

Tesis Doctoral

Métodos robustos en correlación canónica funcional

Alvarez, Agustín

2017-03-22

Este documento forma parte de la colección de tesis doctorales y de maestría de la Biblioteca Central Dr. Luis Federico Leloir, disponible en digital.bl.fcen.uba.ar. 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, available in digital.bl.fcen.uba.ar. It should be used accompanied by the corresponding citation acknowledging the source.

Cita tipo APA:

Alvarez, Agustín. (2017-03-22). Métodos robustos en correlación canónica funcional. Facultad de Ciencias Exactas y Naturales. Universidad de Buenos Aires.

Cita tipo Chicago:

Alvarez, Agustín. "Métodos robustos en correlación canónica funcional". Facultad de Ciencias Exactas y Naturales. Universidad de Buenos Aires. 2017-03-22.

EXACTAS UBA

Facultad de Ciencias Exactas y Naturales



UBA

Universidad de Buenos Aires



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

Métodos robustos en correlación canónica funcional

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

Agustín Alvarez

Directora de tesis: Dra. Graciela Lina Boente Boente

Directora asistente: Dra. Nadia Laura Kudraszow

Lugar de trabajo: Instituto de cálculo, FCEN-UBA

Buenos Aires, Enero de 2017

Fecha de defensa:

Agradecimientos

La elaboración de esta tesis ha sido posible gracias a mucha gente que me rodea y con la que he compartido hermosos momentos estos años. A todos ellos quiero agradecer por ayudarme de una u otra manera en su realización.

En primer lugar, agradezco enormemente a Graciela por aceptar dirigir esta tesis, estar siempre que fue necesario para darme una mano no sólo con la tesis sino también con otras cuestiones académicas y por la paciencia que me tuvo. También, por incorporar a Nadia en la dirección de este trabajo. A Nadia por aceptar dirigirme, por mostrar siempre una gran predisposición para juntarnos y la buena onda con la que pudimos trabajar. A ambas por todo lo que aprendí de ellas.

A Daniela y Mariela que me recibieron con los brazos abiertos cuando me acerqué a ellas y les comenté mi voluntad de hacer el doctorado. Ellas me abrieron las puertas para comenzar este camino y luego me acompañaron en el proceso siempre con la mejor predisposición para ayudarme con todo. A mis compañeros de oficina con los cuales he compartido hermosos momentos y también siempre están ahí con gran generosidad ante cualquier solicitud. A todos los del grupo de estadística, que creo que ninguno se ha salvado de que le pida o le consulte algo en este período. A todos y todas los que integran el Instituto de Cálculo porque es un placer compartir con ellos el lugar de trabajo y los almuerzos. A mis amigos de matemática.

A mi papá y mi mamá, que siempre están presentes y me apoyan en todo lo que emprenda. A mi familia en general y mis amigos de la vida que están en todos los momentos, ayudan a pasar con alegría momentos difíciles y siempre dan buenos consejos. Gracias!!!

Por último quiero agradecer a Javiera, mi amor y compañera de vida, con quien aprendo sobre las cosas más importantes de la vida todos los días. Me ayuda a mejorar como persona y me aguanta, que no es poca cosa.

Resumen

Métodos Robustos en Correlación Canónica Funcional

En diversas aplicaciones, resulta de interés poder medir la asociación entre dos características que se observan sobre los individuos de una población. Por otra parte, muchas veces estas características 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 utilizado para lograr este objetivo es el análisis de correlación canónica funcional. Leurgans *et al.* (1993) prueban que la extensión natural de los estimadores utilizados en el caso multivariado al funcional no resulta consistente para la primera correlación canónica. Por esta razón, dichos autores proponen estimadores que penalizan la rugosidad de las direcciones y prueban que resultan consistentes. Por otro lado, He *et al.* (2003) dan condiciones que garantizan la buena definición de las correlaciones y direcciones canónicas para elementos aleatorios $(X, Y)^T$ a valores en el espacio de Hilbert $\mathcal{L}^2(0, 1) \times \mathcal{L}^2(0, 1)$.

Los trabajos mencionados consideran como medida de asociación la correlación de Pearson. Sin embargo, es bien sabido, que esta medida es muy sensible a la presencia de datos atípicos. Con el fin de lidiar con este problema, en el caso multivariado, Branco *et al.* (2005) y Alfons *et al.* (2016) proponen estimadores utilizando un enfoque de *projection-pursuit*, es decir, maximizando funcionales de asociación robustos de proyecciones de los datos.

En esta tesis, consideramos espacios de Hilbert separables \mathcal{H}_1 y \mathcal{H}_2 y elementos aleatorios $(X, Y)^T \in \mathcal{H} = \mathcal{H}_1 \times \mathcal{H}_2$ y medidas de asociación bivariadas robustas. Para mostrar que la extensión natural del caso multivariado al caso funcional falla cuando se utiliza la extensión de los estimadores de *projection-pursuit* de Branco *et al.* (2005) al caso funcional, extendemos el resultado de Leurgans *et al.* (1993) al caso de medidas de asociación generales. Por otra parte, para proponer estimadores robustos y consistentes consideramos un enfoque que combina *projection-pursuit* con el método de Sieves que aproxima el espacio infinito-

dimensional \mathcal{H} por subespacios de dimensión finita que crecen con el tamaño de muestra. Bajo condiciones de regularidad, obtenemos que los estimadores de las primeras direcciones y de la primera correlación canónica son consistentes al funcional asociado. Para identificar que representa dicho funcional, se muestra la consistencia Fisher de éstos cuando se utilizan medidas de asociación que cumplan ciertas condiciones. En particular, las medidas de correlación robustas de uso habitual permiten obtener estimadores consistentes a la cantidad de interés en el caso de procesos Gaussianos o de procesos elípticos. Mediante un estudio de simulación, se compara el comportamiento de los estimadores robustos con los estimadores basados en la correlación de Pearson, para muestras finitas Gaussianas y contaminadas. Se proponen asimismo un procedimiento de convalidación cruzada robusta para elegir las dimensiones de los subespacios aproximantes y métodos para la detección de datos atípicos. Finalmente, se ilustra en un conjunto de datos reales las ventajas de los estimadores robustos ya que permiten identificar los datos atípicos y proveen estimaciones confiables.

Palabras Claves: Correlación canónica; Datos funcionales; Espacios aproximantes; Robustez

Abstract

Robust Inference in Functional Canonical Correlation

In many applications, measuring the association between two individual features is an important issue. Besides, 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 working with the discretized trajectories as multivariate data samples. A technique widely used to study the dependence between the two functions recorded for a sample of subjects is the functional canonical correlation analysis. Leurgans *et al.* (1993) demonstrated the need for regularization in functional canonical correlation analysis, since the natural extension of the multivariate estimators are not consistent. Leurgans *et al.* (1993) also derived the consistency of the estimators obtained penalizing the roughness. These results are complemented with those obtained in He *et al.* (2003) who provide conditions ensuring the existence and proper definition of the canonical directions and correlations when the random element belongs to $L^2(0, 1) \times L^2(0, 1)$.

The aforementioned papers consider as association measure the Pearson correlation. However, it is well known that the Pearson correlation is not robust, being sensitive to outliers and this sensitivity is inherited by the canonical directions and correlations. In the multivariate case, a robust approach was given by Branco *et al.* (2005) and Alfons *et al.* (2016) who proposed estimators using *projection-pursuit*, that is, maximizing a robust bivariate association measure over the projected data.

In this thesis, we consider separable Hilbert spaces \mathcal{H}_1 and \mathcal{H}_2 and random elements $(X, Y) \in \mathcal{H} = \mathcal{H}_1 \times \mathcal{H}_2$ as well as robust bivariate measures of association. To show that the natural extension from the multivariate case to the functional case fails when using the *projection-pursuit* estimators of Branco *et al.* (2005), we generalize the result of Leurgans *et al.* (1993) to the case of general measures of association. On the other hand, to obtain robust and consistent estimators in a functional framework, we consider an approach that com-

bines *projection–pursuit* with Sieves which approximate the infinite–dimensional space \mathcal{H} by means of finite-dimensional linear spaces growing with the sample size. Under regularity conditions, consistency results for the first canonical correlation and directions are derived. An important point to highlight is what the functional related to the proposed estimators represent, at least in some situations. For that purpose, Fisher consistency is obtained when the bivariate measure of association may be transformed to a Fisher–consistent bivariate association functional. In particular, the usual robust correlations allow to obtain consistent estimators of the quantities of interest when $(X, Y)^T$ is a Gaussian or an elliptical process. Through a simulation study, we compare the performance of the robust proposals with the estimators based on the Pearson correlation, for clean and contaminated data. Furthermore, a robust cross–validation procedure is also proposed to select the dimension of the approximating linear spaces and outlier detection rules are studied. Finally, we illustrate on a real data set the advantages of the robust estimators which allow identify outliers and provide reliable estimates.

Keywords: Canonical correlation; Functional data; Robustness; Sieves

Índice general

1.. <i>Introducción</i>	8
2.. <i>Nociones previas</i>	14
2.1. Elementos aleatorios y operador de covarianza	14
2.2. Elipticidad	16
2.3. Medidas de asociación	19
3.. <i>¿Por qué suavizar?</i>	36
3.1. Introducción	37
3.2. El estimador natural falla en el análisis de correlación canónica funcional	42
3.3. Condiciones para la consistencia de los estimadores basados en Sieves	47
3.4. Datos de alta dimensión	49
4.. <i>Consistencia de los estimadores</i>	50
4.1. Introducción	50
4.2. Supuestos	54
4.3. Consistencia	61

4.4. Apéndice 1	66
4.5. Apéndice 2	75
5.. <i>Estudio de Monte Carlo</i>	79
5.1. Modelos de la simulación	79
5.2. Estimadores utilizados y algoritmos para su cálculo	82
5.3. Selección de parámetros	87
5.4. Resultados de la simulación	88
5.5. Detección de datos atípicos	125
5.6. Conclusiones generales de la simulación	133
5.7. Apéndice	140
6.. <i>Aplicación a un conjunto de datos reales: Escritura de letras</i>	144
6.1. Introducción	144
6.2. Métodos utilizados y resultados	145
6.3. Conclusiones	160

1. INTRODUCCIÓN

El método de correlaciones canónicas introducido en Hotelling (1936) se utiliza para relacionar dos grupos de variables. En general, esta metodología se utiliza cuando un grupo de variables aleatorias, definidas en un mismo espacio, puede dividirse en dos grupos homogéneos y se desea identificar y cuantificar la relación entre ambos grupos. Clásicamente el análisis de correlación canónica consiste en buscar las combinaciones lineales de cada subconjunto de variables que maximizan la correlación. Para ser más precisos, dados dos vectores aleatorios $\mathbf{X} \in \mathbb{R}^p$ e $\mathbf{Y} \in \mathbb{R}^q$ definidos sobre el mismo espacio de probabilidad, se buscan direcciones $\mathbf{u}_1 \in \mathbb{R}^p$ y $\mathbf{v}_1 \in \mathbb{R}^q$ de manera que la correlación entre las proyecciones $U_1 = \mathbf{u}_1^T \mathbf{X}$ y $V_1 = \mathbf{v}_1^T \mathbf{Y}$ sea máxima. Las subsiguientes direcciones canónicas $(\mathbf{u}_j, \mathbf{v}_j)_{j=2}^d$, donde $d = \min(p, q)$, se encuentran de manera análoga maximizando la correlación de las proyecciones con la restricción de que las variables $U_j = \mathbf{u}_j^T \mathbf{X}$ y $V_j = \mathbf{v}_j^T \mathbf{Y}$ sean no correlacionadas con las ya encontradas. Para poder aspirar a que haya unicidad salvo cambio de signo en las direcciones canónicas se agrega habitualmente la restricción de que las varianzas de las proyecciones sea 1. Los subsecuentes valores de las correlaciones entre proyecciones maximizantes, que indicaremos como $\rho_j = \rho_{\text{CL}}(U_j, V_j)$, con $\rho_{\text{CL}}(U, V)$ la correlación de Pearson entre las variables U y V , se denominan correlaciones canónicas y son una medida de asociación entre los vectores \mathbf{X} e \mathbf{Y} . El método de correlaciones canónicas puede ser visto como un método para reducir la dimensión de los vectores \mathbf{X} e \mathbf{Y} cuando lo que nos interesa es estudiar la estructura de correlación entre estos vectores. A este enfoque de definir el análisis de correlación canónica a través de la maximización de correlaciones de proyecciones se lo denomina “Projection

Pursuit”.

Llamando $\Sigma_{\mathbf{ZZ}}$ a la matriz de covarianza del vector $\mathbf{Z} = (\mathbf{X}^T, \mathbf{Y}^T)^T$ y efectuando una partición de la misma según la partición del vector \mathbf{Z} , obtenemos la descomposición

$$\Sigma_{\mathbf{ZZ}} = \begin{pmatrix} \Sigma_{\mathbf{XX}} & \Sigma_{\mathbf{XY}} \\ \Sigma_{\mathbf{YX}} & \Sigma_{\mathbf{YY}} \end{pmatrix}.$$

Cuando las matrices $\Sigma_{\mathbf{XX}}$ y $\Sigma_{\mathbf{YY}}$ son no singulares, el análisis de correlación canónica tiene una resolución sencilla (ver Kshirsagar 1972). Efectivamente, en este caso, las direcciones canónicas $\mathbf{u}_1, \dots, \mathbf{u}_d$ y $\mathbf{v}_1, \dots, \mathbf{v}_d$ son, respectivamente, los autovectores de las matrices

$$\Sigma_A = \Sigma_{\mathbf{XX}}^{-1} \Sigma_{\mathbf{XY}} \Sigma_{\mathbf{YY}}^{-1} \Sigma_{\mathbf{YX}} \quad \text{y} \quad \Sigma_B = \Sigma_{\mathbf{YY}}^{-1} \Sigma_{\mathbf{YX}} \Sigma_{\mathbf{XX}}^{-1} \Sigma_{\mathbf{XY}},$$

asociados a los autovalores $\rho_1^2 \geq \dots \geq \rho_d^2$. Las matrices Σ_A y Σ_B tienen los mismos autovalores ρ_j^2 que son los cuadrados de las correlaciones canónicas. Usando esta propiedad de las correlaciones y direcciones canónicas, los estimadores clásicos de las correlaciones y direcciones canónicas se obtienen como los autovalores y autovectores de las versiones muestrales de Σ_A y Σ_B , que indicaremos por $\widehat{\Sigma}_A$ y $\widehat{\Sigma}_B$. Dada una muestra $\mathbf{Z}_1, \dots, \mathbf{Z}_n$, para definir los estimadores $\widehat{\Sigma}_A$ y $\widehat{\Sigma}_B$ se reemplaza en la definición de Σ_A y Σ_B , las matrices de covarianza por sus versiones muestrales, es decir, las obtenidas a partir de la partición de la matriz $\widehat{\Sigma}_{\mathbf{ZZ}} = \sum_{i=1}^n (\mathbf{Z}_i - \bar{\mathbf{Z}})(\mathbf{Z}_i - \bar{\mathbf{Z}})^T / n$. Anderson (1999) derivó la distribución asintótica de los estimadores de las direcciones y correlaciones canónicas obtenidos de esta forma suponiendo que las observaciones tienen distribución normal. Es bien sabido que este procedimiento es óptimo bajo normalidad, sin embargo, los estimadores resultantes son sensibles a observaciones atípicas debido a la sensibilidad de la matriz de covarianza muestral a un pequeño porcentaje de datos atípicos. Por esta razón, en muchas aplicaciones, cuando es de interés obtener estimadores más fiables es deseable proveer estimadores robustos de las direcciones y correlaciones canónicas.

Un primer enfoque fue dado por Taskinen *et al.* (2006) mediante un procedimiento *plug-in*. Esta propuesta consiste en dar estimadores robustos de la matriz de covarianza de \mathbf{Z}

y a partir de ella, obtener estimadores robustos $\tilde{\Sigma}_A$ y $\tilde{\Sigma}_B$ de Σ_A y Σ_B , respectivamente. Los autovectores y autovalores de $\tilde{\Sigma}_A$ y $\tilde{\Sigma}_B$ permiten definir estimadores de las direcciones y correlaciones canónicas robustos. Taskinen *et al.* (2006) obtuvieron las funciones de influencia para los funcionales asociados a este tipo de estimadores y a partir de estas su eficiencia asintótica.

Otra alternativa para enfrentar el problema de la sensibilidad a datos atípicos es utilizar el procedimiento de projection pursuit para el análisis de correlación canónica. Alfons *et al.* (2016), utilizan distintas medidas de asociación robustas para definir estimadores robustos de las correlaciones y direcciones canónicas, mientras que Jin y Cui (2010) obtienen su distribución asintótica.

Suponiendo que $\mathbb{E}(\mathbf{X}) = \mathbb{E}(\mathbf{Y}) = 0$, en Seber (2004) puede verse un enfoque distinto al problema planteando un problema de predicción. Más precisamente, supongamos que, para $1 \leq k \leq d$, nos interesa encontrar los vectores k -dimensionales $\mathbf{U} = (\mathbf{u}_1^T \mathbf{X}, \dots, \mathbf{u}_k^T \mathbf{X})^T$ y $\mathbf{V} = (\mathbf{v}_1^T \mathbf{Y}, \dots, \mathbf{v}_k^T \mathbf{Y})^T$ que minimizan $\mathbb{E}(\|\mathbf{U} - \mathbf{V}\|^2)$ con la restricción que $\text{VAR}(\mathbf{U}) = \text{VAR}(\mathbf{V}) = \mathbf{I}_k$, donde con VAR nos referimos a la matriz de covarianza e \mathbf{I}_k denota la matriz identidad de $\mathbb{R}^{k \times k}$. La solución a este problema son los vectores canónicas. Siguiendo este enfoque, Adrover y Donato (2015) introducen una familia de estimadores robustos reemplazando la pérdida cuadrática por una M -escala para definir S -estimadores.

Estimadores para las correlaciones canónicas han sido ampliamente estudiados para el caso en que p y q son fijos. Recientemente, ha resultado de interés considerar el caso en que q es fijo y p crece con n , es decir las observaciones en \mathbf{X} corresponden a lo que se denomina datos de alta dimensión. Resultados asintóticos en esta situación pueden verse en Fujikoshi *et al.* (2011). También debido a avances tecnológicos, procesos estocásticos $\{Z(t) = (X(t), Y(t))^T\}_{t \in \mathcal{I}}$ donde \mathcal{I} es un intervalo de la recta pueden ser medidos y guardados en grillas cada vez más finas. Sin embargo, no parece adecuado considerar a dichas observaciones discretizadas como datos multivariados ya que se pierde información del proceso relacionada

con la suavidad del mismo. Por ejemplo, si las trayectorias son continuas, es de esperar que haya alta correlación entre $X(t_1)$ y $X(t_2)$ cuando t_1 y t_2 son cercanos. Supongamos ahora que, en lugar de observar vectores finito dimensionales, cada observación corresponda a un par de funciones $X_i(t)$ e $Y_i(t)$, $1 \leq i \leq n$, $t \in \mathcal{I}$, donde \mathcal{I} es un intervalo de la recta. Más aún, supongamos que $(X_i, Y_i)^T$ tienen la misma distribución que $(X, Y)^T$ donde $X, Y \in L^2(\mathcal{I})$. Podemos pensar que tanto X como Y son vectores aleatorios infinito dimensionales que toman valores en el espacio de Hilbert infinito dimensional $L^2(\mathcal{I})$. En tal caso, X e Y se denominan elementos aleatorios. En Leurgans *et al.* (1993) se estudia el problema de estimar la primera correlación y las primeras direcciones canónicas de elementos aleatorios X e Y en $L^2(\mathcal{I}_1)$ y $L^2(\mathcal{I}_2)$, respectivamente. Estas direcciones y correlaciones canónicas poblacionales se definen de manera análoga a las del caso multivariado. Sin embargo, estos autores muestran como la extensión natural del estimador de la primera correlación canónica multivariada no es en general consistente para el caso en que el elemento aleatorio $(X, Y)^T$ es Gaussiano. Surge entonces la necesidad de penalizar la función objetivo de modo a proveer estimadores suavizados que resulten consistentes. Leurgans *et al.* (1993) proponen modificar la medida de asociación usual basada en la covarianza muestral agregando un término a la varianza muestral de las proyecciones que penalizan la rugosidad de la dirección. De esta forma obtienen estimadores que resultan consistentes bajo condiciones débiles. He *et al.* (2003) dan condiciones sobre el elemento aleatorio $(X, Y)^T$ en $L^2(\mathcal{I}_1) \times L^2(\mathcal{I}_2)$ para que sus direcciones y correlaciones canónicas poblacionales estén bien definidas. Los trabajos mencionados para datos funcionales dan resultados cuando la medida de asociación utilizada es la correlación de Pearson. Sin embargo, es bien sabido que esta medida de asociación es muy sensible a la presencia de datos atípicos.

El objetivo de esta tesis es introducir estimadores de las direcciones y correlaciones canónicas que resulten resistentes cuando la muestra contiene un porcentaje de datos atípicos. Consideraremos un elemento aleatorio $Z = (X, Y)^T : \Omega \rightarrow \mathcal{H}_1 \times \mathcal{H}_2$ donde \mathcal{H}_i son espa-

cios de Hilbert separables y, en general, de dimensión infinita. Nuestra definición se basa en el enfoque projection pursuit para las correlaciones canónicas y utilizaremos funcionales de correlación o asociación preferentemente robustos. Para evitar los problemas de la alta dimensión así como medidas de rugosidad, proponemos estimadores basados en un método de Sieves. Los estimadores se definen como las direcciones que maximizan la asociación de las proyecciones, donde estas direcciones se buscan en subespacios de dimensiones finitas p_n y q_n de \mathcal{H}_1 y \mathcal{H}_2 , respectivamente. Dichas dimensiones van creciendo con el tamaño de muestra n .

La tesis está diagramada de la siguiente manera:

En el Capítulo 2, se presenta una breve introducción a elementos aleatorios, operador de covarianza y elipticidad. A su vez, se recuerdan las distintas medidas de asociación conocidas en la literatura para vectores bivariados analizando en cada caso su consistencia Fisher.

En el Capítulo 3 se define el problema de análisis de correlación canónica funcional. Se extiende la definición de las correlaciones y direcciones canónicas poblacionales para el caso de elementos aleatorios. Se generaliza el resultado dado en Leurgans *et al.* (1993) sobre la problemática de la extensión natural en la estimación de la primera correlación canónica incluyendo nuevos funcionales de correlación además del de Pearson y permitiendo que el elemento aleatorio pertenezca a una familia más general que contiene a los elementos Gaussianos. Se definen los estimadores de Sieves y se prueba que no resultan consistentes si las dimensiones de los subespacios aproximantes son mayores o iguales al tamaño de muestra. Se relaciona este hecho con la inconsistencia de los estimadores basados en Sieves para datos de alta dimensión.

En el Capítulo 4, se muestra que los estimadores propuestos para la primera correlación y las primeras direcciones canónicas son consistentes. Se demuestra además la consistencia Fisher de las distintas propuestas según las características del funcional de asociación que se

utilice.

En el Capítulo 5, se describen los resultados de un estudio de Montecarlo diseñado para estudiar el comportamiento de las propuestas dadas y de los estimadores clásicos para conjuntos de datos con y sin observaciones atípicas. En dicho estudio, se pudo observar como los estimadores basados en funcionales de asociación robustos tienen buenas propiedades de robustez ya que siguen estimando los valores deseados bajo contaminación, no ocurriendo lo mismo con los estimadores clásicamente utilizados. Este primer estudio fue realizado cuando las dimensiones de los subespacios son fijas y no adaptivas. Por esta razón, se complementó el estudio anterior, realizando otro estudio de Montecarlo en el que la dimensión de los subespacios aproximantes se elije mediante convalidación cruzada. Finalmente, en este capítulo, proponemos además tres métodos para la detección de datos atípicos.

En el Capítulo 6 aplicamos los estimadores de Sieves y los métodos para la detección de datos atípicos presentados en el Capítulo 5 a un ejemplo de datos reales a fin de detectar posibles observaciones atípicas que de otra forma no se hubieran identificado.

2. NOCIONES PREVIAS

Este capítulo se divide en tres secciones. En la Sección 2.1 se introducen las nociones de elemento aleatorio y de operador de covarianza. La Sección 2.2 versa sobre elipticidad. Recordaremos primero la definición para vectores aleatorios en \mathbb{R}^p y luego, la extensión de esta noción para elementos aleatorios en un espacio de Hilbert. Finalmente, en la Sección 2.3 se introducen distintas medidas de asociación. Varias de estas medidas son utilizadas luego en el Capítulo 5 para el estudio de Montecarlo.

2.1. Elementos aleatorios y operador de covarianza

En esta sección definimos el concepto de elemento aleatorio a valores en un espacio de Hilbert separable \mathcal{H} con producto interno $\langle \cdot, \cdot \rangle_{\mathcal{H}}$.

Definición 2.1.1. Sea $\Omega = (\Omega, \mathcal{F}, \mathbb{P})$ un espacio de probabilidad, \mathcal{H} un espacio de Hilbert separable y \mathcal{B} la σ -álgebra generada por los abiertos de \mathcal{H} . Diremos que X es un elemento aleatorio en \mathcal{H} si $X : \Omega \rightarrow \mathcal{H}$ es medible Borel, es decir, si $X^{-1}(B) \in \mathcal{F} \forall B \in \mathcal{B}$.

Un elemento aleatorio X en \mathcal{H} induce una probabilidad $P = \mathbb{P}_X$ en $(\mathcal{H}, \mathcal{B})$ definida para $B \in \mathcal{B}$ como $P(B) = \mathbb{P}_X(B) = \mathbb{P}(X^{-1}(B))$.

Observación 2.1.1. Si X es un elemento aleatorio en \mathcal{H} , $\|X\|_{\mathcal{H}} = \langle X, X \rangle_{\mathcal{H}}$ resulta una variable aleatoria. Por otra parte, dado $h \in \mathcal{H}$ fijo cualquiera, $\langle h, X \rangle_{\mathcal{H}}$ también es una variable aleatoria.

Definición 2.1.2. Decimos que el elemento aleatorio X tiene primer momento finito si $\mathbb{E}(\|X\|_{\mathcal{H}}) < \infty$. En tal caso, para todo $h \in \mathcal{H}$ la variable aleatoria $\langle h, X \rangle_{\mathcal{H}}$ tiene esperanza finita y se define la esperanza de X como el único elemento $\mu \in \mathcal{H}$ que cumple

$$\mathbb{E}(\langle h, X \rangle_{\mathcal{H}}) = \langle h, \mu \rangle_{\mathcal{H}} \quad \forall h \in \mathcal{H}.$$

El elemento μ existe debido al Teorema de Representación de Riesz ya que la aplicación $h \mapsto \mathbb{E}(\langle h, X \rangle_{\mathcal{H}})$ resulta lineal y continua. Si X tiene segundo momento finito, es decir, si $\mathbb{E}(\|X\|_{\mathcal{H}}^2) < \infty$, se define la función de covarianza $\mathcal{C}_{XX} : \mathcal{H} \times \mathcal{H} \rightarrow \mathbb{R}$ como

$$\mathcal{C}_{XX}(u, v) = \text{COV}(\langle u, X \rangle_{\mathcal{H}}, \langle v, X \rangle_{\mathcal{H}}) = \mathbb{E}(\langle u, X - \mu \rangle_{\mathcal{H}} \langle v, X - \mu \rangle_{\mathcal{H}}). \quad (2.1)$$

\mathcal{C}_{XX} resulta bilineal y continua en ambas variables. Nuevamente por el Teorema de Representación de Riesz, existe un único elemento $\Gamma_{XX}(v) \in \mathcal{H}$ tal que

$$\mathcal{C}_{XX}(u, v) = \langle u, \Gamma_{XX}(v) \rangle_{\mathcal{H}} \quad \forall u \in \mathcal{H}. \quad (2.2)$$

A Γ_{XX} se lo denomina operador de covarianza de X y es un operador autoadjunto, semi-definido positivo y compacto (ver Kaphle, 2011).

Supongamos tener un elemento aleatorio $Z = (X, Y)^T : \Omega \rightarrow \mathcal{H} = \mathcal{H}_1 \times \mathcal{H}_2$ donde \mathcal{H}_i son espacios de Hilbert separables y consideremos en \mathcal{H} el producto interno $\langle (u_1, v_1), (u_2, v_2) \rangle_{\mathcal{H}} = \langle u_1, u_2 \rangle_{\mathcal{H}_1} + \langle v_1, v_2 \rangle_{\mathcal{H}_2}$. Si $\mathbb{E}\|Z\|_{\mathcal{H}}^2 < \infty$, de manera análoga a como se definió \mathcal{C}_{XX} en (2.1), se definen las funciones \mathcal{C}_{XY} y \mathcal{C}_{YX} como

$$\mathcal{C}_{XY}(u, v) = \text{COV}(\langle u, X \rangle_{\mathcal{H}_1}, \langle v, Y \rangle_{\mathcal{H}_2}) = \mathbb{E}(\langle u, X - \mu_X \rangle_{\mathcal{H}_1} \langle v, Y - \mu_Y \rangle_{\mathcal{H}_2}),$$

$$\mathcal{C}_{YX}(v, u) = \text{COV}(\langle v, Y \rangle_{\mathcal{H}_2}, \langle u, X \rangle_{\mathcal{H}_1}) = \mathbb{E}(\langle v, Y - \mu_Y \rangle_{\mathcal{H}_2} \langle u, X - \mu_X \rangle_{\mathcal{H}_1}).$$

Por otra parte, así como \mathcal{C}_{XX} permite definir el operador de covarianza Γ_{XX} a través de (2.2), utilizando nuevamente el Teorema de Representación de Riesz se pueden definir operadores $\Gamma_{XY} : \mathcal{H}_2 \rightarrow \mathcal{H}_1$ y $\Gamma_{YX} : \mathcal{H}_1 \rightarrow \mathcal{H}_2$ a partir de \mathcal{C}_{XY} y \mathcal{C}_{YX} . Efectivamente, $\Gamma_{XY}(v)$ y $\Gamma_{YX}(u)$

verifican

$$\begin{aligned}\mathcal{C}_{XY}(u, v) &= \langle u, \Gamma_{XY}(v) \rangle_{\mathcal{H}_1} & \forall u \in \mathcal{H}_1, \\ \mathcal{C}_{YX}(v, u) &= \langle v, \Gamma_{YX}(u) \rangle_{\mathcal{H}_2} & \forall v \in \mathcal{H}_2.\end{aligned}$$

Resulta sencillo verificar que $\Gamma_{ZZ} : \mathcal{H} \rightarrow \mathcal{H}$ se puede descomponer como

$$\Gamma_{ZZ}(u, v) = (\Gamma_{XX}(u) + \Gamma_{XY}(v), \Gamma_{YX}(u) + \Gamma_{YY}(v)),$$

por lo que resulta razonable escribir matricialmente a Γ_{ZZ} como

$$\Gamma_{ZZ}(u, v) = \begin{pmatrix} \Gamma_{XX} & \Gamma_{XY} \\ \Gamma_{YX} & \Gamma_{YY} \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix}. \quad (2.3)$$

2.2. Elipticidad

Tanto en el caso multivariado como en el funcional, la familia de distribuciones elípticas resulta una generalización de las distribuciones normales o Gaussianas y es el ámbito adecuado donde estudiar propiedades de robustez de estimadores. A continuación, recordamos las definiciones y principales propiedades que pueden verse en Bilodeau y Brenner (1999).

Definición 2.2.1. Se dice que un vector aleatorio multivariado $\mathbf{Z} : \Omega \rightarrow \mathbb{R}^d$ tiene distribución esférica d -dimensional si su distribución es invariante ante transformaciones ortogonales, es decir, si $\mathbf{Q}\mathbf{Z} \sim \mathbf{Z}$ cualquiera sea $\mathbf{Q} \in \mathbb{R}^{d \times d}$ ortogonal, donde el símbolo \sim denota “tiene la misma distribución que” cuando del lado derecho tenemos un elemento aleatorio y “tiene distribución” cuando del lado derecho tenemos una distribución.

Es fácil caracterizar las distribuciones esféricas a partir de su función característica. Dado un vector aleatorio $\mathbf{Z} \in \mathbb{R}^d$ indicaremos por $\varphi_{\mathbf{Z}}$ la función característica de \mathbf{Z} . Efectivamente, el vector aleatorio \mathbf{Z} tiene distribución esférica si y sólo si existe $g : [0, \infty) \rightarrow \mathbb{R}$ tal que $\varphi_{\mathbf{Z}}(\mathbf{t}_d) = g(\mathbf{t}_d^T \mathbf{t}_d)$ para todo $\mathbf{t}_d \in \mathbb{R}^d$.

Debido a esta caracterización se suele notar $\mathbf{Z} \sim S_d(g)$. Tomando, por ejemplo, $g_{\mathcal{N}}(t) = \exp(-t/2)$, obtenemos que $S_d(g_{\mathcal{N}})$ es la distribución $\mathcal{N}_d(0, \mathbf{I}_d)$.

Las distribuciones elípticas pueden definirse como transformaciones afines de distribuciones esféricas.

Definición 2.2.2. Se dice que el vector $\mathbf{X} : \Omega \rightarrow \mathbb{R}^d$ tiene distribución elíptica $\mathcal{E}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, g)$ si existe un vector aleatorio $\mathbf{Z} \in \mathbb{R}^d$, una matriz $\mathbf{B} \in \mathbb{R}^{d \times d}$ y $\boldsymbol{\mu} \in \mathbb{R}^d$ tales que $\mathbf{Z} \sim S_d(g)$, $\boldsymbol{\Sigma} = \mathbf{B}\mathbf{B}^T$ y $\mathbf{X} \sim \mathbf{B}\mathbf{Z} + \boldsymbol{\mu}$.

Fijada g , la familia de distribuciones elípticas $\mathcal{E}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, g)$ es una clase de distribuciones con parámetro de posición $\boldsymbol{\mu}$ y parámetro de dispersión la matriz simétrica, semi definida positiva $\boldsymbol{\Sigma}$. Cuando \mathbf{X} tiene primer momento, resulta que $\mathbb{E}(\mathbf{X}) = \boldsymbol{\mu}$, mientras que cuando \mathbf{X} admite segundo momento, la matriz de covarianza $\text{COV}(\mathbf{X})$ resulta un múltiplo positivo de $\boldsymbol{\Sigma}$, es decir, existe $c > 0$ tal que $\text{COV}(\mathbf{X}) = c\boldsymbol{\Sigma}$. Como antes, $\mathcal{E}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, g_{\mathcal{N}}) = \mathcal{N}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

Si $\mathbf{Z} \sim S_d(g)$ tiene densidad f , entonces existe $\tilde{f} : \mathbb{R} \rightarrow \mathbb{R}$ tal que $f(\mathbf{z}) = \tilde{f}(\mathbf{z}^T \mathbf{z})$ en casi todo \mathbf{z} . Por otra parte, si $\boldsymbol{\Sigma} = \mathbf{B}\mathbf{B}^T$ es inversible, entonces $\mathbf{X} = \mathbf{B}\mathbf{Z} + \boldsymbol{\mu}$ también tiene densidad y se cumple que $f_{\mathbf{X}}(\mathbf{x}) = \det(\boldsymbol{\Sigma})^{-\frac{1}{2}} \tilde{f}((\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}))$. Finalmente, la función característica de \mathbf{X} es igual a $\varphi_{\mathbf{X}}(\mathbf{t}) = \exp(it^T \boldsymbol{\mu}) g(\mathbf{t}^T \boldsymbol{\Sigma} \mathbf{t})$. Más aún, la distribución $\mathcal{E}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, g)$ puede ser caracterizada por sus marginales, en el siguiente sentido: $\mathbf{X} \sim \mathcal{E}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, g)$ si y sólo si para todo $k < d$, $\mathbf{A}\mathbf{X} \sim \mathcal{E}_k(\mathbf{A}\boldsymbol{\mu}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^T, g)$ para toda matriz $\mathbf{A} \in \mathbb{R}^{k \times d}$.

Esta caracterización de la distribución elíptica fue utilizada en Bali y Boente (2009) para extender la noción de distribuciones elípticas a espacios de Hilbert separables.

Definición 2.2.3. Se dice que un elemento aleatorio Z en un espacio de Hilbert separable \mathcal{H} tiene distribución elíptica con parámetros $\mu \in \mathcal{H}$, $\Gamma : \mathcal{H} \rightarrow \mathcal{H}$ (donde Γ es un operador autoadjunto, semi definido positivo y compacto) y $g : [0, \infty) \rightarrow \mathbb{R}$, si y sólo si para cualquier $d \geq 1$ y cualquier operador lineal $A : \mathcal{H} \rightarrow \mathbb{R}^d$ se tiene que $AZ \sim \mathcal{E}_d(A\mu, A\Gamma A^*, g)$. Indicaremos $Z \sim \mathcal{E}(\mu, \Gamma, g)$.

Definición 2.2.4. Decimos que un elemento aleatorio $X \sim \mathcal{E}(\mu, \Gamma, g)$ tiene rango infinito si Γ tiene rango infinito.

Un elemento aleatorio $V : \Omega \rightarrow \mathcal{H}$ se dice Gaussiano si $V \sim \mathcal{E}(\mu, \Gamma, g_{\mathcal{N}})$ donde Γ representa en este caso el operador de covarianza de V .

La Proposición 2.2.1 presenta la caracterización dada en Boente *et al.* (2014) para elementos aleatorios elípticos de dimensión infinita como mezcla de normales.

Proposición 2.2.1. *Sea $X \sim \mathcal{E}(\mu, \Gamma, g)$ un elemento aleatorio elíptico de rango infinito en un espacio de Hilbert separable \mathcal{H} . Luego, existe un elemento aleatorio Gaussiano $V \in \mathcal{H}$ con $V \sim \mathcal{E}(0, \Gamma, g_{\mathcal{N}})$ y una variable aleatoria $S \geq 0$ independiente de V , tal que $X = \mu + SV$.*

La Proposición 2.2.1 no es cierta si Γ es de rango finito. Se pueden construir elementos aleatorios X elípticos de rango finito que no resultan mezcla de normales. De hecho, en Kano (1994) se muestran ejemplos de vectores aleatorios \mathbf{Z} en \mathbb{R}^q elípticos que no son mezcla de normales. Un ejemplo de distribuciones elípticas $\mathcal{E}_q(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \varphi)$, que no es mezcla de normales, se puede obtener a partir de la familia de distribuciones exponencial con potencias $\varepsilon_{q,\alpha}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ cuya densidad está dada por

$$f(\mathbf{z}) = c_{q,\alpha} (\det(\boldsymbol{\Sigma}))^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} [(\mathbf{z} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{z} - \boldsymbol{\mu})]^\alpha \right\},$$

donde $c_{q,\alpha}$ es una constante normalizadora (ver también Bilodeau y Brenner, 1999). Efectivamente, para $\alpha = 0,5$, $\mathbb{E}(z_1 - \mu_1)^2 = 4(q+1)\boldsymbol{\Sigma}_{11}$, donde $\boldsymbol{\Sigma}_{11}$ es el elemento (1, 1) de la matriz $\boldsymbol{\Sigma}$. Como dicho momento depende de la dimensión, esta familia de distribuciones no es consistente en el sentido de Kano (1994) y por lo tanto, no es mezcla de normales. Esta distribución permite obtener elementos aleatorios X elípticos de rango finito que no resultan mezcla de normales. Efectivamente, sea $\mathbf{Z} = (Z_1, \dots, Z_q)^T \sim \varepsilon_{q,\alpha}(\mathbf{0}, \boldsymbol{\Sigma})$ con $\alpha = 0,5$ y definamos $X = \sum_{j=1}^q Z_j \phi_j$, donde ϕ_j son los primeros q elementos de una base ortonormal de \mathcal{H} . El elemento aleatorio X resulta elíptico, pero no es mezcla de normales. Efectivamente,

supongamos que X es mezcla de normales, es decir, como $\mu = 0$, $X = SV$ donde $S \geq 0$ es una variable aleatoria y $V \in \mathcal{H}$ es un elemento Gaussiano con media $0_{\mathcal{H}}$, con S y V independientes entre sí. Definamos $T : \mathcal{H} \rightarrow \mathbb{R}^q$ como $T(x) = (\langle x, \phi_1 \rangle, \dots, \langle x, \phi_q \rangle)$, luego $T(X) = \mathbf{Z}$ y como $X = SV$ tendríamos que $T(X) = ST(V)$ donde el vector $T(V) \in \mathbb{R}^q$ es normal con media $\mathbf{0}$, llegando a una contradicción ya que \mathbf{Z} no es mezcla de normales.

2.3. Medidas de asociación

Así como suele interesar estudiar medidas de centralidad o de dispersión alternativas a la media y al desvío estándar para el caso de variables aleatorias a valores reales, también resulta de interés el estudio de medidas alternativas de asociación o correlación para vectores aleatorios en \mathbb{R}^2 . La razón de dicho interés es la misma. Las medidas clásicas se basan en esperanzas y por lo tanto, los estimadores relacionados son muy sensibles a la presencia de datos atípicos y por ende muy inestables, lo cual genera la necesidad de medidas más estables o robustas ante la presencia de datos atípicos.

Una medida de asociación ρ es un funcional sobre el espacio de medidas bivariadas que verifica las siguientes propiedades

$$(i) \quad \rho(X, Y) = \rho(Y, X)$$

$$(ii) \quad \rho(aX + b, cY + d) = \text{signo}(ac)\rho(X, Y) \text{ para todo } a, b, c, d \in \mathbb{R}$$

$$(iii) \quad -1 \leq \rho(X, Y) \leq 1$$

donde $(X, Y)^T \in \mathbb{R}^2$ es un vector aleatorio tal que $(X, Y)^T \sim P$ e indicamos indistintamente $\rho(X, Y) = \rho(P)$. Una propiedad deseable de una medida de asociación es que identifique independencia, esto es, que si X e Y son independientes entonces $\rho(X, Y) = 0$ aunque en muchos casos la recíproca no vale, como es bien sabido.

Es importante tener en cuenta que distintas medidas de asociación pueden representar distintas cantidades poblacionales al igual que la media y la mediana son medidas de posición que representan distintas cantidades según el tipo de distribución. En los ejemplos que daremos trataremos de mostrar lo que las medidas de asociación definidas representan en algunos casos particulares, como ser el de las distribuciones normales bivariadas.

Una propiedad deseable es que distintas medidas de asociación midan lo mismo para ciertas familias de distribuciones \mathcal{P} . Si esto ocurre diremos que ρ es Fisher consistente en \mathcal{P} . En general, \mathcal{P} es una familia de distribuciones que contiene a las distribuciones normales bivariadas, como ser la familia de distribuciones elípticas. Esta propiedad nos permitirá garantizar la consistencia de los estimadores de la direcciones canónicas a las direcciones de interés.

En particular, si $(X, Y)^T \sim \mathcal{E}_2(\boldsymbol{\mu}, \boldsymbol{\Sigma}_\kappa, g)$, lo que incluye el caso normal con $g = g_N$, y donde la matriz $\boldsymbol{\Sigma}_\kappa$ está dada por

$$\boldsymbol{\Sigma}_\kappa = \begin{pmatrix} \sigma_1^2 & \kappa \sigma_1 \sigma_2 \\ \kappa \sigma_1 \sigma_2 & \sigma_2^2 \end{pmatrix} \quad (2.4)$$

nuestro objetivo es estimar la cantidad $-1 < \kappa < 1$ que representa una medida de cuán relacionadas están las variables X e Y , por analogía a lo que sucede en el caso normal. Observemos que si la medida de asociación cumple (ii) podemos suponer que $\boldsymbol{\mu} = \mathbf{0}$ y que $\sigma_1 = \sigma_2 = 1$ cuando intentamos estimar κ . Por lo tanto, si por simplicidad llamamos $P_\kappa = \mathcal{E}_2(\mathbf{0}, \boldsymbol{\Sigma}_\kappa, g)$ donde

$$\boldsymbol{\Sigma}_\kappa = \begin{pmatrix} 1 & \kappa \\ \kappa & 1 \end{pmatrix} \quad (2.5)$$

y si definimos $T(\kappa) = \rho(P_\kappa)$, la consistencia Fisher dice que $T(\kappa) = \kappa$.

Recordemos que dados $(X_i, Y_i)^T \in \mathbb{R}^2$, $i = 1 \dots, n$, independientes e idénticamente dis-

tribuidos, la medida empírica se define como

$$P_n(A) = P_n[X, Y](A) = \frac{1}{n} \sum_{i=1}^n \mathbb{I}_A(X_i, Y_i),$$

donde A es un conjunto boreliano de \mathbb{R}^2 e \mathbb{I}_A es la función indicadora del conjunto A que vale 1 en A y 0 fuera de A .

La propiedad $T(\kappa) = \kappa$ es una de las condiciones a pedir para asegurar que el estimador obtenido como $\hat{\kappa} = \rho(P_n[X, Y])$ resulte un estimador consistente de κ .

Recordaremos a continuación algunas medidas de asociación.

2.3.1. Medidas de Asociación no robustas

En esta sección consideraremos medidas de asociación tales que sus versiones muestrales no son resistentes a la presencia de datos atípicos.

2.3.1.1. Correlación de Pearson

Dado $(X, Y)^T : \Omega \rightarrow \mathbb{R}^2$ con distribución conjunta $P = P[X, Y]$, el coeficiente de correlación de Pearson al cual comunmente nos referiremos como correlación clásica se define como:

$$\rho = \rho_{\text{CL}}(P) = \rho_{\text{CL}}(X, Y) = \begin{cases} \frac{\text{COV}(X, Y)}{(\text{VAR}(X)\text{VAR}(Y))^{\frac{1}{2}}} & \text{si } \text{VAR}(X)\text{VAR}(Y) \neq 0 \\ 0 & \text{si } \text{VAR}(X)\text{VAR}(Y) = 0 \end{cases}$$

donde $\text{VAR}(X)$, $\text{VAR}(Y)$ y $\text{COV}(X, Y)$ son las varianzas y covarianzas clásicas de las variables aleatorias X e Y .

A partir de ahora, al considerar otros funcionales de asociación cuya definición involucre una división, entenderemos que el resultado es 0 cuando el denominador se anula.

Observación 2.3.1. Vale la pena mencionar que ρ_{CL} se aplica a las medidas de probabilidad de \mathbb{R}^2 para las cuales se puede calcular la correlación clásica, es decir, a medidas de probabilidad que corresponden a vectores aleatorios con segundo momento finito.

Como mencionamos anteriormente indicaremos indistintamente $\rho_{\text{CL}}(X, Y)$ o $\rho_{\text{CL}}(P[X, Y])$, donde $P[X, Y]$ representa la medida de probabilidad que induce en \mathbb{R}^2 el vector $(X, Y)^T$, según sea conveniente para alivianar notación. El mismo abuso de notación, de aplicar un funcional a un vector o a una variable aleatoria en vez de a la medida de probabilidad que induce, lo cometeremos con otros funcionales.

En principio, es natural pensar que el parámetro de interés que deseamos encontrar es $\rho_{\text{CL}}(P)$ ya que si $(X, Y)^T \sim P_\kappa = \mathcal{E}_2(\mathbf{0}, \Sigma_\kappa, g)$ con Σ_κ dada en (2.5) y además $\mathbb{E}X^2 < \infty$ y $\mathbb{E}Y^2 < \infty$, entonces $\rho_{\text{CL}}(P_\kappa) = \kappa$.

El coeficiente de correlación muestral, denotado por r , no es otra cosa que el funcional ρ_{CL} aplicado a la medida empírica $P_n[X, Y]$.

$$r = \hat{\rho}_{\text{CL}}(P) = \rho_{\text{CL}}(P_n[X, Y]) = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\left(\sum_{i=1}^n (X_i - \bar{X})^2 \sum_{i=1}^n (Y_i - \bar{Y})^2 \right)^{\frac{1}{2}}}.$$

Por un lado r es el estimador de máxima verosimilitud de κ si se supone que la distribución de $(X, Y)^T \sim \mathcal{N}(\mathbf{0}, \Sigma_\kappa)$ donde Σ_κ está dada en (2.4). Por otro lado, bajo este modelo r resulta tener eficiencia asintótica óptima por ser un estimador de máxima verosimilitud, ver Daniels (1961). Sin embargo, en el modelo de contaminación que corresponde a una mezcla de normales bivariadas ($0 \leq \epsilon < 0,5$), es decir, si

$$(X, Y)^T \sim (1 - \epsilon)\mathcal{N}(\mathbf{0}, \Sigma_\kappa) + \epsilon\mathcal{N}(\mathbf{0}, \Sigma_{\kappa^*}),$$

con Σ_κ como en (2.4) y

$$\Sigma_{\kappa^*} = \begin{pmatrix} (\sigma_1^*)^2 & \kappa^* \sigma_1^* \sigma_2^* \\ \kappa^* \sigma_1^* \sigma_2^* & (\sigma_2^*)^2 \end{pmatrix},$$

el coeficiente de correlación muestral puede resultar altamente sesgado respecto del parámetro estimado κ . De hecho, dado $\epsilon > 0$, si $k = \sigma_1^*/\sigma_1 = \sigma_2^*/\sigma_2$ es suficientemente grande, resulta que $\mathbb{E}r \approx \kappa^*$. Este hecho muestra la necesidad de utilizar estimadores de asociación robustos si el modelo central es Gaussiano.

2.3.1.2. Distancia de correlación

Una medida de asociación utilizada con frecuencia para clasificación de datos funcionales es la distancia de correlación o correlación Browniana introducida por Székely *et al.* (2007) y Székely y Rizzo (2009), respectivamente.

Dados vectores aleatorios $\mathbf{X} \in \mathbb{R}^p$ e $\mathbf{Y} \in \mathbb{R}^q$, recordemos que indicábamos por $\varphi_{\mathbf{X}}$ y $\varphi_{\mathbf{Y}}$ a las funciones características de \mathbf{X} e \mathbf{Y} , respectivamente. Sea $\varphi_{\mathbf{X},\mathbf{Y}}$ la función característica conjunta de (\mathbf{X}, \mathbf{Y}) . Es conocido que \mathbf{X} e \mathbf{Y} son independientes si y sólo si $\varphi_{\mathbf{X},\mathbf{Y}} = \varphi_{\mathbf{X}}\varphi_{\mathbf{Y}}$. Por lo tanto, un modo natural de medir la dependencia entre \mathbf{X} e \mathbf{Y} es encontrar una norma razonable para medir la distancia entre $\varphi_{\mathbf{X},\mathbf{Y}}$ y $\varphi_{\mathbf{X}}\varphi_{\mathbf{Y}}$. Para funciones β a valores complejos definidas en \mathbb{R}^{p+q} , $\beta : \mathbb{R}^{p+q} \rightarrow \mathbb{C}$, la norma L^2 pesada $\|\cdot\|_{\omega}$ se define por:

$$\|\beta(\mathbf{t}, \mathbf{s})\|_{\omega}^2 = \int_{\mathbb{R}^{p+q}} |\beta(\mathbf{t}, \mathbf{s})|^2 \omega(\mathbf{t}, \mathbf{s}) d\mathbf{t} d\mathbf{s},$$

donde $\omega(\mathbf{t}, \mathbf{s})$ es una función de peso positiva arbitraria, para la cuál existe la integral. Székely *et al.* (2007) definieron la medida de dependencia $\nu^2(\mathbf{X}, \mathbf{Y}, \omega)$ como

$$\nu^2(\mathbf{X}, \mathbf{Y}, \omega) = \|\varphi_{\mathbf{X},\mathbf{Y}}(\mathbf{t}, \mathbf{s}) - \varphi_{\mathbf{X}}(\mathbf{t})\varphi_{\mathbf{Y}}(\mathbf{s})\|_{\omega}^2.$$

Si notamos $\nu^2(\mathbf{X}, \omega) = \nu^2(\mathbf{X}, \mathbf{X}, \omega)$ y de manera análoga $\nu^2(\mathbf{Y}, \omega)$ se puede definir una medida de asociación sin signo como

$$R_{\omega} = \frac{\nu(\mathbf{X}, \mathbf{Y}; \omega)}{\sqrt{\nu(\mathbf{X}; \omega)\nu(\mathbf{Y}; \omega)}}.$$

Sea $|\cdot|$ la norma euclídea, Székely y Rizzo (2009) prueban que eligiendo

$$\omega(t, s) = (c_p c_q |t|_p^{p+1} |s|_q^{q+1})^{-1}, \quad (2.6)$$

donde

$$c_p = \frac{2\pi^{p/2}\Gamma(1/2)}{2\Gamma((p+1)/2)}, \quad c_q = \frac{2\pi^{q/2}\Gamma(1/2)}{2\Gamma((q+1)/2)} \quad \text{y} \quad \Gamma(z) = \int_0^\infty t^{z-1} \exp(-t) dt,$$

entonces R_ω caracteriza independencia, es decir, $R_\omega \geq 0$ y $R_\omega = 0$ si y sólo si X e Y son independientes y R_ω es invariante por cambios de escala, o sea, $R_\omega(\sigma X, \sigma Y) = R_\omega(X, Y)$ para todo $\sigma > 0$. Por esta razón, introducen la distancia de correlación definida a continuación.

Definición 2.3.1. Sean $\nu(\cdot, \cdot) = \nu(\cdot, \cdot, \omega)$ y $\nu(\cdot) = \nu(\cdot, \omega)$ con ω definido en (2.6). La distancia de correlación se define como la raíz no negativa $\mathcal{R}(\mathbf{X}, \mathbf{Y})$ de

$$\mathcal{R}^2(\mathbf{X}, \mathbf{Y}) = \begin{cases} \frac{\nu^2(\mathbf{X}, \mathbf{Y})}{\sqrt{\nu^2(\mathbf{X})\nu^2(\mathbf{Y})}} & \text{si } \nu^2(\mathbf{X})\nu^2(\mathbf{Y}) > 0 \\ 0 & \text{si } \nu^2(\mathbf{X})\nu^2(\mathbf{Y}) = 0. \end{cases}$$

Como dijimos anteriormente, a diferencia de otras medidas de asociación, la distancia de correlación caracteriza independencia, lo que ha generalizado su uso. Por otra parte, dadas dos variables aleatorias X e Y , el Teorema 3 de Székely y Rizzo (2009) muestra que la distancia de correlación verifica propiedades análogas a las del valor absoluto de una medida de asociación descritas anteriormente, es decir, $\mathcal{R}(X, Y) = \mathcal{R}(Y, X)$, $\mathcal{R}(aX + b, cY + d) = \mathcal{R}(X, Y)$, $0 \leq \mathcal{R}(X, Y) \leq 1$.

Es de interés saber que cantidad representa la distancia de correlación en el caso de vectores aleatorios bivariados. Por la equivariancia dada en el Teorema 3, basta considerar el caso de vectores bivariados cuyas componentes tienen varianza 1. Del Teorema 6 de Székely y Rizzo (2009), se obtiene que si $(X, Y)^T \sim \mathcal{N}(\mathbf{0}, \Sigma_\kappa)$ donde Σ_κ está dada en (2.5), entonces $\mathcal{R}(X, Y) \leq |\kappa|$ y más aún, $\mathcal{R}^2(X, Y) = \gamma(\kappa)$, donde

$$\gamma(\kappa) = \frac{\kappa \arcsen(\kappa) + \sqrt{1 - \kappa^2} - \kappa \arcsen(\kappa/2) - \sqrt{4 - \kappa^2} + 1}{1 + \pi/3 - \sqrt{3}}.$$

La función γ , definida en el intervalo $[-1, 1]$ es una función par y toma valores en el $[0, 1]$. Si restringimos el dominio de γ al intervalo $[0, 1]$, resulta estrictamente creciente con imagen el intervalo $[0, 1]$ y, por lo tanto, existe $\gamma^{-1} : [0, 1] \rightarrow [0, 1]$.

Para identificar en el caso normal al parámetro de interés, dado un vector bivariado $(X, Y)^T$ definimos la medida de asociación ρ_{DIST} como

$$\rho_{\text{DIST}}(X, Y) = \gamma^{-1}(\mathcal{R}^2(X, Y)). \quad (2.7)$$

De esta forma, si $(X, Y)^T \sim \mathcal{N}(\mathbf{0}, \Sigma_\kappa)$ donde Σ_κ está dada en (2.5), se cumple que

$$\rho_{\text{DIST}}(X, Y) = \gamma^{-1}(\mathcal{R}^2(X, Y)) = \gamma^{-1}(\gamma(\kappa)) = \gamma^{-1}(\gamma(|\kappa|)) = |\kappa|. \quad (2.8)$$

Como en correlación canónica, nuestro interés será maximizar correlaciones, nuestro objetivo será el parámetro $|\kappa|$. La ecuación (2.8) dice que ρ_{DIST} es Fisher consistente para $|\kappa|$ en la familia de las normales bivariadas estándar.

2.3.2. Medidas de correlación robustas

Hay distintas maneras de obtener estimadores robustos de asociación. En esta sección, presentaremos algunas de las propuestas existentes en la literatura. Un desarrollo más completo puede verse en Shevlyakov y Vilchevski (2001). Las introduciremos de acuerdo a la propiedad que se utilizó para definir las, es decir, introduciremos primero alternativas al coeficiente de correlación muestral obtenidas podando observaciones grandes. Presentaremos luego algunas medidas de asociación basados en nociones no paramétricas para luego describir las medidas obtenidas mediante funcionales de escala. Finalmente, consideraremos medidas basadas en matrices de dispersión.

En lo que sigue, por simplicidad, indicaremos por $\vec{\mathbf{X}}$ y $\vec{\mathbf{Y}}$ a los vectores de observaciones (X_1, \dots, X_n) y (Y_1, \dots, Y_n) , respectivamente. Recordemos que un funcional de escala $\sigma_{\mathbb{R}}$ es una función del espacio de medidas de probabilidad sobre \mathbb{R} en los reales no-negativos. Es decir, dada una variable aleatoria U con distribución P_U , un funcional de escala cumple que

- $\sigma_{\mathbb{R}}(P_U) \geq 0$,

$$\blacksquare \sigma_{\mathbf{R}}(P_{aU+b}) = |a|\sigma_{\mathbf{R}}(P_U).$$

Como antes indicaremos indistintamente $\sigma_{\mathbf{R}}(U)$ o $\sigma_{\mathbf{R}}(P_U)$.

2.3.2.1. Estimadores basados en medias podadas

Una opción natural para robustificar el coeficiente de correlación muestral es modificar en la definición de $r = \rho_{\text{CL}}(P_n[X, Y])$ el promedio de las observaciones por estimadores robustos de posición y posteriormente, en lugar de calcular el promedio del producto de las observaciones centradas evaluar un promedio donde las observaciones alejadas de su medida de posición tienen menos peso. Para ello, podemos podar un porcentaje prefijado de las observaciones más lejanas o podemos acotar su influencia mediante una función de escores acotada. Presentaremos un enfoque que unifica ambas alternativas.

Sean $\hat{\mu}(\vec{\mathbf{X}})$ y $\hat{\mu}(\vec{\mathbf{Y}})$ estimadores robustos de posición de las muestras $\{X_1, \dots, X_n\}$ y $\{Y_1, \dots, Y_n\}$, respectivamente. Podemos elegir como $\hat{\mu}(\vec{\mathbf{X}})$ la mediana de las observaciones o un M -estimador de posición. Sea $\psi = \psi(z)$ una función monótona como la función de Huber $\psi_k(z) = \max[-k, \min(z, k)]$ e indiquemos \sum_{α} el promedio α -podado. Más precisamente, si $Z_{(1)}, \dots, Z_{(n)}$ son los estadísticos de orden de la muestra Z_1, \dots, Z_n , dado $0 \leq \alpha \leq 1/2$ definimos

$$\sum_{\alpha} Z_i = \frac{1}{n - 2k} \sum_{i=k+1}^{n-k} Z_{(i)} \quad \text{con} \quad k = \lceil \alpha(n - 1) \rceil,$$

donde $\lceil a \rceil$ denota la parte entera de a .

Utilizando las ideas descritas más arriba, obtenemos la siguiente familia de estimadores

$$r_{\alpha}(\psi) = \frac{\sum_{\alpha} \psi(X_i - \hat{\mu}(\vec{\mathbf{X}}))\psi(Y_i - \hat{\mu}(\vec{\mathbf{Y}}))}{\left\{ \sum_{\alpha} \psi^2(X_i - \hat{\mu}(\vec{\mathbf{X}})) \sum_{\alpha} \psi^2(Y_i - \hat{\mu}(\vec{\mathbf{Y}})) \right\}^{\frac{1}{2}}}. \quad (2.9)$$

En particular, tomando $\alpha = 0,5$, $\hat{\mu}(\vec{X}) = \text{med}(\vec{X}) = \text{mediana}_{1 \leq i \leq n}(X_i)$, $\hat{\mu}(\vec{Y}) = \text{med}(\vec{Y}) = \text{mediana}_{1 \leq i \leq n}(Y_i)$ y $\psi(z) = z$ en (2.9) obtenemos el estimador de correlación mediana dado por

$$r_{0,5} = \hat{\rho}_{\text{CM}}(X, Y) = \frac{\text{mediana}_{1 \leq i \leq n} \left\{ \left[X_i - \text{med}(\vec{X}) \right] \left[Y_i - \text{med}(\vec{Y}) \right] \right\}}{\text{MAD}(\vec{X}) \text{MAD}(\vec{Y})}$$

donde $\text{MAD}(\vec{X}) = \text{mediana}_{1 \leq i \leq n} |X_i - \text{med}(\vec{X})|$.

Este estimador está asociado al funcional

$$\rho_{\text{CM}}(X, Y) = \frac{\text{med}(X - \text{med}(X))(Y - \text{med}(Y))}{\text{MAD}(X) \text{MAD}(Y)},$$

donde para una variable aleatoria Z , $\text{med}(Z)$ representa la mediana de la distribución de Z y $\text{MAD}(Z)$ la mediana de la distribución de la variable aleatoria $V = |Z - \text{med}(Z)|$.

Falk (1998) describe distintas propiedades de las comedias. En particular, prueba que existe una función $\gamma : [-1, 1] \rightarrow \mathbb{R}$ estrictamente creciente con derivada continua tal que $\gamma(-1) = -\gamma(1)$ y $\gamma(0) = 0$ de manera tal que si $(X, Y)^T \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}_\kappa)$ donde $\boldsymbol{\Sigma}_\kappa$ dada en (2.4), entonces

$$\rho_{\text{CM}}(X, Y) = \frac{\gamma(\kappa)}{\gamma(1)} \in [-1, 1].$$

A pesar de no obtener una fórmula explícita para la función γ , Falk (1998) muestra que γ cumple la ecuación integral

$$\gamma(\rho) = (\Phi^{-1}(3/4))^2 \exp \left(- \int_{\rho}^1 \frac{\int_{-x}^1 \exp \left(-\frac{\gamma(x)}{u+x} \right) \frac{1}{(u+x)^2(1-u^2)^{1/2}} du}{\int_{-x}^1 \exp \left(-\frac{\gamma(x)}{u+x} \right) \frac{1}{(u+x)(1-u^2)^{1/2}} du} dx \right).$$

Por lo tanto, la función γ se puede aproximar mediante Monte Carlo. Observemos que $\gamma(1) = (\Phi^{-1}(3/4))^2$. Por otra parte, definiendo

$$\rho_{\text{CM}}^*(X, Y) = \gamma^{-1}(\gamma(1)\rho_{\text{CM}}(X, Y)), \quad (2.10)$$

obtenemos que ρ_{CM}^* es Fisher consistente en la familia de distribuciones normales ya que si $(X, Y)^T \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}_\kappa)$ con $\boldsymbol{\Sigma}_\kappa$ dada en (2.4)

$$\rho_{\text{CM}}^*(X, Y) = \kappa.$$

La familia definida en (2.9) no es invariante por cambios de escala cuando la ψ no es la identidad. Por esta razón, otra de las familias introducidas en el contexto de acotar observaciones alejadas de la medida de posición, trunca valores grandes de dichos alejamientos pero respecto de una medida de escala. Más precisamente, dada una función ψ impar, monótona no decreciente y acotada o redescendiente y funcionales robustos de posición y escala, denotados por $\mu(X)$ y $\sigma_{\text{R}}(X)$, respectivamente, se puede definir un funcional de covarianza robusto entre las variables aleatorias X e Y como

$$\text{RCov}(X, Y) = \sigma_{\text{R}}(X)\sigma_{\text{R}}(Y)\mathbb{E}\left[\psi\left(\frac{X - \mu(X)}{\sigma_{\text{R}}(X)}\right)\psi\left(\frac{Y - \mu(Y)}{\sigma_{\text{R}}(Y)}\right)\right]$$

y a partir de él un funcional de correlación como

$$\text{RCORR}(X, Y) = \frac{\text{RCov}(X, Y)}{\{\text{RCov}(X, X)\text{RCov}(Y, Y)\}^{\frac{1}{2}}}.$$

2.3.2.2. Medidas basadas en rangos y signos

Un procedimiento de estimación puede tener propiedades de robustez si se utilizan estadísticos de rango. Los más conocidos son los coeficientes de correlación cuadrante y los coeficientes de correlación de Spearman y de Kendall.

El coeficiente de correlación cuadrante introducido por Blomqvist (1950) se define como

$$r_Q = \frac{1}{n} \sum_{i=1}^n \text{sgn}(X_i - \text{med}(\vec{\mathbf{X}}))\text{sgn}(Y_i - \text{med}(\vec{\mathbf{Y}})),$$

donde $\text{sgn}(t)$ es el signo de t . Observemos que r_Q coincide con calcular el coeficiente de correlación muestral entre los signos de las desviaciones respecto a la mediana.

El coeficiente de correlación de rangos de Spearman (Spearman, 1904) es el coeficiente de correlación muestral de Pearson de los rangos de las muestras \vec{X} e \vec{Y} , es decir,

$$r_{\text{SP}} = \rho_{\text{CL}}(P_n[R(\vec{X}), R(\vec{Y})]),$$

donde $R(\vec{Z})$ es el vector de rangos $(R(Z_1), \dots, R(Z_n))^T$ de la muestra Z_1, \dots, Z_n . Por otra parte, el coeficiente de correlación de signos de Kendall está dado por

$$r_{\text{K}} = \frac{1}{\binom{n}{2}} \sum_{i < j} \text{sgn}((X_i - X_j)(Y_i - Y_j)).$$

El funcional asociado al coeficiente de correlación de rangos de Spearman se define como

$$\tilde{\rho}_{\text{SP}}(X, Y) = \tilde{\rho}_{\text{SP}}(P[X, Y]) = \rho_{\text{CL}}(F_X(X), F_Y(Y)),$$

donde F_X y F_Y son las distribuciones acumuladas de X e Y , respectivamente, mientras que el funcional de correlación de Kendall es

$$\tilde{\rho}_{\text{K}}(X, Y) = \mathbb{E} \{ \text{sgn}((X_1 - X_2)(Y_1 - Y_2)) \},$$

donde $(X_1, Y_1)^T$ y $(X_2, Y_2)^T$ tienen la misma distribución que $(X, Y)^T$ y son independientes.

En los capítulos 9 y 10 de Kendall (1970) se muestra que si $(X, Y)^T \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}_\kappa)$ con $\boldsymbol{\Sigma}_\kappa$ dada en (2.5), entonces

$$\tilde{\rho}_{\text{SP}}(X, Y) = \frac{6}{\pi} \arcsen\left(\frac{\kappa}{2}\right) \quad \tilde{\rho}_{\text{K}}(X, Y) = \frac{2}{\pi} \arcsen(\kappa).$$

Con lo cual, si definimos

$$\rho_{\text{SP}}(X, Y) = 2 \text{sen}\left(\frac{\pi}{6} \tilde{\rho}_{\text{SP}}(X, Y)\right) \quad \text{y} \quad \rho_{\text{K}}(X, Y) = \text{sen}\left(\frac{\pi}{2} \tilde{\rho}_{\text{K}}(X, Y)\right) \quad (2.11)$$

resulta que para todo $(X, Y)^T \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}_\kappa)$

$$\rho_{\text{SP}}(X, Y) = \kappa \quad \text{y} \quad \rho_{\text{K}}(X, Y) = \kappa,$$

es decir, ρ_{SP} y ρ_{K} son Fisher consistentes en la familia de las normales bivariadas.

2.3.2.3. Medidas basadas en funcionales de dispersión robustos

Otra propuesta de estimadores robustos proviene de observar que el funcional de correlación de Pearson se puede pensar como la covarianza de las variables estandarizadas, es decir,

$$\rho_{\text{CL}}(X, Y) = \text{COV} \left(\frac{X}{\text{SD}(X)}, \frac{Y}{\text{SD}(Y)} \right), \quad (2.12)$$

donde $\text{SD}(Z)$ es el desvío estándar de la variable aleatoria Z . Por otro lado, la covarianza se relaciona con el desvío estándar a través de la ecuación

$$\text{COV}(X, Y) = \frac{1}{4}(\text{SD}^2(X + Y) - \text{SD}^2(X - Y)). \quad (2.13)$$

El funcional de escala SD no resulta robusto ya que es muy sensible a la presencia de datos atípicos, especialmente si son asimétricos. Basados en (2.12) y (2.13), Gnanadesikan y Kettenring (1972) introducen la medida de covarianza que presentaremos a continuación, reemplazando el desvío estándar por una escala robusta.

Basados en (2.13), pero usando un funcional robusto de escala σ_{R} , Gnanadesikan y Kettenring (1972) introducen la medida de covarianza

$$\text{COV}_{\text{GK}}(X, Y) = \frac{1}{4}(\sigma_{\text{R}}^2(X + Y) - \sigma_{\text{R}}^2(X - Y)),$$

de donde reemplazando en (2.12) COV por COV_{GK} y SD por σ_{R} , se obtiene el siguiente funcional de correlación robusto

$$\rho_{\text{GK}}(X, Y) = \text{RCORR}_{\text{GK}}(X, Y) = \frac{1}{4} \left\{ \sigma_{\text{R}}^2 \left(\frac{X}{\sigma_{\text{R}}(X)} + \frac{Y}{\sigma_{\text{R}}(Y)} \right) - \sigma_{\text{R}}^2 \left(\frac{X}{\sigma_{\text{R}}(X)} - \frac{Y}{\sigma_{\text{R}}(Y)} \right) \right\}, \quad (2.14)$$

así como una nueva medida de covarianza robusta como

$$\text{RCOV}_{\text{GK}}(X, Y) = \sigma_{\text{R}}(X) \sigma_{\text{R}}(Y) \text{RCORR}_{\text{GK}}(X, Y). \quad (2.15)$$

El funcional ρ_{GK} es Fisher-consistente para distribuciones normales pues, por la propiedad de equivariancia por cambios de escala, el estimador de escala cumple que si $U \sim \mathcal{N}(0, \sigma^2)$,

$\sigma_R(U) = c\sigma$ para alguna constante c y para todo $\sigma^2 > 0$. Más aún, en la sección 6.12.10 de Maronna *et. al.* (2006), se muestra que si $(X, Y)^T \sim \mathcal{E}_2(\boldsymbol{\mu}, \boldsymbol{\Sigma}_\kappa, g)$ con $\boldsymbol{\Sigma}_\kappa$ dada en (2.4) y si existe segundo momento finito, entonces $\rho_{\text{GK}}(X, Y) = \kappa$, es decir, es Fisher consistente también para distribuciones elípticas con segundo momento.

Sea $\hat{\sigma}_R(\vec{Z})$ el estimador robusto asociado a σ_R , es decir, $\hat{\sigma}_R(Z_1, \dots, Z_n) = \sigma_R(P_n[Z])$ donde $P_n[Z]$ es la empírica de la muestra $\vec{Z} = \{Z_1, \dots, Z_n\}$. El estimador robusto de asociación se obtiene como

$$\hat{\rho}_{\text{GK}}(\vec{X}, \vec{Y}) = \frac{1}{4} \left\{ \hat{\sigma}_R^2 \left(\frac{\vec{X}}{\hat{\sigma}_R(\vec{X})} + \frac{\vec{Y}}{\hat{\sigma}_R(\vec{X})} \right) - \hat{\sigma}_R^2 \left(\frac{\vec{X}}{\hat{\sigma}_R(\vec{X})} - \frac{\vec{Y}}{\hat{\sigma}_R(\vec{X})} \right) \right\}.$$

Otra familia de estimadores considerada en Gnanadesikan y Kettenring (1972) se basa en la identidad

$$\rho = \frac{\text{VAR}(U) - \text{VAR}(V)}{\text{VAR}(U) + \text{VAR}(V)}, \quad (2.16)$$

donde

$$U = \frac{1}{\sqrt{2}} \left(\frac{X}{\text{SD}(X)} + \frac{Y}{\text{SD}(Y)} \right) \quad V = \frac{1}{\sqrt{2}} \left(\frac{X}{\text{SD}(X)} - \frac{Y}{\text{SD}(Y)} \right).$$

Nuevamente, utilizando un funcional de escala robusto σ_R , se puede definir la siguiente medida de asociación robusta

$$\rho^*(X, Y) = \frac{\sigma_R^2(U) - \sigma_R^2(V)}{\sigma_R^2(U) + \sigma_R^2(V)},$$

donde ahora

$$U = \frac{1}{\sqrt{2}} \left(\frac{X}{\sigma_R(X)} + \frac{Y}{\sigma_R(Y)} \right) \quad V = \frac{1}{\sqrt{2}} \left(\frac{X}{\sigma_R(X)} - \frac{Y}{\sigma_R(Y)} \right).$$

Por supuesto, usando el estimador robusto $\hat{\sigma}_R$ asociado a σ_R se puede definir un estimador robusto de ρ^* como

$$\hat{\rho}^*(X, Y) = \frac{\hat{\sigma}_R^2(\hat{\mathbf{U}}) - \hat{\sigma}_R^2(\hat{\mathbf{V}})}{\hat{\sigma}_R^2(\hat{\mathbf{U}}) + \hat{\sigma}_R^2(\hat{\mathbf{V}})},$$

donde $\hat{\mathbf{U}} = \{\hat{U}_1, \dots, \hat{U}_n\}$, $\hat{\mathbf{V}} = \{\hat{V}_1, \dots, \hat{V}_n\}$ con

$$\hat{U}_i = \frac{1}{\sqrt{2}} \left(\frac{X_i}{\hat{\sigma}_R(\vec{X})} + \frac{Y_i}{\hat{\sigma}_R(\vec{X})} \right) \quad \hat{V}_i = \frac{1}{\sqrt{2}} \left(\frac{X_i}{\hat{\sigma}_R(\vec{X})} - \frac{Y_i}{\hat{\sigma}_R(\vec{X})} \right).$$

2.3.2.4. Medidas basadas en matrices de dispersión

Dado un vector bivariado $\mathbf{Z} = (X, Y)^T$, si llamamos $\mathbf{\Sigma} = \text{VAR}(X, Y)$ a la matriz de covarianza del vector \mathbf{Z} , el funcional de correlación de Pearson es igual a

$$\rho = \frac{\Sigma_{12}}{(\Sigma_{11}\Sigma_{22})^{\frac{1}{2}}},$$

donde Σ_{ij} es el elemento (i, j) de la matriz $\mathbf{\Sigma}$.

Esta propiedad motivó la definición de un coeficiente de asociación a partir de funcionales robustos de dispersión $\mathbf{V} = \mathbf{V}(P) = \mathbf{V}(X, Y)$, donde $(X, Y)^T \sim P$. De esta manera, se define el coeficiente $\rho_{\mathbf{V}}(P) = \rho_{\mathbf{V}}(X, Y)$ como

$$\rho_{\mathbf{V}}(X, Y) = \frac{\mathbf{V}_{12}(P)}{(\mathbf{V}_{11}(P)\mathbf{V}_{22}(P))^{\frac{1}{2}}}, \quad (2.17)$$

donde $\mathbf{V}_{ij}(P)$ es el elemento (i, j) de la matriz de dispersión $\mathbf{V}(P) = \mathbf{V}(X, Y)$.

Posibles funcionales de dispersión robustos son los M -estimadores de dispersión (Maronna, 1976), el funcional de dispersión de Stahel-Donoho propuesto por Stahel (1981) y Donoho (1982), el funcional de covarianza de determinante mínimo (MCD) o el de elipsoide de mínimo volumen (MVE) (ver Maronna *et al.*, 2006).

Reemplazando en (2.17) \mathbf{V} por $\widehat{\mathbf{V}} = \mathbf{V}(P_n)$, donde $P_n = P_n[X, Y]$ es la distribución empírica de $\mathbf{Z}_1 = (X_1, Y_1)^T, \dots, \mathbf{Z}_n = (X_n, Y_n)^T$, se tiene un estimador robusto $\widehat{\rho}_{\mathbf{V}}$ de asociación.

Una propiedad deseable para un funcional de dispersión \mathbf{V} es la equivarianza afín. Más precisamente, se dice que el funcional $\mathbf{V}(P)$ es afín equivariante si $\mathbf{V}(P_{\mathbf{A}\mathbf{Z}+\mathbf{a}}) = \mathbf{A}\mathbf{V}(P_{\mathbf{Z}})\mathbf{A}^T$ para todo $\mathbf{a} \in \mathbb{R}^2, \mathbf{A} \in \mathbb{R}^{2 \times 2}$ no singular, donde $P_{\mathbf{Z}}$ indica la distribución del vector aleatorio $\mathbf{Z} \in \mathbb{R}^2$. En particular, el M -estimador introducido por Maronna (1976) asociado al M -funcional de dispersión es afín equivariante y es relativamente robusto en dimensión 2 ya que su punto de ruptura es menor que 1/3. Más aún, eligiendo adecuadamente la función de pesos el M -funcional resulta Fisher-consistente en la familia de distribuciones normales.

En Maronna *et al.* (2006) se demuestra la siguiente propiedad.

Propiedad 2.3.1. *Sea $\rho_{\mathbf{V}}$ un funcional de correlación definido como en (2.17) con $\mathbf{V}(P)$ afín equivariante. Si $\mathbf{Z} \in \mathbb{R}^2$ es un vector aleatorio con distribución elíptica, $\mathbf{Z} \sim \mathcal{E}_2(\boldsymbol{\mu}, \boldsymbol{\Sigma}, g)$, entonces existe una constante $c > 0$ tal que*

$$\mathbf{V}(\mathbf{Z}) = c\boldsymbol{\Sigma}.$$

Por lo tanto, si \mathbf{Z} es elíptico,

$$\rho_{\mathbf{V}}(\mathbf{Z}) = \frac{\mathbf{V}_{12}(\mathbf{Z})}{(\mathbf{V}_{11}(\mathbf{Z})\mathbf{V}_{22}(\mathbf{Z}))^{\frac{1}{2}}} = \frac{\boldsymbol{\Sigma}_{12}}{(\boldsymbol{\Sigma}_{11}\boldsymbol{\Sigma}_{22})^{\frac{1}{2}}}.$$

Este resultado muestra que si $\mathbf{Z} = (X, Y)^T \sim \mathcal{E}_2(\boldsymbol{\mu}, \boldsymbol{\Sigma}_{\kappa}, g)$ con $\boldsymbol{\Sigma}_{\kappa}$ dada en (2.5) y el funcional de dispersión $\mathbf{V}(P)$ es afín equivariante, entonces $\rho_{\mathbf{V}}(\mathbf{Z}) = \kappa$, es decir, $\rho_{\mathbf{V}}$ es consistente Fisher para distribuciones elípticas.

Vale la pena observar que ρ_{CL} es un caso particular de $\rho_{\mathbf{V}}$ con $\mathbf{V}(\mathbf{Z}) = \text{COV}(\mathbf{Z})$ la matriz de covarianza del vector \mathbf{Z} y que COV es afín equivariante. Por lo tanto, en distribuciones elípticas con segundo momento $\rho_{\mathbf{V}}$ y ρ_{CL} representan la misma cantidad si el funcional de dispersión robusto $\mathbf{V}(P)$ es afín equivariante.

Una familia de estimadores de dispersión que no es afín equivariante fue introducida por Maronna y Zamar (2002) basándose en el funcional de dispersión RCOV_{GK} dado en (2.15). Más precisamente, dado un funcional de escala σ_{R} , la medida de asociación ρ_{GK} construida a partir de él y un vector aleatorio $\mathbf{U} = (U_1, U_2)^T \in \mathbb{R}^2$, definimos haciendo abuso de notación $\text{COV}_{\text{GK}}(\mathbf{U}) \in \mathbb{R}^{2 \times 2}$ como la matriz

$$\text{COV}_{\text{GK}}(\mathbf{U}) = \begin{pmatrix} \sigma_{\text{R}}^2(U_1) & \sigma_{\text{R}}(U_1)\sigma_{\text{R}}(U_2)\rho_{\text{GK}}(U_1, U_2) \\ \sigma_{\text{R}}(U_1)\sigma_{\text{R}}(U_2)\rho_{\text{GK}}(U_1, U_2) & \sigma_{\text{R}}^2(U_2) \end{pmatrix}.$$

Un problema de este funcional es que no es afín equivariante. Para obtener estimadores equivariantes por transformaciones ortogonales, Maronna y Zamar (2002) introducen un

funcional de dispersión al que denominan covarianza de Gnanadesikan Kettenring ortogonalizada que indicaremos COV_{OGK} . Describimos el procedimiento de construcción de este funcional.

Dado un vector aleatorio $\mathbf{Z} = (X, Y)^T \in \mathbb{R}^2$ y σ_{R} una medida robusta de escala en \mathbb{R} , para definir el funcional de dispersión $\text{COV}_{\text{OGK}}(\mathbf{Z})$ se procede del siguiente modo:

Algoritmo 1 Algoritmo para la definición de COV_{OGK}

- 1: Defina el vector $\mathbf{W} = (W_1, W_2)^T = (X/\sigma_{\text{R}}(X), Y/\sigma_{\text{R}}(Y))^T$, es decir, $\mathbf{W} = \mathbf{D}^{-1}\mathbf{Z}$ donde $\mathbf{D} \in \mathbb{R}^{2 \times 2}$ es la matriz diagonal $\mathbf{D} = \text{diag}\{\sigma_{\text{R}}(X), \sigma_{\text{R}}(Y)\}$.
 - 2: Calcule la matriz $\mathbf{\Upsilon} = \text{COV}_{\text{GK}}\mathbf{W}$. La matriz $\mathbf{\Upsilon}$ resulta simétrica pero no necesariamente semidefinida positiva.
 - 3: Realice la descomposición espectral de la matriz simétrica $\mathbf{\Upsilon}$, es decir, $\mathbf{\Upsilon} = \mathbf{E}\mathbf{\Lambda}\mathbf{E}^T$ donde las columnas de \mathbf{E} son los autovectores \mathbf{e}_1 y \mathbf{e}_2 de $\mathbf{\Upsilon}$ y $\mathbf{\Lambda} = \text{diag}\{\lambda_1, \lambda_2\}$ con $\lambda_1 \geq \lambda_2$ es la matriz de los respectivos autovalores.
 - 4: Construya $\mathbf{U} = (U_1, U_2)^T = \mathbf{E}^T\mathbf{W}$ el vector de “componentes principales” de \mathbf{W} . Sea $\mathbf{\Gamma} \in \mathbb{R}^{2 \times 2}$ la matriz $\text{diag}\{\sigma_{\text{R}}^2(U_1), \sigma_{\text{R}}^2(U_2)\}$. La matriz $\mathbf{\Gamma}$ es semidefinida positiva.
 - 5: Llamando $\mathbf{A} = \mathbf{D}\mathbf{E}$, se tiene que $\mathbf{Z} = \mathbf{A}\mathbf{U}$. Defina entonces $\text{COV}_{\text{OGK}}(\mathbf{Z})$ como $\text{COV}_{\text{OGK}}(\mathbf{Z}) = \mathbf{A}\mathbf{\Gamma}\mathbf{A}^T$.
-

En la sección 6.9.1 de Maronna *et al.* (2006) se encuentran más detalles sobre este funcional de dispersión. Tomando $\mathbf{V} = \text{COV}_{\text{OGK}}(\mathbf{Z})$ se define el funcional de asociación ρ_{OGK} . Más precisamente, sea $\mathbf{Z} = (X, Y)^T \in \mathbb{R}^2$ y $\mathbf{V} = \text{COV}_{\text{OGK}}(\mathbf{Z})$ se define $\rho_{\text{OGK}}(X, Y)$ como

$$\rho_{\text{OGK}}(X, Y) = \frac{\mathbf{V}_{12}}{\sqrt{\mathbf{V}_{11}\mathbf{V}_{22}}}. \quad (2.18)$$

Hemos visto que si el vector \mathbf{Z} es elíptico, o sea, si $\mathbf{Z} = (X, Y)^T \sim \mathcal{E}_2(\boldsymbol{\mu}, \boldsymbol{\Sigma}_{\kappa}, g)$ con $\boldsymbol{\Sigma}_{\kappa}$ dada en (2.4) y \mathbf{Z} tiene segundo momento finito entonces $\rho_{\text{GK}}(X, Y) = \kappa$. De los resultados dados en la sección 6.12.10 de Maronna *et al.* (2006), se deduce que existe una constante $c > 0$ tal que $\text{COV}_{\text{GK}}(\mathbf{Z}) = c \text{COV}\mathbf{Z}$, donde la constante c es tal que $\sigma_{\text{R}}(U_1) = c$ con $\mathbf{U} = (U_1, U_2)$

un vector con distribución esférica y matriz de covarianza identidad. Estas propiedades nos permitirán mostrar el siguiente resultado que muestra la consistencia Fisher de la medida de asociación ρ_{OGK} en la familia de distribuciones elípticas con segundo momento finito.

Proposición 2.3.1. *Sea $\mathbf{Z} = (X, Y)^T \in \mathbb{R}^2$ un vector aleatorio con segundo momento finito. Supongamos que $\mathbf{Z} \sim \mathcal{E}_2(\boldsymbol{\mu}, \boldsymbol{\Sigma}_\kappa, g)$ con $\boldsymbol{\Sigma}_\kappa$ dada en (2.4) tal que $\text{Cov}(\mathbf{Z}) = \boldsymbol{\Sigma}_\kappa$. Entonces $\rho_{\text{OGK}}(X, Y) = \kappa$.*

Demostración. Sin pérdida de generalidad podemos suponer que $\boldsymbol{\mu} = \mathbf{0}$. Utilizando la notación del Algoritmo 1 tenemos que $\boldsymbol{\Upsilon} = \text{COV}_{\text{GK}}(\mathbf{W}) = c \text{Cov} \mathbf{W}$, pues $\mathbf{W} = \mathbf{D}^{-1} \mathbf{Z}$ tiene distribución elíptica $\mathcal{E}_2(\mathbf{0}, \mathbf{D}^{-1} \boldsymbol{\Sigma}_\kappa \mathbf{D}^{-1}, g)$. Por lo tanto, en el paso 3 del Algoritmo 1 estamos consiguiendo una descomposición espectral de $\text{Cov} \mathbf{W} = c^{-1} \boldsymbol{\Upsilon}$ como

$$\text{Cov} \mathbf{W} = \mathbf{E} \left(\frac{\boldsymbol{\Lambda}}{c} \right) \mathbf{E}^T.$$

Luego, para $\mathbf{U} = \mathbf{E}^T \mathbf{W}$, $\text{Cov} \mathbf{U}$ resulta igual a la matriz diagonal $c^{-1} \boldsymbol{\Lambda}$. Por otra parte, como $\mathbf{U} = \mathbf{E}^T \mathbf{D}^{-1} \mathbf{Z}$ y \mathbf{Z} tiene distribución elíptica, \mathbf{U} tiene distribución elíptica. Más precisamente, $\mathbf{U} \sim \mathcal{E}_2(\mathbf{0}, \mathbf{E}^T \mathbf{D}^{-1} \boldsymbol{\Sigma}_\kappa \mathbf{D}^{-1} \mathbf{E}, g)$, por lo tanto como $\boldsymbol{\Sigma}_\kappa = \text{Cov} \mathbf{Z}$, deducimos que $c^{-1} \boldsymbol{\Lambda} = \text{Cov} \mathbf{U} = \mathbf{E}^T \mathbf{D}^{-1} \boldsymbol{\Sigma}_\kappa \mathbf{D}^{-1} \mathbf{E}$. Por otra parte, la constante c es el valor de la escala robusta en vectores esféricos con covarianza identidad, de donde se deduce que $\sigma_{\text{R}}(U_j) = \lambda_j$, para $j = 1, 2$, o sea, $\boldsymbol{\Gamma} = \boldsymbol{\Lambda}$ y $\boldsymbol{\Sigma}_\kappa = c^{-1} \boldsymbol{\Lambda} \boldsymbol{\Lambda}^T$.

Utilizando la definición de $\text{COV}_{\text{OGK}}(\mathbf{Z})$ como $\text{COV}_{\text{OGK}}(\mathbf{Z}) = \boldsymbol{\Lambda} \boldsymbol{\Gamma} \boldsymbol{\Lambda}^T$ y el hecho que $\boldsymbol{\Gamma} = \boldsymbol{\Lambda}$ obtenemos que

$$\text{COV}_{\text{OGK}} \mathbf{Z} = \boldsymbol{\Lambda} \boldsymbol{\Lambda} \boldsymbol{\Lambda}^T = c \boldsymbol{\Sigma}_\kappa. \quad (2.19)$$

Combinando (2.18) y (2.19) se concluye que

$$\rho_{\text{OGK}}(X, Y) = \kappa,$$

como queríamos probar. □

3. ¿POR QUÉ SUAVIZAR?

Este capítulo está dividido en cuatro secciones. En la Sección 3.1 se introducen las definiciones y los principales resultados que se conocen para el análisis de correlación canónica multivariado. Se extiende en forma natural de las definiciones de las direcciones y correlaciones canónicas poblacionales del caso euclídeo al caso de un elemento aleatorio $(X, Y)^T \in \mathcal{H} = \mathcal{H}_1 \times \mathcal{H}_2$ donde \mathcal{H}_i son espacios de Hilbert separables de dimensión posiblemente infinita. También se describen, siguiendo a Leurgans *et al.* (1993), los problemas que surgen al intentar extender del caso multivariado al funcional los estimadores de la primera correlación canónica. En esta sección, se definen además los estimadores basados en Sieves. En la Sección 3.2, se extiende el resultado de Leurgans *et al.* (1993) mostrando que la extensión del procedimiento multivariado para estimar la primera correlación canónica no sólo falla para el funcional de correlación de Pearson sino para una amplia variedad de funcionales de asociación. Además, se prueba este resultado para elementos aleatorios con distribución elíptica y no solamente para elementos Gaussianos. Como corolario de los problemas de los procedimientos de estimación basados en la extensión natural del caso multivariado al caso funcional, en la Sección 3.3, mostramos que los estimadores basados en Sieves no resultan consistentes si las dimensiones de los subespacios aproximantes son mayores o iguales al tamaño de muestra. Finalmente, en la Sección 3.4 se ve como este problema de los estimadores de Sieves, cuando la dimensión de los subespacios aproximantes es mayor o igual que el tamaño de muestra, se puede relacionar con los problemas de inconsistencia en el caso de datos de alta dimensión.

3.1. Introducción

En lo que sigue ρ es un funcional de asociación definido sobre distribuciones de \mathbb{R}^2 , como los vistos en el Capítulo 2.

Sea $(\Omega, \mathcal{A}, \mathbb{P})$ un espacio de probabilidad. Dados $\mathbf{X} : \Omega \rightarrow \mathbb{R}^p$ y $\mathbf{Y} : \Omega \rightarrow \mathbb{R}^q$ vectores aleatorios con distribución conjunta P y vectores fijos $\mathbf{u} \in \mathbb{R}^p$ y $\mathbf{v} \in \mathbb{R}^q$, las proyecciones $\mathbf{u}^T \mathbf{X}$ y $\mathbf{v}^T \mathbf{Y}$ definen el vector aleatorio bidimensional $(\mathbf{u}^T \mathbf{X}, \mathbf{v}^T \mathbf{Y})^T$. Las direcciones, variables y correlaciones canónicas derivadas de un funcional de asociación ρ se definen maximizando la asociación entre proyecciones. Más precisamente, en Alfons *et al.* (2016) se definen las direcciones canónicas como los vectores $(\mathbf{u}_k, \mathbf{v}_k) = (\mathbf{u}_k(P), \mathbf{v}_k(P))$ tales que

$$\begin{aligned} (\mathbf{u}_1, \mathbf{v}_1) &= \operatorname{argmax}_{\|\mathbf{u}\|_p = \|\mathbf{v}\|_q = 1} \rho^2(P[\mathbf{u}^T \mathbf{X}, \mathbf{v}^T \mathbf{Y}]) \\ (\mathbf{u}_k, \mathbf{v}_k) &= \operatorname{argmax}_{(\mathbf{u}, \mathbf{v}) \in \mathcal{B}_k} \rho^2(P[\mathbf{u}^T \mathbf{X}, \mathbf{v}^T \mathbf{Y}]), \end{aligned} \quad (3.1)$$

donde $P[\mathbf{u}^T \mathbf{X}, \mathbf{v}^T \mathbf{Y}]$ indica la medida de probabilidad del vector $(\mathbf{u}^T \mathbf{X}, \mathbf{v}^T \mathbf{Y})^T \in \mathbb{R}^2$ inducida por P , $\|\cdot\|_p$ indica la norma euclídea en \mathbb{R}^p y $\mathcal{B}_k = \{(\mathbf{u}, \mathbf{v}) \in \mathbb{R}^p \times \mathbb{R}^q : \|\mathbf{u}\|_p = \|\mathbf{v}\|_q = 1 \text{ y } \rho(P[\mathbf{u}^T \mathbf{X}, \mathbf{u}_j^T \mathbf{X}]) = \rho(P[\mathbf{v}^T \mathbf{Y}, \mathbf{v}_j^T \mathbf{Y}]) = 0, \text{ para } 1 \leq j \leq k-1\}$. Alfons *et al.* (2016) dan condiciones bajo las cuales los funcionales $(\mathbf{u}_1(P), \mathbf{v}_1(P))$ corresponden a la misma cantidad para distintos funcionales de asociación y analiza cuando resultan Fisher-consistentes.

Las variables $\mathbf{u}_j^T \mathbf{X}$ y $\mathbf{v}_j^T \mathbf{Y}$ se denominan las j -ésimas variables canónicas de \mathbf{X} e \mathbf{Y} , respectivamente, mientras que la j -ésima correlación canónica, $\rho_j \geq 0$, corresponde al valor máximo de correlación obtenido en la j -ésima maximización, es decir, ρ_j cumple

$$\rho_j^2 = \rho^2(P[\mathbf{u}_j^T \mathbf{X}, \mathbf{v}_j^T \mathbf{Y}]). \quad (3.2)$$

El análisis de correlación canónica, consiste justamente en la búsqueda consecutiva de direcciones, variables y correlaciones canónicas, es decir, en maximizar subsecuentemente la correlación entre proyecciones de \mathbf{X} y de \mathbf{Y} con la restricción de que las nuevas proyecciones sean no correlacionadas con las variables canónicas ya encontradas.

Sea $(\mathbf{X}_1, \mathbf{Y}_1), (\mathbf{X}_2, \mathbf{Y}_2), \dots, (\mathbf{X}_n, \mathbf{Y}_n)$ una muestra aleatoria de (\mathbf{X}, \mathbf{Y}) . Sea $P_n[\mathbf{u}^T \mathbf{X}, \mathbf{v}^T \mathbf{Y}]$ la distribución empírica del vector aleatorio $(\mathbf{u}^T \mathbf{X}, \mathbf{v}^T \mathbf{Y})^T$. Más precisamente, si \mathbb{I}_A es la función indicadora del conjunto boreliano $A \subset \mathbb{R}^2$,

$$P_n[\mathbf{u}^T \mathbf{X}, \mathbf{v}^T \mathbf{Y}](A) = \frac{1}{n} \sum_{i=1}^n \mathbb{I}_A(\mathbf{u}^T \mathbf{X}_i, \mathbf{v}^T \mathbf{Y}_i).$$

A partir de la definición poblacional dada anteriormente, está claro que podemos obtener estimadores de las correlaciones canónicas y las direcciones canónicas reemplazando en (3.1) y (3.2) la medida P por la empírica P_n . Es decir, los estimadores de las direcciones canónicas se obtienen como

$$\begin{aligned} (\hat{\mathbf{u}}_1, \hat{\mathbf{v}}_1) &= \operatorname{argmax}_{\|\mathbf{u}\|_p = \|\mathbf{v}\|_q = 1} \rho^2(P_n[\mathbf{u}^T \mathbf{X}, \mathbf{v}^T \mathbf{Y}]) \\ (\hat{\mathbf{u}}_k, \hat{\mathbf{v}}_k) &= \operatorname{argmax}_{(\mathbf{u}, \mathbf{v}) \in \hat{\mathcal{B}}_k} \rho^2(P_n[\mathbf{u}^T \mathbf{X}, \mathbf{v}^T \mathbf{Y}]), \end{aligned} \quad (3.3)$$

donde $\hat{\mathcal{B}}_k = \{(\mathbf{u}, \mathbf{v}) \in \mathbb{R}^p \times \mathbb{R}^q : \|\mathbf{u}\|_p = \|\mathbf{v}\|_q = 1 \text{ y } \rho(P_n[\mathbf{u}^T \mathbf{X}, \hat{\mathbf{u}}_j^T \mathbf{X}]) = \rho(P_n[\mathbf{v}^T \mathbf{Y}, \hat{\mathbf{v}}_j^T \mathbf{Y}]) = 0, \text{ para } 1 \leq j \leq k-1\}$. Los estimadores de las j -ésimas variables canónicas son $\hat{\mathbf{u}}_j^T \mathbf{X}$ y $\hat{\mathbf{v}}_j^T \mathbf{Y}$, mientras que el estimador $\hat{\rho}_j \geq 0$ de la j -ésima correlación canónica cumple la siguiente ecuación

$$\hat{\rho}_j^2 = \rho^2(P_n[\hat{\mathbf{u}}_j^T \mathbf{X}, \hat{\mathbf{v}}_j^T \mathbf{Y}]). \quad (3.4)$$

Cuando $\rho = \rho_{\text{CL}}$ es el coeficiente de correlación de Pearson, la distribución exacta de los estimadores de correlaciones y direcciones canónicas puede verse, por ejemplo, en Muirhead (1982), mientras que su consistencia y distribución asintótica se estudia en Muirhead y Waternaux (1980) y Muirhead (1982).

En esta tesis, consideraremos el caso de elementos aleatorios a valores en un espacio de Hilbert de dimensión infinita. Nuestro objeto de estudio será el análisis de correlación canónica funcional que describiremos a continuación. De ahora en más, \mathcal{H}_j , $j = 1, 2$ indicarán espacios de Hilbert separables con producto interno $\langle \cdot, \cdot \rangle_{\mathcal{H}_j}$ y norma $\|\cdot\|_{\mathcal{H}_j}$.

En ocasiones, para alivianar notación cuando el contexto sea claro, notaremos $\langle \cdot, \cdot \rangle$ en vez de $\langle \cdot, \cdot \rangle_{\mathcal{H}_j}$. De la misma manera usaremos $\| \cdot \|$ en vez de $\| \cdot \|_{\mathcal{H}_j}$.

Sean $\mathcal{S}_j = \{u \in \mathcal{H}_j : \|u\|_{\mathcal{H}_j} = 1\}$, $j = 1, 2$ las esferas en estos espacios. Si $(X, Y)^T : \Omega \rightarrow \mathcal{H}_1 \times \mathcal{H}_2$, es un elemento aleatorio con medida de probabilidad P , se pueden definir las direcciones y correlaciones canónicas del mismo modo que en (3.1) y (3.2) reemplazando el producto interno en el espacio euclídeo por el correspondiente producto interno en el espacio de Hilbert. Denotamos con $P[\langle u, X \rangle_{\mathcal{H}_1}, \langle v, Y \rangle_{\mathcal{H}_2}]$ a la medida de probabilidad del vector $(\langle u, X \rangle_{\mathcal{H}_1}, \langle v, Y \rangle_{\mathcal{H}_2}) \in \mathbb{R}^2$, inducida por P . A partir de una medida de asociación ρ , las direcciones canónicas funcionales pueden ser definidas de la siguiente manera:

$$\begin{aligned} (\Phi_1, \Psi_1) &= \operatorname{argmax}_{\|u\|_{\mathcal{H}_1}=\|v\|_{\mathcal{H}_2}=1} \rho^2(P[\langle u, X \rangle_{\mathcal{H}_1}, \langle v, Y \rangle_{\mathcal{H}_2}]) \\ (\Phi_k, \Psi_k) &= \operatorname{argmax}_{(u,v) \in \mathcal{B}_k} \rho^2(P[\langle u, X \rangle_{\mathcal{H}_1}, \langle v, Y \rangle_{\mathcal{H}_2}]), \end{aligned} \quad (3.5)$$

donde $\mathcal{B}_k = \{(u, v) \in \mathcal{S}_1 \times \mathcal{S}_2 : \rho(P[\langle u, X \rangle_{\mathcal{H}_1}, \langle \Phi_j, X \rangle_{\mathcal{H}_1}]) = \rho(P[\langle v, Y \rangle_{\mathcal{H}_2}, \langle \Psi_j, Y \rangle_{\mathcal{H}_2}]) = 0, \text{ para } 1 \leq j \leq k-1\}$. Como en el caso euclídeo, la j -ésima correlación canónica, $\rho_j \geq 0$, cumple la ecuación

$$\rho_j^2 = \rho^2(P[\langle \Phi_j, X \rangle_{\mathcal{H}_1}, \langle \Psi_j, Y \rangle_{\mathcal{H}_2}]). \quad (3.6)$$

Observemos que si la medida de asociación ρ cumple la condición (ii) dada en la Sección 2.3 tenemos que $\rho(-U, V) = -\rho(U, V)$, por lo tanto si se definen las direcciones canónicas como aquellas que maximizan en (3.5) la correlación en lugar de la correlación al cuadrado se tiene que $\rho(P[\langle \Phi_j, X \rangle_{\mathcal{H}_1}, \langle \Psi_j, Y \rangle_{\mathcal{H}_2}]) \geq 0$.

Cuando $\rho = \rho_{\text{CL}}$ y $\mathcal{H}_j = L^2([0, 1])$, Leurgans *et al.* (1993) muestran que, a diferencia del caso euclídeo, para definir estimadores de las direcciones y correlaciones canónicas es necesario introducir un factor de penalización, es decir, es necesario suavizar. Estos autores probaron que si $(X, Y)^T$ es un proceso Gaussiano y si los operadores de covarianza Γ_{XX} y Γ_{YY} son genuinamente infinito dimensionales, es decir, si tienen infinitos autovalores distintos de cero, entonces con probabilidad uno, la muestra $(X_1, Y_1)^T, \dots, (X_n, Y_n)^T$ admitirá direcciones

u y v (dependientes de la muestra) tales que $\langle u, X_i \rangle_{\mathcal{H}_1} = \langle v, Y_i \rangle_{\mathcal{H}_2}$, para todo $1 \leq i \leq n$, con lo cual las proyecciones $X_u = \langle u, X \rangle_{\mathcal{H}_1}$ e $Y_v = \langle v, Y \rangle_{\mathcal{H}_2}$ tienen correlación de Pearson muestral igual a 1. Más aún, Leurgans *et al.* (1993) muestran que no sólo existen un par de funcionales lineales de los datos con correlación muestral igual a 1, sino que dado cualquier funcional lineal de X existe un funcional lineal de Y tal que la correlación clásica muestral de las proyecciones es igual a 1. Por lo tanto, cualquier función $u \in L^2([0, 1])$ resulta ser una dirección canónica con correlación 1. El hecho contradictorio surge de la observación que, para la correlación de Pearson, la primera correlación canónica poblacional resultará ser 1 solamente en el caso excepcional en que existan funciones $u \in L^2([0, 1])$ y $v \in L^2([0, 1])$ tales que las variables aleatorias $X_u := \langle u, X \rangle_{\mathcal{H}_1}$ e $Y_v := \langle v, Y \rangle_{\mathcal{H}_2}$ mantengan una relación lineal perfecta. Es decir, si existen $u, v \in L^2([0, 1])$ y una recta \mathcal{R} en \mathbb{R}^2 tal que $\mathbb{P}((X_u, Y_v) \in \mathcal{R}) = 1$ y esto en general no ocurre.

Por lo tanto, si se quisiera definir el estimador de la primera correlación canónica poblacional (3.6) como en (3.3) y (3.4), dichos estimadores carecerían de sentido y no serían consistentes a menos que $\rho_1 = 1$. Esto muestra que algún tipo de suavizado es esencial.

En esta sección, mostraremos que el problema de definir los estimadores como en (3.3) y (3.4) persiste aún con hipótesis más débiles sobre el elemento aleatorio siempre que el funcional de asociación ρ definido en \mathbb{R}^2 sea tal que $\rho(Q) = 1$ para toda medida bivariada Q concentrada en la recta $\mathcal{R}_{a,b} = \{(x_1, x_2) \in \mathbb{R}^2 : x_2 = ax_1 + b\}$. Por lo tanto, el estimador de la primera correlación canónica no resulta consistente si $\rho_1 < 1$. Medidas de asociación que cumplen el requerimiento anterior incluyen, por ejemplo, el coeficiente de correlación de Pearson, de Spearman, la correlación de distancias introducida por Székely *et al.* (2007), medidas de correlación definidas a partir de estimadores robustos de la matriz de dispersión del vector bivariado y la correlación de medianas. Para todas excepto la última vale la recíproca, es decir, si $\rho(Q) = 1$ entonces existen a y b tales que $Q(\mathcal{R}_{a,b}) = 1$. La recíproca tampoco vale para el coeficiente de correlación de Kendall como muestran Alfons *et al.*

(2016).

Con las mismas herramientas usadas para probar que no es posible extender de manera directa el método de estimar del caso multivariado al funcional, llegamos a otras dos importantes aplicaciones. Para la primera aplicación, sean $\mathcal{H}_{1,p}$ y $\mathcal{H}_{2,q}$ subespacios de dimensiones p y q de \mathcal{H}_1 y \mathcal{H}_2 , respectivamente. Por ejemplo, $\mathcal{H}_{1,p}$ puede ser el subespacio generado por los primeros p elementos de una base ortonormal de \mathcal{H}_1 . En particular, podría tomarse como base la base de direcciones principales, es decir, si X es un elemento aleatorio con segundo momento finito, la base de autofunciones del operador de covarianza Γ_{XX} . Si tomamos $p = p_n$ y $q = q_n$ tales que $p_n \rightarrow \infty$, $q_n \rightarrow \infty$, la sucesión de subespacios crecientes \mathcal{H}_{1,p_n} y \mathcal{H}_{2,q_n} permite aproximar los espacios \mathcal{H}_1 y \mathcal{H}_2 , respectivamente. Definamos también $\mathcal{S}_{1,p} = \mathcal{S}_1 \cap \mathcal{H}_{1,p}$ y $\mathcal{S}_{2,q} = \mathcal{S}_2 \cap \mathcal{H}_{2,q}$ donde, recordemos, $\mathcal{S}_j = \{u \in \mathcal{H}_j : \|u\|_{\mathcal{H}_j} = 1\}$, $j = 1, 2$. Llamamos estimador basado en Sieves o estimador de Sieves a secas a un estimador definido de la siguiente manera:

$$\begin{aligned} (\widehat{\Phi}_{\text{SI},1}, \widehat{\Psi}_{\text{SI},1}) &= \operatorname{argmax}_{u \in \mathcal{S}_{1,p_n}, v \in \mathcal{S}_{2,q_n}} \rho^2(P_n[\langle u, X \rangle_{\mathcal{H}_1}, \langle v, Y \rangle_{\mathcal{H}_2}]) \\ (\widehat{\Phi}_{\text{SI},k}, \widehat{\Psi}_{\text{SI},k}) &= \operatorname{argmax}_{(u,v) \in \mathcal{B}_{k,p_n,q_n}} \rho^2(P_n[\langle u, X \rangle_{\mathcal{H}_1}, \langle v, Y \rangle_{\mathcal{H}_2}]), \end{aligned} \quad (3.7)$$

donde

$$\begin{aligned} \mathcal{B}_{k,p_n,q_n} &= \{(u, v) \in \mathcal{S}_{1,p_n} \times \mathcal{S}_{2,q_n} : \rho(P_n[\langle u, X \rangle_{\mathcal{H}_1}, \langle \widehat{\Phi}_{\text{SI},j}, X \rangle_{\mathcal{H}_1}]) = \\ &= \rho(P_n[\langle v, Y \rangle_{\mathcal{H}_2}, \langle \widehat{\Psi}_{\text{SI},j}, Y \rangle_{\mathcal{H}_2}]) = 0, \text{ para } 1 \leq j \leq k-1\} \end{aligned}$$

y el estimador de la j -ésima correlación canónica $\widehat{\rho}_j \geq 0$ es tal que

$$\widehat{\rho}_j^2 = \rho^2 \left(P_n[\langle \widehat{\Phi}_{\text{SI},j}, X \rangle_{\mathcal{H}_1}, \langle \widehat{\Psi}_{\text{SI},j}, Y \rangle_{\mathcal{H}_2}] \right). \quad (3.8)$$

Mostraremos que es necesario realizar la maximización sobre espacios de dimensión finita ya que, bajo ciertos supuestos, $\sup_{u \in \mathcal{S}_1, v \in \mathcal{S}_2} \rho^2(P_n[\langle u, X \rangle_{\mathcal{H}_1}, \langle v, Y \rangle_{\mathcal{H}_2}]) = 1$. Por otra parte, veremos que $\widehat{\rho}_1 = 1$ siempre que $p_n \geq n$ o $q_n \geq n$. Vale la pena observar que, como se muestra en Bali *et al.* (2011), en el análisis de componentes principales los estimadores basados en

Sieves resultan consistentes cuando p_n es tal que $p_n \rightarrow \infty$, sin otras restricciones sobre la dimensión del espacio aproximante. Como consecuencia del requerimiento $p_n \leq n$ y $q_n \leq n$ que surge en el análisis de correlación canónica funcional para tener consistencia de los estimadores de Sieves, resulta también la necesidad de suavizar en los problemas de datos de alta dimensión, ésta sería la segunda aplicación a la que nos referíamos.

3.2. El estimador natural falla en el análisis de correlación canónica funcional

Sea $(\Omega, \mathcal{A}, \mathbb{P})$ un espacio de probabilidad y \mathcal{H} un espacio de Hilbert separable con producto interno $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ y norma $\| \cdot \|_{\mathcal{H}}$. El espacio \mathcal{H} puede ser tanto finito como infinito dimensional. Sea $X : \Omega \rightarrow \mathcal{H}$ un elemento aleatorio. En lo que sigue utilizaremos la siguiente condición

C1. $\mathbb{P}(X \in \mathcal{L}) = 0$ para todo subespacio propio \mathcal{L} de \mathcal{H} , con \mathcal{L} de dimensión finita.

Proposición 3.2.1. *Sea $(X_n)_{n \in \mathbb{N}}$ una sucesión de elementos aleatorios independientes e idénticamente distribuidos (i.i.d.) con la misma distribución que X y supongamos que X cumple **C1**. Para cada $n \in \mathbb{N}$, si $\dim(\mathcal{H}) \geq n$, entonces con probabilidad 1, $\{X_1, X_2, \dots, X_n\}$ es un conjunto linealmente independiente (l.i.).*

Demostración. Sea $n \leq \dim(\mathcal{H})$. Probaremos que con probabilidad 0, $\{X_1, X_2, \dots, X_n\}$ es un conjunto linealmente dependiente (l.d.). Indiquemos por $\langle u_1, \dots, u_k \rangle$ el subespacio generado por u_1, \dots, u_k . Sea $\mathcal{A} = \{\omega \in \Omega : \{X_1(\omega), X_2(\omega), \dots, X_n(\omega)\} \text{ es l.d.}\}$, luego $\mathcal{A} \subset \bigcup_{i=0}^n \mathcal{A}_i$, donde $\mathcal{A}_0 = \{\omega : X_i(\omega) = 0 \ \forall i\}$ y para $i \geq 1$, $\mathcal{A}_i = \{\omega \in \Omega : X_i(\omega) \in \langle X_1(\omega), \dots, X_{i-1}(\omega), X_{i+1}(\omega), \dots, X_n(\omega) \rangle\}$. Por un lado, tenemos que $\mathbb{P}(\mathcal{A}_0) \leq \mathbb{P}(X_1 = 0_{\mathcal{H}}) = 0$ ya que X_1 cumple **C1**. Por otro lado, para $1 \leq i \leq n$, como X_1, \dots, X_n son i.i.d., $\mathbb{P}(\mathcal{A}_i) = \mathbb{P}(\mathcal{A}_j) \ \forall i \neq j$, con lo cual basta probar que $\mathbb{P}(\mathcal{A}_1) = 0$.

Los elementos aleatorios $X_1 : \Omega \rightarrow \mathcal{H}$ y $(X_1, \dots, X_n) : \Omega \rightarrow \mathcal{H}^n$ inducen medidas de

probabilidad P y P^n en \mathcal{H} y \mathcal{H}^n respectivamente. Del mismo modo, (X_2, \dots, X_n) induce la medida de probabilidad P^{n-1} sobre \mathcal{H}^{n-1} . Tenemos lo siguiente:

$$\mathbb{P}(\mathcal{A}_1) = \mathbb{P}(X_1 \in \langle X_2, \dots, X_n \rangle) = \mathbb{E}(\mathbb{P}(X_1 \in \langle X_2, \dots, X_n \rangle | (X_2, \dots, X_n))) = 0,$$

donde en la última igualdad hemos usado que para cada x_2, \dots, x_n fijos

$$\mathbb{P}(X_1 \in \langle X_2, \dots, X_n \rangle | (X_2, \dots, X_n) = (x_2, \dots, x_n)) = \mathbb{P}(X_1 \in \langle x_2, \dots, x_n \rangle) = 0,$$

y donde la primera igualdad de la ecuación anterior vale ya que X_1 es independiente de (X_2, \dots, X_n) y la segunda igualdad vale ya que al ser $n \leq \dim(\mathcal{H})$, $\langle x_2, \dots, x_n \rangle$ es un subespacio propio y de dimensión finita de \mathcal{H} y X_1 cumple **C1**. \square

Dada una sucesión de elementos aleatorios $(X_n)_{n \in \mathbb{N}}$ con $X_i : \Omega \rightarrow \mathcal{H}$. Se puede definir para cada $\omega \in \Omega$ y $n \in \mathbb{N}$ la función lineal $T_{n, X(\omega)} : \mathcal{H} \rightarrow \mathbb{R}^n$ de la siguiente manera:

$$T_{n, X(\omega)}(u) = (\langle u, X_1(\omega) \rangle_{\mathcal{H}}, \langle u, X_2(\omega) \rangle_{\mathcal{H}}, \dots, \langle u, X_n(\omega) \rangle_{\mathcal{H}})^T.$$

Corolario 3.2.1. *Sea $(X_n)_{n \in \mathbb{N}}$ una sucesión de elementos aleatorios en \mathcal{H} i.i.d. con la misma distribución que X y supongamos que X cumple **C1**. Para cada $n \in \mathbb{N}$, si $\dim(\mathcal{H}) \geq n$, entonces $\mathbb{P}(\{\omega : T_{n, X(\omega)} \text{ es sobreyectiva}\}) = 1$.*

Demostración. Sea $n \leq \dim(\mathcal{H})$ y definamos, como en la demostración de la Proposición 3.2.1, $\mathcal{A}^c = \{\omega \in \Omega : \{X_1(\omega), X_2(\omega), \dots, X_n(\omega)\} \text{ es l.i.}\}$. Por la Proposición 3.2.1, sabemos que $\mathbb{P}(\mathcal{A}^c) = 1$. Probaremos que para $\omega \in \mathcal{A}^c$, $T_{n, X(\omega)}$ es sobreyectiva. Sea $\omega \in \mathcal{A}^c$, y sea $\mathcal{L}_n(\omega) = \langle X_1(\omega), X_2(\omega), \dots, X_n(\omega) \rangle$. $\mathcal{L}_n(\omega)$ tiene dimensión n por estar ω en \mathcal{A}^c . Bastará ver que si restringimos $T_{n, X(\omega)}$ a $\mathcal{L}_n(\omega)$ resulta inyectiva. En tal caso, la imagen de $T_{n, X(\omega)}$ tendrá que tener dimensión al menos n y será por lo tanto sobreyectiva.

Supongamos que $T_{n, X(\omega)}(\sum_{i=1}^n \alpha_i X_i(\omega)) = T_{n, X(\omega)}(\sum_{i=1}^n \beta_i X_i(\omega))$, entonces, para todo $j = 1, \dots, n$, $\langle \sum_{i=1}^n (\alpha_i - \beta_i) X_i(\omega), X_j(\omega) \rangle_{\mathcal{H}} = 0$. Luego, $\sum_{i=1}^n (\alpha_i - \beta_i) X_i(\omega) \in \mathcal{L}_n(\omega) \cap$

$(\mathcal{L}_n(\omega))^c = 0_{\mathcal{H}}$. Por lo tanto, $\sum_{i=1}^n \alpha_i X_i(\omega) = \sum_{i=1}^n \beta_i X_i(\omega)$ con lo cual $T_{n,X(\omega)}$ restringida a $\mathcal{L}_n(\omega)$ es inyectiva y $T_{n,X(\omega)}$ es sobreyectiva como queríamos probar. \square

Sea $(X, Y)^T : \Omega \rightarrow \mathcal{H}_1 \times \mathcal{H}_2$ un elemento aleatorio bivariado, donde \mathcal{H}_j son espacios de Hilbert separables. Podemos pensar, para fijar ideas, que \mathcal{H}_j es $L^2(\mathcal{I}_j)$ con \mathcal{I}_j un intervalo acotado de la recta real.

Sea $(X_1, Y_1)^T, \dots, (X_n, Y_n)^T$ una muestra aleatoria del elemento aleatorio $(X, Y)^T$. Dados $(u, v) \in \mathcal{H}_1 \times \mathcal{H}_2$, sea

$$\rho_n(u, v) := \rho(P_n[\langle u, X \rangle_{\mathcal{H}_1}, \langle v, Y \rangle_{\mathcal{H}_2}]), \quad (3.9)$$

es decir, la versión empírica de $\rho(P[\langle u, X \rangle_{\mathcal{H}_1}, \langle v, Y \rangle_{\mathcal{H}_2}])$. Si los elementos aleatorios $(X_i, Y_i)^T$ tienen dominio en Ω , para $\omega \in \Omega$, cuando sea necesario, notaremos

$$\rho_{n,\omega}(u, v) := \rho(P_{n,\omega}[\langle u, X \rangle_{\mathcal{H}_1}, \langle v, Y \rangle_{\mathcal{H}_2}]),$$

donde $P_{n,\omega}(A) = (1/n) \sum_{i=1}^n \mathbb{I}_A(X_i(\omega), Y_i(\omega))$.

Corolario 3.2.2. *Supongamos que para todo $v \in \mathcal{H}_2$, $\mathbb{P}(Y = v) = 0$ y que X satisface la condición **C1**. Sea $d = \dim(\mathcal{H}_1)$. Sea ρ una medida de correlación bivariada tal que $\rho(Q) = 1$ si Q está concentrada en la recta $\mathcal{R} = \{(x_1, x_2) \in \mathbb{R}^2 : x_2 = x_1\}$ y $Q(\{x_1 = c\}) = Q(\{x_2 = c\}) < 1/2$ para todo $c \in \mathbb{R}$. Sea $n \in \mathbb{N}$, $2 < n \leq d$ y $(X_1, Y_1)^T, \dots, (X_n, Y_n)^T$ elementos aleatorios i.i.d. con la misma distribución que $(X, Y)^T$. Entonces, con probabilidad uno existen funciones u_n y v_n en \mathcal{H}_1 y \mathcal{H}_2 , respectivamente, tales que $\rho_n(u_n, v_n) = 1$.*

Demostración. Como X cumple **C1** y para todo $v \in \mathcal{H}_2$, $\mathbb{P}(Y = v) = 0$, existe un conjunto $\mathcal{A} \subset \Omega$ con $\mathbb{P}(\mathcal{A}) = 1$ tal que si $\omega \in \mathcal{A}$, $Y_i(\omega) \neq Y_j(\omega) \quad \forall i \neq j$ y además $T_{n,X(\omega)}$ es sobreyectiva. Sea $\omega \in \mathcal{A}$, si $\langle v, Y_i(\omega) \rangle_{\mathcal{H}_2} = \langle v, Y_j(\omega) \rangle_{\mathcal{H}_2}$ entonces $v \in S_{ij}(\omega) = \{Y_i(\omega) - Y_j(\omega)\}^\perp$. Como $Y_i(\omega) \neq Y_j(\omega)$, entonces $S_{ij}(\omega)$ es un subespacio propio de \mathcal{H}_2 , y $S(\omega) = \bigcup_{1 \leq i < j \leq n} S_{ij}(\omega) \neq \mathcal{H}_2$, por ser una unión finita de subespacios propios, o sea de nunca densos. Sea $v_n \in \mathcal{H}_2 - S(\omega)$, entonces $\mathbf{z} = T_{n,Y(\omega)}(v_n)$ tiene n valores distintos. Como $T_{n,X(\omega)}$ es sobreyectiva, existe $u_n \in \mathcal{H}_1$

tal que $T_{n, X(\omega)}(u_n) = \mathbf{z}$. Luego, si llamamos $Q_{n, \omega} = P_{n, \omega}[\langle u_n, X \rangle_{\mathcal{H}_1}, \langle v_n, Y \rangle_{\mathcal{H}_2}]$, se cumple que $Q_{n, \omega}(\mathcal{R}) = 1$. Además, si $\mathbf{z} = (z_1, z_2, \dots, z_n)$, $Q_{n, \omega}(\{x_1 = z_i\}) = Q_{n, \omega}(\{x_2 = z_i\}) = 1/n$ y $Q_{n, \omega}(\{x_1 = c\}) = Q_{n, \omega}(\{x_2 = c\}) = 0$ si $c \neq z_i$ para todo i . Luego $\rho(Q_{n, \omega}) = 1$, o lo que es lo mismo $\rho_{n, \omega}(u_n, v_n) = 1$. \square

Observación 3.2.1. *En el corolario 3.2.2, si $\dim(\mathcal{H}_1) = \infty$, entonces $\sup_{u \in \mathcal{S}_1, v \in \mathcal{S}_2} \rho_n^2(u, v) = 1$ para todo $n \in \mathbb{N}$, por lo que algún tipo de suavizado es necesario.*

Observación 3.2.2. *En el corolario 3.2.2, las hipótesis que se piden sobre los elementos aleatorios X e Y son intercambiables, esto es podría pedirse que para todo $u \in \mathcal{H}_1$, $\mathbb{P}(X = u) = 0$ y que Y satisfaga la condición **C1** y la tesis sería la misma.*

A continuación, encontraremos un amplio conjunto de procesos bivariados $(X, Y)^T$ para los cuales vale **C1** tanto para X como para Y . Primero daremos una nueva condición que resultará más fuerte que **C1** pero permite verificar **C1** fácilmente. Dada una base ortonormal $B = \{\delta_1, \delta_2, \delta_3, \dots\}$ de \mathcal{H} , definimos las funciones $A_{B, k} : \mathcal{H} \rightarrow \mathbb{R}^k$ como

$$A_{B, k}(x) = (\langle x, \delta_1 \rangle_{\mathcal{H}}, \langle x, \delta_2 \rangle_{\mathcal{H}}, \dots, \langle x, \delta_k \rangle_{\mathcal{H}})^T.$$

Dado $X : \Omega \rightarrow \mathcal{H}$, decimos que cumple la condición **C2** si:

C2. Para cualquier base ortonormal $B = \{\delta_1, \delta_2, \delta_3, \dots\}$ de \mathcal{H} y cualquier $k \in \mathbb{N}$ tal que $k \leq \dim(\mathcal{H})$, el vector aleatorio k -dimensional $A_{B, k}(X)$ es un vector aleatorio que admite una función de densidad.

Proposición 3.2.2. *Si $X : \Omega \rightarrow \mathcal{H}$ satisface **C2** entonces satisface **C1**.*

Demostración. Sea \mathcal{L} un subespacio propio de dimensión k de \mathcal{H} , es decir, $k < \dim(\mathcal{H})$ e indiquemos por $\mathbf{e}_1, \dots, \mathbf{e}_{k+1}$ la base canónica de \mathbb{R}^{k+1} .

Sea $B = \{\delta_1, \delta_2, \delta_3, \dots\}$ una base ortonormal de \mathcal{H} tal que $\{\delta_1, \dots, \delta_k\}$ es una base ortonormal de \mathcal{L} . Consideremos el operador $A_{B, k+1}$ y observemos que, si $x \in \mathcal{L}$, la coordenada

$k + 1$ del vector $(k + 1)$ -dimensional $A_{B,k+1}(x)$ es 0, es decir, $\mathbf{e}_{k+1}^T A_{B,k+1}(x) = 0$. Por otra parte, por **C2**, como $k + 1 \leq \dim(\mathcal{H})$, el vector aleatorio $A_{B,k+1}(X)$ admite densidad. Indiquemos por $\mathbf{e}_{k+1}^\perp = \{\mathbf{x} = (x_1, \dots, x_{k+1})^T \in \mathbb{R}^{k+1} : x_{k+1} = 0\}$ al subespacio propio de \mathbb{R}^{k+1} ortogonal a \mathbf{e}_{k+1} . Luego, se cumple que $\mathbb{P}(A_{B,k+1}(X) \in \mathbf{e}_{k+1}^\perp) = 0$ y si $x \in \mathcal{L}$ entonces $A_{B,k+1}(x) \in \mathbf{e}_{k+1}^\perp$, de donde $\mathbb{P}(X \in \mathcal{L}) \leq \mathbb{P}(A_{B,k+1}(X) \in \mathbf{e}_{k+1}^\perp) = 0$, como queríamos probar. \square

Sea ahora $(\tilde{X}, \tilde{Y})^T$ un proceso bivariado Gaussiano en $\mathcal{H}_1 \times \mathcal{H}_2$ con media cero y operador de covarianza Γ . Sean Γ_{11} y Γ_{22} los operadores de covarianza de \tilde{X} y \tilde{Y} , respectivamente y supongamos que el núcleo de Γ_{jj} se reduce al $\{0\}$ para $j = 1, 2$. En este caso, los operadores Γ_{11} y Γ_{22} tienen rango infinito, o sea, que tienen infinitos autovalores no nulos. Sea Z una variable aleatoria positiva, es decir, $\mathbb{P}(Z > 0) = 1$, independiente de (\tilde{X}, \tilde{Y}) . Luego, el elemento aleatorio $(X, Y)^T = Z(\tilde{X}, \tilde{Y})^T$ es un proceso bivariado que es una mezcla de un proceso Gaussiano, por lo tanto, es un proceso elíptico $\mathcal{E}(0, \Gamma)$ como fue definido en Bali y Boente (2009). Más aún, de acuerdo a la caracterización dada en Boente *et al.* (2014) todo proceso elíptico en \mathcal{H} con operador de covarianza de rango infinito es de esta forma ya que es una mezcla de un proceso Gaussiano. Ahora bien, dada una base $B = \{\delta_1, \delta_2, \delta_3, \dots\}$ ortonormal de \mathcal{H}_1 , como el núcleo de Γ_{11} se reduce al $\{0\}$, el vector $A_{B,k}(\tilde{X})$ tiene distribución normal multivariada no singular. Por lo tanto, como $A_{B,k}(X) = Z A_{B,k}(\tilde{X})$, $A_{B,k}(X)$ tiene densidad (ver Muirhead, 1982), es decir, X satisface **C2** y por ende **C1**. En forma similar se obtiene un resultado análogo para Y .

En particular, como los procesos elípticos incluyen a los procesos Gaussianos que fueron considerados por Leurgans *et al.* (1993), el resultado obtenido generaliza el dado por estos autores en dos sentidos. Por un lado considerando otras medidas de probabilidad además de la Gaussiana y por otra parte, ampliando las posibles medidas de asociación más allá de la correlación de Pearson. En cuanto a los posibles funcionales ρ para las cuales no hay consistencia para estos “estimadores naturales”, como hemos dicho, basta que se cumpla la

condición $\rho(Q) = 1$ si Q está concentrada en la recta $\mathcal{R} = \{(x_1, x_2) \in \mathbb{R}^2 : x_2 = x_1\}$ y $Q(\{x_1 = c\}) = Q(\{x_2 = c\}) < 1/2$ para todo $c \in \mathbb{R}$.

3.3. Condiciones para la consistencia de los estimadores basados en Sieves

Por simplicidad de notación consideremos $(X, \mathbf{Y}) : \Omega \rightarrow \mathcal{H} \times \mathbb{R}^q$ donde $\dim(\mathcal{H}) = \infty$. Sea $B = \{\delta_1, \delta_2, \dots\}$ una base ortonormal de \mathcal{H} y $B' = \{\boldsymbol{\eta}_1, \dots, \boldsymbol{\eta}_q\}$ una base ortonormal de \mathbb{R}^q . Sean $\mathcal{H}_{p_n} = \langle \delta_1, \dots, \delta_{p_n} \rangle$, \mathcal{S}_{1,p_n} la esfera unidad de \mathcal{H}_{p_n} y $\mathcal{S}_{2,q}$ la esfera unidad de \mathbb{R}^q . Definimos los estimadores de Sieves como en (3.7), pero teniendo en cuenta que $\mathcal{H}_2 = \mathbb{R}^q$ es de dimensión finita. Por lo tanto, suponiendo $q_n = q \ \forall n$, los estimadores quedan definidos por

$$\begin{aligned} (\widehat{\Phi}_{\text{SI},1}, \widehat{\Psi}_{\text{SI},1}) &= \operatorname{argmax}_{u \in \mathcal{S}_{1,p_n}, \mathbf{v} \in \mathcal{S}_{2,q}} \rho^2(P_n[\langle u, X \rangle_{\mathcal{H}}, \langle \mathbf{v}, \mathbf{Y} \rangle_q]) \\ (\widehat{\Phi}_{\text{SI},k}, \widehat{\Psi}_{\text{SI},k}) &= \operatorname{argmax}_{(u, \mathbf{v}) \in \mathcal{B}_{k,p_n}} \rho^2(P_n[\langle u, X \rangle_{\mathcal{H}}, \langle \mathbf{v}, \mathbf{Y} \rangle_q]), \end{aligned}$$

donde, $\mathcal{B}_{k,p_n} = \{(u, \mathbf{v}) \in \mathcal{S}_{1,p_n} \times \mathcal{S}_{2,q} : \rho(P_n[\langle u, X \rangle_{\mathcal{H}}, \langle \widehat{\Phi}_{\text{SI},j}, X \rangle_{\mathcal{H}}]) = \rho(P_n[\langle \mathbf{v}, \mathbf{Y} \rangle_q, \langle \widehat{\Psi}_{\text{SI},j}, \mathbf{Y} \rangle_q]) = 0, \text{ para } 1 \leq j \leq k-1\}$. El estimador del coeficiente de correlación canónica, $\widehat{\rho}_{\text{SI},j} \geq 0$, se define como

$$\widehat{\rho}_{\text{SI},j}^2 = \rho^2(P_n[\langle \widehat{\Phi}_{\text{SI},j}, X \rangle_{\mathcal{H}}, \langle \widehat{\Psi}_{\text{SI},j}, \mathbf{Y} \rangle_q]). \quad (3.10)$$

Con la intención de probar que este tipo de estimadores de Sieves no son consistentes si $p_n \geq n$, restringimos $T_{n,X(\omega)}$ a \mathcal{H}_{p_n} y probamos el siguiente resultado que sería el análogo del Corolario 3.2.1 en este contexto.

Proposición 3.3.1. *Sea $(X_n)_{n \in \mathbb{N}}$ una sucesión de elementos aleatorios en \mathcal{H} i.i.d. con la misma distribución que X . Si $p_n \geq n$ y X satisface la condición **C2** con $\dim(\mathcal{H}) = \infty$, entonces $T_{n,X(\omega)} : \mathcal{H}_{p_n} \rightarrow \mathbb{R}^n$ es sobreyectiva con probabilidad 1.*

Demostración. Observemos primero que si $u \in \mathcal{H}_{p_n}$

$$T_{n,X(\omega)}(u) = (\langle u, \Pi_{\mathcal{H}_{p_n}} X_1(\omega) \rangle_{\mathcal{H}}, \langle u, \Pi_{\mathcal{H}_{p_n}} X_2(\omega) \rangle_{\mathcal{H}}, \dots, \langle u, \Pi_{\mathcal{H}_{p_n}} X_n(\omega) \rangle_{\mathcal{H}})^{\mathsf{T}},$$

donde $\Pi_{\mathcal{H}_{p_n}}$ es el operador de proyección ortogonal sobre \mathcal{H}_{p_n} .

Sea $Z = \Pi_{\mathcal{H}_{p_n}} X$ y $Z_i = \Pi_{\mathcal{H}_{p_n}} X_i$. Luego, $Z = \sum_{j=1}^{p_n} \tilde{Z}_j \delta_j$, donde $\tilde{Z}_j = \langle X, \delta_j \rangle_{\mathcal{H}}$. Análogamente, podemos escribir $Z_i = \sum_{j=1}^{p_n} \tilde{Z}_{i,j} \delta_j$ y $u = \sum_{j=1}^{p_n} \tilde{u}_j \delta_j$. Por lo tanto, si $\tilde{\mathbf{Z}} = (\tilde{Z}_1, \dots, \tilde{Z}_{p_n})^{\mathsf{T}}$, $\tilde{\mathbf{Z}}_i = (\tilde{Z}_{i,1}, \dots, \tilde{Z}_{i,p_n})^{\mathsf{T}}$ y $\tilde{\mathbf{u}} = (\tilde{u}_1, \dots, \tilde{u}_{p_n})^{\mathsf{T}}$, tenemos que $T_{n,X}(u) = (\tilde{\mathbf{u}}^{\mathsf{T}} \tilde{\mathbf{Z}}_1, \dots, \tilde{\mathbf{u}}^{\mathsf{T}} \tilde{\mathbf{Z}}_n)^{\mathsf{T}} = T_{n,\tilde{\mathbf{Z}}}(\tilde{\mathbf{u}})$. Como la aplicación $u \mapsto \tilde{\mathbf{u}}$ es una biyección entre \mathcal{H}_{p_n} y \mathbb{R}^{p_n} , basta con mostrar que con probabilidad 1, $T_{n,\tilde{\mathbf{Z}}} : \mathbb{R}^{p_n} \rightarrow \mathbb{R}^n$ es sobreyectiva. Como X satisface **C2**, entonces $\tilde{\mathbf{Z}} = (\langle X, \delta_1 \rangle_{\mathcal{H}}, \dots, \langle X, \delta_{p_n} \rangle_{\mathcal{H}})$ es un vector aleatorio p_n -dimensional con función de densidad, y por lo tanto $\tilde{\mathbf{Z}}$ satisface **C1**. Luego, el Corolario 3.2.1 junto con $p_n \geq n$ implican que $T_{n,\tilde{\mathbf{Z}}} : \mathbb{R}^{p_n} \rightarrow \mathbb{R}^n$ es sobreyectiva con probabilidad 1, lo que concluye la demostración. \square

Como corolario obtenemos que, bajo ciertas condiciones, la estimación basada en un método de Sieves no resulta consistente cuando $p_n \geq n$. El siguiente resultado enuncia esta propiedad en el caso que estamos considerando.

Corolario 3.3.1. *Sea $(X, \mathbf{Y}) : \Omega \rightarrow \mathcal{H} \times \mathbb{R}^q$ tal que X satisface **C2**, $\mathbb{P}(\mathbf{Y} = \mathbf{a}) = 0$, para todo $\mathbf{a} \in \mathbb{R}^q$. Supongamos que $p_n \geq n$ para infinitos valores de n y que $\rho_1 < 1$, donde ρ_1 está definido en (3.6) para una medida de correlación bivariada ρ tal que $\rho(Q) = 1$ cuando Q está concentrada en la recta $\mathcal{R} = \{(x_1, x_2) \in \mathbb{R}^2 : x_2 = x_1\}$ y $Q(\{x_1 = c\}) = Q(\{x_2 = c\}) < 1$ para todo $c \in \mathbb{R}$. Entonces el estimador $\hat{\rho}_{\text{SI},1}$ basado en sieves definido en (3.10) no es consistente.*

Demostración. Sea $\Omega_n := \{w \in \Omega : T_{n,X(w)} : \mathcal{H}_{p_n} \rightarrow \mathbb{R}^n \text{ es sobreyectiva}\}$ y $\tilde{\Omega} = \bigcap_{n \in \mathbb{N}: p_n \geq n} \Omega_n$. Por la Proposición 3.3.1, $\tilde{\Omega}$ tiene probabilidad 1. Fijemos $\omega \in \tilde{\Omega}$ y sea $n \in \mathbb{N}$ tal que $p_n \geq n$, luego, argumentos análogos a los utilizados en el Corolario 3.2.2 permiten mostrar la existen-

cia de $u_n \in \mathcal{H}_{p_n}$ y $\mathbf{v} \in \mathbb{R}^q$ tales que $\rho_n(u_n, \mathbf{v}) = 1$. Por lo tanto, con probabilidad 1, $\widehat{\rho}_{1,SI} = 1$ para infinitos valores de n y el estimador no es consistente. \square

Observación 3.3.1. *Si en el Corolario 3.3.1 tenemos que $\dim(\mathcal{H}_2) = \infty$, $\widehat{\rho}_1$ definido como en (3.8) no es consistente si $p_n \geq n$ para infinitos valores de n o si $q_n \geq n$ para infinitos valores de n .*

3.4. Datos de alta dimensión

Se encuentran “datos de alta dimensión” en problemas en los que p es grande y n pequeño. El desarrollo de procedimientos de estimación en el contexto de alta dimensión surgió recientemente debido a grandes avances en el campo del almacenamiento y procesamiento de datos. Esta clase de problemas aparece, por ejemplo, en genética o en procesamiento de imágenes. En el primer caso, los datos son genes y tenemos pocas muestras de la misma persona con miles de genes. En el segundo, los datos son imágenes y hay miles de píxeles en cada imagen 3-d.

Para fijar ideas, supongamos que tenemos $(\mathbf{X}_1, \mathbf{Y}_1), (\mathbf{X}_2, \mathbf{Y}_2), \dots, (\mathbf{X}_n, \mathbf{Y}_n)$, una muestra aleatoria en $\mathbb{R}^{p_n} \times \mathbb{R}^q$ donde $p_n \geq n$. Supongamos que podemos pensar a $\mathbf{X} : \Omega \rightarrow \mathbb{R}^{p_n}$ como las primeras p_n coordenadas de un elemento aleatorio $x : \Omega \rightarrow \ell^2$ donde ℓ^2 es el espacio de sucesiones de cuadrado sumable. Al maximizar la correlación empírica $\rho(P_n[\langle u, \mathbf{X} \rangle_{p_n}, \langle v, \mathbf{Y} \rangle_q])$, estamos calculando el estimador de Sieves de (x, \mathbf{Y}) donde $\mathcal{H} = \ell^2$ y $\mathcal{H}_{p_n} = \langle e_1, e_2, \dots, e_{p_n} \rangle$ donde e_i es el i -ésimo vector canónico de ℓ^2 . Si x satisface **C2** entonces por el Corolario 3.3.1, $\widehat{\rho}_{1,SI} = 1$. Este es otro enfoque para justificar la necesidad de suavizar en este tipo de problemas. Un planteo de este tipo, en donde problemas de datos de alta dimensión son tratados como datos funcionales puede verse en Chen *et al.* (2011).

4. CONSISTENCIA DE LOS ESTIMADORES

Este capítulo contiene los principales resultados de esta tesis. En la Sección 4.2 presentamos las hipótesis necesarias para obtener los resultados de consistencia que describimos en la Sección 4.3. Asimismo, se relacionan estos supuestos con las condiciones dadas por Jin y Cui (2010) y Alfons *et al.* (2016) para obtener el comportamiento asintótico y la función de influencia de los estimadores de projection–pursuit en el caso finito–dimensional, respectivamente. También, se presenta la analogía de algunas de las hipótesis pedidas con las impuestas en Leurgans *et al.* (1993) para probar la consistencia de los estimadores penalizados basados en la covarianza y varianza muestral. En particular, en la Sección 4.2.3 analizamos la consistencia Fisher de dichos estimadores. Finalmente, en los Apéndices se presentan algunos resultados auxiliares útiles para demostrar los principales resultados. Empezaremos en la Sección 4.1 simplificando la notación y recordando la definición de los estimadores de Sieves.

4.1. Introducción

Hemos introducido en el Capítulo 3 los estimadores basados en Sieves. Por otra parte, hemos probado que la estimación no es consistente cuando la dimensión de los subespacios de aproximación es mayor o igual que el tamaño de muestra. Por esta razón, en esta tesis los estimadores propuestos están basados en Sieves pero las dimensiones p_n y q_n de los subespacios aproximantes \mathcal{H}_{1,p_n} \mathcal{H}_{2,q_n} son menores que el tamaño de muestra n . Aunque el hecho que $p_n < n$ y $q_n < n$ no aparece en forma explícita en el enunciado del Teorema 4.3.1,

este supuesto es necesario para garantizar la validez de la hipótesis **S1D** que es uno de los supuestos utilizados en dicho resultado.

A continuación recordaremos la notación que será útil en el resto del capítulo. Como en el Capítulo 3, \mathcal{H}_j , $j = 1, 2$ indicarán espacios de Hilbert separables con producto interno $\langle \cdot, \cdot \rangle_{\mathcal{H}_j}$ y norma $\| \cdot \|_{\mathcal{H}_j}$. Cuando el contexto sea claro omitiremos \mathcal{H}_j en el subíndice. El espacio producto se indicará por $\mathcal{H} = \mathcal{H}_1 \times \mathcal{H}_2$. Sean $\mathcal{S}_j = \{u \in \mathcal{H}_j : \|u\|_{\mathcal{H}_j} = 1\}$, $j = 1, 2$, la esfera unidad y $B_1 = \{\delta_1, \delta_2, \dots\}$ y $B_2 = \{\eta_1, \eta_2, \dots\}$ bases ortonormales de \mathcal{H}_1 y \mathcal{H}_2 , respectivamente. Indicaremos por \mathcal{H}_{1,p_n} y \mathcal{H}_{2,q_n} a los subespacios de dimensiones p_n y q_n de \mathcal{H}_1 y \mathcal{H}_2 generados por $\{\delta_1, \dots, \delta_{p_n}\}$ y $\{\eta_1, \dots, \eta_{q_n}\}$, respectivamente y por $\mathcal{S}_{1,p_n} = \mathcal{S}_1 \cap \mathcal{H}_{1,p_n}$ y $\mathcal{S}_{2,q_n} = \mathcal{S}_2 \cap \mathcal{H}_{2,q_n}$ a las esferas unitarias de \mathcal{H}_{1,p_n} y \mathcal{H}_{2,q_n} , respectivamente.

Dado un funcional de asociación ρ definido sobre el espacio de medidas de \mathbb{R}^2 y un elemento aleatorio $(X, Y)^T : \Omega \rightarrow \mathcal{H}_1 \times \mathcal{H}_2$, por simplicidad de notación indicaremos por

$$\rho_{XY}(u, v) = \rho(\langle u, X \rangle_{\mathcal{H}_1}, \langle v, Y \rangle_{\mathcal{H}_2}) = \rho(P[\langle u, X \rangle_{\mathcal{H}_1}, \langle v, Y \rangle_{\mathcal{H}_2}]) .$$

De modo análogo, definimos

$$\begin{aligned} \rho_{XX}(u_1, u_2) &= \rho(P[\langle u_1, X \rangle_{\mathcal{H}_1}, \langle u_2, X \rangle_{\mathcal{H}_1}]) \\ \rho_{YY}(v_1, v_2) &= \rho(P[\langle v_1, Y \rangle_{\mathcal{H}_2}, \langle v_2, Y \rangle_{\mathcal{H}_2}]) . \end{aligned}$$

De esta forma, si $(X, Y)^T$ es un elemento aleatorio tal que $(X, Y)^T \sim P$, los funcionales asociados a la primera correlación canónica $\rho_1 = \rho_1(P)$ y a las primeras direcciones canónicas $(\Phi_1, \Psi_1) = (\Phi_1(P), \Psi_1(P))$ se definen como

$$\begin{aligned} \rho_1^2 &= \sup_{u \in \mathcal{S}_1, v \in \mathcal{S}_2} \rho_{XY}^2(u, v) \\ (\Phi_1, \Psi_1) &= \operatorname{argmax}_{u \in \mathcal{S}_1, v \in \mathcal{S}_2} \rho_{XY}^2(u, v). \end{aligned} \tag{4.1}$$

Las siguientes direcciones y correlaciones están definidas como en (3.5) y (3.6).

Sea $(X_1, Y_1)^T, \dots, (X_n, Y_n)^T$ una muestra aleatoria de $(X, Y)^T$. Dadas bases ortonormales B_1 y B_2 de \mathcal{H}_1 y \mathcal{H}_2 , respectivamente y sucesiones $(p_n)_{n \in \mathbb{N}}$ y $(q_n)_{n \in \mathbb{N}}$, definimos los estimadores basados en Sieves para las direcciones y las correlaciones canónicas como en (3.7) y (3.8), es decir, maximizando las correlaciones empíricas en esferas de dimensión finita que van creciendo aproximando a los de la definición (4.1). Más precisamente, si ρ_n está dado como en (3.9), es decir,

$$\rho_n(u, v) = \rho(P_n[\langle u, X \rangle_{\mathcal{H}_1}, \langle v, Y \rangle_{\mathcal{H}_2}]),$$

los estimadores que consideraremos en este capítulo, están definidos por

$$\begin{cases} \widehat{\rho}_1^2 = \widehat{\rho}_{n,1}^2 = \sup_{u \in \mathcal{S}_{1,p_n}, v \in \mathcal{S}_{2,q_n}} \rho_n^2(u, v) \\ (\widehat{\Phi}_1, \widehat{\Psi}_1) = (\widehat{\Phi}_{n,1}, \widehat{\Psi}_{n,1}) = \operatorname{argmax}_{u \in \mathcal{S}_{1,p_n}, v \in \mathcal{S}_{2,q_n}} \rho_n^2(u, v). \end{cases} \quad (4.2)$$

Vale la pena observar que los estimadores definidos en (4.2) dependen de las bases de \mathcal{H}_1 y \mathcal{H}_2 elegidas, de las sucesiones $(p_n)_{n \in \mathbb{N}}$ y $(q_n)_{n \in \mathbb{N}}$ y de la medida de asociación ρ que se utilice. En el Capítulo 5, al considerar estimadores basados en distintas medidas de asociación, se calibrarán dichas medidas para que sean comparables, es decir, se las transformará, tal como se explicó en el Capítulo 2, de manera que estimen el mismo valor para procesos Gaussianos.

Observemos además que cuando la medida de asociación es la correlación de Pearson clásicamente utilizada (ρ_{CL}) y el elemento aleatorio $(X, Y)^T$ tiene segundo momento finito, las direcciones y correlaciones canónicas definidas en (4.1) pueden definirse a partir del operador de covarianza definido en (2.3) como

$$\begin{aligned} \rho_{\text{CL},1}^2 &= \sup_{u \in \mathcal{S}_1, v \in \mathcal{S}_2} \frac{\langle u, \Gamma_{XY} v \rangle^2}{\langle u, \Gamma_{XX} u \rangle \langle v, \Gamma_{YY} v \rangle} \\ (\Phi_{\text{CL},1}, \Psi_{\text{CL},1}) &= \operatorname{argmax}_{u \in \mathcal{S}_1, v \in \mathcal{S}_2} \frac{\langle u, \Gamma_{XY} v \rangle^2}{\langle u, \Gamma_{XX} u \rangle \langle v, \Gamma_{YY} v \rangle}, \end{aligned}$$

donde entenderemos que el cociente $\langle u, \Gamma_{XY} v \rangle^2 / \{\langle u, \Gamma_{XX} u \rangle \langle v, \Gamma_{YY} v \rangle\}$ es igual a 0 cuando $\langle u, \Gamma_{XX} u \rangle = 0$ o $\langle v, \Gamma_{YY} v \rangle = 0$.

Este hecho motiva la siguiente definición que servirá en lo que sigue para dar condiciones que garanticen la consistencia Fisher en familias de distribuciones elípticas y para probar bajo que condiciones se cumplen los supuestos que utilizaremos en la Sección 4.3.

Definición 4.1.1. Dado un operador compacto, semi definido positivo y autoadjunto $\Gamma = \mathcal{H}_1 \times \mathcal{H}_2 \rightarrow \mathcal{H}_1 \times \mathcal{H}_2$ con $\Gamma = \begin{pmatrix} \Gamma_{11} & \Gamma_{12} \\ \Gamma_{21} & \Gamma_{22} \end{pmatrix}$, éste induce un funcional (de asociación) $\rho_\Gamma : \mathcal{H}_1 \times \mathcal{H}_2 \rightarrow \mathbb{R}$ definido como

$$\rho_\Gamma(u, v) = \frac{\langle u, \Gamma_{12}v \rangle}{\sqrt{\langle u, \Gamma_{11}u \rangle \langle v, \Gamma_{22}v \rangle}}. \quad (4.3)$$

El funcional ρ_Γ permite definir la primera correlación canónica, $\rho_{\Gamma,1}$, y las primeras direcciones canónicas asociadas a Γ , $(\Phi_{\Gamma,1}, \Psi_{\Gamma,1})$, como

$$\begin{aligned} \rho_{\Gamma,1}^2 &= \sup_{u \in \mathcal{S}_1, v \in \mathcal{S}_2} \rho_\Gamma^2(u, v) \\ (\Phi_{\Gamma,1}, \Psi_{\Gamma,1}) &= \operatorname{argmax}_{u \in \mathcal{S}_1, v \in \mathcal{S}_2} \rho_\Gamma^2(u, v), \end{aligned} \quad (4.4)$$

donde $\rho_\Gamma^2(u, v) = 0$ cuando $\langle u, \Gamma_{11}u \rangle = 0$ o $\langle v, \Gamma_{22}v \rangle = 0$.

Las siguientes correlaciones y direcciones canónicas se definen como

$$\begin{aligned} \rho_{\Gamma,k}^2 &= \sup_{(u,v) \in \mathcal{B}_{\Gamma,k}} \rho_\Gamma^2(u, v) \\ (\Phi_{\Gamma,k}, \Psi_{\Gamma,k}) &= \operatorname{argmax}_{(u,v) \in \mathcal{B}_{\Gamma,k}} \rho_\Gamma^2(u, v), \end{aligned} \quad (4.5)$$

donde $\mathcal{B}_{\Gamma,k} = \{(u, v) \in \mathcal{S}_1 \times \mathcal{S}_2 \text{ y } \langle u, \Gamma_{11}\Phi_{\Gamma,j} \rangle = \langle v, \Gamma_{22}\Psi_{\Gamma,j} \rangle = 0 \text{ si } j \leq k - 1\}$.

Asimismo Γ induce los funcionales de correlación $\rho_{\Gamma_{11}} : \mathcal{H}_1 \times \mathcal{H}_1 \rightarrow \mathbb{R}$ y $\rho_{\Gamma_{22}} : \mathcal{H}_2 \times \mathcal{H}_2 \rightarrow \mathbb{R}$ definidos de la siguiente manera:

$$\begin{cases} \rho_{\Gamma_{11}}(u_1, u_2) = \frac{\langle u_1, \Gamma_{11}u_2 \rangle}{\sqrt{\langle u_1, \Gamma_{11}u_1 \rangle \langle u_2, \Gamma_{11}u_2 \rangle}} \\ \rho_{\Gamma_{22}}(v_1, v_2) = \frac{\langle v_1, \Gamma_{22}v_2 \rangle}{\sqrt{\langle v_1, \Gamma_{22}v_1 \rangle \langle v_2, \Gamma_{22}v_2 \rangle}}, \end{cases} \quad (4.6)$$

que coinciden con ρ_{XX} y ρ_{YY} , respectivamente cuando $(X, Y)^T$ tiene segundo momento finito y $\Gamma = \Gamma_{ZZ}$ es el operador de covarianza definido en (2.3).

4.2. Supuestos

En esta sección presentamos los supuestos necesarios para probar la consistencia de nuestros estimadores. Daremos asimismo algunos comentarios sobre dichas condiciones.

Sea ρ una medida de asociación definida en las medidas bivariadas. Como venimos haciendo, indistintamente aplicaremos ρ a medidas bivariadas o a vectores aleatorios bivariados.

Consideraremos los siguientes supuestos que serán útiles para probar la consistencia de nuestros estimadores.

$$\mathbf{S1F:} \sup_{u \in \mathcal{S}_{1,p_n}, v \in \mathcal{S}_{2,q_n}} |\rho_{XY}^2(u, v) - \rho_n^2(u, v)| \xrightarrow{a.s.} 0.$$

$$\mathbf{S1D:} \sup_{u \in \mathcal{S}_{1,p_n}, v \in \mathcal{S}_{2,q_n}} |\rho_{XY}^2(u, v) - \rho_n^2(u, v)| \xrightarrow{p} 0.$$

$$\mathbf{S2:} \rho_{XY}^2 : \mathcal{H}_1 \times \mathcal{H}_2 \rightarrow \mathbb{R} \text{ es continua en } (\Phi_1, \Psi_1).$$

$\mathbf{S3:}$ Existe un operador $\Gamma : \mathcal{H}_1 \times \mathcal{H}_2 \rightarrow \mathcal{H}_1 \times \mathcal{H}_2$ compacto, definido positivo y auto-adjunto tal que si ρ_Γ , $\rho_{\Gamma_{11}}$ y $\rho_{\Gamma_{22}}$ están definidos por (4.3) y (4.6), entonces

$$\rho_{XY}^2(u, v) = h(\rho_\Gamma^2(u, v)) \tag{4.7}$$

$$\rho_{XX}^2(u_1, u_2) = h_X(\rho_{\Gamma_{11}}^2(u_1, u_2)) \tag{4.8}$$

$$\rho_{YY}^2(v_1, v_2) = h_Y(\rho_{\Gamma_{22}}^2(v_1, v_2)), \tag{4.9}$$

donde si \tilde{h} es cualquiera entre h , h_X o h_Y , $\tilde{h} : [0, 1] \rightarrow [0, 1]$ es estrictamente creciente con $\tilde{h}(0) = 0$, y $\lim_{x \rightarrow 1^-} \tilde{h}(x) = 1$.

$\mathbf{S4:}$ Salvo cambio de signo, tanto en Φ_1 como en Ψ_1 , existe un único $(\Phi_1, \Psi_1) \in \mathcal{H}$ tal que $\|\Phi_1\|_{\mathcal{H}_1} = \|\Psi_1\|_{\mathcal{H}_2} = 1$ y $\rho_1 = \rho_{XY}(\Phi_1, \Psi_1) \geq \rho_{XY}(u, v)$ para todo $(u, v) \in \mathcal{H}_1 \times \mathcal{H}_2$. Además, existe ρ_2 con $0 \leq \rho_2 < \rho_1$ tal que si $\rho_{XX}(u, \Phi_1) = \rho_{YY}(v, \Psi_1) = 0$ entonces $\rho_{XY}^2(u, v) \leq \rho_2^2$.

En las secciones que siguen se comentan las hipótesis **S1F** a **S4** y se dan condiciones bajo las cuales son válidas.

4.2.1. Comentarios sobre **S1F** y **S1D**

La primera observación que debemos hacer es que la hipótesis **S1F** involucra un supremo sobre esferas en los espacios de dimensión finita \mathcal{H}_{1,p_n} y \mathcal{H}_{2,q_n} . Hemos reducido la dimensión del espacio original ya que $\sup_{u \in \mathcal{S}_1, v \in \mathcal{S}_2} |\rho_{XY}^2(u, v) - \rho_n^2(u, v)|$ no converge en casi todo punto a 0 si $\rho_1(P) < 1$. Efectivamente, supongamos que $\sup_{u \in \mathcal{S}_1, v \in \mathcal{S}_2} |\rho_{XY}^2(u, v) - \rho_n^2(u, v)| \xrightarrow{a.s.} 0$. En la Sección 3.2, hemos visto que con probabilidad 1 existían $u_n \in \mathcal{S}_1$, $v_n \in \mathcal{S}_2$ tales que $\rho_n(u_n, v_n) = 1$. Por lo tanto, con probabilidad 1, existen $u_n \in \mathcal{S}_1$, $v_n \in \mathcal{S}_2$ tales que $\rho(u_n, v_n) \rightarrow 1$, lo que contradice el hecho que $\rho_1(P) < 1$.

El siguiente resultado da condiciones suficientes que garantizan el cumplimiento de **S1D**, su demostración se da en el Apéndice 4.4.

Proposición 4.2.1. Sean \mathcal{H}_j , $j = 1, 2$, espacios de Hilbert separables y $Z_i = (X_i, Y_i)^T \sim P$, $i \in \mathbb{N}$, elementos aleatorios independientes sobre $\mathcal{H} = \mathcal{H}_1 \times \mathcal{H}_2$. Dados $u \in \mathcal{S}_1$, $v \in \mathcal{S}_2$, indiquemos por $\mathbf{w} = (u, v) \in \mathcal{S}$ con $\mathcal{S} = \mathcal{S}_1 \times \mathcal{S}_2$. Supongamos que

A1. Existe $\alpha : \mathcal{H} \times \mathcal{S} \rightarrow \mathbb{R}$ tal que

a) $\mathbb{E} \alpha(Z, \mathbf{w}) = 0$, $\forall \mathbf{w} \in \mathcal{S}$, y $\mathbb{E} (\sup_{\mathbf{w} \in \mathcal{S}} |\alpha(Z, \mathbf{w})|) < \infty$,

b) $\alpha(z, \mathbf{w})$ es continua en ambas variables. Más aún, α es equicontinua en \mathbf{w} sobre compactos, es decir, dado $\mathcal{K} \subset \mathcal{H}$ compacto y $\epsilon > 0$, existe $\delta > 0$ tal que si $\mathbf{w}_1, \mathbf{w}_2 \in \mathcal{S}$, $\|\mathbf{w}_1 - \mathbf{w}_2\| < \delta$ entonces $|\alpha(z, \mathbf{w}_1) - \alpha(z, \mathbf{w}_2)| < \epsilon$, para todo $z \in \mathcal{K}$.

c) $\sup_{\mathbf{w} \in \mathcal{S}_n} |\rho_n^2(\mathbf{w}) - \rho_{XY}^2(\mathbf{w}) - 1/n \sum_{j=1}^n \alpha(Z_j, \mathbf{w})| \xrightarrow{p} 0$, donde $\mathcal{S}_n = \mathcal{S}_{1,p_n} \times \mathcal{S}_{2,q_n}$.

A2. $(p_n + q_n)/n \rightarrow 0$,

entonces vale **S1D**, es decir, $\sup_{u \in \mathcal{S}_{1,p_n}, v \in \mathcal{S}_{2,q_n}} |\rho_n^2(u, v) - \rho_{XY}^2(u, v)| \xrightarrow{p} 0$.

Observemos que la hipótesis **A1** se relaciona con el supuesto **S5** de Jin y Cui (2010) quienes piden una condición con tasa, más fuerte, para obtener la distribución asintótica de los estimadores de projection–pursuit en el caso multivariado.

En la Proposición 4.2.2 probaremos que existen sucesiones $(p_n)_{n \in \mathbb{N}}$ y $(q_n)_{n \in \mathbb{N}}$ con $p_n \rightarrow \infty$ y $q_n \rightarrow \infty$, $p_n < n$, $q_n < n$ que verifican **S1D**, es decir, tales que $\sup_{u \in \mathcal{S}_{1,p_n}, v \in \mathcal{S}_{2,q_n}} |\rho(u, v) - \rho_n(u, v)| \xrightarrow{p} 0$. La demostración será una demostración de existencia de dichas sucesiones $(p_n)_{n \in \mathbb{N}}$ y $(q_n)_{n \in \mathbb{N}}$, pero no dará una caracterización de su orden de convergencia como en la hipótesis **A2** de la Proposición 4.2.1.

Como en la Sección 3.1 para cada p y q fijos llamaremos $\mathcal{S}_{1,p} = \mathcal{S}_1 \cap \mathcal{H}_{1,p}$, $\mathcal{S}_{2,p} = \mathcal{S}_2 \cap \mathcal{H}_{2,q}$.

Proposición 4.2.2. Sean $Z = (X, Y)^T \in \mathcal{H}$ un elemento aleatorio y una medida de asociación ρ que cumple **A3D** dada por

A3D. Para cualquier p y q en \mathbb{N} fijos, $\sup_{u \in \mathcal{S}_{1,p}, v \in \mathcal{S}_{2,q}} |\rho_{XY}(u, v) - \rho_n(u, v)| \xrightarrow{p} 0$.

Entonces, existen sucesiones $(p_n)_{n \in \mathbb{N}}$ y $(q_n)_{n \in \mathbb{N}}$ con $p_n \rightarrow \infty$ y $q_n \rightarrow \infty$, $p_n < n$, $q_n < n$, que verifican **S1D**.

La demostración de la Proposición 4.2.2 así como condiciones bajo las cuales se cumple **A3D** se dan en el Apéndice 4.4. Por otra parte, en el Apéndice 4.5 se presentan condiciones para garantizar que **S1F** se cumple.

4.2.2. Comentarios sobre **S2**

Supongamos que la medida de asociación cumple (4.7) para una función h estrictamente creciente y continua. La compacidad de Γ implica su acotación, por lo tanto, las funciones $g_1(u) = \langle u, \Gamma_{11}u \rangle_{\mathcal{H}_1}$, $g_2(v) = \langle v, \Gamma_{22}v \rangle_{\mathcal{H}_2}$ y $g_{12} = \langle u, \Gamma_{12}v \rangle_{\mathcal{H}_1}$ resultan ser funciones continuas.

Luego, como h es estrictamente creciente, continua y $h(0) = 0$, $\rho_1 > 0$ implica que $g_1(\Phi_1) \neq 0$ y $g_2(\Psi_1) \neq 0$, de donde deducimos que ρ_{XY}^2 es continua en (Φ_1, Ψ_1) si $\rho_1 > 0$. Observemos que si el núcleo de Γ_{jj} se reduce al $\{0\}$ para $j = 1, 2$, entonces claramente $g_1(\Phi_1) \neq 0$ y $g_2(\Psi_1) \neq 0$. En particular, si $h(t) = t$, el supuesto **S3** implica **S2** si $\rho_1 > 0$. Por otro lado, por el Lema 4.4.2 que se enuncia más adelante, se obtiene también la validez de **S2** si ρ es débilmente continua.

4.2.3. Comentarios sobre **S3** y la consistencia Fisher

El Teorema 4.3.1 establece la consistencia de los estimadores definidos en (4.2) a las correlaciones y direcciones canónicas definidas en (4.1). Ambas definiciones (4.1) y (4.2) dependen del funcional de asociación ρ que se elija en cada caso. Por esta razón, es importante conocer qué representan la cantidad $\rho_1 = \rho_1(P)$ y la direcciones $(\Phi_1, \Psi_1) = (\Phi_1(P), \Psi_1(P))$, al menos en algunas situaciones particulares.

La primera observación es que si la medida de asociación cumple (4.7), entonces $\rho_1(P) = h(\rho_{\Gamma,1})$, $\Phi_1(P) = \Phi_{\Gamma,1}$ y $\Psi_1(P) = \Psi_{\Gamma,1}$, es decir, las direcciones estimadas definidas en (4.2) son consistentes a las primeras direcciones canónicas asociadas a Γ que son las cantidades naturales de interés en este caso. Por otra parte, si $h(t) = t$, $\hat{\rho}_1$ provee un estimador consistente de la primera correlación canónica asociada a Γ y tenemos que los funcionales definidos en (4.1) son Fisher-consistentes. Vale la pena mencionar que la condición (4.7) del supuesto **S3** es análoga a la hipótesis (iv) de Alfons *et al.* (2016).

Por otra parte, si el elemento aleatorio $Z = (X, Y)^T$ tiene segundo momento finito y tomamos como medida de asociación el coeficiente de correlación de Pearson, o sea, si $\rho = \rho_{CL}$, obtenemos las correlaciones y direcciones canónicas definidas por Leurgans *et al.* (1993) y estudiadas por He *et al.* (2003) quienes dan condiciones suficientes para su existencia. Observemos que en este caso, $\rho_{XY} = \rho_{\Gamma}$ con Γ el operador de covarianza de Z . De ahora

en más al referirnos a estos funcionales los denotaremos por $\rho_{\text{CL},1}$, $\Phi_{\text{CL},1}$ y $\Psi_{\text{CL},1}$ y a los estimadores asociados por $\widehat{\rho}_{\text{CL},1}$, $\widehat{\Phi}_{\text{CL},1}$ y $\widehat{\Psi}_{\text{CL},1}$, respectivamente. Vale la pena mencionar que estos funcionales están bien definidos solamente cuando tanto X como Y tienen segundo momento finito. En particular, están bien definidos para familias Gaussianas o para familias elípticas con segundo momento finito.

Una clase de distribuciones que generaliza las Gaussianas son las elípticas. A continuación, mostraremos que para familias elípticas, los funcionales $\rho_1(P)$, $\Phi_1(P)$ y $\Psi_1(P)$ tienen una interpretación simple para algunos funcionales de asociación descritos en la Sección 2.3. Sea $Z = (X, Y)^T \sim \mathcal{E}(\mu, \Gamma, g)$ un elemento aleatorio en \mathcal{H} donde $\mu = (\mu_1, \mu_2)$, $\mu_j \in \mathcal{H}_j$ y

$$\Gamma = \begin{pmatrix} \Gamma_{11} & \Gamma_{12} \\ \Gamma_{21} & \Gamma_{22} \end{pmatrix}.$$

Supondremos sin pérdida de generalidad que si existe segundo momento finito entonces el operador de covarianza de Z es igual a Γ , lo que ocurre, por ejemplo, en el caso Gaussiano.

Dados $u \in \mathcal{H}_1$ y $v \in \mathcal{H}_2$ definamos $A_{uv} : \mathcal{H} \rightarrow \mathbb{R}^2$ como $A_{uv}(x, y) = (\langle u, x \rangle_{\mathcal{H}_1}, \langle v, y \rangle_{\mathcal{H}_2})^T$. Como $Z \sim \mathcal{E}(\mu, \Gamma, g)$, el vector $\mathbf{Z}_{u,v} = A_{uv}(X, Y)$ tiene distribución F_{uv} elíptica, es decir, $\mathbf{Z}_{uv} \sim \mathcal{E}_2(\boldsymbol{\mu}^{uv}, \boldsymbol{\Sigma}^{uv}, g)$, donde $\boldsymbol{\mu}^{uv} = A_{uv}\mu$ y

$$\boldsymbol{\Sigma}^{uv} = A_{uv}\Gamma A_{uv}^* = \begin{pmatrix} \langle u, \Gamma_{11}u \rangle_{\mathcal{H}_1} & \langle u, \Gamma_{12}v \rangle_{\mathcal{H}_1} \\ \langle v, \Gamma_{21}u \rangle_{\mathcal{H}_2} & \langle v, \Gamma_{22}v \rangle_{\mathcal{H}_2} \end{pmatrix}.$$

Sea ρ una medida de asociación Fisher-consistente para vectores bivariados con distribución elíptica, como por ejemplo, el coeficiente $\rho_{\mathbf{V}}$ definido en (2.17) a partir de un funcional de dispersión afín equivariante o los coeficientes ρ_{GK} y ρ_{OGK} definidos en (2.14) y (2.18), respectivamente, si el vector aleatorio elíptico tiene además segundo momento finito. Como ρ es Fisher-consistente, se cumple que

$$\rho_{XY}^2(u, v) = \rho^2(F_{uv}) = \frac{\langle u, \Gamma_{12}v \rangle_{\mathcal{H}_1}^2}{\langle u, \Gamma_{11}u \rangle_{\mathcal{H}_1} \langle v, \Gamma_{22}v \rangle_{\mathcal{H}_2}} = \rho_{\Gamma}^2(u, v). \quad (4.10)$$

Por lo tanto, ρ cumple (4.7) con $h(t) = t$ y los funcionales definidos en (4.1) son la primera correlación y las primeras direcciones canónicas asociadas a Γ . Más aún, argumentos análogos permiten mostrar que $\rho_{XX}^2(u_1, u_2) = \rho_{\Gamma_{11}}(u_1, u_2)$ y $\rho_{XX}^2(u_1, u_2) = \rho_{\Gamma_{11}}(u_1, u_2)$, es decir, medidas de asociación Fisher-consistentes para familias elípticas bivariadas cumplen **S3** y dan origen a estimadores de las cantidades de interés.

Por otra parte, si consideramos familias Gaussianas y medidas de asociación Fisher-consistentes para vectores bivariados normales, tales como

- la distancia de correlación transformada ρ_{DIST} dada en (2.7),
- la correlación mediana transformada ρ_{CM}^* dada en (2.10),
- las correlaciones de Spearman o de Kendall transformadas, ρ_{SP} y ρ_{K} dadas en (2.11)

por (4.10), obtenemos que se cumple **S3** de donde se deduce la Fisher-consistencia para los funcionales definidos en (4.1).

Por lo tanto, en el caso de procesos Gaussianos, los funcionales $\rho_1(P)$, $\Phi_1(P)$ y $\Psi_1(P)$ son siempre los mismos eligiendo cualquiera de las medidas de asociación ρ_{GK} , ρ_{OGK} , ρ_{V} , ρ_{DIST} , ρ_{CM}^* , ρ_{SP} y ρ_{K} y coinciden con los obtenidos con ρ_{CL} .

4.2.4. Comentarios sobre **S4**

La hipótesis **S4** es una generalización de la condición 3 de Leurgans *et al.* (1993). Como en la Sección 4.2.3, sea $Z = (X, Y)^T \sim \mathcal{E}(\mu, \Gamma, g)$ un elemento aleatorio elíptico Z en \mathcal{H} , con segundo momento finito donde la media $\mu = (\mu_1, \mu_2)$, $\mu_j \in \mathcal{H}_j$ y el operador de dispersión Γ se escribe como $\Gamma = \begin{pmatrix} \Gamma_{11} & \Gamma_{12} \\ \Gamma_{21} & \Gamma_{22} \end{pmatrix}$. Sin pérdida de generalidad, supongamos que Γ es el operador de covarianza de Z y definamos el operador $R = \Gamma_{11}^{-1/2} \Gamma_{12} \Gamma_{22}^{-1/2}$, donde $\Gamma_{11}^{-1/2}$ y $\Gamma_{22}^{-1/2}$ son las inversas generalizadas de las raíces de los operadores de covarianza de X e Y ,

respectivamente. Para la construcción y propiedades de los operadores $\Gamma_{jj}^{-1/2}$, ver por ejemplo, Conway (1985). He *et al.* (2003) consideran la situación $\mathcal{H}_j = L^2(0, 1)$ y la descomposición de Karhunen–Loève de los procesos X y Y , es decir,

$$X = \mu_1 + \sum_{i \geq 1} \xi_{i,1} \nu_{i,1} \quad Y = \mu_2 + \sum_{i \geq 1} \xi_{i,2} \nu_{i,2}$$

donde, para $j = 1, 2$, $\lambda_{i,j}$ y $\nu_{i,j} \in \mathcal{H}_j$ son los autovalores y autofunciones del operador Γ_{jj} y la sucesión de variables aleatorias $\xi_{i,j}$ es tal que $\mathbb{E}\xi_{i,j} = 0$, $\mathbb{E}\xi_{i,j}\xi_{\ell,j} = 0$ si $i \neq \ell$ y $\mathbb{E}\xi_{i,j}^2 = \lambda_{i,j}$ con $\lambda_{1,j} \geq \lambda_{2,j} \geq \dots$ y $\sum_{i \geq 1} \lambda_{i,j}^2 < \infty$. He *et al.* (2000) muestran que si se cumple

$$\sum_{i \geq 1} \frac{\mathbb{E}^2(\xi_{i,1} \xi_{i,2})}{\lambda_{i,1}^2 \lambda_{i,2}} < \infty \quad \text{o} \quad \sum_{i \geq 1} \frac{\mathbb{E}^2(\xi_{i,1} \xi_{i,2})}{\lambda_{i,1} \lambda_{i,2}^2} < \infty, \quad (4.11)$$

el problema de correlación clásico obtenido utilizando como medida de asociación $\rho = \rho_{\text{CL}}$ está bien definido. Más aún, sean ζ_k y v_k , con $\zeta_1 \geq \zeta_2 \geq \dots$, los autovalores no nulos y las autofunciones asociadas de R^*R . He *et al.* (2003) prueban que si el primer autovalor ζ_1 de R^*R tiene multiplicidad 1, entonces **S4** se cumple cuando elegimos el coeficiente de correlación de Pearson ρ_{CL} . Recordemos que si $\rho = \rho_{\text{CL}}$, $\rho_{XY}^2(u, v) = \rho_{\Gamma}^2(u, v)$ y sea $(\Phi_{\Gamma,1}, \Psi_{\Gamma,1})$ el maximizador salvo signo de norma 1 de $\rho_{\Gamma}^2(u, v)$ definido en (4.3). Por lo tanto, si denotamos por v_1 a la autofunción asociada a ζ_1 y por $u_1 = Rv_1/\sqrt{\zeta_1}$, del Teorema 4.8 de He *et al.* (2003) se deduce que $(\Phi_{\Gamma,1}, \Psi_{\Gamma,1}) = (\Gamma_{11}^{-1/2}u_1/\|\Gamma_{11}^{-1/2}u_1\|_{\mathcal{H}_1}, \Gamma_{22}^{-1/2}v_1/\|\Gamma_{22}^{-1/2}v_1\|_{\mathcal{H}_2})$ y es único si ζ_1 tiene multiplicidad 1. Además, He *et al.* (2003) muestran que si $\zeta_2 < \zeta_1$ es el segundo autovalor de R^*R , entonces si $u \in \mathcal{H}_1$ y $v \in \mathcal{H}_2$ son tales que $\rho_{XX}(u, \Phi_1) = \rho_{YY}(v, \Psi_1) = 0$ se tiene que $\rho_{XY}(u, v) \leq \rho_2 = \sqrt{\zeta_2}$.

Como mencionamos en la Sección 4.2.3 si la medida de asociación cumple (4.7), entonces $\rho_1(P) = h(\rho_{\Gamma,1})$, $\Phi_1(P) = \Phi_{\Gamma,1}$ y $\Psi_1(P) = \Psi_{\Gamma,1}$. Recordemos que las medidas de asociación ρ_{GK} y ρ_{OGK} definidas en (2.14) y (2.18), respectivamente, y el coeficiente $\rho_{\mathbf{V}}$ definido en (2.17) a partir de un funcional de dispersión afín equivariante, son coeficientes Fisher–consistentes. Por lo tanto, usando (4.10) obtenemos que ρ_{XY} , ρ_{XX} y ρ_{YY} cumplen **S3** con las funciones

h , h_X y h_Y iguales a la identidad cuando el elemento aleatorio Z tiene segundo momento y es elíptico $Z = (X, Y)^T \sim \mathcal{E}(\mu, \Gamma, g)$ como en el párrafo anterior. De los resultados de He *et al.* (2003), se deduce entonces en forma inmediata que para estas medidas de asociación se cumple **S4**, si el proceso cumple además (4.11).

Por otra parte, si tomamos como medida de asociación, la distancia de correlación transformada ρ_{DIST} dada en (2.7), la correlación mediana transformada ρ_{CM}^* dada en (2.10) o las correlaciones de Spearman o de Kendall transformadas, ρ_{SP} y ρ_{K} dadas en (2.11), bajo (4.11), la hipótesis **S4** se cumple para procesos Gaussianos.

Observemos finalmente que, para las medidas de asociación mencionadas, si el proceso Z cumple (4.11), los funcionales $\rho_1(P) = h(\rho_{\Gamma,1})$, $\Phi_1(P) = \Phi_{\Gamma,1}$ y $\Psi_1(P) = \Psi_{\Gamma,1}$ tienen una interpretación análoga a la del caso finito-dimensional.

4.3. Consistencia

El objetivo de esta sección es probar la consistencia del estimador de la primera correlación canónica y de las primeras direcciones canónicas. La noción de convergencia del estimador de las primeras direcciones canónicas a las primeras direcciones canónicas reales será la de convergencia respecto de la medida de asociación, o más brevemente, convergencia en correlación, análoga a la convergencia en norma Γ que se plantea en Leurgans *et al.* (1993).

Definición 4.3.1. Decimos que $u_n \xrightarrow{\rho} \Phi_1$ y que $v_n \xrightarrow{\rho} \Psi_1$ si $\rho_{XX}^2(u_n, \Phi_1) \rightarrow 1$ y $\rho_{YY}^2(v_n, \Psi_1) \rightarrow 1$, respectivamente.

El siguiente resultado muestra que si $\rho_{XY}^2(u_n, v_n) \rightarrow \rho_1^2$ entonces las direcciones (u_n, v_n) convergen en correlación y será utilizado para probar la consistencia de nuestros estimadores.

Proposición 4.3.1. Sean $(X, Y)^T : \Omega \rightarrow \mathcal{H}_1 \times \mathcal{H}_2$ un elemento aleatorio y ρ una medida de asociación bivariada que cumple **S3** y **S4**. Si $\rho_{XY}^2(u_n, v_n) \rightarrow \rho_1^2$ entonces $\rho_{XX}^2(u_n, \Phi_1) \rightarrow 1$ y

$$\rho_{YY}^2(v_n, \Psi_1) \rightarrow 1.$$

Para probar la Proposición 4.3.1 necesitaremos el siguiente Lema que generaliza el Lema 4 de Leurgans *et al.* (1993) reemplazando la convergencia en norma respecto del operador de covarianza por una convergencia respecto de un operador Γ que puede corresponder al operador de dispersión en familias elípticas sin momentos. Su demostración es análoga a la del Lema 4 de Leurgans *et al.* (1993) y se incluye por completitud.

Lema 4.3.1. *Sea $\Gamma : \mathcal{H}_1 \times \mathcal{H}_2 \rightarrow \mathcal{H}_1 \times \mathcal{H}_2$ un operador compacto, definido positivo y autoadjunto. Sean $\rho_{\Gamma,k}$ definidas como en (4.4) y (4.5) y supongamos que $0 \leq \rho_{\Gamma,2} < \rho_{\Gamma,1}$. Sean $(u_n, v_n) \in \mathcal{H}$, $n \in \mathbb{N}$, tales que $\rho_{\Gamma}^2(u_n, v_n) \rightarrow \rho_{\Gamma,1}^2$, entonces $\rho_{\Gamma_{11}}^2(u_n, \Phi_{\Gamma,1}) \rightarrow 1$ y $\rho_{\Gamma_{22}}^2(v_n, \Psi_{\Gamma,1}) \rightarrow 1$, donde $\Phi_{\Gamma,1}$ y $\Psi_{\Gamma,1}$ son las direcciones definidas en (4.4).*

Demostración. Sin pérdida de generalidad, multiplicando eventualmente $\Phi_{\Gamma,1}$, $\Psi_{\Gamma,1}$, u_n y v_n por constantes adecuadas, podemos suponer que $\langle \Phi_{\Gamma,1}, \Gamma_{11} \Phi_{\Gamma,1} \rangle_{\mathcal{H}_1} = \langle \Psi_{\Gamma,1}, \Gamma_{22} \Psi_{\Gamma,1} \rangle_{\mathcal{H}_2} = \langle u_n, \Gamma_{11} \Phi_{\Gamma,1} \rangle_{\mathcal{H}_1} = \langle v_n, \Gamma_{22} \Psi_{\Gamma,1} \rangle_{\mathcal{H}_2} = 1$.

Definamos $u_n^{(1)} = u_n - \Phi_{\Gamma,1}$ y $v_n^{(1)} = v_n - \Psi_{\Gamma,1}$, o sea, $u_n = \Phi_{\Gamma,1} + u_n^{(1)}$ y $v_n = \Psi_{\Gamma,1} + v_n^{(1)}$, lo que implica que $\langle u_n^{(1)}, \Gamma_{11} \Phi_{\Gamma,1} \rangle_{\mathcal{H}_1} = \langle v_n^{(1)}, \Gamma_{22} \Psi_{\Gamma,1} \rangle_{\mathcal{H}_2} = 0$

Usando que $(\Phi_{\Gamma,1}, \Psi_{\Gamma,1})$ es el máximo de $\rho_{\Gamma}^2(\cdot, \cdot)$ y definiendo $f_1(\lambda) = \rho_{\Gamma}^2(\Phi_{\Gamma,1}, \Psi_{\Gamma,1} + \lambda v_n^{(1)})$ deducimos que $f_1'(0) = 0$ de donde se desprende que $\langle \Phi_{\Gamma,1}, \Gamma_{12} v_n^{(1)} \rangle_{\mathcal{H}_1} = 0$. De manera análoga, obtenemos que $\langle u_n^{(1)}, \Gamma_{12} \Psi_{\Gamma,1} \rangle_{\mathcal{H}_1} = 0$.

Sean $\varepsilon_n^2 = \langle u_n^{(1)}, \Gamma_{11} u_n^{(1)} \rangle_{\mathcal{H}_1}$, $\delta_n^2 = \langle v_n^{(1)}, \Gamma_{22} v_n^{(1)} \rangle_{\mathcal{H}_2}$ y $\sigma = \rho_{\Gamma,2}/\rho_{\Gamma,1} < 1$, entonces vale que

$$\begin{aligned} \rho_{\Gamma,1}^{-2} \rho_{\Gamma}^2(u_n, v_n) &= \frac{\rho_{\Gamma,1}^{-2} \langle u_n, \Gamma_{12} v_n \rangle^2}{(1 + \varepsilon_n^2)(1 + \delta_n^2)} = \frac{\{1 + \rho_{\Gamma,1}^{-1} \varepsilon_n \delta_n \rho_{\Gamma}(u_n^{(1)}, v_n^{(1)})\}^2}{(1 + \varepsilon_n^2)(1 + \delta_n^2)} \\ &\leq \frac{(1 + \sigma \varepsilon_n \delta_n)^2}{(1 + \varepsilon_n^2)(1 + \delta_n^2)} \leq \frac{1 + \sigma(\varepsilon_n^2 + \delta_n^2) + \varepsilon_n^2 \delta_n^2}{1 + \varepsilon_n^2 + \delta_n^2 + \varepsilon_n^2 \delta_n^2} \leq 1. \end{aligned} \quad (4.12)$$

Por estar acotadas superiormente por 1 e inferiormente por $\rho_{\Gamma,1}^{-2} \rho_{\Gamma}^2(u_n, v_n)$ que tiende a 1,

todas las sucesiones de la cadena de desigualdades (4.12) convergen a 1. Sean

$$a_n = \frac{(1 + \sigma \varepsilon_n \delta_n)^2}{(1 + \varepsilon_n^2)(1 + \delta_n^2)} \quad \text{y} \quad b_n = \frac{1 + \sigma(\varepsilon_n^2 + \delta_n^2) + \varepsilon_n^2 \delta_n^2}{1 + \varepsilon_n^2 + \delta_n^2 + \varepsilon_n^2 \delta_n^2}.$$

Es fácil ver que las sucesiones $\{\varepsilon_n\}_{n \in \mathbb{N}}$ y $\{\delta_n\}_{n \in \mathbb{N}}$ son sucesiones acotadas pues $\lim_{n \rightarrow \infty} a_n = 1$. Por lo tanto, usando que $\lim_{n \rightarrow \infty} b_n = 1$ obtenemos que $\lim_{n \rightarrow \infty} \varepsilon_n = \lim_{n \rightarrow \infty} \delta_n = 0$, de donde $\rho_{\Gamma_{11}}^2(u_n, \Phi_{\Gamma,1}) = (1 + \varepsilon_n^2)^{-1} \rightarrow 1$ y $\rho_{\Gamma_{22}}^2(v_n, \Psi_{\Gamma,1}) = (1 + \delta_n^2)^{-1} \rightarrow 1$, lo que concluye la demostración. \square

Probaremos ahora la Proposición 4.3.1.

Demostración de la Proposición 4.3.1. Sean $\lambda_1^2 = h^{-1}(\rho_1^2)$ y $\lambda_2^2 = h^{-1}(\rho_2^2)$. Queremos ver que ρ_Γ cumple los supuestos del Lema 4.3.1.

En primer lugar observemos que **S3** implica que $\rho_1^2 = \rho_{XY}^2(\Phi_1, \Psi_1) = \max_{u,v} \rho_{XY}^2(u, v) = \max_{u,v} h(\rho_\Gamma^2(u, v))$, con lo cual, por ser h estrictamente creciente, $\lambda_1^2 = \rho_\Gamma^2(\Phi_1, \Psi_1) = \max_{u,v} \rho_\Gamma^2(u, v)$, es decir, $\lambda_1 = \rho_{\Gamma,1}$ y $(\Phi_1, \Psi_1) = (\Phi_{\Gamma,1}, \Psi_{\Gamma,1})$. Como $\rho_{XY}^2(u_n, v_n) = h(\rho_\Gamma^2(u_n, v_n)) \rightarrow \rho_1^2$ y h es creciente deducimos que $\rho_\Gamma^2(u_n, v_n) \rightarrow \rho_{\Gamma,1}^2$.

Sean u y v tales que $\langle u, \Gamma_{11}\Phi_1 \rangle = \langle v, \Gamma_{22}\Psi_1 \rangle = 0$. Como $h_X(0) = h_Y(0) = 0$, obtenemos que $\rho_{XX}(u, \Phi_1) = \rho_{YY}(v, \Psi_1) = 0$. Luego, por **S4**, $\rho_{XY}^2(u, v) = h(\rho_\Gamma^2(u, v)) \leq \rho_2^2 = h(\lambda_2^2)$, de donde, usando nuevamente que h es creciente se obtiene que $\lambda_2 = \rho_{\Gamma,2}$ y $\rho_\Gamma^2(u, v) \leq \lambda_2^2 < \lambda_1^2$. Por lo tanto, las sucesiones u_n y v_n junto con ρ_Γ cumplen las hipótesis del Lema 4.3.1, lo que implica que $\rho_{\Gamma_{11}}^2(u_n, \Phi_1) \rightarrow 1$ y $\rho_{\Gamma_{22}}^2(v_n, \Psi_1) \rightarrow 1$, pues $(\Phi_1, \Psi_1) = (\Phi_{\Gamma,1}, \Psi_{\Gamma,1})$. Finalmente, (4.8) y (4.9) nos permiten concluir que $\rho_{XX}^2(u_n, \Phi_1) \rightarrow 1$ y $\rho_{YY}^2(v_n, \Psi_1) \rightarrow 1$. \square

El siguiente Teorema prueba la consistencia débil de nuestros estimadores.

Teorema 4.3.1. (*Consistencia débil*) *Supongamos que se cumplen **S1D** y **S2**. Si $\lim_{n \rightarrow \infty} p_n = \infty$ y $\lim_{n \rightarrow \infty} q_n = \infty$, se tiene que*

$$i) \quad \tilde{\rho}_1^2 \xrightarrow{p} \rho_1^2$$

$$ii) \rho_{XY}^2(\widehat{\Phi}_1, \widehat{\Psi}_1) \xrightarrow{p} \rho_1^2$$

iii) Si además valen **S3** y **S4**, entonces $\rho_{XX}^2(\Phi_1, \widehat{\Phi}_1) \xrightarrow{p} 1$ y $\rho_{YY}^2(\Psi_1, \widehat{\Psi}_1) \xrightarrow{p} 1$.

Demostración. i) Sean $\tilde{\phi}_{1,p_n} = \Pi_{\mathcal{H}_{1,p_n}} \Phi_1 / \|\Pi_{\mathcal{H}_{1,p_n}} \Phi_1\|_{\mathcal{H}_1}$ y $\tilde{\psi}_{1,q_n} = \Pi_{\mathcal{H}_{2,q_n}} \Psi_1 / \|\Pi_{\mathcal{H}_{2,q_n}} \Psi_1\|_{\mathcal{H}_2}$ las proyecciones ortogonales normalizadas de las primeras direcciones canónicas poblacionales sobre \mathcal{H}_{1,p_n} y \mathcal{H}_{2,q_n} , respectivamente. Como $\tilde{\phi}_{1,p_n} \in \mathcal{S}_{1,p_n}$ y $\tilde{\psi}_{1,q_n} \in \mathcal{S}_{2,q_n}$, tenemos que $\widehat{\rho}_1^2 = \rho_n^2(\widehat{\Phi}_1, \widehat{\Psi}_1) \geq \rho_n^2(\tilde{\phi}_{1,p_n}, \tilde{\psi}_{1,q_n})$.

Observemos que $\rho_n^2(\tilde{\phi}_{1,p_n}, \tilde{\psi}_{1,q_n}) = \rho_{XY}^2(\tilde{\phi}_{1,p_n}, \tilde{\psi}_{1,q_n}) + a_n$ y $\rho_{XY}^2(\tilde{\phi}_{1,p_n}, \tilde{\psi}_{1,q_n}) = \rho_{XY}^2(\Phi_1, \Psi_1) + b_n$, donde

$$\begin{aligned} a_n &= \rho_n^2(\tilde{\phi}_{1,p_n}, \tilde{\psi}_{1,q_n}) - \rho_{XY}^2(\tilde{\phi}_{1,p_n}, \tilde{\psi}_{1,q_n}) \\ b_n &= \rho_{XY}^2(\tilde{\phi}_{1,p_n}, \tilde{\psi}_{1,q_n}) - \rho_{XY}^2(\Phi_1, \Psi_1). \end{aligned}$$

Por lo tanto,

$$\widehat{\rho}_1^2 \geq \rho_{XY}^2(\Phi_1, \Psi_1) + a_n + b_n. \quad (4.13)$$

Por otro lado,

$$\begin{aligned} \widehat{\rho}_1^2 &= \rho_n^2(\widehat{\Phi}_1, \widehat{\Psi}_1) = \rho_{XY}^2(\widehat{\Phi}_1, \widehat{\Psi}_1) + \left(\rho_n^2(\widehat{\Phi}_1, \widehat{\Psi}_1) - \rho_{XY}^2(\widehat{\Phi}_1, \widehat{\Psi}_1) \right) \\ &\leq \rho^2(\Phi_1, \Psi_1) + \left(\rho_n^2(\widehat{\Phi}_1, \widehat{\Psi}_1) - \rho_{XY}^2(\widehat{\Phi}_1, \widehat{\Psi}_1) \right) = \rho_{XY}^2(\Phi_1, \Psi_1) + c_n. \end{aligned} \quad (4.14)$$

Recordando que $\rho_1^2 = \rho_{XY}^2(\Phi_1, \Psi_1)$ y usando (4.13) y (4.14) deducimos que para probar i) basta probar que a_n , b_n y c_n convergen en probabilidad a 0. Por **S1D** y usando que $\widehat{\Phi}_1 \in \mathcal{S}_{1,p_n}$ y $\widehat{\Psi}_1 \in \mathcal{S}_{2,q_n}$ obtenemos

$$\begin{aligned} |a_n| &= |\rho_n^2(\tilde{\phi}_{1,p_n}, \tilde{\psi}_{1,p_n}) - \rho_{XY}^2(\tilde{\phi}_{1,p_n}, \tilde{\psi}_{1,p_n})| \leq \sup_{u \in \mathcal{S}_{1,p_n}, v \in \mathcal{S}_{2,q_n}} |\rho_{XY}^2(u, v) - \rho_n^2(u, v)| \xrightarrow{p} 0 \\ |c_n| &= |\rho_n^2(\widehat{\Phi}_1, \widehat{\Psi}_1) - \rho_{XY}^2(\widehat{\Phi}_1, \widehat{\Psi}_1)| \leq \sup_{u \in \mathcal{S}_{1,p_n}, v \in \mathcal{S}_{2,q_n}} |\rho_{XY}^2(u, v) - \rho_n^2(u, v)| \xrightarrow{p} 0. \end{aligned} \quad (4.15)$$

Por otra parte, como $p_n \rightarrow \infty$ y $q_n \rightarrow \infty$ se tiene que $\tilde{\phi}_{1,p_n} \rightarrow \Phi_1$ y $\tilde{\psi}_{1,p_n} \rightarrow \Psi_1$. Por lo tanto, usando la continuidad de ρ_{XY}^2 en (Φ_1, Ψ_1) dada en **S2**, se obtiene que $b_n \rightarrow 0$, lo que concluye la demostración de i).

ii) Tenemos que $\rho_{XY}^2(\hat{\Phi}_1, \hat{\Psi}_1) = \rho_n^2(\hat{\Phi}_1, \hat{\Psi}_1) - c_n$, donde $c_n = \rho_n^2(\hat{\Phi}_1, \hat{\Psi}_1) - \rho_{XY}^2(\hat{\Phi}_1, \hat{\Psi}_1)$. Por (4.15), $c_n \xrightarrow{p} 0$. Por lo tanto, $\rho_{XY}^2(\hat{\Phi}_1, \hat{\Psi}_1) - \rho_1^2 = \rho_n^2(\hat{\Phi}_1, \hat{\Psi}_1) - \rho_1^2 - c_n = d_n - c_n$. El resultado se obtiene ahora en forma inmediata pues $d_n \xrightarrow{p} 0$ por i).

iii) Probaremos que $\rho_{XX}^2(\hat{\Phi}_1, \Phi_1) \xrightarrow{p} 1$ ya que la demostración de que $\rho_{YY}^2(\hat{\Psi}_1, \Psi_1) \xrightarrow{p} 1$ es análoga. Supongamos que $\rho_{XX}^2(\hat{\Phi}_1, \Phi_1)$ no converja en probabilidad a 1. Entonces, existe $\varepsilon > 0$ suficientemente chico de manera que $\mathbb{P}\left(|\rho_{XX}^2(\hat{\Phi}_1, \Phi_1) - 1| > \varepsilon\right)$ no converge a 0. Por lo tanto, existen $\delta > 0$ y una subsucesión $(n_j)_{j \in \mathbb{N}}$ tal que

$$\mathbb{P}\left(|\rho_{XX}^2(\hat{\Phi}_{1,n_j}, \Phi_1) - 1| > \varepsilon\right) > \delta \quad \forall j. \quad (4.16)$$

Por ii) $\rho_{XY}^2(\hat{\Phi}_1, \hat{\Psi}_1) \xrightarrow{p} \rho_1^2$, por lo tanto, $\rho_{XY}^2(\hat{\Phi}_{1,n_j}, \hat{\Psi}_{1,n_j}) \xrightarrow{p} \rho_1^2$. Por lo tanto, existe una subsucesión $(n_{j_k})_{k \in \mathbb{N}}$ tal que $\rho_{XY}^2(\hat{\Phi}_{1,n_{j_k}}, \hat{\Psi}_{1,n_{j_k}}) \xrightarrow{a.s.} \rho_1^2$ (ver Teorema 3.4 en Gut, 2005). Esta convergencia junto con la Proposición 4.3.1 implican que existe \mathcal{N} tal que $\mathbb{P}(\mathcal{N}) = 0$ y para todo $\omega \notin \mathcal{N}$, $\rho_{XX}^2(\hat{\Phi}_{1,n_{j_k}}(\omega), \Phi_1) \rightarrow 1$, con lo cual también vale que $\rho_{XX}^2(\hat{\Phi}_{1,n_{j_k}}, \Phi_1) \xrightarrow{p} 1$ lo que contradice con (4.16). \square

El siguiente Teorema muestra la consistencia fuerte de los estimadores de la primera correlación canónica y de sus direcciones canónicas asociadas.

Teorema 4.3.2. (Consistencia fuerte) Bajo **S1F** y **S2**, si $\lim_{n \rightarrow \infty} p_n = \infty$ y $\lim_{n \rightarrow \infty} q_n = \infty$, tenemos que

$$i) \tilde{\rho}_1^2 \xrightarrow{a.s.} \rho_1^2$$

$$ii) \rho_{XY}^2(\hat{\Phi}_1, \hat{\Psi}_1) \xrightarrow{a.s.} \rho_1^2$$

$$iii) \text{ Si además valen } \mathbf{S3} \text{ y } \mathbf{S4}, \text{ entonces } \rho_{XX}^2(\Phi_1, \hat{\Phi}_1) \xrightarrow{a.s.} 1 \text{ y } \rho_{YY}^2(\Psi_1, \hat{\Psi}_1) \xrightarrow{a.s.} 1$$

Demostración. La demostración de i) y ii) es análoga a la del Teorema 4.3.1, mientras que la demostración de iii) es una consecuencia directa de la Proposición 4.3.1 utilizando ii). \square

4.4. Apéndice 1

4.4.1. Demostración de la Proposición 4.2.1

Para probar la Proposición 4.2.1, nos resultará de utilidad la siguiente notación. Dado un conjunto $\mathcal{A} \subset \mathcal{H}$, $\overline{\mathcal{A}}$ indicará complemento del conjunto \mathcal{A} . Sea

$$\mathcal{B}_{1,p_n}(R) = \{u \in \mathcal{H}_{1,p_n} : \|u\|_{\mathcal{H}_1} \leq R\}$$

y para $u_{p_n} \in \mathcal{H}_{1,p_n}$, denotemos por $\mathcal{B}_{1,p_n}(u_{p_n}, R) = u_{p_n} + \mathcal{B}_{1,p_n}(R)$. En forma análoga, definimos $\mathcal{B}_{2,q_n}(R)$ y $\mathcal{B}_{2,q_n}(v_{q_n}, R)$ para $v_{q_n} \in \mathcal{H}_{2,q_n}$. Sea $\mathcal{B}_{p_n,q_n}(R) = \mathcal{B}_{1,p_n}(R) \times \mathcal{B}_{2,q_n}(R)$ y $\mathcal{B}_{p_n,q_n}(\mathbf{w}_{p_n,q_n}, R) = \mathbf{w}_{p_n,q_n} + \mathcal{B}_{p_n,q_n}(R)$ con $\mathbf{w}_{p_n,q_n} = (u_{p_n}, v_{q_n}) \in \mathcal{H}_{1,p_n} \times \mathcal{H}_{2,q_n}$

Observemos que $\mathcal{B}_{p_n,q_n}(R) \subseteq \{\mathbf{w} = (u, v) \in \mathcal{H}_{1,p_n} \times \mathcal{H}_{2,q_n} : \|\mathbf{w}\|_{\mathcal{H}} \leq \sqrt{2}R\}$. Por lo tanto, como $\mathcal{H}_n = \mathcal{H}_{1,p_n} \times \mathcal{H}_{2,q_n}$ es un espacio finito-dimensional de dimensión $p_n + q_n$, por el Lema 2.5 de van de Geer (2000) obtenemos que dado $\delta < R$, podemos cubrir a $\mathcal{B}_{p_n,q_n}(R)$ por un número finito $\ell_n = C/\delta^{p_n+q_n}$ de conjuntos de la forma $\mathcal{B}_{p_n,q_n}(\mathbf{w}_{p_n,q_n}^{(j)}, \delta)$. Más precisamente, existen $\mathbf{w}_{p_n,q_n}^{(j)} \in \mathcal{H}_n$, $1 \leq j \leq \ell_n$ tales que $\mathcal{B}_{p_n,q_n}(R) \subset \bigcup_{j=1}^{\ell_n} \mathcal{B}_{p_n,q_n}(\mathbf{w}_{p_n,q_n}^{(j)}, \delta)$ donde

$$\ell_n = \left(\frac{4\sqrt{2}R + \delta}{\delta} \right)^{p_n+q_n} \leq \left(\frac{5\sqrt{2}R}{\delta} \right)^{p_n+q_n}. \quad (4.17)$$

Observemos que el mismo resultado vale para conjuntos de la forma $\mathbf{w}_n + \mathcal{B}_{p_n,q_n}(\delta)$ con $\mathbf{w}_n = (u_{p_n}, v_{q_n}) \in \mathcal{H}_{1,p_n} \times \mathcal{H}_{2,q_n}$.

Demostración de la Proposición 4.2.1. Sea $r_n(\mathbf{w}) = \rho_n^2(\mathbf{w}) - \rho_{XY}^2(\theta) - (1/n) \sum_{i=1}^n \alpha(Z_i, \mathbf{w})$,

entonces se cumple que

$$\sup_{\mathbf{w} \in \mathcal{S}_n} |\rho_n^2(\mathbf{w}) - \rho_{XY}^2(\mathbf{w})| \leq \sup_{\mathbf{w} \in \mathcal{S}_n} |r_n(\mathbf{w})| + \sup_{\mathbf{w} \in \mathcal{S}_n} \left| \frac{1}{n} \sum_{i=1}^n \alpha(Z_i, \mathbf{w}) \right|.$$

Por **A1**, $\sup_{\mathbf{w} \in \mathcal{S}_n} |r_n(\mathbf{w})| \xrightarrow{p} 0$, por lo tanto, para probar la Proposición basta ver que

$$\sup_{\mathbf{w} \in \mathcal{S}_n} \left| \frac{1}{n} \sum_{i=1}^n \alpha(Z_i, \mathbf{w}) \right| \xrightarrow{p} 0.$$

Sea $A(z) = \sup_{\mathbf{w} \in \mathcal{S}} |\alpha(z, \mathbf{w})|$. Por **A1a**) $\mathbb{E}A(Z) < \infty$, por lo tanto, dado $\epsilon > 0$, existe $C > 0$ tal que $\mathbb{E}A(Z)\mathbb{I}_{\overline{\mathcal{A}}_C}(Z) < \epsilon/3$ donde $\mathcal{A}_C = \{z \in \mathcal{H} : A(z) \leq C\}$. Sean

$$\begin{aligned} A_{1,n} &= \sup_{\mathbf{w} \in \mathcal{S}_n} \left| \frac{1}{n} \sum_{i=1}^n \alpha(Z_i, \mathbf{w}) \mathbb{I}_{\mathcal{A}_C}(Z_i) - \mathbb{E} \alpha(Z_i, \mathbf{w}) \mathbb{I}_{\mathcal{A}_C}(Z_i) \right| \\ A_{2,n} &= \frac{1}{n} \sum_{i=1}^n \sup_{\mathbf{w} \in \mathcal{S}} |\alpha(Z_i, \mathbf{w})| \mathbb{I}_{\overline{\mathcal{A}}_C}(Z_i). \end{aligned}$$

Observemos que la ley de los grandes números implica que $A_{2,n} \xrightarrow{p} \mathbb{E}A(Z)\mathbb{I}_{\overline{\mathcal{A}}_C}(Z) < \epsilon/3$, con lo cual $\pi_n(\epsilon) = \mathbb{P}(A_{2,n} > 2\epsilon/3) \rightarrow 0$. Usando que $\mathbb{E}\alpha(Z, \mathbf{w}) = 0$, obtenemos

$$\begin{aligned} \sup_{\mathbf{w} \in \mathcal{S}_n} \left| \frac{1}{n} \sum_{i=1}^n \alpha(Z_i, \mathbf{w}) \right| &\leq \sup_{\mathbf{w} \in \mathcal{S}_n} \left| \frac{1}{n} \sum_{i=1}^n \alpha(Z_i, \mathbf{w}) \mathbb{I}_{\mathcal{A}_C}(Z_i) - \mathbb{E} \alpha(Z_i, \mathbf{w}) \mathbb{I}_{\mathcal{A}_C}(Z_i) \right| \\ &\quad + \sup_{\mathbf{w} \in \mathcal{S}_n} \left| \frac{1}{n} \sum_{i=1}^n \alpha(Z_i, \mathbf{w}) \mathbb{I}_{\overline{\mathcal{A}}_C}(Z_i) - \mathbb{E} \alpha(Z_i, \mathbf{w}) \mathbb{I}_{\overline{\mathcal{A}}_C}(Z_i) \right| \\ &\leq A_{1,n} + A_{2,n} + \mathbb{E}A(Z)\mathbb{I}_{\overline{\mathcal{A}}_C}(Z) \leq A_{1,n} + A_{2,n} + \frac{\epsilon}{3}, \end{aligned} \quad (4.18)$$

pues $\mathbb{E}A(Z)\mathbb{I}_{\overline{\mathcal{A}}_C}(Z) < \epsilon/3$. Usando que $\pi_n(\epsilon) \rightarrow 0$ y (4.18), concluimos que

$$\mathbb{P} \left(\sup_{\mathbf{w} \in \mathcal{S}_n} \left| \frac{1}{n} \sum_{i=1}^n \alpha(Z_i, \mathbf{w}) \right| > 4\epsilon \right) \leq \pi_n(\epsilon) + \mathbb{P}(A_{1,n} > 3\epsilon).$$

Por lo tanto, basta probar que

$$\mathbb{P} \left(\sup_{\mathbf{w} \in \mathcal{S}_n} \left| \frac{1}{n} \sum_{i=1}^n \Psi(Z_i, \mathbf{w}) - \mathbb{E} \Psi(Z_i, \mathbf{w}) \right| > 3\epsilon \right) \rightarrow 0, \quad (4.19)$$

donde $\Psi(z, \mathbf{w}) = \alpha(z, \mathbf{w})\mathbb{I}_{\mathcal{A}_C}(z)$ es tal que $|\Psi(z, \mathbf{w})| \leq C$ y $\mathbb{E} \sup_{\mathbf{w} \in \mathcal{S}} |\Psi(Z, \mathbf{w})| < \infty$. Como $\mathcal{H} = \mathcal{H}_1 \times \mathcal{H}_2$ es separable y completo, es σ -compacto, es decir, $\mathcal{H} = \bigcup_{m=1}^{\infty} \mathcal{K}_m$, con $\mathcal{K}_m \subset \mathcal{H}$ conjuntos compactos. Sin pérdida de generalidad podemos suponer que $\mathcal{K}_m \subset \mathcal{K}_{m+1}$. Dado $\epsilon > 0$ existe $M \in \mathbb{N}$ tal que

$$\mathbb{E} \sup_{\mathbf{w} \in \mathcal{S}} |\Psi(Z, \mathbf{w})| \mathbb{I}_{\bar{\mathcal{K}}_M}(Z) < \frac{\epsilon}{3}. \quad (4.20)$$

Descompongamos a la función $\Psi(z, \mathbf{w}) - \mathbb{E}\Psi(Z, \mathbf{w})$ como $\Psi_1(z, \mathbf{w}) + \Psi_2(z, \mathbf{w})$ donde

$$\begin{aligned} \Psi_1(z, \mathbf{w}) &= \Psi(z, \mathbf{w})\mathbb{I}_{\mathcal{K}_M}(z) - \mathbb{E}\Psi(Z, \mathbf{w})\mathbb{I}_{\mathcal{K}_M}(Z) \\ \Psi_2(z, \mathbf{w}) &= \Psi(z, \mathbf{w})\mathbb{I}_{\bar{\mathcal{K}}_M}(z) - \mathbb{E}\Psi(Z, \mathbf{w})\mathbb{I}_{\bar{\mathcal{K}}_M}(Z). \end{aligned}$$

Luego, para probar (4.19) basta demostrar que

$$\mathbb{P} \left(\sup_{\mathbf{w} \in \mathcal{S}_n} \left| \frac{1}{n} \sum_{i=1}^n \Psi_1(Z_i, \mathbf{w}) \right| > 2\epsilon \right) \rightarrow 0 \quad (4.21)$$

$$\mathbb{P} \left(\sup_{\mathbf{w} \in \mathcal{S}_n} \left| \frac{1}{n} \sum_{i=1}^n \Psi_2(Z_i, \mathbf{w}) \right| > \epsilon \right) \rightarrow 0. \quad (4.22)$$

Empezaremos probando (4.22). Usando (4.20), deducimos que

$$\frac{1}{n} \sum_{i=1}^n \sup_{\mathbf{w} \in \mathcal{S}} |\Psi_2(Z_i, \mathbf{w})| \leq \frac{1}{n} \sum_{i=1}^n \sup_{\mathbf{w} \in \mathcal{S}} |\Psi(Z_i, \mathbf{w})| \mathbb{I}_{\bar{\mathcal{K}}_M}(Z_i) + \frac{\epsilon}{3}. \quad (4.23)$$

Por otra parte, usando nuevamente (4.20) y la ley de los grandes números obtenemos que $(1/n) \sum_{i=1}^n \sup_{\mathbf{w} \in \mathcal{S}} |\Psi(Z_i, \mathbf{w})| \mathbb{I}_{\bar{\mathcal{K}}_M}(Z_i) \xrightarrow{p} \mathbb{E} |\Psi(Z, \mathbf{w})| \mathbb{I}_{\bar{\mathcal{K}}_M}(Z) < \epsilon/3$, de donde, usando (4.23), concluimos la demostración de (4.22).

Probaremos ahora (4.21). Por (4.17), existen $\mathbf{w}^{(j)} = \mathbf{w}_{p_n, q_n}^{(j)} \in \mathcal{H}_n$, $1 \leq j \leq \ell_n$ tales que $\mathcal{B}_{p_n, q_n}(1) \subset \bigcup_{j=1}^{\ell_n} \mathcal{B}_{p_n, q_n}(\mathbf{w}^{(j)}, \delta)$ con $\ell_n \leq (5\sqrt{2}/\delta)^{p_n + q_n}$. En particular, $\mathcal{S}_n \subset \bigcup_{j=1}^{\ell_n} \mathcal{V}_j$ donde, por simplicidad, llamamos $\mathcal{V}_j = \mathcal{B}_{p_n, q_n}(\mathbf{w}^{(j)}, \delta)$. Por lo tanto, por **A1b**), tenemos que existe $0 < \delta < 1$ tal que si $\mathbf{w}_1, \mathbf{w}_2 \in \mathcal{S}_n$ son tales que $\|\mathbf{w}_1 - \mathbf{w}_2\|_{\mathcal{H}} < \delta$ entonces $|\alpha(z, \mathbf{w}_1) -$

$\alpha(z, \mathbf{w}_2)| \leq \epsilon/2$ para todo $z \in \mathcal{K}_M$, de donde $|\Psi(z, \mathbf{w}_1) - \Psi(z, \mathbf{w}_2)| \leq \epsilon/2$ para todo $z \in \mathcal{K}_M$ y

$$\begin{aligned} \max_{1 \leq j \leq \ell_n} \sup_{\mathbf{w} \in \mathcal{V}_j} \frac{1}{n} \left| \sum_{i=1}^n (\Psi(Z_i, \mathbf{w}) - \Psi(Z_i, \mathbf{w}^{(j)})) \mathbb{I}_{\mathcal{K}_M}(Z_i) \right| &\leq \frac{\epsilon}{2} \\ \max_{1 \leq j \leq \ell_n} \sup_{\mathbf{w} \in \mathcal{V}_j} \mathbb{E} \left| (\Psi(Z, \mathbf{w}) - \Psi(Z, \mathbf{w}^{(j)})) \mathbb{I}_{\mathcal{K}_M}(Z) \right| &\leq \frac{\epsilon}{2}. \end{aligned}$$

Por lo tanto,

$$\max_{1 \leq j \leq \ell_n} \sup_{\mathbf{w} \in \mathcal{V}_j} \frac{1}{n} \left| \sum_{i=1}^n (\Psi_1(Z_i, \mathbf{w}) - \Psi_1(Z_i, \mathbf{w}^{(j)})) \right| \leq \epsilon,$$

lo que implica

$$\sup_{\mathbf{w} \in \mathcal{S}_n} \frac{1}{n} \left| \sum_{i=1}^n \Psi_1(Z_i, \mathbf{w}) \right| \leq \max_{1 \leq j \leq \ell_n} \frac{1}{n} \left| \sum_{i=1}^n \Psi_1(Z_i, \mathbf{w}^{(j)}) \right| + \epsilon.$$

Luego, para ver (4.21) bastará probar que

$$\mathbb{P} \left(\max_{1 \leq j \leq \ell_n} \frac{1}{n} \left| \sum_{i=1}^n \Psi_1(Z_i, \mathbf{w}^{(j)}) \right| > \epsilon \right) \rightarrow 0.$$

Observemos que

$$\mathbb{P} \left(\max_{1 \leq j \leq \ell_n} \frac{1}{n} \left| \sum_{i=1}^n \Psi_1(Z_i, \mathbf{w}^{(j)}) \right| > \epsilon \right) \leq \ell_n \max_{1 \leq j \leq \ell_n} \mathbb{P} \left(\frac{1}{n} \left| \sum_{i=1}^n \Psi_1(Z_i, \mathbf{w}^{(j)}) \right| > \epsilon \right).$$

Como $\mathbb{E}\Psi_1(Z, \mathbf{w}) = 0$ y $|\Psi_1(Z, \mathbf{w})| \leq 2C$, la desigualdad de Bernstein implica que, para todo $\mathbf{w} \in \mathcal{S}_n$,

$$\mathbb{P} \left(\frac{1}{n} \left| \sum_{i=1}^n \Psi_1(Z_i, \mathbf{w}) \right| > \epsilon \right) \leq 2 \exp\{-C_1 n \epsilon^2\}$$

donde $C_1 = 1/(8C^2)$. Por lo tanto,

$$\begin{aligned} \mathbb{P} \left(\max_{1 \leq j \leq \ell_n} \frac{1}{n} \left| \sum_{i=1}^n \Psi_1(Z_i, \mathbf{w}^{(j)}) \right| > \epsilon \right) &\leq \left(\frac{5\sqrt{2}}{\delta} \right)^{p_n+q_n} \exp\{-C_1 n \epsilon^2\} \\ &= \exp \left\{ -n \left(C_1 \epsilon^2 - C_2 \frac{p_n + q_n}{n} \right) \right\}, \end{aligned}$$

donde $C_2 = \log(5\sqrt{2}/\delta)$. Por **A2**, $(p_n + q_n)/n \rightarrow 0$, por lo tanto, el miembro derecho de la desigualdad converge a 0, lo que concluye la demostración. \square

4.4.2. Demostración de la Proposición 4.2.2 y condiciones que garantizan **A3D**

Demostración de la Proposición 4.2.2. Sea $n_1 \in \mathbb{N}$, con $n_1 \geq 2$ tal que si $n \geq n_1$ entonces

$$\mathbb{P} \left(\sup_{u \in \mathcal{S}_{1,1}, v \in \mathcal{S}_{2,1}} |\rho_n(u, v) - \rho(u, v)| > \frac{1}{2} \right) < \frac{1}{2}.$$

Tomemos $p_j = q_j = 1$ si $j < n_1$. Sea ahora $n_2 > n_1$ tal que si $n \geq n_2$, entonces

$$\mathbb{P} \left(\sup_{u \in \mathcal{S}_{1,2}, v \in \mathcal{S}_{2,2}} |\rho_n(u, v) - \rho(u, v)| > \frac{1}{3} \right) < \frac{1}{3},$$

y definamos $p_j = q_j = 2$ si $n_1 \leq j < n_2$. De esta forma, se construye una sucesión creciente $\{n_k\}_{k \in \mathbb{N}}$ tal que $p_j = q_j = k$ si $n_{k-1} \leq j < n_k$ y para todo $n \geq n_k$

$$\mathbb{P} \left(\sup_{u \in \mathcal{S}_{1,k}, v \in \mathcal{S}_{2,k}} |\rho_n(u, v) - \rho(u, v)| > \frac{1}{k+1} \right) < \frac{1}{k+1}.$$

Para probar que esta sucesión cumple **S1D**, veremos que para todo $0 < \epsilon < 1$ y $0 < \delta < 1$, existe n^* tal que si $n \geq n^*$ entonces

$$\mathbb{P} \left(\sup_{u \in \mathcal{S}_{1,p_n}, v \in \mathcal{S}_{2,q_n}} |\rho_n(u, v) - \rho(u, v)| > \epsilon \right) < \delta.$$

Sean $n_0 = 1$ y $k_0 \in \mathbb{N}$ tal que si $k \geq k_0$ entonces

$$\frac{1}{k} < \min\{\epsilon, \delta\}.$$

Sea $n^* = n_{k_0-1}$. Dado $n \geq n^*$, existe $k \geq k_0$ tal que $n_{k-1} \leq n < n_k$. Por lo tanto, $p_n = q_n = k$

y

$$\begin{aligned} \mathbb{P} \left(\sup_{u \in \mathcal{S}_{1,p_n}, v \in \mathcal{S}_{2,q_n}} |\rho_n(u, v) - \rho(u, v)| > \epsilon \right) &= \mathbb{P} \left(\sup_{u \in \mathcal{S}_{1,k}, v \in \mathcal{S}_{2,k}} |\rho_n(u, v) - \rho(u, v)| > \epsilon \right) \\ &\leq \mathbb{P} \left(\sup_{u \in \mathcal{S}_{1,k}, v \in \mathcal{S}_{2,k}} |\rho_n(u, v) - \rho(u, v)| > \frac{1}{k+1} \right) \\ &< \frac{1}{k+1} < \delta, \end{aligned}$$

lo que concluye la demostración. □

Queremos ver que existen un elemento aleatorio $(X, Y)^T \sim P$ y una medida de correlación ρ para las que vale la condición **A3D**. De ahora en más indicaremos por $d_{\text{PR}, \mathcal{M}}(Q_1, Q_2)$ a la distancia de Prohorov entre dos probabilidades Q_1 y Q_2 definidas sobre el espacio métrico (\mathcal{M}, d) , es decir, dadas dos medidas de probabilidad Q_1 y Q_2 sobre \mathcal{M}

$$d_{\text{PR}, \mathcal{M}}(Q_1, Q_2) = \inf\{\epsilon > 0 : Q_1(B) \leq Q_2(B^\epsilon) + \epsilon \text{ para todo conjunto } B \text{ boreliano}\},$$

donde $B^\epsilon = \{m \in \mathcal{M} : d(m, B) < \epsilon\}$.

Antes de dar condiciones para la validez de **A3D**, fijaremos algo de notación y probaremos los Lemas 4.4.1 y 4.4.2 que son necesarios para la demostración de las Proposiciones 4.4.1 y 4.5.2.

Recordemos que el operador $A_{uv} : \mathcal{H} \rightarrow \mathbb{R}^2$ se definía como $A_{uv}(x, y) = (\langle u, x \rangle_{\mathcal{H}_1}, \langle v, y \rangle_{\mathcal{H}_2})^T$. El operador A_{uv} es un operador lineal acotado, más aún, si $\|u\|_{\mathcal{H}_1} = \|v\|_{\mathcal{H}_2} = 1$, entonces $\|A_{uv}z\|_{\mathbb{R}^2} \leq \|z\|_{\mathcal{H}}$ para todo $z = (x, y)^T \in \mathcal{H}$. Para un operador lineal acotado $A : \mathcal{H} \rightarrow \mathbb{R}^2$, indicaremos por $\|A\|_{\text{OP}}$ la norma del operador A , es decir, $\|A\|_{\text{OP}} = \sup_{\|z\|_{\mathcal{H}}=1} \|Az\|_{\mathbb{R}^2}$. Por lo tanto, $\|A_{uv}\|_{\text{OP}} \leq \|(u, v)\|_{\mathcal{H}}$ y $A_{uv} - A_{u_0v_0} = A_{u-u_0, v-v_0}$.

Dada una medida de probabilidad Q sobre \mathcal{H} y un operador lineal acotado $A : \mathcal{H} \rightarrow \mathbb{R}^2$, definimos una medida de probabilidad en \mathbb{R}^2 que notaremos con QA^{-1} , de la siguiente manera: dado $B \subset \mathbb{R}^2$ boreliano, $QA^{-1}(B) = Q(A^{-1}(B))$. En particular, si $A = A_{uv}$ la indicaremos por Q_{uv} . Por lo tanto, si $(X, Y)^T \sim P$ con esta nueva notación tenemos que $P[\langle u, X \rangle, \langle v, Y \rangle] = PA_{uv}^{-1} = P_{uv}$, es decir, $(\langle u, X \rangle_{\mathcal{H}_1}, \langle v, Y \rangle_{\mathcal{H}_2}) \sim P_{uv}$ y $\rho_{XY}(u, v) = \rho(P_{uv})$.

Lema 4.4.1. Sean Q_1 y Q_2 medidas en \mathcal{H} tales que $d_{\text{PR}, \mathcal{H}}(Q_1, Q_2) \leq \delta$. Si $A : \mathcal{H} \rightarrow \mathbb{R}^2$ es un operador lineal acotado tal que $\|A\|_{\text{OP}} \leq 1$ entonces $d_{\text{PR}, \mathbb{R}^2}(Q_1A^{-1}, Q_2A^{-1}) \leq \delta$.

Demostración. Queremos probar que $d_{\text{PR}, \mathbb{R}^2}(Q_1A^{-1}, Q_2A^{-1}) \leq \delta$. Basta ver que, dado $\epsilon > \delta$, se verifica $Q_1A^{-1}(B) \leq Q_2A^{-1}(B^\epsilon) + \epsilon$ para todo $B \in \mathbb{R}^2$ boreliano. Como $d_{\text{PR}, \mathcal{H}}(Q_1, Q_2) \leq \delta$ tenemos que

$$Q_1A^{-1}(B) = Q_1(A^{-1}(B)) \leq Q_2([A^{-1}(B)]^\epsilon) + \epsilon.$$

Por otro lado, como $\|A\|_{\text{OP}} \leq 1$ deducimos que $[A^{-1}(B)]^\epsilon \subset A^{-1}(B^\epsilon)$. Por lo tanto, se tiene

$$Q_1 A^{-1}(B) \leq Q_2 (A^{-1}(B^\epsilon)) + \epsilon = Q_2 A^{-1}(B^\epsilon) + \epsilon$$

como queríamos probar. \square

Lema 4.4.2. Sean \mathcal{H}_j espacios de Hilbert separables, $\mathcal{H} = \mathcal{H}_1 \times \mathcal{H}_2$ y $B_1 = \{\delta_1, \delta_2, \dots\}$ y $B_2 = \{\eta_1, \eta_2, \dots\}$ bases ortonormales de \mathcal{H}_1 y \mathcal{H}_2 , respectivamente. Sea $Z = (X, Y)^\text{T} \in \mathcal{H}$ un elemento aleatorio tal que $Z \sim P$. Sean $p \in \mathbb{N}$ y $q \in \mathbb{N}$ fijos. Si, para todo $(u_0, v_0) \in \mathcal{S}_{1,p} \times \mathcal{S}_{2,q}$ la medida de asociación ρ es continua respecto de la métrica de Prohorov en la distribución bivariada $P_{u_0 v_0}$, entonces, la restricción de ρ_{XY} a $\mathcal{S}_{1,p} \times \mathcal{S}_{2,q}$ es continua, o sea, para todo $(u_0, v_0) \in \mathcal{S}_{1,p} \times \mathcal{S}_{2,q}$, dado $\epsilon > 0$, existe $\delta = \delta(u_0, v_0) > 0$ tal que

$$\|(u, v) - (u_0, v_0)\|_{\mathcal{H}} < \delta, (u, v) \in \mathcal{S}_{1,p} \times \mathcal{S}_{2,q} \implies |\rho_{XY}(u, v) - \rho_{XY}(u_0, v_0)| < \epsilon. \quad (4.24)$$

Demostración. Sean $(u_0, v_0) \in \mathcal{S}_{1,p} \times \mathcal{S}_{2,q}$ y $\epsilon > 0$. Como ρ es continua en $P_{u_0 v_0}$, existe $\tau = \tau(u_0, v_0) > 0$ tal que, para toda medida de probabilidad Q sobre \mathbb{R}^2 , se verifica

$$d_{\text{PR}, \mathbb{R}^2}(Q, P_{u_0 v_0}) \leq \tau \implies |\rho(Q) - \rho(P_{u_0 v_0})| < \epsilon. \quad (4.25)$$

Como \mathcal{H} es separable, P resulta una medida ajustada, es decir, existe $\mathcal{K} \subset \mathcal{H}$ compacto tal que $P(\mathcal{K}) > 1 - \delta$. Como \mathcal{K} es compacto, es acotado. Supongamos que $\|z\|_{\mathcal{H}} \leq M$ para todo $z \in \mathcal{K}$. Sea $\delta = \tau/M$ y $(u, v) \in \mathcal{S}_{1,p} \times \mathcal{S}_{2,q}$ con $\|(u, v) - (u_0, v_0)\|_{\mathcal{H}} < \delta$ entonces $\|A_{uv} - A_{u_0 v_0}\| < \delta$.

Queremos ver que $d_{\text{PR}, \mathbb{R}^2}(P_{uv}, P_{u_0 v_0}) \leq \tau$ de donde por (4.25), se obtiene que $|\rho_{XY}(u, v) - \rho_{XY}(u_0, v_0)| = |\rho(P_{uv}) - \rho(P_{u_0 v_0})| < \epsilon$, como queríamos probar.

Sea $B \subset \mathbb{R}^2$ un conjunto boreliano. Podemos escribir al conjunto $A_{u_0 v_0}^{-1}(B)$ como

$$A_{u_0 v_0}^{-1}(B) = (A_{u_0 v_0}^{-1}(B) \cap \mathcal{K}) \cup (A_{u_0 v_0}^{-1}(B) \cap \overline{\mathcal{K}}). \quad (4.26)$$

Dado $z \in A_{u_0 v_0}^{-1}(B) \cap \mathcal{K}$ como $z \in \mathcal{K}$, $\|z\|_{\mathcal{H}} \leq M$ de donde

$$\|A_{uv} z - A_{u_0 v_0} z\| < \frac{\tau}{M} \|z\|_{\mathcal{H}} \leq \delta,$$

con lo cual, como $A_{u_0v_0}z \in B$, se obtiene $A_{uv}z \in B^\delta$, es decir, $z \in A_{uv}^{-1}(B^\delta)$. Por lo tanto, tenemos que $P(A_{u_0v_0}^{-1}(B) \cap K) \leq P(A_{uv}^{-1}B^\delta)$, de donde, usando (4.26) y que $P(\overline{K}) < \delta$, obtenemos

$$P(A_{u_0v_0}^{-1}(B)) = P(A_{u_0v_0}^{-1}(B) \cap K) + P(A_{u_0v_0}^{-1}(B) \cap \overline{K}) \leq P_{uv}(B^\delta) + \eta.$$

Por la simetría de la distancia de Prohorov, hemos mostrado que si $\|(u, v) - (u_0, v_0)\|_{\mathcal{H}} < \delta$, entonces $d_{\text{PR}, \mathbb{R}^2}(P_{uv}, P_{u_0v_0}) \leq \tau$ lo que implica que $|\rho_{XY}(u, v) - \rho_{XY}(u_0, v_0)| = |\rho(P_{uv}) - \rho(P_{u_0v_0})| < \epsilon$. \square

Proposición 4.4.1. *Sean \mathcal{H}_j espacios de Hilbert separables, $\mathcal{H} = \mathcal{H}_1 \times \mathcal{H}_2$ y $B_1 = \{\delta_1, \delta_2, \dots\}$ y $B_2 = \{\eta_1, \eta_2, \dots\}$ bases ortonormales de \mathcal{H}_1 y \mathcal{H}_2 , respectivamente. Sea $Z = (X, Y)^T \in \mathcal{H}$ un elemento aleatorio tal que $Z \sim P$ y $p \in \mathbb{N}$ y $q \in \mathbb{N}$ fijos. Supongamos que para todo $(u_0, v_0) \in \mathcal{S}_{1,p} \times \mathcal{S}_{2,q}$, la medida de asociación ρ es continua respecto de la métrica de Prohorov en la distribución bivariada $P_{u_0v_0}$, entonces dados $\epsilon > 0$ y $\mathbf{w}_0 = (u_0, v_0) \in \mathcal{S}_{1,p} \times \mathcal{S}_{2,q}$, existen $0 < \delta = \delta(\mathbf{w}_0, p, q, P)$ y $0 < \tau = \tau(\mathbf{w}_0, p, q, P)$ tales que si $d_{\text{PR}, \mathcal{H}}(Q, P) < \tau$ y $\mathbf{w} = (u, v) \in \mathcal{S}_{1,p} \times \mathcal{S}_{2,q}$ es tal que $\|\mathbf{w} - \mathbf{w}_0\|_{\mathcal{H}} < \delta$, entonces $|\rho(Q_{uv}) - \rho(P_{u_0v_0})| < \epsilon$.*

Demostración. Sea $(u_0, v_0) \in \mathcal{S}_{1,p} \times \mathcal{S}_{2,q}$. Como ρ es continua respecto de la métrica de Prohorov en P_{u_0, v_0} , por el Lema 4.4.2 existe $\delta_1 = \delta(u_0, v_0) > 0$ que verifica (4.24), es decir, existe $\delta_1 = \delta(u_0, v_0) > 0$ tal que

$$\|(u, v) - (u_0, v_0)\|_{\mathcal{H}} < \delta_1, (u, v) \in \mathcal{S}_{1,p} \times \mathcal{S}_{2,q} \implies |\rho(P_{uv}) - \rho(P_{u_0v_0})| < \frac{\epsilon}{2}. \quad (4.27)$$

Por otra parte, hemos visto en la demostración del Lema 4.4.2 que la continuidad ρ en $P_{u_0v_0}$, implica que existe $\tau_1 = \tau(u_0, v_0) > 0$ tal que, para toda medida de probabilidad R sobre \mathbb{R}^2 , se verifica

$$d_{\text{PR}, \mathbb{R}^2}(R, P_{u_0v_0}) \leq \tau_1 \implies |\rho(R) - \rho(P_{u_0v_0})| < \frac{\epsilon}{2}. \quad (4.28)$$

Sea $\tau = \tau_1$ y Q tal que $d_{\text{PR}, \mathcal{H}}(Q, P) < \tau$. Por el Lema 4.4.1, obtenemos que para todo

$(u, v) \in \mathcal{S}_{1,p} \times \mathcal{S}_{2,q}$, $d_{\text{PR}, \mathbb{R}^2}(Q_{uv}, P_{uv}) < \tau$ de donde, por (4.28) tenemos que

$$|\rho(Q_{uv}) - \rho(P_{uv})| < \frac{\epsilon}{2}. \quad (4.29)$$

Por lo tanto, usando (4.27) y (4.29), deducimos que si $\|(u, v) - (u_0, v_0)\|_{\mathcal{H}} < \delta_1$, $(u, v) \in \mathcal{S}_{1,p} \times \mathcal{S}_{2,q}$ y $d_{\text{PR}, \mathcal{H}}(Q, P) < \tau$ entonces $|\rho(Q_{uv}) - \rho(P_{u_0v_0})| < \epsilon$ lo que concluye la demostración. \square

Lema 4.4.3. *Sean \mathcal{H}_j espacios de Hilbert separables, $\mathcal{H} = \mathcal{H}_1 \times \mathcal{H}_2$ y $B_1 = \{\delta_1, \delta_2, \dots\}$ y $B_2 = \{\eta_1, \eta_2, \dots\}$ bases ortonormales de \mathcal{H}_1 y \mathcal{H}_2 , respectivamente. Sea $Z = (X, Y)^T$ un elemento aleatorio en \mathcal{H} tal que $Z \sim P$, ρ una medida de asociación y $p \in \mathbb{N}$ y $q \in \mathbb{N}$ fijos. Supongamos que para todo $(u_0, v_0) \in \mathcal{S}_{1,p} \times \mathcal{S}_{2,q}$, la medida de asociación ρ es continua respecto de la métrica de Prohorov en la distribución bivariada $P_{u_0v_0}$, entonces ρ verifica **A3D**.*

Demostración. Sean $p \in \mathbb{N}$ y $q \in \mathbb{N}$ fijos, como vale la conclusión de la Proposición 4.4.1, dado $\epsilon > 0$, para cada $\mathbf{w}_0 = (u_0, v_0) \in \mathcal{S}_{1,p} \times \mathcal{S}_{2,q}$, existen $0 < \delta_1 = \delta(\mathbf{w}_0)$ y $0 < \tau = \tau(\mathbf{w}_0)$ tales que si $d_{\text{PR}, \mathcal{H}}(Q, P) < \tau$ y $\mathbf{w} = (u, v) \in \mathcal{S}_{1,p} \times \mathcal{S}_{2,q}$ es tal que $\|\mathbf{w} - \mathbf{w}_0\|_{\mathcal{H}} < \delta_1$, entonces $|\rho(Q_{uv}) - \rho(P_{u_0v_0})| < \epsilon/2$. Por otra parte, como ρ es continua en la distribución bivariada $P_{u_0v_0}$, existe $0 < \delta_2 = \delta(\mathbf{w}_0)$ tal que $\|\mathbf{w} - \mathbf{w}_0\|_{\mathcal{H}} < \delta_2$, entonces $|\rho(P_{uv}) - \rho(P_{u_0v_0})| < \epsilon/2$, con lo cual si $\delta_0 = \min(\delta_1, \delta_2)$ se cumple que para toda probabilidad Q tal que $d_{\text{PR}, \mathcal{H}}(Q, P) < \tau$ y $\mathbf{w} = (u, v) \in \mathcal{S}_{1,p} \times \mathcal{S}_{2,q}$ tal que $\|\mathbf{w} - \mathbf{w}_0\|_{\mathcal{H}} < \delta_0$, entonces $|\rho(Q_{uv}) - \rho(P_{uv})| < \epsilon$.

Observemos que

$$\mathcal{S}_{1,p} \times \mathcal{S}_{2,q} \subset \bigcup_{\mathbf{w}=(u,v) \in \mathcal{S}_{1,p} \times \mathcal{S}_{2,q}} \mathcal{B}_{p,q}(\mathbf{w}, \delta_0(\mathbf{w})).$$

Luego, como $\mathcal{S}_{1,p} \times \mathcal{S}_{2,q}$ es compacto en $\mathcal{H}_{1,p} \times \mathcal{H}_{2,q}$, existen $\mathbf{w}_j = (u_j, v_j) \in \mathcal{H}_{1,p} \times \mathcal{H}_{2,q}$, $1 \leq j \leq \ell$, tales que $\mathcal{S}_{1,p} \times \mathcal{S}_{2,q} \subset \bigcup_{j=1}^{\ell} \mathcal{V}_j$ donde, por simplicidad, llamamos $\mathcal{V}_j = \mathcal{B}_{p,q}(\mathbf{w}_j, \delta_j)$ y $\delta_j = \delta_0(\mathbf{w}_j)$. Sean $\tau_j = \tau(\mathbf{w}_j)$, $\delta = \min_{1 \leq j \leq \ell} \delta_j$ y $\tau = \min_{1 \leq j \leq \ell} \tau_j$.

Si $d_{\text{PR},\mathcal{H}}(Q, P) < \tau$ y $\mathbf{w} = (u, v) \in \mathcal{V}_j$ se cumple que $|\rho(Q_{uv}) - \rho(P_{uv})| < \epsilon$, de donde como $\mathcal{S}_{1,p} \times \mathcal{S}_{2,q} \subset \bigcup_{j=1}^{\ell} \mathcal{V}_j$

$$\sup_{(u,v) \in \mathcal{S}_{1,p} \times \mathcal{S}_{2,q}} |\rho(Q_{uv}) - \rho(P_{uv})| = \max_{1 \leq j \leq \ell} \sup_{(u,v) \in \mathcal{V}_j} |\rho(Q_{uv}) - \rho(P_{uv})| < \epsilon \quad (4.30)$$

Por lo tanto, si P_n indica la distribución empírica y $P_{n,u,v} = P_n A_{uv}^{-1}$, tenemos que $\rho_n(u, v) = \rho(P_{n,u,v})$ y usando (4.30) deducimos que

$$\mathbb{P} \left(\sup_{(u,v) \in \mathcal{S}_{1,p} \times \mathcal{S}_{2,q}} |\rho(P_{n,u,v}) - \rho(P_{uv})| > \epsilon \right) \leq \mathbb{P} (d_{\text{PR},\mathcal{H}}(P_n, P) \geq \tau) .$$

Usando que $d_{\text{PR},\mathcal{H}}(P_n, P) \xrightarrow{p} 0$ (ver Lema 6.1. de Bali *et al.*, 2011), se obtiene que $Z = (X, Y)^{\text{T}} \sim P$ y ρ cumplen **A3D**. \square

4.5. Apéndice 2

En lo que sigue probaremos que si ρ es un funcional de correlación débilmente continuo, esto es continuo respecto de la métrica de Prohorov, entonces vale **S1F**. Para dar condiciones que garantizan **S1F** probaremos, en primer lugar, bajo un supuesto que llamaremos **A3F**, la existencia de sucesiones $p_n \xrightarrow{n \rightarrow \infty} \infty$ y $q_n \xrightarrow{n \rightarrow \infty} \infty$ para las cuales vale **S1F** análogamente a como hicéramos en la Proposición 4.2.2. Posteriormente, veremos que si ρ es débilmente continuo entonces vale la condición **A3F**.

Proposición 4.5.1. *Sean $(X, Y)^{\text{T}}$ un elemento aleatorio en \mathcal{H} y ρ una medida de asociación que cumple la condición **A3F** dada por*

$$\mathbf{A3F.} \text{ Para cualquier } p \text{ y } q \text{ en } \mathbb{N} \text{ fijos, } \sup_{u \in \mathcal{S}_{1,p}, v \in \mathcal{S}_{2,q}} |\rho_{XY}(u, v) - \rho_n(u, v)| \xrightarrow{a.s.} 0.$$

*Entonces, existen sucesiones $(p_n)_{n \in \mathbb{N}}$ y $(q_n)_{n \in \mathbb{N}}$ con $p_n \rightarrow \infty$ y $q_n \rightarrow \infty$, $p_n < n$, $q_n < n$ que verifican **S1F**.*

Demostración. Denotemos por $\rho_{n,\omega}(u, v) = \rho(P_{n,\omega}[\langle u, X \rangle_{\mathcal{H}_1}, \langle v, Y \rangle_{\mathcal{H}_2}])$ y consideremos los conjuntos

$$\mathcal{E}_{i,k} = \left\{ \omega \in \Omega : \sup_{u \in \mathcal{S}_{1,i}, v \in \mathcal{S}_{2,i}} |\rho_{XY}(u, v) - \rho_{n,\omega}(u, v)| < \frac{1}{i} \quad \forall n \geq k \right\}.$$

Entonces, para todo i, k tenemos que $\mathcal{E}_{i,k} \subseteq \mathcal{E}_{i,k+1}$, mientras que **A3F** implica que $\mathbb{P}(\bigcup_{k \in \mathbb{N}} \mathcal{E}_{i,k}) = 1$ para todo $i \in \mathbb{N}$. Por lo tanto, para todo $i \in \mathbb{N}$ fijo, $\lim_{k \rightarrow \infty} \mathbb{P}(\mathcal{E}_{i,k}) = 1$. Sea k_i tal que

$$\mathbb{P}(\mathcal{E}_{i,k_i}) \geq 1 - \frac{1}{i^2},$$

donde elegimos k_i de manera que $i < k_i < k_{i+1}$ para todo i . Sea entonces $\mathcal{E}^* = \bigcap_{\ell=1}^{\infty} \bigcup_{i \geq \ell} \mathcal{E}_{i,k_i}$.

Es fácil ver que $\mathbb{P}(\mathcal{E}^*) = 1$. Efectivamente, recordemos que $\bar{\mathcal{A}}$ indica el complemento del conjunto $\mathcal{A} \subset \mathcal{H}$, luego

$$\mathbb{P}(\bar{\mathcal{E}}^*) \leq \mathbb{P}\left(\bigcup_{i \geq \ell} \bar{\mathcal{E}}_{i,k_i}\right) \leq \sum_{i \geq \ell} \mathbb{P}(\bar{\mathcal{E}}_{i,k_i}) \geq \sum_{i \geq \ell} \frac{1}{i^2} \xrightarrow{\ell \rightarrow \infty} 0.$$

Como la desigualdad vale para todo ℓ , se tiene que $\mathbb{P}(\mathcal{E}^*) = 1$.

Definamos $p_n = q_n = 1$ si $n < k_2$. Para $i \geq 2$, si $k_i \leq n < k_{i+1}$, definimos $p_n = q_n = i$. Luego, $p_n < n$ y $q_n < n$. Por otro lado, tenemos que si $\omega \in \mathcal{E}^*$ entonces existe ℓ^* tal que $\omega \in \bigcap_{i \geq \ell^*} \mathcal{E}_{i,k_i}$. Por lo tanto,

$$\sup_{u \in \mathcal{S}_{1,i}, v \in \mathcal{S}_{2,i}} |\rho_{XY}(u, v) - \rho_{n,\omega}(u, v)| < \frac{1}{i} \quad \forall n \geq k_i, \forall i \geq \ell^*.$$

Sea $\epsilon > 0$ y $\ell_1 > \ell^*$ tal que $1/\ell_1 < \epsilon$. Definamos $n_0 = k_{\ell_1}$. Dado $n \geq n_0$, existe $i \geq \ell_1$ tal que $k_i \leq n < k_{i+1}$. Por lo tanto, como $p_n = q_n = i$ se tiene

$$\sup_{u \in \mathcal{S}_{1,p_n}, v \in \mathcal{S}_{2,q_n}} |\rho_{XY}(u, v) - \rho_{n,\omega}(u, v)| < \frac{1}{i} < \epsilon,$$

de donde se concluye que

$$\sup_{u \in \mathcal{S}_{1,p_n}, v \in \mathcal{S}_{2,q_n}} |\rho_{XY}(u, v) - \rho_{n,\omega}(u, v)| \xrightarrow{n \rightarrow \infty} 0.$$

lo que implica la convergencia casi segura **S1F**. □

A continuación probaremos la Proposición 4.5.2 que establece que si la medida de correlación ρ es continua con respecto a la métrica de Prohorov, entonces vale **A3F**. Recordemos que $d_{\text{PR},\mathcal{M}}(Q_1, Q_2)$ indicaba la distancia de Prohorov entre dos probabilidades Q_1 y Q_2 definidas sobre el espacio métrico \mathcal{M} .

Proposición 4.5.2. *Sean \mathcal{H}_j espacios de Hilbert separables, $\mathcal{H} = \mathcal{H}_1 \times \mathcal{H}_2$ y $B_1 = \{\delta_1, \delta_2, \dots\}$ y $B_2 = \{\eta_1, \eta_2, \dots\}$ bases ortonormales de \mathcal{H}_1 y \mathcal{H}_2 , respectivamente. Sea $Z = (X, Y)^T$ un elemento aleatorio en \mathcal{H} tal que $Z \sim P$. Si la medida de asociación ρ es continua respecto de la métrica de Prohorov en la distribución bivariada $P_{u_0 v_0}$, para todo $(u_0, v_0) \in \mathcal{S}_{1,p} \times \mathcal{S}_{2,q}$ con $p \in \mathbb{N}$ y $q \in \mathbb{N}$, entonces ρ cumple la condición **A3F**.*

Demostración. Sean Z_1, \dots, Z_n, \dots elementos aleatorios independientes, $Z_i \sim Z$, y P_n la distribución empírica asociada. El Lema 6.1 de Bali *et al.* (2011) implica que $d_{\text{PR},\mathcal{H}}(P_n, P) \xrightarrow{a.s.} 0$, es decir, $\mathbb{P}(\{\omega \in \Omega : d_{\text{PR},\mathcal{H}}(P_{n,\omega}, P) \rightarrow 0\}) = 1$.

Fijemos $p \in \mathbb{N}$, $q \in \mathbb{N}$, por el Lema 4.4.2 dado $(u_0, v_0) \in \mathcal{S}_{1,p} \times \mathcal{S}_{2,q}$, dado $\epsilon > 0$, existe $\delta = \delta(u_0, v_0) > 0$ tal que si $(u, v) \in \mathcal{S}_{1,p} \times \mathcal{S}_{2,q}$ es tal que

$$\|(u, v) - (u_0, v_0)\|_{\mathcal{H}} < \delta \implies |\rho_{XY}(u, v) - \rho_{XY}(u_0, v_0)| < \epsilon. \quad (4.31)$$

Sea $\omega \in \Omega$ tal que $d_{\text{PR},\mathcal{H}}(P_{n,\omega}, P) \rightarrow 0$ y $(u_0, v_0) \in \mathcal{S}_{1,p} \times \mathcal{S}_{2,q}$. Veamos que dado $\epsilon > 0$, existen $\delta^* = \delta^*(u_0, v_0) > 0$ y $n_0 = n_0(u_0, v_0)$ tales que si $(u, v) \in \mathcal{S}_{1,p} \times \mathcal{S}_{2,q}$ es tal que

$$\|(u, v) - (u_0, v_0)\|_{\mathcal{H}} < \delta^* \text{ y } n \geq n_0 \implies |\rho_{n,\omega}(u, v) - \rho_{XY}(u_0, v_0)| < \epsilon. \quad (4.32)$$

Efectivamente, en la demostración del Lema 4.4.2, hemos visto que por la continuidad de ρ en $P_{u_0 v_0}$, existe $\tau = \tau(u_0, v_0) > 0$ que verifica (4.25). Como $d_{\text{PR},\mathcal{H}}(P_{n,\omega}, P) \rightarrow 0$, existe $n_0 \in \mathbb{N}$ tal que para todo $n \geq n_0$, $d_{\text{PR},\mathcal{H}}(P_{n,\omega}, P) < \tau/2$. Por lo tanto, si $n \geq n_0$, como $\rho_{n,\omega}(u, v) = \rho(P_{n,\omega} A_{uv}^{-1})$, por el Lema 4.4.1 obtenemos que para todo $(u, v) \in \mathcal{S}_{1,p} \times \mathcal{S}_{2,q}$, $\|(u, v) - (u_0, v_0)\|_{\mathcal{H}} < \delta^*$ se cumple que $|\rho_{n,\omega}(u, v) - \rho_{XY}(u_0, v_0)| < \epsilon$.

Dado $\epsilon > 0$ y $\mathbf{w} = (u, v) \in \mathcal{H}_{1,p} \times \mathcal{H}_{2,q}$ definamos $\tilde{\delta}(\mathbf{w}) = \min\{\delta(u, v), \delta^*(u, v)\}$ y $\mathcal{B}_{p,q}(\mathbf{w}, \epsilon) = \{(u_1, v_1) \in \mathcal{H}_{1,p} \times \mathcal{H}_{2,q} : \|(u_1, v_1) - \mathbf{w}\|_{\mathcal{H}} \leq \epsilon\}$.

Observemos que

$$\mathcal{S}_{1,p} \times \mathcal{S}_{2,q} \subset \bigcup_{\mathbf{w}=(u,v) \in \mathcal{S}_{1,p} \times \mathcal{S}_{2,q}} \mathcal{B}_{p,q}(\mathbf{w}, \tilde{\delta}(\mathbf{w})).$$

Luego, como $\mathcal{S}_{1,p} \times \mathcal{S}_{2,q}$ es compacto en $\mathcal{H}_{1,p} \times \mathcal{H}_{2,q}$, existen $\mathbf{w}_j = (u_j, v_j) \in \mathcal{H}_{1,p} \times \mathcal{H}_{2,q}$, $1 \leq j \leq \ell$, tales que $\mathcal{S}_{1,p} \times \mathcal{S}_{2,q} \subset \bigcup_{j=1}^{\ell} \mathcal{V}_j$ donde, por simplicidad, llamamos $\mathcal{V}_j = \mathcal{B}_{p,q}(\mathbf{w}_j, \delta_j)$ y $\delta_j = \tilde{\delta}(\mathbf{w}_j)$.

Sea $N = \max_{1 \leq j \leq \ell} (n_j)$ donde $n_j = n_0(u_j, v_j)$ verifican (4.32). Entonces, si $n \geq N$ y $\mathbf{w} = (u, v) \in \mathcal{S}_{1,p} \times \mathcal{S}_{2,q}$ existe $1 \leq j \leq \ell$, tal que $\mathbf{w} \in \mathcal{V}_j$, por lo tanto, como $\delta_j < \delta(u_j, v_j)$ y $\delta_j < \delta^*(u_j, v_j)$ de (4.31) y (4.32) deducimos que

$$|\rho_n(u, v) - \rho_{XY}(u, v)| \leq |\rho_n(u, v) - \rho_{XY}(u_j, v_j)| + |\rho_{XY}(u_j, v_j) - \rho_{XY}(u, v)| < 2\epsilon.$$

Por lo tanto, hemos probado que fijado $\omega \in \Omega$ tal que $d_{\text{PR}, \mathcal{H}}(P_{n\omega}, P) \rightarrow 0$, se cumple

$$\sup_{(u,v) \in \mathcal{S}_{1,p} \times \mathcal{S}_{2,q}} |\rho_{n,\omega}(u, v) - \rho_{XY}(u, v)| \rightarrow 0,$$

lo que concluye la demostración. □

De las Proposiciones 4.5.1 y 4.5.2 se deduce el siguiente corolario.

Corolario 4.5.1. *Sean \mathcal{H}_j espacios de Hilbert separables, $\mathcal{H} = \mathcal{H}_1 \times \mathcal{H}_2$ y $B_1 = \{\delta_1, \delta_2, \dots\}$ y $B_2 = \{\eta_1, \eta_2, \dots\}$ bases ortonormales de \mathcal{H}_1 y \mathcal{H}_2 , respectivamente. Sea $Z = (X, Y)^T$ un elemento aleatorio en \mathcal{H} tal que $Z \sim P$. Si la medida de asociación ρ es continua respecto de la métrica de Prohorov en la distribución bivariada $P_{u_0 v_0}$, para todo $(u_0, v_0) \in \mathcal{S}_{1,p} \times \mathcal{S}_{2,q}$ con $p \in \mathbb{N}$ y $q \in \mathbb{N}$, entonces existen sucesiones $(p_n)_{n \in \mathbb{N}}$ y $(q_n)_{n \in \mathbb{N}}$ con $p_n \rightarrow \infty$ y $q_n \rightarrow \infty$, $p_n < n$, $q_n < n$ para las que vale **S1F**.*

5. ESTUDIO DE MONTE CARLO

5.1. Modelos de la simulación

En este capítulo, presentamos los resultados de un estudio de Monte Carlo diseñado para comparar el comportamiento de los estimadores de Sieves propuestos en el Capítulo 4.1 para datos generados con y sin contaminación. Las dos características principales que describen a dichos estimadores son la medida de asociación y las bases de los espacios \mathcal{H}_1 y \mathcal{H}_2 que se utilizan. Son también importantes las dimensiones de los subespacios aproximantes p_n y q_n . Los datos sin contaminación se generaron utilizando el mismo modelo que en He *et al.* (2004). En todos los casos, se consideraron muestras de tamaño $n = 100$ y $\mathcal{H}_j = L^2[0, 50]$ y las observaciones se generaron en una grilla discretizada de 50 puntos equiespaciados t_j , $j = 1, \dots, 50$ y los productos internos $\langle X_i, u \rangle_{\mathcal{H}_1}$ y $\langle Y_i, v \rangle_{\mathcal{H}_2}$ se aproximaron mediante sumas de Riemann sobre los puntos de diseño $\{t_j\}_{1 \leq j \leq 50}$.

Para datos no contaminados, las observaciones corresponden a muestras con la misma distribución que el elemento aleatorio $(X, Y)^T \in \mathcal{H}_1 \times \mathcal{H}_2$, dado por

$$X(t) = \sum_{i=1}^m \xi_i f_i(t) \quad \text{e} \quad Y(t) = \sum_{i=1}^m \zeta_i f_i(t) \quad (5.1)$$

donde $\{f_n\}_{n \in \mathbb{N}}$ es la base de Fourier de $L^2[0, 50]$ y $m = 21$. Los escores $\boldsymbol{\xi} = (\xi_1, \dots, \xi_m)^T$ y $\boldsymbol{\zeta} = (\zeta_1, \dots, \zeta_m)^T$ se eligieron con distribución

$$\begin{pmatrix} \boldsymbol{\xi} \\ \boldsymbol{\zeta} \end{pmatrix} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}), \quad \text{con} \quad \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}$$

donde $\Sigma_{22} = \Sigma_{11} = \text{diag}\{10, 10, 10, 10 \times 0,75^{i-3}\}$, $i = 4, \dots, m$, $\Sigma_{12} = \text{diag}(7, 3, 1, 0, \dots, 0)$.

Los valores de las direcciones y correlaciones canónicas poblacionales (4.1) del elemento aleatorio $(X, Y)^T$ se calculan en el Apéndice 5.7 utilizando como medida de asociación el coeficiente de correlación de Pearson ρ_{CL} . Cabe mencionar que como las observaciones corresponden a un proceso Gaussiano, por lo descrito en la Sección 4.2.3, todas las medidas de asociación utilizadas en esta simulación estiman las mismas cantidades que el coeficiente de correlación de Pearson, por ello, se indican como ρ_j . Los valores obtenidos para las correlaciones canónicas son

$$\rho_1 = 0,7, \rho_2 = 0,3, \rho_3 = 0,1 \quad \text{y} \quad \rho_\ell = 0 \text{ si } \ell > 3,$$

mientras que, si $\ell = 1, 2, 3$, las direcciones canónicas son

$$\Phi_\ell(t) = \Psi_\ell(t) = f_\ell(t).$$

Además de muestras correspondientes a datos Gaussianos, se consideraron dos escenarios de contaminación a los que nos referiremos como C_1 y C_2 . En ambos escenarios las muestras contienen un 10% de datos contaminados y se indicarán como $C_{1,0,1}$ y $C_{2,0,1}$, en las tablas, mientras que los datos sin contaminar se indicarán como C_0 , por simplicidad.

En el escenario C_1 las muestras contaminadas se generaron con la misma distribución que el elemento aleatorio (X_{C_1}, Y_{C_1}) donde

$$(X_{C_1}, Y_{C_1})^T = (1 - V)(X, Y)^T + VW(f_2, f_2)^T, \quad (5.2)$$

donde V es una variable aleatoria con distribución Bernoulli de parámetro ε y $W \sim \mathcal{N}(25, 1)$ es independiente de V , X e Y . Se tomó un porcentaje de contaminación $\varepsilon = 0,1$, por lo tanto, el 90% de las veces se observa la muestra original $(X, Y)^T$ mientras que un 10% de las muestras son generadas en la dirección de la segunda dirección canónica de $(X, Y)^T$.

Mientras que en el escenario C_2 las muestras contaminadas se generaron con la misma distribución que el elemento aleatorio (X_{C_2}, Y_{C_2}) donde

$$\begin{aligned} X_{C_2} &= (1 - V)X + V \left(\xi_1 f_1 + W(f_3 + f_4)/\sqrt{2} + \sigma \xi_3 f_3 + \sigma \xi_4 f_4 + \sum_{i=5}^{21} \xi_i f_i \right) \\ Y_{C_2} &= (1 - V)Y + V \left(\zeta_1 f_1 + W(f_3 + f_4)/\sqrt{2} + \sigma \zeta_3 f_3 + \sigma \zeta_4 f_4 + \sum_{i=5}^{21} \zeta_i f_i \right), \end{aligned} \quad (5.3)$$

donde V es una variable aleatoria con distribución Bernoulli de parámetro ε y $W \sim \mathcal{N}(25, 0,01)$ es independiente de V , X e Y y $(\xi^T, \zeta^T)^T \sim \mathcal{N}(0, \Sigma)$. Nuevamente, se presentan los resultados cuando $\varepsilon = 0,1$.

Nos interesa saber como afecta cada contaminación a los valores de la primera correlación canónica y las primeras direcciones canónicas poblacionales cuando se utiliza la correlación de Pearson ρ_{CL} . En ambos escenarios de contaminación, las distintas medidas de asociación dan distintos valores del funcional de correlación y por ello, sólo los calculamos en el caso de la correlación de Pearson, dichos valores poblacionales se indican con el subíndice que indica la contaminación a la que corresponde. Los cálculos se presentan en el Apéndice 5.7 y los valores de las correlaciones y direcciones canónicas bajo los modelos contaminados, están dados por

$$\begin{aligned} \rho_{C_{1,1}} &\approx 0,9, & \Phi_{C_{1,1}}(t) &= \Psi_{C_{1,1}}(t) = f_2(t); \\ \rho_{C_{2,1}} &\approx 0,88, & \Phi_{C_{2,1}}(t) &= \Psi_{C_{2,1}}(t) \approx 0,6f_3 + 0,8f_4(t). \end{aligned} \quad (5.4)$$

La robustez de los estimadores cuando se utiliza una medida de asociación distinta del coeficiente de Pearson, se podrá apreciar en el hecho de que para muestras contaminadas según C_i ($i = 1$ ó 2), el estimador de la primera correlación canónica debe presentar poco sesgo manteniéndose cerca del valor $0,7$ y no a $\rho_{C_{i,1}}$. Por otra parte, los estimadores de las primeras direcciones canónicas en X deberán proveer direcciones cercanas a $\Phi_1 = f_1$ y no a $\Phi_{C_{i,1}}$ y análogamente para los estimadores de las direcciones canónicas en Y .

5.2. Estimadores utilizados y algoritmos para su cálculo

5.2.1. Estimadores considerados

Como mencionamos los estimadores a considerar dependen de la medida de asociación que se utilice y de las bases en las que se proyectan los datos funcionales. Como medidas de asociación consideramos

- la medida de correlación de Pearson, ρ_{CL} ,
- la distancia de correlación, ρ_{DIST} , definida en (2.7),
- la medida de asociación ρ_{OGK} definida en (2.18) usando como estimador de escala un τ -estimador (ver sección 6.9.1 de Maronna *et al.*, 2006),
- la medida de asociación construida a partir de un funcional de dispersión bivariado definida en (2.17). Se utilizó como estimador de dispersión un M -estimador (Maronna, 1976) y, por esta razón, se indicará a la medida de asociación $\rho_{\mathbf{V}_M}$. El M -estimador de dispersión se calculó utilizando como función de peso la asociada a la función de Huber (ver sección 2.2.2 de Maronna *et al.*, 2006) con constante de calibración $\sqrt{\chi_{2,0,9}^2}$, siendo $\chi_{\ell,\alpha}^2$ el cuantil α de una distribución Chi-cuadrado con ℓ grados de libertad, tal como se sugiere en Maronna (1976),
- la medida de correlación mediana, ρ_{CM}^* , definida en (2.10) y
- el coeficiente de correlación de Spearman transformado ρ_{SP} definido en (2.11).

Respecto de las bases consideradas, se utilizaron dos bases fijas y una adaptiva. Más precisamente, consideramos

- la base de Fourier,

- la base de B -Splines cúbicos,
- la base de las direcciones principales.

En el caso de usar la base de las componentes principales se consideraron las autofunciones del operador de covarianza muestral al utilizar la medida de asociación de Pearson ρ_{CL} y la base de las componentes principales esféricas para las demás medidas de asociación de modo a obtener direcciones de rápido cálculo y resistentes a datos atípicos, ver Marden (1999). Las componentes principales esféricas fueron introducidas para datos funcionales por Locantore *et al.* (1999) y estudiadas posteriormente en Gervini (2008).

En todas las Tablas, al reportar las distintas medidas resumen se indica la medida de asociación y la base utilizada.

La dimensión de las bases p_n y q_n se tomaron fijas en los resultados que presentamos en la Sección 5.4.1, mientras que en la Sección 5.4.2 se dan los resultados cuando las dimensiones se eligen mediante el procedimiento de convalidación cruzada que se presenta en la Sección 5.3. Para el caso en que la dimensión es fija, se calcularon estimadores cuando $p_n = q_n = 1, \dots, 11$ para la base de Fourier. Por otra parte, para la base de Splines y de componentes principales se tomaron dimensiones $p_n = q_n = 3, \dots, 11$, ya que en el caso de bases de Splines por haber elegido splines cúbicos la dimensión debe ser mayor o igual a 3 y en el caso de bases de componentes principales, las primeras tres direcciones principales tanto de X como de Y no son identificables. Efectivamente, las direcciones principales de X y de Y coinciden, en este caso, con la base de Fourier y con las direcciones canónicas. Sin embargo, debido a la elección $\Sigma_{22} = \Sigma_{11} = \text{diag}\{10, 10, 10, 10 \times 0,75^{i-3}\}$, $i = 4, \dots, m$, las primeras tres direcciones principales no están adecuadamente identificadas, pudiendo solamente identificar el subespacio principal de dimensión 3, ya que $\text{VAR}(\langle f_i, X \rangle_{\mathcal{H}_1}) = \text{VAR}(\langle f_i, Y \rangle_{\mathcal{H}_2}) = 10$ para $i = 1, 2, 3$. Por esta razón y teniendo en cuenta que las contaminaciones introducidas involucran a f_2 y f_3 , al trabajar con la base de componentes principales se tomó $p \geq 3$.

5.2.2. Algoritmos

Teniendo en cuenta que, al utilizar una medida de asociación distinta del coeficiente de correlación de Pearson, los estimadores dados en (4.2) no pueden ser calculados exactamente surge la necesidad de proveer algoritmos para su cálculo aproximado. En esta sección, describiremos los distintos algoritmos utilizados en este estudio de simulación que permitió además evaluar su performance en cuanto a calidad de aproximación de la solución de (4.2).

Sean $B_1 = \{\delta_1, \delta_2, \dots\}$ y $B_2 = \{\eta_1, \eta_2, \dots\}$ bases ortonormales de \mathcal{H}_1 y \mathcal{H}_2 , respectivamente. Los estimadores propuestos en (4.2) consisten en buscar $u = \sum_{i=1}^{p_n} a_i \delta_i$ y $v = \sum_{i=1}^{q_n} b_i \eta_i$, con $\sum_{i=1}^{p_n} a_i^2 = \sum_{i=1}^{q_n} b_i^2 = 1$, que maximicen $\rho(P_n[\langle u, X \rangle_{\mathcal{H}_1}, \langle v, Y \rangle_{\mathcal{H}_2}])$. Observemos que si definimos $\mathbf{a} = (a_1, \dots, a_{p_n})^T$, $\mathbf{b} = (b_1, \dots, b_{q_n})^T$,

$$\mathbf{x} = (\langle X, \delta_1 \rangle_{\mathcal{H}_1}, \dots, \langle X, \delta_{p_n} \rangle_{\mathcal{H}_1})^T \quad \text{y} \quad \mathbf{y} = (\langle Y, \eta_1 \rangle_{\mathcal{H}_2}, \dots, \langle Y, \eta_{q_n} \rangle_{\mathcal{H}_2})^T$$

se tiene que $\langle u, X \rangle_{\mathcal{H}_1} = \langle \mathbf{a}, \mathbf{x} \rangle_{\mathbb{R}^{p_n}} = \mathbf{a}^T \mathbf{x}$ y $\langle v, Y \rangle_{\mathcal{H}_2} = \langle \mathbf{b}, \mathbf{y} \rangle_{\mathbb{R}^{q_n}} = \mathbf{b}^T \mathbf{y}$.

Los estimadores dados en (4.2) se obtienen encontrando $\widehat{\mathbf{a}}_1 = (\widehat{a}_{11}, \dots, \widehat{a}_{1p_n})^T$ y $\widehat{\mathbf{b}}_1 = (\widehat{b}_{11}, \dots, \widehat{b}_{1q_n})^T$ que maximicen $\rho(Q_n[\mathbf{a}^T \mathbf{x}, \mathbf{b}^T \mathbf{y}])$, donde Q_n es la medida empírica asociada a los datos multivariados $(\mathbf{x}_i, \mathbf{y}_i)$, $1 \leq i \leq n$, con $\mathbf{x}_i = (x_{i1}, \dots, x_{ip_n})^T$ y $\mathbf{y}_i = (y_{i1}, \dots, y_{iq_n})^T$. Por lo tanto, el algoritmo deberá buscar los estimadores, $\widehat{\mathbf{a}}_1$ y $\widehat{\mathbf{b}}_1$, de las primeras direcciones canónicas \mathbf{a}_1 y \mathbf{b}_1 de los datos multivariados $\{(\mathbf{x}_i, \mathbf{y}_i)\}_{1 \leq i \leq n}$ para después reconstruir los elementos de \mathcal{H}_j como $\widehat{\Phi}_1 = \sum_{j=1}^{p_n} \widehat{a}_{1j} \delta_j$ y $\widehat{\Psi}_1 = \sum_{j=1}^{q_n} \widehat{b}_{1j} \eta_j$.

El primer método de estimación, descrito en el Algoritmo 2, está basado en el Algoritmo GRID, introducido en Alfons *et al.* (2016), que aproxima las direcciones que maximizan $\rho(Q_n[\mathbf{a}^T \mathbf{x}, \mathbf{b}^T \mathbf{y}])$ y que está implementado para las medidas de asociación ρ_{CL} , ρ_{VM} y ρ_{SP} en la función “ccaGrid” de la librería “ccaPP” de R. Para ello, la maximización se hace iterativamente en una sucesión de subespacios de dimensión 2. Los detalles se pueden ver en dicho trabajo. Indicaremos a los estimadores obtenidos por $(\widehat{\Phi}_{\text{GRID},1}, \widehat{\Psi}_{\text{GRID},1})$ y $\widehat{\rho}_{\text{GRID},1}$.

Algoritmo 2 Cálculo de direcciones canónicas con un método de Sieves utilizando el Algoritmo GRID de Alfons *et al.* (2016).

- 1: A partir de la muestra $(X_1, Y_1)^T, \dots, (X_n, Y_n)^T$ y las bases B_1 y B_2 construya la muestra $(\mathbf{x}_1, \mathbf{y}_1), \dots, (\mathbf{x}_n, \mathbf{y}_n)$ donde $x_{ij} = \langle X_i, \delta_j \rangle_{\mathcal{H}_1}$ y $y_{ij} = \langle Y_i, \eta_j \rangle_{\mathcal{H}_2}$.
 - 2: Utilizando la función “ccaGrid” de la librería “ccaPP” de R, estime las primeras direcciones canónicas $(\hat{\mathbf{a}}_1, \hat{\mathbf{b}}_1)$ de la muestra $(\mathbf{x}_1, \mathbf{y}_1), \dots, (\mathbf{x}_n, \mathbf{y}_n)$.
 - 3: Construya los estimadores de las direcciones canónicas como $\hat{\Phi}_1 = \sum_{i=1}^{p_n} \hat{a}_{1i} \delta_i$ y $\hat{\Psi}_1 = \sum_{i=1}^{q_n} \hat{b}_{1i} \eta_i$ y el estimador de la correlación canónica como $\hat{\rho}_1 = \rho(Q_n[\hat{\mathbf{a}}_1^T \mathbf{x}, \hat{\mathbf{b}}_1^T \mathbf{y}])$.
-

Siguiendo las ideas utilizadas por Croux y Ruiz-Gazen (1996) para la estimación de la primer componente principal y generalizadas para componentes principales funcionales por Bali y Boente (2014), proponemos el Algoritmo 3 como segundo método para la estimación de las direcciones y correlaciones canónicas.

Algoritmo 3 Cálculo de componentes canónicas con un método de Sieves maximizando en un conjunto de direcciones provenientes de la muestra.

- 1: A partir de la muestra $(X_1, Y_1)^T, \dots, (X_n, Y_n)^T$ y las bases B_1 y B_2 construya la muestra $(\mathbf{x}_1, \mathbf{y}_1), \dots, (\mathbf{x}_n, \mathbf{y}_n)$ donde $x_{ij} = \langle X_i, \delta_j \rangle_{\mathcal{H}_1}$ y $y_{ij} = \langle Y_i, \eta_j \rangle_{\mathcal{H}_2}$.
- 2: Calcule medidas de posición eventualmente robustas multivariadas $\hat{\boldsymbol{\mu}}_{\mathbf{x}}$ y $\hat{\boldsymbol{\mu}}_{\mathbf{y}}$ de la muestra $(\mathbf{x}_1, \mathbf{y}_1), \dots, (\mathbf{x}_n, \mathbf{y}_n)$. Defina las observaciones centradas $\mathbf{x}_i^{(c)} = \mathbf{x}_i - \hat{\boldsymbol{\mu}}_{\mathbf{x}}$ e $\mathbf{y}_i^{(c)} = \mathbf{y}_i - \hat{\boldsymbol{\mu}}_{\mathbf{y}}$.
- 3: Normalice las observaciones, o sea, calcule

$$\boldsymbol{\alpha}_i = \frac{\mathbf{x}_i^{(c)}}{\|\mathbf{x}_i^{(c)}\|_{\mathbb{R}^{p_n}}} \quad \text{y} \quad \boldsymbol{\beta}_i = \frac{\mathbf{y}_i^{(c)}}{\|\mathbf{y}_i^{(c)}\|_{\mathbb{R}^{q_n}}}$$

y considere como conjunto de posibles direcciones $\{(\boldsymbol{\alpha}_i, \boldsymbol{\beta}_j) : 1 \leq i, j \leq n\}$.

- 4: Calcule $(\hat{\mathbf{a}}_1, \hat{\mathbf{b}}_1) = \operatorname{argmax}_{1 \leq i, j \leq n} \rho(Q_n[\boldsymbol{\alpha}_i^T \mathbf{x}, \boldsymbol{\beta}_j^T \mathbf{y}])$.
 - 5: Construya los estimadores de las direcciones canónicas como $\hat{\Phi}_1 = \sum_{i=1}^{p_n} \hat{a}_{1i} \delta_i$ y $\hat{\Psi}_1 = \sum_{i=1}^{q_n} \hat{b}_{1i} \eta_i$ y el estimador de la correlación canónica como $\hat{\rho}_1 = \rho(Q_n[\hat{\mathbf{a}}_1^T \mathbf{x}, \hat{\mathbf{b}}_1^T \mathbf{y}])$.
-

A los estimadores calculados según el Algoritmo 3 los denotaremos por $(\widehat{\Phi}_{\text{CRG},1}, \widehat{\Psi}_{\text{CRG},1})$ y $\widehat{\rho}_{\text{CRG},1}$.

Una elección importante en el Algoritmo 3 es la medida de posición utilizada para centrar los datos. Si utilizamos este algoritmo con el coeficiente de correlación de Pearson, la posición multivariada puede tomarse como la media muestral. Sin embargo, al utilizar medidas de asociación robustas la medida de posición debe ser poco sensible a datos atípicos. Por esta razón, en nuestro estudio de simulación, tomamos como estimador robusto de la posición multivariada de una muestra $\mathbf{x}_1, \dots, \mathbf{x}_n$, $\mathbf{x}_i \in \mathbb{R}^p$ a la mediana espacial o mediana L^1 , $\widehat{\boldsymbol{\mu}}_{\mathbf{x}}$, es decir, $\widehat{\boldsymbol{\mu}}_{\mathbf{x}} = \operatorname{argmin}_{\boldsymbol{\mu} \in \mathbb{R}^p} (1/n) \sum_{i=1}^n \|\mathbf{x}_i - \boldsymbol{\mu}\|_{\mathbb{R}^p}$, donde $\|\cdot\|_{\mathbb{R}^p}$ es la norma usual en \mathbb{R}^p .

Está claro que ambos estimadores $\widehat{\rho}_{\text{GRID},1}$ y $\widehat{\rho}_{\text{CRG},1}$, al maximizar en subconjuntos más pequeños que el considerado en (4.2), cumplen que $\widehat{\rho}_{\text{CRG},1} \leq \widehat{\rho}_1$ y $\widehat{\rho}_{\text{GRID},1} \leq \widehat{\rho}_1$. Entre ambas aproximaciones numéricas del estimador de la primera correlación canónica $\widehat{\rho}_1$ dado en (4.2), será preferible aquel de mayor valor. Debido a esto proponemos un algoritmo cuyo resultado sea la estimación más cercana a $\widehat{\rho}_1$ entre los estimadores presentados previamente. A este nuevo estimador de la primera correlación canónica lo denotamos por $\widehat{\rho}_{\text{COMB},1}$ y a los estimadores de las primeras direcciones canónicas por $(\widehat{\Phi}_{\text{COMB},1}, \widehat{\Psi}_{\text{COMB},1})$ y se obtienen usando el siguiente algoritmo:

Algoritmo 4 Cálculo de componentes canónicas combinando el Algoritmo 2 y el Algoritmo 3

- 1: Calcule los estimadores $\widehat{\rho}_{\text{GRID},1}$, $(\widehat{\Phi}_{\text{GRID},1}, \widehat{\Psi}_{\text{GRID},1})$, $\widehat{\rho}_{\text{CRG},1}$ y $(\widehat{\Phi}_{\text{CRG},1}, \widehat{\Psi}_{\text{CRG},1})$.
 - 2: Si $\widehat{\rho}_{\text{GRID},1} \geq \widehat{\rho}_{\text{CRG},1}$ defina $(\widehat{\Phi}_{\text{COMB},1}, \widehat{\Psi}_{\text{COMB},1}) = (\widehat{\Phi}_{\text{GRID},1}, \widehat{\Psi}_{\text{GRID},1})$ y $\widehat{\rho}_{\text{COMB},1} = \widehat{\rho}_{\text{GRID},1}$.
 - 3: Si $\widehat{\rho}_{\text{GRID},1} < \widehat{\rho}_{\text{CRG},1}$ defina $(\widehat{\Phi}_{\text{COMB},1}, \widehat{\Psi}_{\text{COMB},1}) = (\widehat{\Phi}_{\text{CRG},1}, \widehat{\Psi}_{\text{CRG},1})$ y $\widehat{\rho}_{\text{COMB},1} = \widehat{\rho}_{\text{CRG},1}$.
-

Vale la pena mencionar que si bien los estimadores clásicos, asociados al coeficiente de correlación de Pearson ρ_{CL} pueden calcularse explícitamente, en los estudios numéricos que presentaremos se calcularon utilizando el mismo algoritmo que se usó para computar los estimadores robustos, para hacer comparables los resultados obtenidos. Por otra parte, para

los Algoritmos 2 y 4, sólo utilizaremos las medidas de asociación ρ_{CL} , ρ_{VM} y ρ_{SP} para las cuales está implementado en R el Algoritmo GRID, introducido en Alfons *et al.* (2016).

5.3. Selección de parámetros

Los estimadores de Sieves propuestos dependen de las bases B_1 y B_2 de \mathcal{H}_1 y \mathcal{H}_2 elegidas respectivamente, la medida de asociación ρ a maximizar y de las dimensiones p_n y q_n de las bases sobre las que se proyecta. Una vez elegidos B_1 , B_2 y ρ , deseamos determinar por un procedimiento automático a partir de la muestra observada, las dimensiones de las bases p_n y q_n que utilizaremos para estimar. Para esto usaremos un método de convalidación cruzada adaptado al problema de correlación canónica.

Llamemos $\mathbf{r}_n = (p_n, q_n)$ al parámetro a determinar. Sean $(\widehat{\Phi}_{\mathbf{r}_n,1}^{(i)}, \widehat{\Psi}_{\mathbf{r}_n,1}^{(i)})$ las primeras direcciones canónicas estimadas usando el parámetro \mathbf{r}_n y calculadas con toda la muestra salvo el dato $(X_i, Y_i)^T$

$$(\widehat{\Phi}_{\mathbf{r}_n,1}^{(i)}, \widehat{\Psi}_{\mathbf{r}_n,1}^{(i)}) = \underset{u \in \mathcal{S}_{1,p_n}, v \in \mathcal{S}_{2,q_n}}{\operatorname{argmax}} \rho^2 (P_n^{(i)}[\langle u, X \rangle_{\mathcal{H}_1}, \langle v, Y \rangle_{\mathcal{H}_2}]) ,$$

donde $P_n^{(i)}(A) = (1/n) \sum_{j \neq i} \mathbb{I}_A(X_j, Y_j)$.

Sean $U_{\mathbf{r}_n,1}^{(i)} = \langle \widehat{\Phi}_{\mathbf{r}_n,1}^{(i)}, X_i \rangle_{\mathcal{H}_1}$ y $V_{\mathbf{r}_n,1}^{(i)} = \langle \widehat{\Psi}_{\mathbf{r}_n,1}^{(i)}, Y_i \rangle_{\mathcal{H}_2}$. Como en He *et al.* (2004), definimos

$$RCV_{\mathbf{r}_n,1} = \rho^2 \left(\frac{1}{n} \sum_{i=1}^n \delta_{(U_{\mathbf{r}_n,1}^{(i)}, V_{\mathbf{r}_n,1}^{(i)})} \right) , \quad (5.5)$$

donde indicamos por $\delta_{(a,b)}$ a la medida bivariada de probabilidad que concentra toda su masa en el punto (a, b) . Dado un conjunto de valores posibles \mathcal{R} para el parámetro \mathbf{r}_n , elegimos como parámetro $\widehat{\mathbf{r}}$ al valor

$$\widehat{\mathbf{r}}_{\text{máx}} = \underset{\mathbf{r}_n \in \mathcal{R}}{\operatorname{argmax}} RCV_{\mathbf{r}_n,1} . \quad (5.6)$$

Para que los tiempos de cómputo en la simulación sean razonables, los valores de $\mathbf{r} = (p, q)$ dentro de los cuales maximizamos son tales que $p = q$, con lo cual los posibles valores de \mathbf{r}_n

varían en $\mathcal{R} = \{(p, p) : p \in \mathcal{R}^*\}$. Además, los valores considerados para p son los mismos que usamos en los cálculos en que la dimensión de la base se fijaba de antemano, o sea, para bases de Fourier, $\mathcal{R}^* = \{1, \dots, 11\}$ y para las otras bases, $\mathcal{R}^* = \{3, \dots, 11\}$. El funcional de correlación ρ utilizado en (5.5) es el mismo que se usa para calcular el estimador de las direcciones y correlaciones canónicas. Por lo tanto, este método provee un procedimiento de convalidación cruzada robusta al elegir medidas de asociación robustas.

5.4. Resultados de la simulación

Cuando se utilizaron bases de dimensión fija se realizaron $nr = 1000$ repeticiones para cada una de las situaciones a considerar, mientras que cuando se utilizó convalidación cruzada, se realizaron $nr = 100$ repeticiones debido al tiempo de computo.

Para evaluar la bondad de un estimador $\hat{\rho}_1$ de la correlación canónica ρ_1 , calculamos el promedio sobre las repeticiones de los valores $\hat{\rho}_1$ obtenidos y el error cuadrático medio *ECM*, es decir, el promedio de $(\hat{\rho}_1 - \rho_1)^2$. Por otra parte, para evaluar a un estimador $\hat{\Phi}_1$ de la dirección canónica Φ_1 , se calcularon dos medidas:

- el promedio del valor absoluto del coseno del ángulo $\theta_{\Phi_1, \hat{\Phi}_1}$ entre el estimador $\hat{\Phi}_1$ y la dirección canónica Φ_1 , es decir, el promedio de

$$|\cos(\theta_{\Phi_1, \hat{\Phi}_1})| = \left| \frac{\langle \Phi_1, \hat{\Phi}_1 \rangle_{\mathcal{H}_1}}{\|\Phi_1\|_{\mathcal{H}_1} \|\hat{\Phi}_1\|_{\mathcal{H}_1}} \right|$$

- el promedio sobre las repeticiones de la correlación empírica $\hat{\rho}_{XX}(\Phi_1, \hat{\Phi}_1) = \rho(P_n[\langle \Phi_1, X \rangle, \langle \hat{\Phi}_1, X \rangle])$.

Estas dos medidas se presentan solamente para los estimadores de Φ ya que para los estimadores de Ψ se obtienen resultados semejantes.

Se calculó además una medida global del ajuste dado por las estimaciones de ambas direcciones, que llamaremos AISE como en He *et al.* (2004). Dicha medida es el promedio sobre las nr replicaciones de $\|\widehat{\Phi}_1 - \Phi_1\|^2 + \|\widehat{\Psi}_1 - \Psi_1\|^2$.

5.4.1. Resultados con bases de dimensión fija

En las Tablas 5.2 a 5.20 se presentan los resultados obtenidos cuando la dimensión de la base es fija. Como se mencionó anteriormente, en todos los casos, se tomó $p_n = q_n$. Las Tablas 5.2 a 5.7 presentan los resultados obtenidos con los estimadores $(\widehat{\Phi}_{\text{GRID},1}, \widehat{\Psi}_{\text{GRID},1})$ y $\widehat{\rho}_{\text{GRID},1}$, las Tablas 5.8 a 5.13 corresponden a los estimadores $(\widehat{\Phi}_{\text{CRG},1}, \widehat{\Psi}_{\text{CRG},1})$ y $\widehat{\rho}_{\text{CRG},1}$, mientras que las Tablas 5.15 a 5.20 dan medidas resumen para los estimadores $(\widehat{\Phi}_{\text{COMB},1}, \widehat{\Psi}_{\text{COMB},1})$ y $\widehat{\rho}_{\text{COMB},1}$.

Vale la pena mencionar que si $p = q = 1$ y $(X, Y)^T$ están dados en (5.1), entonces $(\langle X, f_1 \rangle_{\mathcal{H}_1}, \langle Y, f_1 \rangle_{\mathcal{H}_2}) = (\xi_1, \zeta_1)$, mientras que $(\langle X_{C_1}, f_1 \rangle_{\mathcal{H}_1}, \langle Y_{C_1}, f_1 \rangle_{\mathcal{H}_2}) = (1 - V)(\xi_1, \zeta_1)$ y $(\langle X_{C_2}, f_1 \rangle_{\mathcal{H}_1}, \langle Y_{C_2}, f_1 \rangle_{\mathcal{H}_2}) = (\xi_1, \zeta_1) = (\langle X, f_1 \rangle_{\mathcal{H}_1}, \langle Y, f_1 \rangle_{\mathcal{H}_2})$, donde $(X_{C_1}, Y_{C_1})^T$ y $(X_{C_2}, Y_{C_2})^T$ están definidos en (5.2) y (5.3), respectivamente. Luego, para datos con o sin contaminación, si $p = 1$ el estimador de Sieves de las direcciones canónicas será $\widehat{\Phi}_1 = f_1 = \Phi_1$ y $\widehat{\Psi}_1 = f_1 = \Psi_1$. Por esta razón, en las tablas antes mencionadas correspondientes a los estimadores de la primera dirección canónica no se reportan los valores cuando $p = 1$. En la Tabla 5.1 se presentan las medidas resumen para los estimadores de la primera correlación canónica cuando se utiliza la base de Fourier y $p = 1$. Observemos que en este caso, los valores de los estimadores bajo C_0 y C_2 son iguales pues los datos proyectados coinciden y por esa razón se presenta solamente el caso C_0 y $C_{1,0,1}$.

Los valores reportados en las Tablas 5.2 a 5.4 corresponden a la estimación de las primeras direcciones canónicas utilizando el Algoritmo 2 para muestras sin contaminación y para muestras con un 10% de contaminación en los escenarios C_1 y C_2 , respectivamente. Bajo

ρ	C_0		$C_{1,0,1}$	
	Media	$10^4 \times \text{ECM}$	Media	$10^4 \times \text{ECM}$
ρ_{CL}	0.70	26	0.70	28
ρ_{OGK}	0.70	32	0.69	43
$\rho_{\mathbf{V}_M}$	0.70	28	0.70	30
ρ_{DIST}	0.71	27	0.71	28
ρ_{CM}^*	0.69	135	0.63	247
ρ_{SP}	0.70	31	0.68	40

Tab. 5.1: Media y error cuadrático medio de $\hat{\rho}_1$ para distintas medidas de asociación ρ usando la Base de Fourier, cuando $p = 1$.

C_0 y $C_{2,0,1}$, el AISE va creciendo con la dimensión para todas las medidas de asociación consideradas al utilizar las bases de Fourier y de Splines. Por otro lado, bajo $C_{1,0,1}$ se observa un hecho análogo para las bases de Splines y para la Base de Fourier cuando se utilizan como medidas de asociación los funcionales $\rho_{\mathbf{V}_M}$ y ρ_{SP} . Para esta contaminación y para todas las bases, salvo para la base de Fourier cuando $p = 1$ como se mencionó anteriormente, los estimadores basados en el coeficiente de correlación de Pearson ρ_{CL} muestran un notable sesgo. Efectivamente, la media de los valores de $|\cos(\theta_{\Phi_1, \hat{\Phi}_{\text{GRID},1}})|$ y $\hat{\rho}_{XX}(\Phi_1, \hat{\Phi}_{\text{GRID},1})$ se acerca a cero mientras que el AISE es mayor que 3. Notemos que 4 es el valor máximo posible de AISE y se alcanza cuando $\hat{\Phi}_1$ y $\hat{\Psi}_1$ son ortogonales a Φ_1 y Ψ_1 , respectivamente. Bajo $C_{1,0,1}$, los estimadores de la primera dirección canónica basados en funcionales de correlación robustos utilizando la base de Fourier, son más estables ya que los valores reportados son semejantes a los del escenario sin contaminación, aunque presentan un ligero sesgo que es mayor al utilizar el funcional $\rho_{\mathbf{V}_M}$. Algo similar observamos cuando usamos la base de Splines, en las que el estimador $\hat{\Phi}_{\text{GRID},1}$ basado en ρ_{CL} se ve severamente afectado a partir de dimensión $p = 3$, mientras que los estimadores basados en funcionales robustos son mucho más estables frente a esta contaminación mostrando los mejores valores del AISE en

dimensión 3. Por otra parte, bajo $C_{2,0,1}$, al utilizar bases de Fourier, el estimador basado en ρ_{CL} se ve severamente afectado a partir de $p = 4$, que es cuando los subespacios aproximantes contienen a la dirección $\Phi_{C_{2,1}} = \Psi_{C_{2,1}} \approx 0,6f_3 + 0,8f_4$, obteniéndose valores del AISE cercanos o mayores a 3, mientras que los estimadores de la primera dirección canónica basados en funcionales de asociación robustos se ven levemente afectados si $p = 3$ ó 4 produciéndose un desmejoramiento paulatino de las estimaciones al crecer la dimensión.

Bajo C_0 y al utilizar la base de componentes principales el menor valor del AISE, es decir, la mejor dimensión para los subespacios aproximantes, se obtiene cuando $p = 5$ para el estimador clásico y $p = 4$ para las medidas de asociación robustas. Se observa que bajo C_0 para las dimensiones que minimizan el AISE, todos los estimadores presentan un buen desempeño ya que los valores medios de $|\cos(\theta_{\Phi_1, \hat{\Phi}_{GRID,1}})|$ y $\hat{\rho}_{XX}(\Phi_1, \hat{\Phi}_{GRID,1})$ resultan cercanos a 1, mientras que el AISE es próximo a 0. Como es de esperar, para muestras Gaussianas la correlación de Pearson ρ_{CL} da el mejor procedimiento de estimación para todas las bases, ya que minimiza el AISE sobre los demás competidores alcanzando además valores medios de $|\cos(\theta_{\Phi_1, \hat{\Phi}_{GRID,1}})|$ y $\hat{\rho}_{XX}(\Phi_1, \hat{\Phi}_{GRID,1})$ más cercanos a 1. En todas las bases observamos como, a partir de la dimensión en la que se ve la mejor estimación, los resultados comienzan a desmejorar, creciendo paulatinamente el AISE al ir incrementando la dimensión como se mencionó anteriormente. Este crecimiento paulatino del AISE podría deberse a un fenómeno de sobreajuste que relacionaremos con los resultados obtenidos en el Capítulo 3 al analizar el comportamiento de la primera correlación canónica.

Vale la pena remarcar que, bajo C_0 , en cada base, las mejores estimaciones se logran en dimensiones en que los subespacios aproximantes logran estar considerablemente cercanos a f_1 . El desempeño de los estimadores utilizando bases de Splines y bases de Fourier es mejor que el de los estimadores que utilizan las bases de componentes principales, pues los valores de AISE mínimos correspondientes a estos últimos duplican en general los valores obtenidos con las otras dos elecciones de base. Una posible razón es el error que se acumula por la

variabilidad de los estimadores de las componentes principales.

Cuando utilizamos la base de componentes principales, el estimador clásico de las direcciones canónicas muestra su sensibilidad ante la contaminación $C_{1,0,1}$ a partir de $p = 3$, mientras que los estimadores basados en funcionales robustos no se ven mayormente afectados, alcanzando los mejores valores en dimensión 4 y 5 para luego ir desmejorando lentamente al crecer la dimensión. Como mencionamos anteriormente, para la contaminación $C_{1,0,1}$, f_2 es la primera dirección canónica tanto en X como en Y al maximizar ρ_{CL} . Los subespacios aproximantes, al utilizar bases de Fourier, contienen a f_2 a partir de dimensión 2 que es justamente la dimensión a partir de la cual los estimadores basados en ρ_{CL} se ven severamente afectados. Tanto para bases de Splines como para la bases de componentes principales, los estimadores basados en el coeficiente de Pearson muestran sesgo para todas las dimensiones consideradas. Un fenómeno análogo se observa para la contaminación $C_{2,0,1}$, bajo la cual los estimadores clásicos se rompen al utilizar bases de Splines a partir de dimensión 5 y a partir de $p = 3$ para la base de componentes principales. Por otra parte, como se observó bajo $C_{1,0,1}$, bajo $C_{2,0,1}$, los estimadores de la primera dirección canónica basados en funcionales de asociación robustos muestran sus mejores resultados cuando la dimensión es 3 para la base de Splines y 4 para la de componentes principales.

Entre los estimadores basados en funcionales de correlación robustos, el estimador de la primera dirección canónica basado en ρ_{SP} es el que muestra mejor desempeño ya que tiene el menor AISE en todas las bases y dimensiones consideradas, ofreciendo un buen compromiso entre eficiencia y robustez.

Las Tablas 5.5 a 5.7 presentan la media sobre replicaciones y el error cuadrático medio, multiplicado por 10^4 , de los estimadores de la primera correlación canónica cuando se utiliza el Algoritmo 2 para muestras sin contaminación y para muestras con un 10% de contaminación en los escenarios C_1 y C_2 , respectivamente.

Si consideramos que las mejores estimaciones se consiguen cuando el ECM es mínimo, observamos que éstas se dan en dimensión 2, 3 y 4 para bases de Fourier, Splines y componentes principales, respectivamente. Dichas estimaciones presentan valores medios cercanos al valor que se desea estimar, $\rho_1 = 0,7$, y valores pequeños de ECM. Como es de esperar para muestras Gaussianas, la correlación de Pearson ρ_{CL} da el mejor procedimiento de estimación para todas las bases, ya que minimiza el ECM sobre los demás competidores. Sin embargo los valores de ECM conseguidos al utilizar los funcionales robustos son muy similares a los conseguidos utilizando ρ_{CL} . También se observa a partir de la dimensión en que el ECM es mínimo un crecimiento del sesgo de las estimaciones que induce un crecimiento del ECM. Dicho crecimiento podría estar relacionado con el fenómeno descrito en el Capítulo 3. En el Corolario 3.3.1, mostramos que el estimador basado en Sieves $\hat{\rho}_1$ definido en (4.2) es 1 con probabilidad 1 si alguna de las dimensiones de los subespacios aproximantes es mayor o igual que el tamaño de muestra. En nuestra simulación, el tamaño de muestra es 100, por lo tanto, si eligiéramos valores de las dimensiones hasta 100, las medias de $\hat{\rho}_{GRID,1}$ se acercarían a 1 aunque sin llegar necesariamente a 1, ya que el Algoritmo 2 busca en finitas direcciones y es una aproximación a los estimadores definidos en (4.2).

Al contaminar las observaciones, como era de esperarse, el estimador clásico se ve afectado de manera contundente produciendo valores medios superiores a 0,90 bajo $C_{1,0,1}$ para todas las dimensiones consideradas, excepto cuando $p = 1$ y se utiliza la base de Fourier. Por otro lado, bajo $C_{2,0,1}$ el estimador clásico de la correlación canónica presenta valores medios superiores a 0.88 que incrementan su error cuadrático medio, a partir de $p = 4, 5$ y 4 para bases de Fourier, Splines y componentes principales, respectivamente, acorde a lo comentado sobre la Tabla 5.4.

Como para las direcciones canónicas, entre los estimadores de la primera correlación canónica basados en funcionales de correlación robustos y obtenidos por el Algoritmo 2, el estimador basado en ρ_{SP} presenta el mejor desempeño.

Las Tablas 5.8 a 5.10 resumen los resultados obtenidos cuando se utiliza el Algoritmo 3 para estimar la primera dirección canónica y corresponden a muestras sin contaminación y muestras contaminadas bajo $C_{1,0,1}$ y $C_{2,0,1}$, respectivamente. Como en el caso del Algoritmo 2, bajo C_0 al considerar bases de Fourier, el AISE crece lentamente con la dimensión. Las dimensiones en las que se minimiza el AISE para cada base son las mismas que al utilizar el Algoritmo 2 pues dependen básicamente de la capacidad que poseen los subespacios para contener buenos aproximantes de las direcciones canónicas. En todas las bases observamos como el AISE crece paulatinamente luego de la dimensión en que éste se minimiza. Una posible razón de este efecto es que al aumentar la dimensión la esfera unidad se vuelve más difícil de aproximar con una cantidad fija de direcciones, como se explicará más adelante al analizar el comportamiento de los estimadores de la primera correlación canónica. Para la dimensión en que se minimiza el AISE, todos los estimadores tienen un buen desempeño ya que tanto el valor medio del coseno del ángulo como el promedio de las correlaciones $\widehat{\rho}_{XX}(\Phi_1, \widehat{\Phi}_{\text{CRG},1})$ resultan cercanos a 1 mientras que se obtienen valores pequeños de AISE. Los estimadores basados en ρ_{CM}^* son más sesgados que el resto de los estimadores bajo C_0 , aunque provee estimadores aceptables de las direcciones canónicas. Como era de esperar, para datos sin contaminar, el menor valor de AISE en todas las bases se consigue al utilizar el estimador clásico, si bien todas las demás medidas de asociación, salvo la basada en la mediana ρ_{CM}^* , dan valores de AISE similares.

Tanto bajo $C_{1,0,1}$ como bajo $C_{2,0,1}$, el estimador clásico y el que utiliza la distancia de correlación ρ_{DIST} se ven afectados de manera contundente debido a la presencia de observaciones contaminadas. En particular, bajo $C_{2,0,1}$, el mayor cambio para dichos estimadores se observa a partir de las mismas dimensiones que cuando se usó el Algoritmo 2, pues como mencionamos antes los subespacios aproximantes deben estar cercanos a las direcciones canónicas Φ_{C_2} y Ψ_{C_2} . Por otra parte, las dimensiones en las que se minimiza el AISE coinciden con las obtenidas mediante el Algoritmo 2. El resto de los estimadores, basados en

funcionales de asociación robustos sólo se ven levemente afectados y resultan estables para las contaminaciones consideradas. Las dimensiones en las cuales se consiguen los valores de AISE mínimos para los estimadores basados en funcionales robustos son las mismas que cuando se utiliza el Algoritmo 2. Nuevamente, para datos contaminados, el estimador basado en el coeficiente de Spearman ρ_{SP} se destaca mostrando los valores de AISE más pequeños en todas las bases, siendo seguido por el estimador basado en ρ_{OGK} .

Los valores reportados en las Tablas 5.11 a 5.13 corresponden a la estimación de la primera correlación canónica utilizando el Algoritmo 3 para datos sin contaminación y datos contaminados bajo $C_{1,0,1}$ y $C_{2,0,1}$, respectivamente. Las dimensiones en las cuales se minimiza el ECM son las mismas que cuando se utiliza el Algoritmo 2. En dichas dimensiones, bajo C_0 , todos los estimadores, salvo los basadas en ρ_{CM}^* que presentan un mayor sesgo, resultan aceptables pues presentan valores medios cercanos al valor que se desea estimar, $\rho_1 = 0,7$, dando origen a valores pequeños de ECM. Cuando se utilizan bases de Fourier y de Splines, la correlación de Pearson ρ_{CL} resulta ser el mejor procedimiento de estimación, ya que minimiza el ECM sobre los demás competidores, como era de esperar. Sin embargo, los valores de ECM conseguidos al utilizar los funcionales robustos son muy similares a los conseguidos utilizando ρ_{CL} , excepto nuevamente para los estimadores basados en la mediana ρ_{CM}^* . Con el Algoritmo 3, a diferencia de cuando se utiliza el Algoritmo 2, se observa que a partir de la dimensión en que el ECM es mínimo, los valores medios de las estimaciones de ρ_1 decrecen. Como mencionamos anteriormente, a diferencia del Algoritmo 2, el Algoritmo 3 toma una cantidad fija de direcciones, basadas en los propios datos, sobre la cual maximizar la función objetivo. Si bien para el problema de componentes principales, este método da origen a estimadores consistentes, en este caso al aumentar la dimensión la cantidad de direcciones parece ser insuficiente para proveer una buena aproximación del valor máximo dado en (4.2).

Como ya observáramos en las tablas correspondientes a las direcciones canónicas, bajo $C_{1,0,1}$ y $C_{2,0,1}$, los estimadores basados en ρ_{CL} y ρ_{DIST} se ven contundentemente afectados por

la contaminación. En particular, bajo $C_{1,0,1}$, la media sobre replicaciones de los estimadores muestra cercanos a 0,9 a partir de las dimensiones 2, 3 y 7 para bases de Fourier, Splines y componentes principales, respectivamente. Por otra parte, bajo $C_{2,0,1}$, la influencia de la contaminación en el estimador clásico es mayor a partir de $p = 4, 5$ y 3 para bases de Fourier, Splines y componentes principales, respectivamente, tal como ocurría al utilizar el Algoritmo 2. El sesgo de los estimadores basados en funcionales robustos se muestra algo afectado por las contaminaciones $C_{1,0,1}$ y $C_{2,0,1}$ ya que aumentan levemente los valores medios de las estimaciones para las dimensiones en las cuales se minimiza el AISE. La única excepción la constituyen los estimadores basados en la mediana ρ_{CM}^* que muestra estimaciones más sesgadas de ρ_1 . La disminución que se observa en la media de las estimaciones de ρ_1 a partir de las dimensiones en que se minimizaba el AISE en la Tabla 5.10 podría atribuirse al fenómeno descrito previamente sobre la dificultad de aproximar, para dimensiones altas, la esfera unidad con una cantidad fija, independiente de la dimensión, de direcciones.

Cabe mencionar que los valores medios obtenidos para el estimador de la correlación canónica son en general menores a los resultantes al utilizar el Algoritmo 2, lo que hace sospechar que este último provee un mejor método para aproximar los estimadores de las direcciones canónicas. Este hecho se ilustra en la Tabla 5.14 que reporta la cantidad de veces en las 1000 replicaciones en que el valor del estimador de la primera correlación canónica $\hat{\rho}_{CRG,1}$ dado por el Algoritmo 3 es mayor que el valor $\hat{\rho}_{GRID,1}$ obtenido utilizando el Algoritmo 2. Se observa que, a partir de dimensión $p = 4$, prevalece el estimador que utiliza el Algoritmo 2, salvo en el escenario $C_{1,0,1}$ cuando se usan bases de componentes principales y el estimador basado en ρ_{CL} . Combinado este resultado con el hecho que, para dimensiones bajas las medidas resumen obtenidas mediante el Algoritmo 2 son similares a las obtenidas con el Algoritmo 3, resulta razonable que las Tablas 5.15 a 5.20 asociadas a los estimadores provenientes de utilizar el Algoritmo 4, $\hat{\rho}_{COMB,1}$ y $(\hat{\Phi}_{COMB,1}, \hat{\Psi}_{COMB,1})$, muestren valores similares y tengan una interpretación análoga a las Tablas 5.2 a 5.7 que corresponden al

Algoritmo 2.

Vale la pena notar que el Algoritmo 2 da una mejor aproximación que el Algoritmo 3 al valor máximo dado en (4.2) al considerar las medidas de asociación dadas por los coeficientes de Pearson y Spearman o la generada por el funcional de dispersión bivariado definida en (2.17). Sin embargo, el Algoritmo 3 provee resultados razonables para estas medidas y una alternativa simple cuando se consideran otras medidas de asociación. Por esta razón, en las Secciones 5.4.2 y 5.5, al considerar la performance del método de convalidación cruzada para elegir la dimensión del subespacio aproximante y de las medidas de detección de datos atípicos, incluiremos también los resultados obtenidos cuando se calculan los estimadores mediante el Algoritmo 3.

	$\widehat{c}_{\text{GRID}}$	$\widehat{\rho}_{X,\text{GRID}}$	AISE	$\widehat{c}_{\text{GRID}}$	$\widehat{\rho}_{X,\text{GRID}}$	AISE	$\widehat{c}_{\text{GRID}}$	$\widehat{\rho}_{X,\text{GRID}}$	AISE	$\widehat{c}_{\text{GRID}}$	$\widehat{\rho}_{X,\text{GRID}}$	AISE	$\widehat{c}_{\text{GRID}}$	$\widehat{\rho}_{X,\text{GRID}}$	AISE
ρ	$p = 2$			$p = 3$			$p = 4$			$p = 5$			$p = 6$		
	Base de Fourier														
ρ_{CL}	0.99	0.99	0.04	0.99	0.99	0.06	0.98	0.98	0.09	0.97	0.97	0.14	0.95	0.97	0.20
$\rho_{\mathbf{V}_M}$	0.99	0.99	0.04	0.98	0.98	0.08	0.97	0.97	0.12	0.96	0.97	0.18	0.94	0.96	0.25
ρ_{SP}	0.99	0.99	0.05	0.98	0.98	0.09	0.97	0.97	0.13	0.95	0.95	0.21	0.93	0.95	0.28
	Base de Splines														
ρ_{CL}				0.99	0.99	0.06	0.98	0.98	0.10	0.97	0.97	0.14	0.95	0.97	0.21
$\rho_{\mathbf{V}_M}$				0.98	0.98	0.08	0.97	0.97	0.13	0.95	0.96	0.18	0.93	0.96	0.28
ρ_{SP}				0.97	0.97	0.12	0.96	0.96	0.18	0.94	0.95	0.24	0.91	0.94	0.36
	Base de Componentes Principales														
ρ_{CL}				0.91	0.93	0.36	0.96	0.97	0.17	0.96	0.97	0.16	0.95	0.97	0.21
$\rho_{\mathbf{V}_M}$				0.91	0.92	0.38	0.95	0.96	0.19	0.95	0.96	0.20	0.93	0.96	0.27
ρ_{SP}				0.90	0.91	0.41	0.94	0.95	0.23	0.94	0.95	0.25	0.92	0.94	0.34
ρ	$p = 7$			$p = 8$			$p = 9$			$p = 10$			$p = 11$		
	Base de Fourier														
ρ_{CL}	0.93	0.96	0.27	0.91	0.95	0.38	0.88	0.95	0.49	0.84	0.94	0.65	0.79	0.93	0.83
$\rho_{\mathbf{V}_M}$	0.91	0.95	0.34	0.88	0.94	0.48	0.84	0.92	0.64	0.79	0.91	0.82	0.74	0.90	1.05
ρ_{SP}	0.90	0.93	0.41	0.87	0.91	0.54	0.82	0.89	0.72	0.77	0.87	0.91	0.70	0.83	1.18
	Base de Splines														
ρ_{CL}	0.93	0.96	0.28	0.90	0.95	0.40	0.87	0.95	0.50	0.83	0.94	0.68	0.79	0.93	0.85
$\rho_{\mathbf{V}_M}$	0.91	0.95	0.36	0.87	0.94	0.51	0.84	0.92	0.65	0.79	0.91	0.84	0.74	0.90	1.04
ρ_{SP}	0.88	0.92	0.47	0.85	0.91	0.60	0.82	0.90	0.73	0.78	0.90	0.86	0.75	0.88	0.99
	Base de Componentes Principales														
ρ_{CL}	0.93	0.96	0.29	0.90	0.95	0.42	0.86	0.95	0.56	0.81	0.94	0.77	0.75	0.93	0.98
$\rho_{\mathbf{V}_M}$	0.90	0.95	0.38	0.86	0.93	0.55	0.82	0.92	0.73	0.76	0.91	0.96	0.70	0.90	1.19
ρ_{SP}	0.88	0.92	0.47	0.84	0.91	0.62	0.79	0.88	0.82	0.75	0.86	1.01	0.69	0.85	1.22

Tab. 5.2: Medidas resumen correspondientes a la estimación de las primeras direcciones canónicas para distintas medidas de asociación ρ y dimensiones $p = q$ usando el Algoritmo 2. Se presentan el AISE, la media sobre replicaciones del ángulo medido a través del $\widehat{c}_{\text{GRID}} = |\cos(\theta_{\Phi_1, \widehat{\Phi}_{\text{GRID},1}})|$ y del valor $\widehat{\rho}_{X,\text{GRID}} = \widehat{\rho}_{XX}(\Phi_1, \widehat{\Phi}_{\text{GRID},1})$ correspondiente a la medida de asociación usada para datos sin contaminar.

	$\widehat{c}_{\text{GRID}}$	$\widehat{\rho}_{X,\text{GRID}}$	AISE	$\widehat{c}_{\text{GRID}}$	$\widehat{\rho}_{X,\text{GRID}}$	AISE	$\widehat{c}_{\text{GRID}}$	$\widehat{\rho}_{X,\text{GRID}}$	AISE	$\widehat{c}_{\text{GRID}}$	$\widehat{\rho}_{X,\text{GRID}}$	AISE	$\widehat{c}_{\text{GRID}}$	$\widehat{\rho}_{X,\text{GRID}}$	AISE
ρ	$p = 2$			$p = 3$			$p = 4$			$p = 5$			$p = 6$		
	Base de Fourier														
ρ_{CL}	0.18	0.09	3.29	0.20	0.09	3.22	0.20	0.09	3.19	0.21	0.10	3.18	0.21	0.10	3.18
$\rho_{\mathbf{V}_M}$	0.86	0.77	0.55	0.85	0.73	0.61	0.83	0.71	0.69	0.81	0.69	0.77	0.78	0.68	0.87
ρ_{SP}	0.96	0.88	0.15	0.95	0.88	0.22	0.93	0.87	0.29	0.91	0.86	0.38	0.88	0.84	0.50
	Base de Splines														
ρ_{CL}				0.20	0.09	3.18	0.20	0.09	3.19	0.20	0.09	3.19	0.20	0.09	3.20
$\rho_{\mathbf{V}_M}$				0.84	0.71	0.65	0.82	0.70	0.71	0.81	0.68	0.78	0.78	0.67	0.90
ρ_{SP}				0.93	0.81	0.28	0.92	0.80	0.33	0.90	0.79	0.42	0.85	0.76	0.59
	Base de Componentes Principales														
ρ_{CL}				0.18	0.09	3.29	0.20	0.09	3.21	0.20	0.09	3.20	0.20	0.09	3.20
$\rho_{\mathbf{V}_M}$				0.74	0.63	1.04	0.80	0.68	0.79	0.80	0.68	0.81	0.77	0.67	0.91
ρ_{SP}				0.80	0.72	0.79	0.86	0.77	0.53	0.86	0.77	0.57	0.82	0.75	0.71
ρ	$p = 7$			$p = 8$			$p = 9$			$p = 10$			$p = 11$		
	Base de Fourier														
ρ_{CL}	0.21	0.10	3.18	0.20	0.11	3.19	0.20	0.11	3.19	0.19	0.11	3.22	0.19	0.12	3.25
$\rho_{\mathbf{V}_M}$	0.76	0.67	0.98	0.72	0.66	1.10	0.68	0.65	1.27	0.64	0.64	1.43	0.59	0.62	1.65
ρ_{SP}	0.84	0.82	0.66	0.81	0.82	0.76	0.77	0.80	0.94	0.71	0.78	1.15	0.66	0.76	1.36
	Base de Splines														
ρ_{CL}	0.20	0.09	3.22	0.19	0.10	3.24	0.19	0.10	3.24	0.18	0.10	3.28	0.17	0.10	3.31
$\rho_{\mathbf{V}_M}$	0.75	0.66	0.99	0.72	0.64	1.13	0.67	0.62	1.31	0.63	0.61	1.49	0.59	0.59	1.66
ρ_{SP}	0.82	0.75	0.72	0.78	0.73	0.89	0.75	0.73	0.99	0.71	0.71	1.15	0.68	0.71	1.26
	Base de Componentes Principales														
ρ_{CL}	0.20	0.09	3.22	0.19	0.10	3.24	0.19	0.10	3.26	0.17	0.10	3.30	0.16	0.10	3.34
$\rho_{\mathbf{V}_M}$	0.74	0.65	1.03	0.70	0.63	1.21	0.65	0.61	1.41	0.59	0.60	1.63	0.53	0.58	1.86
ρ_{SP}	0.78	0.73	0.87	0.73	0.70	1.09	0.68	0.69	1.27	0.64	0.67	1.46	0.58	0.65	1.67

Tab. 5.3: Medidas resumen correspondientes a la estimación de las primeras direcciones canónicas para distintas medidas de asociación ρ y dimensiones $p = q$ usando el Algoritmo 2 bajo $C_{1,0,1}$. Se presentan el AISE, la media sobre replicaciones del ángulo medido a través del $\widehat{c}_{\text{GRID}} = |\cos(\theta_{\Phi_1, \widehat{\Phi}_{\text{GRID},1}})|$ y del valor $\widehat{\rho}_{X,\text{GRID}} = \widehat{\rho}_{XX}(\Phi_1, \widehat{\Phi}_{\text{GRID},1})$ correspondientes a las medidas de asociación usadas.

	$\widehat{c}_{\text{GRID}}$	$\widehat{\rho}_{X,\text{GRID}}$	AISE	$\widehat{c}_{\text{GRID}}$	$\widehat{\rho}_{X,\text{GRID}}$	AISE	$\widehat{c}_{\text{GRID}}$	$\widehat{\rho}_{X,\text{GRID}}$	AISE	$\widehat{c}_{\text{GRID}}$	$\widehat{\rho}_{X,\text{GRID}}$	AISE	$\widehat{c}_{\text{GRID}}$	$\widehat{\rho}_{X,\text{GRID}}$	AISE
ρ	$p = 2$			$p = 3$			$p = 4$			$p = 5$			$p = 6$		
	Base de Fourier														
ρ_{CL}	0.99	0.99	0.03	0.56	0.44	1.74	0.26	0.17	2.96	0.26	0.17	2.97	0.25	0.17	2.99
$\rho_{\mathbf{V}_M}$	0.99	0.99	0.04	0.88	0.81	0.46	0.82	0.74	0.71	0.81	0.73	0.77	0.79	0.72	0.86
ρ_{SP}	0.99	0.99	0.06	0.96	0.92	0.16	0.92	0.85	0.31	0.89	0.83	0.43	0.86	0.81	0.57
	Base de Splines														
ρ_{CL}				0.56	0.44	1.76	0.55	0.44	1.79	0.26	0.17	2.98	0.25	0.17	2.99
$\rho_{\mathbf{V}_M}$				0.87	0.78	0.52	0.85	0.77	0.61	0.81	0.73	0.77	0.78	0.71	0.88
ρ_{SP}				0.91	0.83	0.36	0.89	0.82	0.42	0.88	0.79	0.48	0.83	0.75	0.67
	Base de Componentes Principales														
ρ_{CL}				0.26	0.17	2.96	0.26	0.17	2.97	0.26	0.17	2.97	0.25	0.17	3
$\rho_{\mathbf{V}_M}$				0.70	0.62	1.20	0.81	0.73	0.76	0.81	0.73	0.77	0.78	0.71	0.87
ρ_{SP}				0.77	0.70	0.93	0.88	0.79	0.47	0.87	0.79	0.51	0.83	0.76	0.67
ρ	$p = 7$			$p = 8$			$p = 9$			$p = 10$			$p = 11$		
	Base de Fourier														
ρ_{CL}	0.25	0.17	3.02	0.24	0.17	3.04	0.23	0.17	3.07	0.22	0.17	3.10	0.22	0.17	3.14
$\rho_{\mathbf{V}_M}$	0.75	0.70	0.99	0.72	0.68	1.14	0.67	0.66	1.31	0.63	0.65	1.49	0.58	0.63	1.69
ρ_{SP}	0.82	0.79	0.71	0.78	0.77	0.89	0.73	0.75	1.08	0.67	0.72	1.31	0.62	0.71	1.52
	Base de Splines														
ρ_{CL}	0.25	0.17	3.02	0.24	0.17	3.05	0.23	0.17	3.07	0.22	0.17	3.11	0.21	0.17	3.15
$\rho_{\mathbf{V}_M}$	0.75	0.70	0.98	0.72	0.69	1.10	0.69	0.67	1.25	0.64	0.66	1.42	0.60	0.65	1.58
ρ_{SP}	0.80	0.74	0.77	0.78	0.74	0.89	0.75	0.73	1	0.72	0.73	1.12	0.68	0.72	1.27
	Base de Componentes Principales														
ρ_{CL}	0.25	0.17	3.02	0.24	0.17	3.05	0.23	0.17	3.09	0.22	0.17	3.14	0.20	0.17	3.19
$\rho_{\mathbf{V}_M}$	0.75	0.70	1	0.71	0.68	1.17	0.66	0.66	1.36	0.61	0.65	1.57	0.55	0.63	1.80
ρ_{SP}	0.78	0.73	0.87	0.72	0.70	1.10	0.66	0.67	1.34	0.61	0.65	1.54	0.55	0.62	1.79

Tab. 5.4: Medidas resumen correspondientes a la estimación de las primeras direcciones canónicas para distintas medidas de asociación ρ y dimensiones $p = q$ usando el Algoritmo 2 bajo $C_{2,0,1}$. Se presentan el AISE, la media sobre replicaciones del ángulo medido a través del $\widehat{c}_{\text{GRID}} = |\cos(\theta_{\Phi_1, \widehat{\Phi}_{\text{GRID},1}})|$ y del valor $\widehat{\rho}_{X,\text{GRID}} = \widehat{\rho}_{XX}(\Phi_1, \widehat{\Phi}_{\text{GRID},1})$ correspondientes a las medidas de asociación usadas.

	Media $10^4 \times \text{ECM}$		Media $10^4 \times \text{ECM}$		Media $10^4 \times \text{ECM}$		Media $10^4 \times \text{ECM}$		Media $10^4 \times \text{ECM}$	
ρ	$p = 2$		$p = 3$		$p = 4$		$p = 5$		$p = 6$	
	Base de Fourier									
ρ_{CL}	0.71	25	0.72	25	0.72	27	0.73	30	0.74	35
ρ_{V_M}	0.71	26	0.72	28	0.73	31	0.74	36	0.75	42
ρ_{SP}	0.71	29	0.73	34	0.74	40	0.75	49	0.76	59
	Base de Splines									
ρ_{CL}			0.72	25	0.72	27	0.73	30	0.74	35
ρ_{V_M}			0.72	28	0.73	30	0.74	36	0.75	43
ρ_{SP}			0.73	34	0.74	39	0.75	48	0.76	55
	Base de Componentes Principales									
ρ_{CL}			0.67	87	0.71	30	0.73	30	0.74	34
ρ_{V_M}			0.67	81	0.72	31	0.74	35	0.75	42
ρ_{SP}			0.68	79	0.73	39	0.75	46	0.76	56
ρ	$p = 7$		$p = 8$		$p = 9$		$p = 10$		$p = 11$	
	Base de Fourier									
ρ_{CL}	0.75	40	0.75	47	0.76	54	0.77	64	0.78	74
ρ_{V_M}	0.76	51	0.77	62	0.77	75	0.78	90	0.79	106
ρ_{SP}	0.76	71	0.77	81	0.78	93	0.78	104	0.78	116
	Base de Splines									
ρ_{CL}	0.75	40	0.75	47	0.76	54	0.77	63	0.78	74
ρ_{V_M}	0.76	51	0.77	62	0.78	74	0.78	88	0.79	105
ρ_{SP}	0.76	67	0.77	75	0.78	84	0.78	93	0.79	102
	Base de Componentes Principales									
ρ_{CL}	0.75	40	0.75	47	0.76	55	0.77	63	0.78	73
ρ_{V_M}	0.76	50	0.77	62	0.78	75	0.78	89	0.79	104
ρ_{SP}	0.76	68	0.77	78	0.77	91	0.78	97	0.78	106

Tab. 5.5: Media y error cuadrático medio, multiplicado por 10^4 , de $\hat{\rho}_{\text{GRID},1}$ para distintas medidas de asociación ρ y dimensiones p usando el Algoritmo 2, para datos sin contaminar.

	Media $10^4 \times \text{ECM}$		Media $10^4 \times \text{ECM}$		Media $10^4 \times \text{ECM}$		Media $10^4 \times \text{ECM}$		Media $10^4 \times \text{ECM}$	
ρ	$p = 2$		$p = 3$		$p = 4$		$p = 5$		$p = 6$	
	Base de Fourier									
ρ_{CL}	0.90	414	0.90	423	0.91	432	0.91	441	0.91	449
ρ_{VM}	0.82	199	0.84	241	0.85	265	0.86	280	0.86	296
ρ_{SP}	0.75	55	0.76	66	0.77	75	0.78	86	0.78	96
	Base de Splines									
ρ_{CL}			0.90	422	0.91	434	0.91	443	0.91	452
ρ_{VM}			0.84	236	0.85	251	0.85	262	0.86	287
ρ_{SP}			0.78	91	0.79	102	0.80	113	0.80	129
	Base de Componentes Principales									
ρ_{CL}			0.90	411	0.90	426	0.91	439	0.91	450
ρ_{VM}			0.82	202	0.85	246	0.85	267	0.86	286
ρ_{SP}			0.73	68	0.77	88	0.79	106	0.80	122
ρ	$p = 7$		$p = 8$		$p = 9$		$p = 10$		$p = 11$	
	Base de Fourier									
ρ_{CL}	0.91	458	0.91	468	0.92	476	0.92	486	0.92	494
ρ_{VM}	0.87	310	0.87	326	0.88	343	0.88	357	0.89	379
ρ_{SP}	0.79	110	0.80	122	0.80	134	0.81	145	0.81	160
	Base de Splines									
ρ_{CL}	0.91	461	0.91	470	0.92	480	0.92	490	0.92	499
ρ_{VM}	0.87	303	0.87	322	0.88	339	0.88	361	0.89	379
ρ_{SP}	0.81	139	0.81	149	0.82	165	0.82	173	0.83	183
	Base de Componentes Principales									
ρ_{CL}	0.91	460	0.91	470	0.92	479	0.92	489	0.92	499
ρ_{VM}	0.87	304	0.87	321	0.88	340	0.88	360	0.89	379
ρ_{SP}	0.80	134	0.81	147	0.81	153	0.82	162	0.82	175

Tab. 5.6: Media y error cuadrático medio, multiplicado por 10^4 , de $\hat{\rho}_{\text{GRID},1}$ para distintas medidas de asociación ρ y dimensiones p usando el Algoritmo 2, bajo $C_{1,0,1}$.

	Media $10^4 \times \text{ECM}$		Media $10^4 \times \text{ECM}$		Media $10^4 \times \text{ECM}$		Media $10^4 \times \text{ECM}$		Media $10^4 \times \text{ECM}$	
ρ	$p = 2$		$p = 3$		$p = 4$		$p = 5$		$p = 6$	
	Base de Fourier									
ρ_{CL}	0.71	25	0.79	98	0.88	352	0.89	361	0.89	371
ρ_{VM}	0.71	26	0.76	56	0.81	141	0.81	157	0.82	169
ρ_{SP}	0.71	29	0.74	44	0.77	72	0.77	81	0.78	94
	Base de Splines									
ρ_{CL}			0.79	98	0.79	107	0.89	357	0.89	371
ρ_{VM}			0.75	52	0.76	61	0.81	146	0.82	162
ρ_{SP}			0.74	47	0.75	56	0.78	84	0.78	94
	Base de Componentes Principales									
ρ_{CL}			0.86	290	0.88	333	0.88	353	0.89	368
ρ_{VM}			0.75	102	0.79	124	0.81	146	0.82	164
ρ_{SP}			0.69	74	0.76	64	0.77	79	0.78	94
ρ	$p = 7$		$p = 8$		$p = 9$		$p = 10$		$p = 11$	
	Base de Fourier									
ρ_{CL}	0.89	381	0.90	391	0.90	402	0.90	412	0.90	422
ρ_{VM}	0.83	186	0.83	202	0.84	219	0.85	236	0.85	252
ρ_{SP}	0.79	104	0.79	113	0.80	122	0.80	132	0.81	143
	Base de Splines									
ρ_{CL}	0.89	382	0.90	392	0.90	402	0.90	412	0.90	423
ρ_{VM}	0.82	175	0.83	189	0.84	207	0.84	223	0.85	245
ρ_{SP}	0.79	102	0.80	116	0.80	127	0.81	136	0.81	146
	Base de Componentes Principales									
ρ_{CL}	0.89	380	0.90	391	0.90	402	0.90	412	0.90	422
ρ_{VM}	0.82	177	0.83	195	0.84	211	0.84	227	0.85	243
ρ_{SP}	0.78	105	0.79	114	0.79	123	0.80	128	0.80	137

Tab. 5.7: Media y error cuadrático medio, multiplicado por 10^4 , de $\hat{\rho}_{\text{GRID},1}$ para distintas medidas de asociación ρ y dimensiones p usando el Algoritmo 2, bajo $C_{2,0,1}$.

	\widehat{c}_{CRG}	$\widehat{\rho}_{X,\text{CRG}}$	AISE	\widehat{c}_{CRG}	$\widehat{\rho}_{X,\text{CRG}}$	AISE	\widehat{c}_{CRG}	$\widehat{\rho}_{X,\text{CRG}}$	AISE	\widehat{c}_{CRG}	$\widehat{\rho}_{X,\text{CRG}}$	AISE	\widehat{c}_{CRG}	$\widehat{\rho}_{X,\text{CRG}}$	AISE
ρ	$p = 2$			$p = 3$			$p = 4$			$p = 5$			$p = 6$		
	Base de Fourier														
ρ_{CL}	0.99	0.99	0.04	0.97	0.97	0.10	0.94	0.95	0.24	0.91	0.92	0.38	0.87	0.90	0.51
ρ_{OGK}	0.98	0.98	0.08	0.96	0.96	0.17	0.92	0.93	0.32	0.88	0.90	0.48	0.85	0.88	0.61
$\rho_{\mathbf{V}_M}$	0.99	0.99	0.05	0.97	0.97	0.12	0.93	0.94	0.26	0.90	0.92	0.41	0.87	0.90	0.54
ρ_{DIST}	0.99	0.99	0.05	0.97	0.97	0.12	0.93	0.94	0.27	0.90	0.92	0.42	0.87	0.90	0.54
ρ_{CM}^*	0.93	0.95	0.26	0.90	0.91	0.43	0.84	0.87	0.65	0.80	0.85	0.81	0.77	0.83	0.94
ρ_{SP}	0.98	0.98	0.06	0.97	0.97	0.13	0.93	0.93	0.28	0.90	0.91	0.42	0.86	0.89	0.55
	Base de Splines														
ρ_{CL}				0.97	0.97	0.10	0.94	0.95	0.23	0.91	0.92	0.38	0.88	0.91	0.50
ρ_{OGK}				0.96	0.96	0.17	0.92	0.93	0.32	0.88	0.90	0.47	0.85	0.89	0.59
$\rho_{\mathbf{V}_M}$				0.97	0.97	0.12	0.94	0.95	0.25	0.90	0.92	0.40	0.87	0.90	0.52
ρ_{DIST}				0.97	0.97	0.13	0.94	0.95	0.25	0.90	0.92	0.40	0.87	0.91	0.53
ρ_{CM}^*				0.89	0.91	0.44	0.84	0.88	0.63	0.79	0.84	0.82	0.77	0.83	0.94
ρ_{SP}				0.97	0.96	0.13	0.94	0.94	0.26	0.90	0.91	0.41	0.86	0.90	0.54
	Base de Componentes Principales														
ρ_{CL}				0.91	0.92	0.37	0.92	0.94	0.30	0.90	0.92	0.40	0.87	0.91	0.51
ρ_{OGK}				0.88	0.90	0.48	0.90	0.92	0.39	0.88	0.90	0.49	0.86	0.89	0.59
$\rho_{\mathbf{V}_M}$				0.90	0.92	0.41	0.92	0.93	0.32	0.89	0.91	0.43	0.87	0.90	0.53
ρ_{DIST}				0.89	0.91	0.43	0.92	0.93	0.33	0.89	0.92	0.43	0.87	0.90	0.54
ρ_{CM}^*				0.82	0.86	0.73	0.82	0.86	0.73	0.79	0.85	0.85	0.76	0.83	0.97
ρ_{SP}				0.89	0.91	0.43	0.92	0.93	0.34	0.89	0.91	0.44	0.86	0.89	0.55
ρ	$p = 7$			$p = 8$			$p = 9$			$p = 10$			$p = 11$		
	Base de Fourier														
ρ_{CL}	0.85	0.89	0.61	0.83	0.89	0.67	0.82	0.88	0.74	0.80	0.88	0.78	0.80	0.88	0.80
ρ_{OGK}	0.83	0.87	0.71	0.80	0.86	0.78	0.79	0.86	0.85	0.78	0.85	0.89	0.77	0.85	0.91
$\rho_{\mathbf{V}_M}$	0.84	0.89	0.63	0.82	0.88	0.70	0.81	0.88	0.76	0.80	0.87	0.80	0.79	0.87	0.83
ρ_{DIST}	0.84	0.89	0.64	0.82	0.88	0.71	0.81	0.88	0.76	0.80	0.88	0.80	0.79	0.88	0.83
ρ_{CM}^*	0.74	0.82	1.05	0.72	0.82	1.15	0.70	0.81	1.19	0.70	0.81	1.22	0.69	0.81	1.24
ρ_{SP}	0.84	0.88	0.65	0.82	0.88	0.72	0.80	0.87	0.78	0.80	0.87	0.81	0.79	0.87	0.84
	Base de Splines														
ρ_{CL}	0.85	0.89	0.60	0.83	0.89	0.66	0.82	0.88	0.73	0.81	0.88	0.77	0.80	0.88	0.80
ρ_{OGK}	0.82	0.87	0.70	0.81	0.87	0.77	0.79	0.86	0.84	0.78	0.86	0.88	0.77	0.85	0.92
$\rho_{\mathbf{V}_M}$	0.84	0.89	0.63	0.83	0.89	0.69	0.81	0.88	0.76	0.80	0.87	0.79	0.79	0.87	0.83
ρ_{DIST}	0.84	0.89	0.64	0.83	0.89	0.69	0.81	0.88	0.76	0.80	0.88	0.80	0.79	0.88	0.83
ρ_{CM}^*	0.74	0.82	1.07	0.73	0.82	1.10	0.71	0.82	1.18	0.70	0.81	1.21	0.69	0.81	1.25
ρ_{SP}	0.84	0.88	0.64	0.82	0.88	0.72	0.81	0.87	0.77	0.80	0.87	0.82	0.79	0.87	0.84
	Base de Componentes Principales														
ρ_{CL}	0.85	0.90	0.59	0.84	0.89	0.65	0.82	0.88	0.71	0.81	0.88	0.75	0.81	0.88	0.77
ρ_{OGK}	0.83	0.87	0.70	0.81	0.87	0.77	0.80	0.86	0.82	0.79	0.86	0.86	0.78	0.86	0.89
$\rho_{\mathbf{V}_M}$	0.85	0.89	0.62	0.83	0.89	0.68	0.82	0.88	0.73	0.81	0.88	0.77	0.80	0.87	0.81
ρ_{DIST}	0.84	0.89	0.63	0.83	0.89	0.68	0.82	0.89	0.73	0.81	0.88	0.78	0.80	0.88	0.81
ρ_{CM}^*	0.74	0.83	1.04	0.73	0.82	1.11	0.71	0.82	1.16	0.70	0.81	1.21	0.69	0.81	1.25
ρ_{SP}	0.84	0.89	0.62	0.83	0.88	0.69	0.81	0.88	0.75	0.80	0.87	0.79	0.79	0.87	0.83

Tab. 5.8: Medidas resumen correspondientes a la estimación de las primeras direcciones canónicas para distintas medidas de asociación ρ y dimensiones $p = q$ usando el Algoritmo 3. Se presentan el AISE, la media sobre replicaciones del ángulo medido a través del $\widehat{c}_{\text{CRG}} = |\cos(\theta_{\Phi_1, \widehat{\Phi}_{\text{CRG},1}})|$ y del valor $\widehat{\rho}_{X,\text{CRG}} = \widehat{\rho}_{XX}(\Phi_1, \widehat{\Phi}_{\text{CRG},1})$ correspondiente a la medida de asociación usada para datos sin contaminar.

	\hat{c}_{CRG}	$\hat{\rho}_{X,\text{CRG}}$	AISE	\hat{c}_{CRG}	$\hat{\rho}_{X,\text{CRG}}$	AISE	\hat{c}_{CRG}	$\hat{\rho}_{X,\text{CRG}}$	AISE	\hat{c}_{CRG}	$\hat{\rho}_{X,\text{CRG}}$	AISE	\hat{c}_{CRG}	$\hat{\rho}_{X,\text{CRG}}$	AISE
ρ	$p = 2$			$p = 3$			$p = 4$			$p = 5$			$p = 6$		
	Base de Fourier														
ρ_{CL}	0.20	0.09	3.19	0.18	0.09	3.28	0.14	0.08	3.45	0.09	0.06	3.62	0.07	0.06	3.70
ρ_{OGK}	0.94	0.89	0.25	0.90	0.84	0.42	0.83	0.77	0.68	0.78	0.73	0.89	0.74	0.70	1.03
$\rho_{\mathbf{V}_M}$	0.85	0.73	0.60	0.82	0.69	0.74	0.76	0.64	0.95	0.71	0.59	1.16	0.68	0.57	1.29
ρ_{DIST}	0.14	0.33	3.44	0.12	0.33	3.54	0.08	0.32	3.70	0.05	0.32	3.79	0.04	0.32	3.83
ρ_{CM}^*	0.94	0.85	0.21	0.88	0.79	0.47	0.82	0.74	0.73	0.76	0.71	0.95	0.73	0.68	1.08
ρ_{SP}	0.96	0.83	0.18	0.93	0.80	0.30	0.88	0.77	0.48	0.83	0.74	0.67	0.80	0.72	0.81
	Base de Splines														
ρ_{CL}				0.18	0.09	3.28	0.14	0.08	3.44	0.09	0.07	3.62	0.08	0.06	3.68
ρ_{OGK}				0.89	0.84	0.44	0.85	0.79	0.63	0.77	0.73	0.90	0.75	0.72	0.99
$\rho_{\mathbf{V}_M}$				0.82	0.69	0.74	0.77	0.65	0.92	0.71	0.60	1.15	0.68	0.58	1.26
ρ_{DIST}				0.12	0.33	3.54	0.09	0.33	3.65	0.05	0.32	3.79	0.04	0.32	3.82
ρ_{CM}^*				0.88	0.79	0.47	0.83	0.76	0.71	0.76	0.70	0.95	0.73	0.69	1.09
ρ_{SP}				0.93	0.81	0.30	0.88	0.78	0.48	0.84	0.74	0.66	0.80	0.73	0.79
	Base de Componentes Principales														
ρ_{CL}				0.38	0.45	2.47	0.47	0.51	2.10	0.29	0.33	2.84	0.16	0.21	3.36
ρ_{OGK}				0.77	0.80	0.91	0.84	0.86	0.63	0.84	0.86	0.63	0.82	0.85	0.73
$\rho_{\mathbf{V}_M}$				0.77	0.80	0.91	0.79	0.81	0.85	0.75	0.78	0.99	0.72	0.75	1.12
ρ_{DIST}				0.75	0.82	1.00	0.66	0.76	1.34	0.45	0.61	2.19	0.22	0.45	3.13
ρ_{CM}^*				0.72	0.73	1.12	0.73	0.73	1.09	0.69	0.69	1.23	0.66	0.67	1.39
ρ_{SP}				0.79	0.82	0.82	0.86	0.87	0.58	0.86	0.88	0.56	0.84	0.86	0.64
ρ	$p = 7$			$p = 8$			$p = 9$			$p = 10$			$p = 11$		
	Base de Fourier														
ρ_{CL}	0.06	0.06	3.74	0.06	0.06	3.78	0.05	0.05	3.80	0.05	0.05	3.81	0.05	0.05	3.81
ρ_{OGK}	0.71	0.69	1.15	0.69	0.67	1.23	0.68	0.67	1.29	0.66	0.66	1.35	0.66	0.66	1.35
$\rho_{\mathbf{V}_M}$	0.65	0.55	1.41	0.63	0.54	1.48	0.62	0.54	1.52	0.61	0.54	1.55	0.60	0.53	1.59
ρ_{DIST}	0.03	0.32	3.86	0.03	0.32	3.87	0.03	0.32	3.88	0.03	0.32	3.88	0.03	0.32	3.88
ρ_{CM}^*	0.69	0.67	1.22	0.68	0.66	1.28	0.67	0.65	1.32	0.66	0.65	1.38	0.65	0.65	1.40
ρ_{SP}	0.77	0.71	0.93	0.74	0.70	1.01	0.73	0.69	1.06	0.72	0.69	1.10	0.72	0.69	1.13
	Base de Splines														
ρ_{CL}	0.06	0.06	3.74	0.06	0.05	3.77	0.05	0.05	3.79	0.05	0.05	3.80	0.05	0.05	3.80
ρ_{OGK}	0.72	0.70	1.13	0.70	0.69	1.19	0.68	0.67	1.28	0.67	0.67	1.29	0.66	0.66	1.36
$\rho_{\mathbf{V}_M}$	0.66	0.56	1.37	0.63	0.55	1.45	0.62	0.54	1.52	0.61	0.54	1.54	0.60	0.53	1.58
ρ_{DIST}	0.03	0.32	3.86	0.03	0.32	3.86	0.03	0.32	3.88	0.03	0.32	3.88	0.03	0.32	3.88
ρ_{CM}^*	0.70	0.67	1.21	0.68	0.66	1.27	0.67	0.66	1.30	0.65	0.65	1.36	0.65	0.65	1.38
ρ_{SP}	0.77	0.71	0.91	0.75	0.70	0.98	0.73	0.70	1.05	0.72	0.69	1.10	0.72	0.69	1.13
	Base de Componentes Principales														
ρ_{CL}	0.06	0.11	3.76	0.05	0.10	3.81	0.04	0.09	3.84	0.04	0.09	3.84	0.04	0.09	3.84
ρ_{OGK}	0.78	0.83	0.86	0.76	0.82	0.93	0.75	0.82	0.99	0.73	0.81	1.06	0.73	0.81	1.09
$\rho_{\mathbf{V}_M}$	0.69	0.73	1.26	0.67	0.72	1.34	0.63	0.70	1.46	0.62	0.70	1.50	0.62	0.70	1.53
ρ_{DIST}	0.11	0.38	3.55	0.06	0.35	3.75	0.05	0.35	3.82	0.04	0.35	3.83	0.04	0.35	3.84
ρ_{CM}^*	0.62	0.65	1.52	0.59	0.63	1.64	0.56	0.62	1.74	0.56	0.62	1.76	0.54	0.61	1.83
ρ_{SP}	0.82	0.85	0.72	0.80	0.84	0.80	0.78	0.83	0.88	0.77	0.83	0.92	0.76	0.83	0.96

Tab. 5.9: Medidas resumen correspondientes a la estimación de las primeras direcciones canónicas para distintas medidas de asociación ρ y dimensiones $p = q$ usando el Algoritmo 3 bajo $C_{1,0,1}$. Se presentan el AISE, la media sobre replicaciones del ángulo medido a través del $\hat{c}_{\text{CRG}} = |\cos(\theta_{\Phi_1, \hat{\Phi}_{\text{CRG},1}})|$ y del valor $\hat{\rho}_{X,\text{CRG}} = \hat{\rho}_{XX}(\Phi_1, \hat{\Phi}_{\text{CRG},1})$ correspondientes a las medidas de asociación usadas.

	\hat{c}_{CRG}	$\hat{\rho}_{X,\text{CRG}}$	AISE	\hat{c}_{CRG}	$\hat{\rho}_{X,\text{CRG}}$	AISE	\hat{c}_{CRG}	$\hat{\rho}_{X,\text{CRG}}$	AISE	\hat{c}_{CRG}	$\hat{\rho}_{X,\text{CRG}}$	AISE	\hat{c}_{CRG}	$\hat{\rho}_{X,\text{CRG}}$	AISE
ρ	$p = 2$			$p = 3$			$p = 4$			$p = 5$			$p = 6$		
	Base de Fourier														
ρ_{CL}	0.99	0.99	0.04	0.51	0.40	1.95	0.17	0.13	3.31	0.14	0.12	3.42	0.13	0.11	3.47
ρ_{OGK}	0.98	0.98	0.08	0.91	0.87	0.38	0.86	0.83	0.55	0.82	0.80	0.72	0.79	0.78	0.86
$\rho_{\mathbf{V}_M}$	0.99	0.99	0.05	0.86	0.78	0.57	0.78	0.69	0.88	0.75	0.66	1.03	0.71	0.64	1.17
ρ_{DIST}	0.99	0.99	0.05	0.31	0.36	2.74	0.11	0.20	3.56	0.10	0.20	3.59	0.10	0.20	3.59
ρ_{CM}^*	0.93	0.94	0.29	0.85	0.83	0.60	0.79	0.77	0.82	0.75	0.74	1.02	0.70	0.71	1.19
ρ_{SP}	0.98	0.98	0.07	0.92	0.85	0.31	0.88	0.78	0.49	0.84	0.76	0.65	0.80	0.74	0.80
	Base de Splines														
ρ_{CL}				0.50	0.39	1.98	0.42	0.34	2.33	0.14	0.12	3.42	0.13	0.12	3.45
ρ_{OGK}				0.91	0.87	0.36	0.86	0.84	0.55	0.82	0.80	0.73	0.79	0.78	0.86
$\rho_{\mathbf{V}_M}$				0.85	0.77	0.59	0.79	0.72	0.82	0.75	0.67	1.03	0.71	0.64	1.17
ρ_{DIST}				0.31	0.36	2.76	0.27	0.33	2.90	0.10	0.20	3.58	0.10	0.20	3.60
ρ_{CM}^*				0.85	0.82	0.61	0.80	0.79	0.81	0.74	0.74	1.04	0.71	0.73	1.16
ρ_{SP}				0.92	0.85	0.33	0.87	0.81	0.51	0.82	0.74	0.73	0.79	0.72	0.85
	Base de Componentes Principales														
ρ_{CL}				0.20	0.14	3.20	0.17	0.13	3.31	0.14	0.12	3.41	0.13	0.12	3.45
ρ_{OGK}				0.74	0.73	1	0.84	0.81	0.65	0.82	0.81	0.72	0.79	0.78	0.86
$\rho_{\mathbf{V}_M}$				0.69	0.62	1.23	0.76	0.68	0.94	0.74	0.66	1.05	0.71	0.64	1.19
ρ_{DIST}				0.14	0.22	3.41	0.11	0.20	3.56	0.10	0.20	3.58	0.10	0.20	3.59
ρ_{CM}^*				0.70	0.71	1.17	0.77	0.76	0.90	0.74	0.74	1.05	0.71	0.72	1.16
ρ_{SP}				0.76	0.70	0.95	0.86	0.77	0.58	0.83	0.76	0.67	0.80	0.74	0.80
ρ	$p = 7$			$p = 8$			$p = 9$			$p = 10$			$p = 11$		
	Base de Fourier														
ρ_{CL}	0.13	0.12	3.48	0.12	0.11	3.49	0.12	0.11	3.51	0.12	0.11	3.51	0.12	0.11	3.51
ρ_{OGK}	0.75	0.76	1	0.73	0.74	1.08	0.72	0.74	1.14	0.70	0.73	1.19	0.70	0.73	1.20
$\rho_{\mathbf{V}_M}$	0.68	0.62	1.29	0.66	0.61	1.37	0.65	0.60	1.43	0.64	0.60	1.46	0.63	0.60	1.49
ρ_{DIST}	0.10	0.20	3.60	0.10	0.20	3.60	0.10	0.20	3.60	0.10	0.20	3.61	0.10	0.20	3.60
ρ_{CM}^*	0.68	0.71	1.29	0.66	0.70	1.37	0.64	0.69	1.45	0.64	0.69	1.46	0.63	0.69	1.49
ρ_{SP}	0.77	0.72	0.92	0.76	0.72	0.99	0.74	0.71	1.05	0.73	0.71	1.08	0.72	0.71	1.12
	Base de Splines														
ρ_{CL}	0.13	0.11	3.48	0.13	0.11	3.49	0.12	0.11	3.50	0.12	0.11	3.51	0.12	0.11	3.51
ρ_{OGK}	0.75	0.76	0.99	0.74	0.75	1.07	0.72	0.74	1.14	0.71	0.74	1.16	0.70	0.73	1.22
$\rho_{\mathbf{V}_M}$	0.68	0.62	1.29	0.66	0.61	1.36	0.65	0.61	1.43	0.64	0.60	1.47	0.63	0.60	1.49
ρ_{DIST}	0.10	0.20	3.60	0.10	0.20	3.60	0.10	0.20	3.60	0.10	0.20	3.60	0.10	0.20	3.61
ρ_{CM}^*	0.68	0.71	1.28	0.67	0.70	1.35	0.64	0.69	1.43	0.64	0.70	1.44	0.63	0.69	1.47
ρ_{SP}	0.76	0.71	0.98	0.74	0.70	1.05	0.72	0.70	1.10	0.72	0.69	1.14	0.71	0.69	1.16
	Base de Componentes Principales														
ρ_{CL}	0.13	0.12	3.46	0.12	0.11	3.49	0.12	0.11	3.50	0.12	0.11	3.50	0.12	0.11	3.51
ρ_{OGK}	0.76	0.76	0.98	0.74	0.75	1.04	0.72	0.74	1.11	0.71	0.74	1.15	0.71	0.73	1.19
$\rho_{\mathbf{V}_M}$	0.68	0.63	1.28	0.66	0.61	1.36	0.65	0.61	1.40	0.65	0.61	1.44	0.64	0.61	1.46
ρ_{DIST}	0.10	0.20	3.60	0.10	0.20	3.60	0.10	0.20	3.60	0.10	0.20	3.61	0.10	0.20	3.60
ρ_{CM}^*	0.68	0.70	1.29	0.66	0.71	1.35	0.65	0.70	1.40	0.65	0.70	1.43	0.64	0.69	1.45
ρ_{SP}	0.78	0.73	0.89	0.76	0.72	0.96	0.75	0.72	1.01	0.74	0.71	1.06	0.73	0.71	1.08

Tab. 5.10: Medidas resumen correspondientes a la estimación de las primeras direcciones canónicas para distintas medidas de asociación ρ y dimensiones $p = q$ usando el Algoritmo 3 bajo $C_{2,0,1}$. Se presentan el AISE, la media sobre replicaciones del ángulo medido a través del $\hat{c}_{\text{CRG}} = |\cos(\theta_{\Phi_1, \hat{\Phi}_{\text{CRG},1}})|$ y del valor $\hat{\rho}_{X,\text{CRG}} = \hat{\rho}_{XX}(\Phi_1, \hat{\Phi}_{\text{CRG},1})$ correspondientes a las medidas de asociación usadas.

	Media $10^4 \times \text{ECM}$		Media $10^4 \times \text{ECM}$		Media $10^4 \times \text{ECM}$		Media $10^4 \times \text{ECM}$		Media $10^4 \times \text{ECM}$	
ρ	$p = 2$		$p = 3$		$p = 4$		$p = 5$		$p = 6$	
	Base de Fourier									
ρ_{CL}	0.71	25	0.70	24	0.68	29	0.67	40	0.66	51
ρ_{OGK}	0.72	32	0.71	29	0.70	28	0.68	33	0.67	39
$\rho_{\mathbf{V}_M}$	0.71	26	0.71	25	0.69	29	0.67	39	0.66	48
ρ_{DIST}	0.72	28	0.72	27	0.70	25	0.68	30	0.67	35
ρ_{CM}^*	0.85	289	0.85	287	0.83	214	0.81	172	0.80	151
ρ_{SP}	0.71	29	0.71	28	0.69	31	0.67	40	0.66	47
	Base de Splines									
ρ_{CL}			0.70	24	0.69	27	0.67	39	0.66	46
ρ_{OGK}			0.71	29	0.70	27	0.68	34	0.67	37
$\rho_{\mathbf{V}_M}$			0.71	25	0.69	27	0.67	38	0.66	44
ρ_{DIST}			0.72	27	0.70	25	0.68	31	0.68	33
ρ_{CM}^*			0.85	287	0.84	244	0.81	172	0.81	164
ρ_{SP}			0.70	28	0.69	29	0.67	40	0.66	45
	Base de Componentes Principales									
ρ_{CL}			0.66	86	0.68	35	0.67	41	0.66	49
ρ_{OGK}			0.67	71	0.69	32	0.68	33	0.67	38
$\rho_{\mathbf{V}_M}$			0.66	88	0.68	35	0.67	39	0.66	46
ρ_{DIST}			0.67	75	0.69	30	0.68	30	0.68	34
ρ_{CM}^*			0.81	215	0.82	202	0.81	175	0.80	157
ρ_{SP}			0.66	91	0.68	38	0.67	41	0.66	46
ρ	$p = 7$		$p = 8$		$p = 9$		$p = 10$		$p = 11$	
	Base de Fourier									
ρ_{CL}	0.65	58	0.64	63	0.64	65	0.64	67	0.64	69
ρ_{OGK}	0.66	44	0.66	48	0.65	50	0.65	52	0.65	53
$\rho_{\mathbf{V}_M}$	0.65	55	0.65	61	0.64	63	0.64	66	0.64	67
ρ_{DIST}	0.67	39	0.66	42	0.66	43	0.66	44	0.66	44
ρ_{CM}^*	0.79	140	0.79	134	0.79	132	0.79	127	0.79	128
ρ_{SP}	0.65	54	0.65	58	0.65	60	0.64	62	0.64	63
	Base de Splines									
ρ_{CL}	0.65	57	0.65	60	0.64	65	0.64	66	0.64	68
ρ_{OGK}	0.66	44	0.66	46	0.66	50	0.65	51	0.65	53
$\rho_{\mathbf{V}_M}$	0.65	54	0.65	58	0.64	63	0.64	64	0.64	67
ρ_{DIST}	0.67	38	0.66	40	0.66	43	0.66	43	0.66	44
ρ_{CM}^*	0.80	139	0.80	138	0.79	133	0.79	131	0.79	126
ρ_{SP}	0.65	53	0.65	55	0.65	60	0.65	61	0.64	62
	Base de Componentes Principales									
ρ_{CL}	0.65	55	0.65	60	0.64	62	0.64	65	0.64	67
ρ_{OGK}	0.66	42	0.66	45	0.66	48	0.65	51	0.65	52
$\rho_{\mathbf{V}_M}$	0.65	51	0.65	57	0.65	60	0.64	64	0.64	65
ρ_{DIST}	0.67	37	0.67	40	0.66	41	0.66	43	0.66	43
ρ_{CM}^*	0.80	146	0.79	139	0.79	136	0.79	131	0.79	128
ρ_{SP}	0.66	52	0.65	55	0.65	57	0.65	61	0.65	61

Tab. 5.11: Media y error cuadrático medio, multiplicado por 10^4 , de $\hat{\rho}_{\text{CRG},1}$ para distintas medidas de asociación ρ y dimensiones p usando el Algoritmo 3, para datos sin contaminar.

	Media $10^4 \times \text{ECM}$		Media $10^4 \times \text{ECM}$		Media $10^4 \times \text{ECM}$		Media $10^4 \times \text{ECM}$		Media $10^4 \times \text{ECM}$	
ρ	$p = 2$		$p = 3$		$p = 4$		$p = 5$		$p = 6$	
	Base de Fourier									
ρ_{CL}	0.90	415	0.90	419	0.90	416	0.90	414	0.90	412
ρ_{OGK}	0.82	168	0.81	140	0.79	105	0.78	88	0.77	78
$\rho_{\mathbf{V}_M}$	0.85	242	0.84	220	0.82	191	0.81	172	0.80	161
ρ_{DIST}	0.91	471	0.91	473	0.91	470	0.91	468	0.91	467
ρ_{CM}^*	0.95	670	0.94	628	0.92	539	0.91	488	0.90	457
ρ_{SP}	0.78	78	0.77	66	0.75	47	0.74	37	0.73	32
	Base de Splines									
ρ_{CL}			0.90	417	0.90	418	0.90	414	0.90	412
ρ_{OGK}			0.81	139	0.79	117	0.78	88	0.77	82
$\rho_{\mathbf{V}_M}$			0.84	220	0.83	199	0.81	173	0.81	166
ρ_{DIST}			0.91	471	0.91	471	0.91	468	0.91	468
ρ_{CM}^*			0.94	630	0.93	568	0.91	484	0.90	462
ρ_{SP}			0.77	66	0.76	52	0.74	38	0.73	34
	Base de Componentes Principales									
ρ_{CL}			0.67	63	0.71	33	0.74	50	0.77	86
ρ_{OGK}			0.69	71	0.70	38	0.70	31	0.69	32
$\rho_{\mathbf{V}_M}$			0.66	98	0.69	44	0.69	47	0.68	58
ρ_{DIST}			0.68	78	0.72	57	0.76	105	0.81	186
ρ_{CM}^*			0.83	294	0.85	314	0.85	274	0.83	228
ρ_{SP}			0.65	109	0.67	48	0.67	43	0.66	49
ρ	$p = 7$		$p = 8$		$p = 9$		$p = 10$		$p = 11$	
	Base de Fourier									
ρ_{CL}	0.90	410	0.90	409	0.90	409	0.90	409	0.90	409
ρ_{OGK}	0.76	72	0.76	67	0.76	68	0.76	66	0.76	66
$\rho_{\mathbf{V}_M}$	0.80	156	0.80	153	0.80	152	0.80	151	0.80	150
ρ_{DIST}	0.91	467	0.91	466	0.91	466	0.91	466	0.91	466
ρ_{CM}^*	0.89	429	0.89	426	0.89	417	0.89	419	0.89	413
ρ_{SP}	0.72	30	0.72	29	0.72	28	0.72	28	0.72	28
	Base de Splines									
ρ_{CL}	0.90	410	0.90	410	0.90	409	0.90	409	0.90	409
ρ_{OGK}	0.76	73	0.76	72	0.76	68	0.76	67	0.76	66
$\rho_{\mathbf{V}_M}$	0.80	157	0.80	155	0.80	152	0.80	151	0.80	151
ρ_{DIST}	0.91	467	0.91	466	0.91	466	0.91	466	0.91	466
ρ_{CM}^*	0.89	436	0.89	430	0.89	415	0.89	417	0.89	415
ρ_{SP}	0.72	30	0.72	29	0.72	28	0.72	28	0.72	28
	Base de Componentes Principales									
ρ_{CL}	0.84	211	0.86	281	0.91	457	0.93	517	0.95	633
ρ_{OGK}	0.68	36	0.67	38	0.67	40	0.67	41	0.67	41
$\rho_{\mathbf{V}_M}$	0.68	73	0.67	82	0.67	89	0.67	94	0.67	96
ρ_{DIST}	0.86	324	0.89	410	0.93	559	0.95	614	0.96	700
ρ_{CM}^*	0.82	191	0.81	180	0.81	176	0.81	174	0.81	179
ρ_{SP}	0.65	58	0.64	64	0.64	65	0.64	67	0.64	69

Tab. 5.12: Media y error cuadrático medio, multiplicado por 10^4 , de $\hat{\rho}_{\text{CRG},1}$ para distintas medidas de asociación ρ y dimensiones p usando el Algoritmo 3, bajo $C_{1,0,1}$.

	Media $10^4 \times \text{ECM}$		Media $10^4 \times \text{ECM}$		Media $10^4 \times \text{ECM}$		Media $10^4 \times \text{ECM}$		Media $10^4 \times \text{ECM}$	
ρ	$p = 2$		$p = 3$		$p = 4$		$p = 5$		$p = 6$	
	Base de Fourier									
ρ_{CL}	0.71	25	0.78	92	0.88	337	0.88	337	0.88	337
ρ_{OGK}	0.72	32	0.74	40	0.74	40	0.72	32	0.71	29
$\rho_{\mathbf{V}_M}$	0.71	26	0.75	46	0.77	94	0.76	82	0.75	75
ρ_{DIST}	0.72	28	0.81	149	0.91	447	0.91	447	0.91	447
ρ_{CM}^*	0.84	268	0.86	310	0.84	254	0.83	216	0.82	184
ρ_{SP}	0.71	29	0.73	32	0.73	33	0.71	29	0.70	27
	Base de Splines									
ρ_{CL}			0.78	92	0.78	91	0.88	333	0.88	337
ρ_{OGK}			0.74	39	0.73	30	0.72	32	0.71	31
$\rho_{\mathbf{V}_M}$			0.75	46	0.74	39	0.76	81	0.75	78
ρ_{DIST}			0.81	149	0.81	149	0.90	442	0.91	447
ρ_{CM}^*			0.86	310	0.85	270	0.83	210	0.82	194
ρ_{SP}			0.73	30	0.72	25	0.71	26	0.70	27
	Base de Componentes Principales									
ρ_{CL}			0.87	311	0.88	330	0.88	333	0.88	336
ρ_{OGK}			0.70	68	0.73	39	0.72	33	0.71	30
$\rho_{\mathbf{V}_M}$			0.74	94	0.77	87	0.76	79	0.75	75
ρ_{DIST}			0.89	385	0.90	426	0.90	437	0.90	443
ρ_{CM}^*			0.81	221	0.84	242	0.83	207	0.82	195
ρ_{SP}			0.69	68	0.72	32	0.71	28	0.71	27
ρ	$p = 7$		$p = 8$		$p = 9$		$p = 10$		$p = 11$	
	Base de Fourier									
ρ_{CL}	0.88	338	0.88	338	0.88	338	0.88	338	0.88	338
ρ_{OGK}	0.71	29	0.70	29	0.70	29	0.70	28	0.70	28
$\rho_{\mathbf{V}_M}$	0.75	72	0.74	71	0.74	70	0.74	69	0.74	70
ρ_{DIST}	0.91	447	0.91	447	0.91	448	0.91	448	0.91	448
ρ_{CM}^*	0.82	177	0.81	165	0.81	154	0.81	159	0.81	156
ρ_{SP}	0.70	27	0.69	28	0.69	28	0.69	28	0.69	28
	Base de Splines									
ρ_{CL}	0.88	338	0.88	338	0.88	338	0.88	338	0.88	338
ρ_{OGK}	0.71	28	0.70	29	0.70	29	0.70	29	0.70	29
$\rho_{\mathbf{V}_M}$	0.75	73	0.74	72	0.74	70	0.74	70	0.74	70
ρ_{DIST}	0.91	447	0.91	447	0.91	447	0.91	448	0.91	448
ρ_{CM}^*	0.81	166	0.81	172	0.81	164	0.81	157	0.81	158
ρ_{SP}	0.69	27	0.69	27	0.69	28	0.69	28	0.69	29
	Base de Componentes Principales									
ρ_{CL}	0.88	337	0.88	337	0.88	338	0.88	338	0.88	338
ρ_{OGK}	0.71	29	0.70	29	0.70	28	0.70	29	0.70	29
$\rho_{\mathbf{V}_M}$	0.75	73	0.75	71	0.74	71	0.74	70	0.74	70
ρ_{DIST}	0.91	445	0.91	447	0.91	447	0.91	447	0.91	448
ρ_{CM}^*	0.82	179	0.81	172	0.81	168	0.81	164	0.81	158
ρ_{SP}	0.70	27	0.70	27	0.70	27	0.69	28	0.69	28

Tab. 5.13: Media y error cuadrático medio, multiplicado por 10^4 , de $\hat{\rho}_{\text{CRG},1}$ para distintas medidas de asociación ρ y dimensiones p usando el Algoritmo 3, bajo $C_{2,0,1}$.

ρ	C_0										$C_{1,0,1}$										$C_{2,0,1}$										
	p																														
	2	3	4	5	6	7	8	9	10	11	2	3	4	5	6	7	8	9	10	11	2	3	4	5	6	7	8	9	10	11	
	Base de Fourier																														
ρ_{CL}	51	0	0	0	0	0	0	0	0	0	315	35	12	10	11	12	9	15	11	16	57	12	0	0	0	0	0	0	0	0	0
ρ_{V_M}	76	5	3	2	1	0	1	2	2	0	405	130	44	32	27	29	28	34	39	34	55	95	61	20	13	3	1	2	0	0	0
ρ_{SP}	39	9	7	11	13	20	24	22	25	42	552	437	271	144	111	65	57	57	50	53	36	199	90	56	53	43	42	33	37	24	0
	Base de Splines																														
ρ_{CL}	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	5	1	1	0	0	0	0	0	0	0	0
ρ_{V_M}	1	0	0	0	0	0	0	0	0	0	218	127	61	0	0	0	0	0	0	0	204	121	52	32	16	7	3	3	1	0	0
ρ_{SP}	18	7	5	0	0	0	0	0	0	0	173	87	44	0	0	0	0	0	0	0	250	106	36	45	34	15	14	14	11	0	0
	Base de Componentes Principales																														
ρ_{CL}	338	20	0	0	0	0	0	0	0	0	0	0	0	0	4	18	399	627	961	849	360	33	1	0	0	0	0	0	0	0	
ρ_{V_M}	2	0	0	0	0	1	1	2	1	0	12	3	2	0	0	0	5	6	21	217	138	47	24	17	8	5	2	1	0	0	
ρ_{SP}	12	7	7	10	13	18	30	26	35	0	61	34	24	15	16	16	12	11	9	224	120	69	54	67	61	63	54	49	0	0	

Tab. 5.14: Cantidad de veces en las 1000 replicaciones en que el valor de $\hat{\rho}_1$ dado por el Algoritmo 3 es mayor que el obtenido utilizando el Algoritmo GRID de Alfons *et al.* (2016).

	$\widehat{c}_{\text{COMB}}$	$\widehat{\rho}_{X,\text{COMB}}$	AISE	$\widehat{c}_{\text{COMB}}$	$\widehat{\rho}_{X,\text{COMB}}$	AISE	$\widehat{c}_{\text{COMB}}$	$\widehat{\rho}_{X,\text{COMB}}$	AISE	$\widehat{c}_{\text{COMB}}$	$\widehat{\rho}_{X,\text{COMB}}$	AISE	$\widehat{c}_{\text{COMB}}$	$\widehat{\rho}_{X,\text{COMB}}$	AISE
ρ	$p = 2$			$p = 3$			$p = 4$			$p = 5$			$p = 6$		
	Base de Fourier														
ρ_{CL}	0.99	0.99	0.04	0.99	0.99	0.06	0.98	0.98	0.09	0.97	0.97	0.14	0.95	0.97	0.20
$\rho_{\mathbf{V}_M}$	0.99	0.99	0.05	0.98	0.98	0.07	0.97	0.97	0.11	0.96	0.97	0.17	0.94	0.96	0.25
ρ_{SP}	0.99	0.98	0.06	0.98	0.98	0.08	0.97	0.97	0.12	0.96	0.96	0.18	0.94	0.95	0.25
	Base de Splines														
ρ_{CL}				0.99	0.99	0.06	0.98	0.98	0.10	0.97	0.97	0.14	0.95	0.97	0.21
$\rho_{\mathbf{V}_M}$				0.98	0.98	0.08	0.97	0.97	0.13	0.95	0.96	0.18	0.93	0.96	0.28
ρ_{SP}				0.98	0.97	0.10	0.96	0.96	0.18	0.94	0.95	0.23	0.91	0.94	0.35
	Base de Componentes Principales														
ρ_{CL}				0.92	0.94	0.30	0.96	0.97	0.17	0.96	0.97	0.16	0.95	0.97	0.21
$\rho_{\mathbf{V}_M}$				0.91	0.92	0.38	0.95	0.96	0.19	0.95	0.96	0.20	0.93	0.96	0.27
ρ_{SP}				0.90	0.92	0.40	0.94	0.95	0.22	0.94	0.95	0.24	0.92	0.94	0.32
ρ	$p = 7$			$p = 8$			$p = 9$			$p = 10$			$p = 11$		
	Base de Fourier														
ρ_{CL}	0.93	0.96	0.27	0.91	0.95	0.38	0.88	0.95	0.49	0.84	0.94	0.65	0.79	0.93	0.83
$\rho_{\mathbf{V}_M}$	0.91	0.95	0.34	0.88	0.94	0.48	0.84	0.92	0.64	0.79	0.91	0.82	0.74	0.90	1.05
ρ_{SP}	0.91	0.94	0.36	0.88	0.93	0.47	0.83	0.90	0.66	0.79	0.89	0.84	0.73	0.86	1.06
	Base de Splines														
ρ_{CL}	0.93	0.96	0.28	0.90	0.95	0.40	0.87	0.95	0.50	0.83	0.94	0.68	0.79	0.93	0.85
$\rho_{\mathbf{V}_M}$	0.91	0.95	0.36	0.87	0.94	0.51	0.84	0.92	0.65	0.79	0.91	0.84	0.74	0.90	1.04
ρ_{SP}	0.89	0.92	0.46	0.85	0.91	0.60	0.82	0.90	0.73	0.79	0.90	0.86	0.75	0.89	0.98
	Base de Componentes Principales														
ρ_{CL}	0.93	0.96	0.29	0.90	0.95	0.42	0.86	0.95	0.56	0.81	0.94	0.77	0.75	0.93	0.98
$\rho_{\mathbf{V}_M}$	0.90	0.95	0.38	0.86	0.94	0.54	0.82	0.92	0.73	0.76	0.91	0.96	0.70	0.90	1.19
ρ_{SP}	0.89	0.93	0.44	0.86	0.92	0.57	0.82	0.91	0.74	0.76	0.88	0.94	0.72	0.87	1.13

Tab. 5.15: Medidas resumen correspondientes a la estimación de las primeras direcciones canónicas para distintas medidas de asociación ρ y dimensiones $p = q$ usando el Algoritmo 4. Se presentan el AISE, la media sobre repeticiones del angulo medido a través del $\widehat{c}_{\text{COMB}} = |\cos(\theta_{\Phi_1, \widehat{\Phi}_{\text{COMB},1}})|$ y del valor $\widehat{\rho}_{X,\text{COMB}} = \widehat{\rho}_{XX}(\Phi_1, \widehat{\Phi}_{\text{COMB},1})$ correspondiente a la medida de asociación usada para datos sin contaminar.

	\hat{c}_{COMB}	$\hat{\rho}_{X,\text{COMB}}$	AISE	\hat{c}_{COMB}	$\hat{\rho}_{X,\text{COMB}}$	AISE	\hat{c}_{COMB}	$\hat{\rho}_{X,\text{COMB}}$	AISE	\hat{c}_{COMB}	$\hat{\rho}_{X,\text{COMB}}$	AISE	\hat{c}_{COMB}	$\hat{\rho}_{X,\text{COMB}}$	AISE
ρ	$p = 2$			$p = 3$			$p = 4$			$p = 5$			$p = 6$		
	Base de Fourier														
ρ_{CL}	0.20	0.09	3.19	0.20	0.09	3.19	0.20	0.09	3.19	0.20	0.09	3.20	0.20	0.09	3.21
$\rho_{\mathbf{V}_M}$	0.85	0.73	0.60	0.84	0.71	0.65	0.82	0.70	0.71	0.80	0.69	0.80	0.78	0.67	0.90
ρ_{SP}	0.96	0.83	0.17	0.94	0.83	0.24	0.93	0.84	0.29	0.92	0.85	0.34	0.89	0.85	0.45
	Base de Splines														
ρ_{CL}				0.20	0.09	3.18	0.20	0.09	3.19	0.20	0.09	3.19	0.20	0.09	3.20
$\rho_{\mathbf{V}_M}$				0.83	0.71	0.67	0.82	0.70	0.74	0.80	0.68	0.80	0.78	0.67	0.90
ρ_{SP}				0.94	0.82	0.24	0.92	0.80	0.33	0.90	0.79	0.40	0.87	0.77	0.54
	Base de Componentes Principales														
ρ_{CL}				0.18	0.09	3.29	0.20	0.09	3.21	0.20	0.09	3.20	0.20	0.09	3.20
$\rho_{\mathbf{V}_M}$				0.74	0.63	1.03	0.80	0.68	0.79	0.80	0.68	0.81	0.77	0.67	0.91
ρ_{SP}				0.81	0.74	0.73	0.89	0.79	0.45	0.88	0.79	0.50	0.83	0.76	0.67
ρ	$p = 7$			$p = 8$			$p = 9$			$p = 10$			$p = 11$		
	Base de Fourier														
ρ_{CL}	0.20	0.10	3.21	0.20	0.10	3.22	0.19	0.10	3.23	0.19	0.10	3.26	0.18	0.10	3.29
$\rho_{\mathbf{V}_M}$	0.75	0.66	1.01	0.72	0.65	1.14	0.68	0.63	1.30	0.63	0.62	1.48	0.58	0.60	1.69
ρ_{SP}	0.86	0.84	0.58	0.83	0.83	0.71	0.79	0.82	0.84	0.74	0.79	1.07	0.69	0.78	1.26
	Base de Splines														
ρ_{CL}	0.20	0.09	3.22	0.19	0.10	3.24	0.19	0.10	3.24	0.18	0.10	3.28	0.17	0.10	3.31
$\rho_{\mathbf{V}_M}$	0.75	0.66	0.99	0.72	0.64	1.14	0.67	0.62	1.30	0.63	0.61	1.49	0.59	0.59	1.66
ρ_{SP}	0.83	0.76	0.68	0.79	0.74	0.84	0.76	0.73	0.95	0.72	0.72	1.11	0.69	0.71	1.24
	Base de Componentes Principales														
ρ_{CL}	0.20	0.09	3.22	0.19	0.10	3.25	0.13	0.10	3.49	0.09	0.09	3.65	0.04	0.09	3.83
$\rho_{\mathbf{V}_M}$	0.74	0.65	1.03	0.70	0.63	1.21	0.65	0.61	1.41	0.59	0.60	1.64	0.53	0.58	1.87
ρ_{SP}	0.80	0.74	0.82	0.74	0.71	1.04	0.69	0.69	1.24	0.64	0.68	1.42	0.59	0.65	1.64

Tab. 5.16: Medidas resumen correspondientes a la estimación de las primeras direcciones canónicas para distintas medidas de asociación ρ y dimensiones $p = q$ usando el Algoritmo 4 bajo $C_{1,0,1}$. Se presentan el AISE, la media sobre repeticiones del angulo medido a través del $\hat{c}_{\text{COMB}} = |\cos(\theta_{\Phi_1, \hat{\Phi}_{\text{COMB},1}})|$ y del valor $\hat{\rho}_{X,\text{COMB}} = \hat{\rho}_{XX}(\Phi_1, \hat{\Phi}_{\text{COMB},1})$ correspondientes a las medidas de asociación usadas.

	$\widehat{c}_{\text{COMB}}$	$\widehat{\rho}_{X,\text{COMB}}$	AISE	$\widehat{c}_{\text{COMB}}$	$\widehat{\rho}_{X,\text{COMB}}$	AISE	$\widehat{c}_{\text{COMB}}$	$\widehat{\rho}_{X,\text{COMB}}$	AISE	$\widehat{c}_{\text{COMB}}$	$\widehat{\rho}_{X,\text{COMB}}$	AISE	$\widehat{c}_{\text{COMB}}$	$\widehat{\rho}_{X,\text{COMB}}$	AISE
ρ	$p = 2$			$p = 3$			$p = 4$			$p = 5$			$p = 6$		
	Base de Fourier														
ρ_{CL}	0.99	0.99	0.04	0.56	0.44	1.75	0.26	0.17	2.96	0.26	0.17	2.97	0.25	0.17	2.99
$\rho_{\mathbf{V}_M}$	0.99	0.99	0.05	0.88	0.80	0.48	0.82	0.74	0.71	0.81	0.73	0.77	0.79	0.72	0.86
ρ_{SP}	0.99	0.98	0.06	0.95	0.89	0.21	0.93	0.85	0.28	0.91	0.84	0.36	0.88	0.83	0.47
	Base de Splines														
ρ_{CL}				0.56	0.44	1.75	0.55	0.44	1.79	0.26	0.17	2.98	0.25	0.17	2.99
$\rho_{\mathbf{V}_M}$				0.87	0.79	0.52	0.85	0.77	0.62	0.81	0.73	0.77	0.78	0.71	0.88
ρ_{SP}				0.93	0.85	0.30	0.90	0.83	0.38	0.89	0.79	0.45	0.85	0.77	0.62
	Base de Componentes Principales														
ρ_{CL}				0.21	0.15	3.15	0.23	0.16	3.06	0.25	0.17	2.98	0.25	0.17	3.00
$\rho_{\mathbf{V}_M}$				0.70	0.63	1.18	0.81	0.73	0.77	0.81	0.73	0.78	0.78	0.71	0.87
ρ_{SP}				0.77	0.70	0.91	0.89	0.80	0.43	0.89	0.80	0.44	0.85	0.78	0.58
ρ	$p = 7$			$p = 8$			$p = 9$			$p = 10$			$p = 11$		
	Base de Fourier														
ρ_{CL}	0.25	0.17	3.02	0.24	0.17	3.04	0.23	0.17	3.07	0.22	0.17	3.10	0.22	0.17	3.14
$\rho_{\mathbf{V}_M}$	0.75	0.70	0.98	0.72	0.68	1.14	0.67	0.67	1.30	0.63	0.65	1.49	0.58	0.63	1.69
ρ_{SP}	0.84	0.81	0.62	0.80	0.79	0.80	0.75	0.77	1.00	0.69	0.74	1.22	0.64	0.72	1.46
	Base de Splines														
ρ_{CL}	0.25	0.17	3.02	0.24	0.17	3.05	0.23	0.17	3.07	0.22	0.17	3.11	0.21	0.17	3.15
$\rho_{\mathbf{V}_M}$	0.75	0.70	0.98	0.72	0.69	1.10	0.69	0.67	1.25	0.64	0.66	1.42	0.60	0.65	1.58
ρ_{SP}	0.81	0.75	0.73	0.78	0.75	0.86	0.75	0.73	0.99	0.72	0.73	1.10	0.68	0.72	1.25
	Base de Componentes Principales														
ρ_{CL}	0.25	0.17	3.02	0.24	0.17	3.05	0.23	0.17	3.09	0.22	0.17	3.14	0.20	0.17	3.19
$\rho_{\mathbf{V}_M}$	0.75	0.70	1.00	0.71	0.68	1.17	0.66	0.66	1.35	0.61	0.65	1.57	0.55	0.63	1.80
ρ_{SP}	0.82	0.76	0.74	0.76	0.73	0.95	0.70	0.70	1.20	0.65	0.68	1.41	0.58	0.65	1.67
—															

5. Estudio de Monte Carlo

Tab. 5.17: Medidas resumen correspondientes a la estimación de las primeras direcciones canónicas para distintas medidas de asociación ρ y dimensiones $p = q$ usando el Algoritmo 4 bajo $C_{2,0,1}$. Se presentan el AISE, la media sobre repeticiones del ángulo medido a través del $\widehat{c}_{\text{COMB}} = |\cos(\theta_{\Phi_1, \widehat{\Phi}_{\text{COMB},1}})|$ y del valor $\widehat{\rho}_{X,\text{COMB}} = \widehat{\rho}_{XX}(\Phi_1, \widehat{\Phi}_{\text{COMB},1})$ correspondientes a las medidas de asociación usadas.

	Media $10^4 \times \text{ECM}$		Media $10^4 \times \text{ECM}$		Media $10^4 \times \text{ECM}$		Media $10^4 \times \text{ECM}$		Media $10^4 \times \text{ECM}$	
ρ	$p = 2$		$p = 3$		$p = 4$		$p = 5$		$p = 6$	
	Base de Fourier									
ρ_{CL}	0.71	25	0.72	25	0.72	27	0.73	30	0.74	35
ρ_{V_M}	0.71	26	0.72	27	0.73	30	0.74	36	0.75	42
ρ_{SP}	0.71	29	0.73	33	0.74	38	0.75	47	0.76	56
	Base de Splines									
ρ_{CL}			0.72	25	0.72	27	0.73	30	0.74	35
ρ_{V_M}			0.72	27	0.73	30	0.74	36	0.75	43
ρ_{SP}			0.73	33	0.74	39	0.75	47	0.76	55
	Base de Componentes Principales									
ρ_{CL}			0.68	65	0.71	29	0.73	30	0.74	34
ρ_{V_M}			0.67	81	0.72	31	0.74	35	0.75	42
ρ_{SP}			0.68	78	0.73	38	0.75	45	0.76	55
ρ	$p = 7$		$p = 8$		$p = 9$		$p = 10$		$p = 11$	
ρ_{CL}	0.75	40	0.75	47	0.76	54	0.77	64	0.78	74
ρ_{V_M}	0.76	51	0.77	62	0.77	75	0.79	90	0.79	106
ρ_{SP}	0.76	68	0.77	79	0.78	91	0.78	102	0.79	113
	Base de Splines									
ρ_{CL}	0.75	40	0.75	47	0.76	54	0.77	63	0.78	74
ρ_{V_M}	0.76	51	0.77	62	0.78	74	0.78	88	0.79	105
ρ_{SP}	0.76	66	0.77	75	0.78	84	0.78	93	0.79	102
	Base de Componentes Principales									
ρ_{CL}	0.75	40	0.75	47	0.76	55	0.77	63	0.78	73
ρ_{V_M}	0.76	50	0.77	62	0.78	75	0.78	89	0.79	104
ρ_{SP}	0.76	66	0.77	75	0.78	88	0.78	95	0.78	104

Tab. 5.18: Media y error cuadrático medio, multiplicado por 10^4 , de $\hat{\rho}_{\text{COMB},1}$ para distintas medidas de asociación ρ y dimensiones p usando el Algoritmo 4 para datos sin contaminar.

	Media $10^4 \times \text{ECM}$		Media $10^4 \times \text{ECM}$		Media $10^4 \times \text{ECM}$		Media $10^4 \times \text{ECM}$		Media $10^4 \times \text{ECM}$	
ρ	$p = 2$		$p = 3$		$p = 4$		$p = 5$		$p = 6$	
	Base de Fourier									
ρ_{CL}	0.90	415	0.90	424	0.91	433	0.91	443	0.91	451
ρ_{VM}	0.85	244	0.85	253	0.85	268	0.86	283	0.87	299
ρ_{SP}	0.78	80	0.78	82	0.78	82	0.78	88	0.79	98
	Base de Splines									
ρ_{CL}			0.90	422	0.91	434	0.91	443	0.91	452
ρ_{VM}			0.85	247	0.85	257	0.85	265	0.86	287
ρ_{SP}			0.79	92	0.79	103	0.80	114	0.80	129
	Base de Componentes Principales									
ρ_{CL}			0.90	411	0.90	426	0.91	439	0.91	450
ρ_{VM}			0.82	202	0.85	247	0.85	267	0.86	286
ρ_{SP}			0.73	65	0.78	85	0.79	105	0.80	121
ρ	$p = 7$		$p = 8$		$p = 9$		$p = 10$		$p = 11$	
ρ_{CL}	0.91	460	0.91	470	0.92	479	0.92	488	0.92	498
ρ_{VM}	0.87	314	0.88	331	0.88	347	0.89	363	0.89	385
ρ_{SP}	0.79	111	0.80	123	0.80	132	0.81	144	0.82	159
	Base de Splines									
ρ_{CL}	0.91	461	0.91	470	0.92	480	0.92	490	0.92	499
ρ_{VM}	0.87	304	0.87	322	0.88	339	0.88	361	0.89	379
ρ_{SP}	0.81	139	0.81	149	0.82	165	0.82	173	0.83	183
	Base de Componentes Principales									
ρ_{CL}	0.91	460	0.92	471	0.92	506	0.93	538	0.95	634
ρ_{VM}	0.87	304	0.87	321	0.88	340	0.88	360	0.89	380
ρ_{SP}	0.80	133	0.81	146	0.81	153	0.82	161	0.82	174

Tab. 5.19: Media y error cuadrático medio, multiplicado por 10^4 , de $\hat{\rho}_{\text{COMB},1}$ para distintas medidas de asociación ρ y dimensiones p usando el Algoritmo 4, bajo $C_{1,0,1}$.

	Media $10^4 \times \text{ECM}$		Media $10^4 \times \text{ECM}$		Media $10^4 \times \text{ECM}$		Media $10^4 \times \text{ECM}$		Media $10^4 \times \text{ECM}$	
ρ	$p = 2$		$p = 3$		$p = 4$		$p = 5$		$p = 6$	
	Base de Fourier									
ρ_{CL}	0.71	25	0.79	98	0.88	352	0.89	361	0.89	371
ρ_{VM}	0.71	26	0.76	58	0.81	143	0.81	157	0.82	169
ρ_{SP}	0.71	29	0.75	45	0.77	72	0.78	81	0.78	92
	Base de Splines									
ρ_{CL}			0.79	98	0.79	107	0.89	357	0.89	371
ρ_{VM}			0.76	56	0.76	62	0.81	148	0.82	163
ρ_{SP}			0.75	44	0.75	53	0.78	83	0.78	92
	Base de Componentes Principales									
ρ_{CL}			0.87	313	0.88	337	0.88	354	0.89	368
ρ_{VM}			0.75	102	0.80	128	0.81	147	0.82	164
ρ_{SP}			0.70	66	0.76	62	0.77	78	0.78	92
ρ	$p = 7$		$p = 8$		$p = 9$		$p = 10$		$p = 11$	
ρ_{CL}	0.89	381	0.90	391	0.90	402	0.90	412	0.90	422
ρ_{VM}	0.83	186	0.83	202	0.84	219	0.85	236	0.85	252
ρ_{SP}	0.79	103	0.79	112	0.80	122	0.80	131	0.81	143
	Base de Splines									
ρ_{CL}	0.89	382	0.90	392	0.90	402	0.90	412	0.90	423
ρ_{VM}	0.82	175	0.83	189	0.84	207	0.84	223	0.85	245
ρ_{SP}	0.79	101	0.80	116	0.80	127	0.81	136	0.81	146
	Base de Componentes Principales									
ρ_{CL}	0.89	380	0.90	391	0.90	402	0.90	412	0.90	422
ρ_{VM}	0.82	178	0.83	195	0.84	211	0.84	227	0.85	243
ρ_{SP}	0.79	103	0.79	112	0.80	122	0.80	128	0.80	137

Tab. 5.20: Media y error cuadrático medio, multiplicado por 10^4 , de $\hat{\rho}_{\text{COMB},1}$ para distintas medidas de asociación ρ y dimensiones p usando el Algoritmo 4, bajo $C_{2,0,1}$.

5.4.2. Resultados eligiendo la dimensión de la base mediante convalidación cruzada

En esta sección, se resumen los resultados obtenidos cuando la dimensión de las bases se elige tal como se describió en la Sección 5.3, de acuerdo al criterio de convalidación cruzada $RCV_{\mathbf{r},1}$ definido en (5.5) donde $\mathbf{r} = (p, q)$. Como se menciona en la Sección 5.3, la búsqueda se realizó tomando $p = q$ y valores de p que variaban en $\mathcal{R}^* = \{1, \dots, 11\}$ para las bases de Fourier y $\mathcal{R}^* = \{3, \dots, 11\}$ para bases de Splines y de componentes principales. Una vez obtenido el valor $\hat{\mathbf{r}}_{\text{máx}} = (\hat{p}, \hat{p})^T$ definido en (5.6), se estiman la primera correlación canónica y las primeras direcciones canónicas utilizando \hat{p} elementos de la base. Tanto esta última estimación como las estimaciones de direcciones canónicas quitando una observación de la muestra anteriores a la maximización (5.6) se realizan con el mismo algoritmo, ya sea el Algoritmo 2 o el 3.

En las tablas presentadas a continuación se reporta, en la columna \hat{p} , la mediana de los valores \hat{p} sobre las nr replicaciones además de las medidas resumen utilizadas anteriormente. En las tablas que corresponden a la estimación de la primera correlación canónica ρ_1 mediante convalidación cruzada, se presentan además el promedio sobre las nr replicaciones y el error cuadrático medio de $\tilde{\rho}_{\hat{p},1} = \sqrt{RCV_{\hat{p},1}}$, es decir, el promedio y ECM de las correlaciones máximas definidas en (5.5). Como en He *et al.* (2004), el valor $\tilde{\rho}_{\hat{p},1}$ puede pensarse como un estimador del verdadero valor ρ_1 . En dicho trabajo, donde se utilizó como medida de asociación la correlación de Pearson, He *et al.* (2004) denominan a $\tilde{\rho}_{\hat{p},1}$ la correlación canónica empírica (*Empirical canonical correlation*) ya que, en las simulaciones allí reportadas, $\tilde{\rho}_{\hat{p},1}$ resultó ser el mejor estimador de la primera correlación canónica. Respecto a las medidas de asociación utilizadas, son las mismas que las utilizadas en la sección previa salvo que aquí no utilizamos la medida ρ_{DIST} pues mostró un comportamiento similar a la clásica en bases fijas.

Como se mencionó anteriormente, si bien para dimensión fija el Algoritmo 2 mostró

ser mejor que el Algoritmo 3 para aproximar los estimadores dados en (4.2), realizaremos la comparación del comportamiento de las estimaciones obtenidas mediante convalidación cruzada para ambos algoritmos ya que el procedimiento introducido por Alfons *et al.* (2016) está implementado en R solamente cuando la medida de asociación es ρ_{CL} , ρ_{SP} o $\rho_{\mathbf{V}_M}$. La implementación de dicho algoritmo para una medida de asociación general que incluya todas las consideradas en esta tesis, excede los límites de la misma.

En la Tabla 5.21 se pueden encontrar los resultados correspondientes a la estimación de las primeras direcciones canónicas usando convalidación cruzada y el Algoritmo 2. En general, para muestras sin contaminar, todos los estimadores tienen un buen desempeño, ya que tanto la media del coseno del ángulo entre los estimadores de la dirección canónica y la dirección real como la medida de asociación $\hat{\rho}_{XX}(\Phi_1, \hat{\Phi}_{GRID,1})$ resultan cercanas a 1. Como era de esperarse en esta situación todos los procedimientos tiene un AISE pequeño, y al elegir en forma adaptiva la dimensión se obtienen valores muy similares a las dimensiones que minimizaban el AISE en la Tabla 5.2. Tal como observáramos en la Tabla 5.2, cuando no hay datos atípicos, la medida de correlación de Pearson ρ_{CL} provee los mejores resultados para las bases de Splines y componentes principales ya que se consigue el mínimo AISE y los valores promedios del $|\cos(\theta_{\Phi_1, \hat{\Phi}_{GRID,1}})|$ y de $\hat{\rho}_{XX}(\Phi_1, \hat{\Phi}_{GRID,1})$ más cercanos a 1. Sin embargo, como era de esperar, las estimaciones usando funcionales robustos dan resultados muy similares a los obtenidos con la correlación de Pearson. Teniendo en cuenta que los datos se generaron usando una base de Fourier, los menores valores de AISE se obtienen cuando se utiliza esta base seguida por la base de Splines.

Como en el caso en que la dimensión de la base es fija, bajo ambos escenarios de contaminación, el estimador basado en la correlación de Pearson, ρ_{CL} , se ve fuertemente afectado por la presencia de datos atípicos. Efectivamente, en este caso, los valores promedios del coseno del ángulo y de la medida de asociación entre direcciones $\hat{\rho}_{XX}(\Phi_1, \hat{\Phi}_{GRID,1})$ se alejan de 1 acercándose a 0, por lo que los valores de AISE crecen notoriamente tomando valores muy

cercanos al máximo valor que es 4. En cambio, la contaminación introducida sólo produce un leve sesgo en los estimadores robustos de la primera dirección canónica que se traduce en una reducción del valor medio del coseno del ángulo. Sin embargo, dicha reducción no es drástica ya que los valores son siempre mayores a 0.75. Asimismo, los valores de AISE crecen pero nunca superan el valor 1. Como en el caso en que la dimensión de la base era fija, entre los estimadores robustos se destaca el basado en la medida de correlación de Spearman ρ_{SP} ya que presenta el menor valor de AISE. Por otra parte, la media sobre replicaciones de $|\cos(\theta_{\Phi_1, \hat{\Phi}_{GRID,1}})|$ y $\hat{\rho}_{XX}(\Phi_1, \hat{\Phi}_{GRID,1})$ es la más cercana a 1 para las bases consideradas. Estos resultados muestran semejanzas con los reportados por Alfons *et al.* (2016) para el caso de datos multivariados. Vale la pena mencionar que al igual que para datos Gaussianos, los procedimientos basados en la base de Fourier con la que se generaron los datos dan los valores más pequeños de AISE, seguidos por los de bases de Splines.

La Tabla 5.22 reporta la media y el error cuadrático medio, multiplicado por 10^4 , de los estimadores de la primera correlación canónica cuando usamos convalidación cruzada y el Algoritmo 2. Tal como comentamos en la Sección 5.4.1 en lo referente a la Tabla 5.5, cuando no hay contaminación, todos los estimadores de la primera correlación canónica presentan un desempeño aceptable. Los valores medios de $\hat{\rho}_{GRID,1}$ son cercanos al valor que se desea estimar ($\rho_1 = 0,7$), con un sesgo positivo en todos los casos, y los ECM resultan bajos. Por otro lado, observamos que el estimador $\tilde{\rho}_{\hat{p},1} = \tilde{\rho}_{\hat{p},GRID,1}$ resulta levemente negativamente sesgado para bases de Splines y componentes principales, dando valores medios apenas por debajo del valor que se desea estimar para todas las medidas de asociación, siendo el sesgo menor en valor absoluto que el del estimador $\hat{\rho}_{GRID,1}$. Los valores del error cuadrático medio de $\hat{\rho}_{GRID,1}$ son muy similares entre sí y apenas mayores que los mínimos ECM observados en la Tabla 5.5. Para todos los funcionales de correlación y para todas las bases, el error cuadrático medio de $\tilde{\rho}_{\hat{p},GRID,1}$ es algo menor que el de $\hat{\rho}_{GRID,1}$. Este fenómeno puede explicarse por el efecto del menor sesgo en valor absoluto de $\tilde{\rho}_{\hat{p},GRID,1}$. Si bien el error cuadrático medio

tiene una componente relacionada con la variabilidad, cuyos valores no reportamos aquí, la reducción en sesgo compensa la mayor dispersión de $\tilde{\rho}_{\hat{p},\text{GRID},1}$ respecto de $\hat{\rho}_{\text{GRID},1}$. Como es usual, para muestras no contaminadas y para todas las bases, al utilizar el coeficiente de correlación de Pearson, ρ_{CL} , se obtienen errores cuadráticos medios inferiores a los obtenidos con los estimadores basados en funcionales robustos de asociación.

Respecto de la dimensión de las bases \hat{p} , el método de convalidación cruzada elige dimensiones bajas y cercanas a la dimensión que minimizaba el AISE cuando p era fijo, ver Tabla 5.2. Observemos que en algunos casos la mediana no da un valor entero por lo que tanto la parte entera de dicho número como la parte entera más uno, resultan dimensiones razonables.

Para muestras contaminadas, los estimadores de la primera correlación canónica $\hat{\rho}_{\text{GRID},1}$ y $\tilde{\rho}_{\hat{p},\text{GRID},1}$ basados en el coeficiente de Pearson se ven fuertemente afectados ya que la media de sus estimaciones se acerca a 0,9 bajo $C_{1,0,1}$ y 0,89 bajo $C_{2,0,1}$. Entre los estimadores de la primera correlación canónica basados en las dos medidas de asociación robustas consideradas, el estimador basado en la correlación de Spearman ρ_{SP} resulta ser el más estable y con mejor comportamiento, como en el caso de dimensión fija. Efectivamente, para este estimador el incremento de su ECM respecto del obtenido para muestras Gaussianas sin contaminar y el sesgo de $\hat{\rho}_{\text{GRID},1}$ resultan moderados y menores que cuando se utiliza como medida de asociación $\rho_{\mathbf{V}_M}$. Por otra parte, el error cuadrático medio de $\tilde{\rho}_{\hat{p},\text{GRID},1}$ es algo menor que el obtenido con $\hat{\rho}_{\text{GRID},1}$, en todos los casos. Podemos explicar este resultado por la naturaleza de la contaminación introducida que tiende a aumentar los valores de los estimadores de la correlación. Recordemos que para muestras Gaussianas $\tilde{\rho}_{\hat{p},\text{GRID},1}$ tenía un sesgo negativo que en valor absoluto era menor que el sesgo positivo de $\hat{\rho}_{\text{GRID},1}$. En este caso, por efecto de la contaminación considerada los estimadores $\tilde{\rho}_{\hat{p},\text{GRID},1}$ y $\hat{\rho}_{\text{GRID},1}$ dan valores mayores resultando los estimadores $\tilde{\rho}_{\hat{p},\text{GRID},1}$ con menor sesgo que los estimadores $\hat{\rho}_{\text{GRID},1}$.

Cuando consideramos estimadores basados en medidas de asociación robustas, como

ocurría para datos Gaussianos, bajo $C_{1,0,1}$ y $C_{2,0,1}$, la mediana de la dimensión \hat{p} obtenida por el procedimiento de convalidación cruzada robusto, está próxima a la dimensión que minimizaba el AISE cuando p era fijo, ver Tablas 5.3 y 5.4.

	C_0				$C_{1,0,1}$				$C_{2,0,1}$			
ρ	\hat{p}	\hat{c}_{GRID}	$\hat{\rho}_{X,\text{GRID}}$	AISE	\hat{p}	\hat{c}_{GRID}	$\hat{\rho}_{X,\text{GRID}}$	AISE	\hat{p}	\hat{c}_{GRID}	$\hat{\rho}_{X,\text{GRID}}$	AISE
	Base de Fourier											
ρ_{CL}	1	0.98	0.99	0.08	2	0.18	0.09	3.26	4	0.21	0.14	3.12
$\rho_{\mathbf{V}_M}$	1	0.98	0.99	0.07	4	0.81	0.68	0.79	4	0.83	0.77	0.66
ρ_{SP}	1	0.98	0.98	0.11	2	0.93	0.84	0.29	3	0.91	0.86	0.34
	Base de Splines											
ρ_{CL}	3	0.96	0.98	0.16	3	0.20	0.10	3.20	5	0.20	0.14	3.14
$\rho_{\mathbf{V}_M}$	4	0.95	0.97	0.23	4	0.80	0.67	0.82	5	0.78	0.72	0.88
ρ_{SP}	4	0.94	0.96	0.24	4	0.92	0.79	0.33	5	0.87	0.80	0.52
	Base de Componentes Principales											
ρ_{CL}	4	0.94	0.97	0.22	3	0.17	0.09	3.28	4	0.19	0.13	3.21
$\rho_{\mathbf{V}_M}$	4.50	0.92	0.95	0.32	4.50	0.76	0.65	0.95	5	0.76	0.69	0.97
ρ_{SP}	5	0.92	0.95	0.31	4	0.86	0.78	0.54	5	0.84	0.78	0.61

Tab. 5.21: Mediana de la dimensión obtenida \hat{p} mediante convalidación cruzada, AISE y medias sobre replicaciones de $\hat{c}_{\text{GRID}} = |\cos(\theta_{\Phi_1, \hat{\Phi}_1})|$ y $\hat{\rho}_{X,\text{GRID}} = \hat{\rho}_{XX}(\Phi_1, \hat{\Phi}_1)$ correspondientes al Algoritmo 2 y distintas medidas de asociación.

Las Tablas 5.23 y 5.24 presentan medidas resumen para los estimadores de la primera dirección canónica en el espacio X y para la primera correlación canónica, respectivamente cuando se utiliza convalidación cruzada para elegir la dimensión del subespacio aproximante combinada con el Algoritmo 3 para calcular los estimadores. Los resultados obtenidos son similares a los obtenidos en la Tabla 5.21 y le caben comentarios análogos, si bien vale la pena observar que, como en el caso de dimensión fija, al utilizar como medidas de asociación

	C_0					$C_{1,0,1}$					$C_{2,0,1}$				
		Media		$10^4 \times \text{ECM}$			Media		$10^4 \times \text{ECM}$			Media		$10^4 \times \text{ECM}$	
ρ	\hat{p}	$\hat{\rho}_1$	$\tilde{\rho}_{\hat{p},1}$	$\hat{\rho}_1$	$\tilde{\rho}_{\hat{p},1}$	\hat{p}	$\hat{\rho}_1$	$\tilde{\rho}_{\hat{p},1}$	$\hat{\rho}_1$	$\tilde{\rho}_{\hat{p},1}$	\hat{p}	$\hat{\rho}_1$	$\tilde{\rho}_{\hat{p},1}$	$\hat{\rho}_1$	$\tilde{\rho}_{\hat{p},1}$
Base de Fourier															
ρ_{CL}	1	0.71	0.71	31	28	2	0.91	0.91	448	451	4	0.89	0.88	393	348
$\rho_{\mathbf{V}_M}$	1	0.72	0.71	40	32	4	0.86	0.83	304	199	4	0.81	0.76	156	61
ρ_{SP}	1	0.73	0.71	53	32	2	0.77	0.75	82	55	3	0.77	0.73	86	38
Base de Splines															
ρ_{CL}	3	0.73	0.69	37	34	3	0.91	0.90	456	443	5	0.90	0.88	398	341
$\rho_{\mathbf{V}_M}$	4	0.74	0.69	50	40	4	0.87	0.83	316	184	5	0.83	0.76	196	65
ρ_{SP}	4	0.75	0.69	59	41	4	0.80	0.76	125	64	5	0.80	0.73	120	45
Base de Componentes Principales															
ρ_{CL}	4	0.73	0.68	37	39	3	0.91	0.91	454	446	4	0.89	0.88	385	353
$\rho_{\mathbf{V}_M}$	4.50	0.75	0.68	57	54	4.50	0.87	0.81	312	161	5	0.82	0.75	186	61
ρ_{SP}	5	0.76	0.68	61	52	4	0.80	0.74	122	50	5	0.79	0.71	108	41

Tab. 5.22: Mediana de la dimensión obtenida \hat{p} mediante convalidación cruzada, media y error cuadrático medio, multiplicado por 10^4 , de los estimadores $\hat{\rho}_1 = \hat{\rho}_{\text{GRID},1}$ y $\tilde{\rho}_{\hat{p},\text{GRID},1} = \tilde{\rho}_{\hat{p},1}$ obtenidos mediante el Algoritmo 2 y distintas medidas de asociación.

ρ_{CL} , $\rho_{\mathbf{V}_M}$ o ρ_{SP} los valores medios obtenidos para el estimador de la primera correlación canónica basados en el Algoritmo 2 son menores que los obtenidos por el Algoritmo 3. Como consecuencia de este hecho, el estimador $\hat{\rho}_{\text{GRID},1}$ resulta más sesgado que $\hat{\rho}_{\text{CRG},1}$ para procesos Gaussianos, en especial para bases de Splines y de componentes principales. Un hecho inverso se observa con $\tilde{\rho}_{\hat{p},\text{GRID},1}$ que resulta menos sesgado que $\tilde{\rho}_{\hat{p},\text{CRG},1}$.

ρ	C_0				$C_{1,0,1}$				$C_{2,0,1}$			
	\hat{p}	\hat{c}_{CRG}	$\hat{\rho}_{X,\text{CRG}}$	AISE	\hat{p}	\hat{c}_{CRG}	$\hat{\rho}_{X,\text{CRG}}$	AISE	\hat{p}	\hat{c}_{CRG}	$\hat{\rho}_{X,\text{CRG}}$	AISE
Base de Fourier												
ρ_{CL}	1	0.99	0.99	0.03	4.50	0.11	0.07	3.56	7	0.12	0.11	3.47
ρ_{OGK}	1	0.96	0.96	0.18	4.50	0.81	0.76	0.75	4	0.88	0.87	0.47
$\rho_{\mathbf{V}_M}$	1	0.98	0.99	0.07	4	0.75	0.62	1	4.50	0.77	0.69	0.94
ρ_{CM}^*	5	0.87	0.91	0.56	5	0.79	0.74	0.82	1.50	0.85	0.86	0.53
ρ_{SP}	1	0.97	0.97	0.10	3	0.89	0.78	0.47	3	0.91	0.85	0.33
Base de Splines												
ρ_{CL}	4	0.94	0.95	0.27	5	0.09	0.06	3.63	8	0.12	0.10	3.51
ρ_{OGK}	4	0.90	0.92	0.40	5	0.77	0.72	0.90	5	0.80	0.78	0.77
$\rho_{\mathbf{V}_M}$	4	0.93	0.94	0.31	4	0.73	0.61	1.07	6	0.73	0.65	1.12
ρ_{CM}^*	7	0.79	0.85	0.88	7	0.73	0.68	1.07	6	0.73	0.74	1.06
ρ_{SP}	4	0.91	0.93	0.35	4	0.87	0.76	0.54	5	0.84	0.77	0.62
Base de Componentes Principales												
ρ_{CL}	5	0.89	0.92	0.44	5	0.09	0.07	3.64	7	0.11	0.10	3.52
ρ_{OGK}	6.50	0.85	0.89	0.62	7	0.72	0.69	1.07	6	0.79	0.78	0.79
$\rho_{\mathbf{V}_M}$	5	0.89	0.91	0.47	5	0.69	0.57	1.20	5	0.71	0.63	1.17
ρ_{CM}^*	8	0.75	0.83	1	8	0.70	0.66	1.18	7	0.66	0.67	1.30
ρ_{SP}	6	0.87	0.91	0.52	6	0.80	0.72	0.77	5	0.82	0.75	0.70

Tab. 5.23: Mediana de la dimensión obtenida \hat{p} mediante convalidación cruzada, AISE y medias sobre replicaciones de $\hat{c}_{\text{CRG}} = |\cos(\theta_{\hat{\Phi}_1, \hat{\Phi}_1})|$ y $\hat{\rho}_{X,\text{CRG}} = \hat{\rho}_{XX}(\Phi_1, \hat{\Phi}_1)$ correspondientes al Algoritmo 3 y distintas medidas de asociación.

ρ	C_0					$C_{1,0,1}$					$C_{2,0,1}$				
	\hat{p}	Media		$10^4 \times \text{ECM}$		\hat{p}	Media		$10^4 \times \text{ECM}$		\hat{p}	Media		$10^4 \times \text{ECM}$	
	\hat{p}	$\hat{\rho}_1$	$\tilde{\rho}_{\hat{p},1}$	$\hat{\rho}_1$	$\tilde{\rho}_{\hat{p},1}$	\hat{p}	$\hat{\rho}_1$	$\tilde{\rho}_{\hat{p},1}$	$\hat{\rho}_1$	$\tilde{\rho}_{\hat{p},1}$	\hat{p}	$\hat{\rho}_1$	$\tilde{\rho}_{\hat{p},1}$	$\hat{\rho}_1$	$\tilde{\rho}_{\hat{p},1}$
Base de Fourier															
ρ_{CL}	1	0.70	0.70	29	29	4.50	0.91	0.91	439	455	7	0.89	0.88	361	348
ρ_{OGK}	1	0.71	0.71	28	28	4.50	0.80	0.78	126	103	4	0.75	0.74	45	41
$\rho_{\mathbf{V}_M}$	1	0.71	0.71	29	29	4	0.84	0.83	232	203	4.50	0.78	0.77	102	81
ρ_{CM}^*	5	0.81	0.76	237	166	5	0.90	0.82	562	346	1.50	0.81	0.76	224	131
ρ_{SP}	1	0.71	0.70	29	29	3	0.76	0.75	58	45	3	0.74	0.73	34	30
Base de Splines															
ρ_{CL}	4	0.69	0.67	33	58	5	0.91	0.91	439	461	8	0.89	0.88	359	344
ρ_{OGK}	4	0.71	0.68	29	50	5	0.80	0.77	131	113	5	0.74	0.72	47	48
$\rho_{\mathbf{V}_M}$	4	0.70	0.68	30	39	4	0.84	0.83	235	204	6	0.78	0.76	99	80
ρ_{CM}^*	7	0.84	0.72	256	212	7	0.93	0.78	589	331	6	0.87	0.71	326	212
ρ_{SP}	4	0.70	0.68	33	49	4	0.76	0.75	60	45	5	0.73	0.71	37	35
Base de Componentes Principales															
ρ_{CL}	5	0.67	0.65	41	77	5	0.91	0.91	438	458	7	0.89	0.88	365	355
ρ_{OGK}	6.50	0.69	0.65	38	84	7	0.80	0.76	116	87	6	0.74	0.70	38	64
$\rho_{\mathbf{V}_M}$	5	0.68	0.66	42	73	5	0.83	0.82	216	174	5	0.78	0.76	99	87
ρ_{CM}^*	8	0.83	0.66	207	231	8	0.93	0.74	559	316	7	0.85	0.67	263	245
ρ_{SP}	6	0.68	0.65	38	66	6	0.75	0.73	44	38	5	0.73	0.70	32	51

Tab. 5.24: Mediana de la dimensión obtenida \hat{p} mediante convalidación cruzada, media y error cuadrático medio, multiplicado por 10^4 , de los estimadores $\hat{\rho}_1 = \hat{\rho}_{\text{GRID},1}$ y $\tilde{\rho}_{\hat{p},\text{GRID},1} = \tilde{\rho}_{\hat{p},1}$ obtenidos mediante el Algoritmo 3 y distintas medidas de asociación.

5.5. Detección de datos atípicos

En esta Sección, se evaluarán tres propuestas para la detección de datos atípicos. El potencial efecto de los datos atípicos al realizar un análisis de datos multivariados ha sido ampliamente estudiado y se han desarrollado diversos métodos que resultan robustos a su influencia, ver por ejemplo Maronna *et al.* (2006). Sin embargo, el estudio de métodos para detección de datos atípicos para datos funcionales es bastante reciente y se ha realizado predominantemente en datos funcionales univariados. En Hubert *et al.* (2015) y en López-Pintado *et al.* (2014) se proponen métodos para la detección de datos atípicos para datos funcionales multivariados. En esta sección propondremos tres métodos para la detección de datos atípicos cuando se tiene una muestra de un elemento aleatorio (X, Y) en el producto de espacios de Hilbert separables $\mathcal{H}_1 \times \mathcal{H}_2$. Este tipo de elementos aleatorios contiene a los datos funcionales bivariados, por lo tanto nuestros métodos entrarían en este grupo. A pesar de no contar con una definición precisa de lo que es un dato atípico, llamaremos observación atípica a una observación que está alejada de la mayoría de las observaciones de la muestra o que puede afectar los resultados de la estimación resultando muy influyentes en esta. En Hubert *et al.* (2015) trabajan con datos funcionales multivariados y dan una taxonomía de los distintos tipos de datos atípicos. Primero distinguen entre curvas atípicas *aisladas* y curvas atípicas *persistentes*. Las aisladas son aquellas que difieren de la mayoría de las curvas en un pequeño intervalo de tiempo. Este tipo de curvas atípicas se puede considerar el análogo a lo que Alqallaf *et al.* (2009) denominan propagación de outliers en el caso de muestras multivariadas $\mathbf{X}_1, \dots, \mathbf{X}_n$ en \mathbb{R}^d en que la probabilidad de error es independiente entre coordenadas. Dicho modelo, cuando la dimensión es grande suele mostrar datos atípicos que difieren de la mayoría por mostrar valores lejanos en una o pocas coordenadas del vector. En el caso de datos funcionales, el efecto nocivo de estos datos atípicos aislados puede amortiguarse realizando previamente un suavizado robusto mediante un M -estimador local basado en núcleos. Las curvas atípicas *persistentes* son las que difieren del resto en la mayor

parte del tiempo o incluso en todo el dominio. Esta clase de curvas atípicas es usual cuando se plantea un modelo de Tukey-Huber de mezcla de distribuciones, como el que planteamos en los escenarios de contaminación C_1 y C_2 . Dentro de este grupo, dichos autores distinguen entre curvas atípicas por *traslación*, por *amplitud* y por *forma*. Llaman datos atípicos por *traslación* a aquellas curvas que tienen la misma forma que la mayoría de las curvas pero están alejadas de estas. Llaman curvas atípicas por *forma* a aquellas cuya forma difiere de la mayoría de las curvas. Dicen que una curva es atípica en *amplitud* cuando puede tener la misma forma que la mayoría de las curvas pero su escala (rango, amplitud) difiere. Claro que puede haber curvas atípicas que muestren una combinación de algunos de los comportamientos atípicos descritos. Vale la pena mencionar que análogamente al caso de vectores aleatorios multivariados, el hecho de que una observación (X_i, Y_i) de una muestra del elemento aleatorio (X, Y) en $\mathcal{H}_1 \times \mathcal{H}_2$ sea atípica no significa que necesariamente X_i será atípico dentro de la muestra X_1, \dots, X_n ni Y_i será atípico dentro de la muestra de Y_1, \dots, Y_n . Tampoco un dato atípico en X o en Y necesariamente será atípico del vector (X, Y) . López-Pintado *et al.* (2014) muestran un ejemplo con las curvas de crecimiento de altura y peso de niños, encontrando un niño que crece poco en altura sin ser atípico su crecimiento en altura y crece mucho en peso sin ser atípico su crecimiento en peso, sin embargo de manera conjunta la observación resulta atípica. Debido a estas cuestiones, los métodos que presentamos se basan en utilizar los estimadores robustos de direcciones y correlaciones canónicas propuestos en la tesis para poder detectar datos que resulten atípicos de manera conjunta y que sean influyentes respecto a estas estimaciones.

Las condiciones de la simulación son las mismas que las consideradas en la Sección 5.4 y la dimensión de la base se elige por convalidación cruzada como en la Sección 5.3. Como es bien sabido las medidas de detección de datos atípicos deben basarse en estimadores robustos para evitar enmascaramiento. Por esta razón, las propuestas descritas en la Sección 5.5.1 y el estudio numérico de la Sección 5.5.4, no incluyen al coeficiente de correlación de Pearson y

a la distancia de correlación ρ_{DIST} . Comenzaremos describiendo los distintos procedimientos de detección de datos atípicos.

5.5.1. Criterios para detección de datos atípicos

Proponemos 3 criterios para la detección de datos atípicos que describiremos a continuación y que evaluaremos mediante un estudio de simulación en la Sección 5.5.4.

5.5.2. Criterios de detección cuando se predice usando el análisis de correlación canónica

Dados vectores aleatorios $\mathbf{X} \in \mathbb{R}^p$ e $\mathbf{Y} \in \mathbb{R}^q$ con media $\mathbf{0}_{\mathbb{R}^p}$ y $\mathbf{0}_{\mathbb{R}^q}$, respectivamente, podemos considerar las proyecciones ortogonales en las primeras direcciones canónicas $\langle \mathbf{X}, \Phi_1 \rangle_{\mathbb{R}^p} \Phi_1$ y $\langle \mathbf{Y}, \Psi_1 \rangle_{\mathbb{R}^q} \Psi_1$ como predictores de \mathbf{X} e \mathbf{Y} , respectivamente. En base a estos predictores definimos los primeros dos criterios de detección de outliers.

Algoritmo 5 Criterio 1 de detección de datos atípicos

- 1: A partir de la muestra $(X_i, Y_i)^T$, $1 \leq i \leq n$, obtenga estimadores robustos de las primeras direcciones canónicas que indicaremos $\hat{\Phi}_1$ y $\hat{\Psi}_1$. Dichos estimadores pueden obtenerse mediante el Algoritmo 2 o el Algoritmo 3.
 - 2: Centre los datos utilizando un estimador robusto de posición. Defina la muestra centrada $X_i^{(c)} = X_i - \hat{\mu}_X$ e $Y_i^{(c)} = Y_i - \hat{\mu}_Y$.
 - 3: Obtenga los predichos de $X^{(c)}$ e $Y^{(c)}$ proyectando en las primeras direcciones canónicas, es decir $\hat{X}_i^{(c)} = \langle X_i^{(c)}, \hat{\Phi}_1 \rangle_{\mathcal{H}_1} \hat{\Phi}_1$ y $\hat{Y}_i^{(c)} = \langle Y_i^{(c)}, \hat{\Psi}_1 \rangle_{\mathcal{H}_2} \hat{\Psi}_1$.
 - 4: Calcule los residuos $R_{X,i}^2 = \|X_i^{(c)} - \hat{X}_i^{(c)}\|_{\mathcal{H}_1}^2$ y $R_{Y,i}^2 = \|Y_i^{(c)} - \hat{Y}_i^{(c)}\|_{\mathcal{H}_2}^2$.
 - 5: Aplique un boxplot ajustado a la muestra $(R_{X,i}^2)_{1 \leq i \leq n}$ y obtenga el grupo de outliers mayores a la mediana al que denominamos \mathcal{G}_X .
Aplique un boxplot ajustado a la muestra $(R_{Y,i}^2)_{1 \leq i \leq n}$ y obtenga el grupo de outliers mayores a la mediana al que denominamos \mathcal{G}_Y .
-

Dentro del procedimiento de detección de outliers dado por el Algoritmo 5 podemos utilizar tres criterios que indicamos en las tablas como se detalla a continuación:

- a) considerar como outliers a los datos en \mathcal{G}_X , indicado AT_X en las Tablas 5.25 y 5.28,
- b) considerar como outliers a los datos en \mathcal{G}_Y , indicado AT_Y en las Tablas 5.25 y 5.28,
- c) considerar como outliers a los datos en la unión $\mathcal{G}_X \cup \mathcal{G}_Y$, indicado como $AT_{X \cup Y}$ en las Tablas 5.25 y 5.28.

Notación 5.5.1. *Nos referiremos a Algoritmo 5 (GRID) o Algoritmo 5 (CRG) dependiendo de si en el paso 1 del algoritmo se utiliza el Algoritmo 2 o el Algoritmo 3, respectivamente, para el cálculo de direcciones canónicas.*

A continuación presentamos el segundo criterio de detección de outliers basado en el bagplot.

Algoritmo 6 Criterio 2 de detección de datos atípicos

- 1: Realice los pasos 1 a 3 del Algoritmo 5 y obtenga $\widehat{X}_i^{(c)}$ e $\widehat{Y}_i^{(c)}$.
 - 2: Haga un bagplot de la muestra bidimensional $(\widehat{X}_i^{(c)}, \widehat{Y}_i^{(c)})$ con factor de dispersión K .
Considere outliers a los que resultan de dicho bagplot.
-

En las simulaciones utilizamos tres factores de dispersión $K = 2, 2,5$ y 3 en el Algoritmo 6, obteniendo 3 criterios distintos de detección de outliers.

Notación 5.5.2. *Nos referiremos a Algoritmo 6 (GRID) o Algoritmo 6 (CRG) dependiendo de si en el paso 1 del algoritmo se utiliza el Algoritmo 5 (GRID) o el Algoritmo 5 (CRG), respectivamente.*

5.5.3. *Criterio de detección con predicción cruzada usando análisis de correlación canónica.*

En Yohai y García Ben (1980) se considera el siguiente problema. Sean vectores aleatorios $\mathbf{X} \in \mathbb{R}^p$ e $\mathbf{Y} \in \mathbb{R}^q$ con media $\mathbf{0}_{\mathbb{R}^p}$ y $\mathbf{0}_{\mathbb{R}^q}$, respectivamente. Dado un vector aleatorio $\mathbf{W} \in \mathbb{R}^s$, indiquemos por $\mathbf{Y}_{\mathbf{W}}^*$ el mejor predictor lineal de \mathbf{Y} basado en \mathbf{W} , es decir, $\mathbf{Y}_{\mathbf{W}}^* = \mathbb{E}(\mathbf{Y}\mathbf{W}^T)\mathbb{E}(\mathbf{W}\mathbf{W}^T)^{-1}\mathbf{W}$. Se intenta encontrar el vector aleatorio $\mathbf{W} = (W_1, \dots, W_s)^T$ donde cada componente es una función lineal de \mathbf{X} tal que, $\mathbf{Y}_{\mathbf{W}}^*$ minimiza el determinante de $\mathbb{E}(\mathbf{Y} - \mathbf{Y}_{\mathbf{W}}^*)(\mathbf{Y} - \mathbf{Y}_{\mathbf{W}}^*)^T$. Yohai y García Ben (1980) muestran que el vector \mathbf{W} que minimiza dicha expresión es el formado por las primeras s variables canónicas en \mathbf{X} . Basados en este hecho, hemos considerado el siguiente Algoritmo para detección de outliers.

Algoritmo 7 Criterio 3 de detección de datos atípicos.

- 1: Aplique los pasos 1 a 4 del Algoritmo 3 y obtenga una muestra de $\tilde{w}_1 = \langle \mathbf{x}^{(c)}, \hat{\mathbf{a}}_1 \rangle$ y $\tilde{\omega}_1 = \langle \mathbf{y}^{(c)}, \hat{\mathbf{b}}_1 \rangle$, u obtenga dichas variables canónicas utilizando el Algoritmo 2.
 - 2: Obtenga el mejor predictor lineal robusto de $\mathbf{y}^{(c)}$ basado en \tilde{w}_1 que indicaremos $\hat{\mathbf{Y}}_{\tilde{w}_1}^*$. Del mismo modo, obtenga el mejor predictor lineal de $\mathbf{x}^{(c)}$ basado en $\tilde{\omega}_1$ que llamaremos $\hat{\mathbf{X}}_{\tilde{\omega}_1}^*$.
 - 3: Calcule los residuos $R_{X,i}^2 = \|\mathbf{x}_i^{(c)} - \hat{\mathbf{X}}_{\tilde{\omega}_1,i}^*\|_{\mathcal{H}_1}^2$ y $R_{Y,i}^2 = \|\mathbf{y}_i^{(c)} - \hat{\mathbf{Y}}_{\tilde{w}_1,i}^*\|_{\mathcal{H}_2}^2$.
 - 4: Realice un boxplot ajustado de la muestra $(R_{X,i}^2)_{1 \leq i \leq n}$ y obtenga el grupo \mathcal{G}_X de outliers mayores a la mediana.
Realice un boxplot ajustado a la muestra $(R_{Y,i}^2)_{1 \leq i \leq n}$ y obtenga el grupo \mathcal{G}_Y de outliers mayores a la mediana.
Aplique un boxplot ajustado a la muestra $(R_{X,i}^2 + R_{Y,i}^2)_{1 \leq i \leq n}$ y obtenga el grupo \mathcal{G}_{X+Y} de outliers mayores a la mediana.
-

Para el método de detección de outliers descrito en el Algoritmo 7 podemos utilizar varios criterios para identificar las observaciones influyentes, que indicaremos en las tablas

como se detalla a continuación:

- a) considerar como outliers el grupo \mathcal{G}_X , indicado AT_X en las Tablas 5.27 y 5.30,
- b) considerar como outliers el grupo \mathcal{G}_Y , indicado AT_Y en las Tablas 5.27 y 5.30,
- c) considerar como outliers el grupo $\mathcal{G}_X \cup \mathcal{G}_Y$, indicado $AT_{X \cup Y}$ en las Tablas 5.27 y 5.30,
- d) considerar como outliers el grupo \mathcal{G}_{X+Y} , indicado AT_{X+Y} en las Tablas 5.27 y 5.30.

Notación 5.5.3. *Nos referiremos a Algoritmo 7 (GRID) o Algoritmo 7 (CRG) dependiendo de si en el paso 1 del algoritmo, para el cálculo de las variables canónicas, se utiliza el Algoritmo 2 o el Algoritmo 3, respectivamente.*

5.5.4. Comparación de los criterios de detección de datos atípicos

Como mencionamos anteriormente, en esta sección presentaremos los resultados obtenidos para los distintos métodos de detección de datos atípicos descritos en la Sección 5.5. Para medir la efectividad del procedimiento se calculó el promedio sobre las $nr = 100$ replicaciones de su sensibilidad y especificidad. La sensibilidad es la proporción de datos atípicos generados que son correctamente identificados como tales, mientras que la especificidad es la proporción de funciones no atípicas que el procedimiento identifica correctamente como no atípicas. Un método ideal debería mantener simultáneamente alta sensibilidad y especificidad. Es claro que para el modelo en que las observaciones eran Gaussianas sólo se puede reportar la especificidad.

Las Tablas 5.25 a 5.27 presentan la especificidad y sensibilidad de los Criterios 1 a 3 de detección de datos atípicos descritos en los Algoritmos 5 a 7, respectivamente, cuando se utiliza el Algoritmo 2 para calcular los estimadores. Para el Criterio 2 se utilizaron tres factores $K = 2, 2,5$ y 3 para la construcción del bagplot.

Para todos los criterios, los niveles de especificidad son buenos en todos los escenarios. Claramente, para los Criterio 1 y 3 la especificidad resulta algo menor cuando se identifica como datos atípicos la unión de los grupos \mathcal{G}_X con \mathcal{G}_Y así como al considerar el procedimiento AT_{X+Y} del Criterio 3. Respecto del Criterio 2, como es de esperar, la especificidad crece con el factor K ya que aumenta la cápsula convexa del bagplot y los valores identificados como atípicos con el bagplot disminuyen. Para los factores $K = 2,5$ y $K = 3$, los niveles de especificidad resultan razonables ya que son mayores al 95 % y resultan similares para las dos medidas de asociación consideradas. Sin embargo, los valores de especificidad del Criterio 2 son inferiores a los obtenidos por los Criterio 1 y 3.

Cuando las muestras tienen datos atípicos, es decir, bajo $C_{1,0,1}$ y $C_{2,0,1}$, las dos medidas de asociación robustas combinadas con los Criterios 1 y 3 tienen alta sensibilidad, independientemente de la base elegida. Para el Criterio 1, el procedimiento basado en la unión de los dos grupos \mathcal{G}_X y \mathcal{G}_Y e indicado como $AT_{X \circ Y}$ parece ser el más adecuado, ya que el lógico descenso en especificidad respecto a las propuestas AT_X y AT_Y es ínfimo en relación al ascenso en sensibilidad respecto a dichas medidas de detección. Un resultado análogo se observa con el Criterio 3 al utilizar las medidas de detección denominada AT_{X+Y} o $AT_{X \circ Y}$ para las bases de Splines y la de Fourier, no ocurriendo lo mismo con la base de Fourier donde la sensibilidad no supera el 70 %. En casi todos los casos, el método $AT_{X \circ Y}$ del Criterio 1 supera en sensibilidad a los métodos AT_{X+Y} o $AT_{X \circ Y}$ del Criterio 3.

Los valores más altos de sensibilidad en el Criterio 1 se obtienen al combinar el método denominado $AT_{X \circ Y}$ con el funcional de correlación ρ_{SP} , para todas las bases consideradas, mientras que el Criterio 3 toma valores muy próximos a 1 con ese funcional de correlación pero para las bases de Splines y de componentes principales tanto con $AT_{X \circ Y}$ como con AT_{X+Y} . A pesar de que para dichos procedimientos la sensibilidad es levemente mejor para las bases de Fourier y Splines, recomendamos el uso de la base de componentes principales por ser adaptiva y más adecuada al análisis de un conjunto de datos reales.

Finalmente, el Criterio 2 que está basado en medidas de correlación robusta, tiene dificultades para detectar curvas atípicas ya que su sensibilidad es baja, por debajo del 40%. La razón de dicho comportamiento puede explicarse por el hecho que, tanto bajo $C_{1,0,1}$ como bajo $C_{2,0,1}$, los outliers correspondientes a X son ortogonales a Φ_1 (similarmente los correspondientes a Y son ortogonales a Ψ_1). Efectivamente, un procedimiento robusto de estimación producirá estimadores $\widehat{\Phi}_1$ y $\widehat{\Psi}_1$ cercanos a Φ_1 y Ψ_1 , respectivamente por lo que los valores $(\widehat{X}_i^{(c)}, \widehat{Y}_i^{(c)})^T$ estarán cercanos a $(0, 0)^T$, correspondiendo a inliers y no pudiendo ser detectados por este procedimiento. Esta medida de detección es útil cuando los datos atípicos corresponden a scores grandes en la dirección de la primera dirección canónica.

En las Tablas 5.28 a 5.30 se pueden encontrar los resultados de especificidad y sensibilidad de los Criterios 1 a 3 de detección de datos atípicos descritos en los Algoritmos 5 a 7, respectivamente, cuando se utiliza el Algoritmo 3 para calcular los estimadores. Como al utilizar el Algoritmo 2, los niveles de especificidad son altos para todos los criterios propuestos, obteniéndose en algunos casos el valor 1 o valores muy cercanos a él. Por otra parte, cuando las muestras tienen datos atípicos, para todas las medidas de asociación robustas, los Criterios 1 y 3 combinan buenos valores de sensibilidad y especificidad, independientemente de la base elegida. Una excepción a este comportamiento es cuando se utiliza la base de Fourier y los procedimientos $AT_{X \circ Y}$ o AT_{X+Y} del Criterio 3, para el que los valores de sensibilidad no superan el 78%. A diferencia de lo observado anteriormente respecto del comportamiento de los estimadores, en este caso, los procedimientos basados en la mediana ρ_{CM}^* dan valores muy altos de sensibilidad al utilizar el método denominado $AT_{X \circ Y}$ del Criterio 1 o los procedimientos $AT_{X \circ Y}$ y AT_{X+Y} del Criterio 3, obteniéndose resultados muy similares a los conseguidos utilizando ρ_{SP} . El procedimiento basado en la medida de asociación ρ_{OGK} presenta asimismo valores altos de sensibilidad en esas situaciones y podría rankearse como el tercer procedimiento en cuanto a su performance en sensibilidad y especificidad. Como para el Algoritmo 2, el procedimiento basado en la unión de los dos grupos \mathcal{G}_X y \mathcal{G}_Y e indicado

como AT_{XOY} combinado con el Criterio 1 parece ser el más adecuado. En este caso, los valores de sensibilidad son semejantes para todas las bases, lo mismo ocurre con el Criterio 3 salvo para la base de Fourier. Por esta razón y teniendo en cuenta que se adapta a los datos a analizar, recomendamos el uso de la base de componentes principales. Respecto del Criterio 2, como se describió anteriormente, este método de detección de datos atípicos no consigue tener valores razonables de sensibilidad ya que sólo puede detectar datos atípicos asociados a altos scores en la dirección de la primera dirección canónica.

5.6. Conclusiones generales de la simulación

Hemos realizado dos estudios de simulación, uno con dimensión de la base fija y el otro cuando dicha dimensión se elige mediante un procedimiento de convalidación cruzada. A su vez, en cada caso hemos generado estimadores en base al Algoritmo 2 y en base al Algoritmo 3. En el primer caso (Tablas 5.2 a 5.13), con ambos algoritmos, se observó que, cuando los datos eran Gaussianos con la misma distribución considerada en He *et al.* (2004), todas las medidas de asociación, salvo la comediana ρ_{CM}^* , daban buenas estimaciones tanto de la primera correlación canónica como de las primeras direcciones canónicas. Respecto de los resultados presentados en He *et al.* (2004), nuestros ECM son algo menores ya que utilizamos muestras de tamaño $n = 100$ cuando en dicho trabajo se consideraron muestras de tamaño $n = 50$. Para las contaminaciones consideradas, las estimaciones basadas en medidas de asociación robustas siguen dando resultados confiables en todos los casos, excepto al utilizar la comediana ρ_{CM}^* . Entre los distintos estimadores robustos, se destaca por su desempeño el estimador que maximiza la correlación de Spearman ρ_{SP} . Para este procedimiento el ECM del estimador de correlación y el AISE de las direcciones canónicas estimadas se mantuvieron muy cercanos a los obtenidos para datos Gaussianos. Por otra parte, si se utilizan la correlación de Pearson ρ_{CL} o la de distancias ρ_{DIST} que son medidas de asociación no robustas,

ρ	C_0			$C_{1,0,1}$						$C_{2,0,1}$					
	Especificidad			Especificidad			Sensibilidad			Especificidad			Sensibilidad		
	AT _X	AT _Y	AT _{XoY}	AT _X	AT _Y	AT _{XoY}	AT _X	AT _Y	AT _{XoY}	AT _X	AT _Y	AT _{XoY}	AT _X	AT _Y	AT _{XoY}
	Base de Fourier														
$\rho_{\mathbf{V}_M}$	0.991	0.990	0.981	0.998	0.999	0.997	0.874	0.871	0.900	0.999	0.999	0.999	0.952	0.924	0.979
ρ_{SP}	0.991	0.991	0.983	0.999	0.999	0.998	0.940	0.957	0.970	1	0.999	0.999	0.992	0.990	0.999
	Base de Splines														
$\rho_{\mathbf{V}_M}$	0.990	0.991	0.981	0.998	0.998	0.997	0.869	0.859	0.900	0.999	0.999	0.998	0.936	0.923	0.970
ρ_{SP}	0.991	0.990	0.981	0.999	1	0.998	0.987	0.994	0.999	0.999	0.999	0.998	0.992	0.974	1
	Base de Componentes Principales														
$\rho_{\mathbf{V}_M}$	0.990	0.990	0.980	0.999	0.999	0.998	0.869	0.860	0.891	0.998	0.999	0.997	0.899	0.901	0.943
ρ_{SP}	0.989	0.992	0.980	0.999	1	0.999	0.958	0.971	0.978	0.999	0.999	0.998	0.976	0.964	0.992

Tab. 5.25: Especificidad y Sensibilidad media sobre repeticiones cuando se usa la medida de detección basada en distancias a valores predichos, Algoritmo 5 (GRID).

ρ		C_0			$C_{1,0,1}$						$C_{2,0,1}$					
		Especificidad			Especificidad			Sensibilidad			Especificidad			Sensibilidad		
	K	2	2,5	3	2	2,5	3	2	2,5	3	2	2,5	3	2	2,5	3
		Base de Fourier														
$\rho_{\mathbf{V}_M}$		0.921	0.974	0.992	0.916	0.966	0.986	0.177	0.001	0	0.927	0.971	0.990	0.392	0.189	0.087
ρ_{SP}		0.920	0.976	0.992	0.915	0.972	0.990	0.017	0	0	0.930	0.978	0.993	0.234	0.115	0.045
		Base de Splines														
$\rho_{\mathbf{V}_M}$		0.916	0.970	0.991	0.918	0.965	0.985	0.228	0.036	0	0.916	0.962	0.985	0.457	0.236	0.104
ρ_{SP}		0.916	0.974	0.993	0.925	0.976	0.993	0.060	0.028	0.010	0.925	0.974	0.993	0.394	0.188	0.066
		Base de Componentes Principales														
$\rho_{\mathbf{V}_M}$		0.912	0.964	0.988	0.916	0.965	0.985	0.325	0.035	0	0.920	0.964	0.987	0.475	0.255	0.112
ρ_{SP}		0.915	0.974	0.992	0.926	0.975	0.993	0.061	0.020	0.011	0.926	0.978	0.993	0.344	0.146	0.052

Tab. 5.26: Especificidad y Sensibilidad media sobre repeticiones cuando se usa la medida de atipicidad basada en bagplot de escores para distintos niveles $K = 2, 2,5$ y 3 de dispersión, Algoritmo 6 (GRID).

ρ	C_0				$C_{1,0,1}$								$C_{2,0,1}$							
	Especificidad				Especificidad				Sensibilidad				Especificidad				Sensibilidad			
	AT_X	AT_Y	AT_{XoY}	AT_{X+Y}	AT_X	AT_Y	AT_{XoY}	AT_{X+Y}	AT_X	AT_Y	AT_{XoY}	AT_{X+Y}	AT_X	AT_Y	AT_{XoY}	AT_{X+Y}	AT_X	AT_Y	AT_{XoY}	AT_{X+Y}
	Base de Fourier																			
ρ_{V_M}	0.994	0.994	0.988	0.994	0.999	0.998	0.996	0.998	0.759	0.732	0.849	0.842	0.998	0.998	0.997	0.996	0.609	0.615	0.674	0.679
ρ_{SP}	0.993	0.994	0.987	0.994	0.999	1.000	0.999	0.999	0.839	0.843	0.865	0.881	0.998	0.998	0.996	0.998	0.500	0.524	0.574	0.583
	Base de Splines																			
ρ_{V_M}	0.992	0.991	0.984	0.992	0.999	0.999	0.998	0.999	0.762	0.795	0.850	0.878	0.999	0.999	0.999	0.999	0.817	0.767	0.896	0.904
ρ_{SP}	0.991	0.992	0.984	0.992	0.999	0.999	0.999	1.000	0.936	0.904	0.978	0.979	0.999	0.999	0.998	0.999	0.846	0.848	0.932	0.959
	Base de Componentes Principales																			
ρ_{V_M}	0.991	0.991	0.983	0.990	0.998	0.997	0.995	0.999	0.669	0.735	0.789	0.782	0.999	0.999	0.998	0.999	0.866	0.865	0.935	0.936
ρ_{SP}	0.990	0.993	0.984	0.989	0.999	0.999	0.997	0.999	0.846	0.834	0.895	0.919	0.999	0.999	0.998	0.999	0.964	0.938	0.992	1.000

Tab. 5.27: Especificidad y Sensibilidad media sobre replicaciones cuando se usa la medida de atipicidad basada en predichos cruzados, Algoritmo 7 (GRID).

ρ	C_0			$C_{1,0,1}$						$C_{2,0,1}$					
	Especificidad			Especificidad			Sensibilidad			Especificidad			Sensibilidad		
	AT _X	AT _Y	AT _{XoY}	AT _X	AT _Y	AT _{XoY}	AT _X	AT _Y	AT _{XoY}	AT _X	AT _Y	AT _{XoY}	AT _X	AT _Y	AT _{XoY}
	Base de Fourier														
ρ_{OGK}	0.989	0.991	0.980	0.998	0.999	0.998	0.952	0.945	0.978	0.999	1	0.998	0.980	0.978	0.999
$\rho_{\mathbf{V}_M}$	0.990	0.992	0.982	0.999	0.999	0.997	0.854	0.843	0.886	0.998	0.999	0.997	0.891	0.884	0.932
ρ_{CM}^*	0.989	0.988	0.978	0.999	0.999	0.998	0.951	0.943	0.995	0.999	1	0.999	0.977	0.963	0.999
ρ_{SP}	0.990	0.992	0.982	0.998	1	0.998	0.981	0.960	1	0.999	1	0.999	0.988	0.971	0.998
	Base de Splines														
ρ_{OGK}	0.990	0.991	0.980	1	1	0.999	0.918	0.942	0.964	0.998	1	0.998	0.962	0.964	0.994
$\rho_{\mathbf{V}_M}$	0.989	0.989	0.979	0.999	0.999	0.998	0.831	0.851	0.877	0.999	0.999	0.998	0.899	0.904	0.943
ρ_{CM}^*	0.989	0.990	0.979	0.999	0.999	0.999	0.936	0.948	1	0.999	0.999	0.999	0.963	0.966	0.990
ρ_{SP}	0.991	0.989	0.980	0.999	0.999	0.999	0.984	0.979	1	0.999	1	0.998	0.981	0.983	0.999
	Base de Componentes Principales														
ρ_{OGK}	0.990	0.991	0.981	0.998	1	0.998	0.938	0.923	0.979	0.998	1	0.998	0.973	0.970	0.998
$\rho_{\mathbf{V}_M}$	0.992	0.990	0.982	0.999	0.999	0.998	0.822	0.810	0.859	0.998	1	0.998	0.879	0.887	0.929
ρ_{CM}^*	0.989	0.987	0.977	0.999	0.999	0.999	0.931	0.957	1	0.998	0.999	0.997	0.951	0.969	0.993
ρ_{SP}	0.989	0.991	0.980	0.999	1	0.998	0.970	0.964	0.997	0.999	0.999	0.998	0.988	0.968	1

Tab. 5.28: Especificidad y Sensibilidad media sobre replicaciones cuando se usa la medida de detección basada en distancias a valores predichos, Algoritmo 5 (CRG).

ρ	K	C_0			$C_{1,0,1}$						$C_{2,0,1}$					
		Especificidad			Especificidad			Sensibilidad			Especificidad			Sensibilidad		
		2	2,5	3	2	2,5	3	2	2,5	3	2	2,5	3	2	2,5	3
		Base de Fourier														
ρ_{OGK}		0.917	0.974	0.992	0.912	0.966	0.988	0.143	0.092	0.043	0.926	0.976	0.992	0.290	0.153	0.065
$\rho_{\mathbf{V}_M}$		0.924	0.975	0.993	0.916	0.971	0.991	0.349	0.164	0.048	0.925	0.974	0.991	0.458	0.278	0.150
ρ_{CM}^*		0.912	0.972	0.992	0.917	0.971	0.990	0.143	0.086	0.031	0.928	0.977	0.993	0.268	0.154	0.089
ρ_{SP}		0.921	0.977	0.994	0.927	0.976	0.992	0.126	0.038	0	0.927	0.979	0.993	0.274	0.127	0.045
		Base de Splines														
ρ_{OGK}		0.905	0.964	0.990	0.912	0.966	0.987	0.206	0.123	0.059	0.923	0.971	0.992	0.389	0.216	0.112
$\rho_{\mathbf{V}_M}$		0.917	0.972	0.993	0.921	0.974	0.993	0.357	0.144	0.049	0.923	0.973	0.993	0.491	0.312	0.178
ρ_{CM}^*		0.910	0.971	0.993	0.922	0.975	0.992	0.256	0.176	0.095	0.926	0.978	0.992	0.415	0.221	0.102
ρ_{SP}		0.918	0.974	0.993	0.928	0.976	0.993	0.108	0.049	0.013	0.927	0.974	0.992	0.397	0.216	0.109
		Base de Componentes Principales														
ρ_{OGK}		0.903	0.965	0.988	0.911	0.964	0.987	0.182	0.123	0.066	0.927	0.977	0.992	0.421	0.243	0.113
$\rho_{\mathbf{V}_M}$		0.915	0.972	0.993	0.924	0.973	0.990	0.370	0.210	0.082	0.926	0.974	0.990	0.570	0.351	0.217
ρ_{CM}^*		0.914	0.970	0.990	0.920	0.972	0.992	0.224	0.146	0.075	0.928	0.974	0.993	0.510	0.345	0.215
ρ_{SP}		0.915	0.973	0.994	0.923	0.976	0.992	0.208	0.107	0.051	0.927	0.978	0.992	0.446	0.261	0.143

Tab. 5.29: Especificidad y Sensibilidad media sobre replicaciones cuando se usa la medida de atipicidad basada en bagplot de escores para distintos niveles K de dispersión, Algoritmo 6 (CRG).

ρ	C_0				$C_{1,0,1}$								$C_{2,0,1}$							
	Especificidad				Especificidad				Sensibilidad				Especificidad				Sensibilidad			
	AT _X	AT _Y	AT _{XoY}	AT _{X+Y}	AT _X	AT _Y	AT _{XoY}	AT _{X+Y}	AT _X	AT _Y	AT _{XoY}	AT _{X+Y}	AT _X	AT _Y	AT _{XoY}	AT _{X+Y}	AT _X	AT _Y	AT _{XoY}	AT _{X+Y}
	Base de Fourier																			
ρ_{OGK}	0.993	0.993	0.987	0.993	0.998	0.998	0.996	0.996	0.636	0.680	0.768	0.765	0.998	0.998	0.996	0.998	0.550	0.525	0.596	0.605
ρ_{V_M}	0.992	0.996	0.989	0.993	0.999	0.998	0.998	0.998	0.643	0.659	0.737	0.749	0.999	0.999	0.998	0.998	0.654	0.641	0.729	0.776
ρ_{CM}^*	0.992	0.990	0.982	0.990	0.998	0.998	0.996	0.998	0.725	0.729	0.800	0.793	0.998	0.997	0.995	0.996	0.419	0.406	0.451	0.473
ρ_{SP}	0.993	0.993	0.987	0.992	0.999	0.999	0.998	0.998	0.738	0.813	0.859	0.857	0.997	0.998	0.995	0.997	0.471	0.456	0.516	0.543
	Base de Splines																			
ρ_{OGK}	0.991	0.993	0.984	0.994	0.999	0.999	0.999	0.999	0.812	0.814	0.889	0.897	0.999	0.999	0.998	0.999	0.856	0.825	0.924	0.938
ρ_{V_M}	0.991	0.993	0.985	0.994	0.999	0.999	0.999	0.999	0.777	0.785	0.859	0.871	0.999	0.999	0.999	0.999	0.807	0.783	0.889	0.904
ρ_{CM}^*	0.990	0.991	0.981	0.990	0.999	0.999	0.998	0.999	0.870	0.855	0.960	0.982	0.999	0.999	0.998	0.999	0.875	0.852	0.950	0.966
ρ_{SP}	0.990	0.992	0.983	0.993	0.999	0.999	0.998	0.999	0.869	0.917	0.961	0.953	1.000	1.000	0.999	0.998	0.826	0.853	0.930	0.944
	Base de Componentes Principales																			
ρ_{OGK}	0.990	0.991	0.982	0.993	0.999	0.999	0.998	0.999	0.716	0.716	0.814	0.807	0.999	0.999	0.998	0.999	0.943	0.945	0.986	0.995
ρ_{V_M}	0.990	0.991	0.981	0.993	0.998	0.998	0.996	0.997	0.656	0.683	0.751	0.752	0.999	0.999	0.999	0.998	0.845	0.863	0.921	0.931
ρ_{CM}^*	0.988	0.990	0.978	0.991	0.998	0.999	0.998	0.999	0.874	0.887	0.989	0.974	1.000	0.999	0.999	0.999	0.949	0.908	0.993	0.989
ρ_{SP}	0.990	0.992	0.982	0.992	0.998	1.000	0.998	0.999	0.791	0.783	0.886	0.874	0.999	0.999	0.998	0.999	0.942	0.946	0.987	0.994

Tab. 5.30: Especificidad y Sensibilidad media sobre replicaciones cuando se usa la medida de atipicidad basada en predichos cruzados, Algoritmo 7 (CRG).

los estimadores de la primera dirección canónica y de la primera correlación canónica se ven fuertemente afectados por la presencia de datos atípicos. El Algoritmo 4, que combina los dos algoritmos mencionados previamente, provee por lo tanto una mejor aproximación a los estimadores definidos en (4.2) que el Algoritmo 2 o 3. Los resultados obtenidos, ver Tablas 5.15 a 5.20, muestra que el comportamiento de los estimadores obtenidos mediante el Algoritmo 4 es muy similar al de los que se obtienen a partir del Algoritmo 2.

Cuando se estima la dimensión de la base mediante un procedimiento de convalidación cruzada semejante al descrito en He *et al.* (2004) (Tablas 5.21 a 5.24), debido al alto costo computacional sólo se realizaron 100 replicaciones en lugar de 1000. Aunque la dimensión se estime en forma adaptiva a partir de las observaciones, los resultados son similares a los conseguidos para dimensión fija si se elige la dimensión p que minimiza el AISE. Los errores cuadráticos medios del estimador de correlación y los valores del AISE de las direcciones estimadas son levemente mayores cuando se utiliza convalidación cruzada, lo que muestra que este método es adecuado para su aplicación práctica. Nuevamente, para dimensiones adaptivas, el mejor comportamiento corresponde a los estimadores basados en la medida de correlación de Spearman. El procedimiento de convalidación cruzada da origen a otro estimador de la primera correlación canónica que indicamos por $\tilde{\rho}_{\hat{p},1}$ y que fue introducido en He *et al.* (2004) bajo el nombre de *Empirical Canonical Correlation*. Este estimador se evaluó y se comparó con las aproximaciones dadas por los distintos algoritmos del estimador $\hat{\rho}_1$ definido en (4.2). Como se observa en las Tablas 5.22 y 5.24, $\tilde{\rho}_{\hat{p},1}$ da resultados aceptables con ambos algoritmos en escenarios con y sin contaminación, aunque el estimador $\hat{\rho}_{\text{CRG},1}$, obtenido mediante el Algoritmo 3, presenta menos sesgo para datos Gaussianos.

Utilizando el método de convalidación cruzada para elegir la dimensión de las bases, se evaluaron tres métodos de detección de datos atípicos propuestos en esta tesis (Tablas 5.25 a 5.30). Para cada método se calculó la media de la especificidad y de la sensibilidad, ésta última sólo en los modelos con contaminación. Cada método a su vez se analizó con

algunas variantes para considerar a los datos como atípicos. El método basado en distancias a predichos dado en el Algoritmo 5 resultó ser el que dio mejores resultados (ver Tablas 5.25 y 5.28). Para todas las medidas de asociación robustas, los valores de especificidad y sensibilidad del Criterio 1 fueron altos, con resultados casi óptimos al utilizar como funcional de asociación el coeficiente de correlación de Pearson ρ_{SP} o la mediana ρ_{CM}^* . El procedimiento basado en la unión de los dos grupos \mathcal{G}_X y \mathcal{G}_Y e indicado como $AT_{X \cup Y}$ parece ser el más adecuado para este criterio. El procedimiento de detección de datos atípicos basado en predicción cruzada usando correlación canónica, descrito en el Algoritmo 7, también dio resultados de especificidad y sensibilidad adecuados (ver Tablas 5.27 y 5.30). Para dicho criterio, se puede apreciar que la sensibilidad resulta más alta al utilizar la propuesta denotada AT_{X+Y} y nuevamente las medidas de asociación que más se destacan son ρ_{CM}^* y ρ_{SP} .

5.7. Apéndice

Queremos ver que para el elemento aleatorio $(X, Y)^T$ construido en (5.1), $\rho_1 = 0,7$, $\rho_2 = 0,3$, $\rho_3 = 0,1$ y que para $1 \leq i \leq 3$ $\Phi_i = \Psi_i = f_i$. Sean $u = \sum_{i=1}^{\infty} a_i f_i$ y $v = \sum_{i=1}^{\infty} b_i f_i$ en S_1 . Como $\mathbb{E}(\langle u, X \rangle_{\mathcal{H}_1}) = \mathbb{E}(\langle v, Y \rangle_{\mathcal{H}_2}) = 0$ vale que

$$\begin{aligned} \text{VAR}(\langle u, X \rangle) &= \mathbb{E}(\langle u, X \rangle^2) = \mathbb{E} \left(\sum_{i=1}^{21} a_i^2 \xi_i^2 \right) = 10 \left(a_1^2 + a_2^2 + a_3^2 + \sum_{i=4}^{21} a_i^2 (0,75)^{i-3} \right) \\ \text{VAR}(\langle v, Y \rangle) &= \mathbb{E}(\langle v, Y \rangle^2) = \mathbb{E} \left(\sum_{i=1}^{21} b_i^2 \zeta_i^2 \right) = 10 \left(b_1^2 + b_2^2 + b_3^2 + \sum_{i=4}^{21} b_i^2 (0,75)^{i-3} \right). \end{aligned}$$

Por otro lado,

$$\begin{aligned} \text{COV}(\langle u, X \rangle_{\mathcal{H}_1}, \langle v, Y \rangle_{\mathcal{H}_2}) \mathbb{E}(\langle u, X \rangle_{\mathcal{H}_1} \langle v, Y \rangle_{\mathcal{H}_2}) &= \mathbb{E} \left[\left(\sum_{i=1}^{21} a_i \xi_i \right) \left(\sum_{i=1}^{21} b_i \zeta_i \right) \right] \\ &= 7a_1b_1 + 3a_2b_2 + a_3b_3, \end{aligned}$$

de donde

$$\begin{aligned}
|\rho_{\text{CL}}(\langle u, X \rangle, \langle v, Y \rangle)| &= \left| \frac{7a_1b_1 + 3a_2b_2 + a_3b_3}{\sqrt{10^2 \left(a_1^2 + a_2^2 + a_3^2 + \sum_{i=4}^{21} a_i^2 (,75)^{i-3} \right)} \left(b_1^2 + b_2^2 + b_3^2 + \sum_{i=4}^{21} b_i^2 (,75)^{i-3} \right)} \right| \\
&\leq \frac{|7a_1b_1 + 3a_2b_2 + a_3b_3|}{10\sqrt{(a_1^2 + a_2^2 + a_3^2)}\sqrt{(b_1^2 + b_2^2 + b_3^2)}}. \tag{5.7}
\end{aligned}$$

Tomando $\tilde{\mathbf{a}} = \lambda_1(a_1, a_2, a_3)$ y $\tilde{\mathbf{b}} = \lambda_2(b_1, b_2, b_3)$ tales que $\tilde{a}_1^2 + \tilde{a}_2^2 + \tilde{a}_3^2 = 1$ y $\tilde{b}_1^2 + \tilde{b}_2^2 + \tilde{b}_3^2 = 1$, vale que

$$\frac{|7a_1b_1 + 3a_2b_2 + a_3b_3|}{10\sqrt{(a_1^2 + a_2^2 + a_3^2)}\sqrt{(b_1^2 + b_2^2 + b_3^2)}} = \frac{|7\tilde{a}_1\tilde{b}_1 + 3\tilde{a}_2\tilde{b}_2 + \tilde{a}_3\tilde{b}_3|}{10(\tilde{a}_1^2 + \tilde{a}_2^2 + \tilde{a}_3^2)^{\frac{1}{2}}(\tilde{b}_1^2 + \tilde{b}_2^2 + \tilde{b}_3^2)^{\frac{1}{2}}},$$

por lo que en (5.7) podemos suponer sin pérdida de generalidad que $a_1^2 + a_2^2 + a_3^2 = b_1^2 + b_2^2 + b_3^2 =$

1. Por lo tanto, obtenemos

$$\begin{aligned}
\frac{|7a_1b_1 + 3a_2b_2 + a_3b_3|}{10\sqrt{(a_1^2 + a_2^2 + a_3^2)}\sqrt{(b_1^2 + b_2^2 + b_3^2)}} &= |0,7a_1b_1 + 0,3a_2b_2 + 0,1a_3b_3| \\
&\leq 0,7|a_1b_1| + 0,3|a_2b_2| + 0,1|a_3b_3| \\
&\leq 0,7 \frac{|a_1b_1|}{|a_1b_1| + |a_2b_2| + |a_3b_3|} + \\
&\quad + 0,3 \frac{|a_2b_2|}{|a_1b_1| + |a_2b_2| + |a_3b_3|} + 0,1 \frac{|a_3b_3|}{|a_1b_1| + |a_2b_2| + |a_3b_3|},
\end{aligned}$$

donde en la última desigualdad, si llamamos $\mathbf{a} := (|a_1|, |a_2|, |a_3|)$ y $\mathbf{b} := (|b_1|, |b_2|, |b_3|)$, usamos que por Cauchy Schwartz, $|\langle \mathbf{a}, \mathbf{b} \rangle| \leq \|\mathbf{a}\| \|\mathbf{b}\| = 1$.

Por lo tanto, obtenemos que $|\rho_{\text{CL}}(\langle u, X \rangle_{\mathcal{H}_1}, \langle v, Y \rangle_{\mathcal{H}_2})|$ es a lo sumo una combinación convexa de los números 0,7, 0,3 y 0,1, de donde $|\rho_{\text{CL}}(\langle u, X \rangle_{\mathcal{H}_1}, \langle v, Y \rangle_{\mathcal{H}_2})| \leq 0,7$. Se consigue la igualdad $\rho_{\text{CL}}(\langle u, X \rangle_{\mathcal{H}_1}, \langle v, Y \rangle_{\mathcal{H}_2}) = 0,7$ sólo si $u = f_1$ y $v = f_1$, salvo eventual cambio de signo. De manera similar, se llega a que $\Phi_2 = \Psi_2 = f_2$, $\rho_2 = 0,3$ y que $\Phi_3 = \Psi_3 = f_3$ y $\rho_3 = 0,1$.

Para el caso en el que se contaminan un 10 % de los datos según el modelo C_1 y se define $(X_{C_1}, Y_{C_1})^T$ como en (5.2), llamamos $\rho_{C_1, i}$ a la i -ésima correlación canónica y $(\Phi_{C_1, i}, \Psi_{C_1, i})$ al i -ésimo par de direcciones canónicas de $(X_{C_1}, Y_{C_1})^T$. Tenemos, como antes, $u = \sum_{i=1}^{\infty} a_i f_i$ y $v = \sum_{i=1}^{\infty} b_i f_i$ en S_1 . Observemos que en este caso

$$\begin{aligned}\mathbb{E}\langle u, X_{C_1} \rangle_{\mathcal{H}_1} &= \mathbb{E}((1-V)\langle u, X \rangle_{\mathcal{H}_1} + a_2 V W) = 25 a_2 \mathbb{E}(V) = 25 a_2 \varepsilon \\ \mathbb{E}\langle v, Y_{C_1} \rangle_{\mathcal{H}_2} &= \mathbb{E}((1-V)\langle v, Y \rangle_{\mathcal{H}_2} + b_2 V W) = 25 b_2 \mathbb{E}(V) = 25 b_2 \varepsilon\end{aligned}$$

mientras que

$$\begin{aligned}\text{VAR}(\langle u, X_{C_1} \rangle_{\mathcal{H}_1}) &= \mathbb{E} \left((1-V) \sum_{i=1}^{21} a_i \xi_i + a_2 V W \right)^2 - (25 a_2 \varepsilon)^2 \\ &= \mathbb{E} \left((1-\varepsilon) \left(\sum_{i=1}^{21} a_i \xi_i \right)^2 + V a_2^2 W^2 \right) - (25 a_2 \varepsilon)^2 \\ &= 9 \left(a_1^2 + a_2^2 + a_3^2 + \sum_{i=4}^{21} a_i^2 (0,75)^{i-3} \right) + a_2^2 \varepsilon \mathbb{E}(W^2) - (25 a_2 \varepsilon)^2.\end{aligned}$$

Como $\mathbb{E}(W^2) = 25^2 + 1$, $\varepsilon = 0,1$ y realizando una cuenta análoga para $\text{VAR}(\langle v, Y_{C_1} \rangle_{\mathcal{H}_2})$, se obtiene que:

$$\begin{aligned}\text{VAR}(\langle u, X_{C_1} \rangle_{\mathcal{H}_1}) &= 9a_1^2 + [9 + 0,1(25^2 + 1) - (2,5)^2] a_2^2 + 9a_3^2 + 9 \sum_{i=4}^{21} a_i^2 (0,75)^{i-3} \\ \text{VAR}(\langle v, Y_{C_1} \rangle_{\mathcal{H}_2}) &= 9b_1^2 + [9 + 0,1(25^2 + 1) - (2,5)^2] b_2^2 + 9b_3^2 + 9 \sum_{i=4}^{21} b_i^2 (0,75)^{i-3}.\end{aligned}$$

Por otro lado, tenemos que

$$\text{COV}(\langle u, X_{C_1} \rangle_{\mathcal{H}_1}, \langle v, Y_{C_1} \rangle_{\mathcal{H}_2}) = 6,3a_1b_1 + 59,05 a_2b_2 + 0,9 a_3b_3,$$

de donde se desprende que

$$|\rho_{\text{CL}}(\langle u, X_c \rangle_{\mathcal{H}_1}, \langle v, Y_c \rangle_{\mathcal{H}_2})| \leq \frac{|6,3a_1b_1 + 59,05a_2b_2 + 0,9a_3b_3|}{\sqrt{9a_1^2 + 65,35a_2^2 + 9a_3^2} \sqrt{9b_1^2 + 65,35b_2^2 + 9b_3^2}}. \quad (5.8)$$

El problema de maximizar la parte derecha de la desigualdad (5.8) es equivalente a encontrar la primera correlación canónica $\tilde{\rho}_1$ del vector aleatorio $(\mathbf{X}, \mathbf{Y}) : \Omega \rightarrow \mathbb{R}^3 \times \mathbb{R}^3$ cuando las matrices de covarianza son $\Sigma_{\mathbf{XY}} = \text{diag}\{6,3, 59,05, 0,9\}$ y $\Sigma_{\mathbf{XX}} = \Sigma_{\mathbf{YY}} = \text{diag}\{9, 65,35, 9\}$. Es bien conocido que $\tilde{\rho}_1^2$ es el primer autovalor de la matriz $\Upsilon = \Sigma_{\mathbf{XX}}^{-1}\Sigma_{\mathbf{XY}}\Sigma_{\mathbf{YY}}^{-1}\Sigma_{\mathbf{YX}}$. En nuestro caso, la matriz Υ es igual a

$$\Upsilon = \begin{pmatrix} 0,49 & 0 & 0 \\ 0 & 0,8165 & 0 \\ 0 & 0 & 0,01 \end{pmatrix}$$

de donde el máximo en (5.8) se alcanza cuando $a_2 = b_2 = 1$ o $a_2 = b_2 = -1$ y además, el valor máximo es

$$\tilde{\rho}_1 = \sqrt{0,8165} = 0,903596.$$

Tomando $u = f_2$ y $v = f_2$ vale que $|\rho_{\text{CL}}(\langle u, X_{C_1} \rangle, \langle v, Y_{C_1} \rangle)| = \tilde{\rho}_1$. Por lo tanto, debido a la desigualdad (5.8), se obtiene que $\tilde{\rho}_1$ es la primera correlación canónica (usando la correlación de Pearson) del elemento aleatorio $(X_{C_1}, Y_{C_1})^T$, esto es, $\tilde{\rho}_1 = \rho_{C_1,1}$ y las primeras direcciones canónicas son $\Phi_{C_1,1} = \Psi_{C_1,1} = f_2$. Con técnicas similares se llega a que $\rho_{C_1,2} = 0,7$ y $\Phi_{C_1,2} = \Psi_{C_1,2} = f_1$ mientras que $\rho_{C_1,3} = 0,1$ y $\Phi_{C_1,3} = \Psi_{C_1,3} = f_3$ y $\rho_{C_1,i} = 0$ para todo $i > 3$.

Usando cálculos análogos se puede mostrar que para el elemento aleatorio $(X_{C_2}, Y_{C_2})^T$ definido en (5.3), las correlaciones y direcciones canónicas están dadas por $\rho_{C_2,1} \approx 0,8848$ y $\Phi_{C_2,1} = \Psi_{C_2,1} = 0,6f_3 + 0,8f_4$.

6. APLICACIÓN A UN CONJUNTO DE DATOS REALES: ESCRITURA DE LETRAS

6.1. Introducción

En este capítulo, se ilustrará, en un conjunto de datos reales, la ventaja de utilizar estimadores robustos de las correlaciones y direcciones canónicas. Así mismo, los métodos de detección de datos atípicos que mejores resultados mostraron en la simulación, permitirán identificar posibles datos influyentes que presentan un comportamiento diferencial.

Los datos a analizar corresponden a la escritura de letras. La muestra completa, extraída de Bache y Lichman (2013), puede encontrarse en <https://archive.ics.uci.edu/ml/datasets/Character+Trajectories> y consta de 2858 observaciones. Cada observación de la muestra corresponde a la velocidad en el eje horizontal, la velocidad en el eje vertical y la fuerza de presión de la punta de una lapicera sobre una tableta WACOM cuando un participante del experimento escribe alguna entre 20 letras del abecedario. Este conjunto de datos fue estudiado por Hubert *et al.* (2016) para ilustrar el buen desempeño del procedimiento de clasificación basado en profundidades que estos autores proponen. En este capítulo, consideraremos los 186 elementos de la muestra que corresponden a la escritura de la letra “e” y solamente utilizamos las velocidades en el eje horizontal y vertical, indicadas X e Y , respectivamente como los datos funcionales.

Uno de los objetivos de este capítulo es evaluar si los criterios para detectar datos atípicos

presentados en la Sección 5.5 permiten efectivamente detectar datos que resulten influyentes en la estimación de la primera correlación y las primeras direcciones canónicas y puedan considerarse como atípicos. También, nos interesa constatar si efectivamente los estimadores que mostraron propiedades de robustez en la simulación siguen resultando fiables para estimar las primeras direcciones canónicas y la primera correlación canónica.

6.2. Métodos utilizados y resultados

Consideramos que las observaciones corresponden a 186 realizaciones de un elemento aleatorio $(X, Y)^T : \Omega \rightarrow L^2([0, 1]) \times L^2([0, 1])$, donde $X(t)$ es la velocidad en el eje horizontal en función del tiempo de una persona que escribe la letra “e” e $Y(t)$ es la velocidad en el eje vertical, donde t varía en $[0, 1]$ tomando como $t = 0$ el momento en que comienza a escribir la letra y $t = 1$ el momento en que termina de escribirla. A través de una interpolación de la muestra original, obtenemos los datos $(X_i(t_j), Y_i(t_j))$ para valores de t_j en una grilla equiespaciada de 111 elementos. En este caso, $t_j = 1/(2 \cdot 111) + (j-1)/111$, con $j = 1, \dots, 111$.

Como es usual en este tipo de análisis, evaluaremos el comportamiento de los estimadores clásicos basados en el coeficiente de correlación de Pearson ρ_{CL} y definidos por (4.2), con el de los estimadores obtenidos mediante un procedimiento resistente a datos atípicos. En base a los resultados obtenidos en el Capítulo 5, tomamos como estimador robusto el estimador basado en el coeficiente de correlación de Spearman ρ_{SP} . Los estimadores asociados a los coeficientes de correlación de Pearson y Spearman se indicarán por $\hat{\rho}_{CL,1}$, $(\hat{\Phi}_{CL,1}, \hat{\Psi}_{CL,1})$ y $\hat{\rho}_{SP,1}$, $(\hat{\Phi}_{SP,1}, \hat{\Psi}_{SP,1})$, respectivamente. Los estimadores se calcularon utilizando el Algoritmo 2 que fue el que resultó aproximar mejor los verdaderos estimadores definidos en (4.2).

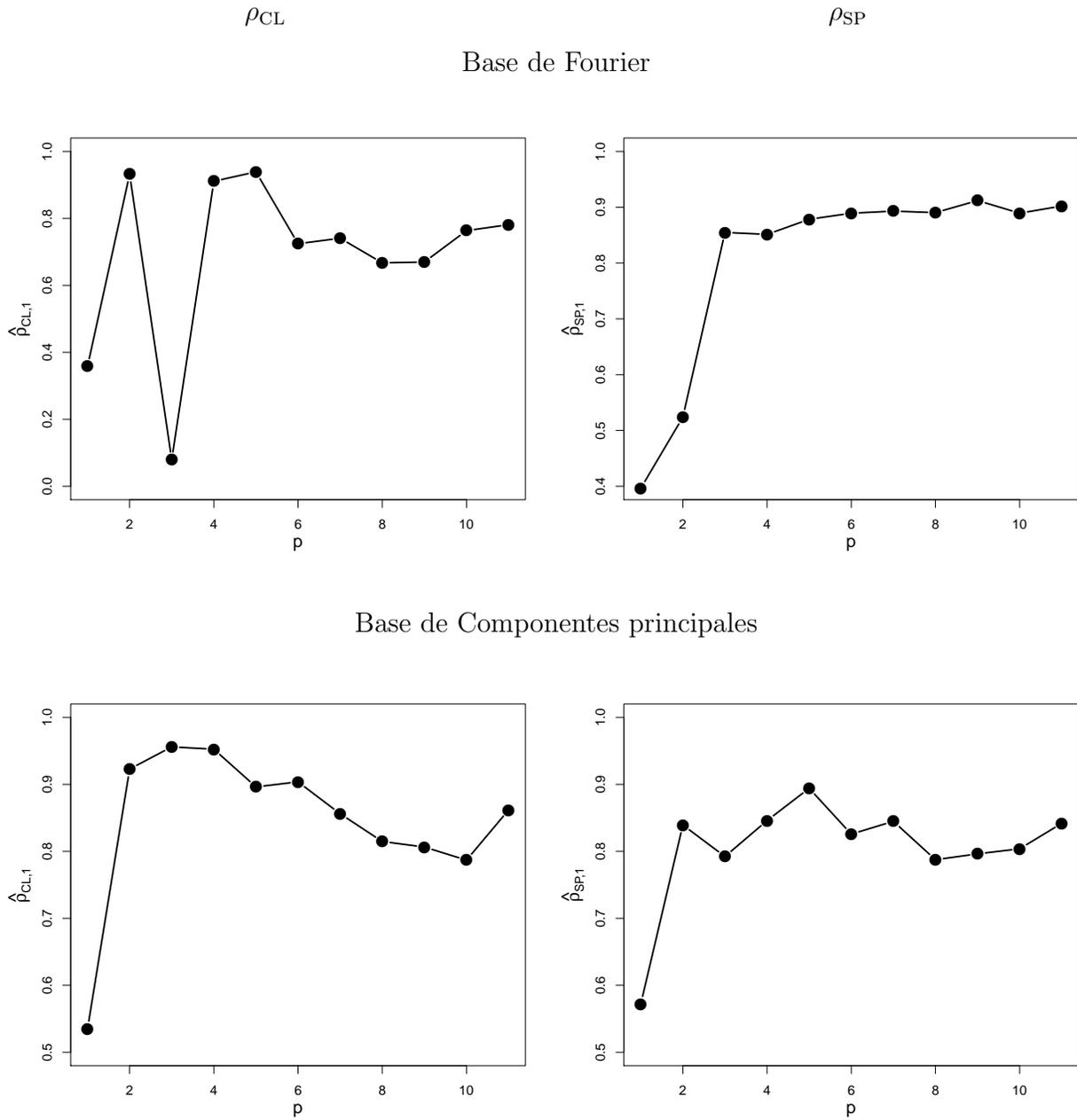
Se consideraron dos bases en $L^2([0, 1])$, una base fija, la base de Fourier, y una base adaptiva, la de componentes principales. Como en el Capítulo 5, las componentes principales se calcularon como las autofunciones del operador de covarianza muestral, cuando se calculaban

los estimadores basados en la correlación de Pearson ρ_{CL} , mientras que para el coeficiente de correlación de Spearman ρ_{SP} , se utilizaron las componentes principales esféricas introducidas por Locantore *et al.* (1999). Si bien los datos no son periódicos en realidad, se utilizó la base de Fourier pues no se introducen efectos espurios en nuestros resultados asumiendo periodicidad.

Para calcular los estimadores y las medidas de detección, elegimos la dimensión de los subespacios aproximantes por el criterio de convalidación cruzada descrito en la Sección 5.3. Recordemos que este procedimiento maximizaba $RCV_{\mathbf{r}_n,1}$, definida en (5.5), en un conjunto \mathcal{R} de posibles valores $\mathbf{r}_n = (p_n, q_n)$. Por simplicidad numérica, realizamos la maximización sobre el conjunto de puntos de la forma $p = q$. Tanto para la base de Fourier como para la base de componentes principales, tomamos $\mathcal{R} = \{(j, j), 1 \leq j \leq 11\}$. La Figura 6.1 presenta el gráfico de la función $\sqrt{RCV(p, 1)} = \sqrt{RCV(\mathbf{r}_n, 1)}$ como función de la dimensión p cuando la base es la base de Fourier. Podemos observar que cuando se utiliza la correlación de Spearman, la función $RCV(p, 1)$ es estable a partir de 5, de hecho el valor máximo de $RCV(p, 1)$ es igual a 0.912 mientras que $RCV(6, 1) = 0,889$, es decir una diferencia relativa del 2.5 %, por lo que tomar $p = 6$ podría dar también resultados confiables. Por otro lado, la función de convalidación cruzada $RCV(p, 1)$ en el caso del coeficiente de Pearson es mucho más irregular y este comportamiento se mantiene aún cuando eliminamos los datos detectados como influyentes en el análisis que sigue. Respecto de la base de componentes principales y el coeficiente de Spearman, la diferencia relativa entre $RCV(\hat{p}, 1)$ y los demás valores de la función supera siempre al 6 %.

Al efectuar el análisis, se observó que las direcciones obtenidas utilizando el coeficiente de correlación de Pearson y de Spearman mostraban diferencias. También se observaron diferencias en los estimadores de los coeficientes de correlación obtenidos con estas dos medidas de asociación. Por esta razón, se calculó el ángulo entre las direcciones clásicas y robustas. Se obtuvo que el coseno del ángulo entre $\hat{\Phi}_{CL,1}$ y $\hat{\Phi}_{SP,1}$, indicado $\cos(\theta_{\hat{\Phi}_{CL,1}, \hat{\Phi}_{SP,1}})$, era igual a

Fig. 6.1: Gráfico de la función $\sqrt{RCV(p, 1)}$ como función de p y dos medidas de asociación, el coeficiente de Pearson y de Spearman.



0.43 al utilizar bases de Fourier y 0.78 cuando se usaba la base de componentes principales y las dimensiones obtenidas por convalidación cruzada. Asimismo, el coseno del ángulo entre las direcciones canónicas correspondientes a la variable Y , $\cos(\theta_{\hat{\Psi}_{CL,1}, \hat{\Psi}_{SP,1}})$, tomaba el valor 0.27 al utilizar bases de Fourier y 0.87 para la base de componentes principales. Este hecho parece indicar que algunas observaciones atípicas podrían estar presentes en la muestra y que la base de Fourier no provee una buena base para aproximar las direcciones canónicas. Un resultado análogo se observó cuando para los estimadores basados en el coeficiente de Spearman y la base de Fourier se utilizó dimensión $p = 6$.

Para detectar esas posibles observaciones influyentes, se utilizaron los procedimientos de detección de datos atípicos descritos en la Sección 5.5. De los métodos allí presentados, se eligieron los Criterios 1 y 3 dados por los Algoritmos 5 y 7, utilizando las bases de componentes principales esféricas y la base de Fourier. Para el Criterio 1 se utilizó la medida de atipicidad AT_{XoY} mientras que para el Criterio 3 se usó AT_{X+Y} ya que estas elecciones fueron las más efectivas en el estudio de simulación realizado. Como los estimadores robustos se calcularon con el coeficiente de correlación de Spearman y las medidas basadas en ρ_{SP} mostraron un buen balance de sensibilidad y especificidad en la Sección 5.5.4, se utilizó esta medida de asociación para efectuar la detección de datos atípicos. Los datos detectados como atípicos se presentan en la Tabla 6.1, donde también se presentan los valores de la dimensión \hat{p} elegida para las bases mediante convalidación cruzada. Las observaciones 33, 38, 139 y 175 son detectados como atípicas con ambas bases y ambos criterios, mientras que el Criterio 1 detecta algunas observaciones adicionales cuando se usa la base de componentes principales.

Para evaluar la influencia de estos datos en el estimador basado en la correlación de Pearson, se calcularon los estimadores clásicos cuando se eliminan las observaciones con índices en alguno de los siguientes grupos $\mathcal{I}_1 = \{33, 38, 139, 175\}$ y $\mathcal{I}_2 = \{7, 33, 38, 113, 139, 154, 175, 137, 140\}$. Observemos que como el Criterio 1 resultó con mayor sensibilidad que el Criterio 3, no consideramos el subconjunto $\{38, 139, 175\}$ identificado por el Criterio 3 cuando se usa la

Base de Fourier		
\hat{p}	Método	Índice de la observación detectada
9	Criterio 1 AT_{XoY}	33, 38, 139, 175
9	Criterio 3 AT_{X+Y}	33, 38, 139, 175
6	Criterio 1 AT_{XoY}	33, 38, 139, 175
6	Criterio 3 AT_{X+Y}	38, 139, 175
Base de Componentes Principales		
\hat{p}	Método	Índice de la observación detectada
5	Criterio 1 AT_{XoY}	7, 33, 38, 113, 139, 154, 175, 137, 140
5	Criterio 3 AT_{X+Y}	33, 38, 139, 175

Tab. 6.1: Datos atípicos detectados mediante los distintos criterios usando el Algoritmo GRID.

base de Fourier y dimensión $p = 6$. Los estimadores así obtenidos se indicarán como $\hat{\rho}_{CL,1}^{-\mathcal{I}_j}$, $\hat{\Phi}_{CL,1}^{-\mathcal{I}_j}$ y $\hat{\Psi}_{CL,1}^{-\mathcal{I}_j}$ en las Tablas 6.3 y 6.4. Vale la pena observar que a diferencia del procedimiento basado en bases de Fourier, para los estimadores de Sieves basados en las bases de componentes principales clásicas, fue necesario en cada situación recalcular los elementos de dicha base.

En la Tabla 6.2, se presentan los valores obtenidos para los estimadores de la primera correlación canónica, junto con la dimensión de los espacios aproximantes elegidos mediante convalidación cruzada. Se reportan los valores obtenidos con el coeficiente de correlación de Spearman y de Pearson cuando se usan todos los datos y los valores obtenidos utilizando ρ_{CL} cuando se eliminan las observaciones detectadas por los Criterio 1 y 3, es decir, los valores $\hat{\rho}_{CL,1}^{-\mathcal{I}_j}$, para $j = 1, 2$. Para el caso del coeficiente de Spearman y base de Fourier, se reportan los valores cuando $p = 6$ y cuando $p = 9$ que es el valor que maximiza $RCV(p, 1)$.

Observemos que, al utilizar bases de Fourier, las estimaciones clásicas luego de quitar los datos detectados como atípicos, $\hat{\rho}_{CL,1}^{-\mathcal{I}_j}$ y $\hat{\rho}_{CL,\hat{p},1}^{-\mathcal{I}_j}$, se acercan a las estimaciones robustas

Indices de las observaciones eliminadas	ρ	Base de Fourier			Base de Comp. Princ.		
		\hat{p}	$\hat{\rho}_1$	$\tilde{\rho}_{\hat{p},1}$	\hat{p}	$\hat{\rho}_1$	$\tilde{\rho}_{\hat{p},1}$
—	ρ_{CL}	2	0.62	0.96	3	0.91	0.96
—	ρ_{SP}	6	0.93	0.89	5	0.92	0.89
—	ρ_{SP}	9	0.94	0.91	5	0.92	0.89
\mathcal{I}_1	ρ_{CL}	4	0.91	0.94	4	0.89	0.92
\mathcal{I}_2	ρ_{CL}	5	0.91	0.89	4	0.88	0.90

Tab. 6.2: Estimación de la primera correlación canónica ρ_1 utilizando convalidación cruzada con todos los datos y cuando se eliminan las observaciones cuyos índices pertenecen a los grupos $\mathcal{I}_1 = \{33, 38, 139, 175\}$ y $\mathcal{I}_2 = \{7, 33, 38, 113, 139, 154, 175, 137, 140\}$.

con la muestra completa, $\hat{\rho}_{SP,1}$ y $\tilde{\rho}_{SP,\hat{p},1}$, respectivamente. Al utilizar bases de componentes principales, se observa el mismo fenómeno para el estimador $\tilde{\rho}_{\hat{p},1}$, que muestra valores mayores y cercanos a uno cuando se consideran todos los datos. Sin embargo, las estimaciones $\hat{\rho}_{CL,1}$ y $\hat{\rho}_{SP,1}$ son muy parecidas aunque se tome el conjunto total de datos al calcular el estimador de Pearson. Recordemos que de acuerdo a nuestro estudio de simulación el coeficiente $\tilde{\rho}_1$ era menos sesgado que $\hat{\rho}_1$, en especial, al utilizar la base de componentes principales.

Las Tablas 6.3 y 6.4 muestran resultados concernientes a las estimaciones de las direcciones canónicas. Se presentan los valores del coseno del ángulo entre las direcciones robustas y las direcciones clásicas calculadas con toda la muestra y sacando las observaciones con índices en \mathcal{I}_j , $j = 1, 2$. Se dan además las correlaciones de las variables canónicas clásicas y robustas obtenidas a partir de dichas direcciones. Las correlaciones se calcularon usando el coeficiente de correlación de Spearman, es decir, se reporta para dos direcciones (u, v) el valor $\hat{\rho}_{XX}(u, v)$ corresponde a $\hat{\rho}_{XX}(u, v) = \rho_{SP}(P_n[\langle u, X \rangle, \langle v, X \rangle])$. En base a los resultados observados en la simulación respecto del comportamiento de los estimadores cuando p es grande y teniendo en cuenta que, en el caso de la base de Fourier, al aumentar la dimensión

de la base se logra claramente un mejor ajuste a los datos pero al mismo tiempo las trayectorias de las direcciones se vuelven más irregulares pues incorporan frecuencias más altas, las comparaciones que haremos a partir de ahora para los estimadores basados en la base de Fourier y el coeficiente de Spearman estarán basadas en $p = 6$. Recordemos que en este caso, la función $RCV(p, 1)$ era estable a partir de $p = 6$ y los valores de $RCV(6, 1)$ y $RCV(\hat{p}, 1)$ eran muy parecidos con una diferencia relativa del 2.5%. Para resaltar que tomamos $p = 6$ en lugar de \hat{p} , en las Tablas 6.3 y 6.4 indicaremos a $\hat{\Phi}_{SP,1}$ y $\hat{\Psi}_{SP,1}$ por $\hat{\Phi}_{SP,1,p=6}$ y $\hat{\Psi}_{SP,1,p=6}$, respectivamente.

$\hat{\rho}_{XX}(\cdot, \cdot)$	$\hat{\Phi}_{CL,1}$	$\hat{\Phi}_{CL,1}^{-\mathcal{I}_1}$	$\hat{\Phi}_{CL,1}^{-\mathcal{I}_2}$	$ \cos(\theta) $	$\hat{\Phi}_{CL,1}$	$\hat{\Phi}_{CL,1}^{-\mathcal{I}_1}$	$\hat{\Phi}_{CL,1}^{-\mathcal{I}_2}$
	Base de Fourier				Base de Fourier		
$\hat{\Phi}_{SP,1,p=6}$	0.63	0.99	0.97	$\hat{\Phi}_{SP,1,p=6}$	0.56	0.95	0.87
	Base de CP				Base de CP		
$\hat{\Phi}_{SP,1}$	0.95	0.97	0.97	$\hat{\Phi}_{SP,1}$	0.79	0.86	0.85

Tab. 6.3: Estimación de correlaciones entre variables canónicas obtenidas con las bases de Fourier y de Componentes Principales (CP) y ángulos θ entre los estimadores de la primera dirección canónica en el espacio X . La correlación se mide mediante el coeficiente de correlación de Spearman. $\hat{\Phi}_{CL,1}^{-\mathcal{I}_j}$ indica el estimador clásico obtenido eliminando las observaciones con índices en \mathcal{I}_j , donde $\mathcal{I}_1 = \{33, 38, 139, 175\}$ y $\mathcal{I}_2 = \{7, 33, 38, 113, 139, 154, 175, 137, 140\}$.

$\hat{\rho}_{YY}(\cdot, \cdot)$	$\hat{\Psi}_{CL,1}$	$\hat{\Psi}_{CL,1}^{-\mathcal{I}_1}$	$\hat{\Psi}_{CL,1}^{-\mathcal{I}_2}$	$ \cos(\theta) $	$\hat{\Psi}_{CL,1}$	$\hat{\Psi}_{CL,1}^{-\mathcal{I}_1}$	$\hat{\Psi}_{CL,1}^{-\mathcal{I}_2}$
	Base de Fourier				Base de Fourier		
$\hat{\Psi}_{SP,1,p=6}$	0.56	0.99	0.97	$\hat{\Psi}_{SP,1,p=6}$	0.49	0.93	0.87
	Base de CP				Base de CP		
$\hat{\Psi}_{SP,1}$	0.98	0.98	0.98	$\hat{\Psi}_{SP,1}$	0.87	0.87	0.89

Tab. 6.4: Estimación de correlaciones entre variables canónicas obtenidas con las bases de Fourier y de Componentes Principales (CP) y ángulos θ entre los estimadores de la primera dirección canónica en el espacio Y . La correlación se mide mediante el coeficiente de correlación de Spearman. $\hat{\Psi}_{CL,1}^{-\mathcal{I}_j}$ indica el estimador clásico obtenido eliminando las observaciones con índices en \mathcal{I}_j , donde $\mathcal{I}_1 = \{33, 38, 139, 175\}$ y $\mathcal{I}_2 = \{7, 33, 38, 113, 139, 154, 175, 137, 140\}$.

En la Tabla 6.3 observamos que, al considerar las estimaciones basadas en todos los datos y la base de Fourier, la correlación entre las primeras variables canónicas en X y el coseno del ángulo entre las direcciones canónicas clásica y robusta toman valores cercanos a 0.60, lo que muestra la diferencia entre ambos estimadores. Por otra parte, la variable canónica $\langle \hat{\Phi}_{SP,1}, X \rangle$ tiene una alta correlación empírica (superior a 0.95) con la variable canónica clásica cuando se eliminan los grupos de observaciones detectadas como atípicas. Se observa asimismo un crecimiento en el coseno del ángulo entre $\hat{\Phi}_{SP,1}$ y los estimadores clásicos luego de eliminar los datos detectados como atípicos, $\hat{\Phi}_{CL,1}^{-\mathcal{I}_j}$.

Un fenómeno diferente se observa al considerar la base de componentes principales. En este caso, la correlación entre las primeras variables canónicas en X estimadas ambas con la muestra completa resulta alta ($\hat{\rho}_{XX}(\hat{\Phi}_{CL,1}, \hat{\Phi}_{SP,1}) \approx 0,95$), aunque el coseno del ángulo correspondería a un ángulo cercano a los 38° . Al eliminar observaciones atípicas, la correlación crece levemente ($\hat{\rho}_{XX}(\hat{\Phi}_{CL,1}^{-\mathcal{I}_j}, \hat{\Phi}_{SP,1}) \approx 0,97$) y el coseno del ángulo también resultando correspondiendo a un ángulo de aproximadamente 30° .

La Tabla 6.4 muestra un comportamiento de las variables canónicas en Y similar al

descrito en la Tabla 6.3. Sin embargo, en este caso, al utilizar las bases de componentes principales la correlación entre las variables canónicas en Y se mantiene prácticamente igual al eliminar los datos atípicos que cuando el estimador basado en el coeficiente de correlación de Pearson se calcula con todos los datos. Por otra parte, los ángulos cambian levemente y al eliminar los datos atípicos del grupo \mathcal{I}_2 ambos estimadores se vuelven más parecidos.

Los resultados obtenidos son consistentes con lo esperado, es decir, los estimadores basados en el coeficiente de correlación de Spearman por ser resistente a datos atípicos da estimaciones similares a las obtenidas con el coeficiente de correlación de Pearson después de haber eliminado las observaciones detectadas como atípicas.

En la Figura 6.2 se presentan las curvas de las direcciones canónicas estimadas. Nuevamente, en el caso de la base de Fourier se presentan los estimadores $\widehat{\Phi}_{\text{SP},1,p=6}$ y $\widehat{\Psi}_{\text{SP},1,p=6}$. Se observa como, para ambas bases, los estimadores clásicos calculados sin las observaciones de los grupos \mathcal{I}_1 e \mathcal{I}_2 se parecen más a las curvas estimadas de manera robusta con la muestra completa que las direcciones canónicas clásicas con todos los datos.

En la Figura 6.3 se resaltan en negrita las trayectorias de los grupos \mathcal{I}_1 y \mathcal{I}_2 . Los demás datos se dibujan en gris. El grupo de observaciones \mathcal{I}_1 es el grupo de observaciones más alejadas de la región en la que se encuentran la mayoría de las curvas, tanto las de la velocidad en el eje horizontal como en el vertical. Asimismo, las curvas que corresponden a \mathcal{I}_2 y no a \mathcal{I}_1 parecen tener un defasaje temporal en el sentido que alcanzan sus máximo y mínimos en tiempos algo alejados respecto del tiempo en que dicho máximo o mínimo es alcanzado para la mayoría de las trayectorias. En particular, dos de las trayectorias de \mathcal{I}_1 , las correspondientes a las observaciones identificadas como 139 y 175, presentan un comportamiento netamente diferente en el intervalo $[0,8, 1]$, en particular en el eje vertical, además del desplazamiento temporal en los máximos y mínimos mencionados. Por otra parte, el máximo de $X(t)$ en el intervalo $[0, 0,5]$ correspondiente a la observación 38 supera claramente al del resto de las trayectorias y se alcanza cerca de $t = 0,3$ mientras que para la mayoría de los datos se

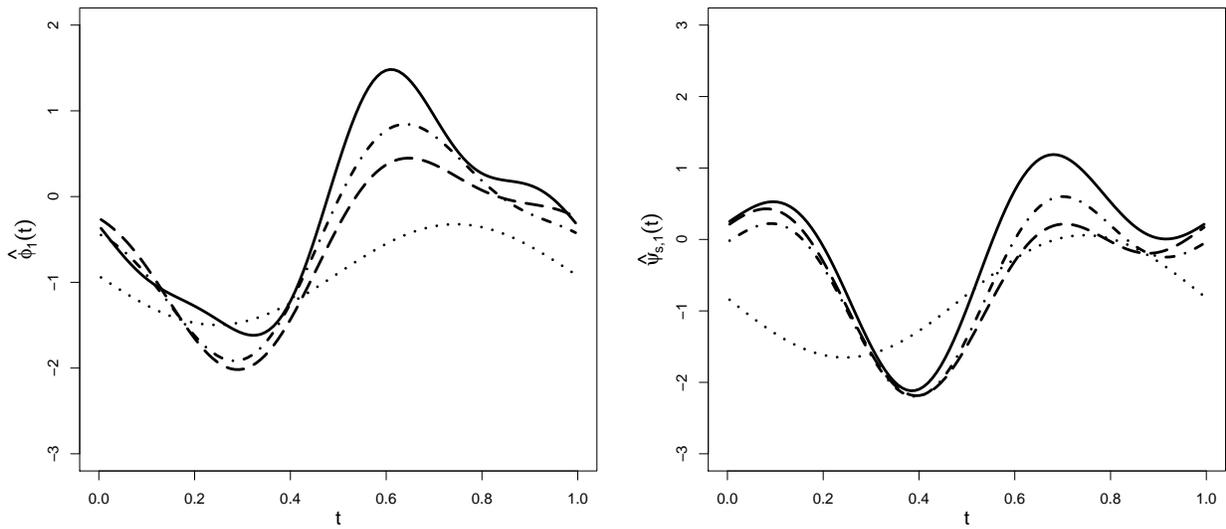
alcanza en valores de tiempo cercanos a 0.2.

En las Figuras 6.4 y 6.5 individualizamos las observaciones atípicas del grupo \mathcal{I}_1 junto al resto de las curvas, permitiendo distinguir que aquellas curvas que se alejan severamente del común de las curvas no son otras que las del grupo \mathcal{I}_1 .

En la Figura 6.6, se presenta el dibujo de la letra que corresponde a cada observación del grupo \mathcal{I}_1 , construido a partir de integrar los datos de las velocidades en cada eje. Denominamos a las posiciones en función del tiempo en los ejes horizontal y vertical $Z_1(t)$ y $Z_2(t)$, respectivamente. Agregamos además una referencia temporal del subíndice j de los tiempos t_j . Esto permite ver como dibujaron la letra “e” las personas cuyas observaciones fueron detectadas como atípicas y confirmar que el método sirvió para detectar las letras “e” dibujadas de una forma distinta al común de la muestra cuyas trayectorias corresponden al de una letra “e” imprenta.

Fig. 6.2: Gráficos de direcciones estimadas: Se grafican con línea rellena la estimación robusta, con línea punteada la estimación clásica con muestra completa, con línea de puntos y guiones la estimación clásica quitando \mathcal{I}_1 y con línea de guiones la estimación clásica quitando \mathcal{I}_2

Utilizando bases de Fourier



Utilizando bases de componentes principales

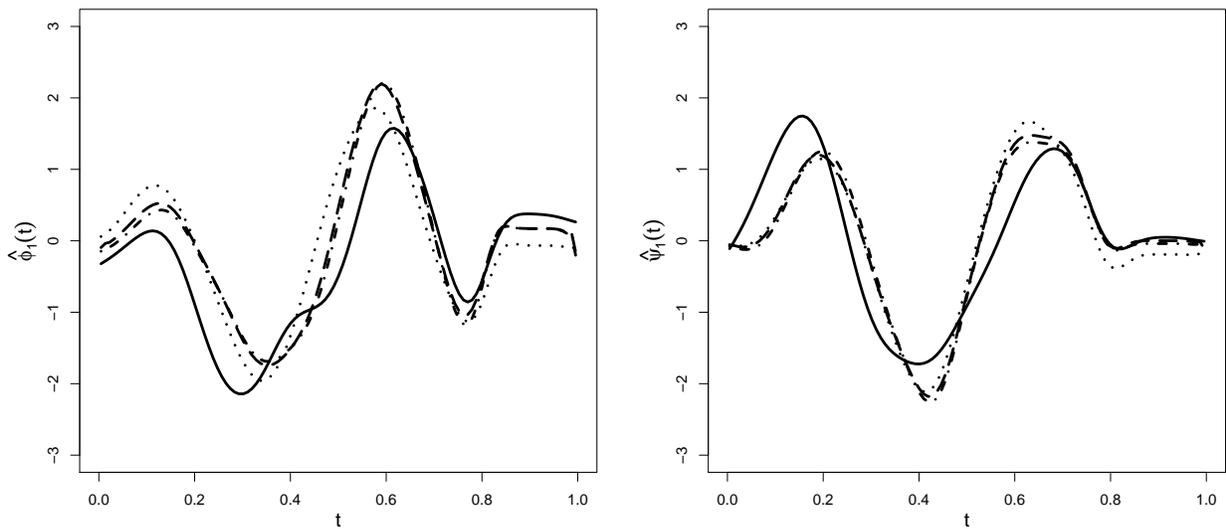


Fig. 6.3: Gráficos de velocidad: Las curvas negras indican las observaciones atípicas de los grupos \mathcal{I}_j , $j = 1, 2$.

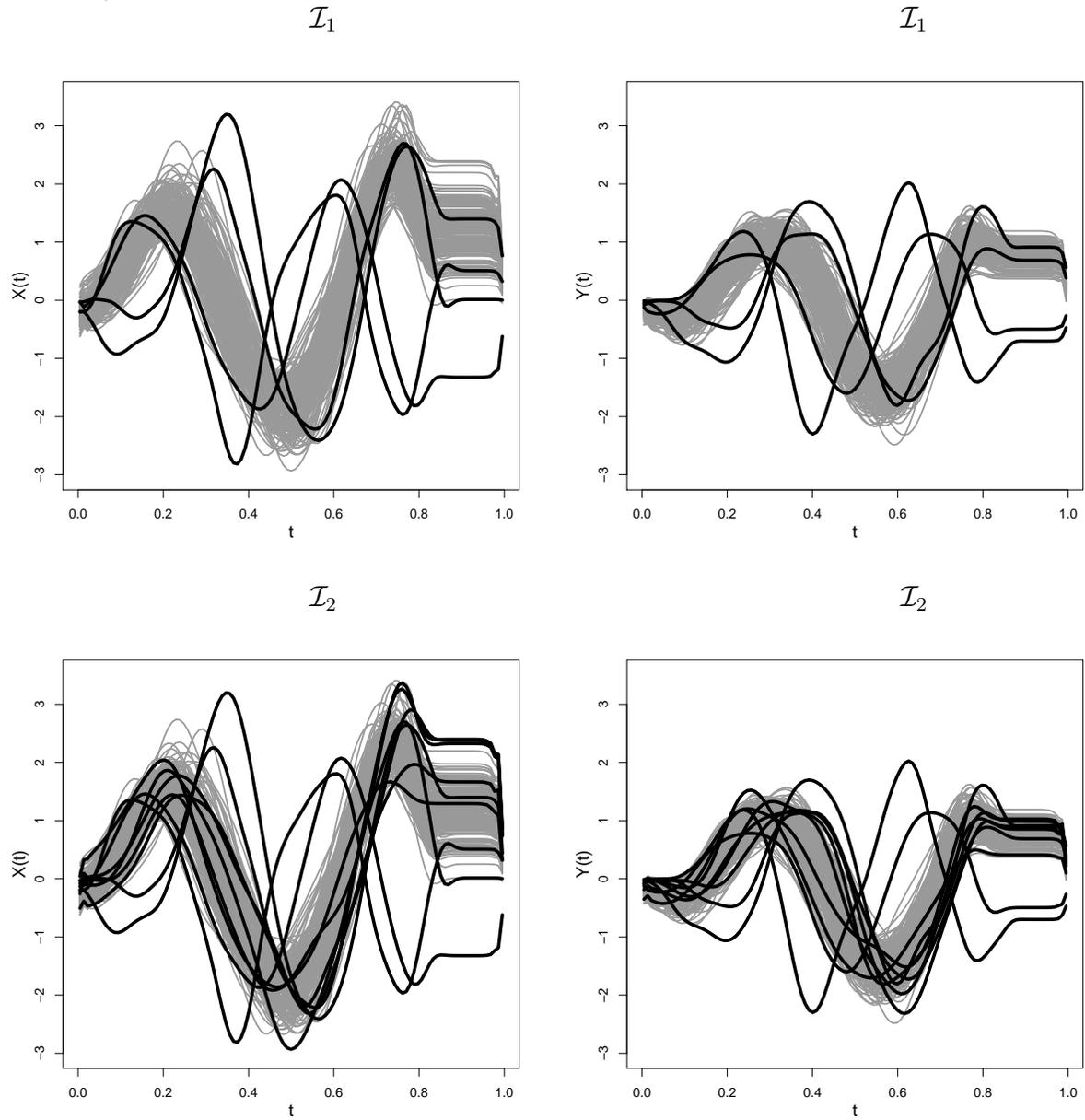


Fig. 6.4: Gráficos de velocidad en el eje horizontal: Las curvas negras indican las observaciones atípicas del grupo \mathcal{I}_1 .

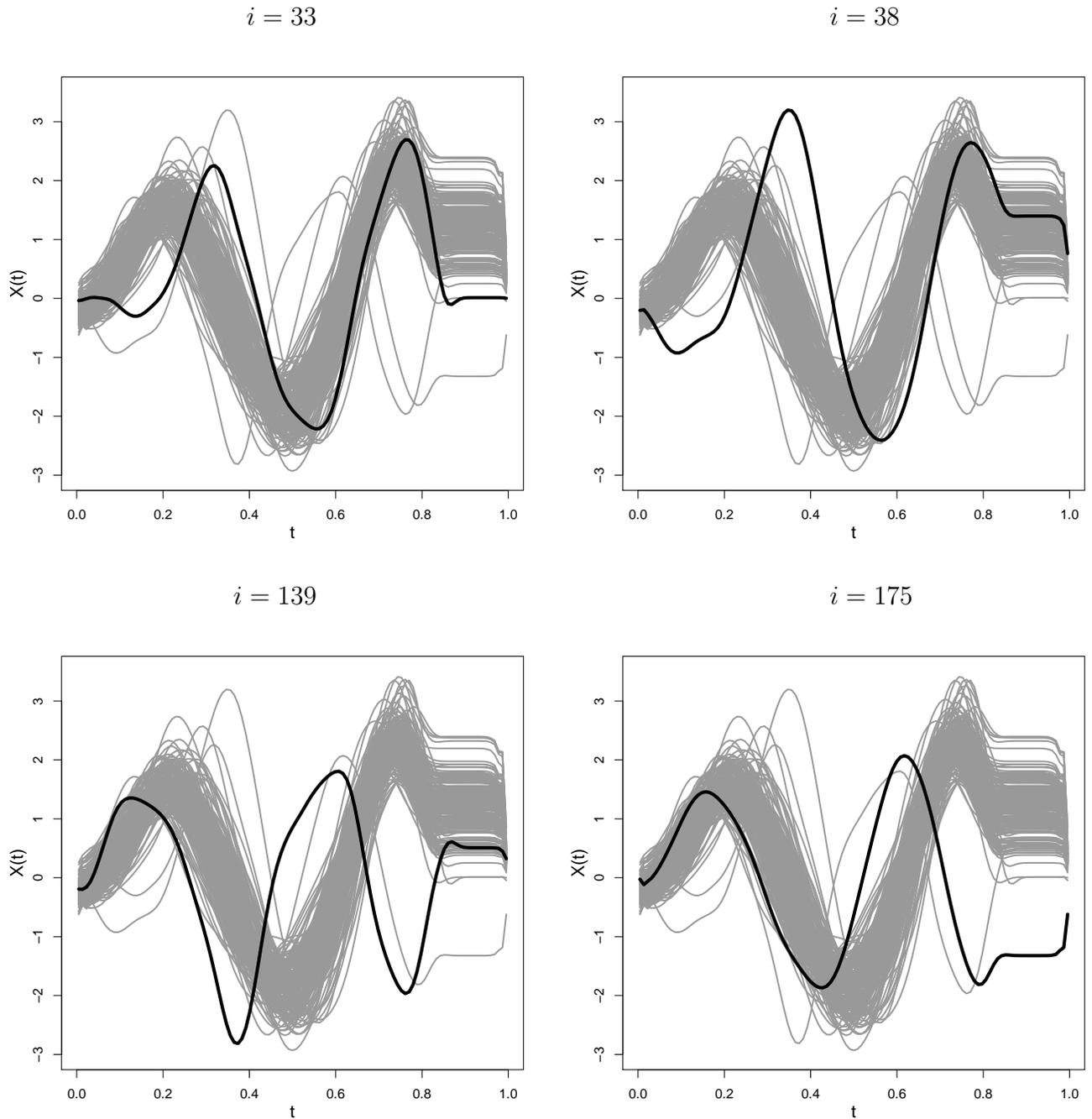


Fig. 6.5: Gráficos de velocidad en el eje vertical: Las curvas negras indican las observaciones atípicas del grupo \mathcal{I}_1 .

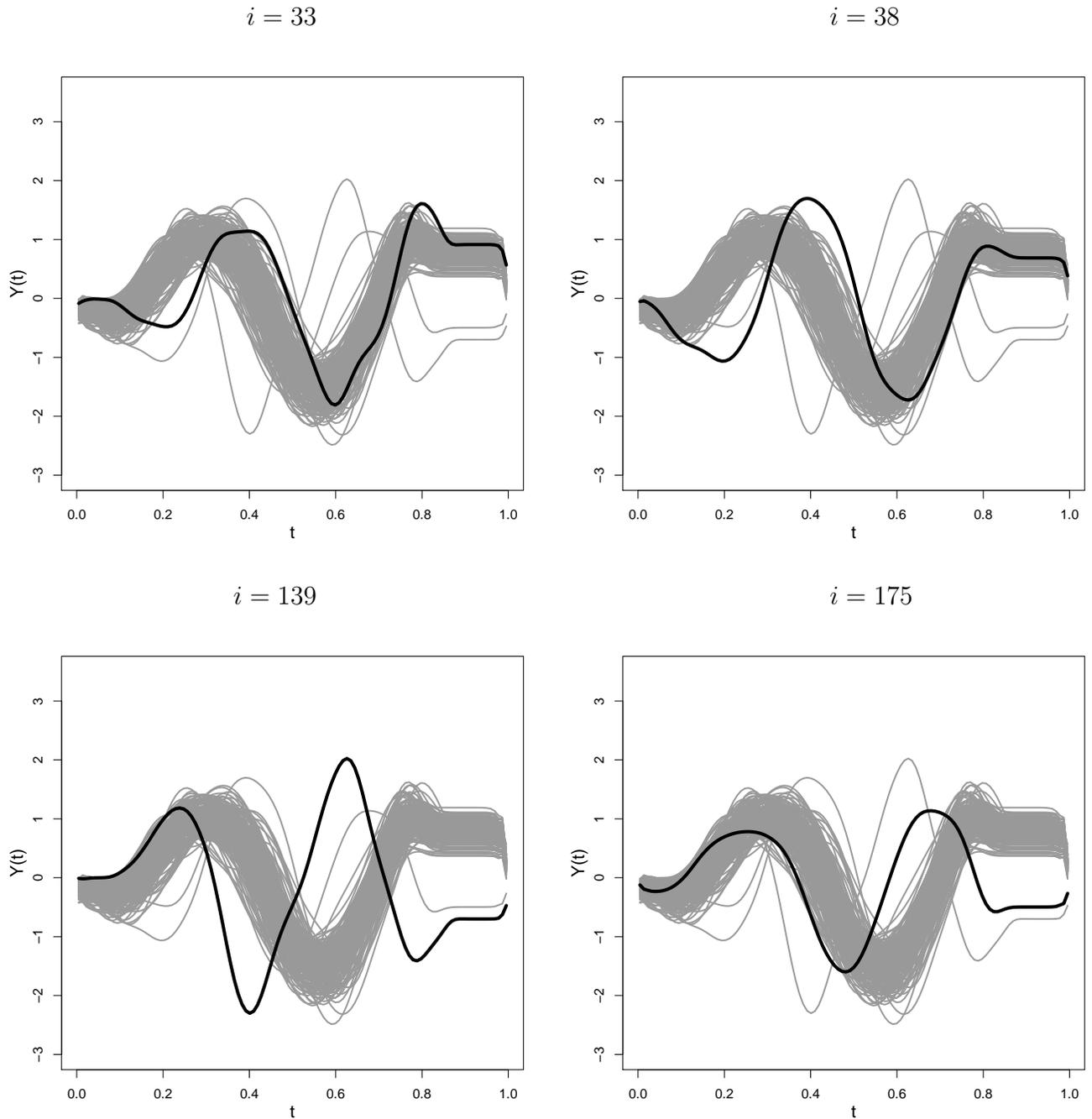
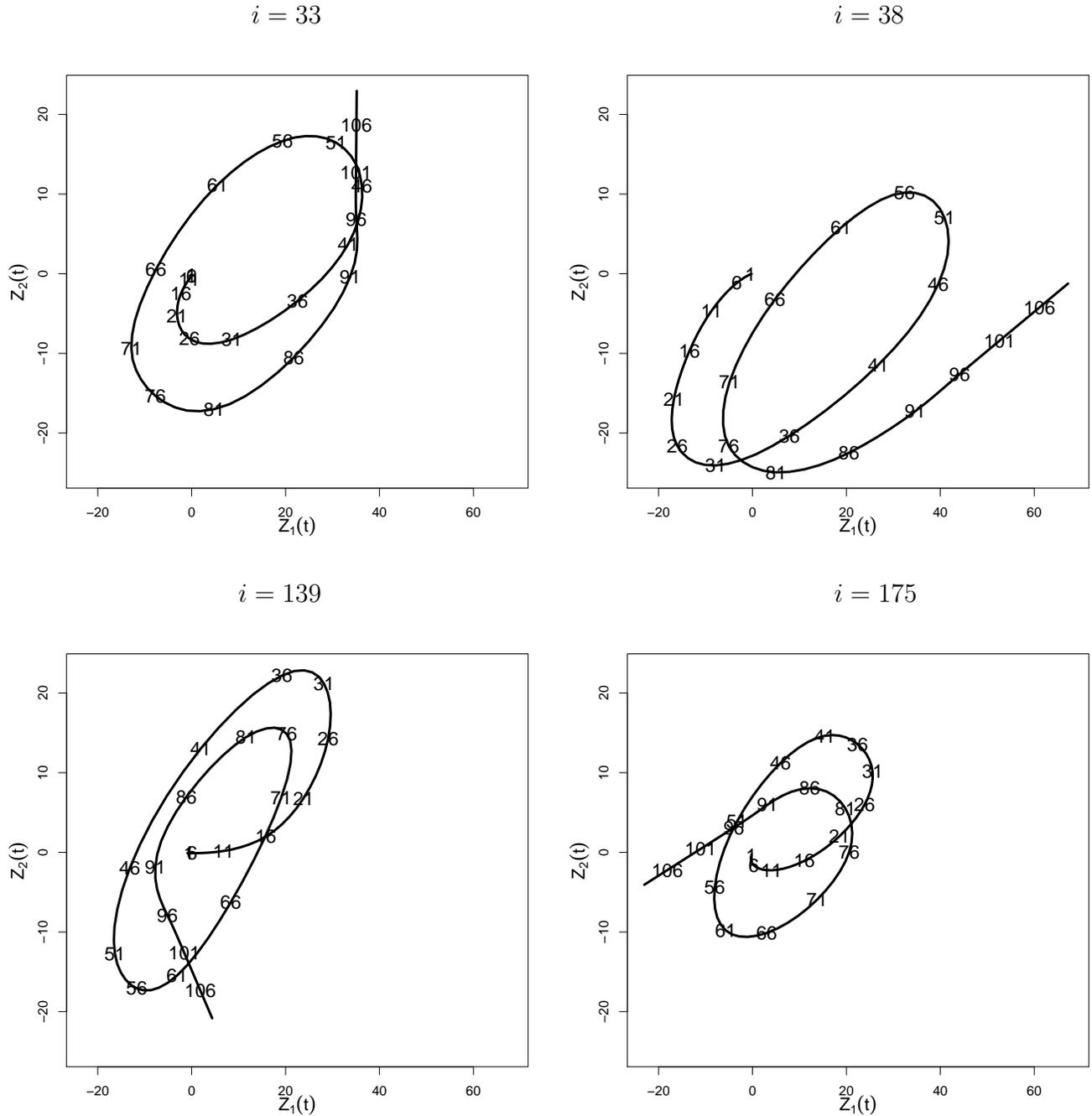


Fig. 6.6: Gráficos de escritura de la letra “e”: Se grafican las observaciones atípicas del grupo \mathcal{I}_1 con referencia temporal.



6.3. Conclusiones

En este ejemplo, hemos utilizado los estimadores robustos basados en Sieves introducidos en esta tesis para estimar la primera correlación y las primeras direcciones canónicas y para detectar datos atípicos. Hemos calculado los estimadores clásicos y robustos de la primera correlación y las primeras direcciones canónicas tanto con bases de Fourier como con bases de componentes principales utilizando convalidación cruzada para la elección de la dimensión y el Algoritmo 2. Basándonos en los estimadores robustos, los procedimientos de detección designados como $AT_{X \circ Y}$ en el Criterio 1 (Algoritmo 5 (GRID)) y como AT_{X+Y} en el Criterio 3 (Algoritmo 7 (GRID)) permitieron detectar dos grupos $\mathcal{I}_1 \subseteq \mathcal{I}_2$ de observaciones influyentes. Quitando de la muestra cada uno de esos grupos, los estimadores $\tilde{\rho}_{\hat{p},1}$ clásicos se acercan a los estimadores $\tilde{\rho}_{\hat{p},1}$ basados en el coeficiente de Spearman ρ_{SP} calculados con la muestra completa. Por otra parte, para la base de Fourier, el estimador clásico de la primera correlación canónica obtenido eliminando los datos atípicos dio notoriamente más cercana a los estimadores basados en el coeficiente de Spearman utilizando todos los datos, que cuando se calcula con la muestra completa. El mismo fenómeno se observó en la estimación de las direcciones canónicas, un acercamiento de las estimaciones clásicas quitando grupos de outliers a la estimación robusta con muestra completa respecto de la estimación clásica con muestra completa, lo que ilustra la confiabilidad de los procedimientos robustos ante la presencia de datos influyentes. Un gráfico de las curvas de velocidad en el eje horizontal y vertical, permitió visualizar como los datos detectados tenían un comportamiento diferente del resto de los datos. Dicho comportamiento estaría relacionado con un desfase temporal en los máximos y mínimos de las trayectorias así como un comportamiento diferencial en el intervalo $[0,8, 1]$. Para el grupo \mathcal{I}_1 , los gráficos de la escritura de las letras permitieron notar que las observaciones detectadas en el grupo \mathcal{I}_1 eran las más alejadas del común de las curvas y que correspondían a letras “e” hechas de una manera extraña comparadas con el resto de la muestra.

Bibliografía

- [1] Adrover J. G. y Donato S. M. (2015). A robust predictive approach for canonical correlation analysis. *Journal of Multivariate Analysis*, **133**, 356-376.
- [2] Alqallaf F., Van Aelst S., Yohai V. J. y Zamar R. H. (2009). Propagation of outliers in multivariate data. *The Annals of Statistics*, 311-331.
- [3] Anderson T. W. (1999). Asymptotic Theory for Canonical Correlation Analysis. *Journal of Multivariate Analysis*, **70**, 1-29.
- [4] Alfons A., Croux C., y Filzmoser P. (2016). Robust Maximum Association Estimators. *Journal of the American Statistical Association*, en prensa, <http://dx.doi.org/10.1080/01621459.2016.1148609>.
- [5] Bache, K. y Lichman, M. (2013). UCI Machine Learning Repository. <https://archive.ics.uci.edu/ml/datasets.html>
- [6] Bali L. y Boente G. (2009). Principal points and elliptical distributions from the multivariate setting to the functional case. *Statistics and Probability Letters*, **79**, 1858-1865.
- [7] Bali L. y Boente G. (2014). Consistency of a numerical approximation to the first principal component projection pursuit estimator. *Statistics and Probability Letters*, **94**, 181-191.
- [8] Bali L., Boente G., Tyler D.E. y Wang J.L. (2011), Robust functional principal components: A projection-pursuit approach, *The Annals of Statistics*, **39**, 2852-2882.

-
- [9] Bilodeau M. y Brenner, D. (1999). *Theory of Multivariate Statistics*. Springer–Verlag, New York.
- [10] Boente G., Barrera M. S. y Tyler D. E. (2014). A characterization of elliptical distributions and some optimality properties of principal components for functional data. *Journal of Multivariate Analysis*, **131**, 254-264.
- [11] Blomqvist N. (1950). On a measure of dependence between two random variables. *The Annals of Mathematical Statistics*, **21**, 593-600.
- [12] Branco J. A., Croux C., Filzmoser P. y Oliveira M. R. (2005). Robust canonical correlations: A comparative study. *Computational Statistics*, **20**, 203-229.
- [13] Chen K., Müller H. G., y Wang J. L. (2011). Stringing high-dimensional data for functional analysis. *Journal of the American Statistical Association*, **106**, 275-284.
- [14] Conway J. (1985). *A Course in Functional Analysis*. Springer, New York.
- [15] Croux C. y 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, pp. 211-217.
- [16] Daniels H. E. (1961). The asymptotic efficiency of a maximum likelihood estimator. In *Fourth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. **1**, pp. 151-163, University of California Press Berkeley.
- [17] Donoho D. L. (1982). Breakdown properties of multivariate location estimators. *Technical report, Harvard University, Boston*. <http://www-stat.stanford.edu/~donoho/Reports/Oldies/BPMLE.pdf>.
- [18] Falk M. (1998). A note on the comedian for elliptical distributions. *Journal of Multivariate Analysis*, **67**, 306-317.

-
- [19] Fujikoshi Y., Ulyanov V. y Shimizu R. (2011). Multivariate statistics: High-dimensional and large-sample approximations. *John Wiley and Sons*, New York.
- [20] Gnanadesikan R. y Kettenring J. R. (1972). Robust estimates, residuals, and outlier detection with multiresponse data. *Biometrics*, 81-124.
- [21] Gervini, D. (2008). Robust functional estimation using the spatial median and spherical principal components. *Biometrika*, **95**, 587-600.
- [22] Gut A. (2005). *Probability: A graduate course*. Springer, Berlin.
- [23] He G., Müller H. G. y Wang J. L. (2000) Extending correlation and regression from multivariate to functional data. En: M. Puri (Ed.), *Asymptotics in Statistics and Probability*, VSP, Zeist (Netherlands), pp. 197-210.
- [24] He G., Müller H. G. y Wang J. L. (2003). Functional canonical analysis for square integrable stochastic processes. *Journal of Multivariate Analysis*, **85**, 54-77.
- [25] He G., Müller H. G. y Wang J. L. (2004), Methods of canonical analysis for functional data, *Journal of Statistical Planning and Inference*, **122**, 141-159.
- [26] Hotelling H. (1936). Relations between two sets of variates. *Biometrika*, **28**, 321-377.
- [27] Hubert M., Rousseeuw P. J. y Segaert P. (2015). Multivariate functional outlier detection. *Statistical Methods & Applications*, **24(2)**, 177-202.
- [28] Hubert M., Rousseeuw P. J. y Segaert P. (2016). Multivariate and functional classification using depth and distance. arXiv preprint arXiv:1504.01128.
- [29] Jin J. y Cui H. (2010). Asymptotic distributions in the projection pursuit based canonical correlation analysis. *Science China Mathematics*, **53**, 485-498.

-
- [30] Kano Y. (1994). Consistency property of elliptical probability density function. *Journal of Multivariate Analysis* **51**, 139-147.
- [31] Kaphle K. (2011). *Perturbations of Operators with Application to Testing Equality of Covariance Operators*. (PhD. dissertation, Texas Tech University).
- [32] Kendall M.G. (1970). *Rank Correlation Methods*. Charles Griffin Books Series, London.
- [33] Kshirsagar A. (1972) *Multivariate Analysis*. Marcel Dekker, New York.
- [34] Leurgans S.E., Moyeed R. A. y Silverman B.W. (1993), Canonical Correlation Analysis when the data are curves, *Journal of the Royal Society, Series B*, **55**, 725-740.
- [35] 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.
- [36] López-Pintado S., Sun Y., Lin J. K. y Genton M. G. (2014). Simplicial band depth for multivariate functional data. *Advances in Data Analysis and Classification*, **8(3)**, 321-338.
- [37] Marden J. I. (1999). Some robust estimates of principal components. *Statistics and Probability Letters*, **43**, 349-359.
- [38] Maronna R. (1976). Robust M -estimators of multivariate location and scatter. *Annals of Statistics*, **4**, 51-67.
- [39] Maronna R. y Zamar R. (2002). Robust estimates of location and dispersion for high-dimensional datasets, *Technometrics*, **44**, 307-317.
- [40] Maronna R., Martin R. D. y Yohai V. (2006). *Robust Statistics: Theory and Methods*, John Wiley & Sons, New York.
- [41] Muirhead R. J. (1982). *Aspects of Multivariate Statistical Theory*, Wiley, New York.

-
- [42] Muirhead R. J. y Waternaux C. M. (1980), Asymptotic distributions in canonical correlation analysis and other multivariate procedures for nonnormal populations, *Biometrika*, **67**, 31-43.
- [43] Seber G.A.F. (2004). *Multivariate observations*, Wiley, New York.
- [44] Shevlyakov G. L. y Vilchevski N. O. (2001). *Robustness in Data Analysis: criteria and methods*. Walter de Gruyter, Utrech.
- [45] Spearman C. (1904). The proof and measurement of association between two things. *The American journal of psychology*, **15**, 72-101.
- [46] Stahel W. A. (1981). *Breakdown of Covariance Estimators*. <http://www.stat.math.ethz.ch/~stahel/publications/resrep31-wst.pdf>.
- [47] Székely G. J., Rizzo M. L. y Bakirov N. K. (2007). Measuring and testing dependence by correlation of distances. *The Annals of Statistics*, **35**, 2769-2794.
- [48] Székely G. J. y Rizzo M. L. (2009). Brownian distance covariance. *The Annals of Applied Statistics*, **3**, 1236-1265.
- [49] Taskinen S., Croux C., Kankainen A., Ollila E. y Oja, H. (2006). Influence functions and efficiencies of the canonical correlation and vector estimates based on scatter and shape matrices. *Journal of Multivariate Analysis*, **97**, 359-384.
- [50] van der Geer S. A. (2000). *Empirical Processes in M-estimation*. Cambridge University Press, Cambridge.
- [51] Yohai V. J. y Ben M. G. (1980). Canonical variables as optimal predictors. *The Annals of Statistics*, **8**, 865-869.