

Métodos de Estimación Adaptativos en Modelos de Regresión



Miguel Angel Padrino Garcia

Supervisor: Prof. Karine Bertin

Advisor: Prof. Lisandro Fermín

Instituto de Estadística
Universidad de Valparaíso

This dissertation is submitted for the degree of
Doctor en Estadística

Resumen Ejecutivo

En el presente trabajo se plantea el problema de estimación no paramétrica (por núcleo) de la función de regresión en un modelo de regresión univariante. Nos interesa estudiar las propiedades asintóticas de los estimadores propuestos usando el error cuadrático medio (error local en un punto fijo $x \in \mathbb{R}$). Dos enfoques pueden ser considerados: caso no-adaptativo y caso adaptativo. En el caso no-adaptativo, se conoce la regularidad de la función a estimar y los estimadores con propiedades óptimas (es decir con error cuadrático medio mínimo) dependen de esta regularidad, Tsybakov (2009). En el caso adaptativo, se desconoce tal regularidad, pero se trata de encontrar un estimador que se adapta a la regularidad y que tiene propiedades óptimas al igual que si uno conociera la regularidad; en este caso, el interés es estudiar estimadores óptimos en error puntual.

La mayoría de los trabajos existentes que proponen estimadores adaptativos lo hacen usando un riesgo integrado (en L_2 por ejemplo), ver Baraud et al. (2001) Baraud (2002). Pocos son los resultados que se tienen de estimación adaptativa de la función de regresión usando un error puntual. Lepski and Serdyukova (2014) proponen un procedimiento de estimación adaptativa de la función de regresión para datos independientes con función de densidad conocida g , común para las variables explicativas. Nguyen (2014) investiga el mismo problema donde se asume que g es la densidad uniforme.

Al igual que ambos artículos, en este trabajo se propone usar el mismo procedimiento presentado por Goldenshluger and Lepski (2011). Tal metodología también recibe el nombre de método Goldenshluger-Lepski. En la última década, el método Goldenshluger-Lepski (GL) se ha utilizado en varios modelos estadísticos (densidad, regresión, densidad condicional, modelos de ruido blanco, entre otros) para derivar desigualdades de oráculo (y, por lo tanto, estimadores adaptativos) especialmente en contextos independientes (ver, por ejemplo, Goldenshluger and Lepski (2011), Goldenshluger and Lepski (2013), Bertin et al. (2016), Chichignoud et al. (2017)) y también en algunos casos dependientes, como en Bertin and Klutchnikoff (2017), Asin and Johannes (2016), Comte et al. (2017), Bertin et al. (2020a) y Bertin et al. (2020b).

Nuestro objetivo en esta tesis es investigar la estimación adaptativa de la función de regresión en riesgo puntual, cuando la función de densidad g es conocida o desconocida y considerar el caso de datos débilmente dependientes (noción de dependencia definida en Doukhan and Louhichi (1999)).

El capítulo 1, *Formulación general del problema y preliminares*, está dividido en dos secciones. En la Sección 1.1, se plantea el problema de estimación no paramétrica adaptativa de la función de regresión para un modelo de regresión univariante, en los casos de interés en este trabajo. En la Sección 1.2, se presentan los distintos estimadores de tipo núcleo que serán usados para la estimación de la función de regresión. Cuando la función de densidad de la variable explicativa es conocida, tanto en contexto de independencia como de dependencia, se presenta una modificación del estimador de Nadaraya-Watson (Nadaraya (1989)) planteada en Lepski and Serdyukova (2014). Además, se plantea otra modificación del estimador de Nadaraya-Watson que será usada para la estimación de la función de regresión cuando la densidad de la variable explicativa es desconocida, en contexto de dependencia. Se definen las propiedades básicas de los estimadores, tales como sesgo, varianza y consistencia. Se demuestra que los estimadores por núcleo de la función de regresión son consistentes, su error cuadrático medio tiende a cero; además, se proporciona un esquema y se dan referencias para obtener la velocidad de convergencia del estimador por núcleo de la regresión. Se plantean las nociones de dependencia débil dadas en Doukhan and Louhichi (1999), Dedecker et al. (2007) y Bertin and Klutchnikoff (2017). Finalmente, se proporciona una desigualdad de Bernstein para datos independientes dada por Massart (2007) y otra para datos débil dependientes dada por Doukhan and Neumann (2007), las cuales serán útiles para obtener cotas del término estocástico de los estimadores no paramétricos y para la calibración de los estimadores.

El capítulo 2, *Estimación adaptativa en modelos de regresión para datos independientes y variable explicativa con densidad conocida*, conforma una introducción a la metodología que será utilizada después en el caso de datos dependientes. En este capítulo, se estudia el problema bajo las hipótesis de que la variable explicativa es independiente, con función de densidad conocida, y los errores del modelo son acotados. Está dividido en cuatro secciones. En la Sección 2.1, *Modelo*, se plantea el modelo de regresión univariante y las distintas hipótesis propuestas. En la Sección 2.2, *Procedimiento estadístico*, se presenta con detalles el método GL para la estimación no paramétrica adaptativa de la función de regresión, el estimador, la familia de ventanas y se describe la metodología para seleccionar las ventanas. En la Sección 2.3, se determina el sesgo y la varianza del estimador, se demuestra la consistencia y la velocidad de convergencia del estimador; además, en la Sección 2.3.2 se plantea una desigualdad de oráculo y se demuestra que el estimador de la regresión satisface tal desigualdad, lo que al final permite demostrar que el estimador es adaptativo. En la Sección 2.4 *Demostraciones*, está reunida la mayoría de las demostraciones de los lemas, proposiciones y teoremas de este capítulo. Se debe destacar que en este capítulo al ser el error del modelo acotado y la función

de regresión acotada, entonces la variable respuesta también es acotada, lo que permite que se pueda usar de forma directa la desigualdad de Bernstein para variables aleatorias independientes, en el control del término estocástico del estimador no paramétrico de la regresión.

El capítulo 3, *Estimación adaptativa en modelos de regresión para datos débilmente dependientes y variable explicativa con densidad conocida*, está dividido en cinco secciones. En la Sección 3.1, *Modelo*, se plantea el modelo de regresión univariante y se dan hipótesis tales como, débil dependencia de la variable explicativa de densidad conocida y acotada, función de regresión acotada, error del modelo i.i.d de distribución normal, junto a otras hipótesis necesarias para la estimación no paramétrica adaptativa de la regresión. Además, se proporciona un ejemplo de proceso débil dependiente. En la Sección 3.2, *Procedimiento estadístico*, se presenta con detalles el método GL para la estimación no paramétrica adaptativa de la función de regresión en contexto de dependencia débil, se da el estimador, la familia de ventanas y la descripción de la metodología para seleccionar las ventanas. En la Sección 3.3, *Resultados*, se determina el sesgo y la varianza del estimador en el contexto de dependencia débil, se demuestra la consistencia y la velocidad de convergencia del estimador, se plantea una desigualdad de oráculo y se demuestra que el estimador de la regresión satisface tal desigualdad, lo que al final permite demostrar que el estimador es adaptativo. En la Sección 3.4, *Estudio de Simulación*, se plantea el esquema de simulación para generar datos del modelo de regresión con función de regresión acotada, variable explicativa de densidad acotada, y débil dependientes de tipo α -mixing con decaimiento exponencial. Se explicita como obtener el estimador no paramétrico, con núcleo gaussiano, de la función de regresión usando el método GL y los datos simulados en el esquema de simulación planteado previamente. Posteriormente se muestra cómo se calibra el método usando los mismos datos simulados. Finalmente, se presenta un estudio comparativo de resultados de estimaciones realizadas, usando errores globales y locales empíricos, para distintos tamaños de muestras simuladas y distintos valores de la desviación estándar del error del modelo de regresión. La Sección 3.5, *Demostraciones*, reúne las demostraciones de los lemas, proposiciones, teoremas y corolarios de este capítulo. Se debe destacar que en este capítulo al no ser acotada la variable respuesta del modelo, se introduce la definición de truncamiento del estimador de la función de regresión, lo que permite descomponer el estimador por núcleo de la función de regresión en dos términos, uno truncado y otro no acotado. El término estocástico no acotado del estimador se controla con la hipótesis de decaimiento exponencial del coeficiente de dependencia débil y el término estocástico del estimador de regresión truncado es controlado con una adaptación que se realizó de la desigualdad de Bernstein para datos débilmente

dependientes de Doukhan and Neumann (2007).

El capítulo 4, *Estimación adaptativa en modelos de regresión para datos débilmente dependientes y variable explicativa con densidad desconocida*, esta dividido en cinco secciones. En la Sección 4.1, *Modelo*, se plantea el modelo de regresión univariante y se dan hipótesis tales como, débil dependencia de la variable explicativa de densidad $g(\cdot)$ desconocida y acotada, función de regresión $r(\cdot)$ acotada, error del modelo i.i.d de distribución normal. Se muestra que la función de regresión se puede escribir de la forma $r(\cdot) = \frac{m(\cdot)}{g(\cdot)}$, donde $m(x) = \int_{\mathbb{R}} y f_{X,Y}(x,y) dy$, con $f_{X,Y}(\cdot, \cdot)$ la densidad conjunta de las variables explicativa y respuesta del modelo, y se plantean mas hipótesis necesarias para la estimación no paramétrica adaptativa de la regresión. En la Sección 4.2, *Procedimiento estadístico*, se proporcionan estimadores no paramétricos $\hat{g}_{h^*}(\cdot)$ y $\hat{m}_h(\cdot)$ de g y m respectivamente, posteriormente se da un estimador $\hat{r}_{h/h^*}(\cdot)$ en términos de los estimadores de m y g . Se presenta con detalles el método GL para la estimación no paramétrica adaptativa de las funciones $m(\cdot)$ y $g(\cdot)$ en contexto de dependencia débil. En la Sección 4.3, *Resultados*, se determina el sesgo, la varianza, consistencia, velocidad de convergencia y adaptatividad de los estimadores $\hat{m}_h(\cdot)$ y $\hat{g}_{h^*}(\cdot)$ en contexto de dependencia, siguiendo el esquema de demostración planteado en la Sección 3.3. Se demuestra que el estimador de regresión es consistente, se determina su velocidad de convergencia y finalmente basados en desigualdades de oráculo, se demuestra que es adaptativo. En la Sección 4.4, *Estudio de Simulación*, se plantea el esquema de simulación para generar datos del modelo de regresión con función de regresión acotada, variable explicativa de densidad acotada, débil dependiente de coeficiente α -mixing con decaimiento exponencial. Se plantea como obtener el estimador no paramétrico adaptativo, con núcleo gaussiano, de la función de regresión usando los datos simulados en el esquema planteado previamente. Posteriormente se muestra cómo se calibra el método usando los mismos datos simulados, y finalmente se presenta un estudio comparativo de resultados de estimaciones realizadas, usando errores globales y locales empíricos para distintos tamaños de muestras simuladas y distintos valores de la desviación estándar del error del modelo de regresión. La Sección 4.5, *Demostraciones*, reúne las demostraciones de los lemas, proposiciones, teoremas y corolarios de este capítulo. Se debe destacar que en este capítulo al no ser acotada la variable respuesta del modelo, se introduce la definición de truncamiento del estimador de la función $m(\cdot)$, lo que permite descomponer el estimador por núcleo de la función $m(\cdot)$ en dos terminos uno truncado y otro no acotado. El término estocástico no acotado del estimador se controla con la hipótesis de decaimiento exponencial del coeficiente de dependencia débil y el término estocástico truncado del estimador de la función $m(\cdot)$ es controlado con una adaptación que se realizó de la desigualdad de Bernstein para datos dependientes de Doukhan

and Neumann (2007).

En el capítulo 5, *Conclusión*; se destacan aportes relevantes del trabajo de grado y se proyectan futuras investigaciones.

Finalmente en anexo A, *Programa en R para la estimación adaptativa de la función de regresión correspondiente al capítulo 3* y anexo B, *Programa en R para la estimación adaptativa de la función de regresión correspondiente al capítulo 4*; se describen las hipótesis, consideraciones, metodología, entradas numéricas, salidas numéricas y salidas gráficas de los programas desarrollados en R que permitieron la implementación práctica, vía simulación, de la estimación adaptativa de la función de regresión tanto en el caso g conocido como en el caso g desconocido. Además, se muestran los códigos de ambos programas.

Agradecimientos

Me gustaría expresar mi más sincero agradecimiento a mi tutora, la Dra. Karine Bertin. Esta investigación no habría sido posible sin su talentosa guía, continua motivación y paciencia. Agradezco a mi co-tutor, el Dr. Lisandro Fermín, por la orientación y apoyo. Agradezco al jurado, por la disposición para revisar y evaluar este trabajo. Agradezco a CONICYT – PFHA / Doctorado Nacional 2019 – 21191358, por financiar el doctorado. Agradezco al proyecto FONDECYT regular 1221373 y finalmente agradezco al CIMFAV, por facilitar los equipos computacionales necesarios para el desarrollo de esta investigación.

Resumen

En el presente trabajo se plantea el problema de estimación no paramétrica (por núcleo) de la función de regresión en un modelo de regresión univariante. La precisión de la estimación se mide utilizando riesgos puntuales. Específicamente se plantea la estimación adaptativa de la función de regresión en contexto de dependencia, considerando que la variable explicativa es un proceso débilmente dependiente cuyo coeficiente de correlación tiene decaimiento exponencial. Se asume que la variable explicativa es idénticamente distribuida con función de densidad acotada, esta función de densidad en un caso se considera conocida (caso 1) y en otro caso desconocida (caso 2). Para estimar la función de regresión, se propone estimación por Núcleo y de selección de ventana por enfoque de Goldenshluger-Lepski (G-L). En ambos casos se demuestran nuevos resultados obteniendo que los estimadores seleccionados satisfacen desigualdades de oráculo y que son adaptativos respecto a la regularidad de la función de regresión. Además, se hace una calibración de los métodos de selección de estimadores en base a desigualdades tipo Bernstein adaptados a datos débilmente dependientes. Finalmente, se implementan los distintos métodos propuestos en el software R y se desarrollan simulaciones para ilustrar el desempeño de los métodos propuestos en riesgo puntual y riesgo integrado.

Abstract

In the present work, the problem of non-parametric estimation of the regression function in a univariate regression model is addressed. The accuracy of the estimation is measured using pointwise risks. Specifically, adaptive estimation of the regression function in a context of dependence is proposed, considering that the explanatory variable is a weakly dependent process with exponentially decaying correlation. It is assumed that the explanatory variable is identically distributed with a bounded density function; this density function is considered known in one case (case 1) and unknown in another case (case 2).

To estimate the regression function, Kernel estimation and bandwidth selection are proposed using the Goldenshluger-Lepski (G-L) approach. In both cases, new results are demonstrated, showing that the selected estimators satisfy oracle inequalities and are adaptive to the smoothness of the regression function. Moreover, calibration of the estimator selection methods is performed based on Bernstein-type inequalities adapted to weakly dependent data. Finally, the different proposed methods are implemented in the software R, and simulations are conducted to illustrate the performance of the proposed methods in terms of pointwise and integrated risk.

Índice general

| | |
|--|-------------|
| Índice de figuras | XIII |
| Índice de cuadros | XVII |
| 1. Formulación general del problema y preliminares | 1 |
| 1.1. Formulación general del problema | 1 |
| 1.2. Preliminares | 3 |
| 1.2.1. Estimadores de núcleo de la regresión | 3 |
| 1.2.2. Propiedades de los estimadores | 5 |
| 1.2.3. Dependencia débil | 8 |
| 1.2.4. Desigualdades para procesos independientes y débil dependientes . | 10 |
| 2. Estimación adaptativa en modelos de regresión para datos independientes y variable explicativa con densidad conocida | 13 |
| 2.1. Modelo | 14 |
| 2.2. Estimador de núcleo de la regresión y sus propiedades (caso no-adaptativo) | 15 |
| 2.3. Selección de ventanas por método GL, desigualdades de oráculo y adaptatividad | 18 |
| 2.3.1. Método GL | 18 |
| 2.3.2. Desigualdad de Oráculo y adaptatividad del estimador | 19 |
| 2.4. Demostraciones | 23 |
| 2.4.1. Constantes. | 23 |
| 2.4.2. Demostración de la Proposición 4: Sesgo del estimador. | 23 |
| 2.4.3. Demostración de la Proposición 5: Varianza del estimador. | 24 |
| 2.4.4. Demostración de la Proposición 7. | 25 |
| 2.4.5. Demostración de la Proposición 8, (i). | 27 |
| 2.4.6. Demostración de la Proposición 8, (ii) | 32 |

| | |
|--|-----------|
| 3. Estimación adaptativa en modelos de regresión para datos débilmente dependientes y variable explicativa con densidad conocida | 37 |
| 3.1. Modelo | 37 |
| 3.2. Procedimiento estadístico | 41 |
| 3.3. Resultados | 42 |
| 3.3.1. Resultados para el estimador \hat{r}_h | 42 |
| 3.3.2. Desigualdad de Oráculo y adaptatividad del estimador | 44 |
| 3.4. Estudio de simulación | 48 |
| 3.4.1. Esquema de simulación | 48 |
| 3.4.2. Estimación GL | 50 |
| 3.4.3. Calibración del método para una muestra. | 50 |
| 3.4.4. Comparación de errores globales y locales empíricos, para distintos tamaños de muestra y valores de σ | 52 |
| 3.5. Demostraciones | 58 |
| 3.5.1. Constantes. | 58 |
| 3.5.2. Enunciados de resultados técnicos. | 59 |
| 3.5.3. Demostraciones de resultados técnicos. | 61 |
| 3.5.4. Demostración de la Proposición 9: Sesgo del estimador \hat{g}_h | 72 |
| 3.5.5. Demostración de la Proposición 10: Varianza del estimador \hat{g}_h | 73 |
| 3.5.6. Demostración de la Proposición 11. | 77 |
| 3.5.7. Demostración de la Proposición 12, (i). | 78 |
| 3.5.8. Demostración de la Proposición 12, (ii). | 86 |
| | |
| 4. Estimación adaptativa en modelos de regresión para datos débilmente dependientes y variable explicativa con densidad desconocida | 91 |
| 4.1. Modelo | 92 |
| 4.2. Procedimiento estadístico | 96 |
| 4.3. Resultados | 99 |
| 4.3.1. Resultados de los estimadores \hat{m}_h y \hat{g}_h | 99 |
| 4.3.2. Desigualdades de Oráculo y adaptatividad del estimador $\hat{m}_{\hat{h}}$ | 101 |
| 4.3.3. Resultados del estimador \hat{r}_{h/h^*} | 102 |
| 4.4. Estudio de Simulación | 103 |
| 4.4.1. Esquema de simulación | 103 |
| 4.4.2. Estimación GL | 105 |
| 4.4.3. Calibración del método para una muestra. | 106 |
| 4.4.4. Comparación de errores globales y locales empíricos, para distintos tamaños de muestra y valores de σ | 113 |

| | |
|---|------------|
| 4.5. Demostraciones | 118 |
| 4.5.1. Constantes. | 118 |
| 4.5.2. Enunciados de resultados técnicos. | 118 |
| 4.5.3. Demostraciones de resultados técnicos. | 121 |
| 4.5.4. Demostración de la Proposición 16: Sesgo del estimador \hat{m}_h | 127 |
| 4.5.5. Demostración de la Proposición 17: Varianza del estimador \hat{m}_h | 128 |
| 4.5.6. Demostración del Teorema 8 | 132 |
| 4.5.7. Demostración de la Proposición 20. | 133 |
| 4.5.8. Demostración del Teorema 11. | 138 |
| 5. Conclusión | 139 |
| Bibliografía | 141 |
| Apéndice A. Programa en R para la estimación adaptativa de la función de regresión correspondiente al capítulo 3 | 143 |
| Apéndice B. Programa en R para la estimación adaptativa de la función de regresión correspondiente al capítulo 4 | 163 |

Índice de figuras

| | | |
|------|---|----|
| 3.1. | Gráfico de dispersión de \mathbb{X} y función de autocorrelación muestral $\gamma_{\mathbb{X}}$ | 49 |
| 3.2. | Se muestran tres estimadores GL de la función de regresión r , para muestras aleatorias $\{(Y_k, X_k)\}_{k=1}^{n+2q}$, con $\sigma = 0,5$ y $q = 100$. Cuando $n = 1000$ el estimador GL se representa con puntos, para $n = 2000$ con rayas largas y para $n = 5000$ con puntos y rayas cortas. Además se representa con una línea continua la función de regresión r | 51 |
| 3.3. | Arriba gráfico de $\{Error(\gamma_i)\}_{i=1}^{21}$ y abajo se muestra la nube de puntos $\{(X_i, Y_i)\}_{i=1}^n$, se representa con una línea continua la función de regresión r y con puntos el gráfico de $\{\hat{r}_{h_i}(x_i)\}_{i=1}^{21}$ en el intervalo $[-1, 1]$, calibrado en γ_3 , que es donde el gráfico de la izquierda alcanza su mínimo. En ambos se tomo $n = 2000$, $q = 100$ y $\sigma = 0,5$ | 53 |
| 3.4. | Arriba se representan tres boxplot de I_j para $j = 1, \dots, N$, con los tamaños de muestra $n = 1000, 2000$ y 5000 de izquierda a derecha respectivamente, sobre cada boxplot hay un punto que representa el $\hat{M}ISE$ y abajo se muestran tres representaciones del $\hat{M}SE(x_i)$ para $i = 1, \dots, s$ con una línea para el tamaño de muestra $n = 1000$, con puntos para $n = 2000$ y con rayas para $n = 5000$. En ambos casos para $\sigma = 0,1$ | 55 |
| 3.5. | Arriba se representan tres boxplot de I_j para $j = 1, \dots, N$, con los tamaños de muestra $n = 1000, 2000$ y 5000 de izquierda a derecha respectivamente, sobre cada boxplot hay un punto que representa el $\hat{M}ISE$ y abajo se muestran tres representaciones del $\hat{M}SE(x_i)$ para $i = 1, \dots, s$ con una línea para el tamaño de muestra $n = 1000$, con puntos para $n = 2000$ y con rayas para $n = 5000$. En ambos casos para $\sigma = 0,5$ | 56 |

3.6. Arriba se representan tres boxplot de I_j para $j = 1, \dots, N$, con los tamaños de muestra $n = 1000, 2000$ y 5000 de izquierda a derecha respectivamente, sobre cada boxplot hay un punto que representa el $M\hat{I}SE$ y abajo se muestran tres representaciones del $M\hat{SE}(x_i)$ para $i = 1, \dots, s$ con una línea para el tamaño de muestra $n = 1000$, con puntos para $n = 2000$ y con rayas para $n = 5000$. En ambos casos para $\sigma = 1$ 57

4.1. Gráfico de dispersión de \mathbb{X} y función de autocorrelación muestral γ_X 104

4.2. Se muestran tres estimadores GL de la función de regresión r , para muestras aleatorias $\{(Y_k, X_k)\}_{k=1}^{n+2q}$, con $\sigma = 0,5$ y $q = 100$. Cuando $n = 1000$ el estimador GL se representa con puntos, para $n = 2000$ con rayas largas y para $n = 5000$ con puntos y rayas cortas. Además se representa con una línea continua la función de regresión r 107

4.3. Histograma realizado con la muestra $\{X_i\}_{i=1}^n$ con $n = 2000$ y función de densidad g evaluada en la malla aleatoria $\{\tilde{x}_i\}_{i=1}^q$, con $q = 100$ 109

4.4. Arriba gráfico de $\{Error_g(\gamma_i^*)\}_{i=1}^{21}$ y abajo el gráfico de g es representado con una línea continua y representado con puntos el gráfico de $\{\hat{g}_{\hat{h}_i^*}(x_i)\}_{i=1}^{41}$ en el intervalo $[-2, 2]$, calibrado en γ_9^* , que es donde el gráfico de la izquierda alcanza su mínimo. 110

4.5. Arriba gráfico de $\{Error_m(\gamma_i)\}_{i=1}^{21}$ y abajo se muestra la nube de puntos $\left\{ \left(\tilde{x}_i, \tilde{y}_i \hat{g}_{\hat{h}_i^* \gamma_9^*}(\tilde{x}_i) \right) \right\}_{i=1}^q$, se representa con una línea continua el gráfico de m en el intervalo $[-2, 2]$ y con puntos el gráfico de $\{\hat{m}_{\hat{h}_i}(x_i)\}_{i=1}^{41}$, calibrado en γ_2 , que es donde el gráfico de la izquierda alcanza su mínimo. 111

4.6. La figura de arriba muestra la nube de puntos $\{(X_i, Y_i)\}_{i=1}^n$, se representa con una línea continua la función de regresión r y con puntos el gráfico de $\{\hat{r}_{\hat{h}_i/\hat{h}_i^*}(x_i)\}_{i=1}^{41}$ en el intervalo $[-2, 2]$, donde $x_i = -2 + 0,1(i - 1)$ y la figura de abajo muestra las mismas representaciones pero en la malla $x_i = -1 + 0,1(i - 1)$, para $i = 1, \dots, 21$, del intervalo $[-1, 1]$. Ambas estimaciones se calibraron en $\gamma_g = \gamma_9^*$ y $\gamma_m = \gamma_2$. Además se tomo $n = 2000, q = 100$ y $\sigma = 0,5$ 112

4.7. Arriba se representan tres boxplot de I_j para $j = 1, \dots, N$, con los tamaños de muestra $n = 1000, 2000$ y 5000 de izquierda a derecha respectivamente, sobre cada boxplot hay un punto que representa el $M\hat{I}SE$ y abajo se muestran tres representaciones del $M\hat{SE}(x_i)$ para $i = 1, \dots, s$ con una línea para el tamaño de muestra $n = 1000$, con puntos para $n = 2000$ y con rayas para $n = 5000$. En ambos casos para $\sigma = 0,1$ 114

-
- 4.8. Arriba se representan tres boxplot de I_j para $j = 1, \dots, N$, con los tamaños de muestra $n = 1000, 2000$ y 5000 de izquierda a derecha respectivamente, sobre cada boxplot hay un punto que representa el $M\hat{I}SE$ y abajo se muestran tres representaciones del $M\hat{S}E(x_i)$ para $i = 1, \dots, s$ con una línea para el tamaño de muestra $n = 1000$, con puntos para $n = 2000$ y con rayas para $n = 5000$. En ambos casos para $\sigma = 0,5$ 115
- 4.9. Arriba se representan tres boxplot de I_j para $j = 1, \dots, N$, con los tamaños de muestra $n = 1000, 2000$ y 5000 de izquierda a derecha respectivamente, sobre cada boxplot hay un punto que representa el $M\hat{I}SE$ y abajo se muestran tres representaciones del $M\hat{S}E(x_i)$ para $i = 1, \dots, s$ con una línea para el tamaño de muestra $n = 1000$, con puntos para $n = 2000$ y con rayas para $n = 5000$. En ambos casos para $\sigma = 1$ 116

Índice de cuadros

| | |
|---|-----|
| 3.1. Valores del $M\hat{I}SE$ para distintos valores de σ y tamaños de la muestra, caso g conocido. | 58 |
| 4.1. Valores del $M\hat{I}SE$ para distintos valores de σ y tamaños de la muestra, caso g desconocido. | 117 |

Capítulo 1

Formulación general del problema y preliminares

1.1. Formulación general del problema

En esta tesis, estudiamos la estimación de la función de regresión en modelo de regresión univariante. Se observa la muestra $\{(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)\}$ idénticamente distribuida donde llamamos a $X_i \in \mathbb{R}$ la variable explicativa y a $Y_i \in \mathbb{R}$ la variable respuesta. Se considera para $x \in \mathbb{R}$ la función de regresión $r(x) = \mathbb{E}[Y|X = x]$. Se considera el siguiente modelo de regresión

$$Y_i = r(X_i) + \varepsilon_i, \quad i = 1, \dots, n \quad (1.1)$$

donde $n \geq 2$, los X_i son idénticamente distribuidos con función de densidad g , los ε_i son independientes e idénticamente distribuidos, centrados y de varianza acotada y los X_i son independientes de los ε_i .

El objetivo es obtener la estimación no paramétrica adaptativa de la función de regresión r que es desconocida.

Se mide la calidad del estimador \hat{r} usando el error cuadrático medio (MSE) en el punto $x \in \mathbb{R}$

$$MSE(\hat{r}, r, x) = \mathbb{E} \left[(\hat{r}(x) - r(x))^2 \right]. \quad (1.2)$$

Observación 1. El error cuadrático medio integrado (o error L_2) considerado en la literatura es el $MISE(\hat{r}, r) = E \left[\int (\hat{r}(x) - r(x))^2 dx \right]$.

En este contexto consideramos un caso introductorio que corresponde a un enfoque estudiado en Lepski and Serdyukova (2014).

0. $\{(X_1, Y_1), \dots, (X_n, Y_n)\}$ son independientes y g es conocido.

Los casos nuevos estudiados en esta tesis en contexto de dependencia son

1. $\{(X_1, Y_1), \dots, (X_n, Y_n)\}$ son dependientes y g es conocido.

2. $\{(X_1, Y_1), \dots, (X_n, Y_n)\}$ son dependientes y g es desconocido.

Se utilizan adaptaciones del estimador de Nadaraya–Watson (ver Nadaraya (1989)), de la función de regresión r . Cuando la densidad g es conocida (caso 0 y 1) se utiliza una adaptación propuesta en el trabajo Lepski and Serdyukova (2014) y cuando la densidad g es desconocida (caso 2) se utiliza una adaptación similar a la propuesta en el trabajo Fermin et al. (2017).

En cualquiera de los casos planteados, dado una familia de estimadores de tipo núcleo $\{\hat{r}_h, h > 0\}$, donde h es la ventana o parámetro de suavizamiento, el interés es determinar la ventana óptima h_{opt} del estimador de regresión que minimice la función de riesgo $MSE(\hat{r}_h, r, x)$ definida en (1.2), o error cometido al estimar la función de regresión r , por $\hat{r}_{h_{opt}}$. En un primer enfoque conociendo la regularidad β de la función de regresión, la ventana óptima se obtiene acotando la función de riesgo y depende de la regularidad β . El estimador resultante converge a una velocidad óptima que también depende de β . Para que la estimación no paramétrica sea adaptativa se debe cumplir que la ventana óptima seleccionada no dependa de la regularidad de las funciones involucradas en el problema y que el estimador resultante converja a la velocidad óptima. Un procedimiento que permite obtener estimadores adaptativos es el método Goldenshluger-Lepski (GL). Este será descrito para introducir esta metodología de estimación de la función de regresión r en el capítulo 2, tomando como referencia el trabajo de Lepski and Serdyukova (2014) considerando independencia de la variable explicativa con función de densidad g conocida (caso 0) y suponiendo la hipótesis de que los errores ε_i del modelo (1.1) son acotados.

En este trabajo de tesis se generaliza este procedimiento para los casos 1 y 2. En el capítulo 3 se desarrolla con detalles el caso 1 extendiendo los resultados del caso 0 al introducir la hipótesis de que la variable explicativa X_i no es independiente. Para este fin se trabaja bajo el contexto de dependencia débil, además se cambia la hipótesis de que los errores ε_i del modelo (1.1) son acotados por la hipótesis de que los errores tienen distribución normal. En el capítulo 4 se desarrolla con detalles el caso 2, extendiendo los resultados del caso 1, al cambiar la hipótesis de que la densidad g de la variable explicativa X_i es conocida, es decir en este caso se supone que no se conoce la densidad g .

1.2. Preliminares

1.2.1. Estimadores de núcleo de la regresión

En esta sección, se presentan distintos estimadores de la función de regresión. Empezamos por presentar el estimador de Nadaraya-Watson que es uno de los estimadores de núcleo más clásicos para la función de regresión en la literatura. Después introducimos los dos estimadores de la regresión que usaremos en los 3 casos que pueden ser deducidos o se asemejan al estimador de Nadaraya-Watson.

- El estimador de Nadaraya-Watson, de la función de regresión r , depende de las estimaciones no paramétricas de:
 - La densidad g de X_i , para $i = 1, \dots, n$.
 - La densidad conjunta $f(x, y)$ de (X_i, Y_i) , para $i = 1, \dots, n$.

Una estimación para la densidad g esta dada (ver Silverman (1986)) por

$$\hat{g}_h(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right)$$

donde $h > 0$ es el ancho de ventana y $K : \mathbb{R} \rightarrow \mathbb{R}$, es una función de núcleo satisfaciendo $\int K(u)du = 1$. Un ejemplo clásico es el núcleo gaussiano $K(u) = \frac{1}{\sqrt{2\pi}}e^{-\frac{u^2}{2}}$.

Se introduce la notación $K_h(\cdot) = \frac{1}{h}K(\frac{\cdot}{h})$, lo que permite reescribir el estimador de la densidad g de la siguiente manera.

$$\hat{g}_h(x) = \frac{1}{n} \sum_{i=1}^n K_h(x - X_i). \quad (1.3)$$

El estimador de kernel para la densidad conjunta $f(x, y)$ es una generalización del caso anterior para \mathbb{R}^2 , el cual esta dado por

$$\hat{f}_h(x, y) = \frac{1}{nh^2} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) K\left(\frac{y - Y_i}{h}\right) = \frac{1}{n} \sum_{i=1}^n K_h(x - X_i) K_h(y - Y_i) \quad (1.4)$$

La función de regresión satisface

$$r(x) = \mathbb{E}[Y|X = x] = \int_{\mathbb{R}} yf(y|x)dy = \int_{\mathbb{R}} y \frac{f(x, y)}{g(x)} dy \quad (1.5)$$

En el caso en que la densidad $g(x)$ es desconocida, se reemplaza en la ecuación anterior, los estimadores $\hat{g}_h(x)$ y $\hat{f}_h(x, y)$ planteados previamente, y asumiendo que $\int wK(w)dw = 0$, se obtiene el estimador de Nadaraya–Watson de la función de regresión, dado por:

$$\hat{r}_h(x) = \begin{cases} \frac{\sum_{i=1}^n K_h(x-X_i)Y_i}{\sum_{j=1}^n K_h(x-X_j)} & \text{si } \sum_{j=1}^n K_h(x-X_j) \neq 0 \\ 0 & \text{si } \sum_{j=1}^n K_h(x-X_j) = 0 \end{cases} \quad (1.6)$$

En esta tesis, no se uso el estimador de Nadaraya-Watson como estimador de regresión dado que no permite usar el procedimiento de Goldenshluger-Lepski que es el que queremos usar para hacer después la selección de ventanas y obtener estimadores adaptativos. Para estimadores de núcleo, el método GL funciona en particular para estimadores que tienen forma de suma y no cociente. Por esta razón, definimos a continuación dos otros estimadores de la regresión que se aparentan al estimador de Nararaya-Watson pero que involucran estimadores suma.

- Para definir este segundo estimador, vamos a usar que la función de regresión dada en la ecuación (1.5) se reescribe de la siguiente forma,

$$r(x) = \frac{m(x)}{g(x)} \quad (1.7)$$

donde $m(x) = \int_{\mathbb{R}} yf(x, y)dy$.

En la ecuación (1.6) se considera que el numerador y denominador de la fracción dependen de ventanas distintas, obteniendo de esta manera el estimador de la función de regresión en las ventanas h y h' , para el caso g desconocido, dado por

$$\hat{r}_{h/h'}(x) = \frac{\hat{m}_h(x)}{\hat{g}_{h'}(x)} \quad (1.8)$$

donde $\hat{m}_h(x) = \frac{1}{n} \sum_{i=1}^n K_h(x-X_i)Y_i$ y $\hat{g}_{h'}(x) = \frac{1}{n} \sum_{j=1}^n K_{h'}(x-X_j)$, son estimadores de $m(\cdot)$ y $g(\cdot)$ respectivamente. Un estimador de este tipo es el que se utiliza cuando g es desconocido (caso 2), en el capítulo 4. En este mismo capítulo, el método GL se aplicará por separado al estimador \hat{m}_h y $\hat{g}_{h'}$.

- Cuando la función de densidad g es conocida (casos 0 y 1), se utiliza una adaptación del estimador de Nadaraya–Watson (1.6), propuesto en el trabajo Lepski and Serdyukova (2014), dado por

$$\hat{r}_h(x) = \frac{1}{n} \sum_{i=1}^n K_h(x - X_i) g^{-1}(X_i) Y_i \quad (1.9)$$

este estimador es utilizado en los capítulos 2 y 3.

1.2.2. Propiedades de los estimadores

En esta sección se darán definiciones y propiedades estadísticas básicas que debe tener un buen estimador, tales como sesgo, varianza, consistencia y velocidad de convergencia. Se van a explicitar estas definiciones para el enfoque de estimación no-paramétrica en riesgo puntual (en $x \in \mathbb{R}$) donde se quiere estimar la función de regresión r del modelo (1.1) basado en una muestra de tamaño n . Estas propiedades estadísticas se demostrarán para los estimadores de la regresión considerados de acuerdo al caso de estudio en los capítulos 2, 3 y 4.

Observación 2. *Para simplificar, escribimos estas definiciones para la función r , pero también son válidas para la estimación de las funciones m y g .*

Una cualidad que se espera de un estimador es que tome valores cercanos al parámetro $r(\cdot)$ a estimar, al menos para muestras grandes.

Definición 1 (Consistencia). *El estimador $\hat{r}_{h_n}(x)$ se dice consistente en $x \in \mathbb{R}$ para la estimación de la función $r(x)$, si $\hat{r}_{h_n}(x) \xrightarrow{\mathbb{P}} r(x)$ cuando $n \rightarrow \infty$, es decir si $\forall \varepsilon > 0$*
 $\lim_{n \rightarrow \infty} \mathbb{P}[|\hat{r}_{h_n}(x) - r(x)| > \varepsilon] = 0$.

Lo que significa que un estimador consistente se aleja del parámetro con una probabilidad débil, si el tamaño n de la muestra es lo suficientemente grande.

Aunque un estimador sea consistente, puede suceder que los valores que toma estén desplazados en promedio, con respecto al verdadero valor del parámetro. Se dice en este caso que el estimador tiene sesgo.

Definición 2 (Sesgo). *Se llama sesgo del estimador $\hat{r}_{h_n}(x)$ en el punto $x \in \mathbb{R}$ con respecto de $r(x)$ a la cantidad,*

$$\text{sesgo}(\hat{r}_{h_n}, x) = \mathbb{E}[\hat{r}_{h_n}(x)] - r(x).$$

El estimador se dice sin sesgo o insesgado en x si $\text{sesgo}(\hat{r}_{h_n}, x) = 0$. El estimador se dice asintóticamente insesgado en x , si $\lim_{n \rightarrow \infty} \text{sesgo}(\hat{r}_{h_n}, x) = 0$.

Además, se define la varianza del estimador.

Definición 3 (Varianza). Se llama *varianza del estimador* $\hat{r}_{h_n}(x)$ en el punto $x \in \mathbb{R}$ con respecto de $r(x)$ a la cantidad,

$$\text{var}(\hat{r}_{h_n}(x)) = \mathbb{E} \left[(\hat{r}_{h_n}(x) - \mathbb{E}[\hat{r}_{h_n}(x)])^2 \right].$$

El error cuadrático medio de $\hat{r}_{h_n}(\cdot)$ con respecto a $r(\cdot)$ en el punto x , $MSE(\hat{r}_{h_n}, r, x) = \mathbb{E} \left[(\hat{r}_{h_n}(x) - r(x))^2 \right]$ da garantía de que los valores que toma el estimador están efectivamente en un radio fijo alrededor del verdadero valor del parámetro, para un tamaño de muestra dado, por eso la calidad del estimador se cuantifica con la noción de error cuadrático. Además, el error cuadrático está ligado a la consistencia por la siguiente proposición.

Proposición 1. Si $MSE(\hat{r}_{h_n}, r, x) \rightarrow 0$, cuando $n \rightarrow \infty$, entonces \hat{r}_{h_n} es un estimador consistente de r .

Demostración. Si $|\hat{r}_{h_n}(x) - r(x)| > \varepsilon$, entonces $(\hat{r}_{h_n}(x) - r(x))^2 > \varepsilon^2$. Por lo tanto,

$$\mathbb{E} \left[(\hat{r}_{h_n}(x) - r(x))^2 \right] > \varepsilon^2 \mathbb{P} [|\hat{r}_{h_n}(x) - r(x)| > \varepsilon]$$

Si $MSE(\hat{r}_{h_n}, r, x) \rightarrow 0$ cuando $n \rightarrow \infty$, entonces también $\mathbb{P} [|\hat{r}_{h_n}(x) - r(x)| > \varepsilon] \rightarrow 0$, cuando $n \rightarrow \infty$. \square

Una relación importante entre el error cuadrático medio del estimador, el sesgo y su varianza es la siguiente,

$$MSE(\hat{r}_{h_n}, r, x) = (\text{sesgo}(\hat{r}_{h_n}, x))^2 + \text{var}(\hat{r}_{h_n}(x))$$

Esta relación permite establecer que, si un estimador es insesgado o asintóticamente insesgado y si su varianza tiende a cero, entonces el estimador es consistente. Siempre en estadística no paramétrica el estudio de la calidad de la estimación pasa por el control del sesgo y la varianza del estimador.

Para obtener propiedades mas finas que la consistencia del estimador, se busca la velocidad de convergencia del estimador. Para esto, se dice que se estudia la calidad de la estimación desde un punto de vista minimax (ver Tsybakov (2009)) y se asume que la función a estimar pertenece a una clase de funciones Σ . Se considera el riesgo máximo sobre esta clase de funciones. En el caso del riesgo puntual, el riesgo máximo del estimador $\hat{r}_{h_n}(\cdot)$ en el punto x esta dado por

$$\sup_{r \in \Sigma} \mathbb{E} \left[(\hat{r}_{h_n}(x) - r(x))^2 \right] = \sup_{r \in \Sigma} MSE(\hat{r}_{h_n}, r, x)$$

El estimador $\hat{r}_{h_n}(\cdot)$ es minimax sobre una clase de funciones Σ si para una constante C se tiene

$$\sup_{r \in \Sigma} \mathbb{E} \left[(\hat{r}_{h_n}(x) - r(x))^2 \right] \leq C \phi_n^2(\Sigma),$$

donde $\phi_n(\Sigma)$ es llamada la velocidad de convergencia minimax sobre Σ si

$$C_1 \phi_n^2(\Sigma) \leq \inf_{\hat{r} \in \Theta} \sup_{r \in \Sigma} \mathbb{E} \left[(\hat{r}(x) - r(x))^2 \right] \leq C_2 \phi_n^2(\Sigma) \quad (1.10)$$

donde $C_1, C_2 > 0$ y Θ es el conjunto de todos los estimadores de r . En (1.10), la primera desigualdad se llama cota inferior y la segunda cota se llama cota superior. Las cotas inferiores requieren el uso de técnicas específicas (ver Tsybakov (2009)).

Generalmente, los estimadores minimax dependen de la clase de función Σ y por ende de la regularidad de la función a estimar. Por esto, se quiere encontrar estimadores que se adaptan a la regularidad o a la clase de funciones. Dado una familia de clases funcionales $(\Sigma_\beta)_{\beta \in B}$, diremos que el estimador \hat{r}_{h_n} es adaptativo para la familia de clases de funciones $(\Sigma_\beta)_{\beta \in B}$ si \hat{r}_{h_n} converge a la velocidad $\phi_n(\Sigma_\beta)$ sobre cada clase de funciones Σ_β :

$$\sup_{r \in \Sigma_\beta} \mathbb{E} \left[(\hat{r}_{h_n}(x) - r(x))^2 \right] \leq C \phi_n^2(\Sigma_\beta),$$

donde $C > 0$.

En esta tesis consideramos como clases de funciones las clases de Hölder.

Definición 4 (Clase de Hölder). Sea $\beta > 0$, $L > 0$. La clase de Hölder $\Sigma(\beta, L)$ se define como el conjunto de todas las funciones $f : \mathbb{R} \rightarrow \mathbb{R}$ tales que la derivada $f^{(l)}$, $l = \lfloor \beta \rfloor$ existe y

$$\left| f^{(l)}(x) - f^{(l)}(y) \right| \leq L |x - y|^{\beta - l}, \forall x, y \in \mathbb{R}$$

donde $\lfloor \beta \rfloor = \max \{n \in \mathbb{N}, n < \beta\}$.

En el caso no-adaptativo para datos independientes, la velocidad de convergencia sobre clases de Hölder de regularidad β para la estimación de la función de regresión en riesgo puntual o integrado es de la forma $n^{-\beta/(2\beta+1)}$, ver Ibragimov and Khasminskii (1979), Stone (1977). El caso de datos dependientes ha sido estudiado por ejemplo en Bosq (2012), Bühlmann et al. (2002).

En el caso adaptativo para datos independientes o dependientes, Baraud (2000), Baraud et al. (2001), Baraud (2002), y Massart (2007) obtuvieron una velocidad de convergencia de la forma $\left(\frac{\log n}{n}\right)^{\beta/(2\beta+1)}$ para el riesgo integrado. Para el riesgo puntual en caso adaptativo, los

resultados de velocidades de convergencia han sido obtenidos solo para datos independientes y con velocidad $\left(\frac{\log n}{n}\right)^{\beta/(2\beta+1)}$ en Lepski and Serdyukova (2014) y Nguyen (2014). En esta tesis, nos interesa en particular ver el caso de riesgo puntual en caso adaptativo para datos dependientes.

1.2.3. Dependencia débil

La dependencia débil introducida por Doukhan and Louhichi (1999), consiste en establecer un control sobre el decaimiento de las colas de la función de covarianza. Antes de dar la definición formal, se recuerda aquí algunos hechos básicos sobre la independencia de las variables aleatorias.

Sean X, Y variables aleatorias definidas en el mismo espacio de probabilidad $(\Omega, \mathcal{A}, \mathbb{P})$ tomando valores en los espacios medibles (E_X, \mathcal{E}_X) y (E_Y, \mathcal{E}_Y) . La independencia de ambas variables aleatorias X, Y se escribe de la siguiente manera

$$\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B), \quad \forall (A, B) \in \mathcal{E}_X \times \mathcal{E}_Y.$$

Extender esta identidad por linealidad a funciones constantes escalonadas y usar límites produce una formulación de esta relación que parece más adaptada para aplicaciones:

$$\text{Cov}(f(X), g(Y)) = 0, \quad \forall (f, g) \in \mathbb{L}(E_X, \mathcal{E}_X) \times \mathbb{L}(E_Y, \mathcal{E}_Y),$$

donde, por ejemplo, $\mathbb{L}(E_X, \mathcal{E}_X)$ denota el espacio de funciones medibles y acotadas $f : (E_X, \mathcal{E}_X) \rightarrow (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$.

Al considerar una serie de tiempo $\mathbb{X} = (X_i)_{i \in \mathbb{Z}}$ con valores en un espacio E se puede considerar un vector $X_{i_1:u}$ del pasado y un vector $X_{j_1:v}$ del futuro:

$$X_{i_1:u} = (X_{i_1}, \dots, X_{i_u}), \quad X_{j_1:v} = (X_{j_1}, \dots, X_{j_v}),$$

donde $i_1 < \dots < i_u \leq i_u + r < j_1 < \dots < j_v$, y $u, v, r \in \mathbb{N}$. La independencia de la serie de tiempo \mathbb{X} se caracteriza como la independencia de $X_{i_1:u}$ y $X_{j_1:v}$. La independencia de la serie de tiempo hasta el tiempo m , también llamada m -dependencia, ahora se caracteriza como la independencia de $X_{i_1:u}$ y $X_{j_1:v}$ si $i_u + m \leq j_1$, es decir si $r \geq m$.

En esta sección se presenta la definición de dependencia débil dada por Doukhan and Louhichi (1999). Esta definición hace explícita la independencia asintótica entre "pasado" y

“futuro”; esto significa que el “pasado” se olvida progresivamente. En términos generales, se asume que

$$\text{Cov}(f(\text{'pasado'}), g(\text{'futuro'})).$$

es pequeño cuando la distancia r entre el “pasado” y el “futuro” es suficientemente grande.

Sea $\Delta^{(k)}$ el conjunto de funciones de Lipschitz acotadas f definidas en \mathbb{R}^k tales que para todos (x_1, \dots, x_k) y (x'_1, \dots, x'_k) en \mathbb{R}^k

$$|f(x_1, \dots, x_k) - f(x'_1, \dots, x'_k)| \leq \text{Lip}(f) \sum_{m=1}^k |x_m - x'_m|.$$

Arriba se denotó la constante de Lipschitz de f por $\text{Lip}(f)$. Sea $\Delta_1^{(k)}$ el conjunto de funciones f en $\Delta^{(k)}$ tales que $\|f\|_\infty \leq 1$.

Sea $\mathbb{X} = (X_i)_{i \in \mathbb{Z}}$ una serie a valores reales estacionaria.

Definición 5 (Doukhan and Louhichi). *El proceso $\mathbb{X} = (X_i)_{i \in \mathbb{Z}}$ es (ε, ψ) -débilmente dependiente si existe una secuencia infinita $\varepsilon(r)$ decreciendo a cero y una función ψ de $\mathbb{N}^2 \times (\mathbb{R}^+)^2$ a \mathbb{R}^+ tal que*

$$|\text{Cov}(f(X_{i_1}, \dots, X_{i_u}), g(X_{j_1}, \dots, X_{j_v}))| \leq \psi(u, v, \text{Lip}(f), \text{Lip}(g))\varepsilon(r),$$

Para todo $r \geq 0$ y todas las $(u+v)$ -tuplas tales que $i_1 < \dots < i_u \leq i_u + r < j_1 < \dots < j_v$, donde $(f, g) \in \Delta_1^{(u)} \times \Delta_1^{(v)}$.

Se introduce ε como el coeficiente de dependencia.

Observación 3. *Los ε -coeficientes dependen de la forma de la función ψ . Sea $\mathcal{J}(u, v, r)$ el conjunto de las $(u+v)$ -tuplas tales que $i_1 < \dots < i_u \leq i_u + r < j_1 < \dots < j_v$. Dada una función ψ de $\mathbb{N}^2 \times (\mathbb{R}^+)^2$ a \mathbb{R}^+ , el ε -coeficiente asociado a ψ es definido por*

$$\varepsilon(r) = \sup_{u, v} \sup_{\mathcal{J}(u, v, r)} \sup_{(f, g) \in \Delta_1^{(u)} \times \Delta_1^{(v)}} \frac{|\text{cov}(f(X_{i_1}, \dots, X_{i_u}), g(X_{j_1}, \dots, X_{j_v}))|}{\psi(u, v, \text{Lip}(f), \text{Lip}(g))}$$

Funciones específicas ψ dan la noción de dependencia débil apropiada para describir diferentes modelos. Diferentes tipos de ψ se utilizan en la bibliografía actual de dependencia débil, ver Dedecker et al. (2007), Doukhan and Louhichi (2001), Doukhan and Wintenberger (2008).

En esta de tesis, se trabaja con procesos débil dependientes en los capítulos 3 y 4, que es donde se abordan los casos en contexto de dependencia. Se sigue la noción de dependencia débil planteada en Bertin and Klutchnikoff (2017), ver sección 2, la cual escrita en términos de la notación que se sigue a lo largo de este trabajo, es la siguiente:

Se asume una estructura de dependencia débil sobre las variables X_i . Mas precisamente, para u y v enteros positivos, denotando $i_{1:u} = (i_1, \dots, i_u) \in \mathbb{Z}^u$ y $j_{1:v} = (j_1, \dots, j_v) \in \mathbb{Z}^v$, se definen los vectores aleatorios $X_{i_{1:u}} = (X_{i_1}, X_{i_2}, \dots, X_{i_u})$ y $X_{j_{1:v}} = (X_{j_1}, X_{j_2}, \dots, X_{j_v})$ a valores en \mathbb{R}^u y \mathbb{R}^v respectivamente. Se define la función $q: \mathbb{Z}^u \times \mathbb{Z}^v \rightarrow \mathbb{Z}$ por $q(i_{1:u}, j_{1:v}) = \min(j_{1:v}) - \max(i_{1:u})$ y Λ_u la clase de funciones $G_u: \mathbb{R}^u \rightarrow \mathbb{R}$ tales que $\|G_u\|_\infty = \sup_{x \in \mathbb{R}^u} |G_u(x)| < \infty$. Para un proceso aleatorio $\mathbb{X} = (X_i)_{i \in \mathbb{Z}}$, se define el coeficiente de correlación $\alpha(\mathbb{X}) = (\alpha_k(\mathbb{X}))_{k \in \mathbb{N}}$ por,

$$\alpha_k(\mathbb{X}) = \sup_{u, v \in \mathbb{N}} \sup_{(i_{1:u}, j_{1:v}) \in \mathbb{Z}^u \times \mathbb{Z}^v, q(i_{1:u}, j_{1:v}) \geq k} \sup_{G_u \in \Lambda_u} \sup_{G_v \in \Lambda_v} \frac{|\text{cov}(G_u(X_{i_{1:u}}), G_v(X_{j_{1:v}}))|}{\Psi(u, v, G_u, G_v)}$$

con $\Psi(u, v, G_u, G_v) = 4\|G_u\|_\infty\|G_v\|_\infty$.

Además, se impone la hipótesis de que el coeficiente α -mixing de las variables X_i cumple que, existe $a \in]0, 1[$ tal que,

$$\alpha_k(\mathbb{X}) \leq a^k, \quad \forall k \in \mathbb{N}.$$

1.2.4. Desigualdades para procesos independientes y débil dependientes

Las desigualdades tipo Bernstein son muy útiles para obtener cotas para el comportamiento del término estocástico de estimadores no paramétricos. Para el problema de regresión con datos independientes desarrollado en el capítulo 2, se usa una desigualdad de Bernstein enunciada y demostrada en Massart (2007), ver sección 2.2.3.

Proposición 2 (Desigualdad de Bernstein; caso independiente). *Sean Z_1, \dots, Z_n variables aleatorias independientes a valores reales. Si existen números positivos v y b tales que $\sum_{i=1}^n \mathbb{E}(Z_i^2) \leq v$ y $|Z_i| \leq b$, entonces para todo $t > 0$*

$$\mathbb{P}\left(|S_n| \geq \sqrt{2vt} + \frac{b}{3}t\right) \leq 2e^{-t}$$

donde $S_n = \sum_{i=1}^n (Z_i - \mathbb{E}(Z_i))$.

Doukhan and Neumann (2007) proponen una desigualdad de Bernstein que puede utilizarse para procesos débilmente dependientes. Esta desigualdad se ha utilizado en particular en el problema de la estimación adaptativa puntual de la densidad en Bertin and Klutchnikoff (2017). En los capítulos 3 y 4, se usa tal desigualdad en el problema de regresión con datos dependientes.

Proposición 3 (Desigualdad de Bernstein; caso débil dependiente). *Suponga que Z_1, \dots, Z_n son variables aleatorias a valores reales con media cero, definidas sobre un espacio de probabilidad $(\Omega, \mathcal{A}, \mathbb{P})$, que $S_n = Z_1 + \dots + Z_n$ y $\text{var}(S_n) = \sigma_n^2$. Sea $\psi : \mathbb{N}^2 \rightarrow \mathbb{N}$ una de las siguientes funciones:*

- (a) $\psi(u, v) = 2v$,
- (b) $\psi(u, v) = u + v$,
- (c) $\psi(u, v) = uv$,
- (d) $\psi(u, v) = \alpha(u + v) + (1 - \alpha)uv$, para cada $\alpha \in]0, 1[$.

Se asume que existen constantes positivas K, M, L_1, L_2, μ, ν y una secuencia no creciente de coeficientes reales $(\rho(n))_{n \geq 0}$ tal que, para toda u -tupla (i_1, \dots, i_u) y toda ν -tupla (j_1, \dots, j_ν) con $1 \leq i_1 \leq \dots \leq i_u \leq j_1 < \dots \leq j_\nu \leq n$. Si se cumplen las siguientes desigualdades

$$|\text{cov}(Z_{i_1} \dots Z_{i_u}, Z_{j_1} \dots Z_{j_\nu})| \leq K^2 M^{u+\nu-2} ((u+\nu)!)^\nu \psi(u, \nu) \rho(j_1 - i_u)$$

donde

$$\sum_{s=1}^{\infty} (s+1)^k \rho(s) \leq L_1 L_2^k (k!)^\mu, \quad \forall k \geq 0$$

y

$$\mathbb{E}|Z_i|^k \leq (k!)^\nu M^k, \quad \forall k \geq 0.$$

Entonces, $\forall t \geq 0$

$$\mathbb{P}(S_n \geq t) \leq \exp\left(-\frac{t^2/2}{A_n + B_n^{1/(\mu+\nu+2)} t^{(2\mu+2\nu+3)/(\mu+\nu+2)}}\right),$$

donde A_n es una cota superior de σ_n^2 y

$$B_n = 2L_2 \max(K, M) \max\left\{1, \frac{2^{4+\mu+\nu} n K^2 L_1}{A_n}\right\}.$$

Capítulo 2

Estimación adaptativa en modelos de regresión para datos independientes y variable explicativa con densidad conocida

Resumen

Este capítulo conforma una introducción a la metodología que será utilizada después en el caso de datos dependientes en los capítulos 3 y 4. Estamos interesados en estudiar la estimación de la función de regresión en el caso de datos independientes en riesgo puntual. Para tal fin, se sigue el procedimiento basado en datos usando estimación por Núcleo. El ancho de ventana se selecciona con el enfoque Goldenshluger-Lepski. Se demuestra que el estimador resultante satisface una desigualdad de tipo oráculo, lo que permite establecer que el estimador GL es adaptativo. Los resultados de este capítulo pueden ser derivados de Lepski and Serdyukova (2014) pero están demostrados bajo hipótesis mas simplificadas a modo de presentar los estimadores utilizados y los esquemas de demostraciones de manera mas simple. En este sentido se plantea un modelo de regresión univariante donde la función de regresión es acotada, los errores del modelo también son acotados además de i.i.d y la variable explicativa es independiente e idénticamente distribuida con función de densidad acotada y conocida.

2.1. Modelo

Observamos $\{(X_1, Y_1), \dots, (X_n, Y_n)\}$ idénticamente distribuidas que satisfacen

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

donde $n \geq 2$, los X_i son idénticamente distribuidos con función de densidad g conocida, los ε_i son acotados, independientes e idénticamente distribuidos, centrados y de varianza acotada, los X_i son independientes de los ε_i , y r es la función de regresión $r(x) = \mathbb{E}(Y|X = x)$.

Nuestro objetivo es estimar la función r en el punto $x \in \mathbb{R}$ usando la muestra observada $\{(X_1, Y_1), \dots, (X_n, Y_n)\}$. Para motivar el problema de tesis, se trabajara en este capítulo con un enfoque de independencia.

Se mide la calidad de la estimación de un estimador \hat{r} usando el error cuadrático en un punto $x \in \mathbb{R}$.

$$R(r, \hat{r}, x) = \mathbb{E} (r(x) - \hat{r}(x))^2.$$

Por un lado, se asume la siguiente hipótesis correspondiente a una cota sobre la densidad de las variables X_i en la vecindad de x :

$$B(x) = [x - 2, x + 2].$$

(H_1) La densidad g de las X_i satisface

$$g_{\inf} \leq g(u) \leq g_{\sup}, \quad \forall u \in B(x)$$

donde g_{\inf} y g_{\sup} son constantes positivas.

Se asume la siguiente hipótesis sobre la función de regresión

(H_2) Existe una constante $r_{\sup} > 0$ tal que

$$|r(u)| \leq r_{\sup}, \quad \forall u \in B(x).$$

Finalmente, se dan condiciones de acotamiento sobre el error de la regresión

(H_3) Existen constantes positivas b_0 y σ tales que que los ε_i satisfacen

$$|\varepsilon_i| \leq b_0 \quad \text{y} \quad \mathbb{E}(\varepsilon^2) = \sigma^2$$

A continuación, en la Sección 2.2, *Estimador de núcleo de la regresión y sus propiedades (caso no-adaptativo)*; se presenta el estimador por Núcleo de la función de regresión correspondiente al caso densidad desconocida, además se dan definiciones y propiedades estadísticas que debe tener para ser un buen estimador, tales como sesgo, varianza, consistencia y velocidad de convergencia. Todo esto en contexto de independencia. Se establece que el estimador converge a una velocidad óptima en una ventana que depende de la regularidad de la función de regresión, algo característico en el caso no-adaptativo. Las demostraciones de los resultados se encuentran en la Sección 2.4. La Sección 2.3 esta dividida en dos partes. Primero, *Método GL*; donde se define una familia de ventanas \mathcal{H} la cual evidentemente no depende de la regularidad de la función de regresión y se describe el procedimiento de selección de ventanas en \mathcal{H} basado en datos, conocido como método Goldenshluger-Lepski. Segundo, ; se establece que el estimador por Núcleo en la ventana seleccionada, satisface una desigualdad de tipo oráculo, lo cual posteriormente permite concluir que tal estimador converge a una velocidad casi óptima en la ventana seleccionada por el método GL. Tal ventana no depende de la regularidad de la función de regresión, lo que hace que el estimador sea adaptativo. Las demostraciones de esta sección se fundamentan en las hipótesis planteadas en las secciones anteriores, en especial la de independencia y la de errores acotados. Los detalles técnicos se pueden encontrar en la Sección 2.4.

2.2. Estimador de núcleo de la regresión y sus propiedades (caso no-adaptativo)

Consideramos el estimador de la función de regresión mencionado en capítulo 1 dado por

$$\hat{r}_h(x) = \frac{1}{n} \sum_{k=1}^n Y_k K_h(x - X_k) g^{-1}(X_k)$$

donde $h > 0$ es el ancho de ventana, $K_h(\cdot) = \frac{1}{h} K(\frac{\cdot}{h})$ y $K : \mathbb{R} \rightarrow \mathbb{R}$, es una función de núcleo satisfaciendo que $\int K(u) du = 1$ y la siguiente hipótesis.

(H₄) K tiene soporte $[-1, 1]$ y , $\|K\|_\infty < +\infty$.

Esta hipótesis implica que $\|K\|_1 < +\infty$ y $\|K\|_2 < +\infty$. Note que bajo (H₁) el estimador esta bien definido dado que $(g(x_k))^{-1} \neq 0$ para cada $x_k \in B(x_k)$.

Vamos a estudiar el sesgo, varianza y velocidad de convergencia de este estimador de regresión. Para poder controlar el sesgo y demostrar que es asintóticamente insesgado se dan las definiciones de clase de Hölder y de kernel de orden m .

Definición 6 (Clase de Hölder). Sea $\beta > 0$, $L > 0$. La clase de Hölder $\Sigma(\beta, L)$ se define como el conjunto de todas las funciones $f : \mathbb{R} \rightarrow \mathbb{R}$ tales que la derivada $f^{(l)}$, $l = \lfloor \beta \rfloor$ existe y

$$\left| f^{(l)}(x) - f^{(l)}(y) \right| \leq L|x - y|^{\beta - l}, \forall x, y \in \mathbb{R}$$

donde $\lfloor \beta \rfloor = \max \{n \in \mathbb{N}, n < \beta\}$.

Definición 7 (Kernel de orden m). Sea $m \in \mathbb{N}$. Se dice que $K : \mathbb{R} \rightarrow \mathbb{R}$ es un kernel de orden m si las funciones $u \rightarrow u^j K(u)$, $j = 0, 1, \dots, m$, satisfacen

$$\int_{\mathbb{R}} K(u) du = 1, \int_{\mathbb{R}} u^j K(u) du = 0, j = 1, \dots, m.$$

A continuación, en las dos siguientes proposiciones se acota el sesgo y la varianza del estimador \hat{r}_h .

Proposición 4. Sea $\beta > 0$ y $L > 0$. Asumimos que K es de orden m , donde $m \geq l = \lfloor \beta \rfloor$, y que satisface $\int_{\mathbb{R}} |u|^\beta |K(u)| du < \infty$. Entonces, bajo la hipótesis (H_1) , si $r \in \Sigma(\beta, L)$, se tiene

$$|E[\hat{r}_h(x)] - r(x)| = |K_h * r(x) - r(x)| \leq h^\beta A_0,$$

donde $A_0 = \frac{1}{\Gamma} \int_{\mathbb{R}} |u|^\beta |K(u)| du$. El estimador \hat{r}_h es asintóticamente insesgado, cuando h tiende a 0.

Proposición 5. Bajo las hipótesis (H_1) , (H_2) , (H_3) y (H_4) se tiene que para $h > 0$

$$\mathbb{E} \left[(\hat{r}_h(x) - \mathbb{E}[\hat{r}_h(x)])^2 \right] \leq \frac{A_1}{nh},$$

donde $A_1 = ((r_{\sup})^2 + \sigma^2)(g_{\inf})^{-1} \|K\|_2^2$.

Proposiciones 4 y 5 implican que cuando $n \rightarrow \infty$, $h \rightarrow 0$ y $nh \rightarrow \infty$, $MSE(\hat{r}_h, r, x) = \mathbb{E} \left[(\hat{r}_h(x) - r(x))^2 \right] \rightarrow 0$, o sea $\hat{r}_h(x)$ es un estimador consistente en media cuadrática (y por tanto en probabilidad) de $r(x)$.

La siguiente Proposición es fundamental para determinar la velocidad de convergencia del estimador de regresión $\hat{r}_h(x)$.

Proposición 6. Bajo las hipótesis de las Proposiciones 4 y 5, se tiene que, si $r \in \Sigma(\beta, L)$ entonces el estimador \hat{r}_{h_*} , con $h_* = n^{-\frac{1}{2\beta+1}}$ satisface

$$\mathbb{E} \left[(\hat{r}_{h_*}(x) - r(x))^2 \right] \leq C_* n^{-\frac{2\beta}{2\beta+1}},$$

donde C_* es una constante positiva que depende del núcleo K , de β , L , r_{\sup} , σ y g_{\inf} .

Demostración. Usando Proposiciones 4 y 5, se deduce que,

$$MSE(\hat{r}_h, r, x) \leq h^{2\beta} A_0^2 + \frac{A_1}{nh}. \quad (2.1)$$

Substituyendo h por h_* en la ecuación anterior se tiene que,

$$MSE(\hat{r}_{h_*}, r, x) \leq (A_0^2 + A_1) n^{-\frac{2\beta}{2\beta+1}}$$

□

Note que cualquier ventana de la forma $An^{-\frac{1}{2\beta+1}}$ permite obtener la misma tasa de convergencia. En particular también la ventana que minimiza en h al termino de la derecha en la desigualdad (2.1), ya que es de la forma $h = \left(\frac{A_1}{2\beta A_0^2}\right)^{\frac{1}{2\beta+1}} n^{-\frac{1}{2\beta+1}}$.

Por el resultado anterior, se tiene

$$\sup_{r \in \Sigma(\beta, L)} E \left[(\hat{r}_{h_*}(x) - r(x))^2 \right] \leq C_* n^{-\frac{2\beta}{2\beta+1}},$$

y

$$\inf_{\hat{r} \in \Theta} \sup_{r \in \Sigma(\beta, L)} E \left[(\hat{r}(x) - r(x))^2 \right] \leq C_* n^{-\frac{2\beta}{2\beta+1}},$$

donde Θ es el conjunto de todos los estimadores de r .

En Tsybakov (2009) (ver Sección 2.5), se demuestra

$$\inf_{\hat{r} \in \Theta} \sup_{r \in \Sigma(\beta, L)} E \left[(\hat{r}(x) - r(x))^2 \right] \geq cn^{-\frac{2\beta}{2\beta+1}},$$

donde c es una constante que depende de β , L , g_{sup} , y condiciones adicionales sobre la densidad de los ε_i .

Esto permite concluir que el estimador \hat{r}_{h_*} converge a la velocidad óptima sobre la clase de Hölder $\Sigma(\beta, L)$ la cual está dada por

$$v_n = n^{-\frac{\beta}{2\beta+1}}. \quad (2.2)$$

El estimador \hat{r}_{h_*} depende de la regularidad β de la función a estimar. En esta tesis nos interesa encontrar estimadores adaptativos, que no dependen de la regularidad y con velocidad de convergencia óptimas. Un procedimiento que permite obtener estimadores adaptativos es el método de Goldenshluger-Lepski. A continuación, lo vamos a presentar y dar las ideas de

por que permite obtener estimadores con buenas propiedades (desigualdades de oráculo y adaptatividad)

2.3. Selección de ventanas por método GL, desigualdades de oráculo y adaptatividad

2.3.1. Método GL

En la última década, el método de Goldenshluger-Lepski (GL) se ha utilizado en varios modelos estadísticos para obtener estimadores adaptativos basados en estimadores de núcleo. Se ha estudiado sobre todo en contextos independientes (ver por ejemplo Goldenshluger and Lepski (2011), Goldenshluger and Lepski (2013), Bertin et al. (2016), Lepski and Serdyukova (2014)). En casos dependientes, Bertin and Klutchnikoff (2017) y Comte et al. (2017) aplican los procedimientos GL en modelos de densidad.

Para seleccionar la ventana h del estimador en nuestro problema, vamos a usar el método de Goldenshluger-Lepski (GL).

Consideramos la familia de anchos de ventanas

$$\mathcal{H} = \{e^{-i}\}_{i=0}^M \cap [h_{\min}, h_{\max}]$$

con $h_{\min} = \frac{\log n}{n}$, $h_{\max} = 1$ y $M = \left\lceil \log \left(\frac{n}{\log n} \right) \right\rceil$. Se define para $h, h' \in \mathcal{H}$ un estimador auxiliar sobreesuavizado

$$\hat{r}_{h,h'}(x) = \frac{1}{n} \sum_{k=1}^n Y_k K_h * K_{h'}(X_k - x) g^{-1}(X_k).$$

Se define para $h \in \mathcal{H}$,

$$A(h, x) = \max_{h' \in \mathcal{H}} \{ |\hat{r}_{h,h'}(x) - \hat{r}_{h'}(x)| - V(h') \}_+$$

donde $V(h)$ esta dado por,

$$V(h) = \left(\sqrt{2\gamma A_1} + \frac{\gamma A_2}{3} \right) (1 + \|K\|_1) \frac{(\log n)^{\frac{1}{2}}}{(nh)^{\frac{1}{2}}}$$

donde $A_1 = ((r_{\sup})^2 + \sigma^2) (g_{\inf})^{-1} \|K\|_2^2$, $A_2 = (r_{\sup} + b_0) (g_{\inf})^{-1} \|K\|_{\infty}$ y $\gamma > 1$.

El procedimiento GL consiste en seleccionar, basado en los datos, un ancho de ventana \hat{h} de la familia \mathcal{H} , dado por

$$\hat{h} = \arg \min_{h \in \mathcal{H}} \{A(h, x) + V(h)\}, \quad (2.3)$$

La idea es que se elige el estimador que minimice la suma del sesgo (estimado) $A(h, x)$ y la desviación estándar $V(h)$.

En efecto, la cantidad $V(h)$ es una cota superior ajustada de la desviación estándar del estimador de regresión \hat{r}_h , dado que tenemos (ver Proposición 5)

$$(E|\hat{r}_h(x) - E(\hat{r}_h(x))|^2)^{1/2} \leq V(h).$$

La cantidad $A(h, x)$ puede verse como un estimador del sesgo del estimador de regresión $\hat{r}_h(\cdot)$, ya que, más adelante en la Proposición 7, se demostrará que $A(h, x) \leq T_1 + T_2 + C(h)\|K\|_1$, donde $C(h) = \max_{u \in B(x)} |K_h * r(u) - r(u)|$ esta en términos del sesgo de $\hat{r}_h(\cdot)$, con $B(x) = [x - 2, x + 2]$, y T_1 y T_2 son variables aleatorias, que posteriormente la Proposición 8, establece que $\mathbb{E}(T_1^2) \leq \frac{A_3}{n}$ y $\mathbb{E}(T_2^2) \leq \mathbb{E}(T_1^2)\|K\|_1$ de lo que se desprende que el momento de orden dos de $A(h, x)$ esta acotado superiormente por el sesgo de $\hat{r}_h(\cdot)$ mas dos cantidades insignificantes con respecto al sesgo.

2.3.2. Desigualdad de Oráculo y adaptatividad del estimador

En la siguiente proposición, obtenemos una primera desigualdad que satisface el estimador $\hat{r}_{\hat{h}}(x)$ usando principalmente la definición de \hat{h} dada en (2.3).

Proposición 7. *Bajo la hipótesis (H_4) , el estimador $\hat{r}_{\hat{h}}(x)$ satisface para todo $h \in \mathcal{H}$,*

$$\begin{aligned} & \left(\mathbb{E} (r(x) - \hat{r}_{\hat{h}}(x))^2 \right)^{\frac{1}{2}} \\ & \leq \left(\mathbb{E} (r(x) - \hat{r}_h(x))^2 \right)^{\frac{1}{2}} + 2 \left(\mathbb{E}(T_1^2) \right)^{\frac{1}{2}} + 2 \left(\mathbb{E}(T_2^2) \right)^{\frac{1}{2}} + 2C(h)\|K\|_1 + 2V(h) \end{aligned}$$

donde

$$T_1 = \max_{h' \in \mathcal{H}} \{ |\hat{r}_{h'}(x) - \mathbb{E}(\hat{r}_{h'}(x))| - V_1(h') \}_+, \quad T_2 = \max_{h' \in \mathcal{H}} \{ |\hat{r}_{h, h'}(x) - \mathbb{E}(\hat{r}_{h, h'}(x))| - V_2(h') \}_+,$$

$$C(h) = \max_{u \in B(x)} |K_h * r(u) - r(u)|$$

y $V_1(h')$, $V_2(h')$ tales que $V(h') = V_1(h') + V_2(h')$, con

$$V_1(h') = \left(\sqrt{2\gamma A_1} + \frac{\gamma A_2}{3} \right) \frac{(\log n)^{\frac{1}{2}}}{(nh')^{\frac{1}{2}}}, \quad V_2(h') = \|K\|_1 V_1(h').$$

En la Proposición anterior se proporciona una cota de la raíz cuadrada del error cuadrático medio puntual del estimador $\hat{r}_h(\cdot)$, en términos de la varianza y el sesgo del estimador $\hat{r}_h(\cdot)$, además de $V_1(h')$, $V_2(h')$, $\mathbb{E}(T_1^2)$ y $\mathbb{E}(T_2^2)$. En la demostración de la Proposición es fundamental el hecho de que $V(h') = V_1(h') + V_2(h')$. La elección precisa que se tomo de las expresiones de $V_1(h')$ y $V_2(h')$ es fundamental para la demostración de la Proposición 8 y permite establecer que $\mathbb{E}(T_1^2)$ y $\mathbb{E}(T_2^2)$ sean despreciables respecto a los términos de sesgo y varianza de cualquier estimador \hat{r}_h para $h \in \mathcal{H}$.

Para controlar los términos $\mathbb{E}(T_1^2)$ y $\mathbb{E}(T_2^2)$, se usa la desigualdad de Bernstein enunciada en la proposición 2, cuya demostración se puede encontrar en Massart (2007), (ver Sección 2.2.3).

Proposición 8. *Bajo las hipótesis (H_1) , (H_2) , (H_3) y (H_4) , se tiene que,*

$$(i) \quad \mathbb{E}(T_1^2) \leq \frac{A_3}{n},$$

$$(ii) \quad \mathbb{E}(T_2^2) \leq \frac{A_3 \|K\|_1^2}{n},$$

donde A_3 es una constante explícita (ver anexo) que depende del núcleo K , de r_{sup} , g_{inf} , σ , b_0 y $\gamma > 1$.

Teorema 1 (Desigualdad de Oráculo). *Bajo las hipótesis (H_1) , (H_2) , (H_3) y (H_4) , el estimador $\hat{r}_h(x)$ satisface la siguiente desigualdad,*

$$\left(\mathbb{E} (r(x) - \hat{r}_h(x))^2 \right)^{\frac{1}{2}} \leq \min_{h \in \mathcal{H}} \left(A_4 C(h) + A_5 \frac{(\log n)^{\frac{1}{2}}}{(nh)^{\frac{1}{2}}} \right) + \frac{A_6}{n^{\frac{1}{2}}} \quad (2.4)$$

donde $A_4 = 1 + 2\|K\|_1$, $A_5 = \sqrt{A_1} + 2 \left(\sqrt{2\gamma A_1} + \frac{\gamma A_2}{3} \right) (1 + \|K\|_1)$ y $A_6 = 2(1 + \|K\|_1)\sqrt{A_3}$.

Demostración. Sea $h \in \mathcal{H}$. Usando que $|\mathbb{E}(\hat{r}_h(x)) - r(x)| = |K_h * r(x) - r(x)| \leq C(h)$ y la Proposición 5 se tiene que,

$$\begin{aligned} \left(\mathbb{E} \left[(r(x) - \hat{r}_h(x))^2 \right] \right)^{\frac{1}{2}} &\leq C(h) + \left(\frac{A_1}{nh} \right)^{\frac{1}{2}} \\ &\leq C(h) + \frac{\sqrt{A_1}}{(nh)^{1/2}} \end{aligned} \quad (2.5)$$

Ahora por la Proposición 7 y la desigualdad (2.5), se tiene que,

$$\begin{aligned}
 & \left(\mathbb{E} (r(x) - \hat{r}_{\hat{h}}(x))^2 \right)^{\frac{1}{2}} \\
 & \leq C(h) + \frac{\sqrt{A_1}}{(nh)^{1/2}} + 2C(h)\|K\|_1 + 2V(h) + 2 \left(\mathbb{E}(T_1^2) \right)^{\frac{1}{2}} + 2 \left(\mathbb{E}(T_2^2) \right)^{\frac{1}{2}} \\
 & \leq (1 + 2\|K\|_1)C(h) + A_5 \frac{(\log n)^{\frac{1}{2}}}{(nh)^{\frac{1}{2}}} + 2 \left(\mathbb{E}(T_1^2) \right)^{\frac{1}{2}} + 2 \left(\mathbb{E}(T_2^2) \right)^{\frac{1}{2}}
 \end{aligned}$$

De la Proposición 8 se obtiene que

$$2 \left(\mathbb{E}(T_1^2) \right)^{\frac{1}{2}} + 2 \left(\mathbb{E}(T_2^2) \right)^{\frac{1}{2}} \leq \frac{2(1 + \|K\|_1)\sqrt{A_3}}{n^{\frac{1}{2}}},$$

lo que permite concluir el resultado. □

Observación 4. *En el Teorema 1, podemos ver en (2.4) que el estimador $\hat{r}_{\hat{h}}$ imita al oráculo, es decir, al mejor estimador posible (pero desconocido) \hat{r}_h en la familia $(\hat{r}_h)_{h \in \mathcal{H}}$, que es el que minimiza la suma del sesgo $C(h)$ y de la desviación estandar. Cuando la familia $(\hat{r}_h)_{h \in \mathcal{H}}$ es lo suficientemente amplia, los estimadores que satisfacen desigualdades de oráculo suelen ser estimadores adaptativos. Es lo que se ve a continuación en el siguiente teorema.*

Teorema 2 (Adaptatividad). *Asumimos que las hipótesis (H_1) , (H_2) , (H_3) y (H_4) están satisfechas. Asumimos que K es de orden m . Entonces para todo $0 < \beta \leq m$ y $L > 0$ el estimador $\hat{r}_{\hat{h}}$ satisface para $r \in \Sigma(\beta, L)$.*

$$\left(\mathbb{E} (r(x) - \hat{r}_{\hat{h}}(x))^2 \right)^{\frac{1}{2}} \leq C^* \left(\frac{n}{\log n} \right)^{-\frac{\beta}{2\beta+1}}$$

donde C^* es una constante positiva que depende del nucleo K , de β , L , r_{sup} , g_{inf} , σ , b_0 y γ .

Observación 5. *El estimador $\hat{r}_{\hat{h}}$ converge a la velocidad $\left(\frac{n}{\log n} \right)^{-\frac{\beta}{2\beta+1}}$ sobre las clases de Hölder de regularidad β y el estimador no depende de β . Esta velocidad es casi la velocidad óptima o minimax salvo por el término multiplicativo logaritmico. Este término logaritmico extra es clásico en el caso adaptativo y aparece también por ejemplo en Lepski and Serdyukova (2014) o Bertin and Klutchnikoff (2014).*

Demostración. Demostración del Teorema 2 Sea $h_i = e^{-i}$ para cada $i \in \{0, 1, 2, \dots, M\}$ y $h_\beta = \left(\frac{\log n}{n}\right)^{\frac{1}{2\beta+1}}$. Para n suficientemente grande, existe $i \in \{0, 1, 2, \dots, M\}$ tal que,

$$h_{i+1} \leq h_\beta \leq h_i \quad (2.6)$$

donde $h_{i+1} = e^{-i-1} = \frac{e^{-i}}{e} = \frac{h_i}{e}$.

Se procede a acotar el primer término dentro del mínimo en la desigualdad (2.4). Por la Proposición 4 se tiene que el sesgo del estimador en la ventana h_{i+1} satisface para $u \in B(x)$,

$$|K_{h_{i+1}} * r(u) - r(u)| \leq A_0 h_{i+1}^\beta$$

Por lo tanto usando la desigualdad (2.6), se tiene

$$C(h_{i+1}) = \max_{u \in B(x)} |K_{h_{i+1}} * r(u) - r(u)| \leq A_0 h_{i+1}^\beta \leq A_0 h_\beta^\beta = A_0 \left(\frac{n}{\log n}\right)^{-\frac{\beta}{2\beta+1}}. \quad (2.7)$$

Ahora se procede a determinar una cota para el segundo término dentro del mínimo en la desigualdad (2.4). De la desigualdad (2.6) se deduce que $\frac{\log n}{nh_i} \leq \frac{\log n}{nh_\beta}$ y además como $h_{i+1} = \frac{h_i}{e}$ se establece que,

$$\frac{\log n}{nh_{i+1}} = \frac{e \log n}{nh_i} \leq \frac{e \log n}{nh_\beta}$$

que implica usando el valor de h_β

$$\left(\frac{\log n}{nh_{i+1}}\right)^{\frac{1}{2}} \leq \frac{e^{\frac{1}{2}} (\log n)^{\frac{1}{2}}}{n^{\frac{1}{2}}} \frac{1}{h_\beta^{\frac{1}{2}}} = e^{\frac{1}{2}} \left(\frac{\log n}{n}\right)^{\frac{\beta}{2\beta+1}} \quad (2.8)$$

Por las desigualdades (2.7) y (2.8) se tiene que para $h_{i+1} \in \mathcal{H}$,

$$A_4 C(h_{i+1}) + A_5 \frac{(\log(n))^{\frac{1}{2}}}{(nh_{i+1})^{\frac{1}{2}}} \leq (A_4 A_0 + A_5 e^{\frac{1}{2}}) \left(\frac{n}{\log n}\right)^{-\frac{\beta}{2\beta+1}}$$

Por esto último y por la desigualdad (2.4) se tiene que,

$$\left(\mathbb{E} (r(x) - \hat{r}_h(x))^2\right)^{\frac{1}{2}} \leq (A_4 A_0 + A_5 e^{\frac{1}{2}}) \left(\frac{n}{\log n}\right)^{-\frac{\beta}{2\beta+1}} + \frac{A_6}{n^{\frac{1}{2}}} \quad (2.9)$$

Usando que $\frac{1}{n^{\frac{1}{2}}} \leq \left(\frac{\log n}{n}\right)^{\frac{1}{2}} \leq \left(\frac{\log n}{n}\right)^{\frac{\beta}{2\beta+1}}$ para $n \geq 3$, se deduce que

$$\left(\mathbb{E}(r(x) - \hat{r}_h(x))^2\right)^{\frac{1}{2}} \leq C^* \left(\frac{n}{\log n}\right)^{-\frac{\beta}{2\beta+1}}$$

donde C^* depende de A_0, A_4, A_5 y A_6 . □

2.4. Demostraciones

2.4.1. Constantes.

Recordamos algunos valores de constantes dados previamente y además a continuación van constantes que usaremos en el desarrollo de las demostraciones.

$$\begin{aligned} A_0 &= \frac{L}{l!} \int_{\mathbb{R}} |u|^\beta |K(u)| du, & A_1 &= ((r_{\text{sup}})^2 + \sigma^2)(g_{\text{inf}})^{-1} \|K\|_2^2, \\ A_2 &= (r_{\text{sup}} + b_0)(g_{\text{inf}})^{-1} \|K\|_\infty, & A_3 &= \frac{32 \max(A_1, A_2^2)}{1 - e^{-(\gamma-1)}}, \\ A_4 &= 1 + 2\|K\|_1, & A_5 &= \sqrt{A_1} + 2 \left(\sqrt{2\gamma A_1} + \frac{\gamma A_2}{3} \right) (1 + \|K\|_1), \\ A_6 &= 2(1 + \|K\|_1) \sqrt{A_3}, & \tilde{A}_1 &= A_1 \|K\|_1^2, & \tilde{A}_2 &= A_2 \|K\|_1, \\ C_* &= A_0^2 + A_1 & \text{y} & C^* &= A_0 A_4 + A_5 e^{\frac{1}{2}} + A_6. \end{aligned}$$

2.4.2. Demostración de la Proposición 4: Sesgo del estimador.

$$\begin{aligned} \mathbb{E}\{\hat{r}_h(x)\} &= \mathbb{E} \left\{ \frac{1}{n} \sum_{i=1}^n K_h(x - X_i) Y_i g^{-1}(X_i) \right\} \\ &= \mathbb{E} \left\{ \frac{1}{n} \sum_{i=1}^n K_h(x - X_i) r(X_i) g^{-1}(X_i) \right\} + \mathbb{E} \left\{ \frac{1}{n} \sum_{i=1}^n K_h(x - X_i) \varepsilon_i g^{-1}(X_i) \right\} \\ &= \mathbb{E} \{ K_h(x - X) r(X) g^{-1}(X) \} + \mathbb{E} \{ K_h(x - X) g^{-1}(X) \} \mathbb{E}\{\varepsilon\} \\ &= \int_{\mathbb{R}} K_h(x - t) r(t) dt = K_h * r(x) \end{aligned}$$

Así se tiene que,

$$\mathbb{E}[\hat{r}_h(x)] - r(x) = K_h * r(x) - r(x) = \int_{\mathbb{R}} (r(x - hu) - r(x)) K(u) du$$

De la fórmula de Taylor-Lagrange al orden l aplicado a $r(x)$ se tiene,

$$r(x - hu) = r(x) - hur'(x) + \frac{(-hu)^2}{2!} r''(x) - \dots + \frac{(-hu)^{l-1}}{(l-1)!} r^{(l-1)}(x) + \frac{(-hu)^l}{l!} r^{(l)}(x - \zeta_u hu)$$

donde $\zeta_u \in [0, 1]$.

Por lo anterior y como K es un núcleo de orden m con $m \geq l$ se tiene que,

$$\mathbb{E}[\hat{r}_h(x)] - r(x) = \int_{\mathbb{R}} \frac{(-hu)^l}{l!} r^{(l)}(x - \zeta_u hu) K(u) du$$

Dado que K es un núcleo de orden l al restar el termino nulo $\int_{\mathbb{R}} \frac{(-uh)^l}{l!} r^{(l)}(x) K(u) du$ a la ecuación anterior, se tiene que,

$$\mathbb{E}[\hat{r}_h(x)] - r(x) = (-1)^l \frac{h^l}{l!} \int_{\mathbb{R}} \left(r^{(l)}(x - \zeta_u hu) - r^{(l)}(x) \right) u^l K(u) du$$

Al tomar valor absoluto en ambos lados y usar la hipótesis $r \in \Sigma(\beta, L)$ se tiene,

$$\begin{aligned} |\mathbb{E}[\hat{r}_h(x)] - r(x)| &\leq \frac{h^\beta}{l!} L \int_{\mathbb{R}} |u|^\beta \zeta_u^{\beta-l} |K(u)| du \\ &\leq \frac{h^\beta}{l!} L \int_{\mathbb{R}} |u|^\beta |K(u)| du, \quad \text{pues } \zeta_u^{\beta-l} < 1 \\ &\leq h^\beta A_0. \end{aligned}$$

2.4.3. Demostración de la Proposición 5: Varianza del estimador.

Al sustituir Y_i por $r(X_i) + \varepsilon_i$, se obtiene,

$$\begin{aligned}
& \mathbb{E} [(\hat{r}_h(x) - \mathbb{E}[\hat{r}_h(x)])^2] \\
&= \mathbb{E} \left[\left\{ \frac{1}{n} \sum_{i=1}^n K_h(x - X_i) Y_i g^{-1}(X_i) - \mathbb{E} \left(\frac{1}{n} \sum_{i=1}^n K_h(x - X_i) Y_i g^{-1}(X_i) \right) \right\}^2 \right] \\
&= \frac{1}{n^2} \sum_{i=1}^n \text{Var} (K_h(x - X_i) Y_i g^{-1}(X_i)) + \frac{1}{n^2} \sum_{i \neq j} \text{Cov} (K_h(x - X_i) Y_i g^{-1}(X_i), \\
&\quad K_h(x - X_j) Y_j g^{-1}(X_j)) \\
&\leq \frac{1}{n^2} \sum_{i=1}^n \mathbb{E} [K_h^2(x - X_i) Y_i^2 g^{-2}(X_i)] \\
&= \frac{1}{n^2} \sum_{i=1}^n \mathbb{E} [K_h^2(x - X_i) (r(x_i))^2 g^{-2}(X_i)] \\
&+ \frac{1}{n^2} \sum_{i=1}^n \mathbb{E} [K_h^2(x - X_i) 2r(X_i) \varepsilon_i g^{-2}(X_i)] + \frac{1}{n^2} \sum_{i=1}^n \mathbb{E} [K_h^2(x - X_i) \varepsilon_i^2 g^{-2}(X_i)] \\
&\leq \frac{((r_{\text{sup}})^2 + \sigma^2) (g_{\text{inf}})^{-1}}{n^2 h^2} \sum_{i=1}^n \mathbb{E} \left[K^2 \left(\frac{x - X_i}{h} \right) g^{-1}(X_i) \right] \\
&= \frac{((r_{\text{sup}})^2 + \sigma^2) (g_{\text{inf}})^{-1}}{n h^2} \int_{\mathbb{R}} K^2 \left(\frac{x - t}{h} \right) dt \\
&= \frac{A_1}{n h}
\end{aligned}$$

donde $A_1 = ((r_{\text{sup}})^2 + \sigma^2) (g_{\text{inf}})^{-1} \|K\|_2^2$.

2.4.4. Demostración de la Proposición 7.

Sea $h \in \mathcal{H}$. Tenemos

$$|r(x) - \hat{r}_{\hat{h}}(x)| \leq |r(x) - \hat{r}_h(x)| + |\hat{r}_h(x) - \hat{r}_{h, \hat{h}}(x)| + |\hat{r}_{\hat{h}}(x) - \hat{r}_{h, \hat{h}}(x)|.$$

Al sumar y restar $V(h)$ y $V(\hat{h})$, tenemos

$$\begin{aligned}
|r(x) - \hat{r}_{\hat{h}}(x)| &\leq |r(x) - \hat{r}_h(x)| + A(h, x) + V(h) + A(\hat{h}, x) + V(\hat{h}) \\
&\leq |r(x) - \hat{r}_h(x)| + 2(A(h, x) + V(h)).
\end{aligned} \tag{2.10}$$

Por otro lado,

$$\begin{aligned} |\hat{r}_{h,h'}(x) - \hat{r}_{h'}(x)| &\leq |\hat{r}_{h,h'}(x) - \mathbb{E}(\hat{r}_{h,h'}(x))| + |\hat{r}_{h'}(x) - \mathbb{E}(\hat{r}_{h'}(x))| \\ &\quad + |\mathbb{E}(\hat{r}_{h,h'}(x)) - \mathbb{E}(\hat{r}_{h'}(x))| \end{aligned}$$

y si restamos $V(h')$ en ambos miembros y tomamos parte positiva tenemos,

$$\begin{aligned} &\{|\hat{r}_{h,h'}(x) - \hat{r}_{h'}(x)| - V(h')\}_+ \\ &\leq \{|\hat{r}_{h,h'}(x) - \mathbb{E}(\hat{r}_{h,h'}(x))| + |\hat{r}_{h'}(x) - \mathbb{E}(\hat{r}_{h'}(x))| - V(h')\}_+ \\ &\quad + |\mathbb{E}(\hat{r}_{h,h'}(x)) - \mathbb{E}(\hat{r}_{h'}(x))| \end{aligned} \tag{2.11}$$

Recuerde que la esperanza del estimador es $\mathbb{E}(\hat{r}_h(x)) = (K_h * r)(x)$, por lo cuál el segundo término del lado derecho de la desigualdad anterior, se expresa como

$$\begin{aligned} |\mathbb{E}(\hat{r}_{h,h'}(x)) - \mathbb{E}(\hat{r}_{h'}(x))| &= |(K_{h'} * r)(x) - (K_{h'} * (K_h * r))(x)| \\ &= |(K_{h'} * (r - (K_h * r)))(x)| \\ &\leq \int |K_{h'}(x-u)(K_h * r(u) - r(u))| du. \end{aligned}$$

Como el soporte de K es $[-1, 1]$, se tiene que $\frac{x-u}{h'} \in [-1, 1]$, de donde $u \in [x-h', x+h'] \subset B(x)$, lo que implica

$$\begin{aligned} |\mathbb{E}(\hat{r}_{h,h'}(x)) - \mathbb{E}(\hat{r}_{h'}(x))| &\leq \int_{B(x)} \left| \frac{1}{h'} K\left(\frac{x-u}{h'}\right) \right| |(K_h * r(u) - r(u))| du \\ &\leq \max_{u \in B(x)} |(K_h * r(u) - r(u))| \int \left| \frac{1}{h'} K\left(\frac{x-u}{h'}\right) \right| du \\ &\leq \|K\|_1 \max_{u \in B(x)} |(K_h * r(u) - r(u))| \\ &\leq C(h) \|K\|_1. \end{aligned}$$

Usando que $V(h') = V_1(h') + V_2(h')$, la desigualdad anterior y la desigualdad (2.11) se tiene que,

$$\begin{aligned} &\{|\hat{r}_{h,h'}(x) - \hat{r}_{h'}(x)| - V(h')\}_+ \\ &\leq \{|\hat{r}_{h,h'}(x) - \mathbb{E}(\hat{r}_{h,h'}(x))| - V_2(h')\}_+ + \{|\hat{r}_{h'}(x) - \mathbb{E}(\hat{r}_{h'}(x))| - V_1(h')\}_+ \\ &\quad + C(h) \|K\|_1. \end{aligned} \tag{2.12}$$

Por la definición de $A(h, x)$ y al tomar en la desigualdad (2.12) máximo sobre h' en \mathcal{H} , se tiene que,

$$A(h, x) \leq T_1 + T_2 + C(h) \|K\|_1. \tag{2.13}$$

Al tomar la norma $\mathbb{E}[\cdot]^2]^{\frac{1}{2}}$ determinada por la esperanza en las ecuaciones (2.10) y (2.13) se obtiene

$$\begin{aligned} & \left(\mathbb{E} (r(x) - \hat{r}_h(x))^2 \right)^{\frac{1}{2}} \\ & \leq \left(\mathbb{E} (r(x) - \hat{r}_h(x))^2 \right)^{\frac{1}{2}} + 2 \left((\mathbb{E} (A(h, x))^2)^{\frac{1}{2}} + V(h) \right) \\ & \leq \left(\mathbb{E} (r(x) - \hat{r}_h(x))^2 \right)^{\frac{1}{2}} + C(h) \|K\|_1 + V(h) + (\mathbb{E} (T_1^2))^{\frac{1}{2}} + (\mathbb{E} (T_2^2))^{\frac{1}{2}} \end{aligned}$$

2.4.5. Demostración de la Proposición 8, (i).

se va a demostrar que,

$$\mathbb{E}(T_1^2) \leq \frac{A_3}{n}$$

donde $A_3 = \frac{32 \max(A_1, A_2^2)}{1 - e^{-(\gamma-1)}}$ con $A_1 = ((r_{\sup})^2 + \sigma^2) (g_{\inf})^{-1} \|K\|_2^2$, $A_2 = (r_{\sup} + b_0) (g_{\inf})^{-1} \|K\|_{\infty}$ y $\gamma > 1$. Para estudiar $\mathbb{E}(T_1^2)$ se usa el siguiente hecho,

$$\begin{aligned} \mathbb{E}(T_1^2) & \leq \max_{h' \in \mathcal{H}} \mathbb{E} \left[\left\{ |\hat{r}_{h'}(x) - \mathbb{E}(\hat{r}_{h'}(x))| - V_1(h') \right\}_+^2 \right] \\ & \leq \sum_{h' \in \mathcal{H}} \mathbb{E} \left[\left\{ |\hat{r}_{h'}(x) - \mathbb{E}(\hat{r}_{h'}(x))| - V_1(h') \right\}_+^2 \right] \end{aligned} \quad (2.14)$$

Ahora se analiza el término,

$$\begin{aligned} & \mathbb{E} \left(\left\{ |\hat{r}_{h'}(x) - \mathbb{E}(\hat{r}_{h'}(x))| - V_1(h') \right\}_+^2 \right) \\ & = \int_0^{+\infty} \mathbb{P} \left(\left\{ |\hat{r}_{h'}(x) - \mathbb{E}(\hat{r}_{h'}(x))| - V_1(h') \right\}_+^2 > t \right) dt \\ & = \int_0^{+\infty} \mathbb{P} \left(\left\{ |\hat{r}_{h'}(x) - \mathbb{E}(\hat{r}_{h'}(x))| - V_1(h') \right\}_+ > t^{\frac{1}{2}} \right) dt \\ & = \int_0^{+\infty} \mathbb{P} \left(|\hat{r}_{h'}(x) - \mathbb{E}(\hat{r}_{h'}(x))| > V_1(h') + t^{1/2} \right) dt \end{aligned} \quad (2.15)$$

pues para $t > 0$, $\mathbb{P}(\{T\}_+ > t) = \mathbb{P}(T > t)$ ya que $\mathbb{P}(T > t \wedge T < 0) = 0$.

En el integrando $\mathbb{P} \left(|\hat{r}_{h'}(x) - \mathbb{E}(\hat{r}_{h'}(x))| > V_1(h') + t^{1/2} \right)$ de la ecuación (2.15), se observa que,

$$\hat{r}_{h'}(x) - \mathbb{E}(\hat{r}_{h'}(x)) = S_n \quad (2.16)$$

donde $S_n = \sum_{i=1}^n (Z_i - \mathbb{E}(Z_i))$ con $Z_i = \frac{1}{n} K_{h'}(x - X_i) Y_i g^{-1}(X_i)$ para $h' \in \mathcal{H}$.

Como,

$$|Z_i| = \frac{1}{nh'} \left| K \left(\frac{x - X_i}{h'} \right) Y_i g^{-1}(X_i) \right| \leq \frac{A_2}{nh'} = b \quad (2.17)$$

donde $A_2 = (r_{\text{sup}} + b_0)(g_{\text{inf}})^{-1} \|K\|_{\infty}$.

y

$$\begin{aligned} & \sum_{i=1}^n \mathbb{E}(Z_i^2) \\ & \leq \frac{(g_{\text{inf}})^{-1}}{n^2(h')^2} \sum_{i=1}^n \mathbb{E} \left(K^2 \left(\frac{x - X_i}{h'} \right) (r^2(X_i) + 2r(X_i)\varepsilon_i + \varepsilon_i^2) g^{-1}(X_i) \right) \\ & = \frac{(g_{\text{inf}})^{-1}}{n^2(h')^2} \sum_{i=1}^n \left\{ \mathbb{E} \left(K^2 \left(\frac{x - X_i}{h'} \right) r^2(X_i) g^{-1}(X_i) \right) + \mathbb{E} \left(K^2 \left(\frac{x - X_i}{h'} \right) \varepsilon_i^2 g^{-1}(X_i) \right) \right\} \\ & \leq \frac{((r_{\text{sup}})^2 + \sigma^2)(g_{\text{inf}})^{-1}}{n^2(h')^2} \sum_{i=1}^n \mathbb{E} \left(K^2 \left(\frac{x - X_i}{h'} \right) g^{-1}(X_i) \right) \\ & \leq \frac{A_1}{nh'} = v \end{aligned} \quad (2.18)$$

donde $A_1 = ((r_{\text{sup}})^2 + \sigma^2)(g_{\text{inf}})^{-1} \|K\|_2^2$.

Entonces, se satisfacen las hipótesis la proposición 2 (desigualdad de Bernstein), en el integrando de la ecuación (2.15).

Tomando en cuenta la proposición 2 se elige $\lambda(u) = \sqrt{2vu} + \frac{b}{3}u$, donde b y v están dados en las ecuaciones (2.17) y (2.18). Se demostrará que $\lambda(\gamma|\log h'|) \leq V_1(h')$, donde $V_1(h')$ se especifica en el enunciado de la Proposición 7.

$$\begin{aligned}
\lambda (\gamma |\log h'|) &= \sqrt{2\nu\gamma |\log h'|} + \frac{b}{3} \gamma |\log h'| \\
&= \sqrt{2 \frac{A_1}{nh'} \gamma |\log h'|} + \frac{A_2}{3nh'} \gamma |\log h'| \\
&= \sqrt{2\gamma A_1} \frac{|\log h'|^{\frac{1}{2}}}{(nh')^{\frac{1}{2}}} + \frac{\gamma A_2}{3} \frac{|\log h'|}{nh'}
\end{aligned} \tag{2.19}$$

Para poder controlar la cantidad anterior es fundamental el hecho de que para cada $h_i \in \mathcal{H}$ con $i = 0, \dots, M$ y $M = \left\lceil \log \left(\frac{n}{\log n} \right) \right\rceil$, se tiene que $\left(\frac{n}{\log n} \right)^{-1} \leq h_i \leq 1$.

Como $\frac{\log n}{n} \leq h'$ entonces $\frac{1}{nh'} \leq \frac{1}{\log n}$ de donde $\frac{|\log h'|}{nh'} \leq \frac{|\log h'|}{\log n}$ de lo que se obtiene que,

$$\frac{|\log h'|^{\frac{1}{2}}}{(nh')^{\frac{1}{2}}} \leq \frac{|\log h'|^{\frac{1}{2}}}{(\log n)^{\frac{1}{2}}} \tag{2.20}$$

Por otra parte también del hecho de que $h' \geq \frac{\log n}{n}$, se tiene que $\log h' \geq \log(\log n) - \log n$, de donde $-\log h' \leq \log n - \log(\log n)$, por lo cual $|\log h'| \leq \log n - \log(\log n) \leq \log n$, es decir,

$$\frac{|\log h'|}{\log n} \leq 1 \tag{2.21}$$

Por las desigualdades (2.20) y (2.21) se tiene que,

$$\frac{|\log h'|^{\frac{1}{2}}}{(nh')^{\frac{1}{2}}} \leq \frac{|\log h'|^{\frac{1}{2}}}{(\log n)^{\frac{1}{2}}} \leq 1$$

De la desigualdad anterior se puede concluir que,

$$\frac{|\log h'|}{nh'} \leq \frac{|\log h'|^{\frac{1}{2}}}{(nh')^{\frac{1}{2}}}$$

Por esta última desigualdad y la desigualdad (2.19), se tiene que,

$$\lambda (\gamma |\log h'|) \leq \left(\sqrt{2\gamma A_1} + \frac{\gamma A_2}{3} \right) \frac{|\log h'|^{\frac{1}{2}}}{(nh')^{\frac{1}{2}}}$$

De la desigualdad anterior y la desigualdad (2.21) se puede concluir que,

$$\begin{aligned}\lambda(\gamma|\log h'|) &\leq \left(\sqrt{2\gamma A_1} + \frac{\gamma A_2}{3}\right) \frac{(\log n)^{\frac{1}{2}}}{(nh')^{\frac{1}{2}}} \\ &= V_1(h')\end{aligned}$$

Por las ecuaciones (2.15) y (2.16), junto con la desigualdad anterior, se tiene que,

$$\mathbb{E}\left(\left\{|\hat{r}_{h'}(x) - \mathbb{E}(\hat{r}_{h'}(x))| - V_1(h')\right\}_+^2\right) \leq \int_0^{+\infty} \mathbb{P}\left(|S_n| > \lambda(\gamma|\log h'|) + t^{1/2}\right) dt \quad (2.22)$$

Se considera el cambio de variable $t^{1/2} = \lambda(u)$, de donde, $t = \lambda^2(u)$, por lo cual $dt = 2\lambda(u)\lambda'(u)du$, luego se sustituye el cambio de variable en el lado derecho de la desigualdad anterior, para obtener,

$$\begin{aligned}&\int_0^{+\infty} \mathbb{P}\left(|S_n| > \lambda(\gamma|\log h'|) + t^{1/2}\right) dt \\ &= \int_0^{+\infty} \mathbb{P}\left\{|S_n| > \lambda(\gamma|\log(h')|) + \lambda(u)\right\} 2\lambda'(u)\lambda(u)du\end{aligned} \quad (2.23)$$

Por el lado derecho de la ecuación anterior y los siguientes hechos:

- Como $\lambda(u) = \sqrt{2vu} + \frac{b}{3}u$ entonces $\lambda(u) = \sqrt{2v}\sqrt{u} + \frac{b}{3}u$, de donde $\lambda'(u) = \sqrt{2v}\frac{1}{2\sqrt{u}} + \frac{b}{3}$, al multiplicar por u se tiene que $\lambda'(u)u = \frac{\sqrt{2vu}}{2} + \frac{b}{3}u \leq \sqrt{2vu} + \frac{b}{3}u = \lambda(u)$, de lo que se concluye que $\lambda'(u)\lambda(u) \leq \frac{\lambda^2(u)}{u}$.
- λ es subaditiva, es decir $\lambda(p+q) \leq \lambda(p) + \lambda(q)$.
- Si a y b son números reales tales que $a < b$ y X es una variable aleatoria entonces $\{X > b\} \subset \{X > a\}$, por lo cual $\mathbb{P}\{X > b\} \leq \mathbb{P}\{X > a\}$

Se tiene que,

$$\begin{aligned}&\int_0^{+\infty} \mathbb{P}\left\{|S_n| > \lambda(\gamma|\log(h')|) + \lambda(u)\right\} 2\lambda'(u)\lambda(u)du \\ &\leq 2 \int_0^{+\infty} \frac{\lambda^2(u)}{u} \mathbb{P}\left\{|S_n| > \lambda(\gamma|\log(h')|) + u\right\} du\end{aligned}$$

Por la ecuación (2.23) y la desigualdad anterior se tiene,

$$\begin{aligned} & \int_0^{+\infty} \mathbb{P} \left(|S_n| > \lambda \left(\gamma |\log h'| \right) + t^{1/2} \right) dt \\ & \leq 2 \int_0^{+\infty} \frac{\lambda^2(u)}{u} \mathbb{P} \{ |S_n| > \lambda \left(\gamma |\log h'| + u \right) \} du \end{aligned} \quad (2.24)$$

Como $\lambda(u) = \sqrt{2vu} + \frac{b}{3}u$ se tiene que $\lambda^2(u) \leq 2 \left((\sqrt{2vu})^2 + \left(\frac{b}{3}u\right)^2 \right) = 4vu + 2\frac{b^2}{9}u^2 \leq 4vu + 2b^2u^2$ es decir $\lambda^2(u) \leq 4vu + 2b^2u^2$, por esta última desigualdad y al aplicar la desigualdad Bernstein (Proposición 2) en la integral del lado derecho de la desigualdad (2.24) se tiene que,

$$\begin{aligned} & \int_0^{+\infty} \frac{\lambda^2(u)}{u} \mathbb{P} \{ |S_n| > \lambda \left(\gamma |\log h'| + u \right) \} du \\ & \leq \int_0^{+\infty} \frac{4vu + 2b^2u^2}{u} 2 \exp\{ -(\gamma |\log h'| + u) \} du \\ & = 2 \exp\{ -\gamma |\log h'| \} \int_0^{+\infty} (4v + 2b^2u) \exp\{ -u \} du \\ & = 2 \exp\{ -\gamma |\log h'| \} (4v + 2b^2) \\ & \leq 8(v + b^2) \exp\{ -\gamma |\log h'| \} \end{aligned} \quad (2.25)$$

Por las ecuaciones (2.14), (2.22), (2.24) y (2.25) se tiene que,

$$\mathbb{E}(T_1^2) \leq \sum_{h' \in \mathcal{H}} 16(v + b^2) \exp\{ -\gamma |\log h'| \}$$

Por la desigualdad anterior y como $v = \frac{A_1}{nh'}$ y $b = \frac{A_2}{nh'}$ entonces,

$$\mathbb{E}(T_1^2) \leq \sum_{h' \in \mathcal{H}} 16 \left(\frac{A_1}{nh'} + \frac{A_2^2}{(nh')^2} \right) \exp\{ -\gamma |\log h'| \} \quad (2.26)$$

Como $h' \in \mathcal{H}$ entonces $\frac{\log n}{n} \leq h' \leq 1$, por lo cual $nh' \geq 1$, de lo que se desprende que $\frac{1}{(nh')^2} \leq \frac{1}{nh'}$, por este último resultado y la desigualdad (2.26) se tiene que,

$$\begin{aligned} \mathbb{E}(T_1^2) & \leq \sum_{h' \in \mathcal{H}} \frac{16}{nh'} (A_1 + A_2^2) \exp\{ -\gamma |\log h'| \} \\ & \leq \frac{16(A_1 + A_2^2)}{n} \sum_{h' \in \mathcal{H}} \frac{1}{h'} \exp\{ -\gamma |\log h'| \} \end{aligned}$$

Como $h' \leq 1$ entonces $|\log(h')| = -\log(h')$, al sustituir esto en el resultado anterior se tiene que,

$$\begin{aligned} \mathbb{E}(T_1^2) &\leq \frac{16(A_1 + A_2^2)}{n} \sum_{h' \in \mathcal{H}} \frac{1}{h'} \exp\{\gamma \log h'\} \\ &= \frac{16(A_1 + A_2^2)}{n} \sum_{h' \in \mathcal{H}} (h')^{\gamma-1} \\ &\leq \frac{16(A_1 + A_2^2)}{n} \sum_{i=0}^{\infty} (e^{-i})^{\gamma-1} \\ &\leq \frac{32 \max(A_1, A_2^2)}{n} \frac{1}{1 - e^{-(\gamma-1)}} \end{aligned}$$

con $\gamma > 1$.

Por la desigualdad anterior se concluye que,

$$\mathbb{E}(T_1^2) \leq \frac{A_3}{n} \quad (2.27)$$

donde $A_3 = \frac{32 \max(A_1, A_2^2)}{1 - e^{-(\gamma-1)}}$ con $A_1 = ((r_{\text{sup}})^2 + \sigma^2) (g_{\text{inf}})^{-1} \|K\|_2^2$, $A_2 = (r_{\text{sup}} + b_0) (g_{\text{inf}})^{-1} \|K\|_{\infty}$ y $\gamma > 1$.

2.4.6. Demostración de la Proposición 8, (ii) .

Para estudiar $\mathbb{E}(T_2^2)$ se usa de manera similar a T_1 que

$$\mathbb{E}(T_2^2) \leq \sum_{h' \in \mathcal{H}} \mathbb{E} \left[\{ |\hat{r}_{hh'}(x) - \mathbb{E}(\hat{r}_{hh'}(x))| - V_2(h') \}_+^2 \right] \quad (2.28)$$

De igual manera a la demostración de la parte (i):

$$\begin{aligned} &\mathbb{E} \left(\{ |\hat{r}_{hh'}(x) - \mathbb{E}(\hat{r}_{hh'}(x))| - V_2(h') \}_+^2 \right) \\ &= \int_0^{+\infty} \mathbb{P} \left(|\hat{r}_{hh'}(x) - \mathbb{E}(\hat{r}_{hh'}(x))| > V_2(h') + t^{1/2} \right) dt \end{aligned} \quad (2.29)$$

En el integrando $\mathbb{P} \left(|\hat{r}_{hh'}(x) - \mathbb{E}(\hat{r}_{hh'}(x))| > V_1(h') + t^{1/2} \right)$ de la ecuación (2.29), se observa que,

$$\hat{r}_{hh'}(x) - \mathbb{E}(\hat{r}_{hh'}(x)) = \tilde{S}_n \quad (2.30)$$

donde $\tilde{S}_n = \sum_{i=1}^n (\tilde{Z}_i - \mathbb{E}(\tilde{Z}_i))$ con $\tilde{Z}_i = \frac{1}{n} K_{hh'}(x - X_i) Y_i g^{-1}(X_i)$ y $K_{hh'}(\cdot) = K_h * K_{h'}(\cdot)$, para $h, h' \in \mathcal{H}$.

Como $\|K_{hh'}\|_\infty \leq \|K_{h'}\|_\infty \|K\|_1$ y $\|K_h\|_\infty \leq \frac{1}{h} \|K\|_\infty$ entonces,

$$\begin{aligned} |\tilde{Z}_i| &= \left| \frac{1}{n} K_{hh'}(x - X_i) Y_i g^{-1}(X_i) \right| \\ &\leq \frac{(r_{\text{sup}} + b_0)(g_{\text{inf}})^{-1} \|K_{hh'}\|_\infty}{n} \\ &\leq \frac{(r_{\text{sup}} + b_0)(g_{\text{inf}})^{-1} \|K_{h'}\|_\infty \|K\|_1}{n} \\ &\leq \frac{(r_{\text{sup}} + b_0)(g_{\text{inf}})^{-1} \|K\|_\infty \|K\|_1}{nh'} \\ &\leq \frac{\tilde{A}_2}{nh'} = \tilde{b} \end{aligned} \quad (2.31)$$

donde $\tilde{A}_2 = A_2 \|K\|_1$.

Por otra parte, como $\|K_{hh'}\|_2 \leq \|K_{h'}\|_2 \|K_h\|_1$, $\|K_{h'}\|_2^2 = \frac{1}{h'} \|K\|_2^2$ y $\|K_h\|_1 = \|K\|_1$ entonces,

$$\begin{aligned}
& \sum_{i=1}^n \mathbb{E}(\tilde{Z}_i^2) \\
& \leq \frac{(g_{\inf})^{-1}}{n^2} \sum_{i=1}^n \mathbb{E} (K_{hh'}^2(x - X_i) (r^2(X_i) + 2r(X_i)\varepsilon_i + \varepsilon_i^2) g^{-1}(X_i)) \\
& = \frac{(g_{\inf})^{-1}}{n^2} \sum_{i=1}^n \{ \mathbb{E} (K_{hh'}^2(x - X_i) r^2(X_i) g^{-1}(X_i)) + \mathbb{E} (K_{hh'}^2(x - X_i) \varepsilon_i^2 g^{-1}(X_i)) \} \\
& \leq \frac{((r_{\sup})^2 + \sigma^2) (g_{\inf})^{-1}}{n^2} \sum_{i=1}^n \mathbb{E} (K_{hh'}^2(x - X_i) g^{-1}(X_i)) \\
& \leq \frac{((r_{\sup})^2 + \sigma^2) (g_{\inf})^{-1}}{n} \int K_{hh'}^2(x - t) dt \\
& = \frac{((r_{\sup})^2 + \sigma^2) (g_{\inf})^{-1} \|K_{hh'}\|_2^2}{n} \\
& \leq \frac{((r_{\sup})^2 + \sigma^2) (g_{\inf})^{-1} \|K_{h'}\|_2^2 \|K_h\|_1^2}{n} \\
& = \frac{((r_{\sup})^2 + \sigma^2) (g_{\inf})^{-1} \|K\|_2^2 \|K\|_1^2}{nh'} \\
& = \frac{\tilde{A}_1}{nh'} = \tilde{\nu}
\end{aligned} \tag{2.32}$$

donde $\tilde{A}_1 = A_1 \|K\|_1^2$.

Así se tiene que, se satisfacen las hipótesis de la proposición 2 (Desigualdad de Bernstein), en el integrando de la ecuación (2.29).

Tomando en cuenta la proposición 2 se elige $\tilde{\lambda}(u) = \sqrt{2\tilde{\nu}u} + \frac{\tilde{b}}{3}u$, donde \tilde{b} y $\tilde{\nu}$ están dados en las ecuaciones (2.31) y (2.32). Se demostrara que $\tilde{\lambda}(\gamma|\log h'|) \leq V_2(h')$, donde $V_2(h')$ se especifica en el enunciado de la Proposición 7.

Por un procedimiento análogo al planteado en la Proposición 8 (i).

$$\begin{aligned}
\tilde{\lambda}(\gamma|\log h'|) & \leq \left(\sqrt{2\gamma\tilde{A}_1} + \frac{\gamma\tilde{A}_2}{3} \right) \frac{|\log n|^{\frac{1}{2}}}{(nh')^{\frac{1}{2}}} \\
& \leq \|K\|_1 \left(\sqrt{2\gamma A_1} + \frac{\gamma A_2}{3} \right) \frac{|\log n|^{\frac{1}{2}}}{(nh')^{\frac{1}{2}}} \\
& = V_2(h')
\end{aligned}$$

Por las ecuaciones (2.29) y (2.30), junto con la desigualdad anterior, se tiene que,

$$\mathbb{E} \left(\left\{ |\hat{r}_{hh'}(x) - \mathbb{E}(\hat{r}_{hh'}(x))| - V_2(h') \right\}_+^2 \right) \leq \int_0^{+\infty} \mathbb{P} \left(|\tilde{S}_n| > \tilde{\lambda} (\gamma |\log h'|) + t^{1/2} \right) dt$$

Siguiendo el esquema de demostración planteado en la Proposición 8 (i), se tiene que,

$$\begin{aligned} \mathbb{E}(T_2^2) &\leq \frac{32 \max(\tilde{A}_1, \tilde{A}_2^2)}{n} \frac{1}{1 - e^{-(\gamma-1)}} \\ &= \frac{32 \max(A_1 \|K\|_1^2, A_2^2 \|K\|_1^2)}{1 - e^{-(\gamma-1)}} \frac{1}{n} \\ &= \frac{32 \max(A_1, A_2^2) \|K\|_1^2}{1 - e^{-(\gamma-1)}} \frac{1}{n} \\ &= \frac{A_3 \|K\|_1^2}{n} \end{aligned}$$

Capítulo 3

Estimación adaptativa en modelos de regresión para datos débilmente dependientes y variable explicativa con densidad conocida

Resumen

Este capítulo está dedicado a la estimación de la función de regresión cuando ésta es acotada, los errores del modelo son variables aleatorias i.i.d de distribución normal y la variable explicativa es un proceso débilmente dependiente cuyo coeficiente de correlación tiene decaimiento exponencial, además la variable explicativa es idénticamente distribuida con función de densidad conocida y acotada. La precisión de la estimación se mide utilizando riesgos puntuales. Se propone un procedimiento basado en datos utilizando estimación por Núcleo. El ancho de ventana se selecciona usando el enfoque de Goldenshluger-Lepski y se demuestra que el estimador resultante satisface una desigualdad tipo oráculo. El procedimiento también ha demostrado ser adaptativo. Algunas simulaciones ilustran el desempeño del método propuesto.

3.1. Modelo

Observamos $\{(X_1, Y_1), \dots, (X_n, Y_n)\}$ una muestra idénticamente distribuidas que satisfacen

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

donde $n \geq 2$, los X_i son idénticamente distribuidos con función de densidad g conocida, los ε_i son independientes e idénticamente distribuidos de distribución normal con media nula y varianza $\sigma^2 > 0$, los X_i son independientes de los ε_i , y r es la función de regresión $r(x) = \mathbb{E}(Y|X = x)$.

Nuestro objetivo es estimar la función r en el punto $x \in \mathbb{R}$ usando la muestra observada $\{(X_1, Y_1), \dots, (X_n, Y_n)\}$ bajo un enfoque de dependencia débil.

Se mide la calidad de la estimación de un estimador \hat{r} usando el error cuadrático en un punto $x \in \mathbb{R}$.

$$R(r, \hat{r}, x) = \mathbb{E}(r(x) - \hat{r}(x))^2.$$

Por un lado, se asumen las siguientes hipótesis correspondientes a cotas sobre las densidades de las variables X_i en la vecindad de x :

$$B(x) = [x - 2/(\log n)^2, x + 2/(\log n)^2].$$

(H₁) La densidad g de las X_i satisface

$$g_{\inf} \leq g(u) \leq g_{\sup}, \quad \forall u \in B(x)$$

donde g_{\inf} y g_{\sup} son constantes positivas.

(H₂) $\forall (i, j)$, la densidad conjunta $g_{i,j}$ de (X_i, X_j) satisface

$$|g_{i,j}(u, v)| \leq Q, \quad \forall u, v \in B(x),$$

donde Q es una constante positiva.

Por otro lado, asumimos una estructura de dependencia débil sobre las variables X_i . Mas precisamente, para u y v enteros positivos, denotamos $i_{1:u} = (i_1, \dots, i_u) \in \mathbb{Z}^u$ y $j_{1:v} = (j_1, \dots, j_v) \in \mathbb{Z}^v$, entonces se definen los vectores aleatorios $X_{i_{1:u}} = (X_{i_1}, X_{i_2}, \dots, X_{i_u})$ y $X_{j_{1:v}} = (X_{j_1}, X_{j_2}, \dots, X_{j_v})$ a valores en \mathbb{R}^u y \mathbb{R}^v respectivamente. Se define la función $q: \mathbb{Z}^u \times \mathbb{Z}^v \rightarrow \mathbb{Z}$ por $q(i_{1:u}, j_{1:v}) = \min(j_{1:v}) - \max(i_{1:u})$ y Λ_u la clase de funciones $G_u: \mathbb{R}^u \rightarrow \mathbb{R}$ tales que $\|G_u\|_\infty = \sup_{x \in \mathbb{R}^u} |G_u(x)| < \infty$. Para un proceso aleatorio $\mathbb{X} = (X_i)_{i \in \mathbb{Z}}$, se define el coeficiente de correlación $\alpha(\mathbb{X}) = (\alpha_k(\mathbb{X}))_{k \in \mathbb{N}}$ por,

$$\alpha_k(\mathbb{X}) = \sup_{u, v \in \mathbb{N}} \sup_{(i_{1:u}, j_{1:v}) \in \mathbb{Z}^u \times \mathbb{Z}^v, q(i_{1:u}, j_{1:v}) \geq k} \sup_{G_u \in \Lambda_u} \sup_{G_v \in \Lambda_v} \frac{|cov(G_u(X_{i_{1:u}}), G_v(X_{j_{1:v}}))|}{\Psi(u, v, G_u, G_v)} \quad (3.1)$$

con $\Psi(u, v, G_u, G_v) = 4\|G_u\|_\infty\|G_v\|_\infty$. Ver Sección 2 en Bertin and Klutchnikoff (2017) .

Asumimos que el coeficiente α -mixing de las variables X_i satisfacen la siguiente hipótesis.

(H₃) Existe $a \in]0, 1[$, tal que

$$\alpha_k(\mathbb{X}) \leq a^k, \quad \forall k \in \mathbb{N}.$$

En lo que sigue vamos a denotar por \mathcal{L} , el conjunto de los procesos \mathbb{X} que satisfacen (H₁), (H₂) y (H₃).

A continuación, se dan dos ejemplos de procesos que satisfacen estas hipótesis.

Ejemplo 1. Consideramos $\mathbb{Z} = \{Z_t\}_{t \geq 1}$, un proceso autoregresivo de orden 1, definido por $Z_t = \phi Z_{t-1} + \rho \xi_t$, con $\xi_t \sim N(0, 1)$, $|\phi| < 1$ y $\rho > 0$. Por recurrencia $Z_t = \phi^n Z_{t-n} + \rho \sum_{i=0}^{n-1} \phi^i \xi_{t-i}$ y al tomar $n \rightarrow \infty$ se tiene $Z_t = \rho \sum_{i=0}^{\infty} \phi^i \xi_{t-i}$ resulta un proceso gaussiano centrado y estacionario con función de covarianza $\gamma_{\mathbb{Z}}(k) = \text{cov}(Z_t, Z_{t+k}) = \frac{\rho^2}{1-\phi^2} \phi^k$ para $k \geq 0$. Se puede demostrar que el proceso \mathbb{Z} es débilmente dependiente, por lo cual satisface la hipótesis (H₃), ver Doukhan (1994).

Ejemplo 2. Al denotar $\phi_{\mu, \sigma^2}(x)$ como la función de densidad $N(\mu, \sigma^2)$ y $\Phi_{\mu, \sigma^2}(x)$ como la función de distribución acumulada $N(\mu, \sigma^2)$. Para $c \in \mathbb{R}^+$, las funciones de densidad y de distribución acumulada normal truncada en $[-c, c]$ de media cero y varianza uno, están dadas por $g(x) = \frac{\phi_{0,1}(x)}{p} \mathbf{1}_{[-c,c]}(x)$, con $p = \Phi_{0,1}(c) - \Phi_{0,1}(-c)$, para cada $x \in \mathbb{R}$ y

$$G(x) = \begin{cases} 0 & \text{si } x < -c \\ \frac{1}{p} (\Phi_{0,1}(x) - \Phi_{0,1}(-c)) & \text{si } -c \leq x < c \\ 1 & \text{si } x \geq c \end{cases}$$

respectivamente. Además, la función inversa de G es $G^{-1}(u) = \Phi_{0,1}^{-1}(pu + \Phi_{0,1}(c))$ para cada $u \in [0, 1]$.

Se puede demostrar que el proceso $\mathbb{X} = \{X_t\}_{t \geq 1}$ definido por $X_t = \left(G^{-1} \circ \Phi_{0, \frac{\rho^2}{1-\phi^2}} \right) (Z_t)$ cumple que $\mathbb{X} \in \mathcal{L}$, donde el proceso $\mathbb{Z} = \{Z_t\}_{t \geq 1}$ con cada Z_t de distribución $\Phi_{0, \frac{\rho^2}{1-\phi^2}}$ es el proceso del Ejemplo 1.

Finalmente, asumimos la siguiente hipótesis sobre la función de regresión

(H_4) Existe una constante $r_{\text{sup}} > 0$ tal que

$$|r(u)| \leq r_{\text{sup}}, \quad \forall u \in B(x).$$

Ahora, en la Sección 3.2, *Procedimiento estadístico*; se presenta el estimador por Núcleo de la función de regresión correspondiente al caso densidad conocida y se describe el procedimiento de selección de ventana basado en el método Goldenhluger-Lepski. Se observa que la familia de ventanas en este caso es mas acotado que en el caso independiente. Posteriormente se presenta la Sección 3.3, *Resultados*; que esta dividido en tres partes. Subsección 3.3.1, *Resultados para el estimador \hat{r}_h* ; donde se dan las definiciones y propiedades usuales que debe tener un buen estimador por Núcleo, sesgo, varianza, consistencia y velocidad de convergencia, el reto ahora es que la variable explicativa es débil dependiente, el coeficiente de dependencia tiene decaimiento exponencial y los errores del modelo ya no son acotados ahora son i.i.d $N(0, \sigma^2)$. La demostración del sesgo es invariante al cambio de hipótesis, pero en el caso de la varianza la demostración se hace un poco mas extensa, luego con procedimientos estándares se determina la consistencia y velocidad de convergencia, detalles de las demostraciones se pueden ver en la Sección 3.5, *Demostraciones*. Subsección 3.3.2, *Desigualdad de Oráculo y adaptatividad del estimador*; acá el desafío del cambio de hipótesis se hace mas fuerte, pues en este punto para demostrar la desigualdad de tipo oráculo surge la necesidad de usar la desigualdad de Bernstein para datos dependientes propuesta en Doukhan and Neumann (2007) y generar una serie de lemas técnicos siguiendo un esquema de demostración propuesto en Bertin and Klutchnikoff (2017). Además, al no ser acotada la variable respuesta del modelo, se descompone el estimador por Núcleo en una parte no acotada y otra acotada, usando el estimador de truncamiento. La parte aleatoria no acotada se controla usando débil dependencia y la parte aleatoria acotada se controla usando la desigualdad de Bernstein para datos dependientes. Esta técnica para controlar el estimador por Núcleo permiten demostrar que el estimador en la ventana seleccionada por el método G-L satisface una desigualdad de tipo oráculo. Esta demostración y los lemas, proposiciones y corolarios técnicos que se generaron, se encuentran en la Sección 3.5. Finalmente por técnicas usuales del método G-L y con ayuda de la desigualdad de tipo oráculo que satisface el estimador por Núcleo en la ventana seleccionada, se determina que el estimador es adaptativo. La Sección 3.4, *Estudio de Simulación*, se plantea el esquema de simulación para generar datos del modelo de regresión con función de regresión acotada, variable explicativa de densidad acotada, y débil dependientes de tipo α -mixing con decaimiento exponencial. Se explicita como obtener el estimador no paramétrico, con núcleo gaussiano, de la función de regresión usando el método GL y los datos simulados en el esquema de simulación planteado

previamente. Posteriormente se muestra cómo se calibra el método usando los mismos datos simulados. Finalmente, se presenta un estudio comparativo de resultados de estimaciones realizadas, usando errores globales y locales empíricos, para distintos tamaños de muestras simuladas y distintos valores de la desviación estándar del error del modelo de regresión. La Sección 3.5, *Demostraciones*, reúne las demostraciones de los lemas, proposiciones, teoremas y corolarios de este capítulo.

3.2. Procedimiento estadístico

Consideramos el estimador de la función de regresión dado por

$$\hat{r}_h(x) = \frac{1}{n} \sum_{k=1}^n Y_k K_h(x - X_k) g^{-1}(X_k)$$

donde $h > 0$ es el ancho de ventana, $K_h(\cdot) = \frac{1}{h} K(\frac{\cdot}{h})$ y $K : \mathbb{R} \rightarrow \mathbb{R}$, es una función de núcleo satisfaciendo $\int K(u) du = 1$, $\int u K(u) du = 0$, y la siguiente hipótesis.

(H₅) K tiene soporte $[-1, 1]$ y, $\|K\|_\infty < +\infty$.

Esta hipótesis implica que $\|K\|_1 < +\infty$ y $\|K\|_2 < +\infty$.

Para seleccionar la ventana h del estimador, vamos a usar el método de Goldenshluger-Lepski (GL). Consideramos la familia de anchos de ventanas

$$\mathcal{H} = \{e^{-i}\}_{i=0}^M \cap [h_{\min}, h_{\max}]$$

con $h_{\min} = \frac{(\log n)^8}{n}$, $h_{\max} = \frac{1}{(\log n)^2}$ y $M = \left\lceil \log \left(\frac{n}{(\log n)^8} \right) \right\rceil$. Se define para $h, h' \in \mathcal{H}$ un estimador auxiliar sobreesuavizado

$$\hat{r}_{h,h'}(x) = \frac{1}{n} \sum_{k=1}^n Y_k K_h * K_{h'}(X_k - x) g^{-1}(X_k).$$

Se define para $h \in \mathcal{H}$,

$$A(h, x) = \max_{h' \in \mathcal{H}} \{|\hat{r}_{h,h'}(x) - \hat{r}_h(x)| - V(h')\}_+$$

donde $V(h)$ esta dado por,

$$V(h) = \sqrt{2\gamma A_1} (\|K\|_1 + 1) (1 + \delta_n) \frac{(\log n)^{\frac{1}{2}}}{(nh)^{\frac{1}{2}}}$$

donde $\gamma > 2$, $A_1 = ((r_{\text{sup}})^2 + \sigma^2)(g_{\text{inf}})^{-1} \|K\|_2^2$ y $\delta_n = (\log n)^{-\frac{1}{5}}$.

El procedimiento GL consiste en seleccionar, basado en los datos, un ancho de ventana \hat{h} de la familia \mathcal{H} , dado por

$$\hat{h} = \arg \min_{h \in \mathcal{H}} \{A(h, x) + V(h)\}, \quad (3.2)$$

El estimador resultante, $\hat{r}_{\hat{h}}(x)$, satisface una desigualdad de oráculo que permite demostrar que es adaptativo.

3.3. Resultados

3.3.1. Resultados para el estimador $\hat{r}_{\hat{h}}$.

En esta sección se estudia el sesgo, varianza y velocidad de convergencia del estimador de regresión. Las demostraciones de los resultados se encuentran en la Sección 3.5. Para poder controlar el sesgo y demostrar que es asintóticamente insesgado se dan las definiciones de clase de Hölder y de kernel de orden m .

Definición 8 (Clase de Hölder). *Sea $\beta > 0$, $L > 0$. La clase de Hölder $\Sigma(\beta, L)$ se define como el conjunto de todas las funciones $f : \mathbb{R} \rightarrow \mathbb{R}$ tales que la derivada $f^{(l)}$, $l = \lfloor \beta \rfloor$ existe y*

$$\left| f^{(l)}(x) - f^{(l)}(y) \right| \leq L |x - y|^{\beta - l}, \forall x, y \in \mathbb{R}$$

donde $\lfloor \beta \rfloor = \max \{n \in \mathbb{N}, n < \beta\}$.

Definición 9 (Kernel de orden m). *Sea $m \in \mathbb{N}$. Se dice que $K : \mathbb{R} \rightarrow \mathbb{R}$ es un kernel de orden m si las funciones $u \rightarrow u^j K(u)$, $j = 0, 1, \dots, m$, satisfacen*

$$\int_{\mathbb{R}} K(u) du = 1, \int_{\mathbb{R}} u^j K(u) du = 0, j = 1, \dots, m.$$

A continuación, en las dos siguientes proposiciones se acota el sesgo y la varianza del estimador $\hat{r}_{\hat{h}}$.

Proposición 9. *Sean $\beta > 0$ y $L > 0$. Asumimos que K es de orden m , donde $m \geq l = \lfloor \beta \rfloor$, y que satisface $\int_{\mathbb{R}} |u|^\beta |K(u)| du < \infty$. Entonces, bajo la hipótesis (H_1) , si $r \in \Sigma(\beta, L)$, se tiene*

$$|E[\hat{r}_{\hat{h}}(x)] - r(x)| = |K_{\hat{h}} * r(x) - r(x)| \leq h^\beta A_0,$$

donde $A_0 = \frac{L}{\Gamma} \int_{\mathbb{R}} |u|^\beta |K(u)| du$. El estimador \hat{r}_h es asintóticamente insesgado, cuando h tiende a 0.

Proposición 10. *Bajo las hipótesis (H_1) , (H_2) , (H_3) , (H_4) y (H_5) se tiene que para $h \in (0, h_{\max})$*

$$\mathbb{E} [(\hat{r}_h(x) - \mathbb{E}[\hat{r}_h(x)])^2] \leq \frac{A_1}{nh} + A_2 \frac{(\log n)^{-\frac{1}{2}}}{nh},$$

donde $A_1 = ((r_{\sup})^2 + \sigma^2)(g_{\inf})^{-1} \|K\|_2^2$, $A_2 = 2A_3 + \frac{4}{|\log a|} A_4$, $A_3 = (r_{\sup})^2 \|K\|_1^2 ((g_{\inf})^{-2} Q + 1)$, y $A_4 = \frac{4}{1-a} (r_{\sup})^2 (g_{\inf})^{-2} \|K\|_\infty^2$.

Proposiciones 9 y 10 implican que cuando $n \rightarrow \infty$, $h \rightarrow 0$ y $nh \rightarrow \infty$, $MSE(\hat{r}_h, r, x) = E[(\hat{r}_h(x) - r(x))^2] \rightarrow 0$, o sea $\hat{r}_h(x)$ es un estimador consistente en media cuadrática (y por tanto en probabilidad) de $r(x)$.

Ahora se determinará la velocidad de convergencia del estimador de regresión $\hat{r}_h(x)$, lo cual se establece en el siguiente teorema.

Teorema 3. *Bajo las hipótesis de las Proposiciones 9 y 10, se tiene que, si $r \in \Sigma(\beta, L)$ entonces el estimador \hat{r}_{h_*} , con $h_* = n^{-\frac{1}{2\beta+1}}$ satisface*

$$E[(\hat{r}_{h_*}(x) - r(x))^2] \leq C_* n^{-\frac{2\beta}{2\beta+1}},$$

donde C_* es una constante positiva que depende del núcleo K , de β , L , r_{\sup} , Q , g_{\inf} , σ y a .

Demostración. Usando Proposiciones 9 y 10, se deduce que para $n \geq 4$,

$$MSE(\hat{r}_h, r, x) \leq h^{2\beta} A_0^2 + \frac{A_1 + A_2}{nh}. \quad (3.3)$$

Substituyendo h por h_* en la ecuación anterior se tiene que,

$$MSE(\hat{r}_{h_*}, r, x) \leq (A_0^2 + A_1 + A_2) n^{-\frac{2\beta}{2\beta+1}}$$

□

Note que cualquier ventana de la forma $An^{-\frac{1}{2\beta+1}}$ permite obtener la misma tasa de convergencia. En particular también la ventana que minimiza en h al término de la derecha en la desigualdad (3.3), ya que es de la forma $h = \left(\frac{A_1 + A_2}{2\beta A_0^2}\right)^{\frac{1}{2\beta+1}} n^{-\frac{1}{2\beta+1}}$.

Por el resultado anterior, se tiene

$$\sup_{r \in \Sigma(\beta, L)} E[(\hat{r}_{h_*}(x) - r(x))^2] \leq C_* n^{-\frac{2\beta}{2\beta+1}},$$

y

$$\inf_{\hat{r} \in \Theta} \sup_{r \in \Sigma(\beta, L), \mathbb{X} \in \mathcal{L}} E \left[(\hat{r}(x) - r(x))^2 \right] \leq C_* n^{-\frac{2\beta}{2\beta+1}},$$

donde Θ es el conjunto de todos los estimadores de r , y recordamos que \mathcal{L} es el conjunto de todos los procesos \mathbb{X} que satisfacen (H_1) , (H_2) y (H_3) .

En Tsybakov (2009) (ver Sección 2.5), se demuestra

$$\inf_{\hat{r} \in \Theta} \sup_{r \in \Sigma(\beta, L), \mathbb{X} \in \mathcal{L}} E \left[(\hat{r}(x) - r(x))^2 \right] \geq cn^{-\frac{2\beta}{2\beta+1}},$$

donde c es una constante que depende de β , L , g_{sup} , y σ^2 .

Esto permite concluir que el estimador \hat{r}_{h^*} converge a la velocidad óptima sobre la clase de Hölder $\Sigma(\beta, L)$ la cual está dada por

$$v_n = n^{-\frac{\beta}{2\beta+1}}. \quad (3.4)$$

3.3.2. Desigualdad de Oráculo y adaptatividad del estimador

En la siguiente proposición, obtenemos una primera desigualdad que satisface el estimador $\hat{r}_{\hat{h}}(x)$ usando principalmente la definición de \hat{h} dada en (3.2).

Proposición 11. *Bajo la hipótesis (H_5) , el estimador $\hat{r}_{\hat{h}}(x)$ satisface para todo $h \in \mathcal{H}$,*

$$\begin{aligned} & \left(\mathbb{E} (r(x) - \hat{r}_{\hat{h}}(x))^2 \right)^{\frac{1}{2}} \\ & \leq \left(\mathbb{E} (r(x) - \hat{r}_h(x))^2 \right)^{\frac{1}{2}} + 2 \left(\mathbb{E}(T_1^2) \right)^{\frac{1}{2}} + 2 \left(\mathbb{E}(T_2^2) \right)^{\frac{1}{2}} + 2C(h) \|K\|_1 + 2V(h) \end{aligned}$$

donde

$$T_1 = \max_{h' \in \mathcal{H}} \{ |\hat{r}_{h'}(x) - \mathbb{E}(\hat{r}_{h'}(x))| - V_1(h') \}_+, \quad T_2 = \max_{h' \in \mathcal{H}} \{ |\hat{r}_{h,h'}(x) - \mathbb{E}(\hat{r}_{h,h'}(x))| - V_2(h') \}_+,$$

$$C(h) = \max_{u \in B(x)} |K_h * r(u) - r(u)|$$

y $V_1(h')$, $V_2(h')$ tales que $V(h') = V_1(h') + V_2(h')$, con

$$V_1(h') = \sqrt{2\gamma A_1 \frac{\log n}{nh'}} (1 + \delta_n), \quad V_2(h') = \|K\|_1 V_1(h').$$

En la proposición anterior se proporciona una cota de la raíz cuadrada del error cuadrático medio puntual del estimador $\hat{r}_{\hat{h}}(\cdot)$, en términos de la varianza y el sesgo del estimador

$\hat{r}_h(\cdot)$, además de $V_1(h')$, $V_2(h')$, $\mathbb{E}(T_1^2)$ y $\mathbb{E}(T_2^2)$. En la demostración de la proposición es fundamental el hecho de que $V(h') = V_1(h') + V_2(h')$. La elección precisa que se tomo de las expresiones de $V_1(h')$ y $V_2(h)$ es fundamental para la demostración de la Proposición 12 y permite establecer que $\mathbb{E}(T_1^2)$ y $\mathbb{E}(T_2^2)$ sean despreciables respecto a los términos de sesgo y varianza de cualquier estimador \hat{r}_h para $h \in \mathcal{H}$.

Para controlar los términos $\mathbb{E}(T_1^2)$ y $\mathbb{E}(T_2^2)$, se usa la desigualdad de Bernstein enunciada en la proposición 3.

Proposición 12. *Bajo las hipótesis (H_1) , (H_2) , (H_3) , (H_4) y (H_5) , se tiene para n suficientemente grande*

$$(i) \quad \mathbb{E}(T_1^2) \leq A_5 \frac{\log n}{n},$$

$$(ii) \quad \mathbb{E}(T_2^2) \leq A_5 \|K\|_1^2 \frac{\log n}{n},$$

donde A_5 es una constante explicita (ver Sección 3.5.1) que depende del núcleo K , de r_{sup} , Q , g_{inf} , σ , a y γ .

Teorema 4. (Desigualdad de Oráculo) *Bajo las hipótesis (H_1) , (H_2) , (H_3) , (H_4) y (H_5) , el estimador $\hat{r}_h(x)$ satisface la siguiente desigualdad,*

$$\left(\mathbb{E} \left(r(x) - \hat{r}_h(x) \right)^2 \right)^{\frac{1}{2}} \leq \min_{h \in \mathcal{H}} \left(A_6 C(h) + A_7 (1 + \delta_n) \frac{(\log n)^{\frac{1}{2}}}{(nh)^{\frac{1}{2}}} \right) + A_8 \frac{(\log n)^{\frac{1}{2}}}{n^{\frac{1}{2}}} \quad (3.5)$$

donde $A_6 = 1 + 2\|K\|_1$, $A_7 = \sqrt{A_1 + A_2} + 2\sqrt{2\gamma A_1}(\|K\|_1 + 1)$ y A_8 es una constante positiva que depende del núcleo K , de r_{sup} , Q , g_{inf} , σ , a y γ .

Observación 6. *En el Teorema 4, podemos ver en (3.5) que el estimador \hat{r}_h imita al oráculo, es decir, al mejor estimador posible (pero desconocido) \hat{r}_h en la familia $(\hat{r}_h)_{h \in \mathcal{H}}$, que es el que minimiza la suma del sesgo $C(h)$ y de la desviación estándar. Cuando la familia $(\hat{r}_h)_{h \in \mathcal{H}}$ es lo suficientemente amplia, los estimadores que satisfacen desigualdades de oráculo suelen ser estimadores adaptativos. Es lo que se ve a continuación en el Teorema 5.*

Demostración. Sea $h \in \mathcal{H}$. Usando que $|\mathbb{E}(\hat{r}_h(x)) - r(x)| = |K_h * r(x) - r(x)| \leq C(h)$ y la Proposición 10 se tiene que,

$$\begin{aligned} \left(\mathbb{E} \left[(r(x) - \hat{r}_h(x))^2 \right] \right)^{\frac{1}{2}} &\leq C(h) + \left(\frac{A_1}{nh} + A_2 \frac{(\log n)^{-\frac{1}{2}}}{nh} \right)^{\frac{1}{2}} \\ &\leq C(h) + (A_1 + A_2)^{1/2} (nh)^{-1/2} \end{aligned} \quad (3.6)$$

Ahora por la Proposición 11 y la desigualdad (3.6), se tiene que,

$$\begin{aligned} & \left(\mathbb{E} (r(x) - \hat{r}_h(x))^2 \right)^{\frac{1}{2}} \\ & \leq C(h) + (A_1 + A_2)^{1/2} (nh)^{-1/2} + 2C(h) \|K\|_1 + 2V(h) + 2 \left(\mathbb{E}(T_1^2) \right)^{\frac{1}{2}} + 2 \left(\mathbb{E}(T_2^2) \right)^{\frac{1}{2}} \\ & \leq A_6 C(h) + A_7 (1 + \delta_n) \frac{(\log n)^{\frac{1}{2}}}{(nh)^{\frac{1}{2}}} + 2 \left(\mathbb{E}(T_1^2) \right)^{\frac{1}{2}} + 2 \left(\mathbb{E}(T_2^2) \right)^{\frac{1}{2}}. \end{aligned}$$

La Proposición 12 cuyo resultado es para n suficientemente grande permite obtener que

$$2 \left(\mathbb{E}(T_1^2) \right)^{\frac{1}{2}} + 2 \left(\mathbb{E}(T_2^2) \right)^{\frac{1}{2}} \leq A_8 \frac{(\log n)^{\frac{1}{2}}}{n^{\frac{1}{2}}}$$

para alguna constante A_8 que depende del núcleo K , de r_{sup} , Q , g_{inf} , σ , a y γ , lo que permite concluir el resultado. \square

Teorema 5. (Adaptatividad) *Asumimos que las hipótesis (H_1) , (H_2) , (H_3) , (H_4) y (H_5) están satisfechas. Asumimos que K es de orden m . Entonces para todo $0 < \beta \leq m$ y $L > 0$ el estimador \hat{r}_h satisface para $r \in \Sigma(\beta, L)$.*

$$\left(\mathbb{E} (r(x) - \hat{r}_h(x))^2 \right)^{\frac{1}{2}} \leq C^* \left(\frac{n}{\log n} \right)^{-\frac{\beta}{2\beta+1}}$$

donde C^* es una constante positiva que depende del núcleo K , de β , L , r_{sup} , Q , g_{inf} , g_{sup} , σ^2 , a y γ .

Observación 7. *El estimador \hat{r}_h converge a la velocidad $\left(\frac{n}{\log n} \right)^{-\frac{\beta}{2\beta+1}}$ sobre las clases de Hölder de regularidad β y el estimador no depende de β . Esta velocidad es casi la velocidad óptima o minimax salvo por el término multiplicativo logarítmico. Este término logarítmico extra es clásico en el caso adaptativo y aparece también por ejemplo en Lepski and Serdyukova (2014). Este resultado de velocidad de convergencia en riesgo puntual en caso adaptativo es nuevo en contexto de datos dependientes y generaliza el resultado de Lepski and Serdyukova (2014) obtenido para datos independientes y densidad de variable explicativa conocida.*

Demostración. Sea $h_i = e^{-i}$ para cada $i \in \{0, 1, 2, \dots, M\}$ y $h_\beta = \left(\frac{\log n}{n} \right)^{\frac{1}{2\beta+1}}$. Para n suficientemente grande, existe $i \in \{0, 1, 2, \dots, M\}$ tal que,

$$h_{i+1} \leq h_\beta \leq h_i \tag{3.7}$$

donde $h_{i+1} = e^{-i-1} = \frac{e^{-i}}{e} = \frac{h_i}{e}$.

Se procede a acotar el primer término dentro del mínimo en la desigualdad (3.5). Por la Proposición 9 se tiene que el sesgo del estimador en la ventana h_{i+1} satisface para $u \in B(x)$,

$$|K_{h_{i+1}} * r(u) - r(u)| \leq A_0 h_{i+1}^\beta$$

Por lo tanto usando la desigualdad (3.7), se tiene

$$C(h_{i+1}) = \max_{u \in B(x)} |K_{h_{i+1}} * r(u) - r(u)| \leq A_0 h_{i+1}^\beta \leq A_0 h_\beta^\beta = A_0 \left(\frac{n}{\log n} \right)^{-\frac{\beta}{2\beta+1}}. \quad (3.8)$$

Ahora se procede a determinar una cota para el segundo término dentro del mínimo en la desigualdad (3.5). De la desigualdad (3.7) se deduce que $\frac{\log n}{nh_i} \leq \frac{\log n}{nh_\beta}$ y además como $h_{i+1} = \frac{h_i}{e}$ se establece que,

$$\frac{\log n}{nh_{i+1}} = \frac{e \log n}{nh_i} \leq \frac{e \log n}{nh_\beta}$$

que implica usando el valor de h_β

$$\left(\frac{\log n}{nh_{i+1}} \right)^{\frac{1}{2}} \leq \frac{e^{\frac{1}{2}} (\log n)^{\frac{1}{2}}}{n^{\frac{1}{2}}} \frac{1}{h_\beta^{\frac{1}{2}}} = e^{\frac{1}{2}} \left(\frac{\log n}{n} \right)^{\frac{\beta}{2\beta+1}} \quad (3.9)$$

Por las desigualdades (3.8) y (3.9) se tiene que para $h_{i+1} \in \mathcal{H}$,

$$A_6 C(h_{i+1}) + A_7 (1 + \delta_n) \frac{(\log(n))^{\frac{1}{2}}}{(nh_{i+1})^{\frac{1}{2}}} \leq \left(A_6 A_0 + A_7 (1 + \delta_n) e^{\frac{1}{2}} \right) \left(\frac{n}{\log n} \right)^{-\frac{\beta}{2\beta+1}}$$

Por esto último y por la desigualdad (3.5) se tiene que,

$$\left(\mathbb{E} (r(x) - \hat{r}_{\hat{h}}(x))^2 \right)^{\frac{1}{2}} \leq \left(A_6 A_0 + A_7 e^{\frac{1}{2}} (1 + \delta_n) \right) \left(\frac{n}{\log n} \right)^{-\frac{\beta}{2\beta+1}} + A_8 \frac{(\log n)^{\frac{1}{2}}}{n^{\frac{1}{2}}} \quad (3.10)$$

Usando que $\left(\frac{\log n}{n} \right)^{\frac{1}{2}} \leq \left(\frac{\log n}{n} \right)^{\frac{\beta}{2\beta+1}}$, se deduce que

$$\left(\mathbb{E} (r(x) - \hat{r}_{\hat{h}}(x))^2 \right)^{\frac{1}{2}} \leq C^* \left(\frac{n}{\log n} \right)^{-\frac{\beta}{2\beta+1}}$$

donde C^* depende de A_0, A_6, A_7 y A_8 , para $n \geq 3$. \square

3.4. Estudio de simulación

En esta sección se muestra el ajuste del procedimiento planteado para datos simulados. Se comparan los errores empíricos globales y locales, para distintos tamaños de la muestra. Además, se esquematiza el procedimiento utilizado para calibrar el método. El objetivo es estimar la función de regresión para un proceso débilmente dependiente, con función de densidad conocida.

3.4.1. Esquema de simulación

Se plantea el modelo de regresión $Y_i = r(X_i) + \varepsilon_i$ para $i = 1, \dots, n$, donde las variables ε_i son independientes e idénticamente distribuidas, con distribución común $N(0, \sigma^2)$ y $\sigma = 0,1, 0,5$ y 1 . La función de regresión a estimar es $r(x) = 0,7x + 2e^{-10x^2}$ que restringida al intervalo $[-2, 2]$ satisface la hipótesis (H_4).

Se genera una muestra $\{X_i\}_{i=1}^{n+2q}$ del proceso \mathbb{X} presentado en el Ejemplo 2, donde $X_i = \left(G^{-1} \circ \Phi_{0, \frac{\rho^2}{1-\phi^2}} \right) (Z_i)$ con $\rho = 1$, $\phi = 0,75$ y $c = 2$. Los X_i tienen distribución normal truncada, específicamente su función de densidad es $g(x) = \frac{\phi_{0,1}(x)}{p} \mathbf{1}_{[-2,2]}(x)$, con $p = \Phi_{0,1}(2) - \Phi_{0,1}(-2)$, para cada $x \in \mathbb{R}$. Como mencionamos en el Ejemplo 2, el proceso cumple que $\mathbb{X} \in \mathcal{L}$, es decir satisface las hipótesis (H_1), (H_2) y (H_3).

La Figura 3.1 muestra el gráfico de dispersión de las observaciones $\{X_i\}_{i=1}^{n+2q}$ para $n = 2000$ y $q = 100$ (superior) y la función de autocorrelación muestral $\gamma_X(k) = \text{corr}(X_i, X_{i+k})$, para $k = 1, \dots, 100$ (inferior). Se puede apreciar el decaimiento exponencial de las correlaciones, característica presente en los procesos débilmente dependientes de tipo α -mixing.

Las primeras n observaciones $\{(X_i, Y_i)\}_{i=1}^n$ conforman la muestra usada en la sección 3.4.2 para estimar aplicando el método GL, y las últimas q observaciones $\{(X_i, Y_i)\}_{i=n+q+1}^{n+2q}$ es la muestra usada para calibrar el método seleccionando un valor de γ adecuado como se describe en la sección 3.4.3. La idea de dejar un salto temporal de tamaño q entre la muestra de estimación y la muestra de calibración es para asegurar que el coeficiente de correlación $\alpha_m(\mathbb{X})$ -mixing sea suficientemente pequeño y así disminuir el efecto de sobreajuste producido por la dependencia entre las muestras, el control del valor del coeficiente de correlación $\alpha_m(\mathbb{X})$ es realizado desde la hipótesis de decrecimiento exponencial (H_3).

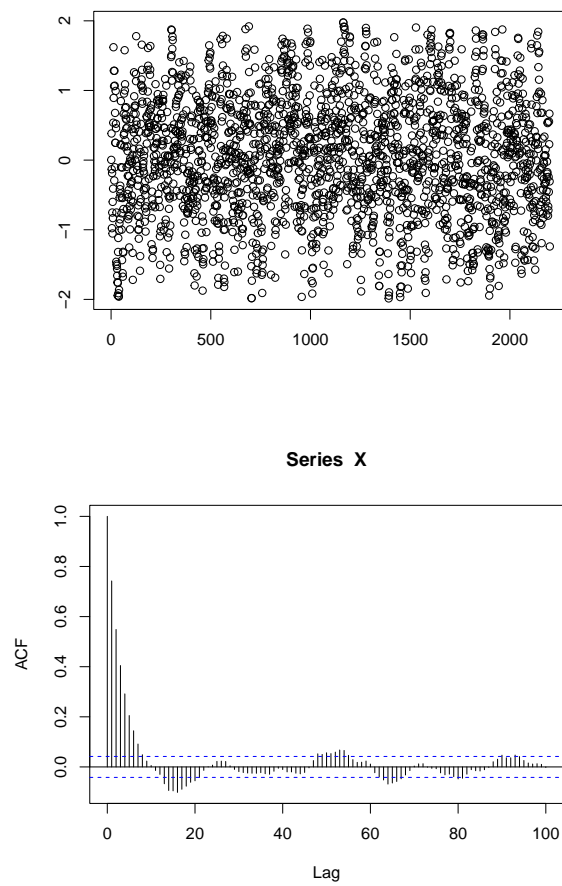


Figura 3.1 Gráfico de dispersión de \mathbb{X} y función de autocorrelación muestral $\gamma_{\mathbb{X}}$.

3.4.2. Estimación GL

La función de regresión $r(\cdot)$ se estima en la malla igualmente espaciada $\{x_i\}_{i=1}^s$ del intervalo $[-1, 1]$, donde $x_i = -1 + 2\left(\frac{i-1}{s-1}\right)$ para $i = 1, \dots, s$, y consideramos $s = 21$. Para aplicar el método GL se considera la muestra aleatoria $\{(X_i, Y_i)\}_{i=1}^n$ como se indicó anteriormente y se determina el estimador $\{\hat{r}_{\hat{h}_i}(x_i)\}_{i=1}^s$ donde,

$$\hat{r}_{\hat{h}_i}(x_i) = \frac{1}{n} \sum_{k=1}^n Y_k K_{\hat{h}_i}(x_i - X_k) g^{-1}(X_k),$$

se usa como Kernel $K(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \forall x \in \mathbb{R}$, que aunque no es de soporte compacto tiene buenas propiedades practicas, luego se toma $K_{\hat{h}_i}(\cdot) = \frac{1}{\hat{h}_i} K\left(\frac{\cdot}{\hat{h}_i}\right)$ donde cada ventana \hat{h}_i se selecciona de la familia de ventanas $\mathcal{H} = \left\{ e^{-0,1j} : j = 0, \dots, \frac{[\log(n)]^{\frac{2}{3}}}{0,1} \right\}$, tomando

$$\hat{h}_i = \arg \min_{h \in \mathcal{H}} \{A(h, x_i) + V(h, x_i)\} \quad (3.11)$$

para cada $i = 1, \dots, s$, con $A(h, x_i) = \max_{h' \in \mathcal{H}} \{|\hat{r}_{h, h'}(x_i) - \hat{r}_{h'}(x_i)| - V(h', x_i)\}_+$ y $V(h, x_i) = \sqrt{2\gamma A_1(i)} (\|K\|_1 + 1) (1 + \delta_n) \frac{(\log n)^{\frac{1}{2}}}{(nh)^{\frac{1}{2}}}$, donde $\delta_n = (\log n)^{\frac{1}{5}}$, $A_1(i) = (g_{\inf}(i))^{-1} \|K\|_2^2 ((\hat{r}_{\sup}(i))^2 + \hat{\sigma}^2)$, $\|K\|_1 = 1$, $\|K\|_2 = \frac{1}{\sqrt{2\sqrt{\pi}}}$, $g_{\inf}(i) = \min \{w > 0 | w = g(X_j) \mathbf{1}_{|x_i - 0,5, x_i + 0,5|}(X_j), \text{ para } j = 1, \dots, n\}$, $\hat{r}_{\sup}(i) = \max \{|Y_j| \mathbf{1}_{|x_i - 0,5, x_i + 0,5|}(X_j) | j = 1, \dots, n\}$ y $\hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^n (Y_i - \tilde{r}_{\bar{h}}(X_i))^2$ donde $\tilde{r}_{\bar{h}}(\cdot)$ es el estimador de Nadaraya-Watson, en la muestra $\{(X_i, Y_i)\}_{i=1}^n$ y la ventana \bar{h} , con \bar{h} la ventana resultante del comando $dpill(\cdot, \cdot)$ en R correspondiente a la librería *KernSmooth*, método plug-in de Ruppert, Sheather y Wand (1995).

El parámetro γ es un parámetro de calibración del método, cuyo valor se determina según la metodología planteada en la Sección 3.4.3.

En la Figura 3.2 se muestra la función de regresión r y tres estimadores GL de la función de regresión para muestras aleatorias $\{(Y_k, X_k)\}_{k=1}^{n+2q}$, con $n = 1000, 2000$ y 5000 tomando $q = 100$ y $\sigma = 0,5$.

3.4.3. Calibración del método para una muestra.

Como la estimación de la función de regresión r se realiza en cada punto de la malla $\{x_j\}_{j=1}^s$ del intervalo $[-1, 1]$, y para cada x_j se selecciona una ventana \hat{h}_j de la familia de ventanas \mathcal{H} , entonces se tiene un vector de ventanas $\hat{h} = (\hat{h}_1, \dots, \hat{h}_s)$. En general el estimador

Estimación Adaptativa Método GL para r

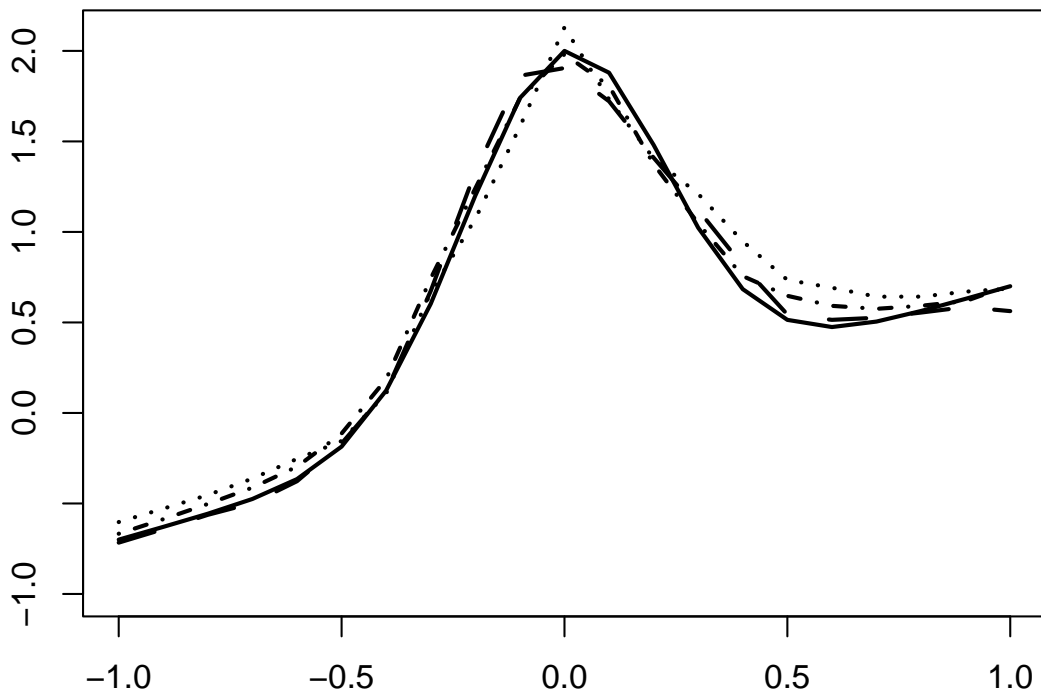


Figura 3.2 Se muestran tres estimadores GL de la función de regresión r , para muestras aleatorias $\{(Y_k, X_k)\}_{k=1}^{n+2q}$, con $\sigma = 0,5$ y $q = 100$. Cuando $n = 1000$ el estimador GL se representa con puntos, para $n = 2000$ con rayas largas y para $n = 5000$ con puntos y rayas cortas. Además se representa con una línea continua la función de regresión r .

GL de r sobre la malla $\{x_i\}_{i=1}^s$ en el vector de ventanas óptimas $\hat{h} = (\hat{h}_1, \dots, \hat{h}_s)$, se denota por $(\hat{r}_{\hat{h}_1}(x_1), \dots, \hat{r}_{\hat{h}_s}(x_s))$, donde cada ventana óptima se obtiene según la ecuación (3.11) con $V(h, x_i)$ dependiente del parámetro $\gamma > 2$. En la practica para calibrar el método se toma γ en una malla igualmente espaciada de un subintervalo I de $]0, 1/10[$. Para cada γ se tiene que el vector de ventanas óptimas depende de γ , tal dependencia se denota de la siguiente manera $\hat{h}^\gamma = (\hat{h}_1^\gamma, \dots, \hat{h}_s^\gamma)$. La elección del intervalo I se toma de forma tal que la curva descrita por los estimadores $\hat{r}_{\hat{h}_i^\gamma}(x_i)$, para $i = 1, \dots, s$, pasen de ser curvas muy irregulares a curvas suaves.

Al calibrar el método, para una muestra $\{(X_i, Y_i)\}_{i=1}^{n+2q}$, se toman los primeros n datos de la muestra para construir el estimador $\hat{r}_h(x)$. De los últimos q datos de la muestra $\{(X_i, Y_i)\}_{i=n+q+1}^{n+2q}$ se construye una malla aleatoria del intervalo $[-1, 1]$ seleccionando aquellos valores que cumplen $-1 \leq X_i \leq 1$ y se ordenan los datos en forma creciente respecto a la primera coordenada, obteniendo así una muestra denotada con letras minúsculas $\{(\tilde{x}_i, \tilde{y}_i)\}_{i=1}^p$ con $0 < p \leq 100$. La malla $\{\tilde{x}_i\}_{i=1}^p$ tiene asociado el vector de ventanas óptimas $\hat{h}^\gamma = (\hat{h}_1^\gamma, \dots, \hat{h}_p^\gamma)$, para cada $\gamma \in I$, de esta manera se tienen los estimadores GL $\hat{r}_{\hat{h}_i^\gamma}(\tilde{x}_i)$ para cada $\gamma \in I$, sobre la malla $\{\tilde{x}_i\}_{i=1}^p$ en el vector de ventanas óptimas \hat{h}^γ .

Se determina $Error(\gamma) = \sum_{i=1}^p d_i \left(\hat{r}_{\hat{h}_i^\gamma}(\tilde{x}_i) - \tilde{y}_i \right)^2$ para cada $\gamma \in I$, donde $d_1 = \frac{\tilde{x}_1 + \tilde{x}_2}{2} + 1$, $d_s = 1 - \frac{\tilde{x}_{s-1} + \tilde{x}_s}{2}$ y $d_i = \frac{\tilde{x}_{i+1} - \tilde{x}_{i-1}}{2}$, para $i \in \{2, \dots, p-1\}$. Finalmente, el método se calibra seleccionando el valor γ en I que minimiza $Error(\gamma)$.

En la practica se tomó $I =]0,00000005, 0,05[$, además, sobre tal intervalo se considera la malla igualmente espaciada $\gamma_i = 0,00000005 + (i-1) \frac{0,05 - 0,00000005}{20}$ para $i = 1, \dots, 21$.

La Figura 3.3, en el gráfico de arriba se muestra que $Error(\gamma)$ se minimiza en γ_3 . Además, abajo se muestra la nube de puntos de la muestra de estimación, se representa con una linea continua la función de regresión r y con puntos el estimador $\hat{r}_{\hat{h}_i}(x_i)$ calibrado en γ_3 , y evaluado en la malla determinística del intervalo $[-1, 1]$, $x_i = -1 + 2 \left(\frac{i-1}{s-1} \right)$ para $i = 1, \dots, s$, con $s = 21$. En ambos se tomo $n = 2000$, $q = 100$ y $\sigma = 0,5$.

3.4.4. Comparación de errores globales y locales empíricos, para distintos tamaños de muestra y valores de σ .

Para mostrar la calidad del estimador de la función de regresión r , se generan $N = 500$ replicas de la muestra $\{(X_i, Y_i)\}_{i=1}^{n+2q}$. Posteriormente se calculan N estimadores GL ya calibrados $\left\{ \hat{r}^{(1)}(x_i) \right\}_{i=1}^s, \dots, \left\{ \hat{r}^{(N)}(x_i) \right\}_{i=1}^s$, con $x_i = -1 + 2 \frac{i-1}{s-1}$, $s = 21$ y $\hat{r}^{(j)}(x_i) = \hat{r}_{\hat{h}_i}(x_i)$

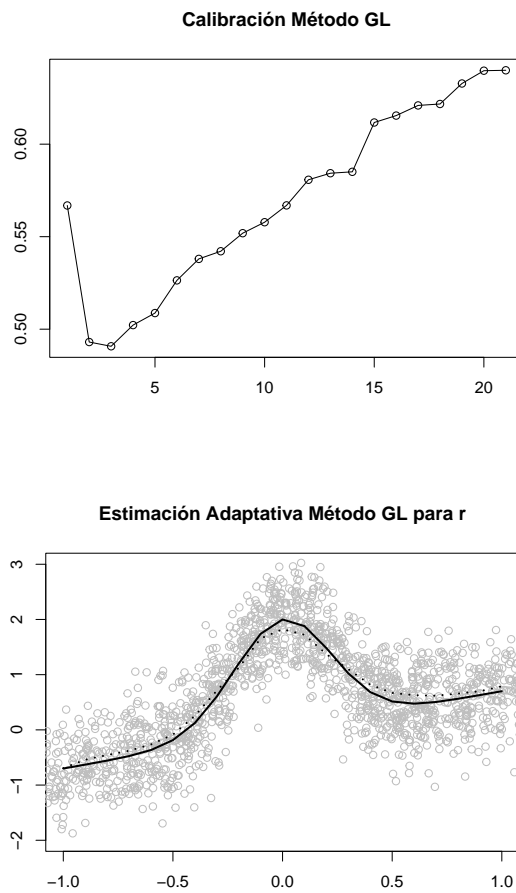


Figura 3.3 Arriba gráfico de $\{Error(\gamma_i)\}_{i=1}^{21}$ y abajo se muestra la nube de puntos $\{(X_i, Y_i)\}_{i=1}^n$, se representa con una línea continua la función de regresión r y con puntos el gráfico de $\{\hat{r}_{h_i}(x_i)\}_{i=1}^{21}$ en el intervalo $[-1, 1]$, calibrado en γ_3 , que es donde el gráfico de la izquierda alcanza su mínimo. En ambos se tomo $n = 2000$, $q = 100$ y $\sigma = 0,5$.

donde \hat{h}_i se obtuvo para la j -ésima replica por el método GL. El error cuadrático medio (MSE) y el error cuadrático medio integrado ($MISE$) se estiman de la siguiente forma

$$\hat{MSE}(x_i) = \frac{1}{N} \sum_{j=1}^N \left[\left(r(x_i) - \hat{r}^{(j)}(x_i) \right)^2 \right]$$

para $i = 1, \dots, s$.

$$\hat{MISE} = \frac{1}{N} \sum_{j=1}^N I_j$$

con $I_j = \sum_{i=1}^s d_i \left(r(x_i) - \hat{r}^{(j)}(x_i) \right)^2$ donde $d_1 = d_s = 0,05$ y $d_i = 0,1$ para $i = 2, \dots, s-1$.

En la Figura 3.4 se representan tres boxplot de I_j para $j = 1, \dots, N$, con los tamaños de muestra $n = 1000, 2000$ y 5000 de izquierda a derecha respectivamente, además sobre cada boxplot hay un punto que representa el \hat{MISE} (superior) y tres representaciones del $\hat{MSE}(x_i)$ para $i = 1, \dots, s$ con una línea para el tamaño de muestra $n = 1000$, con puntos para $n = 2000$ y con rayas para $n = 5000$, (inferior). En los tres casos para $\sigma = 0,1$. En las Figuras 3.5 y 3.6 se hacen las mismas representaciones pero con $\sigma = 0,5$ y $\sigma = 1$ respectivamente.

Al analizar el gráfico que corresponde al \hat{MISE} con $\sigma = 0,1$ en la Figura 3.4 (gráficos de abajo) se observa que los valores del \hat{MISE} son mas elevados para valores del dominio de r cercanos a donde se alcanza el máximo y donde hay cambio de concavidad, ese patrón se observa con tamaños de muestra $n = 1000, 2000$ y 5000 . Como es evidente a medida que aumenta el tamaño de la muestra los valores del \hat{MISE} disminuyen. Comportamientos similares se observan al analizar el \hat{MISE} para $\sigma = 0,5$ y $\sigma = 1$ de las Figuras 3.5 y 3.6 respectivamente (gráficos de abajo). Como es evidente cuando aumenta el valor de σ también aumentan los valores del \hat{MISE} .

En los diagramas de caja de los valores de I_j para $j = 1, \dots, N$ con $\sigma = 0,1$ correspondientes a la Figura 3.4 (gráficos de arriba) se observa que a medida que los valores de n aumentan de 1000, 2000 a 5000 la dispersión de los $\{I_j\}_{j=1}^N$ disminuye, también disminuye la cantidad de datos atípicos y el valor del \hat{MISE} se aproxima cada vez mas al valor de la mediana de los $\{I_j\}_{j=1}^N$, aunque en general la mediana siempre es inferior al \hat{MISE} . En los diagrama de caja de las Figuras 3.5 y 3.6 correspondientes a $\sigma = 0,5$ y $\sigma = 1$ respectivamente se observa que a medida que aumentan los valores de σ aumenta la dispersión de los $\{I_j\}_{j=1}^N$, también aumenta la distancia entre el \hat{MISE} y la mediana de los $\{I_j\}_{j=1}^N$ y que en general la mediana siempre es inferior al valor del \hat{MISE} .

Como se observa en la tabla 3.1 el valor del \hat{MISE} es inversamente proporcional a n y directamente proporcional a σ .

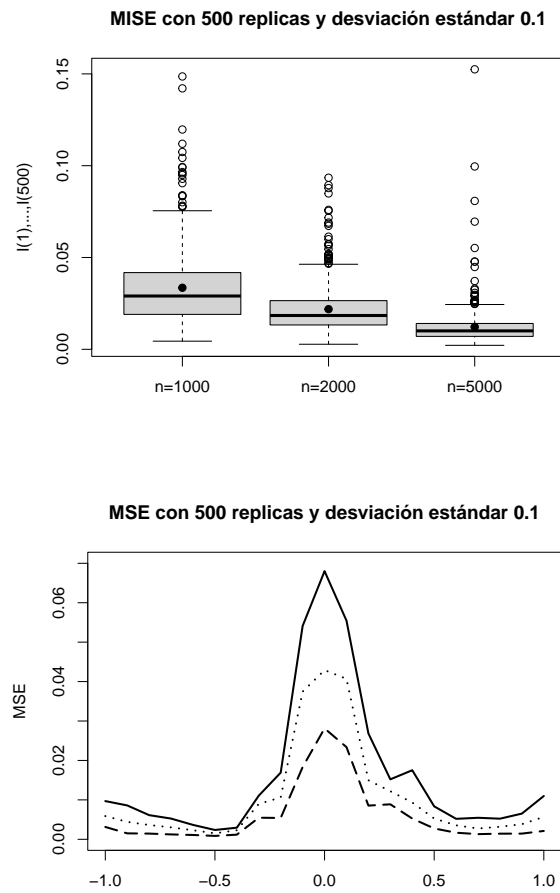


Figura 3.4 Arriba se representan tres boxplot de I_j para $j = 1, \dots, N$, con los tamaños de muestra $n = 1000, 2000$ y 5000 de izquierda a derecha respectivamente, sobre cada boxplot hay un punto que representa el $M\hat{I}SE$ y abajo se muestran tres representaciones del $M\hat{S}E(x_i)$ para $i = 1, \dots, s$ con una línea para el tamaño de muestra $n = 1000$, con puntos para $n = 2000$ y con rayas para $n = 5000$. En ambos casos para $\sigma = 0,1$.

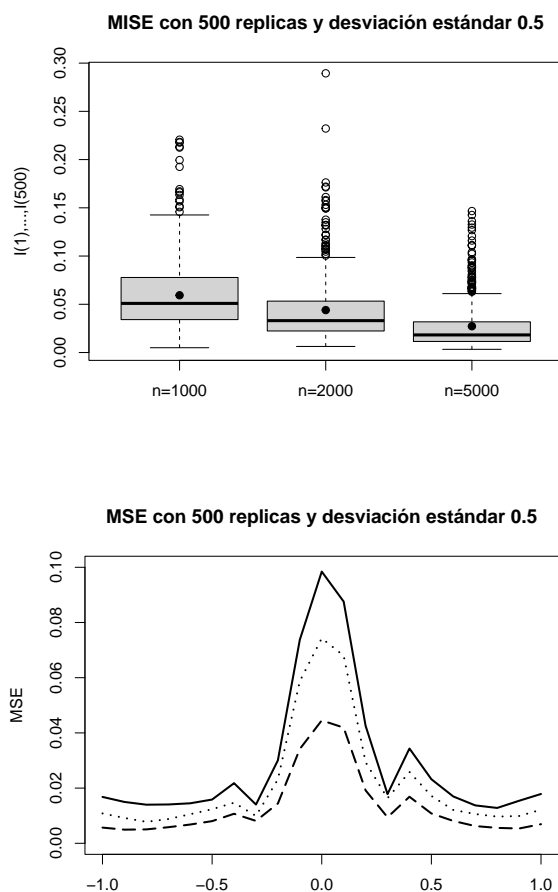


Figura 3.5 Arriba se representan tres boxplot de I_j para $j = 1, \dots, N$, con los tamaños de muestra $n = 1000, 2000$ y 5000 de izquierda a derecha respectivamente, sobre cada boxplot hay un punto que representa el \hat{MISE} y abajo se muestran tres representaciones del $\hat{MSE}(x_i)$ para $i = 1, \dots, s$ con una línea para el tamaño de muestra $n = 1000$, con puntos para $n = 2000$ y con rayas para $n = 5000$. En ambos casos para $\sigma = 0,5$.

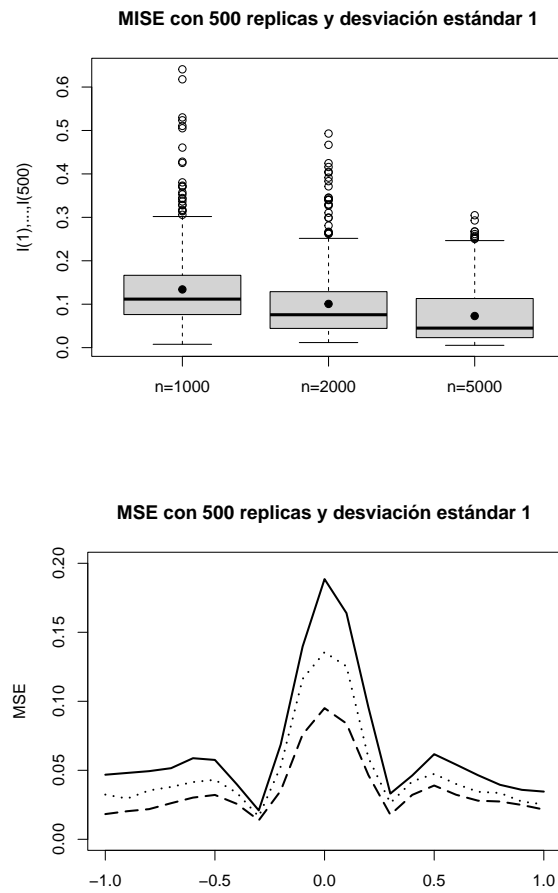


Figura 3.6 Arriba se representan tres boxplot de I_j para $j = 1, \dots, N$, con los tamaños de muestra $n = 1000, 2000$ y 5000 de izquierda a derecha respectivamente, sobre cada boxplot hay un punto que representa el $M\hat{I}SE$ y abajo se muestran tres representaciones del $M\hat{S}E(x_i)$ para $i = 1, \dots, s$ con una línea para el tamaño de muestra $n = 1000$, con puntos para $n = 2000$ y con rayas para $n = 5000$. En ambos casos para $\sigma = 1$.

| | $n = 1000$ | $n = 2000$ | $n = 5000$ |
|----------------|------------|------------|------------|
| $\sigma = 0,1$ | 0.03351511 | 0.02186702 | 0.01219458 |
| $\sigma = 0,5$ | 0.05932621 | 0.04399132 | 0.02723891 |
| $\sigma = 1$ | 0.13400681 | 0.10064675 | 0.07281427 |

Cuadro 3.1 Valores del $M\hat{I}SE$ para distintos valores de σ y tamaños de la muestra, caso g conocido.

3.5. Demostraciones

3.5.1. Constantes.

Recordamos algunos valores de constantes dados previamente y además a continuación van constantes que usaremos en el desarrollo de las demostraciones.

$$\begin{aligned}
 A_0 &= \frac{L}{l!} \int_{\mathbb{R}} |u|^\beta |K(u)| du, \quad A_1 = (g_{\text{inf}})^{-1} \|K\|_2^2 ((r_{\text{sup}})^2 + \sigma^2) = B_1 + B_2, \\
 A_2 &= \frac{2A_3}{|\log a|} + 20A_4, \quad A_3 = (r_{\text{sup}})^2 \|K\|_1^2 ((g_{\text{inf}})^{-2} Q + 1), \\
 A_4 &= \frac{4}{1-a} (r_{\text{sup}})^2 (g_{\text{inf}})^{-2} \|K\|_\infty^2, \quad A_5 = B_9 + 2B_{10}, \\
 A_6 &= 1 + 2\|K\|_1, \quad A_7 = \sqrt{A_1 + A_2} + 2\sqrt{2\gamma A_1} (\|K\|_1 + 1), \quad A_8 = 2(1 + \|K\|_1^2) A_5, \\
 B &= \frac{12(B_6 \vee \sqrt{B_7})}{1 \wedge |\log a|} \left(\frac{2^7 3 B_7 a^{-\frac{1}{3}}}{A_1 (1 \wedge |\log a|)} \vee 1 \right), \\
 B_1 &= \sigma^2 (g_{\text{inf}})^{-1} \|K\|_2^2, \quad B_2 = (r_{\text{sup}})^2 (g_{\text{inf}})^{-1} \|K\|_2^2, \quad B_3 = \left(r_{\text{sup}} + \frac{2\sigma}{\sqrt{2\pi}} \right) \|K\|_1, \\
 B_4 &= \left(r_{\text{sup}} + \frac{2\sigma}{\sqrt{2\pi}} \right)^2 (g_{\text{inf}})^{-2} Q \|K\|_1^2, \quad B_5 = B_4 + B_3^2, \quad B_6 = 2\|K\|_\infty (g_{\text{inf}})^{-1}, \\
 B_7 &= (4B_5 B_6 (\sigma + r_{\text{sup}}))^{\frac{2}{3}}, \quad B_8 = \frac{2B_5}{|\log a|} + \frac{18B_7}{1-a^{\frac{1}{3}}}, \\
 B_9 &= (g_{\text{inf}})^{-2} \|K\|_\infty^2 \{8((r_{\text{sup}})^4 + 3\sigma^4)\}^{\frac{1}{2}} \left(\frac{2}{\sqrt{2\pi}} \right)^{\frac{1}{2}}, \\
 B_{10} &= \frac{2^4 3}{1 - e^{-\frac{\gamma-2}{2}}} (A_1 + B_8) + \frac{5! 2^{18} 3 B^2}{1 - e^{-\frac{\gamma-2}{2}}} (\sigma + r_{\text{sup}})^2, \\
 C_* &= A_0^2 + A_1 + A_2 \quad \text{y} \quad C^* = A_0 A_6 + 2e^{-\frac{1}{2}} A_7 + A_8.
 \end{aligned}$$

3.5.2. Enunciados de resultados técnicos.

Lemma 1. *Bajo las hipótesis H_1, H_2, H_4 y H_5 se tiene que:*

$$i) \mathbb{E} \left[\left(K_h(x - X_i) \varepsilon_i g^{-1}(X_i) \right)^2 \right] \leq \frac{B_1}{h},$$

$$ii) \mathbb{E} \left[\left(K_h(x - X_i) r(X_i) g^{-1}(X_i) \right)^2 \right] \leq \frac{B_2}{h},$$

$$iii) \mathbb{E} \left[|K_h(x - X_i) K_h(x - X_j)| \right] \leq Q \|K\|_1^2,$$

$$iv) \mathbb{E} \left[|K_h(x - X_i) r(X_i) g^{-1}(X_i)| \right] \leq r_{\text{sup}} \|K\|_1,$$

donde $B_1 = \sigma^2 (g_{\text{inf}})^{-1} \|K\|_2^2$, y $B_2 = (r_{\text{sup}})^2 (g_{\text{inf}})^{-1} \|K\|_2^2$.

Lemma 2. *Para cada $h \in \mathcal{H}$, $a \in]0, 1[$ y $n \geq 4$, se cumple que,*

$$\frac{1}{h} a^{\left\{ \frac{1}{h |\log a| (\log n)^{\frac{1}{2}}} \right\}} \leq \frac{10}{(\log n)^{\frac{5}{2}}}.$$

Proposición 13. *Asumimos que los X_i satisfacen la hipótesis (H_3) . Sea $k, u, v \in \mathbb{N}$, tales que $i_{1:u} \in \mathbb{Z}^u$, $j_{1:v} \in \mathbb{Z}^v$ y tal que $i_1 \leq \dots \leq i_u < i_u + k \leq j_1 \leq \dots \leq j_v$. Si $G_u : \mathbb{R}^u \times \mathbb{R}^u \rightarrow \mathbb{R}$ y $G_v : \mathbb{R}^v \times \mathbb{R}^v \rightarrow \mathbb{R}$ son funciones tales que $\|G_u\|_\infty < \infty$ y $\|G_v\|_\infty < \infty$ entonces,*

$$|\text{cov}(G_u(X_{i_{1:u}}, \varepsilon_{i_{1:u}}), G_v(X_{j_{1:v}}, \varepsilon_{j_{1:v}}))| \leq \Psi(u, v, G_u, G_v) a^k$$

donde $\Psi(u, v, G_u, G_v) = 4 \|G_u\|_\infty \|G_v\|_\infty$, con $\|G_u\|_\infty = \sup_{(x,y) \in \mathbb{R}^u \times \mathbb{R}^u} |G_u(x,y)|$ y $\|G_v\|_\infty = \sup_{(x,y) \in \mathbb{R}^v \times \mathbb{R}^v} |G_v(x,y)|$.

A continuación vamos a trabajar con funciones G_u específicas. Mas precisamente consideramos para $u \in \mathbb{N}$, y $i_{1:u} \in \mathbb{Z}^u$, $G_u : \mathbb{R}^u \times \mathbb{R}^u \rightarrow \mathbb{R}$ dada por

$$G_u(X_{i_{1:u}}, \varepsilon_{i_{1:u}}) = \prod_{k=1}^u \{G(X_{i_k}, \varepsilon_{i_k}) - \mathbb{E}(G(X_{i_k}, \varepsilon_{i_k}))\}$$

donde para $i \in \mathbb{N}$,

$$G(X_i, \varepsilon_i) = \frac{1}{n} (r(X_i) + \varepsilon_i) K_h(x - X_i) g^{-1}(X_i) \mathbf{1}_{\{|(r(X_i) + \varepsilon_i)| \leq M_n\}},$$

donde $M_n = \sigma \log n + r_{\text{sup}}$. Las variables $G(X_i, \varepsilon_i)$ satisfacen los siguientes resultados.

Lemma 3. *Bajo las hipótesis (H_1) , (H_2) , (H_4) y (H_5) se tiene que: $\forall i, k \in \mathbb{N}$*

- i) $|\mathbb{E}[G(X_i, \varepsilon_i)]| \leq \frac{B_3}{n}$,
- ii) $|\mathbb{E}[G(X_i, \varepsilon_i)G(X_{i+k}, \varepsilon_{i+k})]| \leq \frac{B_4}{n^2}$,
- iii) $|\text{Cov}(G(X_i, \varepsilon_i), G(X_{i+k}, \varepsilon_{i+k}))| \leq \frac{B_5}{n^2}$,
- iv) $|G(X_i, \varepsilon_i)| \leq \frac{B_6 M_n}{2nh}$

donde

$$B_3 = \left(r_{\text{sup}} + \frac{2\sigma}{\sqrt{2\pi}} \right) \|K\|_1, \quad B_4 = \left(r_{\text{sup}} + \frac{2\sigma}{\sqrt{2\pi}} \right)^2 (g_{\text{inf}})^{-2} Q \|K\|_1^2$$

$$B_5 = \left(r_{\text{sup}} + \frac{2\sigma}{\sqrt{2\pi}} \right)^2 \|K\|_1^2 ((g_{\text{inf}})^{-2} Q + 1) \quad \text{y} \quad B_6 = 2\|K\|_{\infty} (g_{\text{inf}})^{-1}.$$

Proposición 14. Sean $u, v, k \in \mathbb{N}$, si $(i_1, \dots, i_u, j_1, \dots, j_v) \in \mathbb{Z}^{u+v}$ es tal que $i_1 \leq \dots \leq i_u < i_u + k \leq j_1 \leq \dots \leq j_v$. Bajo las hipótesis (H_1) , (H_2) , (H_3) , (H_4) y (H_5) tenemos, $\forall h \in \mathcal{H}$

$$\gamma_h(u, v) = |\text{cov}(G_u(X_{i_{1:u}}, \varepsilon_{i_{1:u}}), G_v(X_{j_{1:v}}, \varepsilon_{j_{1:v}}))| \leq \phi(u, v) (D_1(h))^{u+v-2} D_2(h) a^{\frac{k}{3}}$$

donde $\phi(u, v) = u + v + uv$,

$$D_1(h) = \frac{B_6 M_n}{nh}, \quad D_2(h) = \frac{B_7}{n^2 h} \quad \text{y} \quad B_7 = (4B_5 B_6 (\sigma + r_{\text{sup}}))^{\frac{2}{3}}.$$

Proposición 15. Desigualdad de Bernstein. Sea el proceso $\mathbb{Z} = (Z_i)_{i \in \mathbb{N}}$, donde $Z_i = G(X_i, \varepsilon_i) - \mathbb{E}[G(X_i, \varepsilon_i)]$ y $S_n = \sum_{i=1}^n Z_i$. Bajo las hipótesis (H_1) , (H_2) , (H_3) , (H_4) y (H_5) tenemos para $h \in \mathcal{H}$ y $\forall t \geq 0$

$$\mathbb{P}(|S_n| \geq t) \leq \exp\left(-\frac{\frac{t^2}{2}}{\mathcal{A}_n + \mathcal{B}_n^{\frac{1}{3}} t^{\frac{5}{3}}}\right) \quad (3.12)$$

donde $\mathcal{A}_n = \frac{A_1}{nh} + B_8 \frac{(\log n)^{-\frac{1}{2}}}{nh}$ es una cota superior de $\text{Var}(S_n)$, $\mathcal{B}_n = B \frac{M_n}{nh}$, con

$$B_8 = \frac{2B_5}{|\log a|} + \frac{18B_7}{1 - a^{\frac{1}{3}}}, \quad B = \frac{12(B_6 \vee \sqrt{B_7})}{1 \wedge |\log a|} \left(\frac{2^7 3 B_7 a^{-\frac{1}{3}}}{A_1 (1 \wedge |\log a|)} \vee 1 \right).$$

Siguiendo el esquema de un procedimiento que se plantea en el Lema 3 de Bertin and Klutchnikoff (2017), se puede reescribir la desigualdad de Bernstein de otra forma, como se establece en el siguiente Corolario.

Corolario 1. *Bajo las hipótesis de la Proposición 15, si $\lambda(u) = (2\mathcal{A}_n u)^{\frac{1}{2}} + \mathcal{B}_n(2u)^3$ para cada $u \geq 0$ y se toma el cambio de variable $t = \lambda(u)$ en la desigualdad de Bernstein (3.12), entonces tal desigualdad se puede reescribir de la siguiente forma*

$$\mathbb{P}(|S_n| \geq \lambda(u)) \leq \exp\left(-\frac{u}{2}\right)$$

para cada $u \geq 0$.

3.5.3. Demostraciones de resultados técnicos.

Demostración Lema 1.

i) Tenemos

$$\begin{aligned} \mathbb{E} \left[(K_h(x - X_i) \varepsilon_i g^{-1}(X_i))^2 \right] &= \mathbb{E} \left[K_h^2(x - X_i) \varepsilon_i^2 g^{-2}(X_i) \right] \\ &= \frac{\sigma^2}{h^2} \mathbb{E} \left[K^2 \left(\frac{x - X_i}{h} \right) g^{-2}(X_i) \right] \\ &= \frac{\sigma^2}{h^2} \int_{\mathbb{R}} K^2 \left(\frac{x - t}{h} \right) g^{-1}(t) dt \\ &\leq \frac{\sigma^2 (g_{\inf})^{-1}}{h^2} \int_{\mathbb{R}} K^2(u) h du = \frac{B_1}{h}. \end{aligned}$$

ii) Tenemos

$$\begin{aligned} \mathbb{E} \left[(K_h(x - X_i) r(X_i) g^{-1}(X_i))^2 \right] &= \mathbb{E} \left[\frac{1}{h^2} K^2 \left(\frac{x - X}{h} \right) r^2(X) g^{-2}(X) \right] \\ &= \frac{1}{h^2} \int_{\mathbb{R}} K^2 \left(\frac{x - t}{h} \right) r^2(t) g^{-1}(t) dt \\ &\leq \frac{(r_{\sup})^2 (g_{\inf})^{-1}}{h^2} \int_{\mathbb{R}} K^2(u) h du = \frac{B_2}{h}. \end{aligned}$$

iii) Tenemos

$$\begin{aligned} \mathbb{E} \left[|K_h(x - X_i) K_h(x - X_j)| \right] &= \int \int |K_h(x - t) K_h(x - s)| g_{i,j}(t, s) dt ds \\ &\leq \frac{1}{h^2} \int \int \left| K \left(\frac{x - t}{h} \right) K \left(\frac{x - s}{h} \right) \right| Q dt ds \\ &\leq \frac{Q}{h^2} \int \int |K(u)| |K(v)| h^2 du dv = Q \|K\|_1^2 \end{aligned}$$

iv) Tenemos

$$\begin{aligned} \mathbb{E} [|K_h(x - X_i) r(X_i) g^{-1}(X_i)|] &\leq r_{\text{sup}} \mathbb{E} [|K_h(x - X_i) g^{-1}(X_i)|] \\ &= r_{\text{sup}} \int |K_h(x - t)| dt \\ &= r_{\text{sup}} \int \frac{1}{h} \left| K \left(\frac{x - t}{h} \right) \right| dt = r_{\text{sup}} \|K\|_1 \end{aligned}$$

Demostración Lema 2.

Como $h \in \mathcal{H}$ se tiene que $\frac{(\log n)^8}{n} \leq h < \frac{1}{(\log n)^2}$, por lo cual $\frac{\log n}{n} \leq h \leq \frac{1}{(\log n)^2}$, este conjunto de desigualdades se usan para acotar la expresión $\frac{1}{h} a^{\left\{ \frac{1}{h |\log a| (\log n)^{\frac{1}{2}}} \right\}}$.

$$\begin{aligned} \frac{1}{h} a^{\left\{ \frac{1}{h |\log a| (\log n)^{\frac{1}{2}}} \right\}} &\leq \frac{n}{\log n} a^{\left\{ \frac{1}{h |\log a| (\log n)^{\frac{1}{2}}} \right\}} \\ &\leq \frac{n}{\log n} a^{\left\{ (\log n)^2 \frac{1}{|\log a| (\log n)^{\frac{1}{2}}} \right\}} \\ &= \frac{n}{\log n} a^{\frac{(\log n)^{\frac{3}{2}}}{|\log a|}} \\ &= \frac{1}{\log n} a^{\log_a n} a^{\frac{(\log n)^{\frac{3}{2}}}{|\log a|}} \\ &= \frac{1}{\log n} a^{\frac{(\log n)^{\frac{3}{2}}}{|\log a|} + \frac{\log n}{\log a}} \\ &= \frac{1}{\log n} a^{\frac{(\log n)^{\frac{3}{2}}}{|\log a|}} \left\{ 1 - \frac{1}{(\log n)^{\frac{1}{2}}} \right\} \end{aligned} \tag{3.13}$$

Si $n \geq 4$ entonces $-\frac{9}{10} < -\frac{1}{(\log n)^{\frac{1}{2}}} < 0$.

Al tomar en la desigualdad (3.13) $n \geq 4$ se tiene,

$$\frac{1}{h} a^{\left\{ \frac{1}{h |\log a| (\log n)^{\frac{1}{2}}} \right\}} \leq \frac{1}{\log n} a^{\frac{(\log n)^{\frac{3}{2}}}{|\log a|}} \left\{ 1 - \frac{9}{10} \right\} = \frac{1}{\log n} a^{\frac{1}{10} \frac{(\log n)^{\frac{3}{2}}}{|\log a|}}$$

Como $a \in]0, 1[$ y $e^{-x} \leq \frac{1}{x}, \forall x > 0$ entonces,

$$a^{\frac{1}{10} \frac{(\log n)^{\frac{3}{2}}}{|\log a|}} \leq \left(e^{\log a} \right)^{\frac{1}{10} \frac{(\log n)^{\frac{3}{2}}}{|\log a|}} \leq e^{\left(-\frac{1}{10} \frac{(\log n)^{\frac{3}{2}}}{|\log a|} |\log a| \right)} \leq \frac{10}{(\log n)^{\frac{3}{2}}}$$

De las dos desigualdades anteriores se concluye que,

$$\frac{1}{h} a^{\left\{ \frac{1}{h |\log a| (\log n)^{\frac{1}{2}}} \right\}} \leq \frac{10}{(\log n)^{\frac{5}{2}}}$$

para $n \geq 4$.

Demostración de la Proposición 13.

Tenemos

$$\text{cov}(G_u(X_{i_{1:u}}, \boldsymbol{\varepsilon}_{i_{1:u}}), G_v(X_{j_{1:v}}, \boldsymbol{\varepsilon}_{j_{1:v}})) = \mathbb{E} \left[\text{cov}(G_u(X_{i_{1:u}}, \boldsymbol{\varepsilon}_{i_{1:u}}), G_v(X_{j_{1:v}}, \boldsymbol{\varepsilon}_{j_{1:v}}) | \boldsymbol{\varepsilon}_{i_{1:u}} \boldsymbol{\varepsilon}_{j_{1:v}}) \right]. \quad (3.14)$$

Como en $\text{cov}(G_u(X_{i_{1:u}}, \boldsymbol{\varepsilon}_{i_{1:u}}), G_v(X_{j_{1:v}}, \boldsymbol{\varepsilon}_{j_{1:v}}) | \boldsymbol{\varepsilon}_{i_{1:u}} \boldsymbol{\varepsilon}_{j_{1:v}})$, los $\boldsymbol{\varepsilon}_{i_{1:u}} \boldsymbol{\varepsilon}_{j_{1:v}}$ están dados (observados), entonces las funciones involucradas $G_u(\cdot, \boldsymbol{\varepsilon}_{i_{1:u}})$ y $G_v(\cdot, \boldsymbol{\varepsilon}_{j_{1:v}})$ tienen como variable independiente la primera variable, por lo cuál, usando la ecuación (3.1),

$$\begin{aligned} |\text{cov}(G_u(X_{i_{1:u}}, \boldsymbol{\varepsilon}_{i_{1:u}}), G_v(X_{j_{1:v}}, \boldsymbol{\varepsilon}_{j_{1:v}}) | \boldsymbol{\varepsilon}_{i_{1:u}} \boldsymbol{\varepsilon}_{j_{1:v}})| &\leq 4 \|G_u(\cdot, \boldsymbol{\varepsilon}_{i_{1:u}})\|_{\infty} \|G_v(\cdot, \boldsymbol{\varepsilon}_{j_{1:v}})\|_{\infty} \alpha_k(\mathbb{X}) \\ &\leq 4 \|G_u\|_{\infty} \|G_v\|_{\infty} \alpha_k(\mathbb{X}) \end{aligned}$$

Por la hipótesis (H3), la desigualdad anterior y la ecuación (3.14), se tiene que

$$|\text{cov}(G_u(X_{i_{1:u}}, \boldsymbol{\varepsilon}_{i_{1:u}}), G_v(X_{j_{1:v}}, \boldsymbol{\varepsilon}_{j_{1:v}}))| \leq 4 \|G_u\|_{\infty} \|G_v\|_{\infty} a^k.$$

Demostración del Lema 3.

i) Tenemos

$$\begin{aligned}
|\mathbb{E}(G(X_i, \varepsilon_i))| &\leq \mathbb{E} \left\{ \left| \frac{1}{n} (r(X_i) + \varepsilon_i) K_h(x - X_i) g^{-1}(X_i) \mathbf{1}_{\{|r(X_i) + \varepsilon_i| \leq M_n\}} \right| \right\} \\
&\leq \frac{1}{n} \mathbb{E} \{ |r(X_i) + \varepsilon_i| |K_h(x - X_i) g^{-1}(X_i)| \} \\
&\leq \frac{1}{n} \mathbb{E} \{ |r(X_i) K_h(x - X_i) g^{-1}(X_i)| \} + \frac{1}{n} \mathbb{E} \{ |\varepsilon_i K_h(x - X_i) g^{-1}(X_i)| \} \\
&\leq \frac{1}{n} r_{\text{sup}} \|K\|_1 + \frac{1}{n} \frac{2\sigma}{\sqrt{2\pi}} \int_{\mathbb{R}} |K_h(x - t)| dt = \frac{B_3}{n}
\end{aligned}$$

por la desigualdad iv) del Lema 1.

ii) Se denota $A(X_i) = \frac{1}{n} K_h(x - X_i) g^{-1}(X_i)$ por lo cual,

$$\begin{aligned}
&|\mathbb{E}[G(X_i, \varepsilon_i)G(X_{i+k}, \varepsilon_{i+k})]| \\
&\leq \mathbb{E} [|Y_i A(X_i) \mathbf{1}_{\{|Y_i| \leq M_n\}} Y_{i+k} A(X_{i+k}) \mathbf{1}_{\{|Y_{i+k}| \leq M_n\}}|] \\
&\leq \mathbb{E} [|r(X_i) + \varepsilon_i| |A(X_i)| |r(X_{i+k}) + \varepsilon_{i+k}| |A(X_{i+k})|] \\
&\leq \mathbb{E} [|r(X_i) r(X_{i+k}) A(X_i) A(X_{i+k})|] + \mathbb{E} [|r(X_i) \varepsilon_{i+k} A(X_i) A(X_{i+k})|] \\
&\quad + \mathbb{E} [|\varepsilon_i r(X_{i+k}) A(X_i) A(X_{i+k})|] + \mathbb{E} [|\varepsilon_i \varepsilon_{i+k} A(X_i) A(X_{i+k})|] \\
&\leq (r_{\text{sup}})^2 \mathbb{E} [|A(X_i) A(X_{i+k})|] + r_{\text{sup}} \mathbb{E} [|\varepsilon_{i+k}|] \mathbb{E} [|A(X_i) A(X_{i+k})|] \\
&\quad + r_{\text{sup}} \mathbb{E} [|\varepsilon_i|] \mathbb{E} [|A(X_i) A(X_{i+k})|] + \mathbb{E} [|\varepsilon_i|] \mathbb{E} [|\varepsilon_{i+k}|] \mathbb{E} [|A(X_i) A(X_{i+k})|] \\
&= \left\{ (r_{\text{sup}})^2 + 2r_{\text{sup}} \frac{2\sigma}{\sqrt{2\pi}} + \left(\frac{2\sigma}{\sqrt{2\pi}} \right)^2 \right\} \mathbb{E} [|A(X_i) A(X_{i+k})|] \\
&\leq \left(r_{\text{sup}} + \frac{2\sigma}{\sqrt{2\pi}} \right)^2 (g_{\text{inf}})^{-2} \frac{1}{n^2} \mathbb{E} [|K_h(x - X_i) K_h(x - X_{i+k})|] \\
&\leq \left(r_{\text{sup}} + \frac{2\sigma}{\sqrt{2\pi}} \right)^2 (g_{\text{inf}})^{-2} Q \|K\|_1^2 \frac{1}{n^2} \\
&= \frac{B_4}{n^2}.
\end{aligned}$$

iii) Ahora usando (i) y (ii), se tiene

$$\begin{aligned}
&|\text{Cov}(G(X_i, \varepsilon_i), G(X_{i+k}, \varepsilon_{i+k}))| \\
&\leq |\mathbb{E}[G(X_i, \varepsilon_i)G(X_{i+k}, \varepsilon_{i+k})]| + |\mathbb{E}[G(X_i, \varepsilon_i)]| |\mathbb{E}[G(X_{i+k}, \varepsilon_{i+k})]| \\
&\leq \frac{B_4}{n^2} + \frac{B_3}{n} \frac{B_3}{n} = \frac{B_5}{n^2}.
\end{aligned}$$

iv) La desigualdad (iv) es inmediata usando hipótesis (H_1) y (H_5) .

Demostración de la Proposición 14.

Para probar esta Proposición, se determinan dos cotas diferentes del término

$$\gamma_h(u, v) = |\text{cov}(G_u(X_{i_{1:u}}, \boldsymbol{\varepsilon}_{i_{1:u}}), G_v(X_{j_{1:v}}, \boldsymbol{\varepsilon}_{j_{1:v}}))|$$

la primera cota se obtiene por cálculo directo mientras que la segunda cota se obtiene gracias a la estructura de dependencia de las observaciones. A lo largo de la prueba se denota $l = u + v$.

Cota Directa. La prueba de esta cota esta compuesta por dos pasos. Primero se asume que $l = 2$, después se considera el caso $l \geq 3$.

Para $l = 2$ se tiene que $u = v = 1$, $i_{1:1} = i_1$ y $j_{1:1} = j_1$, por lo cuál, usando la desigualdad (iii) del Lema 3, se tiene

$$\gamma_h(1, 1) = |\text{cov}(G(X_{i_1}, \boldsymbol{\varepsilon}_{i_1}), G(X_{j_1}, \boldsymbol{\varepsilon}_{j_1}))| \leq \frac{B_5}{n^2}. \quad (3.15)$$

Ahora se asume que $l \geq 3$, sin perdida de generalidad se puede tomar $u \geq 2$ y $v \geq 1$. Se tiene

$$\gamma_h(u, v) = |\text{cov}(G_u(X_{i_{1:u}}, \boldsymbol{\varepsilon}_{i_{1:u}}), G_v(X_{j_{1:v}}, \boldsymbol{\varepsilon}_{j_{1:v}}))| \leq \Gamma_1 + \Gamma_2 \quad (3.16)$$

donde

- $\Gamma_1 = |\mathbb{E} [\prod_{k=1}^u \{G(X_{i_k}, \boldsymbol{\varepsilon}_{i_k}) - \mathbb{E}(G(X_{i_k}, \boldsymbol{\varepsilon}_{i_k}))\} \prod_{m=1}^v \{G(X_{j_m}, \boldsymbol{\varepsilon}_{j_m}) - \mathbb{E}(G(X_{j_m}, \boldsymbol{\varepsilon}_{j_m}))\}]|$
- $\Gamma_2 = |\mathbb{E} [\prod_{k=1}^u G(X_{i_k}, \boldsymbol{\varepsilon}_{i_k}) - \mathbb{E}(G(X_{i_k}, \boldsymbol{\varepsilon}_{i_k}))] \mathbb{E} [\prod_{m=1}^v G(X_{j_m}, \boldsymbol{\varepsilon}_{j_m}) - \mathbb{E}(G(X_{j_m}, \boldsymbol{\varepsilon}_{j_m}))]|$

Usando que $|G(X_i, \boldsymbol{\varepsilon}_i) - \mathbb{E}(G(X_i, \boldsymbol{\varepsilon}_i))| \leq B_6 \frac{M_n}{nh}$ (ver (iv) Lema 3) y la desigualdad (3.15), se tiene,

$$\begin{aligned} \Gamma_1 &\leq \left(B_6 \frac{M_n}{nh}\right)^{u+v-2} |\mathbb{E} \{[G(X_{i_1}, \boldsymbol{\varepsilon}_{i_1}) - \mathbb{E}(G(X_{i_1}, \boldsymbol{\varepsilon}_{i_1}))][G(X_{i_2}, \boldsymbol{\varepsilon}_{i_2}) - \mathbb{E}(G(X_{i_2}, \boldsymbol{\varepsilon}_{i_2}))]\}| \\ &\leq \left(B_6 \frac{M_n}{nh}\right)^{l-2} \frac{B_5}{n^2} \end{aligned}$$

Análogamente

$$\Gamma_2 \leq \left(B_6 \frac{M_n}{nh}\right)^{l-2} \frac{B_5}{n^2}$$

Por la desigualdad (3.16) y las cotas obtenidas de Γ_1 y Γ_2 , se tiene $\forall l \geq 2$ que

$$\gamma_h(u, v) \leq 2 \left(B_6 \frac{M_n}{nh} \right)^{l-2} \frac{B_5}{n^2}. \quad (3.17)$$

Cota Estructural. Se dará otra cota para $\gamma_h(u, v)$ usando la estructura de dependencia.

Como $|G(X_{i_k}, \varepsilon_{i_k}) - \mathbb{E}(G(X_{i_k}, \varepsilon_{i_k}))| \leq B_6 \frac{M_n}{nh}$, para cada $k = 1, \dots, u$, entonces

$$|G_u(X_{i_{1:u}}, \varepsilon_{i_{1:u}})| = \prod_{k=1}^u |G(X_{i_k}, \varepsilon_{i_k}) - \mathbb{E}(G(X_{i_k}, \varepsilon_{i_k}))| \leq \left(B_6 \frac{M_n}{nh} \right)^u.$$

Por lo cuál $\|G_u\|_\infty \leq \left(B_6 \frac{M_n}{nh} \right)^u$, análogamente $\|G_v\|_\infty \leq \left(B_6 \frac{M_n}{nh} \right)^v$. Por esto último y por la Proposición 13, se tiene que

$$\begin{aligned} \gamma_h(u, v) &\leq 4 \left(B_6 \frac{M_n}{nh} \right)^l a^k \\ &= \left(B_6 \frac{M_n}{nh} \right)^{l-2} 4B_6^2 \frac{M_n^2}{n^2 h^2} a^k \end{aligned} \quad (3.18)$$

Como $M_n = \sigma \log n + r_{\text{sup}}$ y $h \leq \frac{1}{(\log n)^2}$ entonces $M_n^2 \leq (\sigma + r_{\text{sup}})^2 (\log n)^2 \leq (\sigma + r_{\text{sup}})^2 \frac{1}{h}$. Por esto último y por la desigualdad (3.18) se tiene la cota estructural,

$$\gamma_h(u, v) \leq \left(B_6 \frac{M_n}{nh} \right)^{l-2} \frac{4B_6^2 (\sigma + r_{\text{sup}})^2}{n^2 h^3} a^k \quad (3.19)$$

Combinación de las dos cotas: Ahora se combinan (3.17) y (3.19), de la siguiente forma

$$\begin{aligned} \gamma_h(u, v) &= (\gamma_h(u, v))^{\frac{2}{3}} (\gamma_h(u, v))^{\frac{1}{3}} \\ &\leq \left(2 \left(B_6 \frac{M_n}{nh} \right)^{l-2} \frac{B_5}{n^2} \right)^{\frac{2}{3}} \left(\left(B_6 \frac{M_n}{nh} \right)^{l-2} \frac{4B_6^2 (\sigma + r_{\text{sup}})^2}{n^2 h^3} a^k \right)^{\frac{1}{3}} \\ &= \left(B_6 \frac{M_n}{nh} \right)^{l-2} (2B_5)^{\frac{2}{3}} (2B_6 (\sigma + r_{\text{sup}}))^{\frac{2}{3}} \frac{1}{n^2 h} a^{\frac{k}{3}} \\ &= \left(B_6 \frac{M_n}{nh} \right)^{l-2} \frac{B_7}{n^2 h} a^{\frac{k}{3}} \\ &\leq (D_1(h))^{l-2} D_2(h) \phi(u, v) a^{\frac{k}{3}}. \end{aligned}$$

Demostración de la Proposición 15.

Vamos a usar el teorema 1 en Doukhan and Neumann (2007). Este teorema indica que si existen constantes $D_1, D_2, L_1, L_2 < \infty$ y una sucesión no creciente de coeficientes reales $(\rho_k)_{k \in \mathbb{N}}$ tales que para cada $u, v \in \mathbb{N}, i_{1:u} = (i_1, i_2, \dots, i_u) \in \mathbb{N}^u$ y $j_{1:v} = (j_1, j_2, \dots, j_v) \in \mathbb{N}^v$ con $i_1 \leq i_2 \leq \dots \leq i_u < i_u + k \leq j_1 \leq j_2 \leq \dots \leq j_v$, se satisfacen las siguiente desigualdades

$$\left| \text{cov} \left(\prod_{s=1}^u Z_{i_s}, \prod_{s=1}^v Z_{j_s} \right) \right| \leq D_2 D_1^{u+v-2} \phi(u, v) \rho_k \quad (3.20)$$

$$\sum_{s=0}^{\infty} (s+1)^r \rho_s \leq L_1 L_2^r r!, \quad \forall r \in \mathbb{N}, \quad (3.21)$$

y

$$\mathbb{E}|Z_i|^r \leq D_1^r, \quad \forall r \in \mathbb{N}, \quad (3.22)$$

entonces se tiene que, para todo $t \geq 0$

$$\mathbb{P}(S_n \geq t) \leq \exp \left(- \frac{\frac{t^2}{2}}{\mathcal{A}_n + (\mathcal{B}_n^*)^{\frac{1}{3}} t^{\frac{5}{3}}} \right)$$

donde \mathcal{A}_n es una cota superior de la varianza de S_n y

$$\mathcal{B}_n^* = 2(\sqrt{D_2} \vee D_1) L_2 \left(\frac{2^5 n D_2 L_1}{\mathcal{A}_n} \vee 1 \right). \quad (3.23)$$

A continuación vamos a verificar que se cumplen las tres desigualdades (3.20), (3.21) y (3.22). Vamos a mostrar que $\text{Var}(S_n) \leq \mathcal{A}_n$ y $\mathcal{B}_n^* \leq \mathcal{B}_n$. Esto implica el resultado de la proposición.

Primero, usando Proposición 14, se comprueba que la desigualdad (3.20) esta satisfecha con, $D_1 = D_1(h)$, $D_2 = D_2(h)$ y la sucesión no creciente $(\rho_k)_{k \in \mathbb{N}}$ con $\rho_k = a^{\frac{k}{3}}$ para $k \in \mathbb{N}$.

Ahora vamos a demostrar que la desigualdad (3.21) se cumple. Sea $r \in \mathbb{N}$. Tenemos

$$\begin{aligned} \sum_{s=0}^{\infty} (s+1)^r \rho_s &= \sum_{s=0}^{\infty} (s+1)^r a^{\frac{s}{3}} = \sum_{s=0}^{\infty} \int_s^{s+1} (s+1)^r a^{\frac{s}{3}} dx \\ &\leq \sum_{s=0}^{\infty} \int_s^{s+1} (x+1)^r a^{\frac{x}{3}} a^{-\frac{1}{3}} dx = a^{-\frac{1}{3}} \int_0^{\infty} (x+1)^r a^{\frac{x}{3}} dx, \end{aligned} \quad (3.24)$$

donde usamos que si $s \leq x \leq s + 1$, entonces $a^{\frac{x}{3}} \leq a^{\frac{s}{3}} a^{-\frac{1}{3}}$ y $(s + 1)^r \leq (x + 1)^r$.

Al aplicar integración por partes r -veces, se tiene que.

$$\begin{aligned} & \int_0^\infty (x+1)^r a^{\frac{x}{3}} dx \\ &= -\frac{3}{\log a} + r \left(-\frac{3}{\log a}\right)^2 + r(r-1) \left(-\frac{3}{\log a}\right)^3 + \dots + r! \left(-\frac{3}{\log a}\right)^{r+1} \\ &\leq 3^{r+1} r! \sum_{j=1}^{r+1} \left(\frac{1}{|\log a|}\right)^j \end{aligned}$$

Por está última desigualdad y la desigualdad (3.24) se tiene que,

$$\begin{aligned} \sum_{s=0}^\infty (s+1)^r \rho_s &\leq a^{-\frac{1}{3}} 3^{r+1} r! \sum_{j=0}^{r+1} \left(\frac{1}{|\log a|}\right)^j \\ &= a^{-\frac{1}{3}} 3^{r+1} r! \sum_{j=0}^{r+1} \frac{1}{|\log a|^j} \frac{2^j}{2^j} \\ &\leq a^{-\frac{1}{3}} 3^{r+1} r! 2^{r+1} \sum_{j=0}^{r+1} \frac{1}{|\log a|^j} \frac{1}{2^j} \\ &\leq a^{-\frac{1}{3}} 6^{r+1} r! \frac{1}{(1 \wedge |\log a|)^{r+1}} \sum_{j=0}^{r+1} \frac{1}{2^j} \\ &\leq a^{-\frac{1}{3}} \left(\frac{6}{1 \wedge |\log a|}\right)^{r+1} r! 2 = L_1 L_2^r r! \end{aligned}$$

donde $L_1 = \frac{12a^{-\frac{1}{3}}}{1 \wedge |\log a|}$ y $L_2 = \frac{6}{1 \wedge |\log a|}$, verificándose así la cota planteada en la desigualdad (3.21).

Desigualdad (3.22) es una consecuencia de (iv) del Lema 3.

Ahora se determina la cota \mathcal{A}_n de la varianza S_n Usando que $\mathbb{E}[S_n] = 0$ tenemos

$$\text{Var}[S_n] = \mathbb{E} \left[\left(\sum_{i=1}^n Z_i \right)^2 \right] = \sum_{i=1}^n \mathbb{E}[Z_i^2] + 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^n \mathbb{E}[Z_i Z_j] \leq nJ_1 + 2J_2$$

donde $J_1 = \mathbb{E}[Z_1^2]$ y $J_2 = \sum_{i=1}^{n-1} \sum_{r=1}^{n-i} |\mathbb{E}[Z_i Z_{i+r}]|$. Tenemos usando (i) and (ii) del Lema 1 que

$$\begin{aligned}
J_1 &\leq \mathbb{E} [(G(X_1, \varepsilon_1))^2] \\
&\leq \frac{1}{n^2} \mathbb{E} [(r(X_1) + \varepsilon_1)^2 K_h^2(x - X_1) g^{-2}(X_1)] \\
&\leq \frac{1}{n^2} \mathbb{E} [(r(X_1))^2 K_h^2(x - X_1) g^{-2}(X_1)] + \frac{1}{n^2} \mathbb{E} [(\varepsilon_1)^2 K_h^2(x - X_1) g^{-2}(X_1)] \\
&\leq \frac{1}{n^2} \frac{B_2}{h} + \frac{1}{n^2} \frac{B_1}{h} = \frac{A_1}{n^2 h}.
\end{aligned} \tag{3.25}$$

Sea (u_n) una secuencia tal que $1 \leq u_n \leq n - 1$. Tenemos

$$J_2 = \sum_{i=1}^{n-1} \left(\sum_{r=1}^{u_n} |Cov(Z_i, Z_{i+r})| + \sum_{r=u_n+1}^{n-i} |Cov(Z_i, Z_{i+r})| \right) \tag{3.26}$$

Usando la desigualdad (iii) del Lema 3, se obtiene que

$$\sum_{r=1}^{u_n} |Cov(Z_i, Z_{i+r})| = \sum_{r=1}^{u_n} |Cov(G(X_i, \varepsilon_i), G(X_{i+r}, \varepsilon_{i+r}))| \leq \sum_{r=1}^{u_n} \frac{B_5}{n^2} = B_5 \frac{u_n}{n^2} \tag{3.27}$$

Usando la Proposición 14 se tiene que,

$$|Cov(Z_i, Z_{i+r})| \leq \phi(1, 1) D_1^0(h) D_2(h) \alpha_r^{\frac{1}{3}} = 3 \frac{B_7}{n^2 h} a^{\frac{r}{3}}.$$

Por lo anterior se tiene,

$$\sum_{r=u_n+1}^{n-i} |Cov(Z_i, Z_{i+r})| \leq \frac{3B_7}{n^2 h} \sum_{r=u_n+1}^{\infty} a^{\frac{r}{3}} = \frac{3B_7}{n^2 h} \frac{(a^{\frac{1}{3}})^{u_n+1}}{1 - a^{\frac{1}{3}}} \tag{3.28}$$

Por las desigualdades (3.26), (3.27) y (3.28), se tiene,

$$J_2 \leq \frac{1}{n} \left(B_5 u_n + \frac{3B_7}{1 - a^{\frac{1}{3}}} \frac{(a^{\frac{1}{3}})^{u_n+1}}{h} \right) \tag{3.29}$$

En la demostración de la Proposición 10, se toma la variable u_n como $u_n = \left\lceil \frac{1}{h |\log a| (\log n)^{\frac{1}{2}}} \right\rceil$ y se usan los hechos de que $[x] \leq x \leq [x] + 1$ y $0 < a < 1$, para seguir acotando la desigualdad anterior,

$$\begin{aligned}
J_2 &\leq \frac{1}{n} \left(B_5 \frac{1}{h|\log a|(\log n)^{\frac{1}{2}}} + \frac{3B_7}{1-a^{\frac{1}{3}}} (a^{\frac{1}{3}})^{\left\{ \frac{1}{h|\log a|(\log n)^{\frac{1}{2}}} \right\}} \frac{1}{h} \right) \\
&= \frac{1}{nh} \left(\frac{B_5}{|\log a|(\log n)^{\frac{1}{2}}} + \frac{3B_7}{1-a^{\frac{1}{3}}} (a^{\frac{1}{3}})^{\left\{ \frac{1}{h|\log a|(\log n)^{\frac{1}{2}}} \right\}} \right) \tag{3.30}
\end{aligned}$$

Como $h \in \mathcal{H}$ entonces $h \leq \frac{1}{(\log n)^2}$, además, por el hecho de que $e^{-x} < \frac{1}{x}$ para cada $x \in \mathbb{R}$ y que $a \in]0, 1[$, siguiendo un procedimiento análogo al realizado en el Lema 2. se llega a que,

$$(a^{\frac{1}{3}})^{\left\{ \frac{1}{h|\log a|(\log n)^{\frac{1}{2}}} \right\}} \leq \frac{3}{(\log n)^{\frac{3}{2}}}$$

para cada $n \geq 2$.

Por el resultado anterior y la desigualdad (3.30), se tiene,

$$\begin{aligned}
J_2 &\leq \frac{1}{nh} \left(\frac{B_5}{|\log a|(\log n)^{\frac{1}{2}}} + \frac{3B_7}{1-a^{\frac{1}{3}}} \frac{3}{(\log n)^{\frac{3}{2}}} \right) \\
&\leq \left(\frac{B_5}{|\log a|} + \frac{9B_7}{1-a^{\frac{1}{3}}} \right) \frac{(\log n)^{-\frac{1}{2}}}{nh} \tag{3.31}
\end{aligned}$$

Ahora por las desigualdades (3.25), (3.29) y (3.31), se tiene la siguiente cota sobre la varianza de S_n ,

$$\text{Var}[S_n] \leq n \frac{A_1}{n^2 h} + 2 \left(\frac{B_5}{|\log a|} + \frac{9B_7}{1-a^{\frac{1}{3}}} \right) \frac{(\log n)^{-\frac{1}{2}}}{nh} = \mathcal{A}_n \tag{3.32}$$

donde $\mathcal{A}_n = \frac{A_1}{nh} + B_8 \frac{(\log n)^{-\frac{1}{2}}}{nh}$.

Finalmente, como $\mathcal{A}_n \geq \frac{A_1}{nh}$ y por la expresión de \mathcal{B}_n^* dada en (3.23), se tiene que

$$\mathcal{B}_n^* \leq 2(\sqrt{D_2} \vee D_1) L_2 \left(\frac{2^5 n D_2 L_1}{\frac{A_1}{nh}} \vee 1 \right).$$

Usando que para n grande, $M_n \geq 1$, tenemos

$$\sqrt{D_2} \vee D_1 \leq (\sqrt{B_7} \vee B_6) \frac{M_n}{nh}$$

lo que permite obtener que

$$\mathcal{B}_n^* \leq \mathcal{B}_n.$$

Demostración del Corolario 1.

Como $\lambda(u) = (2\mathcal{A}_n u)^{\frac{1}{2}} + \mathcal{B}_n (2u)^3$ entonces,

$$\lambda^2(u) = \left[(2\mathcal{A}_n u)^{\frac{1}{2}} + \mathcal{B}_n (2u)^3 \right]^2 \geq 2u\mathcal{A}_n \quad (3.33)$$

de forma análoga

$$\lambda^{\frac{1}{3}}(u) \geq 2u\mathcal{B}_n^{\frac{1}{3}}$$

Al multiplicar la desigualdad anterior por $\lambda^{\frac{5}{3}}(u)$ se tiene que,

$$\lambda^2(u) \geq 2u\mathcal{B}_n^{\frac{1}{3}}\lambda^{\frac{5}{3}}(u)$$

Ahora, al sumar la desigualdad (3.33) con la desigualdad anterior se tiene que,

$$\lambda^2(u) \geq u \left[\mathcal{A}_n + \mathcal{B}_n^{\frac{1}{3}}\lambda^{\frac{5}{3}}(u) \right]$$

Por la desigualdad anterior y al tomar el cambio de variable $t = \lambda(u)$ se obtiene que,

$$\frac{t^2}{\mathcal{A}_n + \mathcal{B}_n^{\frac{1}{3}}t^{\frac{5}{3}}} = \frac{1}{2} \frac{\lambda^2(u)}{\mathcal{A}_n + \mathcal{B}_n^{\frac{1}{3}}\lambda^{\frac{5}{3}}(u)} \geq \frac{u}{2}$$

Al considerar la desigualdad de Bernstein planteada en la Proposición 15, la desigualdad anterior y el cambio de variables $t = \lambda(u)$ se tiene que,

$$\mathbb{P}(|S_n| \geq \lambda(u)) \leq \exp\left(-\frac{u}{2}\right)$$

para cada $u \geq 0$.

3.5.4. Demostración de la Proposición 9: Sesgo del estimador \hat{g}_h .

$$\begin{aligned}
 \mathbb{E}\{\hat{r}_h(x)\} &= \mathbb{E}\left\{\frac{1}{n}\sum_{i=1}^n K_h(x-X_i)Y_i g^{-1}(X_i)\right\} \\
 &= \mathbb{E}\left\{\frac{1}{n}\sum_{i=1}^n K_h(x-X_i)r(X_i)g^{-1}(X_i)\right\} + \mathbb{E}\left\{\frac{1}{n}\sum_{i=1}^n K_h(x-X_i)\varepsilon_i g^{-1}(X_i)\right\} \\
 &= \mathbb{E}\{K_h(x-X)r(X)g^{-1}(X)\} + \mathbb{E}\{K_h(x-X)g^{-1}(X)\}\mathbb{E}\{\varepsilon\} \\
 &= \int_{\mathbb{R}} K_h(x-t)r(t)dt = K_h * r(x)
 \end{aligned}$$

Así se tiene que,

$$\mathbb{E}[\hat{r}_h(x)] - r(x) = K_h * r(x) - r(x) = \int_{\mathbb{R}} (r(x-hu) - r(x))K(u)du$$

De la fórmula de Taylor-Lagrange al orden l aplicado a $r(x)$ se tiene,

$$r(x-hu) = r(x) - hur'(x) + \frac{(-hu)^2}{2!}r''(x) - \dots + \frac{(-hu)^{l-1}}{(l-1)!}r^{(l-1)}(x) + \frac{(-hu)^l}{l!}r^{(l)}(x - \zeta_u hu)$$

donde $\zeta_u \in [0, 1]$.

Por lo anterior y como K es un núcleo de orden m con $m \geq l$ se tiene que,

$$\mathbb{E}[\hat{r}_h(x)] - r(x) = \int_{\mathbb{R}} \frac{(-hu)^l}{l!}r^{(l)}(x - \zeta_u hu)K(u)du$$

Dado que K es un núcleo de orden l al restar el termino nulo $\int_{\mathbb{R}} \frac{(-uh)^l}{l!}r^{(l)}(x)K(u)du$ a la ecuación anterior, se tiene que,

$$\mathbb{E}[\hat{r}_h(x)] - r(x) = (-1)^l \frac{h^l}{l!} \int_{\mathbb{R}} \left(r^{(l)}(x - \zeta_u hu) - r^{(l)}(x) \right) u^l K(u) du$$

Al tomar valor absoluto en ambos lados y usar la hipótesis $r \in \Sigma(\beta, L)$ se tiene,

$$\begin{aligned}
|\mathbb{E}[\hat{r}_h(x)] - r(x)| &\leq \frac{h^\beta}{l!} L \int_{\mathbb{R}} |u|^\beta \zeta_u^{\beta-l} |K(u)| du \\
&\leq \frac{h^\beta}{l!} L \int_{\mathbb{R}} |u|^\beta |K(u)| du, \quad \text{pues } \zeta_u^{\beta-l} < 1 \\
&\leq h^\beta A_0.
\end{aligned}$$

3.5.5. Demostración de la Proposición 10: Varianza del estimador \hat{g}_h .

Al sustituir Y_i por $r(X_i) + \varepsilon_i$, se obtiene,

$$\begin{aligned}
&\mathbb{E} [(\hat{r}_h(x) - \mathbb{E}[\hat{r}_h(x)])^2] \\
&= \mathbb{E} \left[\left\{ \frac{1}{n} \sum_{i=1}^n K_h(x - X_i) Y_i g^{-1}(X_i) - \mathbb{E} \left(\frac{1}{n} \sum_{i=1}^n K_h(x - X_i) Y_i g^{-1}(X_i) \right) \right\}^2 \right] \\
&= \mathbb{E} \left\{ \frac{1}{n} \sum_{i=1}^n [K_h(x - X_i) r(X_i) g^{-1}(X_i) - \mathbb{E}(K_h(x - X_i) r(X_i) g^{-1}(X_i))] \right. \\
&\quad \left. + \frac{1}{n} \sum_{i=1}^n K_h(x - X_i) \varepsilon_i g^{-1}(X_i) \right\}^2 \\
&= I_1 + I_2 + I_3
\end{aligned} \tag{3.34}$$

donde,

- $A(X_1, \dots, X_n) = \sum_{i=1}^n [K_h(x - X_i) r(X_i) g^{-1}(X_i) - \mathbb{E}(K_h(x - X_i) r(X_i) g^{-1}(X_i))]$
- $I_1 = \mathbb{E} \left\{ \frac{1}{n} \sum_{i=1}^n [K_h(x - X_i) r(X_i) g^{-1}(X_i) - \mathbb{E}(K_h(x - X_i) r(X_i) g^{-1}(X_i))] \right\}^2$
- $I_2 = \mathbb{E} \left\{ 2 \frac{1}{n} A(X_1, \dots, X_n) \frac{1}{n} \sum_{i=1}^n K_h(x - X_i) \varepsilon_i g^{-1}(X_i) \right\}$
- $I_3 = \mathbb{E} \left\{ \frac{1}{n} \sum_{i=1}^n K_h(x - X_i) \varepsilon_i g^{-1}(X_i) \right\}^2$

Se procede a acotar I_1 , I_2 e I_3 .

$$\begin{aligned}
I_3 &= \frac{1}{n^2} \mathbb{E} \left\{ \sum_{i=1}^n K_h(x - X_i) \varepsilon_i g^{-1}(X_i) \right\}^2 \\
&= \frac{1}{n^2} \sum_{i=1}^n \mathbb{E} \{ K_h(x - X_i) \varepsilon_i g^{-1}(X_i) \}^2 \\
&\quad + \frac{1}{n^2} \sum_{i \neq j} \mathbb{E} \{ K_h(x - X_i) \varepsilon_i g^{-1}(X_i) K_h(x - X_j) \varepsilon_j g^{-1}(X_j) \} \\
&= \frac{1}{n^2} \sum_{i=1}^n \frac{B_1}{h} + \frac{1}{n^2} \sum_{i \neq j} \mathbb{E} \{ K_h(x - X_i) g^{-1}(X_i) K_h(x - X_j) g^{-1}(X_j) \} \mathbb{E}(\varepsilon_i \varepsilon_j) \\
&= \frac{B_1}{nh}
\end{aligned} \tag{3.35}$$

por la primera desigualdad del Lema 1.

$$\begin{aligned}
I_2 &= \frac{2}{n^2} \mathbb{E} \left\{ A(X_1, \dots, X_n) \sum_{i=1}^n K_h(x - X_i) \varepsilon_i g^{-1}(X_i) \right\} \\
&= \frac{2}{n^2} \sum_{i=1}^n \mathbb{E} \{ A(X_1, \dots, X_n) K_h(x - X_i) g^{-1}(X_i) \} \mathbb{E} \{ \varepsilon_i \} = 0
\end{aligned} \tag{3.36}$$

$$\begin{aligned}
I_1 &= \frac{1}{n^2} \mathbb{E} \left\{ \sum_{i=1}^n [K_h(x - X_i) r(X_i) g^{-1}(X_i) - \mathbb{E}(K_h(x - X_i) r(X_i) g^{-1}(X_i))] \right\}^2 \\
&\leq I_4 + I_5
\end{aligned} \tag{3.37}$$

donde,

- $I_4 = \frac{1}{n^2} \sum_{i=1}^n \text{Var} [K_h(x - X_i) r(X_i) g^{-1}(X_i)]$
- $I_5 = \frac{1}{n^2} \sum_{i \neq j} |\text{Cov} [K_h(x - X_i) r(X_i) g^{-1}(X_i), K_h(x - X_j) r(X_j) g^{-1}(X_j)]|$

Se procede a acotar I_4 e I_5 . Por la segunda desigualdad del Lema 1

$$I_4 \leq \frac{1}{n^2} \sum_{i=1}^n \mathbb{E} \left[(K_h(x - X_i) r(X_i) g^{-1}(X_i))^2 \right] \leq \frac{B_2}{nh} \tag{3.38}$$

$$\begin{aligned}
I_5 &= \frac{2}{n^2} \sum_{i=1}^{n-1} \sum_{k=1}^{n-i} |\text{Cov}(\tilde{X}_i, \tilde{X}_{i+k})| \\
&= \frac{2}{n^2} \sum_{i=1}^{n-1} \left(\sum_{k=1}^{u_n} |\text{Cov}(\tilde{X}_i, \tilde{X}_{i+k})| + \sum_{k=u_n+1}^{n-i} |\text{Cov}(\tilde{X}_i, \tilde{X}_{i+k})| \right) \tag{3.39}
\end{aligned}$$

donde $1 \leq u_n \leq n-1$ y $\tilde{X}_i = K_h(x - X_i)r(X_i)g^{-1}(X_i)$, $\forall i$.

Tenemos por la tercera y cuarta desigualdad del Lema 1.

$$\begin{aligned}
&\sum_{k=1}^{u_n} |\text{Cov}(\tilde{X}_i, \tilde{X}_{i+k})| \\
&\leq \sum_{k=1}^{u_n} |\mathbb{E}[\tilde{X}_i \tilde{X}_{i+k}]| + |\mathbb{E}[\tilde{X}_i] \mathbb{E}[\tilde{X}_{i+k}]| \\
&\leq \sum_{k=1}^{u_n} \left\{ \mathbb{E}[|K_h(x - X_i)r(X_i)g^{-1}(X_i)K_h(x - X_{i+k})r(X_{i+k})g^{-1}(X_{i+k})|] \right. \\
&\quad \left. + (\mathbb{E}[|\tilde{X}_i|])^2 \right\} \\
&\leq \sum_{k=1}^{u_n} \left\{ (r_{\text{sup}})^2 (g_{\text{inf}})^{-2} \mathbb{E}[|K_h(x - X_i)K_h(x - X_{i+k})|] + (r_{\text{sup}} \|K\|_1)^2 \right\} \\
&= \sum_{k=1}^{u_n} \left\{ (r_{\text{sup}})^2 (g_{\text{inf}})^{-2} Q \|K\|_1^2 + (r_{\text{sup}})^2 \|K\|_1^2 \right\} = A_3 u_n, \tag{3.40}
\end{aligned}$$

Para acotar el término $\sum_{k=u_n+1}^{n-i} |\text{Cov}(\tilde{X}_i, \tilde{X}_{i+k})|$ de la igualdad (3.39) se usa que los X_i satisfacen (H_3) .

Como

$$\begin{aligned}
|\tilde{X}_i| &= |K_h(x - X_i)r(X_i)g^{-1}(X_i)| \\
&\leq r_{\text{sup}}(g_{\text{inf}})^{-1} \left| \frac{1}{h} K \left(\frac{x - X_i}{h} \right) \right| \\
&\leq \frac{r_{\text{sup}}(g_{\text{inf}})^{-1} \|K\|_{\infty}}{h}
\end{aligned}$$

Por la desigualdad anterior, la ecuación 3.1 y la hipótesis (H_3) ,

$$|\text{cov}(\tilde{X}_i, \tilde{X}_{i+k})| \leq 4a^k \frac{(r_{\text{sup}})^2 (g_{\text{inf}})^{-2} \|K\|_{\infty}^2}{h^2}$$

Por esta última desigualdad se tiene que,

$$\begin{aligned}
\sum_{k=u_n+1}^{n-i} |Cov(\tilde{X}_i, \tilde{X}_{i+k})| &\leq \frac{4(r_{\text{sup}})^2(g_{\text{inf}})^{-2}\|K\|_{\infty}^2}{h^2} \sum_{k=u_n+1}^{\infty} a^k \\
&= \frac{4(r_{\text{sup}})^2(g_{\text{inf}})^{-2}\|K\|_{\infty}^2 a^{u_n+1}}{h^2 (1-a)} \\
&= A_4 \frac{a^{u_n+1}}{h^2}.
\end{aligned} \tag{3.41}$$

Por la ecuación (3.39) y las desigualdades (3.40) y (3.41) se tiene,

$$\begin{aligned}
I_5 &\leq \frac{2}{n^2} \sum_{i=1}^{n-1} \left(A_3 u_n + A_4 \frac{a^{u_n+1}}{h^2} \right) \\
&\leq \frac{1}{n} \left(2A_3 u_n + 2A_4 \frac{a^{u_n+1}}{h^2} \right)
\end{aligned}$$

La variable u_n en las sumatorias de la ecuación (3.39) se toma como $u_n = \left\lceil \frac{1}{h|\log a|(\log n)^{\frac{1}{2}}} \right\rceil$ y se usan los hechos de que $[x] \leq x \leq [x] + 1$ y $0 < a < 1$, para seguir acotando la desigualdad anterior,

$$\begin{aligned}
I_5 &\leq \frac{1}{n} \left(2A_3 \frac{1}{h|\log a|(\log n)^{\frac{1}{2}}} + 2A_4 \frac{a^{\frac{1}{h|\log a|(\log n)^{\frac{1}{2}}}}}{h^2} \right) \\
&= \frac{1}{nh} \left(2A_3 \frac{1}{|\log a|(\log n)^{\frac{1}{2}}} + 2A_4 \frac{1}{h} a^{\frac{1}{h|\log a|(\log n)^{\frac{1}{2}}}} \right)
\end{aligned}$$

Por el Lema 2 y la desigualdad anterior se tiene que, para cada $n \geq 4$,

$$\begin{aligned}
I_5 &\leq \frac{1}{nh} \left(\frac{2A_3}{|\log a|} \frac{1}{(\log n)^{\frac{1}{2}}} + 2A_4 \frac{10}{(\log n)^{\frac{5}{2}}} \right) \\
&\leq \left(\frac{2A_3}{|\log a|} + 20A_4 \right) \frac{(\log n)^{-\frac{1}{2}}}{nh}.
\end{aligned} \tag{3.42}$$

Por las igualdades (3.37), (3.38) y (3.42), para todo $n \geq 4$,

$$I_1 \leq \frac{B_2}{nh} + \left(\frac{2A_3}{|\log a|} + 20A_4 \right) \frac{(\log n)^{-\frac{1}{2}}}{nh}. \tag{3.43}$$

Finalmente, por la ecuación (3.34) y las desigualdades (3.35), (3.36) y (3.43) se tiene que, para cada $n \geq 4$,

$$\mathbb{E} [(\hat{r}_h(x) - \mathbb{E}[\hat{r}_h(x)])^2] = I_1 + I_3 \leq \frac{A_1}{nh} + A_2 \frac{(\log n)^{-\frac{1}{2}}}{nh}.$$

3.5.6. Demostración de la Proposición 11.

Sea $h \in \mathcal{H}$. Tenemos

$$|r(x) - \hat{r}_{\hat{h}}(x)| \leq |r(x) - \hat{r}_h(x)| + |\hat{r}_h(x) - \hat{r}_{h,\hat{h}}(x)| + |\hat{r}_{\hat{h}}(x) - \hat{r}_{h,\hat{h}}(x)|.$$

Al sumar y restar $V(h)$ y $V(\hat{h})$, tenemos

$$\begin{aligned} |r(x) - \hat{r}_{\hat{h}}(x)| &\leq |r(x) - \hat{r}_h(x)| + A(h, x) + V(h) + A(\hat{h}, x) + V(\hat{h}) \\ &\leq |r(x) - \hat{r}_h(x)| + 2(A(h, x) + V(h)). \end{aligned} \quad (3.44)$$

Por otro lado,

$$\begin{aligned} |\hat{r}_{h,h'}(x) - \hat{r}_{h'}(x)| &\leq |\hat{r}_{h,h'}(x) - \mathbb{E}(\hat{r}_{h,h'}(x))| + |\hat{r}_{h'}(x) - \mathbb{E}(\hat{r}_{h'}(x))| \\ &\quad + |\mathbb{E}(\hat{r}_{h,h'}(x)) - \mathbb{E}(\hat{r}_{h'}(x))| \end{aligned}$$

y si restamos $V(h')$ en ambos miembros y tomamos parte positiva tenemos,

$$\begin{aligned} &\{|\hat{r}_{h,h'}(x) - \hat{r}_{h'}(x)| - V(h')\}_+ \\ &\leq \{|\hat{r}_{h,h'}(x) - \mathbb{E}(\hat{r}_{h,h'}(x))| + |\hat{r}_{h'}(x) - \mathbb{E}(\hat{r}_{h'}(x))| - V(h')\}_+ \\ &\quad + |\mathbb{E}(\hat{r}_{h,h'}(x)) - \mathbb{E}(\hat{r}_{h'}(x))| \end{aligned} \quad (3.45)$$

Recuerde que la esperanza del estimador es $\mathbb{E}(\hat{r}_h(x)) = (K_h * r)(x)$, por lo cual el segundo término del lado derecho de la desigualdad anterior, se expresa como

$$\begin{aligned} |\mathbb{E}(\hat{r}_{h,h'}(x)) - \mathbb{E}(\hat{r}_{h'}(x))| &= |(K_{h'} * r)(x) - (K_{h'} * (K_h * r))(x)| \\ &= |(K_{h'} * (r - (K_h * r)))(x)| \\ &\leq \int |K_{h'}(x-u)(K_h * r(u) - r(u))| du. \end{aligned}$$

Como el soporte de K es $[-1, 1]$, se tiene que $\frac{x-u}{h'} \in [-1, 1]$, de donde $u \in B(x)$, lo que implica

$$\begin{aligned} |\mathbb{E}(\hat{r}_{h,h'}(x)) - \mathbb{E}(\hat{r}_{h'}(x))| &\leq \int_{B(x)} \left| \frac{1}{h'} K \left(\frac{x-u}{h'} \right) \right| |(K_h * r(u) - r(u))| du \\ &\leq \max_{u \in B(x)} |(K_h * r(u) - r(u))| \int \left| \frac{1}{h'} K \left(\frac{x-u}{h'} \right) \right| du \\ &\leq \|K\|_1 \max_{u \in B(x)} |(K_h * r(u) - r(u))| \\ &\leq C(h) \|K\|_1. \end{aligned}$$

Usando que $V(h') = V_1(h') + V_2(h')$ y la desigualdad anterior se tiene que,

$$\begin{aligned} &\{|\hat{r}_{h,h'}(x) - \hat{r}_{h'}(x)| - V(h')\}_+ \\ &\leq \{|\hat{r}_{h,h'}(x) - \mathbb{E}(\hat{r}_{h,h'}(x))| - V_2(h')\}_+ + \{|\hat{r}_{h'}(x) - \mathbb{E}(\hat{r}_{h'}(x))| - V_1(h')\}_+ \\ &+ C(h) \|K\|_1. \end{aligned} \tag{3.46}$$

Por la definición de $A(h, x)$ y al tomar en la desigualdad (3.46) máximo sobre h' en \mathcal{H} , se tiene que,

$$A(h, x) \leq T_1 + T_2 + C(h) \|K\|_1. \tag{3.47}$$

Al tomar la norma $\mathbb{E}[\cdot]^{\frac{1}{2}}$ determinada por la esperanza en las ecuaciones (3.44) y (3.47) se obtiene

$$\begin{aligned} &\left(\mathbb{E} (r(x) - \hat{r}_h(x))^2 \right)^{\frac{1}{2}} \\ &\leq \left(\mathbb{E} (r(x) - \hat{r}_h(x))^2 \right)^{\frac{1}{2}} + 2 \left(\left(\mathbb{E} (A(h, x))^2 \right)^{\frac{1}{2}} + V(h) \right) \\ &\leq \left(\mathbb{E} (r(x) - \hat{r}_h(x))^2 \right)^{\frac{1}{2}} + 2 \left(C(h) \|K\|_1 + V(h) + \left(\mathbb{E} (T_1^2) \right)^{\frac{1}{2}} + \left(\mathbb{E} (T_2^2) \right)^{\frac{1}{2}} \right). \end{aligned}$$

3.5.7. Demostración de la Proposición 12, (i).

Vamos a demostrar que para n suficientemente grande

$$\mathbb{E}(T_1^2) \leq A_5 \frac{\log n}{n}$$

donde $A_5 = B_9 + 2B_{10}$, con $B_9 = (g_{\inf})^{-2} \|K\|_\infty^2 \{8((r_{\sup})^4 + 3\sigma^4)\}^{\frac{1}{2}} \left(\frac{2}{\sqrt{2\pi}}\right)^{\frac{1}{2}}$, y $B_{10} = \frac{2^4 3}{1-e^{-\frac{\delta}{2}}}(A_1 + C_6) + \frac{5! 2^{18} 3 B^2}{1-e^{-\frac{\delta}{2}}}(\sigma + r_{\sup})^2$. Para estudiar $\mathbb{E}(T_1^2)$ se usa el siguiente hecho,

$$\begin{aligned} \mathbb{E}(T_1^2) &\leq \max_{h' \in \mathcal{H}} \mathbb{E} \left[\{|\hat{r}_{h'}(x) - \mathbb{E}(\hat{r}_{h'}(x))| - V_1(h')\}_+^2 \right] \\ &\leq \sum_{h' \in \mathcal{H}} \mathbb{E} \left[\{|\hat{r}_{h'}(x) - \mathbb{E}(\hat{r}_{h'}(x))| - V_1(h')\}_+^2 \right] \end{aligned} \quad (3.48)$$

En este caso no se tiene garantía de que el término dentro de la sumatoria en el lado derecho de la desigualdad (3.48) sea acotado, lo que impide que se pueda aplicar la desigualdad de Bernstein para acotarlo. Para solucionar este inconveniente se define el estimador auxiliar de truncamiento de la siguiente manera.

$$\tilde{r}_h(x) = \frac{1}{n} \sum_{k=1}^n Y_k K_h(x - X_k) g^{-1}(X_k) \mathbf{1}_{\{|Y_k| \leq M_n\}}$$

para $h \in \mathcal{H}$ con $M_n > 0, \forall n \in \mathbb{N}$.

Se descompone en dos partes el término $\{|\hat{r}_{h'}(x) - \mathbb{E}(\hat{r}_{h'}(x))| - V_1(h')\}_+^2$ usando el estimador auxiliar de truncamiento y su esperanza, como se muestra a continuación.

$$\hat{r}_{h'}(x) - \mathbb{E}(\hat{r}_{h'}(x)) = \tilde{r}_{h'}(x) - \mathbb{E}(\tilde{r}_{h'}(x)) + \{\hat{r}_{h'}(x) - \tilde{r}_{h'}(x) - \mathbb{E}(\hat{r}_{h'}(x) - \tilde{r}_{h'}(x))\}$$

así se tiene que,

$$|\hat{r}_{h'}(x) - \mathbb{E}(\hat{r}_{h'}(x))| \leq |\tilde{r}_{h'}(x) - \mathbb{E}(\tilde{r}_{h'}(x))| + |\hat{r}_{h'}(x) - \tilde{r}_{h'}(x) - \mathbb{E}(\hat{r}_{h'}(x) - \tilde{r}_{h'}(x))|$$

por lo cual,

$$\begin{aligned} \{|\hat{r}_{h'}(x) - \mathbb{E}(\hat{r}_{h'}(x))| - V_1(h')\}_+^2 &\leq 2\{|\tilde{r}_{h'}(x) - \mathbb{E}(\tilde{r}_{h'}(x))| - V_1(h')\}_+^2 \\ &\quad + 2\{\hat{r}_{h'}(x) - \tilde{r}_{h'}(x) - \mathbb{E}(\hat{r}_{h'}(x) - \tilde{r}_{h'}(x))\}^2 \end{aligned}$$

al aplicar esperanza queda,

$$\mathbb{E}(\{|\hat{r}_{h'}(x) - \mathbb{E}(\hat{r}_{h'}(x))| - V_1(h')\}_+^2) \leq 2\Delta_1(h') + 2\Delta_2(h')$$

donde

$$\begin{aligned}\Delta_1(h') &= \mathbb{E} \left(\{ |\tilde{r}_{h'}(x) - \mathbb{E}(\tilde{r}_{h'}(x))| - V_1(h') \}_+^2 \right), \\ \Delta_2(h') &= \mathbb{E} \left(\{ \hat{r}_{h'}(x) - \tilde{r}_{h'}(x) - \mathbb{E}(\hat{r}_{h'}(x) - \tilde{r}_{h'}(x)) \}^2 \right)\end{aligned}$$

Al sustituir el resultado anterior en la desigualdad (3.48) se tiene que,

$$\mathbb{E}(T_1^2) \leq 2 \left(\sum_{h' \in \mathcal{H}} \Delta_1(h') + \sum_{h' \in \mathcal{H}} \Delta_2(h') \right) \quad (3.49)$$

Se procede a acotar la sumatoria $\sum_{h' \in \mathcal{H}} \Delta_2(h')$, para tal fin se analiza el término $\hat{r}_{h'}(x) - \tilde{r}_{h'}(x)$,

$$\begin{aligned}\hat{r}_{h'}(x) - \tilde{r}_{h'}(x) &= \frac{1}{n} \sum_{k=1}^n Y_k K_{h'}(x - X_k) g^{-1}(X_k) - \frac{1}{n} \sum_{k=1}^n Y_k K_{h'}(x - X_k) g^{-1}(X_k) \mathbf{1}_{\{|Y_k| \leq M_n\}} \\ &= \frac{1}{n} \sum_{k=1}^n \eta_k\end{aligned}$$

donde $\eta_k = Y_k K_{h'}(x - X_k) g^{-1}(X_k) \mathbf{1}_{\{|Y_k| > M_n\}}$.

Entonces, tenemos

$$\Delta_2(h') = \frac{1}{n^2} \text{var} \left(\sum_{k=1}^n \eta_k \right) \leq \frac{1}{n^2} \mathbb{E} \left(\left\{ \sum_{k=1}^n \eta_k \right\}^2 \right) \leq \frac{1}{n^2} \mathbb{E} \left(n \sum_{k=1}^n \eta_k^2 \right) \leq \mathbb{E}(\eta_1^2)$$

y

$$\begin{aligned}\mathbb{E}(\eta_1^2) &= \frac{1}{(h')^2} \mathbb{E} \left\{ Y^2 K^2 \left(\frac{x - X}{h'} \right) g^{-2}(X) \mathbf{1}_{\{|Y| > M_n\}} \right\} \\ &\leq \frac{(g_{\inf})^{-2} \|K\|_\infty^2}{(h')^2} \mathbb{E} \{ (Y^2)^2 \}^{\frac{1}{2}} \mathbb{E} \{ \mathbf{1}_{\{|Y| > M_n\}}^2 \}^{\frac{1}{2}} \\ &\leq \frac{(g_{\inf})^{-2} \|K\|_\infty^2}{(h')^2} \mathbb{E} \{ Y^4 \}^{\frac{1}{2}} \mathbb{P} \{ \{|Y| > M_n\} \}^{\frac{1}{2}},\end{aligned}$$

con

$$\mathbb{E}\{Y^4\} \leq 8\mathbb{E}\{r^4(X) + \varepsilon^4\} \leq 8((r_{sup})^4 + \mathbb{E}\{\varepsilon^4\}) \leq 8((r_{sup})^4 + 3\sigma^4)$$

y

$$\begin{aligned} \mathbb{P}(\{|Y| > M_n\}) &= \mathbb{P}(\{|r(X) + \varepsilon| > M_n\}) \\ &\leq \mathbb{P}(\{r_{sup} + |\varepsilon| > M_n\}) \\ &\leq \mathbb{P}(\{|\varepsilon| > M_n - r_{sup}\}) \\ &\leq 2\mathbb{P}\left(\left\{\frac{\varepsilon}{\sigma} > \frac{M_n - r_{sup}}{\sigma}\right\}\right) \\ &\leq 2\frac{\phi(\log n)}{\log n} \end{aligned}$$

donde $\phi(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$.

Esto implica que

$$\Delta_2(h') \leq \frac{B_9}{(h')^2} \frac{e^{-\frac{(\log n)^2}{4}}}{(\log n)^{\frac{1}{2}}}$$

y por lo tanto,

$$\sum_{h' \in \mathcal{H}} \Delta_2(h') \leq B_9 \frac{e^{-\frac{(\log n)^2}{4}}}{(\log n)^{\frac{1}{2}}} \sum_{h' \in \mathcal{H}} \frac{1}{(h')^2}.$$

Como la familia de anchos de ventanas es $\mathcal{H} = \{e^{-i}\}_{i=0}^M \cap [h_{\min}, h_{\max}]$ con $h_{\min} = \frac{(\log n)^8}{n}$, $h_{\max} = \frac{1}{(\log n)^2}$ y $M = \left\lceil \log \left(\frac{n}{(\log n)^8} \right) \right\rceil$, entonces tenemos

$$\sum_{h' \in \mathcal{H}} \frac{1}{(h')^2} \leq \sum_{i=0}^M e^{2i} \leq \int_0^M e^{2x} dx \leq \frac{e^{2M}}{2} \leq \frac{n^2}{2}.$$

Usando

$$\frac{e^{-\frac{(\log n)^2}{4}}}{(\log n)^{\frac{1}{2}}} n^2 = \frac{e^{-\frac{(\log n)^2}{4}}}{(\log n)^{\frac{1}{2}}} e^{2 \log n} = \frac{e^{-\log n \left(\frac{\log n}{4} - 2 \right)}}{(\log n)^{\frac{1}{2}}} \leq \frac{e^{-\log n}}{(\log n)^{\frac{1}{2}}} \leq \frac{1}{n}$$

obtenemos finalmente que

$$\sum_{h' \in \mathcal{H}} \Delta_2(h') \leq \frac{B_9}{2n}. \quad (3.50)$$

Ahora se procede a acotar la sumatoria $\sum_{h' \in \mathcal{H}} \Delta_1(h')$ de la desigualdad (3.49), donde el término general de la sumatoria $\Delta_1(h') = \mathbb{E}(\{|\tilde{r}_{h'}(x) - \mathbb{E}(\tilde{r}_{h'}(x))| - V_1(h')\}_+^2)$ esta en térmi-

nos del estimador auxiliar de truncamiento, lo que garantiza que tal término es acotado.

Se analiza el término $\Delta_1(h')$, como se muestra a continuación.

$$\begin{aligned}
 \Delta_1(h') &= \mathbb{E} \left(\left\{ |\tilde{r}_{h'}(x) - \mathbb{E}(\tilde{r}_{h'}(x))| - V_1(h') \right\}_+^2 \right) \\
 &= \int_0^{+\infty} \mathbb{P} \left(\left\{ |\tilde{r}_{h'}(x) - \mathbb{E}(\tilde{r}_{h'}(x))| - V_1(h') \right\}_+^2 > t \right) dt \\
 &= \int_0^{+\infty} \mathbb{P} \left(\left\{ |\tilde{r}_{h'}(x) - \mathbb{E}(\tilde{r}_{h'}(x))| - V_1(h') \right\}_+ > t^{1/2} \right) dt \\
 &= \int_0^{+\infty} \mathbb{P} \left(|\tilde{r}_{h'}(x) - \mathbb{E}(\tilde{r}_{h'}(x))| > V_1(h') + t^{1/2} \right) dt \tag{3.51}
 \end{aligned}$$

pues para $t > 0$, $\mathbb{P}(\{T\}_+ > t) = \mathbb{P}(T > t)$ ya que $\mathbb{P}(T > t \wedge T < 0) = 0$.

Del integrando $\mathbb{P}(|\tilde{r}_{h'}(x) - \mathbb{E}(\tilde{r}_{h'}(x))| > V_1(h') + t^{1/2})$ de la ecuación (3.51), se observa que,

$$\tilde{r}_{h'}(x) - \mathbb{E}(\tilde{r}_{h'}(x)) = \sum_{i=1}^n (G(X_i, \varepsilon_i) - \mathbb{E}[G(X_i, \varepsilon_i)]) = \sum_{i=1}^n Z_i = S_n$$

donde los Z_i y S_n están definidos en la Proposición 15. Entonces se satisface la desigualdad (3.12) donde $\mathcal{A}_n = \frac{A_1}{nh'} + B_8 \frac{(\log n)^{-\frac{1}{2}}}{nh'}$, es una cota superior de $\text{Var}(S_n)$ y $\mathcal{B}_n = B \frac{M_n}{nh'}$.

En este punto se detiene el análisis de $\Delta_1(h')$, para mostrar que al tomar $s(u) = \sqrt{2\mathcal{A}_n u} + 4\mathcal{B}_n(2u)^3$ para cada $u \in \mathbb{R}^+$, con $\mathcal{A}_n = \frac{A_1}{nh'} + B_8 \frac{(\log n)^{-\frac{1}{2}}}{nh'}$ y $\mathcal{B}_n = B \frac{M_n}{nh'}$, se cumple que $s(\gamma |\log h'|) \leq V_1(h')$ para n suficientemente grande.

$$\begin{aligned}
s(\gamma|\log h'|) &= \sqrt{2 \left(\frac{A_1}{nh'} + B_8 \frac{(\log n)^{-\frac{1}{2}}}{nh'} \right) \gamma|\log h'| + 4B \frac{M_n}{nh'} (2\gamma|\log h'|)^3} \\
&\leq \sqrt{2 \left(\frac{A_1}{nh'} + B_8 \frac{(\log n)^{-\frac{1}{2}}}{nh'} \right) \gamma \log n + 4(2\gamma)^3 B (\log n)^3 \frac{M_n}{nh'}} \\
&\leq \sqrt{2\gamma A_1 \frac{\log n}{nh'}} + \sqrt{2\gamma B_8 \frac{(\log n)^{\frac{1}{2}}}{nh'}} + 4(2\gamma)^3 B (\log n)^3 (\sigma + r_{\text{sup}}) \frac{\log n}{nh'} \\
&= \sqrt{2\gamma A_1 \frac{\log n}{nh'}} \left(1 + \sqrt{\frac{B_8}{A_1} (\log n)^{-\frac{1}{2}}} \right) + 4(2\gamma)^3 B (\sigma + r_{\text{sup}}) \frac{(\log n)^4}{nh'}
\end{aligned} \tag{3.52}$$

Como $\frac{(\log n)^8}{n} \leq h'$ entonces $\frac{(\log n)^8}{nh'} \leq 1$, de donde $\frac{(\log n)^4}{\sqrt{nh'}} \leq 1$. De esto último se deduce la siguiente desigualdad,

$$4(2\gamma)^3 B (\sigma + r_{\text{sup}}) \frac{(\log n)^4}{nh'} = \frac{(\log n)^4}{\sqrt{nh'}} \frac{4(2\gamma)^3 B (\sigma + r_{\text{sup}})}{\sqrt{nh'}} \leq \frac{4(2\gamma)^3 B (\sigma + r_{\text{sup}})}{\sqrt{nh'}}$$

Por esta última desigualdad y la desigualdad (3.52), se tiene,

$$\begin{aligned}
s(\gamma|\log h'|) &\leq \sqrt{2\gamma A_1 \frac{\log n}{nh'}} \left(1 + \sqrt{\frac{B_8}{A_1} (\log n)^{-\frac{1}{2}}} \right) + \frac{4(2\gamma)^3 B (\sigma + r_{\text{sup}})}{\sqrt{nh'}} \\
&= \sqrt{2\gamma A_1 \frac{\log n}{nh'}} \left(1 + \sqrt{\frac{B_8}{A_1} (\log n)^{-\frac{1}{4}}} + \frac{4(2\gamma)^3 B (\sigma + r_{\text{sup}})}{\sqrt{2\gamma A_1}} (\log n)^{-\frac{1}{2}} \right)
\end{aligned} \tag{3.53}$$

Como $\delta_n = (\log n)^{-\frac{1}{5}}$, es claro que $(\log n)^{-\frac{1}{2}} \leq (\log n)^{-\frac{1}{4}} \leq \delta_n$ para $n \geq 3$, por lo cuál $\frac{(\log n)^{-\frac{1}{2}}}{\delta_n} \searrow 0$ y $\frac{(\log n)^{-\frac{1}{4}}}{\delta_n} \searrow 0$. Por otra parte, para cualquier par de constantes α y β en \mathbb{R}^+ se cumple que,

$$0 \leq \alpha \frac{(\log n)^{-\frac{1}{2}}}{\delta_n} \leq \frac{1}{2} \quad \text{y} \quad 0 \leq \beta \frac{(\log n)^{-\frac{1}{4}}}{\delta_n} \leq \frac{1}{2}$$

para todo n suficientemente grande. Por lo cuál,

$$0 \leq \alpha (\log n)^{-\frac{1}{2}} \leq \frac{\delta_n}{2} \quad \text{y} \quad 0 \leq \beta (\log n)^{-\frac{1}{4}} \leq \frac{\delta_n}{2}$$

Ahora, por este último resultado y por la desigualdad (3.53) se tiene, para n suficientemente grande,

$$\begin{aligned}
 s(\gamma|\log h'|) &\leq \sqrt{2\gamma A_1 \frac{\log n}{nh'}} \left(1 + \frac{\delta_n}{2} + \frac{\delta_n}{2}\right) \\
 &= \sqrt{2\gamma A_1 \frac{\log n}{nh'}} (1 + \delta_n) \\
 &= V_1(h').
 \end{aligned} \tag{3.54}$$

Se toman las funciones $\lambda(u) = \sqrt{2\mathcal{A}_n u} + \mathcal{B}_n(2u)^3$, del Corolario 1, y $s(u) = \sqrt{2\mathcal{A}_n u} + 4\mathcal{B}_n(2u)^3$ para cada $u \in \mathbb{R}^+$. Al realizar en la ecuación (3.51) el cambio de variable $t^{\frac{1}{2}} = s(u)$, de donde $t = s^2(u)$ y $dt = 2s(u)s'(u)du$, se tiene que,

$$\Delta_1(h') = \int_0^{+\infty} \mathbb{P}(|S_n| > V_1(h') + s(u)) 2s(u)s'(u)du$$

Por la ecuación anterior y la desigualdad (3.54), se tiene que,

$$\Delta_1(h') \leq \int_0^{+\infty} \mathbb{P}(|S_n| > s(\gamma|\log h'|) + s(u)) 2s(u)s'(u)du \tag{3.55}$$

Como la función raíz cuadrada es subaditiva y $(p+q)^3 \leq 4(p^3+q^3)$, se tiene que,

$$\begin{aligned}
 s(p) + s(q) &= \sqrt{2\mathcal{A}_n p} + 4\mathcal{B}_n(2p)^3 + \sqrt{2\mathcal{A}_n q} + 4\mathcal{B}_n(2q)^3 \\
 &\geq \sqrt{2\mathcal{A}_n(p+q)} + \mathcal{B}_n(2(p+q))^3 \\
 &= \lambda(p+q)
 \end{aligned}$$

Por la desigualdad (3.55), este último resultado y el Corolario 1, se tiene,

$$\begin{aligned}
\Delta_1(h') &\leq \int_0^{+\infty} \mathbb{P}(|S_n| > \lambda(\gamma|\log h'| + u)) 2s(u)s'(u)du \\
&\leq \int_0^{+\infty} \exp\left(-\frac{\gamma|\log h'| + u}{2}\right) 2s(u)s'(u)du \\
&= \exp\left(-\frac{\gamma|\log h'|}{2}\right) \int_0^{+\infty} \exp\left(-\frac{u}{2}\right) 2s(u)s'(u)du \tag{3.56}
\end{aligned}$$

- Como $s(u) = \sqrt{2\mathcal{A}_n}\sqrt{u} + 4\mathcal{B}_n2^3u^3$ y $s'(u) = \sqrt{2\mathcal{A}_n}\frac{1}{2\sqrt{u}} + 4\mathcal{B}_n2^33u^2$ entonces $s'(u)u \leq 3s(u)$, por lo cuál $2s'(u)s(u) \leq 6\frac{s^2(u)}{u}$.
- Como $s^2(u) = [\sqrt{2\mathcal{A}_n u} + 4\mathcal{B}_n(2u)^3]^2 \leq 2[2\mathcal{A}_n u + 4^2\mathcal{B}_n^2(2u)^6]$ entonces $\frac{s^2(u)}{u} \leq 2^2\mathcal{A}_n + 2^{11}\mathcal{B}_n^2u^5$.

Por estos dos últimos resultados y la desigualdad (3.56), se tiene,

$$\begin{aligned}
\Delta_1(h') &\leq \exp\left(-\frac{\gamma|\log h'|}{2}\right) \int_0^{+\infty} 6(2^2\mathcal{A}_n + 2^{11}\mathcal{B}_n^2u^5) \exp\left(-\frac{u}{2}\right) du \\
&= 3e^{\left\{\frac{\gamma}{2}\log h'\right\}} \left\{ \int_0^\infty 2^3\mathcal{A}_n e^{-\frac{u}{2}} du + \int_0^\infty 2^{12}\mathcal{B}_n^2u^5 e^{-\frac{u}{2}} du \right\} \\
&= 3e^{\left\{\log(h'^{\frac{\gamma}{2}})\right\}} \left\{ 2^3\mathcal{A}_n \lim_{b \rightarrow \infty} \left[-2e^{-\frac{u}{2}}\right]_0^b + 2^{13}\mathcal{B}_n^2 \int_0^\infty (2w)^5 e^{-w} dw \right\} \\
&= 3(h')^{\frac{\gamma}{2}} \left\{ 2^4\mathcal{A}_n + 2^{18}\mathcal{B}_n^2 \int_0^\infty w^5 e^{-w} dw \right\} \\
&= 3 \{2^4\mathcal{A}_n + 5!2^{18}\mathcal{B}_n^2\} (h')^{\frac{\gamma}{2}}
\end{aligned}$$

Al sumar en la desigualdad anterior sobre cada $h' \in \mathcal{H}$, sustituir las expresiones de \mathcal{A}_n y \mathcal{B}_n , dadas en la Proposición 15 y usar que $M_n \leq (\sigma + r_{\text{sup}}) \log n$ y $\frac{1}{nh'} \leq \frac{1}{\log n}$ se tiene,

$$\begin{aligned}
& \sum_{h' \in \mathcal{H}} \Delta_1(h') \\
& \leq \sum_{h' \in \mathcal{H}} 3 \left\{ 2^4 \left[\frac{A_1}{nh'} + B_8 \frac{(\log n)^{-\frac{1}{2}}}{nh'} \right] + 5!2^{18} \left[B \frac{M_n}{nh'} \right]^2 \right\} (h')^{\frac{\gamma}{2}} \\
& \leq \sum_{h' \in \mathcal{H}} \left\{ 2^4 3 \left[\frac{A_1}{n} + B_8 \frac{(\log n)^{-\frac{1}{2}}}{n} \right] (h')^{\frac{\gamma}{2}-1} + 5!2^{18} 3 \frac{B^2(\sigma + r_{\text{sup}})^2 (\log n)^2}{nh'} (h')^{\frac{\gamma}{2}} \right\} \\
& \leq \sum_{h' \in \mathcal{H}} \left\{ 2^4 3 \left[\frac{A_1}{n} + B_8 \frac{(\log n)^{-\frac{1}{2}}}{n} \right] (h')^{\frac{\gamma}{2}-1} + 5!2^{18} 3 B^2 (\sigma + r_{\text{sup}})^2 \frac{\log n}{n} (h')^{\frac{\gamma}{2}-1} \right\} \\
& = \left\{ 2^4 3 \left[\frac{A_1}{n} + B_8 \frac{(\log n)^{-\frac{1}{2}}}{n} \right] + 5!2^{18} 3 B^2 (\sigma + r_{\text{sup}})^2 \frac{\log n}{n} \right\} \sum_{h' \in \mathcal{H}} (h')^{\frac{\gamma}{2}-1}
\end{aligned}$$

Ahora como $\mathcal{H} = \{e^{-i}\}_{i=0}^M \cap [h_{\text{mín}}, h_{\text{máx}}]$ y $\gamma - 2 > 0$ ya que $\gamma > 2$, se reescribe la sumatoria del lado derecho de la desigualdad anterior de la siguiente forma.

$$\begin{aligned}
& \sum_{h' \in \mathcal{H}} \Delta_1(h') \\
& \leq \left\{ 2^4 3 \left[\frac{A_1}{n} + B_8 \frac{(\log n)^{-\frac{1}{2}}}{n} \right] + 5!2^{18} 3 B^2 (\sigma + r_{\text{sup}})^2 \frac{\log n}{n} \right\} \sum_{h' \in \mathcal{H}} (h')^{\frac{\gamma-2}{2}} \\
& \leq \left\{ 2^4 3 \left[\frac{A_1}{n} + B_8 \frac{(\log n)^{-\frac{1}{2}}}{n} \right] + 5!2^{18} 3 B^2 (\sigma + r_{\text{sup}})^2 \frac{\log n}{n} \right\} \sum_{i=0}^{\infty} \left(e^{-\frac{\gamma-2}{2}} \right)^i \\
& \leq \left\{ \frac{2^4 3}{1 - e^{-\frac{\gamma-2}{2}}} \left[\frac{A_1}{n} + B_8 \frac{(\log n)^{-\frac{1}{2}}}{n} \right] + \frac{5!2^{18} 3 B^2}{1 - e^{-\frac{\gamma-2}{2}}} (\sigma + r_{\text{sup}})^2 \frac{\log n}{n} \right\} \\
& \leq B_{10} \frac{\log n}{n} \tag{3.57}
\end{aligned}$$

donde $B_{10} = \frac{2^4 3}{1 - e^{-\frac{\gamma-2}{2}}} (A_1 + B_8) + \frac{5!2^{18} 3 B^2}{1 - e^{-\frac{\gamma-2}{2}}} (\sigma + r_{\text{sup}})^2$.

Por las desigualdades (3.49), (3.50) y (3.57), se tiene,

$$\mathbb{E}(T_1^2) \leq 2 \left(B_{10} \frac{\log n}{n} + \frac{B_9}{2n} \right) = A_5 \frac{\log n}{n}. \tag{3.58}$$

3.5.8. Demostración de la Proposición 12, (ii).

Para estudiar $\mathbb{E}(T_2^2)$ se usa de manera similar a T_1 que

$$\mathbb{E}(T_2^2) \leq \sum_{h' \in \mathcal{H}} \mathbb{E} \left[\left\{ |\hat{r}_{hh'}(x) - \mathbb{E}(\hat{r}_{hh'}(x))| - V_2(h') \right\}_+^2 \right] \quad (3.59)$$

Igual que para la demostración de la Proposición 12 (i), se define el estimador auxiliar de truncamiento de la siguiente manera.

$$\tilde{r}_{hh'}(x) = \frac{1}{n} \sum_{k=1}^n Y_k K_{hh'}(x - X_k) g^{-1}(X_k) \mathbf{1}_{\{|Y_k| \leq M_n\}}$$

donde $K_{hh'}(\cdot) = K_h * K_{h'}(\cdot)$, para $h, h' \in \mathcal{H}$.

De manera similar se tiene

$$\mathbb{E} \left(\left\{ |\hat{r}_{hh'}(x) - \mathbb{E}(\hat{r}_{hh'}(x))| - V_2(h') \right\}_+^2 \right) \leq 2\tilde{\Delta}_1(h') + 2\tilde{\Delta}_2(h')$$

donde

$$\tilde{\Delta}_1(h') = \mathbb{E} \left(\left\{ |\tilde{r}_{hh'}(x) - \mathbb{E}(\tilde{r}_{hh'}(x))| - V_2(h') \right\}_+^2 \right)$$

y

$$\tilde{\Delta}_2(h') = \mathbb{E} \left(\left\{ \hat{r}_{hh'}(x) - \tilde{r}_{hh'}(x) - \mathbb{E}(\hat{r}_{hh'}(x) - \tilde{r}_{hh'}(x)) \right\}^2 \right).$$

Se tiene

$$\mathbb{E}(T_2^2) \leq 2 \left(\sum_{h' \in \mathcal{H}} \tilde{\Delta}_1(h') + \sum_{h' \in \mathcal{H}} \tilde{\Delta}_2(h') \right). \quad (3.60)$$

Se procede a acotar la sumatoria $\sum_{h' \in \mathcal{H}} \tilde{\Delta}_2(h')$, reescribiendo

$$\hat{r}_{hh'}(x) - \tilde{r}_{hh'}(x) = \frac{1}{n} \sum_{k=1}^n \tilde{\eta}_k$$

donde $\tilde{\eta}_k = Y_k K_{hh'}(x - X_k) g^{-1}(X_k) \mathbf{1}_{\{|Y_k| > M_n\}}$.

Se observa que la única diferencia entre el termino $\tilde{\eta}_k$ y el termino η_k obtenido en la Proposición 12 es el Kernel usado, en este caso es $K_{hh'}$ y en el caso anterior es $K_{h'}$, lo cual no influye de manera determinante en las demostraciones que siguen. Sólo hay que tomar en consideración que cada vez que aparezca $\|K_{h'}\|_\infty$ se debe tener en realidad $\|K_{hh'}\|_\infty$, además de lo siguiente,

$$\begin{aligned}
|K_{h'h}(x-X)| &= |K_{h'} * K_h(x-X)| \\
&\leq \int_{\mathbb{R}} |K_{h'}(x-X-t)K_h(t)| dt \\
&= \|K_{h'}\|_{\infty} \int_{\mathbb{R}} |K_h(t)| dt \\
&= \|K_{h'}\|_{\infty} \|K\|_1
\end{aligned}$$

por lo cual se debe sustituir en los resultados obtenidos en la Proposición 12 (i), $\|K\|_{\infty}$ por $\|K\|_{\infty}\|K\|_1$. Por lo planteado previamente y la desigualdad (3.50) se tiene,

$$\sum_{h' \in \mathcal{H}} \tilde{\Delta}_2(h') \leq \frac{B_9 \|K\|_1^2}{2n}. \quad (3.61)$$

Ahora se procede a acotar la sumatoria $\sum_{h' \in \mathcal{H}} \tilde{\Delta}_1(h')$. Tenemos de manera similar a la demostración de la parte (i):

$$\begin{aligned}
&\mathbb{E} \left(\left\{ |\tilde{r}_{hh'}(x) - \mathbb{E}(\tilde{r}_{hh'}(x))| - V_2(h') \right\}_+^2 \right) \\
&= \int_0^{+\infty} \mathbb{P} \left(|\tilde{r}_{hh'}(x) - \mathbb{E}(\tilde{r}_{hh'}(x))| > V_2(h') + t^{1/2} \right) dt \\
&= \int_0^{+\infty} \mathbb{P} \left(|\tilde{S}_n| > V_2(h') + t^{1/2} \right) dt,
\end{aligned}$$

donde $\tilde{S}_n = \sum_{i=1}^n \tilde{Z}_i$, $\tilde{Z}_i = \tilde{G}(X_i, \varepsilon_i) - \mathbb{E}[\tilde{G}(X_i, \varepsilon_i)]$, $\tilde{G}(X_i, \varepsilon_i) = \frac{1}{n} Y_i K_{hh'}(x - X_i) g^{-1}(X_i) \mathbf{1}_{\{|Y_i| \leq M_n\}}$.

De nuevo se observa que en las demostraciones que siguen no es determinante el hecho de que en la definición de $\tilde{G}(\cdot, \cdot)$ este en términos del Kernel $K_{hh'}$ en lugar del Kernel $K_{h'}$. Sólo hay que tomar en consideración que cada vez que aparezca $\|K_{h'}\|_2$, $\|K_{h'}\|_1$ y $\|K_{h'}\|_{\infty}$ se debe tener en realidad $\|K_{hh'}\|_2$, $\|K_{hh'}\|_1$ y $\|K_{hh'}\|_{\infty}$ respectivamente, y usar que

- $\|K_{hh'}\|_{\infty} \leq \|K_{h'}\|_{\infty} \|K\|_1$
- $\|K_{hh'}\|_2 \leq \|K_{h'}\|_2 \|K_h\|_1 = \|K_{h'}\|_2 \|K\|_1$
- $\|K_{hh'}\|_1 \leq \|K_{h'}\|_1 \|K_h\|_1 = \|K\|_1^2$.

Así que cada vez que aparezca en los resultados de la Proposición 12 (i), y la Proposición 15, $\|K\|_2$, $\|K\|_1$ y $\|K\|_{\infty}$ se debe reemplazar por $\|K\|_2 \|K\|_1$, $\|K\|_1 \|K\|_1$ y $\|K\|_{\infty} \|K\|_1$

respectivamente. Por lo mencionado previamente y la desigualdad (3.57), se tiene,

$$\sum_{h' \in \mathcal{H}} \tilde{\Delta}_1(h') \leq B_{10} \|K\|_1^2 \frac{\log n}{n}. \quad (3.62)$$

Ahora por las desigualdades (3.61) y (3.62), se tiene,

$$\mathbb{E}(T_2^2) \leq A_5 \|K\|_1^2 \frac{\log n}{n}.$$

Capítulo 4

Estimación adaptativa en modelos de regresión para datos débilmente dependientes y variable explicativa con densidad desconocida

Resumen

Este capítulo está dedicado a la estimación de la función de regresión, cuando la función de regresión es acotada, los errores del modelo son variables aleatorias i.i.d de distribución normal y la variable explicativa es un proceso débilmente dependiente cuyo coeficiente de correlación tiene decaimiento exponencial, además la variable explicativa es idénticamente distribuida con función de densidad g acotada y desconocida. La precisión de la estimación se mide utilizando riesgos puntuales. La función de regresión se expresa como el cociente entre una función m y la densidad g . Se propone realizar un procedimiento basado en datos utilizando estimación por Núcleo, para estimar la función m . El ancho de ventana se selecciona usando el enfoque de Goldenshluger y Lepski, se demuestra que el estimador resultante de la función m satisface una desigualdad tipo oráculo. Lo que permite establecer que es adaptativo. Por Bertin and Klutchnikoff (2017) se tiene que el estimador GL de la función de densidad g también satisface las propiedades antes mencionadas. Posteriormente se utiliza un estimador del tipo Nadaraya-Watson, similar al planteado en Fermin et al. (2017), con la diferencia de que este estimador depende de dos ventanas distintas una para el numerador y otra para el denominador. Finalmente, se demuestra que el estimador propuesto

para la función de regresión satisface una desigualdad de tipo oráculo y es adaptativo. Algunas simulaciones ilustran el desempeño del método propuesto.

4.1. Modelo

Observamos $\{(X_1, Y_1), \dots, (X_n, Y_n)\}$ idénticamente distribuidas que satisfacen

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

donde $n \geq 2$, los X_i son idénticamente distribuidos con función de densidad g desconocida, los ε_i son independientes e idénticamente distribuidos de distribución $N(0, \sigma^2)$ con $\sigma^2 > 0$, los X_i son independientes de los ε_i , y la función r es la función de regresión $r(x) = \mathbb{E}(Y|X = x)$.

Nuestro objetivo es estimar la función r en el punto $x \in \mathbb{R}$ usando la muestra observada $\{(X_1, Y_1), \dots, (X_n, Y_n)\}$ en un enfoque de dependencia. Vamos a usar que la función r se puede descomponer como

$$r(x) = E[Y|X = x] = \int_{\mathbb{R}} yf(y|x)dy = \int_{\mathbb{R}} y \frac{f_{X,Y}(x,y)}{g(x)} dy = \frac{m(x)}{g(x)}, \quad (4.1)$$

donde $f(\cdot|\cdot)$ es la función de densidad condicional de Y dado X , la cual satisface que $f(y|x) = f_{X,Y}(x,y)g^{-1}(x)$, con $f_{X,Y}$ la densidad conjunta de los $\{(X_k, Y_k)\}_{k=1}^n$ y

$$m(x) = \int_{\mathbb{R}} yf_{X,Y}(x,y)dy. \quad (4.2)$$

Dada (4.1), como estrategia de estimación vamos a buscar estimadores de las funciones m y g por separado y tomar como estimador de r el cociente de ambos estimadores.

Se mide la calidad de la estimación de un estimador \hat{r} usando el error cuadrático en un punto $x \in \mathbb{R}$:

$$R(r, \hat{r}, x) = \mathbb{E}(r(x) - \hat{r}(x))^2.$$

Por un lado, se asumen las siguientes hipótesis correspondientes a cotas sobre las densidades de las variables X_i en la vecindad de x :

$$B(x) = [x - 2/(\log n)^2, x + 2/(\log n)^2].$$

(H₁) La densidad g de las X_i satisface

$$g_{\text{inf}} \leq g(u) \leq g_{\text{sup}}, \quad \forall u \in B(x)$$

donde g_{inf} y g_{sup} son constantes positivas.

(H₂) $\forall(i, j)$, la densidad conjunta $g_{i,j}$ de (X_i, X_j) satisface

$$|g_{i,j}(u, v)| \leq Q, \quad \forall u, v \in B(x),$$

donde Q es una constante positiva.

Por otro lado, asumimos una estructura de dependencia débil sobre las variables X_i . Mas precisamente, para u y v enteros positivos, denotamos $i_{1:u} = (i_1, \dots, i_u) \in \mathbb{Z}^u$ y $j_{1:v} = (j_1, \dots, j_v) \in \mathbb{Z}^v$, entonces se definen los vectores aleatorios $X_{i_{1:u}} = (X_{i_1}, X_{i_2}, \dots, X_{i_u})$ y $X_{j_{1:v}} = (X_{j_1}, X_{j_2}, \dots, X_{j_v})$ a valores en \mathbb{R}^u y \mathbb{R}^v respectivamente. Se define la función $q: \mathbb{Z}^u \times \mathbb{Z}^v \rightarrow \mathbb{Z}$ por $q(i_{1:u}, j_{1:v}) = \text{mín}(j_{1:v}) - \text{máx}(i_{1:u})$ y Λ_u la clase de funciones $G_u: \mathbb{R}^u \rightarrow \mathbb{R}$ tales que $\|G_u\|_\infty = \sup_{x \in \mathbb{R}^u} |G_u(x)| < \infty$. Para un proceso aleatorio $\mathbb{X} = (X_i)_{i \in \mathbb{Z}}$, se define el coeficiente de correlación $\alpha(\mathbb{X}) = (\alpha_k(\mathbb{X}))_{k \in \mathbb{N}}$ por,

$$\alpha_k(\mathbb{X}) = \sup_{u,v \in \mathbb{N}} \sup_{(i_{1:u}, j_{1:v}) \in \mathbb{Z}^u \times \mathbb{Z}^v, q(i_{1:u}, j_{1:v}) \geq k} \sup_{G_u \in \Lambda_u} \sup_{G_v \in \Lambda_v} \frac{|\text{cov}(G_u(X_{i_{1:u}}), G_v(X_{j_{1:v}}))|}{\Psi(u, v, G_u, G_v)}$$

con $\Psi(u, v, G_u, G_v) = 4\|G_u\|_\infty\|G_v\|_\infty$.

Suponemos que el coeficiente α -mixing de las variables X_i satisfacen la siguiente hipótesis.

(H₃) Existe $a \in]0, 1[$, tal que

$$\alpha_k(\mathbb{X}) \leq a^k, \quad \forall k \in \mathbb{N}.$$

En lo siguiente vamos a denotar por \mathcal{L} , el conjunto de los procesos \mathbb{X} que satisfacen (H₁), (H₂) y (H₃).

A continuación, se dan dos ejemplos de procesos que satisfacen estas hipótesis.

Ejemplo 3. Consideramos $\mathbb{Z} = \{Z_t\}_{t \geq 1}$, un proceso autoregresivo de orden 1, definido por $Z_t = \phi Z_{t-1} + \rho \xi_t$, con $\xi_t \sim N(0, 1)$, $|\phi| < 1$ y $\rho > 0$. Por recurrencia $Z_t = \phi^n Z_{t-n} + \rho \sum_{i=0}^{n-1} \phi^i \xi_{t-i}$ y al tomar $n \rightarrow \infty$ se tiene $Z_t = \rho \sum_{i=0}^{\infty} \phi^i \xi_{t-i}$ resulta un proceso gaussiano centrado y estacionario con función de covarianza $\gamma_{\mathbb{Z}}(k) = \text{cov}(Z_t, Z_{t+k}) = \frac{\rho^2}{1-\phi^2} \phi^k$ para

$k \geq 0$. Se puede demostrar que el proceso \mathbb{Z} es débilmente dependiente, por lo cual satisface la Hipótesis (H_3) , ver Doukhan (1994).

Ejemplo 4. Al denotar $\phi_{\mu, \sigma^2}(x)$ como la función de densidad $N(\mu, \sigma^2)$ y $\Phi_{\mu, \sigma^2}(x)$ como la función de distribución acumulada $N(\mu, \sigma^2)$. Para $c \in \mathbb{R}^+$, las funciones de densidad y de distribución acumulada normal truncada en $[-c, c]$ de media cero y varianza uno, están dadas por $g(x) = \frac{\phi_{0,1}(x)}{p} \mathbf{1}_{[-c,c]}(x)$, con $p = \Phi_{0,1}(c) - \Phi_{0,1}(-c)$, para cada $x \in \mathbb{R}$ y

$$G(x) = \begin{cases} 0 & \text{si } x < -c \\ \frac{1}{p} (\Phi_{0,1}(x) - \Phi_{0,1}(-c)) & \text{si } -c \leq x < c \\ 1 & \text{si } x \geq c \end{cases}$$

respectivamente. Además, la función inversa de G es $G^{-1}(u) = \Phi_{0,1}^{-1}(pu + \Phi_{0,1}(c))$ para cada $u \in [0, 1]$.

Se puede demostrar que el proceso $\mathbb{X} = \{X_t\}_{t \geq 1}$ definido por $X_t = \left(G^{-1} \circ \Phi_{0, \frac{\rho^2}{1-\phi^2}} \right) (Z_t)$ cumple que $\mathbb{X} \in \mathcal{L}$, donde el proceso $\mathbb{Z} = \{Z_t\}_{t \geq 1}$ con cada Z_t de distribución $\Phi_{0, \frac{\rho^2}{1-\phi^2}}$ es el proceso del Ejemplo 1.

Finalmente, asumimos las siguiente hipótesis sobre la función de regresión.

(H_4) Existe una constante positiva r_{sup} tal que

$$|r(u)| \leq r_{\text{sup}}, \quad \forall u \in B(x).$$

Observa que dado que $m(\cdot) = r(\cdot)g(\cdot)$, las hipótesis (H_1) y (H_4) implican que $\forall u \in B(x)$

$$|m(u)| \leq r_{\text{sup}}g_{\text{sup}}.$$

A continuación, en la Sección 4.2, *Procedimiento estadístico*; se observa que la función de regresión se puede escribir como el cociente de una función m y la función de densidad g , que en este caso es desconocida. Es decir la función de regresión $r = \frac{m}{g}$, donde tanto las funciones numerador y denominador son desconocidas. Lo que motiva el hecho de que al tomar un estimador por Núcleo \hat{m}_h de m y otro estimador por Núcleo \hat{g}_{h^*} de g , se tendrá un estimador por Núcleo \hat{r}_{h/h^*} de r . para evitar que se anule el denominador se toma el

estimador \bar{g}_{h^*} que por su definición es distinto de cero, lo cual lo hace ser conveniente para definir $\hat{r}_{h/h^*} = \frac{\hat{m}_h}{\bar{g}_{h^*}}$ para $h \in \mathcal{H}$ y $h^* \in \mathcal{H}^*$, pero como se vera mas adelante las propiedades estadísticas del estimador de r al final dependerán de las propiedades estadísticas de \hat{m}_h y \hat{g}_{h^*} . Finalmente para \hat{m}_h con $h \in \mathcal{H}$ y \bar{g}_{h^*} con $h^* \in \mathcal{H}^*$ se describe como seleccionar la ventana siguiendo el procedimiento GL. La Sección 4.3, *Resultados*; se encarga de estudiar cada una de las propiedades estadísticas de los estimadores que forman parte de la definición del estimador de r y su relación entre ellos, por eso esta dividido en tres subsecciones. Subsección 4.3.1, *Resultados de los estimadores \hat{m}_h y \hat{g}_{h^*}* ; Aunque en esta subsección se muestren todos los resultados de sesgo, varianza, consistencia y velocidad de convergencia del estimador \hat{m}_h algunos dentro de la subsección y otros en la Sección 4.5, se debe destacar que todas las demostraciones son similares a las del Capítulo 3, ya que el estimador por Núcleo de r , con g conocido, es muy similar al estimador \hat{m}_h . En el caso del sesgo varianza consistencia y velocidad de convergencia del estimador \hat{g}_{h^*} las propiedades solo se enuncian ya que se demuestran siguiendo un procedimiento análogo al de las propiedades del estimador \hat{m}_h . La Subsección 4.3.2, *Desigualdades de Oráculo y adaptatividad del estimador \hat{m}_h* ; se dedica a establecer que el estimador de m en la ventana seleccionada por el procedimiento G-L satisface una desigualdad de tipo Oráculo y que finalmente el estimador es adaptativo. Ambas demostraciones se hacen siguiendo los procedimientos del Capítulo 3, aunque en la Sección 4.5 se puede encontrar la que corresponde a la desigualdad de tipo Oráculo. La Subsección 4.3.3, *Resultados del estimador \hat{r}_{h/h^*}* ; esta parte es la que establece la relación entre los distintos estimadores involucrados, específicamente que el error cuadrático medio del estimador $\hat{r}_{h/h^*} = \frac{\hat{m}_h}{\bar{g}_{h^*}}$ esta controlado por una combinación lineal entre el error cuadrático medio de \hat{m}_h y el error cuadrático medio de \bar{g}_{h^*} . para realizar esta demostración se crean varios lemas y proposiciones técnicos, además de usar la desigualdad de Bernstein para datos dependientes. De este resultado se desprende de forma inmediata la consistencia y la velocidad de convergencia del estimador de regresión. Finalmente, con esa misma desigualdad y con la ayuda de otro lema técnico se demuestra la adaptatividad del estimador por Núcleo de la función de regresión, ver detalles en la Sección 4.5. En la Sección 4.4, *Estudio de Simulación*, se plantea el esquema de simulación para generar datos del modelo de regresión con función de regresión acotada, variable explicativa de densidad acotada, débil dependiente de coeficiente α -mixing con decaimiento exponencial. Se plantea como obtener el estimador no paramétrico adaptativo, con núcleo gaussiano, de la función de regresión usando los datos simulados en el esquema planteado previamente. Posteriormente se muestra cómo se calibra el método usando los mismos datos simulados, y finalmente se presenta un estudio comparativo de resultados de estimaciones realizadas, usando errores globales y locales empíricos para distintos tamaños de muestras simuladas y distintos valores de la

desviación estándar del error del modelo de regresión. La Sección 4.5, *Demostraciones*, reúne las demostraciones de los lemas, proposiciones, teoremas y corolarios de este capítulo. Se debe destacar que en este capítulo al no ser acotada la variable respuesta del modelo, se introduce la definición de truncamiento del estimador de la función $m(\cdot)$, lo que permite descomponer el estimador por núcleo de la función $m(\cdot)$ en dos términos uno truncado y otro no acotado, el término estocástico no acotado del estimador se controla con la hipótesis de decaimiento exponencial del coeficiente de dependencia débil y el término estocástico truncado del estimador de la función $m(\cdot)$ es controlado con una adaptación que se realizó de la desigualdad de Bernstein para datos dependientes de Doukhan and Neumann (2007).

4.2. Procedimiento estadístico

Para la muestra $\{(X_k, Y_k)\}_{k=1}^n$ se consideran los estimadores por núcleo de la función de densidad g y de la densidad conjunta $f_{X,Y}$ dados para $h > 0$ por:

$$\hat{g}_h(x) = \frac{1}{n} \sum_{i=1}^n K_h(x - X_i) \tag{4.3}$$

y

$$\hat{f}_{X,Y}^{(h)}(x, y) = \frac{1}{n} \sum_{i=1}^n K_h(x - X_i) K_h(y - Y_i)$$

donde $K_h(\cdot) = \frac{1}{h}K(\frac{\cdot}{h})$ y $K : \mathbb{R} \rightarrow \mathbb{R}$, es una función de núcleo satisfaciendo $\int K(u)du = 1$ y la siguiente hipótesis

(H_5) K tiene soporte $[-1, 1]$ y , $\|K\|_\infty < +\infty$.

Esta hipótesis implica que $\|K\|_1 < +\infty$ y $\|K\|_2 < +\infty$.

Al reemplazar en la ecuación (4.2), la densidad conjunta por $\hat{f}_{X,Y}^{(h)}(x, y)$ (y asumiendo $\int uK(u)du = 0$) se obtiene el estimador de m

$$\hat{m}_h(x) = \frac{1}{n} \sum_{i=1}^n K_h(x - X_i) Y_i. \tag{4.4}$$

Nuestra estrategia de estimación consiste en tomar el cociente de los estimadores (4.3) y (4.4). Con el fin de evitar que $\hat{g}_h(x)$ se anule, definimos un estimador alternativo de g definido por

$$\bar{g}_h(x) = \text{máx} \left\{ \hat{g}_h(x), \frac{1}{\log n} \right\}. \tag{4.5}$$

Finalmente, consideramos el estimador de r

$$\hat{r}_{h/h^*} = \frac{\hat{m}_h}{\bar{g}_{h^*}} \quad (4.6)$$

que depende de dos ventanas $h, h^* > 0$. Observa que si $h = h^*$, este estimador está relacionado con el estimador de Nadaraya-Watson (Nadaraya (1989)).

Para seleccionar las ventanas h y h^* del estimador, vamos a usar el método de Goldenshluger-Lepski (GL) y seleccionar por separado h y h^* para el estimador de m y el estimador de g .

Procedimiento de selección de ventana para estimar m : Consideramos la familia de anchos de ventanas

$$\mathcal{H} = \{e^{-i}\}_{i=0}^M \cap [h_{\min}, h_{\max}]$$

con $h_{\min} = \frac{(\log n)^8}{n}$, $h_{\max} = \frac{1}{(\log n)^2}$ y $M = \left\lceil \log \left(\frac{n}{(\log n)^8} \right) \right\rceil$. Se define para $h, h' \in \mathcal{H}$ un estimador auxiliar sobre suavizado

$$\hat{m}_{h,h'}(x) = \sum_{i=1}^n K_h * K_{h'}(x - X_i) Y_i, \quad (4.7)$$

también se define para $h \in \mathcal{H}$,

$$A(h, x) = \max_{h' \in \mathcal{H}} \{|\hat{m}_{h,h'}(x) - \hat{m}_{h'}(x)| - V(h')\}_+ \quad (4.8)$$

donde $\{y\}_+ = \max(0, y)$ denota la parte no negativa de $y \in \mathbb{R}$ y $V(h)$ está dado por,

$$V(h) = \sqrt{2\gamma A_1} (\|K\|_1 + 1) (1 + \delta_n) \frac{(\log n)^{\frac{1}{2}}}{(nh)^{\frac{1}{2}}} \quad (4.9)$$

con $\gamma > 2$ y $A_1 = ((r_{\sup})^2 + \sigma^2) g_{\sup} \|K\|_2^2$ y $\delta_n = (\log n)^{-\frac{1}{5}}$.

El procedimiento consiste en seleccionar, basado en los datos, un ancho de ventana \hat{h} de la familia \mathcal{H} , dado por

$$\hat{h} = \operatorname{argmin}_{h \in \mathcal{H}} \{A(h, x) + V(h)\}$$

Mostramos que el estimador resultante, $\hat{m}_{\hat{h}}$, satisface una desigualdad de oráculo que permite demostrar que es adaptativo.

Procedimiento de selección de ventana para estimar g : Aquí usaremos el método propuesto en Bertin and Klutchnikoff (2017). De forma análoga se define una familia de ventanas

$$\mathcal{H}^* = \{2^{-k} : k \in \mathbb{N}\} \cap [h_{\min}^*, h_{\max}^*]$$

donde $h_{\min}^* = n^{-1} \exp(\sqrt{\log n} - 1)$ y $h_{\max}^* = \exp(-\sqrt{\log n})$.

Se define para $h \in \mathcal{H}^*$

$$A^*(h, x) = \max_{h' \in \mathcal{H}^*} \{|\hat{g}_{h \vee h'}(x) - \hat{g}_{h'}(x)| - V^*(h, h')\}_+ \quad (4.10)$$

y

$$V^*(h) = \sqrt{2\gamma^* |\log h| \left(\hat{J}_n(h) + \frac{\delta_n^*}{nh} \right)} \quad (4.11)$$

donde $\delta_n^* = (\log n)^{-\frac{1}{2}}$ y $h \vee h' = \max\{h, h'\}$, para todo $h, h' \in \mathcal{H}^*$.

También, se considera

$$V^*(h, h') = V^*(h') + V^*(h \vee h')$$

y

$$\hat{J}(h) = \frac{1}{n^2} \sum_{i=1}^n K_h^2(x - X_i)$$

El procedimiento consiste en seleccionar, basado en los datos, un ancho de ventana \hat{h}^* de la familia \mathcal{H}^* , dado por

$$\hat{h}^* = \operatorname{argmin}_{h \in \mathcal{H}^*} \{A^*(h, x) + V^*(h)\}$$

El estimador resultante, $\hat{g}_{\hat{h}^*}$, satisface una desigualdad de oráculo y es adaptativo (Ver Teoremas 1 y 2 en Bertin and Klutchnikoff (2017)).

4.3. Resultados

4.3.1. Resultados de los estimadores \hat{m}_h y \hat{g}_h .

En esta sección se estudia el sesgo, varianza y velocidad de convergencia del estimador \hat{m}_h y se enuncian estos mismos resultados para el estimador de densidad \hat{g}_h . Para poder controlar el sesgo y demostrar que es asintóticamente insesgado se dan las definiciones de clase de Hölder y de kernel de orden m_0 .

Definición 10 (Clase de Hölder). Sea $\beta > 0$, $L > 0$. La clase de Hölder $\Sigma(\beta, L)$ se define como el conjunto de todas las funciones $f : \mathbb{R} \rightarrow \mathbb{R}$ tales que la derivada $f^{(l)}$, $l = \lfloor \beta \rfloor$ existe y

$$\left| f^{(l)}(x) - f^{(l)}(y) \right| \leq L|x - y|^{\beta-l}, \forall x, y \in \mathbb{R}$$

donde $\lfloor \beta \rfloor = \max \{n \in \mathbb{N}, n < \beta\}$.

Definición 11 (Kernel de orden m_0). Sea $m_0 \in \mathbb{N}$. Se dice que $K : \mathbb{R} \rightarrow \mathbb{R}$ es un kernel de orden m_0 si las funciones $u \rightarrow u^j K(u)$, $j = 0, 1, \dots, m_0$, satisfacen

$$\int_{\mathbb{R}} K(u) du = 1, \int_{\mathbb{R}} u^j K(u) du = 0, j = 1, \dots, m_0.$$

A continuación, en las dos siguientes proposiciones se acota el sesgo y la varianza del estimador \hat{m}_h .

Proposición 16. Sea $\beta_1 > 0$ and $L_1 > 0$. Asumimos que K es de orden m_0 , donde $m_0 \geq l_1 = \lfloor \beta_1 \rfloor$, y que satisface $\int_{\mathbb{R}} |u|^{\beta_1} |K(u)| du < \infty$. Entonces, si $m \in \Sigma(\beta_1, L_1)$, se tiene

$$|E[\hat{m}_h(x)] - m(x)| = |K_h * m(x) - m(x)| \leq h^{\beta_1} A_0,$$

donde $A_0 = \frac{L_1}{l_1!} \int_{\mathbb{R}} |u|^{\beta_1} |K(u)| du$. El estimador \hat{m}_h es asintóticamente insesgado, cuando h tiende a 0.

Proposición 17. Bajo las hipótesis (H_1) , (H_2) , (H_3) , (H_4) y (H_5) se tiene que para $h \in (0, h_{max})$

$$\mathbb{E} [(\hat{m}_h(x) - \mathbb{E}[\hat{m}_h(x)])^2] \leq \frac{A_1}{nh} + A_2 \frac{(\log n)^{-\frac{1}{2}}}{nh},$$

para cada $n \geq 4$, donde las constantes son definidas por $A_1 = ((r_{sup})^2 + \sigma^2) g_{sup} \|K\|_2^2$, $A_2 = \frac{2A_3}{|\log a|} + 20A_4$, $A_3 = (r_{sup})^2 \|K\|_1^2 (Q + (g_{sup})^2)$, y $A_4 = \frac{4}{1-a} (r_{sup})^2 \|K\|_{\infty}^2$.

Proposiciones 16 y 17 implican que cuando $n \rightarrow \infty$, $h \rightarrow 0$ y $nh \rightarrow \infty$, $E[(\hat{m}_h(x) - m(x))^2]$ tiende a 0, o sea $\hat{m}_h(x)$ es un estimador consistente en media cuadrática (y en probabilidad) de $m(x)$.

Ahora se determinará la velocidad de convergencia del estimador $\hat{m}_h(x)$, lo cual se establece en el siguiente teorema.

Teorema 6. *Bajo las hipótesis de las Proposiciones 16 y 17, se tiene que, si $m \in \Sigma(\beta_1, L_1)$ entonces el estimador $\hat{m}_{\tilde{h}}$, con $\tilde{h} = n^{-\frac{1}{2\beta_1+1}}$ satisface*

$$E \left[(\hat{m}_{\tilde{h}}(x) - m(x))^2 \right] \leq Cn^{-\frac{2\beta_1}{2\beta_1+1}},$$

donde C es una constante positiva que depende del núcleo K , de β_1 , L_1 , r_{sup} , Q , g_{sup} , σ y a .

Usando Proposiciones 16 y 17, se deduce que para $n \geq 4$,

$$E \left[(\hat{m}_h(x) - m(x))^2 \right] \leq h^{2\beta_1} A_0^2 + \frac{(A_1 + A_2)}{nh}. \quad (4.12)$$

Substituyendo h por \tilde{h} se obtiene el resultado del teorema. Nota que cualquier ventana de la forma $An^{-\frac{1}{2\beta_1+1}}$ permite obtener la misma tasa de convergencia. La ventana que minimiza en h el miembro de derecha de la desigualdad (4.12) es de esa forma con A dependiendo de A_0 , A_1 y A_2 .

Siguiendo el mismo esquema de las demostraciones anteriores, se dan los resultados correspondientes al sesgo, varianza, consistencia y velocidad de convergencia del estimador \hat{g}_h , los cuales se enuncian en las siguientes proposiciones y teoremas (ver también Bertin and Klutchnikoff (2017)).

Proposición 18. *Sea $\beta_2 > 0$ and $L_2 > 0$. Asumimos que K es de orden m_0 , donde $m_0 \geq l_2 = \lfloor \beta_2 \rfloor$, y que satisface $\int_{\mathbb{R}} |u|^{\beta_2} |K(u)| du < \infty$. Entonces, si $g \in \Sigma(\beta_2, L_2)$, se tiene*

$$|\mathbb{E}[\hat{g}_h(x)] - g(x)| \leq h^{\beta_2} \mathfrak{A}_0,$$

donde $\mathfrak{A}_0 = \frac{L_2}{l_2!} \int_{\mathbb{R}} |u|^{\beta_2} |K(u)| du$. El estimador es asintóticamente insesgado, cuando h tiende a 0.

Proposición 19. *Bajo las hipótesis (H_1) , (H_2) , (H_3) , y (H_5) se tiene que para $h \in (0, h_{max}^*)$*

$$\mathbb{E} \left[(\hat{g}_h(x) - \mathbb{E}[\hat{g}_h(x)])^2 \right] \leq \frac{\mathfrak{A}_1}{nh} + \mathfrak{A}_2 \frac{(\log n)^{-\frac{1}{2}}}{nh},$$

para cada $n \geq 4$, donde $\mathfrak{A}_1 = g_{sup} \|K\|_2^2$ y $\mathfrak{A}_2 = \frac{2\mathfrak{A}_3}{|\log a|} + 20\mathfrak{A}_4$ con $\mathfrak{A}_3 = \|K\|_1^2 (Q + (g_{sup})^2)$ y $\mathfrak{A}_4 = \frac{4}{1-a} \|K\|_{\infty}^2$.

Usando proposiciones 18 y 19, se deduce el siguiente teorema.

Teorema 7. *Bajo las hipótesis de las Proposiciones 18 y 19, se tiene que, si $g \in \Sigma(\beta_2, L_2)$ entonces el estimador $\hat{g}_{\tilde{h}^*}$, con $\tilde{h}^* = n^{-\frac{1}{2\beta_2+1}}$ satisface*

$$\mathbb{E} \left[\left(\hat{g}_{\tilde{h}^*}(x) - g(x) \right)^2 \right] \leq \mathfrak{C} n^{-\frac{2\beta_2}{2\beta_2+1}},$$

donde \mathfrak{C} es una constante positiva que depende del núcleo K , de β_2 , L_2 , \mathcal{Q} , g_{sup} y a .

4.3.2. Desigualdades de Oráculo y adaptatividad del estimador $\hat{m}_{\hat{h}}$

Teorema 8. (Desigualdad de Oráculo) *Bajo las hipótesis (H_1) , (H_2) , (H_3) , (H_4) y (H_5) , el estimador $\hat{m}_{\hat{h}}(x)$ satisface la siguiente desigualdad,*

$$\left(\mathbb{E} \left(m(x) - \hat{m}_{\hat{h}}(x) \right)^2 \right)^{\frac{1}{2}} \leq \min_{h \in \mathcal{H}} \left(A_6 C(h) + A_7 (1 + \delta_n) \frac{(\log n)^{\frac{1}{2}}}{(nh)^{\frac{1}{2}}} \right) + A_8 \frac{(\log n)^{\frac{1}{2}}}{n^{\frac{1}{2}}} \quad (4.13)$$

donde $A_6 = 1 + 2\|K\|_1$, $A_7 = \sqrt{A_1 + A_2} + 2\sqrt{2\gamma A_1}(\|K\|_1 + 1)$ y A_8 es una constante positiva que depende del kernel K , de r_{sup} , \mathcal{Q} , g_{sup} , σ , a y γ .

Ahora usando el Teorema 8 y siguiendo la misma demostración que la del Teorema 5 en el capítulo 3 deducimos el siguiente teorema para el estimador $\hat{m}_{\hat{h}}$.

Teorema 9. (Adaptatividad) *Asumimos que las hipótesis (H_1) , (H_2) , (H_3) , (H_4) y (H_5) están satisfechas. Asumimos que K es de orden m_0 . Entonces para todo $0 < \beta_1 \leq m_0$ y $L_1 > 0$ el estimador $\hat{m}_{\hat{h}}$ satisface para $m \in \Sigma(\beta_1, L_1)$.*

$$\left(\mathbb{E} \left(m(x) - \hat{m}_{\hat{h}}(x) \right)^2 \right)^{\frac{1}{2}} \leq C^* \left(\frac{n}{\log n} \right)^{-\frac{\beta_1}{2\beta_1+1}}$$

donde C^* es una constante positiva que depende del kernel K , de β_1 , L_1 , r_{sup} , \mathcal{Q} , g_{sup} , σ^2 , a y γ .

4.3.3. Resultados del estimador \hat{r}_{h/h^*} .

Ahora vamos a estudiar las propiedades del estimador \hat{r}_{h/h^*} . Este estimador satisface

$$\begin{aligned} r(x) - \hat{r}_{h/h^*}(x) &= \frac{m(x)}{g(x)} - \frac{\hat{m}_h(x)}{\bar{g}_{h^*}(x)} = \frac{m(x)}{\bar{g}_{h^*}(x)} - \frac{\hat{m}_h(x)}{\bar{g}_{h^*}(x)} + \frac{m(x)}{g(x)} - \frac{m(x)}{\bar{g}_{h^*}(x)} \\ &= \frac{1}{\bar{g}_{h^*}(x)} (m(x) - \hat{m}_h(x)) + \frac{m(x)}{g(x)} \left(1 - \frac{g(x)}{\bar{g}_{h^*}(x)} \right) \\ &= \frac{1}{\bar{g}_{h^*}(x)} (m(x) - \hat{m}_h(x)) + \frac{r(x)}{\bar{g}_{h^*}(x)} (\bar{g}_{h^*}(x) - g(x)). \end{aligned} \quad (4.14)$$

Por la formula (4.14), vemos que las propiedades de \hat{r}_{h/h^*} pueden ser deducidas de las propiedades de los estimadores de m y g . Tenemos la siguiente proposición y corolario sobre la consistencia de este estimador.

Proposición 20. *Asumimos que las hipótesis (H_1) , (H_2) , (H_4) y (H_5) se cumplen. Sea $\beta_2 > 0$ y $L_2 > 0$. Asumimos que $g \in \Sigma(\beta_2, L_2)$. Para $h \in \mathcal{H}$ y $h^* \in \mathcal{H}^*$ tenemos*

$$\mathbb{E} \left[(r(x) - \hat{r}_{h/h^*}(x))^2 \right] \leq F_1 \mathbb{E} \left[(m(x) - \hat{m}_h(x))^2 \right] + F_2 \mathbb{E} \left[(g(x) - \hat{g}_{h^*}(x))^2 \right] + \frac{F_3}{n^4}$$

donde $F_1 = \frac{8}{(g_{\inf})^2}$, $F_2 = 8 \frac{(r_{\sup})^2}{(g_{\inf})^2}$ y F_3 es una constante positiva que depende del kernel K , de β_2 , L_2 , r_{\sup} , Q , g_{\inf} , g_{\sup} , y a .

Corolario 2. (Consistencia) *Bajo las hipótesis de las Proposiciones 16, 17, 18, 19 y 20, si $n \rightarrow \infty$, $h \rightarrow 0$, $h^* \rightarrow 0$, $nh \rightarrow \infty$ y $nh^* \rightarrow \infty$, entonces $\mathbb{E} \left[(\hat{r}_{h/h^*}(x) - r(x))^2 \right] \rightarrow 0$, es decir $\hat{r}_{h/h^*}(x) \xrightarrow{\mathbb{P}} r(x)$ y el estimador \hat{r}_{h/h^*} es un estimador consistente de r .*

Teorema 10. (Velocidad de convergencia) *Bajo las hipótesis de los Teoremas 6 y 7, si $m \in \Sigma(\beta_1, L_1)$ y $g \in \Sigma(\beta_2, L_2)$, entonces existe una constante $\mathfrak{D} > 0$ tal que el estimador $\hat{r}_{\tilde{h}/\tilde{h}^*}$ con $\tilde{h} = n^{-\frac{1}{2\beta_1+1}}$ y $\tilde{h}^* = n^{-\frac{1}{2\beta_2+1}}$, satisface*

$$\mathbb{E} \left[\left(\hat{r}_{\tilde{h}/\tilde{h}^*}(x) - r(x) \right)^2 \right] \leq \mathfrak{D} n^{-\frac{2(\beta_1 \wedge \beta_2)}{2(\beta_1 \wedge \beta_2)+1}}.$$

donde \mathfrak{D} es una constante positiva que depende del kernel K , de β_1 , L_1 , β_2 , L_2 , r_{\sup} , g_{\inf} , g_{\sup} , Q , σ , a y $\{\alpha_k(\mathbb{X})\}_{k \geq 0}$.

La Proposición 20, los Teoremas 6 y 7, implican que

$$\mathbb{E} \left[\left(\hat{r}_{\tilde{h}/\tilde{h}^*}(x) - r(x) \right)^2 \right] \leq F_1 C n^{-\frac{2\beta_1}{2\beta_1+1}} + F_2 \mathfrak{C} n^{-\frac{2\beta_2}{2\beta_2+1}} + \frac{F_3}{n^4},$$

lo que permite obtener el resultado del teorema.

Ahora vamos a estudiar el estimador $\hat{r}_{\hat{h}/\hat{h}^*} = \frac{\hat{m}_{\hat{h}}}{\hat{g}_{\hat{h}^*}}$, donde $\hat{h} = \operatorname{argmin}_{h \in \mathcal{H}} \{A(h, x) + V(h)\}$ y $\hat{h}^* = \operatorname{argmin}_{h \in \mathcal{H}^*} \{A^*(h, x) + V^*(h)\}$ han sido descritos en la Sección 4.2. Tenemos el siguiente teorema.

Teorema 11. (Adaptatividad) *Asumimos que las hipótesis (H_1) , (H_2) , (H_3) , (H_4) y (H_5) están satisfechas. Asumimos que K es de orden m_0 . Entonces para todo $0 < \beta_1 \leq m_0$, $L_1 > 0$, $0 < \beta_2 \leq m_0$ y $L_2 > 0$ el estimador $\hat{r}_{\hat{h}/\hat{h}^*}$ satisface para $m \in \Sigma(\beta_1, L_1)$ y $g \in \Sigma(\beta_2, L_2)$*

$$\left(\mathbb{E} \left[\left(\hat{r}_{\hat{h}/\hat{h}^*}(x) - r(x) \right)^2 \right] \right)^{\frac{1}{2}} \leq \mathfrak{B} \left(\frac{n}{\log n} \right)^{-\frac{(\beta_1 \wedge \beta_2)}{2(\beta_1 \wedge \beta_2) + 1}}$$

donde \mathfrak{B} es una constante positiva que depende del kernel K , de β_1 , L_1 , β_2 , L_2 , r_{sup} , Q , g_{inf} , g_{sup} , σ^2 y a .

4.4. Estudio de Simulación

En esta sección se muestra el ajuste del procedimiento planteado para datos simulados. Se comparan los errores empíricos globales y locales, para distintos tamaños de la muestra. Además, se esquematiza el procedimiento utilizado para calibrar el método. El objetivo es estimar la función de regresión cuando la variable explicativa es un proceso débilmente dependiente, con función de densidad desconocida.

4.4.1. Esquema de simulación

Se plantea el modelo de regresión $Y_i = r(X_i) + \varepsilon_i$ para $i = 1, \dots, n$, donde las variables ε_i son independientes e idénticamente distribuidas de distribución $N(0, \sigma^2)$ con $\sigma = 0,1, 0,5$ y 1 . La función de regresión a estimar es $r(x) = 0,7x + 2e^{-10x^2}$ que restringida al intervalo $[-2, 2]$ satisface la hipótesis (H_4) .

Se genera una muestra $\{X_i\}_{i=1}^{n+2q}$ del proceso \mathbb{X} presentado en el Ejemplo 4, donde $X_i = \left(G^{-1} \circ \Phi_{0, \frac{\rho^2}{1-\phi^2}} \right) (Z_i)$ con $\rho = 1$, $\phi = 0,75$ y $c = 2$. Los X_i tienen distribución normal truncada, específicamente su función de densidad es $g(x) = \frac{\phi_{0,1}(x)}{p} \mathbf{1}_{[-2,2]}(x)$, con $p = \Phi_{0,1}(2) - \Phi_{0,1}(-2)$, para cada $x \in \mathbb{R}$. Como mencionamos en el Ejemplo 4, el proceso cumple que $\mathbb{X} \in \mathcal{L}$, es decir satisface las hipótesis (H_1) , (H_2) y (H_3) .

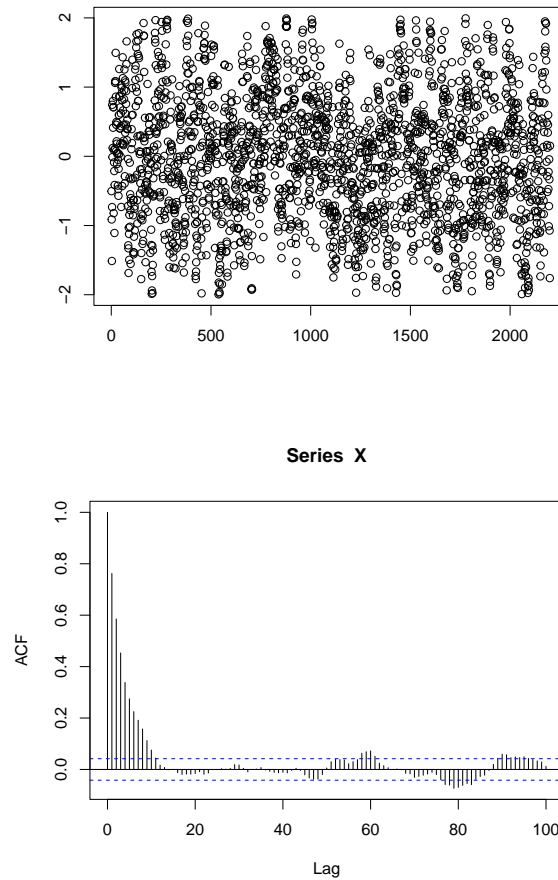


Figura 4.1 Gráfico de dispersión de \mathbb{X} y función de autocorrelación muestral γ_X .

La Figura 4.1 muestra el gráfico de dispersión de las observaciones $\{X_i\}_{i=1}^{n+2q}$ para $n = 2000$ y $q = 100$ (superior) y la función de autocorrelación muestral $\gamma_X(k) = \text{corr}(X_i, X_{i+k})$, para $k = 1, \dots, 100$ (inferior). Se puede apreciar el decaimiento exponencial de las correlaciones, característica presente en los procesos débilmente dependientes de tipo α -mixing.

Las primeras n observaciones $\{(X_i, Y_i)\}_{i=1}^n$ conforman la muestra usada en la sección 4.4.2 para estimar aplicando el método GL, y las últimas q observaciones $\{(X_i, Y_i)\}_{i=n+q+1}^{n+2q}$ es la muestra usada para calibrar el método seleccionando un valor de γ^* y otro valor γ adecuados como se describe en la sección 4.4.3. La idea de dejar un salto temporal de tamaño q entre la muestra de estimación y la muestra de calibración es para asegurar que el coeficiente de correlación $\alpha_m(\mathbb{X})$ -mixing sea suficientemente pequeño y así disminuir el efecto de sobreajuste producido por la dependencia entre las muestras, el control del

valor del coeficiente de correlación $\alpha_m(\mathbb{X})$ es realizado desde la hipótesis de decrecimiento exponencial (H_3).

4.4.2. Estimación GL

La función de regresión r se estima en la malla igualmente espaciada $\{x_i\}_{i=1}^s$ del intervalo $[-1, 1]$, donde $x_i = -1 + 2\left(\frac{i-1}{s-1}\right)$ para $i = 1, \dots, s$, y consideramos $s = 21$. Para estimar la función de regresión $r(\cdot) = \frac{m(\cdot)}{g(\cdot)}$ a partir de la muestra $\{(X_i, Y_i)\}_{i=1}^n$, en cada punto x_i para $i = 1, \dots, s$, se usan estimadores de la forma $\hat{r}_{h_i/h_i^*}(x_i) = \frac{\hat{m}_{h_i}(x_i)}{\hat{g}_{h_i^*}(x_i)}$ donde,

$$\hat{m}_{h_i}(x_i) = \frac{1}{n} \sum_{j=1}^n K_{h_i}(x_i - X_j) Y_j \quad \text{y} \quad \hat{g}_{h_i^*}(x_i) = \frac{1}{n} \sum_{j=1}^n K_{h_i^*}(x_i - X_j)$$

para $h_i, h_i^* \in \mathcal{H}$ con \mathcal{H} la familia de ventanas $\mathcal{H} = \left\{ e^{-0,1j} : j = 0, \dots, \frac{[\log(n)]^{\frac{2}{3}}}{0,1} \right\}$. Se usa como Kernel $K(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$, $x \in \mathbb{R}$, que aunque no es de soporte compacto tiene buenas propiedades practicas, del cual se sabe que $\|K\|_1 = 1$ y $\|K\|_2 = \frac{1}{\sqrt{2\sqrt{\pi}}}$.

El estimador GL de la función de regresión r en el intervalo $[-1, 1]$, esta dado por $\hat{r}_{\hat{h}_i/\hat{h}_i^*}(x_i) = \frac{\hat{m}_{\hat{h}_i}(x_i)}{\hat{g}_{\hat{h}_i^*}(x_i)}$ en cada x_i , para $i = 1, \dots, s$, donde $\{\hat{m}_{\hat{h}_i}(x_i)\}_{i=1}^s$ y $\{\hat{g}_{\hat{h}_i^*}(x_i)\}_{i=1}^s$ son los estimadores GL de m y g respectivamente en el intervalo $[-1, 1]$, los cuales se determinan de la siguiente forma:

La ventana \hat{h}_i^* se selecciona de la familia de ventanas \mathcal{H} , tomando

$$\hat{h}_i^* = \arg \min_{h \in \mathcal{H}} \{A^*(h, x_i) + V^*(h)\} \quad (4.15)$$

para cada $i = 1, \dots, s$, donde $A^*(h, x_i) = \max_{h' \in \mathcal{H}} \{|\hat{g}_{h,h'}(x_i) - \hat{g}_{h'}(x_i)| - V^*(h')\}_+$ con $\hat{g}_{h,h'}(x_i) = \sum_{j=1}^n K_h * K_{h'}(x_i - X_j)$ y $V^*(h) = \sqrt{2\gamma^* \bar{g}_{sup}} \|K\|_2 (\|K\|_1 + 1) (1 + \delta_n^*) \frac{(\log n)^{-\frac{1}{2}}}{(nh)^{\frac{1}{2}}}$, donde $\delta_n^* = (\log n)^{\frac{1}{2}}$ y \bar{g}_{sup} es el maximo del histograma de la muestra $\{X_i\}_{i=1}^n$, construido a partir del comando hist de R.

Observación 8. En la ecuación (4.10) se sustituyo $\hat{g}_{h,h'}$ en lugar de $\hat{g}_{h \vee h'}$, posteriormente se acoto $\hat{J}(h) = \frac{1}{n^2} \sum_{i=1}^n K_h^2(x - X_i)$, esa cota se sustituyo en la ecuación (4.11) y se trabajo esa ecuación con el objetivo de que $V^*(h)$ tuviera una expresión equivalente a la ecuación (4.9).

La ventana \hat{h}_i se selecciona de la familia de ventanas \mathcal{H} , tomando

$$\hat{h}_i = \arg \min_{h \in \mathcal{H}} \{A(h, x_i) + V(h, x_i)\} \quad (4.16)$$

para cada $i = 1, \dots, s$, donde $A(h, x_i) = \max_{h' \in \mathcal{H}} \{|\hat{m}_{h, h'}(x_i) - \hat{m}_{h'}(x_i)| - V(h', x_i)\}_+$ con $\hat{m}_{h, h'}(x_i) = \sum_{j=1}^n K_h * K_{h'}(x_i - X_j) Y_j$ y $V(h, x_i) = \sqrt{2\gamma A_1(i)} (\|K\|_1 + 1) (1 + \delta_n) \frac{(\log n)^{\frac{1}{2}}}{(nh)^{\frac{1}{2}}}$, donde $\delta_n = (\log n)^{-\frac{1}{5}}$, $A_1(i) = ((\hat{r}_{sup}(i))^2 + \hat{\sigma}^2) \hat{g}_{sup} \|K\|_2^2$, $\hat{g}_{sup} = \max\{\hat{g}_{\hat{h}_i}(x_i) | i = 1, \dots, s\}$, $\hat{r}_{sup}(i) = \max\{|Y_j| \mathbf{1}_{|x_i - 0.5, x_i + 0.5|}(X_j) | j = 1, \dots, n\}$ y $\hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^n (Y_i - \tilde{r}_{\bar{h}}(X_i))^2$ donde $\tilde{r}_{\bar{h}}(\cdot)$ es el estimador de Nadaraya-Watson, en la muestra $\{(X_i, Y_i)\}_{i=1}^n$ y la ventana \bar{h} , con \bar{h} la ventana resultante del comando $dpill(\cdot, \cdot)$ en R correspondiente a la librería *KernSmooth*, método plug-in de Ruppert, Sheather y Wand (1995).

Los parámetros γ^* y γ son parámetros de calibración del método, cuyos valores se determinan según la metodología planteada en la Sección 4.4.3.

En la Figura 4.2, se muestra la función de regresión r y tres estimadores GL de la función de regresión para muestras aleatorias $\{(Y_k, X_k)\}_{k=1}^{n+2q}$, con $n = 1000, 2000$ y 5000 tomando $q = 100$ y $\sigma = 0,5$.

4.4.3. Calibración del método para una muestra.

Al calibrar el método, para una muestra $\{(X_i, Y_i)\}_{i=1}^{n+2q}$, se toman los primeros n datos de la muestra para construir estimadores de g y m : \hat{g}_{h^*} y \hat{m}_{h^*} , $h \in \mathcal{H}$. Con los últimos q datos de la muestra $\{(X_i, Y_i)\}_{i=n+q+1}^{n+2q}$ se construye una malla aleatoria del intervalo $[-2, 2]$ reordenando sus elementos de forma creciente respecto a la primera componente, es decir $\{(X_i, Y_i)\}_{i=n+q+1}^{n+2q} = \{(\tilde{x}_i, \tilde{y}_i)\}_{i=1}^q$, donde $\tilde{x}_i < \tilde{x}_j$ si $i < j$. La calibración del método se realiza utilizando la malla determinada por $\{\tilde{x}_i\}_{i=1}^q$ en el intervalo $[-2, 2]$.

El estimador de la función de regresión r se calcula en cada punto de la malla $\{\tilde{x}_i\}_{i=1}^q$ del intervalo $[-2, 2]$, y para cada \tilde{x}_i se seleccionan dos ventanas óptimas \hat{h}_i y \hat{h}_i^* de la familia de ventanas \mathcal{H} , donde la ventana \hat{h}_i se obtiene según la ecuación (4.16) con $V(h, x_i)$ dependiente del parámetro γ y la ventana \hat{h}_i^* se obtiene según la ecuación (4.15) con $V^*(h, x_i)$ dependiente del parámetro γ^* . Por tal motivo las ventanas óptimas \hat{h}_i y \hat{h}_i^* dependen de los parámetros γ y γ^* respectivamente, tal dependencia se denota por \hat{h}_i^γ y $\hat{h}_i^{\gamma^*}$. En la práctica para calibrar el método se toman γ y γ^* en una malla igualmente espaciada de un subintervalo I de $]0, 1/10[$

Estimación Adaptativa Método GL

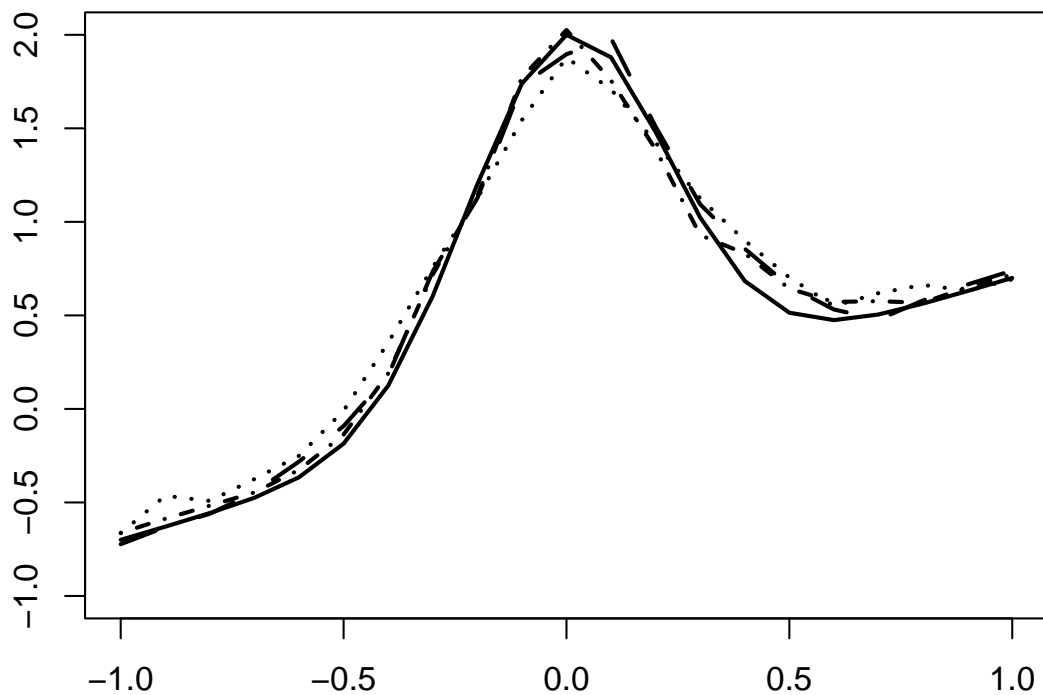


Figura 4.2 Se muestran tres estimadores GL de la función de regresión r , para muestras aleatorias $\{(Y_k, X_k)\}_{k=1}^{n+2q}$, con $\sigma = 0,5$ y $q = 100$. Cuando $n = 1000$ el estimador GL se representa con puntos, para $n = 2000$ con rayas largas y para $n = 5000$ con puntos y rayas cortas. Además se representa con una línea continua la función de regresión r .

y la elección del intervalo I se toma de forma tal que las curvas descritas por los estimadores $\left\{\hat{m}_{\hat{h}_i^\gamma}(x_i)\right\}_{i=1}^q$ y $\left\{\hat{g}_{\hat{h}_i^{\gamma^*}}(x_i)\right\}_{i=1}^q$, pasen de ser curvas muy irregulares a curvas suaves, cuando se varían los valores de γ y γ^* en el intervalo I .

El método de calibración se realiza en tres etapas, primero elegir el γ^* adecuado en I de forma tal que $\left\{\hat{g}_{\hat{h}_i^{\gamma^*}}(x_i)\right\}_{i=1}^q$ sea un buen estimador de $\{g(x_i)\}_{i=1}^q$, segundo elegir el γ adecuado en I de forma tal que $\left\{\hat{m}_{\hat{h}_i^\gamma}(x_i)\right\}_{i=1}^q$ sea un buen estimador de $\{m(x_i)\}_{i=1}^q$ y finalmente con los resultados obtenidos en las etapas 1 y 2, construir el estimador GL de la función de regresión r .

Etapla 1 (Elección de γ^*): Usando el comando `hist` de R, se construye un histograma con los datos $\{X_i\}_{i=1}^n$ y con esa información se define la función histograma $H_g(x)$ para cada $x \in [-2, 2]$. Posteriormente se determina $Error_g(\gamma^*) = \sum_{i=1}^q d_i \left(\hat{g}_{\hat{h}_i^{\gamma^*}}(\tilde{x}_i) - H_g(\tilde{x}_i)\right)^2$ para cada $\gamma^* \in I$, donde $d_1 = \frac{\tilde{x}_1 + \tilde{x}_2}{2} + 2$, $d_q = 2 - \frac{\tilde{x}_{q-1} + \tilde{x}_q}{2}$ y $d_i = \frac{\tilde{x}_{i+1} - \tilde{x}_{i-1}}{2}$, para $i \in \{2, \dots, q-1\}$. Finalmente, la primera etapa de calibración concluye al tomar $\gamma_g = \arg \min_{\gamma^* \in I} \{Error_g(\gamma^*)\}$.

Etapla 2 (Elección de γ): Como el modelo de estudio es $Y_i = \frac{m(X_i)}{g(X_i)} + \varepsilon_i$, $i = 1, \dots, n$, entonces $m(X_i) \approx Y_i g(X_i)$, por tal motivo en esta etapa de calibración se usan los últimos q datos reordenados de la muestra $\{(\tilde{x}_i, \tilde{y}_i)\}_{i=1}^q$ y el estimador $\left\{\hat{g}_{\hat{h}_i^{\gamma_g}}(\tilde{x}_i)\right\}_{i=1}^q$ obtenido en la etapa anterior. Se determina $Error_m(\gamma) = \sum_{i=1}^q d_i \left(\hat{m}_{\hat{h}_i^\gamma}(\tilde{x}_i) - \tilde{y}_i \hat{g}_{\hat{h}_i^{\gamma_g}}(\tilde{x}_i)\right)^2$ para cada $\gamma \in I$, donde $\{d_i\}_{i=1}^q$ se define igual que en la etapa 1. Finalmente, esta etapa de calibración concluye al tomar $\gamma_m = \arg \min_{\gamma \in I} \{Error_m(\gamma)\}$.

Etapla 3 (Estimador de r): Para la estimación GL de la función de regresión $r = \frac{m}{g}$ en la malla igualmente espaciada $\{x_i\}_{i=1}^s$ del intervalo $[-1, 1]$, donde $x_i = -1 + 2 \left(\frac{i-1}{s-1}\right)$ para $i = 1, \dots, s$ con $s = 21$, se considera $\left\{\hat{r}_{\hat{h}_i^{\gamma_m} / \hat{h}_i^{\gamma_g}}(x_i)\right\}_{i=1}^s$ y para simplificar la notación cuando se dice que el estimador GL esta calibrado se omite la dependencia de las ventanas óptimas a γ_m y γ_g , escribiendo simplemente el estimador GL calibrado de la siguiente forma $\left\{\hat{r}_{\hat{h}_i / \hat{h}_i^*}(x_i)\right\}_{i=1}^s$.

En la practica se tomó $I =]0,00000005, 0,08[$, además, sobre tal intervalo se considera la malla igualmente espaciada $\gamma_i = 0,00000005 + (i-1) \frac{0,08 - 0,00000005}{20}$ para $i = 1, \dots, 21$.

En la Figura 4.3, se puede apreciar que el gráfico de la función de densidad g no abarca todo el intervalo $[-2, 2]$ esto se debe a que tal función se evaluó en la malla aleatoria $\{\tilde{x}_i\}_{i=1}^q$

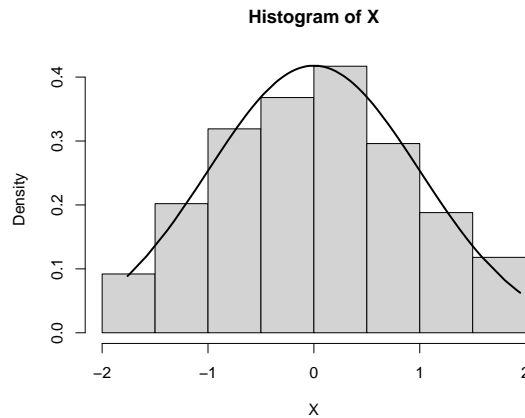


Figura 4.3 Histograma realizado con la muestra $\{X_i\}_{i=1}^n$ con $n = 2000$ y función de densidad g evaluada en la malla aleatoria $\{\tilde{x}_i\}_{i=1}^q$, con $q = 100$.

y para este caso pocas muestras fueron sorteadas en los extremos del intervalo. Este hecho pone en evidencia posibles problemas de estimación de la función de regresión r en los extremos del intervalo $[-2, 2]$.

La Figura 4.4, en el gráfico de arriba, muestra que $Error_g(\gamma^*)$ se minimiza en γ_9^* . Además, en el gráfico de abajo, se muestra con una línea continua la función de densidad g y con puntos el estimador $\hat{g}_{\hat{h}_i^*}(x_i)$ calibrado en γ_9^* y evaluado en la malla $\{x_i\}_{i=1}^{41}$ del intervalo $[-2, 2]$, donde $x_i = -2 + 0,1(i - 1)$. En ambos se tomo $n = 2000$, $q = 100$ y $\sigma = 0,5$.

La Figura 4.5, en el gráfico de arriba, muestra que $Error_m(\gamma)$ se minimiza en γ_2 . Además, en el gráfico de abajo, muestra la nube de puntos $\left\{ \left(\tilde{x}_i, \tilde{y}_i \hat{g}_{\hat{h}_i^*}(\tilde{x}_i) \right) \right\}_{i=1}^q$, se representa con una línea continua la función m y con puntos el estimador $\hat{m}_{\hat{h}_i}(x_i)$ calibrado en γ_2 y evaluado en la malla $\{x_i\}_{i=1}^{41}$ del intervalo $[-2, 2]$, donde $x_i = -2 + 0,1(i - 1)$. En ambos se tomo $n = 2000$, $q = 100$ y $\sigma = 0,5$.

En la Figura 4.6, superior se muestra la nube de puntos $\{(X_i, Y_i)\}_{i=1}^n$, se representa con una línea continua la función de regresión r y con puntos el gráfico de $\{\hat{r}_{\hat{h}_i/\hat{h}_i^*}(x_i)\}_{i=1}^{41}$ en el intervalo $[-2, 2]$, donde $x_i = -2 + 0,1(i - 1)$. Inferior se muestra la nube de puntos $\{(X_i, Y_i)\}_{i=1}^n$, se representa con una línea continua la función de regresión r y con puntos el gráfico de $\{\hat{r}_{\hat{h}_i/\hat{h}_i^*}(x_i)\}_{i=1}^{21}$ en el intervalo $[-1, 1]$, donde $x_i = -1 + 0,1(i - 1)$. En ambos casos se tomo $n = 2000$, $q = 100$ y $\sigma = 0,5$. Se observa que el estimador de la función de regresión se aleja de la función a estimar en los extremos del intervalo $[-2, 2]$ por tal motivo al final se da la estimación de la función de regresión en el intervalo $[-1, 1]$ y los valores del $M\hat{S}E$ y $M\hat{I}S\hat{E}$ que se proporcionan en la siguiente sección se calculan en el intervalo $[-1, 1]$.

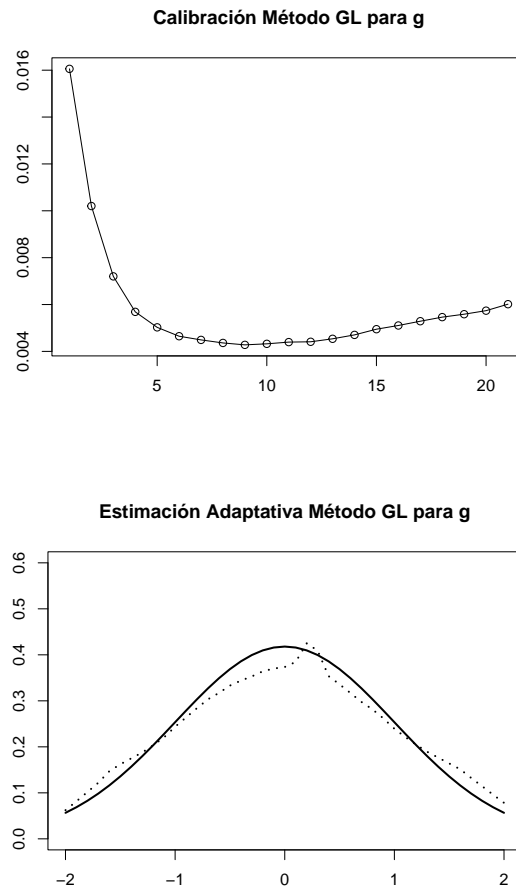


Figura 4.4 Arriba gráfico de $\{Error_g(\gamma_i^*)\}_{i=1}^{21}$ y abajo el gráfico de g es representado con una línea continua y representado con puntos el gráfico de $\{\hat{g}_{\hat{h}_i^*}(x_i)\}_{i=1}^{41}$ en el intervalo $[-2, 2]$, calibrado en γ_9^* , que es donde el gráfico de la izquierda alcanza su mínimo.

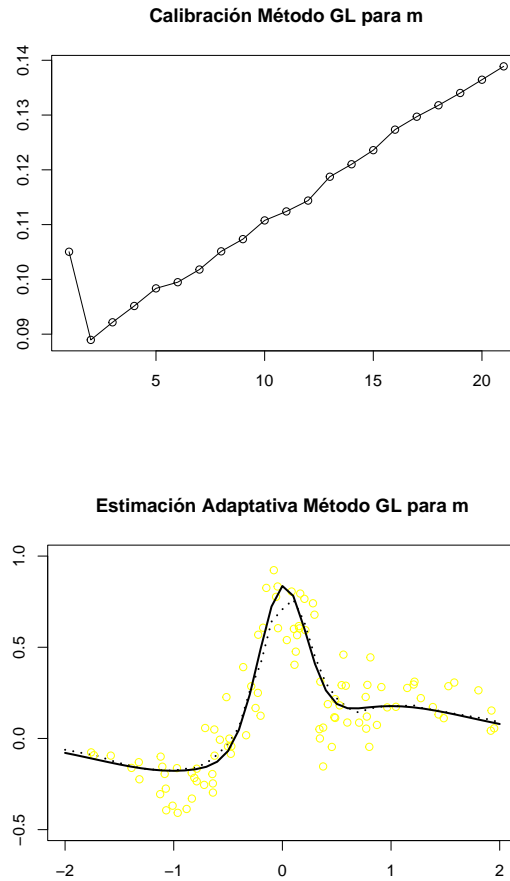


Figura 4.5 Arriba gráfico de $\{Error_m(\gamma)\}_{i=1}^{21}$ y abajo se muestra la nube de puntos $\left\{ \left(\tilde{x}_i, \tilde{y}_i \hat{g}_{\hat{h}_i^{* \gamma_2}}(\tilde{x}_i) \right) \right\}_{i=1}^q$, se representa con una línea continua el gráfico de m en el intervalo $[-2, 2]$ y con puntos el gráfico de $\{\hat{m}_{\hat{h}_i}(x_i)\}_{i=1}^{41}$, calibrado en γ_2 , que es donde el gráfico de la izquierda alcanza su mínimo.

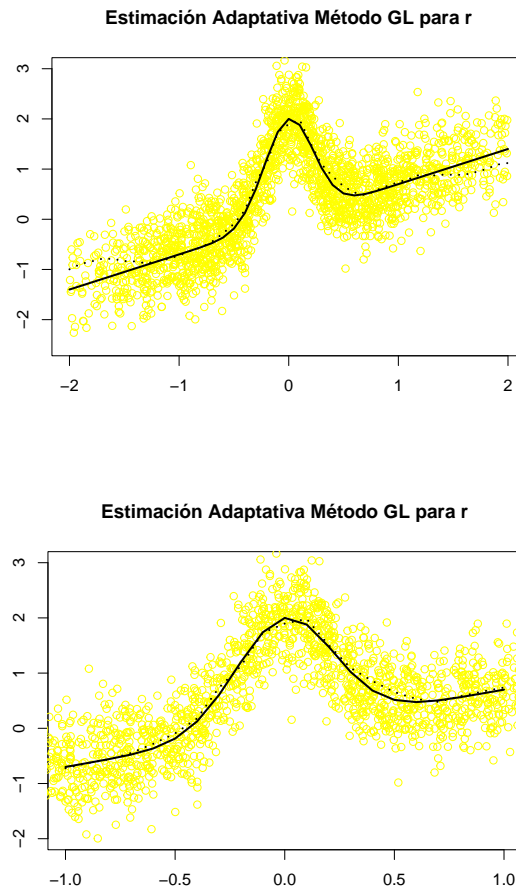


Figura 4.6 La figura de arriba muestra la nube de puntos $\{(X_i, Y_i)\}_{i=1}^n$, se representa con una línea continua la función de regresión r y con puntos el gráfico de $\{\hat{r}_{\hat{h}_i/\hat{h}_i^*}(x_i)\}_{i=1}^{41}$ en el intervalo $[-2, 2]$, donde $x_i = -2 + 0,1(i - 1)$ y la figura de abajo muestra las mismas representaciones pero en la malla $x_i = -1 + 0,1(i - 1)$, para $i = 1, \dots, 21$, del intervalo $[-1, 1]$. Ambas estimaciones se calibraron en $\gamma_g = \gamma_9^*$ y $\gamma_m = \gamma_2$. Además se tomo $n = 2000$, $q = 100$ y $\sigma = 0,5$.

4.4.4. Comparación de errores globales y locales empíricos, para distintos tamaños de muestra y valores de σ .

Para mostrar la calidad del estimador de la función de regresión r , se generan $N = 500$ replicas de la muestra $\{(X_i, Y_i)\}_{i=1}^{n+2q}$. Posteriormente se calculan N estimadores GL ya calibrados $\{\hat{r}^{(1)}(x_i)\}_{i=1}^s, \dots, \{\hat{r}^{(N)}(x_i)\}_{i=1}^s$, con $x_i = -1 + 2\frac{i-1}{s-1}$, $s = 21$ y $\hat{r}^{(j)}(x_i) = \hat{h}_i/\hat{h}_i^*(x_i)$ donde \hat{h}_i y \hat{h}_i^* se obtuvieron para la j -ésima replica por el método GL. El error cuadrático medio (MSE) y el error cuadrático medio integrado ($MISE$) se estiman de la siguiente forma

$$\hat{MSE}(x_i) = \frac{1}{N} \sum_{j=1}^N \left[\left(r(x_i) - \hat{r}^{(j)}(x_i) \right)^2 \right] \text{ para } i = 1, \dots, s$$

$$\hat{MISE} = \frac{1}{N} \sum_{j=1}^N I_j$$

con $I_j = \sum_{i=1}^s 0,1 \left(r(x_i) - \hat{r}^{(j)}(x_i) \right)^2$ para $j = 1, \dots, N$.

En la Figura 4.7 se representan tres boxplot de I_j para $j = 1, \dots, N$, con los tamaños de muestra $n = 1000, 2000$ y 5000 de izquierda a derecha respectivamente, además sobre cada boxplot hay un punto que representa el \hat{MISE} (superior) y tres representaciones del $\hat{MSE}(x_i)$ para $i = 1, \dots, s$ con una línea para el tamaño de muestra $n = 1000$, con puntos para $n = 2000$ y con rayas para $n = 5000$, (inferior). En ambos casos para $\sigma = 0,1$. En las Figuras 4.8 y 4.9 se hacen las mismas representaciones pero con $\sigma = 0,5$ y $\sigma = 1$ respectivamente.

Al analizar el gráfico que corresponde al \hat{MSE} con $\sigma = 0,1$ en la Figura 4.7 (gráficos de abajo) se observa que los valores del \hat{MSE} son mas elevados para valores del dominio de r cercanos a donde se alcanza el máximo y donde hay cambio de concavidad, ese patrón se observa con tamaños de muestra $n = 1000, 2000$ y 5000 . Como es evidente a medida que aumenta el tamaño de la muestra los valores del \hat{MSE} disminuyen, en mayor medida para valores cercanos a donde se alcanza el máximo y en menor medida en valores cercanos donde se da el cambio de concavidad. Comportamientos similares se observan al analizar el \hat{MSE} para $\sigma = 0,5$ y $\sigma = 1$ de las Figuras 4.8 y 4.9 respectivamente (gráficos de abajo). Como es evidente cuando aumenta el valor de σ también aumentan los valores del \hat{MSE} .

En los diagramas de caja de los valores de I_j para $j = 1, \dots, N$ con $\sigma = 0,1$ correspondientes a la Figura 4.7 (gráficos de arriba) se observa que a medida que los valores de n aumentan de $1000, 2000$ a 5000 la dispersión de los $\{I_j\}_{j=1}^N$ disminuye, también disminuye la cantidad de datos atípicos y el valor del \hat{MISE} se aproxima cada vez mas al valor de la mediana de los $\{I_j\}_{j=1}^N$, aunque en general la mediana siempre es inferior al \hat{MISE} . En los

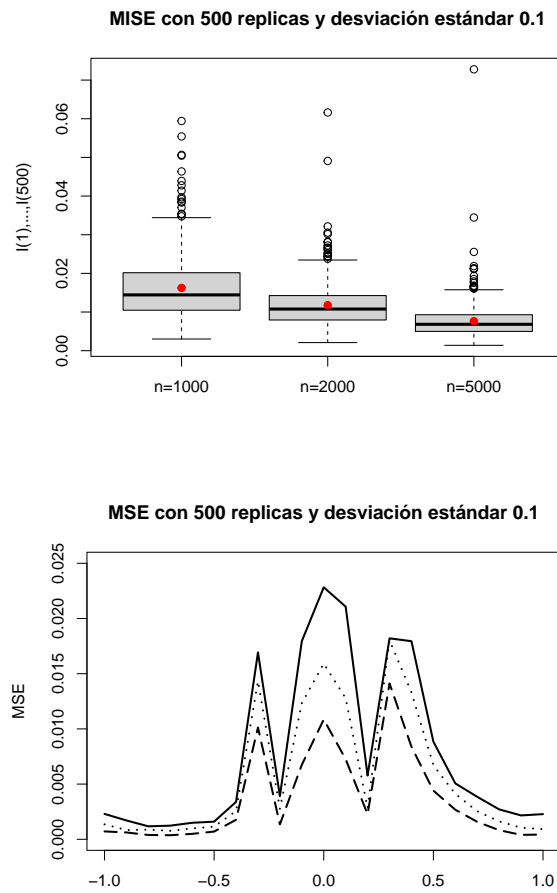


Figura 4.7 Arriba se representan tres boxplot de I_j para $j = 1, \dots, N$, con los tamaños de muestra $n = 1000, 2000$ y 5000 de izquierda a derecha respectivamente, sobre cada boxplot hay un punto que representa el \hat{MISE} y abajo se muestran tres representaciones del $\hat{MSE}(x_i)$ para $i = 1, \dots, s$ con una línea para el tamaño de muestra $n = 1000$, con puntos para $n = 2000$ y con rayas para $n = 5000$. En ambos casos para $\sigma = 0,1$.

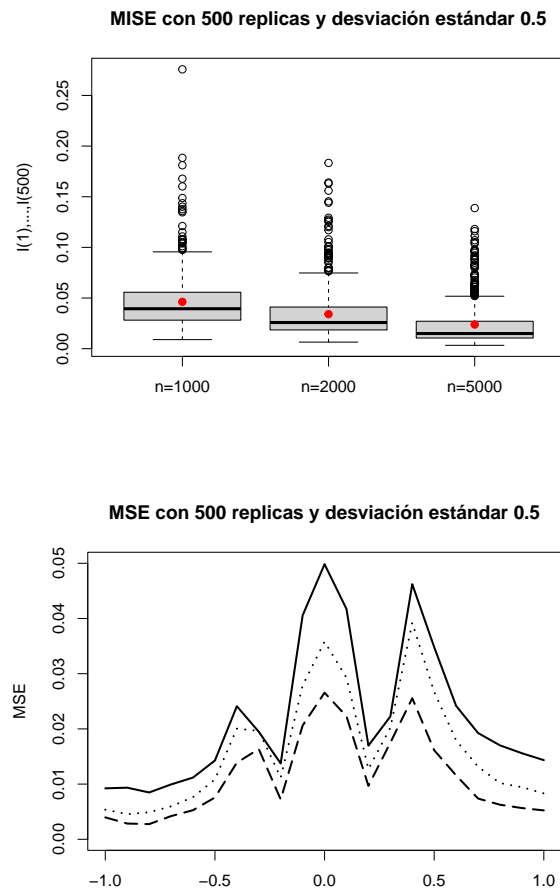


Figura 4.8 Arriba se representan tres boxplot de I_j para $j = 1, \dots, N$, con los tamaños de muestra $n = 1000, 2000$ y 5000 de izquierda a derecha respectivamente, sobre cada boxplot hay un punto que representa el $M\hat{I}SE$ y abajo se muestran tres representaciones del $M\hat{S}E(x_i)$ para $i = 1, \dots, s$ con una línea para el tamaño de muestra $n = 1000$, con puntos para $n = 2000$ y con rayas para $n = 5000$. En ambos casos para $\sigma = 0,5$.

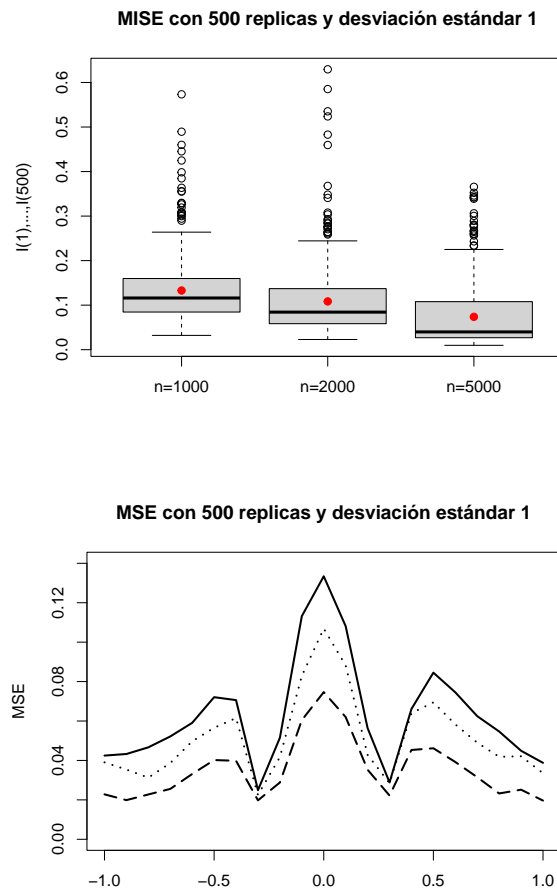


Figura 4.9 Arriba se representan tres boxplot de I_j para $j = 1, \dots, N$, con los tamaños de muestra $n = 1000, 2000$ y 5000 de izquierda a derecha respectivamente, sobre cada boxplot hay un punto que representa el \hat{MISE} y abajo se muestran tres representaciones del $\hat{MSE}(x_i)$ para $i = 1, \dots, s$ con una línea para el tamaño de muestra $n = 1000$, con puntos para $n = 2000$ y con rayas para $n = 5000$. En ambos casos para $\sigma = 1$.

| | $n = 1000$ | $n = 2000$ | $n = 5000$ |
|----------------|---------------|------------|-------------|
| $\sigma = 0,1$ | 0.01624254 | 0.01177675 | 0.007650944 |
| $\sigma = 0,5$ | 0.04625373 | 0.03409679 | 0.023852842 |
| $\sigma = 1$ | 0.13286962055 | 0.10837311 | 0.073751940 |

Cuadro 4.1 Valores del $M\hat{I}SE$ para distintos valores de σ y tamaños de la muestra, caso g desconocido.

diagrama de caja de las Figuras 4.8 y 4.9 correspondientes a $\sigma = 0,5$ y $\sigma = 1$ respectivamente se observa que a medida que aumentan los valores de σ aumenta la dispersión de los $\{I_j\}_{j=1}^N$, también aumenta la distancia entre el $M\hat{I}SE$ y la mediana de los $\{I_j\}_{j=1}^N$ y que en general la mediana siempre es inferior al valor del $M\hat{I}SE$.

Al comparar los resultados obtenidos en las Figuras 4.7, 4.8 y 4.9 con las Figuras equivalentes del caso g conocido, se puede observar que los valores del $M\hat{I}SE$ disminuyen a la mitad en el caso g desconocido, cuando $\sigma = 0,1$ y $0,5$ con n tomando los valores 1000, 2000 y 5000, de igual forma, en el caso g desconocido, los valores del $M\hat{I}SE$ disminuyen aunque en menor medida cuando $\sigma = 1$ con n tomando los valores 1000, 2000 y 5000. También se puede apreciar que en el caso g desconocido es mucho mas evidente los problemas que presenta el estimador de r al estimar valores cercanos a los puntos donde hay cambio de concavidad, pero de igual manera todos los valores del $M\hat{I}SE$ en el caso g desconocido son mucho mejores que los del caso g conocido. En los Diagramas de Caja el caso g desconocido también presenta mejores resultados.

Como se observa en la tabla 4.1 el valor del $M\hat{I}SE$ es inversamente proporcional a n y directamente proporcional a σ . Al comparar estos resultados con los obtenidos en la tabla 3.1 del caso g conocido, se evidencia que los valores del $M\hat{I}SE$ en el caso g desconocido disminuyen a la mitad cuando $\sigma = 0,1$ y n toma los valores 1000, 2000 y 5000, que igual disminuyen de forma considerable pero en menor medida cuando $\sigma = 0,5$ y n toma los valores 1000, 2000 y 5000, y finalmente se observa que en ambos casos los valores del $M\hat{I}SE$ son similares cuando $\sigma = 1$ y n toma los valores 1000, 2000 y 5000.

En general se puede concluir que el estimador de la función de regresión r en el caso g desconocido tiene mejor rendimiento que en el caso g conocido.

4.5. Demostraciones

4.5.1. Constantes.

Recordamos algunos valores de constantes dados previamente y además a continuación se declaran constantes que usaremos en el desarrollo de las demostraciones.

$$\begin{aligned}
 A_0 &= \frac{L_1}{l_1!} \int_{\mathbb{R}} |u|^{\beta_1} |K(u)| du, \quad A_1 = ((r_{\text{sup}})^2 + \sigma^2) g_{\text{sup}} \|K\|_2^2 = B_1 + B_2, \\
 A_2 &= \frac{2A_3}{|\log a|} + 20A_4, \quad A_3 = (r_{\text{sup}})^2 \|K\|_1^2 (Q + (g_{\text{sup}})^2), \quad A_4 = \frac{4}{1-a} (r_{\text{sup}})^2 \|K\|_{\infty}^2, \\
 C &= A_0^2 + A_1 + A_2, \quad \mathfrak{A}_0 = \frac{L_2}{l_2!} \int_{\mathbb{R}} |u|^{\beta_2} |K(u)| du, \quad \mathfrak{A}_1 = g_{\text{sup}} \|K\|_2^2, \\
 \mathfrak{A}_2 &= \frac{2\mathfrak{A}_3}{|\log a|} + 20\mathfrak{A}_4, \quad \mathfrak{A}_3 = \|K\|_1^2 (Q + (g_{\text{sup}})^2), \quad \mathfrak{A}_4 = \frac{4}{1-a} \|K\|_{\infty}^2, \\
 \mathfrak{C} &= \mathfrak{A}_0^2 + \mathfrak{A}_1 + \mathfrak{A}_2, \quad B_1 = \sigma^2 g_{\text{sup}} \|K\|_2^2, \quad B_2 = (r_{\text{sup}})^2 g_{\text{sup}} \|K\|_2^2, \\
 B_3 &= \left(r_{\text{sup}} + \frac{2\sigma}{\sqrt{2\pi}} \right) g_{\text{sup}} \|K\|_1, \quad B_4 = \left(r_{\text{sup}} + \frac{2\sigma}{\sqrt{2\pi}} \right)^2 Q \|K\|_1^2, \\
 B_5 &= \left(r_{\text{sup}} + \frac{2\sigma}{\sqrt{2\pi}} \right)^2 \|K\|_1^2 (Q + (g_{\text{sup}})^2), \quad B_6 = 2\|K\|_{\infty}, \quad B_7 = (4B_5 B_6 (\sigma + r_{\text{sup}}))^{\frac{2}{3}}, \\
 B_8 &= \frac{2B_5}{|\log a|} + \frac{18B_7}{1-a^{\frac{1}{3}}}, \quad B = \frac{12(B_6 \vee \sqrt{B_7})}{1 \wedge |\log a|} \left(\frac{2^7 3 B_7 a^{-\frac{1}{3}}}{A_1 (1 \wedge |\log a|)} \vee 1 \right), \\
 B_9 &= \frac{2^2 \|K\|_{\infty}^2}{(2\pi)^{\frac{1}{4}}} ((r_{\text{sup}})^4 + 3\sigma^4)^{\frac{1}{2}}, \quad B_{10} = \frac{2^4 3}{1 - e^{-\frac{\gamma-2}{2}}} (A_1 + B_8) + \frac{5! 2^{18} 3 B^2}{1 - e^{-\frac{\gamma-2}{2}}} (\sigma + r_{\text{sup}})^2, \\
 A_5 &= B_9 + 2B_{10}, \quad A_6 = 1 + 2\|K\|_1, \quad A_7 = \sqrt{A_1 + A_2} + 2\sqrt{2\gamma A_1} (1 + \|K\|_1), \\
 A_8 &= 2\sqrt{A_5} (1 + \|K\|_1), \quad C^* = A_0 A_6 + 2e^{\frac{1}{2}} A_7 + A_8, \\
 \bar{F}_1 &= 2B_{11} (r_{\text{sup}} g_{\text{sup}})^2 + 2\|K\|_{\infty}^2 B_{11}^{1/2} (2^3 (r_{\text{sup}}^4 + 3\sigma^4))^{1/2}, \quad \bar{F}_2 = 2B_{11} \{ (\|K\|_{\infty}^2 \vee 1) + (g_{\text{sup}})^2 \}, \\
 F_1 &= \frac{8}{(g_{\text{inf}})^2}, \quad F_2 = \frac{8(r_{\text{sup}})^2}{(g_{\text{inf}})^2}, \quad F_3 = 2\bar{F}_1 + 2(r_{\text{sup}})^2 \bar{F}_2, \quad \mathfrak{D} = F_1 C + F_2 \mathfrak{C} + F_3, \\
 \mathfrak{B} &= F_1 (C^*)^2 + F_2 (C^{**})^2 + F_3.
 \end{aligned}$$

4.5.2. Enunciados de resultados técnicos.

Lemma 4. *Bajo las hipótesis (H₁), (H₂), (H₄) y (H₅) se tiene que:*

$$i) \mathbb{E} \left[(K_h(x - X_i) \varepsilon_i)^2 \right] \leq \frac{B_1}{h}$$

$$ii) \mathbb{E} \left[(K_h(x - X_i)r(X_i))^2 \right] \leq \frac{B_2}{h}$$

$$iii) \mathbb{E} \left[|K_h(x - X_i)K_h(x - X_j)| \right] \leq Q \|K\|_1^2$$

$$iv) \mathbb{E} \left[|K_h(x - X_i)r(X_i)| \right] \leq r_{\text{sup}} g_{\text{sup}} \|K\|_1,$$

donde $B_1 = \sigma^2 g_{\text{sup}} \|K\|_2^2$ y $B_2 = (r_{\text{sup}})^2 g_{\text{sup}} \|K\|_2^2$.

Lemma 5. Para cada $h \in \mathcal{H}$, $a \in]0, 1[$ y $n \geq 4$, se cumple que,

$$\frac{1}{h} a \left\{ \frac{1}{h |\log a| (\log n)^{\frac{1}{2}}} \right\} \leq \frac{10}{(\log n)^{\frac{5}{2}}}$$

Proposición 21. Asumimos que los X_i satisfacen la hipótesis (H_3) . Sea $k, u, v \in \mathbb{N}$, tales que $i_{1:u} \in \mathbb{Z}^u$, $j_{1:v} \in \mathbb{Z}^v$ y tal que $i_1 \leq \dots \leq i_u < i_u + k \leq j_1 \leq \dots \leq j_v$. Si $G_u : \mathbb{R}^u \times \mathbb{R}^u \rightarrow \mathbb{R}$ y $G_v : \mathbb{R}^v \times \mathbb{R}^v \rightarrow \mathbb{R}$ son funciones tales que $\|G_u\|_\infty < \infty$ y $\|G_v\|_\infty < \infty$ entonces,

$$|\text{cov}(G_u(X_{i_{1:u}}, \varepsilon_{i_{1:u}}), G_v(X_{j_{1:v}}, \varepsilon_{j_{1:v}}))| \leq \Psi(u, v, G_u, G_v) a^k$$

donde $\Psi(u, v, G_u, G_v) = 4 \|G_u\|_\infty \|G_v\|_\infty$, con $\|G_u\|_\infty = \sup_{(x,y) \in \mathbb{R}^u \times \mathbb{R}^u} |G_u(x,y)|$ y $\|G_v\|_\infty = \sup_{(x,y) \in \mathbb{R}^v \times \mathbb{R}^v} |G_v(x,y)|$.

Como en el capítulo 3, a continuación vamos a trabajar con funciones G_u específicas. Mas precisamente consideramos para $u \in \mathbb{N}$, y $i_{1:u} \in \mathbb{Z}^u$, $G_u : \mathbb{R}^u \times \mathbb{R}^u \rightarrow \mathbb{R}$ dada por

$$G_u(X_{i_{1:u}}, \varepsilon_{i_{1:u}}) = \prod_{k=1}^u \{G(X_{i_k}, \varepsilon_{i_k}) - \mathbb{E}(G(X_{i_k}, \varepsilon_{i_k}))\}$$

donde para $i \in \mathbb{N}$,

$$G(X_i, \varepsilon_i) = \frac{1}{n} (r(X_i) + \varepsilon_i) K_h(x - X_i) \mathbf{1}_{\{|(r(X_i) + \varepsilon_i)| \leq M_n\}},$$

donde $M_n = \sigma \log n + r_{\text{sup}}$. Las variables $G(X_i, \varepsilon_i)$ satisfacen los siguientes resultados.

Lemma 6. Bajo las hipótesis (H_1) , (H_2) , (H_4) y (H_5) se tiene que: $\forall i, k \in \mathbb{N}$

$$i) |\mathbb{E}[G(X_i, \varepsilon_i)]| \leq \frac{B_3}{n},$$

$$ii) |\mathbb{E}[G(X_i, \varepsilon_i)G(X_{i+k}, \varepsilon_{i+k})]| \leq \frac{B_4}{n^2},$$

$$iii) |\text{Cov}(G(X_i, \varepsilon_i), G(X_{i+k}, \varepsilon_{i+k}))| \leq \frac{B_5}{n^2},$$

$$iv) |G(X_i, \varepsilon_i)| \leq \frac{B_6 M_n}{2nh}$$

donde

$$B_3 = \left(r_{\text{sup}} + \frac{2\sigma}{\sqrt{2\pi}} \right) g_{\text{sup}} \|K\|_1, \quad B_4 = \left(r_{\text{sup}} + \frac{2\sigma}{\sqrt{2\pi}} \right)^2 Q \|K\|_1^2$$

$$B_5 = \left(r_{\text{sup}} + \frac{2\sigma}{\sqrt{2\pi}} \right)^2 \|K\|_1^2 (Q + (g_{\text{sup}})^2) \quad \text{y} \quad B_6 = 2\|K\|_\infty.$$

Proposición 22. Sean $u, v, k \in \mathbb{N}$, si $(i_1, \dots, i_u, j_1, \dots, j_v) \in \mathbb{Z}^{u+v}$ es tal que $i_1 \leq \dots \leq i_u < i_u + k \leq j_1 \leq \dots \leq j_v$. Bajo las hipótesis (H_1) , (H_2) , (H_3) , (H_4) y (H_5) tenemos, $\forall h \in \mathcal{H}$

$$\gamma_h(u, v) = |\text{cov}(G_u(X_{i_{1:u}}, \varepsilon_{i_{1:u}}), G_v(X_{j_{1:v}}, \varepsilon_{j_{1:v}}))| \leq \phi(u, v) (D_1(h))^{u+v-2} D_2(h) a^{\frac{k}{3}}$$

donde $\phi(u, v) = u + v + uv$,

$$D_1(h) = \frac{B_6 M_n}{nh}, \quad D_2(h) = \frac{B_7}{n^2 h} \quad \text{y} \quad B_7 = (4B_5 B_6 (\sigma + r_{\text{sup}}))^{\frac{2}{3}}.$$

Proposición 23. Desigualdad de Bernstein. Sea el proceso $\mathbb{Z} = (Z_i)_{i \in \mathbb{N}}$, donde $Z_i = G(X_i, \varepsilon_i) - \mathbb{E}[G(X_i, \varepsilon_i)]$ y $S_n = \sum_{i=1}^n Z_i$. Bajo las hipótesis (H_1) , (H_2) , (H_3) , (H_4) y (H_5) tenemos para $h \in \mathcal{H}$ y $\forall t \geq 0$

$$\mathbb{P}(|S_n| \geq t) \leq \exp\left(-\frac{\frac{t^2}{2}}{\mathcal{A}_n + \mathcal{B}_n^{\frac{1}{3}} t^{\frac{5}{3}}}\right) \quad (4.17)$$

donde $\mathcal{A}_n = \frac{A_1}{nh} + B_8 \frac{(\log n)^{-\frac{1}{2}}}{nh}$ es una cota superior de $\text{Var}(S_n)$, $\mathcal{B}_n = B \frac{M_n}{nh}$, con

$$B_8 = \frac{2B_5}{|\log a|} + \frac{18B_7}{1 - a^{\frac{1}{3}}}, \quad B = \frac{12(B_6 \vee \sqrt{B_7})}{1 \wedge |\log a|} \left(\frac{2^7 3 B_7 a^{-\frac{1}{3}}}{A_1 (1 \wedge |\log a|)} \vee 1 \right).$$

Corolario 3. Bajo las hipótesis de la Proposición 23, si $\lambda(u) = (2\mathcal{A}_n u)^{\frac{1}{2}} + \mathcal{B}_n (2u)^3$ para cada $u \geq 0$ y se toma el cambio de variable $t = \lambda(u)$ en la desigualdad de Bernstein (4.17), entonces tal desigualdad se puede reescribir de la siguiente forma

$$\mathbb{P}(|S_n| \geq \lambda(u)) \leq \exp\left(-\frac{u}{2}\right)$$

para cada $u \geq 0$.

Lemma 7. Tenemos para $i \in \{1, \dots, n\}$

$$i) \mathbb{E}[Y_i^2] \leq 2((r_{\text{sup}})^2 + \sigma^2)$$

$$ii) \mathbb{E}[Y_i^4] \leq 2^3 ((r_{\text{sup}})^4 + 3\sigma^4)$$

Lemma 8. *Tenemos para $i \in \{1, \dots, n\}$*

$$\mathbb{P}[(|Y_i| > M_n)] \leq \frac{2}{\sqrt{2\pi}} \frac{e^{-\frac{\log^2 n}{2}}}{\log n}.$$

Lemma 9. *Sea $\eta_i = Y_i K_h(x - X_i) \mathbf{1}_{\{|Y_i| > M_n\}}$. Entonces,*

$$i) \mathbb{E}[|\eta_i|] \leq \frac{2\|K\|_\infty}{(2\pi)^{\frac{1}{4}}} ((r_{\text{sup}})^2 + \sigma^2)^{\frac{1}{2}} \frac{e^{-\frac{(\log n)^2}{4}}}{h(\log n)^{\frac{1}{2}}}$$

$$ii) \mathbb{E}[\eta_i^2] \leq B_9 \frac{e^{-\frac{(\log n)^2}{4}}}{h^2(\log n)^{\frac{1}{2}}} \text{ donde } B_9 = \frac{2^2\|K\|_\infty^2}{(2\pi)^{\frac{1}{4}}} ((r_{\text{sup}})^4 + 3\sigma^4)^{\frac{1}{2}}.$$

Proposición 24. *Consideramos para $h \in \mathcal{H}^*$, el evento $\Omega_h = \{\hat{g}_h(x) \geq \frac{g_{\text{inf}}}{2}\}$. Asumimos que $g \in \Sigma(\beta_2, L_2)$ con $\beta_2 > 0$ $L_2 > 0$ y K es un núcleo de orden $m_0 \geq l_2 = [\beta_2]$ con $\int_{\mathbb{R}} |u|^{\beta_2} |K(u)| du < \infty$. Bajo (H_1) , (H_2) , (H_3) y (H_5) , se tiene*

$$\mathbb{P}(\Omega_h^c) \leq \frac{B_{11}}{n^{13}}$$

donde B_{11} es una constante positiva que depende del kernel K , de $|\log(a)|$, $\{\alpha_k(\mathbb{X})\}_{k \geq 0}$, g_{sup} y Q .

Proposición 25. *Bajo las hipótesis de la Proposición 24. Si $\hat{h}^* = \operatorname{argmin}_{h \in \mathcal{H}^*} \{A^*(h, x) + V^*(h)\}$ entonces*

$$\mathbb{P}(\Omega_{\hat{h}^*}^c) \leq \frac{B_{11}}{n^{12}}.$$

4.5.3. Demostraciones de resultados técnicos.

Demostración del Lema 4.

i) Tenemos

$$\mathbb{E} \left[(K_h(x - X_i) \varepsilon_i)^2 \right] = \frac{\sigma^2}{h^2} \mathbb{E} \left[K^2 \left(\frac{x - X_i}{h} \right) \right] = \frac{\sigma^2}{h^2} \int_{\mathbb{R}} K^2 \left(\frac{x - t}{h} \right) g(t) dt \leq \frac{B_1}{h}.$$

ii) Tenemos

$$\begin{aligned} \mathbb{E} \left[(K_h(x - X_i)r(X_i))^2 \right] &= \mathbb{E} \left[\frac{1}{h^2} K^2 \left(\frac{x - X}{h} \right) r^2(X) \right] \\ &= \frac{1}{h^2} \int_{\mathbb{R}} K^2 \left(\frac{x - t}{h} \right) r^2(t) g(t) dt \\ &\leq \frac{(r_{\text{sup}})^2 g_{\text{sup}}}{h^2} \int_{\mathbb{R}} K^2(u) h du = \frac{B_2}{h}. \end{aligned}$$

iii) Ha sido demostrado en la parte (iii) del lema 1 en el Capítulo 3, ver demostración en la Sección 3.5.3.

iv) Tenemos

$$\begin{aligned} \mathbb{E} [|K_h(x - X_i)r(X_i)|] &\leq r_{\text{sup}} \mathbb{E} [|K_h(x - X_i)|] \\ &= r_{\text{sup}} \int |K_h(x - t)| g(t) dt \\ &\leq r_{\text{sup}} g_{\text{sup}} \int \frac{1}{h} \left| K \left(\frac{x - t}{h} \right) \right| dt = r_{\text{sup}} g_{\text{sup}} \|K\|_1. \end{aligned}$$

Demostración del Lema 5.

Ha sido demostrado en el lema 2 del Capítulo 3, ver demostración en la Sección 3.5.3.

Demostración de la Proposición 21.

Ha sido demostrado en la proposición 13 del Capítulo 3, ver demostración en la Sección 3.5.3.

Demostración del Lema 6.

i) Tenemos

$$\begin{aligned} |\mathbb{E}(G(X_i, \varepsilon_i))| &\leq \mathbb{E} \left\{ \left| \frac{1}{n} (r(X_i) + \varepsilon_i) K_h(x - X_i) \mathbf{1}_{\{|r(X_i) + \varepsilon_i| \leq M_n\}} \right| \right\} \\ &\leq \frac{1}{n} \mathbb{E} \{ |r(X_i) + \varepsilon_i| |K_h(x - X_i)| \} \\ &\leq \frac{1}{n} \mathbb{E} \{ |r(X_i) K_h(x - X_i)| \} + \frac{1}{n} \mathbb{E} \{ |\varepsilon_i K_h(x - X_i)| \} \\ &\leq \frac{1}{n} r_{\text{sup}} g_{\text{sup}} \|K\|_1 + \frac{1}{n} \frac{2\sigma}{\sqrt{2\pi}} g_{\text{sup}} \|K\|_1 = \frac{B_3}{n} \end{aligned}$$

por la desigualdad iv) del Lema 4.

ii) Se denota $A(X_i) = \frac{1}{n}K_h(x - X_i)$ por lo cuál, usando la desigualdad (iii) del Lema 4, se tiene

$$\begin{aligned}
& |\mathbb{E}[G(X_i, \varepsilon_i)G(X_{i+k}, \varepsilon_{i+k})]| \\
& \leq \mathbb{E}[|Y_i A(X_i) \mathbf{1}_{\{|Y_i| \leq M_n\}} Y_{i+k} A(X_{i+k}) \mathbf{1}_{\{|Y_{i+k}| \leq M_n\}}|] \\
& \leq \mathbb{E}[|r(X_i) + \varepsilon_i| |A(X_i)| |r(X_{i+k}) + \varepsilon_{i+k}| |A(X_{i+k})|] \\
& \leq \mathbb{E}[|r(X_i)r(X_{i+k})A(X_i)A(X_{i+k})|] + \mathbb{E}[|r(X_i)\varepsilon_{i+k}A(X_i)A(X_{i+k})|] \\
& \quad + \mathbb{E}[|\varepsilon_i r(X_{i+k})A(X_i)A(X_{i+k})|] + \mathbb{E}[|\varepsilon_i \varepsilon_{i+k}A(X_i)A(X_{i+k})|] \\
& \leq (r_{\text{sup}})^2 \mathbb{E}[|A(X_i)A(X_{i+k})|] + r_{\text{sup}} \mathbb{E}[|\varepsilon_{i+k}|] \mathbb{E}[|A(X_i)A(X_{i+k})|] \\
& \quad + r_{\text{sup}} \mathbb{E}[|\varepsilon_i|] \mathbb{E}[|A(X_i)A(X_{i+k})|] + \mathbb{E}[|\varepsilon_i|] \mathbb{E}[|\varepsilon_{i+k}|] \mathbb{E}[|A(X_i)A(X_{i+k})|] \\
& = \left\{ (r_{\text{sup}})^2 + 2r_{\text{sup}} \frac{2\sigma}{\sqrt{2\pi}} + \left(\frac{2\sigma}{\sqrt{2\pi}} \right)^2 \right\} \mathbb{E}[|A(X_i)A(X_{i+k})|] \\
& \leq \left(r_{\text{sup}} + \frac{2\sigma}{\sqrt{2\pi}} \right)^2 \frac{1}{n^2} \mathbb{E}[|K_h(x - X_i)K_h(x - X_{i+k})|] \\
& \leq \left(r_{\text{sup}} + \frac{2\sigma}{\sqrt{2\pi}} \right)^2 Q \|K\|_1^2 \frac{1}{n^2} \\
& = \frac{B_4}{n^2}.
\end{aligned}$$

iii) Ahora usando (i) y (ii), se tiene

$$\begin{aligned}
& |\text{Cov}(G(X_i, \varepsilon_i), G(X_{i+k}, \varepsilon_{i+k}))| \\
& \leq |\mathbb{E}[G(X_i, \varepsilon_i)G(X_{i+k}, \varepsilon_{i+k})]| + |\mathbb{E}[G(X_i, \varepsilon_i)]| |\mathbb{E}[G(X_{i+k}, \varepsilon_{i+k})]| \\
& \leq \frac{B_4}{n^2} + \frac{B_3}{n} \frac{B_3}{n} = \frac{B_5}{n^2}.
\end{aligned}$$

La desigualdad (iv) es inmediata usando hipótesis (H_5).

Demostración de la Proposición 22.

Usando el lema 6 y siguiendo la demostración de la Proposición 14 en sección 3.5.3 del capítulo 3, se obtiene el resultado de la Proposición 22.

Demostración Proposición 23.

Por la Proposición 22 y siguiendo la demostración de la Proposición 15 en Sección 3.5.3 del Capítulo 3, se obtiene el resultado de esta proposición.

Demostración Corolario 3.

Por la proposición 23 y siguiendo la demostración del corolario 1 en sección 3.5.3 del capítulo 3, se obtiene el resultado.

Demostración del Lema 7.

Las afirmaciones *i)* y *ii)* de este Lema, son consecuencia directa del hecho que para $k \in \mathbb{N}$

$$\mathbb{E} \left[Y_i^{2^k} \right] \leq \left(\prod_{j=0}^{k-1} 2^{2^j} \right) \mathbb{E} \left[(r(X_i))^{2^k} + \varepsilon_i^{2^k} \right],$$

$$\mathbb{E}[\varepsilon_i^{2^{k-1}}] = 0, \quad \text{y} \quad \mathbb{E}[\varepsilon_i^{2^k}] = \left[\prod_{j=1}^k (2j-1) \right] \sigma^{2^k}.$$

Demostración del Lema 8.

Tenemos

$$\begin{aligned} \mathbb{P}(\{|Y| > M_n\}) &= \mathbb{P}(\{|r(X) + \varepsilon| > M_n\}) \\ &\leq \mathbb{P}(\{r_{\text{sup}} + |\varepsilon| > M_n\}) \\ &\leq \mathbb{P}(\{|\varepsilon| > M_n - r_{\text{sup}}\}) \\ &\leq 2\mathbb{P}\left(\left\{\frac{\varepsilon}{\sigma} > \frac{M_n - r_{\text{sup}}}{\sigma}\right\}\right) \\ &\leq 2 \frac{\phi(\log n)}{\log n} \end{aligned}$$

substituyendo M_n por su valor y donde $\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$. Por lo anterior se tiene que,

$$\mathbb{P}(\{|Y| > M_n\}) \leq \frac{2}{\sqrt{2\pi}} \frac{e^{-\frac{(\log n)^2}{2}}}{\log n}.$$

Demostración del Lema 9.

i)

$$\begin{aligned}
\mathbb{E}[|\eta_i|] &= \mathbb{E} \left[\left| Y_i \frac{1}{h} K \left(\frac{x - X_i}{h} \right) \mathbf{1}_{\{|Y_i| > M_n\}} \right| \right] \\
&\leq \frac{\|K\|_\infty}{h} \mathbb{E} [|Y_i| \mathbf{1}_{\{|Y_i| > M_n\}}] \\
&\leq \frac{\|K\|_\infty}{h} (\mathbb{E}[|Y_i|^2])^{\frac{1}{2}} (\mathbb{P}[\{|Y_i| > M_n\}])^{\frac{1}{2}} \\
&\leq \frac{\|K\|_\infty}{h} (2((r_{\text{sup}})^2 + \sigma^2))^{\frac{1}{2}} \left(\frac{2}{\sqrt{2\pi}} \frac{e^{-\frac{(\log n)^2}{2}}}{\log n} \right)^{\frac{1}{2}} \\
&= \frac{2\|K\|_\infty}{(2\pi)^{\frac{1}{4}}} ((r_{\text{sup}})^2 + \sigma^2)^{\frac{1}{2}} \frac{e^{-\frac{(\log n)^2}{4}}}{h(\log n)^{\frac{1}{2}}}
\end{aligned}$$

por la parte (i) del Lema 7 y el Lema 8.

ii) Usando de nuevo Cauchy-schwartz, la parte (ii) del Lema 7 y el Lema 8, obtenemos:

$$\begin{aligned}
\mathbb{E}[\eta_i^2] &\leq \frac{\|K\|_\infty^2}{h^2} (\mathbb{E}[(Y_i)^4])^{\frac{1}{2}} (\mathbb{P}[\{|Y_i| > M_n\}])^{\frac{1}{2}} \\
&\leq \frac{2^2\|K\|_\infty^2}{(2\pi)^{\frac{1}{4}}} ((r_{\text{sup}})^4 + 3\sigma^4)^{\frac{1}{2}} \frac{e^{-\frac{(\log n)^2}{4}}}{h^2(\log n)^{\frac{1}{2}}}
\end{aligned}$$

Demostración de la Proposición 24.

Sea $h \in \mathcal{H}^*$. Tenemos

$$\begin{aligned}
\mathbb{P}[\Omega_h^c] &= \mathbb{P} \left[\hat{g}_h(x) < \frac{g_{\text{inf}}}{2} \right] \\
&= \mathbb{P} \left[\hat{g}_h(x) - \mathbb{E}[\hat{g}_h(x)] + \mathbb{E}[\hat{g}_h(x)] - g(x) + g(x) < \frac{g_{\text{inf}}}{2} \right] \\
&= \mathbb{P} \left[\frac{1}{nh} \sum_{i=1}^n \left\{ K \left(\frac{x - X_i}{h} \right) - \mathbb{E} \left[K \left(\frac{x - X_i}{h} \right) \right] \right\} + b(x) + g(x) < \frac{g_{\text{inf}}}{2} \right] \\
&= \mathbb{P} \left[S_n < \frac{g_{\text{inf}}}{2} - g(x) - b(x) \right] \tag{4.18}
\end{aligned}$$

donde $S_n = \frac{1}{nh} \sum_{i=1}^n \left\{ K \left(\frac{x - X_i}{h} \right) - \mathbb{E} \left[K \left(\frac{x - X_i}{h} \right) \right] \right\}$ y $b(x) = \mathbb{E}[\hat{g}_h(x)] - g(x)$.

Usando la Proposición 18, se tiene

$$|b(x)| \leq h^{\beta_2} \mathfrak{A}_0$$

por lo cuál

$$-h^{\beta_2}\mathfrak{A}_0 < -b(x) < h^{\beta_2}\mathfrak{A}_0$$

Por la desigualdad anterior y como $\frac{g_{\text{inf}}}{2} - g(x) < -\frac{g_{\text{inf}}}{2}$, se tiene que

$$\frac{g_{\text{inf}}}{2} - g(x) - b(x) < -\frac{g_{\text{inf}}}{2} + h^{\beta_2}\mathfrak{A}_0 < -\frac{g_{\text{inf}}}{2} + (h_{\text{máx}}^*)^{\beta_2}\mathfrak{A}_0$$

Entonces, para n suficientemente grande se tiene que

$$\frac{g_{\text{inf}}}{2} - g(x) - b(x) < -\frac{g_{\text{inf}}}{4}$$

Por la desigualdad (4.18) se tiene entonces que

$$\mathbb{P}[\Omega_h^c] \leq \mathbb{P}\left[S_n < -\frac{g_{\text{inf}}}{4}\right] \leq \mathbb{P}\left[|S_n| > \frac{g_{\text{inf}}}{4}\right] \quad (4.19)$$

En el Lema 3 de Bertin and Klutchnikoff (2017), se establece que

$$\mathbb{P}[|S_n| \geq \lambda^*(t)] \leq e^{-\frac{t}{2}} \quad (4.20)$$

para $t \geq 0$, donde $\lambda^*(t) = \sigma_n^*(h)\sqrt{2t} + B_n^*(h)(2t)^3$, $\sigma_n^*(h) = \sqrt{J_n + C_6^* \frac{(\delta_n^*)^2}{nh}}$, $\delta_n^* = (\log n)^{-\frac{1}{2}}$, $B_n^*(h) = \frac{C_7^*}{nh}$, $J_n(h) = \frac{1}{n} \int K_h^2(x_0 - t)g(t)dt$, C_6^* es una constante positiva que depende del núcleo K y de $|\log(a)|$, $\{\alpha_k(\mathbf{X})\}_{k \geq 0}$, g_{sup} y \mathcal{Q} , y C_7^* es una constante positiva que depende del núcleo K y de $|\log(a)|$, g_{sup} y \mathcal{Q} .

Por la definición de $J_n(h)$, se tiene

$$J_n(h) \leq \frac{g_{\text{sup}}}{nh^2} \int K^2\left(\frac{x_0 - t}{h}\right) dt \leq \frac{g_{\text{sup}}\|K\|_2^2}{nh}$$

Por la cota de $J_n(h)$, las definiciones de $B_n^*(h)$, δ_n^* , $\sigma_n^*(h)$ y $\lambda^*(t)$. Al tomar $t_0 = 26 \log n$, se tiene

$$\begin{aligned} \lambda^*(t_0) &\leq \sqrt{\frac{g_{\text{sup}}\|K\|_2^2}{nh} + C_6^* \frac{(\log n)^{-1}}{nh}} \sqrt{2t_0} + \frac{C_7^*}{nh} (2t_0)^3 \\ &\leq \sqrt{\frac{g_{\text{sup}}\|K\|_2^2}{nh_{\text{mín}}^*} + C_6^* \frac{(\log n)^{-1}}{nh_{\text{mín}}^*}} \sqrt{2(26 \log n)} + \frac{C_7^*}{nh_{\text{mín}}^*} (2(26 \log n))^3 \\ &= \sqrt{\frac{g_{\text{sup}}\|K\|_2^2}{e^{(\sqrt{\log n}-1)}} + C_6^* \frac{(\log n)^{-1}}{e^{(\sqrt{\log n}-1)}}} \sqrt{52 \log n} + \frac{C_7^*}{e^{(\sqrt{\log n}-1)}} (52 \log n)^3 \end{aligned}$$

Es claro que $\lambda^*(t_0) \searrow 0$ cuando $n \rightarrow \infty$ con $t_0 = 26 \log n$, por lo cual, para n suficientemente grande se tiene que $\lambda^*(t_0) < \frac{g_{\text{inf}}}{4}$. Esto último permite establecer la siguiente relación de inclusión de eventos

$$\left[|S_n| > \frac{g_{\text{inf}}}{4} \right] \subseteq [|S_n| > \lambda(t_0)]$$

Por lo anterior y la desigualdad (4.20) se tiene

$$\mathbb{P} \left[|S_n| > \frac{g_{\text{inf}}}{4} \right] \leq e^{-\frac{26 \log n}{2}} = n^{-13}$$

para n suficientemente grande. Por esto último y la desigualdad (4.19) se concluye que

$$\mathbb{P}[\Omega_h^c] \leq B_{11} n^{-13},$$

donde B_{11} depende del kernel K , de $|\log(a)|$, $\{\alpha_k(\mathbb{X})\}_{k \geq 0}$, g_{sup} y Q .

Demostración de la Proposición 25.

Tenemos

$$\mathbb{P} \left[\Omega_{\hat{h}^*}^c \right] \leq \sum_{h \in \mathcal{H}^*} \mathbb{P} \left[\hat{g}_h(x) < \frac{g_{\text{inf}}}{2} \right] \leq B_{11} n^{-12}$$

por la Proposición 24 y usando que $|\mathcal{H}^*| < n$.

4.5.4. Demostración de la Proposición 16: Sesgo del estimador \hat{m}_h .

$$\begin{aligned} \mathbb{E}\{\hat{m}_h(x)\} &= \mathbb{E} \left\{ \frac{1}{n} \sum_{i=1}^n K_h(x - X_i) Y_i \right\} \\ &= \mathbb{E} \left\{ \frac{1}{n} \sum_{i=1}^n K_h(x - X_i) r(X_i) \right\} + \mathbb{E} \left\{ \frac{1}{n} \sum_{i=1}^n K_h(x - X_i) \varepsilon_i \right\} \\ &= \mathbb{E} \{ K_h(x - X) r(X) \} + \mathbb{E} \{ K_h(x - X) \} \mathbb{E} \{ \varepsilon \} \\ &= \int_{\mathbb{R}} K_h(x - t) r(t) g(t) dt \\ &= \int_{\mathbb{R}} K_h(x - t) m(t) dt = K_h * m(x). \end{aligned}$$

Así se tiene que,

$$\mathbb{E}[\hat{m}_h(x)] - m(x) = K_h * m(x) - m(x) = \int_{\mathbb{R}} (m(x - hu) - m(x)) K(u) du.$$

De la fórmula de Taylor-Lagrange al orden l_1 aplicado a $m(x)$ se tiene,

$$m(x - hu) = m(x) - hum'(x) + \frac{(-hu)^2}{2!}m''(x) - \dots + \frac{(-hu)^{l_1}}{l_1!}m^{(l_1)}(x - \zeta_u hu)$$

donde $\zeta_u \in [0, 1]$.

Por lo anterior y como K es un núcleo de orden m_0 con $m_0 \geq l_1$ se tiene que,

$$\mathbb{E}[\hat{m}_h(x)] - m(x) = \int_{\mathbb{R}} \frac{(-hu)^{l_1}}{l_1!}m^{(l_1)}(x - \zeta_u hu)K(u)du$$

Al restar el término nulo $\int_{\mathbb{R}} \frac{(-uh)^{l_1}}{l_1!}m^{(l_1)}(x)K(u)du$ a la ecuación anterior, se tiene que,

$$\mathbb{E}[\hat{m}_h(x)] - m(x) = (-1)^{l_1} \frac{h^{l_1}}{l_1!} \int_{\mathbb{R}} \left(m^{(l_1)}(x - \zeta_u hu) - m^{(l_1)}(x) \right) u^{l_1} K(u) du$$

Al tomar valor absoluto en ambos lados y usar la hipótesis $m \in \Sigma(\beta_1, L_1)$ se tiene,

$$\begin{aligned} |\mathbb{E}[\hat{m}_h(x)] - m(x)| &\leq \frac{h^{\beta_1}}{l_1!} L_1 \int_{\mathbb{R}} |u|^{\beta_1} \zeta_u^{\beta_1 - l_1} |K(u)| du \\ &\leq \frac{h^{\beta_1}}{l_1!} L_1 \int_{\mathbb{R}} |u|^{\beta_1} |K(u)| du, \quad \text{pues } \zeta_u^{\beta_1 - l_1} < 1 \\ &\leq h^{\beta_1} A_0. \end{aligned}$$

4.5.5. Demostración de la Proposición 17: Varianza del estimador \hat{m}_h

Al sustituir Y_i por $r(X_i) + \varepsilon_i$, se obtiene,

$$\begin{aligned} \mathbb{E} [\{\hat{m}_h(x) - \mathbb{E}[\hat{m}_h(x)]\}^2] &= \mathbb{E} \left[\left\{ \frac{1}{n} \sum_{i=1}^n K_h(x - X_i) Y_i - \mathbb{E} \left(\frac{1}{n} \sum_{i=1}^n K_h(x - X_i) Y_i \right) \right\}^2 \right] \\ &= \mathbb{E} \left\{ \frac{1}{n} \sum_{i=1}^n [K_h(x - X_i) r(X_i) - \mathbb{E}(K_h(x - X_i) r(X_i))] \right. \\ &\quad \left. + \frac{1}{n} \sum_{i=1}^n K_h(x - X_i) \varepsilon_i \right\}^2 \\ &= I_1 + I_2 + I_3 \end{aligned} \tag{4.21}$$

donde,

- $A(X_1, \dots, X_n) = \sum_{i=1}^n [K_h(x - X_i)r(X_i) - \mathbb{E}(K_h(x - X_i)r(X_i))]$
- $I_1 = \mathbb{E} \left\{ \frac{1}{n} \sum_{i=1}^n [K_h(x - X_i)r(X_i) - \mathbb{E}(K_h(x - X_i)r(X_i))] \right\}^2$
- $I_2 = \mathbb{E} \left\{ 2 \frac{1}{n} A(X_1, \dots, X_n) \frac{1}{n} \sum_{i=1}^n K_h(x - X_i) \varepsilon_i \right\}$
- $I_3 = \mathbb{E} \left\{ \frac{1}{n} \sum_{i=1}^n K_h(x - X_i) \varepsilon_i \right\}^2$

Se procede a acotar I_1 , I_2 e I_3 .

$$\begin{aligned}
 I_3 &= \frac{1}{n^2} \mathbb{E} \left\{ \sum_{i=1}^n K_h(x - X_i) \varepsilon_i \right\}^2 \\
 &= \frac{1}{n^2} \sum_{i=1}^n \mathbb{E} \{ K_h(x - X_i) \varepsilon_i \}^2 + \frac{1}{n^2} \sum_{i \neq j}^n \mathbb{E} \{ K_h(x - X_i) \varepsilon_i K_h(x - X_j) \varepsilon_j \} \\
 &\leq \frac{1}{n^2} \sum_{i=1}^n \frac{B_1}{h} + \frac{1}{n^2} \sum_{i \neq j}^n \mathbb{E} \{ K_h(x - X_i) K_h(x - X_j) \} \mathbb{E} \{ \varepsilon_i \varepsilon_j \} \\
 &= \frac{B_1}{nh}
 \end{aligned} \tag{4.22}$$

por la primera desigualdad del Lema 4.

$$\begin{aligned}
 I_2 &= \frac{2}{n^2} \mathbb{E} \left\{ A(X_1, \dots, X_n) \sum_{i=1}^n K_h(x - X_i) \varepsilon_i \right\} \\
 &= \frac{2}{n^2} \sum_{i=1}^n \mathbb{E} \{ A(X_1, \dots, X_n) K_h(x - X_i) \varepsilon_i \} \\
 &= \frac{2}{n^2} \sum_{i=1}^n \mathbb{E} \{ A(X_1, \dots, X_n) K_h(x - X_i) \} \mathbb{E} \{ \varepsilon_i \} = 0
 \end{aligned}$$

$$\begin{aligned}
 I_1 &= \frac{1}{n^2} \mathbb{E} \left\{ \sum_{i=1}^n [K_h(x - X_i)r(X_i) - \mathbb{E}(K_h(x - X_i)r(X_i))] \right\}^2 \\
 &\leq I_4 + I_5
 \end{aligned} \tag{4.23}$$

donde,

- $I_4 = \left| \frac{1}{n^2} \sum_{i=1}^n \text{Var} [K_h(x - X_i)r(X_i)] \right|$

$$\blacksquare I_5 = \left| \frac{1}{n^2} \sum_{i \neq j} \text{Cov} [K_h(x - X_i)r(X_i), K_h(x - X_j)r(X_j)] \right|$$

Se procede a acotar I_4 e I_5 . Por la segunda desigualdad del Lema 4, se tiene

$$I_4 \leq \frac{1}{n^2} \sum_{i=1}^n \mathbb{E} \left[(K_h(x - X_i)r(X_i))^2 \right] \leq \frac{B_2}{nh}.$$

$$\begin{aligned} I_5 &\leq \frac{1}{n^2} \sum_{i \neq j}^n |\text{Cov} [K_h(x - X_i)r(X_i), K_h(x - X_j)r(X_j)]| \\ &= \frac{2}{n^2} \sum_{i=1}^{n-1} \sum_{j=i+1}^n |\text{Cov} [K_h(x - X_i)r(X_i), K_h(x - X_j)r(X_j)]| \\ &= \frac{2}{n^2} \sum_{i=1}^{n-1} \sum_{k=1}^{n-i} |\text{Cov}(\tilde{X}_i, \tilde{X}_{i+k})| \\ &= \frac{2}{n^2} \sum_{i=1}^{n-1} \left(\sum_{k=1}^{u_n} |\text{Cov}(\tilde{X}_i, \tilde{X}_{i+k})| + \sum_{k=u_n+1}^{n-i} |\text{Cov}(\tilde{X}_i, \tilde{X}_{i+k})| \right) \end{aligned} \quad (4.24)$$

donde $1 \leq u_n \leq n - 1$ y $\tilde{X}_i = K_h(x - X_i)r(X_i)$, $\forall i$.

Se analiza el término $\sum_{k=1}^{u_n} |\text{Cov}(\tilde{X}_i, \tilde{X}_{i+k})|$ de la igualdad (4.24).

$$\begin{aligned} \sum_{k=1}^{u_n} |\text{Cov}(\tilde{X}_i, \tilde{X}_{i+k})| &\leq \sum_{k=1}^{u_n} |\mathbb{E}[\tilde{X}_i \tilde{X}_{i+k}]| + |\mathbb{E}[\tilde{X}_i] \mathbb{E}[\tilde{X}_{i+k}]| \\ &\leq \sum_{k=1}^{u_n} \mathbb{E}[|\tilde{X}_i \tilde{X}_{i+k}|] + \mathbb{E}[|\tilde{X}_i|] \mathbb{E}[|\tilde{X}_{i+k}|] \\ &= \sum_{k=1}^{u_n} \left\{ \mathbb{E}[|K_h(x - X_i)r(X_i)K_h(x - X_{i+k})r(X_{i+k})|] + (\mathbb{E}[|\tilde{X}_i|])^2 \right\} \\ &\leq \sum_{k=1}^{u_n} \left\{ (r_{\text{sup}})^2 \mathbb{E}[|K_h(x - X_i)K_h(x - X_{i+k})|] + (r_{\text{sup}}g_{\text{sup}}\|K\|_1)^2 \right\} \\ &= \sum_{k=1}^{u_n} \left\{ (r_{\text{sup}})^2 Q \|K\|_1^2 + (r_{\text{sup}})^2 (g_{\text{sup}})^2 \|K\|_1^2 \right\} = A_3 u_n \end{aligned}$$

por la tercera y cuarta desigualdad del Lema 4.

Usando (H_3) y que $|\tilde{X}_i| \leq \frac{r_{\text{sup}}\|K\|_{\infty}}{h}$, se obtiene

$$|\text{Cov}(\tilde{X}_i, \tilde{X}_{i+k})| \leq 4a^k \frac{(r_{\text{sup}})^2 \|K\|_{\infty}^2}{h^2}$$

Por esta última desigualdad se tiene que,

$$\begin{aligned}
\sum_{k=u_n+1}^{n-i} |Cov(\tilde{X}_i, \tilde{X}_{i+k})| &\leq \frac{4(r_{\text{sup}})^2 \|K\|_{\infty}^2}{h^2} \sum_{k=u_n+1}^{\infty} a^k \\
&= \frac{4(r_{\text{sup}})^2 \|K\|_{\infty}^2 a^{u_n+1}}{h^2 (1-a)} \\
&= A_4 \frac{a^{u_n+1}}{h^2}.
\end{aligned} \tag{4.25}$$

Por la ecuación (4.24) y las desigualdades (4.25) y (4.25) se tiene,

$$\begin{aligned}
I_5 &\leq \frac{2}{n^2} \sum_{i=1}^{n-1} \left(A_3 u_n + A_4 \frac{a^{u_n+1}}{h^2} \right) \\
&\leq \frac{1}{n} \left(2A_3 u_n + 2A_4 \frac{a^{u_n+1}}{h^2} \right)
\end{aligned}$$

La variable u_n en las sumatorias de la ecuación (4.24) se toma como $u_n = \left\lceil \frac{1}{h|\log a|(\log n)^{\frac{1}{2}}} \right\rceil$ y se usan los hechos de que $[x] \leq x \leq [x] + 1$ y $0 < a < 1$, para seguir acotando la desigualdad anterior,

$$\begin{aligned}
I_5 &\leq \frac{1}{n} \left(2A_3 \frac{1}{h|\log a|(\log n)^{\frac{1}{2}}} + 2A_4 \frac{a^{\frac{1}{h|\log a|(\log n)^{\frac{1}{2}}}}}{h^2} \right) \\
&= \frac{1}{nh} \left(2A_3 \frac{1}{|\log a|(\log n)^{\frac{1}{2}}} + 2A_4 \frac{1}{h} a^{\frac{1}{h|\log a|(\log n)^{\frac{1}{2}}}} \right)
\end{aligned}$$

Por el Lema 5 y la desigualdad anterior se tiene que, para cada $n \geq 4$,

$$\begin{aligned}
I_5 &\leq \frac{1}{nh} \left(\frac{2A_3}{|\log a|(\log n)^{\frac{1}{2}}} + 2A_4 \frac{10}{(\log n)^{\frac{5}{2}}} \right) \\
&\leq \left(\frac{2A_3}{|\log a|} + 20A_4 \right) \frac{(\log n)^{-\frac{1}{2}}}{nh}.
\end{aligned} \tag{4.26}$$

Por las igualdades (4.23), (4.24) y (4.26), para todo $n \geq 4$,

$$I_1 \leq \frac{B_2}{nh} + \left(\frac{2A_3}{|\log a|} + 20A_4 \right) \frac{(\log n)^{-\frac{1}{2}}}{nh}. \tag{4.27}$$

Finalmente, por las desigualdades (4.22) y (4.27) se tiene que,

$$\mathbb{E} [(\hat{r}_h(x) - \mathbb{E}[\hat{r}_h(x)])^2] = I_1 + I_3 \leq \frac{A_1}{nh} + A_2 \frac{(\log n)^{-\frac{1}{2}}}{nh}.$$

4.5.6. Demostración del Teorema 8

Siguiendo el mismo esquema que la demostración del Teorema 4 en el Capítulo 3, el estimador satisface primero

$$\begin{aligned} & \left(\mathbb{E} (m(x) - \hat{m}_h(x))^2 \right)^{\frac{1}{2}} \\ & \leq \left(\mathbb{E} (m(x) - \hat{m}_h(x))^2 \right)^{\frac{1}{2}} + 2 \left(\mathbb{E}(T_1^2) \right)^{\frac{1}{2}} + 2 \left(\mathbb{E}(T_2^2) \right)^{\frac{1}{2}} + 2C(h)\|K\|_1 + 2V(h) \end{aligned} \quad (4.28)$$

donde

$$\begin{aligned} T_1 &= \max_{h' \in \mathcal{H}} \{ |\hat{m}_{h'}(x) - \mathbb{E}(\hat{m}_{h'}(x))| - V_1(h') \}_+ \\ T_2 &= \max_{h' \in \mathcal{H}} \{ |\hat{m}_{h,h'}(x) - \mathbb{E}(\hat{m}_{h,h'}(x))| - V_2(h') \}_+ \\ C(h) &= \max_{u \in B(x)} |K_h * m(u) - m(u)| \end{aligned}$$

con

$$V_1(h) = \sqrt{2\gamma A_1} (1 + \delta_n) \frac{(\log n)^{\frac{1}{2}}}{(nh)^{\frac{1}{2}}} \quad \text{y} \quad V_2(h) = \sqrt{2\gamma A_1} \|K\|_1 (1 + \delta_n) \frac{(\log n)^{\frac{1}{2}}}{(nh)^{\frac{1}{2}}}.$$

La desigualdad (4.28) se obtiene siguiendo línea por línea la demostración de la Proposición 11 en Sección 3.5.6 del Capítulo 3 substituyendo \hat{r}_h por \hat{m}_h y usando que $\mathbb{E}(\hat{m}_h(x)) = K_h * m(x)$.

Segundo, T_1 y T_2 satisfacen para n suficientemente grande

$$\mathbb{E}(T_1^2) \leq A_5 \frac{\log n}{n} \quad (4.29)$$

$$\mathbb{E}(T_2^2) \leq A_5 \|K\|_1^2 \frac{\log n}{n} \quad (4.30)$$

donde

$$A_5 = B_9 + 2B_{10}$$

con

$$B_9 = \|K\|_\infty^2 \left\{ (8((r_{\text{sup}})^4 + 3\sigma^4)) \right\}^{\frac{1}{2}} \left(\frac{2}{\sqrt{2\pi}} \right)^{\frac{1}{2}}$$

$$B_{10} = \frac{2^4 3}{1 - e^{-\frac{\gamma-2}{2}}} (A_1 + B_8) + \frac{5! 2^{18} 3 B^2}{1 - e^{-\frac{\gamma-2}{2}}} (\sigma + r_{\text{sup}})^2$$

De nuevo, (4.29) y (4.30) se obtienen siguiendo línea por línea las demostraciones ubicadas en secciones 3.5.7 y 3.5.8 en el Capítulo 3 y usando Proposición 23, Corolario 3 y la desigualdad (ii) del Lema 9. Ahora usando que para $h \in \mathcal{H}$, $|\mathbb{E}(\hat{m}_h(x)) - m(x)| = |K_h * m(x) - m(x)| \leq C(h)$ y la Proposición 17 se tiene que,

$$\begin{aligned} \left(\mathbb{E} \left[(r(x) - \hat{r}_h(x))^2 \right] \right)^{\frac{1}{2}} &\leq C(h) + \left(\frac{A_1}{nh} + A_2 \frac{(\log n)^{-\frac{1}{2}}}{nh} \right)^{\frac{1}{2}} \\ &\leq C(h) + (A_1 + A_2)^{1/2} (nh)^{-1/2} \end{aligned} \quad (4.31)$$

Ahora por las desigualdades (4.28) y (4.31), se tiene que,

$$\begin{aligned} &\left(\mathbb{E} (r(x) - \hat{r}_h(x))^2 \right)^{\frac{1}{2}} \\ &\leq C(h) + (A_1 + A_2)^{1/2} (nh)^{-1/2} + 2C(h) \|K\|_1 + 2V(h) + 2 \left(\mathbb{E}(T_1^2) \right)^{\frac{1}{2}} + 2 \left(\mathbb{E}(T_2^2) \right)^{\frac{1}{2}} \\ &\leq A_6 C(h) + A_7 (1 + \delta_n) \frac{(\log n)^{\frac{1}{2}}}{(nh)^{\frac{1}{2}}} + 2 \left(\mathbb{E}(T_1^2) \right)^{\frac{1}{2}} + 2 \left(\mathbb{E}(T_2^2) \right)^{\frac{1}{2}} \end{aligned}$$

Usando desigualdades (4.29) y (4.30) permite obtener que

$$2 \left(\mathbb{E}(T_1^2) \right)^{\frac{1}{2}} + 2 \left(\mathbb{E}(T_2^2) \right)^{\frac{1}{2}} \leq A_8 \frac{(\log n)^{\frac{1}{2}}}{n^{\frac{1}{2}}}$$

para una constante $A_8 = 2\sqrt{A_5}(\|K\|_1 + 1)$ que depende del kernel K , de r_{sup} , Q , g_{sup} , σ , a y γ , lo que permite concluir.

4.5.7. Demostración de la Proposición 20.

Sea n tal que $\frac{1}{\log n} < \frac{g_{\text{inf}}}{2}$. Tenemos para $h \in \mathcal{H}$ y $h^* \in \mathcal{H}^*$

$$(r(x) - \hat{r}_{h/h^*}(x))^2 \leq \frac{2}{(\bar{g}_{h^*}(x))^2} (m(x) - \hat{m}_h(x))^2 + 2 \left(\frac{r(x)}{\bar{g}_{h^*}(x)} \right)^2 (\bar{g}_{h^*}(x) - g(x))^2$$

por lo cuál

$$\begin{aligned} \mathbb{E} \left[(r(x) - \hat{r}_{h/h^*}(x))^2 \right] &\leq 2\mathbb{E} \left[\frac{1}{(\bar{g}_{h^*}(x))^2} (m(x) - \hat{m}_h(x))^2 \right] \\ &\quad + 2(r_{\text{sup}})^2 \mathbb{E} \left[\frac{1}{(\bar{g}_{h^*}(x))^2} (\bar{g}_{h^*}(x) - g(x))^2 \right] \end{aligned} \quad (4.32)$$

Se recuerda que $\bar{g}_{h^*}(x) = \max \left\{ \hat{g}_{h^*}(x), \frac{1}{\log n} \right\}$. Consideremos el evento $\Omega_{h^*} = \left\{ \hat{g}_{h^*}(x) \geq \frac{g_{\text{inf}}}{2} \right\}$. Sobre Ω_{h^*} , tenemos $\bar{g}_{h^*}(x) = \hat{g}_{h^*}(x)$ y

$$\frac{1}{\bar{g}_{h^*}(x)} \leq \frac{2}{g_{\text{inf}}}. \quad (4.33)$$

De la desigualdad (4.32) se analiza el primer término

$$\begin{aligned} \mathbb{E} \left[\frac{1}{(\bar{g}_{h^*}(x))^2} (m(x) - \hat{m}_h(x))^2 \right] &= \mathbb{E} \left[\frac{1}{(\bar{g}_{h^*}(x))^2} (m(x) - \hat{m}_h(x))^2 \mathbf{1}_{\Omega_{h^*}} \right] \\ &\quad + \mathbb{E} \left[\frac{1}{(\bar{g}_{h^*}(x))^2} (m(x) - \hat{m}_h(x))^2 \mathbf{1}_{\Omega_{h^*}^c} \right] \end{aligned} \quad (4.34)$$

Por la desigualdad (4.33) y el primer término de la ecuación (4.34), se tiene que

$$\begin{aligned} \mathbb{E} \left[\frac{1}{(\bar{g}_{h^*}(x))^2} (m(x) - \hat{m}_h(x))^2 \mathbf{1}_{\Omega_{h^*}} \right] &\leq \frac{4}{(g_{\text{inf}})^2} \mathbb{E} \left[(m(x) - \hat{m}_h(x))^2 \mathbf{1}_{\Omega_{h^*}} \right] \\ &\leq \frac{4}{(g_{\text{inf}})^2} \mathbb{E} \left[(m(x) - \hat{m}_h(x))^2 \right] \end{aligned} \quad (4.35)$$

Por definición de \bar{g}_{h^*} y el segundo término de la ecuación (4.34), se tiene que

$$\begin{aligned}
& \mathbb{E} \left[\frac{1}{(\bar{g}_{h^*}(x))^2} (m(x) - \hat{m}_h(x))^2 \mathbf{1}_{\Omega_{h^*}^c} \right] \\
& \leq (\log n)^2 \mathbb{E} \left[(m(x) - \hat{m}_h(x))^2 \mathbf{1}_{\Omega_{h^*}^c} \right] \\
& \leq (\log n)^2 \mathbb{E} \left[(2(m(x))^2 + 2(\hat{m}_h(x))^2) \mathbf{1}_{\Omega_{h^*}^c} \right] \\
& = 2(\log n)^2 \left\{ \mathbb{E} \left[(m(x))^2 \mathbf{1}_{\Omega_{h^*}^c} \right] + \mathbb{E} \left[(\hat{m}_h(x))^2 \mathbf{1}_{\Omega_{h^*}^c} \right] \right\} \\
& = 2(\log n)^2 \left\{ (r_{\text{sup}} g_{\text{sup}})^2 \mathbb{P}[\Omega_{h^*}^c] + I_1 \right\} \\
& \leq 2(r_{\text{sup}} g_{\text{sup}})^2 B_{11} n^{-4} + 2(\log n)^2 I_1
\end{aligned} \tag{4.36}$$

donde la ultima linea es consecuencia de la Proposición 24, y de que $I_1 = \mathbb{E} \left[(\hat{m}_h(x))^2 \mathbf{1}_{\Omega_{h^*}^c} \right]$.

Ahora

$$\begin{aligned}
I_1 & = \mathbb{E} \left[\left(\frac{1}{n} \sum_{i=1}^n Y_i \frac{1}{h} K \left(\frac{x - X_i}{h} \right) \right)^2 \mathbf{1}_{\Omega_{h^*}^c} \right] \\
& \leq \frac{\|K\|_\infty^2}{n^2 (h_{\text{mín}})^2} \mathbb{E} \left[\left(\sum_{i=1}^n Y_i \right)^2 \mathbf{1}_{\Omega_{h^*}^c} \right] \\
& \leq \frac{\|K\|_\infty^2}{n^2 (h_{\text{mín}})^2} \left(\mathbb{E} \left[\left(\sum_{i=1}^n Y_i \right)^4 \right] \right)^{1/2} (\mathbb{P}[\Omega_{h^*}^c])^{1/2} \\
& \leq \frac{\|K\|_\infty^2}{n^2 (h_{\text{mín}})^2} \left(\mathbb{E} \left[n^3 \sum_{i=1}^n Y_i^4 \right] \right)^{1/2} (\mathbb{P}[\Omega_{h^*}^c])^{1/2} \\
& \leq \frac{\|K\|_\infty^2 n^2 (2^3 (r_{\text{sup}}^4 + 3\sigma^4))^{1/2}}{n^2 (h_{\text{mín}})^2} (\mathbb{P}[\Omega_{h^*}^c])^{1/2} \\
& \leq \frac{\|K\|_\infty^2 (2^3 (r_{\text{sup}}^4 + 3\sigma^4))^{1/2}}{(\log n)^2} B_{11}^{1/2} n^{-4},
\end{aligned}$$

donde la ultima linea es consecuencia de la Proposición 24, y de que $h_{\text{mín}} = \frac{(\log n)^8}{n} \geq \frac{\log n}{n}$.

Entonces

$$\mathbb{E} \left[\frac{1}{(\bar{g}_{h^*}(x))^2} (m(x) - \hat{m}_h(x))^2 \mathbf{1}_{\Omega_{h^*}^c} \right] \leq \bar{F}_1 n^{-4}, \tag{4.37}$$

donde $\bar{F}_1 = 2B_{11}(r_{\text{sup}} g_{\text{sup}})^2 + 2\|K\|_\infty^2 B_{11}^{1/2} (2^3 (r_{\text{sup}}^4 + 3\sigma^4))^{1/2}$.

Se da una cota del primer término del lado derecho de la desigualdad (4.32), usando las desigualdades (4.34), (4.35) y (4.37).

$$\mathbb{E} \left[\frac{1}{(\bar{g}_{h^*}(x))^2} (m(x) - \hat{m}_h(x))^2 \right] \leq \frac{4}{(g_{inf})^2} \mathbb{E} [(m(x) - \hat{m}_h(x))^2] + \frac{\bar{F}_1}{n^4}. \quad (4.38)$$

Ahora se da una cota del segundo término del lado derecho de la desigualdad (4.32).

$$\begin{aligned} \mathbb{E} \left[\frac{1}{(\bar{g}_{h^*}(x))^2} (\bar{g}_{h^*}(x) - g(x))^2 \right] &= \mathbb{E} \left[\frac{1}{(\bar{g}_{h^*}(x))^2} (\bar{g}_{h^*}(x) - g(x))^2 \mathbf{1}_{\Omega_{h^*}} \right] \\ &+ \mathbb{E} \left[\frac{1}{(\bar{g}_{h^*}(x))^2} (\bar{g}_{h^*}(x) - g(x))^2 \mathbf{1}_{\Omega_{h^*}^c} \right] \end{aligned} \quad (4.39)$$

Por el primer término del lado derecho de la ecuación (4.39), la ecuación (4.33) y el hecho que sobre Ω_{h^*} , $\bar{g}_{h^*}(x) = \hat{g}_{h^*}(x)$, tenemos

$$\begin{aligned} \mathbb{E} \left[\frac{1}{(\bar{g}_{h^*}(x))^2} (\bar{g}_{h^*}(x) - g(x))^2 \mathbf{1}_{\Omega_{h^*}} \right] &\leq \mathbb{E} \left[\left(\frac{2}{g_{inf}} \right)^2 (\hat{g}_{h^*}(x) - g(x))^2 \mathbf{1}_{\Omega_{h^*}} \right] \\ &\leq \frac{4}{(g_{inf})^2} \mathbb{E} [(\hat{g}_{h^*}(x) - g(x))^2] \end{aligned} \quad (4.40)$$

Por el segundo término del lado derecho de la ecuación (4.39) se tiene que

$$\begin{aligned} \mathbb{E} \left[\frac{1}{(\bar{g}_{h^*}(x))^2} (\bar{g}_{h^*}(x) - g(x))^2 \mathbf{1}_{\Omega_{h^*}^c} \right] &\leq (\log n)^2 \mathbb{E} [(\bar{g}_{h^*}(x) - g(x))^2 \mathbf{1}_{\Omega_{h^*}^c}] \\ &\leq (\log n)^2 \mathbb{E} [(2(\bar{g}_{h^*}(x))^2 + 2(g(x))^2) \mathbf{1}_{\Omega_{h^*}^c}] \\ &\leq 2(\log n)^2 \{J_1 + J_2\} \end{aligned} \quad (4.41)$$

donde $J_1 = \mathbb{E} [(\bar{g}_{h^*}(x))^2 \mathbf{1}_{\Omega_{h^*}^c}]$ y $J_2 = \mathbb{E} [(g(x))^2 \mathbf{1}_{\Omega_{h^*}^c}]$.

Antes de acotar el termino J_1 de la desigualdad (4.41), debemos recordar que $\bar{g}_{h^*}(x) = \max \left\{ \hat{g}_{h^*}(x), \frac{1}{\log n} \right\}$. Usando que

$$|\hat{g}_{h^*}(x)| = \left| \frac{1}{nh^*} \sum_{i=1}^n K \left(\frac{x - X_i}{h^*} \right) \right| \leq \frac{\|K\|_\infty}{h_{\min}^*}$$

obtenemos que $\bar{g}_{h^*}(x) \leq \left(\frac{\|K\|_\infty}{h_{\min}^*} \vee \frac{1}{\log n} \right)$. Al sustituir el resultado anterior en el término J_1 de la desigualdad (4.41), se tiene

$$J_1 \leq \left(\frac{\|K\|_\infty^2}{(h_{\min}^*)^2} \vee \frac{1}{(\log n)^2} \right) \mathbb{P}[\Omega_{h^*}^c] \quad (4.42)$$

Ahora se procede a acotar el término J_2 de la desigualdad (4.41)

$$J_2 \leq \mathbb{E} \left[(g_{\sup})^2 \mathbf{1}_{\Omega_{h^*}^c} \right] = (g_{\sup})^2 \mathbb{P}[\Omega_{h^*}^c] \quad (4.43)$$

Por las desigualdades (4.41), (4.42) y (4.43), además de la Proposición 24 y como $h_{\min}^* = n^{-1} e^{(\sqrt{\log n} - 1)}$, se tiene que

$$\begin{aligned} & \mathbb{E} \left[\frac{1}{(\bar{g}_{h^*}(x))^2} (\bar{g}_{h^*}(x) - g(x))^2 \mathbf{1}_{\Omega_{h^*}^c} \right] \\ & \leq 2(\log n)^2 \left\{ \left(\frac{\|K\|_\infty^2}{(h_{\min}^*)^2} \vee \frac{1}{(\log n)^2} \right) + (g_{\sup})^2 \right\} \mathbb{P}[\Omega_{h^*}^c] \\ & \leq \frac{2B_{11}(\log n)^2}{n^{13}} \left\{ \left(\frac{\|K\|_\infty^2}{(n^{-1} e^{(\sqrt{\log n} - 1)})^2} \vee \frac{1}{(\log n)^2} \right) + (g_{\sup})^2 \right\} \\ & \leq \frac{\bar{F}_2}{n^4} \end{aligned} \quad (4.44)$$

donde $\bar{F}_2 = 2B_{11} \{ (\|K\|_\infty^2 \vee 1) + (g_{\sup})^2 \}$

Por las ecuaciones (4.39), (4.40) y (4.44), se tiene que

$$\mathbb{E} \left[\frac{1}{(\bar{g}_{h^*}(x))^2} (\bar{g}_{h^*}(x) - g(x))^2 \right] \leq \frac{4}{(g_{\inf})^2} \mathbb{E} [(\hat{g}_{h^*}(x) - g(x))^2] + \frac{\bar{F}_2}{n^4}. \quad (4.45)$$

Finalmente por las desigualdades (4.32), (4.38) y (4.45)

$$\mathbb{E} \left[(r(x) - \hat{r}_{h/h^*}(x))^2 \right] \leq F_1 \mathbb{E} [(m(x) - \hat{m}_h(x))^2] + F_2 \mathbb{E} [(\hat{g}_{h^*}(x) - g(x))^2] + \frac{F_3}{n^4} \quad (4.46)$$

donde $F_1 = \frac{8}{(g_{\inf})^2}$, $F_2 = \frac{8(r_{\sup})^2}{(g_{\inf})^2}$ y $F_3 = 2\bar{F}_1 + 2(r_{\sup})^2 \bar{F}_2$.

4.5.8. Demostración del Teorema 11.

Tenemos de manera similar a Proposición 20

$$\mathbb{E} \left[\left(r(x) - \hat{r}_{\hat{h}/\hat{h}^*}(x) \right)^2 \right] \leq F_1 \mathbb{E} \left[\left(m(x) - \hat{m}_{\hat{h}}(x) \right)^2 \right] + F_2 \mathbb{E} \left[\left(g(x) - \hat{g}_{\hat{h}^*}(x) \right)^2 \right] + \frac{F_3}{n^4}$$

usando $\mathbb{P}(\Omega_{\hat{h}^*}^c) \leq \frac{1}{n^{12}}$ (ver Proposición 25).

Por el Teorema 9 y el Teorema 2 en Bertin and Klutchnikoff (2017), se tiene que

$$\begin{aligned} & \mathbb{E} \left[\left(\hat{r}_{\hat{h}/\hat{h}^*}(x) - r(x) \right)^2 \right] \\ & \leq F_1 \left\{ C^* \left(\frac{n}{\log n} \right)^{-\frac{\beta_1}{2\beta_1+1}} \right\}^2 + F_2 \left\{ C^{**} \left(\frac{\log n}{n} \right)^{\frac{\beta_2}{2\beta_2+1}} \right\}^2 + \frac{F_3}{n^4} \\ & = \mathfrak{B} \left(\frac{n}{\log n} \right)^{-\frac{2(\beta_1 \wedge \beta_2)}{2(\beta_1 \wedge \beta_2)+1}} \end{aligned}$$

donde $\mathfrak{B} = F_1(C^*)^2 + F_2(C^{**})^2 + F_3$ con C^{**} una constante positiva que depende del kernel K , de g_{sup} , Q , β_2 , L_2 y $|\log(a)|$.

Capítulo 5

Conclusión

En esta tesis, estudiamos el problema de la estimación de la función de regresión, cuando ésta es acotada, los errores del modelo son variables aleatorias i.i.d de distribución normal y la variable explicativa es un proceso débilmente dependiente cuyo coeficiente de correlación tiene decaimiento exponencial, además la variable explicativa es idénticamente distribuida con función de densidad g acotada. La precisión de la estimación se hace a través de riesgos puntuales. En este contexto consideramos dos casos de interés que son el principal aporte de este trabajo de investigación; cuando la densidad g es conocida y cuando esta es desconocida.

En el caso en que la densidad es conocida; se demostró que al considerar un estimador de la función de regresión del tipo Nadaraya-Watson (1.9); es decir, al tomar un procedimiento basado en datos utilizando Núcleo y seleccionar el ancho de ventana usando el enfoque de Goldenshluger y Lepski se obtiene que el estimador resultante satisface una desigualdad tipo oráculo y es adaptativo.

En el caso en que la densidad g es desconocida; proponemos como estimador de la función de regresión un estimador cociente de dos estimadores. Se selecciona ventanas con método GL para ambos estimadores que conforman el cociente con método GL. Demostramos desigualdades de tipo oráculo lo que permite demostrar que el estimador de la función de regresión es adaptativo.

Ambos casos generalizan los resultados de Lepski and Serdyukova (2014) que estaban en el contexto de datos independientes.

Como aporte final de esta investigación, se implementaron los distintos métodos en el software R y se desarrollaron simulaciones que permitieron ver el desempeño de los métodos

en riesgo puntual y riesgo integrado.

En el futuro, se propone extender los métodos adaptativos desarrollados en esta tesis a los modelos de regresión no paramétricos con cambio de régimen en un contexto de datos dependientes, desarrollar métodos de estimación adaptativos para modelos de regresión no paramétrica ponderada espacio temporal, y extender las técnicas de estimación adaptativa no paramétrica para datos funcionales.

Bibliografía

- N. Asin and J. Johannes. Adaptive non-parametric estimation in the presence of dependence. *arXiv preprint arXiv:1602.00531*, 2016.
- Y. Baraud. Model selection for regression on a fixed design. *Probability Theory and Related Fields*, 117(4):467–493, 2000.
- Y. Baraud. Model selection for regression on a random design. *ESAIM: Probability and Statistics*, 6:127–146, 2002.
- Y. Baraud, F. Comte, and G. Viennet. Model selection for (auto-) regression with dependent data. *ESAIM: Probability and Statistics*, 5:33–49, 2001.
- K. Bertin and N. Klutchnikoff. Adaptive estimation of a density function using beta kernels. *ESAIM: Probability and Statistics*, 18:400–417, 2014.
- K. Bertin and N. Klutchnikoff. Pointwise adaptive estimation of the marginal density of a weakly dependent process. *Journal of Statistical Planning and Inference*, 187:115–129, 2017.
- K. Bertin, C. Lacour, and V. Rivoirard. Adaptive pointwise estimation of conditional density function. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, 52(2):939–980, 2016.
- K. Bertin, N. Klutchnikoff, J. R. Léon, and C. Prieur. Adaptive density estimation on bounded domains under mixing conditions. *Electronic Journal of Statistics*, 14(1):2198–2237, 2020a.
- K. Bertin, N. Klutchnikoff, F. Panloup, and M. Varvenne. Adaptive estimation of the stationary density of a stochastic differential equation driven by a fractional brownian motion. *Statistical inference for stochastic processes*, 2020b.
- D. Bosq. *Nonparametric statistics for stochastic processes: estimation and prediction*, volume 110. Springer Science & Business Media, 2012.
- P. Bühlmann, P. Doukhan, and P. A. Nze. Weak dependence beyond mixing and asymptotics for nonparametric regression. *The Annals of Statistics*, 30(2):397–430, 2002.
- M. Chichignoud, V. H. Hoang, T. M. P. Ngoc, and V. Rivoirard. Adaptive wavelet multivariate regression with errors in variables. *Electronic journal of statistics*, 11(1):682–724, 2017.
- F. Comte, C. Prieur, and A. Samson. Adaptive estimation for stochastic damping hamiltonian systems under partial observation. *Stochastic Processes and their Applications*, 2017.

- J. Dedecker, P. Doukhan, G. Lang, S. León, R. Louhichi, and C. Prieur. Weak dependence: with examples and applications. *Lecture Notes in Statistics*, 190, 2007.
- P. Doukhan. Mixing, volume 85 of lecture notes in statistics, 1994.
- P. Doukhan and S. Louhichi. A new weak dependence condition and applications to moment inequalities. *Stochastic processes and their applications*, 84(2):313–342, 1999.
- P. Doukhan and S. Louhichi. Functional estimation of a density under a new weak dependence condition. *Scandinavian journal of statistics*, 28(2):325–341, 2001.
- P. Doukhan and M. H. Neumann. Probability and moment inequalities for sums of weakly dependent random variables, with applications. *Stochastic Processes and their Applications*, 117(7):878–903, 2007.
- P. Doukhan and O. Wintenberger. Weakly dependent chains with infinite memory. *Stochastic Processes and their Applications*, 118(11):1997–2013, 2008.
- L. J. Fermin, R. Rios, and L. A. Rodriguez. A robbins–monro algorithm for non-parametric estimation of nar process with markov switching: Consistency. *Journal of Time Series Analysis*, 2017.
- A. Goldenshluger and O. Lepski. Bandwidth selection in kernel density estimation: oracle inequalities and adaptive minimax optimality. *The Annals of Statistics*, 39(3):1608–1632, 2011.
- A. Goldenshluger and O. Lepski. General selection rule from a family of linear estimators. *Theory Prob. Appl.*, 57(2):209–226, 2013.
- I. A. Ibragimov and R. Khasminskii. Asymptotic theory of estimation. *Nauka, Mosow*, 1: 979, 1979.
- O. Lepski and N. Serdyukova. Adaptive estimation under single-index constraint in a regression model. *The Annals of Statistics*, 42(1):1–28, 2014.
- P. Massart. *Concentration inequalities and model selection*. Springer, 2007.
- E. A. Nadaraya. *Nonparametric estimation of probability densities and regression curves*. Springer, 1989.
- N. B. Nguyen. *Adaptation via des inégalités d’oracle dans le modèle de regression avec design aléatoire*. PhD thesis, Aix-Marseille, 2014.
- B. W. Silverman. *Density estimation for statistics and data analysis*, volume 26. CRC press, 1986.
- C. J. Stone. Consistent nonparametric regression. *The annals of statistics*, pages 595–620, 1977.
- A. B. Tsybakov. *Introduction to nonparametric estimation*. Springer Series in Statistics. Springer, New York, 2009. ISBN 978-0-387-79051-0. doi: 10.1007/b13794. URL <http://dx.doi.org/10.1007/b13794>. Revised and extended from the 2004 French original, Translated by Vladimir Zaiats.

Apéndice A

Programa en R para la estimación adaptativa de la función de regresión correspondiente al capítulo 3

A continuación se presenta el programa desarrollado en R que se utilizó en la sección 3.4, para estimar la función de regresión $r(x) = 0,7x + 2e^{-10x^2}$ en el modelo $Y_i = r(X_i) + \varepsilon_i$ para $i = 1, \dots, n$, donde $\{X_i\}_{i=1}^n$ es una muestra del proceso \mathbb{X} presentado en el Ejemplo 2 y las variables ε_i son independientes e idénticamente distribuidas de distribución $N(0, \sigma^2)$ con $\sigma > 0$. Se asume que la densidad de la variable explicativa es conocida.

El programa tiene como entradas las variables; n tamaño de la muestra a generar y $sig1$ desviación estándar de los errores $\{\varepsilon_i\}_{i=1}^n$. Con esa información el programa genera la muestra aleatoria $\{(X_i, Y_i)\}_{i=1}^n$. Posteriormente usando la muestra generada, el programa calibra el método y determina la estimación adaptativa de la función de regresión en una malla de longitud 21 del intervalo $[-1, 1]$, esos valores estimados se almacenan en la variable de salida *rchu1*.

Antes de describir las salidas gráficas del programa, se debe destacar que se tomo la familia de ventanas \mathcal{H} como se define en la sección 3.4.2 y como valores del parámetro de calibración del método la secuencia $\{\gamma_j\}_{j=1}^{21}$ como se indica en la sección 3.4.3.

En el proceso de estimación el programa proporciona varias salidas gráficas:

- 1) Gráfico de dispersión de la muestra generada $\{X_i\}_{i=1}^n$.
- 2) Gráfico de la función de correlacion de $\{X_i\}_{i=1}^{100}$.

- 3) Veintiún gráficos que muestran paso a paso el proceso de calibración del método para estimar la función de regresión $r(\cdot)$. Para cada $\{\gamma_j\}_{j=1}^{21}$ se proporciona un gráfico; representando con una línea roja el estimador GL $\{\hat{r}_{\hat{h}_i^{\gamma_j}}(\tilde{x}_i)\}_{i=1}^q$ de $\{r(\tilde{x}_i)\}_{i=1}^q$, con una línea negra la función a estimar, con una línea azul el estimador $\hat{r}_h(\cdot)$ en una ventana fija y con puntos verdes la muestra de entrenamiento $\{(\tilde{x}_i, \tilde{y}_i)\}_{i=1}^p$ con $0 < p \leq 100$, descrita en la sección 3.4.3, que es usada para calibrar el método.
- 4) Gráfico de la función $Error(\gamma)$, descrito en la sección 3.4.3, que muestran el error cometido entre las veintiún estimaciones de la función de regresión $r(\cdot)$ y los puntos de entrenamiento $\{(\tilde{x}_i, \tilde{y}_i)\}_{i=1}^p$ con $0 < p \leq 100$. El valor γ que minimiza la secuencia $\{Error(\gamma_j)\}_{j=1}^{21}$ es el que determina al mejor estimador, el cuál se denota por $\{\hat{r}_{\hat{h}_i^{\gamma}}(\tilde{x}_i)\}_{i=1}^q$.
- 5) Gráfico del estimador adaptativo de la función de regresión $\hat{r}_{\hat{h}_i^{\gamma}}(x_i)$ en una malla igualmente espaciada $\{x_i\}_{i=1}^{21}$ del intervalo $[-1, 1]$, donde el valor γ con el que se aplicó el método GL es el obtenido en el ítem (4); se representa con una línea punteada la estimación adaptativa, con una línea continua la función a estimar y con puntos grises la muestra de estimación.

Observación sobre los valores de n y $sig1$ en el programa. Se debe introducir un valor de n mayor a doscientos, ya que los primeros $n - 200$ datos es la muestra de estimación y los últimos 100 datos es toma la muestra de entrenamiento, como se describe en la sección 3.4.3. Al ser $Sig1$ una desviación estándar su valor tiene que ser mayor a cero.

```
#####
#####
###Programa: Estimacion adaptativa de la funcion de regresion ,
###Usando nucleo gaussiano , con g conocida .
#####
#####

n<-2200      # Longitud de la muestra que se generara .
sig1<-0.5    # Desviacion Estandar del Error de la regresion .

#####
#####
#####      INICIO CALIBRACION      #####
```

```
#####
#####

#####
### Generacion de numeros aleatorios de distribucion  $N(0,1)$ 
### truncada por  $1_{\{-a,a\}}(x)$  dependientes usando datos
### AR(1) con  $\phi=0.75$  y el metodo de la transformada
### inversa.
#####

Z=arima.sim(list(ar=0.75),n=n) # Generacion de AR(1) con
                               #  $\phi=0.75$  y  $\sigma=1$ 
U=pnorm(Z,mean=0,sd=sqrt(1/(1-0.75^2)))
    #  $U=F(Z)$ , donde  $Z\sim N(0,1/(1-0.75^2))$ 
    # y  $F$  es su funcion de distribucion.
    #  $U\sim U(0,1)$  y es dependiente.
a=2; # Soporte para la normal truncada, dependiente.
p=pnorm(a)-pnorm(-a); # Area bajo la normal  $(0,1)$  truncada
    # en  $a$ , tal densidad se denota por  $g$ 
    # y ademas  $g=G'$ .
X=qnorm(p*U+pnorm(-a),mean=0,sd=1) #  $X=G^{-1}(U)=(G^{-1}\circ F)(Z)$ .
    # Normal truncada dependiente.

plot(seq(1,n,1),X,xlab="_",ylab="_") # Grafico nube de puntos
    # de datos generados,
    # variable explicativa.
acf(X,lag=100) # Grafico de la funcion de correlacion de la
    # variable explicativa

T1=X

Y0t=0.7*T1+2*exp(-10*T1^2) # Funcion de regresion a estimar
    #  $r(x)=0.7x + 2\exp(-10x^2)$ .

Yt=Y0t+rnorm(n,0,sig1) # Variable respuesta
```

```

ne<-n-200 # Numero de datos con los que se construye el
          # estimador
X=T1[1:(n-200)] # Datos de la variable explicativa con los que
                # se construye el estimador

Y0=Y0t[1:(n-200)] # Y0=r(X), donde X es normal truncada
                  # dependiente.
Y=Yt[1:(n-200)] # Modelo Y=r(X)+e donde e~N(0,1), datos de
                 # la variable respuesta con los que se
                 # construye el estimador
    
```

```

#####
##### Estimacion desviacion estandar de e_j #####
##### donde e_j es el error de la regresion #####
#####
    
```

```

library (KernSmooth)
h<-dpill(X,Y) # Metodo plug-in de Ruppert, Sheather y
              # Wand (1995), esta ventana obtenida por el
              # metodo del pulgar la sustituir en el
              # estimador de N-W, para asi tener una
              # estimacion de r en cada punto de X en R^2000.
    
```

```

#####
##### MODULO 1 m #####
#####
    
```

```

Estimmm<-function (x,X,Y,h){
  # Estimador de la funcion m con kernel gaussiano, ventana
  # h, muestra aleatoria de la variable explicativa X,
  # muestra aleatoria de la variable respuesta Y, y variable
  # independiente x.
  #  $m(x)=\sum_{i=1}^n (1/(nh*\sqrt{2*\pi}))Y[i]\exp(-((x-
  # X[i])/h)^2)/2)$ .
  n<-length (X);
  mm=0;
    
```

```

for(i in 1:n){
  mm=mm+(1/(n*h*sqrt(2*pi)))*Y[i]*exp(-(((x-X[i])/h)^2)/2);
}
return (mm)
}

#####
##### MODULO 1 g #####
#####
Estimg<-function(x,X,h){
  # Estimador de la densidad g con kernel gaussiano, ventana h
  # para la muestra aleatoria X envaluada en x.
  #  $g(x) = \sum_{i=1}^n (1/(nhsqrt(2*pi))) \exp(-(((x-X[i])/h)^2)/2)$ .
  n<-length(X);
  g=0;
  for(i in 1:n){
    g=g+(1/(n*h*sqrt(2*pi)))*exp(-(((x-X[i])/h)^2)/2);
  }
  return (g)
}

#####
##### Estimador mpul, gpul y rpul #####
#####
mpul<-0;
gpul<-0;
for(i in 1:ne){
  mpul[i]<-Estimmm(X[i],X,Y,h); # Numerador del estimador de
                                # Nadaraya - W
  gpul[i]<-Estimg(X[i],X,h);   # Denominador del estimador de
                                # Nadaraya - W
}
rpul<-mpul/gpul # Estimador de Nadaraya - W, en la ventana h,
                 # estimada por el metodo del pulgar.
#####

```

```
sigqchu<-(1/(ne-1))*sum((Y-rpul)^2) # Estimacion Var(e_j)
sigchu<-sqrt(sigqchu) # Estimacion desviacion estandar de e_j
sig<-sigchu # Desviacion estandar de e_j (Estimacion
# del error de la regresion)
```

```
#####
##### Fin Estimacion desviacion estandar e_j #####
#####
```

```
t1=T1[((n-100)+1):n]
t2=Yt[((n-100)+1):n]
ii=order(t1)
x=t1[ii]
Yct<-t2[ii]
```

```
#####
##### Formula de la densidad g, normal(0,1) truncada #####
#####
```

```
mu<-0;
sigm<-1;
#Extremo del soporte es a=2.
dTnorm<-function(x,mu,sigm,a){
  #Densidad N(mu,sigm) truncada de -a hasta a.
  p<-pnorm(mu+a,mu,sigm,lower.tail=TRUE,log.p=FALSE)-
  pnorm(mu-a,mu,sigm,lower.tail=TRUE,log.p=FALSE);
  if(mu-a<=x && x<=mu+a){g<-dnorm(x,mu,sigm,log=FALSE)/p;}
  else {g=0;}
  return(g)
}
```

```
#####
```

```
#####
##### MODULO 1 #####
#####
```

```
EstimR<-function(x,X,Y,h){
```

```

# Estimador de la regresion con kernel gaussiano, ventana h
# para la muestra aleatoria (X,Y) envaluada en x.
#  $r(x) = \sum_{i=1}^n (1/(n\sqrt{2\pi})) Y[i] \exp(-((x - X[i])/h)^2/2)$ 
#  $g^{-1}(X[i])$ .
n<-length(X);
r=0;
for(i in 1:n){
  r=r+(1/(n*h*sqrt(2*pi)))*Y[i]*exp(-(((x-X[i])/h)^2)/2)
  *dTnorm(X[i],mu, sigma, a)^(-1);
}
return(r)
}
#####

#####
##### MODULO 2 #####
#####
EstimRSS<-function(x,X,Y,h,h1){
# Estimador sobre suavizado con nucleo gaussiano, muestra
# aleatoria simple (X,Y), ventanas h y h1, evaluado en x.
n<-length(X);
rss=0
for(i in 1:n){
  rss=rss+(1/(n*sqrt(2*pi*(h^2+h1^2))))*Y[i]*exp(-((x-
  X[i])^2)/(2*(h^2+h1^2)))*dTnorm(X[i],mu, sigma, a)^(-1);
}
return(rss)
}
#####

#####
##### MODULO 3 #####
#####
# Esta funcion determina el valor absoluto de la diferencia

```

```

# entre el estimador de la regresion sobreesuavizado rhh1 y rh1
# en un valor x del soporte.
VAdifEst<-function(x,X,Y,h,h1){
  VA<-abs(EstimRSS(x,X,Y,h,h1)-EstimR(x,X,Y,h1))
  return(VA)
}
#####

ne<-length(X); # Longitud de la muestra de enterenamiento
I<-0
xt<-0
Yc<-0
for(i in 1:100){
  if(-1<=x[i] & x[i]<=1){I<-I+1; xt[I]<-x[i]; Yc[I]<-Yct[i]}
}
x<-xt

s<-length(x) # Numero de elementos en la particion del soporte
# de la densidad de X
M<-floor(log(ne))*(2/3) # Definicion de M, donde M es la parte
# entera n/log(n),
H<-exp(-seq(0,M,0.1)) # Familia de ventanas, donde H={h_i}_0^M,
# con M definido en
m<-length(H) # Numero de ventanas.

#####
#####
#### Rutina para almacenar los datos
#### usando la funcion VAdifEst
#### en una lista de nombre l que
#### contiene s matrices donde cada
#### matriz es de orden (mxm)
#### s=length(x) y m=length(H).
#### El almacenamiento en cada matriz

```

```

#### se realiza por filas.
#####

l<-list()      # Nombre de la lista que almacenara s matrices
               # donde cada una es de orden mxm.
for(k in 1:s){
  10<-0        # Variable auxiliar que almacenara temporalmente
               # cada matriz de datos.
  100<-0       # Variable auxiliar que almacena temporalmente
               # cada fila de la matriz 10.
  for(j in 1:m){
    100<-VAdifEst(x[k],X,Y,H,H[j])
    if(j==1) {10<-100}
    else {10<-cbind(10,100)}
  }
  l[[k]]<-10
}
#####

#####

##### Cambio Local de rsup y ginf      IMPORTANTE   ###
##### Ahora rsup y ginf son vectores de dimension s   ###
##### Cada rsup[i] y ginf[i] corresponde con x[i],     ###
##### para i=1:s                                       ###
#####

d<-rep(0,s)    #d distancias para el calculo de Int(rchu-Y)^2
d[1]<-((x[1]+x[2])/2)-(-1)
d[s]<-1-((x[s-1]+x[s])/2)
for(i in 2:(s-1)){
  d[i]<-((x[i]+x[i+1])/2)-((x[i-1]+x[i])/2)
}

e<-rep(0,s+1)
se<-length(e)

```

```

e[1]<-(-1)
e[se]<-1

for (i in 2:(se-1)) {
  e[i]<-(x[i-1]+x[i])/2
}

rsup<-rep(0,s)
ginf<-rep(0,s)

for(i in 1:s){
  rsup[i]<-max(abs(Y*(e[i]<X)*(X<e[i+1]))) # Maximo Local de
                                           # la funcion de
                                           # regresion.

  W1<-0
  W1<-X*(e[i]<X)*(X<e[i+1])
  W2<-0
  I1<-1
  for(j in 1:ne){
    if(W1[j]!=0){W2[I1]<-X[j]; I1<-I1+1}
  }
  valorg<-rep(0,length(W2))
  for (j in 1:length(W2)) {
    valorg[j]<-dTnorm(W2[j],0,1,a)
  }
  ginf[i]<-min(valorg) #Minimo de la funcion de densidad de X.
}
##### FIN IMPORTANTE, CAMBIO LOCAL #####
##### FIN IMPORTANTE, CAMBIO LOCAL #####

del<-((log(ne))^(1/5))
N2<-1/sqrt(2*sqrt(pi)) #Norma 2 del nucleo gaussiano.

#####

```

```

##### Cambio Local de A3          IMPORTANTE          ###
##### Ahora A3 es un vector de dimension s          ###
##### Cada A3[i] corresponde con x[i], para cada i=1:s. ###
#####
A3<-rep(0,s)
for(i in 1:s){
  A3[i]<-(rsup[i]^2+sig^2)*ginf[i]^(-1)*N2^2
}
##### FIN IMPORTANTE, CAMBIO LOCAL #####
##### FIN IMPORTANTE, CAMBIO LOCAL #####

GAM1<-0.00000005
GAM21<-0.05
dd<-(GAM21-GAM1)/20
GAM<-seq(GAM1,GAM21,dd) # Valores de la variable gamma,
                        # usados para calibrar el metodo

Norma2Cuadrado<-rep(0,length(GAM));

for(q in 1:length(GAM)){
  print(q)
  gam<-GAM[q] #Parametro gamma > 2. (valor que he cambiado)
  B4<-sqrt(2*gam*A3)*(2) # El 2 que multiplica despues de la
                        # raiz cuadrada es 2=\|K\|_1 +1,
                        # donde \|K\|_1 =1
                        # "IMPORTANTE ahora B4 es una variable LOCAL"
                        # "Es claro que B4 es un vector de dimension s".
                        # Cada B4[i] corresponde a x[i], para cada i=1:s
                        # IMPORTANTE en esta parte no hubo cambios.

#####
##### Calculo del estimador V(x,h), para cada x en      ##
##### la malla de longitud s y h \in H.                  ##
##### Es decir V es una matriz de orden sxm.            ##
#####
##### IMPORTANTE aca se hacen cambios locales,          ##
##### Ahora V tambien depende de x, ademas de           ##

```

```
##### depender de H (como antes) #####
##### sera una matriz de oden sxm, donde cada fila #####
##### corresponde a un valor de x y cada columna a #####
##### los valores de H. V[i,j] donde i=1:s y j=1:m #####
#####
V<-matrix(0 ,nrow=s , ncol=m)
for (k in 1:m){
  V[,k]<-B4*(1+del)*sqrt(log(ne))/sqrt(ne*H[k])
      # Aparece ne, pues es la longitud
      # de la muestra de entrenamiento
}
##### Fin del cambio local de V #####
#####

#####
##### Calculo del Maximo de cada fila #####
##### de la matriz {l[k]-V}_{+}, #####
##### para cada k de 1:s. #####
##### Lo cual se guarda en la lista #####
##### A={A[k]}_{k=1;s} con A[k] en R^m #####
##### Se genera la lista AV, compuesta #####
##### de s vectores m dimensionales #####
##### AV[[k]]=A[[k]]+V en R^m para #####
##### k=1,...,s. #####
##### Se geneta el vector hopt en R^s #####
##### donde hopt[k]=argmin (AV[[k]]+V) #####
#####
A<-list () # Es una lista de s vectores , donde cada
# vector A[k] en R^m, corresponde a un
# valor x[k] en de la malla , y A[k][i]
# es el valor en H[i]
AV<-list () # Es una lista de vectores , donde cada
# vector AV[k] coresponde a la suma de
# vectores A[k]+V[k,].
arg<-0 # Esta Variable sera un vector donde
```

```

        # cada componente es
        #  $arg[k]=argmin_{h \in H} (A[k]+V[k,])$ 
hopt<-0 # Variable que almacena las ventanas optimas
        # para cada  $x[k]$  en la malla.

for(k in 1:s){
  A0<-0 # Esta variable auxiliar, sera una matriz que
        # almacenara temporalmente los datos contenidos
        # en cada matriz  $A[[k]]$  de la lista A
  for(i in 1:m){
    A0[i]<-max(1[[k]][i,]-V[k,])
    if(A0[i]<0) {A0[i]<-0}
  }
  A[[k]]<-A0 # Se almacenan los datos en la lista A
  AV[[k]]<-A0+V[k,] # Se almacenan los datos en la lista AV
  arg[k]<-which.min(AV[[k]]) # Se almacena la posicion de la
        # ventana optima para cada  $x[k]$ 
        # de la malla de x.
}

hopt<-H[arg] # Vector de ventanas optimas,
             # es un vector en  $R^s$ 

#####
##### Estimador rchu #####
#####
rchu<-0;
TMSE0<-0; # Variable auxiliar que almacenara temporalmente
          # los terminos del MSE de la estimacion.
for(i in 1:s){
  rchu[i]<-EstimR(x[i],X,Y,hopt[i]);
  TMSE0[i]<-(rchu[i]-Yc[i])^2;
}
#####

```

```
#####
##### Grafico del estimador rchu para cada GAM[q], #####
##### con q=1:\length(GAM) #####
#####
plot(x, rchu, ylim=c(-1.5, 2.5))
lines(x, rchu, lwd=2, col="red")
points(X, Y0, col="black") #Curva Y0=r(X)

r<-EstimR(x,X,Y,1/6); # Estimador de nucleo en una ventana
# fija MODULO 1
lines(x, r, lwd=2, col='blue')
points(x, Yc, lwd=2, col='green')
#####

Norma2Cuadrado[q]<-sum(d*TMSE0)
}

plot(Norma2Cuadrado, main="Calibracion_Metodo_GL", xlab="_", ylab="_")
lines(Norma2Cuadrado)

i0<-which.min(Norma2Cuadrado)
i0
GAM[i0]
#####
#####
##### Final Calibracion #####
#####
#####

#####
#####
##### INICIO AJUSTE #####
#####
#####
a<-2
```

```

a1<-1      # Limite superior del soporte de la densidad de X
x1<-seq(-a1,a1,0.1)  # Soporte de la densidad de X
s1<-length(x1)  # Numero de elementos en la particion del
                # soporte de la densidad de X

#####
#####
##### Rutina para almacenar los datos #####
##### usando la funcion VAdifEst #####
##### en una lista de nombre l que #####
##### contiene s matrices donde cada #####
##### matriz es de orden (mxm) #####
##### s=length(x) y m=length(H). #####
##### El almacenamiento en cada matriz #####
##### se realiza por filas. #####
#####
l<-0      # Nombre de la lista que almacenara s matrices
          # donde cada una es de orden mxm.

for(k in 1:s1){
  l0<-0    # Variable auxiliar que almacenara temporalmente
           # cada matriz de datos.
  l100<-0  # Variable auxiliar que almacena temporalmente
           # cada fila de la matriz l0.
  for(j in 1:m){
    l100<-VAdifEst(x1[k],X,Y,H,H[j])
    if(j==1) {l10<-l100}
    else {l10<-cbind(l10,l100)}
  }
  if(k==1) {l<-list(l10)}
  else {l[[k]]<-l10}
}

#####
#####
##### Cambio Local de rsup y ginf      IMPORTANTE #####
##### Ahora rsup y ginf son vectores de dimension s #####

```

```
##### Cada rsup[i] y ginf[i] corresponde con x[i], #####
##### para i=1:s #####
#####
#####
rsup<-rep(0,s1)
ginf<-rep(0,s1)
for(i in 1:s1){
  rsup[i]<-max(abs(Y*((x1[i]-0.05)<X)*(X<(x1[i]+0.05))))
  # Maximo Local de la funcion de regresion. #####
  W1<-0
  W1<-X*((x1[i]-0.05)<X)*(X<(x1[i]+0.05))
  W2<-0
  I1<-1
  for(j in 1:ne){
    if(W1[j]!=0){W2[I1]<-X[j]; I1<-I1+1}
  }
  valorg<-rep(0,length(W2))
  for(j in 1:length(W2)) {
    valorg[j]<-dTnorm(W2[j],0,1,a)
  }
  ginf[i]<-min(valorg) #Minimo de la funcion de densidad de X.
}
##### FIN IMPORTANTE, CAMBIO LOCAL #####
##### FIN IMPORTANTE, CAMBIO LOCAL #####
#####

gam<-GAM[i0] # Parametro gamma
del<-((log(ne))^(1/5))
N2<-1/sqrt(2*sqrt(pi)) # Norma 2 del nucleo gaussiano.

#####
##### Cambio Local de A3 IMPORTANTE #####
##### Ahora A3 es un vector de dimension s #####
##### Cada A3[i] corresponde con x[i], #####
##### para cada i=1:s. #####
#####
```

```

A3<-rep(0,s1)
for(i in 1:s1){
  A3[i]<-(rsup[i]^2+sig^2)*ginf[i]^(-1)*N2^2
}
##### FIN IMPORTANTE, CAMBIO LOCAL #####
##### FIN IMPORTANTE, CAMBIO LOCAL #####
#####

B4<-sqrt(2*gam*A3)*(2)      # El 2 que multiplica despues de
                             # la raiz cuadrada es
                             #  $2 = \sqrt{|K| - 1} + 1$ , donde  $\sqrt{|K| - 1} = 1$ 

#####
##### Calculo del estimador  $V(h)$ , para cada  $h \in H$ . #####
##### IMPORTANTE aca se hacen cambios locales, #####
##### Ahora  $V$  tambien depende de  $x$ , ademas de  $h$  #####
#####  $V$  sera una matriz de orden  $s \times m$ , donde cada #####
##### fila corresponde a un valor de  $x$  y cada #####
##### columna a los valores de  $H$ . #####
#####  $V[i, j]$  donde  $i = 1:s$  y  $j = 1:m$  #####
#####

V<-matrix(0,nrow=s1,ncol=m)
for(k in 1:m){
  V[,k]<-B4*(1+del)*sqrt(log(ne))/sqrt(ne*H[k])
}
#####
##### Fin del cambio local de  $V$ . #####
#####

#####
##### Calculo del Maximo de cada fila #####
##### de la matriz  $\{l[k] - V\}_{+}$ , #####
##### para cada  $k$  de  $1:s$ . #####
##### Lo cual se guarda en la lista #####
#####  $A = \{A[k]\}_{k=1:s}$  con  $A[k]$  en  $R^m$  #####
##### Se genera la lista  $AV$ , compuesta #####

```

```
##### de s vectores m dimensionales #####
##### AV[[k]]=A[[k]]+V en R^m para #####
##### k = 1, ..., s. #####
##### Se genera el vector hopt en R^s #####
##### donde hopt[k]=argmin (AV[[k]]+V) #####
#####
A<-list() # Es una lista de s vectores, donde cada
          # vector A[k] en R^m, corresponde a un valor
          # x[k] de la malla, y A[k][i] es el
          # valor en H[i]
AV<-list() # Es una lista de s vectores, donde cada vector
          # AV[k] corresponde a la suma de vectores A[k]+V.
arg<-0 # Esta Variable sera un vector donde cada
        # componente es arg[k]=argmin_{h in H} (A[k]+V)
hopt<-0 # Variable que almacena las ventanas optimas
        # para cada x[k] en la malla.

for(k in 1:s1){
  A0<-0 # Esta variable auxiliar, sera una matriz que
        # almacenara temporalmente los datos contenidos
        # en cada matriz A[[k]] de la lista A
  for(i in 1:m){
    A0[i]<-max(1[[k]][i,]-V[k,])
    if(A0[i]<0) {A0[i]<-0}
  }
  ##### Se genera la lista A y AV #####
  A[[k]]<-A0; # Se guardan resultados en la lista A
  AV[[k]]<-A0+V[k,] # Se guardan resultados en la lista AV
  arg[k]<-which.min(AV[[k]]) # Se almacena la ventana optima
                            # para cada x[k] de la malla de x
}
hopt<-H[arg]
#####
#####
```

```

##### Estimador rchu #####
#####
rchu1<-0;
for(i in 1:s1){
  rchu1[i]<-EstimR(x1[i],X,Y,hopt[i]);
}
#####

#####
#####
##### FINALIZO EL AJUSTE #####
#####
#####

#####
##### GRAFICOS FINALES #####
#####
plot(x1,rchu1,ylim=c(-2,3),type="n",main="Estimacion_Adaptativa
Metodo_GL_para_r",xlab="",ylab="")
points(X,Y,col="gray")
lines(x1,rchu1,lwd=2,lty="dotted")

#### Funcion de regresion a estimar  $r(x)=0.7x+2\exp(-10x^2)$ 
#### evaluada en la malla x1 del intervalo [-1,1]
r=0.7*x1+2*exp(-10*x1^2)
lines(x1,r,lwd=2,col='black')

```


Apéndice B

Programa en R para la estimación adaptativa de la función de regresión correspondiente al capítulo 4

A continuación se presenta el programa desarrollado en R que se utilizó en la sección 4.4, para estimar la función de regresión $r(x) = 0,7x + 2e^{-10x^2}$ en el modelo $Y_i = r(X_i) + \varepsilon_i$ para $i = 1, \dots, n$, donde $\{X_i\}_{i=1}^n$ es una muestra del proceso \mathbb{X} presentado en el Ejemplo 4 y las variables ε_i son independientes e idénticamente distribuidas de distribución $N(0, \sigma^2)$ con $\sigma > 0$. Se asume que la densidad de la variable explicativa es desconocida.

El programa tiene como entradas las variables; n tamaño de la muestra a generar y $sig1$ desviación estándar de los errores $\{\varepsilon_i\}_{i=1}^n$. Con esa información el programa genera la muestra aleatoria $\{(X_i, Y_i)\}_{i=1}^n$. Posteriormente usando la muestra generada, el programa calibra el método y determina la estimación adaptativa de la función de regresión en una malla de longitud 41 del intervalo $[-2, 2]$ esos valores estimados se almacenan en la variable de salida *rchu*. Además, si se escribe el comando *rchu[j2]* se obtienen los valores del estimador adaptativo de la función de regresión en una malla igualmente espaciada de longitud 21 del intervalo $[-1, 1]$.

Antes de describir las salidas gráficas del programa, se debe destacar que se tomo la familia de ventanas $\mathcal{H} = \mathcal{H}^*$ como se define en la sección 4.4.2 y como valores de los parámetros de calibración del método la secuencia $\{\gamma_j\}_{j=1}^{21} = \{\gamma_j^*\}_{j=1}^{21}$ como se indica en la sección 4.4.3.

En el proceso de estimación el programa proporciona varias salidas gráficas:

- 1) Gráfico de dispersión de la muestra generada $\{X_i\}_{i=1}^n$.
- 2) Gráfico de la función de correlación de $\{X_i\}_{i=1}^{100}$.
- 3) Histograma de la muestra $\{X_i\}_{i=1}^n$, junto con la curva de la densidad g (que se pretende estimar).
- 4) Veintiún gráficos que muestran paso a paso el proceso de calibración del método para estimar la función de densidad $g(\cdot)$. Para cada $\{\gamma_j^*\}_{j=1}^{21}$ se proporciona un gráfico; representando con una línea roja el estimador GL $\{\hat{g}_{\hat{h}_i^{\gamma_j^*}}(\tilde{x}_i)\}_{i=1}^q$ de $\{g(\tilde{x}_i)\}_{i=1}^q$, con una línea azul la función histograma de los datos y con una línea negra la función a estimar $g(\cdot)$. La secuencia $\{\tilde{x}_i\}_{i=1}^q$ es una partición del intervalo $[-2, 2]$, como se especifica en la sección 4.4.3 y el programa toma $q = 100$.
- 5) Gráfico de la función $Error_g(\gamma^*)$, descrito en la sección 4.4.3, que muestran el error cometido entre las veintiún estimaciones de la función de densidad y la función histograma. El valor γ_g que minimiza la secuencia $\{Error_g(\gamma_j^*)\}_{j=1}^q$ es el que determina al mejor estimador, el cuál se denota por $\{\hat{g}_{\hat{h}_i^{\gamma_g}}(\tilde{x}_i)\}_{i=1}^q$.
- 6) Gráfico del estimador adaptativo de la función de densidad en el intervalo $[-2, 2]$; con una línea punteada se representa el estimador GL de la densidad $\{\hat{g}_{\hat{h}_i^{\gamma_g}}(\tilde{x}_i)\}_{i=1}^q$ y con una línea continua se representa la función de densidad $g(\cdot)$.
- 7) Veintiún gráficos que muestran paso a paso el proceso de calibración del método para estimar la función $m(\cdot)$. Para cada $\{\gamma_j\}_{j=1}^{21}$ se proporciona un gráfico; representando con una línea roja el estimador GL $\{\hat{m}_{\hat{h}_i^{\gamma_j}}(\tilde{x}_i)\}_{i=1}^q$ de $\{m(\tilde{x}_i)\}_{i=1}^q$ y con puntos amarillos la secuencias de puntos de entrenamiento $\{\tilde{y}_i \hat{g}_{\hat{h}_i^{\gamma_g}}(\tilde{x}_i)\}_{i=1}^q$. Los puntos de entrenamiento están especificados en la sección 4.4.3, como se mencionó previamente el programa toma $q = 100$.
- 8) Gráfico de la función $Error_m(\gamma)$, descrito en la sección 4.4.3, que muestran el error cometido entre las veintiún estimaciones de la función $m(\cdot)$ y los puntos de entrenamiento $\{\tilde{y}_i \hat{g}_{\hat{h}_i^{\gamma_g}}(\tilde{x}_i)\}_{i=1}^q$. El valor γ_m que minimiza la secuencia $\{Error_m(\gamma_j)\}_{j=1}^q$ es el que determina al mejor estimador, el cuál se denota por $\{\hat{m}_{\hat{h}_i^{\gamma_m}}(\tilde{x}_i)\}_{i=1}^q$.
- 9) Gráfico del estimador adaptativo de la función $m(\cdot)$ en el intervalo $[-2, 2]$; con una línea punteada se representa el estimador GL $\{\hat{m}_{\hat{h}_i^{\gamma_m}}(\tilde{x}_i)\}_{i=1}^q$, con una línea continua se representa la función $m(\cdot)$ y con puntos amarillos la secuencia de puntos de entrenamiento $\{\tilde{y}_i \hat{g}_{\hat{h}_i^{\gamma_g}}(\tilde{x}_i)\}_{i=1}^q$.

- 10) Gráfico del estimador adaptativo de la función de regresión $\hat{r}_{\hat{h}_i^{\gamma_m} / \hat{h}_i^{\gamma_g}}(x_i) = \frac{\hat{m}_{\hat{h}_i^{\gamma_m}}(x_i)}{\hat{g}_{\hat{h}_i^{\gamma_g}}(x_i)}$ en una malla igualmente espaciada $\{x_i\}_{i=1}^{41}$ del intervalo $[-2, 2]$, donde $\{\hat{g}_{\hat{h}_i^{\gamma_g}}(x_i)\}_{i=1}^{41}$ es el estimador GL de $\{g(x_i)\}_{i=1}^{41}$ calibrado en el valor γ_g obtenido en el ítem (6) y $\{\hat{m}_{\hat{h}_i^{\gamma_m}}(x_i)\}_{i=1}^{41}$ es el estimador GL de $\{m(x_i)\}_{i=1}^{41}$ calibrado en el valor γ_m obtenido en el ítem (9); se representa con una línea punteada el estimador adaptativo, con una línea continua la función de regresión y con puntos amarillos los datos de estimación generados.
- 11) Gráfico del estimador adaptativo de la función de regresión en una partición $\{X_i\}_{i=1}^{21}$ igualmente espaciada del intervalo $[-1, 1]$, con especificaciones similares a las descritas en el ítem (10).

Observación sobre los valores de n y $sig1$ en el programa. Se debe introducir un valor de n mayor a doscientos, ya que los primeros $n - 200$ datos es la muestra de estimación y los últimos 100 datos determinan la muestra de entrenamiento, como se describe en la sección 4.4.3. Al ser $Sig1$ una desviación estándar su valor tiene que ser mayor a cero.

```
#####
#####  Estimador de r con densidad g desconocida. #####
#####          Calibracion de gchu usando          #####
#####                      || gchu-HHg || _2^2      #####
#####          donde gchu es la estimacion de la    #####
#####          funcion de densidad y HHg un histograma #####
#####          V(h) distinto al articulo           #####
#####          de Bertin y Klutchnikoff            #####
#####
```

```
n<-2200    # Numero de datos a generar , para la muestra
sig1<-0.5  # Desviacion Estandar del Error de la regresion.
```

```
#####
##### CALIBRACION Y ESTIMACION g #####
#####
for (HH1 in 1:1) {
  print(HH1)
  #####
```

```
##### Generacion de numeros aleatorios #####
##### de distribucion N(0,1) truncada por #####
#####  $1_{\{-a,a\}}(x)$  dependientes usando #####
##### datos AR(1) con  $\phi=0.75$  y el #####
##### metodo de la transformada inversa. #####
#####
#####
Z=arima.sim(list(ar=0.75),n=n) # Generacion de AR(1) con
#  $\phi=0.75$  y  $\sigma=1$ 
U=pnorm(Z, mean=0, sd=sqrt(1/(1-0.75^2)))
#  $U=F(Z)$ , donde  $Z\sim N(0,1/(1-0.75^2))$ 
# y  $F$  es su funcion de distribucion.
#  $U\sim U(0,1)$  y es dependiente.
a=2; # Soporte para la normal truncada, dependiente.
p=pnorm(a)-pnorm(-a); # Area bajo la normal (0,1) truncada
# en  $a$ , tal densidad se denota por  $g$ 
# y ademas  $g=G'$ .
X=qnorm(p*U+pnorm(-a), mean=0, sd=1) #  $X=G^{-1}(U)=(G^{-1}\circ F)(Z)$ .
# Normal truncada dependiente.

plot(seq(1,n,1),X, xlab="_", ylab="_")
# Diagrama de dispersion de los
# datos generados  $\{X_i\}_{i=1}^n$ 
acf(X, lag=100) # Grafico de la funcion de correlacion
# de los primeros 100 datos  $X_i$ 

T1=X

### Funcion de regresion a estimar  $r(x)=0.7x + 2\exp(-10x^2)$ .
Y0t=0.7*T1+2*exp(-10*T1^2)

Yt=Y0t+rnorm(n,0, sig1) # Generacion de la variable
# respuesta del modelo
# de regresion

ne<-n-200 # Numero de datos para estimar
```

```

X=T1[1:(n-200)] # Datos de la variable explicativa
                # para estimar

Y0=Y0t[1:(n-200)] # Y0=r(X), donde X es normal
                  # truncada dependiente.

Y=Yt[1:(n-200)] # Modelo Y=r(X)+e, donde e~N(0,1)
                 # o datos de la variable respuesta
                 # para estimar

#####
#####
##### Estimacion desviacion estandar e_j #####
#####
#####
library(KernSmooth)
h<-dpill(X,Y) # Metodo plug-in de Ruppert,
              # Sheather y Wand (1995)
# Esta ventana obtenida por el metodo del pulgar,
# la sustituir en el estimador de N-W para asi
# tener una pre-estimacion de r en cada punto de
# X en R^2000. Lo que permitira estimar e_j
#####

#####
##### MODULO 1 m #####
##### Estimador de m #####
#####
Estimmm<-function(x,X,Y,h){
  # Estimador de m con kernel gausiano, ventana h para la
  # muestra aleatoria X envaluada en x.
  #  $m(x) = \sum_{i=1}^n (1/(nh * \sqrt{2 * \pi})) Y[i]$ 
  #  $\exp(-((x-X[i])/h)^2 / 2)$ .
  n<-length(X);
  mm=0;
  for(i in 1:n){

```

```

        mm=mm+(1/(n*h*sqrt(2*pi)))*Y[i]*exp(-(((x-X[i])/h)^2)/2);
    }
    return(mm)
}
#####

#####
#####          MODULO 1      g          #####
#####          Estimador de g      #####
#####
Estimg<-function(x,X,h){
  # Estimador de g con kernel gaussiano, ventana h para la
  # muestra aleatoria X envaluada en x.
  #  $g(x) = \sum_{i=1}^n (1/(nhsqrt(2*pi)))$ 
  #  $\exp(-(((x-X[i])/h)^2)/2)$ .
  n<-length(X);
  g=0;
  for(i in 1:n){
    g=g+(1/(n*h*sqrt(2*pi)))*exp(-(((x-X[i])/h)^2)/2);
  }
  return(g)
}
#####

#####
#####          Estimador mpul, gpul y rpul #####
#####
mpul<-0;
gpul<-0;
for(i in 1:ne){
  mpul[i]<-Estimmm(X[i],X,Y,h);
  gpul[i]<-Estimg(X[i],X,h);
}
rpul<-mpul/gpul
#####

```

```

sigqchu<-(1/(ne-1))*sum((Y-rpul)^2) # Estimacion Var(e_j)
sigchu<-sqrt(sigqchu) # Estimacion desviacion estandar
# de e_j
sig<-sigchu # Desviacion estandar de e_j
#####
#####
##### Fin Estimacion desviacion estandar e_j #####
#####
#####

t1=T1[((n-100)+1):n] # Ultimos 100 datos de la variable
# explicativa para calibrar.
t2=Yt[((n-100)+1):n] # Ultimos 100 datos de la variable
# respuesta para calibrar,

ii=order(t1) # Comando para ordenar

x=t1[ii] # Ultimos 100 datos de la variable explicativa
# para calibrar, "ORDENADOS"
Yct<-t2[ii] # Ultimos 100 datos de la variable respuesta
# para calibrar, "ORDENADOS"

#####
### Formula de la densidad g, normal(0,1) truncada #####
#####
mu<-0;
sigm<-1;
#Extremo del soporte es a=2.
dTnorm<-function(x,mu,sigm,a){
  #Densidad N(mu,sigm) truncada de -a hasta a.
  p<-pnorm(mu+a,mu,sigm,lower.tail=TRUE,log.p=FALSE)
  -pnorm(mu-a,mu,sigm,lower.tail=TRUE,log.p=FALSE);
  if (mu-a<=x && x<=mu+a){g<-dnorm(x,mu,sigm,log=FALSE)/p;}
  else {g=0;}
}

```

```

    return (g)
}
    ***ACA PARA QUE CORRA***#
#####

#####
##### MODULO 2 g #####
##### Estimador de g sobre suavizado #####
#####
EstimgSS<-function(x,X,h,h1){
  # Estimador sobre suavizado con nucleo gaussiano , muestra
  # aleatoria simple X, ventanas h y h1.
  n<-length(X);
  gss=0
  for(i in 1:n){
    gss=gss+(1/(n*sqrt(2*pi*(h^2+h1^2))))
    *exp(-((x-X[i])^2)/(2*(h^2+h1^2)));
  }
  return(gss)
}
    ***ACA PARA QUE CORRA***#
#####

#####
##### MODULO 3 g #####
#####
# Esta funcion determina el valor absoluto de la diferencia
# entre el estimador de la regresion sobresuavizado rhh1 y rh1
# en un valor x del soporte.
VAdifEstg<-function(x,X,h,h1){
  VA<-abs(EstimgSS(x,X,h,h1)-Estimg(x,X,h1))
  return(VA)
}
#####
#####

```

```

histg<-hist(X, freq=FALSE) # Calculo y grafico del histograma
                             # usado para calibrar la
                             # estimacion de la densidad
s<-length(x) # Numero de elementos en la particion del
              # soporte de la densidad de X
go<-0       # Variable que almacena la funcion de
            # densidad original
for (i in 1:s) {
  go[i]<-dTnorm(x[i], mu, sigm, a)
}
lines(x, go, lwd=2, col='black') # Al histograma se le agrega la
                                  # curva de la densidad original

eh<-histg[[1]];
EstHist<-histg[[3]];
nh<-length(EstHist)
gsup<-max(EstHist) # Se determina el valor maximo del
                   # histohrama

#####
##### MODULO 4 Histograma #####
##### Funcion Histograma #####
#####
Histograma<-function(x, nh, EstHist, eh){
  for (i in 1:(nh-1)) {
    if(eh[i]<=x && x<eh[i+1]){Hg<-EstHist[i]}
  }
  if(eh[nh]<=x && x<=eh[nh+1]){Hg<-EstHist[nh]}
  return(Hg)
}
#####
#####

ne<-length(X); # Numero de datos en la muestra de
               # estimacion (Para construir el Estimador)
#a1<-1 # Limite superior del soporte de la densidad de X

```

```

#x<-seq(-a1,a1,0.1) # Soporte de la densidad de X

M<-floor(log(ne))*(2/3) # Definicion de M, donde M es la
                        # parte entera n/log(n),
H<-exp(-seq(0,M,0.1)) # Familia de ventanas, donde
                       # H={h_i}_0^M, con M definido en
                       # Linea anterior.

m<-length(H) # Numero de ventanas.

#####
#####
#### Rutina para almacenar los datos #####
#### usando la funcion VAdifEst #####
#### en una lista de nombre l que #####
#### contiene s matrices donde cada #####
#### matriz es de orden (mxm) #####
#### s=length(x) y m=length(H). #####
#### El almacenamiento en cada matriz #####
#### se realiza por filas. #####
#####
l<-list() # Nombre de la lista que almacenara s matrices
          # donde cada una es de orden mxm.

for(k in 1:s){
  l0<-0 # Variable auxiliar que almacena temporalmente
        # cada matriz de datos.
  l100<-0 # Variable auxiliar que almacena temporalmente
          # cada fila de la matriz l0.
  for(j in 1:m){
    l100<-VAdifEstg(x[k],X,H,H[j])
    if(j==1) {l10<-100}
    else {l10<-cbind(l10,l100)}
  }
  l[[k]]<-l10

```

```

}
#####
#####

#####
#### Cambio Local de rsup y ginf IMPORTANTE ####
#### Ahora rsup y ginf son vectores de dimension s ####
#### Cada rsup[i] y ginf[i] corresponde con x[i], ####
#### para i=1:s ####
#####
d<-rep(0,s) #d distancias para el calculo de Int(rchu-Y)^2
d[1]<-((x[1]+x[2])/2)-(-2)
d[s]<-2-((x[s-1]+x[s])/2)
for(i in 2:(s-1)){
  d[i]<-((x[i]+x[i+1])/2)-((x[i-1]+x[i])/2)
}

delg<-((log(ne))^(1/2));

N2<-1/sqrt(2*sqrt(pi)); #Norma 2 del nucleo gaussiano.
N1<-1; #Norma 1 del nucleo gasussiano

GAM1<-0.00000005 # Extremo inferior de la malla para calibrar
GAM21<-0.08 # Extremo superior de la malla para calibrar
dd<-(GAM21-GAM1)/20 # Numero de elementos de la malla
GAM<-seq(GAM1,GAM21,dd) # Malla para calibrar el estimador
# de la funcion de densidad g

Norma2gchuHHg<-rep(0,length(GAM)); # Variable que almacena
# la funcion Error(GAM)
GCHU<-matrix(0,nrow=s,ncol=length(GAM)) # Matriz que almacena
# las estimaciones gchu para cada
# valor GAM[i]

#####

```

```
#####
#####          INICIO CALIBRACION          #####
#####
#####
for (q in 1:length(GAM)){
  print(q)
  gam<-GAM[q] #Parametro gamma > 2. (valor que he cambiado)

  #####
  ###   Calculo del estimador V(x,h), para cada x en   ###
  ###   la malla de longitud s y h \in H.           ###
  ###   Es decir V es una matriz de orden sxm       ###
  #####
  ###   IMPORTANTE aca se hacen cambios locales,     ###
  ###   Ahora V tambien depende de x, ademas de     ###
  ###   depender de H(como antes)                   ###
  ###   V sera una matriz de oden sxm, donde cada   ###
  ###   fila corresponde a un valor de x y cada     ###
  ###   columna a los valores de H.                 ###
  ###   V[i,j] donde i=1:s y j=1:m                 ###
  #####
  V<-matrix(0 ,nrow=s , ncol=m)
  for(i in 1:s){
    for(k in 1:m){
      V[i ,k]<-sqrt(2 *gsup*gam)*N2*(1+N1)*(1+delg)
      *((log(ne))/(ne*H[k]))^(1/2)
    }
  }
  #####          ***ACA PARA QUE CORRA***          #####
  ###           Fin del cambio local de V.           ###
  #####

  #####
  #####
  ##### Calculo del Maximo de cada fila #####
  ##### de la matriz {l[k]-V}_{+}, #####
```

```

##### para cada k de 1:s. #####
##### Lo cual se guarda en la lista #####
#####  $A=\{A[k]\}_{k=1;s}$  con  $A[k]$  en  $R^m$ . #####
##### Se genera la lista AV, compuesta #####
##### de s vectores m dimensionales #####
#####  $AV[[k]]=A[[k]]+V$  en  $R^m$  para #####
#####  $k=1, \dots, s$ . #####
##### Se genera el vector hopt en  $R^s$  #####
##### donde  $hopt[k]=\operatorname{argmin}(AV[[k]]+V)$  #####
#####
#####
A<-list() # Es una lista de s vectores, donde cada
# vector A[k] en  $R^m$ , corresponde a un valor
# x[k] en de la malla, y A[k][i] es el
# valor en H[i]
AV<-list() # Es una lista de s, vectores, donde cada
# vector AV[k] coresponde a la suma
# de vectores A[k]+V[k,].
arg<-0 # Esta Variable sera un vector donde cada
# componente es  $\operatorname{arg}[k]=\operatorname{argmin}_{\{h \in H\}}(A[k]+V[k,])$ 
hoptg<-0 # Variable que almacena las ventanas optimas
# para cada x[k] en la malla.

### IMPORTANTE, En la parte que sigue se hicieron ###
### CAMBIOS LOCALES, correspondiente a la variable V. ###
for(k in 1:s){
  A0<-0 # Esta variable auxiliar, sera una matriz que
# almacenara temporalmente los datos contenidos
# en cada matriz A[[k]] de la lista A
  for(i in 1:m){
    A0[i]<-max(l[[k]][i,]-V[k,])
    if(A0[i]<0) {A0[i]<-0}
  }
  A[[k]]<-A0 # Se almacenan los datos en la lista A
  AV[[k]]<-A0+V[k,] # Se almacenan los datos en
# la lista AV

```

```

    arg[k]<-which.min(AV[[k]]) # Se almacena la posicion
                              # de la ventana optima
                              # para cada x[k] de la
                              # malla de x.
}
# FIN CAMBIOS LOCALES, correspondientes a la variable V. #
# FIN CAMBIOS LOCALES, correspondientes a la variable V. #

hoptg<-H[arg] # Vector de ventanas optimas ,
              # es un vector en R^s

#####

#####
##### Estimador gchu #####
#####
gchu<-0;
HHg<-0;
TNorma2<-0
for(i in 1:s){
  gchu[i]<-Estimg(x[i],X,hoptg[i]);
  HHg[i]<-Histograma(x[i],nh,EstHist,eh);
  TNorma2[i]<-d[i]*(gchu[i]-HHg[i])^2;
}

GCHU[,q]<-gchu; # Matriz de estimadores de gchu para
                # cada gam cada columna tiene un
                # estimado de gchu

#####

##### Grafico del estimador gchu para cada #####
##### GAM[q], con q=1:\length(GAM) #####
#####
plot(x,gchu,ylim=c(0,0.6))
lines(x,gchu,lwd=2,col="red")

```

```

lines (x , go , lwd=2 , col=' black ' )
lines (x , HHg , lwd=2 , col=' blue ' )
#####

Norma2gchuHHg [ q ] <-sum ( TNorma2)

}

#####
##### Grafico de la funcion Error (\gamma) #####
##### Que determina el criterio para #####
##### calibrar la funcion de densidad g #####
#####

plot (Norma2gchuHHg , main=" Calibracion _Metodo _GL _para _g"
, xlab=" _" , ylab=" _" )
lines (Norma2gchuHHg)
#####
##### ***ACA PARA QUE CORRA*** #####

i0g<-which . min (Norma2gchuHHg) # Posicion del valor de gamma
# que calibra al estimador
# gchu de g.

#####
#####
##### Final Calibracion #####
##### Del estimador de g #####
#####
#####

#####
#####
##### INICIO AJUSTE #####

```

```
##### CALIBRADO DE g #####
#####
#####
#####
a<-2
x1<-seq(-a,a,0.1); # Soporte para la estimacion.
s1<-length(x1) # Numero de elementos en la particion
# del soporte de la densidad de X

#####
#####
#### Rutina para almacenar los datos #####
#### usando la funcion VAdifEst #####
#### en una lista de nombre l que #####
#### contiene s matrices donde cada #####
#### matriz es de orden (mxm) #####
#### s=length(x) y m=length(H). #####
#### El almacenamiento en cada matriz #####
#### se realiza por filas. #####
#####
l<-0 # Nombre de la lista que almacenara s matrices
# donde cada una es de orden mxm.
for(k in 1:s1){
  l0<-0 # Variable auxiliar que almacenara temporalmente
# cada matriz de datos.
  l100<-0 # Variable auxiliar que almacena temporalmente
# cada fila de la matriz l0.
  for(j in 1:m){
    l100<-VAdifEstg(x1[k],X,H,H[j])
    if(j==1) {l100<-100}
    else {l100<-cbind(l100,l100)}
  }
  if(k==1) {l<-list(l100)}
  else {l[[k]]<-l100}
}
#####
```

```
#####

gam<-GAM[i0g] #Parametro gamma > 2. (valor que he cambiado)
Q<-i0g      # Posicion en la malla GAM, donde se calibro gchu,
            # es decir la columna

# Q de GCHU (GCHU[,Q]) tiene la estimacion calibrada en una
# malla de longitud 100.

#####
####  Calculo del estimador V(h), para cada h \in H.  #####
####  IMPORTANTE aca se hacen cambios locales,      #####
####  Ahora V tambien depende de x, ademas de      #####
####  depender de H(como antes)                     #####
####  V sera una matriz de oden sxm, donde cada     #####
####  fila corresponde a un valor de x y cada       #####
####  columna a los valores de H.                   #####
####  V[i,j] donde i=1:s y j=1:m                    #####
#####
V<-matrix(0,nrow=s1,ncol=m)
for(i in 1:s1){
  for(k in 1:m){
    V[i,k]<-sqrt(2*gsup*gam)*N2*(1+N1)*(1+delg)*((log(ne))/(ne*H[k]))
  }
}
#####
##### Fin del cambio local de V. #####
#####

#####
##### Calculo del Maximo de cada fila #####
##### de la matriz {l[k]-V}_{+}, #####
##### para cada k de 1:s. #####
##### Lo cual se guarda en la lista #####
##### A={A[k]}_{k=1;s} con A[k] en R^m. #####
```

```
##### Se genera la lista AV, compuesta #####
##### de s vectores m dimensionales #####
##### AV[[k]]=A[[k]]+V en R^m para #####
##### k=1,...,s. #####
##### Se genera el vector hopt en R^s #####
##### donde hopt[k]=argmin (AV[[k]]+V) #####
#####
A<-list() # Es una lista de s vectores, donde cada
          # vector A[k] en R^m, corresponde a un valor
          # x[k] de la malla, y A[k][i] es el valor en H[i]
AV<-list() # Es una lista de s, vectores, donde cada
           # vector AV[k] corresponde a la suma de
           # vectores A[k]+V.
arg<-0 # Esta Variable sera un vector donde cada
       # componente es arg[k]=argmin_{h in H} (A[k]+V)
hoptg<-0 # Variable que almacena las ventanas optimas
         # para cada x[k] en la malla.

for(k in 1:s1){
  A0<-0 # Esta variable auxiliar, sera una matriz que
        # almacenara temporalmente los datos contenidos
        # en cada matriz A[[k]] de la lista A
  for(i in 1:m){
    A0[i]<-max(1[[k]][i,]-V[k,])
    if(A0[i]<0) {A0[i]<-0}
  }
  # Se genera la lista A y AV
  A[[k]]<-A0; # Se guardan resultados en la lista A
  AV[[k]]<-A0+V[k,] # Se guardan resultados en la lista AV
  arg[k]<-which.min(AV[[k]]) # Se almacena la ventana optima
                            # para cada x[k] de la malla de x.
}
hoptg<-H[arg]

#####
```

```

#####
##### Estimador gchu #####
#####
#####
gchu<-0;
for(i in 1:s1){
  gchu[i]<-Estimg(x1[i],X,hoptg[i]); #Coloque hopt^{-1}
}
#####

#####
##### FINALIZO EL AJUSTE #####
##### DE LA FUNCION DE DENSIDAD g #####
#####
#####
}
#####
##### GRAFICOS FINALES #####
##### PARA LA FUNCION DE DENSIDAD g #####
#####
plot(x1,gchu,ylim=c(0,0.6),type="n"
,main="Estimacion_Adaptativa_Metodo_GL_para_g",xlab="",ylab="")
lines(x1,gchu,lwd=2,lty="dotted")

go<-0
HHg<-0;
for(i in 1:s1){
  go[i]<-dTnorm(x1[i],mu,sigm,a)
  HHg[i]<-Histograma(x1[i],nh,EstHist,eh)
}
lines(x1,go,lwd=2,col='black')
#lines(x1,HHg,lwd=2,col='blue')

***ACA PARA QUE CORRA***

```

```
#####
##### FIN GRAFICOS FINALES #####
##### DE LA FUNCION DE DENSIDAD g #####
#####
```

```
#####
##### CALIBRACION y ESTIMACION m #####
#####
```

```
for (HH2 in 1:1) {
  print (HH2)
  #####
  ##### MODULO 2 mss #####
  #####
  EstimmmSS<-function (x ,X, Y, h , h1){
    # Estimador sobre suavizado con nucleo gaussiano , muestra
    # aleatoria simple X, ventanas h y h1.
    n<-length (X);
    mmss=0
    for (i in 1:n){
      mmss=mmss+(1 / (n*sqrt (2*pi*(h^2+h1 ^2))))*Y[ i ]
      *exp ( -((x-X[ i ])^2) / (2*(h^2+h1 ^2)));
    }
    return (mmss)
  }
  #**ACA PARA QUE CORRA**#
  #####
  #####
```

```
#####
##### MODULO 3 m #####
#####
# Esta funcion determina el valor absoluto de la diferencia
# entre el estimador de la regresion sobresuavizado rhh1 y rh1
# en un valor x del soporte.
VAdifEstmm<-function (x ,X, Y, h , h1){
  VA<-abs (EstimmmSS (x ,X, Y, h , h1 )-Estimmm (x ,X, Y, h1 ))
```

```

return (VA)
}
#####
#####

#####
#####
#### Rutina para almacenar los datos #####
#### usando la funcion VAdifEst #####
#### en una lista de nombre l que #####
#### contiene s matrices donde cada #####
#### matriz es de orden (mxm) #####
#### s=length(x) y m=length(H). #####
#### El almacenamiento en cada matriz #####
#### se realiza por filas. #####
#####
l<-list() # Nombre de la lista que almacenara s matrices
          # donde cada una es de orden mxm.
for(k in 1:s){
  10<-0 # Variable auxiliar que almacenara temporalmente
        # cada matriz de datos.
  100<-0 # Variable auxiliar que almacena temporalmente
         # cada fila de la matriz 10.
  for(j in 1:m){
    100<-VAdifEstmm(x[k],X,Y,H,H[j])
    if(j==1) {10<-100}
    else {10<-cbind(10,100)}
  }
  l[[k]]<-10
}
#####
#####

#####
## Cambio Local de rsup y ginf IMPORTANTE ##

```

```

##      Ahora rsup y ginf son vectores de dimension s      ##
##Cada rsup[i] y ginf[i] corresponde con x[i], para i=1:s##
#####
e<-rep(0 , s+1)
se<-length(e)
e[1]<-(-2)
e[se]<-2

for (i in 2:(se-1)) {
  e[i]<-(x[i-1]+x[i])/2
}

rsup<-rep(0 , s)
gchusup<-max(gchu)

for(i in 1:s){
  rsup[i]<-max(abs(Y*(e[i]<X)*(X<e[i+1]))) # Maximo Local de la
                                           # funcion de regresion
}
#####
##### FIN IMPORTANTE, CAMBIO LOCAL #####
##### FIN IMPORTANTE, CAMBIO LOCAL #####
#####

delm<- (log(ne))^( -1/5)
N2<-1/sqrt(2*sqrt(pi)) #Norma 2 del nucleo gaussiano.

#####
####      Cambio Local de A3      IMPORTANTE      ####
####      Ahora A3 es un vector de dimension s      ####
####      Cada A3[i] corresponde con x[i], para cada i=1:s      ####
#####
A3<-rep(0 , s)
for(i in 1:s){
  A3[i]<-(rsup[i]^2+ sig ^2)*gchusup*N2^2
}

```

```

#####
##### FIN IMPORTANTE, CAMBIO LOCAL #####
##### FIN IMPORTANTE, CAMBIO LOCAL #####
#####

Norma2mchuYHHg<-rep(0 , length(GAM)); # Variable que almacenara
# los valores de la
# funcion Error_m(gamma)

#####
##### CALIBRACION m #####
#####
for(q in 1:length(GAM)){

  gam<-GAM[q] #Parametro gamma > 2. (valor que he cambiado)

#####
##### Calculo del estimador V(h), para cada h \in H. ###
##### IMPORTANTE aca se hacen cambios locales, ###
##### Ahora V tambien depende de x, ademas de ###
##### depender de H(como antes) ###
##### V sera una matriz de oden sxm, donde ###
##### cada fila corresponde ###
##### a un valor de x y cada columna a los valores de H.
##### V[i,j] donde i=1:s y j=1:m ###
#####
V<-matrix(0 , nrow=s , ncol=m)
for (k in 1:m){
  V[,k]<-sqrt(2*gam*A3)*(N1+1)*(1+delm)
  *sqrt(log(ne))/sqrt(ne*H[k])
}
#####
##### Fin del cambio local de V. #####
#####

```

```
#####
##### Calculo del Maximo de cada fila #####
##### de la matriz  $\{l[k]-V\}_+$ , #####
##### para cada k de 1:s. #####
##### Lo cual se guarda en la lista #####
##### A= $\{A[k]\}_{k=1;s}$  con A[k] en  $R^m$  #####
##### Se genera la lista AV, compuesta #####
##### de s vectores m dimencionales #####
##### AV[[k]]=A[[k]]+V en  $R^m$  para #####
##### k = 1, ..., s. #####
##### Se geneta el vector hopt en  $R^s$  #####
##### donde hopt[k]=argmin (AV[[k]]+V) #####
#####
A<-0 # Es una lista de s vectores , donde cada
# vector A[k] en  $R^m$ , corresponde a un valor  $x[k]$ 
# de la malla , y A[k][i] es el valor en H[i]
AV<-0 # Es una lista de s , vectores , donde cada vector
# AV[k] coresponde a la suma de vectores A[k]+V.
arg<-0 # Esta Variable sera un vector donde cada
# componente es  $arg[k]=argmin_{h \in H} (A[k]+V)$ 
hoptm<-0 # Variable que almacena las ventanas optimas
# para cada  $x[k]$  en la malla.

#####
### IMPORTANTE, En la parte que sigue se hicieron ###
### CAMBIOS LOCALES, correspondiente a la variable V. ###
#####
for(k in 1:s){
  A0<-0 # Esta variable auxiliar , sera una matriz que
# almacenara temporalmente los datos contenidos
# en cada matriz A[[k]] de la lista A
  for(i in 1:m){
    A0[i]<-max(l[[k]][i,]-V[k,])
    if(A0[i]<0) {A0[i]<-0}
  }
}
```

```

if (k==1) {A<-list (A0); AV<-list (A0+V[k,])} # Se genera la
# lista A y AV
else {A[[k]]<-A0; AV[[k]]<-A0+V[k,]} # Se generan las
# listas A y AV
arg[k]<-which.min(AV[[k]]) # Se almacena la ventana
# optima para cada x[k] de
# la malla de x.
}
#####
# FIN CAMBIOS LOCALES, correspondientes a la variable V. #
# FIN CAMBIOS LOCALES, correspondientes a la variable V. #
#####
hoptm<-H[ arg ]

#####
##### Estimador mchu #####
#####
mchu<-0;
TNorma2mchuYHHg<-0; # Variable auxiliar que almacenara
# temporalmente los terminos del
# MSE de la estimacion.
for (i in 1:s){
  mchu[i]<-Estimm(x[i],X,Y,hoptm[i]);
  TNorma2mchuYHHg[i]<-d[i]*(mchu[i]-Yct[i]*GCHU[i,Q])^2;
}
#####

#####
##### Grafico del estimador mchu para cada GAM[q], #####
##### con q=1:length(GAM) #####
#####

plot (x, mchu, ylim=c(-0.5,1))
lines (x, mchu, lwd=2, col="red")
points (x, Yct*GCHU[,Q], col="yellow")
#####

```

```

Norma2mchuYHHg [ q ] <-sum ( TNorma2mchuYHHg );

}
#####
##### Grafico de la funcion Error_m(gamma) que permite #####
##### saber donde se calibra el estimador mchu de la #####
##### funcion m. #####
#####
plot ( Norma2mchuYHHg , main=" Calibracion _Metodo _GL _para _m"
, xlab=" _" , ylab=" _" )
lines ( Norma2mchuYHHg )
#####
***ACA PARA QUE CORRA***
#####

i0m<-which . min ( Norma2mchuYHHg ) # En esta posicion de la malla
# GAM se calibra mchu

#####
#####
##### Final Calibracion m #####
#####
#####

#####
#####
##### INICIO AJUSTE m #####
#####
#####

#####
#####
##### Rutina para almacenar los datos #####

```

```

##### usando la funcion VAdifEst #####
##### en una lista de nombre l que #####
##### contiene s matrices donde cada #####
##### matriz es de orden (mxm) #####
##### s=length(x) y m=length(H). #####
##### El almacenamiento en cada matriz #####
##### se realiza por filas. #####
#####

l<-0 # Nombre de la lista que almacenara s matrices
# donde cada una es de orden mxm.

for(k in 1:s1){
  l0<-0 # Variable auxiliar que almacenara temporalmente
# cada matriz de datos.
  l100<-0 # Variable auxiliar que almacena temporalmente
# cada fila de la matriz l0.
  for(j in 1:m){
    l100<-VAdifEstmm(x1[k],X,Y,H,H[j])
    if(j==1) {l100<-100}
    else {l100<-cbind(l0,l100)}
  }
  if(k==1) {l<-list(l100)}
  else {l[[k]]<-l100}
}
#####
#####

gam<-GAM[i0m] #Parametro gamma > 2. (valor que he cambiado)

#####
##### Calculo del estimador V(h), para cada h \in H. #####
##### IMPORTANTE aca se hacen cambios locales, #####
##### Ahora V tambien depende de x, ademas de #####
##### depender de H(como antes) #####
##### V sera una matriz de oden sxm, donde #####

```

```

##### cada fila corresponde a un valor de x #####
##### y cada columna a los valores de H. #####
##### V[i, j] donde i=1:s y j=1:m #####
#####
e<-rep(0, s1+1)
se<-length(e)
e[1]<-(-2)
e[se]<-2

for (i in 2:(se-1)) {
  e[i]<-(x1[i-1]+x1[i])/2
}

rsup<-rep(0, s1)
gchusup<-max(gchu)

for(i in 1:s1){
  rsup[i]<-max(abs(Y*(e[i]<X)*(X<e[i+1]))) # Maximo Local de la
                                           # funcion de regresion
}

A3<-rep(0, s1)
for(i in 1:s1){
  A3[i]<-(rsup[i]^2+sig^2)*gchusup*N2^2
}

V<-matrix(0, nrow=s1, ncol=m)
for (k in 1:m){
  V[,k]<-sqrt(2*gam*A3)*(N1+1)*(1+delm)*sqrt(log(ne))/sqrt(ne*H[k])
}
#####
##### Fin del cambio local de V. #####
#####
#####
#####

```

```

##### Calculo del Maximo de cada fila #####
##### de la matriz  $\{l[k]-V\}_{+}$ , #####
##### para cada  $k$  de  $1:s$ . #####
##### Lo cual se guarda en la lista #####
#####  $A=\{A[k]\}_{k=1;s}$  con  $A[k]$  en  $R^m$  #####
##### Se genera la lista AV, compuesta #####
##### de  $s$  vectores  $m$  dimensionales #####
#####  $AV[[k]]=A[[k]]+V$  en  $R^m$  para #####
#####  $k=1, \dots, s$ . #####
##### Se genera el vector hopt en  $R^s$  #####
##### donde  $hopt[k]=\operatorname{argmin}(AV[[k]]+V)$  #####
##### #####
A<-list() # Es una lista de  $s$  vectores, donde cada
          # vector  $A[k]$  en  $R^m$ , corresponde a un valor  $x[k]$ 
          # de la malla, y  $A[k][i]$  es el valor en  $H[i]$ 
AV<-list() # Es una lista de  $s$  vectores, donde cada vector
           #  $AV[k]$  corresponde a la suma de vectores  $A[k]+V$ .
arg<-0 # Esta Variable sera un vector donde cada
       # componente es  $\operatorname{arg}[k]=\operatorname{argmin}_{h \in H}(A[k]+V)$ 
hoptm<-0 # Variable que almacena las ventanas optimas
         # para cada  $x[k]$  en la malla.

for(k in 1:s){
  A0<-0 # Esta variable auxiliar, sera una matriz que
        # almacenara temporalmente los datos contenidos
        # en cada matriz  $A[[k]]$  de la lista A
  for(i in 1:m){
    A0[i]<-max(A[[k]][i,]-V[k,])
    if(A0[i]<0) {A0[i]<-0}
  }
  # Se genera la lista A y AV
  A[[k]]<-A0; # Se guardan resultados en la lista A
  AV[[k]]<-A0+V[k,] # Se guardan resultados en la lista AV
  arg[k]<-which.min(AV[[k]]) # Se almacena la ventana optima
                            # para cada  $x[k]$  de la malla de x
}

```

```

hoptm<-H[ arg ]
#####

#####
##### Estimador mchu #####
#####
mchu<-0;
for(i in 1:s1){
  mchu[i]<-Estimmm(x1[i],X,Y,hoptm[i]);
}
#####

#####
##### FINALIZO EL AJUSTE m #####
#####
}

#####
##### Grafico de mchu estimador de m #####
##### Junto a la nube de puntos de entrenamiento #####
##### y la funcion m, en el intervalo [-2,2] #####
#####
plot(x1,mchu,ylim=c(-0.5,1),type="n"
,main="Estimacion_Adaptativa_Metodo_GL_para_m",xlab="",ylab="")
lines(x1,mchu,lwd=2,lty="dotted") # Curva del estimador mchu
points(x,Yct*GCHU[,Q],col="yellow") # Nube de puntos de
# entrenamiento

r=0.7*x1+2*exp(-10*x1^2) # Funcion de regresion a estimar
# r(x)=0.7x +2exp(-10x^2)
# evaluada en la malla x1
# del intervalo [-2,2]

lines(x1,r*go,lwd=2,col='black') # Grafico de m=r*go evaluado
    
```

```

# en la malla x1 del
# intervalo [-2,2]
***ACA PARA QUE CORRA***
#####

#####
### GRAFICO de rchu=mchu/gchu en el intervalo [-2,2] ###
#####

rchu<-mchu/gchu; # Estimador rchu de la funcion de regrion r

plot(x1, rchu, ylim=c(-2.5, 3), type="n"
, main="Estimacin_Adaptativa_Metodo_GL_para_r", xlab="", ylab="")
points(X, Y, col="yellow") # Nube de puntos de estimacion
lines(x1, rchu, lwd=2, lty="dotted") # Curva del estimador mchu
# de la regresion r

lines(x1, r, lwd=2, col='black') # Curva de la funcion de
# de regresion r
***ACA PARA QUE CORRA***
#####

#####
##### Grafico de rchu en el intervalo [-1,1] #####
#####
j2<-seq(11, 31, 1) # Indices para que la malla del soporte
# este en [-1,1]

plot(x1[j2], rchu[j2], ylim=c(-2, 3), type="n"
, main="Estimacion_Adaptativa_Metodo_GL_para_r", xlab="", ylab="")
points(X, Y, col="yellow") # Nube de puntos de estimacion
lines(x1[j2], rchu[j2], lwd=2, lty="dotted") # Curva del
# estimador mchu
# en [-1,1]

```

```
r=0.7*x1[j2]+2*exp(-10*x1[j2]^2) # Funcion de regresion a estimar
# r(x)=0.7x +2exp(-10x^2)
# evaluada en la malla x1[j2]
# del intervalo [-1,1]

lines(x1[j2],r,lwd=2,col='black') # Curva de la funcion de
# regresion r, en el
# intervalo [-1,1]

***ACA PARA QUE CORRA***#
#####
```