



UNIVERSIDAD NACIONAL DE LA PLATA
FACULTAD DE CIENCIAS EXACTAS
DEPARTAMENTO DE MATEMÁTICA

Trabajo de Tesis Doctoral

Estimadores robustos en modelos de regresión
no paramétricos funcionales y en modelos
semi-funcionales parcialmente lineales

Alejandra Valeria Vahnovan

Directora: Dra. Graciela Boente

Co-Director: Dr. Ricardo Maronna

2013

A Olivia y Mariano

Agradecimientos

A Graciela y Ricardo por haber aceptado ser mis directores en este trabajo de tesis, y en especial a Graciela por su paciencia, dedicación y constante apoyo en todo el camino recorrido.

A mis amigos por estar siempre presentes, por los almuerzos en la oficina, las charlas y los mates compartidos.

Al Departamento de Matemática por haberme dado un lugar de trabajo y brindado las condiciones para que pudiera finalizar esta tesis.

A mi familia por apoyarme y acompañarme en cada momento de esta etapa en forma incondicional, como lo hicieron en cada momento de mi vida.

A Mariano y Olivia, las dos personas que desde que llegaron a mi vida me bancaron y acompañaron día y noche, recargando mis energías y haciendo de cada día, el mejor.

Índice general

1. Introducción	4
2. Definiciones y resultados previos	7
2.1. Convergencia en espacios de probabilidad	7
2.2. Robustez	9
2.3. Diferenciabilidad	12
2.4. Entropía y Desigualdad maximal	13
3. Regresión noparamétrica y parcialmente lineal en el caso finito dimensional	17
3.1. Estimadores clásicos	18
3.2. Estimadores robustos	19
4. Regresión noparamétrica funcional	22
4.1. El espacio y la semimétrica	23
4.2. Estimadores clásicos	24
4.2.1. Estimación de la función de regresión	25
4.2.2. Estimación de la función de distribución acumulada condicional . .	26
4.3. Estimadores noparamétricos robustos	27
4.3.1. Definiciones y notación	27
4.3.2. Hipótesis	28
4.3.3. Notación y resultados preliminares	32
4.3.4. Convergencia fuerte uniforme	35
4.4. Tasas de convergencia fuerte uniforme	40
5. Un resultado auxiliar sobre el número de cubrimiento	52
5.1. El resultado	52
6. Modelo parcialmente lineal en dimensión infinita	57
6.1. Estimadores clásicos	58
6.2. Estimadores robustos	59

7. Comportamiento asintótico de los estimadores bajo un modelo PLSF	62
7.1. Consistencia	62
7.2. Distribución Asintótica	64
8. Estudio de Monte Carlo	72
8.1. Modelo y estimadores considerados	72
8.2. Medidas resumen de la simulación	75
8.3. Resultados	76
9. Apéndice	95

Capítulo 1

Introducción

La mayoría de los procedimientos estadísticos clásicos están basados en modelos con hipótesis rígidas, tales como errores normales, observaciones equidistribuidas, etc.. Bajo estas hipótesis, se deducen procedimientos óptimos. Por ejemplo, para el caso de regresión el procedimiento óptimo es el de mínimos cuadrados; para modelos paramétricos en general, los procedimientos óptimos clásicos son los estimadores de máxima verosimilitud. Sin embargo, estos métodos son muy sensibles al incumplimiento de las hipótesis que los generaron, tales como la presencia en la muestra de observaciones atípicas. En efecto, pequeñas desviaciones a la normalidad manifestadas por unas pocas desviaciones atípicas pueden invalidar las conclusiones basadas en estos procedimientos. Los procedimientos estadísticos robustos tienen como objetivo permitir inferencias válidas cuando el modelo no se cumple exactamente y, al mismo tiempo, ser altamente eficientes bajo el modelo central.

Por otra parte, los modelos clásicamente usados son paramétricos y la suposición es que la muestra de observaciones proviene de una familia paramétrica conocida. En estos casos, el problema es estimar los parámetros desconocidos o hallar tests de hipótesis o intervalos de confianza para los parámetros. Esta suposición puede ser relativamente fuerte porque el modelo paramétrico supuesto puede no ser el correcto si existe alguno (los datos pueden ser tales que no exista una familia paramétrica adecuada que dé un buen ajuste). Además, los métodos estadísticos desarrollados para un modelo paramétrico particular pueden llevar a conclusiones erróneas cuando se aplican a un modelo ligeramente perturbado (falta de robustez respecto del modelo). Estos problemas llevaron a la tendencia de desarrollar además de procedimientos estadísticos robustos, métodos noparamétricos o semiparamétricos para analizar los datos finito-dimensionales.

En muchas situaciones, uno se enfrenta con datos que, o bien, son, o bien, parecen provenir de un proceso suave. En tales situaciones, discretizar el proceso para estudiar las componentes de variación mediante las técnicas usuales de componentes principales, regresión paramétrica o noparamétrica no parece ser lo más indicado. Por otra parte, las técnicas de Análisis Multivariado no darán información sobre las derivadas o las integrales

de las funciones que se suponen yacen bajo los datos.

Un problema muy común en estadística consiste en tratar de explicar cómo una variable de interés Y se relaciona con una covariable X . Podemos encontrar diversos autores que en los últimos años han abordado problemas en donde la variable explicativa X toma valores en un espacio de dimensión infinita (o espacio funcional). Este tipo de variables, conocidas como variables funcionales en la literatura existente, nos permite considerar variables como funciones (del tiempo, por ejemplo).

En este contexto general, el estudio de modelos estadísticos adaptados a datos de dimensión infinita ha sido objeto de numerosos trabajos en la literatura estadística reciente. Trabajos previos sobre modelos paramétricos (mayormente lineales) y buenas recopilaciones acerca de esta literatura puede encontrarse en Ramsay y Silverman (2005, 2002). En este último, figuran ejemplos basados en datos reales que ilustran cómo las propuestas del análisis de datos funcionales son adecuadas en situaciones prácticas, en diversas áreas de aplicación como son la criminología, la economía y la psicología, entre otras. Podemos mencionar también diversas técnicas de análisis para datos funcionales que se han desarrollado en los últimos años como el análisis de componentes principales funcional estudiado por Dauxois *et al.* (1982) y por Ramsay y Dalzell (1991) y la regresión paramétrica funcional, ver Yao *et al.* (2005) y Cai y Hall (2006), entre otros (ver también Ramsay y Silverman, 1997). Como es bien sabido, en muchos casos la relación entre la variable respuesta y la explicativa no se conoce exactamente y en ese caso, los modelos no-paramétricos ofrecen una opción más flexible tanto en el caso finito-dimensional como en el caso en que las variables explicativas provienen de un proceso suave. El estudio de modelos noparamétricos para datos funcionales es mucho más reciente. Ferraty *et al.* (2006) consideraron el problema de predecir una variable aleatoria real a partir de variables explicativas funcionales mediante una aproximación noparamétrica basada en núcleos. Para una descripción de distintos procedimientos para datos funcionales ver Ferraty y Vieu (2006) y Ferraty y Romain (2010).

En particular, el *modelo de regresión noparamétrica funcional* puede escribirse como

$$Y = r(X) + \epsilon \tag{1.1}$$

donde Y es una variable aleatoria, X es una variable aleatoria funcional, es decir, toma valores en un espacio infinito-dimensional \mathcal{H} y r es un operador suave que no tiene porqué ser lineal. Supongamos tener n observaciones (Y_i, X_i) , $1 \leq i \leq n$, independientes e idénticamente distribuidas que satisfacen (1.1). De ahora en más, se supondrá que \mathcal{H} es un espacio de Banach separable con norma $\|\cdot\|$. Este contexto contiene el espacio L^2 así como espacios de Sobolev o de Besov. La hipótesis usual sobre los errores ϵ_i es $\mathbb{E}(\epsilon_i|X_i) = 0$ $\mathbb{E}(\epsilon_i^2|X_i = X) = \sigma^2(X)$.

El estimador clásico es un estimador de tipo Nadaraya–Watson y puede definirse como

$$\hat{r}(x) = \sum_{i=1}^n Y_i K\left(\frac{\|X_i - x\|}{h}\right) \left\{ \sum_{i=1}^n K\left(\frac{\|X_i - x\|}{h}\right) \right\}^{-1}$$

Por estar basado en un promedio de las variables respuesta este estimador es muy sensible a observaciones atípicas, particularmente a aquellas que se encuentran en el entorno de x .

Estos estimadores fueron extendidos al caso en que se considera una semimétrica d para medir la cercanía entre x y X_i . Attouch *et al.* (2007) extendieron los M –estimadores robustos dados por Collomb y Härdle (1986) del caso real al caso funcional. Asimismo, Crambes *et al.* (2008) estudiaron sus propiedades asintóticas. Sin embargo, estos autores suponen que la función de escala es conocida. En el Capítulo 4, se considera el estimador definido como la solución $\widehat{r}(x)$ de

$$\sum_{i=1}^n \psi \left(\frac{Y_i - \widehat{r}(x)}{\widehat{s}(x)} \right) K \left(\frac{d(X_i, x)}{h} \right) = 0$$

con $\widehat{s}(x)$ un estimador robusto consistente de la escala $\sigma(x)$ de las observaciones. Para ello, supondremos que los errores $\epsilon_i = \sigma(X_i)u_i$ con u_i independientes e independientes de X_i y tales que u_i tiene distribución simétrica. Se obtendrán resultados de consistencia fuerte y ordenes de convergencia que resultarán de utilidad en el Capítulo 6.

Aneiros-Pérez y Vieu (2006) consideraron una generalización del modelo de regresión no paramétrico en el que además de la variable funcional, se incluye otro conjunto de covariables $\mathbf{Z} \in \mathbb{R}^p$ que se relacionan linealmente con la respuesta. Este modelo, llamado modelo de regresión parcialmente lineal semi–funcional (PLSF), supone se tienen n observaciones i.i.d. (Y_i, \mathbf{Z}_i, X_i) , $i = 1, \dots, n$, tales que

$$Y_i = \boldsymbol{\beta}^T \mathbf{Z}_i + r(X_i) + \epsilon_i$$

donde Z_{ij} ($j = 1, \dots, p$) son variables explicativas reales, X_i es otra variable explicativa pero de naturaleza funcional, los errores ϵ_i son independientes e idénticamente distribuidos e independientes de $(\mathbf{Z}_i^T, X_i)^T$, $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^T$ es un vector de parámetros reales desconocidos y $r : \mathcal{H} \rightarrow \mathbb{R}$ es un operador suave desconocido que no tiene porqué ser lineal. En el Capítulo 6, se introduce una familia de estimadores robustos para $\boldsymbol{\beta}$ y r . Resultados de consistencia fuerte se obtienen en el Capítulo 7 así como la distribución asintótica del estimador de $\boldsymbol{\beta}$. En el Capítulo 8, se presentan los resultados de un estudio de Monte Carlo que permite comparar el comportamiento de la propuesta dada con el estimador clásicamente utilizado.

Algunas definiciones y resultados preliminares, se resumen en el Capítulo 2, mientras que algunas de las propuestas de estimadores existentes en el caso finito–dimensional se describen en el Capítulo 3. Finalmente, en el Capítulo 5 se obtiene un resultado sobre el número de cubrimiento de familias de funciones en espacios funcionales.

Capítulo 2

Definiciones y resultados previos

En este capítulo, se resumen algunas definiciones y resultados que serán utilizados en los capítulos siguientes y que se incluyen por completitud.

2.1. Convergencia en espacios de probabilidad

En esta Sección, se dan algunos resultados sobre convergencia de probabilidades definidas sobre un espacio métrico que pueden verse, por ejemplo, en Billingsley (1968).

Definición 1. Sea \mathcal{P} la colección de las medidas de probabilidad sobre (S, \mathcal{S}) , donde S es un espacio métrico con métrica d y \mathcal{S} es la σ -álgebra de Borel asociada (o sea, aquella generada por los conjuntos abiertos del espacio). Una sucesión $\{P_n\}_{n \geq 1}$ de elementos de \mathcal{P} se dice que converge débilmente a $P \in \mathcal{P}$ y escribimos $P_n \Rightarrow P$ si

$$\lim_{n \rightarrow \infty} \int f dP_n = \int f dP$$

para toda función continua y acotada $f : S \rightarrow \mathbb{R}$.

A \mathcal{P} se la puede dotar de una topología, denominada topología débil, tomando como base los entornos de un punto $P \in \mathcal{P}$ a la colección

$$\{Q \in \mathcal{P} : |\int_S f_i dQ - \int_S f_i dP| < \epsilon \ 1 \leq i \leq k\}$$

donde $f_i : \mathbb{R} \rightarrow \mathbb{R}$, $i = 1, \dots, k$, son funciones continuas y acotadas sobre S y $\epsilon \in \mathbb{R}$ es positivo. De este modo la convergencia débil es la convergencia en esta topología.

Bajo ciertas condiciones, la topología débil sobre el espacio S es metrizable. La siguiente proposición da una de las posibles métricas.

Proposición 2.1.1. Sea $\pi : \mathcal{P} \times \mathcal{P} \longrightarrow \mathbb{R}$ la función definida como

$$\pi(P, Q) = \inf\{\epsilon : P(A) \leq Q(A^\epsilon) + \epsilon \forall A \in \mathcal{S}\},$$

donde $A^\epsilon = \{x \in S : d(x, A) \leq \epsilon\}$.

Entonces π es una métrica sobre \mathcal{P} y si, además, S es separable y completo entonces (\mathcal{P}, π) resulta también un espacio métrico, separable y completo y π metriza la topología débil.

A π se la denomina la métrica o distancia de Prohorov. La métrica de Prohorov no es la única que metriza la topología débil.

Es fácil verificar que la métrica de Prohorov tiene las propiedades de una distancia tomando valores entre 0 y 1:

- $\pi(P, Q) \geq 0$
- $\pi(P, Q) \leq 1$
- $\pi(P, Q) = \pi(Q, P)$
- $\pi(P, Q) = 0 \Leftrightarrow P = Q$.

Lema de Portmanteau. Sean $\{P_n\}_{n \geq 1}$ y P medidas de probabilidad (o funciones de distribución) sobre $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, donde $\mathcal{B}(\mathbb{R})$ es la σ -álgebra de Borel de \mathbb{R} con la métrica usual. Son equivalentes:

- i) $P_n \Rightarrow P$;
- ii) $\lim_{n \rightarrow \infty} \int f dP_n = \int f dP$ para cada $f : \mathbb{R} \rightarrow \mathbb{R}$ uniformemente continua y acotada;
- iii) $\lim_{n \rightarrow \infty} \int f dP_n = \int f dP$ para cada $f : \mathbb{R} \rightarrow \mathbb{R}$ Lipschitz continua y acotada;
- iv) $\lim_{n \rightarrow \infty} \inf P_n(G) \geq P(G)$ para cada conjunto abierto G ;
- v) $\lim_{n \rightarrow \infty} \sup P_n(F) \leq P(F)$ para cada conjunto cerrado F ;
- vi) $\lim_{n \rightarrow \infty} P_n(A) = P(A)$ para cada $A \in \mathcal{B}(\mathbb{R})$ tal que $P(\partial A) = 0$ donde ∂A es la frontera de A .

Observemos que, si consideramos $\{X_n\}_{n \geq 1}$ y X variables aleatorias y F_n y F sus correspondientes funciones de distribución, el item (vi) del Lema de Portmanteau implica el concepto usual de convergencia en distribución, que denotaremos con $X_n \xrightarrow{D} X$, a saber, que $\lim_{n \rightarrow \infty} F_n(x) = F(x)$, para cada punto de continuidad, x , de F .

Definición 2. Sea $(z_n)_{n \in \mathbb{N}^*}$ una sucesión de variables reales; decimos que z_n converge casi completamente ($z_n \xrightarrow{a.co.} 0$) a cero si y solo si $\forall \epsilon > 0, \sum_{n=1}^{\infty} \mathbb{P}(|z_n| > \epsilon) < \infty$. Más aún, dada $(u_n)_{n \in \mathbb{N}^*}$ una sucesión de números reales positivos; decimos que $z_n = O(u_n)$ a.co. si y sólo si $\exists A > 0, \sum_{n=1}^{\infty} \mathbb{P}(|z_n| > A u_n) < \infty$.

Por el Lema de Borel–Cantelli, la convergencia casi completa implica la convergencia en casi todo punto. Por lo tanto, es usual obtener resultados de convergencia completa sobre sucesiones de estimadores, utilizando desigualdades exponenciales, de modo a deducir la consistencia fuerte.

2.2. Robustez

Definición 3. Una función- ρ denotará a una función $\rho(u)$ tal que:

- $\rho(0) = 0$.
- $\rho(-u) = \rho(u)$.
- $0 \leq u \leq v$ implica $\rho(u) \leq \rho(v)$.
- ρ es continua.
- $\sup_u \rho(u) = 1$.
- Si $\rho(u) < 1$ y $0 \leq u < v$ entonces $\rho(u) < \rho(v)$.

Observemos que de acuerdo a la terminología de Maronna *et al.* (2006) ésta sería una “función- ρ acotada”.

Definición 4. Una función- ψ denotará a una función ψ que es la derivada de una función- ρ .

Una función- ρ muy popular es la *función bicuadrada*:

$$\rho_B(u) = \begin{cases} 1 - [1 - (u/k)^2]^3 & \text{si } |u| \leq k \\ 1 & \text{si } |u| > k, \end{cases} \quad (2.1)$$

cuya derivada es $\rho'_B(u) = 6\psi_B(u)/k^2$ donde

$$\psi_B(u) = u \left[1 - \left(\frac{u}{k} \right)^2 \right]^2 I(|u| \leq k).$$

En la Figura 2.1 se muestra el gráfico de la función- ρ bicuadrada y su correspondiente función- ψ para $k = 1$.

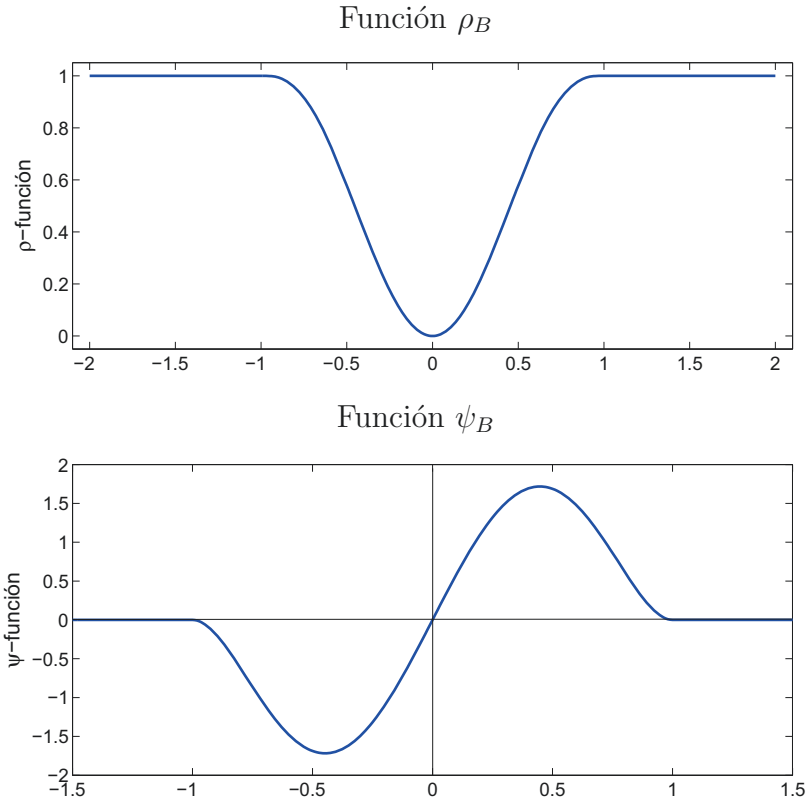


Figura 2.1: Funciones- ρ y $-\psi$ bicuadradas.

Otro tipo de funciones- ρ (respectivamente ψ) muy utilizada es la familia de funciones de Huber definida por

$$\rho_k(u) = \begin{cases} \frac{1}{2}u^2 & \text{si } |u| \leq k \\ k|u| - \frac{1}{2}k^2 & \text{si } |u| > k, \end{cases} \quad (2.2)$$

con derivada $\psi_k(u)$, donde

$$\psi_k(u) = \begin{cases} u & \text{si } |u| \leq k \\ \text{signo}(u)k & \text{si } |u| > k. \end{cases}$$

En la Figura 2.2 se muestra el gráfico de la función- ρ de Huber y su correspondiente función- ψ para $k = 1$.

Definición 5. Sean y_i variables respuesta y $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})^T$ covariables tales que (y_i, \mathbf{x}_i) son independientes. El modelo de regresión lineal puede escribirse como

$$y_i = \mathbf{x}_i^T \boldsymbol{\beta} + \epsilon_i,$$

donde $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^T$, ϵ_i son independientes e independientes de \mathbf{x}_i y $\mathbb{E}(\epsilon_i) = 0$.

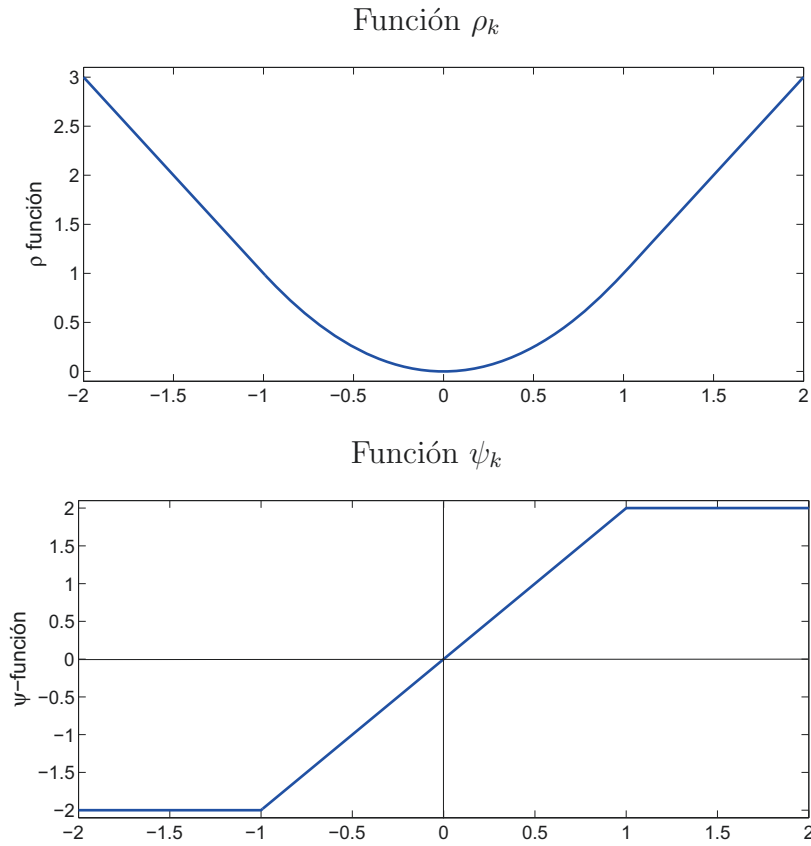


Figura 2.2: Funciones- ρ y $-\psi$ de Huber.

Los valores ajustados \hat{y}_i y los residuos r_i correspondientes a un vector \mathbf{b} están definidos respectivamente como

$$\hat{y}_i(\mathbf{b}) = \mathbf{x}_i^T \mathbf{b} \quad \text{y} \quad r_i(\mathbf{b}) = y_i - \hat{y}_i(\mathbf{b}).$$

Definición 6. Los M -estimadores de regresión $\hat{\boldsymbol{\beta}}$ están definidos como

$$\hat{\boldsymbol{\beta}} = \underset{\mathbf{b} \in \mathbb{R}^p}{\operatorname{argmin}} \sum_{i=1}^n \rho \left(\frac{r_i(\mathbf{b})}{\hat{\sigma}} \right) \quad (2.3)$$

Aquí ρ es una ρ -función, y $\hat{\sigma}$ es un estimador de escala auxiliar que es requerido para que $\hat{\boldsymbol{\beta}}$ resulte equivariante por cambios de escala.

Derivando (2.3) obtenemos la ecuación

$$\sum_{i=1}^n \psi \left(\frac{r_i(\hat{\boldsymbol{\beta}})}{\hat{\sigma}} \right) \mathbf{x}_i = 0, \quad (2.4)$$

donde $\psi = \rho'$.

Se supondrá que las funciones ρ y ψ son respectivamente ρ - y ψ -funciones y que la matriz $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_n)$ tiene rango completo.

Las soluciones de 2.4 con ψ monótona (resp. redescendiente) se llaman M -estimadores de regresión monótonos (resp. redescendientes). La ventaja de los estimadores monótonos es que todas las soluciones de (2.4) son soluciones de (2.3). Más aún, si ψ es creciente entonces la solución es única (ver Maronna *et al.* 2006).

2.3. Diferenciabilidad

Hay varias maneras de definir diferenciabilidad de aplicaciones ϕ entre espacios normados. A continuación se recordará la definición de diferenciabilidad de Hadamard y unos resultados que se pueden encontrar en Van der Vaart (1998).

Definición 7. Una aplicación $\phi : D_\phi \rightarrow E$, definida sobre un subconjunto D_ϕ de un espacio normado D que contiene a θ , se dice que es Hadamard diferenciable en θ si existe una aplicación lineal, continua $\phi'_\theta : D \rightarrow E$ tal que

$$\left\| \frac{\phi(\theta + th_t) - \phi(\theta)}{t} - \phi'_\theta(h) \right\|_E \rightarrow 0$$

cuando $t \downarrow 0$, y para toda $h_t \rightarrow h$. Más precisamente, para cada $h_t \rightarrow h$ tal que $\theta + th_t$ está en el dominio de ϕ para todo $t > 0$ pequeño.

La definición requiere que $\phi'_\theta : D \rightarrow E$ exista como aplicación sobre todo D . Si esto no sucede, pero ϕ'_θ existe sobre un subconjunto D_0 y las sucesiones $h_t \rightarrow h$ se restringen a converger a límites $h \in D_0$, entonces se dice que ϕ es Hadamard diferenciable tangencialmente a ese subconjunto.

La continuidad de la aplicación $\phi'_\theta : D \rightarrow E$ no debe confundirse con continuidad de la dependencia $\theta \mapsto \phi'_\theta$ (si ϕ tiene derivadas en un entorno de θ).

Teorema 2.3.1. (Regla de la cadena): Sean $\phi : D_\phi \rightarrow E$ y $\psi : E_\psi \rightarrow F$ aplicaciones definidas sobre subconjuntos D_ϕ y E_ψ de espacios normados D y E respectivamente. Sea ϕ Hadamard diferenciable en θ tangencialmente a D_0 y sea ψ Hadamard diferenciable en $\phi(\theta)$ tangencialmente a $\phi'_\theta(D_0)$. Entonces $\psi \circ \phi : D_\phi \rightarrow F$ es Hadamard diferenciable en θ tangencialmente a D_0 con derivada $\psi'_{\phi(\theta)} \circ \phi'_\theta$.

La demostración de este teorema y otros resultados relacionados se pueden encontrar, por ejemplo, en Van der Vaart (1998).

2.4. Entropía y Desigualdad maximal

En esta Sección recordaremos la noción de entropía y unas desigualdades que son de utilidad para mostrar resultados de convergencia uniforme. Sea (\mathcal{H}, d) un espacio semimétrico, es decir, \mathcal{H} es un espacio (posiblemente de dimensión infinita) y d define una semimétrica, o sea, $d : \mathcal{H} \times \mathcal{H} \rightarrow \mathbb{R}$ es tal que $d(x, y) \geq 0$, $d(x, y) = d(y, x)$ y $d(x, z) \leq d(x, y) + d(y, z)$. Esta situación incluye el caso de espacios normados. Indicaremos $B(x, \epsilon) = \{y \in \mathcal{H} : d(x, y) \leq \epsilon\}$.

Definición 8. Una ϵ -red para un subconjunto S de \mathcal{H} es un conjunto de puntos $\{x_k\}_{k \in J}$ con la propiedad de que para cada $x \in S$ existe $k \in J$ tal que $x \in B(x_k, \epsilon)$, es decir, $S \subset \cup_{j \in J} B(x_k, \epsilon)$. Obsérvese que los puntos x_k no necesariamente son elementos de S .

Definición 2.4.1. Dado un subconjunto $S \subset \mathcal{H}$ y $\epsilon > 0$, sea $N_\epsilon(S)$ el mínimo número de bolas abiertas de radio ϵ necesarias para cubrir a S . La ϵ -entropía de Kolmogorov del conjunto S está dada por $\psi_S(\epsilon) = \log(N_\epsilon(S))$.

La noción de entropía fue introducida por Kolmogorov (ver, por ejemplo, Kolmogorov y Tikhomirov, 1959) y representa una medida de la complejidad de un conjunto, en el sentido que, una alta entropía significa que se necesita mucha información para describir un elemento con una exactitud ϵ . Por lo tanto, la elección de la estructura topológica (en otras palabras, la elección de la semi-métrica) jugará un papel crucial en la búsqueda de resultados asintóticos uniformes sobre algún subconjunto $S_{\mathcal{H}}$ de \mathcal{H} .

Ejemplos

A continuación se mencionan algunos ejemplos y cálculos de entropía que pueden encontrarse en Ferraty *et al.* (2010).

- *Ejemplo 1 (subconjunto compacto en un espacio de dimensión finita):* Un teorema estandar de topología garantiza que para cada subconjunto compacto S de \mathbb{R}^p y para cada $\epsilon > 0$ existe una ϵ -red y además, se tiene que para cualquier $\epsilon > 0$,

$$\psi_S(\epsilon) \leq Cp \log \left(\frac{1}{\epsilon} \right),$$

donde C es una constante que depende solamente de d y del conjunto S . Más precisamente, Chate y Courbage (1997) han mostrado que, para cualquier $\epsilon > 0$, el poliedro regular en \mathbb{R}^p con longitud r puede ser cubierto por $(\lceil 2r\sqrt{p}/\epsilon \rceil + 1)^p$ bolas, donde $\lceil m \rceil$ denota al mayor entero menor o igual que m . Luego, la ϵ -entropía de Kolmogorov de un poliedro P_r en \mathbb{R}^p con longitud r es

$$\forall \epsilon > 0, \quad \psi_{P_r}(\epsilon) \sim p \log \left(\left(\left\lceil \frac{2r\sqrt{p}}{\epsilon} \right\rceil + 1 \right) \right).$$

- *Ejemplo 2 (Bola cerrada en un espacio de Sobolev):* Kolmogorov y Tikhomirov (1959) obtuvieron algunas cotas superiores e inferiores para la ϵ -entropía de varios subconjuntos funcionales. Sea $\mathcal{H} = \{f : [0, 2\pi) \rightarrow \mathbb{R}\}$ con norma L^2 . Para $r > 0$ fijo, consideremos el subconjunto $W_2^m(r) \subset \mathcal{H}$ de funciones $f \in \mathcal{H}$ con condiciones de acotación periódicas, m veces diferenciables y tales que

$$\frac{1}{2\pi} \int_0^{2\pi} f^2(t) dt + \frac{1}{2\pi} \int_0^{2\pi} f^{(m)^2}(t) dt \leq r.$$

La ϵ -entropía de esta clase (respecto de la norma L^2) es

$$\psi_{W_2^m(r)}(\epsilon) \leq C \left(\frac{1}{\epsilon} \right)^{\frac{1}{m}}.$$

- *Ejemplo 3 (Subconjunto compacto en un espacio de Hilbert con semi-métrica proyección):* Las semi-métricas basadas en proyecciones se construyen de la siguiente manera. Supongamos que \mathcal{H} es un espacio de Hilbert separable, con producto interno $\langle \cdot, \cdot \rangle$ y con base ortonormal $\{e_1, \dots, e_j, \dots\}$, y sea k un entero fijo mayor que cero. Como se muestra en Ferraty y Vieu (2006, Lema 13.6), una semimétrica d_k sobre \mathcal{H} puede definirse como sigue:

$$d_k(x, y) = \sqrt{\sum_{i=1}^k \langle x - y, e_i \rangle^2}. \quad (2.5)$$

Sea $\mathcal{X} : \mathcal{H} \rightarrow \mathbb{R}^k$ el operador definido como

$$\mathcal{X}(x) = (\langle x, e_1 \rangle, \dots, \langle x, e_k \rangle),$$

y sea d_{eucl} la distancia euclídea en \mathbb{R}^k . Denotemos por $B_{eucl}(\cdot, \cdot)$ una bola abierta en \mathbb{R}^k con la topología asociada. Similarmente, llamemos $B_k(\cdot, \cdot)$ a una bola abierta de \mathcal{H} con la semimétrica d_k . Como \mathcal{X} es una aplicación continua de (\mathcal{H}, d_k) en (\mathbb{R}^k, d_{eucl}) , se tiene que para cualquier subconjunto compacto S de (\mathcal{H}, d_k) , $\mathcal{X}(S)$ es un subconjunto compacto de \mathbb{R}^k . Luego, para cada $\epsilon > 0$ se puede cubrir $\mathcal{X}(S)$ con bolas de centro $z_i \in \mathbb{R}^k$ y radio ϵ , es decir,

$$\mathcal{X}(S) \subset \bigcup_{i=1}^{N_\epsilon} B_{eucl}(z_i, \epsilon) \quad \text{con} \quad N_\epsilon = N_\epsilon(\mathcal{X}(S)) = C \epsilon^{-k} \quad \text{para algún } C > 0. \quad (2.6)$$

Para $i = 1, \dots, N_\epsilon$, sea x_i un elemento de \mathcal{H} tal que $\mathcal{X}(x_i) = z_i$. La solución de la ecuación $\mathcal{X}(x) = z_i$ en general no es única, pero se toma x_i como una de esas soluciones. Por (2.5) se tiene que

$$\mathcal{X}^{-1}(B_{eucl}(z_i, \epsilon)) = B_k(x_i, \epsilon). \quad (2.7)$$

Finalmente, (2.6) y (2.7) son suficientes para mostrar que la ϵ -entropía de Kolmogorov de S es

$$\psi_S(\epsilon) = C k \log \left(\frac{1}{\epsilon} \right).$$

Para definir la integral de entropía recordaremos algunas nociones asociadas.

Sea X_1, \dots, X_n una muestra aleatoria de una distribución de probabilidad \mathbb{P} sobre un espacio medible $(\mathcal{X}, \mathcal{A})$. La distribución empírica es la medida uniforme discreta de las observaciones y la denotaremos por $P_n = n^{-1} \sum_{i=1}^n \delta_{X_i}$ donde δ_x es la masa puntual en x . Dada una función medible $f : \mathcal{X} \rightarrow \mathbb{R}$, indicamos por $P_n f$ a la esperanza de f bajo la medida empírica, y $\mathbb{P}f$ a la esperanza de f bajo \mathbb{P} , es decir,

$$P_n f = \frac{1}{n} \sum_{i=1}^n f(X_i), \quad \mathbb{P}f = \int f d\mathbb{P}.$$

El proceso empírico evaluado en f está definido como $G_n f = \sqrt{n}(P_n f - \mathbb{P}f)$. El supremo de esos procesos estocásticos sobre una clase dada \mathcal{F} de funciones medibles se denotará por $\|G_n\|_{\mathcal{F}} = \sup_{f \in \mathcal{F}} |G_n f|$.

Definición 9. Una función envolvente de una clase \mathcal{F} es una función medible $F : \mathcal{X} \rightarrow \mathbb{R}$ tal que $|f| \leq F$ para toda $f \in \mathcal{F}$.

Como sabemos, una forma relativamente simple de medir el tamaño de una clase \mathcal{F} es en términos de la entropía. Consideraremos la entropía con bracketing relativa a la norma $L^r(\mathbb{P})$ dada por $\|f\|_{\mathbb{P},r} = (\mathbb{P}|f|^r)^{1/r}$. Dadas dos funciones l y u , el bracket $[l, u]$ es el conjunto de todas las funciones f tales que $l \leq f \leq u$. Un ϵ -bracket en $L^r(\mathbb{P})$ es un bracket $[l, u]$ con $\|u - l\|_{\mathbb{P},r} < \epsilon$.

Definición 10. El número de bracketing $N_{[\cdot]}(\epsilon, \mathcal{F}, L^r(\mathbb{P}))$ es el mínimo número de ϵ -brackets necesarios para cubrir \mathcal{F} . La entropía con bracketing es el logaritmo del número de bracketing.

Definición 11. La integral de bracketing está definida como

$$J_{[\cdot]}(\delta, \mathcal{F}, L^2(\mathbb{P})) = \int_0^\delta \sqrt{\log N_{[\cdot]}(\epsilon, \mathcal{F}, L^2(\mathbb{P}))} d\epsilon.$$

Definición 12. El número de cubrimiento indicado $N(\epsilon, \mathcal{F}, L^r(\mathbb{Q}))$ es el mínimo número de $L^r(\mathbb{Q})$ bolas de radio ϵ necesarias para cubrir \mathcal{F} . La entropía es el logaritmo del número de cubrimiento.

Definición 13. Si F es una función envolvente de \mathcal{F} con cuadrado integrable diremos que \mathcal{F} satisface la condición de entropía uniforme si

$$\int_0^\infty \sup_{\mathbb{Q}} \sqrt{\log N(\epsilon \|F\|_{\mathbb{Q},2}, \mathcal{F}, L^2(\mathbb{Q}))} d\epsilon < +\infty,$$

donde el supremo está tomado sobre todas las medidas de probabilidad discretas finitas con $\|F\|_{\mathbb{Q},2} > 0$.

Sea $\mathcal{J}(\delta, \mathcal{F})$ la función definida por

$$\mathcal{J}(\delta, \mathcal{F}) = \sup_{\mathbb{Q}} \int_0^\delta \sqrt{1 + \log N(\epsilon \|F\|_{\mathbb{Q},2}, \mathcal{F}, L^2(\mathbb{Q}))} d\epsilon,$$

donde el supremo está tomado sobre todas las medidas de probabilidad discretas \mathbb{Q} con $\|F\|_{\mathbb{Q},2} > 0$. Entonces $\mathcal{J}(1, \mathcal{F}) < \infty$ si \mathcal{F} cumple la condición de entropía uniforme dada en la definición 13 y además $\mathcal{J}(\delta, \mathcal{F}) \rightarrow 0$ si $\delta \rightarrow 0$. Para más detalles ver el Capítulo 2.14 de Van der Vaart y Wellner.

Capítulo 3

Regresión no paramétrica y parcialmente lineal en el caso finito dimensional

En este capítulo, se describen procedimientos de estimación, basados en núcleos, para la función de regresión bajo un modelo de regresión no paramétrico y para el parámetro de regresión y la componente no paramétrica bajo un modelo parcialmente lineal.

Recordemos que el modelo de regresión no paramétrico está definido como

$$Y = r(\mathbf{X}) + \epsilon \quad (3.1)$$

donde Y es una variable aleatoria en \mathbb{R} , \mathbf{X} es un vector aleatorio en \mathbb{R}^p , $r : \mathbb{R}^p \rightarrow \mathbb{R}$ es una función suave, y ϵ es una variable aleatoria en \mathbb{R} y $\epsilon = \sigma(X)u$ con u independiente de \mathbf{X} .

Es de gran interés la estimación de la función de regresión $r(\cdot)$ a partir de una muestra de observaciones $\{(Y_i, \mathbf{X}_i) : i = 1, \dots, n\}$. En este capítulo, describiremos los procedimientos de estimación de la función de regresión r . Comenzaremos describiendo los métodos clásicos de estimación y luego, mencionaremos los procedimientos robustos.

Un inconveniente del modelo (3.1) es que, cuando p es grande, este modelo sufre de la así llamada *maldición de la dimensión* debido a que los entornos de un punto fijo \mathbf{x} se hacen cada vez más malos al crecer p . Por esta razón, se introdujeron los modelos de regresión parcialmente lineales, que combinan una componente de regresión lineal y una componente no paramétrica. En este sentido, estos modelos combinan la flexibilidad de los modelos no paramétricos y la fácil interpretación de los modelos lineales. En el modelo parcialmente lineal, tenemos observaciones $(Y_i, \mathbf{Z}_i^T, X_i)^T \in \mathbb{R}^{p+2}$ tales que

$$Y_i = \boldsymbol{\beta}^T \mathbf{Z}_i + g(X_i) + \epsilon_i, \quad i = 1, \dots, n \quad (3.2)$$

donde $Y_i \in \mathbb{R}$ es la variable respuesta, $\mathbf{Z}_i \in \mathbb{R}^p$ y $X_i \in \mathbb{R}$ son las covariables del estudio y los errores $\epsilon_i \in \mathbb{R}$ son i.i.d. e independientes de $(\mathbf{Z}_i^T, X_i)^T$. El parámetro $\boldsymbol{\beta}$ y la función $g : \mathbb{R} \rightarrow$

\mathbb{R} son desconocidos y se desean estimar. Los modelos parcialmente lineales son mucho más flexibles que los modelos lineales estándares porque tienen una componente paramétrica y una no paramétrica. En las secciones que siguen describiremos procedimientos para estimar r en el modelo de regresión no paramétrico funcional y β y g en el modelo parcialmente lineal.

3.1. Estimadores clásicos

Un estimador no paramétrico clásico de la función de regresión bajo el modelo (3.1) fue introducido por Nadaraya (1964) y Watson (1964) en forma independiente, y está dado por

$$\hat{r}(\mathbf{x}) = \frac{\sum_{i=1}^n \mathcal{K}(\mathbf{H}^{-1}(\mathbf{x} - \mathbf{X}_i)) Y_i}{\sum_{i=1}^n \mathcal{K}(\mathbf{H}^{-1}(\mathbf{x} - \mathbf{X}_i))}$$

donde \mathcal{K} es un núcleo en \mathbb{R}^p , es decir, una función $\mathcal{K} : \mathbb{R}^p \rightarrow \mathbb{R}$ tal que $\mathcal{K}(\mathbf{x}) \geq 0$ para todo $\mathbf{x} \in \mathbb{R}^p$ y $\int_{\mathbb{R}^p} \mathcal{K}(\mathbf{x}) d\mathbf{x} = 1$ y $\mathbf{H} \in \mathbb{R}^{p \times p}$ es una matriz ventana que debe tender a 0 cuando n crece, es decir, su norma de Frobenius tiende a cero. La función \mathcal{K} satisface algunas condiciones de regularidad, como por ejemplo, tener soporte compacto. En general $\mathbf{H} = h_n \mathbf{I}_p$ para alguna sucesión de ventanas h_n y \mathcal{K} es tomado como

$$\mathcal{K}(u) = K(\|u\|)$$

con K un núcleo univariado con soporte $[0, 1]$ que satisface algunas condiciones adicionales y $\|\cdot\|$ la norma Euclídea en \mathbb{R}^p , quedando el estimador

$$\hat{r}(\mathbf{x}) = \frac{\sum_{i=1}^n K\left(\frac{\|\mathbf{x} - \mathbf{X}_i\|}{h}\right) Y_i}{\sum_{i=1}^n K\left(\frac{\|\mathbf{x} - \mathbf{X}_i\|}{h}\right)},$$

donde $h = h_n$ es una sucesión de números reales positivos (llamada ventana) que tiende a cero. Como se puede ver, para cada \mathbf{x} , $\hat{r}(\mathbf{x})$ es un promedio pesado de las respuestas Y_i cuyos pesos dependen de las distancias (con respecto a la norma Euclídea) entre \mathbf{x} y \mathbf{X}_i . Luego, sólo aquellas observaciones con \mathbf{X}_i en la bola centrada en \mathbf{x} de radio h serán utilizadas en la estimación.

Si bien existen otros estimadores de la función de regresión, por su simplicidad de implementación e interpretación, el estimador de Nadaraya-Watson se ha convertido en un referente en el campo de la regresión no paramétrica.

Cuando existen momentos, si $\mathbb{E}(\epsilon|X) = 0$, tenemos que la función de regresión se expresa como $r(x) = \mathbb{E}(Y|\mathbf{X})$. Podemos encontrar en la literatura numerosos resultados teóricos relacionados con la estimación de la función de regresión. Por ejemplo, bajo ciertas hipótesis, que incluyen que $h \rightarrow 0$ y $nh^p \rightarrow \infty$, cuando $n \rightarrow \infty$ se ha demostrado que

$$\mathbb{E}(\hat{r}(\mathbf{x})) - r(x) = O(h^2)$$

$$\text{VAR}(\widehat{r}(\mathbf{x})) = O((nh^p)^{-1})$$

para \mathbf{x} un punto fijo en el interior del dominio de la función r .

Intuitivamente, cuando estimamos r en \mathbf{x} , h debe tender a cero para trabajar sólo con observaciones que están lo más cerca posible de \mathbf{x} para evitar sesgo, pero al mismo tiempo, el número de observaciones usadas en la estimación (es decir, en el entorno de \mathbf{x}) necesita crecer a infinito para asegurar que la varianza del estimador tienda a cero.

Hay numerosos métodos noparamétricos para estimar la función de regresión en el caso de observaciones i.i.d.. Una revisión de dichos procedimientos puede verse, por ejemplo, en Collomb (1981).

Como hemos dicho, cuando p es grande estos modelos sufren de la así llamada *maldición de la dimensión*. Por ello, consideraremos ahora el modelo parcialmente lineal (3.2). Este modelo ha sido ampliamente estudiado por diversos autores tales como Speckman (1988), Robinson (1988), Green *et al.* (1985) y Gao (1992), entre otros.

Speckman (1988) consideró primero una representación paramétrica de g como $\mathbf{W}\boldsymbol{\gamma}$, donde $\mathbf{W} \in \mathbb{R}^{n \times q}$ es una matriz de rango completo y $\boldsymbol{\gamma}$ es un parámetro adicional. El modelo queda en forma matricial

$$\mathbf{Y} = \mathbf{Z}\boldsymbol{\beta} + \mathbf{W}\boldsymbol{\gamma} + \boldsymbol{\epsilon},$$

donde $\mathbf{Y} = (Y_1, \dots, Y_n)^T$, $\mathbf{Z}^T = (\mathbf{Z}_1, \dots, \mathbf{Z}_n)$ y $\boldsymbol{\epsilon} = (\epsilon_1, \dots, \epsilon_n)^T$. Las ecuaciones normales para determinar $\boldsymbol{\beta}$ y $\boldsymbol{\gamma}$ son

$$\begin{aligned} \mathbf{Z}^T \mathbf{Z} \boldsymbol{\beta} &= \mathbf{Z}^T (\mathbf{Y} - \mathbf{W} \boldsymbol{\gamma}), \\ \mathbf{W} \boldsymbol{\gamma} &= P_{\mathbf{W}} (\mathbf{Y} - \mathbf{Z} \boldsymbol{\beta}) \end{aligned}$$

donde $P_{\mathbf{W}} = \mathbf{W}(\mathbf{W}^T \mathbf{W})^{-1} \mathbf{W}^T$ es la proyección sobre el espacio generado por las columnas de \mathbf{W} . Green *et al.* (1985) propusieron reemplazar el operador proyección de Speckman (1988) por un suavizador de núcleos y utilizar el hecho que $Y_i - \mathbb{E}(Y_i|X_i) = \boldsymbol{\beta}^T(\mathbf{Z}_i - \mathbb{E}(\mathbf{Z}_i|X_i)) + \epsilon_i$ pues $\mathbb{E}(Y_i|X_i) = \boldsymbol{\beta}^T \mathbb{E}(\mathbf{Z}_i|X_i) + g(X_i)$. De esta forma, definieron los estimadores como

$$\begin{aligned} \widehat{\boldsymbol{\beta}} &= (\mathbf{Z}^T (\mathbf{I} - \mathbf{W}_h) \mathbf{Z})^{-1} \mathbf{Z}^T (\mathbf{I} - \mathbf{W}_h) \mathbf{Y}, \\ \widehat{g} &= \mathbf{W}_h (\mathbf{Y} - \mathbf{Z} \widehat{\boldsymbol{\beta}}), \end{aligned}$$

donde \mathbf{W}_h es la matriz asociada a un procedimiento de suavización por núcleos.

En Härdle (1984) y en Härdle *et al.* (2000), se puede encontrar una descripción más detallada de estos resultados así como diversos ejemplos en los que se aplican estos procedimientos.

3.2. Estimadores robustos

Dos de los métodos más comunes en regresión noparamétrica son el método de núcleos, descrito en la Sección 3.1 y el método de k -vecinos más cercanos con núcleos, introducidos por Nadaraya (1964) y Watson (1964) y por Collomb (1981), respectivamente. Ambos

son promedios pesados de las variables respuesta y por lo tanto, son altamente sensibles a fluctuaciones en los datos.

Un primer acercamiento a los estimadores robustos en el caso i.i.d. fue dado por Tsybakov (1982) y Härdle (1984), quienes estudiaron propiedades asintóticas puntuales de una versión robusta del método de Nadaraya–Watson cuando la escala es conocida. Más adelante, Härdle y Tsybakov (1988) extendieron sus resultados previos a M –estimadores equivariantes de escala mediante estimación simultánea de la función de regresión y de escala. Boente y Fraiman (1989) consideraron estimadores noparamétricos robustos equivariantes por escala usando pesos de vecinos más cercanos y pesos basados en núcleos aplicando un funcional de posición robusto al estimador de la función de distribución empírica condicional de las variables respuesta. Recordaremos su definición. Indiquemos por $\widehat{F}(y|\mathbf{X} = \mathbf{x})$ la distribución empírica condicional que se define como

$$\widehat{F}(y|\mathbf{X} = \mathbf{x}) = \sum_{i=1}^n w_i(\mathbf{x}) \mathbb{I}_{(-\infty, y]}(Y_i), \quad (3.3)$$

donde $w_i(\mathbf{x})$ son los pesos basados en núcleos tomando la misma ventana para todas las componentes

$$w_i(\mathbf{x}) = \mathcal{K} \left(\frac{\mathbf{X}_i - \mathbf{x}}{h_n} \right) \left\{ \sum_{j=1}^n \mathcal{K} \left(\frac{\mathbf{X}_j - \mathbf{x}}{h_n} \right) \right\}^{-1}. \quad (3.4)$$

$\widehat{F}(y|\mathbf{X} = \mathbf{x})$ provee un estimador de la distribución de $Y|\mathbf{X} = \mathbf{x}$ que indicaremos por $F(y|\mathbf{X} = \mathbf{x})$. El M –estimador local, $\widehat{r}_M(\mathbf{x})$ se define como el M –estimador de posición asociado a $\widehat{F}(y|\mathbf{X} = \mathbf{x})$, es decir, es la solución \widehat{r}_M de

$$\sum_{i=1}^n w_i(\mathbf{x}) \psi \left(\frac{Y_i - \widehat{r}_M(\mathbf{x})}{\widehat{s}(\mathbf{x})} \right) = 0, \quad (3.5)$$

donde $w_i(\mathbf{x})$ están dados en (3.4), ψ es una función impar, acotada y continua y $\widehat{s}(\mathbf{x})$ es un estimador de escala robusto. Elecciones posibles de la función de escores ψ son la funciones de Huber o la función bicuadrada de Tukey, mientras que la escala $\widehat{s}(\mathbf{x})$ puede tomarse como la mediana local de los desvíos absolutos respecto de la mediana local (MAD local), es decir, la MAD (Huber (1981)) respecto de la distribución $\widehat{F}(y|\mathbf{X} = \mathbf{x})$ definida en (3.3). Observemos que la mediana local corresponde a elegir $\psi(t) = \text{sg}(t)$. Las propiedades de consistencia y distribución asintótica fueron estudiadas en Boente y Fraiman (1989) para el caso de observaciones independientes, mientras que ordenes de convergencia fuerte pueden verse en Boente y Fraiman (1990, 1991). El caso particular de las medianas locales fue estudiado en Boente y Fraiman (1995). Por otra parte, Boente y Fraiman (1992) consideraron L –estimadores locales aplicando un funcional asociado a un L –estimador de posición a la función de distribución empírica $\widehat{F}(y|\mathbf{X} = \mathbf{x})$. Una propuesta basada en la norma L^1 puede verse en Wang y Scott (1994) quienes también

consideraron un procedimiento para elegir la ventana h . Es bien conocido el efecto que la ventana h tiene en la estimación de r . Ventanas grandes producen estimadores con varianza pequeña pero alto sesgo, mientras que valores pequeños de h producen curvas más irregulares. Este compromiso entre sesgo y varianza llevó a varias propuestas para elegir la ventana, tales como convalidación cruzada y métodos plug-in. Para el modelo de regresión no paramétrica (3.1), la sensibilidad de las ventanas clásicas a datos atípicos fue descrita por varios autores tales como Leung *et al.* (1993), Wang y Scott (1994), Boente *et al.* (1997), Cantoni y Ronchetti (2001) y Leung (2005).

Por otra parte, es sabido que, tanto en regresión lineal como en regresión no paramétrica, los estimadores de mínimos cuadrados pueden ser seriamente afectados por la presencia de datos anómalos. Lo mismo ocurre para el modelo parcialmente lineal (3.2). Por ese motivo, han surgido en los últimos años procedimientos robustos para este tipo de modelos. En este contexto, Bianco y Boente (2004) propusieron estimadores robustos mediante un procedimiento en tres pasos para el modelo de regresión parcialmente lineal. En ese trabajo, establecieron resultados de consistencia de los estimadores propuestos y también estudiaron la distribución asintótica de los estimadores del parámetro de regresión. Los estimadores considerados en Bianco y Boente (2004) pueden describirse como sigue. Supongamos que se cumple el modelo (3.2) y que $Z_{ij} = \phi_j(X_i) + \eta_{ij}$ $1 \leq j \leq p$, donde η_{ij} son independientes, independientes de X_i y tales que $\eta_{ij} \sim G_j(\cdot/\sigma_{\eta,j})$ con G_j simétrica respecto de 0. Más aún, supongamos que la distribución F_0 de ϵ_1 es simétrica respecto de 0. Indiquemos por $\phi_0(x) = \boldsymbol{\phi}(x)^T \boldsymbol{\beta} + g(x)$ de modo tal que $Y_i = \phi_0(X_i) + \nu_i$ con $\nu_i = \boldsymbol{\eta}_i^T \boldsymbol{\beta} + \epsilon_i$. Los estimadores robustos definidos en Bianco y Boente (2004) pueden describirse como

Paso 1: Estime $\phi_0(t)$ y $\phi_j(t)$, $1 \leq j \leq p$ mediante un estimador robusto, tal como el M -estimador local definido en (3.5). Sean $\hat{\phi}_{0,R}$ and $\hat{\phi}_{j,R}$ los estimadores resultantes y $\hat{\boldsymbol{\phi}}_R(x) = (\hat{\phi}_{1,R}(x), \dots, \hat{\phi}_{p,R}(x))^T$.

Paso 2: Estime el parámetro de regresión aplicando un estimador robusto de regresión, como el descrito en (2.3), a los residuos $\hat{r}_i = Y_i - \hat{\phi}_{0,R}(X_i)$ and $\hat{\boldsymbol{\eta}}_i = \mathbf{Z}_i - \hat{\boldsymbol{\phi}}_R(X_i)$. Sea $\hat{\boldsymbol{\beta}}_R$ el estimador de $\boldsymbol{\beta}$.

Paso 3: Defina el estimador de la función g como $\hat{g}_R(x) = \hat{\phi}_{0,R}(x) - \hat{\boldsymbol{\beta}}_R^T \hat{\boldsymbol{\phi}}_R(x)$.

Métodos robustos de selección de la ventana para estos modelos fueron estudiados por Boente y Rodriguez (2008).

Capítulo 4

Regresión noparamétrica funcional

En este capítulo, estudiaremos métodos de estimación noparamétrica de la función de regresión en el caso funcional, o sea, en el caso en que las covariables se encuentran en un espacio de dimensión infinita. El problema de estimar la función de regresión tiene gran interés y constituye una de las herramientas más importantes para abordar problemas de predicción noparamétricos. La literatura existente es muy variada cuando los datos son reales (ver, por ejemplo, Collomb y Härdle, 1986, Boente y Fraiman, 1989,1990). Sin embargo, la literatura es más reducida en el caso funcional. Cadre (2001) estudió procedimientos de estimación para la mediana L^1 de una variable aleatoria con valores sobre un espacio de Banach. Ferraty y Vieu (2004) consideraron modelos noparamétricos para datos funcionales con aplicaciones a regresión. Por otra parte, utilizando también un enfoque noparamétrico con núcleos, Ferraty *et al.* (2006) consideraron el problema de predicción estudiando el comportamiento asintótico del estimador e incluyendo convergencia con tasas así como su distribución asintótica.

Recientemente, Ferraty *et al.* (2010) obtuvieron resultados de convergencia uniforme casi completa (con tasas) de estimadores con núcleo para modelos noparamétricos que incluían la función de regresión y la distribución condicional acumulada, entre otros, bajo hipótesis que involucran, por ejemplo, la entropía de Kolmogorov y la propiedad de concentración de la pequeña bola de la variable funcional. En Ferraty y Vieu (2006) se pueden encontrar también referencias y resultados importantes sobre el problema de predicción en el caso noparamétrico funcional. Azzedine *et al.* (2008) estudiaron métodos de estimación noparamétrica robusta basado en la clase de M -estimadores, introducido por Huber (1964), cuando la escala es conocida.

Nosotros daremos aquí resultados de convergencia fuerte uniforme de M -estimadores locales, cuando se estima la función de escala, para el caso infinito dimensional extendiendo resultados previos en dos sentidos. Por un lado, se extienden los resultados obtenidos en Boente y Fraiman (1989, 1991) del caso euclídeo al caso funcional y por otra parte, se extienden los resultados dados por Ferraty *et al.* (2010) de convergencia sobre compactos de la distribución empírica condicional, bajo hipótesis similares, lo que permite obtener

tasas de convergencia fuerte uniforme sobre compactos del estimador de la función de regresión cuando se tiene un estimador preliminar consistente de la función de escala.

4.1. El espacio y la semimétrica

Consideremos el modelo de regresión funcional $Y = r(X) + \epsilon$, con $Y \in \mathbb{R}$ y ϵ un escalar aleatorio tal que $\mathbb{E}(\epsilon|X) = 0$ *a.s.*. Uno podría pensar que la curva aleatoria X pertenece a un conjunto apropiado de funciones, como por ejemplo, el conjunto $\mathcal{H} = L^2(\mathcal{J})$ de funciones de cuadrado integrable sobre el intervalo \mathcal{J} acotado, por lo que ahora $r : \mathcal{H} \rightarrow \mathbb{R}$ es un operador de $L^2(\mathcal{J})$ en \mathbb{R} que satisface algunas condiciones de regularidad.

Una distancia natural entre dos elementos X_1 y X_2 de $L^2(\mathcal{J})$ podría obtenerse en términos de la norma usual en $L^2(\mathcal{J})$ de su diferencia

$$\|X_1 - X_2\|_2 = \left(\int_{\mathcal{J}} (X_1(t) - X_2(t))^2 dt \right)^{1/2},$$

y luego una extensión del estimador de Nadaraya Watson al caso funcional con esta distancia nos quedaría

$$\hat{r}(x) = \frac{\sum_{i=1}^n K(\|x - X_i\|_2/h) Y_i}{\sum_{i=1}^n K(\|x - X_i\|_2/h)}.$$

Hemos mencionado, en el caso finito-dimensional, que a medida que crece la dimensión las bolas se hacen cada vez más ralas. Lo mismo ocurre en $L^2(\mathcal{J})$ con la norma $\|\cdot\|_2$ ya que para un x fijo, hay muy pocos elementos X_i de la muestra suficientemente cerca de x en el sentido de la norma $\|\cdot\|_2$ aún cuando el tamaño de la muestra tiende a infinito, resultando insuficientes para tener una estimación confiable de $r(x)$. Es por eso que interesa contar con una medida de proximidad que permita encontrar más curvas cercanas a una dada, y por lo tanto, más observaciones para el cálculo del estimador $\hat{r}(x)$.

Tomemos un ejemplo utilizado en Geenens (2011) para aclarar la situación. Si consideramos dos funciones $X_1(t)$ y $X_2(t) = X_1(t) + C$ con $C \in \mathbb{R}$ constante, nos encontramos con dos funciones que son idénticas salvo traslaciones. En este caso, la norma $\|\cdot\|_2$ se centra sólo en esa traslación y concluye que son totalmente diferentes, cosa acertada si lo importante son los valores que toman las funciones, pero si ese no es el caso (como por ejemplo, si sólo la forma de las funciones está relacionada a la respuesta Y) puede llevarnos a conclusiones erróneas. Una forma de medir la proximidad entre ambas funciones si sólo nos interesa la forma sería considerar

$$d_1(X_1, X_2) = \left(\int_{\mathcal{J}} (X_1'(t) - X_2'(t))^2 dt \right)^{1/2}.$$

En este caso tendríamos dos elementos distintos X_1 y X_2 con $d_1(X_1, X_2) = 0$, es decir, d_1 define una semi-métrica.

Por esta razón, resulta más interesante trabajar con espacios funcionales semimétricos que con los espacios normados usuales. Obviamente, la elección de la semimétrica juega un rol crucial ya que debe extraer toda la información necesaria para relacionar la respuesta con el predictor funcional. En el trabajo previamente mencionado se puede encontrar un análisis detallado de este tema. Por otra parte, en Ferraty y Vieu (2006) y Ferraty *et al.* (2010) se dan ejemplos que muestran como la elección de la semimétrica permite resolver el problema de la *maldición de la dimensión* que es aún peor en espacios de dimensión infinita.

4.2. Estimadores clásicos

Sean $\{(Y_i, X_i) : i = 1, \dots, n\}$ elementos aleatorios independientes con la misma distribución que (Y, X) y que toman valores en $\mathbb{R} \times \mathcal{H}$, donde (\mathcal{H}, d) es un espacio funcional semimétrico (es decir, X es una variable aleatoria funcional y d es una semi-métrica sobre \mathcal{H}). Supongamos que $\mathbb{E}|Y| < \infty$. En este caso, el operador de regresión (no lineal) r de Y sobre X está definido por

$$r(x) = \mathbb{E}(Y|X = x),$$

El modelo de regresión no paramétrico funcional puede escribirse como

$$Y_i = r(X_i) + \epsilon_i, \quad i = 1, \dots, n, \quad (4.1)$$

donde $r : \mathcal{H} \rightarrow \mathbb{R}$ es la función de regresión que, en principio, no tiene porqué ser lineal y $\epsilon_i \in \mathbb{R}$ es una variable aleatoria real. En el caso clásico, se supone que $\mathbb{E}(\epsilon_i|X_i) = 0$, $0 < \mathbb{E}(\epsilon_i^2|X_i = X) = \sigma^2(X) < \infty$. Más aún, si $\epsilon_i = \sigma(X_i)u_i$, en algunas situaciones se suele suponer que u_i es independiente de X_i .

Por otra parte, la función de distribución acumulada condicional (c.d.f.) de Y dado $X = x$ está definida, para cada $x \in \mathcal{H}$ por

$$\forall y \in \mathbb{R}, \quad F_Y^X(x, y) = \mathbb{P}(Y \leq y|X = x).$$

Con esta definición, se está asumiendo implícitamente que existe una versión regular de esta distribución condicional. Discutir condiciones sobre dicha existencia está más allá del ámbito de esta tesis. Sólo mencionaremos que si d es una métrica, la existencia queda asegurada bajo condiciones de separabilidad, ver por ejemplo, Jirina (1984) y Tortrat (1963). En el caso de espacios más generales, condiciones que implican la existencia de una versión regular es aún tema de estudio, para algunos resultados y referencias, recomendamos ver Faden (1985) y Leao *et al.* (2004).

La mediana $m(x)$ de la función de distribución condicional F_Y^X se define como

$$m(x) = \inf \left\{ y \in \mathbb{R} : F_Y^X(x, y) \geq \frac{1}{2} \right\}.$$

Sea $S_{cdf}^X = \{f : \mathcal{H} \times \mathbb{R} \rightarrow \mathbb{R} : f(x, \cdot) \text{ es una función de distribución estrictamente creciente para todo } x\}$.

La condición $F_Y^X \in S_{cdf}^X$ asegura la existencia y unicidad de la mediana condicional que, en ese caso, puede definirse como $m(x) = (F_Y^x)^{-1}(1/2)$, donde $F_Y^x : \mathbb{R} \rightarrow [0, 1]$ está definida por $F_Y^x(y) = F_Y^X(x, y)$.

4.2.1. Estimación de la función de regresión

La estimación de la función de regresión r definida en (4.1) juega un rol crucial en muchos problemas teóricos y aplicados. A continuación, enunciaremos algunos resultados existentes sobre la estimación en este tipo de modelos, consistencia y tasas de convergencia.

El estimador clásico de la función r es el estimador con núcleos funcional definido por

$$\hat{r}(x) = \frac{\sum_{i=1}^n K\left(\frac{d(x, X_i)}{h}\right) Y_i}{\sum_{i=1}^n K\left(\frac{d(x, X_i)}{h}\right)}, \quad (4.2)$$

donde K es una función núcleo y $h = h_n$ una sucesión de números reales estrictamente positiva. El estimador (4.2) es la versión funcional del estimador de Nadaraya Watson introducida previamente para regresión no paramétrica finito dimensional (ver, por ejemplo, Härdle, 1990). Claramente, $\hat{r}(x) = \sum_{i=1}^n w_i(x) Y_i$, donde

$$w_i(x) = \frac{K\left(\frac{d(x, X_i)}{h}\right)}{\sum_{i=1}^n K\left(\frac{d(x, X_i)}{h}\right)}. \quad (4.3)$$

Observemos que los pesos $w_i(x)$ satisfacen $\sum_{i=1}^n w_i(x) = 1$.

Ferraty y Vieu (2006) han obtenido, bajo condiciones de regularidad, resultados de convergencia y tasas de convergencia del estimador $\hat{r}(x)$. Más precisamente, estos autores probaron que $\hat{r}(x) \xrightarrow{a.co.} r(x)$ y que $\hat{r}(x) - r(x) = O(h^\eta) + O_{a.co.}\left(\frac{\sqrt{\log(n)/n\phi_x(h)}}{h}\right)$ donde $\phi_x(h) = \mathbb{P}(X \in B(x, h))$, $\xrightarrow{a.co.}$ y $a.co.$ indican convergencia casi completamente.

Además, Ferraty *et al.* (2010), obtuvieron resultados de consistencia uniforme con tasas para la estimación de la función de regresión generalizada

$$r_\varphi(x) = \mathbb{E}(\varphi(Y)|X = x),$$

donde φ es una función medible Borel a valores reales conocida. El estimador de $r_\varphi(x)$ está dado por $\widehat{r}_\varphi(x) = \sum_{i=1}^n w_i(x)\varphi(Y_i)$. Bajo condiciones de regularidad, Ferraty *et al.* (2010) probaron la convergencia uniforme casi completa de \widehat{r}_φ sobre algún subconjunto compacto $S_{\mathcal{H}}$ de \mathcal{H} . Más aún, dicha convergencia se relaciona con la ϵ -entropía de Kolmogorov $\psi_{S_{\mathcal{H}}}(\epsilon)$ y la función ϕ que controla la probabilidad de la pequeña bola de la variable funcional X , de la siguiente forma

$$\sup_{x \in S_{\mathcal{H}}} |\widehat{r}_\varphi(x) - r_\varphi(x)| = O(h^\eta) + O_{a.co.} \left(\sqrt{\frac{\psi_{S_{\mathcal{H}}}(\log(n)/n)}{n\phi(h)}} \right).$$

4.2.2. Estimación de la función de distribución acumulada condicional

Existen varios estimadores de la función de distribución condicional F_Y^X . Ferraty *et al.* (2006) definen dos estimadores para $F_Y^X(x, y)$. Uno de ellos puede pensarse como un caso particular del estimador con núcleos de la función de regresión generalizada $r_\varphi(x)$ tomando como función $\varphi(Y) = \mathbb{I}_{(-\infty, y]}(Y)$. El otro estimador es un estimador de doble núcleo.

Utilizando la esperanza condicional, podemos redefinir a F_Y^X como

$$F_Y^X(x, y) = \mathbb{E}(\mathbb{I}_{(-\infty, y]}(Y) | X = x),$$

de donde es fácil definir estimadores de $F_Y^X(x, y)$. Por lo tanto, por analogía con el contexto de regresión funcional, el primer estimador considerado se define como un estimador con núcleos de la función de distribución condicional como

$$\widehat{F}_Y^X(x, y) = \frac{\sum_{i=1}^n K\left(\frac{d(x, X_i)}{h}\right) \mathbb{I}_{(-\infty, y]}(Y_i)}{\sum_{i=1}^n K\left(\frac{d(x, X_i)}{h}\right)} = \sum_{i=1}^n w_i(x) \mathbb{I}_{(-\infty, y]}(Y_i), \quad (4.4)$$

donde $w_i(x)$ son los pesos definidos en (4.3).

Ferraty *et al.* (2010) obtienen resultados de convergencia uniforme del estimador \widehat{F}_Y^X de F_Y^X sobre $S_{\mathcal{H}} \times S_{\mathbb{R}}$ donde $S_{\mathcal{H}} \subset \mathcal{H}$ y $S_{\mathbb{R}} \subset \mathbb{R}$ son conjuntos compactos. Para ser más precisos, estos autores muestran que

$$\begin{aligned} \sup_{x \in S_{\mathcal{H}}} |\widehat{F}_Y^x(y) - F_Y^x(y)| &= O(h^\eta) + O_{a.co.} \left(\sqrt{\frac{\psi_{S_{\mathcal{H}}}(\log(n)/n)}{n\phi(h)}} \right) \\ \sup_{x \in S_{\mathcal{H}}} \sup_{y \in S_{\mathbb{R}}} |\widehat{F}_Y^x(y) - F_Y^x(y)| &= O(h^{\eta_1}) + O_{a.co.} \left(\sqrt{\frac{\psi_{S_{\mathcal{H}}}(\log(n)/n)}{n\phi(h)}} \right), \end{aligned}$$

donde ϕ es como antes una función que controla la probabilidad de la pequeña bola de la variable funcional, $\psi_{S_{\mathcal{H}}}(\epsilon)$ es la ϵ -entropía de Kolmogorov del conjunto $S_{\mathcal{H}}$, η y η_1 son constantes positivas

En el segundo caso, el estimador con núcleos propuesto para F_Y^X se define como

$$\tilde{F}_Y^X(x, y) = \frac{\sum_{i=1}^n K\left(\frac{d(x, X_i)}{h}\right) H\left(\frac{y - Y_i}{k}\right)}{\sum_{j=1}^n K\left(\frac{d(x, X_j)}{h}\right)} = \sum_{i=1}^n w_i(x) H\left(\frac{y - Y_i}{k}\right),$$

donde $K : \mathbb{R} \rightarrow \mathbb{R}$ es un núcleo, $h = h_n$ y $k = k_n$ son sucesiones de números reales positivos y H es una función de distribución. Con hipótesis adaptadas a este contexto funcional, Ferraty *et al.* (2006) establecen resultados de convergencia de este estimador.

En este Capítulo trabajaremos con el estimador con núcleos de la función de distribución acumulada condicional, \hat{F}_Y^X , y obtendremos resultados de convergencia fuerte uniforme con tasas de ese estimador sobre $S_{\mathcal{H}}$ y \mathbb{R} , generalizando el resultado obtenido en Ferraty *et al.* (2010).

4.3. Estimadores noparamétricos robustos

Los métodos de estimación robustos proveen un enfoque alternativo al de los métodos clásicos cuando, por ejemplo, los datos se encuentran afectados por la presencia de datos atípicos o outliers. En los últimos años, se han empezado a considerar este tipo de métodos para modelos con datos funcionales. Azzedine *et al.* (2008) extendieron a dimensión infinita los métodos de estimación noparamétrica robusta de la función de regresión dados por Collomb y Härdle (1986) para el caso real. Los estimadores robustos considerados están basados en núcleos y para ellos establecieron la convergencia casi completa con tasas, bajo ciertas hipótesis que involucran la propiedad de concentración de la pequeña bola de la variable funcional, entre otras. En ese mismo contexto, Attouch *et al.* (2007) estudiaron la normalidad asintótica de esos estimadores. Cambres *et al.* (2008) estudiaron también estimadores robustos haciendo mayor hincapié en las propiedades asintóticas de los estimadores noparamétricos condicionales para una variable real con covariable funcional.

4.3.1. Definiciones y notación

Sea (Y, X) un vector aleatorio con la misma distribución que (Y_i, X_i) y denotemos por $g(x)$ a la única solución de $\lambda(x, a, s(x)) = 0$ donde $s(x)$ es una medida robusta de escala condicional y

$$\lambda(x, a, \sigma) = \mathbb{E} \left(\psi \left(\frac{Y - a}{\sigma} \right) | X = x \right), \quad (4.5)$$

con $\psi : \mathbb{R} \rightarrow \mathbb{R}$ una función estrictamente creciente, continua y acotada. La medida de escala condicional puede tomarse como la mediana de los desvíos absolutos de la distribución condicional F_Y^X , es decir,

$$s(x) = \text{mediana}(|Y - m(x)| | X = x) = \text{MAD}_C(F_Y^X(x, \cdot)) \quad (4.6)$$

donde $m(x) = \text{mediana}(Y | X = x)$ es la mediana de la distribución condicional. Observemos que si (Y, X) cumple el modelo (4.1) y los errores $\epsilon_i = \sigma(X_i)u_i$ donde u_i son independientes de X_i y tienen distribución simétrica, entonces $g = r$. La función $g(x)$ corresponde al funcional definido en Boente y Fraiman (1989) y la llamaremos función de regresión ya que coincide con r si existe esperanza.

Por otra parte, observemos que indicamos por $s(x)$ a la función de escala ya que hemos reservado el nombre de $\sigma(x)$ para el caso de la función de dispersión, o sea, la raíz cuadrada de la función de varianza.

Por simplicidad de notación, a partir de ahora, indicaremos por $F(y|X = x)$ a la función de distribución condicional de Y dado X y por $\hat{F}(y|X = x)$ a los estimadores basados en pesos con núcleos de la función de distribución condicional definidos en (4.4). Por otra parte, denotemos por $\hat{s}(x) = \text{MAD}_C(\hat{F}(\cdot|X = x))$, o sea, a la medida de escala definida en (4.6) evaluada en $\hat{F}(y|X = x)$. Con esta notación, el estimador noparamétrico robusto de $g(x)$ está dado por la solución $\hat{g}(x)$ de $\hat{\lambda}(x, a, \hat{s}(x)) = 0$, donde

$$\hat{\lambda}(x, a, \sigma) = \int \psi \left(\frac{y - a}{\sigma} \right) d\hat{F}(y|X = x) = \sum_{i=1}^n w_i(x) \psi \left(\frac{Y_i - a}{\sigma} \right). \quad (4.7)$$

Es decir, $\hat{g}(x)$ es la solución de la ecuación

$$\int \psi \left(\frac{y - \hat{g}(x)}{\hat{s}(x)} \right) d\hat{F}(y|X = x) = \sum_{i=1}^n w_i(x) \psi \left(\frac{Y_i - \hat{g}(x)}{\hat{s}(x)} \right) = 0.$$

4.3.2. Hipótesis

Sea $B(x, h)$ la bola de centro x y radio h para la topología asociada con la semi-métrica d , $B(x, h) = \{y \in \mathcal{H} : d(x, y) \leq h\}$. Como se indica en Ferraty y Vieu (2006), todos los resultados en estadística noparamétrica para variables funcionales están muy relacionados con las propiedades de concentración de la medida de probabilidad de la variable funcional, en este caso, X . Como nos interesa obtener resultados de convergencia uniforme, la hipótesis **H1** tendrá también en cuenta el aspecto uniforme.

De ahora en más, C y C' indicarán constantes genéricas estrictamente positivas, que pueden cambiar de renglón a renglón. A partir de ahora, sean $S_{\mathbb{R}} \subset \mathbb{R}$ y $S_{\mathcal{H}} \subset \mathcal{H}$ conjuntos compactos.

Consideraremos el siguiente conjunto de hipótesis:

H1 Para todo $x \in S_{\mathcal{H}}$ existe $\phi : \mathbb{R} \rightarrow \mathbb{R}_{>0}$ tal que $0 < C\phi(h) \leq \mathbb{P}(X \in B(x, h)) \leq C'\phi(h)$

H2 $F(y|X = x)$ es simétrica alrededor de $g(x)$.

H3 Sea $S_{\mathcal{H}}$ un conjunto compacto de \mathcal{H} tal que

- i) para cada y fijo $F(y|X = x)$ es una función uniformemente continua de x en un entorno de $S_{\mathcal{H}}$
- ii) se verifica la siguiente condición de equicontinuidad:

$$\forall \epsilon > 0 \exists \delta > 0 : |u - v| < \delta \Rightarrow \sup_{x \in S_{\mathcal{H}}} |F(u|X = x) - F(v|X = x)| < \epsilon .$$

H4 El núcleo K es una función no-negativa acotada con soporte $[0, 1]$ tal que $\int K(u)du = 1$ y satisface una condición de Lipschitz de orden uno. Además,

- a) Si $K(1) = 0$, K es derivable con derivada K' y $-\infty < \inf_{u \in \mathbb{R}} K'(u) \leq \sup_{u \in \mathbb{R}} K'(u) = \|K'\|_{\infty} < 0$.
- b) Si $K(1) > 0$, existen $C, C' > 0$ tales que $C\mathbb{I}_{[0,1]}(u) < K(u) < C'\mathbb{I}_{[0,1]}(u)$.

H5 Las funciones ϕ y $\psi_{S_{\mathcal{H}}}$ son tales que:

- a) $\phi : \mathbb{R} \rightarrow \mathbb{R}_{>0}$ es derivable con derivada ϕ' . Más aún, existen $C_{\phi} > 0$ y $\eta_0 > 0$, tales que para todo $\eta < \eta_0$, $\phi'(\eta) < C_{\phi}$.
Si $K(1) = 0$, la función ϕ debe satisfacer la siguiente condición adicional:

$$\exists C > 0, \exists \eta_0 > 0 \text{ tales que } \forall 0 < \eta < \eta_0 \int_0^{\eta} \phi(u)du > C\eta\phi(\eta) .$$

- b) para n suficientemente grande,

$$\frac{(\log(n))^2}{n\phi(h)} < \psi_{S_{\mathcal{H}}} \left(\frac{\log(n)}{n} \right) < \frac{n\phi(h)}{\log(n)} .$$

H6 $\psi : \mathbb{R} \rightarrow \mathbb{R}$ es una función impar, estrictamente creciente, acotada y continuamente diferenciable, con derivada ψ' tal que $u\psi'(u) \leq \psi(u)$.

H7 La sucesión $h = h_n$ es tal que $h_n \rightarrow 0$, $n\phi(h_n) \rightarrow \infty$ y $n\phi(h_n)/\log(n) \rightarrow \infty$ cuando $n \rightarrow \infty$.

H8 $F(y|X = x)$ tiene una única mediana $m(x)$.

H9 La función $F(y|X = x)$ es uniformemente Lipschitz en un entorno $S_{\mathcal{H}}^{\epsilon}$ de $S_{\mathcal{H}}$, es decir, existen constantes $D > 0$ y $\eta_1 > 0$ tales que, $x_1, x_2 \in S_{\mathcal{H}}^{\epsilon}$, se cumple

$$\sup_{y \in \mathbb{R}} |F(y|X = x_1) - F(y|X = x_2)| \leq D d^{\eta_1}(x_1, x_2) .$$

H10 La ϵ -entropía de Kolmogorov de $S_{\mathcal{H}}$ satisfice

$$\sum_{n=1}^{\infty} n^{\frac{1}{2}} \exp \left\{ (1 - \beta) \psi_{S_{\mathcal{H}}} \left(\frac{\log(n)}{n} \right) \right\} < \infty \quad \text{para algún } \beta > 1.$$

Observaciones:

- i) Las hipótesis **H1** a **H6**, se utilizarán para obtener la consistencia fuerte del M -estimador local. Cuando consideremos la mediana local, requeriremos **H8** en lugar de **H6**. La función bicuadrada de Tukey no es creciente como se requiere en **H6**. Sin embargo, la función $\lambda(x, a, s(x))$ tiene un único cambio de signo en un entorno de $g(x)$ por lo que, resultados de consistencia puntual se pueden obtener utilizando argumentos análogos a los considerados en el Teorema 2.1 de Boos y Serfling (1980). Por otra parte, la función ψ bicuadrada es la derivada de una función ρ , lo que garantiza Fisher-consistencia y consistencia fuerte puntual puede deducirse utilizando este hecho. Por otra parte, **H9** y **H10** se pedirán para obtener órdenes de convergencia fuerte.
- ii) La función ϕ definida en **H1** controla la concentración de la medida de probabilidad de la variable funcional sobre una bola pequeña. En regresión noparamétrica multivariada, o sea, si $X \in \mathbb{R}^q$, es usual suponer que la covariable X tiene densidad f_X acotada y estrictamente positiva sobre compactos. La condición $\phi(h) > 0$ pedida en **H1** es una generalización de la hipótesis de densidad positiva pedida en esta situación. En esta tesis, supondremos que $\phi(h) \rightarrow 0$ cuando $h \rightarrow 0$, lo cual es equivalente, por **H1**, a que $\mathbb{P}(X = x) = 0$ para todo $x \in \mathcal{H}$. En particular, Ferraty y Vieu (2006) consideran los así llamados procesos fractales de orden $\tau > 0$ respecto de la semimétrica d para los cuales $\phi(h) \sim C h^\tau$ para alguna constante $C > 0$.
- iii) **H3** i) implica que existen números reales a, b tales que, para todo $x \in S_{\mathcal{H}}$, $F(b|X = x) > 1 - \epsilon$ y $F(a|X = x) < \epsilon$.

Veamos que existe a tal que, para todo $x \in S_{\mathcal{H}}$, $F(a|X = x) < \epsilon$. Efectivamente, sea $x_0 \in S_{\mathcal{H}}$ fijo. Sabemos que dado $\epsilon > 0$, existe $a(x_0) \in \mathbb{R}$ tal que $F(a(x_0)|X = x_0) < \epsilon$. Como $F(y|X = x)$ es una función continua de x , también existe $\delta(x_0) > 0$ tal que si $d(x, x_0) < \delta(x_0)$ entonces $F(a(x_0)|X = x) < \epsilon$.

Como $S_{\mathcal{H}}$ es compacto, existen $x_1, \dots, x_m \in S_{\mathcal{H}}$ tales que $S_{\mathcal{H}} \subset \bigcup_{j=1}^m B(x_j, \delta(x_j))$. Sean $a(x_1), \dots, a(x_m) \in \mathbb{R}$ tales que $F(a(x_j)|X = x_j) < \epsilon$ para $1 \leq j \leq m$. Definamos $a = \min_{1 \leq j \leq m} (a(x_j))$. Dado $x \in S_{\mathcal{H}}$, existe $1 \leq j \leq m$ tal que $x \in B(x_j, \delta(x_j))$ de donde $F(a|X = x) \leq F(a(x_j)|X = x) < \epsilon$.

La otra desigualdad se demuestra análogamente.

iv) La función $\phi(h)$ es decreciente en h , luego, la condición de acotación de la derivada de ϕ , dada en **H5a**), nos permite considerar a ϕ como una función Lipschitz alrededor de cero. Por otra parte, por **H5a**) $\phi(h) \leq Ch$, luego **H7** implica que $nh/\log(n) \rightarrow \infty$.

En el caso finito-dimensional, o sea, si $X \in \mathbb{R}^q$, la función $\phi(h) = h^q$ si la densidad es positiva y acotada sobre compactos, con lo que se cumple **H5a**). Por otra parte, la hipótesis **H5b**) extiende los requerimientos pedidos a la ventana del caso finito-dimensional al caso funcional. Finalmente, como mencionan Ferraty y Vieu (2006), la noción de la pequeña bola está relacionada fuertemente con la semi-métrica d . En el capítulo 13 de Ferraty y Vieu (2006), se dan varios ejemplos de semi-métricas d para las cuales la probabilidad de la pequeña bola y por lo tanto, $\phi(h)$, puede calcularse o al menos ser evaluadas asintóticamente cuando $h \rightarrow 0$, de modo a dar ejemplos que cumplan **H5**.

v) **H5b**) implica que $\log(n)/(n\phi(h)) \rightarrow 0$ cuando $n \rightarrow \infty$.

Más aún, $\psi_{S_{\mathcal{X}}}(\log(n)/n) (n\phi(h))^{-1} \rightarrow 0$. Como se menciona en Ferraty *et al.* (2010), en muchos casos, $\psi_{S_{\mathcal{X}}}(\log(n)/n) \sim C \log(n)$ luego, **H5b**) se cumple si $C(\log(n))^2 < n\phi(h)$.

vi) Supongamos, por ejemplo, que el proceso es fractal de orden τ , es decir, que $\phi(h) \sim h^\tau$ cuando $h \rightarrow 0$. Tomemos $h = n^{-\alpha}$, luego, **H7** se cumple si $\alpha < 1/\tau$ mientras que **H5b**) dice que

$$\frac{(\log(n))^2}{n^{1-\alpha\tau}} < \psi_{S_{\mathcal{X}}}\left(\frac{\log(n)}{n}\right) < \frac{n^{1-\alpha\tau}}{\log(n)}.$$

Si además, $\psi_{S_{\mathcal{X}}}(\log(n)/n) \sim C \log(n)$ y $\alpha < 1/\tau$ **H5b**) se cumplirá.

vii) Los resultados de convergencia uniforme con tasas que obtendremos mostrarán que $\widehat{g}(x) - g(x)$ tiene tasa uniforme casi completa $h^{m_1} + \theta_n$ donde

$$\theta_n^2 = \frac{\psi_{S_{\mathcal{X}}}\left(\frac{\log(n)}{n}\right)}{n\phi(h)}.$$

Supongamos que $h = n^{-\alpha}$ y que el proceso es fractal de orden τ . En aquellas situaciones en las que, como se menciona en Ferraty *et al.* (2010), $\psi_{S_{\mathcal{X}}}(\epsilon) \leq C \log(1/\epsilon)$, si $\alpha < 1/(2\tau)$ tendremos que $n^{1/4}\theta_n \rightarrow 0$. Por otra parte, si $1/(4\eta_1) < \alpha$ entonces $n^{1/4}h^{m_1} \rightarrow 0$, es decir, eligiendo $1/(4\eta_1) < \alpha < 1/(2\tau)$, tendremos que $n^{1/4} \sup_{x \in S_{\mathcal{X}}} |\widehat{g}(x) - g(x)| \xrightarrow{a.co.} 0$.

Observemos que, en este caso, **H10**, se cumplirá si β es suficientemente grande.

viii) Hemos mencionado que si (Y, X) cumple el modelo (4.1) y los errores $\epsilon_i = \sigma(X_i)u_i$ donde u_i son independientes de X_i y tienen distribución simétrica, entonces $g = r$.

Por otra parte, si la distribución F_0 de u_1 tiene única mediana en 0 entonces se cumple **H8**.

Queremos ver que condiciones sobre la distribución de los errores y sobre la función de regresión establecen las hipótesis **H3** y **H9**. Observemos que en el caso que estamos considerando $F(y|X = x) = F_0((y - r(x))/\sigma(x))$ y como F_0 es simétrica alrededor de 0, $g(x) = r(x)$. Luego, si $F_0 : \mathbb{R} \rightarrow [0, 1]$ es continua y si $r : \mathcal{H} \rightarrow \mathbb{R}$ y $\sigma : \mathcal{H} \rightarrow \mathbb{R}$ son funciones uniformemente continuas en un entorno de $S_{\mathcal{H}}$, se cumple **H3i**). Más aún, si $\inf_{x \in S_{\mathcal{H}}} \sigma(x) > 0$ y F_0 es Lipschitz entonces se cumple **H3ii**). Por otra parte, si $\sigma(x) = \sigma > 0$ para todo x , es decir, en el caso homocedástico, la condición **H9** se cumple si F_0 es Lipschitz y r es Lipschitz de orden η_1 , es decir, $|r(x_1) - r(x_2)| \leq C d^{\eta_1}(x_1, x_2)$. En el caso heteroscedástico, necesitamos los requerimientos adicionales $\inf_{x \in S_{\mathcal{H}}} \sigma(x) > 0$ y σ es Lipschitz de orden η_1 , es decir, $|\sigma(x_1) - \sigma(x_2)| \leq C d^{\eta_1}(x_1, x_2)$ y además F_0 debe tener densidad f_0 tal que $y f_0(y)$ es acotada.

4.3.3. Notación y resultados preliminares

En las secciones 4.3.4 y 4.4 obtendremos resultados de convergencia uniforme y tasas de convergencia uniforme sobre compactos para los M -estimadores locales. Para ello, en los Teoremas 4.3.4 y 4.4.1 obtendremos resultados análogos para la función de distribución empírica condicional $\widehat{F}(y|X = x)$, es decir, mostraremos, respectivamente, que

$$\sup_{x \in S_{\mathcal{H}}} \sup_{y \in \mathbb{R}} |\widehat{F}(y|X = x) - F(y|X = x)| \xrightarrow{a.s.} 0 \quad (4.8)$$

y que

$$\sup_{x \in S_{\mathcal{H}}} \sup_{y \in \mathbb{R}} |\widehat{F}(y|X = x) - F(y|X = x)| = O_{a.co.}(h^{\eta_1} + \theta_n) \quad (4.9)$$

donde

$$\theta_n = \sqrt{\frac{\psi_{S_{\mathcal{H}}}(\log(n)/n)}{n\phi(h)}} \rightarrow 0.$$

Para probar esos resultados, introduciremos algo de notación y enunciaremos unos resultados previos. Dado $y \in \mathbb{R}$ fijo denotamos por

$$\begin{aligned} W_i &= \mathbb{I}_{(-\infty, y]}(Y_i) \\ K_i(x) &= K\left(\frac{d(x, X_i)}{h}\right) \\ \widetilde{R}_j(x) &= \frac{1}{n} \sum_{i=1}^n W_i^j \frac{K_i(x)}{\mathbb{E}K_1(x)} \\ R_j(x) &= \frac{1}{n} \sum_{i=1}^n W_i^j \frac{K_i(x)}{\phi(h)} \end{aligned} \quad (4.10)$$

para $j = 0, 1$. Entonces $\widehat{F}(y|X = x) = \widetilde{R}_1(x)/\widetilde{R}_0(x)$.

Como en Collomb (1982), tenemos que

$$\begin{aligned} \sup_{x \in S_{\mathcal{H}}} |\widehat{F}(y|X = x) - F(y|X = x)| &\leq \frac{1}{\inf_{x \in S_{\mathcal{H}}} \widetilde{R}_0(x)} \left[\sup_{x \in S_{\mathcal{H}}} |\widetilde{R}_1(x) - \mathbb{E}\widetilde{R}_1(x)| \right. \\ &\quad \left. + \sup_{x \in S_{\mathcal{H}}} |\widetilde{R}_0(x) - \mathbb{E}\widetilde{R}_0(x)| + \sup_{x \in S_{\mathcal{H}}} |\mathbb{E}\widetilde{R}_1(x) - F(y|X = x)\mathbb{E}\widetilde{R}_0(x)| \right]. \end{aligned} \quad (4.11)$$

Para probar (4.8), mostraremos que para cada $y \in \mathbb{R}$

$$\sup_{x \in S_{\mathcal{H}}} |\widetilde{R}_j(x) - \mathbb{E}\widetilde{R}_j(x)| \xrightarrow{a.co.} 0 \quad (4.12)$$

$$\sup_{x \in S_{\mathcal{H}}} |\mathbb{E}\widetilde{R}_1(x) - F(y|X = x)\mathbb{E}\widetilde{R}_0(x)| \rightarrow 0. \quad (4.13)$$

y que para algún $a > 0$

$$\sum_{n \geq 1} \mathbb{P}(\inf_{x \in S_{\mathcal{H}}} \widetilde{R}_0(x) < a) < \infty, \quad (4.14)$$

Por otra parte, para probar (4.9), mostraremos además

$$\sup_{y \in \mathbb{R}} \sup_{x \in S_{\mathcal{H}}} |\widetilde{R}_j(x) - \mathbb{E}\widetilde{R}_j(x)| = O_{a.co.}(\theta_n) \quad (4.15)$$

$$\sup_{y \in \mathbb{R}} \sup_{x \in S_{\mathcal{H}}} |\mathbb{E}\widetilde{R}_1(x) - F(y|X = x)\mathbb{E}\widetilde{R}_0(x)| = O(h^m). \quad (4.16)$$

Observemos que los Lemas 4.3 y 4.4 de Ferraty y Vieu (2006), **H1**, **H4** y **H5** implican que existen constantes $0 < C < C' < \infty$ tales que

$$\forall x \in S_{\mathcal{H}}, \quad C\phi(h) < \mathbb{E}K_1(x) < C'\phi(h). \quad (4.17)$$

Luego, si $\widetilde{C} = 1/C$, obtenemos que

$$|\widetilde{R}_j(x) - \mathbb{E}\widetilde{R}_j(x)| \leq \widetilde{C} |R_j(x) - \mathbb{E}R_j(x)|.$$

Por lo tanto, para obtener (4.12) será suficiente que mostremos que

$$\sup_{x \in S_{\mathcal{H}}} |R_j(x) - \mathbb{E}R_j(x)| \xrightarrow{a.co.} 0 \quad (4.18)$$

mientras que para probar (4.15) bastará probar

$$\sup_{y \in \mathbb{R}} \sup_{x \in S_{\mathcal{H}}} |R_j(x) - \mathbb{E}R_j(x)| = O_{a.co.}(\theta_n).$$

Observemos que como

$$\inf_{x \in S_{\mathcal{H}}} \tilde{R}_0(x) \geq \inf_{x \in S_{\mathcal{H}}} \mathbb{E} \tilde{R}_0(x) - \sup_{x \in S_{\mathcal{H}}} |\tilde{R}_0(x) - \mathbb{E} \tilde{R}_0(x)|,$$

el hecho que $\mathbb{E} \tilde{R}_0(x) = 1$, la acotación $|\tilde{R}_j(x) - \mathbb{E} \tilde{R}_j(x)| \leq \tilde{C} |R_j(x) - \mathbb{E} R_j(x)|$ y (4.18) implican (4.14).

Utilizaremos los siguientes lemas, el primero corresponde a la desigualdad exponencial enunciada, por ejemplo, en Corolario A.9 de Ferraty y Vieu (2006). La demostración de los Lemas 4.3.2 y 4.3.3 pueden encontrarse en el Apéndice.

Lema 4.3.1. *Sea $\{Z_i\}$ una sucesión de variables aleatorias independientes, $\mathbb{E}(Z_i) = 0$. Si existe $M < \infty$ tal que $|Z_1| \leq M$ y si $\sigma^2 = \mathbb{E} Z_1^2$, entonces para todo $\epsilon > 0$ se tiene que*

$$\mathbb{P} \left(\left| \sum_{i=1}^n Z_i \right| > \epsilon n \right) \leq 2 \exp \left\{ -\frac{\epsilon^2 n}{2\sigma^2 \left(1 + \epsilon \frac{M}{\sigma^2}\right)} \right\}.$$

Lema 4.3.2. *Sea $S_{\mathcal{H}} \subset \mathcal{H}$ un conjunto compacto. Bajo las hipótesis **H1**, **H4** y **H5**, se tiene que, para $j = 0, 1$,*

a) *para todo $n \geq n_0$ y para todo $\epsilon > 0$ fijo*

$$\sup_{y \in \mathbb{R}} \sup_{x \in S_{\mathcal{H}}} \mathbb{P} \{ |R_j(x) - \mathbb{E} R_j(x)| > \epsilon \} \leq 2 \exp \left\{ -\frac{\epsilon^2 n \phi(h)}{2C' \|K\|_{\infty}^2 \left(1 + \frac{\epsilon}{C' \|K\|_{\infty}}\right)} \right\}.$$

b) *existen $a_1 > 0$, $a_2 > 0$ tales que, para todo $n \geq n_0$ y para todo $\epsilon > 0$ fijo*

$$\mathbb{P} \left\{ \sup_{x \in S_{\mathcal{H}}} |R_j(x) - \mathbb{E} R_j(x)| > \epsilon \right\} \leq 8N_{\rho}(S_{\mathcal{H}}) \exp \left\{ -\frac{\epsilon^2 n \phi(h)}{a_1(1 + a_2 \epsilon)} \right\},$$

donde ρ es tal que $\rho/h \rightarrow 0$ y $\rho/\phi(h) \rightarrow 0$ cuando $n \rightarrow \infty$.

c) *existe $c > 2$ tal que, para todo $\epsilon_0 > c$ y $n \geq n_0$, se cumple*

$$\sup_{y \in \mathbb{R}} \mathbb{P} \left\{ \theta_n^{-1} \sup_{x \in S_{\mathcal{H}}} |R_j(x) - \mathbb{E} R_j(x)| > \epsilon_0 \right\} \leq 8 \exp \left\{ \left(1 - \frac{\epsilon_0^2}{8(1 + \epsilon_0)}\right) \psi_{S_{\mathcal{H}}} \left(\frac{\log(n)}{n} \right) \right\},$$

Vale la pena observar que (c) vale también con tasa $(h^n + \theta_n)$.

Lema 4.3.3. Sea $\tilde{R}_j(x)$ definidos en (4.10) para $j = 0, 1$. Bajo las hipótesis **H1** y **H4**, si $h_n \rightarrow 0$, se tiene que

$$a) \sup_{y \in \mathbb{R}} \sup_{x \in S_{\mathcal{H}}} |\mathbb{E}\tilde{R}_1(x) - F(y|X=x)\mathbb{E}\tilde{R}_0(x)| \rightarrow 0 .$$

b) si además se cumple **H9** entonces

$$\sup_{y \in \mathbb{R}} \sup_{x \in S_{\mathcal{H}}} |\mathbb{E}\tilde{R}_1(x) - F(y|X=x)\mathbb{E}\tilde{R}_0(x)| = O(h^n) .$$

Tenemos entonces el siguiente Corolario.

Corolario. Bajo las hipótesis del Lema 4.3.3b),

$$\sup_{y \in \mathbb{R}} \sup_{x \in S_{\mathcal{H}}} |\mathbb{E}\tilde{R}_1(x) - F(y|X=x)\mathbb{E}\tilde{R}_0(x)| = O(h^n + \theta_n) .$$

4.3.4. Convergencia fuerte uniforme

En esta Sección, estableceremos resultados de convergencia uniforme sobre compactos del estimador \hat{g} de la función de regresión g .

El siguiente Teorema generaliza el resultado obtenido en el Lemma 6.5 de Ferraty y Vieu (2006) y es el equivalente funcional del Teorema 3.1 de Boente y Fraiman (1991).

Teorema 4.3.4. Sea $S_{\mathcal{H}} \subset \mathcal{H}$ compacto. Bajo las hipótesis **H1**, **H3**, **H4**, **H5** y **H7** se tiene que

$$\sup_{x \in S_{\mathcal{H}}} \sup_{y \in \mathbb{R}} |\hat{F}(y|X=x) - F(y|X=x)| \xrightarrow{a.s.} 0$$

cuando $n \rightarrow \infty$.

DEMOSTRACIÓN. Por el Lema 4.3.2b) tomando $\rho_n = \log(n)/n$ obtenemos que para cada $\epsilon > 0$

$$\begin{aligned} \mathbb{P} \left\{ \sup_{x \in S_{\mathcal{H}}} |R_j(x) - \mathbb{E}R_j(x)| > \epsilon \right\} &\leq 8 \exp \left\{ \psi_{S_{\mathcal{H}}}(\rho_n) - \frac{\epsilon^2 n \phi(h)}{a_1(1+a_2\epsilon)} \right\} \\ &\leq 8 \exp \left\{ -n\phi(h) \left(\frac{\epsilon^2}{a_1(1+a_2\epsilon)} - \frac{\psi_{S_{\mathcal{H}}}(\rho_n)}{n\phi(h)} \right) \right\} \end{aligned}$$

Luego, usando **H5b)** se deduce que si $n \geq n_0$

$$\frac{\psi_{S_{\mathcal{H}}}(\rho_n)}{n\phi(h)} < \frac{1}{2} \frac{\epsilon^2}{a_1(1+a_2\epsilon)}$$

con lo cual si $n \geq n_0$ y si $c = 2a_1$

$$\mathbb{P} \left\{ \sup_{x \in S_{\mathcal{H}}} |R_j(x) - \mathbb{E}R_j(x)| > \epsilon \right\} \leq 8 \exp \left\{ -n\phi(h) \frac{\epsilon^2}{c(1+a_2\epsilon)} \right\}.$$

Usando que $|\tilde{R}_j(x) - \mathbb{E}\tilde{R}_j(x)| \leq \tilde{C}|R_j(x) - \mathbb{E}R_j(x)|$ se obtiene que

$$\mathbb{P} \left\{ \sup_{x \in S_{\mathcal{H}}} |\tilde{R}_j(x) - \mathbb{E}\tilde{R}_j(x)| > \tilde{C}\epsilon \right\} \leq 8 \exp \left\{ -n\phi(h) \frac{\epsilon^2}{c(1+a_2\epsilon)} \right\}.$$

En particular, si $\mathcal{A} = \{\inf_{x \in S_{\mathcal{H}}} \tilde{R}_0(x) \leq \frac{1}{2}\}$ y usando que $\mathbb{E}\tilde{R}_0(x) = 1$ obtenemos que para $A^{-1} = 2\tilde{C}c(2\tilde{C} + a_2)$

$$\mathbb{P}(\mathcal{A}) \leq \mathbb{P}(\sup_{x \in S_{\mathcal{H}}} |\tilde{R}_0(x) - \mathbb{E}\tilde{R}_0(x)| > \frac{1}{2}) \leq 8 \exp \{-An\phi(h)\}.$$

Por lo tanto, usando que $n\phi(h)/\log(n) \rightarrow \infty$ se obtiene que $\sup_{x \in S_{\mathcal{H}}} |\tilde{R}_j(x) - \mathbb{E}\tilde{R}_j(x)| \xrightarrow{a.co.} 0$ para $j = 0, 1$ y como

$$\mathbb{P} \left(\inf_{x \in S_{\mathcal{H}}} \tilde{R}_0(x) \leq \frac{1}{2} \right) < 8 \exp \{-An\phi(h)\}, \quad (4.19)$$

se obtiene (4.14). Esta conclusión combinada con el Lema 4.3.3a) y (4.11), permite deducir que para cada $y \in \mathbb{R}$,

$$\sup_{x \in S_{\mathcal{H}}} |\hat{F}(y|X = x) - F(y|X = x)| \xrightarrow{a.s.} 0. \quad (4.20)$$

Para cada $q \in \mathbb{Q}$ definamos $\mathcal{N}(q) = \{\omega \in \Omega : \sup_{x \in S_{\mathcal{H}}} |\hat{F}(q|X = x) - F(q|X = x)| \not\rightarrow 0\}$ y $\mathcal{N} = \bigcup_{q \in \mathbb{Q}} \mathcal{N}(q)$. Luego, (4.20) implica que $\mathbb{P}(\mathcal{N}) = 0$.

Sea $\omega \in \Omega$, $\omega \notin \mathcal{N}$, entonces, para todo $q \in \mathbb{Q}$, se cumple que $\sup_{x \in S_{\mathcal{H}}} |\hat{F}(q|X = x) - F(q|X = x)| \rightarrow 0$.

Dado $\epsilon > 0$, por la observación i) tenemos que **H3** implica que existen $a, b \in \mathbb{Q}$ tales que $F(b|X = x) > 1 - \epsilon$ y $F(a|X = x) < \epsilon$, para todo $x \in S_{\mathcal{H}}$. Por otra parte, la equicontinuidad pedida en **H3** ii) implica que existen $a = y_1 < y_2 < \dots < y_\ell = b$, $y_i \in \mathbb{Q}$, tales que $|y_i - y_{i-1}| < \delta$ y $|F(y|X = x) - F(y_i|X = x)| < \epsilon$ para todo $x \in S_{\mathcal{H}}$ y para todo y tal que $|y - y_i| < \delta$.

Sea $n_0 \in \mathbb{N}$ tal que para todo $n \geq n_0$, $\max_{1 \leq i \leq \ell} \sup_{x \in S_{\mathcal{H}}} |\hat{F}(y_i|X = x) - F(y_i|X = x)| < \epsilon$. Entonces, es fácil ver que $\sup_{y \in \mathbb{R}} \sup_{x \in S_{\mathcal{H}}} |\hat{F}(y|X = x) - F(y|X = x)| < 2\epsilon$, para todo $n \geq n_0$, lo que concluye la demostración. \square

Para probar el Teorema 4.3.6 necesitaremos el siguiente lema que es análogo al Lema 3.1 de Boente y Fraiman (1991). Su demostración se da en el Apéndice.

Lema 4.3.5. Sea $F_n(y|X = x)$ una sucesión de funciones de distribución condicional que verifican

$$\sup_{x \in S_{\mathcal{H}}} \sup_{y \in \mathbb{R}} |F_n(y|X = x) - F(y|X = x)| \rightarrow 0. \quad (4.21)$$

Entonces, si F verifica la hipótesis **H3**, existen constantes positivas $A \leq B$ tales que $s_n(x) = \text{MAD}_C(F_n(\cdot|X = x))$ verifica $A \leq s_n(x) \leq B$ para todo $x \in S_{\mathcal{H}}$ compacto y $n \geq n_0$.

Observemos que $s_n(x)$ es la escala definida en (4.6), correspondiente a $F_n(y|X = x)$.

Teorema 4.3.6. Supongamos que valen **H1** a **H5** y **H7**. Entonces, dado $S_{\mathcal{H}} \subset \mathcal{H}$ compacto se cumple

- a) $\sup_{x \in S_{\mathcal{H}}} |\widehat{g}(x) - g(x)| \xrightarrow{a.s.} 0$ cuando $n \rightarrow \infty$, si además **H6** vale.
- b) Si además se cumple **H8**, se tiene que $\sup_{x \in S_{\mathcal{H}}} |\widehat{m}(x) - m(x)| \xrightarrow{a.s.} 0$ cuando $n \rightarrow \infty$, donde $\widehat{m}(x)$ es la mediana de $\widehat{F}(y|X = x)$ y $m(x)$ es la mediana de $F(y|X = x)$.

DEMOSTRACIÓN. La demostración de la parte a) sigue los pasos de la demostración del Teorema 3.3 de Boente y Fraiman (1991).

a) Como $F(y|X = x)$ es simétrica alrededor de $g(x)$ y ψ es impar, tenemos que $\lambda(x, g(x), \sigma) = 0$ para todo $\sigma > 0$, donde $\lambda(x, a, \sigma)$ está definida en (4.5). Observemos que **H3i**) y **H6** implican que $\lambda(\cdot, a, \sigma) : \mathcal{H} \rightarrow \mathbb{R}$ es una función continua.

Queremos ver que $g(x)$ es una función continua sobre $S_{\mathcal{H}}$. Por ser ψ una función creciente, tenemos que

$$\forall \eta > 0, \lambda(x_0, g(x_0) - \eta, \sigma) > \lambda(x_0, g(x_0), \sigma) > \lambda(x_0, g(x_0) + \eta, \sigma).$$

Sea $\sigma = \sigma_0$ fijo y $x_0 \in S_{\mathcal{H}}$. Como $\lambda(\cdot, a, \sigma)$ es continua en $x = x_0$, tenemos que para todo $\epsilon > 0$ existe $\delta > 0$ tal que

$$d(x, x_0) < \delta \Rightarrow |\lambda(x, g(x_0) \pm \eta, \sigma_0) - \lambda(x_0, g(x_0) \pm \eta, \sigma_0)| < \epsilon. \quad (4.22)$$

Sea $\epsilon_\eta = \min\{\lambda(x_0, g(x_0) - \eta, \sigma_0) - \lambda(x_0, g(x_0), \sigma_0), \lambda(x_0, g(x_0), \sigma_0) - \lambda(x_0, g(x_0) + \eta, \sigma_0)\}$. Luego, (4.22) implica que existe $\delta_1 > 0$ tal que si $d(x_0, x) < \delta_1$ entonces

$$\lambda(x, g(x_0) + \eta, \sigma_0) < \lambda(x_0, g(x_0) + \eta, \sigma_0) + \epsilon_\eta < \lambda(x, g(x), \sigma_0),$$

de donde, como ψ es creciente, $g(x) - g(x_0) < \eta$. Análogamente se ve que existe $\delta_2 > 0$ tal que, si $d(x_0, x) < \delta_2$, $g(x) - g(x_0) > -\eta$.

Por lo tanto, dado $\eta > 0$ si $\delta = \min\{\delta_1, \delta_2\} > 0$ se cumple que $d(x_0, x) < \delta \Rightarrow |g(x) - g(x_0)| < \eta$, de donde resulta que g es continua sobre $S_{\mathcal{H}}$ y $\lambda(x, g(x) + a, \sigma)$ es continua para todo a y σ fijos. Más aún, es fácil ver que $\lambda(x, g(x) + a, \sigma)$ es una función continua de (x, σ) para todo a fijo.

Vamos a probar ahora que $\sup_{x \in S_{\mathcal{H}}} |\widehat{g}(x) - g(x)| \xrightarrow{a.s.} 0$. Dado $a \in \mathbb{R}$, sea $\psi_a(y) = \psi((y - g(x) - a)/\widehat{s}(x))$, luego ψ_a es de variación acotada y $\|\psi_a\|_V = \|\psi\|_V/\widehat{s}(x)$ donde $\|\psi\|_V$ indica la norma en variación de la función ψ . El Lema 4.3.5, implica que existen constantes reales positivas A y B tales que $A < \widehat{s}(x) < B$ para todo $x \in S_{\mathcal{H}}$, y todo $n \geq n_0$ casi seguramente. Luego, como

$$|\lambda(x, g(x) + a, \widehat{s}(x)) - \widehat{\lambda}(x, g(x) + a, \widehat{s}(x))| \leq \|\psi_a\|_V \sup_{y \in \mathbb{R}} |F(y|X = x) - \widehat{F}(y|X = x)|,$$

obtenemos que

$$\begin{aligned} \sup_{x \in S_{\mathcal{H}}} |\lambda(x, g(x) + a, \widehat{s}(x)) - \widehat{\lambda}(x, g(x) + a, \widehat{s}(x))| & \quad (4.23) \\ & \leq \frac{1}{A} \|\psi\|_V \sup_{x \in S_{\mathcal{H}}} \sup_{y \in \mathbb{R}} |F(y|X = x) - \widehat{F}(y|X = x)|, \end{aligned}$$

El Teorema 4.3.4 implica que el lado derecho de (4.23) converge a 0 casi seguramente, por lo tanto

$$\sup_{x \in S_{\mathcal{H}}} |\lambda(x, g(x) + a, \widehat{s}(x)) - \widehat{\lambda}(x, g(x) + a, \widehat{s}(x))| \xrightarrow{a.s.} 0 \quad (4.24)$$

Dado $\epsilon > 0$, **H2** y **H6** y la continuidad de $\lambda(x, g(x) \pm \epsilon, \sigma)$ implican que

$$\lambda_1 = \sup_{A \leq \sigma \leq B} \sup_{x \in S_{\mathcal{H}}} \lambda(x, g(x) + \epsilon, \sigma) < 0 < \inf_{A \leq \sigma \leq B} \inf_{x \in S_{\mathcal{H}}} \lambda(x, g(x) - \epsilon, \sigma) = \lambda_2. \quad (4.25)$$

Usando (4.24) y el hecho que $\mathbb{P}(A < \widehat{s}(x) < B) = 1$, de (4.25) deducimos que para n suficientemente grande

$$\widehat{\lambda}(x, g(x) + \epsilon, \widehat{s}(x)) < \frac{\lambda_1}{2} < 0 < \frac{\lambda_2}{2} < \widehat{\lambda}(x, g(x) - \epsilon, \widehat{s}(x)),$$

para todo $x \in S_{\mathcal{H}}$ casi seguramente. Por lo tanto, $\mathbb{P}(\sup_{x \in S_{\mathcal{H}}} |\widehat{g}(x) - g(x)| < \epsilon) = 1$, lo que concluye la demostración de a).

b) Veamos que $m(x)$ es continua en $S_{\mathcal{H}}$. Sea $x_0 \in S_{\mathcal{H}}$ fijo, como existe una única mediana, sabemos que para todo $\eta > 0$ se cumple

$$F(m(x_0) + \eta|X = x_0) > F(m(x_0)|X = x_0) > F(m(x_0) - \eta|X = x_0).$$

Por otra parte, **H3ii** implica que dado $\epsilon > 0$ existe $\delta > 0$ tal que si $d(x_0, x) < \delta$ entonces $|F(m(x_0) \pm \eta|X = x) - F(m(x_0) \pm \eta|X = x_0)| < \epsilon$.

Sea $\epsilon_\eta = \min\{F(m(x_0)|X = x_0) - F(m(x_0) - \eta|X = x_0), F(m(x_0) + \eta|X = x_0) - F(m(x_0)|X = x_0)\}$. Siguiendo la misma idea que en el inciso anterior, llegamos a que si $d(x_0, x) < \delta_1$ entonces

$$F(m(x_0) - \eta|X = x) < F(m(x_0) - \eta|X = x_0) + \epsilon_\eta < F(m(x_0)|X = x_0) = 1/2,$$

de donde tenemos que $m(x_0) - m(x) < \eta$. En forma análoga llegamos a que $m(x_0) - m(x) > -\eta$ si $d(x_0, x) < \delta_2$ para algún $\delta_2 > 0$, con lo cual obtenemos que, para todo $\eta > 0$, existe $\delta = \min(\delta_1, \delta_2) > 0$ tal que $d(x_0, x) < \delta$ implica que $|m(x) - m(x_0)| < \eta$, lo que prueba la continuidad de m en x_0 .

Sea $\epsilon > 0$ por **H3ii)** existe $\delta < \epsilon/2$ tal que

$$|u - v| < \delta \Rightarrow \sup_{x \in S_{\mathcal{H}}} |F(u|X = x) - F(v|X = x)| < \frac{\epsilon}{2},$$

de donde para todo $x \in S_{\mathcal{H}}$

$$\left| F\left(m(x) + \frac{\delta}{4} | X = x\right) - F\left(m(x) - \frac{\delta}{4} | X = x\right) \right| < \frac{\epsilon}{2}.$$

Como $F(m(x)|X = x) = 1/2$ obtenemos

$$\begin{aligned} \frac{1}{2} < F\left(m(x) + \frac{\delta}{4} | X = x\right) < F\left(m(x) - \frac{\delta}{4} | X = x\right) + \frac{\epsilon}{2} < \frac{1}{2} + \frac{\epsilon}{2} \\ \frac{1}{2} - \frac{\epsilon}{2} < F\left(m(x) + \frac{\delta}{4} | X = x\right) - \frac{\epsilon}{2} < F\left(m(x) - \frac{\delta}{4} | X = x\right) < \frac{1}{2}. \end{aligned}$$

Resumiendo para todo $x \in S_{\mathcal{H}}$ se cumple

$$\frac{1}{2} < F\left(m(x) + \frac{\delta}{4} | X = x\right) < \frac{1}{2} + \frac{\epsilon}{2} \quad \frac{1}{2} - \frac{\epsilon}{2} < F\left(m(x) - \frac{\delta}{4} | X = x\right) < \frac{1}{2}.$$

Como m es continua, tenemos que

$$i(\delta) = \inf_{x \in S_{\mathcal{H}}} F\left(m(x) + \frac{\delta}{4} | X = x\right) > 1/2 \quad s(\delta) = \sup_{x \in S_{\mathcal{H}}} F\left(m(x) - \frac{\delta}{4} | X = x\right) < 1/2.$$

Sea $\eta = \min\{i(\delta) - 1/2, 1/2 - s(\delta)\} > 0$ y $\epsilon_1 = \min\{\eta/2, \epsilon/2\}$. Luego, por el Teorema 4.3.4 se cumple que si

$$\mathcal{N} = \{\omega \in \Omega : \sup_{x \in S_{\mathcal{H}}} \sup_{y \in \mathbb{R}} |\widehat{F}(y|X = x) - F(y|X = x)| \not\rightarrow 0\}$$

$\mathbb{P}(\mathcal{N}) = 0$. Sea $\omega \notin \mathcal{N}$, luego existe n_0 tal que si $n \geq n_0$, para todo $x \in S_{\mathcal{H}}$ y para todo $y \in \mathbb{R}$, se verifica que $|\widehat{F}(y|X = x) - F(y|X = x)| < \epsilon_1$. En particular,

$$\begin{aligned} \left| \widehat{F}\left(m(x) + \frac{\delta}{4} | X = x\right) - F\left(m(x) + \frac{\delta}{4} | X = x\right) \right| < \epsilon_1 \\ \left| \widehat{F}\left(m(x) - \frac{\delta}{4} | X = x\right) - F\left(m(x) - \frac{\delta}{4} | X = x\right) \right| < \epsilon_1, \end{aligned}$$

es decir, tenemos que

$$F\left(m(x) - \frac{\delta}{4} | X = x\right) - \epsilon_1 < \widehat{F}\left(m(x) - \frac{\delta}{4} | X = x\right) < F\left(m(x) - \frac{\delta}{4} | X = x\right) + \epsilon_1 \quad (4.26)$$

$$F\left(m(x) + \frac{\delta}{4} | X = x\right) - \epsilon_1 < \widehat{F}\left(m(x) + \frac{\delta}{4} | X = x\right) < F\left(m(x) + \frac{\delta}{4} | X = x\right) + \epsilon_1. \quad (4.27)$$

Ahora bien, para todo $x \in S_{\mathcal{H}}$ tenemos que

$$F\left(m(x) - \frac{\delta}{4} | X = x\right) + \epsilon_1 < F\left(m(x) - \frac{\delta}{4} | X = x\right) + \frac{\eta}{2} < F\left(m(x) - \frac{\delta}{4} | X = x\right) + \frac{1}{4} - \frac{s(\delta)}{2}$$

y $F(m(x) - \delta/4 | X = x) < s(\delta) < 1/2$, de donde

$$F\left(m(x) - \frac{\delta}{4} | X = x\right) + \epsilon_1 < s(\delta) - \frac{s(\delta)}{2} + \frac{1}{4} = \frac{s(\delta)}{2} + \frac{1}{4} < \frac{1}{2}. \quad (4.28)$$

De igual forma tenemos

$$F\left(m(x) + \frac{\delta}{4} | X = x\right) - \epsilon_1 > F\left(m(x) + \frac{\delta}{4} | X = x\right) - \frac{1}{2} \left(i(\delta) - \frac{1}{2}\right) > \frac{1}{4} + \frac{i(\delta)}{2} > \frac{1}{2}. \quad (4.29)$$

Es decir, (4.26), (4.27), (4.28) y (4.29) implican que

$$\widehat{F}\left(m(x) - \frac{\delta}{4} | X = x\right) < \frac{1}{2} \quad \text{y} \quad \widehat{F}\left(m(x) + \frac{\delta}{4} | X = x\right) > \frac{1}{2}.$$

Como $\widehat{F}(\widehat{m}(x) | X = x) = 1/2$, obtenemos que $m(x) - \delta/4 < \widehat{m}(x) < m(x) + \delta/4$, o sea, $\sup_{x \in S_{\mathcal{H}}} |\widehat{m}(x) - m(x)| < \epsilon/8$, lo que concluye la demostración. \square

4.4. Tasas de convergencia fuerte uniforme

Para obtener convergencia en modelos parcialmente lineales necesitaremos tasas en probabilidad o en casi todo punto para la convergencia de los estimadores de la componente no paramétrica. En esta Sección, obtendremos tasas de convergencia con tal fin.

Los Teoremas 4.4.1 y 4.4.2 dan tasas de convergencia casi completa para los estimadores de la distribución empírica condicional y para los M -estimadores locales de la función de regresión g .

Teorema 4.4.1. *Sea $S_{\mathcal{H}} \subset \mathcal{H}$ compacto. Supongamos que se cumplen **H1**, **H3**, **H4**, **H5**, **H7**, **H9** y **H10** donde si $\eta_1 < 1/2$ se cumple además que existe una constante positiva C^* tal que*

$$\left(\frac{n}{\log(n)}\right)^{1-\eta_1} \phi(h) \leq C^*. \quad (4.30)$$

Entonces, se tiene que

$$\sup_{y \in \mathbb{R}} \sup_{x \in S_{\mathcal{J}_c}} |\widehat{F}(y|X = x) - F(y|X = x)| = O_{a.co.} \left(h^{\eta_1} + \sqrt{\frac{\psi_{S_{\mathcal{J}_c}}(\log(n)/n)}{n\phi(h)}} \right)$$

DEMOSTRACIÓN. El Lema 4.3.2c) y el hecho que $\theta_n \leq \tilde{\theta}_n = h^{\eta_1} + \theta_n$ implican que si $\epsilon_0 > c$

$$\sup_{y \in \mathbb{R}} \mathbb{P} \left\{ \tilde{\theta}_n^{-1} \sup_{x \in S_{\mathcal{J}_c}} |R_j(x) - \mathbb{E}R_j(x)| > \epsilon_0 \right\} \leq 8 \exp \left\{ \left(1 - \frac{\epsilon_0^2}{8(1 + \epsilon_0)} \right) \psi_{S_{\mathcal{J}_c}} \left(\frac{\log(n)}{n} \right) \right\}.$$

Por lo tanto, usando que $|\tilde{R}_j(x) - \mathbb{E}\tilde{R}_j(x)| \leq \tilde{C} |R_j(x) - \mathbb{E}R_j(x)|$ y llamando $b_1 = 8\tilde{C}^2$ y $b_2 = 1/\tilde{C}$, obtenemos que si $\epsilon_0 > 2c$

$$\sup_{y \in \mathbb{R}} \mathbb{P} \left\{ \tilde{\theta}_n^{-1} \sup_{x \in S_{\mathcal{J}_c}} |\tilde{R}_j(x) - \mathbb{E}\tilde{R}_j(x)| > \epsilon_0 \right\} \leq 8 \exp \left\{ \left(1 - \frac{\epsilon_0^2}{b_1(1 + b_2\epsilon_0)} \right) \psi_{S_{\mathcal{J}_c}} \left(\frac{\log(n)}{n} \right) \right\}, \quad (4.31)$$

El Corolario del Lema 4.3.3 y el hecho que $\theta_n \leq \tilde{\theta}_n = h^{\eta_1} + \theta_n$ implican que

$$\sup_{y \in \mathbb{R}} \sup_{x \in S_{\mathcal{J}_c}} |\mathbb{E}\tilde{R}_1(x) - F(y|X = x)\mathbb{E}\tilde{R}_0(x)| \leq C\tilde{\theta}_n$$

Por otra parte, (4.19) implica que $\sum_{n \geq 1} \mathbb{P} \left(\inf_{x \in S_{\mathcal{J}_c}} \tilde{R}_0(x) \leq 1/2 \right) < \infty$. Luego, si $\mathcal{A} = \{ \inf_{x \in S_{\mathcal{J}_c}} \tilde{R}_0(x) \leq 1/2 \}$, obtenemos que si $\epsilon_0 > C/2$

$$\sup_{y \in \mathbb{R}} \mathbb{P} \left(\sup_{x \in S_{\mathcal{J}_c}} |\widehat{F}(y|X = x) - F(y|X = x)| > 4\epsilon_0\tilde{\theta} \right) \leq \mathbb{P}(\mathcal{A}) \quad (4.32)$$

$$\begin{aligned} & + \sup_{y \in \mathbb{R}} \mathbb{P} \left(\sum_{j=0}^1 \sup_{x \in S_{\mathcal{J}_c}} |\tilde{R}_j(x) - \mathbb{E}\tilde{R}_j(x)| > \epsilon_0\tilde{\theta} \right) \\ & \leq \mathbb{P}(\mathcal{A}) + \mathbb{P} \left(\sup_{x \in S_{\mathcal{J}_c}} |\tilde{R}_0(x) - \mathbb{E}\tilde{R}_0(x)| > \epsilon_0\tilde{\theta} \right) \\ & + \sup_{y \in \mathbb{R}} \mathbb{P} \left(\sup_{x \in S_{\mathcal{J}_c}} |\tilde{R}_1(x) - \mathbb{E}\tilde{R}_1(x)| > \epsilon_0\tilde{\theta} \right) \end{aligned} \quad (4.33)$$

Luego, usando (4.31) y (4.19) obtenemos que

$$\begin{aligned} \sup_{y \in \mathbb{R}} \mathbb{P} \left(\sup_{x \in S_{\mathcal{J}_c}} |\widehat{F}(y|X = x) - F(y|X = x)| > 4\epsilon_0\tilde{\theta} \right) & \leq 8 \exp \{ -An\phi(h) \} \\ & + 16 \exp \left\{ \left(1 - \frac{\epsilon_0^2}{b_1(1 + b_2\epsilon_0)} \right) \psi_{S_{\mathcal{J}_c}} \left(\frac{\log(n)}{n} \right) \right\} \end{aligned} \quad (4.34)$$

En lo que sigue suponemos, sin pérdida de generalidad, que $D = 1$.

i) Supongamos primero que $\eta_1 > 1/2$.

Sea x_1, \dots, x_ℓ un cubrimiento de $S_{\mathcal{H}}$ por bolas de radio $\rho_n = \log(n)/n$, es decir, $S_{\mathcal{H}} \subset \cup_{j=1}^{\ell} B(x_j, \rho_n)$ donde $\ell = N_{\rho_n}(S_{\mathcal{H}}) = \exp(\psi_{S_{\mathcal{H}}}(\log(n)/n))$.

Definamos $M = \lceil \sqrt{n} \rceil + 1$, donde $\lceil z \rceil$ indica la parte entera de $z \in \mathbb{R}$. Como $F(y|X = x)$ es una función continua de y , dado $M \in \mathbb{N}$, para cada $1 \leq j \leq \ell$ existe y_k^j , $1 \leq k \leq M$, tal que $F(y_k^j|X = x_j) = k/M - 1/(2M)$. Como $\eta_1 > 1/2$ tenemos que $\rho_n^m \leq 1/M$. Luego si ϵ está dado en **H9**, para $n \geq n_0$ tenemos que $\rho_n < \epsilon$ y además si $x \in B(x_j, \rho_n)$ se cumple $|F(y|X = x) - F(y|X = x_j)| \leq \rho_n^m \leq 1/M$. Sea $S_{\mathcal{H},j} = S_{\mathcal{H}} \cap B(x_j, \rho_n)$, entonces para todo $x \in S_{\mathcal{H},j}$

$$\frac{k}{M} - \frac{3}{2M} < F(y_k^j|X = x) < \frac{k}{M} + \frac{1}{2M},$$

lo que implica que

$$\begin{aligned} \tilde{\theta}_n^{-1} \sup_{y \in \mathbb{R}} \sup_{x \in S_{\mathcal{H}}} |\widehat{F}(y|X = x) - F(y|X = x)| &\leq 2 \frac{\theta_n}{\tilde{\theta}_n} \frac{1}{\theta_n M} \\ &+ \tilde{\theta}_n^{-1} \max_{1 \leq j \leq \ell} \sup_{x \in S_{\mathcal{H},j}} \max_{1 \leq k \leq M} |\widehat{F}(y_k^j|X = x) - F(y_k^j|X = x)|. \end{aligned}$$

Luego, usando por **H5b)** obtenemos que

$$M^{-1} \theta_n^{-1} \leq \sqrt{\frac{n \phi(h)}{\psi_{S_{\mathcal{H}}} \left(\frac{\log(n)}{n} \right)} \frac{1}{\sqrt{n}}} = \sqrt{\frac{\phi(h)}{\psi_{S_{\mathcal{H}}} \left(\frac{\log(n)}{n} \right)}} \rightarrow 0,$$

de donde como $\theta_n \leq \tilde{\theta}_n$ obtenemos que si $n \geq n_0$

$$\begin{aligned} \tilde{\theta}_n^{-1} \sup_{y \in \mathbb{R}} \sup_{x \in S_{\mathcal{H}}} |\widehat{F}(y|X = x) - F(y|X = x)| &\leq 1 \\ &+ \tilde{\theta}_n^{-1} \max_{1 \leq k \leq M} \max_{1 \leq j \leq \ell} \sup_{x \in S_{\mathcal{H},j}} |\widehat{F}(y_k^j|X = x) - F(y_k^j|X = x)|. \end{aligned}$$

Por lo tanto, como $\epsilon_0 > 1$ obtenemos que

$$\begin{aligned} \mathbb{P}(\sup_{y \in \mathbb{R}} \sup_{x \in S_{\mathcal{H}}} |\widehat{F}(y|X = x) - F(y|X = x)| > 5\epsilon_0 \tilde{\theta}_n) \\ \leq \mathbb{P} \left(\max_{1 \leq k \leq M} \max_{1 \leq j \leq \ell} \sup_{x \in S_{\mathcal{H},j}} |\widehat{F}(y_k^j|X = x) - F(y_k^j|X = x)| > 4\epsilon_0 \tilde{\theta} \right) \\ \leq \ell M \sup_{y \in \mathbb{R}} \mathbb{P}(\sup_{x \in S_{\mathcal{H}}} |\widehat{F}(y|X = x) - F(y|X = x)| > 4\epsilon_0 \tilde{\theta}) \end{aligned}$$

De donde usando (4.34) y que $\ell = N_{\rho_n}(S_{\mathcal{J}_c}) = \exp(\psi_{S_{\mathcal{J}_c}}(\log(n)/n))$, se deduce que

$$\begin{aligned}
& \mathbb{P}(\sup_{y \in \mathbb{R}} \sup_{x \in S_{\mathcal{J}_c}} |\widehat{F}(y|X = x) - F(y|X = x)| > 5\epsilon_0 \tilde{\theta}_n) \\
& \leq 16\sqrt{n} \exp\left(\psi_{S_{\mathcal{J}_c}}\left(\frac{\log(n)}{n}\right)\right) \left(\exp\{-An\phi(h)\} \right. \\
& \quad \left. + 2 \exp\left\{\left(1 - \frac{\epsilon_0^2}{b_1(1 + b_2\epsilon_0)}\right) \psi_{S_{\mathcal{J}_c}}\left(\frac{\log(n)}{n}\right)\right\}\right) \\
& \leq 16\sqrt{n} \left(\exp\left\{\left(1 - A \frac{n\phi(h)}{\psi_{S_{\mathcal{J}_c}}\left(\frac{\log(n)}{n}\right)}\right) \psi_{S_{\mathcal{J}_c}}\left(\frac{\log(n)}{n}\right)\right\} \right. \\
& \quad \left. + 2 \exp\left\{\left(2 - \frac{\epsilon_0^2}{b_1(1 + b_2\epsilon_0)}\right) \psi_{S_{\mathcal{J}_c}}\left(\frac{\log(n)}{n}\right)\right\}\right)
\end{aligned}$$

Ahora bien **H5b)** implica que

$$\frac{n\phi(h)}{\psi_{S_{\mathcal{J}_c}}\left(\frac{\log(n)}{n}\right)} \rightarrow \infty$$

luego si $n \geq n_0$,

$$A \frac{n\phi(h)}{\psi_{S_{\mathcal{J}_c}}\left(\frac{\log(n)}{n}\right)} - 1 > \beta - 1 > 0$$

donde β está dado en **H10**. Por otra parte, si ϵ_0 es tal que $\epsilon_0^2/(b_1(1 + b_2\epsilon_0)) > 1 + \beta$, entonces

$$2 - \frac{\epsilon_0^2}{b_1(1 + b_2\epsilon_0)} < 1 - \beta$$

Luego, si $n \geq n_0$, existe ϵ_0 tal que

$$\begin{aligned}
& \mathbb{P}(\sup_{y \in \mathbb{R}} \sup_{x \in S_{\mathcal{J}_c}} |\widehat{F}(y|X = x) - F(y|X = x)| > 5\epsilon_0 \tilde{\theta}_n) \\
& \leq 48\sqrt{n} \exp\left\{(1 - \beta) \psi_{S_{\mathcal{J}_c}}\left(\frac{\log(n)}{n}\right)\right\}
\end{aligned}$$

Por lo tanto, $\sum_{n \geq 1} \mathbb{P}(\sup_{y \in \mathbb{R}} \sup_{x \in S_{\mathcal{J}_c}} |\widehat{F}(y|X = x) - F(y|X = x)| > \epsilon_0 \tilde{\theta}_n) < \infty$.

ii) Consideremos ahora el caso $\eta_1 < 1/2$. Definamos $M = \lfloor \rho_n^{-\eta_1} \rfloor$ con $\rho_n = \log(n)/n$, o sea, $M = \lfloor (n/\log(n))^{\eta_1} \rfloor$. Luego, $\rho_n^{\eta_1} \leq 1/M$ y usando (4.30) se obtiene que $\theta_n^{-1} M^{-1} \leq C^*$ pues $\psi_{S_{\mathcal{J}_c}}(\log(n)/n) \geq (\log(n))^2/(n\phi(h))$. Luego si ϵ está dado en **H9**, para $n \geq n_0$ tenemos que $\rho_n < \epsilon$ y además si $x \in B(x_j, \rho_n)$ se cumple $|F(y|X = x) - F(y|X = x_j)| \leq \rho_n^{\eta_1} \leq 1/M$.

Sea $S_{\mathcal{H},j} = S_{\mathcal{H}} \cap B(x_j, \rho_n)$, entonces como en i) tenemos que

$$\begin{aligned} \tilde{\theta}_n^{-1} \sup_{y \in \mathbb{R}} \sup_{x \in S_{\mathcal{H}}} |\hat{F}(y|X = x) - F(y|X = x)| &\leq 2 \frac{\theta_n}{\tilde{\theta}_n} \frac{1}{\theta_n M} \\ &+ \tilde{\theta}_n^{-1} \max_{1 \leq j \leq \ell} \sup_{x \in S_{\mathcal{H},j}} \max_{1 \leq k \leq M} |\hat{F}(y_k^j|X = x) - F(y_k^j|X = x)|. \end{aligned}$$

Como $\theta_n^{-1} M^{-1} \leq C^*$ y $\theta_n \leq \tilde{\theta}_n$ obtenemos

$$\begin{aligned} \tilde{\theta}_n^{-1} \sup_{y \in \mathbb{R}} \sup_{x \in S_{\mathcal{H}}} |\hat{F}(y|X = x) - F(y|X = x)| &\leq 2C^* \\ &+ \tilde{\theta}_n^{-1} \max_{1 \leq k \leq M} \max_{1 \leq j \leq \ell} \sup_{x \in S_{\mathcal{H},j}} |\hat{F}(y_k^j|X = x) - F(y_k^j|X = x)|. \end{aligned}$$

Por lo tanto, como $\epsilon_0 > 1/4$ obtenemos que

$$\begin{aligned} \mathbb{P}(\sup_{y \in \mathbb{R}} \sup_{x \in S_{\mathcal{H}}} |\hat{F}(y|X = x) - F(y|X = x)| > 5\epsilon_0 \tilde{\theta}_n) \\ \leq \mathbb{P} \left(\max_{1 \leq k \leq M} \max_{1 \leq j \leq \ell} \sup_{x \in S_{\mathcal{H},j}} |\hat{F}(y_k^j|X = x) - F(y_k^j|X = x)| > 4\epsilon_0 \tilde{\theta} \right) \\ \leq \ell M \sup_{y \in \mathbb{R}} \mathbb{P}(\sup_{x \in S_{\mathcal{H}}} |\hat{F}(y|X = x) - F(y|X = x)| > 4\epsilon_0 \tilde{\theta}) \end{aligned}$$

Como $\rho_n = \log(n)/n$ luego $\ell = N_{\rho_n}(S_{\mathcal{H}}) = \exp(\psi_{S_{\mathcal{H}}}(\log(n)/n))$. Por lo tanto, como en i) deducimos que

$$\begin{aligned} \mathbb{P}(\sup_{y \in \mathbb{R}} \sup_{x \in S_{\mathcal{H}}} |\hat{F}(y|X = x) - F(y|X = x)| > 5\epsilon_0 \tilde{\theta}_n) \\ \leq 16\theta_n^{-1} \exp \left(\psi_{S_{\mathcal{H}}} \left(\frac{\log(n)}{n} \right) \right) (\exp \{-An\phi(h)\} \\ + 2 \exp \left\{ \left(1 - \frac{\epsilon_0^2}{b_1(1+b_2\epsilon_0)} \right) \psi_{S_{\mathcal{H}}} \left(\frac{\log(n)}{n} \right) \right\}) \\ \leq 16\theta_n^{-1} \left(\exp \left\{ \left(1 - A \frac{n\phi(h)}{\psi_{S_{\mathcal{H}}} \left(\frac{\log(n)}{n} \right)} \right) \psi_{S_{\mathcal{H}}} \left(\frac{\log(n)}{n} \right) \right\} \right. \\ \left. + 2 \exp \left\{ \left(2 - \frac{\epsilon_0^2}{b_1(1+b_2\epsilon_0)} \right) \psi_{S_{\mathcal{H}}} \left(\frac{\log(n)}{n} \right) \right\} \right) \end{aligned}$$

Usando como en i) **H5b)** y si ϵ_0 es tal que $\epsilon_0^2/(b_1(1+b_2\epsilon_0)) > 1 + \beta$, obtenemos que si $n \geq n_0$

$$\begin{aligned} \mathbb{P}(\sup_{y \in \mathbb{R}} \sup_{x \in S_{\mathcal{H}}} |\hat{F}(y|X = x) - F(y|X = x)| > 5\epsilon_0 \tilde{\theta}_n) \\ \leq 48\sqrt{n} \exp \left\{ (1 - \beta) \psi_{S_{\mathcal{H}}} \left(\frac{\log(n)}{n} \right) \right\} \end{aligned}$$

pues $\theta_n^{-1} \leq \sqrt{n}$. Por lo tanto, $\sum_{n \geq 1} \mathbb{P}(\sup_{y \in \mathbb{R}} \sup_{x \in S_{\mathcal{H}}} |\widehat{F}(y|X = x) - F(y|X = x)| > \epsilon_0 \widetilde{\theta}_n) < \infty$. \square

Observación: Vale la pena mencionar que si en lugar de estar interesados en tasas de convergencia uniforme para $x \in S_{\mathcal{H}}$ e $y \in \mathbb{R}$ queremos obtener tasas solamente para $x \in S_{\mathcal{H}}$ basta con requerir que la ϵ -entropía de Kolmogorov de $S_{\mathcal{H}}$ satisfaga

$$\sum_{n=1}^{\infty} \exp \left\{ (1 - \beta) \psi_{S_{\mathcal{H}}} \left(\frac{\log(n)}{n} \right) \right\} < \infty \quad \text{para algún } \beta > 1,$$

en lugar de **H10**. Esta condición es pedida en el Corolario 3 de Ferraty *et al.* (2010). La uniformidad para $y \in \mathbb{R}$ necesaria para obtener las tasas de convergencia uniforme para \widehat{g} exigen una condición más fuerte para la entropía del conjunto $S_{\mathcal{H}}$.

Teorema 4.4.2. *Sea $S_{\mathcal{H}} \subset \mathcal{H}$ compacto. Supongamos que se cumplen **H1** a **H5**, **H6**, **H7**, **H9** y **H10** donde si $\eta_1 < 1/2$ se cumple (4.30). Entonces,*

$$\sup_{x \in S_{\mathcal{H}}} |\widehat{g}(x) - g(x)| = O_{a.co.} \left(h^m + \sqrt{\frac{\psi_{S_{\mathcal{H}}}(\log(n)/n)}{n\phi(h)}} \right).$$

Demostración. Como $A \leq \widehat{s}(x) \leq B$, dado $a \in \mathbb{R}$ tenemos que

$$\sup_{x \in S_{\mathcal{H}}} |\lambda(x, g(x) + a, \widehat{s}(x)) - \widehat{\lambda}(x, g(x) + a, \widehat{s}(x))| \leq \frac{1}{A} \|\psi\|_V \sup_{x \in S_{\mathcal{H}}} \sup_{y \in \mathbb{R}} |F(y|X = x) - \widehat{F}(y|X = x)|.$$

Por lo tanto, si $\widetilde{\theta}_n = h^m + \sqrt{\psi_{S_{\mathcal{H}}}(\log(n)/n)/(n\phi(h))}$, el Teorema 4.4.1 implica que para cada $\delta > 0$

$$\widetilde{\theta}_n^{-1} \sup_{x \in S_{\mathcal{H}}} \sup_{|a| \leq \delta} |\lambda(x, g(x) + a, \widehat{s}(x)) - \widehat{\lambda}(x, g(x) + a, \widehat{s}(x))| = O_{a.s.}(1).$$

Entonces, como $\sup_{x \in S_{\mathcal{H}}} |\widehat{g}(x) - g(x)| \xrightarrow{a.s.} 0$, tenemos que

$$\widetilde{\theta}_n^{-1} \sup_{x \in S_{\mathcal{H}}} |\widehat{\lambda}(x, \widehat{g}(x), \widehat{s}(x)) - \lambda(x, \widehat{g}(x), \widehat{s}(x))| = O_{a.s.}(1). \quad (4.35)$$

Observemos que

$$\lambda(x, g(x), \widehat{s}(x)) - \lambda(x, \widehat{g}(x), \widehat{s}(x)) + \lambda(x, \widehat{g}(x), \widehat{s}(x)) - \widehat{\lambda}(x, \widehat{g}(x), \widehat{s}(x)) = 0$$

luego (4.35) implica que

$$\widetilde{\theta}_n^{-1} \sup_{x \in S_{\mathcal{H}}} |\lambda(x, g(x), \widehat{s}(x)) - \lambda(x, \widehat{g}(x), \widehat{s}(x))| = O_{a.s.}(1). \quad (4.36)$$

Sea

$$\lambda'(x, t, \sigma) = \frac{\partial}{\partial u} \lambda(x, u, \sigma)|_{u=t}.$$

Por la hipótesis **H6** y el Lema 4.3.5 tenemos que

$$\inf_{|a| \leq \delta} \inf_{x \in S_{\mathcal{H}}} -\frac{\partial}{\partial u} \lambda(x, g(x) + u, \widehat{s}(x))|_{u=a} > c_0 > 0, \quad (4.37)$$

para todo $n \geq N_0$ casi seguramente.

Por el Teorema del valor medio, obtenemos que

$$\lambda(x, g(x), \widehat{s}(x)) - \lambda(x, \widehat{g}(x), \widehat{s}(x)) = (\widehat{g}(x) - g(x))[-\lambda'(x, \xi_n, \widehat{s}(x))],$$

donde $\xi_n = \xi_n(x)$ es un punto intermedio entre $g(x)$ y $\widehat{g}(x)$.

Como vale (4.37) y $\sup_{x \in S_{\mathcal{H}}} |\widehat{g}(x) - g(x)| \xrightarrow{a.s.} 0$, existe \mathcal{N} tal que $\mathbb{P}(\mathcal{N}) = 0$ y si $\omega \in \mathcal{N}$, existe N_1 tal que si $n \geq N_1$, $|\xi_n - g(x)| \leq \delta$, $\widetilde{\theta}_n^{-1} \sup_{x \in S_{\mathcal{H}}} |\lambda(x, g(x), \widehat{s}(x)) - \lambda(x, \widehat{g}(x), \widehat{s}(x))| = O(1)$ y (4.37) vale. Luego, si $\omega \in \mathcal{N}$ tenemos que $\inf_{x \in S_{\mathcal{H}}} -\lambda'(x, \xi_n, \widehat{s}(x)) > c_0 > 0$, que junto a (4.36) concluye la demostración. \square

El siguiente teorema establece los mismos resultados que el Teorema 4.4.2 bajo condiciones sobre g y σ sin pedir condiciones Lipschitz sobre la distribución condicional.

Teorema 4.4.3. *Sea $S_{\mathcal{H}} \subset \mathcal{H}$ compacto. Supongamos que se cumplen **H1**, **H4**, **H5**, **H6** y **H7** y que*

$$\sum_{i=1}^{\infty} n \exp \left\{ (1 - \beta) \psi_{S_{\mathcal{H}}} \left(\frac{\log(n)}{n} \right) \right\} < \infty \quad \text{para algún } \beta > 1. \quad (4.38)$$

Supongamos además que σ y g son Lipschitz de orden η_1 y η_2 , respectivamente, donde $\eta = \min(\eta_1, \eta_2)$ es tal que si $\eta < 1$ se cumple (4.39) o (4.40) dadas por

$$h \left(\frac{n}{\log(n)} \right)^{1-\eta} \leq C_{\eta} \quad \text{para todo } n \geq 1 \quad (4.39)$$

$$\phi(h) \left(\frac{n}{\log(n)} \right)^{1-\eta} \leq C_{\eta} \quad \text{para todo } n \geq 1. \quad (4.40)$$

Sea $g(x)$ la única solución de $\lambda(x, a, s(x)) = 0$. Entonces, si $\widehat{g}(x)$ es una solución de (4.7) tal que $\sup_{x \in S_{\mathcal{H}}} |\widehat{g}(x) - g(x)| \xrightarrow{a.s.} 0$ se tiene que

$$\sup_{x \in S_{\mathcal{H}}} |\widehat{g}(x) - g(x)| = O_{a.co.} \left(h^{\eta} + \sqrt{\frac{\psi_{S_{\mathcal{H}}}(\log(n)/n)}{n\phi(h)}} \right).$$

Observación: Si **H2** se verifica entonces, por el Teorema 4.3.6, tenemos que $\sup_{x \in S_{\mathcal{H}}} |\widehat{g}(x) - g(x)| \xrightarrow{a.s.} 0$ se cumple.

Demostración. Como en el Teorema 4.4.2 bastará ver que

$$A_n = \widetilde{\theta}_n^{-1} \sup_{x \in S_{\mathcal{H}}} \sup_{|a| \leq \delta} \sup_{A \leq \sigma \leq B} |\widehat{\lambda}(x, g(x) + a, \sigma) - \lambda(x, g(x) + a, \sigma)| = O_{a.co.}(1).$$

Llamemos

$$W_{i,a,\sigma}(x) = \psi \left(\frac{Y_i - g(x) - a}{\sigma} \right), \quad \widetilde{R}_1(x, a, \sigma) = \frac{1}{n} \sum_{i=1}^n W_{i,a,\sigma}(x) \frac{K_i(x)}{\mathbb{E}K_1(x)}$$

$$R_1(x, a, \sigma) = \frac{1}{n} \sum_{i=1}^n W_{i,a,\sigma}(x) \frac{K_i(x)}{\phi(h)}.$$

Entonces $\widehat{\lambda}(x, g(x) + a, \sigma) = \widetilde{R}_1(x, a, \sigma) / \widetilde{R}_0(x)$, donde $\widetilde{R}_0(x)$ está definido en (4.10) y como antes

$$|\widehat{\lambda}(x, g(x) + a, \sigma) - \lambda(x, g(x) + a, \sigma)| \leq \frac{1}{\inf_{x \in S_{\mathcal{H}}} \widetilde{R}_0(x)} \left[\sup_{x \in S_{\mathcal{H}}} \sup_{|a| \leq \delta} \sup_{A \leq \sigma \leq B} |\widetilde{R}_1(x, a, \sigma) - \mathbb{E}\widetilde{R}_1(x, a, \sigma)| \right. \\ \left. + \sup_{x \in S_{\mathcal{H}}} |\widetilde{R}_0(x) - \mathbb{E}\widetilde{R}_0(x)| + \sup_{x \in S_{\mathcal{H}}} \sup_{|a| \leq \delta} \sup_{A \leq \sigma \leq B} |\mathbb{E}\widetilde{R}_1(x, a, \sigma) - \lambda(x, g(x) + a, \sigma)\mathbb{E}\widetilde{R}_0(x)| \right]. \quad (4.41)$$

Como ψ cumple **H6** y si σ y g son Lipschitz de orden η_1 y η_2 , respectivamente, para alguna constante C , tenemos que si $d(x_1, x) < 1$

$$\sup_{x \in S_{\mathcal{H}}} \sup_{|a| \leq \delta} \sup_{A \leq \sigma \leq B} |\lambda(x_1, g(x) + a, \sigma) - \lambda(x, g(x) + a, \sigma)| \leq C[d(x_1, x)^{\eta_1} + d(x_1, x)^{\eta_2}] \leq Cd(x_1, x)^\eta$$

donde $\eta = \min\{\eta_1, \eta_2\}$. Como $K_1(x) \leq I_{B(x,h)}(x_1)$ y $d(x_1, x) < 1$, para n suficientemente grande, deducimos que

$$|\mathbb{E}\widetilde{R}_1(x, a, \sigma) - \lambda(x, g(x) + a, \sigma)\mathbb{E}\widetilde{R}_0(x)| \leq \mathbb{E}|\lambda(x_1, g(x) + a, \sigma) - \lambda(x, g(x) + a, \sigma)| \frac{K_1(x)}{\mathbb{E}K_1(x)} \\ \leq Ch^\eta.$$

Luego si $\widetilde{\theta}_n = h^\eta + \theta_n$ tenemos que

$$B_{2n} = \sup_{x \in S_{\mathcal{H}}} \sup_{|a| \leq \delta} \sup_{A \leq \sigma \leq B} |\mathbb{E}\widetilde{R}_1(x, a, \sigma) - \lambda(x, g(x) + a, \sigma)\mathbb{E}\widetilde{R}_0(x)| \leq C\widetilde{\theta}_n.$$

Por lo tanto, si $\mathcal{A} = \{\inf_{x \in S_{\mathcal{H}}} \widetilde{R}_0(x) < 1/2\}$ y si $\epsilon_0 > C$ de (4.41) deducimos que

$$\mathbb{P}(A_n > 4\epsilon_0) \leq \mathbb{P}(\mathcal{A}) + \mathbb{P}(\sup_{x \in S_{\mathcal{H}}} |\widetilde{R}_0(x) - \mathbb{E}\widetilde{R}_0(x)| > \epsilon_0\theta_n) + \mathbb{P}(\widetilde{B}_{1n} > \epsilon_0\theta_n),$$

donde $\tilde{B}_{1n} = \sup_{x \in S_{\mathcal{H}}} \sup_{|a| \leq \delta} \sup_{A \leq \sigma \leq B} |\tilde{R}_1(x, a, \sigma) - \mathbb{E}\tilde{R}_1(x, a, \sigma)|$. Por el Lema (4.3.2)b), tomando $\rho_n = \log(n)/n$ y usando (4.38) obtenemos que $\sum_{n \geq 1} \mathbb{P}(\mathcal{A}) < \infty$. Por otra parte el 4.3.2)c) implica que si $\epsilon_0 > c$ entonces $\sum_{n \geq 1} \mathbb{P}(\sup_{x \in S_{\mathcal{H}}} |\tilde{R}_0(x) - \mathbb{E}\tilde{R}_0(x)| > \epsilon_0 \theta_n) < \infty$.

Observemos que $\mathbb{E}K_1(x) \geq C\phi(h)$ implica que

$$\mathbb{P}(\tilde{B}_{1n} > \epsilon_0 \theta_n) \leq \mathbb{P}(B_{1n} > \epsilon_1 \theta_n)$$

donde $\epsilon_1 = \epsilon_0 C$ y $B_{1n} = \sup_{x \in S_{\mathcal{H}}} \sup_{|a| \leq \delta} \sup_{A \leq \sigma \leq B} |R_1(x, a, \sigma) - \mathbb{E}R_1(x, a, \sigma)|$. Por lo tanto, bastará mostrar que para algún c_1 y para todo $\epsilon_1 > c_1$, $\sum_{n \geq 1} \mathbb{P}(B_{1n} > \epsilon_1 \theta_n) < \infty$.

Sea $\rho_n = \log(n)/n$ y sean x_1, \dots, x_ℓ tales que $S_{\mathcal{H}} \subset \cup_{j=1}^{\ell} B(x_j, \rho_n)$ donde $\ell = N_{\rho_n}(S_{\mathcal{H}})$. Luego, si llamamos $S_n(x, a, \sigma) = R_1(x, a, \sigma) - \mathbb{E}R_1(x, a, \sigma)$ y $\tilde{S}_{n,k}(x, a, \sigma) = S_n(x, a, \sigma) - S_n(x_k, a, \sigma)$, tenemos que $S_n(x, a, \sigma) = S_n(x_k, a, \sigma) + \tilde{S}_{n,k}(x, a, \sigma)$ con lo cual

$$\begin{aligned} \mathbb{P}\left(\sup_{\substack{A \leq \sigma \leq B \\ |a| \leq \delta}} \sup_{x \in S_{\mathcal{H}}} S_n(x, a, \sigma) > \epsilon_0 \theta_n\right) &\leq \mathbb{P}\left(\sup_{\substack{A \leq \sigma \leq B \\ |a| \leq \delta}} \max_{1 \leq k \leq \ell} |S_n(x_k, a, \sigma)| > \frac{\epsilon_1}{2} \theta_n\right) \\ &+ \mathbb{P}\left(\sup_{\substack{A \leq \sigma \leq B \\ |a| \leq \delta}} \max_{1 \leq k \leq \ell} \sup_{x \in B(x_k, \rho_n) \cap S_{\mathcal{H}}} |\tilde{S}_{n,k}(x, a, \sigma)| > \frac{\epsilon_1}{2} \theta_n\right). \end{aligned}$$

Observemos que $|\tilde{S}_{n,k}(x, a, \sigma)| \leq |R_1(x, a, \sigma) - R_1(x_k, a, \sigma)| + |\mathbb{E}R_1(x, a, \sigma) - \mathbb{E}R_1(x_k, a, \sigma)|$. Por otra parte, como g es Lipschitz de orden η_1 , tenemos que $|W_{i,a,\sigma}(x_1) - W_{i,a,\sigma}(x_2)| \leq C_g \|\psi'\|_{\infty} d(x_1, x_2)^{\eta_1} / A$ para todo a y $A < \sigma < B$ donde C_g es la constante de Lipschitz de g . Luego, si $d(x_k, x) \leq \rho_n < 1$ tenemos que

$$|R_1(x, a, \sigma) - R_1(x_k, a, \sigma)| \leq \|\psi\|_{\infty} \frac{1}{n\phi(h)} \sum_{i=1}^n |K_i(x) - K_i(x_k)| + \frac{C_g}{A} \|\psi'\|_{\infty} \frac{\rho_n^{\eta_1}}{n\phi(h)} \sum_{i=1}^n |K_i(x_k)|.$$

Luego, si $\tilde{A}_1 = C_g \|K\|_{\infty} \|\psi'\|_{\infty} / A$

$$\begin{aligned} \max_{1 \leq k \leq \ell} \sup_{x \in B(x_k, \rho_n) \cap S_{\mathcal{H}}} \sup_{|a| \leq \delta} \sup_{A \leq \sigma \leq B} |R_1(x, a, \sigma) - R_1(x_k, a, \sigma)| &\leq \tilde{A}_1 \max_{1 \leq k \leq \ell} \frac{\rho_n^{\eta_1}}{n\phi(h)} \sum_{i=1}^n \mathbb{I}_{B(x_k, \rho)}(X_i) \\ &+ \|\psi\|_{\infty} \max_{1 \leq k \leq \ell} \sup_{x \in B(x_k, \rho_n) \cap S_{\mathcal{H}}} \frac{1}{n\phi(h)} \sum_{i=1}^n |K_i(x) - K_i(x_k)|. \end{aligned}$$

Si $\underline{\eta} \geq 1$ obtenemos que $\rho_n^{\eta_1} < \rho_n$

$$\begin{aligned} \max_{1 \leq k \leq \ell} \sup_{x \in B(x_k, \rho_n) \cap S_{\mathcal{H}}} \sup_{|a| \leq \delta} \sup_{A \leq \sigma \leq B} |R_1(x, a, \sigma) - R_1(x_k, a, \sigma)| &\leq \tilde{A}_1 \max_{1 \leq k \leq \ell} \frac{\rho_n}{n\phi(h)} \sum_{i=1}^n \mathbb{I}_{B(x_k, \rho_n)}(X_i) \\ &+ \|\psi\|_{\infty} \max_{1 \leq k \leq \ell} \sup_{x \in B(x_k, \rho_n) \cap S_{\mathcal{H}}} \frac{1}{n\phi(h)} \sum_{i=1}^n |K_i(x) - K_i(x_k)| \end{aligned}$$

que ya estudiamos en la demostración del Lema 4.3.2c). Por lo tanto, tenemos que

$$\sum_{n \geq 1} \mathbb{P} \left(\sup_{A \leq \sigma \leq B} \sup_{|a| \leq \delta} \max_{1 \leq k \leq \ell} \sup_{x \in B(x_k, \rho_n) \cap S_{\mathcal{H}}} |\tilde{S}_{n,k}(x, a, \sigma)| > \epsilon_1 \frac{\theta_n}{2} \right) < \infty. \quad (4.42)$$

Si $\eta < 1$ y se cumple la condición (4.39) entonces, $\rho_n^\eta \leq C_\eta \rho_n / h$ con lo cual

$$\begin{aligned} \max_{1 \leq k \leq \ell} \sup_{x \in B(x_k, \rho_n) \cap S_{\mathcal{H}}} \sup_{|a| \leq \delta} \sup_{A \leq \sigma \leq B} |R_1(x, a, \sigma) - R_1(x_k, a, \sigma)| &\leq \tilde{A}_1 C_\eta \max_{1 \leq k \leq \ell} \frac{\rho_n}{nh\phi(h)} \sum_{i=1}^n \mathbb{I}_{B(x_k, \rho_n)}(X_i) \\ + \|\psi\|_\infty \max_{1 \leq k \leq \ell} \sup_{x \in B(x_k, \rho_n) \cap S_{\mathcal{H}}} \frac{1}{n\phi(h)} \sum_{i=1}^n |K_i(x) - K_i(x_k)|. \end{aligned}$$

y (4.42) se deduce de lo probado en la demostración del Lema 4.3.2c).

Si $\eta < 1$ y se cumple la condición (4.40) entonces la sucesión $\theta_n^{-1} \rho_n^\eta$ es acotada, luego si definimos

$$Z_i = \frac{\rho_n^\eta}{n\phi(h)} \mathbb{I}_{B(x_k, \rho_n)}(X_i)$$

argumentos análogos a los dados en el Lema 4.3.2c) permiten mostrar que existe \tilde{c} tal que para $\epsilon_1 > \tilde{c}$

$$\sum_{n \geq 1} \mathbb{P} \left(\max_{1 \leq k \leq \ell} \frac{\rho_n^\eta}{n\phi(h)} \sum_{i=1}^n \mathbb{I}_{B(x_k, \rho_n)}(X_i) > \frac{\epsilon_1}{4} \theta_n \right) < \infty,$$

lo que prueba (4.42).

Falta acotar solamente π_n donde

$$\pi_n = \mathbb{P} \left(\sup_{A \leq \sigma \leq B} \sup_{|a| \leq \delta} \max_{1 \leq k \leq \ell} |S_n(x_k, a, \sigma)| > \epsilon_1 \frac{\theta_n}{2} \right).$$

Consideremos ahora un cubrimiento de $|a| \leq \delta$ por intervalos $I_j^{(a)} = [a_j, a_{j+1}]$ tales que $|a_{j+1} - a_j| < \nu_n$ con $\nu_n = 1/\sqrt{n}$. Luego, tenemos a lo sumo $N_{1, \nu_n} = 2\delta\nu_n^{-1}$ intervalos. Consideremos también un cubrimiento de $A \leq \sigma \leq B$ con intervalos $I_j^{(\sigma)} = [\sigma_j, \sigma_{j+1}]$ tales que $|\sigma_{j+1} - \sigma_j| < \nu_n$, donde nuevamente tenemos a lo sumo $N_{2, \nu_n} = (B - A)\nu_n^{-1}$ intervalos. Observemos que $\pi_n \leq \pi_{n,1} + \pi_{n,2}$ donde

$$\pi_{n,1} = \mathbb{P} \left(\max_{1 \leq s \leq N_{2, \nu_n}} \max_{1 \leq j \leq N_{1, \nu_n}} \max_{1 \leq k \leq \ell} |S_n(x_k, a_j, \sigma_s)| > \theta_n \frac{\epsilon_1}{4} \right) \quad (4.43)$$

$$\pi_{n,2} = \mathbb{P} \left(\max_{\substack{1 \leq j \leq N_{1, \nu_n} \\ 1 \leq s \leq N_{2, \nu_n}}} \max_{1 \leq k \leq \ell} \sup_{\substack{A \leq \sigma \leq B \\ \sigma \in I_s^{(\sigma)}}} \sup_{\substack{|a| \leq \delta \\ a \in I_j^{(a)}}} |S_n(x_k, a_j, \sigma_s) - S_n(x_k, a, \sigma)| \geq \theta_n \frac{\epsilon_1}{4} \right). \quad (4.44)$$

Como $|W_{i,a,\sigma}(x)| \leq \|\psi\|_\infty$, argumentos análogos a los utilizados en el Lema 4.3.2, permiten mostrar que

$$\sup_{A \leq \sigma \leq B} \sup_{|a| \leq \delta} \max_{1 \leq k \leq \ell} \mathbb{P} \left(|S_n(x_k, a, \sigma)| > \theta_n \frac{\epsilon_1}{4} \right) \leq \exp \left\{ (1 - C\epsilon_1^2) \psi_{S_{\mathcal{H}}} \left(\frac{\log(n)}{n} \right) \right\}$$

Por lo tanto, si $C_2 = 2\delta(B - A)$, tenemos que

$$\begin{aligned} \mathbb{P} \left(\max_{\substack{1 \leq j \leq N_{1,\nu_n} \\ 1 \leq s \leq N_{2,\nu_n}}} \max_{1 \leq k \leq \ell} |S_n(x_k, a_j, \sigma_s)| > \theta_n \frac{\epsilon_1}{4} \right) &\leq 2N_{1,\nu_n} N_{2,\nu_n} N_{\rho_n}(S_{\mathcal{H}}) \exp \left\{ -C\epsilon_1^2 \psi_{S_{\mathcal{H}}} \left(\frac{\log(n)}{n} \right) \right\} \\ &\leq 2C_2 \nu_n^{-2} \exp \left\{ (1 - C\epsilon_1^2) \psi_{S_{\mathcal{H}}} \left(\frac{\log(n)}{n} \right) \right\}. \end{aligned}$$

Si elegimos ϵ_1 tal que $(1 - C\epsilon_1^2) > 1 - \beta$ y como $\nu_n = 1/\sqrt{n}$ tenemos que

$$\pi_{n,1} \leq 2C_2 \nu_n^{-2} \exp \left\{ (1 - \beta) \psi_{S_{\mathcal{H}}} \left(\frac{\log(n)}{n} \right) \right\} \leq 2C_2 n \exp \left\{ (1 - \beta) \psi_{S_{\mathcal{H}}} \left(\frac{\log(n)}{n} \right) \right\}. \quad (4.45)$$

Por lo tanto, (4.38) implica que $\sum_{n \geq 1} \pi_{n,1} < \infty$.

Falta mostrar que $\sum_{n \geq 1} \pi_{n,2} < \infty$. Observemos que si $a \in I_j^{(a)}$, para todo $A < \sigma < B$

$$|R_1(x_k, a_j, \sigma) - R_1(x_k, a, \sigma)| \leq \frac{\|\psi'\|_\infty}{A} \frac{\nu_n}{n\phi(h)} \sum_{i=1}^n K_i(x_k),$$

Por otra parte, si $\sigma \in I_s^{(\sigma)}$ llamando $\zeta(t) = t\psi'(t)$ obtenemos

$$|R_1(x_k, a_j, \sigma_s) - R_1(x_k, a_j, \sigma)| \leq \frac{\|\zeta\|_\infty}{A} \frac{\nu_n}{n\phi(h)} \sum_{i=1}^n K_i(x_k).$$

Por lo tanto, tenemos que

$$\begin{aligned} |R_1(x_k, a, \sigma) - R_1(x_k, a_j, \sigma_s)| &\leq |R_1(x_k, a, \sigma) - R_1(x_k, a_j, \sigma)| + |R_1(x_k, a_j, \sigma) - R_1(x_k, a_j, \sigma_s)| \\ &\leq \left(\frac{\|\psi'\|_\infty}{A} + \frac{\|\zeta\|_\infty}{A} \right) \frac{\nu_n}{n\phi(h)} \sum_{i=1}^n K_i(x_k) \end{aligned}$$

de donde si llamamos

$$C_3 = \left(\frac{\|\psi'\|_\infty}{A} + \frac{\|\zeta\|_\infty}{A} \right)$$

y observamos que $(1/n\phi(h)) \sum_{i=1}^n K_i(x_k) = R_0(x_k)$ deducimos que

$$\begin{aligned} |S_n(x_k, a, \sigma) - S_n(x_k, a_j, \sigma_s)| &\leq |R_1(x_k, a, \sigma) - R_1(x_k, a_j, \sigma_s)| + \mathbb{E}|R_1(x_k, a, \sigma) - R_1(x_k, a_j, \sigma_s)| \\ &\leq C_3 \frac{\nu_n}{n\phi(h)} \sum_{i=1}^n K_i(x_k) + C_3 \frac{\nu_n}{\phi(h)} \mathbb{E}K_1(x_k) \\ &\leq C_3 \nu_n |R_0(x_k) - \mathbb{E}R_0(x_k)| + 2C_3 \frac{\nu_n}{\phi(h)} \mathbb{E}K_1(x_k) \end{aligned}$$

Usando que $\mathbb{E}K_1(x_k) \leq C'\phi(h)$ y que, como $\nu_n = 1/\sqrt{n}$, $\theta_n^{-1}\eta_n \rightarrow 0$, obtenemos que existe n_0 tal que si $n \geq n_0$, $\theta_n^{-1}\nu_n \leq \min\{1/C_3, \epsilon_1/(16C'C_3)\}$. Luego, para $n \geq n_0$ tenemos la acotación

$$\begin{aligned} \theta_n^{-1}|S_n(x_k, a, \sigma) - S_n(x_k, a_j, \sigma_s)| &\leq |R_1(x_k, a, \sigma) - R_1(x_k, a_j, \sigma_s)| + \mathbb{E}|R_1(x_k, a, \sigma) - R_1(x_k, a_j, \sigma_s)| \\ &\leq C_3 \theta_n^{-1} \nu_n |R_0(x_k) - \mathbb{E}R_0(x_k)| + 2C'C_3\theta_n^{-1}\eta_n \\ &\leq |R_0(x_k) - \mathbb{E}R_0(x_k)| + \frac{\epsilon_1}{8}. \end{aligned}$$

Por lo tanto, el miembro derecho de la última desigualdad no depende de a y σ lo que implica que

$$\pi_{n,2} \leq \mathbb{P}\left(\max_{1 \leq k \leq \ell} |R_0(x_k) - \mathbb{E}R_0(x_k)| > \frac{\epsilon_1}{8}\right).$$

Recordemos que para todo $\epsilon_0 > 0$

$$\sum_{n \geq 1} \mathbb{P}\left(\sup_{x \in S_{\mathcal{X}^c}} |R_0(x) - \mathbb{E}R_0(x)| > \epsilon_0\right) < \infty$$

de donde $\sum_{n \geq 1} \pi_{n,2} < \infty$. \square

Capítulo 5

Un resultado auxiliar sobre el número de cubrimiento

En este Capítulo daremos un resultado auxiliar que nos resultará de utilidad para obtener la distribución asintótica de los estimadores robustos en modelos parcialmente lineales funcionales.

5.1. El resultado

Sea $S_{\mathcal{H}} \subset \mathcal{H}$ compacto.

Lema 5.1.1. *Para cualquier $\delta > 0$ y medida de probabilidad \mathbb{Q} , se cumple que*

$$\begin{aligned}\log N(2\delta, \mathcal{C}, L^2(\mathbb{Q})) &\leq \frac{2}{\delta} + C N_{\delta}(S_{\mathcal{H}}), \\ \log N_{[\cdot]}(4\delta, \mathcal{C}, L^2(\mathbb{Q})) &\leq \frac{2}{\delta} + C N_{\delta}(S_{\mathcal{H}}),\end{aligned}$$

donde

$$\mathcal{C} = \{g : S_{\mathcal{H}} \subset \mathcal{H} \rightarrow \mathbb{R} \text{ continua} : \|g\|_{\infty} \leq 1 \text{ y } |g(y) - g(x)| \leq d(x, y) \\ \forall x \in S_{\mathcal{H}} \forall y \in S_{\mathcal{H}}\}.$$

y C es una constante que no depende ni de \mathbb{Q} ni de $S_{\mathcal{H}}$.

DEMOSTRACIÓN. Como $S_{\mathcal{H}}$ es compacto, dado $\delta > 0$, existen $x_1, \dots, x_{\ell} \in \mathcal{H}$ tales que

$$S_{\mathcal{H}} \subset \bigcup_{i=1}^{\ell} B(x_i, \delta),$$

donde $\ell = N_\delta(S_{\mathcal{H}})$. Sin pérdida de generalidad podemos suponer $x_i \in S_{\mathcal{H}}$. Elijamos los índices en x_1, \dots, x_ℓ de modo tal que dado $j > 1 \exists i < j$ tal que $B(x_i, \delta) \cap B(x_j, \delta) \neq \emptyset$, o sea, $d(x_i, x_j) < 2\delta$.

Sea $B_1 = B(x_1, \delta)$; $B_2 = B(x_2, \delta) - B_1$; $B_j = B(x_j, \delta) - \bigcup_{i=1}^{j-1} B_i$. Luego $S_{\mathcal{H}} \subset \bigcup_{i=1}^{\ell} B_i$ y $B_i \cap B_j = \emptyset$ si $i \neq j$. Más aún, $S_{\mathcal{H}} = \bigcup_{i=1}^{\ell} A_i$ con $A_i = B_i \cap S_{\mathcal{H}}$.
Para cada $g \in \mathcal{C}$ definamos la función

$$\tilde{g}(x) = \sum_{k=1}^{\ell} \delta \left\lfloor \frac{g(x_k)}{\delta} \right\rfloor \mathbb{I}_{A_k}(x),$$

donde $\lfloor w \rfloor$ indica la parte entera de w .

Luego, la función \tilde{g} resulta constante sobre cada conjunto A_k y solamente toma valores en el conjunto $\{\delta m, m \in \mathbb{Z}\}$. Además, se tiene que $\|\tilde{g} - g\|_\infty < 2\delta$. Efectivamente, dado $x \in S_{\mathcal{H}}$,

i) Como

$$\frac{g(x_k)}{\delta} = \left\lfloor \frac{g(x_k)}{\delta} \right\rfloor + d \text{ con } 0 \leq d < 1,$$

tenemos que

$$\delta \frac{g(x_k)}{\delta} = \delta \left\lfloor \frac{g(x_k)}{\delta} \right\rfloor + \delta d \text{ con } 0 \leq d < 1.$$

Obtenemos entonces que $g(x_k) = \delta \lfloor g(x_k)/\delta \rfloor + \delta d$ con $0 \leq d < 1$, por lo tanto

$$\left| g(x_k) - \delta \left\lfloor \frac{g(x_k)}{\delta} \right\rfloor \right| < \delta.$$

ii) Sea $x \in A_k$, luego $x \in B(x_k, \delta)$. Por lo tanto, observando que

$$\left| \delta \left\lfloor \frac{g(x_k)}{\delta} \right\rfloor - g(x) \right| \leq \left| \delta \left\lfloor \frac{g(x_k)}{\delta} \right\rfloor - g(x_k) \right| + |g(x_k) - g(x)|,$$

y usando que $|g(x_k) - g(x)| \leq d(x_k, x) < \delta$ y lo probado en i), se tiene que

$$\left| \delta \left\lfloor \frac{g(x_k)}{\delta} \right\rfloor - g(x) \right| < 2\delta.$$

iii) Dado $x \in S_{\mathcal{H}}$, como $\sum_{k=1}^{\ell} \mathbb{I}_{A_k} = \mathbb{I}_{S_{\mathcal{H}}}$, $A_k \cap A_j = \emptyset$, existe un único $k = k_x$ tal que $x \in A_k$ luego por ii)

$$|\tilde{g}(x) - g(x)| = \left| \delta \left\lfloor \frac{g(x_k)}{\delta} \right\rfloor - g(x) \right| < 2\delta.$$

Es decir, para todo $x \in S_{\mathcal{H}}$, $|\tilde{g}(x) - g(x)| < 2\delta$.

Como $|g(x_1)| \leq 1$, $[g(x_1)/\delta] \leq [1/\delta]$, por lo tanto, $[g(x_1)/\delta]$ toma a lo sumo $2[1/\delta] + 1$ valores.

Ahora bien, dado $k > 1$ existe $j < k$ $d(x_j, x_k) < 2\delta$ por lo tanto,

$$|\tilde{g}(x_k) - \tilde{g}(x_j)| \leq |\tilde{g}(x_k) - g(x_k)| + |g(x_k) - g(x_j)| + |g(x_j) - \tilde{g}(x_j)| < 6\delta.$$

Luego, una vez obtenido $\tilde{g}(x_j)$, $\tilde{g}(x_k)$ sólo puede tomar valores entre $\tilde{g}(x_j) - 6\delta$ y $\tilde{g}(x_j) + 6\delta$, es decir, como

$$\tilde{g}(x_k) = \delta \left\lfloor \frac{g(x_k)}{\delta} \right\rfloor \quad \text{y} \quad \tilde{g}(x_j) = \delta \left\lfloor \frac{g(x_j)}{\delta} \right\rfloor,$$

$(1/\delta)\tilde{g}(x_k)$ puede tomar sólo 13 valores enteros entre

$$-\left\lfloor \frac{g(x_j)}{\delta} \right\rfloor - 6 \quad \text{y} \quad \left\lfloor \frac{g(x_j)}{\delta} \right\rfloor + 6.$$

Resumiendo si $M_k = [g(x_k)/\delta]$,

$$\tilde{g}(x) = \sum_{k=1}^{\ell} \delta M_k \mathbb{I}_{B_k}(x)$$

donde M_1 toma $N = 2 \left\lfloor \frac{1}{\delta} \right\rfloor + 1$ valores, una vez elegido M_{k-1} , sólo tenemos 13 posibles elecciones para el valor M_k , es decir, tengo un total de a lo sumo $N13^\ell$. Es decir, el número de posibles funciones \tilde{g} está mayorado por $(2[1/\delta] + 1) 13^\ell$. Por lo tanto,

$$N(2\delta, \mathcal{C}, \|\cdot\|_\infty) \leq \left(2 \left\lfloor \frac{1}{\delta} \right\rfloor + 1 \right) 13^\ell.$$

Usando que $N(2\delta, \mathcal{C}, L^2(\mathbb{Q})) \leq N(2\delta, \mathcal{C}, \|\cdot\|_\infty)$ y como $\ell = N_\delta(S_{\mathcal{H}})$ obtenemos que

$$\log N(2\delta, \mathcal{C}, L^2(\mathbb{Q})) \leq \log \left(\frac{2}{\delta} + 1 \right) + \log(13) N_\delta(S_{\mathcal{H}}) \leq \frac{2}{\delta} + \log(13) N_\delta(S_{\mathcal{H}}).$$

Sea $h = \tilde{g}$ una de las posibles funciones que construimos. Para esa función h , definamos $U_h(x) = h(x) + 2\delta$ y $L_h(x) = h(x) - 2\delta$. Como para cada $g \in \mathcal{C}$, existe \tilde{g} tal que $\|\tilde{g} - g\|_\infty < 2\delta$, de donde $\tilde{g}(x) - 2\delta < g(x) < \tilde{g}(x) + 2\delta$. Luego, con cada una de las funciones h construimos un bracket de tamaño 4δ de la forma $[L_h(x), U_h(x)]$ y con estos brackets cubrimos a \mathcal{C} . Obtenemos entonces que

$$N_{[\cdot]}(4\delta, \mathcal{C}, \|\cdot\|_\infty) \leq \left(2 \left\lfloor \frac{1}{\delta} \right\rfloor + 1 \right) 13^\ell.$$

Como además cada bracket en $\|\cdot\|_\infty$ tiene un tamaño en $L^2(\mathbb{Q})$ de 4δ ,

$$N_{[\cdot]}(4\delta, \mathcal{C}, L^2(\mathbb{Q})) \leq \left(2 \left\lfloor \frac{1}{\delta} \right\rfloor + 1 \right) 13^\ell.$$

y por lo tanto, para cualquier medida de probabilidad \mathbb{Q} ,

$$\log N_{\lfloor \rfloor}(4\delta, \mathcal{C}, L^2(\mathbb{Q})) \leq \log\left(\frac{2}{\delta} + 1\right) + \log(13)N_\delta(S_{\mathcal{H}}) \leq \frac{2}{\delta} + \log(13)N_\delta(S_{\mathcal{H}})$$

lo que concluye la demostración del Lema. \square

Del Lema 5.1.1 se deduce inmediatamente el siguiente resultado

Lema 5.1.2. *Sea $S_{\mathcal{H}} \subset \mathcal{H}$ compacto. Para cualquier $\delta > 0$ y medida de probabilidad \mathbb{Q} , se cumple que*

$$\begin{aligned} \log N(2\delta, \mathcal{C}_M, L^2(\mathbb{Q})) &\leq \frac{2M}{\delta} + C N_{\frac{\delta}{M}}(S_{\mathcal{H}}), \\ \log N_{\lfloor \rfloor}(4\delta, \mathcal{C}_M, L^2(\mathbb{Q})) &\leq \frac{2M}{\delta} + C N_{\frac{\delta}{M}}(S_{\mathcal{H}}), \end{aligned}$$

donde

$$\begin{aligned} \mathcal{C}_M = \{g : S_{\mathcal{H}} \subset \mathcal{H} \rightarrow \mathbb{R} \text{ continua} : \|g\|_\infty \leq M \text{ y } |g(y) - g(x)| \leq M d(x, y) \\ \forall x \in S_{\mathcal{H}} \forall y \in S_{\mathcal{H}}\}. \end{aligned} \quad (5.1)$$

y C es una constante que no depende ni de \mathbb{Q} ni de $S_{\mathcal{H}}$.

Para probar el Lema 5.1.2, basta usar que $N(2\delta, \mathcal{C}_M, L^2(\mathbb{Q})) = N(2\delta/M, \mathcal{C}, L^2(\mathbb{Q}))$ y $N_{\lfloor \rfloor}(4\delta, \mathcal{C}_M, L^2(\mathbb{Q})) = N_{\lfloor \rfloor}(4\delta/M, \mathcal{C}, L^2(\mathbb{Q}))$.

En el caso particular, en que \mathcal{H} sea un espacio normado obtenemos el siguiente resultado.

Lema 5.1.3. *Sea $S_{\mathcal{H}} \subset \mathcal{H}$ compacto y convexo. Para cualquier $\delta > 0$ y medida de probabilidad \mathbb{Q} , se cumple que*

$$\begin{aligned} \log N(2\delta, \mathcal{C}_M, L^2(\mathbb{Q})) &\leq \frac{2M}{\delta} + C N_{\frac{\delta}{M}}(S_{\mathcal{H}}), \\ \log N_{\lfloor \rfloor}(4\delta, \mathcal{C}_M, L^2(\mathbb{Q})) &\leq \frac{2M}{\delta} + C N_{\frac{\delta}{M}}(S_{\mathcal{H}}), \end{aligned}$$

donde

$$\begin{aligned} \mathcal{C}_M = \{g : S_{\mathcal{H}} \subset \mathcal{H} \rightarrow \mathbb{R} \text{ continua} : \|g\|_\infty \leq M \text{ y } \|g'_x\| \leq M \\ \forall x \in S_{\mathcal{H}} \text{ donde } g'_x \text{ es la derivada de Hadamard de } g \text{ en } x\}. \end{aligned}$$

DEMOSTRACIÓN. Sin pérdida de generalidad supongamos $M = 1$. Como $S_{\mathcal{H}}$ es compacto, dado $\delta > 0$, existen $x_1, \dots, x_\ell \in \mathcal{H}$ tales que $S_{\mathcal{H}} \subset \bigcup_{i=1}^{\ell} B(x_i, \delta)$, donde $\ell = N_\delta(S_{\mathcal{H}})$. Como en la demostración del Lema 5.1.1, supongamos que $x_i \in S_{\mathcal{H}}$. Elijamos los índices en x_1, \dots, x_ℓ de modo tal que dado $j > 1 \exists i < j$ tal que $B(x_i, \delta) \cap B(x_j, \delta) \neq \emptyset$, o sea, $d(x_i, x_j) < 2\delta$.

Definamos $\phi_i(t) = tx + (1-t)x_i$, con $t \in [0, 1]$ entonces, $\phi_i(0) = x_i$ y $\phi_i(1) = x$. Para $\psi \in \mathcal{C}_M$ como $Im(\phi_i) \subset S_{\mathcal{H}}$, definamos las funciones $g_i : [0, 1] \rightarrow \mathbb{R}$, $g_i(t) = \psi(\phi_i(t))$.

Observemos que g_i es una función continua y derivable, pues tanto ψ como ϕ_i son Hadamard diferenciables. Luego, por el Teorema del valor medio existe ξ_i entre 0 y 1 tal que $g_i(1) = g_i(0) + g'_i(\xi_i)(1 - 0)$, es decir,

$$\psi(x) = \psi(x_i) + g'_i(\xi_i)$$

donde, por regla de la cadena, $g'_i(\xi_i)1 = \psi'_{\phi_i(\xi_i)}(\phi'_i(\xi_i)1) = \psi'_{\phi_i(\xi_i)}(x - x_i)$. Luego, $|\psi(x) - \psi(x_i)| \leq d(x, x_i)$. La conclusión se obtiene ahora como en el Lema 5.1.1. \square

Capítulo 6

Modelo parcialmente lineal en dimensión infinita

En este capítulo, combinaremos la flexibilidad de un modelado parcialmente lineal con el tratamiento no paramétrico de datos funcionales.

Como en Aneiros-Pérez y Vieu (2006), trabajaremos con el modelo de regresión parcialmente lineal semi-funcional (PLSF), que supone se tienen n observaciones i.i.d. $(Y_i, \mathbf{Z}_i^T, X_i)^T$, $i = 1, \dots, n$, tales que

$$Y_i = \sum_{j=1}^p Z_{ij}\beta_j + g(X_i) + \epsilon_i = \boldsymbol{\beta}^T \mathbf{Z}_i + g(X_i) + \epsilon_i \quad (6.1)$$

donde Z_{ij} ($j = 1, \dots, p$) son variables explicativas reales, X_i es otra variable explicativa pero de naturaleza funcional, los errores ϵ_i son independientes e idénticamente distribuidos e independientes de $(\mathbf{Z}_i^T, X_i)^T$, $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^T$ es un vector de parámetros reales desconocidos y $g : \mathcal{H} \rightarrow \mathbb{R}$ es un operador suave desconocido que no tiene por qué ser lineal.

Ya mencionamos que los espacios más interesantes para modelar con variables funcionales son los espacios semimétricos. Por eso consideraremos que X toma valores en algún espacio abstracto semi-métrico \mathcal{H} y denotamos por $d(\cdot, \cdot)$ a la semimétrica asociada.

En este Capítulo estudiaremos el modelo de regresión PLSF. Más precisamente, estaremos interesados en estimar el operador de regresión g y el parámetro desconocido multivariado $\boldsymbol{\beta}$.

6.1. Estimadores clásicos

Supongamos que $(Y, \mathbf{Z}^T, X)^T$ es un vector aleatorio con la misma distribución que $(Y_i, \mathbf{Z}_i^T, X_i)^T$, es decir,

$$Y = \boldsymbol{\beta}^T \mathbf{Z} + g(X) + \epsilon$$

donde ϵ es independiente de $(\mathbf{Z}^T, X)^T$. Aneiros-Pérez y Vieu (2006) suponen que $\mathbb{E}(\epsilon) = 0$. Luego, usando que ϵ es independiente de $(\mathbf{Z}^T, X)^T$ y definiendo $\phi_0(x) = \mathbb{E}(Y|X = x)$, $\phi_j(x) = \mathbb{E}(Z_j|X = x)$ y $\boldsymbol{\phi}(x) = (\phi_1(x), \dots, \phi_p(x))^T$, obtenemos, tomando esperanza condicional, que $g(x) = \phi_0(x) - \boldsymbol{\beta}^T \boldsymbol{\phi}(x)$, de donde

$$Y - \phi_0(x) = \boldsymbol{\beta}^T (\mathbf{Z} - \boldsymbol{\phi}(x)) + \epsilon.$$

Esto sugiere que si proponemos a priori estimadores de $\phi_0(x)$ y $\boldsymbol{\phi}(x)$, que indicaremos $\hat{\phi}_0(x)$ y $\hat{\boldsymbol{\phi}}(x)$ respectivamente, podemos obtener un estimador para $\boldsymbol{\beta}$ y, finalmente, un estimador para g .

En el caso clásico se estima las esperanzas condicionales por el estimador de Nadaraya-Watson, o sea, por

$$\hat{\phi}_0(x) = \sum_{i=1}^n w_i(x) Y_i \quad \hat{\boldsymbol{\phi}}_j(x) = \sum_{i=1}^n w_i(x) Z_{ij},$$

donde $w_i(x)$ son los pesos definidos en (4.3). Por otra parte, los estimadores de $\boldsymbol{\beta}$ y g propuestos por Aneiros-Pérez y Vieu (2006) se obtienen utilizando un estimador de mínimos cuadrados sobre los residuos, o sea,

$$\begin{cases} \hat{\boldsymbol{\beta}} &= (\tilde{\mathbf{Z}}_h^T \tilde{\mathbf{Z}}_h)^{-1} \tilde{\mathbf{Z}}_h^T \tilde{\mathbf{Y}}_h \\ \hat{g}(x) &= \sum_{i=1}^n w_i(x) (Y_i - \mathbf{Z}_i^T \hat{\boldsymbol{\beta}}), \end{cases}$$

donde $\tilde{\mathbf{Y}}_h = (\mathbf{I}_n - \mathbf{W}_h) \mathbf{Y}$ y $\tilde{\mathbf{Z}}_h = (\mathbf{I}_n - \mathbf{W}_h) \mathbf{Z}$, $\mathbf{I}_n \in \mathbb{R}^{n \times n}$ la matriz identidad y \mathbf{W}_h la matriz de pesos $(\mathbf{W}_h)_{ij} = w_j(X_i)$.

Como en el Capítulo 4, h es el parámetro de suavizado que aparece típicamente en cualquier enfoque de estimación no paramétrico.

Bajo ciertas hipótesis, Aneiros-Pérez y Vieu (2006) obtienen resultados de convergencia sobre compactos del estimador \hat{g} :

$$\sup_{x \in S_{\mathcal{X}}} |\hat{g}(x) - g(x)| = O(h^\alpha) + O_{a.s.} \left(\sqrt{\frac{\log(n)}{n\phi(h)}} \right)$$

donde ϕ es una función que controla la probabilidad de la pequeña bola dada en **H1** y α es una constante positiva relacionada con las hipótesis utilizadas. Además, obtienen resultados sobre el comportamiento asintótico del estimador de la componente paramétrica del modelo extendiendo resultados previos establecidos para el caso no funcional.

Como se puede apreciar el estimador \widehat{g} está basado en un promedio de las variables respuesta por lo que resulta muy sensible a observaciones atípicas, particularmente a aquellas que se encuentran en el entorno del punto X . Por otra parte, para la estimación de β se considera un estimador de mínimos cuadrados sobre los residuos, por lo que dicho estimador también será sensible a la presencia de datos anómalos.

6.2. Estimadores robustos

Los métodos estadísticos robustos tienen como objetivo permitir inferencias válidas cuando el modelo no se cumple exactamente. Teniendo en cuenta la sensibilidad de los estimadores de (β, g) definidos por Aneiros-Pérez y Vieu (2006), introduciremos estimadores robustos para el modelo PLSF extendiendo una propuesta en tres pasos dada en el caso finito dimensional por Bianco y Boente (2004) al caso funcional utilizando el M -estimador local funcional definido en el Capítulo 4.

Este enfoque no requiere momentos de los errores y está basado, como en el Capítulo 4, en la extensión de la noción de esperanza condicional a la situación en que no existen momentos dada en Boente y Fraiman (1989). Sea $\psi : \mathbb{R} \rightarrow \mathbb{R}$ una función de escores acotada y continua y sean $F_0(y|X = x)$ y $F_j(z|X = x)$ las distribuciones condicionales de $Y|X = x$ y de $Z_j|X = x$, respectivamente. Sean $\phi_j(x)$, $0 \leq j \leq p$ los funcionales de posición condicional relacionados a un suavizador robusto. Más precisamente, sean para cada $x \in \mathcal{H}$, $\phi_j(x)$, $0 \leq j \leq p$ las soluciones de

$$\begin{aligned} \mathbb{E} \left[\psi \left(\frac{Y - \phi_0(x)}{\sigma_0(x)} \right) \middle| X = x \right] &= 0 \\ \mathbb{E} \left[\psi \left(\frac{Z_j - \phi_j(x)}{\sigma_j(x)} \right) \middle| X = x \right] &= 0, \quad 1 \leq j \leq p. \end{aligned}$$

donde $\sigma_0(x)$ and $\sigma_j(x)$ son la MAD de las distribuciones condicionales $Y|X = x$ y de $Z_j|X = x$, respectivamente.

Los estimadores en tres pasos pueden definirse como sigue

Paso 1: Estimemos $\phi_0(x)$ y $\phi_j(x)$ con un suavizador robusto. Denotemos con $\widehat{\phi}_0(x)$ y $\widehat{\phi}_j(x)$ a los estimadores obtenidos y $\widehat{\phi}(x) = (\widehat{\phi}_1(x), \dots, \widehat{\phi}_p(x))^T$.

Paso 2: El estimador de β , $\widehat{\beta}$, se obtiene utilizando cualquier estimador de regresión robusto sobre los residuos $Y_i - \widehat{\phi}_0(X_i)$ y $\mathbf{Z}_i - \widehat{\phi}(X_i)$.

Paso 3: El estimador de la función de regresión g se define como $\widehat{g}(x) = \widehat{\phi}_0(x) - \widehat{\beta}^T \widehat{\phi}(x)$.

Como en Boente *et al.* (2006), en el **Paso 3**, podemos considerar un estimador alternativo de la función de regresión g suavizando robustamente los residuos $Y_i - \widehat{\beta}^T \mathbf{Z}_i$.

Esto puede hacerse utilizando también estimadores robustos basados en núcleos como en el Capítulo 4. Para ser más precisos, podemos definir un nuevo estimador de $g(x)$ que indicaremos $\widehat{g}(x)$ como la solución de $\widehat{\lambda}(x, a, \widehat{s}(x)) = 0$, donde

$$\widehat{\lambda}(x, a, \sigma) = \sum_{i=1}^n K\left(\frac{d(x, X_i)}{h}\right) \psi\left(\frac{Y_i - \widehat{\beta}^T \mathbf{Z}_i - a}{\sigma}\right), \quad (6.2)$$

y $\widehat{s}(x)$ es un estimador robusto de escala como, por ejemplo, la MAD de la distribución empírica condicional

$$\widehat{F}(u|X = x) = \sum_{i=1}^n K\left(\frac{d(x, X_i)}{h}\right) \mathbb{I}_{(-\infty, u]}(Y_i - \widehat{\beta}^T \mathbf{Z}_i) \left\{ \sum_{i=1}^n K\left(\frac{d(x, X_i)}{h}\right) \right\}^{-1}.$$

Sin embargo, un parámetro de suavizado h distinto del utilizado en el **Paso 1** puede ser preferible ya que los residuos $Y_i - \widehat{\beta}^T \mathbf{Z}_i$ tienen menor variabilidad que las variables originales Y_i . Estos estimadores son los que consideraremos en el estudio de simulación que describiremos en el Capítulo 8.

En el **Paso 1**, los suavizadores robustos pueden elegirse como las medianas locales $\widehat{\phi}_{0, \text{MED}}(x)$ y $\widehat{\phi}_{j, \text{MED}}(x)$, es decir, como la mediana de las funciones de distribución condicionales empíricas $\widehat{F}_0(y|X = x)$ y $\widehat{F}_j(z|X = x)$, correspondientes a los definidos en (3.3). Es decir, $\widehat{F}_0(y|X = x)$ y $\widehat{F}_j(z|X = x)$ están definidas como

$$\begin{aligned} \widehat{F}_0(y|X = x) &= \sum_{i=1}^n w_i(x) \mathbb{I}_{(-\infty, y]}(Y_i), \\ \widehat{F}_j(z|X = x) &= \sum_{i=1}^n w_i(x) \mathbb{I}_{(-\infty, z]}(Z_{ij}) \quad 1 \leq j \leq p, \end{aligned}$$

donde $w_i(x)$ son los pesos definidos en (4.3). Observemos que $\widehat{F}_0(y|X = x)$ y $\widehat{F}_j(z|X = x)$ proveen estimadores de las funciones de distribución condicional, $F_0(y|T = t)$ y $F_j(z|X = x)$, de $Y|X = t$ y $Z_j|X = x$, respectivamente.

Por otro lado, en el **Paso 1**, también podemos considerar M -estimadores locales de la regresión de Y versus X y de cada componente de Z versus X , $\widehat{\phi}_{0, M}(x)$ y $\widehat{\phi}_{j, M}(x)$ respectivamente. Como en el Capítulo 4, los estimadores están definidos implícitamente como la solución única de $\widehat{\lambda}_0(x, \widehat{\phi}_{0, M}(x), \widehat{s}_0(x)) = 0$ y $\widehat{\lambda}_j(x, \widehat{\phi}_{j, M}(x), \widehat{s}_j(x)) = 0$ donde

$$\begin{aligned} \widehat{\lambda}_0(x, a, \sigma) &= \sum_{i=1}^n w_i(x) \psi\left(\frac{Y_i - a}{\sigma}\right) \\ \widehat{\lambda}_j(x, a, \sigma) &= \sum_{i=1}^n w_i(x) \psi\left(\frac{Z_{ij} - a}{\sigma}\right), \end{aligned}$$

donde \widehat{s}_0 y \widehat{s}_j , $1 \leq j \leq p$, son estimadores locales de escala robustos, tales como la MAD de las distribuciones $\widehat{F}_0(y|X = x)$ y $\widehat{F}_j(z|X = x)$, respectivamente. Posibles elecciones para la función ψ son la función de Huber que es monótona o la bicuadrada que como se mencionó en la observación i) de la Sección 4.3.2, no cumple **H6**.

Como se describe en el **Paso 2**, una vez que uno obtiene estimadores robustos de $\phi_0(x)$ y $\phi_j(x)$, $\widehat{\phi}_0(x)$ y $\widehat{\phi}_j(x)$, el estimador robusto del parámetro de regresión puede obtenerse aplicando a los residuos $\widehat{r}_i = Y_i - \widehat{\phi}_0(X_i)$ y $\widehat{\mathbf{U}}_i = \mathbf{Z}_i - \widehat{\boldsymbol{\phi}}(X_i)$ cualquiera de los métodos robustos propuestos para regresión lineal (ver Maronna *et al.*, 2006).

Capítulo 7

Comportamiento asintótico de los estimadores bajo un modelo PLSF

En este Capítulo probaremos la consistencia del estimador $\hat{\beta}$ y del estimador \hat{g} definidos en los Pasos 1 a 3, bajo condiciones generales. Por otra parte, los resultados sobre la distribución asintótica del estimador de β requerirán como en el caso finito-dimensional que $\mathbb{P}(X \in S_{\mathcal{H}}) = 1$ para algún conjunto compacto $S_{\mathcal{H}} \subset \mathcal{H}$. En este caso, necesitaremos pedir algunas condiciones sobre el número de cubrimiento de $S_{\mathcal{H}}$ para asegurar la convergencia de algunos de los términos del desarrollo lineal del estimador.

7.1. Consistencia

Para obtener resultados de consistencia para los estimadores propuestos necesitamos considerar las siguientes hipótesis adicionales.

H2' Para $0 \leq j \leq p$, $F_j(\cdot|X = x)$ son simétricas alrededor de $\phi_j(x)$, respectivamente.

H3' Sea $S_{\mathcal{H}}$ un conjunto compacto de \mathcal{H} tal que

- i) $F_0(y|X = x)$ y $F_j(z|X = x)$ son funciones continuas de x en un entorno de $S_{\mathcal{H}}$, para $1 \leq j \leq p$
- ii) se verifica la siguiente condición de equicontinuidad:

$$\forall \epsilon > 0 \exists \delta > 0 : |u - v| < \delta \Rightarrow \sup_{x \in S_{\mathcal{H}}} \max_{0 \leq j \leq p} |F_j(u|X = x) - F_j(v|X = x)| < \epsilon.$$

Las conclusiones de los Teoremas 7.1.1 y 7.1.2, que son consecuencia directa de los Teoremas 4.3.4 y 4.3.6 permiten, conjuntamente con el Teorema 7.1.4, establecer la consistencia de los estimadores tanto del parámetro de regresión β como de la función g , cuando el suavizado está basado en las medianas locales o M -suavizadores locales.

Teorema 7.1.1. *Bajo las hipótesis **H1**, **H3'**, **H4**, **H5** y **H7** se tiene que*

$$\sup_{x \in S_{\mathcal{H}}} \sup_{u \in \mathbb{R}} |\widehat{F}_j(u|X=x) - F_j(u|X=z)| \xrightarrow{a.s.} 0$$

cuando $n \rightarrow \infty$, $0 \leq j \leq p$.

Proposición 7.1.2. *Supongamos que se cumplen **H1**, **H2'**, **H3'**, **H4**, **H5** y **H7**. Entonces, dado $S_{\mathcal{H}} \subset \mathcal{H}$ compacto,*

- a) *si se cumple **H6**, entonces $\sup_{x \in S_{\mathcal{H}}} |\widehat{\phi}_{j,M}(x) - \phi_j(x)| \xrightarrow{a.s.} 0$ para $0 \leq j \leq p$,*
- b) *si, $F_j(z|X=x)$ tiene una única mediana en $\phi_j(x)$, para $0 \leq j \leq p$, se tiene que $\sup_{x \in S_{\mathcal{H}}} |\widehat{\phi}_{j,MED}(x) - \phi_j(x)| \xrightarrow{a.s.} 0$, $0 \leq j \leq p$.*

El Lema 7.1.3 y el Teorema 7.1.4 son análogos al Lema 1 y Teorema 1 de Bianco y Boente (2004), respectivamente.

Lema 7.1.3. *Sea \mathcal{H} separable y completo y sean $(r_i, \mathbf{u}_i^T, x_i)^T \in \mathbb{R}^{p+1} \times \mathcal{H}$, $1 \leq i \leq n$ vectores aleatorios i.i.d. sobre (Ω, A, P) tales que $(r_i, \mathbf{u}_i^T)^T$ tienen distribución común P . Sean $\widehat{\eta}_0(x)$ y $\widehat{\eta}(x) = (\widehat{\eta}_1(x), \dots, \widehat{\eta}_p(x))^T$ funciones aleatorias tales que para cualquier conjunto compacto $S_{\mathcal{H}} \subset \mathcal{H}$*

$$\sup_{x \in S_{\mathcal{H}}} |\widehat{\eta}_j(x)| \xrightarrow{a.s.} 0, \quad 0 \leq j \leq p. \quad (7.1)$$

Definimos a P_n y Q_n como las siguientes medidas empíricas sobre \mathbb{R}^{p+1}

$$P_n(A) = \frac{1}{n} \sum_{i=1}^n \mathbb{I}_A(r_i, \mathbf{u}_i) \quad Q_n(A) = \frac{1}{n} \sum_{i=1}^n \mathbb{I}_A(r_i + \widehat{\eta}_0(x_i), \mathbf{u}_i + \widehat{\eta}(x_i)),$$

donde $A \subset \mathbb{R}^{p+1}$ es un conjunto de Borel. Entonces,

- a) *para toda función continua y acotada $f : \mathbb{R}^{p+1} \rightarrow \mathbb{R}$ se tiene que*

$$|\mathbb{E}_{Q_n}(f) - \mathbb{E}_{P_n}(f)| \xrightarrow{a.s.} 0$$

- b) *$\pi(Q_n, P) \xrightarrow{a.s.} 0$, donde π es la distancia de Prohorov.*

La demostración del Lema 7.1.3 se encuentra en el Apéndice.

Teorema 7.1.4. Sean $(Y_i, \mathbf{Z}_i^T, X_i)^T$, $1 \leq i \leq n$ vectores aleatorios independientes que satisfacen el modelo (6.1). Denotemos por P a la distribución de $(r_i, \mathbf{u}_i^T)^T = (Y_i - \phi_0(X_i), \mathbf{Z}_i^T - \boldsymbol{\phi}(X_i)^T)^T$, donde $\phi_0(x)$ y $\boldsymbol{\phi}(x)$ están definidas en (**H2'**) con $\phi_0(x) = \boldsymbol{\beta}^T \boldsymbol{\phi}(x) + g(x)$. Supongamos que \mathcal{H} es separable y completo y que $\hat{\phi}_j(x)$, $0 \leq j \leq p$ son estimadores de $\phi_j(x)$ tales que para cualquier conjunto compacto $S_{\mathcal{H}} \subset \mathcal{H}$

$$\sup_{x \in S_{\mathcal{H}}} |\hat{\phi}_j(x) - \phi_j(x)| \xrightarrow{a.s.} 0, \quad 0 \leq j \leq p. \quad (7.2)$$

Sea $\boldsymbol{\beta}(H)$ un funcional de regresión para el modelo $w = \boldsymbol{\beta}^T \mathbf{v} + \epsilon$ donde ϵ y \mathbf{v} son independientes y $(w, \mathbf{v}^T)^T \sim H$. Supongamos que $\boldsymbol{\beta}(H)$ es continuo en P y que además provee estimadores Fisher consistentes.

Sean $\hat{P}_n(A) = (1/n) \sum_{i=1}^n \mathbb{I}_A(\hat{r}_i, \hat{\mathbf{u}}_i)$ con $\hat{r}_i = Y_i - \hat{\phi}_0(X_i)$ y $\hat{\mathbf{u}}_i = \mathbf{Z}_i - \hat{\boldsymbol{\phi}}(X_i)$, donde $\hat{\boldsymbol{\phi}}(x) = (\hat{\phi}_1(x), \dots, \hat{\phi}_p(x))^T$ y $\hat{\boldsymbol{\beta}}_R = \boldsymbol{\beta}(\hat{P}_n)$. Entonces, se tiene que

$$\hat{\boldsymbol{\beta}}_R \xrightarrow{a.s.} \boldsymbol{\beta}.$$

DEMOSTRACIÓN. Sea

$$P_n(A) = \frac{1}{n} \sum_{i=1}^n \mathbb{I}_A(r_i, \mathbf{u}_i).$$

Usando (7.2) y el Lema 7.1.3, tenemos que $|\mathbb{E}_{\hat{P}_n}(f) - \mathbb{E}_{P_n}(f)| \xrightarrow{a.s.} 0$, más aún, como $\pi(P_n, P) \xrightarrow{a.s.} 0$ obtenemos que $\pi(\hat{P}_n, P) \xrightarrow{a.s.} 0$.

Usando que $\boldsymbol{\beta}(H)$ es continua en P , obtenemos que $\boldsymbol{\beta}(\hat{P}_n) \xrightarrow{a.s.} \boldsymbol{\beta}(P)$. Para concluir la demostración basta observar que $\hat{\boldsymbol{\beta}}_R = \boldsymbol{\beta}(\hat{P}_n)$ y $\boldsymbol{\beta}(P) = \boldsymbol{\beta}$ pues $r_i = \boldsymbol{\beta}^T \mathbf{u}_i + \epsilon_i$. \square

El siguiente Corolario se deduce inmediatamente del Teorema 7.1.4.

Corolario 7.1.5. Sean $(Y_i, \mathbf{Z}_i^T, X_i)^T$, $1 \leq i \leq n$ vectores aleatorios independientes que satisfacen el modelo (6.1) y supongamos que $\hat{\phi}_j(x)$, $1 \leq j \leq p$ son estimadores de $\phi_j(x)$ tales que para cualquier conjunto compacto $S_{\mathcal{H}} \subset \mathcal{H}$

$$\sup_{x \in S_{\mathcal{H}}} |\hat{\phi}_j(x) - \phi_j(x)| \xrightarrow{a.s.} 0, \quad 0 \leq j \leq p.$$

Bajo las condiciones del Teorema 7.1.4, el estimador $\hat{g}(x) = \hat{\phi}_0(x) - \hat{\boldsymbol{\beta}}_R^T \hat{\boldsymbol{\phi}}(x)$ de la función de regresión g cumple que $\sup_{x \in S_{\mathcal{H}}} |\hat{g}(x) - g(x)| \xrightarrow{a.s.} 0$.

7.2. Distribución Asintótica

Sean ψ_1 y w_2 una función de escores y de peso, respectivamente. Sean $\hat{\phi}_0(x)$ y $\hat{\boldsymbol{\phi}}(x)$ estimadores consistentes de $\phi_0(x)$ y $\boldsymbol{\phi}(x)$, respectivamente, donde $\phi_0(x)$ cumple $\phi_0(x) = \boldsymbol{\beta}^T \boldsymbol{\phi}(x) + g(x)$.

En este capítulo, daremos la distribución asintótica de los estimadores del parámetro de regresión cuando están definidos como cualquier solución de

$$\sum_{i=1}^n \psi_1 \left(\frac{\widehat{r}_i - \widehat{\boldsymbol{\beta}}^T \widehat{\mathbf{u}}_i}{s_n} \right) w_2(\|\widehat{\mathbf{u}}_i\|) \widehat{\mathbf{u}}_i = 0, \quad (7.3)$$

con $\widehat{r}_i = Y_i - \widehat{\phi}_0(X_i)$, $\widehat{\mathbf{u}}_i = \mathbf{Z}_i - \widehat{\boldsymbol{\phi}}(X_i)$ y s_n es un estimador de escala de los residuos. Observemos que la ecuación (7.3) corresponde a la ecuación derivada que permite definir un estimador robusto de regresión, ya sea, un M -estimador, GM -estimador o un MM -estimador.

Para obtener la distribución asintótica de los estimadores del parámetro de regresión, pediremos que las covariables X_i estén en un conjunto compacto $S_{\mathcal{X}}$.

Indiquemos por $(R(X), \mathbf{U}(X)^T)^T$ a un vector aleatorio con la misma distribución que $(r_i, \mathbf{u}_i^T)^T = (Y_i - \phi_0(X_i), (\mathbf{Z}_i - \boldsymbol{\phi}(X_i))^T)^T$. Luego, $R(X) - \mathbf{U}(X)^T \boldsymbol{\beta} \sim \epsilon$, donde ϵ está definido en (6.1).

Necesitaremos el siguiente conjunto de hipótesis:

N1 ψ_1 es una función impar, acotada, dos veces continuamente diferenciable con derivadas ψ_1' y ψ_1'' acotadas tales que $\varphi_1(t) = t\psi_1'(t)$ y $\varphi_2(t) = t\psi_1''(t)$ son acotadas.

N2 $\mathbb{E}(w_2(\|\mathbf{U}(X)\|)\|\mathbf{U}(X)\|^2) < \infty$ y la matriz

$$\begin{aligned} \mathbf{A} &= \mathbb{E} \left(\psi_1' \left(\frac{R(X) - \mathbf{U}(X)^T \boldsymbol{\beta}}{\sigma_0} \right) w_2(\|\mathbf{U}(X)\|) \mathbf{U}(X) \mathbf{U}(X)^T \right) \\ &= \mathbb{E} \left(\psi_1' \left(\frac{\epsilon}{\sigma_0} \right) \right) \mathbb{E}(w_2(\|\mathbf{U}(X)\|) \mathbf{U}(X) \mathbf{U}(X)^T) \end{aligned}$$

es no singular.

N3 $w_2(u) = \psi_2(u)u^{-1} > 0$ es una función acotada y Lipschitz de orden 1. Además, ψ_2 es también una función acotada y continuamente diferenciable con derivada acotada ψ_2' tal que $\lambda_2(t) = t\psi_2'(t)$ es acotada.

N4 $\mathbb{E}(w_2(\|\mathbf{U}(X)\|)\mathbf{U}(X)|X = x) = 0$ para casi todo x .

N5 Las funciones $\phi_j(x)$, $0 \leq j \leq p$ son continuas. Más aún, ϕ_j son Lipschitz de orden 1 para todo $x, y \in S_{\mathcal{X}}$, es decir, $|\phi(x) - \phi(y)| \leq Cd(x, y)$.

Lema 7.2.1. Sean $(Y_i, \mathbf{Z}_i^T, X_i)^T$, $1 \leq i \leq n$ vectores independientes que satisfacen el modelo (6.1) con ϵ_i independiente de $(\mathbf{Z}_i^T, X_i)^T$, ϵ_i con distribución simétrica. Supongamos que X_i son elementos aleatorios tales que $\mathbb{P}(X_i \in S_{\mathcal{X}}) = 1$ donde $S_{\mathcal{X}}$ es un conjunto compacto. Llamemos $(R(X), \mathbf{U}(X)^T)^T$ un vector aleatorio con la misma distribución que

$(r_i, \mathbf{u}_i^T)^T = (Y_i - \phi_0(X_i), (\mathbf{Z}_i - \boldsymbol{\phi}(X_i))^T)^T$. Sean $\widehat{\phi}_j(x)$, $0 \leq j \leq p$ estimadores de $\phi_j(x)$ tales que

$$\sup_{x \in S_{\mathcal{X}}} |\widehat{\phi}_j(x) - \phi_j(x)| \xrightarrow{p} 0, \quad 1 \leq j \leq p$$

y supongamos que $\widetilde{\boldsymbol{\beta}} \xrightarrow{p} \boldsymbol{\beta}$ y $s_n \xrightarrow{p} \sigma_0$. Entonces, bajo **N1-N3**, $\mathbf{A}_n \xrightarrow{p} \mathbf{A}$ donde \mathbf{A} está dada en **N2** y

$$\mathbf{A}_n = \frac{1}{n} \sum_{i=1}^n \psi_1' \left(\frac{\widehat{r}_i - \widehat{\mathbf{u}}_i^T \widetilde{\boldsymbol{\beta}}}{s_n} \right) w_2(\|\widehat{\mathbf{u}}_i\|) \widehat{\mathbf{u}}_i \widehat{\mathbf{u}}_i^T.$$

La demostración del Lema 7.2.1 se encuentra en el Apéndice.

Teorema 7.2.2. Sean $(Y_i, \mathbf{Z}_i^T, X_i)^T$, $1 \leq i \leq n$ vectores independientes que satisfacen el modelo (6.1) con ϵ_i independiente de $(\mathbf{Z}_i^T, X_i)^T$, ϵ_i con distribución simétrica. Supongamos que X_i son elementos aleatorios tales que $\mathbb{P}(X \in S_{\mathcal{X}}) = 1$ donde $S_{\mathcal{X}}$ es un conjunto compacto. Llamemos $(R(X), \mathbf{U}(X)^T)^T$ un vector aleatorio con la misma distribución que $(r_i, \mathbf{u}_i^T)^T = (Y_i - \phi_0(X_i), (\mathbf{Z}_i - \boldsymbol{\phi}(X_i))^T)^T$. Sean $\widehat{\phi}_j(x)$, $0 \leq j \leq p$ estimadores de $\phi_j(x)$ tales que $\widehat{\phi}_j(x)$ es continua en $S_{\mathcal{X}}$ y

$$n^{1/4} \sup_{x \in S_{\mathcal{X}}} |\widehat{\phi}_j(x) - \phi_j(x)| \xrightarrow{p} 0, \quad 0 \leq j \leq p, \quad (7.4)$$

Supongamos que $s_n \xrightarrow{p} \sigma_0$ y que **N1** a **N5** se cumplen. Si además para todo $x \in S_{\mathcal{X}}$ y para todo $h \in \mathcal{H}$ tal que $x+h \in S_{\mathcal{X}}$ y $0 \leq j \leq p$ se verifica que existe una constante C_1 independiente de n , x y h tal que

$$|(\widehat{\phi}_j(x+h) - \widehat{\phi}_j(x)) - (\phi_j(x+h) - \phi_j(x))| \leq C_1 d(x, x+h) \quad (7.5)$$

entonces, si $\int_0^\delta \sqrt{1 + N_\nu(S_{\mathcal{X}})} d\nu < \infty$ se cumple

$$n^{1/2}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{D} N(0, \mathbf{A}^{-1} \boldsymbol{\Sigma} (\mathbf{A}^{-1})^T),$$

donde \mathbf{A} está definida en **N2** y

$$\begin{aligned} \boldsymbol{\Sigma} &= \sigma_0^2 \mathbb{E} \left(\psi_1^2 \left(\frac{R(X) - \mathbf{U}(X)^T \boldsymbol{\beta}}{\sigma_0} \right) w_2^2(\|\mathbf{U}(X)\|) \mathbf{U}(X) \mathbf{U}(X)^T \right) \\ &= \sigma_0^2 \mathbb{E} \left(\psi_1^2 \left(\frac{\epsilon}{\sigma_0} \right) \right) \mathbb{E}(w_2^2(\|\mathbf{U}(X)\|) \mathbf{U}(X) \mathbf{U}(X)^T). \end{aligned}$$

DEMOSTRACIÓN. Llamemos

$$\begin{aligned} \mathbf{L}_n(\sigma, \mathbf{b}) &= \frac{\sigma}{n} \sum_{i=1}^n \psi_1 \left(\frac{r_i - \mathbf{u}_i^T \mathbf{b}}{\sigma} \right) w_2(\|\mathbf{u}_i\|) \mathbf{u}_i, \\ \widehat{\mathbf{L}}_n(\sigma, \mathbf{b}) &= \frac{\sigma}{n} \sum_{i=1}^n \psi_1 \left(\frac{\widehat{r}_i - \widehat{\mathbf{u}}_i^T \mathbf{b}}{\sigma} \right) w_2(\|\widehat{\mathbf{u}}_i\|) \widehat{\mathbf{u}}_i. \end{aligned}$$

A partir de un desarrollo de Taylor de orden 1 alrededor de $\widehat{\boldsymbol{\beta}}$, se tiene que

$$\widehat{\mathbf{L}}_n(\sigma, \boldsymbol{\beta}) = \frac{\sigma}{n} \sum_{i=1}^n \psi_1 \left(\frac{\widehat{r}_i - \widehat{\mathbf{u}}_i^T \widehat{\boldsymbol{\beta}}}{\sigma} \right) w_2(\|\widehat{\mathbf{u}}_i\|) \widehat{\mathbf{u}}_i + \frac{1}{n} \sum_{i=1}^n \psi_1' \left(\frac{\widehat{r}_i - \widehat{\mathbf{u}}_i^T \widetilde{\boldsymbol{\beta}}}{\sigma} \right) w_2(\|\widehat{\mathbf{u}}_i\|) \widehat{\mathbf{u}}_i \widehat{\mathbf{u}}_i^T (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}),$$

donde $\widetilde{\boldsymbol{\beta}}$ es un punto intermedio entre $\widehat{\boldsymbol{\beta}}$ y $\boldsymbol{\beta}$. Debido a que $\widehat{\boldsymbol{\beta}}$ es solución de $\widehat{\mathbf{L}}_n(s_n, \mathbf{b}) = 0$, esto implica que

$$\widehat{\mathbf{L}}_n(s_n, \boldsymbol{\beta}) = 0 + \frac{1}{n} \sum_{i=1}^n \psi_1' \left(\frac{\widehat{r}_i - \widehat{\mathbf{u}}_i^T \widetilde{\boldsymbol{\beta}}}{s_n} \right) w_2(\|\widehat{\mathbf{u}}_i\|) \widehat{\mathbf{u}}_i \widehat{\mathbf{u}}_i^T (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})$$

y luego, se tiene que $(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = \mathbf{A}_n^{-1} \widehat{\mathbf{L}}_n(s_n, \boldsymbol{\beta})$ con

$$\mathbf{A}_n = \frac{1}{n} \sum_{i=1}^n \psi_1' \left(\frac{\widehat{r}_i - \widehat{\mathbf{u}}_i^T \widetilde{\boldsymbol{\beta}}}{s_n} \right) w_2(\|\widehat{\mathbf{u}}_i\|) \widehat{\mathbf{u}}_i \widehat{\mathbf{u}}_i^T.$$

Por la consistencia de $\widehat{\boldsymbol{\beta}}$, el Lema 7.2.1 implica que $\mathbf{A}_n \xrightarrow{p} \mathbf{A}$ y por lo tanto, como

$$\widehat{\mathbf{L}}_n(s_n, \boldsymbol{\beta}) = (\widehat{\mathbf{L}}_n(s_n, \boldsymbol{\beta}) - \mathbf{L}_n(s_n, \boldsymbol{\beta})) + (\mathbf{L}_n(s_n, \boldsymbol{\beta}) - \mathbf{L}_n(\sigma_0, \boldsymbol{\beta})) + \mathbf{L}_n(\sigma_0, \boldsymbol{\beta}),$$

a partir de **N2** será suficiente mostrar que

- (a) $n^{1/2} \mathbf{L}_n(\sigma_0, \boldsymbol{\beta}) \xrightarrow{D} N(0, \Sigma)$.
- (b) $n^{1/2} [\widehat{\mathbf{L}}_n(s_n, \boldsymbol{\beta}) - \mathbf{L}_n(s_n, \boldsymbol{\beta})] \xrightarrow{p} 0$.
- (c) $n^{1/2} [\mathbf{L}_n(s_n, \boldsymbol{\beta}) - \mathbf{L}_n(\sigma_0, \boldsymbol{\beta})] \xrightarrow{p} 0$.

(a) Se deduce del partir del Teorema Central del Límite, porque $r_i - \mathbf{u}_i^T \boldsymbol{\beta} = \epsilon_i$.

(b) Llamemos ξ_i a un punto intermedio entre $r_i - \mathbf{u}_i^T \boldsymbol{\beta}$ y $\widehat{r}_i - \widehat{\mathbf{u}}_i^T \widetilde{\boldsymbol{\beta}}$ y $\widehat{\eta}_j(x) = \widehat{\phi}_j(x) - \phi_j(x)$ para $0 \leq j \leq p$, y $\widehat{\boldsymbol{\eta}} = (\widehat{\eta}_1(x), \dots, \widehat{\eta}_p(x))^T$. A partir de un desarrollo de Taylor de segundo orden, se tiene que $\widehat{\mathbf{L}}_n(s_n, \boldsymbol{\beta}) = \mathbf{L}_n(s_n, \boldsymbol{\beta}) + \widehat{\mathbf{L}}_{n,1} + \widehat{\mathbf{L}}_{n,2} + \widehat{\mathbf{L}}_{n,3} + \widehat{\mathbf{L}}_{n,4} + \widehat{\mathbf{L}}_{n,5}$, donde

$$\begin{aligned} \widehat{\mathbf{L}}_{n,1} &= \frac{1}{n} \sum_{i=1}^n \psi_1' \left(\frac{r_i - \mathbf{u}_i^T \boldsymbol{\beta}}{s_n} \right) [\widehat{\boldsymbol{\eta}}^T(X_i) \boldsymbol{\beta} - \widehat{\eta}_0(x_i)] w_2(\|\mathbf{u}_i\|) \mathbf{u}_i, \\ \widehat{\mathbf{L}}_{n,2} &= \frac{s_n}{n} \sum_{i=1}^n \psi_1 \left(\frac{r_i - \mathbf{u}_i^T \boldsymbol{\beta}}{s_n} \right) [w_2(\|\widehat{\mathbf{u}}_i\|) \widehat{\mathbf{u}}_i - w_2(\|\mathbf{u}_i\|) \mathbf{u}_i], \\ \widehat{\mathbf{L}}_{n,3} &= \frac{s_n}{n} \sum_{i=1}^n \left[\psi_1 \left(\frac{\widehat{r}_i - \widehat{\mathbf{u}}_i^T \widetilde{\boldsymbol{\beta}}}{s_n} \right) - \psi_1 \left(\frac{r_i - \mathbf{u}_i^T \boldsymbol{\beta}}{s_n} \right) \right] w_2(\|\widehat{\mathbf{u}}_i\|) (\widehat{\mathbf{u}}_i - \mathbf{u}_i), \\ \widehat{\mathbf{L}}_{n,4} &= \frac{1}{2s_n} \frac{1}{n} \sum_{i=1}^n \psi_1'' \left(\frac{\xi_i}{s_n} \right) [\boldsymbol{\beta}^T \widehat{\boldsymbol{\eta}}(X_i) - \widehat{\eta}_0(X_i)]^2 w_2(\|\widehat{\mathbf{u}}_i\|) \mathbf{u}_i, \\ \widehat{\mathbf{L}}_{n,5} &= \frac{1}{n} \sum_{i=1}^n \psi_1' \left(\frac{r_i - \mathbf{u}_i^T \boldsymbol{\beta}}{s_n} \right) [\boldsymbol{\beta}^T \widehat{\boldsymbol{\eta}}(X_i) - \widehat{\eta}_0(X_i)] [w_2(\|\widehat{\mathbf{u}}_i\|) - w_2(\|\mathbf{u}_i\|)] \mathbf{u}_i. \end{aligned}$$

Es fácil ver que **N3** implica que $|w_2(\|\widehat{\mathbf{u}}_i\|) - w_2(\|\mathbf{u}_i\|)| \leq C\|\widehat{\eta}(X_i)\|/\|\mathbf{u}_i\|$, donde $C = \|w_2\|_\infty + C_{\psi_2}$, luego

$$\begin{aligned} n^{1/2}\|\widehat{\mathbf{L}}_{n,3}\| &\leq p\|w_2\|_\infty\|\psi'_1\|_\infty n^{1/2} \left[\max_{0 \leq j \leq p} \sup_{x \in S_{\mathcal{H}}} |\widehat{\eta}_j(x)| \right]^2 (1 + p\|\boldsymbol{\beta}\|), \\ n^{1/2}\|\widehat{\mathbf{L}}_{n,4}\| &\leq \frac{1}{2} \frac{1}{s_n} \|\psi''_1\|_\infty n^{1/2} \left[\max_{0 \leq j \leq p} \sup_{x \in S_{\mathcal{H}}} |\widehat{\eta}_j(x)| \right]^2 (1 + p\|\boldsymbol{\beta}\|)^2 \\ &\quad \times \left(\|\psi_2\|_\infty + p\|w_2\|_\infty \max_{0 \leq j \leq p} \sup_{x \in S_{\mathcal{H}}} |\widehat{\eta}_j(x)| \right), \\ n^{1/2}\|\widehat{\mathbf{L}}_{n,5}\| &\leq pC\|\psi'_1\|_\infty (1 + p\|\boldsymbol{\beta}\|) n^{1/2} \left[\max_{0 \leq j \leq p} \sup_{x \in S_{\mathcal{H}}} |\widehat{\eta}_j(x)| \right]^2. \end{aligned}$$

Por lo tanto, usando (7.4) y la consistencia de s_n , se tiene que $n^{1/2}\|\widehat{\mathbf{L}}_{n,j}\| \xrightarrow{p} 0$, $3 \leq j \leq 5$.

Falta ver que $n^{1/2}\widehat{\mathbf{L}}_{n,j} \xrightarrow{p} 0$ para $j = 1, 2$. Para eso, mostraremos que

$$n^{1/2} \frac{s_n}{n} \sum_{i=1}^n \psi'_1 \left(\frac{r_i - \mathbf{u}_i^T \boldsymbol{\beta}}{s_n} \right) \widehat{\eta}_\ell(X_i) w_2(\|\mathbf{u}_i\|) \mathbf{u}_i \xrightarrow{p} 0, \quad 0 \leq \ell \leq p, \quad (7.6)$$

$$n^{1/2} \frac{s_n}{n} \sum_{i=1}^n \psi_1 \left(\frac{r_i - \mathbf{u}_i^T \boldsymbol{\beta}}{s_n} \right) [w_2(\|\widehat{\mathbf{u}}_i\|) \widehat{\mathbf{u}}_i - w_2(\|\mathbf{u}_i\|) \mathbf{u}_i] \xrightarrow{p} 0. \quad (7.7)$$

Para esto, usaremos la desigualdad máxima que puede verse en Van der Vart y Wellner (1996), componente a componente, y definiremos clases apropiadas de funciones con entropía uniforme finita.

Fijemos una coordenada j , $1 \leq j \leq p$. Para cualquier función $m : \mathcal{H} \rightarrow \mathbb{R}$, cualquier vector de funciones $\mathbf{m}(x) = (m_1(x), \dots, m_p(x))^T$ y cualquier $\sigma \in (\frac{\sigma_0}{2}, 2\sigma_0)$, si $u_{i,j}$ denota la j -ésima coordenada de \mathbf{u}_i , definamos

$$J_{n,1}(\sigma, m) = n^{1/2} \frac{\sigma}{n} \sum_{i=1}^n \psi'_1 \left(\frac{r_i - \mathbf{u}_i^T \boldsymbol{\beta}}{\sigma} \right) m(X_i) w_2(\|\mathbf{u}_i\|) u_{i,j},$$

$$J_{n,2}(\sigma, \mathbf{m}) = n^{1/2} \frac{\sigma}{n} \sum_{i=1}^n \psi_1 \left(\frac{r_i - \mathbf{u}_i^T \boldsymbol{\beta}}{\sigma} \right) [w_2(\|u_i + \mathbf{m}(x_i)\|) (u_{i,j} + m_j(x_i)) - w_2(\|\mathbf{u}_i\|) u_{i,j}],$$

donde por simplicidad se ha omitido el subíndice j en las definiciones de $J_{n,1}(\sigma, m)$ y $J_{n,2}(\sigma, \mathbf{m})$.

Sea $\mathcal{J} = (\sigma_0/2, 2\sigma_0)$ y sean C_2 y C_3 constantes positivas fijas. Definamos

$$\begin{aligned} \mathcal{C} = \{ &g : S_{\mathcal{H}} \subset \mathcal{H} \rightarrow \mathbb{R} \text{ continuas} : \|g\|_\infty \leq C_2, \text{ y } |g(y) - g(x)| \leq C_3 d(y, x) \\ &\forall x \in S_{\mathcal{H}}, y \in S_{\mathcal{H}} \} \end{aligned}$$

Luego, $\mathcal{C} \subset \mathcal{C}_{C_4}$ con $C_4 = \max(C_2, C_3)$ y \mathcal{C}_M definida en (5.1). Usando el Lema 5.1.2, tenemos que para cualquier medida de probabilidad \mathbb{Q} y $\forall 0 < \nu < 2$, el número de bracketing $N_{[\cdot]}(2\nu, \mathcal{C}, L^2(\mathbb{Q}))$ satisface

$$\log N(\nu, \mathcal{C}, L^2(\mathbb{Q})) \leq C_4 \frac{2}{\nu} + \log(13) N_{\frac{\nu}{2C_4}}(S_{\mathcal{H}})$$

donde $N_\nu(S_{\mathcal{H}})$ es el mínimo número de bolas de radio ν que cubren $S_{\mathcal{H}}$. Consideremos las clases de funciones

$$\begin{aligned} \mathcal{F}_1 &= \left\{ f_{1,\sigma,h}(r, u, x) = \sigma \psi'_1 \left(\frac{r - u^T \boldsymbol{\beta}}{\sigma} \right) w_2(\|\mathbf{u}\|) u_j m(x), \sigma \in \mathcal{J}, h \in \mathcal{C} \right\} \\ \mathcal{F}_2 &= \left\{ f_{1,\sigma,h}(r, u, x) = \sigma \psi_1 \left(\frac{r - u^T \boldsymbol{\beta}}{\sigma} \right) [w_2(\|\mathbf{u} + \mathbf{m}(x)\|)(u_j + m_j(x)) - w_2(\|\mathbf{u}\|)u_j], \right. \\ &\quad \left. \sigma \in \mathcal{J}, \mathbf{m}(x) = (m_1(x), \dots, m_p(x))^T, m_\ell \in \mathcal{C} \right\}, \end{aligned}$$

donde nuevamente por simplicidad se ha omitido el subíndice j . Observar que \mathcal{F}_1 y \mathcal{F}_2 tienen como envolvente las constantes $A_1 = 2\sigma_0 \|\psi'_1\|_\infty \|\psi_2\|_\infty$ y $A_2 = 4\sigma_0 \|\psi_1\|_\infty \|\psi_2\|_\infty$, respectivamente. Por otro lado, la independencia entre los $\epsilon_i = r_i - \mathbf{u}_i^T \boldsymbol{\beta}$ y $(\mathbf{Z}_i^T, X_i)^T$ implica que, para cualquier $f \in \mathcal{F}_1$, $\mathbb{E}f(r_i, \mathbf{u}_i, X_i) = 0$ ya que vale **N4**; mientras que, $\mathbb{E}f(r_i, \mathbf{u}_i, X_i) = 0$ para cualquier $f \in \mathcal{F}_2$, dado que ψ_1 es impar y los errores tienen distribución simétrica.

Llamemos $\psi_{1,s}(t) = s\psi_1(t/s)$ y $\psi'_{1,s}(t) = s\psi'_1(t/s)$. Por **N1**, se tiene que φ_1 y φ_2 son acotadas, por lo tanto

$$\begin{aligned} |\psi'_{1,s_1}(r) - \psi'_{1,s_2}(r)| &\leq (\|\psi'_1\|_\infty + \|\varphi_2\|_\infty) |s_1 - s_2|, \\ |\psi_{1,s_1}(r) - \psi_{1,s_2}(r)| &\leq (\|\psi_1\|_\infty + \|\varphi_1\|_\infty) |s_1 - s_2|. \end{aligned} \quad (7.8)$$

Sean $\|f\|_{\mathbb{Q},2} = (\mathbb{E}_{\mathbb{Q}}(f^2))^{1/2}$, $B_1 = (\|\psi'_1\|_\infty(3+2\sigma_0) + \|\varphi_2\|_\infty) \|\psi_2\|_\infty$ y $B_2 = 2\|\psi_2\|_\infty (\|\psi_1\|_\infty + \|\varphi_1\|_\infty) + 2\sigma_0 p \|\psi_1\|_\infty (\|w_2\|_\infty(1+p) + p\|\psi'_2\|_\infty)$.

Es fácil ver que, dada $m \in \mathcal{C}$, $\sigma \in \mathcal{J}$ y $0 < \nu < 2$, si $\|m_s - m\|_{\mathbb{Q},2} < \nu$ y $|\sigma_\ell - \sigma| < \nu$, entonces

$$\|f_{1,\sigma_\ell,m_s} - f_{1,\sigma,m}\|_{\mathbb{Q},2} \leq B_1 \nu.$$

Por lo tanto,

$$N(\epsilon B_1, \mathcal{F}_1, L^2(\mathbb{Q})) \leq N(\epsilon, \mathcal{C}, L^2(\mathbb{Q})) N(\epsilon, \mathcal{J}, |\cdot|).$$

Análogamente, dado $\mathbf{m} = (m_1, \dots, m_p)^T$ con $m_j \in \mathcal{C}$, $\sigma \in \mathcal{J}$ y $0 < \nu < 2$, si $\mathbf{m}_s = (m_{s,1}, \dots, m_{s,p})^T$ es tal que $\|m_{s,j} - m_j\|_{\mathbb{Q},2} < \nu$ para $0 \leq j \leq p$ y $|\sigma_\ell - \sigma| < \nu$, entonces

$$\|f_{2,\sigma_\ell,\mathbf{m}_s} - f_{2,\sigma,\mathbf{m}}\|_{\mathbb{Q},2} \leq B_2 \nu$$

lo que implica

$$N(\nu B_2, \mathcal{F}_2, L^2(\mathbb{Q})) \leq N(\nu, \mathcal{C}, L^2(\mathbb{Q}))^p N(\nu, \mathcal{J}, |\cdot|).$$

Por lo tanto, las clases de funciones \mathcal{F}_1 y \mathcal{F}_2 tienen entropía uniforme finita pues $\sqrt{1+a+b} \leq \sqrt{a} + \sqrt{1+b}$ si $a, b > 0$.

Para cualquier clase de funciones \mathcal{F} , definamos la integral

$$\mathcal{J}(\delta, \mathcal{F}) = \sup_{\mathbb{Q}} \int_0^\delta \sqrt{1 + \log(N(\nu \|F\|_{\mathbb{Q},2}, \mathcal{F}, L^2(\mathbb{Q})))} d\nu,$$

donde F es la envolvente de \mathcal{F} y el supremo es tomado sobre todas las medidas de probabilidad discretas \mathbb{Q} con $\|F\|_{\mathbb{Q},2} > 0$. La función \mathcal{J} es creciente, $\mathcal{J}(0, \mathcal{F}) = 0$ y $\mathcal{J}(1, \mathcal{F}) < \infty$ y $\mathcal{J}(\delta, \mathcal{F}) \rightarrow 0$ cuando $\delta \rightarrow 0$ para clases de funciones \mathcal{F} que satisfacen la condición de entropía-uniforme. Más aún, si $\mathcal{F}_0 \subset \mathcal{F}$ y si la envolvente F es usada para \mathcal{F}_0 , entonces $\mathcal{J}(\delta, \mathcal{F}_0) \leq \mathcal{J}(\delta, \mathcal{F})$.

Para cualquier $0 < \delta < 1$, consideremos las subclases

$$\mathcal{F}_{1,\delta} = \{f_{1,\delta,\mathbf{m}}(r, \mathbf{u}, x) \in \mathcal{F}_1 \text{ con } \|\mathbf{m}\|_\infty < \delta\} \subset \mathcal{F}_1,$$

$$\mathcal{F}_{2,\delta} = \{f_{2,\delta,\mathbf{m}}(r, \mathbf{u}, x) \in \mathcal{F}_2 \text{ con } \mathbf{m}(x) = (m_1(x), \dots, m_p(x))^T \text{ y } \|m_\ell\|_\infty < \delta, 1 \leq \ell \leq p\} \subset \mathcal{F}_2.$$

Para cualquier $\nu > 0$, sea $0 < \delta < 1$. Usando que $s_n \xrightarrow{p} \sigma_0$ y que (7.4) y (7.5) llevan a que, para $0 \leq j \leq p$

$$\begin{aligned} \sup_{x \in S_{\mathcal{H}}} |\hat{\eta}_j(x)| &= \sup_{x \in S_{\mathcal{H}}} |\hat{\phi}_j(x) - \phi_j(x)| \xrightarrow{p} 0, \\ |\hat{\eta}_j(y) - \hat{\eta}_j(x)| &< C_1 d(x, y), \quad \forall x \in S_{\mathcal{H}} \forall y \in S_{\mathcal{H}}, \end{aligned}$$

se tiene que, para n lo suficientemente grande, $\mathbb{P}(s_n \in \mathcal{J}) > 1 - \delta/2$ y $\mathbb{P}(\hat{\eta}_j \in \mathcal{C} \text{ y } \|\hat{\eta}_j\|_\infty < \delta) > 1 - \delta/2$, para $0 \leq j \leq p$. Sea $A_3 = 2\sigma_0 \|\psi_1\|_\infty (\|w_2\|_\infty (1+p) + p \|\psi'_2\|_\infty)$. Realizando algunos cálculos sencillos se llega a que

$$\begin{aligned} \sup_{f \in \mathcal{F}_{1,\delta}} \frac{1}{n} \sum_{i=1}^n f^2(r_i, \mathbf{u}_i, X_i) &\leq A_1^2 \delta, \\ \sup_{f \in \mathcal{F}_{2,\delta}} \frac{1}{n} \sum_{i=1}^n f^2(r_i, \mathbf{u}_i, X_i) &\leq A_3^2 \delta. \end{aligned}$$

La desigualdad maximal para el número de cubrimiento establece que, para cualquier $0 \leq \ell \leq p$,

$$\begin{aligned} \mathbb{P}(|J_{n,1}(s_n, \hat{\eta}_\ell)| > \epsilon) &\leq \mathbb{P}(|J_{n,1}(s_n, \hat{\eta}_\ell)| > \epsilon, s_n \in \mathcal{J}, \hat{\eta}_\ell \in \mathcal{C} \text{ y } \|\hat{\eta}_\ell\|_\infty < \delta) + \delta \\ &\leq \mathbb{P}\left(\sup_{f \in \mathcal{F}_{1,\delta}} \left| n^{1/2} \frac{1}{n} \sum_{i=1}^n f(r_i, \mathbf{u}_i, X_i) \right| > \epsilon\right) + \delta \\ &\leq \frac{1}{\epsilon} \mathbb{E} \left(\sup_{f \in \mathcal{F}_{1,\delta}} \left| n^{1/2} \frac{1}{n} \sum_{i=1}^n f(r_i, \mathbf{u}_i, X_i) \right| \right) + \delta \\ &\leq \frac{1}{\epsilon} D_1 A_1 \mathcal{J}(\delta, \mathcal{F}_1) + \delta, \end{aligned}$$

donde D_1 es una constante que no depende de n . Similarmente,

$$\mathbb{P}(|J_{n,2}(s_n, \hat{\eta})| > \epsilon) \leq \frac{1}{\epsilon} D_2 A_2 \mathcal{J} \left(\frac{A_3^2}{A_2^2} \delta, \mathcal{F}_2 \right) + \delta.$$

Ahora, (7.6) y (7.7) se deducen del hecho de que $\lim_{\delta \rightarrow 0} \mathcal{J}(\delta, \mathcal{F}_1) = 0$ y $\lim_{\delta \rightarrow 0} \mathcal{J}(\delta, \mathcal{F}_2) = 0$, cuando las clases \mathcal{F}_1 y \mathcal{F}_2 satisfacen la condición de entropía uniforme.

(c) Como

$$n^{1/2}[\mathbf{L}_n(s_n, \boldsymbol{\beta}) - \mathbf{L}_n(\sigma_0, \boldsymbol{\beta})] = n^{-1/2} \sum_{i=1}^n [\psi_{1,s_n}(r_i - \mathbf{u}_i^T \boldsymbol{\beta}) - \psi_{1,\sigma_0}(r_i - \mathbf{u}_i^T \boldsymbol{\beta})] w_2(\|\mathbf{u}_i\|) \mathbf{u}_i,$$

obtenemos el resultado deseado usando (7.8), la acotación de ψ_2 y la desigualdad maximal para el número de cubrimiento, como en (b). \square

La condición $\int_0^\delta \sqrt{1 + N_\nu(S_{\mathcal{H}})} d\nu < \infty$ pedida sobre el conjunto compacto $S_{\mathcal{H}}$ se cumple por ejemplo si $N_\nu(S_{\mathcal{H}}) \leq C\nu^{-\alpha}$ para algún $\alpha > 2$. Como se menciona en Ferraty *et al.* (2010) esta condición se cumple trivialmente en el caso $\mathcal{H} = \mathbb{R}^q$ cuando la semimétrica es la distancia euclídea usual en \mathbb{R}^q . También como se menciona en Henry y Rodriguez (2013) se cumple cuando \mathcal{H} es una variedad diferenciable. En Ferraty y Vieu (2006) y Ferraty *et al.* (2010), se dan ejemplos de semi-métricas que cumplen esa condición en espacios de dimensión infinita. Por ejemplo, si \mathcal{H} es un espacio de Hilbert con producto interno $\langle \cdot, \cdot \rangle$ y si definimos

$$d(x, y) = \sqrt{\sum_{j=1}^k \langle x - y, e_j \rangle^2}$$

donde $\{e_1, \dots, e_j, \dots\}$ es una base ortonormal de \mathcal{H} entonces $N_\nu(S_{\mathcal{H}}) \leq C\nu^{-\alpha}$ se cumple con $\alpha = k$

Por otra parte, si estamos en un espacio normado, la condición (7.5) se cumple si la derivada de Hadamard $\hat{\eta}'_{j,x}$ de $\hat{\eta}_j(x) = \hat{\phi}_j(x) - \phi_j(x)$ en el punto x es tal que $\|\hat{\eta}'_{j,x}\| \leq C_1$, $0 \leq j \leq p$.

Capítulo 8

Estudio de Monte Carlo

En este Capítulo, presentamos los resultados de un estudio de simulación realizado para comparar el comportamiento del estimador robusto introducido con la propuesta de Aneiros-Pérez y Vieu (2006).

8.1. Modelo y estimadores considerados

Se realizó un estudio de simulación cuando el parámetro de regresión tiene dimensión 2 y comparamos el comportamiento del estimador de mínimos cuadrados introducido por Aneiros-Pérez y Vieu (2006) con los diferentes estimadores robustos. Antes de calcular el estimador de β los estimadores de $\phi_0(x)$ y $\phi_j(x)$, $j = 1, 2$ se calcularon mediante un suavizado robusto utilizando

- un M -estimador local con función de escores bicuadrada, con constante 4.685, que da una 95 % de eficiencia para el M -estimador local.
- la mediana local.

Luego de suavizar las variables respuesta Y y las covariables de regresión \mathbf{Z} , se calcularon los siguientes estimadores de regresión de β :

- M -estimadores con función de Huber con constante 1.345.
- M -estimadores con función bicuadrada (Tukey) con constante 4.685.
- GM -estimadores con función de Huber con constante 1.6 para los residuos y con función de peso bicuadrada con constante $c = \chi_{2,0,95}^2$ donde $\mathbb{P}(\chi_2^2 \geq \chi_{2,1-\alpha}^2) = \alpha$.
- LMS -estimadores, o sea, los estimadores de menor mediana de los residuos cuadrados, introducidos por Rousseeuw (1984).

- *LTS*–estimadores con 33% de observaciones podadas, o sea, los estimadores de menor suma de cuadrados podada.

Las constantes de calibración en los residuos elegidas para los estimadores son las utilizadas por Bianco y Boente (2004) en el caso finito–dimensional. Con estas constantes los *M*–estimadores y el *LTS*–estimador tienen una eficiencia asintótica del 95% y 80%, respectivamente. Por otra parte, la eficiencia del *GM*– para el modelo elegido es del 60% mientras que el *LMS*–estimador tiene eficiencia 0 ya que converge a una tasa menor.

En todas las tablas y gráficos LS indica al estimador de mínimos cuadrados, MH y MT a los *M*–estimadores obtenidos con la función de Huber y Tukey, GM los *GM*–estimadores, mientras que LTS y LMS denotan los estimadores de alto punto de ruptura obtenidos usando los estimadores “least trimmed of squares” y “least median of squares”, respectivamente. Cuando se utiliza como suavizador a la mediana local, los estimadores se denotarán como MH.m, MT.m, GM.m, LTS.m y LMS.m, respectivamente.

En el proceso de suavizado, se utilizó el núcleo

$$K(u) = \begin{cases} \frac{35}{16}(1-u^2)^3 & 0 \leq u \leq 1 \\ 0 & \text{caso contrario} \end{cases} .$$

También se hizo una simulación con el núcleo de Epanechnikov, obteniendo resultados similares. El estimador robusto de la función g considerado fue el *M*–estimador local $\widehat{g}(x)$ definido en la Sección 6.2, suavizando los residuos $Y_i - \mathbf{Z}_i^T \widehat{\boldsymbol{\beta}}$, pero también se reportan los resultados correspondientes al estimador $\widehat{g}(x) = \widehat{\phi}_{0,\text{MED}}(x) - \widehat{\phi}_{\text{MED}}(x)^T \widehat{\boldsymbol{\beta}}$. Para los estimadores iniciales, se consideraron las ventanas $h_1 = 0,04$, $h_1 = 0,06$. Por otra parte, para el estimador robusto \widehat{g} solución de (6.2), se consideró como ventana $h_2 = h_1$ y una ventana menor que h_1 . Cuando $h_1 = 0,04$ se tomó $h_2 = 0,03$ mientras que cuando $h_1 = 0,06$ se tomó $h_2 = 0,04$.

Se realizaron $nr = 1000$ replicaciones generando muestras independientes de tamaño $n = 100$ siguiendo el modelo utilizado en Aneiros-Pérez y Vieu (2009)

$$Y_i = \mathbf{Z}_i^T \boldsymbol{\beta} + g(X_i) + \epsilon_i, \quad 1 \leq i \leq n$$

donde $\boldsymbol{\beta} = (-1, 3)^T$, Z_{ij} y ϵ_i son variables aleatorias independientes con distribución $Z_{ij} \sim N(0, 1)$ y $\epsilon_i \sim N(0, \sigma_\epsilon^2)$ donde $\sigma_\epsilon = 0,1(\max_{x \in \mathcal{H}} g(x) - \min_{x \in \mathcal{H}} g(x))$.

El dato funcional considerado es un elemento del espacio de funciones sobre el intervalo $[0, 1]$ y se generó como $X_i(z) = a_i(z - 0,5)^2 + b_i$, $0 \leq z \leq 1$, donde a_i y b_i son variables aleatorias independientes con distribución $a_i \sim U(0, 1)$ y $b_i \sim U(-0,5, 0,5)$.

La función de regresión r considerada fue

$$g(X_i) = \exp(-8f(X_i)) - \exp(-12f(X_i))$$

con $f(X_i) = \text{signo}(X'_i(1) - X'_i(0)) \sqrt{3 \int_0^1 (X'_i(z))^2 dz}$.

La suavidad de las curvas X_i permite considerar una semimétrica basada en la norma L_2 de derivadas de las curvas. Usaremos, como en Aneiros-Perez y Vieu (2009), la semimétrica

$$d(X, X^*) = \left(\int_0^1 (X'(z) - X^{*'}(z))^2 dz \right)^{1/2}.$$

Los resultados para conjuntos de datos normales se indicarán con C_0 , mientras que C_1 y C_2 denotarán las siguientes contaminaciones:

- C_1 : $\epsilon_1, \dots, \epsilon_n$ son i.i.d. $0.9 N(0, \sigma_\epsilon^2) + 0.1 N(0, 25 \sigma_\epsilon^2)$.
- C_2 : $\epsilon_1, \dots, \epsilon_n$ son i.i.d. $0.9 N(0, \sigma_\epsilon^2) + 0.1 N(0, 25 \sigma_\epsilon^2)$ y fueron modificadas artificialmente 10 observaciones \mathbf{Z}_i pero no las variables respuesta, como $(-20, 20)^T$.

En la Figura 8.1 podemos ver las curvas X_i consideradas correspondiente a una de las muestras generadas.

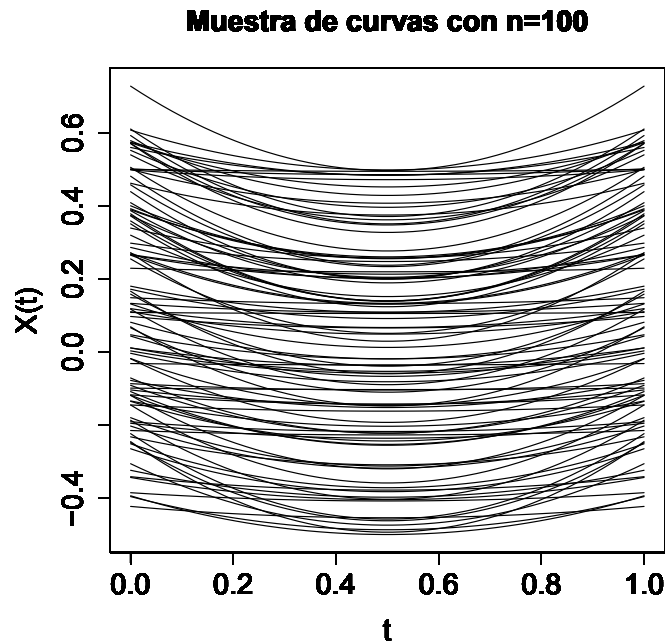


Figura 8.1: Ejemplo de una muestra de datos generados

8.2. Medidas resumen de la simulación

Para cada situación, el comportamiento de un estimador \hat{g} de g fue evaluado utilizando dos medidas:

$$MSE(\hat{g}) = \frac{1}{n} \sum_{i=1}^n [\hat{g}(x_i) - g(x_i)]^2,$$

$$MedSE(\hat{g}) = \text{mediana}([\hat{g}(x_i) - g(x_i)]^2)_{1 \leq i \leq n}.$$

Se consideraron además dos medidas para evaluar el comportamiento en los modelos C_1 y C_2 evaluando el comportamiento de los estimadores sobre los datos contaminados y los datos no contaminados. Sea $\gamma_i = 0$ si (Y_i, \mathbf{Z}_i, X_i) corresponde a un dato no contaminado y $\gamma_i = 1$ si (Y_i, \mathbf{Z}_i, X_i) corresponde a un dato contaminado

$$ECMP_{1,1} = \frac{1}{\sum_{i=1}^n \gamma_i} \frac{\sum_{j=1}^n \gamma_i (Y_j - \hat{Y}_j)^2}{\text{VAR}(Y)} \quad (8.1)$$

$$ECMP_{2,1} = \frac{1}{\sum_{i=1}^n \gamma_i} \frac{\sum_{j=1}^n \gamma_i (Y_j - \hat{Y}_j)^2}{\text{MAD}^2(Y)} \quad (8.2)$$

$$ECMP_{1,0} = \frac{1}{n - \sum_{i=1}^n \gamma_i} \frac{\sum_{j=1}^n (1 - \gamma_i) (Y_j - \hat{Y}_j)^2}{\text{VAR}(Y)} \quad (8.3)$$

$$ECMP_{2,0} = \frac{1}{n - \sum_{i=1}^n \gamma_i} \frac{\sum_{j=1}^n (1 - \gamma_i) (Y_j - \hat{Y}_j)^2}{\text{MAD}^2(Y)}. \quad (8.4)$$

Las Tablas 8.1 a 8.33 resumen los resultados de las simulaciones. Las Tablas 8.1 y 8.2 presentan la media, desviación estándar, error cuadrático medio, mediana y mad de los estimadores de regresión de β_1 y β_2 , respectivamente, mientras que las Figuras 8.2 y 8.3 muestran los boxplots de los estimadores de β_1 y β_2 , respectivamente para el caso en que la ventana es $h_1 = 0,04$. En las Tablas 8.12 y 8.13 se presenta la misma información pero para el caso en que la ventana es $h_1 = 0,06$ y en las Tablas 8.27 y 8.28 para el caso en que las ventanas $h_1 = h_2 = 0,15$. Las Tablas 8.3, 8.14 y 8.29 presentan la media y la mediana sobre las 1000 replicaciones de $\|\hat{\beta} - \beta\|_2$. Por otro lado, las Tablas 8.4, 8.8, 8.19, 8.15 y 8.30 reportan la media y la mediana del $MSE(\hat{g})$ para las diferentes ventanas, mientras que las Tablas 8.5, 8.9, 8.20, 8.16 y 8.31 presentan la media y la mediana del $MedSE(\hat{g})$. En las Tablas 8.6 y 8.7 se muestran la media y mediana de los errores de predicción con las medidas definidas en (8.1) a (8.4) tomados sobre los datos sin contaminar y los datos contaminados, respectivamente, y cuando las ventanas son $h_1 = h_2 = 0,04$. En las tablas 8.10, 8.11, 8.17, 8.18, 8.21, 8.22, 8.32 y 8.33 tenemos estas mismas medidas obtenidas a partir de las otras ventanas consideradas.

8.3. Resultados

El estudio de simulación confirma el esperado comportamiento inadecuado para el estimador clásico ante la presencia de observaciones atípicas. Con respecto a la estimación del parámetro de regresión β , al analizar los resultados para ambas coordenadas que figuran en las Tablas 8.1, 8.2, 8.12 y 8.13 bajo C_0 , se puede observar cierta ventaja del método clásico sobre los estimadores robustos, siendo los M -estimadores MH , MT y GM -estimadores los que obtienen los valores más cercanos a LS por ser los más eficientes. Un aumento de la desviación estándar se puede ver, especialmente bajo C_2 . También bajo esta contaminación se puede apreciar que para todas las ventanas consideradas, el mejor comportamiento para estimar los parámetros de regresión se obtuvo con los GM -estimadores, obteniendo también un muy buen desempeño con los estimadores de least trimmed (LMS) y least median of square (LTS). Por otro lado, el error cuadrático medio (MSE) para los estimadores de β_1 y β_2 basados en el procedimiento clásico es más de cien veces mayor que el valor de los estimadores robustos antes mencionados, siendo mucho mayor esta diferencia en el caso de los GM -estimadores. En las Tablas 8.3 y 8.14 podemos ver que para las distintas ventanas se obtienen valores similares con la medida de resumen utilizada. Además, bajo C_0 la media del GM -estimador de β es aproximadamente trece veces mayor que la del LS -estimador lo que muestra la pérdida en eficiencia, mientras que bajo C_2 la situación se invierte siendo la del caso clásico casi sesenta veces mayor que el del GM -estimador mostrando su falta de robustez. Para el caso de la contaminación C_2 los estimadores clásicos se alejan considerablemente de los parámetros verdaderos y los errores cuadráticos medios aumentan considerablemente. En la Figura 8.2 podemos ver el pobre comportamiento del estimador clásico y los M -estimadores en presencia de outliers en las covariables \mathbf{Z} . Recordemos que los M -estimadores toman valores extremos cuando hay presentes puntos de alta palanca. Por otro lado, al observar los diagramas de caja se ve que tanto los estimadores de LMS como de LTS se comportan robustamente ya que los mismos se ven muy similares para muestras normales y contaminadas. En general, los estimadores basados en M -suavizadores iniciales muestran menor variabilidad que aquellos basados en la mediana local. Si comparamos los estimadores obtenidos para β luego de un suavizado por medianas locales con aquellos obtenidos luego de un suavizado basado en M -estimadores locales, podemos observar que tanto cuando no hay contaminación (C_0) como en los casos donde contaminamos con C_1 y C_2 , el sesgo (entre la media de los estimadores y el valor verdadero) de los estimadores basados en medianas locales es ligeramente mayor que el sesgo de los basados en M -estimadores locales y además, los errores cuadráticos de los basados en M -estimadores son menores que los errores cuadráticos de los estimadores basados en la mediana local.

Con respecto a la estimación de la función de regresión g , al observar las tablas 8.4, 8.5, 8.8, 8.9, 8.15, 8.16, 8.19, 8.20, 8.30 y 8.31, se puede apreciar que bajo C_2 tanto los estimadores clásicos como los M -estimadores estiman inadecuadamente a la función de regresión. Los estimadores basados en GM -estimadores y aquellos basados en LMS -

y *LTS*–estimadores muestran un mejor desempeño. Observemos además que aún para estos estimadores, se evidencian valores grandes del $MSE(\hat{g})$ ya que las medias son considerablemente mayores que las medianas dadas las tablas 8.4 y 8.15. En las tablas 8.6, 8.7, 8.10, 8.11, 8.17, 8.18, 8.21, 8.22, 8.32 y 8.33 también se ve el mejor desempeño de los *GM*–estimadores, *LMS*– y *LTS*–estimadores bajo C_2 . Además, todos los métodos parecen no estar muy afectados ante la presencia de outliers en los errores (contaminación C_1).

En las Tablas 8.23 a 8.26 se presentan los resultados obtenidos por los estimadores robustos en dos pasos para distintas ventanas, lo que facilita su comparación. En las dos primeras tablas se encuentran los errores cuadráticos medios y errores cuadráticos medianos cuando se utiliza en el paso 1 una ventana $h_1 = 0,06$ y en el paso 2 las ventanas $h_2 = 0,06$ y $h_2 = 0,04$; y en las dos últimas tablas encontramos estas mismas medidas para el caso en que se utiliza en el paso 1 una ventana de $h_1 = 0,04$ y en el paso 2 $h_2 = 0,04$ y $h_2 = 0,03$. Para los casos considerados no se observa una mejoría sustancial al tomar la ventana $h_2 < h_1$, lo que podría indicar que la ventana h_1 considerada está cercana a la ventana óptima.

Como era de esperar, podemos concluir a grandes rasgos que el estimador clásico tiene un comportamiento muy pobre ante la presencia de contaminación C_2 , para todas las ventanas consideradas, mientras que los *GM*–estimadores y los estimadores basados en *LMS* y *LTS* logran superarlos.

	LS	GM	LMS	LTS	MH	MT	GM.med	LMS.med	LTS.med	MH.med	MT.med
Media	-1,00004	-0,99243	-0,99171	-0,98997	-0,98733	-0,98739	-0,96165	-0,96881	-0,9615	-0,96644	-0,96696
SD	0,00172	0,02368	0,04053	0,03175	0,01642	0,01702	0,06672	0,13992	0,11546	0,04584	0,04685
MSE	0	0,00062	0,00171	0,00111	0,00043	0,00045	0,00592	0,02053	0,0148	0,00323	0,00328
Mediana	-1,00001	-0,99245	-0,99217	-0,99132	-0,9891	-0,9892	-0,96295	-0,97864	-0,9711	-0,96669	-0,96712
MAD	0,00161	0,02107	0,03239	0,02456	0,01456	0,01544	0,0601	0,11856	0,10057	0,04306	0,0462
Media	-1,00007	-0,99251	-0,99168	-0,99005	-0,98736	-0,98744	-0,9616	-0,96541	-0,9615	-0,96646	-0,96698
SD	0,00317	0,02391	0,04326	0,03283	0,01688	0,01753	0,06686	0,14242	0,11622	0,04576	0,04673
MSE	1,00E-05	0,00063	0,00194	0,00118	0,00044	0,00046	0,00594	0,02146	0,01498	0,00322	0,00327
Mediana	-0,99997	-0,9928	-0,99532	-0,99284	-0,98914	-0,98916	-0,96354	-0,97621	-0,9682	-0,96676	-0,96678
MAD	0,00295	0,02132	0,03514	0,02638	0,01524	0,01601	0,06089	0,11799	0,09904	0,0437	0,04597
Media	0,98331	-0,9978	-0,99284	-0,99256	0,97733	0,97423	-0,83387	-0,83474	-0,8401	0,96346	0,96074
SD	0,23141	0,03793	0,11181	0,07647	0,23553	0,23896	0,11135	0,22008	0,16661	0,22552	0,22782
MSE	3,987	0,00144	0,01254	0,0059	3,96527	3,95464	0,03999	0,0757	0,05331	3,90598	3,89634
Mediana	0,97277	-0,99968	-0,99224	-0,99276	0,96663	0,96799	-0,84497	-0,85201	-0,854	0,96116	0,95778
MAD	0,23011	0,03466	0,09115	0,05616	0,23689	0,23869	0,10518	0,21227	0,1634	0,21936	0,22163

Tabla 8.1: Comportamiento de los estimadores de la primer coordenada β_1 de β para $h_1 = 0,04$ y $h_2 = 0,04$.

	LS	GM	LMS	LTS	MH	MT	GM.med	LMS.med	LTS.med	MH.med	MT.med
Media	2,99997	2,99392	2,99578	2,99607	2,99237	2,99425	2,96119	2,96377	2,96342	2,96403	2,96448
SD	0,00171	0,02417	0,04301	0,03374	0,01872	0,01803	0,06692	0,14114	0,11393	0,04585	0,04657
MSE	0	0,00062	0,00187	0,00115	0,00041	0,00036	0,00629	0,02121	0,0143	0,00339	0,00343
Mediana	3,00004	2,99469	2,99729	2,9985	2,99424	2,99583	2,96206	2,97094	2,96894	2,96675	2,96774
MAD	0,00172	0,02023	0,03243	0,02384	0,01555	0,01528	0,05934	0,11075	0,09395	0,043	0,04423
Media	2,99982	2,99382	2,99564	2,99546	2,99216	2,99401	2,96099	2,96192	2,96347	2,96384	2,96427
SD	0,003	0,02468	0,04435	0,034	0,01911	0,01846	0,06929	0,1406	0,11582	0,04589	0,04658
MSE	1,00E-05	0,00065	0,00198	0,00118	0,00043	0,00038	0,00632	0,0212	0,01473	0,00341	0,00344
Mediana	2,99991	2,99437	2,99669	2,99741	2,99386	2,99546	2,96238	2,97205	2,96975	2,96624	2,96726
MAD	0,00285	0,02033	0,032	0,024	0,01619	0,01578	0,05875	0,11472	0,09302	0,04248	0,04368
Media	1,02873	3,00199	3,00461	3,0076	1,01433	1,01163	2,83036	2,821	2,82331	1,00719	1,00536
SD	0,23275	0,03895	0,11189	0,08062	0,23617	0,23958	0,11556	0,24088	0,18532	0,22763	0,22965
MSE	3,94003	0,00152	0,01253	0,00655	3,99861	4,01096	0,04212	0,09001	0,06553	4,02307	4,03128
Mediana	1,02527	3,0027	3,0039	3,00444	1,00822	1,00614	2,84236	2,85034	2,84419	1,0055	1,00374
MAD	0,23234	0,03301	0,09517	0,06349	0,23033	0,24021	0,11035	0,22118	0,17276	0,22571	0,22746

Tabla 8.2: Comportamiento de los estimadores de la segunda coordenada β_2 de β para $h_1 = 0,04$ y $h_2 = 0,04$.

	LS	GM	LMS	LTS	MH	MT	GM.med	LMS.med	LTS.med	MH.med	MT.med
Media	0,00214	0,02876	0,04787	0,03735	0,02398	0,02334	0,09415	0,16696	0,1404	0,07136	0,0718
Mediana	0,002	0,02451	0,04013	0,03006	0,02093	0,02065	0,08416	0,14179	0,11953	0,06505	0,06566
Media	0,0038	0,02925	0,05009	0,03814	0,02447	0,02392	0,09429	0,16894	0,14187	0,07144	0,07184
Mediana	0,00349	0,02464	0,04028	0,02999	0,02128	0,02083	0,08451	0,14514	0,12113	0,06555	0,0651
Media	2,81512	0,04579	0,13156	0,08972	2,82173	2,82201	0,25585	0,33836	0,29012	2,81554	2,81526
Mediana	2,81181	0,04048	0,11034	0,07243	2,81657	2,81718	0,24236	0,29161	0,25574	2,81217	2,81141

Tabla 8.3: Comportamiento de los estimadores del parámetro de regresión β , medidas resumen de $\|\hat{\beta} - \beta\|$ para $h_1 = 0,04$ y $h_2 = 0,04$.

	LS	GM	LMS	LTS	MH	MT	GM.med	LMS.med	LTS.med	MH.med	MT.med	
Media	5,00E-05	0,00022	0,0005	0,00034	0,00017	0,00017	0,27751	0,27834	0,2767	0,27518	0,27525	C0
Mediana	5,00E-05	0,00013	0,00024	0,00017	0,00011	0,00011	0,25866	0,25946	0,25935	0,25776	0,25826	
Media	0,00012	0,00025	0,00059	0,00038	0,0002	0,0002	0,27785	0,2786	0,27728	0,27552	0,27561	C1
Mediana	0,00011	0,00016	0,0003	0,0002	0,00014	0,00014	0,26061	0,25851	0,25844	0,25732	0,25727	
Media	0,97622	0,25792	0,24637	0,25133	1,01498	1,01667	0,80205	0,79692	0,79456	0,90874	0,90869	C2
Mediana	0,92984	0,00033	0,00174	0,0008	0,98366	0,98259	0,66273	0,64923	0,66027	0,85461	0,85432	

Tabla 8.4: Error cuadrático medio de los estimadores de la función de regresión g para $h_1 = 0,04$ y $h_2 = 0,04$.

	LS	GM	LMS	LTS	MH	MT	GM.med	LMS.med	LTS.med	MH.med	MT.med	
Media	2,00E-05	0,0001	0,00024	0,00015	0,00007	0,00007	0,11972	0,11793	0,11682	0,11827	0,11827	C0
Mediana	1,00E-05	5,00E-05	1,00E-04	7,00E-05	4,00E-05	4,00E-05	0,1068	0,10508	0,10353	0,10689	0,10647	
Media	3,00E-05	0,00011	0,00027	0,00017	0,00008	0,00008	0,11989	0,11772	0,11729	0,11824	0,11823	C1
Mediana	3,00E-05	0,00006	0,00012	0,00008	0,00005	0,00005	0,10725	0,10604	0,10504	0,10587	0,1058	
Media	0,46315	0,00024	0,00169	0,00087	0,47923	0,48126	0,30873	0,29689	0,29918	0,45672	0,45724	C2
Mediana	0,42389	0,00013	0,00073	0,00033	0,43598	0,43796	0,27043	0,26264	0,26854	0,39182	0,39195	

Tabla 8.5: Error cuadrático mediano de los estimadores de la función de regresión g para $h_1 = 0,04$ y $h_2 = 0,04$.

	LS	GM	LMS	LTS	MH	MT	GM.med	LMS.med	LTS.med	MH.med	MT.med		
ECMP1	Media	2,00E-05	0,00012	0,00032	0,00021	0,00009	0,00009	0,02856	0,03122	0,02995	0,02787	0,02789	C0
	Mediana	2,00E-05	0,00007	0,00015	0,00009	0,00006	0,00005	0,02658	0,02909	0,02777	0,02586	0,02586	
ECMP2	Media	2,00E-05	0,00013	0,00034	0,00022	0,00009	0,00009	0,02937	0,03206	0,03077	0,02864	0,02867	C1
	Mediana	2,00E-05	0,00007	0,00015	0,00009	0,00006	0,00006	0,02699	0,02938	0,02839	0,02636	0,02637	
ECMP1	Media	3,00E-05	0,00013	0,00034	0,00022	0,00009	0,00009	0,02858	0,03125	0,03	0,02788	0,0279	C1
	Mediana	3,00E-05	0,00007	0,00015	0,00009	0,00006	0,00006	0,02671	0,02908	0,02786	0,0262	0,0262	
ECMP2	Media	3,00E-05	0,00014	0,00038	0,00024	0,0001	0,00009	0,02949	0,03218	0,0309	0,02874	0,02877	C2
	Mediana	3,00E-05	0,00007	0,00016	0,0001	0,00006	0,027	0,02902	0,0283	0,02638	0,0264	0,0264	
ECMP1	Media	0,63369	0,00025	0,00201	0,00099	0,64823	0,64855	0,079	0,08659	0,08223	0,75277	0,7527	C2
	Mediana	0,6373	0,00015	0,00099	0,00044	0,65049	0,65093	0,07392	0,08044	0,07668	0,75188	0,75246	
ECMP2	Media	0,66397	0,00027	0,00216	0,00107	0,67984	0,68024	0,08175	0,08952	0,08503	0,78826	0,78822	C2
	Mediana	0,64751	0,00015	0,00104	0,00046	0,66133	0,6612	0,07622	0,08317	0,07899	0,76507	0,76607	

Tabla 8.6: Errores cuadráticos medio de predicción $ECMP_{1,0}$ y $ECMP_{2,0}$ basado en datos no contaminados para $h_1 = 0,04$ y $h_2 = 0,04$.

	LS	GM	LMS	LTS	MH	MT	GM.med	LMS.med	LTS.med	MH.med	MT.med		
ECMP1	Media	0,00055	0,00083	0,00104	0,00092	0,00079	0,00078	0,05296	0,07499	0,0696	0,05456	0,05472	C1
	Mediana	0,00038	0,00054	0,00066	0,00059	0,00051	0,00052	0,02693	0,02943	0,02787	0,02636	0,02634	
ECMP2	Media	0,00102	0,00145	0,00189	0,00165	0,00144	0,00145	0,08115	0,10208	0,09761	0,08219	0,08239	C2
	Mediana	0,00045	0,00065	0,0008	0,00072	0,00062	0,00063	0,0301	0,03267	0,03141	0,02974	0,02975	
ECMP1	Media	0,71705	388,3857	388,771	389,1674	0,72995	0,73271	321,9757	322,3287	323,749	0,85828	0,86195	C2
	Mediana	0,70641	345,4569	344,0271	348,0131	0,7237	0,72616	288,1669	290,5742	287,672	0,84211	0,84652	
ECMP2	Media	0,88010	480,8141	480,5384	481,8572	0,90011	0,90452	399,3196	400,5861	401,763	1,05853	1,06435	C2
	Mediana	0,74145	366,2284	365,7533	365,082	0,75504	0,75591	302,7997	304,016	301,744	0,88716	0,89267	

Tabla 8.7: Errores cuadráticos medio de predicción $ECMP_{1,1}$ y $ECMP_{2,1}$ basado en datos contaminados para $h_1 = 0,04$ y $h_2 = 0,04$.

	LS	GM	LMS	LTS	MH	MT	GM.med	LMS.med	LTS.med	MH.med	MT.med
Media	5,00E-05	0,00026	0,00064	0,00042	0,0002	0,00019	0,27751	0,27834	0,2767	0,27518	0,27525
Mediana	5,00E-05	0,00015	0,00031	0,0002	0,00013	0,00012	0,25866	0,25946	0,25935	0,25776	0,25826
Media	0,00012	0,00031	0,00075	0,00047	0,00024	0,00023	0,27785	0,2786	0,27728	0,27552	0,27561
Mediana	0,00011	0,00019	0,00037	0,00024	0,00016	0,00015	0,26061	0,25851	0,25844	0,25732	0,25727
Media	0,97622	1,66595	1,6401	1,62974	1,34005	1,29768	0,80205	0,79692	0,79456	0,90874	0,90869
Mediana	0,92984	0,00043	0,00245	0,00107	1,28893	1,25436	0,66273	0,64923	0,66027	0,85461	0,85432

Tabla 8.8: Error cuadrático medio de los estimadores de la función de regresión g para $h_1 = 0,04$ y $h_2 = 0,03$.

	LS	GM	LMS	LTS	MH	MT	GM.med	LMS.med	LTS.med	MH.med	MT.med
Media	2,00E-05	0,00012	0,00029	0,00019	0,00008	0,00008	0,11972	0,11793	0,11682	0,11827	0,11827
Mediana	1,00E-05	0,00006	0,00013	0,00008	0,00005	0,00005	0,1068	0,10508	0,10353	0,10689	0,10647
Media	3,00E-05	0,00013	0,00033	0,00021	0,0001	0,0001	0,11989	0,11772	0,11729	0,11824	0,11823
Mediana	3,00E-05	0,00007	0,00015	0,00009	0,00006	0,00006	0,10725	0,10604	0,10504	0,10587	0,1058
Media	0,46315	0,00031	0,00214	0,00111	0,60598	0,59206	0,30873	0,29689	0,29918	0,45672	0,45724
Mediana	0,42389	0,00016	0,00096	0,00043	0,54833	0,53867	0,27043	0,26264	0,26854	0,39182	0,39195

Tabla 8.9: Error cuadrático mediano de los estimadores de la función de regresión g para $h_1 = 0,04$ y $h_2 = 0,03$.

	LS	GM	LMS	LTS	MH	MT	GM.med	LMS.med	LTS.med	MH.med	MT.med
ECMP1	Media	2,00E-05	0,00011	0,0003	0,0002	0,00008	0,00008	0,02856	0,03122	0,02995	0,02787
	Mediana	2,00E-05	0,00006	0,00014	0,00009	0,00005	0,00005	0,02658	0,02909	0,02777	0,02586
ECMP2	Media	2,00E-05	0,00013	0,00032	0,00021	0,00009	0,00008	0,02937	0,03206	0,03077	0,02864
	Mediana	2,00E-05	0,00006	0,00014	0,00009	0,00005	0,00005	0,02699	0,02938	0,02839	0,02636
ECMP1	Media	3,00E-05	0,00012	0,00033	0,00021	0,00009	0,00008	0,02858	0,03125	0,03	0,02788
	Mediana	3,00E-05	0,00006	0,00014	0,00009	0,00005	0,00005	0,02671	0,02908	0,02786	0,0262
ECMP2	Media	3,00E-05	0,00013	0,00036	0,00022	0,00009	0,00009	0,02949	0,03218	0,0309	0,02874
	Mediana	3,00E-05	0,00007	0,00015	0,00009	0,00006	0,00005	0,027	0,02902	0,0283	0,02638
ECMP1	Media	0,63369	0,0097	0,01123	0,01026	0,61341	0,60801	0,079	0,08659	0,08223	0,75277
	Mediana	0,6373	0,00014	0,00094	0,00041	0,61288	0,60727	0,07392	0,08044	0,07668	0,75188
ECMP2	Media	0,66397	0,01067	0,01238	0,01131	0,64354	0,63798	0,08175	0,08952	0,08503	0,78822
	Mediana	0,64751	0,00014	0,00099	0,00043	0,6241	0,62006	0,07622	0,08317	0,07899	0,76507

Tabla 8.10: Errores cuadráticos medio de predicción $ECMP_{1,0}$ y $ECMP_{2,0}$ basado en datos no contaminados para $h_1 = 0,04$ y $h_2 = 0,03$.

	LS	GM	LMS	LTS	MH	MT	GM.med	LMS.med	LTS.med	MH.med	MT.med
ECMP1	Media	0,00055	0,0008	0,00099	0,00088	0,00077	0,00075	0,05296	0,07499	0,0696	0,05472
	Mediana	0,00038	0,00051	0,00063	0,00057	0,0005	0,00049	0,02693	0,02943	0,02787	0,02636
ECMP2	Media	0,00102	0,00139	0,00178	0,00156	0,00141	0,0014	0,08115	0,10208	0,09761	0,08219
	Mediana	0,00045	0,00061	0,00077	0,00068	0,00059	0,0006	0,0301	0,03267	0,03141	0,02974
ECMP1	Media	0,71705	387,474	387,867	388,269	0,68681	0,68632	321,976	322,329	323,749	0,85828
	Mediana	0,70641	345,344	344,014	347,902	0,6779	0,67674	288,167	290,574	287,672	0,84211
ECMP2	Media	0,88010	479,727	479,468	480,793	0,84814	0,84886	399,32	400,586	401,763	1,05853
	Mediana	0,774145	365,581	364,524	364,689	0,69983	0,70289	302,8	304,016	301,744	0,88716

Tabla 8.11: Errores cuadráticos medio de predicción $ECMP_{1,1}$ y $ECMP_{2,1}$ basado en datos contaminados para $h_1 = 0,04$ y $h_2 = 0,03$.

	LS	GM	LMS	LTS	MH	MT	GM.med	LMS.med	LTS.med	MH.med	MT.med
Media	-1,00004	-0,99463	-0,99227	-0,99052	-0,99237	-0,99206	-0,96982	-0,96779	-0,97008	-0,97478	-0,97501
SD	0,00174	0,01776	0,03772	0,03003	0,01124	0,01185	0,05542	0,12816	0,10961	0,03945	0,03969
MSE	0	0,00034	0,00148	0,00099	0,00018	0,0002	0,00398	0,01745	0,0129	0,00219	0,0022
Mediana	-1,00003	-0,99544	-0,99357	-0,99293	-0,99328	-0,99313	-0,97162	-0,97714	-0,97386	-0,97545	-0,97603
MAD	0,00168	0,01558	0,02809	0,02352	0,0099	0,01034	0,04997	0,10677	0,09447	0,0382	0,03833
Media	-1,00007	-0,99473	-0,99335	-0,99071	-0,99243	-0,99216	-0,96981	-0,96752	-0,9681	-0,97485	-0,97505
SD	0,00312	0,01812	0,03758	0,03023	0,01163	0,01216	0,05546	0,12787	0,10806	0,0395	0,03975
MSE	1,00E-05	0,00036	0,00146	0,001	0,00019	0,00021	0,00398	0,01739	0,01268	0,00219	0,0022
Mediana	-0,99997	-0,99572	-0,99478	-0,99327	-0,99325	-0,99297	-0,97183	-0,97569	-0,97245	-0,97588	-0,97585
MAD	0,00289	0,01636	0,02937	0,0229	0,00981	0,01082	0,05069	0,10955	0,0966	0,03828	0,03833
Media	0,98172	-0,9982	-0,99373	-0,99201	0,98014	0,97867	-0,85316	-0,82829	-0,83365	0,96806	0,96591
SD	0,22397	0,03107	0,10558	0,07867	0,22606	0,22742	0,10595	0,21674	0,16674	0,22215	0,22334
MSE	3,97733	0,00097	0,01117	0,00625	3,97201	3,96682	0,03278	0,07501	0,05545	3,92257	3,91461
Mediana	0,97025	-0,99897	-0,99441	-0,99357	0,96883	0,96915	-0,86579	-0,84922	-0,84899	0,96263	0,96239
MAD	0,22334	0,02733	0,08581	0,06446	0,23009	0,22566	0,09726	0,20393	0,16923	0,22345	0,2219

Tabla 8.12: Comportamiento de los estimadores de la primer coordenada β_1 de β para $h_1 = 0,06$ y $h_2 = 0,06$.

	LS	GM	LMS	LTS	MH	MT	GM.med	LMS.med	LTS.med	MH.med	MT.med
Media	2,99996	2,99679	2,9973	2,99688	2,99641	2,99715	2,9678	2,96201	2,9686	2,97035	2,97072
SD	0,00173	0,01849	0,03914	0,03252	0,01221	0,01183	0,06169	0,13061	0,1115	0,04169	0,04178
MSE	0	0,00035	0,00154	0,00107	0,00016	0,00015	0,00484	0,01848	0,01341	0,00262	0,0026
Mediana	3,00003	2,997	2,99863	2,9983	2,99756	2,998	2,97272	2,97635	2,97939	2,97257	2,97248
MAD	0,00177	0,01435	0,02899	0,02182	0,01015	0,00927	0,05129	0,10721	0,08696	0,03802	0,03774
Media	2,99982	2,99655	2,99811	2,9965	2,99624	2,99698	2,96771	2,96071	2,96877	2,97021	2,97056
SD	0,00294	0,01885	0,03919	0,03281	0,01252	0,01214	0,06179	0,13069	0,11064	0,04182	0,04192
MSE	1,00E-05	0,00037	0,00154	0,00109	0,00017	0,00016	0,00486	0,01861	0,0132	0,00264	0,00262
Mediana	2,99986	2,99706	2,99845	2,99757	2,9975	2,99824	2,97242	2,9728	2,97906	2,97234	2,97199
MAD	0,00276	0,01462	0,02953	0,02236	0,01042	0,00958	0,05061	0,10981	0,0866	0,03822	0,03813
Media	1,02662	3,0031	3,00951	3,00949	1,02025	1,01932	2,84205	2,8185	2,81207	1,01307	1,01182
SD	0,22522	0,02986	0,10201	0,08161	0,22652	0,22769	0,10832	0,2421	0,19282	0,22489	0,22587
MSE	3,94491	0,0009	0,01049	0,00674	3,97067	3,97488	0,03667	0,0915	0,07246	3,99842	4,00381
Mediana	1,01963	3,00358	3,00607	3,00569	1,00855	1,00959	2,8608	2,85109	2,84441	1,00909	1,01248
MAD	0,22553	0,02707	0,07906	0,06651	0,22343	0,2226	0,0988	0,21893	0,18331	0,22104	0,21795

Tabla 8.13: Comportamiento de los estimadores de la segunda coordenada β_2 de β para $h_1 = 0,06$ y $h_2 = 0,06$.

	LS	GM	LMS	LTS	MH	MT	GM.med	LMS.med	LTS.med	MH.med	MT.med
Media	0,00217	0,02128	0,04336	0,03499	0,01543	0,01544	0,079	0,15484	0,13195	0,05984	0,05983
Mediana	0,00205	0,01756	0,03505	0,02676	0,01295	0,01297	0,06864	0,13656	0,11244	0,0537	0,05422
Media	0,00374	0,0218	0,04383	0,03537	0,01585	0,01584	0,07913	0,1552	0,13139	0,05999	0,05999
Mediana	0,00341	0,01837	0,03588	0,02741	0,01352	0,01336	0,06787	0,13527	0,1132	0,05388	0,05437
Media	2,8143	0,03663	0,12046	0,0939	2,81797	2,81779	0,23289	0,33733	0,30283	2,81411	2,81364
Mediana	2,81162	0,03211	0,10121	0,07982	2,81499	2,81454	0,20962	0,28926	0,26135	2,81009	2,80982

Tabla 8.14: Comportamiento de los estimadores del parámetro de regresión β , medidas resumen de $\|\hat{\beta} - \beta\|$ para $h_1 = 0,06$ y $h_2 = 0,06$.

	LS	GM	LMS	LTS	MH	MT	GM.med	LMS.med	LTS.med	MH.med	MT.med	
Media	8,00E-05	0,00015	0,00035	0,00027	0,00013	0,00013	0,22099	0,22056	0,22033	0,21925	0,21929	C0
Mediana	8,00E-05	0,00012	0,0002	0,00016	0,00011	0,00011	0,20143	0,20126	0,20076	0,20088	0,20103	
Media	0,00012	0,00017	0,00038	0,0003	0,00014	0,00014	0,22123	0,22076	0,22052	0,21949	0,21953	C1
Mediana	0,00012	0,00014	0,00022	0,00018	0,00012	0,00012	0,20216	0,1999	0,20033	0,20043	0,20067	
Media	0,66584	0,00031	0,00218	0,00134	0,68432	0,68528	0,58456	0,57258	0,56834	0,6385	0,63845	C2
Mediana	0,62645	0,00022	0,001	0,00067	0,64033	0,64197	0,53325	0,52612	0,51754	0,58319	0,58287	

Tabla 8.15: Error cuadrático medio de los estimadores de la función de regresión g para $h_1 = 0,06$ y $h_2 = 0,06$.

	LS	GM	LMS	LTS	MH	MT	GM.med	LMS.med	LTS.med	MH.med	MT.med	
Media	1,00E-05	0,00005	0,00015	0,00011	0,00003	0,00003	0,10165	0,10012	0,09957	0,1004	0,10038	C0
Mediana	1,00E-05	3,00E-05	6,00E-05	4,00E-05	2,00E-05	2,00E-05	0,08721	0,08478	0,08556	0,08647	0,08659	
Media	3,00E-05	0,00005	0,00017	0,00012	0,00004	0,00004	0,10191	0,10036	0,09995	0,10071	0,10068	C1
Mediana	3,00E-05	3,00E-05	7,00E-05	5,00E-05	3,00E-05	3,00E-05	0,08785	0,08442	0,08579	0,08674	0,08661	
Media	0,3343	0,00012	0,00109	0,00065	0,34328	0,34399	0,30319	0,29011	0,28857	0,34628	0,34638	C2
Mediana	0,2922	0,00007	0,00044	0,00027	0,30029	0,30119	0,25794	0,25406	0,25108	0,27797	0,27933	

Tabla 8.16: Error cuadrático mediano de los estimadores de la función de regresión g para $h_1 = 0,06$ y $h_2 = 0,06$.

	LS	GM	LMS	LTS	MH	MT	GM.med	LMS.med	LTS.med	MH.med	MT.med
ECMP1	2,00E-05	0,00009	0,00029	0,00021	0,00006	0,00006	0,02266	0,0251	0,02421	0,02216	0,02218
Mediana	2,00E-05	0,00005	0,00013	0,00009	0,00004	0,00004	0,02102	0,02272	0,02206	0,02067	0,02069
ECMP2	3,00E-05	0,00009	0,00031	0,00023	0,00006	0,00006	0,02329	0,02576	0,02488	0,02276	0,02278
Mediana	2,00E-05	0,00006	0,00013	0,00009	0,00004	0,00004	0,02098	0,02303	0,02251	0,02059	0,02063
ECMP1	3,00E-05	0,00009	0,00029	0,00021	0,00006	0,00006	0,0227	0,02512	0,02419	0,02219	0,02221
Mediana	3,00E-05	0,00006	0,00014	0,00009	0,00004	0,00004	0,02076	0,02256	0,02185	0,02023	0,02023
ECMP2	3,00E-05	0,0001	0,00031	0,00023	0,00006	0,00006	0,0234	0,02587	0,02492	0,02287	0,02289
Mediana	3,00E-05	0,00006	0,00015	0,00009	0,00004	0,00005	0,02081	0,02275	0,02215	0,02052	0,02053
ECMP1	0,67889	0,00018	0,00183	0,00111	0,68739	0,6874	0,06483	0,07254	0,06857	0,76337	0,76316
Mediana	0,68207	0,00011	0,00089	0,00057	0,69109	0,69151	0,05871	0,06424	0,06205	0,76391	0,76496
ECMP2	0,71124	0,0002	0,00196	0,00117	0,72058	0,72061	0,06694	0,07489	0,07077	0,7993	0,7991
Mediana	0,69488	0,00011	0,00089	0,00057	0,70075	0,70057	0,06054	0,0666	0,0639	0,78662	0,78702

Tabla 8.17: Errores cuadráticos medio de predicción $ECMP_{1,0}$ y $ECMP_{2,0}$ basado en datos no contaminados para $h_1 = 0,06$ y $h_2 = 0,06$.

	LS	GM	LMS	LTS	MH	MT	GM.med	LMS.med	LTS.med	MH.med	MT.med
ECMP1	0,00062	0,00082	0,00106	0,0011	0,00078	0,00079	0,03481	0,04922	0,04453	0,03617	0,03627
Mediana	0,00042	0,00054	0,00069	0,00063	0,00052	0,00053	0,02188	0,02474	0,0236	0,02141	0,02136
ECMP2	0,00113	0,00145	0,00184	0,00182	0,0014	0,00141	0,06003	0,07647	0,07268	0,06123	0,06134
Mediana	0,00051	0,00063	0,00081	0,00074	0,00063	0,00062	0,02574	0,02736	0,02656	0,02479	0,02454
ECMP1	0,77057	388,8611	389,8314	389,6377	0,7793	0,78192	327,5217	320,8354	319,712	0,87303	0,87624
Mediana	0,76125	347,4315	343,8975	346,9685	0,77413	0,77463	293,3562	282,378	285,495	0,86115	0,86527
ECMP2	0,94734	481,5662	481,98751	482,5103	0,96173	0,96607	406,2275	398,1467	397,605	1,08006	1,085
Mediana	0,79588	369,5897	369,38032	367,3978	0,8123	0,80844	306,9335	301,8186	296,373	0,90451	0,91182

Tabla 8.18: Errores cuadráticos medio de predicción $ECMP_{1,1}$ y $ECMP_{2,1}$ basado en datos contaminados para $h_1 = 0,06$ y $h_2 = 0,06$.

	LS	GM	LMS	LTS	MH	MT	GM.med	LMS.med	LTS.med	MH.med	MT.med
Media	8,00E-05	0,00015	0,00045	0,00033	0,00011	0,00011	0,22099	0,22056	0,22033	0,21925	0,21929
Mediana	8,00E-05	0,00011	0,00021	0,00016	0,00009	0,00009	0,20143	0,20126	0,20076	0,20088	0,20103
Media	0,00012	0,00018	0,00049	0,00036	0,00013	0,00013	0,22123	0,22076	0,22052	0,21949	0,21953
Mediana	0,00012	0,00013	0,00025	0,00019	0,00011	0,00011	0,20216	0,1999	0,20033	0,20043	0,20067
Media	0,66584	0,25626	0,254	0,2609	1,00847	0,9742	0,58456	0,57258	0,56834	0,6385	0,63845
Mediana	0,62645	0,00025	0,00146	0,00094	0,97484	0,93242	0,53325	0,52612	0,51754	0,58319	0,58287

Tabla 8.19: Error cuadrático medio de los estimadores de la función de regresión g para $h_1 = 0,06$ y $h_2 = 0,04$.

	LS	GM	LMS	LTS	MH	MT	GM.med	LMS.med	LTS.med	MH.med	MT.med
Media	1,00E-05	0,00006	0,00021	0,00015	0,00004	0,00004	0,10165	0,10012	0,09957	0,1004	0,10038
Mediana	1,00E-05	3,00E-05	8,00E-05	6,00E-05	3,00E-05	3,00E-05	0,08721	0,08478	0,08556	0,08647	0,08659
Media	3,00E-05	0,00007	0,00022	0,00016	0,00005	0,00005	0,10191	0,10036	0,09995	0,10071	0,10068
Mediana	3,00E-05	4,00E-05	9,00E-05	7,00E-05	3,00E-05	3,00E-05	0,08785	0,08442	0,08579	0,08674	0,08661
Media	0,3343	0,00016	0,00151	0,0009	0,47776	0,46485	0,30319	0,29011	0,28857	0,34628	0,34638
Mediana	0,2922	0,00009	0,00063	0,00038	0,43057	0,41804	0,25794	0,25406	0,25108	0,27797	0,27933

Tabla 8.20: Error cuadrático mediano de los estimadores de la función de regresión g para $h_1 = 0,06$ y $h_2 = 0,04$.

	LS	GM	LMS	LTS	MH	MT	GM.med	LMS.med	LTS.med	MH.med	MT.med	
ECMP1	Mediana	2,00E-05	0,00008	0,00027	0,00019	0,00005	0,00005	0,02266	0,0251	0,02421	0,02216	0,02218
	Media	2,00E-05	0,00005	0,00012	0,00008	0,00003	0,00004	0,02102	0,02272	0,02206	0,02067	0,02069
ECMP2	Mediana	3,00E-05	0,00008	0,00029	0,00021	0,00005	0,00005	0,02329	0,02576	0,02488	0,02276	0,02278
	Media	2,00E-05	0,00005	0,00012	0,00008	0,00004	0,00004	0,02098	0,02303	0,02251	0,02059	0,02063
ECMP1	Mediana	3,00E-05	0,00008	0,00027	0,00019	0,00005	0,00005	0,0227	0,02512	0,02419	0,02219	0,02221
	Media	3,00E-05	0,00005	0,00013	0,00008	0,00004	0,00004	0,02076	0,02256	0,02185	0,02023	0,02023
ECMP2	Mediana	3,00E-05	0,00009	0,00028	0,00021	0,00005	0,00005	0,0234	0,02587	0,02492	0,02287	0,02289
	Media	3,00E-05	0,00005	0,00013	0,00008	0,00004	0,00004	0,02081	0,02275	0,02215	0,02052	0,02053
ECMP1	Mediana	0,67889	0,00017	0,0017	0,00103	0,64738	0,6441	0,06483	0,07254	0,06857	0,76337	0,76316
	Media	0,68207	0,0001	0,00084	0,00053	0,64956	0,64665	0,05871	0,06424	0,06205	0,76391	0,76496
ECMP2	Mediana	0,71124	0,00018	0,00182	0,00109	0,6789	0,67551	0,06694	0,07489	0,07077	0,7993	0,7991
	Media	0,69488	0,0001	0,00082	0,00053	0,65984	0,65895	0,06054	0,0666	0,0639	0,78662	0,78702

Tabla 8.21: Errores cuadráticos medio de predicción $ECMP_{1,0}$ y $ECMP_{2,0}$ basado en datos no contaminados para $h_1 = 0,06$ y $h_2 = 0,04$.

	LS	GM	LMS	LTS	MH	MT	GM.med	LMS.med	LTS.med	MH.med	MT.med	
ECMP1	Mediana	0,00062	0,0008	0,00102	0,00109	0,00076	0,03481	0,04922	0,04453	0,03617	0,03627	
	Media	0,00042	0,00052	0,00066	0,0006	0,00051	0,02188	0,02474	0,0236	0,02141	0,02136	
ECMP2	Mediana	0,00113	0,00141	0,00177	0,0018	0,00137	0,06003	0,07647	0,07268	0,06123	0,06134	
	Media	0,00051	0,00061	0,00079	0,00072	0,00061	0,02574	0,02736	0,02656	0,02479	0,02454	
ECMP1	Mediana	0,77057	388,60651	389,583	389,38378	0,7335	0,73301	327,522	320,835	319,712	0,87303	0,87624
	Media	0,76125	347,3924	343,712	347,00865	0,72745	0,72859	293,356	282,378	285,495	0,86115	0,86527
ECMP2	Mediana	0,94734	481,25555	481,685	482,19816	0,90495	0,90619	406,228	398,147	397,605	1,08006	1,085
	Media	0,79588	367,8427	368,415	366,53723	0,75188	0,74594	306,933	301,819	296,373	0,90451	0,91182

Tabla 8.22: Errores cuadráticos medio de predicción $ECMP_{1,1}$ y $ECMP_{2,1}$ basado en datos contaminados para $h_1 = 0,06$ y $h_2 = 0,04$.

		h ₁ =0,04, h ₂ =0,04					h ₁ =0,04, h ₂ =0,03				
		GM	LMS	LTS	MH	MT	GM	LMS	LTS	MH	MT
Media		0,0022	0,0005	0,00034	0,00017	0,00017	0,00026	0,00064	0,00042	0,0002	0,00019
Mediana		0,00013	0,00024	0,00017	0,00011	0,00011	0,00015	0,00031	0,0002	0,00013	0,00012
Media		0,00025	0,00059	0,00038	0,0002	0,0002	0,00031	0,00075	0,00047	0,00024	0,00023
Mediana		0,00016	0,0003	0,0002	0,00014	0,00014	0,00019	0,00037	0,00024	0,00016	0,00015
Media		0,25792	0,24637	0,25133	1,01498	1,01667	1,66595	1,6401	1,62974	1,34005	1,29768
Mediana		0,00033	0,00174	0,0008	0,98366	0,98259	0,00043	0,00245	0,00107	1,28893	1,25436

Tabla 8.23: Error cuadrático medio de los estimadores robustos en dos pasos con $h_1 = 0,04$ y $h_2 = 0,04$ y $0,03$.

		h ₁ =0,04, h ₂ =0,04					h ₁ =0,04, h ₂ =0,03				
		GM	LMS	LTS	MH	MT	GM	LMS	LTS	MH	MT
Media		0,0001	0,00024	0,00015	0,00007	0,00007	0,00012	0,00029	0,00019	0,00008	0,00008
Mediana		5,00E-05	1,00E-04	7,00E-05	4,00E-05	4,00E-05	0,00006	0,00013	0,00008	0,00005	0,00005
Media		0,00011	0,00027	0,00017	0,00008	0,00008	0,00013	0,00033	0,00021	0,0001	0,0001
Mediana		0,00006	0,00012	0,00008	0,00005	0,00005	0,00007	0,00015	0,00009	0,00006	0,00006
Media		0,00024	0,00169	0,00087	0,47923	0,48126	0,00031	0,00214	0,00111	0,60598	0,59206
Mediana		0,00013	0,00073	0,00033	0,43598	0,43796	0,00016	0,00096	0,00043	0,54833	0,53867

Tabla 8.24: Error cuadrático mediano de los estimadores robustos en dos pasos con $h_1 = 0,04$ y $h_2 = 0,04$ y $0,03$.

		h ₁ =0,06, h ₂ =0,06						h ₁ =0,06, h ₂ =0,04								
		GM	LMS	LTS	MH	MT	GM	LMS	LTS	MH	MT	GM	LMS	LTS	MH	MT
Media		0,00015	0,00035	0,00027	0,00013	0,00013	0,00015	0,00045	0,00033	0,00011	0,00011	0,00015	0,00045	0,00033	0,00011	0,00011
Mediana		0,00012	0,0002	0,00016	0,00011	0,00011	0,00011	0,00021	0,00016	0,00009	0,00009	0,00011	0,00021	0,00016	0,00009	0,00009
Media		0,00017	0,00038	0,0003	0,00014	0,00014	0,00018	0,00049	0,00036	0,00013	0,00013	0,00018	0,00049	0,00036	0,00013	0,00013
Mediana		0,00014	0,00022	0,00018	0,00012	0,00012	0,00013	0,00025	0,00019	0,00011	0,00011	0,00013	0,00025	0,00019	0,00011	0,00011
Media		0,00031	0,00218	0,00134	0,68432	0,68528	0,25626	0,254	0,2609	1,00847	0,9742	0,25626	0,254	0,2609	1,00847	0,9742
Mediana		0,00022	0,001	0,00067	0,64033	0,64197	0,00025	0,00146	0,00094	0,97484	0,93242	0,00025	0,00146	0,00094	0,97484	0,93242

Tabla 8.25: Error cuadrático medio de los estimadores robustos en dos pasos con $h_1 = 0,06$ y $h_2 = 0,06$ y $0,04$.

		h ₁ =0,06, h ₂ =0,06						h ₁ =0,06, h ₂ =0,04								
		GM	LMS	LTS	MH	MT	GM	LMS	LTS	MH	MT	GM	LMS	LTS	MH	MT
Media		0,00005	0,00015	0,00011	0,00003	0,00003	0,00006	0,00021	0,00015	0,00004	0,00004	0,00006	0,00021	0,00015	0,00004	0,00004
Mediana		3,00E-05	6,00E-05	4,00E-05	2,00E-05	2,00E-05	3,00E-05	8,00E-05	6,00E-05	3,00E-05	3,00E-05	3,00E-05	8,00E-05	6,00E-05	3,00E-05	3,00E-05
Media		0,00005	0,00017	0,00012	0,00004	0,00004	0,00007	0,00022	0,00016	0,00005	0,00005	0,00007	0,00022	0,00016	0,00005	0,00005
Mediana		3,00E-05	7,00E-05	5,00E-05	3,00E-05	3,00E-05	4,00E-05	9,00E-05	7,00E-05	3,00E-05	3,00E-05	4,00E-05	9,00E-05	7,00E-05	3,00E-05	3,00E-05
Media		0,00012	0,00109	0,00065	0,34328	0,34399	0,00016	0,00151	0,0009	0,47776	0,46485	0,00016	0,00151	0,0009	0,47776	0,46485
Mediana		0,00007	0,00044	0,00027	0,30029	0,30119	0,00009	0,00063	0,00038	0,43057	0,41804	0,00009	0,00063	0,00038	0,43057	0,41804

Tabla 8.26: Error cuadrático mediano de los estimadores robustos en dos pasos con $h_1 = 0,06$ y $h_2 = 0,06$ y $0,04$.

	LS	GM	LMS	LTS	MH	MT	
Media	-1,00007	-0,9976	-0,9947	-0,9945	-0,9978	-0,9977	C0
SD	0,00205	0,0108	0,03189	0,02683	0,00541	0,0056	
MSE	0	0,00012	0,00104	0,00075	0,00003	0,00004	
Mediana	-1	-0,998	-0,9961	-0,9966	-0,9983	-0,9982	
MAD	0,00199	0,00905	0,02412	0,01862	0,00454	0,0045	
Media	-1,0001	-0,9977	-0,9955	-0,9959	-0,9978	-0,9978	C1
SD	0,00328	0,01125	0,03209	0,0272	0,00596	0,00614	
MSE	0,00001	0,00013	0,00105	0,00076	0,00004	0,00004	
Mediana	-1,00008	-0,9981	-0,9967	-0,997	-0,9982	-0,9981	
MAD	0,00321	0,00958	0,02432	0,0191	0,00544	0,00558	
Media	0,98031	-0,9985	-0,9973	-0,9979	0,98033	0,97952	C2
SD	0,21585	0,01866	0,07729	0,06095	0,21723	0,21761	
MSE	3,96816	0,00035	0,00598	0,00372	3,96886	3,96582	
Mediana	0,97125	-0,9991	-0,9965	-0,9963	0,97112	0,96982	
MAD	0,20812	0,01595	0,05267	0,04749	0,2068	0,21008	

Tabla 8.27: Comportamiento de los estimadores de la primer coordenada β_1 de β para $h_1 = 0,15$ y $h_2 = 0,15$.

	LS	GM	LMS	LTS	MH	MT	
Media	2,99999	2,99866	2,99773	2,99703	2,99913	2,99913	C0
SD	0,00209	0,01193	0,03359	0,02932	0,00545	0,0056	
MSE	0	0,00014	0,00113	0,00087	0,00003	0,00003	
Mediana	3,00009	2,99901	2,99813	2,99853	2,99948	2,9994	
MAD	0,00202	0,00882	0,02215	0,01759	0,00469	0,00485	
Media	2,99983	2,99851	2,99718	2,99645	2,999	2,999	C1
SD	0,00312	0,01222	0,03175	0,02843	0,00585	0,00597	
MSE	0,00001	0,00015	0,00102	0,00082	0,00004	0,00004	
Mediana	2,99987	2,99848	2,99824	2,99761	2,99923	2,99926	
MAD	0,0031	0,00874	0,02069	0,01815	0,00543	0,00532	
Media	1,02711	3,00165	3,00268	3,00592	1,02465	1,02477	C2
SD	0,21738	0,01934	0,07655	0,06394	0,21795	0,21805	
MSE	3,93951	0,00038	0,00586	0,00412	3,94946	3,94902	
Mediana	1,01614	3,00219	3,00473	3,00522	1,01519	1,0169	
MAD	0,20771	0,01646	0,05507	0,04859	0,20282	0,20345	

Tabla 8.28: Comportamiento de los estimadores de la segunda coordenada β_2 de β para $h_1 = 0,15$ y $h_2 = 0,15$.

	LS	GM	LMS	LTS	MH	MT	
Media	0,00259	0,01271	0,03585	0,02998	0,00669	0,00688	C0
Mediana	0,00246	0,01036	0,02699	0,02263	0,00577	0,00583	
Media	0,00397	0,01323	0,03548	0,03001	0,00735	0,00751	C1
Mediana	0,00369	0,01078	0,02765	0,02258	0,00656	0,00661	
Media	2,81173	0,02259	0,08622	0,07084	2,81365	2,81302	C2
Mediana	2,80785	0,01937	0,06942	0,0593	2,80975	2,80979	

Tabla 8.29: Comportamiento de los estimadores del parámetro de regresión β , medidas resumen de $\|\hat{\beta} - \beta\|$ para $h_1 = 0,15$ y $h_2 = 0,15$.

	LS	GM	LMS	LTS	MH	MT	
Media	0,00022	0,00025	0,00031	0,0003	0,00024	0,00024	C0
Mediana	0,00022	0,00024	0,00028	0,00027	0,00024	0,00024	
Media	0,00024	0,00026	0,00032	0,00031	0,00025	0,00025	C1
Mediana	0,00024	0,00025	0,00028	0,00028	0,00024	0,00024	
Media	0,29648	0,00028	0,00074	0,00058	0,3009	0,30133	C2
Mediana	0,25416	0,00027	0,00043	0,00039	0,2598	0,26011	

Tabla 8.30: Error cuadrático medio de los estimadores de la función de regresión g para $h_1 = 0,15$ y $h_2 = 0,15$.

	LS	GM	LMS	LTS	MH	MT	
Media	0,00002	3,00E-05	8,00E-05	6,00E-05	2,00E-05	2,00E-05	C0
Mediana	0,00001	2,00E-05	3,00E-05	3,00E-05	1,00E-05	1,00E-05	
Media	0,00003	3,00E-05	8,00E-05	7,00E-05	2,00E-05	2,00E-05	C1
Mediana	0,00002	2,00E-05	4,00E-05	3,00E-05	2,00E-05	2,00E-05	
Media	0,18143	0,00005	0,00034	0,00025	0,18485	0,18469	C2
Mediana	0,13331	0,00003	0,00012	0,00009	0,13897	0,13736	

Tabla 8.31: Error cuadrático mediano de los estimadores de la función de regresión g para $h_1 = 0,15$ y $h_2 = 0,15$.

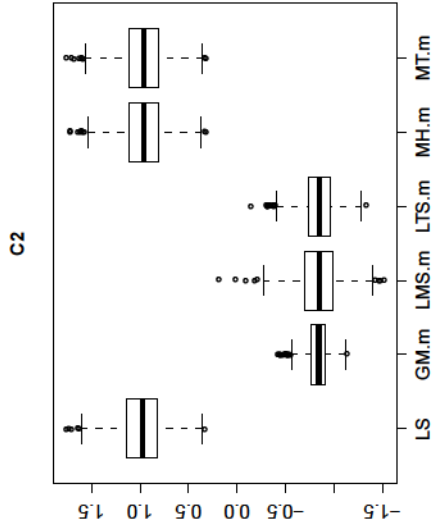
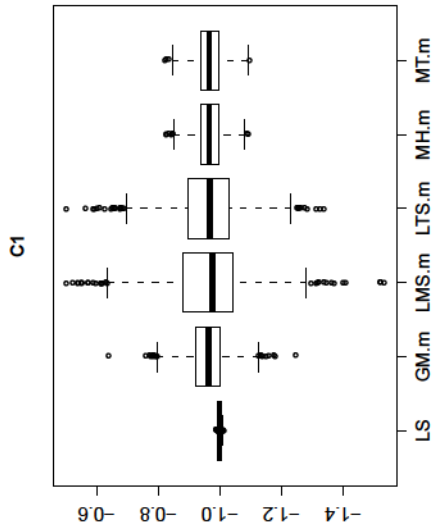
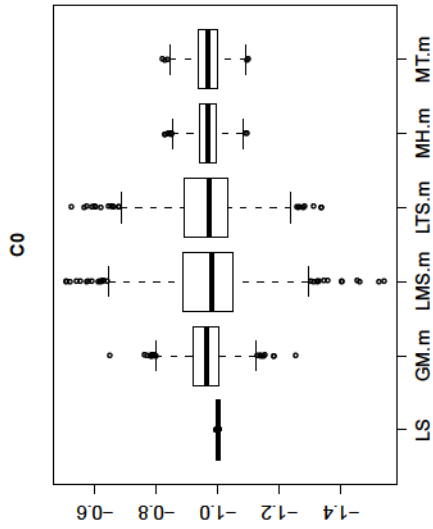
		LS	GM	LMS	LTS	MH	MT
ECMP1	Media	0,00004	0,00007	0,00025	0,00020	0,00005	0,00005
	Mediana	0,00004	0,00006	0,00011	0,00009	0,00005	0,00005
ECMP2	Media	4,00E-05	0,00007	0,00027	0,00021	0,00005	0,00005
	Mediana	4,00E-05	0,00006	0,00011	0,00009	0,00005	0,00005
ECMP1	Media	0,00004	0,00007	0,00024	0,00019	0,00005	0,00005
	Mediana	0,00004	0,00006	0,00012	0,00009	0,00005	0,00005
ECMP2	Media	5,00E-05	0,00007	0,00026	0,00021	0,00005	0,00005
	Mediana	4,00E-05	0,00006	0,00012	0,00009	0,00005	0,00005
ECMP1	Media	0,73354	0,00010	0,00111	0,00075	0,73684	0,73663
	Mediana	0,7366	0,00007	0,00049	0,00035	0,73990	0,73983
ECMP2	Media	0,76835	0,00011	0,00118	0,00080	0,77205	0,77185
	Mediana	0,74952	0,00007	0,00048	0,00036	0,75464	0,75515

Tabla 8.32: Errores cuadráticos medio de predicción $ECMP_{1,0}$ y $ECMP_{2,0}$ basado en datos no contaminados para $h_1 = 0,15$ y $h_2 = 0,15$.

		LS	GM	LMS	LTS	MH	MT
ECMP1	Media	0,00072	0,00079	0,00105	0,00106	0,00078	0,00078
	Mediana	0,00049	0,00055	0,00068	0,00065	0,00053	0,00053
ECMP2	Media	0,0013	0,00142	0,00182	0,00181	0,00141	0,00141
	Mediana	0,00059	0,00065	0,00078	0,00074	0,00063	0,00064
ECMP1	Media	0,84564	388,74368	389,15537	389,41860	0,84782	0,85075
	Mediana	0,83798	345,51315	346,28595	345,82345	0,83975	0,84161
ECMP2	Media	1,04373	481,31043	481,78835	482,14876	1,04797	1,05231
	Mediana	0,88098	369,05868	364,89299	363,85893	0,89383	0,89693

Tabla 8.33: Errores cuadráticos medio de predicción $ECMP_{1,1}$ y $ECMP_{2,1}$ basado en datos contaminados para $h_1 = 0,15$ y $h_2 = 0,15$.

Basados en Mediana Local



Basado en el M -estimador Local

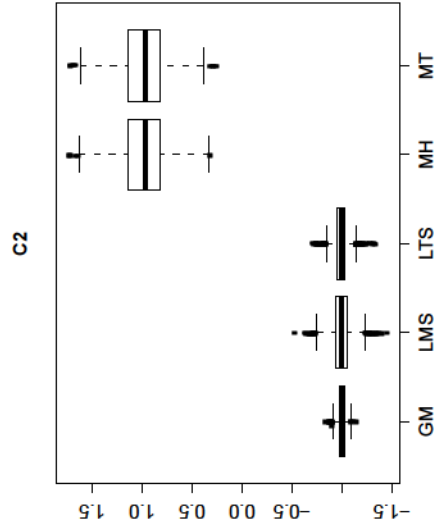
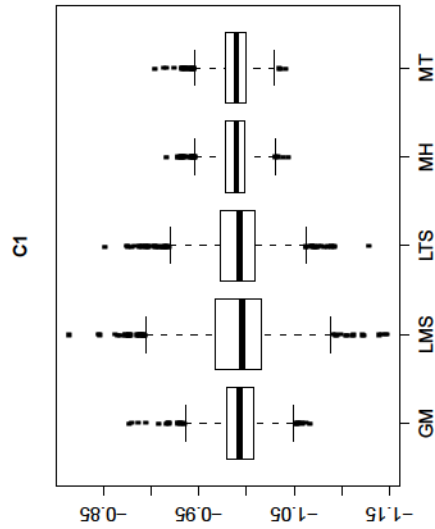
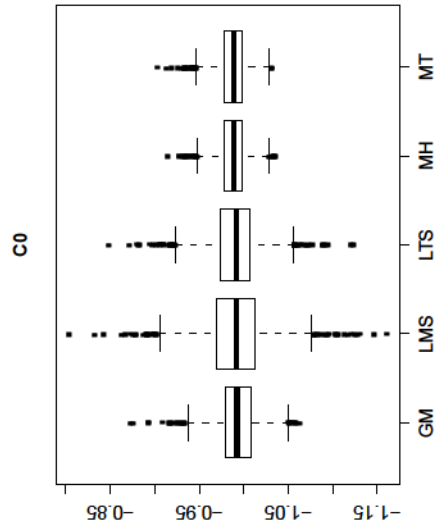
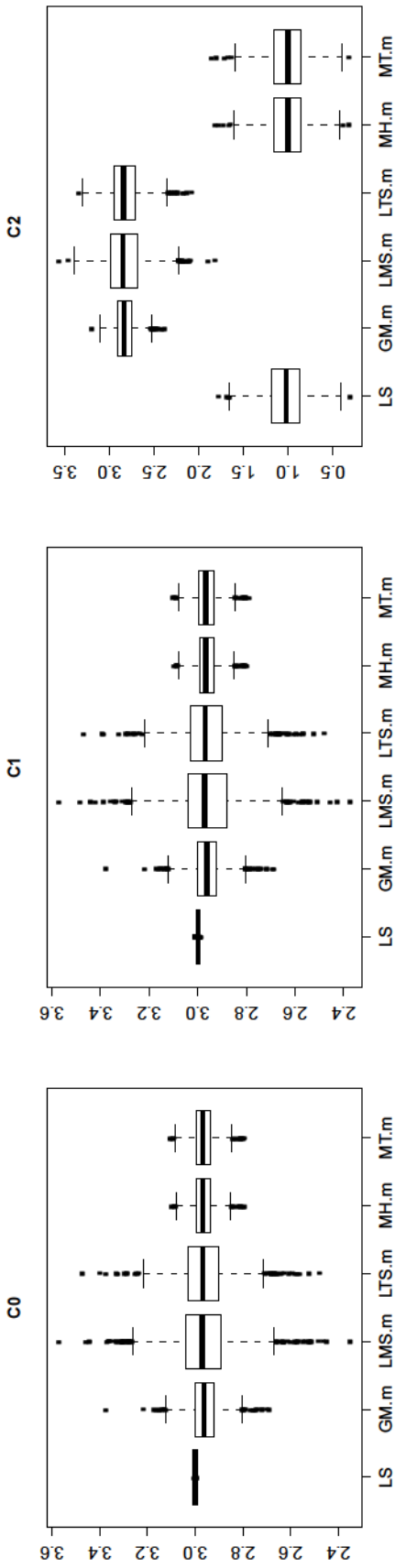


Figura 8.2: Boxplot de los estimadores de β_1 .

Basados en Mediana Local



Basado en el M -estimador Local

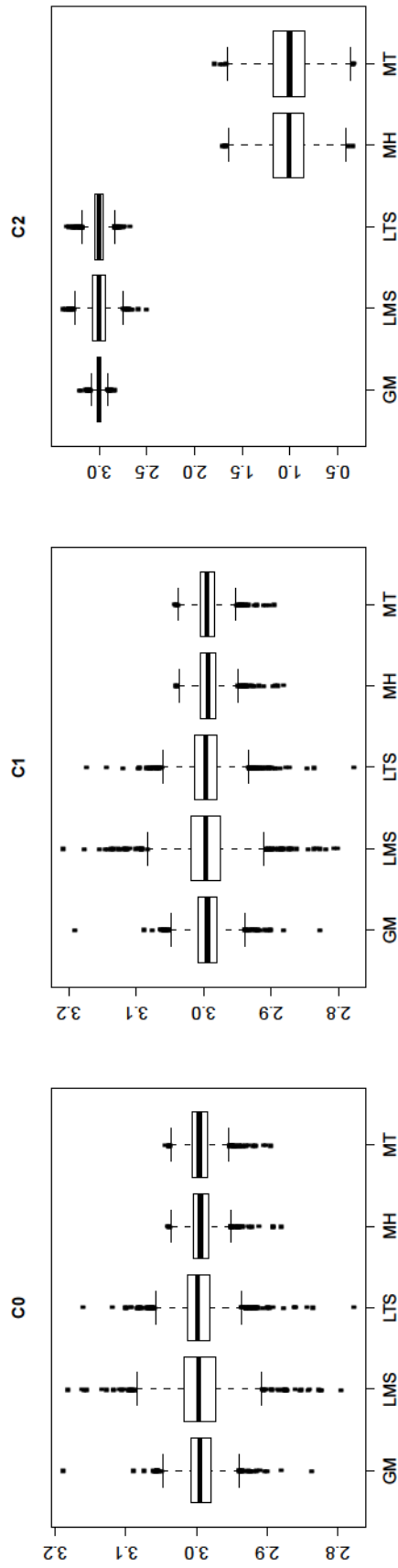


Figura 8.3: Boxplot de los estimadores de β_2 .

Capítulo 9

Apéndice

DEMOSTRACIÓN DEL LEMA 4.3.2. a) Sea $Z_i = U_i \Delta_i - \mathbb{E}(U_1 \Delta_1)$ donde $U_i = W_i^j$ para $j = 0$ ó 1 y $\Delta_i = K_i(x)/\phi(h)$. Observemos que como $|U_i| \leq 1$ y el núcleo K es acotado $|Z_i| \leq 2C_1/\phi(h) = M$, donde $C_1 = \|K\|_\infty$.

Como $\sigma^2 = \text{VAR}(Z_1) \leq \mathbb{E}U_1^2 \Delta_1^2$, para aplicar el Lema 4.3.1, acotaremos $\mathbb{E}U_1^2 \Delta_1^2$. Como K^2 cumple **H4** tenemos que $\mathbb{E}K_1^2(x) \leq C_2 \phi(h)$ donde $C_2 = C'' \|K^2\|_\infty = C' C_1^2$ y C' está dada en **H1**, luego

$$\mathbb{E}U_1^2 \Delta_1^2 \leq \mathbb{E}\Delta_1^2 \leq \frac{1}{\phi(h)^2} \mathbb{E}K_1^2(x) \leq C_2 \frac{1}{\phi(h)} = A.$$

Por lo tanto, como $\mathbb{E}Z_i = 0$, el Lema 4.3.1 implica que para todo $y \in \mathbb{R}$, $x \in S_{\mathcal{H}}$ y $\epsilon > 0$

$$\begin{aligned} \mathbb{P}(|R_j(x) - \mathbb{E}R_j(x)| > \epsilon) &= \mathbb{P}\left(\left|\sum_{i=1}^n Z_i\right| > \epsilon n\right) \leq 2 \exp\left\{-\frac{\epsilon^2 n}{2A\left(1 + \epsilon \frac{M}{A}\right)}\right\} \\ &\leq 2 \exp\left\{-\frac{\epsilon^2 n \phi(h)}{2C_2\left(1 + 2\epsilon/C'C_1\right)}\right\}. \end{aligned}$$

lo que concluye la demostración de a).

(b) Sea $\rho = \rho_n > 0$ una sucesión numérica tal que $\rho_n \rightarrow 0$. Consideremos un cubrimiento de $S_{\mathcal{H}}$ por bolas de radio ρ , o sea, $S_{\mathcal{H}} \subset \bigcup_{k=1}^{\ell} B(x_k, \rho)$ donde $\ell = N_\rho(S_{\mathcal{H}})$. Sea $S_n(x) = R_j(x) - \mathbb{E}R_j(x)$ donde $R_j(x) = (1/n) \sum_{i=1}^n W_i^j K_i(x)/\phi(h)$, para $j = 0$ o 1 .

Fijado $x \in S_{\mathcal{H}}$ existe k tal que $x \in B(x_k, \rho)$. Sea $\tilde{S}_n(x) = S_n(x) - S_n(x_k)$. Luego,

$$\sup_{x \in S_{\mathcal{H}}} |R_j(x) - \mathbb{E}R_j(x)| \leq \max_{1 \leq k \leq \ell} |S_n(x_k)| + \max_{1 \leq k \leq \ell} \sup_{x \in S_{\mathcal{H}} \cap B(x_k, \rho)} |\tilde{S}_n(x)|$$

Luego,

$$\mathbb{P}\left(\sup_{x \in S_{\mathcal{H}}} |R_j(x) - \mathbb{E}R_j(x)| > \epsilon\right) \leq \beta_n + \gamma_n,$$

donde $\beta_n = \mathbb{P}(\max_{1 \leq k \leq \ell} |S_n(x_k)| > \epsilon/2)$ y $\gamma_n = \mathbb{P}\left(\max_{1 \leq k \leq \ell} \sup_{x \in S_{\mathcal{J}_t} \cap B(x_k, \rho)} |\tilde{S}_n(x)| > \epsilon/2\right)$. Por (a), se tiene que si $C_3 = 2/(C' \|K\|_\infty)$ donde C' está dada en **H1** y si $C_2 = C' \|K\|_\infty^2$

$$\beta_n \leq \sum_{k=1}^{\ell} \mathbb{P}(|S_n(x_k)| > \epsilon/2) \leq \ell \sup_{x \in S_{\mathcal{J}_t}} \mathbb{P}(|S_n(x)| > \epsilon/2) \leq 2\ell \exp \left\{ -\frac{\epsilon^2 n \phi(h)}{8C_2 (1 + C_3 \frac{\epsilon}{2})} \right\},$$

Luego, la demostración de b) se deduce del hecho

$$\gamma_n \leq \sum_{k=1}^{\ell} \mathbb{P} \left(\sup_{x \in S_{\mathcal{J}_t} \cap B(x_k, \rho)} |\tilde{S}_n(x)| > \epsilon/2 \right) \leq \ell \max_{1 \leq k \leq \ell} \mathbb{P} \left(\sup_{x \in S_{\mathcal{J}_t} \cap B(x_k, \rho)} |\tilde{S}_n(x)| > \epsilon/2 \right)$$

si probamos que existen $a > 0$ y $b > 0$, tales que para todo $1 \leq k \leq \ell$

$$\mathbb{P} \left(\sup_{x \in S_{\mathcal{J}_t} \cap B(x_k, \rho)} |\tilde{S}_n(x)| > \epsilon/2 \right) \leq 6 \exp \left\{ -\frac{\epsilon^2 n \phi(h)}{b (1 + a \frac{\epsilon}{2})} \right\} \quad (9.1)$$

Consideraremos dos casos i) $K(1) = 0$, ii) $K(1) > 0$.

i) Si $K(1) = 0$ como K es Lipschitz de orden uno en $[0, 1]$ tenemos que

$$\begin{aligned} |\tilde{S}_n(x)| &= \left| \frac{1}{n} \sum_{i=1}^n \left(U_i \frac{K_i(x)}{\phi(h)} - U_i \frac{K_i(x_k)}{\phi(h)} \right) - \mathbb{E} \left[\frac{1}{n} \sum_{i=1}^n \left(U_i \frac{K_i(x)}{\phi(h)} - U_i \frac{K_i(x_k)}{\phi(h)} \right) \right] \right| \\ &\leq \frac{1}{n\phi(h)} \sum_{i=1}^n \{ |K_i(x) - K_i(x_k)| + \mathbb{E} |K_i(x) - K_i(x_k)| \} \\ &\leq \frac{2\mathbb{C}\rho}{\phi(h)h}, \end{aligned}$$

de donde,

$$\sup_{x \in S_{\mathcal{J}_t}} |\tilde{S}_n(x)| \leq 2\mathbb{C} \frac{\rho}{\phi(h)h}.$$

Si $\rho/(\phi(h)h) \rightarrow 0$, podríamos usar esta cota, pero como queremos pedir solamente que $\rho/h \rightarrow 0$ y $\rho/\phi(h) \rightarrow 0$, utilizaremos una cota mejor.

Como se observa en el Lema 8 de Ferraty *et al.* (2010) tenemos la igualdad

$$\frac{1}{n\phi(h)} \sum_{i=1}^n |K_i(x) - K_i(x_k)| = \frac{1}{n\phi(h)} \sum_{i=1}^n |K_i(x) - K_i(x_k)| \mathbb{I}_{B(x_k, h) \cup B(x, h)}(X_i)$$

con lo cual

$$\begin{aligned} \sup_{x \in S_{\mathcal{J}_t} \cap B(x_k, \rho)} |\tilde{S}_n(x)| &\leq \mathbb{C} \sup_{x \in S_{\mathcal{J}_t}} \frac{\rho}{h\phi(h)} \left\{ \frac{1}{n} \sum_{i=1}^n \mathbb{I}_{B(x_k, h) \cup B(x, h)}(X_i) + \frac{1}{n} \sum_{i=1}^n \mathbb{E} (\mathbb{I}_{B(x_k, h) \cup B(x, h)}(X_i)) \right\} \\ &\leq \mathbb{C} \frac{\rho}{h\phi(h)} \left\{ \frac{1}{n} \sum_{i=1}^n \mathbb{I}_{B(x_k, h+\rho)}(X_i) + \mathbb{E} (\mathbb{I}_{B(x_k, h+\rho)}(X_1)) \right\} \end{aligned}$$

Observemos primero que como $\rho/\phi(h) \rightarrow 0$ y se cumple **H5a**), entonces $\phi(h + \rho) \leq \phi(h) + C\rho \leq 2\phi(h)$.

Sea

$$Z_i = \mathbb{I}_{B(x_k, h+\rho)}(X_i) \frac{\rho}{h\phi(h)},$$

luego

$$Z_i \leq \frac{\rho}{h\phi(h)} \quad \mathbb{E}Z_i \leq \phi(h + \rho) \frac{\rho}{h\phi(h)} \leq 2\frac{\rho}{h} \rightarrow 0 \quad (9.2)$$

$$\text{VAR}(Z_i) \leq \mathbb{E}Z_i^2 \leq \phi(h + \rho) \left(\frac{\rho}{h\phi(h)} \right)^2 \leq 2\frac{\rho^2}{h^2\phi(h)}. \quad (9.3)$$

Con lo cual, (9.2) implica que

$$\frac{\rho}{h\phi(h)} \left\{ \frac{1}{n} \sum_{i=1}^n \mathbb{I}_{B(x_k, h+\rho)}(X_i) + \mathbb{E}(\mathbb{I}_{B(x_k, h+\rho)}(X_1)) \right\} \leq \frac{1}{n} \sum_{i=1}^n |Z_i - \mathbb{E}Z_i| + 4\frac{\rho}{h}$$

y como $\rho/h \rightarrow 0$ obtenemos que

$$\mathbb{P} \left(\sup_{x \in S_{\mathcal{I}} \cap B(x_k, \rho)} |\tilde{S}_n(x)| > \epsilon/2 \right) \leq \mathbb{P} \left(\frac{1}{n} \sum_{i=1}^n |Z_i - \mathbb{E}Z_i| > \epsilon/(4\mathbb{C}) \right)$$

Por otra parte, el Lema 4.3.1 implica que para todo $\epsilon > 0$

$$\begin{aligned} \mathbb{P} \left(\frac{1}{n} \sum_{i=1}^n |Z_i - \mathbb{E}Z_i| > \epsilon/(4\mathbb{C}) \right) &\leq 2 \exp \left\{ -\frac{\epsilon^2 n}{32\mathbb{C}^2 \left(2\frac{\rho^2}{h^2\phi(h)} + \epsilon\frac{\rho}{2\mathbb{C}h\phi(h)} \right)} \right\} \\ &\leq 2 \exp \left\{ -\epsilon^2 n \phi(h) \frac{1}{32\mathbb{C}^2 \left(2\frac{\rho^2}{h^2} + \frac{\epsilon}{2\mathbb{C}} \frac{\rho}{h} \right)} \right\} \end{aligned}$$

Luego, como $\rho/h \rightarrow 0$, si $n \geq n_0$ tenemos que $\rho/h < (1/4)(1/(2\|K\|_\infty))$ y $2\rho^2/h^2 < 1/4$ con lo cual

$$32\mathbb{C}^2 \left(2\frac{\rho^2}{h^2} + \frac{\epsilon}{2\mathbb{C}} \frac{\rho}{h} \right) < 8 \left(1 + \epsilon \frac{1}{\|K\|_\infty} \right)$$

con lo que queda demostrado (9.1).

ii) Si $K(1) > 0$, sea

$$\tilde{S}_n(x) = \frac{1}{n} \sum_{i=1}^n \left(U_i \frac{K_i(x)}{\phi(h)} - U_i \frac{K_i(x_k)}{\phi(h)} \right).$$

Luego, $|\tilde{S}_n(x)| \leq |\tilde{\tilde{S}}_n(x)| + \mathbb{E}|\tilde{\tilde{S}}_n(x)|$. Como en el Lema 8 de Ferraty *et al.* (2010) obtenemos que si $x \in S_{\mathcal{J}} \cap B(x_k, \rho)$ entonces

$$\begin{aligned}
|\tilde{\tilde{S}}_n(x)| &\leq \frac{1}{n\phi(h)} \left\{ \sum_{i=1}^n |K_i(x) - K_i(x_k)| \mathbb{I}_{B(x_k, h) \cap B(x, h)}(X_i) + \sum_{i=1}^n |K_i(x)| \mathbb{I}_{B(x_k, h)^c \cap B(x, h)}(X_i) \right. \\
&\quad \left. + \sum_{i=1}^n |K_i(x_k)| \mathbb{I}_{B(x_k, h) \cap B(x, h)^c}(X_i) \right\} \\
&\leq \mathbb{C} \frac{\rho}{h\phi(h)} \frac{1}{n} \sum_{i=1}^n \mathbb{I}_{B(x_k, h) \cap B(x, h)}(X_i) + \|K\|_\infty \frac{1}{\phi(h)} \frac{1}{n} \sum_{i=1}^n \mathbb{I}_{B(x_k, h)^c \cap B(x, h)}(X_i) \\
&\quad + \|K\|_\infty \frac{1}{\phi(h)} \frac{1}{n} \sum_{i=1}^n \mathbb{I}_{B(x, h)^c \cap B(x_k, h)}(X_i) \\
&\leq \mathbb{C} \frac{\rho}{h\phi(h)} \frac{1}{n} \sum_{i=1}^n \mathbb{I}_{B(x_k, h)}(X_i) + \|K\|_\infty \frac{1}{\phi(h)} \frac{1}{n} \sum_{i=1}^n \mathbb{I}_{B(x_k, h)^c \cap B(x_k, h+\rho)}(X_i) \\
&\quad + \|K\|_\infty \frac{1}{\phi(h)} \frac{1}{n} \sum_{i=1}^n \mathbb{I}_{B(x_k, h-\rho)^c \cap B(x_k, h)}(X_i) \leq \mathbb{C} \frac{1}{n} \sum_{i=1}^n Z_i + A_{n,1} + A_{n,2}
\end{aligned}$$

donde Z_i está definido en i). Observemos que podemos escribir a $A_{n,j}$ como $A_{n,j} = (1/n) \sum_{i=1}^n U_{i,j}$, donde

$$U_{i,1} = \|K\|_\infty \frac{2}{\phi(h)} \mathbb{I}_{B(x_k, h)^c \cap B(x_k, h+\rho)}(X_i) \quad U_{i,2} = \|K\|_\infty \frac{2}{\phi(h)} \mathbb{I}_{B(x_k, h-\rho)^c \cap B(x_k, h)}(X_i)$$

con lo cual $U_{i,j} \leq 2\|K\|_\infty/\phi(h)$

$$\mathbb{E}U_{i,1} \leq \|K\|_\infty \frac{2}{\phi(h)} [\phi(h+\rho) - \phi(h)] \leq C \frac{\rho}{\phi(h)} \rightarrow 0 \quad (9.4)$$

$$\mathbb{E}U_{i,2} \leq \|K\|_\infty \frac{2}{\phi(h)} [\phi(h) - \phi(h-\rho)] \leq C \frac{\rho}{\phi(h)} \rightarrow 0 \quad (9.5)$$

$$\text{VAR}(U_{i,j}) \leq \mathbb{E}U_{i,j}^2 \leq C \frac{\rho}{\phi(h)^2}. \quad (9.6)$$

Luego, obtenemos que $\sup_{x \in S_{\mathcal{J}} \cap B(x_k, \rho)} |\tilde{\tilde{S}}_n(x)| \leq (\mathbb{C}/n) \sum_{i=1}^n Z_i + A_{n,1} + A_{n,2}$. Como además $\mathbb{E}Z_1 \leq 2\rho/h$

$$\sup_{x \in S_{\mathcal{J}} \cap B(x_k, \rho)} \mathbb{E}|\tilde{\tilde{S}}_n(x)| \leq \mathbb{E}Z_1 + \mathbb{E}U_{1,1} + \mathbb{E}U_{1,2} \leq 2\frac{\rho}{h} + 2C \frac{\rho}{\phi(h)} \rightarrow 0$$

Tenemos que

$$\mathbb{P} \left(\sup_{x \in S_{\mathcal{J}} \cap B(x_k, \rho)} |\tilde{\tilde{S}}_n(x)| > \frac{\epsilon}{2} \right) \leq \mathbb{P} \left(\sup_{x \in S_{\mathcal{J}} \cap B(x_k, \rho)} |\tilde{\tilde{S}}_n(x)| > \frac{\epsilon}{4} \right) \leq \mathbb{P} \left(\mathbb{C} \frac{1}{n} \sum_{i=1}^n Z_i + A_{n,1} + A_{n,2} > \frac{\epsilon}{4} \right)$$

Por i) tenemos que

$$\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^n Z_i > \frac{\epsilon}{8\mathbb{C}}\right) \leq 2 \exp\left\{-\epsilon^2 n \phi(h) \frac{1}{128\left(1 + \frac{\epsilon}{16\|K\|_\infty}\right)}\right\}$$

Luego, debemos mostrar que tenemos una desigualdad análoga para $\mathbb{P}(A_{n,j} > \epsilon/16)$, $j = 1, 2$. Como $\mathbb{E}U_{1,j} \leq C\rho/\phi(h)$, basta acotar $\mathbb{P}((1/n)\sum_{i=1}^n |U_{i,j} - \mathbb{E}U_{i,j}| > \epsilon/32)$ para $j = 1, 2$. Usando el lema 4.3.1 si $C_1 = 4\|K\|_\infty^2 C$ y $C_2 = 4\|K\|_\infty$, obtenemos que

$$\begin{aligned} \mathbb{P}\left(\frac{1}{n}\sum_{i=1}^n |U_{i,j} - \mathbb{E}U_{i,j}| > \frac{\epsilon}{32}\right) &\leq 2 \exp\left\{-\frac{\epsilon^2 n}{2048\left(C_1 \frac{\rho}{\phi(h)^2} + \epsilon \frac{C_2}{\phi(h)}\right)}\right\} \\ &\leq 2 \exp\left\{-\epsilon^2 n \phi(h) \frac{1}{2048\left(C_1 \frac{\rho}{\phi(h)} + \epsilon C_2\right)}\right\} \end{aligned}$$

Luego, como $\rho/\phi(h) \rightarrow 0$, si $n \geq n_0$ tenemos que $C_1\rho/\phi(h) < 1$ con lo cual

$$2048\left(C_1 \frac{\rho}{\phi(h)} + \epsilon C_2\right) < 2048(1 + \epsilon C_2)$$

con lo que queda demostrado (9.1), concluyendo la demostración de b).

(c) Como en b), consideremos un cubrimiento de $S_{\mathcal{H}}$ por bolas de radio $\rho = \log(n)/n$, o sea, $S_{\mathcal{H}} \subset \bigcup_{k=1}^{\ell} B(x_k, \rho)$ donde $\ell = N_\rho(S_{\mathcal{H}}) = \exp\{\psi_{S_{\mathcal{H}}}(\log(n)/n)\}$. Sea $S_n(x) = R_j(x) - \mathbb{E}R_j(x)$ donde $R_j(x) = (1/n)\sum_{i=1}^n W_i^j K_i(x)/\phi(h)$, para $j = 0$ o 1 .

Fijado $x \in S_{\mathcal{H}}$ existe k tal que $x \in B(x_k, \rho)$. Sea $\tilde{S}_n(x) = S_n(x) - S_n(x_k)$. Como en b) $\mathbb{P}(\sup_{x \in S_{\mathcal{H}}} |R_j(x) - \mathbb{E}R_j(x)| > \epsilon_0 \theta_n) \leq \beta_n + \gamma_n$, donde $\beta_n = \mathbb{P}(\max_{1 \leq k \leq \ell} |S_n(x_k)| > \epsilon_0 \theta_n/2)$ y $\gamma_n = \mathbb{P}(\max_{1 \leq k \leq \ell} \sup_{x \in S_{\mathcal{H}} \cap B(x_k, \rho)} |\tilde{S}_n(x)| > \epsilon_0 \theta_n/2)$ y $\theta_n^2 = \psi_{S_{\mathcal{H}}}(\log(n)/n)/(n\phi(h))$.

Por (a), se tiene que existe $a = 8C'\|K\|_\infty^2 > 0$ y $b > 0$ independientes de n tales que

$$\begin{aligned} \beta_n &\leq \sum_{k=1}^{\ell} \mathbb{P}(|S_n(x_k)| > \epsilon_0 \theta_n/2) \leq \ell \sup_{x \in S_{\mathcal{H}}} \mathbb{P}(|S_n(x)| > \epsilon_0 \theta_n/2) \\ &\leq 2\ell \exp\left\{-\frac{\epsilon_0^2 \theta_n^2 n \phi(h)}{a(1 + b \frac{\epsilon_0 \theta_n}{2})}\right\} \leq 2\ell \exp\left\{-\frac{\epsilon_0^2 \psi_{S_{\mathcal{H}}}\left(\frac{\log(n)}{n}\right)}{a(1 + b \frac{\epsilon_0 \theta_n}{2})}\right\} \\ &\leq 2\ell \exp\left\{-\frac{\epsilon_0^2 \psi_{S_{\mathcal{H}}}\left(\frac{\log(n)}{n}\right)}{a(1 + \epsilon_0)}\right\}, \end{aligned}$$

para $n \geq n_0$ pues $\theta_n \rightarrow 0$.

Luego, la demostración de c) se deduce del hecho

$$\gamma_n \leq \sum_{k=1}^{\ell} \mathbb{P} \left(\sup_{x \in S_{\mathcal{J}_c} \cap B(x_k, \rho)} |\tilde{S}_n(x)| > \epsilon_0 \theta_n / 2 \right) \leq \ell \max_{1 \leq k \leq \ell} \mathbb{P} \left(\sup_{x \in S_{\mathcal{J}_c} \cap B(x_k, \rho)} |\tilde{S}_n(x)| > \epsilon_0 \theta_n / 2 \right)$$

si probamos que para todo $1 \leq k \leq \ell$, existe $c > 0$ tal que si $\epsilon_0 > c$ y $n \geq n_0$

$$\mathbb{P} \left(\sup_{x \in S_{\mathcal{J}_c} \cap B(x_k, \rho)} |\tilde{S}_n(x)| > \epsilon_0 \theta_n / 2 \right) \leq 6 \exp \left\{ -\frac{\epsilon_0^2 \psi_{S_{\mathcal{J}_c}} \left(\frac{\log(n)}{n} \right)}{a(1 + \epsilon_0)} \right\} \quad (9.7)$$

Consideraremos dos casos i) $K(1) = 0$, ii) $K(1) > 0$.

i) Si $K(1) = 0$ como en b)

$$\begin{aligned} \sup_{x \in S_{\mathcal{J}_c} \cap B(x_k, \rho)} |\tilde{S}_n(x)| &\leq \mathbb{C} \sup_{x \in S_{\mathcal{J}_c}} \frac{\rho}{h\phi(h)} \left\{ \frac{1}{n} \sum_{i=1}^n \mathbb{I}_{B(x_k, h) \cup B(x, h)}(X_i) + \mathbb{E} \left(\mathbb{I}_{B(x_k, h+\rho)}(X_1) \right) \right\} \\ &\leq \mathbb{C} \frac{\rho}{h\phi(h)} \frac{1}{n} \sum_{i=1}^n \mathbb{I}_{B(x_k, h+\rho)}(X_i) + \mathbb{C} \frac{\rho}{h\phi(h)} \mathbb{E} \left(\mathbb{I}_{B(x_k, h+\rho)}(X_1) \right) \end{aligned}$$

Como en b) sea

$$Z_i = \mathbb{I}_{B(x_k, h+\rho)}(X_i) \frac{\rho}{h\phi(h)}.$$

Luego, por (9.2) y como $\psi_{S_{\mathcal{J}_c}}(\log(n)/n) \geq (\log(n))^2 / (n\phi(h))$ obtenemos usando **H5a**) que

$$\theta_n^{-1} \mathbb{E} Z_i \leq 2\theta_n^{-1} \frac{\rho}{h} = 2 \sqrt{\frac{n\phi(h)}{\psi_{S_{\mathcal{J}_c}} \left(\frac{\log(n)}{n} \right)}} \frac{\log(n)}{nh} \leq 2 \frac{\phi(h)}{h} \leq C_\phi.$$

Con lo cual, usando que

$$\sup_{x \in S_{\mathcal{J}_c} \cap B(x_k, \rho)} |\tilde{S}_n(x)| \leq \mathbb{C} \left(\frac{1}{n} \sum_{i=1}^n Z_i + \mathbb{E} Z_1 \right) \leq \mathbb{C} \left(\frac{1}{n} \sum_{i=1}^n |Z_i - \mathbb{E} Z_i| + 2\mathbb{E} Z_1 \right)$$

obtenemos que si $\epsilon_0 > 8C_\phi$

$$\mathbb{P} \left(\sup_{x \in S_{\mathcal{J}_c} \cap B(x_k, \rho)} |\tilde{S}_n(x)| > \epsilon_0 \theta_n / 2 \right) \leq \mathbb{P} \left(\frac{1}{n} \sum_{i=1}^n |Z_i - \mathbb{E} Z_i| > \epsilon_0 \theta_n / (4\mathbb{C}) \right)$$

Por otra parte, el Lema 4.3.1 implica que

$$\begin{aligned}
\mathbb{P} \left(\frac{1}{n} \sum_{i=1}^n |Z_i - \mathbb{E}Z_i| > \epsilon_0 \theta_n / (4\mathbb{C}) \right) &\leq 2 \exp \left\{ - \frac{\epsilon_0^2 \theta_n^2 n}{32\mathbb{C}^2 \left(2 \frac{\rho^2}{h^2 \phi(h)} + \frac{\epsilon_0}{4\mathbb{C}} \theta_n \frac{\rho}{h \phi(h)} \right)} \right\} \\
&\leq 2 \exp \left\{ - \epsilon_0^2 \theta_n^2 n \phi(h) \frac{1}{32\mathbb{C}^2 \left(2 \frac{\rho^2}{h^2} + \frac{\epsilon_0}{2\mathbb{C}} \theta_n \frac{\rho}{h} \right)} \right\} \\
&\leq 2 \exp \left\{ - \epsilon_0^2 \psi_{S_{\mathcal{I}}} \left(\frac{\log(n)}{n} \right) \frac{1}{32\mathbb{C}^2 \left(2 \frac{\rho^2}{h^2} + \frac{\epsilon_0}{2\mathbb{C}} \theta_n \frac{\rho}{h} \right)} \right\}
\end{aligned}$$

Luego, como $\rho/h \rightarrow 0$ y $\theta_n \rightarrow 0$ si $n \geq n_0$ tenemos que $\theta_n \rho/h < (\mathbb{C}/4)$ y $2\rho^2/h^2 < 1/4$ con lo cual

$$32\mathbb{C}^2 \left(2 \frac{\rho^2}{h^2} + \frac{\epsilon_0}{\mathbb{C}} \theta_n \frac{\rho}{h} \right) < 8(1 + \epsilon_0),$$

con lo que queda demostrado (9.7).

ii) Si $K(1) > 0$, como en b) definamos

$$\begin{aligned}
Z_i &= \mathbb{I}_{B(x_k, h+\rho)}(X_i) \frac{\rho}{h\phi(h)} \\
U_{i,1} &= \|K\|_\infty \frac{2}{\phi(h)} \mathbb{I}_{B(x_k, h)^c \cap B(x_k, h+\rho)}(X_i) \\
U_{i,2} &= \|K\|_\infty \frac{2}{\phi(h)} \mathbb{I}_{B(x_k, h-\rho)^c \cap B(x_k, h)}(X_i).
\end{aligned}$$

Luego, si $A_{n,j} = (1/n) \sum_{i=1}^n U_{i,j}$ y $\tilde{A}_n = \mathbb{C}(1/n) \sum_{i=1}^n Z_i + A_{n,1} + A_{n,2}$ tenemos que

$$\sup_{x \in S_{\mathcal{I}} \cap B(x_k, \rho)} |\tilde{S}_n(x)| \leq \tilde{A}_n + \mathbb{C}\mathbb{E}Z_1 + \mathbb{E}U_{1,1} + \mathbb{E}U_{1,2}.$$

Por (9.2), (9.4) y (9.5) y como $\psi_{S_{\mathcal{I}}}(\log(n)/n) \geq (\log(n))^2/(n\phi(h))$ tenemos que

$$\theta_n^{-1} \mathbb{E}Z_i \leq 2\theta_n^{-1} \frac{\rho}{h} = \sqrt{\frac{n\phi(h)}{\psi_{S_{\mathcal{I}}} \left(\frac{\log(n)}{n} \right)} \frac{\log(n)}{nh}} \leq \frac{\phi(h)}{h} \leq C_\phi$$

$$\theta_n^{-1} \mathbb{E}U_{i,1} \leq \|K\|_\infty \theta_n^{-1} \frac{2}{\phi(h)} [\phi(h+\rho) - \phi(h)] \leq C\theta_n^{-1} \frac{\rho}{\phi(h)} \leq C \quad (9.8)$$

$$\theta_n^{-1} \mathbb{E}U_{i,2} \leq \|K\|_\infty \theta_n^{-1} \frac{2}{\phi(h)} [\phi(h) - \phi(h-\rho)] \leq C\theta_n^{-1} \frac{\rho}{\phi(h)} \leq C, \quad (9.9)$$

por lo que si $\epsilon_0 > \max(2, 2C_\phi)$

$$\mathbb{P} \left(\sup_{x \in S_{\mathcal{Y}} \cap B(x_k, \rho)} |\tilde{S}_n(x)| > \epsilon_0 \theta_n / 2 \right) \leq \mathbb{P} \left(\tilde{A}_n > \epsilon_0 \theta_n / 4 \right).$$

Por i) tenemos que

$$\mathbb{P} \left(\frac{1}{n} \sum_{i=1}^n Z_i > \epsilon_0 \theta_n / (4\mathbb{C}) \right) \leq 2 \exp \left\{ -\frac{\epsilon_0^2}{8(1 + \epsilon_0)} \psi_{S_{\mathcal{Y}}} \left(\frac{\log(n)}{n} \right) \right\}.$$

Luego, debemos mostrar que tenemos una desigualdad análoga para $\mathbb{P}(A_{n,j} > \epsilon_0 \theta_n / 8)$, $j = 1, 2$. Por (9.8) y (9.9), si $\epsilon_0 > 16$, basta acotar para $j = 1, 2$, $\mathbb{P}((1/n) \sum_{i=1}^n |U_{i,j} - \mathbb{E}U_{i,j}| > \epsilon_0 \theta_n / 16)$. Usando que $U_{i,j} \leq 2\|K\|_\infty / \phi(h)$ y (9.6), si llamamos $C_2 = 2\|K\|_\infty / 16$ tenemos que el lema 4.3.1 implica que

$$\begin{aligned} \mathbb{P} \left(\frac{1}{n} \sum_{i=1}^n |U_{i,j} - \mathbb{E}U_{i,j}| > \frac{\epsilon_0 \theta_n}{16} \right) &\leq 2 \exp \left\{ -\frac{\epsilon_0^2 \theta_n^2 n}{512 \left(C \frac{\rho}{\phi(h)^2} + \epsilon_0 \theta_n \frac{C_2}{\phi(h)} \right)} \right\} \\ &\leq 2 \exp \left\{ -\epsilon_0^2 \theta_n^2 n \phi(h) \frac{1}{512 \left(C \frac{\rho}{\phi(h)} + \epsilon_0 \theta_n C_2 \right)} \right\} \end{aligned}$$

Luego, como $\rho / \phi(h) \rightarrow 0$, y $\theta_n \rightarrow 0$ si $n \geq n_0$ tenemos que $C_1 \rho / \phi(h) < 1/64$ y $\theta_n C_2 < 1/64$ con lo cual

$$512 \left(C \frac{\rho}{\phi(h)} + \epsilon_0 \theta_n C_2 \right) < 8(1 + \epsilon_0)$$

con lo que queda demostrado (9.7), concluyendo la demostración de c). \square

DEMOSTRACIÓN DEL LEMA 4.3.3. Observemos que

$$|\mathbb{E}\tilde{R}_1(x) - F(y|X=x)\mathbb{E}\tilde{R}_0(x)| \leq \frac{1}{n} \sum_{i=1}^n |r_i(x)|$$

donde $r_i(x) = \mathbb{E}[(F(y|X = X_i) - F(y|X = x))K_i(x)] / \mathbb{E}K_1(x) = r_1(x)$. Luego, usando (4.17), tenemos que $|\mathbb{E}\tilde{R}_1(x) - F(y|X=x)\mathbb{E}\tilde{R}_0(x)| \leq |r_1(x)|$.

a) Para probar a) basta mostrar que $\sup_{x \in S_{\mathcal{Y}}} |r_1(x)| \rightarrow 0$, pero

$$\begin{aligned} |\mathbb{E}\tilde{R}_1(x) - F(y|X=x)\mathbb{E}\tilde{R}_0(x)| &\leq |r_1(x)| \\ &\leq \frac{C_1}{\phi(h)} \int |F(y|X=u) - F(y|X=x)| K \left(\frac{d(x,u)}{h} \right) dP_X(u) \\ &\leq \frac{C_1}{\phi(h)} (A_1(x,y) + A_2(x,y)) \end{aligned}$$

donde P_X indica la medida de probabilidad de X y

$$\begin{aligned} A_1(x, y) &= \int_{S_{\mathcal{H}}^\epsilon} |F(y|X = u) - F(y|X = x)| K \left(\frac{d(x, u)}{h} \right) dP_X(u), \\ A_2(x, y) &= \int_{(S_{\mathcal{H}}^\epsilon)^c} |F(y|X = u) - F(y|X = x)| K \left(\frac{d(x, u)}{h} \right) dP_X(u) \end{aligned}$$

y $S_{\mathcal{H}}^\epsilon$ es el entorno de $S_{\mathcal{H}}$ que establece la hipótesis **H3**.

Acotemos primero $A_2(x, y)$. Como $x \in S_{\mathcal{H}}$ y $u \in (S_{\mathcal{H}}^\epsilon)^c$, tenemos que para n lo suficientemente grande $h < \epsilon$ de donde $(S_{\mathcal{H}}^\epsilon)^c \cap S_{\mathcal{H}}^h = \emptyset$. Luego, como K tiene soporte $[0, 1]$, obtenemos que $K(d(x, u)/h) = 0$ si $u \in (S_{\mathcal{H}}^\epsilon)^c$ de donde usando que

$$\sup_{u \in \mathbb{R}} \sup_{x \in S_{\mathcal{H}}} A_2(x, y) \leq 2 \sup_{x \in S_{\mathcal{H}}} \int_{(S_{\mathcal{H}}^\epsilon)^c} K \left(\frac{d(x, u)}{h} \right) dP_X(u)$$

obtenemos que si $n \geq n_0$, para todo $y \in \mathbb{R}$ y $x \in S_{\mathcal{H}}$

$$A_2(x, y) = 0. \quad (9.10)$$

Por otra parte, como $F(y|X = u)$ es uniformemente continua en $S_{\mathcal{H}}^\epsilon$ tenemos que dado $\eta > 0$ existe $\delta > 0$ tal que para todo $x \in S_{\mathcal{H}}$, $d(x, u) < \delta$ implica $|F(y|X = x) - F(y|X = u)| < \eta$. Como K tiene soporte compacto en $[0, 1]$ y es acotado

$$A_1(x, y) \leq \|K\|_\infty \int_{S_{\mathcal{H}}^\epsilon \cap B(x, h)} |F(y|X = u) - F(y|X = x)| dP_X(u)$$

Como $h_n \rightarrow 0$, existe n_0 tal que si $n \geq n_0$ $h < \min(\epsilon, \delta)$. Luego, $S_{\mathcal{H}}^\epsilon \cap B(x, h) = B(x, h)$ y más aún, para todo $u \in B(x, h)$, $|F(y|X = u) - F(y|X = x)| < \eta$. Por lo tanto, usando (9.10) y el hecho que por **H1**, $\mathbb{P}(X \in B(x, h)) \leq C'\phi(h)$, tenemos que si $n \geq n_0$

$$\sup_{x \in S_{\mathcal{H}}} |\mathbb{E}\tilde{R}_1(x) - F(y|X = x)\mathbb{E}\tilde{R}_0(x)| \leq \sup_{x \in S_{\mathcal{H}}} |r_1(x)| \leq \frac{C_1}{\phi(h)} \sup_{x \in S_{\mathcal{H}}} A_1(x, y) \leq \eta$$

de donde se deduce el resultado.

b) En este caso, basta observar que cuando $x, u \in S_{\mathcal{H}}^\epsilon$, por **H9** $|F(y|X = u) - F(y|X = x)| \leq Dd^n(x, u)$. Sea $Z(x)$ la variable aleatoria $Z(x) = d(x, X)/h$, y $P_{Z(x)}$ su función de probabilidad. Luego, usando **H1** y que K es acotado y con soporte $[0, 1]$ obtenemos que

$$\begin{aligned} A_1(x, y) &= D \int_{S_{\mathcal{H}}^\epsilon} d^n(x, u) K \left(\frac{d(x, u)}{h} \right) dP_X(u) \leq Dh^n \int v^n K(v) dP_{Z(x)}(v) \\ &\leq Dh^n \int_0^1 K(v) dP_{Z(x)}(v) \leq D_1 h^n \phi(h). \end{aligned}$$

Luego, obtenemos que para todo $y \in \mathbb{R}$ y $x \in S_{\mathcal{H}}$

$$\frac{C_1}{\phi(h)} A_1(x, y) \leq D_1 h^m \phi(h). \quad (9.11)$$

Luego, a partir de (9.11) y (9.10) obtenemos el resultado deseado. \square

DEMOSTRACIÓN DEL LEMA 4.3.5. La demostración del Lema 4.3.5 es análoga a la del Lema 3.1 de Boente y Fraiman (1991).

Por la hipótesis **H3** existen a, b tales que $F(b|X = x) > \frac{7}{8}$ y $F(a|X = x) < \frac{1}{8}$ para todo $x \in S_{\mathcal{H}}$. Sea $m_n(x)$ es la mediana de $F_n(\cdot|X = x)$, como (4.21) implica que existe $n_0 \in \mathbb{N}$ tal que $F_n(a|X = x) < \frac{1}{4}$ y $F_n(b|X = x) > \frac{3}{4}$ para todo $n \geq n_0$ y $x \in S_{\mathcal{H}}$, se tiene que $a < m_n(x) < b$ para todo $x \in S_{\mathcal{H}}$. Es fácil ver que la elección de a, b y n implican además que $s_n(x) < b - a$ para todo $n \geq n_0$ y $x \in S_{\mathcal{H}}$. Como $F(y|X = x)$ es una función de distribución continua para cada x fijo por **H3** ii), dado $x_0 \in S_{\mathcal{H}}$ existen $a(x_0)$ y $b(x_0)$ tales que

$$F(a(x_0)|X = x_0) = \frac{1}{3}, \quad F(b(x_0)|X = x_0) = \frac{7}{10}.$$

Por la hipótesis **H3ii)**, dado $\epsilon < \frac{1}{60}$ existe $\delta_0 = \delta_0(x_0) > 0$ tal que si $d(x, x_0) \leq \delta_0$ entonces

$$\frac{1}{3} - \epsilon < F(a(x_0)|X = x) < \frac{1}{3} + \epsilon \quad \frac{7}{10} - \epsilon < F(b(x_0)|X = x) < \frac{7}{10} + \epsilon.$$

Denotemos por $\mathcal{V}(x_0, \delta_0) = \{x \in \mathcal{H} : d(x, x_0) < \delta_0\}$. Como $S_{\mathcal{H}}$ es un conjunto compacto, existen $x_1, \dots, x_\ell \in S_{\mathcal{H}}$ tales que $S_{\mathcal{H}} \subset \bigcup_{j=1}^{\ell} \mathcal{V}(x_j, \delta_j)$. Sean $a_j = a(x_j)$ y $b_j = b(x_j)$. Entonces

$$\frac{1}{3} - \epsilon < F(a_j|X = x) < \frac{1}{3} + \epsilon, \quad \frac{7}{10} - \epsilon < F(b_j|X = x) < \frac{7}{10} + \epsilon,$$

para todo $x \in \mathcal{V}(x_j, \delta_j) \cap S_{\mathcal{H}}$, $1 \leq j \leq \ell$.

Por la hipótesis **H3** ii), existe $\eta > 0$ tal que

$$\frac{1}{3} - 2\epsilon < F(a_j - \eta|X = x) < \frac{1}{3} + 2\epsilon, \quad \frac{7}{10} - 2\epsilon < F(b_j + \eta|X = x) < \frac{7}{10} + 2\epsilon,$$

para todo $x \in \mathcal{V}(x_j, \delta_j) \cap S_{\mathcal{H}}$ y $1 \leq j \leq \ell$. Finalmente, (4.21) implica que para todo $n \geq n_0$ se tiene que $F_n(a_j|X = x) < 1/2$, $F_n(b_j|X = x) > 1/2$, $F_n(a_j - \eta|X = x) > 1/4$ y $F_n(b_j + \eta|X = x) < 3/4$, para todo $x \in B(x_j, \delta_j) \cap S_{\mathcal{H}}$, $1 \leq j \leq \ell$. Por lo tanto, $a_j \leq m_n(x) \leq b_j$ para $x \in B(x_j, \delta_j) \cap S_{\mathcal{H}}$, $1 \leq j \leq \ell$, y $s_n(x) > \eta$ para todo $x \in S_{\mathcal{H}}$ si $n \geq n_0$. \square

DEMOSTRACIÓN DEL LEMA 7.1.3. a) Como \mathcal{H} es separable y completo, para todo $\epsilon > 0$, existen conjuntos compactos $\mathcal{K}_1 \subset \mathbb{R}^{p+1}$ y $S_{\mathcal{H}} \subset \mathcal{H}$ tales que, si $\mathcal{K} = \mathcal{K}_1 \times S_{\mathcal{H}}$, $\mathbb{P}(\mathcal{K}) >$

$1 - (\epsilon/8\|f\|_\infty)$. Observemos que $|\mathbb{E}_{Q_n}(f) - \mathbb{E}_{P_n}(f)| \leq A_{1n} + A_{2n}$, donde

$$\begin{aligned} A_{1n} &= \frac{1}{n} \sum_{i=1}^n |f(r_i + \widehat{\eta}_0(x_i), \mathbf{u}_i + \widehat{\boldsymbol{\eta}}(x_i)) - f(r_i, \mathbf{u}_i)| \mathbb{I}_{\mathcal{K}}(r_i, \mathbf{u}_i, x_i), \\ A_{2n} &= 2\|f\|_\infty \frac{1}{n} \sum_{i=1}^n \mathbb{I}_{\mathcal{K}^c}(r_i, \mathbf{u}_i, x_i). \end{aligned}$$

Por (7.1) y la Ley Fuerte de los Grandes Números, tenemos que existe un conjunto $\mathcal{N} \subset \Omega$ tal que $\mathbb{P}(\mathcal{N}) = 0$ y tal que para cualquier $\omega \notin \mathcal{N}$

$$\sup_{x \in S_{\mathcal{K}}} |\widehat{\eta}_0(x)| + \sup_{x \in S_{\mathcal{K}}} \|\widehat{\boldsymbol{\eta}}(x)\| \rightarrow 0 \quad \text{y} \quad \frac{1}{n} \sum_{i=1}^n \mathbb{I}_{\mathcal{K}^c}(r_i, \mathbf{u}_i, x_i) \rightarrow \mathbb{P}(\mathcal{K}^c). \quad (9.12)$$

Por lo tanto, para n suficientemente grande $A_{2n} \leq \epsilon/2$ para $\omega \notin \mathcal{N}$.

Denotemos por \mathcal{C}_1 la clausura de un entorno de radio 1 de \mathcal{K}_1 . La continuidad uniforme de f sobre \mathcal{C}_1 implica que existe δ tal que si $\max_{1 \leq j \leq p+1} |u_j - v_j| < \delta$, $\mathbf{u}, \mathbf{v} \in \mathcal{C}_1$ entonces $|f(\mathbf{u}) - f(\mathbf{v})| < \epsilon/2$. Luego, por (9.12) tenemos que para $\omega \notin \mathcal{N}$ y n lo suficientemente grande, $\max_{0 \leq j \leq p} \sup_{x \in S_{\mathcal{K}}} |\widehat{\eta}_j(x)| < \delta$ y por lo tanto, para $1 \leq i \leq n$,

$$|f(r_i + \widehat{\eta}_0(x_i), \mathbf{u}_i + \widehat{\boldsymbol{\eta}}(x_i)) - f(r_i, \mathbf{u}_i)| \mathbb{I}_{\mathcal{K}}(r_i, \mathbf{u}_i, x_i) < \frac{\epsilon}{2},$$

lo que implica $A_{1n} < \epsilon/2$. Por lo tanto, $|\mathbb{E}_{Q_n}(f) - \mathbb{E}_{P_n}(f)| < \epsilon$ para n lo suficientemente grande y $\omega \notin \mathcal{N}$.

b) Se deduce inmediatamente de a). \square

DEMOSTRACIÓN DEL LEMA 7.2.1. Para cualquier matriz $\mathbf{B} \in \mathbb{R}^{p \times p}$, sea $|\mathbf{B}| = \max_{1 \leq l, j \leq p} |b_{lj}|$. Notemos con ξ_i a puntos intermedios entre $r_i - \mathbf{u}_i^T \widetilde{\boldsymbol{\beta}}$ y $\widehat{r}_i - \widehat{\mathbf{u}}_i^T \widetilde{\boldsymbol{\beta}}$ y $\widehat{\eta}_j(x) = \widehat{\phi}_j(x) - \phi_j(x)$ para $0 \leq j \leq p$, y $\widehat{\boldsymbol{\eta}} = (\widehat{\eta}_1(x), \dots, \widehat{\eta}_p(x))^T$. Usando un desarrollo de Taylor de orden 1, cálculos algebraicos permiten obtener que $\mathbf{A} = \mathbf{A}_n^{(1)} + \mathbf{A}_n^{(2)} + \mathbf{A}_n^{(3)} + \mathbf{A}_n^{(4)}$, donde

$$\begin{aligned} \mathbf{A}_n^{(1)} &= \frac{1}{n} \sum_{i=1}^n \psi'_1 \left(\frac{r_i - \mathbf{u}_i^T \widetilde{\boldsymbol{\beta}}}{s_n} \right) w_2(\|\mathbf{u}_i\|) \mathbf{u}_i \mathbf{u}_i^T, \\ \mathbf{A}_n^{(2)} &= -\frac{1}{n} \sum_{i=1}^n \psi'_1 \left(\frac{\widehat{r}_i - \widehat{\mathbf{u}}_i^T \widetilde{\boldsymbol{\beta}}}{s_n} \right) w_2(\|\widehat{\mathbf{u}}_i\|) [\widehat{\boldsymbol{\eta}}(x_i) \mathbf{u}_i^T + \widehat{\mathbf{u}}_i \widehat{\boldsymbol{\eta}}(X_i)^T], \\ \mathbf{A}_n^{(3)} &= -\frac{1}{n} \sum_{i=1}^n \psi''_1 \left(\frac{\xi_i}{s_n} \right) \left(\frac{\widehat{\eta}_0(X_i) - \widehat{\boldsymbol{\eta}}(X_i)^T \widetilde{\boldsymbol{\beta}}}{s_n} \right) w_2(\|\mathbf{u}_i\|) \mathbf{u}_i \mathbf{u}_i^T, \\ \mathbf{A}_n^{(4)} &= \frac{1}{n} \sum_{i=1}^n \psi'_1 \left(\frac{\widehat{r}_i - \widehat{\mathbf{u}}_i^T \widetilde{\boldsymbol{\beta}}}{s_n} \right) [w_2(\|\widehat{\mathbf{u}}_i\|) - w_2(\|\mathbf{u}_i\|)] \mathbf{u}_i \mathbf{u}_i^T. \end{aligned}$$

Argumentos análogos a los utilizados en el Lema 1 en Bianco y Boente (2002) permiten mostrar que $\mathbf{A}_n^{(1)} \xrightarrow{p} \mathbf{A}$. A partir de **N3**, se puede ver que

$$\|\mathbf{u}_i\|^2 |w_2(\|\widehat{\mathbf{u}}_i\|) - w_2(\|\mathbf{u}_i\|)| \leq \|\widehat{\eta}(X_i)\| (\|\psi_2\|_\infty + \|\widehat{\eta}(X_i)\| (\|w_2\|_\infty + \|\psi'_2\|_\infty) + \|\lambda_2\|_\infty).$$

Luego, el resultado sigue de **N2**, la consistencia de s_n y $\widetilde{\beta}$, la Ley de los Grandes Números y el hecho de que $\max_{0 \leq j \leq p} \sup_{x \in S_{\mathcal{J}_c}} |\widehat{\eta}_j(x)| \xrightarrow{p} 0$, $\mathbb{P}(X \in S_{\mathcal{J}_c}) = 1$ ya que

$$|\mathbf{A}_n^{(2)}| \leq \|\psi'_1\|_\infty \max_{0 \leq j \leq p} \sup_{x \in S_{\mathcal{J}_c}} |\widehat{\eta}_j(x)| \left(2\|\psi_2\|_\infty + \|w_2\|_\infty \max_{0 \leq j \leq p} \sup_{x \in S_{\mathcal{J}_c}} |\widehat{\eta}_j(x)| \right),$$

$$|\mathbf{A}_n^{(3)}| \leq \|\psi''_1\|_\infty \max_{0 \leq j \leq p} \sup_{x \in S_{\mathcal{J}_c}} |\widehat{\eta}_j(x)| \left(\frac{1 + p\|\widetilde{\beta}\|}{s_n} \right) \frac{1}{n} \sum_{i=1}^n w_2(\|\mathbf{u}_i\|) \|\mathbf{u}_i\|^2,$$

$$|\mathbf{A}_n^{(4)}| \leq p\|\psi'_1\|_\infty \max_{0 \leq j \leq p} \sup_{x \in S_{\mathcal{J}_c}} |\widehat{\eta}_j(x)| \times \left(\|\psi_2\|_\infty + p \max_{0 \leq j \leq p} \sup_{x \in S_{\mathcal{J}_c}} |\widehat{\eta}_j(x)| (\|w_2\|_\infty + \|\psi'_2\|_\infty) + \|\lambda_2\|_\infty \right). \square$$

Bibliografía

- [1] Agulló J., Croux C. y Van Aelst S. (2002). The multivariate least trimmed squares estimator. Research Report 0224, K.U. Leuven D.T.E.W.
- [2] Aneiros. G. y Vieu, Ph. (2006). Semi-functional partial linear model. *Statistics & Probability Letters*, **11**, 1102-1110.
- [3] Aneiros. G. y Vieu, Ph. (2009). Automatic estimation procedure in partial linear model with functional data. *Stat Papers*, Springer-Verlag.
- [4] Attouch, M. Laksaci, A. y Ould-Said, E. (2007). Asymptotic distribution of robust estimators for functional nonparametric models. Unpublished Manuscript.
- [5] Azzedine, N., Laksaci, A. y Ould-Said, E. (2008). On robust nonparametric regression estimation for a functional regressor. *Statistics & Probability Letters*, **78**, 3216-3221.
- [6] Bianco, A. y Boente, G. (2002). On the asymptotic behavior of one-step estimates in heteroscedastic regression models. *Statistics & Probability Letters*, **60**, 33-47.
- [7] Bianco, A. y Boente, G. (2004). Robust estimators in semiparametric partly linear regression models. *Journal of Statistical Planning and Inference*, **122**, 229-252.
- [8] Bianco A., Boente G. y Martinez, E. (2006). Robust tests in semiparametric partly linear models. *Scandinavian Journal of Statistics*, **33**, 435-450.
- [9] Billingsley, P. (1968). Convergence of probability measures, Wiley, New York.
- [10] Boente, G. y Fraiman, R. (1989). Robust nonparametric regression estimation for dependent observations. *Annals of Statistics*, **17**, 1242-1256.
- [11] Boente, G. y Fraiman, R. (1990). Asymptotic distribution of robust estimates for nonparametric models from mixing observations. *Annals of Statistics*, **18**, 891-906.
- [12] Boente, G. y Fraiman, R. (1991). Strong Uniform Convergence Rates for Some robust equivariant nonparametric regression estimates for mixing processes. *International Statistical Institute*, **59:3**, 355-372.

- [13] Boente, G. y Fraiman, R. (1994). Local L-estimates for nonparametric regression under dependence. *Journal of Multivariate Analysis*, **4** **1**, 355-372.
- [14] Boente, G. y Fraiman, R. (1995). Asymptotic distribution of smoothers based on local means and local medians under dependence. *Journal of Multivariate Analysis*, **54**, 77-90.
- [15] Boente, G., Fraiman, R. y Meloche, J. (1997). Robust plug-in bandwidth estimators in nonparametric regression. *Journal of Statistical Planning and Inference*, **57**, 109-142.
- [16] Boente, G. y Rodriguez, D. (2008). Robust bandwidth selection in semiparametric partly linear regression models: an influential study. *Computational Statistics and Data Analysis*, **52**, 2808-2828.
- [17] Boos, D. y Serfling, R. (1980). A note on differentials and the CLT and LIL for statistical functions, with application to M -estimates. *Annals of Statistics*, **8**, 618-624.
- [18] Cai, T. y Hall, P. (2006). Prediction in functional linear regression. *Annals of Statistics*, **34**, 2159-2179.
- [19] Cadre, B. (2001). Convergent estimators for the L1-median of a Banach valued random variable. *Statistics: A Journal of Theoretical and Applied Statistics*, **35**, 509-521.
- [20] Cantoni, E. y Ronchetti, E. (2001). Resistant selection of the smoothing parameter for smoothing splines. *Statistics and Computing*, **11**, 141-146.
- [21] Chate, H. y Courbage, M. (1997). Lattice systems. *Physica D*, **103**, N°1-4, 612p, Elsevier.
- [22] Crambes, C., Delsol, L. y Laksaci, A. (2008). Robust nonparametric estimation for functional data. *International Workshop on Functional and Operatorial Statistics 2008 Proceedings, Functional and operatorial statistics*, Dabo-Niang and Ferraty (Eds.), Physica-Verlag, Springer.
- [23] Collomb, A. (1981). Estimation nonparametrique de la regression, revue bibliographique. *International Statistical Review* **49**, 73-93.
- [24] Collomb, G. (1982). Prédiction non paramétrique: Étude de l'erreur quadratique du prédictogramme. *C. R. Acad. Sci. Paris Sér. I Math.* **294** 59-62.
- [25] Collomb, G. y Härdle, W. (1986). Strong uniform convergence rates in robust nonparametric time series analysis and prediction: kernel regression estimation from dependent observations. *Stochastic Processes and Applications*, **23**, 77-89.

- [26] Dauxois, J., Pousse, A. y Romain, Y. (1982). Asymptotic theory for the principal component analysis of a vector random function: Some applications to statistical inference. *Journal of Multivariate Analysis*, **12**, 136-154.
- [27] Faden, A. (1985). The existence of regular conditional probabilities: necessary and sufficient conditions. *Annals of Probability*, **13**, 288-298.
- [28] Ferraty, F. y Vieu, P. (2004). Nonparametric models for functional data, with application in regression, time-series prediction and curve discrimination. *Journal of nonparametric statistics*, **16**, 111-125.
- [29] Ferraty, F., Mas, A. y Vieu, Ph. (2007). Nonparametric regression on functional data: inference and practical aspects. *Australian & New Zealand Journal of Statistics*, **49**, 267-286.
- [30] Ferraty, F. y Vieu, Ph. (2006). Nonparametric Functional Data Analysis: Theory and Practice. Springer Series in Statistics, Springer, New York.
- [31] Ferraty, F. y Romain, Y. (2010). The Oxford Handbook of Functional Data Analysis (Eds), Oxford University Press.
- [32] Ferraty, F., Laksaci, A. y Vieu, Ph. (2006). Estimating some characteristics of the conditional distribution in nonparametric functional models. *Statistical Inference for Stochastic Processes*, **9**, 47-76.
- [33] Ferraty, F., Laksaci, A., Tadj, A. y View, Ph. (2010). Rate of uniform consistency for nonparametric estimates with functional variables. *Journal of Statistical Planning and Inference*, **140**, 335-352.
- [34] Gao, J. (1992). A large sample Theory in Semiparametric Regression Models. Ph.D. Thesis, University of Science and Technology of China, Hefei, China.
- [35] Geenens, G. (2011). Curse of dimensionality and related issues in nonparametric functional regression. *Statistics Surveys*, **5**, 30-43.
- [36] Green, P., Jennison, C. y Seheult, A. (1985). Analysis of field experiments by least squares smoothing. *Journal of the Royal Statistical Society, Series B*, **47**, 299-315.
- [37] Härdle, W. (1984). Robust regression function estimation. *Journal of Multivariate Analysis*, **14**, 169-80.
- [38] Härdle, W., Liang, H. y Gao, J. (2000). Partially linear models. Physica-Verlag.
- [39] Härdle, W. (1990) Applied nonparametric regression. Cambridge University Press.

- [40] Härdle, W. y Tsybakov, A. (1988). Robust Nonparametric Regression with Simultaneous Scale Curve Estimation. *Annals of Statistics*, **16**, 120-135.
- [41] Henry, G. y Rodriguez, D. (2013). Robust estimators in partly linear regression models on Riemannian manifolds. *Communications in Statistics: Theory and Methods*. Aceptado.
- [42] Huber, P.J. (1964). Robust estimation of a location parameter. *Annals of Mathematical Statistics*, **35**, 73-101.
- [43] Jirina, M. (1984). On regular conditional probabilities. *Czechoslovak Mathematical Journal*. **9**, 445-451.
- [44] Kolmogorov, A. N. y Tikhomirov, V. M. (1959). ϵ -entropy and ϵ -capacity of sets in function spaces. *Uspekhi Mat. Nauk.*, **14**, 3-86. (In Russian.) [Published in English in (1961) Amer. Math. Soc. Transl. (2) 17 277-364.]
- [45] Leao, D., Fragoso, M. y Ruffino, P. (2004). Regular conditional probability, integration of probability and Radon spaces. *Proyecciones*, **23**, 15-29.
- [46] Leung, D. (1993). Bandwidth selection in robust smoothing. *Journal of Nonparametric Statistics*, **4**, 333-339.
- [47] Leung, D. (2005). Cross-validation in nonparametric regression with outliers. *Annals of Statistics*, **33**, 2291-2310.
- [48] Maronna, R., Martin, D. y Yohai, V. (2006). Robust statistics, theory and methods. John Wiley & Sons, Ltd.
- [49] Nadaraya, E. A. (1964) On estimating regression. *Theory of Probability and its Applications*, **9**, 141-142.
- [50] Ramsay, J. y Dalzell, C. (1991). Some tools for functional data analysis. *Journal of the Royal Statistical Society. Series B*, **53**, 539-572.
- [51] Ramsay, J. y Silverman, B. (1997). Functional Data Analysis. Springer Series in Statistics, Springer, New York.
- [52] Ramsay, J. y Silverman, B. (2002). Applied functional data analysis: methods and case studies.
- [53] Ramsay, J. y Silverman, B. (2005). Functional Data Analysis (Springer Series in Statistics).
- [54] Robinson, P. (1988). Root-n-consistent semiparametric regression. *Econometrica*, **56**, 931-954.

- [55] Rousseeuw, P.J. (1984), Least Median of Squares Regression. *Journal of the American Statistical Association*, **79**, 871-880.
- [56] Speckman, P. (1988). Kernel smoothing in partial linear models. *Journal of the Royal Statistical Society. Series B* , **50**, 413-436.
- [57] Tortrat, A. (1963). *Calcul des Probabilités* Masson, Paris.
- [58] Tsybakov, A. B. (1982). Robust estimates of a function. *Problems Inform. Transmission*, **18**, 190-201.
- [59] Van der Vaart, A. W. (1998), *Asymptotic Statistics*, Cambridge University Press.
- [60] Van der Vaart, A. W. y Wellner, J. A. (1996). *Weak Convergence and Empirical Processes: With applications to statistics*. Springer Series in Statistics. Springer-Verlag, New York.
- [61] Wang, F. y Scott, D., (1994). The L_1 Method for Robust Nonparametric Regression. *Journal of the American Statistical Association*, **89**, 65-76.
- [62] Watson, G.S. (1964). Smooth Regression Analysis. *Sankhya, The Indian Journal of Statistics. Series A*, **26**, 359-372.
- [63] Yao, F., Müller, H.G. y Wang, J.L. (2005). Functional linear regression analysis for longitudinal data. *Annals of Statistics*, **33**, 2873-2903.