



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

Tesis de Licenciatura

**Estimación en Modelos Parcialmente Lineales para Datos Direccionales
con Respuesta Faltante**

Estefanía Louzau

Directora: Dra. Daniela Rodriguez

27 de Diciembre de 2013

Índice general

1. Introducción	1
2. Modelos parcialmente lineales	4
2.1. Introducción	4
2.2. Estimación	4
2.2.1. Estimación de la densidad	5
2.2.2. Estimación de la función de regresión	8
2.2.3. Estimación del modelo parcialmente lineal	10
3. Estimación en MPL con respuestas faltantes	12
3.1. Introducción	12
3.2. Estimadores basados en métodos de imputación y propiedades asintóticas	12
3.3. Estimadores semiparamétricos de regresión y propiedades asintóticas	15
3.4. Modelos ponderados por la inversa de la probabilidad marginal	17
4. Datos Direccionales	19
4.1. Introducción	19
4.2. Medidas de Posición	20
4.3. Medidas de concentración y dispersión	20
4.4. Distribuciones Circulares	21
4.4.1. Distribución Uniforme	22
4.4.2. Distribución Von Mises	23
4.5. Estimación de la densidad circular	24
4.6. Estimación de la función de regresión circular-lineal	25
5. Estimación del MPL con componente no paramétrica direccional y respuesta faltante	27

5.1. Estudio de simulación	29
5.2. Resultados	31
5.3. Conclusiones	38
6. Un ejemplo real	40
A. Apéndice	43

Capítulo 1

Introducción

Cuando queremos realizar un análisis estadístico sobre la dependencia entre una variable de respuesta y y variables explicativas $\mathbf{X} = (x_1, \dots, x_p)^T$ podemos utilizar un modelo de regresión lineal múltiple. Es decir, a partir de una muestra aleatoria $(\mathbf{X}_i^T, y_i)'$ $1 \leq i \leq n$ podemos expresar dicha relación por medio de

$$y_i = \mathbf{X}_i^T \boldsymbol{\beta} + \varepsilon_i, \quad i = 1, \dots, n \quad (1.1)$$

siendo $\boldsymbol{\beta}$ un vector de parámetros desconocidos que debemos estimar y ε_i un error aleatorio. Si bien el modelo (1.1) resulta adecuado para muchos casos de interés, en otras ocasiones exigir esta dependencia lineal resulta excesiva. Entonces sería interesante relajar esta suposición. Una forma de hacerlo sería trabajar con un modelo de regresión no paramétrico múltiple, es decir

$$y_i = \gamma(\mathbf{X}_i) + \varepsilon_i, \quad i = 1, \dots, n \quad (1.2)$$

con γ una función desconocida a estimar. Es evidente que este modelo engloba prácticamente todas las posibles relaciones entre X e y , incluso la propuesta lineal. Pero, la regresión no paramétrica en varias variables posee dificultades teóricas y computacionales, conocidas como la maldición de la dimensionalidad. Es por eso que surgen los modelos parcialmente lineales es decir,

$$y_i = \mathbf{X}_i^T \boldsymbol{\beta} + g(T_i) + \varepsilon_i, \quad i = 1, \dots, n \quad (1.3)$$

donde $T_i \in \mathbb{R}$, $\boldsymbol{\beta} \in \mathbb{R}^p$ y $g : \mathbb{R} \rightarrow \mathbb{R}$ son ambos desconocidos y los deseamos estimar en función de los datos. Este modelo combina la flexibilidad de los modelos no paramétricos con la fácil interpretación de los modelos lineales.

La mayoría de los métodos de estimación están diseñados para conjuntos de datos completos y surgen problemas cuando faltan observaciones. El ejemplo típico sucede cuando por algún motivo se carece de la variable de respuesta, por ejemplo abandono del paciente a determinado tratamiento, pero existe situaciones en las que tanto la respuesta como las variables explicativas están perdidas.

Por otro lado, en muchas aplicaciones las variables aleatorias involucradas, en lugar de tomar valores en todo \mathbb{R}^p , toman valores en un espacio de dimensión menor, por ejemplo en

una determinada superficie. Tal es el caso en donde, por la particularidad del problema, las variables se encuentran concentradas en la esfera unidad. Este tipo de datos recibe el nombre de datos direccionales. Los datos direccionales tienen una serie de características que los hacen distintos de los datos lineales y por tanto, el análisis direccional es sustancialmente diferente del análisis estadístico “lineal” estándar ya que la propia naturaleza de los datos obliga a replantear aspectos tan básicos como la medición de distancias o la inclusión de su naturaleza periódica. El análisis de los datos direccionales dio lugar a una serie de nuevos problemas estadísticos que propulsaron, en los últimos 20 años, el desarrollo de métodos específicos. Varios modelos paramétricos y no paramétricos han sido ampliamente discutidos en este contexto. Para una revisión sobre este tema ver Mardia y Jupp [7], Batschelet [2], Watson [15], Fisher [4], Rao [10], He [6], Hall *et al.* [5], Bai *et al.* [1], entre otros.

En esta tesis, estudiaremos estimadores para el modelo parcialmente lineal en el caso en el que la variable a ser modelada no parametricamente toma valores en la esfera. Es decir, la variable de respuesta puede ser modelada linealmente en un conjunto de covariables en \mathbb{R}^p y como una función suave de un dato direccional pero además una parte de las variables de respuesta puede estar perdida. En particular, vamos a centrar nuestra atención en los casos en que los datos perdidos sólo se encuentran en las variables respuestas.

El trabajo se organiza en capítulos. En el Capítulo 2 describimos el procedimiento de estimación del modelo parcialmente lineal, en donde se incluye la estimación de la función densidad y las de las funciones de regresión. En el Capítulo 3 presentamos tres métodos de estimación de los parámetros desconocidos bajo un MPL en presencia de datos faltantes. El Capítulo 4 contiene un resumen sobre conceptos definidos en variables aleatorias que toman valores en el círculo unitario. Además estudiamos la estimación no paramétrica tipo núcleo de la función de regresión circular-lineal. En el Capítulo 5, presentamos los estimadores propuestos bajo un modelo parcialmente lineal para datos direccionales con datos faltantes, analizando su comportamiento a través de un estudio de simulación. Por último en el Capítulo 6 mostramos un ejemplo utilizando un conjunto de datos reales.

Capítulo 2

Modelos parcialmente lineales

2.1. Introducción

Como mencionamos anteriormente, un modelo de regresión parcialmente lineal está definido como

$$y_i = \mathbf{X}_i^T \boldsymbol{\beta} + g(T_i) + \varepsilon_i, \quad i = 1, \dots, n \quad (2.1)$$

donde $\mathbf{X}_i = (x_{i1}, \dots, x_{ip})^T$ son observaciones de un vector aleatorio; $T_i \in \mathbb{R}$ puntos fijos del diseño; $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)$ es un vector de parámetros desconocidos, g es un función desconocida de \mathbb{R} a \mathbb{R} , y $\varepsilon_1, \dots, \varepsilon_n$ son errores aleatorios, independientes con media cero y varianza finita $\sigma_i^2 = E(\varepsilon_i^2)$ e independientes de $(\mathbf{X}_i^T, T_i)^T$. Además, supondremos que existe una relación entre \mathbf{X}_i y T_i . Con lo cuál nuestro modelo puede reescribirse como

$$\begin{cases} y_i &= \mathbf{X}_i^T \boldsymbol{\beta} + g(T_i) + \varepsilon_i & 1 \leq i \leq n, \\ x_{ij} &= \phi_j(T_i) + \eta_{ij} & 1 \leq j \leq p, \end{cases} \quad (2.2)$$

donde η_{ij} son errores aleatorios con $E(\eta_{ij}) = 0$ y por lo tanto $\phi_j(T_i) = E(x_{ij}|T_i)$.

2.2. Estimación

Una vez presentado el modelo, nuestro objetivo será obtener estimadores para los parámetros que desconocemos. En el caso particular que la componente no paramétrica del Modelo Parcialmente Lineal (MPL) se asuma conocida, entonces podemos aplicar la teoría de mínimos cuadrados para obtener un estimador de $\boldsymbol{\beta}$. Sin embargo en la mayoría de los casos prácticos esta hipótesis no es válida. Por lo tanto debemos utilizar otras herramientas de estimación como lo son los métodos de suavizados.

Llamemos $\boldsymbol{\phi}(T_i) = (\phi_1(T_i), \dots, \phi_p(T_i))^T$ y sea $\phi_0(T_i) = E(y_i|T_i)$ entonces tomando esperanza condicional respecto a T en (2.1) tenemos,

$$\phi_0(T_i) = \boldsymbol{\phi}^T(T_i) \boldsymbol{\beta} + g(T_i). \quad (2.3)$$

Con lo cuál, restando $\phi_0(T_i)$ a ambos lados de la primera ecuación de (2.2) podemos reescribir al modelo de la siguiente manera

$$y_i - \phi_0(T_i) = (X_i - \phi(T_i))^T \beta + \varepsilon_i. \quad (2.4)$$

Luego, si no conocemos ϕ_0 y ϕ pero tenemos estimadores $\hat{\phi}_0$ y $\hat{\phi}$ respectivamente, podemos obtener por ejemplo, el estimador de mínimos cuadrados, $\hat{\beta}_{LS}$ para β basado en las *pseudo* observaciones $y_i - \hat{\phi}_0(T_i)$ y $\mathbf{X}_i - \hat{\phi}(T_i)$ del siguiente modo,

$$\hat{\beta}_{LS} = \arg \min_{\beta} \sum_{i=1}^n [(y_i - \hat{\phi}_0(T_i)) - (X_i - \hat{\phi}(T_i))^T \beta]^2. \quad (2.5)$$

Finalmente en base a la ecuación (2.3), podemos estimar $g(T_i)$ como,

$$\hat{g}(T_i) = \hat{\phi}_0(T_i) - \hat{\phi}^T(T_i) \hat{\beta}_{LS}. \quad (2.6)$$

Con lo cuál solo nos quedará obtener estimadores de

$$\begin{aligned} \phi_j(T_i) &= E(x_{ij}|T_i) & 1 \leq j \leq p, \\ \phi_0(T_i) &= E(y_i|T_i). \end{aligned}$$

En primer lugar, notemos que si asumimos que tenemos densidad conjunta de x_{ij} y T_i , entonces $\phi_j(T_i)$ la esperanza condicional de x_{ij} dado T_i , puede calcularse como

$$E(x_{ij}|T_i = t) = \int x_{ij} f(x_{ij}|T_i = t) dx_{ij} = \int x_{ij} \frac{f(x_{ij}, t)}{f_{T_i}(t)} dx_{ij}.$$

Por lo tanto como $f(x_{ij}, t)$ y $f_{T_i}(t)$ son desconocidas, una idea sencilla sería reemplazar estas funciones de densidad por estimadores, es decir hacer una estimación plug-in.

2.2.1. Estimación de la densidad

En esta sección resumiremos el método de estimación de una densidad que fue introducido por Rosenblatt (1959). Consideremos X una variable aleatoria con densidad f continua en x y notemos

$$f(x) = F'(x) = \lim_{h \rightarrow 0} \frac{F(x+h) - F(x-h)}{2h} = \lim_{h \rightarrow 0} \frac{P(x-h < X < x+h)}{2h}.$$

Por otro lado un estimador natural de $P(x-h < X < x+h)$ es simplemente considerar la proporción de la muestra que cae en el intervalo $(x-h, x+h)$. Entonces dado un h suficientemente pequeño podemos deducir el siguiente estimador de $f(x)$,

$$\hat{f}(x) = \frac{1}{2h} \frac{\# \{X_i : X_i \in (x-h, x+h)\}}{n}.$$

Esencialmente, este estimador cuenta la cantidad de observaciones que “caen” en un entorno de radio h alrededor de x . Asimismo si consideramos F_n la distribución empírica podemos escribir a $\hat{f}(x)$ como

$$\hat{f}(x) = \frac{F_n(x+h) - F_n(x-h)}{2h}.$$

Otra forma de expresar el estimador $\hat{f}(x)$ es de la siguiente manera,

$$\hat{f}(x) = \frac{1}{2h} \frac{\#\{X_i : X_i \in (x-h, x+h)\}}{n} = \frac{1}{nh} \sum_{i=1}^n \frac{1}{2} I(|x - X_i| < h).$$

Luego si definimos la función w como $w(x) = \frac{1}{2} I_{(|x| < 1)}$, tenemos que $\hat{f}(x)$ es equivalente a

$$\hat{f}(x) = \sum_{i=1}^n \frac{1}{nh} w\left(\frac{x - X_i}{h}\right). \quad (2.7)$$

Notemos que $w \geq 0$, $\int w(s)ds = 1$, además, para cada $1 \leq i \leq n$ tenemos que $w\left(\frac{x - X_i}{h}\right) = \frac{1}{2}$ si y solo si $X_i \in (x - h, x + h)$, es decir la función w le otorga un peso uniforme a cada observación X_i en el entorno $(x - h, x + h)$ y 0 a cada observación fuera del entorno. A la función w se la denomina núcleo uniforme o de Parzen.

Sin embargo, uno podría estar interesado en darle mayor peso a las observaciones más cercanas a x . Esto se lograría fácilmente reemplazando la función de peso o núcleo w por una función K no negativa que verifique la condición $\int K(x)dx = 1$. Además, si consideramos una función de pesos K con mayor suavidad obtendríamos un estimador más suave. En general los pesos utilizados decrecen de manera suave, dándole así menor pesos a las observaciones más alejadas del punto x . Algunas opciones posibles de núcleos, podrían ser

Núcleo Normal:

$$K(t) = e^{-\frac{1}{2}t^2}.$$

Núcleo Epanechnikov:

$$K(t) = \begin{cases} \frac{3}{4}(1 - t^2) & |t| \leq 1 \\ 0 & |t| > 1 \end{cases}$$

Núcleo Tricúbico:

$$K(t) = \begin{cases} (1 - |t|^3)^3 & |t| \leq 1 \\ 0 & |t| > 1 \end{cases}$$

De esta manera obtenemos el estimador de la densidad basado en núcleos, que constituye uno de los estimadores no paramétricos más estudiados,

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) \quad (2.8)$$

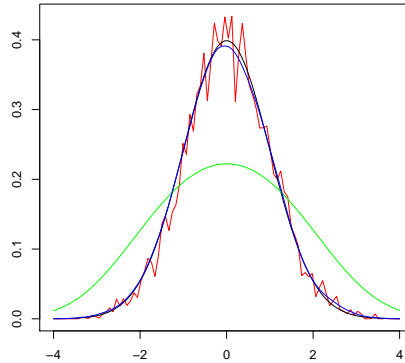


Figura 2.1: Estimación de la densidad de una variable $N(0, 1)$ (negro) por núcleo de Epanechnikov para $h = 0,02$ (rojo), $h = 0,5$ (azul) y $h = 3$ (verde).

donde K es una función núcleo, $h = h_n$ es llamado el *parámetro de suavizado* o *ancho de ventana* y satisface $h_n \rightarrow 0$ si $n \rightarrow \infty$.

Por lo tanto, la estimación de la densidad por núcleos no es otra cosa que un promedio ponderado por la distancia de las observaciones al punto que deseamos estimar. Cuanto mayor sea la distancia de la observación al punto, menor será su peso en la estimación. El peso lo determinarán el núcleo elegido y el valor de h . Ventanas demasiado pequeñas derivarán en estimadores muy variables ya que en cada punto los entornos carecerán de suficientes observaciones en las cuales basar la estimación. Por otra parte, un ventana demasiado grande producirá estimadores muy suaves, que no lograrán captar la estructura local de la densidad dando lugar a estimadores sesgados. Por lo tanto, determinar el valor de h constituye un punto muy importante en el proceso de estimación. (Ver Figura 2.1).

Notemos que si $\int_{-\infty}^{+\infty} K(x)dx = 1$ y $K \geq 0$, entonces el estimador \hat{f} también resulta una función de densidad, pues $\hat{f} \geq 0$ y $\int_{-\infty}^{+\infty} \hat{f}(x)dx = 1$ ya que

$$\int_{-\infty}^{+\infty} \hat{f}(x)dx = \frac{1}{n} \sum_{i=1}^n \int_{-\infty}^{+\infty} \frac{1}{h} K\left(\frac{x - X_i}{h}\right)dx = \frac{1}{n} \sum_{i=1}^n \int_{-\infty}^{+\infty} K(s)ds = 1$$

En lo que sigue, supondremos que K satisface las siguientes condiciones:

- H.1.** $K \geq 0$, $\int K(s)ds = 1$
- H.2.** $\int K(s)sds = 0$ y $\int K^2(s)sds = 0$
- H.3.** $\int |s^i| K^j(s)ds < \infty$, para $i = 1, 2, \frac{5}{2}$ y $j = 1, 2$

Para entender de un modo más técnico la relación entre el tamaño de muestra n y el ancho de banda h , en el Apéndice mostraremos como obtener el error cuadrático medio (ECM) de \hat{f} bajo ciertas hipótesis de regularidad sobre $f(x)$. Más precisamente, podemos ver que

$$|Sesgo(\hat{f}(x))| \leq \frac{h^2}{2} \left| f''(x) \right| \mu_2(K) + o(h^2)$$

si $h \rightarrow 0$ para cada x donde $\mu_2(K) = \int s^2 K(s) ds$ y

$$Var(\hat{f}(x)) \leq \frac{1}{nh} \|K\|^2 f(x) + o\left(\frac{1}{nh}\right).$$

Por lo tanto para valores grandes de ventana el sesgo aumentará. Entonces para obtener menor sesgo habría que considerar ventanas más pequeñas. Además el sesgo depende de $f''(x)$, es decir depende de la curvatura de la función. Por ejemplo, el sesgo será negativo si la derivada segunda es negativa o equivalentemente si la función tiene un máximo local. Por otro lado, si elegimos valores de nh grandes podremos dar un estimador con varianza más pequeña y análogamente si $\|K\|_2^2 = \int K^2$ es pequeña, es decir el núcleo es más bien chato.

Como corolario de los resultados de sesgo y varianza, obtenemos por un lado una cota para el error cuadrático medio del estimador para cada x ,

$$ECM(\hat{f}(x)) \leq h^4 \frac{(f''(x))^2}{4} \mu_2^2(K) + o(h^4) + \frac{1}{nh} \|K\|^2 f(x) + o\left(\frac{1}{nh}\right).$$

Donde nuevamente podemos observar un compromiso entre sesgo y varianza. Pues h pequeños derivarán en estimadores con menor sesgo mientras que al aumentar el ancho de banda lograremos disminuir la varianza. Y finalmente como segundo corolario, se desprende la consistencia débil del estimador, pues si $n \rightarrow \infty$, $h \rightarrow 0$ y $nh \rightarrow \infty$ tenemos que para cada x , $ECM(\hat{f}(x)) \rightarrow 0$ y por lo tanto $\hat{f}(x) \xrightarrow{p} f(x)$ para cada x .

2.2.2. Estimación de la función de regresión

Recordemos que nuestro objetivo es estimar las esperanzas condicionales

$$\begin{aligned} \phi_j(T_i) &= E(x_{ij}|T_i) & 1 \leq j \leq p, \\ \phi_0(T_i) &= E(y_i|T_i), \end{aligned}$$

o también llamadas funciones de regresión. Para simplificar la notación, supongamos que queremos estimar la esperanza condicional de X dado $T=t$, es decir

$$\begin{aligned} E[X|T = t] &= \int x f_{X|T}(x|t) dx \\ &= \int x \frac{f_{XT}(x, t)}{f_T(t)} dx \\ &= \frac{m(t)}{f_T(t)} \end{aligned}$$

siendo $m(t)$

$$m(t) = \int x f_{XT}(x, t) dx$$

La idea que habíamos propuesto anteriormente consistía en reemplazar estas funciones de densidad que desconocemos por estimadores que pueden ser por ejemplo los estudiados en la sección anterior. De esta manera podemos considerar una extensión natural al caso bivariado de la propuesta realizada anteriormente,

$$\widehat{f}_h(x, t) = \frac{1}{n} \sum_{i=1}^n \frac{1}{h^2} K\left(\frac{x - X_i}{h}, \frac{t - T_i}{h}\right)$$

donde K es un núcleo multivariado es decir $K : \mathbb{R}^2 \rightarrow \mathbb{R}$.

En este caso se ha elegido utilizar la misma ventana h , pero no es necesario. Podríamos tomar una ventana distinta en cada componente.

$$\widehat{f}_{h,g}(x, t) = \frac{1}{n} \sum_{i=1}^n \frac{1}{hg} K\left(\frac{x - X_i}{h}, \frac{t - T_i}{g}\right).$$

El núcleo multivariado puede ser elegido como un núcleo multiplicativo. Luego

$$\widehat{f}_{h,g}(x, t) = \frac{1}{n} \sum_{i=1}^n \frac{1}{h} K\left(\frac{x - X_i}{h}\right) \frac{1}{g} K\left(\frac{t - T_i}{g}\right).$$

Calculemos ahora $\widehat{m}_{hg}(t) = \int x \widehat{f}_{h,g}(x, t) dx$,

$$\begin{aligned} \int x \widehat{f}_{h,g}(x, t) dx &= \frac{1}{n} \sum_{i=1}^n \frac{1}{h} K\left(\frac{t - T_i}{g}\right) \int \frac{x}{g} K\left(\frac{x - X_i}{h}\right) dx \\ &= \frac{1}{n} \sum_{i=1}^n \frac{1}{h} K\left(\frac{t - T_i}{g}\right) \int (sg + X_i) K(s) ds \\ &= \frac{1}{n} \sum_{i=1}^n \frac{1}{h} K\left(\frac{t - T_i}{g}\right) X_i. \end{aligned}$$

Por otra parte

$$\widehat{f}_T(t) = \frac{1}{nh} \sum_{k=1}^n K\left(\frac{t - T_i}{h}\right).$$

Entonces tenemos que el estimador de la esperanza condicional queda definido por:

$$\widehat{E}(X|T = t) = \frac{\frac{1}{n} \sum_{i=1}^n \frac{1}{h} K\left(\frac{t - T_i}{h}\right) X_i}{\frac{1}{n} \sum_{i=1}^n \frac{1}{h} K\left(\frac{t - T_i}{h}\right)} = \frac{\sum_{i=1}^n K\left(\frac{t - T_i}{h}\right) X_i}{\sum_{i=1}^n K\left(\frac{t - T_i}{h}\right)}.$$

Luego,

$$\hat{\phi}(t) = \frac{\sum_{i=1}^n K\left(\frac{t-T_i}{h}\right) X_i}{\sum_{i=1}^n K\left(\frac{t-T_i}{h}\right)}.$$

En caso de que el denominador sea 0, el numerador también lo será y el estimador en esos puntos no se encontrará definido.

2.2.3. Estimación del modelo parcialmente lineal

Retomando el procedimiento de estimación descrito en la Sección 2.2 y habiendo ya introducido los estimadores de la función de regresión tenemos ahora si, para el problema que nos interesa que los estimadores de ϕ_0 y ϕ_j quedarán definidos por

$$\begin{aligned}\hat{\phi}_0(t) &= \sum_{k=1}^n w_{nk}(t, h) y_k, \\ \hat{\phi}_j(t) &= \sum_{k=1}^n w_{nk}(t, h) x_{kj} \quad 1 \leq j \leq p.\end{aligned}$$

donde $w_{nk}(t, h) = \frac{K\left(\frac{t-T_k}{h}\right)}{\sum_{j=1}^n K\left(\frac{x-T_j}{h}\right)}$. Ahora reemplazando $\hat{\phi}_0$ y $\hat{\phi}$ en (2.5) podemos estimar β por un estimador de mínimos cuadrados.

Para simplificar la notación será conveniente expresar al modelo (2.2) en forma matricial;

$$\mathbf{Y} = \mathbf{X}\beta + \mathbf{G} + \varepsilon,$$

donde $\mathbf{Y} = (y_1, \dots, y_n)^T$, $\mathbf{X} = (\mathbf{X}_1, \dots, \mathbf{X}_n)$, con $\mathbf{X}_i = (x_{i1}, \dots, x_{ip})^T$, $\mathbf{G} = (g(t_1), \dots, g(t_n))^T$ y $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n)$. Entonces el estimador de β de mínimos cuadrados queda definido por

$$\hat{\beta}_{LS} = (\tilde{\mathbf{X}}^T \tilde{\mathbf{X}})^{-1} \tilde{\mathbf{X}}^T \tilde{\mathbf{Y}},$$

con $\tilde{\mathbf{Y}} = (I - W)\mathbf{Y}$, $\tilde{\mathbf{X}} = (I - W)\mathbf{X}$, W la matriz de suavizado con $w_{ij} = w_{ni}(t_j, h)$ el elemento en la posición i, j e I la matriz identidad de dimensión $n \times n$.

A su vez reemplazando $\hat{\phi}_0$ y $\hat{\phi}$ en (2.6) obtenemos el estimador de la componente no paramétrica,

$$\hat{g}_n(t) = \sum_{k=1}^n w_{nk}(t) (Y_k - \mathbf{X}_k^T \hat{\beta}_{LS}).$$

Observación 2.2.1. *Notemos que necesitaremos que la matriz $\tilde{\mathbf{X}}^T \tilde{\mathbf{X}}$ sea invertible. Bajo ciertas hipótesis, que se detallan en el apéndice se puede probar que*

$$\lim_{n \rightarrow \infty} n^{-1} \tilde{\mathbf{X}}^T \tilde{\mathbf{X}} = \Sigma$$

donde Σ es una matriz definida positiva. Con lo cuál esto nos permite deducir que $n(\tilde{\mathbf{X}}^T \tilde{\mathbf{X}})^{-1}$ para n suficientemente grande puede ser obtenida.

Observación 2.2.2. Otra observación importante es que se puede probar también, bajo las hipótesis detalladas en el apéndice, y si ε es un error homoscedástico con varianza σ^2 que $\hat{\beta}_{LS}$ es un estimador asintóticamente normal, es decir,

$$\sqrt{n} \left(\hat{\beta}_{LS} - \beta \right) \xrightarrow{D} N(0, \sigma^2 \Sigma^{-1}). \quad (2.9)$$

Capítulo 3

Estimación en MPL con respuestas faltantes

3.1. Introducción

Como mencionamos en la Introducción, en la práctica, algunas observaciones pueden faltar debido a distintos acontecimientos. Por ejemplo, la variable respuesta Y puede ser muy cara para medir, o bien la variable Y representa un conjunto de preguntas y algunos individuos muestreados se niegan a suministrar la información deseada. En esta sección, vamos a estudiar el modelo parcialmente lineal con datos incompletos. Más precisamente, sean $(Y_i, \delta_i, \mathbf{X}_i^T, T_i)$, con $i = 1, 2, \dots, n$, donde todas las \mathbf{X}_i 's son observadas y $\delta_i = 0$ si Y_i es desconocido mientras que $\delta_i = 1$ si la observación Y_i es observada. Es decir, a diferencia del modelo (2.2) tenemos una nueva variable δ que corresponde a un marcador de la presencia o ausencia de la variable de respuesta Y .

A lo largo de esta tesis supondremos que Y se clasifica como desaparecida al azar, comunmente denominado MAR (Missing At Random). El supuesto MAR implica que δ e Y son condicionalmente independiente dado \mathbf{X} y T . Es decir, $P(\delta = 1|Y, \mathbf{X}, T) = P(\delta = 1|\mathbf{X}, T)$.

A la hora de generar procedimientos de estimación será necesario tener en cuenta la presencia de los datos faltantes en las propuestas clásicas de estimación. A continuación discutiremos algunos de los métodos de estimación de β y $g(\cdot)$, los parámetros desconocidos de un modelo parcialmente lineal, en el caso de contar con respuestas faltantes.

3.2. Estimadores basados en métodos de imputación y propiedades asintóticas

Wang et al. (2004) desarrollaron un procedimiento de estimación basado en la imputación de datos en modelos parcialmente lineales. Es decir, consideraron imputar un posible

valor para cada dato faltante y luego analizar los resultados como si estuvieran completos.

Comencemos introduciendo notación que necesitaremos a lo largo de este Capítulo. Llamemos \mathbf{Z} al vector de variables explicativas, $\mathbf{Z} = (\mathbf{X}^T, T)$ y $\sigma^2(\mathbf{Z}) = E(\epsilon^2|\mathbf{Z})$. Consideremos las siguientes probabilidades de observar a Y condicional a \mathbf{Z} y T respectivamente, es decir $\Delta(\mathbf{Z}) = P(\delta = 1|\mathbf{Z} = \mathbf{z})$ y $\Delta_T(t) = P(\delta = 1|T = t)$.

Si conociéramos el parámetro β y la función g , podríamos considerar las siguientes observaciones,

$$U_i^{[I]} = \delta_i Y_i + (1 - \delta_i)(\mathbf{X}_i^T \beta + g(T_i)),$$

es decir, $U_i^{[I]} = Y_i$ si contamos con la observación Y_i ($\delta_i = 1$) y $U_i^{[I]} = \mathbf{X}_i^T \beta + g(T_i)$ en caso de estar ausente la observación Y_i . Bajo el supuesto MAR, tenemos que

$$E[U^{[I]}|\mathbf{Z}] = E[\delta Y + (1 - \delta)(\mathbf{X}^T \beta + g(T))|\mathbf{Z}] = \Delta(\mathbf{Z})E(Y|\mathbf{Z}) + (1 - \Delta(\mathbf{Z}))(\mathbf{X}^T \beta + g(T)).$$

Es decir, $E[U^{[I]}|\mathbf{Z}] = \mathbf{X}^T \beta + g(T)$ entonces podemos pensar que $U^{[I]}$ satisface el siguiente modelo

$$U_i^{[I]} = \mathbf{X}_i^T \beta + g(T_i) + e_i, \quad (3.1)$$

donde e_i es una error aleatorio con $E[e_i|\mathbf{Z}_i] = 0$. Por lo tanto, $(U_i^{[I]}, \mathbf{X}_i^T, T_i)$ satisfacen un modelo parcialmente lineal y podemos aplicar el procedimiento de estimación descrito en el Capítulo anterior. Para esto, consideremos los pesos

$$\omega_{ni}(t) = \frac{M\left(\frac{t-T_i}{b_n}\right)}{\sum_{i=1}^n M\left(\frac{t-T_i}{b_n}\right)}$$

con $M(\cdot)$ una función núcleo y b_n un parámetro de suavizado o ventana. Por lo tanto, un estimador $\tilde{\beta}_I$ de β puede ser obtenido como

$$\tilde{\beta}_I = \left[\sum_{i=1}^n (\mathbf{X}_i - \tilde{\phi}(T_i))(\mathbf{X}_i - \tilde{\phi}(T_i))^T \right]^{-1} \sum_{i=1}^n (\mathbf{X}_i - \tilde{\phi}(T_i))(U_i^{[I]} - \tilde{\phi}^{[I]}(T_i)), \quad (3.2)$$

donde $\tilde{\phi}(t)$ y $\tilde{\phi}^{[I]}(t)$ corresponden a los estimadores de las esperanzas condicionales $\phi(t) = E(\mathbf{X}|T = t)$ y $\phi_0(t) = E(U^{[I]}|T = t) = E(Y|T = t)$, respectivamente y están dados por

$$\tilde{\phi}(t) = \sum_{i=1}^n \omega_{ni}(t) \mathbf{X}_i, \quad \tilde{\phi}^{[I]}(t) = \sum_{i=1}^n \omega_{ni}(t) U_i^{[I]}. \quad (3.3)$$

Notemos que del hecho que, $U_i^{[I]}$ depende de β y $g(T_i)$, ambos desconocido, luego $\tilde{\beta}_I$ no es un estimador factible. Una estrategia para resolver este problema, consiste en obtener estimadores previos de β y g , usando solo las observaciones que efectivamente observamos y que notaremos con $\hat{\beta}_C$ y \hat{g}_C . Más precisamente, sea

$$\omega_{ni}^C(t) = \frac{K\left(\frac{t-T_i}{h_n}\right)}{\sum_{j=1}^n \delta_j K\left(\frac{t-T_j}{h_n}\right)}$$

donde $K(\cdot)$ es una función núcleo y h_n una ventana, es decir ω_{ni}^C serían los pesos asociados al suavizado pero solo considerando aquellas observaciones T_i tales que $\delta_i = 1$. Por lo tanto

$$\hat{\beta}_C = \left[\sum_{i=1}^n \delta_i (\mathbf{X}_i - \hat{\phi}_C(T_i)) (\mathbf{X}_i - \hat{\phi}_C(T_i))^T \right]^{-1} \sum_{i=1}^n \delta_i (\mathbf{X}_i - \hat{\phi}_C(T_i)) (Y_i - \hat{\phi}_C^{[I]}(T_i)) \quad (3.4)$$

y

$$\hat{g}_C(t) = \hat{\phi}_C^{[I]}(t) - \hat{\phi}_C(t)^T \hat{\beta}_C \quad (3.5)$$

donde

$$\hat{\phi}_C(t) = \sum_{j=1}^n \delta_j w_{nj}^C(t) \mathbf{X}_j, \quad \hat{\phi}_C^{[I]}(t) = \sum_{j=1}^n \delta_j w_{nj}^C(t) Y_j. \quad (3.6)$$

Ahora, naturalmente reemplazamos en (3.1), β y g por sus respectivos estimadores $\hat{\beta}_C$ y \hat{g}_C y en lugar de considerar las imputaciones $U_i^{[I]}$ tendremos $U_{ni}^{[I]}$ donde

$$U_{ni}^{[I]} = \delta_i Y_i + (1 - \delta_i) (\mathbf{X}_i^T \hat{\beta}_C + \hat{g}_C(T_i)). \quad (3.7)$$

De esta forma, $U_{ni}^{[I]}$ si es factible de cálculo, entonces obtenemos un estimador de β como

$$\hat{\beta}_I = \left[\sum_{i=1}^n (\mathbf{X}_i - \hat{\phi}(T_i)) (\mathbf{X}_i - \hat{\phi}(T_i))^T \right]^{-1} \sum_{i=1}^n (\mathbf{X}_i - \hat{\phi}(T_i)) (U_{ni}^{[I]} - \hat{\phi}^{[I]}(T_i)), \quad (3.8)$$

donde

$$\hat{\phi}(t) = \sum_{i=1}^n \omega_{ni}(t) \mathbf{X}_i, \quad \hat{\phi}^{[I]}(t) = \sum_{i=1}^n \omega_{ni}(t) U_{ni}^{[I]}, \quad (3.9)$$

y finalmente estimaremos $g(t)$ como

$$\hat{g}_n^{[I]}(t) = \hat{\phi}^{[I]}(t) - \hat{\phi}(t)^T \hat{\beta}_I. \quad (3.10)$$

Bajo algunas hipótesis que se detallan en el Apéndice, se puede obtener el comportamiento asintótico de $\hat{\beta}_I$. Más precisamente, si notamos $\tilde{\mathbf{X}} = \mathbf{X} - E(\mathbf{X}|T)$ y $\tilde{\mathbf{X}} = \mathbf{X} - \frac{E(\delta \mathbf{X}|T)}{E(\delta|T)}$ y consideramos las siguientes matrices

$$\Sigma_0 = E[\Delta(\mathbf{Z}) \tilde{\mathbf{X}} \tilde{\mathbf{X}}^T] \quad \Sigma_1 = E[\tilde{\mathbf{X}} \tilde{\mathbf{X}}^T] \quad \Sigma_2 = E[(1 - \Delta(\mathbf{Z})) \tilde{\mathbf{X}} \tilde{\mathbf{X}}^T]$$

se puede probar que

$$\sqrt{n}(\hat{\beta}_I - \beta) \xrightarrow{D} N(0, \Sigma_1^{-1} V_I \Sigma_1^{-1}), \quad (3.11)$$

donde

$$V_I = (\Sigma_2 + \Sigma_0) \Sigma_0^{-1} E[\Delta(\mathbf{Z}) \tilde{\mathbf{X}} \tilde{\mathbf{X}}^T \sigma^2(\mathbf{Z})] \Sigma_0^{-1} (\Sigma_2 + \Sigma_0). \quad (3.12)$$

Teorema 3.2.1. *Bajo los mismos supuestos necesarios para demostrar (3.11), si $b_n = O_p(n^{-\frac{1}{3}})$ y $h_n = O_p(n^{-\frac{1}{3}})$, tenemos que*

$$\hat{g}_n^{[I]}(t) - g(t) = O_p(n^{-\frac{1}{3}}).$$

La demostración se detalla en el Apéndice.

3.3. Estimadores semiparamétricos de regresión y propiedades asintóticas

Matloff (1981) observó que si la forma de la función de regresión se conoce salvo algún parámetro desconocido entonces sustituir la variable de respuesta por los valores de regresión estimados o predichos sin importar si la respuesta fue observada o no, arroja buenos resultados. Wang y Sun (2007) adaptan esta enfoque al caso de la regresión semiparamétrica. Este método sólo utiliza los estimadores semiparamétricos de regresión en vez del valor correspondiente de la respuesta, no importa si fue observada o no. Consideremos entonces

$$U_{ni}^{[R]} = \mathbf{X}_i^T \hat{\boldsymbol{\beta}}_C + \hat{g}_C(T_i) \quad (3.13)$$

donde $\hat{\boldsymbol{\beta}}_C$ y \hat{g}_C fueron introducidos en (3.4) y (3.5), respectivamente. Luego los estimadores de $\boldsymbol{\beta}$ y g que denotaremos por $\hat{\boldsymbol{\beta}}_R$ y \hat{g}_R , se definen igual que $\hat{\boldsymbol{\beta}}_I$ reemplazando $U_{ni}^{[I]}$ por $U_{ni}^{[R]}$ para $i = 1, 2, \dots, n$ y reemplazando $U_{ni}^{[I]}$ y $\hat{\boldsymbol{\beta}}_I$ por $U_{ni}^{[R]}$ y $\hat{\boldsymbol{\beta}}_R$, respectivamente. Es decir,

$$\hat{\boldsymbol{\beta}}_R = \left[\sum_{i=1}^n (\mathbf{X}_i - \hat{\phi}(T_i)) (\mathbf{X}_i - \hat{\phi}(T_i))^T \right]^{-1} \sum_{i=1}^n (\mathbf{X}_i - \hat{\phi}(T_i)) (U_{ni}^{[R]} - \hat{\phi}^{[R]}(T_i)), \quad (3.14)$$

donde

$$\hat{\phi}(t) = \sum_{i=1}^n \omega_{ni}(t) \mathbf{X}_i, \quad \hat{\phi}^{[R]}(t) = \sum_{i=1}^n \omega_{ni}(t) U_{ni}^{[R]}, \quad (3.15)$$

y

$$\hat{g}_n^{[R]}(t) = \hat{\phi}^{[R]}(t) - \hat{\phi}(t)^T \hat{\boldsymbol{\beta}}_R. \quad (3.16)$$

Bajo los mismos supuestos necesarios para probar (3.11), que se detallan en el Apéndice, se puede probar

$$\sqrt{n} \left(\hat{\boldsymbol{\beta}}_R - \boldsymbol{\beta} \right) \xrightarrow{D} N(0, \Sigma_1^{-1} V_R \Sigma_1^{-1}) \quad (3.17)$$

donde

$$V_R = \Sigma_1 \Sigma_0^{-1} E[\tilde{\mathbf{X}} \tilde{\mathbf{X}}^T \Delta(\mathbf{Z}) \sigma^2(\mathbf{Z})] \Sigma_0^{-1} \Sigma_1.$$

Teorema 3.3.1. *Bajo los mismos supuestos necesarios para demostrar (3.11), si $b_n = O_p(n^{-\frac{1}{3}})$ y $h_n = O_p(n^{-\frac{1}{3}})$, tenemos que*

$$\hat{g}_n^{[R]}(t) - g(t) = O_p(n^{-\frac{1}{3}}).$$

La demostración se detalla en el Apéndice.

Observación 3.3.2. *A continuación mostraremos que $\hat{\boldsymbol{\beta}}_R$ tiene la misma varianza asintótica que $\hat{\boldsymbol{\beta}}_I$*

Demostración: Para probar que la observación es válida, bastará ver que $\Sigma_0 + \Sigma_2 = \Sigma_1$.

Recordemos que $\Sigma_0 = E[\Delta(\mathbf{Z})\tilde{\mathbf{X}}\tilde{\mathbf{X}}^T]$, $\Sigma_1 = E[\check{\mathbf{X}}\check{\mathbf{X}}^T]$, $\Sigma_2 = E[(1 - \Delta(\mathbf{Z}))\check{\mathbf{X}}\check{\mathbf{X}}^T]$

donde $\tilde{\mathbf{X}} = \mathbf{X} - \frac{E(\delta\mathbf{X}|T)}{E(\delta|T)}$, $\check{\mathbf{X}} = \mathbf{X} - E(\mathbf{X}|T)$ y $\Delta(\mathbf{Z}) = E(\delta|\mathbf{X}T)$

Sea $\mathbf{H} = \mathbf{H}(T) = E(\mathbf{X}|T) - \frac{E(\delta\mathbf{X}|T)}{E(\delta|T)}$ entonces podemos reescribir $\tilde{\mathbf{X}} = \check{\mathbf{X}} + \mathbf{H}$

Comencemos estudiando $\Delta(\mathbf{Z})\tilde{\mathbf{X}}\tilde{\mathbf{X}}^T + (1 - \Delta(\mathbf{Z}))\check{\mathbf{X}}\check{\mathbf{X}}^T$:

$$\begin{aligned} \Delta(\mathbf{Z})\tilde{\mathbf{X}}\tilde{\mathbf{X}}^T + (1 - \Delta(\mathbf{Z}))\check{\mathbf{X}}\check{\mathbf{X}}^T &= \check{\mathbf{X}}\check{\mathbf{X}}^T + \delta(\mathbf{Z})\tilde{\mathbf{X}}\tilde{\mathbf{X}}^T + (1 - \Delta(\mathbf{Z}))\check{\mathbf{X}}\check{\mathbf{X}}^T - \check{\mathbf{X}}\check{\mathbf{X}}^T \\ &= \check{\mathbf{X}}\check{\mathbf{X}}^T + \Delta(\mathbf{Z})\check{\mathbf{X}}\check{\mathbf{X}}^T + \Delta(\mathbf{Z})\check{\mathbf{X}}\mathbf{H}^T + \Delta(\mathbf{Z})\check{\mathbf{X}}\mathbf{H}^T + \Delta(\mathbf{Z})\mathbf{H}\mathbf{H}^T \\ &\quad + \check{\mathbf{X}}\check{\mathbf{X}}^T + \check{\mathbf{X}}\mathbf{H}^T - \Delta(\mathbf{Z})\check{\mathbf{X}}\check{\mathbf{X}}^T - \Delta(\mathbf{Z})\check{\mathbf{X}}\mathbf{H}^T - \check{\mathbf{X}}\check{\mathbf{X}}^T \\ &= \check{\mathbf{X}}\check{\mathbf{X}}^T + \Delta(\mathbf{Z})\check{\mathbf{X}}\mathbf{H}^T + \Delta(\mathbf{Z})\mathbf{H}\mathbf{H}^T + \check{\mathbf{X}}\mathbf{H}^T \\ &= \check{\mathbf{X}}\check{\mathbf{X}} + (1 + \Delta(\mathbf{Z}))(\check{\mathbf{X}}\mathbf{H}^T) + \Delta(\mathbf{Z})\mathbf{H}\mathbf{H}^T \end{aligned}$$

En segunda instancia veamos que $E\left[(1 + \Delta(\mathbf{Z}))(\check{\mathbf{X}}\mathbf{H}^T)\right] + E\left[\Delta(\mathbf{Z})\mathbf{H}\mathbf{H}^T\right] = E\left[\delta\check{\mathbf{X}}\mathbf{H}^T\right]$:

Usando que

$$E\left[\check{\mathbf{X}}\mathbf{H}^T\right] = E\left[\left(\mathbf{X} - E(\mathbf{X}|T)\right) \overbrace{\left(E(\mathbf{X}|T) - \frac{E(\delta\mathbf{X}|T)}{E(\delta|T)}\right)^T}^{\mathbf{H}=\mathbf{H}(T)}\right] = E\left[\underbrace{E(\mathbf{X} - E(\mathbf{X}|T))}_{=0} \mathbf{H}(T)^T\right] = 0.$$

Se deduce

$$\begin{aligned} E\left[(1 + \Delta(\mathbf{Z}))(\check{\mathbf{X}}\mathbf{H}^T)\right] + E\left[\Delta(\mathbf{Z})\mathbf{H}\mathbf{H}^T\right] &= E\left[\Delta(\mathbf{Z})\check{\mathbf{X}}\mathbf{H}^T + \Delta(\mathbf{Z})\mathbf{H}\mathbf{H}^T\right] = E\left[\Delta(\mathbf{Z})\tilde{\mathbf{X}}\mathbf{H}^T\right] \\ &= E\left[E(\delta|\mathbf{X}T) \underbrace{\left[\mathbf{X} - \frac{E(\mathbf{X}\delta|T)}{E(\delta|T)}\right] \left[E(\mathbf{X}|T) - \frac{E(\delta\mathbf{X}|T)}{E(\delta|T)}\right]^T}_{g(\mathbf{X},T)}\right] \\ &= E\left[E\left[\delta \left[\mathbf{X} - \frac{E(\mathbf{X}\delta|T)}{E(\delta|T)}\right] \left[E(\mathbf{X}|T) - \frac{E(\delta\mathbf{X}|T)}{E(\delta|T)}\right]^T \middle| \mathbf{X}T\right]\right] \\ &= E\left[\delta \left[\mathbf{X} - \frac{E(\mathbf{X}\delta\mathbf{X}|T)}{E(\delta|T)}\right] \left[E(\mathbf{X}|T) - \frac{E(\delta\mathbf{X}|T)}{E(\delta|T)}\right]^T\right] \\ &= E\left[\delta\check{\mathbf{X}}\mathbf{H}^T\right] \end{aligned}$$

Por último veamos que $E\left[\delta\check{\mathbf{X}}\mathbf{H}^T\right] = 0$:

$$\begin{aligned}
E \left[\delta \tilde{\mathbf{X}} \mathbf{H}^T \right] &= E \left[\delta \left[\mathbf{X} - \frac{E(\mathbf{X} \delta \mathbf{X} | T)}{E(\delta | T)} \right] \left[E(\mathbf{X} | T) - \frac{E(\delta \mathbf{X} | T)}{E(\delta | T)} \right]^T \right] \\
&= E \left[E \left[\delta \left[\mathbf{X} - \frac{E(\delta \mathbf{X} | T)}{E(\delta | T)} \right] | T \right] \left[E(\mathbf{X} | T) - \frac{E(\delta \mathbf{X} | T)}{E(\delta | T)} \right]^T \right] = 0
\end{aligned}$$

Pues

$$\begin{aligned}
E \left[\delta \left[\mathbf{X} - \frac{E(\delta \mathbf{X} | T)}{E(\delta | T)} \right] | T \right] &= E(\delta \mathbf{X} | T) - E \left[\delta \frac{E(\delta \mathbf{X} | T)}{E(\delta | T)} | T \right] \\
&= E[\delta \mathbf{X} | T] - \frac{E(\delta \mathbf{X} | T)}{E(\delta | T)} E(\delta | T) = 0
\end{aligned}$$

Luego juntando las tres igualdades tenemos que

$$\begin{aligned}
\Sigma_0 + \Sigma_2 &= E \left[\Delta(\mathbf{Z}) \tilde{\mathbf{X}} \tilde{\mathbf{X}}^T \right] + E \left[(1 - \Delta(\mathbf{Z})) \tilde{\mathbf{X}} \tilde{\mathbf{X}}^T \right] \\
&= E \left[\tilde{\mathbf{X}} \tilde{\mathbf{X}} \right] + E \left[(1 + \Delta(\mathbf{Z})) \tilde{\mathbf{X}} \mathbf{H}^T \right] + E \left[\Delta(\mathbf{Z}) \mathbf{H} \mathbf{H}^T \right] \\
&= E \left[\tilde{\mathbf{X}} \tilde{\mathbf{X}} \right] + E \left[\delta \tilde{\mathbf{X}} \mathbf{H}^T \right] \\
&= E \left[\tilde{\mathbf{X}} \tilde{\mathbf{X}} \right] \\
&= \Sigma_1
\end{aligned}$$

□

3.4. Modelos ponderados por la inversa de la probabilidad marginal

Otro enfoque muy utilizado cuando se trabaja con datos faltantes corresponde al método de la probabilidad inversa ponderada. Entre algunos de los autores que han hecho aportes en este tema se encuentra los trabajos de Zhao, Lipsitz y Lew (1996), Wang et al. (1997), Robins, Rotnitzky y Zhao (1994) y Wang, Lindon y Härdle (2004). Este método en general depende de un suavizado en dimensiones altas para la estimación de la función de *propensity score* y por lo tanto la “maldición de la dimensión” puede restringir su uso. Sin embargo, Wang et al. (2004) realiza una propuesta basada en este método donde esquiva este problema.

En primer lugar, notamos que bajo la condición MAR,

$$E \left[\frac{\delta_i}{\Delta(\mathbf{Z}_i)} Y_i + \left(1 - \frac{\delta_i}{\Delta(\mathbf{Z}_i)} \right) (\mathbf{X}_i^t \boldsymbol{\beta} + g(T_i)) | \mathbf{Z}_i \right] = \mathbf{X}_i^t \boldsymbol{\beta} + g(T_i).$$

Luego análogamente a lo introducido en la Sección 3.2, podríamos usar como variable de respuesta

$$\frac{\delta_i}{\Delta(\mathbf{Z}_i)} Y_i + \left(1 - \frac{\delta_i}{\Delta(\mathbf{Z}_i)} \right) (\mathbf{X}_i^t \boldsymbol{\beta} + g(T_i))$$

para obtener los estimadores β y $g(\cdot)$. Pero además de β y $g(\cdot)$ tenemos que $\Delta(\mathbf{Z})$ es desconocido y no contamos con información a priori acerca de su estructura, por lo tanto necesitaríamos contar con un estimador no paramétrico para estimar Δ . Sin embargo, este es el problema que mencionábamos en el párrafo anterior, donde al estimar $\Delta(\mathbf{Z})$ estaríamos utilizando un suavizado con variables explicativas $\mathbf{Z} = (\mathbf{X}^T, T)$ y tendríamos el problema de la “maldición de la dimensión”. Una alternativa propuesta por Wang et al. (2004) consiste en utilizar

$$U_i^{[IP]} = \frac{\delta_i}{\Delta_t(T_i)} Y_i + \left(1 - \frac{\delta_i}{\Delta_t(T_i)}\right) (\mathbf{X}_i^t \beta + g(T_i)) \quad (3.18)$$

y tomando esperanza condicional respecto a \mathbf{Z} , tenemos

$$E(U_i^{[IP]} | \mathbf{Z}_i) = \mathbf{X}_i^T \beta + g(T_i).$$

Luego tenemos,

$$U_i^{[IP]} = \mathbf{X}_i^t \beta + g(T_i) + \eta_i, \quad (3.19)$$

donde η_1, \dots, η_n satisfacen $E[\eta_i | \mathbf{Z}_i] = 0$. Por lo tanto, la ecuación (3.19) tiene la forma de un modelo parcialmente lineal. Entonces podríamos definir un nuevo estimador, que denotaremos $\hat{\beta}_{IP}$, de forma análoga a $\hat{\beta}_I$ pero reemplazando $U_{ni}^{[I]}$ por $U_{ni}^{[IP]}$ donde

$$U_{ni}^{[IP]} = \frac{\delta_i}{\Delta_t(\hat{T}_i)} Y_i + \left(1 - \frac{\delta_i}{\Delta_t(\hat{T}_i)}\right) (\mathbf{X}_i^t \hat{\beta}_C + g_n^C(T_i))$$

donde $\hat{\Delta}_t(T_i) = \sum_{i=1}^n \tilde{\omega}_{ni}(t) \delta_i$ con $\tilde{\omega}_{ni}(t) = \frac{\Omega\left(\frac{t-T_i}{\gamma_n}\right)}{\sum_{j=1}^n \Omega\left(\frac{t-T_j}{\gamma_n}\right)}$, $\Omega(\cdot)$ una función núcleo y γ_n una secuencia de ventanas.

Asimismo, el estimador de $g(\cdot)$ que notaremos como $\hat{g}_n^{[IP]}$, se puede definir como $\hat{g}_n^{[I]}(\cdot)$ reemplazando $U_{ni}^{[I]}$ y $\hat{\beta}_I$ por $U_{ni}^{[IP]}$ y $\hat{\beta}_{IP}$, respectivamente.

Bajo los mismos supuestos necesarios para las distribuciones asintóticas de los estimadores desarrollados en las secciones anteriores, que se detallan en el apéndice, tenemos que

$$\sqrt{n}(\hat{\beta}_{IP} - \beta) \xrightarrow{d} N(0, \Sigma_1^{-1} V_{IP} \Sigma_1^{-1}),$$

donde $V_{IP} = E\{L(T) \Sigma_0^{-1} \check{\mathbf{X}} \check{\mathbf{X}}^T \Sigma_0^{-1} L(T) \Delta(\mathbf{Z}) \sigma^2(\mathbf{Z})\}$ y $L(T) = \frac{\Sigma_0}{\Delta_t(T)} + E\left(\left(1 - \frac{\delta}{\Delta_t(T)}\right) \check{\mathbf{X}} \check{\mathbf{X}}^T\right)$.

Teorema 3.4.1. *Bajo los mismos supuestos necesarios para demostrar (3.11), y (3.17) si $b_n = O_p(n^{-\frac{1}{3}})$, $h_n = O_p(n^{-\frac{1}{3}})$ y $\gamma_n = O_p(n^{-\frac{1}{3}})$ tenemos que*

$$\hat{g}_n^{[IP]}(t) - g(t) = O_p(n^{-\frac{1}{3}}).$$

La demostración se detalla en el Apéndice.

Capítulo 4

Datos Direccionales

4.1. Introducción

En diversos campos surgen problemas estadísticos donde los datos son recogidos mediante medidas angulares dando la orientación o ángulos en el plano (datos circulares) o en el espacio (datos esféricos). Los datos circulares constituyen el caso más simple de esta categoría de datos llamados datos direccionales, donde la medida no es escalar, sino que es angular o direccional. Los datos circulares aparecen en muchos contextos y se obtienen de diversas formas. Los principales corresponden a los instrumentos de medición circular clásicos: la brújula y el reloj. Entre las observaciones típicas medidas por la brújula se encuentran las direcciones del viento o las direcciones del vuelo migratorio de las aves, como así también, la dirección del polo magnético de la tierra. En el caso del reloj podrían ser de interés, por ejemplo, los horarios más frecuentes de llegada de los pacientes a la guardia de un hospital. Las direcciones en el plano pueden representarse a través de vectores unitarios (i.e. direcciones) o como puntos en el círculo unidad. También podríamos pensar una dirección como un ángulo o bien como un número complejo unitario. En estos dos últimos casos es preciso fijar una dirección inicial y un sentido de rotación para evitar ambigüedades en la representación. De esta manera, dado un ángulo, un punto en el plano puede representarse como,

$$\mathbf{t} = (\cos \theta, \sin \theta)^T$$

$$z = e^{i\theta} = \cos \theta + i \sin \theta$$

En este capítulo presentamos la descriptiva de los datos circulares siguiendo Mardia y Jupp (2000). Comenzamos estudiando las medidas de posición, concentración y dispersión. Luego mencionamos las principales distribuciones circulares. Y finalmente estudiamos la estimación de la función densidad circular y las de regresión circular-lineal, como referencias de estos temas pueden citarse los trabajos de [5], [1], [3] y [10], entre otros.

4.2. Medidas de Posición

Sean $\mathbf{T}_1, \dots, \mathbf{T}_n$ vectores unitarios que se corresponden con los ángulos $\theta_j, j=1, \dots, n$. Se define la *dirección media* $\bar{\theta}$ de $\theta_1, \dots, \theta_n$ como la dirección de la resultante $\mathbf{T}_1 + \dots + \mathbf{T}_n$, que coincide con la dirección del centro de masa $\bar{\mathbf{T}}$ de $\mathbf{T}_1, \dots, \mathbf{T}_n$. Dado que las coordenadas cartesianas de \mathbf{T}_j son $(\cos \theta_j, \sin \theta_j)$ para $j=1, \dots, n$, se tiene que las coordenadas del centro de masa serán (\bar{C}, \bar{S}) , donde

$$\bar{C} = \frac{1}{n} \sum_{j=1}^n \cos \theta_j, \quad \bar{S} = \frac{1}{n} \sum_{j=1}^n \sin \theta_j$$

Si definimos la *longitud media resultante* \bar{R} como $\bar{R} = (\bar{C}^2 + \bar{S}^2)^{1/2}$ resulta que $\bar{\theta}$ es solución de las ecuaciones

$$\bar{C} = \bar{R} \cos \bar{\theta}, \quad \bar{S} = \bar{R} \sin \bar{\theta} \quad (4.1)$$

Esto es válido siempre y cuando $\bar{R} > 0$. En el caso en que $\bar{R} = 0$, $\bar{\theta}$ no está definida.

Teniendo en cuenta estas nuevas definiciones, queda claro que en el contexto de datos circulares $\bar{\theta}$ *no* es lo mismo que $(\theta_1 + \dots + \theta_n)/n$, y que $\bar{\theta}$ constituye una medida más apropiada que $(\theta_1 + \dots + \theta_n)/n$ ya que no depende de la elección de la dirección inicial.

Por otra parte para ciertos casos (por ejemplo, la estimación robusta) es útil tener una versión de la mediana muestral adaptada al caso de datos circulares. Se define la *dirección mediana muestral* $\hat{\theta}$ de $\theta_1, \dots, \theta_n$ como un ángulo ϕ que verifique las siguientes dos condiciones:

- la mitad de los datos caen dentro del arco $[\phi, \phi + \pi]$,
- la mayoría de los datos están más cerca de ϕ que de $\phi + \pi$.

Cuando el tamaño de la muestra es un número impar, la mediana muestral coincide con uno de los datos. Cuando es par, es conveniente tomar como mediana muestral el punto medio entre dos datos adyacentes apropiados.

4.3. Medidas de concentración y dispersión

En la Sección anterior definimos la longitud media resultante \bar{R} como la longitud del centro de masa $\bar{\mathbf{T}}$, dada por

$$\bar{R} = (\bar{C}^2 + \bar{S}^2)^{1/2}.$$

Dado que $\mathbf{T}_1, \dots, \mathbf{T}_n$ son vectores unitarios, se tiene que

$$0 \leq \bar{R} \leq 1.$$

de lo que se deduce que:

- Si las direcciones $\theta_1, \dots, \theta_n$ están estrechamente agrupadas, el valor de \bar{R} será muy cercano a 1.
- Si $\theta_1, \dots, \theta_n$ están muy dispersos, \bar{R} tendrá un valor cercano a 0.

Además esta medida de concentración tiene las siguientes propiedades

- \bar{R} es invariante bajo rotaciones.
- La longitud resultante \bar{R} es la longitud del vector resultante $\mathbf{T}_1 + \dots + \mathbf{T}_n$.

Si bien la longitud media resultante \bar{R} es una medida útil, muchas veces se utilizan otras medidas de dispersión que resultan análogas a las utilizadas para datos en la recta real. La más simple de estas medidas es la *varianza muestral circular* definida por

$$V = 1 - \bar{R},$$

que también satisface

$$0 \leq V \leq 1.$$

También podemos definir la *desviación circular estándar* como

$$v = [-2 \log(1 - V)]^{1/2} = [-2 \log \bar{R}]^{1/2}.$$

Notar que v toma valores en $[0, \infty]$ mientras que V lo hace en $[0, 1]$. También podemos considerar la distancia entre dos ángulos θ y ξ es decir

$$1 - \cos(\theta - \xi).$$

y definir la medida de dispersión de los ángulos $\theta_1, \dots, \theta_n$ con respecto a un α dado como

$$D(\alpha) = \frac{1}{n} \sum_{i=1}^n [1 - \cos(\theta_i - \alpha)]$$

y resulta que $D(\bar{\theta}) = V$.

4.4. Distribuciones Circulares

Una forma de especificar una distribución en el círculo unidad es por medio de su función de distribución. Suponemos que han sido escogidas una dirección y orientación inicial. Luego, podemos considerar un ángulo aleatorio Θ cuya *función de distribución* F es una función definida en toda la recta real dada por

$$F(\theta) = P(0 < \Theta \leq \theta) \quad 0 \leq \theta \leq 2\pi,$$

y

$$F(\theta + 2\pi) - F(\theta) = 1 \quad -\infty < \theta < \infty.$$

Esta última ecuación afirma que todo arco de longitud 2π en el círculo unidad tiene probabilidad 1. Este arco es exactamente el perímetro de todo el círculo unidad.

Para $\alpha \leq \beta \leq \alpha + 2\pi$,

$$P(\alpha < \Theta \leq \beta) = F(\beta) - F(\alpha) = \int_{\alpha}^{\beta} dF(\theta) \quad (4.2)$$

donde la integral es una integral de Lebesgue-Stieltjes. La función de distribución es continua a derecha. A diferencia de las funciones de distribución en la recta se tiene que

$$\lim_{\theta \rightarrow \infty} F(\theta) = \infty, \quad \lim_{\theta \rightarrow -\infty} F(\theta) = -\infty.$$

Por definición,

$$F(0) = 0, \quad F(2\pi) = 1.$$

Observemos que, aunque la función F depende de la elección de la dirección inicial, la ecuación (4.2) muestra que $F(\beta) - F(\alpha)$ es independiente de dicha elección. Luego, cambiar la dirección inicial sólo agrega una constante a F .

Si la función de distribución F es absolutamente continua, F tiene una función de densidad f dada por

$$\int_{\alpha}^{\beta} f(\theta) d\theta = F(\beta) - F(\alpha), \quad -\infty < \alpha \leq \beta < \infty.$$

Una función f es la función de densidad de una distribución absolutamente continua si y sólo si

1. $f(\theta) \geq 0$ en casi todo punto de $(-\infty, \infty)$,
2. $f(\theta + 2\pi) = f(\theta)$ en casi todo punto de $(-\infty, \infty)$,
3. $\int_0^{2\pi} f(\theta) d\theta = 1$.

4.4.1. Distribución Uniforme

La distribución más básica en el círculo, es la única distribución en el círculo que es invariante bajo rotaciones y reflexiones. Su función de densidad es

$$f(\theta) = \frac{1}{2\pi}.$$

Luego, para $\alpha \leq \beta \leq \alpha + 2\pi$

$$P(\alpha < \theta \leq \beta) = \frac{\beta - \alpha}{2\pi},$$

es decir, la probabilidad es proporcional a la longitud de arco.

4.4.2. Distribución Von Mises

Desde el punto de vista de la inferencia estadística, una de las distribuciones de probabilidad circulares más utilizada es la *distribución Von Mises* $VM(\mu, \kappa)$, análoga a las distribuciones normales en la recta. Su función de densidad está dada por

$$g(\theta; \mu, \kappa) = \frac{1}{2\pi I_0(\kappa)} e^{\kappa \cos(\theta - \mu)},$$

donde I_0 denota la función de Bessel modificada de primer tipo y orden 0, definida por

$$I_0(\kappa) = \frac{1}{2\pi} \int_0^{2\pi} e^{\kappa \cos \theta} d\theta.$$

El parámetro $\mu \in [0, 2\pi)$ es la dirección media y el parámetro $\kappa \geq 0$ recibe el nombre de *parámetro de concentración*. Como se puede ver en la Figura 4.1 la distribución von Mises es unimodal y simétrica con respecto a $\theta = \mu$. La moda se encuentra en $\theta = \mu$ y la antimoda en $\theta = \mu + \pi$. El cociente entre el valor de la densidad en la moda y el valor de la densidad en la antimoda es $e^{2\kappa}$, así que cuanto mayor sea el valor de κ , mayor será la concentración alrededor de la moda.

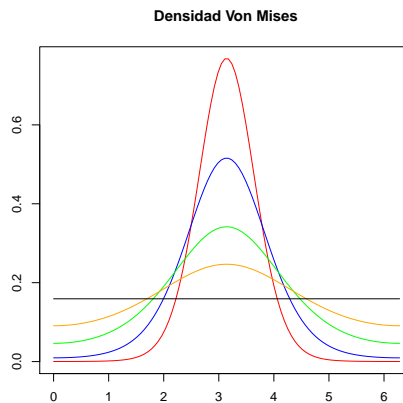


Figura 4.1: Densidad Von Mises $VM(\pi, \kappa)$ para $\kappa = 4$ (rojo), $\kappa = 2$ (azul), $\kappa = 1$ (verde), $\kappa = 0,5$ (naranja), $\kappa = 0$ (negro). Para el valor $\kappa = 0$ esta distribución coincide con la uniforme.

Observación 4.4.1. En la Figura 4.1 también se observa el hecho de que cuanto más grande sea el parámetro κ su función de densidad estará más concentrada.

También hay otras distribuciones conocidas como la distribución Cardicoide, la distribución Normal Proyectada, la distribución Wrapped, entre otras.

4.5. Estimación de la densidad circular

Como vimos en el Capítulo 2, la estimación de la densidad tipo núcleo en el caso lineal viene dada por

$$\hat{f}(t) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{t - T_i}{h}\right) \quad (4.3)$$

donde h es el parámetro de suavizado o ventana, K es una función núcleo no negativa y que verifica la condición $\int K(t)dt = 1$. Si en lugar de tener una muestra de variables aleatorias $T_1 \dots T_n$ tenemos vectores aleatorios $q + 1$ -dimensionales $\mathbf{T}_1, \dots, \mathbf{T}_n$, el estimador (4.3) se escribe como

$$\hat{f}(\mathbf{t}) = \frac{1}{nh^{q+1}} \sum_{i=1}^n K\left(\frac{\mathbf{t} - \mathbf{T}_i}{h}\right), \quad \mathbf{T} \in \mathbb{R}^{q+1}. \quad (4.4)$$

La pregunta natural es como se extiende este estimador en el caso de una muestra de datos circulares. Al trabajar con puntos en el círculo, o con datos direccionales en general, podemos *medir* la distancia entre dos puntos por

$$\begin{aligned} \|\mathbf{t} - \mathbf{T}_i\|^2 &= (\mathbf{t} - \mathbf{T}_i)^T (\mathbf{t} - \mathbf{T}_i) = \mathbf{t}^T \mathbf{T} - 2\mathbf{t}^T \mathbf{T}_i + \mathbf{T}_i^T \mathbf{T}_i \\ &= \|\mathbf{t}\|^2 - 2\mathbf{t}^T \mathbf{T}_i + \|\mathbf{T}_i\|^2. \end{aligned}$$

Como estamos considerando el círculo unitario resulta que

$$\|\mathbf{t} - \mathbf{T}_i\|^2 = 2 - 2\mathbf{t}^T \mathbf{T}_i = 2(1 - \mathbf{t}^T \mathbf{T}_i), \quad (4.5)$$

de esta forma un estimador de densidad, considerando la distancia en el círculo unitario, resulta

$$\hat{f}(\mathbf{t}) = \frac{1}{n} C(h) \sum_{i=1}^n K\left(\frac{1 - \mathbf{t}^T \mathbf{T}_i}{h^2}\right), \quad \mathbf{T} \in \mathbb{R}^{q+1}. \quad (4.6)$$

donde $C(h)$ se elige de manera que el estimador resulte una densidad y K es una función núcleo. Notemos que este estimador también puede escribirse como

$$\hat{f}_n(\mathbf{t}) = \frac{1}{n} d_0(\kappa) \sum_{i=1}^n K[\kappa(1 - \mathbf{t}^T \mathbf{T}_i)]. \quad (4.7)$$

Este estimador fue propuesto y estudiado por Bai et al. [1]. Otra propuesta fue introducida por Hall *et al.* [5], donde estudian un estimador de naturaleza similar para estimar la densidad circular dado por

$$\hat{f}_n(\mathbf{t}) = \frac{1}{n} c_0(\kappa) \sum_{i=1}^n K(\kappa \mathbf{t}^T \mathbf{T}_i), \quad (4.8)$$

donde el nuevo parámetro de suavizado κ reemplaza a la ventana h y $c_0(\kappa)$ se elige de manera tal que el estimador \hat{f}_n integre 1, es decir de manera que resulte también una función de densidad. En el citado trabajo, se demuestra, mediante un cambio de variables, que la integral de $K(\kappa \mathbf{t}^T \mathbf{T}_i)$ no depende de i .

De (4.5) se deduce que si \mathbf{t} está cerca de \mathbf{T}_i , $\mathbf{t} - \mathbf{T}_i$ es cercano a 0, pero $\mathbf{t}^T \mathbf{T}_i$ es cercano a 1. Luego, si bien los estimadores (4.6) y (4.8) son similares tienen características distintas lo que lleva a tener que usar núcleos diferentes.

4.6. Estimación de la función de regresión circular-lineal

En esta Sección veremos como estimar no paramétricamente la función de regresión cuando se tiene una variable respuesta lineal y la variable explicativa es circular. Es decir, consideremos el conjunto de datos (\mathbf{T}_i^T, Y_i) con $i = 1, \dots, n$, donde \mathbf{T}_i e Y_i satisfacen $\|\mathbf{T}_i\| = 1$ y $Y_i \in \mathbb{R}$. Estamos interesados en estimar $E(Y|\mathbf{T} = \mathbf{t})$, pero teniendo en cuenta el hecho que $\|\mathbf{T}\| = 1$.

Al igual que para el caso “lineal” descrito en la Sección 2.2.2 tenemos

$$E(Y|\mathbf{T} = \mathbf{t}) = \int y \frac{f_{y,\mathbf{T}}(y, \mathbf{t})}{f_{\mathbf{T}}(\mathbf{t})} dy.$$

Luego por lo visto en la Sección anterior podemos estimar las funciones de densidad $f_{y,\mathbf{T}}$ y $f_{\mathbf{T}}$, a través de estimadores plug-in pero considerando la naturaleza direccional \mathbf{T} . Por lo tanto, el estimador propuesto, que llamaremos de regresión circular-lineal quedaría;

$$E(\widehat{Y}|\mathbf{T} = \mathbf{t}) = \sum_{k=1}^n w_{nk}(\mathbf{t}) y_k, \tag{4.9}$$

donde los pesos w_{nk} son calculados como en la Sección anterior, por ejemplo

$$w_{nk}(\mathbf{t}) = \frac{K(h^{-2}(1 - \mathbf{t}^T \mathbf{T}_k))}{\sum_{i=1}^n K(h^{-2}(1 - \mathbf{t}^T \mathbf{T}_i))}$$

o

$$w_{nk}(\mathbf{t}) = \frac{K(\kappa \mathbf{t}^T \mathbf{T}_k)}{\sum_{i=1}^n K(\kappa \mathbf{t}^T \mathbf{T}_i)}.$$

Notemos que tanto en la Sección 4.5 como en la Sección 4.6, estamos considerando una muestra aleatoria de vectores unitarios $\mathbf{T}_1, \dots, \mathbf{T}_n$, pero también como notamos al comienzo del Capítulo podríamos haber considerado una muestra de datos angulares $\Theta_1, \dots, \Theta_n$ en $[0, 2\pi)$ donde $\mathbf{T}_i^T = (\cos \Theta_i, \sin \Theta_i)$. En algunas situaciones puede ser conveniente “pensar” en función de los datos angulares en lugar de los datos direccionales. De esta forma si $\mathbf{t} = (\cos \theta, \sin \theta)$ tenemos $\mathbf{t}^T \mathbf{T}_i = \cos(\theta - \Theta_i)$. Luego haciendo un abuso de notación

tenemos que el estimador de (4.6) de la densidad de \mathbf{T} puede escribirse como

$$\widehat{f}_{\Theta}(\theta) = \frac{1}{n} C(h) \sum_{i=1}^n K \left[\frac{1 - \cos(\theta - \Theta_i)}{h^2} \right]$$

y análogamente el estimador de la función de regresión definida en (4.9) puede escribirse como

$$E(\widehat{Y} | \Theta = \theta) = \frac{\frac{1}{n} C(h) \sum_{i=1}^n \frac{1}{h} K \left(\frac{1 - \cos(\theta - \Theta_i)}{h^2} \right) Y_i}{\frac{1}{n} C(h) \sum_{i=1}^n \frac{1}{h} K \left(\frac{1 - \cos(\theta - \Theta_i)}{h^2} \right)} = \frac{\sum_{i=1}^n K \left(\frac{1 - \cos(\theta - \Theta_i)}{h^2} \right) Y_i}{\sum_{i=1}^n K \left(\frac{1 - \cos(\theta - \Theta_i)}{h^2} \right)}.$$

Capítulo 5

Estimación del MPL con componente no paramétrica direccional y respuesta faltante

Como mencionamos en la Introducción el objetivo de esta tesis se orienta al estudio del MPL en el caso en el que la variable a ser modelada noparamétricamente se corresponde con un dato direccional y además cuando parte de las variables de respuesta puede estar *desaparecida al azar*. Como resumen de lo introducido en los Capítulos anteriores, sabemos estimar bajo un modelo parcialmente lineal cuando tenemos respuesta faltantes y bajo un modelo de regresión no paramétrica cuando la naturaleza de la variable explicativa es direccional.

Luego, consideremos $(Y_i, \delta_i, \mathbf{X}_i^T, \mathbf{T}_i^T)$ para $1 \leq i \leq n$ donde $Y_i \in \mathbb{R}$, $\mathbf{X}_i \in \mathbb{R}^p$, $\mathbf{T}_i \in \mathbb{R}^2$ con $\|\mathbf{T}_i\| = 1$ y $\delta_i = 1$ si Y_i es observada y $\delta_i = 0$ en caso contrario, por lo tanto el modelo en el que nos centraremos en este Capítulo puede escribirse como

$$Y_i = \mathbf{X}_i^T \boldsymbol{\beta} + g(\mathbf{T}_i) + \varepsilon_i, \quad i = 1, \dots, n$$

donde $\boldsymbol{\beta}$ y g corresponden a los parámetros a estimar, y ε_i al error aleatorio con $E(\varepsilon_i) = 0$. Además, supondremos que

$$x_{ij} = \phi_j(\mathbf{T}_i) + \eta_{ij} \quad 1 \leq j \leq p,$$

donde $\mathbf{X}_i = (x_{i1}, \dots, x_{ip})^T$ con η_{ij} son errores aleatorios con $E(\eta_{ij}) = 0$.

Entonces en primer lugar consideremos estimadores de $\boldsymbol{\beta}$ y g utilizando solo los datos que observamos en forma completa. Más precisamente, para estimar $\phi_j(t) = E(x_{ij} | \mathbf{T}_i = \mathbf{t})$ para $1 \leq j \leq p$ y $E(Y_i | \mathbf{T}_i = \mathbf{t})$ consideremos los pesos que tiene en cuenta la estructura direccional de la variable \mathbf{T}_i , por ejemplo podemos tomar

$$\omega_{nk}^C(\mathbf{t}) = \frac{K(h^{-2}(1 - \mathbf{t}^T \mathbf{T}_k))}{\sum_{i=1}^n \delta_i K(h^{-2}(1 - \mathbf{t}^T \mathbf{T}_i))},$$

que son los pesos que utilizaremos en el estudio de simulación. De esta manera, tenemos $\hat{\phi}_C(\mathbf{t}) = (\hat{\phi}_{1C}(\mathbf{t}), \dots, \hat{\phi}_{pC}(\mathbf{t}))$ donde

$$\hat{\phi}_{jC}(t) = \sum_{k=1}^n \delta_k \omega_{nk}^C(t) x_{kj} \quad \text{y} \quad \hat{\phi}_C^{[I]}(t) = \sum_{k=1}^n \delta_k \omega_{nk}^C(t) Y_k$$

y por lo tanto

$$\hat{\beta}_C = \left[\sum_{i=1}^n \delta_i (\mathbf{X}_i - \hat{\phi}_C(\mathbf{T}_i)) (\mathbf{X}_i - \hat{\phi}_C(\mathbf{T}_i))^T \right]^{-1} \sum_{i=1}^n \delta_i (\mathbf{X}_i - \hat{\phi}_C(\mathbf{T}_i)) (Y_i - \hat{\phi}_C^{[I]}(\mathbf{T}_i)) \quad (5.1)$$

y

$$\hat{g}_C(\mathbf{t}) = \hat{\phi}_C^{[I]}(\mathbf{t}) - \hat{\phi}_C(\mathbf{t})^T \hat{\beta}_C. \quad (5.2)$$

En el Capítulo 3.2, 3.3 y 3.4 se introducen estimadores de β y g que se basan en los estimadores anteriores $\hat{\beta}_C$ y \hat{g}_C , pero que tienen en cuenta todos los datos. Para esto se definen las siguientes pseudo observaciones,

$$U_{ni}^{[I]} = \delta_i Y_i + (1 - \delta_i) (\mathbf{X}_i^T \hat{\beta}_C + \hat{g}_C(\mathbf{T}_i)) \quad (5.3)$$

$$U_{ni}^{[R]} = \mathbf{X}_i^T \hat{\beta}_C + \hat{g}_C(\mathbf{T}_i) \quad (5.4)$$

$$U_{ni}^{[IP]} = \frac{\delta_i}{\Delta_t(\hat{\mathbf{T}}_i)} Y_i + \left(1 - \frac{\delta_i}{\Delta_t(\hat{\mathbf{T}}_i)} \right) (\mathbf{X}_i^T \hat{\beta}_C + g_n^C(\mathbf{T}_i)) \quad (5.5)$$

donde $\Delta_{\mathbf{t}}(\hat{\mathbf{t}}) = \sum_{k=1}^n \tilde{\omega}_{nk}(\mathbf{t}) \delta_k$ con $\tilde{\omega}_{nk}(\mathbf{t}) = \frac{K(h^{-2}(1 - \mathbf{t}^T \mathbf{T}_k))}{\sum_{i=1}^n K(h^{-2}(1 - \mathbf{t}^T \mathbf{T}_i))}$ donde el núcleo K y la ventana h pueden ser tomados distintos que los utilizados para ω_{nk}^C .

Finalmente, en lugar de considerar la variable Y_i como respuesta, consideremos cada una de las respuestas $U_{ni}^{[I]}$, $U_{ni}^{[R]}$, y $U_{ni}^{[IP]}$ que darán origen a tres diferentes estimadores para cada uno de los parámetros $\hat{\beta}_I$, $\hat{g}^{[I]}$, $\hat{\beta}_R$, $\hat{g}^{[R]}$, $\hat{\beta}_{IP}$ y $\hat{g}^{[IP]}$. Explícitamente, si consideramos

$$\omega_{nk}(\mathbf{t}) = \frac{K(h^{-2}(1 - \mathbf{t}^T \mathbf{T}_k))}{\sum_{i=1}^n K(h^{-2}(1 - \mathbf{t}^T \mathbf{T}_i))}$$

podemos definir

$$\begin{aligned} \hat{\beta}_I &= \left[\sum_{i=1}^n (\mathbf{X}_i - \hat{\phi}(\mathbf{T}_i)) (\mathbf{X}_i - \hat{\phi}(\mathbf{T}_i))^T \right]^{-1} \sum_{i=1}^n (\mathbf{X}_i - \hat{\phi}(\mathbf{T}_i)) (U_{ni}^{[I]} - \hat{\phi}^{[I]}(\mathbf{T}_i)) \\ \hat{\beta}_R &= \left[\sum_{i=1}^n (\mathbf{X}_i - \hat{\phi}(\mathbf{T}_i)) (\mathbf{X}_i - \hat{\phi}(\mathbf{T}_i))^T \right]^{-1} \sum_{i=1}^n (\mathbf{X}_i - \hat{\phi}(\mathbf{T}_i)) (U_{ni}^{[R]} - \hat{\phi}^{[R]}(\mathbf{T}_i)) \quad (5.6) \\ \hat{\beta}_{IP} &= \left[\sum_{i=1}^n (\mathbf{X}_i - \hat{\phi}(\mathbf{T}_i)) (\mathbf{X}_i - \hat{\phi}(\mathbf{T}_i))^T \right]^{-1} \sum_{i=1}^n (\mathbf{X}_i - \hat{\phi}(\mathbf{T}_i)) (U_{ni}^{[IP]} - \hat{\phi}^{[IP]}(\mathbf{T}_i)) \end{aligned}$$

con

$$\begin{aligned}\hat{\phi}(\mathbf{t}) &= \sum_{k=1}^n \omega_{nk}(\mathbf{t}) \mathbf{X}_k, & \hat{\phi}^{[I]}(\mathbf{t}) &= \sum_{k=1}^n \omega_{nk}(\mathbf{t}) U_{nk}^{[I]}, \\ \hat{\phi}^{[R]}(\mathbf{t}) &= \sum_{k=1}^n \omega_{nk}(\mathbf{t}) U_{nk}^{[R]} & \text{y} & \hat{\phi}^{[IP]}(\mathbf{t}) = \sum_{k=1}^n \omega_{nk}(\mathbf{t}) U_{nk}^{[IP]}.\end{aligned}$$

Por último, obtenemos los estimadores de g como

$$\begin{aligned}\hat{g}^{[I]}(\mathbf{t}) &= \hat{\phi}^{[I]}(\mathbf{t}) - \hat{\phi}(\mathbf{t})^T \hat{\beta}_I \\ \hat{g}^{[R]}(\mathbf{t}) &= \hat{\phi}^{[R]}(\mathbf{t}) - \hat{\phi}(\mathbf{t})^T \hat{\beta}_R \\ \hat{g}^{[IP]}(\mathbf{t}) &= \hat{\phi}^{[IP]}(\mathbf{t}) - \hat{\phi}(\mathbf{t})^T \hat{\beta}_{IP}.\end{aligned}\tag{5.7}$$

5.1. Estudio de simulación

Esta Sección contiene los resultados de un estudio de simulación con el fin de analizar y comparar el comportamiento de los estimadores desarrollados para muestras pequeñas.

En todos los procedimientos de suavizados utilizamos el núcleo

$$K(u) = \frac{6}{4}(1 - u^2)I_{(0,1)}(u),$$

notemos que el núcleo no es simétrico ya que será evaluado en una distancia que es siempre positiva. Consideramos el estimador (4.7) como detallamos en el capítulo anterior, estos estimadores dependen de la constante $C(h)$, luego para obtener el valor de $C(h)$, recurrimos al trabajo de Zhao y Wu [17]. En donde se establece que, suponiendo $h = h_n \rightarrow 0$ si $n \rightarrow \infty$,

$$C(h)^{-1} \approx \lambda h^q, \quad \text{con} \quad \lambda = 2^{q/2-1} \omega_{q-1} \int_0^\infty K(t) t^{q/2-1} dt,$$

es decir $C(h)^{-1}/\lambda h^q \rightarrow 1$ si $n \rightarrow \infty$ y $\omega_{q-1} = 2\pi^{q/2}/\Gamma(q/2)$, $q \geq 1$, siendo $\Omega_q \in \mathfrak{R}^{q+1}$ una esfera q -dimensional en la cual toma valores la muestra aleatoria que se está analizando. Como en nuestro caso estamos considerando datos circulares, por lo tanto $q = 1$. Tenemos, $\omega_0 = 2\pi^{1/2}/\Gamma(1/2) = 2\pi^{1/2}/\pi^{1/2} = 2$,

$$\int_0^\infty \frac{3}{4}(1 - t^2)I(|t| < 1)t^{1/2-1} dt = \int_0^1 \frac{6}{4}t^{-1/2} dt = \frac{12}{5}$$

y $\lambda = 2^{1/2-1} 2 \frac{12}{5} = \frac{12}{5} \sqrt{2}$. En cada caso, la simulación se realizó para distintos valores de ancho de ventana prestando atención en la relación sesgo-varianza, estudiada para el caso de la estimación de densidad en el Capítulo 2.

Se generaron 1000 replicaciones de muestras de tamaño $n = 300$ siguiendo el MPL. Más precisamente, las variables $\mathbf{T}_1, \dots, \mathbf{T}_n$ fueron generadas a partir de una muestra de datos angulares siguiendo una distribución Von Mises de parámetros π y 2, es decir se generaron

$\theta_1, \dots, \theta_n$ donde $\theta_i \sim VM(\pi, 2)$, por lo tanto $\mathbf{T}_i = (\cos(\theta_i), \sin(\theta_i))$. A partir de estas variables el modelo considerado es el siguiente:

$$\begin{aligned} y_i &= 2x_i + \cos(\theta_i) + \varepsilon_i & 1 \leq i \leq n \\ x_i &= \frac{5}{\log(5)} \cos(\theta_i) + \eta_i & 1 \leq i \leq n. \end{aligned}$$

Es decir con nuestra notación, $\beta = 2$ y si $\mathbf{T} = (t_1, t_2) = (\cos(\theta), \sin(\theta))$ la función g sería $g(\mathbf{T}) = t_1 = t_1(\theta)$ y $\phi_1(\mathbf{T}) = E(X|\mathbf{T}) = \frac{5}{\log(5)} t_1 = \frac{5}{\log(5)} t_1(\theta)$. Haciendo un abuso de notación, pensaremos indistintamente a g y ϕ_1 en función directamente del ángulo, es decir, escribiremos también $g(\theta) = \cos(\theta)$ y $\phi_1(\theta) = \frac{5}{\log(5)} \cos(\theta)$.

Los errores (ε_i, η_i) se tomaron i.i.d normales con media 0 y matriz de covarianza identidad.

Para generar el mecanismo de perdida consideramos la siguiente función de probabilidad

$$\Delta(T) = \Delta(\Theta) = \begin{cases} 0,6 + \frac{e^{-\theta}}{1+e^{-\theta}} & \text{si } \frac{e^{-\theta}}{1+e^{-\theta}} \leq 0,1 \\ 0,5 & \text{si } \frac{e^{-\theta}}{1+e^{-\theta}} > 0,1 \end{cases}$$

Observación 5.1.1. *La tasa de respuesta utilizando este mecanismo de pérdida es aproximadamente del 60%.*

En las tablas y figuras, notamos $\hat{\beta}_I$, $\hat{\beta}_R$ y $\hat{\beta}_{IP}$ los estimadores de β definidos en (5.6) y $\hat{g}^{[I]}$, $\hat{g}^{[R]}$ y $\hat{g}^{[IP]}$ los estimadores de g definidos en (5.7). Mientras que $\hat{\beta}_C$ y \hat{g}^C denotan los estimadores utilizando solo los datos observados definidos en (5.1) y (5.2), respectivamente y finalmente los estimadores resultantes de utilizar todos los datos, es decir si no tuvieramos pérdida, fueron denotados mediante $\hat{\beta}_{full}$ y \hat{g}^{full} . A fin de resumir los resultados de la simulación en las tablas y figuras, se consideraron las siguientes medidas:

- para los estimadores del parámetro β , en la Tabla 5.3 consideramos el error cuadrático medio sobre las 1000 replicaciones que indicamos por $MSE(\hat{\beta})$, es decir,

$$MSE(\hat{\beta}) = \frac{1}{1000} \sum_{i=1}^{1000} (\hat{\beta}_i - 2)^2,$$

donde $\hat{\beta}_i$ corresponde al estimador obtenido para la muestra i . Por otra parte, en las Tablas 5.1 y 5.2 se presentaron los valores medios y los desvíos estándar calculados sobre las 1000 replicaciones. Estas medidas fueron obtenidas para cada uno de los estimadores propuestos y para los distintos valores de ventana h considerados.

- el comportamiento de los estimadores de g se evaluó utilizando dos medidas

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

$$MedSE(\hat{g}) = median \left([\hat{g}(\theta_i) - g(\theta_i)]^2 \right)$$

La Tabla 5.4 muestra el valor medio de MSE para los distintos estimadores de g , mientras que la Tabla 5.5 resume los valores medianos de $MedSE(\hat{g})$

Finalmente, generamos una muestra bajo el modelo y a modo ilustrativo exhibimos los gráficos de las curvas estimadas de la función g para distintos anchos de ventanas. Esto se puede ver en las Figuras 5.7, 5.8, 5.9.

5.2. Resultados

h	$\hat{\beta}_I$	$\hat{\beta}_R$	$\hat{\beta}_{IP}$	$\hat{\beta}_C$	$\hat{\beta}_{full}$
0,2	2,015	2,021	2,013	2,020	2,009
0,27	2,014	2,019	2,012	2,018	2,008
0,33	2,016	2,022	2,013	2,018	2,010
0,4	2,029	2,017	2,022	2,021	2,015
0,47	2,030	2,040	2,024	2,027	2,022
0,53	2,041	2,055	2,034	2,036	2,032
0,6	2,054	2,070	2,046	2,047	2,044
0,67	2,068	2,087	2,059	2,060	2,058
0,73	2,03	2,102	2,074	2,074	2,072
0,8	2,097	2,117	2,089	2,088	2,087

Tabla 5.1: Media de $\hat{\beta}_I$, $\hat{\beta}_R$, $\hat{\beta}_{IP}$, $\hat{\beta}_C$, y $\hat{\beta}_{full}$ para diferentes valores de h .

h	$\hat{\beta}_I$	$\hat{\beta}_R$	$\hat{\beta}_{IP}$	$\hat{\beta}_C$	$\hat{\beta}_{full}$
0,2	0,074	0,074	0,079	0,076	0,060
0,27	0,073	0,073	0,077	0,074	0,059
0,33	0,072	0,071	0,075	0,073	0,057
0,40	0,069	0,069	0,072	0,071	0,056
0,47	0,067	0,066	0,07	0,069	0,055
0,53	0,065	0,063	0,067	0,066	0,053
0,6	0,062	0,060	0,065	0,064	0,051
0,67	0,060	0,058	0,062	0,062	0,049
0,73	0,058	0,056	0,045	0,060	0,047
0,8	0,056	0,054	0,058	0,057	0,046

Tabla 5.2: Desvíos estándar de $\hat{\beta}_I$, $\hat{\beta}_R$, $\hat{\beta}_{IP}$, $\hat{\beta}_C$, y $\hat{\beta}_{full}$ para diferentes valores de h .

h	$\hat{\beta}_I$	$\hat{\beta}_R$	$\hat{\beta}_{IP}$	$\hat{\beta}_C$	$\hat{\beta}_{full}$
0,2	0,0058	0,0059	0,0064	0,0061	0,0037
0,27	0,0056	0,0057	0,0061	0,0059	0,0036
0,33	0,0054	0,0057	0,0058	0,0057	0,0034
0,4	0,0053	0,0056	0,0055	0,0055	0,0033
0,47	0,0054	0,0059	0,0055	0,0054	0,0035
0,53	0,0059	0,0069	0,0057	0,0057	0,0038
0,6	0,0068	0,0086	0,0063	0,0063	0,0046
0,67	0,0083	0,0108	0,007	0,0075	0,0057
0,73	0,010	0,013	0,009	0,009	0,0074
0,8	0,012	0,016	0,011	0,011	0,009

Tabla 5.3: MSE de $\hat{\beta}_I$, $\hat{\beta}_R$, $\hat{\beta}_{IP}$, $\hat{\beta}_C$, y $\hat{\beta}_{full}$ para diferentes valores de h .

Observación 5.2.1. De las tablas 5.1, 5.2 y 5.3 se observa que $\hat{\beta}_C$ no solo tiene medias, desvíos y MSE's aceptables, para ciertos valores de h , sino también similares a los estimadores imputados e incluso al estimador obtenido utilizando todos los datos. Esto es de esperar, pues dado un MPL

$$Y_i = \mathbf{X}_i\beta + g(\mathbf{T}_i) + \varepsilon \quad 1 \leq i \leq n$$

al perder datos, podemos pensar que tenemos la siguiente redefinición del modelo de la siguiente manera

$$\delta_i Y_i = \delta_i \mathbf{X}_i \beta + \delta_i g(\mathbf{T}_i) + \delta_i \varepsilon \quad 1 \leq i \leq n$$

donde podemos observar que al no variar β , la eficiencia del estimador $\hat{\beta}_C$ (5.1) depende únicamente del tamaño de la muestra de respuesta. Esto no ocurre con $\hat{\beta}_I$, pues $g(\mathbf{T}_i)$ puede variar significativamente de una observación a otra.

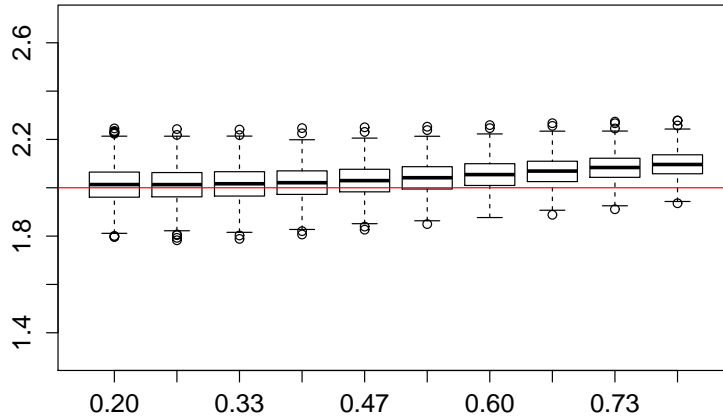


Figura 5.1: Boxplots del estimador $\hat{\beta}_I$ para los distintos valores de h

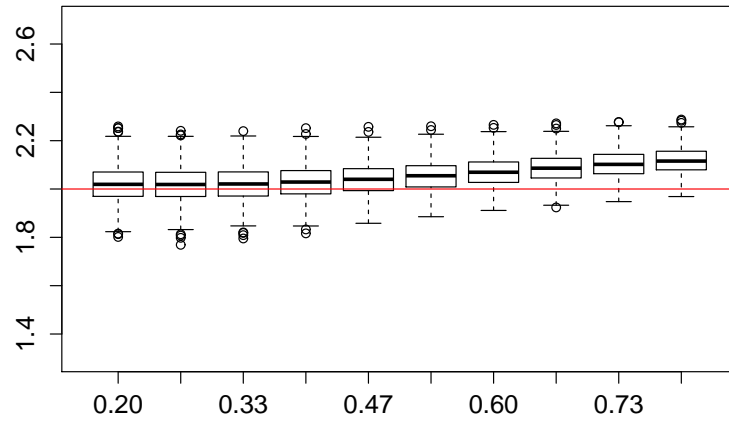


Figura 5.2: Boxplots del estimador $\hat{\beta}_R$ para los distintos valores de h

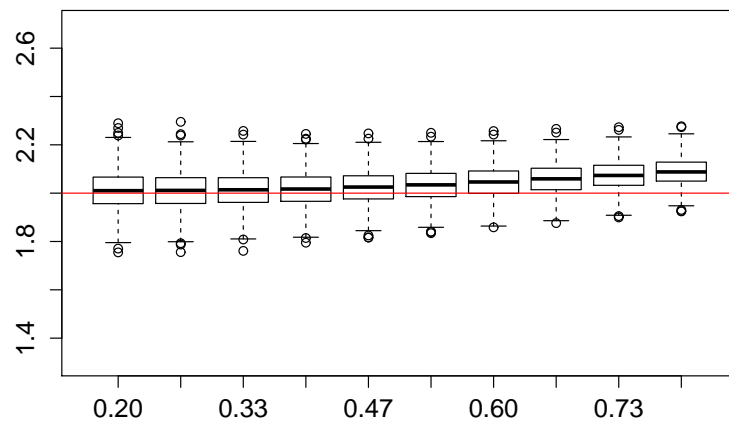


Figura 5.3: Boxplots del estimador $\hat{\beta}_{[IP]}$ para los distintos valores de h

h	\hat{g}_I	\hat{g}_R	\hat{g}_{IP}	\hat{g}_C	\hat{g}_{full}
0,2	0,09	0,085	0,128	0,108	0,068
0,27	0,079	0,073	0,10	0,09	0,056
0,33	0,071	0,068	0,087	0,080	0,049
0,40	0,067	0,068	0,077	0,074	0,045
0,47	0,066	0,073	0,071	0,07	0,045
0,53	0,071	0,085	0,070	0,070	0,047
0,6	0,079	0,100	0,073	0,074	0,054
0,67	0,092	0,122	0,08	0,082	0,064
0,73	0,108	0,144	0,094	0,095	0,079
0,8	0,12	0,16	0,11	0,11	0,097

Tabla 5.4: Valores medios de MSE de \hat{g}_I , \hat{g}_R , \hat{g}_{IP} , \hat{g}_C , y \hat{g}_{full} para diferentes valores de h .

h	\hat{g}_I	\hat{g}_R	\hat{g}_{IP}	\hat{g}_C	\hat{g}_{full}
0,2	0,031	0,029	0,034	0,035	0,021
0,27	0,027	0,026	0,029	0,03	0,018
0,33	0,025	0,0253	0,026	0,026	0,017
0,40	0,024	0,028	0,0253	0,025	0,016
0,47	0,029	0,029	0,0255	0,033	0,0182
0,53	0,032	0,041	0,028	0,03	0,021
0,6	0,040	0,0059	0,034	0,033	0,027
0,67	0,055	0,082	0,043	0,044	0,039
0,73	0,075	0,110	0,06	0,059	0,056
0,8	0,100	0,13	0,08	0,079	0,079

Tabla 5.5: Valores medianos de MedSE de \hat{g}_I , \hat{g}_R , \hat{g}_{IP} , \hat{g}_C , y \hat{g}_{full} para diferentes valores de h .

La tabla siguiente resume el comportamiento de diferentes estimadores para $E(Y)$, que fueron calculados en todos los diferentes escenarios promediando las observaciones completas ($\hat{E}(Y)_{full}$), solo las respuestas observadas ($\hat{E}(Y)_C$), y las pseudo-observaciones $U_{ni}^{[I]}$, $U_{ni}^{[R]}$, y $U_{ni}^{[IP]}$, ($\hat{E}(Y)_I$), ($\hat{E}(Y)_R$) y ($\hat{E}(Y)_{IP}$), respectivamente.

h	$\hat{E}(Y)_I$	$\hat{E}(Y)_R$	$\hat{E}(Y)_{IP}$	$\hat{E}(Y)_C$	$\hat{E}(Y)_{full}$
0,2	-5,037	-5,044	-5,031	-5,179	-5,031
0,27	-5,038	-5,047	-5,031	-5,181	-5,031
0,33	-5,040	-5,052	-5,031	-5,185	-5,031
0,4	-5,043	-5,057	-5,031	-5,190	-5,031
0,47	-5,045	-5,063	-5,031	-5,195	-5,031
0,53	-5,048	-5,068	-5,031	-5,199	-5,031

Tabla 5.6: Valores esperados de Y para diferentes valores de h .

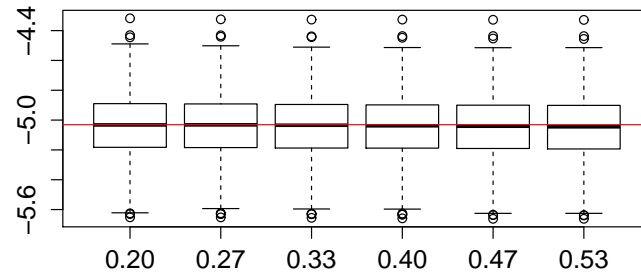


Figura 5.4: Boxplots de $\hat{E}(Y)_I$ para los distintos valores de h

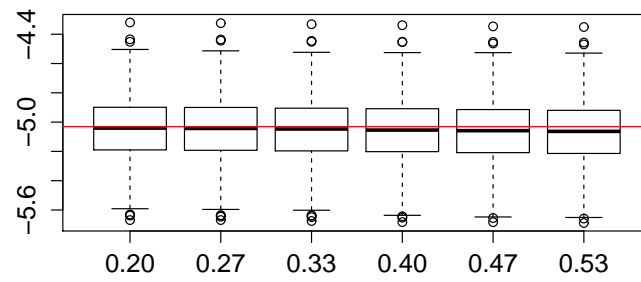


Figura 5.5: Boxplots de $\hat{E}(Y)_R$ para los distintos valores de h

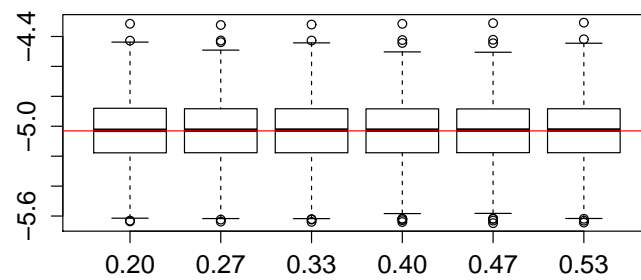


Figura 5.6: Boxplots de $\hat{E}(Y)_{IP}$ para los distintos valores de h

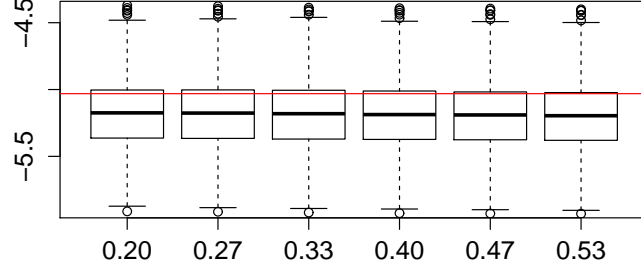


Figura 5.6: Boxplots de $\hat{E}(Y)_C$ para los distintos valores de h

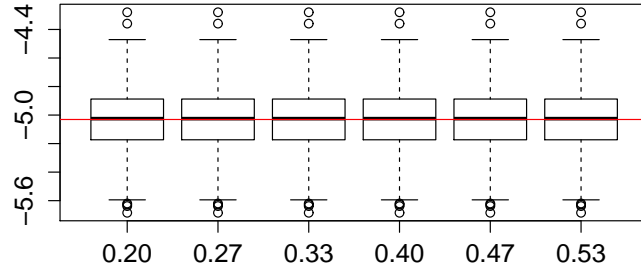


Figura 5.6: Boxplots de $\hat{E}(Y)_{full}$ para los distintos valores de h

Observación 5.2.2. En la Tabla 5.6 se puede observar el comportamiento diferente de $\hat{E}(Y)_C$, respecto al resto de las propuestas. Esto se deduce por lo siguiente: Por la ley de los grandes números y la propiedad MAR tenemos que

$$\begin{aligned} \hat{E}(Y)_C &= \sum_{i=1}^n \frac{Y_i \delta_i}{n} \rightarrow E(\delta Y) = E(E(\delta Y | Y, X, T)) = E(Y E(\delta | X, T)) \\ &= E(Y \Delta(Z)) \end{aligned}$$

Y por lo tanto, no hay razón teórica para suponer que $E(Y \Delta(Z))$ y $E(Y)$ coincidan. Sin embargo, como mostramos en el Capítulo 3, $E(U^{[I]} | \mathbf{Z}) = E(Y | \mathbf{Z})$ y por lo tanto $E(U^{[I]}) = E(E(U^{[I]} | \mathbf{Z})) = E(E(Y | \mathbf{Z})) = E(Y)$. Luego si consideramos el promedio de las observaciones $U^{[I]}$, este sí resultará un estimador consistente de $E(Y)$. Pero nuevamente como al estimar los parámetros del MPL, debemos considerar $U_{ni}^{[I]} = \delta_i Y_i + (1 - \delta_i)(\mathbf{X}^T \hat{\beta}_C + g_C(\mathbf{T}_i))$ en lugar de $U^{[I]}$. Naturalmente suponiendo consistencia de los estimadores de β y g podemos asumir por válida la siguiente convergencia a medida que el tamaño de la muestra aumenta

$U_{ni}^{[I]} \rightarrow U_i^{[I]} = \mathbf{X}_i^T \beta + g(\mathbf{T}_i)$ y por lo tanto se razonable que $\hat{E}(Y)_I = \sum_{i=1}^n \frac{U_{ni}^{[I]}}{n} \rightarrow E(Y)$.

De forma análoga se deduce que los estimadores $\hat{E}(Y)_R$ y $\hat{E}(Y)_{IP}$ basados en $U_{ni}^{[R]}$ y $U_{ni}^{[IP]}$, respectivamente resultan más adecuados para estimar $E(Y)$.

A continuación graficaremos los estimadores de $g(\theta) = \cos(\theta)$ para distintos valores de h . En todos los casos la curva negra corresponde a la g verdadera, la verde a la curva $y = \hat{g}_{full}(\theta)$ y la roja a $Y = \hat{g}_I(\theta)$, $Y = \hat{g}_R(\theta)$ y $Y = \hat{g}_{IP}(\theta)$ respectivamente en el gráfico que corresponda.

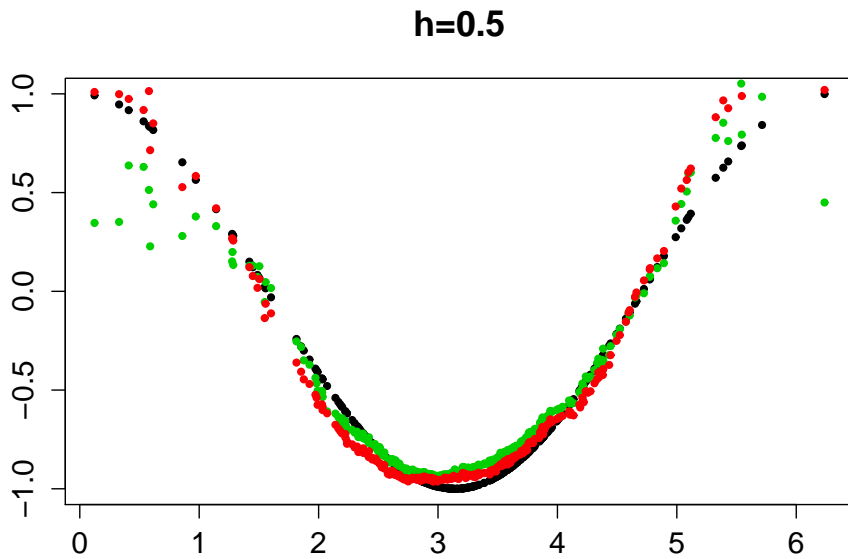


Figura 5.7: $Y = \hat{g}_I(\theta)$

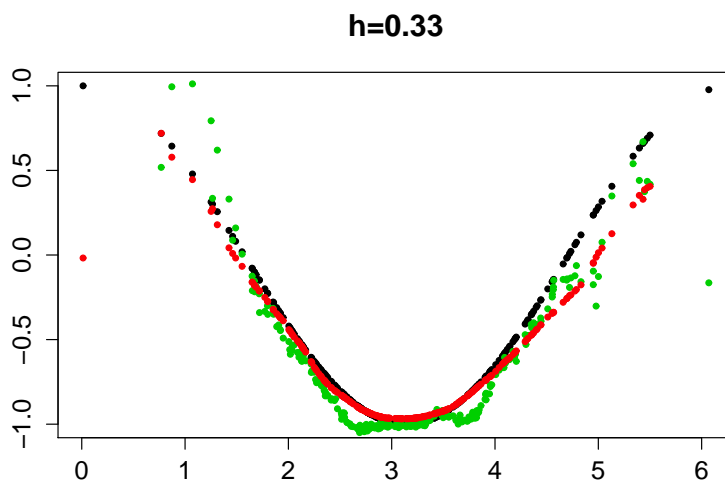


Figura 5.8: $Y = \hat{g}_R(\theta)$

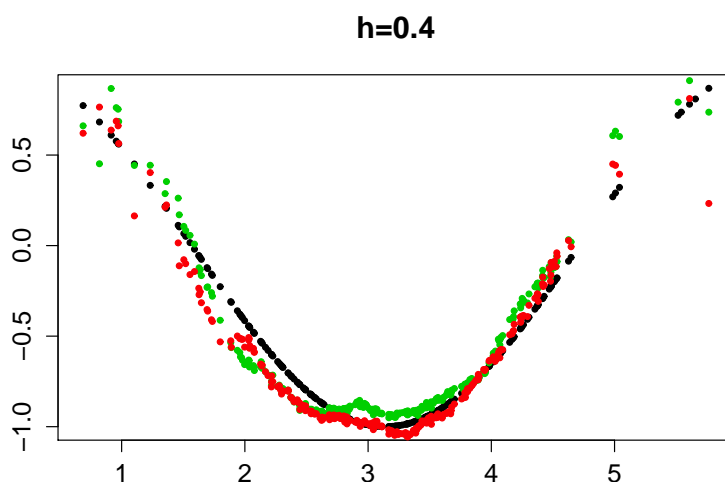


Figura 5.9: $Y = \hat{g}_{IP}(\theta)$

5.3. Conclusiones

De las tablas 5.1, 5.2 y 5.3 observamos que las medias, desvíos y MSE's de los estimadores propuestos para β son similares y aceptables para los valores de h menores a 0.5. Notamos que estos resultados son *ligeramente* más grandes que las medias, desvíos y MSE's correspondientes a $\hat{\beta}_{full}$ y *ligeramente* más chicos que las medidas de resumen que corresponden

a $\hat{\beta}_C$. Esto nos confirma un comportamiento adecuado de los estimadores imputados, $\hat{\beta}_I$, $\hat{\beta}_R$ y $\hat{\beta}_{IP}$.

Además notamos que generalmente para valores grandes de ventana los sesgos de los estimadores propuestos aumentan, pues las medias aumentan y las varianzas disminuyen pues los desvíos disminuyen. Con lo cual al igual que en el estimador de densidad tipo núcleo estudiado Capítulo 2, observamos un compromiso entre sesgo y varianza. Por otra parte los desvíos similares de $\hat{\beta}_I$, y de $\hat{\beta}_R$ que se muestran en la tabla 5.2 reflejan el mismo comportamiento asintótico de la varianza demostrado en la observación 3.3.2.

Las tablas 5.4 y 5.5, y figuras 5.4, 5.5 y 5.6 describen una performance aceptable para valores de h menores a 0.5, de los estimadores de g , \hat{g}_I , \hat{g}_R y \hat{g}_{IP} . Como era de esperar los valores medios de MSE y los valores medianos de Medse son apenas mayores a los de \hat{g}_{full}

Finalmente en la tabla 5.6 notamos que $\hat{E}(Y)_C$ se aleja de $\hat{E}(Y)_{full}$. Esto justifica el planteo de nuevos estimadores, para los cuales la estimación de los valores esperados resultan adecuados.

Capítulo 6

Un ejemplo real

La insolación es la cantidad de energía en forma de radiación solar que llega a un lugar de la Tierra en un día concreto o en un año . Esta variable mide la duración de la luz solar en cuestión de segundos. La Organización Meteorológica Mundial define la insolación como la suma de intervalos de tiempo en los que la irradiación supera el umbral de 120 watts por méetro cuadrado.

La irradiación es la radiación directa normal o perpendicular al sol sobre la superficie de la Tierra. Los valores de la radiación solar en un lugar en particular dependen de las condiciones climáticas y la posición del sol en el horizonte. Por ejemplo, la presencia de nubes aumenta la absorción, la reflexión y la dispersión de la radiación solar. Las zonas desérticas, dada la falta de nubes, tienen los más altos valores de insolación en el planeta.

Para ilustrar los estimadores propuestos, vamos a analizar la relación entre la insolación, la humedad, la radiación y la dirección del viento. Consideramos un conjunto de datos disponibles en <http://meteo.navarra.es/>. Estos datos contienen el promedio diario de la humedad relativa, la dirección del viento, la radiación y la insolación.

La dirección del viento se midió con el punto cero en la dirección norte. Los datos fueron medidos en la estación meteorológica automática de Pamplona-Larrabide GN, en Navarra, España durante el año 2004. En nuestro estudio, consideramos una muestra aleatoria de estos conjuntos de datos. Y generamos una variable aleatoria δ ficticia para generar un mecanismo de pérdida, con función distribución

$$\Delta(\Theta) = \begin{cases} 0,6 + \frac{e^{-\theta}}{1+e^{-\theta}} & \text{si } \frac{e^{-\theta}}{1+e^{-\theta}} \leq 0,1 \\ 0,5 & \text{si } \frac{e^{-\theta}}{1+e^{-\theta}} > 0,1 \end{cases}$$

En la figura 6.1, podemos ver que tanto la humedad y la insolación, como la radiación y la insolación tienen una relación lineal. En nuestro estudio consideramos un modelo parcialmente lineal para explicar la insolación, como una función lineal de la humedad y la radiación más una función noparamétrica de la dirección, una variable circular. En todos los procedimientos de suavizados utilizamos el núcleo

$$K(u) = \frac{6}{4}(1 - u^2)I_{(0,1)}(u),$$

y consideramos el estimador (4.7) como detallamos en el Capítulo 4.

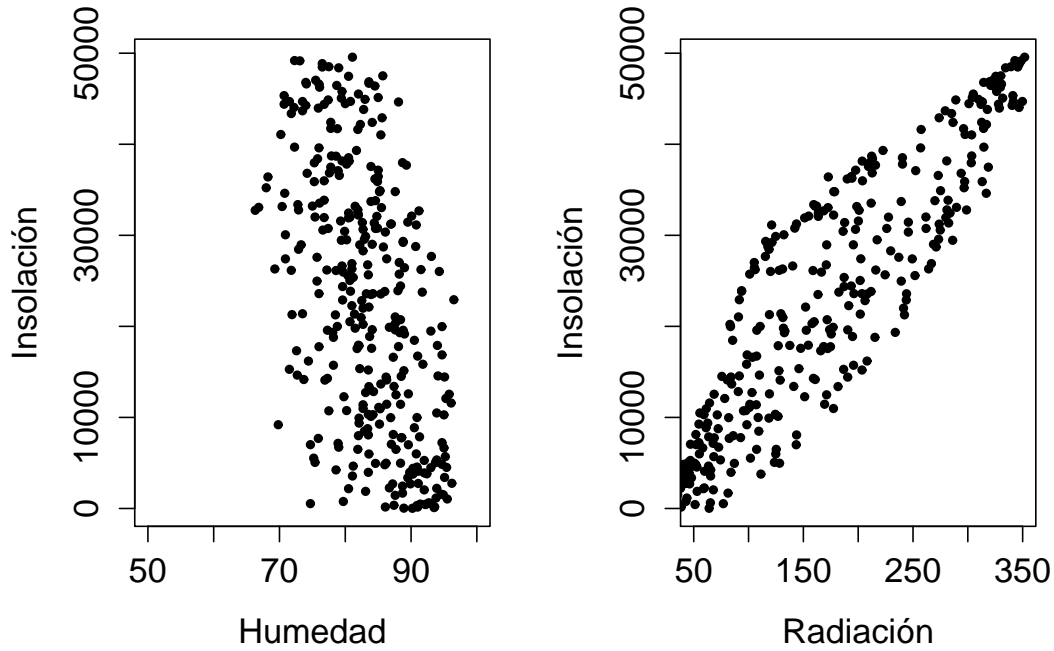


Figura 6.1: Gráficos de puntos

Los Resultados para $h=0.4$ y $n=331$ fueron los siguientes

- $\hat{\beta}_{1I} = -63,63$ y $\hat{\beta}_{2I} = 130,49$
- $\hat{\beta}_{1R} = -62,99$ y $\hat{\beta}_{2R} = 128,84$
- $\hat{\beta}_{1IP} = -64,11$ y $\hat{\beta}_{2IP} = 131,77$

donde β_1 corresponde al parámetro correspondiente a la humedad y β_2 al correspondiente radiación.

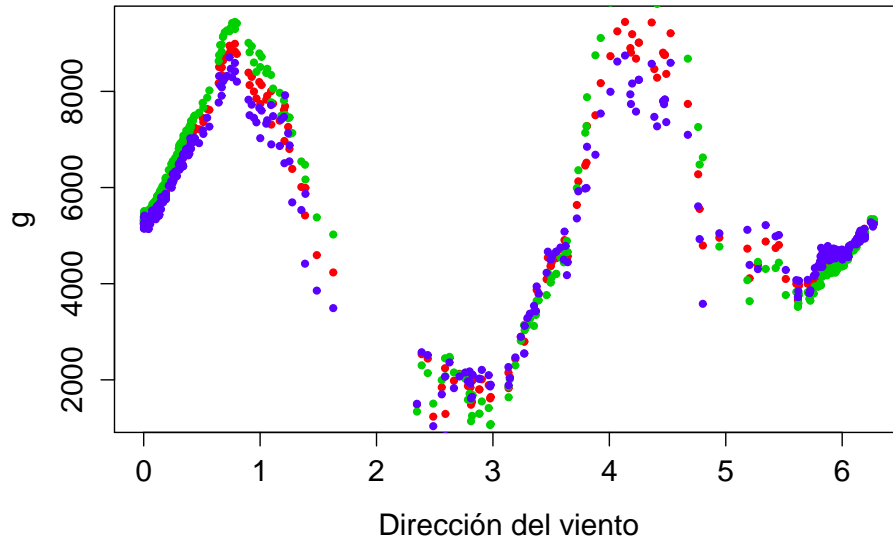


Figura 6.2: Estimación de la función noparamétrica. El estimador \hat{g}_I está representado por los puntos rojos, \hat{g}_R por los verdes y \hat{g}_{IP} por los puntos azules.

Por lo tanto, podemos ver que existe una relación lineal negativa entre la insolación y la humedad. Es decir, altos niveles de humedad derivarán en bajos niveles de insolación. Mientras que altos niveles de insolación estarán ligados a altos niveles de radiación. Por otro lado, cuando la dirección del viento se aproxima al intervalo $[2.49, 2.97]$ los niveles de insolación disminuyen hasta alcanzar su nivel más bajo.

Apéndice A

Apéndice

A continuación se demuestran proposiciones y lemas necesarios para obtener resultados de consistencia en los estimadores de densidad tipo núcleo y de regresión.

Proposición A.0.1. *Sea f una función de densidad $C^2(\mathbb{R})$ con $f''(\cdot)$ absolutamente continua tal que $\|f'''\|_{L^2} < \infty$ y K un núcleo univariado que satisface las hipótesis **H.1**, **H.2** y **H.3**, entonces si $h \rightarrow 0$ para cada x*

$$|\text{Sesgo}(\tilde{f}(x))| \leq \frac{h^2}{2} |f''(x)| \mu_2(K) + o(h^2)$$

donde $\mu_2(K) = \int s^2 K(s) ds$ y

$$\text{Var}(\tilde{f}(x)) \leq \frac{1}{nh} \|K\|_{L^2}^2 f(x) + o\left(\frac{1}{nh}\right).$$

Demostración:

Sesgo($\tilde{f}(x)$):

$$\begin{aligned} E[\tilde{f}(x)] &= E\left(\frac{1}{hn} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right)\right) = \frac{1}{n} \sum_{i=1}^n E\left(\frac{1}{h} K\left(\frac{x - X_i}{h}\right)\right) \\ &= E\left(\frac{1}{h} K\left(\frac{x - X_i}{h}\right)\right) = \frac{1}{h} \int K\left(\frac{x - u}{h}\right) f(u) du \end{aligned}$$

Mediante el cambio de variable $y = \frac{x-u}{h}$ y en vista de que $\int K(u) du = 1$ (**H1**) obtenemos que

$$E[\tilde{f}(x)] - f(x) = \int K(y)(f(x - hy) - f(x)) dy$$

Haciendo el desarrollo de Taylor de f de orden 2 alrededor de x y expresando el resto en la forma integral obtenemos la siguiente cota,

$$f(x - hy) - f(x) = -f'hy + \frac{f''(x)}{2} y^2 h^2 + \int_{x-yh}^x \frac{f'''(t)}{2!} (x - t)^2 dt.$$

Con lo cual,

$$\begin{aligned}
|E(\tilde{f}(x)) - f(x)| &= \left| \int K(y) \left[-f'(x)hy + \frac{f''(x)}{2}y^2h^2 + \int_{x-yh}^x \frac{f'''(t)}{2!}(x-t)^2 dt \right] dy \right| \\
&= \left| -f'(x)h \int K(y)y dy + h^2 \frac{f''(x)}{2} \int K(y)y^2 dy + \int \int_{x-yh}^x \frac{f'''(t)}{2!}(x-t)^2 dt K(y) dy \right| \\
&\leq \left| h^2 \frac{f''(x)}{2} \right| \int K(y)y^2 dy + \left| \int \frac{1}{2} (\|f'''\|_{L^2}) \left(\int_{x-yh}^x |x-t|^4 \right)^{\frac{1}{2}} K(y) dy \right| \\
&\leq \left| h^2 \frac{f''(x)}{2} \right| \int K(y)y^2 dy + \frac{\|f'''\|_{L^2}}{2!\sqrt{5}} h^{\frac{5}{2}} \int |y|^{\frac{5}{2}} K(y) dy \\
&\leq \left| h^2 \frac{f''(x)}{2} \right| \int K(y)y^2 dy + o(h^2)
\end{aligned}$$

Por lo tanto,

$$|\text{Sesgo}(\tilde{f}(x))| = h^2 \left| \frac{f''(x)}{2} \right| \int K(y)y^2 dy + o(h^2).$$

Analicemos $\text{Var}(\tilde{f}(x))$:

$$\text{Var}(\tilde{f}(x)) = \text{Var} \left(\frac{1}{nh} \sum_{i=1}^n K(x - X_i) \right) = \frac{1}{n^2 h^2} n \text{Var} \left(K \left(\frac{x - X_i}{h} \right) \right) \leq \frac{1}{nh^2} E \left(K^2 \left(\frac{x - X}{h} \right) \right)$$

Estudiemos $E \left(\frac{1}{h^2} K^2 \left(\frac{x - X}{h} \right) \right)$:

Haciendo el mismo cambio de variable tenemos que

$$\begin{aligned}
E \left(\frac{1}{h^2} K^2 \left(\frac{x - X}{h} \right) \right) &= \frac{1}{h} \int K^2(y) f(x - hy) dy \\
&\leq \frac{1}{h} \left| \int K^2(y) \left(f(x) - f'(x)hy + \frac{f''(x)}{2}h^2 y^2 dy \right) \right| + Ch^{\frac{5}{2}} \int K^2(y) |y|^{\frac{5}{2}} dy \\
&\leq \frac{1}{h} f(x) \int K^2(y) dy + \frac{1}{2h} |f''(x)| h^2 \int K^2(y) y^2 dy + \frac{1}{h} o(h^2) \int K^2(y) |y|^{\frac{5}{2}} dy.
\end{aligned}$$

Podemos acotar $\text{Var}(\tilde{f}(x))$ de la siguiente forma

$$\text{Var}(\tilde{f}(x)) \leq \frac{1}{nh} f(x) \int K^2(y) dy + o \left(\frac{1}{nh} \right)$$

□

Lema A.0.2. Sean X y T variables aleatorias continuas y K un núcleo univariado que cumplen las siguientes hipótesis **H.1**, **H.2** y **H.3**. Además se satisfacen

- $f_T(t) > 0$, $f(\cdot) \in C^2(\mathbb{R})$, con $f''(\cdot)$ absolutamente continua y $\|f'''\|_{L^2} < \infty$

- $m(\cdot) \in C^2(\mathbb{R})$, con $m''(\cdot)$ absolutamente continua y $\|m'''\|_{L^2} < \infty$
- $s^2(\cdot) = E(X^2|T = \cdot) \in C^2(\mathbb{R})$, con $(s^2)''(\cdot)$ absolutamente continua y $\|(s^2)'''\|_{L^2} < \infty$

Entonces,

$$ECM(\hat{m}_h(t)) \leq h^4 \frac{(m(t)'')^2}{4} \left(\int K(x)x^2 dx \right)^2 + o(h^4) + \frac{s^2(t)f_T(t)}{nh} \int K^2(x)dx + o\left(\frac{1}{nh}\right)$$

Demostración

Analicemos $|Sesgo(\hat{m}_h(t))| = |E(\hat{m}_h(t)) - m(t)|$

$$\begin{aligned} E\left(\frac{1}{nh} \sum_{i=1}^n K\left(\frac{t-T_i}{h}\right) X_i\right) &= \frac{1}{h} E\left(K\left(\frac{t-T}{h}\right) X\right) \\ &= \frac{1}{h} E\left(E(X|T) K\left(\frac{t-T}{h}\right)\right) \\ &= \frac{1}{h} E\left(\phi(T) K\left(\frac{t-T}{h}\right)\right) \\ &= \frac{1}{h} \int \phi(u) K\left(\frac{x-u}{h}\right) f_T(u) du \end{aligned}$$

Si hacemos el desarrollo de Taylor de la función $\phi(u)f_T(u) = m(u)$ alrededor de t como en la proposición A.0.1 obtenemos que

$$|E(\hat{m}_h(t)) - m(t)| = |Sesgo(\hat{m}_h(t))| \leq \frac{h^2}{2} |(m(t)'')| \int x^2 K(x) dx + o(h^2) \quad (\text{A.1})$$

Veamos $Var\left(\frac{1}{nh} \sum_{i=1}^n K\left(\frac{t-T_i}{h}\right) X_i\right)$:

$$\begin{aligned} Var\left(\frac{1}{nh} \sum_{i=1}^n K\left(\frac{t-T_i}{h}\right) X_i\right) &= \frac{1}{nh^2} Var\left(K\left(\frac{t-T_i}{h}\right) X\right) \\ &\leq E\left(\left(K\left(\frac{t-T}{h}\right) X\right)^2\right) \\ &\leq \frac{1}{nh^2} E\left(E\left(K\left(\frac{t-T}{h}\right)^2 X^2|T\right)\right) \\ &= \frac{1}{nh^2} E\left(E(Y^2|X) K\left(\frac{t-T}{h}\right)^2\right) \\ &= \frac{1}{nh^2} \int K\left(\frac{t-T}{h}\right)^2 s^2(u) f_T(u) du \\ &= \frac{1}{nh} f_T(t) s^2(t) \int K^2(s) ds + o\left(\frac{1}{nh}\right) \end{aligned}$$

Por lo tanto, deducimos que

$$\text{Var} \left(\frac{1}{nh} \sum_{i=1}^n K \left(\frac{t - T_i}{h} \right) X_i \right) \leq \frac{1}{nh} f_T(t) s^2(t) \int K^2(s) ds + o \left(\frac{1}{nh} \right) \quad (\text{A.2})$$

Finalmente, $ECM(\hat{m}(t))$ nos queda acotado por:

$$ECM(\hat{m}(t)) \leq \frac{h^4}{4} (m(t)'')^2 \left(\int x^2 K(x) dx \right)^2 + o(h^4) + \frac{1}{nh} f_T(t) s^2(t) \int K^2(s) ds + o \left(\frac{1}{nh} \right) \quad \square$$

Proposición A.0.3. *Bajo las misma hipótesis que el Lema A.0.2, $\hat{\phi}(t)$ resulta un estimador débilmente consistente de $\phi(t)$.*

Demostración: De (A.1) y (A.2) concluimos que el numerador converge en probabilidad a $m(t)$ si $n \rightarrow \infty$, $h \rightarrow 0$ y $nh \rightarrow \infty$. Observemos que en la proposición A.0.1 probamos que el denominador converge en probabilidad a $f_T(t)$ bajo las mismas hipótesis. Finalmente mediante las propiedades de la convergencia en probabilidad con respecto al cociente, siendo $f_T(t) > 0$, podemos asegurar que $\frac{\hat{m}_h(t)}{\hat{f}_T(t)}$ converge en probabilidad a $\phi(t)$ \square

Hipótesis del capítulo 2

A continuación listaremos las hipótesis necesarias para obtener los resultados mencionados en las observaciones 2.2.1 y 2.2.2.

Hipótesis A.0.1.

- i. $\sup_{0 \leq t \leq 1} E(\|X_1\|^3 | T) < \infty$.
- ii. $\Sigma = \text{Cov}(X_1 - E(X_1|T_1))$ es una matriz definida positiva.
- iii. Los errores ϵ_i , $1 \leq i \leq n$ son independientes de (X_i, T_i) .

Hipótesis A.0.2. *Las primeras dos derivadas de $g(\cdot)$ y $\phi = E[X|T]$ son Lipschitz continuas de orden uno.*

Hipótesis A.0.3. *La función de pesos $w_{ni}(\cdot)$ satisface con probabilidad*

- i. $\max_{1 \leq i \leq n} \sum_{j=1}^n \omega_{ni}(T_j) = O(1)$ y $\max_{1 \leq j \leq n} \sum_{i=1}^n \omega_{ni}(T_j) = O(1)$,
- ii. $\max_{1 \leq i, j \leq n} \omega_{ni}(T_j) = O(b_n)$,

$$\text{iii. } \max_{1 \leq i \leq n} \sum_{j=1}^n \omega_{nj}(T_j) I(|T_i - T_j| > c_n) = O(c_n),$$

donde b_n y c_n son dos secuencias que satisfacen $\limsup_{n \rightarrow \infty} nb_n^2 \log^4 n < \infty$, $\liminf_{n \rightarrow \infty} nc_n^2 > 0$, $\limsup_{n \rightarrow \infty} nc_n^4 \log n < \infty$ y $\limsup_{n \rightarrow \infty} nb_n^2 c_n^2 < \infty$.

Observación A.0.4. Nosotros basaremos nuestras estimaciones en la función de pesos

introducidas en el Capítulo 2, $w_{ni} = \frac{K\left(\frac{t-T_i}{h_n}\right)}{\sum_{j=1}^n K\left(\frac{t-T_j}{h_n}\right)}$ que satisface las condiciones A.0.3.

Pero, existen muchas elecciones posibles de funciones de pesos que satisfacen la hipótesis A.0.3.

Hipótesis del Capítulo 3.

(a) Sea $\tilde{X} = X - E[X|T]$ y $\hat{X} = X - \frac{E[\delta X|T]}{E[\delta|T]}$ se tiene que (i) $E[\tilde{X}\hat{X}^T]$ es una matriz definida positiva.

(ii) $E[\Delta(Z)\tilde{X}\tilde{X}^t]$ es una matriz definida positiva.

(b) (i) $\inf_t \Delta_t(T) > 0$.

(ii) $\Delta_t(\cdot)$ tiene derivadas parciales de orden 2 acotadas

(c) (i) $K(\cdot)$ es una función núcleo acotada de orden 2 con soporte acotado.

(ii) $M(\cdot)$ es una función núcleo acotada de orden 2 con soporte acotado.

(iii) $\Omega(\cdot)$ es un función núcleo acotada de orden 2 con soporte acotado.

(d) (i) $\phi(\cdot)$ y $\phi_0(\cdot)$ tiene derivadas acotadas de orden 2.

(ii) $\hat{\phi}_C(\cdot)$ y $\hat{\phi}_C^{[I]}(\cdot)$ tiene derivadas acotadas de orden 2.

(e) (i) $\sup_{x,t} E[Y^2|X=x, T=t] < \infty$,

(ii) $\sup_t E[\|X\|^2|T=t] < \infty$.

(f) La densidad de T, $f_T(T)$, existe y las derivadas de orden 2 están acotadas y satisface

$$0 < \inf_{t \in [0,1]} f_T(t) \leq \sup_{t \in [0,1]} f_T(t) < \infty.$$

(g) $nb_n h_n \rightarrow \infty$; $nh_n^4 \rightarrow 0$, $nb_n^4 \rightarrow 0$ y $\frac{h_n^2}{b_n} \rightarrow 0$.

(h) $n\gamma_n \rightarrow \infty$ y $n\gamma_n^4 \rightarrow 0$.

Demostración de los Teorema 3.1, 3.2 y 3.3 .

Teorema 3.1 Bajo las hipótesis que se detallaron anteriormente, si

$b_n = O_p(n^{-\frac{1}{3}})$ y $h_n = O_p(n^{-\frac{1}{3}})$ tenemos que

$$\hat{g}_n^{[I]}(t) - g(t) = O_p(n^{-\frac{1}{3}}).$$

Demostración Por definición de $\hat{g}_n(t)$, tenemos que

$$\begin{aligned} \hat{g}_n^{[I]}(t) - g(t) &= \hat{\phi}_0(t) - \phi_0(t) - (\hat{\phi}(t) - \phi(t))^T (\hat{\beta}_I - \beta) - \phi(t)^T (\hat{\beta}_I - \beta) \\ &\quad - (\hat{\phi}(t) - \phi(t))^T \beta \end{aligned} \quad (\text{A.3})$$

Primero analicemos $\hat{\phi}_0(t) - \phi_0(t)$

Usando que $\sum_{i=1}^n \omega_{ni}(t) = 1$ y que $U_i^{[I]} = \delta_i Y_i + (1 - \delta_i) (X_i^T \beta + g(T_i))$ obtenemos

$$\begin{aligned} \hat{\phi}(t) - \phi_0(t) &= \sum_{i=1}^n \omega_{ni} U_{ni}^{[I]} - \phi_0(t) \\ &= \sum_{i=1}^n \omega_{ni}(t) \left[\delta_i Y_i + (1 - \delta_i) (X_i^T \hat{\beta}_c + \hat{g}_C(T_i)) - \phi_0(t) \right] \\ &= \sum_{i=1}^n \omega_{ni}(t) (U_i^{[I]} - \phi_0(t)) + \sum_{i=1}^n \omega_{ni}(t) (1 - \delta_i) X_i^T (\hat{\beta}_C - \beta) \\ &\quad + \sum_{i=1}^n \omega_{ni}(t) (1 - \delta_i) (\hat{g}_C(T_i) - g(T_i)) \end{aligned}$$

Notemos que $E[U_i^{[I]} | T_i = t] = \phi_0(t)$ y $E[|(1 - \delta_i) X_i| | T_i] < \infty$. Luego por la teoría estándar de regresión por núcleos, ver por ejemplo [18] ó [19], sabemos que

$$\begin{aligned} \sup_t \left| \sum_{i=1}^n \omega_{ni}(t) (U_i^{[I]} - \phi_0(t)) \right| &= O_P(n b_n)^{\frac{1}{2}} + O_P(b_n) \\ \sup_t |\hat{g}_C(T_i) - g(T_i)| &= O_P\left((n h_n)^{-\frac{1}{2}}\right) + O_P(h_n) \\ \sup_t |\hat{\phi}(t) - \phi| &= O_P\left((n b_n)^{\frac{1}{2}}\right) + O_P(b_n) \end{aligned} \quad (\text{A.4})$$

y que $\sum_{i=1}^n \omega_{ni}(t) (1 - \delta_i) X_i = O_P(1)$ y $\sum_{i=1}^n \omega_{ni}(t) (1 - \delta_i) = O_P(1)$. Juntando todo y del hecho de que $\hat{\beta}_C - \beta = O_P(n^{-\frac{1}{2}})$ y $\hat{\beta}_I - \beta = O_P(n^{-\frac{1}{3}})$ se deduce que

$$\begin{aligned}
\sup_t \left| \hat{g}_n^{[I]}(t) - g(t) \right| &= O_P \left((nb_n)^{\frac{1}{2}} \right) + O_P \left((n)^{-\frac{1}{2}} \right) + O_P(n^{-\frac{1}{2}}) + O_P((nh_n)^{-\frac{1}{2}}) + O_P(h_n) \\
&+ \left[O_P((nb_n)^{-\frac{1}{2}}) + O_P(b_n) \right] O_P(n^{-\frac{1}{3}}) + O_P(n^{-\frac{1}{2}}) \\
&+ O_P((nb_n)^{-\frac{1}{2}}) + O_P(b_n) \\
&= O_P((nb_n)^{-\frac{1}{2}}) + O_P(b_n) + O_P((nh_n)^{-\frac{1}{2}}) + O_P(h_n)
\end{aligned}$$

Luego si $b_n = n^{-\frac{1}{3}}$ y $h_n = n^{-\frac{1}{3}}$ el teorema queda demostrado

□

El Teorema 3.2 y el 3.3 se demuestran de forma análoga.

Bibliografía

- [1] Bai, Z. D., Rao, R. and Zhao, L. C. (1988) Kernel Estimators of Density Function of Directional Data. *J. Multivariate Analysis*. **27**, 24-39.
- [2] Batschelet, E. (1981) Circular Statistics in Biology. *London: Academic Press*.
- [3] Di Marzio, M.; Panzera, A. and Taylor, C. (2009) Local polynomial regression for circular predictors. *Statistics & Probability Letters*, 79 (19). 2066–2075.
- [4] Fisher, N. I. (1993) Statistical Analysis of Circular Data. *Cambridge: Cambridge University Press*.
- [5] Hall, P. , Watson, G. S. and Cabrera, J. (1987) Kernel density estimation with spherical data. *Biometrika*. **74**, 751–762.
- [6] He X. (1992) Robust statistics of directional data: a survey. *Nonparametric Statistics and Related Topics. Amsterdam: North-Holand*. 87–95.
- [7] Mardia, K. V. and Jupp, P. E. (2000) Directional Statistics. John Wiley.
- [8] Matloff, N. (1981) Use of regression functions for improved estimation of means. *Biometrika*, **68**, 685–689.
- [9] Parzen, E. (1962) On estimation of a probability density function and mode. *Ann. Math. Statist.* **33**, 1065–1076.
- [10] Rao, J. S. (1984) Nonparametric methods in directional data analysis. *Handbook of Statistics, Vol. 4. Amsterdam/New York: Elsevier*. 757–770.
- [11] Robins, J.; Rotnitzky, A. and Zhao, L. (1994) Estimation of regression coefficients when some regressors are not always observed. *J. Amer. Statist. Assoc.*, **89**, 846–866.
- [12] Rosenblatt, M. (1956) Remarks on some nonparametric estimates of a density function. *Ann. Math. Statist.* **27**, 832–837
- [13] Wang, Q.; Lindon, O. and Härdle, W. (2004) Semiparametric regression analysis with missing response at random. *J. Amer. Statist. Assoc.* **99**, 334–345.
- [14] Wang, Q. and Sun, Z. (2007) Estimation in partially linear models with missing responses at random. *Journal of Multivariate Analysis*, **98**, 1470–1493.

- [15] Watson, G. S. (1983) *Statistics on Spheres*. *New York: Wiley*.
- [16] Zhao, L. and Wu, C. (2001) Central limit theorem for integrated square error of kernel estimators of spherical density. *Science in China (Serie A)*, **44**, 474-483.
- [17] ang, C.; Wang, S. and Zhao, S. (1997) Weighted semiparametric estimation in regression analysis regression with missing covariates data, *J. Amer. Statist. Assoc.* **92** 512–525.
- [18] Wang, Q., Lindon, O. and Härdle, W. (2004) Semiparametric regression analysis with missing response at random, *J. Amer. Statist. Assoc.* **99** 334–345.
- [19] Zhao, L.; Lipsitz, S. and Lew, D. (1996) Regression analysis with missing covariate data using estimating equations, *Biometrics* **52** 1165–1182.