modelo* CPC. *Sea $\mathbf{s}$ un funcional robusto de escala, equivariante bajo transformaciones de escala y $F = F_1 \times \cdots \times F_k$. Supongamos que*

i) $\mathbf{x}_i = \boldsymbol{\mu}_i + \mathbf{C}_i \, \mathbf{z}_i$ *donde $\mathbf{z}_i$ tienen la misma distribución esférica G para todo $1 \leq i \leq k$.*

ii) $\mathbf{s}(G[\mathbf{e}_1]) = 1$.

Si $\boldsymbol{\Sigma} = \sum_{i=1}^{k} \tau_i \boldsymbol{\Sigma}_i$ no tiene autovalores múltiples, entonces , el funcional $\boldsymbol{\beta}_{\mathbf{s}}(F) = (\boldsymbol{\beta}_{1,\mathbf{s}}(F), \ldots, \boldsymbol{\beta}_{p,\mathbf{s}}(F))$ definido por (3.3.8) *y* (3.3.9) *y el funcional $\lambda_{ij,\mathbf{s}}(F)$ definido por* (3.3.10) *son Fisher–consistentes, es decir, $\boldsymbol{\beta}_{j,\mathbf{s}}(F) = \boldsymbol{\beta}_j$ y $\lambda_{ij,\mathbf{s}}(F) = \lambda_{ij}$ donde las direcciones comunes $\boldsymbol{\beta}_j$ de $\boldsymbol{\Sigma}_i$ se ordenan de modo que $\nu_1 > \ldots > \nu_p$ con $\nu_j = \sum_{i=1}^{k} \tau_i \lambda_{ij}$.*

Capítulo 4

Comportamiento asintótico de los estimadores de *projection–pursuit* para el modelo CPC.

En este capítulo, vamos a obtener las propiedades asintóticas para los estimadores definidos en (3.3.4), (3.3.5) y (3.3.6). Las demostraciones siguen las ideas utilizadas para probar los resultados dados por Cui, He & Ng (2003), presentados en la Sección 2.4.2. Comenzaremos recordando algunos objetos allí definidos:

- $H[\mathbf{a}]$ indica la distribución de la variable aleatoria $\mathbf{z}^{\mathrm{T}}\mathbf{a}$ cuando $\mathbf{z} \sim H$.
- $\mathbf{s}$ es un funcional de escala fijo, utilizado para construir los funcionales y los estimadores de *projection–pursuit* definidos en la Sección 3.3.2,
- $\mathbf{s}_n$ denota al estimador robusto de escala univariado asociado a $\mathbf{s}$, o sea, $\mathbf{s}_n(\mathbf{Y}) = \mathbf{s}(F_{n,\mathbf{Y}})$ donde $\mathbf{Y} = (y_1, \ldots, y_n)$ es una muestra de variables univariadas y $F_{n,\mathbf{Y}}$ es su distribución empírica.

En adelante, para facilitar la notación, cuando no se preste a confusión, omitiremos la dependencia de $\mathbf{s}$ en los estimadores de las $\mathbf{s}$ direcciones y tamaños dados en (3.3.5) y (3.3.6). Además, dado un funcional robusto de escala univariado $\mathbf{s}(\cdot) : \mathcal{F}_1 \to I\!R_{\geq 0}$, para cada $\mathbf{a} \in I\!R^p$ fijo definimos el funcional $\mathbf{s}_{\mathbf{a}} : \mathcal{F}_p \to I\!R_{\geq 0}$ haciendo $\mathbf{s}_{\mathbf{a}}(H) = \mathbf{s}(H[\mathbf{a}])$.

De ahora en más, supondremos que, para $1 \leq i \leq k$, $\mathbf{X}_i = (\mathbf{x}_{i1}, \ldots, \mathbf{x}_{in_i})$ son i.i.d con distribución F_i fija. Por simplicidad de notación, indicaremos por:

- $\varsigma_{n_i} : I\!R^p \to I\!R$ a la función $\varsigma_{n_i}(\mathbf{a}) = \mathbf{s}_{n_i}(\mathbf{a}^{\mathrm{T}}\mathbf{X}_i) = \mathbf{s}\left(F_{n_i,\mathbf{a}^{\mathrm{T}}\mathbf{X}_i}\right)$ donde $F_{n_i,\mathbf{a}^{\mathrm{T}}\mathbf{X}_i}$ es la distribución empírica de $\mathbf{a}^{\mathrm{T}}\mathbf{x}_{i1}, \ldots, \mathbf{a}^{\mathrm{T}}\mathbf{x}_{in_i}$ y $\mathbf{a}^{\mathrm{T}}\mathbf{X}_i = (\mathbf{a}^{\mathrm{T}}\mathbf{x}_{i1}, \ldots, \mathbf{a}^{\mathrm{T}}\mathbf{x}_{in_i})$.
- Análogamente, $\varsigma_i(\mathbf{a})$ indicará a $\varsigma_i(\mathbf{a}) = \mathbf{s}_{\mathbf{a}}(F_i) = \mathbf{s}(F_i[\mathbf{a}])$.
- $\dot{\varsigma}_i(\mathbf{a})$ denotará la derivada con respecto de $\mathbf{a}$ de la función $\varsigma_i : I\!R^p \to I\!R$
- $\dot{\varsigma}_{n_i}(\mathbf{a})$ indicará la derivada de $\varsigma_{n_i}(\mathbf{a})$ con respecto a $\mathbf{a}$, en caso de existir.

Para cada $1 \leq i \leq k$, llamaremos $h_i(\mathbf{x}, \mathbf{a})$ a la función de influencia del funcional $\mathbf{s}_{\mathbf{a}}$, definido en (2.4.4), evaluado en la distribución F_i, es decir,

$$h_i(\mathbf{x}, \mathbf{a}) = \mathrm{IF}(\mathbf{x}, \mathbf{s}_{\mathbf{a}}; F_i) = \mathbf{s}_{\mathbf{a}}(F_i)\, \mathrm{IF}(\mathbf{x}^{\mathrm{T}}\mathbf{a}/\mathbf{s}_{\mathbf{a}}(F_i), \mathbf{s}; F_{i,\mathbf{a}}) \;,$$

donde $F_{i,\mathbf{a}}$ es la distribución de $\mathbf{a}^{\mathrm{T}}\mathbf{x}/\mathbf{s_a}(F_i)$ cuando $\mathbf{x} \sim F_i$. Si bien sólo nos interesa el caso $\mathbf{a} \in \mathcal{S}_p$, consideraremos a $h_i(\mathbf{x}, \mathbf{a})$ como función de $\mathbf{a} \in I\!R^p$ para facilitar la diferenciación, como en la Sección 2.4.2.

Recordemos que los estimadores de definidos en (3.3.4) se obtienen maximizando, sobre $\|\mathbf{a}\| = 1$, la función

$$\rho_N(\mathbf{a}) = \sum_{i=1}^{k} \tau_i \, \varsigma_{n_i}^2(\mathbf{a}) \ , \tag{4.0.1}$$

donde $N = \sum_{i=1}^{k} n_i$, $\tau_i = n_i/N$, mientras que los funcionales dados en 3.3.8 hacen lo propio con

$$\rho(\mathbf{a}) = \sum_{i=1}^{k} \tau_i \varsigma_i^2(\mathbf{a}) \ . \tag{4.0.2}$$

Definamos $\widehat{\nu}_m$ y $\nu_{m,\mathbf{s}}$ como

$$\widehat{\nu}_m \quad = \quad \text{máx}_{\mathbf{a} \in \mathcal{C}_m} \, \rho_N(\mathbf{a}) \tag{4.0.3}$$

$$\nu_{m,\mathbf{s}} \quad = \quad \text{máx}_{\mathbf{a} \in \mathbf{C}_m} \, \rho(\mathbf{a}) \tag{4.0.4}$$

donde $\mathcal{C}_m = \{\mathbf{a} \in I\!R^p \ / \ \|\mathbf{a}\| = 1, \ \ \mathbf{a}^{\mathrm{T}}\widehat{\boldsymbol{\beta}}_i = 0 \ \forall \ i = 1, 2, \ldots, m-1\}$ y $\mathbf{C}_m = \{\mathbf{a} \in I\!R^p \ / \ \|\mathbf{a}\| = 1, \ \ \mathbf{a}^{\mathrm{T}}\boldsymbol{\beta}_{i,\mathbf{s}} = 0 \ \forall \ i = 1, 2, \ldots, m-1\}$, respectivamente.

A continuación, enunciaremos una serie de supuestos bajo los cuales estudiaremos el comportamiento asintóticos de los estimadores **PP** de los parámetros del modelo CPC definidos en el capítulo anterior. Como antes, conciernen al funcional de dispersión tanto en las distribuciones que generan los datos como en diferentes empíricas.

S1: La función $h_i(\mathbf{x}, \mathbf{a})$ es continua en ambas variables, o sea $\mathrm{IF}(u, \mathbf{s}; G)$ es continua. Además para cada $\mathbf{x}$ fijo, es diferenciable a trozos, en el siguiente sentido: existe un entero m_i y una partición $\{\mathcal{R}_j^{(i)}(\mathbf{x}), \ j = 1, \ldots, m_i\}$ de $I\!R^p$ tal que $h_i(\mathbf{x}, \mathbf{a})$ es diferenciable en $\mathbf{a}$ en el interior de cada $\mathcal{R}_j^{(i)}(\mathbf{x})$. Por último se cumple

$$\mathrm{E}\left(\sup_{\mathbf{a} \in \mathcal{S}_p} h_i^2(\mathbf{x}_{i1}, \mathbf{a})\right) < \infty.$$

S2: El siguiente desarrollo $\varsigma_{n_i}(\mathbf{a}) - \varsigma_i(\mathbf{a}) = n_i^{-1} \sum_{j=1}^{n_i} h_i(\mathbf{x}_{ij}, \mathbf{a}) + o_p\left(n_i^{-\frac{1}{2}}\right)$ vale uniformemente en $\mathbf{a} \in \mathcal{S}_p$, con $\mathrm{E}h_i(\mathbf{x}_i, \mathbf{a}) = 0$ y $\mathrm{E}\left(\sup_{\mathbf{a} \in \mathcal{S}_p} |h_i(\mathbf{x}_i, \mathbf{a})|\right) < \infty$.

S3: Supondremos que $\dot{\varsigma}_i(\mathbf{a})$ es diferenciable con derivada continua $\ddot{\varsigma}_i(\mathbf{a})$.

S4: La derivada $\dot{\varsigma}_{n_i}(\mathbf{a})$ de $\varsigma_{n_i}(\mathbf{a})$ con respecto a $\mathbf{a}$ existe en casi todo punto $\mathbf{a} \in \mathcal{S}_p$. Más aún,

$$\dot{\varsigma}_{n_i}(\mathbf{a}) - \dot{\varsigma}_i(\mathbf{a}) = n_i^{-1} \sum_{j=1}^{n_i} h_i^{\star}(\mathbf{x}_{ij}, \mathbf{a}) + o_p\left(n_i^{-\frac{1}{2}}\right)$$

uniformemente en $\mathbf{a} \in \mathcal{S}_p$ con $h_i^{\star}$ tal que

i) $\mathrm{E}h_i^{\star}(\mathbf{x}_i, \mathbf{a}) = 0$ para todo $\mathbf{a} \in I\!R^p$

ii) $\mathrm{E}\left(\sup_{\mathbf{a}\in\mathcal{S}_p}|h_i^\star(\mathbf{x}_i,\mathbf{a})|\right)<\infty.$

S5: Para cada $\mathbf{x}$ fijo, la función $h_i^\star(\mathbf{x},\mathbf{a})$ es continua en $\mathbf{a}$ y diferenciable a trozos, en el siguiente sentido: existe un entero ℓ_i y una partición $\{\mathcal{T}_j^{(i)}(\mathbf{x}),\ j=1,\ldots,\ell_i\}$ de $I\!R^p$ tal que $h_i^\star(\mathbf{x},\mathbf{a})$ es diferenciable en $\mathbf{a}$ en el interior de cada $\mathcal{T}_j^{(i)}(\mathbf{x})$.

S6: La familia de funciones $\mathcal{H}_i=\{\,f(\mathbf{x})=h_i(\mathbf{x},\mathbf{a})\ \mathbf{a}\in\mathcal{S}_p\}\,,1\le i\le k,$ es acotada y tiene número de cubrimiento [1] $N(\,\varepsilon,\,\mathcal{H}_i,\,L^2(Q))\le A\,\varepsilon^{-N}\quad 0<\varepsilon<1$ para toda probabilidad Q.

La condición **S2** es una representación de Bahadur uniforme para los estimadores de escala y, como mencionamos, se verifica para la mayoría de los $M-$estimadores de escala. En la mayoría de las situaciones $h_i^\star(\mathbf{x},\mathbf{a})=(\partial/\partial\mathbf{a})h_i(\mathbf{x},\mathbf{a})=\dot{h}_i(\mathbf{x},\mathbf{a})$, pero no se impone esa restricción para facilitar la verificación de **S4**.

El siguiente Teorema muestra que si el funcional asociado es Fisher–consistente, los estimadores $\widehat{\boldsymbol{\beta}}_m$ y $\widehat{\lambda}_{im}$ de las direcciones comunes y los tamaños de cada población definidos en (3.3.5) y en (3.3.6) son consistentes.

Teorema 4.0.3. *Sean* $\mathbf{X}_i=(\mathbf{x}_{i1},\ldots,\mathbf{x}_{in_i})$ *independientes e idénticamente distribuídas,* $1\le i\le k$, *tales que* $\mathbf{x}_{ij}\sim F_i$. *Supongamos que las matrices de escala* $\boldsymbol{\Sigma}_i$ *de* F_i *cumplen el modelo* CPC, *o sea,* $\boldsymbol{\Sigma}_i=\boldsymbol{\beta}\boldsymbol{\Lambda}_i\boldsymbol{\beta}^{\mathrm{T}}$. *Sean* $\boldsymbol{\beta}_{m,\mathbf{s}}=\boldsymbol{\beta}_{m,\mathbf{s}}(F)$ *y* $\lambda_{im,\mathbf{s}}=\lambda_{im,\mathbf{s}}(F)$ *los funcionales definidos en* (3.3.9) *y* (3.3.10). *Sean* $\widehat{\boldsymbol{\beta}}_m$ *y* $\widehat{\lambda}_{im}$ *los estimadores asociados definidos en* (3.3.5) *y* (3.3.6)*, respectivamente. Supongamos que* $\nu_{1,\mathbf{s}}>\nu_{2,\mathbf{s}}>\ldots>\nu_{q,\mathbf{s}}$ *para algún* $q\le p$ *y que* $\boldsymbol{\beta}_{m,\mathbf{s}}$ *son únicos salvo cambios de signo. Bajo las condiciones **S1** y **S2**, se cumple que para* $1\le m\le q$, $\widehat{\boldsymbol{\beta}}_m\xrightarrow{p}\boldsymbol{\beta}_{m,\mathbf{s}}$ *y* $\widehat{\lambda}_{im}\xrightarrow{p}\lambda_{im,\mathbf{s}}$*, para* $1\le i\le k$.

Recordemos que que si las distribuciones F_i verifican las condiciones de la Proposición 3.3.1, $\boldsymbol{\beta}_{m,\mathbf{s}}$ resultan ser los autovectores de las matrices de escala $\boldsymbol{\Sigma}_i$ y por lo tanto, los estimadores son consistentes para los autovectores comunes de las matrices de dispersión de las k-poblaciones, como se deseaba.

Para obtener una expresión para la distribución asintótica, definamos

- $\mathbf{g}_i(\mathbf{x},\boldsymbol{\beta})=2\varsigma_i(\boldsymbol{\beta})\,h_i^\star(\mathbf{x},\boldsymbol{\beta})+2\dot{\varsigma}_i(\boldsymbol{\beta})\,h_i(\mathbf{x},\boldsymbol{\beta})\quad 1\le i\le k$
- $\mathbf{u}_m=-N^{-\frac{1}{2}}\sum_{i=1}^k\sum_{j=1}^{n_i}\mathbf{g}_i(\mathbf{x}_{ij},\boldsymbol{\beta}_{m,\mathbf{s}})$
- $\mathbf{P}_{m+1}=\mathbf{I}_p-\sum_{j=1}^m\boldsymbol{\beta}_{j,\mathbf{s}}\,\boldsymbol{\beta}_{j,\mathbf{s}}^{\mathrm{T}}$ la matriz de proyección sobre el subespacio $<\boldsymbol{\beta}_{1,\mathbf{s}},\ldots,\boldsymbol{\beta}_{m,\mathbf{s}}>^{\perp}$,
- $\mathbf{B}_{jm}=\boldsymbol{\beta}_{j,\mathbf{s}}^{\mathrm{T}}\,\dot{\rho}(\boldsymbol{\beta}_{m,\mathbf{s}})\,\mathbf{I}_p+\boldsymbol{\beta}_{j,\mathbf{s}}\,\dot{\rho}(\boldsymbol{\beta}_{m,\mathbf{s}})^{\mathrm{T}}$,
- $\mathbf{A}_m=\mathbf{P}_{m+1}\,\ddot{\rho}(\boldsymbol{\beta}_{m,\mathbf{s}})-\boldsymbol{\beta}_{m,\mathbf{s}}^{\mathrm{T}}\,\dot{\rho}(\boldsymbol{\beta}_{m,\mathbf{s}})\mathbf{I}_p-\sum_{j=1}^{m-1}\boldsymbol{\beta}_{j,\mathbf{s}}^{\mathrm{T}}\,\dot{\rho}(\boldsymbol{\beta}_{m,\mathbf{s}})\boldsymbol{\beta}_{m,\mathbf{s}}\,\boldsymbol{\beta}_{j,\mathbf{s}}^{\mathrm{T}}$.
- En forma recurrente, podemos definir para $1\le m\le q$, $\mathbf{Z}_m=\sum_{j=0}^{m-1}\mathbf{A}_m^{-1}\mathbf{B}_{jm}\mathbf{Z}_j+\mathbf{A}_m^{-1}\mathbf{P}_{m+1}\mathbf{u}_m$, donde $\mathbf{Z}_0=0$ y suponemos la existencia de $\mathbf{A}_m^{-1}$.

  Claramente, $\mathbf{Z}_m$ puede ser representado en la forma $\mathbf{Z}_m=\sum_{j=0}^{m-1}\mathbf{C}_{jm}\mathbf{u}_j$, donde las matrices $\mathbf{C}_{jm}$ están determinadas por $\mathbf{A}_m,\ \mathbf{B}_{jm}\ y\ \mathbf{P}_{m+1}$.
- $\boldsymbol{\xi}_{i,m}(\mathbf{x})=\sum_{\ell=1}^m\mathbf{C}_{\ell m}\,\mathbf{g}_i(\mathbf{x},\boldsymbol{\beta}_{\ell,\mathbf{s}})$
- $\eta_{i,m}(\mathbf{x})=2\sqrt{\lambda_{im}}\,\left\{h_i(\mathbf{x},\boldsymbol{\beta}_{m,\mathbf{s}})+\sum_{j=1}^{m-1}\boldsymbol{\beta}_{j,\mathbf{s}}^{\mathrm{T}}\,\dot{\varsigma}_i(\boldsymbol{\beta}_{m,\mathbf{s}})\,\boldsymbol{\beta}_{j,\mathbf{s}}^{\mathrm{T}}\,\left(\sum_{i=1}^k\boldsymbol{\xi}_{i,m}(\mathbf{x})\right)\right\}$

[1] La definición del número de cubrimiento puede verse en Van der Vaart & Wellner (1996).

Teorema 4.0.4. *Bajo las condiciones del Teorema 4.0.3, si además también se verifican las condiciones* ***S3*** *,* ***S4****,* ***S5*** *y* ***S6****, si las matrices* $\mathbf{A}_m$*,* $1 \leq m \leq q$*, son no singulares, tenemos que*

$$\begin{aligned} \widehat{\boldsymbol{\beta}}_m - \boldsymbol{\beta}_{m,\mathbf{s}} &= \frac{1}{N} \sum_{i=1}^{k} \sum_{j=1}^{n_i} \boldsymbol{\xi}_{i,m}(\mathbf{x}_{ij}) + o_p\,(N^{-\frac{1}{2}}) \\ \widehat{\lambda}_{im} - \lambda_{im,\mathbf{s}} &= \frac{1}{n_i} \sum_{j=1}^{n_i} \eta_{i,m}(\mathbf{x}_{ij}) + o_p\,(n_i^{-\frac{1}{2}}) \end{aligned}$$

Como consecuencia del Teorema 4.0.4, la distribución conjunta de

$$N^{-\frac{1}{2}}(\widehat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_{1,\mathbf{s}}, \ldots, \widehat{\boldsymbol{\beta}}_q - \boldsymbol{\beta}_{q,\mathbf{s}}, \widehat{\lambda}_1 - \lambda_{1,\mathbf{s}}, \ldots, \widehat{\lambda}_q - \lambda_{q,\mathbf{s}})$$

converge a una distribución normal multivariada con media 0 y matriz de covarianza dada por $\mathrm{COV}_F\,(\xi_1(\mathbf{X}), \ldots, \xi_q(\mathbf{X}) \ldots, \eta_1(\mathbf{X}), \ldots, \eta_q(\mathbf{X})\ )$.

4.1. Consistencia

En la demostración que haremos del Teorema 4.0.3 probamos también la consistencia de $\widehat{\nu}_m$. Si bien estos estimadores no son de interés estadístico, serán utilizados para demostrar la consistencia de los ejes principales y sus correspondientes tamaños. El siguiente resultado es de crucial importancia en esta Sección.

Lema 4.1.1. *Sea* f_n *una sucesión de funciones aleatorias.* $f_n : \mathcal{S}_p \longrightarrow I\!R$ *y sea* $f : \mathcal{S}_p \longrightarrow I\!R$ *una función tal que*

$$\sup_{\mathbf{a} \in \mathcal{S}_p} |f_n(\mathbf{a}) - f(\mathbf{a})| = \mathbf{o}_p(1).$$

Si llamamos

$$\begin{aligned} \mathbf{w}_n &= \arg \max_{\mathbf{a} \in \mathcal{S}_p} f_n(\mathbf{a}) \\ \mathbf{w} &= \arg \max_{\mathbf{a} \in \mathcal{S}_p} f(\mathbf{a}), \end{aligned}$$

donde, tanto f_n *como* f *son continuas y* $\mathbf{w}$ *es el único máximo de* f *en* $\mathcal{S}_p$ *entonces:*

$$\mathbf{w}_n - \mathbf{w} = \mathbf{o}_p(1).$$

Demostración. Tenemos que

$$|f(\mathbf{w}_n) - f(\mathbf{w})| \leq |f(\mathbf{w}_n) - f_n(\mathbf{w}_n)| + |f_n(\mathbf{w}_n) - f(\mathbf{w})| \leq \sup_{\mathbf{a} \in \mathcal{S}_p} |f_n(\mathbf{a}) - f(\mathbf{a})| + |f_n(\mathbf{w}_n) - f(\mathbf{w})|$$

Ahora bien, supongamos que $|f_n(\mathbf{w}_n) - f(\mathbf{w})| = f_n(\mathbf{w}_n) - f(\mathbf{w})$ como $f(\mathbf{w}_n) \leq f(\mathbf{w})$ luego

$$|f_n(\mathbf{w}_n) - f(\mathbf{w})| = f_n(\mathbf{w}_n) - f(\mathbf{w}) \leq f_n(\mathbf{w}_n) - f(\mathbf{w}_n) \leq \sup_{\mathbf{a} \in \mathcal{S}_p} |f_n(\mathbf{a}) - f(\mathbf{a})| = \mathbf{o}_p(1).$$

Si por el contrario, $|f_n(\mathbf{w}_n) - f(\mathbf{w})| = f(\mathbf{w}) - f_n(\mathbf{w}_n)$ vale también que $|f_n(\mathbf{w}_n) - f(\mathbf{w})| \leq \sup_{\mathbf{a} \in \mathcal{S}_p} |f_n(\mathbf{a}) - f(\mathbf{a})| = \mathbf{o}_p(1)$, con lo cual $|f(\mathbf{w}_n) - f(\mathbf{w})| = \mathbf{o}_p(1)$.

Veamos que si $f(\mathbf{w}_n) - f(\mathbf{w}) = \mathbf{o}_p(1)$, entonces $\mathbf{w}_n - \mathbf{w} = \mathbf{o}_p(1)$. Para ello, notemos que la sucesión de vectores aleatorios $(\mathbf{w}_n)_{n \geq 1}$ verifica

1) Rigidez: toda subsucesión de $(\mathbf{w}_n)_{n\geq 1}$ tiene una sub-subsucesión débilmente convergente. Esto es debido a que $(\mathbf{w}_n)_{n\geq 1}$ es una sucesión de vectores aleatorios tomando valores en la bola p-dimensional $\mathcal{S}_p$, compacta.

2) Siendo f continua y $\mathbf{w}$ su único máximo, la sucesión $(\mathbf{w}_n)_{n\geq 1}$ tiene a $\mathbf{w}$ como único punto de acumulación. En efecto, sea $(\mathbf{w}_{n_k})_{k\geq 1}$ una subsucesiones de $(\mathbf{w}_n)_{n\geq 1}$ débilmente convergente: $\mathbf{w}_{n_k} \to \alpha$, siendo α una variables vector tomando valores en la bola $\mathcal{S}_p$. Como f es continua, resulta que $f(\mathbf{w}_{n_k}) \to f(\alpha)$. Por otra parte, tenemos que $f(\mathbf{w}_{n_k}) \to f(\mathbf{w})$, de donde concluimos que $P(f(\alpha) = f(\mathbf{w})) = 1$. Finalmente, siendo $\mathbf{w}$ el único máximo de la función f, tenemos que $P(\alpha = \mathbf{w}) = 1$, garantizando así que $\mathbf{w}$ es el único punto de acumulación.

De 1) y 2) obtenemos que $\mathbf{w}_n - \mathbf{w} = \mathbf{o}_p(1)$. □

Observación 4.1.1. *Más generalmente, hemos demostrado que si f es continua en un compacto $\mathcal{K}$ con único máximo en $\mathbf{w}$ y $(\mathbf{w}_n)_{n\geq 1}$ es una sucesión de elementos aleatorios tomando valores en $\mathcal{K}$ de forma tal que $f(\mathbf{w}_n) - f(\mathbf{w}) = \mathbf{o}_p(1)$, entonces $\mathbf{w}_n - \mathbf{w} = \mathbf{o}_p(1)$.*

Para probar la consistencia de los estimadores, vamos a necesitar una expresión para $\rho_N(\mathbf{a}) - \rho(\mathbf{a})$ análoga a la dada por el supuesto **S2**.

Proposición 4.1.1. *Bajo la condición **S2**, vale que*

$$\rho_N(\mathbf{a}) - \rho(\mathbf{a}) = N^{-1}\sum_{i=1}^{k} 2\varsigma_i(\mathbf{a})\sum_{j=1}^{n_i} h_i(\mathbf{x}_{ij};\mathbf{a}) + \mathbf{o}_p(1)\,,$$

*uniformemente en $\mathbf{a}$. Más aún, bajo **S6** tenemos que*

$$\rho_N(\mathbf{a}) - \rho(\mathbf{a}) = N^{-1}\sum_{i=1}^{k} 2\varsigma_i(\mathbf{a})\sum_{j=1}^{n_i} h_i(\mathbf{x}_{ij};\mathbf{a}) + \mathbf{o}_p(N^{-\frac{1}{2}})\,,$$

uniformemente en $\mathbf{a}$.

Demostración. Recordando las definiciones para $\rho_N(\mathbf{a})$ y $\rho(\mathbf{a})$ dadas en (4.0.1) y (4.0.2), respectivamente, basta estudiar el comportamiento de $\varsigma^2_{n_i}(\mathbf{a}) - \varsigma^2_i(\mathbf{a})$ para cada $i = 1,\ldots,k$, valiéndonos de la condición **S2**

$$\varsigma_{n_i}(\mathbf{a}) - \varsigma_i(\mathbf{a}) = n_i^{-1}\sum_{j=1}^{n_i} h_i(\mathbf{x}_{ij},\mathbf{a}) + \mathbf{o}_p(n_i^{-\frac{1}{2}})$$

uniformemente en $\mathbf{a} \in \mathcal{S}_p$. Notemos que

$$\varsigma^2_{n_i}(\mathbf{a}) - \varsigma^2_i(\mathbf{a}) \;=\; \{\, \varsigma_{n_i}(\mathbf{a}) - \varsigma_i(\mathbf{a}) \,\}\,\{\, \varsigma_{n_i}(\mathbf{a}) + \varsigma_i(\mathbf{a}) \,\}. \tag{4.1.1}$$

Además, para cada $1 \leq i \leq k$, $h_i(\mathbf{x},\mathbf{a})$ es uniformemente continua en $\mathbf{a} \in \mathcal{S}_p$ compacto, lo que garantiza usando que $\mathrm{E}\, h_i(\mathbf{x}_{i1},\mathbf{a}) = 0$ y el Teorema 2 de Pollard que

$$\sup_{\mathbf{a}\in\mathcal{S}_p} \frac{1}{n_i}\Big|\sum_{j=1}^{n_i} h_i(\mathbf{x}_{ij},\mathbf{a})\Big| = \mathbf{o}_p^{(i)}(1).$$

Por lo tanto, usando la condición **S2** resulta que

$$\varsigma_{n_i}(\mathbf{a}) - \varsigma_i(\mathbf{a}) \;=\; \mathbf{o}_p^{(i)}(1) + \mathbf{o}_p(n_i^{-\frac{1}{2}}) = \mathbf{o}_p^{(i)}(1), \tag{4.1.2}$$

uniformemente en $\mathbf{a} \in \mathcal{S}_p$. Más aún, bajo la condición **S6**

$$\varsigma_{n_i}(\mathbf{a}) - \varsigma_i(\mathbf{a}) = O_p(n_i^{-\frac{1}{2}})$$

uniformemente en $\mathbf{a} \in \mathcal{S}_p$, de donde

$$\varsigma_{n_i}(\mathbf{a}) + \varsigma_i(\mathbf{a}) = 2\varsigma_i(\mathbf{a}) + O_p(n_i^{-\frac{1}{2}})$$

uniformemente en $\mathbf{a} \in \mathcal{S}_p$. Reemplazando la última expresión en (4.1.1), nos queda

$$\varsigma_{n_i}^2(\mathbf{a}) - \varsigma_i^2(\mathbf{a}) \;=\; 2\varsigma_i(\mathbf{a})\{\, \varsigma_{n_i}(\mathbf{a}) - \varsigma_i(\mathbf{a}) \,\} + \mathbf{o}_p(n_i^{-\frac{1}{2}}) \;=\; 2\varsigma_i(\mathbf{a})\, n_i^{-1} \sum_{j=1}^{n_i} h_i(\mathbf{x}_{ij}; \mathbf{a}) + \mathbf{o}_p(n_i^{-\frac{1}{2}})\ . \quad (4.1.3)$$

Multiplicando el miembro izquierdo y derecho de (4.1.3) por τ_i y sumando en $i = 1, \ldots, k$, obtenemos el resultado buscado

$$\rho_N(\mathbf{a}) - \rho(\mathbf{a}) \;=\; \sum_{i=1}^{k} \tau_i \varsigma_{n_i}^2(\mathbf{a}) - \sum_{i=1}^{k} \tau_i \varsigma_i^2(\mathbf{a}) \;=\; \sum_{i=1}^{k} 2\varsigma_i(\mathbf{a})\, \tau_i\, n_i^{-1} \sum_{j=1}^{n_i} h_i(\mathbf{x}_{ij}; \mathbf{a}) + \mathbf{o}_p(N^{-\frac{1}{2}})\ .\square$$

Demostración del teorema 4.0.3.

Etapa 1. Consistencia de $\widehat{\boldsymbol{\beta}}_1$

Recordemos que $\widehat{\boldsymbol{\beta}}_1 = \mathrm{argmax}_{\mathbf{a} \in \mathcal{S}_p}\, \rho_N(\mathbf{a})$ y $\boldsymbol{\beta}_{1,\mathbf{s}} = \mathrm{argmax}_{\mathbf{a} \in \mathcal{S}_p}\, \rho(\mathbf{a})$. Para poder aplicar el Lema 4.1.1, sea $f_N = \rho_N$ y $f = \rho$. Veamos que

$$\sup_{\mathbf{a} \in \mathcal{S}_p} |\rho_N(\mathbf{a}) - \rho(\mathbf{a})| = \mathbf{o}_p(1)\ . \quad (4.1.4)$$

A tal efecto, por la Proposición 4.1.1, tenemos que

$$\rho_N(\mathbf{a}) - \rho(\mathbf{a}) = \sum_{i=1}^{k} \tau_i 2\varsigma_i(\mathbf{a}) n_i^{-1} \sum_{j=1}^{n_i} h_i(\mathbf{x}_{ij}; \mathbf{a}) + \mathbf{o}_p(N^{-\frac{1}{2}})\ ,$$

uniformemente en $\mathbf{a}$. Además, vimos que para cada $1 \leq i \leq k$

$$\sup_{\mathbf{a} \in \mathcal{S}_p} \frac{1}{n_i} \Big| \sum_{j=1}^{n_i} h_i(\mathbf{x}_{ij}, \mathbf{a}) \Big| = \mathbf{o}_p^{(i)}(1)\ .$$

Por último, recordando que $\varsigma_i(\mathbf{a})$ es una función continua en la bola $\mathcal{S}_p$, tenemos que

$$\sup_{\mathbf{a} \in \mathcal{S}_p} |\rho_N(\mathbf{a}) - \rho(\mathbf{a})| \;\leq\; \sum_{i=1}^{k} 2\tau_i\, c_i \sum_{j=1}^{n_i} \mathbf{o}_p^{(i)}(1) + \mathbf{o}_p(N^{-\frac{1}{2}}) = \mathbf{o}_p(1)\ ,$$

donde $c_i = \mathrm{máx}_{\mathbf{a} \in \mathcal{S}_p} \varsigma_i(\mathbf{a})$, con lo cual $\sup\limits_{\mathbf{a} \in \mathcal{S}_p} |\rho_N(\mathbf{a}) - \rho(\mathbf{a})| = \mathbf{o}_p(1)$, lo que muestra (4.1.4). Como $\rho_N(\mathbf{a})$ y $\rho(\mathbf{a})$ son continuas y $\boldsymbol{\beta}_{1,\mathbf{s}}$ es el único máximo de $\rho(\mathbf{a})$ en $\mathcal{S}_p$ entonces, por el Lema 4.1.1

$$\widehat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_{1,\mathbf{s}} = \mathbf{o}_p(1)\ ,$$

tal como queríamos demostrar.

Observación 4.1.2. *Notemos que gracias a* (4.1.4) *obtenemos la consistencia de* $\widehat{\nu}_1$ *ya que*

$$|\widehat{\nu}_1 - \nu_{1,\mathbf{s}}| = |\max_{\mathbf{a}\in\mathcal{S}_p} \rho_N(\mathbf{a}) - \max_{\mathbf{a}\in\mathcal{S}_p} \rho(\mathbf{a})| \le \max_{\mathbf{a}\in\mathcal{S}_p} |\rho_N(\mathbf{a}) - \rho(\mathbf{a})| = \mathbf{o}_p(1).$$

Observación 4.1.3. *Como* ρ_N *y* $\rho(\mathbf{a})$ *son funciones continuas en* $\mathcal{S}_p$ *podemos usar indistintamente* $\sup_{\mathbf{a}\in\mathcal{S}_p} |\rho_N(\mathbf{a}) - \rho(\mathbf{a})|$ *o* $\max_{\mathbf{a}\in\mathcal{S}_p} |\rho_N(\mathbf{a}) - \rho(\mathbf{a})|$.

Etapa 2. Consistencia de $\widehat{\nu}_2$ Definamos $f_1(\mathbf{v}) = \max_{\mathbf{a}\in\mathcal{S}_p,\, \mathbf{a}\perp\mathbf{v}} \rho(\mathbf{a})$. Por la continuidad de $\rho(\mathbf{a})$ resulta que $f_1(\mathbf{v})$ es continua en la variable $\mathbf{v}$. Además,

$$\begin{aligned} f_1(\boldsymbol{\beta}_1) &= \nu_{2,\mathbf{s}} \\ f_1(\widehat{\boldsymbol{\beta}}_1) &= \max_{\mathbf{a}\in\mathcal{S}_p,\, \mathbf{a}\perp\widehat{\beta}_1} \rho(\mathbf{a}). \end{aligned}$$

Luego

$$|\widehat{\nu}_2 - \nu_{2,\mathbf{s}}| = |\max_{\mathbf{a}\perp\widehat{\beta}_1, \mathbf{a}\in\mathcal{S}_p} \rho_N(\mathbf{a}) - \max_{\mathbf{a}\perp\beta_{1,\mathbf{s}}, \in\mathcal{S}_p} \rho(\mathbf{a})| = |\max_{\mathbf{a}\perp\widehat{\beta}_1, \mathbf{a}\in\mathcal{S}_p} \rho_N(\mathbf{a}) - f_1(\widehat{\boldsymbol{\beta}}_1) + f_1(\widehat{\boldsymbol{\beta}}_1) - f_1(\boldsymbol{\beta}_{1,\mathbf{s}})| \le$$

$$\max_{\mathbf{a}\in\mathcal{S}_p,\, \mathbf{a}\perp\widehat{\beta}_1} |\rho_N(\mathbf{a}) - \rho(\mathbf{a})| + |f_1(\widehat{\boldsymbol{\beta}}_1) - f_1(\boldsymbol{\beta}_{1,\mathbf{s}})| \le \max_{\mathbf{a}\in\mathcal{S}_p} |\rho_N(\mathbf{a}) - \rho(\mathbf{a})| + |f_1(\widehat{\boldsymbol{\beta}}_1) - f_1(\boldsymbol{\beta}_{1,\mathbf{s}})| = \mathbf{o}_p(1) \ ,$$

ya que ρ_N converge uniformemente a ρ en $\mathcal{S}_p$ (4.1.4), $\widehat{\boldsymbol{\beta}}_1$ es consistente para $\boldsymbol{\beta}_{1,\mathbf{s}}$ y f_1 es continua.

Etapa 3. Consistencia de $\widehat{\boldsymbol{\beta}}_2$

Tomemos $< \boldsymbol{\beta}_{1,\mathbf{s}} >^{\perp} = < \mathbf{a}_2, \ldots, \mathbf{a}_p >$ de forma tal que $\{\boldsymbol{\beta}_{1,\mathbf{s}},\ \mathbf{a}_2, \ldots,\ \mathbf{a}_p\}$ resulte una base ortonormal de $I\!R^p$, para escribir

$$\widehat{\boldsymbol{\beta}}_2 = < \widehat{\boldsymbol{\beta}}_2, \boldsymbol{\beta}_{1,\mathbf{s}} > \boldsymbol{\beta}_{1,\mathbf{s}} + \sum_{i=2}^{p} < \widehat{\boldsymbol{\beta}}_2, \mathbf{a}_i > \mathbf{a}_i = \widehat{c}_1 \boldsymbol{\beta}_{1,\mathbf{s}} + \mathbf{b}_2 \ ,$$

con $< \boldsymbol{\beta}_{1,\mathbf{s}}, \mathbf{b}_2 >= 0$. Como $\widehat{\boldsymbol{\beta}}_2$ maximiza ρ_N entre los elementos de $\mathcal{S}_p$ ortogonales a $\widehat{\boldsymbol{\beta}}_1$, sabemos que $< \widehat{\boldsymbol{\beta}}_1, \widehat{\boldsymbol{\beta}}_2 >= 0$ y entonces

$$\widehat{c}_1 = < \widehat{\boldsymbol{\beta}}_2, \boldsymbol{\beta}_{1,\mathbf{s}} > = < \widehat{\boldsymbol{\beta}}_2, \boldsymbol{\beta}_{1,\mathbf{s}} - \widehat{\boldsymbol{\beta}}_1 > = \mathbf{o}_p(1)$$

pues $\|\widehat{\boldsymbol{\beta}}_2\| = 1$ y $\widehat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_{1,\mathbf{s}} = \mathbf{o}_p(1)$ por la consistencia de $\widehat{\boldsymbol{\beta}}_1$.

Como $\|\widehat{\boldsymbol{\beta}}_2\| = 1$ y $\widehat{c}_1 \to 0$, tenemos que $\|\mathbf{b}_2\| \to 1$. Pongamos entonces

$$\widehat{\boldsymbol{\beta}}_2 = \widehat{c}_1\, \boldsymbol{\beta}_{1,\mathbf{s}} + \widehat{c}_2\, \widehat{\mathbf{b}}_2 \ ,$$

con $\widehat{c}_2 = \|\mathbf{b}_2\| \to 1$. Además,

$$\widehat{\boldsymbol{\beta}}_2 - \widehat{\mathbf{b}}_2 = \widehat{\boldsymbol{\beta}}_2 - \widehat{\mathbf{b}}_2 + \widehat{\mathbf{b}}_2\, \widehat{c}_2 - \widehat{\mathbf{b}}_2\, \widehat{c}_2 = \widehat{\boldsymbol{\beta}}_2 - \widehat{\mathbf{b}}_2\, \widehat{c}_2 + \widehat{\mathbf{b}}_2\, (\widehat{c}_2 - 1) = \widehat{c}_1\, \boldsymbol{\beta}_{1,\mathbf{s}} + \widehat{\mathbf{b}}_2\, (\widehat{c}_2 - 1) = \mathbf{o}_p(1)$$

ya que $\widehat{c}_1 \to 0$, $\widehat{c}_2 \to 1$ y $\|\widehat{\mathbf{b}}_2\| = 1$. Luego, la consistencia de $\widehat{\boldsymbol{\beta}}_2$ quedará demostrada si probamos que $\widehat{\mathbf{b}}_2 - \boldsymbol{\beta}_{2,\mathbf{s}} = \mathbf{o}_p(1)$. Como $\widehat{\mathbf{b}}_2$, $\boldsymbol{\beta}_{2,\mathbf{s}} \in \mathcal{S}_p \cap < \boldsymbol{\beta}_{1,\mathbf{s}} >^{\perp} = \mathbf{C}_2$ y $\boldsymbol{\beta}_{2,\mathbf{s}}$ es el único máximo de ρ restringida al compacto $\mathbf{C}_2$, por la Observación 4.1.1 basta verificar que $\rho(\widehat{\mathbf{b}}_2) - \rho(\boldsymbol{\beta}_{2,\mathbf{s}}) = \mathbf{o}_p(1)$, hecho que se desprende de la siguiente desigualdad

$$|\rho(\widehat{\mathbf{b}}_2) - \rho(\boldsymbol{\beta}_{2,\mathbf{s}})| \le |\rho(\widehat{\mathbf{b}}_2) - \rho(\widehat{\boldsymbol{\beta}}_2)| + |\rho(\widehat{\boldsymbol{\beta}}_2) - \rho_N(\widehat{\boldsymbol{\beta}}_2)| + |\rho_N(\widehat{\boldsymbol{\beta}}_2) - \rho(\boldsymbol{\beta}_{2,\mathbf{s}})| = \mathbf{o}_p(1) \ ,$$

pues

1. ρ es uniformemente continua en $\mathbf{C}_2$ y $\widehat{\boldsymbol{\beta}}_2 - \widehat{\mathbf{b}}_2 = \mathbf{o}_p(1)$.

2. $\sup_{\mathbf{a}\in\mathcal{S}_p} |\rho_N(\mathbf{a}) - \rho(\mathbf{a})| = \mathbf{o}_p(1)$.

3. $\rho_N(\widehat{\boldsymbol{\beta}}_2) - \rho(\boldsymbol{\beta}_{2,\mathbf{s}}) = \widehat{\nu}_{2,\mathbf{s}} - \nu_{2,\mathbf{s}} = \mathbf{o}_p(1)$, como vimos en la Etapa 2.

Etapa 4. Paso iterativo

Gracias a las etapas anteriores, vimos que $\widehat{\boldsymbol{\beta}}_m - \boldsymbol{\beta}_{m,\mathbf{s}} = \mathbf{o}_p(1) \quad 1 \leq m \leq 2$. Si ahora repetimos el procedimiento hecho en la Etapa 2 adaptando la función f_1 convenientemente, tendremos la consistencia de $\widehat{\nu}_3$ y luego, repitiendo el procedimiento hecho en la Etapa 3, se obtiene que $\widehat{\boldsymbol{\beta}}_3 - \boldsymbol{\beta}_{3,\mathbf{s}} = \mathbf{o}_p(1)$. Se sigue así con el resto de los $\widehat{\boldsymbol{\beta}}_m - \boldsymbol{\beta}_{m,\mathbf{s}}$, en la medida que podamos garantizar que ρ en el espacio $\mathbf{C}_m$ sólo se minimiza en $\boldsymbol{\beta}_{m,\mathbf{s}}$.

Etapa 5. Consistencia de $\widehat{\lambda}_{ij}$

Con la notación introducida en este Capítulo, tenemos que

$$\widehat{\lambda}_{im} = \mathbf{s}^2_{n_i}(\widehat{\boldsymbol{\beta}}^{\mathrm{T}}_m \mathbf{X}_i) = \varsigma^2_{n_i}(\widehat{\boldsymbol{\beta}}_m)\ , \qquad \lambda_{im,\mathbf{s}} = \mathbf{s}^2(F_i[\boldsymbol{\beta}_m]) = \varsigma^2_i(\boldsymbol{\beta}_m).$$

En la demostración de la Proposición 4.1.1, vimos que $\varsigma_{n_i}(\mathbf{a}) - \varsigma_i(\mathbf{a}) = \mathbf{o}^{(i)}_p(1)$ uniformemente para $\mathbf{a} \in \mathcal{S}_p$ y ya hemos demostrado consistencia de $\widehat{\boldsymbol{\beta}}_m$, para $m \leq q$. Podemos entonces concluir que $\widehat{\lambda}_{im} \rightarrow \lambda_{im,\mathbf{s}}$, para $1 \leq i \leq k$, $1 \leq m \leq q$. □

4.2. Distribucion Asintótica.

Queremos buscar una recurrencia para $\widehat{\boldsymbol{\beta}}_m - \boldsymbol{\beta}_{m,\mathbf{s}}$ mediante la cual podamos encontrar, iterativamente la distribución de $\sqrt{N}(\widehat{\boldsymbol{\beta}}_m - \boldsymbol{\beta}_{m,\mathbf{s}})$. Para eso, muy pacientemente, vamos a demostrar una serie de identidades y propiedades, las cuales nos permitirán encontrar dicha recurrencia y el desarrollo dado en el Teorema 4.0.4.

Comenzamos con la siguiente propiedad que nos dará una expresión análoga a la condición **S3** del enunciado para $\dot{\rho_N}(\mathbf{a}) - \dot{\rho}(\mathbf{a})$.

Proposición 4.2.1. *Bajo las condiciones **S2** y **S3** vale que*

$$\begin{aligned}\dot{\rho}_N(\mathbf{a}) - \dot{\rho}(\mathbf{a}) &= \sum_{i=1}^{k} \tau_i \left[2\,\varsigma_i(\mathbf{a})\, n_i^{-1} \sum_{j=1}^{n_i} h_i^{\star}(\mathbf{x}_{ij};\mathbf{a}) + 2\,\dot{\varsigma}_i(\mathbf{a})\, n_i^{-1} \sum_{j=1}^{n_i} h_i(\mathbf{x}_{ij};\mathbf{a}) \right] + \mathbf{o}_P(\, N^{-\frac{1}{2}}\,) \\ &= \sum_{i=1}^{k}\sum_{j=1}^{n_i} \mathbf{g}_i(\mathbf{x}_{ij},\mathbf{a}) + \mathbf{o}_P(\, N^{-\frac{1}{2}}\,),\end{aligned}$$

uniformemente para $\mathbf{a} \in \mathcal{S}_p$.

Demostración. Poniendo $\rho_{n_i}(\mathbf{a}) = \varsigma^2_{n_i}(\mathbf{a}), \quad \rho_i(\mathbf{a}) = \varsigma^2_i(\mathbf{a})$, y usando nuevamente que

$$\varsigma_{n_i}(\mathbf{a}) - \varsigma_i(\mathbf{a}) = n_i^{-1} \sum_{j=1}^{n_i} h_i(\mathbf{x}_{ij},\mathbf{a}) + \mathbf{o}_p(n_i^{-\frac{1}{2}})$$

y que

$$\dot{\varsigma}_{n_i}(\mathbf{a}) - \dot{\varsigma}_i(\mathbf{a}) = n_i^{-1} \sum_{j=1}^{n_i} h_i^{\star}(\mathbf{x}_{ij},\mathbf{a}) + \mathbf{o}_p(n_i^{-\frac{1}{2}})$$

uniformemente para $\mathbf{a} \in \mathcal{S}_p$. Vamos a buscar una expresión para $\dot\rho_{n_i}(\mathbf{a}) - \dot\rho_i(\mathbf{a})$. Por la regla de la cadena, tenemos $\dot\rho_{n_i}(\mathbf{a}) = 2\,\varsigma_{n_i}(\mathbf{a})\,\dot\varsigma_{n_i}(\mathbf{a})$ y $\dot\rho_i(\mathbf{a}) = 2\,\varsigma_i(\mathbf{a})\,\dot\varsigma_i(\mathbf{a})$. Luego $\dot\rho_{n_i}(\mathbf{a}) - \dot\rho_i(\mathbf{a}) = 2\,\{\,\varsigma_{n_i}(\mathbf{a})\,\dot\varsigma_{n_i}(\mathbf{a}) - \varsigma_i(\mathbf{a})\,\dot\varsigma_i(\mathbf{a})\,\}$. Sumando y restando $2\,\dot\varsigma_i(\mathbf{a})\,\varsigma_{n_i}(\mathbf{a})$ obtenemos

$$\dot\rho_{n_i}(\mathbf{a}) - \dot\rho_i(\mathbf{a}) = 2\varsigma_{n_i}(\mathbf{a})\,[\dot\varsigma_{n_i}(\mathbf{a}) - \dot\varsigma_i(\mathbf{a})] + 2\dot\varsigma_i(\mathbf{a})\,[\varsigma_{n_i}(\mathbf{a}) - \varsigma_i(\mathbf{a})].$$

Sumando y restando $2\varsigma_i(\mathbf{a})[\dot\varsigma_{n_i}(\mathbf{a}) - \dot\varsigma_i(\mathbf{a})]$ en la última igualdad, nos queda

$$\dot\rho_{n_i}(\mathbf{a}) - \dot\rho_i(\mathbf{a}) = \mathbf{o}_p(n_i^{-\frac{1}{2}}) + 2\varsigma_i(\mathbf{a})[\dot\varsigma_{n_i}(\mathbf{a}) - \dot\varsigma_i(\mathbf{a})] + 2\dot\varsigma_i(\mathbf{a})[\varsigma_{n_i}(\mathbf{a}) - \varsigma_i(\mathbf{a})],$$

donde

$$\mathbf{o}_p(n_i^{-\frac{1}{2}}) = 2[\varsigma_{n_i}(\mathbf{a}) - \varsigma_i(\mathbf{a})][\dot\varsigma_{n_i}(\mathbf{a}) - \dot\varsigma_i(\mathbf{a})].$$

Por lo tanto, usando las condiciones **S2** y **S3** obtenemos

$$\dot\rho_{n_i}(\mathbf{a}) - \dot\rho_i(\mathbf{a}) \;=\; 2\varsigma_i(\mathbf{a})n_i^{-1}\sum_{j=1}^{n_i} h_i^\star(\mathbf{x}_{ij};\mathbf{a}) + 2\dot\varsigma_i(\mathbf{a})n_i^{-1}\sum_{j=1}^{n_i} h_i(\mathbf{x}_{ij};\mathbf{a}) + \mathbf{o}_p(n_i^{-\frac{1}{2}}) \tag{4.2.1}$$

uniformemente en $\mathbf{a} \in \mathcal{S}_p$. Luego

$$\begin{aligned}
\dot\rho_N(\mathbf{a}) - \dot\rho(\mathbf{a}) &= \sum_{i=1}^{k} \tau_i[\dot\rho_{n_i}(\mathbf{a}) - \dot\rho_i(\mathbf{a})] \\
&= \sum_{i=1}^{k} \tau_i[2\varsigma_i(\mathbf{a})n_i^{-1}\sum_{j=1}^{n_i} h_i^\star(\mathbf{x}_{ij};\mathbf{a}) + 2\dot\varsigma_i(\mathbf{a})n_i^{-1}\sum_{j=1}^{n_i} h_i(\mathbf{x}_{ij};\mathbf{a})] + \mathbf{o}_p(N^{-\frac{1}{2}}) \\
&= N^{-1}\sum_{i=1}^{k}\sum_{j=1}^{n_i} \mathbf{g}_i(\mathbf{x}_{ij},\mathbf{a}) + \mathbf{o}_p(N^{-\frac{1}{2}})
\end{aligned}$$

uniformemente en $\mathbf{a}$, como queríamos probar. $\square$

Ahora vamos a buscar la primera de las identidades necesarias para la demostración del Teorema 4.0.4. Sea $g(\mathbf{u}) = \mathbf{u}^{\mathrm{T}}\mathbf{u} - 1$, entonces $\mathcal{S}_p = \{\mathbf{u} \in I\!R^p : g(\mathbf{u}) = 0\}$. Como $\widehat{\boldsymbol\beta}_1 = \arg\max_{\mathbf{a}\in\mathcal{S}_p} \rho_N(\mathbf{a})$, por la teoría de multiplicadores de Lagrange, sabemos que existe un número real μ_1 tal que

$$\dot\rho_N(\widehat{\boldsymbol\beta}_1) = 2\mu_1\widehat{\boldsymbol\beta}_1.$$

Si reemplazamos el lado izquierdo de la igualdad utilizando la expresión para $\dot\rho_N$ encontrada en (4.2.1), obtenemos

$$\dot\rho(\widehat{\boldsymbol\beta}_1) + N^{-1}\sum_{i=1}^{k}\sum_{j=1}^{n_i} \mathbf{g}_i(\mathbf{x}_{ij},\widehat{\boldsymbol\beta}_1) \;=\; 2\mu_1\widehat{\boldsymbol\beta}_1. \tag{4.2.2}$$

Definamos para $1 \le m \le p$ $\widehat{\mathbf{P}}_{m+1} = \mathbf{I}_p - \sum_{j=1}^{m} \widehat{\boldsymbol\beta}_j\widehat{\boldsymbol\beta}_j^{\mathrm{T}}$, la matriz de proyección sobre $< \widehat{\boldsymbol\beta}_1,\ldots,\widehat{\boldsymbol\beta}_m >^{\perp}$, por lo tanto $\widehat{\mathbf{P}}_2\widehat{\boldsymbol\beta}_1 = 0$. Multiplicando a ambos lados de la igualdad (4.2.2) por $\widehat{\mathbf{P}}_2$, y despejando nos queda

$$\widehat{\mathbf{P}}_2\,[\dot\rho(\widehat{\boldsymbol\beta}_1) + N^{-1}\sum_{i=1}^{k}\sum_{j=1}^{n_i} \mathbf{g}_i(\mathbf{x}_{ij},\widehat{\boldsymbol\beta}_1)] \;=\; 0.$$

En general, para un m arbitrario $1 \leq m \leq p$, por la definición (3.3.4), $\widehat{\boldsymbol{\beta}}_m = \operatorname{argmax}_{\mathbf{a}\in\mathcal{C}_m} \rho_N(\mathbf{a})$, siendo $\mathcal{C}_m = \{\mathbf{a} \in I\!R^p \ / \ \|\mathbf{a}\| = 1, \mathbf{a}^{\mathrm{T}}\widehat{\boldsymbol{\beta}}_i = 0 \ \forall \ 1 \leq i \leq m-1 \ \}$. Luego, si definimos $t_j(\mathbf{u}) = \widehat{\boldsymbol{\beta}}_j^{\mathrm{T}}\mathbf{u}$, $1 \leq j \leq m-1$, resulta que podemos reescribir el conjunto $\mathcal{C}_m$ en términos de las funciones t_j y de la función g en la forma $\mathcal{C}_m = \{\mathbf{a} \in I\!R^p \ / \ t_j(\mathbf{a}) = 0 \ 1 \leq j \leq m-1 \ \wedge \ g(\mathbf{a}) = 0\}$. Luego, nuevamente por multiplicadores de Lagrange, tenemos que existen $\mu_1, \ldots, \mu_m \in I\!R$ tales que

$$\dot{\rho}_N(\widehat{\boldsymbol{\beta}}_m) = \mu_1 \dot{t}_1(\widehat{\boldsymbol{\beta}}_m) + \ldots + \mu_{m-1}\dot{t}_{m-1}(\widehat{\boldsymbol{\beta}}_m) + \mu_m \dot{g}(\widehat{\boldsymbol{\beta}}_m)$$

como $\dot{t}_j(\mathbf{a}) = \mathbf{a}$ para $1 \leq j \leq m-1$ y $\dot{g}(\mathbf{a}) = 2\mathbf{a}$ resulta

$$\dot{\rho}_N(\widehat{\boldsymbol{\beta}}_m) = \sum_{j=1}^{m-1} \mu_j \widehat{\boldsymbol{\beta}}_j + 2\mu_m \widehat{\boldsymbol{\beta}}_m.$$

Usando nuevamente la representación para $\dot{\rho}_N$ dada en la Proposición 4.2.1 tenemos que

$$\dot{\rho}(\widehat{\boldsymbol{\beta}}_m) + N^{-1}\sum_{i=1}^{k}\sum_{j=1}^{n_i} \mathbf{g}_i(\mathbf{x}_{ij}, \widehat{\boldsymbol{\beta}}_m) + \mathbf{o}_P(N^{-\frac{1}{2}}) \ = \ \sum_{j=1}^{m-1} \mu_j \widehat{\boldsymbol{\beta}}_j + 2\mu_m \widehat{\boldsymbol{\beta}}_m.$$

Con lo cual, como $\widehat{\mathbf{P}}_{m+1}\widehat{\boldsymbol{\beta}}_j = 0 \ \ 1 \leq j \leq m$, multiplicando a ambos lados de la igualdad anterior por $\widehat{\mathbf{P}}_{m+1}$, nos queda

$$\widehat{\mathbf{P}}_{m+1}\Big[\dot{\rho}(\widehat{\boldsymbol{\beta}}_m) + N^{-1}\sum_{i=1}^{k}\sum_{j=1}^{n_i} \mathbf{g}_i(\mathbf{x}_{ij}, \widehat{\boldsymbol{\beta}}_m)\Big] = \ 0. \tag{4.2.3}$$

La ecuación (4.2.3) tiene su versión asintótica, para ver eso, necesitamos el siguiente Lema.

Lema 4.2.1. *Bajo las condiciones del Teorema 4.0.4 valen las siguientes igualdades:*

i) $\rho_N(\widehat{\boldsymbol{\beta}}_m) - \rho_N(\boldsymbol{\beta}_{m,\mathbf{s}}) - \rho(\widehat{\boldsymbol{\beta}}_m) + \rho(\boldsymbol{\beta}_{m,\mathbf{s}}) = \mathbf{o}_p(N^{-\frac{1}{2}})$

ii) $\dot{\rho}(\widehat{\boldsymbol{\beta}}_m) + N^{-1}\sum_{i=1}^{k}\sum_{j=1}^{n_i} \mathbf{g}_i(\mathbf{x}_{ij}, \widehat{\boldsymbol{\beta}}_m) - \Big\{\dot{\rho}(\boldsymbol{\beta}_{m,\mathbf{s}}) + N^{-1}\sum_{i=1}^{k}\sum_{j=1}^{n_i} \mathbf{g}_i(\mathbf{x}_{ij}, \boldsymbol{\beta}_{m,\mathbf{s}}\Big\} =$
$\ddot{\rho}(\boldsymbol{\beta}_{m,\mathbf{s}})(\widehat{\boldsymbol{\beta}}_m - \boldsymbol{\beta}_{m,\mathbf{s}}) + \mathbf{o}_p(\|\widehat{\boldsymbol{\beta}}_m - \boldsymbol{\beta}_{m,\mathbf{s}}\|) + \mathbf{o}_P(N^{-\frac{1}{2}})$

Demostración. Vamos a demostrar el item ii) del lema, el item i) sale en forma similar.

Bajo las condiciones puestas sobre $\varsigma_i(\mathbf{a})$, $h_i^\star(\mathbf{x};\mathbf{a})$, $\dot{\varsigma}_i(\mathbf{a})$ y $h_i(\mathbf{x};\mathbf{a})$, tenemos que $\{\mathbf{g}_i(\mathbf{x},\mathbf{a}) : \mathbf{a} \in \mathcal{S}_p\}$ tiene entropía finita. Notar que además, $E\{\mathbf{g}_i(\mathbf{x},\mathbf{a})\} = 0$. Luego por la desigualdad maximal, la clase de funciones es Donsker y por lo tanto

$$\sup_{\mathbf{a},\mathbf{b}\in\mathcal{S}_p \ \|\mathbf{a}-\mathbf{b}\|\leq\alpha_N} \Big|\frac{1}{n_i}\sum_{j=1}^{n_i} \mathbf{g}_i(\mathbf{x}_{ij},\mathbf{a}) - \mathbf{g}_i(\mathbf{x}_{ij},\mathbf{b})\Big| \ = \mathbf{o}_p(n_i^{-\frac{1}{2}}) \ ,$$

si $\alpha_N \to 0$, de donde concluimos que

$$\sup_{\mathbf{a},\mathbf{b}\in\mathcal{S}_p \ \|\mathbf{a}-\mathbf{b}\|\leq\alpha_N} \Big|\sum_{i=1}^{k}\tau_i\frac{1}{n_i}\sum_{j=1}^{n_i}\{\mathbf{g}_i(\mathbf{x}_{ij},\mathbf{a}) - \mathbf{g}_i(\mathbf{x}_{ij},\mathbf{b})\}\Big| \ = \mathbf{o}_p(N^{-\frac{1}{2}}) \ .$$

Luego, la consistencia de $\widehat{\boldsymbol{\beta}}_m$ garantiza que

$$\dot{\rho}(\widehat{\boldsymbol{\beta}}_m)+N^{-1}\sum_{i=1}^{k}\sum_{j=1}^{n_i} g_i(\mathbf{x}_{ij};\widehat{\boldsymbol{\beta}}_m)-\{\dot{\rho}(\boldsymbol{\beta}_{m,\mathbf{s}})+N^{-1}\sum_{i=1}^{k}\sum_{j=1}^{n_i} g_i(\mathbf{x}_{ij};\boldsymbol{\beta}_{m,\mathbf{s}})\} \;=\; \dot{\rho}(\widehat{\boldsymbol{\beta}}_m)-\dot{\rho}(\boldsymbol{\beta}_{m,\mathbf{s}})+\mathbf{o}_p(N^{-\frac{1}{2}})\;.$$

Por último, expandiendo $\dot{\rho}(\mathbf{a})$ en su desarrollo de Taylor usando la continuidad de $\ddot{\rho}(\mathbf{a})$, concluimos el resultado deseado. □

Ahora si veamos la versión asintótica de la ecuación (4.2.3).

Proposición 4.2.2. *Bajo las condiciones del Teorema 4.0.4 vale la siguiente igualdad*

$$\mathbf{P}_{m+1}\dot{\rho}(\boldsymbol{\beta}_{m.\mathbf{s}}) \;=\; 0 \tag{4.2.4}$$

Demostración. Gracias al item ii) del Lema 4.2.1 y a la consistencia de $\widehat{\boldsymbol{\beta}}_{m,\mathbf{s}}$ tenemos que

$$N^{-1}\sum_{i=1}^{k}\sum_{j=1}^{n_i} \mathbf{g}_i(\mathbf{x}_{ij};\widehat{\boldsymbol{\beta}}_m) = N^{-1}\sum_{i=1}^{k}\sum_{j=1}^{n_i} \mathbf{g}_i(\mathbf{x}_{ij};\boldsymbol{\beta}_{m,\mathbf{s}})\} + \mathbf{o}_p(1)$$

Reemplazando esto en la igualdad (4.2.3), obtenemos la siguiente expresión

$$\widehat{\mathbf{P}}_{m+1}\Big[\dot{\rho}(\widehat{\boldsymbol{\beta}}_m) + N^{-1}\sum_{i=1}^{k}\sum_{j=1}^{n_i} \mathbf{g}_i(\mathbf{x}_{ij};\boldsymbol{\beta}_{m,\mathbf{s}})\}\Big] + \mathbf{o}_p(1) = \;0\;.$$

Debido a la expresión de $\mathbf{g}_i$, la ley de los grandes números, la continuidad de $\dot{\rho}$ y la consistencia de $\widehat{\boldsymbol{\beta}}_m$, tomando límite a ambos lados de la igualdad anterior, obtenemos la identidad buscada. □

Paralelamente, tenemos la siguiente relación entre los estimadores de projection pursuit de las direcciones principales bajo CPC y las matrices de proyección

Lema 4.2.2. *Bajo las condiciones del Teorema 4.0.4 se cumple que, para todo* $\mathbf{b} \in I\!R^p$

$$(\widehat{\mathbf{P}}_{m+1}-\mathbf{P}_{m+1})\mathbf{b} \;=\; -\sum_{j=1}^{m}(\;\boldsymbol{\beta}_{j,\mathbf{s}}^{\mathrm{T}}\mathbf{b}\mathbf{I}_p \;+\; \boldsymbol{\beta}_{j,\mathbf{s}}\mathbf{b}^{\mathrm{T}}\;)(\widehat{\boldsymbol{\beta}}_j-\boldsymbol{\beta}_{j,\mathbf{s}})+O_p(\;\|\mathbf{b}\|\sum_{i=1}^{m}\|\widehat{\boldsymbol{\beta}}_i-\boldsymbol{\beta}_{i,\mathbf{s}}\|^2\;). \tag{4.2.5}$$

Demostración. En efecto,

$$\sum_{j=1}^{m}(\;\boldsymbol{\beta}_{j,\mathbf{s}}^{\mathrm{T}}\mathbf{b}\mathbf{I}_p+\boldsymbol{\beta}_{j,\mathbf{s}}\mathbf{b}^{\mathrm{T}}\;)(\widehat{\boldsymbol{\beta}}_j-\boldsymbol{\beta}_{j,\mathbf{s}}) = \sum_{j=1}^{m}\boldsymbol{\beta}_{j,\mathbf{s}}^{\mathrm{T}}\mathbf{b}\widehat{\boldsymbol{\beta}}_j-\sum_{j=1}^{m}\boldsymbol{\beta}_{j,\mathbf{s}}^{\mathrm{T}}\mathbf{b}\boldsymbol{\beta}_{j,\mathbf{s}}+\sum_{j=1}^{m}\boldsymbol{\beta}_{j,\mathbf{s}}\mathbf{b}^{\mathrm{T}}\widehat{\boldsymbol{\beta}}_j-\sum_{j=1}^{m}\boldsymbol{\beta}_{j,\mathbf{s}}\mathbf{b}^{\mathrm{T}}\boldsymbol{\beta}_{j,\mathbf{s}} =$$

$$\sum_{j=1}^{m}<\;\boldsymbol{\beta}_{j,\mathbf{s}},\mathbf{b}\;>\widehat{\boldsymbol{\beta}}_j+\sum_{j=1}^{m}<\;\widehat{\boldsymbol{\beta}}_j,\mathbf{b}\;>\boldsymbol{\beta}_{j,\mathbf{s}}-\sum_{j=1}^{m}<\;\boldsymbol{\beta}_{j,\mathbf{s}},\mathbf{b}\;>\boldsymbol{\beta}_{j,\mathbf{s}}-\sum_{j=1}^{m}<\;\boldsymbol{\beta}_{j,\mathbf{s}},\mathbf{b}\;>\boldsymbol{\beta}_{j,\mathbf{s}}.$$

Por lo tanto,

$$\sum_{j=1}^{m}(\;\boldsymbol{\beta}_{j,\mathbf{s}}^{\mathrm{T}}\mathbf{b}\mathbf{I}_p+\boldsymbol{\beta}_{j,\mathbf{s}}\mathbf{b}^{\mathrm{T}}\;)(\widehat{\boldsymbol{\beta}}_j-\boldsymbol{\beta}_{j,\mathbf{s}}) = \sum_{j=1}^{m}<\;\boldsymbol{\beta}_{j,\mathbf{s}},\mathbf{b}\;>\widehat{\boldsymbol{\beta}}_j+\sum_{j=1}^{m}<\;\widehat{\boldsymbol{\beta}}_j,\mathbf{b}\;>\boldsymbol{\beta}_{j,\mathbf{s}}-2\sum_{j=1}^{m}<\;\boldsymbol{\beta}_{j,\mathbf{s}},\mathbf{b}\;>\boldsymbol{\beta}_{j,\mathbf{s}}.$$

Por otro lado

$$(\widehat{\mathbf{P}}_{m+1} - \mathbf{P}_{m+1})\mathbf{b} \;=\; \sum_{i=1}^{m} \boldsymbol{\beta}_{i,\mathbf{s}}\boldsymbol{\beta}_{i,\mathbf{s}}^{\mathrm{T}}\mathbf{b} - \sum_{i=1}^{m} \widehat{\boldsymbol{\beta}}_i\widehat{\boldsymbol{\beta}}_i^{\mathrm{T}}\mathbf{b} \;=\; \sum_{i=1}^{m} <\; \boldsymbol{\beta}_{i,\mathbf{s}}, \mathbf{b} \;> \boldsymbol{\beta}_{i,\mathbf{s}} - \sum_{i=1}^{m} <\; \widehat{\boldsymbol{\beta}}_i, \mathbf{b} \;> \widehat{\boldsymbol{\beta}}_i.$$

Sumando las dos últimas igualdades, y, utilizando las propiedades del producto escalar, obtenemos

$$\sum_{j=1}^{m} (\; \boldsymbol{\beta}_{j,\mathbf{s}}^{\mathrm{T}}\mathbf{b}\mathbf{I}_p + \boldsymbol{\beta}_{j,\mathbf{s}}\mathbf{b}^{\mathrm{T}}\;)(\widehat{\boldsymbol{\beta}}_j - \boldsymbol{\beta}_{j,\mathbf{s}}) + (\widehat{\mathbf{P}}_{m+1} - \mathbf{P}_{m+1})\mathbf{b} = \sum_{j=1}^{m} <\; \widehat{\boldsymbol{\beta}}_j - \boldsymbol{\beta}_{j,\mathbf{s}}, \mathbf{b} \;> (\boldsymbol{\beta}_{j,\mathbf{s}} - \widehat{\boldsymbol{\beta}}_j).$$

Por otra parte,

$$\sum_{j=1}^{m} <\; \widehat{\boldsymbol{\beta}}_j - \boldsymbol{\beta}_{j,\mathbf{s}}, \mathbf{b} \;> (\boldsymbol{\beta}_{j,\mathbf{s}} - \widehat{\boldsymbol{\beta}}_j) \;=\; O_p(\;\|\mathbf{b}\| \sum_{i=1}^{m} \|\widehat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_{i,\mathbf{s}}\|^2\;).$$

En efecto, usando la desigualdad triangular y la desigualdad de Cauchy- Schwartz, tenemos

$$\begin{aligned} \|\sum_{j=1}^{m} <\; \widehat{\boldsymbol{\beta}}_j - \boldsymbol{\beta}_{j,\mathbf{s}}, \mathbf{b} \;> (\boldsymbol{\beta}_{j,\mathbf{s}} - \widehat{\boldsymbol{\beta}}_j)\| &\leq \sum_{j=1}^{m} |<\; \widehat{\boldsymbol{\beta}}_j - \boldsymbol{\beta}_{j,\mathbf{s}}, \mathbf{b} \;>| \|(\boldsymbol{\beta}_{j,\mathbf{s}} - \widehat{\boldsymbol{\beta}}_j)\| \\ &\leq \sum_{j=1}^{m} \|\widehat{\boldsymbol{\beta}}_j - \boldsymbol{\beta}_{j,\mathbf{s}}\|\|\mathbf{b}\|\|\boldsymbol{\beta}_{j,\mathbf{s}} - \widehat{\boldsymbol{\beta}}_j\| \;=\; \sum_{j=1}^{m} \|\widehat{\boldsymbol{\beta}}_j - \boldsymbol{\beta}_{j,\mathbf{s}}\|^2\|\mathbf{b}\|, \end{aligned}$$

de donde

$$\|\sum_{j=1}^{m} <\; \widehat{\boldsymbol{\beta}}_j - \boldsymbol{\beta}_{j,\mathbf{s}}, \mathbf{b} \;> (\boldsymbol{\beta}_{j,\mathbf{s}} - \widehat{\boldsymbol{\beta}}_j)\| \leq \sum_{j=1}^{m} \|\widehat{\boldsymbol{\beta}}_j - \boldsymbol{\beta}_{j,\mathbf{s}}\|^2\|\mathbf{b}\|$$

como queríamos ver. En suma, demostramos que

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

de donde se deduce la igualdad (4.2.5) □

La próxima propiedad que vamos a desarrollar es la clave para construir una recurrencia para los autovectores.

Proposición 4.2.3. *Bajo las condiciones del Teorema 4.0.4, se verifica*

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

Demostración. Por comodidad llamemos

$$L = \mathbf{P}_{m+1}\ddot{\rho}(\boldsymbol{\beta}_{m,\mathbf{s}})(\widehat{\boldsymbol{\beta}}_m - \boldsymbol{\beta}_{m,\mathbf{s}}) - \boldsymbol{\beta}^{\mathrm{T}}_{m,\mathbf{s}}\dot{\rho}(\boldsymbol{\beta}_{m,\mathbf{s}})(\widehat{\boldsymbol{\beta}}_m - \boldsymbol{\beta}_{m,\mathbf{s}}) - \boldsymbol{\beta}_{m,\mathbf{s}}\dot{\rho}(\boldsymbol{\beta}_{m,\mathbf{s}})^{\mathrm{T}}(\widehat{\boldsymbol{\beta}}_m - \boldsymbol{\beta}_{m,\mathbf{s}}).$$

Si sumamos y restamos $\widehat{\mathbf{P}}_{m+1}\,\ddot{\rho}(\boldsymbol{\beta}_{m,\mathbf{s}})\,(\widehat{\boldsymbol{\beta}}_m - \boldsymbol{\beta}_{m,\mathbf{s}})$ en L obtenemos

$$\begin{aligned} L &= (\mathbf{P}_{m+1} - \widehat{\mathbf{P}}_{m+1})\,\ddot{\rho}(\boldsymbol{\beta}_{m,\mathbf{s}})\,(\widehat{\boldsymbol{\beta}}_m - \boldsymbol{\beta}_{m,\mathbf{s}}) - \boldsymbol{\beta}^{\mathrm{T}}_m\,\dot{\rho}(\boldsymbol{\beta}_m)\,(\widehat{\boldsymbol{\beta}}_m - \boldsymbol{\beta}_{m,\mathbf{s}}) \\ &- \boldsymbol{\beta}_{m,\mathbf{s}}\,\dot{\rho}^{\mathrm{T}}(\boldsymbol{\beta}_{m,\mathbf{s}})\,(\widehat{\boldsymbol{\beta}}_m - \boldsymbol{\beta}_{m,\mathbf{s}}) + \widehat{\mathbf{P}}_{m+1}\,\ddot{\rho}(\boldsymbol{\beta}_{m,\mathbf{s}})\,(\widehat{\boldsymbol{\beta}}_m - \boldsymbol{\beta}_{m,\mathbf{s}}). \end{aligned}$$

Como $(\mathbf{P}_{m+1} - \widehat{\mathbf{P}}_{m+1}) = \mathbf{o}_p(1)$ el primer sumando de L es un $\mathbf{o}_p(\|\widehat{\boldsymbol{\beta}}_m - \boldsymbol{\beta}_{m,\mathbf{s}}\|)$. Aplicando la igualdad ii) del Lema 4.2.1 al último sumando, obtenemos

$$\begin{aligned} L &= \mathbf{o}_p(\|\widehat{\boldsymbol{\beta}}_m - \boldsymbol{\beta}_{m,\mathbf{s}}\|) - \boldsymbol{\beta}^{\mathrm{T}}_m\dot{\rho}(\boldsymbol{\beta}_{m,\mathbf{s}})(\widehat{\boldsymbol{\beta}}_m - \boldsymbol{\beta}_{m,\mathbf{s}}) - \boldsymbol{\beta}_{m,\mathbf{s}}\dot{\rho}(\boldsymbol{\beta}_{m,\mathbf{s}})^{\mathrm{T}}(\widehat{\boldsymbol{\beta}}_m - \boldsymbol{\beta}_{m,\mathbf{s}}) \\ &+ \widehat{\mathbf{P}}_{m+1}\Big[\dot{\rho}(\widehat{\boldsymbol{\beta}}_m) + N^{-1}\sum_{i=1}^{k}\sum_{j=1}^{n_i}\mathbf{g}_i(\mathbf{x}_{ij};\widehat{\boldsymbol{\beta}}_m) - \{\dot{\rho}(\boldsymbol{\beta}_{m,\mathbf{s}}) + N^{-1}\sum_{i=1}^{k}\sum_{j=1}^{n_i}\mathbf{g}_i(\mathbf{x}_{ij};\boldsymbol{\beta}_{m,\mathbf{s}})\}\Big] \\ &+ \mathbf{o}_p(\|\widehat{\boldsymbol{\beta}}_m - \boldsymbol{\beta}_{m,\mathbf{s}}\|) + \mathbf{o}_p(N^{-\frac{1}{2}}). \end{aligned}$$

Ahora bien, por la igualdad (4.2.3) tenemos que

$$\widehat{\mathbf{P}}_{m+1}\Big[\dot{\rho}(\widehat{\boldsymbol{\beta}}_m) + N^{-1}\sum_{i=1}^{k}\sum_{j=1}^{n_i}\mathbf{g}_i(\mathbf{x}_{ij};\widehat{\boldsymbol{\beta}}_m)\Big] = 0.$$

Luego usando la última igualdad, la definición de $\tau_i = n_i/N$ y reordenando los términos, obtenemos

$$\begin{aligned} L &= \mathbf{o}_p(\|\widehat{\boldsymbol{\beta}}_m - \boldsymbol{\beta}_{m,\mathbf{s}}\|) + \mathbf{o}_p(N^{-\frac{1}{2}}) - (\ \boldsymbol{\beta}^{\mathrm{T}}_{m,\mathbf{s}}\dot{\rho}(\boldsymbol{\beta}_{m,\mathbf{s}})\mathbf{I}_p\ +\ \boldsymbol{\beta}_{m,\mathbf{s}}\dot{\rho}(\boldsymbol{\beta}_{m,\mathbf{s}})^{\mathrm{T}}\ )(\widehat{\boldsymbol{\beta}}_m - \boldsymbol{\beta}_{m,\mathbf{s}}) \\ &- \widehat{\mathbf{P}}_{m+1}\Big[\dot{\rho}(\boldsymbol{\beta}_{m,\mathbf{s}}) + N^{-1}\sum_{i=1}^{k}\sum_{j=1}^{n_i}\mathbf{g}_i(\mathbf{x}_{ij};\boldsymbol{\beta}_{m,\mathbf{s}})\Big]. \end{aligned}$$

Si recordamos que $\mathbf{u}_m = -N^{-\frac{1}{2}}\sum_{i=1}^{k}\sum_{j=1}^{n_i}\mathbf{g}_i(\mathbf{x}_{ij};\boldsymbol{\beta}_{m,\mathbf{s}})$, podemos escribir

$$\begin{aligned} L &= \mathbf{o}_p(\|\widehat{\boldsymbol{\beta}}_m - \boldsymbol{\beta}_{m,\mathbf{s}}\|) + \mathbf{o}_p(N^{-\frac{1}{2}}) - (\ \boldsymbol{\beta}^{\mathrm{T}}_{m,\mathbf{s}}\dot{\rho}(\boldsymbol{\beta}_{m,\mathbf{s}})\mathbf{I}_p\ +\ \boldsymbol{\beta}_{m,\mathbf{s}}\dot{\rho}(\boldsymbol{\beta}_{m,\mathbf{s}})^{\mathrm{T}}\ )(\widehat{\boldsymbol{\beta}}_m - \boldsymbol{\beta}_{m,\mathbf{s}}) \\ &+ \widehat{\mathbf{P}}_{m+1}\mathbf{u}_m N^{-\frac{1}{2}} - \widehat{\mathbf{P}}_{m+1}\dot{\rho}(\boldsymbol{\beta}_{m,\mathbf{s}}). \end{aligned}$$

Luego, sumando y restando $\mathbf{P}_{m+1}\mathbf{u}_m N^{-\frac{1}{2}}$ y $\mathbf{P}_{m+1}\dot{\rho}(\boldsymbol{\beta}_{m,\mathbf{s}})$ a la última igualdad y usando (4.2.4) obtenemos,

$$\begin{aligned} L &= \mathbf{o}_p(\|\widehat{\boldsymbol{\beta}}_m - \boldsymbol{\beta}_{m,\mathbf{s}}\|) + \mathbf{o}_p(N^{-\frac{1}{2}}) - (\ \boldsymbol{\beta}^{\mathrm{T}}_{m,\mathbf{s}}\dot{\rho}(\boldsymbol{\beta}_{m,\mathbf{s}})\mathbf{I}_p\ +\ \boldsymbol{\beta}_{m,\mathbf{s}}\dot{\rho}(\boldsymbol{\beta}_{m,\mathbf{s}})^{\mathrm{T}}\ )(\widehat{\boldsymbol{\beta}}_m - \boldsymbol{\beta}_{m,\mathbf{s}}) \\ &+ (\widehat{\mathbf{P}}_{m+1} - \mathbf{P}_{m+1})\mathbf{u}_m N^{-\frac{1}{2}} + \mathbf{P}_{m+1}\mathbf{u}_m N^{-\frac{1}{2}} - (\widehat{\mathbf{P}}_{m+1} - \mathbf{P}_{m+1})\dot{\rho}(\boldsymbol{\beta}_{m,\mathbf{s}}). \end{aligned}$$

Por el Teorema central del limite $\mathbf{u}_m$ converge en distribución a una distribución normal y, por lo tanto, es acotada en probabilidad, de donde al ser $(\widehat{\mathbf{P}}_{m+1} - \mathbf{P}_{m+1}) = \mathbf{o}_p(1)$ resulta que

$$(\widehat{\mathbf{P}}_{m+1} - \mathbf{P}_{m+1})\mathbf{u}_m N^{-\frac{1}{2}} = \mathbf{o}_p(N^{-\frac{1}{2}}).$$

Por lo tanto,

$$\begin{aligned} L &= \mathbf{o}_p(\|\widehat{\boldsymbol{\beta}}_m - \boldsymbol{\beta}_{m,\mathbf{s}}\|) + \mathbf{o}_p(N^{-\frac{1}{2}}) - (\ \boldsymbol{\beta}^{\mathrm{T}}_{m,\mathbf{s}}\dot{\rho}(\boldsymbol{\beta}_{m,\mathbf{s}})\mathbf{I}_p\ +\ \boldsymbol{\beta}_{m,\mathbf{s}}\dot{\rho}(\boldsymbol{\beta}_{m,\mathbf{s}})^{\mathrm{T}}\ )(\widehat{\boldsymbol{\beta}}_m - \boldsymbol{\beta}_{m,\mathbf{s}}) \\ &+ \mathbf{P}_{m+1}\mathbf{u}_m N^{-\frac{1}{2}} - (\widehat{\mathbf{P}}_{m+1} - \mathbf{P}_{m+1})\dot{\rho}(\boldsymbol{\beta}_{m,\mathbf{s}}). \end{aligned}$$

Usando el Lema 4.2.2 obtenemos

$$L = \mathbf{o}_p(\|\widehat{\boldsymbol\beta}_m - \boldsymbol\beta_{m,\mathbf{s}}\|) + \mathbf{o}_p(N^{-\frac{1}{2}}) - (\ \boldsymbol\beta^{\mathrm{T}}_{m,\mathbf{s}}\dot\rho(\boldsymbol\beta_{m,\mathbf{s}})\mathbf{I}_p\ +\ \boldsymbol\beta_{m,\mathbf{s}}\dot\rho(\boldsymbol\beta_{m,\mathbf{s}})^{\mathrm{T}}\ )(\widehat{\boldsymbol\beta}_m - \boldsymbol\beta_{m,\mathbf{s}})$$
$$+\mathbf{P}_{m+1}\mathbf{u}_m N^{-\frac{1}{2}} + \sum_{j=1}^{m}(\ \boldsymbol\beta^{\mathrm{T}}_{j,\mathbf{s}}\dot\rho(\boldsymbol\beta_{m,\mathbf{s}})\mathbf{I}_p\ +\ \boldsymbol\beta_{j,\mathbf{s}}\dot\rho(\boldsymbol\beta_{m,\mathbf{s}})^{\mathrm{T}}\ )(\widehat{\boldsymbol\beta}_j - \boldsymbol\beta_{j,\mathbf{s}}) + O_p(\ \|\dot\rho(\boldsymbol\beta_{m,\mathbf{s}})\|\sum_{j=1}^{m}\|\widehat{\boldsymbol\beta}_j - \boldsymbol\beta_{j,\mathbf{s}}\|^2\ ).$$

Usando que por **S3** $\dot\rho$ es continua y, por lo tanto, acotada en $\mathcal{S}_p$ y que $\|\widehat{\boldsymbol\beta}_j - \boldsymbol\beta_{j,\mathbf{s}}\| = \mathbf{o}_p(1)$ obtenemos

$$O_p(\ \|\dot\rho(\boldsymbol\beta_{m,\mathbf{s}})\|\sum_{j=1}^{m}\|\widehat{\boldsymbol\beta}_j - \boldsymbol\beta_{j,\mathbf{s}}\|^2) = \mathbf{o}_p(\ \sum_{j=1}^{m}\|\widehat{\boldsymbol\beta}_j - \boldsymbol\beta_{j,\mathbf{s}}\|) = \mathbf{o}_p(\ \sum_{j=1}^{m-1}\|\widehat{\boldsymbol\beta}_j - \boldsymbol\beta_{j,\mathbf{s}}\|) + \mathbf{o}_p(\|\widehat{\boldsymbol\beta}_m - \boldsymbol\beta_{m,\mathbf{s}}\|),$$

de donde

$$(\mathbf{P}_{m+1}\ddot\rho(\boldsymbol\beta_{m,\mathbf{s}}) - \boldsymbol\beta^{\mathrm{T}}_{m,\mathbf{s}}\dot\rho(\boldsymbol\beta_{m,\mathbf{s}})\mathbf{I}_p - \boldsymbol\beta_{m,\mathbf{s}}\dot\rho(\boldsymbol\beta_{m,\mathbf{s}})^{\mathrm{T}})(\widehat{\boldsymbol\beta}_m - \boldsymbol\beta_{m,\mathbf{s}}) + \mathbf{o}_p(\ \|\widehat{\boldsymbol\beta}_m - \boldsymbol\beta_{m,\mathbf{s}}\ \|)\ =$$
$$\sum_{j=1}^{m-1}(\ \boldsymbol\beta^{\mathrm{T}}_{j,\mathbf{s}}\dot\rho(\boldsymbol\beta_{m,\mathbf{s}})\mathbf{I}_p\ +\ \boldsymbol\beta_{j,\mathbf{s}}\dot\rho(\boldsymbol\beta_{m,\mathbf{s}})^{\mathrm{T}}\ )(\widehat{\boldsymbol\beta}_j - \boldsymbol\beta_{j,\mathbf{s}}) + N^{-\frac{1}{2}}\mathbf{P}_{m+1}\mathbf{u}_m + \mathbf{o}_p(\ \sum_{j=1}^{m-1}\|\widehat{\boldsymbol\beta}_j - \boldsymbol\beta_{j,\mathbf{s}}\|\ ) + \mathbf{o}_p(N^{-\frac{1}{2}})$$

lo que concluye la demostración. □

Usando la Proposición anterior, demostraremos la siguiente identidad.

Proposición 4.2.4. *Bajo las hipótesis del Teorema 4.0.4*

$$\mathbf{A}_m(\widehat{\boldsymbol\beta}_m - \boldsymbol\beta_{m,\mathbf{s}}) + \mathbf{o}_p(\|\widehat{\boldsymbol\beta}_m - \boldsymbol\beta_{m,\mathbf{s}}\|)\ =\ \sum_{j=1}^{m-1}\mathbf{B}_{jm}(\widehat{\boldsymbol\beta}_j - \boldsymbol\beta_{j,\mathbf{s}})$$
$$+\ \ N^{-\frac{1}{2}}\mathbf{P}_{m+1}\mathbf{u}_m + \mathbf{o}_p(\sum_{j=1}^{m-1}\|\widehat{\boldsymbol\beta}_j - \boldsymbol\beta_{j,\mathbf{s}}\|) + \mathbf{o}_p(N^{-\frac{1}{2}}) \qquad (4.2.6)$$

donde

$$\mathbf{A}_m = \mathbf{P}_{m+1}\ddot\rho(\boldsymbol\beta_{m,\mathbf{s}}) - \boldsymbol\beta^{\mathrm{T}}_{m,\mathbf{s}}\dot\rho(\boldsymbol\beta_{m,\mathbf{s}})\mathbf{I}_p - \sum_{i=1}^{m-1}\boldsymbol\beta^{\mathrm{T}}_{i,\mathbf{s}}\dot\rho(\boldsymbol\beta_{m,\mathbf{s}})\boldsymbol\beta_{m,\mathbf{s}}\boldsymbol\beta^{\mathrm{T}}_{i,\mathbf{s}}$$
$$\mathbf{B}_{jm} = \boldsymbol\beta^{\mathrm{T}}_{j,\mathbf{s}}\dot\rho(\boldsymbol\beta_{m,\mathbf{s}})\mathbf{I}_p + \boldsymbol\beta_{j,\mathbf{s}}\dot\rho(\boldsymbol\beta_{m,\mathbf{s}})^{\mathrm{T}}$$

Demostración. Como $\mathbf{P}_{m+1}\dot\rho(\boldsymbol\beta_{m,\mathbf{s}}) = 0$, tenemos que

$$\dot\rho(\boldsymbol\beta_{m,\mathbf{s}})\ \ =\ \ \sum_{j=1}^{m} < \dot\rho(\boldsymbol\beta_{m,\mathbf{s}}); \boldsymbol\beta_{j,\mathbf{s}} > \boldsymbol\beta_{j,\mathbf{s}}. \qquad (4.2.7)$$

Luego usando la Proposición 4.2.3 bastará mostrar que

$$\boldsymbol\beta^{\mathrm{T}}_m(\widehat{\boldsymbol\beta}_m - \boldsymbol\beta_{m,\mathbf{s}}) = O_p(\|\widehat{\boldsymbol\beta}_m - \boldsymbol\beta_{m,\mathbf{s}}\|^2).$$

Primero notemos que como $\widehat{\boldsymbol\beta}_m$ y $\boldsymbol\beta_{m,\mathbf{s}} \in \mathcal{S}_p$,

$$\|\widehat{\boldsymbol\beta}_m - \boldsymbol\beta_{m,\mathbf{s}}\|^2 =\ < \widehat{\boldsymbol\beta}_m - \boldsymbol\beta_{m,\mathbf{s}}, \widehat{\boldsymbol\beta}_m - \boldsymbol\beta_{m,\mathbf{s}} >=\ \ 1 - 2 < \boldsymbol\beta_{m,\mathbf{s}}, \widehat{\boldsymbol\beta}_m > +1,$$

con lo cual,

$$<\boldsymbol{\beta}_{m,\mathbf{s}}, \widehat{\boldsymbol{\beta}}_m> = 1 - \frac{\|\widehat{\boldsymbol{\beta}}_m - \boldsymbol{\beta}_{m,\mathbf{s}}\|^2}{2},$$

y por lo tanto,

$$\begin{aligned}
\boldsymbol{\beta}_{m,\mathbf{s}}^{\mathrm{T}}(\widehat{\boldsymbol{\beta}}_m - \boldsymbol{\beta}_{m,\mathbf{s}}) &= \; <\boldsymbol{\beta}_{m,\mathbf{s}}, \widehat{\boldsymbol{\beta}}_m - \boldsymbol{\beta}_{m,\mathbf{s}}> = <\boldsymbol{\beta}_{m,\mathbf{s}} - \widehat{\boldsymbol{\beta}}_m, \widehat{\boldsymbol{\beta}}_m - \boldsymbol{\beta}_{m,\mathbf{s}}> + <\widehat{\boldsymbol{\beta}}_m, \widehat{\boldsymbol{\beta}}_m - \boldsymbol{\beta}_{m,\mathbf{s}}> \\
&= -\|\widehat{\boldsymbol{\beta}}_m - \boldsymbol{\beta}_{m,\mathbf{s}}\|^2 + 1 - <\boldsymbol{\beta}_{m,\mathbf{s}}, \widehat{\boldsymbol{\beta}}_m> \\
&= -\|\widehat{\boldsymbol{\beta}}_m - \boldsymbol{\beta}_{m,\mathbf{s}}\|^2 + 1 - (1 - \frac{\|\widehat{\boldsymbol{\beta}}_m - \boldsymbol{\beta}_{m,\mathbf{s}}\|^2}{2}) \\
&= -\frac{\|\widehat{\boldsymbol{\beta}}_m - \boldsymbol{\beta}_{m,\mathbf{s}}\|^2}{2}.
\end{aligned}$$

Es decir, $\boldsymbol{\beta}_{m,\mathbf{s}}^{\mathrm{T}}(\widehat{\boldsymbol{\beta}}_m - \boldsymbol{\beta}_{m,\mathbf{s}}) = O_p(\|\widehat{\boldsymbol{\beta}}_m - \boldsymbol{\beta}_{m,\mathbf{s}}\|^2)$, con lo cual

$$\boldsymbol{\beta}_{m,\mathbf{s}}^{\mathrm{T}}(\widehat{\boldsymbol{\beta}}_m - \boldsymbol{\beta}_{m,\mathbf{s}}) = \mathbf{o}_p(\|\widehat{\boldsymbol{\beta}}_m - \boldsymbol{\beta}_{m,\mathbf{s}}\|) \tag{4.2.8}$$

lo que concluye la demostración. □

De la última identidad podemos deducir la siguiente Proposición.

Proposición 4.2.5. *Bajo las condiciones del Teorema 4.0.4 se cumple*

a) $\widehat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_{1,\mathbf{s}} = O_p(N^{-\frac{1}{2}})$

b) $\widehat{\boldsymbol{\beta}}_m - \boldsymbol{\beta}_{m,\mathbf{s}} = O_p(N^{-\frac{1}{2}})$

Demostración.

a) De la identidad (4.2.6), para $m = 1$ obtenemos $\mathbf{A}_1(\widehat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_{1,\mathbf{s}}) + \mathbf{o}_p(\|\widehat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_{1,\mathbf{s}}\|) = N^{-\frac{1}{2}}\,\mathbf{P}_2\,\mathbf{u}_1 + \mathbf{o}_p(N^{-\frac{1}{2}})$, de donde, despejando $(\widehat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_{1,\mathbf{s}})$ se tiene que $(\widehat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_{1,\mathbf{s}}) = \mathbf{o}_p(\|\widehat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_{1,\mathbf{s}}\|) + \mathbf{A}_1^{-1}\,N^{-\frac{1}{2}}\,\mathbf{P}_2\,\mathbf{u}_1 + \mathbf{o}_p(N^{-\frac{1}{2}})$. Como $\mathbf{P}_2\,\mathbf{u}_1$ converge en distribución a una normal, resulta que $\mathbf{A}_1^{-1}\,N^{-\frac{1}{2}}\,\mathbf{P}_2\,\mathbf{u}_1 = O_p(N^{-\frac{1}{2}})$, de donde se deduce $N^{\frac{1}{2}}(\widehat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_{1,\mathbf{s}}) = N^{\frac{1}{2}}\,\mathbf{o}_p(\|\widehat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_{1,\mathbf{s}}\|) + O_p(1) + \mathbf{o}_p(1)$, es decir,

$$N^{\frac{1}{2}}(\widehat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_{1,\mathbf{s}}) = N^{\frac{1}{2}}\mathbf{o}_p(\|\widehat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_{1,\mathbf{s}}\|) + O_p(1). \tag{4.2.9}$$

Por lo tanto,

$$N^{\frac{1}{2}}\,\|\widehat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_{1,\mathbf{s}}\| \left[\frac{(\widehat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_{1,\mathbf{s}})}{\|\widehat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_{1,\mathbf{s}}\|} - \frac{\mathbf{o}_p(\|\widehat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_{1,\mathbf{s}}\|)}{\|\widehat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_{1,\mathbf{s}}\|}\right] = O_p(1).$$

Como $\left[\frac{(\widehat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_{1,\mathbf{s}})}{\|\widehat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_{1,\mathbf{s}}\|} - \frac{\mathbf{o}_p(\|\widehat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_{1,\mathbf{s}}\|)}{\|\widehat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_{1,\mathbf{s}}\|}\right] \xrightarrow{p} 1$, resulta que $N^{\frac{1}{2}}\,\|\widehat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_{1,\mathbf{s}}\| = O_p(1)$ como queríamos ver.

b) Veamos el paso iterativo, es decir veamos que si $N^{\frac{1}{2}}\,\|\widehat{\boldsymbol{\beta}}_j - \boldsymbol{\beta}_{j,\mathbf{s}}\| = O_p(1)$, para todo $2 \leq j \leq m-1$ entonces resulta que $N^{\frac{1}{2}}\,\|\widehat{\boldsymbol{\beta}}_m - \boldsymbol{\beta}_{m,\mathbf{s}}\| = O_p(1)$. En efecto, en la identidad (4.2.6), obtenemos

$$N^{\frac{1}{2}}(\widehat{\boldsymbol{\beta}}_m - \boldsymbol{\beta}_{m,\mathbf{s}}) = N^{\frac{1}{2}}\mathbf{o}_p(\|\widehat{\boldsymbol{\beta}}_m - \boldsymbol{\beta}_{m,\mathbf{s}}\|) + \sum_{j=1}^{m-1} \mathbf{A}_m^{-1}\mathbf{B}_{jm}N^{\frac{1}{2}}(\widehat{\boldsymbol{\beta}}_j - \boldsymbol{\beta}_{j,\mathbf{s}})$$

$$+ \quad \mathbf{P}_{m+1}\mathbf{u}_m + N^{\frac{1}{2}}\mathbf{o}_p(\sum_{j=1}^{m-1} \|\widehat{\boldsymbol{\beta}}_j - \boldsymbol{\beta}_{j,\mathbf{s}}\|) + \mathbf{o}_p(1) \tag{4.2.10}$$

Como $\mathbf{P}_{m+1}\mathbf{u}_m$ converge en distribución a una normal y usando la hipótesis inductiva obtenemos

$$N^{\frac{1}{2}}(\widehat{\boldsymbol{\beta}}_m - \boldsymbol{\beta}_{m,\mathbf{s}}) \;=\; N^{\frac{1}{2}}\mathbf{o}_p(\|\widehat{\boldsymbol{\beta}}_m - \boldsymbol{\beta}_{m,\mathbf{s}}\|) + O_p(1).$$

Esta última ecuación es análoga a la ecuación (4.2.9), pero para $N^{\frac{1}{2}}(\widehat{\boldsymbol{\beta}}_m - \boldsymbol{\beta}_{m,\mathbf{s}})$. Luego procediendo de igual forma que lo hecho para la expresión $N^{\frac{1}{2}}(\widehat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_{1,\mathbf{s}})$ obtenemos que $\widehat{\boldsymbol{\beta}}_m - \boldsymbol{\beta}_{m,\mathbf{s}} = O_p(N^{-\frac{1}{2}})$ como queríamos. □

Si despejamos $(\widehat{\boldsymbol{\beta}}_m - \boldsymbol{\beta}_{m,\mathbf{s}})$ en la igualdad (4.2.6) y usamos la propiedad anterior obtenemos la más importante de las identidades de esta demostración.Tenemos que

$$(\widehat{\boldsymbol{\beta}}_m - \boldsymbol{\beta}_{m,\mathbf{s}}) \;=\; \sum_{j=1}^{m-1} \mathbf{A}_m^{-1}\,\mathbf{B}_{jm}\,(\widehat{\boldsymbol{\beta}}_j - \boldsymbol{\beta}_{j,\mathbf{s}}) + N^{-\frac{1}{2}}\,\mathbf{A}_m^{-1}\,\mathbf{P}_{m+1}\,\mathbf{u}_m + \mathbf{o}_p(\,N^{-\frac{1}{2}}), \tag{4.2.11}$$

a partir de la cual obtenemos la expresión para $(\widehat{\boldsymbol{\beta}}_m - \boldsymbol{\beta}_{m,\mathbf{s}})$ enunciada en el Teorema 4.0.4.

Nos resta buscar la representación para $\widehat{\lambda}_{im} - \lambda_{im,\mathbf{s}}$. Como en Cui, He & Ng (2003) tenemos que, para cada i,

$$\varsigma_{n_i}^2(\widehat{\boldsymbol{\beta}}_m) - \varsigma_{n_i}^2(\boldsymbol{\beta}_{m,\mathbf{s}}) - \varsigma_i^2(\widehat{\boldsymbol{\beta}}_m) + \varsigma_i^2(\boldsymbol{\beta}_{m,\mathbf{s}}) = \mathbf{o}_p(n_i^{-\frac{1}{2}}),$$

de donde resulta que

$$\begin{aligned}(\widehat{\lambda}_{im} - \lambda_{im,\mathbf{s}}) \;&=\; \varsigma_{n_i}^2(\widehat{\boldsymbol{\beta}}_m) - \varsigma_i^2(\boldsymbol{\beta}_{m,\mathbf{s}}) = \varsigma_{n_i}^2(\boldsymbol{\beta}_{m,\mathbf{s}}) + \varsigma_i^2(\widehat{\boldsymbol{\beta}}_m) - 2\,\varsigma_i^2(\boldsymbol{\beta}_{m,\mathbf{s}}) + \mathbf{o}_p(n_i^{-\frac{1}{2}})\\ &=\; \varsigma_{n_i}^2(\boldsymbol{\beta}_{m,\mathbf{s}}) - \varsigma_i^2(\boldsymbol{\beta}_{m,\mathbf{s}}) + \varsigma_i^2(\widehat{\boldsymbol{\beta}}_m) - \varsigma_i^2(\boldsymbol{\beta}_{m,\mathbf{s}}) + \mathbf{o}_p(n_i^{-\frac{1}{2}})\end{aligned}$$

Ahora bien, haciendo un desarrollo de Taylor para $\varsigma_i^2(\widehat{\boldsymbol{\beta}}_m) - \varsigma_i^2(\boldsymbol{\beta}_{m,\mathbf{s}})$, tenemos que

$$(\widehat{\lambda}_{im} - \lambda_{im,\mathbf{s}}) = \varsigma_{n_i}^2(\boldsymbol{\beta}_{m,\mathbf{s}}) - \varsigma_i^2(\boldsymbol{\beta}_{m,\mathbf{s}}) + 2\,\varsigma_i(\boldsymbol{\beta}_{m,\mathbf{s}})\dot{\varsigma}^{\mathrm{T}}(\boldsymbol{\beta}_{m,\mathbf{s}})(\widehat{\boldsymbol{\beta}}_m - \boldsymbol{\beta}_{m,\mathbf{s}}) + \mathbf{o}_p(\,\|\widehat{\boldsymbol{\beta}}_m - \boldsymbol{\beta}_{m,\mathbf{s}}\|\,) + \mathbf{o}_p(n_i^{-\frac{1}{2}}).$$

Usando que $\mathbf{o}_p(\,\|\widehat{\boldsymbol{\beta}}_m - \boldsymbol{\beta}_{m,\mathbf{s}}\|\,) = \mathbf{o}_p(n_i^{-\frac{1}{2}})$ y que $\varsigma_i(\boldsymbol{\beta}_{m,\mathbf{s}}) = \sqrt{\lambda_{im,\mathbf{s}}}$ obtenemos

$$(\widehat{\lambda}_{im} - \lambda_{im,\mathbf{s}}) = \varsigma_{n_i}^2(\boldsymbol{\beta}_{m,\mathbf{s}}) - \varsigma_i^2(\boldsymbol{\beta}_{m,\mathbf{s}}) + 2\,\sqrt{\lambda_{im,\mathbf{s}}}\dot{\varsigma}_i^{\mathrm{T}}(\boldsymbol{\beta}_{m,\mathbf{s}})(\widehat{\boldsymbol{\beta}}_m - \boldsymbol{\beta}_{m,\mathbf{s}}) + \mathbf{o}_p(n_i^{-\frac{1}{2}})$$

Utilizando la representación para $\varsigma_{n_i}^2(\boldsymbol{\beta}_{m,\mathbf{s}}) - \varsigma_i^2(\boldsymbol{\beta}_{m,\mathbf{s}})$ dada en la identidad (4.1.3) obtenemos

$$\begin{aligned}(\widehat{\lambda}_{im} - \lambda_{im,\mathbf{s}}) \;&=\; 2\,\sqrt{\lambda_{im,\mathbf{s}}}\,n_i^{-1}\sum_{j=1}^{ni} h_i(\mathbf{x}_{ij};\boldsymbol{\beta}_{m,\mathbf{s}}) + 2\,\sqrt{\lambda_{im,\mathbf{s}}}\dot{\varsigma}_i^{\mathrm{T}}(\boldsymbol{\beta}_{m,\mathbf{s}})(\widehat{\boldsymbol{\beta}}_m - \boldsymbol{\beta}_{m,\mathbf{s}})\\ &+\; \mathbf{o}_p(n_i^{-\frac{1}{2}}).\end{aligned} \tag{4.2.12}$$

Además trabajando en forma análoga a lo hecho con ρ para lograr la identidad (4.2.7), pero ahora trabajando con ς_i se obtiene la siguiente expresión para $\dot{\varsigma}_i^{\mathrm{T}}(\boldsymbol{\beta}_{m,\mathbf{s}})(\widehat{\boldsymbol{\beta}}_m - \boldsymbol{\beta}_{m,\mathbf{s}})$

$$\dot{\varsigma}_i^{\mathrm{T}}(\boldsymbol{\beta}_{m,\mathbf{s}})(\widehat{\boldsymbol{\beta}}_m - \boldsymbol{\beta}_{m,\mathbf{s}}) = \sum_{j=1}^{m-1} < \dot{\varsigma}_i(\boldsymbol{\beta}_{m,\mathbf{s}}), \boldsymbol{\beta}_{j,\mathbf{s}} > \boldsymbol{\beta}_{j,\mathbf{s}}(\widehat{\boldsymbol{\beta}}_m - \boldsymbol{\beta}_{m,\mathbf{s}}).$$

Reemplazando esta última expresión en la identidad (4.2.12) obtenemos

$$(\widehat{\lambda}_{im} - \lambda_{im,\mathbf{s}}) = 2\sqrt{\lambda_{im,\mathbf{s}}}\left\{ n_i^{-1}\sum_{j=1}^{n_i} h_i(\mathbf{x}_{ij};\boldsymbol{\beta}_{m,\mathbf{s}}) + \sum_{j=1}^{m-1} < \dot{\varsigma}_i(\boldsymbol{\beta}_{m,\mathbf{s}}), \boldsymbol{\beta}_{j,\mathbf{s}} > \boldsymbol{\beta}_{j,\mathbf{s}}^{\mathrm{T}}(\widehat{\boldsymbol{\beta}}_m - \boldsymbol{\beta}_{m,\mathbf{s}}) \right\} + \mathbf{o}_p(n_i^{-\frac{1}{2}}).$$

De donde se deduce la expresión para $(\widehat{\lambda}_{im} - \lambda_{im,\mathbf{s}})$ enunciada en el Teorema 4.0.4.

Capítulo 5

Caso elíptico y unipoblacional

5.1. Caso elíptico

Bajo hipótesis de elipticidad sobre las poblaciones el desarrollo expuesto en el Teorema 4.0.3 puede simplificarse y coincide con la expansión dada por Boente Pires & Rodrigues (2002) utilizando funciones de influencia parciales. En esta sección, veremos que esto ocurre.

En efecto, supongamos que $\mathbf{x}_i$, $1 \leq i \leq k$, son vectores independientes con distribución elíptica F_i con parámetro de posición $\boldsymbol{\mu}_i = 0$ y matrices escala $\boldsymbol{\Sigma}_i = \mathbf{C}_i \mathbf{C}_i^{\mathrm{T}} = \boldsymbol{\beta} \boldsymbol{\Lambda}_i \boldsymbol{\beta}^{\mathrm{T}}$, o sea, $\boldsymbol{\Sigma}_i$ cumplen el modelo CPC. Supongamos además que

a) $\mathbf{x}_i = \mathbf{C}_i\, \mathbf{z}_i$ donde $\mathbf{z}_i$ tienen la misma distribución esférica G para todo $1 \leq i \leq k$.

b) $\mathbf{s}(G_0) = 1$, donde $G_0 = G[\mathbf{e}_1]$

c) $\boldsymbol{\Sigma} = \sum_{i=1}^{k} \tau_i \boldsymbol{\Sigma}_i$ no tiene autovalores múltiples, es decir, $\nu_1 > \ldots > \nu_p$ donde ν_i es el i-ésimo autovalor de $\boldsymbol{\Sigma}$.

d) La función $(\varepsilon, y) \to \mathbf{s}((1-\varepsilon)G_0 + \varepsilon \Delta_y)$ es dos veces diferenciable en $(0, y)$.

Bajo estos supuestos, $\boldsymbol{\beta}_{m,\mathbf{s}} = \boldsymbol{\beta}_m$, $\lambda_{im,\mathbf{s}} = \lambda_{im}$. Además tenemos

a) $\varsigma_i(\mathbf{a}) = \mathbf{s}(F_i[\mathbf{a}]) = \sqrt{\mathbf{a}^{\mathrm{T}} \boldsymbol{\Sigma}_i \mathbf{a}}$

b) $\dot{\varsigma}_i(\mathbf{a}) = \{\mathbf{a}^{\mathrm{T}}\, \boldsymbol{\Sigma}_i\, \mathbf{a}\}^{-\frac{1}{2}}\, \boldsymbol{\Sigma}_i\, \mathbf{a}$

c) $h_i(\mathbf{x}, \mathbf{a}) = \mathrm{IF}(\mathbf{x}, \mathbf{s}_{\mathbf{a}}; F_i) = \varsigma_i(\mathbf{a})\ \mathrm{IF}\left(\dfrac{\mathbf{x}^{\mathrm{T}}\mathbf{a}}{\varsigma_i(\mathbf{a})}, \mathbf{s}\,; G_0\right)$, ya que para todo $\mathbf{a}$ distinto de cero $\dfrac{\mathbf{x}_i^{\mathrm{T}}\mathbf{a}}{\sqrt{\mathbf{a}^{\mathrm{T}}\boldsymbol{\Sigma}_i\mathbf{a}}}$ tiene la misma distribución G_0.

d)

$$
\begin{aligned}
\dot{h}_i(\mathbf{x}, \mathbf{a}) &= \dot{\varsigma}_i(\mathbf{a})\mathrm{IF}\left(\frac{\mathbf{x}^{\mathrm{T}}\mathbf{a}}{\varsigma_i(\mathbf{a})}, \mathbf{s}\,; G_0\right) + \varsigma_i(\mathbf{a})\mathrm{DIF}\left(\frac{\mathbf{x}^{\mathrm{T}}\mathbf{a}}{\varsigma_i(\mathbf{a})}, \mathbf{s}; G_0\right)\left(\frac{\mathbf{x}\varsigma_i(\mathbf{a}) -\ \mathbf{x}^{\mathrm{T}}\mathbf{a}\,(\mathbf{a}^{\mathrm{T}}\boldsymbol{\Sigma}_i\mathbf{a})^{-\frac{1}{2}}\,\boldsymbol{\Sigma}_i\,\mathbf{a}}{\varsigma_i(\mathbf{a})^2}\right) \\
&= \dot{\varsigma}_i(\mathbf{a})\mathrm{IF}\left(\frac{\mathbf{x}^{\mathrm{T}}\mathbf{a}}{\varsigma_i(\mathbf{a})}, \mathbf{s}\,; G_0\right) + \varsigma_i(\mathbf{a})\mathrm{DIF}\left(\frac{\mathbf{x}^{\mathrm{T}}\mathbf{a}}{\varsigma_i(\mathbf{a})}, \mathbf{s}; G_0\right)\left[\mathbf{x} - \frac{\boldsymbol{\Sigma}_i\mathbf{a}\mathbf{a}^{\mathrm{T}}\mathbf{x}}{\varsigma_i(\mathbf{a})^2}\right],
\end{aligned}
$$

donde $\mathrm{DIF}(y, \mathbf{s}; \mathbf{G}_0)$ indica la derivada de $\mathrm{IF}(y, \mathbf{s}; \mathbf{G}_0)$ con respecto a y. Evaluando cada una de las expresiones en $\boldsymbol{\beta}_m$ obtenemos

a) $\varsigma_i(\boldsymbol{\beta}_m) = \sqrt{\lambda_{im}}$

b) $\dot{\varsigma}_i(\boldsymbol{\beta}_m) = \{\lambda_{im}\}^{-\frac{1}{2}}\,\lambda_{im}\,\boldsymbol{\beta}_m = \sqrt{\lambda_{im}}\;\boldsymbol{\beta}_m$

c) $h_i(\mathbf{x}, \boldsymbol{\beta}_m) = \sqrt{\lambda_{im}}\;\;\mathrm{IF}\left(\dfrac{\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}_m}{\sqrt{\lambda_{im}}}\;,\,\mathbf{s};\,G_0\right)$

d) $\dot{h}_i(\mathbf{x}, \boldsymbol{\beta}_m) = \sqrt{\lambda_{im}}\boldsymbol{\beta}_m\mathrm{IF}\left(\dfrac{\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}_m}{\sqrt{\lambda_{im}}}\;,\,\mathbf{s};\,G_0\right) + \mathrm{DIF}\left(\dfrac{\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}_m}{\sqrt{\lambda_{im}}}\;,\,\mathbf{s};\,G_0\right)\,(\mathbf{I}_p - \boldsymbol{\beta}_m\boldsymbol{\beta}_m^{\mathrm{T}})\mathbf{x}.$

Por (4.2.11) tenemos que

$$N^{\frac{1}{2}}(\widehat{\boldsymbol{\beta}}_m - \boldsymbol{\beta}_m) = \sum_{j=1}^{m-1} \mathbf{A}_m^{-1}\,\mathbf{B}_{jm}\,N^{\frac{1}{2}}\,(\widehat{\boldsymbol{\beta}}_j - \boldsymbol{\beta}_j) + \mathbf{A}_m^{-1}\,\mathbf{P}_{m+1}\,\mathbf{u}_m + \mathbf{o}_p(1).$$

Para reescribir esta igualdad bajo el supuesto elíptico calcularemos $\mathbf{A}_m^{-1}$, $\mathbf{A}_m^{-1}\mathbf{B}_{lm}$, y $\mathbf{A}_m^{-1}\,\mathbf{P}_{m+1}\,\mathbf{u}_m$ en función de las expresiones encontradas para $\varsigma_i(\boldsymbol{\beta}_m)$, su derivada y para h_i y su derivada.

Paso 1. Busco $\mathbf{A}_m^{-1}$

Recordemos que

$$\mathbf{A}_m = \mathbf{P}_{m+1}\ddot{\rho}(\boldsymbol{\beta}_m) - \boldsymbol{\beta}_m^{\mathrm{T}}\,\dot{\rho}(\boldsymbol{\beta}_m)\,\mathbf{I}_p - \sum_{j=1}^{m-1} \boldsymbol{\beta}_j^{\mathrm{T}}\,\dot{\rho}(\boldsymbol{\beta}_m)\,\boldsymbol{\beta}_m\,\boldsymbol{\beta}_j^{\mathrm{T}},$$

por lo tanto, para encontrar a $\mathbf{A}_m$ calculamos $\ddot{\rho}, \dot{\rho}, \rho$.

$\rho(\mathbf{a}) = \sum_{i=1}^{k} \tau_i\varsigma_i^2(\mathbf{a}) = \sum_{i=1}^{k} \tau_i\,\mathbf{a}^{\mathrm{T}}\boldsymbol{\Sigma}_i\mathbf{a} = \mathbf{a}^{\mathrm{T}}\,\Sigma\,\mathbf{a}$

$\dot{\rho}(\mathbf{a}) = 2\sum_{i=1}^{k}\;\tau_i\,\boldsymbol{\Sigma}_i\mathbf{a} = 2\,\boldsymbol{\Sigma}\mathbf{a}$

$\ddot{\rho}(\mathbf{a}) = 2\boldsymbol{\Sigma}$

Evaluando estas funciones en $\boldsymbol{\beta}_m$ obtenemos

$\rho(\boldsymbol{\beta}_m) = \sum_{i=1}^{k} \tau_i\lambda_{im} = \nu_m$

$\dot{\rho}(\boldsymbol{\beta}_m) = 2\nu_m\boldsymbol{\beta}_m$

$\ddot{\rho}(\boldsymbol{\beta}_m) = 2\boldsymbol{\Sigma}$

con lo cual, como $\|\boldsymbol{\beta}_m\| = 1$ $\boldsymbol{\beta}_j^{\mathrm{T}}\boldsymbol{\beta}_m = 0$ $j \neq m$, deducimos que

$$\mathbf{A}_m = \mathbf{P}_{m+1}\,2\boldsymbol{\Sigma} - \boldsymbol{\beta}_m^{\mathrm{T}}\,2\,\nu_m\,\boldsymbol{\beta}_m\mathbf{I}_p - 2\sum_{j=1}^{m-1}\boldsymbol{\beta}_j^{\mathrm{T}}\,\nu_m\,\boldsymbol{\beta}_m\boldsymbol{\beta}_m\boldsymbol{\beta}_j^{\mathrm{T}} = 2\sum_{i=1}^{k}\tau_i(\,\mathbf{P}_{m+1}\,\boldsymbol{\Sigma}_i - \lambda_{im}\mathbf{I}_p).$$

Ahora bien, usando que $\Sigma_i = \sum_{j=1}^{p} \lambda_{ij}\,\boldsymbol{\beta}_j\,\boldsymbol{\beta}_j^{\mathrm{T}}$, $\mathbf{I}_p = \sum_{j=1}^{p}\boldsymbol{\beta}_j\,\boldsymbol{\beta}_j^{\mathrm{T}}$ y que $\mathbf{P}_{m+1}$ es la matriz de proyección sobre $< \boldsymbol{\beta}_1, \ldots \boldsymbol{\beta}_m >^{\perp}$, resulta que $\mathbf{P}_{m+1}\,\boldsymbol{\Sigma}_i - \lambda_{im}\,\mathbf{I}_p = \sum_{j=m+1}^{p}(\lambda_{ij} - \lambda_{im})\,\boldsymbol{\beta}_j\,\boldsymbol{\beta}_j^{\mathrm{T}} - \sum_{j=1}^{m}\lambda_{im}\,\boldsymbol{\beta}_j\,\boldsymbol{\beta}_j^{\mathrm{T}}$, de donde

$$\mathbf{A}_m = 2\sum_{i=1}^{k}\left[\tau_i\sum_{j=m+1}^{p}(\lambda_{ij} - \lambda_{im})\,\boldsymbol{\beta}_j\,\boldsymbol{\beta}_j^{\mathrm{T}} - \tau_i\sum_{j=1}^{m}\lambda_{im}\,\boldsymbol{\beta}_j\,\boldsymbol{\beta}_j^{\mathrm{T}}\right].$$

Cambiando el orden de las sumas obtenemos

$$\mathbf{A}_m = 2\sum_{j=m+1}^{p}\sum_{i=1}^{k}\tau_i(\lambda_{ij}-\lambda_{im})\boldsymbol{\beta}_j\,\boldsymbol{\beta}_j^{\mathrm{T}} - 2\sum_{j=1}^{m}\sum_{i=1}^{k}\tau_i\lambda_{im}\boldsymbol{\beta}_j\,\boldsymbol{\beta}_j^{\mathrm{T}},$$

es decir, recordando que $\nu_m = \sum_{i=1}^{k}\tau_i\,\lambda_{im}$ tenemos que

$$\mathbf{A}_m = 2\sum_{j=m+1}^{p}(\nu_j-\nu_m)\boldsymbol{\beta}_j\,\boldsymbol{\beta}_j^{\mathrm{T}} - 2\sum_{j=1}^{m}\nu_m\boldsymbol{\beta}_j\,\boldsymbol{\beta}_j^{\mathrm{T}}$$

de donde por la escritura de $\mathbf{A}_m$ y el vínculo entre las $\mathbf{s}$-direcciones principales comunes obtenemos

$$\mathbf{A}_m^{-1} \;=\; \frac{1}{2}\sum_{j=m+1}^{p}\frac{1}{(\nu_j-\nu_m)}\boldsymbol{\beta}_j\,\boldsymbol{\beta}_j^{\mathrm{T}} - \frac{1}{2}\sum_{j=1}^{m}\frac{1}{\nu_m}\boldsymbol{\beta}_j\,\boldsymbol{\beta}_j^{\mathrm{T}}. \tag{5.1.1}$$

Paso 2. Calculemos $\mathbf{A}_m^{-1}\,\mathbf{B}_{jm}$ para $j \le m-1$

$$\mathbf{B}_{jm} = \boldsymbol{\beta}_j^{\mathrm{T}}\,2\,\nu_m\,\boldsymbol{\beta}_m\,\mathbf{I}_p + 2\,\boldsymbol{\beta}_j\,\boldsymbol{\beta}_m^{\mathrm{T}}\,\nu_m \quad j \le m-1.$$

Como $j \le m-1$ entonces $\boldsymbol{\beta}_j{}^{\mathrm{T}}\boldsymbol{\beta}_m = 0$, por lo tanto

$$\mathbf{B}_{jm} = 2\,\nu_m\,\boldsymbol{\beta}_j\,\boldsymbol{\beta}_m^{\mathrm{T}} \quad j \le m-1.$$

Usando esta última expresión y (5.1.1) obtenemos

$$\mathbf{A}_m^{-1}\mathbf{B}_{jm} = -\boldsymbol{\beta}_j\,\boldsymbol{\beta}_m^{\mathrm{T}} \quad j \le m-1. \tag{5.1.2}$$

Paso 3. Calculemos $\mathbf{A}_m^{-1}\mathbf{P}_{m+1}\mathbf{u}_m$. Para eso, recordando la definición de $\mathbf{u}_m$ calcularemos

a) $2\,\mathbf{A}_m^{-1}\mathbf{P}_{m+1}\varsigma_i(\boldsymbol{\beta}_m)\dot{h}_i(\mathbf{x}_{ij},\,\boldsymbol{\beta}_m)$

b) $2\,\mathbf{A}_m^{-1}\mathbf{P}_{m+1}\dot{\varsigma}_i(\boldsymbol{\beta}_m)h_i(\mathbf{x}_{ij},\,\boldsymbol{\beta}_m)$

Calculemos a). Por la definición de $\mathbf{P}_{m+1}$ y por (5.1.1) tenemos que $2\,\mathbf{A}_m^{-1}\,\mathbf{P}_{m+1}\,\varsigma_i(\boldsymbol{\beta}_m)\,\dot{h}_i(\mathbf{x}_{ij},\,\boldsymbol{\beta}_m) = \sqrt{\lambda_{im}}\,\sum_{\ell=m+1}^{p}\dfrac{1}{(\nu_\ell-\nu_m)}\,\boldsymbol{\beta}_\ell\,\boldsymbol{\beta}_\ell^{\mathrm{T}}\,\dot{h}_i(\mathbf{x}_{ij},\boldsymbol{\beta}_m)$. Luego, recordando la expresión hallada para $\dot{h}_i(\mathbf{x}_{ij},\boldsymbol{\beta}_m)$ obtenemos

$$2\,\mathbf{A}_m^{-1}\,\mathbf{P}_{m+1}\,\varsigma_i(\boldsymbol{\beta}_m)\,\dot{h}_i(\mathbf{x}_{ij},\,\boldsymbol{\beta}_m) = \sqrt{\lambda_{im}}\,\mathrm{DIF}\left(\frac{\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta}_m}{\sqrt{\lambda_{im}}}\,,\,\mathbf{s};\,G_0\right)\sum_{\ell=m+1}^{p}\frac{1}{(\nu_\ell-\nu_m)}\,(\mathbf{x}_{ij}^{\mathrm{T}}\,\boldsymbol{\beta}_\ell)\,\boldsymbol{\beta}_\ell. \tag{5.1.3}$$

Calculemos b) $2\,\mathbf{A}_m^{-1}\mathbf{P}_{m+1}\dot{\varsigma}_i(\boldsymbol{\beta}_m)h_i(\mathbf{x}_{ij};\,\boldsymbol{\beta}_m) = 2\,\mathbf{A}_m^{-1}\,\mathbf{P}_{m+1}\,\sqrt{\lambda_{im}}\,\boldsymbol{\beta}_m\,\sqrt{\lambda_{im}}\mathrm{IF}\left(\dfrac{\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta}_m}{\sqrt{\lambda_{im}}}\,,\,\mathbf{s};\,G_0\right) = 0$, pues $\mathbf{P}_{m+1}\boldsymbol{\beta}_m = 0$. Recordando la definición de $\mathbf{u}_m$ y usando también (5.1.3) tenemos

$$\mathbf{A}_m^{-1}\,\mathbf{P}_{m+1}\,\mathbf{u}_m = -N^{-\frac{1}{2}}\sum_{i=1}^{k}\sum_{j=1}^{n_i}\sqrt{\lambda_{im}}\,\mathrm{DIF}\left(\frac{\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}_m}{\sqrt{\lambda_{im}}}\,,\,\mathbf{s};\,G_0\right)\sum_{\ell=m+1}^{p}\frac{1}{(\nu_\ell-\nu_m)}(\mathbf{x}_{ij}^{\mathrm{T}}\,\boldsymbol{\beta}_\ell)\,\boldsymbol{\beta}_\ell. \tag{5.1.4}$$

De (4.2.11) y usando las identidades (5.1.2) y (5.1.4) tenemos que

$$\begin{aligned}
N^{\frac{1}{2}}(\widehat{\boldsymbol{\beta}}_m - \boldsymbol{\beta}_m) &= \sum_{\ell=1}^{m-1} \mathbf{A}_m^{-1}\,\mathbf{B}_{\ell m}\,N^{-\frac{1}{2}}\,(\widehat{\boldsymbol{\beta}}_\ell - \boldsymbol{\beta}_\ell) + \mathbf{A}_m^{-1}\,\mathbf{P}_{m+1}\,\mathbf{u}_m + \mathbf{o}_p(1) = -\sum_{\ell=1}^{m-1} \boldsymbol{\beta}_\ell\,\boldsymbol{\beta}_m^{\mathrm{T}}\,N^{\frac{1}{2}}(\widehat{\boldsymbol{\beta}}_\ell - \boldsymbol{\beta}_\ell) \\
&- N^{-\frac{1}{2}}\sum_{i=1}^{k}\sum_{j=1}^{n_i}\sqrt{\lambda_{im}}\,\mathrm{DIF}\left(\frac{\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta}_m}{\sqrt{\lambda_{im}}}\,,\,\mathbf{s};\,G_0\right)\sum_{\ell=m+1}^{p}\frac{1}{(\nu_\ell - \nu_m)}\,(\mathbf{x}_{ij}^{\mathrm{T}}\,\boldsymbol{\beta}_\ell)\,\boldsymbol{\beta}_\ell + \mathbf{o}_p(1),
\end{aligned}$$

de donde recursivamente se obtiene que,

$$\begin{aligned}
N^{\frac{1}{2}}\,(\widehat{\boldsymbol{\beta}}_m - \boldsymbol{\beta}_m) &= N^{-\frac{1}{2}}\sum_{i=1}^{k}\sum_{j=1}^{n_i}\sum_{\ell=1}^{m-1}\sqrt{\lambda_{i\ell}}\,\mathrm{DIF}\left(\frac{\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta}_\ell}{\sqrt{\lambda_{i\ell}}}\,,\,\mathbf{s}\,;\,G_0\right)\frac{1}{(\nu_m - \nu_\ell)}(\mathbf{x}_{ij}^{\mathrm{T}}\,\boldsymbol{\beta}_\ell)\,\boldsymbol{\beta}_\ell \qquad (5.1.5) \\
&+ N^{-\frac{1}{2}}\sum_{i=1}^{k}\sum_{j=1}^{n_i}\sum_{l=m+1}^{p}\sqrt{\lambda_{im}}\,\mathrm{DIF}\left(\frac{\mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta}_m}{\sqrt{\lambda_{im}}}\,,\,\mathbf{s}\,;\,G_0\right)\frac{1}{(\nu_m - \nu_\ell)}(\mathbf{x}_{ij}^{\mathrm{T}}\,\boldsymbol{\beta}_\ell)\,\boldsymbol{\beta}_\ell + \mathbf{o}_p(1).
\end{aligned}$$

La representación para $N^{-\frac{1}{2}}(\widehat{\boldsymbol{\beta}}_m - \boldsymbol{\beta}_m)$ presentada en el Teorema 4.0.4 coincide con la expansión obtenida por Boente Pires & Rodrigues (2002), utilizando funciones de influencia parciales. El desarrollo para los autovectores es inmediato.

5.2. Caso unipoblacional

Si consideramos el caso $k = 1$, es decir, trabajamos con una sola población, resulta que los funcionales dados en la Definición 2.4.2 y sus respectivos estimadores presentados en la Definición 2.4.3, coinciden con los funcionales y los estimadores dados en la Definición 3.3.2 y en la Definición 3.3.1, respectivamente. Es decir, las $\mathbf{s}$-direcciones principales coinciden con las $\mathbf{s}$-direcciones principales comunes, y lo mismo ocurre con sus respectivos estimadores. Por lo tanto el desarrollo expuesto en el Teorema 2.4.4 debe coincidir con el desarrollo dado por el Teorema 4.0.4. En esta sección mostraremos que, en efecto, los desarrollos coinciden.

Siguiendo a Cui, He & Ng (2003) en su demostración del Teorema 2.4.4, puede verse que para $1 \leq m \leq q$

$$N^{-\frac{1}{2}}(\widehat{\boldsymbol{\beta}}_m - \boldsymbol{\beta}_{m,\mathbf{s}}) = \sum_{j=0}^{m-1}\widetilde{\mathbf{A}}_m^{-1}\widetilde{\mathbf{B}}_{jm}N^{-\frac{1}{2}}(\widehat{\boldsymbol{\beta}}_j - \boldsymbol{\beta}_{j,\mathbf{s}}) + \widetilde{\mathbf{A}}_m^{-1}\widetilde{\mathbf{P}}_{m+1}\widetilde{\mathbf{u}}_m, \qquad (5.2.1)$$

recordando que $\widetilde{\mathbf{A}}_m, \widetilde{\mathbf{B}}_{jm}, \widetilde{\mathbf{P}}_{m+1}$ y $\widetilde{\mathbf{u}}_m$ están definidos antes de el enunciado del Teorema 2.4.4. Posteriormente, en (4.2.11) llegamos a que

$$N^{\frac{1}{2}}(\widehat{\boldsymbol{\beta}}_m - \boldsymbol{\beta}_{m,\mathbf{s}}) = \sum_{j=1}^{m-1}\mathbf{A}_m^{-1}\,\mathbf{B}_{jm}\,N^{\frac{1}{2}}\,(\widehat{\boldsymbol{\beta}}_j - \boldsymbol{\beta}_{j,\mathbf{s}}) + \mathbf{A}_m^{-1}\,\mathbf{P}_{m+1}\,\mathbf{u}_m + \mathbf{o}_p(1).$$

Veremos que ambas recurrencias coinciden, con lo cual, debido a que de las mismas se deducen respectivamente los desarrollos dados por los Teoremas 2.4.4 y 4.0.4, tendremos que dichos desarrollos coinciden. Para lograr nuestros fines transformaremos la identidad 4.2.11 en la identidad 5.2.1.

Si $k = 1$, tenemos que $\rho(\mathbf{a}) = \tau_1\,\varsigma_1^2(\mathbf{a}) = \varsigma^2(\mathbf{a})$ ya que $\tau_1 = 1$ y $\varsigma(\boldsymbol{\beta}_{m,\mathbf{s}}) = \lambda_{m,\mathbf{s}}{}^{\frac{1}{2}}$. También tenemos que $N = n_1 = n$ y que $\rho_N(\mathbf{a}) = \tau_1\,\varsigma_{n_1}^2(\mathbf{a}) = \varsigma_n^2(\mathbf{a})$. Además como los funcionales coinciden resulta que $\mathbf{P}_{m+1} = \widetilde{\mathbf{P}}_{m+1}$. Veamos que $\mathbf{A}_m = 2\,\lambda_{m,\mathbf{s}}{}^{\frac{1}{2}}\,\widetilde{\mathbf{A}}_m$ y $\mathbf{B}_{jm} = 2\,\lambda_{m,\mathbf{s}}{}^{\frac{1}{2}}\,\widetilde{\mathbf{B}}_{jm}$. Para ello notemos que

$$\dot{\boldsymbol{\rho}}(\mathbf{a}) = 2\,\varsigma(\mathbf{a})\dot{\varsigma}(\mathbf{a}) \quad \text{luego} \quad \dot{\boldsymbol{\rho}}(\boldsymbol{\beta}_{m,\mathbf{s}}) = 2\,\lambda_{m,\mathbf{s}}{}^{\frac{1}{2}}\,\dot{\varsigma}(\boldsymbol{\beta}_{m,\mathbf{s}})$$

$$\ddot{\boldsymbol{\rho}}(\mathbf{a}) = 2\,\{\dot{\varsigma}^2(\mathbf{a}) + \varsigma(\mathbf{a})\ddot{\varsigma}(\mathbf{a})\ \} \text{ luego } \ddot{\boldsymbol{\rho}}(\boldsymbol{\beta}_m) = 2\,\{\dot{\varsigma}^2(\boldsymbol{\beta}_{m,\mathbf{s}}) + \lambda_{m,\mathbf{s}}{}^{\frac{1}{2}}\,\ddot{\varsigma}(\boldsymbol{\beta}_{m,\mathbf{s}})\},$$

en vista de lo cual, como $\mathbf{P}_{m+1}\dot{\varsigma}(\boldsymbol{\beta}_{m,\mathbf{s}}) = 0$, resulta que $\mathbf{P}_{m+1}\ \ddot{\rho}(\boldsymbol{\beta}_{m,\mathbf{s}}) = \lambda_{m,\mathbf{s}}{}^{\frac{1}{2}}\,\mathbf{P}_{m+1}\ddot{\varsigma}(\boldsymbol{\beta}_m)$. Luego, reemplazando obtenemos

$$\boldsymbol{\beta}^{\mathrm{T}}_{m,\mathbf{s}}\ \dot{\rho}(\boldsymbol{\beta}_m)\mathbf{I}_p = 2\lambda_{m,\mathbf{s}}{}^{\frac{1}{2}}\,(\boldsymbol{\beta}_{m,\mathbf{s}})^{\mathrm{T}}\dot{\varsigma}(\boldsymbol{\beta}_{m,\mathbf{s}})\mathbf{I}_p$$

$$\boldsymbol{\beta}^{\mathrm{T}}_{j,\mathbf{s}}\ \dot{\rho}(\boldsymbol{\beta}_{m,\mathbf{s}})\boldsymbol{\beta}_{m,\mathbf{s}}\ \boldsymbol{\beta}^{\mathrm{T}}_{j,\mathbf{s}} = 2\,\lambda_{m,\mathbf{s}}{}^{\frac{1}{2}}\,\boldsymbol{\beta}^{\mathrm{T}}_{j,\mathbf{s}}\dot{\varsigma}(\boldsymbol{\beta}_{m,\mathbf{s}})(\boldsymbol{\beta}_{m,\mathbf{s}})(\boldsymbol{\beta}_{j,\mathbf{s}})^{\mathrm{T}},$$

de donde, $\mathbf{A}_m = 2\,\lambda_{m,\mathbf{s}}{}^{\frac{1}{2}}\,\widetilde{\mathbf{A}}_m$ y $\mathbf{B}_{jm} = 2\,\lambda_{m,\mathbf{s}}{}^{\frac{1}{2}}\,\widetilde{\mathbf{B}}_{jm}$, con lo cual $\mathbf{A}_m^{-1}\mathbf{B}_{jm} = \widetilde{\mathbf{A}}_m^{-1}\widetilde{\mathbf{B}}_{jm}$. Luego los primeros sumandos de cada recurrencia coinciden. Además para $k = 1$ resulta que

$$\mathbf{u}_m = n^{-\frac{1}{2}}\{2\,\lambda_{m,\mathbf{s}}{}^{\frac{1}{2}}\,\sum_{j=1}^{n}\mathbf{h}^{\star}(\mathbf{x}_i;\boldsymbol{\beta}_{m,\mathbf{s}}) + 2\dot{\varsigma}(\boldsymbol{\beta}_{m,\mathbf{s}})\sum_{j=1}^{n}h(\mathbf{x}_i;\boldsymbol{\beta}_{m,\mathbf{s}})\}.$$

Nuevamente, recordando que $\mathbf{P}_{m+1}\dot{\varsigma}(\boldsymbol{\beta}_{m,\mathbf{s}}) = 0$, obtenemos $\mathbf{P}_{m+1}\,\mathbf{u}_m = \lambda_{m,\mathbf{s}}{}^{\frac{1}{2}}\,\mathbf{P}_{m+1}\,\widetilde{\mathbf{u}}_m$ y luego $\mathbf{A}_m^{-1}\mathbf{P}_{m+1}\,\mathbf{u}_m$ $\widetilde{\mathbf{A}}_m^{-1}\mathbf{P}_{m+1}\,\widetilde{\mathbf{u}}_m$, lo que prueba, finalmente, que las recurrencias coinciden, como queríamos demostrar.

Bibliografía

[1] Anderson, T. W. (1963). Asymptotic Theory for Principal Component Analysis. *Annals Mathematical Statistics*, **34**, 122-148.

[2] Arnold, S. J. & Phillips, P. C. (1999). Hierarchical comparison of genetic variance-covariance matrices. II. Coastal-inland divergence in the garter snake *Thamnophis elegans*. *Evolution*, **53**, 1516-27.

[3] Berrendero, J. R. (1996). *Contribuciones a la Teoría de la Robustez Respecto al Sesgo*. Unpublished PhD Thesis (in spanish), Universidad Carlos III de Madrid.

[4] Boente, G. (1987). Asymptotic Theory for Robust Principal Components. *Journal of Multivariate Analysis*, **21**, 67-78.

[5] Boente, G. & Orellana, L. (2001). A Robust Approach to Common Principal Components. In *Statistics in Genetics and in the Environmental Sciences*, eds. L. T. Fernholz, S. Morgenthaler, and W. Stahel, pp. 117-147. Basel: Birkhauser.

[6] 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.

[7] Boente, G., Pires, A. M. & Rodrigues, I. M. (2007). Robust tests for the common principal components model. Available at `http://www.ic.fcen.uba.ar/preprints/boentepiresrodrigues2.pdf`.

[8] Croux C., Filzmoser P,. & Oliveira M.R. (2007). Algorithms for Projection-Pursuit Robust Principal Component Analysis. *Chemometrics and Intelligent Laboratory Systems*, **87**, 218-225.

[9] Croux, C. & Haesbroeck, G. (2000). Principal component analysis based on robust estimators of the covariance or correlation matrix: Influence functions and efficiencies. *Biometrika*, **87**, 603–618.

[10] Croux, C. & Ruiz–Gazen, A. (1996). A Fast Algorithm for Robust Principal Components based on Projection Pursuit. In *Compstat: Proceedings in Computational Statistics*, ed. A. Prat, Heidelberg: Physica–Verlag, 211- 217.

[11] Croux, C. & Ruiz–Gazen, A. (2005). High Breakdown Estimators for Principal Components: the Projection–Pursuit Approach Revisited. *Journal of Multivariate Analysis*, **95**, 206-226.

[12] Cui, H., He, X. & Ng, K. W. (2003). Asymptotic Distribution of Principal Components Based on Robust Dispersions. *Biometrika*, **90**, 953-966.

[13] Davies PL (1987). Asymptotic behavior of $S-$estimates of multivariate location parameters and dispersion matrices. *Annals of Statistics*, **15**, 1269-1292.

[14] Devlin, S. J., Gnanadesikan, R. & Kettenring, J. R. (1981). Robust Estimation of dispersion Matrices and Principal Components. *Journal of the American Statistical Association*, **76**, 354-362.

[15] Donoho, D. L. (1982). *Breakdown properties of multivariate location estimators*, Qualifying paper, Harvard University, Boston.

[16] Flury, B. K. (1984). Common Principal Components in k Groups. *Journal of the American Statistical Association*, **79**, 892-898.

[17] Flury, B. K. (1986). Asymptotic Theory for Common Principal Component Analysis. *Annals of Statistics*, **14**, 418-430.

[18] Flury, B. K. (1987). A Hierarchy of Relationships Between Covariance Matrices. In *Advances in Multivariate Statistical Analysis* (A.K. Gupta, ed.), 31-43, Boston: Reidel.

[19] Flury, B. K. (1988). *Common Principal Components and Related Multivariate Models*. New York: John Wiley.

[20] Flury, B. K. & Gautschi, W. (1986). An Algorithm for Simultaneous Orthogonal Transformation of Several Positive Definite Symmetric Matrices to Nearly Diagonal Form. *SIAM J. Scient. Statist. Computing*, **7**, 169-184.

[21] Huber, P. (1981). *Robust Statistics*. New York: John Wiley.

[22] Kent, J. & Tyler, D. (1996) Constrained $M-$estimation for multivariate location and scatter. *Annals of Statistics*, **24**, 1346-1370.

[23] Klingerberg, C. P. & Neuenschwander, B. E. (1996). Ontogeny and individual variation: analysis of patterned covariance matrices with common principal components. *Systematic Biology*, **45**, 135-7.

[24] Kshirsagar, A. (1972). *Multivariate Analysis*. Marcel Dekker, New York.

[25] 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.

[26] Locantore, N., Marron, J. S., Simpson, D. G., Tripoli, N., Zhang, J. T. & Cohen, K. L. (1999). Robust Principal Components for functional Data. *Test*, **8**, 1-28.

[27] Maronna, R. (1976). Robust $M-$estimates of multivariate location and scatter. *Annals of Statistics*, **4**, 51-67.

[28] Maronna, R., Martin, D. and Yohai, V. (2006). *Robust Statistics: Theory and Methods*. New York: John Wiley & Sons.

[29] Muirhead, R. (1982). *Aspects of Multivariate Analysis*. Wiley, New York.

[30] Novi Inverardi, P. L. and Flury, B. K. (1992). Robust Estimation of Common Principal Components. *Quaderni di Satistica e Matematica Applicata alle Scienze Economico–Social*, **14**, 49-79.

[31] Patak, Z. (1991). *Robust Principal Components*. Master Thesis of the Department of Statistics, University of British Columbia, Vancouver.

[32] Phillips, P. C. & Arnold, S. J. (1999). Hierarchical comparison of genetic variance-covariance matrices. I. Using the Flury hierarchy. *Evolution*, **53**, 1506-15.

[33] Rousseeuw, P. J. (1985). Multivariate estimation with high breakdown point. In: *Mathematical Statistics and Applications*, W. Grossmann, G. Pflug, I. Vincze and W. Wertz (eds.). Reidel, Dordrecht, pp.283-297.

[34] Stahel, W. A. (1981). *Robuste schätzungen: infinitesimale optimalität und schätzungen von Kovarianzmatrizen*, PhD thesis, ETH Zürich.

[35] Tyler, D. E. (1983). The Asymptotic Distribution of Principal Components Roots Under Local Alternative to Multiple Roots. *Annals of Statistics*, **11**, 1232-1242.

[36] Xie, Y. L., Wang, J. H., Liang, Y. Z., Sun, L. X. Song, X. H. & Yu, R. Q. (1993). Robust Principal Component Analysis by Projection Pursuit. *Journal of Chemometrics*, **7**, 527-541.

[37] Pollard, D. (1984). Convergence of Stochastic Processes. Springer-Verlag

[38] Van der Vaart, A. & Wellner, J.(1996). Weak Convergence and Empirical Processes: With Applications to Statistics (Springer Series in Statistics) (Hardcover)