



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

Una medida de profundidad local para datos en espacios de Banach

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

Lic. Lucas Fernández Piana

Director de tesis: Dra. Marcela Svarc
Consejero de estudios: Dra. Graciela Boente
Lugar de Trabajo: IMAS - UBA - CONICET

Fecha de Defensa: 3 de diciembre 2018

Una medida de profundidad local para datos en espacios de Banach

Resumen

Las profundidades juegan un rol importante cuando analizamos conjuntos de datos complejos, como datos funcionales o datos en dimensión alta. El principal objetivo de una medida de profundidad es dar un orden del centro hacia afuera generalizando el concepto de mediana. Además son útiles para describir distintas características de la distribución subyacente de los datos. Inclusive son utilizadas como herramientas en distintos problemas de inferencia como tests de posición y simetría, clasificación, detección de datos atípicos, etc.

Sin embargo, dado que una de sus principales características es que el valor de la profundidad decrece sobre cada semirrecta con origen en el centro, no pueden capturar información relevante cuando la distribución estudiada es multimodal o no tiene soporte convexo. Con el objetivo de captar estas características locales, en los últimos años, se han introducido distintas definiciones de profundidad local: Agostinelli y Romanazzi (2), Paindaveine y Van Bever (44) y Agostinelli (1). La idea es restringir la profundidad global a un entorno de cada punto del espacio. En este sentido una profundidad local debe comportarse como una profundidad global condicionada al entorno.

Nuestro objetivo es dar una definición general de profundidad local para elementos aleatorios en un espacio de Banach extendiendo la definición de profundidad global dada por Cuevas y Fraiman (14), donde proponen la Profundidad Dual Integrada (IDD). Nosotros definiremos la Profundidad Local Dual Integrada (IDLD). Estudiaremos como las propiedades clásicas, introducidas por Zuo y Serfling (57), deben ser adaptadas en este nuevo contexto. Probaremos que bajo condiciones de regularidad nuestra propuesta cumple esas propiedades. Además, mostraremos resultados de consistencia fuerte para la versión empírica de la IDLD y para las regiones de profundidad local. La mayor ventaja de nuestra propuesta es su flexibilidad para lidiar con datos en contextos generales y su bajo costo computacional, lo cual la vuelve apta para el análisis de datos en alta dimensión.

Como una aplicación natural, propondremos un procedimiento de clusters basado en profundidades locales y mostraremos su muy buen desempeño con datos simulados y datos reales para distintas clases de los mismos: multivariados, funcionales, funcionales multidimensionales y mixtos.

Palabras Claves: Medidas de Profundidad, Análisis de Cluster, Datos Funcionales, Procedimientos basados en Proyecciones.

A local depth measure for data in Banach spaces

Abstract

Data depth measures play an important role when analyzing complex data, such as functional or high dimensional data. The main goal of depth measures is to give center-outer order of the data, generalizing the concept of median to general settings. Depth measures are also useful to describe different features of the underlying distribution of the data. Moreover, they are powerful tools to deal with several inference problems such as location and symmetry tests, classification, outlier detection, etc.

Nonetheless, since one of their major characteristics is that the depth values decrease along any half-line ray from the center, they are not suitable to capture distribution characteristics when data are multimodal or have non-convex support. Hence, with the aim of revealing local features of the underlying distribution in the last few years several definitions of local depths have been introduced: Agostinelli and Romanazzi (2), Paindaveine and Van Bever (44) and Agostinelli (1). The basic idea is to restrict a global depth measure to a neighborhood of every point of the space. In that way, a local depth measure should behave as a global depth measure conditional to neighborhoods of the different points.

Our goal is to give a general definition of local depth for random elements in a Banach space, extending the definition of global depth given by Cuevas and Fraiman (14), where they introduce the Integrated Dual Depth (IDD). We define the Integrated Dual Local Depth (IDLD). We study how the classical properties, introduced by Zou and Serfling (57), should be analyzed within the framework of local depth. We prove, under mild regularity conditions, that our proposal enjoys those properties. Moreover, uniform strong consistency results are exhibited for the definition of the empirical local depth of to the population counterpart, and also for the local depth regions. The main advantages of our proposals are its flexibility in dealing with general data and also its low computational cost, which enables it to work with high-dimensional data.

As a natural application, we propose a clustering procedure based on local depths, and illustrate its promising performance with synthetic and real data, for different kind of data: multivariate, functional, functional multidimensional and mixed.

Key Words: Data Depth, Cluster Analysis, Functional Data, Projection Procedures.

Agradecimientos

A Marcela, mi directora, gracias por el tiempo, las ganas, el compromiso, la contención, por escucharme, por levantarme cuando me caía; en resumen gracias por todo.

A mi familia. Especialmente, a mi madre que se tomó el trabajo de revisar la ortografía.

A mis amigos, por su incondicionalidad y paciencia en este proceso.

A Rodrigo, por su ayuda con los gráficos y contención emocional.

A Guido, Pablo y Julián por haber aportado su punto de vista y conocimiento.

A Graciela por aceptar ser mi consejera de estudios y su compromiso.

A mis compañeros de trabajo.

A Deborah nuestra querida auxiliar en la subcomisión de doctorado por la gran predisposición a la hora de lidiar con trámites administrativos.

A mis compañeros de elenco.

A mis compañeros de estudio.

A esos locos con los que hacemos divulgación científica.

A todas las personas que me permitieron aprender de su experiencia.

A la Universidad de Buenos Aires.

A CONICET por la beca.

Índice general

1. Medidas de Profundidad Globales y Locales	9
1.1. Medidas de Profundidad Globales	9
1.1.1. Definiciones y Propiedades	9
1.1.2. Medidas de Profundidad para Datos Multivariados	11
1.1.3. Estructura general para funciones de profundidad	13
1.1.4. Medidas de Profundidad para Datos Funcionales	16
1.1.5. Profundidad Dual Integrada (IDD)	20
1.2. Aplicaciones	22
1.3. Profundidades Locales	24
1.3.1. Profundidad Local Simplicial y Profundidad Local del Semiespacio para datos multivariados.	26
1.3.2. Construcción general de profundidades locales a partir de profundi- dades para datos multivariados.	28
1.3.3. Profundidad Local del Semiespacio para Datos Funcionales	30
2. Profundidad Local Dual Integrada	33
2.1. Introducción	33
2.2. Marco de Referencia y Definiciones	33
2.2.1. Profundidad Local Dual Integrada	34
2.3. Versión Empírica y Resultados Asintóticos	48
2.4. Regiones de Profundidad Local	56
3. Clusters Basados en Profundidades Locales	61
3.1. Procedimiento de cluster basado en profundidades locales	61
3.2. Otras Aplicaciones	63
3.2.1. Datos Funcionales Mixtos	63
3.3. Simulaciones y Datos Reales	66
3.3.1. Criterios para Comparar Particiones	66
3.3.2. Simulaciones: datos multivariados	68
3.3.3. Simulaciones: datos funcionales	73
3.3.4. Simulaciones Datos Funcionales Multidimensionales	82

3.3.5. Ejemplos: Datos Reales Funcionales	87
3.3.6. Ejemplo: Datos Reales de Estructura Mixta	89

Capítulo 1

Medidas de Profundidad Globales y Locales

1.1. Medidas de Profundidad Globales

1.1.1. Definiciones y Propiedades

El concepto de *Profundidad* surgió primero para datos multivariados, buscando dar una idea de centralidad en los datos, de este modo un dato con un valor de profundidad *alto* debería estar en el centro de la nube de puntos y a medida que se aleja del centro el valor de la profundidad debería disminuir. En el caso de datos univariados podríamos pensar como punto más profundo a la mediana y mediante los cuantiles asignarle profundidad al resto de los puntos. De este modo el concepto de profundidad estaría estrechamente ligado al orden natural de los números reales. Cuando pasamos a un contexto multivariado, no se cuenta con un orden total, sino únicamente con órdenes parciales y esto hace que surjan diversas definiciones de profundidad, cuya interpretación ya no será inmediata.

A grandes rasgos, para una distribución P en \mathbb{R}^d , una función de profundidad es una función $D(x, P)$ que provee un orden del centro hacia afuera para puntos $x \in \mathbb{R}^d$ basado en P . Interpretar “orden del centro hacia afuera” sugiere dos cosas

- Una noción relevante de centro.
- Los puntos cerca del centro deberían tener mayor profundidad que los más lejanos.

De estas consideraciones se desprende que el “centro” debe constar de los puntos que maximicen globalmente la función de profundidad. Naturalmente surge una pregunta: ¿existe una definición general de función de profundidad definida en \mathbb{R}^d definida con respecto a una distribución arbitraria que puede ser continua o discreta? ¿Qué propiedades son deseables para caracterizarlas?

Dada una distribución P en \mathbb{R}^d , nuestra atención quedará confinada a que las medidas consideradas sean acotadas y no negativas. A continuación daremos una definición formal de función de profundidad (57), para ello necesitamos introducir la siguiente notación: llamamos \mathbb{P} a la clase de distribuciones sobre los conjuntos Borelianos de \mathbb{R}^d y llamamos P_ξ la distribución dada por el vector aleatorio ξ .

Definición 1. Sea $D(\cdot, \cdot) : \mathbb{R}^d \times \mathbb{P} \rightarrow \mathbb{R}$ una función acotada y no negativa que satisface:

P.1. $D(Ax + b, P_{AX+b}) = D(x, P_X)$ para todo vector aleatorio $X \in \mathbb{R}^d$, para toda matriz $A \in \mathbb{R}^{d \times d}$ no singular y para cualquier vector $b \in \mathbb{R}^d$.

P.2. $D(\theta, P) = \sup_{x \in \mathbb{R}^d} D(x, P)$, para cualquier $P \in \mathbb{P}$ con centro θ .

P.3. Para toda $P \in \mathbb{P}$ con punto más profundo θ , $D(x, P) \leq D(tx + (1-t)\theta, P)$, para cualquier $t \in [0, 1]$.

P.4. $\lim_{\|x\| \rightarrow +\infty} D(x, P) = 0$, para todo $P \in \mathbb{P}$.

Si además quisieramos que se cumplieran hipótesis de continuidad, deberíamos pedir que

P.5. Fijada $P \in \mathbb{P}$ $D(\cdot, P) : \mathbb{R}^d \rightarrow \mathbb{R}$ es continua.

P.6. Fijado $x \in \mathbb{R}^d$ $D(x, \cdot) : \mathbb{P} \rightarrow \mathbb{R}$ es continua.

Anteriormente hicimos referencia a la palabra *centro* para resaltar un punto de simetría. Es posible tomar varias nociones de simetría multivariada. Sin embargo, las más usadas en la literatura son las siguientes.

Definición 2. Sea $X \in \mathbb{R}^d$ un vector aleatorio decimos que:

- Es **Centralmente simétrico** (C-simétrico) respecto de θ si la distribución de $X - \theta$ es la misma que la de $\theta - X$.
- Es **Angularmente simétrico** (A-simétrico) respecto de θ si $(X - \theta)/\|X - \theta\|$ es centralmente simétrico respecto del origen.
- Es **Semiespacialmente simétrico** (H-simétrico) respecto de θ si $P(X \in H) \geq \frac{1}{2}$ para todo H semiespacio cerrado tal que $\theta \in H$.

Observación 1. Es claro que C-simétrico \Rightarrow A-simétrico \Rightarrow H-simétrico.

1.1.2. Medidas de Profundidad para Datos Multivariados

Las medidas de profundidad para datos multivariados surgieron con el objetivo de extender la noción de los estadísticos de orden. Los mismos son sumamente útiles al emplear técnicas no paramétricas, sobre todo en casos donde la distribución de los datos no es gaussiana. Además, es importante contar con técnicas de “ordenación” para los puntos que tengan en cuenta la geometría de los datos. Lo que no ocurre, por ejemplo si se busca definir la observación central mediante la media o la mediana coordenada a coordenada. Por otra parte, cuando consideramos datos en la recta real el orden usual de los mismos da una noción de ranking entre las observaciones, al pasar a dimensiones mayores esto no se puede hacer de manera natural y por lo tanto, es conveniente poder definir un centro (que se puede hacer de muchas formas distintas) y a partir del mismo orientar las observaciones de adentro hacia afuera.

Por otra parte, juegan un rol importante a la hora de describir características de la distribución subyacente a las observaciones. Verdaderamente, las profundidades no sólo proveen un funcional robusto multivariado para el problema de localización a través del punto más profundo, sino también revelan información sobre la dispersión, forma y simetría de la distribución cuando se contruyen las regiones de profundidad como describe Serfling (2004) (48). Otras posibilidades que ofrecen se pueden encontrar en los llamados métodos basados en profundidades que permiten abordar distintos problemas de inferencia. Por ejemplo, tests de posición y diferencias de escala basados en los *DD*-plots, que fueron introducidos en primera instancia como una herramienta gráfica para exploración de datos (Liu et al. (1999) (37), Li y Liu (2004) (36)); diagnóstico de no normalidad (Liu et al. (1999) (37) y detección de outliers (Chen et al. (2009) (8)). Recientemente, las profundidades se han utilizado extensivamente en el contexto de clasificación como se ve reflejado, entre muchos otros autores, en los artículos de Ghosh y Chaudhuri (2005) (23), Li et al. (2012) (34), Cuevas y Fraiman (2009) (14), Dutta y Ghosh (2012) (15), Paindaveine and Van Bever (2012) (44), Cuesta-Albertos et al. (2017) (11).

En esta sección estudiaremos algunos ejemplos de funciones de profundidad para datos multivariados y trataremos de dar una clasificación de las medidas de profundidad con respecto a las propiedades (P.1.-P.4.). En primer lugar definiremos las dos medidas de profundidad para datos multivariados más clásicas en la literatura.

Profundidad del Semiespacio

La semilla de esta teoría fue plantada por Tukey (53) con la *Profundidad del Semiespacio* (Halfspace Depth, HD).

La profundidad del semiespacio de un punto $x \in \mathbb{R}^d$ con respecto a una medida de probabilidad P en \mathbb{R}^d está definida por la mínima probabilidad de que un semiespacio cerrado que contenga a x , esto es,

$$HD(x, P) = \inf\{P(H) : H \text{ semiespacio cerrado, } x \in H\}, \quad x \in \mathbb{R}^d.$$

La versión muestral es la siguiente, sea $x \in \mathbb{R}^d$, la profundidad de x respecto de una muestra X_1, X_2, \dots, X_n también \mathbb{R}^d es la menor fracción de puntos de la muestra que hay en un semiespacio cerrado que contenga a x , es decir,

$$HD_n(x) = \frac{\min_{u \in \mathbb{R}^d} \#\{i : \langle X_i, u \rangle \geq \langle x, u \rangle\}}{n},$$

donde $\langle \cdot, \cdot \rangle$ denota el producto interno.

Si consideramos P una medida de probabilidad en \mathbb{R} y X una variable aleatoria con función de distribución acumulada F , entonces

$$HD(P, x) = \min\{P(X \leq x), (1 - P(X < x))\} = \min\{F(x), (1 - F(x^-))\}.$$

El principal inconveniente que presenta esta medida de profundidad es que es computacionalmente muy costosa en dimensiones moderadas y altas.

Teorema 1. *La profundidad del semiespacio cumple las propiedades P.1.-P.4.*

Demostración. Es claro que $H(x, P)$ es acotada, no negativa. Sea H_x un subespacio que contiene a x .

P.1. Sea, $X \in \mathbb{R}^d$ vector aleatorio, $A \in \mathbb{R}^{d \times d}$ una matriz no singular y $b \in \mathbb{R}^d$. Sea,

$$H : x \in H \Rightarrow Ax + b \in AH + b$$

y

$$P_{AX+b}(AH + b) = P(H).$$

Entonces $HD(Ax + b, P_{AX+b}) = HD(x, P)$.

P.2. Supongamos que P es H -simétrica respecto de un único punto $\theta \in \mathbb{R}^d$. Tenemos que, $P(H_\theta) \geq \frac{1}{2}$ para todo H_θ tal que $\theta \in H_\theta$. Entonces, $HD(\theta, P) \geq \frac{1}{2}$.

Supongamos que existe $x_0 \in \mathbb{R}^d$ con $HD(x_0, P) \geq \frac{1}{2}$ entonces $P(H) \geq \frac{1}{2}$ para todo semiespacio tal que $x_0 \in H$.

Por lo tanto, P es H -simétrica respecto de x_0 . Esto es una contradicción.

Luego,

$$HD(\theta, P) = \sup_{x \in \mathbb{R}^d} HD(x, P).$$

P.3. Sea $0 < \alpha < 1$, debemos comparar $HD(x, P)$ y $HD((1 - \alpha)\theta + \alpha x, P)$.

Por el Teorema de Separación del hiperplano sabemos que ,

$$\exists H_x \text{ tq } H_x \subseteq H_{(1-\alpha)\theta + \alpha x}.$$

Entonces,

$$HD(x, P) \leq HD((1 - \alpha)\theta + \alpha x, P).$$

P.4. Sabemos que $P(\|X\| \geq \|x\|) \xrightarrow{\|x\| \rightarrow +\infty} 0$. Además, $H_x \subseteq \{\|X\| \geq \|x\|\}$.

Luego, $HD(x, P) \xrightarrow{\|x\| \rightarrow +\infty} 0$.

□

Profundidad Simplicial

Otra noción importante fue planteada por Regina Liu en 1990 (35). La *Profundidad Simplicial* (SD) de un punto $x \in \mathbb{R}^d$ con respecto a una medida de probabilidad P en \mathbb{R}^d es la probabilidad de que x pertenezca a un simplex aleatorio en \mathbb{R}^d , i.e,

$$SD(x, P) = P(x \in S[X_1, \dots, X_{d+1}]), \quad x \in \mathbb{R}^d.$$

donde X_1, \dots, X_{d+1} es una muestra de vectores aleatorios con distribución P y $S[X_1, \dots, X_{d+1}]$ es el simplex d -dimensional con vértices X_1, \dots, X_{d+1} .

A continuación daremos la definición muestral de la profundidad simplicial. Sea X_1, \dots, X_n una muestra aleatoria en \mathbb{R}^d , la profundidad simplicial de un punto $x \in \mathbb{R}^d$ es

$$SD_n(x) = \binom{n}{d+1}^{-1} \sum_{1 \leq i_1 < i_2 < \dots < i_{d+1} \leq n} \mathcal{I}\{x \in S[X_{i_1}, \dots, X_{i_{d+1}}]\},$$

donde \mathcal{I} es la función indicadora.

En el caso univariado, la función de profundidad simplicial es $SD(x, P) = F(x)(1 - F(x^-))$, donde F es la función de distribución acumulada. Zuo y Serfling en un artículo del año 2000 (57) prueban que cumple las propiedades **P.1.-P.4.** para distribuciones continuas A -simétricas.

Sin embargo, estas propiedades no se cumplen en el caso general. Veamos el siguiente contraejemplo, consideramos una distribución discreta unidimensional donde la probabilidad puntual está dada por,

$$P(X = 0) = P(X = \pm 1) = P(X = \pm 2) = \frac{1}{5}.$$

Es claro que X es C -simétrico respecto de 0.

$$SD\left(\frac{1}{2}, P\right) = F\left(\frac{1}{2}\right)(1 - F(0)) = \frac{6}{25}$$

$$SD(1, P) = F(1)(1 - F(0)) = \frac{8}{25}$$

Esto contradice **P.3.**

1.1.3. Estructura general para funciones de profundidad

En el mismo artículo mencionado anteriormente Zou y Serfling (57), proponen cuatro estructuras generales para la construcción de profundidades.

Funciones de profundidad tipo A

Sea $h(x, x_1, \dots, x_r)$ una función acotada y no negativa que mide en algún sentido la cercanía de x a los puntos x_1, \dots, x_r . Una *función de profundidad tipo A* nos queda definida por la cercanía esperada de x a una muestra aleatoria de tamaño r

$$D(x, P) = Eh(x, X_1, \dots, X_r),$$

donde X_1, \dots, X_r es una muestra de vectores aleatorios con distribución P .

Ejemplo 1. Tomando $r = d + 1$ y $h(x, x_1, \dots, x_{d+1}) = I\{x \in S[x_1, \dots, x_{d+1}]\}$ obtenemos la *profundidad simplicial*.

Ejemplo 2. Dados $x_1, \dots, x_d \in \mathbb{R}^d$ determinan un único hiperplano que los contiene, al cual le corresponden dos semiespacios cerrados que lo contienen como frontera. Notamos H_{x_1, \dots, x_d}^P al que tenga probabilidad mayor o igual que $\frac{1}{2}$ bajo la distribución $P \in \mathbb{R}^d$. Definimos la *Profundidad Mayorante* como

$$MJD(x, P) = P(x \in H_{x_1, \dots, x_d}^P), \quad x \in \mathbb{R}^d.$$

Es claro que es una *profundidad tipo A* tomando $r = d$ y $h(x, x_1, \dots, x_d) = I\{x \in H_{x_1, \dots, x_d}^P\}$.

Funciones de profundidad tipo B

Sea $h(x, x_1, \dots, x_r)$ una función no acotada y no negativa que mide en algún sentido la lejanía de x a los puntos x_1, \dots, x_r . Una *función de profundidad tipo B* nos queda definida por

$$D(x, P) = (1 + Eh(x, X_1, \dots, X_r))^{-1},$$

para X_1, \dots, X_r una muestra de vectores aleatorios de distribución P .

Ejemplo 3. Tomando $h(x, x_1, \dots, x_r) = \Delta^\alpha(S[x, x_1, \dots, x_d])$ donde $\Delta(S[x, x_1, \dots, x_d])$ es el volumen del simplex d -dimensional $S[x, x_1, \dots, x_d]$ y $\alpha > 0$. h es una medida de dispersión de la nube de puntos $\{x, x_1, \dots, x_d\}$ y de acuerdo a lo anterior

$$(1 + E[\Delta^\alpha(S[x, x_1, \dots, x_d])])^{-1},$$

define una *función de profundidad tipo B*. Sin embargo, no cumple con la propiedad de la invarianza afín porque para una matriz A no-singunlar y un vector b , sabemos que la ecuación

$$\Delta^\alpha(S[Ax + b, Ax_1 + b, \dots, Ax_d + b]) = |\det(A)|\Delta^\alpha(S[x, x_1, \dots, x_d]),$$

no necesariamente se verifica. Se cumple únicamente si $\det(A) = 1$.

Este problema puede ser rectificado y definimos la Profundidad Simplicial de Volumen

$$SVD^\alpha = \left(1 + E \left[\left(\frac{\Delta(S[x, X_1, \dots, X_d])}{\sqrt{\det(\Sigma)}} \right)^\alpha \right] \right)^{-1},$$

donde Σ es la matriz de covarianzas de P . Esta versión cumple **(P.I.)**

Funciones de profundidad tipo C

Sea $O(x, P)$ una medida de outlyingness para el punto $x \in \mathbb{R}^d$ con respecto al centro o al punto más profundo de la distribución P . Nos queda definida una *función de profundidad tipo C* por

$$D(x, P) = (1 + O(x, P))^{-1}.$$

Ejemplo 4. Sea $x \in \mathbb{R}^d$. Definimos la Profundidad de Proyección $PD(x, P)$, mediante la siguiente función de outlyingness

$$O(x, P) = \sup_{u \in \mathbb{R}^d, \|u\|=1} \left(\frac{u'x - \text{Med}(u'X)}{\text{MAD}(u'X)} \right),$$

donde X tiene distribución P , Med es la mediana univariada, $\text{MAD}(Y) = \text{Med}(|Y - \text{Med}(Y)|)$ y $\|\cdot\|$ es la norma Euclídea.

Funciones de profundidad tipo D

Sea \mathcal{C} una clase de conjuntos cerrados de \mathbb{R}^d y P una medida de probabilidad en \mathbb{R}^d . Definimos la correspondiente *función de profundidad tipo D* como

$$D(x; P, \mathcal{C}) = \inf_{C \in \mathcal{C}} \{P(C) : x \in C\}.$$

Hay una clase distinguida que releva nuestra atención y cumple dos propiedades

- Si $C \in \mathcal{C} \implies \overline{C} \in \mathcal{C}$.
- Para $C \in \mathcal{C}$ y $x \in C^o$, $\exists C_1$ tal que $x \in \partial C_1$, $C_1 \subseteq C^o$.

Observar que la clase de semiespacios cerrados en \mathbb{R}^d satisface ambas propiedades y por lo visto antes, la profundidad del semiespacio es un típico ejemplo de este tipo de medidas de profundidad.

1.1.4. Medidas de Profundidad para Datos Funcionales

La extensión de las funciones de profundidad al contexto de espacios infinito dimensionales no es inmediata. El foco del problema se encuentra en que el análogo natural de algunas funciones de profundidad definidas en \mathbb{R}^d pueden ser nulas en un conjunto de probabilidad uno. Esto se señala en (Dutta et al. 2011) (16) para la profundidad del semiespacio (53) con respecto a ciertas medidas de probabilidad en el espacio de Hilbert ℓ_2 y también para la profundidad de bandas (38) en (Chakraborty y Chaudhuri 2014) (7). Un trabajo más detallado sobre este problema se puede encontrar en (Kuelbs y Zinn 2013) (33).

Cabe señalar que, en el último tiempo, algunos autores han propuesto algunas modificaciones a las propiedades que debe cumplir una función de profundidad en el contexto funcional: (Nieto-Reyes y Battey 2017) (41) (42) y (Gijbels y Nagy 2017) (24). En este trabajo, nosotros no tendremos en cuenta estas nuevas propiedades, nos mantendremos en el camino clásico trazado por Zuo y Serfling (57).

A continuación detallaremos algunos ejemplos de funciones de profundidades definidas para espacios de dimensión infinita; cada una de ellas es posible que cuenten con su propia patología, pero no nos detendremos a analizarlas particularmente. Sin embargo cabe destacar que la profundidad dual integrada (14) parece ser la más inmune a estas dificultades.

Profundidad de Fraiman y Muniz

Es la primer medida de profundidad propuesta para datos funcionales y fue introducida por Fraiman y Muniz en 2001 (18). Sea D_t una medida de profundidad univariada asociada a una distribución univariada F_t y sea $x(t)$ una función continua en el intervalo $[0, 1]$. Llamamos $Z(t) = D_t(x(t))$, que en particular para el caso la profundidad simplicial es $Z(t) = F_t(x(t))[1 - F_t(x(t^-))]$. Luego, la correspondiente medida de profundidad es

$$I(x) = \int Z(t)dt = \int F_t(x(t))[1 - F_t(x(t^-))]dt.$$

A continuación damos la correspondiente definición muestral. Para cada $t \in [0, 1]$, sea $F_{n,t}$ la distribución empírica de la muestra $x_1(t), \dots, x_n(t)$ y sea $Z_i(t)$ la profundidad univariada del dato $x_i(t)$ en esta muestra, dada por $D_i(t) = 1 - \left| \frac{1}{2} - F_{n,t}(x_i(t)) \right|$. Para cada $1 \leq i \leq n$, sea

$$I_i = \int_0^1 D_i(t)dt,$$

luego construiremos un ranking para las observaciones $x_i(t)$ de acuerdo a los valores de I_i . La adaptación al caso multivariado es directa, reemplazando la integral por la suma finita apropiada.

En este caso los autores no prueban ninguna de las propiedades citadas por Zuo y Serfling (57) y consiguen la consistencia fuerte para funciones Lipschitz en el intervalo $[0, 1]$.

Profundidad de Tukey Aleatorizada

Mencionamos en la sección anterior la Profundidad del Semiespacio, $HD(x, P)$, que también es conocida como Profundidad de Tukey. Un grave problema que presenta esta profundidad es que requiere calcular todas las posibles proyecciones univariadas. Cuesta-Albertos y Nieto-Reyes (2008) (12) definen la Profundidad de Tukey Aleatorizada, con el objetivo de superar este problema. A su vez esta definición toma gran importancia, ya que se puede extender a espacios de Hilbert separables.

Una definición equivalente de $HD(x, P)$ es la siguiente: dado $v \in \mathbb{R}^p$, sea Π_v la proyección de \mathbb{R}^p en el subespacio unidimensional generado por v . Entonces, $P \circ \Pi_v^{-1}$ es la distribución marginal de P en ese subespacio y

$$HD(x, P) = \inf\{D_1(\Pi_v(x), P \circ \Pi_v^{-1}) : v \in \mathbb{R}^p\},$$

donde $D_1(x, P) = \min\{F(x), 1 - F(x^-)\}$.

En primer término definen la Profundidad de Tukey Aleatorizada para datos multivariados.

Definición 3. Sea $P \in \mathbb{P}$. Sea $v \in \mathbb{P}$ absolutamente continua y sean v_1, \dots, v_k vectores aleatorios i.i.d. con distribución v . La Profundidad de Tukey Aleatorizada de $x \in \mathbb{R}^p$ con respecto a P basada en k vectores aleatorios elegidos con v es

$$D_{T,k,v}(x, P) = \min\{D_1(\Pi_{v_i}(x), P \circ \Pi_{v_i}^{-1}) : 1 \leq i \leq k\}, \quad x \in \mathbb{R}^p.$$

Según prueban los autores (12), para datos multivariados esta profundidad cumple **P.1.-P.3.** y **P.4.** con convergencia en probabilidad. Una posibilidad interesante que presenta la Profundidad de Tukey Aleatorizada es que puede extenderse a espacios de Hilbert separables. Sin embargo, la extensión pierde la propiedad **P.4.** Veámoslo en el siguiente ejemplo.

Consideramos $\mathbb{H} = L_2[0, 1]$ con el producto interno usual. Sea $\{\delta_n\} \subset \mathbb{R}^+$ con $\lim_n \delta_n = 0$. Sea $\{x_n\} \in \mathbb{H}$, $x_n(t) = \frac{1}{\delta_n} \mathcal{I}_{[0, \delta_n]}(t)$. Tenemos, $\|x_n\| = \delta_n^{-\frac{1}{2}} \Rightarrow \lim_n \|x_n\| = +\infty$. Tomemos v igual a la distribución del movimiento Browniano y sea $P = v$. Si X es un elemento aleatorio con distribución v , entonces $\langle x_n, X \rangle$ converge a cero en probabilidad, pues

$$E|\langle x_n, X \rangle| = E\left|\int_0^{\delta_n} X(t)\delta_n^{-1} dt\right| \leq \int_0^{\delta_n} E|X(t)|\delta_n^{-1} dt = \int_0^{\delta_n} \left(\frac{2t}{\pi}\right)^{\frac{1}{2}} \delta_n^{-1} dt \leq \left(\frac{2\delta_n}{\pi}\right)^{\frac{1}{2}},$$

donde la última igualdad es válida porque la distribución de $X(t)$ es $N(0, t)$.

Luego, si $v_1, \dots, v_k \in \mathbb{H}$ son elegidos aleatoriamente con distribución v tenemos

$$\lim_n D_1(\langle v_i, x_n \rangle, P \circ \Pi_{v_i}^{-1}) = D_1(0, P \circ \Pi_{v_i}^{-1}) = \max_x D_1(x, P \circ \Pi_{v_i}^{-1}) = \frac{1}{2},$$

porque $P \circ \Pi_{v_i}^{-1}$ porque la distribución es una normal centrada.

Profundidad h-Modal

La h-mode puede considerarse como el primer precedente de la idea de “profundidad local” (Cuevas et al. 2007) (13), tema que desarrollaremos a lo largo de esta tesis. La idea es bastante simple e intuitiva: supongamos que los datos viven en un espacio normado y sea K una función núcleo en \mathbb{R} , como por ejemplo los que se utilizan para definir estimadores de la densidad. Sea h una constante positiva, se define la profundidad h -modal de un dato x con respecto a la distribución de un elemento aleatorio X como,

$$D_h(x) = E [K_h(\|x - X\|)], \text{ donde } K_h(t) = \frac{1}{h} K(t/h) \text{ es el núcleo reescalado.} \quad (1.1)$$

Esta noción de profundidad tiene un significado “local” y recuerda, en cierto modo, a la idea de definir la profundidad, en \mathbb{R}^d , en términos de la función de densidad. Dado que en espacios infinito dimensionales no hay un concepto natural de densidad, la función D_h puede servir como un sucedáneo. La versión muestral se construye utilizando el método plug-in; es decir, dada una muestra aleatoria X_1, \dots, X_n se puede estimar la versión poblacional mediante,

$$\hat{D}_h(x) = \frac{1}{n} \sum_{i=1}^n K_h(\|x - X_i\|). \quad (1.2)$$

La profundidad h-modal ha llamado la atención en los últimos años, si bien los autores en el paper original no prueban ninguna resultado, se puede encontrar la consistencia del estimador 1.2 (Nagy 2015) (40) y un estudio exhaustivo de las propiedades **P.1-P.6** para la versión poblacional 1.1 en (Gijbels y Nagy 2017) (24).

Profundidad de Bandas

La propuesta de López-Pintado y Romo (38) se basa en explotar las propiedades del gráfico de una función. Antes de pasar a la definición, recordemos algunos conceptos. Sea x_1, \dots, x_n un subconjunto en el espacio $C(I)$ de funciones reales continuas sobre el intervalo compacto I . El gráfico de una función x es un subconjunto del plano

$$G(x) = \{(t, x(t)) : t \in I\}.$$

Una banda en \mathbb{R}^2 está determinada por las curvas x_{i_1}, \dots, x_{i_k} mediante

$$V(x_{i_1}, \dots, x_{i_k}) = \left\{ (t, y) : t \in I, \min_{1 \leq r \leq k} x_{i_r}(t) \leq y \leq \max_{1 \leq r \leq k} x_{i_r}(t) \right\}. \quad (1.3)$$

Para una función cualquier $x \in \{x_1, \dots, x_n\}$ la cantidad

$$S_n^j = \binom{n}{j}^{-1} \sum_{1 \leq i_1 < i_2 < \dots < i_j \leq n} \mathcal{I} \{G(x) \subset V(x_{i_1}, \dots, x_{i_j})\}, \quad j \geq 2, \quad (1.4)$$

expresa la cantidad de bandas $V(x_{i_1}, \dots, x_{i_j})$ determinadas por j curvas diferentes x_{i_1}, \dots, x_{i_j} conteniendo el gráfico de x .

Definición 4. Para funciones x_1, \dots, x_n , la Profundidad de Bandas muestral de cualquiera de estas curvas x es

$$DB_{n,J}(x) = \sum_{j=2}^J S_n^j(x), \quad J \geq 2. \quad (1.5)$$

Si X_1, \dots, X_n son copias independientes del proceso estocástico X que genera las observaciones x_1, \dots, x_n , la versión poblacional de la Profundidad de Bandas se define

$$DB_J(x) = \sum_{j=2}^J S^j(x) \quad \text{con} \quad S^j(x) = P\left(G(x) \subset V(X_{i_1}, \dots, X_{i_j})\right).$$

La profundidad de bandas DB_J es una profundidad de tipo A, es decir si escribimos,

$$h(x, x_1, x_2, \dots, x_J) = h_2(x, x_1, x_2) + h_3(x, x_1, x_2, x_3) + \dots + h_J(x, x_1, x_2, \dots, x_J)$$

con $h_j(x_1, x_2, \dots, x_j) = I\{x \in V(x_1, x_2, \dots, x_j)\}$. Los autores prueban consistencia fuerte y las propiedades **P.2.-P.3.** en el caso finito dimensional cuando la distribución P es absolutamente continua y **P.4.** para cualquier distribución. Para el caso funcional sólo prueban **P.4** cuando el proceso estocástico X vive en el espacio $C(I)$ y un resultado de consistencia fuerte cuando las funciones se encuentran en un conjunto equicontinuo de $C(I)$. En lugar de considerar la función indicadora en la Definición 4, una definición más flexible se puede introducir tomando la medida del conjunto donde la función está dentro de la banda correspondiente. Para cualquier función x en el conjunto $\{x_1, \dots, x_n\}$, sea

$$A_j(x) = A_{\{x_{i_1}, \dots, x_{i_j}\}}(x) = \left\{ t \in I : \min_{i_1 \leq r \leq i_j} x_r(t) \leq x(t) \leq \max_{i_1 \leq r \leq i_j} x_r(t) \right\}, \quad j \leq 2,$$

el conjunto de puntos en el intervalo I donde la función x está en la banda determinada por las observaciones x_1, \dots, x_n . Si λ es la medida de Lebesgue en I , $\lambda_r(A_j(x)) = \frac{\lambda(A_j(x))}{\lambda(I)}$ representa la proporción del tiempo que x está en la banda. Luego,

$$GS_n^j = \binom{n}{j}^{-1} \sum_{1 \leq i_1 < i_2 < \dots < i_j \leq n} \lambda_r(A(x; x_{i_1}, \dots, x_{i_j})), \quad j \geq 2, \quad (1.6)$$

es una versión generalizada de $S_n^j(x)$. Si x siempre está dentro de la banda, el valor de $\lambda_r(A_j(x))$ es 1 y por lo tanto la Ecuación (1.6) extiende la definición anterior dada por la Ecuación (1.4), dando lugar a la definición de Profundidad de Bandas Modificada.

Definición 5. Sea x_1, \dots, x_n , se define la Profundidad de Bandas Modificada muestral de cualquiera de estas curvas x como

$$MBD_{n,J}(x) = \sum_{j=2}^J GS_n^j(x), \quad J \leq 2.$$

Si X_1, \dots, X_n son copias independientes del proceso X que provee las observaciones x_1, \dots, x_n la versión poblacional de la Profundidad de Bandas Modificada como

$$MBD_J(x) = \sum_{j=2}^J GS^j(x) \quad J \geq 2,$$

donde $GS^j(x) = E[\lambda_r(Ax; X_{i_1}, \dots, X_{i_j})]$, $j \geq 2$.

La principal diferencia entre la Profundidad de Bandas y su generalización es que la primera tiene mayor dependencia en la forma de las curvas y es más restrictiva, generando muchos empates en la versión muestral. La segunda se ve más influenciada por la magnitud de las curvas en vez de la forma. Otra diferencia relevante entre ambas profundidades es el comportamiento que cada una tiene respecto a las curvas que dejan el centro de la muestra por un período muy corto de tiempo, es decir que se mantienen casi todo el tiempo en el centro, pero toman valores extremos en pequeños subintervalos de I . La Profundidad de Bandas es sensible a este tipo de contaminación y les asigna valores chicos, mientras que la Profundidad de Bandas Modificada les permite mantener valores relativamente altos.

1.1.5. Profundidad Dual Integrada (IDD)

Cuevas y Fraiman en 2009 (14), presentan una definición de medidas de profundidad para datos en un espacio general de Banach separable, basada en proyecciones continuas unidimensionales.

Sea Ω un espacio de probabilidad y \mathbb{E} un espacio de Banach separable con dual \mathbb{E}' separable. Sea $X : \Omega \rightarrow \mathbb{E}$ un elemento aleatorio en \mathbb{E} con distribución P . Dada una medida de probabilidad P^1 en \mathbb{R} y $u \in \mathbb{R}$ sea $D(P^1, u)$ el valor que indica la profundidad de u con respecto a P^1 , por ejemplo tomando la Profundidad Simplicial tenemos que,

$$SD(u, P^1) = F^1(u)(1 - F^1(u^-)), \quad (1.7)$$

donde F^1 es la función de distribución acumulada de P^1 . Esta es la profundidad unidimensional que consideraremos de aquí en adelante. También se puede usar la Profundidad de Tukey

$$HD_T(u, P^1) = \min(F^1(u), 1 - F^1(u^-)).$$

Definición 6. Definimos la Profundidad Dual Integrada (IDD) como

$$D(x, P) = \int D(P_f, f(x)) dQ(f), \quad (1.8)$$

donde $f \in \mathbb{E}'$, $x \in \mathbb{E}$ y $P_f = P_{f(X)}$ es la distribución de una variable aleatoria $f(X)$ y Q es una medida de probabilidad en \mathbb{E}' independiente de P .

Observación 2. Si \mathbb{E} es un espacio infinito dimensional Q se puede elegir como una medida Gaussiana no degenerada. En caso de que la dimensión sea finita tomamos Q como la Medida de Haar, que es la distribución uniforme en la bola unitaria.

Observación 3. En el caso que \mathbb{E} es un espacio de Hilbert \mathbb{H} , tenemos una isometría biyectiva $\Psi : \mathbb{H} \rightarrow \mathbb{H}'$ dada por $\Psi(h) = \langle h, \cdot \rangle$. Si $\varphi \in \mathbb{H}'$ y X un elemento aleatorio de \mathbb{H} , entonces $\varphi(X) = \langle h, X \rangle$ para un único $h \in \mathbb{H}$ y $P_{\varphi(X)} = P_h$ es la distribución de $\langle h, X \rangle$, que es la proyección del elemento aleatorio X en la dirección h . Luego, la Ecuación (1.8) queda

$$D(x, P) = \int D(\langle h, x \rangle, P_h) dQ(h), \quad (1.9)$$

donde Q es una medida de probabilidad en \mathbb{H} .

Definición 7. La mediana asociada con la Ecuación (1.8) será el punto más profundo

$$x_0 = \operatorname{argmáx}_{x \in \mathbb{E}} D(P, x).$$

Definición 8. Sean X_1, \dots, X_n elementos aleatorios i.i.d en \mathbb{E} con distribución P , sea P_n la distribución empírica y sea $P_{n,f}$ la distribución empírica de $f(X_1), \dots, f(X_n)$. Definimos la IDD Empírica como

$$D(x, P_n) = \int D(f(x), P_{n,f}) dQ(f), \quad (1.10)$$

y la mediana muestral como

$$\hat{x}_{o,n} = \operatorname{argmáx}_{x \in \mathbb{E}} D(x, P_n).$$

En la práctica para aproximar la Ecuación (1.10) consideramos $f_1, \dots, f_N \in \mathbb{E}'$ con distribución Q independiente de X_1, \dots, X_n y calculamos

$$\tilde{D}(x, P_n) = \frac{1}{N} \sum_{j=1}^N D(f_j(x), P_{n,f_j}). \quad (1.11)$$

Con esto podemos estimar la Ecuación (1.11) mirando el punto más profundo, \tilde{x}_0 , definido por

$$\operatorname{máx}_{x_{1 \leq j \leq n}} \{\tilde{D}(X_j, P_n)\}.$$

Observación 4. Respecto a las propiedades **P.1.-P.4.** nos encontramos con que esta profundidad en cualquier espacio de Banach separable cumple con **P.2-P.4** en el caso que la distribución P sea continua. Sin embargo, la invarianza afín (**P.1.**) no se cumple, lo que se tiene es invarianza bajo transformaciones ortogonales cuando el espacio sobre el que está definida tiene dimensión finita. Por otro lado, cumple con las dos propiedades de continuidad, **P.5.** y **P.6.**, cuando la distribución es continua.

Cabe destacar que con respecto a otras medidas de profundidad tiene una ventaja computacional que la hace viable cuando la dimensión de los datos es alta ya que se basa en proyecciones al azar. Alcanza con generar unos pocos (con respecto a la dimensión del espacio) funcionales en el dual para tener una buena aproximación de la profundidad.

1.2. Aplicaciones

Si bien varias aplicaciones se han desarrollado hasta ahora en el campo de las medidas de profundidad como hemos mencionado anteriormente al comienzo del capítulo, en esta sección haremos una descripción más detallada sobre las posibilidades que ellas ofrecen en el contexto de clasificación supervisada desde un prunto de vista no paramétrico.

A partir de la bibliografía que hemos podido reunir encontramos que el trabajo donde las profundidades se utilizan en este contexto pertenece a Regina Liu (1990) (35) donde se propuso el clasificador de máxima profundidad. La idea es la siguiente, dadas dos probabilidades P , Q y D una función de profundidad se clasifica a un punto x como proveniente de P si $D(x, P) > D(x, Q)$. El procedimiento se encuentra desarrollado en su totalidad en Ghosh y Chaudhuri (2005) (23).

En el sentido más puro lo que se intenta es tener una herramienta para comparar dos distribuciones o muestras multivariadas. Una herramienta que toma protagonismo cuando el enfoque se basa en profundidades es el DD-plot que fue introducido por Liu et al (1999) (37). Consideramos dos conjuntos de datos $\{x_1, \dots, x_n\}$ e $\{y_1, \dots, y_m\}$ generados por las distribuciones P y Q en \mathbb{R}^p respectivamente. Un DD-plot es un gráfico de dispersión entre las profundidades de las observaciones $\{x_1, \dots, x_n, y_1, \dots, y_m\}$ respecto de la profundidad P versus las profundidades del mismo conjunto de observaciones respecto de la profundidad Q , donde las profundidades provenientes de cada muestra deben ser distinguidas, por ejemplo utilizando distinto color. Es decir que sin importar la dimensión del espacio al cual pertenecen las observaciones, el DD-plot siempre es un gráfico bidimensional.

Se muestra en Liu et al. (1999) (37) que algunas diferencias en las distribuciones como medias, escala, kurtosis, etc, están asociadas con patrones característicos en el DD-plot. Es decir, que es un herramienta gráfica simple para comparar dos muestras de cualquier dimensión. Sin embargo, una aplicación del DD-plot que genera mayor atención está relacionada con el problema de clasificación. Si observamos la Figura 1.1, en el panel (a) tenemos un conjunto de datos formado por 400 puntos generado a partir de dos distribuciones distintas $P \sim N_2(\mu_1, I_2)$ y $Q \sim N_2(\mu_2, I_2)$ donde I_2 es la matriz identidad de 2×2 . En el panel (b) de la misma figura, se puede apreciar como funciona el clasificador de máxima profundidad, los puntos serán clasificados como provenientes de P o Q dependiendo si se encuentran por arriba o debajo de la recta punteada. Como se ve en ese mismo gráfico casi todos los puntos quedan bien clasificados mostrando que el procedimiento es óptimo para este caso.

Al observar la Figura 1.2 tenemos un conjunto de datos de 400 puntos, pero esta vez, la diferencia en las distribuciones no se encuentra en la media, si no en la dispersión. Con este

objetivo se generaron 200 puntos cada uno de ellos provenientes de dos normales multivariadas con media $\mu = 0$ y matrices de covarianza $\Sigma_1 = I_2$ y $\Sigma_2 = 4I_2$, respectivamente. Al observar el panel (b) de la Figura 1.2, apreciamos que el DD-plot es capaz de captar la diferencia en las distribuciones, pero el clasificador de máxima profundidad asigna a casi todos los puntos a la distribución con menor varianza. Una conclusión interesante podemos extraer de este último ejemplo, como se menciona en Li et al. (2012) (34), aunque el clasificador de máxima profundidad falle el DD-plot puede contener información que nos permita obtener un buen clasificador. Por ejemplo, para el segundo ejemplo podría utilizarse un clasificador basado en vecinos más cercanos aplicado al DD-plot.

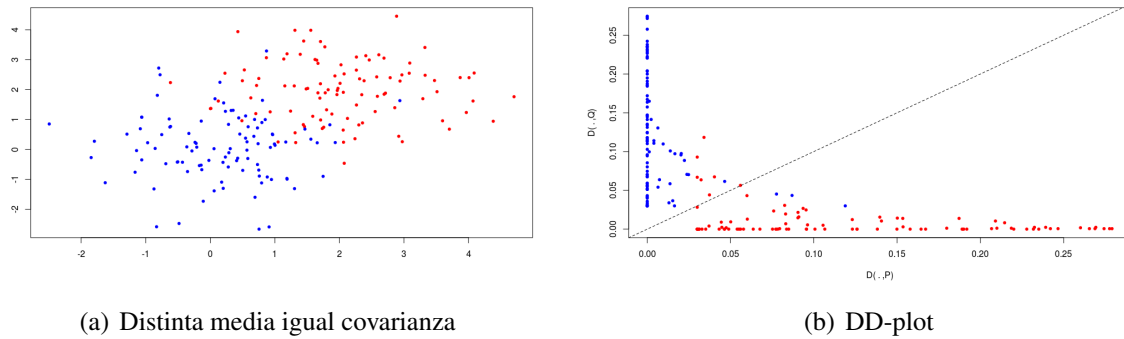


Figura 1.1: En rojo: los datos provienen de un distribución normal bivariada con media $\mu_1 = (0, 0)$ y $\Sigma = I_2$. En azul: los datos provienen de una distribución normal bivariada con media $\mu_2 = (2, 2)$ y $\Sigma = I_2$. Siendo I_2 la matriz identidad de 2×2 .

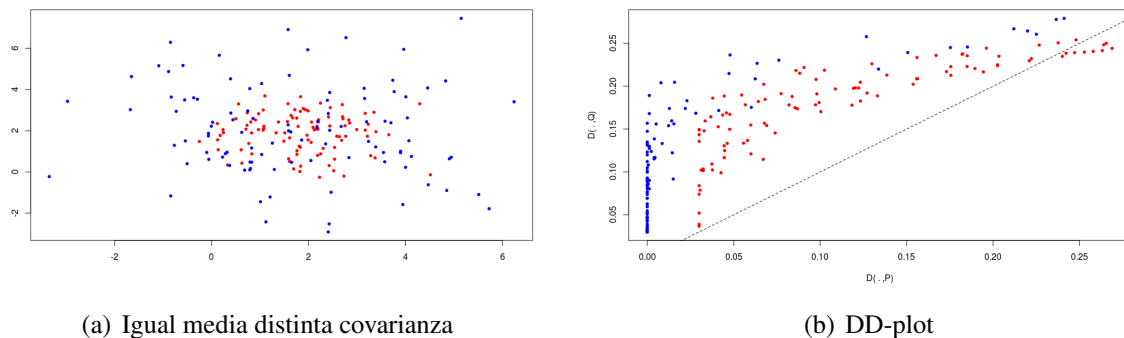


Figura 1.2: En rojo: los datos provienen de un distribución normal bivariada con media $\mu = (0, 0)$ y $\Sigma_1 = I_2$. En azul: los datos provienen de una distribución normal bivariada con media $\mu = (0, 0)$ y $\Sigma_2 = 4I_2$.

Valiéndose de estas ideas Cuesta-Albertos et al (2017) (11) proponen un nuevo clasifica-

dor llamado DD^G -classifier. El mismo, trata de resolver dos inconvenientes del clasificador de máxima profundidad de manera unificada. El primero es extraer un buen clasificador del DD-plot y el segundo es extender esta misma noción al caso en que tenemos más de dos poblaciones. El método puede aplicarse tanto en el caso de datos funcionales como multivariados, la limitación queda impuesta por las funciones de profundidad que se vayan a considerar. Describamos brevemente el clasificador: supongamos que se tienen K distribuciones de probabilidad en un espacio \mathbb{E} que puede ser finito o infinito dimensional y consideremos D_1, \dots, D_g funciones de profundidad definidas en el espacio \mathbb{E} . El DD^G -classifier construye la siguiente función

$$\begin{aligned} M^G : \mathbb{E} &\longrightarrow \mathbb{R}^G \\ x &\longmapsto d = (\vec{D}_1(x), \dots, \vec{D}_g(x)), \end{aligned}$$

donde $G = Kg$, $\vec{D}_j(x) = (D_j(x, P_1), \dots, D_j(x, P_K))$ denota al vector K -dimensional que corresponde a la j -ésima función de profundidad considerada $1 \leq j \leq g$ y P_1, \dots, P_K son las distribuciones de cada población.

La función M^G traduce la información de \mathbb{E} en un vector de dimensión G , una característica que es de especial interés en el contexto de datos funcionales o de alta dimensión. El último paso en el DD^G es seleccionar una regla de clasificación apropiada para la imagen de M^G lo cual es un problema puro de clasificación multivarado en los cuales se conocen una gran variedad de técnicas para llevar a cabo este trabajo.

1.3. Profundidades Locales

Las medidas de profundidad son una herramienta muy útil cuando analizamos un conjunto de datos complejo, como es el caso, cuando las observaciones caen en un espacio funcional o de alta dimensión. Como vimos anteriormente el objetivo principal de las profundidades es dar un orden del centro hacia afuera generalizando el concepto de mediana en un contexto general.

Una de las principales características es que el valor de la profundidad decrece sobre cualquier semirecta que tenga como punto de inicio el centro de la distribución. Lo que lleva a que las funciones de profundidad no puedan detectar multimodalidad (ver Figura 1.3) o no se comporten de manera deseada cuando el soporte de la distribución no es convexo. Sin embargo, la existencia de múltiples centros, su localización e importancia relativa son de interés tanto desde la perspectiva teórica como aplicada. Es el caso de las distribuciones mixtas (por ejemplo, datos reales y datos funcionales) y en análisis de clusters, entre otros.

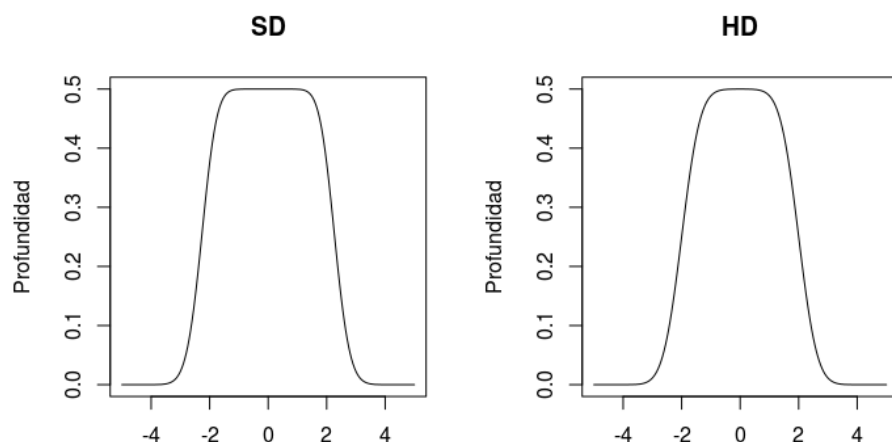


Figura 1.3: Funciones de profundidad del semiespacio (HD) y simplicial (SD) univariadas para una mixtura normal con modas en -2 y 2 . En este gráfico se puede apreciar que el punto más profundo está en el 0 y ninguna de las profundidades tiene la capacidad de detectar la multimodalidad.

Con el espíritu de captar estas características en los últimos años se han introducido definiciones de profundidades locales. Por el momento, en la literatura hay muy pocos trabajos sobre este nuevo concepto: Agostinelli y Romanazzi (2), Pandeveine y Van Bever (45) y Agostinelli (1). Aunque todavía es un campo poco explorado y no se cuenta con una definición formal de profundidad local a diferencia de la profundidad global donde Zou y Serfling (57) dan una definición formal de las mismas y realizan un profundo análisis sobre las propiedades que las mismas deben satisfacer. El punto en común entre estos autores se encuentra en que las profundidades locales deben ser una familia de funciones indexadas por un parámetro de localidad. Valores pequeños del mismo describen centralidad del punto relativo a un pequeño entorno alrededor de sí mismo. Mientras que al aumentar el valor de este parámetro la descripción de la centralidad se parece cada vez más a la que se obtendría al utilizar una función de profundidad global.

Un enfoque alternativo para sortear los mismos problemas es el que proponen Kotík y Hlubinka (32) al considerar una versión pesada de la profundidad del semiespacio, en la cual la función indicadora del semiespacio cerrado se reemplaza por una función de peso más general. Cheng et al. (8), introducen la profundidad espacial basada en núcleos positivos, utilizando el hecho de que los mismos brindan una medida de similaridad entre dos puntos. Luego en lugar de evaluar la profundidad en los puntos del espacio lo hacen en un espacio de características inducido por el núcleo. Estas propuestas pierden la naturaleza de ser medidas de centralidad.

Nuestro enfoque del problema situaría a las profundidades locales como una familia de

funciones $\{LD^\beta\}_{\beta \in [0,1]}$ que tienen un comportamiento relativo a la distribución subyacente, donde β es el índice de localidad. Si existe algún tipo de simetría global en los datos, entonces todos los integrantes de la familia asignan el mismo orden, desde el punto de simetría hacia afuera comportándose para cualquier valor del parámetro de localidad como una profundidad global.

En el caso que haya simetrías locales no todos los integrantes podrán detectar el centro de simetría local, dependerá del valor asignado al parámetro β , el cual tiene una relación estrecha con la concentración de la probabilidad alrededor de dicho punto. Dentro de la familia de profundidades locales siempre está incluida una profundidad global que se obtiene cuando tomamos $\beta = 1$. A continuación haremos una reseña de las propuestas existentes en la literatura.

1.3.1. Profundidad Local Simplicial y Profundidad Local del Semiespacio para datos multivariados.

El trabajo publicado por Agostinelli y Romanazzi en el año 2011 (2) es el primero en esta área. En este artículo se define una versión local para la Profundidad Simplicial y otra para la Profundidad del Semiespacio como describiremos a continuación.

Consideremos un vector aleatorio X , p -dimensional, con función de distribución F . Sean X_1, \dots, X_{p+1} vectores aleatorios independientes con la misma distribución que X . Notemos $S_{p+1} = S[X_1, \dots, X_{p+1}]$ al simplex aleatorio cerrado cuyos vértices son X_1, \dots, X_{p+1} . Recordemos que la Profundidad Simplicial (SD) para un punto x es la probabilidad de que ese simplex cubra a x . La medida común de tamaño para un subconjunto de \mathbb{R}^p es su volumen, notaremos $t(S_{p+1})$ al volumen de S_{p+1} . Naturalmente, es una variable aleatoria, no negativa, con una distribución heredada de F . En la definición de SD el tamaño de los simples no se tiene en cuenta. Es decir, que $SD(x)$ es la probabilidad de que x esté cubierto por simples de cualquier tamaño. Para lograr que la Profundidad Simplicial pueda captar efectos locales es necesario usar simples cuyo volumen no sea mayor a cierto umbral.

Definición 9. Sea τ un número positivo. La Profundidad Local Simplicial $LSD(\cdot, \tau, F)$ se define como

$$LSD(x, \tau, F) = P_F(x \in S_{p+1} \cap \{t(S_{p+1}) \leq \tau\}).$$

En el caso univariado, el simplex $S_2 = S[X_1, X_2]$ es un intervalo cerrado en la recta real con extremos X_1 y X_2 y $t(S_2)$ es la longitud de dicho intervalo. Por lo tanto, la profundidad local simplicial, es la probabilidad de que x esté en un intervalo de longitud τ . En la recta real es fácil ver dos propiedades que nos dejan entender mejor la idea de los autores respecto de lo que esperan de las profundidades locales. La primera es la relación con la profundidad global, pues

$$SD(x, F) = LSD(x, F, \tau) + P_F(S_2 \cap \{t(S_2) > \tau\});$$

lo cual implica que $LDS(x, \tau, F) \leq SD(x, F)$, $\forall \tau > 0$.

El segundo resultado, también se desprende de la Definición 9 puede probarse sin mayor dificultad

$$LDS(x, \tau, F) = \int_{x-\tau}^x (F(y+\tau) - F(x))f(y)dy + \int_x^{x+\tau} (F(x) - F(y-\tau))f(y)dy \quad (1.12)$$

Con lo cual el valor de $LDS(x, \tau, F)$ depende solamente de las propiedades de F dentro de un intervalo cerrado $[x - \tau, x + \tau]$. Por lo tanto, se justifica el uso de la profundidad local para condicionar la distribución subyacente a un entorno simétrico de x . El papel que juega el parámetro τ puede entenderse de las siguientes propiedades, es decir, para cualquier $x \in \mathbb{R}$ fijo:

- Si $0 < \tau_1 < \tau_2$, $LDS(x, F, \tau_1) \leq LDS(x, F, \tau_2)$.
- $\lim_{\tau \rightarrow +0} LDS(x, F, \tau) = P_F(x)$.
- $\lim_{\tau \rightarrow +\infty} LDS(x, F, \tau) = 0$.

Al igual que con la Profundidad Simplicial, en el trabajo definen una versión local para la Profundidad del Semiespacio (53). Notemos por $HS_u(a) = \{z \in \mathbb{R}^p : u'z \geq a\}$ el semiespacio cerrado que es uno de los lados correspondientes al hiperplano $HP_u(a) = \{z \in \mathbb{R}^p : u'z = a\}$. Fijado $\tau > 0$, la plancha cerrada $SL_u(a, a + \tau)$ entre dos hiperplanos paralelos $HP_u(a)$ y $HP_u(a + \tau)$ es la intersección entre el lado positivo de $HP_u(a)$ y la parte negativa de $HP_u(a + \tau)$, es decir

$$SL_u(a, a + \tau) = HS_u(a) \cap HS_{-u}(a + \tau) = \{z \in \mathbb{R}^p : a \leq u'z \leq u'z + a\}.$$

Luego, definen para un $x \in \mathbb{R}^p$ la Profundidad Local del Semiespacio como

$$\begin{aligned} LHD(x, \tau, F) &= \inf_{\{u \in \mathbb{R}^p : \|u\|=1\}} P_F(SL_u(u'x, u'x + \tau)) = \\ &= \inf_{\{u \in \mathbb{R}^p : \|u\|=1\}} P_F(z \in \mathbb{R}^p : u'x \leq u'z \leq u'x + \tau). \end{aligned}$$

Por último, para poder utilizar en la práctica este nuevo concepto definen una versión muestral de la Profundidad Local Simplicial. Sea X un vector aleatorio en \mathbb{R}^p con distribución absolutamente continua F . Sea X_1, \dots, X_n una muestra aleatoria con la misma distribución que X . Sea C el conjunto de todas las submuestras de tamaño $p + 1$ tomadas sin repetición y sea $S_{p+1}^n = S(X_{i_1}, \dots, X_{i_{p+1}})$ un simplex cerrado correspondiente a una submuestra. Se define la Profundidad Local Simplicial Empírica como

$$LDS(x, \tau, F_n) = \binom{n}{p+1}^{-1} \sum_C \mathcal{I}\{x \in \{S_{p+1}^n \cap t(S_{p+1}^n) \leq \tau\}\}.$$

Esta versión empírica resulta también en un estimador de la profundidad local simplicial, dado que los autores prueban la consistencia uniforme para el mismo.

Un problema que plantea este enfoque es la dificultad para poder elegir el parámetro de ajuste τ . No hay una relación entre τ y la distribución dado que el parámetro queda fijo para todos los puntos del espacio. Por ejemplo, si tuviésemos una mixtura normal dada por dos normales con distinta varianza y medias μ_1 y μ_2 no tenemos una guía sobre cómo se determina el valor de τ para poder captar las medias. Por otro lado, hay un problema con la dimensión. Si p es mayor que n la construcción hereda el mismo problema que tiene la Profundidad Simplicial y no puede calcularse. Además, es conocido que el cálculo se vuelve computacionalmente muy lento cuando la dimensión es grande.

1.3.2. Construcción general de profundidades locales a partir de profundidades para datos multivariados.

El trabajo de Paindeveine y Van Bever (45) introduce un nuevo concepto de profundidad local que, a cualquier nivel de localidad, mantiene su naturaleza de profundidad y provee una medida de centralidad local. La construcción planteada permite convertir de manera genérica cualquier profundidad global en \mathbb{R}^p en su correspondiente versión local. Esto le da una ventaja sobre otros competidores que se centran en una función específica. Esquemáticamente se podría decir que la propuesta se define como una profundidad global condicional a un entorno de interés. Para que este concepto se mantenga puramente basado en profundidades, ellos utilizan los entornos introducidos en Paindeveine y Van Bever (44) que describiremos a continuación.

Para cualquier función de profundidad las regiones de α -profundidad, $R_\alpha(P) = \{x \in \mathbb{R}^p : D(x, P) \geq \alpha\}$, son de gran importancia para revelar características de la distribución: posición, dispersión, estructura de dependencia, etc. Claramente, estas regiones están anidadas y sus regiones interiores contienen puntos de alta profundidad. Para arribar a un concepto local es más apropiado indexar la familia $\{R_\alpha\}$ en un sentido que esté más relacionado al contexto probabilístico. Es decir, para $\beta \in [0, 1]$, se define

$$R^\beta(P) = \bigcap_{\alpha \in A(\beta)} R_\alpha(P), \quad \text{con } A(\beta) = \{\alpha \geq 0 : P(R_\alpha(P)) \geq \beta\}.$$

Es decir, $R^\beta(P)$ es la menor región de profundidad con probabilidad mayor o igual a β . Utilizando el subíndice (respectivamente: el superíndice) para notar las regiones de profundidad asociadas con un orden fijo α (respectivamente: con una probabilidad fija β).

A diferencia del contexto global, la propuesta de estos autores involucra entornos para cada $x \in \mathbb{R}^p$. Las regiones $R_\alpha(P)$ o $R^\beta(P)$ proveen entornos del punto más profundo solamente, por lo que no se pueden utilizar para la propuesta. Sin embargo, entornos para cualquier $x \in \mathbb{R}^p$ pueden obtenerse reemplazando P por P_x su versión simetrizada en x . Más específicamente, si llamamos $P = P^X$ a la medida de probabilidad con la que vamos a trabajar y

entendemos que X es un vector asociado a esa distribución, se define $P_x = \frac{1}{2}P^X + \frac{1}{2}P^{2x-X}$. Reemplazando P por P_x , se obtienen entornos $R_\alpha(P_x) = R_{x,\alpha}(P)$ y $R^\beta(P_x) = R_x^\beta(P)$ que tienen naturaleza no paramétrica. Además podemos observar que α o β juegan un papel de parámetros de localidad.

Basados en esta nueva parametrización de la localidad los autores proponen una manera de definir una profundidad local a partir de su contraparte global.

Definición 10. Sea $D(\cdot, P)$ una función de profundidad. La correspondiente profundidad local de nivel $\beta \in (0, 1]$ es

$$LD^\beta(x, P) = D(x, P_x^\beta),$$

donde P_x^β es la medida de distribución condicional de P dado R_x^β .

Es evidente por la definición que la parametrización en β resulta más adecuada, que la parametrización en α . Se detallan dos razones, primero el $\alpha_\star = \max_{x \in \mathbb{R}^p} D(x, P)$ depende de P . Segundo, el entorno $R_x^\alpha(P)$ puede tener probabilidad 0 para α cercano a α_\star . Un ejemplo se puede construir tomando $x = 0 \in \mathbb{R}^p$, X una normal estandar multivariada y P la distribución condicional de X dado $\{\|X\| > 1\}$. En contraste con la situación anterior, la parametrización en β está siempre bien definida y el rango de valores no depende de P o D . Además, para $\beta = 1$ la profundidad local se reduce a su antecesora global, es decir se recupera la función original.

Al igual que con todas las funciones de profundidad se define una versión empírica de la misma que requerirá para cualquier $x \in \mathbb{R}^p$, la distribución empírica simetrizada $P_{x,n}$ que no es más que la distribución empírica asociada con $X_1, \dots, X_n, 2x - X_1, \dots, 2x - X_n$. Se define la versión empírica de la profundidad local como

Definición 11. Sea $D(\cdot, P)$ una función de profundidad global, la correspondiente función de profundidad local empírica de nivel β se define como

$$LD_n^\beta(x, P) = LD^\beta(x, P_n) = D(x, P_{x,n}^\beta).$$

Observemos que por definición, $R_x^\beta(P_n)$ es la menor región de profundidad que contiene a por lo menos una porción β de los $2n$ vectores aleatorios $X_1, \dots, X_n, 2x - X_1, \dots, 2x - X_n$ o equivalentemente una proporción β de las n originales X_1, \dots, X_n .

Bajo esta definición los autores aseguran que el método es consistente en el siguiente sentido: fijado $x \in \mathbb{R}^p$ y $D(\cdot, P)$ que satisface **(P.2)**, **(P.3)** y **(P.5)**, entonces para cualquier P absolutamente continua y cualquier sucesión $\beta_n \rightarrow \beta$ se tiene que

$$LD_n^\beta(x, P_n) \xrightarrow[n \rightarrow +\infty]{P\text{-c.t.p.}} LD^\beta(x, P).$$

Un ejemplo unidimensional para entender como funciona la adaptación global al contexto se puede mostrar a partir de la Profundidad Simplicial univariada. Recordemos $SD(x, F) =$

$2F(x)(1 - F(x))$, donde F es la función de distribución acumulada de X una variable aleatoria (que no es la versión simetrizada para más detalles ver Paindeveine y Van Bever (45)). Sea

$$x_\beta = x - \inf\{h > 0 : F(x + \lambda) - F(x - \lambda) \geq \beta\},$$

resulta

$$LD_S^\beta(x, F) = \frac{2}{\beta^2} (F(x) - F(x_\beta)) (F(2x - x_\beta) - F(x)).$$

Si llamamos $\lambda_x^\beta = \inf\{h > 0 : F(x + \lambda) - F(x - \lambda) \geq \beta\}$ la ecuación anterior se reduce a

$$LD_S^\beta(x, F) = \frac{2}{\beta^2} (F(x) - F(x - \lambda_x^\beta)) (F(x + \lambda_x^\beta) - F(x)),$$

una fórmula que utilizaremos a menudo en el Capítulo 2. Observemos también que λ juega un papel similar al parámetro τ para la definición dada por Agostinelli y Romanazzi (2). La ventaja de este enfoque es que no debe permanecer constante para todos los puntos, se adapta a cada uno de ellos basándose en la parametrización de la localidad, es decir, β . En particular tiene una buena adaptación si los datos provienen de distribuciones con distinta dispersión, solo importa la centralidad local.

Entre las aplicaciones que se proponen en este trabajo podemos destacar la aplicación a problemas de clasificación sobre todo en contextos donde el soporte de la distribución no es convexo, la construcción de test de simetría y la extensión a problemas de regresiones basadas en profundidades locales.

1.3.3. Profundidad Local del Semiespacio para Datos Funcionales

Los dos trabajos que describimos anteriormente se basan en que los datos se encuentren en \mathbb{R}^p , sin embargo hoy en día el estudio de las funciones de profundidad (globales) traspasan la barrera de la dimensión finita. La pregunta natural sería es posible tener una profundidad local para el contexto de datos funcionales o el contexto de alta dimensión. Por el momento en la literatura sólo hay un trabajo escrito al respecto cuyo autor es Claudio Agostinelli (1) que detallamos a continuación, donde se extienden las ideas planteadas por Agostinelli y Romanazzi (2) para datos multivariados.

Un típico modelo para datos funcionales es una función $x = x(t)$, $t \in T$, que pertenece al espacio $C(T)$ de funciones reales continuas de un intervalo compacto T en los reales. El gráfico de una función x es un subconjunto de $G(x) = \{(t, x(t)) : t \in T\}$. El hipografo (hyp) y el epigrafo (epi) de una función x se definen de la siguiente manera:

$$hyp(x) = \{(t, z) : t \in T, z \leq x(t)\} \text{ y } epi(x) = \{(t, z) : t \in T, z \geq x(t)\}.$$

Una versión local de la profundidad del semiespacio se puede obtener imponiendo una restricción en el tamaño de hipografo y el epigrafo de una función x . Más formalmente,

$$hyp(x, \tau) = hyp(x) \cap epi(x - \tau) = \{(t, z) : t \in T, x(t) - \tau \leq z \leq x(t)\},$$

análogamente la misma restricción se puede imponer para el epigrafo

$$epi(x, \tau) = epi(x) \cap epi(x + \tau) = \{(t, z) : t \in T, x(t) \leq z \leq x(t) + \tau\}.$$

Definición 12. Para una proceso estocástico X la Profundidad Local del Semiespacio para una función x se define como

$$LHRD(x, X, \tau) = \min\{P(G(X) \subset hyp(x, \tau)); P(G(X) \subset epi(x, \tau))\}.$$

La versión empírica se obtiene dadas n realizaciones independientes del proceso estocástico X , llamemoslas x_1, \dots, x_n ,

$$LHRD_n(x, X, \tau) = \min\left(\frac{1}{n} \sum_{i=1}^n \mathcal{I}((G(x_i) \subset hyp(x, \tau)); \frac{1}{n} \sum_{i=1}^n \mathcal{I}((G(x_i) \subset epi(x, \tau)))\right).$$

Agostinelli muestra que bajo condiciones de regularidad la profundidad local es fuertemente consistente.

Por otra parte, basado en el concepto de profundidad local, introduce una medida de similaridad local, la cual utiliza para resolver problemas de cluster jerárquico.

Capítulo 2

Profundidad Local Dual Integrada

2.1. Introducción

En este Capítulo nuestro objetivo es dar una definición general para una profundidad local para elementos aleatorios en un espacio de Banach, extendiendo la profundidad global dada por Cuevas y Fraiman (14) donde introducen la Profundidad Dual Integrada (IDD). Recordemos que la misma se basa en combinar proyecciones al azar sobre espacios unidimensionales y luego aplicar alguna profundidad conocida. Más específicamente, sea $(\Omega, \mathcal{A}, \mathbb{P})$ un espacio de probabilidad y sea \mathbb{E} un Banach separable con dual \mathbb{E}' . Sea $X : \Omega \rightarrow \mathbb{E}$ un elemento aleatorio con distribución P y sea Q una medida de probabilidad en \mathbb{E}' independiente de P . En el Capítulo 1 habíamos definido la IDD (ver Definición 1.8), en lo que sigue definimos la Profundidad Dual Local Integrada (IDL), basándonos en reemplazar la profundidad global de la Ecuación (1.8) por una profundidad local unidimensional siguiendo la definición dada por Paindavaine y Van Bever (45). Estudiamos como las propiedades clásicas, introducidas por Zuo y Serfling (57), deben ser analizadas en este nuevo contexto. Probamos que bajo condiciones de regularidad la propuesta cumple con estas propiedades. Además, obtenemos resultados de consistencia fuerte para la versión empírica de la profundidad local con respecto a la versión poblacional de la misma y también para las regiones de profundidad locales empíricas respecto de su contraparte poblacional.

Las principales ventajas de nuestra propuesta son, por un lado, la flexibilidad para trabajar con datos en contextos muy generales y por otra parte, el bajo costo computacional que permite el análisis, en un tiempo razonable, cuando la dimensión es alta.

2.2. Marco de Referencia y Definiciones

En esta sección, primero recordaremos el concepto de profundidad local en el caso univariado. Luego, daremos la definición para la Profundidad Local Dual Integrada y finalmente mostraremos que bajo condiciones de regularidad satisface las propiedades teóricas estable-

cidas en (57).

Sea P^1 una medida de probabilidad en \mathbb{R} , $x \in \mathbb{R}$ y sea $LD(x, P^1)$ una función de profundidad local de x con respecto a P^1 . Por ejemplo, la Profundidad Local Simplicial univariada tiene la siguiente fórmula,

$$LD_S^\beta(x, P^1) = \frac{2}{\beta^2} \left(F^1(x + \lambda_x^\beta) - F^1(x) \right) \left(F^1(x) - F^1(x - \lambda_x^\beta) \right), \quad (2.1)$$

donde F^1 es la función de distribución acumulada de P^1 , β y λ_x^β son parámetros que determinan la profundidad local, los mismos serán definidos a continuación.

Definición 13. Sea F^1 una función de distribución univariada y $x \in \mathbb{R}$. Notaremos $\beta \in (0, 1]$ al nivel de localidat. Dado $\beta \in (0, 1]$, definimos el correspondiente **ancho del entorno de profundidad** λ_x^β como

$$\lambda_x^\beta = \inf \{ \lambda > 0 : F^1(x + \lambda) - F^1(x - \lambda) \geq \beta \}. \quad (2.2)$$

Observación 5. Si F^1 es absolutamente continua el ínfimo de la Ecuación (2.2) se alcanza y de aquí,

$$\lambda_x^\beta = \min \{ \lambda > 0 : F^1(x + \lambda) - F^1(x - \lambda) \geq \beta \}.$$

Más aún, es claro que si $\beta_1 < \beta_2$, entonces $\lambda_x^{\beta_1} < \lambda_x^{\beta_2}$.

El nivel de localidat β , es un parámetro que determina la centralidad de un punto del espacio, x , condicionado a un entorno dado alrededor de sí mismo. Si el valor es alto, este aproxima el valor de la profundidad del punto mientras que si es chico describirá la centralidad en un pequeño intervalo alrededor de x . Cuando β tiende a 1 la profundidad local tiende a la profundidad global.

También definimos en forma análoga, la profundidad local de Tukey,

$$LD_H^\beta(x, P^1) = \frac{1}{\beta} \min \{ F^1(x + \lambda_x^\beta) - F^1(x), F^1(x) - F^1(x - \lambda_x^\beta) \}.$$

En adelante, sin pérdida de generalidad, nos restringiremos al caso de la profundidad local simplicial, LD_S^β .

2.2.1. Profundidad Local Dual Integrada

En esta sección nuestra meta es extender la IDD introducida por Cuevas y Fraiman (14), al contexto local. La IDD es una profundidad definida para elementos aleatorios en un espacio de Banach, la idea es proyectar los datos sobre direcciones aleatorias y calcular la profundidad univariada. Para obtener una profundidad global integramos las profundidades unidimensionales. Bajo condiciones de regularidad la IDD satisface las propiedades básicas para una función de profundidad (57) y es fuertemente consistente. Es importante remarcar que tiene un bajo costo computacional, aún en espacios de alta dimensión, dado que está basada en el cálculo de proyecciones sobre una dimensión.

Definición 14. Sea Ω un espacio de probabilidad y \mathbb{E} un espacio de Banach separable, notemos \mathbb{E}' al dual del Banach. Sea $X : \Omega \rightarrow \mathbb{E}$ un elemento aleatorio en \mathbb{E} con distribución P y Q una medida de probabilidad sobre \mathbb{E}' independiente de P , $\beta \in (0, 1]$ y $x \in \mathbb{E}$. Definimos la **Profundidad Local Dual Integrada (IDLD)**,

$$IDLD^\beta(x, P) = \int LD_S^\beta(f(x), P_f) dQ(f), \quad (2.3)$$

donde LD_S^β es la profundidad local univariada dada por la Ecuación (2.1), $f \in \mathbb{E}'$ y P_f es la distribución univariada de $f(X)$.

Sugieren Cuevas y Fraiman (14) para la IDD que en el caso infinito dimensional, Q puede ser una medida Gaussiana no degenerada y en el caso multivariado se puede tomar la distribución uniforme sobre la esfera unidad. Nosotros hemos decidido utilizar una medida Gaussiana en ambos casos como se verá más adelante en el Capítulo 3 dedicado a las simulaciones y al estudio de ejemplos con datos reales. Por simplicidad, haciendo un ligero abuso de la notación, escribiremos a $F_f = F_{f(X)}$ la función de distribución acumulada de $f(X)$. Específicamente,

$$F_f(t) = F_{f(X)}(t) = P_{f(X)}((-\infty, t]) = P(f(X) \leq t).$$

Cabe aclarar además que de ahora en adelante cuando mencionemos que X es un elemento aleatorio continuo nos estaremos refiriendo a que para todo funcional no nulo en \mathbb{E}' , $f(X)$ es una variable aleatoria absolutamente continua.

Antes de continuar, veamos que la IDLD está bien definida, es decir que es acotada y no negativa. Sea X un elemento aleatorio en \mathbb{E} y $x \in \mathbb{E}$, es evidente que $IDLD^\beta(x, P) \geq 0$. Además, si X es continua para todo $x \in \mathbb{E}$,

$$\begin{aligned} \beta &= F_f(f(x) + \lambda_{f(x)}^\beta) - F_f(f(x) - \lambda_{f(x)}^\beta) \\ &= [F(f(x) + \lambda_{f(x)}^\beta) - F_f(f(x))] + [F_f(f(x)) - F_f(f(x) - \lambda_{f(x)}^\beta)]. \end{aligned} \quad (2.4)$$

Ambos términos en la Ecuación (2.4) son positivos el producto está acotado por $\frac{\beta^2}{4}$. Entonces,

$$\begin{aligned} LD_S^\beta(f(x), P_f) &= \frac{2}{\beta^2} [F_f(f(x)) - F_f(f(x) - \lambda_{f(x)}^\beta)] [F_f(f(x) + \lambda_{f(x)}^\beta) - F_f(f(x))] \\ &= \frac{2}{\beta^2} [F_f(f(x) + \lambda_{f(x)}^\beta) - F_f(f(x))] [\beta - (F_f(f(x) + \lambda_{f(x)}^\beta) - F_f(f(x)))] \leq \\ &\leq \frac{2}{\beta^2} \frac{\beta^2}{4} = \frac{1}{2}. \end{aligned}$$

Con lo cual,

$$|IDLD^\beta(x, P)| \leq \int |LD_S^\beta(f(x), P_f)| dQ(f) \leq \int \frac{1}{2} dQ(f) = \frac{1}{2}. \quad (2.5)$$

Zuo y Serfling (57) establecieron las propiedades generales que una función de profundidad debe satisfacer. Describiremos que propiedades son cumplidas por la IDLD.

La primer propiedad que analizaremos para la IDLD es la invarianza afín. Recordemos que esta propiedad no se cumple en general para la IDD (i.e. sea A una transformación lineal en \mathbb{R}^p y P_{AX} denota la distribución de AX entonces $D(Ax, P_{AX})$ no es igual a $D(x, P_X)$). Cuevas y Fraiman, muestran que es independiente del sistema de coordenadas para el caso finito dimensional. Dado que la IDLD es una generalización de la IDD, tampoco será invariante afín, pero sin embargo heredará la propiedad de ser invariante respecto del sistema de coordenadas. Por otro lado, es claro que IDLD, es invariante bajo translaciones y cambios de escala.

P. 1. (invarianza afín). : Sea $\mathbb{E} = \mathbb{R}^d$ con la norma euclídea y $X \in \mathbb{E}$ un vector aleatorio, sea Q la medida uniforme en la esfera unidad de \mathbb{E}' independiente de P_X . Sea $A : \mathbb{E} \rightarrow \mathbb{E}$ una transformación ortogonal y $\beta \in (0, 1]$. Entonces $IDLD^\beta(Ax, P_{AX}) = IDLD^\beta(x, P_X)$.

Demostración. Sea A una transformación ortogonal y Q_1 denota la medida de probabilidad de $h = f \circ A$.

$$IDLD^\beta(Ax, P_{AX}) = \int LD_S^\beta(f(Ax), P_{f(Ax)})Q(f) = \int LD_S^\beta(h(x), P_{h(x)})Q_1(h).$$

Como Q es la medida uniforme sobre la esfera unidad, $Q = Q_1$ y se cumple la igualdad deseada.

Hay contextos, cuando trabajamos con datos, donde la invarianza afín es una propiedad relevante. Es bien sabido que la mediana espacial tampoco es invariante afín por lo que Chakraborty y Chaudhuri ((5) y (6)), desarrollaron métodos de transformación y retransformación para construir medianas multivariadas invariantes afines. Por otro lado, Kotík and Hlubinka (32) muestran como es posible, para la profundidad que ellos proponen, redefinirla para que sea invariante afín. Una condición necesaria para este proceso es que la profundidad sea invariante por transformaciones ortogonales. Siguiendo las ideas expuestas anteriormente, damos una definición de Profundidad Local Dual Integrada que cumple con la invarianza afín.

Sea X un vector aleatorio que cumple las mismas hipótesis de la Propiedad **(P.1)**. Además supongamos que su matriz de covarianza Σ_X existe y es no singular.

Definición 15. Definimos una versión invariante afín de la Profundidad Local Dual Integrada como

$$IDLD_{af}^\beta(x, P_X) = IDLD^\beta(\Sigma_X^{-1/2}x, P_{\Sigma_X^{-1/2}X}), \quad (2.6)$$

donde $\Sigma_X^{-1/2}$ es la inversa de la raíz cuadrada de la matriz de covarianza de X , que está bien definida si la misma es definida positiva.

Veremos a continuación que nuestra modificación cumple con el objetivo planteado, como primer caso, supongamos que X es un vector aleatorio tal que $Var(X) = I$, I es la matriz identidad de $p \times p$ y A es una matriz en $\mathbb{R}^{p \times p}$ no singular.

Podemos observar que $Var(AX) = \Sigma_{AX} = AA'$ es simétrica y definida positiva, donde A' es la matriz traspuesta. Es más, $\Sigma_{AX}^{-1/2}A$ es una matriz ortogonal, pues

$$(\Sigma_{AX}^{-1/2}A)(\Sigma_{AX}^{-1/2}A)' = (AA')^{-1/2}AA'(AA')^{-1/2} = I.$$

Luego, aplicando la Propiedad **(P.1)**,

$$IDLD_{af}^\beta(AX, P_{AX}) = IDLD^\beta(\Sigma_{AX}^{-1/2}Ax, P_{\Sigma_{AX}^{-1/2}AX}) = IDLD_{af}^\beta(x, P_X). \quad (2.7)$$

Tengamos en consideración el caso general, cuando $Var(X) = \Sigma_X \neq I$, es una matriz simétrica definida positiva. Sabemos que es posible encontrar un vector aleatorio Y y una matriz no singular B tal que $X = BY$ con $Var(Y) = I$. Por ejemplo, tomando $B = \Sigma_X^{1/2}$, es decir la raíz cuadrada de la matriz de covarianza que está bien definida por ser la anterior definida positiva. Por el caso $Var(X) = I$ deducimos rápidamente que

$$IDLD_{af}^\beta(x, P_X) = IDLD_{af}^\beta(By, P_{BY}) = IDLD_{af}^\beta(y, P_Y) \text{ donde } y = B^{-1}x.$$

Finalmente, de la Ecuación (2.7) tenemos que,

$$IDLD_{af}^\beta(AX, P_{AX}) = IDLD_{af}^\beta(ABY, P_{ABY}) = IDLD_{af}^\beta(y, P_Y) = IDLD_{af}^\beta(x, P_X).$$

Las medidas de profundidad son una herramienta útil para el análisis especialmente en los casos donde los elementos aleatorios que se estén considerando tengan alguna propiedad de simetría. La profundidad local (restringida a un entorno) debería heredar esa característica. Como primer paso, para analizar este hecho debemos dar una apropiada definición de simetría local.

Definición 16. Sea X una variable aleatoria en la recta real y $\beta \in (0, 1]$, decimos que X es β -simétrico en θ si la distribución acumulada F satisface que,

$$F^1(\theta + \lambda_\theta^{\beta'}) - F^1(\theta) = \frac{\beta'}{2}, \text{ para todo } 0 < \beta' \leq \beta. \quad (2.8)$$

Un elemento aleatorio X en un espacio de Banach \mathbb{E} es β -simétrico en θ si para todo $f \in \mathbb{E}'$ no nulo,

$$F_f(f(\theta) + \lambda_{f(\theta)}^{\beta'}) - F_f(f(\theta)) = \frac{\beta'}{2}, \text{ para todo } 0 < \beta' \leq \beta.$$

La β -simetría intenta captar el comportamiento de una variable aleatoria multimodal en un entorno de probabilidad β alrededor de θ , el punto de mayor profundidad local. Las Figuras 2.1 (a) y (b) exhiben una distribución bimodal, con modas en $\theta = 1$ and $\theta = 4$. En la Figura 2.1 (a) podemos observar que ambas modas son puntos de simetría local para $\beta = 0,25$. Mientras que en la Figura 2.1 (b) se puede ver que $\theta = 4$ es un punto de simetría local para $\beta = 0,4$, pero $\theta = 1$ no lo es. Esto se deduce rápidamente observando que la región, alrededor de $\theta = 1$, que se requiere para poder alcanzar probabilidad $\beta = 0,4$ no resulta simétrica.

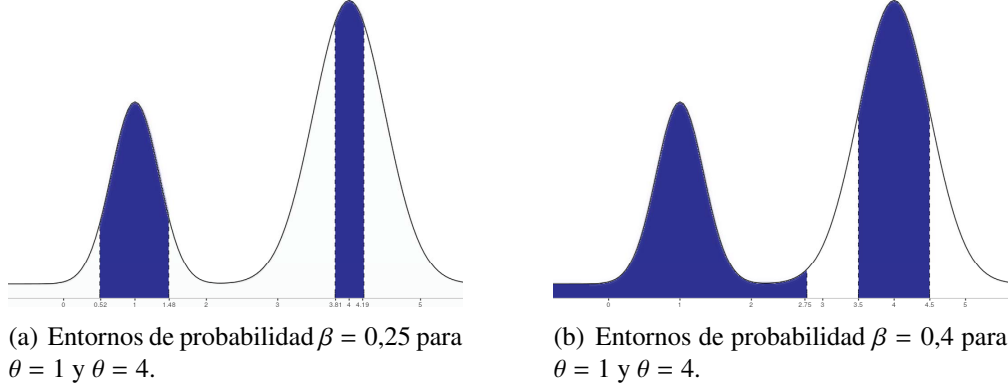


Figura 2.1: Puntos de simetría local.

Una propiedad importante de las medidas de profundidad es que es la maximalidad al centro, es decir, que si P es simétrica en θ entonces $D(x, P)$ consigue su máximo valor en ese punto. Esta cualidad debe ser heredada por las profundidades locales si la distribución P es unimodal y convexa. Además, deberíamos esperar que lo mismo suceda a nivel local, es decir, que se maximice en centros locales, en nuestro caso puntos de β -simetría.

P. 2. (maximiza en el centro). Sea $X \in \mathbb{E}$ un elemento aleatorio continuo y β -simétrico en θ . Para $\beta \in (0, 1]$ tenemos que,

$$IDLD^{\beta'}(\theta, P_X) = \max_{x \in \mathbb{E}} IDLD^{\beta'}(x, P_X), \text{ para todo } 0 < \beta' \leq \beta. \quad (2.9)$$

Demostración. Es suficiente con mostrar que siendo $f \in \mathbb{E}'$ y $\beta' \in (0, \beta]$

$$\begin{aligned} LD_S^{\beta'}(f(\theta), P_f) &= \frac{2}{\beta'^2} \left[F_f(f(\theta)) - F_f(f(\theta) - \lambda_{f(\theta)}^{\beta'}) \right] \left[F_f(f(\theta) + \lambda_{f(\theta)}^{\beta'}) - F_f(f(\theta)) \right] = \\ &= \frac{2}{\beta'^2} \frac{\beta'}{2} \frac{\beta'}{2} = \frac{1}{2}. \end{aligned}$$

Dado que alcanza la cota (2.5), $LD_S^{\beta'}(f(\theta), P_f)$ tiene un máximo global θ . Es claro que $IDLD^{\beta'}(\theta, P)$, también es un máximo global.

La Proposición 1 conecta la β -simetría con la definición usual de C -simetría.

Proposición 1. Sea $X \in \mathbb{E}$ un elemento aleatorio continuo C -simétrico en θ (ver definición 2), entonces X es β -simétrico en θ para cada $\beta \in (0, 1]$.

Demostración. Sea $\beta \in (0, 1]$, X es C -simétrico en θ si para todo $f \in \mathbb{E}'$, $f(X)$ es simétrica en $f(\theta)$. Entonces, para cada $0 < \beta' \leq \beta$,

$$F_f(f(\theta) + \lambda_{f(\theta)}^{\beta'}) - F_f(f(\theta)) = F_f(f(\theta)) - F_f(f(\theta) - \lambda_{f(\theta)}^{\beta'}).$$

Luego,

$$\begin{aligned}\beta' &= F_f(f(\theta) + \lambda_{f(\theta)}^{\beta'}) - F_f(f(\theta) - \lambda_{f(\theta)}^{\beta'}) \\ &= F_f(f(\theta) + \lambda_{f(\theta)}^{\beta'}) - F_f(f(\theta)) + F_f(f(\theta)) - F_f(f(\theta) - \lambda_{f(\theta)}^{\beta'}) \\ &= 2(F_f(f(\theta) + \lambda_{f(\theta)}^{\beta'}) - F_f(f(\theta))).\end{aligned}$$

Finalmente,

$$\frac{\beta'}{2} = F_f(f(\theta) + \lambda_{f(\theta)}^{\beta'}) - F_f(f(\theta)),$$

que es lo que queríamos mostrar.

La Proposición 2 describe la β -simetría de los puntos de X .

Proposición 2. *Sea X un elemento aleatorio en \mathbb{E} con medida de distribución asociada P y $x_0 \in \mathbb{E}$ tal que $LD^{\beta'}(x_0, P) = \frac{1}{2}$, para todo $0 < \beta' \leq \beta$. Entonces x_0 es un punto de β -simetría.*

Demostración. *Primero notemos que, dado $f \in \mathbb{E}'$, $x \in \mathbb{E}$ and $\beta \in (0, 1]$.*

$$\begin{aligned}\beta &= F_f(f(x) + \lambda_{f(x)}^{\beta}) - F_f(f(x) - \lambda_{f(x)}^{\beta}) \\ \beta &= F_f(f(x) + \lambda_{f(x)}^{\beta}) - F_f(f(x)) + F_f(f(x)) - F_f(f(x) - \lambda_{f(x)}^{\beta}) \\ F_f(f(x) + \lambda_{f(x)}^{\beta}) - F_f(f(x)) &= \beta - (F_f(f(x)) - F_f(f(x) - \lambda_{f(x)}^{\beta})).\end{aligned}$$

De la Ecuación (2.1) es claro que,

$$LD_S^{\beta}(f(x), P_f) = \frac{2}{\beta^2} [\beta - (F_f(f(x)) - F_f(f(x) - \lambda_{f(x)}^{\beta}))][F_f(f(x) + \lambda_{f(x)}^{\beta}) - F_f(f(x))].$$

Sea $h : [0, \beta] \rightarrow \mathbb{R}$, $h(t) = \frac{2}{\beta^2}(\beta - t)t$, alcanza un máximo global en $t = \frac{\beta}{2}$, por lo tanto LD_S^{β} que alcanza su máximo cuando $F_f(f(x) + \lambda_{f(x)}^{\beta}) - F_f(f(x)) = \frac{\beta}{2}$. Si esta propiedad se satisface para cualquier $0 < \beta' \leq \beta$, entonces, x es un punto de β -simetría de $f(X)$.

Sea $0 < \beta' \leq \beta$,

$$\frac{1}{2} = IDLD^{\beta'}(x_0, P) = \int LD_S^{\beta'}(f(x_0), P_f) dQ(f) \Rightarrow 0 = \int \left(\frac{1}{2} - LD_S^{\beta'}(f(x_0), P_f) \right) dQ(f).$$

Para todo $f \in \mathbb{E}$, $LD_S^{\beta'}(f(x_0), P_f)$ está acotado por $\frac{1}{2}$; lo que implica que $LD_S^{\beta'}(f(x_0), P_f) = \frac{1}{2}$ Q -c.t.p. La primera parte de la demostración nos dice que $f(x_0)$ es un punto de β -simetría de $f(X)$, por lo tanto x_0 es un punto de β -simetría de X .

Ahora nos enfocaremos en la propiedad que establece que la profundidad local es monótona respecto del punto más profundo. Mostraremos que este resultado se cumple cuando la distribución es unimodal. Pero antes necesitaremos algunos resultado auxiliares.

Observación 6. Una definición de unimodalidad dada por Dharmadhikari y Joag-dev (17) (ver Def. 1.1) es que una variable aleatoria Z o una función de distribución acumulada F se dice unimodal en una moda z_0 si F es convexa en $(-\infty, z_0)$ y concava en $(z_0, +\infty)$. Se desprende de la definición que si F es absolutamente continua, la unimodalidad de F en z_0 es equivalente a la existencia una función de densidad que es creciente en $(-\infty, z_0)$ y decreciente en $(z_0, +\infty)$.

Además aclaramos que en las siguientes resultados y demostraciones, cuando se mencione la función de densidad de una variable aleatoria absolutamente continua estaremos considerando el representante continuo de la densidad.

Lema 1. Sea Z una variable aleatoria absolutamente continua, simétrica y unimodal en $t = 0$ y sea F la función de distribución acumulada. Sea $\beta \in (0, 1]$, definimos las funciones,

- $U(t) = F(t + \lambda_t^\beta) - F(t)$.
- $V(t) = F(t) - F(t - \lambda_t^\beta)$.

Entonces, para todo $t \in \mathbb{R}$ se cumple:

- a) Si $t \geq 0 \implies U(t) \leq \beta/2 \leq V(t)$.
- b) Si $t \leq 0 \implies V(t) \leq \beta/2 \leq U(t)$.

Demostración. a) Es claro que si $t = 0$ se cumple la igualdad por la simetría.

Sea $t > 0$ y sea f_Z la función de densidad de Z . Tenemos dos posibilidades para tener en cuenta:

i) Primer caso, $-t < t - \lambda_t^\beta$:

Como $f_Z(s)$ es decreciente en $(0, +\infty)$, tenemos que

$$\min_{0 \leq s \leq t} f_X(s) \geq \max_{t \leq s \leq t + \lambda_t^\beta} f_Z(s).$$

Observemos que por la simetría $f_Z(-s) = f_Z(s)$ para todo $s \in [0, t]$, entonces

$$\min_{-t \leq s \leq t} f_X(s) = \min_{0 \leq s \leq t} f_Z(s).$$

Por otro lado, $[t - \lambda_t^\beta, t] \subset [-t, t]$, pues $-t < t - \lambda_t^\beta$, lo que implica

$$\min_{-t \leq s \leq t} f_Z(s) \leq \min_{t - \lambda_t^\beta \leq s \leq t} f_Z(s).$$

Luego,

$$\begin{aligned}
U(t) - V(t) &= \int_t^{t+\lambda_t^\beta} f_Z(s) ds - \int_{t-\lambda_t^\beta}^t f_Z(s) ds \leq \int_t^{t+\lambda_t^\beta} \max_{t \leq s \leq t+\lambda_t^\beta} f_Z(s) - \int_{t-\lambda_t^\beta}^t \min_{t-\lambda_t^\beta \leq s \leq t} f_Z(s) = \\
&= \lambda_t^\beta \max_{t \leq s \leq t+\lambda_t^\beta} f_Z(s) - \lambda_t^\beta \min_{t-\lambda_t^\beta \leq s \leq t} f_Z(s) = \\
&= \lambda_t^\beta \left(\max_{t \leq s \leq t+\lambda_t^\beta} f_Z(s) - \min_{t-\lambda_t^\beta \leq s \leq t} f_Z(s) \right) \leq 0
\end{aligned}$$

ii) Segundo caso, $t - \lambda_t^\beta < -t$:

Observemos que si $s \in [t, -t + \lambda_t^\beta]$, tenemos $-s \in [t - \lambda_t^\beta, -t]$ y por la simetría respecto de $t = 0$ de la función de densidad, sabemos que $f_Z(s) = f_Z(-s)$. Por lo tanto,

$$\int_t^{-t+\lambda_t^\beta} f_Z(s) ds = \int_{t-\lambda_t^\beta}^{-t} f_Z(s) ds.$$

f_Z es decreciente, conseguimos las siguientes desigualdades,

$$\max_{\lambda_t^\beta \leq s \leq t+\lambda_t^\beta} f_Z(s) \leq \max_{-t+\lambda_t^\beta \leq s \leq \lambda_t^\beta} f_Z(s) \leq \min_{0 \leq s \leq t} f_Z(s).$$

Luego,

$$\begin{aligned}
U(t) - V(t) &= \int_t^{t+\lambda_t^\beta} f_Z(s) ds - \int_{t-\lambda_t^\beta}^t f_Z(s) ds = \\
&= \int_t^{-t+\lambda_t^\beta} f_Z(s) ds + \int_{-t+\lambda_t^\beta}^{t+\lambda_t^\beta} f_Z(s) ds - \int_{t-\lambda_t^\beta}^{-t} f_Z(s) ds - \int_{-t}^t f_Z(s) ds = \\
&= \int_{-t+\lambda_t^\beta}^{t+\lambda_t^\beta} f_Z(s) ds - \int_{-t}^t f_Z(s) ds = \\
&= \int_{-t+\lambda_t^\beta}^{\lambda_t^\beta} f_Z(s) ds + \int_{\lambda_t^\beta}^{t+\lambda_t^\beta} f_Z(s) ds - 2 \int_0^t f_Z(s) ds \leq \\
&\leq t \max_{-t+\lambda_t^\beta \leq s \leq \lambda_t^\beta} f_Z(s) + t \max_{\lambda_t^\beta \leq s \leq t+\lambda_t^\beta} f_Z(s) - 2t \min_{0 \leq s \leq t} f_Z(s) \leq \\
&\leq 2t \left(\max_{-t+\lambda_t^\beta \leq s \leq \lambda_t^\beta} f_Z(s) - \min_{0 \leq s \leq t} f_Z(s) \right) \leq 0.
\end{aligned}$$

Finalmente, como $U(t) + V(t) = \beta$ y $U(t) \leq V(t) \implies U(t) \leq \beta/2 \leq V(t)$.

b) Consideremos la variable aleatoria $-Z$ que resulta ser absolutamente continua simétrica y unimodal en $t = 0$. Llamemos F_Z a la distribución acumulada de Z y F_{-Z} a la distribución acumulada de $-Z$. Observemos además que si $t \in \mathbb{R}$

$$F_{-Z}(t) = P(-Z \leq t) = P(Z \geq -t) = 1 - F_Z(-t).$$

$$\begin{aligned} U_{-Z}(t) &= F_{-Z}(t + \lambda_t^\beta) - F_{-Z}(t) = 1 - F_Z(-t - \lambda_t^\beta) - (1 - F_Z(-t)) = \\ &= F_Z(-t) - F_Z(-t - \lambda_t^\beta) = V_Z(-t). \end{aligned}$$

Análogamente,

$$\begin{aligned} V_{-Z}(t) &= F_{-Z}(t) - F_{-Z}(t - \lambda_t^\beta) = 1 - F_Z(-t) - (1 - F_Z(-(t - \lambda_t^\beta))) = \\ &= F_Z(-t + \lambda_t^\beta) - F_Z(-t) = U_Z(-t). \end{aligned}$$

Luego, si $t < 0$ tenemos que $-t > 0$ y por la parte (a) de la demostración,

$$U_{-Z}(-t) \leq \beta/2 \leq V_{-Z}(-t) \implies V_Z(t) \leq \beta/2 \leq U_Z(t).$$

Lema 2. Sea Z una variable aleatoria absolutamente continua tal que la función de distribución acumulada F_Z es C^1 . Sea $\beta \in (0, 1]$. Sea $t_0 \in \mathbb{R}$ tal que la densidad f_Z cumple $f(t_0 - \lambda_{t_0}^\beta) \in \text{Sop}(f_Z)$ o $f(t_0 + \lambda_{t_0}^\beta) \in \text{Sop}(f_Z)$. Entonces, existe un intervalo I centrado en t_0 y una función $\lambda^\beta : I \rightarrow \mathbb{R}_{\geq 0}$ tal que λ es C^1 en I , $\lambda^\beta(t_0) = \lambda_{t_0}^\beta$. Más aún, para cada $s \in I$,

$$\frac{\partial \lambda^\beta}{\partial t}(s) = -\frac{f_Z(t + \lambda^\beta(s)) - f_Z(t - \lambda^\beta(s))}{f_Z(t + \lambda^\beta(s)) + f_Z(t - \lambda^\beta(s))}$$

Demostración. La demostración se basa en aplicar el teorema de la función implícita a la función $g : \mathbb{R} \times \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}$ $g(x, \lambda) = F_Z(x + \lambda) - F_Z(x - \lambda) - \beta$.

Donde las hipótesis del teorema se cumplen:

- g es C^1 .
- $g(t_0, \lambda_{t_0}^\beta) = 0$.
- $\frac{\partial g}{\partial \lambda}(t_0, \lambda_{t_0}^\beta) = f_Z(t_0 + \lambda_{t_0}^\beta) + f_Z(t_0 - \lambda_{t_0}^\beta) > 0$.

Por último,

$$\frac{\partial g}{\partial t}(t, \lambda) = \frac{\partial}{\partial t} (F_Z(t + \lambda) - F_Z(t - \lambda) - \beta) = f_Z(t + \lambda) - f_Z(t - \lambda).$$

$$\frac{\partial g}{\partial \lambda}(t, \lambda) = \frac{\partial}{\partial \lambda} (F_Z(t + \lambda) - F_Z(t - \lambda) - \beta) = f_Z(t + \lambda) + f_Z(t - \lambda).$$

Lema 3. Sea Z una variable aleatoria absolutamente continua, simétrica y unimodal en $t = 0$ tal que la función de distribución acumulada F_Z es C^1 . Sea $\beta \in (0, 1]$. Sea f_Z la función de densidad tal que $f_Z(t + \lambda_t^\beta)f_Z(t - \lambda_t^\beta) > 0$ y cumple la siguiente desigualdad

$$f_Z(t) \geq 2 \frac{f_Z(t + \lambda_t^\beta)f_Z(t - \lambda_t^\beta)}{f_Z(t + \lambda_t^\beta) + f_Z(t - \lambda_t^\beta)} \quad \forall t \in \mathbb{R}. \quad (2.10)$$

Entonces,

a) $LD_S^\beta(t, F_Z)$ es no creciente si $t > 0$.

b) $LD_S^\beta(t, F_Z)$ es no decreciente si $t < 0$.

Demostración. Utilizando el resultado del Lema 2 y por simplicidad notando $\lambda_t^\beta = \lambda^\beta(t)$. Aplicando la Ecuación (2.4) escribimos,

$$LD_S^\beta(t, F_Z) = \frac{2}{\beta^2} [F_Z(t + \lambda(t)) - F_Z(t)] [\beta - (F_Z(t + \lambda(t)) - F_Z(t))] = \frac{2}{\beta^2} U(t)(\beta - U(t)).$$

Calculemos la derivada de LD_S^β respecto de t :

$$\frac{\partial LD_S^\beta}{\partial t}(t, F_Z) = \frac{2}{\beta^2} \left[\beta \frac{\partial U}{\partial t}(t) - 2U(t) \frac{\partial U}{\partial t}(t) \right] = \frac{2}{\beta^2} \frac{\partial U}{\partial t}(t) [\beta - 2U(t)].$$

Observemos la derivada de $U(t)$ respecto de t y apliquemos el Lema 2:

$$\begin{aligned} \frac{\partial U}{\partial t}(t) &= \frac{\partial}{\partial t} (F_Z(t + \lambda(t)) - F_Z(t)) = f_Z(t + \lambda(t)) \left(1 + \frac{\partial \lambda}{\partial t}(t) \right) - f_Z(t) = \\ &= f_Z(t + \lambda(t)) + f_Z(t + \lambda(t)) \frac{\partial \lambda}{\partial t}(t) - f_Z(t) = \\ &= f_Z(t + \lambda(t)) - f_Z(t) - f_Z(t + \lambda(t)) \frac{f_Z(t + \lambda(t)) - f_Z(t - \lambda(t))}{f_Z(t + \lambda(t)) + f_Z(t - \lambda(t))} = \\ &= \frac{f_Z(t + \lambda(t)) (f_Z(t + \lambda(t)) + f_Z(t - \lambda(t))) - (f_Z(t + \lambda(t)))^2 + f_Z(t + \lambda(t)) f_Z(t - \lambda(t))}{f_Z(t + \lambda(t)) + f_Z(t - \lambda(t))} - f_Z(t) = \\ &= 2 \frac{f_Z(t + \lambda(t)) f_Z(t - \lambda(t))}{f_Z(t + \lambda(t)) + f_Z(t - \lambda(t))} - f_Z(t) \leq 0. \end{aligned}$$

A partir del Lema 1 podemos deducir:

$$a) \text{ Si } t < 0, U(t) > \frac{\beta}{2} \Rightarrow \frac{\partial LD_S^\beta}{\partial t}(t, F_Z) = \frac{2}{\beta^2} \frac{\partial U}{\partial t}(t) [\beta - 2U(t)] \geq 0.$$

$$b) \text{ Si } t > 0, U(t) < \frac{\beta}{2} \Rightarrow \frac{\partial LD_S^\beta}{\partial t}(t, F_Z) = \frac{2}{\beta^2} \frac{\partial U}{\partial t}(t) [\beta - 2U(t)] \leq 0.$$

Lema 4. Sea Z una variable aleatoria y $\mu \in \mathbb{R}$. Sea $\beta \in (0, 1]$ y $Y = Z - \mu$. Llamemos F_Z y F_Y a las respectivas funciones de distribución acumulada, entonces, $LD_S^\beta(t, F_Z) = LD_S^\beta(t - \mu, F_Y)$.

Demostración. Sea $t \in \mathbb{R}$, tenemos que $F_Z(t) = F_Y(t - \mu)$. Por lo tanto,

$$U_Z(t) = F_Z(t + \lambda_t^\beta) - F_Z(t) = F_Y(t - \mu - \lambda_t^\beta) - F_Y(t - \mu) = U_Y(t - \mu).$$

y resulta la igualdad deseada.

Proposición 3. Sea Z una variable aleatoria absolutamente continua, simétrica y unimodal en θ . Supongamos que la función de distribución acumulada es C^1 y la función de densidad cumple la desigualdad (2.10). Entonces, dado $t \in \mathbb{R}$,

$$LD_S^\beta(t, F_Z) \leq LD_S^\beta((1-s)\theta + st, F_Z), \text{ para todo } s \in [0, 1].$$

Demostración. Sea $t \in \mathbb{R}$ e $Y = Z - \theta$. Supongamos que $t > \theta$ con lo cual $t - \theta > 0$. Por otro lado, $(1-s)\theta + st = \theta + s(t - \theta)$ y $s(t - \theta) < t - \theta$. Entonces, por los Lemas 3 y 4

$$LD_S^\beta(t, F_Z) = LD_S^\beta(t - \theta, F_Y) \leq LD_S^\beta(s(t - \theta), F_Y) = LD_S^\beta(s(t - \theta) + \theta, F_Z) = LD_S^\beta((1-s)\theta + st, F_Z).$$

P. 3. (monotonía respecto al punto más profundo). Sea \mathbb{E} un espacio de Banach separable con dual \mathbb{E}' . Sea X un elemento aleatorio continuo C -simétrico en θ con medida de probabilidad asociada P . Sea Q una medida de probabilidad en \mathbb{E}' independiente de P y supongamos que para todo $f \in \mathbb{E}'$ se cumple que $f(X)$ tiene función de densidad unimodal en $f(\theta)$ y cumple la desigualdad (2.10) Q -c.t.p. Entonces, para todo $x \in \mathbb{E}$ y $\beta \in (0, 1]$

$$IDLD^\beta(x, P) \leq IDLD^\beta((1-t)\theta + xt, P) \text{ para todo } t \in [0, 1].$$

Demostración. Sean $\beta \in (0, 1]$, $x \in \mathbb{E}$ y $t \in [0, 1]$. Entonces, aplicando el Lema (3),

$$\begin{aligned} IDLD^\beta(x, P) &= \int LD_S^\beta(f(x), P_f) dQ(f) \leq \int LD_S^\beta(f((1-t)\theta + tx), P_f) dQ(f) = \\ &= \int LD_S^\beta((1-t)f(\theta) + tf(x), P_f) dQ(f) = IDLD^\beta((1-t)\theta + tx, P). \end{aligned}$$

Ahora mostraremos que la distribución normal univariada cumple la desigualdad (2.10) condición impuesta por el Lema 3. Con lo cual estaríamos diciendo que las proyecciones de un proceso Gaussiano cumplen la desigualdad anteriormente mencionada y por consiguiente la propiedad **P.3**.

Veamos primero que vale la desigualdad (2.10) para la $N(0, 1)$ para todo $t \in \mathbb{R}$ y $\lambda > 0$. Llamemos φ su función de densidad.

$$\begin{aligned} \varphi(t) \geq 2 \frac{\varphi(t + \lambda)\varphi(t - \lambda)}{\varphi(t + \lambda) + \varphi(t - \lambda)} &\Leftrightarrow \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} \geq \frac{\frac{1}{\sqrt{2\pi}} e^{-\frac{(t+\lambda)^2}{2}} \frac{1}{\sqrt{2\pi}} e^{-\frac{(t-\lambda)^2}{2}}}{\frac{1}{\sqrt{2\pi}} e^{-\frac{(t+\lambda)^2}{2}} + \frac{1}{\sqrt{2\pi}} e^{-\frac{(t-\lambda)^2}{2}}} \Leftrightarrow \\ &\Leftrightarrow e^{-\frac{t^2}{2}} \geq 2 \frac{e^{-\frac{(t+\lambda)^2}{2}} e^{-\frac{(t-\lambda)^2}{2}}}{e^{-\frac{(t+\lambda)^2}{2}} + e^{-\frac{(t-\lambda)^2}{2}}} \Leftrightarrow 2e^{-\frac{t^2}{2}} \geq \frac{e^{-t^2 - \lambda^2}}{e^{-\frac{(t+\lambda)^2}{2}} + e^{-\frac{(t-\lambda)^2}{2}}} \Leftrightarrow \\ &\Leftrightarrow e^{-\frac{t^2}{2}} \geq 2 \frac{e^{-t^2} e^{-\lambda^2}}{e^{-\frac{t^2}{2}} e^{-\frac{\lambda^2}{2}} (e^{-\frac{t\lambda}{2}} + e^{\frac{t\lambda}{2}})} \Leftrightarrow 1 \geq 2 \frac{e^{-\frac{\lambda^2}{2}}}{e^{-\frac{t\lambda}{2}} + e^{\frac{t\lambda}{2}}} \Leftrightarrow \\ &\Leftrightarrow \frac{e^{-\frac{\lambda^2}{2}}}{e^{-\frac{t\lambda}{2}} + e^{\frac{t\lambda}{2}}} \leq \frac{1}{2}. \quad (2.11) \end{aligned}$$

Es claro que $e^{-\frac{\lambda^2}{2}} \leq 1$. Por lo tanto, la desigualdad (2.10) se cumple porque la función $h(t) = \frac{1}{e^{-\frac{t}{2}} + e^{\frac{t}{2}}}$ tiene un máximo global en $t = 0$ y para todo $\lambda > 0$, $h(0) = \frac{1}{2}$.

Veamos que si $Z \sim N(\mu, \sigma)$ su densidad, f_Z también cumple la desigualdad (2.10). Recordemos que podemos escribir $f_Z(t) = \frac{1}{\sigma} f\left(\frac{t-\mu}{\sigma}\right)$. Sean $t \in \mathbb{R}$ y $\lambda > 0$.

$$\begin{aligned} f_Z(t) &= \frac{1}{\sigma} \varphi\left(\frac{t-\mu}{\sigma}\right) \geq \frac{2}{\sigma} \frac{\varphi\left(\frac{t-\mu}{\sigma} + \frac{\lambda}{\sigma}\right) \varphi\left(\frac{t-\mu}{\sigma} - \frac{\lambda}{\sigma}\right)}{\varphi\left(\frac{t-\mu}{\sigma} + \frac{\lambda}{\sigma}\right) + \varphi\left(\frac{t-\mu}{\sigma} - \frac{\lambda}{\sigma}\right)} = 2 \frac{\sigma}{\sigma^2} \frac{\varphi\left(\frac{t+\lambda-\mu}{\sigma}\right) \varphi\left(\frac{t-\lambda-\mu}{\sigma}\right)}{\varphi\left(\frac{t+\lambda-\mu}{\sigma}\right) + \varphi\left(\frac{t-\lambda-\mu}{\sigma}\right)} = \\ &= 2 \frac{f_Z(t+\lambda) f_Z(t-\lambda)}{f_Z(t+\lambda) + f_Z(t-\lambda)}. \end{aligned}$$

La Propiedad 4 muestra, que al igual que la profundidad global, la profundidad local tiende a cero cuando nos acercamos a infinito.

P. 4. (desvanecerse en infinito). *Supongamos que*

$$\sup_{\|u\|=1} \{f : f(u) \leq \epsilon\} = O(\epsilon),$$

donde $O(\epsilon)$ es una función tal que $\lim_{\epsilon \rightarrow 0} O(\epsilon) = 0$ luego

$$\lim_{\|x\| \rightarrow +\infty} IDLD^\beta(x, P) = 0.$$

Demostración. Sea $\beta \in (0, 1]$. Por la Ecuación (2.4) tenemos que, $0 \leq F_f(f(x)) - F_f\left(f(x) - \lambda_{f(x)}^\beta\right) \leq \beta$. Como F_f es un función de distribución acumulada es claro que, para todo $f \in \mathbb{E}'$, $F_f\left(f(x) - \lambda_{f(x)}^\beta\right) \leq 1$.

Entonces,

$$\begin{aligned} IDLD^\beta(x, P) &= \int \frac{2}{\beta^2} \left[F_f(f(x)) - F_f\left(f(x) - \lambda_{f(x)}^\beta\right) \right] \left[F_f\left(f(x) + \lambda_{f(x)}^\beta\right) - F_f(f(x)) \right] dQ(f) \\ &\leq \int \frac{2}{\beta^2} \beta \left[F_f\left(f(x) + \lambda_{f(x)}^\beta\right) - F_f(f(x)) \right] dQ(f) \\ &\leq \frac{2}{\beta} \int \left[1 - F_f(f(x)) \right] dQ(f) = \frac{2}{\beta} \int P(f(X) > f(x)) dQ(f). \end{aligned}$$

Sea $\epsilon > 0$, $M > 0$ y $x \in \mathbb{E}$ tal que $\|x\| \leq M$,

$$\begin{aligned} \int P(f(X) > f(x)) dQ(f) &\leq \int \mathcal{I}_{\{f: f\left(\frac{x}{\|x\|}\right) \leq \epsilon\}} dQ(f) + \int P(f(X) > f(x)) \mathcal{I}_{\{f: f\left(\frac{x}{\|x\|}\right) \geq \epsilon\}} dQ(f) \\ &\leq \epsilon + \int P(f(X) \geq M\epsilon) dQ(f). \end{aligned}$$

Luego $\lim_{M \rightarrow +\infty} P(f(X) \geq M\epsilon) = 0$ para todo $f \in \mathbb{E}'$, por el Teorema de la Convergencia Mayorada obtenemos $\lim_{M \rightarrow +\infty} \int P(f(X) \geq M\epsilon) = 0$.

La Propiedad 5 muestra que $IDLD^\beta(x, P)$ es continua como función de x . Para obtener este resultado necesitaremos antes el siguiente Lema 5.

Lema 5. *Sea Z una variable aleatoria absolutamente continua con función de distribución acumulada F . Sea $(z_n)_{n \geq 1}$ una sucesión de números reales tales que $z_n \xrightarrow{n \rightarrow +\infty} z$ y $\beta \in (0, 1]$. Entonces, $LD_S^\beta(z_n, F) \xrightarrow{n \rightarrow +\infty} LD_S^\beta(z, F)$.*

Demostración. Recordemos antes de comenzar la demostración que si F es una función de distribución acumulada, $F^{-1}(s) = \inf\{t \in \mathbb{R} : F(t) \geq s\}$.

Dado que F es continua es suficiente con mostrar que $\lambda_{z_n}^\beta \xrightarrow{n \rightarrow +\infty} \lambda_z^\beta$.

Sea $t \in \mathbb{R}$, notemos $F_z(t) = \frac{1}{2}F(t) + \frac{1}{2}(1 - F(2z - t))$ a la versión simetrizada de F alrededor z .

$$\begin{aligned} \lambda_z^\beta &: F(z + \lambda_z^\beta) - F(z - \lambda_z^\beta) = \beta \\ \lambda_{z_n}^\beta &: F(z_n + \lambda_{z_n}^\beta) - F(z_n - \lambda_{z_n}^\beta) = \beta \text{ para todo } n \in \mathbb{N} \end{aligned}$$

Entonces,

$$\begin{aligned} F_z(z + \lambda_z^\beta) &= \frac{1}{2}F(z + \lambda_z^\beta) + \frac{1}{2}(1 - F(2z - (z + \lambda_z^\beta))) = \frac{1}{2}F(z + \lambda_z^\beta) + \frac{1}{2}(1 - F(z - \lambda_z^\beta)) \\ &= \frac{1}{2}[F(z + \lambda_z^\beta) - F(z - \lambda_z^\beta)] + \frac{1}{2} = \frac{\beta}{2} + \frac{1}{2}. \end{aligned}$$

Implicando que $z + \lambda_z^\beta = F_z^{-1}\left(\frac{\beta}{2} + \frac{1}{2}\right)$, análogamente $z_n + \lambda_{z_n}^\beta = F_z^{-1}\left(\frac{\beta}{2} + \frac{1}{2}\right)$. Dada que $F_{z_n}(t) \xrightarrow{n \rightarrow +\infty} F_z(t)$ es claro que

$$z_n + \lambda_{z_n}^\beta = F_{z_n}^{-1}\left(\frac{\beta}{2} + \frac{1}{2}\right) \xrightarrow{n \rightarrow +\infty} F_z^{-1}\left(\frac{\beta}{2} + \frac{1}{2}\right) = z + \lambda_z^\beta.$$

P. 5. (continua como función de x). *Sea $X \in \mathbb{E}$ un elemento aleatorio continuo con medida de probabilidad asociada P y $\beta \in (0, 1]$, entonces $IDLD^\beta(\cdot, P) : \mathbb{E} \rightarrow \mathbb{R}$ es continua.*

Demostración. *Sea $(x_n)_{n \geq 1}$ una sucesión en \mathbb{E} tal que $x_n \xrightarrow[n \rightarrow +\infty]{\|\cdot\|_{\mathbb{E}}} x$, específicamente $f(x_n) \xrightarrow{n \rightarrow +\infty} f(x)$ para todo $f \in \mathbb{E}'$. Para cada f fijo, el Lema 5 nos dice*

$$LD_S^\beta(f(x_n), P_f) \xrightarrow{n \rightarrow +\infty} LD_S^\beta(f(x), P_f).$$

El resultado final se obtiene como consecuencia del Teorema de la Convergencia Mayorada,

$$LD^\beta(x_n, P) \xrightarrow{n \rightarrow +\infty} LD_S^\beta(x, P).$$

Finalmente, probaremos que $IDLD^\beta(x, P)$ es continua como funcional de P . Es decir, es continua respecto a la topología que induce la convergencia en distribución; también llamada débil en contextos probabilísticos.

Bilingsley (3) da la siguiente definición: sea S un espacio métrico separable, y sea $(P_n)_{n \geq 1}$ una sucesión de medidas de probabilidad definidas en S ; decimos que P_n converge debilmente a una medida de probabilidad P si se satisface,

$$\int_S h dP_n \xrightarrow{n \rightarrow +\infty} \int_S h dP, \quad \forall h : S \rightarrow \mathbb{R} \text{ continua y acotada.}$$

Utilizando esta definición decimos que una sucesión de elementos aleatorios $(X_n)_{n \geq 1}$ en S convergen en distribución a un elemento X si las medidas P_n asociadas a X_n convergen debilmente a P , la medida asociada a X . Una notación que se menciona es $X_n \xrightarrow[n \rightarrow +\infty]{D} X$.

Por otro lado, hay una serie de condiciones equivalentes que se mencionan en el Teorema 2.1 de Bilingsley (3). Haremos hincapié en una de ellas que nos será de utilidad más adelante para probar la propiedad **P.6**.

$$X_n \xrightarrow[n \rightarrow +\infty]{D} X \text{ si } E(h(X_n)) \xrightarrow[n \rightarrow +\infty} E(h(X)) \quad \forall h : S \rightarrow \mathbb{R} \text{ uniformemente continua y acotada.} \quad (2.12)$$

P. 6. (continua como funcional de P). Sea \mathbb{E} un Banach separable y sea $(P_n)_{n \geq 1}$ una sucesión de medidas de probabilidad definidas en E tales que P_n converge debilmente a P . Entonces, para todo $\beta \in (0, 1]$, $IDLD^\beta(x, P_n) \xrightarrow[n \rightarrow +\infty} IDLD^\beta(x, P)$.

Demostración. Sea $x \in \mathbb{E}$ y $\beta \in (0, 1]$. Nuestro objetivo es probar que si P_n es una sucesión de medidas de probabilidad que convergen P , entonces

$$IDLD^\beta(x, P_n) \xrightarrow[n \rightarrow +\infty} IDLD^\beta(x, P).$$

Sea $(X_n)_{n \geq 1} \subset \mathbb{E}$ una sucesión de elementos aleatorios continuos en \mathbb{E} con medida de probabilidad asociada P_n , tal que, $X_n \xrightarrow[n \rightarrow +\infty} X$ en distribución. Veamos que

$$f(X_n) \xrightarrow[n \rightarrow +\infty} f(X) \text{ en distribución para todo } f \in \mathbb{E}'.$$

Basta mostrar que para toda $h : \mathbb{R} \rightarrow \mathbb{R}$ acotada y uniformemente continua

$$E(h(f(X_n))) \xrightarrow[n \rightarrow +\infty} E(h(f(X))).$$

Sin embargo, es inmediato a partir de 2.12 y que $f : \mathbb{E} \rightarrow \mathbb{R}$ es uniformemente continua por ser lineal y continua, entonces $h \circ f$ es uniformemente continua y acotada,

Dado que $F_{n,f}$ es la función de distribución acumulada de $f(X_n)$ la cual converge puntualmente a F_f , obtenemos que

$$LD_S^\beta(f(x), P_{n,f}) \xrightarrow[n \rightarrow +\infty} LD_S^\beta(f(x), P).$$

La demostración sigue de la aplicación del Teorema de la Convergencia Mayorada.

2.3. Versión Empírica y Resultados Asintóticos

En esta sección introduciremos la contraparte empírica de la IDLD y daremos algunos resultados asintóticos.

Antes de comenzar recordemos la definición de Paindavaine y Van Bever (45) de la Profundidad Simplicial univariada empírica, sea $ELD_S^{\beta(k)}(\cdot, F_n) : \mathbb{R} \rightarrow [0, 1/2]$, entonces

$$ELD_S^{\beta(k)}(z, F_n) = \frac{2}{\beta(k)^2} \left[F_n(z + \lambda_{z,n}^{\beta(k)}) - F_n(z) \right] \left[F_n(z) - F_n(z - \lambda_{z,n}^{\beta(k)}) \right],$$

donde

$$\lambda_{z,n}^{\beta(k)} = \inf_{\lambda > 0} \left\{ F_n(z + \lambda) - F_n(z - \lambda) = \beta(k) \right\}.$$

La siguiente observación implica la buena definición del nivel de localidad empírico, $\lambda_{z,n}^{\beta(k)}$.

Observación 7. Sea $\beta \in (0, 1]$ y Z_1, \dots, Z_n variables aleatoria de i.i.d. con distribución F . Dado $z \in \mathbb{R}$, para cada $1 \leq j \leq n$ notamos $d_j(z) = |Z_j - z|$ y $d^j(z)$ al j -ésimo estadístico de orden de $d_1(z), \dots, d_n(z)$. Sea $k = [n\beta]$, donde $[\cdot]$ representa la función parte entera, por lo tanto, $\#\{Z_j : Z_j \in [z - d^k(z), z + d^k(z)]\} = k$. Luego, $F_n(z + d^k(z)) - F_n(z - d^k(z)) = \frac{[n\beta]}{n} = \beta(k)$, y el nivel de localidad empírico es $\lambda_{z,n}^{\beta(k)} = d^k(z)$.

A partir de lo que hemos observado definimos la contraparte empírica de la IDLD como sigue.

Definición 17. Sea $\beta \in (0, 1]$, $X : \Omega \rightarrow \mathbb{E}$ un elemento aleatorio continuo y X_1, \dots, X_n una muestra aleatoria con la misma distribución que X . Sea $k = [n\beta]$, para cada $x \in \mathbb{E}$ y $f \in \mathbb{E}'$ escribimos

$$\lambda_{f(x),n}^{\beta(k)} = \inf \left\{ \lambda > 0 : F_{f,n}(f(x) + \lambda) - F_{f,n}(f(x) - \lambda) = \frac{k}{n} \right\}. \quad (2.13)$$

Sea $\beta(k) = \frac{k}{n}$, la versión empírica de la IDLD de nivel $\beta(k)$ se define como,

$$EIDLD^{\beta(k)}(x, P_n) = IDLD^{\beta(k)}(x, P_n). \quad (2.14)$$

Primero, estableceremos la convergencia fuerte de la profundidad local simplicial. Asumiendo que se mantienen las condiciones de la observación (7) notemos algunos hechos relevantes que se cumplen. Para mantener sencilla la notación llamemos, $\lambda = \lambda_z^{\beta}$,

$p_+ = F(z + \lambda)$, $p_- = F(z - \lambda)$ y $p = F(z)$.

Sea $p \in (0, 1)$ entonces,

$$(i) \quad Q_{p,n} = Z_{([np]+1)}, \quad Q_{p+,n} = Z_{([np_+]+1)} \text{ y } Q_{p-,n} = Z_{([np_-]+1)}.$$

$$(ii) F_n(Q_{p_+,n}) - F_n(Q_{p_-,n}) = \frac{[np_+] + 1}{n} - \frac{[np_-] + 1}{n} = \frac{[np_+] - [np_-]}{n}.$$

$$\text{Más aún, } \frac{[np_+ - np_-]}{n} \leq \frac{[np_+] - [np_-]}{n} \leq \frac{[np_+] - [np_-] + 1}{n}.$$

$[np_+ - np_-] = [n(p_+ - p_-)] = [n\beta]$. Entonces,

$$\frac{[n\beta]}{n} \leq F_n(Q_{p_+,n}) - F_n(Q_{p_-,n}) \leq \frac{[n\beta] + 1}{n}.$$

$$(iii) Z_{([np_-]+1)} \leq z \leq Z_{([np_+]+1)}.$$

$$(iv) [z - d^{(k)}(z), z + d^{(k)}(z)] \subset [Z_{([np_-]+1)}, Z_{([np_+]+1)}].$$

$$(v) d_{(k)}(z) = \min\{z - Q_{p_-,n}, z + Q_{p_+,n}\}.$$

$$(vi) F_n(Q_{p_+,n}) - F_n(z + d_{(k)}(z)) \leq \frac{1}{n} \text{ y } F_n(Q_{p_-,n}) - F_n(z - d_{(k)}(z)) \leq \frac{1}{n}.$$

$$(vii) \beta(k) \leq \beta \leq \beta(k) + 1.$$

Lema 6. Sea Z una variable aleatoria absolutamente continua con distribución F y sean Z_1, \dots, Z_n una sucesión de variables aleatorias i.i.d con distribución F . Sea $z_p = F^{-1}(p)$ el cuantil $p \in (0, 1)$ de F y $Q_{p,n}$ el cuantil p de F_n (función de distribución empírica de Z_1, \dots, Z_n .) Entonces,

$$(i) Q_{p,n} = Z_{([np]+1)}.$$

$$(ii) |F_n(Q_{p,n}) - F(z_p)| \leq \frac{1}{n} \forall p \in (0, 1).$$

$$(iii) |F(Q_{p,n}) - F(z_p)| \leq \|F_n - F\|_\infty + \frac{1}{n}.$$

Demostración. (i) Es directo por la definición.

(ii) Sea $p \in (0, 1)$,

$$|F_n(Q_{p,n}) - F(z_p)| = |F_n(Q_{p,n}) - p| = \frac{[np] + 1}{n} - p = \frac{[np] - np + 1}{n} \leq \frac{1}{n}.$$

(iii) Sea $p \in (0, 1)$,

$$\begin{aligned} |F(Q_{p,n}) - F(z_p)| &\leq |F(Q_{p,n}) - F_n(Q_{p,n})| + |F_n(Q_{p,n}) - F(z_p)| \\ &\leq \sup_{t \in \mathbb{R}} |F_n(t) - F(t)| + \frac{1}{n} = \|F_n - F\|_\infty + \frac{1}{n}. \end{aligned}$$

Proposición 4. Sea Z_1, \dots, Z_n una sucesión de variables aleatorias i.i.d. con función de distribución acumulada F . Sea $\beta \in (0, 1]$ y $z \in \mathbb{R}$. Entonces,

$$\left| ELD_S^{\beta(k)}(z, F_n) - LD_S^\beta(z, F) \right| \leq \frac{1}{2} \left(1 - \left(\frac{\beta(k)}{\beta} \right)^2 \right) + \frac{2}{\beta^2} \left(\frac{8}{n} + 4\|F_n - F\|_\infty \right) \quad (2.15)$$

Demostración. Para mantener la escritura simple notaremos $\lambda = \lambda_z^\beta$ y $d^k = d^{(k)}(z)$.

$$\begin{aligned} & \left| (F(z + \lambda) - F(z))(F(z) - F(z - \lambda)) - (F_n(z + d^k) - F_n(z))(F_n(z) - F_n(z - d^k)) \right| = \\ & = \left| \left[F(z + \lambda)F(z) - F(z + \lambda)F(z - \lambda) - F(z)^2 + F(z)F(z - \lambda) \right] - \right. \\ & \left. - \left[F_n(z + d^k)F_n(z) - F_n(z + d^k)F_n(z - d^k) - F_n(z)^2 + F_n(z)F_n(z - d^k) \right] \right| = \\ & = \left| F(z + \lambda)F(z) - F(z + \lambda)F(z - \lambda) - F(z)^2 + F(z)F(z - \lambda) - \right. \\ & \left. - F_n(z + d^k)F_n(z) + F_n(z + d^k)F_n(z - d^k) + F_n(z)^2 - F_n(z)F_n(z - d^k) \right| = \\ & \leq \left| F(z + \lambda)F(z) - F_n(z + d^k)F_n(z) \right| + \left| F_n(z + d^k)F_n(z - d^k) - F(z + \lambda)F(z - \lambda) \right| + \\ & + \left| F_n(z)^2 - F(z)^2 \right| + \left| F(z - \lambda)F(z) - F_n(z - d^k)F_n(z) \right| \end{aligned} \quad (2.16)$$

Analicemos cada término de la Ecuación (2.16) por separado,

(a)

$$\begin{aligned} & \left| F(z + \lambda)F(z) - F_n(z + d^k)F_n(z) \right| = \\ & = \left| F(z + \lambda)F(z) - F(z)F_n(Q_{p+,n}) + F(z)F_n(Q_{p+,n}) - F_n(z + d^k)F_n(z) \right| \leq \\ & \leq F(z) \left| F(z + \lambda) - F_n(Q_{p+,n}) \right| + \\ & + \left| F(z)F_n(Q_{p+,n}) - F(z)F_n(z + d^k) + F(z)F_n(z + d^k) - F_n(z + d^k)F_n(z) \right| \leq \\ & \leq \frac{1}{n} + F(z) \left| F_n(Q_{p+,n}) - F_n(z + d^k) \right| + \left| F(z) - F_n(z) \right| F_n(z + d^k) \leq \\ & \leq \frac{1}{n} + \frac{1}{n} + \|F - F_n\|_\infty = \frac{2}{n} + \|F - F_n\|_\infty. \end{aligned}$$

(b)

$$\begin{aligned}
& \left| F_n(z + d^k)F_n(z - d^k) - F(z + \lambda)F(z - \lambda) \right| = \\
& = \left| F_n(z + d^k)F_n(z - d^k) - F(z + \lambda)F_n(z - d^k) + \right. \\
& \quad \left. + F(z + \lambda)F_n(z - d^k) - F(z + \lambda)F(z - \lambda) \right| \leq \\
& \leq F_n(z - d^k) \left| F_n(z + d^k) - F(z + \lambda) \right| + F(z + \lambda) \left| F_n(z - d^k) - F(z - \lambda) \right| \leq \\
& \leq \left| F_n(z + d^k) - F(z + \lambda) \right| + \left| F_n(z - d^k) - F(z - \lambda) \right| \leq \\
& \leq \left| F_n(z + d^k) - F_n(Q_{p+,n}) \right| + \left| F_n(Q_{p+,n}) - F(z + \lambda) \right| + \\
& \quad + \left| F_n(Q_{p-,n}) - F_n(z - d^k) \right| + \left| F_n(Q_{p-,n}) - F(z - \lambda) \right| \leq \frac{1}{n} + \frac{1}{n} + \frac{1}{n} + \frac{1}{n} = \frac{4}{n}.
\end{aligned}$$

(c)

$$\begin{aligned}
& \left| F_n(z)^2 - F(z)^2 \right| = \left| F_n(z) - F(z) \right| \left| F_n(z) + F(z) \right| \leq \\
& \leq 2 \left| F_n(z) - F(z) \right| \leq 2 \|F_n - F\|_\infty.
\end{aligned}$$

(d) Análogo al ítem (a).

Finalmente, llamemos

- $H = (F(z + \lambda)F(z)) (F(z)F(z - \lambda))$ y
- $G = (F_n(z + d^k) - F_n(z)) (F_n(z) - F_n(z - \lambda))$.

Luego,

$$\begin{aligned}
& \left| ELD_S^\beta(z, F_n) - LD_S^\beta(z, F) \right| = \left| \frac{2}{\beta(k)^2} G - \frac{2}{\beta^2} H \right| \leq \left| \frac{2}{\beta(k)^2} G - \frac{2}{\beta^2} G \right| + \left| \frac{2}{\beta^2} G - \frac{2}{\beta^2} H \right| \leq \\
& \leq \left(\frac{2}{\beta(k)^2} - \frac{2}{\beta^2} \right) |G| + \frac{2}{\beta^2} |G - H|.
\end{aligned}$$

Dado que se cumplen las hipótesis de la Proposición 2 tenemos que el primer término es menor o igual $\frac{\beta(k)^2}{2}$. Entonces,

$$\left(\frac{2}{\beta(k)^2} - \frac{2}{\beta^2} \right) |G| \leq \left(\frac{2}{\beta(k)^2} - \frac{2}{\beta^2} \right) \frac{\beta(k)^2}{4} = \frac{1}{2} \left(1 - \left(\frac{\beta(k)}{\beta} \right)^2 \right). \quad (2.17)$$

Por otro lado, ya sabíamos que

$$|G - H| \leq \frac{8}{n} + 4 \|F - F_n\|_\infty. \quad (2.18)$$

De las desigualdades (2.17) y (2.18) probamos (2.15).

Los teoremas que siguen a continuación establecen, en primer lugar, la convergencia uniforme para la profundidad local simplicial unidimensional empírica respecto de su contraparte poblacional. Luego basándonos en ese resultado conseguiremos la consistencia uniforme para la profundidad local dual integrada empírica.

Teorema 2. *Sea \mathbb{E} un espacio de Banach Separable con dual \mathbb{E}' . Dados X_1, \dots, X_n elementos aleatorios independientes e idénticamente distribuidos en \mathbb{E} con medida de probabilidad asociada P y sea $\beta \in (0, 1]$. Entonces, tenemos que*

(a)

$$E \left(\sup_{x \in \mathbb{E}} \left| ELD_S^{\beta(k)}(f(x), P_{n,f}) - LD_S^\beta(f(x), P_f) \right| \right) \xrightarrow{n \rightarrow +\infty} 0 \text{ para todo } f \in \mathbb{E}'. \quad (2.19)$$

(b)

$$E \left(\sup_{x \in \mathbb{E}} \left| EIDL D^{\beta(k)}(x, P_n) - IDLD^\beta(x, P) \right| \right) \xrightarrow{n \rightarrow +\infty} 0. \quad (2.20)$$

Demostración. (a) *Sea $f \in \mathbb{E}'$ y $x \in \mathbb{E}$. Notamos P_f a la medida de probabilidad asociada a $f(X)$ y F_f a la función de distribución acumulada; donde X es un elemento aleatorio en \mathbb{E} con medida de probabilidad P . Análogamente notamos $P_{n,f}$ a la medida de probabilidad empírica de P_f basada en $f(X_1), \dots, f(X_n)$ y $F_{f,n}$ a la función de distribución acumulada empírica.*

Por la Proposición 4 tenemos,

$$\left| ELD_S^{\beta(k)}(f(x), P_{f,n}) - LD_S^\beta(f(x), P_f) \right| \leq \frac{1}{2} \left(1 - \left(\frac{\beta(k)}{\beta} \right)^2 \right) + \frac{2}{\beta^2} \left(\frac{8}{n} + 4 \|F_{f,n} - F_f\|_\infty \right).$$

Observamos que,

$$\frac{1}{2} \left(1 - \left(\frac{\beta(k)}{\beta} \right)^2 \right) = \frac{1}{2} \frac{\beta^2 - \beta(k)^2}{\beta^2} = \frac{1}{2} (\beta - \beta(k)) \frac{(\beta + \beta(k))}{\beta^2} \leq \frac{2}{2n\beta^2} = \frac{1}{n\beta^2}.$$

Entonces,

$$\left| ELD_S^{\beta(k)}(f(x), P_{f,n}) - LD_S^\beta(f(x), P_f) \right| \leq \frac{1}{\beta^2} \left(\frac{17}{n} + 8 \|F_{f,n} - F_f\|_\infty \right). \quad (2.21)$$

Como la desigualdad (2.21) no depende de x , ésta sigue valiendo si tomamos supremo.

$$\sup_{x \in \mathbb{E}} \left| ELD_S^{\beta(k)}(f(x), P_{f,n}) - LD_S^\beta(f(x), P_f) \right| \leq \frac{1}{\beta^2} \left(\frac{17}{n} + 8 \|F_{f,n} - F_f\|_\infty \right).$$

Por lo tanto,

$$\begin{aligned}
E \left[\sup_{x \in \mathbb{E}} \left| ELD_S^{\beta(k)}(f(x), P_{f,n}) - LD_S^\beta(f(x), P_f) \right| \right] &\leq E \left[\frac{1}{\beta^2} \left(\frac{17}{n} + 8 \|F_{f,n} - F_f\|_\infty \right) \right] = \\
&= \frac{1}{\beta^2} \left(\frac{17}{n} + 8 E \left[\|F_{f,n} - F_f\|_\infty \right] \right). \tag{2.22}
\end{aligned}$$

Se desprende de (2.22) que basta mostrar que $E \left[\|F_{f,n} - F_f\|_\infty \right] \xrightarrow{n \rightarrow +\infty} 0$.

Sea $\epsilon > 0$, existe un $n_0 \in \mathbb{N}$ tal que $\forall n \geq n_0$, $4e^{-n\frac{\epsilon^2}{2}} < \frac{\epsilon}{2}$;

$$\begin{aligned}
E \left[\|F_{f,n} - F_f\|_\infty \right] &= \\
&= E \left[\|F_{f,n} - F_f\|_\infty \mathcal{I} \left\{ \|F_{f,n} - F_f\|_\infty \leq \frac{\epsilon}{2} \right\} \right] + E \left[\|F_{f,n} - F_f\|_\infty \mathcal{I} \left\{ \|F_{f,n} - F_f\|_\infty > \frac{\epsilon}{2} \right\} \right] \leq \\
&\leq \epsilon \mathbb{P} \left(\|F_{f,n} - F_f\|_\infty \leq \frac{\epsilon}{2} \right) + 2 \mathbb{P} \left(\|F_{f,n} - F_f\|_\infty > \frac{\epsilon}{2} \right) \leq \\
&\leq \frac{\epsilon}{2} + 2 \mathbb{P} \left(\|F_{f,n} - F_f\|_\infty > \frac{\epsilon}{2} \right).
\end{aligned}$$

Por la desigualdad de Dvoretzky-Kiefer-Wolfowitz (39) obtenemos que

$$E \left[\|F_{f,n} - F_f\|_\infty \right] \leq \frac{\epsilon}{2} + 4e^{-n\frac{\epsilon^2}{2}} < \epsilon \tag{2.23}$$

Una observación importante es que la convergencia no depende del funcional f que se considere; lo cual no será de utilidad para probar la parte (b).

(b) Es directo de la parte (a) del teorema y del hecho de que es la integral de una función medible, positiva y acotada. Es decir, dado $\epsilon > 0$, existe un $n_1 \in \mathbb{N}$ tal que $\forall n \geq n_1$,

$$E \left[\sup_{x \in \mathbb{E}} \left| ELD_S^{\beta(k)}(f(x), P_{f,n}) - LD_S^\beta(f(x), P_f) \right| \right] < \epsilon. \tag{2.24}$$

Luego,

$$\begin{aligned}
&E \left[\sup_{x \in \mathbb{E}} \left| EIDL D^{\beta(k)}(x, P_n) - IDLD^\beta(x, P) \right| \right] \leq \\
&\leq E \left[\sup_{x \in \mathbb{E}} \int \left| ELD_S^{\beta(k)}(f(x), P_{f,n}) - LD_S^\beta(f(x), P_f) \right| dQ(f) \right] = \\
&= E \left[\int \sup_{x \in \mathbb{E}} \left| ELD_S^{\beta(k)}(f(x), P_{f,n}) - LD_S^\beta(f(x), P_f) \right| dQ(f) \right] = \\
&= \int E \left[\sup_{x \in \mathbb{E}} \left| ELD_S^{\beta(k)}(f(x), P_{f,n}) - LD_S^\beta(f(x), P_f) \right| \right] dQ(f) \leq \\
&= \int \epsilon dQ(f) = \epsilon \text{ si } n > n_1.
\end{aligned}$$

Teorema 3. *Sea X un elemento aleatorio en \mathbb{E} un espacio de Banach separable con medida de probabilidad asociada P tal que $E(f(X)^2) < +\infty$ para todo $f \in \mathbb{E}'$. Sea X_1, \dots, X_n una muestra aleatoria con la misma distribución que X y sea $\beta \in (0, 1]$. Entonces,*

$$\mathbb{P}\left(\sup_{x \in \mathbb{E}} |EIDLD^{\beta(k)}(x, P_n) - IDLD^{\beta}(x, P)| \xrightarrow{n \rightarrow +\infty} 0\right) = 1.$$

Demostración. *Observemos que,*

$$\begin{aligned} & \mathbb{P}\left(\sup_{x \in \mathbb{E}} |EIDLD^{\beta}(x, P_n) - IDLD^{\beta}(x, P)| \xrightarrow{n \rightarrow +\infty} 0\right) = \\ & = \mathbb{P}\left(\bigcap_{\epsilon > 0} \bigcup_{n \in \mathbb{N}} \bigcap_{l \geq n} \left\{ \sup_{x \in \mathbb{E}} |EIDLD^{\beta(k)}(x, P_l) - IDLD^{\beta}(x, P)| < \epsilon \right\}\right) = \\ & = 1 - \mathbb{P}\left(\bigcup_{\epsilon > 0} \bigcap_{n \in \mathbb{N}} \bigcup_{l \geq n} \left\{ \sup_{x \in \mathbb{E}} |EIDLD^{\beta(k)}(x, P_l) - IDLD^{\beta}(x, P)| > \epsilon \right\}\right). \end{aligned}$$

Luego, es suficiente mostrar que

$$\mathbb{P}\left(\bigcup_{\epsilon > 0} \bigcap_{n \in \mathbb{N}} \bigcup_{l \geq n} \left\{ \sup_{x \in \mathbb{E}} |EIDLD^{\beta(k)}(x, P_l) - IDLD^{\beta}(x, P)| > \epsilon \right\}\right) = 0.$$

Por el lema de Borell-Cantelli es basta con probar que si la probabilidad de los conjuntos

$$A_n = \left\{ \sup_{x \in \mathbb{E}} |EIDLD^{\beta(k)}(x, P_n) - IDLD^{\beta}(x, P)| > \epsilon \right\},$$

es sumable, entonces, para todo $\epsilon > 0$,

$\mathbb{P}\left(\bigcap_{n \in \mathbb{N}} \bigcup_{l \geq n} \left\{ \sup_{x \in \mathbb{E}} |EIDLD^{\beta(k)}(x, P_l) - IDLD^{\beta}(x, P)| > \epsilon \right\}\right) = 0$ *y la demostración estará terminada.*

Sea $\epsilon > 0$,

$$\begin{aligned} & \sup_{x \in \mathbb{E}} |EIDLD^{\beta(k)}(x, P_n) - IDLD^{\beta}(x, P)| = \\ & = \sup_{x \in \mathbb{E}} \left| \int ELD_S^{\beta(k)}(f(x), P_{f,n}) - LD_S^{\beta}(f(x), P_f) dQ \right| \leq \\ & \leq \sup_{x \in \mathbb{E}} \int |ELD_S^{\beta(k)}(f(x), P_{f,n}) - LD_S^{\beta}(f(x), P_f)| dQ = \\ & = \int \sup_{x \in \mathbb{E}} |ELD_S^{\beta(k)}(f(x), P_{f,n}) - LD_S^{\beta}(f(x), P_f)| dQ \leq \\ & \leq \int \frac{1}{\beta^2} \left(\frac{8}{n} + 4 \|F_{f,n} - F_f\|_{\infty} \right) dQ = \frac{1}{\beta^2} \frac{8}{n} + \frac{1}{\beta^2} 4 \int \|F_{f,n} - F_f\|_{\infty} dQ \leq \\ & \leq \frac{8}{n\beta^2} + \frac{1}{2\beta^2} \sup_{f \in \mathbb{E}'} \|F_{f,n} - F_f\|_{\infty}. \end{aligned}$$

Dado que $\sup_{f \in \mathbb{B}'} \|F_{f,n} - F_f\|_\infty < +\infty$ existe $f_0 \in \mathbb{B}'$ tal que

$$\sup_{f \in \mathbb{B}'} \|F_{f,n} - F_f\|_\infty \leq \|F_{f_0,n} - F_{f_0}\|_\infty + \beta^2 \epsilon,$$

entonces

$$\frac{8}{n\beta^2} + \frac{1}{2\beta^2} \sup_{f \in \mathbb{B}'} \|F_{f,n} - F_f\|_\infty \leq \frac{8}{n\beta^2} + \frac{1}{2\beta^2} \|F_{f_0,n} - F_{f_0}\|_\infty + \frac{\beta^2 \epsilon}{2\beta^2}.$$

Por la desigualdad de Dvoretzky-Kiefer-Wolfowitz,

$$\begin{aligned} \mathbb{P}(A_n) &\leq \mathbb{P}\left(\frac{8}{n\beta^2} + \frac{1}{2\beta^2} \|F_{f_0,n} - F_{f_0}\|_\infty + \frac{\epsilon}{2} > \epsilon\right) = \mathbb{P}\left(\|F_{n,f_0} - F_{f_0}\|_\infty > \epsilon\beta^2 - \frac{16}{n}\right) \leq \\ &\leq 2 \exp\left\{-2n\left(\epsilon\beta^2 - \frac{16}{n}\right)^2\right\}. \end{aligned}$$

Con lo cual la suma nos queda acotada que es lo que queríamos ver, es decir,

$$\sum_{n \in \mathbb{N}} \mathbb{P}(A_n) \leq 2 \sum_{n \in \mathbb{N}} \exp\left\{-2n\left(\epsilon\beta^2 - \frac{16}{n}\right)^2\right\} < +\infty. \quad (2.25)$$

2.4. Regiones de Profundidad Local

En esta sección definiremos la *región de profundidad local* α con nivel de localidad β , la cual será de utilidad para las aplicaciones. Idealmente, estas regiones deberían ser invariantes respecto del sistema de coordenadas y acotadas. También estudiaremos bajo qué condiciones tendremos propiedades asintóticas.

Notemos LD^β a una función de profundidad local y ELD^β a su versión empírica, en particular podemos considerar La Profundidad Local Dual Integrada definida en la Sección 2.2.

Definición 18. Sea \mathbb{E} un espacio de Banach separable, sea $X : \Omega \rightarrow \mathbb{E}$ un elemento aleatorio con medida de probabilidad asociada P . Sea $\beta \in (0, 1]$ y $\alpha \in (0, \frac{1}{2}]$, llamemos *región de profundidad local* α con nivel de localidad β , al conjunto

$$R_\beta^\alpha = \{x \in \mathbb{E} : LD^\beta(x, P) \geq \alpha\}. \quad (2.26)$$

Sean X_1, \dots, X_n una muestra aleatoria en \mathbb{E} , la versión empírica de R_β^α es

$$R_n^\alpha = R_{n,\beta}^\alpha = \{x \in \mathbb{E} : ELD^\beta(x, P_n) \geq \alpha\}.$$

En toda esta sección el nivel de localidad β quedará fijo, por lo tanto notaremos R^α (resp. R_n^α) y R_β^α (resp. $R_{n,\beta}^\alpha$) sin ambigüedad.

Para ilustrar un poco más este concepto y poder entender mejor la diferencia con las regiones globales veremos dos ejemplos gráficos. En la Figura 2.2 hemos calculado la profundidad local para una grilla. Se puede apreciar que hay tres zonas con forma de elipse teñidas por un azul más oscuro que el resto del plano. A medida que nos alejamos del centro de las elipses se genera un gradiente de color hacia tonos más claros. Las dos elipses que se encuentran más cerca del borde rodean a la moda y la que queda en el medio a la antimoda. Una buena analogía puede hacerse comparando este concepto a las fosas que se encuentran en los océanos donde el color del agua se vuelve más oscuro a medida que aumenta la profundidad siendo, en este caso un océano definido por la medida de probabilidad. El parámetro β determina, siguiendo la analogía, la forma del océano y el parámetro α un límite para poder fijar estas regiones, es decir un corte para poder determinar si estamos dentro o fuera de fosa. Esto se puede ver en la Figura 2.3 donde los puntos de color azul quedan dentro de la región empírica y los puntos negros quedan fuera.

Observación 8. ■ Si \mathbb{E} es finito dimensional, entonces R^α es invariante por transformaciones ortogonales.

- Si $\alpha_1 \leq \alpha_2$ entonces $R_\beta^{\alpha_2} \subset R_\beta^{\alpha_1}$.

El teorema que sigue nos mostrará que la región de profundidad local α con nivel de localidad β empírica es fuertemente consistente respecto a su contraparte poblacional bajo condiciones de regularidad.

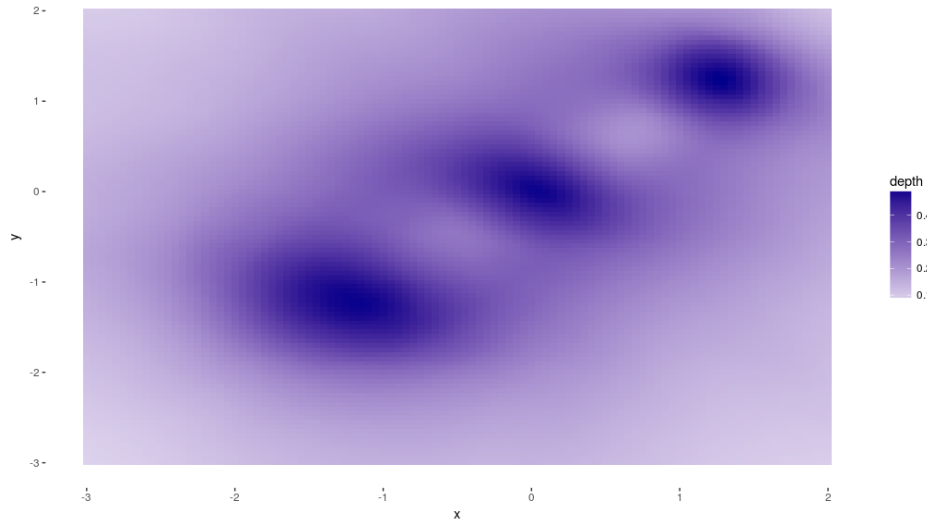


Figura 2.2: Mapa de profundidades según una escala de azules.

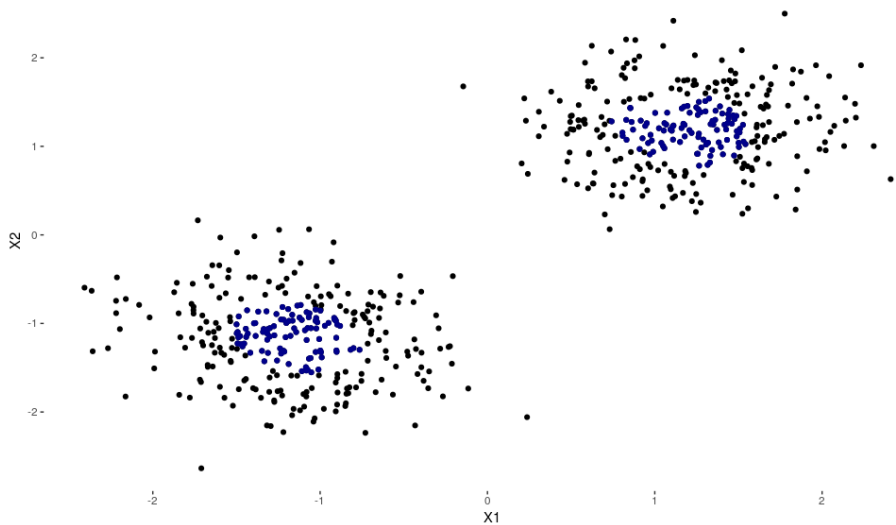


Figura 2.3: Regiones empíricas de profundidad local para $\alpha = 0,4$ y $\beta(k) = 0,45$.

Teorema 4. Sea \mathbb{E} un espacio de Banach separable, sea $X : \Omega \rightarrow \mathbb{E}$ una elemento aleatorio con medida de probabilidad asociada P . Supongamos que,

$$a) \quad LD^\beta(x, P) \xrightarrow{\|x\| \rightarrow +\infty} 0.$$

$$b) \quad \sup_{x \in \mathbb{E}} |ELD^\beta(x, P_n) - LD^\beta(x, P)| \xrightarrow{n \rightarrow +\infty} 0 \text{ c.s.}$$

Entonces, para todo $\epsilon > 0$, $0 < \delta < \epsilon$, $0 < \alpha$ y una sucesión de números reales $(\alpha_n)_{n \geq 1}$ tal que $\alpha_n \rightarrow \alpha$,

(I) Existe $n_0 \in \mathbb{N}$ que cumple $R^{\alpha+\epsilon} \subset R_n^{\alpha_n+\delta} \subset R_n^{\alpha_n} \subset R_n^{\alpha_n-\delta} \subset R^{\alpha-\epsilon}$.

(II) Si $P(x \in \mathbb{E} : LD^\beta(x) = \alpha) = 0$, entonces $R_n^{\alpha_n} \xrightarrow{n \rightarrow +\infty} R^\alpha$ c.s.

Demostración. (I) Sea $\epsilon > 0$, $0 < \delta < \epsilon$, $\alpha > 0$ y una sucesión $\alpha_n \rightarrow \alpha$. Es claro por la Observación 8 que

$$R_n^{\alpha_n-\delta} \subset R_n^{\alpha_n} \subset R_n^{\alpha_n+\delta}.$$

Queremos ver que $R_n^{\alpha_n-\delta} \subset R^{\alpha-\epsilon}$. Sin pérdida de generalidad asumimos que $\alpha - \epsilon > 0$, en caso contrario la inclusión se cumple siempre.

Dado que $\alpha_n \rightarrow \alpha$, existe $n_1 \in \mathbb{N}$ tal que para $n \geq n_1$, $|\alpha_n - \alpha| < \frac{\epsilon - \delta}{2}$. Más aún, de la hipótesis (b) obtenemos que existe $n_2 > n_1$ tal que si $n \geq n_2$ entonces,

$$\sup_{x \in \mathbb{E}} |ELD^\beta(x, P_n) - LD^\beta(x, P)| \leq \frac{\epsilon - \delta}{2} \quad \text{c.s.}$$

Sea $x \in R_n^{\alpha_n-\delta} \cap (R^{\alpha-\epsilon})^c$, si $n \geq n_2$, tenemos

$$ELD^\beta(x, P_n) - LD^\beta(x, P) > \alpha_n - \delta - (\alpha - \epsilon) = \alpha_n - \alpha - \delta + \epsilon \geq -\frac{(\epsilon - \delta)}{2} + \epsilon - \delta = \frac{\epsilon - \delta}{2}.$$

Lo cual es una contradicción, lo que nos lleva a concluir que la intersección es vacía. Luego,

$$\text{si } x \in R_n^{\alpha_n-\delta} \Rightarrow x \in R^{\alpha-\epsilon} \Rightarrow R_n^{\alpha_n-\delta} \subset R^{\alpha-\epsilon}.$$

Nos faltaría para terminar esta primera parte probar la otra inclusión, $R^{\alpha+\epsilon} \subset R_n^{\alpha_n+\delta}$. La prueba es análoga, sea $x \in R^{\alpha+\epsilon} \cap R_n^{\alpha_n+\delta}$, si $n \geq N_2$

$$LD^\beta(x, P) - ELD^\beta(x, P_n) > \epsilon + \alpha - \alpha_n - \alpha = \epsilon - \delta - (\alpha_n - \alpha) \geq \epsilon - \delta - \frac{(\epsilon - \delta)}{2} = \frac{\epsilon - \delta}{2}.$$

Otra vez nos encontramos con una contradicción. Por lo tanto,

$$R^{\alpha+\epsilon} \cap (R_n^{\alpha_n+\delta})^c = \emptyset \Rightarrow R^{\alpha+\epsilon} \subset R_n^{\alpha_n+\delta}.$$

(II) Sabemos que

$$\{x \in \mathbb{E} : LD^\beta(x, P) > \alpha\} = \bigcup_{\epsilon \in \mathbb{Q}^+} R^{\alpha+\epsilon} \subset \bigcap_{\epsilon \in \mathbb{Q}^+} R^{\alpha-\epsilon} = \{x \in \mathbb{E} : LD^\beta(x, P) \geq \alpha\} \quad (2.27)$$

Queremos probar que $\bigcup_{\epsilon \in \mathbb{Q}^+} R^{\alpha+\epsilon} \subset \liminf R_n^{\alpha_n} = \bigcup_{n \geq 1} \bigcap_{k \geq n} R_k^{\alpha_k}$ c.s.

Sea $\epsilon \in \mathbb{Q}^+$ and $x \in R^{\alpha+\epsilon}$. La parte (I) establece que existe n_0 tal que

$$R^{\alpha+\epsilon} \subset R_k^{\alpha_k} \text{ para todo } k \geq n_0 \text{ c.s.} \Rightarrow R^{\alpha+\epsilon} \subset \bigcap_{k \geq n_0} R_k^{\alpha_k} \text{ c.s.}$$

Entonces, para todo

$$\forall \epsilon \in \mathbb{Q}^+, \quad R^{\alpha+\epsilon} \subset \bigcup_{n \leq 1} \bigcap_{k \geq n} R_k^{\alpha_k} \text{ c.s.} \Rightarrow \bigcup_{\epsilon \in \mathbb{Q}^+} R^{\alpha+\epsilon} \subset \bigcup_{n \geq 1} \bigcap_{k \geq n} R_k^{\alpha_k} \text{ c.s.}$$

Falta probar que $\limsup R_n^{\alpha_n} = \bigcap_{n \geq 1} \bigcup_{k \geq n} R_k^{\alpha_k} \subset \bigcap_{\epsilon \in \mathbb{Q}^+} R^{\alpha-\epsilon}$ c.s.

De la parte (I) del teorema, para $\epsilon \in \mathbb{Q}^+$ existe n_0 tal que $R_k^{\alpha_k} \subset R^{\alpha-\epsilon}$ para todo $k \geq n_0$ c.s. Con lo cual, $\bigcup_{k \geq n_0} R_k^{\alpha_k} \subset R^{\alpha-\epsilon}$ c.s.

Luego,

$$\bigcap_{n \geq 1} \bigcup_{k \geq n} R_k^{\alpha_k} \subset \bigcup_{k \geq n_0} R_k^{\alpha_k} \subset R^{\alpha-\epsilon} \text{ c.s.}$$

Como se cumple para todo $\epsilon \in \mathbb{Q}^+$, $\bigcap_{n \geq 1} \bigcup_{k \geq n} R_k^{\alpha_k} \subset R^{\alpha-\epsilon}$ c.s.

Finalmente y utilizando que $\liminf R_n^{\alpha_n} \subset \limsup R_n^{\alpha_n}$,

$$\bigcup_{\epsilon \in \mathbb{Q}^+} R^{\alpha+\epsilon} \subset \liminf R_n^{\alpha_n} \subset \limsup R_n^{\alpha_n} \subset \bigcap_{\epsilon \in \mathbb{Q}^+} R^{\alpha-\epsilon} \text{ c.s.} \quad (2.28)$$

Juntando los resultados (2.27) y (2.28),

$$\begin{aligned} P(\limsup R_n^{\alpha_n} \neq \liminf R_n^{\alpha_n}) &= P(\limsup R_n^{\alpha_n} - \liminf R_n^{\alpha_n}) \leq \\ &\leq P(R^\alpha - \{x \in \mathbb{E} : LD^\beta(x, P) > \alpha\}) = \\ &= P(\{x \in \mathbb{E} : LD^\beta(x, P) = \alpha\}) = 0. \end{aligned}$$

Luego el límite existe y procediendo análogamente,

$$\begin{aligned} P(R^\alpha \neq \lim_{n \rightarrow +\infty} R_n^{\alpha_n}) &= P(R^\alpha - \lim_{n \rightarrow +\infty} R_n^{\alpha_n}) \leq \\ &\leq P(R^\alpha - \{x \in \mathbb{E} : LD^\beta(x, P) > \alpha\}) = 0. \end{aligned}$$

Capítulo 3

Clusters Basados en Profundidades Locales

En este capítulo proponemos un método de clusters basado en profundidades locales. El procedimiento se puede aplicar a una amplia variedad de datos, dependiendo de la capacidad de la profundidad local para adaptarse al espacio donde se encuentren las observaciones.

Mostraremos como se puede extender inmediatamente la IDLD al contexto de datos funcionales mutidimensionales y datos mixtos. Además mostraremos como puede construirse una profundidad local combinando otras funciones de profundidad local dadas y como puede utilizarse para sortear el problema de las escalas cuando se combinan distintos tipos de datos.

Pondremos a prueba el método de cluster basado en profundidades con simulaciones para datos multivariados, funcionales y datos funcionales multidimensionales. Por último aplicaremos el procedimiento de clusters para conjuntos clásicos de datos funcionales y un ejemplo con datos reales de datos mixtos.

3.1. Procedimiento de cluster basado en profundidades locales

En esta sección introducimos un procedimiento de Clusters basado en Profundidades Locales (LDC). Es un procedimiento de partición del espacio, basado en la definición de centroides, donde cada uno de ellos en lugar de ser un punto es una región del espacio, que llamamos regiones núcleo. Aquellos puntos que no pertenecen a las regiones núcleo se asignan al grupo determinado por la región núcleo más cercana. A continuación explicaremos el método en detalle.

Sea X un elemento aleatorio en un espacio de Banach separable \mathbb{E} , con distribución P .

Etapas 1: Regiones Núcleo.

- a) Considerar la región interior de profundidad local α con nivel de localidad β , R_β^α , definida por la Ecuación (2.26).
- b) Realizar una partición de R_β^α , en K clusters, $\tilde{C}_1^\alpha, \dots, \tilde{C}_K^\alpha$, tales que $R_\beta^\alpha = \bigcup_{i=1}^K \tilde{C}_i^\alpha$, y $P(\tilde{C}_i^\alpha \cap \tilde{C}_j^\alpha) = 0$, para $i \neq j$.

Etapa 2: Asignación final.

Basado en los clusters definidos en la Etapa 1, parte b), para los puntos en R_β^α , realizar la asignación de los puntos restantes siguiendo la regla de la mínima distancia a cada región núcleo, i.e.

$$C_i^\alpha = \{x \in \mathbb{E} : d(x, \tilde{C}_i^\alpha) \leq d(x, \tilde{C}_j^\alpha) \text{ para todo, } j \neq i\},$$

donde $d(x, \tilde{C}_j^\alpha) = \inf_{y \in \tilde{C}_j^\alpha} d(x, y)$.

La idea de esta propuesta es determinar el centro de los clusters como una región del espacio en vez de utilizar un único punto. Aunque, es conocido que no existe un procedimiento de clusters que sea adecuado para toda configuración de clusters y que la elección del procedimiento adecuado debería estar fuertemente relacionada con la distribución subyacente de los datos. Nuestra idea es tener centros con una forma flexible que nos permitan capturar mejor la distribución de los clusters. Típicamente, los métodos de clusters basados en centros tienen muy buen desempeño bajo distribuciones esféricas. Cuanto más flexibilidad tenga la región central debería resultar en una mejor performance detectando la verdadera estructura de clusters bajo mayor número de distribuciones, incluyendo distribuciones elípticas. Además, dado que las medidas de profundidad tienen una relación cercana con la robustez, es esperable que las regiones núcleo sean resistentes a la presencia de datos atípicos.

Algunas consideraciones importantes a la hora de implementar el procedimiento de LDC. En la Etapa 1, las regiones núcleo pueden ser seleccionadas considerando cualquier profundidad local. Siempre que se cumplan las hipótesis del Teorema 4, el procedimiento resultará consistente. En el Etapa 1, parte b), se puede considerar cualquier procedimiento de clusters, por simplicidad en adelante nosotros usaremos el clásico algoritmo k -medias. Asumimos que el número de clusters, K , está dado de antemano, sin embargo si esto no ocurriera puede estimarse utilizando cualquier procedimiento existente en la literatura.

A continuación daremos la contraparte empírica de nuestra propuesta, que se basa en aplicar el método plug-in a la versión poblacional.

Sean X_1, \dots, X_n observaciones i.i.d. en \mathbb{E} , un espacio de Banach separable, con una estructura de K clusters. Notemos R_n^α , a la región interior de profundidad α con nivel de localidad β , y $\tilde{C}_{n,1}^\alpha, \dots, \tilde{C}_{n,K}^\alpha$, a la partición inicial obtenida al aplicar la Etapa 1, parte b). La asignación final, está dado por,

$$C_{n,i}^\alpha = \{x \in \mathbb{E} : d(x, \tilde{C}_{n,i}^\alpha) \leq d(x, \tilde{C}_{n,j}^\alpha) \text{ para toda, } j \neq i\},$$

donde $d(x, \tilde{C}_{n,j}^\alpha) = \min_{y \in \tilde{C}_{n,j}^\alpha} d(x, y)$.

3.2. Otras Aplicaciones

3.2.1. Datos Funcionales Mixtos

En esta sección analizamos en detalle la implementación del método de clusters para el caso de datos mixtos, donde por ejemplo tendremos variables funcionales y otras reales. Consideramos que este caso es de suma importancia, ya que conjuntos de datos con estas características aparecen cada vez con mayor frecuencia en problemas reales. Sin embargo, la literatura en este ámbito es escasa.

Consideramos el conjunto de datos AEMET, que se encuentra en la librería de R *fda.usc*. Este conjunto de datos es provisto por la Agencia Estatal de Meteorología Española. Cuenta con series diarias anuales para la temperatura media, precipitación media y velocidad media del viento para 73 estaciones meteorológicas desplegadas en el territorio español durante el período 1980-2009. El conjunto de datos contiene además la información geográfica de cada estación junto al año al que corresponden las series.

Estudiaremos el problema de cluster para las estaciones meteorológicas, basándonos en una variable funcional, la serie anual de velocidades medias del viento, y un conjunto de datos multivariados con tres variables geográficas: latitud, longitud, altura de las estaciones meteorológicas. Analizar estas variables de manera conjunta es relevante dado que la altura influye en la intensidad de los vientos. Si se considerasen únicamente los datos multivariados o los datos funcionales para analizar la conformación del cluster llevaría a un análisis menos rico. Aunque los sensores estén ubicados a la misma altura es posible que fenómenos relacionados con el clima de la región generen deformaciones en las curvas dadas por la intensidad del viento. Aplicaremos la técnica de clustering que introdujimos en esta tesis, para ello necesitamos ser precisos en la definición de la profundidad local dual integrada en espacios que tengan estas características.

Presentamos dos propuestas para definir profundidades locales en este contexto. Primero veremos como puede extenderse directamente la IDLD y luego una modificación para cualquier otra profundidad local que esté definida en espacios de dimensión finita e infinita.

Para formalizar esta idea podemos asumir que tenemos dos espacios de Hilbert separables, $(H_f, \langle \cdot, \cdot \rangle_f)$, de dimensión infinita para las variables funcionales y $(H_m, \langle \cdot, \cdot \rangle_m)$, de dimensión finita, para el vector multivariado que vayamos a considerar. El elemento aleatorio X pertenece a $H = H_f \oplus H_m$ que también es un espacio de Hilbert separable. El producto interno es $\langle \cdot, \cdot \rangle_{H_f \oplus H_m}$ y caracteriza los funcionales lineales continuos. Es decir, si $f \in H'$ entonces $f = f_f \oplus f_m$ donde $f_f \in H'_f$ y $f_m \in H'_m$. Más aún, existen $h_f \in H_f$ y $h_m \in H_m$ tal que $f = \langle \cdot, h_f \rangle_f + \langle \cdot, h_m \rangle_m$. El cálculo de la IDLD resulta directo basta considerar una medida Q independiente de la medida de probabilidad asociada con X . En particular podría tomarse Q como la medida producto considerando Q_f una medida para H'_f y Q_m una medida para H'_m . Ambas independientes de P la medida de probabilidad asociada a X .

En este enfoque hay que tener mucho cuidado con la escala de las variables. La suma de los productos internos respectivos caracteriza las proyecciones en las cuales se basa la IDLD.

Dado que las variables funcionales en general tienen mayor escala que las multivariadas, la variabilidad de las mismas podría terminar absorbiendo la información que queremos incorporar. En sentido contrario también puede ocurrir; una variable multivariada con una escala mucho mayor que el resto tendría el mismo efecto. Una solución es estandarizar las variables, sin embargo no es muy claro en la literatura como hacerlo.

El segundo enfoque trata de sortear el problema que mencionamos anteriormente. Se basa en que la IDLD puede calcularse en ambos espacios y una vez obtenido el resultado, no hay un problema de escalas. Luego de calcular las profundidades por separado para cada espacio necesitamos una forma de combinarlas. Nuestra propuesta es tomar una combinación convexa. Trataremos de mostrar que esta combinación tiene sentido, es decir que volvemos a encontrarnos con una función de profundidad local.

Generalizamos estas ideas al marco de espacios de Banach. Supongamos que tenemos una colección finita de espacios de Banach $(\mathbb{E}_i, \|\cdot\|_{\mathbb{E}_i})$, $1 \leq i \leq L$ consideramos el espacio producto $\mathbb{E} = \mathbb{E}_1 \times \cdots \times \mathbb{E}_L$ y notamos $x \in \mathbb{E}$ como $x = (x_1, \dots, x_L)$. Dotamos al espacio \mathbb{E} con la norma:

$$\|x\|_{\mathbb{E}} = \sum_{i=1}^L \|x_i\|_{\mathbb{E}_i}.$$

Utilizando esta norma nos queda caracterizado el dual, $\mathbb{E}' = \mathbb{E}'_1 \times \cdots \times \mathbb{E}'_L$ y los elementos en este conjunto son $f(x) = \sum_{i=1}^L f_i(x_i)$. La norma para un elemento del dual se calcula $\|f\|_{\mathbb{E}'} = \max_{1 \leq i \leq L} \|f_i\|_{\mathbb{E}_i}$, pues

$$\begin{aligned} |f(x)| &= \sum_{i=1}^L |f_i(x_i)|_{\mathbb{E}_i} \leq \sum_{i=1}^L \|f_i\|_{\mathbb{E}'_i} \|x_i\|_{\mathbb{E}_i} \leq \sup_{1 \leq j \leq L} \|f_j\|_{\mathbb{E}'_j} \sum_{i=1}^L \|x_i\|_{\mathbb{E}_i} = \max_{1 \leq j \leq L} \|f_j\|_{\mathbb{E}'_j} \|x\|_{\mathbb{E}}, \\ \|f\|_{\mathbb{E}'} &= \sup_{\|x\|_{\mathbb{E}}=1} |f(x)| \leq \max_{1 \leq i \leq L} \|f_i\|_{\mathbb{E}'_i}. \end{aligned}$$

Luego, la igualdad se alcanza considerando la proyección correcta que logra el máximo.

Definición 19. Sean $\mathbb{E}_1, \dots, \mathbb{E}_L$ espacios de Banach separables y $\mathbb{E} = \mathbb{E}_1 \times \cdots \times \mathbb{E}_L$. Sea X un elemento aleatorio con \mathbb{E} con medida de probabilidad asociada P . Sea $\Pi_i : \mathbb{E} \rightarrow \mathbb{E}_i$ la proyección sobre \mathbb{E}_i , Q una medida de probabilidad en \mathbb{E}' independiente de P y $\beta \in (0, 1]$ Definimos la Profundidad Local Dual Integrada Mixta (MIDLD) como

$$MIDLD^\beta(x, P) = \sum_{i=1}^L \delta_i IDLD^\beta(x_i, P_i) \quad (3.1)$$

donde $(\delta_i)_{i=1}^L \subset \mathbb{R}_{\geq 0}$ tal que $\sum_{i=1}^L \delta_i = 1$, $X_i = \Pi_i(X)$ y P_i es la medida de probabilidad asociada a X_i dada por las proyección.

Hemos definido una nueva versión de la IDLD, sin embargo, para ser consistentes deberíamos chequear que la MIDLD satisface las propiedades establecidas por Serfling (57)

para ser una función de profundidad local. De ahora en adelante mantenemos la notación de la definición.

P.1.: La propiedad debería cumplirse para transformaciones ortogonales en espacios de dimensión finita. Bajo las mismas condiciones de la Definición 19 si tenemos $\mathbb{E}_1, \dots, \mathbb{E}_L$ espacios de dimensión finita y $A : \mathbb{E} \rightarrow \mathbb{E}$ es ortogonal, tal que $Ax = (A_1x, \dots, A_Lx)$ con $A_i : \mathbb{E}_i \rightarrow \mathbb{E}_i$ ortogonal.

$$\begin{aligned} MIDLD^\beta(Ax, P_{AX}) &= \sum_{i=1}^L \delta_i IDLD^\beta(A_i x_i, P_{A_i x_i}) = \sum_{i=1}^L \delta_i IDLD^\beta(x_i, P_{x_i}) = \\ &= MIDLD^\beta(x, P_X). \end{aligned}$$

P.2.: Supongamos que $\theta = (\theta_1, \dots, \theta_L) \in \mathbb{E}$ es un punto de β -simetría. Recordemos que el dual nos queda caracterizado por $\mathbb{E}' = \mathbb{E}'_1 \times \dots \times \mathbb{E}'_L$. Esperaríamos que la proyección de dicho punto sobre alguno de los \mathbb{E}_i sea también un punto de β -simetría. Para recuperar el resultado para \mathbb{E}_i basta con considerar los funcionales lineales de la forma $(0, \dots, 0, f_i, 0, \dots, 0)$. Luego,

$$MIDLD^\beta(\theta, P) = \sum_{i=1}^L \delta_i IDLD^\beta(\theta_i, P_i) = \sum_{i=1}^L \delta_i \frac{1}{2} = \frac{1}{2}.$$

P.3: Consideremos un punto de C -simetría, $\theta \in \mathbb{E}$. Razonando análogamente a la propiedad (**P.2**), θ_i resulta un punto de C -simetría en \mathbb{E}_i . Sea $t \in [0, 1]$, $x \in \mathbb{E}$ y $\beta \in (0, 1]$, luego

$$\begin{aligned} MIDLD^\beta(x, P) &= \sum_{i=1}^L \delta_i IDLD^\beta(x_i, P_i) \leq \sum_{i=1}^L \delta_i IDLD^\beta((1-t)\theta_i + tx_i, P_i) \\ &= MIDLD^\beta((1-t)\theta + tx, P_X). \end{aligned}$$

P.4: Es inmediata por las propiedades de la suma y el límite,

$$IDLD^\beta(x_i, P_i) \xrightarrow{\|x\| \rightarrow +\infty} 0 \text{ para todo } 1 \leq i \leq L \implies MIDLD^\beta(x, P) \xrightarrow{\|x\| \rightarrow +\infty} 0.$$

P.5.-P.6.: Inmediatas por las propiedades de la continuidad respecto a la suma.

Proposición 5. Sean $\mathbb{E}_1, \dots, \mathbb{E}_L$ espacios de Banach separables y $\mathbb{E} = \mathbb{E}_1 \times \dots \times \mathbb{E}_L$. Sea X un elemento aleatorio con \mathbb{E} con medida de probabilidad asociada P . Sea $\Pi_i : \mathbb{E} \rightarrow \mathbb{E}_i$ la proyección sobre \mathbb{E}_i , Q una medida de probabilidad en \mathbb{E}' independiente de P y $\beta \in (0, 1]$. Entonces, la $MIDLD$ cumple las propiedades **P.1.** - **P.6.** mencionadas en la Sección 2.2.1.

Es claro que se puede construir una medida de profundidad local mixta a partir de medidas de profundidad local definidas en cada uno de los espacios de Banach originales, las

propiedades de esta nueva medida de profundidad local serán heredadas de las medidas de profundidad local originales. Es inmediato que aplicando el método plug-in se puede obtener una versión empírica de la MIDLD a partir de las versiones empíricas de la profundidad que se desee utilizar. En este escenario la convergencia dependerá de la profundidad local elegida.

Cabe mencionar que la MIDLD tiene una debilidad con respecto a la IDLD. Si bien hereda las propiedades mencionadas por Serfling (57), no podemos asegurar que si la MIDLD se maximiza en un punto, entonces este punto es de β – simetría. Podemos ilustrarlo con un ejemplo muy sencillo:

Consideremos el vector aleatorio $(X, Y) \in \mathbb{R}^2$ con función de densidad:

$$f_{XY}(x, y) = (2x + 2y - 4xy)I_{[0,1]}(x)I_{[0,1]}(y).$$

Podemos observar que $X \sim U[0, 1]$ e $Y \sim U[0, 1]$ por lo que $x = 1/2$ e $y = 1/2$ son puntos de β – simetría para X e Y respectivamente, pero no es cierto que $(1/2, 1/2)$ sea un punto de β – simetría para (X, Y) dado que su distribución conjunta no es simétrica. La raíz del problema se encuentra en que la MIDLD depende de la distribución marginal de cada dato mixto y no de la distribución conjunta; con lo cual no tiene la capacidad de captar toda la información.

3.3. Simulaciones y Datos Reales

En esta sección analizaremos numéricamente el desempeño del procedimiento de clustering que introdujimos en la Sección 3.1. Las simulaciones comprenden los casos finito e infinito dimensional. En el último, tuvimos en cuenta datos funcionales univariados y multivariados. Además hemos puesto a prueba el método en varios conjuntos reales de datos funcionales muy conocidos en la literatura. El LDC fue implementado considerando, no solamente la IDLD, sino también otras propuestas para medidas de profundidad locales disponibles.

3.3.1. Criterios para Comparar Particiones

En el análisis de clusters no existe un criterio común aceptado para evaluar el desempeño de los procedimientos, ni una cota inferior que indique la dificultad del problema como ocurre en la clasificación supervisada, donde el error de Bayes cumple ese papel. En general, en problemas de clasificación supervisada basta con contar la proporción de observaciones correctamente clasificadas en una muestra de validación. Por el contrario, en problemas de cluster los datos no tienen este tipo de información. En este contexto, se pueden encontrar dos situaciones, por un lado comparar la performance de un método de cluster versus resultados estándares (conocidos) y por otro lado necesitar compararlos con los resultados de otros métodos.

Si bien en la literatura hay muchas propuestas para medir la performance de un procedimiento de cluster (ver (25)), nosotros desarrollaremos dos de ellas que son las que mayor

relevancia tienen, en ambos casos subyace la idea de contar la proporción de observaciones mal clasificadas. En ambos casos el objetivo es comparar dos particiones distintas.

Tasa correcta de clasificación (CCR)

Esta medida es adecuada cuando una de las dos particiones es conocida, esto suele ocurrir cuando los datos son simulados, o en el caso de que hubiera un antecedente confiable de partición de los datos, por ejemplo en conjuntos de datos que son conocidos y se utilizan para validar un nuevo procedimiento de cluster. En general esta medida es muy utilizada en problemas de aprendizaje supervisado.

Sean y_1, \dots, y_n las etiquetas que le corresponden a la partición conocida de n observaciones en K clusters. Por otro lado, se tienen $\widehat{y}_1, \dots, \widehat{y}_n$, las etiquetas asignadas por el procedimiento de cluster del cual se quiere evaluar su desempeño, donde también se segmenta en K clusters. El CCR está definido como:

$$CCR = 1 - \min_{\sigma \in \Sigma} \frac{1}{n} \sum_{i=1}^n \mathcal{I}_{\{y_i \neq \sigma(\widehat{y}_i)\}}, \quad (3.2)$$

donde Σ es el conjunto de todas las posibles permutaciones de las etiquetas predichas. En este caso es necesario tomar todas las posibles permutaciones de las etiquetas predichas, porque a diferencia del problema de clasificación las etiquetas en cluster no tienen un significado. Veámoslo en un ejemplo sencillo. Supongamos que tenemos un conjunto de cinco observaciones, donde las tres primeras pertenecen a un grupo y las dos últimas a otro grupo, las particiones $\{1, 1, 1, 2, 2\}$ y $\{2, 2, 2, 1, 1\}$ son iguales, solo difieren en la etiqueta asignada, permutar las etiquetas es necesario para detectar estos casos.

Este criterio tiene dos inconvenientes, el primero es que hay que contar con una partición conocida y claramente en problemas reales esto nunca ocurre. El segundo es que si el número de clusters es grande el problema se vuelve difícil de tratar computacionalmente porque la cantidad de permutaciones que hay que considerar es alta. Este último problema se puede resolver en tiempo polinomial utilizando técnicas para apareamiento bipartito mediante el algoritmo Húngaro (43).

Indice de Rand Ajustado (ARI)

Hubert y Arabie (25) introducen este criterio, que es más flexible dado que permite comparar dos particiones en forma simétrica, típicamente son dos particiones provenientes de dos criterios de cluster y se quiere estudiar el grado de concordancia entre ambas. Más formalmente, un conjunto de n datos, sean $U = \{U_1, \dots, U_K\}$ y $V = \{V_1, \dots, V_L\}$ dos particiones del mismo. Es importante observar que el número de clusters de la dos particiones no necesariamente coincide. La idea del método consiste en considerar todos los pares posibles de observaciones analizar si ambos pertenecen a un mismo grupo o a grupos distintos en las dos particiones. La superposición entre las dos particiones puede resumirse en una tabla de contingencia de K filas por L columnas $M = [n_{i,j}]_{1 \leq i \leq K, 1 \leq j \leq L}$, donde $n_{i,j}$ representa el número de datos que son comunes a U_i y V_j . Luego, los $\binom{n}{2}$ pares de observaciones pueden clasificarse en 4 categorías:

- N_{11} : el número de pares de observaciones donde ambas observaciones pertenecen a un mismo cluster en U y en V .
- N_{10} : el número de pares de observaciones donde las dos observaciones se encuentran en el mismo cluster en U , pero en diferente cluster en V .
- N_{01} : el número de pares de observaciones donde las dos observaciones se encuentran en distinto cluster en U , pero en el mismo cluster en V .
- N_{00} : el número de pares de observaciones donde ambas observaciones pertenecen a un mismo cluster en U y en V .

Si N_{00} y N_{11} son altos en comparación con N_{01} y N_{10} las dos particiones tendrán un alto grado de concordancia. Finalmente, se define el ARI como

$$ARI(U, V) = \frac{2(N_{00}N_{11} - N_{01}N_{10})}{(N_{00} + N_{01})(N_{01} + N_{11}) + (N_{00}N_{10})(N_{10}N_{11})}.$$

Este es un número acotado por 0 y 1. Cuando el $ARI = 1$ nos dice que la concordancia entre las particiones es perfecta. Este procedimiento no tiene un alto costo computacional.

3.3.2. Simulaciones: datos multivariados

El propósito de esta sección es evaluar el desempeño de nuestra propuesta bajo una basta diversidad de distribuciones, que presenten desafíos usuales en problemas de aprendizaje no supervisado. Específicamente, analizamos los casos donde los datos son esparsos, presentan observaciones atípicas o los tamaños de los clusters son heterogéneos. Para este objetivo trabajamos con catorce escenarios diferentes en los cuales la distribución original de las variables fue propuesta por Witten y Tibshirani (55) y luego extendida por Kondo et al (31). En todos los casos, comparamos con otros métodos de clusters conocidos, los cuales describiremos brevemente.

En los catorce modelos considerados los datos tienen una estructura de tres grupos. Cada uno de ellos tiene 300 observaciones, en los doce primeros modelos los clusters tienen todos el mismo tamaño. Los conjuntos de datos fueron generados de la siguiente forma.

Modelo 1: Los datos son esféricos y fueron generados a partir de distribuciones normales, $N(\mu_i, \Sigma)$, para $i = 1, 2, 3$, con medias $(-3, -3, 0)$, $(0, 0, 0)$, $(3, 3, 0)$. Los tres grupos tienen como matriz de covarianza la identidad. La Figura 3.1 muestra las dos primeras coordenadas de un conjunto de datos generados siguiendo el Modelo 1.

Modelo 2: Los datos son elípticos y fueron generados a partir de distribuciones normales $N(\mu_i, \Sigma)$, para $i = 1, 2, 3$, con medias $(-3, -3, 0)$, $(0, 0, 0)$, $(3, 3, 0)$, las mismas que en el Modelo 1. Mientras, que si bien la matriz de covarianza es la misma para los tres grupos en este caso "a dada por $\Sigma = \text{diag}(3, 0, 25, 1)$.

En estos dos modelos, las primeras dos variables son informativas para determinar la estructura de cluster, mientras que la tercera es ruido.

El Modelo 3, (respectivamente Modelo 4) tiene cinco variables. Las primeras tres siguen la distribución como en el Modelo 1 (respectivamente Modelo 2). El resto de las variables son independientes con distribución normal estándar y sólo son ruido.

Por otra parte, estudiamos el desempeño del procedimiento en presencia de outliers. Para ello consideramos dos configuraciones distintas para la contaminación para cada uno de los modelos. En cada una de ellas reemplazamos una coordenada de cinco observaciones por datos atípicos que siguen una distribución uniforme en el intervalo $[25, 25,01]$. En la primera configuración reemplazamos cinco observaciones de la primer coordenada (informativa) correspondiente al primer cluster (Modelos 5 a 8,) el resto de la distribución la mantuvimos sin modificaciones siguiendo en los Modelos 1 a 4. En los Modelos 9 a 12 el esquema es el mismo, pero la contaminación es alojada en la última coordenada, que no es informativa.

Los dos modelos restantes, Modelo 13 y 14, tienen clusters con tamaño desbalanceado, se siguen las mismas distribuciones que en los Modelos 1 y 2, pero en lugar de tener 100 observaciones cada cluster, el primer cluster tiene el 60 % de las observaciones, mientras que los dos clusters restantes tienen el 20 % cada uno de ellos.

Tomamos como punto de referencia los siguientes métodos de cluster para la comparación:

- *k*-medias. Es el procedimiento de cluster por partición del espacio más difundido por su buen desempeño en ejemplos concretos, sus buenas propiedades teóricas y la facilidad de uso ya que se encuentra incorporado en todos los programas de estadística. La idea central es hallar *k* centros y asignar las observaciones a su centro más cercano, de modo tal de minimizar la varianza dentro de cada grupo. Desde el punto de vista numérico si bien el problema es *NP*-completo, hay algoritmos que permiten converger rápidamente a mínimos locales. Estos algoritmos dependen fuertemente de la inicialización de los centros, para mejorar la identificación de clusters se aconseja realizar 10 inicializaciones aleatorias.
- El procedimiento para datos esparzos basado en *k*-medias (Sparse *k*-means clustering procedure, SKM), introducido por Witten y Tibshirani (55). El método es un extensión de *k*-medias específicamente diseñada para tratar problemas con datos en alta dimensión donde la proporción de variables informativas en relación con la dimensión es pequeña. Las características relevantes son seleccionadas vía procedimientos del tipo Lasso. El algoritmo depende de un parámetro, la cota L_1 , que fue elegida como sugiere la literatura ($s = 3, 7$). Al igual que en *k*-medias el algoritmo depende de la inicialización de los centros, por lo tanto se consideraron 5 inicializaciones aleatorias.
- El procedimiento para datos esparzos y robustos basado en *k*-medias (Robust and Sparse *k*-means clustering procedure, RSKM), propuesto por Kondo et al. (31). Es una extensión de SKM diseñada para tratar no sólo con conjuntos de datos esparzos sino

también bajo contaminación. Los dos parámetros de inicialización los hemos ajustado como se sugiere en la literatura, tomando el parámetro correspondiente a la cota de la norma L_1 igual a 4 y la proporción de corte en 0,1.

- Fraley y Raftery ((19),(20)), propusieron el criterio de cluster MCLUST. El mismo asume que los datos fueron generados por una mixtura normal k con matrices de covarianza elipsoidales con forma, volumen y orientación variables. Los parámetros para cada población se estiman mediante máxima y el modelo más apropiado es seleccionado basándose en el Criterio de Información Bayesiano.

EL procedimiento LDC introducido en la Sección 3.1 lo implementamos considerando tres definiciones para profundidades locales. El parámetro α representa la proporción de datos que contendrán las regiones núcleo de los clusters, si este valor es muy pequeño el procedimiento tendrá un comportamiento muy similar a k -medias, no pudiendo capturar la estructura central de los clusters. Si toma valores altos, las regiones núcleo tendrán observaciones con profundidad local moderada, eso puede llevar a errores en las asignaciones. Por estos motivos sugerimos tomar valores entre 0,15 y 0,4. Para fijar el parámetro α , realizamos un análisis de la sensibilidad de los parámetros, siguiendo las ideas de remuestreo propuestas por Hennig (21), a partir de ellas pudimos ver que en todos los casos el método es estable, como en la mayor parte de los casos $\alpha = 0,4$ mostró ligeramente mejor desempeño nos quedamos con este valor a lo largo del estudio.

- La Profundidad Local Simplicial (LDCS) introducida por Agostinelli y Romanazzi (2). Utilizamos el paquete de R *localdepth*, El parámetro τ , fue calculado con la función *quantile.localdepth* como sugiere la librería, el cuantil de los estadísticos de orden se fijó en *probs = 0,1*.
- La versión local de profundidad con nivel de localidad β (LDCPV) de acuerdo con la propuesta de Paindaveine y Van Bever (45). Utilizamos el paquete de R *DepthProc*, y fijamos el parámetro de localidad β igual a 0,2.
- La Profundidad Dual Local Integrada con nivel de localidad β , (LDCI) que introdujimos en la Sección 2.2.1. Al igual que para LDCPV, fijamos $\beta = 0,2$ y consideramos $N = 50$ proyecciones al azar con distribución normal estándar.

En los dos últimos, a la hora de seleccionar el parámetro de localidad β tuvimos en cuenta, que si se consideran valores altos, la profundidad local se irá pareciendo mucho a la profundidad global, dejando de capturar modas locales, mientras que si el valor es bajo se potenciarán efectos locales pequeños pudiéndose detectar modas espúreas, por lo tanto consideramos $\beta = 0,2, 0,3$ y $0,4$, y siguiendo las ideas de remuestreo de Hennig, pudimos observar que el procedimiento es estable y tomamos $\beta = 0,2$ para todos los casos, que era el valor donde globalmente se observaba mejor desempeño.

Realizamos $M = 500$ réplicas para cada modelo.

El resultado de la simulación se muestra en la Tabla 3.1. Era esperado que todos los métodos tuvieran una muy buena performance bajos los Modelos 1 y 3, donde todos los clusters son esféricos sin datos atípicos. Bajo los Modelos 2 y 4 los clusters tiene distribución elíptica, MCLUST tiene un desempeño excepcional y es claro que LDC, con cualquier alternativa, es mejor que las otras tres alternativas. En los Modelos 5 a 12, dado que k -medias, SKM y MCLUST no son robustos fallan, típicamente los 5 datos atípicos conforman un grupo y el cluster con media $(0, \dots, 0)$ es dividido entre los otros grupos. LDC y RSKM están basados en criterios robustos, lo que los lleva a tener buen desempeño. RKSM parece funcionar mejor bajo distribuciones esféricas y LCD cuando la distribución es elíptica. En los doce primeros modelos LDC tiene buenos resultados sin importar la elección de la profundidad local. Sin embargo, cuando los tamaños de los clusters no son parejos, los únicos criterios capaces de detectar la estructura de cluster son MCLUST y la LDC considerando la profundidad local dual integrada. Es claro que LDC en combinación con las otras dos propuestas de profundidades locales no es capaz de detectar correctamente los centros de los clusters. En cuanto los otros criterios de cluster tienen buena performance en el caso esférico, pero fallan en el caso elíptico.

En resumen, se puede ver que LDCI es el único procedimiento de cluster lo suficientemente versátil como para detectar correctamente los clusters bajo situaciones adversas (datos esparsos, presencia de outliers y cluster desbalanceados en tamaño).

Cuadro 3.1: CCR para cada criterio y configuración de la distribución.

Modelo	k -medias	SKM	RSKM	MCLUST	LDCS	LDCPV	LDCI
1	0,98	0,98	0,98	0,98	0,96	0,95	0,97
2	0,87	0,80	0,86	0,99	0,91	0,87	0,91
3	0,98	0,98	0,98	0,98	0,96	0,96	0,97
4	0,87	0,80	0,85	0,99	0,89	0,90	0,90
5	0,66	0,70	0,96	0,65	0,95	0,92	0,95
6	0,65	0,62	0,84	0,66	0,90	0,85	0,87
7	0,67	0,70	0,96	0,65	0,94	0,94	0,95
8	0,65	0,62	0,84	0,66	0,88	0,89	0,87
9	0,65	0,68	0,98	0,65	0,95	0,94	0,95
10	0,65	0,65	0,86	0,66	0,91	0,84	0,89
11	0,65	0,67	0,98	0,65	0,95	0,86	0,96
12	0,65	0,66	0,85	0,66	0,88	0,95	0,90
13	0,97	0,98	0,98	0,97	0,54	0,46	0,96
14	0,74	0,70	0,69	0,98	0,52	0,43	0,82

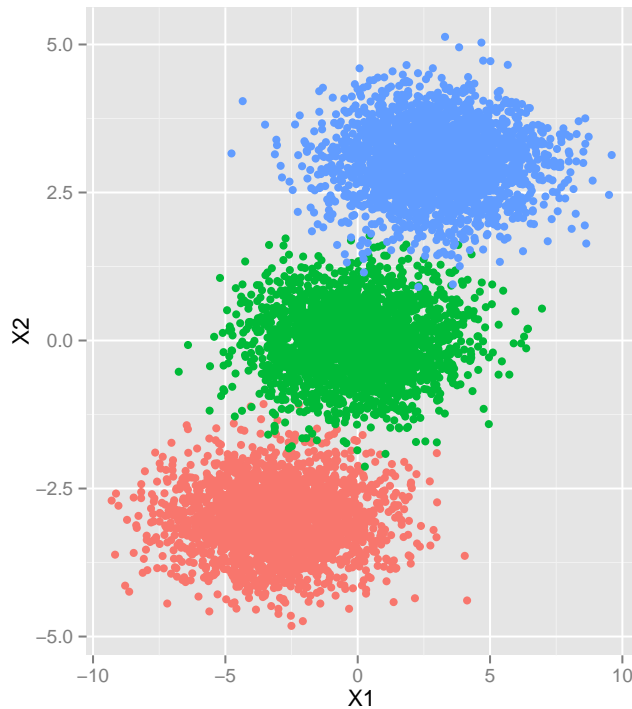


Figura 3.1: Scatter plot de las dos primeras variables para datos generados con el Modelo 1.

Por otra parte, analizamos el desempeño computacional de las tres definiciones de profundidades locales, considerando diferentes tamaños muestrales y dimensión de los datos. Generamos datos de acuerdo al Modelo 3, pero en lugar de tener 3 variables no informativas agregamos $p - 2$, ($p = 5, 35, 65$) variables aleatorias independientes no informativas con distribución normal estándar. También consideramos diferentes tamaños de muestra, $n = 300, 2100, 3900$ y 5700 . Para la ILDL generamos 50 direcciones al azar. El tiempo de cómputo crece exponencialmente con el aumento de la dimensión, por esta razón, sólo realizamos $M = 50$ replicas bajo cada escenario.

En la Tabla 3.2 se presentan los resultados de la simulación. Se observa que en todos los casos IDLD es el procedimiento más rápido, más aún, no se ve afectado por la dimensión del conjunto de datos. Sin embargo LDS y LDPV tienen un aumento en el tiempo computacional dramático cuando p crece. LDPV es por lejos el procedimiento más lento. Aunque todos los procedimientos demandan más tiempo cuando aumenta el tamaño muestral, IDLD tiene la menor tasa de crecimiento.

Cuadro 3.2: Tiempo de cómputo medio para LDS, LDPV and IDLD.

p		n			
		300	2100	3900	5700
5	LDS	0,785	38,27	131,65	280,66
	LDPV	4,236	100,08	292,67	624,91
	IDLD	0,397	20,74	73,74	160,43
35	LDS	1,770	86,88	299,03	638,38
	LDPV	7,840	200,94	629,07	1363,97
	IDLD	0,402	20,68	74,41	160,29
65	LDS	3,788	184,92	641,01	1368,31
	LDPV	10,934	288,79	982,79	2094,89
	IDLD	0,406	20,66	75,07	164,40

3.3.3. Simulaciones: datos funcionales

Esta sección es dedicada al estudio empírico de nuestro método cuando los datos son funcionales. Las nuevas tecnologías nos permiten recopilar datos con exactitud y alta frecuencia en una variable temporal o espacial generando grandes volúmenes de información. El aumento en la velocidad de procesamiento de las computadoras hizo que el análisis de esta clase de datos sea viable no sólo en grandes centros de cómputos sino también en computadoras personales. Esta democratización hizo que los datos funcionales se conviertan en un tema de investigación estadística muy activo. Además cabe mencionar que se utilizan para modelar fenómenos en la ciencia, ingeniería, medicina, economía y cualquier tópico donde tenga sentido registrar una curva. En este caso las dos medidas de profundidad local que conocemos son la IDLD, propuesta en esta tesis, y la profundidad Local del Semiespacio para datos funcionales de Agostinelli (1) (LDCH).

Al implementar el procedimiento de cluster LDC considerando como medida de profundidad local IDLD, los parámetros α y β toman los mismos valores que en la simulación para datos multivariados y fueron seleccionados siguiendo el mismo criterio. El número de proyecciones al azar, que se generaron siguiendo una distribución Browniano, para cada réplica lo fijamos en $N = 50$. Por otra parte al implementar el procedimiento de cluster LDC considerando la profundidad local LDCH, la elección del parámetro τ se realizó del mismo modo que en el caso multivariado.

Llevamos a cabo un estudio de simulación para modelos sintéticos que ya han sido analizados por Justel and Svarc (30) cuando introdujeron el método de clusters para datos funcionales DivClusFD.

Primero, analizamos tres conjuntos de datos que presentan warping y el último de ellos

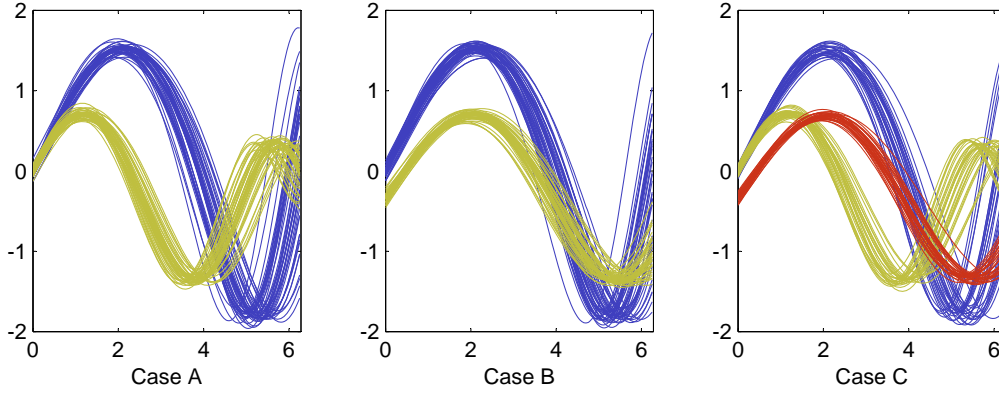


Figura 3.2: Ejemplo de un conjunto de datos simulado con los Modelos A, B y C.

presenta errores de muestreo puntuales.

Modelo A: Dos clusters con $n/2$ funciones generados por,

$$X_i(t) = (1 + \epsilon_{1i}) \sin(\epsilon_{3i} + \epsilon_{4i}t) + (1 + \epsilon_{2i}) \sin\left(\frac{(\epsilon_{3i} + \epsilon_{4i}t)^2}{2\pi}\right), \quad (3.3)$$

$$t \in [0, 2\pi], \text{ para } i = 1, \dots, n/2,$$

$$X_i(t) = (1 + \epsilon_{1i}) \sin(\epsilon_{3i} + \epsilon_{4i}t) - (1 + \epsilon_{2i}) \sin\left(\frac{(\epsilon_{3i} + \epsilon_{4i}t)^2}{2\pi}\right), \quad (3.4)$$

$$t \in [0, 2\pi], \text{ para } i = n/2 + 1, \dots, n.$$

Modelo B: Dos clusters con $n/2$ funciones generadas de la siguiente forma. En el primer grupo siguiendo la distribución dada por la Ecuación (3.3) y en el segundo grupo,

$$X_i(t) = (1 + \epsilon_{1i}) \sin\left(\epsilon_{3i} + \epsilon_{4i}\left(-\frac{1}{3} + \frac{3}{4}t\right)\right) - (1 + \epsilon_{2i}) \sin\left(\frac{(\epsilon_{3i} + \epsilon_{4i}\left(-\frac{1}{3} + \frac{3}{4}t\right))^2}{2\pi}\right), \quad (3.5)$$

$$t \in [0, 2\pi], \text{ for } i = n/2 + 1, \dots, n.$$

Modelo C: Tres clusters con $n/3$ funciones generadas cada uno de ellos. El el primer grupo siguiendo la distribución dada por la Ecuación (3.3), en el segundo grupo definido como en la Ecuación (3.4) y por último el tercer grupo sigue la distribución dada por la Ecuación (3.5).

Modelo D: Cuatro grupos con $n/4$ funciones cada uno de ellos, generadas de la siguiente manera:

$$X_{ij}(t) = f_j(t) + \epsilon_i(t), \quad (3.6)$$

$$\text{para } t \in [0, 1], i = 1, \dots, n/4 \text{ y } j = 1, \dots, 4,$$

donde

$$f_1(t) = \min\left(\frac{2-5t}{2}, \left(\frac{2-5t^2}{2} \sin\left(\frac{5\pi t}{2}\right)\right)\right),$$

$$f_2(t) = -f_1(t), \quad f_3(t) = \cos(2\pi t) \quad \text{y} \quad f_4(t) = -f_4(t).$$

Los conjuntos de datos tienen $n = 90$ observaciones para los Modelos A, B y C. Para el Modelo D el tamaño muestral es $n = 600$. Todos los errores $\epsilon_{1i}, \dots, \epsilon_{4i}$ son independientes y normales con media 0 y desvío estándar 0.05. La Figura 3.2 muestra tres ejemplos de 90 curvas generadas de acuerdo con los Modelos A, B y C. En la Ecuación (3.6) los errores, $\epsilon(t)$, son normales con media 0,4, desvío estándar 0,9 y con estructura de covarianza dada por,

$$\rho(s, t) = 0,3 \exp\left(-\frac{(s-t)^2}{0,3}\right), \quad \text{for } s, t \in [0, 1].$$

Recientemente, Yassouridis y Leisch (56) han hecho una reseña de varios procedimientos de clusters para datos funcionales que hemos utilizado como puntos de referencia para analizar el desempeño de nuestra propuesta, además incluimos el método DivClusFD. Antes de exhibir los resultados de la simulación hacemos una un breve reseña de cada método de los procedimientos de cluster alternativos considerados.

Fitclust (James y Sugar (29)): el método proyecta cada curva $Y_i(t)$, $1 \leq i \leq n$ en una base de splines de dimensión q . Asume que los coeficientes de la base provienen de una distribución Gaussiana con parámetros específicos que determinan cada cluster. Resultando

$$Y_i(t_j) = \sum_{j=1}^q s_{ij}(t_j)\eta_j + \epsilon_i(t_j).$$

Sea η_i el vector de la base de coeficientes de la curva i . Entonces η_i puede ser reemplazado por

$$\eta_i = \mu_{z_i} + \gamma_i.$$

μ_{z_i} es una media específica del cluster, γ_i es la desviación de la curva de esta media y $z_i \in \{1, \dots, K\}$ identifica el grupo al que corresponde la curva i . La dimensión de la base puede reducirse asumiendo que las medias de cada cluster son el resultado de sumar una media global con un término específico para cada cluster, que vive en un subespacio de dimensión $h \leq \min(q, K-1)$ tal que $\mu_K = \lambda_0 + \Lambda\alpha_K$. Además, se supone que $\epsilon_i \sim N(0, \sigma)$ y γ_i proviene de una distribución normal multivariada con media cero y con matriz de covarianza Γ .

Finalmente, la distribución conjunta con parámetros $\theta_i = (\Gamma, \sigma, \lambda_0, \Lambda, \alpha_i)$ con probabilidad π_i de pertenecer al cluster i puede escribirse

$$L(\theta_1, \dots, \theta_K | y_1, \dots, y_n) = \prod_{i=1}^n \sum_{k=1}^K \pi_k f_k(y_i | \theta_k)$$

Lo que lleva a estimar los parámetros por el estimador de máxima verosimilitud.

Distclus (Peng and Mueller (46)): la idea central es representar las curvas en una base de funciones truncada, a continuación aplicar algún procedimiento de cluster, para datos multivariados, a los coeficientes en los cuales fueron expresadas las funciones. Las observaciones Y_i se asumen como resultado de un proceso estocástico,

$$Y_i(t) = X_i(t) + \epsilon_i(t).$$

Considerar la distancia $L_2(T)$ entre las curvas Y_i e Y_j , i.e.,

$$D(Y_i, Y_j) = \left(\int_T (X_i(t) - X_j(t))^2 dt \right).$$

Se asume que $X(t)$ tiene función de covarianza continua dada por $C(s, t) = \text{cov}(X(s), X(t))$. $C(s, t)$ es un núcleo y por lo tanto existe una base ortonormal de autofunciones ϕ_j de $C(s, t)$ tales que sus correspondientes autovalores ξ_j son no negativos. Es posible escribir el proceso $X(t)$ en la expansión de Karhunen-Loeve (K-L)

$$X_i(t) = \mu(t) + \sum_{j \geq 1} \xi_{ij} \phi_j(t),$$

donde $\mu = E(X(\cdot))$ y $\xi_{ij} \sim N(0, \Lambda_j)$.

Usualmente se obtiene una buena representación truncando una cantidad q de autofunciones. Luego para calcular los clusters se aplica cualquier método de partición sobre los coeficientes.

IterSubspace (Chiou and Li (10)): la idea de este método es similar a Distclus, ya que se busca representar a las curvas en una base de funciones, en este caso a cada curva se asigna a su subespacio más cercano. Utilizando la expansión K-L se calculan los scores ξ_1, \dots, ξ_n que se clusterizan en un primer paso utilizando k -medias. El subespacio está definido para cada cluster c por la media específica del subespacio $\mu^{(c)}$ y sus autofunciones $\phi_j^{(c)}$. Cada curva es proyectada en todos los subespacios c :

$$Y_i^{(c)} = \mu^{(c)}(t) + \sum_{j=1}^{q_c} \xi_{ij}^{(c)} \phi_j^{(c)}(t). \quad (3.7)$$

Si y es una realización de (3.7) se asigna al subespacio de acuerdo a

$$c^*(y) = \operatorname{argmin}_{c \in \{1, \dots, k\}} \|y - y^{(c)}\|_{L^2}.$$

Luego hay un paso de reclasificación para cada curva i asignada anteriormente al cluster c ,

$$y_{-i}^{(c)} = \mu_{-i}^{(c)}(t) + \sum_{j=1}^{q_c} \hat{\xi}_{ij}^{(c)} \phi_{j,-i}^{(c)}(t),$$

donde $\mu_{-i}^{(c)}$, $\hat{\xi}_{ij}^{(c)}$ y $\phi_{j,-i}^{(c)}$ son la media, los scores y las autofunciones, basadas en las curvas del cluster c , pero habiendo sacado la i -ésima curva para el cálculo. Luego, la curva y_i es clasificada en el cluster c^* de acuerdo con

$$\arg \min_k \left(\min_{k \in 1, \dots, G, i \in k} \|y_i - y_{-i}^k\|_{L^2}, \min_{k \in 1, \dots, G, i \notin k} \|y_i - y^k\|_{L^2} \right).$$

Waveclust (Giacofci et al. (22)): nuevamente la idea central de este método es expresar a las curvas en una base de funciones, similar al método *Fitfclust*, pero en lugar de utilizar la base spline se usa una base wavelet. El modelo subyacente que supone este método es:

$$Y_i(t) = \mu_k(t) + U_i(t) + \epsilon_i(t),$$

donde μ_k es la media del cluster y U_i es una desviación relacionada específicamente a la curva con $U_i(t) \sim N(0, \Gamma_k(s, t))$ y $\Gamma_k(s, t)$ es la función de covarianza relacionada al cluster k . La proyección de una función f en una base wavelet se hace mediante el funcional W :

$$Wf(t) = \sum_{l=0}^{2^{j_0}-1} c_{j_0 l}^* \phi_{j_0 l}(t) + \sum_{j \geq j_0} \sum_{l=0}^{2^j-1} d_{jl}^* \psi_{jl}(t).$$

Este operador se aplica al modelo resultando en un modelo lineal con efectos mixtos con coeficientes,

$$c_i = \alpha_k + v_i + \epsilon_i$$

$$d_i = \beta_k + \theta_i + \epsilon_i$$

donde (α_k, β_k) son los coeficientes para μ_k y (v_i, θ_i) son los coeficientes aleatorios para $U_i(t)$. Se asume distribución normal para (v_i, θ_i) y al igual que *Fitfclust* se estima por máxima verosimilitud.

FSCM (Serban y Jiang (51)): el problema estadístico en el cual se basa este método es el de poder agrupar múltiples funciones observadas con error,

$$Y_{i,j} = f_{s_j}(t_{i,j}) + \sigma_\epsilon \epsilon_{ij},$$

donde s_j para $1 \leq j \leq n$ son coordenadas espaciales de un dominio espacial d -dimensional, f_{s_j} es una función aleatoria que varía con el tiempo indexada por las unidades espaciales s_j observadas en los tiempos $(t_{1,j}, \dots, t_{m,j})$ y ϵ_{ij} son errores espaciales aleatorios correlacionados. Por simplicidad para presentar el método asumiremos que los tiempos observados t_{ij} son los mismos para todas funciones.

Dado el vector de pertenencia al cluster $Z_j = (Z_{j1}, \dots, Z_{jC})$, $j = 1, \dots, n$ donde $Z_{jk} = 1$ si Y_{ij} pertenece al cluster k y 0 en caso contrario, asumimos que $Y_j|Z_j$ sigue una distribución multinomial con representación funcional

$$Y_j(t_i)|(Z_{jk} = 1) = \mu(t_i) + \mu_k(t_i) + e_{ij}$$

donde $\mu(t)$ es una media global, $\mu_k(t)$ es el efecto del cluster k y e_{ij} es un error aleatorio el cual se supone comunmente i.i.d con distribución $N(0, \sigma_\epsilon^2)$. Similarmente al ANOVA funcional, restringimos $\sum_{k=1}^C \mu_k(t) = 0$. Siguiendo la representación funcional descomponemos

$$\mu(t) = \sum_{l=1}^{L_0} \alpha_l \phi_l(t) \text{ y } \mu_k = \sum_{l=1}^{L_k} \beta_{k,l} \phi_l(t),$$

donde $\phi_l(t)$ provienen de una base de $L^2(T)$.

La estructura de correlación del error espacial e_{ij} se descompone en un efecto aleatorio $\tau(s)$ y error independientes $\delta_{ij} \sim N(0, \sigma_\epsilon^2)$. Luego el modelo funcional con errores correlacionados nos queda

$$Y_{ij}|(Z_{ik} = 1, \tau) = \mu(t_i) + \mu_k(t_i) + \tau(s_j) + \delta_{ij},$$

donde $\tau(s) = \sum_{r=1}^n \gamma_r \psi_r(s)$ es el efecto espacial con $\gamma_r \sim N(0, \sigma_s^2)$ y $\{\psi_r(s)\}_{r \geq 1}$ es una base. Finalmente, los parámetros del modelo se estiman maximizando la función de verosimilitud.

Funclust (Jacques and Preda (27)): Sea X un funcional aleatorio que toma valores en $L^2(T)$ y asumimos que X es un proceso estocástico en L^2 -continuo. Este método se basa en intentar una aproximación para densidad de X usando la expansión K-L. Sin pérdida de generalidad podemos suponer que X tiene media cero y podemos representarlo

$$X(t) = \sum_{j \geq 1} C_j \psi_j(t),$$

donde $C_j = \int_T X(t) \psi_j(t) dt$, son las componentes principales de X y ψ_j forman una base ortonormal de autofunciones del operador de covarianza de X . Observemos que las componentes principales C_j son variables aleatorias no correlacionadas con varianza λ_j . Sea $X^{(q)}$ la aproximación de X por los primeros q términos de la base de componentes principales. Suponiendo que X proviene de un proceso gaussiano las componentes principales C_j resultan variables aleatorias independientes normales con media cero y varianza λ_j . La densidad conjunta es

$$f_X^{(q)} = \prod_{j=1}^q f_{C_j}(c_j(x)) \text{ con } c_j(x) = \langle x, \psi_j \rangle_{L^2}$$

que es utilizada como una aproximación de la densidad de X . Finalmente se construye la función de verosimilitud para el problema de cluster dado, suponiendo que X tiene una distribución que proviene de una mixtura normal funcional.

FunHDDC (Bouveyron y Jacques (4)): este método también se basa en maximizar la verosimilitud. Consideremos un conjunto de curvas $\{x_1, \dots, x_n\}$ con $x_i = \{x_i(t)\}_{t \in [0, T]}$ que se quieren clusterizar en k grupos homogéneos. Asumimos que las curvas son realizaciones

independientes de un proceso estocástico X en $L^2[T]$. Suponemos que el proceso X admite la siguiente expansión en una base

$$X = \sum_{j=1}^p \gamma_j \psi_j(t),$$

donde el número p se considera conocido y fijo. Por otro lado, asumimos que existen variables $Z = (Z_1, \dots, Z_k) \in \{0, 1\}^k$ tal que z_{ir} es el valor de Z_r que indica si la curva x_i pertenece al r -ésimo grupo o no. Además, se debe asumir que existen k subespacios de $L^2, F_1[T], \dots, F_k[T]$ a los cuales pertenecen las curvas x_i dependiendo del cluster al que pertenecen. Llamemos y_i a la representación de x_i en $F_r[T]$ si $z_{ir} = 1$. También asumimos que y_i dado $z_{ir} = 1$ proviene de un proceso estocástico Y que admite una expansión en una base dependiendo del cluster, es decir,

$$Y(t)|(Z_r = 1) = \sum_{j=1}^{d_r} \alpha_{r,j} \psi_j(t).$$

Por último, se asume que Y está vinculado con X condicional a Z a través de un operador lineal Υ_r de $L^2[T]$ en $F_r[T]$,

$$X|(Z_r = 1) = \Upsilon_r Y|(Z_r = 1) + \epsilon|(Z_r = 1),$$

donde ϵ es una función que modela el error y admite una descomposición en la base $\epsilon(t) = \sum_{j=1}^p \beta_j \psi_j(t)$. Se puede observar además que Υ_r puede representarse por una matriz U_r de tamaño $p \times d_r$. El método requiere además hacer suposiciones sobre la distribución. Los coeficientes $\alpha_1, \dots, \alpha_n$ de Y se asumen condicionales a Z distribuidos de acuerdo a una normal multivariada,

$$\alpha|(Z_r = 1) \sim N(m_r, S_r),$$

con $S_r = \text{diag}(a_{k1}, \dots, a_{kd_r})$. Los coeficientes β_1, \dots, β_n relacionados al error ϵ se asumen condicionales a Z de acuerdo a

$$\beta|(Z_r = 1) \sim N(0, \Gamma_r).$$

A partir de estas suposiciones sobre la distribución podemos inferir la distribución condicional de los coeficientes de X ,

$$\gamma|\alpha, (Z_r = 1) \sim N(U_r \alpha, \gamma_r),$$

y su distribución marginal es una mixtura normal

$$p(\gamma) = \sum_{r=1} \pi_r \phi(\gamma, \mu_r, \sigma_r),$$

donde ϕ es una función de densidad Gaussiana, $\mu_r = U_r m_r$, $\Sigma_r = U_r S_r U_r^t + \Gamma_r$ y $\pi_r = P(Z_r = 1)$. Se define la matriz de $p \times p$ $Q_r = [U_r, V_r]$ que satisface $Q_r^t Q_r = Q_r Q_r^t = I_p$. Finalmente, se

asume que Γ_k es tal que $\Delta_r = Q_r^t \Sigma_r Q_r$ es diagonal con los primeros elementos de la diagonal a_{r1}, \dots, a_{rd_r} y los últimos $p - d_r$ son b_r con $a_{rj} > b_r$ para $1 \leq j \leq d_r$ y $r = 1, \dots, k$.

Los parámetros del modelo se estiman mediante un procedimiento de máxima verosimilitud. Dada los coeficientes $\gamma_1, \dots, \gamma_n$ de las curvas observadas x_1, \dots, x_n en la base, la función de log-verosimilitud tiene la siguiente forma,

$$\text{loglik}(\theta, \gamma_1, \dots, \gamma_n) = -\frac{1}{2} \sum_{r=1}^k n_r \left[\sum_{j=1}^{d_r} \left(\log(a_r) + \frac{q_{rj}^t W_k q_{rj}}{a_{rj}} \right) + \sum_{j=d_r+1}^p \left(\log(b_r) + \frac{q_{rj}^t W_k q_{rj}}{b_r} \right) - 2 \log(\pi_r) \right] + \xi,$$

donde q_{rj} es la j -ésima columna de Q_r , $W_r = \frac{1}{n_r} \sum_{i=1}^n n_{z_{ir}} (\gamma_i - \mu_r)^t (\gamma_i - \mu_r)$, $n_r = \sum_{i=1}^n z_{ir}$ y ξ es un término constante.

DivFunFD (Justel y Svarc (30)): es un método divisivo jerárquico de cluster, la idea central se basa en detectar regiones de las funciones o sus derivadas en las cuales se identifiquen más claramente la presencia de grupos, una vez detectadas estas regiones se dividen los grupos inicialmente con un criterio local, que es revisado con un criterio global para dar la conformación final de los mismos. Este esquema se repite hasta que no se encuentran regiones donde la presencia de grupos sea evidente. Sean x_1, \dots, x_n un conjunto de funciones suaves definidas en el intervalo cerrado T . El método DivClusFD comienza buscando si las funciones observadas (o sus derivadas) difieren en alguna subregión de T y se puede proporcionar un agrupamiento local en tales subregiones; en particular, buscamos la subregión donde el número de los grupos inducidos (nivel o forma) se maximizan. Si los hallazgos indican un solo grupo en todas las subregiones de las funciones y sus derivadas, la búsqueda termina. De lo contrario, existen al menos g clusters, donde g es el mayor número de clusters encontrados. La partición $\{C_1, \dots, C_g\}$ de $\{1, \dots, n\}$ identifica los índices de las funciones que pertenecen al mismo cluster. Después de asignar el funciones a los clusters, el método DivClusFD busca divisiones dentro de cada grupo. El mismo procedimiento se repite en cada paso de división. Este paso es repetido sobre cada nuevo cluster hasta que no haya evidencia de una división de cluster adicional. Para realizar estas particiones considera en una misma grilla de puntos en T , donde se evalúan las funciones y sus derivadas. Para cada instante fijo de la grilla y para la funciones o cada una de sus derivadas, se tiene una muestra unidimensional de n observaciones donde mediante el gap statistics (54) se identifica el punto en el dominio T donde la función o una de sus derivadas presenta mayor evidencia de cluster y se realiza mediante k -medias unidimensional la partición inicial. Esta partición inicial se realiza siguiendo únicamente un criterio local, por este motivo es revisada en una segunda etapa con un criterio global que considera la información que tienen las curvas completas. Para ello se considera cada uno de los grupos determinados por $\{C_1, \dots, C_g\}$, y en cada uno de ellos se construye el boxplot funcional (ver (49),) que permite identificar posibles outliers (si la partición se realizó con información correspondiente a las funciones entonces el boxplot funcional se realiza utilizando las funciones, mientras que si la partición se realizó utilizando una

derivada el boxplot funcional se analiza en esa derivada). El potencial outlier es cambiado de grupo si su trayectoria se atraviesa el boxplot funcional de otro grupo durante más tiempo del que atraviesa el propio, sino se deja en el mismo grupo.

A diferencia de los otros procedimientos explicados, este procedimiento además de asignar las observaciones a cluster estima el número de grupos que hay en una muestra.

A continuación procedemos con el estudio de Montecarlo. Es importante aclarar que todos los procedimientos de cluster considerados dependen de parámetros que hay que fijar para llevar a cabo el análisis. Por ejemplo, en los métodos que se basan en expresar las curvas en una base funciones, que se trunca adecuadamente y luego se realiza un procedimiento de cluster multivariados con los coeficientes de la base es determinante fijar el donde truncar la base. Fijar esos parámetros no es tarea sencilla, sin embargo, el paquete de R *funcy* (ver Yassouridis y Leisch (56),) que tiene implementados todos los procedimientos de cluster menos DivFunFD, ajusta estos parámetros automáticamente siguiendo las recomendaciones dadas por los diferentes autores, nosotros seguimos las recomendaciones dadas en el mismo. En el caso de DivFunFD, seguimos las recomendaciones dadas por Justel y Svarc ((30)).

La Tabla 3.3.3 reporta el promedio del CCR para cada modelo y procedimiento de clusters, hemos corrido 200 replicas para cada uno.

Para DivClusFD sólo reportamos el CCR promedio cuando el número de clusters es estimado correctamente, lo cual sucede en todos los casos más del 75 % de las réplicas.

Los Modelos A and B son fáciles de clasificar, de hecho todos los procedimientos lo hacen casi a la perfección. La clasificación del Modelo C es una tarea un poco más desafiante, lo cual se ve en los resultados de waveclust. En general los métodos con peor desempeño son Funclust y funHDDC. Es notable como el CCR de waveclust disminuye a medida que la dificultad presentada por los modelos aumenta. EL Modelo D sigue un patrón totalmente diferente que los otros modelos, en este caso, los únicos que presentan un desempeño sobresaliente son LDCI y DivClusFD.

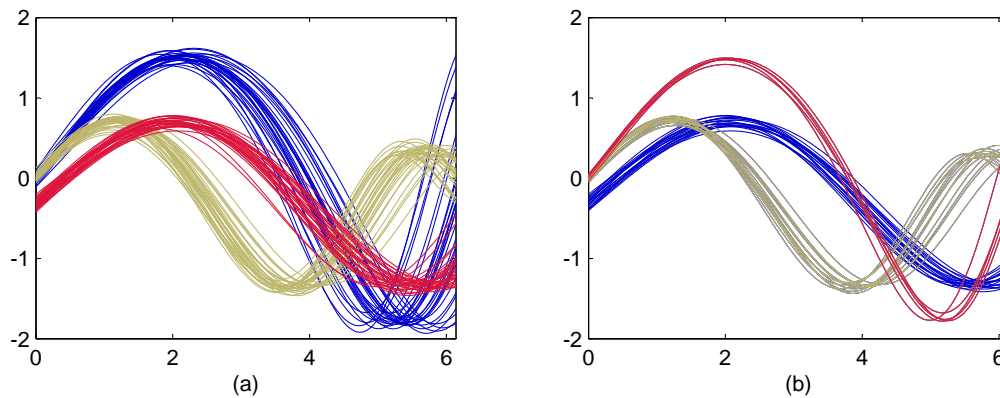
En la Figura 3.3 se pueden ver las observaciones núcleo para un conjunto de datos generado siguiendo el Modelo A. Si bien el diseño de los clusters es balanceado es claro que el número de observaciones del núcleo de cada cluster no es el mismo.

De la Tabla 3.3.3 se puede ver claramente que el desempeño de LDCI y LDCH de es diferente, LDCH no siempre consigue detectar el centro de los clusters. A modo de ejemplo, generamos una muestra siguiendo la distribución del Modelo C, ver Figura 3.4, donde el LDCH falla. La paleta de colores empleada en la Figura 3.4 panels (a) (respectivamente (b)) es descendente en respecto de IDLD (respectivamente LDH), es claro que todo grupo tiene curvas con alto valor de IDLD, sin embargo no todo grupo tiene curvas con alto valor de LDH. La Figura 3.4 (c) presenta el gráfico de dispersión entre IDLD y LDH, donde se puede ver que prácticamente no hay estructura, más aún el coeficiente de correlación entre LDH y IDLD en este caso particular es bajo, $\rho = 0,27$. Esto explica la diferencia entre el desempeño de LDC al modificar la profundidad local considerada.

Por último en la Figura 3.5(a) mostramos un conjunto de datos generado siguiendo el

	Modelo A	Modelo B	Modelo C	Modelo D
LDCI	99,98	99,94	99,07	99,39
LDCH	93,11	100	66,21	41,51
DivClusFD	99,67	99,96	99,45	99,45
fitfclust	99,99	99,93	98,95	74,65
distclust	99,79	100	99,82	74,61
iterSubspace	99,80	100	98,89	73,92
funclust	81,55	79,26	61,64	38,97
funHDDC	87,88	95,99	72,02	51,15
fscm	97,79	99,79	99,64	74,48
waveclust	99,88	95,94	89,18	72,98

Cuadro 3.3: CCR promedio para los distintos procedimientos de cluster considerados.

Figura 3.3: (a) Datos simulados según el *Modelo A*. (b) Núcleo correspondiente a los datos en (a)

Modelo D, cada cluster está representado por un color distinto. La Figura 3.5(b) muestra las observaciones núcleo para el mismo conjunto de datos, se respeta el código de colores. En este modelo los grupos no se distinguen por simple inspección ocular.

3.3.4. Simulaciones Datos Funcionales Multidimensionales

En esta sección consideramos el análisis de datos funcionales multidimensionales, es decir que los datos se pueden representar como vectores donde cada coordenada es un dato

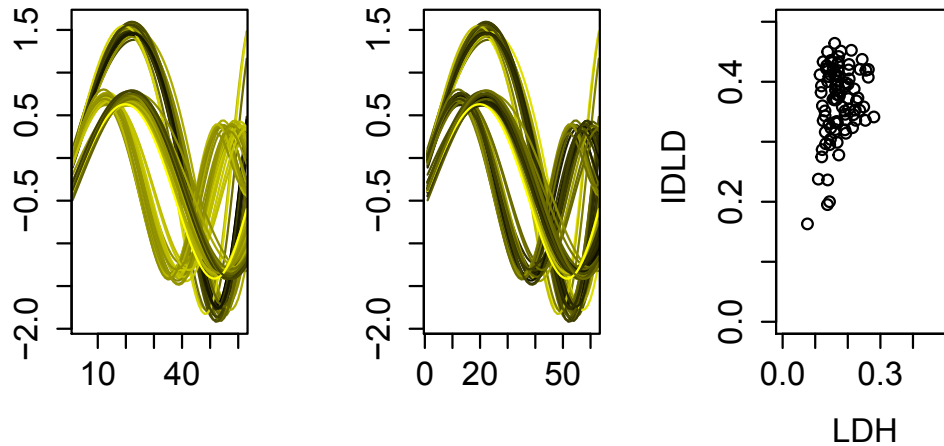


Figura 3.4: (a) y (b) Un conjunto de datos simulado siguiendo el Modelo C, el color de las curvas está dado por una paleta descendente en el valor de profundidad local de cada curva para IDLD y LDH. (c) Scatter plot IDLD versus LDH.

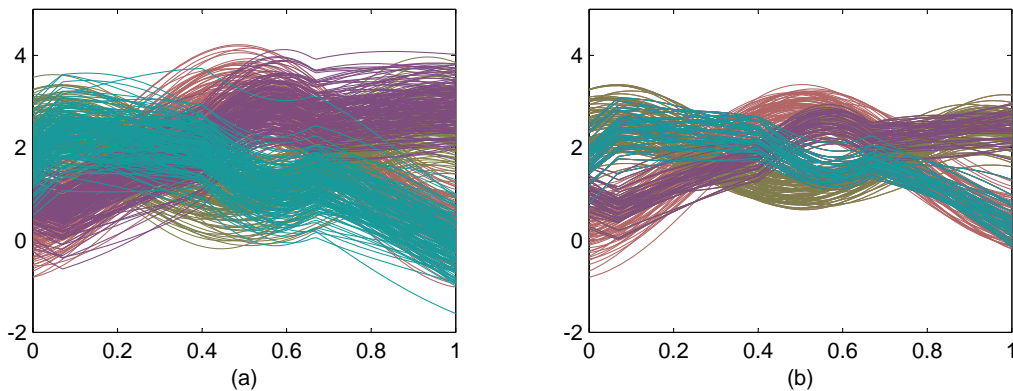


Figura 3.5: (a) Un conjunto de datos generados siguiendo el Modelo D (b) Observaciones núcleo para el mismo conjunto de datos.

funcional. El modelo estadístico subyacente, en este contexto, es representado por un proceso estocástico continuo, $X = \{X_t : t \in T\}$, donde T es un intervalo acotado. La mayoría de los enfoques para datos funcionales consideran el caso univariado, es decir cuando $X_t \in \mathbb{R}$. Si bien el análisis del caso multivariado, $X_t = (X_1(t), \dots, X_p(t)) \in \mathbb{R}^p$, no es el que predomina en la literatura en los últimos tiempos su estudio ha sido considerado en mayor profundidad. Un ejemplo de datos funcionales bivariados y el contexto de aplicación se encuentra en Ramsay y Silverman (47). Los autores utilizan el proceso $X_t = (X_1(t), X_2(t)) \in \mathbb{R}^2$ para modelizar me-

didadas de rodilla y cadera respectivamente, en el contexto de componentes principales, como una extensión del caso univariado.

Entre los pocos métodos de cluster que se pueden hallar en la literatura encontramos la propuesta de Ieva et al. (26), que utilizan un algoritmo del tipo k -medias basado en distancias específicas entre los datos funcionales multivariados, más específicamente los autores utilizan como d_1 la distancia natural del espacio $H^1[T, \mathbb{R}^8]$ y como d_2 la distancia en $L^2[T, \mathbb{R}^8]$ que involucra la primer derivada de cada función en la coordenada correspondiente. Utilizando estas dos distancias Jacques y Preda (28) extienden el método Funclust cuya idea se centra en considerar una aproximación de la densidad para datos funcionales basándose en calcular las componentes principales para el caso de datos funcionales multivariados. Schmutz et al. (50) proponen un procedimiento análogo al método funHDDC basados en el cálculo de las componentes principales que hacen Jacques y Preda (28).

Nuestra propuesta también se aplica al caso multifuncional gracias a que la IDLD está definida en espacios de Banach separables los cuales comprenden este contexto. Particularmente, para la simulación hemos supuesto que los datos provienen de $H = L^2[0, 1] \times L^2[0, 1]$ que es un espacio de Hilbert reflexivo, es decir, el espacio dual es él mismo. Hemos generado una medida independiente de las distribución de los datos $Q = Q_1 \times Q_2$ donde Q_1 y Q_2 son dos procesos Gaussianos independientes que toman valores en $L^2[0, 1]$. De ahí en forma inmediata se puede calcular la IDLD. Las distancia involucrada para el LDCI es la usual dada por el producto. La elección de los parámetros α y β se la hicimos del mismo modo que en la Sección 3.3.2 y tomamos los mismos valores.

Replicamos la simulación hecha por Schmutz et al. (50) para el caso de datos funcionales multivariados. Ellos presentan tres escenarios diferentes, en cada caso los datos son datos funcionales bivariados.

Modelo E. Tres grupos, cada uno con 100 observaciones. donde $a_1 \sim N(0, 0.2)$, $a_2 \sim$

$$\begin{aligned} \text{Group 1: } & X_1(t) = \sin((10 + a_1)t) + (1 + a_1) + e_1(t) \\ & X_2(t) = \sin((5 + a_2)t) + (0.5 + a_2) + e_2(t) \\ \text{Group 2: } & X_1(t) = \sin((5 + a_2)t) + (0.5 + a_2) + e_2(t) \\ & X_2(t) = \sin((15 + a_1)t) + (1 + a_1) + e_1(t) \\ \text{Group 3: } & X_1(t) = \sin((15 + a_1)t) + (1 + a_1) + e_1(t) \\ & X_2(t) = \sin((10 + a_1)t) + (1 + a_1) + e_1(t), \end{aligned}$$

$N(0, 0.3)$, $e_1(t)$ es el ruido blanco con varianza $|\frac{a_1}{2}|$, y $e_2(t)$ también es ruido blanco con varianza $|\frac{a_2}{2}|$. Las curvas son generadas en 101 puntos equidistantes en el intervalo $[0, 1]$. Cada coordenada se encuentra representada en la Figura 3.6. Una animación de las curvas se puede encontrar en <https://rawgit.com/lucasferpiana/ModelA/master/ModelA.html>.

Modelo F. Cuatro grupos, cada uno de ellos con 250 observaciones. donde $t \in [1, 21]$, $U \sim U(0, 0.1)$, $e_1(t)$ es ruido blanco independiente de U con varianza 0.25. Las funciones son $h_1(t) = (6 - |t - 7|)_+$ y $h_2(t) = (6 - |t - 15|)_+$ donde $(\cdot)_+$ representa parte positiva. Las curvas son generadas en 101 puntos equidistantes en el intervalo $[1, 21]$. Cada coordenada se

$$\begin{aligned}
\text{Group 1: } & X_1(t) = U + (1 - U)h_1(t) + e(t), \\
& X_2(t) = U + (0,5 - U)h_1(t) + e(t), \\
\text{Group 2: } & X_1(t) = U + (1 - U)h_2(t) + e(t), \\
& X_2(t) = U + (0,5 - U)h_2(t) + e(t), \\
\text{Group 3: } & X_1(t) = U + (0,5 - U)h_1(t) + e(t), \\
& X_2(t) = U + (1 - U)h_2(t) + e(t), \\
\text{Group 4: } & X_1(t) = U + (0,5 - U)h_2(t) + e(t), \\
& X_2(t) = U + (1 - U)h_1(t) + e(t),
\end{aligned}$$

encuentra representada en la Figura 3.7. Una animación de las curvas se puede encontrar en <https://rawgit.com/lucasferpiana/ModelB/master/ModelB.html>

Modelo G. Cuatro grupos, cada uno de ellos con 250 observaciones. donde $t \in [1, 21]$,

$$\begin{aligned}
\text{Group 1: } & X_1(t) = U + (1 - U)h_1(t) + e(t), \\
& X_2(t) = U + (0,5 - U)h_1(t) + e(t), \\
\text{Group 2: } & X_1(t) = U + (1 - U)h_2(t) + e(t), \\
& X_2(t) = U + (0,5 - U)h_2(t) + e(t), \\
\text{Group 3: } & X_1(t) = U + (1 - U)h_1(t) + e(t), \\
& X_2(t) = U + (1 - U)h_1(t) + e(t), \\
\text{Group 4: } & X_1(t) = U + (0,5 - U)h_2(t) + e(t), \\
& X_2(t) = U + (0,5 - U)h_1(t) + e(t),
\end{aligned}$$

$U, e(t), h_1$ y h_2 son definidos como en el modelo anterior. Las curvas son generadas en 101 puntos equidistantes en el intervalo $[1, 21]$. Cada coordenada se encuentra representada en la Figura 3.8. Una animación de las curvas se puede encontrar en <https://rawgit.com/lucasferpiana/ModelC/master/ModelC.html>

A diferencia del caso de datos funcionales unidimensionales no contamos con acceso a los procedimientos de cluster que se encuentran en la literatura. Por este motivo, informamos los resultados que presentan Schmutz et al. (50). En el mismo el desempeño de los diferentes procedimientos de cluster se mide mediante el ARI entre la partición teórica y la obtenida por cada método. Schmutz et al. (50), informan el ARI para cuatro propuestas propias, para *funclust* (28), *kmeans - d₁* and *kmeans - d₂*, que son dos propuestas de Ieva et al (26). Por este motivo, nosotros también calculamos el ARI para nuestra propuesta utilizando la función *AdjustedRandIndex* del paquete de R *mclust*. Mantenemos el mismo número de réplicas que en el trabajo original, $M = 50$.

En la Tabla 3.3.4 informamos el máximo valor del ARI para Schmutz et al y el resto de los métodos. Es claro que LDCI tiene la mejor performance por lejos dado que no clasifica mal ninguna de las observaciones en el estudio de simulación.

	Modelo E	Modelo F	Modelo G
LDCI	1	1	1
Schmutz (mejor)	0,96	0,92	0,80
funclust	0,23	0,36	0,45
<i>kmeans</i> – d_1	0,90	0,37	0,32
<i>kmeans</i> – d_2	0,90	0,37	0,32

Cuadro 3.4: ARI para los distintos procedimientos de cluster para datos funcionales multivariados.

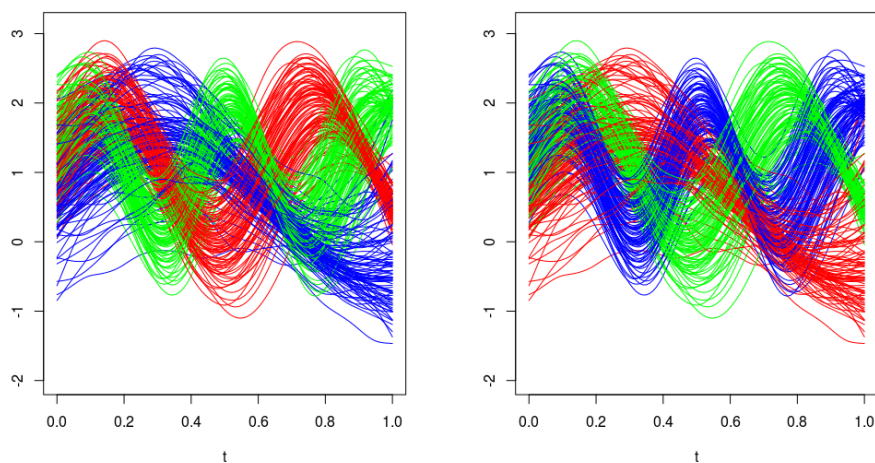


Figura 3.6: Modelo E. Rojo: Grupo 1. Azul: Grupo 2. Verde: Grupo 3.

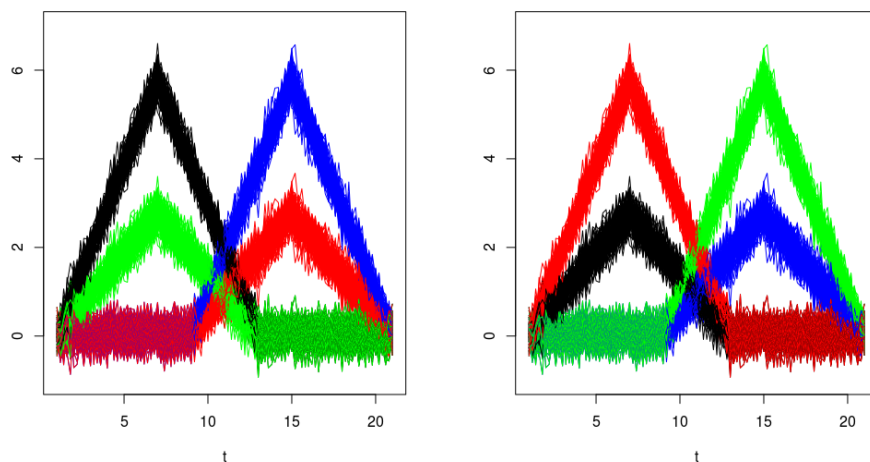


Figura 3.7: Modelo F. Rojo: Grupo 1. Azul: Grupo 2. Verde: Grupo 3. Negro: Grupo 4.

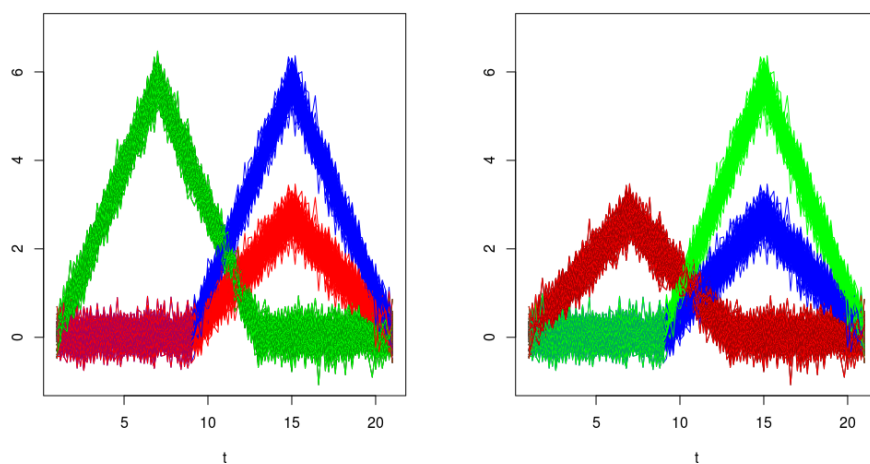


Figura 3.8: Modelo G. Negro: Grupo 1. Azul: Grupo 2. Verde: Grupo 3. Rojo: Grupo 4.

3.3.5. Ejemplos: Datos Reales Funcionales

En esta sección analizamos la performance de nuestro método con varios conjuntos de datos funcionales. Únicamente consideramos la profundidad local IDLD ya que presentó mejor desempeño en el ejercicio de simulación. Estos conjuntos son ampliamente citados en la literatura y están disponibles en internet. Como benchmark utilizamos los procedimientos de

clusters de la librería *funcit* de R. Fueron considerados los siguientes datasets: Growth, Canadian Weather, ECG200 y Tecator. Estos ejemplos provienen del problema de clasificación por lo tanto la configuración de los clusters es conocida. Reportamos el CCR como en los estudios de simulación. Los parámetros para los distintos procedimientos de clustering permanecen fijos e iguales que en las simulaciones. Los ejemplos analizados son considerados un reto para cualquier propuesta de cluster.

El estudio de crecimiento de la universidad de Berkeley (Growth) es conocido como una de las investigaciones de desarrollo humano a largo plazo más completos que hay. Fue conducida por Tuddenham y Snyder (52). Se midieron, desde el primer año de vida hasta los dieciocho años de edad, en intervalos de tiempo irregulares, la altura de 54 niñas y 39 niños. Las mediciones se hicieron con mayor frecuencia durante la adolescencia y los últimos años de la niñez cuando el crecimiento es más rápido. La frecuencia fue menor en los primeros años cuando el crecimiento es más estable. Este conjunto de datos se encuentra en el paquete R *fd*.

El conjunto de datos Canadian Weather fue introducido por Ramsay y Silverman (47) y también está disponible en el paquete de R *fd*. Los datos contienen la temperatura media diaria medida en el curso de un año monitoreada por 35 estaciones climáticas canadienses. Los datos están agrupados en cuatro regiones geográficas.

El conjunto de datos ECG200 consiste de 200 electrocardiogramas que pueden ser encontrados en el sitio web “UCR Time Series Classification and Clustering” (Chen et al. (9)). En estos datos encontramos dos grupos distintos. Uno de ellos con 133 electrocardiogramas y en el otro 67, cada uno de ellos registrados en 96 puntos equidistantes.

El conjunto de datos Tecator consiste de 215 curvas espectrométricas obtenidas de muestras de carne, las variables que las describen son la grasa, el agua y las proteínas obtenidas con procedimientos analíticos. Las curvas se clasifican en dos grupos, uno de ellos con gran contenido de grasa (mayor 15) y el otro con bajo contenido (debajo de 15), estos parámetros fueron fijados en la literatura. Nuestro objetivo es agrupar los datos en estos dos grupos basados en las curvas espectrométricas.

Cuando analizamos los resultados presentados, en la Tabla 3.3.5, se puede observar que ningún procedimiento se destaca del resto en todos los casos. Lo que nos lleva a concluir, como es esperable, que el método más apropiado dependerá de las características de los datos. Sin embargo, se puede observar que LDCI tiene una muy buena performance en todos los ejemplos analizados, siendo el que mejor clasifica las observaciones para los conjuntos Growth y ECG200, y el segundo mejor clasificador para Tecator.

Si bien no está reportado en la Tabla 3.3.5, también estudiamos la presencia de cluster con el criterio DivFunFD. Para los conjuntos de datos Canadian Weather, ECG200 y Tecator, no pudo identificar el número correcto de grupos, conformando clusters espúreos. Mientras que para el conjunto de datos Growth la tasa de observaciones bien clasificadas es 89,25 %

	Growth	Canadian Weather	ECG200	Tecator
LDCI	91,40	62,86	76,00	85,12
fitfclust	60,22	60,00	65,00	66,98
distclust	66,67	68,57	64,00	66,05
iterSubspace	56,99	64,28	64,00	73,95
funclust	77,42	60,00	52,00	59,07
funHDDC	90,32	71,42	70,00	50,23
fscm	65,59	34,29	67,00	76,28
waveclust	78,95	57,14	75,00	90,7

Cuadro 3.5: CCR para los distintos conjuntos de datos considerados agrupados siguiendo diferentes métodos de cluster.

3.3.6. Ejemplo: Datos Reales de Estructura Mixta

Consideramos el conjunto de datos EAMET, que describimos en la Sección 3.2.1, buscaremos analizar la estructura de cluster de los mismos mediante el procedimiento LDC con profundidad local MIDLD. Al no encontrar, en la literatura otro procedimiento de cluster específico para datos mixtos, nuestro análisis se limita a interpretar los resultados que obtenemos.

Antes de realizar el análisis de cluster es necesario fijar los parámetros del procedimiento. El conjunto de datos analizado tiene una componente funcional y otra multivariada, es decir que los datos pertenecen al espacio $\mathbb{E} = C[0, 365] \times \mathbb{R}^3$. En primer lugar determinamos los pesos de la MIDLD, dándole igual valor a cada componente del conjuntos de datos, i.e. $\delta_1 = \delta_2 = 1/2$, ya que no tenemos ninguna información adicional para hacerlo de otro modo. En cuanto al procedimiento de cluster propiamente dicho, tenemos que fijar los parámetros α y β . Los datos funcionales presentan una alta variabilidad entre instantes próximos de tiempo, esta información es importante a la hora de realizar el análisis y no debemos perderla, pero dificulta la visualización. Consideramos $\alpha = 0,15$ fijo a lo largo del problema, de este modo se favorece la visualización de los datos de las regiones centrales, ya que son pocos.

En la Figura 3.9 se muestran las regiones de profundidad local $R_{0,15}^\beta$ para $\beta = 0,9, 0,5, 0,3$, en este caso las curvas graficadas fueron previamente suavizadas mediante splines-cúbicos, esta técnica se encuentra disponible en el paquete base de R. Es importante remarcar, que el suavizado se realizó únicamente para permitir una visualización más clara, pero que a lo largo del análisis se trabajó con los datos funcionales crudos.

Puede observarse que para $\beta = 0,9$, donde la profundidad local toma valores más próximos a la profundidad global, la disposición de las curvas en la región es la más uniforme de las tres, y no hay una estructura de cluster aparente. Para $\beta = 0,5$ (figura central), las curvas

tienden a agruparse al comienzo de las series, pero luego este efecto se desvanece, manteniendo un patrón más similar al que tenían las curvas con $\beta = 0,9$, pero menos marcado. Esto nos lleva a pensar que $\beta = 0,9$ está detectando un centro espúreo. Es decir, un centro entre dos centros, un problema que hereda de la IDLD. Por último analizamos $\beta = 0,3$, la distribución espacial de las curvas es más espaciada, concentrándose este fenómeno sobre dos o tres niveles, especialmente en el comienzo de las series. Finalmente, optamos por tomar $\beta = 0,3$ dado que nos permite separar el núcleo en dos grupos y aplicar luego el método de cluster, LDC. Es importante destacar, que tampoco es esperable observar claramente la estructura de cluster en los datos multivariados, porque las profundidades locales consideran la información multivariada, además de la información funcional.

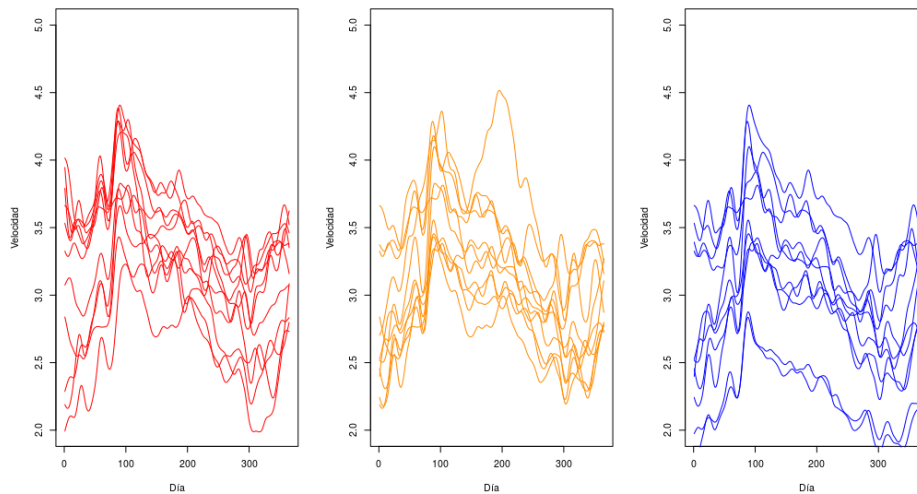


Figura 3.9: Regiones núcleo para las series diarias anuales para la temperatura media. En el panel izquierdo en rojo $R_{0,15}^{0,9}$, en el panel central en naranja $R_{0,15}^{0,5}$, y por último a la derecha en azul $R_{0,15}^{0,3}$.

Buscamos dos clusters aplicando el método LDC basado en la $MIDLD^{0,3}$ apoyándonos en el análisis anterior. En la Figura 3.10 se muestra la disposición geográfica de las estaciones meteorológicas indicando en diferentes colores las estaciones correspondientes a cada grupo. Claramente uno de los grupos representa a las estaciones que se encuentran en zonas costeras (marcadas en naranja) mientras que el otro grupo corresponde a las estaciones meteorológicas de regiones continentales, más alejadas del océano o mar. Esta clasificación se corresponde con el conocido hecho de que la velocidad del viento presenta menor variabilidad sobre las zonas costeras. Un ejemplo puede encontrarse en el aprovechamiento que se hace de las granjas eólicas. Las Figuras 3.12 y 3.11, muestran esta caracterización con más detalle, reforzando la idea de que el mismo efecto continental se da tanto tanto en la región peninsular e insular Mediterránea como en las Islas Baleares. Para visualizar mejor el mapa

usar el siguiente link <http://rpubs.com/fplucas/373813>.

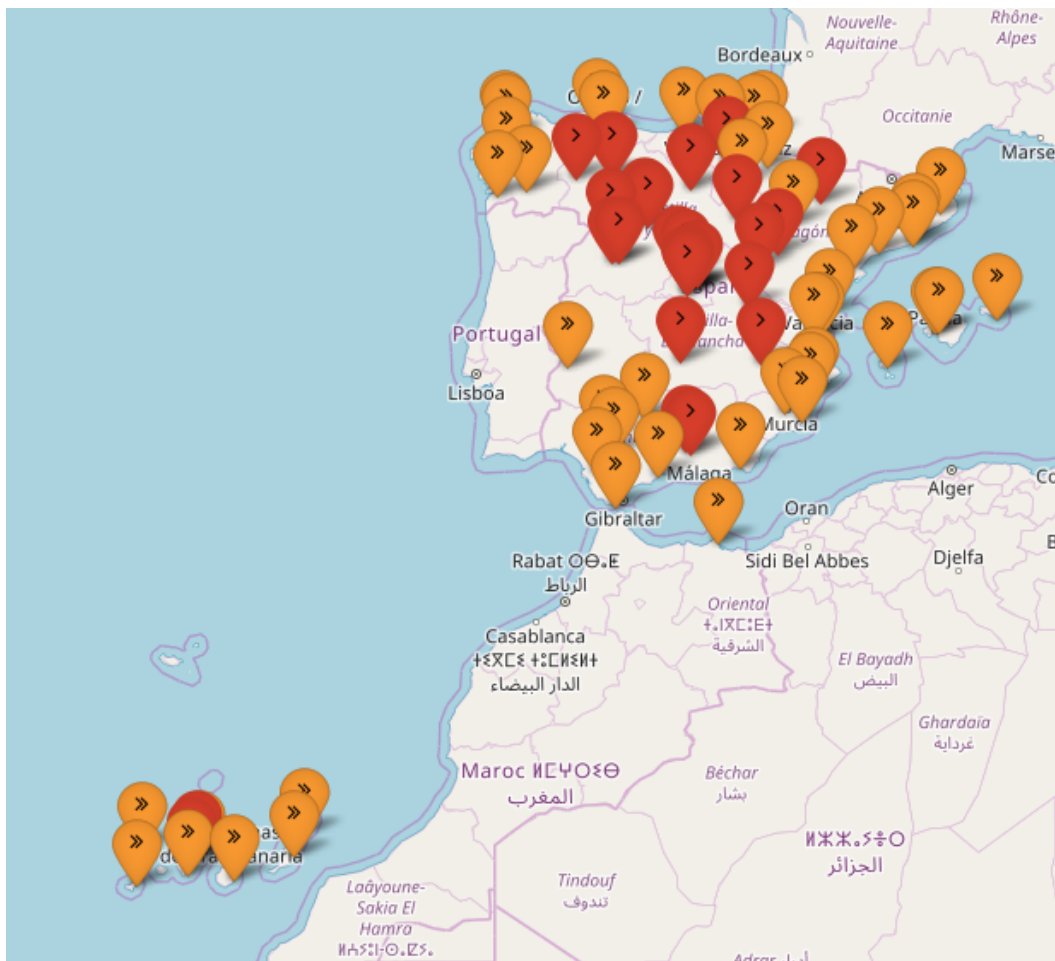


Figura 3.10: Posición geográfica de cada estación meteorológica. Las estaciones correspondientes al grupo costero están señaladas en naranja, mientras que las estaciones meteorológicas correspondientes al grupo continental están marcadas en rojo.

Por último, para comprender la conformación de los grupos en forma integral, es conveniente analizar las regiones centrales para la velocidad media de los vientos y de la altura. Se puede ver que las estaciones correspondientes a la región central continental se encuentran a mayor altura, sufren más variabilidad en la intensidad del viento, como muestran los paneles izquierdo y derecho de la Figura 3.13. Sin embargo, las estaciones costeras que están en zonas más bajas tienen menor variabilidad diaria y aparentemente el viento tiene mayor intensidad, como se puede ver en los paneles central y derecho de la Figura 3.13.

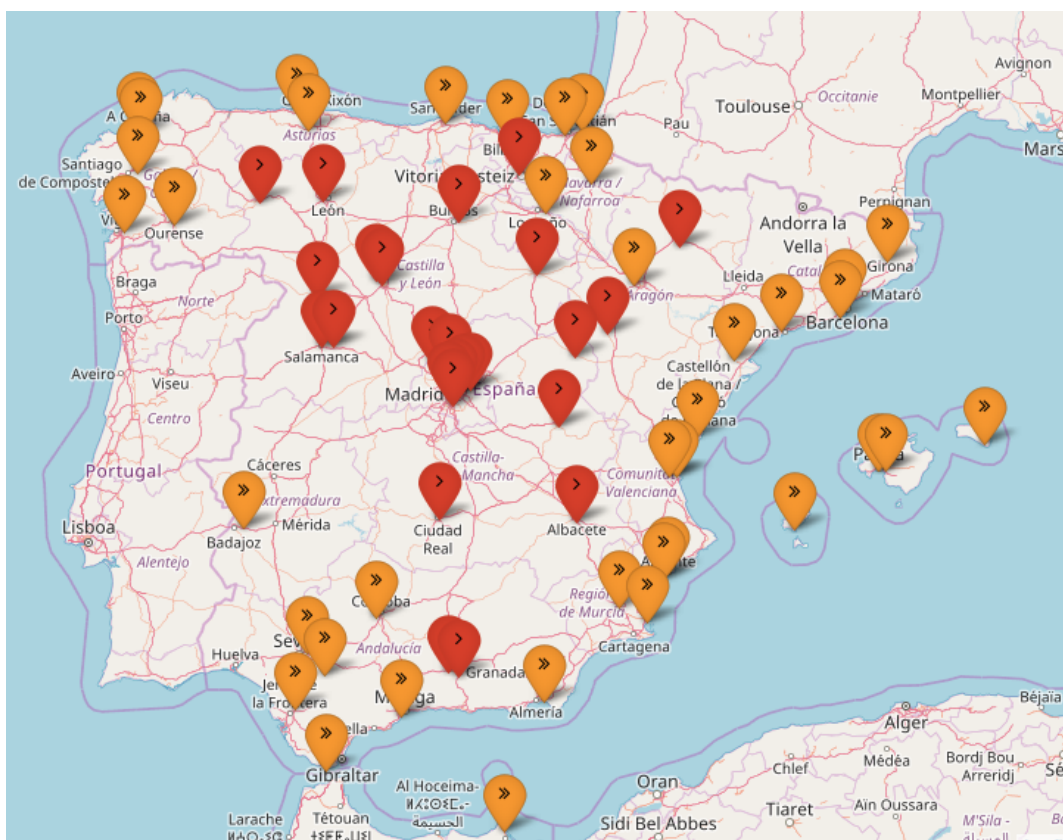


Figura 3.11: Estaciones meteorológicas para España peninsular e insular Mediterránea. En naranja están indicadas las estaciones del grupo costero y en rojo las del grupo continental.

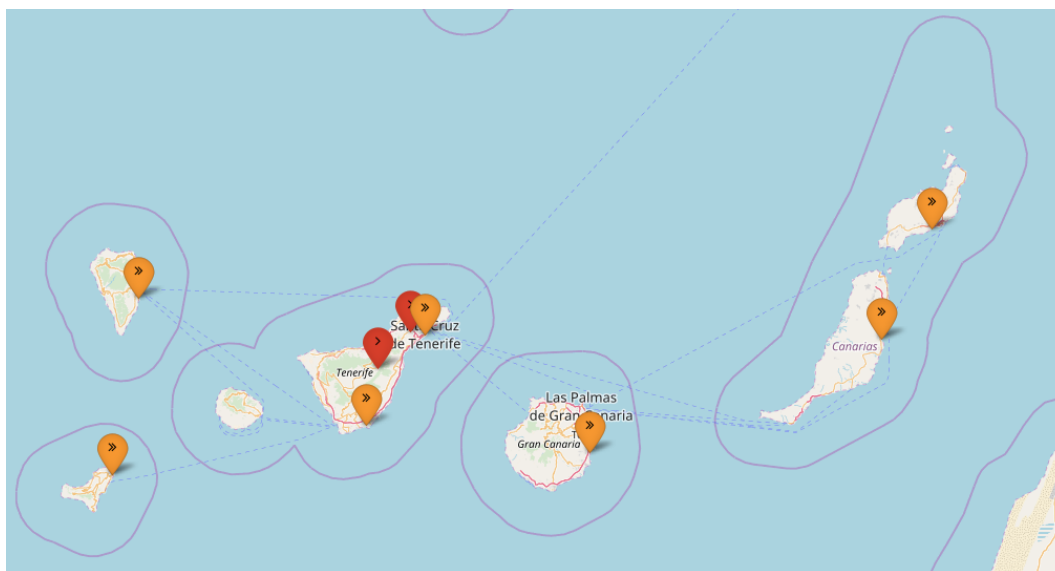


Figura 3.12: Estaciones meteorológicas para regiones insulares sobre el Océano Atlántico. En naranja están indicadas las estaciones del grupo costero y en rojo las del grupo continental.

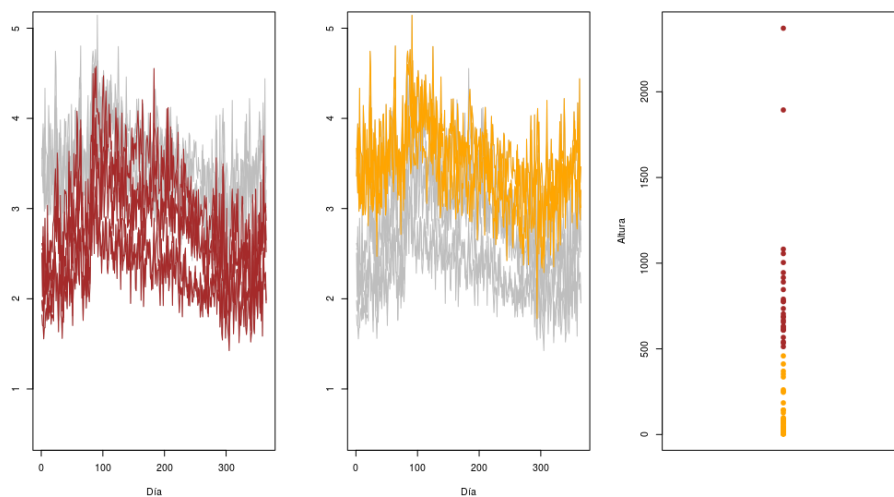


Figura 3.13: Izq: Núcleo del primer grupo elegido por el LCD para la velocidad del viento. Centro: Núcleo del segundo grupo elegido por el LCD para la velocidad del viento. Derecha: Representación de los clusters en la altura.

Bibliografía

- [1] AGOSTINELLI, C. “Local half-region depth for functional data”. Journal of Multivariate Analysis (2018), 163, 67–79.
- [2] AGOSTINELLI, C. Y ROMANAZZI, M. “Local depth”. Journal of Statistical Planning and Inference (2011), 141, 817–830.
- [3] BILINGSLEY, P. “Convergence of Probability Measures”. Jhon Wiley and Sons, New York (1968).
- [4] BOUVEYRON, C. Y JACQUES, J. “Model-based clustering of time series in group-specific functional subspaces”. Advances in Data Analysis and Classification (2011), 5 (4), 281–300.
- [5] CHAKRABORTY, B. Y CHAUDHURI, P. “On transformation and retransformation technique for constructing an affine equivariant multivariate median”. Proceedings of the American Mathematical Society (1996), 8, 2539–2547.
- [6] CHAKRABORTY, B. Y CHAUDHURI, P. “Operating transformation retransformation on spatial median angle test”. Statistica Sinica (1998), 8, 767–784.
- [7] CHAKRABORTY, B. Y CHAUDHURI, P. “On data depth in infinite dimensional spaces”. Annals of the Institute of Statistical Mathematics (2014), 66, 303–324.
- [8] CHEN, Y., DANG, X., PENG, H., Y BART, H., L., J. “Outlier detection with the kernelized spatial depth function”. IEEE Transactions in Pattern Analysis and Machine Intelligence (2009), 31, 288–305.
- [9] CHEN, Y., KEOGH, E., HU, B., BEGUM, N., BAGNALL, A., MUEEN, A. Y BATISTA, G. “The ucr time series classification archive”. http://www.cs.ucr.edu/~eamonn/time_series_data/ (2015).
- [10] CHIOU, J. Y LI, P. “Functional clustering and identifying substructures of longitudinal data”. Journal of the Royal Statistical Society: Series B (2007), 69 (4), 679–699.

- [11] CUESTA-ALBERTOS, J.A., FEBRERO-BANDE, M. Y OVIEDO DE LA FUENTE, M. “The DD^G -classifier in the functional setting”. Test (2017), 26 (1), 119–142.
- [12] CUESTA-ALBERTOS, J. A. Y NIETO-REYES, A. “The random tukey depth”. Computational Statistics and Data Analysis (2008), 52, 4979–4988.
- [13] CUEVAS, A. FEBRERO, M. Y FRAIMAN, R. “Robust estimation and classification for functional data via projection-based depth notions”. Computational Statistics (2007), 22, 481–496.
- [14] CUEVAS, A. Y FRAIMAN, R. “On depth measure and dual statistics. A methodology for dealing with general data”. Journal of Multivariate Analysis (2009), 100, 753–766.
- [15] DUTTA, S. Y GHOSH, A, K. “On robust classification using projection depth”. Annals of the Institute of Statistical Mathematics (2012), 64, 657–676.
- [16] DUTTA, S., GHOSH, A, K. Y CHAUDHURI, P. “Some intriguing properties of Tukeys half-space depth. ”. Bernoulli (2011), 17, 1420–1434.
- [17] DHARMADHIKARI, S. Y JOAG-DEV, K. “Unimodality, Convexity and Applications”. Academic Press, Nueva York (1988).
- [18] FRAIMAN, R. Y MUNIZ, G. “Trimmed means for functional data”. Test (2001), 10, 419–440.
- [19] FRALEY, C. Y RAFTERY, A. “Model-based clustering, discriminant analysis, and density estimation”. Journal of the American Statistical Association (2002), 97, 611–631.
- [20] FRALEY, C. Y RAFTERY, A. “MCLUST Version 3 for r: Normal mixture modeling and model-based clustering”. Technical Report No. 504, Department of Statistics, University of Washington (2009).
- [21] HENNING, C. “Cluster-wise assesment of cluster stability”. Computational Statistics and Data Analysis (2007), 52, 258–271.
- [22] GIACOFI, M., LAMBERT-LACROIX, S., MAROT, G. Y PICARD, F. “Wavelet-based clustering for mixed-effects functional models in high dimension”. Biometrics (2011), 69 (1), 31–40.
- [23] GHOSH, A. K. Y CHAUDHURI, P. “On maximum depth and related classifiers”. Scandinavian Journal of Statistics (2005), 32 (2), 327350.
- [24] GIJBELS, I. Y NAGY, S. “On a general definition of depth for functional data”. Statistical Science (2017), 32 , 630639.

- [25] HUBERT, L. Y ARABIE, P. “Comparing partitions”. Journal of Classification (1985), 2 (1), 193–218.
- [26] IEVA, F., PAGANONI, A. M., PIGOLI, D. Y. VITELLI, V. “Multivariate functional clustering for the morphological analysis of electrocardiograph curves”. Journal of the Royal Statistical Society: Series C (Applied Statistics) (2013), 62, 401–418.
- [27] JACQUES, J. Y PREDÀ, C. “Funclust: a curves clustering method using functional random variables density approximation”. Neurocomputing (2013), 112, 164–171.
- [28] JACQUES, J. Y PREDÀ, C. “Model-based clustering of functional data”. Computational Statistics and Data Analysis (2014), 71, 92–106.
- [29] JAMES, GM. Y SUGAR, C. “Clustering for sparsely sampled functional data”. Journal of the American Statistical Association (2003), 98, 397–408.
- [30] JUSTEL, A. Y SVARC, M. “A divisive clustering method for functional data with special consideration of outliers”. Advances in Data Analysis and Classification (2017), DOI: <https://doi.org/10.1007/s11634-017-0290-1>.
- [31] KONDO, Y., SALIBIAN-BARRERA, M. Y ZAMAR, R. H. “A robust and sparse k-means clustering algorithm”. Journal of Statistical Software (2016), 72 (5), 1–26.
- [32] KOTÍK, L. Y HLUBINKA, D. “A weighted localization of halfspace depth and its properties”. Journal of Multivariate Analysis (2017), 157, 53–69.
- [33] KUELBS, J. Y ZINN, J. “Concerns with functional depth”. ALEA (2013), 10, 831–855.
- [34] LI, J., CUESTA-ALBERTOS, J. A., Y LIU, R. Y. “DDclassifier: Nonparametric classification procedure based on DDplot”. Journal of the American Statistical Association (2012), 107 (498), 737–753.
- [35] LIU, R. Y. “On a notion of data depth based on random simplices”. The Annals of Statistics (1990), 18 (1), 405–414.
- [36] LI, J. Y LIU, R. Y. “New nonparametric tests of multivariate locations and scales using data depth”. Statistical Science (2004), 19 (4), 686–696.
- [37] LIU, R. Y., PARELIUS, J. M. Y SINGH, K. “Multivariate analysis by data depth: Descriptive statistics, graphics and inference, (with discussion and a rejoinder by Liu and Singh)”. The Annals of Statistics (1990), 27 (3), 783–858.
- [38] LOPEZ-PINTADO, S. Y ROMO, J. “On the concept of depth for functional data”. Journal of the American Statistical Association (2009), 104 (486), 718–734.

- [39] MASSART, M. “The tight constant in the Dvoretzky-Kiefer-Wolfowitz inequality”. Annals of Probability (1990), 18 (3), 1269–1283.
- [40] NAGY, S. “Consistency of h-mode depth”. Journal of Statistical Planning and Inference (2015), 165, 91–103.
- [41] NIETO-REYES, A. Y BATTEY, H. “A topologically valid definition of depth for functional data”. Statistical Science (2017), 31, 61–79.
- [42] NIETO-REYES, A. Y BATTEY, H. Correction to “A topologically valid definition of depth for functional data”. Statistical Science (2017), 32 (4), 640.
- [43] PAPADIMITRIOU, C, Y STEIGLITZ, K. “Combinatorial optimization: algorithms and complexity”. Prentice Hall, Englewood Cliffs (1982).
- [44] PAINDAVEINE, D. Y VAN BEVER, G. “Nonparametrically consistent depth-based classifiers”. Bernoulli (2015), 21 (1), 62–82.
- [45] PAINDAVEINE, D. Y VAN BEVER, G. “From depth to local depth : A focus on centrality”. Journal of the American Statistical Association (2013), 108, 1105–1119.
- [46] PENG, J. Y MULLER, H. “Distance-based clustering of sparsely observed stochastic processes, with applications to online auctions”. Annals Applied Statistics (2008), 2, 1056–1077.
- [47] RAMSAY, J. Y SILVERMAN, B., W. “Functional data analysis”. Springer (2005).
- [48] SERFLING, R. “Nonparametric multivariate descriptive measures based on spatial quantile”. Journal of Statistical Planning Inference (2004), 123, 259–278.
- [49] SUN Y, Y GENTON, M., G. “Functional boxplots”. Journal of Computational and Graphical Statistics (2011), 20, 316–334.
- [50] SCHMUTZ, A., JACQUES, J., BOUVEYRON, C., CHEZE, L. Y MARTIN, P. “Clustering multivariate functional data in group-specific functional subspaces”. Preprint (2017), <https://hal.inria.fr/hal--01652467/file/Clusteringmultivariatefunctionaldata.pdf>.
- [51] SERBAN, N. Y JIANG, H. “Clustering random curves under spatial interdependence with application to service accessibility”. Technometrics (2012), 54 (2), 108–119.
- [52] TUDDENHAM, RD. Y SNYDER, M. “Physical growth of California boys and girls from birth to eighteen 630 years”. Tech. Rep. 1, University of California Publications in Child Development (1954).

- [53] TUKEY, J. W. “Mathematics and the picturing of data”. Proceedings of the International Congress Math Vancouver (1975), 2, 523–531.
- [54] TIBSHIRANI, R., WALTHER, G. Y HASTIE, T. “Estimating the number of data clusters via the gap statistic”. Journal of the Royal Statistical Society, B (2001), 62, 411–423.
- [55] WITTEN, D. Y TIBSHIRANI, R. “A framework for feature selection in clustering”. Journal of the American Statistical Association (2010), 490, 713–726.
- [56] YASSOURIDIS, C. Y LEISCH, F. “Benchmarking different clustering algorithms functional data”. Advances in Data Analysis and Classification (2017), 11 (3), 467–492.
- [57] ZUO, Y. Y SERFLING, R. “General notion of statistical depth function”. The Annals of Statistics (2000), 28 (2), 461–482.