Biblioteca Digital
F C E N - U B A

BIBLIOTECA CENTRAL LUIS F LELOIR
BIBLIOTECA CENTRAL LELOIR
FACULTAD DE CIENCIAS EXACTAS Y NATURALES UBA

Tesis Doctoral

Inferencia en modelos aditivos

Martínez, Alejandra Mercedes

2014-03-11

Este documento forma parte de la colección de tesis doctorales y de maestría de la Biblioteca Central Dr. Luis Federico Leloir, disponible en digital.bl.fcen.uba.ar. Su utilización debe ser acompañada por la cita bibliográfica con reconocimiento de la fuente.

This document is part of the doctoral theses collection of the Central Library Dr. Luis Federico Leloir, available in digital.bl.fcen.uba.ar. It should be used accompanied by the corresponding citation acknowledging the source.

Cita tipo APA:
Martínez, Alejandra Mercedes. (2014-03-11). Inferencia en modelos aditivos. Facultad de Ciencias Exactas y Naturales. Universidad de Buenos Aires.

Cita tipo Chicago:
Martínez, Alejandra Mercedes. "Inferencia en modelos aditivos". Facultad de Ciencias Exactas y Naturales. Universidad de Buenos Aires. 2014-03-11.

EXACTAS UBA
Facultad de Ciencias Exactas y Naturales

UBA
Universidad de Buenos Aires

Dirección: Biblioteca Central Dr. Luis F. Leloir, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires. Intendente Güiraldes 2160 - C1428EGA - Tel. (++54 +11) 4789-9293

Contacto: digital@bl.fcen.uba.ar

UNIVERSIDAD DE BUENOS AIRES

Facultad de Ciencias Exactas y Naturales

Departamento de Matemática

Inferencia en modelos aditivos

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

Alejandra Mercedes Martínez

Directores: Dra. Graciela L. Boente Boente y Dr. Matías Salibián–Barrera

Consejera: Dra. Graciela L. Boente Boente

Inferencia en Modelos Aditivos

La inferencia estadística comúnmente utiliza modelos paramétricos y el supuesto es que las observaciones de la muestra pertenecen a una familia paramétrica conocida. En este caso, el problema consiste en estimar o hacer inferencia sobre los parámetros desconocidos, permitiendo llegar a conclusiones precisas cuando el modelo supuesto es cierto pero llevando posiblemente a conclusiones equivocadas cuando se aplica a un modelo ligeramente perturbado. Por esta razón, se han desarrollado modelos noparamétricos y semiparamétricos para analizar los datos. Recientemente, los modelos noparamétricos han ganado una importante atención en el estudio de fenómenos naturales con comportamiento de complejidad no lineal. Si bien estos modelos tienen menor precisión, están asociados con una alta estabilidad. En esta tesis nos enfocaremos en los modelos de regresión noparamétricos.

Para los modelos de regresión noparamétricos multivariados, los estimadores de la función de regresión multivariada, tales como el estimador de Nadaraya–Watson, sufren de la bien conocida maldición de la dimensión, debido a que en entornos de radio fijo la cantidad de observaciones disminuye exponencialmente. Para evitar este problema, se introdujeron los modelos aditivos, que generalizan los modelos lineales, son de fácil interpretación y además resuelven la maldición de la dimensión. La mayoría de los procedimientos para estimar las componentes de un modelo aditivo se basan en promedios o polinomios locales usando ajustes por mínimos cuadrados. Por esta razón, tienen la desventaja de ser muy sensibles a la presencia de datos atípicos. Por otro lado, en muchas situaciones, sobre todo en estudios biomédicos, puede haber un conjunto de puntos del diseño con respuestas faltantes.

En esta tesis, introducimos estimadores robustos basados en polinomios locales para resolver tanto la maldición de la dimensión, como la presencia de datos atípicos y así como también la existencia de respuestas faltantes. Estos estimadores están basados en un procedimiento de integración marginal adaptado a la situación de datos faltantes. Dichas propuestas resultaron ser consistentes y asintóticamente normalmente distribuidas bajo condiciones de regularidad. Además, se consideró una familia de estimadores robustos basados en el procedimiento de backfitting cuando no hay observaciones faltantes. Finalmente, se realizó un estudio de simulación para comparar el procedimiento de las propuestas bajo diferentes escenarios.

Palabras Claves: Backfitting; Datos Faltantes; Integración Marginal; Modelos Aditivos; Pesos basados en Núcleos; Polinomios locales; Regresión noparamétrica; Robustez

Inference in Additive Models

Most commonly used models in statistical inference are parametric and the assumption is that the observations in the sample belong to a known parametric family. In this case, the problem consists in estimating or making inference on the unkown parameters, leading to accurate conclusions when the model is correct, but they can lead to wrong conclusions when they are applied to a slightly disturbed models. For this reason, nonparametric and semiparametric models have been developed for data analysis. Recently, nonparametric regression models have gain importance when studying natural phenomenons with non lineal complexity behaviour. Even though these models are less accurate, they are associated with high stability. In this thesis, we will focus on nonparametric regression models.

For the nonparamatric multivariate regression models, estimators of the multivariate regression function, such as the Nadaraya–Watson estimator, suffer from the well–known curse of dimensionality, due to in fixed-radius neiborhoods the amount of observations decreases exponentially. To overcome this drawback, additive regression models have been introduced. They generalize linear models, are easily interpretable, and they also solve the problem of the curse of the dimensionality. Most methods to estimate the components under an additive model are based on local averages or local polynomials, being sensitive to outliers. On the other hand, in many applied statistical analysis, for example in many biological experiments, it can be part of the design points on which the responses are missing.

In this thesis, we introduce robust estimators based on local polynomials to solve either the curse of dimensionality, the sensitivity to atypical observations and also the existence of missing responses. The estimators are based on marginal integration adapted to the missing situation. The proposed estimators turn out to be consistent and asymptotically normally distributed under mild assumptions. Besides, a family of robust estimators based on backfitting is also considered for the situation where no missing responses arise. A simulation study allows to compare the behaviour of our procedures, under different scenarios.

Keywords: Additive models; Backfitting; Kernel weights; Local polynomials; Nonparametric regression; Marginal integration; Missing Data; Robustness Data

Agradecimientos

No hubiese podido llegar a esta instancia sin la ayuda y cariño de mucha gente. Este agradecimiento es para todos ellos.

A mi mamá, por su infinita contención de siempre, sus buenos consejos y haberme motivado y relajado cada vez que lo necesité,

A mi papá, por haberme pensado doctora aún cuando no lo era y ahora disfrutarlo conmigo desde el cielo,

A Santi y a Agus, por haberme permitido compartir con ellos mis éxitos y fracasos,

A Dani, por su enorme paciencia y comprensión en esta última y difícil etapa,

A mi tía Chiquita, por sus consejos y ayuda, y por haber sido toda mi vida el ejemplo de matemático a seguir,

A mi abuela, simplemente por quererme y dejarme compartir con ella todas mis alegrías,

A mi directora Graciela, por haberme permitido crecer como matemático junto a ella, empapándome de su genialidad,

A mi director Matías, por confiar en mí permitiéndome trabajar junto a él, y haberme abierto las puertas de su casa dándome la posibilidad de compartir gratos momentos con su bella familia,

A mis tíos y primos, por haberse interesado y preocupado por mí,

A mis amigas y amigos: Pau, Aguis, None, Pao, Guada, Li, Pointy, David, Berli, Renée, Vero y Gaby, por haberme alentado, escuchado y dado fuerzas cuando más lo necesité,

A mis compañeras y compañeros del Instituto: Pau, Stella, Mari, Juli, Andrés, Lucas, Agustín, Maru, Lucía, Juan, Saveli, Dani, Marie, Oscar, Marta, Víctor, Ana, Diana, Susana, Willy, Flor, Vero y Vera, por su compañía, charlas y agradables encuentros en el comedor,

Y a todos aquellos que forman y formaron parte de mi vida y que agradezco muchísimo que sea así.

Indice

1 Introducción **1**

1.1 El modelo aditivo 1

1.2 ¿Por qué usar estimadores robustos? 3

1.3 Propósito de la tesis 4

1.4 Estructura de la tesis 4

2 Estimadores para los modelos aditivos con respuestas faltantes **6**

2.1 Introducción 6

2.1.1 Estimación de los efectos marginales 7

2.1.2 Estimadores cuando hay respuestas faltantes 9

2.2 $M-$estimadores basados en polinomios locales de orden 0 11

2.3 $M-$estimadores basados en polinomios locales de orden 1 en todas las coordenadas de $\mathbf{X}$ 13

2.4 $M-$estimadores de regresión usando polinomios locales de orden q 14

3 Consistencia de los $M-$estimadores **16**

3.1 Consistencia de los $M-$estimadores basados en polinomios locales de orden 0 16

3.2 Consistencia de los $M-$estimadores locales basados en polinomios locales de orden 1 19

3.2.1 Notación y Supuestos 19

3.3 Consistencia de los $M-$estimadores locales basados en polinomios locales de orden q 34

3.4 Apéndice 43

4 Distribución asintótica de los $M-$estimadores **58**

4.1 Introducción 58

4.2 Distribución asintótica de los M-estimadores basados en polinomios locales de orden 0 59

4.3 Distribución asintótica de los M-estimadores basados en polinomios locales de orden q en la dirección de interés α . . . 77

4.4 Apéndice . . . 99

4.5 Comentarios Generales . . . 102

5 Estimadores robustos basados en backfitting **104**

5.1 Introducción . . . 104

5.2 El funcional de Backfitting . . . 105

5.2.1 Fisher–consistencia y descripción del algoritmo . . . 106

5.2.2 Convergencia del algoritmo de Backfitting robusto si $\sigma(\mathbf{X}) = \sigma$ . . . 112

5.3 Estimadores de las componentes aditivas por el algoritmo de Backfitting robusto si $\sigma(\mathbf{X}) = \sigma$ . . . 115

5.4 Ejemplo . . . 117

6 Estudio Monte Carlo **120**

6.1 Condiciones de la simulación . . . 120

6.1.1 Selección preliminar de $\widehat{g}_{\mathrm{M}_0}$ . . . 123

6.2 Resultados . . . 125

Capítulo 1

Introducción

1.1 El modelo aditivo

La inferencia estadística comúnmente se focaliza sobre funciones de distribución que son puramente paramétricas o puramente noparamétricas. Un modelo paramétrico razonable produce inferencias precisas, pero un modelo paramétrico erróneo posiblemente conducirá a conclusiones equivocadas. Paralelamente, los modelos noparamétricos, si bien están asociados con alta estabilidad, tienen menor precisión. Recientemente, los modelos noparamétricos han ganado una importante atención en el estudio de fenómenos naturales con comportamiento de complejidad no lineal. El caso más general de estos modelos consiste en suponer que se tienen $(\mathbf{X}_i^{\mathrm{T}}, Y_i)^{\mathrm{T}}$ observaciones independientes idénticamente distribuidas (i.i.d) tales que $Y_i \in \mathbb{R}$, $\mathbf{X}_i \in \mathbb{R}^d$ y que se satisface

$$Y_i = g(\mathbf{X}_i) + \sigma(\mathbf{X}_i)\varepsilon_i \qquad 1 \leq i \leq n.$$

donde los errores ε_i son independientes e independientes de $\mathbf{X}_i$, y $g : \mathbb{R}^d \to \mathbb{R}$ es la función a ser estimada.

El análisis de este modelo requiere técnicas de suavizado multivariadas para la función g y, por lo tanto, encuentra en aplicaciones la dificultad conocida como la "maldición de la dimensión", que está asociada al hecho de que al aumentar la dimensión, los entornos de un punto $\mathbf{x}$ se hacen cada vez más ralos. Es decir, se necesita un número exponencialmente mayor de datos para que dichos entornos contengan observaciones de la muestra. Por ejemplo, si tenemos observaciones $\mathbf{X}_i \sim \mathcal{U}([-1,1]^d)$, $1 \leq i \leq n$, independientes, o sea, $X_{ij} \sim \mathcal{U}([-1,1])$ son variables independientes e idénticamente distribuídas, el número esperado de observaciones que yacen en $[-0.1, 0.1]^d$ es igual a $n/10^d$. Este fenómeno se traduce en la tasa de convergencia de los estimadores que no es $\sqrt{n}$ como en el caso paramétrico sino que, por ejemplo, para el caso de estimadores basados en núcleos es de orden $(nh_n^d)^{\frac{1}{2}}$ siendo h_n la ventana o parámetro de suavizado utilizado en el cómputo del estimador. En los últimos años, para resolver este problema, diversos autores han tratado el problema de reducción de la dimensión de las covariables en modelos de regresión noparamétrica. Hastie y Tibshirani (1990) introdujeron los modelos aditivos que generalizan los modelos lineales, resuelven el problema de "la maldición de la dimensión"y además son de fácil interpretación. Este nuevo planteo combina la flexibilidad de los modelos noparamétricos con la simple interpretación del modelo lineal estándar. En el mismo, se supone que $g(\mathbf{x}) = \sum_{j=1}^d g_j(x_j)$ donde las funciones

o componentes aditivas $g_j : \mathbb{R} \to \mathbb{R}$ son las incógnitas. Es decir, el modelo que consideraremos en esta tesis puede escribirse como

$$Y = g(\mathbf{X}) + \sigma(\mathbf{X})\varepsilon = \mu + \sum_{j=1}^{d} g_j(X_j) + \sigma(\mathbf{X})\varepsilon \ , \tag{1.1}$$

Para estimar las componentes g_j existen dos métodos comúnmente usados, uno basado en el procedimiento de *integración marginal* y otro basado en el método iterativo *backfitting*.

Por ejemplo, Harrison y Rubinfeld (1978) analizaron el precio de las casas en Boston. Para ello, consideraron 506 observaciones obtenidas por el censo realizado en el área metropolitana de Boston, y 13 variables explicativas para modelar los precios medianos de los hogares ocupados por sus dueños (medidos en 1.000 U$S), que llamaremos Y. Más adelante, en Härdle et. al (2004) se seleccionaron 10 de las variables explicativas, agrupadas en $\mathbf{X}$, y supusieron el siguiente modelo aditivo $\mathbb{E}(Y|\mathbf{X} = \mathbf{x}) = \mu + \sum_{j=1}^{10} g_j(\log(x_j))$. Para estimar las componentes g_j (para $j = 1, \cdots, 10$), utilizaremos el método de *backfitting* y elegiremos las ventanas para cada dimensión j proporcional a su desviación estándar ($h_j = 0.5\widehat{\sigma}_j$) como fue hecho en Härdle et. al (2004). Algunas de las estimaciones obtenidas se pueden ver en la Figura 1.1.

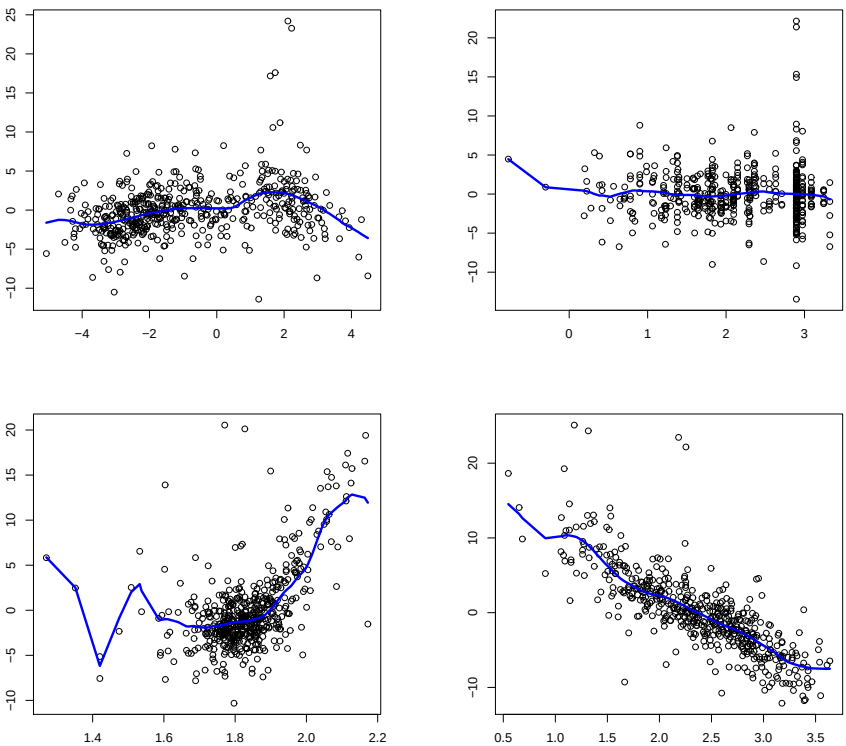

Figura 1.1: Gráficos de los ajustes obtenidos para las componentes g_1, g_2, g_4 y g_{10} (de izquierda a derecha y de arriba a abajo) mediante el procedimiento de *backfitting*, y sus respectivos residuos parciales.

Por otro lado, en muchas situaciones, sobre todo en estudios biomédicos, puede haber un conjunto de los puntos del diseño con respuestas faltantes. La inferencia básica en estos modelos

considera una muestra aleatoria $(\mathbf{X}_i^{\mathrm{T}}, Y_i, \delta_i)^{\mathrm{T}}$, $1 \leq i \leq n$ donde la covariable $\mathbf{X}_i$ es observada, mientras que la variable respuesta Y_i no es completamente observada. Aún cuando en muchas ocasiones, tanto las respuestas como la variables explicativas son faltantes, en esta tesis nos concentraremos en el caso en que los datos faltantes ocurren sólo en las respuestas. Esta situación ocurre en muchos estudios biológicos, cuando las covariables pueden ser controladas.

1.2 ¿Por qué usar estimadores robustos?

La mayoría de los procedimientos estadísticos clásicos están basados en modelos con hipótesis rígidas, tales como errores normales, observaciones equidistribuidas, etc. 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.

Un ejemplo de esto lo podemos observar en la Figura 1.2 que presenta el ajuste obtenido para el mismo conjunto de datos que el descripto en la sección anterior, pero en el que hemos modificado las respuestas de sólo un 2% de las observaciones.

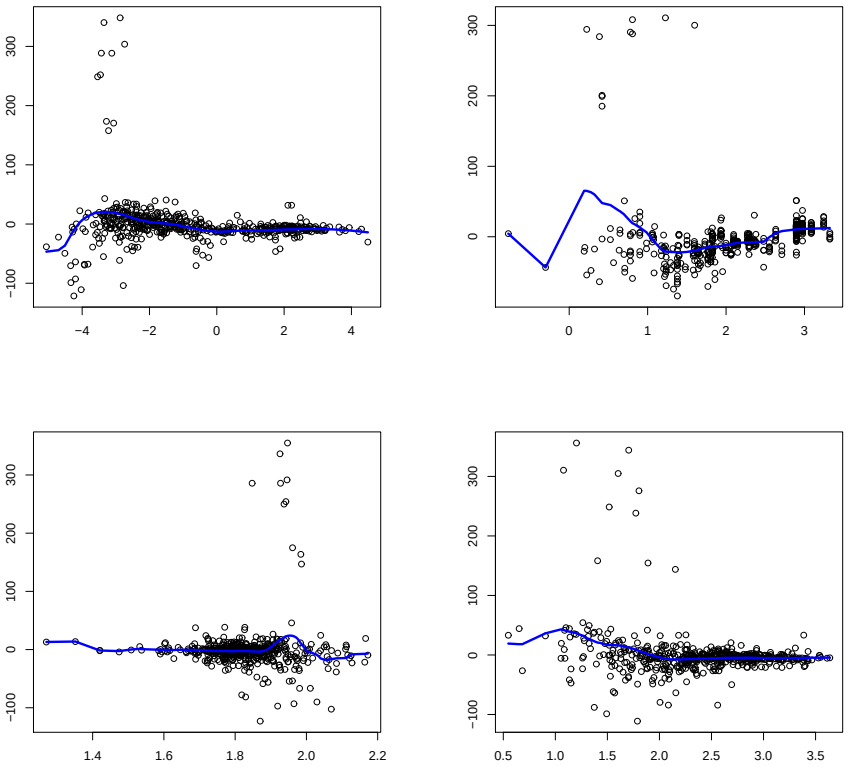

Figura 1.2: Gráficos de los ajustes obtenidos para las componentes g_1, g_2, g_4 y g_{10} (de izquierda a derecha y de arriba a abajo) mediante el procedimiento de *backfitting*, y sus respectivos residuos parciales, con datos contaminados.

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.

1.3 Propósito de la tesis

En el presente trabajo, se unificarán los aspectos antes mencionados. Consideraremos observaciones $(\mathbf{X}_i^{\mathrm{T}}, Y_i)^{\mathrm{T}}$ que satisfacen un modelo aditivo, donde algunas de las respuestas pueden no haber sido observadas, y en base a los datos disponibles estimaremos las componentes aditivas de dicho modelo. Sin embargo, con el objetivo de obtener resistencia frente a la posible presencia de *datos atípicos* en las respuestas, propondremos alternativas robustas.

Uno de los aportes de esta tesis es dar un marco de trabajo unificado al problema de estimación robusta de las componentes aditivas en un contexto en el que puede o no haber respuestas faltantes, sin necesidad de hacer supuestos relacionados con la *casi* independencia de las componentes del vector $\mathbf{X}$. Cabe aclarar que además de incluir respuestas faltantes, en esta tesis se mejoran las siguientes propuestas presentes en la literatura: a) los estimadores clásicos definidos en Severance–Lossin y Sperlich (1999) pues permitimos cambiar la pérdida cuadrática por una pérdida general; b) los estimadores considerados por Bianco y Boente (1998), que son una versión robusta de los definidos por Baek y Wehrly (1993), pues no requerimos para obtener consistencia que $Y - g_\alpha(X_\alpha)$ sea independiente de X_α y c) los estimadores definidos por Li, Zheng y Zheng (2012) ya que no necesitamos que la dimensión d de la covariables sea menor que 4 para tener distribución asintótica normal de las componentes aditivas, es decir, resolvemos el problema de la *maldición de la dimensión*. Para mostrar que las propuestas consideradas aproximan efectivamente las componentes aditivas, se prueba la consistencia fuerte de las alternativas consideradas, así como la consistencia uniforme sobre compactos. Por otra parte, se obtienen resultados de distribución asintótica para la familia de estimadores locales basados en polinomios de orden q que definiremos y que permiten resolver el problema de la alta dimensión en las covariables. Mediante un estudio de Monte Carlo, se muestra para muestras de tamaño moderado, el buen comportamiento de los estimadores robustos ante la presencia de distintos tipos de contaminación.

Por otro lado, hemos modificado el algoritmo de backfitting de modo a introducir una versión robusta del mismo y para dicha propuesta hemos obtenido resultados sobre su convergencia.

Por último, en esta tesis se desarrolló en R un algoritmo eficiente para el cálculo de los estimadores propuestos. La complejidad numérica que involucra el cálculo del estimador inicial combinada con el problema de tener que evaluarlo en una grilla de puntos conveniente de modo a poder aproximar la integral involucrada en el proceso de integración marginal que los define, era un desafío, en particular, en dimensión $d > 2$. Para resolverlo, se desarrollaron rutinas en C que se procesan en R de modo a acelerar el cálculo de los estimadores. Para el estimador robusto basado en backfitting también se desarrolló un algoritmo eficiente.

1.4 Estructura de la tesis

Esta tesis se organiza de la siguiente forma.

En el **Capítulo 2**, recordaremos los principales métodos de estimación en los modelos adi-

tivos basados en el método de integración marginal. En este mismo capítulo se describen diversos procedimientos para estimar la función de regresión en los modelos noparamétricos cuando hay respuestas faltantes, y se describen los últimos procedimientos existentes para la estimación de los efectos marginales en los modelos aditivos en presencia de respuestas faltantes. Finalmente, se dan las propuestas robustas de esta tesis.

En el **Capítulo 3**, se prueban resultados de consistencia puntual y uniforme para los $M-$estimadores locales robustos propuestos, mientras que en el **Capítulo 4**, se presentan resultados sobre su distribución asintótica.

En el **Capítulo 5**, se describen los procedimientos clásicos de estimación basados en *backfitting* y se introduce el estimador robusto basado en el procedimiento de *backfitting*. En este capítulo se supondrá que se cuenta con todas la respuestas, o sea, no se considerará el caso de respuestas faltantes.

Finalmente, en el **Capítulo 6**, se presentan los resultados de un estudio de simulación con el objetivo de analizar el comportamiento de las propuestas robustas respecto de sus versiones clásicas. Consideramos tanto datos sin contaminar como distintos tipos de contaminaciones que producen datos atípicos aislados o aglutinados.

Capítulo 2

Estimadores para los modelos aditivos con respuestas faltantes

2.1 Introducción

Sea Y la variable dependiente y $\mathbf{X}$ un vector d-dimensional de variables explicativas. Consideremos la estimación de la función de regresión $g(\mathbf{X}) = \mathbb{E}(Y|\mathbf{X})$. Cuando $d = 1$ y la función de regresión tiene segunda derivada continua, la tasa del error cuadrático medio asintótico es de orden $n^{4/5}$. Para el caso multidimensional, Stone (1985) mostró que la tasa de convergencia óptima para estimar g es $n^{\kappa/(2\kappa+d)}$ con κ el índice de suavidad de g. Por lo tanto, altos valores de d llevan a una tasa de convergencia más lenta. Este hecho está asociado al fenómeno de entornos cada vez más ralos al aumentar la dimensión. Como hemos mencionado, para resolver este problema, se introdujeron los modelos aditivos (1.1). Stone (1985) mostró que bajo un modelo aditivo la tasa óptima para estimar g es la tasa uno–dimensional $n^{\kappa/(2\kappa+1)}$. Hablamos entonces de una reducción de la dimensión. Los modelos aditivos combinan la flexibilidad de los modelos noparamétricos con la simple interpretación del modelo lineal estándar.

Buja, Hastie y Tibshirani (1989) y Hastie y Tibshirani (1990) propusieron un procedimiento iterativo denominado *backfitting* para estimar las componentes aditivas. Más recientemente, Tjøstheim y Auestad (1994) y Linton y Nielsen (1995) introdujeron un método no iterativo para estimar los efectos marginales que definimos como la esperanza de Y respecto del error aleatorio ε y de todas las covariables excepto X_α que se deja fija. El efecto marginal dice cómo varía Y en promedio al variar X_α. Notemos que los efectos marginales coinciden con las componentes aditivas g_α, excepto por la constante μ, si la verdadera función de regresión g es efectivamente aditiva. La idea de este método consiste en primero estimar el funcional multidimensional g y luego usar el procedimiento denominado *integración marginal* que describiremos más abajo para obtener los efectos marginales.

Consideraremos el problema de estimar, basados en la muestra aleatoria $\{(\mathbf{X}_i^{\mathrm{T}}, Y_i)^{\mathrm{T}}\}_{i=1}^n$, la esperanza condicional de Y dado $\mathbf{X} = \mathbf{x}$, o sea, la función $g(\mathbf{x})$ que supondremos satisface el modelo aditivo (1.1). Las funciones componentes $g_\alpha(\cdot)$ de (1.1) explican el impacto específico de la componente particular X_α en la respuesta Y. Para que el modelo sea identificable, necesitamos

agregar alguna restricción. Supongamos, por ejemplo, que

$$\mathbb{E}g_\alpha(X_\alpha) = 0 \qquad \forall\, \alpha = 1, \cdots, d\,. \tag{2.1}$$

De donde, $\mathbb{E}(Y) = \mu$. La constante μ puede ser fácilmente estimada con un orden de convergencia $\sqrt{n}$ mediante el estimador $\widehat{\mu} = \overline{Y} = (1/n)\sum_{i=1}^n Y_i$, por lo que en cuanto al comportamiento asintótico de los estimadores de g_α no hay pérdida de generalidad al suponer que $\mu = 0$.

El procedimiento de integración marginal busca estimar los efectos marginales de las variables regresoras X_α. Indiquemos por $\mathbf{X}_{\underline{\alpha}}$ al vector de todas las variables explicativas menos la X_α, es decir, $\mathbf{X}_{\underline{\alpha}} = (X_1, \ldots, X_{\alpha-1}, X_{\alpha+1}, \ldots, X_d)^{\mathrm{T}}$ y por $f_{\underline{\alpha}}$ su densidad conjunta que suponemos existe. El efecto marginal de una variable explicativa dice cómo cambia Y en promedio si esa variable varía. En otras palabras, el efecto marginal representa la esperanza "condicional" que indicaremos $\mathbb{E}_{\varepsilon, X_{\underline{\alpha}}}(Y|X_\alpha)$, donde la esperanza no es sólo respecto del error de distribución sino también de todas las variables regresoras excepto X_α que permanece fija y de allí la notación. Observemos que hasta este momento no habíamos indicado respecto de que variables tomábamos esperanza en las esperanzas condicionales, éste es el único caso donde necesitamos mencionar explícitamente respecto de cual distribución estamos calculando la esperanza. De ahora en más sólo lo indicaremos cuando sea preciso, para evitar confusión.

Como ya lo hemos mencionado, cuando se cumple el modelo aditivo (1.1), los efectos marginales corresponden a las funciones componentes aditivas $g_\alpha + \mu$. El estimador clásico está basado en la idea de integración, proveniente de la siguiente observación. Notemos por f_α la densidad marginal de X_α. Por (2.1) tenemos que $\mathbb{E}\left\{g_\alpha(X_\alpha)\right\} = \int g_\alpha(t) f_\alpha(t)\, dt = 0$, para todo $1 \le \alpha \le d$. Si ahora $g(\mathbf{X})$ cumple el modelo aditivo $g(\mathbf{x}) = \mu + \sum_{j=1}^d g_j(x_j)$, tenemos que

$$\int g(\mathbf{x}) f_{\underline{\alpha}}(\mathbf{x}_{\underline{\alpha}}) \prod_{k \neq \alpha} dX_k = \mathbb{E}_{\mathbf{X}_{\underline{\alpha}}} \left\{ \mu + g_\alpha(X_\alpha) + \sum_{k \neq \alpha} g_k(X_k) \right\} = \mu + g_\alpha(X_\alpha)\,. \tag{2.2}$$

Notemos que la expresión anterior coincide con $\mathbb{E}_{\varepsilon, \mathbf{X}_{\underline{\alpha}}}(Y|X_\alpha)$.

Para estimar simultáneamente las funciones y sus derivadas se combina el procedimiento de integración marginal con aproximaciones localmente polinomiales. Una aproximación basada en polinomios de orden 0, o sea, ajustando localmente una constante, es posible si sólo interesa estimar la función de regresión y no las derivadas de cada componente, esta aproximación es la que presentaremos en la Sección 2.1.1. Estimadores en presencia de respuestas faltantes se describen en la Sección 2.1.2.

2.1.1 Estimación de los efectos marginales

A fin de estimar el efecto marginal $g_\alpha(x)$, la ecuación (2.2) sugiere el siguiente procedimiento. En primer lugar, estimar la función g con un suavizador multidimensional $\widetilde{g}$ y luego integrar sobre todas las variables salvo X_α respecto de un estimador de la densidad $f_{\underline{\alpha}}$. En la estimación, el procedimiento de integración puede ser reemplazado por un promedio tomado en todas las componentes que no son la de interés, es decir, sobre $\mathbf{X}_{\underline{\alpha}}$. El estimador resulta ser entonces

$$\widehat{\{g_\alpha(x) + \mu\}} = \frac{1}{n} \sum_{i=1}^n \widetilde{g}(x, \mathbf{X}_{i\underline{\alpha}}),$$

donde por simplicidad de notación, indicamos por $(x, \mathbf{X}_{i\underline{\alpha}})$ al vector que tiene la componente α igual a x y todas las demás iguales a X_{ij}, $j \neq \alpha$. Notemos que para obtener los efectos marginales, sólo integramos $\widetilde{g}$ sobre todas las otras direcciones $\underline{\alpha}$. En caso de aditividad, estos efectos marginales son las funciones componentes aditivas g_α más la constante μ que, como ya hemos mencionado, puede ser estimada consistentemente con tasa $\sqrt{n}$. Por lo tanto, un posible estimador para g_α es

$$\widehat{g}_\alpha(x) = \frac{1}{n}\sum_{i=1}^{n} \widetilde{g}(x, \mathbf{X}_{i\underline{\alpha}}) - \widehat{\mu}\,, \tag{2.3}$$

y el estimador de la función de regresión g se obtiene como

$$\widehat{g}(\mathbf{x}) = \widehat{\mu} + \sum_{j=1}^{d} \widehat{g}_j(x_j)\ . \tag{2.4}$$

El estimador obtenido centrando las marginales, es decir, si definimos el estimador de g_α como

$$\widehat{g}_\alpha(x) = \{\widehat{g_\alpha(x) + \mu}\} - \frac{1}{n}\sum_{i=1}^{n}\{\widehat{g_\alpha(X_{i\alpha})} + \mu\} = \frac{1}{n}\sum_{i=1}^{n}\widetilde{g}(x, \mathbf{X}_{i\underline{\alpha}}) - \frac{1}{n}\sum_{i=1}^{n}\frac{1}{n}\sum_{\ell=1}^{n}\widetilde{g}(X_{i\alpha}, \mathbf{X}_{\ell\underline{\alpha}})$$

tiene el mismo comportamiento asintótico que el obtenido centrando usando $\widehat{\mu}$.

Nos queda por discutir cómo obtener un estimador razonable $\widetilde{g}(x_\alpha, \mathbf{x}_{\underline{\alpha}})$. En principio, éste podría ser cualquier estimador noparamétrico multivariado. En particular, si usamos el estimador de núcleos

$$\widetilde{g}_n(x, \mathbf{x}_{\underline{\alpha}}) = \frac{\displaystyle\sum_{i=1}^{n} K_{h_n}(X_{i\alpha} - x)\mathcal{L}_{\mathbf{H}_{d-1}}(\mathbf{X}_{i\underline{\alpha}} - \mathbf{x}_{\underline{\alpha}})Y_i}{\displaystyle\sum_{j=1}^{n} K_{h_n}(X_{j\alpha} - x)\mathcal{L}_{\mathbf{H}_{d-1}}(\mathbf{X}_{j\underline{\alpha}} - \mathbf{x}_{\underline{\alpha}})}\ ,$$

donde $K : \mathbb{R} \to \mathbb{R}$ es una función núcleo, $K_h(u) = K(u/h)/h$, $\mathbf{H}_{d-1} \in \mathbb{R}^{(d-1)\times(d-1)}$ es una matriz no singular y $\mathcal{L} : \mathbb{R}^{d-1} \to \mathbb{R}$ es tal que $\int \mathcal{L}(\mathbf{u})\, d\mathbf{u} = 1$ y $\mathcal{L}_{\mathbf{H}}(\mathbf{u}) = (\det(\mathbf{H}))^{-1}\mathcal{L}(\mathbf{H}^{-1}\mathbf{u})$, obtenemos

$$\widehat{g}_\alpha(x) = \frac{1}{n}\sum_{\ell=1}^{n} \frac{\sum_{i=1}^{n} K_{h_n}(X_{i\alpha} - x)\mathcal{L}_{\mathbf{H}_{d-1}}(\mathbf{X}_{i\underline{\alpha}} - \mathbf{X}_{\ell\underline{\alpha}})Y_i}{\sum_{j=1}^{n} K_{h_n}(X_{j\alpha} - x)\mathcal{L}_{\mathbf{H}_{d-1}}(\mathbf{X}_{j\underline{\alpha}} - \mathbf{X}_{\ell\underline{\alpha}})} - \overline{Y}.$$

Otra familia de estimadores se obtiene considerando una medida producto Q sobre $\mathbb{R}^d$. Sean $Q_{\underline{\alpha}}(\mathbf{x}_{\underline{\alpha}}) = Q(\mathbb{R}, \mathbf{x}_{\underline{\alpha}})d\mathbf{x}_{\underline{\alpha}}$, $q d\mathbf{x} = dQ(\mathbf{x})$, $q_{\underline{\alpha}} d\mathbf{x}_{\underline{\alpha}} = dQ_{\underline{\alpha}}(\mathbf{x}_{\underline{\alpha}})$. En este caso, en lugar de identificar las componentes en (1.1) mediante la condición $\mathbb{E}(g_\alpha(X_\alpha)) = 0$, se las identifica pidiendo que

$$\int g_\alpha(x)\, q_\alpha(x) dx = 0\,. \tag{2.5}$$

Los estimadores se definen entonces como

$$\widehat{\widehat{g}}_\alpha(x_\alpha) = \int \widetilde{g}(x_\alpha, \mathbf{u}_{\underline{\alpha}}) q_{\underline{\alpha}}(\mathbf{u}_{\underline{\alpha}})\, d\mathbf{u}_{\underline{\alpha}} - \widehat{\mu} \tag{2.6}$$

Estimadores basado en polinomios locales pueden verse en Härdle, Müller, Sperlich y Werwatz (2004).

Para el caso bivariado, Linton y Nielsen (1995) mostraron que integrar el estimador de Nadaraya-Watson produce estimadores de las componentes marginales que son asintóticamente normales con tasa de convergencia óptima. Algunos desarrollos heurísticos, basados en la consistencia del estimador piloto, sugieren que el estimador no convergería con tasa de convergencia óptima en presencia de más de cuatro covariables. Para resolver este problema, Hengartner y Sperlich (2005) propusieron un estimador inicial $\widetilde{g}$ de la función de regresión internamente corregido tal que la integración marginal aplicada a este estimador da como resultado estimadores de las componentes marginales con tasa óptima. La demostración dada por dichos autores revela que el estimador piloto sobresuavizará las variables a ser integradas y que el estimador resultante es en sí mismo un suavizador de regresión de menor dimensión. Por otra parte, Severance–Lossin y Sperlich (1999) propusieron un estimador basado en polinomios locales adaptivo a la componente a estimar, de modo a obtener tasas uniparamétricas óptimas. Si nos interesa estimar g_α, el procedimiento propuesto por estos autores considera el desarrollo sólo de la función g_α, es decir, el polinomio local está basado en un polinomio de grado q en $(X_{i\alpha} - x_\alpha)$ y utiliza núcleos de orden mayor a 2 en las componentes X_{ij} con $j \neq \alpha$. Una versión robusta de este estimador se discute en la Sección 2.4.

2.1.2 Estimadores cuando hay respuestas faltantes

Los estimadores anteriores suponen que se observan todas las respuestas. Como se mencionó en la Introducción en algunas situaciones, en algunas situaciones las respuestas no se observan completamente, ya sea por diseño, como en el caso de estudio en dos etapas, o por azar. El simple patrón de datos faltantes considerado típicamente se asocia con el esquema de muestra doble propuesto por Neyman (1938) para componer la falta de información de la respuesta Y considerando más (y posiblemente menos costosas) observaciones en las covariables $\mathbf{X}$. En la práctica, estos datos faltantes usualmente ocurren en forma de no–respuestas en muestras basadas en encuestas. En otros casos, los entrevistados se niegan a responder, por ejemplo sobre temas privados, o las respuestas pueden ser muy caras de obtener. En otros casos, la falta de datos puede ser causada por una falta de información debido a factores no controlables o por fallas en la recolección de los datos. Nosotros nos concentraremos en el caso en que existen observaciones faltantes entre las respuestas, pero las covariables se observan completamente.

Consideraremos, por lo tanto, una muestra aleatoria $(\mathbf{X}_i^{\mathrm{T}}, Y_i, \delta_i)^{\mathrm{T}}$, $1 \leq i \leq n$. Aquí las covariables $\mathbf{X}_i$ son observadas, mientras que la variable respuesta Y_i no es completamente observada. Para indicar la presencia o ausencia de respuesta introdujimos la variable indicadora δ_i que toma el valor 1 si Y_i se observa mientras que si Y_i está faltante $\delta_i = 0$, para $1 \leq i \leq n$. Supondremos que $(\mathbf{X}_i^{\mathrm{T}}, Y_i, \delta_i)^{\mathrm{T}}$ para $1 \leq i \leq n$ tiene la misma distribución que $(\mathbf{X}^{\mathrm{T}}, Y, \delta)^{\mathrm{T}}$ donde $(\mathbf{X}^{\mathrm{T}}, Y)^{\mathrm{T}}$ satisface el modelo aditivo (1.1).

Un tema de fundamental interés es estudiar el impacto de las observaciones faltantes en el funcionamiento de los estimadores de las componentes aditivas.

En esta tesis, supondremos que los datos son faltantes al azar (MAR). Con una aproximación puramente noparamétrica para discutir los datos faltantes, la suposición de MAR requiere la

existencia de un mecanismo de aleatoridad, denotado por $p(\mathbf{X})$, tal que

$$P(\delta = 1|\mathbf{X}, Y) = P(\delta = 1|\mathbf{X}) = p(\mathbf{X}) \tag{2.7}$$

vale casi seguramente. Por otro lado, MCAR demandaría que δ sea independiente tanto de $\mathbf{X}$ como de Y, es decir, $p(\mathbf{X})$ es idénticamente igual a una constante p entre 0 y 1.

En Boente y Martínez (2013) mostraron que aplicar el procedimiento descripto en la Sección 2.1.1 a la submuestra de datos completos $\{(\mathbf{X}_i^{\mathrm{T}}, Y_i)^{\mathrm{T}}_{\{i:\delta_i=1\}}\}$ introduce un sesgo en la estimación de las componentes aditivas. Para corregirlo, utilizan todas las covariables $\mathbf{X}_i$ y no sólo aquellas en las que la respuesta está presente. Estos autores propusieron dos familias de estimadores para los modelos aditivos con respuestas faltantes. Ambas propuestas se basan en aplicar el procedmiento de integración marginal a dos distintos estimadores preliminares de la función de regresión multivariada g. El primero, denotado por $\widetilde{g}_{\mathrm{S}}^{(1)}$, usa pesos con núcleos multiplicados por la indicadora de respuestas faltantes, mientras que el segundo, denotado por $\widetilde{g}_{\mathrm{S}}^{(2)}$, está relacionado con el estimador internamente corregido considerado en Hengartner y Sperlich (2005). Es decir, los estimadores $\widetilde{g}_{\mathrm{S}}^{(1)}$ y $\widetilde{g}_{\mathrm{S}}^{(2)}$ son los siguientes

$$\widetilde{g}_{\mathrm{S}}^{(1)}(\mathbf{x}) = \sum_{i=1}^{n} w_{i,\mathbf{H}_d}(\mathbf{x})\, Y_i \left(\sum_{j=1}^{n} w_{j,\mathbf{H}_d}(\mathbf{x})\right)^{-1}, \quad \widetilde{g}_{\mathrm{S}}^{(2)}(\mathbf{x}) = \sum_{i=1}^{n} \frac{w_{i,\mathbf{H}_d}(\mathbf{x})\, Y_i}{\widehat{f}(\mathbf{X}_i)} \left(\sum_{k=1}^{n} \frac{w_{k,\mathbf{H}_d}(\mathbf{x})}{\widehat{f}(\mathbf{X}_k)}\right)^{-1}, \tag{2.8}$$

donde $w_{i,\mathbf{H}_d}(\mathbf{x}) = \mathcal{K}_{\mathbf{H}_d}(\mathbf{x} - \mathbf{X}_i)\delta_i$, $\widehat{f}(\mathbf{x}) = (1/n)\sum_{j=1}^{n} \mathcal{K}_{\mathbf{H}_d}(\mathbf{x} - \mathbf{X}_j)$ es el estimador de densidad basado en núcleos y la matriz $\mathbf{H}_d = \mathrm{diag}(h_1, \ldots, h_d)$ es la matriz de ventanas.

Una vez elegido el estimador de regresión preliminar $\widetilde{g}_{\mathrm{S}}^{(1)}$ o $\widetilde{g}_{\mathrm{S}}^{(2)}$, mediante el procedimiento de integración marginal se obtienen los estimadores de las componentes aditivas $\widehat{g}_{\alpha,\mathrm{S}}^{(1)}$ $(\alpha = 1, \cdots, d)$ o $\widehat{g}_{\alpha,\mathrm{S}}^{(2)}$ $(\alpha = 1, \cdots, d)$ según corresponda y, finalmente, sumando los estimadores de las componentes aditivas se obtiene el estimador de la función de regresión final $\widehat{g}_{\mathrm{S}}^{(1)}$ o $\widehat{g}_{\mathrm{S}}^{(2)}$, como en (2.4), según qué estimador de la función de regresión preliminar se haya usado. Por estar basados en promedios pesados, estos estimadores son muy sensibles a datos atípicos en las variables respuestas y de allí surge la necesidad de introducir procedimientos menos sensibles a este tipo de datos.

En el caso de datos completos, o sea, cuando no hay datos faltantes, una propuesta robusta para modelos aditivos fue dada por Bianco y Boente (1998). Los estimadores considerados en Bianco y Boente (1998) son una versión robusta de los definidos por Baek y Wehrly (1993) y se obtienen sumando $M-$estimadores locales calculados condicionando solamente a la componente α de $\mathbf{X}$. La desventaja de este procedimiento es que para obtener consistencia, requiere que $Y - g_\alpha(X_\alpha)$ sea independiente de X_α, lo que restringe su aplicación debido al sesgo que introduce. Recientemente, Li, Zheng y Zheng (2012) dieron una propuesta robusta basada en polinomios locales de orden uno tomando el mismo núcleo K para todas las coordenadas y aplicando luego integración marginal. La propuesta de estos autores no resulta equivariante por escala y no considera el caso de observaciones faltantes. Como discutiremos más adelante, al considerar el mismo núcleo para todas las componentes surgen dos problemas. El primero y menos grave es que el sesgo de g_α depende de todas las componentes de g. El segundo es que para que el sesgo no vaya a infinito es necesario pedir tasas no unidimensionales.

Cuando hay respuestas faltantes, estimadores robustos para la función de regresión $g(\mathbf{x})$ basados en un $M-$estimador local, fueron considerados por Boente, González–Manteiga y Pérez–González

(2009). El $M-$estimador simplificado definido por Boente, González–Manteiga y Pérez–González (2009) usa la información accesible, es decir, el subconjunto de datos completos, y no supone un modelo aditivo. Indiquemos por $\widehat{F}(y|\mathbf{X}=\mathbf{x})$ la distribución empírica condicional que se define como

$$\widehat{F}(y|\mathbf{X}=\mathbf{x}) \quad = \quad \sum_{i=1}^{n} w_i(\mathbf{x}) I_{(-\infty,y]}(Y_i) \left(\sum_{j=1}^{n} w_j(\mathbf{x}) \right)^{-1} , \tag{2.9}$$

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

$$w_i(\mathbf{x}) = \mathcal{K}\left(\frac{\mathbf{X}_i - \mathbf{x}}{h_n} \right) \delta_i \, . \tag{2.10}$$

$\widehat{F}(y|\mathbf{X}=\mathbf{x})$ provee un estimador de la distribución de $Y|\mathbf{X}=\mathbf{x}$ que indicaremos por $F(y|\mathbf{X}=\mathbf{x})$ y fue estudiado por Cheng y Chu (1996). Como en (2.8), los pesos basados en núcleos fueron modificados multiplicándolos por el indicador de la variabe faltante para adaptarlos a la muestra completa y evitar sesgo. Entre otros estimadores, Boente, González–Manteiga y Pérez–González (2009) consideraron la mediana local calculada como la mediana de $\widehat{F}(y|\mathbf{X}=\mathbf{x})$. Un hecho interesante de la mediana local es que, a diferencia de otros estimadores robustos, no necesita un estimador consistente de escala cuando la función de escala $\sigma(\mathbf{x})$ es desconocida. Por otra parte, el $M-$estimador local $\widehat{g}_{\mathrm{M}}(\mathbf{x})$ se define como el $M-$estimador de posición asociado a $\widehat{F}(y|\mathbf{X}=\mathbf{x})$, es decir, es la solución de

$$\sum_{i=1}^{n} w_i(\mathbf{x}) \psi\left(\frac{Y_i - \widetilde{g}_{\mathrm{M}}(\mathbf{x})}{\widehat{s}(\mathbf{x})} \right) = 0 \, , \tag{2.11}$$

donde los pesos $w_i(\mathbf{x})$ están dados en (2.10), 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 función 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 (2.9). Observemos que la mediana local corresponde a elegir $\psi(t) = \mathrm{sg}(t)$. Este estimador, si bien está adaptado al hecho de tener observaciones faltantes, sufre el problema de la *maldición de la dimensión*. Por esta razón, en la Sección 2.2 describiremos cómo construir estimadores robustos para modelos aditivos basados en constantes locales, extenderemos esta definición al caso de polinomios locales en la Sección 2.4. En la Sección 2.3 consideraremos una extensión de los estimadores definidos por Li, Zheng y Zheng (2012) que resulta equivariante por escala y resuelve el problema del sesgo asintótico.

2.2 $M-$estimadores basados en polinomios locales de orden 0

Como antes, el parámetro μ puede estimarse con tasa $\sqrt{n}$, por ejemplo, mediante $\widehat{\mu} = \frac{1}{n}\sum_{i=1}^{n} \widetilde{g}_{\mathrm{M}}(\mathbf{X}_i)$ o $\widehat{\mu} = \int \widetilde{g}_{\mathrm{M}}(\mathbf{x}) q(\mathbf{x}) d\mathbf{x}$ según la condición de identificabilidad impuesta, donde $\widetilde{g}_{\mathrm{M}}$ es un estimador robusto inicial. Por esta razón, supondremos, de ahora en adelante, que $\mu = 0$. Tengamos en cuenta que si $\mu \neq 0$ en las expresiones de los estimadores de las componentes marginales que daremos en esta sección y las siguientes, se debe restar $\widehat{\mu}$ para obtener consistencia.

Para generalizar el estimador dado en (2.11), sea $\xi_i = \xi(\mathbf{X}_i) = \delta_i/t(\mathbf{X}_i)$ donde la función $t(\mathbf{x})$ se determinará más adelante. Sea $\widehat{s}(\mathbf{x})$ un estimador preliminar de escala robusto cuando $\sigma(\mathbf{x})$ es desconocido. Si la función de varianza es conocida tomamos $\widehat{s}(\mathbf{x}) = \sigma(\mathbf{x})$. Sea $\rho : \mathbb{R} \to \mathbb{R}$ una función ρ como se define en Maronna, Martin y Yohai (2006, Capítulo 2), es decir, una función ρ tal que $\rho(x)$ es nodecreciente en $|x|$, $\rho(0) = 0$, $\rho(x)$ es creciente para $x > 0$ cuando $\rho(x) < \|\rho\|_\infty$. Cuando ρ es acotada, supondremos que $\|\rho\|_\infty = 1$.

Definamos el $M-$estimador inicial $\widetilde{g}_{\mathrm{M}_0}(\mathbf{x})$ basado en polinomios locales de orden 0, donde M indica que es un $M-$estimador local y el subíndice 0 el orden del polinomio, como la solución del siguiente problema

$$\widetilde{g}_{\mathrm{M}_0}(\mathbf{x}) = \underset{a}{\operatorname{argmin}} \sum_{i=1}^{n} \rho\left(\frac{Y_i - a}{\widehat{s}(\mathbf{x})}\right) \mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i - \mathbf{x})\xi_i\,, \tag{2.12}$$

con $\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i - \mathbf{x}) = (\det(\mathbf{H}_d))^{-1}\mathcal{K}(\mathbf{H}_d^{-1}(\mathbf{X}_i - \mathbf{x}))$ y $\mathcal{K} : \mathbb{R}^d \to \mathbb{R}$ es una función núcleo producto de dimensión d, es decir, $\mathcal{K}(\mathbf{x}) = \prod_{j=1}^{d} K_j(x_j)$ con $K_j : \mathbb{R} \to \mathbb{R}$ núcleos univariados. Por otra parte, $\mathbf{H}_d = \operatorname{diag}(h_1, h_2, \ldots, h_d) \in \mathbb{R}^{d\times d}$ es una matriz de ventanas. Cabe mencionar que la matriz de ventanas se tomará en forma diferente según la componente a estimar. Así, si nos interesa estimar la componente α, el elemento diagonal α de $\mathbf{H}_d$ será igual a h_n y los demás serán iguales a $\widetilde{h}_n$. Como se menciona en Härdle, Müller, Sperlich y Werwatz (2004), la ventana $\widetilde{h}_n$ corresponde a las direcciones que pueden considerarse como *perturbadoras* en la estimación.

Luego, si $\rho : \mathbb{R} \to \mathbb{R}$ es derivable con derivada $\psi : \mathbb{R} \to \mathbb{R}$, $\widetilde{g}_{\mathrm{M}_0}(\mathbf{x})$ es el cero de la siguiente ecuación

$$\Psi_n(a) = \sum_{i=1}^{n} \psi\left(\frac{Y_i - a}{\widehat{s}(\mathbf{x})}\right) \mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i - \mathbf{x})\xi_i \;= 0\,. \tag{2.13}$$

Como en Boente, González–Manteiga y Pérez–González (2009) el estimador $\widetilde{g}_{\mathrm{M}_0}(\mathbf{x})$ corresponde al $M-$estimador de posición asociado a

$$\widehat{F}(y|\mathbf{X} = \mathbf{x}) = \frac{\displaystyle\sum_{i=1}^{n} \mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i - \mathbf{x})\xi_i\, I_{(-\infty,y]}(Y_i)}{\displaystyle\sum_{i=1}^{n} \mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i - \mathbf{x})\xi_i}\,.$$

Observemos que cuando $t(\mathbf{X}_i) \equiv 1$ y $h_1 = h_2 = \ldots, h_d = h$, $\widetilde{g}_{\mathrm{M}_0}(\mathbf{x}) = \widetilde{g}_{\mathrm{M}}(\mathbf{x})$ definido en (2.11). Por otra parte, la elección $t(\mathbf{X}_i) = p(\mathbf{X}_i)$ adapta al estimador al hecho de tener datos faltantes, si bien ambas elecciones de t proveen estimadores consistentes de $g(\mathbf{x})$.

El estimador $\widetilde{g}_{\mathrm{M}_0}(\mathbf{x})$ puede calcularse iterativamente usando ponderaciones iteradas, como se describe para posición en el Capítulo 2 de Maronna, Martin and Yohai (2006). El estimador inicial puede tomarse como la mediana local. En el paso $(\ell + 1)$ y dado $\widetilde{g}^{(\ell)}(\mathbf{x})$, los pesos involucrados en el algoritmo de ponderación iterada se calculan como $\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i - \mathbf{x})\xi_i\, W\left(\left(Y_i - \widetilde{g}^{(\ell)}(\mathbf{x})\right)/\widehat{s}(\mathbf{x})\right)$, con $W(t) = \psi(t)/t$ for $t \neq 0$ y $W(0) = \psi'(0)$. Luego, en el paso $(\ell + 1)$ obtenemos

$$\widetilde{g}^{(\ell+1)}(\mathbf{x}) = \frac{\displaystyle\sum_{i=1}^{n} \mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i - \mathbf{x})\xi_i\, W\left(\frac{Y_i - \widetilde{g}^{(\ell)}(\mathbf{x})}{\widehat{s}(\mathbf{x})}\right) Y_i}{\displaystyle\sum_{i=1}^{n} \mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i - \mathbf{x})\xi_i\, W\left(\frac{Y_i - \widetilde{g}^{(\ell)}(\mathbf{x})}{\widehat{s}(\mathbf{x})}\right)}$$

e iteramos hasta convergencia.

También podemos utilizar un algoritmo de Newton–Raphson para resolver (2.13), pero ambos procedimientos involucran un costo computacional elevado para cada punto $\mathbf{x}$. Para reducir el costo computacional, se puede utilizar un estimador de un paso de Newton. Específicamente, dado un estimador inicial $\widetilde{g}_{\mathrm{INI}}(\mathbf{x})$, que podría ser la mediana local $\widetilde{g}_{\mathrm{MED}}(\mathbf{x})$, el estimador de un paso, indicado con el subíndice OS por *one–step*, se define como

$$\widetilde{g}_{\mathrm{OS,M_0}}(\mathbf{x}) = \widetilde{g}_{\mathrm{INI}}(\mathbf{x}) - W_n^{-1}\Psi_n(\widetilde{g}_{\mathrm{INI}}(\mathbf{x}))$$

con

$$W_n = \Psi_n'(\widetilde{g}_{\mathrm{INI}}(\mathbf{x})) = -\frac{1}{\widehat{s}(\mathbf{x})}\sum_{i=1}^{n}\psi'\left(\frac{Y_i - \widetilde{g}_{\mathrm{INI}}(\mathbf{x})}{\widehat{s}(\mathbf{x})}\right)\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i - \mathbf{x})\xi_i\,.$$

Con el objetivo de obtener estimadores de las componentes del modelo aditivo, usaremos el método de integración marginal. Como antes, indiquemos por $\mathbf{X}_{\underline{\alpha}}$ al vector de todas las variables explicativas menos la X_α, es decir, $\mathbf{X}_{\underline{\alpha}} = (X_1, \ldots, X_{\alpha-1}, X_{(\alpha+1)}, \ldots, X_d)^{\mathrm{T}}$ y por $(x, \mathbf{X}_{i\underline{\alpha}})$ al vector que tiene la componente α igual a x y todas las demás iguales a X_{ij}, $j \neq \alpha$. Como en la Sección 2.1.1, sean Q una medida producto sobre $\mathbb{R}^d$ con $Q_{\underline{\alpha}}(\mathbf{x}_{\underline{\alpha}}) = Q(\mathbb{R}, \mathbf{x}_{\underline{\alpha}})d\mathbf{x}_{\underline{\alpha}}$, $qd\mathbf{x} = dQ$ y $q_{\underline{\alpha}}d\mathbf{x}_{\underline{\alpha}} = dQ_{\underline{\alpha}}$. Supongamos que las componentes en (1.1) cumplen (2.5).

El estimador robusto de $g_\alpha(x_\alpha)$ basado en polinomios locales de orden 0 se define como

$$\widehat{g}_{\alpha,\mathrm{M_0}}(x_\alpha) = \int \widetilde{g}_{\mathrm{M_0}}(x_\alpha, \mathbf{u}_{\underline{\alpha}})q_{\underline{\alpha}}(\mathbf{u}_{\underline{\alpha}})\,d\mathbf{u}_{\underline{\alpha}}\,. \tag{2.14}$$

Con el objetivo de reducir el costo computacional, podemos introducir el estimador robusto basado en el $M-$estimador local de un paso $\widetilde{g}_{\mathrm{OS,M_0}}(\mathbf{x})$. El estimador resultante de $g_\alpha(x_\alpha)$, indicado por $\widehat{g}_{\alpha,\mathrm{OS,M_0}}(x_\alpha)$, está dado por

$$\widehat{g}_{\alpha,\mathrm{OS,M_0}}(x_\alpha) = \int \widetilde{g}_{\mathrm{OS,M_0}}(x_\alpha, \mathbf{u}_{\underline{\alpha}})q_{\underline{\alpha}}(\mathbf{u}_{\underline{\alpha}})\,d\mathbf{u}_{\underline{\alpha}}\,. \tag{2.15}$$

2.3 $M-$estimadores basados en polinomios locales de orden 1 en todas las coordenadas de X

El problema de minimización dado en (2.12) utiliza polinomios locales de orden 0 para aproximar la respuesta en el modelo de regresión robusto, utilizaremos ahora polinomios locales de orden 1 que suelen tener menos efecto de frontera. Sea $\mathbf{a}(\mathbf{x}) = (g(\mathbf{x}), \nabla^{\mathrm{T}}g(\mathbf{x}))^{\mathrm{T}}$ donde $\nabla g(\mathbf{x}) = (g_1'(x_1), \ldots, g_d'(x_d))^{\mathrm{T}}$ es el gradiente de $g(\mathbf{x})$ y $\widehat{s}(\mathbf{x})$ un estimador preliminar de escala robusto cuando $\sigma(\mathbf{x})$ es desconocido, si la función de varianza es conocida tomamos $\widehat{s}(\mathbf{x}) = \sigma(\mathbf{x})$. Como en Li, Zheng y Zheng (2010), consideramos el problema de minimización

$$\widehat{\mathbf{a}}(\mathbf{x}) = (\widehat{a}_0(\mathbf{x}), \widehat{\mathbf{a}}_1(\mathbf{x})^{\mathrm{T}})^{\mathrm{T}} = \underset{(a_0,\mathbf{a}^{\mathrm{T}})^{\mathrm{T}}}{\operatorname{argmin}}\sum_{i=1}^{n}\rho\left(\frac{Y_i - [a_0 + \mathbf{a}_1^{\mathrm{T}}(\mathbf{X}_i - \mathbf{x})]}{\widehat{s}(\mathbf{x})}\right)\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i - \mathbf{x})\xi_i \tag{2.16}$$

con $\mathbf{a}_1 \in \mathbb{R}^d$ y donde las funciones ρ, $\mathcal{K}$ y la matriz $\mathbf{H}_d$ son como en la Sección 2.2. Luego, el estimador tipo $M-$estimador local basado en polinomios de orden 1 con respuestas faltantes

de $\mathbf{a}(\mathbf{x}) = (g(\mathbf{x}), \nabla^{\mathrm{T}} g(\mathbf{x}))^{\mathrm{T}}$ es $\widehat{\mathbf{a}}(\mathbf{x})$. En particular, $\widetilde{g}_{\mathrm{M}_1}(\mathbf{x}) = \widehat{a}_0(\mathbf{x})$ es un estimador de $g(\mathbf{x})$. Equivalentemente, cuando ρ tiene derivada ψ, $\widehat{\mathbf{a}}(\mathbf{x})$ satisface las siguientes ecuaciones

$$\mathbf{\Psi}_n(\widehat{\mathbf{a}}(\mathbf{x}), \mathbf{x}, \widehat{s}(\mathbf{x})) = \mathbf{0}_{d+1}\,, \tag{2.17}$$

con $\mathbf{0}_s \in \mathbb{R}^s$ un vector con sus componentes iguales a 0 y $\mathbf{\Psi}_n(\mathbf{a}, \mathbf{x}, \sigma) = (\Psi_{n0}(\mathbf{a}, \mathbf{x}, \sigma), \ldots, \Psi_{nd}(\mathbf{a}, \mathbf{x}, \sigma))^{\mathrm{T}}$ está definida por

$$\begin{aligned}
\Psi_{n0}(\mathbf{a}, \mathbf{x}, \sigma) &= \sum_{i=1}^{n} \psi\left(\frac{Y_i - [a_0 + \mathbf{a}_1^{\mathrm{T}}(\mathbf{X}_i - \mathbf{x})]}{\sigma}\right) \mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i - \mathbf{x})\xi_i\,, \\
\Psi_{n\ell}(\mathbf{a}, \mathbf{x}, \sigma) &= \sum_{i=1}^{n} \psi\left(\frac{Y_i - [a_0 + \mathbf{a}_1^{\mathrm{T}}(\mathbf{X}_i - \mathbf{x})]}{\sigma}\right) \mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i - \mathbf{x})\xi_i(X_{i,\ell} - x_\ell)\,, \quad 1 \le \ell \le d\,.
\end{aligned}$$

Como en la Sección 2.2, podemos considerar un estimador de un paso. Para ello, dado un estimador inicial de $\mathbf{a}(\mathbf{x})$, es decir de g y de su gradiente, que indicaremos por $\widehat{\mathbf{a}}_{\mathrm{INI}}(\mathbf{x})$, el estimador de un paso de $\mathbf{a}(\mathbf{x})$ se define como

$$\widehat{\mathbf{a}}_{\mathrm{OS}}(\mathbf{x}) = \widehat{\mathbf{a}}_{\mathrm{INI}}(\mathbf{x}) - \mathbf{W}_n^{-1}\mathbf{\Psi}_n(\widehat{\mathbf{a}}_{\mathrm{INI}}(\mathbf{x}), \mathbf{x}, \widehat{s}(\mathbf{x}))\,, \tag{2.18}$$

donde $\mathbf{W}_n = (w_{km}) \in \mathbb{R}^{(d+1)\times(d+1)}$ es la matriz Hessiana evaluada en el estimador inicial, es decir, con elementos dados por

$$w_{km} = \left.\frac{\partial \Psi_{nk}(\mathbf{a}, \mathbf{x}, \widehat{s}(\mathbf{x}))}{\partial a_m}\right|_{\mathbf{a}=\widehat{\mathbf{a}}_{\mathrm{INI}}(\mathbf{x})} \qquad 0 \le k, m \le d\,.$$

Como antes, la primera componente, $\widehat{a}_{0,\mathrm{OS}}(\mathbf{x})$, de $\widehat{\mathbf{a}}_{\mathrm{OS}}(\mathbf{x})$ es un estimador de $g(\mathbf{x})$ que indicaremos por $\widetilde{g}_{\mathrm{OS},\mathrm{M}_1}(\mathbf{x})$.

Como en la Sección 2.2, el estimador robusto de $g_\alpha(x_\alpha)$ basado en polinomios locales de orden 1 se define como

$$\widehat{g}_{\alpha,\mathrm{M}_1}(x_\alpha) = \int \widetilde{g}_{\mathrm{M}_1}(x_\alpha, \mathbf{u}_{\underline{\alpha}}) q_{\underline{\alpha}}(\mathbf{u}_{\underline{\alpha}})\, d\mathbf{u}_{\underline{\alpha}} = \int \mathbf{e}_1^{\mathrm{T}}\, \widehat{\mathbf{a}}(x_\alpha, \mathbf{u}_{\underline{\alpha}}) q_{\underline{\alpha}}(\mathbf{u}_{\underline{\alpha}})\, d\mathbf{u}_{\underline{\alpha}}\,. \tag{2.19}$$

donde $\mathbf{e}_j \in \mathbb{R}^{d+1}$ es el vector con su $j-$ésima coordenada igual a 1 y las restantes iguales a 0. Mientras que el estimador robusto basado en el $M-$estimador local de un paso $\widetilde{g}_{\mathrm{OS}}(\mathbf{x})$ está dado por

$$\widehat{g}_{\alpha,\mathrm{OS},\mathrm{M}_1}(x_\alpha) = \int \widetilde{g}_{\mathrm{OS},\mathrm{M}_1}(x_\alpha, \mathbf{u}_{\underline{\alpha}}) q_{\underline{\alpha}}(\mathbf{u}_{\underline{\alpha}})\, d\mathbf{u}_{\underline{\alpha}} = \int \mathbf{e}_1^{\mathrm{T}} \widehat{\mathbf{a}}_{\mathrm{OS}}(x_\alpha, \mathbf{u}_{\underline{\alpha}}) q_{\underline{\alpha}}(\mathbf{u}_{\underline{\alpha}})\, d\mathbf{u}_{\underline{\alpha}}\,. \tag{2.20}$$

2.4 $M-$estimadores de regresión usando polinomios locales de orden q

Una mejora en el ajuste se puede obtener usando polinomios de orden superior en todas las coordenadas salvo la correspondiente a la componente a estimar. Para ello, como en Severance–Lossin y Sperlich (1999), haremos uso de un tipo especial de estimadores locales adaptándolos al hecho de tener respuestas faltantes y proveyendo una versión robusta de la propuesta dada por esos autores.

Supongamos que nos interesa estimar la componente α, es decir, $g_\alpha(x)$. El estimador a definir trata en forma diferente las covariables correspondientes a esa componente y calcula un polinomio local de orden q en esa dirección. Para ser más precisos, sea $\boldsymbol{\beta}(\mathbf{x}) = (g(\mathbf{x}), g'_\alpha(x_\alpha), \ldots, g_\alpha^{(q)}(x_\alpha)/q!)^{\mathrm{T}}$ donde $g_\alpha^{(\nu)}(x_\alpha)$ es la derivada de orden ν de la componente g_α de $g(\mathbf{x})$ y sea $\widehat{s}(\mathbf{x})$ un estimador preliminar de escala robusto cuando $\sigma(\mathbf{x})$ es desconocido.

El estimador preliminar $\widetilde{g}_{\mathrm{M}_{q,\alpha}}(\mathbf{x})$ se obtendrá a partir del problema de minimización

$$\widehat{\boldsymbol{\beta}}(\mathbf{x}) = \underset{(\beta_0,\beta_1,\ldots,\beta_q)}{\operatorname{argmin}} \sum_{i=1}^{n} \rho\left(\frac{Y_i - \left[\beta_0 + \sum_{j=1}^{q} \beta_j (X_{i\alpha} - x_\alpha)^j\right]}{\widehat{s}(\mathbf{x})}\right) \mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i - \mathbf{x})\xi_i \tag{2.21}$$

donde $\widehat{\boldsymbol{\beta}}(\mathbf{x}) = (\widehat{\beta}_0(\mathbf{x}), \widehat{\beta}_1(\mathbf{x}), \ldots, \widehat{\beta}_q(\mathbf{x}))^{\mathrm{T}}$, $\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i - \mathbf{x}) = (\det(\mathbf{H}_d))^{-1}\mathcal{K}(\mathbf{H}_d^{-1}(\mathbf{X}_i - \mathbf{x}))$, $\mathcal{K}(\mathbf{x}) = \prod_{j=1}^{d} K_j(x_j)$ con $K_j : \mathbb{R} \to \mathbb{R}$ núcleos univariados. Por otra parte, $\mathbf{H}_d$ es una matriz diagonal de ventanas.

Cuando ρ tiene derivada ψ, $\widehat{\boldsymbol{\beta}}(\mathbf{x})$ satisface las siguientes ecuaciones

$$\boldsymbol{\Psi}_{n,\alpha}(\widehat{\boldsymbol{\beta}}(\mathbf{x}), \mathbf{x}, \widehat{s}(\mathbf{x})) = \mathbf{0}_{d+1}\,, \tag{2.22}$$

donde $\boldsymbol{\Psi}_{n,\alpha}(\boldsymbol{\beta}, \mathbf{x}, \sigma) = (\Psi_{n,\alpha,0}(\boldsymbol{\beta}, \mathbf{x}, \sigma), \ldots, \Psi_{n,\alpha,q}(\boldsymbol{\beta}, \mathbf{x}, \sigma))^{\mathrm{T}}$ está definida por

$$\Psi_{n,\alpha,0}(\boldsymbol{\beta}, \mathbf{x}, \sigma) = \sum_{i=1}^{n} \psi\left(\frac{Y_i - \beta_0 - \sum_{j=1}^{q} \beta_j (X_{i\alpha} - x_\alpha)^j}{\sigma}\right) \mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i - \mathbf{x})\xi_i\,,$$

$$\Psi_{n,\alpha,\ell}(\boldsymbol{\beta}, \mathbf{x}, \sigma) = \sum_{i=1}^{n} \psi\left(\frac{Y_i - \beta_0 - \sum_{j=1}^{q} \beta_j (X_{i\alpha} - x_\alpha)^j}{\sigma}\right) \mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i - \mathbf{x})\xi_i (X_{i\alpha} - x_\alpha)^\ell\,,\ 1 \le \ell \le d\,.$$

De esta forma, obtenemos que el estimador inicial de $g(\mathbf{x})$ está dado por $\widetilde{g}_{\mathrm{M}_{q,\alpha}}(\mathbf{x}) = \widehat{\beta}_0(\mathbf{x})$, donde M indica que es un $M-$estimador local y el subíndice q el orden del polinomio usando sobre la coordenada α de $\mathbf{x}$. Observemos que si $q = 1$, este estimador no coincide con el estimador $\widetilde{g}_{\mathrm{M}_1}(\mathbf{x})$ definido en la Sección 2.3, ya que este último utiliza todas las coordenadas de $\mathbf{X}_i - \mathbf{x}$ al utilizar la aproximación lineal de la función g, mientras que $\widetilde{g}_{\mathrm{M}_{q,\alpha}}(\mathbf{x})$ sólo desarrolla en la dirección de la coordenada de interés permitiendo de esta forma, al utilizar el modelo aditivo, obtener estimadores para g_α y sus derivadas. El estimador robusto de la componente α se obtiene como en (2.19), es decir,

$$\widehat{g}_{\alpha,\mathrm{M}_{q,\alpha}}(x_\alpha) = \int \widetilde{g}_{\mathrm{M}_{q,\alpha}}(x_\alpha, \mathbf{u}_{\underline{\alpha}}) q_{\underline{\alpha}}(\mathbf{u}_{\underline{\alpha}})\, d\mathbf{u}_{\underline{\alpha}} = \int \mathbf{e}_1^{\mathrm{T}}\, \widehat{\boldsymbol{\beta}}(x_\alpha, \mathbf{u}_{\underline{\alpha}}) q_{\underline{\alpha}}(\mathbf{u}_{\underline{\alpha}})\, d\mathbf{u}_{\underline{\alpha}}\,. \tag{2.23}$$

donde $\mathbf{e}_j \in \mathbb{R}^{d+1}$ es el vector con su $j-$ésima coordenada igual a 1 y las restantes iguales a 0. Más aún, el estimador de la derivada de orden ν, $1 \le \nu \le q$ de g_α está dado por

$$\widehat{g}^{(\nu)}_{\alpha,\mathrm{M}_{q,\alpha}}(x_\alpha) = \nu! \int \widehat{\beta}_\nu(x_\alpha, \mathbf{u}_{\underline{\alpha}}) q_{\underline{\alpha}}(\mathbf{u}_{\underline{\alpha}})\, d\mathbf{u}_{\underline{\alpha}} = \nu! \int \mathbf{e}_{\nu+1}^{\mathrm{T}}\, \widehat{\boldsymbol{\beta}}(x_\alpha, \mathbf{u}_{\underline{\alpha}}) q_{\underline{\alpha}}(\mathbf{u}_{\underline{\alpha}})\, d\mathbf{u}_{\underline{\alpha}}\,.$$

Para calcular la distribución asintótica de estos estimadores, supondremos que $\mathbf{H}_d$ es una matriz diagonal de ventanas tal que el elemento diagonal α es igual a h_α y los demás son iguales a $\widetilde{h}$. Más aún, al estimar la componente α tomaremos un núcleo K positivo, acotado, simétrico para esa componente, o sea, $K_\alpha = K$ mientras que para las restantes tomaremos $K_j = L$, $j \neq \alpha$, donde L es un núcleo de orden ℓ mayor a 2, es decir, $\int L(u)du = 1$, $\int u^s L(u)du = 0$, para $s = 1, \ldots, \ell - 1$ y $\int u^\ell L(u)du \neq 0$.

Capítulo 3

Consistencia de los $M-$estimadores

En este capítulo, probaremos la consistencia de los estimadores robustos definidos en las secciones 2.2 a 2.4. Como en esas secciones, supondremos que Q es una medida producto con densidad $q(\mathbf{x})$. Indicaremos por $q_\alpha(x)\,dx = dQ_\alpha(x)$ donde Q_α es la α-ésima marginal de la medida Q. Por otra parte, como mencionamos en la Sección 2.2, el parámetro μ puede estimarse con tasa $\sqrt{n}$, por lo tanto, de ahora en adelante, supondremos que $\mu = 0$.

Consideraremos una muestra aleatoria $(\mathbf{X}_i^{\mathrm{T}}, Y_i, \delta_i)^{\mathrm{T}}$, $1 \leq i \leq n$ con la misma distribución que $(\mathbf{X}^{\mathrm{T}}, Y, \delta)^{\mathrm{T}}$, donde el mecanismo de pérdida de datos es independiente de las respuestas Y condicional a las covariables $\mathbf{X}$, o sea, se cumple (2.7). Por otra parte, $(\mathbf{X}^{\mathrm{T}}, Y)^{\mathrm{T}}$ cumplen el modelo aditivo

$$Y = \sum_{j=1}^{d} g_j(X_j) + \sigma(\mathbf{X})\varepsilon \ , \tag{3.1}$$

donde $Y \in \mathbb{R}$ es la variable respuesta, X_j $(j = 1, \cdots, d)$ es la j-ésima componente del vector aleatorio $\mathbf{X} \in \mathbb{R}^d$, ε es una variable aleatoria no observable independiente de $\mathbf{X}$ con distribución simétrica F_0 y $\sigma : \mathbb{R}^d \to \mathbb{R}_{>0}$ es la función de varianza desconocida. Para identificar las componentes aditivas supondremos que

$$\int g_\alpha(x) q_\alpha(x)\,dx = 0 \qquad \alpha = 1, \ldots, d \tag{3.2}$$

3.1 Consistencia de los $M-$estimadores basados en polinomios locales de orden 0

Como hemos mencionado, cuando $h_1 = h_2 = \cdots = h_d = h$ y $t(\mathbf{x}) \equiv 1$, el estimador $\widetilde{g}_{\mathrm{M}_0}$ coincide con el estimador $\widetilde{g}_{\mathrm{M}}$ introducido en Boente, González–Manteiga y Pérez–González (2009). Estos autores mostraron bajo condiciones de regularidad, la convergencia uniforme sobre compactos. El resultado que daremos en esta sección, generaliza el dado por estos autores.

Como mencionamos en la Sección 2.2, el núcleo $\mathcal{K}$ y las ventanas h_j deben adaptarse a la componentes a estimar. Para darle un marco más general probaremos consistencia uniforme de $\widetilde{g}_{\mathrm{M}_0}$ cuando las ventanas y los núcleos son diferentes. Observemos que bajo (3.1), la función de

distribución $F(y|\mathbf{X}=\mathbf{x})$ de $Y|\mathbf{X}=\mathbf{x}$ cumple

$$F(y|\mathbf{X}=\mathbf{x}) = F_0\left(\frac{y-\sum_{j=1}^d g_j(x_j)}{\sigma(\mathbf{x})}\right),$$

con lo que resulta simétrica alrededor de $g(\mathbf{x}) = \sum_{j=1}^d g_j(x_j)$ si F_0 es simétrica. De ahora en más indicaremos por $r(\mathbf{x}) = p(\mathbf{x})/t(\mathbf{x})$.

Consideraremos los siguientes supuestos donde $\mathcal{C}$ es un conjunto compacto.

A0 La medida producto Q tiene soporte compacto $\mathcal{S}_Q$ contenido en el soporte de $f_{\mathbf{X}}$, $\mathcal{S}_f$. Indicaremos por $\mathcal{C}_\alpha$ el soporte de q_α.

A1 $(\mathbf{X}_i^{\mathrm{T}}, Y_i, \delta_i)^{\mathrm{T}}$, $1 \le i \le n$ son i.i.d. cumplen (2.7) y $(\mathbf{X}_i^{\mathrm{T}}, Y_i)^{\mathrm{T}}$ cumple el modelo aditivo (3.1) con ε_i independiente de $\mathbf{X}_i$ con distribución F_0 simétrica alrededor de 0 y Lipschitz continua. Las componentes g_α verifican (3.2).

A2 $\psi : \mathbb{R} \to \mathbb{R}$ es una función impar, estrictamente creciente, acotada y de variación acotada.

A3 La función de densidad de $\mathbf{X}$, $f_{\mathbf{X}}(\mathbf{x})$, y las funciones $t(\mathbf{x})$ y $p(\mathbf{x})$ son funciones no–negativas y acotadas en el compacto $\mathcal{C} \subset \mathcal{S}_f$. Más aún, $t(\mathbf{x}) > 0$ para todo $\mathbf{x} \in \mathcal{S}_f$ y las funciones p y t son funciones continuas en un entorno de $\mathcal{C}$. Además, se cumple que $i(p) = \inf_{\mathbf{x}\in\mathcal{C}} p(\mathbf{x}) > 0$, $i(t) = \inf_{\mathbf{x}\in\mathcal{C}} t(\mathbf{x}) > 0$ y $i(f_{\mathbf{X}}) = \inf_{\mathbf{x}\in\mathcal{C}} f_{\mathbf{X}}(\mathbf{x}) > 0$.

A4 $\sigma(\mathbf{x})$ y $g(\mathbf{x})$ son funciones continuas de $\mathbf{x}$ en un entorno de $\mathcal{C}$. Más aún, $i(\sigma) > 0$.

A5 El núcleo $\mathcal{K} : \mathbb{R}^d \to \mathbb{R}$ es una función acotada, no–negativa tal que $0 < \int \mathcal{K}(\mathbf{u})\, d\mathbf{u} < \infty$, $\int |\mathbf{u}|\mathcal{K}(\mathbf{u})\, d\mathbf{u} < \infty$, $\|\mathbf{u}\|^d\mathcal{K}(\mathbf{u}) \to 0$ cuando $\|\mathbf{u}\| \to \infty$ y cumple una condición de Lipschitz de orden 1.

A6 La sucesión de ventanas $h_j = h_{j,n} > 0$ es tal que $h_{j,n} \to 0$ y $n\prod_{j=1}^d h_{j,n}/\log n \to \infty$.

A7 El estimador robusto de escala cumple $A < \widehat{s}(\mathbf{x}) < B$, para todo $\mathbf{x} \in \mathcal{C}$ y para $n \ge n_0$ casi seguramente.

Observación 3.1.1. La hipótesis **A1** dice que el modelo considerado es un modelo aditivo y asegura que las componentes aditivas g_j son identificables. Las hipótesis **A3** y **A4** establecen condiciones de regularidad en la densidad marginal de $\mathbf{X}$ y en la función de distribución condicional. Observemos que **A3** implica que algunas variables respuestas son observadas para todo $\mathbf{x} \in \mathcal{C}$ y que $r(\mathbf{x}) = p(\mathbf{x})/t(\mathbf{x})$ es continua en un entorno de $\mathcal{C}$ y por lo tanto, acotada en $\mathcal{C}$. Por otra parte, como t es acotada en $\mathcal{C}$ resulta que $i(r) = \inf_{\mathbf{x}\in\mathcal{C}} r(\mathbf{x}) > 0$. La continuidad de g y la hipótesis **A5** son supuestos muy usados para suavizados basados en núcleos. **A5** incluye el caso en que $\mathcal{K}(\mathbf{u}) = \prod_{j=1}^d K_j(u_j)$ donde los núcleos K_j tienen todos soporte compacto como en las secciones siguientes. La hipótesis **A1** y **A2** implican que $\mathbb{E}[\psi(\varepsilon)] = 0$. Por último, alguna relación entre las ventanas y el tamaño de muestra es siempre necesario.

El siguiente resultado se obtiene usando argumentos análogos a los considerados en la demostración de la Proposición 3.2.1 de Boente, González–Manteiga y Pérez–González (2009) definiendo

$r(y,\mathbf{x}) = p(\mathbf{x})\, F(y|\mathbf{X}=\mathbf{x})/t(\mathbf{x})$, $\widehat{F}(y|\mathbf{X}=\mathbf{x}) = \widehat{r}(y,\mathbf{x})/\widehat{r}(\mathbf{x})$ donde $\widehat{r}(y,\mathbf{x}) = \sum_{i=1}^n W_{i,n}(\mathbf{x})\xi_i I_{(-\infty,y]}(Y_i)$, $\widehat{r}(\mathbf{x}) = \sum_{i=1}^n W_{i,n}(\mathbf{x})\xi_i$ y

$$W_{i,n}(\mathbf{x}) = \mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i - \mathbf{x}) \left\{ \sum_{j=1}^n \mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_j - \mathbf{x}) \right\}^{-1}, \tag{3.3}$$

y probando que

$$\sup_{y\in\mathbb{R}} \sup_{\mathbf{x}\in\mathcal{C}} |\widehat{r}(y,\mathbf{x}) - r(y,\mathbf{x})| \xrightarrow{a.s.} 0 \qquad \text{y} \qquad \sup_{\mathbf{x}\in\mathcal{C}} |\widehat{r}(\mathbf{x}) - r(\mathbf{x})| \xrightarrow{a.s.} 0\,. \tag{3.4}$$

Proposición 3.1.1. *Supongamos que se cumplen **A1** a **A6** entonces*

a) $\sup_{y\in\mathbb{R}} \sup_{\mathbf{x}\in\mathcal{C}} |\widehat{F}(y|\mathbf{X}=\mathbf{x}) - F(y|\mathbf{X}=\mathbf{x})| \xrightarrow{a.s.} 0$,

*b) Si además vale **A7**, entonces* $\sup_{\mathbf{x}\in\mathcal{C}} |\widetilde{g}_{\mathrm{M}_0}(\mathbf{x}) - g(\mathbf{x})| \xrightarrow{a.s.} 0$.

c) Sea $\widetilde{g}_{\mathrm{MED}}(\mathbf{x})$ *la mediana de* $\widehat{F}(y|\mathbf{X}=\mathbf{x})$. *Si* F_0 *tiene una única mediana en* 0 *entonces,* $\sup_{\mathbf{x}\in\mathcal{C}} |\widetilde{g}_{\mathrm{MED}}(\mathbf{x}) - g(\mathbf{x})| \xrightarrow{a.s.} 0$.

El supuesto **A7**, se cumple por ejemplo si consideramos la MAD respecto de la distribución $\widehat{F}(y|\mathbf{X}=\mathbf{x})$ como se deduce del Lema 3.1 de Boente y Fraiman (1991) y de la Proposición 3.1.1 a).

El siguiente resultado muestra la consistencia del estimador de la componente g_α definido por (2.14). En particular, en **A6** podemos tomar, $h_{\alpha,n} = h_n$ y $h_{j,n} = \widetilde{h}_n$ al estimar g_α. Omitimos la demostración ya que es inmediata y se sigue como en la demostración del Teorema 3.2.3 en Boente y Martínez (2013).

Teorema 3.1.1. *Supongamos que se cumplen **A0** a **A7** con* $\mathcal{C} = \mathcal{S}_Q$, *entonces*

a) $\sup_{x\in\mathcal{C}_\alpha} |\widehat{g}_{\alpha,\mathrm{M}_0}(x) - g_\alpha(x)| \xrightarrow{a.s.} 0$,

b) $\sup_{\mathbf{x}\in\mathcal{C}} |\widehat{g}_{\mathrm{M}_0}(\mathbf{x}) - g(\mathbf{x})| \xrightarrow{a.s.} 0$, *donde* $\widehat{g}_{\mathrm{M}_0}(\mathbf{x}) = \sum_{j=1}^d \widehat{g}_{j,\mathrm{M}_0}(x_j)$.

Finalmente, el siguiente resultado muestra la consistencia del estimador de un paso de la componente g_α definido por (2.15).

Teorema 3.1.2. *Supongamos que* $\widetilde{g}_{\mathrm{INI}}(\mathbf{x})$ *es un estimador tal que se cumplen* $\sup_{\mathbf{x}\in\mathcal{C}} |\widetilde{g}_{\mathrm{INI}}(\mathbf{x}) - g(\mathbf{x})| \xrightarrow{a.s.} 0$ *y que se cumplen **A0** a **A7** con* $\mathcal{C} = \mathcal{S}_Q$. *Si además* ψ *es derivable con derivada* ψ' *acotada, Lipschitz, de variación acotada y tal que* $\mathbb{E}(\psi'(\varepsilon)) \neq 0$, *entonces*

a) $\sup_{\mathbf{x}\in\mathcal{C}} |\widetilde{g}_{\mathrm{OS},\mathrm{M}_0}(\mathbf{x}) - g(\mathbf{x})| \xrightarrow{a.s.} 0$

b) $\sup_{x\in\mathcal{C}_\alpha} |\widehat{g}_{\alpha,\mathrm{OS},\mathrm{M}_0}(x) - g_\alpha(x)| \xrightarrow{a.s.} 0$,

c) $\sup_{\mathbf{x}\in\mathcal{C}} |\widehat{g}_{\mathrm{OS},\mathrm{M}_0}(\mathbf{x}) - g(\mathbf{x})| \xrightarrow{a.s.} 0$, *donde* $\widehat{g}_{\mathrm{OS},\mathrm{M}_0}(\mathbf{x}) = \sum_{j=1}^d \widehat{g}_{j,\mathrm{OS},\mathrm{M}_0}(x_j)$.

DEMOSTRACIÓN. Basta probar a) pues b) y c) se deducen en forma inmediata. Para probar a) definamos

$$\begin{aligned}\Lambda_n(\mathbf{x},a,\sigma) &= \sum_{i=1}^{n} W_{i,n}(\mathbf{x})\xi_i\psi\left(\frac{Y_i-a}{\sigma}\right)\\ \Lambda_{1,n}(\mathbf{x},a,\sigma) &= \sum_{i=1}^{n} W_{i,n}(\mathbf{x})\xi_i\psi'\left(\frac{Y_i-a}{\sigma}\right)\end{aligned}$$

donde $W_{i,n}(\mathbf{x})$ están definidos en (3.3). Observemos que

$$W_n^{-1}\Psi_n(\widetilde{g}_{\mathrm{INI}}(\mathbf{x})) = -\widehat{s}(\mathbf{x})\frac{\Lambda_n(\mathbf{x},\widetilde{g}_{\mathrm{INI}}(\mathbf{x}),\widehat{s}(\mathbf{x}))}{\Lambda_{1,n}(\mathbf{x},\widetilde{g}_{\mathrm{INI}}(\mathbf{x}),\widehat{s}(\mathbf{x}))}\,.$$

Luego, como $\sup_{\mathbf{x}\in\mathcal{C}}|\widetilde{g}_{\mathrm{INI}}(\mathbf{x})-g(\mathbf{x})| \xrightarrow{a.s.} 0$, $\mathbb{E}\psi'(\varepsilon)\neq 0$, $i(r)>0$ y $A<\widehat{s}(\mathbf{x})<B$ basta con probar que

$$\sup_{\mathbf{x}\in\mathcal{C}}|\Lambda_n(\mathbf{x},\widetilde{g}_{\mathrm{INI}}(\mathbf{x}),\widehat{s}(\mathbf{x}))| \xrightarrow{a.s.} 0 \tag{3.5}$$

$$\sup_{\mathbf{x}\in\mathcal{C}}|\Lambda_{1,n}(\mathbf{x},\widetilde{g}_{\mathrm{INI}}(\mathbf{x}),\widehat{s}(\mathbf{x}))-r(\mathbf{x})\mathbb{E}(\psi'(\varepsilon))| \xrightarrow{a.s.} 0\,. \tag{3.6}$$

Como en la demostración del Teorema 3.3 de Boente y Fraiman (1991), tenemos que para todo a

$$\sup_{\mathbf{x}\in\mathcal{C}}|\Lambda_n(\mathbf{x},a,\widehat{s}(\mathbf{x}))-\Lambda(\mathbf{x},a,\widehat{s}(\mathbf{x}))| \leq \|\psi\|_{\mathrm{V}}\sup_{y\in\mathbb{R}}\sup_{\mathbf{x}\in\mathcal{C}}|\widehat{r}(y,\mathbf{x})-r(y,\mathbf{x})|\,,$$

donde

$$\Lambda(\mathbf{x},a,\sigma) = r(\mathbf{x})\mathbb{E}\left\{\psi\left(\frac{Y-a}{\sigma}\right)|\mathbf{X}=\mathbf{x}\right\} = r(\mathbf{x})\mathbb{E}\left\{\psi\left(\frac{\sigma(\mathbf{x})\varepsilon+g(\mathbf{x})-a}{\sigma}\right)\right\}\,.$$

Observemos que como ψ es derivable con derivada acotada, $|\Lambda(\mathbf{x},\widetilde{g}_{\mathrm{INI}}(\mathbf{x}),\sigma)-\Lambda(\mathbf{x},g(\mathbf{x}),\sigma)| \leq \|\psi'\|_\infty\, r(\mathbf{x})\,|\widetilde{g}_{\mathrm{INI}}(\mathbf{x})-g(\mathbf{x})|\,/\sigma$. Luego, (3.5) se obtiene de (3.4) y del hecho que $\Lambda(\mathbf{x},g(\mathbf{x}),\sigma)=0$ para todo σ y que $\inf_{\mathbf{x}\in\mathcal{C}}\widehat{s}(\mathbf{x})>A$. La demostración de (3.6) se obtiene en forma análoga. □

3.2 Consistencia de los $M-$estimadores locales basados en polinomios locales de orden 1

Para obtener la consistencia de los $M-$estimadores locales basados en polinomios locales de orden 1 necesitaremos algunos supuestos adicionales e introduciremos en la próxima sección notación que nos resultará de utilidad.

3.2.1 Notación y Supuestos

Además de algunos de los supuestos dados en la sección 3.1 necesitaremos las siguientes hipótesis.

C1 Para todo $j=1,\cdots,d$, la componente marginal g_j es dos veces continuamente diferenciable en un entorno del soporte, $\mathcal{S}_j$, de la densidad de X_j y además g_j'' es acotada en $\mathcal{S}_j$.

C2 La función de núcleos $\mathcal{K} : \mathbb{R}^d \to \mathbb{R}$ es tal que $\mathcal{K}(\mathbf{x}) = \prod_{j=1}^d K_j(x_j)$, donde $K_j : \mathbb{R} \to \mathbb{R}$ son funciones pares, acotadas, con soporte acotado, digamos $[-1, 1]$. Supondremos sin pérdida de generalidad que $\int K_j(u)du = 1$ y sea $0 < s_j = \int u^2 K_j(u)\, du < \infty$.

C3 a) La función ρ es una función par, dos veces continuamente diferenciable con derivada primera ψ acotada y tal que la derivada segunda, ψ' es acotada y cumple $\mathbb{E}(\psi'(\varepsilon)) > 0$ y $\zeta(u) = u\psi'(u)$ es acotada.

b) Para algún $0 < \gamma < 1$, la función ψ' satisface

$$\begin{aligned} \mathbb{E}\left[\sup_{|a-1|\le\gamma}\sup_{|t|\le\delta} |\psi'(\varepsilon a + t) - \psi'(\varepsilon a)|\right] &= o(1)\,, \\ \mathbb{E}\left[\sup_{|a-1|\le\gamma}\sup_{|t|\le\delta} |\psi(\varepsilon a + t) - \psi(\varepsilon a) - \psi'(\varepsilon a)\, t|\right] &= o(\delta) \ \text{ cuando } \delta \to 0\,. \end{aligned}$$

c) La función ψ' satisface

$$\mathbb{E}\left[\sup_{|a-1|\le\delta} |\psi'(\varepsilon a) - \psi'(\varepsilon)|\right] = o(\delta) \ \text{ cuando } \delta \to 0\,.$$

C4 La sucesión de ventanas $h_j = h_{j,n} > 0$ es tal que $h_{j,n} \to 0$ y $n\prod_{j=1}^d h_{j,n} \to \infty$.

C5 La función $f_{\mathbf{X}}$ es una función continua en un entorno del compacto $\mathcal{C}$.

Observemos que **C3**b) y c) se cumplen si existe ψ es dos veces diferenciable y si su derivada segunda ψ'' es acotada y tal que $u\psi''(u)$ es acotada. La hipótesis de continuidad de ψ' en **C3**b) se puede debilitar a que ψ' exista y sea continua salvo en un número finito de puntos si, por ejemplo, ε tiene densidad.

Sea $s_{ij} = \int u_i u_j \mathcal{K}(\mathbf{u})\, d\mathbf{u}$ $0 \le i, j \le d$, donde $u_0 = 1$ and $\mathbf{u} = (u_1, \ldots, u_d)^{\mathrm{T}}$. Observemos que bajo **C2**, tenemos que $s_{ij} = 0$ si $i \ne j$, $s_{00} = 1$ y $s_{jj} = s_j = \int u^2 K_j(u)\, du$ para $1 \le j \le d$, con lo cual la matriz $\mathbf{S} = (s_{ij}) \in \mathbb{R}^{(d+1)\times(d+1)}$ es una matriz diagonal definida positiva, $\mathbf{S} = \mathrm{diag}(1, s_1, \cdots, s_d)$. Sea $U_i = \sigma(\mathbf{X}_i)\varepsilon_i$, luego, $Y_i = g(\mathbf{X}_i) + U_i$. Definamos $R(\mathbf{X}_i) = g(\mathbf{X}_i) - g(\mathbf{x}) - \sum_{j=1}^d g'_j(x_j)(X_{i,j} - x_j)$,

$$\begin{aligned} \breve{\mathbf{x}}_i &= (1, (X_{i,1} - x_1)/h_1, (X_{i,2} - x_2)/h_2, \cdots, (X_{i,d} - x_d)/h_d)^{\mathrm{T}}\,, \\ \mathbf{H} &= \mathrm{diag}(1, h_1, h_2, \cdots, h_d) \in \mathbb{R}^{(d+1)\times(d+1)}\,. \end{aligned}$$

La hipótesis **C2** asegura que la matriz $\mathbf{S}$ es no singular ya que $\int u^2 K_j(u)\, du > 0$. Esta hipótesis se utiliza para obtener la consistencia de los estimadores y por lo tanto, no permite utilizar núcleos de orden superior en las coordenadas $j \ne \alpha$ cuando estimo g_α.

Proposición 3.2.1. *Sea* $\mathbf{x}$ *un punto del interior de* $\mathcal{S}_f$ *y* $\mathcal{C}_0$ *la clausura de un entorno de* $\mathbf{x}$ *para el que se cumple* ***A3*** *con* $\mathcal{C} = \mathcal{C}_0$. *Supongamos que se cumplen* ***A1***, ***A4***, ***A7*** *y* ***C1*** *a* ***C5***. *Entonces, si* $\widehat{s}(\mathbf{x}) \xrightarrow{p} \sigma(\mathbf{x})$ *existe una solución* $\widehat{\mathbf{a}}(\mathbf{x})$ *de (2.17), más aún,* $\mathbf{H}\{\widehat{\mathbf{a}}(\mathbf{x}) - \mathbf{a}(\mathbf{x})\} \xrightarrow{p} 0$ *donde* $\mathbf{a}(\mathbf{x}) = (g(\mathbf{x}), \nabla^{\mathrm{T}} g(\mathbf{x}))^{\mathrm{T}}$.

Demostración. Sea $\mathbf{r} = (a_0, h_1 a_1, h_2 a_2, \cdots, h_d a_d)^{\mathrm{T}}$ y $\mathbf{r}_0 = (g(\mathbf{x}), h_1\, g_1'(x_1), h_2\, g_2'(x_2), \cdots, h_d\, g_d'(x_d))^{\mathrm{T}}$. Recordemos que $\mathcal{K}_{\mathbf{H}_d}(\mathbf{u}) = \left(\prod_{j=1}^d h_j\right)^{-1} \prod_{j=1}^d K_j(u_j/h_j)$. Notemos por

$$\begin{aligned} \ell_n(\mathbf{r}) &= \frac{1}{n}\sum_{i=1}^n \rho\left(\frac{Y_i - [a_0 + \mathbf{a}_1^{\mathrm{T}}(\mathbf{X}_i - \mathbf{x})]}{\widehat{s}(\mathbf{x})}\right)\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i - \mathbf{x})\xi_i \\ &= \frac{1}{n}\sum_{i=1}^n \rho\left(\frac{Y_i - \mathbf{r}^{\mathrm{T}}\breve{\mathbf{x}}_i}{\widehat{s}(\mathbf{x})}\right)\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i - \mathbf{x})\xi_i\,. \end{aligned} \tag{3.7}$$

Sea $\mathcal{V}_\tau(\mathbf{r}_0) = \{\mathbf{r} : \|\mathbf{r} - \mathbf{r}_0\| = \tau\}$ la esfera de centro $\mathbf{r}_0$ y de radio τ. Basta ver que dado $\nu > 0$, existe $0 < \tau_\nu < 1$ suficientemente pequeño, tal que para cualquier $0 < \tau < \tau_\nu$ si $n \geq n_0$,

$$\mathbb{P}\left(\inf_{\mathbf{r}\in\mathcal{V}_\tau(\mathbf{r}_0)} \ell_n(\mathbf{r}) > \ell_n(\mathbf{r}_0)\right) > 1 - \nu\,. \tag{3.8}$$

Efectivamente, en el conjunto $\{\inf_{\mathbf{r}\in\mathcal{V}_\tau(\mathbf{r}_0)} \ell_n(\mathbf{r}) > \ell_n(\mathbf{r}_0)\}$ se tiene que la función $\ell_n(\mathbf{r})$ tiene un mínimo local en el interior de $\mathcal{B}_\tau(\mathbf{r}_0) = \{\mathbf{r} : \|\mathbf{r} - \mathbf{r}_0\| \leq \tau\}$. Por ser mínimo local y estar en el interior de $\mathcal{B}_\tau$, es una solución de (2.17), es decir, con probabilidad mayor que $1 - \nu$, existe una solución de (2.17) en el interior de $\mathcal{B}_\tau$. De donde, dado $\nu > 0$, y τ suficientemente pequeño, si $n \geq n_0$

$$\mathbb{P}\left(\|\mathbf{H}[\hat{\mathbf{a}}(\mathbf{x}) - \mathbf{a}(\mathbf{x})]\| \leq \tau\right) > 1 - \nu\,,$$

lo que implica que $\|\mathbf{H}[\hat{\mathbf{a}}(\mathbf{x}) - \mathbf{a}(\mathbf{x})]\| \xrightarrow{p} 0$.

Para mostrar (3.8), observemos que $Y_i - \mathbf{r}^{\mathrm{T}}\breve{\mathbf{x}}_i = U_i + g(\mathbf{X}_i) - \mathbf{r}^{\mathrm{T}}\breve{\mathbf{x}}_i = U_i + R(\mathbf{X}_i) + \mathbf{r}_0^{\mathrm{T}}\breve{\mathbf{x}}_i - \mathbf{r}^{\mathrm{T}}\breve{\mathbf{x}}_i$ de donde tenemos que

$$\begin{aligned} \ell_n(\mathbf{r}) - \ell_n(\mathbf{r}_0) &= \frac{1}{n}\sum_{i=1}^n \mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i - \mathbf{x})\xi_i\left[\rho\left(\frac{U_i + R(\mathbf{X}_i) - (\mathbf{r} - \mathbf{r}_0)^{\mathrm{T}}\breve{\mathbf{x}}_i}{\widehat{s}(\mathbf{x})}\right) - \rho\left(\frac{U_i + R(\mathbf{X}_i)}{\widehat{s}(\mathbf{x})}\right)\right] \\ &= \frac{1}{n}\sum_{i=1}^n \mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i - \mathbf{x})\xi_i\left[\rho\left(\widehat{Z}_i - \widehat{\Delta}_i\right) - \rho\left(\widehat{Z}_i\right)\right]\,, \end{aligned}$$

donde $\widehat{Z}_i = (U_i + R(\mathbf{X}_i))/\widehat{s}(\mathbf{x})$ y $\widehat{\Delta}_i = (\mathbf{r} - \mathbf{r}_0)^{\mathrm{T}}\breve{\mathbf{x}}_i/\widehat{s}(\mathbf{x})$. Luego, usando que $\rho(b) - \rho(a) = \int_a^b \psi(u)du$, obtenemos que

$$\ell_n(\mathbf{r}) - \ell_n(\mathbf{r}_0) = \frac{1}{n}\sum_{i=1}^n \mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i - \mathbf{x})\xi_i \int_{\widehat{Z}_i}^{\widehat{Z}_i - \widehat{\Delta}_i} \psi\left(t\right)\, dt = K_{n1} + K_{n2} + K_{n3}\,, \tag{3.9}$$

con

$$\begin{aligned} K_{n1} &= \frac{1}{n}\sum_{i=1}^n \mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i - \mathbf{x})\xi_i \int_{\widehat{Z}_i}^{\widehat{Z}_i - \widehat{\Delta}_i} \psi(\widehat{V}_i)dt\,, \\ K_{n2} &= \frac{1}{n}\sum_{i=1}^n \mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i - \mathbf{x})\xi_i \int_{\widehat{Z}_i}^{\widehat{Z}_i - \widehat{\Delta}_i} \psi'(\widehat{V}_i)(t - \widehat{V}_i)dt\,, \\ K_{n3} &= \frac{1}{n}\sum_{i=1}^n \mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i - \mathbf{x})\xi_i \int_{\widehat{Z}_i}^{\widehat{Z}_i - \widehat{\Delta}_i} \left[\psi\left(t\right) - \psi(\widehat{V}_i) - \psi'(\widehat{V}_i)(t - \widehat{V}_i)\right]\, dt\,, \end{aligned}$$

donde $\widehat{V}_i = U_i/\widehat{s}(\mathbf{x}) = \sigma(\mathbf{X}_i)\varepsilon_i/\widehat{s}(\mathbf{x})$ y $\mathbf{r} \in \mathcal{V}_\tau(\mathbf{r}_0)$.

Sea $V_i = V_i(\mathbf{x}) = U_i/\sigma(\mathbf{x}) = \sigma(\mathbf{X}_i)\varepsilon_i/\sigma(\mathbf{x})$, luego $\widehat{V}_i = V_i\,\widehat{a}(\mathbf{x})$ con $\widehat{a}_\sigma(\mathbf{x}) = \sigma(\mathbf{x})/\widehat{s}(\mathbf{x})$. Definamos

$$\mathbf{J}_n(\mathbf{x}, a) = \frac{1}{n}\sum_{i=1}^{n} \mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i - \mathbf{x})\xi_i\psi\left(\frac{\sigma(\mathbf{X}_i)\varepsilon_i}{\sigma(\mathbf{x})}\,a\right)\breve{\mathbf{x}}_i = \frac{1}{n}\sum_{i=1}^{n} \mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i - \mathbf{x})\xi_i\psi\left(V_i\,a\right)\breve{\mathbf{x}}_i\,. \tag{3.10}$$

Observemos que

$$K_{n1} = -\frac{1}{\widehat{s}(\mathbf{x})}(\mathbf{r} - \mathbf{r}_0)^{\mathrm{T}}\frac{1}{n}\sum_{i=1}^{n} \mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i - \mathbf{x})\xi_i\psi(\widehat{V}_i)\breve{\mathbf{x}}_i = -\frac{1}{\widehat{s}(\mathbf{x})}(\mathbf{r} - \mathbf{r}_0)^{\mathrm{T}}\mathbf{J}_n(\mathbf{x}, \widehat{a}_\sigma(\mathbf{x}))\,.$$

Por **A1** y **C3** tenemos que $\mathbb{E}(\psi(V_i\,a)|\mathbf{X}_i) = 0$, para todo $a > 0$ de donde $\mathbb{E}\mathbf{J}_n(\mathbf{x}, a) = \mathbf{0}$. Por otra parte,

$$\operatorname{tr}\left(\textsc{Var}(\mathbf{J}_n(\mathbf{x}, a))\right) = \frac{1}{n}\operatorname{tr}\left(\textsc{Var}(\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_1 - \mathbf{x})\xi_1\psi\left(V_1 a\right)\breve{\mathbf{x}}_1)\right) = \frac{1}{n}\mathbb{E}(\mathcal{K}^2_{\mathbf{H}_d}(\mathbf{X}_1 - \mathbf{x})\xi_1^2\psi^2\left(V_1 a\right)\|\breve{\mathbf{x}}_1\|^2)\,.$$

Luego, tenemos

$$\begin{aligned}\operatorname{tr}\left(\textsc{Var}(\mathbf{J}_n(\mathbf{x}, a))\right) &\leq \|\psi\|^2_\infty\frac{1}{n}\mathbb{E}\left(\mathcal{K}^2_{\mathbf{H}_d}(\mathbf{X}_1 - \mathbf{x})\xi_1^2\|\breve{\mathbf{x}}_1\|^2\right) \leq \|\psi\|^2_\infty\frac{1}{n}\mathbb{E}\left(\mathcal{K}^2_{\mathbf{H}_d}(\mathbf{X}_1 - \mathbf{x})\frac{p(\mathbf{X}_1)}{t^2(\mathbf{X}_1)}\|\breve{\mathbf{x}}_1\|^2\right)\\ &\leq \|\psi\|^2_\infty\frac{1}{n}\mathbb{E}\left(\mathcal{K}^2_{\mathbf{H}_d}(\mathbf{X}_1 - \mathbf{x})\|\breve{\mathbf{x}}_1\|^2\frac{r(\mathbf{X}_1)}{t(\mathbf{X}_1)}\right) = \|\psi\|^2_\infty\frac{1}{n\prod_{j=1}^d h_j}A_n(\mathbf{x})\end{aligned}$$

donde

$$A_n(\mathbf{x}) = \mathbb{E}\left(\frac{1}{\prod_{j=1}^d h_j}\prod_{j=1}^d K_j^2\left(\frac{X_{1j} - x_j}{h_j}\right)\left[1 + \sum_{j=1}^d\frac{(X_{1j} - x_j)^2}{h_j^2}\right]\frac{r(\mathbf{X}_1)}{t(\mathbf{X}_1)}\right)\,.$$

Es fácil ver que **C2** y **A3** implican que $A_n(\mathbf{x}) \leq C$ para $n \geq n_0$. Luego, $\operatorname{tr}\left(\textsc{Var}(\mathbf{J}_n(\mathbf{x}, a))\right) = O\left(\left\{n\prod_{j=1}^d h_j\right\}^{-1}\right)$, con lo cual como $\mathbb{E}\mathbf{J}_n(\mathbf{x}, 1) = \mathbf{0}$ se tiene que la desigualdad de Chebishev implica $\mathbf{J}_n(\mathbf{x}, 1) = O_{\mathbb{P}}\left(\left\{n\prod_{j=1}^d h_j\right\}^{-\frac{1}{2}}\right)$.

Por el Lema 3.4.3

$$\widehat{\mathbf{D}}_n(\mathbf{x}) = \sqrt{n\prod_{j=1}^d h_j}\left(\widehat{s}(\mathbf{x})\mathbf{J}_n(\mathbf{x}, \widehat{a}_\sigma(\mathbf{x})) - \sigma(\mathbf{x})\mathbf{J}_n(\mathbf{x}, 1)\right) \xrightarrow{p} 0\,.$$

De esta manera, como $\widehat{s}(\mathbf{x}) \xrightarrow{p} \sigma(\mathbf{x})$ y $\sqrt{n\prod_{j=1}^d h_j}\mathbf{J}_n(\mathbf{x}, 1) = O_p(1)$, resulta que

$$\sqrt{n\prod_{j=1}^n h_j}\left(\mathbf{J}_n(\mathbf{x}, \widehat{a}_\sigma(\mathbf{x})) - \mathbf{J}_n(\mathbf{x}, 1)\right) = \frac{1}{\widehat{s}(\mathbf{x})}\left[\widehat{\mathbf{D}}_n(\mathbf{x}) + (\widehat{s}(\mathbf{x}) - \sigma(\mathbf{x}))\sqrt{n\prod_{j=1}^d h_j}\mathbf{J}_n(\mathbf{x}, 1)\right] \xrightarrow{p} 0\,.$$

Por otra parte, como $\mathbf{J}_n(\mathbf{x},1) = O_{\mathbb{P}}\left(\left\{n\prod_{j=1}^d h_j\right\}^{-\frac{1}{2}}\right)$, el hecho que $\mathbf{r} \in \mathcal{V}_\tau(\mathbf{r}_0)$ implica que $(\mathbf{r}-\mathbf{r}_0)^{\mathrm{T}}\mathbf{J}_n(\mathbf{x},1) = \tau\, O_{\mathbb{P}}\left(\left\{\sqrt{n\prod_{j=1}^d h_j}\right\}^{-1}\right)$, de donde

$$K_{n1} = \tau\,\frac{1}{\sqrt{n\prod_{j=1}^d h_j}}\,\widehat{\Upsilon}_{1,n}\,, \tag{3.11}$$

donde $\widehat{\Upsilon}_{1,n} = O_{\mathbb{P}}(1)$ no depende de $\mathbf{r}$ y por lo tanto, de τ. Recordemos que $\widehat{Z}_i = \widehat{V}_i + R(\mathbf{X}_i)\widehat{s}(\mathbf{x})^{-1}$ y $\widehat{\Delta}_i = (\mathbf{r}-\mathbf{r}_0)^{\mathrm{T}}\breve{\mathbf{x}}_i/\widehat{s}(\mathbf{x})$. Por el teorema del valor medio integral,

$$\begin{aligned} K_{n3} &= \frac{1}{n}\sum_{i=1}^n \mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i-\mathbf{x})\xi_i \int_{\widehat{V}_i+R(\mathbf{X}_i)\widehat{s}(\mathbf{x})^{-1}}^{\widehat{V}_i+R(\mathbf{X}_i)\widehat{s}(\mathbf{x})^{-1}-\widehat{\Delta}_i} \left[\psi(t)-\psi(\widehat{V}_i)-\psi'(\widehat{V}_i)(t-\widehat{V}_i)\right] dt \\ &= -(\mathbf{r}-\mathbf{r}_0)^{\mathrm{T}}\frac{1}{n\widehat{s}(\mathbf{x})}\sum_{i=1}^n \mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i-\mathbf{x})\xi_i\left[\psi\left(\widehat{V}_i+\widehat{\theta}_i\right)-\psi(\widehat{V}_i)-\psi'(\widehat{V}_i)\widehat{\theta}_i\right]\breve{\mathbf{x}}_i\,, \end{aligned}$$

donde $\widehat{\theta}_i$ es un punto intermedio entre $R(\mathbf{X}_i)/\widehat{s}(\mathbf{x})$ y $\{R(\mathbf{X}_i)-(\mathbf{r}-\mathbf{r}_0)^{\mathrm{T}}\breve{\mathbf{x}}_i\}/\widehat{s}(\mathbf{x})$. Por simplicidad, si $i \in \{i : \mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i-\mathbf{x}) = 0\}$, redefinamos $\widehat{\theta}_i = 0$ ya que no cambia el resultado de la suma. Observando que $|X_{i,j}-x_j| \le h_j$ para $i = 1,\cdots,n$ y $j = 1,\cdots,d$, cuando $\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i-\mathbf{x}) \neq 0$ resulta $\max_{i:\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i-\mathbf{x})\neq 0}|R(\mathbf{X}_i)| = \max_{1\le j\le d}\sup_{u\in\mathcal{S}_j}|g_j''(u)|\left(\sum_{j=1}^d h_j^2\right)/2 = A_g\sum_{j=1}^d h_j^2$, de donde como se cumple **A7**

$$\max_{1\le i\le n}|\widehat{\theta}_i| \le \frac{(d+1)\tau + A_g\left(\sum_{j=1}^d h_j^2\right)}{\widehat{s}(\mathbf{x})} \le \frac{(d+1)\tau + A_g\left(\sum_{j=1}^d h_j^2\right)}{A} \le A_1\left(\tau+\sum_{j=1}^d h_j^2\right),$$

donde $A_1 = \max((d+1),A_g)/A$.

Sea $\boldsymbol{\theta}_n = (\theta_1,\ldots,\theta_n)^{\mathrm{T}}$ y para cualquier $\eta > 0$ dado, definamos $D_\eta = \{\boldsymbol{\theta}_n : |\theta_i| < \eta, \forall i \le n\}$. Llamemos

$$V(\boldsymbol{\theta}_n,a) = \frac{1}{n}\sum_{i=1}^n \mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i-\mathbf{x})\xi_i\left[\psi(V_i a+\theta_i)-\psi(V_i a)-\psi'(V_i a)\,\theta_i\right]\breve{\mathbf{x}}_i\,.$$

Luego

$$\sup_{D_\eta}|V(\boldsymbol{\theta}_n,a)| \le \frac{1}{n}\sum_{i=1}^n \sup_{D_\eta}|\psi(V_i a+\theta_i)-\psi(V_i a)-\psi'(V_i a)\,\theta_i|\,\|\breve{\mathbf{x}}_i\|\,|\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i-\mathbf{x})|\,\xi_i\,.$$

Observemos que $aV_i = c_i(\mathbf{x})\varepsilon_i$ donde $c_i(\mathbf{x}) = a\sigma(\mathbf{X}_i)/\sigma(\mathbf{x})$. Como σ es una función continua y $|X_{ij}-x_j| \le h_j$ si $\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i-\mathbf{x}) \neq 0$, pues K_j tiene soporte en $[-1,1]$, tenemos que $|\sigma(\mathbf{X}_i)-\sigma(\mathbf{x})| \le i(\sigma)\gamma/4$, para todo $i = 1,\ldots,n$ tal que $\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i-\mathbf{x}) \neq 0$ y n suficientemente grande. Por lo tanto, si n es suficientemente grande y $|a-1| \le \gamma/2$, se cumple que $|c_i(\mathbf{x})-1| \le \gamma$. Es decir, si $|a-1| \le \gamma/2$ por **C3**b) tenemos que, para todo $i = 1,\ldots,n$ tal que $\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i-\mathbf{x}) \neq 0$,

$$\begin{aligned} \mathbb{E}\left(\sup_{D_\eta}|\psi(V_i a+\theta_i)-\psi(V_i a)-\psi'(V_i a)\,\theta_i|\,|\mathbf{X}_i\right) &\le \mathbb{E}\left(\sup_{|c-1|\le\gamma}\sup_{|\theta|\le\eta}|\psi(c\varepsilon+\theta)-\psi(c\varepsilon)-\psi'(c\varepsilon)\,\theta|\right) \\ &= \eta\, b_\eta\,, \end{aligned}$$

donde b_η es una sucesión de números positivos que tiende a cero cuando $\eta \to 0$. Por lo tanto, tomando $\eta = 2A_1\tau$ y usando que $r(\mathbf{x})$ es continua y acotada en $\mathcal{C}$ obtenemos

$$\mathbb{E}\left[\sup_{|a-1|\leq\gamma/2}\sup_{D_\eta}\|V(\boldsymbol{\theta}_n,a)\|\right] \leq \mathbb{E}\left[\|\breve{\mathbf{x}}_1\|\,|\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_1-\mathbf{x})|\,r(\mathbf{X}_1)\right]\eta b_\eta \leq \eta B_\eta$$

con $B_\eta = C\,b_\eta$ para alguna constante $C > 0$ independiente de η y de n. Por lo tanto, para todo $\phi > 0$

$$\mathbb{P}\left(\sup_{|a-1|\leq\gamma/2}\sup_{D_\eta}\|V(\boldsymbol{\theta}_n,a)\| > \phi\eta\right) \leq \frac{B_\eta}{\phi}\,.$$

Como $\max_{1\leq i\leq n}|\widehat{\theta}_i| \leq A_1\left(\tau + \sum_{j=1}^d h_j^2\right)$ y se cumple **C4**, podemos elegir n_τ tal que si $n \geq n_\tau$ $\sum_{j=1}^d h_j^2 \leq \tau$, de donde $\max_{1\leq i\leq n}|\widehat{\theta}_i| \leq \eta$. Como $\widehat{s}(\mathbf{x}) \xrightarrow{p} \sigma(\mathbf{x})$ se sigue que para todo $\phi > 0$

$$\begin{aligned}\mathbb{P}\left(\|V(\widehat{\boldsymbol{\theta}}_n,\widehat{a}_\sigma(\mathbf{x}))\| > \phi\eta\right) &\leq \mathbb{P}\left(|\widehat{a}_\sigma(\mathbf{x})-1| > \frac{\gamma}{2}\right) + \mathbb{P}\left(\sup_{|a-1|\leq\gamma/2}\sup_{D_\eta}\|V(\boldsymbol{\theta}_n,a)\| > \phi\eta\right)\\ &\leq \mathbb{P}\left(|\widehat{a}_\sigma(\mathbf{x})-1| > \frac{\gamma}{2}\right) + \frac{B_\eta}{\phi}\,,\end{aligned}$$

con $\widehat{a}_\sigma(\mathbf{x}) = \sigma(\mathbf{x})/\widehat{s}(\mathbf{x})$ y $\widehat{\boldsymbol{\theta}}_n = (\widehat{\theta}_1,\ldots,\widehat{\theta}_n)^{\mathrm{T}}$. De esta manera, como $K_{n3} = -(\mathbf{r}-\mathbf{r}_0)^{\mathrm{T}}V(\widehat{\boldsymbol{\theta}}_n,\widehat{a}_\sigma(\mathbf{x}))/\widehat{s}(\mathbf{x})$ y $\widehat{s}(\mathbf{x}) > A$ tenemos que

$$\mathbb{P}\left(\sup_{\mathbf{r}\in\mathcal{V}_\tau}|K_{n3}| > \tau\,\frac{\phi\,\eta}{A}\right) \leq \mathbb{P}\left(|\widehat{a}_\sigma(\mathbf{x})-1| > \frac{\gamma}{2}\right) + \frac{B_\eta}{\phi}\,, \tag{3.12}$$

Usando que $\widehat{Z}_i = \widehat{V}_i + R(\mathbf{X}_i)\widehat{s}(\mathbf{x})^{-1}$ y $\widehat{\Delta}_i = (\mathbf{r}-\mathbf{r}_0)^{\mathrm{T}}\breve{\mathbf{x}}_i/\widehat{s}(\mathbf{x})$ obtenemos que

$$\begin{aligned}K_{n2} &= \frac{1}{n}\sum_{i=1}^n \mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i-\mathbf{x})\xi_i\int_{\widehat{Z}_i}^{\widehat{Z}_i-\widehat{\Delta}_i}\psi'(\widehat{V}_i)(t-\widehat{V}_i)\\ &= \frac{1}{2\widehat{s}^2(\mathbf{x})}\frac{1}{n}\sum_{i=1}^n\psi'(\widehat{V}_i)\left[\mathbf{r}_i^{\mathrm{T}}\mathbf{r}_i - 2R(\mathbf{X}_i)\mathbf{r}_i\right]\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i-\mathbf{x})\xi_i\\ &= \frac{1}{2\widehat{s}^2(\mathbf{x})}\left((\mathbf{r}-\mathbf{r}_0)^{\mathrm{T}}\widehat{\mathbf{M}}_{n1}(\mathbf{r}-\mathbf{r}_0) - 2(\mathbf{r}-\mathbf{r}_0)^{\mathrm{T}}\widehat{\mathbf{M}}_{n2}\right)\,,\end{aligned}$$

donde $\mathbf{r}_i = (\mathbf{r}-\mathbf{r}_0)^{\mathrm{T}}\breve{\mathbf{x}}_i$, $\widehat{V}_i = \sigma(\mathbf{X}_i)\varepsilon_i/\widehat{s}(\mathbf{x})$. Definamos

$$\mathbf{M}_{n,1} = \frac{1}{n}\sum_{i=1}^n\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i-\mathbf{x})\xi_i\psi'(V_i)\breve{\mathbf{x}}_i\breve{\mathbf{x}}_i^{\mathrm{T}}\,.$$

Usando el Lema 3.4.1 (con $s(\mathbf{x}) = \sigma(\mathbf{x})$ y $\eta_i = 0$ para todo $i = 1,\cdots,n$) y recordando que $\mathbb{E}(\xi_1|\mathbf{x}_1 = \mathbf{u}) = r(\mathbf{u}) = p(\mathbf{u})/t(\mathbf{u})$ tenemos que

$$\mathbf{M}_{n1} = \mathbb{E}(\psi'(\varepsilon))\,f_{\mathbf{X}}(\mathbf{x})\,r(\mathbf{x})\,\mathbf{S}(1+o_{\mathbb{P}}(1))\,.$$

Por otra parte, como ψ' y g_j'' son acotadas, $r(\mathbf{u})$ es acotada en un entorno de $\mathbf{x}$ y K_j tiene soporte compacto $[-1,1]$ implican que

$$\widehat{\mathbf{M}}_{n,2} = \frac{1}{n}\sum_{i=1}^{n} \mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i - \mathbf{x})\xi_i\psi'(\widehat{V}_i)R(\mathbf{X}_i)\check{\mathbf{x}}_i = O_{\mathbb{P}}\left(\sum_{j=1}^{d} h_j^2\right) .$$

De esta manera, usando la consistencia de $\widehat{s}(\mathbf{x})$ y el hecho que $\widehat{\mathbf{M}}_{n,1} - \mathbf{M}_{n,1} = o_{\mathbb{P}}(1)$ (ver Lema 3.4.2) obtenemos que

$$K_{n2} = \frac{1}{2\sigma^2(\mathbf{x})}\mathbb{E}(\psi'(\varepsilon))f_{\mathbf{X}}(\mathbf{x})\frac{p(\mathbf{x})}{t(\mathbf{x})}(\mathbf{r}-\mathbf{r}_0)^{\mathrm{T}}\mathbf{S}(\mathbf{r}-\mathbf{r}_0)(1+\widehat{\zeta}_n) + \tau\sum_{j=1}^{d} h_j^2\,\widehat{\Upsilon}_{2,n} ,$$

donde $\widehat{\zeta}_n = o_{\mathbb{P}}(1)$ y $\widehat{\Upsilon}_{2,n} = O_{\mathbb{P}}(1)$ no dependen de $\mathbf{r}$ y por lo tanto, de τ. Como $\mathbf{S}$ es una matriz diagonal definida positiva, llamemos λ al menor elemento diagonal de $\mathbf{S}$. Luego, como $\mathbb{E}(\psi'(\varepsilon)) > 0$ para cualquier $\mathbf{r} \in \mathcal{V}_\tau$ se cumple

$$Q(\mathbf{r}) = \frac{1}{2\sigma^2(\mathbf{x})}\mathbb{E}(\psi'(\varepsilon))f_{\mathbf{X}}(\mathbf{x})r(\mathbf{x})(\mathbf{r}-\mathbf{r}_0)^{\mathrm{T}}\mathbf{S}(\mathbf{r}-\mathbf{r}_0) \geq \frac{\lambda}{2\sigma(\mathbf{x})}\mathbb{E}(\psi'(\varepsilon))f_{\mathbf{X}}(\mathbf{x})r(\mathbf{x})\tau^2 = M(\mathbf{x})\tau^2 > 0 .$$

Luego, como $\widehat{\Upsilon}_{2,n} = O_{\mathbb{P}}(1)$, se tiene que existe $\widetilde{A}_1$ y n_1 tal que si $n \geq n_1$

$$\mathbb{P}\left(|\widehat{\Upsilon}_{2,n}| \leq \widetilde{A}_1\right) > 1 - \frac{\nu}{8} \qquad\qquad \mathbb{P}\left(|\widehat{\zeta}_n| \leq \frac{1}{2}\right) > 1 - \frac{\nu}{8}$$

Sea n_τ tal que si $n \geq n_\tau$, $\sum_{j=1}^{d} h_j^2 \leq \tau \min\{M(\mathbf{x})/(4\widetilde{A}_1), 1\}$. Por lo tanto, como $K_{n2} \geq Q(\mathbf{r})(1 - |\widehat{\zeta}_n|) - \tau\sum_{j=1}^{d} h_j^2\,|\widehat{\Upsilon}_{2,n}|$, si $n \geq n_{1,\tau} = \max(n_\tau, n_1)$ tenemos que

$$\mathbb{P}\left(\inf_{\mathbf{r}\in\mathcal{V}_\tau} K_{n2} > \frac{M(\mathbf{x})}{4}\tau^2\right) > 1 - \frac{\nu}{4} . \tag{3.13}$$

Recordemos que por (3.11) $K_{n1} = \tau\left\{n\prod_{j=1}^{d} h_j\right\}^{-1/2}\widehat{\Upsilon}_{1,n}$ de donde, usando que $n\prod_{j=1}^{d} h_j \to \infty$ se deduce que si $n \geq n_{2,\tau}$,

$$\frac{1}{\sqrt{n\prod_{j=1}^{d} h_j}} \leq \tau\frac{M(\mathbf{x})}{8\widetilde{A}_2} \qquad \text{donde } \widetilde{A}_2 \text{ es tal que} \qquad \mathbb{P}\left(|\widehat{\Upsilon}_{1,n}| \leq \widetilde{A}_2\right) > 1 - \frac{\nu}{4} ,$$

de donde usando (3.13), obtenemos que si $n \geq \max(n_{1,\tau}, n_{2,\tau})$

$$\mathbb{P}\left(\inf_{\mathbf{r}\in\mathcal{V}_\tau}(K_{n1} + K_{n2}) > \frac{M(\mathbf{x})}{8}\tau^2\right) > 1 - \frac{\nu}{2} . \tag{3.14}$$

Recordemos que por (3.12) como $\eta = 2A_1\tau$,

$$\mathbb{P}\left(\sup_{\mathbf{r}\in\mathcal{V}_\tau}|K_{n3}| > \tau^2\frac{2\phi\,A_1}{A}\right) \leq \mathbb{P}\left(|\widehat{a}_\sigma(\mathbf{x}) - 1| > \frac{\gamma}{2}\right) + \frac{B_\eta}{\phi} ,$$

donde $\lim_{\eta\to 0} B_\eta = 0$. Luego, si tomamos $\phi = M(\mathbf{x})A/(32\,A_1) > 0$ obtenemos que

$$\mathbb{P}\left(\sup_{\mathbf{r}\in\mathcal{V}_\tau} |K_{n3}| > \frac{M(\mathbf{x})}{16}\tau^2\right) \le \mathbb{P}\left(|\widehat{a}_\sigma(\mathbf{x}) - 1| > \frac{\gamma}{2}\right) + B_\eta \frac{32\,A_1}{M(\mathbf{x})A}\,,$$

Como $\widehat{s}(\mathbf{x}) \xrightarrow{p} \sigma(\mathbf{x})$ tenemos que si $n \ge n_3$,

$$\mathbb{P}\left(|\widehat{a}_\sigma(\mathbf{x}) - 1| > \frac{\gamma}{2}\right) \le \frac{\nu}{4}$$

Por otra parte, como $\lim_{\eta\to 0} B_\eta = 0$, si $0 < \tau < \tau_\nu$, $32\,B_\eta\,A_1/(M(\mathbf{x})A) \le \nu/4$. Luego, para todo $0 < \tau < \tau_\nu$, si $n \ge \max(n_{1,\tau}, n_{2,\tau}, n_3)$, se cumple

$$\mathbb{P}\left(\sup_{\mathbf{r}\in\mathcal{V}_\tau} |K_{n3}| \le \frac{M(\mathbf{x})}{16}\tau^2\right) \ge 1 - \frac{\nu}{2}$$

de donde usando (3.9) y (3.14) deducimos (3.8). □

Observemos que la Proposición 3.2.1 permite mostrar que el estimador inicial $\widetilde{g}_{\mathrm{M}_1}$ de g es consistente. Por otra parte, muestra que $h_j(\widehat{\mathbf{a}}_{1,j}(\mathbf{x}) - g_j'(x_j)) \xrightarrow{p} 0$. Para obtener resultados de consistencia sobre $\widehat{g}_{\alpha,\mathrm{M}_1}(x)$ necesitaremos mostrar resultados de convergencia uniforme sobre compactos para $\widetilde{g}_{\mathrm{M}_1}$. Necesitaremos las siguiente condiciones adicionales sobre la función de escala $\sigma(\cdot)$, su estimador $\widehat{s}(\cdot)$ y sobre la función de escores ψ.

C6 El estimador de escala $\widehat{s}(\cdot)$ satisface que $\sup_{\mathbf{x}\in\mathcal{C}} |\widehat{s}(\mathbf{x}) - \sigma(\mathbf{x})| \xrightarrow{a.s.} 0$.

C7 La función ψ es dos veces continuamente diferenciable con derivada ψ' tal que $\mathbb{E}(\psi'(\varepsilon)) > 0$ y $\zeta(u) = u\psi'(u)$ es acotada. Más aún, la derivada segunda ψ'' es acotada y tal que $\zeta_2(u) = u\psi''(u)$ es acotada.

C8 Las funciones $K_j : [-1, 1] \to \mathbb{R}$ cumplen una condición de Lipschitz de orden 1.

Observación 3.2.1. Está claro que **C6** y **A4** implican que **A7** se cumple, es decir, existen constantes A y B tales que

$$\mathbb{P}\left(\exists\, n_0 \text{ tal que para todo } n \ge n_0 \text{ y para todo } \mathbf{x} \in \mathcal{C} \quad A < \widehat{s}(\mathbf{x}) < B\right) = 1\,. \tag{3.15}$$

Efectivamente, basta tomar $A = \inf_{\mathbf{x}\in\mathcal{C}} \sigma(\mathbf{x})/2$ y $B = (3/2)\sup_{\mathbf{x}\in\mathcal{C}} \sigma(\mathbf{x})$.

Por otra parte, si $\widehat{a}_\sigma(\mathbf{x}) = \sigma(\mathbf{x})/\widehat{s}(\mathbf{x})$ y valen **A4** y **C6**, entonces

$$\sup_{\mathbf{x}\in\mathcal{C}} |\widehat{a}_\sigma(\mathbf{x}) - 1| \xrightarrow{a.s.} 0\,, \tag{3.16}$$

ya que

$$|\widehat{a}_\sigma(\mathbf{x}) - 1| = \frac{|\sigma(\mathbf{x}) - \widehat{s}(\mathbf{x})|}{|\widehat{s}(\mathbf{x}) - \sigma(\mathbf{x}) + \sigma(\mathbf{x})|} \le \frac{\sup_{\mathbf{x}\in\mathcal{C}} |\sigma(\mathbf{x}) - \widehat{s}(\mathbf{x})|}{\sigma(\mathbf{x}) - \sup_{\mathbf{x}\in\mathcal{C}} |\widehat{s}(\mathbf{x}) - \sigma(\mathbf{x})|} \le \frac{\sup_{\mathbf{x}\in\mathcal{C}} |\sigma(\mathbf{x}) - \widehat{s}(\mathbf{x})|}{i(\sigma) - \sup_{\mathbf{x}\in\mathcal{C}} |\sigma(\mathbf{x}) - \widehat{s}(\mathbf{x})|}\,.$$

De ahora en más, indicaremos por $A_h = 1/\min_{1\le j\le d} h_j$ y para reforzar la dependencia en $\mathbf{x}$ indicaremos por $R(\mathbf{X}_i, \mathbf{x}) = g(\mathbf{X}_i) - g(\mathbf{x}) - \sum_{j=1}^d g_j'(x_j)(X_{i,j} - x_j)$.

Proposición 3.2.2. *Sea $\mathcal{C} \subset \mathcal{S}_f$ un conjunto compacto para el que cumple **A3**. Supongamos que se cumplen **A1**, **A4**, **A6**, **C1** a **C3**a) y **C5** a **C8**. Entonces, existe una solución $\widehat{\mathbf{a}}(\mathbf{x})$ de (2.17) tal que $\sup_{\mathbf{x}\in\mathcal{C}} |\mathbf{H}\{\widehat{\mathbf{a}}(\mathbf{x}) - \mathbf{a}(\mathbf{x})\}| \xrightarrow{a.s.} 0$ donde $\mathbf{a}(\mathbf{x}) = (g(\mathbf{x}), \nabla^{\mathrm{T}} g(\mathbf{x}))^{\mathrm{T}}$.*

Demostración. La demostración sigue pasos análogos a los dados en la Proposición 3.2.1, pero tratando de obtener resultados uniformes.

Sea $\mathbf{r} = (a_0, h_1 a_1, h_2 a_2, \cdots, h_d a_d)^{\mathrm{T}}$ y $\mathbf{r}_0(\mathbf{x}) = (g(\mathbf{x}), h_1\, g_1'(x_1), h_2\, g_2'(x_2), \cdots, h_d\, g_d'(x_d))^{\mathrm{T}}$. Recordemos que $\mathcal{K}_{\mathbf{H}_d}(\mathbf{u}) = \left(\prod_{j=1}^d h_j\right)^{-1} \prod_{j=1}^d K_j(u_j/h_j)$ y sea $\ell_n(\mathbf{r})$ definida en (3.7). Sea $\mathcal{V}_\tau = \{\mathbf{r} : \|\mathbf{r}\| = \tau\}$ la esfera de centro 0 y de radio τ. Basta ver que existe $\mathcal{N}$ tal que $\mathbb{P}(\mathcal{N}) = 0$ y tal que para todo $\omega \notin \mathcal{N}$ se cumple que dado $\nu > 0$, existe $0 < \tau_\nu < 1$ suficientemente pequeño, tal que para cualquier $0 < \tau < \tau_\nu$ si $n \geq n_0$,

$$\inf_{\mathbf{r}\in\mathcal{V}_\tau} \inf_{\mathbf{x}\in\mathcal{C}} \{\ell_n(\mathbf{r} + \mathbf{r}_0(\mathbf{x})) - \ell_n(\mathbf{r}_0(\mathbf{x}))\} > 0\,. \tag{3.17}$$

Efectivamente, en el conjunto $\{\inf_{\mathbf{r}\in\mathcal{V}_\tau} \inf_{\mathbf{x}\in\mathcal{C}} [\ell_n(\mathbf{r} + \mathbf{r}_0(\mathbf{x})) - \ell_n(\mathbf{r}_0(\mathbf{x}))] > 0\}$ se tiene que para todo $\mathbf{x} \in \mathcal{C}$, la función $L_n(\mathbf{r}) = \ell_n(\mathbf{r}+\mathbf{r}_0(\mathbf{x})) - \ell_n(\mathbf{r}_0(\mathbf{x}))$ tiene un mínimo local $\widetilde{\mathbf{r}}(\mathbf{x})$ en el interior de $\mathcal{V}_\tau$ pues $\inf_{\mathbf{r}\in\mathcal{V}_\tau} \ell_n(\mathbf{r}+\mathbf{r}_0(\mathbf{x})) > \ell_n(\mathbf{r}_0(\mathbf{x}))$. Observemos que entonces, para todo $\mathbf{x} \in \mathcal{C}$, $\widetilde{\mathbf{r}}(\mathbf{x}) + \mathbf{r}_0(\mathbf{x})$ es mínimo local de $\ell_n(\mathbf{r})$ y $\widetilde{\mathbf{r}}(\mathbf{x}) + \mathbf{r}_0(\mathbf{x})$ está en el interior de $\mathcal{B}_\tau(\mathbf{r}_0(\mathbf{x})) = \{\mathbf{r} : \|\mathbf{r} - \mathbf{r}_0(\mathbf{x})\| \leq \tau\}$, con lo cual $\widehat{\mathbf{a}}(\mathbf{x}) = \widetilde{\mathbf{r}}(\mathbf{x}) + \mathbf{r}_0(\mathbf{x})$ es una solución de (2.17). Es decir, con probabilidad 1, para todo $\mathbf{x} \in \mathcal{C}$, existe una solución $\widehat{\mathbf{a}}(\mathbf{x})$ de (2.17) en el interior de $\mathcal{B}_\tau(\mathbf{r}_0(\mathbf{x}))$. De donde, para $\omega \notin \mathcal{N}$, dado $\nu > 0$, y τ suficientemente pequeño, si $n \geq n_0$

$$\sup_{\mathbf{x}\in\mathcal{C}} \|\mathbf{H}[\widehat{\mathbf{a}}(\mathbf{x}) - \mathbf{a}(\mathbf{x})]\| \leq \tau\,,$$

lo que implica que $\sup_{\mathbf{x}\in\mathcal{C}} \|\mathbf{H}[\widehat{\mathbf{a}}(\mathbf{x}) - \mathbf{a}(\mathbf{x})]\| \xrightarrow{a.s.} 0$.

Para mostrar (3.17), observemos que $Y_i - (\mathbf{r} + \mathbf{r}_0(\mathbf{x}))^{\mathrm{T}}\breve{\mathbf{x}}_i = U_i + g(\mathbf{X}_i) - \mathbf{r}^{\mathrm{T}}\breve{\mathbf{x}}_i - \mathbf{r}_0^{\mathrm{T}}\breve{\mathbf{x}}_i = U_i + R(\mathbf{X}_i, \mathbf{x})\breve{\mathbf{x}}_i - \mathbf{r}^{\mathrm{T}}\breve{\mathbf{x}}_i$, por lo tanto, tenemos que

$$\begin{aligned}
\ell_n(\mathbf{r} + \mathbf{r}_0(\mathbf{x})) - \ell_n(\mathbf{r}_0(\mathbf{x})) &= \frac{1}{n}\sum_{i=1}^n \mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i - \mathbf{x})\xi_i \left[\rho\left(\frac{U_i + R(\mathbf{X}_i, \mathbf{x}) - \mathbf{r}^{\mathrm{T}}\breve{\mathbf{x}}_i}{\widehat{s}(\mathbf{x})}\right) - \rho\left(\frac{U_i + R(\mathbf{X}_i, \mathbf{x})}{\widehat{s}(\mathbf{x})}\right)\right] \\
&= \frac{1}{n}\sum_{i=1}^n \mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i - \mathbf{x})\xi_i \left[\rho\left(\widehat{Z}_i(\mathbf{x}) - \widehat{\Delta}_i(\mathbf{x})\right) - \rho\left(\widehat{Z}_i(\mathbf{x})\right)\right]\,,
\end{aligned}$$

donde $\widehat{Z}_i(\mathbf{x}) = \widehat{V}_i(\mathbf{x}) + R(\mathbf{X}_i, \mathbf{x})\widehat{s}(\mathbf{x})^{-1}$ con $\widehat{V}_i(\mathbf{x}) = U_i/\widehat{s}(\mathbf{x}) = \sigma(\mathbf{X}_i)\varepsilon_i/\widehat{s}(\mathbf{x})$ y $\widehat{\Delta}_i(\mathbf{x}) = \mathbf{r}^{\mathrm{T}}\breve{\mathbf{x}}_i/\widehat{s}(\mathbf{x})$. Luego, usando que $\rho(b) - \rho(a) = \int_a^b \psi(u)du$, obtenemos que $\rho(\widehat{Z}_i(\mathbf{x}) - \widehat{\Delta}_i(\mathbf{x})) - \rho(\widehat{Z}_i(\mathbf{x})) = \int_{\widehat{Z}_i(\mathbf{x})}^{\widehat{Z}_i(\mathbf{x})-\widehat{\Delta}_i(\mathbf{x})} \psi\,(t)\; dt$. Por lo tanto, como en la Proposición 3.2.1, tenemos que para todo $\mathbf{r} \in \mathcal{V}_\tau$

$$\ell_n(\mathbf{r} + \mathbf{r}_0(\mathbf{x})) - \ell_n(\mathbf{r}_0(\mathbf{x})) = K_{n1}(\mathbf{x}) + K_{n2}(\mathbf{x}) + K_{n3}(\mathbf{x})\,, \tag{3.18}$$

donde

$$
\begin{aligned}
K_{n1}(\mathbf{x}) &= \frac{1}{n}\sum_{i=1}^{n}\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i-\mathbf{x})\xi_i\int_{\widehat{Z}_i(\mathbf{x})}^{\widehat{Z}_i(\mathbf{x})-\widehat{\Delta}_i(\mathbf{x})}\psi(\widehat{V}_i)dt = -\frac{1}{\widehat{s}(\mathbf{x})}\mathbf{r}^{\mathrm{T}}\frac{1}{n}\sum_{i=1}^{n}\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i-\mathbf{x})\xi_i\psi(\widehat{V}_i(\mathbf{x}))\breve{\mathbf{x}}_i\ ,\\
K_{n2}(\mathbf{x}) &= \frac{1}{n}\sum_{i=1}^{n}\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i-\mathbf{x})\xi_i\int_{\widehat{Z}_i(\mathbf{x})}^{\widehat{Z}_i(\mathbf{x})-\widehat{\Delta}_i(\mathbf{x})}\psi'(\widehat{V}_i(\mathbf{x}))(t-\widehat{V}_i(\mathbf{x}))dt\\
&= \frac{1}{2\widehat{s}^2(\mathbf{x})}\frac{1}{n}\sum_{i=1}^{n}\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i-\mathbf{x})\xi_i\psi'(\widehat{V}_i(\mathbf{x}))\left[\left(\mathbf{r}^{\mathrm{T}}\breve{\mathbf{x}}_i\right)^2-2R(\mathbf{X}_i,\mathbf{x})\mathbf{r}^{\mathrm{T}}\breve{\mathbf{x}}_i\right]\ ,\\
K_{n3}(\mathbf{x}) &= \frac{1}{n}\sum_{i=1}^{n}\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i-\mathbf{x})\xi_i\int_{\widehat{Z}_i(\mathbf{x})}^{\widehat{Z}_i(\mathbf{x})-\widehat{\Delta}_i(\mathbf{x})}\left[\psi(t)-\psi(\widehat{V}_i(\mathbf{x}))-\psi'(\widehat{V}_i(\mathbf{x}))(t-\widehat{V}_i(\mathbf{x}))\right]dt\ .
\end{aligned}
$$

Supongamos haber probado que

$$
\sup_{\mathbf{r}\in\mathcal{V}_\tau}\sup_{\mathbf{x}\in\mathcal{C}}\|K_{n1}(\mathbf{x})\| = \tau\sqrt{\frac{\log n}{n\prod_{j=1}^{d}h_j}}\,\widehat{\Upsilon}_{1,n}\ , \tag{3.19}
$$

$$
K_{n2}(\mathbf{x}) = \frac{1}{2\sigma^2(\mathbf{x})}\mathbb{E}(\psi'(\varepsilon))f_{\mathbf{X}}(\mathbf{x})r(\mathbf{x})\ \mathbf{r}^{\mathrm{T}}\mathbf{S}\mathbf{r}(1+\widehat{\zeta}_n)+\tau\sum_{j=1}^{d}h_j^2\,\widehat{\Upsilon}_{2,n}\,, \tag{3.20}
$$

$$
\sup_{\mathbf{r}\in\mathcal{V}_\tau}\sup_{\mathbf{x}\in\mathcal{C}}\|\mathcal{K}_{n3}(\mathbf{x})\| = \tau\left(\tau+\sum_{j=1}^{d}h_j^2\right)^2\widehat{\Upsilon}_{3,n}\ . \tag{3.21}
$$

con $\widehat{\Upsilon}_{1,n}=O_{\text{a.co.}}(1)$, $\widehat{\Upsilon}_{3,n}=O_{\text{a.co.}}(1)$ y además $\widehat{\Upsilon}_{1,n}$ y $\widehat{\Upsilon}_{3,n}$ no dependen de τ y donde además $\widehat{\zeta}_n=o_{\text{a.co.}}(1)$ y $\widehat{\Upsilon}_{2,n}=O_{\text{a.co.}}(1)$ no dependen de $\mathbf{x}\in\mathcal{C}$ ni de $\mathbf{r}$ y por lo tanto, de τ. Veremos que esto permite obtener (3.17).

Efectivamente, como $\mathbf{S}$ es una matriz diagonal definida positiva, sea $\lambda>0$ el menor elemento de $\mathbf{S}$. Luego, como $\mathbb{E}\psi'(\varepsilon)>0$, $i(f_{\mathbf{X}})>0$, $i(r)>0$ y σ está acotada sobre $\mathbf{x}\in\mathcal{C}$, si $M=\lambda\mathbb{E}\psi'(\varepsilon)i(f_{\mathbf{X}})\,i(r)/(2\sup_{\mathbf{x}\in\mathcal{C}}\sigma^2(\mathbf{x}))$ tenemos que para todo $\mathbf{x}\in\mathcal{C}$ y $\mathbf{r}\in\mathcal{V}_\tau$,

$$
Q(\mathbf{r},\mathbf{x})=\frac{1}{2\sigma^2(\mathbf{x})}\mathbb{E}\psi'(\varepsilon)f_{\mathbf{X}}(\mathbf{x})r(\mathbf{x})\,\mathbf{r}^{\mathrm{T}}\mathbf{S}\mathbf{r}\geq\frac{\lambda}{2\sigma^2(\mathbf{x})}\mathbb{E}\psi'(\varepsilon)f_{\mathbf{X}}(\mathbf{x})r(\mathbf{x})\tau^2\geq M\tau^2>0\,.
$$

Como $\widehat{\Upsilon}_{2,n}=O_{\text{a.co.}}(1)$ y $\widehat{\zeta}_n=o_{\text{a.co.}}(1)$, dado $\nu>0$ existe $\widetilde{A}_1$ tal que

$$
\sum_{n\geq1}\mathbb{P}\left(|\widehat{\Upsilon}_{2,n}|>\widetilde{A}_1\right)<\infty \qquad\qquad \sum_{n\geq1}\mathbb{P}\left(|\widehat{\zeta}_n|>\frac{1}{2}\right)<\infty\,.
$$

Sea n_τ tal que si $n\geq n_\tau$, $\sum_{j=1}^{d}h_j^2\leq\tau\min\{M/(4\widetilde{A}_1),1\}$. Por lo tanto, existe $\mathcal{N}_1$ tal que $\mathbb{P}(\mathcal{N}_1)=0$ y para todo $\omega\notin\mathcal{N}_1$, existe $n_1\in\mathbb{N}$ tal que para todo $n\geq n_1$

$$
|\widehat{\Upsilon}_{2,n}|<\widetilde{A}_1 \qquad\qquad |\widehat{\zeta}_n|<\frac{1}{2}\,.
$$

Como $K_{n2}(\mathbf{x})\geq Q(\mathbf{r},\mathbf{x})(1-|\widehat{\zeta}_n|)-\tau\sum_{j=1}^{d}h_j^2\,|\widehat{\Upsilon}_{2,n}|$, si $\omega\notin\mathcal{N}_1$ y $n\geq n_{1,\tau}=\max(n_\tau,n_1)$ tenemos que

$$
\inf_{\mathbf{r}\in\mathcal{V}_\tau}\inf_{\mathbf{x}\in\mathcal{C}}K_{n2}(\mathbf{x})>\frac{M}{4}\tau^2\ . \tag{3.22}
$$

Por (3.19) y dado que $n\prod_{j=1}^d h_j/\log n \to \infty$ se deduce que existe $\widetilde{A}_2$ y $\mathcal{N}_2$ tal que $\mathbb{P}(\mathcal{N}_2)=0$ y para todo $\omega \notin \mathcal{N}_2$ existe $n_{2,\tau}$ tal que si $n \geq n_{2,\tau}$,

$$|\widehat{\Upsilon}_{1,n}| \leq \widetilde{A}_2 \qquad \text{y} \qquad \sqrt{\frac{\log n}{n\prod_{j=1}^d h_j}} \leq \tau \min\left\{\frac{M}{8\widetilde{A}_2}, 1\right\} ,$$

de donde usando (3.22), obtenemos que si $\omega \notin \mathcal{N}_1 \cup \mathcal{N}_2$ y $n \geq \max(n_{1,\tau}, n_{2,\tau})$

$$\inf_{\mathbf{r}\in\mathcal{V}_\tau} \inf_{\mathbf{x}\in\mathcal{C}} (K_{n1}(\mathbf{x}) + K_{n2}(\mathbf{x})) > \frac{M}{8}\tau^2 . \tag{3.23}$$

Por otra parte, K_{n3} cumple (3.21) con $\widehat{\Upsilon}_{3,n} = O_{\text{a.co.}}(1)$, luego existe $\widetilde{A}_3$ y $\mathcal{N}_3$ tal que $\mathbb{P}(\mathcal{N}_3)=0$ y para todo $\omega \notin \mathcal{N}_3$ existe n_3 tal que si $n \geq n_3$, $|\widehat{\Upsilon}_{3,n}| \leq \widetilde{A}_3$. Además si $n \geq n_{3,\tau}$, $\sum_{j=1}^d h_j^2 \leq \tau$, de donde se tiene que para todo $\omega \notin \mathcal{N}_3$ y $n \geq n_{4,\tau} = \max\{n_{3,\tau}, n_3\}$, $\sup_{\mathbf{r}\in\mathcal{V}_\tau} \sup_{\mathbf{x}\in\mathcal{C}} |K_{n3}(\mathbf{x})| \leq \widetilde{A}_3\tau^3$. Tomando $\tau_\nu < \min\{1, M/(16\,\widetilde{A}_3)\}$, obtenemos que si $\tau < \tau_\nu$, $\omega \notin \mathcal{N}_3$ y $n \geq n_{4,\tau}$

$$\sup_{\mathbf{r}\in\mathcal{V}_\tau} \sup_{\mathbf{x}\in\mathcal{C}} |K_{n3}(\mathbf{x})| \leq \frac{M}{16}\tau^2 . \tag{3.24}$$

Luego, para todo $0 < \tau < \tau_\nu < 1$, si $\omega \notin \mathcal{N}_3$ y $n \geq \max(n_{1,\tau}, n_{2,\tau}, n_{4,\tau})$, se cumplen (3.24) y (3.23), de donde conjuntamente con (3.18), deducimos (3.17).

Por lo tanto, la demostración estará completa si probamos (3.19), (3.20) y (3.21).

• Comenzaremos probando (3.19). Sea $V_i(\mathbf{x}) = U_i/\sigma(\mathbf{x}) = \sigma(\mathbf{X}_i)\varepsilon_i/\sigma(\mathbf{x})$, luego $\widehat{V}_i(\mathbf{x}) = V_i(\mathbf{x})\,\widehat{a}(\mathbf{x})$ con $\widehat{a}_\sigma(\mathbf{x}) = \sigma(\mathbf{x})/\widehat{s}(\mathbf{x})$. Definamos $\mathbf{J}_n(\mathbf{x}, a)$ como en (3.10) y observemos que

$$K_{n1}(\mathbf{x}) = -\frac{1}{\widehat{s}(\mathbf{x})}\mathbf{r}^{\mathrm{T}}\mathbf{J}_n(\mathbf{x}, \widehat{a}_\sigma(\mathbf{x})) .$$

Por (3.15) tenemos que

$$\mathbb{P}\left(\exists n_0 \text{ tal que } \forall n \geq n_0 \sup_{\mathbf{r}\in\mathcal{V}_\tau} \sup_{\mathbf{x}\in\mathcal{C}} \|K_{n1}(\mathbf{x})\| < \frac{\tau}{A} \sup_{\mathbf{x}\in\mathcal{C}} \|\mathbf{J}_n(\mathbf{x}, \widehat{a}_\sigma(\mathbf{x}))\|\right) = 1 .$$

Por el Lema 3.4.7, para la $j-$ésima coordenada, $J_{n,j}(\mathbf{x}, \widehat{a}_\sigma(\mathbf{x}))$, de $\mathbf{J}_n(\mathbf{x}, \widehat{a}_\sigma(\mathbf{x}))$ se tiene que $\sup_{\mathbf{x}\in\mathcal{C}} |J_{n,j}(\mathbf{x}, \widehat{a}_\sigma(\mathbf{x}))| = O_{\text{a.s.}}(\theta_n)$ con $\theta_n = \left(\log n/(n\prod_{j=1}^d h_j)\right)^{1/2}$, y por lo tanto $\sup_{\mathbf{x}\in\mathcal{C}} \|\mathbf{J}_n(\mathbf{x}, \widehat{a}_\sigma(\mathbf{x}))\| = O_{\text{a.s.}}(\theta_n)$. Luego, hemos probado (3.19).

• Probaremos ahora (3.21). Recordemos que $\widehat{Z}_i(\mathbf{x}) = \widehat{V}_i(\mathbf{x}) + R(\mathbf{X}_i, \mathbf{x})\widehat{s}(\mathbf{x})^{-1}$ y $\widehat{\Delta}_i(\mathbf{x}) = \mathbf{r}^{\mathrm{T}}\breve{\mathbf{x}}_i/\widehat{s}(\mathbf{x})$. Por el teorema del valor medio integral,

$$\begin{aligned} K_{n3}(\mathbf{x}) &= \frac{1}{n}\sum_{i=1}^n \mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i - \mathbf{x})\xi_i \int_{\widehat{V}_i(\mathbf{x})+R(\mathbf{X}_i,\mathbf{x})\widehat{s}(\mathbf{x})^{-1}}^{\widehat{V}_i(\mathbf{x})+R(\mathbf{X}_i,\mathbf{x})\widehat{s}(\mathbf{x})^{-1}-\widehat{\Delta}_i(\mathbf{x})} \left[\psi(t) - \psi(\widehat{V}_i(\mathbf{x})) - \psi'(\widehat{V}_i(\mathbf{x}))(t - \widehat{V}_i(\mathbf{x}))\right] dt \\ &= -\mathbf{r}^{\mathrm{T}}\frac{1}{n\widehat{s}(\mathbf{x})}\sum_{i=1}^n \mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i - \mathbf{x})\xi_i \left[\psi\left(\widehat{V}_i(\mathbf{x}) + \widehat{\theta}_i(\mathbf{x})\right) - \psi(\widehat{V}_i(\mathbf{x})) - \psi'(\widehat{V}_i(\mathbf{x}))\widehat{\theta}_i(\mathbf{x})\right]\breve{\mathbf{x}}_i , \end{aligned}$$

donde $\widehat{\theta}_i(\mathbf{x})$ es un punto intermedio entre $R(\mathbf{X}_i,\mathbf{x})/\widehat{s}(\mathbf{x})$ y $\{R(\mathbf{X}_i,\mathbf{x})-\mathbf{r}^{\mathrm{T}}\breve{\mathbf{x}}_i\}/\widehat{s}(\mathbf{x})$. Por simplicidad, si $i\in\{i:\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i-\mathbf{x})=0\}$, redefinamos $\widehat{\theta}_i(\mathbf{x})=0$ ya que no cambia el resultado de la suma. Observando que $|X_{i,j}-x_j|\le h_j$ para $i=1,\cdots,n$ y $j=1,\cdots,d$, cuando $\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i-\mathbf{x})\neq 0$ y utilizando que g_j es dos veces continuamente diferenciable y que g_j'' es acotada en $\mathcal{S}_j$, resulta $\max_{i:\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i-\mathbf{x})\neq 0}|R(\mathbf{X}_i,\mathbf{x})|=\max_{1\le j\le d}\sup_{u\in\mathcal{S}_j}|g_j''(u)|\left(\sum_{j=1}^d h_j^2\right)/2=A_g\sum_{j=1}^d h_j^2$. Por lo tanto,

$$\sup_{\mathbf{x}\in\mathcal{C}}\max_{i:\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i-\mathbf{x})\neq 0}|R(\mathbf{X}_i,\mathbf{x})|\le A_g\sum_{j=1}^d h_j^2\,, \tag{3.25}$$

de donde si $A<\widehat{s}(\mathbf{x})$ y $A_1=\max((d+1),A_g)/A$ tenemos que

$$\max_{1\le i\le n}|\widehat{\theta}_i(\mathbf{x})|\le\frac{(d+1)\tau+A_g\left(\sum_{j=1}^d h_j^2\right)}{\widehat{s}(\mathbf{x})}\le\frac{(d+1)\tau+A_g\left(\sum_{j=1}^d h_j^2\right)}{A}\le A_1\left(\tau+\sum_{j=1}^d h_j^2\right).$$

Es decir, de (3.15) deducimos que

$$\mathbb{P}\left(\exists n_0 \text{ tal que } \forall n\ge n_0\quad \sup_{\mathbf{x}\in\mathcal{C}}\max_{1\le i\le n}|\widehat{\theta}_i(\mathbf{x})|\le A_1\left(\tau+\sum_{j=1}^d h_j^2\right)\right)=1\,.$$

Por la Observación 3.4.1, si $\mathcal{K}^\star(\mathbf{u})=|\mathcal{K}(\mathbf{u})|/\int|\mathcal{K}(\mathbf{u})|d\mathbf{u}$ el estimador de la densidad basado en $\mathcal{K}^\star$, $\widehat{f}(\mathbf{x})=(1/n)\sum_{j=1}^n\mathcal{K}^\star_{\mathbf{H}_d}(\mathbf{x}-\mathbf{X}_j)$, converge a $f_{\mathbf{X}}$ uniformemente en casi todo punto (ver (3.60)). Por lo tanto, por **A3** obtenemos que $\sup_{\mathbf{x}\in\mathcal{C}}\widehat{f}(\mathbf{x})=O_{\text{a.s.}}(1)$, de donde, usando que ψ'' está acotada y que cada coordenada de $\breve{\mathbf{x}}_i$ es menor o igual que 1 cuando $\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i-\mathbf{x})\neq 0$, se deduce que si $A<\widehat{s}(\mathbf{x})$ para todo $\mathbf{x}\in\mathcal{C}$, entonces

$$\begin{aligned}\|\mathcal{K}_{n3}(\mathbf{x})\| &\le \|\mathbf{r}\|\frac{\|\psi''\|_\infty}{A\,i(t)}\max_{1\le i\le n}|\widehat{\theta}_i(\mathbf{x})|^2\frac{1}{n}\sum_{i=1}^n|\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i-\mathbf{x})|\\ \sup_{\mathbf{r}\in\mathcal{V}_\tau}\sup_{\mathbf{x}\in\mathcal{C}}\|\mathcal{K}_{n3}(\mathbf{x})\| &\le \tau\frac{\|\psi''\|_\infty}{A\,i(t)}\sup_{\mathbf{x}\in\mathcal{C}}\max_{1\le i\le n}|\widehat{\theta}_i(\mathbf{x})|^2\sup_{\mathbf{x}\in\mathcal{C}}\frac{1}{n}\sum_{i=1}^n|\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i-\mathbf{x})|\,.\end{aligned}$$

Por lo tanto, de (3.15) deducimos que

$$\mathbb{P}\left(\exists n_0 \text{ tal que } \forall n\ge n_0\quad \sup_{\mathbf{r}\in\mathcal{V}_\tau}\sup_{\mathbf{x}\in\mathcal{C}}\|\mathcal{K}_{n3}(\mathbf{x})\|\le\tau\frac{\|\psi''\|_\infty}{A\,i(t)}A_1\left(\tau+\sum_{j=1}^d h_j^2\right)^2\widehat{\Upsilon}^\star_{3,n}\right)=1$$

donde $\widehat{\Upsilon}^\star_{3,n}=\sup_{\mathbf{x}\in\mathcal{C}}(1/n)\sum_{i=1}^n|\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i-\mathbf{x})|=O_{\text{a.s.}}(1)$ no depende de τ, lo que concluye la demostración de (3.21).

• Por último, probaremos (3.20). Tenemos que

$$\begin{aligned}K_{n2}(\mathbf{x}) &= \frac{1}{2\widehat{s}^2(\mathbf{x})}\frac{1}{n}\sum_{i=1}^n\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i-\mathbf{x})\xi_i\psi'(\widehat{V}_i(\mathbf{x}))\left[\left(\mathbf{r}^{\mathrm{T}}\breve{\mathbf{x}}_i\right)^2-2R(\mathbf{X}_i,\mathbf{x})\mathbf{r}^{\mathrm{T}}\breve{\mathbf{x}}_i\right]\\ &= \frac{1}{2\widehat{s}^2(\mathbf{x})}\left(\mathbf{r}^{\mathrm{T}}\widehat{\mathbf{M}}_{n1}\mathbf{r}-2\mathbf{r}^{\mathrm{T}}\widehat{\mathbf{M}}_{n2}\right).\end{aligned} \tag{3.26}$$

Para $\mathbf{x} \in \mathcal{C}$ y $a \in \mathcal{I}_\delta = [1-\delta, 1+\delta]$ (con $0 < \delta < 1$), definamos

$$\mathbf{M}_{n1}(\mathbf{x}, a) = \frac{1}{n} \sum_{i=1}^{n} \mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i - \mathbf{x}) \xi_i \psi'(V_i(\mathbf{x})a) \breve{\mathbf{x}}_i \breve{\mathbf{x}}_i^{\mathrm{T}}$$

y $\mathbf{M}(\mathbf{x}, a) = \mathbb{E}\psi'(\varepsilon a) f_{\mathbf{X}}(\mathbf{x}) r(\mathbf{x}) \mathbf{S}$, donde $\mathbf{S} = \mathrm{diag}(1, s_1, \ldots, s_d)$ y $s_j = \int u^2 K_j(u) du$. Luego, $\mathbf{M}(\mathbf{x}, 1) = \mathbb{E}\psi'(\varepsilon) f_{\mathbf{X}}(\mathbf{x}) r(\mathbf{x}) \mathbf{S}$. Queremos mostrar que

$$\sup_{\mathbf{x} \in \mathcal{C}} \|\widehat{\mathbf{M}}_{n1}(\mathbf{x}) - \mathbb{E}\psi'(\varepsilon) f_{\mathbf{X}}(\mathbf{x}) r(\mathbf{x}) \mathbf{S}\| = o_{\text{a.s.}}(1) \tag{3.27}$$

$$\sup_{\mathbf{x} \in \mathcal{C}} \|\widehat{\mathbf{M}}_{n2}(\mathbf{x})\| = O_{\text{a.s.}}\left(\sum_{j=1}^{d} h_j^2\right). \tag{3.28}$$

Efectivamente, si (3.27) y (3.28) valen, usando (3.15) y que $\sup_{\mathbf{x} \in \mathcal{C}} |\widehat{s}(\mathbf{x}) - \sigma(\mathbf{x})| \xrightarrow{a.s.} 0$, σ es acotada en $\mathcal{C}$, $i(\sigma) > 0$ y reemplazando (3.27) y (3.28) en (3.26) obtenemos que

$$K_{n2}(\mathbf{x}) = \frac{1}{2\sigma^2(\mathbf{x})} \mathbb{E}(\psi'(\varepsilon)) f_{\mathbf{X}}(\mathbf{x}) r(\mathbf{x})\, \mathbf{r}^{\mathrm{T}} \mathbf{S} \mathbf{r} (1 + \widehat{\zeta}_n) + \tau \sum_{j=1}^{d} h_j^2\, \widehat{\Upsilon}_{2,n}\,,$$

donde $\widehat{\zeta}_n = o_{\text{a.s.}}(1)$ y $\widehat{\Upsilon}_{2,n} = O_{\text{a.s.}}(1)$ no dependen de $\mathbf{r}$ y por lo tanto, de τ, ni de $\mathbf{x} \in \mathcal{C}$ pues las convergencias son uniformes sobre $\mathbf{x}$, lo que prueba (3.20).

Comenzaremos probando (3.27). Para $1 \leq j, k \leq d+1$, llamemos $\widehat{M}_{n1,jk}(\mathbf{x})$, $M_{n1,jk}(\mathbf{x}, a)$ y $M_{jk}(\mathbf{x}, a)$ a la componente (j, k) de las matrices $\widehat{\mathbf{M}}_{n1}(\mathbf{x})$, $\mathbf{M}_{n1}(\mathbf{x}, a)$ y $\mathbf{M}(\mathbf{x}, a)$, respectivamente. Luego, para probar (3.27) bastará mostrar que

$$\sup_{\mathbf{x} \in \mathcal{C}} |\widehat{M}_{n1,jk}(\mathbf{x}) - M_{jk}(\mathbf{x}, 1)| = o_{\text{a.s.}}(1)\,. \tag{3.29}$$

Como $\widehat{a}_\sigma(\mathbf{x}) = \sigma(\mathbf{x})/\widehat{s}(\mathbf{x})$, $\widehat{M}_{n1,jk}(\mathbf{x}) = M_{n1,jk}(\mathbf{x}, \widehat{a}_\sigma(\mathbf{x}))$. Por lo tanto, como

$$\sup_{\mathbf{x} \in \mathcal{C}} |\widehat{M}_{n1,jk}(\mathbf{x}) - M_{jk}(\mathbf{x}, 1)| \leq \sup_{\mathbf{x} \in \mathcal{C}} |M_{n1,jk}(\mathbf{x}, \widehat{a}_\sigma(\mathbf{x})) - M_{jk}(\mathbf{x}, \widehat{a}_\sigma(\mathbf{x}))| + \sup_{\mathbf{x} \in \mathcal{C}} |M_{jk}(\mathbf{x}, \widehat{a}_\sigma(\mathbf{x})) - M_{jk}(\mathbf{x}, 1)|\,,$$
$$\sup_{\mathbf{x} \in \mathcal{C}} |M_{n1,jk}(\mathbf{x}, a) - M_{jk}(\mathbf{x}, a)| \leq \sup_{\mathbf{x} \in \mathcal{C}} |M_{n1,jk}(\mathbf{x}, a) - \mathbb{E}M_{n1,jk}(\mathbf{x}, a)| + \sup_{\mathbf{x} \in \mathcal{C}} |\mathbb{E}M_{n1,jk}(\mathbf{x}, a) - M_{jk}(\mathbf{x}, a)|\,,$$

para probar (3.29) probaremos que

(i) $\sup_{\mathbf{x} \in \mathcal{C}} \sup_{a \in \mathcal{I}_\delta} |M_{n1,jk}(\mathbf{x}, a) - \mathbb{E}M_{n1,jk}(\mathbf{x}, a)| = O_{\text{a.s.}}\left(\sqrt{\log n / (n \prod_{j=1}^{d} h_j)}\right)$.

(ii) $\sup_{\mathbf{x} \in \mathcal{C}} \sup_{a \in \mathcal{I}_\delta} |\mathbb{E}M_{n1,jk}(\mathbf{x}, a) - M_{jk}(\mathbf{x}, a)| = o(1)$

(iii) $\sup_{\mathbf{x} \in \mathcal{C}} |M_{n1,jk}(\mathbf{x}, \widehat{a}_\sigma(\mathbf{x})) - M_{jk}(\mathbf{x}, \widehat{a}_\sigma(\mathbf{x}))| = o_{\text{a.s.}}(1)$

(iv) $\sup_{\mathbf{x} \in \mathcal{C}} |M_{jk}(\mathbf{x}, \widehat{a}_\sigma(\mathbf{x})) - M_{jk}(\mathbf{x}, 1)| = o_{\text{a.s.}}(1)$.

(i) se deduce inmediamente del Lema 3.4.6 y de la Observación 3.4.1. Basta considerar la sucesión de variables aleatorias independientes $W_i(a) = \psi'\left(\sigma(\mathbf{X}_i)\varepsilon_i a/\sigma(\mathbf{x})\right)\xi_i$, observar que $|W_i(a)| \leq \|\psi'\|_\infty/i(t)$ para todo a y que $|W_i(a_1) - W_i(a_2)| \leq \|\zeta_2\|_\infty |a_1 - a_2|$ y tomar $m = \widetilde{m} = 1$.

Para probar (ii), sea $\mathcal{C}_0$ indica el entorno compacto de $\mathcal{C}$ de las hipótesis **C5**, **A3** y **A4**. Luego, $f_{\mathbf{X}}$, r y σ son uniformemente continuas en $\mathcal{C}_0$. Llamemos

$$\gamma_{\mathbf{x},a}(\mathbf{t}) = \mathbb{E}\left(\psi'\left(\frac{\sigma(\mathbf{X}_1)\varepsilon_1}{\sigma(\mathbf{x})}a\right)|\mathbf{X}_1 = \mathbf{t}\right).$$

Como los errores ε_i son independientes de las covariables, resulta que

$$|\gamma_{\mathbf{x},a}(\mathbf{t}) - \mathbb{E}\psi'(\varepsilon_1 a)| \leq \mathbb{E}\left|\psi'\left(\frac{\sigma(\mathbf{t})}{\sigma(\mathbf{x})}\varepsilon_1 a\right) - \psi'(\varepsilon_1 a)\right| = \mathbb{E}\left|\zeta_2(\varepsilon_1\, a\,\theta)\right|\left|\frac{\sigma(\mathbf{t})}{\sigma(\mathbf{x})} - 1\right|\frac{1}{\theta},$$

donde θ es un valor intermedio entre $\sigma(\mathbf{t})/\sigma(\mathbf{x})$ y 1. Luego, tenemos que

$$\frac{1}{\theta} \leq \max\left\{1, \frac{\sigma(\mathbf{x})}{\sigma(\mathbf{t})}\right\},$$

de donde usando que $\zeta_2(u) = u\psi''(u)$ es acotada obtenemos que

$$|\gamma_{\mathbf{x},a}(\mathbf{t}) - \mathbb{E}\psi'(\varepsilon_1 a)| \leq \|\zeta_2\|_\infty \left|\frac{\sigma(\mathbf{t})}{\sigma(\mathbf{x})} - 1\right| \max\left\{1, \frac{\sigma(\mathbf{x})}{\sigma(\mathbf{t})}\right\} = \|\zeta_2\|_\infty \left|\sigma(\mathbf{t}) - \sigma(\mathbf{x})\right| \max\left\{\frac{1}{\sigma(\mathbf{x})}, \frac{1}{\sigma(\mathbf{t})}\right\}.$$

Usando que $\inf_{\mathbf{t}\in\mathcal{C}_0}\sigma(\mathbf{t}) > 0$, $\sup_{\mathbf{t}\in\mathcal{C}_0}\sigma(\mathbf{t}) < \infty$, se tiene que existe c_1 tal que para todo $\mathbf{x} \in \mathcal{C}$ y $\mathbf{t} \in \mathcal{C}_0$

$$|\gamma_{\mathbf{x},a}(\mathbf{t}) - \mathbb{E}\psi'(\varepsilon_1 a)| \leq c_1\left|\sigma(\mathbf{t}) - \sigma(\mathbf{x})\right|. \tag{3.30}$$

Observemos que

$$\begin{aligned}
\mathbb{E}M_{n1,jk}(\mathbf{x}, a) &= \mathbb{E}\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_1 - \mathbf{x})r(\mathbf{X}_1)\gamma_{\mathbf{x},a}(\mathbf{X}_1)\breve{\mathbf{x}}_{1,j}\breve{\mathbf{x}}_{1,k} \\
&= \int \mathcal{K}_{\mathbf{H}_d}(\mathbf{u} - \mathbf{x})r(\mathbf{u})\gamma_{\mathbf{x},a}(\mathbf{u})\left[\frac{u_{j-1} - x_{j-1}}{h_{j-1}}\right]\left[\frac{u_{k-1} - x_{k-1}}{h_{k-1}}\right] f_{\mathbf{X}}(\mathbf{u})\, d\mathbf{u}
\end{aligned}$$

donde si $j = 1$, definimos $(u_{j-1} - x_{j-1})/h_{j-1} = 1$ y si $k = 1$, $(u_{k-1} - x_{k-1})/h_{k-1} = 1$. Haciendo cambio de variables con $\mathbf{y} = \mathbf{H}_d^{-1}(\mathbf{u} - \mathbf{x})$, se tiene que

$$\begin{aligned}
\mathbb{E}M_{n1_{jk}}(\mathbf{x}, a) - M_{jk}(\mathbf{x}, a) &= \int r(\mathbf{H}_d\mathbf{y} + \mathbf{x})\gamma_{\mathbf{x},a}(\mathbf{H}_d\mathbf{y} + \mathbf{x})f_{\mathbf{X}}(\mathbf{H}_d\mathbf{y} + \mathbf{x})y_{j-1}y_{k-1}\mathcal{K}(\mathbf{y})\, d\mathbf{y} \\
&\quad -\mathbb{E}(\psi'(\varepsilon_1 a))f_{\mathbf{X}}(\mathbf{x})r(\mathbf{x})S_{jk},
\end{aligned}$$

donde $\mathbf{S} = (S_{jk})$ con $S_{jk} = \int y_{j-1}y_{k-1}\mathcal{K}(\mathbf{y})d\mathbf{y}$, donde entendemos que $y_0 = 1$. Por lo tanto, si $j \neq k$ $S_{jk} = 0$ y $S_{jj} = s_{j-1,j-1}$. Luego, obtenemos que

$$\begin{aligned}
\mathbb{E}M_{n1,jk}(\mathbf{x}, a) - M_{jk}(\mathbf{x}, a) &= \int r(\mathbf{H}_d\mathbf{y} + \mathbf{x})\gamma_{\mathbf{x},a}(\mathbf{H}_d\mathbf{y} + \mathbf{x})f_{\mathbf{X}}(\mathbf{H}_d\mathbf{y} + \mathbf{x})y_{j-1}y_{k-1}\mathcal{K}(\mathbf{y})\, d\mathbf{y} \\
&\quad -\mathbb{E}(\psi'(\varepsilon_1 a))f_{\mathbf{X}}(\mathbf{x})r(\mathbf{x})\int y_{j-1}y_{k-1}\mathcal{K}(\mathbf{y})d\mathbf{y} \\
&= \int \left\{r(\mathbf{H}_d\mathbf{y} + \mathbf{x})\gamma_{\mathbf{x},a}(\mathbf{H}_d\mathbf{y} + \mathbf{x})f_{\mathbf{X}}(\mathbf{H}_d\mathbf{y} + \mathbf{x}) - \mathbb{E}(\psi'(\varepsilon_1 a))f_{\mathbf{X}}(\mathbf{x})r(\mathbf{x})\right\} \\
&\quad \times y_{j-1}y_{k-1}\mathcal{K}(\mathbf{y})\, d\mathbf{y}.
\end{aligned}$$

Como $f_{\mathbf{X}}$, σ y r son uniformemente continuas en $\mathcal{C}_0$, dado $\epsilon > 0$ existe $\eta > 0$ tal que para todo $\mathbf{x} \in \mathcal{C}$, $\mathbf{u} \in \mathcal{C}_0$ si $\|\mathbf{u} - \mathbf{x}\| < \eta$ entonces $|f_{\mathbf{X}}(\mathbf{u})r(\mathbf{u}) - f_{\mathbf{X}}(\mathbf{x})r(\mathbf{x})| < \epsilon$ y $|\sigma(\mathbf{u}) - \sigma(\mathbf{x})| < \epsilon$. Como K_j tiene soporte compacto $[-1, 1]$, si $\mathcal{K}(\mathbf{y}) \neq 0$ entonces $\|\mathbf{y}\| \leq \sqrt{d}$. Luego, como $\max_{1\leq j\leq d} h_{j,n} \to 0$, existe n_0 tal que si $n \geq n_0$, $\|\mathbf{H}_d\mathbf{y}\| < \eta$ para todo $\mathbf{y}$ tal que $\mathcal{K}(\mathbf{y}) \neq 0$ y además, $\mathbf{H}_d\mathbf{y} + \mathbf{x} \in \mathcal{C}_0$. Por lo tanto, si $n \geq n_0$ para todo $\mathbf{x} \in \mathcal{C}$ y para todo $\mathbf{y}$ tal que $\mathcal{K}(\mathbf{y}) \neq 0$, tenemos que $|f_{\mathbf{X}}(\mathbf{H}_d\mathbf{y} + \mathbf{x})r(\mathbf{H}_d\mathbf{y} + \mathbf{x}) - f_{\mathbf{X}}(\mathbf{x})r(\mathbf{x})| < \epsilon$ y $|\sigma(\mathbf{H}_d\mathbf{y} + \mathbf{x}) - \sigma(\mathbf{x})| \leq \epsilon$, de donde, como $\gamma_{\mathbf{x},a}(\mathbf{x}) = \mathbb{E}(\psi'(\varepsilon_1 a))$, usando (3.30) y que $|\gamma_{\mathbf{x},a}(\mathbf{x})| \leq \|\psi'\|_\infty$ obtenemos que para todo $\mathbf{x} \in \mathcal{C}$, $a \in \mathcal{I}_\delta$ e $\mathbf{y}$ tal que $\mathcal{K}(\mathbf{y}) \neq 0$

$$\begin{aligned}
|r(\mathbf{H}_d\mathbf{y} + \mathbf{x})\gamma_{\mathbf{x},a}(\mathbf{H}_d\mathbf{y} + \mathbf{x})f_{\mathbf{X}}(\mathbf{H}_d\mathbf{y} + \mathbf{x}) \; & - \mathbb{E}(\psi'(\varepsilon_1 a))f_{\mathbf{X}}(\mathbf{x})r(\mathbf{x})| \\
\leq \; & r(\mathbf{H}_d\mathbf{y} + \mathbf{x})f_{\mathbf{X}}(\mathbf{H}_d\mathbf{y} + \mathbf{x})\,|\gamma_{\mathbf{x},a}(\mathbf{H}_d\mathbf{y} + \mathbf{x}) - \gamma_{\mathbf{x},a}(\mathbf{x})| \\
& + |\gamma_{\mathbf{x},a}(\mathbf{x})|\,|r(\mathbf{H}_d\mathbf{y} + \mathbf{x})f_{\mathbf{X}}(\mathbf{H}_d\mathbf{y} + \mathbf{x}) - f_{\mathbf{X}}(\mathbf{x})r(\mathbf{x})| \\
\leq \; & c_1 \sup_{\mathbf{u}\in\mathcal{C}_0} r(\mathbf{u})f_{\mathbf{X}}(\mathbf{u})\;|\sigma(\mathbf{H}_d\mathbf{y} + \mathbf{x}) - \sigma(\mathbf{x})| \\
& + \|\psi'\|_\infty\,|r(\mathbf{H}_d\mathbf{y} + \mathbf{x})f_{\mathbf{X}}(\mathbf{H}_d\mathbf{y} + \mathbf{x}) - f_{\mathbf{X}}(\mathbf{x})r(\mathbf{x})| \\
\leq \; & \left(c_1 \sup_{\mathbf{u}\in\mathcal{C}_0} r(\mathbf{u})f_{\mathbf{X}}(\mathbf{u}) + \|\psi'\|_\infty\right)\epsilon = c_2\,\epsilon\,.
\end{aligned}$$

Luego, si $n \geq n_0$ tenemos que

$$\sup_{\mathbf{x}\in\mathcal{C}} \sup_{a\in\mathcal{I}_\delta} |\mathbb{E}M_{n1,jk}(\mathbf{x}, a) - M_{jk}(\mathbf{x}, a)| \;\leq\; c_2\epsilon \int y_{j-1}y_{k-1}\mathcal{K}(\mathbf{y})\,d\mathbf{y} = c_3\epsilon\,,$$

lo que concluye la demostración de (ii).

De esta manera, por (i) y (ii) resulta que

$$\sup_{\mathbf{x}\in\mathcal{C}} \sup_{a\in\mathcal{I}_\delta} |M_{n1,jk}(\mathbf{x}, a) - M_{jk}(\mathbf{x}, a)| = O_{\text{a.s.}}\left(\sqrt{\frac{\log n}{n\prod_{j=1}^d h_j}}\right) + o\,(1) = o_{\text{a.s.}}(1)\,. \tag{3.31}$$

En particular,

$$\sup_{\mathbf{x}\in\mathcal{C}} |M_{n1,jk}(\mathbf{x}, 1) - M_{jk}(\mathbf{x}, 1)| = O_{\text{a.s.}}\left(\sqrt{\frac{\log n}{n\prod_{j=1}^d h_j}}\right) + o\,(1) = o_{\text{a.s.}}(1)\,. \tag{3.32}$$

Probemos ahora (iii). Por (3.16) y (3.31), dado $\eta > 0$, existe $\mathcal{N}$ tal que $\mathbb{P}(\mathcal{N}) = 0$ y para todo $\omega \notin \mathcal{N}$, existe n_1 tal que para todo $n \geq n_1$, $|\widehat{a}_\sigma(\mathbf{x}) - 1| < 1/2$ y $\sup_{\mathbf{x}\in\mathcal{C}} \sup_{a\in\mathcal{I}_\delta} |M_{n1,jk}(\mathbf{x}, a) - M_{jk}(\mathbf{x}, a)| < \eta$, con $\delta = 1/2$. Luego, para todo $\omega \notin \mathcal{N}$ y $n \geq n_1$ tenemos que

$$\sup_{\mathbf{x}\in\mathcal{C}} |M_{n1,jk}(\mathbf{x}, \widehat{a}_\sigma(\mathbf{x})) - M_{jk}(\mathbf{x}, \widehat{a}_\sigma(\mathbf{x}))| \leq \sup_{\mathbf{x}\in\mathcal{C}} \sup_{a\in\mathcal{I}_\delta} |M_{n1,jk}(\mathbf{x}, a) - M_{jk}(\mathbf{x}, a)| < \eta\,,$$

con lo que queda demostrado (iii).

Para demostrar (iv), observemos que como $\mathbf{S} = \text{diag}(1, s_1, \ldots, s_d)$, $M_{jk}(\mathbf{x}, a) = 0$ si $j \neq k$. Es decir, basta probar (iv) cuando $j = k$. Usando que $f_{\mathbf{X}}$ y r están acotadas en $\mathcal{C}$, si $c_4 = \sup_{\mathbf{x}\in\mathcal{C}} f_{\mathbf{X}}(\mathbf{x})r(\mathbf{x}) \max(1, s_1, \ldots, s_d)$ tenemos que

$$\sup_{\mathbf{x}\in\mathcal{C}} |M_{jj}(\mathbf{x}, \widehat{a}_\sigma(\mathbf{x})) - M_{jj}(\mathbf{x}, 1)| \leq c_4 \sup_{\mathbf{x}\in\mathcal{C}} \left|\mathbb{E}\psi'\left(\varepsilon\widehat{a}_\sigma(\mathbf{x})\right) - \mathbb{E}\psi'(\varepsilon)\right|\,.$$

Mediante un razonamiento análogo al realizado para acotar $\gamma_{\mathbf{x},a}$ en (ii), si llamamos $\lambda_1(a) = \mathbb{E}\psi'(\varepsilon a)$. Como ψ' es continua, resulta que λ_1 es continua y además se cumple

$$|\lambda_1(a) - \lambda_1(1)| = \left|\mathbb{E}\left[\psi'(\varepsilon a) - \psi'(\varepsilon)\right]\right| = \left|\mathbb{E}\psi''(\varepsilon\theta)\varepsilon(a-1)\right| \leq \|\zeta_2\|_\infty \frac{1}{\theta}(a-1)\,,$$

donde θ es un punto intermedio entre a y 1, es decir,

$$|\lambda_1(a) - \lambda_1(1)| \leq \|\zeta_2\|_\infty (a-1)\max(1, \frac{1}{a})\,,$$

de donde,

$$\sup_{\mathbf{x}\in\mathcal{C}} |M_{jj}(\mathbf{x}, \widehat{a}_\sigma(\mathbf{x})) - M_{jj}(\mathbf{x}, 1)| \leq c_4\|\zeta_2\|_\infty \sup_{\mathbf{x}\in\mathcal{C}} |\widehat{a}_\sigma(\mathbf{x}) - 1| \sup_{\mathbf{x}\in\mathcal{C}} \left(1 + \frac{1}{\widehat{a}_\sigma(\mathbf{x})}\right)\,.$$

(iv) se deduce ahora del hecho que $\sup_{\mathbf{x}\in\mathcal{C}} |\widehat{a}_\sigma(\mathbf{x}) - 1| \xrightarrow{a.s.} 0$. Es decir, hemos concluído la demostración de (3.27).

Por (3.25) y usando que ψ' está acotada y que $i(t) > 0$, se tiene que

$$\sup_{\mathbf{x}\in\mathcal{C}} \|\widehat{\mathbf{M}}_{n2}(\mathbf{x})\| \leq \frac{\|\psi'\|_\infty}{i(t)} A_g \left(\sum_{j=1}^d h_j^2\right) \sup_{\mathbf{x}\in\mathcal{C}} \frac{1}{n}\sum_{i=1}^n |\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i - \mathbf{x})|\,.$$

Como antes, por la Observación 3.4.1, $\widehat{f}(\mathbf{x}) = (1/n)\sum_{j=1}^n \mathcal{K}^\star_{\mathbf{H}_d}(\mathbf{x} - \mathbf{X}_j)$, converge a $f_{\mathbf{X}}$ uniformemente. Por lo tanto, usando **A3** obtenemos que $\sup_{\mathbf{x}\in\mathcal{C}} \widehat{f}(\mathbf{x}) = O_{\text{a.s.}}(1)$ lo que concluye la demostración de (3.28) y por lo tanto, la de la Proposición. □

De la Proposición 3.2.2 obtenemos en forma inmediata el siguiente Teorema que muestra la consistencia del estimador de integración marginal basado en polinomios de orden 1.

Teorema 3.2.1. *Supongamos que se cumple **A0** y supongamos que se cumplen **A3**, **A4** y **C5** con $\mathcal{C} = \mathcal{S}_Q \subset \mathcal{S}_f$. Si además valen **A1**, , **A6**, **C1** a **C3**a), **C6** y **C7**, entonces*

a) $\sup_{x\in\mathcal{C}_\alpha} |\widehat{g}_{\alpha,\mathrm{M}_1}(x) - g_\alpha(x)| \xrightarrow{a.s.} 0$,

b) $\sup_{\mathbf{x}\in\mathcal{C}} |\widehat{g}_{\mathrm{M}_1}(\mathbf{x}) - g(\mathbf{x})| \xrightarrow{a.s.} 0$, *donde* $\widehat{g}_{\mathrm{M}_1}(\mathbf{x}) = \sum_{j=1}^d \widehat{g}_{j,\mathrm{M}_1}(x_j)$.

3.3 Consistencia de los $M-$estimadores locales basados en polinomios locales de orden q

Para obtener la consistencia de los $M-$estimadores locales basados en polinomios locales de orden q necesitaremos reemplazar los supuestos **C1** y **C2**. Recordemos que el $M-$estimador local basado en polinomios locales de orden q inicial $\widetilde{g}_{\mathrm{M}_{q,\alpha}}(\mathbf{x})$ se adapta a la componente α es estimar. Por eso en esta sección fijaremos $\alpha = 1, \ldots, d$. Para darle mayor generalidad permitiremos que las ventanas de las distintas componentes sean distintas, es decir, no nos restingiremos al caso $h_{\alpha,n} = h_n$ mientras que $h_{j,n} = \widetilde{h}_n$ si $j \neq \alpha$.

C1′ Para todo $j = 1, \cdots, d$, la componente marginal g_j es continuamente diferenciable en un entorno del soporte, $\mathcal{S}_j$, de la densidad de X_j con derivada $g_j' = g_j^{(1)}$ acotada. Además, g_α es continuamente diferenciable hasta el orden $q+1$ y la derivada de orden $q+1$, $g_\alpha^{(q+1)}$, es acotada en $\mathcal{S}_\alpha$.

C2′ La función de núcleos $\mathcal{K} : \mathbb{R}^d \to \mathbb{R}$ es tal que $\mathcal{K}(\mathbf{x}) = \prod_{j=1}^d K_j(x_j)$, donde $K_j : \mathbb{R} \to \mathbb{R}$ son funciones pares y acotadas con soporte acotado, digamos $[-1, 1]$ y tales que $\int K_j(u)du = 1$. El núcleo K_α cumple que $0 < \int |u|^j K_\alpha(u)\, du < \infty$ para $j = 1, \ldots, 2q$ y $\int u^j K_\alpha(u)\, du > 0$ si j es par. Más aún, la matriz $\mathbf{S}^{(\alpha)} = (\int u^{i+j} K_\alpha(u)du)_{0 \le i,j \le q}$ es definida positiva.

Sea $s_{ij}^{(\alpha)} = \int u_\alpha^{i+j} \mathcal{K}(\mathbf{u})\, d\mathbf{u} = \int u^{i+j} K_\alpha(u)du$, $0 \le i, j \le q$, donde $\mathbf{u} = (u_1, \ldots, u_d)^{\mathrm{T}}$. Observemos que bajo **C2′**, tenemos que $s_{ij}^{(\alpha)} = 0$ si $i + j$ es impar y $s_{00}^{(\alpha)} = 1$, con lo cual $\mathbf{S}^{(\alpha)} = (s_{ij}^{(\alpha)})$. Como antes sea $U_i = \sigma(\mathbf{X}_i)\varepsilon_i$, luego, $Y_i = g(\mathbf{X}_i) + U_i$. Definamos $R_\alpha(X_{i,\alpha}) = g_\alpha(X_{i,\alpha}) - g_\alpha(x_\alpha) - \sum_{j=1}^q g_\alpha^{(j)}(x_\alpha)(X_{i,\alpha} - x_\alpha)^j/j!$ y sea $R(\mathbf{X}_i) = \sum_{j \ne \alpha} \{g_j(X_{ij}) - g_j(x_j)\} + R_\alpha(X_{i,\alpha})$,

$$\begin{aligned} \breve{\mathbf{x}}_{i,\alpha} &= \left(1, \frac{X_{i,\alpha} - x_\alpha}{h_\alpha}, \frac{(X_{i,\alpha} - x_\alpha)^2}{h_\alpha^2}, \ldots, \frac{(X_{i,\alpha} - x_\alpha)^q}{h_\alpha^q}\right)^{\mathrm{T}}, \\ \mathbf{H}_\alpha &= \operatorname{diag}(1, h_\alpha, h_\alpha^2, \ldots, h_\alpha^q) \in \mathbb{R}^{(q+1)\times(q+1)}. \end{aligned}$$

Proposición 3.3.1. *Sea* $\mathbf{x}$ *un punto del interior de* $\mathcal{S}_f$ *y* $\mathcal{C}_0$ *la clausura de un entorno de* $\mathbf{x}$ *para el que se cumple* ***A3*** *con* $\mathcal{C} = \mathcal{C}_0$. *Supongamos que se cumplen* ***A1***, ***A4***, ***A7*** *y* ***C1′***, ***C2′***, ***C3*** *a* ***C5***. *Entonces, si* $\widehat{s}(\mathbf{x}) \xrightarrow{p} \sigma(\mathbf{x})$ *existe una solución* $\widehat{\boldsymbol{\beta}}(\mathbf{x})$ *de* $\boldsymbol{\Psi}_{n,\alpha}(\boldsymbol{\beta}, \mathbf{x}, \widehat{s}(\mathbf{x})) = \mathbf{0}_{q+1}$ *donde* $\boldsymbol{\Psi}_{n,\alpha}(\boldsymbol{\beta}, \mathbf{x}, \sigma) = (\Psi_{n,\alpha,0}(\boldsymbol{\beta}, \mathbf{x}, \sigma), \ldots, \Psi_{n,\alpha,q}(\boldsymbol{\beta}, \mathbf{x}, \sigma))^{\mathrm{T}}$ *está definida en (2.22). Más aún,* $\mathbf{H}_\alpha\{\widehat{\boldsymbol{\beta}}(\mathbf{x}) - \boldsymbol{\beta}(\mathbf{x})\} \xrightarrow{p} 0$, *donde* $\boldsymbol{\beta}(\mathbf{x}) = (g(\mathbf{x}), g_\alpha^{(1)}(x_\alpha), \ldots, g_\alpha^{(q)}(x_\alpha)/q!)^{\mathrm{T}}$ *y* $g_\alpha^{(1)} = g_\alpha'$.

Demostración. La demostración es análoga a la de la Proposición 3.2.1.

Sea $\mathbf{r} = (\beta_0, h_\alpha\beta_1, h_\alpha^2\beta_2, \cdots, h_\alpha^q\beta_q)^{\mathrm{T}}$ y $\mathbf{r}_0 = (g(\mathbf{x}), h_\alpha\, g_\alpha^{(1)}(x_\alpha), h_\alpha^2\, g_\alpha^{(2)}(x_\alpha), \ldots, h_\alpha^q\, g_\alpha^{(q)}(x_\alpha))^{\mathrm{T}}$. Definamos

$$\begin{aligned} \ell_n(\mathbf{r}) &= \frac{1}{n}\sum_{i=1}^n \rho\left(\frac{Y_i - \beta_0 - \sum_{j=1}^q \beta_j(X_{i\alpha} - x_\alpha)^j}{\widehat{s}(\mathbf{x})}\right)\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i - \mathbf{x})\xi_i \\ &= \frac{1}{n}\sum_{i=1}^n \rho\left(\frac{Y_i - \mathbf{r}^{\mathrm{T}}\breve{\mathbf{x}}_{i,\alpha}}{\widehat{s}(\mathbf{x})}\right)\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i - \mathbf{x})\xi_i\,. \end{aligned} \tag{3.33}$$

Sea $\mathcal{V}_\tau = \{\mathbf{r} : \|\mathbf{r} - \mathbf{r}_0\| = \tau\}$ la esfera de centro $\mathbf{r}_0$ y de radio τ. Como en la Proposición 3.2.1, basta mostrar que dado $\nu > 0$, existe $0 < \tau_\nu < 1$ suficientemente pequeño, tal que para cualquier $0 < \tau < \tau_\nu$ si $n \ge n_0$,

$$\mathbb{P}\left(\inf_{\mathbf{r}\in\mathcal{V}_\tau} \ell_n(\mathbf{r}) > \ell_n(\mathbf{r}_0)\right) > 1 - \nu\,. \tag{3.34}$$

Observemos que $Y_i - \mathbf{r}^{\mathrm{T}}\breve{\mathbf{x}}_{i,\alpha} = U_i + g(\mathbf{X}_i) - \mathbf{r}^{\mathrm{T}}\breve{\mathbf{x}}_{i,\alpha} = U_i + R(\mathbf{X}_i) + (\mathbf{r}_0 - \mathbf{r})^{\mathrm{T}}\breve{\mathbf{x}}_{i,\alpha}$ de donde, como en la Proposición 3.2.1, si $\widehat{Z}_i = (U_i + R(\mathbf{X}_i))/\widehat{s}(\mathbf{x})$ y $\widehat{\Delta}_i = (\mathbf{r} - \mathbf{r}_0)^{\mathrm{T}}\breve{\mathbf{x}}_{i,\alpha}/\widehat{s}(\mathbf{x})$ tenemos que

$$\ell_n(\mathbf{r}) - \ell_n(\mathbf{r}_0) = \frac{1}{n}\sum_{i=1}^n \mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i - \mathbf{x})\xi_i \int_{\widehat{Z}_i}^{\widehat{Z}_i - \widehat{\Delta}_i} \psi(t)\, dt = K_{n1} + K_{n2} + K_{n3}\,, \tag{3.35}$$

con

$$\begin{aligned}
K_{n1} &= \frac{1}{n}\sum_{i=1}^{n}\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i-\mathbf{x})\xi_i\int_{\widehat{Z}_i}^{\widehat{Z}_i-\widehat{\Delta}_i}\psi(\widehat{V}_i)dt\\
K_{n2} &= \frac{1}{n}\sum_{i=1}^{n}\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i-\mathbf{x})\xi_i\int_{\widehat{Z}_i}^{\widehat{Z}_i-\widehat{\Delta}_i}\psi'(\widehat{V}_i)(t-\widehat{V}_i)dt\\
K_{n3} &= \frac{1}{n}\sum_{i=1}^{n}\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i-\mathbf{x})\xi_i\int_{\widehat{Z}_i}^{\widehat{Z}_i-\widehat{\Delta}_i}\left[\psi(t)-\psi(\widehat{V}_i)-\psi'(\widehat{V}_i)(t-\widehat{V}_i)\right]dt\,,
\end{aligned}$$

donde $\widehat{V}_i = U_i/\widehat{s}(\mathbf{x}) = \sigma(\mathbf{X}_i)\varepsilon_i/\widehat{s}(\mathbf{x})$ y $\mathbf{r}\in\mathcal{V}_\tau$.

Como en la Proposición 3.2.1, sea $V_i = V_i(\mathbf{x}) = U_i/\sigma(\mathbf{x}) = \sigma(\mathbf{X}_i)\varepsilon_i/\sigma(\mathbf{x})$, luego $\widehat{V}_i = V_i\,\widehat{a}_\sigma(\mathbf{x})$ con $\widehat{a}_\sigma(\mathbf{x}) = \sigma(\mathbf{x})/\widehat{s}(\mathbf{x})$. Entonces $K_{n1} = -(1/\widehat{s}(\mathbf{x}))(\mathbf{r}-\mathbf{r}_0)^{\mathrm{T}}\mathbf{J}_{n,\alpha}(\mathbf{x},\widehat{a}_\sigma(\mathbf{x}))$, donde $\mathbf{J}_{n,\alpha}(\mathbf{x},a)$ está definido por

$$\mathbf{J}_{n,\alpha}(\mathbf{x},a) = \frac{1}{n}\sum_{i=1}^{n}\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i-\mathbf{x})\xi_i\psi\left(\frac{\sigma(\mathbf{X}_i)\varepsilon_i}{\sigma(\mathbf{x})}\,a\right)\breve{\mathbf{x}}_{i,\alpha} = \frac{1}{n}\sum_{i=1}^{n}\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i-\mathbf{x})\xi_i\psi\left(V_i\,a\right)\breve{\mathbf{x}}_{i,\alpha}\,. \quad (3.36)$$

Como en la Proposición 3.2.1, **A1** y **C3** implican $\mathbb{E}\mathbf{J}_{n,\alpha}(\mathbf{x},a) = \mathbf{0_{q+1}}$. Por otra parte,

$$\mathrm{tr}\left(\mathrm{VAR}(\mathbf{J}_{n,\alpha}(\mathbf{x},a))\right) = \frac{1}{n}\mathrm{tr}\left(\mathrm{VAR}(\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_1-\mathbf{x})\xi_1\psi\left(V_1a\right)\breve{\mathbf{x}}_{1,\alpha})\right) = \frac{1}{n}\mathbb{E}(\mathcal{K}^2_{\mathbf{H}_d}(\mathbf{X}_1-\mathbf{x})\xi_1^2\psi^2\left(V_1a\right)\|\breve{\mathbf{x}}_{1,\alpha}\|^2)\,.$$

Luego, tenemos

$$\mathrm{tr}\left(\mathrm{VAR}(\mathbf{J}_{n,\alpha}(\mathbf{x},a))\right) \le \|\psi\|^2_\infty\frac{1}{n}\mathbb{E}\left(\mathcal{K}^2_{\mathbf{H}_d}(\mathbf{X}_1-\mathbf{x})\|\breve{\mathbf{x}}_{1,\alpha}\|^2\frac{r(\mathbf{x}_1)}{t(\mathbf{X}_1)}\right) = \|\psi\|^2_\infty\frac{1}{n\prod_{j=1}^d h_j}A_n(\mathbf{x})\,,$$

donde ahora

$$A_n(\mathbf{x}) = \mathbb{E}\left(\frac{1}{\prod_{j=1}^d h_j}\prod_{j=1}^d K_j^2\left(\frac{X_{1j}-x_j}{h_j}\right)\left[1+\sum_{j=1}^q\frac{(X_{1\alpha}-x_\alpha)^j}{h_\alpha^j}\right]\frac{r(\mathbf{x}_1)}{t(\mathbf{X}_1)}\right)\,.$$

Es fácil ver que **C2′** y **A3** implican que $A_n(\mathbf{x})\le C$ para $n\ge n_0$. Luego, $\mathrm{tr}\left(\mathrm{VAR}(\mathbf{J}_{n,\alpha}(\mathbf{x},a))\right) = O\left(\left\{n\prod_{j=1}^d h_j\right\}^{-1}\right)$.

Como en la Proposición 3.2.1, $\sqrt{n\prod_{j=1}^d h_j}\left(\mathbf{J}_{n,\alpha}(\mathbf{x},\widehat{a}_\sigma(\mathbf{x}))-\mathbf{J}_{n,\alpha}(\mathbf{x},1)\right)\xrightarrow{p}0$ pues $\widehat{s}(\mathbf{x})\xrightarrow{p}\sigma(\mathbf{x})$. Por otra parte, como $\mathbf{r}\in\mathcal{V}_\tau$, $(\mathbf{r}-\mathbf{r}_0)^{\mathrm{T}}\mathbf{J}_{n,\alpha}(\mathbf{x},\sigma(\mathbf{x})) = O_{\mathbb{P}}\left(\{\sqrt{n\prod_{j=1}^d h_j}\}^{-1}\right)\tau$, de donde

$$K_{n1} = \tau\,\frac{1}{\sqrt{n\prod_{j=1}^d h_j}}\,\widehat{\Upsilon}_{1,n}\,, \quad (3.37)$$

donde $\widehat{\Upsilon}_{1,n} = O_{\mathbb{P}}(1)$ que no depende de $\mathbf{r}$. Recordemos que $\widehat{Z}_i = \widehat{V}_i + R(\mathbf{X}_i)\widehat{s}(\mathbf{x})^{-1}$ y $\widehat{\Delta}_i = (\mathbf{r}-\mathbf{r}_0)^{\mathrm{T}}\breve{\mathbf{x}}_{i,\alpha}/\widehat{s}(\mathbf{x})$. Por el teorema del valor medio integral,

$$\begin{aligned} K_{n3} &= \frac{1}{n}\sum_{i=1}^{n}\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i-\mathbf{x})\xi_i\int_{\widehat{V}_i+R(\mathbf{X}_i)\widehat{s}(\mathbf{x})^{-1}}^{\widehat{V}_i+R(\mathbf{X}_i)\widehat{s}(\mathbf{x})^{-1}-\widehat{\Delta}_i}\left[\psi(t)-\psi(\widehat{V}_i)-\psi'(\widehat{V}_i)(t-\widehat{V}_i)\right]dt \\ &= -(\mathbf{r}-\mathbf{r}_0)^{\mathrm{T}}\frac{1}{n\widehat{s}(\mathbf{x})}\sum_{i=1}^{n}\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i-\mathbf{x})\xi_i\left[\psi\left(\widehat{V}_i+\widehat{\theta}_i\right)-\psi(\widehat{V}_i)-\psi'(\widehat{V}_i)\widehat{\theta}_i\right]\breve{\mathbf{x}}_{i,\alpha}\,, \end{aligned}$$

donde $\widehat{\theta}_i$ es un punto intermedio entre $R(\mathbf{X}_i)/\widehat{s}(\mathbf{x})$ y $\{R(\mathbf{X}_i)-(\mathbf{r}-\mathbf{r}_0)^{\mathrm{T}}\breve{\mathbf{x}}_{i,\alpha}\}/\widehat{s}(\mathbf{x})$. Por simplicidad, si $i\in\{i:\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i-\mathbf{x})=0\}$, redefinamos $\widehat{\theta}_i=0$ ya que no cambia el resultado de la suma. Observando que $|X_{i,j}-x_j|\le h_j$ para $j=1,\cdots,d$, si $\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i-\mathbf{x})\neq 0$, resulta $\max_{i:\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i-\mathbf{x})\neq 0}|R(\mathbf{X}_i)|\le A_g\left(\sum_{j\neq\alpha}h_j+h_\alpha^{q+1}\right)$, donde A_g es una constante que depende solamente de $\|g_j^{(1)}\|_\infty$ para $j\neq\alpha$ y de $\|g_\alpha^{(q+1)}\|_\infty$. Luego, como se cumple **A7**

$$\max_{1\le i\le n}|\widehat{\theta}_i|\le\frac{(d+1)\tau+A_g\left(\sum_{j\neq\alpha}h_j+h_\alpha^{q+1}\right)}{A}\le A_1\left(\tau+h_\alpha^{q+1}+\sum_{j\neq\alpha}h_j\right).$$

Sea $\boldsymbol{\theta}_n=(\theta_1,\ldots,\theta_n)^{\mathrm{T}}$ y para cualquier $\eta>0$ dado, definamos $D_\eta=\{\boldsymbol{\theta}_n:|\theta_i|<\eta,\forall i\le n\}$. Llamemos

$$V(\boldsymbol{\theta}_n,a)=\frac{1}{n}\sum_{i=1}^{n}\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i-\mathbf{x})\xi_i\left[\psi\left(V_ia+\theta_i\right)-\psi\left(V_ia\right)-\psi'\left(V_ia\right)\theta_i\right]\breve{\mathbf{x}}_{i,\alpha}\,.$$

Luego

$$\sup_{D_\eta}|V(\boldsymbol{\theta}_n,a)|\le\frac{1}{n}\sum_{i=1}^{n}\sup_{D_\eta}\left|\psi\left(V_ia+\theta_i\right)-\psi\left(V_ia\right)-\psi'\left(V_ia\right)\theta_i\right|\|\breve{\mathbf{x}}_{i,\alpha}\|\,|\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i-\mathbf{x})|\,\xi_i\,.$$

Observemos que $aV_i=c_i(\mathbf{x})\varepsilon_i$ donde $c_i(\mathbf{x})=a\sigma(\mathbf{X}_i)/\sigma(\mathbf{x})$. Como σ es una función continua y $|X_{ij}-x_j|\le h_j$ si $\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i-\mathbf{x})\neq 0$, pues K_j tiene soporte en $[-1,1]$, tenemos que $|\sigma(\mathbf{X}_i)-\sigma(\mathbf{x})|\le i(\sigma)\gamma/4$, para todo $i=1,\ldots,n$ tal que $\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i-\mathbf{x})\neq 0$ y n suficientemente grande. Por lo tanto, si n es suficientemente grande y $|a-1|\le\gamma/2$, se cumple que $|c_i(\mathbf{x})-1|\le\gamma$. Es decir, si $|a-1|\le\gamma/2$ por **C3**b) tenemos que, para todo $i=1,\ldots,n$ tal que $\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i-\mathbf{x})\neq 0$,

$$\begin{aligned}\mathbb{E}\left(\sup_{D_\eta}\left|\psi\left(V_ia+\theta_i\right)-\psi\left(V_ia\right)-\psi'\left(V_ia\right)\theta_i\right|\,|\mathbf{X}_i\right) &\le \mathbb{E}\left(\sup_{|c-1|\le\gamma}\sup_{|\theta|\le\eta}\left|\psi\left(c\varepsilon+\theta\right)-\psi\left(c\varepsilon\right)-\psi'\left(c\varepsilon\right)\theta\right|\right) \\ &= \eta\, b_\eta\,,\end{aligned}$$

donde b_η es una sucesión de números positivos que tiende a cero cuando $\eta\to 0$. Por lo tanto, tomando $\eta=2A_1\tau$ y usando que $r(\mathbf{x})$ es continua y acotada en un entorno del punto $\mathbf{x}$ obtenemos

$$\begin{aligned}\mathbb{E}\left[\sup_{|a-1|\le\gamma/2}\sup_{D_\eta}\|V(\boldsymbol{\theta}_n,a)\|\right] &\le b_\eta\mathbb{E}\left[\|\breve{\mathbf{x}}_{1,\alpha}\|\,|\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_1-\mathbf{x})|\,r(\mathbf{X}_1)\right]\eta \\ &\le \eta B_\eta\,,\end{aligned}$$

con $B_\eta = C\, b_\eta$ para alguna constante $C > 0$ independiente de η y de n. Por lo tanto, como en la Proposición 3.2.1 como $\max_{1\le i\le n} |\widehat{\theta}_i| = A_1 \left(\tau + h_\alpha^{q+1} + \sum_{j\neq\alpha} h_j\right)$ y se cumple **C4**, podemos elegir n_τ tal que para $n \ge n_\tau$, se verifica $h_\alpha^{q+1} + \sum_{j\neq\alpha} h_j \le \tau$, de donde $\max_{1\le i\le n} |\widehat{\theta}_i| \le \eta$. Como $\widehat{s}(\mathbf{x}) \xrightarrow{p} \sigma(\mathbf{x})$ se sigue que para todo $\phi > 0$

$$\begin{aligned}
\mathbb{P}\left(\|V(\widehat{\boldsymbol{\theta}}_n, \widehat{a}_\sigma(\mathbf{x}))\| > \phi\eta\right) &\le \mathbb{P}\left(|\widehat{a}_\sigma(\mathbf{x}) - 1| > \frac{\gamma}{2}\right) + \mathbb{P}\left(\sup_{|a-1|\le\gamma/2} \sup_{D_\eta} \|V(\boldsymbol{\theta}_n, a)\|\right) \\
&\le \mathbb{P}\left(|\widehat{a}_\sigma(\mathbf{x}) - 1| > \frac{\gamma}{2}\right) + \frac{B_\eta}{\phi}\,,
\end{aligned}$$

con $\widehat{a}_\sigma(\mathbf{x}) = \sigma(\mathbf{x})/\widehat{s}(\mathbf{x})$ y $\widehat{\boldsymbol{\theta}}_n = (\widehat{\theta}_1, \ldots, \widehat{\theta}_n)^{\mathrm{T}}$. De esta manera, como $K_{n3} = -(\mathbf{r}-\mathbf{r}_0)^{\mathrm{T}} V(\widehat{\boldsymbol{\theta}}_n, \widehat{a}_\sigma(\mathbf{x}))/\widehat{s}(\mathbf{x})$ y $\widehat{s}(\mathbf{x}) > A$ tenemos que

$$\mathbb{P}\left(\sup_{\mathbf{r}\in\mathcal{V}_\tau} |K_{n3}| > \tau\,\frac{\phi\,\eta}{A}\right) \le \mathbb{P}\left(|\widehat{a}_\sigma(\mathbf{x}) - 1| > \frac{\gamma}{2}\right) + \frac{B_\eta}{\phi}\,, \tag{3.38}$$

Usando que $\widehat{Z}_i = \widehat{V}_i + R(\mathbf{X}_i)\widehat{s}(\mathbf{x})^{-1}$ y $\widehat{\Delta}_i = (\mathbf{r} - \mathbf{r}_0)^{\mathrm{T}}\breve{\mathbf{x}}_{i,\alpha}/\widehat{s}(\mathbf{x})$ obtenemos que

$$\begin{aligned}
K_{n2} &= \frac{1}{n}\sum_{i=1}^n \mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i - \mathbf{x})\xi_i \int_{\widehat{Z}_i}^{\widehat{Z}_i - \widehat{\Delta}_i} \psi'(\widehat{V}_i)(t - \widehat{V}_i) \\
&= \frac{1}{2\widehat{s}^{\,2}(\mathbf{x})}\frac{1}{n}\sum_{i=1}^n \psi'(\widehat{V}_i)\left[r_i^2 - 2R(\mathbf{X}_i)r_i\right]\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i - \mathbf{x})\xi_i \\
&= \frac{1}{2\widehat{s}^{\,2}(\mathbf{x})}\left((\mathbf{r} - \mathbf{r}_0)^{\mathrm{T}}\widehat{\mathbf{M}}_{n1}(\mathbf{r} - \mathbf{r}_0) - 2(\mathbf{r} - \mathbf{r}_0)^{\mathrm{T}}\widehat{\mathbf{M}}_{n2}\right)\,,
\end{aligned}$$

donde $r_i = (\mathbf{r} - \mathbf{r}_0)^{\mathrm{T}}\breve{\mathbf{x}}_{i,\alpha}$, $\widehat{V}_i = \sigma(\mathbf{X}_i)\varepsilon_i/\widehat{s}(\mathbf{x})$. Definamos

$$\mathbf{M}_{n,1} = \frac{1}{n}\sum_{i=1}^n \mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i - \mathbf{x})\xi_i\psi'(V_i)\breve{\mathbf{x}}_{i,\alpha}\breve{\mathbf{x}}_{i,\alpha}^{\mathrm{T}}\,.$$

Argumentos análogos a los el Lema 3.4.1 (con $s(\mathbf{x}) = \sigma(\mathbf{x})$ y $\eta_i = 0$ para todo i) permiten mostrar que

$$\mathbf{M}_{n1} = \mathbb{E}(\psi'(\varepsilon))\, f_{\mathbf{X}}(\mathbf{x})\, r(\mathbf{x})\, \mathbf{S}^{(\alpha)}(1 + o_{\mathbb{P}}(1))\,.$$

Como en la Proposición 3.2.1, por la acotación de ψ', de g_j'' y de r en un entorno de $\mathcal{C}$ es inmediato que

$$\widehat{\mathbf{M}}_{n,2} = \frac{1}{n}\sum_{i=1}^n \mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i - \mathbf{x})\xi_i\psi'(\widehat{V}_i)R(\mathbf{X}_i)\breve{\mathbf{x}}_{i,\alpha} = O_{\mathbb{P}}\left(h_\alpha^{q+1} + \sum_{j\neq\alpha}^d h_j\right)\,.$$

Usando la consistencia de $\widehat{s}(\mathbf{x})$ y el hecho que argumentos análogos a los del Lema 3.4.2 permiten mostrar que $\widehat{\mathbf{M}}_{n,1} - \mathbf{M}_{n,1} = o_{\mathbb{P}}(1)$ nos permiten deducir que

$$K_{n2} = \frac{1}{2\sigma(\mathbf{x})}\mathbb{E}(\psi'(\varepsilon)) f_{\mathbf{X}}(\mathbf{x})\, r(\mathbf{x})\, (\mathbf{r} - \mathbf{r}_0)^{\mathrm{T}}\mathbf{S}^{(\alpha)}(\mathbf{r} - \mathbf{r}_0)(1 + o_{\mathbb{P}}(1)) + O_{\mathbb{P}}\left(h_\alpha^{q+1} + \sum_{j\neq\alpha} h_j\right)\tau\,. \tag{3.39}$$

Como $\mathbf{S}^{(\alpha)}$ es una matriz simétrica definida positiva, llamemos $\lambda > 0$ al menor autovalor $\mathbf{S}^{(\alpha)}$. Luego, como $\mathbb{E}(\psi'(\varepsilon)) > 0$ para cualquier $\mathbf{r} \in \mathcal{V}_\tau$ se cumple

$$\frac{1}{2\sigma(\mathbf{x})}\mathbb{E}(\psi'(\varepsilon))f_{\mathbf{X}}(\mathbf{x})\, r(\mathbf{x})\, (\mathbf{r}-\mathbf{r}_0)^{\mathrm{T}}\mathbf{S}^{(\alpha)}(\mathbf{r}-\mathbf{r}_0) \geq \frac{\lambda}{2\sigma(\mathbf{x})}\mathbb{E}(\psi'(\varepsilon))f_{\mathbf{X}}(\mathbf{x})\, r(\mathbf{x})\, \tau^2 > 0\,. \tag{3.40}$$

La demostración concluye como en la Proposición 3.2.1 usando **C4**, que $\lim_{\eta\to 0} B_\eta = 0$ y combinando (3.35) a (3.40). □

Como en la sección anterior, la Proposición 3.3.1 permite mostrar que el estimador inicial $\widetilde{g}_{\mathrm{M}_q,\alpha}$ de g es consistente. Por otra parte, muestra que $h_\alpha^s(\widehat{\boldsymbol{\beta}}_{s+1}(\mathbf{x}) - g_\alpha^{(s)}(x_\alpha)) \xrightarrow{p} 0$. Para obtener resultados de consistencia sobre $\widehat{g}_{\alpha,\mathrm{M}_q,\alpha}(x)$ necesitaremos mostrar resultados de convergencia uniforme sobre compactos para $\widetilde{g}_{\mathrm{M}_q,\alpha}$.

Para reforzar la dependencia en x_α indicaremos por

$$R_\alpha(X_{i,\alpha}, x_\alpha) = g_\alpha(X_{i,\alpha}) - g_\alpha(x_\alpha) - \sum_{j=1}^{q} g_\alpha^{(j)}(x_\alpha)(X_{i,\alpha} - x_\alpha)^j$$

y por $R(\mathbf{X}_i, \mathbf{x}) = \sum_{j\neq\alpha}\{g_j(X_{i,j}) - g_j(x_j)\} + R_\alpha(X_{i,\alpha}, x_\alpha)$.

Proposición 3.3.2. *Sea $\mathcal{C} \subset \mathcal{S}_f$ un conjunto compacto para el que cumple **A3**. Supongamos que se cumplen **A1**, **A4**, **A6**, **C1′**, **C2′**, **C3**a) y **C5** a **C8**. Entonces, existe una solución $\widehat{\boldsymbol{\beta}}(\mathbf{x})$ de (2.22) tal que $\sup_{\mathbf{x}\in\mathcal{C}} \|\mathbf{H}_\alpha\{\widehat{\boldsymbol{\beta}}(\mathbf{x}) - \boldsymbol{\beta}(\mathbf{x})\}\| \xrightarrow{a.s.} 0$ donde $\boldsymbol{\beta}(\mathbf{x}) = (g(\mathbf{x}), g_\alpha^{(1)}(x_\alpha), \ldots, g_\alpha^{(q)}(x_\alpha))^{\mathrm{T}}$ y $g_\alpha^{(1)} = g'_\alpha$.*

Demostración. La demostración sigue pasos análogos a los dados en la Proposición 3.2.2 adaptándolos a nuestro contexto como en la Proposición 3.3.1.

Sea $\mathbf{r} = (\beta_0, h_\alpha\beta_1, h_\alpha^2\beta_2, \cdots, h_\alpha^q\beta_q)^{\mathrm{T}}$ y $\mathbf{r}_0(\mathbf{x}) = \Big(g(\mathbf{x}), h_\alpha g_\alpha^{(1)}(x_\alpha), \ldots, h_\alpha^q g_\alpha^{(q)}(x_\alpha)\Big)^{\mathrm{T}}$. Recordemos que $\mathcal{K}_{\mathbf{H}_d}(\mathbf{u}) = \Big(\prod_{j=1}^d h_j\Big)^{-1}\prod_{j=1}^d K_j(u_j/h_j)$ y sea $\ell_n(\mathbf{r})$ definida en (3.33). Sea $\mathcal{V}_\tau = \{\mathbf{r} : \|\mathbf{r}\| = \tau\}$ la esfera de centro 0 y de radio τ. Como en la Proposición 3.2.2, basta ver que existe $\mathcal{N}$ tal que $\mathbb{P}(\mathcal{N}) = 0$ y tal que para todo $\omega \notin \mathcal{N}$ se cumple que dado $\nu > 0$, existe $0 < \tau_\nu < 1$ suficientemente pequeño, tal que para cualquier $0 < \tau < \tau_\nu$ si $n \geq n_0$,

$$\inf_{\mathbf{r}\in\mathcal{V}_\tau}\inf_{\mathbf{x}\in\mathcal{C}}\{\ell_n(\mathbf{r}+\mathbf{r}_0(\mathbf{x})) - \ell_n(\mathbf{r}_0(\mathbf{x}))\} > 0\,. \tag{3.41}$$

Observemos que $Y_i - (\mathbf{r}+\mathbf{r}_0(\mathbf{x}))^{\mathrm{T}}\breve{\mathbf{x}}_{i,\alpha} = U_i + g(\mathbf{X}_i) - \mathbf{r}^{\mathrm{T}}\breve{\mathbf{x}}_{i,\alpha} - \mathbf{r}_0^{\mathrm{T}}\breve{\mathbf{x}}_{i,\alpha} = U_i + R(\mathbf{X}_i,\mathbf{x})\breve{\mathbf{x}}_{i,\alpha} - \mathbf{r}^{\mathrm{T}}\breve{\mathbf{x}}_{i,\alpha}$. Llamemos $\widehat{Z}_i(\mathbf{x}) = (U_i + R(\mathbf{X}_i,\mathbf{x})/\widehat{s}(\mathbf{x}) = \widehat{V}_i(\mathbf{x}) + R(\mathbf{X}_i,\mathbf{x})\widehat{s}(\mathbf{x})^{-1}$ con $\widehat{V}_i(\mathbf{x}) = U_i/\widehat{s}(\mathbf{x}) = \sigma(\mathbf{X}_i)\varepsilon_i/\widehat{s}(\mathbf{x})$ y $\widehat{\Delta}_i(\mathbf{x}) = \mathbf{r}^{\mathrm{T}}\breve{\mathbf{x}}_i/\widehat{s}(\mathbf{x})$. Luego, como en la Proposición 3.2.2, tenemos que para todo $\mathbf{r} \in \mathcal{V}_\tau$

$$\ell_n(\mathbf{r}+\mathbf{r}_0(\mathbf{x})) - \ell_n(\mathbf{r}_0(\mathbf{x})) = K_{n1}(\mathbf{x}) + K_{n2}(\mathbf{x}) + K_{n3}(\mathbf{x})\,, \tag{3.42}$$

con

$$
\begin{aligned}
K_{n1}(\mathbf{x}) &= \frac{1}{n}\sum_{i=1}^{n}\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i-\mathbf{x})\xi_i\int_{\widehat{Z}_i(\mathbf{x})}^{\widehat{Z}_i(\mathbf{x})-\widehat{\Delta}_i(\mathbf{x})}\psi(\widehat{V}_i)dt = -\frac{1}{\widehat{s}(\mathbf{x})}\mathbf{r}^{\mathrm{T}}\frac{1}{n}\sum_{i=1}^{n}\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i-\mathbf{x})\xi_i\psi(\widehat{V}_i(\mathbf{x}))\breve{\mathbf{x}}_{i,\alpha}\ ,\\
K_{n2}(\mathbf{x}) &= \frac{1}{n}\sum_{i=1}^{n}\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i-\mathbf{x})\xi_i\int_{\widehat{Z}_i(\mathbf{x})}^{\widehat{Z}_i(\mathbf{x})-\widehat{\Delta}_i(\mathbf{x})}\psi'(\widehat{V}_i(\mathbf{x}))(t-\widehat{V}_i(\mathbf{x}))dt\\
&= \frac{1}{2\widehat{s}^{\,2}(\mathbf{x})}\frac{1}{n}\sum_{i=1}^{n}\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i-\mathbf{x})\xi_i\psi'(\widehat{V}_i(\mathbf{x}))\left[\left(\mathbf{r}^{\mathrm{T}}\breve{\mathbf{x}}_{i,\alpha}\right)^2-2R(\mathbf{X}_i,\mathbf{x})\mathbf{r}^{\mathrm{T}}\breve{\mathbf{x}}_{i,\alpha}\right]\ ,\\
K_{n3}(\mathbf{x}) &= \frac{1}{n}\sum_{i=1}^{n}\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i-\mathbf{x})\xi_i\int_{\widehat{Z}_i(\mathbf{x})}^{\widehat{Z}_i(\mathbf{x})-\widehat{\Delta}_i(\mathbf{x})}\left[\psi(t)-\psi(\widehat{V}_i(\mathbf{x}))-\psi'(\widehat{V}_i(\mathbf{x}))(t-\widehat{V}_i(\mathbf{x}))\right]dt\ .
\end{aligned}
$$

Supongamos haber probado que

$$
\sup_{\mathbf{r}\in\mathcal{V}_\tau}\sup_{\mathbf{x}\in\mathcal{C}}\|K_{n1}(\mathbf{x})\| = \tau\sqrt{\frac{\log n}{n\prod_{j=1}^{d}h_j}}\,\widehat{\Upsilon}_{1,n}\ , \tag{3.43}
$$

$$
K_{n2}(\mathbf{x}) = \frac{1}{2\sigma^2(\mathbf{x})}\mathbb{E}(\psi'(\varepsilon))f_{\mathbf{X}}(\mathbf{x})r(\mathbf{x})\ \mathbf{r}^{\mathrm{T}}\mathbf{S}^{(\alpha)}\mathbf{r}(1+\widehat{\zeta}_n)+\tau\left(h_\alpha^{q+1}+\sum_{j\neq\alpha}h_j\right)\widehat{\Upsilon}_{2,n}\ , \tag{3.44}
$$

$$
\sup_{\mathbf{r}\in\mathcal{V}_\tau}\sup_{\mathbf{x}\in\mathcal{C}}\|\mathcal{K}_{n3}(\mathbf{x})\| = \tau\left(\tau+h_\alpha^{q+1}+\sum_{j\neq\alpha}h_j\right)^2\widehat{\Upsilon}_{3,n}\ . \tag{3.45}
$$

con $\widehat{\Upsilon}_{1,n}=O_{\text{a.co.}}(1)$, $\widehat{\Upsilon}_{3,n}=O_{\text{a.co.}}(1)$ y además $\widehat{\Upsilon}_{1,n}$ y $\widehat{\Upsilon}_{3,n}$ no dependen de τ y donde además $\widehat{\zeta}_n=o_{\text{a.co.}}(1)$ y $\widehat{\Upsilon}_{2,n}=O_{\text{a.co.}}(1)$ no dependen de $\mathbf{x}\in\mathcal{C}$ ni de $\mathbf{r}$ y por lo tanto, de τ. Luego, como $\mathbf{S}^{(\alpha)}$ es una matriz simétrica definida positiva, llamando $\lambda>0$ al menor autovalor de $\mathbf{S}^{(\alpha)}$, el resultado se obtiene utilizando argumentos análogos a los dados en la Proposición 3.2.2 pero usando una acotación como en (3.40).

Luego, la demostración estará completa si probamos (3.43), (3.44) y (3.45).

• Comenzaremos probando (3.43). Definamos $\mathbf{J}_{n,\alpha}(\mathbf{x},a)$ como en (3.36) y observemos que

$$
K_{n1}(\mathbf{x}) = -\frac{1}{\widehat{s}(\mathbf{x})}\mathbf{r}^{\mathrm{T}}\mathbf{J}_{n,\alpha}(\mathbf{x},\widehat{a}_\sigma(\mathbf{x}))\ .
$$

Nuevamente, por (3.15) tenemos que

$$
\mathbb{P}\left(\exists n_0 \text{ tal que } \forall n\geq n_0\ \sup_{\mathbf{r}\in\mathcal{V}_\tau}\sup_{\mathbf{x}\in\mathcal{C}}\|K_{n1}(\mathbf{x})\|<\frac{\tau}{A}\sup_{\mathbf{x}\in\mathcal{C}}\|\mathbf{J}_{n,\alpha}(\mathbf{x},\widehat{a}_\sigma(\mathbf{x}))\|\right)=1\,.
$$

Dado que el Lema 3.4.7, vale también para la $j-$ésima coordenada de $\mathbf{J}_{n,\alpha}(\mathbf{x},\widehat{a}_\sigma(\mathbf{x}))$ se tiene que $\sup_{\mathbf{x}\in\mathcal{C}}\|\mathbf{J}_{n,\alpha}(\mathbf{x},\widehat{a}_\sigma(\mathbf{x}))\|=O_{\text{a.s.}}(\theta_n)$. Luego, hemos probado (3.43).

• Probaremos ahora (3.45). Recordemos que $\widehat{Z}_i(\mathbf{x}) = \widehat{V}_i(\mathbf{x}) + R(\mathbf{X}_i, \mathbf{x})\widehat{s}(\mathbf{x})^{-1}$ y $\widehat{\Delta}_i(\mathbf{x}) = \mathbf{r}^{\mathrm{T}}\check{\mathbf{x}}_{i,\alpha}/\widehat{s}(\mathbf{x})$. Por el teorema del valor medio integral,

$$\begin{aligned} K_{n3}(\mathbf{x}) &= \frac{1}{n}\sum_{i=1}^{n}\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i - \mathbf{x})\xi_i \int_{\widehat{V}_i(\mathbf{x})+R(\mathbf{X}_i,\mathbf{x})\widehat{s}(\mathbf{x})^{-1}}^{\widehat{V}_i(\mathbf{x})+R(\mathbf{X}_i,\mathbf{x})\widehat{s}(\mathbf{x})^{-1}-\widehat{\Delta}_i(\mathbf{x})} \left[\psi(t) - \psi(\widehat{V}_i(\mathbf{x})) - \psi'(\widehat{V}_i(\mathbf{x}))(t - \widehat{V}_i(\mathbf{x}))\right] dt \\ &= -\mathbf{r}^{\mathrm{T}}\frac{1}{n\widehat{s}(\mathbf{x})}\sum_{i=1}^{n}\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i - \mathbf{x})\xi_i \left[\psi\left(\widehat{V}_i(\mathbf{x}) + \widehat{\theta}_i(\mathbf{x})\right) - \psi(\widehat{V}_i(\mathbf{x})) - \psi'(\widehat{V}_i(\mathbf{x}))\widehat{\theta}_i(\mathbf{x})\right]\check{\mathbf{x}}_{i,\alpha}\,, \end{aligned}$$

donde $\widehat{\theta}_i(\mathbf{x})$ es un punto intermedio entre $R(\mathbf{X}_i, \mathbf{x})/\widehat{s}(\mathbf{x})$ y $\{R(\mathbf{X}_i, \mathbf{x}) - \mathbf{r}^{\mathrm{T}}\check{\mathbf{x}}_{i,\alpha}\}/\widehat{s}(\mathbf{x})$. Por simplicidad, si $i \in \{i : \mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i - \mathbf{x}) = 0\}$, redefinamos $\widehat{\theta}_i(\mathbf{x}) = 0$ ya que no cambia el resultado de la suma. Observando que $|X_{i,j} - x_j| \le h_j$ para $i = 1, \cdots, n$ y $j = 1, \cdots, d$, cuando $\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i - \mathbf{x}) \neq 0$ resulta que

$$\sup_{\mathbf{x}\in\mathcal{C}} \max_{i:\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i-\mathbf{x})\neq 0} |R(\mathbf{X}_i, \mathbf{x})| \le A_g \left(\sum_{j\neq\alpha} h_j + h_\alpha^{q+1}\right), \tag{3.46}$$

donde A_g es una contante que depende solamente de $\|g_j^{(1)}\|_\infty$ para $j \neq \alpha$ y $\|g_\alpha^{(q+1)}\|_\infty$. Luego, como se cumple (3.15),

$$\max_{1\le i\le n} |\widehat{\theta}_i(\mathbf{x})| \le A_1 \left(\tau + h_\alpha^{q+1} + \sum_{j\neq\alpha} h_j\right)$$

y se deduce que

$$\mathbb{P}\left(\exists n_0 \text{ tal que } \forall n \ge n_0 \quad \sup_{\mathbf{x}\in\mathcal{C}} \max_{1\le i\le n} |\widehat{\theta}_i(\mathbf{x})| \le A_1 \left(\tau + h_\alpha^{q+1} + \sum_{j\neq\alpha} h_j\right)\right) = 1\,.$$

Como en la Proposición 3.2.2, tenemos que $\widehat{f}(\mathbf{x}) = (1/n)\sum_{j=1}^{n}\mathcal{K}^\star_{\mathbf{H}_d}(\mathbf{x} - \mathbf{X}_j)$ converge a $f_{\mathbf{X}}$ uniformemente en casi todo punto (ver (3.60)) y por **A3** obtenemos que $\sup_{\mathbf{x}\in\mathcal{C}} \widehat{f}(\mathbf{x}) = O_{\text{a.s.}}(1)$, de donde, usando que ψ'' está acotada y que cada coordenada de $\check{\mathbf{x}}_{i,\alpha}$ es menor o igual que 1 cuando $\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i - \mathbf{x}) \neq 0$, se deduce que si $A < \widehat{s}(\mathbf{x})$ para todo $\mathbf{x} \in \mathcal{C}$, entonces

$$\sup_{\mathbf{r}\in\mathcal{V}_\tau}\sup_{\mathbf{x}\in\mathcal{C}} \|\mathcal{K}_{n3}(\mathbf{x})\| \le \tau\frac{\|\psi''\|_\infty}{A\,i(t)} \sup_{\mathbf{x}\in\mathcal{C}} \max_{1\le i\le n} |\widehat{\theta}_i(\mathbf{x})|^2 \sup_{\mathbf{x}\in\mathcal{C}} \frac{1}{n}\sum_{i=1}^{n} |\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i - \mathbf{x})|\,.$$

Por lo tanto, de (3.15) deducimos que

$$\mathbb{P}\left(\exists n_0 \text{ tal que } \forall n \ge n_0 \quad \sup_{\mathbf{r}\in\mathcal{V}_\tau}\sup_{\mathbf{x}\in\mathcal{C}} \|\mathcal{K}_{n3}(\mathbf{x})\| \le \tau\frac{\|\psi''\|_\infty}{A\,i(t)} A_1 \left(\tau + h_\alpha^{q+1} + \sum_{j\neq\alpha} h_j\right)^2 \widehat{\Upsilon}^\star_{3,n}\right) = 1$$

donde $\widehat{\Upsilon}^\star_{3,n} = \sup_{\mathbf{x}\in\mathcal{C}}(1/n)\sum_{i=1}^{n} |\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i - \mathbf{x})| = O_{\text{a.s.}}(1)$ no depende de τ, lo que concluye la demostración de (3.45).

• Por último, probaremos (3.44). Tenemos que

$$\begin{aligned} K_{n2}(\mathbf{x}) &= \frac{1}{2\widehat{s}^2(\mathbf{x})}\frac{1}{n}\sum_{i=1}^{n}\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i-\mathbf{x})\xi_i\psi'(\widehat{V}_i(\mathbf{x}))\left[\left(\mathbf{r}^{\mathrm{T}}\breve{\mathbf{x}}_{i,\alpha}\right)^2-2R(\mathbf{X}_i,\mathbf{x})\mathbf{r}^{\mathrm{T}}\breve{\mathbf{x}}_{i,\alpha}\right] \\ &= \frac{1}{2\widehat{s}^2(\mathbf{x})}\left(\mathbf{r}^{\mathrm{T}}\widehat{\mathbf{M}}_{n1}\mathbf{r}-2\mathbf{r}^{\mathrm{T}}\widehat{\mathbf{M}}_{n2}\right). \end{aligned} \tag{3.47}$$

Para $\mathbf{x}\in\mathcal{C}$ y $a\in\mathcal{I}_\delta=[1-\delta,1+\delta]$ (con $0<\delta<1$), definamos

$$\mathbf{M}_{n1}(\mathbf{x},a)=\frac{1}{n}\sum_{i=1}^{n}\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i-\mathbf{x})\xi_i\psi'(V_i(\mathbf{x})a)\breve{\mathbf{x}}_{i,\alpha}\breve{\mathbf{x}}_{i,\alpha}^{\mathrm{T}}$$

y $\mathbf{M}(\mathbf{x},a)=\mathbb{E}\psi'(\varepsilon a)f_{\mathbf{X}}(\mathbf{x})r(\mathbf{x})\mathbf{S}^{(\alpha)}$. Luego, $\mathbf{M}(\mathbf{x},1)=\mathbb{E}\psi'(\varepsilon)f_{\mathbf{X}}(\mathbf{x})r(\mathbf{x})\mathbf{S}^{(\alpha)}$. Queremos mostrar que

$$\sup_{\mathbf{x}\in\mathcal{C}}\|\widehat{\mathbf{M}}_{n1}(\mathbf{x})-\mathbb{E}\psi'(\varepsilon)f_{\mathbf{X}}(\mathbf{x})r(\mathbf{x})\mathbf{S}^{(\alpha)}\| = o_{\text{a.s.}}(1) \tag{3.48}$$

$$\sup_{\mathbf{x}\in\mathcal{C}}\|\widehat{\mathbf{M}}_{n2}(\mathbf{x})\| = O_{\text{a.s.}}\left(h_\alpha^{q+1}+\sum_{j\neq\alpha}h_j\right). \tag{3.49}$$

Efectivamente, si (3.48) y (3.49) valen, usando (3.15) y que $\sup_{\mathbf{x}\in\mathcal{C}}|\widehat{s}(\mathbf{x})-\sigma(\mathbf{x})|\xrightarrow{a.s.}0$, σ es acotada en $\mathcal{C}$, $i(\sigma)>0$ y reemplazando (3.48) y (3.49) en (3.47) obtenemos que

$$K_{n2}(\mathbf{x})=\frac{1}{2\sigma^2(\mathbf{x})}\mathbb{E}(\psi'(\varepsilon))f_{\mathbf{X}}(\mathbf{x})r(\mathbf{x})\ \mathbf{r}^{\mathrm{T}}\mathbf{S}^{(\alpha)}\mathbf{r}(1+\widehat{\zeta}_n)+\tau\left(h_\alpha^{q+1}+\sum_{j\neq\alpha}h_j\right)\widehat{\Upsilon}_{2,n}\,,$$

donde $\widehat{\zeta}_n=o_{\text{a.s.}}(1)$ y $\widehat{\Upsilon}_{2,n}=O_{\text{a.s.}}(1)$ no dependen de $\mathbf{r}$ y por lo tanto, de τ, ni de $\mathbf{x}\in\mathcal{C}$ pues las convergencias son uniformes sobre $\mathbf{x}$, lo que prueba (3.44).

Como en la Proposición 3.2.2, para probar (3.48) basta ver que para todo $1\le j,k\le d+1$

$$\sup_{\mathbf{x}\in\mathcal{C}}|\widehat{M}_{n1,jk}(\mathbf{x})-M_{jk}(\mathbf{x},1)|=o_{\text{a.s.}}(1)\,, \tag{3.50}$$

donde llamamos $\widehat{M}_{n1,jk}(\mathbf{x})$, $M_{n1,jk}(\mathbf{x},a)$ y $M_{jk}(\mathbf{x},a)$ a la componente (j,k) de las matrices $\widehat{\mathbf{M}}_{n1}(\mathbf{x},a)$, $\mathbf{M}_{n1}(\mathbf{x},a)$ y $\mathbf{M}(\mathbf{x},a)$, respectivamente.

Como $\widehat{a}_\sigma(\mathbf{x})=\sigma(\mathbf{x})/\widehat{s}(\mathbf{x})$, $\widehat{M}_{n1,jk}(\mathbf{x})=M_{n1,jk}(\mathbf{x},\widehat{a}_\sigma(\mathbf{x}))$. Por lo tanto, como

$$\sup_{\mathbf{x}\in\mathcal{C}}|\widehat{M}_{n1,jk}(\mathbf{x})-M_{jk}(\mathbf{x},1)|\le\sup_{\mathbf{x}\in\mathcal{C}}|M_{n1,jk}(\mathbf{x},\widehat{a}_\sigma(\mathbf{x}))-M_{jk}(\mathbf{x},\widehat{a}_\sigma(\mathbf{x}))|+\sup_{\mathbf{x}\in\mathcal{C}}|M_{jk}(\mathbf{x},\widehat{a}_\sigma(\mathbf{x}))-M_{jk}(\mathbf{x},1)|\,,$$
$$\sup_{\mathbf{x}\in\mathcal{C}}|M_{n1,jk}(\mathbf{x},a)-M_{jk}(\mathbf{x},a)|\le\sup_{\mathbf{x}\in\mathcal{C}}|M_{n1,jk}(\mathbf{x},a)-\mathbb{E}M_{n1,jk}(\mathbf{x},a)|+\sup_{\mathbf{x}\in\mathcal{C}}|\mathbb{E}M_{n1,jk}(\mathbf{x},a)-M_{jk}(\mathbf{x},a)|\,,$$

para probar (3.50) basta ver que

(i) $\sup_{\mathbf{x}\in\mathcal{C}}\sup_{a\in\mathcal{I}_\delta}|M_{n1,jk}(\mathbf{x},a)-\mathbb{E}M_{n1,jk}(\mathbf{x},a)|=O_{\text{a.s.}}\left(\sqrt{\log n/(n\prod_{j=1}^{d}h_j)}\right)$.

(ii) $\sup_{\mathbf{x}\in\mathcal{C}} \sup_{a\in\mathcal{I}_\delta} |\mathbb{E}M_{n1,jk}(\mathbf{x},a) - M_{jk}(\mathbf{x},a)| = o(1)$

(iii) $\sup_{\mathbf{x}\in\mathcal{C}} |M_{n1,jk}(\mathbf{x},\widehat{a}_\sigma(\mathbf{x})) - M_{jk}(\mathbf{x},\widehat{a}_\sigma(\mathbf{x}))| = o_{\text{a.s.}}(1)$

(iv) $\sup_{\mathbf{x}\in\mathcal{C}} |M_{jk}(\mathbf{x},\widehat{a}_\sigma(\mathbf{x})) - M_{jk}(\mathbf{x},1)| = o_{\text{a.s.}}(1)$.

(i), (ii), (iii) y (iv) se demuestran utilizando argumentos análogos a los del Lema 3.4.6, la Observación 3.4.1 y a los utilizados en la Proposición 3.2.2.

Por (3.46) y usando que ψ' está acotada y que $i(t) > 0$, se tiene que

$$\sup_{\mathbf{x}\in\mathcal{C}} \|\widehat{\mathbf{M}}_{n2}(\mathbf{x})\| \leq \frac{\|\psi'\|_\infty}{i(t)} A_g \left(h_\alpha^{q+1} + \sum_{j\neq\alpha} h_j \right) \sup_{\mathbf{x}\in\mathcal{C}} \frac{1}{n}\sum_{i=1}^n |\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i - \mathbf{x})| \ .$$

Como antes, por la Observación 3.4.1, $\widehat{f}(\mathbf{x}) = (1/n)\sum_{j=1}^n \mathcal{K}^\star_{\mathbf{H}_d}(\mathbf{x} - \mathbf{X}_j)$ converge a $f_\mathbf{X}$ uniformemente. Por lo tanto, usando **A3** obtenemos que $\sup_{\mathbf{x}\in\mathcal{C}} \widehat{f}(\mathbf{x}) = O_{\text{a.s.}}(1)$ lo que concluye la demostración de (3.28) y por lo tanto, la de la Proposición. □

De la Proposición 3.3.2 obtenemos en forma inmediata el siguiente Teorema que muestra la consistencia del estimador de integración marginal basado en polinomios de orden q.

Teorema 3.2.1. *Supongamos que se cumple **A0** y supongamos que se cumplen **A3**, **A4** y **C5** con $\mathcal{C} = \mathcal{S}_Q \subset \mathcal{S}_f$. Si además valen **A1**, , **A6**, **C1** a **C3**a), **C6** a **C8**, entonces*

a) $\sup_{x\in\mathcal{C}_\alpha} |\widehat{g}_{\alpha,\mathrm{M}_{q,\alpha}}(x) - g_\alpha(x)| \xrightarrow{a.s.} 0$,

b) $\sup_{\mathbf{x}\in\mathcal{C}} |\widehat{g}_{\mathrm{M}_q}(\mathbf{x}) - g(\mathbf{x})| \xrightarrow{a.s.} 0$, *donde* $\widehat{g}_{\mathrm{M}_q}(\mathbf{x}) = \sum_{j=1}^d \widehat{g}_{j,\mathrm{M}_{q,\alpha}}(x_j)$,

c) $\sup_{\mathbf{x}\in\mathcal{C}} h_\alpha^\nu \, |\widehat{g}^{(\nu)}_{\alpha,\mathrm{M}_{q,\alpha}}(x) - g^{(\nu)}_\alpha(x)| \xrightarrow{a.s.} 0$.

3.4 Apéndice

En esta Sección incluímos tres Lemas cuyas conclusiones se utilizaron para probar la Proposición 3.2.1.

Lema 3.4.1. *Sea $\mathbf{x}$ un punto del interior de $\mathcal{S}_f$ y $\mathcal{C}_0$ la clausura de un entorno de $\mathbf{x}$ para el que se cumple **A3** con $\mathcal{C} = \mathcal{C}_0$. Sean $\{\eta_i\}_{i=1}^n$, variables aleatorias tales que $\max_{1\leq i\leq n} |\eta_i| = o_\mathbb{P}(1)$ y sea $V_i = V_i(\mathbf{x}) = s(\mathbf{X}_i)\varepsilon_i/s(\mathbf{x})$ donde $s : \mathbb{R}^d \to \mathbb{R}$ es una función positiva, continua en un entorno de $\mathcal{C}_0$ y sea $i(s) = \inf_{\mathbf{u}\in\mathcal{C}_0} s(\mathbf{u}) > 0$. Supongamos que se cumplen **A1**, **A4** y **C2**, **C3**a) y b), **C4** y **C5**, entonces*

$$\frac{1}{n}\sum_{i=1}^n \psi'(V_i+\eta_i)\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i - \mathbf{x})\xi_i\breve{\mathbf{x}}_i\breve{\mathbf{x}}_i^{\mathrm{T}} = \mathbb{E}\left(\psi'(\varepsilon_1)\right) f_\mathbf{X}(\mathbf{x})\, r(\mathbf{x})\, \mathbf{S}(1+o_\mathbb{P}(1)) \,, \tag{3.51}$$

$$\frac{1}{n}\sum_{i=1}^n \psi'(V_i+\eta_i)R(\mathbf{X}_i)\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i - \mathbf{x})\xi_i\breve{\mathbf{x}}_i = \frac{1}{2}\mathbb{E}\left(\psi'(\varepsilon_1)\right) f_\mathbf{X}(\mathbf{x})\, r(\mathbf{x})\, \mathbf{b}(\mathbf{x})(1+o_\mathbb{P}(1)) \,, \tag{3.52}$$

donde $r(\mathbf{x}) = p(\mathbf{x})/t(\mathbf{x})$ *y* $\mathbf{b}(\mathbf{x}) = \sum_{j=1}^{d} g_j''(x_j)\, h_j^2 \int u_j^2\, \mathcal{K}(\mathbf{u})(1, \mathbf{u}^{\mathrm{T}})^{\mathrm{T}}\, d\mathbf{u}$.

Observemos que si se cumple **C2**, los núcleos K_j son simétricos y por lo tanto, $\int u_j^2\, \mathcal{K}(\mathbf{u})(1, \mathbf{u}^{\mathrm{T}})^{\mathrm{T}}\, d\mathbf{u} = (s_j, 0, \ldots, 0)^{\mathrm{T}}$ con lo cual $\mathbf{b}(\mathbf{x}) = (\sum_{j=1}^{d} g_j''(x_j)\, h_j^2 s_j, 0, \ldots, 0)$.

DEMOSTRACIÓN. Sólo probaremos (3.51) porque (3.52) puede ser demostrada usando los mismos argumentos. Sea $m = 0, 1$ y fijemos $1 \leq j, k \leq d$, entonces

$$\begin{aligned}
& \sum_{i=1}^{n} \psi'(V_i + \eta_i) \mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i - \mathbf{x}) \xi_i \left(\frac{(X_{ij} - x_j)(X_{ik} - x_k)}{h_j h_k} \right)^m \\
= \; & \sum_{i=1}^{n} \psi'(V_i) \mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i - \mathbf{x}) \xi_i \left(\frac{(X_{ij} - x_j)(X_{ik} - x_k)}{h_j h_k} \right)^m \\
& + \sum_{i=1}^{n} [\psi'(V_i + \eta_i) - \psi'(V_i)] \mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i - \mathbf{x}) \xi_i \left(\frac{(X_{ij} - x_j)(X_{ik} - x_k)}{h_j h_k} \right)^m \\
= \; & n\, (L_{n1} + L_{n2}) \; .
\end{aligned}$$

Observemos que $L_{n1} - \mathbb{E}(L_{n1}) \xrightarrow{p} 0$, luego basta calcular $\mathbb{E}(L_{n1})$ y mostrar que $L_{n2} \xrightarrow{p} 0$.

Sea

$$\gamma_{\mathbf{x}}(\mathbf{v}) = \mathbb{E}\left(\psi' \left(\frac{s(\mathbf{X}_1)\varepsilon_1}{s(\mathbf{x})} \right) | \mathbf{X}_1 = \mathbf{v} \right) = \mathbb{E}\left(\psi' \left(\frac{s(\mathbf{v})\varepsilon_1}{s(\mathbf{x})} \right) \right) ,$$

luego $\gamma_{\mathbf{x}}(\mathbf{x}) = \mathbb{E}\psi'(\varepsilon_1)$. Como $\xi_i = \delta_i / t(\mathbf{X}_i)$ y δ_i y ε_i son condicionalmente independientes dado $\mathbf{X}_i$, resulta que

$$\begin{aligned}
\mathbb{E}L_{n1} &= \frac{1}{n} \mathbb{E}\left[\sum_{i=1}^{n} \psi'(V_i) \mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i - \mathbf{x}) \xi_i \left(\frac{(X_{ij} - x_j)(X_{ik} - x_k)}{h_j h_k} \right)^m \right] \\
&= \mathbb{E}\left[\mathbb{E}\left\{ \psi'(V_1) \mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_1 - \mathbf{x}) \xi_1 \left(\frac{(X_{1j} - x_j)(X_{1k} - x_k)}{h_j h_k} \right)^m | \mathbf{X}_1 \right\} \right] \\
&= \mathbb{E}\left[\gamma_{\mathbf{x}}(\mathbf{X}_1) \frac{p(\mathbf{X}_1)}{t(\mathbf{X}_1)} \mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_1 - \mathbf{x}) \left(\frac{(X_{1j} - x_j)(X_{1k} - x_k)}{h_j h_k} \right)^m \right] \\
&= \int \gamma_{\mathbf{x}}(\mathbf{v})\, r(\mathbf{v})\, \mathcal{K}_{\mathbf{H}_d}(\mathbf{v} - \mathbf{x}) \left(\frac{(v_j - x_j)(v_k - x_k)}{h_j h_k} \right)^m f_{\mathbf{X}}(\mathbf{v})\, d\mathbf{v} \\
&= \int \gamma_{\mathbf{x}}(\mathbf{H}_d\, \mathbf{u} + \mathbf{x})\, r(\mathbf{H}_d\, \mathbf{u} + \mathbf{x})\, f_{\mathbf{X}}(\mathbf{H}_d\, \mathbf{u} + \mathbf{x})\, u_j^m u_k^m \prod_{\ell=1}^{d} K_\ell(u_\ell)\, d\mathbf{u} \; .
\end{aligned}$$

Usando que $f_{\mathbf{X}}$, r y $\gamma_{\mathbf{x}}$ son funciones continuas y que $\int K_j(u) du = 1$, obtenemos que

$$\mathbb{E}L_{n1} = \gamma_{\mathbf{x}}(\mathbf{x})\, r(\mathbf{x})\, f_{\mathbf{X}}(\mathbf{x}) \int u_j^m u_k^m \mathcal{K}(\mathbf{u})\, d\mathbf{u} (1 + o_{\mathbb{P}}(1)) = \mathbb{E}\psi'(\varepsilon_1)\, r(\mathbf{x})\, f_{\mathbf{X}}(\mathbf{x}) \nu_{jk}^{(m)} (1 + o(1))$$

donde $\nu_{jk}^{(0)} = s_{00} = 1$ y $\nu_{jk}^{(1)} = s_{jk} = \int u_j u_k \int \mathcal{K}(\mathbf{u}) d\mathbf{u}$. Luego,

$$L_{n1} = \mathbb{E}\psi'(\varepsilon_1)\, r(\mathbf{x})\, f_{\mathbf{X}}(\mathbf{x}) \nu_{jk}^{(m)} (1 + o_{\mathbb{P}}(1)) \, .$$

Falta ver que

$$L_{n2} = o_{\mathbb{P}}(1) \ . \tag{3.53}$$

Para cualquier $\eta > 0$, sea $\boldsymbol{\theta}_n = (\theta_1, \ldots, \theta_n)^{\mathrm{T}}$ y $D_\eta = \{\boldsymbol{\theta}_n : |\theta_i| < \eta, \forall i \leq n\}$,

$$V(\boldsymbol{\theta}_n) = \frac{1}{n}\sum_{i=1}^{n}\left[\psi'(V_i+\theta_i) - \psi'(V_i)\right]\xi_i \mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i - \mathbf{x})\left(\frac{(X_{ij}-x_j)(X_{ik}-x_k)}{h_j h_k}\right)^m .$$

Luego, usando que K_j tiene soporte en $[-1,1]$, obtenemos que $|\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i - \mathbf{x})(X_{ij}-x_j)(X_{ik}-x_k)| \leq h_j h_k |\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i - \mathbf{x})|$ de donde

$$\sup_{D_\eta}|V(\boldsymbol{\theta}_n)| \leq \frac{1}{n}\sum_{i=1}^{n}\sup_{D_\eta}\left[\psi'(V_i+\theta_i) - \psi'(V_i)\right]\ |\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i - \mathbf{x})|\,\xi_i \ .$$

Observemos que $V_i = a_i(\mathbf{x})\varepsilon_i$ donde $a_i(\mathbf{x}) = s(\mathbf{X}_i)/s(\mathbf{x})$. Como s es una función continua y $|X_{ij} - x_j| \leq h_j$ si $\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i - \mathbf{x}) \neq 0$, pues K_j tiene soporte en $[-1,1]$, tenemos que $|s(\mathbf{X}_i) - s(\mathbf{x})| \leq i(\sigma)\gamma$, para todo $i = 1, \ldots, n$ tal que $\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i - \mathbf{x}) \neq 0$ y n suficientemente grande. Por lo tanto, si n es suficientemente grande, se cumple que $|a_i(\mathbf{x}) - 1| \leq \gamma$. Por lo tanto, usando la hipótesis **C3**b), tenemos que si n es suficientemente grande

$$\mathbb{E}\left[\sup_{D_\eta}|V(\boldsymbol{\theta}_n)|\right] \leq b_\eta \mathbb{E}\,|\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_1 - \mathbf{x})|\ r(\mathbf{X}_1) \leq A b_\eta$$

donde $A = 2\ r(\mathbf{x}) f_X(\mathbf{x}) \int |\mathcal{K}(\mathbf{u})| d\mathbf{u}$ y b_η es una sucesión de números positivos que converge a cero cuando $\eta \to 0$.

Dado que $\max_{1\leq i\leq n}|\eta_i| = o_{\mathbb{P}}(1)$, se sigue que $L_{n2} = V(\boldsymbol{\eta}_n) = o_{\mathbb{P}}(1)$ con $\boldsymbol{\eta}_n = (\eta_1, \ldots, \eta_n)^{\mathrm{T}}$ de donde se obtiene (3.53). □

Lema 3.4.2. *Sea* $\mathbf{x}$ *un punto del interior de* $\mathcal{S}_f$ *y* $\mathcal{C}_0$ *la clausura de un entorno de* $\mathbf{x}$ *para el que se cumple* ***A3*** *con* $\mathcal{C} = \mathcal{C}_0$. *Sea* $V_i = V_i(\mathbf{x}) = \sigma(\mathbf{X}_i)\varepsilon_i/\sigma(\mathbf{x})$, $\widehat{V}_i = \sigma(\mathbf{X}_i)\varepsilon_i/\widehat{s}(\mathbf{x})$, *donde* $\widehat{s}(\mathbf{x}) \xrightarrow{p} \sigma(\mathbf{x})$. *Supongamos que se cumplen* ***A1***, ***A4*** *y* ***C2*** *a* ***C4***, *entonces*

$$\frac{1}{n}\sum_{i=1}^{n}\psi'(\widehat{V}_i)\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i - \mathbf{x})\xi_i\breve{\mathbf{x}}_i\breve{\mathbf{x}}_i^{\mathrm{T}} - \frac{1}{n}\sum_{i=1}^{n}\psi'(V_i)\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i - \mathbf{x})\xi_i\breve{\mathbf{x}}_i\breve{\mathbf{x}}_i^{\mathrm{T}} = o_{\mathbb{P}}(1) \ . \tag{3.54}$$

DEMOSTRACIÓN. Sea $m = 0, 1$ y fijemos $1 \leq j, k \leq d$ y definamos

$$\begin{aligned}
\widehat{M}_{m,jk} &= \frac{1}{n}\sum_{i=1}^{n}\psi'(\widehat{V}_i)\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i - \mathbf{x})\xi_i\left(\frac{(X_{ij}-x_j)(X_{ik}-x_k)}{h_j h_k}\right)^m \\
M_{m,jk} &= \frac{1}{n}\sum_{i=1}^{n}\psi'(V_i)\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i - \mathbf{x})\xi_i\left(\frac{(X_{ij}-x_j)(X_{ik}-x_k)}{h_j h_k}\right)^m \\
E_{m,jk} &= \frac{1}{n}\sum_{i=1}^{n}\psi'(\varepsilon_i)\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i - \mathbf{x})\xi_i\left(\frac{(X_{ij}-x_j)(X_{ik}-x_k)}{h_j h_k}\right)^m .
\end{aligned}$$

Usando el Lema 3.4.1 con $s(\mathbf{x}) = \sigma(\mathbf{x})$ y $s(\mathbf{x}) = 1$ obtenemos que $M_{m,jk} - E_{m,jk} \xrightarrow{p} 0$. Por lo tanto, basta probar que $\widehat{M}_{m,jk} - E_{m,jk} \xrightarrow{p} 0$.

Observemos que $\widehat{M}_{m,jk} - E_{m,jk} = V_n(\widehat{\boldsymbol{\theta}}(\mathbf{x}))$ donde $\widehat{\boldsymbol{\theta}}(\mathbf{x}) = \left(\widehat{\theta}_1(\mathbf{x}), \ldots, \widehat{\theta}_n(\mathbf{x})\right)^{\mathrm{T}}$ con $\widehat{\theta}_i(\mathbf{x}) = \sigma(\mathbf{X}_i)/\widehat{s}(\mathbf{x})$ y

$$V_n(\boldsymbol{\theta}) = \sum_{i=1}^{n} \left[\psi'(\theta_i\,\varepsilon_i) - \psi'(\varepsilon_i)\right] \mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i - \mathbf{x})\xi_i \left(\frac{(X_{ij} - x_j)(X_{ik} - x_k)}{h_j h_k}\right)^m .$$

Dado $\eta > 0$, consideremos como en el Lema 3.4.1, el conjunto $D_\eta = \{\boldsymbol{\theta}_n : |\theta_i| < \eta, \forall\, i \leq n\}$, donde $\boldsymbol{\theta}_n = (\theta_1, \ldots, \theta_n)^{\mathrm{T}}$. Luego, usando que K_j tiene soporte en $[-1, 1]$, obtenemos que $|\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i - \mathbf{x})(X_{ij} - x_j)(X_{ik} - x_k)| \leq h_j h_k |\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i - \mathbf{x})|$ con lo cual

$$\sup_{D_\eta} |V(\boldsymbol{\theta}_n)| \leq \frac{1}{n} \sum_{i=1}^{n} \sup_{D_\eta} \left[\psi'(\theta_i\,\varepsilon_i) - \psi'(\varepsilon_i)\right] \, |\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i - \mathbf{x})|\xi_i \, .$$

Por lo tanto, usando la hipótesis **C3**c), tenemos que si n es suficientemente grande

$$\mathbb{E}\left[\sup_{D_\eta} |V(\boldsymbol{\theta}_n)|\right] \leq b_\eta \mathbb{E}\left[\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_1 - \mathbf{x}) r(\mathbf{X}_1)\right] \leq B_\eta$$

donde $B_\eta = A b_\eta$ con A una constante independiente de η y de n y b_η es una sucesión de números positivos que tiende a cero cuando $\eta \to 0$. Por lo tanto, para todo $\nu > 0$ se cumple que

$$\mathbb{P}\left(\sup_{D_\eta} |V(\boldsymbol{\theta}_n)| > \nu\right) \leq \frac{B_\eta}{\nu} \, .$$

Observemos que $\widehat{\theta}_i(\mathbf{x}) = a_i(\mathbf{x})\,\sigma(\mathbf{x})/\widehat{s}(\mathbf{x})$ donde $a_i(\mathbf{x}) = \sigma(\mathbf{X}_i)/\sigma(\mathbf{x})$. Como σ es una función continua y $|X_{ij} - x_j| \leq h_j$ si $\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i - \mathbf{x}) \neq 0$, pues K_j tiene soporte en $[-1, 1]$, tenemos que $|\sigma(\mathbf{X}_i) - \sigma(\mathbf{x})| \leq i(\sigma)\eta/4$, para todo $i = 1, \ldots, n$ tal que $\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i - \mathbf{x}) \neq 0$ y n suficientemente grande. Por lo tanto, si n es suficientemente grande, se cumple que $|a_i(\mathbf{x}) - 1| \leq \eta/4$. Por otra parte, $\sigma(\mathbf{x})/\widehat{s}(\mathbf{x}) \xrightarrow{p} 1$, con lo cual, para todo $0 < \eta \leq 1/2$ se cumple que

$$\lim_{n\to\infty} \mathbb{P}\left(\left|\frac{\sigma(\mathbf{x})}{\widehat{s}(\mathbf{x})} - 1\right| > \frac{\eta}{2}\right) = 0 \tag{3.55}$$

de donde

$$\lim_{n\to\infty} \mathbb{P}(\widehat{\boldsymbol{\theta}}(\mathbf{x}) \notin D_\eta) = \lim_{n\to\infty} \mathbb{P}\left(\max_{1\leq i\leq n} |\widehat{\theta}_i(\mathbf{x}) - 1| > \eta\right) = 0 \, .$$

Sean $0 < \nu \leq 1/2$ y sea $0 < \eta_k$ una sucesión que converge a 0, entonces tenemos que para todo $k \geq 1$,

$$\begin{aligned}
\mathbb{P}\left(|V(\widehat{\boldsymbol{\theta}}_n(\mathbf{x}))| > \nu\right) &\leq \mathbb{P}\left(\sup_{D_{\eta_k}} |V(\boldsymbol{\theta}_n)| > \nu\right) + \mathbb{P}\left(\max_{1\leq i\leq n} |\widehat{\theta}_i(\mathbf{x}) - 1| > \eta_k\right) \\
&\leq 2\frac{B_{\eta_k}}{\nu} + \mathbb{P}\left(\max_{1\leq i\leq n} |\widehat{\theta}_i(\mathbf{x}) - 1| > \eta_k\right) .
\end{aligned}$$

Usando (3.55) obtenemos que

$$\limsup_{n\to\infty} \mathbb{P}\left(|V(\widehat{\boldsymbol{\theta}}_n(\mathbf{x}))| > \nu\right) \leq \frac{B_{\eta_k}}{\nu}\,,$$

de donde se obtiene que $V(\widehat{\boldsymbol{\theta}}_n(\mathbf{x})) \xrightarrow{p} 0$ pues $a_{\eta_k} \to 0$ cuando $k \to \infty$. □

Lema 3.4.3. *Sea* $\mathbf{x}$ *un punto del interior de* $\mathcal{S}_f$ *y* $\mathcal{C}_0$ *la clausura de un entorno de* $\mathbf{x}$ *para el que se cumple* ***A3*** *con* $\mathcal{C} = \mathcal{C}_0$. *Sean* $V_i = V_i(\mathbf{x}) = \sigma(\mathbf{X}_i)\varepsilon_i/\sigma(\mathbf{x})$, $\widehat{a}_\sigma(\mathbf{x}) = \sigma(\mathbf{x})/\widehat{s}(\mathbf{x})$ *donde* $\widehat{s}(\mathbf{x}) \xrightarrow{p} \sigma(\mathbf{x})$. *Supongamos que se cumplen* ***A1***, ***A4*** *y* ***C2***, ***C3***a) *y* ***C4***, *entonces*

$$\widehat{\mathbf{D}}_n(\mathbf{x}) = \sqrt{n\prod_{j=1}^{d} h_j}\,\{\widehat{s}(\mathbf{x})\mathbf{J}_n(\mathbf{x},\widehat{a}_\sigma(\mathbf{x})) - \sigma(\mathbf{x})\mathbf{J}_n(\mathbf{x},1)\} \xrightarrow{p} 0\,. \tag{3.56}$$

DEMOSTRACIÓN. La demostración sigue argumentos análogos a los considerados por Boente y Fraiman (1990). Sea $U_i = \sigma(\mathbf{X}_i)\varepsilon_i$ y definamos $\psi_a(t) = a\psi(t/a)$. Luego como $V_i = U_i/\sigma(\mathbf{x})$ y $V_i\widehat{a}_\sigma(\mathbf{x}) = U_i/\widehat{s}(\mathbf{x})$ tenemos que

$$\widehat{s}(\mathbf{x})\mathbf{J}_n(\mathbf{x},\widehat{a}_\sigma(\mathbf{x})) - \sigma(\mathbf{x})\mathbf{J}_n(\mathbf{x},1) = \frac{1}{n}\sum_{i=1}^{n}\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i-\mathbf{x})\xi_i\,\left(\psi_{\widehat{s}(\mathbf{x})}(U_i) - \psi_{\sigma(\mathbf{x})}(U_i)\right)\breve{\mathbf{x}}_i$$

de donde

$$\widehat{\mathbf{D}}_n(\mathbf{x}) = \left(n\prod_{j=1}^{d} h_j\right)^{-\frac{1}{2}}\sum_{i=1}^{n}\left\{\prod_{j=1}^{d} K_j\left(\frac{X_{ij}-x_j}{h_j}\right)\right\}\xi_i\left[\psi_{\widehat{s}(\mathbf{x})}(U_i) - \psi_{\sigma(\mathbf{x})}(U_i)\right]\breve{\mathbf{x}}_i\,.$$

Probaremos que cada componente de $\widehat{\mathbf{D}}_n(\mathbf{x})$ converge a 0 en probabilidad. Indiquemos por $\breve{\mathbf{x}}_i = (\breve{x}_{i,1},\ldots,\breve{x}_{i,d+1})^{\mathrm{T}}$. Observemos que si $\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i-\mathbf{x}) \neq 0$, $|\breve{x}_{i,m}| \leq 1$ pues K_j tienen soporte en $[-1,1]$.

Sea $1 \leq m \leq d+1$ fijo, e indiquemos por $\widehat{\mathbf{D}}_{n,m}(\mathbf{x})$ a la componente m de $\widehat{\mathbf{D}}_n(\mathbf{x})$. Mostraremos que $\widehat{\mathbf{D}}_{n,m}(\mathbf{x}) \xrightarrow{p} 0$.

Para todo $\lambda > 0$, definamos $H_\lambda(u) = \psi_{\sigma(\mathbf{x})+\lambda}(u) - \psi_{\sigma(\mathbf{x})}(u)$, $I_\lambda(u) = \psi_{\sigma(\mathbf{x})-\lambda}(u) - \psi_{\sigma(\mathbf{x})}(u)$,

$$\begin{aligned} J_n^+(\lambda) &= \frac{1}{\sqrt{n\prod_{j=1}^{d} h_j}}\sum_{i=1}^{n}\left\{\prod_{j=1}^{d} K_j\left(\frac{X_{ij}-x_j}{h_j}\right)\right\}H_\lambda(U_i)\xi_i\breve{x}_{i,m}\,,\\ J_n^-(\lambda) &= \frac{1}{\sqrt{n\prod_{j=1}^{d} h_j}}\sum_{i=1}^{n}\left\{\prod_{j=1}^{d} K_j\left(\frac{X_{ij}-x_j}{h_j}\right)\right\}I_\lambda(U_i)\xi_i\breve{x}_{i,m}\,. \end{aligned}$$

Observemos que $\mathbb{E}(J_n^+(\lambda)) = \mathbb{E}(J_n^-(\lambda)) = 0$ pues vale (2.7) y $\mathbb{E}\psi(\varepsilon_i a) = 0$ para todo a.

Observemos que si $\widehat{\lambda}(\mathbf{x}) = \widehat{s}(\mathbf{x}) - \sigma(\mathbf{x}) > 0$, entonces $\widehat{\mathbf{D}}_{n,m}(\mathbf{x}) = J_n^+(\widehat{\lambda}(\mathbf{x}))$, mientras que si $\widehat{\lambda}(\mathbf{x}) = \widehat{s}(\mathbf{x}) - \sigma(\mathbf{x}) < 0$, $\widehat{\mathbf{D}}_{n,m}(\mathbf{x}) = J_n^-(\widehat{\lambda}(\mathbf{x}))$. Luego, como $\widehat{\lambda}(\mathbf{x}) \xrightarrow{p} 0$, basta probar que

$$\lim_{n\to\infty} \lim_{d\to 0} P\left(\sup_{0\le\lambda\le d} \|J_n^+(\lambda)\| > \varepsilon \right) = 0 \tag{3.57}$$

$$\lim_{n\to\infty} \lim_{d\to 0} P\left(\sup_{0\le\lambda\le d} \|J_n^-(\lambda)\| > \varepsilon \right) = 0 \tag{3.58}$$

Probaremos solamente (3.57) pues (3.58) se obtiene en forma análoga.

Para probar (3.57) basta con ver que, la sucesión de variables aleatorias en el espacio $\mathcal{C}([0,1])$, $J_n^+(\lambda)$ es ajustada.

De acuerdo al Teorema 12.3 de Billingsley (1968) es suficiente probar que

a) La sucesión $\{J_n^+(0)\}$ es ajustada.

b) Existen constantes $\gamma > 0$ y $\alpha > 1$ y una función no decreciente $\widetilde{F}$ en $[0,1]$ tal que

$$\mathbb{E}\left(\left|J_n^+(\lambda_2) - J_n^+(\lambda_1)\right|^\gamma\right) \le \left(\widetilde{F}(\lambda_2) - \widetilde{F}(\lambda_1)\right)^\alpha$$

para todo $0 \le \lambda_1 < \lambda_2$ y n suficientemente grande.

a) vale pues $J_n^+(0) = 0$.

b) Observemos que como $\psi_a(t) = a\psi(t/a)$ y $\zeta(t) = u\psi'(u)$

$$\frac{\partial \psi_{\sigma(\mathbf{x})+\lambda}(u)}{\partial\lambda} = \psi\left(\frac{u}{\sigma(\mathbf{x})+\lambda}\right) - (\sigma(\mathbf{x})+\lambda)\psi'\left(\frac{u}{\sigma(\mathbf{x})+\lambda}\right)\frac{u}{(\sigma(\mathbf{x})+\lambda)^2} = \psi\left(\frac{u}{\sigma(\mathbf{x})+\lambda}\right) - \zeta\left(\frac{u}{\sigma(\mathbf{x})+\lambda}\right),$$

por lo tanto, si $0 < \lambda_1 < \lambda_2$ y si $\widetilde{\lambda}$ es un punto intermedio entre λ_1 y λ_2

$$\begin{aligned} |H_{\lambda_2}(u) - H_{\lambda_1}(u)| &= |\psi_{\sigma(\mathbf{x})+\lambda_2}(u) - \psi_{\sigma(\mathbf{x})+\lambda_1}(u)| \\ &= \left| \frac{\partial \psi_{\sigma(\mathbf{x})+\lambda}(u)}{\partial\lambda}\Big|_{\lambda=\widetilde{\lambda}} (\lambda_2 - \lambda_1) \right| \\ &\le (\|\psi\|_\infty + \|\zeta\|_\infty)(\lambda_2 - \lambda_1) = \widetilde{F}(\lambda_2) - \widetilde{F}(\lambda_1) \end{aligned}$$

donde $\widetilde{F}(\lambda) = (\|\zeta\|_\infty + \|\psi\|_\infty)\lambda$.

Denotemos por

$$A_{i,n} = \prod_{\ell=1}^{d} K_\ell\left(\frac{X_{i\ell} - x_\ell}{h_\ell}\right) \xi_i \left(H_{\lambda_2}(U_i) - H_{\lambda_1}(U_i)\right) \breve{x}_{im}\,.$$

Como $|\breve{x}_{i,m}| \le 1$ si $\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i - \mathbf{x}) \ne 0$. Por otra parte, si $n \ge n_0$, tenemos que $\{\mathbf{u} : |u_j - x_j| \le h_j\} \subset \mathcal{C}_0$ por lo tanto, si $\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i - \mathbf{x}) \ne 0$ y $n \ge n_0$, $\mathbf{X}_i \in \mathcal{C}_0$ de donde $\xi_i \le 1/t(\mathbf{X}_i) \le 1/i(t)$. Por lo tanto, como K_j es acotada, si llamamos $C = (1/i(t)) \prod_{\ell=1}^{d} \|K_\ell\|_\infty$

$$|A_{i,n}| \le C(\widetilde{F}(\lambda_2) - \widetilde{F}(\lambda_1))\,.$$

Por otra parte, como

$$\mathbb{E}\left[\prod_{\ell=1}^{d} K_\ell\left(\frac{X_{i\ell}-x_\ell}{h_\ell}\right)\right] = \prod_{\ell=1}^{d} h_\ell \int \prod_{\ell=1}^{d} K_\ell\left(u_\ell\right) f_{\mathbf{X}}(\mathbf{x}+\mathbf{H}_d\,\mathbf{u})d\mathbf{u}\,,$$

$\mathbf{y}=\mathbf{x}+\mathbf{H}_d\,\mathbf{u}\in\mathcal{C}_0$ si $n\geq n_0$, $f_{\mathbf{X}}$ es acotada en $\mathcal{C}_0$ y $\int K_j(u)du=1$ tenemos que

$$\begin{aligned}
\mathbb{E}\left(A_{i,n}^2\right) &\leq \left(\widetilde{F}(\lambda_2)-\widetilde{F}(\lambda_1)\right)^2 \mathbb{E}\left[\prod_{\ell=1}^{d} K_\ell^2\left(\frac{X_{i\ell}-x_\ell}{h_\ell}\right)\xi_i^2\breve{x}_{im}^2\right]\\
&\leq \left(\widetilde{F}(\lambda_2)-\widetilde{F}(\lambda_1)\right)^2 \frac{1}{i^2(t)}\prod_{\ell=1}^{d}\|K_\ell\|_\infty\,\mathbb{E}\left[\prod_{\ell=1}^{d} K_\ell\left(\frac{X_{i\ell}-x_\ell}{h_\ell}\right)\right]\\
&\leq C_1\left(\widetilde{F}(\lambda_2)-\widetilde{F}(\lambda_1)\right)^2\prod_{\ell=1}^{d} h_\ell
\end{aligned}$$

donde $C_1=(1/i^2(t))\prod_{\ell=1}^{d}\|K_\ell\|_\infty \sup_{\mathbf{u}\in\mathcal{C}_0} f_{\mathbf{X}}(\mathbf{u})$ para todo $n\geq n_0$.

Finalmente, como $\mathbb{E}(A_{i,n})=0$, por la independencia tenemos que

$$\mathbb{E}\left(J_n^+(\lambda_2)-J_n^+(\lambda_1)\right)^2=\frac{1}{n\prod_{\ell=1}^{d}h_\ell}\mathrm{VAR}\left(A_{i,n}\right)=\frac{1}{n\prod_{\ell=1}^{d}h_\ell}\sum_{i=1}^{n}\mathbb{E}\left(A_{i,n}^2\right)\leq C_1\left(\widetilde{F}(\lambda_2)-\widetilde{F}(\lambda_1)\right)^2$$

para $n\geq n_0$, lo que concluye la demostración b) como queríamos. □

El siguiente Lema corresponde a la desigualdad exponencial para variables acotadas y puede verse por ejemplo en el Corolario A.9 de Ferraty y Vieu (2006) y será necesaria para probar el Lema 3.4.5 que es un paso previo en la demostración del Lema 3.4.7.

Lema 3.4.4. *Sea $\{Z_i\}_{i\geq 1}$ una sucesión de variables aleatorias independientes, $\mathbb{E}Z_i=0$. Si existe $M<\infty$ tal que $|Z_1|\leq M$ y si $\sigma^2=\mathbb{E}Z_i^2$, entonces para todo $\varepsilon>0$ se tiene que*

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

Dado un conjunto compacto $\mathcal{C}\subset\mathbb{R}^d$, indicaremos por $N_\rho(\mathcal{C})$ el mínimo número de bolas de radio ρ necesarias para cubrir a $\mathcal{C}$, es decir, $\mathcal{C}\in\bigcup_{k=1}^{N_\rho(\mathcal{C})}\mathcal{B}(\mathbf{x}_k,\rho)$donde $\mathcal{B}(\mathbf{x}_k,\rho)=\{\mathbf{y}\in\mathbb{R}^d:\|\mathbf{y}-\mathbf{x}_k\|\leq\rho\}$ indica la bola de centro $\mathbf{x}_k$ y radio ρ. Es sabido que $N_\rho(\mathcal{C})\leq A_1/\rho^d$ donde la constante A_1 no depende de ρ.

Lema 3.4.5. *Sea $\mathcal{C}\subset\mathbb{R}^d$un conjunto compacto de interior no vacío. Sea $W_i=f(Y_i,\mathbf{X}_i,\delta_i)$ una sucesión de variables aleatorias tales que $|W_i|\leq M$, para todo i. Definamos para cada $\mathbf{x}\in\mathbb{R}^d$ $S_n(\mathbf{x})=(1/n)\sum_{i=1}^{n}\left(G_i(\mathbf{x})-\mathbb{E}G_i(\mathbf{x})\right)$ donde, para cada $1\leq i\leq n$, $G_i(\mathbf{x})=\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i-\mathbf{x})W_i\breve{x}_{i,j}^{m}\breve{x}_{i,\ell}^{\widetilde{m}}$ con $m,\widetilde{m}=0,1$ y $1\leq j,\ell\leq d+1$ fijos y $\breve{\mathbf{x}}_i=(\breve{x}_{i,1},\ldots,\breve{x}_{i,d+1})^{\mathrm{T}}$. Entonces, bajo **A3**, **A6**, **C2** y **C8** se tiene que*

a) Existen $b_1>0$ y $b_2>0$ tal que para todo $\eta>0$

$$\sup_{\mathbf{x}\in\mathcal{C}}\mathbb{P}\left(|S_n(\mathbf{x})|>\eta\right)\leq 2\exp\left\{-\frac{\eta^2 n\prod_{j=1}^{d}h_j}{b_1+b_2\eta}\right\}$$

b) *Sean θ_n y $\rho = \rho_n$ sucesiones numéricas no negativas que tienden a cero, tales que $\theta_n^{-1} A_h \rho \leq M_2$, para todo $n \geq 1$ y $\rho_n \left\{\prod_{j=1}^n h_j\right\}^{-1} \to 0$. Luego, existen $b_3 > 0$ y $b_4 > 0$ y una constante $C_0 > 0$ tales que para toda $C > C_0$ y para todo $n \geq n_0$,*

$$\mathbb{P}\left(\theta_n^{-1} \sup_{\mathbf{x}\in\mathcal{B}(\mathbf{x}_k,\rho)\cap\mathcal{C}} \left|\widetilde{S}_k(\mathbf{x})\right| > C\right) \leq 2\exp\left\{-\frac{C^2\theta_n^2 n \prod_{j=1}^d h_j}{b_3 A_h^2 \rho^2 + b_4 C\theta_n A_h \rho}\right\}$$

donde $\widetilde{S}_k(\mathbf{x}) = (1/n)\sum_{i=1}^n G_i(\mathbf{x}) - \mathbb{E}G_i(\mathbf{x}) - G_i(\mathbf{x}_k) + \mathbb{E}G_i(\mathbf{x}_k)$

c) *Sean θ_n y $\rho = \rho_n$ sucesiones numéricas no negativas que tienden a cero, tales que $\theta_n^{-1} A_h \rho \leq M_2$ y $\rho_n \left\{\prod_{j=1}^n h_j\right\}^{-1} \to 0$. Luego, existen b_1, b_2, b_3 y b_4 constantes positivas y una constante $C_0 > 0$ tales que para toda $C > C_0$ y para todo $n \geq n_0$,*

$$\mathbb{P}\left(\theta_n^{-1} \sup_{\mathbf{x}\in\mathcal{C}} |S_n(\mathbf{x})| > C\right) \leq 2\, N_\rho(\mathcal{C}) \left\{\exp\left\{-\frac{C^2\theta_n^2 n \prod_{j=1}^d h_j}{4b_1 + 2b_2 C\theta_n}\right\} + \exp\left\{-\frac{C^2\theta_n^2 n \prod_{j=1}^d h_j}{4b_3 A_h^2\rho^2 + 2b_4\, C\theta_n A_h \rho}\right\}\right\}$$

d) *Sea $\theta_n = \sqrt{\log n/(n\prod_{j=1}^n h_j)}$. Luego, existe C tal que*

$$\sum_{n\geq 1} \mathbb{P}\left(\theta_n^{-1} \sup_{\mathbf{x}\in\mathcal{C}} |S_n(\mathbf{x})| > C\right) < \infty\,,$$

es decir, $\sup_{\mathbf{x}\in\mathcal{C}} |S_n(\mathbf{x})| = O_{a.co.}(\theta_n)$.

Demostración. a) Usando que los núcleos K_j tienen soporte en el intervalo $[-1, 1]$ resulta que $|G_i(\mathbf{x})| \leq M\|\mathcal{K}\|_\infty \left(\prod_{j=1}^d h_j\right)^{-1}$, de donde $|\mathbb{E}G_i(\mathbf{x})| \leq M\|\mathcal{K}\|_\infty \left(\prod_{j=1}^d h_j\right)^{-1}$. Más aún, tenemos que $|\mathbb{E}G_i(\mathbf{x})| \leq M\|f_\mathbf{X}\|_\infty \int |\mathcal{K}(\mathbf{u})| d\mathbf{u}$. Luego,

$$|G_i(\mathbf{x}) - \mathbb{E}G_i(\mathbf{x})| \leq \frac{2M\|\mathcal{K}\|_\infty}{\prod_{j=1}^d h_j}\,.$$

Observemos que $Z_i = G_i(\mathbf{x}) - \mathbb{E}G_i(\mathbf{x})$ son tales que $\mathbb{E}Z_i = 0$ y

$$\text{VAR}(Z_i) \leq EG_i^2(\mathbf{x}) \leq \frac{c_1 M^2}{\prod_{j=1}^d h_j}\,,$$

con $c_1 = \|f_\mathbf{X}\|_\infty \int |\mathcal{K}^2(\mathbf{u})| d\mathbf{u}$. Luego, usando el Lema 3.2.1. tenemos que

$$\begin{aligned}\mathbb{P}\left(\frac{1}{n}\left|\sum_{i=1}^n [G_i(\mathbf{x}) - \mathbb{E}G_i(\mathbf{x})]\right| > \eta\right) &\leq 2\exp\left\{-\frac{\eta^2 n}{2\frac{c_1 M^2}{\prod_{j=1}^d h_j} + 2\eta\frac{2M\|\mathcal{K}\|_\infty}{\prod_{j=1}^d h_j}}\right\} \\ &\leq 2\exp\left\{-\frac{\eta^2 n \prod_{j=1}^d h_j}{b_1 + b_2\eta}\right\}\end{aligned}$$

con $b_1 = 2c_1M^2$ y $b_2 = 4M\|\mathcal{K}\|_\infty$.

b) Sea $\breve{\mathbf{x}}_{k_i} = (1, (X_{i,1} - x_{k,1})/h_1, \ldots, (X_{i,d} - x_{k,d})/h_d)^{\mathrm{T}}$ y para $1 \le j, \ell \le d+1$ definamos $\mathcal{K}^{(j\ell)}(\mathbf{u}) = \mathcal{K}(\mathbf{u}) u_{j-1}^m u_{\ell-1}^{\widetilde{m}}$ donde $u_0 = 1$. Luego, tenemos que $\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i - \mathbf{x}) \breve{x}_{i,j}^m \breve{x}_{i,\ell}^{\widetilde{m}} = \mathcal{K}_{\mathbf{H}_d}^{(j\ell)}(\mathbf{X}_i - \mathbf{x})$ y $\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i - \mathbf{x}_k) \breve{x}_{k_i,j}^m \breve{x}_{k_i,\ell}^{\widetilde{m}} = \mathcal{K}_{\mathbf{H}_d}^{(j\ell)}(\mathbf{X}_i - \mathbf{x}_k)$. Por lo tanto, usando que los núcleos K_j tienen soporte en $[-1, 1]$ obtenemos

$$\begin{aligned}
|\widetilde{S}_k(\mathbf{x})| &\le \frac{1}{n}\left|\sum_{i=1}^n (G_i(\mathbf{x}) - G_i(\mathbf{x}_k))\right| + \frac{1}{n}\sum_{i=1}^n |\mathbb{E}G_i(\mathbf{x}_k) - \mathbb{E}G_i(\mathbf{x})| \\
&\le \frac{M}{n}\sum_{i=1}^n \Big\{ \left|\mathcal{K}_{\mathbf{H}_d}^{(j\ell)}(\mathbf{X}_i - \mathbf{x}) - \mathcal{K}_{\mathbf{H}_d}^{(j\ell)}(\mathbf{X}_i - \mathbf{x}_k)\right| \\
&\quad + \mathbb{E}\left|\mathcal{K}_{\mathbf{H}_d}^{(j\ell)}(\mathbf{X}_i - \mathbf{x}) - \mathcal{K}_{\mathbf{H}_d}^{(j\ell)}(\mathbf{X}_i - \mathbf{x}_k)\right| \Big\} \mathbb{I}_{B(\mathbf{x},\mathbf{h}) \cup B(\mathbf{x}_k,\mathbf{h})}(\mathbf{X}_i)
\end{aligned}$$

con $\mathbf{h} = (h_1, \ldots, h_d)^{\mathrm{T}}$ y $B(\mathbf{x}, \mathbf{h}) = \{\mathbf{y} \in \mathbb{R}^d : |y_j - x_j| \le h_j \text{ para } 1 \le j \le d\}$. Como K_j son Lipschitz de orden uno, tenemos que $\mathcal{K}_{\mathbf{H}_d}^{(j\ell)}$ es Lipschitz pues si $\mathcal{K}(\mathbf{u}) \ne 0$ entonces $|u_{j-1}^m u_{\ell-1}^{\widetilde{m}}| \le 1$. Si c_2 indica la constante de Lipschitz de $\mathcal{K}_{\mathbf{H}_d}^{(j\ell)}$ y si $A_h = 1/\min_{1\le j\le d} h_j$ tenemos que

$$\left|\mathcal{K}_{\mathbf{H}_d}^{(j\ell)}(\mathbf{X}_i - \mathbf{x}) - \mathcal{K}_{\mathbf{H}_d}^{(j\ell)}(\mathbf{X}_i - \mathbf{x}_k)\right| \le c_2 \frac{\|\mathbf{H}_d^{-1}(\mathbf{x} - \mathbf{x}_k)\|}{\prod_{j=1}^d h_j} \le c_2 \frac{A_h \rho}{\prod_{j=1}^d h_j}\,.$$

Luego, si llamamos $\mathbf{h} + \rho = (h_1 + \rho, \ldots, h_d + \rho)^{\mathrm{T}}$

$$\begin{aligned}
|\widetilde{S}_k(\mathbf{x})| &\le 2M\, c_2 \frac{A_h \rho}{\prod_{j=1}^d h_j} \frac{1}{n}\sum_{i=1}^n \mathbb{I}_{B(\mathbf{x}_k,\mathbf{h}) \cup B(\mathbf{x},\mathbf{h})}(\mathbf{X}_i) \\
&\le c_3 \frac{A_h \rho}{\prod_{j=1}^d h_j} \frac{1}{n}\sum_{i=1}^n \mathbb{I}_{B(\mathbf{x}_k,\mathbf{h}+\rho)}(\mathbf{X}_i)\,.
\end{aligned}$$

Dado que esta cota no depende de $\mathbf{x}$, tenemos que

$$\theta_n^{-1} \sup_{\mathbf{x} \in \mathcal{B}(\mathbf{x}_k,\rho) \cap \mathcal{C}} |\widetilde{S}_k(\mathbf{x})| \le \theta_n^{-1} c_3 \frac{A_h \rho}{\prod_{j=1}^d h_j} \frac{1}{n}\sum_{i=1}^n \mathbb{I}_{B(\mathbf{x}_k,\mathbf{h}+\rho)}(\mathbf{X}_i)\,.$$

Como $\rho/\prod_{j=1}^n h_j \to 0$, existe n_1 tal que para todo $n \ge n_1$, $\prod_{j=1}^d (h_j + \rho) \le \prod_{j=1}^d h_j + c_4\rho \le 2\prod_{j=1}^d h_j$. Observemos que como para n suficientemente grande $h_j < 1$ para todo j, $A_h\rho \le \rho/\prod_{j=1}^d h_j \to 0$. Por otra parte, también tenemos por hipótesis que $A_h\rho \le M_2\theta_n \to 0$.

Sea $Z_i = \mathbb{I}_{B(\mathbf{x}_k,\mathbf{h}+\rho)}(\mathbf{X}_i)\, A_h\rho/\prod_{j=1}^d h_j$. Luego, como $\mathbb{E}\,\mathbb{I}_{B(\mathbf{x}_k,\mathbf{h}+\rho)}(\mathbf{X}_i) \le c_5 \prod_{j=1}^d (h_j + \rho)$ donde $c_5 = \|f_{\mathbf{X}}\|_\infty$ tenemos que para $n \ge n_1$,

$$|Z_i| \le \frac{A_h\rho}{\prod_{j=1}^d h_j} \qquad \mathbb{E}Z_i \le c_5 \prod_{j=1}^d (h_j + \rho) \frac{A_h\rho}{\prod_{j=1}^d h_j} \le 2c_5 A_h \rho \to 0\,.$$

Por lo tanto, $|Z_i - \mathbb{E}Z_i| \le 2\, A_h\rho/\prod_{j=1}^d h_j$ y

$$\mathrm{VAR}(Z_i) \le \mathbb{E}Z_i^2 \le c_5 \prod_{j=1}^d (h_j + \rho) \left(\frac{A_h\rho}{\prod_{j=1}^d h_j}\right)^2 \le 2c_5 \frac{A_h^2\rho^2}{\prod_{j=1}^d h_j}\,.$$

Con lo cual,

$$\frac{A_h\rho}{\prod_{j=1}^d h_j}\left(\frac{1}{n}\sum_{i=1}^n \mathbb{I}_{B(\mathbf{x}_k,\mathbf{h}+\rho)}(\mathbf{X}_i)\right) = \frac{1}{n}\sum_{i=1}^n Z_i \leq \frac{1}{n}\left|\sum_{i=1}^n (Z_i - \mathbb{E}Z_i)\right| + 2c_5 A_h\rho \tag{3.59}$$

Como por hipótesis $\theta_n^{-1}A_h\rho \leq M_2$, para cualquier $C > C_0 = 4c_3\, c_5\, M_2$ se tiene que

$$\mathbb{P}\left(\theta_n^{-1}\sup_{\mathbf{x}\in\mathcal{B}(\mathbf{x}_k,\rho)}\left|\widetilde{S}_k(\mathbf{x})\right| > C\right) \leq \mathbb{P}\left(\frac{1}{n}\left|\sum_{i=1}^n (Z_i - \mathbb{E}Z_i)\right| > \frac{C\theta_n}{2c_3}\right)$$

Finalmente, por el Lema 3.2.1, para todo $n \geq n_1$ tenemos que

$$\begin{aligned}
\mathbb{P}\left(\theta_n^{-1}\sup_{\mathbf{x}\in\mathcal{B}(\mathbf{x}_k,\rho)\cap\mathcal{C}}\left|\widetilde{S}_k(\mathbf{x})\right| > C\right) &\leq \mathbb{P}\left(\left|\sum_{i=1}^n (Z_i - \mathbb{E}Z_i)\right| > \frac{C\ \theta_n}{2c_3}\ n\right) \\
&\leq 2\exp\left\{-\frac{C^2\theta_n^2 n}{8c_3^2\left(2c_5\frac{A_h^2\rho^2}{\prod_{j=1}^d h_j} + \frac{C\theta_n}{2c_3}\,2\frac{A_h\rho}{\prod_{j=1}^d h_j}\right)}\right\} \\
&\leq 2\exp\left\{-\frac{C^2\theta_n^2 n\prod_{j=1}^d h_j}{b_3A_h^2\rho^2 + b_4C\theta_nA_h\rho}\right\}
\end{aligned}$$

con $b_3 = 16c_3^3c_5$ y $b_4 = 4c_3$, concluyendo la demostración de b).

c) Recordemos que podemos cubrir al compacto $\mathcal{C}$ como $\mathcal{C} \subset \bigcup_{k=1}^{N_\rho(\mathcal{C})}\mathcal{B}(\mathbf{x}_k,\rho)$. Luego, dado $\mathbf{x}\in\mathcal{C}$ existe k tal que $\mathbf{x}\in\mathcal{B}(\mathbf{x}_k,\rho)$. Luego, si $\mathbf{x}\in\mathcal{B}(\mathbf{x}_k,\rho)$, tenemos que $|S_n(\mathbf{x})| \leq |S_n(\mathbf{x}_k)| + \left|\widetilde{S}_k(\mathbf{x})\right|$, por lo que

$$\sup_{\mathbf{x}\in\mathcal{C}}|S_n(\mathbf{x})| = \max_{1\leq k\leq N_\rho(\mathcal{C})}\sup_{\mathbf{x}\in\mathcal{B}(\mathbf{x}_k,\rho)\cap\mathcal{C}}|S_n(\mathbf{x})| \leq \max_{1\leq k\leq N_\rho(\mathcal{C})}|S_n(\mathbf{x}_k)| + \max_{1\leq k\leq N_\rho(\mathcal{C})}\sup_{\mathbf{x}\in\mathcal{B}(\mathbf{x}_k,\rho)\cap\mathcal{C}}\left|\widetilde{S}_k(\mathbf{x})\right|$$

de donde $\mathbb{P}\left(\theta_n^{-1}\sup_{\mathbf{x}\in\mathcal{C}}|S_n(\mathbf{x})| > C\right) \leq \beta_n + \gamma_n$ donde

$$\begin{aligned}
\beta_n &= \mathbb{P}\left(\max_{1\leq k\leq N_\rho(\mathcal{C})}|S_n(\mathbf{x}_k)| > \frac{C\theta_n}{2}\right) \leq N_\rho(\mathcal{C})\sup_{\mathbf{x}\in\mathcal{C}}\mathbb{P}\left(|S_n(\mathbf{x})| > \frac{C\theta_n}{2}\right) \\
\gamma_n &= \mathbb{P}\left(\max_{1\leq k\leq N_\rho(\mathcal{C})}\sup_{\mathbf{x}\in\mathcal{B}(\mathbf{x}_k,\rho)\cap\mathcal{C}}\left|\widetilde{S}_k(\mathbf{x})\right| > \frac{C\theta_n}{2}\right)
\end{aligned}$$

Por a), tenemos que

$$\beta_n \leq 2N_\rho(\mathcal{C})\exp\left\{-\frac{C^2\theta_n^2 n\prod_{j=1}^d h_j}{4b_1 + 2b_2C\theta_n}\right\}$$

Usando b), resulta que para todo $C > 2\,C_0 > 0$ y para todo $n \geq n_0$

$$\gamma_n \leq N_\rho(\mathcal{C})\max_{1\leq k\leq N_\rho(\mathcal{C})}\mathbb{P}\left(\sup_{\mathbf{x}\in\mathcal{B}(\mathbf{x}_k,\rho)\cap\mathcal{C}}\left|\widetilde{S}_k(\mathbf{x})\right| > \frac{C\theta_n}{2}\right) \leq 2\,N_\rho(\mathcal{C})\exp\left\{-\frac{C^2\theta_n^2 n\prod_{j=1}^d h_j}{4b_3A_h^2\rho^2 + 2b_4\,C\theta_nA_h\rho}\right\}.$$

De esta manera, para toda $C > 2C_0 > 0$ y para todo $n \geq n_0$ se tiene que

$$\mathbb{P}\left(\theta_n^{-1}\sup_{\mathbf{x}\in\mathcal{C}}|S_n(\mathbf{x})| > C\right) \leq 2N_\rho(\mathcal{C})\exp\left\{-\frac{C^2\theta_n^2 n\prod_{j=1}^d h_j}{4b_1+2b_2C\theta_n}\right\}+2N_\rho(\mathcal{C})\exp\left\{-\frac{C^2\theta_n^2 n\prod_{j=1}^d h_j}{4b_3A_h^2\rho^2+2b_4\,C\theta_nA_h\rho}\right\}$$

lo que concluye la demostración de c).

d) Observemos que, por **A6**, $\theta_n \to 0$. Tomemos $\rho = \log n/n$ y veamos que se cumplen las condiciones de c). Es claro que $\rho \to 0$ y además por **A6**, $\rho_n\left\{\prod_{j=1}^n h_j\right\}^{-1} = \theta_n^2 \to 0$. Por otra parte,

$$\left(\theta_n^{-1}A_h\rho\right)^2 = \left(\frac{1}{\min_{1\leq j\leq d}\{h_j\}}\frac{\log n}{n}\sqrt{\frac{n\prod_{j=1}^d h_j}{\log n}}\right)^2 \leq \frac{\log n}{n\prod_{j=1}^d h_j} \to 0\,,$$

por lo que $\theta_n^{-1}A_h\rho \leq 1$ si n es suficientemente grande. Observando que $\theta_n^2 n\prod_{j=1}^d h_j = \log(n)$ y que existe A_1 tal que $N_\rho(\mathcal{C}) \leq A_1\rho^{-d}$, tenemos que c) implica que

$$\begin{aligned}\mathbb{P}\left(\theta_n^{-1}\sup_{\mathbf{x}\in\mathcal{C}}|S_n(\mathbf{x}))| > C\right) &\leq 2A_1\rho^{-d}\left[\exp\left\{-\frac{C^2\theta_n^2 n\prod_{j=1}^d h_j}{4b_1+2b_2C\theta_n}\right\}+\exp\left\{-\frac{C^2\theta_n^2 n\prod_{j=1}^d h_j}{4b_3A_h^2\rho^2+2b_4\,C\theta_nA_h\rho}\right\}\right]\\ &\leq 2A_1\rho^{-d}\left[\exp\left\{-\frac{C^2\log(n)}{4b_1+2b_2C\theta_n}\right\}+\exp\left\{-\frac{C^2\log(n)}{4b_3A_h^2\rho^2+2b_4\,C\theta_nA_h\rho}\right\}\right]\end{aligned}$$

Por último, como $\theta_n \to 0$ y $A_h\rho \to 0$ existe n_1 tal que para todo $n \geq n_1$, $2b_2C\theta_n \leq 4b_1$ y $4b_3A_h^2\rho^2 + 2b_4\,C\theta_nA_h\rho \leq 8b_1$, luego

$$\mathbb{P}\left(\theta_n^{-1}\sup_{\mathbf{x}\in\mathcal{C}}|S_n(\mathbf{x}))| > C\right) \leq 4\,A_1\rho^{-d}\exp\left\{-\frac{C^2\log(n)}{8b_1}\right\} = 4\,A_1\left(\frac{1}{\log(n)}\right)^d n^{d-\frac{C^2}{8b_1}} \leq 4\,A_1\,n^{d-\frac{C^2}{8b_1}}$$

Por lo tanto, tenemos que $\sum_{n\geq 1}\mathbb{P}\left(\theta_n^{-1}\sup_{\mathbf{x}\in\mathcal{C}}|S_n(\mathbf{x}))| > C\right) < \infty$ si $C > C_1 = \max\{C_0, \sqrt{8b_1d+3}\}$, lo cual concluye la demostración. □

Observación 3.4.1.

a) En el caso en que $W_i = W_i(a)$ dependiera de un parámetro a, si $|W_i(a)| \leq M$ y $|W_i(a_1) - W_i(a_2)| \leq C|a_1 - a_2|$ donde M no depende de a y C no depende de i, entonces la cota hallada en (3.59) seguirá valiendo aún tomando supremo sobre a en el lado izquierdo de esta desigualdad. De esta manera, todos los resultados obtenidos en el Lema seguirán valiendo aún si tomamos supremo sobre $a \in \mathcal{I}_\delta = [1-\delta, 1+\delta]$ dentro de las probabilidades que se acotan cambiando $N_\rho(\mathcal{C})$ por $N_\rho(\mathcal{C})N_\rho(\mathcal{I}_\delta)$ como veremos en el Lema 3.4.6.

b) Si consideramos tomamos $m = \widetilde{m} = 0$ y $W_i \equiv 1$, de d) tenemos la convergencia uniforme para el estimador de núcleos de la densidad, es decir, obtenemos que

$$\sup_{\mathbf{x}\in\mathcal{C}}\frac{1}{n}\left|\sum_{i=1}^n(\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i-\mathbf{x}) - \mathbb{E}\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i-\mathbf{x}))\right| = O_{\text{a.co.}}(\theta_n)\,,$$

y si $W_i = Y_i$ obtenemos una generalización del resultado dado por Mack y Silverman (1982). Como $\sup_{\mathbf{x}\in\mathcal{C}} |f_{\mathbf{X}}(\mathbf{x})| < \infty$, $\mathcal{K}$ tiene soporte compacto y $f_{\mathbf{X}}$ es uniformemente continua en $\mathcal{C}$,

$$\sup_{\mathbf{x}\in\mathcal{C}} \left| \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i - \mathbf{x}) - f_{\mathbf{X}}(\mathbf{x}) \right| = \sup_{\mathbf{x}\in\mathcal{C}} \left| \int \mathcal{K}(\mathbf{u}) \left(f_{\mathbf{X}}(\mathbf{H}_d\mathbf{u} + \mathbf{x}) - f_{\mathbf{X}}(\mathbf{x}) \right) \right| \to 0$$

se tiene que

$$\sup_{\mathbf{x}\in\mathcal{C}} \left| \frac{1}{n} \sum_{i=1}^{n} \mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i - \mathbf{x}) - f_{\mathbf{X}}(\mathbf{x}) \right| = o_{\text{a.s.}}(1) \tag{3.60}$$

Lema 3.4.6. *Sea $\mathcal{C} \subset \mathbb{R}^d$ un conjunto compacto de interior no vacío, $\mathcal{C} \subset \bigcup_{k=1}^{N_\rho(\mathcal{C})} \mathcal{B}(\mathbf{x}_k, \rho)$ y para $\delta \leq 1/2$ sea $\mathcal{I}_\delta = [1-\delta, 1+\delta] \subset \bigcup_{k=1}^{N_\rho(\mathcal{I}_\delta)} \mathcal{I}_s$ con $\mathcal{I}_s = [a_s - \rho, a_s + \rho]$. Sea $W_i = W_i(a) = f(Y_i, \mathbf{X}_i, \delta_i, a)$ una sucesión de variables aleatorias tales que $|W_i| \leq M$, para todo i y $|W_i(a_1) - W_i(a_2)| \leq M_1|a_1 - a_2|$. Definamos para cada $\mathbf{x} \in \mathbb{R}^d$ $S_n(\mathbf{x}, a) = (1/n)\sum_{i=1}^{n} (G_i(\mathbf{x}, a) - \mathbb{E}G_i(\mathbf{x}, a))$ donde, para cada $1 \leq i \leq n$, $G_i(\mathbf{x}, a) = \mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i - \mathbf{x}) W_i(a) \breve{x}_{i,j}^{m} \breve{x}_{i,\ell}^{\widetilde{m}}$ con $m, \widetilde{m} = 0, 1$ y $1 \leq j, \ell \leq d+1$ fijos y $\breve{\mathbf{x}}_i = (\breve{x}_{i,1}, \ldots, \breve{x}_{i,d+1})^{\mathrm{T}}$. Supongamos que se cumplen además **A3**, **A6**, **C2** y **C8**.*

a) *Sean θ_n y $\rho = \rho_n$ sucesiones numéricas no negativas que tienden a cero, tales que $\theta_n^{-1} A_h \rho \leq M_2$, para todo $n \geq 1$ y $\rho_n \left\{ \prod_{j=1}^{n} h_j \right\}^{-1} \to 0$. Luego, existen $b_1 > 0$ y $b_2 > 0$ y una constante $C_0 > 0$ tales que para toda $C > C_0$ y para todo $n \geq n_0$,*

$$\mathbb{P}\left(\theta_n^{-1} \sup_{\mathbf{x}\in\mathcal{B}(\mathbf{x}_k,\rho)\cap\mathcal{C}} \sup_{a\in\mathcal{I}_s\cap\mathcal{I}_\delta} \left| \widetilde{S}_{k,s}(\mathbf{x}, a) \right| > C \right) \leq 4 \exp\left\{ -\frac{C^2\theta_n^2 n \prod_{j=1}^{d} h_j}{b_1 A_h^2 \rho^2 + b_2 C \theta_n A_h \rho} \right\}$$

donde $\widetilde{S}_{k,s}(\mathbf{x}, a) = S_n(\mathbf{x}, a) - S_n(\mathbf{x}_k, a_s)$.

b) *Sean θ_n y $\rho = \rho_n$ sucesiones numéricas no negativas que tienden a cero, tales que $\theta_n^{-1} A_h \rho \leq M_2$ y $\rho_n \left\{ \prod_{j=1}^{n} h_j \right\}^{-1} \to 0$. Luego, existen b_1, b_2, b_3 y b_4 constantes positivas y una constante $C_0 > 0$ tales que para toda $C > C_0$ y para todo $n \geq n_0$,*

$$\mathbb{P}\left(\theta_n^{-1} \sup_{\substack{\mathbf{x}\in\mathcal{C} \\ a\in\mathcal{I}_\delta}} |S_n(\mathbf{x}, a)| > C \right) \leq \frac{4A_1}{\rho^{d+1}} \left\{ \exp\left\{ -\frac{C^2\theta_n^2 n \prod_{j=1}^{d} h_j}{4b_3 + 2b_4 C \theta_n} \right\} + \exp\left\{ -\frac{C^2\theta_n^2 n \prod_{j=1}^{d} h_j}{4b_1 A_h^2 \rho^2 + 2b_2\, C \theta_n A_h \rho} \right\} \right\}$$

c) *Sea $\theta_n = \sqrt{\log n / (n \prod_{j=1}^{n} h_j)}$. Luego, existe C tal que*

$$\sum_{n\geq 1} \mathbb{P}\left(\theta_n^{-1} \sup_{\mathbf{x}\in\mathcal{C}} \sup_{a\in\mathcal{I}_\delta} |S_n(\mathbf{x}, a)| > C \right) < \infty\,,$$

es decir, $\sup_{\mathbf{x}\in\mathcal{C}} \sup_{a\in\mathcal{I}_\delta} |S_n(\mathbf{x}, a)| = O_{\text{a.co.}}(\theta_n)$.

Demostración. a) Como en Lema 3.4.5, sea $\breve{\mathbf{x}}_{k_i} = (1, (X_{i,1} - x_{k,1})/h_1, \ldots, (X_{i,d} - x_{k,d})/h_d)^{\mathrm{T}}$ y para $1 \leq j, \ell \leq d+1$ definamos $\mathcal{K}^{(j\ell)}(\mathbf{u}) = \mathcal{K}(\mathbf{u}) u_{j-1}^m u_{\ell-1}^{\widetilde{m}}$ donde $u_0 = 1$. Luego, tenemos que $\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i - \mathbf{x}) \breve{x}_{i,j}^m \breve{x}_{i,\ell}^{\widetilde{m}} = \mathcal{K}_{\mathbf{H}_d}^{(j\ell)}(\mathbf{X}_i - \mathbf{x})$ y $\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i - \mathbf{x}_k) \breve{x}_{k_i,j}^m \breve{x}_{k_i,\ell}^{\widetilde{m}} = \mathcal{K}_{\mathbf{H}_d}^{(j\ell)}(\mathbf{X}_i - \mathbf{x}_k)$. Por lo tanto, usando que los núcleos K_j tienen soporte en $[-1, 1]$ obtenemos

$$\begin{aligned} |\widetilde{S}_{k,s}(\mathbf{x}, a)| &\leq \frac{1}{n} \left| \sum_{i=1}^n (G_i(\mathbf{x}, a) - G_i(\mathbf{x}_k, a_s)) \right| + \frac{1}{n} \sum_{i=1}^n |\mathbb{E} G_i(\mathbf{x}_k, a_s) - \mathbb{E} G_i(\mathbf{x}, a)| \\ &\leq \widetilde{\widetilde{S}}_k(\mathbf{x}, a) + \widetilde{\widetilde{S}}_{k,s}(\mathbf{x}_k, a) \end{aligned}$$

donde

$$\begin{aligned} \widetilde{\widetilde{S}}_k(\mathbf{x}, a) &= \frac{1}{n} \left| \sum_{i=1}^n (G_i(\mathbf{x}, a) - G_i(\mathbf{x}_k, a)) \right| + \frac{1}{n} \sum_{i=1}^n |\mathbb{E} G_i(\mathbf{x}_k, a) - \mathbb{E} G_i(\mathbf{x}, a)| \\ &\leq \frac{M}{n} \sum_{i=1}^n \left\{ \left| \mathcal{K}_{\mathbf{H}_d}^{(j\ell)}(\mathbf{X}_i - \mathbf{x}) - \mathcal{K}_{\mathbf{H}_d}^{(j\ell)}(\mathbf{X}_i - \mathbf{x}_k) \right| \right. \\ &\quad \left. + \left| \mathbb{E} \mathcal{K}_{\mathbf{H}_d}^{(j\ell)}(\mathbf{X}_i - \mathbf{x}) - \mathbb{E} \mathcal{K}_{\mathbf{H}_d}^{(j\ell)}(\mathbf{X}_i - \mathbf{x}_k) \right| \right\} \mathbb{I}_{B(\mathbf{x},\mathbf{h}) \cup B(\mathbf{x}_k,\mathbf{h})}(\mathbf{X}_i) \\ \widetilde{\widetilde{S}}_{k,s}(\mathbf{x}_k, a) &= \frac{1}{n} \left| \sum_{i=1}^n (G_i(\mathbf{x}_k, a) - G_i(\mathbf{x}_k, a_s)) \right| + \frac{1}{n} \sum_{i=1}^n |\mathbb{E} G_i(\mathbf{x}_k, a_s) - \mathbb{E} G_i(\mathbf{x}_k, a)| \end{aligned}$$

con $\mathbf{h} = (h_1, \ldots, h_d)^{\mathrm{T}}$ y $B(\mathbf{x}, \mathbf{h}) = \{\mathbf{y} \in \mathbb{R}^d : |y_j - x_j| \leq h_j \text{ para } 1 \leq j \leq d\}$. Luego, como en el Lema 3.4.5, tenemos que

$$\theta_n^{-1} \sup_{\mathbf{x} \in \mathcal{B}(\mathbf{x}_k, \rho) \cap \mathcal{C}} \sup_{a \in \mathcal{I}_s \cap \mathcal{I}_\delta} |\widetilde{\widetilde{S}}_k(\mathbf{x}, a)| \leq \theta_n^{-1} c_3 \frac{A_h \rho}{\prod_{j=1}^d h_j} \frac{1}{n} \sum_{i=1}^n \mathbb{I}_{B(\mathbf{x}_k, \mathbf{h}+\rho)}(\mathbf{X}_i) = \widetilde{A}_{k,n} \,,$$

con lo cual, para todo $n \geq n_1$ si $C > C_{0,1}$ tenemos que existen b_1 y b_2

$$\mathbb{P}\left(\theta_n^{-1} \sup_{\substack{\mathbf{x} \in \mathcal{B}(\mathbf{x}_k, \rho) \cap \mathcal{C} \\ a \in \mathcal{I}_s \cap \mathcal{I}_\delta}} |\widetilde{\widetilde{S}}_k(\mathbf{x}, a)| > \frac{C}{2} \right) \leq \mathbb{P}\left(\theta_n^{-1} \widetilde{A}_{k,n} > \frac{C}{2} \right) \leq 2 \exp\left\{ -\frac{C^2 \theta_n^2 n \prod_{j=1}^d h_j}{b_1 A_h^2 \rho^2 + b_2 C \theta_n A_h \rho} \right\} \quad (3.61)$$

Por otra parte, de $|W_i(a_1) - W_i(a_2)| \leq M_1 |a_1 - a_2|$ y del hecho que $|a - a_s| < \rho$ se deduce que

$$|G_i(\mathbf{x}_k, a) - G_i(\mathbf{x}_k, a_s)| \leq M_1 \rho \, \mathcal{K}_{\mathbf{H}_d}^{(j\ell)}(\mathbf{X}_i - \mathbf{x}_k) \leq M_1 \, \rho \, \frac{1}{\prod_{j=1}^d h_j} \mathbb{I}_{B(\mathbf{x}_k, \mathbf{h})}(\mathbf{X}_i) \leq M_1 \, \rho \, \frac{1}{\prod_{j=1}^d h_j} \mathbb{I}_{B(\mathbf{x}_k, \mathbf{h}+\rho)}(\mathbf{X}_i)$$

Luego, si $n \geq n_2$ tenemos que

$$\widetilde{\widetilde{S}}_{k,s}(\mathbf{x}_k, a) \leq 2 \, M_1 \rho \frac{1}{n} \sum_{i=1}^n \mathbb{I}_{B(\mathbf{x}, \mathbf{h}+\rho)}(\mathbf{X}_i) \leq \widetilde{A}_{k,n}$$

pues $A_h \to \infty$. Por lo tanto, si $n \geq \max\{n_1, n_2\}$

$$\mathbb{P}\left(\theta_n^{-1} \sup_{\substack{\mathbf{x} \in \mathcal{B}(\mathbf{x}_k, \rho) \cap \mathcal{C} \\ a \in \mathcal{I}_s \cap \mathcal{I}_\delta}} |\widetilde{\widetilde{S}}_{k,\ell}(\mathbf{x}, a)| > \frac{C}{2} \right) \leq \mathbb{P}\left(\theta_n^{-1} \widetilde{A}_{k,n} > \frac{C}{2} \right) \leq 2 \exp\left\{ -\frac{C^2 \theta_n^2 n \prod_{j=1}^d h_j}{b_1 A_h^2 \rho^2 + b_2 C \theta_n A_h \rho} \right\} ,$$

lo que conjuntamente con (3.61) concluye la demostración de a).

b) Recordemos que $\mathcal{C} \subset \bigcup_{k=1}^{N_\rho(\mathcal{C})} \mathcal{B}(\mathbf{x}_k, \rho)$ con $N_\rho(\mathcal{C}) \leq A_1/\rho^d$ y $\mathcal{I}_\delta = [1-\delta, 1+\delta] \subset \bigcup_{k=1}^{N_\rho(\mathcal{I}_\delta)} \mathcal{I}_s$ con $N_\rho(\mathcal{I}_\delta) \leq 2/\rho$. Luego, dado $\mathbf{x} \in \mathcal{C}$ y $a \in \mathcal{I}_\delta$ existen k, s tales que $\mathbf{x} \in \mathcal{B}(\mathbf{x}_k, \rho)$, $a \in \mathcal{I}_s$. Luego, si $\mathbf{x} \in \mathcal{B}(\mathbf{x}_k, \rho)$, tenemos que $|S_n(\mathbf{x}, a)| \leq |S_n(\mathbf{x}_k, a_s)| + \left|\widetilde{S}_{k,s}(\mathbf{x}, a)\right|$, por lo que

$$\sup_{\mathbf{x}\in\mathcal{C}} \sup_{a\in\mathcal{I}_\delta} |S_n(\mathbf{x}, a)| \leq \max_{\substack{1\leq k\leq N_\rho(\mathcal{C}) \\ 1\leq s\leq N_\rho(\mathcal{I}_\delta)}} |S_n(\mathbf{x}_k, a_s)| + \max_{\substack{1\leq k\leq N_\rho(\mathcal{C}) \\ 1\leq s\leq N_\rho(\mathcal{I}_\delta)}} \sup_{\substack{\mathbf{x}\in\mathcal{B}(\mathbf{x}_k,\rho)\cap\mathcal{C} \\ a\in\mathcal{I}_s\cap\mathcal{I}_\delta}} \left|\widetilde{S}_{k,s}(\mathbf{x}, a)\right|$$

de donde $\mathbb{P}\left(\theta_n^{-1} \sup_{\mathbf{x}\in\mathcal{C}} |S_n(\mathbf{x})| > C\right) \leq \beta_n + \gamma_n$ donde

$$\begin{aligned}
\beta_n &= \mathbb{P}\left(\max_{\substack{1\leq k\leq N_\rho(\mathcal{C}) \\ 1\leq s\leq N_\rho(\mathcal{I}_\delta)}} |S_n(\mathbf{x}_k, a_s)| > \frac{C\theta_n}{2}\right) \leq N_\rho(\mathcal{C})\, N_\rho(\mathcal{I}_\delta) \sup_{\substack{\mathbf{x}\in\mathcal{C} \\ a\in\mathcal{I}_\delta}} \mathbb{P}\left(|S_n(\mathbf{x}, a)| > \frac{C\theta_n}{2}\right) \\
\gamma_n &= \mathbb{P}\left(\max_{\substack{1\leq k\leq N_\rho(\mathcal{C}) \\ 1\leq s\leq N_\rho(\mathcal{I}_\delta)}} \sup_{\substack{\mathbf{x}\in\mathcal{B}(\mathbf{x}_k,\rho)\cap\mathcal{C} \\ a\in\mathcal{I}_s\cap\mathcal{I}_\delta}} \left|\widetilde{S}_{k,s}(\mathbf{x}, a)\right| > \frac{C\theta_n}{2}\right)
\end{aligned}$$

Por el Lema 3.4.5 a), tenemos que

$$\beta_n \leq 2N_\rho(\mathcal{C})N_\rho(\mathcal{I}_\delta) \exp\left\{-\frac{C^2\theta_n^2 n \prod_{j=1}^d h_j}{4b_3 + 2b_4 C\theta_n}\right\} \leq 2\frac{A_1}{\rho^{d+1}} \exp\left\{-\frac{C^2\theta_n^2 n \prod_{j=1}^d h_j}{4b_3 + 2b_4 C\theta_n}\right\}$$

Usando a), resulta que para todo $C > 2\,C_0 > 0$ y para todo $n \geq n_0$

$$\gamma_n \leq N_\rho(\mathcal{C})N_\rho(\mathcal{I}_\delta)\mathbb{P}\left(\sup_{\substack{\mathbf{x}\in\mathcal{B}(\mathbf{x}_k,\rho)\cap\mathcal{C} \\ a\in\mathcal{I}_s\cap\mathcal{I}_\delta}} \left|\widetilde{S}_{k,s}(\mathbf{x}, a)\right| > \frac{C\theta_n}{2}\right) \leq 4\,\frac{A_1}{\rho^{d+1}} \exp\left\{-\frac{C^2\theta_n^2 n \prod_{j=1}^d h_j}{4b_1 A_h^2\rho^2 + 2b_2\, C\theta_n A_h \rho}\right\}.$$

De esta manera, para toda $C > 2C_0 > 0$ y para todo $n \geq n_0$ se tiene que

$$\mathbb{P}\left(\theta_n^{-1} \sup_{\substack{\mathbf{x}\in\mathcal{C} \\ a\in\mathcal{I}_\delta}} |S_n(\mathbf{x}, a)| > C\right) \leq 4\frac{A_1}{\rho^{d+1}} \left\{\exp\left\{-\frac{C^2\theta_n^2 n \prod_{j=1}^d h_j}{4b_3 + 2b_4 C\theta_n}\right\} + \exp\left\{-\frac{C^2\theta_n^2 n \prod_{j=1}^d h_j}{4b_1 A_h^2\rho^2 + 2b_2\, C\theta_n A_h \rho}\right\}\right\}$$

lo que concluye la demostración de b).

c) Se obtiene en forma análoga a la demostración de Lema 3.4.5 d) usando b). □

Lema 3.4.7. *Supongamos que se cumplen* ***A1***, ***A4***, ***A6***, ***C1*** *a* ***C3****a) y* ***C5*** *a* ***C8*** *entonces*

$$\sup_{\mathbf{x}\in\mathcal{C}} |J_{n,j}(\mathbf{x}, \widehat{a}_\sigma(\mathbf{x}))| = O_{a.s.}(\theta_n)$$

con $\theta_n = \sqrt{\log n/(n\prod_{j=1}^d h_j)}$, *donde* $j = 1, \ldots, d+1$ *y* $J_{n,j}$ *es la componente* j*-ésima del vector* $\mathbf{J}_n$.

Demostración. Por (3.16), tomando $\delta = 1/2$, tenemos que existe $\mathcal{N}$, $\mathbb{P}(\mathcal{N}) = 0$ tal que si $\omega \notin \mathcal{N}$, existe $n_1 \in \mathbb{N}$ tal que para todo $n \geq n_1$ $\sup_{\mathbf{x}\in\mathcal{C}} |\widehat{a}_\sigma(\mathbf{x}) - 1| \leq \delta$. De esta manera, si $\omega \notin \mathcal{N}$ y $n \geq n_1$ tenemos que

$$\sup_{\mathbf{x}\in\mathcal{C}} |J_{n,j}(\mathbf{x}, \widehat{a}_\sigma(\mathbf{x}))| \leq \sup_{\mathbf{x}\in\mathcal{C}} \sup_{a\in[1-\delta,1+\delta]} |J_{n,j}(\mathbf{x}, a)|$$

por lo que bastará ver que $\sup_{\mathbf{x}\in\mathcal{C}} \sup_{a\in[1-\delta,1+\delta]} |J_{n,j}(\mathbf{x}, a)| = O_{\text{a.s.}}(\theta_n)$. Recordemos que

$$\mathbf{J}_n(\mathbf{x}, a) = \frac{1}{n}\sum_{i=1}^{n} \mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i - \mathbf{x})\xi_i\psi\left(\frac{\sigma(\mathbf{X}_i)\varepsilon_i}{\sigma(\mathbf{x})}\, a\right)\breve{\mathbf{x}}_i = \frac{1}{n}\sum_{i=1}^{n} \mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i - \mathbf{x})\xi_i\psi\left(V_i\, a\right)\breve{\mathbf{x}}_i\ .$$

luego,

$$J_{n,j}(\mathbf{x}, a) = \frac{1}{n}\sum_{i=1}^{n} \mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i - \mathbf{x})\xi_i\psi\left(V_i\, a\right)\breve{x}_{i,j}\ .$$

Por lo tanto, la conclusión se deduce del Lema 3.4.6 c) tomando $m = 1$, $\widetilde{m} = 0$ y $W_i(a) = \xi_i\psi\left(V_i\, a\right)$ y observando que $|W_i(a)| \leq \|\psi\|_\infty/i(t)$ y $|W_i(a_1) - W_i(a_2)| \leq (\|\zeta\|_\infty/i(t))\,|a_1 - a_2|$. $\square$

Capítulo 4

Distribución asintótica de los $M-$estimadores

4.1 Introducción

En esta sección deduciremos la distribución asintótica de los estimadores iniciales $\widetilde{g}_{\mathrm{M}_0}(\mathbf{x})$ y $\widetilde{g}_{\mathrm{M}_{q,\alpha}}(\mathbf{x})$ y de los estimadores de las componentes aditivas obtenidos a partir de ellos. Vale mencionar que el estimador $\widetilde{g}_{\mathrm{M}_0}(\mathbf{x})$ es un caso particular del estimador $\widetilde{g}_{\mathrm{M}_{q,\alpha}}(\mathbf{x})$ cuando $q = 0$. Sin embargo, presentaremos primero la demostración de su distribución asintótica en la Sección 4.2 ya que permite entender las ideas que subyacen en la misma.

En primer lugar, explicaremos porqué no daremos la distribución del estimador $\widetilde{g}_{\mathrm{M}_1}(\mathbf{x})$ y de las componentes basadas en él. Como dijimos en la Sección 2.1.2 este estimador tiene problemas cuando la dimensión es mayor a 4. Efectivamente, supongamos que $\sigma(\mathbf{x})$ es conocida y no depende de $\mathbf{x}$, o sea, $\sigma(\mathbf{x}) \equiv \sigma$, más aún supongamos que $\sigma = 1$. En este caso, el estimador $\widetilde{g}_{\mathrm{M}_1}(\mathbf{x})$ coincide con el introducido por Li, Zheng y Zheng (2012) si $K_j = K$ para todo j. Consideremos por simplicidad el problema de estimar g_1 y supongamos como en Li, Zheng y Zheng (2012) que $K_j = K$, para todo j, $h_{1,n} = h_n$ y $h_{j,n} = \widetilde{h}_n$. Esta restricción no cambia los comentarios que haremos ya que si utilizamos distintos núcleos K_j es necesario de todas formas que $\int u^2 K_j(u)du > 0$ ya que la demostración de estos autores, involucra invertir la matriz $\mathbf{S} = \mathrm{diag}(1, s_1, \ldots, s_d)$. En este caso, estos autores mostraron que, bajo condiciones de regularidad que incluyen las hipótesis **C1**, **C2** y **C3** y la condición de identificabilidad (3.2), si $\widetilde{h}_n \to 0$, $nh_n\widetilde{h}_n^{d-1} \to \infty$, $n\widetilde{h}_n^{d+4} \to \infty$, $h_n \to 0$ y $nh_n \to \infty$ entonces

$$\sqrt{nh_n}\left[\widehat{g}_{1,\mathrm{M}_1}(x) - g_1(x) - \beta_1(x)\right] \xrightarrow{D} N\left(0, \Sigma_1(x)\right)$$

donde

$$\beta_1(x) = \frac{1}{2}\int u^2 K(u)du \left(h_n^2 g_1''(x) + \widetilde{h}_n^2 \sum_{j=2}^{d} \int g_j''(u)\, q_j(u)du\right)(1 + o_{\mathbb{P}}(1))$$

$$\Sigma_1(x) = \frac{\mathbb{E}\psi^2(\varepsilon)}{\left(\mathbb{E}\psi'(\varepsilon)\right)^2} \int K^2(u)du \int \frac{1}{f_{\mathbf{X}}(\mathbf{x})} q_{\underline{1}}^2(\mathbf{x}_{\underline{1}})\, d\mathbf{x}_{\underline{1}}\,.$$

Por lo tanto, el término $b_1(x) = \sqrt{nh_n}\beta_1(x)$ representa el sesgo asintótico. Observemos que

$$b_1(x) = \frac{1}{2}\int u^2 K(u)du \left((nh_n^5)^{\frac{1}{2}} g_1''(x) + (nh_n \widetilde{h}_n^4)^{\frac{1}{2}} \sum_{j=2}^{d} \int g_j''(u)\, q_j(u)du \right) (1 + o_{\mathbb{P}}(1))$$

El primer sumando puede mantenerse acotado si $nh_n^5 \to \beta$, de hecho, Li, Zheng y Zheng (2012) piden en una de sus hipótesis que $h_n = \beta n^{-\frac{1}{5}}$. Sin embargo, para que el segundo sumando no vaya a infinito sería necesario que $nh_n\widetilde{h}_n^4 \to \widetilde{\beta} \geq 0$ lo que contradice la hipótesis $nh_n\widetilde{h}_n^{d-1} \to \infty$ si $d > 4$. Es decir, estos estimadores sufren en problema de la dimensión descripto por varios autores. Entre otros, Hengartner and Sperlich (2005) y Kong, Linton y Xia (2010) mencionan que para evitar el problema antes mencionado es necesario que el orden q del polinomio utilizado crezca con la dimensión de las covariables $\mathbf{X}$ ya que sería necesario que $3d < 2q + 5$, lo que aumenta la complejidad numérica al crecer la dimensión.

Por estas razones, consideraremos solamente el caso de los estimadores robustos $\widehat{g}_{\alpha,\mathrm{M}_0}(x)$ y $\widehat{g}_{\alpha,\mathrm{M}_{q,\alpha}}(x)$. En este caso, las condiciones sobre las ventanas dependerán solamente del orden del polinomio y del orden $\ell > 2$ del núcleo L que se considera para suavizar las componentes $j \neq \alpha$. Por ejemplo, si elegimos un polinomio de orden $q = 1$ para estimar g_α entonces, eligiendo $h_n = \beta n^{\frac{1}{5}}$ y $\widetilde{h}_n = \widetilde{\beta} n^{-\nu}$ con $2/(5\,\ell) < \nu < 2/(5\,(d-1))$ se cumplen las condiciones que garantizan la distribución asintótica. En particular, eso dice que el orden ℓ del núcleo L debe adaptarse a la dimensión de las covariables ya que es necesario que $\ell \geq d$.

4.2 Distribución asintótica de los M-estimadores basados en polinomios locales de orden 0

El siguiente resultado generaliza la Proposición 3.3.1 de Boente, González-Manteiga y Pérez-González (2009) al caso en que los núcleos no son iguales y g_j es dos veces diferenciable como se supuso en **C1**. Si suponemos solamente que g_j, $r(\mathbf{u}) = p(\mathbf{u})/t(\mathbf{u})$ y $\sigma(\mathbf{u})$ son diferenciables y que $n\prod_{j=1}^{d} h_j \sum_{j=1}^{d} h_j^2 \to \beta$, una demostración análoga permite ver que el sesgo asintótico es 0 ya que $\int tK_j(t)dt = 0$.

Teorema 4.2.1. *Sea* $\mathbf{x}$ *un punto del interior de* $\mathcal{S}_f$ *tal que se cumple* ***A3*** *en un entorno compacto de* $\mathbf{x}$. *Supongamos que se cumplen* ***A1***, ***A4***, ***C1***, ***C2***, ***C7*** *y que además* ψ'' *es Lipschitz. Más aún, supongamos que* $f_{\mathbf{X}}(\mathbf{u})$ *es continuamente diferenciable y que* $r(\mathbf{u}) = p(\mathbf{u})/t(\mathbf{u})$ *y* $\sigma(\mathbf{u})$ *son dos veces diferenciables. Entonces, si* $\widehat{s}(\mathbf{x}) \xrightarrow{p} \sigma(\mathbf{x})$ *y* $n\prod_{j=1}^{d} h_j\, h_\ell^4 \to \beta_\ell$, $1 \leq \ell \leq d$, *tenemos que*

$$\sqrt{n\prod_{j=1}^{d} h_j}\,(\widetilde{g}_{\mathrm{M}_0}(\mathbf{x}) - g(\mathbf{x})) \xrightarrow{D} N\,(b_0(\mathbf{x}), v_0(\mathbf{x}))$$

donde

$$\begin{aligned} b_0(\mathbf{x}) &= \frac{1}{2}\sum_{1\le\ell\le d}\beta_\ell^{\frac{1}{2}} s_\ell\, g_\ell''(x_\ell) \\ &\quad + \frac{1}{r(\mathbf{x})f_{\mathbf{X}}(\mathbf{x})}\sum_{1\le\ell\le d}\beta_\ell^{\frac{1}{2}} s_\ell\, g_\ell'(x_\ell)\left[\frac{\partial}{\partial u_\ell}r(\mathbf{u})\Big|_{\mathbf{u}=\mathbf{x}} f_{\mathbf{X}}(\mathbf{x}) + \frac{\partial}{\partial u_\ell}f_{\mathbf{X}}(\mathbf{u})\Big|_{\mathbf{u}=\mathbf{x}} r(\mathbf{x})\right] \\ v_0(\mathbf{x}) &= \frac{\mathbb{E}\psi^2(\varepsilon)}{\mathbb{E}^2\psi'(\varepsilon)}\,\frac{\sigma^2(\mathbf{x})}{p(\mathbf{x})f_{\mathbf{X}}(\mathbf{x})}\int \mathcal{K}^2(\mathbf{u})d\mathbf{u}\,. \end{aligned}$$

DEMOSTRACIÓN. La demostración sigue argumentos análogos a los considerados por Boente, González-Manteiga y Pérez-González (2009). Recordemos que $\widetilde{g}_{\mathrm{M}_0}(\mathbf{x})$ es solución de $\Psi_n(a) = 0$ donde $\Psi_n(a)$ está definida en (2.13). Usando un desarrollo de Taylor de orden uno, obtenemos que

$$\widetilde{g}_{\mathrm{M}_0}(\mathbf{x}) - g(\mathbf{x}) = \widehat{s}(\mathbf{x})A_{0,n}^{-1}(\mathbf{x})A_{1,n}(\mathbf{x},\widehat{s}(\mathbf{x}))$$

donde

$$\begin{aligned} A_{0,n}(\mathbf{x}) &= A_{01,n}(\mathbf{x}) + A_{02,n}(\mathbf{x}) \\ A_{01,n}(\mathbf{x}) &= \frac{1}{n}\sum_{i=1}^n \mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i-\mathbf{x})\,\xi_i\psi'\left(\frac{Y_i-g(\mathbf{x})}{\widehat{s}(\mathbf{x})}\right) \\ A_{02,n}(\mathbf{x}) &= \frac{1}{2\widehat{s}(\mathbf{x})}(\widetilde{g}_{\mathrm{M}_0}(\mathbf{x})-g(\mathbf{x}))\frac{1}{n}\sum_{i=1}^n \mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i-\mathbf{x})\,\xi_i\psi''\left(\frac{Y_i-\widehat{\theta}(\mathbf{x})}{\widehat{s}(\mathbf{x})}\right) \\ A_{1,n}(\mathbf{x},\sigma) &= \frac{1}{n}\sum_{i=1}^n \mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i-\mathbf{x})\,\xi_i\psi\left(\frac{Y_i-g(\mathbf{x})}{\sigma}\right) \end{aligned}$$

donde $\widehat{\theta}(\mathbf{x})$ es un punto intermedio entre $\widetilde{g}_{\mathrm{M}_0}(\mathbf{x})$ y $g(\mathbf{x})$. Sea $A_0(\psi) = \mathbb{E}(\psi'(\varepsilon))$. Bastará probar que

a) $A_{01,n}(\mathbf{x}) \xrightarrow{p} r(\mathbf{x})f_{\mathbf{X}}(\mathbf{x})A_0(\psi)$, $A_{02,n}(\mathbf{x}) \xrightarrow{p} 0$

b) $\widehat{D}_n(\mathbf{x}) = \left(n\prod_{j=1}^d h_j\right)^{\frac{1}{2}}(\widehat{s}(\mathbf{x})A_{1,n}(\mathbf{x},\widehat{s}(\mathbf{x})) - \sigma(\mathbf{x})A_{1,n}(\mathbf{x},\sigma(\mathbf{x}))) \xrightarrow{p} 0$

c) $\widetilde{C}_n(\mathbf{x}) = \left(n\prod_{j=1}^d h_j\right)^{\frac{1}{2}}\left(A_{1,n}(\mathbf{x},\sigma(\mathbf{x})) - \widetilde{A}_{1,n}(\mathbf{x},\sigma(\mathbf{x}))\right) \xrightarrow{p} a_1$

d) $\left(n\prod_{j=1}^d h_j\right)^{\frac{1}{2}}\widetilde{A}_{1,n}(\mathbf{x},\sigma(\mathbf{x})) \xrightarrow{D} N(0,\sigma_1)$,

con

$$\widetilde{A}_{1,n}(\mathbf{x},\sigma) = \frac{1}{n}\sum_{i=1}^n \mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i-\mathbf{x})\,\xi_i\psi\left(\frac{Y_i-g(\mathbf{X}_i)}{\sigma}\right)$$

$$
\begin{aligned}
a_1 &= b_0(\mathbf{x})\frac{p(\mathbf{x})f_{\mathbf{X}}(\mathbf{x})A_0(\psi)}{\sigma(\mathbf{x})t(\mathbf{x})} = b_0(\mathbf{x})\frac{r(\mathbf{x})f_{\mathbf{X}}(\mathbf{x})A_0(\psi)}{\sigma(\mathbf{x})} \\
&= \frac{1}{2}A_0(\psi)\frac{r(\mathbf{x})f_{\mathbf{X}}(\mathbf{x})}{\sigma(\mathbf{x})}\sum_{1\le\ell\le d}\beta_\ell^{\frac{1}{2}}s_\ell\, g''_\ell(x_\ell) \\
&\quad + A_0(\psi)\frac{1}{\sigma(\mathbf{x})}\sum_{1\le\ell\le d}\beta_\ell^{\frac{1}{2}}s_\ell\, g'_\ell(x_\ell)\left[\frac{\partial}{\partial u_\ell}r(\mathbf{u})\Big|_{\mathbf{u}=\mathbf{x}}f_{\mathbf{X}}(\mathbf{x}) + \frac{\partial}{\partial u_\ell}f_{\mathbf{X}}(\mathbf{u})\Big|_{\mathbf{u}=\mathbf{x}}r(\mathbf{x})\right] \qquad (4.1) \\
\sigma_1 &= \frac{p(\mathbf{x})f_{\mathbf{X}}(\mathbf{x})}{t^2(\mathbf{x})}\int \mathcal{K}^2(\mathbf{u})d\mathbf{u}\,\mathbb{E}(\psi^2(\varepsilon))\,. \qquad (4.2)
\end{aligned}
$$

a) Como ψ'' es acotada, $\mathbf{x}$ es tal que se cumple **A3** en un entorno compacto $\mathcal{C}$ de $\mathbf{x}$ y K_j tiene soporte compacto, tenemos que si n es suficientemente grande, si $\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i-\mathbf{x})\neq 0$ entonces $\mathbf{X}_i\in\mathcal{C}$. Por lo tanto, tenemos que

$$
|A_{02,n}(\mathbf{x})| \le \frac{\|\psi''\|_\infty}{2\widehat{s}(\mathbf{x})}\frac{1}{i(t)}|\widetilde{g}_{\mathrm{M}_0}(\mathbf{x})-g(\mathbf{x})|\frac{1}{n}\sum_{i=1}^n|\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i-\mathbf{x})|\,,
$$

de donde usando la consistencia de $\widetilde{g}_{\mathrm{M}_0}(\mathbf{x})$ y el hecho que $\sum_{i=1}^n|\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i-\mathbf{x})|/n$ es acotado en probabilidad se deduce que $A_{02,n}(\mathbf{x})\xrightarrow{p}0$.

Por otra parte, si llamamos como antes $\widehat{V}_i=\sigma(\mathbf{X}_i)\varepsilon_i/\widehat{s}(\mathbf{x})$ usando que $Y_i=g(\mathbf{X}_i)+\sigma(\mathbf{X}_i)\varepsilon_i$ y utilizando nuevamente un desarrollo de Taylor de orden uno obtenemos que

$$
\begin{aligned}
A_{01,n}(\mathbf{x}) &= \frac{1}{n}\sum_{i=1}^n\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i-\mathbf{x})\,\xi_i\psi'\left(\widehat{V}_i+\frac{g(\mathbf{X}_i)-g(\mathbf{x})}{\widehat{s}(\mathbf{x})}\right) \\
&= \frac{1}{n}\sum_{i=1}^n\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i-\mathbf{x})\,\xi_i\psi'\left(\widehat{V}_i\right)+\frac{1}{\widehat{s}(\mathbf{x})}\frac{1}{n}\sum_{i=1}^n\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i-\mathbf{x})\,\xi_i(g(\mathbf{X}_i)-g(\mathbf{x}))\psi''\left(\widehat{V}_i+\widehat{\nu}_i\right) \\
&= B_{1,n}(\mathbf{x})+B_{2,n}(\mathbf{x})
\end{aligned}
$$

donde $\widehat{\nu}_i(\mathbf{x})$ es un punto intermedio entre 0 y $(g(\mathbf{X}_i)-g(\mathbf{x}))/\widehat{s}(\mathbf{x})$. Observemos que como g_j es continuamente diferenciable en $\mathcal{S}_j$, es Lipschitz, de donde $|g(\mathbf{X}_i)-g(\mathbf{x})|\le C\sum_{j=1}^d|X_{i,j}-x_j|\le C\sum_{j=1}^d h_j$ si $\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i-\mathbf{x})\neq 0$. Por lo tanto,

$$
|B_{2,n}(\mathbf{x})|\le C\sum_{j=1}^d h_j\frac{\|\psi''\|_\infty}{\widehat{s}(\mathbf{x})}\frac{1}{i(t)}\frac{1}{n}\sum_{i=1}^n|\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i-\mathbf{x})|\,,
$$

y el hecho que $h_j\to 0$ implica que $B_{2,n}(\mathbf{x})\xrightarrow{p}0$.

Por otra parte, los Lemas 3.4.1 y 3.4.2 implican que $B_{1,n}(\mathbf{x})\xrightarrow{p}r(\mathbf{x})f_{\mathbf{X}}(\mathbf{x})A_0(\psi)$, lo que concluye la demostración de a).

b) La demostración de b) sigue los mismos pasos que la del Lema 3 de Boente y Fraiman (1990). Efectivamente, como en el Lema 3.4.3, definamos $\psi_a(t)=a\psi(t/a)$ y sea $Y_{i,\mathbf{x}}=Y_i-g(\mathbf{x})$, luego

$$
\widehat{D}_n(\mathbf{x})=\widehat{s}(\mathbf{x})A_{1,n}(\mathbf{x},\widehat{s}(\mathbf{x}))-\sigma(\mathbf{x})A_{1,n}(\mathbf{x},\sigma(\mathbf{x})=\frac{1}{n}\sum_{i=1}^n\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i-\mathbf{x})\xi_i\left(\psi_{\widehat{s}(\mathbf{x})}(Y_{i,\mathbf{x}})-\psi_{\sigma(\mathbf{x})}(Y_{i,\mathbf{x}})\right)\,.
$$

Para todo $s > 0$, definamos $H_\lambda(u) = \psi_{\sigma(\mathbf{x})+\lambda}(u) - \psi_{\sigma(\mathbf{x})}(u)$, $I_\lambda(u) = \psi_{\sigma(\mathbf{x})-\lambda}(u) - \psi_{\sigma(\mathbf{x})}(u)$,

$$\begin{aligned}
J_n^+(\lambda) &= \frac{1}{\sqrt{n\prod_{j=1}^d h_j}} \sum_{i=1}^n \left\{ \prod_{j=1}^d K_j\left(\frac{X_{ij}-x_j}{h_j}\right)\right\} H_\lambda(Y_{i,\mathbf{x}})\xi_i\,, \\
\widetilde{J}_n^+(\lambda) &= J_n^+(\lambda) - \mathbb{E}(J_n^+(\lambda)) \\
J_n^-(\lambda) &= \frac{1}{\sqrt{n\prod_{j=1}^d h_j}} \sum_{i=1}^n \left\{ \prod_{j=1}^d K_j\left(\frac{X_{ij}-x_j}{h_j}\right)\right\} I_\lambda(Y_{i,\mathbf{x}})\xi_i \\
\widetilde{J}_n^-(\lambda) &= J_n^-(\lambda) - \mathbb{E}(J_n^-(\lambda))\,.
\end{aligned}$$

Observemos que si $\widehat{\lambda}(\mathbf{x}) = \widehat{s}(\mathbf{x}) - \sigma(\mathbf{x}) > 0$, entonces $\widehat{D}_n(\mathbf{x}) = J_n^+(\widehat{\lambda}(\mathbf{x}))$, mientras que si $\widehat{\lambda}(\mathbf{x}) = \widehat{s}(\mathbf{x}) - \sigma(\mathbf{x}) < 0$, $\widehat{D}_n(\mathbf{x}) = J_n^-(\widehat{\lambda}(\mathbf{x}))$. Luego, como $\widehat{\lambda}(\mathbf{x}) \xrightarrow{p} 0$, basta probar que

$$\lim_{n\to\infty}\lim_{d\to 0} P\left(\sup_{0\le\lambda\le d} |\widetilde{J}_n^+(\lambda)| > \varepsilon\right) = 0 \tag{4.3}$$

$$\lim_{n\to\infty}\lim_{d\to 0}\sup_{0\le\lambda\le d} \mathbb{E}(J_n^+(\lambda)) = 0 \tag{4.4}$$

y un resultado análogo para $J_n^-(\lambda)$ en lugar de $J_n^+(\lambda)$. Probaremos solamente (4.3) y (4.4) ya que el resultado para $J_n^-(\lambda)$ es análogo.

Usando que ψ es continua y acotada, por el Teorema de convergencia mayorada obtenemos que $\lim_{d\to 0}\sup_{0\le\lambda\le d}\mathbb{E}(J_n^+(\lambda)) = 0$ para todo n, de donde vale (4.4).

Para probar (4.3), como en el Lema 3.4.3, basta con ver que la sucesión de variables aleatorias, $\widetilde{J}_n^+(\lambda)$, en el espacio $\mathcal{C}([0,1])$ es ajustada.

De acuerdo al Teorema 12.3 de Billingsley (1968) es suficiente probar que

i) La sucesión $\left\{\widetilde{J}_n^+(0)\right\}$ es ajustada.

ii) Existen constantes $\gamma > 0$ y $\alpha > 1$ y una función no decreciente $\widetilde{F}$ en $[0,1]$ tal que

$$\mathbb{E}\left(\left|\widetilde{J}_n^+(\lambda_2) - \widetilde{J}_n^+(\lambda_1)\right|^\gamma\right) \le \left(\widetilde{F}(\lambda_2) - \widetilde{F}(\lambda_1)\right)^\alpha$$

para todo $0 \le \lambda_1 < \lambda_2$ y n suficientemente grande.

i) vale pues $\widetilde{J}_n^+(0) = 0$.

ii) Sea $0 < \lambda_1 < \lambda_2$, usando que $\psi_a(t) = a\psi(t/a)$ y $\zeta(t) = u\psi'(u)$, como en el Lema 3.4.3 obtenemos que

$$|H_{\lambda_2}(u) - H_{\lambda_1}(u)| \le (\|\psi\|_\infty + \|\zeta\|_\infty)(\lambda_2 - \lambda_1) = \widetilde{F}(\lambda_2) - \widetilde{F}(\lambda_1)$$

donde $\widetilde{F}(\lambda) = (\|\zeta\|_\infty + \|\psi\|_\infty)\lambda$.

Denotemos por

$$M_{i,n} = \prod_{\ell=1}^{d} K_\ell \left(\frac{X_{i\ell} - x_\ell}{h_\ell} \right) \xi_i \left(H_{\lambda_2}(Y_{i,\mathbf{x}}) - H_{\lambda_1}(Y_{i,\mathbf{x}}) \right) .$$

Como K_j es acotada, si llamamos $C = (1/i(t)) \prod_{\ell=1}^{d} \|K_\ell\|_\infty$ argumentos análogos a los del Lema 3.4.3 permiten mostrar que

$$|M_{i,n}| \leq C(\widetilde{F}(\lambda_2) - \widetilde{F}(\lambda_1)) .$$

Por otra parte, como

$$\mathbb{E}\left[\prod_{\ell=1}^{d} K_\ell \left(\frac{X_{i\ell} - x_\ell}{h_\ell} \right) \right] = \prod_{\ell=1}^{d} h_\ell \int \prod_{\ell=1}^{d} K_\ell \left(u_\ell \right) f_{\mathbf{X}}(\mathbf{x} + \mathbf{H}_d \, \mathbf{u}) d\mathbf{u} ,$$

$\mathbf{y} = \mathbf{x} + \mathbf{H}_d \, \mathbf{u} \in \mathcal{C}_0$ si $n \geq n_0$, $f_{\mathbf{X}}$ es acotada en $\mathcal{C}_0$ y $\int K_j(u)du = 1$ tenemos que

$$\begin{aligned}
\mathbb{E}\left(M_{i,n}^2\right) &\leq \left(\widetilde{F}(\lambda_2) - \widetilde{F}(\lambda_1)\right)^2 \mathbb{E}\left[\prod_{\ell=1}^{d} K_\ell^2 \left(\frac{X_{i\ell} - x_\ell}{h_\ell} \right) \xi_i^2 \right] \\
&\leq \left(\widetilde{F}(\lambda_2) - \widetilde{F}(\lambda_1)\right)^2 \frac{1}{i^2(t)} \prod_{\ell=1}^{d} \|K_\ell\|_\infty \, \mathbb{E}\left[\prod_{\ell=1}^{d} K_\ell \left(\frac{X_{i\ell} - x_\ell}{h_\ell} \right) \right] \\
&\leq C_1 \left(\widetilde{F}(\lambda_2) - \widetilde{F}(\lambda_1)\right)^2 \prod_{\ell=1}^{d} h_\ell
\end{aligned}$$

donde $C_1 = (1/i^2(t)) \prod_{\ell=1}^{d} \|K_\ell\|_\infty \sup_{\mathbf{u} \in \mathcal{C}_0} f_{\mathbf{X}}(\mathbf{u})$ para todo $n \geq n_0$.

Finalmente, como $\mathbb{E}(\widetilde{J}_n^+(\lambda)) = 0$, $\mathbb{E}\left(\widetilde{J}_n^+(\lambda_2) - \widetilde{J}_n^+(\lambda_1)\right)^2 = \text{VAR}\left(\widetilde{J}_n^+(\lambda_2) - \widetilde{J}_n^+(\lambda_1)\right)$, de donde por la independencia tenemos que

$$\mathbb{E}\left(\widetilde{J}_n^+(\lambda_2) - \widetilde{J}_n^+(\lambda_1)\right)^2 = \frac{1}{n \prod_{\ell=1}^{d} h_\ell} \text{VAR}\left(M_{i,n}\right) \leq \frac{1}{n \prod_{\ell=1}^{d} h_\ell} \sum_{i=1}^{n} \mathbb{E}\left(M_{i,n}^2\right) \leq C_1 \left(\widetilde{F}(\lambda_2) - \widetilde{F}(\lambda_1)\right)^2$$

para $n \geq n_0$, lo que concluye la demostración ii) y por lo tanto la de b) como queríamos.

c) Observemos que

$$\widetilde{C}_n(\mathbf{x}) = \frac{1}{\sqrt{n \prod_{j=1}^{d} h_j}} \sum_{i=1}^{n} \prod_{j=1}^{d} K_j \left(\frac{X_{i,j} - x_j}{h_j} \right) \xi_i \left[\psi\left(\frac{Y_i - g(\mathbf{x})}{\sigma(\mathbf{x})} \right) - \psi\left(\frac{Y_i - g(\mathbf{X}_i)}{\sigma(\mathbf{x})} \right) \right] .$$

Por lo tanto,

$$\begin{aligned}
\text{VAR}(\widetilde{C}_n(\mathbf{x})) &= \frac{1}{\prod_{j=1}^{d} h_j} \text{VAR}\left\{ \prod_{j=1}^{d} K_j \left(\frac{X_{1,j} - x_j}{h_j} \right) \xi_1 \left[\psi\left(\frac{Y_1 - g(\mathbf{x})}{\sigma(\mathbf{x})} \right) - \psi\left(\frac{Y_1 - g(\mathbf{X}_1)}{\sigma(\mathbf{x})} \right) \right]^2 \right\} \\
&\leq \frac{1}{\prod_{j=1}^{d} h_j} \mathbb{E}\left\{ \prod_{j=1}^{d} K_j^2 \left(\frac{X_{1,j} - x_j}{h_j} \right) \frac{p(\mathbf{X}_1)}{t(\mathbf{X}_1)} \Phi(\mathbf{X}_1, \mathbf{x}) \right\} ,
\end{aligned}$$

donde

$$\begin{aligned}\Phi(\mathbf{u},\mathbf{x}) &= \mathbb{E}\left\{\left[\psi\left(\frac{Y_1-g(\mathbf{x})}{\sigma(\mathbf{x})}\right)-\psi\left(\frac{Y_1-g(\mathbf{X}_1)}{\sigma(\mathbf{x})}\right)\right]^2|\mathbf{X}_1=\mathbf{u}\right\}\\ &= \mathbb{E}\left[\psi\left(\frac{\sigma(\mathbf{u})\epsilon_1+g(\mathbf{u})-g(\mathbf{x})}{\sigma(\mathbf{x})}\right)-\psi\left(\frac{\sigma(\mathbf{u})\epsilon_1}{\sigma(\mathbf{x})}\right)\right]^2.\end{aligned}$$

Luego, como $\Phi(\mathbf{x},\mathbf{x})=0$, obtenemos fácilmente que $\mathrm{VAR}(\widetilde{C}_n(\mathbf{x}))\to 0$. Por lo tanto para probar que $\widetilde{C}_n(\mathbf{x})\xrightarrow{p}a_1$ bastará ver que $\mathbb{E}(\widetilde{C}_n(\mathbf{x}))\to a_1$.

Como $\mathbb{E}\psi\left((Y_1-g(\mathbf{X}_1))/\sigma(\mathbf{x})\right)=\mathbb{E}\psi\left(V_1\right)=0$, donde $V_1=\sigma(\mathbf{X}_1)\epsilon_1/\sigma(\mathbf{x})$ tenemos que

$$\begin{aligned}\mathbb{E}(\widetilde{C}_n(\mathbf{x})) &= \frac{n}{\sqrt{n\prod_{j=1}^d h_j}}\mathbb{E}\left\{\prod_{j=1}^d K_j\left(\frac{X_{1,j}-x_j}{h_j}\right)\xi_1\psi\left(\frac{Y_1-g(\mathbf{x})}{\sigma(\mathbf{x})}\right)\right\}\\ &= \frac{n}{\sqrt{n\prod_{j=1}^d h_j}}\mathbb{E}\left\{\prod_{j=1}^d K_j\left(\frac{X_{1,j}-x_j}{h_j}\right)\frac{p(\mathbf{X}_1)}{t(\mathbf{X}_1)}\lambda(\mathbf{X}_1,\mathbf{x})\right\},\end{aligned}$$

donde

$$\lambda(\mathbf{u},\mathbf{x})=\mathbb{E}\left[\psi\left(\frac{\sigma(\mathbf{X}_1)\epsilon_1+g(\mathbf{X}_1)-g(\mathbf{x})}{\sigma(\mathbf{x})}\right)|\mathbf{X}_1=\mathbf{u}\right]=\mathbb{E}\left[\psi\left(\frac{\sigma(\mathbf{u})\varepsilon_1+g(\mathbf{u})-g(\mathbf{x})}{\sigma(\mathbf{x})}\right)\right]$$

es tal que $\lambda(\mathbf{x},\mathbf{x})=0$ y $\lambda(\mathbf{u},\mathbf{x})$ es diferenciable respecto de $\mathbf{u}$. Más aún, como

$$\begin{aligned}\psi\left(\frac{\sigma(\mathbf{u})\varepsilon_1+g(\mathbf{u})-g(\mathbf{x})}{\sigma(\mathbf{x})}\right) &= \psi\left(\frac{\sigma(\mathbf{u})\varepsilon_1}{\sigma(\mathbf{x})}\right)+\frac{1}{\sigma(\mathbf{x})}\psi'\left(\frac{\sigma(\mathbf{u})\varepsilon_1}{\sigma(\mathbf{x})}\right)(g(\mathbf{u})-g(\mathbf{x}))\\ &\quad+\frac{1}{2}\frac{1}{\sigma^2(\mathbf{x})}\psi''\left(\frac{\sigma(\mathbf{u})\varepsilon_1}{\sigma(\mathbf{x})}\right)(g(\mathbf{u})-g(\mathbf{x}))^2\\ &\quad+\frac{1}{2}\frac{1}{\sigma^2(\mathbf{x})}\left[\psi''\left(\frac{\sigma(\mathbf{u})\varepsilon_1}{\sigma(\mathbf{x})}+\theta(\mathbf{u})\right)-\psi''\left(\frac{\sigma(\mathbf{u})\varepsilon_1}{\sigma(\mathbf{x})}\right)\right](g(\mathbf{u})-g(\mathbf{x}))^2,\end{aligned}$$

donde $\theta(\mathbf{u})$ es un punto intermedio entre 0 y $(g(\mathbf{u})-g(\mathbf{x}))/\sigma(\mathbf{x})$, y como ψ'' es impar pues ψ lo es, tenemos que

$$\begin{aligned}\lambda(\mathbf{u},\mathbf{x}) &= \frac{g(\mathbf{u})-g(\mathbf{x})}{\sigma(\mathbf{x})}\mathbb{E}\psi'\left(\frac{\sigma(\mathbf{u})\varepsilon_1}{\sigma(\mathbf{x})}\right)+\frac{1}{2}\frac{(g(\mathbf{u})-g(\mathbf{x}))^2}{\sigma^2(\mathbf{x})}\mathbb{E}\left[\psi''\left(\frac{\sigma(\mathbf{u})\varepsilon_1}{\sigma(\mathbf{x})}+\theta(\mathbf{u})\right)-\psi''\left(\frac{\sigma(\mathbf{u})\varepsilon_1}{\sigma(\mathbf{x})}\right)\right]\\ &= \frac{g(\mathbf{u})-g(\mathbf{x})}{\sigma(\mathbf{x})}\lambda_1(\mathbf{u},\mathbf{x})+\frac{1}{2}\frac{(g(\mathbf{u})-g(\mathbf{x}))^2}{\sigma^2(\mathbf{x})}\lambda_2(\mathbf{u},\mathbf{x})\end{aligned}$$

donde $\lambda_1(\mathbf{x},\mathbf{x})=A_0(\psi)$ y $|\lambda_2(\mathbf{u},\mathbf{x})|\le C|\theta(\mathbf{u})|\le C|g(\mathbf{u})-g(\mathbf{x})|/\sigma(\mathbf{x})$ pues ψ'' es Lipschitz.

Por lo tanto, tenemos que $\mathbb{E}(\widetilde{C}_n(\mathbf{x}))=C_{1,n}/\sigma(\mathbf{x})+C_{2,n}/\sigma^2(\mathbf{x})$ donde

$$\begin{aligned}C_{1,n} &= \frac{n}{\sqrt{n\prod_{j=1}^d h_j}}\mathbb{E}\left\{\prod_{j=1}^d K_j\left(\frac{X_{1,j}-x_j}{h_j}\right)\frac{p(\mathbf{X}_1)}{t(\mathbf{X}_1)}\left[g(\mathbf{X}_1)-g(\mathbf{x})\right]\lambda_1(\mathbf{X}_1,\mathbf{x})\right\}\\ C_{2,n} &= \frac{1}{2}\frac{n}{\sqrt{n\prod_{j=1}^d h_j}}\mathbb{E}\left\{\prod_{j=1}^d K_j\left(\frac{X_{1,j}-x_j}{h_j}\right)\frac{p(\mathbf{X}_1)}{t(\mathbf{X}_1)}(g(\mathbf{X}_1)-g(\mathbf{x}))^2\lambda_2(\mathbf{X}_1,\mathbf{x})\right\}\end{aligned}$$

Como g_j es diferenciable, $|g(\mathbf{X}_1)-g(\mathbf{x})|\leq\sum_{j=1}^d h_j\leq d\max_{1\leq j\leq d}h_j$ si $\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_1-\mathbf{x})\neq 0$ de donde, usando que $|\lambda_2(\mathbf{u},\mathbf{x})|\leq C|g(\mathbf{u})-g(\mathbf{x})|/\sigma(\mathbf{x})$, $p(\mathbf{u})\leq 1$, $t(\mathbf{u})\geq i(t)$ y $f_{\mathbf{X}}$ es acotada por digamos A_f, en un entorno compacto de $\mathbf{x}$, obtenemos que

$$\begin{aligned}
|C_{2,n}| &\leq \frac{C}{2}\frac{n}{\sqrt{n\prod_{j=1}^d h_j}}\mathbb{E}\left\{\prod_{j=1}^d\left|K_j\left(\frac{X_{1,j}-x_j}{h_j}\right)\right|\frac{p(\mathbf{X}_1)}{t(\mathbf{X}_1)}\frac{|g(\mathbf{X}_1)-g(\mathbf{x})|^3}{\sigma(\mathbf{x})}\right\}\\
&\leq \frac{Cd^3}{2}\frac{1}{i(t)\,\sigma(\mathbf{x})}\frac{n}{\sqrt{n\prod_{j=1}^d h_j}}\max_{1\leq j\leq d}h_j^3\mathbb{E}\left\{\prod_{j=1}^d\left|K_j\left(\frac{X_{1,j}-x_j}{h_j}\right)\right|\right\}\\
&\leq \frac{A_f\,Cd^3}{2}\frac{1}{i(t)\,\sigma(\mathbf{x})}\left(n\prod_{j=1}^d h_j\sum_{j=1}^d h_j^4\right)^{\frac{1}{2}}\max_{1\leq j\leq d}h_j\int\prod_{j=1}^d|K_j(u_j)|\,d\mathbf{u}\,.
\end{aligned}$$

Por lo tanto, usando que $h_j\to 0$ y $n\prod_{j=1}^d h_j\,\sum_{j=1}^d h_j^4\to\beta$ deducimos que $C_{2,n}\to 0$.

Por otra parte, recordemos que $r(\mathbf{u})=p(\mathbf{u})/t(\mathbf{u})$

$$\begin{aligned}
C_{1,n} &= \sqrt{n\prod_{j=1}^d h_j}\int\prod_{j=1}^d K_j(u_j)\,r(\mathbf{u}\mathbf{H}_d+\mathbf{x})\,[g(\mathbf{u}\mathbf{H}_d+\mathbf{x})-g(\mathbf{x})]\,\lambda_1(\mathbf{u}\mathbf{H}_d+\mathbf{x},\mathbf{x})f_{\mathbf{X}}(\mathbf{u}\mathbf{H}_d+\mathbf{x})d\mathbf{u}\\
&= \sqrt{n\prod_{j=1}^d h_j}\int\prod_{j=1}^d K_j(u_j)\,m(\mathbf{u}\mathbf{H}_d+\mathbf{x})\upsilon(\mathbf{u}\mathbf{H}_d+\mathbf{x})d\mathbf{u}
\end{aligned}$$

donde $m(\mathbf{u})=g(\mathbf{u})-g(\mathbf{x})$ y $\upsilon(\mathbf{u})=r(\mathbf{u})f_{\mathbf{X}}(\mathbf{u})\lambda_1(\mathbf{u},\mathbf{x})$ y por lo tanto, $m(\mathbf{x})=0$. Más aún, como $\lambda_1(\mathbf{x},\mathbf{x})=A_0(\psi)$ y $\mathbb{E}\psi''(a\epsilon)=0$ tenemos que $\upsilon(\mathbf{x})=r(\mathbf{x})f_{\mathbf{X}}(\mathbf{x})A_0(\psi)$ y

$$\frac{\partial}{\partial u_j}\upsilon(\mathbf{u})\Big|_{\mathbf{u}=\mathbf{x}} = A_0(\psi)\left[\frac{\partial}{\partial u_j}r(\mathbf{u})\Big|_{\mathbf{u}=\mathbf{x}}f_{\mathbf{X}}(\mathbf{x})+\frac{\partial}{\partial u_j}f_{\mathbf{X}}(\mathbf{u})\Big|_{\mathbf{u}=\mathbf{x}}r(\mathbf{x})\right],$$

de donde como $m(\mathbf{u}\mathbf{H}_d+\mathbf{x})=\sum_{\ell=1}^d g'_\ell(x_\ell)u_\ell h_\ell+(1/2)\sum_{\ell=1}^d g''_\ell(\xi_\ell)u_\ell^2h_\ell^2$ donde ξ_ℓ es un punto intermedio entre x_ℓ y $x_\ell+u_\ell h_\ell$, obtenemos que

$$\begin{aligned}
m(\mathbf{u}\mathbf{H}_d+\mathbf{x})\upsilon(\mathbf{u}\mathbf{H}_d+\mathbf{x}) &= \upsilon(\mathbf{x})\sum_{\ell=1}^d g'_\ell(x_\ell)u_\ell h_\ell+\sum_{1\leq\ell,s\leq d}g'_\ell(x_\ell)\frac{\partial}{\partial u_s}\upsilon(\mathbf{u})\Big|_{\mathbf{u}=\mathbf{x}}u_\ell u_s h_\ell h_s\\
&\quad+\sum_{1\leq\ell,s\leq d}g'_\ell(x_\ell)\left[\frac{\partial}{\partial u_s}\upsilon(\mathbf{u})\Big|_{\mathbf{u}=\boldsymbol{\xi}^{(1)}}-\frac{\partial}{\partial u_s}\upsilon(\mathbf{u})\Big|_{\mathbf{u}=\mathbf{x}}\right]u_\ell u_s h_\ell h_s\\
&\quad+\frac{1}{2}\upsilon(\mathbf{x})\sum_{\ell=1}^d g''_\ell(x_\ell)u_\ell^2h_\ell^2+\frac{1}{2}\upsilon(\mathbf{x})\sum_{\ell=1}^d\left[g''_\ell(\xi_\ell)-g''_\ell(x_\ell)\right]u_\ell^2h_\ell^2\\
&\quad+\frac{1}{2}\sum_{1\leq\ell,s\leq d}g''_\ell(\xi_\ell)\frac{\partial}{\partial u_s}\upsilon(\mathbf{u})\Big|_{\mathbf{u}=\boldsymbol{\xi}^{(1)}}u_\ell^2u_s h_\ell^2h_s\,,
\end{aligned}$$

con $\boldsymbol{\xi}^{(1)}$ un punto intermedio entre $\mathbf{x}$ y $\mathbf{u}\mathbf{H}_d+\mathbf{x}$. Por lo tanto, si escribimos a $C_{1,n}=\sum_{j=1}^6 C_{1j,n}$

con

$$
\begin{aligned}
C_{11,n} &= \upsilon(\mathbf{x})\sqrt{n\prod_{j=1}^{d}h_j}\sum_{\ell=1}^{d}g'_\ell(x_\ell)h_\ell\int\prod_{j=1}^{d}K_j\left(u_j\right)u_\ell d\mathbf{u}\\
C_{12,n} &= \sqrt{n\prod_{j=1}^{d}h_j}\sum_{1\le\ell,s\le d}g'_\ell(x_\ell)\frac{\partial}{\partial u_s}\upsilon(\mathbf{u})\Big|_{\mathbf{u}=\mathbf{x}}h_\ell h_s\int\prod_{j=1}^{d}K_j\left(u_j\right)u_\ell u_s\,d\mathbf{u}\\
C_{13,n} &= \sqrt{n\prod_{j=1}^{d}h_j}\sum_{1\le\ell,s\le d}g'_\ell(x_\ell)h_\ell h_s\int\prod_{j=1}^{d}K_j\left(u_j\right)\left[\frac{\partial}{\partial u_s}\upsilon(\mathbf{u})\Big|_{\mathbf{u}=\boldsymbol{\xi}^{(1)}}-\frac{\partial}{\partial u_s}\upsilon(\mathbf{u})\Big|_{\mathbf{u}=\mathbf{x}}\right]u_\ell u_s\,d\mathbf{u}\\
C_{14,n} &= \frac{1}{2}\upsilon(\mathbf{x})\sqrt{n\prod_{j=1}^{d}h_j}\sum_{\ell=1}^{d}g''_\ell(x_\ell)h_\ell^2\int\prod_{j=1}^{d}K_j\left(u_j\right)u_\ell^2 d\mathbf{u}\\
C_{15,n} &= \frac{1}{2}\upsilon(\mathbf{x})\sqrt{n\prod_{j=1}^{d}h_j}\sum_{\ell=1}^{d}h_\ell^2\int\prod_{j=1}^{d}K_j\left(u_j\right)\left[g''_\ell(\xi_\ell)-g''_\ell(x_\ell)\right]u_\ell^2 d\mathbf{u}\\
C_{16,n} &= \frac{1}{2}\upsilon(\mathbf{x})\sqrt{n\prod_{j=1}^{d}h_j}\sum_{1\le\ell,s\le d}h_\ell^2 h_s\int\prod_{j=1}^{d}K_j\left(u_j\right)g''_\ell(\xi_\ell)\frac{\partial}{\partial u_s}\upsilon(\mathbf{u})\Big|_{\mathbf{u}=\boldsymbol{\xi}^{(1)}}u_\ell^2 u_s d\mathbf{u}
\end{aligned}
$$

con $\boldsymbol{\xi}^{(1)}$ un punto intermedio entre $\mathbf{x}$ y $\mathbf{u}\mathbf{H}_d+\mathbf{x}$. Luego, como $\int K_j(t)tdt=0$ tenemos que $C_{11,n}=0$. Además, como $n\prod_{j=1}^{d}h_j\,h_\ell^4\to\beta_\ell$, $h_s\to 0$ y υ es diferenciable y g_j es dos veces diferenciable tenemos que $C_{15,n}\to 0$ y $C_{16,n}\to 0$. Por otra parte, tenemos que si $C_g=\max_{1\le\ell\le d}|g'_\ell(x_\ell)|$ tenemos la cota

$$
|C_{13,n}|\le C_g\max_{1\le s\le d}\left(n\prod_{j=1}^{d}h_jh_s^4\right)^{\frac{1}{2}}\int\prod_{j=1}^{d}|K_j\left(u_j\right)|\sum_{1\le\ell,s\le d}\left|\frac{\partial}{\partial u_s}\upsilon(\mathbf{u})\Big|_{\mathbf{u}=\boldsymbol{\xi}^{(1)}}-\frac{\partial}{\partial u_s}\upsilon(\mathbf{u})\Big|_{\mathbf{u}=\mathbf{x}}\right|\left|u_\ell u_s\right|d\mathbf{u}
$$

de donde como $\boldsymbol{\xi}^{(1)}$ es un punto intermedio entre $\mathbf{x}$ y $\mathbf{u}\mathbf{H}_d+\mathbf{x}$ y $n\prod_{j=1}^{d}h_j\,h_\ell^4\to\beta_\ell$, obtenemos que $C_{13,n}\to 0$.

Usando que $\int\prod_{j=1}^{d}K_j\left(u_j\right)u_su_\ell\,d\mathbf{u}=0$ si $s\neq\ell$ y $\int\prod_{j=1}^{d}K_j\left(u_j\right)u_su_\ell\,d\mathbf{u}=s_\ell$ si $s=\ell$ obtenemos que

$$
\begin{aligned}
C_{12,n} &= \sqrt{n\prod_{j=1}^{d}h_j}\sum_{\ell=1}^{d}g'_\ell(x_\ell)\frac{\partial}{\partial u_\ell}\upsilon(\mathbf{u})\Big|_{\mathbf{u}=\mathbf{x}}h_\ell^2 s_\ell\longrightarrow\sum_{\ell=1}^{d}\beta_\ell^{\frac{1}{2}}g'_\ell(x_\ell)\frac{\partial}{\partial u_\ell}\upsilon(\mathbf{u})\Big|_{\mathbf{u}=\mathbf{x}}s_\ell\\
C_{14,n} &= \frac{1}{2}\upsilon(\mathbf{x})\sqrt{n\prod_{j=1}^{d}h_j}\sum_{\ell=1}^{d}g''_\ell(x_\ell)h_\ell^2 s_\ell\longrightarrow\frac{1}{2}\upsilon(\mathbf{x})\sum_{\ell=1}^{d}\beta_\ell^{\frac{1}{2}}g''_\ell(x_\ell)s_\ell
\end{aligned}
$$

de donde obtenemos que $\mathbb{E}(\widetilde{C}_n(\mathbf{x}))\to a_1$, lo que concluye la demostración de c).

Para probar d) indiquemos por $Z_i=\xi_i\psi\left(\dfrac{Y_i-g(\mathbf{X}_i)}{\sigma(\mathbf{x})}\right)$, $E(Z_1|\mathbf{X}_1)=0$. Luego, el resultado se sigue usando la distribución asintótica del estimador clásico de Nadaraya–Watson para variables acotadas aplicado a $(\mathbf{X}_i,Z_i)$. $\square$

Teorema 4.2.2. *Sea* $\mathbf{x}$ *un punto del interior de* $\mathcal{S}_f$ *tal que se cumple* ***A3*** *en un entorno compacto de* $\mathbf{x}$. *Supongamos que el estimador inicial* $\widetilde{g}_{\text{INI}}(\mathbf{x})$ *es tal que* $\widetilde{g}_{\text{INI}}(\mathbf{x}) - g(\mathbf{x}) = O_{\mathbb{P}}\left(\sum_{j=1}^d h_j^4\right)$. *Más aún, supongamos que* $f_{\mathbf{X}}(\mathbf{u})$, $r(\mathbf{u}) = p(\mathbf{u})/t(\mathbf{u})$ *y* $\sigma(\mathbf{u})$ *son diferenciables. Luego bajo* ***A1***, ***A4***, ***C1***, ***C2*** *y* ***C7***, *si además* ψ'' *es Lipschitz,* $\widehat{s}(\mathbf{x}) \xrightarrow{p} \sigma(\mathbf{x})$ *y* $n\prod_{j=1}^d h_j\, h_\ell^4 \to \beta_\ell$ *tenemos que*

$$\sqrt{n\prod_{j=1}^d h_j}\left(\widetilde{g}_{\text{OS},\text{M}_0}(\mathbf{x}) - g(\mathbf{x})\right) \xrightarrow{D} N\left(b_0(\mathbf{x}), v_0(\mathbf{x})\right)$$

donde $b_0(\mathbf{x})$ *y* $v_0(\mathbf{x})$ *están dados en el Teorema 4.2.1.*

DEMOSTRACIÓN. Observemos que

$$\sqrt{n\prod_{j=1}^d h_j}\left(\widetilde{g}_{\text{OS},\text{M}_0}(\mathbf{x}) - g(\mathbf{x})\right) = \sqrt{n\prod_{j=1}^d h_j}\left(\widetilde{g}_{\text{INI}}(\mathbf{x}) - g(\mathbf{x})\right) - W_n^{-1}\sqrt{n\prod_{j=1}^d h_j}\Psi_n\left(\widetilde{g}_{\text{INI}}(\mathbf{x})\right).$$

Por lo tanto, bastará probar que

a) $W_n \xrightarrow{p} -r(\mathbf{x})f_{\mathbf{X}}(\mathbf{x})A_0(\psi)/\sigma(\mathbf{x})$,

b) $\sqrt{n\prod_{j=1}^d h_j}\Psi_n\left(\widetilde{g}_{\text{INI}}(\mathbf{x})\right) \xrightarrow{D} N(a_1, \sigma_1)$ donde a_1 y σ_1 están definidas en (4.1) y (4.2), respectivamente

c) $\sqrt{n\prod_{j=1}^d h_j}\left(\widetilde{g}_{\text{INI}}(\mathbf{x}) - g(\mathbf{x})\right) \xrightarrow{p} 0$,

Como $\widetilde{g}_{\text{INI}}(\mathbf{x}) - g(\mathbf{x}) = O_{\mathbb{P}}\left(\sum_{j=1}^d h_j^4\right)$, c) es inmediato. Por otra parte, a) y b) se deducen usando argumentos análogos a los considerados en la demostración del Teorema 4.2.1. □

Para obtener la distribución asintótica del estimador de la componente marginal, supondremos un modelo homocedástico, o sea, que $\sigma(\mathbf{x}) \equiv \sigma$, y que tenemos un estimador robusto de escala $\widehat{s}$ de σ con tasa $n^{1/2}$. En el caso de observaciones completas, ejemplos de estimadores que cumplen esta condición son los $M-$estimadores de escala robustos de los residuos $Y_i - \widetilde{g}(\mathbf{X}_i)$, donde $\widetilde{g}(\mathbf{X}_i)$ es un estimador robusto preliminar consistente de g. También, como en el caso $d = 1$, podemos considerar un $M-$estimador de escala robusto de las diferencias $Y_i - Y_{i,[1]}$ donde $Y_{i,[1]}$ indica la respuesta asociada a la covariable $\mathbf{X}_j$ más cercana a $\mathbf{X}_i$, o sea, $\|\mathbf{X}_i - \mathbf{X}_{i,[1]}\| = \min_{j\neq i}\|\mathbf{X}_i - \mathbf{X}_j\|$. En el caso de tener respuestas faltantes, como en el caso en que se desea estimar el parámetro de posición μ, se deberá trabajar con un estimador basado en la muestra observada corregido por el *propensity score*, o sea, pesando a cada observación con $\delta_i = 1$ con el inverso de un estimador de la probabilidad $p(\mathbf{X}_i)$.

Teorema 4.2.3. *Sea* $\mathbf{x}$ *un punto del interior de* $\mathcal{S}_f$. *Supongamos que* $\sigma(\mathbf{x}) \equiv \sigma$, $\sup_{\mathbf{x}\in\mathcal{S}_Q}|\widetilde{g}_{\text{M}_0}(\mathbf{x}) - g(\mathbf{x})| \xrightarrow{p} 0$ *y se cumplen* ***A0***, ***A1***, ***C1***, ***C2***, ***C7*** *y* ***C8*** *y que* ψ'' *es Lipschitz. Más aún, supongamos que* $q_{\underline{1}}(\mathbf{u})$ *es continua,* $f_{\mathbf{X}}(\mathbf{u})$ *y* $r(\mathbf{u}) = p(\mathbf{u})/t(\mathbf{u})$ *son diferenciables y que* $\sup_{\mathbf{x}\in\mathcal{S}_Q}|g(\mathbf{x})| < \infty$

y $\sup_{\mathbf{x}\in\mathcal{S}_Q} f_{\mathbf{X}}(\mathbf{x}) < \infty$, $\inf_{\mathbf{x}\in\mathcal{C}} f_{\mathbf{X}}(\mathbf{x}) > 0$, $\inf_{\mathbf{x}\in\mathcal{C}} t(\mathbf{x}) > 0$ *y* $\inf_{\mathbf{x}\in\mathcal{C}} p(\mathbf{x}) > 0$ *para algún entorno compacto,* $\mathcal{C} \subset \mathcal{S}_f$*, de* $\mathcal{S}_Q$*. Si además,* $\sqrt{n}(\widehat{s}-\sigma) = O_{\mathbb{P}}(1)$*,* $h_1 = \beta n^{-1/5}$*,* $h_j = \widetilde{h}$ *para* $j \neq 1$ *es tal que* $\widetilde{h} = o(h_1)$ *y* $\log n/(n^{2/5}\widetilde{h}^{d-1}) \to 0$*, entonces*

$$\sqrt{nh_1}\,[\widehat{g}_{1,\mathrm{M}_0}(x_1) - g_1(x_1)] \xrightarrow{D} N\left(b_{0,1}(x_1), \sigma_1^2(x_1)\right)$$

donde

$$\begin{aligned} b_{0,1}(x_1) &= \beta^{5/2} s_1 \left[g_1'(x_1) \int \frac{1}{r(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x})} \frac{\partial}{\partial u_1} v(\mathbf{u})\Big|_{\mathbf{u}=\mathbf{x}} q_{\underline{1}}(\mathbf{x}_{\underline{1}}) d\mathbf{x}_{\underline{1}} + \frac{1}{2} g_1''(x_1) \right] \\ \sigma_1^2(x_1) &= \sigma^2 \frac{\mathbb{E}(\psi^2(\varepsilon))}{[\mathbb{E}\psi'(\varepsilon)]^2} \int K_1^2(u) du \int \frac{1}{f_{\mathbf{X}}(x_1, \mathbf{x}_{\underline{1}}) p(x_1, \mathbf{x}_{\underline{1}})} q_{\underline{1}}^2(\mathbf{x}_{\underline{1}}) d\mathbf{x}_{\underline{1}}\,, \end{aligned}$$

$\mathbf{x}_{\underline{1}} = (x_2, \ldots, x_d)^{\mathrm{T}}$ *y* $v(\mathbf{u}) = r(\mathbf{u}) f_{\mathbf{X}}(\mathbf{u})$.

Observación 4.2.1. En el caso en que $\widetilde{h} = \gamma n^{-\tau}$, para que $\widetilde{h} = o(h_1)$, es necesario que $\tau > 1/5$ ya que $h_1 = \beta n^{-1/5}$. Por otra parte, la condición $\sqrt{nh_1}\left(nh_1\widetilde{h}^{d-1}/\log n\right)^{-3/4} \to 0$ es equivalente a $\log(n)/n^{2/5}\widetilde{h}^{d-1} \to 0$, luego se cumplirá si $2/5 - \tau(d-1) > 0$. Por lo tanto, debemos tener

$$\frac{1}{5} < \tau < \frac{2}{5(d-1)}$$

de donde, $d \leq 2$.

DEMOSTRACIÓN. Como en el Teorema 4.2.1, $\widetilde{g}_{\mathrm{M}_0}(\mathbf{x})$ es solución de $\Psi_n(a) = 0$ donde $\Psi_n(a)$ está definida en (2.13) vimos que $\widetilde{g}_{\mathrm{M}_0}(\mathbf{x})$. Para reforzar la dependencia en el estimador de escala sea

$$\Psi_n(a, s) = \frac{1}{n} \sum_{i=1}^{n} \psi\left(\frac{Y_i - a}{s}\right) \mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i - \mathbf{x}) \xi_i\,.$$

Luego, $\Psi_n(\widetilde{g}_{\mathrm{M}_0}(\mathbf{x}), \widehat{s}) = 0$. Usando que ψ es Lispchitz y que $\zeta(u) = u\psi'(u)$ es acotada, es fácil ver que $\widehat{D}_n = \sup_{\mathbf{x}\in\mathcal{S}_Q} \sup_a |\widehat{s}\Psi_n(a, \widehat{s}) - \sigma\Psi_n(a, \sigma)| = O_{\mathbb{P}}(1/\sqrt{n})$, por lo tanto, $\Psi(\widetilde{g}_{\mathrm{M}_0}(\mathbf{x}), \sigma) = O_{\mathbb{P}}(1/\sqrt{n})$.

Definamos $\widetilde{g}(\mathbf{x})$ como la solución de $\Psi_n(a, \sigma) = 0$ (es decir, cuando suponemos la escala conocida). Por la Proposición 3.1.1, $\sup_{\mathbf{x}\in\mathcal{C}} |\widetilde{g}(\mathbf{x}) - g(\mathbf{x})| = o_{\mathbb{P}}(1)$. Luego, tenemos que

$$\widehat{D}_n = \sup_{\mathbf{x}\in\mathcal{S}_Q} |\widetilde{g}_{\mathrm{M}_0}(\mathbf{x}) - \widetilde{g}(\mathbf{x})| = O_{\mathbb{P}}(1/\sqrt{n})\,. \tag{4.5}$$

Efectivamente, para probar (4.5) observemos que un desarrollo de Taylor de orden uno y el hecho que $\Psi_n(\widetilde{g}(\mathbf{x}), \sigma) = 0$ permite obtener $\Psi_n(\widetilde{g}_{\mathrm{M}_0}(\mathbf{x}), \sigma) = \Psi_n(\widetilde{g}(\mathbf{x}), \sigma) + \widehat{D}_{1,n}(\mathbf{x})(\widetilde{g}_{\mathrm{M}_0}(\mathbf{x}) - \widetilde{g}(\mathbf{x})) = \widehat{D}_{1,n}(\mathbf{x})(\widetilde{g}_{\mathrm{M}_0}(\mathbf{x}) - \widetilde{g}(\mathbf{x}))$ donde

$$\widehat{D}_{1,n}(\mathbf{x}) = \frac{\partial \Psi_n(a, \sigma)}{\partial a}\Big|_{a=\xi_n} = -\frac{1}{\sigma}\frac{1}{n}\sum_{i=1}^{n} \psi'\left(\frac{Y_i - \xi_n}{\sigma}\right) \mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i - \mathbf{x}) \xi_i$$

con $\xi_n = \xi_n(\mathbf{x})$ un punto intermedio entre $\widetilde{g}_{\mathrm{M}_0}(\mathbf{x})$ y $\widetilde{g}(\mathbf{x})$. Observemos que como $\sup_{\mathbf{x}\in\mathcal{S}_Q} |\widetilde{g}(\mathbf{x}) - g(\mathbf{x})| = o_{\mathbb{P}}(1)$ y $\sup_{\mathbf{x}\in\mathcal{S}_Q} |\widetilde{g}_{\mathrm{M}_0}(\mathbf{x}) - g(\mathbf{x})| = o_{\mathbb{P}}(1)$ entonces, $\sup_{\mathbf{x}\in\mathcal{S}_Q} |\xi_n(\mathbf{x}) - g(\mathbf{x})| = o_{\mathbb{P}}(1)$. Sea $D_0(\mathbf{x}) = -A_0(\psi)r(\mathbf{x})f_{\mathbf{X}}(\mathbf{x})/\sigma$. Observando que $i(|D_0(\mathbf{x})|) > 0$, $\Psi(\widetilde{g}_{\mathrm{M}_0}(\mathbf{x}), \sigma) = O_{\mathbb{P}}(1/\sqrt{n})$ y $\Psi_n(\widetilde{g}_{\mathrm{M}_0}(\mathbf{x}), \sigma) = \widehat{D}_{1,n}(\mathbf{x})(\widetilde{g}_{\mathrm{M}_0}(\mathbf{x}) - \widetilde{g}(\mathbf{x}))$ para obtener (4.5) bastará mostrar que

$$\sup_{\mathbf{x}\in\mathcal{S}_Q} |\widehat{D}_{1,n}(\mathbf{x}) - D_0(\mathbf{x})| \;=\; o_{\mathbb{P}}(1)\,. \tag{4.6}$$

Tenemos que $\widehat{D}_{1,n}(\mathbf{x}) = (1/\sigma)\left(\widehat{D}_{11,n}(\mathbf{x}) + \widehat{D}_{12,n}(\mathbf{x}) + \widehat{D}_{13,n}(\mathbf{x})\right)$ donde

$$\begin{aligned}
\widehat{D}_{11,n}(\mathbf{x}) &= \frac{1}{n}\sum_{i=1}^{n} \mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i - \mathbf{x})\xi_i \left[\psi'\left(\frac{Y_i - g(\mathbf{x})}{\sigma}\right) - \psi'\left(\frac{Y_i - \xi_n}{\sigma}\right)\right] \\
\widehat{D}_{12,n}(\mathbf{x}) &= \frac{1}{n}\sum_{i=1}^{n} \mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i - \mathbf{x})\xi_i \left[\psi'\left(\frac{Y_i - g(\mathbf{X}_i)}{\sigma}\right) - \psi'\left(\frac{Y_i - g(\mathbf{x})}{\sigma}\right)\right] \\
\widehat{D}_{13,n}(\mathbf{x}) &= -\frac{1}{n}\sum_{i=1}^{n} \mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i - \mathbf{x})\xi_i \psi'\left(\frac{Y_i - g(\mathbf{X}_i)}{\sigma}\right).
\end{aligned}$$

Usando que ψ' es Lipschitz, que $\sup_{\mathbf{x}\in\mathcal{S}_Q} |\xi_n(\mathbf{x}) - g(\mathbf{x})| = o_{\mathbb{P}}(1)$, $\sup_{\mathbf{x}\in\mathcal{S}_Q} (1/n)\sum_{i=1}^{n} |\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i - \mathbf{x})| = O_{\mathbb{P}}(1)$ y $\inf_{\mathbf{x}\in\mathcal{C}} t(\mathbf{x}) > 0$, con $\mathcal{C}$ un entorno de $\mathcal{S}_Q$ deducimos que $\sup_{\mathbf{x}\in\mathcal{S}_Q} |\widehat{D}_{11,n}(\mathbf{x})| = o_{\mathbb{P}}(1)$. Por otra parte, como $\sup_{|\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i - \mathbf{x})|\neq 0} |g(\mathbf{X}_i) - g(\mathbf{x})| \leq C\sum_{j=1}^{d} h_j$, tenemos que $\sup_{\mathbf{x}\in\mathcal{S}_Q} |\widehat{D}_{12,n}(\mathbf{x})| = o_{\mathbb{P}}(1)$. Por último, usando el Lema 3.4.6, resulta que $\sup_{\mathbf{x}\in\mathcal{S}_Q} |\widehat{D}_{13,n}(\mathbf{x}) - D_0(\mathbf{x})| = o_{\mathbb{P}}(1)$, concluyendo la demostración de (4.6).

Integrando respecto de todas las variables menos la x_1, se tiene que

$$\begin{aligned}
&\left|\sqrt{nh_1}(\widehat{g}_{1,\mathrm{M}_0}(\mathbf{x}) - g_1(\mathbf{x})) - \sqrt{nh_1}\int [\widetilde{g}(\mathbf{x}) - g(\mathbf{x})]\, q_{\underline{1}}(\mathbf{x}_{\underline{1}})\, d\mathbf{x}_{\underline{1}}\right| \\
&= \left|\sqrt{nh_1}\int [\widetilde{g}_{\mathrm{M}_0}(\mathbf{x}) - g(\mathbf{x}) - (\widetilde{g}(\mathbf{x}) - g(\mathbf{x}))]\, q_{\underline{1}}(\mathbf{x}_{\underline{1}})\, d\mathbf{x}_{\underline{1}}\right| \leq \sqrt{h_1}\sqrt{n}\widehat{D}_n \xrightarrow{p} 0\,,
\end{aligned} \tag{4.7}$$

de donde, si $\widehat{g}_1(x_1) = \int \widetilde{g}(\mathbf{x}) q_{\underline{1}}(\mathbf{x}_{\underline{1}})\, d\mathbf{x}_{\underline{1}}$, bastará hallar la distribución asintótica de

$$\sqrt{nh_1}[\widehat{g}_1(x_1) - g_1(x_1)] = \sqrt{nh_1}\int [\widetilde{g}(\mathbf{x}) - g(\mathbf{x})]\, q_{\underline{1}}(\mathbf{x}_{\underline{1}})\, d\mathbf{x}_{\underline{1}}\,,$$

es decir, nos hemos reducido a probar el resultado en el caso en que la escala es conocida.

Como en el Teorema 4.2.1, usando un desarrollo de Taylor de orden uno, obtenemos que

$$\widetilde{g}(\mathbf{x}) - g(\mathbf{x}) = \sigma\, \widehat{A}_{0,n}^{-1}(\mathbf{x})\widehat{A}_{1,n}(\mathbf{x})$$

con

$$\begin{aligned}
\widehat{A}_{0,n}(\mathbf{x}) &= \frac{1}{n}\sum_{i=1}^{n}\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i-\mathbf{x})\,\xi_i\psi'\left(\frac{Y_i-\widehat{\theta}_1(\mathbf{x})}{\sigma}\right)=\widehat{A}_{01,n}(\mathbf{x})+\widehat{A}_{02,n}(\mathbf{x})\\
\widehat{A}_{01,n}(\mathbf{x}) &= \frac{1}{n}\sum_{i=1}^{n}\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i-\mathbf{x})\,\xi_i\psi'\left(\frac{Y_i-g(\mathbf{x})}{\sigma}\right)\\
\widehat{A}_{02,n}(\mathbf{x}) &= \frac{1}{2\sigma}(\widetilde{g}(\mathbf{x})-g(\mathbf{x}))\frac{1}{n}\sum_{i=1}^{n}\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i-\mathbf{x})\,\xi_i\psi''\left(\frac{Y_i-\widehat{\theta}_2(\mathbf{x})}{\sigma}\right)\\
\widehat{A}_{1,n}(\mathbf{x}) &= \frac{1}{n}\sum_{i=1}^{n}\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i-\mathbf{x})\,\xi_i\psi\left(\frac{Y_i-g(\mathbf{x})}{\sigma}\right)
\end{aligned}$$

donde $\widehat{\theta}_1(\mathbf{x})$ y $\widehat{\theta}_2(\mathbf{x})$ son puntos intermedios entre $\widetilde{g}(\mathbf{x})$ y $g(\mathbf{x})$. Sea $A_0(\psi)=\mathbb{E}(\psi'(\varepsilon))$, $A_0(\mathbf{u})=v(\mathbf{u})A_0(\psi)$ donde $v(\mathbf{u})=r(\mathbf{u})f_{\mathbf{X}}(\mathbf{u})$. Como $\sup_{\mathbf{x}\in\mathcal{S}_Q}|\widehat{A}_{0,n}(\mathbf{x})-A_0(\mathbf{x})|=o_{\mathbb{P}}(1)$, empezaremos estudiando el comportamiento de

$$\widehat{B}_n=\sqrt{nh_1}\sigma\int A_0^{-1}(\mathbf{x})\widehat{A}_{1,n}(\mathbf{x})q_{\underline{1}}(\mathbf{x}_{\underline{1}})\,d\mathbf{x}_{\underline{1}}\,.$$

Observemos que $\widehat{B}_n=B_{n,1}+\widehat{B}_{n,2}$ donde

$$\begin{aligned}
B_{n,1} &= \sigma\sqrt{nh_1}\int A_0^{-1}(\mathbf{x})\mathbb{E}(\widehat{A}_{1,n}(\mathbf{x}))q_{\underline{1}}(\mathbf{x}_{\underline{1}})d\mathbf{x}_{\underline{1}}\\
\widehat{B}_{n,2} &= \sigma\sqrt{nh_1}\int A_0^{-1}(\mathbf{x})\left[\widehat{A}_{1,n}(\mathbf{x})-\mathbb{E}(\widehat{A}_{1,n}(\mathbf{x}))\right]q_{\underline{1}}(\mathbf{x}_{\underline{1}})d\mathbf{x}_{\underline{1}}
\end{aligned}$$

Veremos que

a) $B_{n,1}\to b_{0,1}(x_1)$

b) $\widehat{B}_{n,2}\xrightarrow{D}N(0,\sigma_1^2(x_1))$.

<u>Empezaremos calculando $\mathbb{E}\widehat{A}_{1,n}(\mathbf{x})$ ya que la usaremos en los pasos a seguir.</u>

Definamos $\lambda(\mathbf{u},\mathbf{x})$ como

$$\lambda(\mathbf{u},\mathbf{x})=\mathbb{E}\left[\psi\left(\varepsilon_1+\frac{g(\mathbf{X}_1)-g(\mathbf{x})}{\sigma}|\mathbf{X}_1=\mathbf{u}\right)\right]=\mathbb{E}\left[\psi\left(\varepsilon_1+\frac{g(\mathbf{u})-g(\mathbf{x})}{\sigma}\right)\right]\,.$$

Tenemos que

$$\lambda(\mathbf{u},\mathbf{x})=\frac{g(\mathbf{u})-g(\mathbf{x})}{\sigma}A_0(\psi)+\frac{1}{2\sigma}(g(\mathbf{u})-g(\mathbf{x}))^2\lambda_2(\mathbf{u},\mathbf{x})$$

con $\lambda_2(\mathbf{u},\mathbf{x})=\mathbb{E}\psi''(\epsilon+\theta(\mathbf{u}))-\mathbb{E}\psi''(\epsilon)$ con $\theta(\mathbf{u})$ un punto intermedio entre 0 y $(g(\mathbf{u})-g(\mathbf{x}))/\sigma$. Como en el Teorema 4.2.1 tenemos que $|\lambda_2(\mathbf{u},\mathbf{x})|\le C|\theta(\mathbf{u})|\le C|g(\mathbf{u})-g(\mathbf{x})|/\sigma$ pues ψ'' es Lipschitz. Por lo tanto, $\mathbb{E}\widehat{A}_{1,n}(\mathbf{x})=\mathbb{E}\left[\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_1-\mathbf{x})\psi\left((Y_1-g(\mathbf{x}))/\sigma\right)\xi_1\right]=\mathbb{E}\left[\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_1-\mathbf{x})r(\mathbf{X}_1)\lambda(\mathbf{X}_1,\mathbf{x})\right]$ puede escribirse como

$$\mathbb{E}\widehat{A}_{1,n}(\mathbf{x})=\frac{A_0(\psi)}{\sigma}A_{11,n}+\frac{1}{2\sigma^2}A_{12,n}$$

donde

$$\begin{aligned}
A_{11,n} &= \frac{1}{h_1\widetilde{h}^{d-1}}\mathbb{E}\left[K_1\left(\frac{X_{i,1}-x_1}{h_1}\right)\prod_{j=2}^{d}K_j\left(\frac{X_{1,j}-x_j}{\widetilde{h}}\right)r(\mathbf{X}_1)(g(\mathbf{X}_1)-g(\mathbf{x}))\right]\\
&= \int\prod_{j=1}^{d}K_j(u_j)v(\mathbf{x}+\mathbf{H}_d\mathbf{u})m(\mathbf{x}+\mathbf{H}_d\mathbf{u})d\mathbf{u}\\
A_{12,n} &= \frac{1}{h_1\widetilde{h}^{d-1}}\mathbb{E}\left[K_1\left(\frac{X_{i,1}-x_1}{h_1}\right)\prod_{j=2}^{d}K_j\left(\frac{X_{i,j}-x_j}{\widetilde{h}}\right)r(\mathbf{X}_1)(g(\mathbf{X}_1)-g(\mathbf{x}))^2\lambda_2(\mathbf{X}_1,\mathbf{x})\right]\\
&= \int\prod_{j=1}^{d}K_j(u_j)v(\mathbf{x}+\mathbf{H}_d\mathbf{u})m^2(\mathbf{x}+\mathbf{H}_d\mathbf{u})\lambda_2(\mathbf{x}+\mathbf{H}_d\mathbf{u},\mathbf{x})d\mathbf{u}
\end{aligned}$$

con $m(\mathbf{u})=g(\mathbf{u})-g(\mathbf{x})$, luego, $A_{11,n}=\sum_{j=1}^{6}A_{11j,n}$ donde

$$\begin{aligned}
A_{111,n} &= v(\mathbf{x})\left\{\sum_{\ell=1}^{d}g'_\ell(x_\ell)h_\ell\int\prod_{j=1}^{d}K_j(u_j)u_\ell\,d\mathbf{u}\right\}\\
A_{112,n} &= \sum_{1\le\ell,s\le d}g'_\ell(x_\ell)\left.\frac{\partial}{\partial u_s}v(\mathbf{u})\right|_{\mathbf{u}=\mathbf{x}}h_\ell h_s\int\prod_{j=1}^{d}K_j(u_j)u_\ell u_s\,d\mathbf{u}\\
A_{113,n} &= \sum_{1\le\ell,s\le d}g'_\ell(x_\ell)h_\ell h_s\int\prod_{j=1}^{d}K_j(u_j)\left[\left.\frac{\partial}{\partial u_s}v(\mathbf{u})\right|_{\mathbf{u}=\boldsymbol{\xi}^{(1)}}-\left.\frac{\partial}{\partial u_s}v(\mathbf{u})\right|_{\mathbf{u}=\mathbf{x}}\right]u_\ell u_s\,d\mathbf{u}\\
A_{114,n} &= \frac{1}{2}v(\mathbf{x})\left\{\sum_{\ell=1}^{d}g''_\ell(x_\ell)h_\ell^2\int\prod_{j=1}^{d}K_j(u_j)u_\ell^2\,d\mathbf{u}\right\}=\frac{1}{2}v(\mathbf{x})\left\{\sum_{\ell=1}^{d}g''_\ell(x_\ell)h_\ell^2\,s_\ell\right\}\\
A_{115,n} &= \frac{1}{2}v(\mathbf{x})\left\{\sum_{\ell=1}^{d}h_\ell^2\int\prod_{j=1}^{d}K_j(u_j)\left[g''_\ell(\xi_\ell)-g''_\ell(x_\ell)\right]u_\ell^2\,d\mathbf{u}\right\}\\
A_{116,n} &= \frac{1}{2}\left\{\sum_{1\le\ell,s\le d}h_\ell^2h_s\int\prod_{j=1}^{d}K_j(u_j)g''_\ell(\xi_\ell)\left.\frac{\partial}{\partial u_s}v(\mathbf{u})\right|_{\mathbf{u}=\boldsymbol{\xi}^{(1)}}u_\ell^2u_s\,d\mathbf{u}\right\}
\end{aligned}$$

con $\boldsymbol{\xi}^{(1)}$ un punto intermedio entre $\mathbf{x}$ y $\mathbf{x}+\mathbf{H}_d\mathbf{u}$. Luego, como $\int K_j(t)t\,dt=0$ tenemos que $A_{111,n}=0$. Usando que $\int\prod_{j=1}^{d}K_j(u_j)u_su_\ell\,d\mathbf{u}=0$ si $\ell\neq s$ y es igual a s_ℓ si $\ell=j$ obtenemos que

$$A_{112,n} = \sum_{\ell=1}^{d}g'_\ell(x_\ell)\left.\frac{\partial}{\partial u_\ell}v(\mathbf{u})\right|_{\mathbf{u}=\mathbf{x}}h_\ell^2\,s_\ell\,.$$

Por otra parte, como g''_ℓ es acotada y uniformemente continua en $\mathcal{S}_f$ y f_X y r son continuamente diferenciables, tenemos que $(\partial v(\mathbf{u})/\partial u_s)|_{\mathbf{u}=\boldsymbol{\xi}^{(1)}}$ es acotada de donde

$$\sup_{\mathbf{x}\in\mathcal{S}_Q}|A_{115,n}|=(h_1^2+\widetilde{h}^2)\,o(1)\qquad\sup_{\mathbf{x}\in\mathcal{S}_Q}|A_{116,n}|=O(h_1^3+h_1^2\widetilde{h}+h_1\widetilde{h}^2+\widetilde{h}^3)\tag{4.8}$$

Por otra parte, como tenemos que $(\partial v(\mathbf{u})/\partial u_s)|_{\mathbf{u}}$ es uniformemente continua y K_j tienen soporte en $[-1,1]$, de modo tal que $\mathbf{x}+\mathbf{H}_d\mathbf{u}$ pertenece a un entorno de radio $\max_j h_j$ de $\mathbf{x}$, tenemos que

$$\sup_{\mathbf{x}\in\mathcal{S}_Q}\left|\frac{\partial}{\partial u_s}v(\mathbf{u})\Big|_{\mathbf{u}=\boldsymbol{\xi}^{(1)}}-\frac{\partial}{\partial u_s}v(\mathbf{u})\Big|_{\mathbf{u}=\mathbf{x}}\right|=o(1)$$

con lo cual

$$\sup_{\mathbf{x}\in\mathcal{S}_Q}|A_{113,n}|=(h_1+\widetilde{h})^2\,o(1)\,. \tag{4.9}$$

Por otra parte, como $|\lambda_2(\mathbf{u},\mathbf{x})|\le C|\theta(\mathbf{u})|\le C|g(\mathbf{u})-g(\mathbf{x})|/\sigma=Cm(\mathbf{u})/\sigma$ y g_j' y v son acotadas

$$\begin{aligned}\sup_{\mathbf{x}\in\mathcal{S}_Q}|A_{12,n}(\mathbf{x})| &\le \sup_{\mathbf{x}\in\mathcal{S}_Q}\int\prod_{j=1}^d|K_j|(u_j)v(\mathbf{x}+\mathbf{H}_d\mathbf{u})m^2(\mathbf{x}+\mathbf{H}_d\mathbf{u})|\lambda_2(\mathbf{x}+\mathbf{H}_d\mathbf{u},\mathbf{x})|d\mathbf{u}\\ &\le \frac{C}{\sigma}\sup_{\mathbf{x}\in\mathcal{S}_Q}\int\prod_{j=1}^d|K_j|(u_j)v(\mathbf{x}+\mathbf{H}_d\mathbf{u})m^3(\mathbf{x}+\mathbf{H}_d\mathbf{u})d\mathbf{u}\le\frac{\widetilde{C}}{\sigma}(\sum_{j=1}^d h_j)^3\le C_2(h_1+\widetilde{h})^3\end{aligned}$$

Resumiendo, tenemos que, como $h_1\to 0$ y $\widetilde{h}\to 0$, (4.8) y 4.9) implican que

$$\begin{aligned}\mathbb{E}\widehat{A}_{1,n}(\mathbf{x}) &= \frac{A_0(\psi)}{\sigma}A_{11,n}+\frac{1}{2\sigma^2}A_{12,n}\\ &= \frac{A_0(\psi)}{\sigma}\left\{\sum_{\ell=1}^d g_\ell'(x_\ell)\,\frac{\partial}{\partial u_\ell}v(\mathbf{u})\Big|_{\mathbf{u}=\mathbf{x}}h_\ell^2\,s_\ell+\frac{1}{2}v(\mathbf{x})\left[\sum_{\ell=1}^d g_\ell''(x_\ell)h_\ell^2\,s_\ell\right]\right\}+\nu_n(\mathbf{x})\,,\end{aligned} \tag{4.10}$$

donde $\sup_{\mathbf{x}\in\mathcal{S}_Q}|\nu_n(\mathbf{x})|=(h_1+\widetilde{h})^2\,o(1)$.

• Probaremos ahora a). Por (4.10) y usando que $A_0(\mathbf{u})=v(\mathbf{u})A_0(\psi)$ tenemos que

$$\begin{aligned}B_{n,1} &= \sigma\sqrt{nh_1}\int A_0^{-1}(\mathbf{x})\mathbb{E}\widehat{A}_{1,n}(\mathbf{x})q_{\underline{1}}(\mathbf{x}_{\underline{1}})\,d\mathbf{x}_{\underline{1}}=B_{11,n}+B_{12,n}\\ B_{11,n} &= \sqrt{nh_1}\int v^{-1}(\mathbf{x})\left\{\sum_{\ell=1}^d g_\ell'(x_\ell)\,\frac{\partial}{\partial u_\ell}v(\mathbf{u})\Big|_{\mathbf{u}=\mathbf{x}}h_\ell^2\,s_\ell+\frac{1}{2}v(\mathbf{x})\left[\sum_{\ell=1}^d g_\ell''(x_\ell)h_\ell^2\,s_\ell\right]\right\}q_{\underline{1}}(\mathbf{x}_{\underline{1}})\,d\mathbf{x}_{\underline{1}}\\ B_{12,n} &= \sqrt{nh_1}\int v^{-1}(\mathbf{x})\nu_n(\mathbf{x})q_{\underline{1}}(\mathbf{x}_{\underline{1}})\,d\mathbf{x}_{\underline{1}}\end{aligned}$$

Como $\sup_{\mathbf{x}\in\mathcal{S}_Q}|\nu_n(\mathbf{x})|=(h_1+\widetilde{h})^2\,o(1)$, $h_1=\beta n^{-1/5}$ y $\widetilde{h}=o(h_1)$ tenemos que $B_{12,n}\to 0$.

Por otra parte, $B_{11,n}=B_{111,n}+B_{112,n}$ donde

$$\begin{aligned}B_{111,n} &= \sqrt{nh_1}h_1^2\int v^{-1}(\mathbf{x})\left\{g_1'(x_1)\,\frac{\partial}{\partial u_1}v(\mathbf{u})\Big|_{\mathbf{u}=\mathbf{x}}s_1+\frac{1}{2}v(\mathbf{x})g_1''(x_1)\,s_1\right\}q_{\underline{1}}(\mathbf{x}_{\underline{1}})\,d\mathbf{x}_{\underline{1}}\\ &= \sqrt{nh_1}h_1^2s_1\int\left\{v^{-1}(\mathbf{x})g_1'(x_1)\,\frac{\partial}{\partial u_1}v(\mathbf{u})\Big|_{\mathbf{u}=\mathbf{x}}+\frac{1}{2}g_1''(x_1)\right\}q_{\underline{1}}(\mathbf{x}_{\underline{1}})\,d\mathbf{x}_{\underline{1}}\\ B_{112,n} &= \sqrt{nh_1}\widetilde{h}^2\int v^{-1}(\mathbf{x})\left\{\sum_{\ell=2}^d g_\ell'(x_\ell)\,\frac{\partial}{\partial u_\ell}v(\mathbf{u})\Big|_{\mathbf{u}=\mathbf{x}}s_\ell+\frac{1}{2}v(\mathbf{x})\left[\sum_{\ell=2}^d g_\ell''(x_\ell)\,s_\ell\right]\right\}q_{\underline{1}}(\mathbf{x}_{\underline{1}})\,d\mathbf{x}_{\underline{1}}\end{aligned}$$

Como $\widetilde{h} = o(h_1)$, $\sqrt{nh_1}\widetilde{h}^2 \to 0$ de donde $B_{112,n} \to 0$. Por otra parte, como $h_1 = \beta n^{-1/5}$, $\sqrt{nh_1}h_1^2 = \beta^{5/2}$ de donde $B_{111,n} = b_{0,1}(x_1)$.

• Probaremos ahora b).

Llamemos $S_{1,n}(\mathbf{x}) = (1/n)\sum_{i=1}^n \mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i - \mathbf{x})\xi_i\psi(\varepsilon_i)$ y sea $V_i = \psi\left((Y_i - g(\mathbf{x}))/\sigma\right) - \psi((Y_i - g(\mathbf{X}_i))/\sigma)$, tenemos que

$$\widehat{A}_{1,n}(\mathbf{x},\sigma) - S_{1,n}(\mathbf{x},\sigma) = \frac{1}{n}\sum_{i=1}^n \mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i - \mathbf{x})\xi_i V_i\,.$$

Supongamos ahora que hemos probado los dos siguientes resultados

$$\sqrt{nh_1}\sup_{\mathbf{x}\in\mathcal{S}_Q}\left|\frac{1}{n}\sum_{i=1}^n \mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i - \mathbf{x})\xi_i\, V_i - \mathbb{E}\left[\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i - \mathbf{x})\xi_i\, V_i\right]\right| = o_{\mathbb{P}}(1) \tag{4.11}$$

$$\sqrt{nh_1}\sigma\int A_0^{-1}(\mathbf{x})S_{1,n}(\mathbf{x},\sigma)q_{\underline{1}}(\mathbf{x}_{\underline{1}})\,d\mathbf{x}_{\underline{1}} \xrightarrow{D} N(0,\sigma_1^2(x_1))\,. \tag{4.12}$$

Como $\mathbb{E}\left[\psi\left((Y_i - g(\mathbf{X}_i))/\sigma\right)|\mathbf{X}_i\right] = 0$, (4.11) es equivalente a

$$\sqrt{nh_1}\sup_{\mathbf{x}\in\mathcal{S}_Q}\left|\widehat{A}_{1,n}(\mathbf{x},\sigma) - \mathbb{E}\widehat{A}_{1,n}(\mathbf{x},\sigma) - S_{1,n}(\mathbf{x},\sigma)\right| = o_{\mathbb{P}}(1)\,,$$

de donde resulta que

$$\left|\sigma\int A_0^{-1}(\mathbf{x})\left\{\sqrt{nh_1}\left(\widehat{A}_{1,n}(\mathbf{x},\sigma) - \mathbb{E}\widehat{A}_{1,n}(\mathbf{x},\sigma) - S_{1,n}(\mathbf{x},\sigma)\right)\right\}q_{\underline{1}}(\mathbf{x}_{\underline{1}})\,d\mathbf{x}_{\underline{1}}\right| \le \sigma\frac{1}{i(r)i(f_{\mathbf{X}})A_0(\psi)}o_{\mathbb{P}}(1)\,.$$

La desigualdad anterior y (4.12) permiten obtener que

$$\begin{aligned}
\widehat{B}_{n,2} &= \sigma\sqrt{nh_1}\int A_0^{-1}(\mathbf{x})\left[\widehat{A}_{1,n}(\mathbf{x}) - \mathbb{E}(\widehat{A}_{1,n}(\mathbf{x}))\right]q_{\underline{1}}(\mathbf{x}_{\underline{1}})d\mathbf{x}_{\underline{1}}\\
&= \sigma\int A_0^{-1}(\mathbf{x})\left\{\sqrt{nh_1}\left(\widehat{A}_{1,n}(\mathbf{x},\sigma) - \mathbb{E}\widehat{A}_{1,n}(\mathbf{x},\sigma) - S_{1,n}(\mathbf{x},\sigma)\right)\right\}q_{\underline{1}}(\mathbf{x}_{\underline{1}})\,d\mathbf{x}_{\underline{1}}\\
&\quad + \sqrt{nh_1}\sigma\int A_0^{-1}(\mathbf{x})S_{1,n}(\mathbf{x},\sigma)\,q_{\underline{1}}(\mathbf{x}_{\underline{1}})\,d\mathbf{x}_{\underline{1}}\\
&\xrightarrow{D} N(0,\sigma_1^2(x_1))
\end{aligned}$$

dando por probado b).

• Probaremos ahora (4.11). Observemos que $|V_i| \le C_1|g(\mathbf{X}_i) - g(\mathbf{x})|/\sigma$, luego como g es Lipschitz, $h_j = \widetilde{h}$ si $j \ne 1$ y $\widetilde{h} = o(h_1)$ tenemos que $|W_i| \le C_2$ donde $W_i = \xi_i V_i \prod_{j=1}^d \mathbb{I}_{|x_j - X_{ij}|\le h_j}/h_1$. Por otra parte, como K_j tienen soporte compacto en $[-1,1]$ resulta que $\sum_{i=1}^n \mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i - \mathbf{x})\xi_i\, V_i = h_1\sum_{i=1}^n \mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i - \mathbf{x})\,W_i$. Por lo tanto, aplicando el Lema 3.4.5 a $(1/n)\sum_{i=1}^n \mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i - \mathbf{x})\,W_i$ se obtiene que

$$\sup_{\mathbf{x}\in\mathcal{S}_Q}\left|\frac{1}{n}\sum_{i=1}^n \mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i - \mathbf{x})\xi_i\, V_i - \mathbb{E}\left[\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i - \mathbf{x})\xi_i\, V_i\right]\right| = h_1\left(\frac{\log n}{nh_1\widetilde{h}^{d-1}}\right)^{1/2} O_{\mathbb{P}}(1)$$

de donde

$$\sqrt{nh_1}\sup_{\mathbf{x}\in\mathcal{S}_Q}\left|\frac{1}{n}\sum_{i=1}^{n}\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i-\mathbf{x})\xi_i\,V_i-\mathbb{E}\left[\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i-\mathbf{x})\xi_i\,V_i\right]\right|\leq\sqrt{nh_1}h_1\left(\frac{\log n}{nh_1\widetilde{h}^{d-1}}\right)^{1/2}O_{\mathbb{P}}(1)$$

y el resultado se obtiene del hecho que $\log n/(n^{2/5}\widetilde{h}^{d-1})\to 0$.

• Por último, probaremos (4.12). Basta mostrar que

$$\sqrt{nh_1}\int v^{-1}(\mathbf{x})S_{1,n}(\mathbf{x},\sigma)q_{\underline{1}}(\mathbf{x}_{\underline{1}})\,d\mathbf{x}_{\underline{1}}\xrightarrow{D}N(0,\sigma_{11}^2(x_1))$$

donde

$$\sigma_{11}^2(x_1)=\mathbb{E}(\psi^2(\varepsilon))\int K_1^2(u)du\int\frac{1}{f_{\mathbf{X}}(x_1,\mathbf{x}_{\underline{1}})p(x_1,\mathbf{x}_{\underline{1}})}q_{\underline{1}}^2(\mathbf{x}_{\underline{1}})d\mathbf{x}_{\underline{1}}\,.$$

Tenemos que

$$\begin{aligned}\int v^{-1}(\mathbf{x})S_{1,n}(\mathbf{x},\sigma)\,q_{\underline{1}}(\mathbf{x}_{\underline{1}})\,d\mathbf{x}_{\underline{1}} &= \sigma\int v^{-1}(x_1,\mathbf{x}_{\underline{1}})S_{1,n}(x_1,\mathbf{x}_{\underline{1}},\sigma)\,q_{\underline{1}}(\mathbf{x}_{\underline{1}})\,d\mathbf{x}_{\underline{1}}\\ &=\frac{1}{nh_1\widetilde{h}^{d-1}}\sum_{i=1}^{n}\xi_i\psi(\varepsilon_i)K_1\left(\frac{X_{i,1}-x_1}{h_1}\right)\int\prod_{j=2}^{d}K_j\left(\frac{X_{ij}-x_j}{\widetilde{h}}\right)\frac{q_{\underline{1}}(\mathbf{x}_{\underline{1}})}{v(x_1,\mathbf{x}_{\underline{1}})}\,d\mathbf{x}_{\underline{1}}\\ &=\frac{1}{nh_1}\sum_{i=1}^{n}\xi_i\psi(\varepsilon_i)K_1\left(\frac{X_{i,1}-x_1}{h_1}\right)\int\prod_{j=2}^{d}K_j(u_j)\frac{q_{\underline{1}}(\mathbf{X}_{i,\underline{1}}+\mathbf{H}_{d-1}\mathbf{u})}{v(x_1,\mathbf{X}_{i,\underline{1}}+\mathbf{H}_{d-1}\mathbf{u})}\,d\mathbf{u}\\ &=\frac{1}{nh_1}\sum_{i=1}^{n}K_1\left(\frac{X_{i,1}-x_1}{h_1}\right)Z_{i,\mathbf{H}_{d-1}}\end{aligned}$$

donde $Z_{i,\mathbf{H}_{d-1}}=\xi_i\psi(\varepsilon_i)\int\prod_{j=2}^{d}K_j(u_j)\nu(x_1,\mathbf{X}_{i,\underline{1}}+\mathbf{H}_{d-1}\mathbf{u})\,d\mathbf{u}$, $\mathbf{H}_{d-1}=\mathrm{diag}(\widetilde{h},\ldots,\widetilde{h})$ es la submatriz de $\mathbf{H}_d$ de dimensión $(d-1)\times(d-1)$, $\mathbf{X}_{i,\underline{1}}=(X_{i,2},\ldots,X_{i,d})$, $\nu(x_1,\mathbf{u})=q_{\underline{1}}(\mathbf{u})/v(x_1,\mathbf{u})$. Sea $W_{i,n}=K_1\left((X_{i,1}-x_1)/h_1\right)Z_{i,\mathbf{H}_{d-1}}/\sqrt{nh_1}$, luego $\sqrt{nh_1}\int v^{-1}(\mathbf{x})S_{1,n}(\mathbf{x},\sigma)q_{\underline{1}}(\mathbf{x}_{\underline{1}})\,d\mathbf{x}_{\underline{1}}=\sum_{i=1}^{n}W_{i,n}$.

Observemos que por ser t, ψ y $q_{\underline{1}}$ acotadas y como $\inf_{\mathbf{x}\in\mathcal{C}}f_{\mathbf{X}}(\mathbf{x})>0$, $\inf_{\mathbf{x}\in\mathcal{C}}t(\mathbf{x})>0$ y $\inf_{\mathbf{x}\in\mathcal{C}}p(\mathbf{x})>0$, $|Z_{i,\mathbf{H}_{d-1}}|\leq C$ para todo i y $\widetilde{h}$. Por otra parte, la continuidad de $q_{\underline{1}}$ y v implican que $\int\prod_{j=2}^{d}K_j(u_j)\nu(x_1,\mathbf{X}_{i,\underline{1}}+\mathbf{H}_{d-1}\mathbf{u})\,d\mathbf{u}\to\nu(x_1,\mathbf{X}_{i,\underline{1}})$. Luego, como ϵ_1 es independiente de $\mathbf{X}_1$, tenemos que $\mathbb{E}(Z_{1,\mathbf{H}_{d-1}}|\mathbf{X}_1)=0$ de donde $\mathbb{E}W_{i,n}=0$ y

$$\begin{aligned}\mathrm{VAR}(\sum_{i=1}^{n}W_{i,n}) &= n\mathbb{E}(W_{1,n}^2)\to\mathbb{E}(\psi^2(\varepsilon))\int K_1^2(u)du\int\frac{q_{\underline{1}}^2(\mathbf{u})}{p(x_1,\mathbf{u})f_{\mathbf{X}}(x_1,\mathbf{u})}\,d\mathbf{u}=\sigma_{11}^2(x_1)\\ \sum_{i=1}^{n}\mathbb{E}|W_{i,n}|^3 &\leq \frac{n}{(nh_1)^{3/2}}\mathbb{E}\left|K_1\left(\frac{X_{1,1}-x_1}{h_1}\right)Z_{1,\mathbf{H}_{d-1}}\right|^3\\ &\leq C^3\frac{1}{(nh_1)^{1/2}}\frac{1}{h_1}\mathbb{E}\left|K_1^3\left(\frac{X_{1,1}-x_1}{h_1}\right)\right|\leq A\frac{1}{(nh_1)^{1/2}}\,,\end{aligned}$$

pues la densidad f_{X_1} es continua. Por lo tanto, el resultado se obtiene aplicando la versión de Lyapunov del Teorema Central del Límite, lo que concluye la demostración de b).

Falta mostrar que $\sqrt{nh_1}[\widehat{g}_1(x_1) - g_1(x_1)] - \widehat{B}_n \xrightarrow{p} 0$.

Observemos que $\widetilde{g}(\mathbf{x}) - g(\mathbf{x}) - \sigma\, A_0^{-1}(\mathbf{x})\widehat{A}_{1,n}(\mathbf{x}) = \sigma\,(\widehat{A}_{0,n}^{-1}(\mathbf{x}) - A_0^{-1}(\mathbf{x}))\widehat{A}_{1,n}(\mathbf{x}) = \widehat{D}_1(\mathbf{x}) + \widehat{D}_2(\mathbf{x})$ con

$$\begin{aligned}
\widehat{D}_1(\mathbf{x}) &= \sigma\,(\widehat{A}_{0,n}^{-1}(\mathbf{x}) - A_0^{-1}(\mathbf{x}))\mathbb{E}\widehat{A}_{1,n}(\mathbf{x}) \\
\widehat{D}_2(\mathbf{x}) &= \sigma\,(\widehat{A}_{0,n}^{-1}(\mathbf{x}) - A_0^{-1}(\mathbf{x}))\left(\widehat{A}_{1,n}(\mathbf{x}) - \mathbb{E}\widehat{A}_{1,n}(\mathbf{x})\right)
\end{aligned}$$

Probaremos que

$$\sqrt{nh_1}\sup_{\mathbf{x}\in\mathcal{S}_Q}|\widehat{D}_1(\mathbf{x})| \xrightarrow{p} 0 \tag{4.13}$$

$$\sqrt{nh_1}\sup_{\mathbf{x}\in\mathcal{S}_Q}|\widehat{D}_2(\mathbf{x})| \xrightarrow{p} 0 \tag{4.14}$$

ya que entonces

$$\sqrt{nh_1}[\widehat{g}_1(x_1) - g_1(x_1)] - \widehat{B}_n = \sqrt{nh_1}\int\left(\widehat{D}_1(\mathbf{x}) + \widehat{D}_2(\mathbf{x})\right)q_{\underline{1}}(\mathbf{x}_{\underline{1}})d\mathbf{x}_{\underline{1}} \xrightarrow{p} 0$$

lo que concluye la demostracion.

Para probar (4.13), observemos que usando (4.10) y que $\widetilde{h} = o(h_1)$ obtenemos que

$$\sup_{\mathbf{x}\in\mathcal{S}_Q}|\widehat{D}_1(\mathbf{x})| \leq \sigma\sup_{\mathbf{x}\in\mathcal{S}_Q}\left|\widehat{A}_{0,n}^{-1}(\mathbf{x}) - A_0^{-1}(\mathbf{x})\right| O(h_1^2)\,,$$

donde el término $O(h_1^2)$ no depende de $\mathbf{x}$. Por lo tanto, como $\sup_{\mathbf{x}\in\mathcal{S}_Q}\left|\widehat{A}_{0,n}^{-1}(\mathbf{x}) - A_0^{-1}(\mathbf{x})\right| \xrightarrow{p} 0$ ya que $\inf_{\mathbf{x}\in\mathcal{S}_Q} A_0(\mathbf{x}) > 0$ y $\sqrt{nh_1}h_1^2 = \beta^{5/2}$ obtenemos que $\sqrt{nh_1}\sup_{\mathbf{x}\in\mathcal{S}_Q}|\widehat{D}_1(\mathbf{x})| \xrightarrow{p} 0$.

Probaremos ahora (4.14). El Lema 3.4.5 con $\theta_n = \sqrt{\log n/(nh_1\widetilde{h}^{d-1})}$, implica que

$$\sup_{\mathbf{x}\in\mathcal{S}_Q}|\widehat{A}_{1,n}(\mathbf{x}) - \mathbb{E}\widehat{A}_{1,n}(\mathbf{x})| = O_{\mathbb{P}}\left(\left(\frac{\log n}{nh_1\widetilde{h}^{d-1}}\right)^{1/2}\right).$$

Por otra parte, usando (4.10) y que $\widetilde{h} = o(h_1)$ obtenemos que

$$\sup_{\mathbf{x}\in\mathcal{S}_Q}\left|\mathbb{E}\widehat{A}_{1,n}(\mathbf{x})\right| = O(h_1^2)\,.$$

Por lo tanto, como $\sup_{\mathbf{x}\in\mathcal{S}_Q}\left|\widehat{A}_{0,n}^{-1}(\mathbf{x}) - A_0^{-1}(\mathbf{x})\right| \xrightarrow{p} 0$ y $\inf_{\mathbf{x}\in\mathcal{S}_Q}|A_0(\mathbf{x})| > 0$ deducimos que $\sup_{\mathbf{x}\in\mathcal{S}_Q}\left|\widehat{A}_{0,n}^{-1}(\mathbf{x})\right| = O_{\mathbb{P}}(1)$ de donde

$$\sup_{\mathbf{x}\in\mathcal{S}_Q}|\widetilde{g}(\mathbf{x}) - g(\mathbf{x})| \leq O_{\mathbb{P}}(1)\sup_{\mathbf{x}\in\mathcal{S}_Q}\left|\widehat{A}_{1,n}(\mathbf{x})\right| \leq O_{\mathbb{P}}(h_1^2) + O_{\mathbb{P}}\left(\left(\frac{\log n}{nh_1\widetilde{h}^{d-1}}\right)^{1/2}\right). \tag{4.15}$$

Tenemos que

$$\begin{aligned}\sqrt{nh_1}\sup_{\mathbf{x}\in\mathcal{S}_Q}|\widehat{D}_2(\mathbf{x})| &\leq \sqrt{nh_1}\sigma\frac{\sup_{\mathbf{x}\in\mathcal{S}_Q}|\widehat{A}_{0,n}(\mathbf{x})-A_0(\mathbf{x})|}{\inf_{\mathbf{x}\in\mathcal{S}_Q}|A_0(\mathbf{x})|\inf_{\mathbf{x}\in\mathcal{S}_Q}|\widehat{A}_{0,n}(\mathbf{x})|}\sup_{\mathbf{x}\in\mathcal{S}_Q}\left|\widehat{A}_{1,n}(\mathbf{x})-\mathbb{E}\widehat{A}_{1,n}(\mathbf{x})\right| \\ &\leq \sup_{\mathbf{x}\in\mathcal{S}_Q}|\widehat{A}_{0,n}(\mathbf{x})-A_0(\mathbf{x})|O_{\mathbb{P}}\left(\left(\frac{\log n}{\widetilde{h}^{d-1}}\right)^{1/2}\right) \qquad (4.16)\end{aligned}$$

Definamos

$$\widetilde{A}_n(\mathbf{x},a)=\frac{1}{n}\sum_{i=1}^{n}\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i-\mathbf{x})\,\xi_i\psi'\left(\frac{Y_i-g(\mathbf{x})-a}{\sigma}\right)$$

y sea $\lambda_1(t)=\mathbb{E}\psi'(\epsilon_1+t)$. Luego, $\lambda_1(0)=A_0(\psi)$ y

$$\begin{aligned}\lambda_{1,n}(\mathbf{x},a) &= \mathbb{E}\widetilde{A}_n(\mathbf{x},a)=\mathbb{E}\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_1-\mathbf{x})\,r(\mathbf{X}_1)\lambda_1\left(\frac{g(\mathbf{X}_1)-g(\mathbf{x})-a}{\sigma}\right) \\ &= \int\prod_{j=1}^{d}K_j(u_j)r(\mathbf{x}+\mathbf{H}_d\mathbf{u})f_{\mathbf{X}}(\mathbf{x}+\mathbf{H}_d\mathbf{u})\lambda_1\left(\frac{g(\mathbf{x}+\mathbf{H}_d\mathbf{u})-g(\mathbf{x})-a}{\sigma}\right)d\mathbf{u}\,.\end{aligned}$$

Argumentos análogos a los considerados en el Lema 3.4.6 con $\theta_n=\sqrt{\log n/(nh_1\widetilde{h}^{d-1})}$ implican que

$$\sup_{\mathbf{x}\in\mathcal{S}_Q}\sup_{|a|\leq 1}|\widetilde{A}_n(\mathbf{x},a)-\mathbb{E}\widetilde{A}_n(\mathbf{x},a)|=O_{\mathbb{P}}\left(\left(\frac{\log n}{nh_1\widetilde{h}^{d-1}}\right)^{1/2}\right).$$

Luego, como $\widehat{\theta}_1(\mathbf{x})$ es un punto intermedio entre $\widetilde{g}(\mathbf{x})$ y $g(\mathbf{x})$, $\sup_{\mathbf{x}\in\mathcal{S}_Q}|\widetilde{g}(\mathbf{x})-g(\mathbf{x})|=o_{\mathbb{P}}(1)$ (ver 4.15)) tenemos que $\sup_{\mathbf{x}\in\mathcal{S}_Q}|\widehat{a}(\mathbf{x})|=o_{\mathbb{P}}(1)$ donde $\widehat{a}(\mathbf{x})=\widehat{\theta}_1(\mathbf{x})-g(\mathbf{x})$. Usando que $\widehat{A}_{0,n}(\mathbf{x})=\widetilde{A}_n(\mathbf{x},\widehat{a}(\mathbf{x}))$ tenemos que

$$\sup_{\mathbf{x}\in\mathcal{S}_Q}|\widehat{A}_{0,n}(\mathbf{x})-\lambda_{1,n}(\mathbf{x},\widehat{a}(\mathbf{x}))|=O_{\mathbb{P}}\left(\left(\frac{\log n}{nh_1\widetilde{h}^{d-1}}\right)^{1/2}\right). \qquad (4.17)$$

Luego (4.16) y (4.17) y el hecho que $\log n/(n^{2/5}\widetilde{h}^{d-1})\to 0$ implican que

$$\begin{aligned}\sqrt{nh_1}\sup_{\mathbf{x}\in\mathcal{S}_Q}|\widehat{D}_2(\mathbf{x})| &\leq \sup_{\mathbf{x}\in\mathcal{S}_Q}|\widehat{A}_{0,n}(\mathbf{x})-\lambda_{1,n}(\mathbf{x},\widehat{a}(\mathbf{x}))|O_{\mathbb{P}}\left(\left(\frac{\log n}{\widetilde{h}^{d-1}}\right)^{1/2}\right) \\ &\quad+\sup_{\mathbf{x}\in\mathcal{S}_Q}|\lambda_{1,n}(\mathbf{x},\widehat{a}(\mathbf{x}))-A_0(\mathbf{x})|O_{\mathbb{P}}\left(\left(\frac{\log n}{\widetilde{h}^{d-1}}\right)^{1/2}\right) \\ &\leq O_{\mathbb{P}}\left(\frac{1}{\sqrt{nh_1}}\frac{\log n}{\widetilde{h}^{d-1}}\right)+\sup_{\mathbf{x}\in\mathcal{S}_Q}|\lambda_{1,n}(\mathbf{x},\widehat{a}(\mathbf{x}))-A_0(\mathbf{x})|O_{\mathbb{P}}\left(\left(\frac{\log n}{\widetilde{h}^{d-1}}\right)^{1/2}\right) \\ &\leq o_{\mathbb{P}}(1)+\sup_{\mathbf{x}\in\mathcal{S}_Q}|\lambda_{1,n}(\mathbf{x},\widehat{a}(\mathbf{x}))-A_0(\mathbf{x})|O_{\mathbb{P}}\left(\left(\frac{\log n}{\widetilde{h}^{d-1}}\right)^{1/2}\right) \qquad (4.18)\end{aligned}$$

Sea $A_n(\mathbf{x}) = A_0(\psi)\int\prod_{j=1}^d K_j(u_j)r(\mathbf{x}+\mathbf{H}_d\mathbf{u})f_{\mathbf{X}}(\mathbf{x}+\mathbf{H}_d\mathbf{u})\,d\mathbf{u}$. Como $f_{\mathbf{X}}$ y ψ'' son acotadas λ_1 es Lipschitz de donde usando (4.15) obtenemos

$$\begin{aligned}|\lambda_{1,n}(\mathbf{x},\widehat{a}(\mathbf{x})) - \lambda_{1,n}(\mathbf{x},0)| &\leq C_1|\widehat{a}(\mathbf{x})| \leq C_1\,|\widetilde{g}(\mathbf{x}) - g(\mathbf{x})| \\ &\leq O_{\mathbb{P}}(h_1^2) + O_{\mathbb{P}}\left(\left(\frac{\log n}{nh_1\widetilde{h}^{d-1}}\right)^{1/2}\right), \end{aligned} \tag{4.19}$$

Finalmente, como como $f_{\mathbf{X}}$ y r son funciones Lipschitz y acotadas, entonces $v(\mathbf{x}) = f_{\mathbf{X}}(\mathbf{x})r(\mathbf{x})$ también es una función Lipschitz, y por lo tanto como g_j es continuamente diferenciable y $\widetilde{h} = o(h_1)$

$$\begin{aligned}|\lambda_{1,n}(\mathbf{x},0)) - A_n(\mathbf{x})| &\leq C\frac{1}{\sigma}\int\prod_{j=1}^d |K_j(u_j)|r(\mathbf{x}+\mathbf{H}_d\mathbf{u})f_{\mathbf{X}}(\mathbf{x}+\mathbf{H}_d\mathbf{u})\,|g(\mathbf{x}+\mathbf{H}_d\mathbf{u}) - g(\mathbf{x})|\,d\mathbf{u} \\ &\leq C_2\max\{h_1,\widetilde{h}\} \leq C_2h_1 \end{aligned} \tag{4.20}$$

$$\begin{aligned}\sup_{\mathbf{x}\in\mathcal{S}_Q}|A_n(\mathbf{x}) - A_0(\mathbf{x})| &= |A_0(\psi)|\sup_{\mathbf{x}\in\mathcal{S}_Q}\left|\int\mathcal{K}(\mathbf{u})[v(\mathbf{x}+\mathbf{H}_d\mathbf{u}) - v(\mathbf{x})]\,d\mathbf{u}\right| \leq C_3\int|\mathcal{K}(\mathbf{u})|\|\mathbf{H}_d\mathbf{u}\|\,d\mathbf{u} \\ &\leq C_4\max\{h_1,\widetilde{h}\} \leq C_4h_1\,. \end{aligned} \tag{4.21}$$

Las cotas dadas en (4.19) a (4.21) y el hecho que $h_1 = \beta n^{-1/5}$ permiten obtener

$$\begin{aligned}\sup_{\mathbf{x}\in\mathcal{S}_Q}|\lambda_{1,n}(\mathbf{x},\widehat{a}(\mathbf{x})) - A_0(\mathbf{x})|O_{\mathbb{P}}\left(\left(\frac{\log n}{\widetilde{h}^{d-1}}\right)^{1/2}\right) &\leq h_1O_{\mathbb{P}}\left(\left(\frac{\log n}{\widetilde{h}^{d-1}}\right)^{1/2}\right) \\ &+ \left(O_{\mathbb{P}}(h_1^2) + O_{\mathbb{P}}\left(\left(\frac{\log n}{nh_1\widetilde{h}^{d-1}}\right)^{1/2}\right)\right)O_{\mathbb{P}}\left(\left(\frac{\log n}{\widetilde{h}^{d-1}}\right)^{1/2}\right) \\ &\leq O_{\mathbb{P}}\left(\left(\frac{\log n}{n^{2/5}\widetilde{h}^{d-1}}\right)^{1/2}\right) + O_{\mathbb{P}}\left(\frac{1}{\sqrt{nh_1}}\,\frac{\log n}{\widetilde{h}^{d-1}}\right)\end{aligned}$$

lo que junto con (4.18) y el hecho que $\log n/(n^{2/5}\widetilde{h}^{d-1}) \to 0$ concluyen la demostración de (4.14) y del Teorema. □

4.3 Distribución asintótica de los M-estimadores basados en polinomios locales de orden q en la dirección de interés α

En esta sección estudiaremos el comportamiento de los estimadores definidos en la Sección 2.4 como la solución de (2.22). Para ello, recordemos que hemos definido $R_\alpha(X_{i,\alpha}) = g_\alpha(X_{i,\alpha}) - g_\alpha(x_\alpha) - \sum_{j=1}^q g_\alpha^{(j)}(x_\alpha)(X_{i,\alpha} - x_\alpha)^j/j!$, $R(\mathbf{X}_i) = \sum_{j\neq\alpha}\{g_j(X_{ij}) - g_j(x_j)\} + R_\alpha(X_{i,\alpha})$ y

$$\begin{aligned}\breve{\mathbf{x}}_{i,\alpha} &= \left(1, \frac{X_{i,\alpha} - x_\alpha}{h_\alpha}, \frac{(X_{i,\alpha} - x_\alpha)^2}{h_\alpha^2}, \ldots, \frac{(X_{i,\alpha} - x_\alpha)^q}{h_\alpha^q}\right)^{\mathrm{T}}, \\ \mathbf{H}_\alpha &= \mathrm{diag}(1, h_\alpha, h_\alpha^2, \ldots, h_\alpha^q) \in \mathbb{R}^{(q+1)\times(q+1)}\,.\end{aligned}$$

Más aún, sea $\mathbf{V}_\alpha = \left(v_{sm}^{(\alpha)}\right)_{1\leq s,m\leq q+1}$ con $v_{sm}^{(\alpha)} = \int u^{s+m-2}K_\alpha^2(u)du$.

Teorema 4.3.1 *Sea* $\mathbf{x}$ *un punto del interior de* $\mathcal{S}_f$ *tal que se cumple* ***A3*** *en un entorno compacto* $\mathcal{C}$ *de* $\mathbf{x}$. *Supongamos que se cumplen* ***A1***, ***A4***, ***C1′***, ***C2′***, ***C4***, ***C5***, ***C7*** *y* ***C8*** *y que además* $f_{\mathbf{X}}$ *es continua,* ψ'' *es Lipschitz y* $\widehat{s}(\mathbf{x}) \xrightarrow{p} \sigma(\mathbf{x})$. *Sea* $\widehat{\boldsymbol{\beta}}(\mathbf{x})$ *una solución de* $\boldsymbol{\Psi}_{n,\alpha}(\boldsymbol{\beta}, \mathbf{x}, \widehat{s}(\mathbf{x})) = \mathbf{0}_{q+1}$ *tal que* $\mathbf{H}_\alpha\{\widehat{\boldsymbol{\beta}}(\mathbf{x}) - \boldsymbol{\beta}(\mathbf{x})\} \xrightarrow{p} 0$, *donde* $\boldsymbol{\Psi}_{n,\alpha}(\boldsymbol{\beta}, \mathbf{x}, \sigma) = (\Psi_{n,\alpha,0}(\boldsymbol{\beta}, \mathbf{x}, \sigma), \ldots, \Psi_{n,\alpha,q}(\boldsymbol{\beta}, \mathbf{x}, \sigma))^{\mathrm{T}}$ *está definida en (2.22),* $\boldsymbol{\beta}(\mathbf{x}) = (g(\mathbf{x}), g_\alpha^{(1)}(x_\alpha), \ldots, g_\alpha^{(q)}(x_\alpha)/q!)^{\mathrm{T}}$ *y* $g_\alpha^{(1)} = g'_\alpha$. *Luego, si* $n\prod_{j=1}^d h_j\, h_\ell^2 \to \beta_\ell$ *para* $\ell \neq \alpha$ *y* $n\prod_{j=1}^d h_j\, h_\alpha^{2(q+1)} \to \beta_\alpha$, *se tiene que*

$$\sqrt{n\prod_{j=1}^d h_j}\left[\mathbf{H}_\alpha[\widehat{\boldsymbol{\beta}}(\mathbf{x}) - \boldsymbol{\beta}(\mathbf{x}) - \mathbf{b}_{\alpha,q}(\mathbf{x})]\right] \xrightarrow{D} N\left(0, \boldsymbol{\Sigma}_{\alpha,q}(\mathbf{x})\right)$$

donde

$$\mathbf{b}_{\alpha,q}(\mathbf{x}) = \mathbf{H}_\alpha^{-1}(\mathbf{S}^{(\alpha)})^{-1}\mathbf{b}^{(\alpha)}(\mathbf{x}) \qquad \boldsymbol{\Sigma}_{\alpha,q}(\mathbf{x}) = \frac{\mathbb{E}(\psi^2(\varepsilon))}{(\mathbb{E}(\psi'(\varepsilon)))^2}\frac{\sigma^2(\mathbf{x})}{f_{\mathbf{X}}(\mathbf{x})p(\mathbf{x})}(\mathbf{S}^{(\alpha)})^{-1}\mathbf{V}_\alpha(\mathbf{S}^{(\alpha)})^{-1}\,.$$

con $\mathbf{b}^{(\alpha)}(\mathbf{x}) = (b_1^{(\alpha)}(\mathbf{x}), \ldots, b_{q+1}^{(\alpha)}(\mathbf{x}))$ *donde* $b_j^{(\alpha)}(\mathbf{x}) = h_\alpha^{q+1} s_{jq}^{(\alpha)} g_\alpha^{(q+1)}(x_\alpha)/(q+1)!$ *para* $1 \leq j \leq q+1$. *Con lo cual, si* $\mathbf{s}_q = (s_{1q}^{(\alpha)}, \ldots, s_{q+1\,q}^{(\alpha)})$ *tenemos que*

$$\sqrt{n\prod_{j=1}^d h_j}\mathbf{H}_\alpha\left[\widehat{\boldsymbol{\beta}}(\mathbf{x}) - \boldsymbol{\beta}(\mathbf{x})\right] \xrightarrow{D} N\left(\beta_\alpha^{1/2}\frac{g_\alpha^{(q+1)}(x_\alpha)}{(q+1)!}(\mathbf{S}^{(\alpha)})^{-1}\mathbf{s}_q, \boldsymbol{\Sigma}_{\alpha,q}(\mathbf{x})\right)\,.$$

Demostración. Como $R(\mathbf{X}_i) = \sum_{j\neq\alpha}\{g_j(X_{ij}) - g_j(x_j)\} + R_\alpha(X_{i,\alpha})$, por **C1′** tenemos que

$$R(\mathbf{X}_i) = \sum_{j\neq\alpha} g'_j(\widetilde{\theta}_{ij})(X_{i,j} - x_j) + \frac{1}{(q+1)!}g_\alpha^{(q+1)}(\widetilde{\theta}_{i\alpha})(X_{i,\alpha} - x_\alpha)^{q+1}\,, \tag{4.22}$$

donde $\widetilde{\boldsymbol{\theta}}_i = (\widetilde{\theta}_{i1}, \ldots, \widetilde{\theta}_{id})^{\mathrm{T}}$ es un punto intermedio entre $\mathbf{X}_i$ y $\mathbf{x}$.

Llamemos $\widehat{\eta}_i = R(\mathbf{X}_i) - \breve{\mathbf{x}}_{i,\alpha}^{\mathrm{T}}\mathbf{H}_\alpha[\widehat{\boldsymbol{\beta}}(\mathbf{x}) - \boldsymbol{\beta}(\mathbf{x})]$. Luego, las ecuaciones (2.22) pueden reescribirse como

$$\frac{1}{n}\sum_{i=1}^n \psi\left(\widehat{V}_i + \frac{\widehat{\eta}_i}{\widehat{s}(\mathbf{x})}\right)\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i - \mathbf{x})\xi_i\,\breve{\mathbf{x}}_{i,\alpha} = \mathbf{0}_{q+1}\,, \tag{4.23}$$

donde $\widehat{V}_i = \sigma(\mathbf{X}_i)\varepsilon_i/\widehat{s}(\mathbf{x}) = V_i\sigma(\mathbf{x})/\widehat{s}(\mathbf{x})$. Luego haciendo un desarrollo de Taylor de orden dos obtenemos $\mathbf{0}_{q+1} = \mathbf{J}_{n,\alpha}(\mathbf{x}, \widehat{a}_\sigma(\mathbf{x})) + \widehat{\mathbf{I}}_{n,1} + \widehat{\mathbf{I}}_{n,2}\,\mathbf{H}_\alpha\left(\widehat{\boldsymbol{\beta}}(\mathbf{x}) - \boldsymbol{\beta}(\mathbf{x})\right) + \widehat{\mathbf{I}}_{n,3}$ donde $\widehat{a}_\sigma(\mathbf{x}) = \sigma(\mathbf{x})/\widehat{s}(\mathbf{x})$ y

$$\begin{aligned}
\widehat{\mathbf{I}}_{n,1} &= \frac{1}{\widehat{s}(\mathbf{x})}\frac{1}{n}\sum_{i=1}^n \psi'(\widehat{V}_i)\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i - \mathbf{x})\xi_i\, R(\mathbf{X}_i)\,\breve{\mathbf{x}}_{i,\alpha}\,,\\
\widehat{\mathbf{I}}_{n,2} &= -\frac{1}{\widehat{s}(\mathbf{x})}\frac{1}{n}\sum_{i=1}^n \psi'(\widehat{V}_i)\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i - \mathbf{x})\xi_i\,\breve{\mathbf{x}}_{i,\alpha}\breve{\mathbf{x}}_{i,\alpha}^{\mathrm{T}}\,,\\
\widehat{\mathbf{I}}_{n,3} &= \frac{1}{2\,\widehat{s}^2(\mathbf{x})}\frac{1}{n}\sum_{i=1}^n \psi''(\widehat{V}_i + \widehat{\theta}_i)\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i - \mathbf{x})\xi_i\widehat{\eta}_i^2\,\breve{\mathbf{x}}_{i,\alpha}\,,
\end{aligned}$$

con $\min(0, \widehat{\eta}_i/\widehat{s}(\mathbf{x})) \leq \widehat{\theta}_i \leq \max(0, \widehat{\eta}_i/\widehat{s}(\mathbf{x}))$. Por otra parte, $\widehat{\mathbf{I}}_{n,3} = \mathbf{L}_{n,1} - \mathbf{L}_{n,2}\,\mathbf{H}_\alpha\left(\widehat{\boldsymbol{\beta}}(\mathbf{x}) - \boldsymbol{\beta}(\mathbf{x})\right)$ donde

$$\begin{aligned}
\mathbf{L}_{n,1} &= \frac{1}{2\widehat{s}(\mathbf{x})}\frac{1}{n}\sum_{i=1}^{n}\psi''(\widehat{V}_i + \widehat{\theta}_i)\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i - \mathbf{x})\xi_i\, R^2(\mathbf{X}_i)\,\breve{\mathbf{x}}_{i,\alpha}\,,\\
\mathbf{L}_{n,2} &= \frac{1}{2\widehat{s}(\mathbf{x})}\frac{1}{n}\sum_{i=1}^{n}\psi''(\widehat{V}_i + \widehat{\theta}_i)\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i - \mathbf{x})\xi_i\,\left(\widehat{\eta}_i + R(\mathbf{X}_i)\right)\,\breve{\mathbf{x}}_{i,\alpha}\breve{\mathbf{x}}_{i,\alpha}^{\mathrm{T}}\,.
\end{aligned}$$

Por lo tanto, obtenemos

$$-\left(n\prod_{j=1}^{d}h_j\right)^{\frac{1}{2}}\mathbf{J}_{n,\alpha}(\mathbf{x}, \widehat{a}_\sigma(\mathbf{x})) = \left(n\prod_{j=1}^{d}h_j\right)^{\frac{1}{2}}\left(\widehat{\mathbf{I}}_{n,1} + \mathbf{L}_{n,1} + \left(\widehat{\mathbf{I}}_{n,2} - \mathbf{L}_{n,2}\right)\mathbf{H}_\alpha\left(\widehat{\boldsymbol{\beta}}(\mathbf{x}) - \boldsymbol{\beta}(\mathbf{x})\right)\right) \tag{4.24}$$

Definamos $\lambda_1(\mathbf{u}, a) = \mathbb{E}(\psi'(V_1\, a)|\mathbf{X}_1 = \mathbf{u}) = \mathbb{E}(\psi'(\sigma(\mathbf{u})\varepsilon\, a/\sigma(\mathbf{x})))$ y

$$\begin{aligned}
\mathbf{D}_{n,1}(a) &= \frac{1}{\widehat{s}(\mathbf{x})}\frac{1}{n}\sum_{i=1}^{n}\left(\psi'(V_i\, a) - \lambda_1(\mathbf{X}_i, a)\right)\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i - \mathbf{x})\xi_i\, R(\mathbf{X}_i)\,\breve{\mathbf{x}}_{i,\alpha}\,,\\
\mathbf{D}_{n,2}(a) &= -\frac{1}{\widehat{s}(\mathbf{x})}\frac{1}{n}\sum_{i=1}^{n}\left(\psi'(V_i\, a) - \lambda_1(\mathbf{X}_i, a)\right)\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i - \mathbf{x})\xi_i\,\breve{\mathbf{x}}_{i,\alpha}\breve{\mathbf{x}}_{i,\alpha}^{\mathrm{T}}\,.
\end{aligned}$$

Sean además

$$\begin{aligned}
\mathbf{I}_{n,1} &= \frac{1}{\widehat{s}(\mathbf{x})}\frac{1}{n}\sum_{i=1}^{n}\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i - \mathbf{x})\lambda_1(\mathbf{X}_i, \widehat{a}_\sigma(\mathbf{x}))\xi_i\, R(\mathbf{X}_i)\,\breve{\mathbf{x}}_{i,\alpha}\,,\\
\mathbf{I}_{n,2} &= -\frac{1}{\widehat{s}(\mathbf{x})}\frac{1}{n}\sum_{i=1}^{n}\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i - \mathbf{x})\lambda_1(\mathbf{X}_i, \widehat{a}_\sigma(\mathbf{x}))\xi_i\,\breve{\mathbf{x}}_{i,\alpha}\breve{\mathbf{x}}_{i,\alpha}^{\mathrm{T}}\,,
\end{aligned}$$

Luego, $\mathbf{D}_{n,\ell}(\widehat{a}_\sigma(\mathbf{x})) = \widehat{\mathbf{I}}_{n,\ell} - \mathbf{I}_{n,\ell}$.

Queremos ver que $\sqrt{n\prod_{j=1}^{d}h_j}\,(\mathbf{D}_{n,\ell}(\widehat{a}_\sigma(\mathbf{x})) - \mathbf{D}_{n,\ell}(1)) \xrightarrow{p} 0$ para $\ell = 1, 2$. Probaremos esto para cada coordenada j del vector $\mathbf{D}_{n,1}$ y para cada coordenada (j, k) de la matriz $\mathbf{D}_{n,2}$. Para cada $0 \leq m, \tilde{m}, \bar{m} \leq 1$, $1 \leq j, k \leq q+1$ y $1 \leq \alpha \leq d$ enteros definamos

$$\widehat{\mathbf{D}}_n(a) = \frac{1}{n}\sum_{i=1}^{n}(\psi'(V_i a) - \lambda_1(\mathbf{X}_i, a))\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i - \mathbf{x})\xi_i R(\mathbf{X}_i)^{\bar{m}}\breve{x}_{i,j,\alpha}^{m}\breve{x}_{i,k,\alpha}^{\tilde{m}}\,.$$

Observemos que la coordenada j del vector $\mathbf{D}_{n,1}(a)$ es $1/\widehat{s}(\mathbf{x})\widehat{\mathbf{D}}_n(a)$ tomando $\bar{m} = m = 1$ y $\tilde{m} = 0$, mientras que la coordenada (j, k) de la matriz $\mathbf{D}_{n,2}(a)$ es $-1/\widehat{s}(\mathbf{x})\widehat{\mathbf{D}}_n(a)$ con $\bar{m} = 0$ y $m = \tilde{m} = 1$. Por el Lema 4.4.1 tenemos que

$$\widehat{\mathbf{C}}_n(\mathbf{x}) = \sqrt{n\prod_{j=1}^{d}h_j}(\widehat{s}(\mathbf{x})\widehat{\mathbf{D}}_n(\widehat{a}_\sigma(\mathbf{x})) - \sigma(\mathbf{x})\widehat{\mathbf{D}}_n(1)) \xrightarrow{p} 0\,.$$

Luego, como $\widehat{s}(\mathbf{x}) \xrightarrow{p} \sigma(\mathbf{x})$ con $\mathbf{x} \in \mathcal{C}$ y además $\sqrt{n\prod_{j=1}^d}\mathbf{D}_{n,\ell}(1) = O_p(1)$ para $\ell = 1,2$, resulta que, tomando $\bar{m} = m = 1$ y $\tilde{m} = 0$

$$\sqrt{n\prod_{j=1}^d h_j}\left(\mathbf{D}_{n,1_j}(\widehat{a}_\sigma(\mathbf{x})) - \mathbf{D}_{n,1_j}(1)\right) = \frac{1}{\widehat{s}(\mathbf{x})}\left[\widehat{\mathbf{C}}_n(\mathbf{x}) + (\widehat{s}(\mathbf{x}) - \sigma(\mathbf{x}))\sqrt{n\prod_{j=1}^d h_j}\mathbf{D}_{n,1_j}(1)\right] \xrightarrow{p} 0$$

y tomando $\bar{m} = 0$ y $m = \tilde{m} = 1$

$$\sqrt{n\prod_{j=1}^d h_j}\left(\mathbf{D}_{n,2_{j,k}}(\widehat{a}_\sigma(\mathbf{x})) - \mathbf{D}_{n,2_{j,k}}(1)\right) = -\frac{1}{\widehat{s}(\mathbf{x})}\left[\widehat{\mathbf{C}}_n(\mathbf{x}) + (\widehat{s}(\mathbf{x}) - \sigma(\mathbf{x}))\sqrt{n\prod_{j=1}^d h_j}\mathbf{D}_{n,2_{j,k}}(1)\right] \xrightarrow{p} 0\,.$$

Por otra parte, usando que

$$\max_{i:\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i-\mathbf{x})\neq 0}|R(\mathbf{X}_i)| \leq A_g\left(\sum_{j\neq\alpha} h_j + h_\alpha^{q+1}\right), \tag{4.25}$$

donde A_g es una constante que depende solamente de $\|g_j^{(1)}\|_\infty$ para $j \neq \alpha$ y de $\|g_\alpha^{(q+1)}\|_\infty$, se puede ver que $\sqrt{n\prod_{j=1}^d h_j}\mathbf{D}_{n,1}(1) \xrightarrow{p} 0$ ya que $n\prod_{j=1}^d h_j \mathrm{VAR}(\widehat{s}(\mathbf{x})\mathbf{D}_{n,1}(1)) \to 0$. Por lo tanto, tenemos que $\sqrt{n\prod_{j=1}^d h_j}\left(\widehat{\mathbf{I}}_{n,1} - \mathbf{I}_{n,1}\right) \xrightarrow{p} 0$. Por lo tanto,

$$-\left(n\prod_{j=1}^d h_j\right)^{\frac{1}{2}}\mathbf{J}_{n,\alpha}(\mathbf{x},\widehat{a}_\sigma(\mathbf{x})) = \left(n\prod_{j=1}^d h_j\right)^{\frac{1}{2}}\left[\mathbf{I}_{n,1} + \mathbf{L}_{n,1} + \left(\widehat{\mathbf{I}}_{n,2} - \mathbf{L}_{n,2}\right)\mathbf{H}_\alpha\left(\widehat{\boldsymbol{\beta}}(\mathbf{x}) - \boldsymbol{\beta}(\mathbf{x})\right)\right] + o_{\mathbb{P}}(1)$$

Sea $\mathbf{A}_n = \widehat{\mathbf{I}}_{n,2} - \mathbf{L}_{n,2} = \mathbf{I}_{n,2} - \mathbf{L}_{n,2} + \mathbf{D}_{n,2}(\widehat{a}_\sigma(\mathbf{x}))$. Como $\sqrt{n\prod_{j=1}^d h_j}\left(\mathbf{D}_{n,\ell}(\widehat{a}_\sigma(\mathbf{x})) - \mathbf{D}_{n,\ell}(1)\right) \xrightarrow{p} 0$ para $\ell = 1,2$, tenemos en particular que $(\mathbf{D}_{n,2}(\widehat{a}_\sigma(\mathbf{x})) - \mathbf{D}_{n,2}(1)) \xrightarrow{p} 0$. De donde usando que $\mathbb{E}\left[\psi'(V_i\,a) - \lambda_1(\mathbf{X}_i,a)\middle|\mathbf{X}_i\right] = 0$ implica $\mathbf{D}_{n,2}(1) \xrightarrow{p} 0$ se deduce que $\mathbf{D}_{n,2}(\widehat{a}_\sigma(\mathbf{x})) \xrightarrow{p} 0$. Por otra parte, como $\mathbf{H}_\alpha\{\widehat{\boldsymbol{\beta}}(\mathbf{x}) - \boldsymbol{\beta}(\mathbf{x})\} \xrightarrow{p} 0$, g_j', $j \neq \alpha$ y g_α^{q+1} son acotadas, tenemos que $\mathbf{L}_{n,2} \xrightarrow{p} 0$. Por lo tanto, tenemos que $\mathbf{A}_n = \mathbf{I}_{n,2} + o_{\mathbb{P}}(1)$. Observemos que si $\widetilde{\lambda}_1(a) = \mathbb{E}\psi'(a\varepsilon)$ entonces $\lambda_1(\mathbf{u},a) = \widetilde{\lambda}_1(a\sigma(\mathbf{u})/\sigma(\mathbf{x}))$, por lo que

$$\mathbf{I}_{n,2} = -\frac{1}{\widehat{s}(\mathbf{x})}\frac{1}{n}\sum_{i=1}^n \mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i - \mathbf{x})\widetilde{\lambda}_1\left(\frac{\sigma(\mathbf{X}_i)}{\sigma(\mathbf{x})}\widehat{a}_\sigma(\mathbf{x})\right)\xi_i\;\breve{\mathbf{x}}_{i,\alpha}\breve{\mathbf{x}}_{i,\alpha}^{\mathrm{T}}$$

Argumentos análogos a los del Lema 3.4.2, permiten mostrar que $\mathbf{I}_{n,2} - \mathbf{I}_{n,2,1} \xrightarrow{p} 0$ donde

$$\mathbf{I}_{n,2,1} = -\frac{1}{\widehat{s}(\mathbf{x})}\frac{1}{n}\sum_{i=1}^n \mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i - \mathbf{x})\widetilde{\lambda}_1\left(\frac{\sigma(\mathbf{X}_i)}{\sigma(\mathbf{x})}\right)\xi_i\;\breve{\mathbf{x}}_{i,\alpha}\breve{\mathbf{x}}_{i,\alpha}^{\mathrm{T}}\,,$$

de donde se deduce fácilmente que $\mathbf{A}_n \xrightarrow{p} \mathbf{A}$ donde

$$\mathbf{A} = -\mathbb{E}\psi'(\varepsilon)\,\frac{f_{\mathbf{X}}(\mathbf{x})\,r(\mathbf{x})}{\sigma(\mathbf{x})}\,\mathbf{S}^{(\alpha)}\,.$$

Por lo tanto, tenemos que

$$-\mathbf{A}_n^{-1}\left(n\prod_{j=1}^{d}h_j\right)^{\frac{1}{2}}\mathbf{J}_{n,\alpha}(\mathbf{x},\widehat{a}_\sigma(\mathbf{x}))=\left(n\prod_{j=1}^{d}h_j\right)^{\frac{1}{2}}\left\{\mathbf{H}_\alpha\left(\widehat{\boldsymbol{\beta}}(\mathbf{x})-\boldsymbol{\beta}(\mathbf{x})\right)+\mathbf{A}_n^{-1}\left(\mathbf{I}_{n,1}+\mathbf{L}_{n,1}\right)\right\}+o_{\mathbb{P}}(1)\,. \tag{4.26}$$

Hemos visto que $\sqrt{n\prod_{j=1}^{d}h_j}\,(\mathbf{J}_{n,\alpha}(\mathbf{x},\widehat{a}_\sigma(\mathbf{x}))-\mathbf{J}_{n,\alpha}(\mathbf{x},1))\xrightarrow{p}0$ pues $\widehat{s}(\mathbf{x})\xrightarrow{p}\sigma(\mathbf{x})$. Luego, tenemos que $\sqrt{n\prod_{j=1}^{d}h_j}\mathbf{J}_{n,\alpha}(\mathbf{x},\widehat{a}_\sigma(\mathbf{x}))$ tiene la misma distribución asintótica que $\sqrt{n\prod_{j=1}^{d}h_j}\,\mathbf{J}_{n,\alpha}(\mathbf{x},1)$ cuyo comportamiento asintótico se deduce de los resultados sobre la distribución asintótica del estimador de núcleos clásico, o sea, como por **A1** y **C3**, $\mathbb{E}\mathbf{J}_{n,\alpha}(\mathbf{x},a)=\mathbf{0_{q+1}}$, tenemos que

$$\sqrt{n\prod_{j=1}^{d}h_j}\,\mathbf{J}_{n,\alpha}(\mathbf{x},1)\xrightarrow{D}N(\mathbf{0}_{q+1},\boldsymbol{\Sigma}_\alpha(\mathbf{x}))$$

con $\boldsymbol{\Sigma}_\alpha(\mathbf{x})=\mathbb{E}(\psi^2(\varepsilon))\,(p(\mathbf{x})f_{\mathbf{X}}(\mathbf{x})/t^2(\mathbf{x}))\,\mathbf{V}_\alpha=\mathbb{E}(\psi^2(\varepsilon))\,(r(\mathbf{x})f_{\mathbf{X}}(\mathbf{x})/t(\mathbf{x}))\,\mathbf{V}_\alpha$, de donde

$$-\mathbf{A}_n^{-1}\left(n\prod_{j=1}^{d}h_j\right)^{\frac{1}{2}}\mathbf{J}_{n,\alpha}(\mathbf{x},\widehat{a}_\sigma(\mathbf{x}))\xrightarrow{D}N(\mathbf{0}_{q+1},\mathbf{A}^{-1}\,\boldsymbol{\Sigma}_\alpha(\mathbf{x})\mathbf{A}^{-1})$$

y como

$$\mathbf{A}^{-1}\,\boldsymbol{\Sigma}_\alpha(\mathbf{x})\mathbf{A}^{-1}=\frac{\mathbb{E}(\psi^2(\varepsilon))}{(\mathbb{E}\psi'(\varepsilon_1))^2}\frac{\sigma^2(\mathbf{x})}{p(\mathbf{x})\,f_{\mathbf{X}}(\mathbf{x})}(\mathbf{S}^{(\alpha)})^{-1}\mathbf{V}_\alpha(\mathbf{S}^{(\alpha)})^{-1}$$

tenemos que

$$-\mathbf{A}_n^{-1}\left(n\prod_{j=1}^{d}h_j\right)^{\frac{1}{2}}\mathbf{J}_{n,\alpha}(\mathbf{x},\widehat{a}_\sigma(\mathbf{x}))\xrightarrow{D}N(\mathbf{0}_{q+1},\boldsymbol{\Sigma}_{\alpha,q}(\mathbf{x}))\,.$$

Por lo tanto, la demostración quedará terminada si mostramos que

$$\left(n\prod_{j=1}^{d}h_j\right)^{\frac{1}{2}}\{\mathbf{I}_{n,1}+\mathbf{L}_{n,1}+\mathbf{A}\,\mathbf{H}_\alpha\mathbf{b}_{\alpha,q}(\mathbf{x})\}\xrightarrow{p}0\,.$$

Observemos que $\mathbf{L}_{n,1}=(\mathbf{L}_{n,1,1}+\mathbf{L}_{n,1,2}+\mathbf{L}_{n,1,3})/\widehat{s}(\mathbf{x})$ donde

$$\begin{aligned}
\mathbf{L}_{n,1,1} &= \frac{1}{n}\sum_{i=1}^{n}\psi''(V_i)\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i-\mathbf{x})\xi_i\,R^2(\mathbf{X}_i)\,\breve{\mathbf{x}}_{i,\alpha}\\
\mathbf{L}_{n,1,2} &= \frac{1}{n}\sum_{i=1}^{n}\left[\psi''(\widehat{V}_i)-\psi''(V_i)\right]\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i-\mathbf{x})\xi_i\,R^2(\mathbf{X}_i)\,\breve{\mathbf{x}}_{i,\alpha}\\
\mathbf{L}_{n,1,3} &= \frac{1}{n}\sum_{i=1}^{n}\left[\psi''(\widehat{V}_i+\widehat{\theta}_i)-\psi''(\widehat{V}_i)\right]\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i-\mathbf{x})\xi_i\,R^2(\mathbf{X}_i)\,\breve{\mathbf{x}}_{i,\alpha}
\end{aligned}$$

Es fácil ver que como $\mathbb{E}\psi''(V_i)=0$ entonces $\left(n\prod_{j=1}^{d}h_j\right)^{\frac{1}{2}}\mathbf{L}_{n,1,1}\xrightarrow{p}0$ pues se cumple (4.25). Por otra parte, como $\mathbb{E}\psi''(a\,V_i)=0$ para todo a, argumentos análogos a los utilizados Lema 3.4.3

permiten ver que $\left(n\prod_{j=1}^d h_j\right)^{\frac{1}{2}}\mathbf{L}_{n,1,2}\xrightarrow{p}0$. Finalmente, como ψ'' es Lipschitz, $|\check{x}_{i,j,\alpha}|\leq 1$ si $\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i-\mathbf{x})\neq 0$, $\min(0,\widehat{\eta}_i/\widehat{s}(\mathbf{x}))\leq\widehat{\theta}_i\leq\max(0,\widehat{\eta}_i/\widehat{s}(\mathbf{x}))$, con $\widehat{\eta}_i=R(\mathbf{X}_i)-\check{\mathbf{x}}_{i,\alpha}^{\mathrm{T}}\mathbf{H}_\alpha[\widehat{\boldsymbol{\beta}}(\mathbf{x})-\boldsymbol{\beta}(\mathbf{x})]$ y se cumple (4.25), tenemos que

$$\|\mathbf{L}_{n,1,3}\|\leq A_g^2\,C_{\psi''}\frac{1}{i(t)\,\widehat{s}(\mathbf{x})}\,\nu_n(\mathbf{H}_d,\widehat{\boldsymbol{\beta}}(\mathbf{x}))\left(\sum_{j\neq\alpha}h_j+h_\alpha^{(q+1)}\right)^2\frac{1}{n}\sum_{i=1}^n|\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i-\mathbf{x})|$$

donde $\nu_n(\mathbf{H}_d,\widehat{\boldsymbol{\beta}}(\mathbf{x}))=A_g\left(\sum_{j\neq\alpha}h_j+h_\alpha^{q+1}\right)+(q+1)\|\mathbf{H}_\alpha[\widehat{\boldsymbol{\beta}}(\mathbf{x})-\boldsymbol{\beta}(\mathbf{x})]\|$. Luego, como $\sum_{i=1}^n|\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i-\mathbf{x})|/n=O_{\mathbb{P}}(1)$, $n\prod_{j=1}^d h_j\,h_\ell^4\to 0$ para $\ell\neq\alpha$ y $n\prod_{j=1}^d h_j\,h_\alpha^{4(q+1)}\to 0$ (pues $n\prod_{j=1}^d h_j\,h_\ell^2\to\beta_\ell$ para $\ell\neq\alpha$ y $n\prod_{j=1}^d h_j\,h_\alpha^{2(q+1)}\to\beta_\alpha$) y $\nu_n(\mathbf{H}_d,\widehat{\boldsymbol{\beta}}(\mathbf{x}))\xrightarrow{p}0$, tenemos que $\left(n\prod_{j=1}^d h_j\right)^{\frac{1}{2}}\mathbf{L}_{n,1,3}\xrightarrow{p}0$.

Como $\lambda_1(\mathbf{u},1)=\widetilde{\lambda}_1(\sigma(\mathbf{u})/\sigma(\mathbf{x}))$. Para concluir la demostración, bastará mostrar que

a) $\left(n\prod_{j=1}^d h_j\right)^{\frac{1}{2}}\{\mathbf{I}_{n,1}-\mathbf{I}_{n,1,1}\}\xrightarrow{p}0$ donde

$$\mathbf{I}_{n,1,1}=\frac{1}{\sigma(\mathbf{x})}\frac{1}{n}\sum_{i=1}^n\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i-\mathbf{x})\xi_i\,\widetilde{\lambda}_1\left(\frac{\sigma(\mathbf{X}_i)}{\sigma(\mathbf{x})}\right)R(\mathbf{X}_i)\,\check{\mathbf{x}}_{i,\alpha}$$

b) $\mathbf{M}_n=(M_{n,1},\ldots,M_{n,q+1})^{\mathrm{T}}=\left(n\prod_{j=1}^d h_j\right)^{\frac{1}{2}}\{\mathbf{I}_{n,1,1}+\mathbf{A}\,\mathbf{H}_\alpha\mathbf{b}_{\alpha,q}(\mathbf{x})\}\xrightarrow{p}0$.

a) se deduce del Lema 4.4.1.

Por otro lado, como vale (4.25) y ψ' es acotada, tenemos que $\mathrm{VAR}(M_{n,j})=\mathrm{VAR}(\mathbf{I}_{n,1,1,j})\to 0$, para $1\leq j\leq q+1$. Por lo tanto, para probar b) bastará mostrar que $\mathbb{E}\mathbf{M}_n\to\mathbf{0}$. Observemos que

$$\mathbf{m}(\mathbf{x})=\mathbf{A}\,\mathbf{H}_\alpha\mathbf{b}_{\alpha,q}(\mathbf{x})=-\mathbb{E}\psi'(\varepsilon)\,\frac{f_{\mathbf{X}}(\mathbf{x})\,r(\mathbf{x})}{\sigma(\mathbf{x})}\,\mathbf{b}^{(\alpha)}(\mathbf{x})$$

por lo que

$$\mathbf{m}_j(\mathbf{x})=-\mathbb{E}\psi'(\varepsilon)\,\frac{f_{\mathbf{X}}(\mathbf{x})\,r(\mathbf{x})}{\sigma(\mathbf{x})}\frac{h_\alpha^{q+1}}{(q+1)!}s_{jq}^{(\alpha)}g_\alpha^{(q+1)}(x_\alpha)\,,$$

de donde como $n\prod_{\ell=1}^d h_\ell\,h_\alpha^{2(q+1)}\to\beta_\ell$ concluímos que

$$\left(n\prod_{\ell=1}^d h_\ell\right)^{1/2}\mathbf{m}_j(\mathbf{x})\to\mu_j=-\mathbb{E}\psi'(\varepsilon)\,\frac{f_{\mathbf{X}}(\mathbf{x})\,r(\mathbf{x})}{\sigma(\mathbf{x})}\frac{\beta_\alpha^{1/2}}{(q+1)!}s_{jq}^{(\alpha)}g_\alpha^{(q+1)}(x_\alpha)\,.$$

Calculemos ahora $\mathbb{E}(\mathbf{I}_{n,1,1,j})$, $1\leq j\leq q+1$. Observemos que

$$\begin{aligned}\mathbb{E}(\mathbf{I}_{n,1,1,j}) &= \frac{1}{\sigma(\mathbf{x})}\frac{1}{n}\sum_{i=1}^n\mathbb{E}\left(\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i-\mathbf{x})\,\widetilde{\lambda}_1\left(\frac{\sigma(\mathbf{X}_i)}{\sigma(\mathbf{x})}\right)R(\mathbf{X}_i)\,\left(\frac{X_{i,\alpha}-x_\alpha}{h_\alpha}\right)^{j-1}\right)\\ &= \frac{1}{\sigma(\mathbf{x})}\int\mathcal{K}_{\mathbf{H}_d}(\mathbf{u}-\mathbf{x})\widetilde{\lambda}_1\left(\frac{\sigma(\mathbf{u})}{\sigma(\mathbf{x})}\right)R(\mathbf{u})r(\mathbf{u})\left(\frac{u_\alpha-x_\alpha}{h_\alpha}\right)^{j-1}f_{\mathbf{X}}(\mathbf{u})\,d\mathbf{u}\\ &= \frac{1}{\sigma(\mathbf{x})}\int\mathcal{K}(\mathbf{v})\widetilde{\lambda}_1\left(\frac{\sigma(\mathbf{x}+\mathbf{H}_d\mathbf{v})}{\sigma(\mathbf{x})}\right)R(\mathbf{x}+\mathbf{H}_d\mathbf{v})r(\mathbf{x}+\mathbf{H}_d\mathbf{v})v_\alpha^{j-1}f_{\mathbf{X}}(\mathbf{x}+\mathbf{H}_d\mathbf{v})\,d\mathbf{v}\end{aligned}$$

Por (4.22) obtenemos que si $\widetilde{\boldsymbol{\theta}}_i = (\widetilde{\theta}_{i1}, \ldots, \widetilde{\theta}_{id})$ es un punto intermedio entre $\mathbf{x}$ y $\mathbf{x} + \mathbf{H}_d\mathbf{v}$

$$\begin{aligned}\mathbb{E}(\mathbf{I}_{n,1,1,j}) &= \frac{1}{\sigma(\mathbf{x})}\sum_{\ell\neq\alpha} h_\ell \int \mathcal{K}(\mathbf{v})\, v_\alpha^{j-1}\, v_\ell \widetilde{\lambda}_1\left(\frac{\sigma(\mathbf{x}+\mathbf{H}_d\mathbf{v})}{\sigma(\mathbf{x})}\right) g_j'(\widetilde{\theta}_{i\ell})\, r(\mathbf{x}+\mathbf{H}_d\mathbf{v}) f_{\mathbf{X}}(\mathbf{x}+\mathbf{H}_d\mathbf{v})\, d\mathbf{v} \\ &\quad + \frac{1}{(q+1)!} h_\alpha^{q+1} \int \mathcal{K}(\mathbf{v})\, v_\alpha^{q+j}\, \widetilde{\lambda}_1\left(\frac{\sigma(\mathbf{x}+\mathbf{H}_d\mathbf{v})}{\sigma(\mathbf{x})}\right) g_\alpha^{(q+1)}(\widetilde{\theta}_{i\alpha}) r(\mathbf{x}+\mathbf{H}_d\mathbf{v}) f_{\mathbf{X}}(\mathbf{x}+\mathbf{H}_d\mathbf{v})\, d\mathbf{v}\,.\end{aligned}$$

Por lo tanto, usando que g_j' $(j \neq \alpha)$, $g_\alpha^{(q+1)}$, $\widetilde{\lambda}_1$, r y $f_{\mathbf{X}}$ son continuas, $n\prod_{\ell=1}^d h_\ell\, h_j^2 \to \beta_j$ para $j \neq \alpha$, $n\prod_{\ell=1}^d h_\ell\, h_\alpha^{2(q+1)} \to \beta_\alpha$ y $\int K_j(u) u du = 0$, el Teorema de Convergencia Mayorada, implica que

$$\left(n\prod_{\ell=1}^d h_\ell\right)^{\frac{1}{2}} \mathbb{E}(\mathbf{I}_{n,1,1,j}) \to \frac{1}{\sigma(\mathbf{x})}\beta_\alpha^{\frac{1}{2}}\mathbb{E}\psi'(\varepsilon) r(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x}) \frac{1}{(q+1)!} g_\alpha^{(q+1)}(x_\alpha)\int K_\alpha(v) v^{q+j}\, dv\,. \tag{4.27}$$

Finalmente, como $\int K_\alpha(v) v^{q+j}\, dv = s_{jq}^{(\alpha)}$ el miembro derecho de (4.27) es igual a $-\mu_j$ lo que concluye la demostración de b). □

Corolario 4.3.1 *Sea* $\mathbf{x}$ *un punto del interior de* $\mathcal{S}_f$ *tal que se cumple* ***A3*** *en un entorno compacto* $\mathcal{C}$ *de* $\mathbf{x}$. *Supongamos que se cumplen* ***A1***, ***A4***, ***C1′***, ***C2′***, ***C4***, ***C5***, ***C7*** *y* ***C8*** *y que además* $f_{\mathbf{X}}$ *es continua,* ψ'' *es Lipschitz y* $\widehat{s}(\mathbf{x}) \xrightarrow{p} \sigma(\mathbf{x})$. *Sea* $\widehat{\boldsymbol{\beta}}(\mathbf{x})$ *una solución de* $\boldsymbol{\Psi}_{n,\alpha}(\boldsymbol{\beta}, \mathbf{x}, \widehat{s}(\mathbf{x})) = \mathbf{0}_{q+1}$ *tal que* $\mathbf{H}_\alpha\{\widehat{\boldsymbol{\beta}}(\mathbf{x}) - \boldsymbol{\beta}(\mathbf{x})\} \xrightarrow{p} 0$, *donde* $\boldsymbol{\Psi}_{n,\alpha}(\boldsymbol{\beta}, \mathbf{x}, \sigma) = (\Psi_{n,\alpha,0}(\boldsymbol{\beta}, \mathbf{x}, \sigma), \ldots, \Psi_{n,\alpha,q}(\boldsymbol{\beta}, \mathbf{x}, \sigma))^{\mathrm{T}}$ *está definida en (2.22),* $\boldsymbol{\beta}(\mathbf{x}) = (g(\mathbf{x}), g_\alpha^{(1)}(x_\alpha), \ldots, g_\alpha^{(q)}(x_\alpha)/q!)^{\mathrm{T}}$ *y* $g_\alpha^{(1)} = g_\alpha'$. *Luego, si* $n\prod_{j=1}^d h_j\, h_\ell^2 \to \beta_\ell$ *para* $\ell \neq \alpha$ *y* $n\prod_{j=1}^d h_j\, h_\alpha^{2(q+1)} \to \beta_\alpha$, *se tiene que*

$$\sqrt{n\prod_{j=1}^d h_j}\left(\widetilde{g}_{\mathrm{M}_{q,\alpha}}(\mathbf{x}) - g(\mathbf{x})\right) \xrightarrow{D} N\left(\beta_\alpha^{1/2}\frac{g_\alpha^{(q+1)}(x_\alpha)}{(q+1)!}u_1, \sigma_{\alpha,q}(\mathbf{x})\right)$$

donde u_1 *es la primer coordenada del vector* $\mathbf{S}^{(\alpha)})^{-1}\mathbf{s}_q$ *con* $\mathbf{s}_q = (s_{1q}^{(\alpha)}, \ldots, s_{q+1\,q}^{(\alpha)})$ $\sigma_{\alpha,q}(\mathbf{x})$ *es el elemento* $(1,1)$ *de la matriz* $\boldsymbol{\Sigma}_{\alpha,q}(\mathbf{x})$, *o sea,* $\mathbf{e}_1^{\mathrm{T}}\boldsymbol{\Sigma}_{\alpha,q}(\mathbf{x})\mathbf{e}_1$.

Para obtener la distribución asintótica de las componentes aditivas del M-estimador local basado en polinomios locales de orden q en la dirección de interés α, como en el Teorema 4.2.3, supondremos un modelo homocedástico, $\sigma(\mathbf{x}) \equiv \sigma$, y que tenemos un estimador robusto de escala $\widehat{s}$ de σ, con tasa $n^{1/2}$. Por otra parte, en lugar de los supuestos **C1′** y **C2′** usaremos las hipótesis **N1** y **N2**, respectivamente, que se dan a continuación

N1 Para todo $j = 1, \ldots, d$ y $j \neq \alpha$, la componente marginal g_j es ℓ veces continuamente diferenciable en un entorno del soporte $\mathcal{S}_j$ de la densidad de X_j y $g_j^{(\ell)}$ es acotada. Además, g_α es continuamente diferenciable hasta el orden $q+1$ y la derivada $q+1$, $g_\alpha^{(q+1)}$, es acotada en $\mathcal{S}_\alpha$.

N2 a) La función de núcleos $\mathcal{K} : \mathbb{R}^d \to \mathbb{R}$ es tal que $\mathcal{K}(\mathbf{x}) = \prod_{j=1}^d K_j(x_j)$ donde $K_j : \mathbb{R} \to \mathbb{R}$ son funciones pares, acotadas con soporte compacto, digamos $[-1, 1]$, Lipschitz continuas y tales que $\int K_j(u)\, du = 1$.

b) El núcleo K_α cumple que $0 < \int |u|^j K_\alpha(u)\,du < \infty$ para $1 \le j \le 2q+1$ y $\int u^j K_\alpha(u)\,du > 0$ si j es par. Más aún, la matriz $\mathbf{S}^{(\alpha)} = \left(\int u_\alpha^{i+j} K_\alpha(u)\,du\right)_{0\le i,j\le q}$ es definida positiva.

c) Para todo $1 \le j \le d$ y $j \ne \alpha$, $K_j = L$ con L una función de núcleos de orden ℓ, es decir, $\int L(u)du = 1$, $\int u^j L(u)\,du = 0$ si $1 \le j \le \ell - 1$ y $\int u^\ell L(u)\,du \ne 0$.

Por simplicidad de notación, indicaremos por $(x_\alpha, \mathbf{x}_{\underline{\alpha}})$ al vector que tiene la componente α igual a x_α y todas las demás iguales a x_j, $j \ne \alpha$. Sea además $\lambda(a) = \mathbb{E}\psi(\varepsilon_1 + a)$ y $\lambda_1(a) = \mathbb{E}\psi'(\varepsilon_1 + a)$. Dada una matriz simétrica $\mathbf{A} \in \mathbb{R}^{m\times m}$ indicaremos por $\nu_1(\mathbf{A}) \le \cdots \le \nu_m(\mathbf{A})$ los autovalores de $\mathbf{A}$.

Teorema 4.3.2. *Sea* $\mathbf{x}$ *un punto del interior de* $\mathcal{S}_f$. *Supongamos que* $\sigma(\mathbf{x}) \equiv \sigma$ *y que* $\widehat{s}$ *es un estimador consistente de* σ *tal que* $\sqrt{n}(\widehat{s} - \sigma) \xrightarrow{p} 0$. *Sea* $\widehat{\boldsymbol{\beta}}(\mathbf{x})$ *una solución de (2.22) tal que* $\sup_{\mathbf{x}\in\mathcal{S}_Q} \|\mathbf{H}_\alpha[\widehat{\boldsymbol{\beta}}(\mathbf{x}) - \boldsymbol{\beta}(\mathbf{x})]\| \xrightarrow{p} 0$ *donde* $\boldsymbol{\beta}(\mathbf{x}) = (g(\mathbf{x}), g_\alpha^{(1)}(x_\alpha), \ldots, g_\alpha^{(q)}(x_\alpha)/q!)^{\mathrm{T}}$ *y* $g_\alpha^{(1)} = g'_\alpha$. *Supongamos que se cumplen* ***A0***, ***A1***, ***C4***, ***C5***, ***C7***, ***N1*** *y* ***N2*** *y que la función* $\lambda(a)$ *tiene derivadas hasta el orden* $\ell - 1$ *Lipschitz en un entorno de* 0. *Más aún, supongamos que* $q_{\underline{\alpha}}(\mathbf{u})$ *es continua,* $f_{\mathbf{X}}(\mathbf{u})$ *y* $r(\mathbf{u}) = p(\mathbf{u})/t(\mathbf{u})$ *son* ℓ *veces continuamente diferenciables y que* $\sup_{\mathbf{x}\in\mathcal{S}_Q} |g(\mathbf{x})| < \infty$ *y* $\sup_{\mathbf{x}\in\mathcal{S}_Q} f_{\mathbf{X}}(\mathbf{x}) < \infty$, $\inf_{\mathbf{x}\in\mathcal{C}} f_{\mathbf{X}}(\mathbf{x}) > 0$, $\inf_{\mathbf{x}\in\mathcal{C}} t(\mathbf{x}) > 0$ *y* $\inf_{\mathbf{x}\in\mathcal{C}} p(\mathbf{x}) > 0$ *para algún entorno compacto,* $\mathcal{C} \subset \mathcal{S}_f$, *de* $\mathcal{S}_Q$. *Si además,* $h_\alpha = \beta n^{-1/(2q+3)}$, $h_j = \widetilde{h}\ \forall j \ne \alpha$, *y tal que* $\widetilde{h}$ *satisface* $\widetilde{h} = o\left(n^{-(q+1)/(\ell(2q+3))}\right)$ *y* $n^{(q+1)/(2q+3)}\widetilde{h}^{d-1}/\log n \to \infty$ *se tiene que*

$$\sqrt{nh_\alpha}[\widehat{g}_{\alpha,\mathrm{M}_{q,\alpha}}(x_\alpha) - g_\alpha(x_\alpha)] \xrightarrow{D} N\left(b_{q,\alpha}(x_\alpha), \sigma^2_{q,\alpha}(x_\alpha)\right)$$

donde

$$\begin{aligned} b_{q,\alpha}(x_\alpha) &= \beta^{\frac{2q+3}{2}} \frac{1}{(q+1)!} g_\alpha^{(q+1)}(x_\alpha)\mathbf{e}_1^{\mathrm{T}} \left(\mathbf{S}^{(\alpha)}\right)^{-1} \mathbf{s}_q^{(\alpha)} \\ \sigma^2_{q,\alpha}(x_\alpha) &= \frac{\mathbb{E}(\psi^2(\varepsilon))}{[\mathbb{E}\psi'(\varepsilon)]^2}\sigma^2 \int \frac{q^2_{\underline{\alpha}}(\mathbf{x}_{\underline{\alpha}})}{f_{\mathbf{X}}(x_\alpha, \mathbf{x}_{\underline{\alpha}})p(x_\alpha, \mathbf{x}_{\underline{\alpha}})}\, d\mathbf{x}_{\underline{\alpha}} \mathbf{e}_1^{\mathrm{T}} (\mathbf{S}^{(\alpha)})^{-1}\mathbf{V}_\alpha(\mathbf{S}^{(\alpha)})^{-1}\mathbf{e}_1 \end{aligned}$$

con $\mathbf{s}_q^{(\alpha)} = (s_{q,1}^{(\alpha)}, \ldots, s_{q,q+1}^{(\alpha)})^{\mathrm{T}}$ *con* $s_{q,j}^{(\alpha)} = \int K_\alpha(t)t^{q+j}\,dt$ *para* $j = 1, \ldots, q+1$, $\mathbf{V}_\alpha = \left(v_{sm}^{(\alpha)}\right)_{1\le s,m\le q+1}$ *con* $v_{sm}^{(\alpha)} = \int u^{s+m-2}K_\alpha^2(u)du$.

Observación 4.3.1. Observemos que como el núcleo L es par su orden ℓ es un número par.

Supongamos que el parámetro de suavizado en las direcciones que no son de interés sea tal que $\widetilde{h} = \gamma n^{-\tau}$. Luego, la hipótesis $\widetilde{h} = o\left(n^{-(q+1)/(\ell(2q+3))}\right)$ se cumple si y sólo si $\tau > (q+1)/(\ell(2q+3))$. Por otro parte, la condición $n^{(q+1)/(2q+3)}\widetilde{h}^{d-1}/\log n \to \infty$ se cumplirá si $\tau < (q+1)/((2q+3)(d-1))$. Por lo tanto, el orden de la ventana $\widetilde{h}$ deberá ser tal que

$$\frac{q+1}{\ell(2q+3)} < \tau < \frac{q+1}{(2q+3)(d-1)}\,.$$

Para que esta condición se verifique el orden ℓ del núcleo en la dirección de no interés debe ser mayor o igual que la dimensión, es decir, $\ell \ge d$.

Demostración. La demostración sigue argumentos análogos a los utilizados para probar el Teorema 4.2.3.

Recordemos que $\widehat{\boldsymbol{\beta}}(\mathbf{x})$ es solución de (2.22), luego $\boldsymbol{\Psi}^{\star}_{n,\alpha}(\boldsymbol{\beta},\mathbf{x},\widehat{s}) = \mathbf{0}_{q+1}$, donde $\boldsymbol{\Psi}^{\star}_{n,\alpha}(\boldsymbol{\beta},\mathbf{x},s) = \big(\Psi^{\star}_{n,\alpha,0}(\boldsymbol{\beta},\mathbf{x},s),\ldots,\Psi^{\star}_{n,\alpha,q}(\boldsymbol{\beta},\mathbf{x},s)\big)$ con

$$\begin{aligned}\boldsymbol{\Psi}^{\star}_{n,\alpha}(\mathbf{b},\mathbf{x},s) &= \frac{1}{n}\sum_{i=1}^{n}\psi\left(\frac{Y_i - b_0 - \sum_{m=1}^{q} b_m(X_{i,\alpha}-x_\alpha)^m}{s}\right)\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i-\mathbf{x})\xi_i\breve{\mathbf{x}}_{i,\alpha}\\ &= \frac{1}{n}\sum_{i=1}^{n}\psi\left(\frac{Y_i - \breve{\mathbf{x}}^{\mathrm{T}}_{i,\alpha}\mathbf{H}_d\mathbf{b}}{s}\right)\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i-\mathbf{x})\xi_i\breve{\mathbf{x}}_{i,\alpha}\,.\end{aligned}$$

Usando que ψ es Lipschitz y que $\zeta(u) = u\psi'(u)$ es acotada, es fácil ver que para $j = 0,\ldots,q$, $\widehat{D}_{n,j} = \sup_{\mathbf{x}\in\mathcal{S}_Q}\sup_{\mathbf{b}}|\widehat{s}\Psi^{\star}_{n,\alpha,j}(\mathbf{b},\mathbf{x},\widehat{s}) - \sigma\Psi^{\star}_{n,\alpha,j}(\mathbf{b},\mathbf{x},\sigma)| = O_{\mathbb{P}}(1/\sqrt{n})$. Por lo tanto, $\boldsymbol{\Psi}^{\star}_{n,\alpha}(\widehat{\boldsymbol{\beta}}(\mathbf{x}),\mathbf{x},\sigma) = O_{\mathbb{P}}(1/\sqrt{n})$, para todo $j = 1,\ldots,q+1$.

Definamos $\widetilde{\boldsymbol{\beta}}(\mathbf{x})$ como la solución de $\boldsymbol{\Psi}^{\star}_{n,\alpha}(\mathbf{b},\mathbf{x},\sigma) = \mathbf{0}_{q+1}$ (es decir, cuando suponemos que la escala es conocida). Luego, por la Proposición 3.3.2, $\sup_{\mathbf{x}\in\mathcal{C}}\|\mathbf{H}_\alpha[\widetilde{\boldsymbol{\beta}}(\mathbf{x}) - \boldsymbol{\beta}\mathbf{x})]\| = o_{\mathbb{P}}(1)$. Luego, para $1 \le j \le q+1$ tenemos que

$$\widehat{D}_n = \sup_{\mathbf{x}\in\mathcal{S}_Q}\left\|\mathbf{H}_\alpha[\widehat{\beta}(\mathbf{x}) - \widetilde{\beta}(\mathbf{x})]\right\| = O_{\mathbb{P}}(1/\sqrt{n})\,. \tag{4.28}$$

Efectivamente, para probar (4.28) observemos que un desarrollo de Taylor de orden uno y el hecho que $\boldsymbol{\Psi}^{\star}_{n,\alpha}(\widetilde{\boldsymbol{\beta}}(\mathbf{x}),\mathbf{x},\sigma) = \mathbf{0}$ permite obtener $\boldsymbol{\Psi}^{\star}_{n,\alpha}(\widehat{\boldsymbol{\beta}}(\mathbf{x}),\mathbf{x},\sigma) = \boldsymbol{\Psi}^{\star}_{n,\alpha}(\widetilde{\boldsymbol{\beta}}(\mathbf{x}),\mathbf{x},\sigma) + \widehat{\mathbf{D}}_{1,n}(\mathbf{x})\mathbf{H}_\alpha(\widehat{\boldsymbol{\beta}}(\mathbf{x}) - \widetilde{\boldsymbol{\beta}}(\mathbf{x})) = \widehat{\mathbf{D}}_{1,n}(\mathbf{x})\mathbf{H}_\alpha(\widehat{\boldsymbol{\beta}}(\mathbf{x}) - \widetilde{\boldsymbol{\beta}}(\mathbf{x}))$ donde

$$\widehat{\mathbf{D}}_{1,n}(\mathbf{x}) = -\frac{1}{\sigma}\frac{1}{n}\sum_{i=1}^{n}\psi'\left(\frac{Y_i - \breve{\mathbf{x}}^{\mathrm{T}}_{i,\alpha}\mathbf{H}_\alpha\boldsymbol{\xi}_n}{\sigma}\right)\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i-\mathbf{x})\xi_i\breve{\mathbf{x}}_{i,\alpha}\breve{\mathbf{x}}^{\mathrm{T}}_{i,\alpha}$$

con $\boldsymbol{\xi}_n = \boldsymbol{\xi}_n(\mathbf{x})$ un punto intermedio entre $\widehat{\boldsymbol{\beta}}(\mathbf{x})$ y $\widetilde{\boldsymbol{\beta}}(\mathbf{x})$. Observemos que como $\sup_{\mathbf{x}\in\mathcal{S}_Q}\|\mathbf{H}_\alpha[\widehat{\beta}(\mathbf{x}) - \beta(\mathbf{x})]\| = o_{\mathbb{P}}(1)$ y $\sup_{\mathbf{x}\in\mathcal{S}_Q}\|\mathbf{H}_\alpha[\widetilde{\beta}(\mathbf{x}) - \beta(\mathbf{x})]\| = o_{\mathbb{P}}(1)$, entonces, $\sup_{\mathbf{x}\in\mathcal{S}_Q}\|\mathbf{H}_\alpha[\boldsymbol{\xi}_n(\mathbf{x}) - \beta(\mathbf{x})]\| = o_{\mathbb{P}}(1)$. Sea

$$\mathbf{D}_0(\mathbf{x}) = -\frac{1}{\sigma}A_0(\psi)r(\mathbf{x})f_{\mathbf{X}}(\mathbf{x})\mathbf{S}^{(\alpha)}\,.$$

Luego, como $\mathbf{S}^{(\alpha)}$ es no singular, $\inf_{\mathbf{x}\in\mathcal{C}} f_{\mathbf{X}}(\mathbf{x}) > 0$, $\inf_{\mathbf{x}\in\mathcal{C}} r(\mathbf{x}) > 0$ y $A_0(\psi) \neq 0$ obtenemos que $\inf_{\mathbf{x}\in\mathcal{C}}\lambda_1(\mathbf{D}_0(\mathbf{x})) > 0$. Por lo tanto, como $\boldsymbol{\Psi}^{\star}_{n,\alpha}(\widehat{\boldsymbol{\beta}}(\mathbf{x}),\mathbf{x},\sigma) = O_{\mathbb{P}}(1/\sqrt{n})$ y $\boldsymbol{\Psi}^{\star}_{n,\alpha}(\widehat{\boldsymbol{\beta}}(\mathbf{x}),\mathbf{x},\sigma) = \widehat{\mathbf{D}}_{1,n}(\mathbf{x})\mathbf{H}_\alpha(\widehat{\boldsymbol{\beta}}(\mathbf{x}) - \widetilde{\boldsymbol{\beta}}(\mathbf{x}))$ para obtener (4.28) bastará mostrar que

$$\sup_{\mathbf{x}\in\mathcal{S}_Q}\|\widehat{\mathbf{D}}_{1,n}(\mathbf{x}) - \mathbf{D}_0(\mathbf{x})\| = o_{\mathbb{P}}(1)\,. \tag{4.29}$$

Como en la demostración del Teorema 4.2.3, tenemos que $\sigma\widehat{\mathbf{D}}_{1,n}(\mathbf{x}) = \widehat{\mathbf{D}}_{11,n}(\mathbf{x}) + \widehat{\mathbf{D}}_{12,n}(\mathbf{x}) + \widehat{\mathbf{D}}_{13,n}(\mathbf{x})$

donde

$$\begin{aligned}
\widehat{\mathbf{D}}_{11,n}(\mathbf{x}) &= \frac{1}{n}\sum_{i=1}^{n}\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i-\mathbf{x})\xi_i\left[\psi'\left(\frac{Y_i-\breve{\mathbf{x}}_{i,\alpha}^{\mathrm{T}}\mathbf{H}_\alpha\boldsymbol{\beta}(\mathbf{x})}{\sigma}\right)-\psi'\left(\frac{Y_i-\breve{\mathbf{x}}_{i,\alpha}^{\mathrm{T}}\mathbf{H}_\alpha\boldsymbol{\xi}_n}{\sigma}\right)\right]\breve{\mathbf{x}}_{i,\alpha}\breve{\mathbf{x}}_{i,\alpha}^{\mathrm{T}}\\
\widehat{\mathbf{D}}_{12,n}(\mathbf{x}) &= \frac{1}{n}\sum_{i=1}^{n}\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i-\mathbf{x})\xi_i\left[\psi'\left(\frac{Y_i-\breve{\mathbf{x}}_{i,\alpha}^{\mathrm{T}}\mathbf{H}_\alpha\boldsymbol{\beta}(\mathbf{X}_i)}{\sigma}\right)-\psi'\left(\frac{Y_i-\breve{\mathbf{x}}_{i,\alpha}^{\mathrm{T}}\mathbf{H}_\alpha\mathbf{e}(\mathbf{x})}{\sigma}\right)\right]\breve{\mathbf{x}}_{i,\alpha}\breve{\mathbf{x}}_{i,\alpha}^{\mathrm{T}}\\
\widehat{\mathbf{D}}_{13,n}(\mathbf{x}) &= -\frac{1}{n}\sum_{i=1}^{n}\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i-\mathbf{x})\xi_i\psi'\left(\frac{Y_i-\breve{\mathbf{x}}_{i,\alpha}^{\mathrm{T}}\mathbf{H}_\alpha\boldsymbol{\beta}(\mathbf{X}_i)}{\sigma}\right)\breve{\mathbf{x}}_{i,\alpha}\breve{\mathbf{x}}_{i,\alpha}^{\mathrm{T}}\,.
\end{aligned}$$

Usando que ψ' es Lipschitz, $\sup_{\mathbf{x}\in\mathcal{S}_Q}\|\mathbf{H}_\alpha[\boldsymbol{\xi}_n(\mathbf{x})-\beta(\mathbf{x})]\|=o_{\mathbb{P}}(1)$, $\sup_{\mathbf{x}\in\mathcal{C}}\sum_{i=1}^{n}|\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i-\mathbf{x})|/n=O_{\mathbb{P}}(1)$, $\sup_{|\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i-\mathbf{x})|\neq 0}|\breve{\mathbf{x}}_{i,\alpha}|\leq 1$ y que $\inf_{\mathbf{x}\in\mathcal{C}}t(\mathbf{x})>0$ deducimos que, para $1\leq j,m\leq q+1$, $\sup_{\mathbf{x}\in\mathcal{S}_Q}|\widehat{D}_{11,n,j,m}(\mathbf{x})|=o_{\mathbb{P}}(1)$, donde $\widehat{D}_{11,n,j,m}(\mathbf{x})$ es el elemento (j,m) de la matriz $\widehat{\mathbf{D}}_{11,n}(\mathbf{x})$.

Por otra parte, como $\sup_{|\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i-\mathbf{x})|\neq 0}|\breve{\mathbf{x}}_{i,\alpha}^{\mathrm{T}}\mathbf{H}_\alpha(\boldsymbol{\beta}(\mathbf{X}_i)-\boldsymbol{\beta}(\mathbf{x}))|\leq C(\widetilde{h}+h_\alpha^{q+1})$, para $1\leq j,m\leq q+1$ tenemos que $\sup_{\mathbf{x}\in\mathcal{S}_Q}|\widehat{D}_{12,n,j,m}(\mathbf{x})|=o_{\mathbb{P}}(1)$. Por último, usando el Lema 3.4.6, para $1\leq j,m\leq q+1$ resulta que $\sup_{\mathbf{x}\in\mathcal{S}_Q}|\widehat{D}_{13,n,j,m}(\mathbf{x})-\mathbb{E}\widehat{D}_{13,n,j,m}(\mathbf{x})|=o_{\mathbb{P}}(1)$, de donde como $\sup_{\mathbf{x}\in\mathcal{S}_Q}|\mathbb{E}\widehat{D}_{13,n,j,m}(\mathbf{x})-D_{0,j,m}(\mathbf{x})|=o_{\mathbb{P}}(1)$ concluímos la demostración de (4.29).

Observemos que como

$$\widehat{g}_{\alpha,\mathrm{M}_{q,\alpha}}(x_\alpha)=\int\mathbf{e}_1^{\mathrm{T}}\,\widehat{\boldsymbol{\beta}}(x_\alpha,\mathbf{u}_{\underline{\alpha}})q_{\underline{\alpha}}(\mathbf{u}_{\underline{\alpha}})\,d\mathbf{u}_{\underline{\alpha}}=\int\mathbf{e}_1^{\mathrm{T}}\,\mathbf{H}_\alpha\widehat{\boldsymbol{\beta}}(x_\alpha,\mathbf{u}_{\underline{\alpha}})q_{\underline{\alpha}}(\mathbf{u}_{\underline{\alpha}})\,d\mathbf{u}_{\underline{\alpha}}$$

tenemos que

$$\begin{aligned}
&\left|\sqrt{nh_\alpha}(\widehat{g}_{\alpha,\mathrm{M}_{q,\alpha}}(x_\alpha)-g_\alpha(x_\alpha))-\sqrt{nh_\alpha}\int\mathbf{e}_1^{\mathrm{T}}\mathbf{H}_\alpha[\widetilde{\boldsymbol{\beta}}(\mathbf{x})-\boldsymbol{\beta}(\mathbf{x})]q_{\underline{\alpha}}(\mathbf{x}_{\underline{\alpha}})\,d\mathbf{x}_{\underline{\alpha}}\right|\\
&=\left|\sqrt{nh_\alpha}\int\mathbf{e}_1^{\mathrm{T}}\mathbf{H}_\alpha[\widehat{\boldsymbol{\beta}}(\mathbf{x})-\widetilde{\boldsymbol{\beta}}(\mathbf{x})]q_{\underline{\alpha}}(\mathbf{x}_{\underline{\alpha}})\,d\mathbf{x}_{\underline{\alpha}}\right|\leq\sqrt{h_\alpha}\sqrt{n}\widehat{D}_n\xrightarrow{p}0\,.
\end{aligned}$$

Por lo tanto, si llamamos $\widehat{g}_\alpha(x_\alpha)=\int\mathbf{e}_1^{\mathrm{T}}\mathbf{H}_\alpha\widetilde{\boldsymbol{\beta}}(\mathbf{x})q_{\underline{\alpha}}(\mathbf{x}_{\underline{\alpha}})\,d\mathbf{x}_{\underline{\alpha}}$, bastará hallar la distribución asintótica de

$$\sqrt{nh_\alpha}\,[\widehat{g}_\alpha(x_\alpha)-g_\alpha(x_\alpha)]=\sqrt{nh_\alpha}\int\mathbf{e}_1^{\mathrm{T}}\mathbf{H}_\alpha[\widetilde{\boldsymbol{\beta}}(\mathbf{x})-\boldsymbol{\beta}(\mathbf{x})]q_{\underline{\alpha}}(\mathbf{x}_{\underline{\alpha}})\,d\mathbf{x}_{\underline{\alpha}}\,,$$

es decir, nos hemos reducido a probar el resultado en el caso en que la escala es conocida.

Haciendo un desarrollo de Taylor de orden uno alrededor de $\boldsymbol{\beta}(\mathbf{x})$, obtenemos que

$$\mathbf{H}_\alpha[\widetilde{\beta}(\mathbf{x})-\beta(\mathbf{x})]=\sigma\,\widehat{\mathbf{A}}_{0,n}^{-1}(\mathbf{x})\,\widehat{\mathbf{A}}_{1,n}(\mathbf{x})\tag{4.30}$$

donde

$$\begin{aligned}
\widehat{\mathbf{A}}_{0,n}(\mathbf{x}) &= \frac{1}{n}\sum_{i=1}^{n}\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i-\mathbf{x})\xi_i\psi'\left(\frac{Y_i-\breve{\mathbf{x}}_{i,\alpha}^{\mathrm{T}}\widehat{\boldsymbol{\theta}}(\mathbf{x})}{\sigma}\right)\breve{\mathbf{x}}_{i,\alpha}\breve{\mathbf{x}}_{i,\alpha}^{\mathrm{T}}\\
\widehat{\mathbf{A}}_{1,n}(\mathbf{x}) &= \frac{1}{n}\sum_{i=1}^{n}\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i-\mathbf{x})\xi_i\psi\left(\frac{Y_i-\breve{\mathbf{x}}_{i,\alpha}^{\mathrm{T}}\mathbf{H}_\alpha\boldsymbol{\beta}(\mathbf{x})}{\sigma}\right)\breve{\mathbf{x}}_{i,\alpha}
\end{aligned}$$

donde $\widehat{\boldsymbol{\theta}}(\mathbf{x})$ es un punto intermedio entre $\mathbf{H}_\alpha\boldsymbol{\beta}(\mathbf{x})$ y $\mathbf{H}_\alpha\widetilde{\boldsymbol{\beta}}(\mathbf{x})$. Sea $\mathbf{A}_0(\mathbf{u}) = v(\mathbf{u})A_0(\psi)\mathbf{S}^{(\alpha)}$ donde $v(\mathbf{u}) = r(\mathbf{u})f_{\mathbf{X}}(\mathbf{u})$ y $A_0(\psi) = \mathbb{E}(\psi'(\varepsilon))$ como antes. Es fácil ver que $\sup_{\mathbf{x}\in\mathcal{S}_Q} |\widehat{\mathbf{A}}_{0,n}(\mathbf{x}) - \mathbf{A}_0(\mathbf{x})| = o_{\mathbb{P}}(1)$, luego empezaremos estudiando el comportamiento de

$$\widehat{\mathbf{B}}_n = \sqrt{nh_\alpha}\sigma \int \mathbf{A}_0^{-1}(\mathbf{x})\widehat{\mathbf{A}}_{1,n}(\mathbf{x})q_{\underline{\alpha}}(\mathbf{x}_{\underline{\alpha}})\, d\mathbf{x}_{\underline{\alpha}}\,.$$

Para destacar la dependencia en $\mathbf{x}$ definamos

$$\begin{aligned} R_\alpha(X_{i,\alpha}, x_\alpha) &= g_\alpha(X_{i,\alpha}) - g_\alpha(x_\alpha) - \sum_{j=1}^{q} \frac{1}{j!} g_\alpha^{(j)}(x_\alpha)(X_{i,\alpha} - x_\alpha)^j \\ R(\mathbf{X}_i, \mathbf{x}) &= \sum_{j\neq\alpha} \{g_j(X_{ij}) - g_j(x_j)\} + R_\alpha(X_{i,\alpha}, x_\alpha) \end{aligned}$$

Luego, $Y_i - \breve{\mathbf{x}}_{i,\alpha}\mathbf{H}_\alpha\boldsymbol{\beta}(\mathbf{x}) = \epsilon_i + R(\mathbf{X}_i, \mathbf{x})/\sigma$. Por otra parte, **N1** y el hecho que $h_j = \widetilde{h}$ si $j \neq \alpha$ y $R(\mathbf{x}, \mathbf{x}) = 0$ implican que

$$\begin{aligned} R(\mathbf{x} + \mathbf{H}_d\mathbf{u}, \mathbf{x}) &= \sum_{j\neq\alpha}\sum_{s=1}^{\ell} \widetilde{h}^s \frac{g_j^{(s)}(x_j)}{s!} u_j^s + \widetilde{h}^\ell \sum_{j\neq\alpha} \frac{g_j^{(\ell)}(\xi_j) g_j^{(\ell)}(x_j)}{\ell!} u_j^\ell \\ &\quad + h_\alpha^{q+1} \frac{g_\alpha^{(q+1)}(x_\alpha)}{(q+1)!} u_\alpha^{q+1} + h_\alpha^{q+1} \frac{g_\alpha^{(q+1)}(\xi_\alpha) - g_\alpha^{(q+1)}(x_\alpha)}{(q+1)!} u_\alpha^{q+1} \end{aligned} \tag{4.31}$$

Definamos

$$\widetilde{\mathbf{A}}_{1,n}(\mathbf{x}) = \frac{1}{n}\sum_{i=1}^{n} \mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i - \mathbf{x}) r(\mathbf{X}_i) \lambda\left(\frac{R(\mathbf{X}_i, \mathbf{x})}{\sigma}\right) \breve{\mathbf{x}}_{i,\alpha}\,.$$

Observemos que $\widehat{\mathbf{B}}_n = \mathbf{B}_{n,1} + \widehat{\mathbf{B}}_{n,2}$ donde

$$\begin{aligned} \widehat{\mathbf{B}}_{n,1} &= \sigma\sqrt{nh_\alpha} \int \mathbf{A}_0^{-1}(\mathbf{x})\widetilde{\mathbf{A}}_{1,n}(\mathbf{x})\, q_{\underline{\alpha}}(\mathbf{x}_{\underline{\alpha}})\, d\mathbf{x}_{\underline{\alpha}} \\ \widehat{\mathbf{B}}_{n,2} &= \sigma\sqrt{nh_\alpha} \int \mathbf{A}_0^{-1}(\mathbf{x}) \left[\widehat{\mathbf{A}}_{1,n}(\mathbf{x}) - \widetilde{\mathbf{A}}_{1,n}(\mathbf{x})\right] q_{\underline{\alpha}}(\mathbf{x}_{\underline{\alpha}})\, d\mathbf{x}_{\underline{\alpha}}\,. \end{aligned}$$

Para concluir la demostración del Teorema deberemos mostrar que

a) $\widehat{\mathbf{B}}_{n,1} \xrightarrow{p} \mathbf{b}_{q,\alpha}(x_\alpha)$ con $\mathbf{b}_{q,\alpha}(x_\alpha) = \beta^{(2q+3)/2}/((q+1)!)g_\alpha^{(q+1)}(x_\alpha)(\mathbf{S}^{(\alpha)})^{-1}\mathbf{s}_q^{(\alpha)}$

b) $\widehat{\mathbf{B}}_{n,2} \xrightarrow{D} N_{q+1}(\mathbf{0}, \boldsymbol{\Sigma}_{q,\alpha}(x_\alpha))$ con

$$\boldsymbol{\Sigma}_{q,\alpha}(x_\alpha) = \sigma^2 \frac{\mathbb{E}(\psi^2(\varepsilon))}{A_0^2(\psi)} \int \frac{q_{\underline{\alpha}}^2(\mathbf{x}_{\underline{\alpha}})}{f_{\mathbf{X}}(\mathbf{x})p(\mathbf{x})}\, d\mathbf{x}_{\underline{\alpha}} (\mathbf{S}^{(\alpha)})^{-1}\mathbf{V}_\alpha(\mathbf{S}^{(\alpha)})^{-1}\,.$$

c) $\sqrt{nh_\alpha}[\widehat{g}_\alpha(x_\alpha) - g_\alpha(x_\alpha)] - \widehat{B}_n \xrightarrow{p} 0$ donde $\widehat{B}_n = \mathbf{e}_1^{\mathrm{T}}\widehat{\mathbf{B}}_n$

Luego, para obtener el resultado deseado bastará con multiplicar a izquierda por $\mathbf{e}_1^{\mathrm{T}}$.

Observemos que $\mathbb{E}\widetilde{\mathbf{A}}_{1,n}(\mathbf{x}) = \mathbb{E}\widehat{\mathbf{A}}_{1,n}(\mathbf{x})$. Empezaremos calculando $\mathbb{E}\widetilde{\mathbf{A}}_{1,n}(\mathbf{x})$ ya que la usaremos en los pasos a seguir.

Como λ es $\ell-1$ veces diferenciable y $\lambda^{(\ell-1)}$ es Lipschitz, haciendo un desarrollo de Taylor de orden $\ell-1$ y usando que $\lambda(0)=0$, $\lambda'(0)=\mathbb{E}\psi'(\varepsilon_1)=A_0(\psi)$ resulta que

$$\begin{aligned}
\lambda\left(\frac{R(\mathbf{u},\mathbf{x})}{\sigma}\right) &= \lambda(0)+\sum_{k=1}^{\ell-1}\frac{1}{k!}\lambda^{(k)}(0)\left(\frac{R(\mathbf{u},\mathbf{x})}{\sigma}\right)^k+\frac{1}{(\ell-1)!}\left[\lambda^{(\ell-1)}(\theta(\mathbf{u}))-\lambda^{(\ell-1)}(0)\right]\left(\frac{R(\mathbf{u},\mathbf{x})}{\sigma}\right)^{\ell-1}\\
&= A_0(\psi)\frac{R(\mathbf{u},\mathbf{x})}{\sigma}+\sum_{k=2}^{\ell-1}\frac{1}{k!}\lambda^{(s)}(0)\left(\frac{R(\mathbf{u},\mathbf{x})}{\sigma}\right)^k+\frac{1}{(\ell-1)!}\widetilde{\lambda}(\mathbf{u},\mathbf{x})\left(\frac{R(\mathbf{u},\mathbf{x})}{\sigma}\right)^{\ell-1}
\end{aligned}$$

donde $\widetilde{\lambda}(\mathbf{u},\mathbf{x})=\lambda^{(\ell-1)}(\theta(\mathbf{u}))-\lambda^{(\ell-1)}(0)$ con $\theta(\mathbf{u})$ un punto intermedio entre 0 y $R(\mathbf{u},\mathbf{x})/\sigma$. Observemos además que $|\widetilde{\lambda}(\mathbf{u},\mathbf{x})|\le C|\theta(\mathbf{u})|\le C|R(\mathbf{u},\mathbf{x})|/\sigma$ pues $\lambda^{(\ell-1)}$ es Lipschitz. Luego, tenemos que

$$\begin{aligned}
\mathbb{E}\left(\widehat{\mathbf{A}}_{1,n}(\mathbf{x})\right) &= \mathbb{E}\left[\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_1-\mathbf{x})r(\mathbf{X}_1)\lambda\left(\frac{R(\mathbf{X}_1,\mathbf{x})}{\sigma}\right)\breve{\mathbf{x}}_{i,\alpha}\right]\\
&= \frac{A_0(\psi)}{\sigma}\mathbf{A}_{11,n}+\sum_{k=2}^{\ell-1}\frac{\lambda^{(k)}(0)}{k!\sigma^k}\mathbf{A}_{1k,n}+\frac{1}{(\ell-1)!\,\sigma^{\ell-1}}\mathbf{A}_{1\ell,n}
\end{aligned}$$

donde

$$\begin{aligned}
\mathbf{A}_{11,n} &= \frac{1}{h_\alpha\widetilde{h}^{d-1}}\mathbb{E}\left[\prod_{s=1}^{d}K_s\left(\frac{X_{1,s}-x_s}{h_s}\right)r(\mathbf{X}_1)R(\mathbf{X}_1,\mathbf{x})\breve{\mathbf{x}}_{1,\alpha}\right]\\
&= \int\mathcal{K}(\mathbf{u})v(\mathbf{x}+\mathbf{H}_d\mathbf{u})R(\mathbf{x}+\mathbf{H}_d\mathbf{u},\mathbf{x})\breve{\mathbf{u}}_\alpha\,d\mathbf{u}\\
\mathbf{A}_{1k,n} &= \frac{1}{h_\alpha\widetilde{h}^{d-1}}\mathbb{E}\left[\prod_{s=1}^{d}K_s\left(\frac{X_{1,s}-x_s}{h_s}\right)r(\mathbf{X}_1)R^k(\mathbf{X}_1,\mathbf{x})\breve{\mathbf{x}}_{1,\alpha}\right]\\
&= \int\mathcal{K}(\mathbf{u})v(\mathbf{x}+\mathbf{H}_d\mathbf{u})R^k(\mathbf{x}+\mathbf{H}_d\mathbf{u},\mathbf{x})\breve{\mathbf{u}}_\alpha\,d\mathbf{u}\qquad k=2,\dots,\ell-1\\
\mathbf{A}_{1\ell,n} &= \frac{1}{h_\alpha\widetilde{h}^{d-1}}\mathbb{E}\left[\prod_{s=1}^{d}K_s\left(\frac{X_{1,s}-x_s}{h_s}\right)r(\mathbf{X}_1)\widetilde{\lambda}(\mathbf{X}_1,\mathbf{x})R^{\ell-1}(\mathbf{X}_1,\mathbf{x})\breve{\mathbf{x}}_{1,\alpha}\right]\\
&= \int\mathcal{K}(\mathbf{u})v(\mathbf{x}+\mathbf{H}_d\mathbf{u})\widetilde{\lambda}(\mathbf{x}+\mathbf{H}_d\mathbf{u},\mathbf{x})R^{\ell}(\mathbf{x}+\mathbf{H}_d\mathbf{u},\mathbf{x})\breve{\mathbf{u}}_\alpha\,d\mathbf{u}
\end{aligned}$$

donde $\breve{\mathbf{u}}_\alpha=(1,u_\alpha,\dots,u_\alpha^q)^{\mathrm{T}}\in\mathbb{R}^{q+1}$. Mediante un desarrollo de Taylor ℓ de v alrededor de $\mathbf{x}$ obtenemos que

$$v(\mathbf{x}+\mathbf{H}_d\mathbf{u})=v(\mathbf{x})+\sum_{0<|\mathbf{m}|\le\ell}D^{\mathbf{m}}v(\mathbf{x})\mathbf{h}^{\mathbf{m}}\mathbf{u}^{\mathbf{m}}+\sum_{|\mathbf{m}|=\ell}\frac{1}{\mathbf{m}!}\left[D^{\mathbf{m}}v(\boldsymbol{\xi}^{(1)})-D^{\mathbf{m}}v(\mathbf{x})\right]\mathbf{h}^{\mathbf{m}}\mathbf{u}^{\mathbf{m}}\qquad(4.32)$$

donde hemos utilizado la notación de Bourbaki para el desarrollo, $\mathbf{h}=\mathrm{diag}(h_1,\dots,h_d)$, $\mathbf{m}=(m_1,\dots,m_d)$ con $m_i\in\mathbb{N}$, $|\mathbf{m}|=\sum_{j=1}^d m_j$, $\mathbf{u}^{\mathbf{m}}=(u_1^{m_1},\dots,u_d^{m_d})$, $\mathbf{m}!=m_1!\dots m_d!$ y $D^{\mathbf{m}}v=\partial^{|\mathbf{m}|}v/\partial u_1^{m_1}\dots\partial u_d^{m_d}$.

Utilizando (4.31) y (4.32) en $\mathbf{A}_{1k,n}$, $k = 1, \ldots, \ell$ obtenemos que $\mathbf{A}_{11,n} = \sum_{j=1}^{18} \mathbf{A}_{11,j,n}$ donde

$$\mathbf{A}_{11,1,n} = v(\mathbf{x}) \left\{ \sum_{j \neq \alpha} g_j'(x_j) \widetilde{h} \int \prod_{t=1}^{d} K_t(u_t) u_j \breve{\mathbf{u}}_\alpha \, d\mathbf{u} \right\}$$

$$\mathbf{A}_{11,2,n} = v(\mathbf{x}) \frac{1}{(q+1)!} g_\alpha^{(q+1)}(x_\alpha) h_\alpha^{q+1} \int \prod_{t=1}^{d} K_t(u_t) u_\alpha^{q+1} \breve{\mathbf{u}}_\alpha \, d\mathbf{u}$$

$$\mathbf{A}_{11,3,n} = \sum_{k=1}^{\ell-1} \sum_{|\mathbf{m}|=k} \frac{1}{\mathbf{m}!} D^{\mathbf{m}} v(\mathbf{x}) \sum_{j \neq \alpha} g_j'(x_j) \, \widetilde{h} \, \mathbf{h}^{\mathbf{m}} \int \prod_{t=1}^{d} K_t(u_t) u_j \mathbf{u}^{\mathbf{m}} \breve{\mathbf{u}}_\alpha \, d\mathbf{u}$$

$$\mathbf{A}_{11,4,n} = \sum_{k=1}^{\ell-1} \sum_{|\mathbf{m}|=k} \frac{1}{\mathbf{m}!} D^{\mathbf{m}} v(\mathbf{x}) \frac{1}{(q+1)!} g_\alpha^{(q+1)}(x_\alpha) h_\alpha^{q+1} \mathbf{h}^{\mathbf{m}} \int \prod_{t=1}^{d} K_t(u_t) u_\alpha^{q+1} \mathbf{u}^{\mathbf{m}} \breve{\mathbf{u}}_\alpha \, d\mathbf{u}$$

$$\mathbf{A}_{11,5,n} = \sum_{|\mathbf{m}|=\ell} \frac{1}{\mathbf{m}!} D^{\mathbf{m}} v(\mathbf{x}) \sum_{j \neq \alpha} g_j'(x_j) \, \widetilde{h} \, \mathbf{h}^{\mathbf{m}} \int \prod_{t=1}^{d} K_t(u_t) u_j \mathbf{u}^{\mathbf{m}} \breve{\mathbf{u}}_\alpha \, d\mathbf{u}$$

$$\mathbf{A}_{11,6,n} = \sum_{|\mathbf{m}|=\ell} \frac{1}{\mathbf{m}!} D^{\mathbf{m}} v(\mathbf{x}) \frac{1}{(q+1)!} g_\alpha^{(q+1)}(x_\alpha) h_\alpha^{q+1} \mathbf{h}^{\mathbf{m}} \int \prod_{t=1}^{d} K_t(u_t) u_\alpha^{q+1} \mathbf{u}^{\mathbf{m}} \breve{\mathbf{u}}_\alpha \, d\mathbf{u}$$

$$\mathbf{A}_{11,7,n} = \sum_{|\mathbf{m}|=\ell} \sum_{j \neq \alpha} g_j'(x_j) \, \widetilde{h} \, \frac{1}{\mathbf{m}!} \mathbf{h}^{\mathbf{m}} \int \left[D^{\mathbf{m}} v(\boldsymbol{\xi}^{(1)}) - D^{\mathbf{m}} v(\mathbf{x}) \right] \prod_{t=1}^{d} K_t(u_t) u_j \mathbf{u}^{\mathbf{m}} \breve{\mathbf{u}}_\alpha \, d\mathbf{u}$$

$$\mathbf{A}_{11,8,n} = \sum_{|\mathbf{m}|=\ell} \frac{1}{\mathbf{m}!} \frac{1}{(q+1)!} g_\alpha^{(q+1)}(x_\alpha) h_\alpha^{q+1} \mathbf{h}^{\mathbf{m}} \int \left[D^{\mathbf{m}} v(\boldsymbol{\xi}^{(1)}) - D^{\mathbf{m}} v(\mathbf{x}) \right] \prod_{t=1}^{d} K_t(u_t) u_\alpha^{q+1} \mathbf{u}^{\mathbf{m}} \breve{\mathbf{u}}_\alpha \, d\mathbf{u}$$

$$\mathbf{A}_{11,9,n} = v(\mathbf{x}) \left\{ \sum_{j \neq \alpha} \sum_{m=2}^{\ell-1} \frac{1}{m!} g_j^{(m)}(x_j) \widetilde{h}^m \int \prod_{t=1}^{d} K_t(u_t) u_j^m \breve{\mathbf{u}}_\alpha \, d\mathbf{u} \right\}$$

$$\mathbf{A}_{1110,n} = v(\mathbf{x}) \left\{ \sum_{j \neq \alpha} \frac{1}{\ell!} g_j^{(\ell)}(\boldsymbol{\xi}_j) \widetilde{h}^\ell \int \prod_{t=1}^{d} K_t(u_t) u_j^\ell \breve{\mathbf{u}}_\alpha \, d\mathbf{u} \right\}$$

$$\mathbf{A}_{11,11,n} = v(\mathbf{x}) \left\{ \frac{1}{(q+1)!} h_\alpha^{q+1} \int \left[g_\alpha^{(\ell)}(\boldsymbol{\xi}_\alpha) - g_\alpha^{(\ell)}(x_\alpha) \right] \prod_{t=1}^{d} K_t(u_t) u_\alpha^{q+1} \breve{\mathbf{u}}_\alpha \, d\mathbf{u} \right\}$$

$$\mathbf{A}_{11,12,n} = \sum_{k=1}^{\ell-1} \sum_{|\mathbf{m}|=k} \frac{1}{\mathbf{m}!} D^{\mathbf{m}} v(\mathbf{x}) \sum_{j \neq \alpha} \sum_{m=2}^{\ell} \frac{1}{m!} g_j^{(m)}(x_j) \widetilde{h}^m \mathbf{h}^{\mathbf{m}} \int \prod_{t=1}^{d} K_t(u_t) u_j^m \mathbf{u}^{\mathbf{m}} \breve{\mathbf{u}}_\alpha \, d\mathbf{u}$$

$$\mathbf{A}_{1113,n} = \sum_{k=1}^{\ell-1} \sum_{|\mathbf{m}|=k} \frac{1}{\mathbf{m}!} D^{\mathbf{m}} v(\mathbf{x}) \sum_{j \neq \alpha} \widetilde{h}^\ell \mathbf{h}^{\mathbf{m}} \int [g_j^{(\ell)}(\boldsymbol{\xi}_j) - g_j^{(\ell)}(x_j)] \prod_{t=1}^{d} K_t(u_t) u_j^\ell \mathbf{u}^{\mathbf{m}} \breve{\mathbf{u}}_\alpha \, d\mathbf{u}$$

$$\mathbf{A}_{11,14,n} = \sum_{|\mathbf{m}|=\ell} \frac{1}{\mathbf{m}!} \sum_{j \neq \alpha} \widetilde{h}^\ell \mathbf{h}^{\mathbf{m}} \int D^{\mathbf{m}} v(\boldsymbol{\xi}^{(1)}) \left[g_j^{(\ell)}(\boldsymbol{\xi}_j) - g_j^{(\ell)}(x_j) \right] \prod_{t=1}^{d} K_t(u_t) u_j^\ell \mathbf{u}^{\mathbf{m}} \breve{\mathbf{u}}_\alpha \, d\mathbf{u}$$

$$\mathbf{A}_{11,15,n} = \sum_{k=1}^{\ell-1} \sum_{|\mathbf{m}|=k} \frac{1}{\mathbf{m}!} D^{\mathbf{m}} v(\mathbf{x}) \frac{1}{(q+1)!} h_\alpha^{q+1} \mathbf{h}^{\mathbf{m}} \int \left[g_\alpha^{(q+1)}(\boldsymbol{\xi}_\alpha) - g_\alpha^{(q+1)}(x_\alpha) \right] \prod_{t=1}^{d} K_t(u_t) u_\alpha^{q+1} \mathbf{u}^{\mathbf{m}} \breve{\mathbf{u}}_\alpha \, d\mathbf{u}$$

$$\mathbf{A}_{11,16,n} = \sum_{|\mathbf{m}|=\ell} \frac{1}{\mathbf{m}!} \frac{1}{(q+1)!} h_\alpha^{q+1} \mathbf{h}^{\mathbf{m}} \int D^{\mathbf{m}} v(\boldsymbol{\xi}^{(1)}) \left[g_\alpha^{(q+1)}(\boldsymbol{\xi}_\alpha) - g_\alpha^{(q+1)}(x_\alpha) \right] \prod_{t=1}^{d} K_t(u_t) u_\alpha^{q+1} \mathbf{u}^{\mathbf{m}} \breve{\mathbf{u}}_\alpha \, d\mathbf{u}$$

$$\mathbf{A}_{11,17,n} = \sum_{|\mathbf{m}|=\ell} \frac{1}{\mathbf{m}!} D^{\mathbf{m}} v(\mathbf{x}) \sum_{j \neq \alpha} \sum_{m=2}^{\ell} \frac{1}{m!} g_j^{(m)}(x_j) \widetilde{h}^m \mathbf{h}^{\mathbf{m}} \int \prod_{t=1}^{d} K_t(u_t) u_j^m \mathbf{u}^{\mathbf{m}} \breve{\mathbf{u}}_\alpha \, d\mathbf{u}$$

$$\mathbf{A}_{11,18,n} = \sum_{|\mathbf{m}|=\ell} \frac{1}{\mathbf{m}!} \sum_{j \neq \alpha} \sum_{m=2}^{\ell} \frac{1}{m!} g_j^{(m)}(x_j) \widetilde{h}^m \mathbf{h}^{\mathbf{m}} \int \left[D^{\mathbf{m}} v(\boldsymbol{\xi}^{(1)}) - D^{\mathbf{m}} v(\mathbf{x}) \right] \prod_{t=1}^{d} K_t(u_t) u_j^m \mathbf{u}^{\mathbf{m}} \breve{\mathbf{u}}_\alpha \, d\mathbf{u} \,,$$

con $\boldsymbol{\xi}^{(1)}$ un punto intermedio entre $\mathbf{x}$ y $\mathbf{x} + \mathbf{H}_d \mathbf{u}$. Como $\int K_j(t) t \, dt = 0$ tenemos que $\mathbf{A}_{11,1,n} = \mathbf{0}$. Como $K_j = L$ es un núcleo de orden ℓ, si $j \neq \alpha$, entonces $\int \mathcal{K}(\mathbf{u}) u_j \mathbf{u}^{\mathbf{m}} \breve{\mathbf{u}}_\alpha \, d\mathbf{u} = 0$ si $|\mathbf{m}| \leq \ell - 2$ o bien si $|\mathbf{m}| = \ell - 1$ pero cuando $\mathbf{m} \neq (\ell-1)\mathbf{e}_j$, y $\int \mathcal{K}(\mathbf{u}) \mathbf{u}^{\mathbf{m}} u_\alpha^{q+1} \breve{\mathbf{u}}_\alpha \, d\mathbf{u} = 0$ para todo $\mathbf{m}$ con al menos una coordenada $j \neq \alpha$ distinta de 0. Luego, $\mathbf{A}_{11,9,n} = \mathbf{0}$.

Por otra parte, tenemos que

$$\begin{aligned} \mathbf{A}_{11,2,n} &= h_\alpha^{q+1} v(\mathbf{x}) \frac{1}{(q+1)!} g_\alpha^{(q+1)}(x_\alpha) \int K_\alpha(u_\alpha) u_\alpha^{q+1} \breve{\mathbf{u}}_\alpha \, du_\alpha \\ &= h_\alpha^{q+1} v(\mathbf{x}) \frac{1}{(q+1)!} g_\alpha^{(q+1)}(x_\alpha) \mathbf{s}_q^{(\alpha)} \,. \end{aligned}$$

Como $D^{\mathbf{m}} v(\mathbf{u})$ con $|\mathbf{m}| = k$ y $g_\alpha^{(q+1)}$ son continuas y acotadas, y g_j es ℓ veces derivable, con derivadas acotadas para todo $j \neq \alpha$, resulta que

$$\sup_{\mathbf{x} \in \mathcal{S}_Q} \|\mathbf{A}_{11,3,n}\| = \widetilde{h}^\ell O(1) \,, \qquad \sup_{\mathbf{x} \in \mathcal{S}_Q} \|\mathbf{A}_{11,4,n}\| = h_\alpha^{q+2} O(1) \qquad \text{y} \qquad \sup_{\mathbf{x} \in \mathcal{S}_Q} \|\mathbf{A}_{11,10,n}\| = \widetilde{h}^\ell O(1) \,.$$

De la misma manera se puede ver que como ℓ es par, $\int \prod_{t=1}^{d} K_t(u_t) u_j \mathbf{u}^{\mathbf{m}} \breve{\mathbf{u}}_\alpha \, d\mathbf{u} = 0$ salvo cuando $\mathbf{m}$ tenga un 1 en la posición correspondiente a u_α, ℓ en la posición j y 0 en las demás, de donde

$$\begin{aligned} \mathbf{A}_{11,5,n} &= \sum_{j \neq \alpha} g_j'(x_j) \frac{1}{(\ell-1)!} \left. \frac{\partial^\ell v(\mathbf{u})}{\partial u_j^{\ell-1} \partial u_\alpha} \right|_{\mathbf{u}=\mathbf{x}} \widetilde{h}^\ell h_\alpha \int \mathcal{K}(\mathbf{u}) u_j^\ell \breve{\mathbf{u}}_\alpha \, d\mathbf{u} \\ &\quad + \sum_{j \neq \alpha} g_j'(x_j) \frac{1}{\ell!} \left. \frac{\partial^\ell v(\mathbf{u})}{\partial u_j^\ell} \right|_{\mathbf{u}=\mathbf{x}} \widetilde{h}^{\ell+1} \int \mathcal{K}(\mathbf{u}) u_j^{\ell+1} \breve{\mathbf{u}}_\alpha \, d\mathbf{u} \\ \mathbf{A}_{11,6,n} &= \frac{1}{(q+1)!} g_\alpha^{(q+1)}(x_\alpha) h_\alpha^{q+1} \sum_{j \neq \alpha} \frac{1}{\ell!} \left. \frac{\partial^\ell v(\mathbf{u})}{\partial u_j^\ell} \right|_{\mathbf{u}=\mathbf{x}} h_j^\ell \int \mathcal{K}(\mathbf{u}) u_\alpha^{q+1} u_j^\ell \breve{\mathbf{u}}_\alpha \, d\mathbf{u} \\ &\quad + \frac{1}{(q+1)!} g_\alpha^{(q+1)}(x_\alpha) h_\alpha^{q+1} \frac{1}{\ell!} \left. \frac{\partial^\ell v(\mathbf{u})}{\partial u_\alpha^\ell} \right|_{\mathbf{u}=\mathbf{x}} h_\alpha^\ell \int \mathcal{K}(\mathbf{u}) u_\alpha^{q+1+\ell} \breve{\mathbf{u}}_\alpha \, d\mathbf{u} \,, \end{aligned}$$

con lo cual

$$\sup_{\mathbf{x} \in \mathcal{S}_Q} \|\mathbf{A}_{11,5,n}\| = \widetilde{h}^\ell (\widetilde{h} + h_\alpha) O(1) \qquad \sup_{\mathbf{x} \in \mathcal{S}_Q} \|\mathbf{A}_{11,6,n}\| = h_\alpha^{q+1} (\widetilde{h}^\ell + h_\alpha^\ell) O(1) \,.$$

Por otra parte, como tenemos que $D^{\mathbf{m}}v(\mathbf{u})$ con $|\mathbf{m}| = k \leq \ell$, $k \leq \ell$, es uniformemente continua K_j tienen soporte en $[-1,1]$, de modo tal que $\mathbf{x} + \mathbf{H}_d\mathbf{u}$ pertenece a un entorno de radio $\max_j h_j$ de $\mathbf{x}$, tenemos que

$$\sup_{\mathbf{x}\in\mathcal{S}_Q} \left| D^{\mathbf{m}}v(\boldsymbol{\xi}^{(1)}) - D^{\mathbf{m}}v(\mathbf{x}) \right| = o(1)$$

con lo cual

$$\sup_{\mathbf{x}\in\mathcal{S}_Q} \|\mathbf{A}_{11,7,n}\| = \widetilde{h}^{\ell} o(1)\,, \qquad \sup_{\mathbf{x}\in\mathcal{S}_Q} \|\mathbf{A}_{11,8,n}\| = h_\alpha^{q+1} o(1) \qquad \text{y} \qquad \sup_{\mathbf{x}\in\mathcal{S}_Q} \|\mathbf{A}_{11,18,n}\| = \widetilde{h}^{\ell} O(1)\,.$$

De la misma manera, como $g_j^{(\ell)}$ es uniformemente continua y acotada, se tiene que $\sup_{\mathbf{x}\in\mathcal{S}_Q} |g_j^{(\ell)}(\boldsymbol{\xi}_j) - g_j^{(\ell)}(x_j)| = o(1)$ y por lo tanto

$$\begin{aligned}
\sup_{\mathbf{x}\in\mathcal{S}_Q} \|\mathbf{A}_{11,11,n}\| &= h_\alpha^{q+1} o(1) & \sup_{\mathbf{x}\in\mathcal{S}_Q} \|\mathbf{A}_{11,12,n}\| &= \widetilde{h}^{\ell} O(1) \\
\sup_{\mathbf{x}\in\mathcal{S}_Q} \|\mathbf{A}_{11,13,n}\| &= \widetilde{h}^{\ell}(\widetilde{h} + h_\alpha) o(1) & \sup_{\mathbf{x}\in\mathcal{S}_Q} \|\mathbf{A}_{11,14,n}\| &= \widetilde{h}^{\ell} o(1) \\
\sup_{\mathbf{x}\in\mathcal{S}_Q} \|\mathbf{A}_{11,15,n}\| &= h_\alpha^{q+2} o(1) & \sup_{\mathbf{x}\in\mathcal{S}_Q} \|\mathbf{A}_{11,16,n}\| &= h_\alpha^{q+1} o(1) \\
\sup_{\mathbf{x}\in\mathcal{S}_Q} \|\mathbf{A}_{1117,n}\| &= \widetilde{h}^{\ell} O(1) & \sup_{\mathbf{x}\in\mathcal{S}_Q} \|\mathbf{A}_{1118,n}\| &= \widetilde{h}^{\ell} o(1)\,.
\end{aligned}$$

Mediante un razonamiento similar, usando (4.31), que K_j es un núcleo de orden ℓ si $j \neq \alpha$ y que $\widetilde{h}^{\ell} = o(h_\alpha^{q+1})$, obtenemos que para todo $k = 2, \ldots, \ell - 1$

$$\sup_{\mathbf{x}\in\mathcal{S}_Q} \|\mathbf{A}_{1k,n}\| = h_\alpha^{q+1} o(1)\,.$$

Sea $A_{1\ell,n,s}$ indica la coordenada s de $\mathbf{A}_{1\ell,n}$. Usando que $|\widetilde{\lambda}(\mathbf{u},\mathbf{x})| \leq C|R(\mathbf{u},\mathbf{x})|/\sigma$, K_j tiene soporte en $[-1,1]$, que v es acotada y (4.31), tenemos que para $s = 1, \ldots q+1$

$$\begin{aligned}
\sup_{\mathbf{x}\in\mathcal{S}_Q} |A_{1\ell,n,s}| &\leq \sup_{\mathbf{x}\in\mathcal{S}_Q} \int |\mathcal{K}(\mathbf{u})||v(\mathbf{x}+\mathbf{H}_d\mathbf{u})||\widetilde{\lambda}(\mathbf{x}+\mathbf{H}_d\mathbf{u},\mathbf{x})|\, |R(\mathbf{x}+\mathbf{H}_d\mathbf{u},\mathbf{x})|^{\ell-1}|u_\alpha|^{s-1}\, d\mathbf{u} \\
&\leq \frac{C}{\sigma} \int |\mathcal{K}(\mathbf{u})||v(\mathbf{x}+\mathbf{H}_d\mathbf{u})|\, |R(\mathbf{x}+\mathbf{H}_d\mathbf{u},\mathbf{x})|^{\ell}\, d\mathbf{u} \leq c_2(\widetilde{h} + h_\alpha^{q+1})^{\ell} = o\left(h_\alpha^{q+1}\right)\,,
\end{aligned}$$

de donde

$$\sup_{\mathbf{x}\in\mathcal{S}_Q} \|\mathbf{A}_{1\ell,n}\| = h_\alpha^{q+1} o(1)\,.$$

Resumiendo, tenemos que, como $h_\alpha \to 0$ y $\widetilde{h} \to 0$,

$$\begin{aligned}
\mathbb{E}\widetilde{\mathbf{A}}_{1,n}(\mathbf{x}) = \mathbb{E}\widehat{\mathbf{A}}_{1,n}(\mathbf{x}) &= \frac{A_0(\psi)}{\sigma}\mathbf{A}_{11,n} + \sum_{k=2}^{\ell-1} \frac{\lambda^{(k)}(0)}{k!\sigma^k}\mathbf{A}_{1k,n} + \frac{1}{(\ell-1)!\,\sigma^{\ell-1}}\mathbf{A}_{1\ell,n} \\
&= \frac{A_0(\psi)}{\sigma} h_\alpha^{q+1} v(\mathbf{x}) \frac{1}{(q+1)!} g_\alpha^{(q+1)}(x_\alpha)\mathbf{s}_q^{(\alpha)} + \boldsymbol{\nu}_n(\mathbf{x}) \qquad (4.33)
\end{aligned}$$

donde $\sup_{\mathbf{x}\in\mathcal{S}_Q} \|\boldsymbol{\nu}_n(\mathbf{x})\| = \widetilde{h}^{\ell} O(1) + h_\alpha^{q+1} o(1) = h_\alpha^{q+1} o(1)$.

• Probaremos ahora a). Para ver que $\widehat{\mathbf{B}}_{n,1} \xrightarrow{p} \mathbf{b}_{q,\alpha}$, basta ver que $\mathbb{E}\widehat{\mathbf{B}}_{n,1} \to \mathbf{b}_{q,\alpha}$ y que para todo $1 \leq j \leq q+1$, $\text{VAR}(\widehat{B}_{n,1,j}) \to 0$. Por (4.33) y usando que $\mathbf{A}_0(\mathbf{u}) = v(\mathbf{u})A_0(\psi)\mathbf{S}^{(\alpha)}$ y que $\sqrt{nh_\alpha}h_\alpha^{q+1} = \beta^{(2q+3)/2}$ tenemos que

$$\begin{aligned}
\mathbb{E}\widehat{\mathbf{B}}_{n,1} &= \sigma\sqrt{nh_\alpha}\int \mathbf{A}_0(\mathbf{x})^{-1}\mathbb{E}\widetilde{\mathbf{A}}_{1,n}(\mathbf{x})q_{\underline{\alpha}}(\mathbf{x}_{\underline{\alpha}})\,d\mathbf{x}_{\underline{\alpha}} = \mathbf{B}_{11,n} + \mathbf{B}_{12,n} \\
\mathbf{B}_{11,n} &= \beta^{(2q+3)/2}\frac{1}{q+1}g_\alpha^{(q+1)}(x_\alpha)(\mathbf{S}^{(\alpha)})^{-1}\mathbf{s}_q^{(\alpha)} \\
\mathbf{B}_{12,n} &= \sqrt{nh_\alpha}\int \mathbf{A}_0^{-1}(\mathbf{x})\boldsymbol{\nu}_n(\mathbf{x})q_{\underline{\alpha}}(\mathbf{x}_{\underline{\alpha}})\,d\mathbf{x}_{\underline{\alpha}}\,.
\end{aligned}$$

Como $\sup_{\mathbf{x}\in\mathcal{S}_Q}\|\boldsymbol{\nu}_n(\mathbf{x})\| = h_\alpha^{q+1}o(1)$ y $\sqrt{nh_\alpha}h_\alpha^{q+1} = \beta^{(2q+3)/2}$ tenemos que $\mathbf{B}_{12,n} \to \mathbf{0}$ y por lo tanto $\mathbb{E}\widehat{\mathbf{B}}_{1,n} \to \mathbf{b}_{q,\alpha}(x_\alpha)$.
Veamos ahora que $\text{VAR}(\widehat{B}_{n,1,j}) \to 0$, para $1 \leq j \leq q+1$. Como $\mathbf{S}^{(\alpha)}$, σ y $A_0(\psi)$ son constantes, bastará con probar que la varianza de $\widetilde{B}_j$ donde $\widetilde{B}_j$ es la coordenada $j-$ésima de $\widetilde{\mathbf{B}} = A_0(\psi)\mathbf{S}^{(\alpha)}\widehat{\mathbf{B}}_{n,1}/\sigma$ tiende a 0. Observemos primero que como $\mathbf{A}_0(\mathbf{u}) = v(\mathbf{u})A_0(\psi)\mathbf{S}^{(\alpha)}$ entonces

$$\begin{aligned}
\widetilde{B}_j &= \sqrt{nh_\alpha}\int \frac{1}{v(\mathbf{x})}\widetilde{A}_{1,n,j}(\mathbf{x})q_{\underline{\alpha}}(\mathbf{x}_{\underline{\alpha}})\,d\mathbf{x}_{\underline{\alpha}} \\
&= \frac{1}{\sqrt{n\,h_\alpha}}\sum_{i=1}^n K_\alpha\left(\frac{X_{i,\alpha}-x_\alpha}{h_\alpha}\right)r(\mathbf{X}_i)\zeta(\mathbf{H}_{d-1},\mathbf{X}_i,x_\alpha)\left(\frac{X_{i,\alpha}-x_\alpha}{h_\alpha}\right)^{j-1}
\end{aligned}$$

con

$$\begin{aligned}
\zeta(\mathbf{H}_{d-1},\mathbf{X}_i,x_\alpha) &= \frac{1}{\widetilde{h}^{d-1}}\int\prod_{s\neq\alpha}K_s\left(\frac{X_{i,s}-x_s}{\widetilde{h}}\right)q_{\underline{\alpha}}(\mathbf{x}_{\underline{\alpha}})\lambda\left(\frac{R(\mathbf{X}_i,\mathbf{x})}{\sigma}\right)d\mathbf{x}_{\underline{\alpha}} \\
&= \int\prod_{s\neq\alpha}K_s(u_s)q_{\underline{\alpha}}(\mathbf{X}_{i,\underline{\alpha}}+\mathbf{H}_{d-1}\mathbf{u}_{\underline{\alpha}})\lambda\left(\frac{R(\mathbf{X}_i,(x_\alpha,\mathbf{X}_{i,\underline{\alpha}}+\mathbf{H}_{d-1}\mathbf{u}_{\underline{\alpha}}))}{\sigma}\right)d\mathbf{u}_{\underline{\alpha}}
\end{aligned}$$

donde $\mathbf{H}_{d-1} = \text{diag}(\widetilde{h},\ldots,\widetilde{h}) \in \mathbb{R}^{(d-1)\times(d-1)}$. Luego, como ψ es acotada, $q_{\underline{\alpha}}$ es continua y acotada tenemos que $|\zeta(\mathbf{H}_{d-1},\mathbf{X}_i,x_\alpha)| \leq C$, para todo i. Usando que r es acotada y que $|X_{1,\alpha}-x_\alpha| \leq h_\alpha$ si $K_\alpha\left((X_{1,\alpha}-x_\alpha)/h_\alpha\right) \neq 0$ obtenemos que

$$\begin{aligned}
\text{VAR}(\widetilde{B}_j) &= \frac{1}{h_\alpha}\text{VAR}\left(K_\alpha\left(\frac{X_{1,\alpha}-x_\alpha}{h_\alpha}\right)r(\mathbf{X}_i)\zeta(\mathbf{H}_{d-1},\mathbf{X}_i,x_\alpha)\left(\frac{X_{1,\alpha}-x_\alpha}{h_\alpha}\right)^{j-1}\right) \\
&\leq \frac{1}{h_\alpha}\mathbb{E}\left[K_\alpha^2\left(\frac{X_{1,\alpha}-x_\alpha}{h_\alpha}\right)r^2(\mathbf{X}_i)\zeta^2(\mathbf{H}_{d-1},\mathbf{X}_i,x_\alpha)\left(\frac{X_{1,\alpha}-x_\alpha}{h_\alpha}\right)^{2(j-1)}\right]. \\
&\leq \frac{1}{h_\alpha}\int K_\alpha^2\left(\frac{v_\alpha-x_\alpha}{h_\alpha}\right)\zeta^2(\mathbf{H}_{d-1},\mathbf{v},x_\alpha)f_{\mathbf{X}}(\mathbf{v})\,d\mathbf{v} \\
&\leq \int K_\alpha^2(u_\alpha)\zeta^2(\mathbf{H}_{d-1},(x_\alpha+h_\alpha u_\alpha,\mathbf{v}_{\underline{\alpha}}),x_\alpha)f_{\mathbf{X}}(x_\alpha+h_\alpha u_\alpha,\mathbf{v}_{\underline{\alpha}})\,du_\alpha d\mathbf{v}_{\underline{\alpha}}
\end{aligned}$$

Luego, como si $n\to\infty$, $\mathbf{H}_d \to \mathbf{0}$ y como $\zeta^2(\mathbf{0}_{d-1},(x_\alpha,\mathbf{v}_{\underline{\alpha}}),x_\alpha)$ pues $\lambda(0)=0$, usando el Teorema de la Convergencia Mayorada resulta que $\text{VAR}(\widetilde{B}_j) \to 0$ dando por concluido a).

• Probaremos ahora b). Observemos primero que bastará probar que

$$\mathbf{B}_{n,2} = \frac{A_0(\psi)}{\sigma}\mathbf{S}^{(\alpha)}\widehat{\mathbf{B}}_{n,2} \xrightarrow{D} N_{q+1}(\mathbf{0}, \mathbf{\Sigma}_{11}(x_\alpha))$$

donde

$$\mathbf{\Sigma}_{11}(x_\alpha) = \mathbb{E}\psi^2(\varepsilon)\int \frac{q^2_{\underline{\alpha}}(\mathbf{x}_{\underline{\alpha}})}{f_{\mathbf{X}}(x_\alpha, \mathbf{x}_{\underline{\alpha}})p(x_\alpha, \mathbf{x}_{\underline{\alpha}})}\, d\mathbf{x}_{\underline{\alpha}}\mathbf{V}_\alpha\,.$$

Llamemos

$$V(\epsilon_i, \mathbf{H}_d, \mathbf{X}_i, \mathbf{x}) = \xi_i\psi\left(\varepsilon_i + \frac{R(\mathbf{X}_i, \mathbf{x})}{\sigma}\right) - r(\mathbf{X}_i)\lambda\left(\frac{R(\mathbf{X}_i, \mathbf{x})}{\sigma}\right).$$

Es claro que

$$\widehat{\mathbf{A}}_{1,n}(\mathbf{x}) - \widetilde{\mathbf{A}}_{1,n}(\mathbf{x}) = \frac{1}{n}\sum_{i=1}^n \mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i - \mathbf{x})\breve{\mathbf{x}}_{i,\alpha}V(\epsilon_i, \mathbf{H}_d, \mathbf{X}_i, \mathbf{x})\,,$$

de donde

$$\begin{aligned}\mathbf{B}_{n,2} &= \sqrt{nh_\alpha}\int v^{-1}(\mathbf{x})\left[\widehat{\mathbf{A}}_{1,n}(\mathbf{x}) - \widetilde{\mathbf{A}}_{1,n}(\mathbf{x})\right] q_{\underline{\alpha}}(\mathbf{x}_{\underline{\alpha}})\, d\mathbf{x}_{\underline{\alpha}} \\ &= \frac{1}{\sqrt{nh_\alpha}}\sum_{i=1}^n K_\alpha\left(\frac{X_{i,\alpha} - x_\alpha}{h_\alpha}\right)\breve{\mathbf{x}}_{i,\alpha}\gamma(\epsilon_i, \mathbf{H}_{d-1}, \mathbf{X}_i, x_\alpha)\,,\end{aligned}$$

con

$$\gamma(\epsilon_i, \mathbf{H}_{d-1}, \mathbf{X}_i, x_\alpha) = \frac{1}{\widetilde{h}^{d-1}}\int \frac{1}{v(\mathbf{x})}\prod_{j\neq\alpha} K_j\left(\frac{X_{i,j} - x_j}{\widetilde{h}}\right) q_{\underline{\alpha}}(\mathbf{x}_{\underline{\alpha}})V(\epsilon_i, \mathbf{H}_d, \mathbf{X}_i, (x_\alpha, \mathbf{x}_{\underline{\alpha}}))\, d\mathbf{x}_{\underline{\alpha}}$$

$$= \int \frac{1}{v(x_\alpha, \mathbf{X}_{i,\underline{\alpha}} + \mathbf{H}_{d-1}\mathbf{u}_{\underline{\alpha}})}\prod_{j\neq\alpha} K_j(u_j)q_{\underline{\alpha}}(\mathbf{X}_{i,\underline{\alpha}} + \mathbf{H}_{d-1}\mathbf{u}_{\underline{\alpha}})V(\epsilon_i, \mathbf{H}_d, \mathbf{X}_i, (x_\alpha, \mathbf{X}_{i,\underline{\alpha}} + \mathbf{H}_{d-1}\mathbf{u}_{\underline{\alpha}}))\, d\mathbf{u}_{\underline{\alpha}}$$

Sea

$$\mathbf{W}_{i,n} = \frac{1}{\sqrt{nh_\alpha}}K_\alpha\left(\frac{X_{i,\alpha} - x_\alpha}{h_\alpha}\right)\breve{\mathbf{x}}_{i,\alpha}\gamma(\epsilon, \mathbf{H}_{d-1}, \mathbf{X}_i, x_\alpha)\,.$$

Tenemos que $\mathbf{B}_{n,2} = \sum_{i=1}^n \mathbf{W}_{i,n}$.

Para probar que $\mathbf{B}_{n,2} \xrightarrow{D} N_{q+1}(\mathbf{0}, \mathbf{\Sigma}_{11}(x_\alpha))$, bastará mostrar que para todo $\mathbf{c} \in \mathbb{R}^{q+1}$, $\mathbf{c} \neq \mathbf{0}$ se tiene que $\mathbf{c}^{\mathrm{T}}\mathbf{B}_{n,2} = \sum_{i=1}^n \mathbf{c}^{\mathrm{T}}\mathbf{W}_{i,n} \xrightarrow{D} N(0, \mathbf{c}^{\mathrm{T}}\mathbf{\Sigma}_{11}(x_\alpha)\mathbf{c})$.

Sea $\mathbf{c} \in \mathbb{R}^{q+1}$, $\mathbf{c} \neq \mathbf{0}$. Como $\mathbb{E}(V(\epsilon_i, \mathbf{H}_d, \mathbf{X}_i, \mathbf{x})|\mathbf{X}_i) = \mathbf{0}$ para todo $\mathbf{x}$ tenemos que $\mathbb{E}\mathbf{W}_{i,n} = \mathbf{0}$ de donde $\mathbb{E}\mathbf{c}^{\mathrm{T}}\mathbf{W}_{i,n} = 0$. Además, como ψ, p y r son acotadas y $|\xi_i| \leq 1/i(t)$, $|V(\epsilon_i, \mathbf{H}_d, \mathbf{X}_i, \mathbf{x})| \leq C$ con lo cual para alguna constante genérica $C_1 > 0$

$$\begin{aligned}\sum_{i=1}^n \mathbb{E}|\mathbf{c}^{\mathrm{T}}\mathbf{W}_{i,n}|^3 &\leq C_1 n\frac{1}{(nh_\alpha)^{3/2}}\mathbb{E}\left|K_\alpha\left(\frac{X_{i,\alpha} - x_\alpha}{h_\alpha}\right)\right|^3 = C_1\frac{1}{\sqrt{nh_\alpha}}\frac{1}{h_\alpha}\int\left|K^3_\alpha\left(\frac{u - x_\alpha}{h_\alpha}\right)\right| f_{X_\alpha}(u)\, du \\ &\leq C_2\frac{1}{\sqrt{nh_\alpha}} \to 0\,.\end{aligned}$$

Por lo tanto, la demostración de b) quedará concluída si mostramos que $\lim_{n\to\infty} \mathrm{VAR}(\mathbf{c}^{\mathrm{T}} \sum_{i=1}^{n} \mathbf{W}_{i,n}) = \mathbf{c}^{\mathrm{T}}\mathbf{\Sigma}_{11}(x_\alpha)\mathbf{c}$ ya que basta aplicar la versión de Lyapunov del Teorema Central del Límite a las variables $\mathbf{c}^{\mathrm{T}}\mathbf{W}_{i,n}$ que forman un sistema triangular de variables independientes pues $\mathbf{W}_{1,n}, \ldots, \mathbf{W}_{n,n}$ son independientes para cada n fijo.

Calculemos $\mathrm{VAR}(\mathbf{c}^{\mathrm{T}} \sum_{i=1}^{n} \mathbf{W}_{i,n})$, indiquemos por $\mathrm{VAR}(\mathbf{W}_{i,n})$ la matriz de covarianza de $\mathbf{W}_{i,n}$, luego sólo basta ver que $\lim_{n\to\infty} \mathrm{VAR}(\sum_{i=1}^{n} \mathbf{W}_{i,n}) = \mathbf{\Sigma}_{11}(x_\alpha)$. Como $\mathbf{W}_{1,n}, \ldots, \mathbf{W}_{n,n}$ son independientes, $\mathrm{VAR}(\sum_{i=1}^{n} \mathbf{W}_{i,n}) = \sum_{i=1}^{n} \mathrm{VAR}(\mathbf{W}_{i,n}) = n\mathrm{VAR}(\mathbf{W}_{1,n})$. Por otra parte, como $\mathbb{E}\mathbf{W}_{1,n} = 0$, $\mathrm{VAR}(\mathbf{W}_{1,n}) = \mathbb{E}\left(\mathbf{W}_{1,n}\mathbf{W}_{1,n}^{\mathrm{T}}\right)$. Calcularemos entonces, para $1 \le s, m \le q+1$, $E_{sm} = n\mathbb{E}\left(W_{1,n,s}W_{1,n,m}\right)$ donde $W_{1,n,m}$ es la coordenada m de $\mathbf{W}_{1,n}$,

$$\begin{aligned}
E_{sm} = &= \frac{1}{h_\alpha}\mathbb{E}\left[K_\alpha^2\left(\frac{X_{1,\alpha}-x_\alpha}{h_\alpha}\right)\gamma^2(\epsilon_1, \mathbf{H}_{d-1}, \mathbf{X}_1, x_\alpha)\left(\frac{X_{1,\alpha}-x_\alpha}{h_\alpha}\right)^{s+m-2}\right] \\
&= \frac{1}{h_\alpha}\int K_\alpha^2\left(\frac{u_\alpha - x_\alpha}{h_\alpha}\right)\mathbb{E}\left[\gamma^2(\epsilon_1, \mathbf{H}_{d-1}, \mathbf{u}, x_\alpha)|\mathbf{X}_1 = \mathbf{u}\right]\left(\frac{u_\alpha - x_\alpha}{h_\alpha}\right)^{s+m-2} f_{\mathbf{X}}(u_\alpha, \mathbf{u}_{\underline{\alpha}})\, d\mathbf{u} \\
&= \int K_\alpha^2(u_\alpha)u_\alpha^{s+m-2}\mathbb{E}\left[\gamma^2(\epsilon_1, \mathbf{H}_{d-1}, (u_\alpha h_\alpha + x_\alpha, \mathbf{u}_{\underline{\alpha}}), x_\alpha)|\mathbf{X}_1 = (u_\alpha h_\alpha + x_\alpha, \mathbf{u}_{\underline{\alpha}})\right] \\
&\quad \times f_{\mathbf{X}}(u_\alpha h_\alpha + x_\alpha, \mathbf{u}_{\underline{\alpha}})\, d\mathbf{u}\,. \qquad (4.34)
\end{aligned}$$

Calculemos

$$\begin{aligned}
M(\mathbf{H}_{d-1}, \mathbf{X}_1, x_\alpha) &= \mathbb{E}\left[\gamma^2(\epsilon_1, \mathbf{H}_{d-1}, \mathbf{X}_1, x_\alpha)\big|\,\mathbf{X}_1\right] \\
&= \frac{1}{\widetilde{h}^{2(d-1)}}\int\int \frac{1}{v(x_\alpha, \mathbf{x}_{\underline{\alpha}})}\frac{1}{v(x_\alpha, \mathbf{z}_{\underline{\alpha}})}\prod_{j\neq\alpha} K_j\left(\frac{X_{1,j}-x_j}{\widetilde{h}}\right)K_j\left(\frac{X_{1,j}-z_j}{\widetilde{h}}\right) \\
\times \quad & q_{\underline{\alpha}}(\mathbf{x}_{\underline{\alpha}})q_{\underline{\alpha}}(\mathbf{z}_{\underline{\alpha}})\mathbb{E}\left[V(\epsilon_1, \mathbf{H}_d, \mathbf{X}_1, (x_\alpha, \mathbf{x}_{\underline{\alpha}}))V(\epsilon_1, \mathbf{H}_d, \mathbf{X}_1, (x_\alpha, \mathbf{z}_{\underline{\alpha}}))|\mathbf{X}_1\right]\, d\mathbf{x}_{\underline{\alpha}}d\mathbf{z}_{\underline{\alpha}} \qquad (4.35)
\end{aligned}$$

Es fácil ver que

$$\begin{aligned}
\mathbb{E}\left[V(\epsilon_1, \mathbf{H}_d, \mathbf{X}_1, (x_\alpha, \mathbf{x}_{\underline{\alpha}}))V(\epsilon_1, \mathbf{H}_d, \mathbf{X}_1, (x_\alpha, \mathbf{z}_{\underline{\alpha}}))|\mathbf{X}_1\right] &= \frac{p(\mathbf{X}_1)}{t^2(\mathbf{X}_1)}\bar{\lambda}\left(R(\mathbf{X}_1, (x_\alpha, \mathbf{x}_{\underline{\alpha}})), R(\mathbf{X}_1, (x_\alpha, \mathbf{z}_{\underline{\alpha}}))\right) \\
&\quad - r^2(\mathbf{X}_1)\lambda\left(\frac{R(\mathbf{X}_1, (x_\alpha, \mathbf{x}_{\underline{\alpha}}))}{\sigma}\right)\lambda\left(\frac{R(\mathbf{X}_1, (x_\alpha, \mathbf{z}_{\underline{\alpha}}))}{\sigma}\right) \qquad (4.36)
\end{aligned}$$

donde $\bar{\lambda}(a, b) = \mathbb{E}\left[\psi(\varepsilon_1 + a/\sigma)\psi(\varepsilon_1 + b/\sigma)\right]$. Luego, reeemplazando (4.36) en (4.35) y haciendo cambio de variables queda

$$\begin{aligned}
M(\mathbf{H}_{d-1}, \mathbf{X}_1, x_\alpha) = &\int\int \frac{1}{v(x_\alpha, \mathbf{X}_{1,\underline{\alpha}} + \mathbf{H}_{d-1}\mathbf{v}_{\underline{\alpha}})}\frac{1}{v(x_\alpha, \mathbf{X}_{1,\underline{\alpha}} + \mathbf{H}_{d-1}\mathbf{z}_{\underline{\alpha}})}\prod_{j\neq\alpha} K_j(v_j)K_j(z_j) \\
&\times q_{\underline{\alpha}}(\mathbf{X}_{1,\underline{\alpha}} + \mathbf{H}_{d-1}\mathbf{v}_{\underline{\alpha}})q_{\underline{\alpha}}(\mathbf{X}_{1,\underline{\alpha}} + \mathbf{H}_{d-1}\mathbf{z}_{\underline{\alpha}}) \\
&\times \left[\frac{p(\mathbf{X}_1)}{t^2(\mathbf{X}_1)}\bar{\lambda}(R(\mathbf{X}_1, (x_\alpha, \mathbf{X}_{1,\underline{\alpha}} + \mathbf{H}_{d-1}\mathbf{v}_{\underline{\alpha}})), R(\mathbf{X}_1, (x_\alpha, \mathbf{X}_{1,\underline{\alpha}} + \mathbf{H}_{d-1}\mathbf{z}_{\underline{\alpha}}))) \right. \\
- \quad & \left. r^2(\mathbf{X}_1)\lambda\left(\frac{R(\mathbf{X}_1, (x_\alpha, \mathbf{X}_{i,\underline{\alpha}} + \mathbf{H}_{d-1}\mathbf{v}_{\underline{\alpha}}))}{\sigma}\right)\lambda\left(\frac{R(\mathbf{X}_1, (x_\alpha, \mathbf{X}_{i,\underline{\alpha}} + \mathbf{H}_{d-1}\mathbf{z}_{\underline{\alpha}}))}{\sigma}\right)\right] d\mathbf{v}_{\underline{\alpha}}d\mathbf{z}_{\underline{\alpha}}\,.
\end{aligned}$$

Por último, (4.34) puede escribirse como

$$E_{sm} = \int K_\alpha^2(u_\alpha)u_\alpha^{s+m-2}M(\mathbf{H}_{d-1}, (u_\alpha h_\alpha + x_\alpha, \mathbf{u}_{\underline{\alpha}}), x_\alpha)f_{\mathbf{X}}(u_\alpha h_\alpha + x_\alpha, \mathbf{u}_{\underline{\alpha}})\, d\mathbf{u}\,. \qquad (4.37)$$

Tomando límite en (4.37) y usando el Teorema de la Convergencia Mayorada, se tiene que

$$\lim_{n\to\infty} E_{sm} = \int K_\alpha^2(u_\alpha) u_\alpha^{s+m-2} M(\mathbf{0},(x_\alpha,\mathbf{u}_{\underline{\alpha}}),x_\alpha) f_{\mathbf{X}}(x_\alpha,\mathbf{u}_{\underline{\alpha}})\, d\mathbf{u}\,.$$

Sea $\widetilde{\lambda}_2(a) = \mathbb{E}\psi^2(\varepsilon + a)$ y observemos que $\bar{\lambda}(a,a) = \widetilde{\lambda}_2(a)$ y $\widetilde{\lambda}_2(0) = \mathbb{E}\psi^2(\varepsilon)$. Luego usando que $R(\mathbf{x},\mathbf{x}) = 0$ y $\lambda(0) = 0$ deducimos que

$$\begin{aligned} M(\mathbf{0},(x_\alpha,\mathbf{u}_{\underline{\alpha}}),x_\alpha) &= \frac{q_{\underline{\alpha}}^2(\mathbf{u}_{\underline{\alpha}})}{v^2(x_\alpha,\mathbf{u}_{\underline{\alpha}})}\,\frac{p(x_\alpha,\mathbf{u}_{\underline{\alpha}})}{t^2(x_\alpha,\mathbf{u}_{\underline{\alpha}})}\widetilde{\lambda}_2(R((x_\alpha,\mathbf{u}_{\underline{\alpha}}),(x_\alpha,\mathbf{u}_{\underline{\alpha}})))\int\prod_{j\neq\alpha}K_j(v_j)\,d\mathbf{v}_{\underline{\alpha}}\int\prod_{j\neq\alpha}K_j(z_j)\,d\mathbf{z}_{\underline{\alpha}} \\ &= \frac{q_{\underline{\alpha}}^2(\mathbf{u}_{\underline{\alpha}})}{v^2(x_\alpha,\mathbf{u}_{\underline{\alpha}})}\,\frac{p(x_\alpha,\mathbf{u}_{\underline{\alpha}})}{t^2(x_\alpha,\mathbf{u}_{\underline{\alpha}})}\mathbb{E}\psi^2(\varepsilon)\,. \end{aligned}$$

Es decir, hemos mostrado que

$$\begin{aligned} \lim_{n\to\infty} E_{sm} = \lim_{n\to\infty}\sum_{i=1}^n \mathrm{VAR}\,(\mathbf{W}_{i,n})_{s,m} &= \int K_\alpha^2(u_\alpha)u_\alpha^{s+m-2} f_{\mathbf{X}}(x_\alpha,\mathbf{u}_{\underline{\alpha}})\frac{q_{\underline{\alpha}}^2(\mathbf{u}_{\underline{\alpha}})}{v^2(x_\alpha,\mathbf{u}_{\underline{\alpha}})}\,\frac{r^2(x_\alpha,\mathbf{u}_{\underline{\alpha}})}{p(x_\alpha,\mathbf{u}_{\underline{\alpha}})}\mathbb{E}\psi^2(\varepsilon) \\ &= \mathbb{E}\psi^2(\varepsilon)\int\frac{q_{\underline{\alpha}}^2(\mathbf{u}_{\underline{\alpha}})}{f_{\mathbf{X}}(x_\alpha,\mathbf{u}_{\underline{\alpha}})p(x_\alpha,\mathbf{u}_{\underline{\alpha}})}\,d\mathbf{u}_{\underline{\alpha}} v_{sm}^{(\alpha)} \end{aligned}$$

donde $v_{sm}^{(\alpha)}$ es el elemento (s,m) de la matriz $\mathbf{V}_\alpha$, lo que concluye la demostración de b).

• Probaremos ahora c).

Queremos ver que $\sqrt{nh_\alpha}[\widehat{g}_\alpha(x_\alpha) - g_\alpha(x_\alpha)] - \widehat{B}_n \xrightarrow{p} 0$ donde $\widehat{B}_n = \mathbf{e}_1^{\mathrm{T}}\widehat{\mathbf{B}}_n$. Observemos que $\mathbf{H}_\alpha[\widetilde{\boldsymbol{\beta}}(\mathbf{x}) - \boldsymbol{\beta}(\mathbf{x})] - \sigma\mathbf{A}_0^{-1}(\mathbf{x})\widehat{\mathbf{A}}_{1,n}(\mathbf{x}) = \sigma(\widehat{\mathbf{A}}_0^{-1}(\mathbf{x}) - \mathbf{A}_0^{-1}(\mathbf{x}))\widehat{\mathbf{A}}_{1,n}(\mathbf{x}) = \widehat{\mathbf{D}}_1(\mathbf{x}) + \widehat{\mathbf{D}}_2(\mathbf{x})$ con

$$\begin{aligned} \widehat{\mathbf{D}}_1(\mathbf{x}) &= \sigma(\widehat{\mathbf{A}}_{0,n}^{-1}(\mathbf{x}) - \mathbf{A}_0^{-1}(\mathbf{x}))\mathbb{E}\widehat{\mathbf{A}}_{1,n}(\mathbf{x}) \\ \widehat{\mathbf{D}}_2(\mathbf{x}) &= \sigma(\widehat{\mathbf{A}}_{0,n}^{-1}(\mathbf{x}) - \mathbf{A}_0^{-1}(\mathbf{x}))\left(\widehat{\mathbf{A}}_{1,n}(\mathbf{x}) - \mathbb{E}\widehat{\mathbf{A}}_{1,n}(\mathbf{x})\right)\,. \end{aligned}$$

Probaremos que, para todo $1 \le j \le q+1$

$$\sqrt{nh_\alpha}\sup_{\mathbf{x}\in\mathcal{S}_Q}|\widehat{D}_{1,j}(\mathbf{x})| \xrightarrow{p} 0 \tag{4.38}$$

$$\sqrt{nh_\alpha}\sup_{\mathbf{x}\in\mathcal{S}_Q}|\widehat{D}_{2,j}(\mathbf{x})| \xrightarrow{p} 0\,, \tag{4.39}$$

donde $\widehat{D}_{\ell,j}(\mathbf{x})$ es la coordenada j de $\widehat{\mathbf{D}}_\ell(\mathbf{x})$, $\ell = 1,2$. Efectivamente, si se cumplen (4.38) y (4.39) tenemos que

$$\sqrt{nh_\alpha}\,[\widehat{g}_\alpha(x_\alpha) - g_\alpha(x_\alpha)] - \mathbf{e}_1^{\mathrm{T}}\widehat{\mathbf{B}}_n = \sqrt{nh_\alpha}\int\mathbf{e}_1^{\mathrm{T}}\left(\widehat{\mathbf{D}}_1(\mathbf{x}) + \widehat{\mathbf{D}}_2(\mathbf{x})\right)q_{\underline{\alpha}}(\mathbf{x}_{\underline{\alpha}})\,d\mathbf{x}_{\underline{\alpha}} \xrightarrow{p} 0$$

lo que concluiría la demostración.

Sea $1 \le j \le q+1$. Para probar (4.38), observemos que como $\widetilde{h}^\ell = o(h_\alpha^{q+1})$, usando (4.33) y la desigualdad de Cauchy-Schwartz, se obtiene que

$$\sup_{\mathbf{x}\in\mathcal{S}_Q}|\widehat{D}_{1,j}(\mathbf{x})| \le \sigma\sup_{\mathbf{x}\in\mathcal{S}_Q}\left\|\mathbf{e}_j^{\mathrm{T}}(\widehat{\mathbf{A}}_{0,n}^{-1}(\mathbf{x}) - \mathbf{A}_0^{-1})\right\|\,o(h_\alpha^{q+1})\,,$$

donde le término $o(h_\alpha^{q+1})$ no depende de $\mathbf{x}$. Por otra parte, como $\mathbf{S}^{(\alpha)}$ es no singular, $\inf_{\mathbf{x}\in\mathcal{S}_Q}|v(\mathbf{x})| > 0$, $A_0(\psi)\neq 0$ y $\sup_{\mathbf{x}\in\mathcal{S}_Q}\left\|\widehat{\mathbf{A}}_{0,n}(\mathbf{x})-\mathbf{A}_0(\mathbf{x})\right\| \xrightarrow{p} 0$ tenemos que $\sup_{\mathbf{x}\in\mathcal{S}_Q}\left\|\widehat{\mathbf{A}}_{0,n}^{-1}(\mathbf{x})-\mathbf{A}_0^{-1}(\mathbf{x})\right\| \xrightarrow{p} 0$. Por lo tanto, como $\sqrt{nh_\alpha}h_\alpha^{q+1}=\beta^{(2q+3)/2}$ obtenemos que $\sqrt{nh_\alpha}\sup_{\mathbf{x}\in\mathcal{S}_Q}|\widehat{D}_{1,j}(\mathbf{x})| \xrightarrow{p} 0$ para todo $1\le j\le q+1$, lo que concluye la demostración de (4.38).

Probaremos ahora (4.39). El Lema 3.4.6 con $\theta_n=\sqrt{\log n/(nh_\alpha\widetilde{h}^{d-1})}$ aplicado a cada coordenada del vector $\widehat{\mathbf{A}}_{1,n}(\mathbf{x})=(\widehat{A}_{1,n,1}(\mathbf{x}),\ldots,\widehat{A}_{1,n,q+1}(\mathbf{x}))^{\mathrm{T}}$, implica que, para $1\le j\le q+1$,

$$\sup_{\mathbf{x}\in\mathcal{S}_Q}\left|\widehat{A}_{1,n,j}(\mathbf{x})-\mathbb{E}\widehat{A}_{1,n,j}(\mathbf{x})\right| = O_{\mathbb{P}}\left(\left(\frac{\log n}{nh_\alpha\widetilde{h}^{d-1}}\right)^{1/2}\right).$$

Por otra parte, usando (4.33) y que $\widetilde{h}^{\ell}=o(h_\alpha^{q+1})$ obtenemos que

$$\sup_{\mathbf{x}\in\mathcal{S}_Q}\left|\mathbb{E}\widehat{A}_{1,n,j}(\mathbf{x})\right| = o(h_\alpha^{q+1}).$$

Por lo tanto, como $\sup_{\mathbf{x}\in\mathcal{S}_Q}\left\|\widehat{\mathbf{A}}_{0,n}^{-1}(\mathbf{x})-\mathbf{A}_0^{-1}(\mathbf{x})\right\| \xrightarrow{p} \mathbf{0}$ y $\inf_{\mathbf{x}\in\mathcal{S}_Q}\nu_1(\mathbf{A}_0(\mathbf{x}))>0$ deducimos que $\sup_{\mathbf{x}\in\mathcal{S}_Q}\nu_{q+1}(\widehat{A}_{0,n}^{-1}(\mathbf{x}))=O_{\mathbb{P}}(1)$, de donde usando (4.30) obtenemos que

$$\sup_{\mathbf{x}\in\mathcal{S}_Q}\left\|\mathbf{H}_\alpha[\widetilde{\beta}(\mathbf{x})-\beta(\mathbf{x})]\right\| \le o_{\mathbb{P}}(h_\alpha^{q+1})+O_{\mathbb{P}}\left(\left(\frac{\log n}{nh_\alpha\widetilde{h}^{d-1}}\right)^{1/2}\right). \tag{4.40}$$

Usando nuevamente Cauchy-Schwartz, tenemos que

$$\begin{aligned}\sqrt{nh_\alpha}\sup_{\mathbf{x}\in\mathcal{S}_Q}|\widehat{D}_{2,j}(\mathbf{x})| &\le \sigma\tilde{C}\sqrt{nh_\alpha}\sup_{\mathbf{x}\in\mathcal{S}_Q}\|\mathbf{e}_j^{\mathrm{T}}(\widehat{\mathbf{A}}_{0,n}(\mathbf{x})-\mathbf{A}_0(\mathbf{x}))\|\sup_{\mathbf{x}\in\mathcal{S}_Q}\left\|\widehat{\mathbf{A}}_{1,n}(\mathbf{x})-\mathbb{E}\widehat{\mathbf{A}}_{1,n}(\mathbf{x})\right\| \\ &\le C\sup_{\mathbf{x}\in\mathcal{S}_Q}\|\mathbf{e}_j^{\mathrm{T}}(\widehat{\mathbf{A}}_{0,n}(\mathbf{x})-\mathbf{A}_0(\mathbf{x}))\|O_{\mathbb{P}}\left(\left(\frac{\log n}{\widetilde{h}^{d-1}}\right)^{1/2}\right) \end{aligned}\tag{4.41}$$

Definamos $\widetilde{\mathbf{A}}_n$ como

$$\widetilde{\mathbf{A}}_n(\mathbf{x},a)=\frac{1}{n}\sum_{i=1}^n\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i-\mathbf{x})\xi_i\psi'\left(\frac{Y_i-\breve{\mathbf{x}}_{i,\alpha}^{\mathrm{T}}(\mathbf{H}_\alpha\boldsymbol{\beta}(\mathbf{x})+\mathbf{a})}{\sigma}\right)\breve{\mathbf{x}}_{i,\alpha}\breve{\mathbf{x}}_{i,\alpha}^{\mathrm{T}}.$$

Recordemos que $\lambda_1(t)=\mathbb{E}\psi'(\varepsilon_1+t)$ y $\lambda_1(0)=A_0(\psi)$. Llamemos

$$\begin{aligned}\mathbf{\Lambda}_{1,n}(\mathbf{x},\mathbf{a}) &= \mathbb{E}\widetilde{\mathbf{A}}_n(\mathbf{x},\mathbf{a})=\mathbb{E}\left[\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_1-\mathbf{x})r(\mathbf{X}_1)\lambda_1\left(\frac{R(\mathbf{X}_1,\mathbf{x})-\breve{\mathbf{x}}_{i,\alpha}^{\mathrm{T}}\mathbf{a}}{\sigma}\right)\breve{\mathbf{x}}_{1,\alpha}\breve{\mathbf{x}}_{1,\alpha}^{\mathrm{T}}\right] \\ &= \int\mathcal{K}(\mathbf{u})v(\mathbf{x}+\mathbf{H}_d\mathbf{u})\lambda_1\left(\frac{R(\mathbf{x}+\mathbf{H}_d\mathbf{u},\mathbf{x})-\breve{\mathbf{u}}_\alpha\mathbf{a}}{\sigma}\right)\breve{\mathbf{u}}_\alpha\breve{\mathbf{u}}_\alpha^{\mathrm{T}}\,d\mathbf{u}.\end{aligned}$$

Sea $\widetilde{A}_{n,j,m}(\mathbf{x},\mathbf{a})$ el elemento (j,m) de la matriz $\widetilde{\mathbf{A}}_n$. Argumentos análogos a los considerados en el Lema 3.4.6. con $\theta_n=\sqrt{\log n/(nh_\alpha\widetilde{h}^{d-1})}$ implican que, para $1\le j,m\le q+1$

$$\sup_{\mathbf{x}\in\mathcal{S}_Q}\sup_{\|\mathbf{a}\|\le 1}\left|\widetilde{A}_{n,j,m}(\mathbf{x},\mathbf{a})-\mathbb{E}\widetilde{A}_{n,j,m}(\mathbf{x},\mathbf{a})\right| = O_{\mathbb{P}}\left(\left(\frac{\log n}{nh_\alpha\widetilde{h}^{d-1}}\right)^{1/2}\right).$$

Recordemos que $\widehat{\boldsymbol{\theta}}(\mathbf{x})$ era un punto intermedio entre $\mathbf{H}_\alpha\widetilde{\boldsymbol{\beta}}(\mathbf{x})$ y $\mathbf{H}_\alpha\boldsymbol{\beta}(\mathbf{x})$, luego, usando que $\sup_{\mathbf{x}\in\mathcal{S}_Q}\|\mathbf{H}_\alpha[\widetilde{\boldsymbol{\beta}}(\mathbf{x})-\boldsymbol{\beta}(\mathbf{x})]\| = o_{\mathbb{P}}(1)$ (ver (4.40)) tenemos que $\sup_{\mathbf{x}\in\mathcal{S}_Q}\|\widehat{\boldsymbol{\theta}}(\mathbf{x})-\mathbf{H}_\alpha\boldsymbol{\beta}(\mathbf{x})\| = o_{\mathbb{P}}(1)$. Sea $\widehat{\mathbf{a}}(\mathbf{x}) = \widehat{\boldsymbol{\theta}}(\mathbf{x})-\mathbf{H}_\alpha\boldsymbol{\beta}(\mathbf{x})$, luego tenemos que $\sup_{\mathbf{x}\in\mathcal{S}_Q}\|\widehat{\mathbf{a}}(\mathbf{x})\| = o_{\mathbb{P}}(1)$. Usando que $\widehat{\mathbf{A}}_{0,n}(\mathbf{x}) = \widetilde{\mathbf{A}}_n(\mathbf{x},\widehat{\mathbf{a}}(\mathbf{x}))$, tenemos que, para todo $1\le j,m\le q+1$

$$\sup_{\mathbf{x}\in\mathcal{S}_Q}\left|\widehat{A}_{0,n,j,m}(\mathbf{x})-\lambda_{1,n,j,m}(\mathbf{x},\widehat{\mathbf{a}}(\mathbf{x}))\right| = O_{\mathbb{P}}\left(\left(\frac{\log n}{nh_\alpha\widetilde{h}^{d-1}}\right)^{1/2}\right), \tag{4.42}$$

donde $\lambda_{1,n,j,m}(\mathbf{x},\mathbf{a})$ es el elemento (j,m) de la matriz $\boldsymbol{\Lambda}_{1,n}(\mathbf{x},\mathbf{a})$. Luego (4.41) y (4.42) y el hecho que $\log n/(n^{(q+1)/(2q+3)}\widetilde{h}^{d-1})\to 0$ implican que

$$\begin{aligned}
\sqrt{nh_\alpha}\sup_{\mathbf{x}\in\mathcal{S}_Q}|\widehat{D}_{2,j}(\mathbf{x})| &\le \sup_{\mathbf{x}\in\mathcal{S}_Q}\left\|\mathbf{e}_j^{\mathrm{T}}(\widehat{\mathbf{A}}_{0,n}(\mathbf{x})-\boldsymbol{\Lambda}_{1,n}(\mathbf{x},\widehat{\mathbf{a}}(\mathbf{x})))\right\| O_{\mathbb{P}}\left(\left(\frac{\log n}{\widetilde{h}^{d-1}}\right)^{1/2}\right)\\
&\quad+\sup_{\mathbf{x}\in\mathcal{S}_Q}\left\|\mathbf{e}_j^{\mathrm{T}}(\boldsymbol{\Lambda}_{1,n}(\mathbf{x},\widehat{\mathbf{a}}(\mathbf{x}))-\mathbf{A}_0(\mathbf{x}))\right\| O_{\mathbb{P}}\left(\left(\frac{\log n}{\widetilde{h}^{d-1}}\right)^{1/2}\right)\\
&\le O_{\mathbb{P}}\left(\frac{1}{\sqrt{nh_\alpha}}\frac{\log n}{\widetilde{h}^{d-1}}\right)+\sup_{\mathbf{x}\in\mathcal{S}_Q}\left\|\mathbf{e}_j^{\mathrm{T}}(\boldsymbol{\Lambda}_{1,n}(\mathbf{x},\widehat{\mathbf{a}}(\mathbf{x}))-\mathbf{A}_0(\mathbf{x}))\right\| O_{\mathbb{P}}\left(\left(\frac{\log n}{\widetilde{h}^{d-1}}\right)^{1/2}\right)\\
&\le o_{\mathbb{P}}(1)+\sup_{\mathbf{x}\in\mathcal{S}_Q}\left\|\mathbf{e}_j^{\mathrm{T}}(\boldsymbol{\Lambda}_{1,n}(\mathbf{x},\widehat{\mathbf{a}}(\mathbf{x}))-\mathbf{A}_0(\mathbf{x}))\right\| O_{\mathbb{P}}\left(\left(\frac{\log n}{\widetilde{h}^{d-1}}\right)^{1/2}\right).
\end{aligned}\tag{4.43}$$

Sea $\mathbf{A}_n(\mathbf{x}) = A_0(\psi)\int\mathcal{K}(\mathbf{u})v(\mathbf{x}+\mathbf{H}_d\mathbf{u})\check{\mathbf{u}}_\alpha\check{\mathbf{u}}_\alpha^{\mathrm{T}}\,d\mathbf{u}$ y llamamemos $A_{n,j,m}(\mathbf{x})$ su elemento (j,m). Como v y ψ'' son acotadas, λ_1 es Lipschitz y K_j tiene soporte en $[-1,1]$, usando (4.40) obtenemos, para $1\le j,m\le q+1$

$$\begin{aligned}
|\lambda_{1,n,j,m}(\mathbf{x},\widehat{\mathbf{a}}(\mathbf{x}))-\lambda_{1,n,j,m}(\mathbf{x},\mathbf{0})| &\le C_1\|\widehat{\mathbf{a}}(\mathbf{x})\|\le C_1\left\|\mathbf{H}_\alpha[\widetilde{\beta}(\mathbf{x})-\beta(\mathbf{x})\right\|\\
&\le o_{\mathbb{P}}(h_\alpha^{q+1})+O_{\mathbb{P}}\left(\left(\frac{\log n}{nh_\alpha\widetilde{h}^{d-1}}\right)^{1/2}\right).
\end{aligned}\tag{4.44}$$

Finalmente, como λ_1 es Lipschitz y $|R(\mathbf{x}+\mathbf{H}_d\mathbf{u},\mathbf{x})|\le C\left(\widetilde{h}^\ell+h_\alpha^{q+1}\right)$ pues g_j y $g_\alpha^{(\ell)}$ son continuamente diferenciables, obtenemos para $1\le j,m\le q+1$

$$\begin{aligned}
\sup_{\mathbf{x}\in\mathcal{S}_Q}|\lambda_{1,n,j,m}(\mathbf{x},0)-A_{n,j,m}(\mathbf{x})| &\le C\sup_{\mathbf{x}\in\mathcal{S}_Q}\int|\mathcal{K}(\mathbf{u})||v(\mathbf{x}+\mathbf{H}_d\mathbf{u})||R(\mathbf{x}+\mathbf{H}_d\mathbf{u},\mathbf{x})|\,d\mathbf{u}\\
&\le C_2\left(\widetilde{h}^\ell+h_\alpha^{q+1}\right)\le C_3h_\alpha^{q+1}
\end{aligned}\tag{4.45}$$

Por otra parte, como $\mathbf{A}_0(\mathbf{u}) = v(\mathbf{u})A_0(\psi)\mathbf{S}^{(\alpha)}$, $v(\mathbf{u}) = f_{\mathbf{X}}(\mathbf{u})r(\mathbf{u})$ es ℓ veces diferenciable y $\widetilde{h}^\ell = o(h_\alpha^{q+1})$, por (4.32) se tiene para

$$\begin{aligned}
\sup_{\mathbf{x}\in\mathcal{S}_Q}|A_{n,j,m}(\mathbf{x})-A_{0,j,m}(\mathbf{x})| &= |A_0(\psi)|\sup_{\mathbf{x}\in\mathcal{S}_Q}\left|\int\mathcal{K}(\mathbf{u})[v(\mathbf{x}+\mathbf{H}_d\mathbf{u})-v(\mathbf{x})]u_\alpha^{j-1}u_\alpha^{m-1}\,d\mathbf{u}\right|\\
&\le C_4\left(\widetilde{h}^\ell+h_\alpha^{q+1}\right)\le C_5h_\alpha^{q+1}
\end{aligned}\tag{4.46}$$

La cotas obtenidas en (4.44) a (4.46) nos permiten deducir que, para $1 \leq j, m \leq q+1$,

$$\sup_{\mathbf{x}\in\mathcal{S}_Q} |\lambda_{1,n,j,m}(\mathbf{x}, \widehat{\mathbf{a}}(\mathbf{x})) - A_{0,j,m}(\mathbf{x})| O_{\mathbb{P}}\left(\left(\frac{\log n}{\widetilde{h}^{d-1}}\right)^{1/2}\right) \leq h_\alpha^{q+1} O_{\mathbb{P}}\left(\left(\frac{\log n}{\widetilde{h}^{d-1}}\right)^{1/2}\right)$$

$$+\left(o_{\mathbb{P}}(h_\alpha^{q+1}) + O_{\mathbb{P}}\left(\left(\frac{\log n}{nh_\alpha \widetilde{h}^{d-1}}\right)^{1/2}\right)\right) O_{\mathbb{P}}\left(\left(\frac{\log n}{\widetilde{h}^{d-1}}\right)^{1/2}\right)$$

$$+O_{\mathbb{P}}\left(\left(\frac{\log n}{h_\alpha^{-2(q+1)}\widetilde{h}^{d-1}}\right)^{1/2}\right) + O_{\mathbb{P}}\left(\frac{1}{\sqrt{nh_\alpha}}\frac{\log n}{\widetilde{h}^{d-1}}\right)$$

lo que junto con (4.43) y el hecho que $\log n/(n^{(q+1)/(2q+3)}\widetilde{h}^{d-1}) \to 0$ concluyen la demostración de (4.39) y del Teorema. □

Corolario 4.3.2 *Sea* $\mathbf{x}$ *un punto del interior de* $\mathcal{S}_f$. *Supongamos que* $\sigma(\mathbf{x}) \equiv \sigma$ *y que* $\widehat{s}$ *es un estimador consistente de* σ *tal que* $\sqrt{n}(\widehat{s} - \sigma) \xrightarrow{p} 0$. *Sea* $\widehat{\boldsymbol{\beta}}(\mathbf{x})$ *una solución de (2.22) tal que* $\sup_{\mathbf{x}\in\mathcal{S}_Q} \|\mathbf{H}_\alpha \left[\widehat{\boldsymbol{\beta}}(\mathbf{x}) - \boldsymbol{\beta}(\mathbf{x})\right]\| \xrightarrow{p} 0$ *donde* $\boldsymbol{\beta}(\mathbf{x}) = (g(\mathbf{x}), g_\alpha^{(1)}(x_\alpha), \ldots, g_\alpha^{(q)}(x_\alpha)/q!)^{\mathrm{T}}$ *y* $g_\alpha^{(1)} = g'_\alpha$. *Supongamos que se cumplen* ***A0***, ***A1***, ***C4***, ***C5***, ***C7***, ***N1*** *y* ***N2*** *y que la función* $\lambda(a)$ *tiene derivadas hasta el orden* $\ell - 1$ *Lipschitz en un entorno de* 0. *Más aún, supongamos que* $q_{\underline{\alpha}}(\mathbf{u})$ *es continua,* $f_{\mathbf{X}}(\mathbf{u})$ *y* $r(\mathbf{u}) = p(\mathbf{u})/t(\mathbf{u})$ *son* ℓ *veces continuamente diferenciables y que* $\sup_{\mathbf{x}\in\mathcal{S}_Q} |g(\mathbf{x})| < \infty$ *y* $\sup_{\mathbf{x}\in\mathcal{S}_Q} f_{\mathbf{X}}(\mathbf{x}) < \infty$, $\inf_{\mathbf{x}\in\mathcal{C}} f_{\mathbf{X}}(\mathbf{x}) > 0$, $\inf_{\mathbf{x}\in\mathcal{C}} t(\mathbf{x}) > 0$ *y* $\inf_{\mathbf{x}\in\mathcal{C}} p(\mathbf{x}) > 0$ *para algún entorno compacto,* $\mathcal{C} \subset \mathcal{S}_f$, *de* $\mathcal{S}_Q$. *Si además,* $h_\alpha = \beta n^{-1/(2q+3)}$, $h_j = \widetilde{h}\ \forall j \neq \alpha$, *y tal que* $\widetilde{h}$ *satisface* $\widetilde{h} = o\left(n^{-(q+1)/(\ell(2q+3))}\right)$ *y* $n^{(q+1)/(2q+3)}\widetilde{h}^{d-1}/\log n \to \infty$ *se tiene que para* $\nu = 1, \ldots, q$

$$\sqrt{nh_\alpha} h_\alpha^\nu [\widehat{g}^{(\nu)}_{\alpha,\mathrm{M}_{q,\alpha}}(x_\alpha) - g_\alpha^{(\nu)}(x_\alpha)] \xrightarrow{D} N\left(b^{(\nu)}_{q,\alpha}(x_\alpha), \sigma^2_{\nu,q,\alpha}(x_\alpha)\right)$$

donde

$$\begin{aligned} b^{(\nu)}_{q,\alpha}(x_\alpha) &= \nu!\beta^{\frac{2q+3}{2}} \frac{1}{(q+1)!} g_\alpha^{(q+1)}(x_\alpha) \mathbf{e}^{\mathrm{T}}_{\nu+1} \left(\mathbf{S}^{(\alpha)}\right)^{-1} \mathbf{s}_q^{(\alpha)} \\ \sigma^2_{\nu,q,\alpha}(x_\alpha) &= \nu! \frac{\mathbb{E}(\psi^2(\varepsilon))}{(\mathbb{E}(\psi'(\varepsilon)))^2} \sigma^2 \int \frac{q^2_{\underline{\alpha}}(x_{\underline{\alpha}})}{f_{\mathbf{X}}(\mathbf{x}) p(\mathbf{x})} dx_{\underline{\alpha}} \mathbf{e}^{\mathrm{T}}_{\nu+1} (\mathbf{S}^{(\alpha)})^{-1} \mathbf{V} (\mathbf{S}^{(\alpha)})^{-1} \mathbf{e}_{\nu+1} \end{aligned}$$

con $\mathbf{s}_q^{(\alpha)}$ *como en el Teorema 4.3.2.*

Demostración. La demostración sigue argumentos análogos a los utilizados en el Teorema 4.3.2, observando que

$$\begin{aligned} \widehat{g}^{(\nu)}_{\alpha,\mathrm{M}_{q,\alpha}}(x_\alpha) - g_\alpha^{(\nu)}(x_\alpha) &= \nu! \int \mathbf{e}^{\mathrm{T}}_{\nu+1} \left[\widehat{\boldsymbol{\beta}}(x_\alpha, \mathbf{u}_{\underline{\alpha}}) - \boldsymbol{\beta}(x_\alpha, \mathbf{u}_{\underline{\alpha}})\right] q_{\underline{\alpha}}(\mathbf{u}_{\underline{\alpha}})\, d\mathbf{u}_{\underline{\alpha}} \\ &= \nu! h_\alpha^{-\nu} \int \mathbf{e}^{\mathrm{T}}_{\nu+1} \mathbf{H}_\alpha \left[\widehat{\boldsymbol{\beta}}(x_\alpha, \mathbf{u}_{\underline{\alpha}}) - \boldsymbol{\beta}(x_\alpha, \mathbf{u}_{\underline{\alpha}})\right] q_{\underline{\alpha}}(\mathbf{u}_{\underline{\alpha}})\, d\mathbf{u}_{\underline{\alpha}} \end{aligned}$$

pero multiplicando por $\nu!\mathbf{e}^{\mathrm{T}}_{\nu+1}$ en lugar de $\mathbf{e}^{\mathrm{T}}_1$ en los items a) a c). □

4.4 Apéndice

En esta Sección incluímos los Lemas cuyas conclusiones se utilizaron para probar el Teorema 4.3.1.

Lema 4.4.1. *Sea* $\mathbf{x}$ *un punto del interior de* $\mathcal{S}_f$ *y supongamos que se cumplen las hipótesis* ***C1′***, ***C2′***, ***C3*** *a* ***C5***, ***C7*** *y* ***C8***. *Sean* $V_i = V_i(\mathbf{x}) = \sigma(\mathbf{X}_i)\varepsilon_i/\sigma(\mathbf{x})$, $\lambda_1(\mathbf{u}, a) = \mathbb{E}(\psi'(\sigma(\mathbf{u})\varepsilon/\sigma(\mathbf{x})))$, $R(\mathbf{X}_i) = \sum_{j\neq\alpha}\{g_j(X_{ij}) - g_j(x_j)\} + R_\alpha(X_{i,\alpha})$ *y* $\widehat{a}_\sigma(\mathbf{x}) = \sigma(\mathbf{x})/\widehat{s}(\mathbf{x})$ *donde* $\widehat{s}(\mathbf{x}) \xrightarrow{p} \sigma(\mathbf{x})$. *Para* $0 \leq m, \tilde{m}, \bar{m} \leq 1$, $1 \leq j, k \leq q$ *y* $1 \leq \alpha \leq d$ *enteros definamos*

$$\widehat{\mathbf{D}}_n(a) = \frac{1}{n}\sum_{i=1}^{n}(\psi'(V_i a) - \lambda_1(\mathbf{X}_i, a))\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i - \mathbf{x})\xi_i R(\mathbf{X}_i)^{\bar{m}}\breve{x}_{i,j,\alpha}^{m}\breve{x}_{i,k,\alpha}^{\tilde{m}}\,.$$

Luego, se tiene que

a) $\widehat{\mathbf{C}}_n(\mathbf{x}) = \sqrt{n\prod_{j=1}^{d} h_j}(\widehat{s}(\mathbf{x})\widehat{\mathbf{D}}_n(\widehat{a}_\sigma(\mathbf{x})) - \sigma(\mathbf{x})\widehat{\mathbf{D}}_n(1)) \xrightarrow{p} 0,$

b) $\sqrt{n\prod_{j=1}^{n} h_j}\widehat{\mathbf{F}}_n(\mathbf{x}) \xrightarrow{p} 0$ *donde*

$$\widehat{\mathbf{F}}_n(\mathbf{x}) = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{\widehat{s}(\mathbf{x})}\lambda_1(\mathbf{X}_i, \widehat{a}_\sigma(\mathbf{x})) - \frac{1}{\sigma(\mathbf{x})}\lambda_1(\mathbf{X}_i, 1)\right)\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i - \mathbf{x})\xi_i R(\mathbf{X}_i)^{\bar{m}}\breve{x}_{i,j,\alpha}^{m}\breve{x}_{i,k,\alpha}^{\tilde{m}}$$

DEMOSTRACIÓN. La siguiente demostración sigue argumentos análogos a los utilizados en Boente y Fraiman (1990) y en el Lema 3.4.3. Sea $U_i = \sigma(\mathbf{X}_i)\varepsilon_i$, definamos $\psi_{1,a}(t) = a\psi'(t/a)$ y $\lambda_{1,a}(\mathbf{t}) = a\mathbb{E}(\psi'(\sigma(\mathbf{t})\varepsilon_1/a))$. Luego como $V_i = U_i/\sigma(\mathbf{x})$ y $V_i\widehat{a}_\sigma(\mathbf{x}) = U_i/\widehat{s}(\mathbf{x})$ tenemos que a) se deduce de manera análoga a la demostración del Lema 3.4.3 pero usando además que $\max_{i:\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i-\mathbf{x})}|R(\mathbf{X}_i)| \leq A_g\left(\sum_{j\neq\alpha} h_j + h_\alpha^{q+1}\right) \leq 2A_g$ si n es suficientemente grande y acotando $(X_{i\alpha} - x_\alpha)^j/h_\alpha^j$ por 1.

b) Definamos ahora definamos $\psi_{1,a}(t) = (1/a)\psi'(t/a)$ y $\lambda_{1,a}(\mathbf{t}) = (1/a)\mathbb{E}(\psi'(\sigma(\mathbf{t})\varepsilon_1/a))$. Para todo $s > 0$, sea $H_s(\mathbf{u}) = \lambda_{1,\sigma(\mathbf{x})+s}(\mathbf{u}) - \lambda_{1,\sigma(\mathbf{x})}(\mathbf{u})$, $I_s(\mathbf{u}) = \lambda_{1,\sigma(\mathbf{x})-s}(\mathbf{u}) - \lambda_{1,\sigma(\mathbf{x})}(\mathbf{u})$,

$$\begin{aligned} J_n^+(s) &= \frac{1}{\sqrt{n\prod_{j=1}^{d} h_j}}\sum_{i=1}^{n}\left\{\prod_{j=1}^{d} K_j\left(\frac{X_{ij} - x_j}{h_j}\right)\right\} H_s(\mathbf{X}_i)\xi_i R(\mathbf{X}_i)^{\bar{m}}\breve{x}_{i,j,\alpha}^{m}\breve{x}_{i,k,\alpha}^{\tilde{m}}\,, \\ J_n^-(s) &= \frac{1}{\sqrt{n\prod_{j=1}^{d} h_j}}\sum_{i=1}^{n}\left\{\prod_{j=1}^{d} K_j\left(\frac{X_{ij} - x_j}{h_j}\right)\right\} I_s(\mathbf{X}_i)\xi_i R(\mathbf{X}_i)^{\bar{m}}\breve{x}_{i,j,\alpha}^{m}\breve{x}_{i,k,\alpha}^{\tilde{m}}\,, \end{aligned}$$

$\widetilde{J}_n^+(s) = J_n^+(s) - \mathbb{E}J_n^+(s)$ y $\widetilde{J}_n^-(s) = J_n^-(s) - \mathbb{E}J_n^-(s)$. Observemos que si $s_n(\mathbf{x}) = \widehat{s}(\mathbf{x}) - \sigma(\mathbf{x}) > 0$, entonces $\widehat{\mathbf{F}}_n(\mathbf{x}) = J_n^+(s_n(\mathbf{x}))$, mientras que si $s_n(\mathbf{x}) = \widehat{s}(\mathbf{x}) - \sigma(\mathbf{x}) < 0$, $\widehat{\mathbf{F}}_n(\mathbf{x}) = J_n^-(s_n(\mathbf{x}))$. Luego,

como $s_n(\mathbf{x}) \xrightarrow{p} 0$, basta probar que

$$\lim_{n\to\infty}\lim_{d\to 0}\sup_{0\le s\le d}\mathbb{E}J_n^+(s) = 0 \tag{4.47}$$

$$\lim_{n\to\infty}\lim_{d\to 0}\sup_{0\le s\le d}\mathbb{E}J_n^-(s) = 0 \tag{4.48}$$

$$\lim_{n\to\infty}\lim_{d\to 0}P\left(\sup_{0\le s\le d}\|\widetilde{J}_n^+(s)\| > \varepsilon\right) = 0 \tag{4.49}$$

$$\lim_{n\to\infty}\lim_{d\to 0}P\left(\sup_{0\le s\le d}\|\widetilde{J}_n^-(s)\| > \varepsilon\right) = 0\,. \tag{4.50}$$

Probaremos sólo (4.47) y (4.49) ya que (4.48) y (4.50) se obtienen en forma análoga.

Para ver (4.47) observemos que

$$\mathbb{E}J_n^+(s) = \left(n\prod_{j=1}^{d}h_j\right)^{\frac{1}{2}}\int \mathcal{K}_{\mathbf{H}_d}(\mathbf{t})H_s(\mathbf{x}+\mathbf{H}_d\mathbf{t})r(\mathbf{x}+\mathbf{H}_d\mathbf{t})R(\mathbf{x}+\mathbf{H}_d\mathbf{t})t_\alpha^{jm+k\tilde{m}}f_{\mathbf{X}}(\mathbf{x}+\mathbf{H}_d\mathbf{t})\,d\mathbf{t}$$

y que

$$|H_s(\mathbf{x}+\mathbf{H}_d\mathbf{t})| \le \frac{1}{\sigma(\mathbf{x})}\frac{s}{\sigma(\mathbf{x})+s}\left[\frac{1}{\sigma(\mathbf{x})}\|\psi'\|_\infty + \|\psi''\|_\infty\right] \le C_{\mathbf{x},\psi}d$$

si $0 \le s \le d$, donde la constante $C_{\mathbf{x},\psi}$ depende de $\sigma(\mathbf{x}) > 0$ y de ψ. Acotando la integral anterior con la cota hallada, tenemos que el Teorema de Convergencia Mayorada implica que $\mathbb{E}\sup_{0\le s\le d}|J_n^+(s)| \to 0$ si $d \to 0$, por lo que $\lim_{d\to 0}\sup_{0\le s\le d}\mathbb{E}J_n^+(s) = 0$, de donde se deduce (4.47).

Para probar (4.49) basta con ver que la sucesión de variables aleatorias $\bar{J}_n^+(s)$ en el espacio $\mathcal{C}([0,1])$ es ajustada. De acuerdo al Teorema 12.3 de Billingsley (1968) es suficiente probar que

a) La sucesión $\{\widetilde{J}_n^+(0)\}$ es ajustada.

b) Existen constantes $\gamma > 0$ y $\alpha > 1$ y una función no decreciente $\bar{F}$ en $[0,1]$ tal que

$$\mathbb{E}\left(|\widetilde{J}_n^+(s_2) - \widetilde{J}_n^+(s_1)|^\gamma\right) \le \left(\bar{F}(s_2) - \bar{F}(s_1)\right)^\alpha$$

para todo $0 \le s_1 < s_2$ y n suficientemente grande.

a) vale pues $\widetilde{J}_n^+(0) = 0$.

b) Observemos que como $\lambda_{1,a}(\mathbf{u}) = (1/a)\mathbb{E}\psi'(\sigma(\mathbf{u})\varepsilon/a)$, $\zeta_2(t) = t\psi''(t)$ y vale **C7**,

$$\begin{aligned}\frac{\partial\lambda_{1,\sigma(\mathbf{x})+s}(\mathbf{u})}{\partial s} &= -\frac{1}{(\sigma(\mathbf{x})+s)^2}\mathbb{E}\left(\psi'\left(\frac{\sigma(\mathbf{u})\varepsilon}{\sigma(\mathbf{x})+s}\right)\right) - \frac{1}{\sigma(\mathbf{x})+s}\mathbb{E}\left(\psi''\left(\frac{\sigma(\mathbf{u})\varepsilon}{\sigma(\mathbf{x})+s}\right)\frac{\sigma(\mathbf{u})\varepsilon}{(\sigma(\mathbf{x})+s)^2}\right)\\ &= -\frac{1}{(\sigma(\mathbf{x})+s)^2}\left[\mathbb{E}\left(\psi'\left(\frac{\sigma(\mathbf{u})\varepsilon}{\sigma(\mathbf{x})+s}\right)\right) + \mathbb{E}\left(\zeta_2\left(\frac{\sigma(\mathbf{u})\varepsilon}{\sigma(\mathbf{x})+s}\right)\right)\right]\,.\end{aligned}$$

Por lo tanto, si $0 \leq s_1 < s_2$ y $\tilde{s}$ es un punto intermedio entre s_1 y s_2

$$\begin{aligned} |H_{s_2}(\mathbf{u}) - H_{s_1}(\mathbf{u})| &= \left| \frac{\partial \lambda_{1,\sigma(\mathbf{x})+s}(\mathbf{u})}{\partial s} \bigg|_{s=\tilde{s}} (s_2 - s_1) \right| \\ &\leq \frac{1}{\sigma^2(\mathbf{x})} (\|\psi'\|_\infty + \|\zeta_2\|_\infty)(s_2 - s_1) = \widetilde{F}(s_2) - \widetilde{F}(s_1) \end{aligned}$$

donde $\widetilde{F}(s) = \sigma^{-2}(\mathbf{x})(\|\psi'\|_\infty + \|\zeta_2\|_\infty)s$.

Denotemos por

$$A_{i,n} = \prod_{\ell=1}^{d} K_\ell \left(\frac{X_{i\ell} - x_\ell}{h_\ell} \right) (H_{s_2}(\mathbf{X}_i) - H_{s_1}(\mathbf{X}_i)) \xi_i R(\mathbf{X}_i)^{\bar{m}} \breve{x}_{i,j,\alpha}^{m} \breve{x}_{i,k,\alpha}^{\tilde{m}} \,.$$

Recordemos que $\breve{x}_{i,j,\alpha}^{m} \breve{x}_{i,k,\alpha}^{\tilde{m}} \leq 1$ si $\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i - \mathbf{x}) \neq 0$. Por otro lado, si $n \geq n_0$, tenemos que $\{\mathbf{u} : |u_j - x_j| \leq h_j\} \subset \mathcal{C}$, por lo tanto, si $\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i - \mathbf{x}) \neq 0$ y $n \geq n_0$, $\mathbf{X}_i \in \mathcal{C}$, de donde $\xi_i \leq 1/t(\mathbf{X}_i) \leq 1/i(t)$. Si además $n \geq n_1$, $\max_{i:\mathcal{K}_{\mathbf{H}_d}(\mathbf{X}_i - \mathbf{x})} |R(\mathbf{X}_i)| \leq A_g \left(\sum_{j \neq \alpha} h_j + h_\alpha^{q+1} \right) \leq 2A_g$. Por lo tanto, como K_j es acotada, si llamamos $C = 2A_g/i(t) \prod_{\ell=1}^{d} \|K_\ell\|_\infty$

$$|A_{i,n}| \leq C(\widetilde{F}(s_2) - \widetilde{F}(s_1)) \,.$$

Por otra parte, como

$$\mathbb{E}\left[\prod_{\ell=1}^{d} K_\ell \left(\frac{X_{i\ell} - x_\ell}{h_\ell} \right) \right] = \prod_{\ell=1}^{d} h_\ell \int \prod_{\ell=1}^{d} K_\ell(u_\ell) f_{\mathbf{X}}(\mathbf{x} + \mathbf{H}_d \mathbf{u}) \, d\mathbf{u} \,,$$

donde $\mathbf{y} = \mathbf{x} + \mathbf{H}_d \mathbf{u} \in \mathcal{C}$ si $n \geq n_1$, usando que $f_{\mathbf{X}}$ es acotada en $\mathcal{C}$ y $\int K_j(u) \, du = 1$ tenemos que

$$\begin{aligned} \mathbb{E}(A_{i,n}^2) &\leq \left(\widetilde{F}(s_2) - \widetilde{F}(s_1) \right)^2 \mathbb{E}\left[\prod_{\ell=1}^{d} K_\ell^2 \left(\frac{X_{1\ell} - x_\ell}{h_\ell} \right) \xi_1^2 R(\mathbf{X}_1)^{2\bar{m}} \breve{x}_{1,j,\alpha}^{2m} \breve{x}_{1,k,\alpha}^{2\tilde{m}} \right] \\ &\leq \left(\widetilde{F}(s_2) - \widetilde{F}(s_1) \right)^2 \frac{4A_g^2}{i^2(t)} \prod_{\ell=1}^{d} \|K_\ell\|_\infty \mathbb{E}\left[\prod_{\ell=1}^{d} K_\ell \left(\frac{X_{1\ell} - x_\ell}{h_\ell} \right) \right] \\ &\leq C_1 \left(\widetilde{F}(s_2) - \widetilde{F}(s_1) \right)^2 \prod_{\ell=1}^{d} h_\ell \end{aligned}$$

donde $C_1 = 4A_g^2/i^2(t) \prod_{\ell=1}^{d} \|K_\ell\|_\infty \sup_{\mathbf{u} \in \mathcal{C}} f_{\mathbf{X}}(\mathbf{u})$ para todo $n \geq \max\{n_0, n_1\} = n_2$.

Finalmente, por la independencia de los $A_{i,n}$, tenemos que

$$\begin{aligned} \mathbb{E}(\widetilde{J}_n^+(s_2) - \widetilde{J}_n^+(s_1))^2 &= \frac{1}{n \prod_{j=1}^{d} h_j} \mathrm{VAR}(\sum_{i=1}^{n} A_{i,n}) = \frac{1}{n \prod_{j=1}^{d} h_j} \sum_{i=1}^{n} \mathrm{VAR} A_{i,n} \\ &\leq \frac{1}{n \prod_{j=1}^{d} h_j} \sum_{i=1}^{n} \mathbb{E} A_{i,n}^2 \leq C_1 (\widetilde{F}(s_2) - \widetilde{F}(s_1))^2 \end{aligned}$$

para $n \geq n_2$, con lo cual tomando $\bar{F}(s) = \sqrt{C_1} \widetilde{F}(s)$ queda demostrado b). $\square$

4.5 Comentarios Generales

Además de que este trabajo complementa trabajos existentes cuando no hay datos faltantes, en el caso de datos faltantes, el método de integración marginal, a diferencia del método que promedia sobre las covariables el estimador inicial, está asociado al Principio de Transferencia descripto en Koul *et al.* (2012).

Consideremos entonces el estimador que, en lugar de considerar la integral respecto de una medida q_α, considera los promedios. Queremos saber si al ignorar los vectores cuyas respuestas son faltantes, es decir, al quedarnos con la submuestra completa, seguimos obteniendo estimadores consistentes de los efectos marginales. Llamemos $\{\mathbf{Z}_{i_1}, \ldots, \mathbf{Z}_{i_N}\} = \{(\mathbf{X}_i^{\mathrm{T}}, Y_i)^{\mathrm{T}}_{\{i:\delta_i=1\}}\}$ donde $N = \sum_{i=1}^n \delta_i$. Usando estas observaciones, los estimadores de las componentes aditivas basados en el procedimiento de promediar el estimador inicial sobre las covariables están dados por

$$\widehat{g}_{c,\mathrm{M},\alpha}(x_\alpha) = \frac{1}{N}\sum_{j=1}^{N} \widetilde{g}_{c,\mathrm{M}}(x_\alpha, \mathbf{X}_{\underline{\alpha},i_j}) - \widehat{\mu}_c = \frac{1}{\sum_{i=1}^n \delta_i}\sum_{i=1}^{n} \delta_i \widetilde{g}_{c,\mathrm{M}}(x_\alpha, \mathbf{X}_{\underline{\alpha},i}) - \widehat{\mu}_c$$

donde $\widehat{\mu}_c = (1/N)\sum_{j=1}^N \widetilde{g}_{c,\mathrm{M}}(\mathbf{X}_{i_j}) = (1/\sum_{i=1}^n \delta_i)\sum_{i=1}^n \delta_i \widetilde{g}_{c,\mathrm{M}}(\mathbf{X}_i)$ y $\widetilde{g}_{c,\mathrm{M}}$ es el estimador que usa sólo la muestra completa para calcular el estimador inicial y, por lo tanto, corresponde a cualquiera de los estimadores robustos propuestos en el Capítulo 2. Hemos indicado con un subíndice c al hecho de usar sólo la muestra *completa* para calcular el estimador. Observemos además que para el cálculo de estimador preliminar $\widetilde{g}_{\mathrm{M}}$ sólo hemos usado los datos completos, por lo que $\widetilde{g}_{c,\mathrm{M}} = \widetilde{g}_{\mathrm{M}}$.

Tal como ya lo hemos mencionado, bajo condiciones de regularidad hemos probado que $\widetilde{g}_{c,\mathrm{M}}(\mathbf{x}) \xrightarrow{c.s.} g(\mathbf{x})$, es decir, que los estimadores de la función de regresión basados sólo en las observaciones de la muestra completa son fuertemente consistentes. Más aún, hemos probado también que todas las propuestas de estimadores preliminares $\widetilde{g}_{c,\mathrm{M}}$ son uniformemente fuertemente consistentes. Luego, como la función g satisface el modelo aditivo (1.1), se puede ver que

$$\frac{1}{N}\sum_{j=1}^{N} \widetilde{g}_{c,\mathrm{M}}(x_\alpha, \mathbf{x}_{\underline{\alpha},i_j}) \xrightarrow{c.s.} \mu + g_\alpha(x_\alpha) + \frac{1}{p}\sum_{j=1, j\neq\alpha}^{d} \mathbb{E}\left\{p(\mathbf{X})g_j(X_j)\right\},$$

con $p = \mathbb{E}p(\mathbf{X})$. Por otro lado, como $\widehat{\mu}_c \xrightarrow{c.s.} p^{-1}\mathbb{E}\left\{p(\mathbf{X})g(\mathbf{x})\right\} = \mu + p^{-1}\sum_{j=1}^d \mathbb{E}\left\{p(\mathbf{X})g_j(X_j)\right\}$, tenemos que

$$\widetilde{g}_{c,\mathrm{M},\alpha}(x_\alpha) \xrightarrow{c.s.} g_\alpha(x_\alpha) - \frac{1}{p}\mathbb{E}\left\{p(\mathbf{X})g_\alpha(X_\alpha)\right\},$$

por lo que los estimadores son asintóticamente sesgados. Por lo tanto, el Principio de Transferencia no puede ser utilizado para estimar los efectos marginales a través del procedimiento basado en promediar el estimador preliminar.

Observemos que los resultados del Capítulo 3 muestran que el procedimiento de integración marginal utilizando la función $q_{\underline{\alpha}}$ en lugar de usar el promedio sobre las direcciones de no interés $\mathbf{X}_{\underline{\alpha}}$, resuelve el problema aunque los funcionales que utilizamos no son lineales. Para ser más precisos y suponiendo μ conocido e igual a cero por simplicidad, los estimadores introducidos en esta tesis pueden escribirse como $\int \widetilde{g}_{c,\mathrm{M}}(x_\alpha, \mathbf{u}_{\underline{\alpha}})q_{\underline{\alpha}}(\mathbf{u}_{\underline{\alpha}})d\mathbf{u}_{\underline{\alpha}}$ si tomarámos $q_{\underline{\alpha}} = f_{\mathbf{X}_{\underline{\alpha}}}$ que es desconocida obtenemos que los resultados obtenidos permiten obtener la consistencia y distribución

asintótica del estimador $\int \widetilde{g}_{c,\mathrm{M}}(x_\alpha, \mathbf{u}_{\underline{\alpha}}) f_{\mathbf{X}_{\underline{\alpha}}}(\mathbf{u}_{\underline{\alpha}}) d\mathbf{u}_{\underline{\alpha}}$. Ahora bien, como $f_{\mathbf{X}_{\underline{\alpha}}}$ es desconocida esta última integral no se puede calcular, pero podemos aproximarla por $(1/n)\sum_{i=1}^{n} \widetilde{g}_{c,\mathrm{M}}(x_\alpha, \mathbf{x}_{\underline{\alpha},i})$ lo que explica porqué es necesario promediar sobre todas las covariables para tener estimadores consistentes y no sólo sobre las covariables correspondientes a las respuestas existentes. La consistencia del estimador $(1/n)\sum_{i=1}^{n} \widetilde{g}_{c,\mathrm{M}}(x_\alpha, \mathbf{x}_{\underline{\alpha},i})$ se deduce en forma inmediata de los resultados de convergencia uniforme dados en el Capítulo 3, mientras que argumentos análogos a los considerados en este Capítulo permiten obtener su distribución asintótica que coincide con la del estimador $\int \widetilde{g}_{c,\mathrm{M}}(x_\alpha, \mathbf{u}_{\underline{\alpha}}) f_{\mathbf{X}_{\underline{\alpha}}}(\mathbf{u}_{\underline{\alpha}}) d\mathbf{u}_{\underline{\alpha}}$. Por esta razón, todos los resultados se presentan en el caso de respuestas faltantes y no solamente en el de observaciones sin presencia de datos ausentes.

Capítulo 5

Estimadores robustos basados en backfitting

5.1 Introducción

El procedimiento de backfitting fue originalmente propuesto por Buja, Hastie y Tibshirani (1989) y Hastie y Tibshirani (1990) y es ampliamente aceptado para estimar los modelos aditivos por su inmediata implementación y su intuitiva definición. Si bien tanto en estudios de simulación como con datos reales tiene un muy buen comportamiento, debido a la naturaleza del procedimiento iterativo que lo subyace, sus propiedades estadísticas teóricas son muy difíciles de analizar. Debido a este inconviente es que los procedimientos de integración marginal han sido más prometedores y han atraído la atención de los investigadores. Hoy en día, es sabido que el procedimiento de backfitting y el de integración marginal estiman diferentes cosas cuando el modelo no es exactamente el aditivo. Mientras que el procedimiento de backfitting busca el ajuste óptimo en el espacio de los modelos aditivos, el procedimiento de integración marginal estima los *efectos marginales*, ver, por ejemplo, Sperlich, Linton y Härdle (1999) y Sperlich, Tjøstheim y Yang (2002).

El algoritmo de backfitting considera el problema de hallar la proyección de g en el subespacio de las funciones aditivas en L_2. En términos muestrales, este algoritmo requiere resolver un sistema de ecuaciones normales de dimensión $nd \times nd$, y para ello se usa el algoritmo de backfitting o *Gauss–Seidel*. Este algoritmo converge muy rápido pero tiene, en comparación con la solución directa de un sistema de ecuaciones lineales grande, la pequeña desventaja de que tiene una matriz estimada más complicada. Respecto a las propiedades teóricas del algoritmo de backfitting, Opsomer y Ruppert (1997) presentaron resultados sobre el sesgo y varianza condicional para la versión 2–dimensional con núcleos de este estimador que fue extendida al caso $d > 2$ por Opsomer (2000). Estos autores observaron que el estimador de backfitting clásico no alcanza la cota de la eficiencia oracle, es decir, las componentes aditivas no son estimadas con la misma precisión que si las otras componentes fueran conocidas, o sea, el sesgo depende fuertemente de las otras direcciones y el estimador se *rompe* en el sentido de perder la tasa de convergencia univariada, en presencia de correlaciones moderadas entre las covariables. Esto es una consecuencia del hecho de que el comportamiento asintótico de una componente estimada con el procedimiento de backfitting depende tanto de las otras componentes como del diseño de las variables dependientes involucradas. Más

aún, el procedimiento de backfitting clásico requiere de condiciones bastante fuertes, entre ellas, la correlación limitada entre las variables regresoras. Esta es una de las principales desventajas del método de backfitting que sin embargo, continua siendo ampliamente utilizado.

En un extenso estudio de simulación Sperlich, Linton y Härdle (1999) muestran que el estimador de la integración marginal sufre más que el basado en backfitting frente al esparcimiento de las observaciones. El método de backfitting parece funcionar en ese estudio mejor en los puntos de frontera y, a pesar de los resultados de Opsomer (2000), cuando existe una alta correlación entre las covariables. Por otra parte, el método de integración marginal funciona mejor en la mayoría de los demás casos y, especialmente, en la estimación de las componentes aditivas en oposición a la propia función de regresión.

Tal como ocurre en el caso de modelos de regresión lineal, es fácil ver que aquellos estimadores de las componentes aditivas que están basados en la función de pérdida cuadrática se ven severamente afectados por una pequeña proporción de outliers. En este capítulo propondremos estimadores robustos para los modelos aditivos basados en el algoritmo de backfitting cuando no hay respuestas faltantes.

5.2 El funcional de Backfitting

Supongamos que tenemos $(\mathbf{X}^{\mathrm{T}}, Y)^{\mathrm{T}} \sim P$ con $Y \in \mathbb{R}$ y $\mathbf{X} = (X_1, \ldots, X_d)^{\mathrm{T}}$ que satisfacen el modelo de regresión aditivo (1.1) con la condición de identificabilidad (2.1), es decir,

$$Y = g(\mathbf{x}) + U = \mu + \sum_{j=1}^{d} g_j(X_j) + \sigma(\mathbf{X})\varepsilon \qquad \mathbb{E}g_j(X_j) = 0 \quad 1 \leq j \leq d\,, \tag{5.1}$$

donde el error ε independiente de las covariables $X_j's$, $\mu \in \mathbb{R}$ para $1 \leq j \leq d$, $g_j : \mathbb{R} \to \mathbb{R}$ son funciones suaves que cumple (2.1) como se indica.

Cuando $\sigma(\mathbf{X}) \equiv \sigma$, el funcional asociado a los estimadores clásicos de backfitting para el modelo aditivo (5.1), bajo las condiciones de identificabilidad consideradas en (2.1) resuelve el siguiente problema de minimización

$$\min_{(\nu,m)\in\mathbb{R}\times\mathcal{H}^{ad}} \mathbb{E}\left(Y - \nu - \sum_{j=1}^{d} m_j(X_j)\right)^2 \tag{5.2}$$

donde $m(\mathbf{x}) = \sum_{j=1}^{d} m_j(x_j)$,

$$\mathcal{H}^{ad} = \left\{ m(\mathbf{x}) = \sum_{j=1}^{d} m_j(x_j), m_j \in \mathcal{H}_j \right\} \subset \mathcal{H} = \{r(\mathbf{x}) : \mathbb{E}(r(\mathbf{X})) = 0 \text{ y } \mathbb{E}(r^2(\mathbf{X})) < \infty\}\,,$$

con $\mathcal{H}_j$ el espacio de Hilbert de las funciones medibles $f(X_j)$ con media cero y segundo momento finito, o sea, tales que $\mathbb{E}f(X_j) = 0$ y $\mathbb{E}f^2(X_j) < \infty$. Es fácil ver que la solución a este problema está caracterizada por residuos $Y - \mu - g(\mathbf{X})$ que son ortogonales al espacio de ajuste, o sea, como

$\mathcal{H}^{ad}$ está generado por $\mathcal{H}_j$ se satisface

$$\begin{cases} \mathbb{E}\left(Y-\mu-\sum_{j=1}^d g_j(X_j)\right) & = 0\,, \\ \mathbb{E}\left(Y-\mu-\sum_{j=1}^d g_j(X_j)\middle| X_\ell\right) & = 0\,, \quad 1\le \ell\le d\,. \end{cases}$$

Luego, tenemos que $\mu = \mathbb{E}(Y)$ y además, para $1 \le j \le d$, se cumple que $g_\ell(X_\ell) = \mathbb{E}(Y - \mu - \sum_{j\neq \ell} g_j(X_j)|X_\ell)$. El algoritmo de backfitting clásico encuentra estimadores para la solución del sistema de ecuaciones anterior, reemplazando la esperanza condicional por un suavizado basado en polinomios locales.

Tal como ya hemos mencionado a la hora de hallar estimadores robustos basados en el procedimiento de integración marginal, para obtener estimadores robustos de μ y de las funciones g_j en (5.1) proponemos reemplazar la función cuadrática en (5.2) por una $\rho-$función que penalice aquellos residuos grandes. Más específicamente, dada $m \in \mathcal{M}$ con $\mathcal{M} = \{m(\mathbf{x}) = \sum_{j=1}^d m_j(x_j), m_j \in \mathcal{M}_j\}$ y $\mathcal{M}_j$ es el conjunto de las funciones medibles $f(X_j)$ con $\mathbb{E}f(X_j) = 0$, definamos

$$\Upsilon(\nu, m) = \mathbb{E}\,\rho\left(\frac{Y-\nu-\sum_{j=1}^d m_j(X_j)}{\sigma(\mathbf{X})}\right)\,.$$

Consideraremos estimadores del funcional $(\mu(P), g(P))$ definido como la solución del siguiente problema de optimización

$$(\mu(P), g(P)) = \underset{(\nu,m)\in\mathbb{R}\times\mathcal{M}}{\operatorname{argmin}}\ \mathbb{E}\,\rho\left(\frac{Y-\nu-\sum_{j=1}^d m_j(X_j)}{\sigma(\mathbf{X})}\right)\,. \tag{5.3}$$

5.2.1 Fisher–consistencia y descripción del algoritmo

Para obtener los resultados de Fisher consistencia del funcional definido por (5.3) y encontrar un sistema de ecuaciones funcional análogo al que permite definir los estimadores de backfitting clásicos para el enfoque de backfitting robusto, deberemos suponer la siguiente hipótesis adicional sobre los errores que ya fue considerada, por ejemplo, en Maronna, Martin y Yohai (2006) para asegurar la Fisher–consistencia del funcional de posición.

B1 La variable aleatoria ε tiene función de densidad f_0 que es par, nocreciente en $|t|$ y estrictamente decreciente en $|t|$ en un entorno del 0.

Además, la función $\rho : \mathbb{R} \to \mathbb{R}$ es una función ρ como se define en Maronna, Martin y Yohai (2006, Capítulo 2), es decir, ρ satisface:

B2 ρ es continua, $\rho(0) = 0$, $\rho(u) = \rho(-u)$, $0 \le u \le v < c = \operatorname{argsup}_t \rho(t)$ implica que $\rho(u) < \rho(v)$.

Por otra parte, necesitamos la siguiente condición sobre las covariables que permite escribir la representación (5.3) en forma única

B3 Dadas funciones $r_j \in \mathcal{M}_j$, si $\mathbb{P}(\sum_{j=1}^d r_j(X_j) = 0) = 1$ entonces, para todo $1 \leq j \leq d$, se tiene $\mathbb{P}(r_j(X_j) = 0) = 1$

El siguiente teorema muestra que bajo las condiciones mencionadas anteriormente, la única solución de (5.3) está dada por μ y $\sum_{j=1}^d g_j$, en (5.1). A esta propiedad se la denomina Fisher–consistencia.

Teorema 5.2.1. *Supongamos que el vector aleatorio* $(\mathbf{X}^{\mathrm{T}}, Y)^{\mathrm{T}} \in \mathbb{R}^{d+1}$ *cumple el modelo (5.1) y es tal que* $(\mathbf{X}^{\mathrm{T}}, Y)^{\mathrm{T}} \sim P$.

*a) Si se satisfacen **B1** y **B2**, se cumple que* $\Upsilon(\nu, m)$ *alcanza su único mínimo en* $(\nu, m) \in \mathbb{R} \times \mathcal{M}$ *cuando* $\nu = \mu$ *y* $\mathbb{P}(m(\mathbf{X}) = \sum_{j=1}^d g_j(X_j)) = 1$.

*b) Si además se cumple **B3**, el único mínimo se alcanza cuando* $\nu = \mu$ *y* $\mathbb{P}(m_j(X_j) = g_j(X_j)) = 1$, $1 \leq j \leq d$, *es decir,* $\mathbb{P}((\mu(P), \mathbf{g}(P)(\mathbf{X})) = (\mu, \mathbf{g}(\mathbf{X}))) = 1$ *donde* $\mathbf{g}(P) = (g_1(P), \ldots, g_d(P))^{\mathrm{T}}$ *y* $\mathbf{g} = (g_1, \ldots, g_d)^{\mathrm{T}}$.

DEMOSTRACIÓN. Sea $(\nu, m) \in \mathbb{R} \times \mathcal{M}$ tal que $\nu \neq \mu$ o $\mathbb{P}(\sum_{j=1}^d m_j(X_j) \neq \sum_{j=1}^d g_j(X_j)) > 0$. Luego

$$\begin{aligned}
\mathbb{E}\rho\left(\frac{Y - \nu - \sum_{j=1}^d m_j(X_j)}{\sigma(\mathbf{X})}\right) &= \mathbb{E}\rho\left(\frac{\sigma(\mathbf{X})\varepsilon - \left[\nu - \mu + \sum_{j=1}^d (m_j(X_j) - g_j(X_j))\right]}{\sigma(\mathbf{X})}\right) \\
&= \mathbb{E}_{\mathbf{X}}\mathbb{E}\left\{\rho\left(\varepsilon - a(\mathbf{X})\right)|\mathbf{X}\right\},
\end{aligned}$$

donde $a(\mathbf{x}) = b(\mathbf{x})/\sigma(\mathbf{x})$ con $b(\mathbf{x}) = \nu - \mu + \sum_{j=1}^d (m_j(x_j) - g_j(x_j))$. Como ε es independiente de $\mathbf{X}$, $\mathbb{E}\left\{\rho\left(\varepsilon - a(\mathbf{X})\right)|\mathbf{X} = \mathbf{x}\right\} = \mathbb{E}_{F_0}\left\{\rho\left(\varepsilon - a(\mathbf{x})\right)\right\}$. Por lo tanto, alcanza con probar que

$$\mathbb{E}_{F_0}\rho(\varepsilon) < \mathbb{E}_{F_0}\rho(\varepsilon - a) \quad \forall a \neq 0\,. \tag{5.4}$$

$$\mathbb{P}\left(b(\mathbf{X}) = 0\right) < 1\,, \tag{5.5}$$

donde F_0 indica la función de distribución de ε.

Para mostrar (5.5), observemos que si $\mathbb{P}(b(\mathbf{X}) = 0) = 1$ entonces $\mathbb{E}b(\mathbf{X}) = 0$ implica que $\nu = \mu$ pues $\mathbb{E}m_j(X_j) = 0$ para todo j y g_j cumple (2.1).

Por lo tanto, la condición $\mathbb{P}(b(\mathbf{X}) = 0) = 1$ es equivalente a $\mathbb{P}(\sum_{j=1}^d r_j(X_j) = 0) = 1$ donde $r_j = m_j - g_j \in \mathcal{M}_j$, lo que contradice nuestra hipótesis.

Para probar b), supongamos que $\mathbb{P}(\sum_{j=1}^d m_j(X_j) \neq \sum_{j=1}^d g_j(X_j)) > 0$ para algún $1 \leq j \leq d$. Observemos que como la condición $\mathbb{P}(b(\mathbf{X}) = 0) = 1$ es equivalente a $\mathbb{P}(\sum_{j=1}^d r_j(X_j) = 0) = 1$, donde $r_j = m_j - g_j \in \mathcal{M}_j$, usando **B3** obtenemos que $\mathbb{P}(m_j(X_j) = g_j(X_j)) = 1$ para todo $1 \leq j \leq d$. Por lo tanto, si $\nu \neq \mu$ o si $\mathbb{P}(m_j(X_j) \neq g_j(X_j)) > 0$, para algún $1 \leq j \leq d$ se cumple (5.5).

Para ver (5.4), alcanza con ver que $g(\lambda) = \mathbb{E}_{F_0}\rho(\varepsilon - \lambda)$ tiene un único mínimo en $\lambda = 0$. Esto se obtiene gracias a las hipótesis **B1** y **B2**, usando el Lema 3.1 de Yohai (1985). Más específicamente, observemos que, para cualquier $\lambda \neq 0$, la función de distribución F_λ de $\varepsilon_\lambda = |\varepsilon - \lambda|$ satisface:

i. $F_\lambda(u) \leq F_0(u)$ para toda $u > 0$, es decir, $\varepsilon_0 \leq_s \varepsilon_\lambda$, y

ii. existe $\delta > 0$ tal que $F_\lambda(u) < F_0(u)$ para $0 \leq u \leq \delta$,

donde la notación $U \leq_s V$ significa que la variable aleatoria U es estocásticamente menor que V. Luego, se obtiene la desigualdad en (5.4) porque ρ es nodecreciente en $|u|$ y escrictamente creciente en un entorno del 0.

Finalmente, usando (5.5) y (5.4) y si indicamos por $\mathcal{A}_0 = \{\mathbf{X} : a(\mathbf{X}) = 0\}$, tenemos que

$$\begin{aligned}\mathbb{E}\rho\left(\frac{Y-\mu-\sum_{j=1}^d g_j(X_j)}{\sigma(\mathbf{X})}\right) &= \int_{\mathcal{A}_0} \mathbb{E}_{F_0}\rho\left(\varepsilon - a(\mathbf{x})\right)\, dF_{\mathbf{X}}(\mathbf{x}) + \int_{\mathcal{A}_0^c} \mathbb{E}_{F_0}\rho\left(\varepsilon - a(\mathbf{x})\right)\, dF_{\mathbf{X}}(\mathbf{x}) \\ &> \int_{\mathcal{A}_0} \mathbb{E}_{F_0}\rho(\varepsilon)\, dF_{\mathbf{X}}(\mathbf{x}) + \int_{\mathcal{A}_0^c} \mathbb{E}_{F_0}\rho(\varepsilon)\, dF_{\mathbf{X}}(\mathbf{x}) = \mathbb{E}_{\mathbf{X}}\mathbb{E}_{\varepsilon|\mathbf{X}}\rho(\varepsilon) = \mathbb{E}\rho(\varepsilon)\,,\end{aligned}$$

y la demostración finaliza observando que $\mathbb{E}\rho(\varepsilon) = \mathbb{E}\rho\left((Y-\mu-\sum_{j=1}^d g_j(X_j))/\sigma(\mathbf{X})\right)$. □

Observación 5.2.1. Claramente la condición **B3** se cumple si las componentes del vector $\mathbf{X}$ son independientes. Efectivamente, indiquemos por $(x, \mathbf{X}_{\underline{\alpha}})$ al vector con la coordenada α igual a x y las demás coordenadas iguales a X_j, $j \neq \alpha$ y por $r(\mathbf{x}) = \sum_{j=1}^d r_j(x_j)$. Luego, si fijado $1 \leq \alpha \leq d$, integramos respecto de la función de distribución conjunta de $\mathbf{X}_{\underline{\alpha}}$ dejando fija la coordenada α igual al valor x_α tenemos que el hecho que $\mathbb{E}r_j(X_j) = 0$ para todo j, implican que

$$\int r(x_\alpha, \mathbf{u}_{\underline{\alpha}}) dF_{\mathbf{X}_{\underline{\alpha}}}(\mathbf{u}) = r_\alpha(x_\alpha) + \int \sum_{j\neq\alpha} r_j(u_j) dF_{\mathbf{X}_{\underline{\alpha}}}(\mathbf{u}) = r_\alpha(x_\alpha)$$

Por otra parte, la condición $\mathbb{P}(r(\mathbf{X}) = 0) = 1$ implica que para casi todo x_α, $\mathbb{P}(r(x_\alpha, \mathbf{X}_{\underline{\alpha}}) = 0 | X_\alpha = x_\alpha) = 1$. Luego como las componentes de $\mathbf{X}$ son independientes, tenemos que $\mathbb{P}(r(x_\alpha, \mathbf{X}_{\underline{\alpha}}) = 0) = 1$ con lo cual, $\int r(x_\alpha, \mathbf{u}_{\underline{\alpha}}) dF_{\mathbf{X}_{\underline{\alpha}}}(\mathbf{u}) = 0$ de donde $r_\alpha(x_\alpha) = 0$, para casi todo x_α, como queríamos.

En situaciones de dependencia de la componentes de $\mathbf{X}$, la condición $\mathbb{P}(r(x_\alpha, \mathbf{X}_{\underline{\alpha}}) = 0 | X_\alpha = x_\alpha) = 1$ no implica $\int r(x_\alpha, \mathbf{u}_{\underline{\alpha}}) dF_{\mathbf{X}_{\underline{\alpha}}}(\mathbf{u}) = 0$. Esta situación es análoga al comentario realizado en Hastie y Tibshirani (1986) (ver también, Hastie y Tibshirani, 1990, pág. 103) en el el caso de pérdida cuadrática. Efectivamente, estos autores mencionan que el espacio $\mathcal{H}^{ad}$ es cerrado en $\mathcal{H}$ bajo ciertos supuestos. Sin embargo, a pesar de que el mínimo de $\mathbb{E}(Y - r(\mathbf{x}))^2$ sobre $\mathcal{H}^{ad}$ existe y es único, las funciones individuales $r_j(x_j)$ pueden no estar univocamente determinadas ya que la dependencia entre las covariables pueden llevar a más de una representación para la misma superficie (ver también Breiman y Friedman, 1985). De hecho la condición **B3** es análogo al supuesto 5.1 de Breiman y Friedman (1985). El algoritmo de backfitting clásico identifica esas componentes a través de la condición $g_\alpha(X_\alpha) = \mathbb{E}(Y - \mu - \sum_{j\neq\alpha} g_j(X_j) | X_\alpha)$, el Teorema 5.2.2 da una identificiación análoga para el funcional robusto cuando $\sigma(\mathbf{X}) = \sigma$.

Vale la pena mencionar que Stone (1985) da condiciones para que **B3** valga. Efectivamente, el Lema 1 en Stone (1985) permite deducir la siguiente Proposición que da condiciones débiles para que la representación sea única y por lo tanto, para que el funcional $g_j(P)$ sea Fisher–consistente.

Proposición 5.2.1. *Supongamos que* $\mathbf{X}$ *tiene soporte compacto* $\mathcal{S}_f$*, que por simplicidad supondremos* $\mathcal{S}_f = [0,1]^d$*. Más aún, supongamos que* $\mathbf{X}$ *tiene densidad* $f_{\mathbf{X}}$ *acotada en* $\mathcal{S}_f$ *y tal que*

$\inf_{\mathbf{x}\in\mathcal{S}_f} f_{\mathbf{X}}(\mathbf{x}) > 0$. *Luego, si* $V_j = r_j(X_j)$ *son variables aleatorias tales que* $\mathbb{P}(\sum_{j=1}^d V_j = 0) = 1$ *y* $\mathbb{E}(V_j) = 0$ *entonces* $\mathbb{P}(V_j = 0) = 1$.

Aunque no pidamos la condición **B1**, es siempre posible encontrar un minimizador $(\mu(P), g(P))$ de (5.3) si la función ρ es estrictamente convexa. Si además se cumple **B3**, el minimizador tendrá una única representación. El siguiente teorema muestra que podemos elegir la solución $g(P)$ de (5.3) de modo que sus componentes $g_j = g_j(P)$ satisfagan el sistema de ecuaciones

$$\mathbf{\Gamma}(\mu, \mathbf{g}(P), \mathbf{x}) = \mathbf{0} \tag{5.6}$$

donde $\mathbf{g}(P) = (g_1(P), \ldots, g_d(P))^{\mathrm{T}}$ y $\mathbf{\Gamma}(\nu, \mathbf{m}, \mathbf{x}) = (\Gamma_0(\nu, \mathbf{m}), \Gamma_1(\nu, \mathbf{m}, x_1), \ldots, \Gamma_d(\nu, \mathbf{m}, x_d))^{\mathrm{T}}$ con $\mathbf{m} = (m_1, \ldots, m_d)^{\mathrm{T}}$

$$\begin{aligned}
\Gamma_0(\nu, \mathbf{m}) &= \mathbb{E}\left[\psi\left(\frac{Y - \nu - \sum_{j=1}^d m_j(X_j)}{\sigma}\right)\right] \\
\Gamma_\ell(\nu, \mathbf{m}, x_\ell) &= \mathbb{E}\left[\psi\left(\frac{Y - \mu - \sum_{j\neq\ell} g_j(X_j) - m_\ell(X_\ell)}{\sigma}\right)\middle| X_\ell = x_\ell\right] \quad 1 \le \ell \le d\,.
\end{aligned}$$

Teorema 5.2.2. *Sea* ρ *una función derivable que satisface* ***B2*** *y tal que su derivada* $\rho' = \psi$ *es estrictamente creciente, acotada y continua con* $\lim_{t\to+\infty}\psi(t) > 0$ *y* $\lim_{t\to-\infty}\psi(t) < 0$, *de manera tal que, para cualquier distribución* F *y* $V \sim F$, *existe un único* $t \in \mathbb{R}$ *tal que* $\mathbb{E}_F\psi(V - t) = 0$. *Supongamos que* $\sigma(\mathbf{X}) \equiv \sigma$ *y que se cumple* ***B3***. *Supongamos que existe una solución de (5.3), o sea, llamaremos* $\mu(P) \in \mathbb{R}$, $g(P) = \sum_{j=1}^d g_j(P) \in \mathcal{M}$ *a los minimizadores de* $\Upsilon(\nu, m)$ *sobre* $\mathbb{R}\times\mathcal{M}$, *es decir, los mejores aproximantes aditivos robustos de la respuesta* Y. *Luego, si (5.6) tiene una solución notada* $\mathbf{m}_{\mathbf{\Gamma}} = (m_{1,\mathbf{\Gamma}}, \ldots, m_{d,\mathbf{\Gamma}})$ *tal que* $\mathbb{E}m_{j,\mathbf{\Gamma}}(X_j) = 0$, *podemos elegir* $g_j(P)$ *de modo tal que* $\mathbf{g}(P) = (g_1(P), \ldots, g_d(P))^{\mathrm{T}}$ *satisfaga (5.6).*

DEMOSTRACIÓN. Por simplicidad de notación, sea $\mu = \mu(P)$ y $g_j = g_j(P)$. Observemos que como $\Upsilon(\mu, g) \le \Upsilon(\nu, m)$ también se cumple que $\Upsilon(\mu, g) \le \Upsilon(\nu, g)$. Más aún, como ρ es estrictamente convexa, si $\nu \neq \mu$ la desigualdad es estricta. Luego, estamos resolviendo en este caso

$$\min_{\nu\in\mathbb{R}} L(\nu) = \min_{\nu\in\mathbb{R}} \Upsilon(\nu, g) \equiv \min_{\nu\in\mathbb{R}} \mathbb{E}\,\rho\left(\frac{Y - \nu - g(\mathbf{X})}{\sigma}\right).$$

por lo que diferenciando con respecto a μ, obtenemos la primera de las ecuaciones en (5.6).

Fijado $1 \le j \le d$, consideremos el problema de minimizar $\Upsilon(\mu, m)$ con respecto a m_j cuando todas las componentes del vector $\mathbf{m} = (m_1, \ldots, m_d)$ excepto la j se mantienen fijas e iguales a uno de los representantes del valor mínimo, esto es, minimizaremos sólo cuando $m(\mathbf{x}) = m_j(X_j) + \sum_{s\neq j}^d g_s(X_s)$. Sea

$$L_j(m_j) = \Upsilon(\mu, m) = \mathbb{E}\,\rho\left(\frac{Y - \mu - m_j(X_j) - \sum_{s\neq j} g_s(X_s)}{\sigma}\right) \quad \text{con} \quad m(\mathbf{x}) = m_j(X_j) + \sum_{s\neq j}^d g_s(X_s)$$

Nuevamente, como $L_j(g_j) \le L_j(m_j)$, ρ es convexa y se cumple **B3** tenemos que si $\mathbb{P}(g_j(X_j) \neq m_j(X_j)) > 0$ entonces la desigualdad es estricta.

La condición de estacionaridad dada en (5.6), se obtendrá tras igualar a cero la diferencial de Gateaux calculada sobre una dirección arbitraria $\eta_j \in \mathcal{M}_j$, que indicamos $\partial L_j(g_j;\eta_j)$. Veamos que

$$\partial L_j(g_j;\eta_j) = -\frac{1}{\sigma}\,\mathbb{E}\left[\psi\left(\frac{R_j - g_j(X_j)}{\sigma}\right)\eta_j(X_j)\right], \tag{5.7}$$

donde $R_j = Y - \mu - \sum_{k\neq j} g_k(X_k)$ es el residuo para g_j.

Sea $\eta \in \mathcal{M}_j$ fija y $\nu_\eta(t) = L_j(g_j + t\eta)$. Observemos que $\partial L_j(g_j;\eta) = \nu'_\eta(0)$

$$\nu'_\eta(0) = \lim_{t\to\infty}\frac{L_j(g_j+t\eta) - L_j(g_j)}{t} = \lim_{t\to\infty}\frac{1}{t}\,\mathbb{E}\left[\rho\left(\frac{R_j - g_j(X_j) - t\eta(X_j)}{\sigma}\right) - \rho\left(\frac{R_j - g_j(X_j)}{\sigma(\mathbf{X})}\right)\right].$$

Luego, usando el Teorema de la Convergencia Dominada y el hecho de que para todo $\sigma > 0$

$$\frac{\partial}{\partial t}\left\{\rho\left(\frac{R_j - g_j(x) - t\eta(x)}{\sigma}\right)\right\} = \psi\left(\frac{R_j - g_j(x) - t\eta(x)}{\sigma}\right)\left(-\frac{\eta(x)}{\sigma}\right)$$

se tiene que

$$\nu'_\eta(t) = \mathbb{E}\frac{\partial}{\partial t}\left\{\rho\left(\frac{R_j - g_j(X_j) - t\eta(X_j)}{\sigma}\right)\right\} = -\mathbb{E}\left[\frac{1}{\sigma}\Psi\left(\frac{R_j - g_j(X_j) - t\,\eta(X_j)}{\sigma}\right)\eta(X_j)\right],$$

por lo que evaluando en 0, se concluye la demostración de (5.7). Observemos que esto muestra que la función $\nu'_\eta : \mathbb{R} \to \mathbb{R}$ es creciente y por lo tanto, $\nu'_\eta(t) = 0$ tiene a lo sumo una solución.

Como $g_j = \operatorname{argmin}_{m_j\in\mathcal{M}_j} L_j(m_j)$ tenemos que $\partial L_j(g_j;\eta_j) = 0$ para toda función $\eta_j \in \mathcal{M}_j$, es decir, $\nu'_{\eta_j}(0) = 0$

$$\mathbb{E}\left[\psi\left(\frac{Y - \mu - \sum_{k\neq j} g_k(X_k) - g_j(X_j)}{\sigma}\right)\eta_j(X_j)\right] = \mathbb{E}\left[\psi\left(\frac{R_j - g_j(X_j)}{\sigma}\right)\eta_j(X_j)\right] = 0 \quad \forall\,\eta_j \in \mathcal{M}_j \tag{5.8}$$

Por otra parte, como ψ es creciente, acotada y continua con $\lim_{t\to+\infty}\psi(t) > 0$ y $\lim_{t\to-\infty}\psi(t) < 0$, aplicando el Teorema 2.1 de Boente and Fraiman (1989) con $Y = Y - \mu - \sum_{k\neq j} g_k(X_k)$ y $X = X_j$, deducimos que existe una única solución medible $\widetilde{g}_j(x)$ de $\lambda(x,a) = 0$ donde

$$\lambda(x,a) = \mathbb{E}\left\{\psi\left(\frac{Y - \mu - \sum_{k\neq j} g_k(X_k) - a}{\sigma}\right)\middle| X_j = x\right\}.$$

Como supusimos que el sistema (5.6) tiene una solución en $\mathcal{M}_j$, tenemos que $\widetilde{g}_j \in \mathcal{M}_j$. Sea entonces $\eta_j = \widetilde{g}_j(X_j) - g_j(X_j) \in \mathcal{M}_j$. Supongamos que $\mathbb{P}(\eta_j(X_j) \neq 0) > 0$. Como $\lambda(X_j, \widetilde{g}_j(X_j)) = 0$ tenemos que

$$0 = \mathbb{E}\left[\lambda(X_j, \widetilde{g}_j(X_j))\eta_j(X_j)\right] = \mathbb{E}\left[\Psi\left(\frac{Y - \mu - \sum_{k\neq j} g_k(X_k) - g_j(X_j) + \eta_j(X_j)}{\sigma}\right)\eta_j(X_j)\right],$$

o sea, $\widetilde{g}_j$ es tal que $\nu'_{\eta_j}(1) = 0$ lo cual es un absurdo, por lo que $\mathbb{P}(\eta_j(X_j) \neq 0) = 0$, o sea, $g_j = \widetilde{g}_j$, es decir, es solución de (5.6). $\square$

Observación 5.2.2. Vale la pena mencionar que si $(\mathbf{X}^{\mathrm{T}}, Y)^{\mathrm{T}}$ cumple (5.1) con $\sigma(\mathbf{X}) = \sigma$ para todo $\mathbf{x}$, ψ es impar y los errores ε tiene distribución simétrica, el sistema (5.6) tiene una solución m_j tal que $\mathbb{E}m_j(X_j) = 0$. Efectivamente, μ y g_j satisfacen el sistema (5.6) ya que $\Gamma_0(\mu, \mathbf{g}) = \mathbb{E}\psi(\varepsilon) = 0$ y

$$\mathbf{\Gamma}(\mu, \mathbf{g}, x) = \mathbb{E}\left[\psi\left(\frac{Y - \mu - \sum_{j\neq\ell} g_j(X_j) - g_\ell(X_\ell)}{\sigma}\right)\middle| X_\ell = x\right] = \mathbb{E}\psi(\varepsilon) = 0\,.$$

Más aún, en el caso heterocedástico, es fácil ver que como ε es independiente de $\mathbf{X}$, μ y g_j cumplen el sistema

$$\begin{cases} \mathbb{E}\left[\psi\left(\dfrac{Y - \mu - \sum_{j=1}^d g_j(X_j)}{\sigma(\mathbf{X})}\right)\right] & = 0 \\ \mathbb{E}\left[\dfrac{1}{\sigma(\mathbf{X})}\,\psi\left(\dfrac{Y - \mu - \sum_{j\neq\ell} g_j(X_j) - g_\ell(X_\ell)}{\sigma(\mathbf{X})}\right)\middle| X_\ell\right] & = 0 \quad 1 \leq \ell \leq d\,, \end{cases}$$

que provee una forma de generalizar el algoritmo al caso de modelos de función de escala no constante.

Observación 5.2.3. En el caso de tener respuestas faltantes, o sea, cuando tenemos una muestra $(\mathbf{X}_i^{\mathrm{T}}, Y_i, \delta_i)^{\mathrm{T}}$, $1 \leq i \leq n$ donde se cumple (2.7), definamos

$$\Upsilon_\delta(\nu, m) = \mathbb{E}\,\delta\rho\left(\frac{Y - \nu - \sum_{j=1}^d m_j(X_j)}{\sigma(\mathbf{X})}\right) = \mathbb{E}\,p(\mathbf{X})\rho\left(\frac{Y - \nu - \sum_{j=1}^d m_j(X_j)}{\sigma(\mathbf{X})}\right) \tag{5.9}$$

$$(\mu(P), g(P)) = \underset{(\nu,m)\in\mathbb{R}\times\mathcal{M}}{\operatorname{argmin}}\ \Upsilon_\delta(\nu, m)\,. \tag{5.10}$$

Argumentos análogos a los utilizados para probar el Teorema 5.2.1, permiten mostrar que, si $(\mathbf{X}^{\mathrm{T}}, Y)^{\mathrm{T}} \in \mathbb{R}^{d+1}$ cumple el modelo (5.1), se satisfacen **B1** y **B2**, entonces $\Upsilon_\delta(\nu, m)$ alcanza su único mínimo en $(\nu, m) \in \mathbb{R} \times \mathcal{M}$ cuando $\nu = \mu$ y $\mathbb{P}(m(\mathbf{X}) = \sum_{j=1}^d g_j(X_j)) = 1$. Por otra parte, si además se cumple **B3**, el único mínimo $(\mu(P), g(P)(\mathbf{X}) = \sum_{j=1}^d g_j(P)(X_j))$ cumple $\mathbb{P}((\mu(P), \mathbf{g}(P)(\mathbf{X})) = (\mu, \mathbf{g}(\mathbf{X}))) = 1$, con $\mathbf{g} = (g_1, \ldots, g_d)^{\mathrm{T}}$ y $\mathbf{g}(P) = (g_1(P), \ldots, g_d(P))^{\mathrm{T}}$, es decir, tenemos Fisher–consistencia.

Por otra parte, la demostración del Teorema 5.2.2 también puede extenderse al caso de respuestas faltantes ya que como $p(\mathbf{X}) \geq 0$, es posible mostrar que existe una única solución medible $\widetilde{g}_j(x)$ de $\lambda_\delta(x, a) = 0$ donde

$$\begin{aligned} \lambda_\delta(x, a) &= \mathbb{E}\left\{\delta\,\psi\left(\frac{Y - \mu - \sum_{k\neq j} g_k(X_k) - a}{\sigma}\right)\middle| X_j = x\right\} \\ &= \mathbb{E}\left\{p(\mathbf{X})\,\psi\left(\frac{Y - \mu - \sum_{k\neq j} g_k(X_k) - a}{\sigma}\right)\middle| X_j = x\right\}\,. \end{aligned}$$

Para ser más precisos, sea $\mathbf{\Gamma}_\delta(\nu, \mathbf{m}, \mathbf{x}) = (\Gamma_{0,\delta}(\nu, \mathbf{m}), \Gamma_{1,\delta}(\nu, \mathbf{m}, x_1), \ldots, \Gamma_{d,\delta}(\nu, \mathbf{m}, x_d))^{\mathrm{T}}$ con $\mathbf{m} = (m_1, \ldots, m_d)^{\mathrm{T}}$ y

$$\begin{aligned}
\Gamma_{0,\delta}(\nu, \mathbf{m}) &= \mathbb{E}\left[\delta\psi\left(\frac{Y - \nu - \sum_{j=1}^d m_j(X_j)}{\sigma}\right)\right] = \mathbb{E}\left[p(\mathbf{X})\psi\left(\frac{Y - \nu - \sum_{j=1}^d m_j(X_j)}{\sigma}\right)\right] \\
\Gamma_{\ell,\delta}(\nu, \mathbf{m}, x_\ell) &= \mathbb{E}\left[\delta\ \psi\left(\frac{Y - \mu - \sum_{j\neq\ell} g_j(X_j) - m_\ell(X_\ell)}{\sigma}\right)\middle| X_\ell = x_\ell\right] \quad 1 \leq \ell \leq d \\
&= \mathbb{E}\left[p(\mathbf{X})\ \psi\left(\frac{Y - \mu - \sum_{j\neq\ell} g_j(X_j) - m_\ell(X_\ell)}{\sigma}\right)\middle| X_\ell = x_\ell\right] \quad 1 \leq \ell \leq d\,.
\end{aligned}$$

Utilizando argumentos análogos a los considerados en la demostración del Teorema 5.2.2, es posible mostrar que si existe una solución $\mu(P) \in \mathbb{R}$, $g(P) = \sum_{j=1}^d g_j(P) \in \mathcal{M}$ de (5.10) y si $\mathbf{\Gamma}_\delta(\nu, \mathbf{m}, \mathbf{x}) = \mathbf{0}$ tiene una solución $\mathbf{m}_{\mathbf{\Gamma}} = (m_{1,\mathbf{\Gamma}}, \ldots, m_{d,\mathbf{\Gamma}})^{\mathrm{T}}$ tal que $\mathbb{E}m_{j,\mathbf{\Gamma}}(X_j) = 0$, entonces podemos elegir $g_j(P)$ de modo tal que satisfagan $\mathbf{\Gamma}_\delta(\mu, \mathbf{g}(P), \mathbf{x}) = \mathbf{0}$.

5.2.2 Convergencia del algoritmo de Backfitting robusto si $\sigma(\mathbf{X}) = \sigma$

En particular, si el modelo es homocedástico, o sea, si $\sigma(\mathbf{x}) \equiv \sigma$ para todo $\mathbf{x}$ tenemos que el funcional robusto $(\mu(P), \mathbf{g}(P))$ resuelve el sistema

$$\left\{\begin{array}{rl}
\mathbb{E}\left\{\psi\left(\dfrac{Y - \mu - \sum_{j=1}^d g_j(X_j)}{\sigma}\right)\right\} & = 0 \\
\mathbb{E}\left\{\psi\left(\dfrac{Y - \mu - \sum_{j\neq\ell} g_j(X_j)}{\sigma}\right)\middle| X_\ell\right\} & = 0 \quad 1 \leq \ell \leq d\,.
\end{array}\right. \tag{5.11}$$

Para resolverlo, proponemos el siguiente algoritmo de backfitting robusto

- **Paso 0:** Elija $\mathbf{g}^{(0)} = (g_1^{(0)}, \ldots, g_d^{(0)})^{\mathrm{T}} \in \mathcal{M}$, por ejemplo, $\mathbf{g}^{(0)} = \mathbf{0}$, $\mu^{(0)}$ un funcional robusto preliminar de μ, y σ_0 un estimador preliminar de σ.
- **Paso ℓ:** En la $\ell-$ésima iteración resolvemos el siguiente sistema de ecuaciones

$$\begin{aligned}
\mathbb{E}\Psi\left(\frac{Y - \sum_{j=1}^d g_j^{(\ell-1)}(X_j) - \mu^{(\ell)}}{\sigma_0}\right) &= 0 \\
\mathbb{E}\left[\Psi\left(\frac{Y - \mu^{(\ell)} - \sum_{j=2}^d g_j^{(\ell-1)}(X_j) - g_1^{(\ell)}(X_1)}{\sigma_0}\right)\middle| X_1\right] &= 0 \\
\mathbb{E}\left[\Psi\left(\frac{Y - \mu^{(\ell)} - g_1^{(\ell)}(X_1) - \sum_{j=3}^d g_j^{(\ell-1)}(X_j) - g_2^{(\ell)}(X_2)}{\sigma_0}\right)\middle| X_2\right] &= 0 \\
&\vdots \\
\mathbb{E}\left[\Psi\left(\frac{Y - \mu^{(\ell)} - \sum_{j=1}^{d-1} g_j^{(\ell)}(X_j) - g_d^{(\ell)}(X_d)}{\sigma_0}\right)\middle| X_d\right] &= 0
\end{aligned} \tag{5.12}$$

- Itere el **Paso** ℓ hasta que la convergencia es alcanzada.

Observemos que la primera ecuación en la $\ell-$ésima iteración del algoritmo es equivalente a encontrar

$$\mu^{(\ell)} = \underset{\mu\in\mathbb{R}}{\operatorname{argmin}}\, \mathbb{E}\rho\left(\frac{R_0^{(\ell)}-\mu}{\sigma_0}\right)$$

donde $R_0^{(\ell)} = Y - \sum_{j=1}^{d} g_j^{(\ell-1)}(X_j)$.

Por otra parte, en la $(k+1)-$ésima ecuación de la $\ell-$ésima iteración encontramos la solución $a = g_k(X_k) \in \mathcal{M}_k$ de

$$\mathbb{E}\left[\Psi\left(\frac{Y-\mu^{(\ell)}-\sum_{j=1}^{k-1} g_j^{(\ell)}(X_j)-\sum_{j=k+1}^{d} g_j^{(\ell-1)}(X_j)-a}{\sigma_0}\right)\middle| X_k\right] = 0$$

lo que corresponde a encontrar un $M-$funcional condicional de posición de los residuos parciales

$$g_k^{(\ell)}(X_k) = \underset{m_k\in\mathcal{M}_k}{\operatorname{argmin}}\, \mathbb{E}\left[\rho\left(\frac{R_k^{(\ell)}-m_k(X_k)}{\sigma_0}\right)\middle| X_k\right],$$

donde $R_k^{(\ell)} = Y - \mu^{(\ell)} - \sum_{j=1}^{k-1} g_j^{(\ell)}(X_j) - \sum_{j=k+1}^{d} g_j^{(\ell-1)}(X_j)$.

Es decir, el sistema de ecuaciones en (5.12) para la $\ell-$ésima iteración es equivalente al siguiente sistema de ecuaciones

$$\begin{aligned}
\mu^{(\ell)} &= \underset{\mu\in\mathbb{R}}{\operatorname{argmin}}\, \mathbb{E}\rho\left(\frac{R_0^{(\ell)}-\mu}{\sigma_0}\right)\\
g_1^{(\ell)}(X_1) &= \underset{m_1\in\mathcal{M}_1}{\operatorname{argmin}}\, \mathbb{E}\left[\rho\left(\frac{R_1^{(\ell)}-m_1(X_1)}{\sigma_0}\right)\middle| X_1\right]\\
g_2^{(\ell)}(X_2) &= \underset{m_2\in\mathcal{M}_2}{\operatorname{argmin}}\, \mathbb{E}\left[\rho\left(\frac{R_2^{(\ell)}-m_2(X_2)}{\sigma_0}\right)\middle| X_2\right]\\
&\vdots\\
g_d^{(\ell)}(X_d) &= \underset{m_d\in\mathcal{M}_d}{\operatorname{argmin}}\, \mathbb{E}\left[\rho\left(\frac{R_d^{(\ell)}-m_d(X_d)}{\sigma_0}\right)\middle| X_d\right]
\end{aligned} \tag{5.13}$$

Por lo tanto, la sucesión de funciones $\left(\mu^{(\ell)}, \mathbf{g}^{(\ell)}\right)_{\ell\geq 1} = \left(\mu^{(\ell)}, g_1^{(\ell)}, \ldots, g_d^{(\ell)}\right)_{\ell\geq 1}$ produce una sucesión de valores no crecientes de la función objetivo (5.3).

Efectivamente,

Consideremos la iteración ℓ=1. Primero obtenemos $\mu^{(1)}$ usando la primera ecuación en (5.13), es decir,

$$\mu^{(1)} = \underset{\mu\in\mathbb{R}}{\operatorname{argmin}}\, \mathbb{E}\rho\left(\frac{R_0^{(1)}-\mu}{\sigma_0}\right) = \arg\min_{\mu\in\mathbb{R}} \Phi\left(\mu, \mathbf{g}^{(0)}\right).$$

Entonces, para cualquier $\mu \in \mathbb{R}$ tenemos que

$$\Phi\left(\mu^{(1)}, \mathbf{g}^{(0)}\right) \leq \Phi\left(\mu, \mathbf{g}^{(0)}\right) .$$

En particular, vale para $\mu = \mu^{(0)}$ con lo cual $\Phi\left(\mu^{(1)}, \mathbf{g}^{(0)}\right) \leq \Phi\left(\mu^{(0)}, \mathbf{g}^{(0)}\right)$.

El segundo paso de la primera iteración es estimar la primera componente aditiva, esto es,

$$g_1^{(1)}(X_1) = \underset{m_1 \in \mathcal{M}_1}{\operatorname{argmin}} \mathbb{E}\rho\left(\frac{R_1^{(1)} - m_1(X_1)}{\sigma_0}\right) = \underset{m_1 \in \mathcal{M}_1}{\operatorname{argmin}} \Phi\left(\mu^{(1)}, m_1, g_2^{(0)}, \ldots, g_d^{(0)}\right) .$$

Luego, para cualquier función $m_1 \in \mathcal{M}_1$ obtenemos que

$$\Phi\left(\mu^{(1)}, g_1^{(1)}, g_2^{(0)}, \ldots, g_d^{(0)}\right) \leq \Phi\left(\mu^{(1)}, m_1, g_2^{(0)}, \ldots, g_d^{(0)}\right) .$$

En particular, para $m_1 = g_1^{(0)}$ obtenemos $\Phi\left(\mu^{(1)}, g_1^{(1)}, g_2^{(0)}, \ldots, g_d^{(0)}\right) \leq \Phi\left(\mu^{(1)}, g_1^{(0)}, g_2^{(0)}, \ldots, g_d^{(0)}\right) = \Phi\left(\mu^{(1)}, \mathbf{g}^{(0)}\right)$.

El tercer paso de la primer iteración estima la segunda componente aditiva g_2, esto es,

$$g_2^{(1)}(X_1) = \underset{m_2 \in \mathcal{M}_2}{\operatorname{argmin}} \mathbb{E}\rho\left(\frac{R_2^{(1)} - m_2(X_2)}{\sigma_0}\right) = \underset{m_2 \in \mathcal{M}_2}{\operatorname{argmin}} \Phi\left(\mu^{(1)}, g_1^{(1)}, m_2, g_3^{(0)}, \ldots, g_d^{(0)}\right) .$$

Luego, para cualquier función $m_2 \in \mathcal{M}_2$ tenemos que

$$\Phi\left(\mu^{(1)}, g_1^{(1)}, g_2^{(1)}, g_3^{(0)}, \ldots, g_d^{(0)}\right) \leq \Phi\left(\mu^{(1)}, g_1^{(1)}, m_2, g_3^{(0)}, \ldots, g_d^{(0)}\right) .$$

En particular, para $m_2 = g_2^{(0)}$ obtenemos $\Phi\left(\mu^{(1)}, g_1^{(1)}, g_2^{(1)}, g_3^{(0)} \ldots, g_d^{(0)}\right) \leq \Phi\left(\mu^{(1)}, g_1^{(1)}, g_2^{(0)}, \ldots, g_d^{(0)}\right)$.

En general, para el $(k+1)-$ésimo paso de la primer iteración obtenemos

$$\Phi\left(\mu^{(1)}, g_1^{(1)}, \ldots, g_k^{(1)}, g_{k+1}^{(0)} \ldots, g_d^{(0)}\right) \leq \Phi\left(\mu^{(1)}, g_1^{(1)}, \ldots, g_{k-1}^{(1)}, g_k^{(0)}, \ldots, g_d^{(0)}\right) . \tag{5.14}$$

Luego, para $k = d+1$ obtenemos

$$\Phi\left(\mu^{(1)}, \mathbf{g}^{(1)}\right) = \Phi\left(\mu^{(1)}, g_1^{(1)}, \ldots, g_d^{(1)}\right) \leq \Phi\left(\mu^{(1)}, g_1^{(1)}, \ldots, g_{d-1}^{(1)}, g_d^{(0)}\right) .$$

Finalmente, cuando la primera iteración termina, usando (5.14) sucesivamente, obtenemos que

$$\Phi\left(\mu^{(1)}, \mathbf{g}^{(1)}\right) \leq \Phi\left(\mu^{(0)}, \mathbf{g}^{(0)}\right) .$$

Consideremos la iteración ℓ. Usando los mismos argumentos que en la primera iteración, en el $(k+1)-$ésimo paso tenemos que

$$\Phi\left(\mu^{(\ell)}, g_1^{(\ell)}, \ldots, g_k^{(\ell)}, g_{k+1}^{(\ell-1)}, \ldots, g_d^{(\ell-1)}\right) \leq \Phi\left(\mu^{(\ell)}, g_1^{(\ell)}, \ldots, g_{k-1}^{(\ell)}, g_k^{(\ell-1)}, g_{k+1}^{(\ell-1)}, \ldots, g_d^{(\ell-1)}\right) , \tag{5.15}$$

y al final de esta iteración y usando iterativamete (5.15), tenemos que

$$\Phi\left(\mu^{(\ell)}, \mathbf{g}^{(\ell)}\right) \leq \Phi\left(\mu^{(\ell-1)}, \mathbf{g}^{(\ell-1)}\right).$$

Argumentos estándares muestran que el algoritmo converge y que el límite satisface (5.6). La convexidad de la función objetivo implica que éste es el mínimo de $\Upsilon(\nu, m) = \Phi(\nu, \mathbf{m}) = \mathbb{E}\rho\left(\left(Y - \nu - \sum_{j=1}^{d} m_j(X_j)\right)/\sigma\right)$ sobre ν y las funciones $m(\mathbf{x}) = \sum_{j=1}^{d} m_j(x_j)$ tales que $m_j \in \mathcal{M}_j$.

5.3 Estimadores de las componentes aditivas por el algoritmo de Backfitting robusto si $\sigma(\mathbf{X}) = \sigma$

En la práctica, dada una muestra aleatoria $(\mathbf{X}_i^{\mathrm{T}}, Y_i)^{\mathrm{T}}$ $1 \leq i \leq n$ que cumple el modelo aditivo (5.1) aplicamos el algoritmo (5.12) reemplazando esperanzas por medias muestrales y estimamos las esperanzas condicionales usando suavizadores univariados, es decir, usamos $M-$estimadores locales. Sea $K : \mathbb{R} \to \mathbb{R}$ una función núcleo. Posibles elecciones de los estimadores es considerar estimadores basados en núcleos o en polinomios locales de orden q. En este último caso, la versión muestral de (5.12) es

- **Paso 0:** Sean un estimador inicial de las componentes aditivas $\widehat{\mathbf{g}}^{(0)} = (\widehat{g}_1^{(0)}, \ldots, \widehat{g}_d^{(0)})^{\mathrm{T}} \in \mathcal{M}$, por ejemplo, $\widehat{\mathbf{g}} = \mathbf{0}$, $\widehat{\mu}^{(0)}$ un estimador robusto preliminar de μ y $\widehat{\sigma}_0$ es un estimador preliminar de σ.
- **Paso ℓ:** Sea $\mathbf{Z}_{i,k} = (1, X_{i,k} - x, \ldots, (X_{i,k} - x)^q)^{\mathrm{T}}$. En la iteración ℓ resolvemos el siguiente sistema de ecuaciones

$$\left\{\begin{array}{rl}
\dfrac{1}{n}\displaystyle\sum_{i=1}^{n} \Psi\left(\dfrac{Y_i - \sum_{j=1}^{d} \widehat{g}_j^{(\ell-1)}(X_{i,j}) - \widehat{\mu}^{(\ell)}}{\widehat{\sigma}_0}\right) & = 0 \\
\dfrac{1}{n}\displaystyle\sum_{i=1}^{n} K_h(X_{i,1} - x)\Psi\left(\dfrac{Y_i - \widehat{\mu}^{(\ell)} - \sum_{j=2}^{d} \widehat{g}_j^{(\ell-1)}(X_{i,j}) - \sum_{s=1}^{q} \beta_{s,1}(X_{i,1} - x)^s}{\widehat{\sigma}_0}\right)\mathbf{Z}_{i,1} & = \mathbf{0} \\
\text{luego} \qquad \widehat{g}_1^{(\ell)}(x) = \beta_{0,1} & \\
\dfrac{1}{n}\displaystyle\sum_{i=1}^{n} K_h(X_{i,2} - x)\Psi\left(\dfrac{Y_i - \mu^{(\ell)} - \widehat{g}_1^{(\ell)}(X_{i,1}) - \sum_{j=3}^{d} \widehat{g}_j^{(\ell-1)}(X_{i,j}) - \sum_{s=1}^{q} \beta_{s,2}(X_{i,1} - x)^s}{\widehat{\sigma}_0}\right)\mathbf{Z}_{i,2} & = \mathbf{0} \\
\text{luego} \qquad \widehat{g}_2^{(\ell)}(x) = \beta_{0,2} & \\
\vdots & \\
\dfrac{1}{n}\displaystyle\sum_{i=1}^{n} K_h(X_{i,d} - x)\Psi\left(\dfrac{Y_i - \widehat{\mu}^{(\ell)} - \sum_{j=1}^{d-1} \widehat{g}_j^{(\ell)}(X_{i,j}) - \sum_{s=1}^{q} \beta_{s,d}(X_{i,d} - x)^s}{\widehat{\sigma}_0}\right)\mathbf{Z}_{i,d} & = \mathbf{0} \\
\text{luego} \qquad \widehat{g}_d^{(\ell)}(x) = \beta_{0,d} &
\end{array}\right. \tag{5.16}$$

- Repita el **Paso** ℓ hasta convergencia o hasta un número máximo de iteraciones.

En particular, si consideramos $M-$estimadores locales obtenemos el sistema

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n}\Psi\left(\frac{Y_i-\sum_{j=1}^{d}\widehat{g}_j^{(\ell-1)}(X_{i,j})-\widehat{\mu}^{(\ell)}}{\widehat{\sigma}_0}\right) &= 0\\
\frac{1}{n}\sum_{i=1}^{n}K_h(X_{i,1}-x)\Psi\left(\frac{Y_i-\widehat{\mu}^{(\ell)}-\sum_{j=2}^{d}\widehat{g}_j^{(\ell-1)}(X_{i,j})-\beta_{0,1}}{\widehat{\sigma}_0}\right) &= 0\\
\text{luego}\qquad \widehat{g}_1^{(\ell)}(x)=\beta_{0,1}&\\
\frac{1}{n}\sum_{i=1}^{n}K_h(X_{i,2}-x)\Psi\left(\frac{Y_i-\widehat{\mu}^{(\ell)}-\widehat{g}_1^{(\ell)}(X_{i,1})-\sum_{j=3}^{d}\widehat{g}_j^{(\ell-1)}(X_{i,j})-\beta_{0,2}}{\widehat{\sigma}_0}\right) &= 0\\
\text{luego}\qquad \widehat{g}_2^{(\ell)}(x)=\beta_{0,2}&\\
&\vdots\\
\frac{1}{n}\sum_{i=1}^{n}K_h(X_{i,d}-x)\Psi\left(\frac{Y_i-\widehat{\mu}^{(\ell)}-\sum_{j=1}^{d-1}\widehat{g}_j^{(\ell)}(X_{i,j})-\beta_{0,d}}{\widehat{\sigma}_0}\right) &= 0\\
\text{luego}\qquad \widehat{g}_d^{(\ell)}(x)=\beta_{0,d}&
\end{aligned}$$

Es decir, definiendo $W_{n,i}^{(k)}(x) = K_h(X_{i,k}-x)/\sum_{j=1}^{n}K_h(X_{j,k}-x)$, tenemos que por ser ψ creciente, $(\widehat{\mu}^{(\ell)},\widehat{\mathbf{g}}^{(\ell)}(x))$, con $\widehat{\mathbf{g}}^{(\ell)}(x)) = (\widehat{g}_1^{(\ell)}(x),\ldots,\widehat{g}_d^{(\ell)}(x))^{\mathrm{T}}$, es solución del siguiente problema de minimización

$$\begin{aligned}
\widehat{\mu}^{(\ell)} &= \operatorname*{argmin}_{\mu\in\mathbb{R}}\frac{1}{n}\sum_{i=1}^{n}\rho\left(\frac{Y_i-\sum_{j=1}^{d}\widehat{g}_j^{(\ell-1)}(X_{i,j})-\mu}{\widehat{\sigma}_0}\right)\\
\widehat{g}_k^{(\ell)}(x) &= \operatorname*{argmin}_{a\in\mathbb{R}}\sum_{i=1}^{n}W_{n,i}^{(k)}(x)\rho\left(\frac{Y_i-\widehat{\mu}^{(\ell)}-\sum_{j=1}^{k-1}\widehat{g}_j^{(\ell)}(X_{i,j})-\sum_{j=k+1}^{d}\widehat{g}_j^{(\ell-1)}(X_{i,j})-a}{\widehat{\sigma}_0}\right)\qquad 1\le k\le d\,,
\end{aligned}$$

donde si $k=1$, $\sum_{j=1}^{k-1}\widehat{g}_j^{(\ell)}(X_{i,j})=0$ mientras que si $k=d$, $\sum_{j=k+1}^{d}\widehat{g}_j^{(\ell-1)}(X_{i,j})=0$. Por lo tanto, usando que $\sum_{j=1}^{n}W_{n,i}^{(k)}(x)=1$, se obtiene que en cada paso de la iteración se reduce la función objetivo

$$\widehat{\Phi}(\nu,\mathbf{m}) = \frac{1}{n}\sum_{i=1}^{n}\rho\left(\frac{Y_i-\nu-\sum_{j=1}^{d}m_j(X_{i,j})}{\widehat{\sigma}_0}\right)$$

con lo que, como en la Sección 5.2.2, se obtiene que el algoritmo converge al valor mínimo de $\widehat{\Phi}(\nu,\mathbf{m})$.

Observación 5.3.1. Este enfoque también puede ser motivado de otra manera. Si bien tomando $\rho(t)=t^2$ en (5.2) el algoritmo (5.12) se reduce al algoritmo de backfitting estándar, el procedimiento dado en (5.12) se puede motivar observando que, si la muestra $(\mathbf{X}_i^{\mathrm{T}},Y_i)^{\mathrm{T}}$, $1\le i\le n$ satisface (5.1),

entonces para $1 \leq k \leq d$ tenemos que

$$\widetilde{Y}_{i,k} = Y_i - \mu - \sum_{j=1, j\neq k} g_j(X_{i,j}) = g_k(X_{i,k}) + \sigma\varepsilon_i \,.$$

Por lo tanto, la muestra $(X_{i,k}, \widetilde{Y}_{ik})_{i=1}^n$ satisface un modelo de regresión noparamétrico univariado en el que si existen momentos

$$\mathbb{E}\left(\widetilde{Y}_{i,k}|X_{i,k}\right) = g_k(X_{i,k}) \,. \tag{5.17}$$

Luego, nuestro enfoque puede también ser pensado como la utilización de suavizadores robustos para estimar las esperanzas condicionales (5.17) o más generalmente el funcional definido en Boente y Fraiman (1989).

Observación 5.3.2. Una posible elección del estimador preliminar de escala $\widehat{\sigma}_0$ se obtiene calculando la MAD de los residuos obtenidos estimando con medianas locales, es decir,

$$\widehat{\sigma}_0 = \mathrm{mad}_{1\leq i\leq n}\left\{Y_i - \widehat{Y}_i\right\}$$

where $\widehat{Y}_i = \mathrm{median}\,\{Y_j : |X_{j,k} - X_{i,k}| \leq h_k \,\forall\, k = 1, \ldots, d\}$.

5.4 Ejemplo

En esta Sección, estudiaremos el Ejemplo de los precios de las casas en Boston que ya hemos mencionado en la Sección 1.2 de la Introducción. Habíamos visto que el estimador de backfitting clásico varía drásticamente su comportamiento cuando un pequeño porcentaje de las observaciones es modificado. Para ver con más detalle las diferencias en el comportamiento del estimador de backfitting clásico antes y después de contaminar las observaciones, en la Figura 5.1 podemos ver los ajustes obtenidos por este estimador al aplicarse a los datos originales (en líneas punteadas) y a los datos contaminados (en líneas enteras).

En este Capítulo, hemos introducido el estimador de backfitting robusto para obtener estimadores menos sensibles a la presencia de observaciones atípicas. En la Figura 5.2 podemos ver los ajustes obtenidos con este estimador cuando los datos han sido contaminados artificialmente en sus respuestas. Mientras que el estimador clásico ajusta de manera errática al conjunto de datos con observaciones contaminadas, los residuos del estimador robusto permiten detectar sin problemas cuáles son las observaciones contaminadas.

La Figura 5.3 muestra los ajustes obtenidos con ambos estimadores para los datos contaminados artificialmente. Las líneas punteadas en azul corresponden al estimador de backfitting clásico mientras que las líneas rojas al estimador de backfitting robusto. En este gráfico se ve claramente las diferencias de los ajustes obtenidos por ambos procedimientos.

Por último, la Figura 5.4 presenta los ajustes obtenidos por el estimador robusto para el conjunto de datos original (en líneas punteadas) y para el conjunto de datos con respuestas contaminadas (en líneas enteras). Se observa la estabilidad del procedimiento ante un porcentaje de respuestas atípicas ya que el estimador robusto produce ajustes similares en ambos casos.

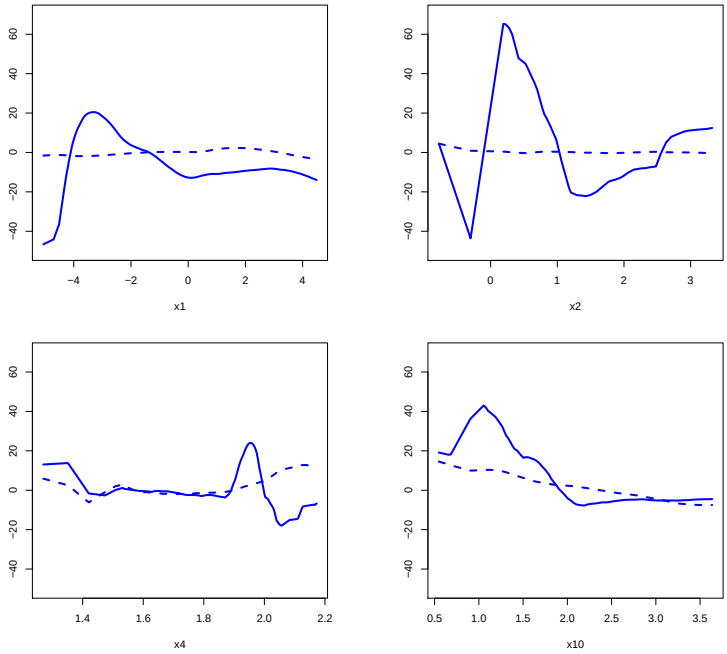

Figura 5.1: Gráficos de los ajustes obtenidos para las componentes g_1, g_2, g_4 y g_{10} (de izquierda a derecha y de arriba a abajo), por el estimador de backfitting clásico con los datos originales (en líneas punteadas) y cuando introducimos artificialmente respuestas contaminadas.

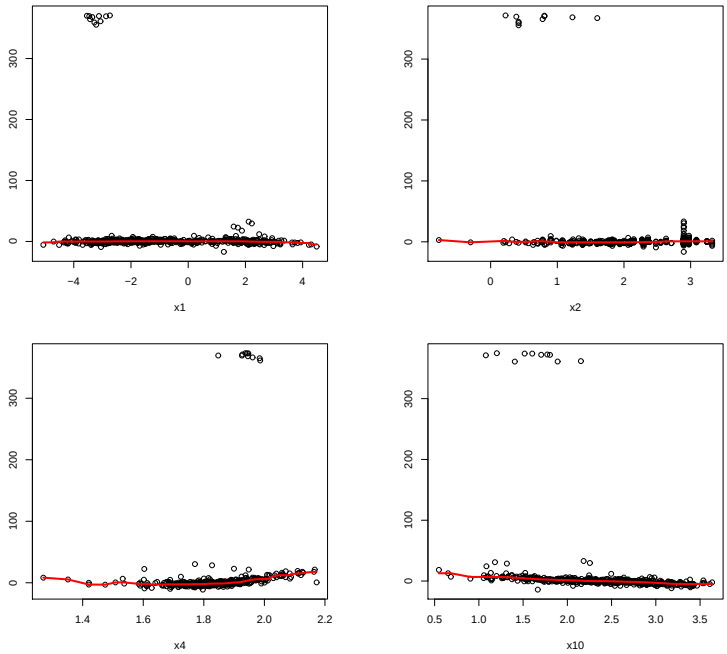

Figura 5.2: Gráficos de los ajustes obtenidos para las componentes g_1, g_2, g_4 y g_{10} (de izquierda a derecha y de arriba a abajo), por el estimador de backfitting robusto cuando hay contaminación en las respuestas.

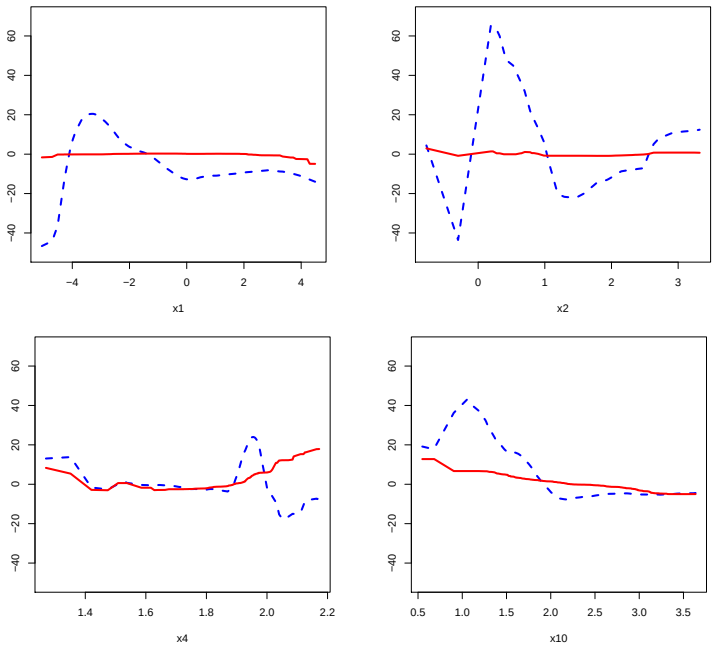

Figura 5.3: Gráficos de los estimadores las componentes g_1, g_2, g_4 y g_{10} (de izquierda a derecha y de arriba a abajo) por el estimador clásico (en líneas punteadas azules) y por el estimador robusto (en líneas enteras rojas) cuando introducimos artificialmente respuestas contaminadas.

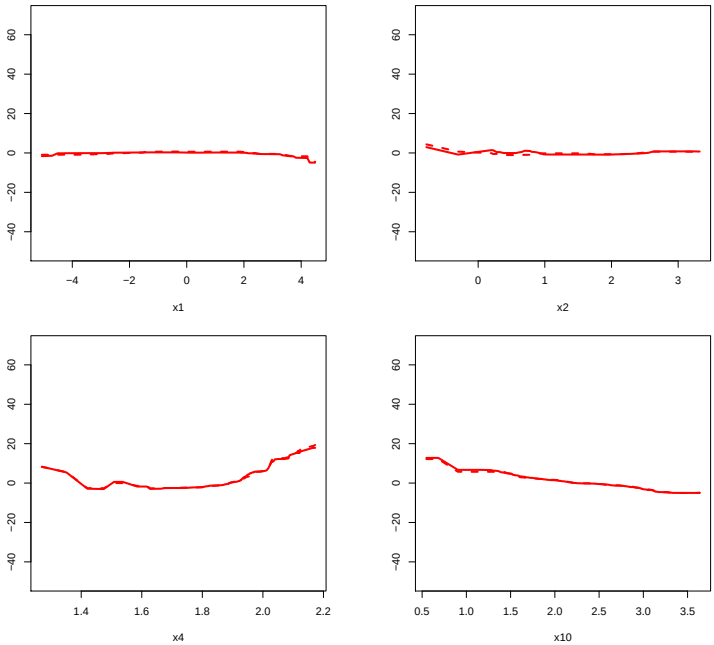

Figura 5.4: Gráficos de los estimadores de las componentes g_1, g_2, g_4 y g_{10} (de izquierda a derecha y de arriba a abajo) utilizando el estimador de backfitting robusto para el conjunto de datos originales (en líneas punteadas) y contaminadas.

Capítulo 6

Estudio Monte Carlo

6.1 Condiciones de la simulación

En este Capítulo, se describen los resultados de un estudio de simulación cuyo objetivo es comparar el comportamiento de los estimadores propuestos en el Capítulo 2 entre ellos y versus sus versiones clásicas, cuando se utilizan dintintos mecanismos de pérdida de datos y diferentes modelos de contaminación.

En un primer estudio de simulación se realizaron 500 replicaciones en las que se generaron muestras aleatorias independientes $\{(\mathbf{X}_i^{\mathrm{T}}, Y_i, \delta_i)\}_{i=1}^n$ de tamaño $n = 100$. Para ello, primero generamos observaciones $(\mathbf{X}_i^{\mathrm{T}}, Z_i)$ tales que

$$Z_i = m(\mathbf{X}_i) + \varepsilon_i, \qquad 1 \leq i \leq n\,,$$

donde $\mathbf{X}_i = (X_{i1}, X_{i2}) \sim U([0,1] \times [0,1])$, $\varepsilon_i = \sigma\, U_i$, $U_i \sim N(0,1)$, $\sigma = 0.5$, $g : \mathbb{R}^2 \to \mathbb{R}$ es una función aditiva de la forma

$$g(x_1, x_2) = 24\left(x_1 - \frac{1}{2}\right)^2 + 2\pi \sin(\pi x_2) - 6\,. \tag{6.1}$$

y $\mathbf{X}_i$ y U_i son independientes entre sí.

Los modelos para generar las respuestas faltantes pueden describirse como siguen. Definimos $Y_i = Z_i$ si $\delta_i = 1$ y faltante en otro caso, donde $\{\delta_i\}_{i=1}^n$ se generaron según un mecanismo de pérdida de observaciones ignorable, MAR, es decir, $P\left(\delta_i = 1 | Y_i, \mathbf{X}_i\right) = P\left(\delta_i = 1 | \mathbf{X}_i\right) = p\left(\mathbf{X}_i\right)$ con p alguna de las siguientes funciones

- $p_1(\mathbf{x}) \equiv 1$ que corresponde a la situación de muestras completas,
- $p_2(\mathbf{x}) = 0.4 + 0.5(\cos(x_1 + 0.2))^2$.

La probabilidad $p_1(\mathbf{x})$ permite comparar los estimadores propuestos, simplificados y completos, con el estimador de regresión que podría ser evaluado si contáramos con el conjunto total de datos. Observemos que este estimador, al que nos referiremos como *estimador de datos completos*, no

puede calcularse en la práctica. El objetivo es detectar qué estimador, simplificado o completo, da errores cuadráticos más cercanos a los obtenidos si no hubiese respuestas faltantes.

Con el mecanismo de pérdida p_2 se pierde aproximadamente un 31.5% de las respuestas.

Para identificar las funciones componentes marginales debemos tener en cuenta que, por **A1**, $\int g_\alpha(x) q_\alpha(x)\, dx = 0$. En esta oportunidad hemos considerado $q_\alpha(x) = I_{(0,1)}(x)$, para todo $1 \leq \alpha \leq d$. Luego, para el modelo (6.1), tenemos que $c = 0$ y las componentes aditivas son

$$g_1(x_1) = 24\left(x_1 - \frac{1}{2}\right)^2 - 2 \qquad\qquad g_2(x_2) = 2\pi \sin(\pi x_2) - 4\ .$$

Respecto del procedimiento de suavizado, hemos utilizado el núcleo de Epanechnikov producto. Recordemos que esta función núcleo es de la forma $\mathcal{K}(\mathbf{x}) = K(x_1)K(x_2)$ donde $K(u) = \frac{3}{4}(1 - u^2)\mathrm{I}_{[-1,1]}(u)$.

El comportamiento de un estimador $\widehat{g}$ de g se midió utilizando una aproximación del error cuadrático medio integrado y calculado para cada una de las replicaciones como

$$\text{ISE}(\widehat{g}) = \frac{1}{\ell^2} \sum_{s=1}^{\ell} \sum_{j=1}^{\ell} \left(g\left(\mathbf{u}_{js}\right) - \widehat{g}\left(\mathbf{u}_{js}\right)\right)^2\ ,$$

donde $\mathbf{u}_{js} = (j/\ell, s/\ell)$, $1 \leq j, s \leq \ell$, $\ell = 50$. Una aproximación del MISE se obtuvo promediando sobre las 500 replicaciones del ISE. Medidas similares se utilizaron para comparar los estimadores de las componentes aditivas g_α.

En una primera etapa de simulación, se han fijado las ventanas $h_j = 0.10$ para $j = 1, 2$.

Los estimadores que hemos considerado son los siguientes.

- $\widetilde{g}_{\text{S}}$: Estimador de Nadaraya-Watson adaptado a respuestas faltantes, definido en (2.8),
- $\widetilde{g}_{\text{LS}_1}$: Estimador preliminar clásico de la función de regresión g basado en polinomios locales de orden 1 en todas las coordenadas de $\mathbf{X}$,
- $\widetilde{g}_{\text{M}_0}$: $M-$ estimador preliminar basado en polinomios locales de orden 0 definido en (2.12) considerando $t(\mathbf{x}) \equiv 1$ y con función ρ la función de Huber con constante $c = 1.345$,
- $\widetilde{g}_{\text{M}_1}$: $M-$ estimador preliminar basado en polinomios locales de orden 1 en todas las coordenadas de $\mathbf{X}$, definido como la primera coordenada de la solución en (2.16), y considerando $t(\mathbf{x}) \equiv 1$ y con función ρ la función de Huber con constante $c = 1.345$,
- $\widehat{g}_{\text{S}}$: Estimador basado en integración marginal que utiliza como estimador preliminar $\widetilde{g}_{\text{S}}$,
- $\widehat{g}_{\text{LS}_1}$: Estimador clásico que utiliza el procedimiento de integración marginal aplicado sobre el estimador de regresión preliminar $\widetilde{g}_{\text{LS}_1}$,
- $\widehat{g}_{\text{LS}_{q,\alpha}}$: Estimador que utiliza el procedimiento de integración marginal sobre el estimador clásico basado en polinomios locales de orden $q = 1$ en la dirección de interés α,
- $\widehat{g}_{\text{BC}}$: Estimador de backfitting clásico adaptado a respuestas faltantes, definido en el Capítulo 5,

- $\widehat{g}_{\mathrm{M}_0}$: Estimador robusto basado en polinomios locales de orden 0, definido en (2.14), que utiliza como estimador preliminar $\widetilde{g}_{\mathrm{M}_0}$,
- $\widehat{g}_{\mathrm{M}_1}$: Estimador robusto basado en polinomios locales de orden 1 en todas las coordenadas $\mathbf{X}$, definido en (2.19), que utiliza como estimador preliminar $\widetilde{g}_{\mathrm{M}_1}$,
- $\widehat{g}_{\mathrm{M}_{q,\alpha}}$: Estimador robusto basado en polinomios locales de orden $q = 1$ en la dirección de interés α, definido en (2.23), y considerando $t(\mathbf{x}) \equiv 1$ y con función ρ la función de Huber con constante $c = 1.345$,
- $\widehat{g}_{\mathrm{BR}}$: Estimador backfitting robusto, definido en el Capítulo 5.

Vale la pena notar que los estimadores clásicos $\widetilde{g}_{\mathrm{S}}$, $\widetilde{g}_{\mathrm{LS}_1}$ y $\widehat{g}_{\mathrm{LS}_{q,\alpha}}$ corresponden a elegir $t(\mathbf{x}) \equiv 1$ y como función ρ la pérdida cuadrática en (2.12), (2.16) y (2.21), respectivamente.

En las tablas y gráficos se reportan los resultados de los MISE promediando sobre las NREP $= 500$ replicaciones y considerando muestras de tamaño $n = 500$.

Notemos por $\mathcal{D}_\eta = [0.2, 0.2 + \eta] \times [0.2, 0.2 + \eta]$. En las tablas y gráficos indicaremos por C_0 a las muestras sin contaminar. Por otra parte, se consideraron cuatro mecanismos de contaminación diferentes que se aplicaron a cada muestra simplificada generada, es decir, tomando solamente los datos completos $\{(\mathbf{X}_i^{\mathrm{T}}, Y_i)^{\mathrm{T}}\}_{\delta_i=1}$ cuando se utiliza un mecanismo de pérdida de datos. Las contaminaciones indicadas C_j, $j = 1, \ldots 4$, se generaron de la siguiente forma

- C_1: es un modelos de errores groseros ya que se generaron los errores con distribución $\varepsilon_i \sim (1-\nu)N(0, \sigma^2) + \nu N(15, 0.01^2)$ con $\nu = 0.15$.
- C_2: es un modelo de contaminación en el que se concentran los datos atípicos en la región $\mathcal{D}_\eta$, con $\eta = 0.3$, más precisamente se tomó $\varepsilon_i \sim N(15, 0.01)$ para todo i tal que $\mathbf{X}_i \in \mathcal{D}_{0.3}$,
- C_3: es un modelo de contaminación concentrada en la región $\mathcal{D}_\eta$, con $\eta = 0.09$, es decir, $\varepsilon_i \sim N(10, 0.01)$ para todo i tal que $\mathbf{X}_i \in \mathcal{D}_{0.09}$,
- C_4: en este caso, se considera un modelo de errores groseros con un alto porcentaje pero en $\mathcal{D}_\eta$, con $\eta = 0.3$, más precisamente, $\varepsilon_i \sim (1-\nu)N(0, \sigma^2) + \nu N(15, 0.01)$, con $\nu = 0.30$, para todo i tal que $\mathbf{X}_i \in \mathcal{D}_{0.3}$.

La situación C_0 en que no hay datos contaminados nos permitirá analizar el balance entre el efecto de las contaminaciones sobre los distintos estimadores estudiados y su eficiencia. La contaminación C_1 corresponde a un *modelo de errores groseros*. Éste es un caso de contaminación esparcida de manera homogénea entre todas las respuestas de la muestra. C_2 corresponde a un caso de contaminación patológica, en el que para todos los puntos pertenecientes al cuadrado $[0.2, 0.5] \times [0.2, 0.5]$, las respuestas son contaminadas. La contaminación C_3 es similar a la contaminación C_2 pero para los puntos pertenecientes al cuadrado $[0.2, 0.29] \times [0.2, 0.29]$. La diferencia más importante entre estas dos contaminaciones es que el primero de los cuadrados ($\mathcal{D}_{0.3}$) tiene lado de longitud 0.3, mayor al tamaño de ventana considerada ($h_j = 0.10$ para $j = 1, 2$), mientras que el segundo ($\mathcal{D}_{0.09}$) es de 0.09. Éste es un caso de contaminación concentrada y por lo tanto, se espera que resulte perjudicial aún para el estimador $\widehat{g}_{\mathrm{M}_0}$. Por último, la contaminación C_4 corresponde al caso en que sólo el 30% de los puntos que yacen en el cuadrado $\mathcal{D}_{0.3}$ son contaminados.

A modo de ejemplo, en la Figura 6.1, se grafican en *azul* los pares (x_1, x_2) cuyas respuestas y han sido contaminadas, para las contaminaciones C_2, C_3 y C_4, considerando una de las muestras de tamaño 500.

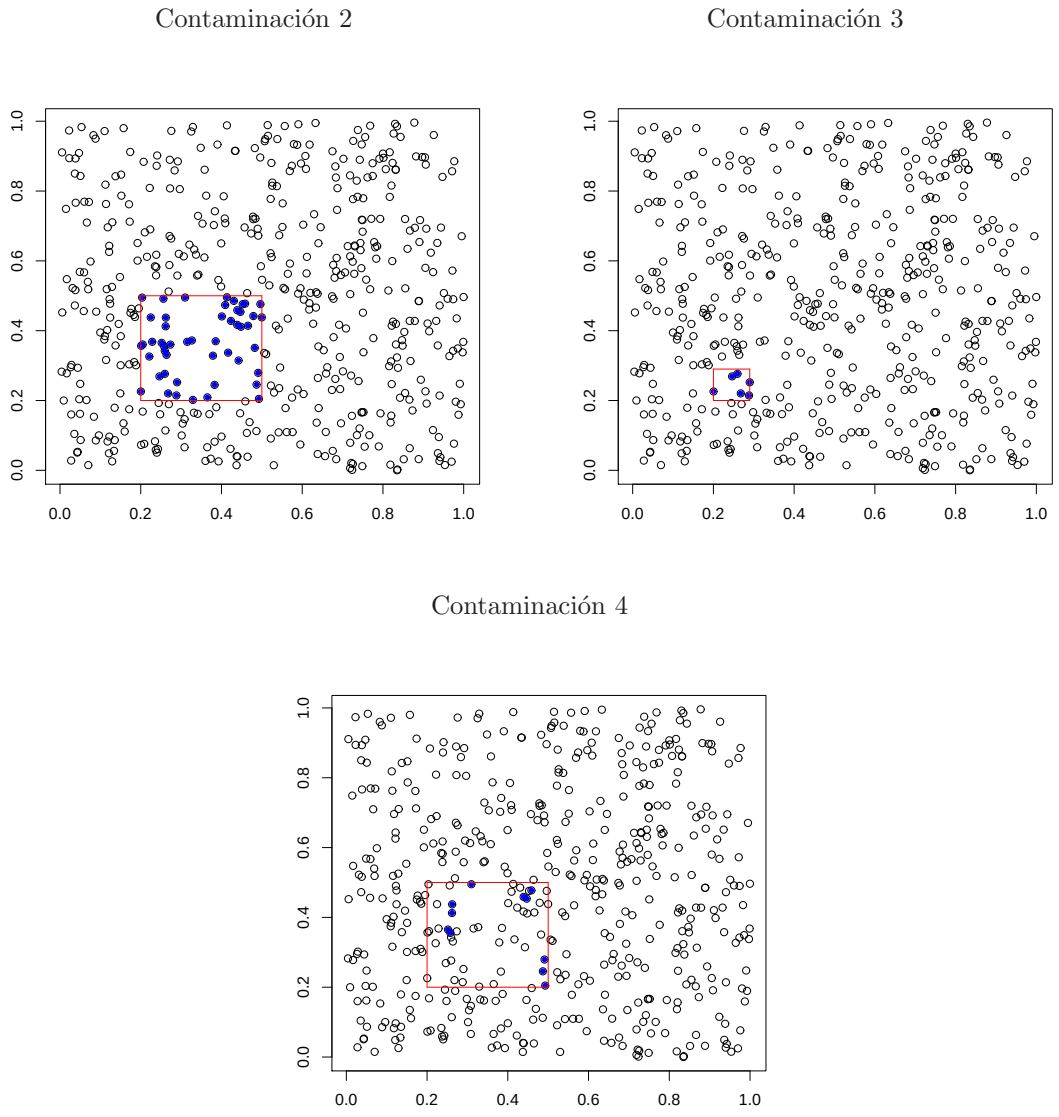

Figura 6.1: Gráficos que indican cuáles son los pares (x_1, x_2) cuyas respuestas y han sido contaminadas, para los últimos tres mecanismos de contaminación.

6.1.1 Selección preliminar de $\widehat{g}_{\mathrm{M}_0}$

En el Capítulo 3, vimos que el estimador $\widehat{g}_{\mathrm{M}_0}$ es un estimador uniformemente consistente para toda función t continua tal que $i(t) > 0$. Luego, en una primera instancia de estudio es posible elegir entre los siguientes dos estimadores, definidos como en 2.14,

- $\widehat{g}_{\alpha,\mathrm{M}_0}$: estimador que considera $t(\mathbf{x}) \equiv 1$,
- $\widehat{g}_{\alpha,\mathrm{M}_0,2}$: estimador que toma $t(\mathbf{x}) = p(\mathbf{x})$, es decir, que la función $t(\cdot)$ es igual al mecanismo de pérdida utilizado y que en realidad, es desconocido.

Observemos que para el caso en el que $p = p_1$, es decir, cuando no hay pérdida de datos, ambos estimadores coinciden.

Las Tablas 6.1 y 6.2 se obtuvieron mediante un estudio de simulación preliminar en el que se arrojaron NREP= 100 muestras de tamaño $n = 500$ siguiendo el procedimiento descripto en la Sección 6.1 y con el mecanismo de pérdida p_2.

	$\widehat{g}_{\mathrm{M}_0,2} < \widehat{g}_{\mathrm{M}_0}$	$\widehat{g}_{1,\mathrm{M}_0,2} < \widehat{g}_{1,\mathrm{M}_0}$	$\widehat{g}_{2,\mathrm{M}_0,2} < \widehat{g}_{2,\mathrm{M}_0}$
C_0	47.4%	50.2%	48.4%
C_1	52.4%	50.4%	55.2%
C_2	44.0%	47.6%	42.4%
C_3	47.8%	53.4%	48.0%
C_4	52.0%	53.4%	52.0%

Tabla 6.1: Porcentaje de veces, entre las 100 replicaciones, que los estimadores que utilizan $t = p$ tienen menores ISE que aquellos estimadores que usan $t \equiv 1$. La columna de la izquierda corresponde al estimador de la función de regresión mientras que las otras dos columnas corresponden a los estimadores de las componentes aditivas g_1 y g_2.

	$\widehat{g}_{\mathrm{M}_0}$	$\widehat{g}_{\mathrm{M}_0,2}$	$\widehat{g}_{1,\mathrm{M}_0}$	$\widehat{g}_{1,\mathrm{M}_0,2}$	$\widehat{g}_{2,\mathrm{M}_0}$	$\widehat{g}_{2,\mathrm{M}_0,2}$
C_0	0.3785	0.3797	0.2622	0.2615	0.2407	0.2446
C_1	16.5560	16.5336	4.9844	4.9737	4.9543	4.9470
C_2	16.0584	16.1461	5.5624	5.5883	6.2723	6.3139
C_3	0.5967	0.5970	0.3553	0.3534	0.3372	0.3402
C_4	2.3138	2.3069	1.0285	1.0229	1.0676	1.0680

Tabla 6.2: MISE de los estimadores bajo distintos tipos de contaminación. La columna de la izquierda corresponde a los MISE obtenidos al estimar la función de regresión, mientras que las dos columnas restantes corresponden a los MISE obtenidos al estimar las componentes aditivas g_1 y g_2.

En la Tabla 6.1 podemos ver que la cantidad de veces que un estimador es superior al otro en términos de ISE es la misma. Mientras que para estimar la primera componente aditiva g_1 pareciera ser mejor el estimador $\widehat{g}_{1,\mathrm{M}_0,2}$, para estimar la segunda componente aditiva g_2 el estimador $\widehat{g}_{2,\mathrm{M}_0}$ tiene menores ISE. Si miramos sólo los estimadores de la función de regresión g, mientras que la Tabla 6.1 refleja que en tres de los cinco mecanismos de contaminación el estimador $\widehat{g}_{\mathrm{M}_0,2}$ tiene menores ISE, en la columna de la izquierda de la Tabla 6.2 se puede ver que el promedio de los ISE, es decir los MISE, son menores cuando se utiliza el estimador $\widehat{g}_{\mathrm{M}0}$. En las otras dos columnas de la Tabla 6.2 se puede ver que el comportamiento de ambos estimadores en términos de los MISE es similar.

De esta manera, para el resto del estudio de simulación elegimos quedarnos con el representante $\widehat{g}_{1,\mathrm{M}_0}$ de los estimadores robustos basados en integración marginal definidos en 2.14, dado que en el estudio preliminar realizado se pudo ver que ambas propuestas se comportan de manera similar, pero el estimador que considera el verdadero mecanismo de pérdida de datos p puede considerarse como un estimador más artificial que aquel que no usa esta información.

6.2 Resultados

A modo de ejemplo y dado que los estimadores clásicos se comportan de manera similar entre sí, así como también los $M-$ estimadores robustos se comportan de manera similar entre sí, en las Figuras 6.2, 6.3 y 6.4, los colores corresponden a los ajustes dados por algunos de los estimadores estudiados sobre la muestra considerada en la Figura 6.1,

- en *rosa*, el estimador $\widetilde{g}_{\mathrm{S}}$,
- en *rojo*, el estimador $\widehat{g}_{\mathrm{S}}$,
- en *marrón*, el estimador $\widehat{g}_{\mathrm{BC}}$,
- en *azul*, el estimador $\widehat{g}_{\mathrm{M}_0}$ y
- en *verde*, el estimador $\widehat{g}_{\mathrm{BR}}$.

Las Figuras 6.2, 6.3 y 6.4 muestran el comportamiento de estos estimadores sobre la variable respuesta y y sobre las componentes del modelo aditivo, g_1 y g_2, para cada uno de estos estimadores y para cada uno de los diferentes mecanismos de contaminación, cuando no hay datos faltantes. Como se puede observar, cuando no hay contaminación los estimadores tienen un comportamiento similar y ajustan adecuadamente las funciones $g : 1$ y g_2. Respecto de los estimadores robustos el estimador de Backfitting robusto parece ser el más estable en todas las contaminaciones consideradas.

Las Tablas 6.3, 6.4 y 6.5 presentan los MISE para los 12 estimadores estudiados bajo los distintos esquemas de pérdida de datos y para los diferentes mecanismos de contaminación utilizados, al estimar la función de regresión, la función aditiva g_1 y la función aditiva g_2, respectivamente. Todos los siguientes estimadores se calcularon usando el núcleo de Epanechnikov.

	$p_1(\mathbf{x}) \equiv 1$											
	$\widetilde{g}_{\mathrm{S}}$	$\widetilde{g}_{\mathrm{LS}_1}$	$\widetilde{g}_{\mathrm{M}_0}$	$\widetilde{g}_{\mathrm{M}_1}$	$\widehat{g}_{\mathrm{S}}$	$\widehat{g}_{\mathrm{LS}_1}$	$\widehat{g}_{\mathrm{LS}_{q,\alpha}}$	$\widehat{g}_{\mathrm{BC}}$	$\widehat{g}_{\mathrm{M}_0}$	$\widehat{g}_{\mathrm{M}_1}$	$\widehat{g}_{\mathrm{M}_{q,\alpha}}$	$\widehat{g}_{\mathrm{BR}}$
C_0	0.0976	0.0306	0.1062	0.0318	0.0738	0.0100	0.0109	0.0704	0.0790	0.0107	0.0113	0.0706
C_1	7.8772	8.8536	0.6083	2.2497	6.1693	6.3800	6.2519	6.1410	0.2526	0.9682	0.8573	0.1532
C_2	14.4716	14.4008	17.6840	15.0068	8.3862	8.2827	8.3343	8.4660	12.0867	8.3349	8.3339	0.2488
C_3	0.3614	0.2938	0.2321	0.2149	0.1573	0.0899	0.0939	0.1576	0.1096	0.0588	0.0511	0.0731
C_4	1.6789	1.6546	0.3986	0.7569	0.9003	0.8459	0.8459	0.9044	0.2090	0.2773	0.2431	0.0782
	$p_2(\mathbf{x}) = 0.4 + 0.5\cos^2(x_1 + 0.2)$											
	$\widetilde{g}_{\mathrm{S}}$	$\widetilde{g}_{\mathrm{LS}_1}$	$\widetilde{g}_{\mathrm{M}_0}$	$\widetilde{g}_{\mathrm{M}_1}$	$\widehat{g}_{\mathrm{S}}$	$\widehat{g}_{\mathrm{LS}_1}$	$\widehat{g}_{\mathrm{LS}_{q,\alpha}}$	$\widehat{g}_{\mathrm{BC}}$	$\widehat{g}_{\mathrm{M}_0}$	$\widehat{g}_{\mathrm{M}_1}$	$\widehat{g}_{\mathrm{M}_{q,\alpha}}$	$\widehat{g}_{\mathrm{BR}}$
C_0	0.1380	0.0708	0.1317	0.0741	0.0860	0.0260	0.0281	0.0785	0.0930	0.0261	0.0283	0.0806
C_1	9.2740	11.1653	2.2914	5.2306	6.5290	6.9394	6.7483	6.4416	0.8896	2.1564	1.7846	0.1681
C_2	14.6767	14.6164	17.9347	15.2247	8.4855	8.3481	8.4418	8.6032	13.5514	8.4196	8.3339	0.2832
C_3	0.4089	0.3432	0.2999	0.2724	0.1739	0.1075	0.1142	0.1678	0.1366	0.0785	0.0731	0.0831
C_4	1.8306	1.8436	0.6199	1.0114	0.9596	0.9060	0.9086	0.9472	0.3175	0.3829	0.3326	0.0880

Tabla 6.3: MISE de los estimadores de la función de regresión bajo distintos esquemas de pérdida de datos y distintos tipos de contaminación.

Para analizar los resultados de las simulaciones realizadas, comenzaremos por el caso en que el esquema de pérdida de datos utilizado es p_1, es decir, cuando no hay datos faltantes. Cuando no hay datos contaminados, los estimadores que utilizan el procedimiento de integración marginal tienen

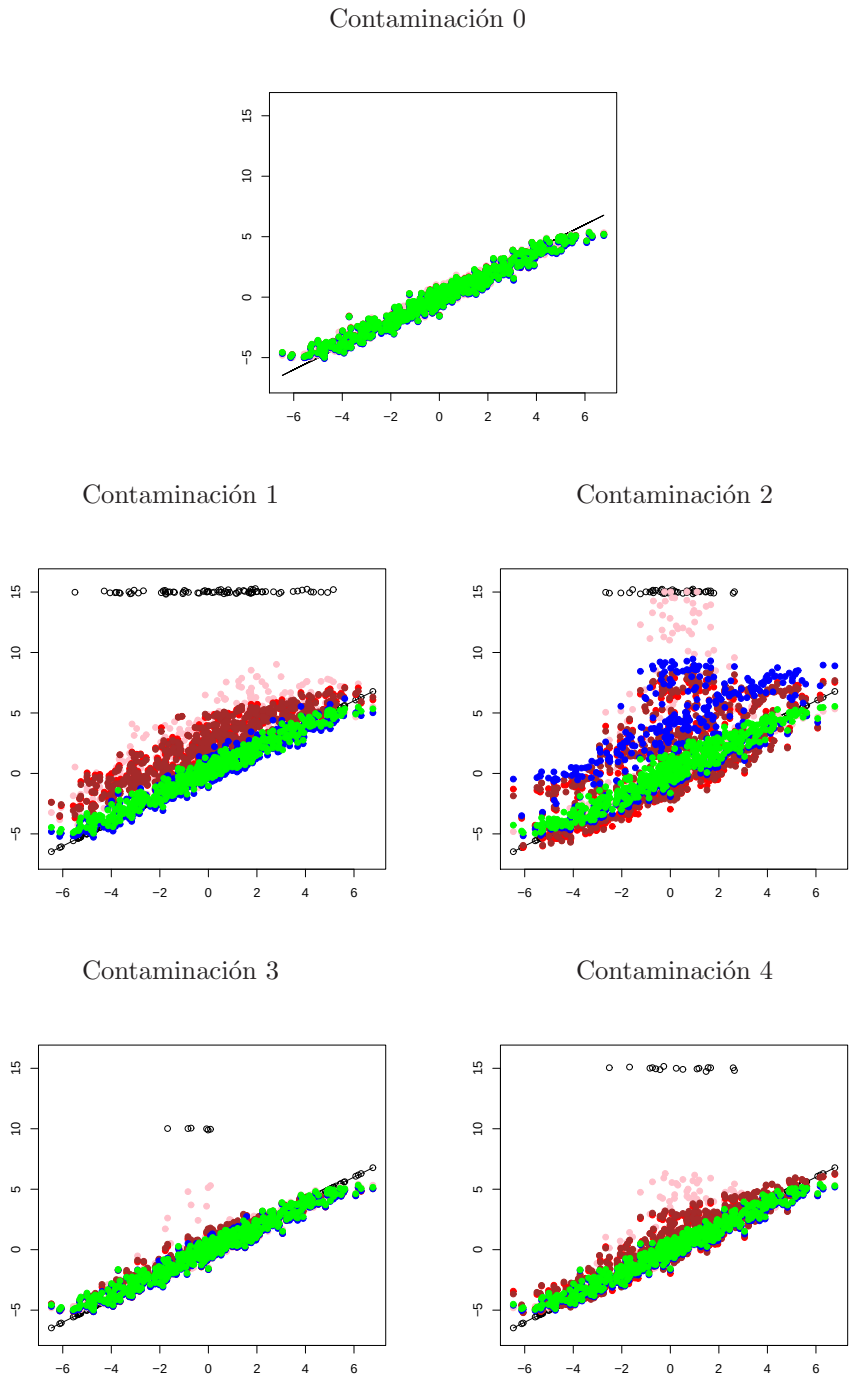

Figura 6.2: Gráficos de los valores predichos obtenidos para los diferentes estimadores, bajo los diferentes mecanismos de contaminación. En el eje horizontalse encuentran los valores y dados en la muestra, mientras que en el eje vertical se encuentran los valores obtenidos por los diferentes estimadores. Los puntos en negro corresponden a los puntos que han sido contaminados.

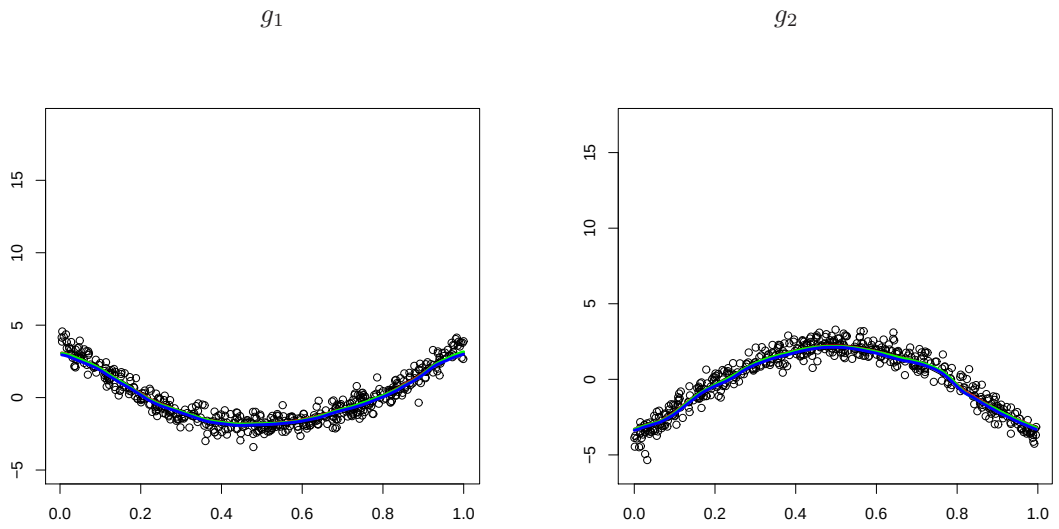

Figura 6.3: Gráficos de los ajustes obtenidos por los distintos estimadores para las dos componentes aditivas g_1 y g_2 cuando no hay datos contaminados.

	$p_1(\mathbf{x}) \equiv 1$							
	$\widehat{g}_{1,\mathrm{S}}$	$\widehat{g}_{1,\mathrm{LS}_1}$	$\widehat{g}_{1,\mathrm{LS}_{q,\alpha}}$	$\widehat{g}_{1,\mathrm{BC}}$	$\widehat{g}_{1,\mathrm{M}_0}$	$\widehat{g}_{1,\mathrm{M}_1}$	$\widehat{g}_{1,\mathrm{M}_{q,\alpha}}$	$\widehat{g}_{1,\mathrm{BR}}$
C_0	0.0466	0.0109	0.0146	0.0445	0.0513	0.0112	0.0158	0.0445
C_1	0.5099	0.5503	0.5034	0.4896	0.1860	0.2534	0.2148	0.0503
C_2	3.4949	3.2939	3.3179	3.5207	4.6340	3.3658	3.3736	0.1309
C_3	0.0953	0.0480	0.0531	0.0943	0.0702	0.0339	0.0344	0.0468
C_4	0.4333	0.3544	0.3589	0.4319	0.1220	0.1301	0.1193	0.0514
	$p_2(\mathbf{x}) = 0.4 + 0.5\cos^2(x_1 + 0.2)$							
	$\widehat{g}_{1,\mathrm{S}}$	$\widehat{g}_{1,\mathrm{LS}_1}$	$\widehat{g}_{1,\mathrm{LS}_{q,\alpha}}$	$\widehat{g}_{1,\mathrm{BC}}$	$\widehat{g}_{1,\mathrm{M}_0}$	$\widehat{g}_{1,\mathrm{M}_1}$	$\widehat{g}_{1,\mathrm{M}_{q,\alpha}}$	$\widehat{g}_{1,\mathrm{BR}}$
C_0	0.0533	0.0209	0.0245	0.0508	0.0601	0.0207	0.0253	0.0519
C_1	0.6873	0.8114	0.7330	0.6570	0.3533	0.5663	0.4604	0.0619
C_2	3.5645	3.3625	3.4136	3.6193	4.6978	3.4343	3.3736	0.1447
C_3	0.1056	0.0605	0.0673	0.1023	0.0834	0.0463	0.0478	0.0543
C_4	0.4715	0.3986	0.4049	0.4625	0.1661	0.1866	0.1676	0.0592

Tabla 6.4: MISE de los estimadores de la función de la componente aditiva g_1 bajo distintos esquemas de pérdida de datos y distintos tipos de contaminación.

MISE menores que sus versiones preliminares, previas al procedimiento de integración (ver Tabla 6.3). Podemos observar también que los estimadores basados en polinomios locales de orden 1 en todas las coordenadas y de orden 1 en la dirección de interés tienen un mejor comportamiento que sus respectivas versiones basadas en polinomios locales de grado 0. El MISE del estimador $\widehat{g}_{\mathrm{LS}_1}$ es menor a los MISE de los restantes estimadores, seguido por su versión robusta $\widehat{g}_{\mathrm{M}_1}$ y por el estimador clásico basado en polinomios locales de orden 1 en la dirección de interés $\widehat{g}_{\mathrm{LS}_{q,\alpha}}$. Además, se puede observar que los MISE de los dos estimadores basados en backfitting son ligeramente inferiores a los de los estimadores de la función de regresión basados en polinomios locales de orden 0 que utilizan el procedimiento de integración marginal. Por otra parte, como es de esperar los estimadores robustos han perdido eficiencia, medida en términos de los MISE, comparados con sus versiones clásicas, o

	$p_1(\mathbf{x}) \equiv 1$							
	$\widehat{g}_{1,\mathrm{S}}$	$\widehat{g}_{1,\mathrm{LS}_1}$	$\widehat{g}_{1,\mathrm{LS}_{q,\alpha}}$	$\widehat{g}_{1,\mathrm{BC}}$	$\widehat{g}_{1,\mathrm{M}_0}$	$\widehat{g}_{1,\mathrm{M}_1}$	$\widehat{g}_{1,\mathrm{M}_{q,\alpha}}$	$\widehat{g}_{1,\mathrm{BR}}$
C_0	0.0422	0.0117	0.0158	0.0405	0.0464	0.0120	0.0172	0.0406
C_1	0.5316	0.6405	0.6023	0.5174	0.2209	0.3317	0.2663	0.0494
C_2	3.0783	3.1853	3.2032	3.1294	4.3202	3.2565	3.2718	0.0903
C_3	0.0690	0.0461	0.0517	0.0688	0.0583	0.0387	0.0349	0.0403
C_4	0.3157	0.3349	0.3374	0.3181	0.0991	0.1244	0.1147	0.0391
	$p_2(\mathbf{x}) = 0.4 + 0.5\cos^2(x_1 + 0.2)$							
	$\widehat{g}_{1,\mathrm{S}}$	$\widehat{g}_{1,\mathrm{LS}_1}$	$\widehat{g}_{1,\mathrm{LS}_{q,\alpha}}$	$\widehat{g}_{1,\mathrm{BC}}$	$\widehat{g}_{1,\mathrm{M}_0}$	$\widehat{g}_{1,\mathrm{M}_1}$	$\widehat{g}_{1,\mathrm{M}_{q,\alpha}}$	$\widehat{g}_{1,\mathrm{BR}}$
C_0	0.0473	0.0177	0.0213	0.0469	0.0562	0.0179	0.0223	0.0483
C_1	0.7093	0.9188	0.8472	0.6727	0.3798	0.6698	0.5371	0.0610
C_2	3.1084	3.1897	3.2142	3.2261	5.3807	3.2655	3.2718	0.1185
C_3	0.0756	0.0519	0.0568	0.0778	0.0770	0.0403	0.0412	0.0482
C_4	0.3344	0.3502	0.3512	0.3440	0.1568	0.1617	0.1452	0.0471

Tabla 6.5: MISE de los estimadores de la función de la componente aditiva g_2 bajo distintos esquemas de pérdida de datos y distintos tipos de contaminación.

lineales.

Las Tablas 6.4 y 6.5 permiten concluir que los MISE de los estimadores de las componentes aditivas g_1 y g_2 basados en polinomios locales de orden 1, tanto en todas las coordenadas como en la de interés, son los menores. También podemos ver que los estimadores robustos han perdido eficiencia respecto de sus versiones clásicas. En la Figura 6.2 se grafica en el eje horizontal los valores obtenidos en una de las 500 muestras generadas y en el eje vertical los valores predichos obtenidos por cada uno de los 5 estimadores mencionados al principio de esta Sección. En la Figura 6.3 se pueden ver las curvas correspondientes a las estimaciones realizadas por estos 5 estimadores para las componentes aditivas g_1 y g_2. En ambas figuras, correspondientes a C_0, los puntos y curvas quedaron superpuestas. Esto se debe a que, en este caso, todos estimadores se comportan de manera similar.

Para el caso de contaminación C_1 los estimadores clásicos de la función de regresión g, $\widetilde{g}_{\mathrm{S}}$, $\widetilde{g}_{\mathrm{LS}_1}$, $\widehat{g}_{\mathrm{S}}$, $\widehat{g}_{\mathrm{LS}_1}$, $\widehat{g}_{\mathrm{LS}_{q,\alpha}}$ y $\widehat{g}_{\mathrm{BC}}$, tienen valores de MISE hasta más de 24 veces mayores que los de sus versiones robustas. Para las estimaciones de las componentes aditivas, esta diferencia sigue siendo importante, pero se reduce a casi 3 veces más. Se puede sospechar que la falta de robustez de los estimadores de posición utilizados en los estimadores clásicos tiene un importante efecto sobre los altos MISE obtenidos en estos estimadores. Entre los estimadores robustos estudiados, tanto para estimar la función de regresión g como para estimar sus componentes aditivas g_1 y g_2, el estimador robusto basado en backfitting posee menores errores MISE que los estimadores basados en integración marginal, seguido por el $M-$estimador basado en polinomios locales de grado 0 que utiliza el procedimiento de integración.

Para el caso de la muestra considerada, las Figuras 6.2 y 6.4 permiten clarificar los resultados obtenidos en la Tablas, observando el comportamiento de algunos de los estimadores utilizados. En *azul* y *verde* tenemos los estimadores robustos $\widehat{g}_{\mathrm{M}_0}$ y $\widehat{g}_{\mathrm{BR}}$, respectivamente. Para el caso de contaminación C_1 se puede apreciar que los puntos y curvas de estos dos colores se encuentran más cercanas a la recta identidad en la Figura 6.2 y más cercana a las curvas g_1 y g_2 en la Figura 6.4. Un análisis similar se puede hacer para las contaminaciones C_3 y C_4. En ambos casos, la Tabla 6.3 refleja que los estimadores robustos arrojaron valores de MISE menores que sus versiones

clásicas, manteniendo los errores cercanos a los obtenidos bajo C_0 y que los estimadores que no hacen uso de que el modelo es aditivo tienen errores mayores que los restantes estimadores que sí hacen uso de esta información. Las Figuras 6.2 y 6.4 muestran un comportamiento similar para los 5 estimadores graficados que lo estudiado para el caso C_1. Para el caso de contaminación C_2 se puede ver que tanto los estimadores clásicos como los $M-$estimadores robustos tienen valores de MISE mayores que el estimador de backfitting robusto. Lo mismo se puede observar en los MISE de las componentes aditivas. En la Figura 6.2 se puede ver que, para el caso de una muestra, para este tipo de contaminación, el estimador robusto $\widehat{g}_{\mathrm{M}_0}$ (en *azul*) se alejó notoriamente de la recta identidad, y en la Figura 6.4 se puede ver que este mismo comportamiento errado lo tuvo también a la hora de estimar las componentes aditivas, dirigiéndose la curva, en ambos casos, hacia los puntos que habían sido contaminados.

En conclusión, cuando no hay datos faltantes, bajo los mecanismos de contaminación C_1, C_3 y C_4, los estimadores robustos tienen valores menores de MISE que sus versiones clásicas. Entre los estimadores robustos, el estimador $\widehat{g}_{\mathrm{BR}}$ tiene menor MISE que los estimadores basados en integración marginal. Esto muestra la sensibilidad de los estimadores clásicos a la presencia de datos atípicos, y que los menores errores del estimador robusto basado en backfitting pueden deberse a que este estimador utiliza iteraciones de estimadores robustos unidimensionales. Por otro lado, en el caso de la contaminación patológica C_2, los $M-$estimadores no son "resistentes"ante el efecto de esta contaminación. Esto ocurre porque todas las respuestas correspondientes a covariables en el cuadrado de lado 0.30 están contaminadas y, por lo tanto, que el procedimiento de integración marginal no resuelve los valores grandes arrojados por los $M-$estimadores preliminares que utilizan como ventana $h = 0.10$. Respecto al comportamiento de los estimadores sobre las componentes aditivas g_1 y g_2, se puede realizar un análisis similar para los cuatro tipo de contaminación considerados.

Para el caso en que hay datos faltantes, se puede hacer el mismo análisis que el realizado para el caso en que no hay datos faltantes para comparar a los distintos estimadores y, por lo tanto, no lo repetiremos para este caso. Si bien el estimador de backfitting robusto se introdujo en el caso de datos completos, en este caso, como se menciona en la Observación 5.2.3 el estimador simplificado que se obtiene utilizando solamente las respuestas observadas continúa siendo Fisher–consistente y por ello, se incluyó en esta comparación. Por otra parte, es de interés comparar los errores obtenidos para estos estimadores cuando no hay respuestas faltantes, es decir, cuando $p = p_1$ y cuando hay respuestas faltantes, en este caso, cuando usamos el mecanismo de pérdida p_2. Las Tablas 6.3, 6.4 y 6.5 permiten observar que los estimadores tienen mayores de MISE que cuando $p = p_1 \equiv 1$ indicando pérdida de eficiencia debido a las observaciones faltantes. Por otra parte, los estimadores $\widehat{g}_{\mathrm{M}_0}$, $\widehat{g}_{\mathrm{M}_1}$ y $\widehat{g}_{\mathrm{M}_{q,\alpha}}$ han aumentado en mayor medida su MISE que los demás estimadores, es decir, han tenido una mayor pérdida de eficiencia, mientras que el estimador que ha aumentado en menor medida sus errores es $\widehat{g}_{\mathrm{BC}}$. Esto mismo se puede observar cuando se analizan los errores obtenidos al estimar las componentes aditivas en las Tablas 6.4 y 6.5.

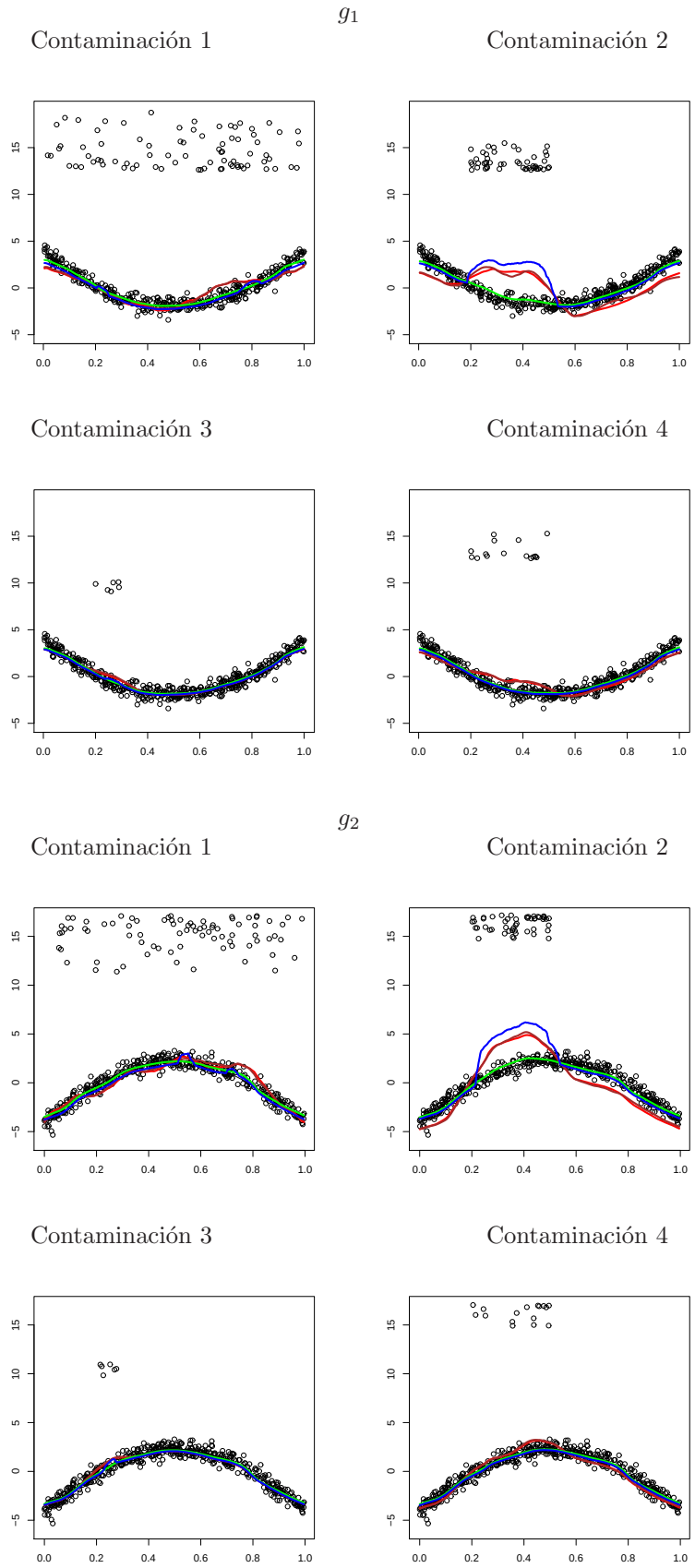

Figura 6.4: Gráficos de los ajustes obtenidos bajo los diferentes tipos de contaminación. Los cuatro gráficos superiores corresponden a las estimaciones obtenidas para la primera componente aditiva g_1, mientras que los cuatro gráficos inferiores corresponden a las estimaciones obtenidas para la segunda componente aditiva g_2.

Bibliografía

[1] Baek, J. y Wehrly, T. (1993). Kernel estimation for additive models under dependence. *Stochastic Processes and their Applications*, **47**, 95-112.

[2] Bianco A., y Boente, G. (1998). Robust Kernel Estimators for Additive Models with Dependent Observations. *The Canadian Journal of Statistics*, **6**, 239-255.

[3] Billingsley, P. (1968). *Convergence of Probability Measures*. Wiley, New York.

[4] Boente, G. y Fraiman, R. (1989). Robust nonparametric regression estimation. *Journal of Multivariate Analysis*, **29**, 180-198.

[5] Boente, G. y Fraiman, R. (1990). Asymptotic distribution of robust estimates for nonparametric models from mixing observations. *Annals of Statistics*, **18**, 2, 891-906.

[6] Boente, G. y Fraiman, R. (1991). Strong Uniform Convergence Rates for Some Robust Equivariant Nonparametric Regression Estimates for Mixing Processes. *International Satistical Review*, **59**, 355-372.

[7] Boente, G., González-Manteiga, W. y Pérez-González, A. (2009). Robust nonparametric estimation with missing data. *Journal of Statistical Planning and Inference*, **139**, 571-592.

[8] Boente, G. y Martínez, A. (2013). Estimating additive models with missing responses. Disponible en `www.ic.fcen.uba.ar/preprints/Paper_aditivo_Boente_Martinez.pdf`

[9] Breiman, L. and Friedman, J. (1985). Estimating optimal transformations for multiple regression and correlation. *J. Amer. Statist. Assoc.*, **809**, 580-597.

[10] Buja, A., Hastie, T. y Tibshirani, R. (1989). Linear smoothers and additive models (with discussion). *Annals of Statistics*, **17**. 453-555.

[11] Cheng, P. E. y Chu, C.K. (1996). Kernel estimation of distribution functions and quantiles with missing data. *Statist. Sinica*, **6**, 63-78.

[12] Ferraty, F. y Vieu, P. (2006). *Nonparametric Functional Data Analysis*. Springer Series in Statistics.

[13] Härdle, W., Müller, M., Sperlich, S. y Werwatz, A. (2004). *Nonparametric and Semiparametric Models*. Springer.

[14] Harrison, D. y Rubinfeld, D. L. (1978). Hedonic prices and the demand for clean air. *J. Environ. Economics and Management*, **5**, 81-102.

[15] Hastie, T.J. y Tibshirani, R.J. (1986). Generalized Additive Models. *Statistical Science*, **1**, 297-310.

[16] Hastie, T.J. y Tibshirani, R.J. (1990). *Generalized Additive Models*. Monographs on Statistics and Applied Probability No. 43. Chapman and Hall, London.

[17] Hengartner, N. y Sperlich, S. (2005). Rate optimal estimation with the integration method in the presence of many covariates. *Journal of Multivariate Analysis*, **95**, 246-272.

[18] Huber, P. J. (1981) *Robust Statistics*. New York: John Wiley & Sons, Inc.

[19] Kong, E., Linton, O. y Xia, Y. (2010). Uniform Bahadur representation for local polynomial estimates of $M-$regression and its application to the additive model. *Econometric Theory*, **26**, 1529-1564.

[20] Koul, H., Müller, U. y Schick, A. (2012). The Transfer Principle: a tool for complete case analysis. *The Annals of Statistics*. **40**, 6, 3031-3049.

[21] Li, J., Zheng, Z. y Zheng, M. (2012) Robust estimation of additive models based on marginal integration. Publicación previa, disponible en `http://www.math.pku.edu.cn:8000/var/preprint/7065.pdf`

[22] Linton, O. y Nielsen, J. (1995). A kernel method of estimating structured nonparametric regression based on marginal integration. *Biometrika*, **82**, 93-101.

[23] Mack, Y. and Silverman, B. (1982). Weak and strong consistency of kernel regression estimates. *Z. Wahrsch. verw. Gebiete*, **61**, 405-415.

[24] Maronna, R., Martin, R. y Yohai, V. (2006) *Robust Statistics, Theory and Methods*. John Wiley & Sons, Ltd.

[25] Neyman, J. (1938). Contribution to the theory of sampling human populations. *J. Amer. Statist. Assoc.*, **33**, 101-116.

[26] Opsomer, J.D. (2000). Asymptotic properties of backfitting estimators. *J. Multivariate Anal.*, **73**, 166-179.

[27] Opsomer, J.D. y Ruppert, D. (1997). Fitting a bivariate additive model by local polynomial regression. *The Annals of Statistics*, **25**, 186-211.

[28] Severance-Lossin, E. y Sperlich, S. (1999). Estimation of derivatives for additive separable models, *Statistics*, **33**, 241-265.

[29] Sperlich, S., Linton, O. y Härdle, W. (1999). Integration and backfitting methods in additive models-finite sample properties and comparison. *TEST*, **8**, 419-458.

[30] Sperlich, S. Tjøstheim, D. y Yang, L. (2002) Nonparametric estimation and testing of interaction in additive models. *Econometric Theory*, **18**, 197-251.

[31] Stone, C.J. (1985). Additive regression and other nonparametric models. *Annals of Statistics*, **13**, 689-705.

[32] Tjøstheim, D. y Auestad, B. (1994) Nonparametric identification of nonlinear time series: Selecting significant lags. *Journal of the American Statistical Association*, **89**, 1410-1430.