



Universidad de Buenos Aires
Facultad de Ciencias Exactas y Naturales

Biblioteca Digital FCEN-UBA

Métodos robustos de estimación de componentes principales funcionales y el modelo de componentes principales comunes

Bali, Juan Lucas
2012

Tesis Doctoral

Facultad de Ciencias Exactas y Naturales
Universidad de Buenos Aires

www.digital.bl.fcen.uba.ar

Contacto: digital@bl.fcen.uba.ar

Este documento forma parte de la colección de tesis doctorales de la Biblioteca Central Dr. Luis Federico Leloir. Su utilización debe ser acompañada por la cita bibliográfica con reconocimiento de la fuente.

This document is part of the doctoral theses collection of the Central Library Dr. Luis Federico Leloir. It should be used accompanied by the corresponding citation acknowledging the source.

Fuente / source:

Biblioteca Digital de la Facultad de Ciencias Exactas y Naturales - Universidad de Buenos Aires



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

Métodos robustos de estimación de componentes principales funcionales y el modelo de componentes principales comunes

Tesis presentada para optar al título de Doctor de la Universidad de Buenos Aires en el área Ciencias Matemáticas

Juan Lucas Bali

Director de tesis: Dra. Graciela Boente
Consejero de estudios: Dra. Graciela Boente
Lugar de Trabajo: Instituto de Cálculo, FCEN-UBA.

Buenos Aires, Mayo 2012

Resumen

En muchas situaciones prácticas, los datos se registran sobre un período de tiempo o corresponden a imágenes. Por esta razón, conviene considerarlos como realizaciones de un proceso estocástico en lugar de discretizarlos y estudiarlos como realizaciones de datos multivariados.

Un método ampliamente usado para describir los principales modos de variación de las observaciones es el de componentes principales funcionales. Sin embargo, los procedimientos clásicos basados en el desvío estándar o el operador de covarianza muestral son muy sensibles a datos atípicos. Para resolver este problema, en esta tesis, se introducen estimadores robustos para las componentes principales mediante un enfoque basado en el método de *projection-pursuit*. Estimadores de *projection-pursuit*, para el caso finito-dimensional, fueron inicialmente propuestos en Huber (1981) como una forma de obtener estimadores robustos de la matriz de covarianza. Posteriormente, fueron considerados por Li y Chen (1985) para el problema de componentes principales como una aproximación directa al problema sin necesidad de obtener estimadores robustos de la matriz de escala (ver, también Huber, 1985).

En esta tesis, se extiende dicha propuesta al caso funcional. Por otra parte, se consideran diversas variaciones de la misma que permiten obtener estimadores suaves mediante una penalización en la escala o en la norma, o bien, mediante una aproximación finito-dimensional sobre una base con elementos suaves. Bajo condiciones débiles de regularidad, se obtiene la continuidad del funcional asociado, de lo que se deduce la consistencia de los estimadores sin suavizar. Por otra parte, se prueba la consistencia de los estimadores que penalizan la norma y/o la escala así como la de una propuesta basada en aproximaciones por subespacios crecientes, bajo condiciones débiles. Mediante un estudio de simulación, se compara la performance de las propuestas clásicas y robustas bajo diferentes tipos de contaminación.

Se complementa el estudio de los estimadores de proyección de las direcciones principales, obteniendo una expresión para la función de influencia del funcional asociado a la dirección principal y su tamaño.

Las propuestas anteriores de estimación se extendieron al caso de k poblaciones que cumplen un modelo de componentes principales comunes funcional. Bajo dicho modelo, las diversas poblaciones presentan una estructura común en cuanto a su operador de dispersión compartiendo sus autofunciones y el orden de las mismas. Para ser más precisos, se definen estimadores robustos de *projection-pursuit* de las direcciones principales bajo un modelo de componentes principales comunes. Para las distintas propuestas, algunas de las cuales incluyen como antes diversos tipos de penalización, se prueba su consistencia, bajo condiciones débiles. Por otra parte, mediante un estudio de Monte Carlo, se estudia el comportamiento de los estimadores introducidos para muestras finitas, bajo diferentes modelos y tipos de contaminación.

Finalmente, se proponen algoritmos que permiten calcular aproximaciones de cada una de las propuestas dadas. Se prueba la consistencia del estimador obtenido a través de dicho algoritmo. Mediante un estudio Monte Carlo, se analiza la velocidad de convergencia de dicha propuesta en el caso de procesos con trayectorias suaves e irregulares. Para estas últimas, se propone un método alternativo de cálculo.

Palabras clave: Análisis de Componentes Principales; Componentes Principales Comunes; Datos Funcionales; Estimadores Robustos; Función de Influencia; Penalización.

Robust methods for estimation of functional principal components and for the model of common principal components

Abstract

In many practical situations, data are recorded over a period of time or may correspond to images. In these situations, it is preferable to consider them as samples of an stochastic process instead of discretizing and studying them as multivariate data samples. A technique widely used to describe the observations principal modes of variations is the functional principal components analysis. The classical procedures based on the standard deviation or the sample covariance are very sensitive to atypical data. To overcome this problem, in this thesis, we introduce robust estimators for the principal components using a projection–pursuit approach. In the multivariate setting, projection–pursuit estimators were initially proposed in Huber(1981) as a way to obtain robust estimators of the covariance matrix. They were then considered by Li and Chen (1985) to perform a robust principal component analysis without passing by a robust estimate of the covariance matrix (see also, Huber, 1985).

In this thesis, we extend that proposal to the functional case. On the other hand, we consider other alternatives which allow us to obtain smooth estimators of the principal directions using either a penalization on the scale or the norm or through a sieve approach which consists in considering an increasing sequence of finite–dimensional spaces spanned by a basis with smooth elements. Under mild regularity conditions, we derive the continuity of the associated functional which entails the consistency of the non–smoothed estimators. On the other hand, we show that the estimators defined penalizing the norm and/or the scale and those based on the sieve approach are consistent. Through a simulation study, we compare the performance of the classical and robust proposals under different contamination models.

We complete the study of the projection–based estimators of principal directions, obtaining an expression for the influence function of the functional related to the principal direction and its size.

The previous proposals are extended to k populations satisfying a functional common principal component model. Under this model, the populations have a common dispersion structure, since the dispersion operators share the same eigenfunctions and their order. More precisely, we introduce robust projection–pursuit estimators of the principal directions under functional common principal components model. We derive the consistency of the the different proposals, some of which include penalization terms. On the other hand, through a Monte Carlo study, we the performance of the robust estimators is compared to that of the classical one, under different models and contaminations.

Finally, we propose algorithms that allow to compute approximations for each of the robust proposals given. The estimators obtained through that algorithm turn out to be consistent. Using a Monte Carlo study, we analyse the convergence rate of the proposals when the observations have smooth or rugged trajectories. For the latter, we propose an alternative method of computation.

Keywords: Common Principal Components Model; Functional Data; Influence Function; Penalization; Principal Component Analysis; Robust Estimators.

Agradecimientos

En primer lugar quisiera agradecer a mi directora, Graciela Boente, por todo lo que ha hecho por mí en estos años, no solo desde lo matemático sino desde lo personal, por siempre estar pensando en lo que es mejor para mi avance. ¡Gracias por tu comprensión y tus consejos!

A los miembros del jurado, Matías, Ricardo y Liliana, por su interés y sus aportes.

A mis padres, Marga y Naren, por haberme aguantado incondicionalmente, en las buenas y en las malas, y por siempre estar conmigo. A Valentina, a Rafael, a mis sobrinos, Pako y Dante, a Malvina y a Flor.

A mi otra familia, la "académica", mis colegas del Instituto: Alejandra, Julián (espero ansioso tu vuelta), Juli, Agus, Andrés, Marina, Pau, Dani, Mariela, Vera, Maru y Lucía.

A mis compañeros de la facu, que sin ellos hubiese sido imposible llegar hasta donde estoy ahora: RomiCor, Luli, Verito, Paulilla, Xime, Ema, Yanu, Marian, Lucho, Flor, Quimey, Pablo E., Pablo C., Carla, Ceci, Nico C., Meli, Nico B., Isa, Antonella, Alexia, Hernán, Rodo y Gerardo.

A mis coleguillas del mundo del rol, por tantas y fructíferas partidas: José Luis, Gaby, Fran, Betiel y Guille.

También agradezco a la ANPCyT y al CONICET por haber financiado el desarrollo de mi doctorado, así como a la UBA por haberme ofrecido la gratificante oportunidad de ejercer la docencia.

Finalmente, pero no por eso en menor importancia, a Mercedes, por todas las alegrías sin las cuales hubiese sido absolutamente imposible concluir esta etapa. ¡Gracias por tu infinita paciencia! Sos lo mejor que me ha pasado en la vida.

Índice

1	Introducción	8
1.1	Datos Funcionales	8
1.2	Análisis de Componentes Principales (PCA)	9
1.3	Inferencia Robusta	9
1.4	Propósito de la tesis	12
1.5	Estructura de la tesis	13
2	Preliminares	14
2.1	Nociones básicas en datos funcionales	14
2.1.1	Elemento Aleatorio	15
2.1.2	Esperanza	15
2.1.3	Operador de Covarianza	16
2.1.4	Elementos Gaussianos y Familias Elípticas	16
2.1.5	Componentes Principales Funcionales	18
2.2	Diferenciabilidad en espacios normados	19
2.3	Estimación Robusta	24
2.3.1	Función de Influencia	26
2.3.2	Funcionales de escala	26
3	Estimadores robustos para el análisis de componentes principales funcionales	29
3.1	Introducción	29
3.2	Las propuestas de estimación	30
3.2.1	Los funcionales de direcciones y valores principales	31
3.2.2	Estimadores “crudos” de <i>projection-pursuit</i>	31
3.2.3	Estimadores robustos penalizados de las direcciones principales	31
3.2.4	Estimadores por <i>sieves</i> de las direcciones principales	33
3.2.5	Una formulación unificada para las alternativas robustas mediante <i>projection pursuit</i>	33
3.3	Resultados Auxiliares	34
3.4	Consistencia de los estimadores robustos de las direcciones principales	38
3.4.1	Supuestos	38

3.4.2	Consistencia Fisher	39
3.4.3	Continuidad de los operadores	40
3.4.4	Consistencia de los estimadores “crudos”	44
3.4.5	Consistencia de los estimadores suavizados mediante penalización de la norma	44
3.4.6	Consistencia de los estimadores suavizados mediante penalización en la escala	50
3.4.7	Consistencia de los estimadores robustos basados en <i>sieves</i>	53
3.5	Elección del parámetro de penalización	55
3.6	Estudio Monte Carlo	56
3.6.1	Descripción del Algoritmo y notación	56
3.6.2	Modelo central y contaminaciones	58
3.6.3	Resultados de la simulación	59
3.6.4	Simulación por <i>K-th fold</i>	85
3.7	Aplicación a un ejemplo real: movimiento del labio	86
4	Función de Influencia	89
4.1	Función de Influencia	89
4.2	La función de influencia para estimadores basados en una M -escala	100
5	Estimadores robustos bajo un modelo de Componentes Principales Comunes funcionales	103
5.1	Introducción al modelo	103
5.2	Notación y preliminares	104
5.3	Consistencia Fisher	106
5.4	Estimadores robustos bajo el modelo CPCF	108
5.4.1	Estimador “crudo” mediante <i>projection-pursuit</i>	108
5.4.2	Estimadores penalizados de las direcciones principales comunes	109
5.4.3	Estimadores mediante <i>sieves</i> de las direcciones principales comunes funcionales	110
5.5	Consistencia de los estimadores de <i>projection-pursuit</i> bajo el modelo CPCF	110
5.5.1	Resultados preliminares	110
5.5.2	Supuestos	112
5.5.3	Consistencia	113
5.6	Estudio de Monte Carlo	132
5.6.1	Modelo CPC de rango finito	132
5.6.2	Modelo CPC de rango infinito	146
6	Aspectos Algorítmicos	152
6.1	Introducción	152
6.2	Lemas Previos	153
6.3	Caso de una población	155
6.3.1	Caso del estimador sin penalizar	155
6.3.2	Caso de penalización en la norma	156

6.3.3	Caso de penalización en la escala	157
6.4	Caso de varias poblaciones bajo un modelo CPCF	159
6.4.1	Caso de estimadores sin penalizar	159
6.4.2	Caso de estimadores con penalización en la norma	162
6.4.3	Caso de estimadores con penalización en la escala	165
6.4.4	Caso de estimadores basados en <i>sieves</i>	167
6.5	Estudio de Monte Carlo	168
6.5.1	Comportamiento de los estimadores	170

Capítulo 1

Introducción

1.1 Datos Funcionales

Existe un importante número de situaciones provenientes de las ciencias aplicadas (química, biométrica, medicina, economía, etc) en donde los datos recolectados resultan ser curvas. En efecto, el desarrollo de las herramientas computacionales actuales nos permiten trabajar con conjuntos grandes de datos, en donde “grande” se debe entender por un lado con respecto a la cantidad de datos y por otro lado por el tamaño de cada uno de ellos. Es decir, para un fenómeno particular podríamos llegar a observar un gran número de variables, teniendo así una alta dimensionalidad de la información. Por ejemplo, si tenemos un proceso estocástico indexado en tiempo, tendríamos entonces $\{X(t_j)\}_{j=1,\dots,J}$. La grilla de puntos en donde se observan los datos podría ser cada vez más fina y los puntos consecutivos cada vez más cercanos entre sí. Una manera de considerar este problema es pensar que nuestro dato es una observación discretizada del proceso $\{X(t), t \in (t_{min}, t_{max})\}$. En general, la idea es que este elemento X presenta alguna información de regularidad que puede ser explotado de forma funcional, por ejemplo, que sea continua o diferenciable. Como en Ferraty y Vieu (2006) diremos que un elemento aleatorio $X : \Omega \rightarrow \mathcal{H}$ es una variable funcional si \mathcal{H} , el espacio de posibles valores de X , tiene dimensión infinita. Una observación de X se dirá un dato funcional

Ejemplos de un dato funcional podría ser la temperatura registrada a lo largo de un período por una estación climática posicionada en un determinado lugar. Si consideramos todas las estaciones climáticas, tendríamos entonces un conjunto de datos funcionales. Ramsay y Silverman (2005) proveen un análisis de este tipo de observaciones.

Cabe aclarar que el hecho que el dato sea funcional no implica necesariamente que tenga que ser una curva en el sentido estricto de la palabra. Podríamos estar considerando también imágenes o superficies, aspectos quizás más complejos. Lo importante en cualquier caso es que el dato es un elemento de un espacio infinito-dimensional, y cuenta, preferentemente, con algún tipo de regularidad que se pueda aprovechar.

Por ejemplo, Ramsay y Silverman (2005) consideran el seguimiento anual de las temperaturas de 35 estaciones climáticas de Canadá; dichas temperaturas se muestran en la Figura 1.1.

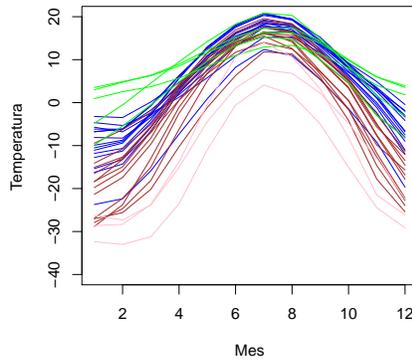


Figura 1.1: Temperatura de 35 estaciones climáticas en Canadá

1.2 Análisis de Componentes Principales (PCA)

El Análisis de Componentes Principales es un procedimiento que utiliza una transformación ortogonal para convertir un vector aleatorio \mathbf{x} , con coordenadas posiblemente correlacionadas, en otro vector cuyas coordenadas, llamadas componentes principales, resultan no correlacionadas. La cantidad de componentes principales es menor o igual al número de variables originales. Esta transformación se define de forma tal que la primer componente principal tiene la mayor varianza posible y las subsiguientes maximizan la varianza sujetas a la condición de ser ortogonales a las precedentemente obtenidas. De esta forma, el subespacio generado por las primeras q componentes da la mejor aproximación en media cuadrática del vector aleatorio \mathbf{x} entre todos los subespacios de dimensión q . Por esta razón, esta herramienta se utiliza fundamentalmente para poder efectuar reducción de la dimensión de observaciones independientes $\mathbf{x}_1, \dots, \mathbf{x}_n$ del vector \mathbf{x} .

Podemos vincular ambos aspectos, el de componentes principales y el de datos funcionales definiendo las componentes principales funcionales, como en Ramsay y Silverman (2005). Por ejemplo, en el caso de las estaciones climáticas de Canadá, en la Figura 1.2 se grafica en negro la media punto a punto de la temperatura, mientras que las líneas roja y azul corresponden a la media más un múltiplo de la primer componente principal y menos dicho múltiplo, respectivamente. Tenemos una idea así del comportamiento general climático en Canadá. La curva de la media nos informa que en los meses de invierno la temperatura es menor, mientras que la primer componente principal da a entender que existe mayor variabilidad de la temperatura en los meses invernales.

1.3 Inferencia Robusta

En términos informales, los procedimientos de inferencia robustos se basan en el hecho que varios supuestos normalmente hechos en estadística (como la normalidad, la linealidad y la independencia de los datos) son, como mucho, una aproximación de la realidad. En muchas situaciones, pueden

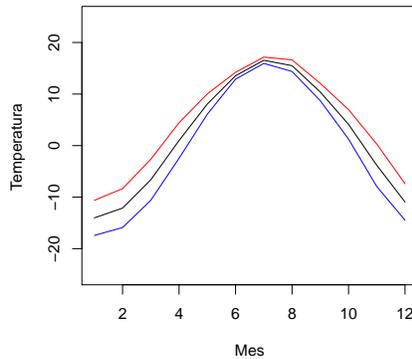


Figura 1.2: Temperatura de 35 estaciones climáticas en Canadá. Media (en negro) y efecto de sumar (en rojo) y restar (en azul) la primer componente principal

ocurrir errores groseros, como los que provienen del simple hecho de que un dato sea incorrectamente transcrito al ser tipeado. Normalmente, estos errores se expresan como “outliers”, datos atípicos, que suelen estar muy alejados del resto de los datos.

Otras razones que llevan a desviaciones de los supuestos idealizados del modelo incluyen el carácter empírico de los mismos y el hecho que se tiene, generalmente, un proceso de aproximación. Así, por ejemplo, al invocar el teorema central del límite, siendo como es un teorema con un resultado asintótico (o sea, límite), en muchas aplicaciones, se piensa los errores experimentales son gaussianos ya que corresponden a una suma de pequeños errores experimentales y ésta puede ser una mala aproximación al problema.

Dada esta situación, el problema con las teorías clásicas es que derivan procedimientos óptimos bajo modelos paramétricos exactos y determinados. Sin embargo, estos procedimientos pueden ser muy ineficaces cuando se presentan desviaciones de dichos modelos, es decir, pueden tener un comportamiento pobre frente a alejamientos del modelo “central” o esperado.

Los procedimientos estadísticos robustos tienen como objetivo permitir inferencias válidas cuando el modelo no se cumple exactamente y al mismo tiempo ser altamente eficientes bajo el modelo central.

Para ilustrar las repercusiones que pueden tener la presencia de “outliers” sobre los datos, consideremos observaciones $\mathbf{x}_1, \dots, \mathbf{x}_{100}$ i.i.d., $\mathbf{x}_i \in \mathbb{R}^2$, tales que $\mathbf{x}_i \sim N(0, \Sigma)$ con

$$\Sigma = \begin{pmatrix} 2 & 1 \\ 1 & 3 \end{pmatrix}.$$

La Figura 1.3 muestra las componentes principales estimadas para estos datos.

Consideremos ahora el modelo de contaminación siguiente. Sean $\mathbf{x}_i^c \in \mathbb{R}^2$ independientes tales que

$$\mathbf{x}_i^c \sim (1 - \epsilon)N(0, \Sigma) + \epsilon N\left(\begin{pmatrix} 10 \\ 0 \end{pmatrix}, \begin{pmatrix} 0.001^2 & 0 \\ 0 & 0 \end{pmatrix}\right)$$

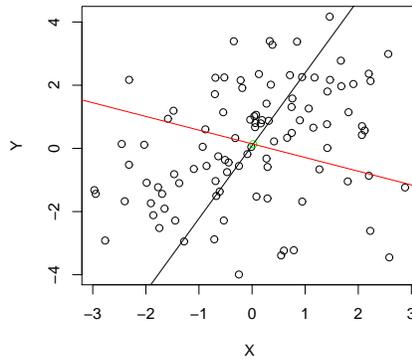


Figura 1.3: Componentes principales para 100 observaciones $\mathbf{x}_i \sim N(0, \Sigma)$ en dimensión 2.

con $\epsilon = 0.1$. Es decir, contaminamos un 10% de las observaciones anteriores con una distribución altamente concentrada en $\begin{pmatrix} 10 \\ 0 \end{pmatrix}$. La Figura 1.4 muestra el resultado de hacer un análisis de componentes principales. La recta negra representa la estimación obtenida mediante los procedimientos clásicos que asumen la normalidad de las observaciones, mientras que la recta azul es una estimación mediante un procedimiento robusto utilizando como escala la mediana de los desvíos absolutos (MAD). Más adelante detallaremos los aspectos técnicos necesarios, el objetivo de esto es apreciar los efectos que pueden tener las contaminaciones sobre los modelos clásicos. Se aprecia claramente la sensibilidad del estimador clásico a la presencia de los datos atípicos que aparecen a la derecha de la figura, cosa que no ocurre con el método robusto.

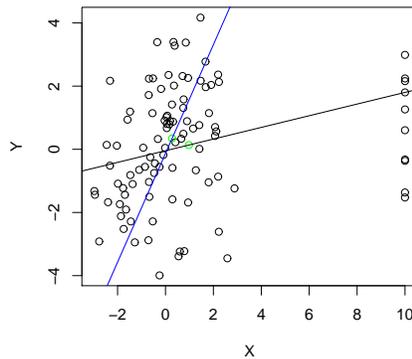


Figura 1.4: Componentes principales para datos en dimensión 2 con 10% de contaminación

1.4 Propósito de la tesis

En el presente trabajo se unificarán los tres aspectos antes mencionados. Consideraremos datos funcionales, es decir, elementos aleatorios a valores en un espacio de Hilbert separable, y en base a ellos estimaremos las direcciones principales de dichos datos. Sin embargo, en aras de obtener resistencia frente a la presencia de *outliers*, propondremos alternativas robustas.

El problema de componentes principales para datos funcionales fue estudiado por Dauxois *et al.* (1982), Pezzulli y Silverman (1993) y Silverman (1996). Sin embargo, en dichos trabajos no se toma en cuenta la eventual presencia de datos atípicos, o *outliers*. Por otra parte, en el caso finito-dimENSIONAL, Li y Chen (1985), Croux y Ruiz-Gazen (1996, 2005), propusieron alternativas robustas para el análisis de componentes principales.

Uno de los aportes de esta tesis es dar un marco de trabajo unificado al problema de estimación robusta de componentes principales en un contexto funcional. Cabe aclarar que, de esta manera, se generalizan ambas aproximaciones: a) los estimadores de Pezzulli y Silverman (1993) y Silverman (1996) pues permitimos cambiar la varianza por una medida de dispersión (eventualmente robusta); b) los estimadores considerados por Li y Chen (1985) y Croux y Ruiz-Gazen (1996, 2005) pues se pasa de un enfoque multivariado (\mathbb{R}^p) a uno sobre un espacio de Hilbert separable, que incluye, como caso particular, aquel sobre el cual trabajaron los autores antes mencionados. En este sentido, proponemos diversas alternativas para poder estimar de forma robusta las componentes principales funcionales que incluyen estimadores sin suavizar y estimadores suaves obtenidos penalizando la escala y/o la norma así como una propuesta basada en aproximaciones por subespacios crecientes de dimensión finita. Para mostrar que dichas propuestas aproximan efectivamente las direcciones principales, se prueba la consistencia de todas esas alternativas. Mediante un estudio de Monte Carlo, se muestra para muestras de tamaño moderado, el buen comportamiento de los estimadores robustos ante la presencia de distintos tipos de contaminación.

Por otro lado, efectuamos el cálculo de la función de influencia de algunas de las propuestas de estimación robusta para las direcciones principales funcionales. Hasta el momento, la función de influencia sólo fue estudiada en el caso finito-dimENSIONAL (ver Croux y Ruiz-Gazen, 2005) y claramente sin considerar penalizaciones. Más aún, demostramos la existencia de dicha función de influencia bajo condiciones generales, algo que ni siquiera está hecho en el caso finito-dimENSIONAL.

Extendemos los estimadores del caso unipoblacional al caso k -poblacional, bajo una extensión al caso funcional del modelo de Componentes Principales Comunes introducido por Flury (1984). En general, cuando se trabaja con varias poblaciones de datos funcionales se suele asumir la igualdad de los operadores de covarianza. En esta tesis, siguiendo la aproximación de Flury (1984) para el caso finito-dimENSIONAL, debilitamos dicha hipótesis de forma tal que sólo se pide igualdad de las autofunciones pero no necesariamente de los autovalores entre las distintas poblaciones. Proponemos distintos estimadores para las autofunciones y autovalores, que como antes incluyen eventualmente penalización en la norma y/o escala y que resultan robustos si la escala elegida es robusta. Obtenemos resultados de consistencia bajo condiciones generales. Nuevamente, los estudios de Monte Carlo avalan el buen funcionamiento de las distintas propuestas robustas ante la presencia de contaminación, para muestras moderadas.

Por último, en esta tesis proponemos un algoritmo que permite calcular en forma práctica los distintos estimadores propuestos, tanto para el caso unipoblacional como para el k -poblacional. La versión finito-dimensional de dicho algoritmo se encuentra desarrollado en Croux y Ruiz-Gazen (1996). Sin embargo, la diferencia con dicho trabajo es que, en esta tesis, se extiende el algoritmo, por un lado, al caso funcional y, por otro lado, se consideran distintos tipos de penalización ya sea en la escala o en la norma. Por otra parte, consideramos también la situación de k poblaciones. Finalmente, una de las contribuciones más importantes y originales de este apartado, es que probamos la consistencia fuerte del algoritmo para todas las propuestas, algo que tampoco estaba hecho en el caso finito-dimensional.

1.5 Estructura de la tesis

Esta tesis se organiza de la siguiente forma.

En el **Capítulo 2**, se dan las herramientas preliminares de trabajo que se usarán en el resto de la tesis. Se introducen los elementos aleatorios junto con todo lo que ello conlleva: definición, esperanza, operador de covarianza y un apartado sobre familias elípticas, pieza clave para algunos de los resultados de esta tesis. Posteriormente, se introduce las componentes principales funcionales. A continuación, se recuerdan distintas nociones de diferenciabilidad en espacios normados que resultarán de utilidad. Finalmente, se repasan algunas nociones útiles en inferencia robusta y se recuerdan algunos estimadores de escala robustos ampliamente utilizados.

En el **Capítulo 3**, introducimos los estimadores robustos de las componentes principales. Presentamos, además, algunos resultados auxiliares vinculados a la convergencia débil y sus consecuencias, que serán útiles para el resto del capítulo. Posteriormente, establecemos la consistencia fuerte, bajo condiciones generales, de las distintas propuestas antes exhibidas. Describimos, asimismo, un procedimiento de selección de parámetro de suavizado. Finalmente, mediante un estudio Monte Carlo se estudia el comportamiento de las distintas propuestas.

En el **Capítulo 4**, se calcula la función de influencia del funcional asociado a algunas propuestas robustas dadas en el Capítulo 3.

En el **Capítulo 5**, consideramos el problema de varias poblaciones. Se introduce la noción de componentes principales comunes robustas y funcionales. Se definen los funcionales que servirán de definición de estas componentes. A continuación, se describen los estimadores robustos de las componentes principales comunes, y se demuestra la consistencia fuerte de dichas propuestas. Concluimos, como en el Capítulo 3, con un estudio de Monte Carlo para analizar el comportamiento de los estimadores ante la presencia de datos atípicos.

En el **Capítulo 6** efectuamos una discusión sobre algunos aspectos algorítmicos. Se demuestra la consistencia de la generalización que proponemos del método de Croux y Ruiz-Gazen, para la primer dirección principal, tanto en el caso de una población como en el de k poblaciones que cumplan el modelo CPC. Se generaliza además el método para adecuarlo a las distintas alternativas de penalización introducidas antes y se prueba su consistencia. Mediante un estudio de Monte Carlo se observa el comportamiento en distintos modelos de dimensión infinita.

Capítulo 2

Preliminares

2.1 Nociones básicas en datos funcionales

Empezaremos la exposición de este capítulo con el repaso de algunos conceptos de Análisis Funcional que usaremos continuamente. Para más detalles, referimos al lector al Conway (1990).

- $(\Omega, \mathcal{A}, \mathbb{P})$ es un espacio de probabilidad sobre Ω con sigma álgebra \mathcal{A} y medida de probabilidad \mathbb{P} .
- \mathcal{X} denotará un espacio de Banach, típicamente separable.
- \mathcal{H} denotará un espacio de Hilbert, siendo ésta nuestra principal herramienta de trabajo. De ahora en más $\langle \cdot, \cdot \rangle$ indicará el producto interno en \mathcal{H} y $\|\alpha\|^2 = \langle \alpha, \alpha \rangle$ la norma asociada. En general consideraremos que \mathcal{H} es separable y, más aún, en algunos casos, $\mathcal{H} = L^2(\mathcal{I})$ siendo \mathcal{I} un intervalo acotado de \mathbb{R} .
- Dados E_1, E_2 espacios de Banach, notamos $B(E_1, E_2)$ al conjunto de operadores lineales y continuos entre ambos espacios.
- Dado $K : E_1 \rightarrow E_2$, notamos $\|K\| = \sup_{x \in E_1, \|x\|=1} \|Kx\|$.
- Si \mathcal{H} es un espacio de Hilbert separable y $(e_j)_{j \in \mathbb{N}}$ es una base ortonormal de \mathcal{H} , notamos como $\|K\|_{\text{hs}}$ a la norma Hilbert–Schmidt de K , definida como

$$\|K\|_{\text{hs}} = \left(\sum_j \|Ke_j\|^2 \right)^{1/2}.$$

Esta norma bien podría dar infinito. En los casos en donde la suma converge y la norma es finita diremos que K es un operador de Hilbert–Schmidt.

- \mathcal{H}^* es el dual topológico y algebraico de \mathcal{H} . Es decir, el espacio de aplicaciones continuas y acotadas $B(\mathcal{H}, \mathbb{R})$.

2.1.1 Elemento Aleatorio

La idea en este capítulo es extender las definiciones clásicas de variable aleatoria al contexto funcional, recurriendo en la medida de lo posible a las definiciones que ya se tienen para el caso real. En todo lo que sigue se supondrá, en un principio, que trabajaremos en un espacio de Banach y en particular, consideraremos después el caso de un espacio de Hilbert separable.

Definimos ahora la extensión de la noción de variable aleatoria para un espacio de Banach.

Definición 2.1.1. Sea $(\Omega, \mathcal{A}, \mathbb{P})$ un espacio de probabilidad y sea \mathcal{X} un espacio de Banach con su σ -álgebra de Borel $\mathcal{B}(\mathcal{X})$ y \mathcal{X}^* el dual topológico de \mathcal{X} . Sea $X : \Omega \rightarrow \mathcal{X}$. Diremos que

- a) X es un elemento aleatorio en sentido débil si para todo $f \in \mathcal{X}^*$, $f(X) : \Omega \rightarrow \mathbb{R}$ es una variable aleatoria real.
- b) X es un elemento aleatorio en sentido fuerte si para todo conjunto $A \in \mathcal{B}(\mathcal{X})$, o sea, para todo A boreliano en \mathcal{X} vale que $X^{-1}(A) \in \mathcal{A}$.

Es claro que un elemento aleatorio en sentido fuerte lo será en sentido débil, los elementos del dual son continuos, por ende medibles, y la medibilidad se preserva bajo la composición. Sin embargo, si \mathcal{X} es separable, todo elemento aleatorio en sentido débil lo será en el sentido fuerte.

Como nuestro contexto de trabajo serán los espacios de Hilbert separables, diremos que X es un elemento aleatorio a secas, sin especificar el sentido, puesto que ambos coinciden.

2.1.2 Esperanza

Definición 2.1.2. Dado \mathcal{X} un espacio de Banach separable y X un elemento aleatorio, diremos que X tiene esperanza $\mathbb{E}(X) \in \mathcal{X}$ si

1. $\mathbb{E}(|f(X)|) < \infty \forall f \in \mathcal{X}^*$
2. $\mathbb{E}(X)$ es el elemento de \mathcal{X} que verifica que $f(\mathbb{E}(X)) = \mathbb{E}(f(X)) \forall f \in \mathcal{X}^*$.

Es fácil ver que si existe el valor esperado de una variable aleatoria entonces es único, basta para eso utilizar el teorema de Hahn-Banach.

La idea de esta definición es la de extender la noción de vector de esperanzas del caso multivariado al caso funcional. Se obtienen entonces extensiones de los resultados que ya se conocen para el caso finito-dimensional, como ser la linealidad de la esperanza. Por otra parte, también vale que la esperanza conmuta con operadores lineales acotados. Para ser más precisos, consideremos dos espacios de Banach separables \mathcal{X} y \mathcal{Y} y un operador lineal acotado $h : \mathcal{X} \rightarrow \mathcal{Y}$. Sea además V un elemento aleatorio en \mathcal{X} con esperanza $\mathbb{E}(X)$. Entonces, vale que $\mathbb{E}(h(X))$ está bien definida y además $\mathbb{E}(h(X)) = h(\mathbb{E}(X))$.

Así como se define la esperanza, es posible también definir la varianza de un elemento aleatorio.

Definición 2.1.3. Sea \mathcal{X} un espacio de Banach separable y X un elemento aleatorio en \mathcal{X} para el que exista $\mathbb{E}(X)$. Se define la varianza de X como $\text{var}(X) = \mathbb{E}(\|X - \mathbb{E}(X)\|^2)$, siempre que $\text{var}(X) < \infty$.

2.1.3 Operador de Covarianza

Nos centraremos en lo que sigue en \mathcal{H} un espacio de Hilbert separable, pues precisaremos la estructura ofrecida por el producto interno. Recordemos que $\langle \cdot, \cdot \rangle$ indica el producto interno en \mathcal{H} y sea $\|\alpha\|^2 = \langle \alpha, \alpha \rangle$. Supondremos además que el elemento aleatorio tiene segundo momento finito, o sea, $\mathbb{E}(\|X\|^2) < \infty$, cosa de tener bien definida la noción de varianza. Vamos a querer extender la idea de matriz de covarianza al caso de elementos aleatorios que toman valores en un espacio de Hilbert separable.

Definición 2.1.4. Sea \mathcal{H} un espacio de Hilbert separable y sea X un elemento aleatorio definido en \mathcal{H} tal que $\mathbb{E}(\|X\|^2) < \infty$. Definimos $a_X : \mathcal{H} \times \mathcal{H} \rightarrow \mathbb{R}$ el operador bilineal de covarianza puntual como $a_X(u, v) = \text{cov}(\langle u, X \rangle, \langle v, X \rangle)$.

Es fácil ver que a_X es bilineal y continuo, luego aplicando el Teorema de Representación de Riesz existen únicos operadores A y B acotados tales que $a_X(u, v) = \langle Au, v \rangle = \langle u, Bv \rangle$, $\forall u, v \in \mathcal{H}$.

Definición 2.1.5. Sea a_X un operador de covarianza puntual y sea A tal que $a_X(u, v) = \langle Au, v \rangle$, $\forall u, v \in \mathcal{H}$. Diremos entonces que A es el operador de covarianza de X y lo notaremos como $\mathbf{\Gamma}_X$.

Sea \otimes el producto tensorial en \mathcal{H} , es decir, para $u, v \in \mathcal{H}$ el operador $u \otimes v : \mathcal{H} \rightarrow \mathcal{H}$ se define como $(u \otimes v)w = \langle v, w \rangle u$. Es decir, si $\|\alpha\| = 1$, $\alpha \otimes \alpha$ es la proyección ortogonal sobre el subespacio generado por α . Con esta notación, podemos escribir al operador de covarianza $\mathbf{\Gamma}_X = \mathbb{E}\{(X - \mu_X) \otimes (X - \mu_X)\}$ siendo μ_X la esperanza de X . El operador $\mathbf{\Gamma}_X$ es autoadjunto (simétrico), semidefinido positivo y continuo. Más aún, es un operador de Hilbert-Schmidt con una cantidad numerable de autovalores, todos ellos reales mayores o iguales a cero. Si $\{\phi_i\}_{i \geq 1}$ es la base ortonormal de autofunciones de $\mathbf{\Gamma}_X$ y sus autovalores asociados son $\{\lambda_i\}_{i \geq 1}$, entonces $\mathbf{\Gamma}_X = \sum_{i \geq 1} \lambda_i \phi_i \otimes \phi_i$.

Un resultado muy importante en análisis multivariado es el que permite obtener la matriz de covarianza de un vector aleatorio compuesto con una transformación lineal, obteniendo que $\mathbf{\Sigma}_{AX} = A\mathbf{\Sigma}_X A^T$. En el caso funcional vale un resultado análogo. Para ser más precisos, sean \mathcal{H}_1 y \mathcal{H}_2 espacios de Hilbert separables y un elemento aleatorio X en \mathcal{H}_1 tal que $\mathbb{E}(\|X\|^2) < \infty$. Sea $A : \mathcal{H}_1 \rightarrow \mathcal{H}_2$ un operador lineal acotado entonces, se cumple que $\mathbf{\Gamma}_{AX} : \mathcal{H}_2 \rightarrow \mathcal{H}_2$ es igual a $\mathbf{\Gamma}_{AX} = A\mathbf{\Gamma}_X A^*$, siendo $A^* : \mathcal{H}_2 \rightarrow \mathcal{H}_1$ el adjunto de A .

Es posible relacionar el operador de covarianza con la varianza, según se definió en la sección anterior. En efecto, si $\{e_n\}_{n \in \mathbb{N}}$ es una base ortonormal de un espacio de Hilbert separable \mathcal{H} y si X es un elemento aleatorio con varianza finita, entonces $\text{var}(X) = \sum_{n=1}^{\infty} \langle \mathbf{\Gamma}_X e_n, e_n \rangle$.

2.1.4 Elementos Gaussianos y Familias Elípticas

Buena parte del estudio clásico en el análisis de datos multivariados, se concentró en el caso de la normal multivariada. Sin embargo, no siempre es posible validar hipótesis de normalidad, tema de especial interés en lo que concierne a la robustez. La idea será entonces, ofrecer una generalización de esta distribución que permita probar la validez de los procedimientos de estimación bajo algún

alejamiento del modelo clásico normal. Las distribuciones elípticas cumplen de forma satisfactoria este papel. Procederemos a recordar la definición dada por Bali y Boente (2009) para el caso funcional. Mencionaremos además algunas propiedades de las mismas.

Definición 2.1.6. Diremos que $X \in \mathcal{H}$ es elípticamente distribuida si existe $\mu \in \mathcal{H}$, $\mathbf{\Gamma} : \mathcal{H} \rightarrow \mathcal{H}$ un operador lineal, simétrico, acotado y semidefinido positivo y $\phi : \mathbb{R}_+ \rightarrow \mathbb{R}$ tal que el funcional característico de $X - \mu$ es

$$\varphi_{X-\mu}(t) = \mathbb{E}(e^{itX}) = \phi(\langle t, \mathbf{\Gamma}t \rangle).$$

En tal caso, diremos que $X \sim \varepsilon(\mu, \mathbf{\Gamma}, \phi)$.

El siguiente resultado sobre las distribuciones elípticas pueden verse en Bali y Boente (2009).

Lema 2.1.1. Si $X \sim \varepsilon(\mu, \mathbf{\Gamma}, \phi)$, luego para todo $A : \mathcal{H} \rightarrow \mathcal{H}_2$ lineal y acotado, $AX \sim \varepsilon(A\mu, A\mathbf{\Gamma}A^*, \phi)$.

Dado un elemento aleatorio $X \in \mathcal{H}$ tal que $X \sim P$, indiquemos por $P[\alpha]$ a la distribución de $\langle \alpha, X \rangle$.

Lema 2.1.2. Sea $X \in \mathcal{H}$ tal que $X \sim \mathcal{E}(\mu, \mathbf{\Gamma}, \phi)$. Luego, existe una distribución univariada simétrica F_0 tal que, para todo $\alpha \notin \ker(\mathbf{\Gamma})$, es decir, tal que $\langle \alpha, \mathbf{\Gamma}\alpha \rangle \neq 0$,

$$P[\alpha]((-\infty, y]) = F_0\left(\frac{y - \langle \mu, \alpha \rangle}{\sqrt{\langle \alpha, \mathbf{\Gamma}\alpha \rangle}}\right),$$

es decir, la variable aleatoria

$$Z_\alpha = \frac{\langle \alpha, X - \mu \rangle}{\sqrt{\langle \alpha, \mathbf{\Gamma}\alpha \rangle}}.$$

tiene distribución F_0 . Por otro lado, si $\langle \alpha, \mathbf{\Gamma}\alpha \rangle = 0$, o sea, $\mathbb{P}(\langle \alpha, X - \mu \rangle = 0) = 1$.

DEMOSTRACIÓN. Primero notemos que puesto que $X \sim \mathcal{E}(\mu, \mathbf{\Gamma}, \phi)$, tenemos que $X - \mu \sim \mathcal{E}(0, \mathbf{\Gamma}, \phi)$, entonces $\langle \alpha, X - \mu \rangle$ tiene una distribución simétrica. Luego, basta ver que la distribución de $\langle \alpha, X - \mu \rangle / \sqrt{\langle \alpha, \mathbf{\Gamma}\alpha \rangle}$ no depende de α , cuando $\langle \alpha, \mathbf{\Gamma}\alpha \rangle \neq 0$, que saldrá del hecho que su función característica no depende de α .

Si definimos $A : \mathcal{H} \rightarrow \mathbb{R}$ como $Ax = \langle \alpha, x \rangle$, tenemos que $X_\alpha = A(X - \mu) \sim \mathcal{E}(0, A\mathbf{\Gamma}A^*, \phi) = \mathcal{E}(0, \langle \alpha, \mathbf{\Gamma}\alpha \rangle, \phi)$. Luego, la función característica de X_α está dada por $\varphi_{X_\alpha}(y) = \phi(\langle \alpha, \mathbf{\Gamma}\alpha \rangle y^2)$.

Observemos que cuando $\langle \alpha, \mathbf{\Gamma}\alpha \rangle = 0$, tenemos que $\varphi_{X_\alpha}(y) = 1$ para todo y , luego $\mathbb{P}(X_\alpha = 0) = 1$ como queríamos probar.

Por otro lado, si $\sigma_\alpha = \langle \alpha, \mathbf{\Gamma}\alpha \rangle \neq 0$, $Z_\alpha = X_\alpha / \sigma_\alpha$. De donde, obtenemos fácilmente que

$$\varphi_{Z_\alpha}(y) = \varphi_{X_\alpha}\left(\frac{y}{\sigma_\alpha}\right) = \phi(y^2)$$

que no depende de α , concluyendo la demostración. \square

La siguiente transformación puede ser utilizada para obtener elementos aleatorios elípticos. Sea V_1 un elemento Gaussiano en \mathcal{H} con media cero y operador de covarianza $\mathbf{\Gamma}_{V_1}$, y sea Z una variable aleatoria real independiente de V_1 . Dado $\mu \in \mathcal{H}$, definamos $X = \mu + ZV_1$. Luego, sale que X tiene una distribución elíptica $\mathcal{E}(\mu, \mathbf{\Gamma})$ con el operador $\mathbf{\Gamma}$ proporcional a $\mathbf{\Gamma}_{V_1}$, sin necesidad de imponer condiciones sobre los momentos. La recíproca vale si el rango de $\mathbf{\Gamma}$ es infinito. Para elementos aleatorios vale que $X(t) = \mu(t) + \sum_{j=1}^q \lambda_j^{\frac{1}{2}} U_j \phi_j(t)$, luego el supuesto de que X sea elíptica es análogo a suponer que $\mathbf{U} = (U_1, \dots, U_q)$ tiene una distribución esférica. Esta expansión finita fue considerado por Gervini (2008) para mostrar propiedades de las componentes esféricas introducidas por Locantore *et al*(1999).

2.1.5 Componentes Principales Funcionales

El análisis de componentes principales es una de las técnicas más populares para analizar datos multivariados. Consiste en explorar la estructura de un conjunto de datos de alta dimensión proyectándolos en las primeras componentes principales que se obtienen calculando los autovectores de la matriz de covarianza muestral. Los autovalores asociados permiten medir la cantidad de información explicada por las componentes principales.

Esta metodología ha sido exitosamente extendida para poder tratar datos funcionales y, usualmente, se la denomina como análisis de componentes principales funcionales. En el caso de un espacio de Hilbert general el procedimiento se puede describir como sigue. Sea $X \in \mathcal{H}$ un elemento aleatorio en un espacio de Hilbert \mathcal{H} definido en $(\Omega, \mathcal{A}, \mathbb{P})$. Como antes, sea $\langle \cdot, \cdot \rangle$ el producto interno en \mathcal{H} y sea $\|\alpha\|^2 = \langle \alpha, \alpha \rangle$. Supongamos que X tenga segundo momento finito $\mathbb{E}(\|X\|^2) < \infty$. Luego, está bien definido el operador de covarianza, $\mathbf{\Gamma}_X$, que resulta lineal, compacto, autoadjunto, semi-definido positivo y continuo.

A pesar de que se considera en este trabajo el caso general en el que $X \in \mathcal{H}$, para introducir las ideas básicas supongamos que $X \in L^2(\mathcal{I})$ donde $\mathcal{I} \subset \mathbb{R}$ es un intervalo finito. Tomamos el producto interno usual de $L^2(\mathcal{I})$, es decir $\langle \alpha, \beta \rangle = \int_{\mathcal{I}} \alpha(t)\beta(t)dt$. En este caso, si indicamos por $\gamma_X(t, s) = \text{cov}(X(t), X(s))$, la función de covarianza, tenemos que el operador de covarianza $\mathbf{\Gamma}_X : L^2(\mathcal{I}) \rightarrow L^2(\mathcal{I})$ está dado por $\mathbf{\Gamma}_X(\alpha)(t) = \int_{\mathcal{I}} \gamma_X(t, s)\alpha(s)ds$. Se supone que la función de covarianza cumple $\int_{\mathcal{I}} \int_{\mathcal{I}} \gamma_X^2(t, s)dt ds < \infty$, de donde se obtiene que $\mathbf{\Gamma}_X$ será un operador de Hilbert–Schmidt.

Como dijimos, los operadores de Hilbert–Schmidt tienen una cantidad numerable de autovalores, siendo todos ellos reales. \mathcal{F} será el espacio de Hilbert de dichos operadores con producto interno definido por $\langle \mathbf{\Gamma}_1, \mathbf{\Gamma}_2 \rangle_{\mathcal{F}} = \sum_{j=1}^{\infty} \langle \mathbf{\Gamma}_1 u_j, \mathbf{\Gamma}_2 u_j \rangle$, donde $\{u_j : j \geq 1\}$ es una base ortonormal de \mathcal{H} . Más aún, puesto que el operador de covarianza $\mathbf{\Gamma}_X$ es también semi-definido positivo, sus autovalores son no negativos. Como pasa con las matrices simétricas en espacios euclídeos finito dimensionales, uno puede elegir las autofunciones de un operador de Hilbert–Schmidt de tal manera de que formen una base ortonormal de \mathcal{H} . Sean $\{\phi_j : j \geq 1\}$ una base ortonormal de autofunciones del operador de covarianza $\mathbf{\Gamma}_X$ y $\{\lambda_j : j \geq 1\}$ sus correspondientes autovalores tales que $\lambda_j \geq \lambda_{j+1}$. La descomposición espectral para $\mathbf{\Gamma}_X$ puede ser expresada como $\mathbf{\Gamma}_X = \sum_{j=1}^{\infty} \lambda_j \phi_j \otimes \phi_j$. En particular, si $\mathcal{H} = L^2(\mathcal{I})$, tenemos que $\gamma_X(t, s) = \sum_{j=1}^{\infty} \lambda_j \phi_j(t)\phi_j(s)$, con $\sum_{j=1}^{\infty} \lambda_j^2 = \|\mathbf{\Gamma}_X\|_{\mathcal{F}}^2 = \int_{\mathcal{I}} \int_{\mathcal{I}} \gamma_X^2(t, s)dt ds$.

Definición 2.1.7. La j -ésima variable principal se define como $Z_j = \langle \phi_j, X \rangle$.

Las componentes principales tienen la propiedad de ser aquellas direcciones que explican la mayor variabilidad de los datos. Para ser más precisos, si $Y = \langle \alpha, X \rangle$, que es una función lineal de X , tenemos $\text{var}(Y) = \langle \alpha, \mathbf{\Gamma}_X \alpha \rangle$. La propiedad de optimalidad de la primer componente principal mencionada es que puede ser definida como la variable $Z_1 = \langle \alpha_1, X \rangle$ tal que

$$\text{var}(Z_1) = \sup_{\{\alpha: \|\alpha\|=1\}} \text{var}(\langle \alpha, X \rangle) = \sup_{\{\alpha: \|\alpha\|=1\}} \langle \alpha, \mathbf{\Gamma}_X \alpha \rangle. \quad (2.1)$$

Cualquier solución de (2.1), es decir, cualquier α para el cual se alcanza el supremo corresponde a una autofunción asociada al autovalor más grande el operador de covarianza $\mathbf{\Gamma}_X$, o sea, $\alpha_1 = \phi_1$ y $\text{var}(Z_1) = \lambda_1$. Si $\lambda_1 > \lambda_2$, entonces α_1 es único salvo cambios de signo.

Como en el caso multivariado, las otras componentes se pueden obtener sucesivamente a través de (2.1), pero bajo la condición de ser ortogonales a las anteriores. Es decir,

$$\begin{cases} \text{var}(\langle \alpha_1, X \rangle) = \sup_{\|\alpha\|=1} \text{var}(\langle \alpha, X \rangle) \\ \text{var}(\langle \alpha_m, X \rangle) = \sup_{\|\alpha\|=1, \alpha \in \mathcal{B}_m} \text{var}(\langle \alpha, X \rangle), \quad 2 \leq m, \end{cases}$$

donde $\mathcal{B}_m = \{\alpha : \langle \alpha, \alpha_k \rangle = 0 \text{ para } k < m\}$. Además, si $\lambda_1 > \dots > \lambda_q > \lambda_{q+1}$ las primeras q componentes principales son únicas y $\alpha_j = \phi_j$, para $j \leq q$.

Esta propiedad será la base para los estimadores robustos que definiremos en el Capítulo 3.

Por otra parte, la descomposición de Karhunen–Loève permite expresar al elemento aleatorio X a través de sus componentes principales, o sea, utilizando la base de autofunciones de $\mathbf{\Gamma}_X$. Efectivamente, se tiene que $X = \mu + \sum_{j=1}^{\infty} Z_j \phi_j$, con $\mu = \mathbb{E}(X)$. Los escores $Z_j = \langle X - \mu, \phi_j \rangle$ son variables aleatorias con media 0, no correlacionados con varianzas λ_j en orden decreciente.

En este sentido, esta descomposición muestra la importancia de tener una estimación precisa de las componentes principales como una forma de predecir las observaciones y examinar su atipicidad.

2.2 Diferenciabilidad en espacios normados

Las nociones de esta sección serán de utilidad en el capítulo 4 para obtener la función de influencia. Existen diversas nociones de diferenciabilidad en espacios normados, la más débil de ellas siendo la diferenciabilidad Gateaux que ha sido usada fuertemente en Estadística debido a su vínculo con la función de influencia y la más fuerte de ellas siendo la diferenciabilidad Fréchet. En muchas situaciones, la diferenciabilidad Fréchet es demasiado fuerte para los funcionales considerados en estadística. En particular, para obtener la normalidad asintótica de los estimadores para un funcional dado, la diferenciabilidad Gateaux no es suficiente mientras que la Fréchet muchas veces no se cumple. Una solución de compromiso es la noción de la derivada de Hadamard, que resulta ser útil para aplicar el método delta. Recordamos las diversas definiciones.

Definición 2.2.1. Sean \mathbb{D}_1 y \mathbb{D}_2 dos espacios normados y consideremos la aplicación $\Upsilon : \mathbb{D}_\Upsilon \subset \mathbb{D}_1 \rightarrow \mathbb{D}_2$.

- a) La función Υ se dice *Fréchet diferenciable* en $\theta \in \mathbb{D}_\Upsilon$ si existe un funcional lineal y acotado $\Upsilon'_\theta : \mathbb{D}_1 \rightarrow \mathbb{D}_2$ tal que

$$\lim_{h \rightarrow 0} \frac{\|\Upsilon(\theta + h) - \Upsilon(\theta) - \Upsilon'_\theta(h)\|}{\|h\|} = 0.$$

- b) La función Υ se dice *Hadamard diferenciable* en $\theta \in \mathbb{D}_\Upsilon$ si existe un funcional lineal y acotado $\Upsilon'_\theta : \mathbb{D}_1 \rightarrow \mathbb{D}_2$ tal que para cualquier compacto $K \subset \mathbb{D}_1$

$$\lim_{t \rightarrow 0} \sup_{h \in K, \theta + th \in \mathbb{D}_\Upsilon} \left\| \frac{\Upsilon(\theta + th) - \Upsilon(\theta)}{t} - \Upsilon'_\theta(h) \right\| = 0.$$

- c) La aplicación Υ se dice *Gateaux diferenciable* en $\theta \in \mathbb{D}_\Upsilon$, en la dirección h , si existe un valor $\Upsilon'_\theta(h)$ tal que, para toda sucesión $t_n \rightarrow 0$ con $n \rightarrow \infty$,

$$\lim_{n \rightarrow \infty} \frac{\Upsilon(\theta + t_n h) - \Upsilon(\theta)}{t_n} = \Upsilon'_\theta(h).$$

Las derivadas de mayor orden se pueden definir de forma similar. Incluso si usamos la misma notación en las tres definiciones, será claro en cada situación qué tipo de derivada es la que se está considerando.

Observación 2.2.1

- a) La derivada de Hadamard es el tipo más débil de derivada que satisface la regla de la cadena. Cuando $\mathbb{D}_1 = \mathbb{R}$ las tres nociones son equivalentes. Por otro lado, si \mathbb{D}_1 es un espacio de Banach, la diferenciabilidad continua de Gateaux es equivalente a la diferenciabilidad Hadamard, donde la continuidad se entiende mientras θ varíe en \mathbb{D}_1 . Para ver vínculos entre estas nociones consultar, por ejemplo, Flett (1980).
- b) Sea $B(\mathbb{D}_1, \mathbb{D}_2)$ el conjunto de aplicaciones lineales y acotadas (y por lo tanto continuas) de \mathbb{D}_1 a \mathbb{D}_2 . Si \mathbb{D}_Υ es un conjunto abierto y Υ es *Hadamard diferenciable* en cada $\theta \in \mathbb{D}_\Upsilon$, podemos considerar la aplicación $\Upsilon' : \mathbb{D}_\Upsilon \rightarrow B(\mathbb{D}_1, \mathbb{D}_2)$ definida como $\Upsilon'(\theta) = \Upsilon'_\theta$ que será llamada la *derivada de Hadamard* de Υ . Cuando Υ' es continua, diremos que Υ es \mathcal{C}^1 -Hadamard.
- c) Cuando \mathbb{D}_1 es un espacio de Hilbert separable, como el que estamos considerando, si $\Upsilon : \mathbb{D}_\Upsilon \rightarrow \mathbb{R}$ es *Fréchet diferenciable* en $\theta \in \mathbb{D}_\Upsilon$, con \mathbb{D}_Υ un abierto de \mathbb{D}_1 , el teorema de Riesz nos permite representar al operador lineal y acotado $\Upsilon'_\theta : \mathbb{D}_1 \rightarrow \mathbb{R}$ a través de un elemento del espacio de Hilbert \mathbb{D}_1 , que será llamado el gradiente de Υ y lo notaremos $\nabla \Upsilon_\theta$, eso es, tendremos que $\Upsilon'_\theta \alpha = \langle \nabla \Upsilon_\theta, \alpha \rangle$.

- d) Indiquemos como $\mathbb{D}_1^* = B(\mathbb{D}_1, \mathbb{R})$ al espacio dual. si $\Upsilon : \mathbb{D}_\Upsilon \rightarrow \mathbb{R}$ es dos veces *Fréchet diferenciable* en $\theta \in \mathbb{D}_\Upsilon$, $\mathbb{D}_\Upsilon \subset \mathbb{D}_1$, usando que $B(\mathbb{D}_1, \mathbb{D}_1^*)$ puede ser identificado con el espacio de operadores bilineales y continuos de $\mathbb{D}_1 \times \mathbb{D}_1$ a \mathbb{R} , notado $BL(\mathbb{D}_1 \times \mathbb{D}_1, \mathbb{R})$, obtenemos que Υ''_θ define un operador bilineal y continuo de $\mathbb{D}_1 \times \mathbb{D}_1$ a \mathbb{R} , a través de la identificación $\Upsilon''_\theta(\alpha, \beta) = (\Upsilon''_\theta(\alpha))\beta$ para $(\alpha, \beta) \in \mathbb{D}_1 \times \mathbb{D}_1$. En particular, si \mathbb{D}_1 es un espacio de Hilbert, Υ''_θ puede ser representado a través de un operador acotado, $H\Upsilon_\theta : \mathbb{D}_1 \rightarrow \mathbb{D}_1$ como $\Upsilon''_\theta(\alpha, \beta) = \langle H\Upsilon_\theta\alpha, \beta \rangle$. Además, si definimos $\Psi(\theta) = \nabla\Upsilon_\theta$, luego $H\Upsilon_\theta = \Psi'_\theta$.
- e) Argumentos similares se pueden usar para aplicaciones diferenciables Hadamard.

Observación 2.2.2

- a) Como un ejemplo simple, que será usado después, consideremos las aplicaciones $\Upsilon : \mathcal{H} \rightarrow \mathbb{R}$, donde \mathcal{H} es un espacio de Hilbert separable, definido como $\Upsilon(\alpha) = \langle \alpha, \Gamma\alpha \rangle$ y $\Psi : \mathcal{H} \rightarrow \mathbb{R}$, definido como $\Psi(\alpha) = \langle \beta_0, \alpha \rangle$ con Γ un operador auto-adjunto, compacto y semi-definido positivo y $\beta_0 \in \mathcal{H}$ fijo. Notemos que Ψ es un operador lineal y acotado. La derivada de Gateaux de Υ y Ψ en un dado α_0 puede ser fácilmente computado y resultan ser operadores lineales y acotados $\Upsilon'_{\alpha_0} : \mathcal{H} \rightarrow \mathbb{R}$ y $\Psi'_{\alpha_0} : \mathcal{H} \rightarrow \mathbb{R}$ dados por $\Upsilon'_{\alpha_0}\alpha = 2\langle \Gamma\alpha_0, \alpha \rangle$ y $\Psi'_{\alpha_0}\alpha = \langle \beta_0, \alpha \rangle$. De hecho, algunas cuentas fáciles, nos permiten ver que Υ y Ψ son Fréchet diferenciable con derivadas Υ'_{α_0} y Ψ'_{α_0} , respectivamente y gradientes $\nabla\Upsilon_{\alpha_0} = 2\Gamma\alpha_0$ y $\nabla\Upsilon_{\alpha_0} = \beta_0$. Más aún, $H\Upsilon_{\alpha_0} = 2\Gamma$.
- b) Usaremos la siguiente propiedad.

Sean las aplicaciones $\Upsilon : \mathcal{U} \subset \mathcal{H} \rightarrow \mathbb{R}$ y $\Psi : \mathcal{U} \subset \mathcal{H} \rightarrow \mathbb{R}$, donde \mathcal{H} es un espacio de Hilbert separable y \mathcal{U} es un conjunto abierto. Definamos $\tilde{\Upsilon}(\theta) = \Upsilon(\theta)\Psi(\theta)$. Luego, si Υ y Ψ son Fréchet (Hadamard o Gateaux) diferenciable en $\theta_0 \in \mathcal{U}$, tenemos que $\tilde{\Upsilon}$ es Fréchet (Hadamard o Gateaux) diferenciable en θ_0 y $\nabla\tilde{\Upsilon}_{\theta_0} = \Psi(\theta_0)\nabla\Upsilon_{\theta_0} + \Upsilon(\theta_0)\nabla\Psi_{\theta_0}$.

En espacios producto, se puede definir también la diferenciabilidad parcial.

Definición 2.2.2. Sea $\mathbb{D}_1, \mathbb{D}_2$ y \mathbb{D} espacios de Banach, $\mathcal{U}_i \subset \mathbb{D}_i$, $i = 1, 2$ conjuntos abiertos, $\Upsilon : \mathcal{U}_1 \times \mathcal{U}_2 \rightarrow \mathbb{D}$ y $\theta_0 = (\theta_{01}, \theta_{02}) \in \mathcal{U}_1 \times \mathcal{U}_2$ fijados. Decimos que Υ es parcialmente Fréchet diferenciable con respecto a la primer coordenada en θ_0 si $\Upsilon_{\theta_{02}} : \mathcal{U}_1 \rightarrow \mathbb{D}$ definido como $\Upsilon_{\theta_{02}}(\theta_1) = \Upsilon(\theta_1, \theta_{02})$, es Fréchet diferenciable en θ_{01} . Notaremos su *derivada parcial de Fréchet* en θ_0 , $D_{1,\theta_0}\Upsilon = \frac{\partial\Upsilon}{\partial\mathbb{D}_1}(\theta_0)$. Entonces, $D_{1,\theta_0}\Upsilon$ es un operador lineal y acotado $D_{1,\theta_0}\Upsilon : \mathbb{D}_1 \rightarrow \mathbb{D}$ que satisface

$$\lim_{h_1 \rightarrow 0} \frac{\|\Upsilon(\theta_0 + (h_1, 0)) - \Upsilon(\theta_0) - (D_{1,\theta_0}\Upsilon)h_1\|}{\|h_1\|} = 0.$$

Análogamente definimos $D_{2,\theta_0}\Upsilon$. En particular, la aplicación $D_1\Upsilon : \mathcal{U}_1 \times \mathcal{U}_2 \rightarrow B(\mathbb{D}_1, \mathbb{D})$ definida como $(D_1\Upsilon)\theta = D_{1,\theta}\Upsilon$ se define cuando Υ es parcialmente Fréchet diferenciable en cualquier $\theta \in \mathcal{U}_1$.

Observación 2.2.3

- a) Es fácil probar que si Υ es Fréchet diferenciable en θ_0 , luego su derivada parcial con respecto a cada coordenada en θ_0 y para todo $h = (h_1, h_2) \in \mathbb{D}_1 \times \mathbb{D}_2$, tenemos

$$\Upsilon'_{\theta_0} h = (D_{1,\theta_0} \Upsilon)h_1 + (D_{2,\theta_0} \Upsilon)h_2 .$$

Más aún, si definimos $J_i : \mathbb{D}_1 \rightarrow \mathbb{D}$ como $J_1(h_1) = (h_1, 0)$ y $J_2(h_2) = (0, h_2)$, luego $D_{i,\theta_0} \Upsilon = \Upsilon'_{\theta_0} \circ J_i$.

- b) Los diferenciales parciales de Hadamard y Gateaux se pueden definir de la misma manera, al igual que las derivadas de orden mayor.

Por ejemplo, supongamos que $D_{1,\theta} \Upsilon$ existe para un θ en un vecindario de θ_0 y definamos $\Psi : \mathcal{U}_1 \times \mathcal{U}_2 \rightarrow B(\mathbb{D}_1, \mathbb{D})$ como $\Psi(\theta) = D_{1,\theta} \Upsilon$. Luego, diremos que Υ es parcialmente Fréchet diferenciable con respecto a la primer coordenada en θ_0 si existe un operador lineal y acotado $D_{11,\theta_0}^2 \Upsilon : \mathbb{D}_1 \rightarrow B(\mathbb{D}_1, \mathbb{D})$ tal que

$$\lim_{h_1 \rightarrow 0} \frac{\|\Psi(\theta_0 + (h_1, 0)) - \Psi(\theta_0) - (D_{11,\theta_0}^2 \Upsilon)h_1\|}{\|h_1\|} = 0 .$$

donde la norma en el numerador es la norma de operador en $B(\mathbb{D}_1, \mathbb{D}_2)$.

Luego, $D_{11,\theta_0}^2 \Upsilon \in B(\mathbb{D}_1, B(\mathbb{D}_1, \mathbb{D}))$. Observando que $B(\mathbb{D}_1, B(\mathbb{D}_1, \mathbb{D}))$ se puede identificar con el espacio de operador bilineales y continuos de $\mathbb{D}_1 \times \mathbb{D}_1$ to \mathbb{D} , denotado $BL(\mathbb{D}_1 \times \mathbb{D}_1, \mathbb{D})$, tenemos que $D_{11,\theta_0}^2 \Upsilon$ define un operador bilineal y continuo de $\mathbb{D}_1 \times \mathbb{D}_1$ a \mathbb{D} , a través de la identificación $D_{11,\theta_0}^2 \Upsilon(h_1, v_1) = ((D_{11,\theta_0}^2 \Upsilon)h_1)v_1$ para $(h_1, v_1) \in \mathbb{D}_1 \times \mathbb{D}_1$. En particular, si \mathbb{D}_i son espacios de Hilbert y $\mathbb{D} = \mathbb{R}$, $D_{11,\theta_0}^2 \Upsilon$ puede ser representado a través de un operador acotado, $\mathbf{G}_{11} : \mathbb{D}_1 \rightarrow \mathbb{D}_1$ como $D_{11,\theta_0}^2 \Upsilon(\alpha, \beta) = \langle \mathbf{G}_{11}\alpha, \beta \rangle$.

Podemos también definir $D_{11}^2 \Upsilon : \mathcal{U}_1 \times \mathcal{U}_2 \rightarrow BL(\mathbb{D}_1 \times \mathbb{D}_1, \mathbb{D})$ como $(D_{11}^2 \Upsilon)\theta_0 = D_{11,\theta_0}^2 \Upsilon$. Argumentos similares se pueden usar para las otras derivadas parciales segundas $D_{12,\theta_0}^2 \Upsilon$, $D_{21,\theta_0}^2 \Upsilon$ y $D_{22,\theta_0}^2 \Upsilon$ donde para prevenir confusión en la notación, $D_{21,\theta_0}^2 \Upsilon$ significará la derivada en la segunda coordenada de $\Psi(\theta) = D_{1,\theta} \Upsilon$, mientras que $D_{12,\theta_0}^2 \Upsilon$ representará la derivada en la primer coordenada de $D_{2,\theta} \Upsilon$.

- c) Como en cálculo real, si Υ es dos veces continuamente Fréchet diferenciable en θ_0 , luego Υ''_{θ_0} es simétrica, implicando que $D_{12,\theta_0}^2 \Upsilon(\theta_1, \theta_2) = D_{21,\theta_0}^2 \Upsilon(\theta_2, \theta_1)$.

Resultados similares valen para aplicaciones Hadamard diferenciables.

- d) En particular, tenemos la siguiente propiedad.

Dados $\Upsilon : \mathbb{D}_1 \times \mathbb{D}_2 \rightarrow \mathbb{F}$, $\eta : \mathbb{D} \rightarrow \mathbb{D}_1$, $v : \mathbb{D} \rightarrow \mathbb{D}_2$ definamos $\Lambda : \mathbb{D} \rightarrow \mathbb{F}$ como $\Lambda(\theta) = \Upsilon(\eta(\theta), v(\theta))$ y $\zeta : \mathbb{D} \rightarrow \mathbb{D}_1 \times \mathbb{D}_2$ como $\zeta(\theta) = (\eta(\theta), v(\theta))$. Si Υ y ζ son Fréchet (Hadamard) diferenciable en $\alpha_0 = (\zeta(\theta_0), v(\theta_0))$ y θ_0 , respectivamente, luego tenemos que Λ es Fréchet (Hadamard) diferenciable en θ_0 y para todo $h \in \mathbb{D}$

$$\Lambda'_{\theta_0} h = \Upsilon'_{\alpha_0} \circ \zeta'_{\alpha_0} h = (D_{1,\alpha_0} \Upsilon)\eta'_{\theta_0} h + (D_{2,\alpha_0} \Upsilon)v'_{\theta_0} h .$$

Por una cuestión de completitud, recordamos el Teorema de la Función Implícita y el método de los multiplicadores de Lagrange que serán usados en lo que sigue. Sus demostraciones pueden verse, por ejemplo, en Flett (1980).

Teorema 2.2.1. *Sea \mathbb{D}_1 y \mathbb{D}_2 espacios de Banach y $\Upsilon : \mathbb{D}_1 \times \mathbb{D}_2 \rightarrow \mathbb{D}$ una aplicación \mathcal{C}^1 -Fréchet diferenciable map. Sea $\theta_0 = (x_0, y_0) \in \mathbb{D}_1 \times \mathbb{D}_2$ tal que $\Upsilon(\theta_0) = 0$ y $D_{2,\theta_0}\Upsilon : \mathbb{D}_2 \rightarrow \mathbb{D}$ es un operador inversible. Luego, existen vecindades abiertas $\mathcal{U}_{x_0} \subset \mathbb{D}_1$ de x_0 y $\mathcal{V}_{y_0} \subset \mathbb{D}_2$ de y_0 y una aplicación \mathcal{C}^1 -Fréchet diferenciable $v : \mathcal{U}_{x_0} \rightarrow \mathcal{V}_{y_0}$ tal que $y_0 = v(x_0)$ y para todo $x \in \mathcal{U}_{x_0}$ e $y \in \mathcal{V}_{y_0}$, tenemos que $\Upsilon(x, y) = 0$ si y solo si $y = v(x)$. Más aún, para todo $x \in \mathcal{U}_{x_0}$, si $\theta = (x, v(x))$ y si $D_{2,\theta}\Upsilon : \mathbb{D}_2 \rightarrow \mathbb{D}$ es un operador inversible, la derivada de v está dada por*

$$v'_x = (D_{2,\theta}\Upsilon)^{-1} D_{1,\theta}\Upsilon ,$$

y v'_x es continua en x_0 .

Sean \mathbb{D} y \mathbb{F} espacios de Banach, $\mathcal{U} \subset \mathbb{D}$ un conjunto abierto, $\Upsilon : \mathcal{U} \times \mathbb{R}$ y $\Gamma : \mathcal{U} \rightarrow \mathbb{F}$. Decimos que Υ tiene un máximo local en $\theta_0 \in \mathcal{U}$ sujeto a la condición $\Psi(\theta) = 0$ si θ_0 pertenece a la superficie de nivel $S = \Psi^{-1}(\{0\}) = \{\theta : \Psi(\theta) = 0\}$ y existe un vecindario $\mathcal{U}_{\theta_0} \subset \mathcal{U}$ de θ_0 tal que $\Upsilon(\theta) \leq \Upsilon(\theta_0)$ para todo $\theta \in S \cap \mathcal{U}_{\theta_0}$. Tenemos una definición análoga para mínimo local condicioando.

Dado $\Psi : \mathcal{U} \rightarrow \mathbb{F}$, decimos que Ψ es sobreyectiva sobre \mathbb{F} si para todo $y \in \mathbb{F}$ existe $\theta \in \mathcal{U}$ tal que $\Psi(\theta) = y$.

Teorema 2.2.2. *Sea \mathbb{D} y \mathbb{F} espacios de Banach, $\mathcal{U} \subset \mathbb{D}$ un conjunto abierto, $\Upsilon : \mathcal{U} \times \mathbb{R}$ y $\Psi : \mathcal{U} \rightarrow \mathbb{F}$. Sea $\theta_0 \in \mathcal{U}$ tal que $\Psi(\theta_0) = 0$. Supongamos que*

- i) Υ tiene un máximo local (o mínimo) en θ_0 condicionado a que $\Psi(\theta) = 0$.
- ii) Υ es Hadamard diferenciable en θ_0
- iii) Ψ es Hadamard diferenciable en un vecindario de θ_0 , Ψ'_θ es sobreyectiva sobre \mathbb{F} y es continua en θ_0 .

Luego, $\ker(\Psi'_\theta) \subset \ker(\Upsilon'_\theta)$ y existe $\Lambda \in \mathbb{F}^*$, donde \mathbb{F}^* significa el espacio dual, tal que $\Upsilon'_\theta = \Lambda \circ \Psi'_\theta$.

Observación 2.2.4. En particular, si $\mathbb{F} = \mathbb{R}^q$, tenemos que existen $\gamma = (\gamma_1, \dots, \gamma_q)^T \in \mathbb{R}^q$, tales que $\Upsilon'_\theta = \gamma^T \Psi'_\theta$.

Más aún, si $\mathbb{F} = \mathbb{R}$ y \mathbb{D}_1 es un espacio de Hilbert separable, existe $\gamma \in \mathbb{R}$, tal que $\Upsilon'_\theta = \gamma \Psi'_\theta$ y $\nabla \Upsilon_\theta = \gamma \nabla \Psi_\theta$.

2.3 Estimación Robusta

En esta sección introduciremos algunos de los aspectos de robustez que utilizamos en esta tesis de doctorado. Muchos de estos resultados pueden verse en Maronna *et al* (2006) y Hampel *et al* (1986).

La cuantificación del grado de robustez de los nuevos procedimientos robustos fue posible debido a la teoría de robustez que tuvo inicio con Huber (1964) y Hampel (1968), posteriormente extendida en otros trabajos y libros, como los de Huber (1981), Hampel *et al* (1986) y Maronna *et al* (2006).

Sea X_1, \dots, X_n una muestra de elementos (variables) independientes e idénticamente distribuidas, tales que, $X_i \in \mathcal{H}$, $X_i \sim P$. En el caso de un modelo paramétrico se suele suponer que $P \in \{P_\theta, \theta \in \Theta\}$ siendo θ desconocido y perteneciente a un espacio paramétrico Θ . Uno de los objetivos es poder estimar (eficientemente) al verdadero valor de θ .

Los procedimientos de inferencia robusta suponen que la medida P pertenece a una vecindad de radio ϵ de P_0 . Una forma de definir una vecindad es considerar los entornos de contaminación introducidos por Tukey (1960)

$$GE(P, \epsilon) = \{P : P = (1 - \epsilon)P_0 + \epsilon Q, Q \in \mathcal{F}\}$$

con \mathcal{F} una familia de probabilidades sobre \mathcal{H} . Por lo tanto, para observaciones generadas según $P = (1 - \epsilon)P_0 + \epsilon Q$, lo que termina ocurriendo es que aproximadamente una proporción ϵ provendrán de una distribución Q .

Para incluir tanto errores de redondeo como errores groseros, es necesario definir entornos mediante topologías débiles en el espacio de las medidas \mathcal{F} como, por ejemplo, la métrica de Prohorov definida como

$$d_{\text{PR}}(P, Q) = \inf\{\epsilon > 0, Q(A) \leq P(A^\epsilon) + \epsilon, \forall A \in \mathcal{B}\}$$

donde \mathcal{B} es la σ -álgebra generada por los conjuntos abiertos de \mathcal{H} y A^ϵ es el entorno de radio ϵ de A , es decir, el conjunto de elementos que se encuentran a distancia menor o igual de ϵ de A . Con esta métrica podemos definir otro tipo de vecindades dadas por $B_\epsilon(P) = \{Q, d_{\text{PR}}(P, Q) < \epsilon\}$. La métrica de Prohorov genera la topología débil, es decir, aquella asociada a la convergencia en distribución o débil de probabilidades y juega un papel fundamental para definir la noción de robustez cualitativa.

Funcionales y Estimadores

Un concepto importante en estimación robusta es la idea de “funcional”. Sea \mathcal{P} el espacio de medidas de probabilidad sobre \mathcal{H} y $\mathcal{P}_0 \subset \mathcal{P}$.

Un funcional es una aplicación $T : \mathcal{P}_0 \rightarrow \mathbb{S}$, siendo \mathbb{S} en principio cualquier conjunto posible, pero, en general, debería estar dotado de alguna estructura aunque sea métrica. Por ejemplo, en el caso paramétrico $\mathbb{S} = \Theta$ mientras que, en el caso de las componentes principales, tendremos $\mathbb{S} = \{\|x\| = 1\}$. Esta aplicación podría estar parcialmente definida, es decir $\mathcal{P}_0 \neq \mathcal{P}$, o sea, quizás T no tiene valor asignado para toda medida de probabilidad. Un ejemplo de esta situación es

la media, $\mu : \mathcal{P}_0 \rightarrow \mathbb{R}$, $\mu(P) = \int_{\Omega} x dP(x)$. Este funcional solamente está definido para aquellas medidas que tengan primer momento finito.

Otro ejemplo, es el funcional de la mediana, definido en el caso $\mathcal{H} = \mathbb{R}$ como $\tilde{\mu}(P) = \inf\{x : P((-\infty, x]) \geq 1/2\}$. En este caso, $\mathcal{P}_0 = \mathcal{P}$.

Por último, también en $\mathcal{H} = \mathbb{R}$, y para el modelo de posición–escala con escala σ conocida, un funcional muy utilizado en robustez es el M –funcional de posición, que es un funcional $T(P)$ definido como la solución de la ecuación $\int \psi((x - T(P))/\sigma) dP(x) = 0$, donde ψ debe cumplir ciertas restricciones (ver, Huber, 1981, para más detalles). Si σ es desconocido, es posible definirlo utilizando un funcional de escala, que definiremos en la Sección 2.3.2.

Si \mathbb{S} está dotada de una topología, cosa que pasa si, por ejemplo, es un espacio métrico, podemos estudiar en qué casos es el funcional continuo. Esto estará vinculado con una medida de robustez que se denomina “robustez cualitativa”. El funcional $T : \mathcal{P}_0 \rightarrow \mathbb{S}$ se dice cualitativamente robusto en $P = P_0 \in \mathcal{P}_0$ si es continuo en P_0 , con respecto de la métrica de Prohorov.

Si X_1, \dots, X_n es una muestra aleatoria, con $X_i \sim P$, la medida empírica se define como

$$P_n = \frac{1}{n} \sum_{i=1}^n \delta_{X_i}$$

donde δ_{x_i} es la medida puntual en x_i . Si estamos interesados en estimar $T(P)$, podemos utilizar $T_n = T(P_n) = T(X_1, \dots, X_n)$ como estimador (“estimador plug-in”) de nuestra cantidad de interés.

Es fácil ver que $\mu(P_n) = \bar{X}$ y que $\tilde{\mu}(P_n) = \text{mediana}_{1 \leq i \leq n} X_i$ que son los estimadores de la media y la mediana, respectivamente.

Existen distintas maneras de medir la robustez de un estimador. Entre ellas, la robustez cualitativa está íntimamente relacionada con la continuidad del funcional respecto de la topología débil.

Definición 2.3.1. (Hampel, 1971) Diremos que una sucesión de estimadores $\{T_n = T(P_n), n \in \mathbb{N}\}$ es **asintóticamente cualitativamente robusto** en T si, para todo $\epsilon > 0$, existe $\delta > 0$, $n_0 \in \mathbb{N}$ tal que

$$d_{\text{PR}}(P, Q) < \delta \Rightarrow d_{\text{PR}}(T(P_n), T(Q_n)) < \epsilon \quad \forall n \geq n_0.$$

Este siguiente teorema cuya demostración puede verse en Hampel(1971) relaciona este concepto con el de continuidad.

Teorema 2.3.1. Si $T : \mathcal{P} \rightarrow \mathbb{S}$ es continua con la métrica de Prohorov entonces los estimadores $T_n = T(P_n)$ son *asintóticamente cualitativamente robustos*.

Es decir, para asegurar la robustez cualitativa basta con verificar la continuidad del funcional T .

Un punto importante en la definición de un estimador es que, efectivamente, estime la cantidad de interés. Esto se relaciona con una propiedad del funcional llamada consistencia en el sentido de Fisher. En el caso paramétrico, o sea, cuando $P \in \{P_{\theta}, \theta \in \Theta\} \subset \mathcal{P}_0$, diremos que un funcional $T : \mathcal{P}_0 \rightarrow \mathbb{S}$ (con $\Theta \subset \mathbb{S}$) es **Fisher–consistente** si, para todo $\theta \in \Theta$, $T(P_{\theta}) = \theta$, es decir, si estimamos la cantidad de interés.

2.3.1 Función de Influencia

La función de influencia es una de las herramientas heurísticas más utilizadas para medir la robustez de un estimador. Da una idea de la sensibilidad infinitesimal del estimador considerado y se define como sigue

Definición 2.3.2. (Hampel, 1968, 1974) La **función de influencia** de un funcional T en P está dada por

$$\text{IF}(x, T, P) = \lim_{\epsilon \rightarrow 0} \frac{T((1 - \epsilon)P + \epsilon\delta_x) - T(P)}{\epsilon}$$

para los puntos $x \in \mathcal{H}$ para los cuales el límite existe.

Es decir, $\text{IF}(x, T, P)$ es la derivada de Gateaux del funcional T en el punto P sobre la dirección $\delta_x - P$. Da una idea aproximada de la influencia que tiene el reemplazo de una observación de la muestra por el valor x en el estimador.

En el caso finito-dimensional, una de las utilidades de la función de influencia es el cálculo heurístico de la varianza asintótica del estimador asociado a un funcional. Supongamos que $\mathcal{H} = \mathbb{R}^d$ y $\mathbb{S} = \mathbb{R}^k$. Sea X_1, \dots, X_n una muestra aleatoria y P_n su medida empírica. Entonces, bajo condiciones de regularidad (ver Fernholz, 1983) se tiene el desarrollo siguiente

$$T(P_n) = T(P) + \int \text{IF}(x, T, P) dP_n(x) + R(P_n - P) = T(P) + \frac{1}{n} \sum_{i=1}^n \text{IF}(x_i, T, P) + R(P_n - P).$$

Cuando $\sqrt{n}R(P_n - P) \xrightarrow{p} 0$, utilizando el teorema central del límite junto con el lema de Slutsky obtenemos que $n^{1/2}(T(P_n) - T(P))$ tiene distribución asintótica normal con media cero y varianza asintótica (ASVAR) dada por

$$\text{ASVAR}(T, P) = \frac{1}{n} \text{var}\left(\sum_{i=1}^n \text{IF}(X_i, T, P)\right) = \mathbb{E}_P(\text{IF}(X, T, P)\text{IF}(X, T, P)^T).$$

Algunos ejemplos de funciones de influencia pueden verse en Huber (1981).

2.3.2 Funcionales de escala

La idea fundamental de nuestra propuesta consiste en considerar las componentes principales como están definidas en la Sección 2.1.5 pero cambiando la varianza por un funcional robusto de escala. Por esta razón recordaremos su definición..

Consideremos \mathcal{G} el conjunto de medidas de probabilidad univariadas. Un funcional de escala $\sigma_R : \mathcal{G} \rightarrow [0, +\infty)$ es una aplicación que es invariante por cambios de posición y equivariante por cambios de escala. Es decir, si $G_{a,b}$ corresponde a la distribución de $aY + b$ con $Y \sim G$, entonces $\sigma_R(G_{a,b}) = |a|\sigma_R(G)$, para cualquier par de números reales a y b . Indistintamente, evaluaremos al funcional de escala σ_R en una variable aleatoria Y , en cuyo caso se lo debe interpretar como siendo evaluada en la distribución de Y . Así pues, pedimos que $\sigma_R(aY + b) = |a|\sigma_R(Y)$.

Un ejemplo conocido de funcional de escala es el desvío estándar, definido por $SD(Y) = \{\mathbb{E}(Y - \mathbb{E}(Y))^2\}^{1/2}$. Otro ejemplo muy usado es la mediana de los desvíos absolutos (median absolute deviation, o MAD), definido como $MAD(Y) = c \text{ mediana}(|Y - \text{mediana}(Y)|)$. La constante de normalización c se usa para conseguir que los estimadores obtenidos a través de este funcional sean consistentes al parámetro de interés. En el caso del desvío estándar usual en la distribución Gaussiana, tomamos $c = 1/\Phi^{-1}(0.75)$, así la MAD se interpreta igual que el desvío estándar en el caso de la distribución normal.

El punto de ruptura, una medida de robustez global, es 0 para el desvío estándar, mientras que para la MAD es igual a 0.5. Sin embargo, la MAD presenta una función de influencia discontinua, reflejando algunos problemas de estabilidad local. Más aún, la versión muestral de la MAD (es decir, $\hat{\sigma}_n = c \text{ mediana}_{1 \leq i \leq n}(|Y_i - \text{mediana}(Y_i)|)$), es conocida por ser relativamente ineficiente sobre la distribución normal (ver, Huber, 1981). En el caso de dimensión finita, como es indicado por Cui *et al.*(2003), el impacto de una función de influencia discontinua en la eficiencia de un estimador de las direcciones principales es particularmente severo.

Una alternativa amplia de funcionales robustos de escala son los M -funcionales de escala. Esta clase incluye al SD y a la MAD. M -funcionales de escala con una función de score continua y acotada podrían tener alto punto de ruptura y una función continua y acotada de escala. Además, su versión muestral, es decir, el M -estimador de escala, puede ser calibrado de forma tal de tener buena eficiencia sobre una clase amplia de distribuciones. Dado un parámetro de posición μ , un M -funcional de escala $\sigma_R(G)$ con una función de score continua $\chi : \mathbb{R} \rightarrow \mathbb{R}$ se define como la solución de la ecuación

$$\mathbb{E} \left[\chi \left(\frac{Y - \mu}{\sigma_R(G)} \right) \right] = \delta. \quad (2.2)$$

Dado un estadístico de posición $\hat{\mu}_n$, el correspondiente M -estimador de escala será la solución s_n de la ecuación

$$\frac{1}{n} \sum_{i=1}^n \chi \left(\frac{Y_i - \mu}{s_n} \right) = \delta. \quad (2.3)$$

Si la función de score es discontinua, como es el caso con la MAD, es preciso un cambio sobre (2.2) y (2.3) para tener bien definido el funcional y estimador, ver, por ejemplo, Martin y Zamar (1993). Típicamente, la función de score χ es par con $\chi(0) = 0$, no decreciente en \mathbb{R}_+ y tal que $0 < \sup_{x \in \mathbb{R}} \chi(x) = \chi(+\infty) = \lim_{x \rightarrow +\infty} \chi(x)$. Cuando $\chi(+\infty) = 2\delta$, el M -estimador de escala tiene un punto de ruptura del 50%, y eligiendo adecuadamente la función χ es posible obtener además un estimador altamente eficiente (ver Croux, 1994).

Para funciones de score estrictamente crecientes, las soluciones de (2.2) y (2.3) son únicas y el siguiente algoritmo iterado

$$\{\hat{\sigma}_n^{(k+1)}\} = \frac{1}{n\delta} \sum_{i=1}^n w \left(\frac{Y_i - \hat{\mu}}{\hat{\sigma}_n^{(k)}} \right) (Y_i - \hat{\mu})^2 \quad (2.4)$$

con $w(y) = \chi(y)/y^2$ para $y \neq 0$ y $w(0) = \chi''(0)$, siempre convergerá la solución única de (2.3) sin importar el valor inicial de $\hat{\sigma}_n^{(0)}$. M -estimadores re-descendentes son también recomendados en

los trabajos de robustez puesto que tienden a eliminar automáticamente los *outliers*. Una elección típica, que usaremos en los estudios Monte Carlo, es la función de score introducida por Beaton y Tukey (1974),

$$\chi_c(y) = \min(3(y/c)^2 - 3(y/c)^4 + (y/c)^6, 1) \quad (2.5)$$

donde c es un constante de calibración que garantiza que los estimadores vinculados sean consistentes al parámetro de interés. Por ejemplo, la elección $c = 1.56$ con $\delta = 1$ asegura que los M -funcionales de escala son Fisher-consistentes en la distribución normal y con punto de ruptura $1/2$.

Al usar un M -estimador redescendiente, el resultado del algoritmo (2.4) podría llegar a depender del valor inicial. En la práctica, el valor inicial $\hat{\sigma}_n^{(0)}$ se suele elegir como la MAD. Un estudio sobre las condiciones que aseguran la convergencia del algoritmo se puede encontrar en Maronna *et al.* (2006).

Para una función acotada de score χ , si la solución $\sigma_R(G_0)$ de (2.2) es única, como ocurre, por ejemplo, si χ es estrictamente creciente, el funcional σ_R es débilmente continuo en G_0 , es decir, es continuo en G_0 con la topología débil en \mathcal{G} dada por la métrica de Prohorov. La continuidad débil en G_0 y la consistencia en un entorno de G_0 aseguran la robustez en G_0 . Más detalles, pueden verse en Huber(1981) y Hampel (1971).

El siguiente resultado vinculan los funcionales de escala con las distribuciones elípticas y será de utilidad en los próximos capítulos.

Lema 2.3.1. *Sea $X \sim P$ tal que $X \sim \mathcal{E}(\mu, \mathbf{\Gamma}, \phi)$, es decir, X está elípticamente distribuida con posición μ y operador de dispersión $\mathbf{\Gamma}$. Sea $P[\alpha]$ la medida de probabilidad de $\langle \alpha, X \rangle$. Entonces, $\sigma_R(P[\alpha]) = c\langle \alpha, \mathbf{\Gamma}\alpha \rangle$ con c una constante que depende exclusivamente del funcional de escala σ_R , más aún, $c = \sigma_R(F_0)$ donde F_0 es la función de distribución del Lema 2.1.2.*

Para su demostración basta usar el Lema 2.1.2.

Capítulo 3

Estimadores robustos para el análisis de componentes principales funcionales

3.1 Introducción

Análogamente al caso de componentes principales clásicas, el método de PCA mediante *projection-pursuit* está basado en encontrar las proyecciones de los datos que tienen la mayor dispersión posible. En vez de usar la varianza como medida de dispersión, se utiliza un estimador robusto de escala s_n para la maximización. Este método fue introducido por Li y Chen (1985), quienes propusieron un método basado en maximizar (o minimizar) una escala robusta. En este sentido, dada una muestra $\mathbf{x}_i \in \mathbb{R}^d$, $1 \leq i \leq n$, la primer componente principal robusta se define como

$$\hat{\mathbf{a}} = \underset{\{\mathbf{a} \in \mathbb{R}^d : \mathbf{a}^T \mathbf{a} = 1\}}{\operatorname{argmax}} s_n(\mathbf{a}^T \mathbf{x}_1, \dots, \mathbf{a}^T \mathbf{x}_n).$$

Los siguientes vectores de componentes principales se obtienen al imponer condiciones de ortogonalidad sobre las componentes previamente calculadas y maximizando nuevamente.

Croux y Ruiz-Gazen (2005) obtienen la función de influencia para las componentes principales, mientras que su distribución asintótica fue estudiada por Cui *et al.* (2003). Un algoritmo de maximización para obtener $\hat{\mathbf{a}}$ fue propuesto en Croux y Ruiz-Gazen (1996) y fue adaptado para datos de altas dimensiones por Croux *et al.* (2007).

El objetivo de este capítulo es adaptar el procedimiento de *projection-pursuit* al caso funcional. Nos enfocamos en datos funcionales $X_i \in \mathcal{H}$, que podrían corresponder a datos que se registran en un período de tiempo y que tomados como realizaciones de un proceso estocástico, se suponen normalmente en el espacio $L^2(\mathcal{I})$, con \mathcal{I} un intervalo real. Consideraremos y compararemos distintas elecciones de escalas robustas, incluyendo la mediana de los desvíos absolutos (MAD) y M -estimadores de escala.

Los procedimientos clásicos de análisis de componentes principales funcional (PCAF) utilizan los autovalores y autofunciones del operador de covarianza muestral. Dauxois *et al.* (1982) estudiaron las propiedades asintóticas de estas componentes principales funcionales muestrales. Boente y Fraiman (2000) consideraron un método basado en núcleos. Rice y Silverman (1991) muestran una forma de suavizar las componentes principales utilizando un método de penalización por falta de suavizado. La consistencia de este método fue estudiado por Pezzulli y Silverman (1993). Por otro lado, también basándose en métodos de penalización por falta de suavidad, Silverman (1996) propone un método en donde se penaliza la norma en vez de la varianza muestral. Trabajos más recientes de estimación de las componentes principales y la función de covarianza incluyen los trabajos de Gervini (2006), Hall y Hosseini–Nasab (2006), Hall *et al.* (2006) y Yao y Lee (2006).

La literatura en componentes principales robustas en el contexto funcional es, sin embargo, bastante escasa. Hasta donde se tiene conocimiento, el primer intento de obtener estimadores de las componentes principales que fuesen menos sensibles a observaciones anómalas se encuentra en Locantore *et al.* (1999), pero su método es, esencialmente, multivariado. Gervini (2008) ofrece un método totalmente funcional para la estimación de componentes principales al considerar una versión funcional de las componentes principales esféricas definidas en Locantore *et al.* (1999). Recientemente, Sawant *et al.* (2011) consideran una aproximación robusta basado en un análisis de componentes principales robusto utilizando matrices de covarianza robustas de los coeficientes de los datos observados en una base dada. Por otra parte, Hyndman y Ullah (2007) dan una aplicación de un método robusto de *projection-pursuit*, aplicado a trayectorias suaves, pero no estudiaron las propiedades de su propuesta.

En este capítulo, se introducen distintos estimadores robustos de las componentes principales en el contexto funcional. Los métodos considerados utilizan un procedimiento robusto de *projection-pursuit* combinado con varias formas de suavizado. Un punto importante del capítulo es el de proveer una base teórica de este método para las componentes principales robustas funcionales. En particular, establecemos bajo condiciones muy generales la consistencia fuerte de los estimadores propuestos.

Primero, se introducen los estimadores robustos de las componentes principales, basados en alternativas con y sin suavizado. Probamos, a continuación, algunos lemas auxiliares necesarios para las demostraciones posteriores. Posteriormente, se dan las hipótesis necesarias para obtener los resultados de consistencia y se verifica, en particular, como dichos supuestos implican la consistencia Fisher de la propuesta no suavizada. La consistencia y la robustez asintótica de los distintos estimadores introducidos se establecen a continuación. Terminamos el capítulo con un estudio de Monte Carlo y el análisis de un conjunto de datos reales.

3.2 Las propuestas de estimación

Vamos a considerar diferentes alternativas robustas de estimación, las mismas se efectuaran sobre un espacio de Hilbert separable \mathcal{H} , teniendo en cuenta que el principal caso a tomar en cuenta es cuando $\mathcal{H} = L^2(\mathcal{I})$. A partir de ahora, $\{X_i, 1 \leq i \leq n\}$ representará las realizaciones de un proceso estocástico $X \sim P$ en un espacio de Hilbert separable \mathcal{H} . Así pues, $X_i \sim P$ son procesos

estocásticos independientes que siguen la misma ley de probabilidad. Dado un funcional de escala σ_{rob} , definamos la aplicación $\sigma : \mathcal{H} \rightarrow \mathbb{R}$ como $\sigma(\alpha) = \sigma_R(P[\alpha])$, donde $P[\alpha]$ es la distribución de $\langle X, \alpha \rangle$ cuando $X \sim P$.

3.2.1 Los funcionales de direcciones y valores principales

Basándonos en la propiedad de la primer componente principal y dado σ_R un funcional robusto de escala, los funcionales de las direcciones principales se definen como

$$\begin{cases} \phi_{R,1}(P) &= \operatorname{argmax}_{\|\alpha\|=1} \sigma_R(P[\alpha]) = \operatorname{argmax}_{\|\alpha\|=1} \sigma(\alpha) \\ \phi_{R,m}(P) &= \operatorname{argmax}_{\|\alpha\|=1, \alpha \in \mathcal{B}_m} \sigma_R(P[\alpha]) = \operatorname{argmax}_{\|\alpha\|=1, \alpha \in \mathcal{B}_m} \sigma(\alpha), \quad 2 \leq m, \end{cases} \quad (3.1)$$

donde $P[\alpha]$ es la distribución de $\langle \alpha, X \rangle$ con $X \sim P$, y $\mathcal{B}_m = \{\alpha \in \mathcal{H} : \langle \alpha, \phi_{R,j}(P) \rangle = 0, 1 \leq j \leq m-1\}$. Denotaremos al m -ésimo valor principal como

$$\lambda_{R,m}(P) = \sigma_R^2(P[\phi_{R,m}]) = \max_{\|\alpha\|=1, \alpha \in \mathcal{B}_m} \sigma_R^2(P[\alpha]) . \quad (3.2)$$

Dado que la bola unitaria es débilmente compacta, el máximo antes planteado se alcanzará si el funcional σ_R es (débilmente) continuo. En la Sección 3.4.2 se discutirá la consistencia Fisher de los funcionales definidos por (3.1) y (3.2).

3.2.2 Estimadores “crudos” de *projection-pursuit*

Denotemos por $s_n^2 : \mathcal{H} \rightarrow \mathbb{R}$ a la función $s_n^2(\alpha) = \sigma_R^2(P_n[\alpha])$, donde P_n es la medida empírica de X_1, \dots, X_n , y $P_n[\alpha]$ es la medida empírica de $\langle X_1, \alpha \rangle, \dots, \langle X_n, \alpha \rangle$. Luego, las direcciones definidas en (3.1) se estimarán empíricamente por estimadores “crudos” o sin suavizar dados por

$$\begin{cases} \hat{\phi}_{RAW,1} &= \operatorname{argmax}_{\|\alpha\|=1} s_n(\alpha) \\ \hat{\phi}_{RAW,m} &= \operatorname{argmax}_{\alpha \in \hat{\mathcal{B}}_m} s_n(\alpha) \quad 2 \leq m, \end{cases} \quad (3.3)$$

con $\hat{\mathcal{B}}_m = \{\alpha \in \mathcal{H} : \|\alpha\| = 1, \langle \alpha, \hat{\phi}_{RAW,j} \rangle = 0, \forall 1 \leq j \leq m-1\}$. Los estimadores de los valores principales se calculan como

$$\hat{\lambda}_{RAW,m} = s_n^2(\hat{\phi}_{RAW,m}), \quad 1 \leq m. \quad (3.4)$$

3.2.3 Estimadores robustos penalizados de las direcciones principales

A veces, en vez de componentes principales funcionales “crudas”, alternativas más suaves son preferibles. Las ventajas de componentes suaves se pueden consultar, por ejemplo, en Rice y Silverman (1991) y Ramsay y Silverman (2005). Un argumento convincente para su aplicación

es que el suavizado es una herramienta de regularización que podría revelar características más interpretables e interesantes de los modos de variación de los datos funcionales. Rice y Silverman (1991) y Silverman (1996) propusieron dos alternativas de suavizado que consisten en penalizar la varianza y la norma, respectivamente. Para ser más específicos, Rice y Silverman (1991) estiman la primer componente principal mediante una maximización sobre $\|\alpha\| = 1$ de la función objetivo $\widehat{\text{var}}(\langle \alpha, X \rangle) - \rho[\alpha, \alpha]$, donde $\widehat{\text{var}}$ es la varianza muestral y $[\alpha, \beta] = \int_{\mathcal{I}} \alpha''(t)\beta''(t)dt$. Resultados de consistencia para este tipo de estimadores fueron obtenidos por Pezzulli y Silverman (1993).

Por otro lado, Silverman (1996) propuso penalizar la rugosidad a través de una norma definida mediante el producto interno penalizado $\langle \alpha, \beta \rangle_r = \langle \alpha, \beta \rangle + r[\alpha, \beta]$. El estimador de la primer dirección principal, $\widehat{\phi}_1$, es entonces el valor α que maximiza $\widehat{\text{var}}(\langle \alpha, X \rangle)$ sobre el conjunto $\|\alpha\|_r = 1$. Silverman(1996) obtuvo resultados de consistencia para estos estimadores bajo el supuesto de que ϕ_j tenga “rugosidad” finita, es decir, $[\phi_j, \phi_j] < \infty$. Claramente, el parámetro de suavizado r habrá de tender a 0 si se quiere obtener resultados de consistencia.

Vamos a describir nuestras propuestas. Tomemos \mathcal{H}_S el subconjunto de “elementos suaves” de \mathcal{H} . Para algunas propuestas habremos de pedir que $\phi_j(P) \in \mathcal{H}_S$. Sea $D : \mathcal{H}_S \rightarrow \mathcal{H}$ un operador lineal, que llamaremos “diferenciador”. Usando D , podemos entonces definir la forma bilineal simétrica semidefinida positiva $[\cdot, \cdot] : \mathcal{H}_S \times \mathcal{H}_S \rightarrow \mathbb{R}$, dada por $[\alpha, \beta] = \langle D\alpha, D\beta \rangle$. El “operador de penalización” se define entonces como $\Psi : \mathcal{H}_S \rightarrow \mathbb{R}$ como $\Psi(\alpha) = [\alpha, \alpha]$, y el producto interno penalizado se define como $\langle \alpha, \beta \rangle_r = \langle \alpha, \beta \rangle + r[\alpha, \beta]$. Entonces, $\|\alpha\|_r^2 = \|\alpha\|^2 + r\Psi(\alpha)$.

Observación 3.1. El caso más común para datos funcionales resulta cuando $\mathcal{H} = L^2(\mathcal{I})$, $\mathcal{H}_S = \{\alpha \in L^2(\mathcal{I}), \alpha \text{ es dos veces diferenciable, y } \int_{\mathcal{I}} (\alpha''(t))^2 dt < \infty\}$, $D\alpha = \alpha''$ y $[\alpha, \beta] = \int_{\mathcal{I}} \alpha''(t)\beta''(t)dt$ de forma tal que $\Psi(\alpha) = \int_{\mathcal{I}} (\alpha''(t))^2 dt$.

Sea $\sigma_R(F)$ un funcional robusto de escala y consideremos $s_n(\alpha)$ como en la sección anterior. Luego, podemos adaptar el procedimiento clásico definiendo los estimadores de direcciones principales funcionales robustas

a) penalizando la norma:

$$\left\{ \begin{array}{l} \widehat{\phi}_{\text{PN},1} = \underset{\|\alpha\|_r=1}{\text{argmax}} s_n^2(\alpha) = \underset{\beta \neq 0}{\text{argmax}} \frac{s_n^2(\beta)}{\langle \beta, \beta \rangle + r\Psi(\beta)} \\ \widehat{\phi}_{\text{PN},m} = \underset{\alpha \in \widehat{\mathcal{B}}_{m,r,\text{PN}}}{\text{argmax}} s_n^2(\alpha) \quad 2 \leq m, \end{array} \right. \quad (3.5)$$

donde $\widehat{\mathcal{B}}_{m,r,\text{PN}} = \{\alpha \in \mathcal{H} : \|\alpha\|_r = 1, \langle \alpha, \widehat{\phi}_{\text{PN},j} \rangle_r = 0, \forall 1 \leq j \leq m-1\}$,

o bien

b) penalizando la escala:

$$\begin{cases} \widehat{\phi}_{\text{PS},1} &= \operatorname{argmax}_{\|\alpha\|=1} \{s_n^2(\alpha) - \rho\Psi(\alpha)\} \\ \widehat{\phi}_{\text{PS},m} &= \operatorname{argmax}_{\alpha \in \widehat{\mathcal{B}}_{m,\text{PS}}} \{s_n^2(\alpha) - \rho\Psi(\alpha)\} \quad 2 \leq m, \end{cases} \quad (3.6)$$

donde $\widehat{\mathcal{B}}_{m,\text{PS}} = \{\alpha \in \mathcal{H} : \|\alpha\| = 1, \langle \alpha, \widehat{\phi}_{\text{PS},j} \rangle = 0, \forall 1 \leq j \leq m-1\}$.

Los correspondientes estimadores de valores principales se definen respectivamente como

$$\widehat{\lambda}_{\text{PS},m} = s_n^2(\widehat{\phi}_{\text{PS},m}) \quad \text{and} \quad \widehat{\lambda}_{\text{PN},m} = s_n^2(\widehat{\phi}_{\text{PN},m}). \quad (3.7)$$

3.2.4 Estimadores por *sieves* de las direcciones principales

Vinculado a B -splines, y más generalmente, al método de *sieves*, podemos definir una propuesta alternativa. El método de *sieves* consisten en aproximar un espacio infinito-dimensional Θ por una sucesión de espacios finito-dimensionales Θ_n , con dimensión dependiente del tamaño muestral n , y usa estos espacios para estimar los parámetros en el espacio Θ_n , en vez de en el espacio original, Θ .

Sea $\{\delta_i\}_{i \geq 1}$ una base de \mathcal{H} y definamos \mathcal{H}_{p_n} el espacio vectorial generado por $\delta_1, \dots, \delta_{p_n}$ y por $\mathcal{S}_{p_n} = \{\alpha \in \mathcal{H}_{p_n} : \|\alpha\| = 1\}$, o sea, $\mathcal{H}_{p_n} = \{\alpha \in \mathcal{H} : \alpha = \sum_{j=1}^{p_n} a_j \delta_j\}$ y $\mathcal{S}_{p_n} = \{\alpha \in \mathcal{H} : \alpha = \sum_{j=1}^{p_n} a_j \delta_j, \text{ tal que } \|\alpha\|^2 = \sum_{j=1}^{p_n} \sum_{s=1}^{p_n} a_j a_s \langle \delta_j, \delta_s \rangle = 1\}$. Notemos que \mathcal{S}_{p_n} aproxima la esfera unidad $\mathcal{S} = \{\alpha \in \mathcal{H} : \|\alpha\| = 1\}$. Además, cuando $\{\delta_i\}_{i \geq 1}$ es una base ortonormal, $\|\alpha\|^2 = \sum_{j=1}^{p_n} a_j^2 = \mathbf{a}^T \mathbf{a}$ donde $\mathbf{a} = (a_1, \dots, a_{p_n})^T$, es decir, la norma de α es igual a la norma euclídea del vector \mathbf{a} . Definimos pues los estimadores robustos mediante *sieves* de las componentes principales como

$$\begin{cases} \widehat{\phi}_{\text{SI},1} &= \operatorname{argmax}_{\alpha \in \mathcal{S}_{p_n}} s_n(\alpha) \\ \widehat{\phi}_{\text{SI},m} &= \operatorname{argmax}_{\alpha \in \widehat{\mathcal{B}}_{n,m}} s_n(\alpha) \quad 2 \leq m, \end{cases} \quad (3.8)$$

donde $\widehat{\mathcal{B}}_{n,m} = \{\alpha \in \mathcal{S}_{p_n} : \langle \alpha, \widehat{\phi}_{\text{SI},j} \rangle = 0, \forall 1 \leq j \leq m-1\}$, y sean los estimadores de valores principales

$$\widehat{\lambda}_{\text{SI},m} = s_n^2(\widehat{\phi}_{\text{SI},m}). \quad (3.9)$$

Algunas de las bases más frecuentemente usadas en el análisis de datos funcionales son las bases de Fourier, las polinomiales, las de *splines* y las de ondoletas. Consultar Ramsay y Silverman (2005) por más detalles.

3.2.5 Una formulación unificada para las alternativas robustas mediante *projection pursuit*

De forma tal de proveer un acercamiento unificado a todas las propuestas anteriores, definiremos el producto $\rho\Psi(\alpha)$ o $r\Psi(\alpha)$ como 0 cuando $\rho = 0$ o $r = 0$, incluso si $\alpha \notin \mathcal{H}_s$, es decir, incluso si $\Psi(\alpha) = \infty$. Más aún, cuando $p_n = \infty$, definiremos $\mathcal{H}_{p_n} = \mathcal{H}$.

Todas las propuestas de antes pueden ser vistas como casos especiales de la siguiente clase general de estimadores.

$$\begin{cases} \hat{\phi}_1 &= \operatorname{argmax}_{\alpha \in \mathcal{H}_{p_n}, \|\alpha\|_r=1} \{s_n^2(\alpha) - \rho\Psi(\alpha)\} \\ \hat{\phi}_m &= \operatorname{argmax}_{\alpha \in \hat{\mathcal{B}}_{m,r}} \{s_n^2(\alpha) - \rho\Psi(\alpha)\} \quad 2 \leq m, \end{cases} \quad (3.10)$$

donde $\hat{\mathcal{B}}_{m,r} = \{\alpha \in \mathcal{H}_{p_n} : \|\alpha\|_r = 1, \langle \alpha, \hat{\phi}_j \rangle_r = 0, \forall 1 \leq j \leq m-1\}$.

Tomando $p_n = \infty$, los estimadores “crudos” se obtienen cuando $\rho = r = 0$, mientras que $\hat{\phi}_{\text{PN},m}$ y $\hat{\phi}_{\text{PS},m}$ corresponden a $\rho = 0$ y $r = 0$, respectivamente. Por otro lado, los estimadores de *sieves* corresponden a elecciones finitas para p_n y $r = \rho = 0$.

Antes de probar la consistencia de los estimadores propuestos, en la próxima sección daremos algunos resultados auxiliares necesarios en la demostración. Por otra parte, la demostración de la consistencia de (3.10) puede verse en Bali *et al.* (2011). Probaremos la consistencia de (3.3), (3.5), (3.6) y (3.8) ya que algunas de las condiciones dadas en Bali *et al.* (2011) pueden debilitarse al considerar cada caso separadamente.

De ahora en más, indicaremos por $\xrightarrow{a.s.}$ a la convergencia en casi todo punto y por $\xrightarrow{\omega}$ a la convergencia débil y por $o_{a.s.}(1)$ a una magnitud que converge a 0 en casi todo punto.

3.3 Resultados Auxiliares

Enunciaremos a continuación algunos lemas que nos resultarán útiles para derivar resultados concernientes a la consistencia de las propuestas y a algunas de las hipótesis que se pedirán para obtener dicha consistencia.

Sea \mathcal{M} un espacio métrico separable y completo y \mathcal{B} la σ -álgebra de Borel de \mathcal{M} . Recordemos que la distancia de Prohorov entre dos medidas de probabilidad P y Q sobre \mathcal{M} se define como $d_{\text{PR}}(P, Q) = \inf\{\epsilon, P(A) \leq Q(A^\epsilon) + \epsilon, \forall A \in \mathcal{B}\}$, donde $A^\epsilon = \{x \in \mathcal{M}, d(x, A) < \epsilon\}$. Luego $P_n \xrightarrow{\omega} P$ si y sólo si $d_{\text{PR}}(P_n, P) \rightarrow 0$. Los Teoremas 3.3.1 y 3.3.2 muestran que, análogamente al Teorema de Glivenko–Cantelli en el caso finito dimensional, en un espacio normado completo y separable las medidas empíricas convergen casi seguramente de forma débil a la medida de probabilidad que genera las observaciones, en forma uniforme sobre los elementos del dual.

Teorema 3.3.1. *Sea $(\Omega, \mathcal{A}, \mathbb{P})$ un espacio de probabilidad y $X_n : \Omega \rightarrow \mathcal{M}$, $n \in \mathbb{N}$, una sucesión de elementos aleatorios independientes e idénticamente distribuidos tales que $X_i \sim P$. Supongamos que \mathcal{M} es un espacio polaco y denotemos por P_n a la medida empírica, eso es, $P_n(A) = \frac{1}{n} \sum_1^n I_A(X_i)$ con $I_A(X_i) = 1$ si $X_i \in A$ y 0 en otro lado. Entonces, $P_n \xrightarrow{\omega} P$ casi seguramente, es decir, $d_{\text{PR}}(P_n, P) \xrightarrow{a.s.} 0$.*

DEMOSTRACIÓN. Notemos que la ley fuerte de los grandes números implica que para cualquier conjunto boreliano A , $P_n(A) \xrightarrow{a.s.} P(A)$. Es decir, $P_n(A) \rightarrow P(A)$ excepto por un conjunto $\mathcal{N}_A \subset \Omega$ de \mathbb{P} -medida cero.

Probemos que dado $j \in \mathbb{N}$, existe $\mathcal{N}_j \subset \Omega$ tal que $\mathbb{P}(\mathcal{N}_j) = 0$ y, para cualquier $\omega \notin \mathcal{N}_j$, existe $n_j(\omega) \in \mathbb{N}$ tal que si $n \geq n_j(\omega)$ se cumple que $d_{\text{PR}}(P_n, P) < 1/j$.

Como \mathcal{M} es un espacio polaco, existe una clase finita de conjuntos $\{A_i, 1 \leq i \leq k\}$ con diámetro menor que $\frac{1}{2j}$ tal que

$$P\left(\bigcup_{i=1}^k A_i\right) > 1 - \frac{1}{2j}. \quad (3.11)$$

Indiquemos por \mathcal{A} a la clase de todos los conjuntos que se obtienen como unión finita de los A_i , es decir, $B \in \mathcal{A}$ si y sólo si existen $A_{i_1}, \dots, A_{i_\ell}$ tal que $B = \bigcup_{j=1}^{\ell} A_{i_j}$. Notemos que \mathcal{A} tiene una cantidad finita de elementos s . Para cada $1 \leq i \leq s$, y $B_i \in \mathcal{A}$, sea $\mathcal{N}_{B_i} \subset \Omega$ con $\mathbb{P}(\mathcal{N}_{B_i}) = 0$ tal que si $\omega \notin \mathcal{N}_{B_i}$, luego $|P_n(B_i) - P(B_i)| \rightarrow 0$. Definimos $\mathcal{N}_j = \bigcup_{i=1}^s \mathcal{N}_{B_i}$, luego $\mathbb{P}(\mathcal{N}_j) = 0$.

Sea $\omega \notin \mathcal{N}_j$, luego tendremos que $|P_n(B_i) - P(B_i)| \rightarrow 0$, para $1 \leq i \leq s$. Por lo tanto, existirá entonces $n_j(\omega) \in \mathbb{N}$ tal que para $n \geq n_j(\omega)$, tendremos que $|P_n(B) - P(B)| < \frac{1}{2j}$ para todo $B \in \mathcal{A}$. Probaremos ahora que si $n \geq n_j(\omega)$ entonces $d_{\text{PR}}(P_n, P) < 1/j$.

Consideremos B un conjunto boreliano y sea A la unión de todos los conjuntos A_i que intersecan con B . Notemos que $A \in \mathcal{A}$ y luego $|P_n(A) - P(A)| < \frac{1}{2j}$. Por ende, $B \subset A \cup \left(\bigcup_{i=1}^k A_i\right)^c$ y $A \subset B^{1/j}$. Esta última inclusión vale pues los conjuntos A_i tienen diámetro menor que $\frac{1}{2j}$. Luego, usando (3.11), obtendremos que

$$P(B) \leq P(A) + P\left[\left(\bigcup_{i=1}^k A_i\right)^c\right] = P(A) + 1 - P\left[\left(\bigcup_{i=1}^k A_i\right)\right] < P(A) + \frac{1}{2j},$$

que, junto con $|P_n(A) - P(A)| < \frac{1}{2j}$, implica que $P(B) \leq P(A) + \frac{1}{2j} < P_n(A) + 1/j$. Usando que $A \subset B^{1/j}$, obtenemos que $P_n(A) + 1/j \leq P_n(B^{1/j}) + 1/j$, de donde $P(B) < P_n(B^{1/j}) + 1/j$ y esto vale para cualquier conjunto B boreliano. Entonces, $d_{\text{PR}}(P_n, P) < 1/j$, como se deseaba.

Para concluir la demostración, demostraremos que $d_{\text{PR}}(P_n, P) \rightarrow 0$ excepto por un conjunto de \mathbb{P} -medida cero. Consideremos todos los conjuntos \mathcal{N}_j antes definidos y sea $\mathcal{N} = \bigcup_{j \in \mathbb{N}} \mathcal{N}_j$. Es claro que $\mathbb{P}(\mathcal{N}) = 0$. Entonces, para todo $\omega \notin \mathcal{N}$, tenemos que, para cada j , existe $n_h = n_j(\omega)$ tal que $d(P_n, P) < 1/j$ si $n \geq n_j$. Esto finaliza la demostración. \square

Sea P una medida de probabilidad en \mathcal{M} , un espacio de Banach separable. Luego, dado $f \in \mathcal{M}^*$, con \mathcal{M}^* el dual algebraico-topológico de \mathcal{M} , definamos $P[f]$ como la medida univariada de la variable aleatoria $f(X)$, con $X \sim P$. El siguiente teorema muestra el análogo a Glivenko-Cantelli al caso de dimensión infinita.

Teorema 3.3.2. *Sea $\{P_n\}_{n \in \mathbb{N}}$ y P medidas de probabilidad definidas en \mathcal{M} tal que $P_n \xrightarrow{\omega} P$, es decir, $d_{\text{PR}}(P_n, P) \rightarrow 0$. Entonces, $\sup_{\|f\|_* = 1} d_{\text{PR}}(P_n[f], P[f]) \rightarrow 0$.*

DEMOSTRACIÓN. Fijemos $\epsilon > 0$ y sea n_0 tal que $d_{\text{PR}}(P_n, P) < \epsilon$, para $n \geq n_0$. Probaremos que $\sup_{\|f\|_* = 1} d_{\text{PR}}(P_n[f], P[f]) < \epsilon$, para $n \geq n_0$.

Usando que $d_{\text{PR}}(P_n, P) < \epsilon$ y el Teorema de Strassen, obtenemos que existen $\{X_n\}_{n \in \mathbb{N}}$ y X en \mathcal{M} tal que $X_n \sim P_n$, $X \sim P$ y $\mathbb{P}(\|X_n - X\| \leq \epsilon) > 1 - \epsilon$.

Notemos que, para cualquier $f \in \mathcal{M}^*$, con $\|f\|_* = 1$, $f(X_n) \sim P_n[f]$ y $f(X) \sim P[f]$. Usando que $|f(X_n) - f(X)| = |f(X_n - X)| \leq \|f\|_* \|X_n - X\| \leq \|X_n - X\|$, obtenemos que para todo $f \in \mathcal{M}^*$, tal que $\|f\|_* = 1$,

$$\{\|X_n - X\| \leq \epsilon\} \subseteq \{|f(X_n) - f(X)| \leq \epsilon\}$$

lo que implica

$$1 - \epsilon < \mathbb{P}(\|X_n - X\| \leq \epsilon) \leq \mathbb{P}(|f(X_n) - f(X)| \leq \epsilon), \quad \forall f \in \mathcal{M}^*, \quad \|f\|_* = 1.$$

Entonces, $\mathbb{P}(|f(X_n) - f(X)| \leq \epsilon) > 1 - \epsilon$, y luego, usando nuevamente el teorema de Strassen, obtenemos que

$$P_n[f](A) \leq P[f](A^\epsilon) + \epsilon, \quad \forall A \in \mathcal{B}, \quad \forall f \in \mathcal{M}^*, \quad \|f\|_* = 1.$$

Entonces, para todo $f \in \mathcal{M}^*$ tal que $\|f\|_* = 1$, tendremos que $d_{\text{PR}}(P_n[f], P[f]) \leq \epsilon$, es decir, $\sup_{\|f\|_* = 1} d_{\text{PR}}(P_n[f], P[f]) \leq \epsilon$, concluyendo la prueba. \square

En particular, al considerar un espacio de Hilbert separable \mathcal{H} , si $f \in \mathcal{H}^*$ es tal que $\|f\|_* = 1$, existe $\alpha \in \mathcal{H}$ tal que $f(X) = \langle \alpha, X \rangle$ con $\|\alpha\| = 1$. El siguiente resultado establece que cuando σ_{R} es un funcional continuo de escala, entonces se alcanza la convergencia uniforme sobre la esfera unidad de las medidas de proyecciones.

Teorema 3.3.3. *Sea $\{P_n\}_{n \in \mathbb{N}}$ y P medidas de probabilidad definidas en un espacio de Hilbert separable \mathcal{H} , tal que $P_n \xrightarrow{\omega} P$, es decir, $d_{\text{PR}}(P_n, P) \rightarrow 0$. Sea σ_{R} un funcional continuo de escala. Entonces $\sup_{\|\alpha\|=1} |\sigma_{\text{R}}(P_n[\alpha]) - \sigma_{\text{R}}(P[\alpha])| \rightarrow 0$.*

DEMOSTRACIÓN. Indiquemos por $a_n = \sup_{\|\alpha\|=1} |\sigma_{\text{R}}(P_n[\alpha]) - \sigma_{\text{R}}(P[\alpha])|$, basta ver que $L = \limsup_{n \rightarrow \infty} a_n = 0$.

Primero notemos que dado que $\mathcal{V}_1 = \{\alpha \in \mathcal{H} : \|\alpha\| \leq 1\}$ es débilmente compacto y σ_{R} es un funcional continuo, para cada n fijo tal que $a_n \neq 0$, existe $\alpha_n \in \mathcal{S}$ tal que

$$a_n = |\sigma_{\text{R}}(P_n[\alpha_n]) - \sigma_{\text{R}}(P[\alpha_n])|. \quad (3.12)$$

En efecto, sea $\gamma_\ell \in \mathcal{S}$ tal que $|\sigma_{\text{R}}(P_n[\gamma_\ell]) - \sigma_{\text{R}}(P[\gamma_\ell])| \rightarrow a_n$, luego la compacidad débil de \mathcal{S} , implica la existencia de una subsucesión γ_{ℓ_s} tal que γ_{ℓ_s} converge débilmente a $\gamma \in \mathcal{H}$ tal que $\|\gamma\| \leq 1$. Además, usando que σ_{R} es continuo obtenemos que $|\sigma_{\text{R}}(P_n[\gamma_{\ell_s}]) - \sigma_{\text{R}}(P[\gamma_{\ell_s}])| \rightarrow |\sigma_{\text{R}}(P_n[\gamma]) - \sigma_{\text{R}}(P[\gamma])|$, si $s \rightarrow \infty$. Luego, $|\sigma_{\text{R}}(P_n[\gamma]) - \sigma_{\text{R}}(P[\gamma])| = a_n$ lo que implica $\gamma \neq 0$. Sea $\tilde{\gamma} = \gamma/\|\gamma\|$, entonces $\tilde{\gamma} \in \mathcal{S}$ y, por ende, $|\sigma_{\text{R}}(P_n[\tilde{\gamma}]) - \sigma_{\text{R}}(P[\tilde{\gamma}])| \leq a_n$. Por otro lado, usando que σ_{R} es un funcional de escala obtenemos que

$$|\sigma_{\text{R}}(P_n[\tilde{\gamma}]) - \sigma_{\text{R}}(P[\tilde{\gamma}])| = \frac{|\sigma_{\text{R}}(P_n[\gamma]) - \sigma_{\text{R}}(P[\gamma])|}{\|\gamma\|} = \frac{a_n}{\|\gamma\|}$$

implicando que $\|\gamma\| \geq 1$ y llevándonos así a $\|\gamma\| = 1$ y a la existencia de una sucesión $\alpha_n \in \mathcal{S}$ que satisface (3.12).

Sea a_{n_k} una subsucesión tal que $a_{n_k} \rightarrow L$, supondremos que $a_{n_k} \neq 0$. Luego, usando (3.12), tendremos que existe $\alpha_{n_k} \in \mathcal{S}$ tal que $a_{n_k} = |\sigma_{\mathbb{R}}(P_{n_k}[\alpha_{n_k}]) - \sigma_{\mathbb{R}}(P[\alpha_{n_k}])| \rightarrow L$. Usando que $\mathcal{S} \subset \mathcal{V}_1$ es débilmente compacto, podemos elegir una subsucesión $\beta_j = \alpha_{n_{k_j}}$ tal que β_j converge débilmente a β , es decir, para todo $\alpha \in \mathcal{H}$, $\langle \beta_j, \alpha \rangle \rightarrow \langle \beta, \alpha \rangle$. Puesto que $\|\beta_j\| = 1$, entonces $\|\beta\| \leq 1$ y luego

$$a_{n_{k_j}} = |\sigma_{\mathbb{R}}(P_{n_{k_j}}[\beta_j]) - \sigma_{\mathbb{R}}(P[\beta_j])| \rightarrow L \quad (3.13)$$

Para simplificar la notación, escribiremos $P^{(j)} = P_{n_{k_j}}$. Luego, el teorema 3.3.2 implica que

$$d_{\text{PR}}(P^{(j)}[\beta_j], P[\beta_j]) \leq \sup_{\|\alpha\|=1} d_{\text{PR}}(P^{(j)}[\alpha], P[\alpha]) \rightarrow 0$$

mientras que el hecho que β_j converge débilmente a β implica $d_{\text{PR}}(P[\beta_j], P[\beta]) \rightarrow 0$, concluyendo así que $d_{\text{PR}}(P^{(j)}[\beta_j], P[\beta]) \rightarrow 0$. La continuidad de $\sigma_{\mathbb{R}}$ nos lleva a

$$\sigma_{\mathbb{R}}(P^{(j)}[\beta_j]) \rightarrow \sigma_{\mathbb{R}}(P[\beta]) . \quad (3.14)$$

Usando nuevamente que β_j converge débilmente a β y la continuidad débil de $\sigma_{\mathbb{R}}$ obtenemos que

$$\sigma_{\mathbb{R}}(P[\beta_j]) \rightarrow \sigma_{\mathbb{R}}(P[\beta]) . \quad (3.15)$$

Luego, (3.14) y (3.15) implican que $\sigma_{\mathbb{R}}(P^{(j)}[\beta_j]) - \sigma_{\mathbb{R}}(P[\beta_j]) \rightarrow 0$ y entonces, de (3.13), $L = 0$, concluyendo así la demostración. \square

Más aún, usando el Teorema 3.3.1, obtenemos el siguiente resultado que posteriormente utilizaremos para mostrar que se cumplen las hipótesis necesarias para garantizar la consistencia de las propuestas, si $\sigma_{\mathbb{R}}$ es un funcional continuo.

Corolario 3.3.1. *Sean P una medida de probabilidad en un espacio de Hilbert separable \mathcal{H} , P_n la medida empírica de una muestra aleatoria X_1, \dots, X_n con $X_i \sim P$, y $\sigma_{\mathbb{R}}$ un funcional continuo de escala. Luego, tenemos que*

$$\sup_{\|\alpha\|=1} |\sigma_{\mathbb{R}}(P_n[\alpha]) - \sigma_{\mathbb{R}}(P[\alpha])| \xrightarrow{a.s.} 0.$$

3.4 Consistencia de los estimadores robustos de las direcciones principales

3.4.1 Supuestos

Recordemos que habíamos definido, dado un P medida de probabilidad en \mathcal{H} y un σ_R funcional unariado de escala, las siguientes aplicaciones:

$$\sigma : \mathcal{H} \rightarrow \mathbb{R}, \quad \sigma(\alpha) = \sigma_R(P[\alpha])$$

y

$$s_n : \mathcal{H} \rightarrow \mathbb{R}, \quad \sigma(\alpha) = \sigma_R(P_n[\alpha])$$

donde P_n es la medida empírica de P en una muestra X_1, \dots, X_n , $X \sim P$.

Para poder derivar la consistencia de los estimadores propuestos, precisaremos los siguientes supuestos.

S0. Para algún $q \geq 2$, $\phi_{R,j}(P)$ es único salvo elección de signo para $1 \leq j \leq q$.

S1. $\sup_{\|\alpha\|=1} |s_n^2(\alpha) - \sigma^2(\alpha)| \xrightarrow{a.s.} 0$

S2. $\sigma : \mathcal{H} \rightarrow \mathbb{R}$ es una función débilmente continua. Es decir, continua con respecto a la topología débil en \mathcal{H} .

Observación 3.4.1.1. Discutiremos a continuación la validez de las hipótesis anteriores.

- El supuesto **S0** vale si y sólo si $\lambda_{R,1} > \dots > \lambda_{R,q} > \lambda_{R,q+1}$.
- El supuesto **S1** vale para los estimadores clásicos basados en la varianza muestral puesto que el operador de covarianza empírico, $\widehat{\Gamma}$, es consistente. En efecto, como se prueba en Dauxois *et al.* (1982), $\|\widehat{\Gamma} - \Gamma_X\| \xrightarrow{a.s.} 0$, lo que implica que $\sup_{\|\alpha\|=1} |s_n^2(\alpha) - \sigma^2(\alpha)| \leq \|\widehat{\Gamma} - \Gamma_X\| \xrightarrow{a.s.} 0$. Sin embargo, este supuesto puede ser más difícil de verificar para otros funcionales de escala puesto que la esfera unitaria $\mathcal{S} = \{\|\alpha\| = 1\}$ no es compacta. Las condiciones más débiles $\sup_{\|\alpha\|_r=1} |s_n^2(\alpha) - \sigma^2(\alpha)| \xrightarrow{a.s.} 0$ o bien $\sup_{\alpha \in \mathcal{S}_{p_n}} |s_n^2(\alpha) - \sigma^2(\alpha)| \xrightarrow{a.s.} 0$ pueden en cambio ser consideradas, puesto que los conjuntos $\{\alpha \in \mathcal{S}_{p_n}\}$ y $\{\alpha, \|\alpha\|_r = 1\}$ son compactos. El Corolario 3.3.1 muestra que si σ_R es un funcional débilmente continuo (cualitativamente robusto) entonces vale **S1**. Por lo tanto, la hipótesis **S1** se cumple para una clase amplia de estimadores de escala.
- Si el funcional de escala σ_R es continuo también se cumple la condición **S2**. Efectivamente, si $\alpha_k \rightarrow \alpha$, cuando $k \rightarrow \infty$, entonces $\langle \alpha_k, X \rangle \xrightarrow{\omega} \langle \alpha, X \rangle$ y por lo tanto, $\sigma_R(P[\alpha_k]) \rightarrow \sigma_R(P[\alpha])$.

En el caso en que el funcional de escala es el desvío estándar y la medida de probabilidad subyacente P admite un operador de covarianza $\mathbf{\Gamma}_X$, se cumple que $\sigma^2(\alpha) = \langle \alpha, \mathbf{\Gamma}_X \alpha \rangle$. Por lo tanto, a pesar de que el desvío estándar no es un funcional débilmente continuo, la condición **S2** se verifica también ya que $\mathbf{\Gamma}_X$ es un operador compacto, .

Si X tiene distribución elíptica $\mathcal{E}(\mu, \mathbf{\Gamma})$, el Lema 2.3.1 implica que existe una constante c positiva tal que c para todo $\alpha \in \mathcal{H}$, $\sigma_{\mathbf{R}}^2(P[\alpha]) = c\langle \alpha, \mathbf{\Gamma} \alpha \rangle$. Por lo tanto, la función σ será débilmente continua.

- Observemos además que como existe una métrica d que genera la topología débil en \mathcal{H} y la bola cerrada $\mathcal{V}_r = \{\alpha : \|\alpha\| \leq r\}$ es débilmente compacta, **S2** implica que $\sigma(\alpha)$ es uniformemente continua con respecto a d y, por lo tanto, con respecto a la topología débil, sobre \mathcal{H} . Usaremos la continuidad uniforme débil en los resultados que siguen.

3.4.2 Consistencia Fisher

En esta sección haremos algunos comentarios con respecto a la Fisher-consistencia de los funcionales definidos en la Sección 3.2.1. Hemos definido en (3.1) y (3.2) una serie de funcionales y nos interesa saber lo que dichos funcionales representan, es decir, bajo qué condiciones coinciden con cantidades conocidas o de interés.

Supongamos que la medida de probabilidad P admita un operador de dispersión $\mathbf{\Gamma}$ con autofunciones ϕ_j y autovalores asociados λ_j tales que $\lambda_1 \geq \lambda_2 \geq \dots$. Sería deseable que $\phi_{\mathbf{R},j}(P) = \phi_j$ y $\lambda_{\mathbf{R},j}(P) = \lambda_j$. El siguiente supuesto nos permitirá deducir esta propiedad:

- S3.** Existe una constante $c > 0$ y un operador auto-adjunto, semidefinido positivo y compacto $\mathbf{\Gamma}$, tal que para todo $\alpha \in \mathcal{H}$, tenemos $\sigma^2(\alpha) = c\langle \alpha, \mathbf{\Gamma} \alpha \rangle$.

Es fácil ver, si vale la hipótesis **S3**, entonces el funcional de las direcciones principales definido en (3.1) es Fisher-consistente, en el sentido que $\phi_{\mathbf{R},j}(P) = \phi_j$, donde ϕ_j es la autofunción de $\mathbf{\Gamma}$ con autovalor asociado λ_j tal que $\lambda_1 \geq \lambda_2 \geq \dots$. No ocurre lo mismo con el funcional de valores principales, debido a la constante c . Efectivamente, $\lambda_{\mathbf{R},j}(P) = c\lambda_j$. Sin embargo, es posible calibrar el funcional de escala, dividiéndolo por c , para obtener Fisher-consistencia para el funcional de valores principales en la probabilidad de interés P .

Por ejemplo, si $P = \mathcal{E}(\mu, \mathbf{\Gamma})$, es decir, bajo elipticidad funcional, vale **S3** y por lo tanto, los funcionales de direcciones principales resultan Fisher-consistentes, mientras que los de los valores principales lo son si $\sigma_{\mathbf{R}}(F_0) = 1$ con F_0 la distribución dada en el Lema 2.1.2.

El supuesto **S3** asegura en el fondo de que estamos estimando las direcciones deseadas. Parece ser algo restrictivo puesto que es difícil de verificar fuera del caso de las familias elípticas, excepto cuando la escala es el desvío estándar. Sin embargo, incluso en el caso finito-dimensional, las propiedades asintóticas se obtuvieron solamente bajo restricciones similares. Por ejemplo, Li y Chen (1985) y Croux y Ruiz-Gazen (2005) suponen una distribución subyacente elíptica para poder deducir la consistencia de los estimadores de *projection-pursuit* y/o calcular su función de influencia, respectivamente. Por otra parte, Cui *et al.* (2003) suponen que la función de influencia

de los datos proyectados es de la forma $h(\mathbf{x}, \mathbf{a}) = 2\sigma(F[\mathbf{a}])\text{IF}(\mathbf{x}, \sigma_{\mathbf{a}}; F_0)$, donde $F[\mathbf{a}]$ es la distribución de $\mathbf{a}^T \mathbf{x}$ cuando $\mathbf{x} \sim F$. Esta condición, sin embargo, vale fundamentalmente cuando la distribución es elíptica.

3.4.3 Continuidad de los operadores

Antes de obtener los resultados de consistencia, fijaremos alguna notación y probaremos la continuidad de los funcionales de direcciones y valores principales. En el resto de este capítulo, $\phi_{R,j}(P)$ y $\lambda_{R,j}(P)$ serán los funcionales definidos mediante (3.1) y (3.2). Para simplificar la notación, denotemos por $\lambda_{R,j} = \lambda_{R,j}(P)$ y $\phi_{R,j} = \phi_{R,j}(P)$.

Indiquemos por \mathcal{L}_{m-1} al subespacio lineal generado por $\{\phi_{R,1}, \dots, \phi_{R,m-1}\}$ y sea $\widehat{\mathcal{L}}_{m-1}$ el subespacio lineal generado por los primeros $m-1$ estimadores de las direcciones principales. Es decir, por $\{\widehat{\phi}_1, \dots, \widehat{\phi}_{m-1}\}$, $\{\widehat{\phi}_{\text{PS},1}, \dots, \widehat{\phi}_{\text{PS},m-1}\}$, $\{\widehat{\phi}_{\text{PN},1}, \dots, \widehat{\phi}_{\text{PN},m-1}\}$ o $\{\widehat{\phi}_{\text{SI},1}, \dots, \widehat{\phi}_{\text{SI},m-1}\}$, donde será claro en cada caso que espacio lineal se estará considerando. Finalmente, para cualquier espacio lineal \mathcal{L} , sea $\pi_{\mathcal{L}} : \mathcal{H} \rightarrow \mathcal{L}$ el operador de proyección ortogonal sobre el espacio lineal \mathcal{L} , que existirá si \mathcal{L} es un subespacio cerrado. En particular, $\pi_{\mathcal{L}_{m-1}}$, $\pi_{\widehat{\mathcal{L}}_{m-1}}$ y $\pi_{\mathcal{H}_{p_m}}$ estarán bien definidos pues son proyecciones sobre un subespacio de dimensión finita, por ende cerrado.

El siguiente lema nos resultará útil para obtener la consistencia y la continuidad de los estimadores de autofunciones. En el siguiente lema y en las proposiciones subsiguientes, se deberá notar que $\langle \widehat{\phi}, \phi \rangle^2 \rightarrow 1$ implica, bajo el mismo modo de convergencia, que el signo de $\widehat{\phi}$ se puede elegir de forma tal que $\widehat{\phi} \rightarrow \phi$.

Lema 3.4.1. *Sea $\widehat{\phi}_m \in \mathcal{S}$ tal que $\langle \widehat{\phi}_m, \widehat{\phi}_j \rangle = 0$ para $j \neq m$. Si **S0** y **S2** valen, se cumple*

a) *Si $\sigma^2(\widehat{\phi}_1) \xrightarrow{a.s.} \sigma^2(\phi_{R,1})$, entonces, $\langle \widehat{\phi}_1, \phi_{R,1} \rangle^2 \xrightarrow{a.s.} 1$.*

b) *Dados $2 \leq m \leq q$, si $\sigma^2(\widehat{\phi}_m) \xrightarrow{a.s.} \sigma^2(\phi_{R,m})$ y $\widehat{\phi}_s \xrightarrow{a.s.} \phi_{R,s}$, para $1 \leq s \leq m-1$, entonces para $1 \leq m \leq q$, $\langle \widehat{\phi}_m, \phi_{R,m} \rangle^2 \xrightarrow{a.s.} 1$.*

DEMOSTRACIÓN. a) Sea $\mathcal{N} = \{\omega : \sigma^2(\widehat{\phi}_1(\omega)) \not\rightarrow \sigma^2(\phi_{R,1})\}$ y fijemos $\omega \notin \mathcal{N}$, entonces $\sigma^2(\widehat{\phi}_1(\omega)) \rightarrow \sigma^2(\phi_{R,1})$. Usando que $\mathcal{S} \subset \{\alpha : \|\alpha\| \leq 1\}$ que es débilmente compacto, tenemos que para cualquier subsucesión γ_ℓ de $\widehat{\phi}_1(\omega)$ existe una subsucesión γ_{ℓ_s} tal que γ_{ℓ_s} converge débilmente a $\gamma \in \mathcal{H}$ tal que $\|\gamma\| \leq 1$.

Además, usando que $\sigma^2(\widehat{\phi}_1(\omega)) \rightarrow \sigma^2(\phi_{R,1})$, obtenemos que $\sigma^2(\gamma_{\ell_s}) \rightarrow \sigma^2(\phi_{R,1})$ mientras que por otro lado, la continuidad débil de σ implica que $\sigma^2(\gamma_{\ell_s}) \rightarrow \sigma^2(\gamma)$, con $s \rightarrow \infty$. Luego, $\sigma^2(\gamma) = \sigma^2(\phi_{R,1})$, implicando que $\gamma \neq 0$. Sea $\tilde{\gamma} = \gamma/\|\gamma\|$, entonces $\tilde{\gamma} \in \mathcal{S}$ y por lo tanto $\sigma^2(\tilde{\gamma}) \leq \sigma^2(\phi_{R,1})$. Por otro lado, usando que σ_R es un funcional de escala obtenemos que

$$\sigma(\tilde{\gamma}) = \frac{\sigma(\gamma)}{\|\gamma\|} = \frac{\sigma(\phi_{R,1})}{\|\gamma\|}$$

lo que implica que $\|\gamma\| \geq 1$, teniendo así que $\|\gamma\| = 1$ y por ende, usando la unicidad de $\phi_{R,1}$ obtenemos que $\langle \gamma, \phi_{R,1} \rangle^2 = 1$. Por lo tanto, dado que cualquier subsucesión de $\widehat{\phi}_1(\omega)$ tendrá un

límite convergiendo o bien a $\phi_{R,1}$ o bien a $-\phi_{R,1}$, obtenemos a).

b) Escribamos $\widehat{\phi}_m$ como $\widehat{\phi}_m = \sum_{j=1}^{m-1} \widehat{a}_j \phi_{R,j} + \widehat{\gamma}_m$, con $\langle \widehat{\gamma}_m, \phi_{R,j} \rangle = 0$, $1 \leq j \leq m-1$. Para obtener b) solamente debemos demostrar que $\langle \widehat{\gamma}_m, \phi_{R,m} \rangle^2 \xrightarrow{a.s.} 1$. Notemos que $\langle \widehat{\phi}_m, \widehat{\phi}_j \rangle = 0$, para $j \neq m$, implicando que $\widehat{a}_j = \langle \widehat{\phi}_m, \phi_{R,j} \rangle = \langle \widehat{\phi}_m, \phi_{R,j} - \widehat{\phi}_j \rangle + \langle \widehat{\phi}_m, \widehat{\phi}_j \rangle = \langle \widehat{\phi}_m, \phi_{R,j} - \widehat{\phi}_j \rangle$. Luego, usando que $\widehat{\phi}_j \xrightarrow{a.s.} \phi_{R,j}$, $1 \leq j \leq m-1$, y $\|\widehat{\phi}_m\| = 1$, obtenemos que $\widehat{a}_j \xrightarrow{a.s.} 0$ para $1 \leq j \leq m-1$ y entonces, $\|\widehat{\phi}_m - \widehat{\gamma}_m\| \xrightarrow{a.s.} 0$. Notemos que $1 = \|\widehat{\phi}_m\|^2 = \sum_{j=1}^{m-1} \widehat{a}_j^2 + \|\widehat{\gamma}_m\|^2$, luego, $\|\widehat{\gamma}_m\|^2 \xrightarrow{a.s.} 1$ implicando que $\|\widehat{\phi}_m - \widehat{\gamma}_m\| \xrightarrow{a.s.} 0$, donde $\widetilde{\gamma}_m = \widehat{\gamma}_m / \|\widehat{\gamma}_m\|$. Usando que $\sigma(\alpha)$ es una función débilmente continua y que la bola unitaria es débilmente compacta, obtenemos que

$$\sigma(\widetilde{\gamma}_m) - \sigma(\widehat{\phi}_m) \xrightarrow{a.s.} 0. \quad (3.16)$$

Efectivamente, sea $\mathcal{N} = \Omega - \{\omega : \|\widehat{\phi}_m - \widetilde{\gamma}_m\| \xrightarrow{a.s.} 0\}$, luego $\mathbb{P}(\mathcal{N}) = 1$. Fijemos $\omega \notin \mathcal{N}$ y sea $b_n = \sigma(\widetilde{\gamma}_m) - \sigma(\widehat{\phi}_m) = \sigma(\widetilde{\gamma}_{n,m}) - \sigma(\widehat{\phi}_{n,m})$. Basta con ver que toda subsucesión de $\{b_n\}$ converge a 0. Sea $\{b_{n'}\}$ una subsucesión, luego por la compacidad débil de \mathcal{V}_1 , existirá una subsucesión $\{n_j\} \subset \{n'\}$ tal que $\widetilde{\gamma}_{n_j,m}$ y $\widehat{\phi}_{n_j,m}$ converge débilmente a γ y ϕ , respectivamente. Del hecho que $\|\widehat{\phi}_m - \widetilde{\gamma}_m\| \rightarrow 0$, obtenemos que $\gamma = \phi$ y entonces la continuidad débil de σ implica que $b_{n_j} \rightarrow 0$.

Usando que $\sigma^2(\widehat{\phi}_m) \xrightarrow{a.s.} \sigma^2(\phi_{R,m})$ y que vale (3.16) obtenemos que $\sigma(\widetilde{\gamma}_m) \xrightarrow{a.s.} \sigma(\phi_{R,m})$. La demostración sigue como en a) usando que $\widetilde{\gamma}_m \in \mathcal{C}_m$, con $\mathcal{C}_m = \{\alpha \in \mathcal{S} : \langle \alpha, \phi_{R,j} \rangle = 0, 1 \leq j \leq m-1\}$ y $\phi_{R,m}$ el único maximizador de $\sigma(\alpha)$ sobre \mathcal{C}_m . \square

La próxima proposición establece la continuidad de los funcionales definidos en (3.1) y (3.2). Por lo tanto, de acuerdo al Teorema 2.3.1, los estimadores definidos por (3.3) y (3.4) resultan cualitativamente robustos. Por otra parte, resultarán asimismo consistentes. Es decir, la propuesta dada en esta tesis efectivamente provee una sucesión de estimadores robustos de las direcciones principales, como deseábamos.

Proposición 3.4.1. *Supongamos que:*

i) *valen S0 y S2.*

ii) $\sup_{\|\alpha\|=1} |\sigma_R(P_n[\alpha]) - \sigma_R(P[\alpha])| \rightarrow 0$ *siempre que* $P_n \xrightarrow{\omega} P$.

Luego, para toda sucesión P_n *tal que* $P_n \xrightarrow{\omega} P$, *tenemos que*

a) $\lambda_{R,1}(P_n) \rightarrow \lambda_{R,1}$ *y* $\sigma^2(\phi_{R,1}(P_n)) \rightarrow \sigma^2(\phi_{R,1})$.

b) $\langle \phi_{R,1}(P_n), \phi_{R,1} \rangle^2 \rightarrow 1$.

c) *Si, para todo* $2 \leq m \leq q$, *se cumple que* $\phi_{R,s}(P_n) \rightarrow \phi_{R,s}$, *para* $1 \leq s \leq m-1$, *entonces,* $\lambda_{R,m}(P_n) \rightarrow \sigma^2(\phi_{R,m}) = \lambda_{R,m}$ *y* $\sigma^2(\phi_{R,m}(P_n)) \rightarrow \sigma^2(\phi_{R,m})$.

d) Para $1 \leq m \leq q$, $\langle \phi_{R,m}(P_n), \phi_{R,m} \rangle^2 \rightarrow 1$.

Observemos que el supuesto ii) se cumple por el Corolario 3.3.1, si σ_R es un funcional univariado de escala robusto.

DEMOSTRACIÓN. A fines de simplificar la notación, vamos a considerar $\sigma_n(\alpha) = \sigma_R(P_n[\alpha])$, $\hat{\phi}_m = \phi_{R,m}(P_n)$ y $\hat{\lambda}_m = \lambda_{R,m}(P_n) = \sigma_n(\hat{\phi}_m)$. Más aún, sea $\hat{\mathcal{B}}_m = \{\alpha \in \mathcal{H} : \|\alpha\| = 1, \langle \alpha, \hat{\phi}_j \rangle = 0, \forall 1 \leq j \leq m-1\}$ y $\hat{\mathcal{L}}_{m-1}$ el subespacio generado por $\hat{\phi}_1, \dots, \hat{\phi}_{m-1}$.

a) Usando ii), obtenemos que $a_{n,1} = \sigma_n^2(\hat{\phi}_1) - \sigma^2(\hat{\phi}_1) \rightarrow 0$ y $b_{n,1} = \sigma_n^2(\phi_{R,1}) - \sigma^2(\phi_{R,1}) \rightarrow 0$ lo que implica que

$$\sigma^2(\phi_{R,1}) = \sigma_n^2(\phi_{R,1}) - b_{n,1} \leq \sigma_n^2(\hat{\phi}_1) - b_{n,1} = \sigma^2(\hat{\phi}_1) + a_{n,1} - b_{n,1} \leq \sigma^2(\phi_{R,1}) + a_{n,1} - b_{n,1} = \sigma^2(\phi_{R,1}) + o(1),$$

donde $o(1)$ significa un término que converge a 0. Luego, $\sigma^2(\phi_{R,1}) \leq \sigma^2(\hat{\phi}_1) + o(1) \leq \sigma^2(\phi_{R,1}) + o(1)$, implica que $\sigma^2(\hat{\phi}_1) \xrightarrow{a.s.} \sigma^2(\phi_{R,1})$, concluyendo así la demostración de a) pues vale ii).

Cabe aclarar que no se ha utilizado la continuidad débil de $\sigma : \mathcal{H} \rightarrow \mathbb{R}$ para obtener a).

b) Se deduce de a) y del Lema 3.4.1 a).

c) Sea $2 \leq m \leq q$ fijo y supongamos que $\hat{\phi}_s \rightarrow \phi_{R,s}$, para $1 \leq s \leq m-1$. Empezaremos probando que $\hat{\lambda}_m \rightarrow \lambda_{R,m}$.

$$\begin{aligned} |\hat{\lambda}_m - \sigma^2(\phi_{R,m})| &= \left| \max_{\alpha \in \hat{\mathcal{B}}_m} \sigma_n^2(\alpha) - \max_{\alpha \in \mathcal{B}_m} \sigma^2(\alpha) \right| \leq \max_{\alpha \in \hat{\mathcal{B}}_m} |\sigma_n^2(\alpha) - \sigma^2(\alpha)| + \left| \max_{\alpha \in \hat{\mathcal{B}}_m} \sigma^2(\alpha) - \max_{\alpha \in \mathcal{B}_m} \sigma^2(\alpha) \right| \\ &\leq \max_{\|\alpha\|=1} |\sigma_n^2(\alpha) - \sigma^2(\alpha)| + \left| \max_{\alpha \in \hat{\mathcal{B}}_m} \sigma^2(\alpha) - \max_{\alpha \in \mathcal{B}_m} \sigma^2(\alpha) \right| \\ &\leq o(1) + \left| \max_{\alpha \in \hat{\mathcal{B}}_m} \sigma^2(\alpha) - \max_{\alpha \in \mathcal{B}_m} \sigma^2(\alpha) \right| = o(1) + \left| \max_{\alpha \in \hat{\mathcal{B}}_m} \sigma^2(\alpha) - \sigma^2(\phi_{R,m}) \right|. \end{aligned}$$

Luego, para obtener el resultado deseado, basta ver que $\max_{\alpha \in \hat{\mathcal{B}}_m} \sigma^2(\alpha) \rightarrow \sigma^2(\phi_{R,m})$. Probaremos que

$$\sigma^2(\phi_{R,m}) \leq \max_{\alpha \in \hat{\mathcal{B}}_m} \sigma^2(\alpha) + o(1) \quad (3.17)$$

$$\max_{\alpha \in \hat{\mathcal{B}}_m} \sigma^2(\alpha) \leq o(1) + \sigma^2(\phi_{R,m}) \quad (3.18)$$

Usando que $\hat{\phi}_s \rightarrow \phi_{R,s}$ para $1 \leq s \leq m-1$, obtenemos que $\|\pi_{\hat{\mathcal{L}}_{m-1}} - \pi_{\mathcal{L}_{m-1}}\| \rightarrow 0$. En particular, tenemos que $\|\pi_{\hat{\mathcal{L}}_{m-1}} \phi_{R,m} - \pi_{\mathcal{L}_{m-1}} \phi_{R,m}\| \rightarrow 0$, lo que, junto con el hecho que $\pi_{\mathcal{L}_{m-1}} \phi_{R,m} = 0$, implican que $\pi_{\hat{\mathcal{L}}_{m-1}} \phi_{R,m} \rightarrow 0$ y entonces, $\phi_{R,m} - \pi_{\hat{\mathcal{L}}_{m-1}} \phi_{R,m} \rightarrow \phi_{R,m}$. Usando que $\phi_{R,m} = \pi_{\hat{\mathcal{L}}_{m-1}} \phi_{R,m} + (\phi_{R,m} - \pi_{\hat{\mathcal{L}}_{m-1}} \phi_{R,m})$, obtenemos que $\|\phi_{R,m} - \pi_{\hat{\mathcal{L}}_{m-1}} \phi_{R,m}\| \rightarrow \|\phi_{R,m}\| = 1$. Denotemos por $\hat{\alpha}_m = (\phi_{R,m} - \pi_{\hat{\mathcal{L}}_{m-1}} \phi_{R,m}) / \|\phi_{R,m} - \pi_{\hat{\mathcal{L}}_{m-1}} \phi_{R,m}\|$, cabe aclarar que $\hat{\alpha}_m \in \hat{\mathcal{B}}_m$. Entonces, usando

que $\|\phi_{R,m} - \pi_{\widehat{\mathcal{L}}_{m-1}} \phi_{R,m}\| \rightarrow 1$ y que $\|\pi_{\widehat{\mathcal{L}}_{m-1}} \phi_{R,m}\| \rightarrow 0$, obtenemos que $\phi_{R,m} = \widehat{\alpha}_m + o(1)$ que junto con la continuidad de σ , implican que $\sigma^2(\widehat{\alpha}_m) \rightarrow \sigma^2(\phi_{R,m})$. Por lo tanto,

$$\sigma^2(\phi_{R,m}) = \sigma^2(\widehat{\alpha}_m) + o(1) \leq \max_{\alpha \in \widehat{\mathcal{B}}_m} \sigma^2(\alpha) + o(1),$$

donde hemos usado que $\widehat{\alpha}_m$ pertenece a $\widehat{\mathcal{B}}_m$, concluyendo la demostración de (3.17).

Para obtener (3.18), notemos que

$$\begin{aligned} \max_{\alpha \in \widehat{\mathcal{B}}_m} \sigma^2(\alpha) &= \max_{\alpha \in \widehat{\mathcal{B}}_m} (\sigma^2(\alpha) - \sigma_n^2(\alpha) + \sigma_n^2(\alpha)) \leq \max_{\alpha \in \widehat{\mathcal{B}}_m} |\sigma^2(\alpha) - \sigma_n^2(\alpha)| + \sigma_n^2(\widehat{\phi}_m) \\ &\leq \max_{\alpha \in \widehat{\mathcal{B}}_m} (\sigma^2(\alpha) - \sigma_n^2(\alpha)) + \sigma_n^2(\widehat{\phi}_m) - \sigma^2(\widehat{\phi}_m) + \sigma^2(\widehat{\phi}_m) \\ &\leq 2 \max_{\|\alpha\|=1} |\sigma^2(\alpha) - \sigma_n^2(\alpha)| + \sigma^2(\widehat{\phi}_m) = o(1) + \sigma^2(\widehat{\phi}_m). \end{aligned}$$

Usando que $\pi_{\widehat{\mathcal{L}}_{m-1}} \widehat{\phi}_m = 0$ y $\|\pi_{\widehat{\mathcal{L}}_{m-1}} \widehat{\phi}_m - \pi_{\mathcal{L}_{m-1}} \widehat{\phi}_m\| \rightarrow 0$ (pues $\|\widehat{\phi}_m\| = 1$) obtenemos que

$$\widehat{\phi}_m = \widehat{\phi}_m - \pi_{\mathcal{L}_{m-1}} \widehat{\phi}_m + (\pi_{\mathcal{L}_{m-1}} - \pi_{\widehat{\mathcal{L}}_{m-1}}) \widehat{\phi}_m = \widehat{\phi}_m - \pi_{\mathcal{L}_{m-1}} \widehat{\phi}_m + o(1).$$

Denotemos $\widehat{b}_m = \widehat{\phi}_m - \pi_{\mathcal{L}_{m-1}} \widehat{\phi}_m$, luego tendremos que $\widehat{\phi}_m = \widehat{b}_m + o(1)$, implicando que $\|\widehat{b}_m\| \rightarrow 1$. Sea $\widehat{\beta}_m$ que representará a $\widehat{\beta}_m = \widehat{b}_m / \|\widehat{b}_m\|$. Notemos que $\widehat{\beta}_m \in \mathcal{B}_m$, luego $\sigma(\widehat{\beta}_m) \leq \sigma(\phi_{R,m})$. Por otro lado, usando que $\widehat{\phi}_m - \widehat{\beta}_m = o(1)$ y que σ es débilmente continua y que \mathcal{S} es débilmente compacta, obtenemos, como en el Lema 3.4.1, que $\sigma(\widehat{\phi}_m) - \sigma(\widehat{\beta}_m) = o(1)$. Entonces,

$$\max_{\alpha \in \widehat{\mathcal{B}}_m} \sigma^2(\alpha) \leq o(1) + \sigma^2(\widehat{\phi}_m) = o(1) + \sigma^2(\widehat{\beta}_m) \leq o(1) + \sigma^2(\phi_{R,m}),$$

concluyendo la demostración de (3.18) y entonces, $\widehat{\lambda}_m = \max_{\alpha \in \widehat{\mathcal{B}}_m} \sigma_n^2(\alpha) \rightarrow \sigma^2(\phi_{R,m}) = \lambda_{R,m}$.

Probemos que $\sigma^2(\widehat{\phi}_m) \rightarrow \sigma^2(\phi_{R,m})$. Tenemos que

$$\begin{aligned} |\sigma^2(\widehat{\phi}_m) - \sigma^2(\phi_{R,m})| &\leq |\sigma^2(\widehat{\phi}_m) - \sigma_n^2(\widehat{\phi}_m)| + |\sigma_n^2(\widehat{\phi}_m) - \sigma^2(\phi_{R,m})| \\ &\leq |\sigma^2(\widehat{\phi}_m) - \sigma_n^2(\widehat{\phi}_m)| + |\widehat{\lambda}_m - \sigma^2(\phi_{R,m})| \\ &\leq \sup_{\|\alpha\|=1} |\sigma^2(\alpha) - \sigma_n^2(\alpha)| + |\widehat{\lambda}_m - \sigma^2(\phi_{R,m})|. \end{aligned}$$

y la demostración sigue usando ii) y que $\widehat{\lambda}_m \rightarrow \sigma^2(\phi_{R,m})$.

d) Ya hemos probado que cuando $m = 1$ el resultado vale. Procedemos por inducción, supongamos que $\langle \widehat{\phi}_j, \phi_{R,j} \rangle^2 \rightarrow 1$ para $1 \leq j \leq m-1$ y demostraremos que $\langle \widehat{\phi}_m, \phi_{R,m} \rangle^2 \rightarrow 1$. Usando c) tenemos que $\sigma^2(\widehat{\phi}_m) \rightarrow \sigma^2(\phi_{R,m})$ y luego, por el Lema 3.4.1 b) concluimos la prueba. \square

3.4.4 Consistencia de los estimadores “crudos”

El Teorema 3.4.1 establece la consistencia de los estimadores “crudos” de componentes principales. Se deduce de la Proposición 3.4.1 usando que si P_n es la medida empírica inducida por X_1, \dots, X_n , entonces $P_n \xrightarrow{\omega} P$ casi seguramente.

Teorema 3.4.1. *Sea $\widehat{\phi}_m$ y $\widehat{\lambda}_m$ los estimadores definidos en (3.3) y (3.4), respectivamente. Bajo **S0**, **S1** y **S2**, tenemos que*

- a) $\widehat{\lambda}_1 \xrightarrow{a.s.} \sigma^2(\phi_{R,1})$ y $\sigma^2(\widehat{\phi}_1) \xrightarrow{a.s.} \sigma^2(\phi_{R,1})$.
- b) $\langle \widehat{\phi}_1, \phi_{R,1} \rangle^2 \xrightarrow{a.s.} 1$.
- c) *Dados $2 \leq m \leq q$, si $\widehat{\phi}_s \xrightarrow{a.s.} \phi_{R,s}$, para $1 \leq s \leq m-1$, entonces $\widehat{\lambda}_m \xrightarrow{a.s.} \sigma^2(\phi_{R,m})$ y $\sigma^2(\widehat{\phi}_m) \xrightarrow{a.s.} \sigma^2(\phi_{R,m})$.*
- d) *Para $1 \leq m \leq q$, $\langle \widehat{\phi}_m, \phi_{R,m} \rangle^2 \xrightarrow{a.s.} 1$.*

3.4.5 Consistencia de los estimadores suavizados mediante penalización de la norma

Recordemos que \mathcal{H}_S es el subespacio de \mathcal{H} de elementos suaves α tal que $\Psi(\alpha) = [\alpha, \alpha] = \|D\alpha\|^2 < \infty$. Para obtener la consistencia de las propuestas dadas en (3.5) y en (3.6), precisaremos alguno de los siguientes supuestos en **S4**.

- S4.** a) $\phi_{R,j} \in \mathcal{H}_S, \forall j$ o b) $\phi_{R,j} \in \overline{\mathcal{H}_S}, \forall j$.

La condición **S4b**) generaliza el supuesto de suavidad requerido por Silverman (1996) y vale, por ejemplo, cuando \mathcal{H}_S es un conjunto denso de \mathcal{H} .

A fines de simplificar, indiquemos $\mathcal{T}_k = \mathcal{L}_k^\perp$ el subespacio lineal ortogonal a ϕ_1, \dots, ϕ_k y $\pi_k = \pi_{\mathcal{T}_k}$ la proyección ortogonal con respecto al producto interno definido en \mathcal{H} . Por otro lado, sea $\widehat{\pi}_{r,k}$ la proyección sobre el subespacio lineal ortogonal a $\widehat{\phi}_{PN,1}, \dots, \widehat{\phi}_{PN,k}$ en el espacio \mathcal{H}_S respecto del producto interno $\langle \cdot, \cdot \rangle_r$, es decir, para todo $\alpha \in \mathcal{H}_S$, $\widehat{\pi}_{r,k}(\alpha) = \alpha - \sum_{j=1}^k \langle \alpha, \widehat{\phi}_{PN,j} \rangle_r \widehat{\phi}_{PN,j}$. Más aún, sea $\widehat{\mathcal{T}}_{r,k}$ el subespacio ortogonal a $\widehat{\mathcal{L}}_k$ respecto del producto interno $\langle \cdot, \cdot \rangle_r$. Entonces, $\widehat{\pi}_{r,k}$ es la proyección ortogonal sobre $\widehat{\mathcal{T}}_{r,k}$ con respecto a este producto interno.

Teorema 3.4.2. *Sea $\widehat{\phi}_{PN,m}$ y $\widehat{\lambda}_{PN,m}$ los estimadores definidos en (3.5) y (3.7), respectivamente. Más aún, supongamos que valen las condiciones **S0**, **S1**, **S2** y **S4b**). Si $r = r_n \rightarrow 0$, $r_n \geq 0$, se verifica*

- a) $\widehat{\lambda}_{PN,1} \xrightarrow{a.s.} \sigma^2(\phi_{R,1})$ y $\sigma^2(\widehat{\phi}_{PN,1}) \xrightarrow{a.s.} \sigma^2(\phi_{R,1})$
- b) $r[\widehat{\phi}_{PN,1}, \widehat{\phi}_{PN,1}] \xrightarrow{a.s.} 0$ y $\|\widehat{\phi}_{PN,1}\| \xrightarrow{a.s.} 1$.

$$c) \langle \widehat{\phi}_{\text{PN},1}, \phi_{\text{R},1} \rangle^2 \xrightarrow{\text{a.s.}} 1.$$

d) Damos $2 \leq m \leq q$, si $\widehat{\phi}_{\text{PN},\ell} \xrightarrow{\text{a.s.}} \phi_{\text{R},\ell}$ y $r[\widehat{\phi}_{\text{PN},\ell}, \widehat{\phi}_{\text{PN},\ell}] \xrightarrow{\text{a.s.}} 0$, for $1 \leq \ell \leq m-1$, entonces, $\widehat{\lambda}_{\text{PN},m} \xrightarrow{\text{a.s.}} \sigma^2(\phi_{\text{R},m})$, $\sigma^2(\widehat{\phi}_{\text{PN},m}) \xrightarrow{\text{a.s.}} \sigma^2(\phi_{\text{R},m})$, $r[\widehat{\phi}_{\text{PN},m}, \widehat{\phi}_{\text{PN},m}] \xrightarrow{\text{a.s.}} 0$ y, por lo tanto, $\|\widehat{\phi}_{\text{PN},m}\| \xrightarrow{\text{a.s.}} 1$.

$$e) \text{ Para } 1 \leq m \leq q, \langle \widehat{\phi}_{\text{PN},m}, \phi_{\text{R},m} \rangle^2 \xrightarrow{\text{a.s.}} 1.$$

DEMOSTRACIÓN. A fin de evitar sobrecarga en la escritura, indiquemos por $\widehat{\phi}_j = \widehat{\phi}_{\text{PN},j}$ y $\widehat{\lambda}_j = \widehat{\lambda}_{\text{PN},j}$.

a) Vamos a probar que

$$\sigma^2(\phi_{\text{R},1}) \geq \widehat{\lambda}_1 + o_{\text{a.s.}}(1) \quad (3.19)$$

y que bajo **S4a)** se tiene

$$\sigma^2(\phi_{\text{R},1}) \leq \widehat{\lambda}_1 + o_{\text{a.s.}}(1), \quad (3.20)$$

Una desigualdad más débil que (3.20) se puede obtener bajo **S4b)**.

Probemos la primer desigualdad. Usando que σ es un funcional de escala, y que $\|\widehat{\phi}_1\| \leq 1$, obtenemos fácilmente que

$$\sigma^2(\phi_{\text{R},1}) = \sup_{\alpha \in \mathcal{S}} \sigma^2(\alpha) \geq \sigma^2\left(\frac{\widehat{\phi}_1}{\|\widehat{\phi}_1\|}\right) = \frac{\sigma^2(\widehat{\phi}_1)}{\|\widehat{\phi}_1\|^2} \geq \sigma^2(\widehat{\phi}_1).$$

Por otro lado, **S1** implica que $\widehat{a}_{n,1} = s_n^2(\widehat{\phi}_1) - \sigma^2(\widehat{\phi}_1) \xrightarrow{\text{a.s.}} 0$ y entonces, $\sigma^2(\phi_{\text{R},1}) \geq \sigma^2(\widehat{\phi}_1) = s_n^2(\widehat{\phi}_1) + o_{\text{a.s.}}(1) = \widehat{\lambda}_1 + o_{\text{a.s.}}(1)$, concluyendo la demostración de (3.19).

Obtengamos a). Como claramente **S4b)** implica **S4a)**, empezamos probando el resultado primero para **S4a)** así podemos disponer de una idea sobre los argumentos que se utilizarán. La extensión a **S4b)** se puede hacer después haciendo algunas argumentaciones técnicas.

i) Supongamos que vale **S4a)**, entonces $\phi_{\text{R},1} \in \mathcal{H}_{\mathcal{S}}$, por lo tanto $\|\phi_{\text{R},1}\|_r < \infty$. Notemos que $\|\phi_{\text{R},1}\|_r \geq \|\phi_{\text{R},1}\| = 1$, luego, definiendo $\beta_1 = \phi_{\text{R},1}/\|\phi_{\text{R},1}\|_r$, tenemos que $\|\beta_1\|_r = 1$, implicando que $\widehat{\lambda}_1 = s_n^2(\widehat{\phi}_1) \geq s_n^2(\beta_1)$. Nuevamente, usando **S1** obtenemos que $\widehat{b}_{n,1} = s_n^2(\beta_1) - \sigma^2(\beta_1) \xrightarrow{\text{a.s.}} 0$, por ende,

$$\widehat{\lambda}_1 \geq s_n^2(\beta_1) = \sigma^2(\phi_{\text{R},1}/\|\phi_{\text{R},1}\|_r) + o_{\text{a.s.}}(1) = \frac{\sigma^2(\phi_{\text{R},1})}{\|\phi_{\text{R},1}\|_r^2} + o_{\text{a.s.}}(1) = \sigma^2(\phi_{\text{R},1}) + o_{\text{a.s.}}(1)$$

donde, en la última desigualdad, hemos usado que $\|\phi_{\text{R},1}\|_r \rightarrow \|\phi_{\text{R},1}\| = 1$ pues $r \rightarrow 0$, concluyendo la demostración de a) para este caso.

ii) Supongamos que vale **S4b)**. En este caso, no podemos considerar $\|\phi_{\text{R},1}\|_r$ pues $\phi_{\text{R},1}$ no pertenece a $\mathcal{H}_{\mathcal{S}}$, sino se podría argumentar como en i). Pero como $\phi_{\text{R},1} \in \overline{\mathcal{H}_{\mathcal{S}}}$, entonces podemos elegir una

sucesión $\tilde{\phi}_{1,k} \in \mathcal{H}_S$ tal que $\tilde{\phi}_{1,k} \rightarrow \phi_{R,1}$, $\|\tilde{\phi}_{1,k}\| = 1$ y $|\sigma^2(\tilde{\phi}_{1,k}) - \sigma^2(\phi_{R,1})| < 1/k$. Notemos que para cualquier k fijo, $\|\tilde{\phi}_{1,k}\|_r \geq \|\tilde{\phi}_{1,k}\| = 1$ y $\|\tilde{\phi}_{1,k}\|_r \rightarrow \|\tilde{\phi}_{1,k}\| = 1$ pues $r_n \rightarrow 0$.

Entonces, usando que $\hat{\lambda}_1 = \max_{\|\alpha\|_r=1} s_n^2(\alpha)$ y definiendo $\beta_{1,k} = \tilde{\phi}_{1,k}/\|\tilde{\phi}_{1,k}\|_r$, obtenemos que $\|\beta_{1,k}\|_r = 1$ y $\hat{\lambda}_1 = s_n^2(\hat{\phi}_1) \geq s_n^2(\beta_{1,k})$.

Obsrvemos que **S1** implica que $\hat{b}_{n,1} = s_n^2(\beta_{1,k}) - \sigma^2(\beta_{1,k}) \xrightarrow{a.s.} 0$, por ende,

$$\hat{\lambda}_1 \geq s_n^2(\beta_{1,k}) = \sigma^2(\beta_{1,k}) + o_{a.s.}(1) = \frac{\sigma^2(\tilde{\phi}_{1,k})}{\|\tilde{\phi}_{1,k}\|_r^2} + o_{a.s.}(1) = \frac{\sigma^2(\phi_{R,1}) - 1/k}{\|\tilde{\phi}_{1,k}\|_r} + o_{a.s.}(1).$$

Entonces, usando (3.19) y que $\|\tilde{\phi}_{1,k}\|_r \geq 1$, se deduce que

$$\sigma^2(\phi_{R,1}) \geq \hat{\lambda}_1 + O_{1,n} \geq \frac{\sigma^2(\phi_{R,1}) - 1/k}{\|\tilde{\phi}_{1,k}\|_r} + o_{a.s.}(1) \geq \sigma^2(\phi_{R,1}) - \left(1 - \frac{1}{\|\tilde{\phi}_{1,k}\|_r}\right) \sigma^2(\phi_{R,1}) - \frac{1}{k} + O_{2,n},$$

donde $O_{i,n} = o_{a.s.}(1)$, $i = 1, 2$. Sea $\mathcal{N} = \cup_{i=1,2} \{\omega : O_{i,n}(\omega) \not\rightarrow 0\}$ y fijemos $\omega \notin \mathcal{N}$. Dado $\epsilon > 0$, fijamos k_0 tal que $1/k_0 < \epsilon$. Sea n_0 tal que, para $n \geq n_0$, $O_{i,n}(\omega) < \epsilon$, $i = 1, 2$ y

$$0 \leq \left(1 - \frac{1}{\|\tilde{\phi}_{1,k_0}\|_r}\right) \sigma^2(\phi_{R,1}) < \epsilon,$$

donde hemos usado que $r_n \rightarrow 0$ y entonces $\|\tilde{\phi}_{1,k_0}\|_r \rightarrow 1$. Luego, usando

$$|\hat{\lambda}_1(\omega) - \sigma^2(\phi_{R,1})| \leq \max\{|O_{1,n}|, |O_{2,n}| + 2\epsilon\} \leq 3\epsilon$$

lo que implica que $\hat{\lambda}_1 \xrightarrow{a.s.} \sigma^2(\phi_{R,1})$, como se deseaba.

Usando **S1**, obtenemos que $\hat{\lambda}_1 - \sigma^2(\hat{\phi}_1) = s_n^2(\hat{\phi}_1) - \sigma^2(\hat{\phi}_1) \xrightarrow{a.s.} 0$, usando que $\hat{\lambda}_1 \xrightarrow{a.s.} \sigma^2(\phi_{R,1})$, obtenemos que $\sigma^2(\hat{\phi}_1) \xrightarrow{a.s.} \sigma^2(\phi_{R,1})$, concluyendo la demostración de a).

Vale la pena notar que como consecuencia de los resultados anteriores, obtenemos que las siguientes desigualdades convergen a igualdades

$$\sigma^2(\phi_{R,1}) \geq \sigma^2\left(\frac{\hat{\phi}_1}{\|\hat{\phi}_1\|}\right) \geq \sigma^2(\hat{\phi}_1) = \hat{\lambda}_1 + o_{a.s.}(1),$$

en particular,

$$\sigma^2\left(\frac{\hat{\phi}_1}{\|\hat{\phi}_1\|}\right) \xrightarrow{a.s.} \sigma^2(\phi_{R,1}) \quad \text{y} \quad \sigma^2(\hat{\phi}_1) \xrightarrow{a.s.} \sigma^2(\phi_{R,1}). \quad (3.21)$$

b) Notemos que

$$r[\hat{\phi}_1, \hat{\phi}_1] = 1 - \|\hat{\phi}_1\|^2 = 1 - \frac{\sigma^2(\hat{\phi}_1)}{\sigma^2(\hat{\phi}_1/\|\hat{\phi}_1\|)}.$$

Luego, usando (3.21) tendremos que el segundo término es $1 + o_{\text{a.s.}}(1)$, concluyendo la demostración b).

c) Dado que $\|\widehat{\phi}_1\|_r = 1$, tenemos que $\|\widehat{\phi}_1\| \leq 1$. Más aún, por b) $\|\widehat{\phi}_1\| \xrightarrow{\text{a.s.}} 1$. Sea $\widetilde{\phi}_1 = \widehat{\phi}_1 / \|\widehat{\phi}_1\|$. Luego, tenemos que $\widetilde{\phi}_1 \in \mathcal{S}$ y $\sigma(\widetilde{\phi}_1) = \sigma(\widehat{\phi}_1) / \|\widehat{\phi}_1\|$. Usando que $\sigma^2(\widehat{\phi}_1) \xrightarrow{\text{a.s.}} \sigma^2(\phi_{R,1})$ y $\|\widehat{\phi}_1\| \xrightarrow{\text{a.s.}} 1$, obtenemos que $\sigma^2(\widetilde{\phi}_1) \xrightarrow{\text{a.s.}} \sigma^2(\phi_{R,1})$ y entonces, la demostración sigue del Lema 3.4.1.

d) Probemos que $\widehat{\lambda}_m \xrightarrow{\text{a.s.}} \sigma^2(\phi_{R,m})$. Empezamos probando la siguiente extensión de **S1**

$$\sup_{\|\alpha\|_r \leq 1} |\sigma^2(\pi_{m-1}\alpha) - s_n^2(\widehat{\pi}_{r,m-1}\alpha)| \xrightarrow{\text{a.s.}} 0. \quad (3.22)$$

Usando **S1** y que, por ser s_n un estimador de escala entonces $s_n(\alpha) = \|\alpha\|_r s_n(\alpha/\|\alpha\|_r)$, obtenemos que

$$\sup_{\|\alpha\|_r \leq 1} |s_n^2(\alpha) - \sigma^2(\alpha)| \xrightarrow{\text{a.s.}} 0. \quad (3.23)$$

Observemos que

$$\sup_{\|\alpha\|_r \leq 1} |\sigma^2(\pi_{m-1}\alpha) - s_n^2(\widehat{\pi}_{r,m-1}\alpha)| \leq \sup_{\|\alpha\|_r \leq 1} |\sigma^2(\pi_{m-1}\alpha) - \sigma^2(\widehat{\pi}_{r,m-1}\alpha)| + \sup_{\|\alpha\|_r \leq 1} |\sigma^2(\widehat{\pi}_{r,m-1}\alpha) - s_n^2(\widehat{\pi}_{r,m-1}\alpha)|.$$

Usando (3.23) y que $\|\alpha\|_r \leq 1$ implica $\|\widehat{\pi}_{r,m-1}\alpha\|_r \leq 1$, obtenemos que el segundo término del lado derecho converge a 0 casi seguramente.

Para concluir la demostración de (3.22), basta ver que

$$\sup_{\|\alpha\|_r \leq 1} |\sigma^2(\pi_{m-1}\alpha) - \sigma^2(\widehat{\pi}_{r,m-1}\alpha)| \xrightarrow{\text{a.s.}} 0. \quad (3.24)$$

Como en Silverman (1996), usando que $\widehat{\phi}_j \xrightarrow{\text{a.s.}} \phi_{R,j}$ y que $r\Psi(\widehat{\phi}_j) = r[\widehat{\phi}_j, \widehat{\phi}_j] \xrightarrow{\text{a.s.}} 0$, para $1 \leq j \leq m-1$, obtenemos que

$$\sup_{\|\alpha\|_r \leq 1} \|\langle \alpha, \phi_{R,j} \rangle \phi_{R,j} - \langle \alpha, \widehat{\phi}_j \rangle_r \widehat{\phi}_j\| \xrightarrow{\text{a.s.}} 0 \quad \text{for } 1 \leq j \leq m-1 \quad (3.25)$$

Efectivamente, para todo $\alpha \in \mathcal{H}_S$ tal que $\|\alpha\|_r^2 = \|\alpha\|^2 + r\Psi(\alpha) \leq 1$, tendremos que

$$\begin{aligned} \|\langle \alpha, \phi_{R,j} \rangle \phi_{R,j} - \langle \alpha, \widehat{\phi}_j \rangle_r \widehat{\phi}_j\| &\leq \|\alpha\| \|\phi_{R,j} - \widehat{\phi}_j\| + \|\widehat{\phi}_j\| \left| \langle \alpha, \phi_{R,j} \rangle - \langle \alpha, \widehat{\phi}_j \rangle_r \right| \\ &\leq \|\phi_{R,j} - \widehat{\phi}_j\| + \left| \langle \alpha, \phi_{R,j} - \widehat{\phi}_j \rangle + r[\alpha, \widehat{\phi}_j] \right| \\ &\leq \|\phi_{R,j} - \widehat{\phi}_j\| + \left\{ \|\phi_{R,j} - \widehat{\phi}_j\| + (r\Psi(\alpha))^{\frac{1}{2}} \left(r\Psi(\widehat{\phi}_j) \right)^{\frac{1}{2}} \right\} \\ &\leq \|\phi_{R,j} - \widehat{\phi}_j\| + \left\{ \|\phi_{R,j} - \widehat{\phi}_j\| + \left(r\Psi(\widehat{\phi}_j) \right)^{\frac{1}{2}} \right\} \end{aligned}$$

y entonces, (3.25) vale implicando que $\sup_{\|\alpha\|_r \leq 1} \|\widehat{\pi}_{r,m-1}\alpha - \pi_{m-1}\alpha\| \xrightarrow{a.s.} 0$. Entonces, usando que σ es débilmente continua y que la bola unitaria es débilmente compacta, fácilmente se obtiene que (3.24) vale, concluyendo la demostración de (3.22).

Como en a), probaremos que

$$\sigma^2(\phi_{R,m}) \geq \widehat{\lambda}_m + o_{a.s.}(1) \quad (3.26)$$

y que cuando **S4a**) entonces

$$\sigma^2(\phi_{R,m}) \leq \widehat{\lambda}_m + o_{a.s.}(1). \quad (3.27)$$

vale. Una desigualdad más débil que (3.27) se obtendrá bajo **S4b**).

Usando nuevamente que σ es un funcional de escala, fácilmente se obtiene que $\sup_{\alpha \in \mathcal{S} \cap \mathcal{T}_{m-1}} \sigma^2(\alpha) = \sup_{\alpha \in \mathcal{S}} \sigma^2(\pi_{m-1}\alpha)$ y, por lo tanto,

$$\sigma^2(\phi_{R,m}) = \sup_{\alpha \in \mathcal{S} \cap \mathcal{T}_{m-1}} \sigma^2(\alpha) = \sup_{\alpha \in \mathcal{S}} \sigma^2(\pi_{m-1}\alpha) \geq \sigma^2\left(\pi_{m-1} \frac{\widehat{\phi}_m}{\|\widehat{\phi}_m\|}\right).$$

De (3.22) obtenemos que $\widehat{b}_m = \sigma^2(\pi_{m-1}\widehat{\phi}_m) - s_n^2(\widehat{\pi}_{r,m-1}\widehat{\phi}_m) \xrightarrow{a.s.} 0$ y luego, como $\widehat{\pi}_{r,m-1}\widehat{\phi}_m = \widehat{\phi}_m$ y $\|\widehat{\phi}_m\| \leq 1$, obtenemos que

$$\begin{aligned} \sigma^2(\phi_{R,m}) &\geq \sigma^2\left(\pi_{m-1} \frac{\widehat{\phi}_m}{\|\widehat{\phi}_m\|}\right) = \frac{\sigma^2(\pi_{m-1}\widehat{\phi}_m)}{\|\widehat{\phi}_m\|^2} \\ &\geq \sigma^2(\pi_{m-1}\widehat{\phi}_m) = s_n^2(\widehat{\pi}_{r,m-1}\widehat{\phi}_m) + o_{a.s.}(1) = s_n^2(\widehat{\phi}_m) + o_{a.s.}(1) = \widehat{\lambda}_m + o_{a.s.}(1), \end{aligned}$$

concluyendo la demostración de (3.26).

Probemos que (3.27) vale si **S4a**) se cumple.

i) Si **S4a**) vale, $\phi_{R,m} \in \mathcal{H}_S$, con lo cual $\|\phi_{R,m}\|_r < \infty$ y $\|\phi_{R,m}\|_r \rightarrow \|\phi_{R,m}\| = 1$. Usando que s_n es un estimador de escala y el hecho que para todo $\alpha \in \mathcal{H}_S$ tal que $\|\alpha\|_r = 1$ tenemos $\|\widehat{\pi}_{r,m-1}\alpha\|_r \leq 1$, obtenemos sin dificultad que

$$\widehat{\lambda}_m = s_n^2(\widehat{\phi}_m) = \sup_{\|\alpha\|_r=1, \alpha \in \widehat{\mathcal{T}}_{r,m-1}} s_n^2(\alpha) = \sup_{\|\alpha\|_r=1} s_n^2(\widehat{\pi}_{r,m-1}\alpha) \geq s_n^2\left(\frac{\widehat{\pi}_{r,m-1}\phi_{R,m}}{\|\phi_{R,m}\|_r}\right)$$

que junto con (3.22) y con $\|\phi_{R,m}\|_r \rightarrow \|\phi_{R,m}\| = 1$, se obtiene que

$$\begin{aligned} \widehat{\lambda}_m &\geq s_n^2\left(\frac{\widehat{\pi}_{r,m-1}\phi_{R,m}}{\|\phi_{R,m}\|_r}\right) = \sigma^2\left(\frac{\pi_{m-1}\phi_{R,m}}{\|\phi_{R,m}\|_r}\right) + o_{a.s.}(1) \\ &\geq \sigma^2\left(\frac{\phi_{R,m}}{\|\phi_{R,m}\|_r}\right) + o_{a.s.}(1) = \frac{\sigma^2(\phi_{R,m})}{\|\phi_{R,m}\|_r^2} + o_{a.s.}(1) \geq \sigma^2(\phi_{R,m}) + o_{a.s.}(1) \end{aligned}$$

concluyendo la demostración de (3.27) en este caso y entonces, cuando **S4a**) vale $\widehat{\lambda}_m \xrightarrow{a.s.} \sigma^2(\phi_{R,m})$.

ii) Supongamos que vale **S4b**). Como en a), consideremos una sucesión $\tilde{\phi}_{m,k} \in \mathcal{H}_S$ tal que $\tilde{\phi}_{m,k} \rightarrow \phi_{R,m}$, con $k \rightarrow \infty$, $\|\tilde{\phi}_{m,k}\| = 1$ y $|\sigma^2(\pi_{m-1}\tilde{\phi}_{m,k}) - \sigma^2(\phi_{R,m})| < 1/k$, pues $\pi_{m-1}\phi_{R,m} = \phi_{R,m}$. Luego, para cada k fijo, tenemos que $\|\tilde{\phi}_{m,k}\|_r \rightarrow \|\tilde{\phi}_{m,k}\| = 1$ ya que $r \rightarrow 0$.

Usando que s_n es un estimador de escala y el hecho que, para todo $\alpha \in \mathcal{H}_S$ tal que $\|\alpha\|_r = 1$, se tiene $\|\hat{\pi}_{r,m-1}\alpha\|_r \leq 1$, obtenemos que

$$\hat{\lambda}_m = s_n^2(\hat{\phi}_m) = \sup_{\|\alpha\|_r=1, \alpha \in \hat{\mathcal{T}}_{r,m-1}} s_n^2(\alpha) = \sup_{\|\alpha\|_r=1} s_n^2(\hat{\pi}_{r,m-1}\alpha) \geq s_n^2\left(\frac{\hat{\pi}_{r,m-1}\tilde{\phi}_{m,k}}{\|\tilde{\phi}_{m,k}\|_r}\right)$$

Usando (3.22), que $|\sigma^2(\pi_{m-1}\tilde{\phi}_{m,k}) - \sigma^2(\phi_{R,m})| < 1/k$ y que $\|\tilde{\phi}_{m,k}\|_r \geq 1$, obtenemos

$$\begin{aligned} \hat{\lambda}_m &\geq s_n^2\left(\frac{\hat{\pi}_{r,m-1}\tilde{\phi}_{m,k}}{\|\tilde{\phi}_{m,k}\|_r}\right) = \sigma^2\left(\frac{\pi_{m-1}\tilde{\phi}_{m,k}}{\|\tilde{\phi}_{m,k}\|_r}\right) + o_{a.s.}(1) \\ &\geq \frac{\sigma^2(\phi_{R,m}) - 1/k}{\|\tilde{\phi}_{m,k}\|_r^2} + o_{a.s.}(1) \\ &\geq \sigma^2(\phi_{R,m}) - \sigma^2(\phi_{R,m})\left(1 - \frac{1}{\|\tilde{\phi}_{m,k}\|_r^2}\right) - \frac{1}{k} + o_{a.s.}(1) \end{aligned}$$

Por ende, argumentando como en a) obtenemos que $\hat{\lambda}_m \xrightarrow{a.s.} \sigma^2(\phi_{R,m})$.

Por otro lado, como en a), usando **S1**, obtenemos que $\hat{\lambda}_m - \sigma^2(\hat{\phi}_m) = s_n^2(\hat{\phi}_m) - \sigma^2(\hat{\phi}_m) \xrightarrow{a.s.} 0$, usando que $\hat{\lambda}_m \xrightarrow{a.s.} \sigma^2(\phi_{R,m})$, se concluye luego que $\sigma^2(\hat{\phi}_m) \xrightarrow{a.s.} \sigma^2(\phi_{R,m})$.

Por lo tanto, queda tan solo probar que $r[\hat{\phi}_m, \hat{\phi}_m] \xrightarrow{a.s.} 0$. Como en a), tenemos que las siguientes desigualdades convergen a igualdades:

$$\sigma^2(\phi_{R,m}) \geq \sigma^2\left(\pi_{m-1}\frac{\hat{\phi}_m}{\|\hat{\phi}_m\|}\right) \geq \sigma^2(\pi_{m-1}\hat{\phi}_m) \geq \hat{\lambda}_m + o_{a.s.}(1). \quad (3.28)$$

Notemos que puesto que σ es un estimador de escala, entonces tenemos que

$$r[\hat{\phi}_m, \hat{\phi}_m] = 1 - \|\hat{\phi}_m\|^2 = 1 - \frac{\sigma^2(\pi_{m-1}\hat{\phi}_m)}{\sigma^2(\pi_{m-1}\hat{\phi}_m/\|\hat{\phi}_m\|)},$$

que junto con (3.28) implica que el segundo término en el lado derecho es $1 + o_{a.s.}(1)$, concluyendo d).

e) Para $m = 1$ el resultado se obtuvo en c). Supongamos que para $1 \leq j \leq m-1$, $\hat{\phi}_j \xrightarrow{a.s.} \phi_{R,j}$ y que $r[\hat{\phi}_j, \hat{\phi}_j] \xrightarrow{a.s.} 0$, probaremos que $\langle \hat{\phi}_m, \phi_{R,m} \rangle^2 \xrightarrow{a.s.} 1$, es decir, haremos uso de un argumento inductivo. Por d) ya sabemos que $r[\hat{\phi}_m, \hat{\phi}_m] \xrightarrow{a.s.} 0$ implicando que $\|\hat{\phi}_m\| \xrightarrow{a.s.} 1$. Indiquemos por $\tilde{\phi}_j = \hat{\phi}_j/\|\hat{\phi}_j\|$, basta ver que $\langle \phi_{R,m}, \tilde{\phi}_m \rangle^2 \xrightarrow{a.s.} 1$. Tenemos que, $\langle \tilde{\phi}_m, \tilde{\phi}_j \rangle_r = 0$. Usando que

$r[\widehat{\phi}_j, \widehat{\phi}_j] \xrightarrow{a.s.} 0$, for $1 \leq j \leq m-1$, tenemos que $r[\widehat{\phi}_j, \widehat{\phi}_m] \xrightarrow{a.s.} 0$ para $1 \leq j \leq m-1$ y entonces, $\langle \widetilde{\phi}_m, \widetilde{\phi}_j \rangle \xrightarrow{a.s.} 0$. Luego, argumentando como en el Lema 3.4.1, podemos escribir $\widetilde{\phi}_m$ como $\widetilde{\phi}_m = \sum_{j=1}^{m-1} \widehat{a}_j \phi_{R,j} + \widehat{\gamma}_m$, con $\langle \widehat{\gamma}_m, \phi_{R,j} \rangle = 0$, $1 \leq j \leq m-1$. Para obtener e) queda demostrar que $\langle \widehat{\gamma}_m, \phi_{R,m} \rangle^2 \xrightarrow{a.s.} 1$. Notemos que $\langle \widetilde{\phi}_m, \widetilde{\phi}_j \rangle \xrightarrow{a.s.} 0$, para $j \neq m$, implicando que $\widehat{a}_j = \langle \widetilde{\phi}_m, \phi_{R,j} \rangle = \langle \widetilde{\phi}_m, \phi_{R,j} - \widetilde{\phi}_j \rangle + \langle \widetilde{\phi}_m, \widetilde{\phi}_j \rangle = \langle \widetilde{\phi}_m, \phi_{R,j} - \widetilde{\phi}_j \rangle + o_{a.s.}(1)$. Luego, usando que $\widetilde{\phi}_j \xrightarrow{a.s.} \phi_{R,j}$, $1 \leq j \leq m-1$, y $\|\widehat{\phi}_m\| \xrightarrow{a.s.} 1$, obtenemos que $\widehat{a}_j \xrightarrow{a.s.} 0$ para $1 \leq j \leq m-1$ y entonces, $\|\widetilde{\phi}_m - \widehat{\gamma}_m\| \xrightarrow{a.s.} 0$. Notemos que $1 = \|\widetilde{\phi}_m\|^2 = \sum_{j=1}^{m-1} \widehat{a}_j^2 + \|\widehat{\gamma}_m\|^2$, luego, $\|\widehat{\gamma}_m\|^2 \xrightarrow{a.s.} 1$ implicando que $\|\widetilde{\phi}_m - \widetilde{\gamma}_m\| \xrightarrow{a.s.} 0$, donde $\widetilde{\gamma}_m = \widehat{\gamma}_m / \|\widehat{\gamma}_m\|$. Usando que $\sigma(\alpha)$ es débilmente continua y que \mathcal{S} es débilmente compacta, obtenemos que $\sigma(\widetilde{\gamma}_m) - \sigma(\widetilde{\phi}_m) \xrightarrow{a.s.} 0$ que junto con el hecho que $\sigma^2(\widetilde{\phi}_m) \xrightarrow{a.s.} \sigma^2(\phi_{R,m})$ y $\|\widetilde{\phi}_m\| \xrightarrow{a.s.} 1$, implican que $\sigma(\widetilde{\gamma}_m) \xrightarrow{a.s.} \sigma(\phi_{R,m})$. La demostración ahora sigue como en 3.4.1 usando que $\widetilde{\gamma}_m \in \mathcal{C}_m$, con $\mathcal{C}_m = \{\alpha \in \mathcal{S} : \langle \alpha, \phi_{R,j} \rangle = 0, 1 \leq j \leq m-1\}$ y que $\phi_{R,m}$ es el único maximizador de $\sigma(\alpha)$ sobre \mathcal{C}_m . \square

3.4.6 Consistencia de los estimadores suavizados mediante penalización en la escala

El Teorema siguiente muestra que los estimadores obtenidos penalizando la escala son consistentes.

Teorema 3.4.3. Sean $\widehat{\phi}_{PS,m}$ y $\widehat{\lambda}_{PS,m}$ los estimadores definidos en (3.6) y (3.6), respectivamente. Más aún, supongamos que valen las condiciones **S0**, **S1**, **S2** y **S4a**). Si $\rho = \rho_n \rightarrow 0$, $\rho_n \geq 0$, entonces se verifica

- a) $\widehat{\lambda}_{PS,1} \xrightarrow{a.s.} \sigma^2(\phi_{R,1})$ y $\sigma^2(\widehat{\phi}_{PS,1}) \xrightarrow{a.s.} \sigma^2(\phi_{R,1})$. Más aún, $\rho[\widehat{\phi}_{PN,1}, \widehat{\phi}_{PN,1}] \xrightarrow{a.s.} 0$.
- b) $\langle \widehat{\phi}_{PS,1}, \phi_{R,1} \rangle^2 \xrightarrow{a.s.} 1$.
- c) Dado $2 \leq m \leq q$, si $\widehat{\phi}_{PS,\ell} \xrightarrow{a.s.} \phi_{R,\ell}$, y $\rho[\widehat{\phi}_{PN,\ell}, \widehat{\phi}_{PN,\ell}] \xrightarrow{a.s.} 0$, para $1 \leq \ell \leq m-1$, luego $\widehat{\lambda}_{PS,m} \xrightarrow{a.s.} \sigma^2(\phi_{R,m})$, $\sigma^2(\widehat{\phi}_{PS,m}) \xrightarrow{a.s.} \sigma^2(\phi_{R,m})$ y $\rho[\widehat{\phi}_{PN,m}, \widehat{\phi}_{PN,m}] \xrightarrow{a.s.} 0$.
- d) Para $1 \leq m \leq q$, $\langle \widehat{\phi}_{PS,m}, \phi_{R,m} \rangle^2 \xrightarrow{a.s.} 1$.

DEMOSTRACIÓN. Para evitar sobrecargar la notación, como antes escribimos $\widehat{\phi}_j = \widehat{\phi}_{PS,j}$ y $\widehat{\lambda}_j = \widehat{\lambda}_{PS,j}$.

a) Probaremos que

$$\sigma^2(\phi_{R,1}) \geq \widehat{\lambda}_1 + o_{a.s.}(1) \tag{3.29}$$

$$\sigma^2(\phi_{R,1}) \leq \widehat{\lambda}_1 + o_{a.s.}(1), \tag{3.30}$$

valen.

Probemos la primer desigualdad. Fácilmente obtenemos que

$$\sigma^2(\phi_{R,1}) = \sup_{\alpha \in \mathcal{S}} \sigma^2(\alpha) \geq \sigma^2(\widehat{\phi}_1).$$

Por otro lado, **S1** implica que $\hat{a}_{n,1} = s_n^2(\hat{\phi}_1) - \sigma^2(\hat{\phi}_1) \xrightarrow{a.s.} 0$ y entonces, $\sigma^2(\phi_{R,1}) \geq \sigma^2(\hat{\phi}_1) = s_n^2(\hat{\phi}_1) + o_{a.s.}(1) = \hat{\lambda}_1 + o_{a.s.}(1)$, concluyendo la demostración de (3.29).

Vamos a probar (3.30). Puesto que **S4a**) vale, tenemos que $\phi_{R,1} \in \mathcal{H}_S$, luego $\|\phi_{R,1}\|_\rho < \infty$. Notemos que

$$\hat{\lambda}_1 = s_n^2(\hat{\phi}_1) \geq s_n^2(\hat{\phi}_1) - \rho[\hat{\phi}_1, \hat{\phi}_1] = \sup_{\alpha \in \mathcal{S}} \{s_n^2(\alpha) - \rho[\alpha, \alpha]\} \geq s_n^2(\phi_{R,1}) - \rho[\phi_{R,1}, \phi_{R,1}]. \quad (3.31)$$

Usando que $\rho \rightarrow 0$, obtenemos que $\rho[\phi_{R,1}, \phi_{R,1}] \rightarrow 0$. Además, usando **S1** nos sale que $s_n^2(\phi_{R,1}) = \sigma^2(\phi_{R,1}) + o_{a.s.}(1)$. Por lo tanto, (3.31) se puede escribir como

$$\hat{\lambda}_1 \geq s_n^2(\phi_{R,1}) - \rho[\phi_{R,1}, \phi_{R,1}] = \sigma^2(\phi_{R,1}) + o_{a.s.}(1)$$

Por ende, se sigue (3.30). Junto con (3.29) esto implica que $\hat{\lambda}_1 \xrightarrow{a.s.} \sigma^2(\phi_{R,1})$. Usando **S1** tenemos que

$$\hat{\lambda}_1 - \sigma^2(\hat{\phi}_1) = s_n^2(\hat{\phi}_1) - \sigma^2(\hat{\phi}_1) \xrightarrow{a.s.} 0$$

por lo tanto, también obtenemos que $\sigma^2(\hat{\phi}_1) \xrightarrow{a.s.} \sigma^2(\phi_{R,1})$. Como $\hat{\lambda}_1 \xrightarrow{a.s.} \sigma^2(\phi_{R,1})$, $s_n^2(\phi_{R,1}) \xrightarrow{a.s.} \sigma^2(\phi_{R,1})$, $\rho \rightarrow 0$ y

$$\hat{\lambda}_1 \geq s_n^2(\hat{\phi}_1) - \rho[\hat{\phi}_1, \hat{\phi}_1] \geq s_n^2(\phi_{R,1}) - \rho[\phi_{R,1}, \phi_{R,1}]$$

obtenemos entonces que $\rho[\hat{\phi}_1, \hat{\phi}_1] \xrightarrow{a.s.} 0$, concluyendo la demostración de a).

b) Se deduce fácilmente del Lema 3.4.1 a), usando que $\sigma^2(\hat{\phi}_1) \xrightarrow{a.s.} \sigma^2(\phi_{R,1})$ y $\|\hat{\phi}_1\| = 1$.

c) Probaremos que

$$\sigma^2(\phi_{R,m}) \geq \hat{\lambda}_m + o_{a.s.}(1) \quad (3.32)$$

y que bajo **S4a**) vale

$$\sigma^2(\phi_{R,m}) \leq \hat{\lambda}_m + o_{a.s.}(1), \quad (3.33)$$

Para poder deducir (3.32), notemos que

$$\sigma^2(\phi_{R,1}) = \sup_{\alpha \in \mathcal{S} \cap \mathcal{T}_{m-1}} \sigma^2(\alpha) = \sup_{\alpha \in \mathcal{S}} \sigma^2(\pi_{m-1}\alpha) \geq \sigma^2(\pi_{m-1}\hat{\phi}_m). \quad (3.34)$$

Probemos que $\sigma^2(\pi_{m-1}\hat{\phi}_m) = s_n^2(\hat{\pi}_{m-1}\hat{\phi}_m) + o_{a.s.}(1)$. En efecto, si $\alpha \in \mathcal{S}$, luego

$$|\sigma^2(\pi_{m-1}\alpha) - s_n^2(\hat{\pi}_{m-1}\alpha)| \leq |\sigma^2(\pi_{m-1}\alpha) - \sigma^2(\hat{\pi}_{m-1}\alpha)| + |\sigma^2(\hat{\pi}_{m-1}\alpha) - s_n^2(\hat{\pi}_{m-1}\alpha)|. \quad (3.35)$$

El segundo término en el lado derecho de(3.35) será $o_{a.s.}(1)$ pues vale **S1**. Probemos que el primero también será $o_{a.s.}(1)$.

Usando que $\|\widehat{\pi}_{m-1} - \pi_{m-1}\| \xrightarrow{a.s.} 0$, obtenemos que $\widehat{\pi}_{m-1}\alpha \xrightarrow{a.s.} \pi_{m-1}\alpha$. Finalmente, de **S2**, es decir, de la continuidad de σ , obtenemos que el primer término del lado derecho de (3.35) será $o_{a.s.}(1)$. Luego, $\sigma^2(\pi_{m-1}\widehat{\phi}_m) = s_n^2(\widehat{\pi}_{m-1}\widehat{\phi}_m) + o_{a.s.}(1)$. Entonces, (3.34) implica que

$$\sigma^2(\phi_{R,1}) \geq \sigma^2(\pi_{m-1}\widehat{\phi}_m) = s_n^2(\widehat{\pi}_{m-1}\widehat{\phi}_m) + o_{a.s.}(1) = s_n^2(\widehat{\phi}_m) + o_{a.s.}(1) = \widehat{\lambda}_m + o_{a.s.}(1)$$

concluyendo la demostración de (3.32).

Vamos a probar que, bajo **S4a**, vale (3.33).

$$\begin{aligned} \widehat{\lambda}_m = s_n^2(\widehat{\phi}_m) &\geq s_n^2(\widehat{\phi}_m) - \rho[\widehat{\phi}_m, \widehat{\phi}_m] = \sup_{\alpha \in \mathcal{S} \cap \widehat{\mathcal{T}}_{m-1}} \{s_n^2(\alpha) - \rho[\alpha, \alpha]\} \\ &\geq \sup_{\alpha \in \mathcal{S}} \{s_n^2(\widehat{\pi}_{m-1}\alpha) - \rho[\widehat{\pi}_{m-1}\alpha, \widehat{\pi}_{m-1}\alpha]\} \end{aligned} \quad (3.36)$$

$$\geq s_n^2(\widehat{\pi}_{m-1}\phi_{R,m}) - \rho[\widehat{\pi}_{m-1}\phi_{R,m}, \widehat{\pi}_{m-1}\phi_{R,m}]. \quad (3.37)$$

Probemos que $\sup_{\alpha \in \mathcal{S}} |s_n^2(\widehat{\pi}_{m-1}\alpha) - s_n^2(\pi_{m-1}\alpha)| \xrightarrow{a.s.} 0$. En efecto,

$$\begin{aligned} &\sup_{\alpha \in \mathcal{S}} |s_n^2(\widehat{\pi}_{m-1}\alpha) - s_n^2(\pi_{m-1}\alpha)| \\ &\leq \sup_{\alpha \in \mathcal{S}} |s_n^2(\widehat{\pi}_{m-1}\alpha) - \sigma^2(\widehat{\pi}_{m-1}\alpha)| + \sup_{\alpha \in \mathcal{S}} |\sigma^2(\widehat{\pi}_{m-1}\alpha) - \sigma^2(\pi_{m-1}\alpha)| + \sup_{\alpha \in \mathcal{S}} |\sigma^2(\pi_{m-1}\alpha) - s_n^2(\pi_{m-1}\alpha)| \\ &\leq \sup_{\alpha \in \mathcal{S}} |s_n^2(\alpha) - \sigma^2(\alpha)| + \sup_{\alpha \in \mathcal{S}} |\sigma^2(\widehat{\pi}_{m-1}\alpha) - \sigma^2(\pi_{m-1}\alpha)| + \sup_{\alpha \in \mathcal{S}} |\sigma^2(\pi_{m-1}\alpha) - s_n^2(\pi_{m-1}\alpha)| \end{aligned}$$

El primer y el tercer término de la última desigualdad convergen a cero casi seguramente puesto que vale **S1**.

Luego solamente debemos probar que $\sup_{\alpha \in \mathcal{S}} |\sigma^2(\widehat{\pi}_{m-1}\alpha) - \sigma^2(\pi_{m-1}\alpha)| = o_{a.s.}(1)$. Usando que $\widehat{\phi}_j \xrightarrow{a.s.} \phi_{R,j}$ for $1 \leq j \leq m-1$, es fácil probar que $\|\widehat{\pi}_{m-1} - \pi_{m-1}\| \xrightarrow{a.s.} 0$ puesto que se reduce a la diferencia de proyecciones finito-dimensionales. Por lo tanto, tenemos convergencia uniforme en el conjunto $\{\alpha, \|\alpha\| \leq 1\}$. Usando que σ es débilmente continua en $\mathcal{S} = \{\alpha, \|\alpha\| \leq 1\}$, que es débilmente compacta, obtenemos que $\sup_{\alpha \in \mathcal{S}} |\sigma^2(\widehat{\pi}_{m-1}\alpha) - \sigma^2(\pi_{m-1}\alpha)|$ converge a 0 casi seguramente. En conclusión,

$$\sup_{\alpha \in \mathcal{S}} |s_n^2(\widehat{\pi}_{m-1}\alpha) - s_n^2(\pi_{m-1}\alpha)| = o_{a.s.}(1).$$

Usando que $\rho[\widehat{\phi}_\ell, \widehat{\phi}_\ell] \xrightarrow{a.s.} 0$, $1 \leq \ell \leq m-1$, argumentos análogos a los empleados en Pezzulli y Silverman (1993) y el hecho que $\rho \rightarrow 0$ implica que $\rho[\phi_{R,m}, \phi_{R,m}] = o(1)$, no es difícil ver que

$$\rho_n[\widehat{\pi}_{m-1}\phi_{R,m}, \widehat{\pi}_{m-1}\phi_{R,m}] \xrightarrow{a.s.} 0.$$

Estos dos resultados nos permiten reemplazar $\widehat{\pi}_{m-1}\alpha$ por $\pi_{m-1}\alpha$ en (3.37). Luego,

$$\begin{aligned} \widehat{\lambda}_m &\geq s_n^2(\pi_{m-1}\phi_{R,m}) + o_{a.s.}(1) = \sup_{\alpha \in \mathcal{S} \cap \widehat{\mathcal{T}}_{m-1}} \{s_n^2(\alpha) - \rho[\alpha, \alpha]\} + o_{a.s.}(1) \\ &\geq s_n^2(\phi_{R,m}) + o_{a.s.}(1) = \sigma^2(\phi_{R,m}) + o_{a.s.}(1) + o_{a.s.}(1) = \sigma^2(\phi_{R,m}) + o_{a.s.}(1) \end{aligned}$$

donde hemos empleado **S1**. Usando que

$$\widehat{\lambda}_m \geq s_n^2(\widehat{\phi}_m) - \rho[\widehat{\phi}_m, \widehat{\phi}_m] \geq \sigma^2(\phi_{R,m}) + o_{a.s.}(1)$$

y el hecho que $\widehat{\lambda}_m = s_n^2(\widehat{\phi}_m) \xrightarrow{a.s.} \sigma^2(\phi_{R,m})$ implica que $\rho[\widehat{\phi}_m, \widehat{\phi}_m] \xrightarrow{a.s.} 0$, concluyendo c).

d) Ya habíamos probado que cuando $m = 1$ el resultado valía. Procedemos por inducción, supondremos que $\langle \widehat{\phi}_j, \phi_{R,j} \rangle^2 \xrightarrow{a.s.} 1$ para $1 \leq j \leq m-1$ y probaremos que $\langle \widehat{\phi}_m, \phi_{R,m} \rangle^2 \xrightarrow{a.s.} 1$. Por definición $\langle \widehat{\phi}_m, \widehat{\phi}_j \rangle = 0$, para $j \neq m$ y $\widehat{\phi}_m \in \mathcal{S}$ luego, usando que c) implica que $\sigma^2(\widehat{\phi}_m) \xrightarrow{a.s.} \sigma^2(\phi_{R,m})$ y el Lema 3.4.1 b), se sigue la demostración. \square

Vale la pena mencionar que los resultados dados en los Teoremas 3.4.2 y 3.4.3 siguen siendo válidos si r y ρ son aleatorios y $r_n \xrightarrow{a.s.} 0$ y $\rho_n \xrightarrow{a.s.} 0$ lo que permite aplicarlos cuando los parámetros de penalización son adaptativos.

3.4.7 Consistencia de los estimadores robustos basados en *sieves*

El siguiente teorema establece la consistencia de los estimadores definidos en (3.8).

Teorema 3.4.4. Sean $\widehat{\phi}_{SI,m}$ y $\widehat{\lambda}_{SI,m}$ los estimadores definidos en (3.8) y (3.9), respectivamente. Bajo **S0**, **S1** y **S2**, si $p_n \rightarrow \infty$, se verifica

a) $\widehat{\lambda}_{SI,1} \xrightarrow{a.s.} \sigma^2(\phi_{R,1})$ y $\sigma^2(\widehat{\phi}_{SI,1}) \xrightarrow{a.s.} \sigma^2(\phi_{R,1})$

b) Dados $2 \leq m \leq q$, si $\widehat{\phi}_{SI,\ell} \xrightarrow{a.s.} \phi_{R,\ell}$, para $1 \leq \ell \leq m-1$, entonces $\widehat{\lambda}_{SI,m} \xrightarrow{a.s.} \sigma^2(\phi_{R,m})$ y $\sigma^2(\widehat{\phi}_{SI,m}) \xrightarrow{a.s.} \sigma^2(\phi_{R,m})$

c) Para $1 \leq m \leq q$, $\langle \widehat{\phi}_{SI,m}, \phi_m \rangle^2 \xrightarrow{a.s.} 1$.

DEMOSTRACIÓN. A fines de simplificar, vamos a omitir el subíndice SI e indicaremos como $\widehat{\phi}_j = \widehat{\phi}_{SI,j}$ y $\widehat{\lambda}_j = \widehat{\lambda}_{SI,j}$.

a) La demostración se obtiene usando argumentos similares a los de la Proposición 3.4.1. Usando **S1** obtenemos que

$$\widehat{a}_{n,1} = s_n^2(\widehat{\phi}_1) - \sigma^2(\widehat{\phi}_1) \xrightarrow{a.s.} 0. \quad (3.38)$$

Sea $\widetilde{\phi}_{1,p_n} = \pi_{\mathcal{H}_{p_n}} \phi_{R,1} / \|\pi_{\mathcal{H}_{p_n}} \phi_{R,1}\|$, entonces, $\widetilde{\phi}_{1,p_n} \in \mathcal{S}_{p_n}$ y $\widetilde{\phi}_{1,p_n} \rightarrow \phi_{R,1}$. Por ende, **S2** implica que $\sigma(\widetilde{\phi}_{1,p_n}) \rightarrow \sigma(\phi_1)$ mientras que usando **S1**, obtenemos que $s_n^2(\widetilde{\phi}_{1,p_n}) - \sigma^2(\widetilde{\phi}_{1,p_n}) \xrightarrow{a.s.} 0$. Luego, $\widehat{b}_{n,1} = s_n^2(\widetilde{\phi}_{1,p_n}) - \sigma^2(\phi_1) \xrightarrow{a.s.} 0$. Como

$$\sigma^2(\phi_{R,1}) = s_n^2(\widetilde{\phi}_{1,p_n}) - \widehat{b}_{n,1} \leq s_n^2(\widehat{\phi}_1) - \widehat{b}_{n,1} = \sigma^2(\widehat{\phi}_1) + \widehat{a}_{n,1} - \widehat{b}_{n,1} \leq \sigma^2(\phi_{R,1}) + \widehat{a}_{n,1} - \widehat{b}_{n,1},$$

es decir, $\sigma^2(\phi_{R,1}) - \widehat{a}_{n,1} + \widehat{b}_{n,1} \leq \sigma^2(\widehat{\phi}_1) \leq \sigma^2(\phi_{R,1})$, de donde

$$\sigma^2(\widehat{\phi}_1) \xrightarrow{a.s.} \sigma^2(\phi_{R,1}), \quad (3.39)$$

que junto con (3.38) implica que $\widehat{\lambda}_1 \xrightarrow{a.s.} \sigma^2(\phi_{R,1})$.

b) Tenemos que

$$\begin{aligned}
|\widehat{\lambda}_m - \sigma^2(\phi_{R,m})| &= |s_n^2(\widehat{\phi}_m) - \sigma^2(\phi_{R,m})| = \left| \max_{\alpha \in \widehat{\mathcal{B}}_{n,m}} s_n^2(\alpha) - \max_{\alpha \in \mathcal{B}_m} \sigma^2(\alpha) \right| \\
&\leq \max_{\alpha \in \widehat{\mathcal{B}}_{n,m}} |s_n^2(\alpha) - \sigma^2(\alpha)| + \left| \max_{\alpha \in \widehat{\mathcal{B}}_{n,m}} \sigma^2(\alpha) - \max_{\alpha \in \mathcal{B}_m} \sigma^2(\alpha) \right| \\
&\leq \max_{\alpha \in \mathcal{S}_{p_n}} |s_n^2(\alpha) - \sigma^2(\alpha)| + \left| \max_{\alpha \in \widehat{\mathcal{B}}_{n,m}} \sigma^2(\alpha) - \max_{\alpha \in \mathcal{B}_m} \sigma^2(\alpha) \right|. \quad (3.40)
\end{aligned}$$

Usando **S1**, obtenemos que el primer término en el lado derecho de (3.40) converge a 0 casi seguramente. Entonces, bastará con demostrar que $\max_{\alpha \in \widehat{\mathcal{B}}_{n,m}} \sigma^2(\alpha) \xrightarrow{a.s.} \sigma^2(\phi_{R,m})$, puesto que $\max_{\alpha \in \mathcal{B}_m} \sigma^2(\alpha) = \sigma^2(\phi_{R,m})$. Usando que $\widehat{\phi}_s \xrightarrow{a.s.} \phi_{R,s}$, for $1 \leq s \leq m-1$, obtenemos que

$$\|\pi_{\widehat{\mathcal{L}}_{m-1}} - \pi_{\mathcal{L}_{m-1}}\| \xrightarrow{a.s.} 0, \quad (3.41)$$

lo que implica que

$$\|\pi_{\widehat{\mathcal{L}}_{m-1}} \widetilde{\phi}_m - \pi_{\mathcal{L}_{m-1}} \widetilde{\phi}_m\| \xrightarrow{a.s.} 0, \quad (3.42)$$

donde $\widetilde{\phi}_m = \pi_{\mathcal{H}_{p_n}} \phi_{R,m}$. Por otro lado, usando que $p_n \rightarrow \infty$, obtenemos que $\widetilde{\phi}_m \rightarrow \phi_{R,m}$, entonces $\|\pi_{\mathcal{L}_{m-1}}(\widetilde{\phi}_m - \phi_{R,m})\| \rightarrow 0$, que junto con (3.42) y el hecho que $\pi_{\mathcal{L}_{m-1}} \phi_{R,m} = 0$, implican que $\pi_{\widehat{\mathcal{L}}_{m-1}} \widetilde{\phi}_m \xrightarrow{a.s.} 0$ y $\widetilde{\beta}_m = \widetilde{\phi}_m - \pi_{\widehat{\mathcal{L}}_{m-1}} \widetilde{\phi}_m \xrightarrow{a.s.} \phi_{R,m}$. Observando que $a_m = \|\widetilde{\beta}_m\|$, $a_m \xrightarrow{a.s.} 1$, obtenemos que $\widehat{\alpha}_m = \widetilde{\beta}_m/a_m \xrightarrow{a.s.} \phi_{R,m}$. Más aún, $\widehat{\alpha}_m \in \widehat{\mathcal{B}}_{n,m}$ puesto que $\widetilde{\phi}_m \in \mathcal{H}_{p_n}$ y $\widehat{\phi}_j \in \mathcal{H}_{p_n}$, $1 \leq j \leq m-1$, implicando que $\max_{\alpha \in \widehat{\mathcal{B}}_{n,m}} \sigma^2(\alpha) \geq \sigma^2(\widehat{\alpha}_m)$. Usando la continuidad débil de σ , obtenemos que $\sigma(\widehat{\alpha}_m) \xrightarrow{a.s.} \sigma(\phi_{R,m})$, es decir,

$$\max_{\alpha \in \widehat{\mathcal{B}}_{n,m}} \sigma^2(\alpha) \geq \sigma^2(\widehat{\alpha}_m) = \sigma^2(\phi_{R,m}) + o_{a.s.}(1).$$

Por otro lado,

$$\begin{aligned}
\max_{\alpha \in \widehat{\mathcal{B}}_{n,m}} \sigma^2(\alpha) &\leq \max_{\alpha \in \widehat{\mathcal{B}}_{n,m}} |\sigma^2(\alpha) - s_n^2(\alpha)| + s_n^2(\widehat{\phi}_m) \\
&\leq 2 \max_{\alpha \in \widehat{\mathcal{B}}_{n,m}} |\sigma^2(\alpha) - s_n^2(\alpha)| + \sigma^2(\widehat{\phi}_m) = o_{a.s.}(1) + \sigma^2(\widehat{\phi}_m)
\end{aligned}$$

Usando (3.41), $\widehat{\phi}_m \in \widehat{\mathcal{B}}_{n,m}$ y $\|\widehat{\phi}_m\| = 1$, obtenemos que $\widehat{\phi}_m = \widehat{\phi}_m - \pi_{\widehat{\mathcal{L}}_{m-1}} \widehat{\phi}_m = \widehat{b}_m + o_{a.s.}(1)$, donde $\widehat{b}_m = \widehat{\phi}_m - \pi_{\mathcal{L}_{m-1}} \widehat{\phi}_m$. Entonces $\|\widehat{b}_m\| \xrightarrow{a.s.} 1$. Llamemos $\widehat{\beta}_m = \widehat{b}_m/\|\widehat{b}_m\|$, luego $\widehat{\beta}_m \in \mathcal{B}_m$ lo que implica que $\sigma(\widehat{\beta}_m) \leq \sigma(\phi_{R,m})$. Además, la continuidad débil de σ y la compacidad débil de la bola unitaria conllevan, como en el Lema 3.4.1, que $\sigma(\widehat{\beta}_m) - \sigma(\widehat{\phi}_m) \xrightarrow{a.s.} 0$, puesto que $\widehat{\phi}_m - \widehat{\beta}_m = o_{a.s.}(1)$. Resumiendo,

$$\begin{aligned}
\sigma^2(\phi_{R,m}) + o_{a.s.}(1) = \sigma^2(\widehat{\alpha}_m) &\leq \max_{\alpha \in \widehat{\mathcal{B}}_{n,m}} \sigma^2(\alpha) \leq o_{a.s.}(1) + \sigma^2(\widehat{\phi}_m) = o_{a.s.}(1) + \sigma^2(\widehat{\beta}_m) \\
&\leq o_{a.s.}(1) + \sigma^2(\phi_{R,m})
\end{aligned}$$

concluyendo que $\max_{\alpha \in \widehat{\mathcal{B}}_{n,m}} \sigma^2(\alpha) \xrightarrow{a.s.} \sigma^2(\phi_{R,m})$ y así la prueba de que $\widehat{\lambda}_m \xrightarrow{a.s.} \sigma^2(\phi_{R,m})$. Más aún, es fácil ver que **S1** y el hecho que $\widehat{\lambda}_m \xrightarrow{a.s.} \sigma^2(\phi_{R,m})$ implican que $\sigma^2(\widehat{\phi}_m) \xrightarrow{a.s.} \sigma^2(\phi_{R,m})$, puesto que

$$\begin{aligned} |\sigma^2(\widehat{\phi}_m) - \sigma^2(\phi_{R,m})| &\leq |\sigma^2(\widehat{\phi}_m) - s_n^2(\widehat{\phi}_m)| + |s_n^2(\widehat{\phi}_m) - \sigma^2(\phi_{R,m})| \\ &= |\sigma^2(\widehat{\phi}_m) - s_n^2(\widehat{\phi}_m)| + |\widehat{\lambda}_m - \sigma^2(\phi_{R,m})| \\ &\leq \sup_{\alpha \in \mathcal{S}_{p_n}} |\sigma^2(\alpha) - s_n^2(\alpha)| + |\widehat{\lambda}_m - \sigma^2(\phi_{R,m})|. \end{aligned}$$

c) Empezamos probando $\widehat{\phi}_1 \xrightarrow{a.s.} \phi_{R,1}$. Usando $\sigma^2(\widehat{\phi}_1) \xrightarrow{a.s.} \sigma^2(\phi_{R,1})$ y el Lema 3.4.1 a), el resultado se sigue fácilmente puesto que $\widehat{\phi}_1 \in \mathcal{S}$. Probemos que si $\widehat{\phi}_j \xrightarrow{a.s.} \phi_{R,j}$ para $1 \leq j \leq m-1$ luego $\langle \widehat{\phi}_m, \phi_{R,m} \rangle^2 \xrightarrow{a.s.} 1$, lo que nos llevará a c). Como $\widehat{\phi}_j \xrightarrow{a.s.} \phi_{R,j}$ para $1 \leq j \leq m-1$, tenemos que $\sigma^2(\widehat{\phi}_m) \xrightarrow{a.s.} \sigma^2(\phi_{R,m})$. Más aún, usando que $\langle \widehat{\phi}_m, \widehat{\phi}_j \rangle = 0$, $j \neq m$ y $\widehat{\phi}_m \in \mathcal{S}$, el Lema 3.4.1 b) concluye la prueba. \square

3.5 Elección del parámetro de penalización

La elección del parámetro de penalización es un asunto de gran importancia práctica ya que el mismo determina el compromiso entre sesgo y varianza realizado. El método más popular para resolver esta cuestión es el de validación cruzada (“cross-validation”). En regresión no paramétrica, la sensibilidad de los métodos de validación cruzada basados en distancia L^2 fueron indicados por Wang y Scott (1994) y por Cantoni y Ronchetti (2001), entre otros. Estos últimos además propusieron alternativas más robustas a la validación cruzada L^2 .

La idea de validación cruzada se puede adaptar al problema en estudio. Supongamos por un momento que estamos interesados en una cantidad fija, ℓ , de componentes. Proponemos entonces el siguiente procedimiento:

1. Centramos los datos, es decir, definimos $\widetilde{X}_i = X_i - \widehat{\mu}$ donde $\widehat{\mu}$ es un estimador robusto de posición como las medias podadas (según propuesto por Fraiman y Muniz (2001)), los estimadores de profundidad de Cuevas *et al.* (2007) y López-Pintado y Romo (2007), o la mediana funcional definida en Gervini (2008).
2. Para los estimadores que utilizan penalización por rugosidad, sea $\tau_n = r$ o ρ según se penalice la norma o la escala. Para cada m en el rango $1 \leq m \leq \ell$ y $0 < \tau_n$, sea $\widehat{\phi}_{m,\tau_n}^{(-j)}$ el estimador robusto de la componente principal m -ésima, calculada sin la j -ésima observación.
3. Definamos $X_j^\perp(\tau_n) = \widetilde{X}_j - \pi_{\widehat{\mathcal{L}}_\ell^{(-j)}}(\widetilde{X}_j)$, donde $\pi_{\mathcal{H}_s}(X)$ corresponde a la proyección ortogonal de X sobre el subespacio lineal cerrado \mathcal{H}_s y $\widehat{\mathcal{L}}_\ell^{(-j)}$ indicará el subespacio lineal generado por $\widehat{\phi}_{1,\tau_n}^{(-j)}, \dots, \widehat{\phi}_{\ell,\tau_n}^{(-j)}$.
4. Dada una escala robusta alrededor del cero σ_n , proponemos minimizar $RCV_\ell(\tau_n) = \sigma_n^2(\|X_1^\perp(\tau_n)\|, \dots, \|X_n^\perp(\tau_n)\|)$.

Por escala robusta alrededor del cero, entendemos que no se utiliza ningún estimador de posición para centrar los datos. Por ejemplo, en el contexto clásico, tomaríamos $\sigma_n^2(z_1, \dots, z_n) = (1/n) \sum_{i=1}^n z_i^2$ mientras que en la situación robusta, podríamos considerar $\sigma_n(z_1, \dots, z_n) = \text{mediana}(|z_1|, \dots, |z_n|)$ o la solución de $\sum_{i=1}^n \chi(|z_i|/\sigma_n) = n/2$.

Para tamaños de muestra grandes, es conocido que los métodos de validación cruzada pueden incurrir en costos computacionales prohibitivos. En tal caso, validación cruzada por K -fold suele proveer una alternativa adecuada. Describimos a continuación brevemente una propuesta mediante K -folds para los estimadores propuestos.

1. Se centran los datos, como antes, usando $\tilde{X}_i = X_i - \hat{\mu}$.
2. Particionamos los datos centrados $\{\tilde{X}_i\}$ de forma aleatoria en K subconjuntos disjuntos de aproximadamente el mismo tamaño con el j -ésimo subconjunto teniendo tamaño $n_j \geq 2$, $\sum_{j=1}^K n_j = n$. Sea $\{\tilde{X}_i^{(j)}\}_{1 \leq i \leq n_j}$ los elementos del j -ésimo subconjunto, y $\{\tilde{X}_i^{(-j)}\}_{1 \leq i \leq n-n_j}$ denotará los elementos en el complemento del j -ésimo subconjunto. El conjunto $\{\tilde{X}_i^{(-j)}\}_{1 \leq i \leq n-n_j}$ será el conjunto de entrenamiento y $\{\tilde{X}_i^{(j)}\}_{1 \leq i \leq n_j}$ el de validación.
3. Procedemos de forma similar al Paso 2 en el procedimiento puro de validación cruzada, pero dejando el j -ésimo subconjunto de validación $\{\tilde{X}_i^{(j)}\}_{1 \leq i \leq n_j}$ afuera en vez de la j -ésima observación.
4. Definimos $X_j^{(j)\perp}(\tau_n)$ como antes, usando el conjunto de validación. Es decir, $X_i^{(j)\perp}(\tau_n) = \tilde{X}_i^{(j)} - \pi_{\hat{\mathcal{L}}_\ell^{(-j)}}(\tilde{X}_i^{(j)})$, $1 \leq i \leq n_j$, donde $\hat{\mathcal{L}}_\ell^{(-j)}$ es el subespacio lineal generado por $\hat{\phi}_{1,\tau_n}^{(-j)}, \dots, \hat{\phi}_{\ell,\tau_n}^{(-j)}$.
5. Dado un estimador de escala alrededor del cero σ_n , el método robusto de validación cruzada con K -fold elige el parámetro de suavizado que minimiza

$$RCV_{\ell, \text{KCV}}(\tau_n) = \sum_{j=1}^K \sigma_n^2(\|X_1^{(j)\perp}(\tau_n)\|, \dots, \|X_{n_j}^{(j)\perp}(\tau_n)\|).$$

Un procedimiento similar se puede implementar para elegir p_n al utilizar los estimadores de sieves.

3.6 Estudio Monte Carlo

3.6.1 Descripción del Algoritmo y notación

Todos los métodos aquí considerados son adaptaciones al caso funcional del algoritmo propuesto por Croux y Ruiz-Gazen (1996) para el cálculo de componentes principales utilizando *projection-pursuit*. El algoritmo se aplica en principio a datos multivariados, por ejemplo m -dimensionales, y requiere una búsqueda sobre todas las direcciones unitarias de \mathbb{R}^m . Para aplicar este algoritmo

a datos funcionales, discretizamos el dominio de las funciones observadas sobre $m = 50$ puntos equiespaciados en $\mathcal{I} = [-1, 1]$. Hemos también adaptado el algoritmo para permitir la búsqueda de estimadores penalizados de las direcciones principales y distintos métodos de centrado. En ese sentido, hay tres características principales que distinguen a los distintos estimadores: el funcional de escala, el método de centrado, el tipo y la cantidad de suavizado utilizado.

- **Funcional de escala:** Consideramos tres funcionales: el desvío estándar clásico (SD), el desvío de las medianas absolutas (MAD) y un M -estimador de escala (M -escala). Los dos últimos son funcionales robustos de escala. El M -estimador combina la robustez de la MAD con la suavidad del desvío estándar. Para el M -estimador, se utilizó como función de *score* $\chi_c(y) = \min\left(3(y/c)^2 - 3(y/c)^4 + (y/c)^6, 1\right)$, introducida por Beaton y Tukey (1974), con $c = 1.56$ y punto de ruptura $1/2$. Para computar la M -escala, el estimador inicial de escala fue la MAD.
- **Centrado:** Para los procedimientos clásicos, es decir, aquellos basados en SD, se utilizó una media punto a punto como esquema de centrado. Para los procedimientos robustos, es decir, aquellos basados en la MAD o en una M -escala, utilizamos la mediana L^1 , que es comúnmente referida como la mediana espacial, o la mediana punto a punto para centrar los datos. Esto evita la complejidad adicional asociada a los funcionales basados en medianas podadas o estimadores basados en profundidad. Se evidenció empíricamente que ambos esquemas de centrado producían resultados similares, por lo tanto para el esquema clásico se utilizó únicamente la media punto a punto y para los robustos la mediana L^1 .
- **Nivel de penalización r y ρ :** Para ambos procedimientos clásicos y robustos, una penalización basada en la norma L^2 de la segunda derivada es utilizada, multiplicada por un factor de suavizado. Notemos que, como se detalló en la exposición de los estimadores, tenemos dos formas de aplicar la penalización, o bien afectando la función objetivo o bien redefiniendo el producto interno de forma tal de contemplar el suavizado. Ambos procedimientos fueron puestos en práctica. Observemos que cuando $r = \rho = 0$, obtenemos las propuestas “crudas”. También se consideró un suavizado de las observaciones en el algoritmo, utilizando un suavizador por núcleos para los procedimientos clásicos y una mediana local para los robustos. Sin embargo, se pudo evidenciar que este procedimiento era extremadamente costoso en tiempo, sin una ventaja apreciable en los resultados. Por eso, se omiten los resultados.
- **Bases para el método basado en Sieves:** Consideramos dos bases diferentes de *sieves*: la base de Fourier, es decir, tomando δ_j como las funciones de la base de Fourier, y las funciones que conforman la base de un B -spline cúbico. La base de Fourier utilizada en el método de *sieve* es la misma base que la que se utiliza para generar los datos.

En todas las Figuras y Tablas, los estimadores correspondientes a cada elección de escala son indicadas como SD, MAD, M -ESCALA. Para cada escala, consideramos cuatro estimadores, el estimador sin suavizado, aquel que se obtiene al penalizar la función de escala tal como se define en

(3.6), aquellos que se obtienen penalizando la norma como se define en (3.5), y los estimadores de *sieve* definidos en (3.8). En todas las tablas, el j -ésimo estimador se indica como $\hat{\phi}_j$, $\hat{\phi}_{\text{PS},j}$, $\hat{\phi}_{\text{PN},j}$ y $\hat{\phi}_{\text{SI},j}$, respectivamente.

Cuando se usan los estimadores penalizados, se consideran distintos valores para los parámetros de penalizado r y ρ . Puesto que grandes valores del parámetro de suavizado hacen que el término de penalizado domine sobre la escala en la elección del estimador de la componente principal, elegimos los valores de r y ρ iguales a $an^{-\alpha}$ para $\alpha = 3$ y 4 y a igual a $0, 0.05, 0.10, 0.15, 0.25, 0.5, 0.75, 1, 1.5$ y 2 . Sin embargo, los boxplots y los estimadores de densidad de los estimadores solamente se grafican para $\alpha = 3$ y $a = 0, 1.5$ para $\hat{\phi}_{\text{PS}}$ y 0.75 para $\hat{\phi}_{\text{PN}}$.

Para el método de *sieves* basado en la base de Fourier, ordenada como $\{1, \cos(\pi x), \sin(\pi x), \dots, \cos(q_n \pi x), \sin(q_n \pi x), \dots\}$, utilizamos los valores $p_n = 2q_n$ con $q_n = 5, 10$ y 15 , mientras que para los estimadores de *sieves* basados en B -splines, elegimos la dimensión del subespacio lineal como $p_n = 10, 20$ y 50 . La base de B -splines fue generada utilizando la función de R *cSplineDes*, con los nodos equiespaciados en el intervalo $[-1, 1]$ y siendo su cantidad igual a $p_n + 1$. La base B -spline resultante, sin embargo, no es ortonormal. Puesto que es más fácil aplicar el algoritmo para los estimadores de *sieves* cuando se utiliza una base ortonormal, aplicamos un procedimiento de Gram-Schmidt a la base de B -splines para obtener una nueva base ortonormal que genere el mismo subespacio.

3.6.2 Modelo central y contaminaciones

La muestra fue generada utilizando una expansión finita de Karhunen-Loève con las funciones, $\phi_i : [-1, 1] \rightarrow \mathbb{R}$, $i = 1, 2, 3$, con

$$\begin{aligned}\phi_1(x) &= \sin(4\pi x) \\ \phi_2(x) &= \cos(7\pi x) \\ \phi_3(x) &= \cos(15\pi x) .\end{aligned}$$

Vale la pena notar que, cuando se consideran los estimadores de *sieve* basados en la base de Fourier, la tercer componente no podrá ser detectada cuando $q_n < 15$, ya que, en este caso, $\phi_3(x)$ es ortogonal al espacio de estimación. De igual forma, la segunda componente no podrá ser detectada cuando $q_n < 7$.

Se realizaron $NR = 1000$ replicaciones generando muestras independientes $\{X_i\}_{i=1}^n$ de tamaño $n = 100$ siguiendo el modelo $X_i = Z_{i1}\phi_1 + Z_{i2}\phi_2 + Z_{i3}\phi_3$, donde Z_{ij} son variables aleatorias independientes cuya distribución depende de la situación considerada. El modelo central, notado C_0 , corresponde a muestras Gaussianas. También hemos considerado cuatro contaminaciones del modelo central, denotadas $C_2, C_{3,a}, C_{3,b}$ y C_{23} dependiendo de la componente a ser contaminada. El modelo central y las contaminaciones pueden ser descriptas como sigue. Para cada modelo, tomamos $\sigma_1 = 4, \sigma_2 = 2$ y $\sigma_3 = 1$.

- C_0 : $Z_{i1} \sim N(0, \sigma_1^2)$, $Z_{i2} \sim N(0, \sigma_2^2)$ y $Z_{i3} \sim N(0, \sigma_3^2)$.

- C_2 : Z_{i2} independientes e idénticamente distribuidas como $0.8 N(0, \sigma_2^2) + 0.2 N(10, 0.01)$, mientras que $Z_{i1} \sim N(0, \sigma_1^2)$ y $Z_{i3} \sim N(0, \sigma_3^2)$. Este caso se corresponde a una fuerte contaminación en la segunda componente y cambia el valor medio de los datos generados Z_{i2} y también la primer componente principal. Notemos que $\text{var}(Z_{i2}) = 19.202 > \text{var}(Z_{i1}) = 16$, luego la segunda componente “gana” a la primer componente y el orden queda ϕ_2, ϕ_1, ϕ_3 . Luego, el procedimiento clásico no robusto estimaría mal las dos primeras componentes pero bien la tercera.
- $C_{3,a}$: $Z_{i1} \sim N(0, \sigma_1^2)$, $Z_{i2} \sim N(0, \sigma_2^2)$ y $Z_{i3} \sim 0.8 N(0, \sigma_3^2) + 0.2 N(15, 0.01)$. Este modelo corresponde a una fuerte contaminación en la tercer componente. Notemos que $\text{var}(Z_{i3}) = 36.802 > \text{var}(Z_{i1}) = 16$. Luego, el orden entre las autofunciones queda ϕ_3, ϕ_1, ϕ_2 .
- $C_{3,b}$: $Z_{i1} \sim N(0, \sigma_1^2)$, $Z_{i2} \sim N(0, \sigma_2^2)$ y $Z_{i3} \sim 0.8 N(0, \sigma_3^2) + 0.2 N(6, 0.01)$. Este modelo corresponde a una fuerte contaminación en la tercer componente. Notemos que $\text{var}(Z_{i3}) = 6.562 > \text{var}Z_{i2} = 4$. El nuevo orden queda ϕ_1, ϕ_3, ϕ_2 .
- C_{23} : Z_{ij} son independientes y tales que $Z_{i1} \sim N(0, \sigma_1^2)$, $Z_{i2} \sim 0.9N(0, \sigma_2^2) + 0.1N(15, 0.01)$ y $Z_{i3} \sim 0.9N(0, \sigma_3^2) + 0.1N(20, 0.01)$. Este modelo corresponde a una contaminación suave en las dos últimas componentes. Notemos que $\text{var}(Z_{i2}) = 23.851$, y $\text{var}(Z_{i3}) = 36.901$. Luego el nuevo orden queda ϕ_3, ϕ_2, ϕ_1 .

También consideramos un escenario en donde los datos se generan mediante una distribución de Cauchy, indicado C_c , definido al tomar $(Z_{i1}, Z_{i2}, Z_{i3}) \sim \mathcal{C}_3(0, \Sigma)$ con $\Sigma = \text{diag}(\sigma_1^2, \sigma_2^2, \sigma_3^2)$, donde $\mathcal{C}_p(0, \Sigma)$ corresponde a la distribución elíptica p -dimensional de una distribución de Cauchy centrado en 0 y con matriz de dispersión Σ . Para este escenario, el operador de covarianza no existe y por ende las componente principales en el sentido clásico no estarán definidas.

Vale la pena destacar que las direcciones ϕ_1, ϕ_2 y ϕ_3 corresponden a las componentes principales clásicas para el caso C_0 , pero no necesariamente para los otros casos. Como se mencionó anteriormente, $C_{3,a}$ intercambia el orden entre ϕ_1 y ϕ_3 , es decir, ϕ_3 es ahora la primer componente clásica, o sea aquella que se obtiene mediante el operador de covarianza, mientras que ϕ_1 es la segunda y ϕ_3 es la tercera.

3.6.3 Resultados de la simulación

Para cada situación calculamos los estimadores de las tres primeras componentes principales y la distancia al cuadrado entre la dirección verdadera y la estimada (normalizadas para que tengan norma L^2 igual a 1), es decir,

$$D_j = \left\| \frac{\hat{\phi}_j}{\|\hat{\phi}_j\|} - \phi_j \right\|^2.$$

Las Tablas 3.5 a 3.10 dan la media de D_j sobre las replicaciones para los estimadores “crudos” y penalizados. La Tabla 3.5 corresponde al estimador “crudo” y penalizado bajo C_0 para diferentes valores del parámetro de penalización. Esta tabla permite ver que una mejor performance se alcanza

en casi todos los casos con $\alpha = 3$. Luego, como se mencionó antes, todas las Figuras corresponden a casos en donde el parámetro de penalización vale an^{-3} . Para ser más precisos, los resultados en la Tabla 3.5 muestran que la mejor elección para $\hat{\phi}_{PS,j}$ es $r = 2n^{-3}$ para todo j . Notemos que $\rho = 1.5n^{-3}$ ofrece resultados muy similares, cuando se usa la M -escala, reduciendo el error en un medio y un tercio para $j = 2$ y $j = 3$, respectivamente.

Cuando penalizamos en la norma, es decir, cuando consideramos $\hat{\phi}_{PN,j}$ la elección del parámetro de penalizado parece depender tanto de la componente a estimar como en el estimador utilizado. Por ejemplo, cuando se utiliza el desvío estándar, la mejor elección es $0.10n^{-3}$, para $j = 1$ y $j = 2$ mientras que para $j = 3$ un orden menor es necesitado para obtener una mejora sobre los estimadores “crudos”. El valor $r = 0.75n^{-4}$ nos lleva a una ganancia menor sobre los estimadores “crudos”. Para los procedimientos robustos, valores más grandes son necesarios para ver la ventaja del acercamiento penalizado sobre los estimadores “crudos”. Por ejemplo, para $j = 1$, la reducción más grande se observa en con $r = 2n^{-3}$ mientras que para $j = 2$, la mejor elección corresponde a $r = 0.5n^{-3}$ y $r = 0.25n^{-3}$ cuando se utiliza la MAD y una M -escala, respectivamente. Por ejemplo, cuando se usa una M -escala, eligiendo $r = 0.75n^{-3}$ lleva a una reducción de aproximadamente el 30% y 50% para la primera y la segunda dirección principal, respectivamente. Por otro lado, cuando se estima la tercer componente, nuevamente valores más chicos para r son necesarios.

Las Tablas 3.2 y 3.3 reportan la media de D_j sobre replicaciones para diferentes tamaños de la grilla bajo C_0 para algunos valores de los parámetros de penalizado mientras que la Tabla 3.1 corresponde a los estimadores sin penalizar. El tamaño $m = 50$ elegido en nuestro estudio ofrece un compromiso entre la performance de los estimadores y el tiempo de cómputo. Como puede ser visto, se observan mejoras cuando se utiliza $m = 250$ en vez de 50 puntos pero a expensas de multiplicar por cinco el costo computacional.

Además, las Tablas 3.11 hasta 3.16 ofrecen la media de D_j sobre replicaciones del estimador de *sieves*. Las Figuras 3.1 a 3.4 muestran los estimadores de densidad de D_j , para $j = 1, 2$ y 3 , cuando $r = \rho = 0$ y cuando el parámetro de penalización es igual a an^{-3} donde $a = 1.5$ y 0.75 para los estimadores que penalizan la escala y la norma, respectivamente. Los estimadores de densidad fueron evaluados usando un núcleo normal con ancho de banda 0.6 en todos los casos. Las líneas en negro corresponden a las densidades de D_j evaluados sobre $NR = 1000$ muestras normalmente distribuidas, mientras que las líneas rojas, grises, azules y verdes corresponden a C_2 , $C_{3,a}$, $C_{3,b}$ y C_{23} , respectivamente. Finalmente, las Figuras 3.8 hasta 3.13 muestran los boxplots del cociente $\hat{\lambda}_m/\lambda_m$ para los estimadores de autovalores. Los estimadores clásicos están indicados como SD, mientras que los robustos se indican por MAD y MS. Para los estimadores penalizados por la escala o la norma, el parámetro de penalizado se indica después del nombre del tipo de estimador utilizado, mientras que para el método de *sieves* el parámetro p_n sigue al nombre del estimador de escala considerado. Para la distribución Cauchy, los estimadores clásicos de los valores principales dan valores muy grandes que oscurecen las diferencias con los procedimientos robustos, por esta razón, las Figuras 3.8 hasta 3.13 dan solamente los boxplots para los estimadores robustos.

La simulación confirma el comportamiento inadecuado de los estimadores clásicos ante la presencia de *outliers*. Se observa un sesgo al estimar los autovalores. La peor eficiencia de los estimadores de autovalores “crudos” se obtiene usando el método de *projection-pursuit* combinado con el es-

timador de escala MAD. Vale la pena notar además que el nivel de suavizado parece afectar los estimadores de autovalores, introduciendo un sesgo incluso en el caso de muestras Gaussianas. Notemos que para algunas combinaciones, los estimadores robustos también son sesgados. Sin embargo, el orden entre ellos es preservado y, por ende, la autofunción objetivo es en la mayoría de los casos bien estimada.

Con respecto a la estimación de la dirección principal, bajo contaminación, los estimadores clásicos no estiman la autofunción objetivo de forma muy acertada, cosa que puede verse con el corrimiento de la densidad de la norma hacia el 2. Notemos que cuando se considera la distribución de Cauchy, el efecto principal es observado en los estimadores de los autovalores ya que, incluso cuando no existe el operador de covarianza, las direcciones parecen ser recuperadas al usar el desvío estándar. Los estimadores robustos de autofunciones parecen verse poco afectados en todos los casos de contaminación salvo bajo $C_{3,a}$. En particular, los estimadores *projection-pursuit* basados en una M -escala parecen más afectados por esta contaminación. Por otro lado, $C_{3,a}$ afecta a los estimadores de la tercer autofunción cuando se penaliza la norma. Con respecto a ese caso, los estimadores robustos obtenidos por penalización de la norma $\hat{\phi}_{PN,j}$ muestran el menor efecto sobre todos sus competidores. Cabe notar que incluso si el orden de las autofunciones es modificado, como se mencionó antes, los estimadores robustos de la primer dirección principal no son afectados por esta contaminación.

Vale la pena notar que los estimadores clásicos de la primer componente no son afectados por $C_{3,a}$ para algunos valores del parámetro de suavizado, cuando se penaliza por la norma, ya que la penalización domina sobre las varianzas contaminadas. El mismo fenómeno es observado bajo $C_{3,b}$ cuando se usa los estimadores clásicos por la cantidad elegida de penalización. Para los estimadores “crudos”, la sensibilidad de los estimadores clásicos bajo esta contaminación puede ser observada en la Tabla 3.4.

Tal como fue notado en Silverman (1996), para los estimadores clásicos, algún nivel de suavizado en el procedimiento basado en penalizar la norma ofrecerá una mejor estimación de ϕ_j en el sentido de L^2 bajo condiciones suaves. En particular, ambos procedimientos de penalización (por escala y por norma) proveen una cierta mejora con respecto a las alternativas “crudas” si $\Psi(\phi_j) < \Psi(\phi_\ell)$, cuando $j < \ell$. Esto significa que las direcciones principales son más “rugosas” a medida que decrece el autovalor (ver Pezzulli y Silverman, 1993, y Silverman, 1996), lo que está también reflejado en nuestro estudio de simulación. Las ventajas del procedimiento suave de *projection-pursuit* son más impactantes cuando se estima ϕ_2 y ϕ_3 con una M -escala y cuando se utiliza el esquema de penalización por la escala.

Como era esperado, cuando se usan los estimadores por *sieves*, la base de Fourier ofrece el mejor comportamiento sobre todos los métodos bajo C_0 , puesto que nuestros datos fueron generados con dicha base (ver Tabla 3.11). La elección de la base B -spline da resultados similares a aquellos obtenidos con $\hat{\phi}_{PS,j}$ cuando $p_n = 50$.

Estimador de escala	m	$j = 1$	$j = 2$	$j = 3$	m	$j = 1$	$j = 2$	$j = 3$
SD	50	0.0080	0.0117	0.0100	100	0.0078	0.0113	0.0079
MAD	50	0.0744	0.1288	0.0879	100	0.0700	0.1212	0.0827
M -ESCALA	50	0.0243	0.0424	0.0295	100	0.0237	0.0416	0.0271
SD	150	0.0077	0.0112	0.0075	200	0.0077	0.0112	0.0073
MAD	150	0.0703	0.1216	0.0824	200	0.0705	0.1223	0.0825
M -ESCALA	150	0.0234	0.0414	0.0268	200	0.0233	0.0416	0.0269
SD	250	0.0077	0.0112	0.0073				
MAD	250	0.0701	0.1212	0.0815				
M -ESCALA	250	0.0233	0.0414	0.0267				

Tabla 3.1: Valores medios de $\|\widehat{\phi}_j - \phi_j\|^2$ bajo C_0 , para los estimadores “crudos”, para distintos tamaños m de la grilla.

Método	Estimador de escala	m	$a = 0.15$			$a = 0.5$			$a = 1$		
			$j = 1$	$j = 2$	$j = 3$	$j = 1$	$j = 2$	$j = 3$	$j = 1$	$j = 2$	$j = 3$
$\widehat{\phi}_{PS,j}$	SD	50	0.0078	0.0106	0.0090	0.0077	0.0090	0.0074	0.0077	0.0081	0.0064
	MAD	50	0.0737	0.1187	0.0780	0.0720	0.1061	0.0663	0.0702	0.0929	0.0531
	M -ESCALA	50	0.0240	0.0377	0.0249	0.0239	0.0317	0.0187	0.0232	0.0270	0.0136
$\widehat{\phi}_{PS,j}$	SD	100	0.0076	0.0095	0.0061	0.0075	0.0079	0.0043	0.0073	0.0069	0.0032
	MAD	100	0.0698	0.1092	0.0697	0.0687	0.0927	0.0529	0.0668	0.0782	0.0380
	M -ESCALA	100	0.0234	0.0345	0.0198	0.0226	0.0269	0.0121	0.0221	0.0231	0.0078
$\widehat{\phi}_{PS,j}$	SD	150	0.0076	0.0094	0.0057	0.0075	0.0077	0.0038	0.0072	0.0068	0.0027
	MAD	150	0.0695	0.1088	0.0692	0.0678	0.0883	0.0483	0.0663	0.0758	0.0346
	M -ESCALA	150	0.0231	0.0340	0.0190	0.0224	0.0262	0.0111	0.0218	0.0223	0.0068
$\widehat{\phi}_{PS,j}$	SD	200	0.0075	0.0093	0.0054	0.0074	0.0076	0.0036	0.0071	0.0067	0.0025
	MAD	200	0.0699	0.1080	0.0678	0.0680	0.0880	0.0475	0.0663	0.0751	0.0337
	M -ESCALA	200	0.0230	0.0336	0.0186	0.0223	0.0259	0.0106	0.0217	0.0221	0.0065
$\widehat{\phi}_{PS,j}$	SD	250	0.0075	0.0093	0.0054	0.0074	0.0075	0.0035	0.0071	0.0066	0.0024
	MAD	250	0.0695	0.1080	0.0679	0.0679	0.0881	0.0474	0.0661	0.0750	0.0333
	M -ESCALA	250	0.0228	0.0333	0.0184	0.0223	0.0258	0.0105	0.0216	0.0219	0.0063
$\widehat{\phi}_{PN,j}$	SD	50	0.0075	0.0075	0.0161	0.0087	0.0095	0.0490	0.0093	0.0113	0.1197
	MAD	50	0.0619	0.0731	0.1465	0.0552	0.0650	0.2687	0.0511	0.0658	0.4073
	M -ESCALA	50	0.0203	0.0216	0.0310	0.0193	0.0213	0.0715	0.0192	0.0233	0.1470
$\widehat{\phi}_{PN,j}$	SD	100	0.0075	0.0086	0.0078	0.0073	0.0073	0.0099	0.0075	0.0074	0.0151
	MAD	100	0.0675	0.1012	0.0988	0.0617	0.0799	0.1226	0.0603	0.0706	0.1499
	M -ESCALA	100	0.0220	0.0293	0.0250	0.0205	0.0227	0.0257	0.0198	0.0206	0.0297
$\widehat{\phi}_{PN,j}$	SD	150	0.0075	0.0100	0.0073	0.0074	0.0087	0.0073	0.0073	0.0079	0.0077
	MAD	150	0.0694	0.1153	0.0864	0.0676	0.1072	0.0994	0.0652	0.0914	0.0997
	M -ESCALA	150	0.0229	0.0371	0.0263	0.0221	0.0306	0.0246	0.0213	0.0264	0.0245
$\widehat{\phi}_{PN,j}$	SD	200	0.0076	0.0107	0.0072	0.0075	0.0099	0.0072	0.0075	0.0091	0.0072
	MAD	200	0.0699	0.1183	0.0821	0.0689	0.1138	0.0865	0.0677	0.1119	0.0953
	M -ESCALA	200	0.0232	0.0396	0.0262	0.0226	0.0360	0.0253	0.0222	0.0327	0.0250
$\widehat{\phi}_{PN,j}$	SD	250	0.0076	0.0109	0.0072	0.0075	0.0105	0.0072	0.0075	0.0101	0.0071
	MAD	250	0.0700	0.1208	0.0829	0.0695	0.1176	0.0831	0.0690	0.1153	0.0859
	M -ESCALA	250	0.0231	0.0404	0.0263	0.0229	0.0387	0.0259	0.0228	0.0368	0.0255

Tabla 3.2: Valores medios de $\|\widehat{\phi}_j/\|\widehat{\phi}_j\| - \phi_j\|^2$ bajo C_0 cuando $\rho = an^{-3}$ y $r = an^{-3}$ para distintos tamaños m de la grilla.

Método	Estimador de escala	m	$\alpha = 0.15$			$\alpha = 0.5$			$\alpha = 1$		
			$j = 1$	$j = 2$	$j = 3$	$j = 1$	$j = 2$	$j = 3$	$j = 1$	$j = 2$	$j = 3$
$\hat{\phi}_{PS,j}$	SD	50	0.0080	0.0117	0.0100	0.0080	0.0116	0.0100	0.0080	0.0116	0.0100
	MAD	50	0.0744	0.1288	0.0879	0.0743	0.1281	0.0872	0.0743	0.1278	0.0869
	M -ESCALA	50	0.0243	0.0423	0.0294	0.0243	0.0422	0.0294	0.0243	0.0420	0.0291
$\hat{\phi}_{PS,j}$	SD	100	0.0078	0.0113	0.0079	0.0078	0.0113	0.0079	0.0077	0.0112	0.0078
	MAD	100	0.0700	0.1211	0.0825	0.0700	0.1211	0.0825	0.0703	0.1212	0.0821
	M -ESCALA	100	0.0237	0.0415	0.0271	0.0237	0.0413	0.0268	0.0237	0.0412	0.0267
$\hat{\phi}_{PS,j}$	SD	150	0.0077	0.0112	0.0075	0.0077	0.0111	0.0074	0.0077	0.0111	0.0073
	MAD	150	0.0703	0.1214	0.0822	0.0704	0.1213	0.0821	0.0703	0.1210	0.0819
	M -ESCALA	150	0.0234	0.0413	0.0267	0.0234	0.0411	0.0265	0.0234	0.0408	0.0261
$\hat{\phi}_{PS,j}$	SD	200	0.0077	0.0112	0.0073	0.0077	0.0111	0.0073	0.0076	0.0110	0.0072
	MAD	200	0.0705	0.1221	0.0823	0.0705	0.1221	0.0823	0.0705	0.1221	0.0823
	M -ESCALA	200	0.0233	0.0415	0.0268	0.0233	0.0410	0.0263	0.0233	0.0405	0.0258
$\hat{\phi}_{PS,j}$	SD	250	0.0076	0.0111	0.0072	0.0076	0.0111	0.0072	0.0076	0.0109	0.0070
	MAD	250	0.0701	0.1210	0.0812	0.0701	0.1205	0.0807	0.0701	0.1204	0.0807
	M -ESCALA	250	0.0233	0.0413	0.0265	0.0233	0.0410	0.0263	0.0232	0.0402	0.0255
$\hat{\phi}_{PN,j}$	SD	50	0.0079	0.0113	0.0100	0.0078	0.0106	0.0098	0.0078	0.0098	0.0096
	MAD	50	0.0737	0.1262	0.0885	0.0732	0.1234	0.0927	0.0720	0.1176	0.0965
	M -ESCALA	50	0.0240	0.0407	0.0291	0.0239	0.0384	0.0288	0.0233	0.0350	0.0281
$\hat{\phi}_{PN,j}$	SD	100	0.0077	0.0113	0.0079	0.0077	0.0111	0.0079	0.0077	0.0109	0.0078
	MAD	100	0.0702	0.1215	0.0829	0.0701	0.1212	0.0839	0.0699	0.1195	0.0838
	M -ESCALA	100	0.0237	0.0414	0.0271	0.0236	0.0409	0.0271	0.0235	0.0402	0.0270
$\hat{\phi}_{PN,j}$	SD	150	0.0077	0.0112	0.0075	0.0077	0.0111	0.0075	0.0077	0.0111	0.0075
	MAD	150	0.0704	0.1213	0.0822	0.0704	0.1213	0.0825	0.0703	0.1214	0.0827
	M -ESCALA	150	0.0234	0.0414	0.0268	0.0234	0.0413	0.0268	0.0234	0.0412	0.0268
$\hat{\phi}_{PN,j}$	SD	200	0.0077	0.0112	0.0073	0.0077	0.0111	0.0073	0.0077	0.0111	0.0073
	MAD	200	0.0705	0.1223	0.0826	0.0705	0.1224	0.0827	0.0705	0.1226	0.0830
	M -ESCALA	200	0.0233	0.0416	0.0269	0.0233	0.0416	0.0269	0.0233	0.0415	0.0269
$\hat{\phi}_{PN,j}$	SD	250	0.0077	0.0112	0.0073	0.0076	0.0111	0.0073	0.0076	0.0111	0.0073
	MAD	250	0.0701	0.1212	0.0814	0.0701	0.1212	0.0815	0.0701	0.1211	0.0815
	M -ESCALA	250	0.0233	0.0414	0.0267	0.0233	0.0414	0.0267	0.0233	0.0414	0.0267

Tabla 3.3: Valores medios de $\|\hat{\phi}_j/\|\hat{\phi}_j\| - \phi_j\|^2$ bajo C_0 cuando $\rho = an^{-4}$ y $r = an^{-4}$, para distintos tamaños m de la grilla.

Estimador de escala	Modelo	$j = 1$	$j = 2$	$j = 3$	Modelo	$j = 1$	$j = 2$	$j = 3$
SD	C_0	0.0080	0.0117	0.0100	$C_{3,b}$	0.0254	1.6314	1.6554
MAD		0.0744	0.1288	0.0879		0.1183	0.6177	0.5971
M -ESCALA		0.0243	0.0424	0.0295		0.0730	0.6274	0.6346
SD	C_2	1.2308	1.2307	0.0040	C_{23}	1.7825	0.3857	1.7563
MAD		0.3730	0.4016	0.0638		0.2590	0.4221	0.2847
M -ESCALA		0.4231	0.4271	0.0173		0.2879	0.4655	0.3053
SD	$C_{3,a}$	1.7977	1.8885	1.9139	C_{Cauchy}	0.3071	0.4659	0.2331
MAD		0.2729	0.8004	0.7922		0.0854	0.1592	0.1100
M -ESCALA		0.3014	0.9660	0.9849		0.0502	0.0850	0.0542

Tabla 3.4: Valores medios de $\|\hat{\phi}_j - \phi_j\|^2$ para los estimadores “crudos”.

Método	Estimador de Escala	a	$\alpha = 3$			$\alpha = 4$		
			$j = 1$	$j = 2$	$j = 3$	$j = 1$	$j = 2$	$j = 3$
$\widehat{\phi}_{\text{RAW},j}$	SD		0.0080	0.0117	0.0100	0.0080	0.0117	0.0100
	MAD		0.0744	0.1288	0.0879	0.0744	0.1288	0.0879
	M -ESCALA		0.0243	0.0424	0.0295	0.0243	0.0424	0.0295
$\widehat{\phi}_{\text{PS},j}$	SD	0.05	0.0079	0.0113	0.0097	0.0080	0.0117	0.0100
	MAD	0.05	0.0739	0.1259	0.0848	0.0744	0.1288	0.0879
	M -ESCALA	0.05	0.0242	0.0405	0.0277	0.0243	0.0424	0.0295
$\widehat{\phi}_{\text{PS},j}$	SD	0.10	0.0079	0.0109	0.0093	0.0080	0.0117	0.0100
	MAD	0.10	0.0739	0.1233	0.0823	0.0744	0.1288	0.0879
	M -ESCALA	0.10	0.0241	0.0390	0.0262	0.0243	0.0423	0.0295
$\widehat{\phi}_{\text{PS},j}$	SD	0.15	0.0078	0.0106	0.0090	0.0080	0.0117	0.0100
	MAD	0.15	0.0737	0.1187	0.0780	0.0744	0.1288	0.0879
	M -ESCALA	0.15	0.0240	0.0377	0.0249	0.0243	0.0423	0.0294
$\widehat{\phi}_{\text{PS},j}$	SD	0.25	0.0078	0.0099	0.0082	0.0080	0.0117	0.0100
	MAD	0.25	0.0730	0.1145	0.0744	0.0744	0.1288	0.0879
	M -ESCALA	0.25	0.0239	0.0353	0.0224	0.0243	0.0423	0.0294
$\widehat{\phi}_{\text{PS},j}$	SD	0.5	0.0077	0.0090	0.0074	0.0080	0.0116	0.0100
	MAD	0.5	0.0720	0.1061	0.0663	0.0743	0.1281	0.0872
	M -ESCALA	0.5	0.0239	0.0317	0.0187	0.0243	0.0422	0.0294
$\widehat{\phi}_{\text{PS},j}$	SD	0.75	0.0077	0.0084	0.0067	0.0080	0.0116	0.0100
	MAD	0.75	0.0710	0.0982	0.0588	0.0743	0.1280	0.0871
	M -ESCALA	0.75	0.0234	0.0287	0.0155	0.0243	0.0422	0.0293
$\widehat{\phi}_{\text{PS},j}$	SD	1	0.0077	0.0081	0.0064	0.0080	0.0116	0.0100
	MAD	1	0.0702	0.0929	0.0531	0.0743	0.1278	0.0869
	M -ESCALA	1	0.0232	0.0270	0.0136	0.0243	0.0420	0.0291
$\widehat{\phi}_{\text{PS},j}$	SD	1.5	0.0075	0.0075	0.0058	0.0080	0.0116	0.0099
	MAD	1.5	0.0689	0.0845	0.0449	0.0742	0.1275	0.0866
	M -ESCALA	1.5	0.0228	0.0246	0.0112	0.0243	0.0416	0.0287
$\widehat{\phi}_{\text{PS},j}$	SD	2	0.0074	0.0071	0.0053	0.0080	0.0115	0.0099
	MAD	2	0.0688	0.0797	0.0391	0.0742	0.1272	0.0863
	M -ESCALA	2	0.0222	0.0229	0.0094	0.0243	0.0413	0.0284
$\widehat{\phi}_{\text{PN},j}$	SD	0.05	0.0076	0.0080	0.0103	0.0079	0.0116	0.0100
	MAD	0.05	0.0660	0.0911	0.1130	0.0741	0.1273	0.0875
	M -ESCALA	0.05	0.0214	0.0254	0.0265	0.0242	0.0417	0.0293
$\widehat{\phi}_{\text{PN},j}$	SD	0.10	0.0074	0.0074	0.0128	0.0079	0.0115	0.0100
	MAD	0.10	0.0644	0.0801	0.1321	0.0739	0.1267	0.0880
	M -ESCALA	0.10	0.0209	0.0228	0.0285	0.0242	0.0412	0.0291
$\widehat{\phi}_{\text{PN},j}$	SD	0.15	0.0075	0.0075	0.0161	0.0079	0.0113	0.0100
	MAD	0.15	0.0619	0.0731	0.1465	0.0737	0.1262	0.0885
	M -ESCALA	0.15	0.0203	0.0216	0.0310	0.0240	0.0407	0.0291
$\widehat{\phi}_{\text{PN},j}$	SD	0.25	0.0080	0.0081	0.0239	0.0078	0.0111	0.0100
	MAD	0.25	0.0583	0.0678	0.1848	0.0735	0.1249	0.0892
	M -ESCALA	0.25	0.0198	0.0210	0.0410	0.0240	0.0401	0.0290
$\widehat{\phi}_{\text{PN},j}$	SD	0.5	0.0087	0.0095	0.0490	0.0078	0.0106	0.0098
	MAD	0.5	0.0552	0.0650	0.2687	0.0732	0.1234	0.0927
	M -ESCALA	0.5	0.0193	0.0213	0.0715	0.0239	0.0384	0.0288
$\widehat{\phi}_{\text{PN},j}$	SD	0.75	0.0089	0.0103	0.0834	0.0078	0.0101	0.0096
	MAD	0.75	0.0526	0.0651	0.3458	0.0728	0.1200	0.0939
	M -ESCALA	0.75	0.0190	0.0220	0.1081	0.0235	0.0366	0.0284
$\widehat{\phi}_{\text{PN},j}$	SD	1	0.0093	0.0113	0.1197	0.0078	0.0098	0.0096
	MAD	1	0.0511	0.0658	0.4073	0.0720	0.1176	0.0965
	M -ESCALA	1	0.0192	0.0233	0.1470	0.0233	0.0350	0.0281
$\widehat{\phi}_{\text{PN},j}$	SD	1.5	0.0100	0.0134	0.1905	0.0078	0.0094	0.0097
	MAD	1.5	0.0462	0.0648	0.4897	0.0714	0.1107	0.0960
	M -ESCALA	1.5	0.0190	0.0255	0.2241	0.0230	0.0327	0.0275
$\widehat{\phi}_{\text{PN},j}$	SD	2	0.0109	0.0160	0.2608	0.0076	0.0089	0.0096
	MAD	2	0.0440	0.0677	0.5736	0.0704	0.1053	0.0971
	M -ESCALA	2	0.0184	0.0271	0.2990	0.0228	0.0312	0.0273

Tabla 3.5: Valores medios de $\|\widehat{\phi}_j/\widehat{\phi}_j\| - \phi_j\|^2$, bajo C_0 cuando ρ o $r = an^{-\alpha}$.

Método	Estimador de escala	a	$\alpha = 3$			$\alpha = 4$		
			$j = 1$	$j = 2$	$j = 3$	$j = 1$	$j = 2$	$j = 3$
$\widehat{\phi}_{RAW,j}$	SD		1.2308	1.2307	0.0040	1.2308	1.2307	0.0040
	MAD		0.3730	0.4016	0.0638	0.3730	0.4016	0.0638
	M -ESCALA		0.4231	0.4271	0.0173	0.4231	0.4271	0.0173
$\widehat{\phi}_{PS,j}$	SD	0.05	1.2301	1.2300	0.0039	1.2308	1.2307	0.0040
	MAD	0.05	0.3730	0.3998	0.0619	0.3730	0.4016	0.0638
	M -ESCALA	0.05	0.4229	0.4266	0.0169	0.4231	0.4271	0.0173
$\widehat{\phi}_{PS,j}$	SD	0.10	1.2296	1.2295	0.0039	1.2308	1.2307	0.0040
	MAD	0.10	0.3725	0.3992	0.0616	0.3730	0.4016	0.0638
	M -ESCALA	0.10	0.4228	0.4264	0.0168	0.4231	0.4271	0.0173
$\widehat{\phi}_{PS,j}$	SD	0.15	1.2289	1.2288	0.0039	1.2308	1.2307	0.0040
	MAD	0.15	0.3722	0.3981	0.0608	0.3730	0.4016	0.0638
	M -ESCALA	0.15	0.4219	0.4254	0.0166	0.4231	0.4271	0.0173
$\widehat{\phi}_{PS,j}$	SD	0.25	1.2271	1.2271	0.0039	1.2308	1.2307	0.0040
	MAD	0.25	0.3720	0.3963	0.0589	0.3730	0.4016	0.0638
	M -ESCALA	0.25	0.4219	0.4249	0.0162	0.4231	0.4271	0.0173
$\widehat{\phi}_{PS,j}$	SD	0.5	1.2228	1.2227	0.0038	1.2308	1.2307	0.0040
	MAD	0.5	0.3706	0.3906	0.0541	0.3730	0.4016	0.0638
	M -ESCALA	0.5	0.4212	0.4233	0.0151	0.4231	0.4270	0.0173
$\widehat{\phi}_{PS,j}$	SD	0.75	1.2177	1.2175	0.0037	1.2306	1.2305	0.0040
	MAD	0.75	0.3687	0.3867	0.0507	0.3730	0.4015	0.0638
	M -ESCALA	0.75	0.4201	0.4219	0.0143	0.4231	0.4270	0.0173
$\widehat{\phi}_{PS,j}$	SD	1	1.2122	1.2120	0.0037	1.2303	1.2303	0.0040
	MAD	1	0.3684	0.3852	0.0490	0.3730	0.4014	0.0637
	M -ESCALA	1	0.4190	0.4204	0.0137	0.4231	0.4270	0.0173
$\widehat{\phi}_{PS,j}$	SD	1.5	1.2017	1.2015	0.0036	1.2302	1.2301	0.0039
	MAD	1.5	0.3662	0.3804	0.0445	0.3730	0.4013	0.0636
	M -ESCALA	1.5	0.4154	0.4161	0.0125	0.4231	0.4270	0.0173
$\widehat{\phi}_{PS,j}$	SD	2	1.1921	1.1919	0.0036	1.2302	1.2301	0.0039
	MAD	2	0.3631	0.3739	0.0403	0.3730	0.4013	0.0636
	M -ESCALA	2	0.4131	0.4136	0.0117	0.4230	0.4269	0.0172
$\widehat{\phi}_{PN,j}$	SD	0.05	1.2177	1.2198	0.0056	1.2303	1.2303	0.0040
	MAD	0.05	0.3666	0.3820	0.0837	0.3730	0.4000	0.0629
	M -ESCALA	0.05	0.4170	0.4206	0.0226	0.4229	0.4269	0.0174
$\widehat{\phi}_{PN,j}$	SD	0.10	1.2078	1.2123	0.0074	1.2303	1.2303	0.0040
	MAD	0.10	0.3637	0.3789	0.1088	0.3726	0.3990	0.0629
	M -ESCALA	0.10	0.4128	0.4184	0.0293	0.4228	0.4265	0.0173
$\widehat{\phi}_{PN,j}$	SD	0.15	1.1934	1.2002	0.0095	1.2303	1.2303	0.0040
	MAD	0.15	0.3605	0.3752	0.1297	0.3723	0.3983	0.0632
	M -ESCALA	0.15	0.4092	0.4173	0.0368	0.4226	0.4262	0.0174
$\widehat{\phi}_{PN,j}$	SD	0.25	1.1736	1.1850	0.0148	1.2301	1.2302	0.0040
	MAD	0.25	0.3560	0.3736	0.1657	0.3720	0.3966	0.0625
	M -ESCALA	0.25	0.4042	0.4174	0.0513	0.4226	0.4260	0.0175
$\widehat{\phi}_{PN,j}$	SD	0.5	1.1092	1.1319	0.0307	1.2288	1.2289	0.0041
	MAD	0.5	0.3379	0.3654	0.2499	0.3719	0.3959	0.0637
	M -ESCALA	0.5	0.3865	0.4121	0.0979	0.4220	0.4248	0.0174
$\widehat{\phi}_{PN,j}$	SD	0.75	1.0430	1.0767	0.0470	1.2277	1.2279	0.0041
	MAD	0.75	0.3254	0.3633	0.3172	0.3724	0.3947	0.0645
	M -ESCALA	0.75	0.3707	0.4073	0.1436	0.4220	0.4249	0.0181
$\widehat{\phi}_{PN,j}$	SD	1	0.9727	1.0174	0.0653	1.2279	1.2282	0.0042
	MAD	1	0.3111	0.3586	0.3807	0.3726	0.3950	0.0674
	M -ESCALA	1	0.3531	0.3996	0.1832	0.4218	0.4246	0.0185
$\widehat{\phi}_{PN,j}$	SD	1.5	0.8598	0.9248	0.1084	1.2263	1.2269	0.0044
	MAD	1.5	0.2876	0.3521	0.4924	0.3723	0.3925	0.0690
	M -ESCALA	1.5	0.3253	0.3895	0.2613	0.4227	0.4254	0.0186
$\widehat{\phi}_{PN,j}$	SD	2	0.7401	0.8205	0.1525	1.2264	1.2271	0.0045
	MAD	2	0.2664	0.3454	0.5727	0.3703	0.3922	0.0736
	M -ESCALA	2	0.2987	0.3780	0.3363	0.4205	0.4234	0.0192

Tabla 3.6: Valores medios de $\|\widehat{\phi}_j/\|\widehat{\phi}_j\| - \phi_j\|^2$, bajo C_2 cuando ρ o $r = an^{-\alpha}$.

Método	Estimador de escala	a	$\alpha = 3$			$\alpha = 4$		
			$j = 1$	$j = 2$	$j = 3$	$j = 1$	$j = 2$	$j = 3$
$\hat{\phi}_{RAW,j}$	SD		1.7977	1.8885	1.9139	1.7977	1.8885	1.9139
	MAD		0.2729	0.8004	0.7922	0.2729	0.8004	0.7922
	M -ESCALA		0.3014	0.9660	0.9849	0.3014	0.9660	0.9849
$\hat{\phi}_{PS,j}$	SD	0.05	1.7976	1.8886	1.9136	1.7977	1.8885	1.9139
	MAD	0.05	0.2714	0.7840	0.7761	0.2729	0.8004	0.7922
	M -ESCALA	0.05	0.3000	0.9555	0.9746	0.3014	0.9659	0.9849
$\hat{\phi}_{PS,j}$	SD	0.10	1.7967	1.8888	1.9135	1.7977	1.8885	1.9139
	MAD	0.10	0.2707	0.7763	0.7697	0.2729	0.8004	0.7922
	M -ESCALA	0.10	0.2990	0.9444	0.9656	0.3014	0.9657	0.9847
$\hat{\phi}_{PS,j}$	SD	0.15	1.7956	1.8890	1.9136	1.7977	1.8885	1.9139
	MAD	0.15	0.2689	0.7668	0.7609	0.2729	0.8004	0.7922
	M -ESCALA	0.15	0.2982	0.9384	0.9598	0.3014	0.9653	0.9844
$\hat{\phi}_{PS,j}$	SD	0.25	1.7930	1.8889	1.9128	1.7977	1.8885	1.9139
	MAD	0.25	0.2660	0.7449	0.7400	0.2729	0.8004	0.7922
	M -ESCALA	0.25	0.2964	0.9235	0.9461	0.3014	0.9653	0.9844
$\hat{\phi}_{PS,j}$	SD	0.5	1.7872	1.8896	1.9117	1.7977	1.8885	1.9139
	MAD	0.5	0.2622	0.7069	0.7063	0.2729	0.7999	0.7913
	M -ESCALA	0.5	0.2912	0.8651	0.8933	0.3014	0.9652	0.9843
$\hat{\phi}_{PS,j}$	SD	0.75	1.7823	1.8900	1.9112	1.7977	1.8885	1.9139
	MAD	0.75	0.2577	0.6645	0.6678	0.2721	0.7962	0.7884
	M -ESCALA	0.75	0.2851	0.8120	0.8457	0.3014	0.9652	0.9843
$\hat{\phi}_{PS,j}$	SD	1	1.7752	1.8904	1.9096	1.7977	1.8885	1.9139
	MAD	1	0.2555	0.6402	0.6456	0.2721	0.7941	0.7867
	M -ESCALA	1	0.2806	0.7728	0.8132	0.3011	0.9632	0.9823
$\hat{\phi}_{PS,j}$	SD	1.5	1.7612	1.8906	1.9071	1.7977	1.8886	1.9139
	MAD	1.5	0.2469	0.5819	0.5936	0.2721	0.7928	0.7854
	M -ESCALA	1.5	0.2701	0.6999	0.7479	0.3009	0.9613	0.9806
$\hat{\phi}_{PS,j}$	SD	2	1.7394	1.8903	1.9067	1.7977	1.8885	1.9139
	MAD	2	0.2365	0.5259	0.5408	0.2721	0.7925	0.7848
	M -ESCALA	2	0.2570	0.6314	0.6856	0.3005	0.9587	0.9778
$\hat{\phi}_{PN,j}$	SD	0.05	1.1991	1.8687	1.9247	1.7961	1.8884	1.9144
	MAD	0.05	0.1654	0.6677	0.9259	0.2712	0.7955	0.7911
	M -ESCALA	0.05	0.1720	0.8166	1.0945	0.2997	0.9664	0.9886
$\hat{\phi}_{PN,j}$	SD	0.10	0.3089	1.7563	1.9172	1.7929	1.8881	1.9146
	MAD	0.10	0.1164	0.5668	0.9717	0.2691	0.7926	0.7920
	M -ESCALA	0.10	0.0986	0.5954	1.0570	0.2988	0.9702	0.9959
$\hat{\phi}_{PN,j}$	SD	0.15	0.0959	1.6179	1.9006	1.7917	1.8882	1.9151
	MAD	0.15	0.0884	0.4555	0.9694	0.2677	0.7908	0.7944
	M -ESCALA	0.15	0.0619	0.4484	1.0327	0.2974	0.9675	0.9972
$\hat{\phi}_{PN,j}$	SD	0.25	0.0740	1.3619	1.8378	1.7863	1.8886	1.9148
	MAD	0.25	0.0648	0.3332	0.9896	0.2645	0.7836	0.7952
	M -ESCALA	0.25	0.0364	0.2517	0.9864	0.2946	0.9643	1.0008
$\hat{\phi}_{PN,j}$	SD	0.5	0.0812	1.0008	1.6349	1.7720	1.8890	1.9165
	MAD	0.5	0.0526	0.2080	0.9771	0.2593	0.7855	0.8131
	M -ESCALA	0.5	0.0276	0.1046	0.9265	0.2873	0.9660	1.0190
$\hat{\phi}_{PN,j}$	SD	0.75	0.0882	0.9942	1.6186	1.7586	1.8885	1.9184
	MAD	0.75	0.0468	0.1606	0.9668	0.2533	0.7831	0.8254
	M -ESCALA	0.75	0.0271	0.0631	0.8130	0.2783	0.9566	1.0271
$\hat{\phi}_{PN,j}$	SD	1	0.0921	1.0504	1.6565	1.7354	1.8887	1.9200
	MAD	1	0.0457	0.1375	0.9365	0.2467	0.7826	0.8418
	M -ESCALA	1	0.0267	0.0527	0.7608	0.2721	0.9635	1.0493
$\hat{\phi}_{PN,j}$	SD	1.5	0.0980	1.1618	1.7271	1.6922	1.8870	1.9218
	MAD	1.5	0.0456	0.1120	0.9501	0.2338	0.7712	0.8603
	M -ESCALA	1.5	0.0266	0.0507	0.7417	0.2553	0.9394	1.0591
$\hat{\phi}_{PN,j}$	SD	2	0.1031	1.2519	1.7678	1.6482	1.8871	1.9239
	MAD	2	0.0446	0.1152	0.9982	0.2216	0.7533	0.8699
	M -ESCALA	2	0.0266	0.0552	0.7756	0.2401	0.9237	1.0732

Tabla 3.7: Valores medios de $\|\hat{\phi}_j/\|\hat{\phi}_j\| - \phi_j\|^2$, bajo $C_{3,a}$ cuando ρ o $r = an^{-\alpha}$.

Método	Estimador de escala	a	$\alpha = 3$			$\alpha = 4$		
			$j = 1$	$j = 2$	$j = 3$	$j = 1$	$j = 2$	$j = 3$
$\hat{\phi}_{RAW,j}$	SD		0.0254	1.6314	1.6554	0.0254	1.6314	1.6554
	MAD		0.1183	0.6177	0.5971	0.1183	0.6177	0.5971
	M -ESCALA		0.0730	0.6274	0.6346	0.0730	0.6274	0.6346
$\hat{\phi}_{PS,j}$	SD	0.05	0.0251	1.6117	1.6360	0.0254	1.6314	1.6553
	MAD	0.05	0.1180	0.6116	0.5903	0.1183	0.6177	0.5971
	M -ESCALA	0.05	0.0717	0.6097	0.6167	0.0730	0.6274	0.6346
$\hat{\phi}_{PS,j}$	SD	0.10	0.0247	1.5937	1.6182	0.0254	1.6313	1.6553
	MAD	0.10	0.1177	0.6022	0.5808	0.1183	0.6177	0.5971
	M -ESCALA	0.10	0.0703	0.5947	0.6019	0.0730	0.6270	0.6342
$\hat{\phi}_{PS,j}$	SD	0.15	0.0244	1.5743	1.5990	0.0254	1.6312	1.6551
	MAD	0.15	0.1171	0.5936	0.5724	0.1183	0.6177	0.5971
	M -ESCALA	0.15	0.0685	0.5785	0.5867	0.0730	0.6268	0.6340
$\hat{\phi}_{PS,j}$	SD	0.25	0.0235	1.5300	1.5559	0.0254	1.6305	1.6545
	MAD	0.25	0.1156	0.5622	0.5418	0.1183	0.6177	0.5971
	M -ESCALA	0.25	0.0667	0.5488	0.5580	0.0730	0.6266	0.6338
$\hat{\phi}_{PS,j}$	SD	0.5	0.0222	1.3781	1.4045	0.0254	1.6296	1.6537
	MAD	0.5	0.1109	0.5045	0.4840	0.1183	0.6177	0.5971
	M -ESCALA	0.5	0.0603	0.4819	0.4909	0.0730	0.6257	0.6330
$\hat{\phi}_{PS,j}$	SD	0.75	0.0209	1.1843	1.2106	0.0254	1.6288	1.6529
	MAD	0.75	0.1083	0.4522	0.4347	0.1183	0.6174	0.5969
	M -ESCALA	0.75	0.0570	0.4196	0.4296	0.0729	0.6248	0.6320
$\hat{\phi}_{PS,j}$	SD	1	0.0198	0.9222	0.9475	0.0254	1.6276	1.6516
	MAD	1	0.1033	0.4005	0.3844	0.1183	0.6168	0.5962
	M -ESCALA	1	0.0532	0.3641	0.3742	0.0729	0.6244	0.6316
$\hat{\phi}_{PS,j}$	SD	1.5	0.0183	0.4274	0.4473	0.0253	1.6253	1.6493
	MAD	1.5	0.0950	0.3119	0.2947	0.1183	0.6150	0.5945
	M -ESCALA	1.5	0.0452	0.2371	0.2462	0.0729	0.6233	0.6304
$\hat{\phi}_{PS,j}$	SD	2	0.0173	0.1491	0.1640	0.0252	1.6231	1.6472
	MAD	2	0.0891	0.2514	0.2368	0.1183	0.6150	0.5945
	M -ESCALA	2	0.0413	0.1650	0.1714	0.0729	0.6231	0.6302
$\hat{\phi}_{PN,j}$	SD	0.05	0.0141	0.3471	0.5910	0.0250	1.6211	1.6480
	MAD	0.05	0.0750	0.3075	0.5051	0.1180	0.6150	0.5974
	M -ESCALA	0.05	0.0318	0.2159	0.4374	0.0717	0.6201	0.6321
$\hat{\phi}_{PN,j}$	SD	0.10	0.0131	0.0397	0.1720	0.0246	1.6136	1.6429
	MAD	0.10	0.0644	0.1890	0.4423	0.1177	0.6148	0.6011
	M -ESCALA	0.10	0.0264	0.0926	0.3101	0.0709	0.6125	0.6291
$\hat{\phi}_{PN,j}$	SD	0.15	0.0135	0.0206	0.1199	0.0244	1.6049	1.6370
	MAD	0.15	0.0617	0.1394	0.4084	0.1169	0.6121	0.6025
	M -ESCALA	0.15	0.0254	0.0516	0.2415	0.0686	0.6048	0.6273
$\hat{\phi}_{PN,j}$	SD	0.25	0.0150	0.0177	0.1176	0.0232	1.5875	1.6256
	MAD	0.25	0.0576	0.0988	0.3934	0.1146	0.5981	0.5954
	M -ESCALA	0.25	0.0254	0.0350	0.2000	0.0667	0.5938	0.6255
$\hat{\phi}_{PN,j}$	SD	0.5	0.0170	0.0207	0.1981	0.0220	1.5442	1.5978
	MAD	0.5	0.0542	0.0748	0.4397	0.1101	0.5758	0.5913
	M -ESCALA	0.5	0.0254	0.0304	0.1958	0.0610	0.5627	0.6147
$\hat{\phi}_{PN,j}$	SD	0.75	0.0191	0.0246	0.2803	0.0206	1.4916	1.5622
	MAD	0.75	0.0532	0.0707	0.4945	0.1078	0.5644	0.5967
	M -ESCALA	0.75	0.0247	0.0302	0.2358	0.0571	0.5427	0.6149
$\hat{\phi}_{PN,j}$	SD	1	0.0205	0.0284	0.3638	0.0197	1.4343	1.5216
	MAD	1	0.0520	0.0737	0.5518	0.1030	0.5395	0.5857
	M -ESCALA	1	0.0242	0.0308	0.2833	0.0534	0.5158	0.6054
$\hat{\phi}_{PN,j}$	SD	1.5	0.0229	0.0399	0.5143	0.0180	1.3033	1.4270
	MAD	1.5	0.0493	0.0745	0.6525	0.0954	0.4867	0.5626
	M -ESCALA	1.5	0.0241	0.0339	0.3798	0.0471	0.4621	0.5835
$\hat{\phi}_{PN,j}$	SD	2	0.0240	0.0486	0.6355	0.0172	1.1662	1.3250
	MAD	2	0.0473	0.0822	0.7511	0.0892	0.4435	0.5449
	M -ESCALA	2	0.0235	0.0369	0.4719	0.0428	0.4154	0.5639

Tabla 3.8: Valores medios de $\|\hat{\phi}_j/\|\hat{\phi}_j\| - \phi_j\|^2$, bajo $C_{3,b}$ cuando ρ o $r = an^{-\alpha}$.

Método	Estimador de escala	a	$\alpha = 3$			$\alpha = 4$		
			$j = 1$	$j = 2$	$j = 3$	$j = 1$	$j = 2$	$j = 3$
$\hat{\phi}_{RAW,j}$	SD		1.7825	0.3857	1.7563	1.7825	0.3857	1.7563
	MAD		0.2590	0.4221	0.2847	0.2590	0.4221	0.2847
	M -ESCALA		0.2879	0.4655	0.3053	0.2879	0.4655	0.3053
$\hat{\phi}_{PS,j}$	SD	0.05	1.7824	0.3887	1.7550	1.7825	0.3857	1.7563
	MAD	0.05	0.2587	0.4195	0.2823	0.2590	0.4220	0.2845
	M -ESCALA	0.05	0.2875	0.4628	0.3029	0.2879	0.4655	0.3052
$\hat{\phi}_{PS,j}$	SD	0.10	1.7804	0.3911	1.7529	1.7825	0.3857	1.7563
	MAD	0.10	0.2577	0.4131	0.2770	0.2590	0.4220	0.2845
	M -ESCALA	0.10	0.2872	0.4592	0.2998	0.2879	0.4655	0.3052
$\hat{\phi}_{PS,j}$	SD	0.15	1.7791	0.3953	1.7518	1.7824	0.3874	1.7564
	MAD	0.15	0.2563	0.4092	0.2735	0.2591	0.4223	0.2844
	M -ESCALA	0.15	0.2860	0.4555	0.2966	0.2879	0.4655	0.3052
$\hat{\phi}_{PS,j}$	SD	0.25	1.7779	0.4008	1.7490	1.7824	0.3874	1.7563
	MAD	0.25	0.2553	0.4037	0.2694	0.2591	0.4223	0.2844
	M -ESCALA	0.25	0.2843	0.4485	0.2905	0.2879	0.4655	0.3052
$\hat{\phi}_{PS,j}$	SD	0.5	1.7723	0.4196	1.7363	1.7824	0.3875	1.7563
	MAD	0.5	0.2530	0.3872	0.2546	0.2591	0.4223	0.2844
	M -ESCALA	0.5	0.2823	0.4378	0.2813	0.2879	0.4649	0.3047
$\hat{\phi}_{PS,j}$	SD	0.75	1.7663	0.4342	1.7215	1.7825	0.3877	1.7563
	MAD	0.75	0.2517	0.3775	0.2454	0.2591	0.4223	0.2844
	M -ESCALA	0.75	0.2797	0.4242	0.2694	0.2879	0.4648	0.3046
$\hat{\phi}_{PS,j}$	SD	1	1.7632	0.4571	1.7112	1.7825	0.3878	1.7562
	MAD	1	0.2496	0.3649	0.2342	0.2591	0.4222	0.2843
	M -ESCALA	1	0.2771	0.4138	0.2611	0.2879	0.4646	0.3044
$\hat{\phi}_{PS,j}$	SD	1.5	1.7526	0.4907	1.6835	1.7825	0.3878	1.7562
	MAD	1.5	0.2435	0.3419	0.2137	0.2591	0.4222	0.2843
	M -ESCALA	1.5	0.2728	0.3959	0.2455	0.2879	0.4644	0.3042
$\hat{\phi}_{PS,j}$	SD	2	1.7408	0.5360	1.6523	1.7823	0.3881	1.7559
	MAD	2	0.2394	0.3232	0.1963	0.2591	0.4216	0.2837
	M -ESCALA	2	0.2669	0.3777	0.2299	0.2878	0.4641	0.3039
$\hat{\phi}_{PN,j}$	SD	0.05	1.5841	1.0675	1.3787	1.7794	0.3924	1.7554
	MAD	0.05	0.2120	0.3628	0.4227	0.2583	0.4222	0.2872
	M -ESCALA	0.05	0.2361	0.4120	0.4700	0.2875	0.4657	0.3085
$\hat{\phi}_{PN,j}$	SD	0.10	1.4791	1.3522	0.8009	1.7785	0.3951	1.7542
	MAD	0.10	0.1927	0.3281	0.5028	0.2579	0.4224	0.2892
	M -ESCALA	0.10	0.2114	0.3694	0.5512	0.2864	0.4653	0.3111
$\hat{\phi}_{PN,j}$	SD	0.15	1.4571	1.4101	0.4440	1.7745	0.4018	1.7516
	MAD	0.15	0.1813	0.3015	0.5521	0.2572	0.4227	0.2928
	M -ESCALA	0.15	0.1932	0.3242	0.5842	0.2861	0.4645	0.3126
$\hat{\phi}_{PN,j}$	SD	0.25	1.4344	1.4216	0.3038	1.7701	0.4189	1.7453
	MAD	0.25	0.1642	0.2648	0.6089	0.2550	0.4167	0.2926
	M -ESCALA	0.25	0.1713	0.2642	0.6139	0.2839	0.4628	0.3172
$\hat{\phi}_{PN,j}$	SD	0.5	1.3906	1.3926	0.4061	1.7568	0.4502	1.7320
	MAD	0.5	0.1523	0.2298	0.6664	0.2518	0.4160	0.3055
	M -ESCALA	0.5	0.1540	0.2113	0.6139	0.2816	0.4595	0.3275
$\hat{\phi}_{PN,j}$	SD	0.75	1.3716	1.3846	0.5333	1.7446	0.4832	1.7208
	MAD	0.75	0.1429	0.2126	0.7109	0.2496	0.4077	0.3094
	M -ESCALA	0.75	0.1463	0.1951	0.5917	0.2773	0.4576	0.3400
$\hat{\phi}_{PN,j}$	SD	1	1.3303	1.3475	0.6386	1.7347	0.5234	1.7066
	MAD	1	0.1354	0.2025	0.7465	0.2462	0.4032	0.3156
	M -ESCALA	1	0.1387	0.1837	0.5833	0.2751	0.4572	0.3516
$\hat{\phi}_{PN,j}$	SD	1.5	1.2728	1.3010	0.8004	1.7113	0.5988	1.6797
	MAD	1.5	0.1261	0.1933	0.7833	0.2396	0.3920	0.3269
	M -ESCALA	1.5	0.1299	0.1827	0.6111	0.2691	0.4540	0.3747
$\hat{\phi}_{PN,j}$	SD	2	1.2048	1.2419	0.9162	1.6962	0.6816	1.6488
	MAD	2	0.1194	0.1938	0.8525	0.2365	0.3959	0.3493
	M -ESCALA	2	0.1233	0.1845	0.6364	0.2620	0.4498	0.3917

Tabla 3.9: Valores medios de $\|\hat{\phi}_j/\hat{\phi}_j\| - \phi_j\|^2$, bajo C_{23} cuando ρ o $r = an^{-\alpha}$.

Método	Estimador de escala	a	$\alpha = 3$			$\alpha = 4$		
			$j = 1$	$j = 2$	$j = 3$	$j = 1$	$j = 2$	$j = 3$
$\hat{\phi}_{RAW,j}$	SD		0.3071	0.4659	0.2331	0.3071	0.4659	0.2331
	MAD		0.0854	0.1592	0.1100	0.0854	0.1592	0.1100
	M -ESCALA		0.0502	0.0850	0.0542	0.0502	0.0850	0.0542
$\hat{\phi}_{PS,j}$	SD	0.05	0.3071	0.4658	0.2329	0.3071	0.4659	0.2331
	MAD	0.05	0.0855	0.1590	0.1092	0.0854	0.1590	0.1098
	M -ESCALA	0.05	0.0501	0.0837	0.0529	0.0502	0.0849	0.0540
$\hat{\phi}_{PS,j}$	SD	0.10	0.3071	0.4656	0.2328	0.3071	0.4659	0.2331
	MAD	0.10	0.0850	0.1577	0.1082	0.0854	0.1590	0.1098
	M -ESCALA	0.10	0.0500	0.0818	0.0512	0.0502	0.0849	0.0540
$\hat{\phi}_{PS,j}$	SD	0.15	0.3071	0.4654	0.2326	0.3071	0.4659	0.2331
	MAD	0.15	0.0850	0.1561	0.1067	0.0854	0.1590	0.1098
	M -ESCALA	0.15	0.0497	0.0806	0.0502	0.0502	0.0849	0.0540
$\hat{\phi}_{PS,j}$	SD	0.25	0.3071	0.4655	0.2323	0.3071	0.4659	0.2331
	MAD	0.25	0.0849	0.1542	0.1049	0.0854	0.1590	0.1098
	M -ESCALA	0.25	0.0497	0.0795	0.0490	0.0502	0.0849	0.0540
$\hat{\phi}_{PS,j}$	SD	0.5	0.3069	0.4645	0.2314	0.3071	0.4659	0.2330
	MAD	0.5	0.0846	0.1469	0.0976	0.0854	0.1590	0.1098
	M -ESCALA	0.5	0.0491	0.0749	0.0444	0.0502	0.0848	0.0540
$\hat{\phi}_{PS,j}$	SD	0.75	0.3069	0.4637	0.2306	0.3071	0.4659	0.2330
	MAD	0.75	0.0837	0.1391	0.0900	0.0854	0.1590	0.1098
	M -ESCALA	0.75	0.0489	0.0724	0.0415	0.0502	0.0848	0.0539
$\hat{\phi}_{PS,j}$	SD	1	0.3069	0.4629	0.2298	0.3071	0.4659	0.2330
	MAD	1	0.0829	0.1332	0.0841	0.0854	0.1592	0.1097
	M -ESCALA	1	0.0487	0.0694	0.0384	0.0501	0.0846	0.0538
$\hat{\phi}_{PS,j}$	SD	1.5	0.3069	0.4610	0.2280	0.3071	0.4659	0.2330
	MAD	1.5	0.0824	0.1216	0.0728	0.0854	0.1590	0.1096
	M -ESCALA	1.5	0.0481	0.0639	0.0333	0.0501	0.0845	0.0537
$\hat{\phi}_{PS,j}$	SD	2	0.3069	0.4593	0.2262	0.3071	0.4659	0.2330
	MAD	2	0.0815	0.1134	0.0655	0.0855	0.1591	0.1093
	M -ESCALA	2	0.0474	0.0606	0.0300	0.0501	0.0841	0.0533
$\hat{\phi}_{PN,j}$	SD	0.05	0.2750	0.3631	0.2238	0.3066	0.4644	0.2331
	MAD	0.05	0.0723	0.1034	0.1273	0.0857	0.1589	0.1103
	M -ESCALA	0.05	0.0436	0.0550	0.0553	0.0499	0.0833	0.0532
$\hat{\phi}_{PN,j}$	SD	0.10	0.2670	0.3332	0.2298	0.3052	0.4631	0.2339
	MAD	0.10	0.0701	0.0891	0.1474	0.0849	0.1574	0.1106
	M -ESCALA	0.10	0.0420	0.0488	0.0627	0.0497	0.0828	0.0535
$\hat{\phi}_{PN,j}$	SD	0.15	0.2649	0.3258	0.2418	0.3047	0.4592	0.2321
	MAD	0.15	0.0684	0.0837	0.1693	0.0848	0.1569	0.1111
	M -ESCALA	0.15	0.0415	0.0468	0.0711	0.0493	0.0814	0.0530
$\hat{\phi}_{PN,j}$	SD	0.25	0.2544	0.3096	0.2708	0.3042	0.4575	0.2323
	MAD	0.25	0.0677	0.0808	0.2166	0.0844	0.1549	0.1118
	M -ESCALA	0.25	0.0408	0.0453	0.0841	0.0491	0.0804	0.0529
$\hat{\phi}_{PN,j}$	SD	0.5	0.2387	0.2927	0.3304	0.3020	0.4486	0.2310
	MAD	0.5	0.0665	0.0798	0.3079	0.0830	0.1503	0.1127
	M -ESCALA	0.5	0.0390	0.0445	0.1245	0.0488	0.0779	0.0535
$\hat{\phi}_{PN,j}$	SD	0.75	0.2317	0.2941	0.3921	0.2986	0.4404	0.2306
	MAD	0.75	0.0643	0.0791	0.3714	0.0830	0.1468	0.1140
	M -ESCALA	0.75	0.0356	0.0426	0.1745	0.0482	0.0755	0.0535
$\hat{\phi}_{PN,j}$	SD	1	0.2266	0.2903	0.4480	0.2962	0.4374	0.2337
	MAD	1	0.0623	0.0801	0.4391	0.0826	0.1401	0.1125
	M -ESCALA	1	0.0338	0.0421	0.2202	0.0475	0.0729	0.0533
$\hat{\phi}_{PN,j}$	SD	1.5	0.2147	0.2980	0.5370	0.2919	0.4197	0.2259
	MAD	1.5	0.0574	0.0810	0.5517	0.0802	0.1306	0.1116
	M -ESCALA	1.5	0.0313	0.0430	0.3212	0.0468	0.0692	0.0536
$\hat{\phi}_{PN,j}$	SD	2	0.1992	0.2883	0.6208	0.2880	0.4057	0.2242
	MAD	2	0.0526	0.0826	0.6515	0.0789	0.1263	0.1156
	M -ESCALA	2	0.0298	0.0449	0.4050	0.0462	0.0660	0.0533

Tabla 3.10: Valores medios de $\|\hat{\phi}_j/\|\hat{\phi}_j\| - \phi_j\|^2$, bajo C_c cuando ρ o $r = an^{-\alpha}$.

Estimador de escala	p_n	$j = 1$	$j = 2$	$j = 3$	p_n	$j = 1$	$j = 2$	$j = 3$
		Base de Fourier				Base de B -splines		
SD	10	0.0000	1.8304	1.9991	10	1.8377	1.6408	1.4733
MAD	10	0.0053	1.8307	1.9948	10	1.8073	1.6106	1.4981
M -ESCALA	10	0.0002	1.8304	1.9989	10	1.8212	1.6212	1.4769
SD	20	0.0063	0.0242	1.7979	20	1.0294	0.9978	1.1164
MAD	20	0.0707	0.0875	1.7986	20	1.0535	1.0297	1.1559
M -ESCALA	20	0.0224	0.0383	1.7979	20	1.0391	1.0091	1.1227
SD	30	0.0097	0.0484	0.0340	50	0.0111	0.0831	0.0446
MAD	30	0.0827	0.1523	0.1060	50	0.0888	0.1780	0.1163
M -ESCALA	30	0.0286	0.0773	0.0528	50	0.0316	0.1100	0.0631

Tabla 3.11: Valores medios de $\|\widehat{\phi}_{SI,j}/\|\widehat{\phi}_{SI,j}\| - \phi_j\|^2$ bajo C_0 .

Estimador de escala	p_n	$j = 1$	$j = 2$	$j = 3$	p_n	$j = 1$	$j = 2$	$j = 3$
		Base de Fourier				Base de B -splines		
SD	10	0.0002	1.8304	1.9969	10	1.9439	1.9384	1.4668
MAD	10	0.0138	1.8312	1.9923	10	1.8346	1.6160	1.4924
M -ESCALA	10	0.0007	1.8305	1.9976	10	1.8471	1.6406	1.4740
SD	20	1.3656	1.3398	1.7978	20	1.6283	1.6683	1.0875
MAD	20	0.3972	0.3817	1.7983	20	1.2734	1.1758	1.1297
M -ESCALA	20	0.4546	0.4311	1.7979	20	1.3084	1.1952	1.1008
SD	30	1.3822	1.3516	0.0275	50	1.4368	1.3927	0.0391
MAD	30	0.4186	0.4274	0.0825	50	0.4363	0.4451	0.0978
M -ESCALA	30	0.4672	0.4481	0.0402	50	0.4928	0.4684	0.0517

Tabla 3.12: Valores medios de $\|\widehat{\phi}_{SI,j}/\|\widehat{\phi}_{SI,j}\| - \phi_j\|^2$ bajo C_2 .

Estimador de escala	p_n	$j = 1$	$j = 2$	$j = 3$	p_n	$j = 1$	$j = 2$	$j = 3$
		Base de Fourier				Base de B -splines		
SD	10	0.0003	1.8304	1.9949	10	1.9032	1.5169	1.8562
MAD	10	0.0184	1.8315	1.9914	10	1.8539	1.6334	1.6979
M -ESCALA	10	0.0014	1.8305	1.9971	10	1.8759	1.6442	1.7321
SD	20	0.0102	0.0502	1.8028	20	1.8972	1.6689	1.9018
MAD	20	0.0782	0.1366	1.8056	20	1.1961	1.2319	1.3661
M -ESCALA	20	0.0306	0.0701	1.8029	20	1.2167	1.2783	1.4478
SD	30	1.8193	1.8792	1.9129	50	1.8196	1.8762	1.9134
MAD	30	0.3000	0.7820	0.7817	50	0.3069	0.7863	0.7905
M -ESCALA	30	0.3325	0.9290	0.9573	50	0.3393	0.9220	0.9560

Tabla 3.13: Valores medios de $\|\widehat{\phi}_{SI,j}/\|\widehat{\phi}_{SI,j}\| - \phi_j\|^2$ bajo $C_{3,a}$.

Estimador de escala	p_n	$j = 1$	$j = 2$	$j = 3$	p_n	$j = 1$	$j = 2$	$j = 3$
		Base de Fourier				Base de B -splines		
SD	10	0.0001	1.8304	1.9979	10	1.8540	1.6745	1.5379
MAD	10	0.0090	1.8309	1.9941	10	1.8163	1.6216	1.5648
M -ESCALA	10	0.0004	1.8304	1.9983	10	1.8208	1.6255	1.5412
SD	20	0.0068	0.0274	1.7985	20	1.0433	1.1832	1.5002
MAD	20	0.0725	0.0999	1.7999	20	1.1144	1.1855	1.3856
M -ESCALA	20	0.0268	0.0449	1.7985	20	1.0867	1.1719	1.3902
SD	30	0.0324	1.5035	1.6046	50	0.0338	1.4611	1.5803
MAD	30	0.1414	0.6281	0.6120	50	0.1461	0.6315	0.6167
M -ESCALA	30	0.0894	0.6556	0.6614	50	0.0916	0.6614	0.6683

Tabla 3.14: Valores medios de $\|\widehat{\phi}_{\text{SI},j}/\|\widehat{\phi}_{\text{SI},j}\| - \phi_j\|^2$ bajo $C_{3,b}$.

Estimador de escala	p_n	$j = 1$	$j = 2$	$j = 3$	p_n	$j = 1$	$j = 2$	$j = 3$
		Base de Fourier				Base de B -splines		
SD	10	0.0005	1.8304	1.9925	10	1.9573	1.8919	1.7526
MAD	10	0.0201	1.8316	1.9909	10	1.8653	1.6676	1.5764
M -ESCALA	10	0.0020	1.8305	1.9962	10	1.8825	1.6810	1.6134
SD	20	1.5778	1.5594	1.7982	20	1.9041	1.6095	1.6431
MAD	20	0.1814	0.2151	1.8038	20	1.2141	1.1535	1.2040
M -ESCALA	20	0.1879	0.2038	1.8022	20	1.2521	1.1752	1.2061
SD	30	1.8044	0.3783	1.7560	50	1.8094	0.4074	1.7541
MAD	30	0.3039	0.4485	0.2849	50	0.3188	0.4633	0.2891
M -ESCALA	30	0.3406	0.4970	0.3064	50	0.3604	0.5154	0.3114

Tabla 3.15: Valores medios de $\|\widehat{\phi}_{\text{SI},j}/\|\widehat{\phi}_{\text{SI},j}\| - \phi_j\|^2$ bajo C_{23} .

Estimador de escala	p_n	$j = 1$	$j = 2$	$j = 3$	p_n	$j = 1$	$j = 2$	$j = 3$
		Base de Fourier				Base de B -splines		
SD	10	0.0088	1.8310	1.9941	10	1.7544	1.5472	1.4988
MAD	10	0.0072	1.8308	1.9943	10	1.7966	1.5910	1.5029
M -ESCALA	10	0.0008	1.8305	1.9979	10	1.8109	1.6112	1.4832
SD	20	0.2815	0.2832	1.7983	20	1.1905	1.1669	1.1749
MAD	20	0.0826	0.1012	1.7989	20	1.0752	1.0466	1.1603
M -ESCALA	20	0.0478	0.0610	1.7980	20	1.0518	1.0228	1.1319
SD	30	0.3352	0.4906	0.2542	50	0.3492	0.5064	0.2633
MAD	30	0.0977	0.1923	0.1380	50	0.1035	0.2102	0.1419
M -ESCALA	30	0.0578	0.1189	0.0790	50	0.0631	0.1481	0.0881

Tabla 3.16: Valores medios de $\|\widehat{\phi}_{\text{SI},j}/\|\widehat{\phi}_{\text{SI},j}\| - \phi_j\|^2$ bajo C_c .

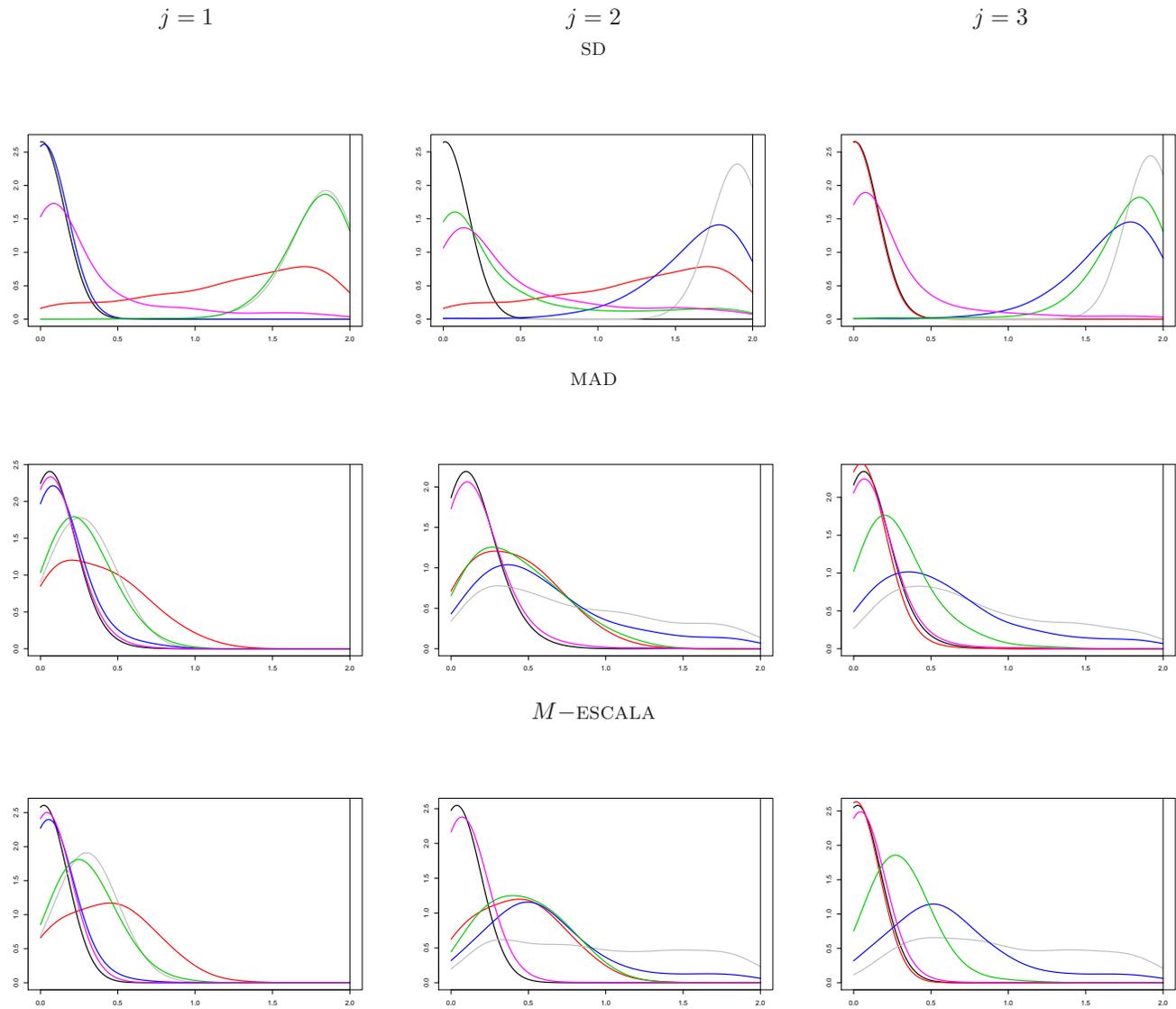


Figura 3.1: Estimadores de densidad de $C(\hat{\phi}_j, \phi_j)$ con ancho de banda 0.6 cuando se usan los estimadores “crudos”. La línea negra corresponde a C_0 , mientras que las líneas roja, gris, azul, verde y rosa corresponden a C_2 , $C_{3,a}$, $C_{3,b}$, C_{23} y C_c , respectivamente.

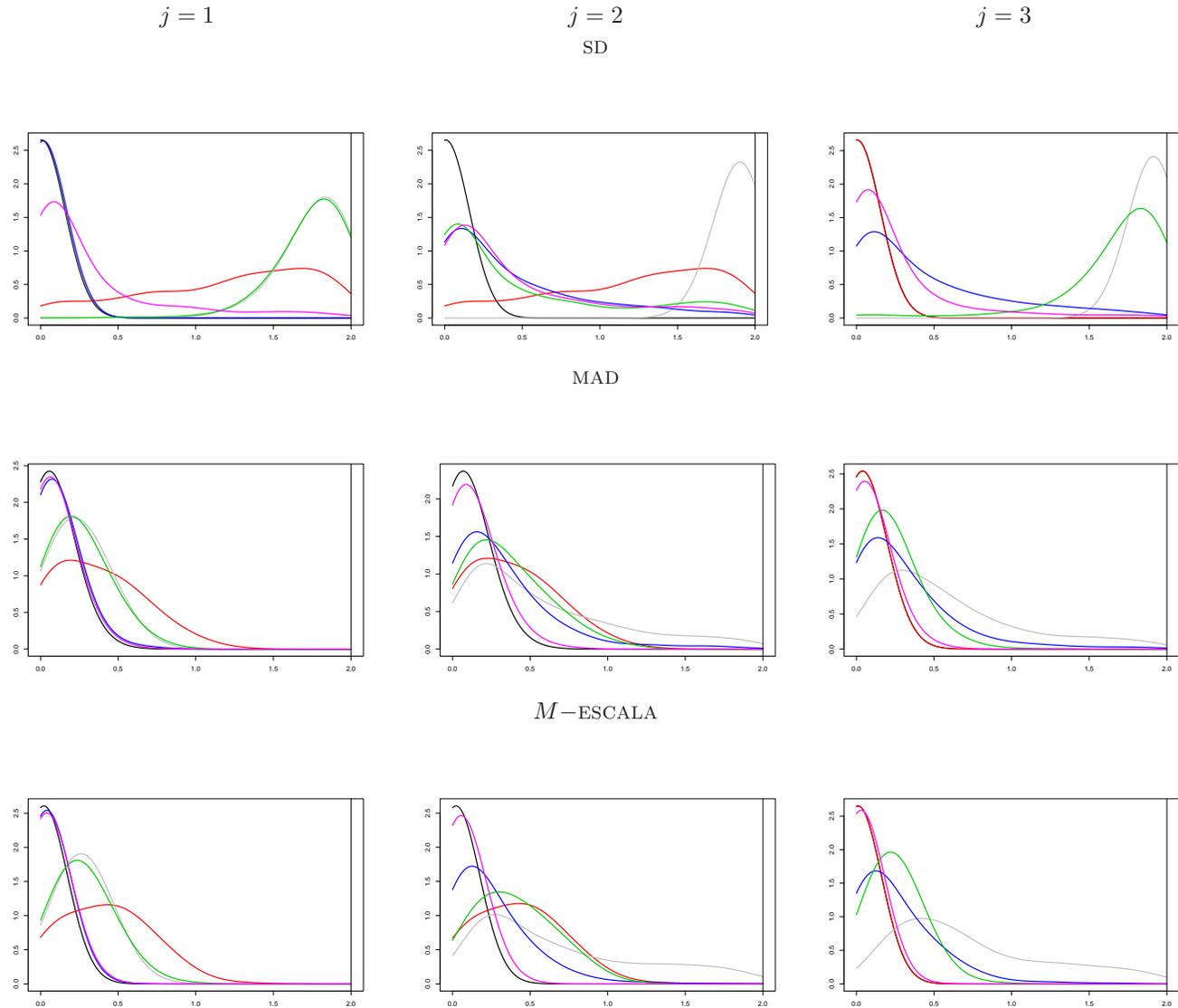


Figura 3.2: Estimadores de densidad $C(\hat{\phi}_j, \phi_j)$ con ancho de banda 0.6 cuando se usan los estimadores que penalizan por la escala, $\hat{\phi}_{PS,j}$, con una penalización de $\rho = 1.50n^{-3}$. La línea negra corresponde a C_0 , mientras que las líneas roja, gris, azul, verde y rosa corresponden a C_2 , $C_{3,a}$, $C_{3,b}$, C_{23} y C_c , respectivamente.

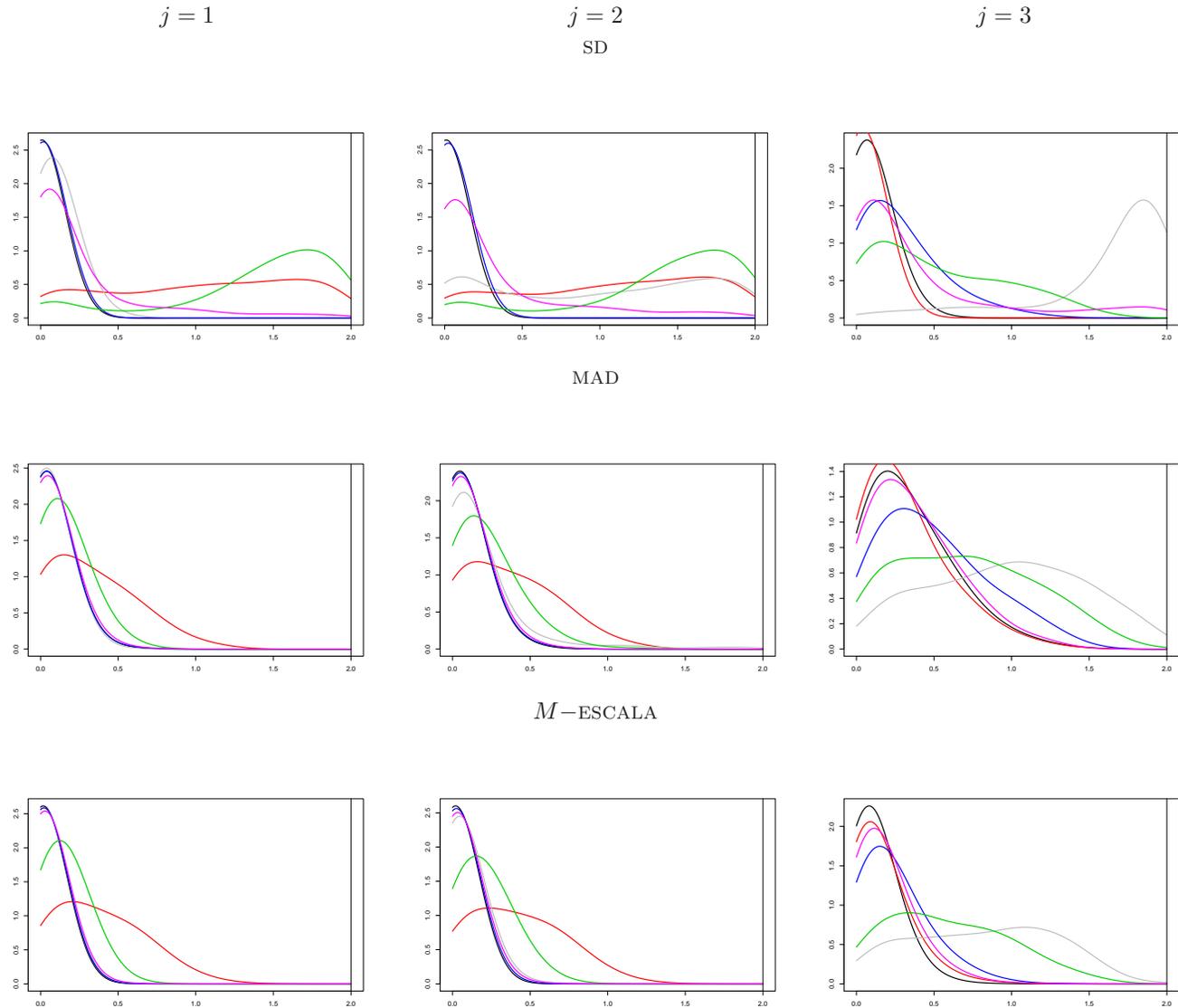


Figura 3.3: Estimadores de densidad de $C(\hat{\phi}_j, \phi_j)$ con ancho de banda 0.6 cuando se usan los estimadores que penalizan por la norma, $\hat{\phi}_{PN,j}$, con una penalización de $r = 0.75n^{-3}$. La línea negra corresponde a C_0 , mientras que las líneas roja, gris, azul, verde y rosa corresponden a C_2 , $C_{3,a}$, $C_{3,b}$, C_{23} y C_c , respectivamente.

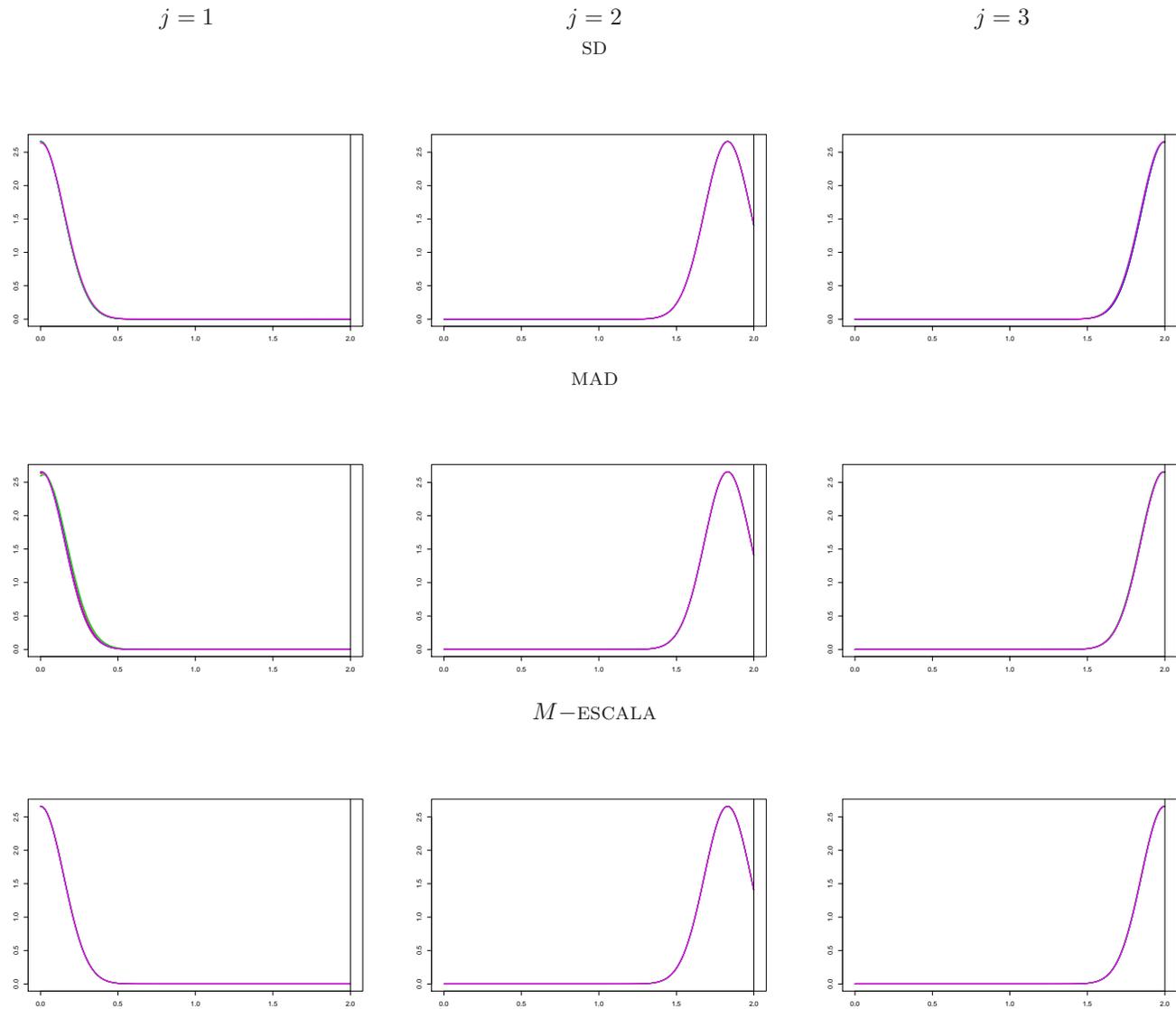


Figura 3.4: Estimadores de la densidad de $C(\hat{\phi}_j, \phi_j)$ con ancho de banda 0.6 para los estimadores $\hat{\phi}_{SI,j}$ defined in (3.8), cuando $p_n = 10$ usando la base de Fourier. La línea negra corresponde a C_0 , mientras que las líneas roja, gris, azul, verde y rosa corresponden a C_2 , $C_{3,a}$, $C_{3,b}$, C_{23} y C_c , respectivamente.

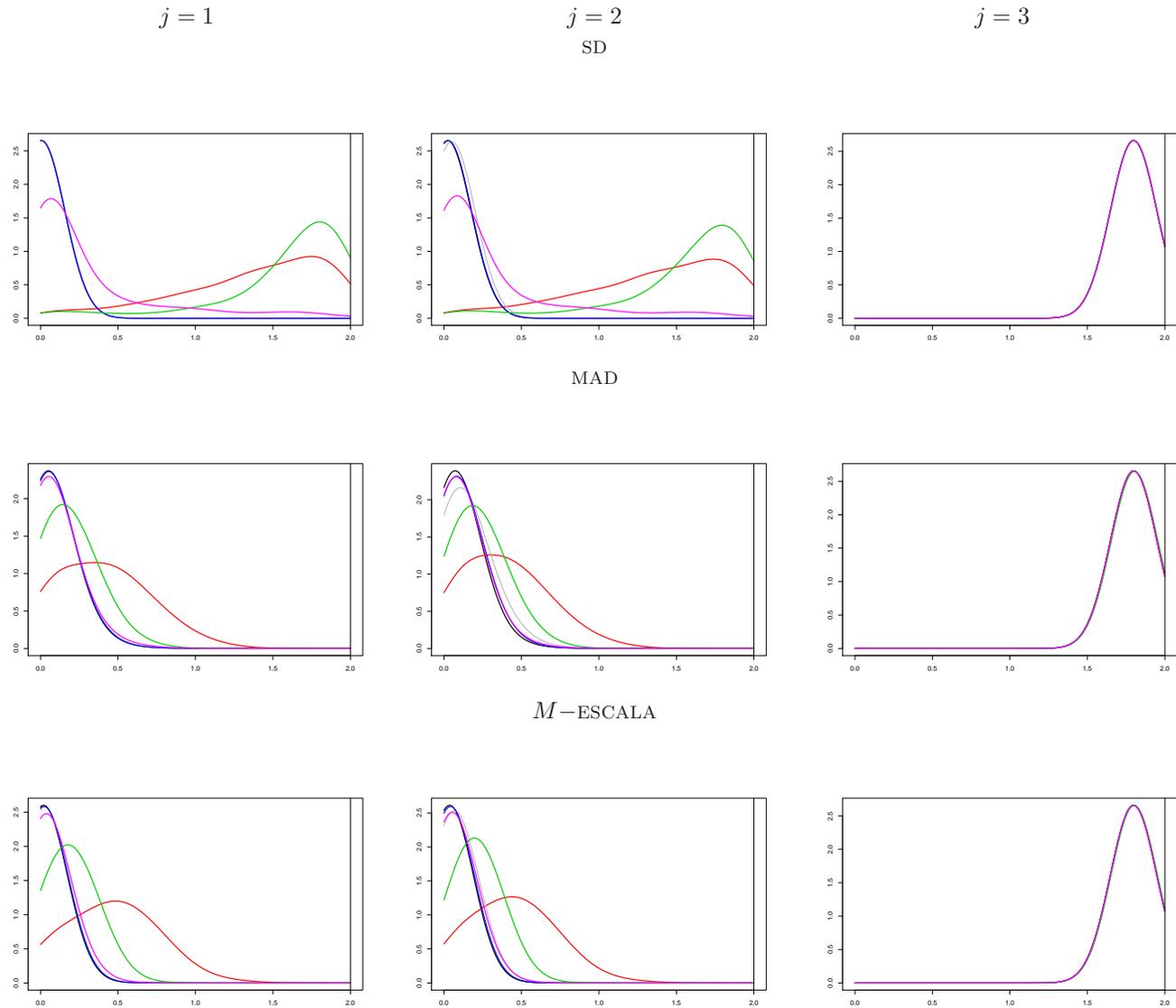


Figura 3.5: Estimadores de la densidad de $C(\hat{\phi}_j, \phi_j)$ con ancho de banda 0.6 para los estimadores $\hat{\phi}_{SI,j}$ defined in (3.8), cuando $p_n = 20$ usando la base de Fourier. La línea negra corresponde a C_0 , mientras que las líneas roja, gris, azul, verde y rosa corresponden a C_2 , $C_{3,a}$, $C_{3,b}$, C_{23} y C_c , respectivamente.

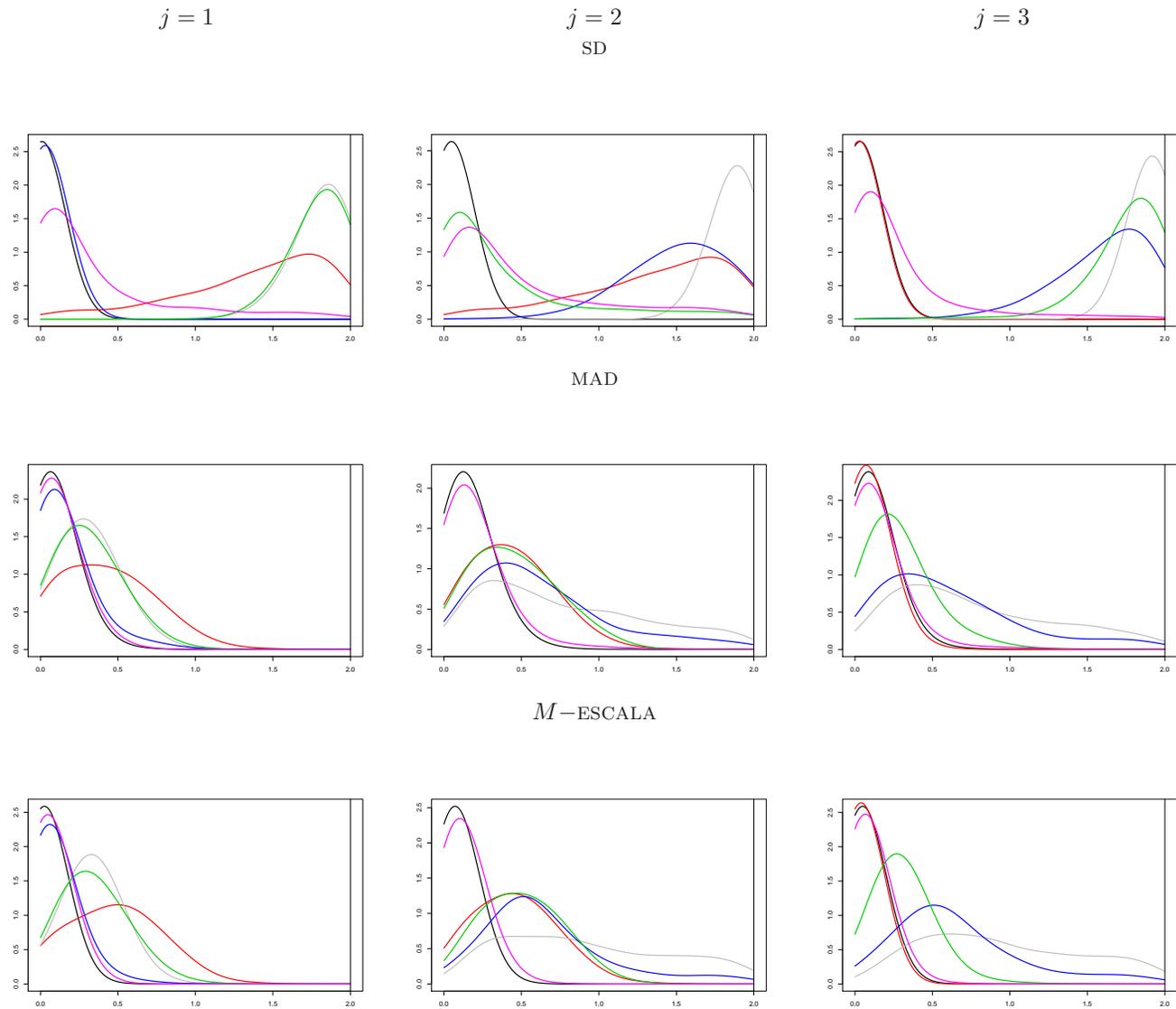


Figura 3.6: Estimadores de la densidad de $C(\hat{\phi}_j, \phi_j)$ con ancho de banda 0.6 para los estimadores $\hat{\phi}_{SI,j}$ defined in (3.8), cuando $p_n = 30$ usando la base de Fourier. La línea negra corresponde a C_0 , mientras que las líneas roja, gris, azul, verde y rosa corresponden a C_2 , $C_{3,a}$, $C_{3,b}$, C_{23} y C_c , respectivamente.

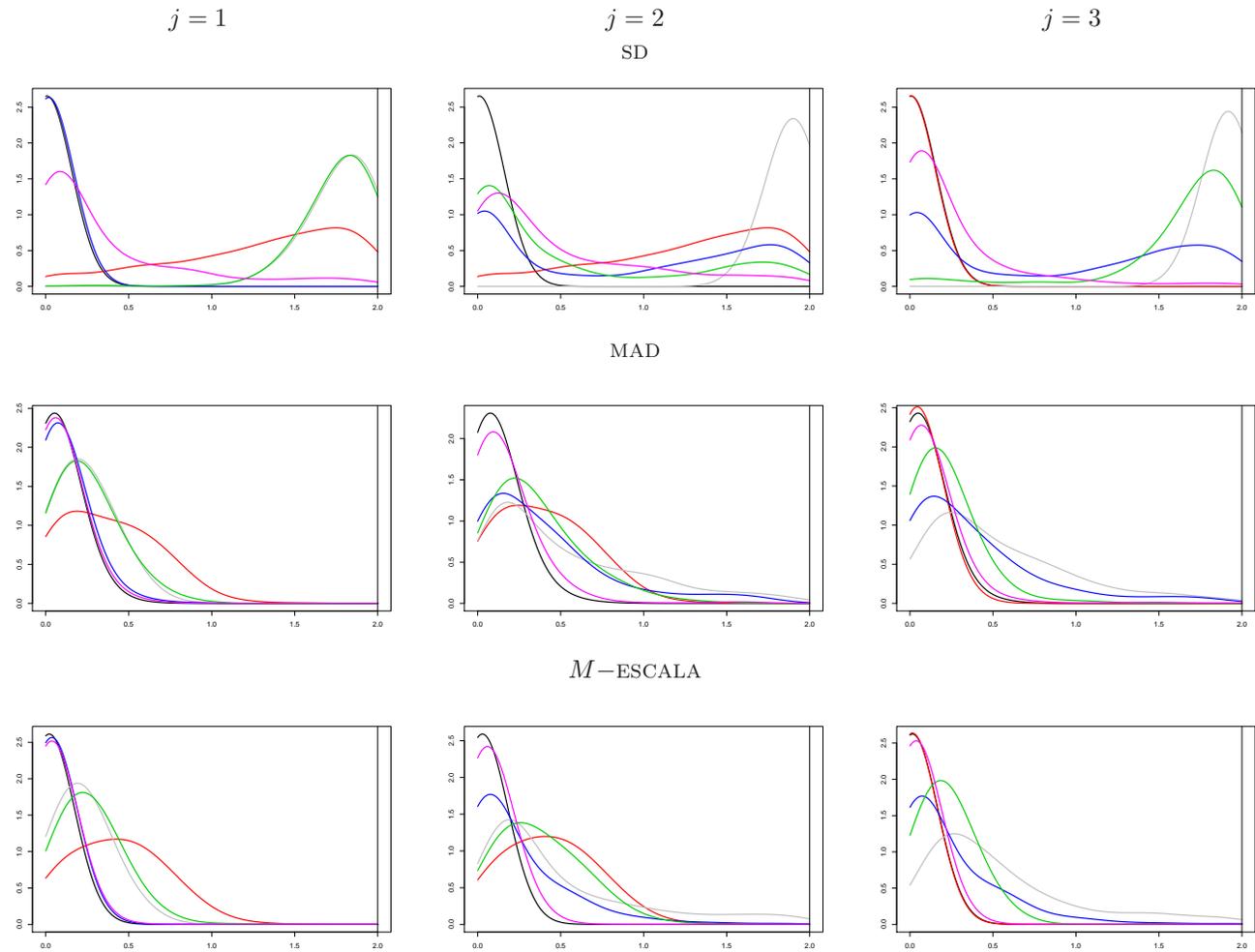


Figura 3.7: Estimadores de la densidad de $C(\hat{\phi}_j, \phi_j)$ cuando la penalización τ es elegida mediante una validación cruzada por K -fold y los estimadores son aquellos que penalizan por escala, $\hat{\phi}_{PS,j}$. La línea negra corresponde a C_0 , mientras que las líneas roja, gris, azul, verde y rosa corresponden a C_2 , $C_{3,a}$, $C_{3,b}$, C_{23} y C_c , respectivamente.

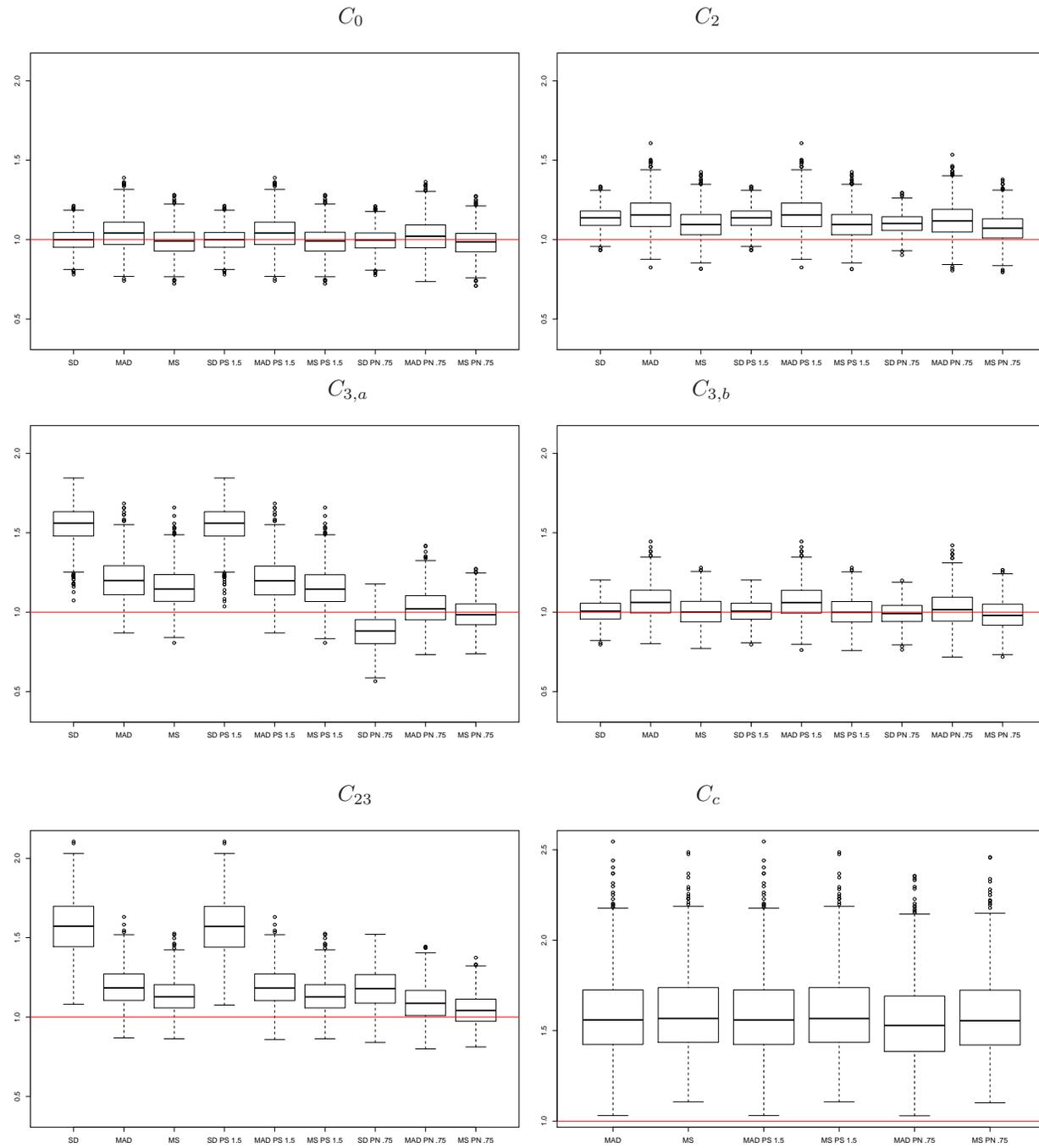


Figura 3.8: Boxplots del cociente $\hat{\lambda}_1/\lambda_1$ para los estimadores “crudos” y penalizados.

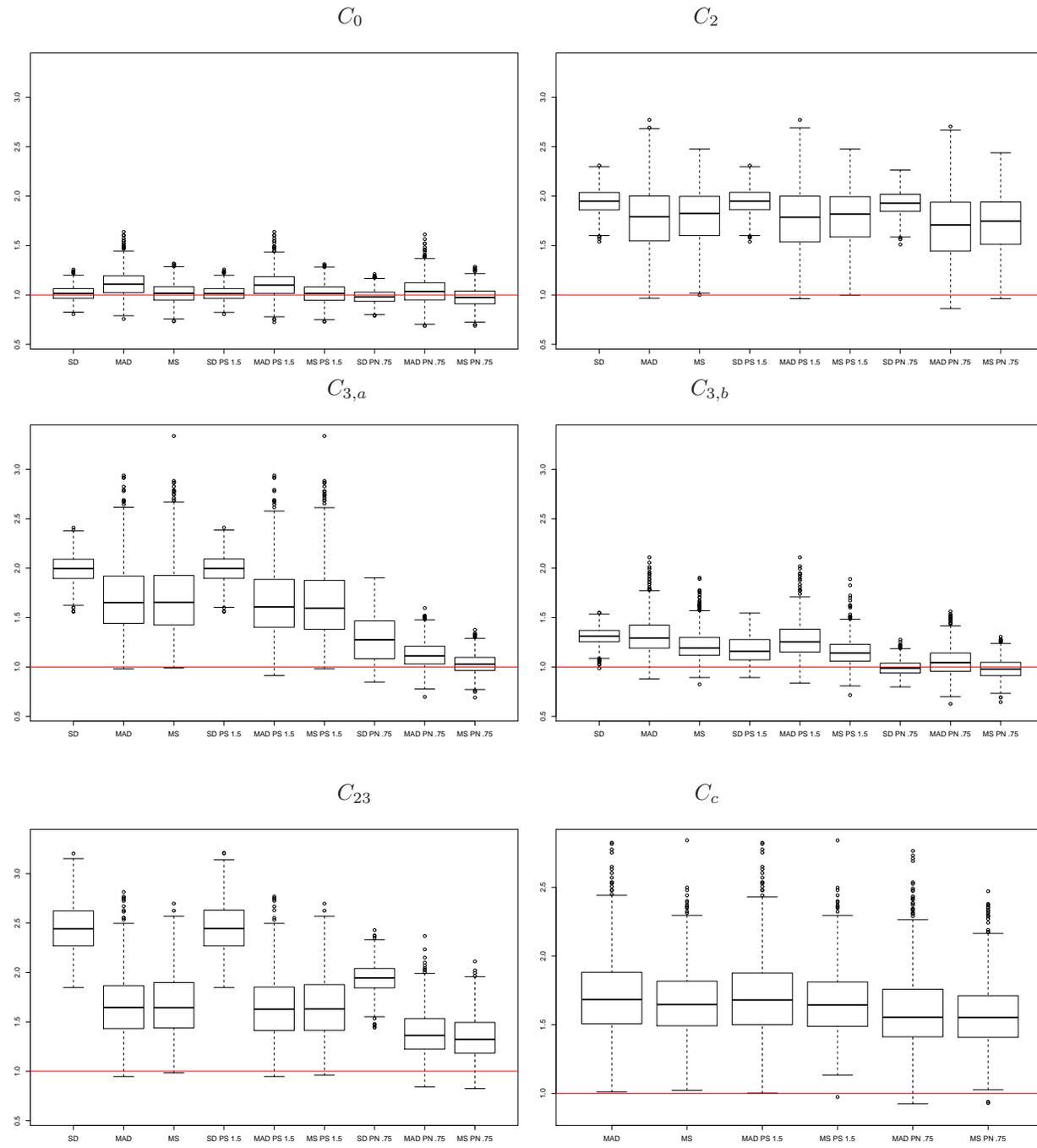


Figura 3.9: Boxplots del cociente $\hat{\lambda}_2/\lambda_2$ para los estimadores “crudos” y penalizados.

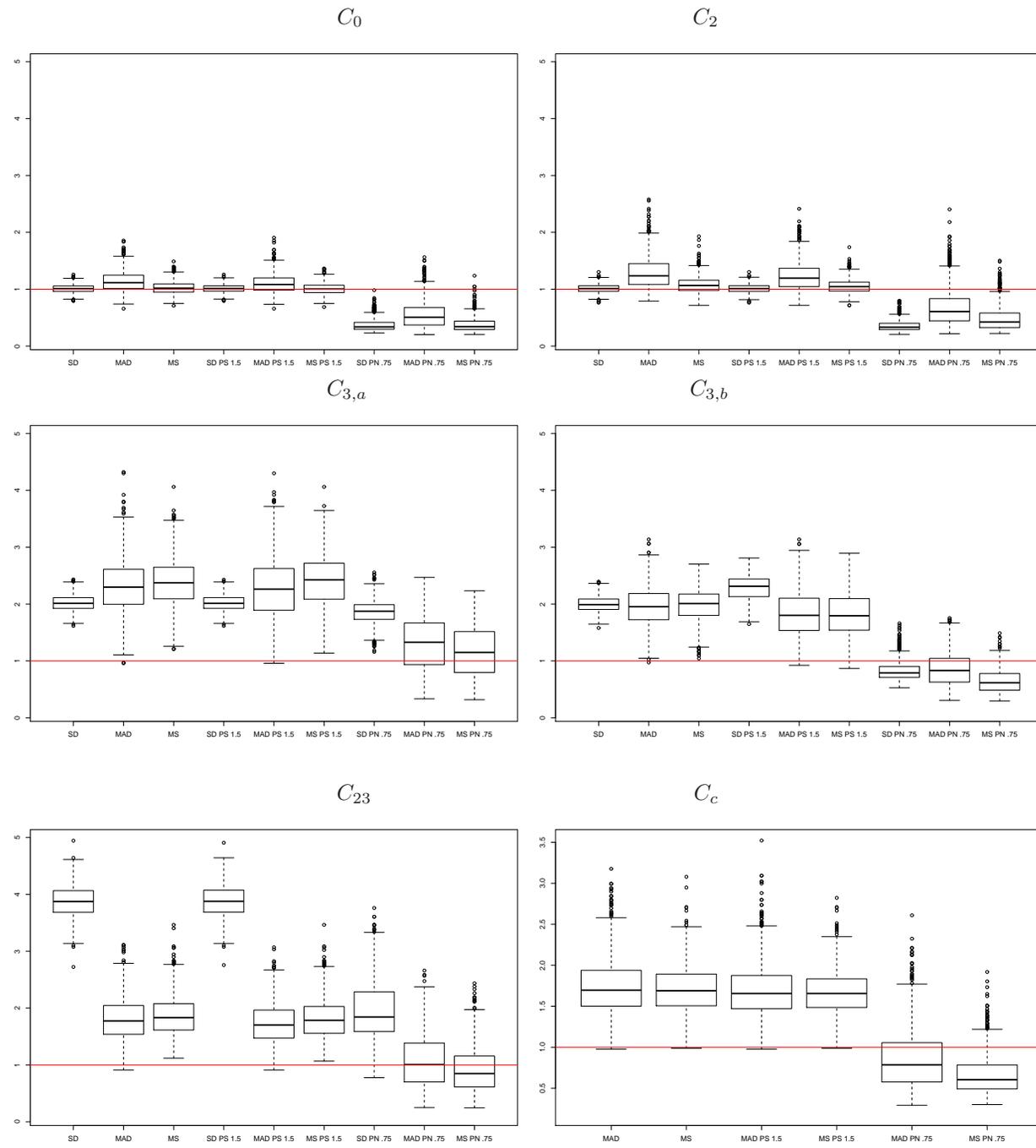


Figura 3.10: Boxplots del cociente $\hat{\lambda}_3/\lambda_3$ para los estimadores “crudos” y penalizados.

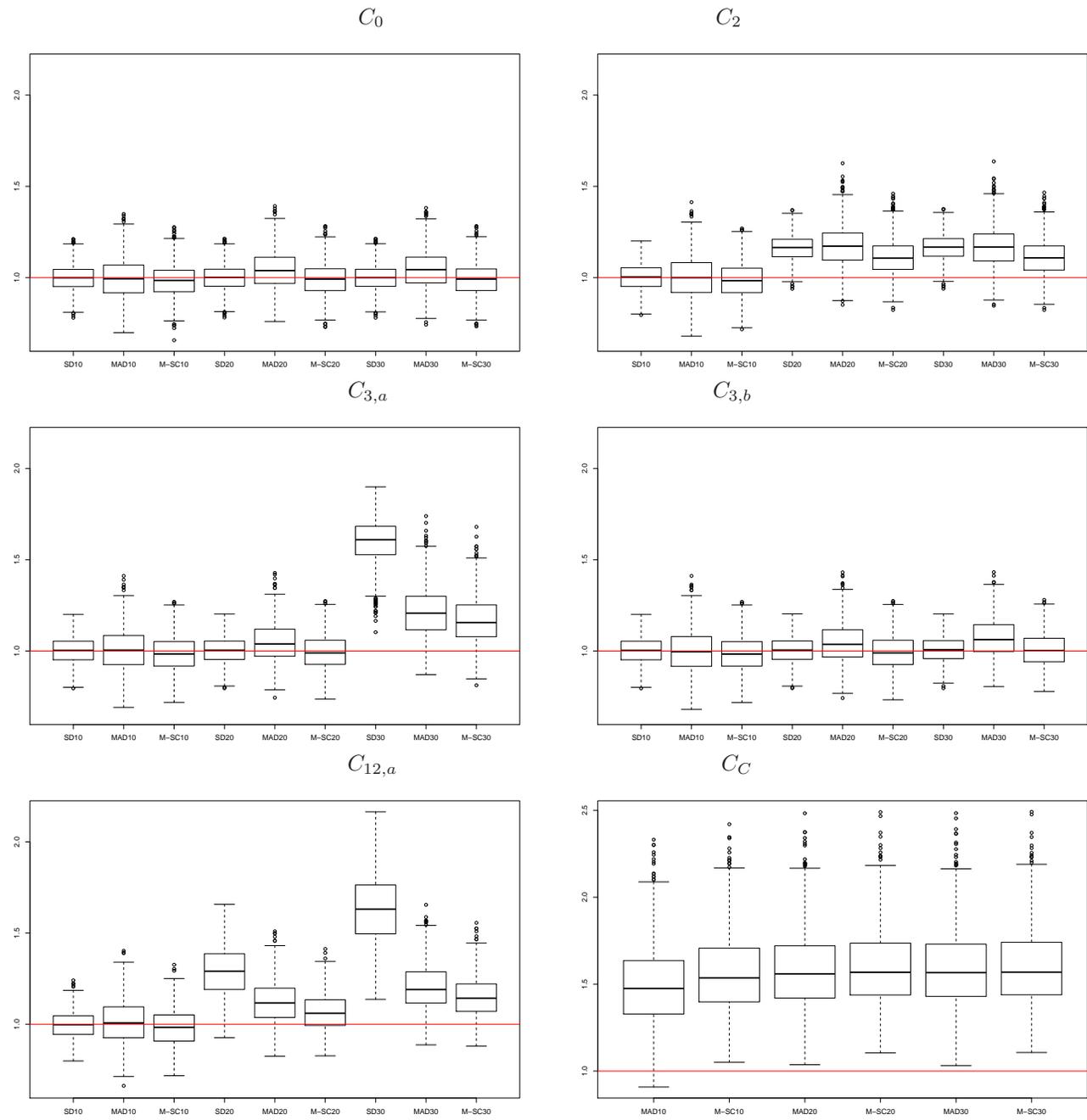


Figura 3.11: Boxplots para el cociente $\hat{\lambda}_1/\lambda_1$ para los estimadores de sieves al utilizar la base de Fourier.

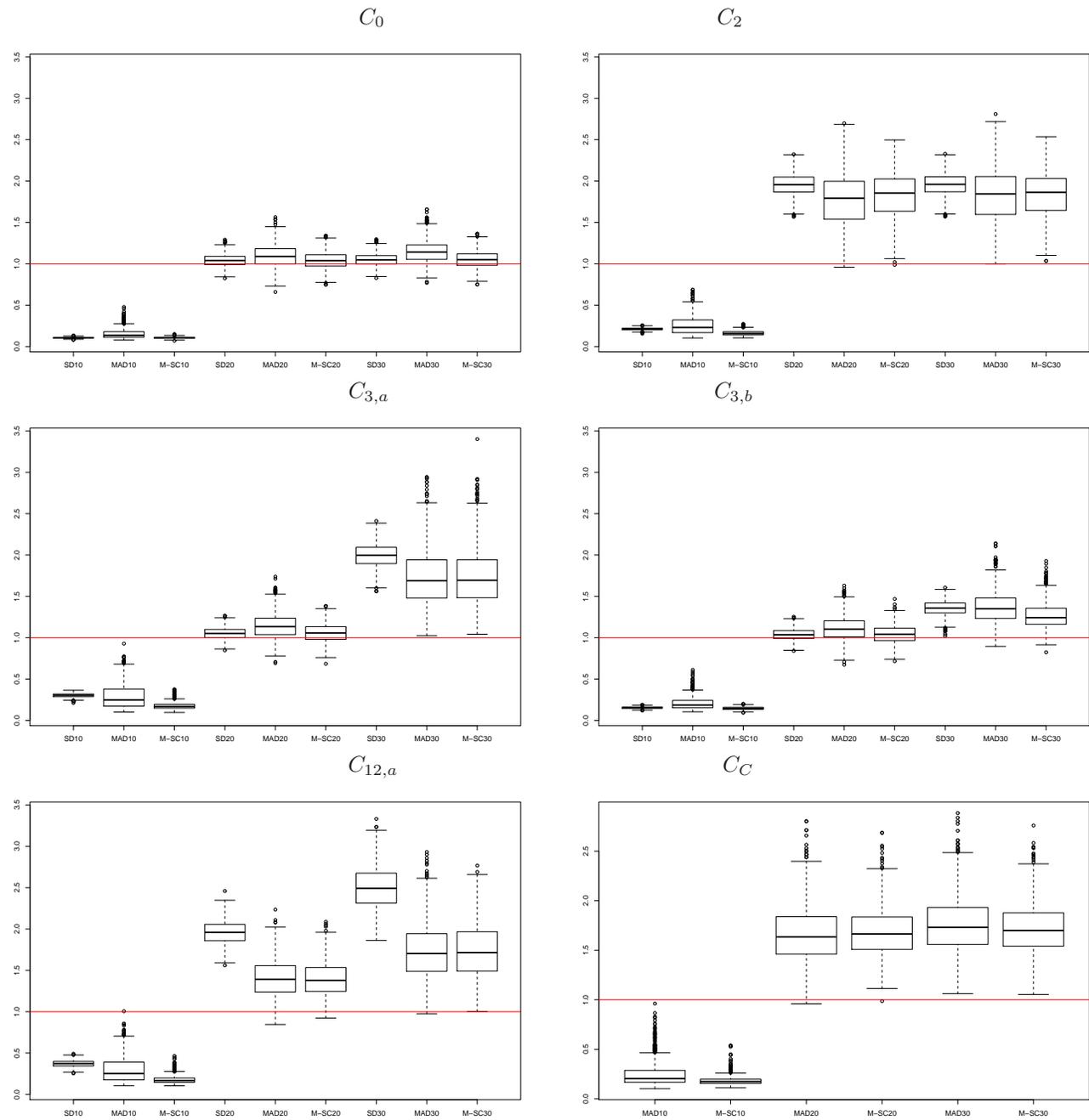


Figura 3.12: Boxplots para el cociente $\hat{\lambda}_2/\lambda_2$ para los estimadores de sieves al utilizar la base de Fourier.

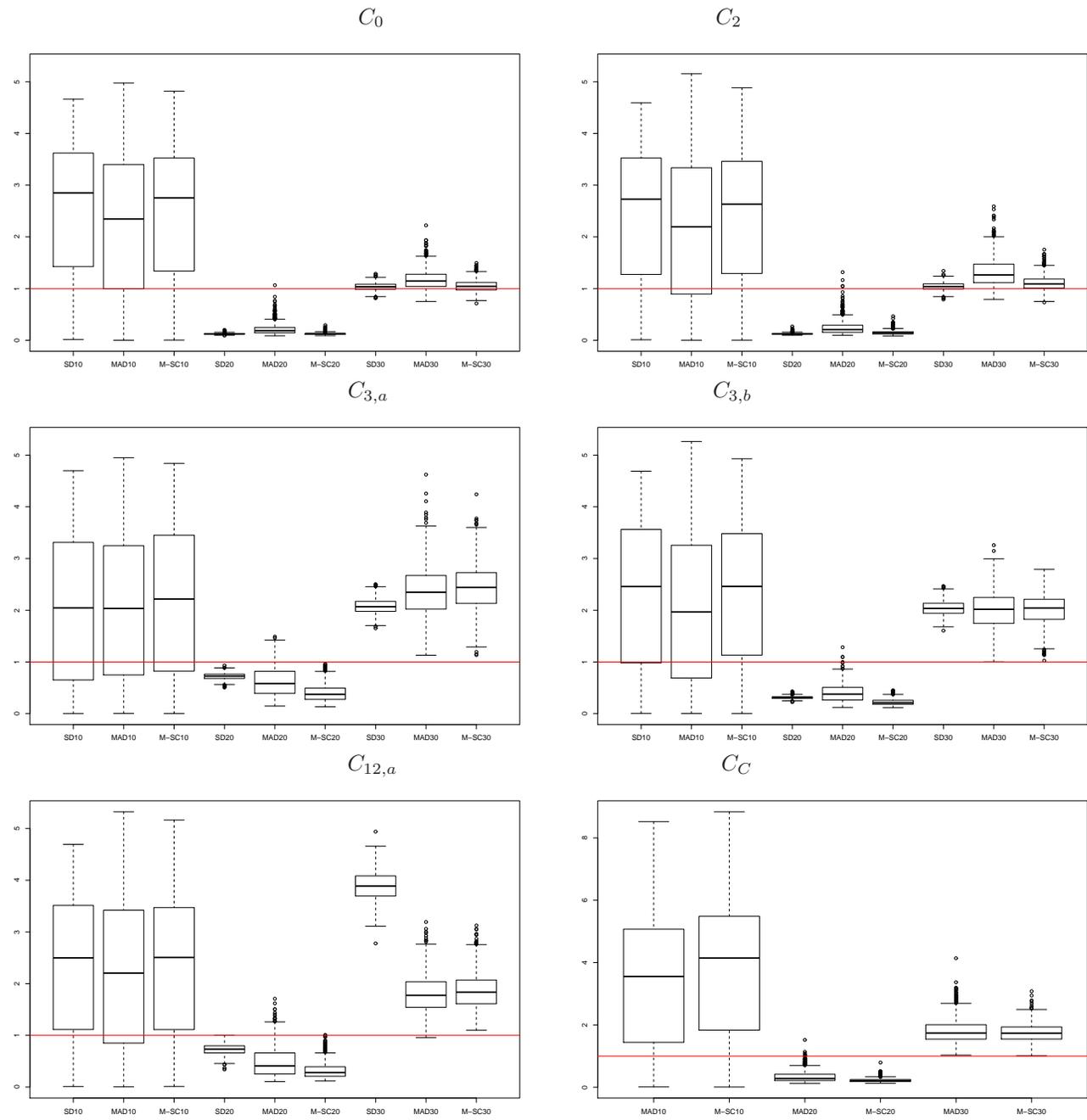


Figura 3.13: Boxplots para el cociente $\hat{\lambda}_3/\lambda_3$ para los estimadores de sieves al utilizar la base de Fourier.

3.6.4 Simulación por K -th fold

La Tabla 3.17 reporta los tiempos de cómputo en minutos para 1000 replicaciones y para un valor fijo del parámetro de penalizado. Esto sugiere que el procedimiento de validación cruzada eliminando un dato a la vez es difícil de ejecutar y, por eso, un esquema de K -fold fue adoptado en su lugar. Se llevó a cabo un estudio de simulación donde el parámetro de suavizado fue elegido usando el procedimiento descrito en la sección 3.5 con $K = 4$, $\ell = 1$. Se realizaron 500 replicaciones.

	SD	MAD	M -ESCALA
$\hat{\phi}_{RAW,j}$	5.62	6.98	17.56
$\hat{\phi}_{PS,j}$	7.75	9.00	20.18
$\hat{\phi}_{PN,j}$	31.87	33.21	44.04

Tabla 3.17: Tiempos de ejecución en minutos para 1000 replicaciones y un valor fijo de r o ρ .

Los resultados cuando se penaliza la función de escala, es decir, para los estimadores definidos en (3.6), son reportados en Tabla 3.18 y en la Figura 3.7. Los estimadores clásicos son sensibles a las contaminaciones consideradas y excepto para contaminaciones en la tercer componente, las alternativas robustas muestran su ventaja. Notemos que, ambos C_{3a} y C_{3b} afectan a los estimadores robustos, cuando el parámetro de suavizado es elegido por el método robusto de validación cruzada K -fold.

Modelo	Estimador de escala	$j = 1$	$j = 2$	$j = 3$
			$\hat{\phi}_{PS,j}$	
C_0	SD	0.0073	0.0094	0.0078
	MAD	0.0662	0.0993	0.0634
	M -ESCALA	0.0225	0.0311	0.0172
C_2	SD	1.2840	1.2837	0.0043
	MAD	0.3731	0.3915	0.0504
	M -ESCALA	0.4261	0.4286	0.0153
C_{3A}	SD	1.7840	1.8901	1.9122
	MAD	0.2271	0.5227	0.5450
	M -ESCALA	0.2176	0.4873	0.5437
C_{3B}	SD	0.0192	0.8350	0.8525
	MAD	0.0986	0.3930	0.3820
	M -ESCALA	0.0404	0.2251	0.2285
C_{23}	SD	1.7645	0.5438	1.6380
	MAD	0.2407	0.3443	0.2064
	M -ESCALA	0.2613	0.3707	0.2174
C_{Cauchy}	SD	0.3580	0.4835	0.2287
	MAD	0.0788	0.1511	0.1082
	M -ESCALA	0.0444	0.0707	0.0434

Tabla 3.18: Valores medios de $\|\hat{\phi}_j/\|\hat{\phi}_j\| - \phi_j\|^2$ cuando el parámetro de penalizado se elige mediante un procedimiento de convalidación cruzada por K -folds.

3.7 Aplicación a un ejemplo real: movimiento del labio

El siguiente ejemplo fue considerado por Gervini (2008) para mostrar el efecto de *outliers* sobre las componentes funcionales principales. Se solicitó a un sujeto que pronunciara la palabra *bob* 32 veces y se registró la posición del labio inferior en cada instante de tiempo. Los datos de movimiento del labio fueron originalmente analizados por Malfait y Ramsay (2003). En la Figura 3.14 se grafican las 32 trayectorias del labio inferior versus el tiempo. Tres de estas curvas (graficadas con una línea gruesa) en la figura de la izquierda parece estar fuera de rango, mostrando un corrimiento en el segundo pico que aparece retrasados respecto del resto. Estas tres trayectorias fueron consideradas como *outliers* por Gervini (2008). Para determinar si estas curvas están o no en el rango normal de variabilidad, es necesario estimar correctamente las componentes principales.

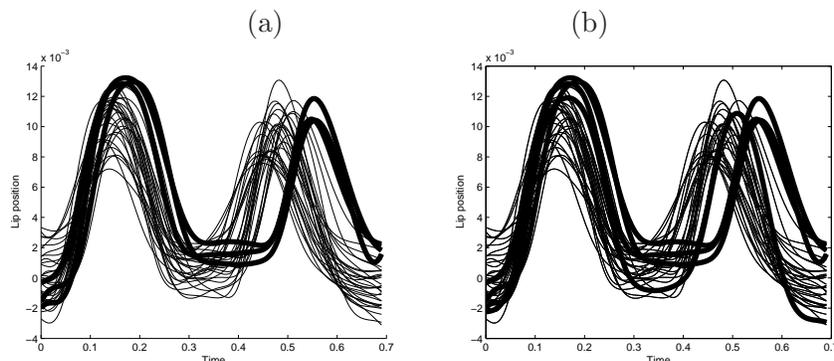


Figura 3.14: Datos de movimiento del labio. (a) Las trayectorias 24,25 y 27 se indican con líneas gruesas (b) Las trayectorias 14, 24,25 y 27 se indican con líneas gruesas.

Como en Gervini (2008), estimamos las 5 primeras direcciones principales usando los estimadores robustos sin penalizar definidos en (3.3) utilizando una M -escala con función de Tukey.

La Figura 3.15 muestra los gráficos de los estimadores clásicos y robustos de las direcciones principales que aparecen como una línea sólida y entrecortada (en rojo), respectivamente. Gervini (2008) describe el tipo de variabilidad explicada por estas componentes. Además, como se describe ahí mismo, un escor alto en la primer dirección, es decir, un valor positivo de $\hat{s}_{i,1} = \langle X_i - \hat{\mu}, \hat{\phi}_1 \rangle$, está asociado a curvas que muestran un primer pico importante y un segundo pico retrasado, como pasa en las tres curvas atípicas.

La Figura 3.16 a) muestra los boxplots de los *escores* $\hat{s}_{i,j} = \langle X_i - \hat{\mu}, \hat{\phi}_j \rangle$ cuando $\hat{\phi}_j$ son los estimadores robustos de projection-pursuit junto con la gráfica de la norma al cuadrado de los residuos definidos por $\hat{r}_i^{(q)} = X_i - \hat{\mu} - \sum_{j=1}^q \hat{s}_{i,j} \hat{\phi}_j$ donde $\hat{\mu} = \operatorname{argmin}_{\theta \in \mathcal{H}} \sum_{i=1}^n (\|X_i - \theta\| - \|X_i\|) / n$. Solamente mostramos gráficos para el ajuste robusto puesto que al considerar estimadores clásicos suele aparecer un efecto de enmascaramiento.

El gráfico de residuos correspondiente a la M -escala, dado en la Figura 3.16 b), muestra claramente que la norma al cuadrado de los residuos correspondientes a las observaciones atípicas

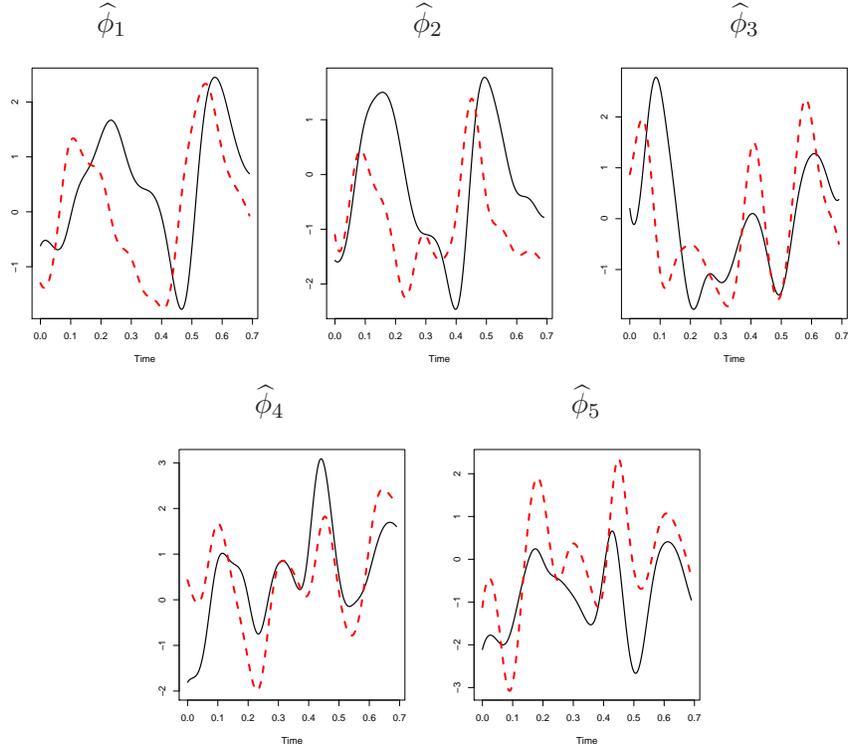


Figura 3.15: Estimadores de las direcciones principales para los datos de movimiento del labio. Las líneas sólida y entrecortada corresponden a los estimadores clásicos y robustos, respectivamente.

están fuera de rango. La Figura 3.16 también presenta los boxplots de $\|\hat{r}_i^{(q)}\|^2$. Debido a la asimetría de la distribución de la norma, consideramos los boxplots ajustados definidos en Hubert y Vandervieren (2008), en vez de los usuales.

Los dos *outliers* que aparecen en el boxplot de los residuos robustos $\|\hat{r}_i^{(1)}\|^2$ y $\|\hat{r}_i^{(5)}\|^2$ corresponden a las observaciones 24 y 25. Vale la pena mencionar que la trayectoria indicada por 14 también corresponde a un *outlier* negativo en el *escore* $\hat{s}_{i,4}$, mientras que las observaciones 24, 25 y 27 aparecen como *outliers* con gran puntaje negativo en $\hat{s}_{i,2}$. La trayectoria 14 está prácticamente explicada por las primeras cuatro componentes, ya que el mínimo y el máximo de $\hat{r}_i^{(4)}$ es -1.084×10^{-18} y 9.758×10^{-19} , respectivamente. La Figura 3.16 (d) muestra las curvas de residuos $\hat{r}_i^{(4)}$ las que sugieren que no alcanza una representación 4–dimensional de Karhunen–Loève para explicar el comportamiento de los datos, mientras que la observación 14 podría ser explicada por $\hat{\phi}_1, \dots, \hat{\phi}_4$ con los *escores* más grandes en valor absoluto en la primera y en la cuarta componente. La Figura 3.14(b) indica con trazos gruesos las observaciones 14, 24, 25 y 27. De este gráfico se observa que la curva relacionada a la observación 14 tiene un primer pico importante y un segundo pico muy suave mientras que su cuarto pico es claramente menor que el del resto de las trayectorias

y ocurre antes de la mayor parte de los datos.

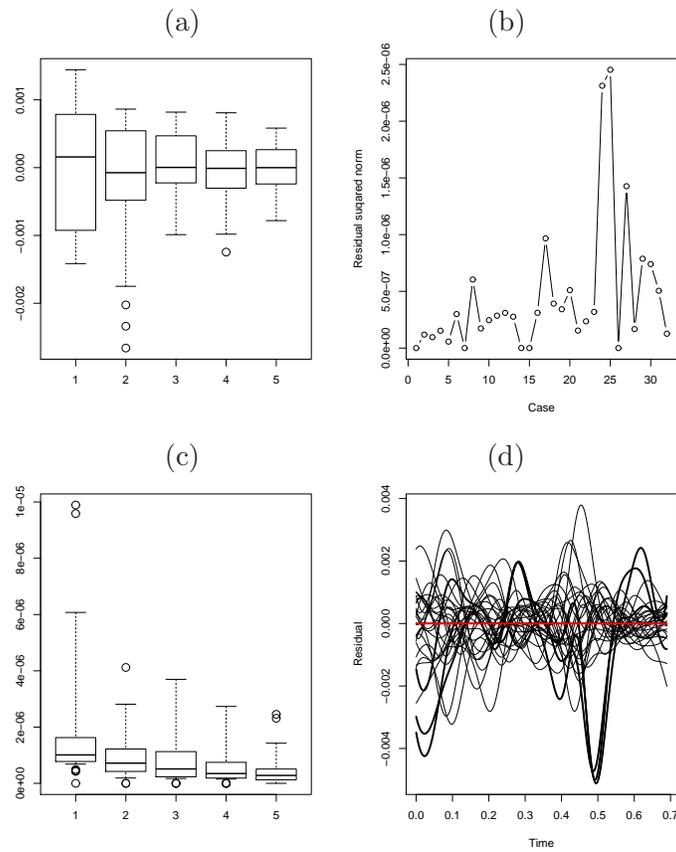


Figura 3.16: Datos de movimiento del labio. (a) Boxplot de los *escores* $\hat{s}_{i,j} = \langle X_i - \hat{\mu}, \hat{\phi}_j \rangle$, (b) Gráficos de la norma al cuadrado de los residuos $\| \hat{r}_i^{(5)} \|^2$, (c) Boxplots ajustados de $\| \hat{r}_i^{(q)} \|^2$, $1 \leq q \leq 5$ y (d) Gráficos de los residuos $\hat{r}_i^{(4)}$ basados en un ajuste robusto. Las líneas gruesas corresponden a las observaciones 24, 25 y 27 mientras que la línea roja horizontal es la trayectoria 14.

Capítulo 4

Función de Influencia

4.1 Función de Influencia

Como se había comentado en el capítulo 2, la función de influencia es una de las herramientas heurísticas más utilizadas para medir el nivel de robustez de un funcional o estimador. Da una idea de la sensibilidad infinitesimal del estimador considerado y se define como

$$\text{IF}(x, T, P) = \lim_{\epsilon \rightarrow 0} \frac{T((1 - \epsilon)P + \epsilon\delta_x) - T(P)}{\epsilon}$$

para los puntos $x \in \Omega$ para los cuales el límite existe. Es decir, $\text{IF}(x, T, P)$ es la derivada de Gateaux del funcional T en el punto P sobre la dirección $\delta_x - P$. Una de las utilidades de la función de influencia es el cálculo de la varianza asintótica de un funcional (ver Capítulo 2 para más detalles).

El siguiente teorema provee una expresión para la función de influencia de los funcionales definidos en (3.1) en una distribución elíptica $\mathcal{E}(\mu, \mathbf{\Gamma})$. Queremos recordar que, en el caso finito dimensional, un resultado análogo bajo elipticidad fue obtenido por Croux y Ruiz-Gazen (2005). Luego, nuestros resultados extienden los de estos autores al caso de un espacio de Hilbert separable. Se prueba además la existencia de la función de influencia.

Como $X \sim \mathcal{E}(\mu, \mathbf{\Gamma})$, si $\lambda_1 \geq \lambda_2 \geq \dots$ son los autovalores de $\mathbf{\Gamma}$ y ϕ_j sus autofunciones asociadas, tenemos que $\phi_{R,j}(P) = \phi_j$ y $\lambda_{R,j} = c\lambda_j$, donde $c = \sigma_R(F_0)$ con F_0 la distribución dada en el Lema 2.1.2. Entonces, si $\lambda_1 > \lambda_2 > \dots > \lambda_q > \lambda_{q+1}$, para $1 \leq j \leq q$, $\phi_{R,j}(P)$ son únicos salvo un cambio de signo. Sin pérdida de generalidad, supondremos que el parámetro de posición vale 0 y que $c = 1$.

Teorema 4.1.1. *Sea P una medida elíptica de probabilidad $P = \mathcal{E}(\mu, \mathbf{\Gamma})$, donde $\mu = 0$ y $\mathbf{\Gamma}$ es un operador auto-adjunto, semi-definido positivo y compacto con autovalores $\lambda_1 \geq \lambda_2 \geq \dots$ y sea ϕ_j la autofunción asociada a λ_j . Supongamos que $\lambda_1 > \lambda_2 > \dots > \lambda_q > \lambda_{q+1}$.*

Sea F_0 la medida univariada de escala definida en el Lema 2.1.2. Supongamos que la aplicación $S : [0, 1] \times \mathbb{R} \rightarrow \mathbb{R}$ definida por $S(\epsilon, y) = \sigma_R((1 - \epsilon)F_0 + \epsilon\delta_y)$ es dos veces continuamente diferenciable en cualquier $(0, y)$. En particular, $\text{IF}(y; \sigma_R, F_0)$ tiene que ser diferenciable y su derivada será notada por $\text{IF}_1(y; \sigma_R, F_0)$.

Además, supongamos que existen $\epsilon_0 < 1$, tal que para cada $\epsilon \in [0, \epsilon_0]$ fijo y $x \in \mathcal{H}$, la aplicación $\Upsilon_{x,\epsilon} : \mathcal{H} \rightarrow \mathbb{R}$ definida como $\Upsilon_{x,\epsilon}(\alpha) = \sigma_{\mathbb{R}}^2(P_{x,\epsilon}[\alpha])$ es Hadamard diferenciable, donde $P_{x,\epsilon} = (1 - \epsilon)P + \epsilon\delta_x$ es la medida de probabilidad contaminada. Supongamos además que, para $k \leq q$, la aplicación $\Upsilon : [0, \epsilon_0] \times \mathcal{H} \rightarrow \mathbb{R}$ definida como $\Upsilon(\epsilon, \alpha) = \sigma_{\mathbb{R}}^2(P_{x,\epsilon}[\alpha])$ es dos veces continuamente Hadamard diferenciable en $(0, \phi_k)$.

Sin pérdida de generalidad, supongamos que $\lambda_{R,j} = \lambda_j$ lo que significa que $\sigma_{\mathbb{R}}$ es Fisher-consistente en F_0 . Luego, para todo $k \leq q$, la función de influencia del funcional de las direcciones principales definido en (3.1) existe y está dada por

$$\begin{aligned} IF(x; \phi_{R,k}, P) &= \sum_{j \geq k+1} \frac{\sqrt{\lambda_k}}{\lambda_k - \lambda_j} IF_1 \left(\frac{\langle x, \phi_k \rangle}{\sqrt{\lambda_k}}; \sigma_{\mathbb{R}}, F_0 \right) \langle x, \phi_j \rangle \phi_j \\ &\quad + \sum_{j=1}^{k-1} \frac{\sqrt{\lambda_j}}{\lambda_k - \lambda_j} IF_1 \left(\frac{\langle x, \phi_j \rangle}{\sqrt{\lambda_j}}; \sigma_{\mathbb{R}}, F_0 \right) \langle x, \phi_k \rangle \phi_j. \end{aligned} \quad (4.1)$$

mientras que la función de influencia de los valores principales definidos por (3.2) está dada por

$$IF(x; \lambda_{R,k}, P) = 2 \lambda_k IF \left(\frac{\langle x, \phi_k \rangle}{\sqrt{\lambda_k}}; \sigma_{\mathbb{R}}, F_0 \right). \quad (4.2)$$

Observemos que si x es colineal con ϕ_k , entonces $IF(x; \phi_k, P) = 0$.

Daremos primero algunas definiciones y cálculos auxiliares que serán usados en las demostraciones que siguen.

Si V indica a la variable aleatoria independiente de $X \sim P$ tal que $\mathbb{P}(V = 1) = \epsilon$ y $\mathbb{P}(V = 0) = 1 - \epsilon$, tenemos que $(1 - V)X + V\delta_x \sim P_{x,\epsilon}$.

Definamos $\phi_{k,\epsilon} = \phi_{R,j}(P_{x,\epsilon})$ y $\lambda_{j,\epsilon} = \lambda_{R,j}(P_{x,\epsilon})$ para todo $\epsilon < \epsilon_0$. Tenemos que $\|\phi_{j,\epsilon}\| = 1$ y $\langle \phi_{j,\epsilon}, \phi_{\ell,\epsilon} \rangle = 0$, para $\ell \neq j$.

Sea $k \leq q$ y definamos las restricciones $\Psi : \mathcal{H} \rightarrow \mathbb{R}^k$ como $\Psi(\alpha) = (\Psi_0(\alpha), \dots, \Psi_{k-1}(\alpha))$ con $\Psi_0(\alpha) = \|\alpha\|^2 - 1$ y $\Psi_j(\alpha) = \langle \alpha, \phi_{j,\epsilon} \rangle$, para $1 \leq j \leq k - 1$, donde entendemos que cuando $k = 1$, $\Psi(\alpha) = \Psi_0(\alpha)$. Por la Observación 2.2.2, Ψ es \mathcal{C}^1 -Fréchet diferenciable. Más aún, $\Psi'_{\phi_{k,\epsilon}}$ es sobreyectiva sobre \mathbb{R}^k ya que $\Psi'_{\phi_{k,\epsilon}} = (\Psi'_{0,\phi_{k,\epsilon}}, \dots, \Psi'_{k-1,\phi_{k,\epsilon}})$ con

$$\Psi'_{0,\phi_{k,\epsilon}}(\alpha) = \langle 2\phi_{k,\epsilon}, \alpha \rangle \quad \Psi'_{j,\phi_{k,\epsilon}}(\alpha) = \langle \phi_{j,\epsilon}, \alpha \rangle, \text{ for } 1 \leq j \leq k - 1.$$

Efectivamente, dados $\mathbf{y} = (y_1, \dots, y_k)^T \in \mathbb{R}^k$, consideremos $\alpha = (y_1/2)\phi_{k,\epsilon} + \sum_{j=1}^{k-1} y_{j+1}\phi_{j,\epsilon}$, luego $\Psi'_{\phi_{k,\epsilon}}(\alpha) = \mathbf{y}$. Entonces, el supuesto (iii) en el Teorema 2.2.2 se satisface.

De (3.1), tenemos que, para todo ϵ fijo, $\phi_{k,\epsilon}$ maximiza $\Upsilon_{x,\epsilon}(\alpha) = \sigma_{\mathbb{R}}^2(P_{x,\epsilon}[\alpha])$ sobre la superficie $\{\alpha : \Psi(\alpha) = 0\}$, eso es, $\Upsilon_{x,\epsilon}(\alpha)$ tiene un máximo local en $\phi_{k,\epsilon}$ sujeto a la condición $\Psi(\alpha) = 0$. Además, $\Upsilon_{x,\epsilon} : \mathcal{H} \rightarrow \mathbb{R}$ es una función Hadamard diferenciable.

Luego, teniendo presente que $\Upsilon_{x,\epsilon}$ depende de ϵ y llamando, para ϵ fijo, $\Lambda(\alpha) = \Upsilon(\epsilon, \alpha) = \Upsilon_{x,\epsilon}(\alpha)$, tenemos que, el Teorema 2.2.2 implica que existen $\gamma_0, \gamma_1, \dots, \gamma_{k-1} \in \mathbb{R}$ (dependiendo de ϵ) tales que

$$\Lambda'_{\phi_{k,\epsilon}} = \sum_{j=0}^{k-1} \gamma_j \Psi'_{j,\phi_{k,\epsilon}}.$$

Vale la pena notar que $\Lambda'_{\phi_{k,\epsilon}} = D_{2,\theta_\epsilon} \Upsilon$ con $\theta_\epsilon = (\epsilon, \phi_{k,\epsilon}) \in [0, \epsilon_0] \times \mathcal{H}$ donde $D_{2,\theta} \Upsilon$ indica la primer derivada Hadamard con respecto a la segunda componente de Υ en $\theta = (\epsilon, \alpha)$. Sea $I_{\mathcal{H}}$ el operador identidad, o sea, $I_{\mathcal{H}}(\alpha) = \alpha$.

Además, del hecho que $\Lambda'_{\phi_{k,\epsilon}}$ es un operador lineal y continuo (ver la Observación 2.2.1), tenemos que existe un único $\nabla \Lambda_{\phi_{k,\epsilon}} \in \mathcal{H}$ tal que $\Lambda'_{\phi_{k,\epsilon}}(\alpha) = \langle \nabla \Lambda_{\phi_{k,\epsilon}}, \alpha \rangle$ y similarmente, $\Psi'_{j,\phi_{k,\epsilon}}(\alpha) = \langle \nabla \Psi_{j,\phi_{k,\epsilon}}, \alpha \rangle$. Luego, $\nabla \Lambda_{\phi_{k,\epsilon}} = \sum_{j=0}^{k-1} \gamma_j \nabla \Psi_{j,\phi_{k,\epsilon}}$. Por otra parte, por la observación 2.2.2, $\nabla \Psi_{0,\phi_{k,\epsilon}} = 2\phi_{k,\epsilon}$ mientras $\nabla \Psi_{j,\phi_{k,\epsilon}} = \phi_{j,\epsilon}$, para $1 \leq j \leq k-1$ lo que implica que

$$\nabla \Lambda_{\phi_{k,\epsilon}} = 2\gamma_0 \phi_{k,\epsilon} + \sum_{j=1}^{k-1} \gamma_j \phi_{j,\epsilon}. \quad (4.3)$$

De (4.3), obtenemos que $2\gamma_0 = \langle \nabla \Lambda_{\phi_{k,\epsilon}}, \phi_{k,\epsilon} \rangle$ mientras que $\gamma_j = \langle \nabla \Lambda_{\phi_{k,\epsilon}}, \phi_{j,\epsilon} \rangle$. Luego, (4.3) puede ser escrito como

$$\nabla \Lambda_{\phi_{k,\epsilon}} = \sum_{j=1}^k \langle \nabla \Lambda_{\phi_{k,\epsilon}}, \phi_{j,\epsilon} \rangle \phi_{j,\epsilon}, \quad (4.4)$$

o sea,

$$(D_{2,\theta_\epsilon} \Upsilon)(\alpha) = \Lambda'_{\phi_{k,\epsilon}} \alpha = \sum_{j=1}^k \langle \nabla \Lambda_{\phi_{k,\epsilon}}, \phi_{j,\epsilon} \rangle \langle \phi_{j,\epsilon}, \alpha \rangle \quad (4.5)$$

Usaremos esta expresión varias veces en la demostración del Teorema 4.1.1.

Precisaremos el siguiente lema, que es una extensión de un resultado análogo que se obtiene en el contexto multivariado.

Lema 4.1.1. *Sea P una medida de probabilidad elíptica $P = \mathcal{E}(\mu, \Gamma)$, con $\mu = 0$ y Γ un operador auto-adjunto, semi-definido positivo y compacto con autovalores $\lambda_1 \geq \lambda_2 \geq \dots$ y sea ϕ_j la autofunción asociada a λ_j . Supongamos que $\lambda_1 > \lambda_2 > \dots > \lambda_q > \lambda_{q+1}$.*

Sea F_0 la medida univariada definida en el Lema 2.1.2. Supongamos que la aplicación $S : [0, 1] \times \mathbb{R} \rightarrow \mathbb{R}$ definida por $S(\epsilon, y) = \sigma_{\mathbb{R}}((1-\epsilon)F_0 + \epsilon\delta_y)$ es dos veces continuamente diferenciable en todo $(0, y)$. En particular, $IF(y; \sigma_{\mathbb{R}}, F_0)$ debe ser diferenciable con respecto a y y su derivada se escribirá como $IF_1(y; \sigma_{\mathbb{R}}, F_0)$.

Luego, para todo $k \leq q$, existe una vecindad \mathcal{U}_k de ϕ_k tal que para todo $\alpha \in \mathcal{U}_k$,

$$IF(\langle \alpha, x \rangle; \sigma_{\mathbb{R}}^2, P[\alpha]) = \langle \alpha, \Gamma \alpha \rangle IF\left(\frac{\langle \alpha, x \rangle}{\sqrt{\langle \alpha, \Gamma \alpha \rangle}}; \sigma_{\mathbb{R}}^2, F_0\right). \quad (4.6)$$

Más aún, si $\Lambda : \mathcal{H} \rightarrow \mathbb{R}$ es la aplicación $\Lambda(\alpha) = \text{IF}(\langle \alpha, x \rangle; \sigma_{\mathbb{R}}^2, P[\alpha])$, se tiene que Λ es Hadamard diferenciable en ϕ_k y

$$\nabla \Lambda_{\phi_k} = 2\lambda_k \text{IF} \left(\frac{\langle \phi_k, x \rangle}{\sqrt{\lambda_k}}; \sigma_{\mathbb{R}}^2, F_0 \right) \phi_k + \lambda_k \text{IF}_1 \left(\frac{\langle \phi_k, x \rangle}{\sqrt{\lambda_k}}; \sigma_{\mathbb{R}}^2, F_0 \right) \left(\frac{x - \langle x, \phi_k \rangle \phi_k}{\sqrt{\lambda_k}} \right), \quad (4.7)$$

o sea,

$$\Lambda'_{\phi_k}(\alpha) = 2\lambda_k \text{IF} \left(\frac{\langle \phi_k, x \rangle}{\sqrt{\lambda_k}}; \sigma_{\mathbb{R}}^2, F_0 \right) \langle \phi_k, \alpha \rangle + \lambda_k \text{IF}_1 \left(\frac{\langle \phi_k, x \rangle}{\sqrt{\lambda_k}}; \sigma_{\mathbb{R}}^2, F_0 \right) \left\langle \alpha, \frac{x - \langle x, \phi_k \rangle \phi_k}{\sqrt{\lambda_k}} \right\rangle.$$

DEMOSTRACIÓN. Usando que $\lambda_k > 0$, obtenemos que existe un vecindario \mathcal{U}_k de ϕ_k tal que para todo $\alpha \in \mathcal{U}_k$, $\alpha \notin \ker(\mathbf{\Gamma})$, luego el Lema 2.1.2 implica que $Z_\alpha = \langle \alpha, X \rangle / \sqrt{\langle \alpha, \mathbf{\Gamma} \alpha \rangle} \sim F_0$. Recordemos que

$$\text{IF}(\langle \alpha, x \rangle; \sigma_{\mathbb{R}}^2, P[\alpha]) = \lim_{\epsilon \rightarrow 0} \frac{\sigma_{\mathbb{R}}^2((1-\epsilon)P[\alpha] + \epsilon\delta_{\langle x, \alpha \rangle}) - \sigma_{\mathbb{R}}^2(P[\alpha])}{\epsilon}. \quad (4.8)$$

Usando que $\sigma_{\mathbb{R}}$ es un funcional de escala y el hecho que $\langle \alpha, X \rangle \sim P[\alpha]$, obtenemos que $\sigma_{\mathbb{R}}(P[\alpha]) = \sqrt{\langle \alpha, \mathbf{\Gamma} \alpha \rangle} \sigma_{\mathbb{R}}(F_0)$.

Por otro lado, consideremos $P_{x,\epsilon}[\alpha] = (1-\epsilon)P[\alpha] + \epsilon\delta_{\langle x, \alpha \rangle}$. Recordando como se construía $P_{x,\epsilon}$, usando la variable Bernoulli V y usando que $(1-V)\langle X, \alpha \rangle + V\langle x, \alpha \rangle = \sqrt{\langle \alpha, \mathbf{\Gamma} \alpha \rangle} ((1-V)Z_\alpha + Vz_\alpha)$ con $z_\alpha = \langle \alpha, x \rangle / \sqrt{\langle \alpha, \mathbf{\Gamma} \alpha \rangle}$ y que $(1-V)Z_\alpha + Vz_\alpha \sim (1-\epsilon)F_0 + \epsilon\delta_{z_\alpha}$, obtenemos que

$$\sigma_{\mathbb{R}}((1-\epsilon)P[\alpha] + \epsilon\delta_{\langle x, \alpha \rangle}) = \sqrt{\langle \alpha, \mathbf{\Gamma} \alpha \rangle} \sigma_{\mathbb{R}}((1-\epsilon)F_0 + \epsilon\delta_{z_\alpha}),$$

con lo cual por (4.8) y la equivarianza del estimador de escala, obtenemos (4.6).

Falta ver que (4.7) vale. Definamos las aplicaciones $\Upsilon : \mathcal{U}_k \rightarrow \mathbb{R}$ y $\Psi : \mathcal{U}_k \rightarrow \mathbb{R}$ como $\Upsilon(\alpha) = \langle \alpha, \mathbf{\Gamma} \alpha \rangle$ y

$$\Psi(\alpha) = \text{IF} \left(\frac{\langle \alpha, x \rangle}{\sqrt{\langle \alpha, \mathbf{\Gamma} \alpha \rangle}}; \sigma_{\mathbb{R}}^2, F_0 \right).$$

Luego, $\Lambda(\alpha) = \text{IF}(\langle \alpha, x \rangle; \sigma_{\mathbb{R}}^2, P[\alpha]) = \Upsilon(\alpha)\Psi(\alpha)$. La Observación 2.2.2 implica que Υ es Hadamard diferenciable con $\nabla \Upsilon_\alpha = 2\mathbf{\Gamma}\alpha$. Luego, si probamos que Ψ es Hadamard diferenciable en ϕ_k , obtenemos que Λ es también Hadamard diferenciable ϕ_k y

$$\nabla \Lambda_{\phi_k} = \Upsilon(\phi_k)\nabla \Psi_{\phi_k} + \Psi(\phi_k)\nabla \Upsilon_{\phi_k} \quad (4.9)$$

Usando que ϕ_k es la autofunción de $\mathbf{\Gamma}$ asociada a λ_k , obtenemos que $\Upsilon(\phi_k) = \lambda_k$, $\nabla \Upsilon_{\phi_k} = 2\mathbf{\Gamma}\phi_k = 2\lambda_k$ y

$$\Psi(\phi_k) = \text{IF} \left(\frac{\langle \phi_k, x \rangle}{\sqrt{\lambda_k}}; \sigma_{\mathbb{R}}^2, F_0 \right),$$

que junto con (4.9) implican que

$$\nabla \Lambda_{\phi_k} = \lambda_k \nabla \Psi_{\phi_k} + 2\lambda_k \text{IF} \left(\frac{\langle \phi_k, x \rangle}{\sqrt{\lambda_k}}; \sigma_{\mathbb{R}}^2, F_0 \right) \phi_k.$$

Luego, para concluir la demostración, queda probar que Ψ es Hadamard diferenciable en ϕ_k y

$$\nabla \Psi_{\phi_k} = \text{IF}_1 \left(\frac{\langle \phi_k, x \rangle}{\sqrt{\lambda_k}}; \sigma_R^2, F_0 \right) \left(\frac{x - \langle x, \phi_k \rangle \phi_k}{\sqrt{\lambda_k}} \right).$$

Definamos $f : \mathbb{R} \rightarrow \mathbb{R}$ como $f(y) = \text{IF}(y; \sigma_R, F_0)$ y $\Phi : \mathcal{U}_k \rightarrow \mathbb{R}$ como $\Phi(\alpha) = \langle x, \alpha \rangle / \sqrt{\langle \alpha, \Gamma \alpha \rangle} = \langle x, \alpha \rangle / \sqrt{\Upsilon(\alpha)}$. Entonces, la diferenciability Hadamard de Ψ en ϕ_k se sigue fácilmente de la regla de la cadena y del hecho que f es diferenciable, Φ es Hadamard diferenciable en ϕ_k , puesto que $\Upsilon(\phi_k) = \lambda_k \neq 0$, y $\Psi = f \circ \Phi$. Más aún, la regla de la cadena implica que $\Psi'_{\phi_k} = f'_{\Phi(\phi_k)} \circ \Phi'_{\phi_k}$, que junto con $f'_{\Phi(\phi_k)}(v) = f'(\Phi(\phi_k))v$, $\Phi'(\phi_k) = \langle \phi_k, x \rangle / \sqrt{\lambda_k}$ y $f'(y) = \text{IF}_1(y; \sigma_R, F_0)$ implican que

$$\begin{aligned} \Psi'_{\phi_k}(\alpha) &= f'_{\Phi(\phi_k)}(\Phi'_{\phi_k}(\alpha)) = f'(\Phi(\phi_k))\Phi'_{\phi_k}(\alpha) \\ &= \text{IF}_1 \left(\frac{\langle \phi_k, x \rangle}{\sqrt{\lambda_k}}; \sigma_R, F_0 \right) \Phi'_{\phi_k}(\alpha) \end{aligned}$$

Usando nuevamente la Observación 2.2.2 y el hecho que $\Gamma \phi_k = \lambda_k$, obtenemos que

$$\Phi'_{\phi_k}(\alpha) = \frac{\langle x, \alpha \rangle \sqrt{\lambda_k} - \frac{\langle \alpha, \Gamma \phi_k \rangle}{\sqrt{\lambda_k}} \langle x, \phi_k \rangle}{\lambda_k} = \frac{\langle x, \alpha \rangle - \langle \phi_k, \alpha \rangle \langle x, \phi_k \rangle}{\sqrt{\lambda_k}} = \left\langle \frac{x - \langle x, \phi_k \rangle \phi_k}{\sqrt{\lambda_k}}, \alpha \right\rangle,$$

concluyendo la demostración. \square

DEMOSTRACIÓN DEL TEOREMA 4.1.1. Probaremos la existencia de la función de influencia. Comenzaremos probando que existe la función de influencia del funcional $\phi_{R,1}(P)$ asociado a la primer dirección principal.

Recordemos que $\Upsilon : [0, \epsilon_0] \times \mathcal{H} \rightarrow \mathbb{R}_+$ estaba definido como $\Upsilon(\epsilon, \alpha) = \sigma_R^2(P_{x,\epsilon}[\alpha])$. Tenemos que $D_{2,(\epsilon, \phi_{1,\epsilon})} \Upsilon : \mathcal{H} \rightarrow \mathbb{R}$ definido por (4.5) cumple $(D_{2,(\epsilon, \phi_{1,\epsilon})} \Upsilon)(h) = \langle \nabla \Lambda_{\phi_{1,\epsilon}, \phi_{1,\epsilon}} \rangle \langle \phi_{1,\epsilon}, h \rangle$. Pero, $(D_{2,(\epsilon, \phi_{1,\epsilon})} \Upsilon)(\phi_{1,\epsilon}) = \langle \nabla \Lambda_{\phi_{1,\epsilon}, \phi_{1,\epsilon}} \rangle$ de donde

$$(D_{2,(\epsilon, \phi_{1,\epsilon})} \Upsilon)(h) = (D_{2,(\epsilon, \phi_{1,\epsilon})} \Upsilon)(\phi_{1,\epsilon}) \langle \phi_{1,\epsilon}, h \rangle.$$

Definimos la aplicación $\mathcal{L} : [0, \epsilon_0] \times \mathcal{H} \rightarrow B(\mathcal{H}, \mathbb{R})$ como sigue:

$$\mathcal{L}(\epsilon, \alpha) = D_{2,(\epsilon, \alpha)} \Upsilon - (D_{2,(\epsilon, \alpha)} \Upsilon)(\alpha) \langle \alpha, \cdot \rangle = D_{2,(\epsilon, \alpha)} \Upsilon - u(\epsilon, \alpha).$$

donde $u(\epsilon, \alpha)(h) = (D_{2,(\epsilon, \alpha)} \Upsilon)(\alpha) \langle \alpha, h \rangle$.

Notemos que $\mathcal{L}(\epsilon, \phi_{1,\epsilon}) = 0$, en particular tenemos que $\mathcal{L}(0, \phi_1) = 0$. Para aplicar el teorema de la función implícita, debemos probar que $D_{2,(0, \phi_1)} \mathcal{L} : \mathcal{H} \rightarrow B(B(\mathcal{H}, \mathbb{R}), \mathbb{R})$ es un isomorfismo, puesto que ya sabemos que Υ es dos veces diferenciable.

Observemos que $D_{2,(0, \phi_1)} \mathcal{L} = D_{22,(0, \phi_1)}^2 \Upsilon - D_{2,(0, \phi_1)} u$, $\Upsilon(0, \alpha) = \langle \alpha, \Gamma \alpha \rangle$, luego $D_{2,(0, \alpha)} \Upsilon = 2\langle \Gamma \alpha, \cdot \rangle$. Por lo tanto, $D_{22,(0, \phi_1)}^2 \Upsilon = 2\Gamma$ y

$$u(0, \alpha) = (D_{2,(0, \alpha)} \Upsilon)(\alpha) \langle \alpha, \cdot \rangle = 2\langle \alpha, \Gamma \alpha \rangle \langle \alpha, \cdot \rangle.$$

Por lo tanto, $u(0, \alpha) = 2 \Upsilon(0, \alpha) \langle \alpha, \cdot \rangle = \langle 2 \Upsilon(0, \alpha) \alpha, \cdot \rangle = \langle \tilde{u}(\alpha), \cdot \rangle$, de donde

$$D_{2,(0,\alpha)}u(h) = 4\langle \Gamma\alpha, h \rangle \alpha + 2\langle \alpha, \Gamma\alpha \rangle h.$$

y deducimos que $D_{2,(0,\phi_1)}(u)(h) = 4\lambda_1 \langle \phi_1, h \rangle \phi_1 + 2\lambda_1 h$. Entonces,

$$\begin{aligned} D_{2,(0,\phi_1)}\mathcal{L}(h) &= 2\Gamma h - 4\lambda_1 \langle \phi_1, h \rangle \phi_1 - 2\lambda_1 h \\ &= 2 \left\{ \sum_{j \geq 1} \lambda_j \phi_j \otimes \phi_j - 2\lambda_1 \phi_1 \otimes \phi_1 - \lambda_1 I_{\mathcal{H}} \right\} h. \end{aligned}$$

Usando que el operador identidad $I_{\mathcal{H}}$ es igual a $\sum_{j \geq 1} \phi_j \otimes \phi_j$, obtenemos

$$T = \frac{1}{2} D_{2,(0,\phi_1)}\mathcal{L} = -2\lambda_1 \phi_1 \otimes \phi_1 - \sum_{j \geq 2} (\lambda_1 - \lambda_j) \phi_j \otimes \phi_j.$$

Como $\lambda_1 > 0$ y $\lambda_1 - \lambda_j > 0$ para $j \geq 2$, es fácil ver entonces que T es un monomorfismo. Falta ver que T es un epimorfismo, o sea que para todo $y \in \mathcal{H}$, existe $x \in \mathcal{H}$ tal que $Tx = y$.

Primero, veamos que

$$\sum_{j \geq 2} \frac{1}{(\lambda_1 - \lambda_j)^2} \langle y, \phi_j \rangle^2 < \infty.$$

En efecto, $\lambda_1 - \lambda_j \geq \lambda_1 - \lambda_2 > 0$ para $j \geq 2$, luego

$$\sum_{j \geq 2} \frac{1}{(\lambda_1 - \lambda_j)^2} \langle y, \phi_j \rangle^2 \leq \frac{1}{(\lambda_1 - \lambda_2)^2} \|y\|^2 < \infty.$$

Definamos

$$x = \frac{-1}{2\lambda_1} \langle y, \phi_1 \rangle \phi_1 - \sum_{j \geq 2} \frac{1}{(\lambda_1 - \lambda_j)} \langle y, \phi_j \rangle \phi_j.$$

Es fácil ver que $x \in \mathcal{H}$ (eso es, tiene norma finita) y que $Tx = y$. Entonces, T es un isomorfismo y por ende, podemos aplicar el teorema de la función implícita en la ecuación $\mathcal{L}(0, \phi_1) = 0$, obteniendo que existe $\frac{\partial}{\partial \epsilon} \phi_{1,\epsilon} |_{\epsilon=0}$.

Ahora probaremos la existencia de la función de influencia del funcional $\phi_{R,k}$. Procederemos inductivamente, supongamos que $\phi_{1,\epsilon}, \dots, \phi_{k-1,\epsilon}$ son diferenciables a ϵ en $\epsilon = 0$. Queremos ver que lo mismo pasa con $\phi_{k,\epsilon}$.

Teníamos que por la ortogonalidad entre $\phi_{j,\epsilon}$ y $\phi_{\ell,\epsilon}$ para $\ell \neq j$, de donde

$$D_{2,(\epsilon,\phi_{k,\epsilon})}\Upsilon(h) = \left(D_{2,(\epsilon,\phi_{k,\epsilon})}\Upsilon \right) (\phi_{k,\epsilon}) \langle \phi_{k,\epsilon}, h \rangle + \sum_{j=1}^{k-1} \left(D_{2,(\epsilon,\phi_{k,\epsilon})}\Upsilon \right) (\phi_{j,\epsilon}) \langle \phi_{j,\epsilon}, h \rangle.$$

Sea $\mathcal{L} : [0, \epsilon_0] \times \mathcal{H} \rightarrow \mathcal{B}(\mathcal{H}, \mathbb{R})$ definida como

$$\mathcal{L}(\epsilon, \alpha) = D_{2,(\epsilon, \alpha)} \Upsilon - u_0(\epsilon, \alpha) - \sum_{j=1}^{k-1} u_j(\epsilon, \alpha)$$

con $u_0(\epsilon, \alpha) = (D_{2,(\epsilon, \alpha)} \Upsilon)(\alpha) \langle \alpha, \cdot \rangle$ y $u_j(\epsilon, \alpha) = (D_{2,(\epsilon, \alpha)} \Upsilon)(\phi_{j, \epsilon}) \langle \phi_{j, \epsilon}, \cdot \rangle$.

Luego $\mathcal{L}(\epsilon, \phi_{k, \epsilon}) = \mathcal{L}(0, \phi_k) = 0$ y \mathcal{L} es diferenciable. Para aplicar el teorema de la función implícita basta ver que $D_{2,(0, \phi_k)} \mathcal{L} : \mathcal{H} \rightarrow \mathcal{B}(\mathcal{B}(\mathcal{H}, \mathbb{R}), \mathbb{R})$ es un isomorfismo.

Observemos que

$$D_{2,(0, \phi_k)} \mathcal{L} = D_{22,(0, \phi_k)}^2 \Upsilon - D_{2,(0, \phi_k)} u_0 - \sum_{j=1}^{k-1} D_{2,(0, \phi_k)} u_j.$$

donde $D_{22,(0, \phi_k)}^2 \Upsilon = 2\mathbf{\Gamma}$, $D_{2,(0, \phi_k)} u_0(h) = 4\lambda_k \langle \phi_k, h \rangle \phi_k + 2\lambda_k h$ y

$$D_{2,(0, \phi_k)} u_j = \left(D_{22,(0, \phi_k)}^2 \Upsilon \right) (\phi_j) \langle \phi_j, \cdot \rangle = 2\mathbf{\Gamma} \phi_j \langle \phi_j, \cdot \rangle = 2\lambda_j \phi_j \langle \phi_j, \cdot \rangle.$$

Por lo tanto,

$$(D_{2,(0, \phi_k)} \mathcal{L})(h) = 2\mathbf{\Gamma} h - (4\lambda_k \langle \phi_k, h \rangle \phi_k + 2\lambda_k h) - 2 \sum_{j=1}^{k-1} \lambda_j \phi_j \langle \phi_j, h \rangle.$$

O sea,

$$D_{2,(0, \phi_k)} \mathcal{L} = 2\mathbf{\Gamma} - 4\lambda_k \phi_k \otimes \phi_k - 2\lambda_k I_{\mathcal{H}} - 2 \sum_{j=1}^{k-1} \lambda_j \phi_j \otimes \phi_j.$$

Definamos,

$$T = \frac{1}{2} D_{2,(0, \phi_k)} \mathcal{L} = \sum_{s \geq 1} \lambda_s \phi_s \otimes \phi_s - 2\lambda_k \phi_k \otimes \phi_k - \lambda_k \sum_{s \geq 1} \phi_s \otimes \phi_s - \sum_{j=1}^{k-1} \lambda_j \phi_j \otimes \phi_j.$$

Entonces, se tiene

$$T = - \sum_{s > k} (\lambda_k - \lambda_s) \phi_s \otimes \phi_s - 2\lambda_k (\phi_k \otimes \phi_k) - \lambda_k \sum_{s=1}^{k-1} \phi_s \otimes \phi_s$$

y argumentando como antes, T será inversible puesto que $\lambda_k - \lambda_s > \lambda_k - \lambda_{k+1} > 0$.

Hemos probado pues la existencia de la función de influencia. Probaremos ahora (4.1).

Recordemos que habíamos supuesto que el funcional de escala $\sigma_{\mathbb{R}}$ estaba calibrado de forma tal que $c = 1$, es decir, $\sigma_{\mathbb{R}}(F_0) = 1$, tenemos que $\sigma(\alpha) = \langle \alpha, \mathbf{\Gamma} \alpha \rangle$.

Recordemos también que

$$(D_{2,\theta_\epsilon} \Upsilon)\alpha = \Lambda'_{\phi_k, \epsilon} \alpha = \sum_{j=1}^k \langle \nabla \Lambda_{\phi_k, \epsilon}, \phi_{j, \epsilon} \rangle \langle \phi_{j, \epsilon}, \alpha \rangle.$$

Notemos que como $\Upsilon(\epsilon, \alpha) = \sigma_{\mathbb{R}}^2(P_{x, \epsilon}[\alpha])$ es dos veces continuamente Hadamard diferenciable en $(0, \phi_k)$, obtenemos que $\nabla \Upsilon_{\epsilon, \phi_k, \epsilon}$ es diferenciable con respecto a ϵ , en $\epsilon = 0$, donde $\nabla \Upsilon_{\epsilon, \phi_k, \epsilon} \in \mathcal{H}$ es tal que $\Upsilon'_{\epsilon, \phi_k, \epsilon}(\alpha) = \langle \nabla \Upsilon_{\epsilon, \phi_k, \epsilon}, \alpha \rangle$ con $\Upsilon'_{\epsilon, \alpha} = D_{2, (\epsilon, \alpha)} \Upsilon$.

Definamos ahora la función $f : [0, \epsilon_0] \rightarrow \mathcal{H}$ como $f(\epsilon) = \nabla \Lambda_{\phi_k, \epsilon} = \nabla \Upsilon_{\epsilon, \phi_k, \epsilon}$. Los argumentos anteriores implican que f es diferenciable con respecto a ϵ , en $\epsilon = 0$. A fines de simplificar, llamemos $f'(0) = f'_0 \in \mathcal{H}$ su derivada en 0. Usando que $h_j(\epsilon) = \phi_{j, \epsilon}$ y $g_j(\epsilon) = \langle \nabla \Lambda_{\phi_k, \epsilon}, \phi_{j, \epsilon} \rangle = \langle f(\epsilon), h_j(\epsilon) \rangle$ también son diferenciables con respecto a ϵ , en $\epsilon = 0$, y (4.4), obtenemos que $f'(0) = \sum_{j=1}^k g'_j(0)h_j(0) + g_j(0)h'_j(0)$. Usando las Observaciones 2.2.2 y 2.2.3, tenemos que

$$\begin{aligned} h'_j(0) &= \text{IF}(x; \phi_{\mathbb{R}, j}, P) \\ g'_j(0) &= \langle f'(0), h_j(0) \rangle + \langle f(0), h'_j(0) \rangle = \langle f'(0), \phi_j \rangle + \langle f(0), \text{IF}(x; \phi_{\mathbb{R}, j}, P) \rangle. \end{aligned}$$

Observando que $f(0) = \nabla \Upsilon_{0, \phi_k} = 2\lambda_k \phi_k$, ya que $\Upsilon(0, \alpha) = \sigma_{\mathbb{R}}^2(P[\alpha]) = \langle \alpha, \Gamma \alpha \rangle$, y que $g_j(0) = 2\langle \phi_k, \phi_j \rangle$ obtenemos

$$\begin{aligned} f'(0) &= \sum_{j=1}^k g'_j(0)h_j(0) + g_j(0)h'_j(0) \\ &= \sum_{j=1}^k \{ [\langle f'(0), \phi_j \rangle + 2\lambda_k \langle \phi_k, \text{IF}(x; \phi_{\mathbb{R}, j}, P) \rangle] \phi_j + 2\lambda_k \langle \phi_k, \phi_j \rangle \text{IF}(x; \phi_{\mathbb{R}, j}, P) \} \end{aligned}$$

De $\langle \phi_k, \phi_j \rangle = 0$ para $j \neq k$, se deduce que

$$f'(0) = 2\lambda_k \text{IF}(x; \phi_{\mathbb{R}, k}, P) + \sum_{j=1}^k [\langle f'(0), \phi_j \rangle + 2\lambda_k \langle \phi_k, \text{IF}(x; \phi_{\mathbb{R}, j}, P) \rangle] \phi_j.$$

Indiquemos por \mathcal{L}_k el subespacio lineal generado por $\phi_1 \dots \phi_k$ y por π_k la proyección ortogonal sobre \mathcal{L}_k^\perp . Luego, $\pi_k = I_{\mathcal{H}} - \sum_{j=1}^k \phi_j \otimes \phi_j$. Entonces,

$$\pi_k f'(0) = 2\lambda_k \text{IF}(x; \phi_{\mathbb{R}, k}, P) + \sum_{j=1}^k 2\lambda_k \langle \phi_k, \text{IF}(x; \phi_{\mathbb{R}, j}, P) \rangle \phi_j. \quad (4.10)$$

Para calcular $f'(0)$, consideremos $\epsilon_1 < \epsilon_0$ y \mathcal{U}_k una vecindad de ϕ_k tal que $\Upsilon(\epsilon, \alpha)$ es continuamente Hadamard diferenciable en $(\epsilon, \alpha) \in [0, \epsilon_1] \times \mathcal{U}_k$. Definamos $\Phi : [0, \epsilon_1] \times \mathcal{U}_k \rightarrow \mathcal{H}^*$ como $\Phi(\epsilon, \alpha) = D_{2, \theta} \Upsilon = \Upsilon'_{\epsilon, \alpha}$ con $\theta = (\epsilon, \alpha)$ y $g : [0, \epsilon_1] \rightarrow \mathcal{H}^*$ como $g(\epsilon) = \Phi(\epsilon, \phi_{k, \epsilon})$. Observemos que $g(\epsilon)$

puede ser identificado con $\nabla \Upsilon_{\epsilon, \phi_k, \epsilon} = f(\epsilon)$. Luego, por la Observación 2.2.3 (d), obtenemos que $g'_0 \in \mathcal{H}^{**} = \mathcal{H}$ y

$$g'_0 = D_{1,(0,\phi_k)}\Phi + (D_{2,(0,\phi_k)}\Phi)\text{IF}(x; \phi_{R,k}, P).$$

Usando nuevamente la Observación 2.2.2, obtenemos que $D_{2,(0,\phi_k)}\Phi = D_{22,(0,\phi_k)}^2\Upsilon = 2\mathbf{\Gamma}$ puesto que $\Upsilon(0, \alpha) = \sigma_R^2(P[\alpha]) = \langle \alpha, \mathbf{\Gamma}\alpha \rangle$. Por otro lado, para todo $(\epsilon, \alpha) \in \mathbb{R} \times \mathcal{H}$, $(D_{1,(0,\phi_k)}\Phi)(\epsilon, \alpha) = D_{12,(0,\phi_k)}^2\Upsilon(\epsilon, \alpha)$. Usando que Υ es dos veces continuamente diferenciable en $(0, \phi_k)$ deducimos que

$$D_{12,(0,\phi_k)}^2\Upsilon(\epsilon, \alpha) = D_{21,(0,\phi_k)}^2\Upsilon(\alpha, \epsilon) = (D_{2,(0,\phi_k)}\xi)(\epsilon, \alpha)$$

con

$$\xi(\epsilon, \alpha) = D_{1,(\epsilon,\alpha)}\Upsilon \tag{4.11}$$

Observando que $\xi(0, \alpha) = D_{1,(0,\alpha)}\Upsilon = \text{IF}(\langle \alpha, x \rangle; \sigma_R^2, P)$ y usando el Lema 4.1.1 obtenemos que

$$D_{12,(0,\phi_k)}^2\Upsilon(\epsilon, \alpha) = \epsilon \left[2\lambda_k \text{IF} \left(\frac{\langle \phi_k, x \rangle}{\sqrt{\lambda_k}}; \sigma_R^2, F_0 \right) \langle \phi_k, \alpha \rangle + \lambda_k \text{IF}_1 \left(\frac{\langle \phi_k, x \rangle}{\sqrt{\lambda_k}}; \sigma_R^2, F_0 \right) \left\langle \alpha, \frac{x - \langle x, \phi_k \rangle \phi_k}{\sqrt{\lambda_k}} \right\rangle \right],$$

lo que nos lleva a

$$\begin{aligned} (g'_0(\epsilon))(\alpha) &= (D_{1,(0,\phi_k)}\Phi)(\epsilon, \alpha) + \epsilon \langle (D_{2,(0,\phi_k)}\Phi)\text{IF}(x; \phi_{R,k}, P), \alpha \rangle \\ &= \epsilon \langle \nabla_k, \alpha \rangle + \epsilon \langle 2\mathbf{\Gamma} \text{IF}(x; \phi_{R,k}, P), \alpha \rangle \\ &= \epsilon \langle \nabla_k + 2\mathbf{\Gamma} \text{IF}(x; \phi_{R,k}, P), \alpha \rangle \end{aligned}$$

donde

$$\nabla_k = 2\lambda_k \text{IF} \left(\frac{\langle \phi_k, x \rangle}{\sqrt{\lambda_k}}; \sigma_R^2, F_0 \right) \phi_k + \lambda_k \text{IF}_1 \left(\frac{\langle \phi_k, x \rangle}{\sqrt{\lambda_k}}; \sigma_R^2, F_0 \right) \left(\frac{x - \langle x, \phi_k \rangle \phi_k}{\sqrt{\lambda_k}} \right).$$

Observando que $(g'_0(\epsilon))(\alpha) = \epsilon \langle f'(0), \alpha \rangle$, obtenemos que $f'(0) = 2\mathbf{\Gamma} \text{IF}(x; \phi_{R,k}, P) + \nabla_k$, de donde al reemplazarlo en (4.10) se deduce que

$$\pi_k (2\mathbf{\Gamma} \text{IF}(x; \phi_{R,k}, P) + \nabla_k) = 2\lambda_k \text{IF}(x; \phi_{R,k}, P) + \sum_{j=1}^k 2\lambda_k \langle \phi_k, \text{IF}(x; \phi_{R,j}, P) \rangle \phi_j.$$

Usando que $\pi_k \phi_k = 0$, tenemos que $\pi_k \nabla_k = \sqrt{\lambda_k} \text{IF}_1 \left(\frac{\langle \phi_k, x \rangle}{\sqrt{\lambda_k}}; \sigma_R^2, F_0 \right) \pi_k x$. Entonces,

$$\pi_k \mathbf{\Gamma} \text{IF}(x; \phi_{R,k}, P) - \lambda_k \text{IF}(x; \phi_{R,k}, P) = -\frac{\sqrt{\lambda_k}}{2} \text{IF}_1 \left(\frac{\langle \phi_k, x \rangle}{\sqrt{\lambda_k}}; \sigma_R^2, F_0 \right) \pi_k x + \sum_{j=1}^k \lambda_k \langle \phi_k, \text{IF}(x; \phi_{R,j}, P) \rangle \phi_j,$$

luego

$$(\pi_k \mathbf{\Gamma} - \lambda_k I_{\mathcal{H}}) \text{IF}(x; \phi_{R,k}, P) = -\frac{\sqrt{\lambda_k}}{2} \text{IF}_1 \left(\frac{\langle \phi_k, x \rangle}{\sqrt{\lambda_k}}; \sigma_R^2, F_0 \right) \pi_k x + \sum_{j=1}^k \lambda_k \langle \phi_k, \text{IF}(x; \phi_{R,j}, P) \rangle \phi_j. \tag{4.12}$$

Notemos que para todo $\alpha \in \mathcal{H}$, $\pi_k \Gamma \alpha = \Gamma \alpha - \sum_{j=1}^k \langle \Gamma \alpha, \phi_j \rangle \phi_j = \Gamma \alpha - \sum_{j=1}^k \lambda_j \langle \alpha, \phi_j \rangle \phi_j = \sum_{j=k+1}^{\infty} \lambda_j \langle \alpha, \phi_j \rangle \phi_j$, luego

$$\pi_k \Gamma - \lambda_k I_{\mathcal{H}} = \sum_{j=k+1}^{\infty} (\lambda_j - \lambda_k) \phi_j \otimes \phi_j - \lambda_k \sum_{j=1}^k \phi_j \otimes \phi_j$$

Definamos $\tau_k : \mathcal{H} \rightarrow \mathcal{H}$ como

$$\tau_k = \sum_{j \geq k+1} \frac{1}{\lambda_j - \lambda_k} \phi_j \otimes \phi_j - \frac{1}{\lambda_k} \sum_{j=1}^k \phi_j \otimes \phi_j = \tilde{\tau}_k - \frac{1}{\lambda_k} \sum_{j=1}^k \phi_j \otimes \phi_j.$$

Debemos ver que la aplicación está bien definida, es decir, que $\|\tau_k(\alpha)\| < \infty$ para todo $\alpha \in \mathcal{H}$. Como $\sum_{j=1}^k \phi_j \otimes \phi_j$ claramente tiene rango finito, basta ver que $\|\tilde{\tau}_k(\alpha)\| < \infty$. Primero observemos que, como $k \leq q$, $\lambda_k > \lambda_{k+1}$ entonces, $\lambda_k - \lambda_j \geq \lambda_k - \lambda_{k+1} = M_k > 0$, para $j \geq k+1$, luego, $1/(\lambda_k - \lambda_j) \leq 1/M_k$. Así pues, para todo $N > k+1$ tenemos que

$$\begin{aligned} \left\| \sum_{j=k+1}^N \frac{1}{\lambda_k - \lambda_j} \langle \phi_j, \alpha \rangle \phi_j \right\|^2 &= \sum_{j=k+1}^N \frac{1}{(\lambda_k - \lambda_j)^2} \langle \phi_j, \alpha \rangle^2 \\ &\leq \frac{1}{M_k^2} \sum_{j=k+1}^N \langle \phi_j, \alpha \rangle^2 \leq \frac{1}{M_k^2} \|\alpha\|^2. \end{aligned}$$

Entonces, $\tilde{\tau}_k(\alpha) \in \mathcal{H}$. Falta probar que $\tau_k \circ (\pi_k \Gamma - \lambda_k I_{\mathcal{H}})(\alpha) = \alpha$. Efectivamente,

$$\begin{aligned} \tau_k \circ (\pi_k \Gamma - \lambda_k I_{\mathcal{H}})(\alpha) &= \tau_k \left(\sum_{j=k+1}^{\infty} (\lambda_j - \lambda_k) \langle \phi_j, \alpha \rangle \phi_j - \lambda_k \sum_{j=1}^k \langle \phi_j, \alpha \rangle \phi_j \right) \\ &= \sum_{j=k+1}^{\infty} (\lambda_j - \lambda_k) \langle \phi_j, \alpha \rangle \tau_k \phi_j - \lambda_k \sum_{j=1}^k \langle \phi_j, \alpha \rangle \tau_k \phi_j. \end{aligned}$$

Notando que $\tau_k \phi_j = \phi_j / (\lambda_j - \lambda_k)$ para $j \geq k+1$ y $\tau_k \phi_j = -\phi_j / \lambda_k$ cuando $j \leq k$, obtenemos que

$$\begin{aligned} \tau_k \circ (\pi_k \Gamma - \lambda_k I_{\mathcal{H}})(\alpha) &= \sum_{j=k+1}^{\infty} (\lambda_j - \lambda_k) \langle \phi_j, \alpha \rangle \frac{1}{\lambda_j - \lambda_k} \phi_j - \lambda_k \sum_{j=1}^k \langle \phi_j, \alpha \rangle \left(-\frac{1}{\lambda_k}\right) \phi_j \\ &= \sum_{j \geq 1} \langle \phi_j, \alpha \rangle \phi_j = \alpha. \end{aligned}$$

Luego, aplicando τ_k en ambos lados de (4.12), deducimos que

$$\text{IF}(x; \phi_{R,k}, P) = \tau_k \left(-\frac{\sqrt{\lambda_k}}{2} \text{IF}_1 \left(\frac{\langle \phi_k, x \rangle}{\sqrt{\lambda_k}}; \sigma_R^2, F_0 \right) \pi_k x + \sum_{j=1}^k \lambda_k \langle \phi_k, \text{IF}(x; \phi_{R,j}, P) \rangle \phi_j \right). \quad (4.13)$$

Usando que $1 = \|\phi_{1,\epsilon}\|^2 = \langle \phi_{k,\epsilon}, \phi_{k,\epsilon} \rangle$ y la regla de la cadena, tenemos que $\langle \text{IF}(x; \phi_{R,k}, P), \phi_k \rangle = 0$. Luego, usando que $\tau_k \phi_j = -\phi_j/\lambda_k$ cuando $j \leq k$ y que $\tau_k \circ \pi_k = \tilde{\pi}_k$, (4.13) se puede reescribir como

$$\begin{aligned} \text{IF}(x; \phi_{R,k}, P) &= \tau_k \left(-\frac{\sqrt{\lambda_k}}{2} \text{IF}_1 \left(\frac{\langle \phi_k, x \rangle}{\sqrt{\lambda_k}}; \sigma_R^2, F_0 \right) \pi_k x + \sum_{j=1}^{k-1} \lambda_k \langle \phi_k, \text{IF}(x; \phi_{R,j}, P) \rangle \phi_j \right) \\ &= -\frac{\sqrt{\lambda_k}}{2} \text{IF}_1 \left(\frac{\langle \phi_k, x \rangle}{\sqrt{\lambda_k}}; \sigma_R^2, F_0 \right) \tilde{\pi}_k x - \sum_{j=1}^{k-1} \langle \phi_k, \text{IF}(x; \phi_{R,j}, P) \rangle \phi_j \\ &= \frac{\sqrt{\lambda_k}}{2} \text{IF}_1 \left(\frac{\langle \phi_k, x \rangle}{\sqrt{\lambda_k}}; \sigma_R^2, F_0 \right) \sum_{j=k+1}^{\infty} \frac{\langle \phi_j, x \rangle}{\lambda_k - \lambda_j} \phi_j - \sum_{j=1}^{k-1} \langle \phi_k, \text{IF}(x; \phi_{R,j}, P) \rangle \phi_j. \end{aligned} \quad (4.14)$$

Entonces, usando que $\text{IF}(x; \phi_{R,k}, P) = \sum_{j \geq 1} \langle \text{IF}(x; \phi_{R,k}, P), \phi_j \rangle \phi_j$ y (4.14) se obtiene que $\langle \text{IF}(x; \phi_{R,k}, P), \phi_k \rangle = 0$ y para $j \geq k+1$

$$\langle \text{IF}(x; \phi_{R,k}, P), \phi_j \rangle = \frac{\sqrt{\lambda_k}}{2} \text{IF}_1 \left(\frac{\langle \phi_k, x \rangle}{\sqrt{\lambda_k}}; \sigma_R^2, F_0 \right) \frac{\langle \phi_j, x \rangle}{\lambda_k - \lambda_j}, \quad (4.15)$$

mientras que para $j < k$ tenemos que

$$\begin{aligned} \langle \text{IF}(x; \phi_{R,k}, P), \phi_j \rangle &= -\langle \phi_k, \text{IF}(x; \phi_{R,j}, P) \rangle = -\frac{\sqrt{\lambda_j}}{2} \text{IF}_1 \left(\frac{\langle \phi_j, x \rangle}{\sqrt{\lambda_j}}; \sigma_R^2, F_0 \right) \frac{\langle \phi_k, x \rangle}{\lambda_j - \lambda_k} \\ &= \frac{\sqrt{\lambda_j}}{2} \text{IF}_1 \left(\frac{\langle \phi_j, x \rangle}{\sqrt{\lambda_j}}; \sigma_R^2, F_0 \right) \frac{\langle \phi_k, x \rangle}{\lambda_k - \lambda_j}, \end{aligned}$$

donde hemos usado (4.15). La demostración de (4.1) se concluye del hecho que $\text{IF}_1(a; \sigma_R^2, F_0) = 2 \text{IF}_1(a; \sigma_R, F_0)$.

Queda por probar (4.2). Indiquemos por $\lambda_{k,\epsilon} = \lambda_{R,k}(P_{x,\epsilon})$. Luego, $h(\epsilon) = \lambda_{k,\epsilon} = \Upsilon(\epsilon, \phi_{k,\epsilon})$. La regla de la cadena implica que

$$h'(0) = \text{IF}(x; \lambda_{R,k}, P) = D_{1,(0,\phi_k)} \Upsilon + (D_{2,(0,\phi_k)} \Upsilon) \text{IF}(x; \phi_{R,k}, P).$$

Recordemos que $\Lambda'_{\phi_{k,\epsilon}} = D_{2,(\epsilon,\phi_{k,\epsilon})} \Upsilon$ y $\Lambda'_{\phi_{k,\epsilon}}(\alpha) = \langle \nabla \Lambda_{\phi_{k,\epsilon}}, \alpha \rangle$ donde $f(\epsilon) = \nabla \Lambda_{\phi_{k,\epsilon}}$ es tal que $f(0) = \nabla \Lambda_{\phi_k} = 2\lambda_k \phi_k$. Entonces, $\langle \text{IF}(x; \phi_{R,k}, P), \phi_k \rangle = 0$ implica que $(D_{2,(0,\phi_k)} \Upsilon) \text{IF}(x; \phi_{R,k}, P) = 2\lambda_k \langle \text{IF}(x; \phi_{R,k}, P), \phi_k \rangle = 0$.

Por otro lado (4.11) implica que $D_{1,(0,\phi_k)} \Upsilon = \xi(0, \phi_k)$. Ya teníamos que $\xi(0, \alpha) = \text{IF}(\langle \alpha, x \rangle; \sigma_R^2, P)$ luego, $\text{IF}(x; \lambda_{R,k}, P) = \text{IF}(\langle x, \phi_k \rangle; \sigma_R^2, P)$. La prueba se concluye ahora usando el Lema 4.1.1 y el hecho que $\text{IF}(a; \sigma_R^2, F_0) = 2 \text{IF}(a; \sigma_R, F_0)$. \square

4.2 La función de influencia para estimadores basados en una M -escala

En esta sección, damos una expresión para la función de influencia de los funcionales $\phi_{R,k}$ y $\lambda_{R,k}$ para distribuciones Gaussianas, en el caso particular en que el funcional de escala sea un M -estimador. En particular, la expresión incluye el desvío estándar y el M -funcional de escala asociado a la función bicuadrada de Tukey.

Sea $X \sim P$ un proceso Gaussiano con media 0 y operador de covarianza $\mathbf{\Gamma}_X = \sum_{j \geq 1} \lambda_j \phi_j \otimes \phi_j$ donde $\lambda_1 \geq \lambda_2 \geq \dots$ y $\{\phi_j\}_{j \in \mathbb{N}}$ es una base ortonormal de \mathcal{H} . Luego, ϕ_j representan las componentes principales de X y por la expansión de Karhunen–Loève, $X = \sum_{j \geq 1} Z_j \phi_j$, con $Z_j \sim N(0, \lambda_j)$ independientes.

Además, para todo α tal que $\alpha \notin \ker(\mathbf{\Gamma}_X)$, tenemos que $Z_\alpha = \langle \alpha, X - \mu \rangle / \sqrt{\langle \alpha, \mathbf{\Gamma}_X \alpha \rangle} \sim N(0, 1)$. Por lo tanto, la distribución F_0 del Lema 2.1.2 es igual a la distribución normal estándar Φ . En este caso, el funcional definido a través de (3.1) es Fisher-consistente, o sea $\phi_{R,k}(P) = \phi_k$, mientras que $\lambda_{R,k} = \lambda_k$ si el funcional de escala es Fisher-consistente en $F_0 = \Phi$.

Por otro lado, la función de influencia de los estimadores de escala $\sigma_R(G)$ definidos en (2.3.2) está dado por

$$\text{IF}(x; \sigma_R, G) = \sigma_R(G) \left\{ \chi \left(\frac{x - \mu(G)}{\sigma_R(G)} - b \right) \right\} \left\{ \int \chi' \left(\frac{y - \mu(G)}{\sigma_R(G)} \right) \frac{y - \mu(G)}{\sigma_R(G)} dG(y) \right\}^{-1}$$

PDe donde deducimos que $\text{IF}_1(x; \sigma_R, G)$ está dada por

$$\text{IF}_1(x; \sigma_R, G) = \chi' \left(\frac{x - \mu(G)}{\sigma_R(G)} \right) \left\{ \int \chi' \left(\frac{y - \mu(G)}{\sigma_R(G)} \right) \frac{y - \mu(G)}{\sigma_R(G)} dG(y) \right\}^{-1},$$

con χ' la derivada de χ . Usualmente, $\sigma_R(G)$ está calibrado de forma tal de tener la Fisher-consistencia en la distribución normal, es decir, $\sigma_R(\Phi) = 1$ y el funcional de posición es también Fisher-consistente, o sea $\mu(\Phi) = 0$. Entonces, si notamos por $Y \sim N(0, 1)$, tenemos que la expresión de arriba se reduce a

$$\begin{aligned} \text{IF}(x; \sigma_R, F_0) &= \{\chi(x) - b\} \{\mathbb{E}_\Phi \chi'(Y) Y\}^{-1} \\ \text{IF}_1(x; \sigma_R, F_0) &= \chi'(x) \{\mathbb{E}_\Phi \chi'(Y) Y\}^{-1}. \end{aligned}$$

Por ejemplo, cuando consideramos el desvío estándar, usando (4.1), tenemos que

$$\begin{aligned} \text{IF}(x; \phi_{\text{SD},k}, P) &= \left(\sum_{j \neq k} \frac{1}{\lambda_k - \lambda_j} \langle \phi_j, x \rangle \langle x, \phi_k \rangle \phi_j \right) \\ \text{IF}(x; \lambda_{\text{SD},k}, P) &= \langle x, \phi_k \rangle^2 - \lambda_k. \end{aligned}$$

Como se mencionó antes, si x is colineal con ϕ_k se cumple que $\text{IF}(x; \phi_{\text{SD},k}, P) = 0$. Además, cualquier contaminación puntual en la dirección de la dirección principal llevará a una influencia

de cero. Por lo tanto, para poder producir modificaciones significativas en los estimadores clásicos es necesario que la contaminación se propague en varias componentes. Las expresiones anteriores son la versión funcional de las expresiones obtenidas por Croux y Ruiz-Gazen (2005) en el caso finito-dimensional y permiten definir reglas de detección de *outliers*.

Como ejemplo, consideremos $X = Z_1\phi_1 + Z_2\phi_2 + Z_3\phi_3$ donde $\{\phi_k\}_{k \geq 1}$ es una base ortonormal y Z_j son variables aleatorias independientes $Z_j \sim N(0, \lambda_j)$, $\lambda_1 = 16$, $\lambda_2 = 4$ y $\lambda_3 = 1$. Las Figuras 4.1 y 4.2 dan los gráficos de $\|\text{IF}(x; \phi_{R,1}, P)\|^2$ en función de $x_1 = \langle \phi_1, x \rangle$ y $x_2 = \langle \phi_2, x \rangle$ cuando $x_3 = \langle \phi_3, x \rangle$ vale 0 y 3, respectivamente. Se consideraron el estimador clásico, basado en el desvío estándar y el robusto, calculado con una M -escala robusta utilizando la función bicuadrada de Tukey con $c = 1.56$ y punto de ruptura $1/2$. Como se esperaba, estos gráficos son análogos a los que se obtienen en el contexto finito-dimensional. La forma de la función de influencia para el estimador de *projection-pursuit* basado en la M -escala con función de Tukey es comparable al estimador clásico en el centro de la distribución. Las observaciones lejanas del centro de la distribución tienen una influencia mucho menor al usar el estimador robusto, incluso si, para las autofunciones, $\|\text{IF}(x; \phi_{R,1}, P)\|^2$ pueden alcanzar valores altos, pero solamente para valores reducidos de x_1 combinados con valores altos de x_2 . Por otro lado, el gráfico confirma que $\text{IF}(x; \lambda_{R,k}, P)$ estará acotada cuando se usa una escala robusta.

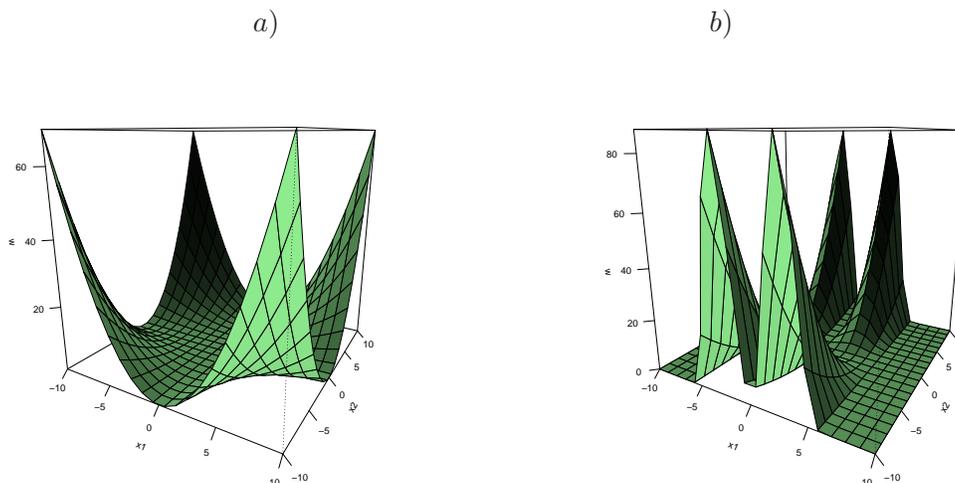


Figura 4.1: Norma de la función de influencia de la autofunción principal para (a) σ_R el desvío estándar y (b) σ_R siendo un M -estimador de escala con función de Tukey, ambos bajo $x_3 = 0$.

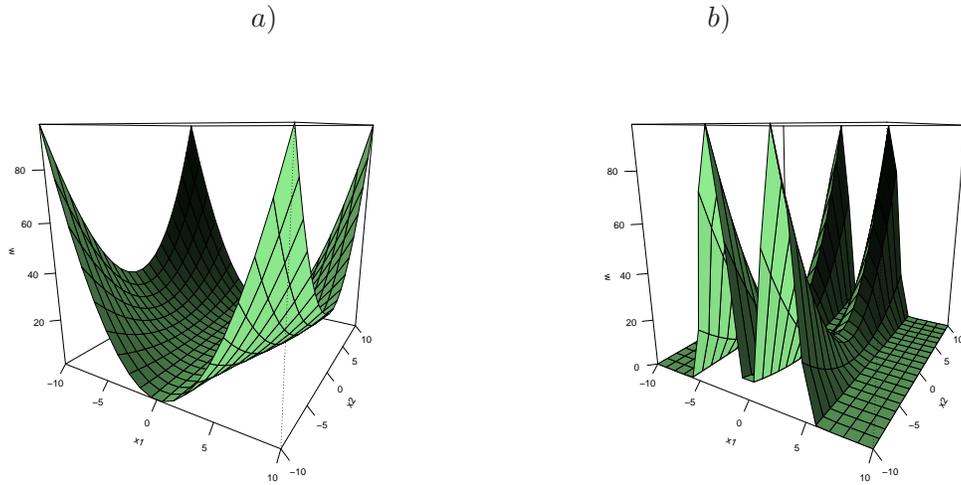


Figura 4.2: Norma de la función de influencia de la autofunción principal para (a) σ_R el desvío estándar y (b) σ_R siendo un M -estimador de escala con función de Tukey, ambos bajo $x_3 = 4$.

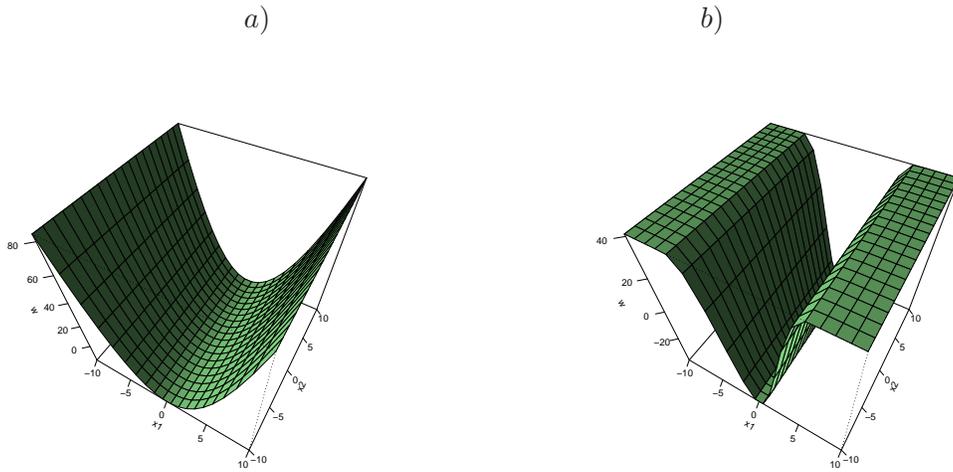


Figura 4.3: Función de influencia para el autovalor más grande para (a) σ_R el desvío estándar y (b) σ_R siendo un M -estimador de escala con función de Tukey, ambos bajo $x_3 = 0$.

Capítulo 5

Estimadores robustos bajo un modelo de Componentes Principales Comunes funcionales

5.1 Introducción al modelo

En este capítulo, se generalizarán los estimadores propuestos en el Capítulo 3 al caso en donde consideramos más de una población. En diversas situaciones, tendremos observaciones independientes $X_{i,1}, \dots, X_{i,n_i}$ provenientes de k poblaciones independientes de elementos aleatorios de un espacio de Hilbert \mathcal{H} con media μ_i y distintos operadores de covarianza $\mathbf{\Gamma}_i$. Sin embargo, como en el caso finito-dimensional, los operadores de covarianza podrían exhibir una estructura en común que es conveniente usar a la hora de efectuar estimaciones. Una generalización simple consiste en suponer que son todos proporcionales, es decir, que $\mathbf{\Gamma}_i = \rho_i \mathbf{\Gamma}_1$, para $1 \leq i \leq k$ y $\rho_1 = 1$.

El modelo de componentes principales comunes, introducido por Flury (1984) para datos finito-dimensionales, generaliza el modelo proporcional suponiendo que las matrices de covarianza tienen los mismos autovectores pero, eventualmente, podrían tener distintos autovalores. Este modelo puede ser visto como una generalización del modelo de componentes principales a k grupos porque provee una diagonalización simultánea para todas las matrices de covarianza. En aplicaciones biométricas, las componentes principales suelen ser interpretadas como factores independientes que determinan el crecimiento o la forma de un organismo. Parece pues razonable considerar un modelo común en donde los mismos factores surgen en especies diferentes pero relacionadas. El modelo de componentes principales comunes claramente sirve para este propósito.

Una extensión natural al caso funcional del modelo de componentes principales comunes es suponer que los operadores de covarianza $\mathbf{\Gamma}_i$ tienen autofunciones comunes ϕ_j pero diferentes autovalores $\lambda_{i,j}$. Es decir,

$$\mathbf{\Gamma}_i = \sum_{j=1}^{\infty} \lambda_{i,j} \phi_j \otimes \phi_j . \quad (5.1)$$

Este modelo fue definido para dos poblaciones por Benko y Hrdle (2009) y posteriormente por Boente *et al.* (2010) para el caso de k poblaciones. Como en Boente *et al.* (2010), vamos a suponer que los autovalores preservan el orden entre las poblaciones, es decir, $\lambda_{i,1} \geq \lambda_{i,2} \geq \dots \geq \lambda_{i,j} \geq \lambda_{i,j+1} \dots$, para $1 \leq i \leq k$, de modo de poder identificar las direcciones sin necesidad de suponer que los operadores $\mathbf{\Gamma}_i$ tengan rango finito.

En este sentido, el proceso $X_{i,1}$, $1 \leq i \leq k$, se podrá escribir como

$$X_{i,1} = \mu_i + \sum_{j=1}^{\infty} \lambda_{i,j}^{\frac{1}{2}} Z_{ij} \phi_j$$

con $\lambda_{i1} \geq \lambda_{i2} \geq \dots \geq 0$ y Z_{ij} variables aleatorias tales que $\mathbb{E}(Z_{ij}) = 0$, $\mathbb{E}(Z_{ij}^2) = 1$, $\mathbb{E}(Z_{ij} Z_{is}) = 0$ para $j \neq s$ y entonces las autofunciones comunes, como en el caso unipoblacional, exhiben los mismos modos mayores de variación en todas las poblaciones. Este modelo es usualmente denominado el modelo de componentes principales comunes funcionales (CPCF). Como en el caso del análisis de componentes principales, el modelo CPCF puede ser utilizado para reducir la dimensionalidad de los datos, guardando la mayor cantidad posible de variabilidad presente en cada población. Además, este modelo provee un marco de trabajo para el análisis de datos provenientes de diferentes poblaciones que comparten sus modos principales de variación ϕ_1, ϕ_2, \dots . Cabe aclarar que cuando se considera un modelo proporcional funcional, $X_{i,1}$, $1 \leq i \leq k$, pueden ser escritos como $X_{i,1} = \mu_i + \rho_i^{\frac{1}{2}} \sum_{j=1}^{\infty} \lambda_j^{\frac{1}{2}} Z_{ij} \phi_j$, con $\rho_1 = 1$, $\lambda_1 \geq \lambda_2 \geq \dots \geq 0$ y Z_{ij} variables aleatorias como se describió arriba. Un problema similar fue estudiado por Benko *et al.* (2009) quienes consideraron el caso de $k = 2$ poblaciones y propusieron tests para igualdad de medias y de una cantidad fija de autofunciones. Boente *et al.* (2010) también trabajaron en esta línea, proponiendo estimadores bajo un modelo CPCF general. Todos estas propuestas se basan en el operador de covarianza muestral de cada población y por lo tanto, no son resistentes a observaciones atípicas.

El objetivo, entonces, es proveer estimadores robustos de las direcciones principales bajo un modelo CPCF y estudiar su comportamiento asintótico. Primero introduciremos la notación básica, similar a la que ya se introdujo para el caso de una población, obviando como antes la necesidad de tener que contar con condiciones sobre los segundos momentos. Posteriormente, presentaremos propuestas robustas de estimación de las direcciones comunes a través de un método de projection-pursuit con diferentes criterios de suavizado. Probaremos la consistencia fuerte de cada una de estas propuestas.

5.2 Notación y preliminares

Consideraremos k muestras de n_i elementos aleatorios en \mathcal{H} , o sea, $X_{i,1}, \dots, X_{i,n_i}$ son independientes y tales que $X_{i,j} \sim X_{i,1} \sim P_i$. Sea $N = \sum_{i=1}^k n_i$ y $\hat{\tau}_i = n_i/N$. Supondremos que $\hat{\tau}_i \rightarrow \tau_i$ con $0 < \tau_i < 1$, para $1 \leq i \leq k$, y $\sum_{i=1}^k \tau_i = 1$.

Definición 5.2.1. Diremos que P_1, \dots, P_k son *débilmente-CPCF* bajo $\sigma_{\mathbf{R}}$ si $\phi_{\mathbf{R},j}(P_i) = \phi_{\mathbf{R},j}(P_m)$ (excepto por un cambio de signo) para todo $j \geq 1$ y $1 \leq i, m \leq k$, donde $\phi_{\mathbf{R},j}(P)$ está definido en

(3.1).

Diremos que P_1, \dots, P_k son *fuertemente*-CPCF si existe una constante $c > 0$ y operadores auto-adjuntos, semidefinidos positivo y compactos Γ_i tal que para todo $\alpha \in \mathcal{H}$, tenemos $\sigma_{\mathbb{R}}^2(P_i[\alpha]) = c_i \langle \alpha, \Gamma_i \alpha \rangle$, donde Γ_i satisface (5.1), es decir, $\Gamma_i = \sum_{j=1}^{\infty} \lambda_{i,j} \phi_j \otimes \phi_j$, con $\lambda_{i,1} \geq \lambda_{i,2} \geq \dots \geq \lambda_{i,j} \geq \lambda_{i,j+1} \dots$, para $1 \leq i \leq k$. Luego, Γ_i tienen las mismas autofunciones y en el mismo orden según los autovalores, que podrían diferir en cada población.

Claramente, fuertemente-FCPC implica débilmente-FCPC. Más aún, si existen los segundos momentos y los operadores de covarianza satisfacen (5.1), luego, P_1, \dots, P_k son fuertemente-FCPC si $\sigma_{\mathbb{R}}$ es el desvío estándar.

Es claro que, cuando las diferentes observaciones de cada población están elípticamente distribuidas (es decir, $X_{i,1} \sim \mathcal{E}(\mu_i, \Gamma_i)$), con operadores de dispersion Γ_i satisfaciendo (5.1), entonces por el Lema 2.3.1, P_1, \dots, P_k son fuertemente-FCPC para cualquier escala $\sigma_{\mathbb{R}}$.

Las direcciones principales pueden ser estimadas aplicando una versión muestral de (3.1) para cada población. Sin embargo, en la mayoría de los casos, los estimadores obtenidos para cada población darán distintos resultados aún bajo el modelo FCPC. Luego, es necesario un acercamiento unificado.

Definamos $\sigma_i : \mathcal{H} \rightarrow [0, +\infty)$ como $\sigma(\alpha) = \sigma_{\mathbb{R}}(P_i[\alpha])$. Notemos que si $c \in \mathbb{R}$, usando que $\sigma_{\mathbb{R}}$ es un funcional de escala obtenemos que $\sigma_i^2(c\alpha) = c^2 \sigma_i^2(\alpha)$. Como en Boente *et al.* (2006), sea $f : \mathbb{R}_{>0} \rightarrow \mathbb{R}_{\geq 0}$ una función creciente de *score* y definamos $\varsigma(\alpha) = \sum_{i=1}^k \tau_i f(\sigma_i^2(\alpha))$. Más aún, indiquemos por P a la medida producto $P = P_1 \times \dots \times P_k$.

Es claro que cuando P_1, \dots, P_k son débilmente-FCPC bajo $\sigma_{\mathbb{R}}$, entonces, para todo $1 \leq i \leq k$, $\phi_{\mathbb{R},1}(P_i)$ maximizará $\sum_{i=1}^k \tau_i \sigma_{\mathbb{R}}^2(P_i[\alpha])$ sobre $\mathcal{S} = \{\alpha \in \mathcal{H} : \|\alpha\| = 1\}$. Más generalmente, maximizará $\varsigma(\alpha)$ sobre \mathcal{S} . Esto motiva la definición de los funcionales de direcciones comunes principales por *projection pursuit* como sigue:

$$\left\{ \begin{array}{l} \phi_{f,1}(P) = \underset{\|\alpha\|=1}{\operatorname{argmax}} \varsigma(\alpha) = \underset{\|\alpha\|=1}{\operatorname{argmax}} \sum_{i=1}^k \tau_i f(\sigma_{\mathbb{R}}^2(P_i[\alpha])) \\ \phi_{f,m}(P) = \underset{\|\alpha\|=1, \alpha \in \mathcal{B}_{f,m}}{\operatorname{argmax}} \varsigma(\alpha) = \underset{\|\alpha\|=1, \alpha \in \mathcal{B}_{f,m}}{\operatorname{argmax}} \sum_{i=1}^k \tau_i f(\sigma_{\mathbb{R}}^2(P_i[\alpha])) \quad 2 \leq m, \end{array} \right. \quad (5.2)$$

donde $\mathcal{B}_{f,m} = \{\alpha \in \mathcal{H} : \langle \alpha, \phi_{f,j}(P) \rangle = 0, 1 \leq j \leq m-1\}$. También definimos los funcionales de valores principales robustos como

$$\lambda_{f,i,m}(P) = \sigma_{\mathbb{R}}^2(P_i[\phi_{f,m}(P)]) = \sigma_i^2(\phi_{f,m}(P)). \quad (5.3)$$

Los siguientes lemas prueban la buena definición de los funcionales

Lema 5.2.1. *Si $f : [0, \infty] \rightarrow \mathbb{R}$ es una función continua y σ_i es débilmente continuo, entonces $\sup_{\|\alpha\|=1} \varsigma(\alpha)$ se alcanza para algún $\alpha \in \mathcal{S}$ y por ende, el funcional está bien definido.*

DEMOSTRACIÓN. Usando que f es una función continua y que σ_i es débilmente continua, obtenemos fácilmente que $f \circ \sigma_i : \mathcal{H} \rightarrow \mathbb{R}$ y $\varsigma : \mathcal{H} \rightarrow \mathbb{R}$ son funciones continuas con respecto a la topología débil en \mathcal{H} . El resultado ahora sigue del hecho que la bola unitaria $\{\|\alpha\| = 1\}$ es débilmente compacta, puesto que toda función continua alcanza su máximo en un conjunto compacto. \square

Notemos que la conclusión del lema anterior todavía vale cuando consideramos $\sup_{\|\alpha\|=1, \alpha \in \mathcal{B}_{f,m}} \varsigma(\alpha)$ puesto que seguimos maximizando en un conjunto débilmente compacto.

El siguiente lema asegura la existencia de $\phi_{f,1}(P)$ cuando $f = \log$.

Lema 5.2.2. *Si $f = \log$, σ_i es una función débilmente continua y existe α_0 tal que $\sigma_i(\alpha_0) > 0$ para todo i , entonces $\sup_{\|\alpha\|=1} \varsigma(\alpha)$ se alcanza para algún $\alpha \in \mathcal{S}$ y por ende, el funcional está bien definido.*

DEMOSTRACIÓN. Sea $\alpha_n \in \mathcal{S}$ una sucesión tal que $\varsigma(\alpha_n) \rightarrow \sup_{\|\alpha\|=1} \varsigma(\alpha)$.

Vamos a empezar probando que $\liminf_{n \rightarrow \infty} \sigma_i(\alpha_n) = d_i > 0$ para todo $1 \leq i \leq k$. Supongamos que esto no vale, es decir, que existe un i y una subsucesión α_{n_j} tal que $\sigma_i(\alpha_{n_j}) \rightarrow 0$ cuando $n_j \rightarrow \infty$. Luego, $\varsigma(\alpha_{n_j}) \rightarrow \sum_{i=1}^k \tau_i f(\sigma_i^2(0)) = -\infty$ de donde $\sup_{\|\alpha\|=1} \varsigma(\alpha) = -\infty$. Por otro lado, usando que existe α_0 tal que $\sigma_i(\alpha_0) > 0$ para todo i , obtenemos $-\infty < \varsigma(\alpha_0/\|\alpha_0\|) \leq \sup_{\|\alpha\|=1} \varsigma(\alpha)$ llevándonos a una contradicción.

Luego, $\liminf_{n \rightarrow \infty} \sigma_i(\alpha_n) = d_i > 0$ para todo i . Sin pérdida de generalidad, supongamos que $\sigma_i(\alpha_n) \rightarrow d_i$. Luego, existirá n_0 tal que para $n \geq n_0$, tenemos que $\sigma_i(\alpha_n) > d_i/2 > 0$. Podemos suponer que $\sigma_i(\alpha_n) > A > 0$ para todo i y n sin pérdida de generalidad.

Usando que \mathcal{S} es débilmente compacto tendremos que existirá una subsucesión α_{n_m} convergiendo a $\beta \in \mathcal{H}$ y $\|\beta\| \leq 1$. Probemos que $\|\beta\| \neq 0$. Si $\beta = 0$, tenemos que $\sigma_i(\beta) = 0$. Pero, la continuidad débil de σ_i implica que $\sigma_i(\beta) = \lim_m \sigma_i(\alpha_{n_m}) > A > 0$ llevándonos a una contradicción. Luego $\|\beta\| \neq 0$.

Entonces, $\alpha_{n_m} \rightarrow \beta$ y $\sigma_i(\alpha_{n_m}) > A$ para todo i y n . Como $f = \log : [A, \infty] \rightarrow \mathbb{R}$ es una función continua, obtenemos que $\varsigma(\alpha_{n_m}) \rightarrow \varsigma(\beta)$. Por otro lado, $\varsigma(\alpha_{n_m}) \rightarrow \sup_{\|\alpha\|=1} \varsigma(\alpha)$, luego $\sup_{\|\alpha\|=1} \varsigma(\alpha) = \varsigma(\beta)$. Basta ver que $\|\beta\| = 1$. Supongamos que $\|\beta\| < 1$ y definamos $\gamma = \beta/\|\beta\|$. Tenemos que $\log(\|\beta\|^2) < 0$ y entonces

$$\varsigma(\gamma) = \varsigma(\beta) - \sum_{i=1}^k \tau_i \log(\|\beta\|^2) > \varsigma(\beta) = \sup_{\|\alpha\|=1} \varsigma(\alpha)$$

lo que resulta una contradicción. Así pues, $\|\beta\| = 1$ y el supremo se alcanza en β . \square

Las mismas ideas se pueden utilizar para probar que $\sup_{\|\alpha\|=1, \alpha \in \mathcal{B}_{f,m}} \varsigma(\alpha)$ se alcanza, requiriendo que exista $\alpha \in \mathcal{B}_{f,m}$ tal que $\sigma_i(\alpha) > 0$ para todo i .

5.3 Consistencia Fisher

Es fácil ver que si las medidas son débilmente-FCPC los pesos τ_i y la función de *score* no jugarán un papel mayor a la hora de definir $\phi_{f,j}(P)$, como muestra el siguiente lema.

Lema 5.3.1. Sean $\tau_i \geq 0$, $\sum_{i=1}^k \tau_i = 1$ y $f : \mathbb{R} \rightarrow \mathbb{R}$ una función estrictamente creciente. Supongamos que P_1, \dots, P_k son débilmente-FCPC bajo σ_R . Entonces, $\phi_{R,j}(P_1)$ cumple (5.2).

DEMOSTRACIÓN. El hecho que P_1, \dots, P_k sean débilmente-FCPC bajo σ_R implica que $\phi_{R,j}(P_i) = \phi_{R,j}(P_m) = \phi_{R,j}$ para todo $j \geq 1$ y $1 \leq i, m \leq k$. Entonces, para cualquier $\alpha \in \mathcal{S}$, tenemos que $\sigma_R^2(P_i[\phi_{R,1}]) \geq \sigma_R^2(P_i[\alpha])$ que junto con el hecho que f es estrictamente creciente nos lleva a que $f(\sigma_R^2(P_i[\phi_{R,1}])) \geq f(\sigma_R^2(P_i[\alpha]))$. Luego, $\varsigma(\alpha) \leq \varsigma(\phi_{R,1})$, lo que implica que $\phi_{f,1} = \phi_{R,1}$.

La demostración ahora sigue por un argumento inductivo. Supongamos que $\phi_{f,j}(P) = \phi_{R,j}(P_1)$ for $1 \leq j \leq m$, queremos ver que $\phi_{f,m+1}(P) = \phi_{R,m+1}(P_1)$. Primero notemos que $\mathcal{B}_{m+1} = \mathcal{B}_{f,m+1}$, luego para todo $\alpha \in \mathcal{S} \cap \mathcal{B}_{f,m+1}$, tenemos que $\sigma_i(\alpha) \leq \sigma_i(\phi_{R,m+1})$, de donde $\varsigma(\alpha) \leq \varsigma(\phi_{R,m+1})$, concluyendo la prueba. \square

Vale la pena notar que el resultado anterior no asegura unicidad de la solución. Un punto importante a destacar es lo que representan las funciones $\phi_{f,m}$, al menos en algunas situaciones particulares. Esta sección se enfoca en demostrar que, para familiar elípticamente distribuidas, los funcionales $\phi_{f,m}(P)$ y $\lambda_{f,i,m}(P)$ tienen una interpretación simple. En particular, nuestros resultados valen para el caso de una familia elíptica, pero no están restringidas a ellas.

El próximo lema establece la consistencia Fisher de los funcionales definidos mediante (5.2) bajo el siguiente supuesto

A1. Existe una constante $c_i > 0$ y un operador auto-adjunto, semidefinido positivo y compacto $\mathbf{\Gamma}_{i,0}$, tal que para todo α , tenemos $\sigma_i^2(\alpha) = c_i \langle \alpha, \mathbf{\Gamma}_{i,0} \alpha \rangle$.

Notemos que **A1** implica que la función $\sigma_i : \mathcal{H} \rightarrow \mathbb{R}$ definida como $\sigma_i(\alpha) = \sigma_R(P_i[\alpha])$ será débilmente continua, independientemente de si σ_R lo es.

Lema 5.3.2. Sea $\phi_{f,m}$ y $\lambda_{f,i,m}$ los funcionales definidos en (5.2) y (5.3), respectivamente. Sea $X_{i,1} \sim P_i$ elementos aleatorios tales que **A1** vale. Supongamos que $\mathbf{\Gamma}_{i,0}$ tiene las mismas autofunciones para todo i . Más aún, supongamos que $\mathbf{\Gamma}_{i,0}$ satisface (5.1). Indiquemos por $\lambda_{i,j}$ al autovalor de $\mathbf{\Gamma}_{i,0}$ relacionado a la autofunción ϕ_j , tal que $\lambda_{i,1} \geq \lambda_{i,2} \geq \dots$. Supongamos que para algún $1 \leq i_0 \leq k$ existe $q \geq 2$ tal que para todo $1 \leq j \leq q$, $\lambda_{i_0,1} > \lambda_{i_0,2} > \dots > \lambda_{i_0,q} > \lambda_{i_0,q+1}$. Luego, si f es una función estrictamente creciente y $\tau_{i_0} > 0$, tenemos que, para todo $1 \leq j \leq q$, $\phi_{f,j}(P)$ es única, $\phi_{f,j}(P) = \phi_j$ y $\lambda_{f,j}(P) = c_i \lambda_{i,j}$.

DEMOSTRACIÓN. Empezamos probando el resultado para $j = 1$. Notemos que **A1** implica que P_1, \dots, P_k son fuertemente-FCPC y por lo tanto, débilmente-FCPC bajo σ_R . Luego, para todo $\alpha \in \mathcal{S}$, tenemos que $\sigma_R^2(P_i[\phi_1]) = \sigma_i^2(\phi_1) = c_i \lambda_{i,1} \geq c_i \langle \alpha, \mathbf{\Gamma}_0 \alpha \rangle = \sigma_i^2(\alpha) = \sigma_R^2(P_i[\alpha])$ y la desigualdad es estricta cuando $i = i_0$ y $\alpha \neq \phi_1$. Entonces, $f(\sigma_{i_0}^2(\phi_1)) \geq f(\sigma_{i_0}^2(\alpha))$ para todo $1 \leq i \leq k$ y $f(\sigma_{i_0}^2(\phi_1)) > f(\sigma_{i_0}^2(\alpha))$ ya que f es estrictamente creciente, que junto con el hecho que $\tau_i \geq 0$ y $\tau_{i_0} > 0$ implican que $\varsigma(\phi_1) > \varsigma(\alpha)$ para todo $\alpha \in \mathcal{S}$, $\alpha \neq \phi_1$. Luego, $\phi_{f,1}(P) = \phi_1$.

La demostración ahora sigue fácilmente mediante un argumento inductivo. Supongamos que $\phi_{f,j}(P) = \phi_j$, para $1 \leq j \leq m-1$, $m \leq q$, queremos ver que $\phi_{f,m}(P) = \phi_m$. El conjunto $\mathcal{B}_{f,m}$

vale $\{\alpha : \langle \alpha, \phi_j \rangle = 0, 1 \leq j \leq m-1\}$, luego, para todo $\alpha \in \mathcal{S} \cap \mathcal{B}_{f,m}$, tenemos que $\langle \alpha, \mathbf{\Gamma}_{i,0}\alpha \rangle \leq \langle \phi_m, \mathbf{\Gamma}_{i,0}\phi_m \rangle = \lambda_{i,m}$ con desigualdad estricta cuando $i = i_0$ y $\alpha \neq \phi_m$. Esto implica que para todo $\alpha \in \mathcal{S} \cap \mathcal{B}_{f,m}$, tenemos $f(\sigma_i^2(\phi_m)) = f(c_i \langle \phi_m, \mathbf{\Gamma}_{i,0}\phi_m \rangle) \geq f(c_i \langle \alpha, \mathbf{\Gamma}_{i,0}\alpha \rangle) = f(\sigma_i^2(\alpha))$ para todo $1 \leq i \leq k$ y $f(\sigma_{i_0}^2(\phi_m)) > f(\sigma_{i_0}^2(\alpha))$ lo que implica que $\zeta(\phi_m) > \zeta(\alpha)$, entonces $\phi_{f,m}(P) = \phi_m$, concluyendo la prueba.

El resultado concerniente a los autovalores seguirá ahora fácilmente ya que $\lambda_{f,i,j} = \sigma_{\mathbb{R}}^2(P_i[\phi_{f,j}]) = \sigma_i^2(\phi_j) = c_i \langle \phi_j, \mathbf{\Gamma}_{i,0}\phi_j \rangle = c_i \lambda_{i,j}$. \square

Si $\mathbf{\Gamma}_{i,0}$ en **A1** es el operador de covarianza de la distribución P_i , los funcionales de las direcciones principales comunes $\phi_{f,m}$ son las componentes principales comunes. Además, tenemos que $\lambda_{f,j} = \sigma_i^2(\phi_{f,j}) = c_i \lambda_{i,j}$ donde $\lambda_{i,j}$ es el autovalor j -ésimo del operador de covarianza de la población i -ésima, eso es, el valor principal tradicional del esquema clásico. Luego, el funcional robusto de autovalores será Fisher-consistente excepto por una constante multiplicativa c_i que puede ser elegida para que valga 1 para todas las poblaciones bajo un modelo central Gaussiano de forma tal de asegurar la consistencia Fisher de estos funcionales también.

5.4 Estimadores robustos bajo el modelo CPCF

Para definir los estimadores de las direcciones principales comunes, definamos primero la versión empírica de $\sigma_i^2, s_{i,n_i}^2 : \mathcal{H} \rightarrow \mathbb{R}$ como $s_{i,n_i}^2(\alpha) = \sigma_{\mathbb{R}}^2(P_{i,n_i}[\alpha])$ y los estimadores de $\zeta(\alpha)$, $\hat{\zeta} : \mathcal{H} \rightarrow \mathbb{R}$, como $\hat{\zeta}(\alpha) = \sum_{i=1}^k \hat{\tau}_i f(s_{i,n_i}^2(\alpha))$ con $\hat{\tau}_i = n_i/N$.

5.4.1 Estimador “crudo” mediante *projection-pursuit*

Análogamente a la propuesta dada por en Boente *et al.* (2006) en el caso finito-dimensional, podemos definir el estimador general “crudo” de direcciones principales comunes se define como

$$\left\{ \begin{array}{l} \hat{\phi}_1 = \operatorname{argmax}_{\|\alpha\|=1} \sum_{i=1}^k \hat{\tau}_i f(s_{i,n_i}^2(\alpha)) = \operatorname{argmax}_{\|\alpha\|=1} \hat{\zeta}(\alpha) \\ \hat{\phi}_m = \operatorname{argmax}_{\|\alpha\|=1, \alpha \in \hat{\mathcal{B}}_{f,m}} \sum_{i=1}^k \hat{\tau}_i f(s_{i,n_i}^2(\alpha)) = \operatorname{argmax}_{\|\alpha\|=1, \alpha \in \hat{\mathcal{B}}_{f,m}} \hat{\zeta}(\alpha) \quad 2 \leq m, \end{array} \right. \quad (5.4)$$

donde $\hat{\mathcal{B}}_m = \{\alpha \in \mathcal{H} : \langle \alpha, \hat{\phi}_j \rangle = 0, 1 \leq j \leq m-1\}$ y el estimador de su tamaño en la población i -ésima como $\hat{\lambda}_{f,i,m} = s_{i,n_i}^2(\hat{\phi}_m)$.

En el caso finito-dimensional, si los operadores de dispersión son proporcionales, es decir, bajo el segundo nivel de la jerarquía definida por Flury (1984), la función de score $f = \log$ minimiza la varianza asintótica sobre una familia de funciones (ver Boente *et al.*, 2006).

La principal desventaja del logaritmo es que $\zeta(\alpha)$ y $\hat{\zeta}(\alpha)$ no están definidas en $\alpha = 0$. Más aún, no serán débilmente continuas en \mathcal{H} debido a la singularidad en $\alpha = 0$. Por esta razón, la mayor parte de las demostraciones requieren un tratamiento especial en dicho caso, distinguiéndolo

del caso de una función continua $f : [0, \infty) \rightarrow \mathbb{R}$ ya que, a diferencia del caso finito-dimensional, la convergencia débil de una sucesión α_n de elementos en \mathcal{S} a α no garantiza que $\alpha \neq 0$.

5.4.2 Estimadores penalizados de las direcciones principales comunes

Referimos a las mismas definiciones de penalización dadas en la sección 3.2.3 cuando definimos los estimadores suavizados para el caso de FPCA.

Los estimadores de las direcciones principales comunes penalizadas pueden definirse

a) por penalización de la escala

$$\left\{ \begin{array}{l} \hat{\phi}_{\text{PS},1} = \operatorname{argmax}_{\|\alpha\|=1} \left\{ \sum_{i=1}^k \hat{\tau}_i f(s_{i,n_i}^2(\alpha)) - \rho[\alpha, \alpha] \right\} = \operatorname{argmax}_{\|\alpha\|=1} \hat{\zeta}(\alpha) - \rho\Psi(\alpha) \\ \hat{\phi}_{\text{PS},m} = \operatorname{argmax}_{\alpha \in \hat{\mathcal{B}}_{\mathcal{S},m}} \left\{ \sum_{i=1}^k \hat{\tau}_i f(s_{i,n_i}^2(\alpha)) - \rho[\alpha, \alpha] \right\} = \operatorname{argmax}_{\alpha \in \hat{\mathcal{B}}_{\mathcal{S},m}} \hat{\zeta}(\alpha) - \rho\Psi(\alpha) \quad 2 \leq m \end{array} \right. \quad (5.5)$$

$$\text{con } \hat{\mathcal{B}}_{\text{PS},m} = \{\alpha \in \mathcal{H} : \|\alpha\| = 1, \langle \alpha, \hat{\phi}_{\text{PS},j} \rangle = 0, \forall 1 \leq j \leq m-1\}.$$

b) o bien penalizando la norma

$$\left\{ \begin{array}{l} \hat{\phi}_{\text{PN},1} = \operatorname{argmax}_{\|\alpha\|_r=1} \sum_{i=1}^k \hat{\tau}_i f(s_{i,n_i}^2(\alpha)) = \operatorname{argmax}_{\|\alpha\|_r=1} \hat{\zeta}(\alpha) \\ \hat{\phi}_{\text{PN},m} = \operatorname{argmax}_{\alpha \in \hat{\mathcal{B}}_{\text{PN},m,r}} \sum_{i=1}^k \hat{\tau}_i f(s_{i,n_i}^2(\alpha)) = \operatorname{argmax}_{\alpha \in \hat{\mathcal{B}}_{\text{PN},m,r}} \hat{\zeta}(\alpha) \quad 2 \leq m \end{array} \right. \quad (5.6)$$

$$\text{con } \hat{\mathcal{B}}_{\text{PN},m,r} = \{\alpha \in \mathcal{H} : \|\alpha\|_r = 1, \langle \alpha, \hat{\phi}_{\text{PN},j} \rangle_r = 0, \forall 1 \leq j \leq m-1\}.$$

Los estimadores de autovalores se definen entonces como

$$\hat{\lambda}_{\text{PS},i,m} = s_{i,n_i}^2(\hat{\phi}_{\text{PS},m}) \quad (5.7)$$

$$\hat{\lambda}_{\text{PN},i,m} = s_{i,n_i}^2(\hat{\phi}_{\text{PN},m}). \quad (5.8)$$

Para tener una formulación unificada para los distintos estimadores considerados antes, permitamos a los productos $\rho\Psi(\alpha)$ o $r\Psi(\alpha)$ valer 0 cuando $\rho = 0$ o $r = 0$ respectivamente, incluso cuando $\alpha \notin \mathcal{H}_{\mathcal{S}}$, en cuyo caso $\Psi(\alpha) = \infty$. Todos los estimadores de *projection-pursuit* considerados antes pueden ser vistos como un caso especial de la siguiente clase general de estimadores:

$$\left\{ \begin{array}{l} \hat{\phi}_{f,1} = \operatorname{argmax}_{\|\alpha\|_r=1} \{\hat{\zeta}(\alpha) - \rho\Psi(\alpha)\} \\ \hat{\phi}_{f,m} = \operatorname{argmax}_{\alpha \in \hat{\mathcal{B}}_{f,m,r}} \{\hat{\zeta}(\alpha) - \rho\Psi(\alpha)\} \quad 2 \leq m, \end{array} \right. \quad (5.9)$$

donde $\widehat{\mathcal{B}}_{f,m,r} = \{\alpha \in \mathcal{H} : \|\alpha\|_r = 1, \langle \alpha, \widehat{\phi}_{f,j} \rangle_r = 0, \forall 1 \leq j \leq m-1\}$.

Con esta definición, el estimador “crudo” se obtiene cuando $\rho = r = 0$, mientras que $\widehat{\phi}_{\text{PN},m}$ y $\widehat{\phi}_{\text{PS},m}$ corresponden a $\rho = 0$ y $r = 0$, respectivamente.

5.4.3 Estimadores mediante *sieves* de las direcciones principales comunes funcionales

Seguimos las mismas definiciones e ideas hechas para el caso PCA.

Sea $\{\delta_i\}_{i \geq 1}$ una base de \mathcal{H} y definamos \mathcal{H}_{p_n} el subespacio lineal generado por $\delta_1, \dots, \delta_{p_n}$ y definamos también $\mathcal{S}_{p_n} = \{\alpha \in \mathcal{H}_{p_n} : \|\alpha\| = 1\}$, es decir,

$$\begin{aligned} \mathcal{H}_{p_n} &= \{\alpha \in \mathcal{H} : \alpha = \sum_{j=1}^{p_n} a_j \delta_j\} \\ \mathcal{S}_{p_n} &= \{\alpha \in \mathcal{H} : \alpha = \sum_{j=1}^{p_n} a_j \delta_j, \mathbf{a} = (a_1, \dots, a_{p_n})^\top \text{ such that } \|\alpha\|^2 = \sum_{j=1}^{p_n} \sum_{s=1}^{p_n} a_j a_s \langle \delta_j, \delta_s \rangle = 1\}. \end{aligned}$$

Como antes, \mathcal{S}_{p_n} aproxima a la esfera unitaria $\mathcal{S} = \{\alpha \in \mathcal{H} : \|\alpha\| = 1\}$. Definamos el estimador robusto de sieves de componentes principales comunes como

$$\begin{cases} \widehat{\phi}_{\text{SI},1} = \operatorname{argmax}_{\alpha \in \mathcal{S}_{p_n}} \sum_{i=1}^k \widehat{\tau}_i f(s_{i,n_i}^2(\alpha)) \\ \widehat{\phi}_{\text{SI},m} = \operatorname{argmax}_{\alpha \in \widehat{\mathcal{B}}_{n,m}} \sum_{i=1}^k \widehat{\tau}_i f(s_{i,n_i}^2(\alpha)) \quad 2 \leq m \end{cases} \quad (5.10)$$

donde $\widehat{\mathcal{B}}_{n,m} = \{\alpha \in \mathcal{S}_{p_n} : \langle \alpha, \widehat{\phi}_{\text{SI},j} \rangle = 0, \forall 1 \leq j \leq m-1\}$ y sus correspondientes estimadores de autofunciones como

$$\widehat{\lambda}_{\text{SI},m} = s_{i,n_i}^2(\widehat{\phi}_{\text{SI},m}). \quad (5.11)$$

5.5 Consistencia de los estimadores de *projection-pursuit* bajo el modelo CPCF

5.5.1 Resultados preliminares

Lema 5.5.1. *Sea $\sigma_{\mathbb{R}}$ un funcional continuo de escala y sea $f : [0, +\infty) \rightarrow \mathbb{R}$ una función estrictamente creciente tal que $f : [0, +\infty) \rightarrow \mathbb{R}$ es una función continua.*

- a) *Sea $\{P_n\}_{n \in \mathbb{N}}$ y P medidas de probabilidad definidas en un espacio de Hilbert separable \mathcal{H} , tal que $d_{\text{PR}}(P_n, P) \rightarrow 0$. Luego, $\sup_{\|\alpha\|=1} |f(\sigma_{\mathbb{R}}^2(P_n[\alpha])) - f(\sigma_{\mathbb{R}}^2(P[\alpha]))| \rightarrow 0$.*

- b) Sea $\{P_{i,n_i}\}_{n_i \in \mathbb{N}}$ y P_i , $1 \leq i \leq k$, medidas de probabilidad definidas en un espacio de Hilbert separable \mathcal{H} , tal que $d_{\text{PR}}(P_{i,n_i}, P_i) \rightarrow 0$ y sea τ_{i,n_i} tal que $0 \leq \tau_{i,n_i}$ y $\tau_{i,n_i} \rightarrow \tau_i$ con $0 \leq \tau_i \leq 1$, $\sum_{i=1}^k \tau_i = 1$. Luego, $\sup_{\|\alpha\|=1} \left| \sum_{i=1}^k \tau_{i,n_i} f(\sigma_{\mathbb{R}}^2(P_{i,n_i}[\alpha])) - \sum_{i=1}^k \tau_i f(\sigma_{\mathbb{R}}^2(P_i[\alpha])) \right| \rightarrow 0$.

DEMOSTRACIÓN. a) Notemos que existe una métrica d que genera la topología débil en \mathcal{H} , la bola cerrada $\mathcal{V}_r = \{\alpha : \|\alpha\| \leq r\}$ es débilmente compacta y entonces, compacta con respecto a d . Por otro lado, $\sigma(\alpha) = \sigma_{\mathbb{R}}(P[\alpha])$ es débilmente continuo en α perteneciente a \mathcal{H} , luego continua con respecto a d . Estos hechos implican que el conjunto $\mathcal{A} = \{\sigma^2(\alpha) : \|\alpha\| \leq 1\}$ es compacto en $[0, +\infty)$, y luego acotado. Supongamos que $\mathcal{A} \subset [0, A] \subset \mathbb{R}$. El hecho que f es continua en $[0, \infty)$ implica que es uniformemente continua en $[0, A + 1]$. Luego, para todo $\epsilon > 0$ existe $\delta > 0$ tal que $u, v \in [0, A + 1]$, $|u - v| \leq \delta$ implica $|f(u) - f(v)| < \epsilon$.

El Lema 3.3.3 implica que $\sup_{\|\alpha\|=1} |\sigma_{\mathbb{R}}(P_n[\alpha]) - \sigma_{\mathbb{R}}(P[\alpha])| \rightarrow 0$, luego existe $n_o \in \mathbb{N}$ tal que, para todo $n \geq n_o$, $\sup_{\|\alpha\|=1} |\sigma_{\mathbb{R}}(P_n[\alpha]) - \sigma_{\mathbb{R}}(P[\alpha])| \leq \min(\delta/(2(A + 1)), \eta)$, donde $\eta \leq \min(1/\sqrt{A}, 1)/4$. Entonces, usando que $\sigma_{\mathbb{R}}(P[\alpha]) \in [0, \sqrt{A}]$, para todo $\alpha \in \mathcal{S}$, obtenemos que, para todo $\alpha \in \mathcal{S}$, $u_n = \sigma_{\mathbb{R}}^2(P_n[\alpha]) \in [0, A + 1]$, $v = \sigma_{\mathbb{R}}^2(P[\alpha]) \in [0, A + 1]$ y $|u_n - v| \leq \delta$, implicando que $|f(\sigma_{\mathbb{R}}^2(P_n[\alpha])) - f(\sigma_{\mathbb{R}}^2(P[\alpha]))| < \epsilon$, para todo $\alpha \in \mathcal{S}$, concluyendo la demostración.

b) Usando a) obtenemos fácilmente que $\sup_{\|\alpha\|=1} \left| \sum_{i=1}^k \tau_{i,n_i} (f(\sigma_{\mathbb{R}}^2(P_{i,n_i}[\alpha])) - f(\sigma_{\mathbb{R}}^2(P_i[\alpha]))) \right| \rightarrow 0$, puesto que $\tau_{i,n_i} \leq 2$, para n_i lo suficientemente grande. Queda ver que $\sup_{\|\alpha\|=1} \left| \sum_{i=1}^k (\tau_{i,n_i} - \tau_i) f(\sigma_{\mathbb{R}}^2(P_i[\alpha])) \right| \rightarrow 0$. Como la bola cerrada $\mathcal{V}_1 = \{\alpha : \|\alpha\| \leq 1\}$ es débilmente compacta y $g_i(\alpha) = f(\sigma_{\mathbb{R}}^2(\alpha)) = f(\sigma_{\mathbb{R}}^2(P_i[\alpha]))$ son funcionales débilmente continuos sobre α en \mathcal{H} , obtenemos que los conjuntos $\mathcal{B}_i = \{f(\sigma_{\mathbb{R}}^2(\alpha)) : \|\alpha\| \leq 1\}$ son compactos, luego acotados, que junto con el hecho que $\tau_{i,n_i} \rightarrow \tau_i$ se concluye la demostración de b). \square

Usando lo anterior, obtenemos

Corolario 5.5.1. Sea $\sigma_{\mathbb{R}}$ un funcional continua de escala y sea $f : [0, +\infty) \rightarrow \mathbb{R}$ una función estrictamente creciente tal que $f : [0, +\infty) \rightarrow \mathbb{R}$ es continua.

- a) Dada P una medida de probabilidad en un espacio de Hilbert separable \mathcal{H} y P_n la medida empírica de una muestra aleatoria X_1, \dots, X_n con $X_i \sim P$, tenemos que $\sup_{\|\alpha\|=1} |f(\sigma_{\mathbb{R}}^2(P_n[\alpha])) - f(\sigma_{\mathbb{R}}^2(P[\alpha]))| \xrightarrow{a.s.} 0$.
- b) Dados P_i , $1 \leq i \leq k$, medidas de probabilidad definidas en un espacio de Hilbert separable \mathcal{H} y $\{P_{i,n_i}\}_{n_i \in \mathbb{N}}$ las medida empíricas de muestras aleatorias $X_{i,1}, \dots, X_{i,n_i}$ con $X_{i,1} \sim P_i$, tenemos que $\sup_{\|\alpha\|=1} \left| \sum_{i=1}^k \tau_{i,n_i} f(\sigma_{\mathbb{R}}^2(P_{i,n_i}[\alpha])) - \sum_{i=1}^k \tau_i f(\sigma_{\mathbb{R}}^2(P_i[\alpha])) \right| \xrightarrow{a.s.} 0$, para toda sucesión τ_{i,n_i} tal que $0 \leq \tau_{i,n_i}$ y $\tau_{i,n_i} \xrightarrow{a.s.} \tau_i$ con $0 \leq \tau_i \leq 1$, $\sum_{i=1}^k \tau_i = 1$.

El siguiente lema será utilizado posteriormente cuando consideremos funciones generales de *score* definidas en 0. La demostración se omite puesto que sigue argumentos análogos a los expuestos en el lema anterior.

Lema 5.5.2. Sea $\sigma_{\mathbb{R}}$ un funcional continua de escala y sea $f : [0, +\infty) \rightarrow \mathbb{R}$ una función estrictamente creciente tal que $f : [0, +\infty) \rightarrow \mathbb{R}$ es continua. Sea $\{P_{i,n_i}\}_{n_i \in \mathbb{N}}$ y P_i , $1 \leq i \leq k$, medidas de probabilidad definidas en un espacio de Hilbert separable \mathcal{H} , tales que

$$\sup_{\alpha \in \mathcal{A}_N} |\sigma_{\mathbb{R}}^2(P_{i,n_i}[\alpha]) - \sigma_{\mathbb{R}}^2(P_i[\alpha])| \longrightarrow 0,$$

donde $\mathcal{A}_N \subset \mathcal{V}_1 = \{\alpha : \|\alpha\| \leq 1\}$, y sea τ_{i,n_i} tal que $0 \leq \tau_{i,n_i}$ y $\tau_{i,n_i} \rightarrow \tau_i$ with $0 \leq \tau_i \leq 1$, $\sum_{i=1}^k \tau_i = 1$. Luego,

$$\sup_{\alpha \in \mathcal{A}_N} \left| \sum_{i=1}^k \tau_{i,n_i} f(\sigma_{\mathbb{R}}^2(P_{i,n_i}[\alpha])) - \sum_{i=1}^k \tau_i f(\sigma_{\mathbb{R}}^2(P_i[\alpha])) \right| \longrightarrow 0.$$

5.5.2 Supuestos

Consideraremos los siguientes supuestos:

- C0.** Para algún $q \geq 2$ y $1 \leq j \leq q$, $\phi_{f,j}(P)$ son únicos salvo un cambio de signo, con $P = P_1 \times \dots \times P_k$.
- C1.** $\sigma_i : \mathcal{H} \rightarrow \mathbb{R}$ es débilmente continuo, es decir, continuo con respecto a la topología débil en \mathcal{H} .
- C2.** $f : [0, +\infty) \rightarrow \mathbb{R}$ es una función estrictamente creciente y continua.
- C3.** $\sup_{\|\alpha\|=1} |s_{i,n_i}^2(\alpha) - \sigma_i^2(\alpha)| \xrightarrow{a.s.} 0$, para todo $1 \leq i \leq k$.
- C4.** $\widehat{\tau}_i \rightarrow \tau_i$.
- C5.** $\sup_{\|\alpha\|=1} |\widehat{\varsigma}(\alpha) - \varsigma(\alpha)| \xrightarrow{a.s.} 0$.

Es claro que **C0** vale si y solamente si, para algún $1 \leq i \leq k$, $\lambda_{f,i,1} > \lambda_{f,i,2} > \dots > \lambda_{f,i,q} > \lambda_{f,i,q+1}$.

Observación 5.1 Vale la pena notar que **C1** y **C2** implican que $\varsigma : \mathcal{H} \rightarrow \mathbb{R}$ es una función débilmente continua. Además **C1** vale si el funcional univariado de escala $\sigma_{\mathbb{R}}$ es cualitativamente robusto, eso es, continuo con respecto de la topología débil en el espacio de medidas de probabilidad inducidas por la distancia de Prohorov, d_{PR} . Sin embargo, esto no es estrictamente necesario. Por ejemplo, si el funcional de escala satisface **A1**, también tendremos la continuidad débil de σ_i . Como se mencionó antes, **A1** se cumple por ejemplo si $\sigma_{\mathbb{R}}$ es el desvío estándar, que no es un funcional continuo de escla. Notemos que **A1** implica también que el funcional σ_i^2 es débilmente uniformemente continuo en la esfera unitaria \mathcal{S} .

Nótese además que si las medidas de probabilidad P_i son elípticas, tenemos que **A1** vale, implicando que **C3** se cumplirá. Por otro lado, **C2** a **C4** implican **C5**.

Por simplicidad notacional, definimos $\widehat{s}_{i,n_i}(\alpha) = f(s_{i,n_i}^2(\alpha))$ y $\widehat{\varsigma}_i(\alpha) = f(\sigma_i^2(\alpha))$.

5.5.3 Consistencia

El siguiente lema será útil para obtener la consistencia de los estimadores generales de las direcciones principales comunes.

Lema 5.5.3. *Sea $P = P_1 \times \dots \times P_k$, $\phi_{f,m} = \phi_{f,m}(P)$ y $\lambda_{f,i,m} = \lambda_{f,i,m}(P)$ definidos como en (5.2) y (5.3) y sea $\widehat{\phi}_m \in \mathcal{V}_1$ tal que $\widehat{\phi}_m \neq 0$, $\|\widehat{\phi}_m\| \xrightarrow{a.s.} 1$ y $\langle \widehat{\phi}_m, \widehat{\phi}_j \rangle \xrightarrow{a.s.} 0$. Supongamos que valen **C0** a **C2**. Luego tenemos que*

- a) *Si $\varsigma(\widehat{\phi}_1) \xrightarrow{a.s.} \varsigma(\phi_{f,1})$, entonces, $\langle \widehat{\phi}_1, \phi_{f,1} \rangle^2 \xrightarrow{a.s.} 1$.*
- b) *Dados $2 \leq m \leq q$, si $\varsigma(\widehat{\phi}_m) \xrightarrow{a.s.} \varsigma(\phi_{f,m})$, $\widehat{\phi}_s \xrightarrow{a.s.} \phi_{f,s}$ para $1 \leq s \leq m-1$, entonces $\langle \widehat{\phi}_m, \phi_{f,m} \rangle^2 \xrightarrow{a.s.} 1$ y por lo tanto, podemos elegir el signo de $\widehat{\phi}_m$ de manera tal que $\widehat{\phi}_m \xrightarrow{a.s.} \phi_{f,m}$.*

DEMOSTRACIÓN. Primero, notemos que **C1** y **C2** implican que $\varsigma : \mathcal{H} \rightarrow \mathbb{R}$ es una función débilmente continua. Luego, es uniformemente débilmente continua en \mathcal{V}_1 que es débilmente compacta.

a) Sea $\mathcal{N} = \{\omega : \varsigma(\widehat{\phi}_1(\omega)) \not\rightarrow \varsigma(\phi_{f,1})\}$ y fijemos $\omega \notin \mathcal{N}$, luego $\varsigma(\widehat{\phi}_1(\omega)) \rightarrow \varsigma(\phi_{f,1})$. Usando que $\{\|\alpha\| \leq 1\}$ es débilmente compacta, tenemos que para cualquier subsucesión γ_ℓ de la sucesión $\widehat{\phi}_1(\omega)$ existe una subsucesión γ_{ℓ_s} tal que $\gamma_{\ell_s} \rightarrow \gamma \in \mathcal{H}$ con $\|\gamma\| \leq 1$.

Por otro lado, usando que $\varsigma(\widehat{\phi}_1(\omega)) \rightarrow \varsigma(\phi_{f,1})$, obtenemos que $\varsigma(\gamma_{\ell_s}) \rightarrow \varsigma(\phi_{f,1})$ mientras que además, la continuidad débil de ς implica que $\varsigma(\gamma_{\ell_s}) \rightarrow \varsigma(\gamma)$, cuando $s \rightarrow \infty$. Luego, $\varsigma(\gamma) = \varsigma(\phi_{f,1})$ y por lo tanto, $\gamma \neq 0$.

Efectivamente, supongamos que $\gamma = 0$. Luego, tenemos que $\sigma_{\mathbb{R}}(P_i[\gamma]) = \sigma_{\mathbb{R}}(P_i[0]) = 0$ lo que nos lleva a que $\varsigma(\gamma) = f(0)$, ya que $\sum_{i=1}^k \tau_i = 1$. Entonces, $\varsigma(\phi_{f,1}) = f(0)$ y $\varsigma(\phi_{f,1}) = \sum_{i=1}^k \tau_i f(\lambda_{f,i,1})$. Usando que f es estrictamente creciente y como **S0** implica que existe i tal que $\lambda_{f,i,1} > 0$, obtenemos que $\sum_{i=1}^k \tau_i f(\lambda_{f,i,1}) > f(0)$ llevándonos a una contradicción. Entonces, $\gamma \neq 0$.

Supongamos que $\|\gamma\| < 1$ y sea $\widetilde{\gamma} = \gamma/\|\gamma\|$, luego $\widetilde{\gamma} \in \mathcal{S}$ lo que implica que $\varsigma(\widetilde{\gamma}) \leq \varsigma(\phi_{f,1})$. Por otro lado, usando que $\sigma_{\mathbb{R}}$ es un funcional de escala, $\|\gamma\| < 1$ y f es estrictamente creciente, obtenemos que

$$\varsigma(\widetilde{\gamma}) = \sum_{i=1}^k \tau_i f(\sigma_{\mathbb{R}}^2(P_i[\widetilde{\gamma}])) = \sum_{i=1}^k \tau_i f\left(\frac{\sigma_{\mathbb{R}}^2(P_i[\gamma])}{\|\gamma\|^2}\right) > \sum_{i=1}^k \tau_i f(\sigma_{\mathbb{R}}^2(P_i[\gamma])) = \varsigma(\gamma) = \varsigma(\phi_{f,1})$$

lo que contradice el hecho que $\varsigma(\phi_{f,1}) = \max_{\|\alpha\|=1} \varsigma(\alpha)$. Por lo tanto, $\|\gamma\| = 1$ y **C0** implica que $\gamma = \phi_{f,1}$ salvo tal vez por un cambio de signo, eso es, $\langle \gamma, \phi_{f,1} \rangle^2 = 1$. Entonces, toda subsucesión de $\widehat{\phi}_1(\omega)$ tendrá un límite convergente o bien a $\phi_{\mathbb{R},1}$ o bien a $-\phi_{\mathbb{R},1}$, concluyendo la demostración de a).

b) Escribamos $\widehat{\phi}_m$ como $\widehat{\phi}_m = \sum_{j=1}^{m-1} \widehat{a}_j \phi_{f,j} + \widehat{\gamma}_m$, con $\langle \widehat{\gamma}_m, \phi_{f,j} \rangle = 0$, $1 \leq j \leq m-1$.

Para obtener b) solamente tenemos que probar que $\langle \widehat{\gamma}_m, \phi_{f,m} \rangle^2 \xrightarrow{a.s.} 1$. Notemos que $\langle \widehat{\phi}_m, \widehat{\phi}_j \rangle \xrightarrow{a.s.} 0$, para $j \neq m$, implica que $\widehat{a}_j = \langle \widehat{\phi}_m, \phi_{f,j} \rangle = \langle \widehat{\phi}_m, \phi_{f,j} - \widehat{\phi}_j \rangle + \langle \widehat{\phi}_m, \widehat{\phi}_j \rangle = \langle \widehat{\phi}_m, \phi_{f,j} - \widehat{\phi}_j \rangle + o_{a.s.}(1)$.

Luego, usando que $\widehat{\phi}_j \xrightarrow{a.s.} \phi_{f,j}$, $1 \leq j \leq m-1$, y $\|\widehat{\phi}_m\| \xrightarrow{a.s.} 1$, obtenemos que $\widehat{a}_j \xrightarrow{a.s.} 0$ para $1 \leq j \leq m-1$. Entonces, $\|\widehat{\phi}_m - \widehat{\gamma}_m\|^2 \xrightarrow{a.s.} 0$. Más aún, usando que $\|\widehat{\phi}_m\|^2 = \sum_{j=1}^{m-1} \widehat{a}_j^2 + \|\widehat{\gamma}_m\|^2$ y $\|\widehat{\phi}_m\|^2 \xrightarrow{a.s.} 1$, obtenemos que $\|\widehat{\gamma}_m\|^2 \leq 1$ y $\|\widehat{\gamma}_m\|^2 \xrightarrow{a.s.} 1$, de donde $\|\widehat{\phi}_m - \widehat{\gamma}_m\| \xrightarrow{a.s.} 0$, con $\widehat{\gamma}_m = \widehat{\gamma}_m / \|\widehat{\gamma}_m\|$.

Usando ahora que $\varsigma(\alpha)$ es una función uniformemente débilmente continua de α en \mathcal{V}_1 , obtenemos que $\varsigma(\widehat{\gamma}_m) - \varsigma(\widehat{\phi}_m) \xrightarrow{a.s.} 0$ que junto con el hecho que $\varsigma(\widehat{\phi}_m) \xrightarrow{a.s.} \varsigma(\phi_{f,m})$ implica que $\varsigma(\widehat{\gamma}_m) \xrightarrow{a.s.} \varsigma(\phi_{f,m})$. La demostración ahora sigue como en a) usando que $\widetilde{\gamma}_m \in \mathcal{C}_m$, con $\mathcal{C}_m = \{\alpha \in \mathcal{S} : \langle \alpha, \phi_{f,j} \rangle = 0, 1 \leq j \leq m-1\}$ y que $\phi_{f,m}$ es el único maximizador de $\varsigma(\alpha)$ sobre \mathcal{C}_m . \square

El siguiente lema será útil para derivar la consistencia de los estimadores de las direcciones principales comunes cuando se usa al logaritmo como función de *score*. Una condición adicional sobre los valores principales $\lambda_{f,i,j}$ es necesario para evitar singularidades.

Lema 5.5.4. *Sea $P = P_1 \times \dots \times P_k$, $\phi_{f,m} = \phi_{f,m}(P)$ y $\lambda_{f,i,m} = \lambda_{f,i,m}(P)$ definidos como en (5.2) y (5.3), respectivamente, con $f = \log$ y $\widehat{\phi}_m \in \mathcal{V}_1$ tal que $\widehat{\phi}_m \neq 0$, $\|\widehat{\phi}_m\| \xrightarrow{a.s.} 1$ y $\langle \widehat{\phi}_m, \widehat{\phi}_j \rangle \xrightarrow{a.s.} 0$. Supongamos que **C0** y **C1** valen y que para todo $1 \leq i \leq k$, $\lambda_{f,i,1} > \lambda_{f,i,2} > \dots > \lambda_{f,i,q} > \lambda_{f,i,q+1}$. Tenemos que*

a) *Si $\varsigma(\widehat{\phi}_1) \xrightarrow{a.s.} \varsigma(\phi_{f,1})$, luego, $\langle \widehat{\phi}_1, \phi_{f,1} \rangle^2 \xrightarrow{a.s.} 1$.*

b) *Dados $2 \leq m \leq q$, si $\varsigma(\widehat{\phi}_m) \xrightarrow{a.s.} \varsigma(\phi_{f,m})$, $\widehat{\phi}_s \xrightarrow{a.s.} \phi_{f,s}$ para $1 \leq s \leq m-1$, luego $\langle \widehat{\phi}_m, \phi_{f,m} \rangle^2 \xrightarrow{a.s.} 1$ y entonces podemos elegir el signo de $\widehat{\phi}_m$ de forma tal que $\widehat{\phi}_m \xrightarrow{a.s.} \phi_{f,m}$.*

DEMOSTRACIÓN. Primero, notemos que **C1** implica que $\sigma_i : \mathcal{H} \rightarrow \mathbb{R}$ es una función débilmente continua sobre \mathcal{V}_1 que es débilmente compacta. La demostración es muy similar a la del Lema 5.5.3.

a) Como en el Lema 5.5.3, sea $\mathcal{N} = \{\omega : \varsigma(\widehat{\phi}_1(\omega)) \not\rightarrow \varsigma(\phi_{f,1})\}$ y fijemos $\omega \notin \mathcal{N}$, luego $\varsigma(\widehat{\phi}_1(\omega)) \rightarrow \varsigma(\phi_{f,1})$.

Usando que $\{\|\alpha\| \leq 1\}$ es débilmente compacto, tenemos que para toda subsucesión γ_ℓ de la sucesión $\widehat{\phi}_1(\omega)$ existe una subsucesión γ_{ℓ_s} tal que $\gamma_{\ell_s} \rightarrow \gamma \in \mathcal{H}$ de forma tal que $\|\gamma\| \leq 1$.

Además, usando que $\varsigma(\widehat{\phi}_1(\omega)) \rightarrow \varsigma(\phi_{f,1})$, tenemos que $\varsigma(\gamma_{\ell_s}) \rightarrow \varsigma(\phi_{f,1})$. Por otro lado, la continuidad débil de σ_i implica que $\sigma_i(\gamma_{\ell_s}) \rightarrow \sigma_i(\gamma)$, con $s \rightarrow \infty$, para $1 \leq i \leq k$.

Si existe $1 \leq i \leq k$ tal que $\sigma_i(\gamma) = 0$, (lo que incluye la situación $\gamma = 0$), el hecho que $\tau_i \geq 0$ implica que $\varsigma(\gamma_{\ell_s}) \rightarrow -\infty$ y entonces, $\varsigma(\phi_{f,1}) = -\infty$ lo que contradice que $\varsigma(\phi_{f,1}) = \sum_{i=1}^k \tau_i f(\lambda_{f,i,1})$ y $\lambda_{f,i,1} > 0$ para todo $1 \leq i \leq k$.

Entonces $\sigma_i(\gamma) \neq 0$, para todo $1 \leq i \leq k$, (implicando que $\gamma \neq 0$), la continuidad de $f = \log$ implica que $\varsigma(\gamma_{\ell_s}) \rightarrow \varsigma(\gamma)$ y entonces, $\varsigma(\gamma) = \varsigma(\phi_{f,1})$. La demostración ahora sigue igual que en el Lema 5.5.3.

b) Escribamos $\widehat{\phi}_m$ como $\widehat{\phi}_m = \sum_{j=1}^{m-1} \widehat{a}_j \phi_{f,j} + \widehat{\gamma}_m$, con $\langle \widehat{\gamma}_m, \phi_{f,j} \rangle = 0$, $1 \leq j \leq m-1$.

Para obtener b) basta con probar que $\langle \widehat{\gamma}_m, \phi_{f,m} \rangle^2 \xrightarrow{a.s.} 1$. Notemos que $\langle \widehat{\phi}_m, \widehat{\phi}_j \rangle \xrightarrow{a.s.} 0$, para $j \neq m$, implica que $\widehat{a}_j = \langle \widehat{\phi}_m, \phi_{f,j} \rangle = \langle \widehat{\phi}_m, \phi_{f,j} - \widehat{\phi}_j \rangle + \langle \widehat{\phi}_m, \widehat{\phi}_j \rangle = \langle \widehat{\phi}_m, \phi_{f,j} - \widehat{\phi}_j \rangle + o_{a.s.}(1)$. Luego,

usando que $\widehat{\phi}_j \xrightarrow{a.s.} \phi_{f,j}$, $1 \leq j \leq m-1$, y $\|\widehat{\phi}_m\| \xrightarrow{a.s.} 1$, obtenemos que $\widehat{a}_j \xrightarrow{a.s.} 0$ para $1 \leq j \leq m-1$. Entonces, $\|\widehat{\phi}_m - \widehat{\gamma}_m\|^2 \xrightarrow{a.s.} 0$. Más aún, usando que $\|\widehat{\phi}_m\|^2 = \sum_{j=1}^{m-1} \widehat{a}_j^2 + \|\widehat{\gamma}_m\|^2$ y $\|\widehat{\phi}_m\|^2 \xrightarrow{a.s.} 1$, obtenemos que $\|\widehat{\gamma}_m\|^2 \leq 1$ y $\|\widehat{\gamma}_m\|^2 \xrightarrow{a.s.} 1$, de donde $\|\widehat{\phi}_m - \widetilde{\gamma}_m\| \xrightarrow{a.s.} 0$, con $\widetilde{\gamma}_m = \widehat{\gamma}_m / \|\widehat{\gamma}_m\|$.

Usando ahora que, para todo $1 \leq i \leq k$, $\sigma_i^2(\alpha)$ es una función uniformemente débilmente continuo, de α en \mathcal{V}_1 , obtenemos que $\sigma_i^2(\widetilde{\gamma}_m) - \sigma_i^2(\widehat{\phi}_m) \xrightarrow{a.s.} 0$ que junto con el hecho que $\sigma_i^2(\widehat{\phi}_m) \xrightarrow{a.s.} \sigma_i^2(\phi_{f,m})$ implica que $\sigma_i^2(\widetilde{\gamma}_m) \xrightarrow{a.s.} \sigma_i^2(\phi_{f,m})$. Entonces, notando que $\lambda_{f,i,m} > 0$ implica que $\sigma_i(\phi_{f,m}) > 0$, para todo $1 \leq i \leq k$, obtenemos de la continuidad de $f = \log$ sobre $(0, \infty)$ que $\varsigma(\widetilde{\gamma}_m) \xrightarrow{a.s.} \varsigma(\phi_{f,m})$.

La demostración ahora sigue como en a) usando que $\widetilde{\gamma}_m \in \mathcal{C}_m$, con $\mathcal{C}_m = \{\alpha \in \mathcal{S} : \langle \alpha, \phi_{f,j} \rangle = 0, 1 \leq j \leq m-1\}$ y que $\phi_{f,m}$ es el único maximizador de $\varsigma(\alpha)$ sobre \mathcal{C}_m . □

Los Teoremas 5.5.1 y 5.5.2 establecen la continuidad de los funcionales definidos en (5.2) y (5.3), para funciones generales de *score* definidas en 0 y para $f = \log$, respectivamente, y, por ende, la robustez asintótica de los estimadores que de ellos se derivan, tal como se define en Hampel (1971). Esto puede verse con apenas reemplazar convergencia casi segura por convergencia en su formulación y tomando P_{i,n_i} , $1 \leq i \leq k$, sucesiones fijas de medidas de probabilidad en vez de medidas aleatorias. Como se ha probado en la Sección 5.5.1, la convergencia uniforme requerida por el supuesto ii) más abajo se satisface, por ejemplo, si σ_R es un funcional continuo de escala cada vez que $P_{i,n_i} \xrightarrow{\omega} P_i$.

Teorema 5.5.1. Sean P_{i,n_i} , $1 \leq i \leq k$ sucesiones aleatorias de medidas de probabilidad, $N = \sum_{i=1}^k n_i$, $\widehat{\tau}_{i,n_i}$ variables aleatorias tales que $\widehat{\tau}_{i,n_i} \xrightarrow{a.s.} \tau_i$ con $0 < \tau_i < 1$, $\sum_{i=1}^k \tau_i = 1$. Sea $r = r_N \geq 0$, $\rho = \rho_N \geq 0$ parámetros aleatorios de suavizado. Sea $\sigma_{i,n_i}^2(\alpha) = \sigma_R^2(P_{i,n_i}[\alpha])$ y $\varsigma_N(\alpha) = \sum_{i=1}^k \widehat{\tau}_i f(\sigma_{i,n_i}^2(\alpha))$ with $f : [0, +\infty) \rightarrow \mathbb{R}$. Definamos $\widehat{\lambda}_{i,m} = \sigma_{i,n_i}^2(\widehat{\phi}_m)$ con

$$\begin{cases} \widehat{\phi}_1 = \operatorname{argmax}_{\|\alpha\|_r=1} \{\varsigma_N(\alpha) - \rho\Psi(\alpha)\} \\ \widehat{\phi}_m = \operatorname{argmax}_{\alpha \in \widehat{\mathcal{B}}_{m,r}} \{\varsigma_N(\alpha) - \rho\Psi(\alpha)\} \quad 2 \leq m, \end{cases}$$

donde $\widehat{\mathcal{B}}_{m,r} = \{\alpha \in \mathcal{H} : \|\alpha\|_r = 1, \langle \alpha, \widehat{\phi}_j \rangle_r = 0, \forall 1 \leq j \leq m-1\}$. Sea $P = P_1 \times \dots \times P_k$ una medida de probabilidad que cumple **C0** y sean $\phi_{f,m} = \phi_{f,m}(P)$ y $\lambda_{f,i,m} = \lambda_{f,i,m}(P)$ definidos como en (5.2) y (5.3), respectivamente. Supongamos que

i) **C1** y **C2** valen,

ii) $\sup_{\|\alpha\|=1} \left| \sigma_{i,n_i}^2(\alpha) - \sigma_i^2(\alpha) \right| \xrightarrow{a.s.} 0$.

iii) $r_N \xrightarrow{a.s.} 0$ y $\rho_N \xrightarrow{a.s.} 0$.

iv) Más aún, si $r_N > 0$ o $\rho_N > 0$, para todo $N \geq N_0$, supongamos que $\phi_{f,j} \in \mathcal{H}_S$, es decir, $\Psi(\phi_{f,j}) < \infty$, para todo $1 \leq j \leq q$.

Luego,

- a) $\varsigma_N(\widehat{\phi}_1) \xrightarrow{a.s.} \varsigma(\phi_{f,1})$ y $\varsigma(\widehat{\phi}_1) \xrightarrow{a.s.} \varsigma(\phi_{f,1})$. Más aún, $\rho\Psi(\widehat{\phi}_1) \xrightarrow{a.s.} 0$ y $r[\widehat{\phi}_1, \widehat{\phi}_1] \xrightarrow{a.s.} 0$, y entonces, $\|\widehat{\phi}_1\| \xrightarrow{a.s.} 1$.
- b) $\langle \widehat{\phi}_1, \phi_{f,1} \rangle^2 \xrightarrow{a.s.} 1$ y $\widehat{\lambda}_{i,1} \xrightarrow{a.s.} \lambda_{f,i,1}$.
- c) Para todo $2 \leq m \leq q$, si $\widehat{\phi}_\ell \xrightarrow{a.s.} \phi_{f,\ell}$, $r\Psi(\widehat{\phi}_\ell) \xrightarrow{a.s.} 0$ y $\rho\Psi(\widehat{\phi}_\ell) \xrightarrow{a.s.} 0$ para $1 \leq \ell \leq m-1$, luego, $\varsigma_N(\widehat{\phi}_m) \xrightarrow{a.s.} \varsigma(\phi_{f,m})$ y $\varsigma(\widehat{\phi}_m) \xrightarrow{a.s.} \varsigma(\phi_{f,m})$. Más aún, $\rho\Psi(\widehat{\phi}_m) \xrightarrow{a.s.} 0$, $r\Psi(\widehat{\phi}_m) \xrightarrow{a.s.} 0$ y entonces, $\|\widehat{\phi}_m\| \xrightarrow{a.s.} 1$.
- d) Para $1 \leq m \leq q$, $\langle \widehat{\phi}_m, \phi_{f,m} \rangle^2 \xrightarrow{a.s.} 1$ y $\widehat{\lambda}_{i,m} \xrightarrow{a.s.} \sigma_i^2(\phi_{f,m})$.

Notemos que el supuesto ii) corresponde a **C3** cuando P_{i,n_i} es la medida empírica de probabilidad de la población i -ésima. Por otro lado, cuando $\sigma_R(\cdot)$ es un funcional continuo de escala, el Lema 5.5.1 implica que ii) valdrá cuando $d_{PR}(P_{i,n_i}, P_i) \xrightarrow{a.s.} 0$. Más aún, si $\sigma_R(\cdot)$ es un funcional continuo de escala y P_i satisface **C0**, el Teorema 5.5.1 implica la continuidad de los funcionales $\phi_{f,j}(\cdot)$ y $\lambda_{f,i,j}(\cdot)$ en P_i , para $1 \leq j \leq q$, y entonces los estimadores propuestos serán consistentes y cualitativamente robustos. En particular, los estimadores son robustos si las poblaciones son independientes y cada una de ellas tiene una distribución elíptica $\mathcal{E}(\mu_1, \mathbf{\Gamma}_1) \times \dots \times \mathcal{E}(\mu_k, \mathbf{\Gamma}_k)$, tal que, para algún $1 \leq i \leq k$, los q mayores autovalores de los operadores $\mathbf{\Gamma}_i$ son todos distintos.

El Teorema 5.5.1 establece en particular la consistencia de los estimadores “crudos” de componentes principales bajo **C0-C5** si se toma $\rho = r = 0$. Además, prueba que las propuestas (3.6) y (3.5) dan estimadores consistentes si $\phi_{f,j} \in \mathcal{H}_S$, $1 \leq j \leq q$.

Precisamos algunas definiciones auxiliares. Indiquemos por \mathcal{L}_k al subespacio generado por $\{\phi_{f,1}, \dots, \phi_{f,k}\}$ y sea $\widehat{\mathcal{L}}_k$ el subespacio generado por los estimadores de las primeras direcciones principales comunes $\widehat{\phi}_1, \dots, \widehat{\phi}_k$. Para cualquier subespacio lineal \mathcal{L} , sea $\pi_{\mathcal{L}} : \mathcal{H} \rightarrow \mathcal{L}$ la proyección ortogonal sobre \mathcal{L} , que estará bien definida puesto que \mathcal{L} será un conjunto cerrado, por ejemplo, si \mathcal{L} es finito-dimensional.

Sea $\mathcal{T}_k = \mathcal{L}_k^\perp$ el subespacio lineal ortogonal a \mathcal{L}_k y sea $\pi_k = \pi_{\mathcal{T}_k}$ la proyección ortogonal con respecto al producto interno definido en \mathcal{H} . Por otro lado, sea $\widehat{\pi}_{r,k}$ la proyección ortogonal en el subespacio ortogonal a $\widehat{\phi}_1, \dots, \widehat{\phi}_k$ en el espacio \mathcal{H}_S con el producto interno $\langle \cdot, \cdot \rangle_r$, o sea, $\widehat{\pi}_{r,k}(\alpha) = \alpha - \sum_{j=1}^k \langle \alpha, \widehat{\phi}_j \rangle_r \widehat{\phi}_j$.

Más aún, $\widehat{\mathcal{T}}_{r,k}$ representará el subespacio ortogonal a $\widehat{\mathcal{L}}_k$ con el producto interno $\langle \cdot, \cdot \rangle_r$. Luego, $\widehat{\pi}_{r,k}$ es la proyección ortogonal sobre $\widehat{\mathcal{T}}_{r,k}$ con respecto a este producto interno.

DEMOSTRACIÓN. Primero notemos que el supuesto ii), **C2** y $\tau_{i,n_i} \xrightarrow{a.s.} \tau_i$ implican que

$$\sup_{\|\alpha\|=1} |\varsigma_N(\alpha) - \varsigma(\alpha)| \xrightarrow{a.s.} 0. \quad (5.12)$$

Además, el hecho que σ_R es un funcional de escala implica que $\sigma_{i,n_i}(\alpha) = \|\alpha\| \sigma_{i,n_i}(\alpha/\|\alpha\|)$. Entonces, del supuesto ii) y de $\|\alpha\| \leq \|\alpha\|_r$, obtenemos que

$$\sup_{\|\alpha\| \leq 1} |\sigma_{i,n_i}^2(\alpha) - \sigma_i^2(\alpha)| \xrightarrow{a.s.} 0 \quad \text{and} \quad \sup_{\|\alpha\|_r \leq 1} |\sigma_{i,n_i}^2(\alpha) - \sigma_i^2(\alpha)| \xrightarrow{a.s.} 0. \quad (5.13)$$

Entonces, usando que los conjuntos $\mathcal{B}_i = \{\sigma_i^2(\alpha), \|\alpha\| \leq 1\}$ son compactos en $[0, \infty)$ y que f es una función continua en $[0, \infty)$ y por lo tanto, uniformemente continua en cualquier vecindad cerrada de \mathcal{B}_i , obtenemos fácilmente del Lema 5.5.2 que

$$\sup_{\|\alpha\| \leq 1} |\varsigma_N(\alpha) - \varsigma(\alpha)| \xrightarrow{a.s.} 0 \quad \text{and} \quad \sup_{\|\alpha\|_r \leq 1} |\varsigma_N(\alpha) - \varsigma(\alpha)| \xrightarrow{a.s.} 0. \quad (5.14)$$

Notemos además de que puesto que f está definida en 0 podemos suponer sin pérdida de generalidad que $f(0) = 0$ lo que implica que $\varsigma(\phi_{f,m}) > 0$ para $1 \leq m \leq q$.

a) Para probar $\varsigma(\widehat{\phi}_1) \xrightarrow{a.s.} \varsigma(\phi_{f,1})$, usando (5.14) y que $\|\widehat{\phi}_1\| \leq \|\widehat{\phi}_1\|_r = 1$, basta ver que $\varsigma_N(\widehat{\phi}_1) \xrightarrow{a.s.} \varsigma(\phi_{f,1})$, para lo cual basta ver que

$$\varsigma(\phi_{f,1}) \geq \varsigma_N(\widehat{\phi}_1) + o_{a.s.}(1), \quad (5.15)$$

$$\varsigma(\phi_{f,1}) \leq \varsigma_N(\widehat{\phi}_1) + o_{a.s.}(1), \quad (5.16)$$

Usando (5.14), obtenemos que $\widehat{a}_{N,1} = \varsigma_N(\widehat{\phi}_1) - \varsigma(\widehat{\phi}_1) \xrightarrow{a.s.} 0$, $\widehat{b}_{N,1} = \varsigma_N(\phi_{f,1}) - \varsigma(\phi_{f,1}) \xrightarrow{a.s.} 0$ y $\widehat{c}_{N,1} = \varsigma_N(\phi_{f,1}/\|\phi_{f,1}\|_r) - \varsigma(\phi_{f,1}/\|\phi_{f,1}\|_r) \xrightarrow{a.s.} 0$.

Usando que σ_R es un funcional de escala, f es estrictamente creciente, $\varsigma(\phi_{f,1}) = \sup_{\alpha \in \mathcal{S}} \varsigma(\alpha)$ y que $\|\widehat{\phi}_1\| \leq \|\widehat{\phi}_1\|_r = 1$, obtenemos fácilmente que

$$\varsigma(\phi_{f,1}) \geq \varsigma\left(\frac{\widehat{\phi}_1}{\|\widehat{\phi}_1\|}\right) \geq \varsigma(\widehat{\phi}_1) = \varsigma_N(\widehat{\phi}_1) - \widehat{a}_{N,1} = \varsigma_N(\widehat{\phi}_1) + o_{a.s.}(1),$$

concluyendo la demostración de (5.15).

Para obtener (5.16), notemos que ya que $\phi_{f,1} \in \mathcal{H}_s$, $\|\phi_{f,1}\|_r < \infty$ y $\|\phi_{f,1}\|_r \geq \|\phi_{f,1}\| = 1$. Luego, usando que $\widehat{\phi}_1 = \operatorname{argmax}_{\|\alpha\|_r=1} \{\varsigma_N(\alpha) - \rho\Psi(\alpha)\}$ y definiendo $\beta_1 = \phi_{f,1}/\|\phi_{f,1}\|_r$, tenemos que $\|\beta_1\|_r = 1$ y

$$\varsigma_N(\widehat{\phi}_1) \geq \varsigma_N(\widehat{\phi}_1) - \rho\Psi(\widehat{\phi}_1) \geq \varsigma_N(\beta_1) - \rho\Psi(\beta_1) = \varsigma_N\left(\frac{\phi_{f,1}}{\|\phi_{f,1}\|_r}\right) - \rho\Psi\left(\frac{\phi_{f,1}}{\|\phi_{f,1}\|_r}\right).$$

Luego, usando que $\Psi(a\alpha) = a^2\Psi(\alpha)$, for any $a \in \mathbb{R}$, tenemos que

$$\begin{aligned} \varsigma_N(\widehat{\phi}_1) &\geq \varsigma_N\left(\frac{\phi_{f,1}}{\|\phi_{f,1}\|_r}\right) - \rho\Psi\left(\frac{\phi_{f,1}}{\|\phi_{f,1}\|_r}\right) = \varsigma_N\left(\frac{\phi_{f,1}}{\|\phi_{f,1}\|_r}\right) - \rho\frac{\Psi(\phi_{f,1})}{\|\phi_{f,1}\|_r^2} \\ &\geq \varsigma\left(\frac{\phi_{f,1}}{\|\phi_{f,1}\|_r}\right) + \widehat{c}_{N,1} - \rho\frac{\Psi(\phi_{f,1})}{\|\phi_{f,1}\|_r^2}. \end{aligned}$$

Cuando $\rho = 0$, habíamos definido $\rho\Psi(\phi_{f,1}) = 0$ y de forma similar cuando $r = 0$ en cuyo caso $\|\phi_{f,1}\|_r = \|\phi_{f,1}\| = 1$. Entonces a partir de ahora suponderemos $r_N > 0$ y $\rho_N > 0$. Como $r \xrightarrow{a.s.} 0$, tenemos que $\|\phi_{f,1}\|_r \xrightarrow{a.s.} \|\phi_{f,1}\| = 1$. Entonces, usando que $\varsigma : \mathcal{H} \rightarrow \mathbb{R}$ es débilmente continua, tenemos que

$$\varsigma\left(\frac{\phi_{f,1}}{\|\phi_{f,1}\|_r}\right) \xrightarrow{a.s.} \varsigma(\phi_{f,1}) ,$$

que junto con $\widehat{c}_{N,1} = o_{a.s.}(1)$, $\|\phi_{f,1}\|_r \xrightarrow{a.s.} 1$, $\rho \xrightarrow{a.s.} 0$ y por lo tanto, $\rho\Psi(\phi_{f,1}) \xrightarrow{a.s.} 0$, implican que

$$\varsigma_N(\widehat{\phi}_1) \geq \varsigma\left(\frac{\phi_{f,1}}{\|\phi_{f,1}\|_r}\right) + \widehat{c}_{N,1} - \rho \frac{\Psi(\phi_{f,1})}{\|\phi_{f,1}\|_r^2} \geq \varsigma(\phi_{f,1}) + o_{a.s.}(1) ,$$

concluyendo la prueba de (5.16). Entonces, $\varsigma_N(\widehat{\phi}_1) \xrightarrow{a.s.} \varsigma(\phi_{f,1})$.

Como se mencionó antes, de (5.14) y $\|\widehat{\phi}_1\| \leq 1$, obtenemos que $\varsigma_N(\widehat{\phi}_1) - \varsigma(\widehat{\phi}_1) \xrightarrow{a.s.} 0$. Luego, usando que $\varsigma_N(\widehat{\phi}_1) \xrightarrow{a.s.} \varsigma(\phi_{f,1})$, tenemos que

$$\varsigma(\widehat{\phi}_1) \xrightarrow{a.s.} \varsigma(\phi_{f,1}) . \quad (5.17)$$

Más aún, las desigualdades

$$\varsigma(\phi_{f,1}) \geq \varsigma\left(\frac{\widehat{\phi}_1}{\|\widehat{\phi}_1\|}\right) \geq \varsigma(\widehat{\phi}_1)$$

obtenidas arriba también implican que

$$\varsigma\left(\frac{\widehat{\phi}_1}{\|\widehat{\phi}_1\|}\right) \xrightarrow{a.s.} \varsigma(\phi_{f,1}) . \quad (5.18)$$

Debemos probar que $\rho\Psi(\widehat{\phi}_1) \xrightarrow{a.s.} 0$, lo que se deduce fácilmente del hecho que $\varsigma_N(\widehat{\phi}_1) \xrightarrow{a.s.} \varsigma(\phi_{f,1})$, $\|\phi_{f,1}\|_r \xrightarrow{a.s.} 1$, $\varsigma(\phi_{f,1}/\|\phi_{f,1}\|_r) \xrightarrow{a.s.} \varsigma(\phi_{f,1})$, $\widehat{c}_{N,1} \xrightarrow{a.s.} 0$ y $\rho \xrightarrow{a.s.} 0$ puesto que

$$\varsigma_N(\widehat{\phi}_1) \geq \varsigma_N(\widehat{\phi}_1) - \rho\Psi(\widehat{\phi}_1) \geq \varsigma\left(\frac{\phi_{f,1}}{\|\phi_{f,1}\|_r}\right) + \widehat{c}_{N,1} - \rho \frac{\Psi(\phi_{f,1})}{\|\phi_{f,1}\|_r^2} .$$

Queda probar que $\rho\Psi(\widehat{\phi}_1) \xrightarrow{a.s.} 0$. Usando que $\|\widehat{\phi}_1\|_r = 1$, obtenemos que $\rho\Psi(\widehat{\phi}_1) = 1 - \|\widehat{\phi}_1\|^2$. Entonces, solamente debemos probar que $\|\widehat{\phi}_1\| \xrightarrow{a.s.} 1$. Notemos que $0 < \|\widehat{\phi}_1\| \leq 1$. Sea $\mathcal{N} = \{\varsigma(\widehat{\phi}_1) \not\rightarrow \varsigma(\phi_{f,1}), \varsigma(\widehat{\phi}_1/\|\widehat{\phi}_1\|) \not\rightarrow \varsigma(\phi_{f,1})\}$, fijemos $\omega \notin \mathcal{N}$ y llamemos $\gamma_N = \widehat{\phi}_1(\omega)$, entonces $0 < y_N = \|\gamma_N\| \leq 1$,

$$\varsigma(\gamma_N) \rightarrow \varsigma(\phi_{f,1}) \quad \text{and} \quad \varsigma(\gamma_N/\|\gamma_N\|) \rightarrow \varsigma(\phi_{f,1}) . \quad (5.19)$$

Queremos probar que $y_N \rightarrow 1$. Dada cualquier subsucesión $\{y_{N'}\}_{N' \in \mathbb{N}}$ de $\{y_N\}_{N \in \mathbb{N}}$, existe una subsucesión $\{y_{N'_\ell}\}_{\ell \in \mathbb{N}}$ tal que $y_{N'_\ell} \rightarrow y$. Claramente, $y \neq 0$. Efectivamente, si $y = 0$, $\gamma_{N'_\ell} \rightarrow 0$ y de la continuidad débil de ς , obtenemos que $\varsigma(\gamma_{N'_\ell}) \rightarrow \varsigma(0)$. Sin embargo, $\varsigma(\gamma_N) \rightarrow \varsigma(\phi_{f,1}) > \varsigma(0)$ puesto que f es estrictamente creciente y $\lambda_{f,i,1} > 0$ para algún i , llevando a una contradicción. Entonces,

$y_{N'_\ell} \rightarrow y \neq 0$. Usando que la bola unitaria \mathcal{V}_1 es débilmente compacta y que $\gamma_{N'_\ell} \in \mathcal{V}_1$, existe una subsucesión tal que $\gamma_{N'_{\ell_s}}$ converge débilmente a $\gamma \in \mathcal{V}_1$. Notemos que $\gamma \neq 0$, ya que la continuidad débil de ς implica que $\varsigma(\gamma_{N'_{\ell_s}}) \rightarrow \varsigma(\gamma)$ mientras que por otro lado, $\varsigma(\gamma_{N'_{\ell_s}}) \rightarrow \varsigma(\phi_{f,1}) > \varsigma(0)$. También, $\|\gamma_{N'_{\ell_s}}\| \rightarrow y$ y entonces, la continuidad débil de ς y el hecho que $\gamma_{N'_{\ell_s}}/\|\gamma_{N'_{\ell_s}}\|$ converge débilmente a γ/y implican que

$$\varsigma\left(\frac{\gamma_{N'_{\ell_s}}}{\|\gamma_{N'_{\ell_s}}\|}\right) \rightarrow \varsigma\left(\frac{\gamma}{y}\right) \quad \text{y} \quad \varsigma(\gamma_{N'_{\ell_s}}) \rightarrow \varsigma(\gamma).$$

Usando (5.19), obtenemos que

$$\varsigma\left(\frac{\gamma}{y}\right) = \varsigma(\gamma) = \varsigma(\phi_{f,1})$$

de donde $y = 1$ puesto que f es estrictamente creciente, concluyendo la demostración de a).

b) La demostración de $\langle \widehat{\phi}_1, \phi_{f,1} \rangle^2 \xrightarrow{a.s.} 1$ sigue inmediatamente del Lema 5.5.3. Como σ_{i,n_i}^2 y σ_i^2 son invariantes por cambio de signo, podemos suponer que $\widehat{\phi}_1 \xrightarrow{a.s.} \phi_{f,1}$.

Usando que $\|\widehat{\phi}_1\| \leq 1$ y (5.13), obtenemos que $\widehat{\lambda}_{i,1} - \sigma_i^2(\widehat{\phi}_1) = \sigma_{i,n_i}^2(\widehat{\phi}_1) - \sigma_i^2(\widehat{\phi}_1) \xrightarrow{a.s.} 0$. Por otro lado, el hecho que $\widehat{\phi}_1 \xrightarrow{a.s.} \phi_{f,1}$ junto con la continuidad débil de σ_i implican que $\sigma_i^2(\widehat{\phi}_1) \xrightarrow{a.s.} \sigma_i^2(\phi_{f,1}) = \lambda_{f,i,1}$, concluyendo la demostración de b).

c) La demostración se llevará a cabo en varios pasos:

PASO 1:

$$\sup_{\|\alpha\|_r \leq 1} |\varsigma(\pi_{m-1}\alpha) - \varsigma_N(\widehat{\pi}_{r,m-1}\alpha)| \xrightarrow{a.s.} 0. \quad (5.20)$$

PASO 2:

$$\varsigma(\phi_{f,m}) \geq \varsigma_N(\widehat{\phi}_m) + o_{a.s.}(1). \quad (5.21)$$

PASO 3:

$$\varsigma(\phi_{f,m}) \leq \varsigma_N(\widehat{\phi}_m) + o_{a.s.}(1). \quad (5.22)$$

Empezamos probando (5.20). Notemos que

$$\sup_{\|\alpha\|_r \leq 1} |\varsigma(\pi_{m-1}\alpha) - \varsigma_N(\widehat{\pi}_{r,m-1}\alpha)| \leq \sup_{\|\alpha\|_r \leq 1} |\varsigma(\pi_{m-1}\alpha) - \varsigma(\widehat{\pi}_{r,m-1}\alpha)| + \sup_{\|\alpha\|_r \leq 1} |\varsigma_N(\widehat{\pi}_{r,m-1}\alpha) - \varsigma(\widehat{\pi}_{r,m-1}\alpha)|.$$

Usando que (5.14) y el hecho que $\|\alpha\|_r \leq 1$ implica que $\|\widehat{\pi}_{r,m-1}\alpha\| \leq 1$, obtenemos que el segundo término en el lado derecho de la desigualdad de arriba converge a cero casi seguramente. Luego, solamente debemos probar que $\sup_{\|\alpha\|_r \leq 1} |\varsigma(\pi_{m-1}\alpha) - \varsigma(\widehat{\pi}_{r,m-1}\alpha)| \xrightarrow{a.s.} 0$.

Para todo $\alpha \in \mathcal{H}_S$, tenemos que $\langle \alpha, \phi_{f,j} \rangle \phi_{f,j} = \langle \alpha, \phi_{f,j} \rangle (\phi_{f,j} - \widehat{\phi}_j) + \langle \alpha, \phi_{f,j} \rangle \widehat{\phi}_j$. Entonces, si $\|\alpha\|_r^2 = \|\alpha\|^2 + r\Psi(\alpha) \leq 1$, obtenemos que

$$\begin{aligned} \|\langle \alpha, \phi_{f,j} \rangle \phi_{f,j} - \langle \alpha, \widehat{\phi}_j \rangle_r \widehat{\phi}_j\| &\leq \|\alpha\| \|\phi_{f,j} - \widehat{\phi}_j\| + \|\widehat{\phi}_j\| \|\langle \alpha, \phi_{f,j} \rangle - \langle \alpha, \widehat{\phi}_j \rangle_r\| \\ &\leq \|\phi_{f,j} - \widehat{\phi}_j\| + |\langle \alpha, \phi_{f,j} - \widehat{\phi}_j \rangle - r[\alpha, \widehat{\phi}_j]| \end{aligned}$$

$$\begin{aligned}
&\leq \|\phi_{f,j} - \widehat{\phi}_j\| + \left\{ \|\phi_{f,j} - \widehat{\phi}_j\| + (r\Psi(\alpha))^{\frac{1}{2}}(r\Psi(\widehat{\phi}_j))^{\frac{1}{2}} \right\} \\
&\leq \|\phi_{f,j} - \widehat{\phi}_j\| + \left\{ \|\phi_{f,j} - \widehat{\phi}_j\| + (r\Psi(\widehat{\phi}_j))^{\frac{1}{2}} \right\}.
\end{aligned}$$

Usando que para $1 \leq j \leq m-1$, $\widehat{\phi}_j \xrightarrow{a.s.} \phi_{f,j}$ y $r\Psi(\widehat{\phi}_j) \xrightarrow{a.s.} 0$, obtenemos que

$$\sup_{\|\alpha\|_r \leq 1} \|\langle \alpha, \phi_{f,j} \rangle \phi_{f,j} - \langle \alpha, \widehat{\phi}_j \rangle_r \widehat{\phi}_j\| \xrightarrow{a.s.} 0.$$

lo que implica que

$$\sup_{\|\alpha\|_r \leq 1} \|\widehat{\pi}_{r,m-1}\alpha - \pi_{m-1}\alpha\| \xrightarrow{a.s.} 0. \quad (5.23)$$

Luego, usando que **C1** y **C2** implican que ς es débilmente uniformemente continuo sobre \mathcal{V}_1 , obtenemos que $\sup_{\|\alpha\|_r \leq 1} |\varsigma(\pi_{m-1}\alpha) - \varsigma(\widehat{\pi}_{r,m-1}\alpha)| \xrightarrow{a.s.} 0$, concluyendo la demostración de (5.20).

PASO 2: Como en a), probaremos ahora que (5.21) vale.

Recordemos que $\varsigma(\phi_{f,m}) = \sup_{\alpha \in \mathcal{S} \cap \mathcal{T}_{m-1}} \varsigma(\alpha)$. Como en a), usando que para cualquier $c \geq 1$ tenemos que $\varsigma(c\alpha) \geq \varsigma(\alpha)$, tenemos que para todo $\alpha \in \mathcal{H}$ tal que $\|\alpha\| \leq 1$, tenemos que $\varsigma(\alpha/\|\alpha\|) \geq \varsigma(\alpha)$. Entonces, $\sup_{\alpha \in \mathcal{S} \cap \mathcal{T}_{m-1}} \varsigma(\alpha) = \sup_{\alpha \in \{\|\alpha\| \leq 1\} \cap \mathcal{T}_{m-1}} \varsigma(\alpha)$. Más aún, para todo $\alpha \in \mathcal{V}_1$, tenemos que $\|\pi_{m-1}\alpha\| \leq 1$ y $\pi_{m-1}\alpha \in \mathcal{T}_{m-1}$, entonces

$$\sup_{\alpha \in \{\|\alpha\| \leq 1\} \cap \mathcal{T}_{m-1}} \varsigma(\alpha) \geq \sup_{\|\alpha\| \leq 1} \varsigma(\pi_{m-1}\alpha) \geq \sup_{\alpha \in \mathcal{S}} \varsigma(\pi_{m-1}\alpha). \quad (5.24)$$

Ahora, usando que $\widehat{\phi}_m/\|\widehat{\phi}_m\| \in \mathcal{S}$, obtenemos que

$$\sup_{\alpha \in \{\|\alpha\| \leq 1\} \cap \mathcal{T}_{m-1}} \varsigma(\alpha) \geq \varsigma\left(\pi_{m-1}\left(\frac{\widehat{\phi}_m}{\|\widehat{\phi}_m\|}\right)\right) \geq \varsigma\left(\frac{1}{\|\widehat{\phi}_m\|}\pi_{m-1}\widehat{\phi}_m\right) \geq \varsigma(\pi_{m-1}\widehat{\phi}_m),$$

ya que $\|\widehat{\phi}_m\| \leq 1$. Sumando todo esto, tenemos que

$$\varsigma(\phi_{f,m}) \geq \varsigma(\pi_{m-1}\widehat{\phi}_m).$$

Usando que (5.20) y el hecho que $\|\widehat{\phi}_m\|_r = 1$, obtenemos que $b_m = \varsigma(\pi_{m-1}\widehat{\phi}_m) - \varsigma_N(\widehat{\pi}_{r,m-1}\widehat{\phi}_m) \xrightarrow{a.s.} 0$. Notando que $\widehat{\pi}_{r,m-1}\widehat{\phi}_m = \widehat{\phi}_m$, obtenemos que

$$\varsigma(\phi_{f,m}) \geq \varsigma(\pi_{m-1}\widehat{\phi}_m) = \varsigma_N(\widehat{\pi}_{r,m-1}\widehat{\phi}_m) + o_{a.s.}(1) = \varsigma_N(\widehat{\phi}_m) + o_{a.s.}(1),$$

concluyendo la demostración de (5.21).

PASO 3: Vamos a probar (5.22). Como $\phi_{f,m} \in \mathcal{H}_s$, tenemos que $1 \leq \|\phi_{f,m}\|_r < \infty$ y usando que $r \xrightarrow{a.s.} 0$, también tenemos que $\|\phi_{f,m}\|_r \xrightarrow{a.s.} \|\phi_{f,m}\| = 1$. Luego,

$$\begin{aligned}
\varsigma_N(\widehat{\phi}_m) &\geq \varsigma_N(\widehat{\phi}_m) - \rho\Psi(\widehat{\phi}_m) = \sup_{\|\alpha\|_r=1, \alpha \in \widehat{\mathcal{T}}_{r,m-1}} \{\varsigma_N(\alpha) - \rho\Psi(\alpha)\} \\
&\geq \varsigma_N\left(\frac{\widehat{\pi}_{r,m-1}\phi_{f,m}}{\|\widehat{\pi}_{r,m-1}\phi_{f,m}\|_r}\right) - \rho\Psi\left(\frac{\widehat{\pi}_{r,m-1}\phi_{f,m}}{\|\widehat{\pi}_{r,m-1}\phi_{f,m}\|_r}\right).
\end{aligned}$$

Usando que $\varsigma_N(c\alpha) \geq \varsigma_N(\alpha)$ si $c \geq 1$ y el hecho que $\|\widehat{\pi}_{r,m-1}\phi_{f,m}\|_r \leq \|\phi_{f,m}\|_r$, obtenemos fácilmente que

$$\begin{aligned} \varsigma_N(\widehat{\phi}_m) &\geq \varsigma_N\left(\frac{\widehat{\pi}_{r,m-1}\phi_{f,m}}{\|\phi_{f,m}\|_r}\right) - \rho\Psi\left(\frac{\widehat{\pi}_{r,m-1}\phi_{f,m}}{\|\widehat{\pi}_{r,m-1}\phi_{f,m}\|_r}\right) \\ &\geq \varsigma_N\left(\frac{\widehat{\pi}_{r,m-1}\phi_{f,m}}{\|\phi_{f,m}\|_r}\right) - \rho\frac{\Psi(\widehat{\pi}_{r,m-1}\phi_{f,m})}{\|\widehat{\pi}_{r,m-1}\phi_{f,m}\|_r^2}. \end{aligned}$$

Notemos que (5.20) implica que $c_m = \varsigma(\pi_{m-1}\phi_{f,m}/\|\phi_{f,m}\|_r) - \varsigma_N(\widehat{\pi}_{r,m-1}\phi_{f,m}/\|\phi_{f,m}\|_r) \xrightarrow{a.s.} 0$, luego,

$$\begin{aligned} \varsigma_N(\widehat{\phi}_m) &\geq \varsigma_N\left(\frac{\widehat{\pi}_{r,m-1}\phi_{f,m}}{\|\phi_{f,m}\|_r}\right) - \rho\Psi\left(\frac{\widehat{\pi}_{r,m-1}\phi_{f,m}}{\|\widehat{\pi}_{r,m-1}\phi_{f,m}\|_r}\right) \\ &\geq \varsigma\left(\frac{\pi_{m-1}\phi_{f,m}}{\|\phi_{f,m}\|_r}\right) + o_{a.s.}(1) - \rho\frac{\Psi(\widehat{\pi}_{r,m-1}\phi_{f,m})}{\|\widehat{\pi}_{r,m-1}\phi_{f,m}\|_r^2}. \end{aligned}$$

Por otro lado, la continuidad débil de ς , el hecho que $\|\phi_{f,m}\|_r \xrightarrow{a.s.} \|\phi_{f,m}\| = 1$ y $\pi_{m-1}\phi_{f,m} = \phi_{f,m}$, implican que

$$\varsigma_N(\widehat{\phi}_m) \geq \varsigma(\phi_{f,m}) + o_{a.s.}(1) - \frac{\rho\Psi(\widehat{\pi}_{r,m-1}\phi_{f,m})}{\|\widehat{\pi}_{r,m-1}\phi_{f,m}\|_r^2}. \quad (5.25)$$

Argumentos análogos a los considerados en Pezzulli y Silverman (1993), permiten probar que ya que $r\Psi(\widehat{\phi}_j) \xrightarrow{a.s.} 0$ y $\rho\Psi(\widehat{\phi}_j) \xrightarrow{a.s.} 0$ luego

$$\rho\Psi(\widehat{\pi}_{r,m-1}\phi_{f,m}) = \rho[\widehat{\pi}_{r,m-1}\phi_{f,m}, \widehat{\pi}_{r,m-1}\phi_{f,m}] \xrightarrow{a.s.} 0, \quad (5.26)$$

$$r\Psi(\widehat{\pi}_{r,m-1}\phi_{f,m}) = r[\widehat{\pi}_{r,m-1}\phi_{f,m}, \widehat{\pi}_{r,m-1}\phi_{f,m}] \xrightarrow{a.s.} 0. \quad (5.27)$$

Además, como en la demostración de (5.23), tenemos que

$$\begin{aligned} \|\langle\phi_{f,m}, \phi_{f,j}\rangle\phi_{f,j} - \langle\phi_{f,m}, \widehat{\phi}_j\rangle_r\widehat{\phi}_j\| &\leq \|\phi_{f,j} - \widehat{\phi}_j\| + \|\widehat{\phi}_j\| |\langle\phi_{f,m}, \phi_{f,j}\rangle - \langle\phi_{f,m}, \widehat{\phi}_j\rangle_r| \\ &\leq \|\phi_{f,j} - \widehat{\phi}_j\| + |\langle\phi_{f,m}, \phi_{f,j} - \widehat{\phi}_j\rangle - r[\phi_{f,m}, \widehat{\phi}_j]| \\ &\leq \|\phi_{f,j} - \widehat{\phi}_j\| + \left\{ \|\phi_{f,j} - \widehat{\phi}_j\| + (r\Psi(\phi_{f,m}))^{\frac{1}{2}} (r\Psi(\widehat{\phi}_j))^{\frac{1}{2}} \right\}. \end{aligned}$$

Usando que para $1 \leq j \leq m-1$, $\widehat{\phi}_j \xrightarrow{a.s.} \phi_{f,j}$, $r\Psi(\widehat{\phi}_j) \xrightarrow{a.s.} 0$ y $r\Psi(\phi_{f,m}) \xrightarrow{a.s.} 0$, obtenemos que $\|\langle\phi_{f,m}, \phi_{f,j}\rangle\phi_{f,j} - \langle\phi_{f,m}, \widehat{\phi}_j\rangle_r\widehat{\phi}_j\| \xrightarrow{a.s.} 0$. Luego, usando que $\langle\phi_{f,m}, \phi_{f,j}\rangle = 0$ para $j < m$ y que $\|\phi_{f,m}\| = 1$ obtenemos que

$$\|\widehat{\pi}_{r,m-1}\phi_{f,m}\| = \|\phi_{f,m} - \sum_{j=1}^{m-1} \langle\phi_{f,m}, \widehat{\phi}_j\rangle_r\widehat{\phi}_j\| \xrightarrow{a.s.} 1.$$

Luego, usando (5.27), obtenemos que

$$\|\widehat{\pi}_{r,m-1}\phi_{f,m}\|_r^2 = \|\widehat{\pi}_{r,m-1}\phi_{f,m}\|^2 + r\Psi(\widehat{\pi}_{r,m-1}\phi_{f,m}) \xrightarrow{a.s.} 1.$$

Entonces, (5.25) junto con (5.26) implican que $\varsigma_N(\widehat{\phi}_m) \geq \varsigma(\phi_{f,m}) + o_{\text{a.s.}}(1)$, concluyendo la demostración de (5.22).

Notemos que (5.21) y (5.22) implican que

$$\varsigma_N(\widehat{\phi}_m) \xrightarrow{\text{a.s.}} \varsigma(\phi_{f,m}), \quad (5.28)$$

como se deseaba. Por otro lado, el hecho que $\|\widehat{\phi}_m\| \leq 1$ y (5.14) implican que $\varsigma_N(\widehat{\phi}_m) - \varsigma(\widehat{\phi}_m) \xrightarrow{\text{a.s.}} 0$, que junto con (5.28) nos llevan a $\varsigma(\widehat{\phi}_m) \xrightarrow{\text{a.s.}} \varsigma(\phi_{f,m})$. Luego, el Lema 5.5.3 nos permite concluir que $\langle \widehat{\phi}_m, \phi_{f,m} \rangle^2 \xrightarrow{\text{a.s.}} 1$.

PASO 4: Probaremos que $\rho\Psi(\widehat{\phi}_m) \xrightarrow{\text{a.s.}} 0$. Notemos que las desigualdades de arriba implican que

$$\varsigma_N(\widehat{\phi}_m) \geq \varsigma_N(\widehat{\phi}_m) - \rho\Psi(\widehat{\phi}_m) \geq \varsigma(\phi_{f,m}) + o_{\text{a.s.}}(1).$$

Luego, (5.28) implican que $0 \leq \rho\Psi(\widehat{\phi}_m) \leq \varsigma_N(\widehat{\phi}_m) - \varsigma(\phi_{f,m}) + o_{\text{a.s.}}(1) \xrightarrow{\text{a.s.}} 0$, como se deseaba.

PASO 5: Para concluir la demostración de c) queda demostrar que

$$r\Psi(\widehat{\phi}_m) \xrightarrow{\text{a.s.}} 0. \quad (5.29)$$

Recordemos que como se probó antes,

$$\varsigma(\phi_{f,m}) \geq \varsigma\left(\frac{\pi_{m-1}\widehat{\phi}_m}{\|\widehat{\phi}_m\|}\right) \geq \varsigma(\pi_{m-1}\widehat{\phi}_m) = \varsigma_N(\widehat{\phi}_m) + o_{\text{a.s.}}(1).$$

Luego,

$$\varsigma\left(\frac{\pi_{m-1}\widehat{\phi}_m}{\|\widehat{\phi}_m\|}\right) \xrightarrow{\text{a.s.}} \varsigma(\phi_{f,m}) \quad \text{y} \quad \varsigma(\pi_{m-1}\widehat{\phi}_m) \xrightarrow{\text{a.s.}} \varsigma(\phi_{f,m})$$

Ahora la demostración se sigue como en a) tomando $\gamma_N = \pi_{m-1}\widehat{\phi}_m(\omega)$, $y_N = \|\widehat{\phi}_m(\omega)\| \leq 1$ para $\omega \notin \mathcal{N}$ con

$$\mathcal{N} = \{\omega \in \Omega : \varsigma(\widehat{\phi}_m(\omega)) \not\rightarrow \varsigma(\phi_{f,m}), \varsigma\left(\frac{\pi_{m-1}\widehat{\phi}_m(\omega)}{\|\widehat{\phi}_m(\omega)\|}\right) \not\rightarrow \varsigma(\phi_{f,m}) \text{ y } \varsigma(\pi_{m-1}\widehat{\phi}_m(\omega)) \not\rightarrow \varsigma(\phi_{f,m})\}$$

y usando que los autovalores están ordenados de forma estrictamente decreciente y $\lambda_{f,i,m} > 0$ para algún i , ya que $m < q$ y **C0** vale.

d) Ya probamos que cuando $m = 1$ el resultado vale. Procedemos por inducción, supongamos que $\langle \widehat{\phi}_\ell, \phi_{f,\ell} \rangle^2 \rightarrow 1$, $r\Psi(\widehat{\phi}_\ell) \xrightarrow{\text{a.s.}} 0$ y que $\rho\Psi(\widehat{\phi}_\ell) \xrightarrow{\text{a.s.}} 0$ para $1 \leq \ell \leq m-1$, queremos ver que $\langle \widehat{\phi}_m, \phi_{f,m} \rangle^2 \rightarrow 1$. Sin pérdida de generalidad, podemos suponer $\widehat{\phi}_\ell \xrightarrow{\text{a.s.}} \phi_{f,\ell}$, para $1 \leq \ell \leq m-1$. Usando c) tenemos que $\varsigma(\widehat{\phi}_m) \xrightarrow{\text{a.s.}} \varsigma(\phi_{f,m})$ y que $\|\widehat{\phi}_m\| \xrightarrow{\text{a.s.}} 1$ y entonces, usando el el Lema 5.5.3 tenemos que $\langle \widehat{\phi}_m, \phi_{f,m} \rangle^2 \xrightarrow{\text{a.s.}} 1$. Sin pérdida de generalidad podemos suponer que $\widehat{\phi}_m \xrightarrow{\text{a.s.}} \phi_{f,m}$, ya que σ_{i,n_i}^2 y σ_i^2 son invariantes por cambios de signo. Luego, como en b) usando que $\|\widehat{\phi}_m\| \leq 1$ y (5.13), obtenemos que $\widehat{\lambda}_{i,m} - \sigma_i^2(\widehat{\phi}_m) = \sigma_{i,n_i}^2(\widehat{\phi}_m) - \sigma_i^2(\widehat{\phi}_m) \xrightarrow{\text{a.s.}} 0$. Por otro lado, el hecho que $\widehat{\phi}_m \xrightarrow{\text{a.s.}} \phi_{f,m}$ junto con la continuidad débil de σ_i implican que $\sigma_i^2(\widehat{\phi}_m) \xrightarrow{\text{a.s.}} \sigma_i^2(\phi_{f,m}) = \lambda_{f,i,m}$, concluyendo la prueba de d). \square

El siguiente Teorema establece un resultado análogo para el caso en que se use al logaritmo como función de *score*.

Teorema 5.5.2. Sean P_{i,n_i} , $1 \leq i \leq k$, sucesiones de medidas de probabilidad aleatorias, $N = \sum_{i=1}^k n_i$, $\widehat{\tau}_{i,n_i}$ variables aleatorias tales que $\widehat{\tau}_{i,n_i} \xrightarrow{a.s.} \tau_i$ con $0 < \tau_i < 1$, $\sum_{i=1}^k \tau_i = 1$. Sea $r = r_N \geq 0$, $\rho = \rho_N \geq 0$ parámetros aleatorios de suavizado. Sea $\sigma_{i,n_i}^2(\alpha) = \sigma_R^2(P_{i,n_i}[\alpha])$ y $\varsigma_N(\alpha) = \sum_{i=1}^k \widehat{\tau}_i f(\sigma_{i,n_i}^2(\alpha))$ con $f = \log$ y definamos $\widehat{\lambda}_{i,m} = \sigma_{i,n_i}^2(\widehat{\phi}_m)$ con

$$\begin{cases} \widehat{\phi}_1 = \operatorname{argmax}_{\|\alpha\|_r=1} \{\varsigma_N(\alpha) - \rho\Psi(\alpha)\} \\ \widehat{\phi}_m = \operatorname{argmax}_{\alpha \in \widehat{\mathcal{B}}_{m,r}} \{\varsigma_N(\alpha) - \rho\Psi(\alpha)\} \quad 2 \leq m, \end{cases}$$

donde $\widehat{\mathcal{B}}_{m,r} = \{\alpha \in \mathcal{H} : \|\alpha\|_r = 1, \langle \alpha, \widehat{\phi}_j \rangle_r = 0, \forall 1 \leq j \leq m-1\}$. Sea $P = P_1 \times \dots \times P_k$ una medida de probabilidad que satisface **C0**, $\phi_{f,m} = \phi_{f,m}(P)$ y $\lambda_{f,i,m} = \lambda_{f,i,m}(P)$ definidos como en (5.2) y (5.3), respectivamente, con $f = \log$ y supongamos que para todo $1 \leq i \leq k$, $\lambda_{f,i,1} > \lambda_{f,i,2} > \dots > \lambda_{f,i,q} > \lambda_{f,i,q+1}$. Más aún, supongamos que

i) **C1** vale,

ii) $\sup_{\|\alpha\|=1} \left| \sigma_{i,n_i}^2(\alpha) - \sigma_i^2(\alpha) \right| \xrightarrow{a.s.} 0$.

iii) $r_N \xrightarrow{a.s.} 0$ y $\rho_N \xrightarrow{a.s.} 0$.

iv) Más aún, si $r_N > 0$ o $\rho_N > 0$, para todo $N \geq N_0$, supongamos que $\phi_{f,j} \in \mathcal{H}_s$, es decir, $\Psi(\phi_{f,j}) < \infty$, para todo $1 \leq j \leq q$.

Luego,

a) $\varsigma_N(\widehat{\phi}_1) \xrightarrow{a.s.} \varsigma(\phi_{f,1})$ y $\varsigma(\widehat{\phi}_1) \xrightarrow{a.s.} \varsigma(\phi_{f,1})$. Más aún, $\rho\Psi(\widehat{\phi}_1) \xrightarrow{a.s.} 0$ y $r[\widehat{\phi}_1, \widehat{\phi}_1] \xrightarrow{a.s.} 0$, y entonces, $\|\widehat{\phi}_1\| \xrightarrow{a.s.} 1$.

b) $\langle \widehat{\phi}_1, \phi_{f,1} \rangle^2 \xrightarrow{a.s.} 1$ y $\widehat{\lambda}_{i,1} \xrightarrow{a.s.} \lambda_{f,i,1}$.

c) Para todo $2 \leq m \leq q$, si $\widehat{\phi}_\ell \xrightarrow{a.s.} \phi_{f,\ell}$, $r\Psi(\widehat{\phi}_\ell) \xrightarrow{a.s.} 0$ y $\rho\Psi(\widehat{\phi}_\ell) \xrightarrow{a.s.} 0$ para $1 \leq \ell \leq m-1$, luego, $\varsigma_N(\widehat{\phi}_m) \xrightarrow{a.s.} \varsigma(\phi_{f,m})$ y $\varsigma(\widehat{\phi}_m) \xrightarrow{a.s.} \varsigma(\phi_{f,m})$. Más aún, $\rho\Psi(\widehat{\phi}_m) \xrightarrow{a.s.} 0$, $r\Psi(\widehat{\phi}_m) \xrightarrow{a.s.} 0$ y entonces, $\|\widehat{\phi}_m\| \xrightarrow{a.s.} 1$.

d) Para $1 \leq m \leq q$, $\langle \widehat{\phi}_m, \phi_{f,m} \rangle^2 \xrightarrow{a.s.} 1$ y $\widehat{\lambda}_{i,m} \xrightarrow{a.s.} \sigma_i^2(\phi_{f,m})$.

DEMOSTRACIÓN. Primero notemos que como en el Teorema 5.5.1, del supuesto ii) y el hecho que $\|\alpha\| \leq \|\alpha\|_r$ obtenemos que (5.13) vale.

a) Como en el Teorema 5.5.1, para probar que $\varsigma_N(\widehat{\phi}_1) \xrightarrow{a.s.} \varsigma(\phi_{f,1})$, basta ver que

$$\varsigma(\phi_{f,1}) \geq \varsigma_N(\widehat{\phi}_1) + o_{a.s.}(1), \quad (5.30)$$

$$\varsigma(\phi_{f,1}) \leq \varsigma_N(\widehat{\phi}_1) + o_{a.s.}(1), \quad (5.31)$$

Usando que (5.13), obtenemos que $\widehat{a}_{N,i} = \sigma_{i,n_i}^2(\widehat{\phi}_1) - \sigma_i^2(\widehat{\phi}_1) \xrightarrow{a.s.} 0$, $\widehat{b}_{N,i} = \sigma_{i,n_i}^2(\phi_{f,1}) - \sigma_i^2(\phi_{f,1}) \xrightarrow{a.s.} 0$. Por otro lado, usando que $\sigma_i^2(\phi_{f,1}) > 0$ tenemos que $\log(\sigma_{i,n_i}^2(\phi_{f,1})) \xrightarrow{a.s.} \log(\sigma_i^2(\phi_{f,1}))$ lo que implica que $\varsigma_N(\phi_{f,1}) \xrightarrow{a.s.} \varsigma(\phi_{f,1})$. Además, usando que $r \xrightarrow{a.s.} 0$ y $\|\phi_{f,1}\|_r \xrightarrow{a.s.} 1$ tenemos que

$$\begin{aligned} \log\left(\sigma_{i,n_i}^2\left(\frac{\phi_{f,1}}{\|\phi_{f,1}\|_r}\right)\right) &= \log\left(\frac{\sigma_{i,n_i}^2(\phi_{f,1})}{\|\phi_{f,1}\|_r^2}\right) \\ &= \log(\sigma_{i,n_i}^2(\phi_{f,1})) - 2\log(\|\phi_{f,1}\|_r) \xrightarrow{a.s.} \log(\sigma_i^2(\phi_{f,1})), \end{aligned}$$

entonces,

$$\varsigma_N\left(\frac{\phi_{f,1}}{\|\phi_{f,1}\|_r}\right) \xrightarrow{a.s.} \varsigma(\phi_{f,1}). \quad (5.32)$$

Para obtener (5.31), notemos que, como $\phi_{f,1} \in \mathcal{H}_S$, $\|\phi_{f,1}\|_r < \infty$ y $\|\phi_{f,1}\|_r \geq \|\phi_{f,1}\| = 1$. Luego, usando que $\widehat{\phi}_1 = \operatorname{argmax}_{\|\alpha\|_r=1} \{\varsigma_N(\alpha) - \rho\Psi(\alpha)\}$ y definiendo $\beta_1 = \phi_{f,1}/\|\phi_{f,1}\|_r$, tenemos que $\|\beta_1\|_r = 1$ y

$$\varsigma_N(\widehat{\phi}_1) \geq \varsigma_N(\widehat{\phi}_1) - \rho\Psi(\widehat{\phi}_1) \geq \varsigma_N(\beta_1) - \rho\Psi(\beta_1) = \varsigma_N\left(\frac{\phi_{f,1}}{\|\phi_{f,1}\|_r}\right) - \rho\Psi\left(\frac{\phi_{f,1}}{\|\phi_{f,1}\|_r}\right).$$

Entonces, usando que $\Psi(a\alpha) = a^2\Psi(\alpha)$, para cualquier $a \in \mathbb{R}$, y (5.32), tenemos que

$$\begin{aligned} \varsigma_N(\widehat{\phi}_1) &\geq \varsigma_N\left(\frac{\phi_{f,1}}{\|\phi_{f,1}\|_r}\right) - \rho\Psi\left(\frac{\phi_{f,1}}{\|\phi_{f,1}\|_r}\right) = \varsigma_N\left(\frac{\phi_{f,1}}{\|\phi_{f,1}\|_r}\right) - \rho\frac{\Psi(\phi_{f,1})}{\|\phi_{f,1}\|_r^2} \\ &\geq \varsigma(\phi_{f,1}) + o_{a.s.}(1) - \rho\frac{\Psi(\phi_{f,1})}{\|\phi_{f,1}\|_r^2}. \end{aligned} \quad (5.33)$$

Cuando $\rho = 0$, habíamos definido $\rho\Psi(\phi_{f,1}) = 0$ y similarmente cuando $r = 0$ en cuyo caso $\|\phi_{f,1}\|_r = \|\phi_{f,1}\| = 1$. Entonces, partir de ahora supondremos que $r_N > 0$ y $\rho_N > 0$. Como $r \xrightarrow{a.s.} 0$, tenemos que $\|\phi_{f,1}\|_r \xrightarrow{a.s.} \|\phi_{f,1}\| = 1$. Entonces, usando que $\rho \xrightarrow{a.s.} 0$ implica que $\rho\Psi(\phi_{f,1})/\|\phi_{f,1}\|_r^2 \xrightarrow{a.s.} 0$, tenemos que (5.31) vale.

Sea $\omega \notin \mathcal{N}$ donde \mathcal{N} es el conjunto de probabilidad 0 donde no vale la convergencia casi segura de los supuestos ii) a iv) y veamos primero que, para todo $1 \leq i \leq k$, $\liminf \sigma_i^2(\widehat{\phi}_1(\omega)) > 0$. En efecto, supongamos que existe i_0 tal que $\liminf \sigma_{i_0}^2(\widehat{\phi}_1(\omega)) = 0$. Luego, existirá una subsucesión de $\gamma_N = \widehat{\phi}_1(\omega)$ tal que $\sigma_{i_0}^2(\gamma_{N_\ell}) \rightarrow 0$, usando que $\widehat{a}_{N,i_0} = \sigma_{i_0,n_{i_0}}^2(\widehat{\phi}_1(\omega)) - \sigma_{i_0}^2(\widehat{\phi}_1(\omega)) \rightarrow 0$, obtenemos que $\sigma_{i_0,n_{i_0}}^2(\gamma_{N_\ell}) \rightarrow 0$ y entonces, $\varsigma_{N_\ell}(\gamma_{N_\ell}) \rightarrow -\infty$. Luego, usando (5.33), obtenemos que $\varsigma(\phi_{f,1}) = -\infty$ lo que contradice el hecho que $\lambda_{f,i,1} > 0$. Entonces, para todo $1 \leq i \leq k$, $\liminf \sigma_i^2(\widehat{\phi}_1(\omega)) >$

0, entonces existe $\epsilon > 0$ tal que $\sigma_i^2(\widehat{\phi}_1(\omega)) \in [\epsilon, +\infty)$. Más aún, usando que σ_i es débilmente continuo y que la esfera unitaria es débilmente compacta, tenemos que $\{\sigma_i(\alpha) : \alpha \in \mathcal{V}_1\}$ es un conjunto acotado. Luego existe $\epsilon > 0$ y $A > 0$ tal que $\sigma_i^2(\widehat{\phi}_1(\omega)) \in [\epsilon, A]$, para todo $1 \leq i \leq k$. Usando que $\widehat{a}_{N,i}(\omega) = \sigma_{i,n_i}^2(\widehat{\phi}_1(\omega)) - \sigma_i^2(\widehat{\phi}_1(\omega)) \rightarrow 0$, obtenemos que para N suficientemente grande, $\sigma_{i,n_i}^2(\widehat{\phi}_1(\omega)) \in [\epsilon/2, 2A]$ que junto con la continuidad uniforme del logaritmo en $[\epsilon/2, 2A]$ implican que

$$\widehat{a}_N = \varsigma_N(\widehat{\phi}_1) - \varsigma(\widehat{\phi}_1) \xrightarrow{a.s.} 0. \quad (5.34)$$

Usando que σ_R es un funcional de escala, f es estrictamente creciente, $\varsigma(\phi_{f,1}) = \sup_{\alpha \in \mathcal{S}} \varsigma(\alpha)$ y el hecho que $\|\widehat{\phi}_1\| \leq \|\widehat{\phi}_1\|_r = 1$, obtenemos fácilmente que

$$\varsigma(\phi_{f,1}) \geq \varsigma\left(\frac{\widehat{\phi}_1}{\|\widehat{\phi}_1\|}\right) \geq \varsigma(\widehat{\phi}_1) = \varsigma_N(\widehat{\phi}_1) - \widehat{a}_N = \varsigma_N(\widehat{\phi}_1) + o_{a.s.}(1)$$

concluyendo la demostración de (5.30) y entonces, $\varsigma_N(\widehat{\phi}_1) \xrightarrow{a.s.} \varsigma(\phi_{f,1})$.

Como se mencionó antes, de (5.34), y utilizando que $\varsigma_N(\widehat{\phi}_1) \xrightarrow{a.s.} \varsigma(\phi_{f,1})$, obtenemos que

$$\varsigma(\widehat{\phi}_1) \xrightarrow{a.s.} \varsigma(\phi_{f,1}). \quad (5.35)$$

Más aún, las desigualdades

$$\varsigma(\phi_{f,1}) \geq \varsigma\left(\frac{\widehat{\phi}_1}{\|\widehat{\phi}_1\|}\right) \geq \varsigma(\widehat{\phi}_1)$$

obtenidas antes implican que

$$\varsigma\left(\frac{\widehat{\phi}_1}{\|\widehat{\phi}_1\|}\right) \xrightarrow{a.s.} \varsigma(\phi_{f,1}). \quad (5.36)$$

Debemos probar que $\rho\Psi(\widehat{\phi}_1) \xrightarrow{a.s.} 0$, que sigue como en la demostración del Teorema 5.5.1.

Queda ver que $r\Psi(\widehat{\phi}_1) \xrightarrow{a.s.} 0$. Usando que $\|\widehat{\phi}_1\|_r = 1$, obtenemos que $r\Psi(\widehat{\phi}_1) = 1 - \|\widehat{\phi}_1\|^2$. Luego, solamente debemos probar que $\|\widehat{\phi}_1\| \xrightarrow{a.s.} 1$. Ahora bien,

$$\varsigma\left(\frac{\widehat{\phi}_1}{\|\widehat{\phi}_1\|}\right) = \varsigma(\widehat{\phi}_1) - \log(\|\widehat{\phi}_1\|^2)$$

y entonces,

$$\varsigma(\widehat{\phi}_1) - \varsigma\left(\frac{\widehat{\phi}_1}{\|\widehat{\phi}_1\|}\right) = \log(\|\widehat{\phi}_1\|^2).$$

Usando (5.35) y (5.36), obtenemos que $\log(\|\widehat{\phi}_1\|^2) \xrightarrow{a.s.} 0$, entonces $\|\widehat{\phi}_1\|^2 \xrightarrow{a.s.} 1$ concluyendo la demostración de a).

b) La demostración de $\langle \widehat{\phi}_1, \phi_{f,1} \rangle^2 \xrightarrow{a.s.} 1$ se sigue inmediatamente del Lema 5.5.4, mientras que el hecho que $\widehat{\lambda}_{i,1} \xrightarrow{a.s.} \lambda_{f,i,1}$ se obtiene como en la demostración del Teorema 5.5.1.

c) Primero notemos que, como en la demostración del Teorema 5.5.1, usando que para $1 \leq j \leq m-1$, $\widehat{\phi}_j \xrightarrow{a.s.} \phi_{f,j}$ y $r\Psi(\widehat{\phi}_j) \xrightarrow{a.s.} 0$, obtenemos que (5.23) vale.

Usando que (5.13), (5.23) y el hecho que $\|\widehat{\phi}_m\|_r = 1$ y $\|\phi_{f,m}\| = 1$, obtenemos que

$$\begin{aligned}\widehat{a}_{N,i} &= \sigma_{i,n_i}^2(\widehat{\phi}_m) - \sigma_i^2(\widehat{\phi}_m) \xrightarrow{a.s.} 0, \\ \widehat{b}_{N,i} &= \sigma_{i,n_i}^2(\pi_{m-1}\widehat{\phi}_m) - \sigma_i^2(\pi_{m-1}\widehat{\phi}_m) \xrightarrow{a.s.} 0, \\ \widehat{c}_{N,i} &= \sigma_{i,n_i}^2(\phi_{f,m}) - \sigma_i^2(\phi_{f,m}) \xrightarrow{a.s.} 0, \\ \widehat{A}_{N,i} &= \sigma_{i,n_i}^2(\widehat{\pi}_{r,m-1}\phi_{f,m}) - \sigma_i^2(\phi_{f,m}) \\ &= \sigma_{i,n_i}^2(\widehat{\pi}_{r,m-1}\phi_{f,m}) - \sigma_i^2(\widehat{\pi}_{r,m-1}\phi_{f,m}) + \sigma_i^2(\widehat{\pi}_{r,m-1}\phi_{f,m}) - \sigma_i^2(\phi_{f,m}) \xrightarrow{a.s.} 0.\end{aligned}$$

Más aún, usando (5.23), obtenemos que

$$\|\pi_{m-1}\widehat{\phi}_m - \widehat{\phi}_m\| = \|\pi_{m-1}\widehat{\phi}_m - \widehat{\pi}_{r,m-1}\widehat{\phi}_m\| \leq \sup_{\|\alpha\|_r \leq 1} \|\pi_{m-1}\alpha - \widehat{\pi}_{r,m-1}\alpha\| \xrightarrow{a.s.} 0.$$

Luego, $\pi_{m-1}\widehat{\phi}_m - \widehat{\phi}_m \xrightarrow{a.s.} 0$. Apelando a que $\pi_{m-1}\widehat{\phi}_m \in \mathcal{V}_1$ y que $\widehat{\phi}_m \in \mathcal{V}_1$ y el hecho que σ_i es uniformemente débilmente continuo en \mathcal{V}_1 , obtenemos que $\widehat{B}_{N,i} = \sigma_{i,n_i}^2(\pi_{m-1}\widehat{\phi}_m) - \sigma_i^2(\widehat{\phi}_m) \xrightarrow{a.s.} 0$.

Argumentando como en a), y usando que para todo $1 \leq i \leq k$, $\sigma_i^2(\phi_{f,m}) > 0$, deducimos que

$$\widehat{a}_N = \varsigma_N(\widehat{\phi}_m) - \varsigma(\widehat{\phi}_m) \xrightarrow{a.s.} 0, \quad (5.37)$$

$$\widehat{A}_N = \varsigma_N(\widehat{\pi}_{r,m-1}\phi_{f,m}) - \varsigma(\phi_{f,m}) \xrightarrow{a.s.} 0, \quad (5.38)$$

$$\widehat{b}_N = \varsigma_N(\pi_{m-1}\widehat{\phi}_m) - \varsigma(\pi_{m-1}\widehat{\phi}_m) \xrightarrow{a.s.} 0, \quad (5.39)$$

$$\widehat{B}_N = \varsigma(\pi_{m-1}\widehat{\phi}_m) - \varsigma(\widehat{\phi}_m) \xrightarrow{a.s.} 0. \quad (5.40)$$

Más aún, también obtenemos que $\log(\sigma_{i,n_i}^2(\phi_{f,m})) \xrightarrow{a.s.} \log(\sigma_i^2(\phi_{f,m}))$ lo que implica que $\varsigma_N(\phi_{f,m}) \xrightarrow{a.s.} \varsigma(\phi_{f,m})$.

Como en a) y en el Teorema 5.5.1, primero debemos probar que

$$\varsigma(\phi_{f,m}) \geq \varsigma_N(\widehat{\phi}_m) + o_{a.s.}(1), \quad (5.41)$$

$$\varsigma(\phi_{f,m}) \leq \varsigma_N(\widehat{\phi}_m) + o_{a.s.}(1), \quad (5.42)$$

lo que asegura que $\varsigma_N(\widehat{\phi}_m) \xrightarrow{a.s.} \varsigma(\phi_{f,m})$, como se deseaba.

Vamos a probar (5.41). Como en la demostración del Paso 2 de la demostración del Teorema 5.5.1, tenemos que

$$\varsigma(\phi_{f,m}) = \sup_{\alpha \in \mathcal{S} \cap \mathcal{T}_{m-1}} \varsigma(\alpha) = \sup_{\alpha \in \mathcal{V}_1 \cap \mathcal{T}_{m-1}} \varsigma(\alpha).$$

Luego, usando que $\|\widehat{\phi}_m\| \leq 1$ y que $f(t) = \log(t)$ es creciente, obtenemos

$$\varsigma(\phi_{f,m}) = \sup_{\alpha \in \mathcal{V}_1 \cap \mathcal{T}_{m-1}} \varsigma(\alpha) \geq \varsigma\left(\pi_{m-1}\left(\frac{\widehat{\phi}_m}{\|\widehat{\phi}_m\|}\right)\right) \geq \varsigma\left(\frac{1}{\|\widehat{\phi}_m\|}\pi_{m-1}\widehat{\phi}_m\right) \geq \varsigma(\pi_{m-1}\widehat{\phi}_m).$$

Por otro lado, (5.37) y (5.40) implican que $\varsigma(\pi_{m-1}\widehat{\phi}_m) = \varsigma_N(\widehat{\phi}_m) - \widehat{a}_N - \widehat{B}_N = \varsigma_N(\widehat{\phi}_m) + o_{a.s.}(1)$ concluyendo la demostración de (5.41).

Procedemos ahora a probar (5.42). Como en el Paso 3 de la demostración del Teorema 5.5.1, tenemos que

$$\begin{aligned} \varsigma_N(\widehat{\phi}_m) &\geq \varsigma_N(\widehat{\phi}_m) - \rho\Psi(\widehat{\phi}_m) = \sup_{\|\alpha\|_r=1, \alpha \in \widehat{\mathcal{T}}_{r,m-1}} \{\varsigma_N(\alpha) - \rho\Psi(\alpha)\} \\ &\geq \varsigma_N\left(\frac{\widehat{\pi}_{r,m-1}\phi_{f,m}}{\|\widehat{\pi}_{r,m-1}\phi_{f,m}\|_r}\right) - \rho\Psi\left(\frac{\widehat{\pi}_{r,m-1}\phi_{f,m}}{\|\widehat{\pi}_{r,m-1}\phi_{f,m}\|_r}\right) \\ &\geq \varsigma_N\left(\frac{\widehat{\pi}_{r,m-1}\phi_{f,m}}{\|\phi_{f,m}\|_r}\right) - \rho\Psi\left(\frac{\widehat{\pi}_{r,m-1}\phi_{f,m}}{\|\widehat{\pi}_{r,m-1}\phi_{f,m}\|_r}\right) \\ &\geq \varsigma_N\left(\frac{\widehat{\pi}_{r,m-1}\phi_{f,m}}{\|\phi_{f,m}\|_r}\right) - \rho\frac{\Psi(\widehat{\pi}_{r,m-1}\phi_{f,m})}{\|\widehat{\pi}_{r,m-1}\phi_{f,m}\|_r^2}. \end{aligned}$$

Notemos que $\varsigma_N(\widehat{\pi}_{r,m-1}\phi_{f,m}/\|\phi_{f,m}\|_r) = \varsigma_N(\widehat{\pi}_{r,m-1}\phi_{f,m}) - 2\log(\|\phi_{f,m}\|_r)$, luego (5.38) y el hecho que $\|\phi_{f,m}\|_r \xrightarrow{a.s.} 1$ implican que

$$\begin{aligned} \varsigma_N(\widehat{\phi}_m) &\geq \varsigma_N\left(\frac{\widehat{\pi}_{r,m-1}\phi_{f,m}}{\|\phi_{f,m}\|_r}\right) - \rho\frac{\Psi(\widehat{\pi}_{r,m-1}\phi_{f,m})}{\|\widehat{\pi}_{r,m-1}\phi_{f,m}\|_r^2} \\ &\geq \varsigma(\phi_{f,m}) - \widehat{A}_N - 2\log(\|\phi_{f,m}\|_r) - \rho\frac{\Psi(\widehat{\pi}_{r,m-1}\phi_{f,m})}{\|\widehat{\pi}_{r,m-1}\phi_{f,m}\|_r^2} \\ &\geq \varsigma(\phi_{f,m}) + o_{a.s.}(1) - \rho\frac{\Psi(\widehat{\pi}_{r,m-1}\phi_{f,m})}{\|\widehat{\pi}_{r,m-1}\phi_{f,m}\|_r^2}. \end{aligned}$$

Finalmente, la demostración de (5.42) se sigue como en la demostración del Teorema 5.5.1 de (5.26), (5.27) y el hecho que $\|\widehat{\pi}_{r,m-1}\phi_{f,m}\|_r^2 \xrightarrow{a.s.} 1$.

Usando que $\varsigma_N(\widehat{\phi}_m) \xrightarrow{a.s.} \varsigma(\phi_{f,m})$ y que $\varsigma_N(\widehat{\phi}_m) - \varsigma(\widehat{\phi}_m) \xrightarrow{a.s.} 0$, obtenemos fácilmente que $\varsigma(\widehat{\phi}_m) \xrightarrow{a.s.} \varsigma(\phi_{f,m})$.

Queda ver que $\rho\Psi(\widehat{\phi}_m) \xrightarrow{a.s.} 0$ y $r\Psi(\widehat{\phi}_m) \xrightarrow{a.s.} 0$. La demostración de que $\rho\Psi(\widehat{\phi}_m) \xrightarrow{a.s.} 0$ es análoga a la del PASO 4 dada en la demostración del Teorema 5.5.1 mientras que la de que $r\Psi(\widehat{\phi}_m) \xrightarrow{a.s.} 0$ se obtiene usando argumentos análogos a aquellos considerados en el PASO 5.

d) Se obtiene igual que en la demostración del Teorema 5.5.1 usando el Lema 5.5.4 en vez del Lema 5.5.3. \square

Las ventajas de usar una función de *score* es que nos permitirá ajustar la eficiencia del estimador mediante la selección de dicha función. Más detalles para el caso finito dimensional y bajo el modelo proporcional se pueden ver en Rodrigues (2003). Conjeturamos que un resultado análogo al obtenido en el caso finito-dimensional vale en este contexto. Es decir que si el modelo es proporcional, para una escala dada, $f = \log$ minimiza la varianza asintótica de los estimadores de las direcciones principales comunes.

A continuación, se muestra la consistencia de los estimadores basados en *sieves*.

Teorema 5.5.3. Sean P_{i,n_i} , $1 \leq i \leq k$, sucesiones aleatorias de medidas de probabilidad, $N = \sum_{i=1}^k n_i$, $\hat{\tau}_{i,n_i}$ variables aleatorias tales que $\hat{\tau}_{i,n_i} \xrightarrow{a.s.} \tau_i$ con $0 < \tau_i < 1$, $\sum_{i=1}^k \tau_i = 1$. Sea $\sigma_{i,n_i}^2(\alpha) = \sigma_{\mathbb{R}}^2(P_{i,n_i}[\alpha])$ y $\varsigma_N(\alpha) = \sum_{i=1}^k \hat{\tau}_i f(\sigma_{i,n_i}^2(\alpha))$ con $f : [0, +\infty) \rightarrow \mathbb{R}$ y definamos $\hat{\lambda}_{i,m} = \sigma_{i,n_i}^2(\hat{\phi}_m)$ con

$$\begin{cases} \hat{\phi}_1 = \operatorname{argmax}_{\alpha \in \mathcal{H}_{p_n}, \|\alpha\|=1} \varsigma_N(\alpha) \\ \hat{\phi}_m = \operatorname{argmax}_{\alpha \in \hat{\mathcal{B}}_{m,n}} \varsigma_N(\alpha) \quad 2 \leq m, \end{cases}$$

donde $\hat{\mathcal{B}}_{m,n} = \{\alpha \in \mathcal{H}_{p_n} : \|\alpha\| = 1, \langle \alpha, \hat{\phi}_j \rangle = 0, \forall 1 \leq j \leq m-1\}$. Sea $P = P_1 \times \dots \times P_k$ una medida de probabilidad que satisface **C0** y $\phi_{f,m} = \phi_{f,m}(P)$ y $\lambda_{f,i,m} = \lambda_{f,i,m}(P)$ definidos como en (5.2) y (5.3), respectivamente. Supongamos que

i) **C1** y **C2** valen,

ii) $\sup_{\|\alpha\|=1} \left| \sigma_{i,n_i}^2(\alpha) - \sigma_i^2(\alpha) \right| \xrightarrow{a.s.} 0$.

iii) $p_n \rightarrow \infty$.

Luego,

a) $\varsigma_N(\hat{\phi}_1) \xrightarrow{a.s.} \varsigma(\phi_{f,1})$ y $\varsigma(\hat{\phi}_1) \xrightarrow{a.s.} \varsigma(\phi_{f,1})$.

b) $\langle \hat{\phi}_1, \phi_{f,1} \rangle^2 \xrightarrow{a.s.} 1$ y $\hat{\lambda}_{i,1} \xrightarrow{a.s.} \lambda_{f,i,1}$.

c) Para todo $2 \leq m \leq q$, si $\hat{\phi}_\ell \xrightarrow{a.s.} \phi_{f,\ell}$ para $1 \leq \ell \leq m-1$, luego $\varsigma_N(\hat{\phi}_m) \xrightarrow{a.s.} \varsigma(\phi_{f,m})$ y $\varsigma(\hat{\phi}_m) \xrightarrow{a.s.} \varsigma(\phi_{f,m})$.

d) Para $1 \leq m \leq q$, $\langle \hat{\phi}_m, \phi_{f,m} \rangle^2 \xrightarrow{a.s.} 1$ y $\hat{\lambda}_{i,m} \xrightarrow{a.s.} \sigma_i^2(\phi_{f,m})$.

A partir de ahora, y como en la sección anterior, sea \mathcal{L}_k y $\hat{\mathcal{L}}_k$ el subespacio lineal generado por $\{\phi_{f,1}, \dots, \phi_{f,k}\}$ y $\{\hat{\phi}_1, \dots, \hat{\phi}_k\}$, respectivamente. Indiquemos por $\pi_{\mathcal{L}} : \mathcal{H} \rightarrow \mathcal{L}$ a la proyección ortogonal en el subespacio lineal \mathcal{L} y sea $\pi_k = \pi_{\mathcal{T}_k}$ y $\hat{\pi}_k = \pi_{\hat{\mathcal{T}}_k}$ donde $\mathcal{T}_k = \mathcal{L}_k^\perp$ y $\hat{\mathcal{T}}_k = \hat{\mathcal{L}}_k^\perp$. Más aún, π_0 y $\hat{\pi}_0$ se consideran como la identidad.

DEMOSTRACIÓN. Recordemos que **C2** implica que ς es un uniformemente débilmente continuo en $\mathcal{V}_1 = \{\|\alpha\| \leq 1\}$. Como veníamos haciendo en las demostraciones anteriores, tenemos que las siguientes convergencias valen

$$\sup_{\|\alpha\| \leq 1} |\varsigma_N(\alpha) - \varsigma(\alpha)| \xrightarrow{a.s.} 0. \quad (5.43)$$

Notemos además que puesto que f está definido en 0 podemos suponer sin pérdida de generalidad que $f(0) = 0$ lo que implica que $\varsigma(\phi_{f,m}) > 0$ para $1 \leq m \leq q$.

Probaremos simultáneamente a) y c), probando primero que

$$\varsigma_N(\widehat{\phi}_m) \geq \varsigma(\phi_{f,m}) + o_{\text{a.s.}}(1), \quad (5.44)$$

$$\varsigma_N(\widehat{\phi}_m) \leq \varsigma(\phi_{f,m}) + o_{\text{a.s.}}(1), \quad (5.45)$$

Obtengamos primero que (5.45). Notemos que (5.24) implica que

$$\varsigma(\phi_{f,m}) = \sup_{\|\alpha\|=1, \alpha \in \mathcal{T}_{m-1}} \varsigma(\alpha) \geq \sup_{\|\alpha\| \leq 1} \varsigma(\pi_{m-1}\alpha) \geq \varsigma(\pi_{m-1}\widehat{\phi}_m).$$

Usando que $\widehat{\phi}_\ell \xrightarrow{\text{a.s.}} \phi_{f,\ell}$ for $1 \leq \ell \leq m-1$, obtenemos que $\|\widehat{\pi}_{m-1} - \pi_{m-1}\| \xrightarrow{\text{a.s.}} 0$ de donde $\widehat{\pi}_{m-1}\widehat{\phi}_m - \pi_{m-1}\widehat{\phi}_m \xrightarrow{\text{a.s.}} 0$ ya que $\|\widehat{\pi}_{m-1}\widehat{\phi}_m - \pi_{m-1}\widehat{\phi}_m\| \leq \|\widehat{\pi}_{m-1} - \pi_{m-1}\| \|\widehat{\phi}_m\| \leq \|\widehat{\pi}_{m-1} - \pi_{m-1}\|$. La continuidad uniforme débil de ς in \mathcal{V}_1 implica que $\varsigma(\widehat{\pi}_{m-1}\widehat{\phi}_m) - \varsigma(\pi_{m-1}\widehat{\phi}_m) \xrightarrow{\text{a.s.}} 0$ que junto con el hecho que $\varsigma(\widehat{\pi}_{m-1}\widehat{\phi}_m) = \varsigma(\widehat{\phi}_m)$ nos conduce a

$$\varsigma(\phi_{f,m}) \geq \varsigma(\widehat{\phi}_m) + o_{\text{a.s.}}(1),$$

concluyendo la demostración de (5.45).

Probemos (5.44). Definamos $\widetilde{\phi}_m = \pi_{\mathcal{H}_{p_n}} \phi_{f,m} / \|\pi_{\mathcal{H}_{p_n}} \phi_{f,m}\|$. Usando que $p_n \rightarrow \infty$, obtenemos que $\widetilde{\phi}_m \rightarrow \phi_{f,m}$, mientras que el hecho que $\|\widehat{\pi}_{m-1} - \pi_{m-1}\| \xrightarrow{\text{a.s.}} 0$ implica que $\widehat{\pi}_{m-1}\widetilde{\phi}_m \xrightarrow{\text{a.s.}} \pi_{m-1}\phi_{f,m} = \phi_{f,m}$ ya que

$$\|\widehat{\pi}_{m-1}\widetilde{\phi}_m - \pi_{m-1}\phi_{f,m}\| \leq \|\widehat{\pi}_{m-1}\widetilde{\phi}_m - \pi_{m-1}\widetilde{\phi}_m\| + \|\pi_{m-1}\widetilde{\phi}_m - \pi_{m-1}\phi_{f,m}\| \leq \|\widehat{\pi}_{m-1} - \pi_{m-1}\| + \|\widetilde{\phi}_m - \phi_{f,m}\|.$$

Luego, la continuidad de ς implica que $a_n = \varsigma(\widehat{\pi}_{m-1}\widetilde{\phi}_m) - \varsigma(\phi_{f,m}) \xrightarrow{\text{a.s.}} 0$. Por otro lado, (5.43) implica que $b_n = \varsigma(\widehat{\pi}_{m-1}\widetilde{\phi}_m) - \varsigma_N(\widehat{\pi}_{m-1}\widetilde{\phi}_m) \xrightarrow{\text{a.s.}} 0$.

Usando que $\widehat{\phi}_j \in \mathcal{H}_{p_n}$, obtenemos que para todo $\alpha \in \mathcal{H}_{p_n}$, $\widehat{\pi}_{m-1}\alpha \in \mathcal{H}_{p_n}$ que junto con el hecho que f es creciente nos lleva a

$$\varsigma_N(\widehat{\phi}_m) = \sup_{\|\alpha\|=1, \alpha \in \widehat{\mathcal{B}}_{m,n}} \varsigma_N(\alpha) \geq \sup_{\|\alpha\| \leq 1, \alpha \in \mathcal{H}_{p_n}} \varsigma_N(\widehat{\pi}_{m-1}\alpha) \geq \varsigma_N(\widehat{\pi}_{m-1}\widetilde{\phi}_m).$$

Ahora, (5.44) se sigue fácilmente del hecho que $\varsigma_N(\widehat{\pi}_{m-1}\widetilde{\phi}_m) = \varsigma(\phi_{f,m}) + a_n + b_n$.

Luego, tenemos que $\varsigma_N(\widehat{\phi}_m) \xrightarrow{\text{a.s.}} \varsigma(\phi_{f,m})$, y por lo tanto $\varsigma(\widehat{\phi}_m) \xrightarrow{\text{a.s.}} \varsigma(\phi_{f,m})$ se deduce de (5.43) concluyendo la demostración de a) y c).

b) y d) La demostración de que $\langle \widehat{\phi}_m, \phi_{f,m} \rangle^2 \xrightarrow{\text{a.s.}} 1$ se deduce inmediatamente del Lema 5.5.3 y el hecho que $\varsigma(\widehat{\phi}_m) \xrightarrow{\text{a.s.}} \varsigma(\phi_{f,m})$, mientras que la convergencia $\widehat{\lambda}_{i,m} \xrightarrow{\text{a.s.}} \lambda_{f,i,m}$ se obtiene como en el Teorema 5.5.1 usando el supuesto ii). \square

El siguiente teorema establece la consistencia del estimador de *sieves* cuando se usa el logaritmo como función de *score*.

Teorema 5.5.4. Sea P_{i,n_i} , $1 \leq i \leq k$, una sucesión de medidas de probabilidad, $N = \sum_{i=1}^k n_i$, $\widehat{\tau}_{i,n_i}$ variables aleatorias tales que $\widehat{\tau}_{i,n_i} \xrightarrow{a.s.} \tau_i$ con $0 < \tau_i < 1$, $\sum_{i=1}^k \tau_i = 1$. Sea $r = r_N \geq 0$, $\rho = \rho_N \geq 0$ parámetros aleatorios de suavizado. Indiquemos por $\sigma_{i,n_i}^2(\alpha) = \sigma_R^2(P_{i,n_i}[\alpha])$ y $\varsigma_N(\alpha) = \sum_{i=1}^k \widehat{\tau}_i f(\sigma_{i,n_i}^2(\alpha))$ con $f = \log$ y definamos $\widehat{\lambda}_{i,m} = \sigma_{i,n_i}^2(\widehat{\phi}_m)$ con

$$\begin{cases} \widehat{\phi}_1 &= \operatorname{argmax}_{\alpha \in \mathcal{H}_{p_n}, \|\alpha\|=1} \varsigma_N(\alpha) \\ \widehat{\phi}_m &= \operatorname{argmax}_{\alpha \in \widehat{\mathcal{B}}_{m,n}} \varsigma_N(\alpha) \quad 2 \leq m, \end{cases}$$

donde $\widehat{\mathcal{B}}_{m,n} = \{\alpha \in \mathcal{H}_{p_n} : \|\alpha\| = 1, \langle \alpha, \widehat{\phi}_j \rangle = 0, \forall 1 \leq j \leq m-1\}$. Sea $P = P_1 \times \dots \times P_k$ una medida de probabilidad que satisface **C0** y $\phi_{f,m} = \phi_{f,m}(P)$ y $\lambda_{f,i,m} = \lambda_{f,i,m}(P)$ definidos como en (5.2) y (5.3), respectivamente, con $f = \log$ y supongamos que para todo $1 \leq i \leq k$, $\lambda_{f,i,1} > \lambda_{f,i,2} > \dots > \lambda_{f,i,q} > \lambda_{f,i,q+1}$. Más aún, supongamos que

i) **C1** vale.

ii) $\sup_{\|\alpha\|=1} \left| \sigma_{i,n_i}^2(\alpha) - \sigma_i^2(\alpha) \right| \xrightarrow{a.s.} 0$.

iii) $p_n \rightarrow \infty$.

Luego,

a) $\varsigma_N(\widehat{\phi}_1) \xrightarrow{a.s.} \varsigma(\phi_{f,1})$ y $\varsigma(\widehat{\phi}_1) \xrightarrow{a.s.} \varsigma(\phi_{f,1})$.

b) $\langle \widehat{\phi}_1, \phi_{f,1} \rangle^2 \xrightarrow{a.s.} 1$ y $\widehat{\lambda}_{i,1} \xrightarrow{a.s.} \lambda_{f,i,1}$.

c) Para todo $2 \leq m \leq q$, si $\widehat{\phi}_\ell \xrightarrow{a.s.} \phi_{f,\ell}$ con $1 \leq \ell \leq m-1$, luego $\varsigma_N(\widehat{\phi}_m) \xrightarrow{a.s.} \varsigma(\phi_{f,m})$ y $\varsigma(\widehat{\phi}_m) \xrightarrow{a.s.} \varsigma(\phi_{f,m})$.

d) Para $1 \leq m \leq q$, $\langle \widehat{\phi}_m, \phi_{f,m} \rangle^2 \xrightarrow{a.s.} 1$ y $\widehat{\lambda}_{i,m} \xrightarrow{a.s.} \sigma_i^2(\phi_{f,m})$.

DEMOSTRACIÓN. Notemos que

$$\sup_{\|\alpha\| \leq 1} |\sigma_{i,n_i}^2(\alpha) - \sigma_i^2(\alpha)| \xrightarrow{a.s.} 0. \quad (5.46)$$

consideremos, como se hizo en el Teorema 5.5.2, el conjunto \mathcal{N} de probabilidad 0 donde no vale la convergencia casi segura del supuesto (ii) y elijamos $\omega \notin \mathcal{N}$. Como en la demostración del Teorema 5.5.2, es fácil ver, usando $\lambda_{f,i,m} > 0$, que $\liminf \sigma_i^2(\widehat{\phi}_m(\omega)) > 0$ y que existirá $\epsilon > 0$ y $A > 0$ tal que para N suficientemente grande ambos $\sigma_i^2(\widehat{\phi}_m(\omega))$ y $\sigma_{i,n_i}(\widehat{\phi}_m(\omega))$ pertenecen a $[\epsilon, A]$. La continuidad del logaritmo en $[\epsilon, A]$ implica que $\varsigma_N(\widehat{\phi}_m) - \varsigma(\widehat{\phi}_m) \xrightarrow{a.s.} 0$.

Es también fácil ver que, en general, si γ es una sucesión aleatoria tal que $\gamma(\omega) - \widehat{\phi}_m(\omega) \rightarrow 0$ existirá $\epsilon > 0$ y A tal que para N suficientemente grande, se cumple que $\sigma_i^2(\widehat{\phi}_m(\omega))$, $\sigma_{i,n_i}^2(\widehat{\phi}_m(\omega))$,

$\sigma_i^2(\gamma(\omega))$ y $\sigma_{i,n_i}^2(\gamma(\omega))$ estarán en $[\epsilon, A]$ y la diferencia entre dos de ellos convergerá a cero. Esta última aseveración se sigue utilizando la continuidad débil uniforme de σ_i y el supuesto ii). Como en $[\epsilon, A]$ la función logaritmo es uniformemente continua, obtenemos que $\varsigma(\gamma(\omega)) - \varsigma(\widehat{\phi}_m(\omega)) \rightarrow 0$ y $\varsigma_N(\gamma(\omega)) - \varsigma(\widehat{\phi}_m(\omega)) \rightarrow 0$.

A modo de resumen, si γ es una sucesión aleatoria tal que $\gamma - \widehat{\phi}_m \xrightarrow{a.s.} 0$, entonces $\varsigma(\gamma) - \varsigma(\widehat{\phi}_m) \xrightarrow{a.s.} 0$, $\varsigma_N(\gamma) - \varsigma(\widehat{\phi}_m) \xrightarrow{a.s.} 0$ y $\varsigma_N(\gamma) - \varsigma(\widehat{\phi}_m) \xrightarrow{a.s.} 0$.

Ahora estamos listos para empezar con la demostración propiamente dicha. Demostraremos simultáneamente a) y c) probando que valen las siguientes desigualdades

$$\varsigma_N(\widehat{\phi}_m) \geq \varsigma(\phi_{f,m}) + o_{a.s.}(1), \quad (5.47)$$

$$\varsigma_N(\widehat{\phi}_m) \leq \varsigma(\phi_{f,m}) + o_{a.s.}(1). \quad (5.48)$$

Probemos que vale (5.48). Como en la demostración del Teorema 5.5.3, tenemos que

$$\varsigma(\phi_{f,m}) = \sup_{\|\alpha\|=1, \alpha \in \mathcal{T}_{m-1}} \varsigma(\alpha) \geq \sup_{\|\alpha\| \leq 1} \varsigma(\pi_{m-1}\alpha) \geq \varsigma(\pi_{m-1}\widehat{\phi}_m)$$

y $\|\widehat{\pi}_{m-1} - \pi_{m-1}\| \xrightarrow{a.s.} 0$. Luego, $\|\widehat{\pi}_{m-1}\widehat{\phi}_m - \pi_{m-1}\widehat{\phi}_m\| \leq \|\widehat{\pi}_{m-1} - \pi_{m-1}\| \|\widehat{\phi}_m\| \leq \|\widehat{\pi}_{m-1} - \pi_{m-1}\| \xrightarrow{a.s.} 0$. Usando que $\widehat{\pi}_{m-1}\widehat{\phi}_m = \widehat{\phi}_m$, obtenemos que $\widehat{\phi}_m - \pi_{m-1}\widehat{\phi}_m \xrightarrow{a.s.} 0$. Luego, tomando γ como $\pi_{m-1}\widehat{\phi}_m$, obtenemos que $\varsigma(\widehat{\phi}_m) - \varsigma(\pi_{m-1}\widehat{\phi}_m) \xrightarrow{a.s.} 0$ de donde

$$\varsigma(\phi_{f,m}) \geq \varsigma(\widehat{\phi}_m) + o_{a.s.}(1),$$

concluyendo la demostración de (5.48).

Probemos (5.47). Como en la demostración del Teorema 5.5.3, definamos $\widetilde{\phi}_m = \pi_{\mathcal{H}_{p_n}} \phi_{f,m} / \|\pi_{\mathcal{H}_{p_n}} \phi_{f,m}\|$ y recordemos que $\widetilde{\phi}_m \rightarrow \phi_{f,m}$. Luego, valen las desigualdades

$$\varsigma_N(\widehat{\phi}_m) = \sup_{\|\alpha\|=1, \alpha \in \widehat{\mathcal{B}}_{m,n}} \varsigma_N(\alpha) \geq \sup_{\|\alpha\| \leq 1, \alpha \in \mathcal{H}_{p_n}} \varsigma_N(\widehat{\pi}_{m-1}\alpha) \geq \varsigma_N(\widehat{\pi}_{m-1}\widetilde{\phi}_m).$$

Usando que $\|\widehat{\pi}_{m-1} - \pi_{m-1}\| \xrightarrow{a.s.} 0$ y $\widetilde{\phi}_m \rightarrow \phi_{f,m}$, obtenemos que $\widehat{\pi}_{m-1}\widetilde{\phi}_m \xrightarrow{a.s.} \pi_{m-1}\phi_{f,m} = \phi_{f,m}$. Notando que $\varsigma_N(\widehat{\pi}_{m-1}\widetilde{\phi}_m) = \varsigma(\phi_{f,m}) + (\varsigma_N(\widehat{\pi}_{m-1}\widetilde{\phi}_m) - \varsigma(\phi_{f,m}))$, la demostración de (5.47) se concluirá si probamos que $\varsigma_N(\widehat{\pi}_{m-1}\widetilde{\phi}_m) - \varsigma(\phi_{f,m}) \xrightarrow{a.s.} 0$.

Sea \mathcal{N} el conjunto de probabilidad cero donde el supuesto ii) no vale y donde $\widehat{\tau}_{i,n_i}$ no converge a τ_i , entonces para $\omega \notin \mathcal{N}$, $\widehat{\pi}_{m-1}\widetilde{\phi}_m \rightarrow \phi_{f,m}$ de donde, $\sigma_{i,n_i}^2(\widehat{\pi}_{m-1}\widetilde{\phi}_m(\omega)) \rightarrow \sigma_i^2(\phi_{f,m})$. Usando que $\sigma_i^2(\phi_{f,m}) > 0$, tenemos que, para N suficientemente grande, ambos $\sigma_i^2(\phi_{f,m})$ y $\sigma_{i,n_i}^2(\widehat{\pi}_{m-1}\widetilde{\phi}_m(\omega))$ pertenecen a $[\epsilon, A]$ para algún $\epsilon > 0$ y A . Usando que $f = \log$ es uniformemente continua sobre $[\epsilon, A]$, obtenemos que $\varsigma_N(\widehat{\pi}_{m-1}\widetilde{\phi}_m(\omega)) - \varsigma(\phi_{f,m}) \rightarrow 0$, como se deseaba.

De (5.47) y (5.48), obtenemos que $\varsigma_N(\widehat{\phi}_m) \xrightarrow{a.s.} \varsigma(\phi_{f,m})$.

Queda ver que $\varsigma(\widehat{\phi}_m) \xrightarrow{a.s.} \varsigma(\phi_{f,m})$. Esto sigue fácilmente del hecho que $\varsigma_N(\widehat{\phi}_m) \xrightarrow{a.s.} \varsigma(\phi_{f,m})$ y que $\varsigma_N(\widehat{\phi}_m) - \varsigma(\widehat{\phi}_m) \xrightarrow{a.s.} 0$.

b) y d) La demostración de que $\langle \widehat{\phi}_m, \phi_{f,m} \rangle^2 \xrightarrow{a.s.} 1$ se sigue inmediatamente del Lema 5.5.3 y el hecho que $\varsigma(\widehat{\phi}_m) \xrightarrow{a.s.} \varsigma(\phi_{f,m})$, mientras que el hecho que $\widehat{\lambda}_{i,m} \xrightarrow{a.s.} \lambda_{f,i,m}$ se sigue como en el Teorema 5.5.2 usando el supuesto ii). \square

5.6 Estudio de Monte Carlo

5.6.1 Modelo CPC de rango finito

Descripción del modelo

Consideramos observaciones de $k = 3$ poblaciones tales que

$$X_{i,\ell} = Z_{i1,\ell}\phi_1 + Z_{i2,\ell}\phi_2 + Z_{i3,\ell}\phi_3, \quad 1 \leq \ell \leq n_i, \quad 1 \leq i \leq 3$$

donde $\phi_i : [-1, 1] \rightarrow \mathbb{R}$ están dados por

$$\phi_1(x) = \sin(4\pi x) \quad \phi_2(x) = \cos(7\pi x) \quad \phi_3(x) = \cos(15\pi x).$$

y $Z_{ij,\ell}$ son independientes. Bajo el modelo central, que seguiremos indicando por C_0 , tendremos que

$$\left\{ \begin{array}{l} (Z_{i1,\ell}, Z_{i2,\ell}, Z_{i3,\ell})^T \sim N(\mathbf{0}, \text{diag}(\sigma_{i1}^2, \sigma_{i2}^2, \sigma_{i3}^2)), \quad 1 \leq \ell \leq n_i \\ \sigma_{i1} = 4 \quad \sigma_{i2} = 2 \quad \sigma_{i3} = 1 \quad i = 1 \\ \sigma_{i1} = 6 \quad \sigma_{i2} = 3 \quad \sigma_{i3} = 1 \quad i = 2 \\ \sigma_{i1} = 2 \quad \sigma_{i2} = 1 \quad \sigma_{i3} = 0.5 \quad i = 3 \end{array} \right.$$

Para cada población, el tamaño muestral es $n_i = 100$, $1 \leq i \leq k = 3$, o sea $N = 300$. Luego, el modelo central C_0 supone que $\mathbf{Z}_i = (Z_{i1}, Z_{i2}, Z_{i3})^T \sim N(\mathbf{0}, \mathbf{\Sigma}_i)$, donde $\mathbf{\Sigma}_i = \text{diag}(\lambda_{i,1}, \lambda_{i,2}, \lambda_{i,3})$ y $\boldsymbol{\lambda}_i = (\lambda_{i,1}, \lambda_{i,2}, \lambda_{i,3})^T$ están dados por $\boldsymbol{\lambda}_1 = (16, 4, 1)^T$, $\boldsymbol{\lambda}_2 = (36, 9, 1)^T$, $\boldsymbol{\lambda}_3 = (4, 1, 1/4)^T$.

Se consideran las siguientes contaminaciones, donde $X_{i,j}^{(c)}$ representará a las observaciones contaminadas

- $C_{2,\epsilon}$ es una contaminación en la segunda componente y las observaciones se definen como sigue,

$$X_{i,\ell}^{(c)} = Z_{i1,\ell}^{(c)}\phi_1 + Z_{i2,\ell}^{(c)}\phi_2 + Z_{i3,\ell}^{(c)}\phi_3, \quad 1 \leq \ell \leq n_i, \quad 1 \leq i \leq 3$$

donde $Z_{i1,\ell}^{(c)} \sim Z_{i1,\ell}$, $Z_{i3,\ell}^{(c)} \sim Z_{i3,\ell}$ y $Z_{i2,j}^{(c)} \sim (1 - \epsilon)N(0, \sigma_{i2}^2) + \epsilon N(\mu_{i,2}, (\sigma_{i2}^{(c)})^2)$. En todos los casos, los $Z_{ij,\ell}^{(c)}$ independientes, $(\sigma_{i2}^{(c)})^2 = 0.01$ y $\mu_{1,2} = 10$, $\mu_{2,2} = 15$ y $\mu_{3,2} = 5$. Indicamos por $P_{C_{2,\epsilon}}$ la medida de probabilidad conjunta de $(X_{1,1}^{(c)}, X_{2,1}^{(c)}, X_{3,1}^{(c)})$ bajo $C_{2,\epsilon}$.

El principal efecto es tener a las dos componentes ϕ_1 y ϕ_2 intercambiadas al considerar el funcional clásico. Lo mismo pasará en las otras dos poblaciones. En efecto, cuando $\epsilon = 0.2$, $\text{var}(Z_{2,2}) = 43.202 > 36 = \text{var}(Z_{2,1})$ y $\text{var}(Z_{3,2}) = 4.802 > 4\text{var}(Z_{3,1})$.

- Se consideraron dos contaminaciones en la tercer componente.

★ $C_{3,a,\epsilon}$ es una contaminación fuerte en la tercer componente y las observaciones serán generadas entonces como

$$X_{i,\ell}^{(c)} = Z_{i1,\ell}^{(c)}\phi_1 + Z_{i2,\ell}^{(c)}\phi_2 + Z_{i3,\ell}^{(c)}\phi_3, \quad 1 \leq \ell \leq n_i, \quad 1 \leq i \leq 3$$

donde $Z_{i1,\ell}^{(c)} \sim Z_{i1,\ell}$, $Z_{i2,\ell}^{(c)} \sim Z_{i2,\ell}$ y $Z_{i3,j}^{(c)} \sim (1 - \epsilon)N(0, \sigma_{i3}^2) + \epsilon N(\mu_{i,3}, (\sigma_{i3}^{(c)})^2)$. En todos los casos $Z_{ij,\ell}^{(c)}$ son independientes, $(\sigma_{i3}^{(c)})^2 = 0.01$ y $\mu_{1,3} = 15$, $\mu_{2,3} = 20$ y $\mu_{3,3} = 7$. Indicamos $P_{C_{3,a,\epsilon}}$ a la distribución conjunta de $(X_{1,1}^{(c)}, X_{2,1}^{(c)}, X_{3,1}^{(c)})$ bajo $C_{3,a,\epsilon}$.

★ $C_{3,b,\epsilon}$: las observaciones son tales que

$$X_{i,\ell}^{(c)} = Z_{i1,\ell}^{(c)}\phi_1 + Z_{i2,\ell}^{(c)}\phi_2 + Z_{i3,\ell}^{(c)}\phi_3, \quad 1 \leq \ell \leq n_i, \quad 1 \leq i \leq 3$$

donde ahora $Z_{i1,\ell}^{(c)} \sim Z_{i1,\ell}$, $Z_{i2,\ell}^{(c)} \sim Z_{i2,\ell}$ y $Z_{i3,j}^{(c)} \sim (1 - \epsilon)N(0, \sigma_{i3}^2) + \epsilon N(\mu_{i,3}, (\sigma_{i3}^{(c)})^2)$. En todos los casos, $Z_{ij,\ell}^{(c)}$ son independientes, $(\sigma_{i3}^{(c)})^2 = 0.01$ y $\mu_{1,3} = 6$, $\mu_{2,3} = 10$ y $\mu_{3,3} = 3$.

Como antes, $P_{C_{3,b,\epsilon}}$ indica la distribución conjunta de $(X_{1,1}^{(c)}, X_{2,1}^{(c)}, X_{3,1}^{(c)})$ bajo $C_{3,b,\epsilon}$.

- $C_{23,\epsilon}$: es una contaminación en la segunda y en la tercer componente, y según el nivel de contaminación todas las direcciones son modificadas cuando se usa el desvío estándar. En este caso,

$$X_{i,\ell}^{(c)} = Z_{i1,\ell}^{(c)}\phi_1 + Z_{i2,\ell}^{(c)}\phi_2 + Z_{i3,\ell}^{(c)}\phi_3, \quad 1 \leq \ell \leq n_i, \quad 1 \leq i \leq 3$$

donde $Z_{i1,\ell}^{(c)} \sim Z_{i1,\ell}$, $Z_{i2,j}^{(c)} \sim (1 - \epsilon)N(0, \sigma_{i2}^2) + \epsilon N(\mu_{i,2}, (\sigma_{i2}^{(c)})^2)$ y $Z_{i3,j}^{(c)} \sim (1 - \epsilon)N(0, \sigma_{i3}^2) + \epsilon N(\mu_{i,3}, (\sigma_{i3}^{(c)})^2)$. En todos los casos, $Z_{ij,\ell}^{(c)}$ son independientes, $(\sigma_{i2}^{(c)})^2 = 0.01$, $(\sigma_{i3}^{(c)})^2 = 0.01$ y los valores medios son $\mu_{1,2} = 10$, $\mu_{2,2} = 15$ y $\mu_{3,2} = 5$ para la segunda componente y $\mu_{1,3} = 15$, $\mu_{2,3} = 20$ y $\mu_{3,3} = 7$ para la tercera. $P_{C_{23,\epsilon}}$ es la distribución conjunta de $(X_{1,1}^{(c)}, X_{2,1}^{(c)}, X_{3,1}^{(c)})$ bajo $C_{23,\epsilon}$.

Consideramos dos porcentajes de contaminación: $\epsilon = 0.1$ y 0.2 . En un primer lugar, veamos como las distintas contaminaciones afectan las direcciones principales comunes definidas en (5.2) para $f = \text{id}$ y $f = \log$, cuando σ_R es el desvío estándar. Por otra parte, nos interesará saber si la eventual modificación de las direcciones se mantiene al penalizar la escala.

La Tabla 5.1 reporta los valores de $\Psi(\phi_j)$, donde $\Psi(\alpha) = \int_{-1}^1 (\alpha''(s))^2 ds$, y aquellos de $\zeta(\phi_j)$ para $f = \text{id}$ y $f = \log$ y $\rho = r = 0$ bajo C_0 , cuando el desvío estándar es tomado como funcional de escala. Llamemos $\zeta_\rho(\alpha) = \sum_{i=1}^3 (1/3) f(\text{var}(P_i[\alpha])) - \rho \Psi(\alpha)$ donde el subíndice ρ es para reforzar la dependencia en la penalización. Cuando la penalización es 0 indicaremos simplemente ζ . Es claro que, bajo C_0 , para cualquier valor de ρ , $\zeta_\rho((\phi_1)) > \zeta_\rho((\phi_2)) > \zeta_\rho((\phi_3))$, puesto que $\Psi(\phi_1) < \Psi(\phi_2) < \Psi(\phi_3)$.

Para las contaminaciones consideradas con $\epsilon = 0.2$, la Tabla 5.2 da la varianza de $\langle X_{i,1}^{(c)}, \phi_j \rangle = Z_{ij,1}^{(c)}$ mientras que la Tabla 5.3 reporta los valores de $\zeta(\phi_j)$, es decir cuando $\rho = r = 0$, así como el valor de $\Psi(\phi_j)$. Por otro lado, las Tablas 5.4 y 5.5 reportan las mismas magnitudes cuando $\epsilon = 0.1$.

Notemos que cuando se usa el desvío estándar, si $\rho = 0$, tenemos que

- Para $\epsilon = 0.2$

$\varsigma(\phi_j)$	$f = \text{id}$	$f = \log$	$\Psi(\phi_j)$
$j = 1$	$\frac{56}{3}$	$\frac{1}{3} \log(2304) \simeq 2.580$	$(16\pi^2)^2$
$j = 2$	$\frac{14}{3}$	$\frac{1}{3} \log(36) \simeq 1.195$	$(49\pi^2)^2$
$j = 3$	$\frac{9}{12}$	$-\frac{1}{3} \log(4) \simeq -0.462$	$(225\pi^2)^2$

Tabla 5.1: Valores de $\Psi(\phi_j)$ y de $\varsigma(\phi_j)$ cuando $\rho = r = 0$ bajo C_0 .

$\text{var}(Z_{ij,1}^{(c)})$	$i = 1$	$i = 2$	$i = 3$	$i = 1$	$i = 2$	$i = 3$
	$C_{2,\epsilon}$			$C_{23,\epsilon}$		
$j = 1$	16	36	4	16	36	4
$j = 2$	19.202	43.202	4.802	19.202	43.202	4.802
$j = 3$	1	1	1/4	36.802	64.802	8.042
	$C_{3,a,\epsilon}$			$C_{3,b,\epsilon}$		
$j = 1$	16	36	4	16	36	4
$j = 2$	4	9	1	4	9	1
$j = 3$	36.802	64.802	8.042	6.562	16.802	1.642

Tabla 5.2: Varianza de $\langle X_{i,1}^{(c)}, \phi_j \rangle = Z_{ij,1}^{(c)}$ bajo $C_{2,\epsilon}$, $C_{3,a,\epsilon}$, $C_{3,b,\epsilon}$ y $C_{23,\epsilon}$ cuando $\epsilon = 0.2$.

$\varsigma(\phi_j)$	$f = \text{id}$	$f = \log$	$f = \text{id}$	$f = \log$	$\Psi(\phi_j)$
	$C_{2,\epsilon}$		$C_{23,\epsilon}$		
$j = 1$	$\frac{56}{3}$	$\frac{1}{3} \log(2304) \simeq 2.580$	$\frac{56}{3}$	$\frac{1}{3} \log(2304) \simeq 2.580$	$(16\pi^2)^2$
$j = 2$	$\frac{67.206}{3}$	$\frac{1}{3} \log(3983.57) \simeq 2.763$	$\frac{67.206}{3}$	$\frac{1}{3} \log(3983.57) \simeq 2.763$	$(49\pi^2)^2$
$j = 3$	$\frac{9}{12}$	$-\frac{1}{3} \log(4) \simeq -0.462$	$\frac{109.646}{3}$	$\frac{1}{3} \log(19178.91) \sim 3.287$	$(225\pi^2)^2$
	$C_{3,a,\epsilon}$		$C_{3,b,\epsilon}$		
$j = 1$	$\frac{56}{3}$	$\frac{1}{3} \log(2304) \simeq 2.580$	$\frac{56}{3}$	$\frac{1}{3} \log(2304) \simeq 2.580$	$(16\pi^2)^2$
$j = 2$	$\frac{14}{3}$	$\frac{1}{3} \log(36) \simeq 1.195$	$\frac{14}{3}$	$\frac{1}{3} \log(36) \simeq 1.195$	$(49\pi^2)^2$
$j = 3$	$\frac{109.646}{3}$	$\frac{1}{3} \log(19178.91) \sim 3.287189$	$\frac{25.006}{3}$	$\frac{1}{3} \log(181.0383) \sim 1.733$	$(225\pi^2)^2$

Tabla 5.3: Valores de $\varsigma(\phi_j)$ para $\rho = r = 0$ bajo $C_{2,\epsilon}$, $C_{3,a,\epsilon}$, $C_{3,b,\epsilon}$ y $C_{23,\epsilon}$ cuando $\epsilon = 0.2$ y valores de $\Psi(\phi_j)$.

★ bajo $C_{2,0.2}$, $\varsigma(\phi_2) > \varsigma(\phi_1) > \varsigma(\phi_3)$. Entonces, cuando $\rho = 0$ tenemos que $\phi_{f,1}(PC_{2,0.2}) = \phi_2$, $\phi_{f,21}(PC_{2,0.2}) = \phi_1$ y $\phi_{f,3}(PC_{2,0.2}) = \phi_3$.

Similarmente, para las otras contaminaciones, tenemos que

★ $\phi_{f,1}(PC_{23,0.2}) = \phi_3$, $\phi_{f,2}(PC_{23,0.2}) = \phi_2$ y $\phi_{f,3}(PC_{23,0.2}) = \phi_1$

★ $\phi_{f,1}(PC_{3a,0.2}) = \phi_3$, $\phi_{f,2}(PC_{3a,0.2}) = \phi_1$ y $\phi_{f,3}(PC_{3a,0.2}) = \phi_2$

★ $\phi_{f,1}(PC_{3b,0.2}) = \phi_1$, $\phi_{f,2}(PC_{3b,0.2}) = \phi_3$ y $\phi_{f,3}(PC_{3b,0.2}) = \phi_2$

• Para $\epsilon = 0.1$

★ $\phi_{f,1}(PC_{2,0.1}) = \phi_1$, $\phi_{f,21}(PC_{2,0.1}) = \phi_2$ y $\phi_{f,3}(PC_{2,0.1}) = \phi_3$. Luego, esta cantidad de

$\text{var}(Z_{ij,1}^{(c)})$	$i = 1$	$i = 2$	$i = 3$	$i = 1$	$i = 2$	$i = 3$
	$C_{2,\epsilon}$			$C_{23,\epsilon}$		
$j = 1$	16	36	4	16	36	4
$j = 2$	12.601	28.351	3.151	12.601	28.351	3.151
$j = 3$	1	1	1/4	21.151	36.901	4.636
	$C_{3,a,\epsilon}$			$C_{3,b,\epsilon}$		
$j = 1$	16	36	4	16	36	4
$j = 2$	4	9	1	4	9	1
$j = 3$	21.151	36.901	4.636	4.141	9.901	1.036

Tabla 5.4: Varianza de $\langle X_{i,1}^{(c)}, \phi_j \rangle = Z_{ij,1}^{(c)}$ bajo $C_{2,\epsilon}$, $C_{3,a,\epsilon}$, $C_{3,b,\epsilon}$ y $C_{23,\epsilon}$ cuando $\epsilon = 0.1$.

$\zeta(\phi_j)$	$f = \text{id}$	$f = \log$	$f = \text{id}$	$f = \log$	$\Psi(\phi_j)$
	$C_{2,\epsilon}$		$C_{23,\epsilon}$		
$j = 1$	$\frac{56}{3}$	$\frac{1}{3} \log(2304) \simeq 2.580$	$\frac{56}{3}$	$\frac{1}{3} \log(2304) \simeq 2.580$	$(16\pi^2)^2$
$j = 2$	$\frac{44.103}{3}$	$\frac{1}{3} \log(1125.698) \simeq 2.342$	$\frac{44.103}{3}$	$\frac{1}{3} \log(1125.698) \simeq 2.342$	$(49\pi^2)^2$
$j = 3$	$\frac{3}{12}$	$-\frac{1}{3} \log(4) \simeq -0.462$	$\frac{62.688}{3}$	$\frac{1}{3} \log(3618.366) \simeq 2.731259$	$(225\pi^2)^2$
	$C_{3,a,\epsilon}$		$C_{3,b,\epsilon}$		
$j = 1$	$\frac{56}{3}$	$\frac{1}{3} \log(2304) \simeq 2.580$	$\frac{56}{3}$	$\frac{1}{3} \log(2304) \simeq 2.580$	$(16\pi^2)^2$
$j = 2$	$\frac{14}{3}$	$\frac{1}{3} \log(36) \simeq 1.195$	$\frac{14}{3}$	$\frac{1}{3} \log(36) \simeq 1.195$	$(49\pi^2)^2$
$j = 3$	$\frac{62.688}{3}$	$\frac{1}{3} \log(3618.366) \simeq 2.731259$	$\frac{15.078}{3}$	$\frac{1}{3} \log(42.47604) \simeq 1.249$	$(225\pi^2)^2$

Tabla 5.5: Valores de $\zeta(\phi_j)$ para $\rho = r = 0$ bajo $C_{2,\epsilon}$, $C_{3,a,\epsilon}$, $C_{3,b,\epsilon}$ y $C_{23,\epsilon}$ cuando $\epsilon = 0.1$ y valores de $\Psi(\phi_j)$.

contaminación podría no afectar a los estimadores clásicos y por lo tanto no se reportarán resultados para esta contaminación.

★ $\phi_{f,1}(P_{C_{23,0.1}}) = \phi_3$, $\phi_{f,2}(P_{C_{23,0.1}}) = \phi_1$ y $\phi_{f,3}(P_{C_{23,0.1}}) = \phi_2$.

En particular, cuando $f = \log$ los valores de la función objetivo son muy cercanos entre sí haciendo difícil la estimación de las direcciones.

★ $\phi_{f,1}(P_{C_{3a,0.1}}) = \phi_3$, $\phi_{f,2}(P_{C_{3a,0.1}}) = \phi_1$ y $\phi_{f,3}(P_{C_{3a,0.1}}) = \phi_2$.

★ $\phi_{f,1}(P_{C_{3b,0.1}}) = \phi_1$, $\phi_{f,2}(P_{C_{3b,0.1}}) = \phi_3$ y $\phi_{f,3}(P_{C_{3b,0.1}}) = \phi_2$.

Notmos que los valores $\zeta(\phi_2)$ y $\zeta(\phi_3)$ son muy cercanos, lo que puede llevar a un problema de identificación propia de las direcciones cuando se las estima.

Si incluimos ahora el término de penalización, debemos asegurarnos que la penalización ρ sea tal que el mismo orden se preservará cuando se considera el desvío estándar, puesto que un valor alto de ρ podría producir que el término dominante sea $\Psi(\alpha)$. Por esta razón solo consideramos valores de $\rho < \rho_{\max}$ donde ρ_{\max} está dado en la Tabla 5.6.

ϵ	$f = \text{id}$	$f = \log$	$f = \text{id}$	$f = \log$
0.2	$C_{2,\epsilon}$		$C_{23,\epsilon}$	
	$1.787733 * 10^{-5}$	$8.734965 * 10^{-7}$	$3.011559 * 10^{-6}$	$1.115237 * 10^{-7}$
0.2	$C_{3,a,\epsilon}$		$C_{3,b,\epsilon}$	
	$3.644629 * 10^{-6}$	$1.439728 * 10^{-7}$	$7.809901 * 10^{-7}$	$1.146145 * 10^{-7}$
0.1	$C_{2,\epsilon}$		$C_{23,\epsilon}$	
	—	—	$4.543727 * 10^{-7}$	$3.066579 * 10^{-8}$
0.1	$C_{3,a,\epsilon}$		$C_{3,b,\epsilon}$	
	$4.543727 * 10^{-7}$	$3.066579 * 10^{-8}$	$7.649531 * 10^{-8}$	$1.173836 * 10^{-8}$

Tabla 5.6: Valores de ρ_{\max} bajo $C_{2,\epsilon}$, $C_{3,a,\epsilon}$, $C_{3,b,\epsilon}$ y $C_{23,\epsilon}$ para $\epsilon = 0.1$ y 0.2 .

Algoritmo y notación

Como en el Capítulo 3, todos los métodos considerados aquí son modificaciones del método básico presentado por Croux y Ruiz–Gazen (1996) para el cómputo de las componentes principales utilizando *projection-pursuit*. En el Capítulo 6, se describe en detalle dicho procedimiento. Hemos adaptado también el algoritmo para permitir componentes principales comunes penalizadas. Las principales características de los algoritmos son exactamente las mismas que en el caso unipoblacional (funcional de escala, centrado, nivel de suavizado), omitimos su descripción y remitimos al lector a la Sección 3.6.1.

Tal como se hizo en el caso unipoblacional, para aplicar el algoritmo a los datos funcionales se discretiza el dominio de la función observada sobre $m = 50$ puntos equiespaciados en $\mathcal{I} = [-1, 1]$.

Como antes, indicamos a los estimadores correspondientes a cada tipo de escala como SD, MAD, M –ESCALA. Para cada escala, consideramos tres tipos de estimadores, el “crudo” que no tiene penalización, los estimadores obtenidos por penalización en la escala, tal como se definen en (3.6) y aquellos que se obtienen al penalizar en la norma, descritos en (3.5). En todas las tablas, el estimador de la j –ésima dirección principal vinculado a cada método están indicados como $\hat{\phi}_{\text{RAW},j}$, $\hat{\phi}_{\text{PS},j}$ y $\hat{\phi}_{\text{PN},j}$, respectivamente.

Cuando se utilizan los estimadores penalizados, se eligieron distintos valores para los parámetros de penalizado ρ y r . Como los valores grandes de dichos parámetros hacen que domine el término de penalización independientemente de la contaminación considerada, elegimos ρ y r iguales a $aN^{-\alpha}$ para $\alpha = 2, 3$ y 4 y a desde 0.1 hasta 5.5 . Como en el Capítulo 3, se reportan los valores medios sobre replicaciones de $\|\hat{\phi}_j / \|\hat{\phi}_j\| - \phi_j\|$ sobre 1000 replicaciones.

Resultados

Las Tablas 5.7 hasta 5.10 reportan los resultados obtenidos bajo C_0 para los diferentes estimadores cuando $f = \text{id}$ mientras que la Tabla 5.11 reporta los resultados cuando $f = \log$. Debido al hecho de que las trayectorias son suaves, el penalizado no tiende a mejorar la eficacia de los estimadores. Por otra parte, los estimadores robustos tienen un comportamiento semejante al de los estimadores

clásicos observándose una pérdida de eficiencia mayor al utilizar la MAD como en el caso de 1 población.

Las Tablas 5.12 y 5.13 reportan el comportamiento de los estimadores bajo las diferentes contaminaciones consideradas, cuando $f = \text{id}$ y cuando $f = \log$, respectivamente. Se informan los resultados para los estimadores RAW y para los penalizados cuando la penalización vale $3n^{-3}$. Como en el caso de una población, los estimadores robustos se comportan mucho mejor que los clásicos. Por otra parte, se obtienen menores errores cuadráticos medios cuando se utilizan estimadores suavizados combinados con la función de *score* $f = \log$.

Estimador de escala	a	$\widehat{\phi}_{PS,j}$			$\widehat{\phi}_{PN,j}$		
		$j = 1$	$j = 2$	$j = 3$	$j = 1$	$j = 2$	$j = 3$
SD	0	0.0033	0.0039	0.0037	0.0033	0.0039	0.0037
MAD	0	0.0436	0.0634	0.0368	0.0436	0.0634	0.0368
M -ESCALA	0	0.0108	0.0141	0.0077	0.0108	0.0141	0.0077
SD	0.1	0.0032	0.0033	0.0033	0.0025	0.0035	0.0055
MAD	0.1	0.0418	0.0494	0.0231	0.0291	0.0439	0.0664
M -ESCALA	0.1	0.0103	0.0111	0.0051	0.0068	0.0096	0.0093
SD	0.5	0.0029	0.0029	0.0032	0.0017	0.0053	0.0220
MAD	0.5	0.0384	0.0384	0.0110	0.0144	0.0475	0.1448
M -ESCALA	0.5	0.0089	0.0088	0.0038	0.0033	0.0105	0.0275
SD	1	0.0027	0.0028	0.0032	0.0015	0.0085	0.0534
MAD	1	0.0333	0.0326	0.0074	0.0091	0.0541	0.1986
M -ESCALA	1	0.0079	0.0078	0.0043	0.0023	0.0137	0.0613
SD	1.5	0.0026	0.0026	0.0032	0.0014	0.0122	0.0863
MAD	1.5	0.0307	0.0300	0.0060	0.0069	0.0605	0.2464
M -ESCALA	1.5	0.0070	0.0070	0.0035	0.0021	0.0180	0.0939
SD	2	0.0024	0.0025	0.0032	0.0014	0.0158	0.1165
MAD	2	0.0271	0.0265	0.0053	0.0057	0.0657	0.2738
M -ESCALA	2	0.0064	0.0064	0.0034	0.0019	0.0221	0.1270
SD	2.5	0.0023	0.0024	0.0032	0.0014	0.0195	0.1447
MAD	2.5	0.0255	0.0250	0.0049	0.0049	0.0701	0.3050
M -ESCALA	2.5	0.0059	0.0059	0.0034	0.0018	0.0257	0.1576
SD	3	0.0022	0.0023	0.0032	0.0014	0.0230	0.1688
MAD	3	0.0240	0.0235	0.0046	0.0046	0.0784	0.3341
M -ESCALA	3	0.0053	0.0054	0.0034	0.0017	0.0290	0.1849
SD	3.5	0.0021	0.0022	0.0032	0.0013	0.0261	0.1921
MAD	3.5	0.0222	0.0217	0.0044	0.0043	0.0832	0.3537
M -ESCALA	3.5	0.0050	0.0051	0.0034	0.0017	0.0326	0.2061
SD	4	0.0020	0.0022	0.0032	0.0013	0.0293	0.2132
MAD	4	0.0212	0.0208	0.0044	0.0039	0.0886	0.3784
M -ESCALA	4	0.0046	0.0047	0.0034	0.0016	0.0366	0.2265
SD	4.5	0.0020	0.0021	0.0032	0.0013	0.0325	0.2326
MAD	4.5	0.0201	0.0198	0.0042	0.0036	0.0949	0.3994
M -ESCALA	4.5	0.0044	0.0045	0.0034	0.0016	0.0398	0.2430
SD	5	0.0019	0.0021	0.0032	0.0013	0.0362	0.2481
MAD	5	0.0187	0.0184	0.0041	0.0035	0.1017	0.4205
M -ESCALA	5	0.0042	0.0043	0.0034	0.0016	0.0428	0.2607
SD	5.5	0.0019	0.0020	0.0032	0.0013	0.0395	0.2630
MAD	5.5	0.0173	0.0171	0.0040	0.0034	0.1028	0.4300
M -ESCALA	5.5	0.0040	0.0041	0.0034	0.0016	0.0460	0.2759

Tabla 5.7: Valores medios de $\|\widehat{\phi}_j/\|\widehat{\phi}_j\| - \phi_j\|^2$, bajo C_0 cuando ρ o $r = aN^{-2}$ y $f = \text{id}$.

Escala	a	$\widehat{\phi}_{PS,j}$			$\widehat{\phi}_{PN,j}$		
		$j = 1$	$j = 2$	$j = 3$	$j = 1$	$j = 2$	$j = 3$
SD	0	0.0033	0.0039	0.0037	0.0033	0.0039	0.0037
MAD	0	0.0436	0.0634	0.0368	0.0436	0.0634	0.0368
M -ESCALA	0	0.0108	0.0141	0.0077	0.0108	0.0141	0.0077
SD	0.1	0.0033	0.0039	0.0037	0.0033	0.0038	0.0037
MAD	0.1	0.0436	0.0634	0.0368	0.0436	0.0637	0.0377
M -ESCALA	0.1	0.0108	0.0140	0.0077	0.0108	0.0140	0.0076
SD	0.15	0.0033	0.0039	0.0037	0.0033	0.0038	0.0037
MAD	0.15	0.0436	0.0634	0.0368	0.0436	0.0635	0.0377
M -ESCALA	0.15	0.0108	0.0140	0.0076	0.0107	0.0139	0.0076
SD	0.25	0.0033	0.0038	0.0037	0.0033	0.0038	0.0037
MAD	0.25	0.0436	0.0633	0.0368	0.0434	0.0628	0.0376
M -ESCALA	0.25	0.0108	0.0140	0.0076	0.0107	0.0138	0.0076
SD	0.5	0.0033	0.0038	0.0037	0.0032	0.0038	0.0037
MAD	0.5	0.0436	0.0632	0.0367	0.0423	0.0611	0.0376
M -ESCALA	0.5	0.0108	0.0140	0.0076	0.0106	0.0136	0.0076
SD	0.75	0.0033	0.0038	0.0037	0.0032	0.0037	0.0037
MAD	0.75	0.0436	0.0629	0.0364	0.0419	0.0606	0.0381
M -ESCALA	0.75	0.0108	0.0139	0.0075	0.0106	0.0134	0.0076
SD	1	0.0033	0.0038	0.0037	0.0032	0.0037	0.0037
MAD	1	0.0436	0.0628	0.0363	0.0419	0.0603	0.0389
M -ESCALA	1	0.0108	0.0139	0.0075	0.0104	0.0132	0.0076
SD	1.5	0.0033	0.0038	0.0036	0.0032	0.0037	0.0037
MAD	1.5	0.0436	0.0624	0.0360	0.0417	0.0590	0.0390
M -ESCALA	1.5	0.0108	0.0138	0.0074	0.0103	0.0129	0.0075
SD	2	0.0033	0.0038	0.0036	0.0032	0.0036	0.0037
MAD	2	0.0436	0.0624	0.0360	0.0415	0.0581	0.0395
M -ESCALA	2	0.0107	0.0136	0.0073	0.0102	0.0126	0.0075
SD	2.5	0.0033	0.0038	0.0036	0.0032	0.0036	0.0038
MAD	2.5	0.0436	0.0620	0.0355	0.0410	0.0577	0.0406
M -ESCALA	2.5	0.0107	0.0135	0.0072	0.0100	0.0123	0.0074
SD	3	0.0033	0.0038	0.0036	0.0031	0.0036	0.0038
MAD	3	0.0435	0.0611	0.0345	0.0407	0.0566	0.0410
M -ESCALA	3	0.0107	0.0135	0.0072	0.0099	0.0121	0.0073
SD	3.5	0.0033	0.0037	0.0036	0.0031	0.0035	0.0038
MAD	3.5	0.0435	0.0609	0.0344	0.0406	0.0558	0.0416
M -ESCALA	3.5	0.0107	0.0134	0.0071	0.0097	0.0118	0.0074
SD	4	0.0033	0.0037	0.0036	0.0031	0.0035	0.0038
MAD	4	0.0435	0.0606	0.0340	0.0404	0.0552	0.0423
M -ESCALA	4	0.0107	0.0133	0.0070	0.0096	0.0117	0.0074
SD	4.5	0.0033	0.0037	0.0036	0.0031	0.0035	0.0039
MAD	4.5	0.0434	0.0602	0.0336	0.0403	0.0548	0.0432
M -ESCALA	4.5	0.0107	0.0133	0.0069	0.0095	0.0116	0.0076

Tabla 5.8: Valores medios de $\|\widehat{\phi}_j/\|\widehat{\phi}_j\| - \phi_j\|^2$, bajo C_0 cuando ρ o $r = aN^{-3}$ y $f = \text{id}$.

Escala	a	$\widehat{\phi}_{PS,j}$			$\widehat{\phi}_{PN,j}$		
SD	5	0.0033	0.0037	0.0036	0.0030	0.0034	0.0039
MAD	5	0.0434	0.0596	0.0330	0.0399	0.0543	0.0440
M -ESCALA	5	0.0107	0.0132	0.0069	0.0095	0.0115	0.0076
SD	5.5	0.0033	0.0037	0.0035	0.0030	0.0034	0.0039
MAD	5.5	0.0432	0.0591	0.0326	0.0397	0.0536	0.0441
M -ESCALA	5.5	0.0107	0.0131	0.0068	0.0094	0.0114	0.0076
SD	6	0.0032	0.0037	0.0035	0.0030	0.0034	0.0039
MAD	6	0.0432	0.0587	0.0321	0.0394	0.0532	0.0449
M -ESCALA	6	0.0107	0.0130	0.0068	0.0093	0.0113	0.0076
SD	6.5	0.0032	0.0036	0.0035	0.0030	0.0034	0.0040
MAD	6.5	0.0432	0.0582	0.0317	0.0393	0.0530	0.0458
M -ESCALA	6.5	0.0107	0.0130	0.0067	0.0092	0.0112	0.0077
SD	7	0.0032	0.0036	0.0035	0.0030	0.0034	0.0040
MAD	7	0.0431	0.0575	0.0310	0.0391	0.0525	0.0457
M -ESCALA	7	0.0107	0.0129	0.0067	0.0092	0.0111	0.0076
SD	7.5	0.0032	0.0036	0.0035	0.0030	0.0034	0.0040
MAD	7.5	0.0428	0.0568	0.0306	0.0385	0.0517	0.0458
M -ESCALA	7.5	0.0107	0.0128	0.0066	0.0091	0.0110	0.0076
SD	8	0.0032	0.0036	0.0035	0.0029	0.0034	0.0041
MAD	8	0.0428	0.0566	0.0304	0.0382	0.0514	0.0465
M -ESCALA	8	0.0106	0.0128	0.0065	0.0090	0.0110	0.0076
SD	8.5	0.0032	0.0036	0.0035	0.0029	0.0034	0.0041
MAD	8.5	0.0427	0.0564	0.0303	0.0380	0.0510	0.0472
M -ESCALA	8.5	0.0106	0.0127	0.0065	0.0090	0.0109	0.0075

Tabla 5.9: Valores medios de $\|\widehat{\phi}_j/\|\widehat{\phi}_j\| - \phi_j\|^2$, bajo C_0 cuando ρ o $r = aN^{-3}$ y $f = \text{id}$.

Escala	a	$\widehat{\phi}_{PS,j}$			$\widehat{\phi}_{PN,j}$		
		$j = 1$	$j = 2$	$j = 3$	$j = 1$	$j = 2$	$j = 3$
SD	0	0.0033	0.0039	0.0037	0.0033	0.0039	0.0037
MAD	0	0.0436	0.0634	0.0368	0.0436	0.0634	0.0368
M -ESCALA	0	0.0108	0.0141	0.0077	0.0108	0.0141	0.0077
SD	0.1	0.0033	0.0039	0.0037	0.0033	0.0039	0.0037
MAD	0.1	0.0436	0.0634	0.0368	0.0436	0.0634	0.0368
M -ESCALA	0.1	0.0108	0.0141	0.0077	0.0108	0.0141	0.0077
SD	0.15	0.0033	0.0039	0.0037	0.0033	0.0039	0.0037
MAD	0.15	0.0436	0.0634	0.0368	0.0436	0.0634	0.0368
M -ESCALA	0.15	0.0108	0.0141	0.0077	0.0108	0.0141	0.0077
SD	0.25	0.0033	0.0039	0.0037	0.0033	0.0039	0.0037
MAD	0.25	0.0436	0.0634	0.0368	0.0436	0.0634	0.0368
M -ESCALA	0.25	0.0108	0.0141	0.0077	0.0108	0.0141	0.0077
SD	0.5	0.0033	0.0039	0.0037	0.0033	0.0039	0.0037
MAD	0.5	0.0436	0.0634	0.0368	0.0436	0.0634	0.0368
M -ESCALA	0.5	0.0108	0.0141	0.0077	0.0108	0.0141	0.0077
SD	0.75	0.0033	0.0039	0.0037	0.0033	0.0039	0.0037
MAD	0.75	0.0436	0.0634	0.0368	0.0436	0.0634	0.0369
M -ESCALA	0.75	0.0108	0.0141	0.0077	0.0108	0.0141	0.0077
SD	1	0.0033	0.0039	0.0037	0.0033	0.0039	0.0037
MAD	1	0.0436	0.0634	0.0368	0.0436	0.0634	0.0369
M -ESCALA	1	0.0108	0.0141	0.0077	0.0108	0.0141	0.0077
SD	1.5	0.0033	0.0039	0.0037	0.0033	0.0039	0.0037
MAD	1.5	0.0436	0.0634	0.0368	0.0436	0.0634	0.0369
M -ESCALA	1.5	0.0108	0.0141	0.0077	0.0108	0.0141	0.0077
SD	2	0.0033	0.0039	0.0037	0.0033	0.0039	0.0037
MAD	2	0.0436	0.0634	0.0368	0.0436	0.0634	0.0369
M -ESCALA	2	0.0108	0.0141	0.0077	0.0108	0.0141	0.0077
SD	2.5	0.0033	0.0039	0.0037	0.0033	0.0039	0.0037
MAD	2.5	0.0436	0.0634	0.0368	0.0436	0.0634	0.0369
M -ESCALA	2.5	0.0108	0.0141	0.0077	0.0108	0.0141	0.0077

Tabla 5.10: Valores medios de $\|\widehat{\phi}_j/\|\widehat{\phi}_j\| - \phi_j\|^2$, bajo C_0 cuando ρ o $r = aN^{-4}$ y $f = \text{id}$.

Estimador de escala	a	$\widehat{\phi}_{PS,j}$			$\widehat{\phi}_{PN,j}$		
		$j = 1$	$j = 2$	$j = 3$	$j = 1$	$j = 2$	$j = 3$
SD	0	0.0026	0.0033	0.0041	0.0026	0.0033	0.0041
MAD	0	0.0352	0.0528	0.0341	0.0352	0.0528	0.0341
M -ESCALA	0	0.0078	0.0119	0.0088	0.0078	0.0119	0.0088
SD	0.1	0.0026	0.0032	0.0040	0.0026	0.0032	0.0041
MAD	0.1	0.0351	0.0523	0.0334	0.0351	0.0526	0.0341
M -ESCALA	0.1	0.0078	0.0117	0.0086	0.0078	0.0117	0.0087
SD	0.5	0.0025	0.0031	0.0040	0.0025	0.0031	0.0041
MAD	0.5	0.0349	0.0515	0.0327	0.0349	0.0519	0.0353
M -ESCALA	0.5	0.0077	0.0113	0.0082	0.0077	0.0114	0.0087
SD	1	0.0025	0.0031	0.0039	0.0025	0.0031	0.0041
MAD	1	0.0347	0.0496	0.0306	0.0347	0.0508	0.0358
M -ESCALA	1	0.0077	0.0110	0.0077	0.0077	0.0111	0.0086
SD	1.5	0.0025	0.0030	0.0038	0.0025	0.0030	0.0041
MAD	1.5	0.0340	0.0480	0.0288	0.0340	0.0490	0.0358
M -ESCALA	1.5	0.0076	0.0105	0.0073	0.0076	0.0106	0.0085
SD	2	0.0025	0.0029	0.0038	0.0025	0.0030	0.0041
MAD	2	0.0338	0.0465	0.0271	0.0338	0.0480	0.0364
M -ESCALA	2	0.0075	0.0102	0.0069	0.0075	0.0104	0.0084
SD	2.5	0.0024	0.0028	0.0037	0.0024	0.0029	0.0041
MAD	2.5	0.0334	0.0458	0.0261	0.0334	0.0475	0.0372
M -ESCALA	2.5	0.0074	0.0099	0.0066	0.0074	0.0101	0.0084
SD	3	0.0024	0.0028	0.0037	0.0024	0.0029	0.0041
MAD	3	0.0331	0.0444	0.0244	0.0331	0.0466	0.0373
M -ESCALA	3	0.0073	0.0096	0.0063	0.0073	0.0099	0.0084
SD	3.5	0.0024	0.0027	0.0036	0.0024	0.0028	0.0042
MAD	3.5	0.0328	0.0435	0.0232	0.0328	0.0460	0.0378
M -ESCALA	3.5	0.0072	0.0092	0.0060	0.0072	0.0096	0.0083
SD	4	0.0023	0.0027	0.0036	0.0023	0.0028	0.0042
MAD	4	0.0323	0.0422	0.0220	0.0324	0.0450	0.0380
M -ESCALA	4	0.0071	0.0090	0.0058	0.0072	0.0095	0.0084
SD	4.5	0.0023	0.0026	0.0035	0.0023	0.0027	0.0042
MAD	4.5	0.0321	0.0413	0.0210	0.0321	0.0445	0.0387
M -ESCALA	4.5	0.0070	0.0088	0.0056	0.0071	0.0093	0.0082
SD	5	0.0023	0.0026	0.0035	0.0023	0.0027	0.0042
MAD	5	0.0318	0.0406	0.0204	0.0319	0.0441	0.0398
M -ESCALA	5	0.0070	0.0086	0.0055	0.0070	0.0091	0.0083
SD	5.5	0.0023	0.0025	0.0035	0.0023	0.0027	0.0042
MAD	5.5	0.0315	0.0396	0.0194	0.0315	0.0433	0.0398
M -ESCALA	5.5	0.0069	0.0085	0.0053	0.0070	0.0090	0.0083

Tabla 5.11: Valores medios de $\|\widehat{\phi}_j/\|\widehat{\phi}_j\| - \phi_j\|^2$, bajo C_0 cuando ρ o $r = an^{-3}$ y $f = \log$.

Modelo	Estimador de Escala	$\hat{\phi}_{\text{RAW},j} (a = 0)$			$\hat{\phi}_{\text{PS},j} (a = 3)$			$\hat{\phi}_{\text{PN},j} (a = 3)$		
		$j = 1$	$j = 2$	$j = 3$	$j = 1$	$j = 2$	$j = 3$	$j = 1$	$j = 2$	$j = 3$
C_0	SD	0.0033	0.0039	0.0037	0.0033	0.0038	0.0036	0.0031	0.0036	0.0038
	MAD	0.0436	0.0634	0.0368	0.0435	0.0611	0.0345	0.0407	0.0566	0.0410
	M -ESCALA	0.0108	0.0141	0.0077	0.0107	0.0135	0.0072	0.0099	0.0121	0.0073
$C_{2,0.2}$	SD	1.4496	1.4496	0.0023	1.4470	1.4470	0.0022	1.3987	1.4060	0.0022
	MAD	0.3780	0.3903	0.0271	0.3779	0.3899	0.0267	0.3678	0.3853	0.0306
	M -ESCALA	0.4347	0.4354	0.0058	0.4340	0.4347	0.0057	0.4225	0.4325	0.0061
$C_{3,a,0.2}$	SD	1.8275	1.9238	1.9405	1.8247	1.9239	1.9402	1.7071	1.9241	1.9437
	MAD	0.2598	0.7858	0.8071	0.2572	0.7619	0.7855	0.2081	0.7528	0.8699
	M -ESCALA	0.2997	1.0419	1.0977	0.2960	1.0138	1.0722	0.2391	1.0190	1.1590
$C_{3,b,0.2}$	SD	0.0140	1.7596	1.7905	0.0136	1.7409	1.7725	0.0089	1.5634	1.6750
	MAD	0.0874	0.5092	0.5013	0.0854	0.4850	0.4774	0.0651	0.4010	0.4910
	M -ESCALA	0.0427	0.4782	0.4966	0.0413	0.4513	0.4704	0.0253	0.3441	0.4699
$C_{23,0.2}$	SD	1.8303	0.1930	1.8255	1.8295	0.1964	1.8233	1.7296	0.3014	1.7622
	MAD	0.9350	1.0709	0.6005	0.9305	1.0663	0.5936	0.8700	1.0133	0.6702
	M -ESCALA	1.0599	1.2001	0.7009	1.0564	1.1928	0.6944	0.9940	1.0974	0.7602
$C_{3,a,0.1}$	SD	1.1438	1.9082	1.9172	1.1096	1.9059	1.9155	0.4586	1.8630	1.9160
	MAD	0.0929	0.2257	0.2339	0.0919	0.2145	0.2229	0.0747	0.1976	0.2627
	M -ESCALA	0.0847	0.2123	0.2628	0.0832	0.2020	0.2528	0.0597	0.1778	0.2843
$C_{3,b,0.1}$	SD	0.0063	0.9221	0.9459	0.0061	0.7742	0.7973	0.0047	0.3208	0.4022
	MAD	0.0545	0.1691	0.1547	0.0540	0.1581	0.1442	0.0469	0.1329	0.1599
	M -ESCALA	0.0200	0.1237	0.1294	0.0196	0.1106	0.1163	0.0154	0.0773	0.1132
$C_{23,0.1}$	SD	1.1762	1.3131	1.5685	1.1477	1.3071	1.5515	0.5761	1.1000	1.1980
	MAD	0.1808	0.3205	0.2415	0.1785	0.3131	0.2349	0.1534	0.2877	0.2706
	M -ESCALA	0.1850	0.3601	0.2862	0.1834	0.3525	0.2795	0.1524	0.3156	0.3141

Tabla 5.12: Valores medios de $\|\hat{\phi}_j/\|\hat{\phi}_j\| - \phi_j\|^2$ cuando ρ o $r = an^{-3}$ para $a = 0, 3$ y $f = \text{id}$.

Modelo	Estimador de Escala	$\widehat{\phi}_{\text{RAW},j}$			$\widehat{\phi}_{\text{PS},j}$			$\widehat{\phi}_{\text{PN},j}$		
		$j = 1$	$j = 2$	$j = 3$	$j = 1$	$j = 2$	$j = 3$	$j = 1$	$j = 2$	$j = 3$
C_0	SD	0.0026	0.0033	0.0041	0.0024	0.0028	0.0037	0.0024	0.0029	0.0041
	MAD	0.0352	0.0528	0.0341	0.0331	0.0444	0.0244	0.0331	0.0466	0.0373
	M -ESCALA	0.0078	0.0119	0.0088	0.0073	0.0096	0.0063	0.0073	0.0099	0.0084
$C_{2,0.2}$	SD	1.5317	1.5316	0.0023	1.4808	1.4808	0.0024	1.4828	1.4898	0.0024
	MAD	0.3724	0.3820	0.0250	0.3596	0.3656	0.0189	0.3596	0.3768	0.0295
	M -ESCALA	0.4350	0.4359	0.0062	0.4193	0.4196	0.0050	0.4193	0.4295	0.0065
$C_{3,a,0.2}$	SD	1.8608	1.9362	1.9453	1.7376	1.9369	1.9229	1.7766	1.9372	1.9495
	MAD	0.2649	0.8389	0.8763	0.2111	0.5327	0.5904	0.2138	0.8055	0.9377
	M -ESCALA	0.3058	1.1085	1.1667	0.2429	0.6651	0.7634	0.2480	1.0856	1.2310
$C_{3,b,0.2}$	SD	0.0096	1.7664	1.8026	0.0067	1.4635	1.5061	0.0067	1.5608	1.6813
	MAD	0.0608	0.4761	0.4801	0.0444	0.3102	0.3156	0.0451	0.3543	0.4597
	M -ESCALA	0.0274	0.4591	0.4827	0.0159	0.3099	0.3316	0.0159	0.3351	0.4722
$C_{23,0.2}$	SD	1.8585	0.1266	1.8542	1.7549	0.2603	1.7420	1.7857	0.1954	1.8153
	MAD	0.9219	1.1274	0.5867	0.8667	1.0157	0.4736	0.8699	1.0779	0.6581
	M -ESCALA	1.0565	1.2364	0.6868	1.0028	1.1224	0.5483	1.0052	1.1709	0.7590
$C_{3,a,0.1}$	SD	1.2444	1.9202	1.9291	0.3998	1.8529	1.8860	0.4865	1.8703	1.9251
	MAD	0.0858	0.2198	0.2400	0.0631	0.1417	0.1568	0.0636	0.1776	0.2533
	M -ESCALA	0.0838	0.2088	0.2630	0.0523	0.1285	0.1709	0.0531	0.1590	0.2713
$C_{3,b,0.1}$	SD	0.0043	0.8526	0.8813	0.0036	0.1738	0.1894	0.0036	0.2166	0.2883
	MAD	0.0425	0.1551	0.1491	0.0364	0.0993	0.0906	0.0364	0.1115	0.1421
	M -ESCALA	0.0135	0.1177	0.1267	0.0104	0.0662	0.0720	0.0104	0.0694	0.1070
$C_{23,0.1}$	SD	1.2966	1.4357	1.6548	0.5249	1.1194	1.1114	0.6028	1.1933	1.3057
	MAD	0.1664	0.3169	0.2501	0.1394	0.2471	0.1837	0.1398	0.2857	0.2830
	M -ESCALA	0.1688	0.3476	0.2947	0.1336	0.2605	0.2130	0.1342	0.2972	0.3210

Tabla 5.13: Valores medios de $\|\widehat{\phi}_j/\|\widehat{\phi}_j\| - \phi_j\|^2$ cuando ρ o $r = an^{-3}$ para $a = 0, 3$ y $f = \log$.

Validación cruzada

Un procedimiento de validación cruzada mediante K -th fold fue implementado para seleccionar automáticamente el parámetro de penalizado. Es, en esencia, una modificación de la propuesta hecha para el caso de una población, con la diferencia de que ahora, si estamos haciendo penalización por la escala, permitimos distintos valores ρ de penalizado para cada componente. Para estimar la primer componente y determinar el ρ_1 (el ρ de la primer componente), puede ser descripto como sigue:

1. Se centran los datos como se hacía en el caso unipoblacional, usando $\tilde{X}_{i,j} = X_{i,j} - \hat{\mu}_i$.
2. Particionar los datos centrados $\{\tilde{X}_{i,j}\}$ de cada población forma aleatoria en K subconjuntos disjuntos de aproximadamente el mismo tamaño con el ℓ -ésimo subconjunto de la población i teniendo tamaño $n_{\ell,i} \geq 2$, $\sum_{\ell=1}^K n_{\ell,i} = n_i$. Sea $\{\tilde{X}_{i,j}^{(\ell)}\}_{1 \leq j \leq n_{\ell,i}}$ los elementos del ℓ -ésimo subconjunto, y $\{\tilde{X}_{i,j}^{(-\ell)}\}_{1 \leq j \leq n-n_{\ell,i}}$ denotará los elementos en el complemento del ℓ -ésimo subconjunto. El conjunto $\{\tilde{X}_{i,j}^{(-\ell)}\}_{1 \leq j \leq n-n_{\ell,i}}$ será el conjunto de entrenamiento y $\{\tilde{X}_{i,j}^{(\ell)}\}_{1 \leq j \leq n_{\ell,i}}$ el de validación.
3. Recorremos los valores ρ posibles en el rango en consideración. Para cada uno de ellos, calculamos el estimador de la primer dirección utilizando los elementos del conjunto de entrenamiento, llamamos $\hat{\phi}_{1,\rho}^{(-\ell)}$ al estimador robusto de la primer componente calculada sin los subconjuntos ℓ -ésimos $\{X_{i,j}^{(\ell)}\}_{1 \leq j \leq n_{\ell,i}}$.
4. Definimos $X_{i,j}^{(\ell)\perp}(\rho)$ como antes, usando el conjunto de validación. Es decir, $X_{i,j}^{(\ell)\perp}(\rho) = \tilde{X}_{i,j}^{(\ell)} - \pi_{\tilde{\mathcal{L}}_\ell^{(-\ell)}}(\tilde{X}_{i,j}^{(\ell)})$, $1 \leq j \leq n_{\ell,i}$, donde $\tilde{\mathcal{L}}_\ell^{(-\ell)}$ es el subespacio lineal generado por $\hat{\phi}_{1,\rho}^{(-\ell)}$.
5. Dado un estimador de escala alrededor del cero σ_n , el método robusto de validación cruzada con K -fold elige el parámetro de suavizado que minimiza

$$RCV_{\ell, KCV}(\tau_n) = \sum_{\ell=1}^K \sum_{i=1}^k \sigma_n^2 (\|X_{i,1}^{(\ell)\perp}(\tau_n)\|, \dots, \|X_{i,n_{i,\ell}}^{(\ell)\perp}(\tau_n)\|).$$

Para las otras componentes, por ejemplo la m -ésima, lo que hacemos es utilizar los parámetros de penalización $\rho_1, \dots, \rho_{m-1}$ ya previamente obtenidos y sus componentes $\hat{\phi}_1, \dots, \hat{\phi}_{m-1}$. No se vuelven a calcular estas componentes, proyectamos los datos en el subestacio ortogonal generados por las mismas y repetimos el procedimiento de antes, pero ahora haciendo en el paso 4 la proyección sobre la nueva estimación $\hat{\phi}_{m,\rho}$.

Los resultados obtenidos sobre 100 replicaciones se dan en la Tabla 5.14. Como en los casos anteriores, no se observa una gran mejoría respecto de los estimadores sin suavizar y esto puede atribuirse a que las trayectorias son suaves.

Modelo	Escala	$\widehat{\phi}_{\text{RAW},j}$			$\widehat{\phi}_{\text{PS},j}$		
		$j = 1$	$j = 2$	$j = 3$	$j = 1$	$j = 2$	$j = 3$
C_0	SD	0.0033	0.0039	0.0037	0.0031	0.0040	0.0038
	MAD	0.0436	0.0634	0.0368	0.0524	0.0667	0.0339
	M -ESCALA	0.0108	0.0141	0.0077	0.0122	0.0152	0.0074
$C_{2,0.2}$	SD	1.4496	1.4496	0.0023	1.5410	1.5410	0.0020
	MAD	0.3780	0.3903	0.0271	0.3880	0.3972	0.0266
	M -ESCALA	0.4347	0.4354	0.0058	0.4591	0.4604	0.0062
$C_{3,a,0.2}$	SD	1.8275	1.9238	1.9405	1.8253	1.9180	1.9367
	MAD	0.2598	0.7858	0.8071	0.2742	0.7945	0.8090
	M -ESCALA	0.2997	1.0419	1.0977	0.2997	1.1054	1.1594
$C_{3,b,0.2}$	SD	0.0140	1.7596	1.7905	0.0145	1.7478	1.7745
	MAD	0.0874	0.5092	0.5013	0.0855	0.5329	0.5262
	M -ESCALA	0.0427	0.4782	0.4966	0.0384	0.4378	0.4631
$C_{23,0.2}$	SD	1.8303	0.1930	1.8255	1.8196	0.1684	1.8113
	MAD	0.9350	1.0709	0.6005	0.9081	1.0608	0.5985
	M -ESCALA	1.0599	1.2001	0.7009	1.0753	1.1905	0.6969

Tabla 5.14: Valores medios de $\|\widehat{\phi}_j/\|\widehat{\phi}_j\| - \phi_j\|^2$, cuando $f = \text{id}$ y el parámetro de suavizado es elegido mediante K -fold.

5.6.2 Modelo CPC de rango infinito

Descripción del modelo

Tenemos $N = \sum_{i=1}^k n_i$ observaciones provenientes de $k = 3$ poblaciones, con $n_i = 100$, $1 \leq i \leq 3$.

Población 1

La primera de ellas, $X_{1,\ell} \sim P_1$, consiste en un movimiento Browniano en el intervalo $[0, 1]$. Es decir, $X_{1,\ell}$ es un proceso Gaussiano con núcleo de covarianza $\gamma_1(s, t) = 10 \min(s, t)$. Las autofunciones (y por ende las direcciones principales) están dadas por

$$\phi_n(t) = \sqrt{2} \sin\left(\frac{1}{2}(2n-1)\pi t\right).$$

Mientras que, los autovalores (y por ende los valores principales) son

$$\lambda_{1,n} = 10 \left(\frac{2}{(2n-1)\pi}\right)^2.$$

Población 2

La segunda población es también un proceso Gaussiano pero con núcleo de covarianza dado por $\gamma_2(s, t) = 20 \min(s, t)$. Las autofunciones son las mismas que las de la población 1 y los autovalores son el doble de los anteriores, es decir,

$$\lambda_{2,n} = 20 \left(\frac{2}{(2n-1)\pi}\right)^2.$$

Población 3

Los datos se generan como

$$X_{3,\ell} = Z_{1,\ell}\phi_1 + Z_{2,\ell}\phi_2 + Z_{3,\ell}\phi_3 ,$$

en donde $Z_{k,\ell} \sim N(0, \sigma_k^2)$, con $\sigma_1 = 3$, $\sigma_2 = 1$ y $\sigma_3 = 1/2$. Luego, las autofunciones serán iguales a las anteriores y los autovalores son $\lambda_{3,1} = \sigma_1^2 = 9$, $\lambda_{3,2} = \sigma_2^2 = 1$ y $\lambda_{3,3} = \sigma_3^2 = 1/4$ y $\lambda_{3,j} = 0$ para $j \geq 4$.

Es claro que estas tres poblaciones conjuntamente respetan el modelo CPC. Los primeros cuatro autovalores de cada población figuran en la Tabla 5.15.

$\lambda_{i,j}$	$i = 1$	$i = 2$	$i=3$
$j = 1$	4.053	8.106	9.000
$j = 2$	0.450	0.901	1.000
$j = 3$	0.162	0.324	0.250
$j = 4$	0.083	0.165	0.000

Tabla 5.15: Resumen de autovalores

Nos interesa para el desvío estándar como afecta la contaminación y la penalización el funcional, es decir, tomamos $\sigma_R^2(P_i[\alpha]) = \text{var}(P_i[\alpha])$ y $\tau_i = 1/3$. Si $f : [0, \infty] \rightarrow \mathbb{R}$ una función creciente. La función objetivo es

$$\varsigma(\alpha) = \sum_{i=1}^3 (1/3)f(\text{var}(P_i[\alpha])).$$

de donde

$$\begin{aligned} \varsigma(\phi_1) &= \frac{1}{3} (f(4.053) + f(8.106) + f(9)) \\ \varsigma(\phi_2) &= \frac{1}{3} (f(0.450) + f(0.901) + f(1)) \\ \varsigma(\phi_3) &= \frac{1}{3} (f(0.162) + f(0.324) + f(0.250)) . \end{aligned}$$

Por lo tanto, si $\Psi(\alpha) = \int_0^1 (\alpha''(s))^2 ds$ y $\log(\exp(1)) = 1$ tenemos así los resultados de la Tabla 5.16.

$\varsigma(\phi_j)$	$f = \text{id}$	$f = \log$	$\Psi(\phi_j)$
$j = 1$	7.053	1.896	6.089
$j = 2$	0.784	-0.301	493.183
$j = 3$	0.245	-1.444	3805.423

Tabla 5.16: Resumen de la función objetivo con suavizado.

Como en la sección 5.6.1 sea $\varsigma_\rho(\alpha) = \sum_{i=1}^3 (1/3)f(\text{var}(P_i[\alpha])) - \rho\Psi(\alpha)$. Para preservar el orden debemos elegir ρ de forma tal que $\varsigma_\rho(\phi_1) > \varsigma_\rho(\phi_2) > \varsigma_\rho(\phi_3)$, desigualdades que valen pues las funciones son menos suaves a medida que los autovalores crecen.

Descripción del modelo de contaminación

Cada una de las tres poblaciones P_1, P_2 y P_3 será contaminada como se detalla a continuación. Si $X_i \sim P_i$, las observaciones contaminadas X_i^c serán $X_i^c = (1 - V)X_i + VW_i$, donde $V \sim Bi(1, \epsilon)$ y $W_i \sim N(\mu_i, \sigma_c)\phi_4$ independientes de X_i , o sea es una contaminación en la cuarta componente. Sea P_i^c la medida de cada proceso contaminado. Buscamos μ_i de forma tal que en cada población la función objetivo en ϕ_4 supere a lo obtenido en ϕ_1 . Necesitamos así que $\text{var}(P_i^c[\phi_4]) > \text{var}(P_i^c[\phi_1])$ para cada población.

Tenemos que $X_i = \sum_{j \geq 1} Z_{i,j}\phi_j$ con $Z_{i,j} \sim N(0, \lambda_{i,j})$. En el caso del modelo contaminado, se verifica que $X_i^c = (1 - V)\sum_{j \geq 1} Z_{i,j}\phi_j + VW_i$, de donde $\langle X_i^c, \alpha \rangle = (1 - V)\sum_{j \geq 1} Z_{i,j}\langle \alpha, \phi_j \rangle + V\langle \alpha, \phi_4 \rangle U_i$ con $U_i \sim N(\mu_i, \sigma_c)$. Por lo tanto, $\langle \alpha, \phi_4 \rangle \mathbb{E}(V)\mu_i = \langle \alpha, \phi_4 \rangle \epsilon \mu_i$, luego, $\mathbb{E}^2(\langle X_i^c, \alpha \rangle) = \epsilon^2 \mu_i^2$.

Es fácil ver que $\mathbb{E}(\langle X_i^c, \alpha \rangle^2) = (1 - \epsilon)\sum_{j \geq 1} \lambda_{i,j} \langle \alpha, \phi_j \rangle^2 + \epsilon \langle \alpha, \phi_4 \rangle^2 (\sigma_c^2 + \mu_i^2)$. En el caso particular de $\alpha = \phi_\ell$, con $\ell \neq 4$, nos queda $\mathbb{E}(\langle X_i^c, \phi_\ell \rangle^2) = (1 - \epsilon)\lambda_{i,\ell}$. Por otro lado, $\mathbb{E}^2(\langle X_i^c, \phi_\ell \rangle) = 0$. Así pues, si $\ell \neq 4$, tendremos que

$$\text{var}(P_i^c[\phi_\ell]) = \text{var}(\langle X_i^c, \phi_\ell \rangle) = (1 - \epsilon)\lambda_{i,\ell}.$$

Cuando $\epsilon = 0$, se tiene $\text{var}(\langle X_i^c, \phi_\ell \rangle) = \lambda_{i,\ell}$.

En cambio, si $\ell = 4$, $\mathbb{E}^2(\langle X_i^c, \alpha \rangle) = \epsilon^2 \mu_i^2$ y $\mathbb{E}(\langle X_i^c, \phi_\ell \rangle^2) = \mathbb{E}(((1 - V)Z_{i,4} + VU_i)^2) = (1 - \epsilon)\lambda_{i,4} + \epsilon(\mu_i^2 + \sigma_c^2)$ de donde

$$\text{var}(P_i^c[\phi_4]) = \text{var}(\langle X_i^c, \phi_4 \rangle) = (1 - \epsilon)\lambda_{i,4} + \epsilon(\mu_i^2 + \sigma_c^2) - \epsilon^2 \mu_i^2 = (1 - \epsilon)\lambda_{i,4} + \epsilon(1 - \epsilon)\mu_i^2 + \epsilon\sigma_c^2.$$

Fijemos $\epsilon = 0.2$ y $\sigma_c^2 = 0.01$, o sea que es una contaminación cuasi-puntual en $\mu_i\phi_4$. Lo que queremos es que $\text{var}(P_i^c[\phi_4]) > \text{var}(P_i^c[\phi_1])$, para cada población. La Tabla 5.17 muestra los valores obtenidos para $\mu_1 = 10$, $\mu_2 = 15$ y $\mu_3 = 20$.

i	μ_i	$\text{var}(P_i^c[\phi_1])$	$\text{var}(P_i^c[\phi_4])$
$i = 1$	10	3.242	16.068
$i = 2$	15	6.485	36.134
$i = 3$	20	7.200	64.002

Tabla 5.17: Valores de la varianza del modelo contaminado

Los valores de la función objetivo en cada uno de las funciones ϕ_1 a ϕ_4 se dan en la Tabla 5.18.

Por lo tanto si $\rho = 0$ y σ_R es el desvío estándar para los datos contaminados tendremos que $\varsigma(\phi_4) > \varsigma(\phi_1)$.

$\varsigma(\phi_j)$	$f = \text{id}$	$f = \log$	$\Psi(\phi_j)$
$j = 1$	5.642	1.673	6.089
$j = 2$	0.627	-0.524	493.183
$j = 3$	0.196	-1.668	3805.423
$j = 4$	38.735	3.508	14618.91

Tabla 5.18: Tabla de valores de la función objetivo

Ahora incorporamos la penalización por suavizado, debemos garantizar que la penalización no sea tan alta de forma que el valor de la función objetivo en ϕ_1 supere al de ϕ_4 .

Debemos pedir que $\varsigma_\rho(\phi_4) > \varsigma_\rho(\phi_1)$. En el caso de la función identidad, obtendremos que $\rho < 0.0023$ mientras que si $f = \log$ obtenemos $\rho < 0.00013$.

Resultados

Se realizaron simulaciones de 1000 replicaciones con el modelo sin contaminar y con el modelo contaminado C_4 para valores de $\epsilon = 0.1$ y $\epsilon = 0.2$. Se consideraron los estimadores definidos en 5.4: sin penalizar, penalizando la escala y penalizando la norma, con $\rho = 10^{-7}$, con $f = \text{id}$ y $f = \log$. Como antes, indicamos a los estimadores correspondientes a cada tipo de escala como SD, MAD y M -ESCALA. En todas las tablas, el estimador de la j -ésima dirección principal vinculado a cada método están indicados como $\hat{\phi}_{\text{RAW},j}$, $\hat{\phi}_{\text{PS},j}$ y $\hat{\phi}_{\text{PN},j}$, respectivamente.

Se utilizó el mismo algoritmo que en la Sección 5.6.1 y que se describe con detalle en el Capítulo 6.

Las Tablas 5.19, 5.20 y 5.21 muestran los resultados obtenidos.

Se confirma el hecho que los estimadores robustos (MAD y M -ESCALA) son más resistentes que el estimador clásico basado en el desvío estándar frente a la presencia de contaminación. Cabe notar que los estimadores robustos son sensibles a un 20% de contaminación sobre todo al tratar de estimar las componentes más altas. Evidentemente, en esos puntos estamos tendiendo a acercarnos al punto de ruptura del estimador debido a la cercanía de los autovalores. Una aproximación al punto de ruptura en el caso finito-dimensional fue dada por Boente y Orellana (2001). Sin embargo, en el caso de datos funcionales dicho enfoque del problema resulta más complejo y será objeto de estudio posterior. Sin embargo, es bien sabido que la sensibilidad de los procedimientos está relacionada con el tamaño relativo de los autovalores y, en el caso de varias poblaciones, cuando $f = \text{id}$ con el cociente entre los autovalores del operador de covarianza pesado $\sum_{i=1}^3 \tau_i \Gamma_i$.

Se desprende de observar las tablas que la función f de *score* juega un papel relevante. Se puede observar que, en general, la performance del estimador es superior cuando se utiliza como *score* la función logaritmo. Se observa una pequeña mejoría al introducir penalización por suavizado, aunque es un poco menos pronunciada que con el cambio de la función de *score*. Como la tercera población tiene trayectorias suaves, es razonable entonces que la introducción de una penalización por suavizado no mejore demasiado la performance del estimador. En el Capítulo 6, se discute con más detalle este comportamiento.

Modelo	Escala	$\phi_{\text{RAW},1}$		$\phi_{\text{PS},1} (\rho = 10^{-7})$		$\phi_{\text{PN},1} (r = 10^{-7})$	
		$f = \text{id}$	$f = \text{log}$	$f = \text{id}$	$f = \text{log}$	$f = \text{id}$	$f = \text{log}$
C_0	SD	0.0037	0.0036	0.0037	0.0037	0.0037	0.0037
	MAD	0.0279	0.0279	0.0274	0.0268	0.0262	0.0268
	M -ESCALA	0.0061	0.0065	0.0061	0.0065	0.0061	0.0065
$C_{4,0.1}$	SD	1.9274	1.9050	1.9274	1.9058	1.9265	1.9058
	MAD	0.0870	0.0845	0.0844	0.0726	0.0743	0.0727
	M -ESCALA	0.0934	0.0897	0.0896	0.0778	0.0790	0.0781
$C_{4,0.2}$	SD	1.9475	1.9418	1.9475	1.9406	1.9469	1.9406
	MAD	0.1946	0.1864	0.1932	0.1780	0.1853	0.1782
	M -ESCALA	0.2108	0.2003	0.2096	0.1945	0.2031	0.1946

Tabla 5.19: Valores medios de $\|\widehat{\phi}_1/\|\widehat{\phi}_1\| - \phi_1\|^2$.

Modelo	Escala	$\phi_{\text{RAW},2}$		$\phi_{\text{PS},2} (\rho = 10^{-7})$		$\phi_{\text{PN},2} (r = 10^{-7})$	
		$f = \text{id}$	$f = \text{log}$	$f = \text{id}$	$f = \text{log}$	$f = \text{id}$	$f = \text{log}$
C_0	SD	0.0043	0.0043	0.0043	0.0045	0.0045	0.0045
	MAD	0.0585	0.0576	0.0551	0.0562	0.0550	0.0565
	M -ESCALA	0.0130	0.0130	0.0130	0.0131	0.0132	0.0131
$C_{4,0.1}$	SD	1.9298	1.9291	1.9284	1.9271	1.9252	1.9271
	MAD	0.2588	0.2360	0.2105	0.1763	0.2317	0.2154
	M -ESCALA	0.2635	0.2292	0.2121	0.1768	0.2394	0.2181
$C_{4,0.2}$	SD	1.9308	1.9295	1.9292	1.9264	1.9260	1.9264
	MAD	0.7966	0.7280	0.7551	0.6460	0.7761	0.7133
	M -ESCALA	0.9394	0.8511	0.8974	0.7649	0.9012	0.8259

Tabla 5.20: Valores medios de $\|\widehat{\phi}_2/\|\widehat{\phi}_2\| - \phi_2\|^2$.

Modelo	Escala	$\phi_{\text{RAW},3}$		$\phi_{\text{PS},3} (\rho = 10^{-7})$		$\phi_{\text{PN},3} (r = 10^{-7})$	
		$f = \text{id}$	$f = \text{log}$	$f = \text{id}$	$f = \text{log}$	$f = \text{id}$	$f = \text{log}$
C_0	SD	0.0036	0.0036	0.0035	0.0036	0.0036	0.0036
	MAD	0.0752	0.0672	0.0440	0.0460	0.0523	0.0520
	M -ESCALA	0.0109	0.0101	0.0103	0.0100	0.0105	0.0100
$C_{4,0.1}$	SD	1.9209	1.9251	1.9206	1.9239	1.9229	1.9242
	MAD	0.8667	0.8521	0.7169	0.6329	0.7594	0.7443
	M -ESCALA	1.1009	1.0773	0.9432	0.8297	0.9607	0.9474
$C_{4,0.2}$	SD	1.9174	1.9222	1.9164	1.9230	1.9181	1.9231
	MAD	1.5080	1.4755	1.4697	1.4103	1.4467	1.4362
	M -ESCALA	1.6430	1.6300	1.6229	1.5674	1.6009	1.5873

Tabla 5.21: Valores medios de $\|\widehat{\phi}_3/\|\widehat{\phi}_3\| - \phi_3\|^2$.

Capítulo 6

Aspectos Algorítmicos

6.1 Introducción

La definición de los estimadores tanto en su variante “cruda” como cuando se penaliza la escala requiere hallar el máximo en la esfera unitaria, que en el caso infinito dimensional resulta no ser compacto. Por lo tanto, esta maximización presenta desafíos numéricos relacionados con poder aproximar el máximo de la función objetivo de manera satisfactoria. Para el caso finito dimensional, Croux y Ruiz-Gazen (1996) propusieron un algoritmo para poder calcular los estimadores de *projection-pursuit*. Para ser precisos, Croux y Ruiz-Gazen (1996) propusieron el siguiente algoritmo.. Dado un conjunto de observaciones multivariadas X_1, \dots, X_n , $X_i \in \mathbb{R}^p$, para obtener una aproximación al estimador $\hat{\phi}_1 = \operatorname{argmax}_{\|\alpha\|=1} \sigma_{\mathbb{R}}^2(P_n[\alpha])$, Croux y Ruiz-Gazen (1996) sugieren

- 1) Calcular una posición μ (eventualmente robusta) de las observaciones. Sea $Y_i = X_i - \mu$.
- 2) Normalizar las observaciones $\alpha_i = Y_i/\|Y_i\|$ y considerar como conjunto de posibles direcciones $\mathcal{A}_n = \{\alpha_i = Y_i/\|Y_i\|, 1 \leq i \leq n\} \subset \mathcal{S}$.
- 3) Estimar $\phi_{\mathbb{R},1}$ mediante $\hat{\phi}_{\text{CR},1} = \operatorname{argmax}_i \sigma_{\mathbb{R}}^2(P_n[\alpha_i]) = \operatorname{argmax}_{\alpha \in \mathcal{A}_n} \sigma_{\mathbb{R}}^2(P_n[\alpha])$.
- 4) Proyectar las observaciones Y_i en el ortogonal de $\hat{\phi}_{\text{CR},1}$. Sea Y_i las observaciones proyectadas, repetir el procedimiento a partir de 2) con los nuevos Y_i para estimar la segunda y el resto de las componentes principales.

El objetivo de esta sección es describir la extensión al caso infinito dimensional del procedimiento anterior y su adaptación al caso de penalizaciones en la norma y en la escala y al de varias poblaciones. Por otra parte probamos, bajo condiciones generales, la efectividad del mismo. Es decir, que el estimador de la primer dirección principal resultante es consistente. Este resultado no ha sido probado hasta ahora ni siquiera en el caso finito-dimensional. Nos restringiremos al caso en que $\mu = 0$ y conocido y, por ende, no es necesario estimarlo.

6.2 Lemas Previos

Empezaremos probando algunos lemas auxiliares:

Lema 6.2.1. *Sea $X \in \mathcal{H}$ donde \mathcal{H} es un espacio de Hilbert separable \mathcal{H} . Supongamos que $X \sim P$ donde $P \sim \mathcal{E}(0, \mathbf{\Gamma})$ con $\mathbf{\Gamma}$ un operador compacto tal que $\mathbf{\Gamma} = \sum_{i \geq 1} \lambda_i \phi_i \otimes \phi_i$, siendo λ_i los autovalores ordenados de modo tal que $\lambda_1 \geq \lambda_2 \geq \dots$ y ϕ_i las autofunciones de $\mathbf{\Gamma}$, con $\sum_{i \geq 1} \lambda_i < \infty$. Supongamos que $\mathbb{P}(X = 0) = 0$. Luego, para cualquier $0 < \epsilon < 1$, se tiene que $p = \mathbb{P}(\langle X/\|X\|, \phi_1 \rangle^2 < 1 - \epsilon) < 1$.*

DEMOSTRACIÓN. Consideraremos dos situaciones

- a) $\mathbf{\Gamma}$ tiene infinitos autovalores positivos
- b) $\mathbf{\Gamma}$ tiene rango finito.

a) Si $\mathbf{\Gamma}$ no tiene rango finito, entonces $X = VY$ donde Y es un proceso Gaussiano con media 0 y operador de covarianza, $\mathbf{\Gamma}_Y$ proporcional a $\mathbf{\Gamma}$ y V es una variable aleatoria real independiente de Y . Además $\lambda_1 > 0$ pues $P(X = 0) = 0$. Por otra parte, como $\mathbb{P}(X = 0) = 0$ e Y es Gaussiano, $\mathbb{P}(V = 0) = 0$. Como $\mathbf{\Gamma}_Y = a\mathbf{\Gamma}$, tenemos por la descomposición de Karhunen-Loève que $Y = a \sum_{j \geq 1} \lambda_j^{1/2} Z_j \phi_j$ en L^2 donde $Z_j \sim N(0, 1)$ independientes. Más aún, $Y - a\lambda_1^{1/2} Z_1 \phi_1$ es independiente de Z_1 . Sin pérdida de generalidad supondremos $a = 1$. Por lo tanto,

$$\left\langle \frac{X}{\|X\|}, \phi_1 \right\rangle^2 = \left\langle \frac{VY}{|V|\|Y\|}, \phi_1 \right\rangle^2 = \frac{V^2}{|V|^2} \left\langle \frac{Y}{\|Y\|}, \phi_1 \right\rangle^2 = \left\langle \frac{Y}{\|Y\|}, \phi_1 \right\rangle^2 = \frac{\lambda_1 Z_1^2}{\|Y\|^2}.$$

Luego,

$$p = \mathbb{P}\left(\frac{\lambda_1 Z_1^2}{\|Y\|^2} < 1 - \epsilon\right) = \mathbb{P}(\lambda_1 Z_1^2 - (1 - \epsilon)\|Y\|^2 < 0) = \mathbb{P}\left(\epsilon \lambda_1 Z_1^2 - (1 - \epsilon)\|Y - \lambda_1^{1/2} Z_1 \phi_1\|^2 < 0\right).$$

Observemos que $Z_j^2 \sim \chi_1^2$ independientes, luego Z_1^2 y $U_2 = (1 - \epsilon)\|Y - \lambda_1^{1/2} Z_1 \phi_1\|^2$ son independientes con densidad soportada en $(0, \infty)$. Sea $g : \mathbb{R}_{>0} \rightarrow [0, 1]$ la función

$$g(u) = \mathbb{P}\left(Z_1^2 < \frac{u}{\epsilon \lambda_1}\right),$$

como $\lambda_1 > 0$, $0 < g(u) < 1$ para todo $u > 0$. Tenemos que $p = \mathbb{E}\{\mathbb{P}(\epsilon \lambda_1 Z_1^2 < U_2 | U_2)\}$ la independencia entre Z_1 y U_2 implican que $p = \mathbb{E}g(U_2)$, de donde $p < 1$.

Observemos que si X es Gaussiano, el resultado es válido aún cuando el rango de $\mathbf{\Gamma}$ sea finito e igual a q , es decir, si $Z_j = \langle X, \phi_j \rangle / \lambda_j^{1/2} \sim N(0, 1)$ independientes, entonces,

$$\mathbb{P}\left(\frac{\lambda_1 Z_1^2}{\sum_{j=1}^q \lambda_j Z_j^2} < 1 - \epsilon\right) < 1. \quad (6.1)$$

b) Notemos que como $\mathbb{P}(X = 0) = 0$, $\mathbf{\Gamma}$ tiene mayor o igual a 1. Supongamos que $\mathbf{\Gamma}$ tiene rango finito q , es decir, es como estar en el caso multivariado ya que si $\mathbf{Y} = (Y_1, \dots, Y_q)$ con $Y_j = \langle X, \phi_j \rangle$, se cumple que $\mathbf{Y} \sim \mathcal{E}_q(0, \mathbf{\Lambda})$ donde $\mathbf{\Lambda} = \text{diag}(\lambda_1, \dots, \lambda_q)$ y

$$\left\langle \frac{X}{\|X\|}, \phi_1 \right\rangle^2 = \frac{Y_1^2}{\sum_{j=1}^q Y_j^2}$$

pues $\mathbb{P}(X = \pi(X)) = 1$ donde π es la proyección ortogonal en el subespacio generado por ϕ_1, \dots, ϕ_q . Por otra parte, como $\mathbb{P}(X = 0) = 0$ entonces $\mathbb{P}(\mathbf{Y} = \mathbf{0}) = 0$. Sea $\mathbf{Z} = \mathbf{\Lambda}^{-1/2} \mathbf{Y}$, luego \mathbf{Z} tiene distribución esférica, o sea, $\mathbf{U} = \mathbf{Z}/\|\mathbf{Z}\|$ es uniforme en la esfera y $\mathbf{U} \sim \mathbf{V}/\|\mathbf{V}\|$ donde $\mathbf{V} \sim N(\mathbf{0}, \mathbf{I})$

$$\frac{Y_1^2}{\sum_{j=1}^q Y_j^2} = \frac{\lambda_1 Z_1^2}{\sum_{j=1}^q \lambda_j Z_j^2} = \frac{\lambda_1 U_1^2}{\sum_{j=1}^q \lambda_j U_j^2} \sim \frac{\lambda_1 V_1^2}{\sum_{j=1}^q \lambda_j V_j^2}$$

Luego, por (6.1) tenemos que

$$\mathbb{P} \left(\frac{\lambda_1 V_1^2}{\sum_{j=1}^q \lambda_j V_j^2} < 1 - \epsilon \right) < 1$$

de donde $p = \mathbb{P}(\langle X/\|X\|, \phi_1 \rangle^2 < 1 - \epsilon) < 1$. \square

Sea $\mathcal{H}_s \subset \mathcal{H}$ el subconjunto de elementos suaves de \mathcal{H} , o sea, $\alpha \in \mathcal{H}_s$ si $\Psi(\alpha) < \infty$.

Lema 6.2.2. Sea $X \in \mathcal{H}$ donde \mathcal{H} es un espacio de Hilbert separable \mathcal{H} . Supongamos que $X \sim P$ donde $P \sim \mathcal{E}(0, \mathbf{\Gamma})$ con $\mathbf{\Gamma}$ un operador compacto y definido semi-positivo tal que $\mathbf{\Gamma} = \sum_{i \geq 1} \lambda_i \phi_i \otimes \phi_i$, siendo λ_i los autovalores, ordenados de modo tal que $\lambda_1 \geq \lambda_2 \geq \dots$ y ϕ_i las autofunciones de $\mathbf{\Gamma}$, con $\sum_{i \geq 1} \lambda_i < \infty$. Supongamos que $\mathbb{P}(X = 0) = 0$, $\mathbb{P}(X \in \mathcal{H}_s) = 1$. Sea r_n una sucesión numérica tal que $r = r_n \rightarrow 0$. Dado $0 < \epsilon < 1$, sea $p_n = \mathbb{P}(\langle X_1/\|X_1\|_r, \phi_1 \rangle^2 < 1 - \epsilon)$. Luego, existe $n_0 \in \mathbb{N}$ y $q = q(\epsilon) < 1$ tal que $p_n \leq q < 1$ para todo $n \geq n_0$.

DEMOSTRACIÓN. Sea $p(\epsilon) = \mathbb{P}(\langle X/\|X\|, \phi_1 \rangle^2 < 1 - \epsilon)$. Por el Lema 6.2.1, $p(\epsilon) < 1$ para todo $0 < \epsilon < 1$. Sea $q = q(\epsilon) = (1 + p(\epsilon/2))/2 < 1$, veremos que $p_n \leq q$ para n suficientemente grande.

Usando que

$$\left\langle \frac{X_1}{\|X_1\|_r}, \phi_1 \right\rangle = \frac{\|X_1\|}{\|X_1\|_r} \left\langle \frac{X_1}{\|X_1\|}, \phi_1 \right\rangle$$

tenemos que

$$p_n = \mathbb{P} \left(\left\langle \frac{X_1}{\|X_1\|_r}, \phi_1 \right\rangle^2 < (1 - \epsilon) \frac{\|X_1\|_r^2}{\|X_1\|^2} \right).$$

La convergencia $r \rightarrow 0$ implica que $(1 - \epsilon) \|X_1\|_r^2 / \|X_1\|^2 \xrightarrow{a.s.} (1 - \epsilon)$. Por lo tanto, como $\|X_1\|_r^2 > \|X_1\|^2$

$$\lim_{n \rightarrow \infty} \mathbb{P} \left(1 - \epsilon < (1 - \epsilon) \frac{\|X_1\|_r^2}{\|X_1\|^2} < 1 - \epsilon/2 \right) = 1$$

de donde, si $\mathcal{B}_n = \{\omega \in \Omega : 1 - \epsilon < (1 - \epsilon) \|X_1(\omega)\|_r^2 / \|X_1(\omega)\|^2 < 1 - \epsilon/2\}$, se tiene que dado $\delta = \eta/2$ donde $\eta < 1 - p(\epsilon/2)$ existe n_0 tal que para todo $n \geq n_0$, $\mathbb{P}(\mathcal{B}_n) \geq 1 - \delta$.

Por lo tanto,

$$\begin{aligned}
p_n &= \mathbb{P} \left(\left\langle \frac{X_1}{\|X_1\|}, \phi_1 \right\rangle^2 < (1 - \epsilon) \frac{\|X_1\|_r^2}{\|X_1\|^2} \cap \mathcal{B}_n \right) + \mathbb{P} \left(\left\langle \frac{X_1}{\|X_1\|}, \phi_1 \right\rangle^2 < (1 - \epsilon) \frac{\|X_1\|_r^2}{\|X_1\|^2} \cap \mathcal{B}_n^c \right) \\
&\leq \mathbb{P} \left(\left\langle \frac{\|X_1\|}{\|X_1\|}, \phi_1 \right\rangle^2 < 1 - \epsilon/2 \right) + \mathbb{P}(\mathcal{B}_n^c) = p(\epsilon/2) + \delta < \frac{1 + p(\epsilon/2)}{2} = q < 1.
\end{aligned}$$

6.3 Caso de una población

En esta sección estudiaremos la consistencia del estimador de la primer dirección principal obtenido mediante una generalización del algoritmo de Croux y Ruiz-Gazen (1996) para cada una de las tres situaciones que hemos considerado en el Capítulo 3: estimadores sin penalizar, estimadores obtenidos penalizando la norma y estimadores obtenidos penalizando la escala. Consideremos $X \sim P$, con posición igual a 0 conocida. Dado σ_R un funcional robusto de escala univariado, el objetivo es estimar $\phi_{R,1} = \phi_{R,1}(P) = \operatorname{argmax}_{\|\alpha\|=1} \sigma_R^2(P[\alpha])$, en base a una muestra aleatoria X_1, \dots, X_n .

6.3.1 Caso del estimador sin penalizar

Como hemos dicho, nuestros estimadores sin penalizar están definidos por (3.3) lo que hace difícil su cálculo efectivo. Por esta razón, proponemos la siguiente generalización del algoritmo de Croux y Ruiz-Gazen (1996) al caso funcional, que hemos considerado en nuestra simulación y consiste en

1. Normalizar las observaciones $\alpha_i = X_i/\|X_i\|$ y considerar el conjunto de posibles direcciones $\mathcal{A}_n = \{\alpha_i = X_i/\|X_i\|, 1 \leq i \leq n\} \subset \mathcal{S}$.
2. Estimar $\phi_{R,1}$ mediante $\hat{\phi}_{CR,1} = \operatorname{argmax}_i s_n^2(\alpha_i) = \operatorname{argmax}_{\alpha \in \mathcal{A}_n} s_n^2(\alpha)$ con $s_n(\alpha) = \sigma_R(P_n[\alpha])$.

Como en el Capítulo 3, consideraremos el supuesto **S2** y una versión relajada del supuesto **S0** más una condición de elipticidad para obtener la consistencia de $\hat{\phi}_{CR,1}$.

Teorema 6.3.1.1. *Sea $X \in \mathcal{H}$ donde \mathcal{H} es un espacio de Hilbert separable \mathcal{H} . Supongamos que $X \sim P$ donde $P \sim \mathcal{E}(0, \Gamma)$ donde Γ un operador compacto tal que $\Gamma = \sum_{i \geq 1} \lambda_i \phi_i \otimes \phi_i$, siendo λ_i los autovalores, ordenados de modo tal que $\lambda_1 \geq \lambda_2 \geq \dots$ y ϕ_i las autofunciones de Γ , con $\sum_{i \geq 1} \lambda_i < \infty$. Supongamos que $\mathbb{P}(X = 0) = 0$. Sea X_1, \dots, X_n una muestra aleatoria con distribución P y supongamos que $\phi_{R,1}(P)$ es único salvo cambio de signo y que vale **S2**. Luego, $\langle \hat{\phi}_{CR,1}, \phi_{R,1} \rangle^2 \xrightarrow{a.s.} 1$, es decir, $\hat{\phi}_{CR,1} \xrightarrow{a.s.} \phi_{R,1}$ salvo eventual cambio de signo.*

DEMOSTRACIÓN. Como X es elíptico, se cumple que existe c tal que $\sigma^2(\alpha) := \sigma_R^2(P[\alpha]) = c < \alpha, \Gamma \alpha >$ y $\phi_{R,1}(P) = \phi_1$. Como estamos bajo el supuesto de unicidad de $\phi_{R,1}$ salvo signo, entonces $c > 0$ y además $\lambda_1 > \lambda_2$. Sin pérdida de generalidad, supondremos que $c = 1$.

Dados X_1, \dots, X_n definimos $\mathcal{A}_n = \{X_i/\|X_i\|, 1 \leq i \leq n\}$ y sea

$$\tilde{\phi}_1 = \tilde{\phi}_{1,n} = \operatorname{argmin}_{\alpha \in \mathcal{A}_n} 1 - \langle \alpha, \phi_1 \rangle^2,$$

es decir, $\langle \tilde{\phi}_1, \phi_1 \rangle^2 = \max_{1 \leq i \leq n} \langle X_i / \|X_i\|, \phi_1 \rangle^2$. Veamos que $\langle \tilde{\phi}_1, \phi_1 \rangle^2 \xrightarrow{a.s.} 1$, o sea, que salvo por elección del signo $\tilde{\phi}_1 \xrightarrow{a.s.} \phi_1$. Para ello como $\langle \tilde{\phi}_1, \phi_1 \rangle^2 \leq 1$, debemos probar que, para todo $\epsilon > 0$, $\lim_{N \rightarrow \infty} \mathbb{P} \left(\cup_{n \geq N} (1 - \langle \tilde{\phi}_1, \phi_1 \rangle^2) > \epsilon \right) = 0$ es decir, basta ver que $\sum_{n \geq 1} \mathbb{P}(1 - \langle \tilde{\phi}_1, \phi_1 \rangle^2) > \epsilon) < \infty$. Observemos que por ser X_1, \dots, X_n independientes e idénticamente distribuidas

$$\begin{aligned} \mathbb{P}(1 - \langle \tilde{\phi}_1, \phi_1 \rangle^2) > \epsilon) &= \mathbb{P} \left(\max_{1 \leq i \leq n} \langle \frac{X_i}{\|X_i\|}, \phi_1 \rangle^2 < 1 - \epsilon \right) \\ &= \prod_{i=1}^n \mathbb{P} \left(\langle \frac{X_i}{\|X_i\|}, \phi_1 \rangle^2 < 1 - \epsilon \right) = \mathbb{P} \left(\langle \frac{X_1}{\|X_1\|}, \phi_1 \rangle^2 < 1 - \epsilon \right)^n. \end{aligned}$$

Por el Lema 6.2.1, tenemos que $p = \mathbb{P}(\langle X/\|X\|, \phi_1 \rangle^2 < 1 - \epsilon) < 1$ y por lo tanto, $\sum_{n \geq 1} \mathbb{P}(1 - \langle \tilde{\phi}_1, \phi_1 \rangle^2) > \epsilon) = \sum_{n \geq 1} p^n < \infty$. En definitiva, hemos demostrado que $\langle \tilde{\phi}_1, \phi_1 \rangle^2 \xrightarrow{a.s.} 1$.

El hecho que, $\sigma^2(\alpha) = \langle \alpha, \Gamma \alpha \rangle$ implica que $\sigma : \mathcal{H} \rightarrow \mathbb{R}$ es débilmente continuo, es decir, vale **S1**. Por lo tanto, $\sigma^2(\tilde{\phi}_1) - \sigma^2(\phi_1) \xrightarrow{a.s.} 0$. Además, **S2** implica que $s_n^2(\tilde{\phi}_1) - \sigma^2(\tilde{\phi}_1) \xrightarrow{a.s.} 0$, es decir, $\hat{b}_{n,1} = s_n^2(\tilde{\phi}_1) - \sigma^2(\tilde{\phi}_1) \xrightarrow{a.s.} 0$. Usando nuevamente **S2** tendremos que

$$\hat{a}_{n,1} = s_n^2(\hat{\phi}_{\text{CR},1}) - \sigma^2(\hat{\phi}_{\text{CR},1}) \xrightarrow{a.s.} 0.$$

Por otra parte, como $\hat{\phi}_{\text{CR},1}$ se obtiene maximizando $s_n^2(\alpha)$ sobre \mathcal{A}_n y $\tilde{\phi}_1 \in \mathcal{A}_n$, tenemos $s_n^2(\tilde{\phi}_1) \leq s_n^2(\hat{\phi}_{\text{CR},1})$. Además, como $\phi_1 = \operatorname{argmax}_{\alpha \in \mathcal{S}} \sigma^2(\alpha)$, se tiene que $\sigma^2(\hat{\phi}_{\text{CR},1}) \leq \sigma^2(\phi_1)$. De donde,

$$\sigma^2(\phi_1) = s_n^2(\tilde{\phi}_1) - \hat{b}_{n,1} \leq s_n^2(\hat{\phi}_{\text{CR},1}) - \hat{b}_{n,1} = \sigma^2(\hat{\phi}_{\text{CR},1}) + \hat{a}_{n,1} - \hat{b}_{n,1} \leq \sigma^2(\phi_1) + \hat{a}_{n,1} - \hat{b}_{n,1}.$$

Por lo tanto, $\sigma^2(\phi_1) - \hat{a}_{n,1} + \hat{b}_{n,1} \leq \sigma^2(\hat{\phi}_{\text{CR},1}) \leq \sigma^2(\phi_1)$, de donde $\sigma^2(\hat{\phi}_{\text{CR},1}) \xrightarrow{a.s.} \sigma^2(\phi_1)$. Usando el Lema 3.4.1, obtenemos que $\langle \hat{\phi}_{\text{CR},1}, \phi_1 \rangle^2 \xrightarrow{a.s.} 1$. \square

6.3.2 Caso de penalización en la norma

Cuando penalizamos en la norma para estimar $\phi_{\text{R},1}$ los estimadores propuestos estaban dados por (3.5). Para poder efectuar el cómputo cuando las observaciones X_1, \dots, X_n son tales que $\mathbb{P}(X \in \mathcal{H}_s) = 1$, una generalización del algoritmo descrito en 6.3.1 consiste en

1. Normalizar las observaciones $\alpha_i = X_i / \|X_i\|_r$ y considerar el conjunto de posibles direcciones $\mathcal{A}_{n,r} = \{\alpha_i = X_i / \|X_i\|_r, 1 \leq i \leq n\} \subset \mathcal{S}$
2. Estimar $\phi_{\text{R},1}$ mediante $\hat{\phi}_{\text{CRN},1} = \operatorname{argmax}_{1 \leq i \leq n} s_n^2(\alpha_i) = \operatorname{argmax}_{\alpha \in \mathcal{A}_{n,r}} s_n^2(\alpha)$.

El siguiente resultado muestra la consistencia de $\hat{\phi}_{\text{CRN},1}$

Teorema 6.3.1.2. *Sea $X \in \mathcal{H}$ donde \mathcal{H} es un espacio de Hilbert separable \mathcal{H} tal que $\mathbb{P}(X \in \mathcal{H}_s) = 1$. Supongamos que $X \sim P$ donde $P \sim \mathcal{E}(0, \Gamma)$ con Γ un operador compacto tal que $\Gamma = \sum_{i \geq 1} \lambda_i \phi_i \otimes \phi_i$, siendo λ_i los autovalores, ordenados de modo tal que $\lambda_1 \geq \lambda_2 \geq \dots$ y ϕ_i*

las autofunciones de Γ , con $\sum_{i \geq 1} \lambda_i < \infty$. Supongamos que $\mathbb{P}(X = 0) = 0$, $\mathbb{P}(X \in \mathcal{H}_S) = 1$. Sea r_n una sucesión numérica tal que $r = r_n \rightarrow 0$. Sea X_1, \dots, X_n una muestra aleatoria con distribución P y supongamos que $\phi_{R,1}(P)$ es único salvo cambio de signo y que se cumple **S2**. Luego, $\langle \widehat{\phi}_{\text{CRN},1}, \phi_{R,1} \rangle^2 \xrightarrow{a.s.} 1$, es decir, $\widehat{\phi}_{\text{CRN},1} \xrightarrow{a.s.} \phi_{R,1}$ salvo eventual cambio de signo.

DEMOSTRACIÓN. Vamos a marcar las diferencias con la demostración del Teorema 6.3.1.1. Definamos

$$\widetilde{\phi}_1 = \widetilde{\phi}_{1,n} = \underset{\alpha \in \mathcal{A}_{n,r}}{\operatorname{argmin}} 1 - \langle \alpha, \phi_1 \rangle^2,$$

Observemos primero que $\|\widehat{\phi}_{\text{CRN},1}\| \leq 1$, $\|\widetilde{\phi}_1\| \leq 1$ y

$$\left| \left\langle \frac{X_i}{\|X_i\|_r}, \phi_1 \right\rangle \right| \leq \frac{\|X_i\|}{\|X_i\|_r} \leq 1.$$

Como en el Teorema 6.3.1.1, $\phi_{R,1}(P) = \phi_1$ y nos basta ver que $\langle \widetilde{\phi}_1, \phi_1 \rangle^2 \xrightarrow{a.s.} 1$, que como vimos se reduce a probar que

$$\sum_{n=1}^{\infty} \mathbb{P}(1 - \langle \widetilde{\phi}_1, \phi_1 \rangle^2 > \epsilon) < \infty.$$

Como en el Teorema 6.3.1.1, obtenemos que, por ser r_n no aleatorio y X_1, \dots, X_n independientes,

$$\mathbb{P}(1 - \langle \widetilde{\phi}_1, \phi_1 \rangle^2 > \epsilon) = \mathbb{P} \left(\left\langle \frac{X}{\|X\|_r}, \phi_1 \right\rangle^2 < 1 - \epsilon \right)^n,$$

de donde para ver que $\sum_{n=1}^{\infty} p_n^n < \infty$ donde $p_n = \mathbb{P} \left(\left\langle \frac{X}{\|X\|_r}, \phi_1 \right\rangle^2 < 1 - \epsilon \right)$. Por el Lema 6.2.2 existe n_0 tal que $p_n \leq q < 1$ para todo $n \geq n_0$, luego

$$\sum_{n=1}^{\infty} \mathbb{P}(1 - \langle \widetilde{\phi}_1, \phi_1 \rangle^2 > \epsilon) \leq \sum_{n=1}^{n_0-1} \mathbb{P}(1 - \langle \widetilde{\phi}_1, \phi_1 \rangle^2 > \epsilon) + \sum_{n=n_0}^{\infty} q^n < \infty.$$

La demostración sigue como en el Teorema 6.3.1.1 usando que $\phi_1 = \operatorname{argmax}_{\alpha \in \mathcal{S}} \sigma^2(\alpha)$, se tiene que $\sigma^2(\widehat{\phi}_{\text{CRN},1} / \|\widehat{\phi}_{\text{CRN},1}\|) \leq \sigma^2(\phi_1)$, es decir, $\sigma^2(\widehat{\phi}_{\text{CRN},1}) \leq \|\widehat{\phi}_{\text{CRN},1}\|^2 \sigma^2(\phi_1) \leq \sigma^2(\phi_1)$. \square

6.3.3 Caso de penalización en la escala

Cuando penalizábamos en la escala para estimar $\phi_{R,1}$ los estimadores resolvían (3.6). Para efectuar una aproximación a los mismos y efectuar un cálculo efectivo proponemos el siguiente algoritmo cuando X_1, \dots, X_n son tales que $P(X \in \mathcal{H}_S) = 1$.

1. Normalizar las observaciones $\alpha_i = X_i / \|X_i\|$ y considerar el conjunto de posibles direcciones $\mathcal{A}_n = \{\alpha_i = X_i / \|X_i\|, 1 \leq i \leq n\} \subset \mathcal{S}$
2. Estimar $\phi_{R,1}$ mediante $\widehat{\phi}_{\text{CRS},1} = \operatorname{argmax}_{\alpha \in \mathcal{A}_n} \sigma_R^2(P_n[\alpha]) - \rho\Psi(\alpha)$.

Teorema 6.3.1.3. Sea $X \in \mathcal{H}$ donde \mathcal{H} es un espacio de Hilbert separable \mathcal{H} tal que $\mathbb{P}(X \in \mathcal{H}_s) = 1$. Supongamos que $X \sim P$ donde $P \sim \mathcal{E}(0, \Gamma)$ con Γ un operador compacto tal que $\Gamma = \sum_{i \geq 1} \lambda_i \phi_i \otimes \phi_i$, siendo λ_i los autovalores, ordenados de modo tal que $\lambda_1 \geq \lambda_2 \geq \dots$ y ϕ_i las autofunciones de Γ , con $\sum_{i \geq 1} \lambda_i < \infty$. Supongamos que $\mathbb{P}(X = 0) = 0$, $\mathbb{P}(X \in \mathcal{H}_s) = 1$. Sea X_1, \dots, X_n una muestra aleatoria con distribución P y supongamos que $\phi_{R,1}(P)$ es único salvo cambio de signo y que se cumple **S2**. Supongamos que $\rho = \rho_n$ es una sucesión numérica tal que $\rho \rightarrow 0$. Luego, $\langle \hat{\phi}_{\text{CRS},1}, \phi_{R,1} \rangle^2 \xrightarrow{a.s.} 1$, es decir, $\hat{\phi}_{\text{CRS},1} \xrightarrow{a.s.} \phi_{R,1}$ salvo eventual cambio de signo y $\rho \Psi(\hat{\phi}_{\text{CRS},1}) \xrightarrow{a.s.} 0$.

DEMOSTRACIÓN. Como en la demostración del Teorema 6.3.1.2, definamos $\mathcal{A}_{n,\rho} = \{\alpha_i = X_i / \|X_i\|_\rho, 1 \leq i \leq n\} \subset \mathcal{S}$

$$\tilde{\phi}_1 = \underset{1 \leq i \leq n}{\operatorname{argmin}} 1 - \langle \phi_1, \frac{\|X_i\|}{\|X_i\|_\rho} \rangle^2 = \underset{\alpha \in \mathcal{A}_{n,\rho}}{\operatorname{argmin}} 1 - \langle \phi_1, \alpha \rangle^2.$$

Entonces, se tiene que $\langle \tilde{\phi}_1, \phi_1 \rangle^2 \xrightarrow{a.s.} 1$, o sea, $\tilde{\phi}_1 \xrightarrow{a.s.} \phi_1$ salvo eventual signo, sin pérdida de generalidad supondremos que $\tilde{\phi}_1 \xrightarrow{a.s.} \phi_1$. En particular, como $\|X_i\| / \|X_i\|_\rho \leq 1$, tenemos que $\|\tilde{\phi}_1\| \leq 1$, de donde

$$\langle \tilde{\phi}_1, \phi_1 \rangle^2 \leq \|\tilde{\phi}_1\|^2 \|\phi_1\| = \|\tilde{\phi}_1\|^2 \leq 1$$

por lo tanto, $\|\tilde{\phi}_1\|^2 \xrightarrow{a.s.} 1$. Como $\|\tilde{\phi}_1\|_\rho = 1$, tendremos que $\rho \Psi(\tilde{\phi}_1) \xrightarrow{a.s.} 0$.

Como en el Teorema 3.4.1 tenemos que

$$\sup_{\|\alpha\| \leq 1} |s_n^2(\alpha) - \sigma^2(\alpha)| \xrightarrow{a.s.} 0,$$

de donde como $\|\tilde{\phi}_1\| \leq 1$, se cumple que $a_n = s_n^2(\tilde{\phi}_1) - \sigma^2(\tilde{\phi}_1) \xrightarrow{a.s.} 0$. Por otra parte, como $\tilde{\phi}_1 \xrightarrow{a.s.} \phi_1$, por la continuidad débil de σ tendremos que $b_n = \sigma^2(\tilde{\phi}_1) - \sigma^2(\phi_1) \xrightarrow{a.s.} 0$, de donde $c_n = s_n^2(\tilde{\phi}_1) - \sigma^2(\phi_1) \xrightarrow{a.s.} 0$. Por otra parte, como $\|\hat{\phi}_{\text{CRS},1}\| = 1$, $d_n = s_n^2(\hat{\phi}_{\text{CRS},1}) - \sigma^2(\hat{\phi}_{\text{CRS},1}) \xrightarrow{a.s.} 0$.

Ahora bien, como $\tilde{\phi}_1 / \|\tilde{\phi}_1\| \in \mathcal{A}_n$, tenemos las desigualdades

$$\begin{aligned} s_n^2(\hat{\phi}_{\text{CRS},1}) &\geq s_n^2(\hat{\phi}_{\text{CRS},1}) - \rho \Psi(\hat{\phi}_{\text{CRS},1}) = \max_{1 \leq i \leq n} \left\{ s_n^2 \left(\frac{X_i}{\|X_i\|} \right) - \rho \Psi \left(\frac{X_i}{\|X_i\|} \right) \right\} \\ &\geq s_n^2 \left(\frac{\tilde{\phi}_1}{\|\tilde{\phi}_1\|} \right) - \rho \Psi \left(\frac{\tilde{\phi}_1}{\|\tilde{\phi}_1\|} \right) = \frac{s_n^2(\tilde{\phi}_1) - \rho \Psi(\tilde{\phi}_1)}{\|\tilde{\phi}_1\|^2} \\ &\geq \frac{\sigma^2(\phi_1) + c_n - \rho \Psi(\tilde{\phi}_1)}{\|\tilde{\phi}_1\|^2}. \end{aligned}$$

Pero $\|\tilde{\phi}_1\|^2 = 1 + o_{a.s.}(1)$ y $\rho \Psi(\tilde{\phi}_1) = o_{a.s.}(1)$, con lo cual obtenemos que

$$s_n^2(\hat{\phi}_{\text{CRS},1}) \geq \frac{\sigma^2(\phi_1) + c_n - \rho \Psi(\tilde{\phi}_1)}{\|\tilde{\phi}_1\|^2} = \sigma^2(\phi_1) + o_{a.s.}(1).$$

Veamos la desigualdad inversa. Como $\|\widehat{\phi}_{\text{CRS},1}\| = 1$,

$$\sigma^2(\phi_1) \geq \sigma^2(\widehat{\phi}_{\text{CRS},1}) = s_n^2(\widehat{\phi}_{\text{CRS},1}) - d_n.$$

Resumiendo,

$$s_n^2(\widehat{\phi}_{\text{CRS},1}) \geq \sigma^2(\phi_1) + o_{\text{a.s.}}(1) \quad \sigma^2(\phi_1) \geq s_n^2(\widehat{\phi}_{\text{CRS},1}) + o_{\text{a.s.}}(1).$$

Es decir, $s_n^2(\widehat{\phi}_{\text{CRS},1}) \xrightarrow{\text{a.s.}} \sigma^2(\phi_1)$, de donde por **S2** $\sigma(\widehat{\phi}_{\text{CRS},1}) \xrightarrow{\text{a.s.}} \sigma(\phi_1)$. La demostración se concluye utilizando el Lema 3.4.1, es decir, $\phi_{\text{CRS},1} \xrightarrow{\text{a.s.}} \phi_1$ salvo un signo.

Por otra parte, como

$$s_n^2(\widehat{\phi}_{\text{CRS},1}) \geq s_n^2(\widehat{\phi}_{\text{CRS},1}) - \rho\Psi(\widehat{\phi}_{\text{CRS},1}) \geq \sigma^2(\phi_1) + o_{\text{a.s.}}(1),$$

tenemos que $\rho\Psi(\widehat{\phi}_{\text{CRS},1}) \xrightarrow{\text{a.s.}} 0$. \square

6.4 Caso de varias poblaciones bajo un modelo CPCF

Consideremos ahora k poblaciones independientes con medidas de probabilidad P_i tales que su parámetro de posición $\mu_i = 0$ y una muestra aleatoria de P_i , es decir $X_{i,1}, \dots, X_{i,n_i}$ independientes tales que $X_{i,j} \sim P_i$, $1 \leq j \leq n_i$ como en el Capítulo 5. Indicaremos por P_{i,n_i} la medida empírica asociada a P_i , es decir,

$$P_{i,n_i}(A) = \frac{1}{n_i} \sum_{j=1}^{n_i} I_A(X_{i,j}),$$

mientras que $P_{i,n_i}[\alpha]$ será la medida empírica de las variables aleatorias $\{\langle X_{i,1}, \alpha \rangle, \dots, \langle X_{i,n_i}, \alpha \rangle\}$.

Dado σ_{R} un funcional robusto de escala univariado, definamos $\sigma_i : \mathcal{H} \rightarrow [0, +\infty)$ como $\sigma_i(\alpha) = \sigma_{\text{R}}(P_i[\alpha])$, $\varsigma(\alpha) = \sum_{i=1}^k \tau_i f(\sigma_i^2(\alpha))$, donde $f : \mathbb{R}_{>0} \rightarrow \mathbb{R}_{\geq 0}$ es una función de *scores* creciente. Más aún, sea P la medida producto $P = P_1 \times \dots \times P_k$.

En el capítulo 5 hemos introducido estimadores de $\phi_{f,1}(P) = \operatorname{argmax}_{\|\alpha\|=1} \sum_{i=1}^k f(\sigma_{\text{R}}^2(P_i[\alpha]))$, en base a una muestra aleatoria $X_{i,j}$, $1 \leq i \leq k$, $1 \leq j \leq n_i$. Recordemos que definimos $s_{i,n_i}^2 : \mathcal{H} \rightarrow \mathbb{R}$ como $s_{i,n_i}^2(\alpha) = \sigma_{\text{R}}^2(P_{i,n_i}[\alpha])$ y $\widehat{\varsigma} = \widehat{\varsigma}_N : \mathcal{H} \rightarrow \mathbb{R}$, como $\widehat{\varsigma}(\alpha) = \sum_{i=1}^k \widehat{\tau}_i f(s_{i,n_i}^2(\alpha))$ donde $\widehat{\tau}_i = n_i/N$. Los estimadores sin suavizar se obtenían mediante (5.4) mientras que los que penalizaban la escala y la norma estaban definidos por (5.5) y (5.6), respectivamente. En esta sección, daremos una generalización del algoritmo de Croux y Ruiz-Gazen (1996) al caso de varias poblaciones que satisfacen un modelo CPCF de modo a aproximar a cada uno de los estimadores descritos en el Capítulo 5.

6.4.1 Caso de estimadores sin penalizar

Como en el caso de una población, el estimador $\widehat{\phi}_1 = \operatorname{argmax}_{\|\alpha\|=1} \widehat{\varsigma}(\alpha)$ no pueden calcularse efectivamente. Por esta razón consideremos el siguiente algoritmo

1. Normalizar las observaciones $\alpha_{i,j} = X_{i,j}/\|X_{i,j}\|$ y considerar el conjunto de posibles direcciones $\mathcal{A}_N = \{\alpha_{i,j} = X_{i,j}/\|X_{i,j}\|, 1 \leq j \leq n_i, 1 \leq i \leq k\} \subset \mathcal{S}$, donde $N = \sum_{i=1}^k n_i$.
2. Estimar $\phi_{f,1}(P)$ mediante

$$\hat{\phi}_{f,\text{CR},1} = \underset{1 \leq j \leq n_s, 1 \leq s \leq k}{\operatorname{argmax}} \sum_{i=1}^k \hat{\tau}_i f(\sigma_{\mathbb{R}}^2(P_{i,n_i}[\alpha_{s,j}])) = \operatorname{argmax}_{\alpha \in \mathcal{A}_N} \hat{\zeta}(\alpha).$$

El siguiente Teorema muestra la consistencia de $\hat{\phi}_{f,\text{CR},1}$ bajo la hipótesis **C3** del Capítulo 5 cuando $\phi_{f,1}(P)$ es único salvo signo.

Teorema 6.4.1. Sean $X_{i,1} \in \mathcal{H}$, $1 \leq i \leq k$ independientes, donde \mathcal{H} es un espacio de Hilbert separable \mathcal{H} . Supongamos que $X_{i,j} \sim P_i$ con $P_i \sim \mathcal{E}(0, \mathbf{\Gamma}_i)$ donde $\mathbf{\Gamma}_i$ un operador compacto tal que $\mathbf{\Gamma}_i = \sum_{\ell \geq 1} \lambda_{i,\ell} \phi_\ell \otimes \phi_\ell$, siendo $\lambda_{i,\ell}$ los autovalores, ordenados de modo tal que $\lambda_{i,1} \geq \lambda_{i,2} \geq \dots$ y ϕ_ℓ las autofunciones comunes de los $\mathbf{\Gamma}_i$, con $\sum_{\ell \geq 1} \lambda_{i,\ell} < \infty$. Supongamos que $\mathbb{P}(X_{i,1} = 0) = 0$ para $1 \leq i \leq k$. Sean $X_{i,1}, \dots, X_{i,n_i}$ muestras aleatorias independientes con distribución P_i y supongamos que se cumple **C3** y $\phi_{f,1}(P)$ es único salvo signo. Más aún, supongamos que $n_i = \hat{\tau}_i N$ donde $0 < \hat{\tau}_i < 1$ es una sucesión numérica tal que $\hat{\tau}_i \rightarrow \tau_i$ con $\sum_{i=1}^k \tau_i = 1$. Supongamos que $f : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}$ cumple **C2** o bien $f = \log$. Luego, $\langle \hat{\phi}_{f,\text{CR},1}, \phi_{f,1}(P) \rangle^2 \xrightarrow{\text{a.s.}} 1$, es decir, $\hat{\phi}_{f,\text{CR},1} \xrightarrow{\text{a.s.}} \phi_{f,1}(P)$ salvo eventual cambio de signo.

DEMOSTRACIÓN. Como $X_{i,j}$ es elíptico, se cumple que $\sigma_i^2(\alpha) = c_i < \alpha, \mathbf{\Gamma}_i \alpha >$ de donde $\phi_{f,1}(P) = \phi_1$. Por otra parte, como $\sigma_i^2(\alpha) = c_i < \alpha, \mathbf{\Gamma}_i \alpha >$ y $\mathbf{\Gamma}_i$ es un operador compacto, se cumple también **C1**. Finalmente, como en el Teorema 6.3.1.1, tenemos que $\mathbb{P}(X_{i,1} = 0) = 0$ implica que $\lambda_{i,1} > 0$, de donde $\lambda_{f,i,1}(P) = \sigma_i^2(\phi_1) = c_i \lambda_{i,1} > 0$ pues $\phi_{f,1}(P)$ es único salvo signo.

Dados $X_{i,j}$, $1 \leq i \leq k$, $1 \leq j \leq n_i$ definimos $\mathcal{A}_N = \{X_{i,j}/\|X_{i,j}\|, 1 \leq i \leq k, 1 \leq j \leq n_i\}$ y sea

$$\tilde{\phi}_1 = \tilde{\phi}_{1,N} = \underset{\alpha \in \mathcal{A}_N}{\operatorname{argmin}} 1 - \langle \alpha, \phi_1 \rangle^2,$$

es decir, $\langle \tilde{\phi}_1, \phi_1 \rangle^2 = \max_{1 \leq i \leq k, 1 \leq j \leq n_i} \langle X_{i,j}/\|X_{i,j}\|, \phi_1 \rangle^2$.

Veamos que $\langle \tilde{\phi}_1, \phi_1 \rangle^2 \xrightarrow{\text{a.s.}} 1$, o sea, que salvo por elección del signo, $\tilde{\phi}_1 \xrightarrow{\text{a.s.}} \phi_1$. Para ello como $\langle \tilde{\phi}_1, \phi_1 \rangle^2 \leq 1$, debemos probar que, para todo $\epsilon > 0$, $\lim_{M \rightarrow \infty} \mathbb{P}\left(\bigcup_{N \geq M} (1 - \langle \tilde{\phi}_1, \phi_1 \rangle^2) > \epsilon\right) = 0$ es decir, basta ver que $\sum_{N \geq 1} \mathbb{P}(1 - \langle \tilde{\phi}_{1,N}, \phi_1 \rangle^2) > \epsilon) < \infty$. Tenemos lo siguiente:

$$\begin{aligned} \mathbb{P}(1 - \langle \tilde{\phi}_1, \phi_1 \rangle^2) > \epsilon) &= \mathbb{P}\left(\max_{1 \leq i \leq k, 1 \leq j \leq n_i} \left\langle \frac{X_{i,j}}{\|X_{i,j}\|}, \phi_1 \right\rangle^2 < 1 - \epsilon\right) \\ &= \prod_{i=1}^k \prod_{j=1}^{n_i} \mathbb{P}\left(\left\langle \frac{X_{i,j}}{\|X_{i,j}\|}, \phi_1 \right\rangle^2 < 1 - \epsilon\right) = \prod_{i=1}^k \mathbb{P}\left(\left\langle \frac{X_{i,1}}{\|X_{i,1}\|}, \phi_1 \right\rangle^2 < 1 - \epsilon\right)^{n_i} \\ &= \left[\prod_{i=1}^k \mathbb{P}\left(\left\langle \frac{X_{i,1}}{\|X_{i,1}\|}, \phi_1 \right\rangle^2 < 1 - \epsilon\right)^{\hat{\tau}_i} \right]^N \end{aligned}$$

con $X_{i,1} \sim P_i$.

Sea $p_i = \mathbb{P}(\langle X_{i,1}/\|X_{i,1}\|, \phi_1 \rangle^2 < 1 - \epsilon)$. Entonces

$$\mathbb{P}(1 - \langle \tilde{\phi}_1, \phi_1 \rangle^2) > \epsilon = \left[\prod_{i=1}^k p_i^{\hat{\tau}_i} \right]^N = \hat{q}^N,$$

con lo cual bastará ver que existe $1 \leq i \leq k$ tal que $p_i < 1$. Por el Lemma 6.2.1, tenemos que $\mathbb{P}(X_{i,1} = 0) = 0$ implica que $p_i < 1$ para $1 \leq i \leq k$. Como $\hat{\tau}_i \rightarrow \tau_i$ tenemos que $\frac{1}{2}\tau_i < \hat{\tau}_i < 2\tau_i$ si $N \geq N_0$ de donde $\hat{q} = \prod_{i=1}^k p_i^{\hat{\tau}_i} < \prod_{i=1}^k p_i^{\tau_i/2} = q < 1$, con lo cual $\langle \tilde{\phi}_1, \phi_1 \rangle^2 \xrightarrow{a.s.} 1$. Sin pérdida de generalidad, supondremos que $\tilde{\phi}_1 \xrightarrow{a.s.} \phi_1$.

Como σ_i es continuo, $\sigma_i^2(\tilde{\phi}_1) - \sigma_i^2(\phi_1) \xrightarrow{a.s.} 0$. Además, **C3** implica que $s_{i,n_i}^2(\tilde{\phi}_1) - \sigma_i^2(\tilde{\phi}_1) \xrightarrow{a.s.} 0$, de donde

$$s_{i,n_i}^2(\tilde{\phi}_1) \xrightarrow{a.s.} \sigma_i^2(\phi_1) = \lambda_{f,i,1}(P) > 0. \quad (6.2)$$

Está claro que, si se cumple **C2**, entonces (6.2) implica que

$$\hat{\varsigma}(\tilde{\phi}_1) \xrightarrow{a.s.} \varsigma(\phi_1). \quad (6.3)$$

Más aún, como $\lambda_{f,i,1}(P) > 0$ y el logaritmo es una función continua en $(0, \infty)$, se tiene también que (6.3) vale cuando $f = \log$. Es decir, en ambos casos,

$$\hat{b}_N = \hat{\varsigma}(\tilde{\phi}_1) - \varsigma(\phi_1) \xrightarrow{a.s.} 0. \quad (6.4)$$

Por otra parte, como $\tilde{\phi}_1 \in \mathcal{A}_N$, $\hat{\phi}_{f,CR,1} = \operatorname{argmax}_{\alpha \in \mathcal{A}_N} \hat{\varsigma}(\alpha)$, $\phi_1 = \operatorname{argmax}_{\|\alpha\|=1} \varsigma(\alpha)$ y $\|\hat{\phi}_{f,CR,1}\| = 1$, se tienen las siguientes desigualdades

$$\hat{\varsigma}(\tilde{\phi}_1) \leq \hat{\varsigma}(\hat{\phi}_{f,CR,1}) \quad (6.5)$$

$$\varsigma(\tilde{\phi}_1) \leq \varsigma(\phi_1) \quad (6.6)$$

$$\varsigma(\hat{\phi}_{f,CR,1}) \leq \varsigma(\phi_1). \quad (6.7)$$

i) Analicemos primero el caso en que f cumple **C2**. En ese caso, el Corolario 5.5.1 implica que $\sup_{\|\alpha\|=1} |\hat{\varsigma}(\alpha) - \varsigma(\alpha)| \xrightarrow{a.s.} 0$, de donde $\hat{a}_N = \hat{\varsigma}(\hat{\phi}_{f,CR,1}) - \varsigma(\hat{\phi}_{f,CR,1}) \xrightarrow{a.s.} 0$. Por lo tanto, de (6.4), (6.5) y (6.7) obtenemos

$$\varsigma(\phi_1) = \hat{\varsigma}(\tilde{\phi}_1) - \hat{b}_N \leq \hat{\varsigma}(\hat{\phi}_{f,CR,1}) - \hat{b}_N = \varsigma(\hat{\phi}_{f,CR,1}) + \hat{a}_N - \hat{b}_N \leq \varsigma(\phi_1) + \hat{a}_N - \hat{b}_N. \quad (6.8)$$

De esta desigualdad, y usando que $\hat{a}_N = o_{a.s.}(1)$ y $\hat{b}_N = o_{a.s.}(1)$, se llega a que $\varsigma(\hat{\phi}_{f,CR,1}) \xrightarrow{a.s.} \varsigma(\phi_1)$, con lo cual usando el Lema 5.5.3 se obtiene el resultado.

ii) Consideremos ahora el caso $f = \log$. Como antes tendremos,

$$\varsigma(\phi_1) = \hat{\varsigma}(\tilde{\phi}_1) - \hat{b}_N \leq \hat{\varsigma}(\hat{\phi}_{f,CR,1}) - \hat{b}_N.$$

Sea \mathcal{N}_1 es el conjunto de medida 0 donde no se cumple **C3**, $\mathcal{N}_2 = \{\omega \in \Omega : \widehat{b}_N \not\rightarrow 0\}$ y $\mathcal{N}_3 = \{\omega \in \Omega : \widetilde{\phi}_1 \not\rightarrow \phi_1\}$ y definamos $\mathcal{N} = \cup_{i=1}^3 \mathcal{N}_i$. Luego, si $\omega \notin \mathcal{N}$ se cumple

$$\sup_{\|\alpha\|=1} |s_{i,n_i}^2(\alpha) - \sigma_i^2(\alpha)| \rightarrow 0, \quad \widehat{b}_N \rightarrow 0, \quad \widetilde{\phi}_1 \rightarrow \phi_1. \quad (6.9)$$

Luego, como $\varsigma(\phi_1) + \widehat{b}_N \leq \widehat{\varsigma}(\widehat{\phi}_{f,\text{CR},1})$, obtenemos que

$$\liminf_{N \rightarrow \infty} \widehat{\varsigma}(\widehat{\phi}_{f,\text{CR},1}(\omega)) \geq \varsigma(\phi_1) \text{ con probabilidad 1.} \quad (6.10)$$

Como $\sigma_i^2(\phi_1) > 0$, para todo $1 \leq i \leq k$, $|\varsigma(\phi_1)| < \infty$. De aquí se deduce fácilmente que, para todo $1 \leq i \leq k$, $\liminf_{N \rightarrow \infty} \sigma_i^2(\widehat{\phi}_{f,\text{CR},1}(\omega)) > 0$. Efectivamente, supongamos que no, entonces, existe i_0 tal que $\liminf_{N \rightarrow \infty} \sigma_{i_0}^2(\widehat{\phi}_{f,\text{CR},1}(\omega)) = 0$, es decir existe una subsucesión γ_{N_ℓ} de $\gamma_N = \widehat{\phi}_{f,\text{CR},1}(\omega)$ tal que $\sigma_{i_0}^2(\gamma_{N_\ell}) \rightarrow 0$. Como se cumple (6.9), tenemos que $s_{i_0, \tau_{i_0, N_\ell}}^2(\gamma_{N_\ell}) \rightarrow 0$ de donde $\widehat{\varsigma}_{N_\ell}(\gamma_{N_\ell}) \rightarrow -\infty$ lo que contradice (6.10).

Luego, para todo $1 \leq i \leq k$, $\liminf_{N \rightarrow \infty} \sigma_i^2(\widehat{\phi}_{f,\text{CR},1}(\omega)) > 0$, de donde usando argumentos análogos a los del Teorema 5.5.2 se deduce que $\widehat{a}_N = \widehat{\varsigma}(\widehat{\phi}_{f,\text{CR},1}) - \varsigma(\widehat{\phi}_{f,\text{CR},1}) \xrightarrow{a.s.} 0$ y la demostración se concluye como en i) usando (6.8). \square

6.4.2 Caso de estimadores con penalización en la norma

Cuando penalizamos en la norma para estimar $\phi_{\text{R},1}(P) = \operatorname{argmax}_{\|\alpha\|=1} \sum_{i=1}^k \tau_i f(\sigma_{\text{R}}^2(P))$, en base a k muestras aleatorias $X_{1,i}, \dots, X_{n_i,i}$, $1 \leq i \leq k$, los estimadores propuestos estaban dados por (5.6). Para efectuar su cálculo, si existe alguna población i_0 tal que $\mathbb{P}(X_{i_0} \in \mathcal{H}_{\text{S}}) = 1$, proponemos el siguiente algoritmo. Sea $\mathcal{J} = \{i, \mathbb{P}(X_{i,1} \in \mathcal{H}_{\text{S}}) = 1\}$.

1. Para $i \in \mathcal{J}$, normalizar las observaciones $\alpha_{i,j} = X_{i,j} / \|X_{i,j}\|_r$ y considerar el conjunto de posibles direcciones $\mathcal{A}_{N,r} = \{\alpha_{i,j} = X_{i,j} / \|X_{i,j}\|_r, 1 \leq j \leq n_i, i \in \mathcal{J}\} \subset \mathcal{S}$ donde $N = \sum_{i=1}^k n_i$.
2. Estimar $\phi_{f,1}(P)$ mediante

$$\widehat{\phi}_{f,\text{CRN},1} = \operatorname{argmax}_{1 \leq j \leq n_s, s \in \mathcal{J}} \sum_{i=1}^k \widehat{\tau}_i f(\sigma_{\text{R}}^2(P_{i,n_i}[\alpha_{s,j}])) = \operatorname{argmax}_{\alpha \in \mathcal{A}_{N,r}} \widehat{\varsigma}(\alpha).$$

Tenemos ahora el siguiente resultado

Teorema 6.4.2. Sean $X_{i,1} \in \mathcal{H}$, $1 \leq i \leq k$ independientes, donde \mathcal{H} es un espacio de Hilbert separable \mathcal{H} tales que $\mathbb{P}(X_{i,1} \in \mathcal{H}_{\text{S}}) = 1$, para todo $i \in \mathcal{J}$. Supongamos que $X_{i,j} \sim P_i$ con $P_i \sim \mathcal{E}(0, \mathbf{\Gamma}_i)$ donde $\mathbf{\Gamma}_i$ un operador compacto tal que $\mathbf{\Gamma}_i = \sum_{\ell \geq 1} \lambda_{i,\ell} \phi_\ell \otimes \phi_\ell$, siendo $\lambda_{i,\ell}$ los autovalores, ordenados de modo tal que $\lambda_{i,1} \geq \lambda_{i,2} \geq \dots$ y ϕ_ℓ las autofunciones comunes de los $\mathbf{\Gamma}_i$, con $\sum_{\ell \geq 1} \lambda_{i,\ell} < \infty$. Supongamos que $\mathbb{P}(X_{i,1} = 0) = 0$ para $1 \leq i \leq k$. Sean $X_{i,1}, \dots, X_{i,n_i}$ muestras aleatorias independientes con distribución P_i y supongamos que se cumple **C3** y que $\phi_{f,1}(P)$ es

único salvo signo. Más aún, supongamos que $n_i = \hat{\tau}_i N$ donde $\hat{\tau}_i$ son tales que convergen a τ_i y $0 < \tau_i < 1$ son números fijos con $\sum_{i=1}^k \tau_i = 1$. Supongamos que $f : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}$ cumple **C2** o bien $f = \log$. Sea r_N una sucesión numérica tal que $r = r_N \rightarrow 0$. Luego, $\langle \hat{\phi}_{f, \text{CRN}, 1}, \phi_{f, 1}(P) \rangle^2 \xrightarrow{a.s.} 1$, es decir, $\hat{\phi}_{f, \text{CRN}, 1} \xrightarrow{a.s.} \phi_{f, 1}(P)$ salvo eventual cambio de signo.

DEMOSTRACIÓN. Como $X_{i,j}$ es elíptico, se cumple que $\sigma_i^2(\alpha) = c_i < \alpha, \mathbf{\Gamma}_i \alpha >$ de donde $\phi_{f, 1}(P) = \phi_1$. Por otra parte, como $\sigma_i^2(\alpha) = c_i < \alpha, \mathbf{\Gamma}_i \alpha >$ y $\mathbf{\Gamma}_i$ es un operador compacto, se cumple también **C1**. Finalmente, como en el Teorema 6.3.1.1, tenemos que $\mathbb{P}(X_{i,1} = 0) = 0$ implica que $\lambda_{i,1} > 0$, de donde $\lambda_{f, i, 1}(P) = \sigma_i^2(\phi_1) = c_i \lambda_{i,1} > 0$ pues $\phi_{f, 1}(P)$ es único salvo signo.

Sea

$$\tilde{\phi}_1 = \tilde{\phi}_{1, N} = \underset{\alpha \in \mathcal{A}_{N, r}}{\operatorname{argmin}} 1 - \langle \alpha, \phi_1 \rangle^2,$$

es decir, $\langle \tilde{\phi}_1, \phi_1 \rangle^2 = \max_{1 \leq j \leq n_i, i \in \mathcal{J}} \langle X_{i,j} / \|X_{i,j}\|_r, \phi_1 \rangle^2$. Veamos que $\langle \tilde{\phi}_1, \phi_1 \rangle^2 \xrightarrow{a.s.} 1$, o sea, que salvo por elección del signo $\tilde{\phi}_1 \xrightarrow{a.s.} \phi_1$. Observemos primero que

$$\left| \left\langle \frac{X_{i,j}}{\|X_{i,j}\|_r}, \phi_1 \right\rangle \right| \leq \frac{\|X_{i,j}\|}{\|X_{i,j}\|_r} \leq 1,$$

de donde $\langle \tilde{\phi}_1, \phi_1 \rangle^2 \leq 1$.

Debemos probar que, para todo $\epsilon > 0$, $\lim_{M \rightarrow \infty} \mathbb{P}\left(\bigcup_{N \geq M} (1 - \langle \tilde{\phi}_1, \phi_1 \rangle^2) > \epsilon\right) = 0$, es decir, basta ver que $\sum_{N \geq 1} \mathbb{P}(1 - \langle \tilde{\phi}_{1, N}, \phi_1 \rangle^2 > \epsilon) < \infty$. Usando que $r = r_N$ es no aleatorio y que $n_i = \hat{\tau}_i N$ tenemos que

$$\begin{aligned} \mathbb{P}(1 - \langle \tilde{\phi}_1, \phi_1 \rangle^2 > \epsilon) &= \mathbb{P}\left(\max_{1 \leq j \leq n_i, i \in \mathcal{J}} \left\langle \frac{X_{i,j}}{\|X_{i,j}\|_r}, \phi_1 \right\rangle^2 < 1 - \epsilon\right) \\ &= \prod_{i \in \mathcal{J}} \prod_{j=1}^{n_i} \mathbb{P}\left(\left\langle \frac{X_{i,j}}{\|X_{i,j}\|_r}, \phi_1 \right\rangle^2 < 1 - \epsilon\right) = \prod_{i \in \mathcal{J}} \mathbb{P}\left(\left\langle \frac{X_{i,1}}{\|X_{i,1}\|_r}, \phi_1 \right\rangle^2 < 1 - \epsilon\right)^{n_i} \\ &= \left[\prod_{i \in \mathcal{J}} \mathbb{P}\left(\left\langle \frac{X_{i,1}}{\|X_{i,1}\|_r}, \phi_1 \right\rangle^2 < 1 - \epsilon\right)^{\hat{\tau}_i} \right]^N = \left[\prod_{i \in \mathcal{J}} p_{N,i}^{\hat{\tau}_i} \right]^N \end{aligned}$$

con $X_{i,1} \sim P_i$ y $p_{N,i} = \mathbb{P}(\langle X_{i,1} / \|X_{i,1}\|_r, \phi_1 \rangle^2 < 1 - \epsilon)$. Por el Lema 6.2.2, para todo $i \in \mathcal{J}$, existe un $N_{0,i}$ y un q_i tal que si $N \geq N_{0,i}$ entonces $p_{N,i} \leq q_i < 1$. Por otra parte, como $\hat{\tau}_i \rightarrow \tau_i$, existe N_1 tal que si $N \geq N_1$, $\frac{1}{2}\tau_i < \hat{\tau}_i < 2\tau_i$, $i \in \mathcal{J}$. Luego, si $N_0 = \max\{\max_{i \in \mathcal{J}}(N_{0,i}), N_1\}$ y $q = \max_{i \in \mathcal{J}}\{q_i\} < 1$ tenemos para todo $N \geq N_0$ y para todo $i \in \mathcal{J}$, $p_{N,i} \leq q < 1$, y entonces $\prod_{i \in \mathcal{J}} p_{N,i}^{\hat{\tau}_i} \leq \prod_{i \in \mathcal{J}} q^{\frac{1}{2}\tau_i} = q^{\frac{1}{2} \sum_{i \in \mathcal{J}} \tau_i} = q_0 < 1$, de donde

$$\begin{aligned} \sum_{N=1}^{\infty} \mathbb{P}(1 - \langle \tilde{\phi}_1, \phi_1 \rangle^2 > \epsilon) &\leq \sum_{N=1}^{N_0-1} \mathbb{P}(1 - \langle \tilde{\phi}_1, \phi_1 \rangle^2 > \epsilon) + \sum_{N=N_0}^{\infty} \left[\prod_{i=1}^k q_i^{\hat{\tau}_i} \right]^N \\ &\leq \sum_{N=1}^{N_0-1} \mathbb{P}(1 - \langle \tilde{\phi}_1, \phi_1 \rangle^2 > \epsilon) + \sum_{N=N_0}^{\infty} q_0^N < \infty \end{aligned}$$

Por lo tanto, $\langle \tilde{\phi}_1, \phi_1 \rangle^2 \xrightarrow{a.s.} 1$, de donde se deduce que $r\Psi(\tilde{\phi}_1) \xrightarrow{a.s.} 0$ pues $\|\tilde{\phi}_1\|_r = 1$. Sin pérdida de generalidad, se puede suponer que $\tilde{\phi}_1 \xrightarrow{a.s.} \phi_1$.

Observemos que **C3** implica que

$$\sup_{\|\alpha\| \leq 1} |s_{i,n_i}^2(\alpha) - \sigma_i^2(\alpha)| \xrightarrow{a.s.} 0, \quad (6.11)$$

para todo $1 \leq i \leq k$, de donde como $\|\alpha\|_r = 1$ implica que $\|\alpha\| \leq 1$, tenemos que $\sup_{\|\alpha\|_r=1} |s_{i,n_i}^2(\alpha) - \sigma_i^2(\alpha)| \xrightarrow{a.s.} 0$. Como en el Teorema 6.4.1, la continuidad débil de σ_i implica que $\sigma_i^2(\tilde{\phi}_1) - \sigma_i^2(\phi_1) \xrightarrow{a.s.} 0$. Además, (6.11) implica que $s_{i,n_i}^2(\tilde{\phi}_1) - \sigma_i^2(\tilde{\phi}_1) \xrightarrow{a.s.} 0$, de donde (6.2) vale. Por lo tanto, bajo **C2**, se verifica (6.3) y usando que $\lambda_{f,i,1}(P) > 0$ y el logaritmo es una función continua en $(0, \infty)$, se tiene también que (6.3) vale cuando $f = \log$. Es decir, en ambos casos, (6.4) se verifica. Por otra parte, como $\tilde{\phi}_1 \in \mathcal{A}_{N,r}$, $\|\tilde{\phi}_1\| \leq 1$ $\hat{\phi}_{f,CRN,1} = \operatorname{argmax}_{\alpha \in \mathcal{A}_{N,r}} \hat{\varsigma}(\alpha)$, $\phi_1 = \operatorname{argmax}_{\|\alpha\|=1} \varsigma(\alpha)$, $\|\hat{\phi}_{f,CRN,1}\| \leq \|\hat{\phi}_{f,CRN,1}\|_r = 1$ y f es estrictamente creciente se tienen las siguientes desigualdades

$$\hat{\varsigma}(\tilde{\phi}_1) \leq \hat{\varsigma}(\hat{\phi}_{f,CR,1}) \quad (6.12)$$

$$\varsigma(\tilde{\phi}_1) \leq \varsigma(\tilde{\phi}_1/\|\tilde{\phi}_1\|) \leq \varsigma(\phi_1) \quad (6.13)$$

$$\varsigma(\hat{\phi}_{f,CRN,1}) \leq \varsigma(\hat{\phi}_{f,CRN,1}/\|\hat{\phi}_{f,CRN,1}\|) \leq \varsigma(\phi_1). \quad (6.14)$$

i) Analicemos primero el caso en que f cumple **C2**. En ese caso, el Corolario 5.5.1 implica que $\sup_{\|\alpha\| \leq 1} |\hat{\varsigma}(\alpha) - \varsigma(\alpha)| \xrightarrow{a.s.} 0$, de donde $\hat{a}_N = \hat{\varsigma}(\hat{\phi}_{f,CR,1}) - \varsigma(\hat{\phi}_{f,CR,1}) \xrightarrow{a.s.} 0$. Por lo tanto, de (6.4), (6.12) y (6.14) obtenemos como en el Teorema 6.4.1 que (6.8) vale de donde $\varsigma(\hat{\phi}_{f,CR,1}) \xrightarrow{a.s.} \varsigma(\phi_1)$, con lo cual usando el Lema 5.5.3 se obtiene el resultado.

ii) Consideremos ahora el caso $f = \log$. Como antes tendremos,

$$\varsigma(\phi_1) = \hat{\varsigma}(\tilde{\phi}_1) - \hat{b}_N \leq \hat{\varsigma}(\hat{\phi}_{f,CR,1}) - \hat{b}_N$$

Como en el Teorema 6.4.1, sea \mathcal{N}_1 es el conjunto de medida 0 donde no se cumple (6.11), $\mathcal{N}_2 = \{\omega \in \Omega : \hat{b}_N \not\rightarrow 0\}$ y $\mathcal{N}_3 = \{\omega \in \Omega : \tilde{\phi}_1 \not\rightarrow \phi_1\}$ y definamos $\mathcal{N} = \cup_{i=1}^3 \mathcal{N}_i$. Luego, si $\omega \notin \mathcal{N}$ se cumple

$$\sup_{\|\alpha\| \leq 1} |s_{i,n_i}^2(\alpha) - \sigma_i^2(\alpha)| \rightarrow 0, \quad \hat{b}_N \rightarrow 0, \quad \tilde{\phi}_1 \rightarrow \phi_1. \quad (6.15)$$

Luego, como $\varsigma(\phi_1) + \hat{b}_N \leq \hat{\varsigma}(\hat{\phi}_{f,CR,1})$, obtenemos que

$$\liminf_{N \rightarrow \infty} \hat{\varsigma}(\hat{\phi}_{f,CRN,1}(\omega)) \geq \varsigma(\phi_1) \text{ con probabilidad 1} \quad (6.16)$$

y como $\sigma_i^2(\phi_1) > 0$, para todo $1 \leq i \leq k$, $|\varsigma(\phi_1)| < \infty$. De aquí se deduce fácilmente que, para todo $1 \leq i \leq k$, $\liminf_{N \rightarrow \infty} \sigma_i^2(\hat{\phi}_{f,CRN,1}(\omega)) > 0$, como en el Teorema 6.4.1, de donde usando argumentos análogos a los del Teorema 5.5.2 se deduce que $\hat{a}_N = \hat{\varsigma}(\hat{\phi}_{f,CRN,1}) - \varsigma(\hat{\phi}_{f,CRN,1}) \xrightarrow{a.s.} 0$ y la demostración se concluye como en i) usando (6.8). \square

6.4.3 Caso de estimadores con penalización en la escala

Cuando penalizamos en la escala para estimar $\phi_{f,1}(P) = \operatorname{argmax}_{\|\alpha\|=1} \sum_{i=1}^k \tau_i f(\sigma_R^2(P))$, los estimadores resolvían (5.5). Como en la sección 6.4.2, supondremos que existe alguna población i_0 tal que $\mathbb{P}(X_{i_0,1} \in \mathcal{H}_s) = 1$ y llamamos $\mathcal{J} = \{i : \mathbb{P}(X_{i,1} \in \mathcal{H}_s) = 1\}$. El algoritmo que proponemos es el siguiente

1. Normalizar las observaciones $\alpha_{i,j} = X_{i,j}/\|X_{i,j}\|$ y considerar el conjunto de posibles direcciones $\tilde{\mathcal{A}}_N = \{\alpha_{i,j} = X_{i,j}/\|X_{i,j}\|, 1 \leq j \leq n_i, i \in \mathcal{J}\} \subset \mathcal{S}$ donde $N = \sum_{i=1}^k n_i$.
2. Estimar $\phi_{f,1}(P)$ mediante

$$\hat{\phi}_{f,\text{CRS},1} = \operatorname{argmax}_{1 \leq j \leq n_s, s \in \mathcal{J}} \sum_{i=1}^k \hat{\tau}_i f(\sigma_R^2(P_{i,n_i}[\alpha_{s,j}])) - \rho \Psi(\alpha) = \operatorname{argmax}_{\alpha \in \tilde{\mathcal{A}}_N} \hat{\zeta}(\alpha) - \rho \Psi(\alpha).$$

Observemos que podríamos maximizar en \mathcal{A}_N definido en la Sección 6.4.2 pues para $i \notin \mathcal{J}$ $\Psi(X_{i,j}) = \infty$ y el máximo se toma sobre $\tilde{\mathcal{A}}_N$.

El siguiente resultado muestra la consistencia de $\hat{\phi}_{f,\text{CRS},1}$.

Teorema 6.4.3. Sean $X_{i,1} \in \mathcal{H}$, $1 \leq i \leq k$ independientes, donde \mathcal{H} es un espacio de Hilbert separable \mathcal{H} tales que $\mathbb{P}(X_{i,1} \in \mathcal{H}_s) = 1$, para todo $i \in \mathcal{J}$. Supongamos que $X_{i,j} \sim P_i$ con $P_i \sim \mathcal{E}(0, \mathbf{\Gamma}_i)$ donde $\mathbf{\Gamma}_i$ un operador compacto tal que $\mathbf{\Gamma}_i = \sum_{\ell \geq 1} \lambda_{i,\ell} \phi_\ell \otimes \phi_\ell$, siendo $\lambda_{i,\ell}$ los autovalores, ordenados de modo tal que $\lambda_{i,1} \geq \lambda_{i,2} \geq \dots$ y ϕ_ℓ las autofunciones comunes de los $\mathbf{\Gamma}_i$, con $\sum_{\ell \geq 1} \lambda_{i,\ell} < \infty$. Supongamos que $\mathbb{P}(X_{i,1} = 0) = 0$ para $1 \leq i \leq k$. Sean $X_{i,1}, \dots, X_{i,n_i}$ muestras aleatorias independientes con distribución P_i y supongamos que se cumple **C3** y $\phi_{f,1}(P)$ es único salvo signo. Más aún, supongamos que $n_i = \hat{\tau}_i N$ donde $\hat{\tau}_i \rightarrow \tau_i$ con $0 < \tau_i < 1$ y $\sum_{i=1}^k \tau_i = 1$. Supongamos que $f : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}$ cumple **C2** o bien $f = \log$. Sea ρ_N una sucesión numérica tal que $\rho = \rho_N \rightarrow 0$. Luego, $\langle \hat{\phi}_{f,\text{CRS},1}, \phi_{f,1}(P) \rangle^2 \xrightarrow{\text{a.s.}} 1$, es decir, $\hat{\phi}_{f,\text{CRS},1} \xrightarrow{\text{a.s.}} \phi_{f,1}(P)$ salvo eventual cambio de signo.

DEMOSTRACIÓN. Como en la demostración del Teorema 6.4.2, definamos $\mathcal{A}_{n,\rho} = \{X_{i,j}/\|X_{i,j}\|_\rho, 1 \leq j \leq n_i, i \in \mathcal{J}\} \subset \mathcal{S}$ y

$$\tilde{\phi}_1 = \operatorname{argmin}_{1 \leq j \leq n_i, i \in \mathcal{J}} 1 - \langle \phi_1, \frac{\|X_{i,j}\|}{\|X_{i,j}\|_\rho} \rangle^2 = \operatorname{argmin}_{\alpha \in \mathcal{A}_{N,\rho}} 1 - \langle \phi_1, \alpha \rangle^2.$$

Usando que $\rho \rightarrow 0$, como en la demostración del Teorema 6.4.2, obtenemos que $\langle \tilde{\phi}_1, \phi_1 \rangle^2 \xrightarrow{\text{a.s.}} 1$, o sea, podemos suponer que $\tilde{\phi}_1 \xrightarrow{\text{a.s.}} \phi_1$ salvo eventual signo. En particular, como para todo $i \in \mathcal{J}$, $1 \leq j \leq n_i$, se cumple que $\|X_{i,j}\|/\|X_{i,j}\|_\rho \leq 1$, tenemos $\|\tilde{\phi}_1\| \leq 1$, de donde

$$\langle \tilde{\phi}_1, \phi_1 \rangle^2 \leq \|\tilde{\phi}_1\|^2 \|\phi_1\|^2 = \|\tilde{\phi}_1\|^2 \leq 1,$$

por lo tanto, $\|\tilde{\phi}_1\|^2 \xrightarrow{\text{a.s.}} 1$. Como $\|\tilde{\phi}_1\|_\rho = 1$, tendremos que $\rho \Psi(\tilde{\phi}_1) \xrightarrow{\text{a.s.}} 0$.

Observemos que como $\tilde{\phi}_1/\|\tilde{\phi}_1\| \in \tilde{\mathcal{A}}_N$,

$$\begin{aligned}\widehat{\varsigma}(\widehat{\phi}_{f,\text{CRS},1}) &\geq \widehat{\varsigma}(\widehat{\phi}_{f,\text{CRS},1}) - \rho\Psi(\widehat{\phi}_{f,\text{CRS},1}) = \max_{\alpha \in \tilde{\mathcal{A}}_N} \{\widehat{\varsigma}(\alpha) - \rho\Psi(\alpha)\} \\ &\geq \widehat{\varsigma}\left(\frac{\tilde{\phi}_1}{\|\tilde{\phi}_1\|}\right) - \rho\Psi\left(\frac{\tilde{\phi}_1}{\|\tilde{\phi}_1\|}\right) = \widehat{\varsigma}\left(\frac{\tilde{\phi}_1}{\|\tilde{\phi}_1\|}\right) - \frac{\rho\Psi(\tilde{\phi}_1)}{\|\tilde{\phi}_1\|^2}.\end{aligned}$$

Por otra parte, el hecho que f sea estrictamente creciente y $\|\tilde{\phi}_1\| \leq 1$, implican que

$$\widehat{\varsigma}\left(\frac{\tilde{\phi}_1}{\|\tilde{\phi}_1\|}\right) \geq \widehat{\varsigma}(\tilde{\phi}_1),$$

de donde se deduce que

$$\widehat{\varsigma}(\widehat{\phi}_{f,\text{CRS},1}) \geq \widehat{\varsigma}(\tilde{\phi}_1) - \rho\frac{\Psi(\tilde{\phi}_1)}{\|\tilde{\phi}_1\|^2}. \quad (6.17)$$

Usando (6.11) y que $\|\tilde{\phi}_1\| \leq 1$, se obtiene que $a_{i,n_i} = s_{i,n_i}^2(\tilde{\phi}_1) - \sigma_i^2(\tilde{\phi}_1) \xrightarrow{a.s.} 0$. Por otra parte, como $\tilde{\phi}_1 \xrightarrow{a.s.} \phi_1$, por la continuidad débil de σ_i tendremos que $b_{i,n_i} = \sigma_i^2(\tilde{\phi}_1) - \sigma_i^2(\phi_1) \xrightarrow{a.s.} 0$, de donde $c_{i,n_i} = s_{i,n_i}^2(\tilde{\phi}_1) - \sigma_i^2(\phi_1) \xrightarrow{a.s.} 0$. Luego, como en el Teorema 6.4.1, obtenemos que

$$\widehat{b}_N = \widehat{\varsigma}(\tilde{\phi}_1) - \varsigma(\phi_1) \xrightarrow{a.s.} 0, \quad (6.18)$$

con lo cual usando (6.17) se deduce que

$$\widehat{\varsigma}(\widehat{\phi}_{f,\text{CRS},1}) \geq \varsigma(\phi_1) - \widehat{b}_N - \rho\frac{\Psi(\tilde{\phi}_1)}{\|\tilde{\phi}_1\|^2}.$$

Recordemos que $\|\tilde{\phi}_1\|^2 \xrightarrow{a.s.} 1$ y $\rho\Psi(\tilde{\phi}_1) \xrightarrow{a.s.} 0$, luego

$$\widehat{\varsigma}(\widehat{\phi}_{f,\text{CRS},1}) \geq \varsigma(\phi_1) + o_{a.s.}(1). \quad (6.19)$$

Queremos probar ahora que

$$\widehat{\varsigma}(\widehat{\phi}_{f,\text{CRS},1}) \leq \varsigma(\phi_1) + o_{a.s.}(1). \quad (6.20)$$

Como $\|\widehat{\phi}_{f,\text{CRS},1}\| = 1$

$$\varsigma(\phi_1) \geq \varsigma(\widehat{\phi}_{f,\text{CRS},1}) = \widehat{\varsigma}(\widehat{\phi}_{f,\text{CRS},1}) - d_N$$

donde $d_N = \widehat{\varsigma}(\widehat{\phi}_{f,\text{CRS},1}) - \varsigma(\widehat{\phi}_{f,\text{CRS},1}) \xrightarrow{a.s.} 0$. Bastaría probar que $d_N \xrightarrow{a.s.} 0$ para obtener (6.20). Observemos que (6.11) y el hecho que $\|\widehat{\phi}_{f,\text{CRS},1}\| = 1$ implican que $d_{i,n_i} = s_{i,n_i}^2(\widehat{\phi}_{f,\text{CRS},1}) - \sigma_i^2(\widehat{\phi}_{f,\text{CRS},1}) \xrightarrow{a.s.} 0$. Bajo **C2**, el Lema 5.5.2, implica que

$$\sup_{\|\alpha\| \leq 1} |\widehat{\varsigma}(\alpha) - \varsigma(\alpha)| \xrightarrow{a.s.} 0$$

de donde se obtiene que $d_N = \widehat{\varsigma}(\widehat{\phi}_{f,\text{CRS},1}) - \varsigma(\widehat{\phi}_{f,\text{CRS},1}) \xrightarrow{a.s.} 0$.

Si $f = \log$, argumentando como en el Teorema 5.5.2, se obtiene la misma conclusión.

De (6.19) y (6.20) se deduce que $\widehat{\varsigma}(\widehat{\phi}_{f,\text{CRS},1}) \xrightarrow{a.s.} \varsigma(\phi_1)$ de donde como $d_N \xrightarrow{a.s.} 0$, tenemos $\varsigma(\widehat{\phi}_{f,\text{CRS},1}) \xrightarrow{a.s.} \varsigma(\phi_1)$ y la demostración se concluye utilizando el Lema 5.5.3. \square

6.4.4 Caso de estimadores basados en *sieves*

Sea $\{\delta_i\}_{i \in \mathbb{N}}$ una base ortonormal de \mathcal{H} . Sea $p_N \rightarrow \infty$, definimos π_{p_N} como la proyección ortogonal sobre el subespacio generado por $\delta_1, \dots, \delta_{p_N}$. Los estimadores basados en *sieves* cumplan (5.10), o sea, $\hat{\phi}_{\text{SI},1} = \operatorname{argmax}_{\alpha \in \mathcal{S}_{p_N}} \varsigma(\alpha)$. Para poder aproximar estos estimadores de modo de efectuar su cálculo, proponemos

1. Proyectar y normalizar las observaciones $\alpha_{i,j} = \pi_{p_N} X_{i,j} / \|\pi_{p_N} X_{i,j}\|$ y considerar el conjunto de posibles direcciones $\mathcal{A}_{N,p_N} = \{\alpha_{i,j} = \pi_{p_N} X_{i,j} / \|\pi_{p_N} X_{i,j}\|, 1 \leq j \leq n_i, 1 \leq i \leq k\} \subset \mathcal{S}$, donde $N = \sum_{i=1}^k n_i$. Si $\pi_{p_N} X_{i,j} = 0$, se define $\alpha_{i,j} = 0$.
2. Estimar $\phi_{f,1}(P)$ mediante

$$\hat{\phi}_{f,\text{CR},1} = \operatorname{argmax}_{1 \leq j \leq n_i, 1 \leq i \leq k} \widehat{\varsigma}(\alpha_{i,j}) = \operatorname{argmax}_{\alpha \in \mathcal{A}_{N,p_N}} \widehat{\varsigma}(\alpha).$$

El siguiente Teorema muestra que esta aproximación es consistente.

Teorema 6.4.4. Sean $X_{i,1} \in \mathcal{H}$, $1 \leq i \leq k$ independientes, donde \mathcal{H} es un espacio de Hilbert separable \mathcal{H} . Supongamos que $X_{i,j} \sim P_i$ con $P_i \sim \mathcal{E}(0, \mathbf{\Gamma}_i)$ donde $\mathbf{\Gamma}_i$ un operador compacto tal que $\mathbf{\Gamma}_i = \sum_{\ell \geq 1} \lambda_{i,\ell} \phi_\ell \otimes \phi_\ell$, siendo $\lambda_{i,\ell}$ los autovalores, ordenados de modo tal que $\lambda_{i,1} \geq \lambda_{i,2} \geq \dots$ y ϕ_ℓ las autofunciones comunes de los $\mathbf{\Gamma}_i$, con $\sum_{\ell \geq 1} \lambda_{i,\ell} < \infty$. Supongamos que $\mathbb{P}(X_{i,1} = 0) = 0$ para $1 \leq i \leq k$. Más aún, supongamos que para $1 \leq i \leq k$, $\mathbb{P}(\pi_{p_N} X_{i,1} = 0) = 0$ si $N \geq N_0$. Sean $X_{i,1}, \dots, X_{i,n_i}$ muestras aleatorias independientes con distribución P_i y supongamos que se cumplen **C3** y $\phi_{f,1}(P)$ es único salvo signo. Más aún, supongamos que $n_i = \hat{\tau}_i N$ donde $\hat{\tau}_i \rightarrow \tau_i$ con $0 < \tau_i < 1$ y $\sum_{i=1}^k \tau_i = 1$. Supongamos que $f : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}$ cumple **C2** o bien $f = \log$ y que $p_N \rightarrow \infty$. Luego, $(\hat{\phi}_{f,\text{CR},1}, \phi_{f,1}(P))^2 \xrightarrow{a.s.} 1$, es decir, $\hat{\phi}_{f,\text{CR},1} \xrightarrow{a.s.} \phi_{f,1}(P)$ salvo eventual cambio de signo.

DEMOSTRACIÓN. Como $X_{i,j}$ es elíptico, se cumple que $\sigma_i^2(\alpha) = c_i \langle \alpha, \mathbf{\Gamma}_i \alpha \rangle$ de donde $\phi_{f,1}(P) = \phi_1$. Por otra parte, como $\sigma_i^2(\alpha) = c_i \langle \alpha, \mathbf{\Gamma}_i \alpha \rangle$ y $\mathbf{\Gamma}_i$ es un operador compacto, se cumple también **C1**. Finalmente, como en el Teorema 6.3.1.1, tenemos que $\mathbb{P}(X_{i,1} = 0) = 0$ implica que $\lambda_{i,1} > 0$, de donde $\lambda_{f,i,1}(P) = \sigma_i^2(\phi_1) = c_i \lambda_{i,1} > 0$ pues $\phi_{f,1}(P)$ es único salvo signo.

Sea

$$\tilde{\phi}_{\pi,1} = \tilde{\phi}_{\pi,1,N} = \operatorname{argmin}_{1 \leq j \leq n_i, 1 \leq i \leq k} 1 - \left\langle \frac{\pi_{p_N} X_{i,j}}{\|\pi_{p_N} X_{i,j}\|}, \phi_1 \right\rangle^2 = \operatorname{argmin}_{\alpha \in \mathcal{A}_{N,p_N}} 1 - \langle \alpha, \phi_1 \rangle^2, ,$$

es decir, $\langle \tilde{\phi}_{\pi,1}, \phi_1 \rangle^2 = \max_{1 \leq j \leq n_i, 1 \leq i \leq k} \langle \alpha_{i,j}, \phi_1 \rangle^2$. Veamos que $\langle \tilde{\phi}_{\pi,1}, \phi_1 \rangle^2 \xrightarrow{a.s.} 1$, o sea, que salvo por elección del signo $\tilde{\phi}_{\pi,1} \xrightarrow{a.s.} \phi_1$.

Podemos definir, como antes en el Teorema 6.4.1, $\tilde{\phi}_1 = \tilde{\phi}_{1,N} = \operatorname{argmin}_{\alpha \in \mathcal{A}_N} 1 - \langle \alpha, \phi_1 \rangle^2$. Ya sabemos que $\langle \tilde{\phi}_1, \phi_1 \rangle^2 \xrightarrow{a.s.} 1$. Usemos eso para probar que $a_N = \langle \tilde{\phi}_{\pi,1}, \phi_1 \rangle^2 \xrightarrow{a.s.} 1$. Por Cauchy-Schwartz sabemos que esto es menor o igual que 1, por lo tanto basta ver que a_N está acotado inferiormente por una sucesión que tiende casi seguramente a 1.

En efecto, si $N \geq N_0$, $\mathbb{P}(\pi_{p_N} X_{i,j} \neq 0) = 1$, luego usando que $\|\pi_{p_N}\|^2 = 1$

$$\begin{aligned} \langle \tilde{\phi}_{\pi,1}, \phi_1 \rangle^2 &= \max_{1 \leq j \leq n_i, 1 \leq i \leq k} \left\langle \frac{\pi_{p_N} X_{i,j}}{\|\pi_{p_N} X_{i,j}\|}, \phi_1 \right\rangle^2 = \max_{1 \leq j \leq n_i, 1 \leq i \leq k} \frac{\|X_{i,j}\|^2}{\|\pi_{p_N} X_{i,j}\|^2} \left\langle \frac{X_{i,j}}{\|X_{i,j}\|}, \pi_{p_N} \phi_1 \right\rangle^2 \\ &\geq \max_{1 \leq j \leq n_i, 1 \leq i \leq k} \frac{\|X_{i,j}\|^2}{\|\pi_{p_N}\|^2 \|X_{i,j}\|^2} \left\langle \frac{X_{i,j}}{\|X_{i,j}\|}, \pi_{p_N} \phi_1 \right\rangle^2 \\ &\geq \max_{1 \leq j \leq n_i, 1 \leq i \leq k} \left\langle \frac{X_{i,j}}{\|X_{i,j}\|}, \pi_{p_N} \phi_1 \right\rangle^2 = \max_{1 \leq j \leq n_i, 1 \leq i \leq k} \left(\left\langle \frac{X_{i,j}}{\|X_{i,j}\|}, \phi_1 \right\rangle + \left\langle \frac{X_{i,j}}{\|X_{i,j}\|}, \pi_{p_N} \phi_1 - \phi_1 \right\rangle \right)^2 \end{aligned}$$

Luego, como por Cauchy–Schwartz

$$\max_{1 \leq j \leq n_i, 1 \leq i \leq k} \left| \left\langle \frac{X_{i,j}}{\|X_{i,j}\|}, \pi_{p_N} \phi_1 - \phi_1 \right\rangle \right| \leq \|\pi_{p_N} \phi_1 - \phi_1\|$$

y $p_N \rightarrow \infty$, se tiene que $\max_{1 \leq j \leq n_i, 1 \leq i \leq k} |\langle X_{i,j}/\|X_{i,j}\|, \pi_{p_N} \phi_1 - \phi_1 \rangle| = o_{a.s.}(1)$, de donde

$$\begin{aligned} |\langle \tilde{\phi}_{\pi,1}, \phi_1 \rangle| &\geq \max_{1 \leq j \leq n_i, 1 \leq i \leq k} \left| \left\langle \frac{X_{i,j}}{\|X_{i,j}\|}, \phi_1 \right\rangle + \left\langle \frac{X_{i,j}}{\|X_{i,j}\|}, \pi_{p_N} \phi_1 - \phi_1 \right\rangle \right| \\ &\geq \max_{1 \leq j \leq n_i, 1 \leq i \leq k} \left| \left\langle \frac{X_{i,j}}{\|X_{i,j}\|}, \phi_1 \right\rangle \right| - \max_{1 \leq i \leq k, 1 \leq j \leq n_i} \left| \left\langle \frac{X_{i,j}}{\|X_{i,j}\|}, \pi_{p_N} \phi_1 - \phi_1 \right\rangle \right| \\ &\geq \max_{1 \leq j \leq n_i, 1 \leq i \leq k} \left| \left\langle \frac{X_{i,j}}{\|X_{i,j}\|}, \phi_1 \right\rangle \right| + o_{a.s.}(1) = |\langle \tilde{\phi}_1, \phi_1 \rangle|, \end{aligned}$$

como deseábamos. Luego, $\tilde{\phi}_{\pi,1} \xrightarrow{a.s.} \phi_1$ salvo eventualmente un signo. El resto de la demostración sigue como la demostración del Teorema 6.4.1, cambiando $\tilde{\phi}_1$ por $\tilde{\phi}_{\pi,1}$. \square

6.5 Estudio de Monte Carlo

Para estudiar el comportamiento de los estimadores definidos por los algoritmos anteriores para muestras moderadas, hemos realizado un estudio de simulación para una población con diferentes tamaños muestrales. Sea X_1, \dots, X_N , i.i.d., $X_i \in L^2(0, 1)$. Consideramos como tamaños muestrales $N = 50, 100, 250, 500, 1000$. Es claro que, si las trayectorias no son suaves los algoritmos descriptos para los estimadores penalizados no garantizarán consistencia. Para cubrir situaciones con trayectorias suaves y no suaves hemos considerado dos núcleos de covarianza para procesos Gaussianos y también introducimos un tipo de contaminación que produce trayectorias irregulares al introducir un pico de forma aleatoria. Realizamos 100 replicaciones.

La distribución de las observaciones serán indicadas como C_0 y C_1 correspondiendo a las siguientes situaciones

- C_0 : En este caso los $X_i(t)$ son Gaussianos con núcleo de covarianza igual a
 - **Modelo 1**: $\gamma_X(s, t) = (1/2)(1/2)^{0.9|s-t|}$, correspondiendo a un proceso de Ornstein Uhlenbeck.

– **Modelo 2:** $\gamma_X(s, t) = (1/2)(1/2)^{0.9(s-t)^2}$, que genera trayectorias suaves.

- C_1 : Esta contaminación corresponde a un pico y fue considerado por Sawant *et al.* (2011). Las contaminaciones observadas $X_i^{(c)}$ se definen como

$$X_i^{(c)}(s) = X_i(s) + V_i D_i M \mathbb{I}_{\{T_i < s < T_i + \ell\}}$$

donde $V_i \sim Bi(1, p)$, D_i es tal que $\mathbb{P}(D_i = 1) = \mathbb{P}(D_i = -1) = 1/2$, $T_i \sim \mathcal{U}(0, 1 - \ell)$, $\ell < 1/2$ y V_i, X_i, D_i y T_i son independientes.

Cálculos más o menos directos, nos permiten obtener que

$$\gamma_{X^{(c)}}(s, t) = \gamma_X(s, t) + M^2 p \left\{ \frac{\min(\min(s, t), 1 - \ell) - \max\{\max(s - \ell, t - \ell), 0\}}{1 - \ell} \right\} I_{\max(s - \ell, t - \ell) < \min(s, t)}.$$

Elegimos $\ell = 1/15$, $M = 30$ y $p = 0.1$.

Cuando consideramos el desvío estándar, la primer dirección objetivo $\phi_{SD,1}(P_{C_1})$ es tal que

- **Modelo 1:** $|\cos(\phi_{SD,1}(P_{C_1}), \phi_{SD,1}(P_{C_0}))| \simeq 0.9871$, entonces $\|\phi_{SD,1}(P_{C_1}) - \phi_{SD,1}(P_{C_0})\|^2 = 0.0258$.
- **Modelo 2:** $|\cos(\phi_{SD,1}(P_{C_1}), \phi_{SD,1}(P_{C_0}))| \simeq 0.9992$, entonces $\|\phi_{SD,1}(P_{C_1}) - \phi_{SD,1}(P_{C_0})\|^2 = 0.0015$.

Además, si notamos por $SD_\alpha(P)$ al desvío $\langle \alpha, X \rangle$ cuando $X \sim P$, tenemos que

- **Model 1:** $SD_{\phi_{SD,1}^2(P_{C_1})}(P_{C_1}) = 1.245387$ mientras $SD_{\phi_{SD,1}^2(P_{C_0})}(P_{C_1}) = 1.219656$.
- **Model 2:** $SD_{\phi_{SD,1}^2(P_{C_1})}(P_{C_1}) = 4.872657$ mientras $SD_{\phi_{SD,1}^2(P_{C_0})}(P_{C_1}) = 4.865712$.

Luego, para tamaños muestrales reducidos, $\phi_{SD,1}(P_{C_1})$ será difícil de estimar ya que los valores de la función objetivo son bastante similares a los de $\phi_{SD,1}(P_{C_0})$. Es claro que uno puede estimar las direcciones principales usando las autofunciones del operador muestral de covarianza, lo que lleva a los estimadores *raw*. Sin embargo, para poder calcular los estimadores penalizados clásicos tenemos que calcular una aproximación sobre la bola unitaria. Por esa razón, calculamos los estimadores *raw* usando el algoritmo propuesto en la sección Sección 6.3.

Además de considerar el estimador clásico basado en el desvío estándar, calculamos estimadores robustos basados en *projection-pursuit* utilizando un M -estimador de escala (M -ESCALA). Para el M -estimador, utilizamos como función de *score* $\chi_c(y) = \min\left(3(y/c)^2 - 3(y/c)^4 + (y/c)^6, 1\right)$ con $c = 1.56$ y punto de ruptura $1/2$. Para calcular la M -escala, el estimador inicial de escala fue la MAD. Para implementar los estimadores las observaciones fueron computadas sobre una grilla de M puntos.

6.5.1 Comportamiento de los estimadores

Debido a la complejidad numérica, se reportan solamente los resultados correspondientes a los estimadores *raw* $\hat{\phi}_{\text{CR},1}$ y los estimadores obtenidos mediante penalización de la escala $\hat{\phi}_{\text{CRS},1}$. Para los últimos, el parámetro de penalizado ρ se tomó fijo e igual a $\rho = 10^{-5}$ para evitar el alto costo computacional del procedimiento de convalidación cruzada.

Las Tablas 6.1 a 6.2 reportan el valor medio sobre las repeticiones de $\|\hat{\phi}_{\text{CR},1} - \phi_1\|^2$ y $\|\hat{\phi}_{\text{CRS},1} - \phi_1\|^2$ bajo C_0 mientras que las Tablas 6.3 a 6.4 reportan la misma medida resumen bajo C_1 .

Los resultados muestran que la convergencia es más rápida si las trayectorias son suaves. en particular, bajo C_1 , cuando se consideran estimadores basados en el desvío estándar, se obtienen resultados pobres para tamaños muestrales menores a 500, para los modelos 1 y 3, lo que se podría confundir con una falta de robustez cuando de hecho lo que pasa es una falta de eficiencia en el algoritmo para aproximar el verdadero máximo sobre la bola unitaria.

El nivel de discretización de la grilla, M , no mejora los resultados obtenidos. Con respecto al comportamiento de $\hat{\phi}_{\text{CR},1}$ y $\hat{\phi}_{\text{CRS},1}$, bajo C_0 , el último se comporta como el primero bajo el Modelo 2, puesto que las trayectorias son muy suaves y ofrecen así un buen conjunto de candidatos $\mathcal{A}_N = \{\alpha_i = X_i/\|X_i\|, 1 \leq i \leq N\} \subset \mathcal{S}$. Por otro lado, para el Modelo 1, la penalización ofrece peores resultados que los estimadores “puros” bajo C_0 debido a la falta de suavidad de las trayectorias.

Como se mencionó antes, el algoritmo descrito en la Sección 6.3 no está bien definido si las trayectorias no son suaves cuando se considera estimadores penalizados, lo que explica el pobre comportamiento obtenido en esta situación. Para resolver este problema, sugerimos combinar dicho algoritmo agregando como candidatos direcciones aleatorias como las propuestas en Cuesta Albertos *et al.* (2009). El algoritmo para generar direcciones adicionales es el siguiente:

Indiquemos por $Be(p_1, p_2)$ la distribución beta de parámetros p_1, p_2

- a) Se generan $b_0 \in [0, 1]$, $b_0 \sim Be(p_1, p_2)$.
- b) Dados $m \geq 1$, $b_m \in [0, 1 - \sum_{i=0}^{m-1} b_i]$, $b_m \sim (1 - \sum_{i=0}^{m-1} b_i)Be(p_1, p_2)$

Definamos para cualquier base δ_i de \mathcal{H} , $\alpha = \sum_{i \geq 1} b_{i-1} \delta_i$, luego $\|\alpha\| = 1$ casi seguramente.

En esta simulación, elegimos $p_1 = 100$, $p_2 = 1$ y como base la base de Fourier y generamos 500 direcciones aleatorias α_i como se describe en a) y b). Estas direcciones se agregan al conjunto \mathcal{A}_N para así obtener un conjunto más grande de candidatos $\tilde{\mathcal{A}}_N = \mathcal{A}_N \cup \{\alpha_i, 1 \leq i \leq 500\} \subset \mathcal{S}$. Esperamos que las direcciones aleatorias produzcan una mejora en la eficiencia del algoritmo, en el sentido de que debieran proveer una mejor aproximación al estimador verdadero que se obtendría maximizando sobre toda la bola unitaria, cuando las trayectorias no son suaves.

Los estimadores obtenidos mediante este procedimiento son denotados $\hat{\phi}_{\text{CRB},1}$. Cuando consideramos un acercamiento penalizado en la escala los referiremos como $\hat{\phi}_{\text{CRSB},1}$.

Las Tablas 6.5 a 6.6 reportan el valor medio sobre todas las repeticiones de $\|\hat{\phi}_{\text{CRB},1} - \phi_1\|^2$ y $\|\hat{\phi}_{\text{CRSB},1} - \phi_1\|^2$, bajo C_0 mientras que las Tablas 6.7 a 6.8 reportan las mismas medidas de resumen bajo C_1 .

El nuevo conjunto de direcciones llevan a resultados más precisos para todos los tamaños muestrales. En particular, para el Modelo 1, cuando se penaliza en la escala, bajo C_0 , $D_1 = \|\hat{\phi}_1 - \phi_1\|^2$ es diez veces menor cuando se incluyen las direcciones beta. Bajo C_1 , los estimadores clásicos “puros” igualmente llevan a resultados de baja precisión mientras que en los robustos la media sobre todas las replicaciones de $D_1 = \|\hat{\phi}_1 - \phi_1\|^2$ es casi la mitad cuando se usa este nuevo procedimiento. Para el Modelo 2, los resultados son casi iguales independientemente de si se incluyen o no las nuevas direcciones, dado que las trayectorias son suaves. Algunas diferencias se pueden apreciar bajo C_1 cuando se considera el estimador robusto penalizado.

Queremos terminar este capítulo, mencionando que es necesario profundizar la investigación en esta área para tener así mejores algoritmos.

N	Escala	$M = 50$	$M = 100$	$M = 250$	$M = 50$	$M = 100$	$M = 250$
		$\hat{\phi}_{CR,1}$			$\hat{\phi}_{CRS,1} (\rho = 10^{-5})$		
50	SD	0.0226	0.0234	0.0219	0.0386	0.0445	0.0411
	M -ESCALA	0.0342	0.0333	0.0326	0.0410	0.0421	0.0418
100	SD	0.0147	0.0149	0.0155	0.0290	0.0349	0.0369
	M -ESCALA	0.0183	0.0197	0.0212	0.0294	0.0351	0.0374
250	SD	0.0109	0.0115	0.0102	0.0269	0.0336	0.0253
	M -ESCALA	0.0122	0.0144	0.0127	0.0265	0.0347	0.0266
500	SD	0.0084	0.0084	0.0079	0.0260	0.0274	0.0247
	M -ESCALA	0.0090	0.0100	0.0088	0.0245	0.0281	0.0247
1000	SD	0.0068	0.0067	0.0064	0.0210	0.0231	0.0208
	M -ESCALA	0.0075	0.0076	0.0071	0.0213	0.0238	0.0204

Tabla 6.1: Valores medios de $\|\hat{\phi}_1 - \phi_1\|^2$ cuando $\gamma_X(s, t) = (1/2)(1/2)^{0.9|s-t|}$ bajo C_0 .

N	Escala	$M = 50$	$M = 100$	$M = 250$	$M = 50$	$M = 100$	$M = 250$
		$\hat{\phi}_{CR,1}$			$\hat{\phi}_{CRS,1} (\rho = 10^{-5})$		
50	SD	0.0034	0.0042	0.0039	0.0036	0.0042	0.0040
	M -ESCALA	0.0108	0.0099	0.0108	0.0108	0.0099	0.0107
100	SD	0.0020	0.0020	0.0020	0.0020	0.0020	0.0021
	M -ESCALA	0.0049	0.0044	0.0059	0.0049	0.0044	0.0060
250	SD	0.0010	0.0008	0.0007	0.0010	0.0008	0.0007
	M -ESCALA	0.0019	0.0025	0.0022	0.0019	0.0025	0.0023
500	SD	0.0004	0.0003	0.0003	0.0005	0.0003	0.0003
	M -ESCALA	0.0008	0.0012	0.0008	0.0008	0.0012	0.0008
1000	SD	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002
	M -ESCALA	0.0006	0.0007	0.0004	0.0006	0.0006	0.0004

Tabla 6.2: Valores medios de $\|\hat{\phi}_1 - \phi_1\|^2$ cuando $\gamma_X(s, t) = (1/2)(1/2)^{0.9(s-t)^2}$, bajo C_0 .

N	Escala	$M = 50$	$M = 100$	$M = 250$	$M = 50$	$M = 100$	$M = 250$
		$\widehat{\phi}_{CR,1}$			$\widehat{\phi}_{CRS,1} (\rho = 10^{-5})$		
50	SD	1.4037	1.4231	1.3850	0.1252	0.1287	0.1267
	M -ESCALA	0.0447	0.0375	0.0324	0.0494	0.0444	0.0462
100	SD	1.3779	1.3308	1.3617	0.0790	0.0724	0.0704
	M -ESCALA	0.0248	0.0233	0.0226	0.0338	0.0386	0.0355
250	SD	0.8164	0.8397	0.8195	0.0423	0.0433	0.0372
	M -ESCALA	0.0158	0.0141	0.0148	0.0257	0.0297	0.0259
500	SD	0.3027	0.1935	0.2244	0.0327	0.0350	0.0325
	M -ESCALA	0.0113	0.0113	0.0096	0.0212	0.0311	0.0258
1000	SD	0.0901	0.0831	0.0791	0.0250	0.0277	0.0246
	M -ESCALA	0.0086	0.0087	0.0082	0.0213	0.0253	0.0211

Tabla 6.3: Valores medios de $\|\widehat{\phi}_1 - \phi_1\|^2$ cuando $\gamma_X(s, t) = (1/2)(1/2)^{0.9|s-t|}$ bajo C_1 .

N	Escala	$M = 50$	$M = 100$	$M = 250$	$M = 50$	$M = 100$	$M = 250$
		$\widehat{\phi}_{CR,1}$			$\widehat{\phi}_{CRS,1} (\rho = 10^{-5})$		
50	SD	1.3846	1.4136	1.3686	0.0993	0.1004	0.0989
	M -ESCALA	0.0183	0.0153	0.0108	0.0138	0.0133	0.0180
100	SD	1.3699	1.3265	1.3460	0.0542	0.0513	0.0507
	M -ESCALA	0.0068	0.0076	0.0070	0.0058	0.0104	0.0165
250	SD	0.7453	0.7611	0.7945	0.0257	0.0279	0.0223
	M -ESCALA	0.0036	0.0033	0.0030	0.0034	0.0043	0.0096
500	SD	0.2592	0.1353	0.1535	0.0163	0.0181	0.0172
	M -ESCALA	0.0017	0.0017	0.0018	0.0015	0.0024	0.0075
1000	SD	0.0496	0.0352	0.0340	0.0125	0.0164	0.0106
	M -ESCALA	0.0011	0.0001	0.0001	0.0001	0.0012	0.0037

Tabla 6.4: Valores medios de $\|\widehat{\phi}_1 - \phi_1\|^2$ cuando $\gamma_X(s, t) = (1/2)(1/2)^{0.9(s-t)^2}$, bajo C_1 .

N	Escala	$M = 50$	$M = 100$	$M = 250$	$M = 50$	$M = 100$	$M = 250$
		$\widehat{\phi}_{CRB,1}$			$\widehat{\phi}_{CRSB,1} (\rho = 10^{-5})$		
50	SD	0.0045	0.0061	0.0056	0.0028	0.0027	0.0027
50	M -escala	0.0174	0.0158	0.0137	0.0043	0.0039	0.0040
100	SD	0.0032	0.0039	0.0035	0.0027	0.0027	0.0026
100	M -escala	0.0083	0.0075	0.0088	0.0031	0.0030	0.0029
250	SD	0.0030	0.0028	0.0028	0.0027	0.0027	0.0026
250	M -escala	0.0045	0.0051	0.0046	0.0028	0.0027	0.0027
500	SD	0.0027	0.0027	0.0027	0.0027	0.0027	0.0026
500	M -escala	0.0033	0.0031	0.0035	0.0027	0.0027	0.0027
1000	SD	0.0027	0.0027	0.0027	0.0027	0.0027	0.0026
1000	M -escala	0.0029	0.0032	0.0028	0.0027	0.0027	0.0026

Tabla 6.5: Valores medios de $\|\widehat{\phi}_1 - \phi_1\|^2$ cuando $\gamma_X(s, t) = (1/2)(1/2)^{0.9|s-t|}$, bajo C_0 .

N	Escala	$M = 50$	$M = 100$	$M = 250$	$M = 50$	$M = 100$	$M = 250$
		$\hat{\phi}_{\text{CRB},1}$			$\hat{\phi}_{\text{CRSB},1} (\rho = 10^{-5})$		
50	SD	0.0033	0.0040	0.0038	0.0035	0.0040	0.0040
50	M -escala	0.0108	0.0098	0.0108	0.0108	0.0098	0.0107
100	SD	0.0019	0.0020	0.0020	0.0019	0.0019	0.0020
100	M -escala	0.0049	0.0044	0.0059	0.0049	0.0044	0.0059
250	SD	0.0010	0.0008	0.0007	0.0010	0.0008	0.0007
250	M -escala	0.0019	0.0025	0.0022	0.0019	0.0025	0.0023
500	SD	0.0004	0.0003	0.0003	0.0005	0.0003	0.0003
500	M -escala	0.0008	0.0012	0.0008	0.0008	0.0012	0.0008
1000	SD	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002
1000	M -escala	0.0006	0.0007	0.0004	0.0006	0.0006	0.0004

Tabla 6.6: Valores medios de $\|\hat{\phi}_1 - \phi_1\|^2$ cuando $\gamma_X(s, t) = (1/2)(1/2)^{0.9(s-t)^2}$, bajo C_0 .

N	Escala	$M = 50$	$M = 100$	$M = 250$	$M = 50$	$M = 100$	$M = 250$
		$\hat{\phi}_{\text{CRB},1}$			$\hat{\phi}_{\text{CRSB},1} (\rho = 10^{-5})$		
50	SD	1.3888	1.4231	1.3850	0.0184	0.0211	0.0246
50	M -escala	0.0237	0.0172	0.0150	0.0049	0.0050	0.0052
100	SD	1.3779	1.3286	1.3605	0.0135	0.0125	0.0157
100	M -escala	0.0120	0.0124	0.0100	0.0040	0.0036	0.0041
250	SD	0.8159	0.8397	0.8190	0.0068	0.0062	0.0109
250	M -escala	0.0078	0.0067	0.0057	0.0028	0.0028	0.0029
500	SD	0.3027	0.1935	0.2244	0.0038	0.0033	0.0048
500	M -escala	0.0043	0.0034	0.0042	0.0027	0.0027	0.0027
1000	SD	0.0901	0.0831	0.0791	0.0030	0.0028	0.0028
1000	M -escala	0.0034	0.0031	0.0035	0.0027	0.0027	0.0026

Tabla 6.7: Valores medios de $\|\hat{\phi}_1 - \phi_1\|^2$ cuando $\gamma_X(s, t) = (1/2)(1/2)^{0.9|s-t|}$, bajo C_1 .

N	Escala	$M = 50$	$M = 100$	$M = 250$	$M = 50$	$M = 100$	$M = 250$
		$\hat{\phi}_{\text{CRB},1}$			$\hat{\phi}_{\text{CRSB},1} (\rho = 10^{-5})$		
50	SD	1.3838	1.4136	1.3686	0.0167	0.0186	0.0225
50	M -escala	0.0175	0.0146	0.0110	0.0067	0.0051	0.0046
100	SD	1.3699	1.3258	1.3458	0.0123	0.0116	0.0139
100	M -escala	0.0067	0.0076	0.0073	0.0045	0.0030	0.0036
250	SD	0.7453	0.7611	0.7943	0.0060	0.0056	0.0097
250	M -escala	0.0036	0.0034	0.0029	0.0035	0.0026	0.0027
500	SD	0.2593	0.1352	0.1533	0.0035	0.0030	0.0040
500	M -escala	0.0017	0.0017	0.0017	0.0017	0.0026	0.0025
1000	SD	0.0496	0.0352	0.0340	0.0028	0.0026	0.0026
1000	M -escala	0.0011	0.00010	0.00010	0.0011	0.0024	0.0025

Tabla 6.8: Valores medios de $\|\hat{\phi}_1 - \phi_1\|^2$ cuando $\gamma_X(s, t) = (1/2)(1/2)^{0.9(s-t)^2}$ bajo C_1 .

Bibliografía

- [1] Bali, J. and Boente, G. (2009). Principal points and elliptical distributions from the multivariate setting to the functional case. *Statist. Probab. Lett.*, **79**, 1858-1865.
- [2] Bali, J., Boente, G., Tyler, D. E. and Wang, J. L. (2011). Robust Functional Principal Components: a projection-pursuit approach. *Annals of Statistics*, **39**, 2852-2882.
- [3] Beaton, A. E. and Tukey, J. W. (1974). The fitting of power series, meaning polynomials, illustrated on band-spectroscopic data. *Technometrics*, **16**, 147-185.
- [4] Benko, M.; Härdle, P. and Kneip, A. (2009). Common Functional Principal Components. *Ann. Statist.*, **37**, 1-34.
- [5] Boente, G., Rodriguez, D. and Sued, M. (2010). Inference under functional proportional and common principal components models. *Journal of Multivariate Analysis*, **101**, 464-475.
- [6] Boente, G. and Fraiman, R. (2000). Kernel-based functional principal components. *Statist. Probab. Lett.*, **48**, 335-345.
- [7] Boente G. L., Orellana L.C. (2001), A robust approach to common principal components. *Statistics in Genetics and in Environmental Sciences*. pp 117-46.
- [8] Boente G., Pires A., Rodrigues I. (2006) General projection-pursuit estimators for the common principal components model: influence functions and Monte Carlo study. *Journal of Multivariate Analysis*, **97**, 124-147.
- [9] Cantoni, E. and Ronchetti, E. (2001). Resistant selection of the smoothing parameter for smoothing splines. *Statistics and Computing*, **11**, 141-146.
- [10] Conway J.B. (2000). A Course in Functional Analysis. Springer.
- [11] Croux, C. (1994). Efficient high-breakdown M -estimators of scale. *Statist. Prob. Letters*, **19**, 371-379.
- [12] Croux, C., Filzmoser, P. and Oliveira M.R. (2007). Algorithms for Projection-Pursuit Robust Principal Component Analysis. *Chemometrics and Intelligent Laboratory Systems*, **87**, 218-225.
- [13] Croux, C. and 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.
- [14] Croux, C. and Ruiz-Gazen, A. (2005). High Breakdown Estimators for Principal Components: the Projection-Pursuit Approach Revisited. *J. Multivar. Anal.*, **95**, 206-226.

- [15] Cuevas, A., Febrero, M. and Fraiman, R. (2007). Robust estimation and classification for functional data via projection-based depth notions. *Comp. Statist.*, **22**, 481 - 96.
- [16] Cui, H., He, X. and Ng, K. W. (2003). Asymptotic Distribution of Principal Components Based on Robust Dispersions. *Biometrika*, **90**, 953-966.
- [17] Dauxois, J., Pousse, A. and Romain, Y. (1982). Asymptotic theory for the principal component analysis of a vector random function: Some applications to statistical inference. *J. Multivar. Anal.*, **12**, 136-154.
- [18] Fernholz, L. T. (1983). *von Mises calculus for statistical functionals*. Lecture Note in Statistics 19, Springer-Verlag, New York.
- [19] Flett, T. (1980). *Differential Analysis*. Cambridge University Press. Cambridge.
- [20] Flury, B. (1984) Common principal components in k groups. *Journal of the American Statistical Association* **79**, 892-898.
- [21] Fraiman, R. and Muñiz, G. (2001). Trimmed means for functional data. *Test*, **10**, 419-40.
- [22] Gervini, D. (2006). Free-knot spline smoothing for functional data. *J. Roy. Statist. Soc. Ser. B*, **68**, 671-87.
- [23] Gervini, D. (2008). Robust functional estimation using the spatial median and spherical principal components. *Biometrika*, **95**, 587-600.
- [24] Hall, P. and Hosseini-Nasab, M. (2006). On properties of functional principal components analysis. *J. Roy. Statist. Soc. Ser. B*, **68**, 109-26.
- [25] Hall, P., Müller, H.-G. and Wang, J.-L. (2006). Properties of principal component methods for functional and longitudinal data analysis. *Ann. Statist.*, **34**, 1493-517.
- [26] Hampel, F. (1971). A general qualitative definition of robustness. *Ann. Math. Statist.*, **42**, 1887-1896.
- [27] Huber, P.J. (1981). *Robust Statistics*. John Wiley, New York.
- [28] Huber, P.J. (1985). Projection pursuit. *Annals of Statistics*, **13**, 435-525.
- [29] Hubert, M. and Vandervieren, E. (2008). An adjusted boxplot for skewed distributions. *Computational Statistics and Data Analysis*, **52**, 5186-5201.
- [30] Hyndman, R. J. and S. Ullah (2007). Robust forecasting of mortality and fertility rates: A functional data approach. *Comp. Statist. Data Anal.*, **51**, 4942-4956.
- [31] Locantore, N., Marron, J. S., Simpson, D. G., Tripoli, N., Zhang, J. T. and Cohen, K. L. (1999). Robust principal components for functional data (with Discussion). *Test*, **8**, 1-73.
- [32] Li, G. and Chen, Z. (1985). Projection-Pursuit Approach to Robust Dispersion Matrices and Principal Components: Primary Theory and Monte Carlo. *J. Amer. Statist. Assoc.*, **80**, 759-766.
- [33] López-Pintado, S. and Romo, J. (2007). Depth-based inference for functional data. *Comp. Statist. Data Anal.*, **51**, 4957-68.
- [34] Malfait, N. and Ramsay, J. O. (2003). The historical functional linear model. *Can. J. Statist.*, **31**, 115-128.
- [35] Maronna, R., Martin, R. D. and Yohai, V. (2006). *Robust Statistics: Theory and Methods*, John Wiley & Sons.

- [36] Martin, R. D. and Zamar, R. (1993). Bias robust estimation of scale. *Ann. Statist.*, **21**, 991-1017.
- [37] Pezzulli, S. D. and Silverman, B. W. (1993). Some properties of smoothed principal components analysis for functional data. *Comput. Statist.*, **8**, 1-16.
- [38] Ramsay, J. O. and Silverman, B. W. (2005). *Functional Data Analysis*, 2nd ed. New York: Springer-Verlag.
- [39] Rice, J. and Silverman, B. W. (1991). Estimating the mean and covariance structure nonparametrically when the data are curves. *J. Roy. Statist. Soc. Ser. B*, **53**, 233-243.
- [40] Rodrigues, I. M. (2003). *Métodos Robustos em Análise de Componentes Principais Comuns*. Unpublished PhD Thesis (in portuguese), Universidade Técnica de Lisboa. Available on <http://www.math.ist.utl.pt/~apires/phd.html>.
- [41] Sawant, P., Billor, N. and Shin, H. (2011) Functional outlier detection with robust functional principal component analysis. *Computational Statistics*, **27**, 83-102.
- [42] Silverman, B. W. (1996). Smoothed functional principal components analysis by choice of norm. *Ann. Statist.*, **24**, 1-24.
- [43] Wang, F. and Scott, D. (1994). The L_1 method for robust nonparametric regression. *J. Amer. Stat. Assoc.*, **89**, 65-76.
- [44] Yao, F. and Lee, T. C. M. (2006). Penalized spline models for functional principal component analysis. *J. R. Statist. Soc. Ser. B*, **68**, 3-25.