

Capítulo 3

Modelo no paramétrico multidimensional

3.1. Introducción

En este capítulo presentaremos dos modelos no paramétricos multidimensionales. Para esto se utilizarán estimadores de regresión no paramétrica con variables independientes multidimensionales. Existen varias formas de encarar el problema de la regresión no paramétrica, los más populares son aquellos basados en núcleos, los basados en splines y los que utilizan wavelets. Cada uno de estos métodos tienen sus propias fortalezas y debilidades pero los que están basados en núcleos se caracterizan por su simplicidad matemática, además de tomar un formato intuitivo, como ser promedios ponderados locales.

En principio es natural pensar en aplicar un estimador semejante al que utilizó Ramsay para el caso unidimensional para ítems multidimensionales, con una dificultad adicional al tener que estimar un rasgo que ahora es un vector.

Consideremos entonces variables de respuestas dicotómicas que dependen de un rasgo multidimensional Θ . Es claro que las respuestas a un ítem condicionadas a un valor del rasgo son variables aleatorias Bernoulli, por lo que la probabilidad $p(\theta)$ es la esperanza condicional de la respuesta. Esta esperanza de la respuesta condicionada a θ es lo que conocemos por regresión. Por lo tanto es razonable pensar en estimar la CCI mediante la regresión no paramétrica. Hardle (1990) menciona que la estimación no paramétrica de las curvas de regresión tiene cuatro objetivos fundamentales: “Primero provee un método versátil para explorar la relación entre dos variables. Segundo, da predicciones de las observaciones sin hacer referencia a un modelo paramétrico fijo. Tercero, provee una herramienta para encontrar observaciones espúreas estudiando la influencia de puntos aislados. Cuarto, constituye un método flexible para sustituir valores perdidos o interpolar entre valores de X ”.

La idea general de la regresión no paramétrica es la del promedio local, es decir, estimar la esperanza condicional mediante un promedio ponderado, donde los pesos son una función que depende del punto donde se va a estimar y de los valores observados.

En este capítulo se presentará una familia de estimadores mediante polinomios locales con núcleos. Esta familia incluye al estimador de Ramsay (Ramsay 1991) en el caso unidimensional y su generalización multidimensional (Luzardo y Forteza 2014), y adicionalmente genera una gran variedad de posibilidades. El modelo de Ramsay y su generalización corresponderán al modelo polinomial de grado 0. Observaremos con atención al polinomio de grado 1 ya que en el caso de variables observables tiene mejores propiedades asintóticas y de frontera que los estimadores de

Nadaraya-Watson.

Lo primero a notar es que los rasgos se ignoran por lo tanto deben estimarse. Para ello se recurre a un procedimiento no paramétrico valiéndose de la función de distribución empírica de variables auxiliares. Se necesitan tantos estadísticos para ordenar a los sujetos como rasgos y esto presenta complicaciones técnicas que detallaremos en el capítulo de consistencia. En el caso unidimensional se puede utilizar el promedio o el puntaje total, pero en el caso multidimensional no es inmediato determinar cómo deben ser los estadísticos y que condiciones deben cumplir para que la estimación sea consistente.

Comenzaremos presentando el modelo indicando como se estimarán los rasgos y como se estimarán las curvas características del ítem y luego obtendremos tres tipos de resultados estadísticos para este modelo.

Un primer tipo de resultados es obtener estimadores para la distribución del rasgo concretamente de la densidad conjunta y la cópula subyacente. Creemos que este punto es muy interesante pues el método solo va a presuponer el conocimiento de las marginales (ni siquiera esto es necesario) y a partir de esto se puede reconstruir la distribución conjunta. Para lograr la demostración de la consistencia del estimador de la densidad y su distribución asintótica deberemos suponer que el rasgo es medido sin error aunque seguiremos suponiendo que los sujetos son elegidos aleatoriamente de la población. En el capítulo de simulaciones veremos que el estimador de la densidad conjunta funciona muy bien aun cuando el rasgo sea medido con error.

También presentamos la consistencia conjunta de la estimación del rasgo y de la CCI en el caso multidimensional encontrando las condiciones que deben cumplirse para que esta ocurra. También encontraremos una acotación exponencial para el error cuadrático medio. Esta sección la cual consideramos central, presenta dificultades técnicas importantes, debiéndose usar fuertemente las desigualdades de concentración de desarrollo reciente. Mencionamos que como producto adicional al investigar las condiciones antes dichas se logró definir una familia de funciones a las que llamaremos asintóticamente Liptchiz.

Finalmente para el caso de la habilidad medida sin error se deducirá la distribución asintótica del estimador de las CCI así como la consistencia completa, lo cual constituirá el tercer bloque de resultados.

3.2. Presentación del modelo multidimensional

Comencemos viendo la estructura general del modelo y como se estimará el rasgo, luego se mostrará como se estimaran las CCI, lo cual se basa en la regresión multivariada usando núcleos.

A los efectos de estimar cada componente del vector que representa un rasgo específico, utilizaremos estadísticos para ordenar a los sujetos y así tomar como estimación de dicha componente del rasgo en cada sujeto el cuartil empírico correspondiente. Expongamos primeramente el método y luego lo formularemos en forma precisa. Se definirá para cada rasgo un estadístico (al que notaremos $g_{n,l}$). Estos estadísticos dependerán solo de un rasgo, al cual estimarán, o sea son independientes de los restantes rasgos. En cada componente se ordenarán los puntajes obtenidos en el estadístico $g_{n,l}$ y luego se tomará el cuartil empírico para obtener el rasgo estimado del sujeto.

Formalmente consideramos ahora que el rasgo es un vector aleatorio en R^d

$$\Theta = (\Theta_1, \dots, \Theta_d).$$

Sea (Ω, A, P) un espacio de probabilidad y consideremos sin pérdida de generalidad que

el rasgo latente Θ tiene distribuciones marginales uniformes $U[0, 1]$. Es claro que todo lo que deduzcamos para este supuesto funcionará para cualquier tipo de distribución marginal.

Es interesante independizarse de la posible distribución del rasgo, ya que esta, usualmente, es desconocida.

Comencemos viendo esto en el caso unidimensional, para ello notemos por un momento al rasgo como τ y supongamos que tiene distribución F . Sabemos que $F(\tau)$ se distribuye uniforme $[0, 1]$. Por lo tanto, podríamos estimar en vez de τ una transformación monótona de ella $\theta = F(\tau)$.

Así pues, se puede suponer que el rasgo tiene distribución uniforme, y entonces, este rasgo se estima mediante la proporción empírica.

Si conocemos la distribución F del rasgo, obtenemos la escala original, mediante $\tau = F^{-1}(\theta)$, que tiene distribución F . En el caso multidimensional recurrimos a las cópulas.

Sea $(\Theta_1, \dots, \Theta_d)$ y supongamos conocidas sus marginales y que son continuas. El vector $(U_1, U_2, \dots, U_d) = (F_1(\Theta_1), \dots, F_d(\Theta_d))$ tiene marginales uniforme. La cópula de $(\Theta_1, \dots, \Theta_d)$ se define como

$$C(u_1, u_2, \dots, u_d) = \mathbb{P}[U_1 \leq u_1, U_2 \leq u_2, \dots, U_d \leq u_d] = P(\Theta_1 \leq F_1^{-1}(u_1), \dots, \Theta_d \leq F_d^{-1}(u_d))$$

Esta función contiene toda la información de dependencia del vector de rasgos. Aplicando el teorema de Sklar podemos reconstruir a partir de las marginales la función de distribución de los rasgos.

Si notamos por $H(\theta_1, \dots, \theta_d) = P[\Theta_1 \leq \theta_1, \dots, \Theta_d \leq \theta_d]$ la distribución conjunta del rasgo entonces $H(\theta_1, \dots, \theta_d) = C(F_1(\theta_1), \dots, F_d(\theta_d))$. Por lo que de aquí en más trabajaremos con marginales uniformes.

Consideremos una sucesión de funciones vectoriales, Borel medibles, \mathbf{g}_n en R^n a valores en $[0, 1]^d$. Sea $g_{n,l}$ la componente l -ésima de \mathbf{g}_n , para $x \in R^n$ $\mathbf{g}_n(x) = (g_{n,1}(x), \dots, g_{n,d}(x))$.

Consideremos un test de n ítems contestados por N sujetos y las variables aleatorias $X_{i,k}$, $i = 1, \dots, n$, $k = 1, \dots, N$ que indican la respuesta del k -ésimo sujeto al i -ésimo ítem. Estas variables dependen de d rasgos latentes (Θ_l) y sea $\mathbf{X} = (X_1, \dots, X_n)$ el vector aleatorio que indica las respuestas a los n ítems. Específicamente, \mathbf{X}_k indicará el vector de respuestas del sujeto k .

Supongamos que las funciones $g_{n,l}(X)$ son independientes de $\Theta_1, \dots, \Theta_{l-1}, \Theta_{l+1}, \dots, \Theta_n$ para todo $1 \leq l \leq n$. Definamos la sucesión de funciones $\{G_{g_n}\} : R^d \rightarrow R^d$ tal que para cada $x = (x_1, \dots, x_d)$ se tiene:

$$(3.1) \quad G_{g_n}(x) = (P(g_{n,1}(X) \leq x_1), \dots, P(g_{n,d}(X) \leq x_d))$$

Sean las distribuciones empíricas: $\hat{F}_{N,l}(x) = \frac{\#\{g_{n,l}(X) \leq x\}}{N} = \frac{\sum_{i=1}^N \chi_{\{g_{n,l}(X_i) \leq x\}}}{N}$ con $l = 1, \dots, d$ y definamos la función:

$$(3.2) \quad \hat{G}_N(x_1, \dots, x_d) = (\hat{F}_{N,1}(x_1), \dots, \hat{F}_{N,d}(x_d))$$

A partir de lo anterior, para estimar la componente l -ésima del rasgo se utiliza la función $g_{n,l}$ para ordenar los sujetos. Luego tomamos la función de distribución empírica, es decir,

$$(3.3) \quad \hat{\Theta}_n = \hat{G}_N(g_n(X))$$

Observemos que no depende este método de estimación del tipo de respuesta obtenido, pudiendo estas ser dicotómicas, politómicas o continuas. Si conocemos las marginales, podemos tener estimaciones en la escala original aplicando la inversa de cada una de estas distribuciones al rasgo estimado mediante la distribución uniforme.

3.2.1. Estimación de la CCI

Dado que $P(\theta_i) = E(X|\Theta = \theta_i)$ podemos considerar que el modelo subyacente es:

$$P(\theta_i) = m(\theta_i) + \sigma(\theta_i)\varepsilon_i \text{ con } i = 1, \dots, N$$

y donde los ε_i condicionados a $\Theta_1, \dots, \Theta_N$ son variables aleatorias independientes con media 0 y varianza 1. Supongamos que la densidad d-dimensional de Θ es $f(\theta)$ y que tenemos N sujetos con rasgos $\theta_1, \dots, \theta_N$. Sea H una matriz de dimension d, simétrica y no singular.

Se puede estimar esta densidad no paramétricamente mediante:

$$\hat{f}_{n,H}(\theta) = \frac{1}{N|H|} \sum_{i=1}^N K(H^{-1}(\theta_i - \theta)) = \frac{1}{N} \sum_{i=1}^N K_H(\theta_i - \theta)$$

donde $|H|$ es el determinante de H y $K_H(\theta_i - \theta) = \frac{1}{|H|} K(H^{-1}(\theta_i - \theta))$.

La matriz H contiene los parámetros de ancho de la ventana, es decir, controla cuán rápidamente los pesos decrecen a cero. Si queremos utilizar un solo ancho de ventana podemos tomar $H = hI_d$ donde I es la matriz identidad $d \times d$ en tanto si se quiere tener una ventana diferente para cada dimensión se puede tomar $H = \text{diag}(h_1, \dots, h_d)$.

Quedan en estos casos

$$\hat{f}_{n,h}(\theta) = \frac{1}{Nh^d} \sum_{i=1}^N K(h^{-1}(\theta_i - \theta))$$

$$\hat{f}_n(\theta) = \frac{1}{Nh_1 h_2 \dots h_d} \sum_{i=1}^N K(h_1^{-1}(\theta_{i1} - \theta_1), \dots, h_d^{-1}(\theta_{id} - \theta_d))$$

Supongamos que tenemos N respuestas a un ítem (cuya respuesta sin pérdida de generalidad supondremos continua) $(\theta_1, X_1), \dots, (\theta_N, X_N)$, que son independientes, idénticamente distribuidas con densidad conjunta $f_{\Theta,X}(\theta, x)$.

La densidad condicional de X dada Θ es: $f_{X|\Theta} = \frac{f_{\Theta,X}(\theta, x)}{f_{\Theta}(\theta)} = \frac{f_{\Theta,X}(\theta, x)}{\int f_{\Theta,X}(\theta, x) dx}$

Así tenemos que:

$$(3.4) \quad E(X | \Theta = \theta) = \frac{\int x f_{\Theta,X}(\theta, x) dx}{\int f_{\Theta,X}(\theta, x) dx}$$

Por lo tanto, para estimar la esperanza condicional utilizamos:

$$(3.5) \quad E(\widehat{X | \Theta} = \theta) = \frac{\int x \widehat{f_{\Theta,X}}(\theta, x) dx}{\int \widehat{f_{\Theta,X}}(\theta, x) dx}$$

Consideremos ahora un núcleo R^{d+1} para estimar la densidad condicional definido como $G_h(\cdot)K_H(\cdot)$ donde G es un núcleo unidimensional y K un núcleo de dimensión d.

El estimador por núcleos de la densidad conjunta es:

$$\widehat{f_{\Theta,X}}(\theta, x) = \frac{1}{N} \sum_{i=1}^N G_h(X_i - x) K_H(\theta_i - \theta)$$

realizando el cambio de variable, $-u = \frac{X_i - x}{h}$

$$N^{-1} \sum_{i=1}^N \int x \frac{1}{h} G\left(\frac{X_i - x}{h}\right) K_H(\theta_i - \theta) dx = N^{-1} \sum_{i=1}^N X_i K_H(\theta_i - \theta)$$

Además,

$$\int \widehat{f_{\Theta,X}}(\theta, x) dx = N^{-1} \sum_{i=1}^N K_H(\theta_i - \theta)$$

Obtenemos finalmente como estimador:

$$(3.6) \quad E(\widehat{X | \Theta} = \theta) = \frac{N^{-1} \sum_{i=1}^N X_i K_H(\theta_i - \theta)}{N^{-1} \sum_{i=1}^N K_H(\theta_i - \theta)} = \frac{\sum_{i=1}^N X_i K_H(\theta_i - \theta)}{\sum_{i=1}^N K_H(\theta_i - \theta)}$$

Notemos que el numerador es la convolución de K con los datos y el denominador es la convolución de K con la unidad. Se puede mostrar que si las respuestas son dicotómicas el estimador se mantiene.

Este es un estimador del tipo de promedio local ponderado con pesos, $w_i(x) = \frac{K_H(\theta_i - \theta)}{\sum_{i=1}^N K_H(\theta_i - \theta)}$. Es decir, es un promedio ponderado de las respuestas de aquellos cuya θ_i vive en una bola o cubo alrededor de θ .

Es claro que en el caso general, se ignoran los valores de $\theta_1 \dots \theta_N$. Pero si se tiene una sucesión de estimadores $\hat{\theta}_{n,i}$ que converge a θ_i para todo $i = 1, \dots, N$ en probabilidad, casi seguramente o completamente es inmediato pensar en sustituir θ_i por $\hat{\theta}_i$. Probaremos luego que esto se puede hacer con los estimadores del rasgo definidos en la sección anterior.

Por lo tanto, un estimador de la CCI del ítem j es:

$$(3.7) \quad \widehat{P_j}(\theta) = \frac{\sum_{i=1}^N X_{i,j} K_H(\hat{\theta}_i - \theta)}{\sum_{i=1}^N K_H(\hat{\theta}_i - \theta)}$$

Es inmediato verificar que el estimador encontrado es la solución de

$$(3.8) \quad \min_{\beta_0} \sum_{i=1}^N (X_{i,j} - \beta_0)^2 K_H(\hat{\theta}_i - \theta)$$

O sea, es un estimador localmente constante. Reemplazando β_0 por un polinomio podemos encontrar estimadores de regresión basados en núcleos del tipo polinomios locales. Un caso importante es el polinomio de grado 1 el cual es la solución de:

$$(3.9) \quad \min_{\beta_0, \beta_1} \sum_{i=1}^N (X_{i,j} - \beta_0 - (\hat{\theta}_i - \theta)^t \beta_1)^2 K_H(\hat{\theta}_i - \theta)$$

Usando mínimos cuadrados ponderados la solución queda:

$$(3.10) \quad \hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1^t)^t = (\Sigma^t W \Sigma)^{-1} \Sigma^t W X \text{ donde}$$

$$\Sigma = \begin{pmatrix} 1 & (\hat{\theta}_1 - \theta)^t \\ \vdots & \vdots \\ 1 & (\hat{\theta}_N - \theta)^t \end{pmatrix}, X = \begin{pmatrix} X_{1,j} \\ \vdots \\ X_{N,j} \end{pmatrix}$$

$$W = \begin{pmatrix} K_H(\hat{\theta}_1 - \theta) & 0 & \dots & 0 \\ 0 & K_H(\hat{\theta}_2 - \theta) & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & \dots & \dots & K_H(\hat{\theta}_N - \theta) \end{pmatrix}$$

En este caso β_0 es el estimador de la CCI y β_1 las derivadas parciales con respecto a las componentes de θ . Al estimador de la CCI obtenido mediante polinomios de grado 1 lo notaremos como

$$\widehat{P_j^1}(\theta) = \hat{\beta}_0$$

Utilizando el vector $e_1 = (1, 0, 0, \dots, 0)$ de la base canónica entonces

$$(3.11) \quad \widehat{P_j^1}(\theta) = e_1^t (\hat{\beta}_0, \hat{\beta}_1^t)^t = (\Sigma^t W \Sigma)^{-1} \Sigma^t W X$$

Hemos entonces obtenido dos estimadores de las CCI dados por las ecuaciones 3.7 y 3.11

3.2.2. Respuestas politómicas y continuas

Si las respuestas son politómicas podemos aplicar el método generando tantas variables dicotómicas como opciones tiene el ítem. Es decir X_{ij}^m valdrá 1 si el sujeto i en el ítem j eligió la opción m . Luego se aplica el estimador.

Si tenemos un modelo de respuesta graduada podemos considerar variables aleatorias dicotómicas como: $Y_{ij}^m = \chi_{(X_{ij} \leq m)}$, luego aplicar el modelo y calcular:

$$(3.12) \quad P(X = m | (\theta_1, \dots, \theta_d)) = P(X \leq m) - P(X \leq (m - 1))$$

Si hay respuestas continuas entonces podemos definir para un $x \in R$ variables $Y_{ij} = \chi_{(X_{ij} \leq x)}$. Con estas variables podemos estimar la distribución condicionada al rasgo de las respuestas.

También en el caso continuo se puede estimar la esperanza condicional de la respuesta, usando el estimador con la respuesta original (es decir no convirtiéndola en una indicatriz para cada x).

3.3. Estimación de la densidad conjunta del rasgo

Surge como primer objetivo estimar la densidad conjunta de las habilidades a partir de las habilidades estimadas. Es claro que el problema que se presenta pretende contestar lo siguiente: Si $\Theta = \hat{\Theta} + \epsilon$ como hallar la densidad de Θ a partir de la densidad de $\hat{\Theta}$. Hay trabajos en este sentido, en especial un artículo de Sperlich (2009). Sin embargo presupone conocimiento sobre el error de estimación elemento del que carecemos. En el apartado de limitaciones trataremos nuevamente este tema. Ahora lo abordaremos desde la perspectiva simplificada que la habilidad es estimada sin error, o que si estimamos la densidad de $\hat{\Theta}$ esta se encontrará cercana a la densidad de Θ . Como en este capítulo demostraremos la consistencia de la estimación del rasgo esta es una aproximación aceptable, aunque es claro que no estamos tomando en cuenta un aumento del sesgo y la varianza del error.

La estructura puede ser estimada tanto con los rasgos estimados en el cubo $[0, 1]^d$ o ya transformados mediante la inversa de las marginales. En lo que sigue supongamos que F_i , $i=1, \dots, d$ son las marginales y que Θ_{ij} $i=1, \dots, d$ $j=1, \dots, N$ ha sido obtenido aplicando F_i^{-1} al estimado del rasgo a partir del cuartil empírico.

Es claro que solo tiene sentido práctico aplicar la estimación no paramétrica de la densidad del rasgo en el caso de dimensión mayor o igual a dos, pues es en este caso que solo suponemos el conocer las marginales (aunque solo se supongan uniformes) en tanto en el caso unidimensional se supone conocida la distribución. Igualmente comenzaremos presentando el estimador unidimensional por la simplicidad de los cálculos para luego generalizar.

La gráfica que sigue muestra la proximidad de la densidad real y la estimada usando los rasgos estimados, esto nos indica que la aproximación usada es adecuada.

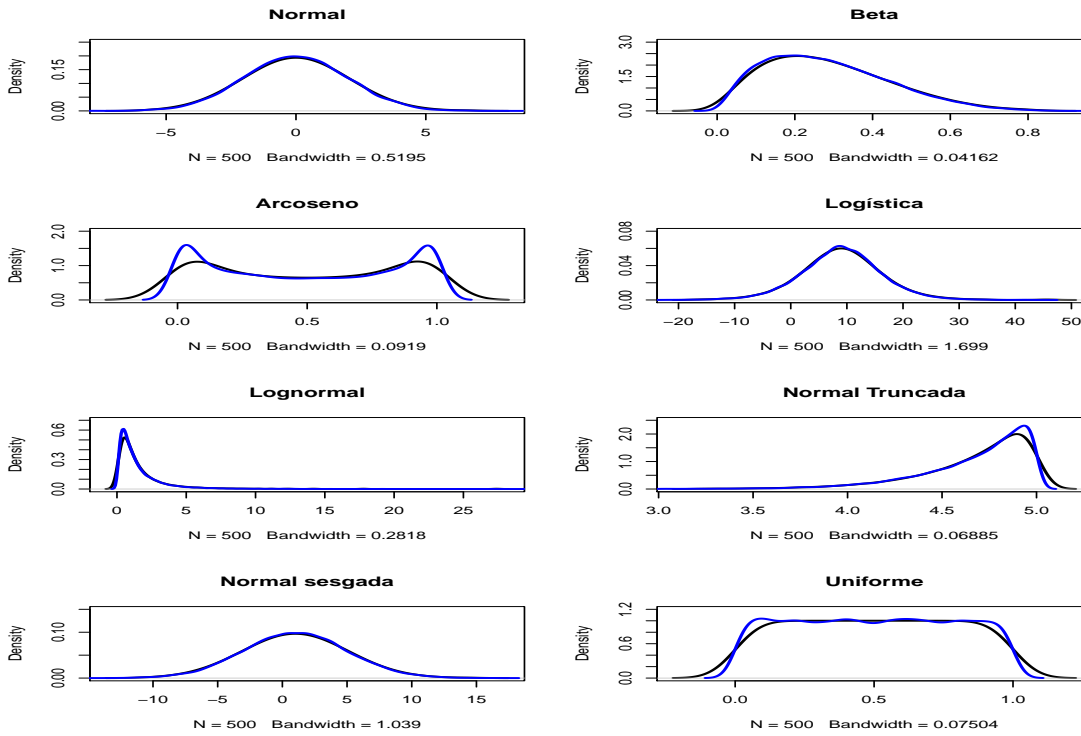


Figura 3.1: Proximidad de la densidad real y estimada por tipo de distribución

Caso unidimensional

Vamos a usar el estimador usual de la densidad por núcleos pero reemplazando la variable independiente por el estimador obtenido recién.

Supongamos que nuestros N sujetos constituyen una muestra del rasgo, o sea $\Theta_1, \dots, \Theta_N$ son variables aleatorias independientes idénticamente distribuidas y estas variables tienen una densidad común $f(\theta)$. En este caso consideramos que Θ está medido sin error por $\hat{\Theta}$

Nuestro objetivo es, a partir de esta muestra, estimar $f(\theta)$. Para ello consideremos la familia \mathfrak{F} de las funciones de distribución absolutamente continuas con densidad f la cual queremos encontrar.

Recordemos que $\hat{f}_{N,h}(\theta)$ se reduce a una convolución $\hat{f}_{N,h} = \bar{K}_h * F_N$ donde $\bar{K}_h(z)$ es un núcleo. El estimador toma la forma

$$(3.13) \quad \hat{f}_{N,h}(\theta) = \frac{1}{Nh} \sum_{i=1}^N K\left(\frac{\theta - \hat{\theta}_i}{h}\right)$$

Como el estimador de la densidad es la convolución de un núcleo con la distribución empírica, para demostrar la consistencia del estimador es clave obtener resultados sobre la convolución de un núcleo con una función. El siguiente lema nos da condiciones para que la convolución converja puntualmente cuando la ventana tiende a 0. Este resultado será luego usado en el teorema 3.3.2

Lema 3.3.1. Sea K una función real que cumple:

- i) K es acotada y simétrica
- ii) $\int |K(t)| dt = C_1 < \infty$
- iii) $\lim_{\theta \rightarrow \pm\infty} |\theta K(\theta)| = 0$

Sea g una función real tal que $\int |g(t)| dt = C_2 < \infty$ y consideremos la convolución:

$$g_h(\theta) = K_h * g(\theta) = \frac{1}{h} \int K\left(\frac{\theta-t}{h}\right) g(t) dt \text{ donde } K_h(z) = \frac{1}{h} K(z/h) \text{ entonces}$$

$$g_h(\theta) \longrightarrow g(\theta) \int K(t) dt \text{ si } h \rightarrow 0 \text{ para todo punto } \theta \text{ de continuidad de } g$$

Demostración: Sea θ un punto de continuidad de g . Dado $\epsilon > 0$ existe $\delta > 0$ tal que

$$|g(u) - g(\theta)| < \frac{\epsilon}{3C_1} \text{ si } |\theta - u| < \delta.$$

Como K es simétrico: $|g_h(\theta) - g(\theta) \int K(t) dt| = \left| \frac{1}{h} \int K\left(\frac{\theta-t}{h}\right) g(t) dt - g(\theta) \int K(t) dt \right| =$
 $\left| \int K(-u) g(\theta + uh) du - g(\theta) \int K(u) du \right| \leq \int |K(u)| |g(\theta + uh) - g(\theta)| du$. Ahora

$$\int |K(u)| |g(\theta + uh) - g(\theta)| du < \frac{\epsilon}{3C_1} \int_{|u| < \delta/h} |K(u)| du +$$

$$\int_{|u| \geq \delta/h} |K(u)| |g(\theta)| du + \int_{|u| \geq \delta/h} |K(u)| |g(\theta + uh)| du$$

Sabemos que θ es un punto de continuidad de g , entonces $|g(\theta)| < C_3$, además

$$\int_{|u| \geq \delta/h} |K(u)| du < \frac{\epsilon}{3C_3} \text{ pues es la cola de una integral convergente. Por último,}$$

$$\int_{|u| \geq \delta/h} |K(u)| |g(\theta + uh)| du \leq \frac{h}{\delta} \int_{|u| \geq \delta/h} |uK(u)| |g(\theta + uh)| du \leq$$

$$\frac{h}{\delta} \frac{\epsilon \delta}{3C_2} \int_{|u| \geq \delta/h} |g(\theta + uh)| du \leq \frac{\epsilon}{3C_2} \int |g(t)| dt = \frac{\epsilon}{3} \text{ pues } |uK(u)| < \frac{\epsilon \delta}{3C_2}$$

para h suficientemente chico.

A partir de estas tres desigualdades obtenemos que $|g_h(\theta) - g(\theta) \int K(t) dt| < \epsilon$ para $h \rightarrow 0$

Veremos ahora un teorema que nos garantiza la consistencia del estimador de la densidad.

Para ello se probará que $E(f_N(\theta)) \xrightarrow{h} f(\theta)$ y $Var(f_N(\theta)) \rightarrow 0$ lo cual garantiza la consistencia

o sea que $f_N(\theta) \xrightarrow{P} f(\theta)$.

Teorema 3.3.2. Supongamos que el rasgo es medido sin error o sea $\widehat{\Theta}_i = \Theta_i$ para todo i . Sea θ un punto de continuidad de f , K un núcleo simétrico y acotado y tal que $\int K(t) dt = 1$. Además $|\theta K(\theta)| \rightarrow 0$ si $|\theta| \rightarrow +\infty$. Entonces

$$f_N(\theta) = \frac{1}{Nh} \sum_{i=1}^N K\left(\frac{\theta - \hat{\Theta}_i}{h}\right) = \frac{1}{Nh} \sum_{i=1}^N K\left(\frac{\theta - \Theta_i}{h}\right)$$

es un estimador asintóticamente insesgado de $f(\theta)$ si $h \rightarrow 0$. Si además $Nh \rightarrow \infty$ es estimador consistente de $f(\theta)$.

Demostración: Por el lema 3.3.1 $E(f_N(\theta)) = \frac{1}{h} \int K\left(\frac{\theta-t}{h}\right) f(t) dt \xrightarrow{h} f(\theta)$

$$Var(f_N(\theta)) = \frac{1}{Nh^2} Var(K\left(\frac{\theta - X_1}{h}\right)) \leq \frac{1}{Nh^2} E(K^2\left(\frac{\theta - X_1}{h}\right)) = \frac{1}{Nh} \frac{1}{h} \int K^2\left(\frac{\theta-t}{h}\right) f(t) dt \rightarrow \frac{f(\theta)}{Nh} \int K^2(t) dt$$

Esta convergencia es válida ya que se cumple $|\theta K(\theta)| \rightarrow 0$ si $|\theta| \rightarrow +\infty$, entonces, $|\theta^2 K^2(\theta)| \rightarrow 0$ si $|\theta| \rightarrow +\infty$, lo que implica que $|\theta K^2(\theta)| \rightarrow 0$ si $|\theta| \rightarrow +\infty$, pues $|\theta K^2(\theta)| \leq |\theta^2 K^2(\theta)|$ para $|\theta| > 1$. Además $\int |K^2(t)| dt \leq \|K\|_\infty \int |K(t)| dt < \infty$. Entonces

si $Nh \rightarrow \infty$ $Var(f_N(\theta)) \rightarrow 0$. Hemos demostrado que si $Nh \rightarrow \infty$ y $h \rightarrow 0$, entonces,

$$f_N(\theta) \xrightarrow{P} f(\theta)$$

Observación: En realidad hemos probado la convergencia en L^2 , es decir que $E(f_N(\theta) - f(\theta))^2 \rightarrow 0$

Observación: el orden de la varianza es $\frac{1}{Nh}$, pues,

$$Var(\sqrt{Nh}f_N(\theta)) = NhVar(f_N(\theta)) = \frac{1}{h}Var(K(\frac{\theta-X_1}{h})) = \frac{1}{h}E(K^2(\frac{\theta-X_1}{h})) - \frac{1}{h}E^2(K(\frac{\theta-X_1}{h}))$$

luego,

$$Var(\sqrt{Nh}f_N(\theta)) \rightarrow f(\theta) \int K^2(u)du \text{ si } h \rightarrow 0 \text{ ya que}$$

$$\frac{1}{h^2}E^2(K(\frac{\theta-X_1}{h})) \rightarrow (f(\theta) \int K(u)du)^2 \text{ y } \frac{1}{h}E(K^2(\frac{\theta-X_1}{h})) \rightarrow f(\theta) \int K^2(u)du$$

de donde $\frac{Var(f_N(\theta))}{1/Nh} \rightarrow C$

Las proposiciones 3.3.3 y 3.3.4 que siguen dan aproximaciones al error cuadrático medio (ECM) calculando el sesgo y la varianza y a partir de esto se obtiene la ventana óptima en cada caso. La proposición 3.3.3 es más exigente que la 3.3.4.

Proposición 3.3.3. Sea θ un punto de continuidad de f y que K cumple las hipótesis del lema 3.3.1. Además suponemos que:

- i) existe f'' y $\sup|f''| < M$ y que el núcleo cumple:
- ii) $\int uK(u)du = 0$ y
- iii) $\int u^2K(u)du = a < \infty$

Entonces la ventana óptima cumple: $h \cong C_\theta N^{-1/5}$ y $ECM \cong CN^{4/5}$

Demostración: Sabemos que $ECM = B^2 + Var$ donde B indica el sesgo. El sesgo es:

$$B(f_N(\theta)) = E(f_N(\theta)) - f(\theta) = \frac{1}{h} \int K(\frac{\theta-t}{h})f(t)dt - f(\theta) =$$

$$\int K(u)(f(\theta + hu) - f(\theta))du = \int K(u) \left[f'(\theta)hu + f''(\xi)\frac{h^2u^2}{2} \right] du = \frac{h^2}{2} \int u^2K(u)f''(\xi)du \leq \frac{Mah^2}{2}$$

Si tomamos $f(\theta) = \theta^2$ entonces la desigualdad se transforma en igualdad, y $B_N(f_N(\theta)) \cong Ch^2$

Tenemos pues que $ECM = \frac{C_1}{Nh} + C_2h^4 + o(\frac{1}{Nh} + h^4)$ entonces el h óptimo sería $h \cong CN^{-1/5}$ y

$$ECM \cong CN^{-4/5}$$

.

Proposición 3.3.4. Consideremos que θ es un punto de continuidad de f y K cumple las hipótesis de los lemas anteriores. Además

- i) f es Lipchitz de constante L , es decir pido menos regularidad que en los lemas anteriores.
- ii) $\int |K(u)u|du = a_3 < \infty$

Entonces la ventana óptima será $h = CN^{-1/2}$ y el $ECM(f_N(\theta)) \cong CN^{-2/3}$

Demostración: A partir de las hipótesis ya sabemos que:

$$Var(\sqrt{Nh}f_N(\theta)) \rightarrow f(\theta) \int K^2(t)dt$$

$$\begin{aligned} \text{Ya que } \int K = 1 \text{ el sesgo es: } |B(f_N(\theta))| &= |E(f_N(\theta)) - f(\theta)| = \left| \frac{1}{h} \int K(\frac{\theta-\mu}{h})f(u)du - f(\theta) \right| \leq \\ &\leq \int |K(t)||f(\theta + th) - f(\theta)|dt \leq L \int |th||K(t)|dt = Lh \int |tK(t)|dt = La_3h \end{aligned}$$

así pues $|B(f_N(\theta))| \leq La_3h$

Si utilizo una función lineal $f(t) = LtI_{[\theta-c, \theta+c]}(t)$ obtengo la igualdad, entonces:

$$B(f_N(\theta)) \cong hC$$

luego $ECM = \frac{C_1}{Nh} + C_2h^2 + o(\frac{1}{Nh} + h^2)$ entonces el orden óptimo es: $\frac{1}{Nh} \cong h^2$ de donde $h = CN^{-1/3}$ así pues

$$ECM(f_N(\theta)) \cong CN^{-2/3}$$

Observación: A menor regularidad se obtiene menor velocidad de convergencia. Esto no tiene nada que ver con la parte aleatoria del problema.

Observemos que al sesgo le conviene que la ventana sea muy chica en tanto a la varianza que sea grande, es decir, existe un compromiso permanente entre sesgo y varianza.

3.3.1. Caso del rasgo multidimensional

La forma general del estimador de la densidad del rasgo multidimensional se obtiene como :

$$(3.14) \quad \hat{f}_{N,H}(\theta) = \frac{1}{N} \sum_{i=1}^N K_H(\theta - \hat{\Theta}_i) \text{ donde } K_H(\theta) = |H|^{-1} K(H^{-1}\theta)$$

y H es una matriz $d \times d$ simétrica definida positiva llamada matriz de ancho de ventanas o de banda y K es un núcleo de dimension d . La matriz H tiene $d(d+1)/2$ coeficientes independientes lo cual para dimensiones altas incrementa sustancialmente el número de parámetros. Hay dos casos particulares que se obtienen tomando $H = \text{diag}(h_1, \dots, h_d)$ y $H = hI$ con I la matriz identidad $d \times d$. En el primer caso se obtiene el estimador

$$\hat{f}_{N,H}(\theta) = \frac{1}{N \prod_{j=1}^d h_j} \sum_{i=1}^N K\left(\frac{\theta_1 - \Theta_{i1}}{h_1}, \dots, \frac{\theta_d - \Theta_{id}}{h_d}\right)$$

En el segundo caso queda:

$$\hat{f}_{N,H}(\theta) = \frac{1}{Nh^d} \sum_{i=1}^N K\left(\frac{\theta_1 - \Theta_{i1}}{h}, \dots, \frac{\theta_d - \Theta_{id}}{h}\right) = \frac{1}{Nh^d} \sum_{i=1}^N K\left(\frac{\theta - \Theta_i}{h}\right)$$

Veamos ahora una versión multidimensional del lema de Bochner. Se puede usar este resultado para obtener los teoremas vistos para cuando el rasgo es unidimensional en el caso multidimensional más sencillo. Este lema es el análogo al lema 3.3.1 para el caso multidimensional.

Lema 3.3.5. Sea $K_h = \frac{1}{h^d} K(\frac{\bullet}{h})$. Consideremos f en $L^1(R^d)$ y K en $L^1(R^d) \cap L^\infty(R^d)$. Además:

i) $\int_{R^d} K(t) dt = 1$

ii) $K(\theta) = o(\|\theta\|^{-d})$ si $\|\theta\| \rightarrow +\infty$

Sea $f_h = f * K_h$ entonces $f_h \rightarrow f$ si $h \rightarrow 0$ en cada punto de continuidad θ de f .

Demostración: Al ser θ un punto de continuidad de f , dado $\epsilon > 0$, existe $\delta > 0$ tal que $|f(\theta - t) - f(\theta)| < \frac{\epsilon}{3\|K\|_1}$ si $\|t\| < \delta$. En primer lugar veamos que $\int_{R^d} K_h(t) dt = 1$ ya que

$$\int_{R^d} K_h(t) dt = \int_{R^d} \frac{1}{h^d} K\left(\frac{t}{h}\right) dt = \int_{R^d} K(u) du = 1$$

Tomando la diferencia $|f_h(\theta) - f(\theta)| = \left| \int_{R^d} f(\theta - t) K_h(t) dt - f(\theta) \int_{R^d} K_h(t) dt \right| =$

$$\left| \int_{R^d} (f(\theta - t) - f(\theta)) K_h(t) dt \right| \leq \frac{\epsilon}{3\|K\|_1} \int_{\|t\| < \delta} |K_h(t)| dt + \int_{\|t\| \geq \delta} |f(\theta - t)| |K_h(t)| dt + \int_{\|t\| \geq \delta} |f(\theta)| |K_h(t)| dt$$

Además $\int_{|t| \geq \delta} |K_h(t)| dt < \frac{\epsilon}{3|f(\theta)|}$ si h es suficientemente chico por ser la cola de una integral convergente. Definamos la función $\ell : R^d \rightarrow R$ mediante $\ell(\theta) = |K(\theta)| \|\theta\|^d$. Por ii) vemos que $\ell(\theta) \rightarrow 0$ si $\|\theta\| \rightarrow +\infty$ luego

$$\int_{\|t\| \geq \delta} |f(\theta - t)| |K_h(t)| dt = \int_{\|t\| \geq \delta} |f(\theta - t)| \frac{\ell(t/h)}{\|t\|^d} dt \leq \frac{1}{\delta^d} \sup_{\|t\| \geq \delta} \ell(t/h) \int_{\|t\| \geq \delta} |f(\theta - t)| dt$$

Como $\sup_{\|t\| \geq \delta} \ell(t/h) \rightarrow 0$ si $h \rightarrow 0$ pues $\|t/h\| \rightarrow +\infty$ y $\int_{\|t\| \geq \delta} |f(\theta - t)| dt \leq \|f\|_1$ entonces $\int_{\|t\| \geq \delta} |f(\theta - t)| |K_h(t)| dt < \frac{\epsilon}{3}$ luego obtenemos que $|f_h(\theta) - f(\theta)| < \epsilon$.

Usando el teorema anterior obtenemos resultados similares a los del caso unidimensional. Debe hacerse una serie de elecciones, en principio el núcleo d dimensional y luego la matriz de anchos de banda, H lo que implica dos elecciones la parametrización y la elección de los h .

Planteemos las siguientes hipótesis:

H1) Todas las componentes de la matriz Hessiana de f (\mathfrak{H}_f) son continuas a trozos y de cuadrado integrable.

H2) $H = H_N$ es una sucesión de matrices tales que todos los coeficientes de H tienden a 0 cuando $N \rightarrow \infty$ y $\lim_{N \rightarrow \infty} N |H| = \infty$. La razón entre el mayor valor propio y el menor valor propio de H permanece acotado para todo n .

H3) K es un núcleo d dimensional acotado de soporte compacto y cumple $\int z K(z) dz = 0$, y $\int z K(z) dz = 0$, $\int z z^t K(z) dz = \mu_2(K) I_d$ con $\mu_2(K) = \int z_i^2 K(z) dz$ para todo i .

Bajo estas condiciones se obtiene aplicando el desarrollo de Taylor multidimensional al igual que en el caso unidimensional (al que vimos en detalle) que:

$$E(f_{N,H}(\theta)) = \int K_H(z - \theta) f(z) dz = \int K_H(s) f(\theta + Hs) ds \approx \int K_H(s) \{f(\theta) + s^t H^t \nabla f(\theta) + \frac{1}{2} s^t H^t \mathfrak{H}_f \nabla H s\} ds$$

$$E(f_{N,H}(\theta)) \approx f(\theta) + \frac{1}{2} \mu_2(K) \text{tr}(H^t \mathfrak{H}_f(\theta) H)$$

$$V(f_{N,H}) = \frac{1}{N} \int (K_H(z - \theta))^2 dz - \frac{1}{N} (E(f_{N,H}(\theta)))^2 \approx \frac{1}{N|H|} \int (K_H(s))^2 f(\theta + Hs) ds$$

$$V(f_{N,H}) \approx \frac{1}{N|H|} \int (K_H(s))^2 \{f(\theta) + s^t H^t \nabla f(\theta)\} ds \approx \frac{1}{N|H|} \|K\|_2^2 f(\theta)$$

El sesgo integrado se define como $SI(H) = \int (E(f_{N,H}(\theta)) - f(\theta))^2 d\theta$ y definimos el sesgo integrado asintótico $SIA(H)$ como el término de primer orden de $SI(H)$ o sea $\frac{SI(H) - SIA(H)}{SIA(H)} = o(1)$ cuando $N \rightarrow \infty$, $H \rightarrow 0$ y $NH \rightarrow \infty$.

La varianza integrada se define como $VI(H) = \int (E(f_{N,H}(\theta)) - E(f_{N,H}(\theta)))^2 d\theta$ e igualmente definimos la varianza integrada asintótica. El error cuadrático integrado medio asintótico (ECIMA) se define $ECIMA(H) = SIA(H) + VIA(H)$ Tenemos a partir de lo anterior que:

$$SIA(H) = \frac{1}{4} \mu_2(K)^2 \int \text{tr}^2(H^t \mathfrak{H}_f(\theta) H) d\theta$$

$$VIA(H) = \frac{1}{N|H|} \|K\|_2^2$$

$$ECIMA(f_{N,H}) = \frac{1}{4} \mu_2(K)^2 \int \text{tr}^2(H^t \mathfrak{H}_f(\theta) H) d\theta + \frac{1}{N|H|} \|K\|_2^2$$

En el caso particular $H = hI$ se obtiene

$$ECIMA(f_{N,H}) = \frac{\int K(z)^2 dz}{N h^d} + \frac{1}{4} h^4 \mu_2(K)^2 \int (\nabla^2 f(\theta))^2 d\theta$$

En este caso se puede obtener una fórmula explícita para la ventana óptima que da

$$(3.15) \quad h = \left(\frac{d \int K(z)^2 dz}{N \mu_2(K)^2 \int (\nabla^2 f(\theta))^2 d\theta} \right)^{1/(d+4)}$$

Si calculamos

$$\inf_{h>0} ECIMA(f_{N,H}) = \frac{d+4}{4d} (\mu_2(K)^{2d} d^4 (\int K(z)^2 dz)^4 \int (\nabla^2 f(\theta))^2 d\theta)^d N^{-4})^{1/(d+4)}$$

o sea, el orden de convergencia es $N^{-4/(d+4)}$ el cual disminuye a medida que crece la dimensión, justamente este orden manifiesta claramente la llamada "maldición de la dimensionalidad".

Distribución asintótica del estimador

Estos estimadores adecuadamente normalizados convergen a la distribución normal. Comencemos dando sin demostración la conocida desigualdad de Jensen.

Teorema 3.3.6. Desigualdad de Jensen

Sea μ una medida positiva sobre una σ -álgebra M en un conjunto Ω tal que $\mu(\Omega) = 1$. Sea f una función real en $L^1(\mu)$ con $a < f(x) < b$ para todo $x \in \Omega$. Si φ es convexa en (a,b) entonces

$$\varphi\left(\int_{\Omega} f d\mu\right) \leq \int_{\Omega} \varphi(f) d\mu$$

Lema 3.3.7. Suponiendo que $\widehat{\Theta}_i = \Theta_i$ para todo $i, i=1, \dots, N$ y bajo las condiciones del lema 3.3.1.

$$\frac{f_N(\theta) - E f_N(\theta)}{\sigma(f_N(\theta))} \xrightarrow{w} N(0, 1)$$

Demostración: El estimador de la densidad de la habilidad lo podemos escribir como

$$f_N(\theta) = \frac{1}{Nh} \sum_{k=1}^N (K \frac{X - \Theta_k}{h}) = \frac{1}{N} \sum_{k=1}^N W_{N_k} \text{ con } W_{N_k} = \frac{1}{h} K\left(\frac{X - \Theta_k}{h}\right)$$

O sea, se tiene un sistema triangular de variables idénticamente distribuidas por filas. Utilizaremos la condición de Liapunov para la convergencia a la Normal, de más está decir que esta condición implica las de Lindeberg. Una condición suficiente para que:

$$\frac{\hat{f}_N(\theta) - E \hat{f}_N(\theta)}{\sigma(\hat{f}_N(\theta))} \xrightarrow{w} Z \sim N(0, 1) \text{ es que } \frac{E(|W_{N_1} - E(W_{N_1})|^{2+\delta})}{N^{\delta/2} \sigma(W_{N_1})^{2+\delta}} \longrightarrow 0 \text{ con } N_1 \rightarrow \infty \text{ para algún } \delta > 0$$

$$E(|W_{N_1}|^{2+\delta}) = \frac{1}{h^{2+\delta}} E(K^{2+\delta} \left(\frac{\theta - \Theta_1}{h}\right)) = \frac{1}{h^{1+\delta}} \frac{1}{h} \int K^{2+\delta} \left(\frac{\theta - t}{h}\right) f(t) dt$$

Si se cumplen las hipótesis del lema 3.3.1 $\frac{1}{h} \int K^{2+\delta} \left(\frac{\theta - t}{h}\right) f(t) dt \rightarrow f(\theta) \int K^{2+\delta}(t) dt$ así pues,

$$h^{1+\delta} E(|W_{N_1}|^{2+\delta}) \rightarrow f(\theta) \int K^{2+\delta}(t) dt$$

De igual modo: $\sigma^2(W_{N_1}) = \frac{1}{h^2} Var \left(K \left(\frac{\theta - \Theta_1}{h}\right)\right) = \frac{1}{h^2} \int K^2 \left(\frac{\theta - t}{h}\right) f(t) dt$.

Como $\frac{1}{h} \int K^2 \left(\frac{\theta - t}{h}\right) f(t) dt$ se tiene que: $h \sigma^2(W_{N_1}) \rightarrow f(\theta) \int K^2(t) dt$

También $E(W_{N_1}) \rightarrow f(\theta)$ si $\int K(u) du = 1$

$$E(|W_{N_1} - E(W_{N_1})|^{2+\delta}) \leq E(|W_{N_1}| + |E(W_{N_1})|)^{2+\delta}$$

usando la desigualdad: $\left(\frac{a+b}{2}\right)^{2+\delta} \leq \frac{a^{2+\delta} + b^{2+\delta}}{2}$ obtenemos:

$$(|W_{N_1}| + |E(W_{N_1})|)^{2+\delta} \leq \frac{2^{2+\delta}}{2} (|W_{N_1}|^{2+\delta} + |E(W_{N_1})|^{2+\delta}) = 2^{1+\delta} (|W_{N_1}|^{2+\delta} + |E(W_{N_1})|^{2+\delta})$$

Aplicando la desigualdad de Jensen a la función $\varphi = \theta^{2+\delta}$ tenemos que:

$$(E|W_{N_1}|)^{2+\delta} \leq E(|W_{N_1}|^{2+\delta}) \text{ así pues}$$

$$E(|W_{N_1}| + |E(W_{N_1})|)^{2+\delta} \leq E(2^{1+\delta} (|W_{N_1}|^{2+\delta} + |E(W_{N_1})|^{2+\delta})) \leq$$

$$\leq E(2^{1+\delta}(|W_{N1}|^{2+\delta} + E(|W_{N1}|^{2+\delta}))) = 2^{1+\delta}(E|W_{N1}|^{2+\delta} + E|W_{N1}|^{2+\delta}) = 2^{2+\delta}E(|W_{N1}|^{2+\delta})$$

luego

$$\frac{E(|W_{N1} - E(W_{N1})|^{2+\delta})}{N^{\frac{\delta}{2}}\sigma(W_{N1})^{2+\delta}} \leq \frac{2^{2+\delta}E(|W_{N1}|^{2+\delta})}{N^{\frac{\delta}{2}}\sigma(W_{N1})^{2+\delta}} = \frac{2^{2+\delta}}{h^{1+\delta}}h^{1+\delta}E(|W_{N1}|^{2+\delta})/N^{\frac{\delta}{2}}\frac{h^{1+\frac{\delta}{2}}\sigma(W_{N1})^{2+\delta}}{h^{1+\frac{\delta}{2}}}$$

$$\leq \frac{C}{h^{1+\delta}}/\frac{N^{\delta/2}}{h^{1+\delta/2}} = \frac{C}{(Nh)^{\delta/2}} \rightarrow 0 \text{ si } Nh \rightarrow \infty \text{ pues}$$

$$h^{1+\delta}E(|W_{N1}|^{2+\delta}) \rightarrow f(\theta) \int K^2(t)dt \text{ y } h^{1+d/2}\sigma(w_{N1})^{2+\delta} \rightarrow cte$$

así pues se cumple la condición de Liapunov y hemos probado que:

$$\frac{f_N(\theta) - Ef_N(\theta)}{\sigma(f_N(\theta))} \xrightarrow{w} N(0, 1)$$

Teorema 3.3.8. Supongamos que $\widehat{\Theta}_i = \Theta_i$ para todo $i, i=1, \dots, N$; que K es acotada y que $\int K(u)du = 1$, además se cumple que $\mu^2 K(u)$ es acotada, $Nh \rightarrow \infty$ y $Nh^3 \rightarrow 0$. También la densidad f es Lipchitz en un entorno de θ , entonces:

- a) $Var(\sqrt{Nh}\widehat{f_N}(\theta)) \rightarrow f(\theta) \int K^2(u)du$
- b) $\sqrt{Nh} \left(\frac{1}{Nh} \sum_{i=1}^N E(K(\frac{\theta - \Theta_i}{h})) - f(\theta) \right) \rightarrow 0$
- c) $\sqrt{Nh} (\widehat{f_N}(\theta) - f(\theta)) \rightarrow N(0, f(\theta) \int K^2(u)du)$

Demostración: a) Como K es acotado, $\int K(u)du = 1$ entonces $\int K^2(u)du < \infty$.

Sabemos que $|uK(u)| \rightarrow 0$ si $|u| \rightarrow \infty$ pues $u^2 K(u)$ acotado luego se cumplen las hipótesis del lema 3.3.1, así que a) se cumple,

- b) existe $\delta > 0$ tal que $|f(\theta + s) - f(\theta)| \leq c|s|$; $|s| < \delta$

$$|\sqrt{Nh} \left(\frac{1}{Nh} \sum_{i=1}^N E(K(\frac{\theta - \Theta_i}{h})) - f(\theta) \right)| = |\sqrt{Nh} \left(\int \frac{1}{h} K(\frac{\theta - t}{h}) f(t)dt - f(\theta) \right)| =$$

$$|\sqrt{Nh} \left(\int K(u)f(\theta + uh)du - f(\theta) \int K(u)du \right)| = |\sqrt{Nh} \int K(u)|f(\theta + uh) - f(\theta)||du| \leq$$

$$\leq \sqrt{Nh} \left[hC \int_{|hu| < \delta} |uK(u)|du + \int_{|u| > \delta/h} |K(u)||f(\theta + uh)|du + |f(\theta)| \int_{|u| > \delta/h} |K(u)|du \right] =$$

Sea $C_1 = \|u^2 K(u)\|_\infty$

$$= \sqrt{Nh} \left[Ch \int_{|hu| < \delta} |uK(u)|du + \int_{|u| > \delta/h} \frac{|u^2 K(u)|}{|u^2|} f(\theta + uh)du + |f(\theta)| \int_{|u| > \delta/h} \frac{|u^2 K(u)|}{u^2} \right] \leq$$

$$\leq \sqrt{Nh} \left[Ch \int_{|hu| < \delta} |uK(u)|du + \frac{h^2}{\delta^2} C_1 \int_{|u| > \delta/h} f(\theta + uh)du + C_1 f(\theta) \int_{|u| > \delta/h} \frac{du}{u^2} \right] \leq$$

$$\leq \sqrt{Nh} \left[Ch \int_{|hu| < \delta} |uK(u)|du + h \frac{C_1}{\delta^2} \int_R f(t) + C_1 f(\theta) h \int_{|t| > \delta} \frac{dt}{t^2} \right] = \sqrt{Nh^3} [C_2 + C_3 + C_4] \rightarrow 0$$

$$\text{si } Nh^3 \rightarrow 0 \text{ pues } \int_{|u| > \delta/h} \frac{du}{u^2} = h \int_{|t| > \delta} \frac{dt}{t^2} \int_{|u| > \delta/h} f(\theta + uh)du = \int_R f(t) \frac{dt}{h} \text{ y } \int_{|hu| < \delta} |uK(u)|du$$

está acotado al igual que: $f(\theta) \int_{|t| > \delta} \frac{dt}{t^2}$

$$\begin{aligned}
c) \sqrt{Nh} (f_N(\theta) - f(\theta)) &= \sqrt{Nh} (f_N(\theta) - Ef_N(\theta)) + \sqrt{Nh} [Ef_N(\theta) - f(\theta)] \\
\text{sabemos que: } \sqrt{Nh} [Ef_N(\theta) - f(\theta)] &\rightarrow 0 \text{ consideremos, } \sigma(f_N(\theta)) \frac{(f_N(\theta) - Ef_N(\theta))}{\sigma(f_N(\theta))} \\
\text{se tiene que: } \frac{(f_N(\theta) - Ef_N(\theta))}{\sigma(f_N(\theta))} &\sim N(0, 1) \text{ además, } Nh\sigma^2(f_N(\theta)) \rightarrow f(\theta) \int K^2(u) du \\
\text{entonces: } \sqrt{Nh}\sigma(f_N(\theta)) &\rightarrow (f(\theta) \int K^2(u) du)^{1/2} \text{ así pues,} \\
\sqrt{Nh} (f_N(\theta) - f(\theta)) &\rightarrow N(0, f(\theta) \int K^2(u) du)
\end{aligned}$$

Convergencia completa del estimador

El estimador no solo converge en probabilidad o sea es debilmente consistente sino que también o hace completamente lo cual implica convergencia casi segura o sea la consistencia fuerte. Recordemos que una sucesión de variables aleatorias X_n converge completamente a otra variable aleatoria X si para todo $\varepsilon > 0$

$$\sum_{i=1}^{\infty} P(|X_n - X| > \varepsilon) < \infty$$

El teorema que prueba la convergencia completa utiliza una desigualdad exponencial (desigualdad de Bernstein) para acotar superiormente la serie de probabilidades mediante una serie convergente.

Teorema 3.3.9. Sea $f_N(\theta) = \frac{1}{Nh} \sum_{i=1}^N K\left(\frac{\theta - \hat{\Theta}_i}{h}\right)$, con $\hat{\Theta}_1, \dots, \hat{\Theta}_N$ estimadores de $\Theta_1, \dots, \Theta_N$ independientes y que tienen densidad f . Supongamos que el rasgo se mide sin error es decir $\widehat{\Theta}_i = \Theta_i$ para todo $i, i=1, \dots, N$. El núcleo K cumple las hipótesis del lema 3.3.1, e $\int K = 1$

En estas condiciones el estimador basado en núcleos converge completamente a la densidad f en cada punto de continuidad de f .

Demostración: Queremos probar que $\sum_{i=1}^{\infty} P(|f_N(\theta) - f(\theta)| > \epsilon) < \infty$.

$$\{|f_N(\theta) - f(\theta)| > \epsilon\} = \left\{ \left| \frac{1}{Nh} \sum_{i=1}^N K\left(\frac{\theta - \Theta_i}{h}\right) - f(\theta) \right| > \epsilon \right\} = \left\{ \left| \frac{1}{N} \left[\sum_{i=1}^N \left(\frac{1}{h} K\left(\frac{\theta - \Theta_i}{h}\right) - f(\theta) \right) \right] \right| > \epsilon \right\}$$

$$\begin{aligned}
\text{entonces } P(|f_N(\theta) - f(\theta)| > \epsilon) &= P\left(\left| \frac{1}{N} \sum_{i=1}^N \left(\frac{1}{h} K\left(\frac{\theta - \Theta_i}{h}\right) - f(\theta) \right) \right| > \epsilon \right) \leq \\
&\leq P\left(\frac{1}{N} \sum_{i=1}^N \left| \frac{1}{h} K\left(\frac{\theta - \Theta_i}{h}\right) - f(\theta) \right| > \epsilon \right)
\end{aligned}$$

Consideremos la variable $Y_i = \frac{1}{h} K\left(\frac{\theta - \Theta_i}{h}\right) - f(\theta)$. Claramente las Y_i son independientes si las Θ_i lo son:

$$E(Y_i) = E\left(\frac{1}{h} K\left(\frac{\theta - \Theta_i}{h}\right)\right) - f(\theta) = \frac{1}{h} \int K\left(\frac{\theta - t}{h}\right) f(t) dt - f(\theta) \rightarrow 0 \text{ pues } \frac{1}{h} \int K\left(\frac{\theta - t}{h}\right) f(t) dt \rightarrow f(\theta)$$

$$\text{Además, } |Y_i| = \left| \frac{1}{h} K\left(\frac{\theta - \Theta_i}{h}\right) - f(\theta) \right| < C \text{ pues } K \text{ es acotada.}$$

$$E(Y_i^2) = E\left(\frac{1}{h^2} K^2\left(\frac{\theta - \Theta_i}{h}\right)\right) - 2f(\theta)E\left(\frac{1}{h} K\left(\frac{\theta - \Theta_i}{h}\right)\right) + f^2(\theta) =$$

$$\frac{1}{h} \frac{1}{h} \int K^2\left(\frac{\theta-t}{h}\right) f(t) dt - 2f(\theta) \frac{1}{h} \int K\left(\frac{\theta-t}{h}\right) f(t) dt + f^2(\theta) \rightarrow \frac{1}{h} f(\theta) \int K^2(t) dt - f^2(\theta)$$

$$\text{ya que } \frac{1}{h} \int K^2\left(\frac{\theta-t}{h}\right) f(t) dt \rightarrow f(\theta) \int K^2(t) dt \text{ y } (1/h) \int K\left(\frac{\theta-t}{h}\right) f(t) dt \rightarrow f(\theta)$$

Utilizando la desigualdad de Bernstein tenemos:

$$P\left(\frac{1}{N} \sum_{i=1}^N |Y_i| > \epsilon\right) \leq 2 \exp\left(\frac{-N^2 \epsilon^2}{2\left(N\left(\frac{1}{h} f(\theta) \int K^2(t) dt - f^2(\theta)\right) + cN\epsilon\right)}\right)$$

Operando en el exponente

$$\frac{-N^2 \epsilon^2}{2\left(N\left(\frac{f(\theta)}{h} \int K^2(t) dt - f^2(\theta) + c\epsilon\right)\right)} = \frac{-N\epsilon^2}{2\left(\frac{C_2}{h} - C_1 + c\epsilon\right)} = \frac{-N\epsilon^2}{2\left(\frac{C_2}{h} + C_3\right)} \leq \frac{-N\epsilon^2}{2\frac{C_2}{h}} = -NhC'$$

$$\text{Así pues } P\left(\frac{1}{N} \sum_{i=1}^N |Y_i| > \epsilon\right) \leq 2 \exp(-NhC') \text{ entonces } \sum P\left(\frac{1}{N} \sum_{i=1}^N |Y_i| > \epsilon\right) \leq \sum 2e^{-NhC'}$$

$$\text{Si } Q(N) = \frac{Nh}{\log N} \text{ tenemos } Nh = Q(N) \log N \text{ entonces } 2e^{-C' \log N Q(N)} = 2e^{\log N^{-C' Q(N)}} = 2N^{-C' Q(N)}$$

Si $Q(N) \rightarrow \infty$ resulta que $\sum 2N^{-C' Q(N)}$ converge, así pues hay convergencia completa.

Estimación de la cópula

Una forma sencilla de estimar la cópula es considerar la cópula empírica. Supongamos que tenemos los estimados $(\widehat{\theta}_{1i}, \dots, \widehat{\theta}_{di})$, $i = 1, \dots, N$ de $(\theta_{1i}, \dots, \theta_{di})$, $i = 1, \dots, N$

Sabemos que si $(U_{1i}, \dots, U_{di}) = (F_1(\theta_{1i}), \dots, F_d(\theta_{di}))$, $i = 1, \dots, n$. entonces

$$C(u_1, u_2, \dots, u_d) = \mathbb{P}[U_1 \leq u_1, U_2 \leq u_2, \dots, U_d \leq u_d]$$

Consideremos $(\widehat{U}_{1i}, \dots, \widehat{U}_{di}) = (F_1(\widehat{\theta}_{1i}), \dots, F_d(\widehat{\theta}_{di}))$, $i = 1, \dots, n$. Observemos que este vector coincide con los estimados originales de los rasgos en cubo $[0, 1]^d$. Luego el candidato a estimador de la cópula es

$$C(u_1, \widehat{u_2}, \dots, u_d) = \mathbb{P}[\widehat{U}_1 \leq u_1, \dots, \widehat{U}_d \leq u_d]$$

Este lo podemos calcular empíricamente como

$$(3.16) \quad C(u_1, \widehat{u_2}, \dots, u_d) = \frac{1}{N} \sum_{i=1}^N \mathbf{1}(\widehat{U}_{1i} \leq u_1, \dots, \widehat{U}_{di} \leq u_d)$$

Es claro que falta estudiar el comportamiento estadístico de este estimador es decir si es consistente.

3.4. Núcleos

En esta sección queremos presentar los núcleos que pueden usarse para estimar los modelos. Primeramente veamos núcleos univariados comúnmente usados, pues a partir de estos suelen contruirse los multivariados.

Los núcleos más comunes son:

Uniforme	$K(u) = \frac{1}{2} 1_{(u \leq 1)}$
Triangular	$K(u) = (1 - u) 1_{(u \leq 1)}$
Epanechnikov	$K(u) = \frac{3}{4}(1 - u^2) 1_{(u \leq 1)}$
Quartic o Biweight	$K(u) = \frac{15}{16}(1 - u^2)^2 1_{(u \leq 1)}$
Triweight	$K(u) = \frac{35}{32}(1 - u^2)^3 1_{(u \leq 1)}$
Tricubo	$K(u) = \frac{70}{81}(1 - u ^3)^3 1_{\{ u \leq 1\}}$
Gaussiano	$K(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2}$
Coseno	$K(u) = \frac{\pi}{4} \cos\left(\frac{\pi}{2}u\right) 1_{(u \leq 1)}$
Logístico	$K(u) = \frac{1}{e^u + 2 + e^{-u}}$
Silverman	$K(u) = \frac{1}{2} e^{-\frac{ u }{\sqrt{2}}} \cdot \sin\left(\frac{ u }{\sqrt{2}} + \frac{\pi}{4}\right)$
Sigmoideo	$K(u) = \frac{2}{\pi} \frac{1}{e^u + e^{-u}}$

Cuadro 3.1: Tipos de núcleo

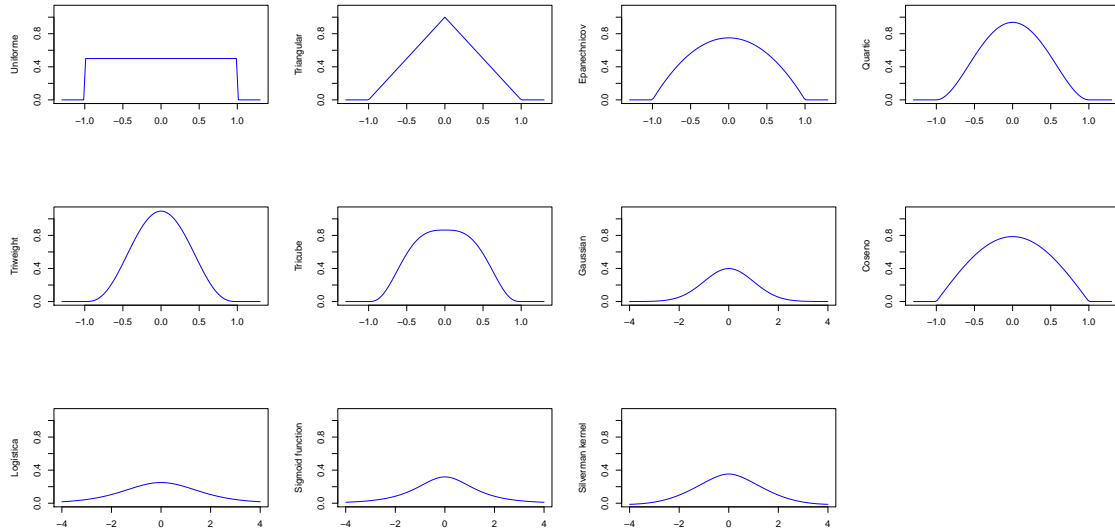


Figura 3.2: Principales núcleos unidimensionales

Hay dos formas clásicas de construir núcleos multivariados, la primera, es construyendo el llamado núcleo multiplicativo, y la segunda, es construir núcleos esféricos o simétricos radialmente.

El núcleo producto se crea mediante $K(u) = K_1(u_1)K_2(u_2)\dots K_d(u_d)$ donde cada K_i es un núcleo unidimensional. Como ejemplo de este tipo de núcleo tenemos el llamado núcleo multiplicativo de Epanechnikov $(\frac{3}{4})^d(1 - u_1^2)(1 - u_2^2)\dots(1 - u_d^2)1_{(|u_1| \leq 1)}\dots 1_{(|u_d| \leq 1)}$

Los núcleos radiales simétricos son de la forma $K^*(u) = \frac{K(\|u\|)}{\int_R K(t)}$ con $\|u\| = u^T u$ y K un núcleo unidimensional.

Por ejemplo el núcleo de Epanechnikov esférico $K(u) = \propto (1 - u^T u)1_{(u^T u \leq 1)}$ y el núcleo normal multivariado es: $K_{\mathbf{H}}(\mathbf{u}) = (2\pi)^{-d/2} |\mathbf{H}|^{-1/2} e^{-\frac{1}{2} \mathbf{u}^T \mathbf{H}^{-1} \mathbf{u}}$ donde \mathbf{H} es la matriz de varianzas covarianzas.

También está el núcleo de Bartlett-Epanechnikov donde si $c_d = \frac{\pi^{d/2}}{\Gamma((d/2)+1)}$ es el volumen de la esfera unitaria en el espacio R^d viene dado por $K(x) = \frac{(1-x^t x)(d+2)}{2c_d} 1_{(x^t x \leq 1)}$

Otros núcleos a ser usados son $K_2(x) = 3\pi^{-1}(1-x^t x)^2 1_{(x^t x \leq 1)}$ y $K_3(x) = 4\pi^{-1}(1-x^t x)^3 1_{(x^t x \leq 1)}$

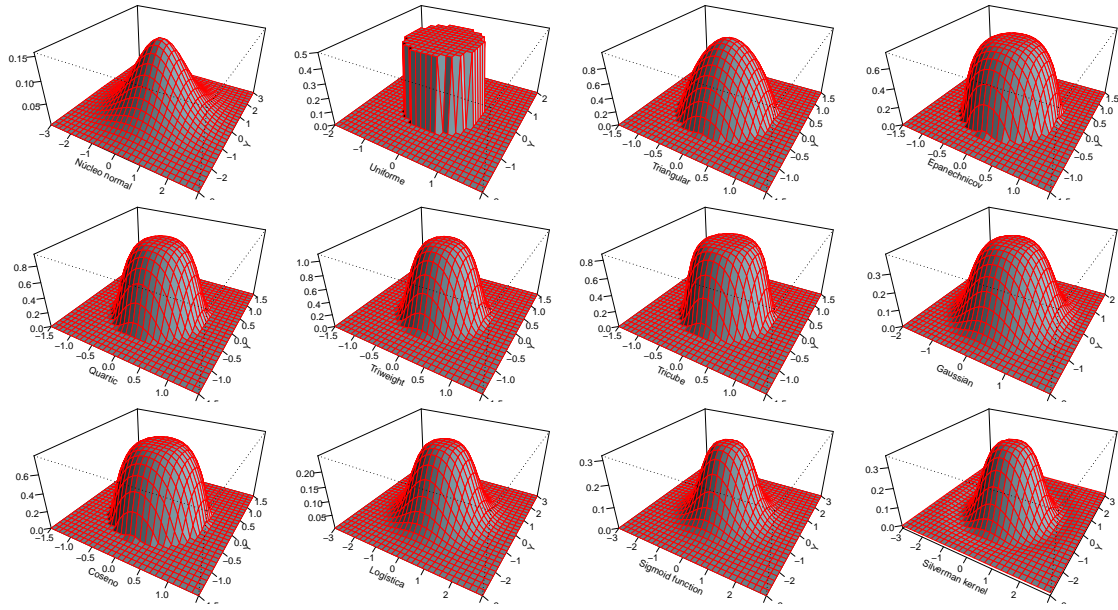


Figura 3.3: Núcleos multiplicativos de dimensión 2

Cabe preguntarse que tipo de núcleo elegir (multiplicativo o esférico) si queremos estimar la densidad del rasgo. Se puede probar que para comparar dos núcleos multivariados las magnitudes a comparar son $(\mu_2(K)^{2d}(\int K(z)^2 dz)^4)^{1/(d+4)}$. Entonces tomando el cociente de estas cantidades para el núcleo esférico sobre el núcleo multiplicativo podemos medir sus eficiencias relativas. Wand y Jones (1995) realizan estas comparación para estos núcleos basados en el núcleo unidimensional beta o sea $K(x) \propto (1-x^2)^p 1_{|x|<1}$.

p	d=2	d=3	d=4
0	0.955	0.888	0.811
1	0.982	0.953	0.916
2	0.983	0.953	0.915
3	0.984	0.956	0.919

Cuadro 3.2: Núcleo beta

Se observa que el núcleo esférico tiene un comportamiento asintótico mejor que el multiplicativo, siendo en particular muy importante para el núcleo uniforme ($p=0$) y que esta diferencia crece con la dimensión. Si queremos encontrar el núcleo óptimo en este sentido se debe encontrar el núcleo que perteneciendo a una familia minimice la cantidad $(\mu_2(K)^{2d}(\int K(z)^2 dz)^4)^{1/(d+4)}$. El núcleo multiplicativo que minimiza esta expresión es $K(x) = (\frac{3}{4})^d \prod_{i=1}^d (1-x_i^2) 1_{|x|<1}$ en el caso esférico es $K(x) = \frac{d\Gamma(d/2)}{4\pi^{d/2}}(d+2)(1-x^t x) 1_{x^t x \leq 1}$.

3.5. Elección del ancho de banda para densidad

Elegir el ancho de banda H a utilizar es un tema importante para la estimación, tanto de la densidad como de la CCI, debido a su efecto compensatorio entre el sesgo y la varianza. En una o dos dimensiones se puede elegir el ancho de banda apropiado observando sucesivas estimaciones para diferentes anchos de banda. Cuando queremos utilizar un método para elegir el ancho de banda suelen usarse dos enfoques: el método plug-in, en particular la llamada “regla del pulgar” y métodos de remuestreo, como la validación cruzada y el bootstrap.

Härdle y Müller (1997) introducen generalizaciones a la “regla del pulgar” de Silverman y a la validación cruzada por mínimos cuadrados.

Método plug-in de Härdle y Müller

Supongamos que el núcleo K es normal típico con dimensión d , luego se tiene $\|K\|_2^2 = 2^{-d}\pi^{-d/2}$; $\mu_2(K) = 1$ y

$$\int [tr \{H^T \mathfrak{H}_f(\theta) H\}]^2 dt = \frac{1}{2^{d+2}\pi^{d/2} |\Sigma|^{1/2}} \left[2tr(H^T \Sigma^{-1} H)^2 + \{tr(H^T \Sigma^{-1} H)\}^2 \right]$$

Se puede derivar fácilmente la fórmula de la “regla del pulgar” suponiendo diferentes condiciones sobre H y Σ .

En 3.15 se mostro que

$$(3.17) \quad h = \left(\frac{d \int K(z)^2 dz}{N \mu_2(K)^2 \int (\nabla^2 f(\theta))^2 d\theta} \right)^{1/(d+4)}$$

Si los datos provienen de una distribución normal multivariante se obtiene que

$$h = Cn^{-1/(d+4)}$$

donde la constante C depende del núcleo.

Si el núcleo es gaussiano de dimensión d $C = \frac{4}{d+2}^{1/(d+4)}$, si el núcleo es K_2 , $C = 2.78$ y K_3 , $C = 3.12$, en tanto si se usa el núcleo de Epanechnikov de dimensión d , $C = (8c_d^{-1}(d+2)2^d\pi^{d/2})^{1/(d+4)}$

Si H y Σ son matrices diagonales $H = diag(h_1, \dots, h_d)$ y $\Sigma = diag(\sigma_1^2, \dots, \sigma_d^2)$, entonces siguiendo el desarrollo de Silverman (1986) se obtiene $h_j = \left(\frac{4}{d+2} \right)^{1/(d+4)} N^{-1/(d+4)} \sigma_j$

Si se reemplaza σ_j por su estimación y el primer factor por 1 se tiene la regla de Scott (1992), $h_j = \left(\frac{4}{d+2} \right)^{1/(d+4)} N^{-1/(d+4)} \hat{\sigma}_j$.

Scott (1992) también halla para el caso del núcleo producto de dos normales y los datos normales bivariadas $h_i = \sigma_i(1 - \rho^2)^{5/12}(1 + \rho^2/2)^{-1/6}n^{-1/6}$ con $i = 1, 2$

Es difícil derivar la regla para el caso general de H y Σ , pero tomando la matriz de ventanas proporcional a la raíz de la matriz de covarianzas se generaliza la fórmula de Scott obteniendo $\hat{H} = N^{-1/(d+4)} \hat{\Sigma}^{1/2}$. Si se desea usar un núcleo no Gaussiano, consideramos un núcleo K y las funciones equivalentes $K_\delta = \delta^{-1}K(\bullet/\delta)$ con $\delta \geq 0$. Tenemos que $\|K_\delta\|_2^2 = \delta^{-d}\|K\|_2^2$ y $\mu_2(K_\delta) = \delta^2\mu_2(K)$. Se elige δ de tal manera que la relación entre la varianza y el sesgo en el $ECIMA(H, K_\delta)$ sea independiente de K_δ . Esto se cumple si

$$\mu_2^2(K_{\delta_0}) = \|K_\delta\|_2^2 \iff \delta_0 \left\{ \frac{\|K_\delta\|_2^2}{\mu_2^2(K)} \right\}^{1/(d+4)}$$

δ_0 se llama ancho de banda canónico del núcleo K . Si queremos calcular el ancho de banda para diferentes núcleos Härdle y Müller (1997) deducen la fórmula $H_B = \frac{\delta_0^B}{\delta_0^A} H_A$ donde δ_0^A y δ_0^B son los anchos de banda canónicos para los núcleos K^A y K^B y siendo H_A y H_B las matrices usadas con K^A y K^B respectivamente. Con esta elección se obtiene que

$$ECIMA(H_A, K^A) \approx ECIMA(H_B, K^B)$$

En Wand y Jones (1994) se encuentran detallados otros métodos plug-in.

Método de validación cruzada

El método de validación cruzada por mínimos cuadrados tiene por objeto encontrar el ancho de banda óptimo para el ECI. Tenemos que

$$ECI(H) = \int \{\hat{f}_H(t) - f(t)\}^2 dt = \int \hat{f}_H^2(t) dt - 2 \int \{\hat{f}_H(t) f(t)\} dt + \int f^2(t) dt$$

Ignoramos el último término de la ecuación, el primer término puede calcularse a partir de los datos. Para el segundo término puede ser estimado observemos que $\int \hat{f}_H(t) f(t) dt = E\hat{f}_H(T)$, donde ahora T es d -dimensional y puede estimarse mediante

$$E\hat{f}_H(T) = \frac{1}{n} \sum_{i=1}^n \hat{f}_{H,-1}(T_i) \text{ donde } \hat{f}_{H,-i}(t) = \frac{1}{n-1} \sum_{i \neq j=1}^n K_H(T_j - t)$$

Entonces tenemos el criterio de validación cruzada multivariada como:

$$CV(H) = \frac{1}{n^2 \det(H)} \sum_{i=j}^n \sum_{i=j}^n K \star K \{H^{-1}(T_j - t_i)\} - \frac{2}{n(n-1)} \sum_{i=j}^n \sum_{i=jj \neq i}^n K_H(T_j - T_i)$$

La dificultad viene porque el ancho de banda es ahora una matriz H $d \times d$ lo que implica tenemos que minimizar en $d(d+1)/2$ parámetros. Si se asume H como una matriz diagonal, esto se convierte en un problema de optimización d -dimensional.

3.6. Consistencia conjunta del modelo multivariado

Esta sección está dedicado a encontrar las condiciones que deben cumplir los estadísticos para obtener estimaciones consistentes del rasgo y la relación entre el tamaño muestral N y la cantidad de ítems n para obtener estimaciones consistentes de las CCI. A nuestro juicio esta sección es de capital importancia para justificar el uso del estimador.

Como antecedente tenemos a Douglas (1997) que demuestra la consistencia conjunta para el caso unidimensional. Douglas asume ciertas hipótesis sobre los ítems, las CCI, los tamaños muestrales N_n , los núcleos y los anchos de banda h_n . Considera 10 hipótesis que enumeramos a continuación:

1. Independencia local y unidimensionalidad.
2. Θ tiene distribución uniforme $(0,1)$.
3. \overline{P}'_n existe en $(0,1)$ y existe un número m tal que $0 < m < \overline{P}'_n(\theta)$ para todo $\theta \in (0,1)$.
4. Existe $C > 0$ tal que $np_{nk} < C$ para todo k donde $p_{nk} = P(\overline{U}_n = k)$.
5. Existen constantes M_L y M_U tal que $M_L n^{3/2} < N_n < M_U n^r$.

6. $P'_{n,i}$ existe y es continua en $(0,1)$ para todo n,i .
7. Para cada intervalo compacto $[a,b] \subset (0,1)$ existe una constante $M_{a,b}$ tal que para todo $\theta \in [a,b]$ y todo n, i $|P'_{n,i}(\theta)| < M_{a,b}$.
8. K tiene soporte acotado, en particular existe una constante C_K , tal que $K(x) > 0$, cuando $|x| < C_K$ y $K(x) = 0$, cuando $|x| > C_K$.
9. K es Lipschitziana de constante L_K .
10. Existe un número $\alpha \in (0, 1/2)$ y números positivos L_α y U_α tal que para todo n ,
$$L_\alpha n^{-\alpha} < h_n < U_\alpha n^{-\alpha}$$

A partir de estas hipótesis se obtiene primeramente:

Teorema 3.6.1. Dadas las hipótesis A1 a A7 $\max_{j=1,\dots,J_n} |\hat{\Theta}_{n,j} - \Theta_{n,j}| \rightarrow 0$ con probabilidad 1.

También se obtiene un resultado sobre el error cuadrático.

Teorema 3.6.2. Dadas las hipótesis 1 a 10 para $\theta_0 \in (0,1)$ fijo

$$(3.18) \quad \max\{ECM(\hat{P}_{n,i}(\theta_0)) : 1 \leq i \leq n\} \leq O(n^{-2\alpha})$$

donde α es la constante de la hipótesis 10.

Definiendo $\delta_{n,a,b} = \sup_{\theta \in (a,b)} \max_{1 \leq i \leq n} |\hat{P}_{n,i}(\theta) - P_{n,i}(\theta)|$ se obtiene:

Teorema 3.6.3. Dadas las hipótesis 1 a 10, $\delta_{n,a,b} \rightarrow 0$ si $n \rightarrow +\infty$ con probabilidad 1.

Finalmente prueba la consistencia conjunta pues a partir de los teoremas 3.6.1 y 3.6.3, y notando que la probabilidad del conjunto donde falla esta convergencia es la unión de dos conjuntos cada uno con probabilidad 0.

Teorema 3.6.4. $\max\{\Delta_n, \delta_{n,a,b}\} \rightarrow 0$ si $n \rightarrow +\infty$ con probabilidad 1.

3.6.1. Preliminares

La consistencia conjunta requiere que el tamaño muestral y la longitud del test crezcan juntos.

Esto puede ejemplificarse mediante arreglos triangulares. Consideremos, primero, un test de n ítems, administrado a N_n examinados, elegidos al azar. Se estiman las CCI de los n ítems y los N_n rasgos. Entonces ahora, se considera un nuevo test de $n+1$ ítems, que no necesariamente contengan los n anteriores, y se administra a N_{n+1} examinados.

Haremos la estimación para cada fila en el arreglo triangular así obtenido y estudiaremos la adecuación de las CCI y de los estimados del rasgo a medida que n tiende a infinito.

Tenemos entonces la sucesión de test:

$$P_{n,1}, P_{n,2}, \dots, P_{n,n}$$

$$P_{n+1,1}, P_{n+1,2}, \dots, P_{n+1,n}, P_{n+1,n+1}$$

$$P_{n+2,1}, P_{n+2,2}, \dots, P_{n+2,n+1}, P_{n+2,n+2}$$

.....

.....

y la sucesión de rasgos:

$$\theta_{n,1}, \theta_{n,2}, \dots, \theta_{n,N_n}$$

$$\theta_{n+1,1}, \theta_{n+1,2}, \dots, \theta_{n+1,N_n}, \theta_{n+1,N_{n+1}}$$

$$\theta_{n+2,1}, \theta_{n+2,2}, \dots, \theta_{n+2,N_{n+1}}, \theta_{n+2,N_{n+2}}$$

.....

.....

Se quiere entonces encontrar las condiciones para que la sucesión de estimadores de las curvas converjan a las verdaderas CCI y la sucesión de rasgos converja a los rasgos verdaderos conjuntamente.

Se obtiene bajo condiciones generales un resultado para la consistencia del rasgo donde si se cumplen las hipótesis 1) a 7) (ver la sección de hipótesis) $\max_{k=1,\dots,N} \|\hat{\Theta}_{n,k} - \Theta_{n,k}\| \rightarrow 0$ en probabilidad.

Si, además, se cumple 8), entonces, $\max_{k=1,\dots,N} \|\hat{\Theta}_{n,k} - \Theta_{n,k}\| \rightarrow 0$ completamente (en particular con probabilidad 1).

Se encuentra también una desigualdad para el error cuadrático medio Se prueba que cuando se cumplen las hipótesis 1) a 12) y para $\theta_0 \in (0, 1)^d$ fijo

$$(3.19) \quad \max\{ECM(\hat{P}_{n,i}(\theta_0)) : 1 \leq i \leq n\} \leq O(\max_l (\sum_{i=1}^n c_{n,l,i}^2)^{2\alpha}) + e^{-\eta}$$

El tercer resultado importante es sobre la consistencia de la estimación de las CCI. Dadas las hipótesis 1) a 12): $\sup_{\theta \in (a,b)^d} \max_{1 \leq i \leq n} |\hat{P}_{n,i}(\theta) - P_{n,i}(\theta)| \rightarrow 0$ si $n \rightarrow +\infty$ completamente con $(a,b)^d \subset (0, 1)^d$.

Finalmente, se obtiene la consistencia conjunta de los rasgos y las CCI. Bajo las hipótesis 1) a 12): $\max\{\max_{k=1,\dots,N} \|\hat{\Theta}_{n,k} - \Theta_{n,k}\|, \sup_{\theta \in (a,b)^d} \max_{1 \leq i \leq n} |\hat{P}_{n,i}(\theta) - P_{n,i}(\theta)|\} \rightarrow 0$ si $n \rightarrow +\infty$ con probabilidad 1.

3.6.2. Sucesión de funciones asintóticamente Lipschitz

Una de las condiciones encontradas sobre la sucesión de las funciones de distribución de los estimadores de los rasgos es lo que llamaremos que estas deben ser asintóticamente Lipschitz.

Definición 3.6.5. Diremos que una sucesión de funciones $f_n : V \rightarrow R^d$ con $V \subset R^k$ acotado es asintóticamente Lipschitz si existen sucesiones L_n y $\gamma_n > 0$ tales que: existen $\delta > 0$ y $\lambda > 0$ tal que $\delta < L_n < \lambda$ para todo n y $\gamma_n \rightarrow 0$ y $\|f_n(x) - f_n(y)\| \leq L_n \|x - y\| + \gamma_n \forall x, y \in V$.

Esta definición está motivada por la necesidad de aproximarse al concepto de funciones Lipschitz cuando se tienen sucesiones de funciones de distribución de variables discretas o distribuciones empíricas. Es claro que estas funciones al ser escalonadas no pueden ser Lipschitz pero pueden cumplir la desigualdad anterior que es muy útil. A partir de la definición si la sucesión de funciones tiene límite, entonces, a partir de la continuidad de la norma el límite es una función lipschitziana. Veamos algunos ejemplos para clarificar.

Ejemplo 3.6.6. Supongamos que tenemos n variables aleatorias Bernoulli X_i , y consideremos la media muestral $\bar{X} = \frac{\sum_{i=1}^n X_i}{n}$. Si para algún $K > 0$, $nP(\sum_{i=1}^n X_i = k) < K$, $k = 0, 1, \dots, n$, (esta

condición implica que el histograma es uniformemente acotado), entonces para $0 \leq x \leq x' \leq 1$, la sucesión de funciones de distribución de \bar{X} son asintóticamente Lipschitz.

$$\begin{aligned} F_n(x') - F_n(x) &= \sum_{k=[nx]+1}^{[nx']} P(\sum_{i=1}^n X_i = k) \leq \sum_{k=[nx]+1}^{[nx']} \frac{K}{n} = \\ &= \frac{K}{n}([nx'] - [nx]) \leq \frac{K}{n}(n(x' - x) + 1) = K[(x' - x) + \frac{1}{n}] = K(x' - x) + \frac{K}{n} \end{aligned}$$

Ejemplo 3.6.7. Consideremos ahora como estimador la media ponderada $\bar{X} = \sum_{i=1}^n \omega_i X_i$.

Representemos por $(X_1^j, \dots, X_i^j, \dots, X_n^j)$ $j = 1, \dots, 2^n$ las posibles enuplas de valores que pueden tomarse y sea $I_n = \min_{j \neq k} | \sum_{i,j,k} \omega_i (X_i^j - X_i^k) |$. Si $\frac{P(\sum \omega_i X_i = k)}{I_n} \leq K$ entonces,

$$F_n(x') - F_n(x) = \sum_{k \in (x, x']} P(\sum \omega_i X_i = k) \leq K I_n (\frac{x' - x}{I_n} + 1) = K(x' - x) + I_n.$$

Luego, si $I_n \rightarrow 0$ la sucesión de funciones de distribución son asintóticamente de Lipschitz.

Ejemplo 3.6.8. Observemos también que si X_n es una sucesión de variables discretas con función de distribución F_n y si existen sucesiones $f(n) > 0$, $g(n) > 0$ y $h(n) > 0$ y constantes K_1 y K_2 tales que, $\sup_k \{P(X_n = k)\} \leq \frac{K_1}{f(n)}$ con $k \in \text{Rec}\{X_n\}$, para todo $0 \leq x \leq x' \leq 1$

$$\sum_{k \in \text{Rec}\{X_n\}} \chi_{(x, x']}(k) \leq K_2 g(n)(x' - x) + h(n)$$

con $\frac{h(n)}{f(n)} \rightarrow 0$ y para algún $\delta > 0$ y $\lambda > 0$; $\delta < \frac{g(n)}{f(n)} < \lambda$ para todo n entonces:

$$F_n(x') - F_n(x) \leq \sup_k \{P(X_n = k)\} \sum_{k \in \text{Rec}\{X_n\}} \chi_{(x, x']}(k) \leq \frac{K_1 K_2 g(n)}{f(n)}(x' - x) + \frac{h(n)}{f(n)}$$

lo que implica que F_n son asintóticamente lipschizianas.

Como un primer resultados se tiene el siguiente lema:

Lema 3.6.9. Consideremos una sucesión de estadísticos X_n con función de distribución F_n , que converge completamente a F , la cual es lipschitziana de constante L . Entonces F_n es asintóticamente Lipschitz.

Demostración: Consideremos $0 \leq x \leq x' \leq 1$.

$$|F_n(x') - F_n(x)| \leq |F_n(x') - F(x')| + |F(x') - F(x)| + |F(x) - F_n(x)| \leq L \|x - x'\| + 2 \sup_x |F_n(x) - F(x)|$$

Como $|F_n(x) - F(x)|$ está acotado para todo x , notemos por $\gamma_n = \sup_x |F_n(x) - F(x)|$, entonces, $|F_n(x') - F_n(x)| \leq L \|x - x'\| + 2\gamma_n$ donde $\gamma_n \rightarrow 0$ pues $F_n \rightarrow F$ completamente.

Corolario 3.6.10. Si tomamos como estimador la media ponderada $\bar{X} = \sum_{i=1}^n \omega_i X_i$ donde las X_i son variables aleatorias independientes idénticamente distribuidas (iid).

Sea $a_n = \sum_{i=1}^n \omega_i^2$. Utilizando el resultado de Fisher (1992) tenemos que si $\limsup_{t \rightarrow \infty} \frac{1}{t} \#\{n : \frac{a_n}{\omega_n^2} < t\} < \infty$, entonces $\frac{S_n}{\sqrt{n}}$ converge en distribución a la normal típica de donde las distribuciones de $\sum_{i=1}^n \frac{\omega_i}{\sqrt{n}} X_i$ son asintóticamente Lipschitz.

3.7. Hipótesis para la consistencia conjunta

Presentamos a continuación, a los efectos de simplificar la lectura, las hipótesis necesarias para obtener la consistencia conjunta de los estimadores de los rasgos y de las CCI.

1) Independencia Local.

2) El rasgo Θ tiene distribuciones marginales uniformes $U[0, 1]$.

3) Para las funciones $g_{n,l}$ existen constantes $c_{n,l,1}, \dots, c_{n,l,n}$ tales que:

$$\sup_{x_1, \dots, x_n, x'_1, \dots, x'_n} \sum_{i=1}^n |g_{n,l}(x_1, \dots, x_i, \dots, x_n) - g_{n,l}(x_1, \dots, x'_i, \dots, x_n)|^2 \leq \sum_{i=1}^n c_{n,l,i}^2 \text{ para } 1 \leq l \leq d$$

4) Las funciones $F_{n,l} = P(g_{n,l}(X) \leq x)$ son asintóticamente Lipschitz, (L_n, γ_n) para toda l con $1 \leq l \leq n$

$$5) \frac{\partial E(g_{n,l}(X)|\Theta_l)}{\partial \Theta_l} > M_l > 0.$$

$$6) \text{ Para } 0 < \alpha < 1/2, \frac{\gamma_n}{\max_l (\sum_{i=1}^n c_{n,l,i}^2)^\alpha} \rightarrow 0$$

$$7) H \max_l (\sum_{i=1}^n c_{n,l,i}^2)^{-\frac{d+2}{2}} < N < Mn^r \text{ con } r > d+1$$

$$8) \lim \log n \max_l (\sum_{i=1}^n c_{n,l,i}^2)^{1-2\alpha} = 0$$

9) El núcleo K es simétrico, acotado y con soporte compacto.

10) Las CCI tienen derivadas parciales continuas en $(0,1)^d$ y además en cada compacto $[a,b]^d \subset (0,1)^d$ y todo par (n,i) están uniformemente acotadas.

$$11) \text{ Para } 0 < \alpha < 1/2, \max_l (\sum_{i=1}^n c_{n,l,i}^2)^\alpha = O(h_n)$$

12) El núcleo K es Lipschitz de constante J.

3.8. Prueba de la consistencia conjunta

Antes de continuar, veamos algunos resultados sobre desigualdades de concentración, que daremos sin demostración y nos ayudarán en lo que sigue. Para una descripción detallada de estos resultados recomendamos leer el excelente trabajo de Boucheron, Lugosi y Massart (2003).

Consideremos un conjunto \mathcal{X} y $g : \mathcal{X}^n \rightarrow R$ una función medible y sea $Z = g(X_1, \dots, X_n)$ donde X_1, \dots, X_n son variables aleatorias independientes (no tienen por qué ser idénticamente distribuidas) que toman valores en \mathcal{X} . Consideremos $E(Z)$ y $E_i(Z) = E(Z | X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n)$.

Sea X'_1, \dots, X'_n una copia independiente de X_1, \dots, X_n y $Z'_i = g(X_1, \dots, X_{i-1}, X'_i, X_{i+1}, \dots, X_n)$.

Se obtienen unos primeros resultados:

Teorema 3.8.1.

$$V(Z) \leq \sum_{i=1}^n E[(Z - E_i(Z))^2]$$

Teorema 3.8.2. Desigualdad de Efron-Stein.

$$V(Z) \leq \frac{1}{2} \sum_{i=1}^n E[(Z - Z'_i)^2]$$

Definición 3.8.3. Diremos que una función $g : \mathcal{X}^n \rightarrow R$ tiene la propiedad de diferencias acotadas si existen constantes no negativas c_1, \dots, c_n , tales que,

$$\sup_{x_1, \dots, x_n, x'_i \in \mathcal{X}} |g(x_1, \dots, x_i, \dots, x_n) - g(x_1, \dots, x'_i, \dots, x_n)| \leq c_i \text{ para } 1 \leq c_i \leq n.$$

Para funciones de diferencias acotadas se tiene el siguiente corolario:

Corolario 3.8.4. Si g tiene la propiedad de diferencias acotadas con constantes c_1, \dots, c_n

$$V(Z) \leq \frac{1}{2} \sum_{i=1}^n c_i^2$$

A partir de la desigualdad logarítmica simétrica de Sobolev, debida a Massart (2000), se prueba que:

Teorema 3.8.5. Si existe una constante C positiva, tal que, casi seguramente

$$\sum_{i=1}^n (Z - Z'_i)^2 \leq C$$

entonces para todo $t > 0$, $P[|Z - E(Z)| > t] \leq 2e^{-\frac{t^2}{4C}}$

Para las funciones de diferencias acotadas se tiene el corolario:

Corolario 3.8.6. Si g tiene la propiedad de diferencias acotadas con constantes c_1, \dots, c_n entonces, $P[|Z - E(Z)| > t] \leq 2e^{-\frac{t^2}{4\sum_{i=1}^n c_i^2}}$.

Esta cota puede ser mejorada y McDiarmid (1989) prueba que, bajo las condiciones del corolario 3.8.6, $P[|Z - E(Z)| > t] \leq 2e^{-\frac{2t^2}{\sum_{i=1}^n c_i^2}}$

En realidad, tampoco es necesario que g sea de diferencias acotadas, pues se puede aplicar el teorema y obtenemos una condición mucho más débil.

Corolario 3.8.7. Si para g existen constantes c_1, \dots, c_n tales que:

$$\sup_{x_1, \dots, x_n, x'_1, \dots, x'_n} \sum_{i=1}^n |g(x_1, \dots, x_i, \dots, x_n) - g(x_1, \dots, x'_i, \dots, x_n)|^2 \leq \sum_{i=1}^n c_i^2$$

$$P[|Z - E(Z)| > t] \leq 2e^{-\frac{t^2}{4\sum_{i=1}^n c_i^2}}$$

Consideremos de ahora en adelante en R^d la norma $\|x\| = \max_{i=1 \dots d} |x_i|$

Lema 3.8.8. Sea $G_{g_n}(x)$ definida por 3.1. Supongamos que para las funciones $g_{n,l}$ existen constantes $c_{n,l,1}, \dots, c_{n,l,n}$ tales que:

$$(3.20) \quad \sup_{x_1, \dots, x_n, x'_1, \dots, x'_n} \sum_{i=1}^n |g_{n,l}(x_1, \dots, x_i, \dots, x_n) - g_{n,l}(x_1, \dots, x'_i, \dots, x_n)|^2 \leq \sum_{i=1}^n c_{n,l,i}^2 \text{ para } 1 \leq l \leq d$$

Supongamos, además, que las $F_{n,l}$ correspondientes son asintóticamente Lipschitz, (L_n, γ_n) para toda l con $1 \leq l \leq n$. Entonces para $\epsilon > 0$:

$P(\|G_{g_n}(g_n(X)) - G_{g_n}(E(g_n(X)))\| > \epsilon) \leq \sum_{l=1}^d 2e^{-\frac{(\epsilon - \gamma_n)^2}{4L_n^2 \sum_{i=1}^n c_{n,l,i}^2}}$. En particular si $c_n = \max_{1 \leq l \leq d, 1 \leq i \leq n} c_{n,l,i}$

$$P(\|G_{g_n}(g_n(X)) - G_{g_n}(E(g_n(X)))\| > \epsilon) \leq 2de^{-\frac{(\epsilon - \gamma_n)^2}{4L_n^2 nc_n^2}}$$

Observación: En el caso unidimensional $g_n(x_1, \dots, x_n) = \bar{x} = \frac{\sum_{i=1}^n x_i}{n}$. Ahora $|g_{n,l}(x_1, \dots, x_i, \dots, x_n) - g_{n,l}(x_1, \dots, x'_i, \dots, x_n)| = \frac{1}{n}$, entonces,

$$\sup_{x_1, \dots, x_n, x'_1, \dots, x'_n} \sum_{i=1}^n |g_{n,l}(x_1, \dots, x_i, \dots, x_n) - g_{n,l}(x_1, \dots, x'_i, \dots, x_n)|^2 \leq \sum_{i=1}^n \frac{1}{n^2} = \frac{1}{n} \text{ luego,}$$

$$P(\|G_{g_n}(g_n(X)) - G_{g_n}(E(g_n(X)))\| > \epsilon) = P(\|G_{g_n}(\bar{X}) - G_{g_n}(E(\bar{X}))\| > \epsilon) \leq 2e^{-\frac{(\epsilon - \frac{K}{n})^2}{4K^2 n \frac{1}{n^2}}} = 2e^{-\frac{n(\epsilon - \frac{K}{n})^2}{4K^2}}$$

$$\begin{aligned} \text{Demostración: } P(\|G_{g_n}(g_n(X)) - G_{g_n}(E(g_n(X)))\| > \epsilon) &\leq P(\|g_n(X) - E(g_n(X))\| > \frac{\epsilon - \gamma_n}{L_n}) = \\ &= P(\max_{1 \leq l \leq d} |g_{n,l}(X) - E(g_{n,l}(X))| > \frac{\epsilon - \gamma_n}{L_n}) = P(\bigcup_{1 \leq l \leq d} |g_{n,l}(X) - E(g_{n,l}(X))| > \frac{\epsilon - \gamma_n}{L_n}) \leq \\ &\leq \sum_{l=1}^d P(|g_{n,l}(X) - E(g_{n,l}(X))| > \frac{\epsilon - \gamma_n}{L_n}) = \end{aligned}$$

$$= \sum_{l=1}^d \int_{[0,1]^d} P(|g_{n,l}(X) - E(g_{n,l}(X))| > \frac{\epsilon - \gamma_n}{L_n} \mid \Theta = \theta) f_\theta(\theta) d\theta \leq \sum_{l=1}^d 2e^{-\frac{(\epsilon - \gamma_n)^2}{4L_n^2 \sum_{i=1}^n c_{n,l,i}^2}}$$

Si $c_n = \max_{1 \leq l \leq d, 1 \leq i \leq n} c_{n,l,i}$, entonces,

$$P(\|G_{g_n}(g_n(X)) - G_{g_n}(E(g_n(X)))\| > \epsilon) \leq \sum_{l=1}^d 2e^{-\frac{(\epsilon - \gamma_n)^2}{2L_n^2 nc_n^2}} = 2de^{-\frac{(\epsilon - \gamma_n)^2}{2L_n^2 nc_n^2}}$$

Lema 3.8.9. Si para toda función $g_{n,l}$ $1 \leq l \leq d$ se cumple la condición dada por la ecuación 3.33, y además, $\frac{E(g_{n,l}(X)|\Theta_l)}{\partial \Theta_l} > M_l > 0$. Sea λ_n una sucesión positiva que tiende a 0, tal que para $0 < \alpha < 1/2$ $\max_l (\sum_{i=1}^n c_{n,l,i}^2)^\alpha = O(\lambda_n)$ entonces para todo $\theta \in [0, 1]^d$:

$$(3.21) \quad \| G_{g_n}(E(g_n(X)|\Theta = \theta)) - \theta \| < \lambda_n$$

Demostración: Sea $\Theta = (\Theta_1, \dots, \Theta_d)$ y $\theta = (\theta_1, \dots, \theta_d)$. Tenemos que:

$$G_{g_n}(E(g_n(X)|\Theta = \theta)) = (P(g_{n,1} \leq E(g_{n,1}(X)|\Theta = \theta)), \dots, P(g_{n,d} \leq E(g_{n,d}(X)|\Theta = \theta))) = \\ = (P(g_{n,1} \leq E(g_{n,1}(X)|\Theta_1 = \theta_1)), \dots, P(g_{n,d} \leq E(g_{n,d}(X)|\Theta_d = \theta_d)))$$

$$\text{Luego, } \| G_{g_n}(E(g_n(X)|\Theta = \theta)) - \theta \| = \max_{1 \leq l \leq d} |P(g_{n,l} \leq E(g_{n,l}(X)|\Theta_l = \theta_l)) - \theta_l|$$

Trabajando con la componente l -ésima y considerando que las marginales son uniformes:

$$P(g_{n,l} \leq E(g_{n,l}(X)|\Theta_l = \theta_l)) = \int_0^1 P(g_{n,l} \leq E(g_{n,l}(X)|\Theta_l = \theta_l) \mid \Theta_l = \theta_l^*) d\theta_l^* = \\ = \int_0^{\theta_l + \frac{\lambda_n}{2}} P(g_{n,l} \leq E(g_{n,l}(X)|\Theta_l = \theta_l) \mid \Theta_l = \theta_l^*) d\theta_l^* + \int_{\theta_l + \frac{\lambda_n}{2}}^1 P(g_{n,l} \leq$$

$$E(g_{n,l}(X)|\Theta_l = \theta_l) \mid \Theta_l = \theta_l^*) d\theta_l^* \leq \theta_l + \frac{\lambda_n}{2} + \int_{\theta_l + \frac{\lambda_n}{2}}^1 P(g_{n,l} \leq E(g_{n,l}(X)|\Theta_l = \theta_l) \mid \Theta_l = \theta_l^*) d\theta_l^*$$

Tenemos que: $\theta_l^* - \theta_l > \frac{\lambda_n}{2}$ y $E(g_{n,l}(X)|\Theta_l = \theta_l) - E(g_{n,l}(X)|\Theta_l = \theta_l^*) < 0$ pues $E(g_{n,l}(X)|\Theta_l)$ es creciente.

Además, $E(g_{n,l}(X)|\Theta_l = \theta_l^*) - E(g_{n,l}(X)|\Theta_l = \theta_l) = [\frac{\partial E(g_{n,l}(X)|\Theta_l)}{\partial \Theta_l}](c)(\theta_l^* - \theta_l) > M_l \frac{\lambda_n}{2}$ con $c \in (\theta_l, \theta_l^*)$, de donde

$$g_{n,l}(X) - E(g_{n,l}(X)|\Theta_l = \theta_l^*) \leq E(g_{n,l}(X)|\Theta_l = \theta_l) - E(g_{n,l}(X)|\Theta_l = \theta_l^*) \text{ si y solo si}$$

$$|g_{n,l}(X) - E(g_{n,l}(X)|\Theta_l = \theta_l^*)| \geq E(g_{n,l}(X)|\Theta_l = \theta_l^*) - E(g_{n,l}(X)|\Theta_l = \theta_l) > M_l \frac{\lambda_n}{2}$$

$$\text{Luego, } \int_{\theta_l + \frac{\lambda_n}{2}}^1 P(g_{n,l}(X) \leq E(g_{n,l}(X)|\Theta_l = \theta_l) \mid \Theta_l = \theta_l^*) d\theta_l^* =$$

$$= \int_{\theta_l + \frac{\lambda_n}{2}}^1 P(g_{n,l}(X) - E(g_{n,l}(X)|\Theta_l = \theta_l^*) \leq E(g_{n,l}(X)|\Theta_l = \theta_l) - E(g_{n,l}(X)|\Theta_l = \theta_l^*) \mid \Theta_l = \theta_l^*) d\theta_l^* <$$

$$< \int_{\theta_l + \frac{\lambda_n}{2}}^1 P(|g_{n,l}(X) - E(g_{n,l}(X)|\Theta_l = \theta_l^*)| > M_l \frac{\lambda_n}{2} \mid \Theta_l = \theta_l^*) d\theta_l^* \leq 2e^{-\frac{M_l^2 \lambda_n^2}{16 \sum_{i=1}^n c_{n,l,i}^2}} (1 - \theta_l - \frac{\lambda_n}{2})$$

De esto obtenemos que:

$$P(g_{n,l} \leq E(g_{n,l}(X)|\Theta_l = \theta_l)) < \theta_l + \frac{\lambda_n}{2} + 2e^{-\frac{M_l^2 \lambda_n^2}{16 \sum_{i=1}^n c_{n,l,i}^2}} (1 - \theta_l - \frac{\lambda_n}{2})$$

$$P(g_{n,l} \leq E(g_{n,l}(X)|\Theta_l = \theta_l)) - \theta_l < \frac{\lambda_n}{2} + 2e^{-\frac{M_l^2 \lambda_n^2}{16 \sum_{i=1}^n c_{n,l,i}^2}}$$

para todo l con $1 \leq l \leq d$. Así pues, como $\max_l (\sum_{i=1}^n c_{n,l,i}^2)^\alpha = O(\lambda_n)$ con $0 < \alpha < 1/2$, existe K , tal que $\frac{\max_l (\sum_{i=1}^n c_{n,l,i}^2)^\alpha}{\lambda_n} < K$, entonces para todo l , tenemos $\lambda_n > \frac{1}{K} (\sum_{i=1}^n c_{n,l,i}^2)^\alpha$, de donde,

$$\frac{M_l^2 \lambda_n^2}{16 \sum_{i=1}^n c_{n,l,i}^2} > \frac{M_l^2}{16 K^2} (\sum_{i=1}^n c_{n,l,i}^2)^{2\alpha-1}$$

A partir de esto, $e^{-\frac{M_l^2 \lambda_n^2}{16 \sum_{i=1}^n c_{n,l,i}^2}} < e^{-\frac{M_l^2}{16 K^2} (\sum_{i=1}^n c_{n,l,i}^2)^{2\alpha-1}}$ y luego

$$\frac{e^{-\frac{M_l^2 \lambda_n^2}{16 \sum_{i=1}^n c_{n,l,i}^2}}}{\lambda_n} < K \frac{e^{-\frac{M_l^2}{16 K^2} (\sum_{i=1}^n c_{n,l,i}^2)^{2\alpha-1}}}{(\sum_{i=1}^n c_{n,l,i}^2)^\alpha} \rightarrow 0, \text{ ya que } 2\alpha - 1 < 0 \text{ y } \sum_{i=1}^n c_{n,l,i}^2 \rightarrow 0.$$

Así para n suficientemente grande: $2e^{-\frac{M_l^2 \lambda_n^2}{16 \sum_{i=1}^n c_{n,l,i}^2}} < \frac{\lambda_n}{2}$

Entonces para $n \geq n_0$: $P(g_{n,l} \leq E(g_{n,l}(X)|\Theta_l = \theta_l)) - \theta_l < \lambda_n$.

Teorema 3.8.10. Supongamos que se cumplen las hipótesis del lema 3.8.8 y del lema 3.8.9 para $0 < \alpha < 1/2$ y $\frac{\gamma_n}{\max_l(\sum_{i=1}^n c_{n,l,i}^2)^\alpha} \rightarrow 0$

1) Si $\max_l(\sum_{i=1}^n c_{n,l,i}^2)^{-1} < N < e^{\frac{32L_n^2+1}{144K^2L_n^2} \max_l(\sum_{i=1}^n c_{n,l,i}^2)^{\alpha-\frac{1}{2}}}$

$\max_{k=1,\dots,N} \|\hat{\Theta}_{n,k} - \Theta_{n,k}\| \rightarrow 0$ en probabilidad.

2) Si además $\lim \log n \max_l(\sum_{i=1}^n c_{n,l,i}^2)^{1-2\alpha} = 0$, entonces, $\max_{k=1,\dots,N} \|\hat{\Theta}_{n,k} - \Theta_{n,k}\| \rightarrow 0$ completamente (en particular con probabilidad 1).

Demostración: Consideremos una sucesión $\epsilon_n \rightarrow 0$ positiva, tal que $\max_l(\sum_{i=1}^n c_{n,l,i}^2)^\alpha = O(\epsilon_n)$. (siendo K una de las cotas).

$$\begin{aligned} P(\max_{k=1,\dots,N} \|\hat{\Theta}_{n,k} - \Theta_{n,k}\| > \epsilon_n) &= P(\bigcup_{k=1}^N \|\hat{\Theta}_{n,k} - \Theta_{n,k}\| > \epsilon_n) \leq \\ &\leq \sum_{k=1}^N P(\|\hat{\Theta}_{n,k} - \Theta_{n,k}\| > \epsilon_n) = \sum_{k=1}^N P(\|\hat{G}_N(g_n(X_k)) - \Theta_{n,k}\| > \epsilon_n) \leq \\ &\leq \sum_{k=1}^N P(\|\hat{G}_N(g_n(X_k)) - G_{g_n}(g_n(X_k))\| > \frac{\epsilon_n}{3}) + \\ &+ \sum_{k=1}^N P(\|G_{g_n}(g_n(X_k)) - G_{g_n}(E(g_n(X_k)))\| > \frac{\epsilon_n}{3}) + \sum_{k=1}^N P(\|G_{g_n}(E(g_n(X_k))) - \Theta_{n,k}\| > \frac{\epsilon_n}{3}) \end{aligned}$$

Ahora:

$$\begin{aligned} P(\|\hat{G}_N(g_n(X_k)) - G_{g_n}(g_n(X_k))\| > \frac{\epsilon_n}{3}) &= P(\max_{l=1,\dots,d} |\hat{G}_{N,l}(g_n(X_k)) - G_{g_n,l}(g_n(X_k))| > \frac{\epsilon_n}{3}) \leq \\ &\leq \sum_{l=1}^d P(|\hat{G}_{N,l}(g_n(X_k)) - G_{g_n,l}(g_n(X_k))| > \frac{\epsilon_n}{3}) = \end{aligned}$$

$$\sum_{l=1}^d P(|\hat{F}_{N,l}(g_n(X_k)) - P(g_{n,l}(X) \leq g_{n,l}(X_k))| > \frac{\epsilon_n}{3}) \leq dCe^{-2\frac{\epsilon_n^2 N}{9}}$$

Para el segundo término, aplicamos el lema 3.8.8.

$$P(\|G_{n,k}(g_n(X)) - G_{n,k}(E(g_n(X)))\| > \frac{\epsilon_n}{3}) \leq 2 \sum_{l=1}^d e^{-\frac{(\frac{\epsilon_n}{3} - \gamma_n)^2}{4L_n^2 \sum_{i=1}^n c_{n,l,i}^2}}$$

El último término, se anula para $n \geq n_0$. Así hemos obtenido que:

$$(3.22) \quad P(\max_{k=1,\dots,N} \|\hat{\Theta}_{n,k} - \Theta_{n,k}\| > \epsilon_n) \leq NdCe^{-2\frac{\epsilon_n^2 N}{9}} + 2N \sum_{l=1}^d e^{-\frac{(\frac{\epsilon_n}{3} - \gamma_n)^2}{4L_n^2 \sum_{i=1}^n c_{n,l,i}^2}} \text{ o sea,}$$

$$(3.23) \quad P(\max_{k=1,\dots,N} \|\hat{\Theta}_{n,k} - \Theta_{n,k}\| > \epsilon_n) \leq Nd \left(Ce^{-2\frac{\epsilon_n^2 N}{9}} + 2e^{-\frac{\epsilon_n^2 (1-3\frac{\gamma_n}{\epsilon_n})^2}{36L_n^2 \max_l \sum_{i=1}^n c_{n,l,i}^2}} \right).$$

Si se cumple cualquiera de las dos condiciones, entonces: $e^{-2\frac{\epsilon_n^2 N}{9}} < e^{-\frac{2 \max_l(\sum_{i=1}^n c_{n,l,i}^2)^{2\alpha-1}}{9K^2}}$, ya que para n suficientemente grande $3\frac{\gamma_n}{\epsilon_n} < 1/2$ para el segundo término tenemos:

$$e^{-\frac{\epsilon_n^2 (1-3\frac{\gamma_n}{\epsilon_n})^2}{36L_n^2 \max_l \sum_{i=1}^n c_{n,l,i}^2}} < e^{-\frac{\max_l(\sum_{i=1}^n c_{n,l,i}^2)^{2\alpha-1}}{144K^2L_n^2}}. \text{ Así pues,}$$

$$P(\max_{k=1,\dots,N} \|\hat{\Theta}_{n,k} - \Theta_{n,k}\| > \epsilon_n) < N \left(C_1 e^{-\frac{32L_n^2+1}{144K^2L_n^2} \max_l(\sum_{i=1}^n c_{n,l,i}^2)^{2\alpha-1}} \right)$$

En el caso 1: $P(\max_{k=1,\dots,N} \|\hat{\Theta}_{n,k} - \Theta_{n,k}\| > \epsilon_n) < C_1 e^{-\frac{32L_n^2+1}{144K^2L_n^2} (\max_l(\sum_{i=1}^n c_{n,l,i}^2)^{2\alpha-1} - \max_l(\sum_{i=1}^n c_{n,l,i}^2)^{\alpha-\frac{1}{2}})}$

o sea, $P(\max_{k=1,\dots,N} \|\hat{\Theta}_{n,k} - \Theta_{n,k}\| > \epsilon_n) \rightarrow 0$, lo cual implica la convergencia en probabilidad.

Si se cumple 2, obtenemos que:

$$\lim_{n \rightarrow \infty} \frac{\frac{32L_n^2+1}{144K^2L_n^2} (\max_l (\sum_{i=1}^n c_{n,l,i}^2)^{2\alpha-1} - \max_l (\sum_{i=1}^n c_{n,l,i}^2)^{\alpha-\frac{1}{2}})}{\log n} = +\infty$$

lo que implica la convergencia completa pues si tenemos una serie: $\sum_{n=1}^{\infty} e^{-f(n)}$ y

$Q(n) = \frac{f(n)}{\log n} \rightarrow +\infty$ entonces la serie converge, pues, $e^{-f(n)} = e^{\log n^{-Q(n)}} = n^{-Q(n)}$ y si $Q(n) \rightarrow +\infty$ la $\sum_{n=1}^{\infty} n^{-Q(n)}$ es convergente.

Las series exponenciales convergen y la convergencia con probabilidad 1 se deduce de aplicar el lema de Borel Cantelli.

Observación: Si consideramos el caso unidimensional con $g_n(X) = \bar{X}$ y $nP(\sum_{i=1}^n X_i = k) < L$ vemos que: $\gamma_n = \frac{1}{n}$ y $c_n = \frac{L}{n}$, por lo que la condición pedida se cumple ya que $\frac{1/n}{1/n^\alpha} \rightarrow 0$.

Si consideramos ϵ_n tal que $\frac{1}{n} = O(\epsilon_n)$ entonces si $n < N < e^{\frac{32L^2+1}{144K^2L^2} n^{\frac{1}{2}-\alpha}}$ y dado que $\frac{\log n}{n^{1-2\alpha}} \rightarrow 0$ entonces, $\max_{k=1, \dots, N} \|\hat{\Theta}_{n,k} - \Theta_{n,k}\| \rightarrow 0$ completamente.

Por la forma de los estimadores se pueden producir empates; romperemos los empates sumando a $g_{n,l}$ una variable aleatoria uniforme.

Consideremos una sucesión a_n por ejemplo $a_n = \min_l \min_{X, X'} |g_{n,l}(X) - g_{n,l}(X')|$ con X y X' posibles vectores de respuestas. La sucesión debe cumplir $a_n < \max_l (\sum_{i=1}^n c_{n,l,i}^2)$

Sea: $Z_n = g_n(X) + W$ con W una variable aleatoria con distribución uniforme $U[0, a(n)]^d$ y consideremos las empíricas:

$$\hat{J}_{N,l}(x) = \frac{\#\{Z_{n,l} \leq x\}}{N} = \frac{\sum_{i=1}^N \chi_{\{Z_{n,l} \leq x\}}}{N}$$

con $l = 1, \dots, d$. Definamos la función: $\hat{T}_N(x_1, \dots, x_d) = (\hat{J}_{N,1}(x_1), \dots, \hat{J}_{N,d}(x_d))$

Entonces:

$$(3.24) \quad \hat{\Theta}_n^* = \hat{T}_N(Z_n)$$

Lema 3.8.11. Supongamos que para $0 < \alpha < 1/2$ $\frac{\gamma_n}{\max_l (\sum_{i=1}^n c_{n,l,i}^2)^\alpha} \rightarrow 0$

Si $\epsilon_n \rightarrow 0$ es una sucesión positiva, tal que: $\max_l (\sum_{i=1}^n c_{n,l,i}^2)^\alpha = O(\epsilon_n)$ (siendo K una de las cotas). Entonces existe una constante C_1 independiente de n , tal que:

$$P(\|\hat{\Theta}_n - \hat{\Theta}_n^*\| > \epsilon_n) \leq dC_1 e^{-\frac{2\epsilon_n^2 N}{9}}$$

Demostración:

$$P(\|\hat{\Theta}_n - \hat{\Theta}_n^*\| > \epsilon_n) = P(\|\hat{G}_N(g_n(X)) - \hat{T}_N(Z_n)\| > \epsilon_n) = P(\max_{1 \leq l \leq d} |\hat{F}_{N,l}(g_n(X)) - \hat{J}_{N,l}(Z_n)| > \epsilon_n).$$

Trabajemos con cada componente.

$$\begin{aligned} \hat{F}_{N,l}(x - a_n) &= \frac{\#\{g_{n,l}(X) \leq x - a_n\}}{N} \leq \frac{\#\{Z_{n,l} \leq x\}}{N} = \hat{J}_{N,l}(x) \leq \hat{J}_{N,l}(x + a_n) = \\ &= \frac{\#\{Z_{n,l} \leq x + a_n\}}{N} \leq \frac{\#\{g_{n,l}(X) \leq x + a_n\}}{N} = \hat{F}_{N,l}(x + a_n) \end{aligned}$$

Luego, $\hat{F}_{N,l}(g_{n,l}(X) - a_n) \leq \hat{J}_{N,l}(Z_n) \leq \hat{F}_{N,l}(g_{n,l}(X) + a_n)$. De aquí,

$$|\hat{J}_{N,l}(Z_n) - \hat{F}_{N,l}(g_{n,l}(X))| \leq \hat{F}_{N,l}(g_{n,l}(X) + a_n) - \hat{F}_{N,l}(g_{n,l}(X) - a_n)$$

A partir de esto,

$$\begin{aligned} P(|\hat{J}_{N,l}(Z_n) - \hat{F}_{N,l}(g_{n,l}(X))| > \epsilon_n) &\leq P(\hat{F}_{N,l}(g_{n,l}(X) + a_n) - \hat{F}_{N,l}(g_{n,l}(X) - a_n) > \epsilon_n) \leq \\ &\leq P(|\hat{F}_{N,l}(g_{n,l}(X) + a_n) - F_{N,l}(g_{n,l}(X) + a_n)| > \frac{\epsilon_n}{3}) + P(|F_{N,l}(g_{n,l}(X) + a_n) - F_{N,l}(g_{n,l}(X) - a_n)| > \frac{\epsilon_n}{3}) + \end{aligned}$$

$$P(|\widehat{F}_{N,l}(g_{n,l}(X) - a_n) - F_{N,l}(g_{n,l}(X) - a_n)| > \frac{\epsilon_n}{3})$$

Sabemos que:

$$P(|\widehat{F}_{N,l}(g_{n,l}(X) + a_n) - F_{N,l}(g_{n,l}(X) + a_n)| > \frac{\epsilon}{3}) \leq Ce^{\frac{-2\epsilon_n^2 N}{9}} y$$

$$P(|\widehat{F}_{N,l}(g_{n,l}(X) - a_n) - F_{N,l}(g_{n,l}(X) - a_n)| > \frac{\epsilon}{3}) \leq Ce^{\frac{-2\epsilon_n^2 N}{9}}$$

$$F_{N,l}(g_{n,l}(X) + a_n) - F_{N,l}(g_{n,l}(X) - a_n) \leq L_n(g_{n,l}(X) + a_n - g_{n,l}(X) - a_n) + \gamma_n = 2a_n L_n + \gamma_n$$

Pero para $n > n_0$ tenemos que: $\gamma_n < \frac{\max_l(\sum_{i=1}^n c_{n,l,i}^2)^\alpha}{6K} < \frac{\epsilon_n}{6}$ y $a(n) < \frac{\epsilon_n}{12L_n}$, luego para $n > n_0$, el último término se anula, así pues tenemos que existe C_1 tal que para todo n :

$$P(\|\widehat{\Theta}_n - \widehat{\Theta}_n^*\| > \epsilon_n) \leq dC_1 e^{\frac{-2\epsilon_n^2 N}{9}}.$$

Obtenemos como corolario que:

Corolario 3.8.12. Supongamos que para $0 < \alpha < 1/2$ $\frac{\gamma_n}{\max_l(\sum_{i=1}^n c_{n,l,i}^2)^\alpha} \rightarrow 0$.

1) Si $\max_l(\sum_{i=1}^n c_{n,l,i}^2)^{-1} < N < e^{\frac{32L_n^2+1}{144K^2L_n^2} \max_l(\sum_{i=1}^n c_{n,l,i}^2)^{\alpha-\frac{1}{2}}}$ en probabilidad

$$\max_{k=1,\dots,N} \|\widehat{\Theta}_{n,k}^* - \Theta_{n,k}\| \rightarrow 0$$

2) Si además $\lim \log n \max_l(\sum_{i=1}^n c_{n,l,i}^2)^{1-2\alpha} = 0$, entonces

$$\max_{k=1,\dots,N} \|\widehat{\Theta}_{n,k}^* - \Theta_{n,k}\| \rightarrow 0$$

completamente (en particular con probabilidad 1).

Demostración: Consideremos una sucesión $\epsilon_n \rightarrow 0$ positiva, tal que: $\max_l(\sum_{i=1}^n c_{n,l,i}^2)^\alpha = O(\epsilon_n)$ (siendo K una de las cotas).

$$\|\widehat{\Theta}_{n,k}^* - \Theta_{n,k}\| \leq \|\widehat{\Theta}_{n,k}^* - \widehat{\Theta}_{n,k}\| + \|\widehat{\Theta}_{n,k} - \Theta_{n,k}\| \text{ luego}$$

$$\begin{aligned} P(\|\widehat{\Theta}_{n,k}^* - \Theta_{n,k}\| > \epsilon_n) &\leq P(\|\widehat{\Theta}_{n,k}^* - \widehat{\Theta}_{n,k}\| > \epsilon_n) + P(\|\widehat{\Theta}_{n,k} - \Theta_{n,k}\| > \epsilon_n) \leq \\ &\leq 2dCe^{\frac{-2\epsilon_n^2 N}{9}} + P(\|\widehat{\Theta}_{n,k} - \Theta_{n,k}\| > \epsilon_n) \end{aligned}$$

Aplicamos el teorema 3.8.10 a los términos $dC_1 e^{\frac{-2\epsilon_n^2 N}{9}}$ y $P(\|\widehat{\Theta}_{n,k} - \Theta_{n,k}\| > \epsilon_n)$ de donde obtenemos el resultado.

Lema 3.8.13. Consideremos que se cumplen las hipótesis del teorema 3.8.10 y que el núcleo K es simétrico acotado y con soporte compacto C_K . Supongamos que las CCI son Lipchitz de constante M_V en un conjunto V compacto con $V \subset [0, 1]^d$. Además, para $0 < \alpha < 1/2$

$$N > \max_l(\sum_{i=1}^n c_{n,l,i}^2)^{-\alpha(d+2)} \text{ y } \max_l(\sum_{i=1}^n c_{n,l,i}^2)^\alpha = O(h_n)$$

Existe un entero n_0 tal que para todo $n > n_0$ y siendo $r = \max\{\|x\| \mid x \in C_K\}$ la siguiente desigualdad ocurre para todo ítem i y $\theta_0 \in U \subset V$, U abierto.

(3.25)

$$P(|\widehat{P}_{n,i}(\theta_0) - P_{n,i}(\theta_0)| > 3M_V h_n r) \leq e^{-8 \frac{m_l^2 M_V^2 N n T h_n^{d+2}}{9 \|K\|_\infty^2 L^2 r^{d-2}}} + N_n(C_1 e^{-\frac{32L_n^2+1}{144K^2L_n^2} \max_l(\sum_{i=1}^n c_{n,l,i}^2)^{2\alpha-1}})$$

Demostración: Como $\frac{\sum_{k=1}^{N_n} K(\frac{\theta - \hat{\theta}_{n,i,k}}{h_n})}{N_n h_n^d}$ tiene límite finito con probabilidad 1. Existen m_l y m_u tal que para n suficientemente grande y $\theta \in (0, 1)^d$:

$$(3.26) \quad m_l N_n h_n^d < \sum_{k=1}^{N_n} K(\frac{\theta - \hat{\theta}_{n,i,k}}{h_n}) < m_u N_n h_n^d$$

en un conjunto de probabilidad 1.

Sabemos que $K(\frac{\theta - \hat{\theta}}{h_n}) > 0$ si y solo si $\frac{\theta - \hat{\theta}}{h_n} \in C_K$. Como C_K es compacto $C_K \subseteq B_K = \overline{B(0, r)}$ con $r = \max\{\|x\| \mid x \in C_K\}$

Sea: $A_{n,i,\theta} = \{k : \frac{\theta - \hat{\theta}_{n,i,k}}{h_n} \in B_K\} = \{k : \hat{\theta}_{n,i,k} \in \theta - h_n B_K\} = \{k : \hat{\theta}_{n,i,k} \in \overline{B(\theta, h_n r)}\}$ entonces,

$$(3.27) \quad \#A_{n,i,\theta} \rightarrow N_n \mu_{\hat{\theta}}(\overline{B(\theta, h_n r)})$$

donde, $\mu_{\hat{\theta}}$ es la distribución de $\hat{\theta}$, así pues en todo abierto U de $(0, 1)^d$ existe n_0 tal que para todo $\theta \in U$ y $n > n_0$, $|\#A_{n,i,\theta} - N_n \mu_{\hat{\theta}}(\overline{B(\theta, h_n r)})| < \epsilon$ de donde

$$(3.28) \quad \frac{N_n}{2} \mu_{\hat{\theta}}(\overline{B(\theta, h_n r)}) < \#A_{n,i,\theta} < \frac{3N_n}{2} \mu_{\hat{\theta}}(\overline{B(\theta, h_n r)})$$

Sea $\theta_0 \in U$ y el suceso $B_{n,i} = \{\hat{\theta}_j - \theta_j \in h_n B_K \text{ para } j = 1, \dots, N_n\}$

$$P(|\hat{P}_{n,i}(\theta_0) - P_{n,i}(\theta_0)| > 3M_V h_n r) = P(|\hat{P}_{n,i}(\theta_0) - P_{n,i}(\theta_0)| > 3M_V h_n r | B_{n,i}) P(B_{n,i}) + \\ + P(|\hat{P}_{n,i}(\theta_0) - P_{n,i}(\theta_0)| > 3M_V h_n r | B_{n,i}^c) P(B_{n,i}^c)$$

luego, $P(|\hat{P}_{n,i}(\theta_0) - P_{n,i}(\theta_0)| > 3M_V h_n r) \leq P(|\hat{P}_{n,i}(\theta_0) - P_{n,i}(\theta_0)| > 3M_V h_n r | B_{n,i}) + P(B_{n,i}^c)$

Como $\max_l (\sum_{i=1}^n c_{n,l,i}^2)^\alpha = O(h_n)$ sabemos que:

$$P(B_{n,i}^c) = P(\max_{k=1, \dots, N} \|\hat{\Theta}_{n,k} - \Theta_{n,k}\| > h_n r) < N \left(C_1 e^{-\frac{32L_n^2+1}{144K^2L_n^2} \max_l (\sum_{i=1}^n c_{n,l,i}^2)^{2\alpha-1}} \right)$$

Tenemos, entonces: $P(|\hat{P}_{n,i}(\theta_0) - P_{n,i}(\theta_0)| > 3M_V h_n r) \leq$

$$\leq P(|\hat{P}_{n,i}(\theta_0) - P_{n,i}(\theta_0)| > 3M_V h_n r | B_{n,i}) + N \left(C_1 e^{-\frac{32L_n^2+1}{144K^2L_n^2} \max_l (\sum_{i=1}^n c_{n,l,i}^2)^{2\alpha-1}} \right)$$

Sea $\Theta_n = (\Theta_{n,1}, \dots, \Theta_{n,N_n})$ y análogamente $\hat{\Theta}_{n,i}$ y sean θ_n y $\hat{\theta}_{n,i}$ realizaciones de estos vectores. Sea W la distribución conjunta de estos vectores condicionado a $B_{n,i}$ Tenemos que:

$$P(|\hat{P}_{n,i}(\theta_0) - P_{n,i}(\theta_0)| > 3M_V h_n r | B_{n,i}) = \int P(|\hat{P}_{n,i}(\theta_0) - P_{n,i}(\theta_0)| > 3M_V h_n r | B_{n,i}, \hat{\theta}_{n,i}, \theta_n) dW(\hat{\theta}_{n,i}, \theta_n)$$

$$\text{Ahora: } P(|\hat{P}_{n,i}(\theta_0) - P_{n,i}(\theta_0)| > 3M_V h_n r | B_{n,i}, \hat{\theta}_{n,i}, \theta_n) \leq \\ \leq P(|\hat{P}_{n,i}(\theta_0) - E(\hat{P}_{n,i}(\theta_0))| + |E(\hat{P}_{n,i}(\theta_0)) - P_{n,i}(\theta_0)| > 3M_V h_n r | B_{n,i}, \hat{\theta}_{n,i}, \theta_n)$$

Esta esperanza condicional la podemos calcular como:

$$E(\hat{P}_{n,i}(\theta_0) | B_{n,i}, \hat{\theta}_{n,i}, \theta_n) = \frac{1}{\sum_{j \in A_{n,i,\theta_0}} K(\frac{\theta_0 - \hat{\theta}_{n,i,j}}{h_n})} \sum_{j \in A_{n,i,\theta_0}} K(\frac{\theta_0 - \hat{\theta}_{n,i,j}}{h_n}) E(X_{n,i,k} | \theta_{n,j}) = \\ = \frac{1}{\sum_{j \in A_{n,i,\theta_0}} K(\frac{\theta_0 - \hat{\theta}_{n,i,j}}{h_n})} \sum_{k \in A_{n,i,\theta_0}} K(\frac{\theta_0 - \hat{\theta}_{n,i,k}}{h_n}) P_{n,j}(\theta_{n,k})$$

Para todo $k \in A_{n,i,\theta_0}$ $\|\theta_0 - \theta_{n,k}\| \leq \|\theta_0 - \hat{\theta}_{n,-i,k}\| + \|\hat{\theta}_{n,-i,k} - \theta_{n,k}\| \leq h_n r + h_n r = 2h_n r$

Eligiendo n suficientemente grande para que $2h_nr < \inf\{\|x - y\| \mid x \in \partial U, y \in \partial V\}$ entonces que $\|\theta_0 - \theta_{n,k}\| \leq 2h_nr$ implica que para todo $k \in A_{n,i,\theta_0}$, $\theta_{n,k} \in V$. Luego, por ser las CCI Lipchitz en V :

$$|P_{n,i}(\theta_0) - P_{n,j}(\theta_{n,k})| \leq M_V \|\theta_0 - \theta_{n,k}\| \leq 2M_V h_n r$$

Esto implica que: $|E(\hat{P}_{n,i}(\theta_0)|B_{n,i}, \hat{\theta}_{n,i}, \theta_n) - P_{n,i}(\theta_0)| =$

$$= \left| \frac{1}{\sum_{k \in A_{n,i,\theta_0}} K(\frac{\theta_0 - \hat{\theta}_{n,i,k}}{h_n})} \sum_{k \in A_{n,i,\theta_0}} K(\frac{\theta_0 - \hat{\theta}_{n,i,k}}{h_n}) [P_{n,j}(\theta_{n,k}) - P_{n,i}(\theta_0)] \right| < 2M_V h_n r$$

Entonces para n suficiente grande:

$$P(|\hat{P}_{n,i}(\theta_0) - P_{n,i}(\theta_0)| > 3M_V h_n r | B_{n,i}, \hat{\theta}_{n,i}, \theta_n) \leq P(|\hat{P}_{n,i}(\theta_0) - E(\hat{P}_{n,i}(\theta_0))| > M_V h_n r | B_{n,i}, \hat{\theta}_{n,i}, \theta_n)$$

Una expresión alternativa para $\hat{P}_{n,i}(\theta_0)$ viene dada por:

$$\hat{P}_{n,i}(\theta_0) = \frac{1}{\#A_{n,i,\theta_0}} \sum_{j \in A_{n,i,\theta_0}} w_j X_{n,i,k} \text{ con } w_j = \frac{\#A_{n,i,\theta_0} K(\frac{\theta_0 - \hat{\theta}_{n,i,j}}{h_n})}{\sum_{j \in A_{n,i,\theta_0}} K(\frac{\theta_0 - \hat{\theta}_{n,i,j}}{h_n})}$$

A partir de 3.26 y 3.28 y como K es acotado para n suficientemente grande se tiene:

$$w_j \leq \frac{\frac{3N_n \|K\|_\infty}{2} \mu_{\hat{\theta}}(B(\theta, h_n r))}{m_l N_n h_n^d} \leq \frac{3r^d h_n^d \|K\|_\infty L}{2m_l h_n^d} = \frac{3r^d \|K\|_\infty L}{2m_l}$$

donde aplicamos el resultado debido a Devroye que $\frac{\mu_{\hat{\theta}}(B(\theta, h_n r))}{h_n^d r^d}$ tiene límite finito para casi todo θ con respecto a μ cuando h tiende a 0. Por lo tanto,

$$Th_n^d r^d < \mu_{\hat{\theta}}(B(\theta, h_n r)) < Lh_n^d r^d$$

Aplicando Hoeffding se obtiene:

$$\begin{aligned} & P(|\hat{P}_{n,i}(\theta_0) - E(\hat{P}_{n,i}(\theta_0))| > M_V h_n r | B_{n,i}, \hat{\theta}_{n,i}, \theta_n) = \\ & P\left(\left| \frac{1}{\#A_{n,i,\theta_0}} \sum_{k \in A_{n,i,\theta_0}} w_k Y_{n,i,k} - \frac{1}{\#A_{n,i,\theta_0}} \sum_{k \in A_{n,i,\theta_0}} w_k E(X_{n,i,k} | \theta_{n,k}) \right| > M_V h_n r | B_{n,i}, \hat{\theta}_{n,i}, \theta_n\right) = \\ & \leq P\left(\left| \sum_{k \in A_{n,i,\theta_0}} [X_{n,i,k} - E(X_{n,i,k} | \theta_{n,k})] \right| > \#A_{n,i,\theta_0} \frac{2m_l M_V h_n}{3\|K\|_\infty L r^{d-1}} | B_{n,i}, \hat{\theta}_{n,i}, \theta_n\right) \\ & \leq 2e^{-\frac{8\#A_{n,i,\theta_0} m_l^2 M_V^2 h_n^2}{9\|K\|_\infty^2 L^2 r^{2(d-1)}}} \leq 2e^{-8\frac{m_l^2 M_V^2 N_n T h_n^{d+2}}{9\|K\|_\infty^2 L^2 r^{d-2}}} \end{aligned}$$

Así pues, para n suficientemente grande y para todo $\theta_0 \in U$:

$$P(|\hat{P}_{n,i}(\theta_0) - P_{n,i}(\theta_0)| > 3M_V h_n r) \leq e^{-8\frac{m_l^2 M_V^2 N_n T h_n^{d+2}}{9\|K\|_\infty^2 L^2 r^{d-2}}} + N_n \left(C_1 e^{-\frac{32L_n^2 + 1}{144K^2 L_n^2} \max_l (\sum_{i=1}^n c_{n,l,i}^2)^{2\alpha-1}} \right)$$

Teorema 3.8.14. Dadas las hipótesis anteriores y para $\theta_0 \in (0, 1)^d$ fijo y

$$N > \eta \frac{9\|K\|_\infty^2 L^2 r^{d-2}}{m_l^2 M_V^2 T} \max_l (\sum_{i=1}^n c_{n,l,i}^2)^{-\alpha(d+2)}$$

$$(3.29) \quad \max\{ECM(\hat{P}_{n,i}(\theta_0)) : 1 \leq i \leq n\} \leq O(\max_l (\sum_{i=1}^n c_{n,l,i}^2)^{2\alpha}) + e^{-\eta}$$

Observación: Si suponemos $N > \max_l (\sum_{i=1}^n c_{n,l,i}^2)^{-\frac{d+2}{2}}$, entonces,

$$(3.30) \quad \max\{ECM(\hat{P}_{n,i}(\theta_0)) : 1 \leq i \leq n\} \leq O(\max_l (\sum_{i=1}^n c_{n,l,i}^2)^{2\alpha})$$

Demostración: $ECM(\hat{P}_{n,i}(\theta_0)) = E((\hat{P}_{n,i}(\theta_0) - P_{n,i}(\theta_0))^2) =$

$$= \int_0^1 P[(\hat{P}_{n,i}(\theta_0) - P_{n,i}(\theta_0))^2 > t] dt \leq (3M_V h_n r)^2 + P[(\hat{P}_{n,i}(\theta_0) - P_{n,i}(\theta_0))^2 > (3M_V h_n r)^2]$$

Esto es, $ECM(\hat{P}_{n,i}(\theta_0)) \leq (3M_{c,d} C_k h_n)^2 + P[|\hat{P}_{n,i}(\theta_0) - P_{n,i}(\theta_0)| > 3M_V h_n r]$.

Aplicando el lema anterior tenemos para n suficientemente grande y todo ítem i :

$$ECM(\hat{P}_{n,i}(\theta_0)) \leq (3M_V h_n r)^2 + e^{-8 \frac{m_l^2 M_V^2 N_n T h_n^{d+2}}{9 \|K\|_\infty^2 L^2 r^{d-2}}} + N_n \left(C_1 e^{-\frac{32L_n^2+1}{144K^2 L_n^2} \max_l (\sum_{i=1}^n c_{n,l,i}^2)^{2\alpha-1}} \right) \\ \leq O(\max_l (\sum_{i=1}^n c_{n,l,i}^2)^{2\alpha}) + e^{-\eta}$$

En el caso unidimensional como $0 < \alpha < 1/2$ si se elige α próximo a $1/2$ entonces el ECM se hace cercano a $O(\frac{1}{n})$. Tenemos entonces el siguiente resultado:

Teorema 3.8.15. Dadas las hipótesis anteriores, suponiendo además que el núcleo es Lipchitz de constante J y, además, $N < Mn^r$ con $r > d + 1$

$$\sup_{\theta \in (a,b)^d} \max_{1 \leq i \leq n} |\hat{P}_{n,i}(\theta) - P_{n,i}(\theta)| \rightarrow 0$$

si $n \rightarrow +\infty$ completamente con $(a, b)^d \subset (0, 1)^d$.

Demostración: Consideremos una grilla de $2^d N^{2d}$ puntos tal que la distancia (con la norma del máximo) entre dos puntos sea menor o igual a $1/N^2$. Sea θ un punto arbitrario de $[0, 1]^d$ entonces existe k tal que la distancia de θ_k^n a θ no es mayor a $1/N^2$.

Para todo θ tenemos

$$|P_{n,i}(\theta) - \hat{P}_{n,i}(\theta)| \leq |P_{n,i}(\theta) - P_{n,i}(\theta_k^n)| + |P_{n,i}(\theta_k^n) - \hat{P}_{n,i}(\theta_k^n)| + |\hat{P}_{n,i}(\theta_k^n) - \hat{P}_{n,i}(\theta)|$$

Consideremos

$$|\hat{P}_{n,i}(\theta_k^n) - \hat{P}_{n,i}(\theta)| = \left| \frac{\sum_{i=1}^N K(\frac{\theta_k^n - \hat{\theta}_i}{h}) X_{ij}}{\sum_{i=1}^N K(\frac{\theta_k^n - \hat{\theta}_i}{h})} - \frac{\sum_{i=1}^N K(\frac{\theta - \hat{\theta}_i}{h}) X_{ij}}{\sum_{i=1}^N K(\frac{\theta - \hat{\theta}_i}{h})} \right| \leq \left| \frac{\sum_{i=1}^N K(\frac{\theta_k^n - \hat{\theta}_i}{h}) X_{ij}}{\sum_{i=1}^N K(\frac{\theta_k^n - \hat{\theta}_i}{h})} - \frac{\sum_{i=1}^N K(\frac{\theta - \hat{\theta}_i}{h}) X_{ij}}{\sum_{i=1}^N K(\frac{\theta - \hat{\theta}_i}{h})} \right| + \\ + \left| \frac{\sum_{i=1}^N K(\frac{\theta - \hat{\theta}_i}{h}) X_{ij}}{\sum_{i=1}^N K(\frac{\theta - \hat{\theta}_i}{h})} - \frac{\sum_{i=1}^N K(\frac{\theta_k^n - \hat{\theta}_i}{h}) X_{ij}}{\sum_{i=1}^N K(\frac{\theta_k^n - \hat{\theta}_i}{h})} \right| = \frac{\sum_{i=1}^N K(\frac{\theta - \hat{\theta}_i}{h}) X_{ij}}{\sum_{i=1}^N K(\frac{\theta - \hat{\theta}_i}{h}) \sum_{i=1}^N K(\frac{\theta_k^n - \hat{\theta}_i}{h})} \left| \sum_{i=1}^N K(\frac{\theta_k^n - \hat{\theta}_i}{h}) - \sum_{i=1}^N K(\frac{\theta - \hat{\theta}_i}{h}) \right| + \\ + \frac{1}{\left| \sum_{i=1}^N K(\frac{\theta_k^n - \hat{\theta}_i}{h}) \right|} \left| \sum_{i=1}^N K(\frac{\theta - \hat{\theta}_i}{h}) X_{ij} - \sum_{i=1}^N K(\frac{\theta_k^n - \hat{\theta}_i}{h}) X_{ij} \right| \leq \frac{2 \sum_{i=1}^N |K(\frac{\theta_k^n - \hat{\theta}_i}{h}) - K(\frac{\theta - \hat{\theta}_i}{h})|}{\sum_{i=1}^N K(\frac{\theta_k^n - \hat{\theta}_i}{h})} \leq \frac{2 \sum_{i=1}^N J \|\theta_k^n - \theta\|}{h \sum_{i=1}^N K(\frac{\theta_k^n - \hat{\theta}_i}{h})}$$

Como $\sum_{i=1}^N K(\frac{\theta_k^n - \hat{\theta}_i}{h}) > Nh^d m_l$ obtenemos que:

$$|\hat{P}_{n,i}(\theta_k^n) - \hat{P}_{n,i}(\theta)| < \frac{2JN\|\theta_k^n - \theta\|}{Nh^{d+1}m_l} < \frac{2J}{N^2 h^{d+1} m_l}$$

Como existe C_1 tal que: $h > C_1 \max_l (\sum_{i=1}^n c_{n,l,i}^2)^\alpha$ y $N > \max_l (\sum_{i=1}^n c_{n,l,i}^2)^{-\alpha(d+2)}$

$$|\hat{P}_{n,i}(\theta_k^n) - \hat{P}_{n,i}(\theta)| = o(\frac{1}{n})$$

A partir de $\sup_{\theta \in [0,1]^d} \left| \frac{\partial P_i}{\partial \theta_j} \right| < c_j$ para todo j entonces si $c = \max_j c_j$ con $j=1, \dots, d$

$$|P_{n,i}(\theta) - P_{n,i}(\theta_k^n)| \leq c \|\theta - \theta_k^n\| \leq c \frac{1}{N^2}$$

Sea $\epsilon > 0$ mostraremos que: $\sum_{n=1}^\infty P(\sup_{\theta \in (a,b)^d} \max_{1 \leq i \leq n} |\hat{P}_{n,i}(\theta) - P_{n,i}(\theta)| > \epsilon) < \infty$

Para probar esto es suficiente probar que:

$$(3.31) \quad \sum_{n=1}^\infty \sum_{i=1}^n \sum_{k=1}^{2^d N^{2d}} P[|P_{n,i}(\theta_k^n) - \hat{P}_{n,i}(\theta_k^n)| > \epsilon/3] < \infty$$

ya que para $n > n_0$ suficientemente grande, $|P_{n,i}(\theta) - P_{n,i}(\theta_k^n)| \leq \frac{\epsilon}{3}$ y $|\hat{P}_{n,i}(\theta_k^n) - \hat{P}_{n,i}(\theta)| \leq \frac{\epsilon}{3}$, entonces,

$$P\left(\sum_{n=1}^{\infty} P(\sup_{\theta \in (a,b)^d} \max_{1 \leq i \leq n} |\hat{P}_{n,i}(\theta) - P_{n,i}(\theta)| > \epsilon) < \sum_{n=1}^{\infty} \sum_{i=1}^n \sum_{k=1}^{2^d N^{2d}} P[|P_{n,i}(\theta_k^n) - \hat{P}_{n,i}(\theta_k^n)| > \epsilon/3]$$

Para este n_0 la triple suma es menor que:

$$(3.32) \quad 2^d n_0^2 N_{n_0}^{2d} + \sum_{n > n_0}^{\infty} \sum_{i=1}^n \sum_{k=1}^{2N_n^{2d}} P[|P_{n,i}(\theta_k^n) - \hat{P}_{n,i}(\theta_k^n)| > 3M_V h_n r]$$

Se puede aplicar el lema 3.8.13 y entonces 3.32 es menor que:

$$2^d n_0^2 N_{n_0}^{2d} + \sum_{n > n_0}^{\infty} 2^d n N_n^{2d} \left(e^{-8 \frac{m_l^2 M_V^2 N_n T h_n^{d+2}}{9 \|K\|_{\infty}^2 L^{2r} d^{-2}}} + N_n (C_1 e^{-\frac{32 L_n^2 + 1}{144 K^2 L_n^2} \max_l (\sum_{i=1}^n c_{n,l,i}^2)^{2\alpha-1}}) \right)$$

Reemplazando N_n por su cota superior $M n^r$ entonces lo anterior queda acotado por

$$2^d n_0^2 N_{n_0}^{2d} + \sum_{n > n_0}^{\infty} 2^d M^{2d} n^{2dr+1} \left(e^{-8 \frac{m_l^2 M_V^2 T \max_l (\sum_{i=1}^n c_{n,l,i}^2)^{2\alpha-1} - [d(1/2-\alpha)+1-2\alpha]}{9 \|K\|_{\infty}^2 L^{2r} d^{-2}}} + \right. \\ \left. + M n^r (C_1 e^{-\frac{32 L_n^2 + 1}{144 K^2 L_n^2} \max_l (\sum_{i=1}^n c_{n,l,i}^2)^{2\alpha-1}}) \right)$$

Como $0 < \alpha < 1/2$ y $\lim \log n \max_l (\sum_{i=1}^n c_{n,l,i}^2)^{1-2\alpha} = 0$ la suma anterior es finita.

Ahora considerando en conjunto los resultados anteriores

Teorema 3.8.16. Si se cumplen las hipótesis:

- 1) Independencia Local
- 2) El rasgo Θ tiene distribuciones marginales uniformes $U[0, 1]$
- 3) para las funciones $g_{n,l}$ existen constantes $c_{n,l,1}, \dots, c_{n,l,n}$ tales que:

$$(3.33) \quad \sup_{x_1, \dots, x_n, x'_1, \dots, x'_n} \sum_{i=1}^n |g_{n,l}(x_1, \dots, x_i, \dots, x_n) - g_{n,l}(x_1, \dots, x'_i, \dots, x_n)|^2 \leq \sum_{i=1}^n c_{n,l,i}^2$$

para $1 \leq l \leq d$.

- 4) Las $F_{n,l}$ correspondientes son asintóticamente Lipschitz, (L_n, γ_n) para toda l con $1 \leq l \leq n$

$$5) \quad \frac{\partial E(g_{n,l}(X) | \Theta_l)}{\partial \Theta_l} > M_l > 0.$$

$$6) \quad \text{Para } 0 < \alpha < 1/2 \quad \frac{\gamma_n}{\max_l (\sum_{i=1}^n c_{n,l,i}^2)^{\alpha}} \rightarrow 0$$

$$7) \quad H \max_l (\sum_{i=1}^n c_{n,l,i}^2)^{-\alpha(d+2)} < N < M n^r \text{ con } r > d + 1$$

$$8) \quad \lim \log n \max_l (\sum_{i=1}^n c_{n,l,i}^2)^{1-2\alpha} = 0.$$

- 9) El núcleo K es simétrico, acotado y con soporte compacto.

10) Las CCI tienen derivadas parciales continuas en $(0, 1)^d$ y en cada compacto $[a, b]^d \subset (0, 1)^d$ son acotadas

$$11) \quad \text{Para } 0 < \alpha < 1/2 \quad \max_l (\sum_{i=1}^n c_{n,l,i}^2)^{\alpha} = O(h_n)$$

- 12) El núcleo K es Lipschitz de constante J .

entonces $\max\{\max_{k=1, \dots, N} \|\hat{\Theta}_{n,k} - \Theta_{n,k}\|, \sup_{\theta \in (a,b)^d} \max_{1 \leq i \leq n} |\hat{P}_{n,i}(\theta) - P_{n,i}(\theta)|\} \rightarrow 0$ si $n \rightarrow +\infty$ con probabilidad 1

Demostración:

Sale directamente de los teoremas 3.6.1 y 3.6.3 notando que la probabilidad del conjunto donde falla esta convergencia es la unión de dos conjuntos cada uno con probabilidad 0.

3.9. Propiedades del estimador de la CCI si la habilidad es medida sin error

Si bien hemos demostrado la consistencia conjunta de los estimadores, en esta sección consideramos el caso más sencillo que es cuando los rasgos son conocidos o cuando son medidos sin error o sea $\widehat{\Theta}_i = \Theta_i$. Con esta simplificación podremos encontrar la distribución asintótica de las CCI. Ramsay (1991) aproxima la distribución para el caso de rasgo medido sin error pero con diseño fijo. En nuestro caso tomaremos en cuenta la distribución del rasgo en la población. El estimador en este caso es:

$$(3.34) \quad \widehat{P_j(\theta)} = \frac{\sum_{i=1}^N X_{i,j} K_H(\theta - \widehat{\Theta}_i)}{\sum_{i=1}^N K_H(\theta - \widehat{\Theta}_i)} = \frac{\sum_{i=1}^N X_{i,j} K_H(\theta - \Theta_i)}{\sum_{i=1}^N K_H(\theta - \Theta_i)}$$

Primeramente veamos que en este caso el estimador es consistente con hipótesis más simples y con un núcleo no necesariamente con soporte compacto. Lo que veamos servirá también para respuestas continuas cuando queremos calcular $E(X|\Theta)$

Teorema 3.9.1. Supongamos que $\Theta \sim f_\theta$, $f_\theta(\theta) > 0$ y continua en θ además la habilidad es medida sin error entonces si $h_N \rightarrow 0$, $Nh_N \rightarrow \infty$

$\int K(u)du = 1$ y $|\theta K(\theta)| \rightarrow 0$ cuando $|\theta| \rightarrow \infty$, $Nh^2 \rightarrow \infty$ y K acotada entonces el estimador de la CCI 3.34 es consistente.

Demostración:

$$\widehat{P}(\theta) = \frac{\frac{1}{Nh} \sum_{i=1}^N X_i K\left(\frac{\theta - \Theta_i}{h}\right)}{\frac{1}{Nh} \sum_{j=1}^N K\left(\frac{\theta - \Theta_j}{h}\right)} = \frac{R_{1N}(\theta)}{R_{0N}(\theta)}$$

Bajo estas condiciones se tiene que, $R_{0N}(\theta) \rightarrow f(\theta)$

Trabajemos ahora con el término $R_{1N}(\theta)$,

$$E(R_{1N}(\theta)) = E\left(\frac{1}{h} K\left(\frac{\theta - \Theta_i}{h}\right) X_i\right) = \frac{1}{h} E\left(K\left(\frac{\theta - \Theta_i}{h}\right) E(X_i|\Theta)\right) = \frac{1}{h} \int K\left(\frac{\theta - t}{h}\right) P(t) f_\Theta(t) dt \rightarrow P(\theta) f(\theta)$$

en cada punto de continuidad de Pf por el lema 3.3.1

$P.f \in L^1$, pues $\int |P(t)| f(t) dt = E(|P(\theta)|) = E(|E(X/\Theta)|) \leq E(E(|X|/\Theta)) = E(|X|)$ o sea basta que $E(|X|) < \infty$ lo cual claramente se cumple (por el tipo de variable que consideramos) y las hipótesis sobre K para aplicar el lema 3.3.1.

Probemos ahora que la varianza tiende a cero:

$$\begin{aligned} Var(R_{1N}(\theta)) &= \frac{1}{Nh^2} Var\left(K\left(\frac{\theta - \Theta_1}{h}\right) X_1\right) \leq \frac{1}{Nh^2} E\left(X_1^2 K^2\left(\frac{\theta - \Theta_1}{h}\right)\right) = \\ &= \frac{1}{Nh^2} E\left(K^2\left(\frac{\theta - \Theta_1}{h}\right) E(X_1^2|\Theta_1)\right) = \frac{1}{Nh^2} \int K^2\left(\frac{\theta - t}{h}\right) m_2(t) f(t) dt \text{ donde} \\ m_2(\Theta) &= E(X^2|\Theta) \text{ y } E(m_2(\Theta)) = E(E(X^2|\Theta)) = E(X^2) \end{aligned}$$

Analicemos ahora el término $\frac{1}{Nh^2} \int K^2\left(\frac{\theta - t}{h}\right) m_2(t) f(t) dt$.

Como $E(X^2) < \infty$, $Nh^2 \rightarrow \infty$ y K está acotado tenemos,

$$\frac{1}{Nh^2} \int K^2\left(\frac{\theta - t}{h}\right) m_2(t) f(t) dt \rightarrow 0$$

Veamos ahora dos resultados, uno de convergencia en probabilidad y otro de convergencia completa.

Sean $(\Theta_1, X_1) \dots (\Theta_N, X_N)$ vectores aleatorios independientes, idénticamente distribuidas, con $\Theta_i \in R^d, X_i \in R$. Consideremos las hipótesis:

$$H1) h_N \rightarrow 0 \text{ y } Nh_N^d \rightarrow \infty$$

Para el núcleo $K : R^d \rightarrow R^+$ pidamos:

H2) existen $C_1, C_2 > 0$ y H tales que $C_1 H(\|\Theta\|) \leq K(\theta) \leq C_2 H(\|\Theta\|)$ con H acotada, decreciente y tal que $t^d H(t) \rightarrow 0$ si $|t| \rightarrow \infty$

H3) Existen $c, r > 0$ tal que $K(\theta) \geq c \chi_{\|\theta\| \leq r}(\theta)$

Si f es integrable y Θ tiene distribución μ entonces definimos:

$$\mu_h = \int K\left(\frac{\Theta-t}{h}\right) f(t) \mu(dt) / \int K\left(\frac{\Theta-t}{h}\right) \mu(dt) = E\left(K\left(\frac{\Theta-\bar{\Theta}}{h}\right) f(\bar{\Theta})\right) / E\left(K\left(\frac{\Theta-\bar{\Theta}}{h}\right)\right)$$

Observaciones:

Lo que le pedimos al núcleo es menos que suponer continuidad y soporte compacto, o radial, comúnmente se toma $K(\|t\|)$. La segunda hipótesis sobre el núcleo implica menos que pedir que $K(0) \neq 0$ y K continua en 0, podemos ver que esta condición la cumplen los núcleos usualmente utilizados. También debemos observar que no hay hipótesis sobre la distribución de las Θ . Presentamos primero dos lemas, el primero un resultado de teoría de la medida y el otro un lema de convolución.

Lema 3.9.2. Sea S_r la bola de radio r centrada en θ , y sea $a_h(\theta) = \frac{h^d}{\mu(S_h)}$ entonces $a_h(\theta)$ tiene un límite finito ctp $\theta(\mu)$.

Demostración: Sea λ la medida de Lebesgue, descompongámosla en la parte absolutamente continua respecto a μ en la parte singular. Notemos por λ_1 a la parte absolutamente continua de λ respecto a μ y λ_2 la parte singular, luego

$$\frac{\lambda(S_h)}{\mu(S_h)} \longrightarrow \frac{d\lambda_1}{d\mu}(\theta) \text{ ctp } \theta(\mu) \text{ como } \lambda(S_h) = h^d \lambda(B_1)$$

$$\text{entonces } \frac{h^d}{\mu(S_h)} = \frac{\lambda(S_h)}{\mu(S_h)} \frac{1}{\lambda(B_1)} \longrightarrow \ell(\theta) \text{ ctp } \theta(\mu) \text{ donde } \ell \text{ es no negativa y finita.}$$

Lema 3.9.3. Sea K no negativo verificando H2) y H3) y f integrable, entonces,

$$\mu_h(\theta) \rightarrow f(\theta) \text{ c.t.p. } \theta(\mu) \text{ si } h \rightarrow 0$$

Si K es el indicador de la bola unitaria, tenemos el teorema de diferenciación respecto a la medida μ . Lo que presentamos es cómo mezclar un teorema de convolución con un teorema de diferenciación.

$$\begin{aligned} \text{Demostración: } |\mu_h(\theta) - f(\theta)| &= \left| \frac{\int K\left(\frac{\Theta-y}{h}\right) f(y) \mu(dy)}{\int K\left(\frac{\Theta-y}{h}\right) \mu(dy)} - f(\theta) \right| = \left| \frac{\int K\left(\frac{\Theta-y}{h}\right) (f(y) - f(\theta)) \mu(dy)}{\int K\left(\frac{\Theta-y}{h}\right) \mu(dy)} \right| \leq \\ &\leq \frac{C_2}{C_1} \int H\left(\frac{\|\theta-y\|}{h}\right) |f(y) - f(\theta)| \mu(dy) / \int H\left(\frac{\|\theta-y\|}{h}\right) \mu(dy) \text{ por H2)} \end{aligned}$$

$$H(t) = \int_0^{H(t)} ds = \int_0^{+\infty} I_{\{H(t) > s\}} ds \text{ y si } A_{t,h} = \left\{ y : H\left(\frac{\|\theta-y\|}{h}\right) > t \right\} \text{ entonces } H\left(\frac{\|\theta-y\|}{h}\right) = \int_0^{+\infty} I_{A_{t,h}}(t) dt$$

$$\text{Sea } \int H\left(\frac{\|\theta-y\|}{h}\right) \mu(dy) = \int_{-\infty}^{+\infty} \int_0^{+\infty} I_{A_{t,h}}(t) dt \mu(dy) = \int_0^{+\infty} \mu(A_{t,h}) dt \text{ ya que } I_{A_{t,h}} \text{ es integrable, aplico Fubini.}$$

$$\text{Ahora } \int H\left(\frac{\|\theta-y\|}{h}\right) |f(y) - f(\theta)| \mu(dy) = \int_{-\infty}^{+\infty} \int_0^{+\infty} I_{A_{t,h}}(t) |f(y) - f(\theta)| dt \mu(dy) = \int_0^{+\infty} \int_{A_{t,h}} |f(y) - f(\theta)| \mu(dy) dt$$

aplicando nuevamente Fubini.

Dividiremos el cociente en dos partes, I y II. Dado $\epsilon > 0$, sea $\delta = \epsilon h^d$

$$I) \frac{\int_{\delta}^{\infty} \int_{A_{t,h}} |f(y) - f(\theta)| \mu(dy) dt}{\int_0^{\infty} \mu(A_{t,h}) dt} \leq \frac{\int_{\delta}^{\infty} \int_{A_{t,h}} |f(y) - f(\theta)| \mu(dy) dt}{\int_{\delta}^{\infty} \mu(A_{t,h}) dt} \leq \sup_{t \geq \delta} \frac{\int_{A_{t,h}} |f(y) - f(\theta)| \mu(dy)}{\mu(A_{t,h})}$$

La última desigualdad es cierta ya que si consideramos dos funciones $g \geq 0$ y $h \geq 0$ con sus

integrales finitas entonces, $\frac{\int_{\delta}^{\infty} g dt}{\int_{\delta}^{\infty} h dt} \leq \sup_{t \geq \delta} \{g/h\}$ Supongamos que $\frac{P(t)}{h(t)} < A$ para todo $t \geq \delta$.

Como g y h son no negativas

$$\int_{\delta}^{+\infty} g dt \leq A \int_{\delta}^{+\infty} h(t) dt \text{ luego } \frac{\int_{\delta}^{+\infty} g(t) dt}{\int_{\delta}^{+\infty} h(t) dt} \leq A$$

Cualquier cota de g/h cumple la desigualdad, así pues lo cumple el supremo de g/h . Si $t \geq \delta$, consideremos los conjuntos $A_{t,h}$ y $A_{\delta,h}$, se tiene $radio(A_{t,h}) < radio(A_{\delta,h})$. Tenemos que $radio(A_{\delta,h}) \rightarrow 0$ pues: $\left\{y : H\left(\frac{\|\theta - y\|}{h}\right) > \delta\right\} = \{y : \|\theta - y\| < H^{-1}(\delta)h\}$ y $hH^{-1}(\delta) = hH^{-1}(\epsilon h^d) \rightarrow 0$ pues $t^d H(t) \xrightarrow{|t| \rightarrow \infty} 0$

Veamos lo anterior. Supongamos que $H^{-1}(\epsilon h^d) = t$, entonces se tiene que $h = \left(\frac{H(t)}{\epsilon}\right)^{1/d}$ luego $|hH^{-1}(\epsilon h^d)| = |t \left(\frac{H(t)}{\epsilon}\right)^{1/d}| \rightarrow 0$ si $h \rightarrow 0$ pues $t \rightarrow \infty$. Lo último se deduce ya que $t^d H(t) \rightarrow 0$ $|t| \rightarrow \infty$ y es claro que $t(H(t))^{1/d} \rightarrow 0$ si $|t| \rightarrow \infty$. Utilizamos un resultado de desigualdades para funciones maximales debido a Hardy y Littlewood. Concretamente, el resultado nos dice que si $radio(A_{t,h}) \rightarrow 0$ entonces

$$\sup_{t \geq \delta} \frac{\int_{A_{t,h}} |f(y) - f(\theta)| \mu(dy)}{\mu(A_{t,h})} \rightarrow 0 \text{ ctp } \theta(\mu)$$

Veamos ahora la segunda parte

$$II) \frac{\int_0^{\delta} \int_{A_{t,h}} |f(y) - f(\theta)| \mu(dy) dt}{\int_0^{\infty} \mu(A_{t,h}) dt}$$

Llamemos $C = \int |f(y)| \mu(dy)$, entonces $\int_0^{\delta} \int_{\Delta_{t,h}} f(y) \mu(dy) \mu \leq C\delta$ y $\int_0^{\delta} \int_{A_{t,h}} f(\theta) \mu(dy) dt \leq \delta f(\theta)$

así pues $\int_0^{\delta} \int_{A_{t,h}} |f(y) - f(\theta)| \mu(dy) dt \leq \delta(C + f(\theta))$

También $\int_0^{\infty} \mu(A_{t,h}) dt = \int H\left(\frac{\|\theta - y\|}{h}\right) \mu(dy) \geq C_1 \mu(S_{rh}) = \frac{C_1 h^d r^d}{a_{rh}(\theta)}$ luego

$$\frac{\int_0^{\delta} \int_{A_{t,h}} |f(y) - f(\theta)| \mu(dy)}{\int_0^{\infty} \mu(A_{t,h}) dt} \leq \frac{\delta(C + f(\theta)) a_{rh}(\theta)}{C_1 r^d h^d} = \frac{(f(\theta) + C) \epsilon a_{rh}(\theta)}{C_1 r^d} < \epsilon' \text{ ctp } \theta(\mu)$$

Aplicamos en la última desigualdad el lema 3.9.2

Teorema 3.9.4. Bajo H1), H2) y H3) entonces:

$$P_N(\theta) \xrightarrow{P} P(\theta) \text{ en ctp } \theta(\mu) \text{ donde } P_N(\theta) = \frac{\sum_{i=1}^N X_i K\left(\frac{\theta - \Theta_i}{h}\right)}{\sum_{j=1}^N K\left(\frac{\theta - \Theta_j}{h}\right)}$$

Obs: En el caso que consideramos respuestas continuas y queremos estimar su esperanza condicional debemos pedir además que $E(|X|) < \infty$ condición necesaria para que exista la esperanza condicional.

Demostración: dividamos por $E\left(K\left(\frac{\theta - \Theta_i}{h}\right)\right)$ obteniendo:

$$P_N(\theta) = \frac{\sum_{i=1}^N X_i K\left(\frac{\theta - \Theta_i}{h}\right) / E\left(K\left(\frac{\theta - \Theta_i}{h}\right)\right)}{\sum_{j=1}^N K\left(\frac{\theta - \Theta_j}{h}\right) / E\left(K\left(\frac{\theta - \Theta_j}{h}\right)\right)}$$

Notemos cómo:

$$V_{Ni} = X_i K\left(\frac{\theta - \Theta_i}{h}\right) / E\left(K\left(\frac{\theta - \Theta_i}{h}\right)\right) \quad Z_{Ni} = K\left(\frac{\theta - \Theta_i}{h}\right) / E\left(K\left(\frac{\theta - \Theta_i}{h}\right)\right) \quad B_{1N} = \frac{1}{N} \sum_{i=1}^N (V_{iN} - E(V_{iN}))$$

$$B_{2N} = \frac{1}{N} \sum_{i=1}^N (Z_{iN} - E(Z_{iN})) \quad A_N = E(V_{Ni}) = E\left(X_i K\left(\frac{\theta - \Theta_i}{h}\right)\right) / E\left(K\left(\frac{\theta - \Theta_i}{h}\right)\right)$$

Observamos que $E(Z_{Ni}) = 1$ y $P_N(\theta) = \frac{(B_{1N} + A_N)}{(B_{2N} + 1)}$. Ahora,

$$A_N = \frac{E\left(K\left(\frac{\theta - \Theta_i}{h}\right) E(X/\Theta)\right)}{E\left(K\left(\frac{\theta - \Theta_i}{h}\right)\right)}$$

así pues, aplico el lema 3.9.3, tomando $f = P(\theta)$; tenemos $P(\theta) \in L^1$, por ser $E(|X|) < \infty$; $K \geq 0$ por H3 y $A_N = \mu_h(\theta)$, entonces, $A_N \rightarrow P(\theta)$ ctp $\theta(\mu)$

Entonces el problema se reduce a probar que $B_{1N} \xrightarrow{P} 0$ y $B_{2N} \xrightarrow{P} 0$ ctp $\theta(\mu)$

Probemos primero que $B_{1N} \xrightarrow{P} 0$. Sean $N > 0$ $Y' = X \chi_{\{|X| \leq N\}}$ e $Y'' = X - Y'$

$$B'_{1N} = \sum_{i=1}^N [Y'_i K\left(\frac{\theta - \Theta_i}{h}\right) - E(Y' K\left(\frac{\theta - \Theta_i}{h}\right))] / n E\left(K\left(\frac{\theta - \Theta_i}{h}\right)\right)$$

$$B''_{1N} = \sum_{i=1}^N [Y''_i K\left(\frac{\theta - \Theta_i}{h}\right) - E(Y'' K\left(\frac{\theta - \Theta_i}{h}\right))] / n E\left(K\left(\frac{\theta - \Theta_i}{h}\right)\right)$$

Se tiene por la linealidad de la esperanza que $B'_{1N} + B''_{1N} = B_{1N}$. Aplicando Chebichev

$$P(|B'_{1N}| > t) \leq \frac{1}{N} \text{Var}[Y'_i K\left(\frac{\theta - \Theta_i}{h}\right)] / t^2 E^2\left(K\left(\frac{\theta - \Theta_i}{h}\right)\right) \leq \frac{1}{N t^2} \frac{E(Y'^2 K^2\left(\frac{\theta - \Theta_i}{h}\right))}{E^2(K\left(\frac{\theta - \Theta_i}{h}\right))} \leq \frac{N^2 \|K\|_\infty}{N t^2} \frac{1}{E(K\left(\frac{\theta - \Theta_i}{h}\right))}$$

Por H3: $E\left(K\left(\frac{\theta - \Theta_i}{h}\right)\right) \geq c E\left(\chi_{\{\|u\| \leq r\}}\left(\frac{\theta - \Theta}{h}\right)\right) = c \int_{\|\theta - t\| \leq rh} \mu(dt) = c \mu(S_{rh})$ luego

$$\frac{N^2 \|K\|_\infty}{Nt^2} \frac{1}{E(K(\frac{\theta-\Theta_i}{h}))} \leq \frac{N^2 \|K\|_\infty}{Nt^2} \frac{r^d h^d}{c\mu(S_{rh})r^d h^d} = \frac{N^2 \|K\|_\infty}{Nt^2 c r^d h^d} a_{rh} = \frac{cte}{N h^d} \rightarrow 0 \text{ ctp } \theta(\mu)$$

para cada N , pues a_{rh} tiene límite finito. Así pues para cada N fijo, $B'_{1N} \xrightarrow{P} 0$ ctp $\theta(\mu)$

Aplicando Markov a B''_{1N} y siendo $g_N(\theta) = E(|Y''|/\Theta)$

$$P(|B''_{1N}| > t) \leq \frac{E(|B''_{1N}|)}{t} \leq \frac{2}{t} \frac{E(|Y''|K(\frac{\theta-\Theta}{h}))}{E(K(\frac{\theta-\Theta}{h}))} = \frac{2}{t} \frac{E(K(\frac{\theta-\Theta}{h}) E(|Y''|/\theta))}{E(K(\frac{\theta-\Theta}{h}))} = \frac{2}{t} g_{n,N}(\theta)$$

Por el lema 3.9.3 con $f = P_N$ se tiene que $\frac{2}{t} g_{n,N}(\theta) \rightarrow P_N(\theta) \frac{2}{t}$ ctp $\theta(\mu)$,

$P_N(\theta) = E(|Y''|/\Theta) = E(|X - Y'|/\Theta), |X - Y'| < 2|X|$, y $|X - Y'| \rightarrow 0$ c.s. $N \rightarrow \infty$

Así pues, $E(P_N(\theta)) = E(E(|X - Y'|/\Theta)) = E(|X - Y'|) \rightarrow 0$ si $N \rightarrow \infty$, además, $P_N \geq 0$ y P_N decreciente, luego $P_N(\theta) \rightarrow 0$ ctp $\theta(\mu)$ con $N \rightarrow \infty$,

Elijamos N tal que $P_N(\theta)$ sea pequeño, entonces si $N \rightarrow \infty$ $B_{1N} \xrightarrow{P} 0$

Esto vale en $A_0 = \{\theta : P_N(\theta) \rightarrow 0\}$ y $A_N = \{\theta : g_{Nh} \rightarrow P_N\}$ y vale en $\bigcap_{N=0}^{\infty} A_N$

Para B_{2N} se hace lo mismo pero como no están las X no hay que truncar.

Teorema 3.9.5. Convergencia completa

Si $|X| < \gamma < \infty$, y se verifican H1), H2), H3), $\frac{Nh^d}{\log N} \rightarrow \infty$ entonces $g_N(\theta) \xrightarrow{c} P(\theta)$ ctp $\theta(\mu)$

Demostración: Escribimos $P_N(\theta) = \frac{(B_{1N} + A_N)}{(B_{2N} + 1)}$. Sabemos que $A_N \rightarrow P(\theta)$ ctp $\theta(\mu)$ por el teorema anterior, luego basta ver que $B_{1N} \xrightarrow{c} 0$ y $B_{2N} \xrightarrow{c} 0$ ctp $\theta(\mu)$. Aplicaremos la desigualdad de Bernstein a V_{Ni}

$$|V_{Ni}| \leq \frac{\gamma \|K\|_\infty}{E(K(\frac{\theta-\Theta}{h}))} \leq \frac{\gamma \|K\|_\infty a_{rh}(\theta)}{c r^d h^d}$$

que como ya vimos tiene límite finito.

$$Var(V_{Ni}) \leq E(V_{Ni}^2) = \frac{E(X_i^2 K^2(\frac{\theta-\Theta}{h}))}{E^2(X_i^2 K^2(\frac{\theta-\Theta}{h}))} \leq \frac{\gamma^2 \|K\|_\infty^2}{E(K(\frac{\theta-\Theta}{h}))} \leq \frac{\gamma^2 \|K\|_\infty^2 a_{rh}(\theta)}{c r^d h^d} \text{ entonces}$$

$$\begin{aligned} P(|B_{1N}| > t) &\leq 2 \exp \left[\frac{-Nt^2}{2 \left(\frac{\gamma^2 \|K\|_\infty^2 a_{rh}(\theta)}{c r^d h^d} + \frac{\gamma \|K\|_\infty a_{rh}(\theta) t}{c r^d h^d} \right)} \right] = 2 \exp \left[\frac{-Nt^2 c r^d h^d}{2 \|K\|_\infty a_{rh}(\theta) \gamma (\gamma + t)} \right] \leq \\ &\leq 2 \exp \left(\frac{-C_{1N} h^d}{C_2 (\gamma + t)} \right) \text{ ctp } \theta(\mu). \text{ llamando a } \frac{c_1}{C_2 (\gamma + t)} = C \text{ obtenemos:} \end{aligned}$$

$$P(|B_{1N}| > t) \leq 2 \exp(-C(\theta) N h^d) \text{ y como } \frac{N h^d}{\log N} \rightarrow \infty \text{ entonces: } \sum 2 \exp(-C(\theta) N h^d)$$

converge, luego $\sum P(|B_{1N}| > t)$ converge, por lo que B_{1N} converge completamente.

Análogamente se procede con B_{2N} .

Veamos ahora los resultados para el caso general de los estimadores. Daremos sin demostración los resultados cuando las habilidades son conocidas.

Teorema 3.9.6. Si las habilidades $\Theta_1, \dots, \Theta_N$ son conocidos o medidas sin error se tiene para el estimador $\hat{P}(\theta)$ que

$$E(\hat{P}(\theta)|\Theta_1, \dots, \Theta_N) - P(\theta) \approx \mu_2(K) \frac{\nabla_P(\theta)^t H H^t \nabla_P(\theta)}{f(\theta)} + \frac{1}{2} \mu_2(K) \text{tr}(H^t W_P(\theta) H)$$

$$V(\hat{P}(\theta)|\Theta_1, \dots, \Theta_N) \approx \frac{1}{N|H|} \|K\|_2^2 \frac{\sigma^2(\theta)}{f(\theta)}$$

Teorema 3.9.7. Si las habilidades $\Theta_1, \dots, \Theta_N$ son conocidos o medidas sin error se tiene para el estimador $\hat{P}^1(\theta)$ que

$$E(\hat{P}^1(\theta)|\Theta_1, \dots, \Theta_N) - P(\theta) \approx \frac{1}{2} \mu_2(K) \text{tr}(H^t W_P(\theta) H)$$

$$V(\hat{P}^1(\theta)|\Theta_1, \dots, \Theta_N) \approx \frac{1}{N|H|} \|K\|_2^2 \frac{\sigma^2(\theta)}{f(\theta)}$$

3.9.1. Distribución asintótica de la CCI

Daremos un resultado para los rasgos conocidos o medidos sin error

Teorema 3.9.8. Sean $\{(\Theta_i, X_i) i \geq 1\}$ el rasgo y la respuesta a cada ítem y que son variables aleatorias independientes idénticamente distribuidas con $\Theta_i \in R^d$ y $X_i \in R$.

Notemos por $F(X_i/\Theta_i = u)$ a la distribución condicional de X_i , dado $\Theta_i = u$ y por $P(u) = E(X_1/\Theta_1 = u)$ y $\sigma^2(u) = E((X_1 - P(u))^2/\Theta_1 = u)$. Supongamos que:

- i) P es Lipchitz, y además $\lim_{\epsilon \rightarrow 0} \frac{P(\theta + \epsilon u) - P(\theta)}{\epsilon} = P'(\theta, u)$ es decir, P tiene derivada en la dirección de u (si P es diferenciable se cumplen las 2 condiciones)
- ii) $\sigma^2(\cdot)$ es continua en θ
- iii) $\Theta_i \sim f$ continua en θ y acotada
- iv) $K : R^d \rightarrow R$ acotado no negativo, $\int K = 1$; $|t|^d K(t) \rightarrow 0$ si $|t| \rightarrow \infty$ y $t^2 K^2(t)$ integrable
- v) existe $0 \leq \beta < \infty$ tal que $Nh^{d+2} \rightarrow \beta$

Sea $P_N(\theta) = \sum \omega_{N_i}(\theta) X_i$ con $\omega_{N_i}(\theta) = \frac{K \frac{\theta - \Theta_i}{h}}{\sum_{i=1}^N K \left(\frac{\theta - \Theta_j}{h} \right)}$, entonces,

$$(Nh^d)^{1/2} (P_N(\theta) - P(\theta)) \xrightarrow{W} N \left(b, \frac{\sigma^2(\theta)}{f(\theta)} \int K^2(u) du \right) \text{ con } b = \beta^{1/2} \int K(t) g'(\theta, t) dt$$

Observación: notando que $Nh^{d+2} \rightarrow \beta$ es $Nh^d h^2 \rightarrow \beta$ podemos observar que como $h^2 \rightarrow 0$ debe cumplirse que $Nh^d \rightarrow \infty$.

Demostración: Nuestro objetivo es calcular la distribución asintótica de $P_N(\theta) - P(\theta)$.

$$\begin{aligned} (Nh^d)^{1/2} \left[\frac{\sum_i K \left(\frac{\theta - \Theta_i}{h} \right) X_i}{\sum_j K \left(\frac{\theta - \Theta_j}{h} \right)} - P(\theta) \right] &= (Nh^d)^{1/2} \left[\frac{\sum_i K \left(\frac{\theta - \Theta_i}{h} \right) X_i - P(\theta) \sum_j K \left(\frac{\theta - \Theta_j}{h} \right)}{\sum_j K \left(\frac{\theta - \Theta_j}{h} \right)} \right] = \\ &= (Nh^d)^{1/2} \frac{\left[\frac{1}{Nh^d} \sum_i K \left(\frac{\theta - \Theta_i}{h} \right) X_i - P(\theta) \frac{1}{Nh^d} \sum_j K \left(\frac{\theta - \Theta_j}{h} \right) \right]}{\frac{1}{Nh^d} \sum_j K \left(\frac{\theta - \Theta_j}{h} \right)} \end{aligned}$$

Como $Nh^d \rightarrow \infty$ por la observación previa, $h \rightarrow 0$ y por iii), iv) y v) obtenemos

$$\frac{1}{Nh^d} \sum_j K\left(\frac{\theta - \Theta_j}{h}\right) \xrightarrow{P} f(\theta)$$

Analicemos ahora el término

$$\begin{aligned} & (Nh^d)^{1/2} \left[\frac{1}{Nh^d} \sum_i K\left(\frac{\theta - \Theta_i}{h}\right) X_i - P(\theta) \frac{1}{Nh^d} \sum_j K\left(\frac{\theta - \Theta_j}{h}\right) \right] = \\ & \frac{1}{(Nh^d)^{1/2}} \left[\sum_i K\left(\frac{\theta - \Theta_i}{h}\right) (X_i - P(\theta)) \right] = \frac{1}{(Nh^d)^{1/2}} \left[\sum_i K\left(\frac{\theta - \Theta_i}{h}\right) (X_i - P(\Theta_i) + P(\Theta_i) - P(\theta)) \right] = \\ & \frac{1}{(Nh^d)^{1/2}} \sum_i K\left(\frac{\theta - \Theta_i}{h}\right) (X_i - P(\Theta_i)) + \frac{1}{(Nh^d)^{1/2}} \sum_i K\left(\frac{\theta - \Theta_i}{h}\right) (P(\Theta_i) - P(\theta)) \end{aligned}$$

Notemos $V_{Ni} = K\left(\frac{\theta - \Theta_i}{h}\right) (X_i - P(\Theta_i))$. Las variables V_{Ni} tienen media 0, y son independientes. Tenemos un sistema triangular y se verifican las condiciones de Lindeberg. Calculemos la varianza límite. Condicionando $(X_i - P(\theta))^2$ a Θ

$$\frac{1}{h^d} E(V_{Ni}^2) = \frac{1}{h^d} \int K^2\left(\frac{\theta - t}{h}\right) \sigma^2(t) f(t) dt \rightarrow \sigma^2(\theta) f(\theta) \int K^2$$

donde hemos aplicado el lema 3.3.1 ya que $\sigma^2 f$ es continua en θ por las hipótesis ii) y iii)

$$\text{Var} \left(\frac{1}{(Nh^d)^{1/2}} \sum_i V_{Ni} \right) = \frac{1}{h^d} E(V_{Ni}^2) \rightarrow \sigma^2 f(\theta) \int K^2(t) dt$$

Así pues, aplicando el teorema de Lindeberg tenemos que:

$$\frac{1}{(Nh^d)^{1/2}} \sum_i V_{Ni} \rightarrow N(0, \sigma^2 f \int K^2) \text{ luego,}$$

$$\frac{1}{(Nh^d)^{1/2}} \sum_i V_{Ni} / \frac{1}{Nh^d} \sum_j K\left(\frac{\theta - \Theta_j}{h}\right) \rightarrow N\left(0, \frac{\sigma^2 f(\theta) \int K^2}{f(\theta)}\right)$$

$$\text{Veamos que: } \frac{1}{(Nh^d)^{1/2}} \sum_i K\left(\frac{\theta - \Theta_i}{h}\right) (P(\Theta_i) - P(\theta)) \xrightarrow{P} \beta^{1/2} f(\theta) \int K(t) g'(\theta, t) dt$$

$$\begin{aligned} & E \left(\frac{1}{(Nh^d)^{1/2}} \sum_i K\left(\frac{\theta - \Theta_i}{h}\right) (P(\Theta_i) - P(\theta)) \right) = \frac{1}{(Nh^d)^{1/2}} N \frac{h^d}{h^d} \int K\left(\frac{\theta - t}{h}\right) (P(t) - P(\theta)) f(t) dt = \\ & = (Nh^d)^{1/2} \int K(u) (P(\theta - uh) - P(\theta)) f(\theta - uh) du = (Nh^d)^{1/2} \int K(u) \frac{(P(\theta - uh) - P(\theta))}{h} f(\theta - uh) h du \end{aligned}$$

$$\text{Ahora si } h \rightarrow 0 \quad K(u) \frac{P(\theta - uh) - P(\theta)}{h} f(\theta - uh) \rightarrow K(u) P'(\theta, u) f(\theta)$$

Además f es acotada, $K(u) \frac{P(\theta - uh) - P(\theta)}{h} f(\theta - uh) \leq C |K(u)| |u|$ y esta es una función integrable, pues $u^2 K^2$ es integrable, luego puedo entonces aplicar convergencia dominada.

Nota: En vez de pedir f acotada puedo pedir K de soporte compacto, entonces fuera de un compacto $K(u)$ vale cero, por lo tanto allí puedo hacer uh suficientemente pequeño, de donde $f(\theta - uh) \leq |f(\theta) + \epsilon|$ y la acotación sería $C |k(u)| |u| |f(\theta) + \epsilon|$.

$$\text{Aplicando convergencia dominada: } \int K(u) \frac{(P(\theta - uh) - P(\theta))}{h} f(\theta - uh) du \rightarrow f(\theta) \int K(t) P'(\theta, t) dt$$

$$\text{y } (Nh^d)^{1/2} h = (Nh^{d+2})^{1/2} \rightarrow \beta^{1/2} \text{ así pues,}$$

$$(Nh^d)^{1/2} h \int K(u) \frac{(P(\theta - uh) - P(\theta))}{h} f(\theta - uh) du \rightarrow \beta^{1/2} f(\theta) \int K(t) P'(\theta, t) dt \text{ luego}$$

$$(Nh^d)^{1/2} \sum_j K\left(\frac{\theta - \Theta_j}{h}\right) (P(\theta - uh) - P(\theta)) / \frac{1}{Nh^d} \sum_j K\left(\frac{\theta - \Theta_j}{h}\right) \rightarrow \beta^{1/2} \int K(t) P'(\theta, t) dt$$

Así pues tengo $\frac{a_N + b_N}{c_N}$ y $c_N \xrightarrow{P} c$; $a_N \xrightarrow{D} N$ y aplicando Slutsky $\frac{b_N}{c_N} \xrightarrow{P} \frac{b}{c}$

Veamos ahora que la varianza tiende a 0:

$$Var\left(\frac{1}{(Nh^d)^{1/2}} \sum_j K\left(\frac{\theta - \Theta_j}{h}\right) (P(\Theta_j) - P(\theta))\right) = \frac{1}{(Nh^d)} NVar\left(K\left(\frac{\theta - \Theta_j}{h}\right) (P(\Theta_j) - P(\theta))\right) =$$

$$\frac{1}{h^d} E\left(K^2\left(\frac{\theta - \Theta_j}{h}\right) (P(\Theta_j) - P(\theta))^2\right) = \frac{1}{h^d} \int K^2\left(\frac{\theta - t}{h}\right) (P(t) - P(\theta))^2 f(t) dt =$$

$$\int K^2(u) (P(\theta - uh) - P(\theta))^2 f(\theta - uh) dt = h^2 \int K^2(u) \frac{(P(\theta - uh) - P(\theta))^2}{h^2} f(\theta - uh) du$$

ahora $\int K^2(u) \frac{(P(\theta - uh) - P(\theta))^2}{h^2} f(\theta - uh) du \rightarrow f(\theta) \int K^2(u) (g'(\theta, u))^2 du$ la integral del lado derecho existe pues g es Lipchitz y además $K^2 u^2$ es integrable, así pues aplico convergencia dominada y se tiene que si $h^2 \rightarrow 0$ la varianza $\rightarrow 0$.

3.10. Sesgo y varianza del estimador multidimensional

Siguiendo el mismo razonamiento que en el caso unidimensional tanto para el estimador no paramétrico multidimensional (basado en Nadaraya-Watson) como el estimador por polinomios locales de grado 1 la varianza condicional queda

$$V(P(\theta) | \widehat{\Theta_1, \dots, \Theta_N}) = \frac{P(\theta)Q(\theta)}{n|H|f(\theta)} \|K\|_2^2 (1 + o_p(1))$$

Para determinar el sesgo se aplica desarrollo de Taylor obteniéndose

$$(P(\Theta_1), \dots, P(\Theta_N))^t = P(\theta) Id_N + \begin{pmatrix} (\Theta_1 - \theta)^t \nabla_P(\theta) \\ \vdots \\ (\Theta_N - \theta)^t \nabla_P(\theta) \end{pmatrix} + \frac{1}{2} \begin{pmatrix} (\Theta_1 - \theta)^t \mathfrak{H}_P(\theta) (\Theta_1 - \theta) \\ \vdots \\ (\Theta_N - \theta)^t \mathfrak{H}_P(\theta) (\Theta_N - \theta) \end{pmatrix}$$

o lo que es igual

$$(P(\Theta_1), \dots, P(\Theta_N))^t = \begin{pmatrix} 1 & (\Theta_1 - \theta)^t \\ \vdots & \vdots \\ 1 & (\Theta_N - \theta)^t \end{pmatrix} \begin{pmatrix} P(\theta) \\ \nabla_P(\theta) \end{pmatrix} + \frac{1}{2} \begin{pmatrix} (\Theta_1 - \theta)^t \mathfrak{H}_P(\theta) (\Theta_1 - \theta) \\ \vdots \\ (\Theta_N - \theta)^t \mathfrak{H}_P(\theta) (\Theta_N - \theta) \end{pmatrix}$$

Reemplazando cada término como en dimensión 1 obtenemos:

$$E(\widehat{P(\theta)}) \approx \frac{1}{f(\theta) + o_p(1)} \left(\frac{\sum_{i=1}^N K_H(\Theta_i - \theta) P(\theta)}{n} + \sum_{i=1}^N K_H(\Theta_i - \theta) (\Theta_i - \theta)^t \nabla_P(\theta) \right. \\ \left. + \sum_{i=1}^N K_H(\Theta_i - \theta) (\Theta_i - \theta)^t \mathfrak{H}_P(\theta) (\Theta_i - \theta) \right)$$

o sea

$$E(\widehat{P(\theta)}) \approx \frac{1}{f(\theta)} \left(f(\theta)P(\theta) + \mu_2(K)\nabla_P H H^t \nabla_f + \frac{1}{2}\mu_2(K)f(\theta)\text{tr}(H^t \mathfrak{H}_P(\theta)H) \right)$$

esto último dado que

$$(1/n) \sum_{i=1}^N K_H(\Theta_i - \theta) = (f(\theta) + o_p(1))$$

$$(1/n) \sum_{i=1}^N K_H(\Theta_i - \theta)(\Theta_i - \theta) = \mu_2(K)HH^t \nabla_f + o_p(HH^t Id_d)$$

$$(1/n) \sum_{i=1}^N K_H(\Theta_i - \theta)(\Theta_i - \theta)(\Theta_i - \theta)^t = \mu_2(K)f(\theta)HH^t \nabla_f + o_p(HH^t)$$

por lo tanto el sesgo queda aproximadamente igual a $\frac{\mu_2(K)\nabla_P H H^t \nabla_f}{f(\theta)} + \frac{1}{2}\mu_2(K)\text{tr}(H^t \mathfrak{H}_P(\theta)H)$

En el caso del polinomio local de grado 1 operando de igual manera sobre la fórmula que utiliza e_1 se tiene que $e_1^t(P(\theta), \nabla_P(\theta))^t = P(\theta)$ luego el sesgo esta constituido por el segundo sumando o sea el sesgo es

$$\frac{1}{2}\mu_2(K)\text{tr}(H^t \mathfrak{H}_P(\theta)H)$$

3.10.1. Métodos de selección de ventanas

Veremos el método de validación cruzada. Recordemos que la suma de los cuadrados residuales es:

$$SCR(H) = N^{-1} \sum_{i=1}^N \{X_i - \hat{m}_H(\Theta_i)\}^2$$

Existe el problema que en el SCR el X_i es usado en el cálculo de $\hat{m}_H(\Theta_i)$ para predecirse a el mismo. Como consecuencia, la SCR puede hacerse tan pequeño como se quiera haciendo $H \rightarrow 0$. Esto conduce a la función de validación cruzada.

$$CV(H) = N^{-1} \sum_{i=1}^N \{X_i - \hat{m}_{H,-i}(\Theta_i)\}^2$$

En esta función se reemplaza $\hat{m}_H(\Theta_i)$ con el estimador

$$\hat{m}_{H,-i}(\Theta_i) = \frac{\sum_{j \neq i} K_H(\Theta_i - \Theta_j) X_j}{\sum_{j \neq i} K_H(\Theta_i - \Theta_j)}$$

lo cual es equivalente a otros enfoques que multiplican cada término por una función de penalización. Para el estimador de Nadaraya-Watson queda

$$CV(H) = \frac{1}{N} \sum_{i=1}^N \{X_i - \hat{m}_{H,-i}(\Theta_i)\}^2 = \frac{1}{N} \sum_{i=1}^N \{X_i - \hat{m}_{H,-i}(\Theta_i)\}^2 \left\{ \frac{X_i - \hat{m}_{H,-i}(\Theta_i)}{X_i - \hat{m}_H(\Theta_i)} \right\}^2$$

y

$$\begin{aligned} \frac{X_i - \hat{m}_H(\Theta_i)}{X_i - \hat{m}_{H,-i}(\Theta_i)} &= \frac{\sum_j K_H(\Theta_i - \Theta_j) X_j - X_i \sum_j K_H(\Theta_i - \Theta_j)}{\sum_{j \neq i} K_H(\Theta_i - \Theta_j) X_j - X_i \sum_{j \neq i} K_H(\Theta_i - \Theta_j)} \cdot \frac{\sum_{j \neq i} K_H(\Theta_i - \Theta_j)}{\sum_j K_H(\Theta_i - \Theta_j)} \\ &= 1 - \frac{K_H(0)}{\sum_j K_H(\Theta_i - \Theta_j)} \end{aligned}$$

Entonces el enfoque de validación cruzada es equivalente al de la función de penalización y comparte las mismas propiedades asintóticas. La validación cruzada es equivalente a la validación cruzada generalizada (Craven y Wahba, 1979). Härdle, Hall y Marron (1988) muestran la optimalidad asintótica del ancho de banda obtenido, aunque la tasa de convergencia es lenta. Una mejora en la selección del ancho de banda es discutida en Härdle, Hall y Marron (1992).

Métodos de selección de anchos de banda distintos a los de validación cruzada pueden ser encontrados en Fan y Gijbels (1996).

Capítulo 4

Modelo isótono

En este capítulo presentaremos un método no paramétrico para estimar las curvas características del ítem (CCI) cuando deben ser monótonas. Primero, se mostrará el caso unidimensional para luego generalizarlo al caso multidimensional.

Se trata de un procedimiento de dos etapas, en la primera etapa se utiliza un estimador no paramétrico de la curva característica del ítem mediante regresión no paramétrica basada en núcleos, y en la segunda etapa, usando el resultado anterior, se estimará la función de densidad de la inversa de la CCI.

Al integrar esta función de densidad obtenemos un estimador isótono de la inversa de la CCI, que se simetriza respecto a la bisectriz del cuadrado unidad para obtener el estimador de la CCI. También presentamos el caso multidimensional donde se procede coordenada a coordenada para obtener un estimador de la CCI isótono en cada dimensión.

En lo que se refiere a estimar funciones de regresión monótonas encontramos gran variedad de métodos como los de Cheng y Lin (1981), Wright (1981), Friedman y Tibshirani (1984), Delecroix y Thomas-Agnan (2000), Gijbels (2005) y Mammen (1991). Brunk (1955) propone un método que es modificado por Mukerjee (1988) para obtener estimadores con propiedades similares a los de regresión no paramétrica; Ramsay (1998), Kelly y Rice (1990) proponen utilizar splines. El procedimiento de Ramsay (1998) es semiparamétrico y es válido cuando la función de regresión satisface una ecuación diferencial de segundo orden por lo que es en general no consistente. Finalmente, Hall y Huang (2001) utilizan el estimador basado en núcleos con los pesos modificados y Dette, Neumeyer y Pilz (2006) utilizan un estimador en dos etapas.

Lee (2007) estudió el desempeño de tres métodos no paramétricos para estimar la CCI y compara el modelo de Ramsay con la regresión isótona y la regresión isótona suavizada. Los métodos de regresión isótona y de regresión isótona suavizada estiman la CCI bajo restricciones de monotonidad.

Bajo la restricción de monotonía de las CCI, Lee (2002) propuso el uso de estimaciones basadas en regresiones isótonas, las cuales son motivadas a partir de Barlow et al. (1972) y Robertson, Wright y Dykstra (1988). La regresión isótona es un método de mínimos cuadrados para ajustar datos bajo restricciones de orden. Afirmar que la función estimada es isótona implica que las funciones elegidas para ajustar los datos son funciones no decrecientes de la variable independiente. La regresión isótona para la estimación de la CCI la definen como sigue:

Definición 4.0.1. Consideremos en la habilidad de los examinados el siguiente orden simple $\theta_1 \leq \theta_2 \leq \dots \leq \theta_k$. Sea $P(\theta)$ una función dada de θ y sea F la familia de todas las funciones isótonas de θ . Entonces, $P^*(\theta)$ es la CCI estimada mediante regresión isótona si y solo si $P^*(\theta)$ es isótona y minimiza $\sum \left[P(\theta) - \hat{P}(\theta) \right]^2$, donde $\hat{P}(\theta)$ es un miembro de F . De aquí, $P^*(\theta)$ cumple

la restricción $P^*(\theta_1) \leq P^*(\theta_2) \leq \dots \leq P^*(\theta_k)$.

La solución más ampliamente usada para computar la regresión isótona es el algoritmo Pool-Adjacent-Violators (PAV) (Barlow et al., 1972; Hanson, Pledger and Wright, 1973). La solución de la regresión isótona a través del algoritmo de PAV se obtiene como sigue:

Paso 1: Ordenar los pares $(\theta_j, P(\theta_j))$ de todos los sujetos mediante la primera componente θ .

Paso 2: Se comienza con $P(\theta_{(1)})$, y se sigue en orden creciente o sea nos movemos a la derecha. Se para si el par $(P(\theta_{(j)}), P(\theta_{(j+1)}))$ viola la restricción de monotonía. Se toma $P(\theta_{(j)})$ y su adyacente $P(\theta_{(j+1)})$ y se remplazan ambos por el promedio:

$$P^*(\theta_{(j)}) = P^*(\theta_{(j+1)}) = \frac{P(\theta_{(j)}) + P(\theta_{(j+1)})}{2}$$

Paso 3: Se verifica que $P(\theta_{(j-1)}) \leq P^*(\theta_{(j)})$. Si no es así, el conjunto $\{P(\theta_{(j-1)}), P(\theta_{(j)}), P(\theta_{(j+1)})\}$ se reemplaza por el promedio

$$P^*(\theta_{(j)}) = P^*(\theta_{(j-1)}) = P^*(\theta_{(j+1)}) = \frac{P(\theta_{(j-1)}) + P(\theta_{(j)}) + P(\theta_{(j+1)})}{3}$$

Se continúa a la izquierda tanto como sea necesario, hasta que los requisitos de monotonía son satisfechos y entonces se continúa nuevamente a la derecha. La solución final es la estimación de la CCI isótona. Sin embargo, el método puede provocar que la función isótona resultante sea una función escalera si se analiza un pequeño grupo de examinados, o la monotonía en algún $P(\theta)$ es violada porque la probabilidad de las CCI fue remplazada por sus promedios. Bajo tales situaciones, la regresión isótona de las CCI son usualmente no muy suaves, pero no decrecientes.

Los estimados de las CCI mediante regresión isótona suavizada son obtenidos primero isotoniando los datos a través del algoritmo PAV y luego suavizando las funciones de regresión isótona usando un núcleo con un ancho de banda apropiado (Lee, 2002). La CCI resultante es generalmente más suave que la obtenida usando la estimación de la CCI mediante regresión isótona.

Si observamos detenidamente el procedimiento vemos que es esencialmente constructivo y nada garantiza la consistencia del estimador.

En este capítulo nos basaremos en el estimador propuesto por Dette, Neumeyer y Pilz (2006) para estimar regresiones monótonas en el intervalo $[0, 1]$. Consideramos que el método propuesto es una alternativa atractiva para estimar las CCI en casos en que estas deban ser crecientes, como por ejemplo en los contextos educativos.

Es un método de mucha flexibilidad, muy elegante y no iterativo. Esta es una ventaja importante respecto a los métodos paramétricos, tanto cuando se trata de modelos unidimensionales como en los modelos multidimensionales. Es justamente en el caso multidimensional donde aparece toda su potencia respecto a los métodos paramétricos ya que la cantidad de iteraciones en estos últimos es elevada. También se obtienen funciones suaves a partir del método original cosa que no es así en otros procedimientos no paramétricos.

Podemos mencionar como limitante que al trabajar en el intervalo $[0, 1]$ no obtenemos la distribución original del rasgo y la estimación de las CCI en el dominio de la habilidad. Sin embargo, esta dificultad es aparente dado que las escalas son equivalentes ante transformaciones monótonas. Si conocemos la distribución del rasgo en el caso unidimensional o las marginales en el caso multidimensional se obtendrán las escalas originales. Otra característica importante es que la dimensión del rasgo latente debe ser reducida ya que la cantidad de observaciones necesarias para una correcta estimación crece potencialmente con el número de dimensiones.

4.1. Modelo isótono

Como dijimos, nuestro enfoque se basa en el método desarrollado por Dette, Neumeyer y Pilz (2006). Mostraremos un estimador no paramétrico para las CCI cuando son isótonas. Como se explicó previamente, este es un supuesto habitual en TRI.

Dette et al. (2006) propusieron un método para estimar funciones monótonas a partir de estimadores no monótonos. Estos autores consideran funciones estrictamente monótonas en $[0, 1]$ con derivadas positivas.

Supongamos que la habilidad τ tiene distribución $F(\tau)$, y sea $P^*(\tau)$ la CCI la cual es monótona creciente y derivable. Notemos por $\theta = F(\tau)$, y $P(\theta) = P^*(F^{-1}(\theta)) = P^*(\tau)$. Es claro que $P(\theta)$ es creciente en $[0, 1]$, por lo tanto podemos suponer sin pérdida de generalidad que el rasgo tiene distribución uniforme en el intervalo unidad.

En este contexto, sean U_1, \dots, U_T una muestra de variables aleatorias con distribución uniforme en $[0, 1]$, K_d es un núcleo, y h_d una ventana. Entonces,

$$\frac{1}{Th_d} \sum_{i=1}^T K_d\left(\frac{P(U_i) - u}{h_d}\right)$$

es el estimador de la densidad de $P(U)$.

Para $u \in [P(0), P(1)]$ tenemos que

$$P(P(U) \leq u) = P(U \leq P^{-1}(u)) = \int_0^{P^{-1}(u)} dt = \int_{P(0)}^u (P^{-1})'(z) dz$$

luego la densidad de $P(U)$ es $(P^{-1})'(u) \chi_{[P(0), P(1)]}(u)$. Entonces integrando,

$$\frac{1}{Th_d} \int_{-\infty}^{\theta} \sum_{i=1}^T K_d\left(\frac{P(U_i) - u}{h_d}\right) du$$

es un estimador consistente de P^{-1} en θ .

Es natural reemplazar $P(U_i)$ mediante un estimador de la CCI valorado en un conjunto de puntos igualmente espaciados en $[0, 1]$, lo cual realizaremos en nuestro caso mediante el modelo no paramétrico basado en el estimador de Nadaraya-Watson de regresión no paramétrica o en el modelo basado en el estimador mediante núcleos de polinomios locales de grado 1. Veamos el proceso con el estimador de Ramsay, entonces, en el contexto de nuestro enfoque, tomemos una grilla $0, \frac{1}{T}, \dots, \frac{i}{T}, \dots, 1$, y usemos el estimador no paramétrico de la regresión en cada punto

$$(4.1) \quad \widehat{P}\left(\frac{i}{T}\right) = \frac{\sum_{j=1}^N K_r\left(\frac{\frac{i}{T} - \widehat{\theta}_j}{h_r}\right) X_j}{\sum_{j=1}^N K_r\left(\frac{\frac{i}{T} - \widehat{\theta}_j}{h_r}\right)}$$

donde K_r y h_r son el núcleo y la ventana del estimador de regresión.

Entonces, el inverso de la CCI monótona en θ será

$$(4.2) \quad \widehat{P^{-1}}(\theta) = \frac{1}{Th_d} \int_{-\infty}^{\theta} \sum_{i=1}^T K_d\left(\frac{\widehat{P}(\frac{i}{T}) - u}{h_d}\right) du$$

Finalmente, el estimador de \widehat{P} se obtiene por reflexión de $\widehat{P^{-1}}$ con respecto a $y = x$.

Para ilustrar el caso multidimensional veremos como proceder en dos dimensiones. Supongamos una CCI creciente en cada rasgo y que el estimador de la CCI se obtuvo como se vio previamente usando regresión no paramétrica en dos dimensiones.

$$(4.3) \quad \widehat{P}_i(\theta_1, \theta_2) = \frac{\sum_{k=1}^N K_r\left(\frac{\widehat{\theta}_{1i} - \theta_1}{h_1}, \frac{\widehat{\theta}_{2i} - \theta_2}{h_2}\right) X_{ik}}{\sum_{k=1}^N K_r\left(\frac{\widehat{\theta}_{1k} - \theta_1}{h_1}, \frac{\widehat{\theta}_{2k} - \theta_2}{h_2}\right)}$$

donde $e K_r$ es un núcleo de dimensión dos, con soporte compacto $C \subset [0, 1]^2$. También suponemos que tanto el núcleo, las ventanas y las funciones utilizadas para estimar los rasgos cumplen las hipótesis requeridas para que este estimador sea consistente.

Tomemos una grilla $0, \frac{1}{T}, \dots, \frac{j}{T}, \dots, 1$ para θ_1 y una grilla $0, \frac{1}{T}, \dots, \frac{j}{T}, \dots, 1$ para θ_2

Para cada $\theta_2 \in (0, 1)$ fijo, consideremos el estimador

$$(4.4) \quad \widehat{H}^{-1}(\theta_1|\theta_2) = \frac{1}{Th_d} \int_{-\infty}^{\theta_1} \sum_{i=1}^T K_d\left(\frac{\widehat{P}(\frac{i}{T}, \theta_2) - u}{h_d}\right) du$$

La función $\widehat{H}^{-1}(\theta_1|\theta_2)$ es estrictamente creciente en θ_1 para cada θ_2 fijo.

Podemos calcular el inverso (como función de θ_1) para obtener $\widehat{H}(\theta_1|\theta_2)$.

Para cada $\theta_1 \in (0, 1)$ fijo, calculamos

$$(4.5) \quad \widehat{P^{*-1}}(\theta_1, \theta_2) = \frac{1}{Th_d} \sum_{j=1}^T \int_{-\infty}^{\theta_2} K_d\left(\frac{H(\theta_1|\frac{j}{T}) - u}{h_d}\right) du$$

Finalmente, $\widehat{P^*}(\theta_1, \theta_2)$ se calcula mediante la inversa (como función de θ_2).

El algoritmo sería:

Paso 1 Usando el procedimiento estandar para cada $1 \leq i \leq T$, y $1 \leq j \leq T$, calcular

$$\widehat{P}\left(\frac{i}{T}, \frac{j}{T}\right)$$

Paso 2 Para cada $\frac{j}{T}$ con $1 \leq j \leq T$, calcular $H^{-1}(\theta_1|\frac{j}{T})$ mediante

$$(4.6) \quad \widehat{H}^{-1}(\theta_1|\frac{j}{T}) = \frac{1}{Th_d} \int_{-\infty}^{\theta_1} \sum_{i=1}^T K_d\left(\frac{\widehat{P}(\frac{i}{T}, \frac{j}{T}) - u}{h_d}\right) du$$

Paso 3 Invirtiendo con respecto a θ_1 , obtenemos $\widehat{H}(\theta_1|\frac{j}{T})$ with $1 \leq j \leq T$

Paso 4 Calcular

$$\widehat{P^{*-1}}(\theta_1, \theta_2) = \frac{1}{Th_d} \sum_{j=1}^T \int_{-\infty}^{\theta_2} K_d\left(\frac{H(\theta_1|\frac{j}{T}) - u}{h_d}\right) du$$

Paso 5 Invirtiendo respecto a θ_2 , obtenemos $\widehat{P^*}(\theta_1, \theta_2)$.

4.2. Propiedades del estimador cuando el rasgo es medido sin error

Veremos las propiedades del estimador en el caso más simple donde se puede aplicar directamente el modelo de regresión. Supondremos que el rasgo es conocido o medido sin error y sin pérdida de generalidad que las respuestas son continuas en $[0, 1]$ y se quiere estimar la esperanza condicional de las respuestas con respecto al rasgo. Trabajaremos en una dimensión pues la extensión de los resultados a una dimensión mayor es clara. Para mantener la notación pondremos $E(X_i|\Theta) = P(\Theta)$. Consideremos el modelo

$$X_i = P(\Theta_i) + \sigma(\Theta_i)\varepsilon_i, \quad i = 1, \dots, N$$

donde $\{\Theta_i, X_i\}$ $i = 1, \dots, N$ es una muestra bivariada de observaciones independientes idénticamente distribuidas tales que Θ_i tiene densidad f dos veces continuamente diferenciable, con soporte compacto $[0, 1]$. Asumiremos que las variables aleatorias ε_i son i.i.d. con $E(\varepsilon_i) = 0$,

$E(\varepsilon_i^2) = 1$ y momento cuarto finito. También asumiremos que la función $\sigma : [0, 1] \rightarrow \mathbb{R}^+$ y la función de P son continuas y dos veces continuamente diferenciables, respectivamente.

Vimos más arriba que si $\Theta_i \sim f$ entonces $\frac{1}{Th_d} \sum_{i=1}^T K_d(\frac{\Theta_i - \theta}{h_d}) = \hat{f}(\theta)$ es un estimador de la densidad f y

$$(4.7) \quad \widehat{P_I^{-1}}(\theta) = \int_{-\infty}^{\theta} \frac{1}{Th_d} \sum_{i=1}^T K_d\left(\frac{\hat{P}(\frac{i}{T}) - u}{h_d}\right) du$$

Definamos

$$(4.8) \quad P_T^{-1}(\theta) = \int_{-\infty}^{\theta} \frac{1}{Th_d} \sum_{i=1}^T K_d\left(\frac{P(\frac{i}{T}) - u}{h_d}\right) du$$

Supondremos que K_d y K_r tienen soporte compacto, por ejemplo, sin pérdida de generalidad en $[-1, 1]$.

Hay dos resultados principales uno sobre una aproximación de $P_T(\theta)$ a $P(\theta)$ y otro que obtiene la distribución asintótica.

El primer resultado establece que si las hipótesis mencionadas se cumplen y además P es estrictamente creciente entonces para todo $\theta \in [0, 1]$ con $P'(\theta) > 0$

$$(4.9) \quad P_T(\theta) = P(\theta) + \kappa_2(K_d)h_d^2 \frac{P''(\theta)}{(P'(\theta))^2} + o(h_d) + O\left(\frac{1}{Th_d}\right)$$

donde $\kappa_2(K) = \frac{1}{2} \int_{-1}^1 v^2 K(v) dv$

El resultado asintótico establece que si \hat{P}_I y P_T denotan las inversas de las funciones \hat{P}_I^{-1} y P_T^{-1} . Si $\lim_{n \rightarrow \infty} \frac{h_r}{h_d} = c \in [0, \infty)$, existe entonces tenemos, que para $\theta \in (0, 1)$ con $P'(\theta) > 0$,

$$\sqrt{Nh_d} \left(\hat{P}_I(\theta) - P_T(\theta) - \kappa_2(K_r)h_r^2 \left(\frac{P''f + 2P'f'}{f} \right) (\theta) \right) \Rightarrow N(0, s^2(\theta))$$

donde la varianza asintótica está dada por

$$s^2(\theta) = \frac{\sigma^2(\theta)P'(\theta)}{f(\theta)} \int \int \int K_d(w + cP'(\theta)(v - u)) K_d(w) K_r(u) K_r(v) dw du dv$$

Si $\lim_{n \rightarrow \infty} \frac{h_r}{h_d} = \infty$ tenemos que para todo $\theta \in (0, 1)$ con $P'(\theta) > 0$,

$$\sqrt{Nh_d} \left(\hat{P}_I(\theta) - P_T(\theta) - \kappa_2(K_r)h_r^2 \left(\frac{P''f + 2P'f'}{f} \right) (\theta) \right) \Rightarrow N(0, \tilde{s}^2(\theta))$$

donde la varianza asintótica está dada por $\tilde{s}^2(\theta) = \frac{\sigma^2(\theta)}{f(\theta)} \int K_r^2(u) du$.

Observación 4.2.1. $\widehat{P_I^{-1}} = 1$ si $\theta > \max_{i=1}^T \hat{P}(\frac{i}{T}) + h_d$ y $\widehat{P_I^{-1}} = 0$ si $\theta < \min_{i=1}^T \hat{P}(\frac{i}{T}) - h_d$

Dado que $\widehat{P_I^{-1}} = \sum_{i=1}^T \int_{-\infty}^{\theta} \frac{1}{Th_d} K_d\left(\frac{\hat{P}(\frac{i}{T}) - u}{h_d}\right) du$ como K_d tiene soporte $[-1, 1]$ entonces $K_d = 0$

si para todo i se cumple alguna de las desigualdades $\begin{cases} \frac{\hat{P}(\frac{i}{T}) - u}{h_d} > 1 \\ \frac{\hat{P}(\frac{i}{T}) - u}{h_d} < -1 \end{cases}$ o lo que es lo mismo $\hat{P}(\frac{i}{T}) - h_d > u$, o $\hat{P}(\frac{i}{T}) + h_d < u$ para todo i .

Si $\theta < \min_{i=1}^T \hat{P}(\frac{i}{T}) - h_d$, para todo $u \in (-\infty, \theta]$ y para todo i se cumple $u < \hat{P}(\frac{i}{T}) - h_d$ entonces $\hat{P}_I^{-1}(\theta) = 0$. Si $\theta > \max_{i=1}^T \hat{P}(\frac{i}{T}) + h_d$ entonces tenemos que

$$\int_{-\infty}^{\theta} \frac{1}{Th_d} \sum_{i=1}^T K_d\left(\frac{\hat{P}(\frac{i}{T})-u}{h_d}\right) du = \int_{-\infty}^{\min \hat{P}(\frac{i}{T})-h_d} \frac{1}{Th_d} \sum_{i=1}^T K_d\left(\frac{\hat{P}(\frac{i}{T})-u}{h_d}\right) du +$$

$$\int_{\min \hat{P}(\frac{i}{T})-h_d}^{\max \hat{P}(\frac{i}{T})+h_d} \frac{1}{Th_d} \sum_{i=1}^T K_d\left(\frac{\hat{P}(\frac{i}{T})-u}{h_d}\right) du + \int_{\max \hat{P}(\frac{i}{T})+h_d}^{+\infty} \frac{1}{Th_d} \sum_{i=1}^T K_d\left(\frac{\hat{P}(\frac{i}{T})-u}{h_d}\right) du = 1$$

ya que la primera y última integral valen 0 por verificarse las desigualdades vistas más arriba y la segunda integral vale 1, ya que el núcleo tiene soporte $[-1, 1]$.

Observación 4.2.2. Tenemos que $P_T^{-1}(\theta) = \int_0^1 \chi_{\{P(x) \leq \theta\}} dx + o(1)$ dado que

$$\int_{-\infty}^{\theta} \frac{1}{Th_d} \sum_{i=1}^T K_d\left(\frac{P(\frac{i}{T})-u}{h_d}\right) du \rightarrow \int_{-\infty}^{\theta} (P^{-1}(u))' du = P^{-1}(\theta) = \int_0^{P^{-1}(\theta)} d\theta =$$

$$\int_0^1 \chi_{\{P^{-1}(\theta) \geq x\}} dx = \int_0^1 \chi_{\{P(x) \leq \theta\}} dx$$

Observación 4.2.3. También se cumple que $\int_0^1 \chi_{\{P(x) \geq \theta\}} dx = P^{-1}(\theta) = \inf\{u : P(u) > \theta\}$

Observación 4.2.4. P_T^{-1} es estrictamente creciente si $\max_{i=1..T-1} [P(\frac{i+1}{T}) - P(\frac{i}{T})] < 2h_d$.

Sea $\theta_2 > \theta_1$ $P_T(\theta_2) - P_T(\theta_1) = \frac{1}{Th_d} \int_{\theta_1}^{\theta_2} \sum_{i=1}^T K_d\left(\frac{P(\frac{i}{T})-u}{h_d}\right) du$ y este término es positivo excepto si $K_d\left(\frac{P(\frac{i}{T})-u}{h_d}\right) = 0$ para todo i y $u \in (\theta_1, \theta_2)$. Como K_d tiene soporte $[-1, 1]$ entonces esto se verifica si $|\frac{P(\frac{i}{T})-u}{h_d}| > 1 \forall i$ y $u \in (\theta_1, \theta_2)$ o sea $|P(\frac{i}{T}) - u| > h_d$.

Si $\max_{i=1..T-1} [P(\frac{i+1}{T}) - P(\frac{i}{T})] < 2h_d$, entonces, si $u \in [P(\frac{i}{T}), \frac{P(\frac{i+1}{T})+P(\frac{i}{T})}{2}]$, entonces, $|P(\frac{i}{T}) - u| < h_d$ y si $u \in [\frac{P(\frac{i}{T})+P(\frac{i+1}{T})}{2}, P(\frac{i+1}{T})]$, entonces $|P(\frac{i+1}{T}) - u| < h_d$ luego; $\forall u \in (\theta_1, \theta_2)$, existe i tal que $K_d\left(\frac{P(\frac{i}{T})-u}{h_d}\right) > 0$.

Igualmente \hat{P}_I^{-1} es estrictamente creciente si $\max_{i=1..T-1} [\hat{P}(\frac{i+1}{T}) - \hat{P}(\frac{i}{T})] < 2h_d$.

Las condiciones anteriores se cumplen para T y N suficientemente grandes por la continuidad de P y su estimación \hat{P} .

Consideremos para hacer más simple la notación que el número T de puntos en la grilla es igual al tamaño muestral N por lo que de ahora en adelante $T=N$

Lema 4.2.5. Si el núcleo K es acotado y P es derivable entonces

$$(4.10) \quad P_T^{-1}(\theta) = \int_0^1 \int_{-\infty}^{\theta} K\left(\frac{P(x)-u}{h_d}\right) \frac{du}{h_d} dx + O\left(\frac{1}{Th_d}\right)$$

Demostración. Como $T=N$ sea $g(N, h) = \int_{-\infty}^{\theta} \frac{1}{h_d} \sum_{i=1}^N \frac{1}{N} K\left(\frac{P(\frac{i}{N})-u}{h_d}\right) du - \int_{-\infty}^{\theta} \frac{1}{h_d} \int_0^1 K\left(\frac{P(x)-u}{h_d}\right) dx du$ luego tenemos

$$(4.11) \quad Ng(N, h) = \int_{-\infty}^{\theta} \frac{1}{h_d} \sum_{i=1}^N K\left(\frac{P(\frac{i}{N})-u}{h_d}\right) du - N \int_{-\infty}^{\theta} \frac{1}{h_d} \int_0^1 K\left(\frac{P(x)-u}{h_d}\right) dx du$$

Como $\int_{-\infty}^{\theta} \frac{1}{h_d} K\left(\frac{P(x)-u}{h_d}\right) du = \int_{P(x)-h_d}^{\theta} \frac{1}{h_d} K\left(\frac{P(x)-u}{h_d}\right) du = \int_{-1}^{\frac{\theta-P(x)}{h_d}} K(z) dz$ entonces 4.11 se transforma en

$$(4.12) \quad \sum_{i=1}^N \left[\int_0^1 \int_{-1}^{\frac{\theta-P(\frac{i}{N})}{h_d}} K(z) dz - \int_0^1 \int_{-1}^{\frac{\theta-P(x)}{h_d}} K(z) dz dx \right] \leq \sum_{i=1}^N \int_0^1 \int_{\frac{\theta-P(x)}{h_d}}^{\frac{\theta-P(\frac{i}{N})}{h_d}} K(z) dz$$

Como $K(z)$ está acotado por C y usando 4.12 resulta que:

$$\begin{aligned} \int_{\frac{\theta-P(x)}{h_d}}^{\frac{\theta-P(\frac{i}{N})}{h_d}} K(z)dz &\leq \frac{C}{h_d}(P(\frac{i}{N})-P(x)) \leq \frac{C}{h_d} \sum_{i=1}^N \int_0^1 (P(\frac{i}{N})-P(x))dx = \frac{C}{h_d} \sum_{i=1}^N \int_0^1 C_1((\frac{i}{N})-x)dx \leq \\ &\leq \frac{CC_1}{h_d} \sum_{i=1}^N ((\frac{i}{N}) - \frac{1}{2}) = \frac{C_2}{h_d} (\frac{N(N+1)}{2N} - N\frac{1}{2}) = \frac{C_2}{2h_d} \end{aligned}$$

Entonces $Nhg(N, h) \leq C_2$ o sea $g(N, h) = O(\frac{1}{Nh})$

Si tomamos N y T cualesquiera también

$$P_T^{-1}(\theta) = \int_0^1 \int_{-\infty}^{\theta} K(\frac{P(x)-u}{h_d}) \frac{du}{h_d} dx + O(\frac{1}{Th_d})$$

Lema 4.2.6. Si P es estrictamente creciente y las hipótesis ya vistas entonces:

para todo $\theta \in (P(0), P(1))$ con $(P^{-1})'(\theta) > 0$

$$(4.13) \quad P_T^{-1}(\theta) = P^{-1}(\theta) + \kappa_2(K_d)h_d^2(P^{-1})'' + o(h_d^2) + O(\frac{1}{Th_d})$$

$$\text{con } \kappa_2(K) = \frac{1}{2} \int_{-1}^1 v^2 K(v) dv.$$

Demostración. K_d tiene soporte $[-1, 1]$ usando el lema anterior debe ser entonces

$-1 \leq \frac{P(x)-u}{h_d} \leq 1$ o sea $u \geq P(x)-h_d$ y $u \leq P(x)+h_d$. Para ese x se cumple que $P(x)-h_d = \theta$ lo que implica que varía entre 0 y $P^{-1}(\theta+h_d)$ y u varía entre $P(x)-h_d$ y θ , entonces llamando

$$\begin{aligned} A(h_d) &= \int_0^1 \int_{-\infty}^{\theta} K_d(\frac{P(x)-u}{h_d}) \frac{du}{h_d} dx = \int_0^{P^{-1}(\theta+h_d)} \int_{P(x)+h_d}^{\theta} K_d(\frac{P(x)-u}{h_d}) \frac{du}{h_d} dx = \\ &= \int_0^{P^{-1}(\theta-h_d)} \int_{P(x)-h_d}^{\theta} K_d(\frac{P(x)-u}{h_d}) \frac{du}{h_d} dx + \int_{P^{-1}(\theta-h_d)}^{P^{-1}(\theta+h_d)} \int_{P(x)-h_d}^{\theta} K_d(\frac{P(x)-u}{h_d}) \frac{du}{h_d} dx \end{aligned}$$

En la primera integral haciendo el cambio de variable $-w = \frac{P(x)-u}{h_d}$ y luego aplicando partes

$$\begin{aligned} \int_0^{P^{-1}(\theta-h_d)} \int_{P(x)-h_d}^{\theta} K_d(\frac{P(x)-u}{h_d}) \frac{du}{h_d} dx &= \int_0^{P^{-1}(\theta-h_d)} \int_{-1}^{\frac{-P(x)+\theta}{h_d}} K_d(w) dw dx = \\ (4.14) \quad &= x \int_{-1}^{\frac{\theta-P(x)}{h_d}} K(w) dw \Big|_0^{P^{-1}(\theta-h_d)} + \int_0^{P^{-1}(\theta+h_d)} x K_d(\frac{\theta-P(x)}{h_d}) \frac{P'(x)}{h_d} dx \end{aligned}$$

donde la última integral vale cero ya que $x \in [0, P^{-1}(\theta-h_d)]$.

Recordemos que $K_d \neq 0$ si $|\frac{\theta-P(x)}{h_d}| < 1$ o sea $P(x) \in [\theta-h_d, \theta+h_d]$. Como P es monótona debe ser $x \in [P^{-1}(\theta-h_d), P^{-1}(\theta+h_d)]$ pero en nuestro caso $x \in [0, P^{-1}(\theta-h_d)]$, entonces $K_d = 0$ luego la última integral vale cero.

La segunda integral de 4.14 es igual a:

$$\begin{aligned} \int_{P^{-1}(\theta-h_d)}^{P^{-1}(\theta+h_d)} \int_{P(x)-h_d}^{\theta} K_d(\frac{P(x)-u}{h_d}) \frac{du}{h_d} dx &= \\ \int_0^1 \chi\{P^{-1}(\theta-h_d) \leq x \leq P^{-1}(\theta+h_d)\} \int_{P(x)-h_d}^{\theta} K_d(\frac{P(x)-u}{h_d}) \frac{du}{h_d} dx \end{aligned}$$

Haciendo el cambio de variables $v = \frac{P(x)-u}{h_d}$ resulta que:

$$\begin{aligned} \int_0^1 \chi\{P^{-1}(\theta - h_d) \leq x \leq P^{-1}(\theta + h_d)\} \int_1^{\frac{P(x)-\theta}{h_d}} -K_d(v)dvdx = \\ = \int_0^1 \chi\{P^{-1}(\theta - h_d) \leq x \leq P^{-1}(\theta + h_d)\} \int_{\frac{P(x)-\theta}{h_d}}^1 K_d(v)dvdx \end{aligned}$$

Haciendo el cambio de variables $z = \frac{P(x)-\theta}{h_d}$ en las indicatrices tenemos:

$$\int_{\frac{P(0)-\theta}{h_d}}^{\frac{P(1)-\theta}{h_d}} \chi\{-1 \leq z \leq 1\} \int_z^1 h_d K_d(v) (P^{-1})'(\theta + h_d z) dz = h_d \int_{-1}^1 \int_z^1 K_d(v) (P^{-1})'(\theta + h_d z) dz$$

Entonces, $A(h_d) = P^{-1}(\theta - h_d) + h_d \int_{-1}^1 (P^{-1})'(\theta + h_d z) \int_z^1 K_d(v) dv dz$. Aplicamos Taylor y

$$P^{-1}(\theta - h_d) = P^{-1}(\theta) - (P^{-1})'(\theta)(h_d) + (P^{-1})''(\theta) \frac{h_d^2}{2} + o(h_d^2)$$

$$(P^{-1})'(\theta + zh_d) = (P^{-1})'(\theta) + (P^{-1})''(\theta)(zh_d) + o(h_d)$$

Remplazando resulta $A(h_d) = P^{-1}(\theta) - (P^{-1})'(\theta)(h_d) + (P^{-1})''(\theta) \frac{h_d^2}{2} + o(h_d^2) +$

$$h_d \int_{-1}^1 [(P^{-1})'(\theta) + (P^{-1})''(\theta)(zh_d) + o(h_d)] \int_z^1 K_d(v) dv dz$$

Como $\int_{-1}^1 \int_z^1 K_d(v) dv dz = 1$ (observación 4.2.7) resulta $h_d \int_{-1}^1 (P^{-1})'(\theta) \int_z^1 K_d(v) dv dz = h_d (P^{-1})'(\theta)$. Entonces,

$$A(h_d) = P^{-1}(\theta) + (P^{-1})''(\theta) \frac{h_d^2}{2} + o(h_d^2) + h_d \int_{-1}^1 (P^{-1})''(\theta) zh_d \int_z^1 K_d(v) dv dz + h_d(o(h_d)) \int_{-1}^1 \int_z^1 K_d(v) dv dz$$

$$\text{luego } A(h_d) = P^{-1}(\theta) + h_d^2 (P^{-1})''(\theta) \left(\frac{1}{2} + \int_{-1}^1 z \int_z^1 K_d(v) dv dz \right) + o(h_d^2)$$

Apliquemos partes a:

$$\int_{-1}^1 z \int_z^1 K_d(v) dz = \frac{z^2}{2} \int_z^1 K_d(v) dv \Big|_{-1}^1 + \int_1^1 \frac{z^2}{2} K_d(z) dz = \frac{-1}{2} + \int_{-1}^1 \frac{z^2}{2} K_d(z) dz =$$

$$\text{entonces obtenemos } \int_{-1}^1 \frac{z^2}{2} K_d(z) dz = \frac{1}{2} + \int_{-1}^1 z \int_z^1 K_d(v) dv dz.$$

$$\text{Ahora llamamos } \kappa_2(K_d) = \int_{-1}^1 \frac{z^2}{2} K_d(z) dz = \frac{1}{2} + \int_{-1}^1 z \int_z^1 K_d(v) dv dz \text{ luego}$$

$$A(h_d) = P^{-1}(\theta) + P^{-1}''(\theta) h_d^2 \kappa_2(K_d) + o(h_d^2)$$

Por el lema 4.2.5 sabemos que

$$P_T^{-1}(\theta) = \int_0^1 \int_{-\infty}^{\theta} K_d\left(\frac{P(x)-u}{h_d}\right) \frac{du}{h_d} dx + O\left(\frac{1}{Th_d}\right) = P^{-1}(\theta) + P^{-1}''(\theta) h_d^2 \kappa_2(K_d) + o(h_d^2) + O\left(\frac{1}{Th_d}\right)$$

lo que prueba el lema.

Observación 4.2.7. Probemos aplicando partes que $\int_{-1}^1 \int_z^1 K_d(v) dv dz = 1$

$$\int_{-1}^1 K_d(v) dv dz = 1 = z \int_z^1 K_d(v) dv \Big|_{-1}^1 - \int_{-1}^1 z(-K_d(z)) dz = 1 - K_d(1) \int_1^1 z dz + \int_{-1}^1 z K_d(z) dz = 1$$

Definición 4.2.8. Sea θ fijo. Definamos $M = \{H \in C^2[0, 1] : H' > 0 \text{ en } [0, 1] \text{ el cual contiene a } \theta \text{ en el interior de la imagen, es decir } \theta \in \text{int}(H[0, 1])\}$

Definición 4.2.9. Definimos $\phi : M \rightarrow [0, 1]$ dada por $\phi(H) = H^{-1}(t)$.

También definimos para $H_1, H_2 \in M$ $Q[0, 1] \rightarrow \mathbb{R}$ dada por $\lambda \rightarrow \phi(H_1 + \lambda(H_2 - H_1))$

o sea $Q(\lambda) = \phi(H_1 + \lambda(H_2 - H_1)) = (H_1 + \lambda(H_2 - H_1))^{-1}(t)$

Lema 4.2.10. El mapa $Q : [0, 1] \rightarrow \mathbb{R}$ es dos veces diferenciable con

$$Q'(\lambda) = \frac{H_2 - H_1}{H'_1 + \lambda(H'_2 - H'_1)} \circ (H_1 + \lambda(H_2 - H_1))^{-1}(t)$$

$$Q''(\lambda) = Q'(\lambda) \cdot \left\{ -2 \frac{H'_2 - H'_1}{H'_1 + \lambda(H'_2 - H'_1)} + \frac{(H_2 - H_1)(H''_1 + \lambda(H''_2 - H''_1))}{(H'_1 + \lambda(H'_2 - H'_1))^2} \right\} \circ Q(\lambda)$$

Demostración. Sea $F(x, y) = (H_1 + x(H_2 - H_1))(y) - t$ entonces

$$F(\lambda, Q(\lambda)) = (H_1 + \lambda(H_2 - H_1))(H_1 + \lambda(H_2 - H_1))^{-1}(t) - t = t - t = 0$$

Es fácil ver que el dominio de Q se puede extender a un entorno de $[0, 1]$ entonces aplicamos el teorema de la Función Implícita.

Definimos $y = Q(x)$, o sea, consideramos $F(x, Q(x))$. En λ tenemos que $F(\lambda, Q(\lambda)) = 0$ entonces existe $Q'(\lambda)$ y se puede calcular como $\frac{\partial Q(\lambda)}{\partial \lambda} = \frac{-F_x}{F_y}$ evaluado en $(\lambda, Q(\lambda))$.

Resulta $\frac{\partial F}{\partial x} = (H_2 - H_1)(y)$ y también $\frac{\partial F}{\partial y} = \frac{\partial H_1}{\partial y} + x \frac{\partial (H_2 - H_1)(y)}{\partial y} = H'_1(y) + x(H'_2 - H'_1)(y)$

Entonces $\frac{\partial Q(x)}{\partial x}(\lambda) = -\frac{(H_2 - H_1)(y)}{H'_1(y) + x(H'_2 - H'_1)(y)} \mid_{(\lambda, Q(\lambda))}$

$$Q'(\lambda) = -\frac{(H_2 - H_1)(Q(\lambda))}{H'_1(Q(\lambda)) + \lambda(H'_2 - H'_1)(Q(\lambda))} = -\frac{(H_2 - H_1)}{H'_1 + \lambda(H'_2 - H'_1)} \circ (H_1 + \lambda(H_2 - H_1))^{-1}(t)$$

Para la derivada segunda derivamos la derivada primera. A partir de

$$Q'(\lambda) = -\frac{(H_2 - H_1)(Q(\lambda))}{H'_1(Q(\lambda)) + \lambda(H'_2 - H'_1)(Q(\lambda))}$$

La derivada del numerador es: $[(H_2 - H_1)(Q(\lambda))]' = (H'_2 - H'_1)(Q(\lambda))Q'(\lambda)$, la derivada del denominador es:

$$[H'_1(Q(\lambda)) + \lambda(H'_2 - H'_1)(Q(\lambda))]' = H''_1(Q(\lambda))Q'(\lambda) + (H'_2 - H'_1)(Q(\lambda)) + \lambda(H''_2 - H''_1)(Q(\lambda))Q'(\lambda)$$

$$\text{Entonces, } Q''(\lambda) = -\left[\frac{(H'_2 - H'_1)(Q(\lambda))Q'(\lambda)(H'_1(Q(\lambda)) + \lambda(H'_2 - H'_1)(Q(\lambda)))}{[H'_1(Q(\lambda)) + \lambda(H'_2 - H'_1)(Q(\lambda))]^2} - \frac{(H_2 - H_1)(Q(\lambda))(H''_1(Q(\lambda))Q'(\lambda) + (H'_2 - H'_1)(Q(\lambda)) + \lambda(H''_2 - H''_1)(Q(\lambda))Q'(\lambda))}{[H'_1(Q(\lambda)) + \lambda(H'_2 - H'_1)(Q(\lambda))]^2} \right]$$

$$\text{Agrupando resulta que: } Q''(\lambda) = \frac{(H_2 - H_1)(Q(\lambda))(H'_2 - H'_1)(Q(\lambda))}{[H'_1(Q(\lambda)) + \lambda(H'_2 - H'_1)(Q(\lambda))]^2} - \frac{Q'(\lambda)[((H'_2 - H'_1)(H'_1 + \lambda(H'_2 - H'_1))(Q(\lambda)) - (H_2 - H_1)(Q(\lambda))[H''_1(Q(\lambda)) + \lambda(H''_2 - H''_1)(Q(\lambda))])]}{[H'_1(Q(\lambda)) + \lambda(H'_2 - H'_1)(Q(\lambda))]^2}$$

Entonces tenemos que:

$$Q''(\lambda) = \frac{-Q'(\lambda)(H'_2 - H'_1)(Q(\lambda))}{[H'_1(Q(\lambda)) + \lambda(H'_2 - H'_1)(Q(\lambda))]} - \frac{Q'(\lambda)(H'_2 - H'_1)(Q(\lambda))}{[H'_1(Q(\lambda)) + \lambda(H'_2 - H'_1)(Q(\lambda))]} + \frac{Q'(\lambda)(H_2 - H_1)(Q(\lambda))[H''_1(Q(\lambda)) + \lambda(H''_2 - H''_1)(Q(\lambda))]}{[H'_1(Q(\lambda)) + \lambda(H'_2 - H'_1)(Q(\lambda))]^2}$$

$$\text{Finalmente resulta } Q(\lambda)'' = Q'(\lambda) \left(\frac{-2(H'_2 - H'_1)}{[H'_1 + \lambda(H'_2 - H'_1)]} + \frac{[H_2 - H_1][H''_1 + \lambda(H''_2 - H''_1)]}{[H'_1 + \lambda(H'_2 - H'_1)]^2} \right) (Q(\lambda))$$

Lema 4.2.11. Si P es estrictamente creciente y las hipótesis anteriores se cumplen entonces para todo $\theta \in [0, 1]$ con $P'(\theta) > 0$

$$(4.15) \quad P_T(\theta) = P(\theta) + \kappa_2(K_d)h_d^2 \frac{P''(\theta)}{(P'(\theta))^2} + o(h_d) + O\left(\frac{1}{Th_d}\right)$$

donde $\kappa_2(K) = \frac{1}{2} \int_{-1}^1 v^2 K(v) dv$

Demostración. Como estamos suponiendo que $T=N$ probaremos

$$P_N(\theta) = P(\theta) + \kappa_2(K_d)h_d^2 \frac{P''(\theta)}{(P'(\theta))^2} + o(h_d) + O\left(\frac{1}{Nh_d}\right)$$

Por definición tenemos que $P_N(\theta) = \phi(P_N^{-1}(\theta))$ entonces,

$$P_N(\theta) - P(\theta) = \phi(P_N^{-1}(\theta)) - \phi(P^{-1}(\theta)) = Q(1) - Q(0) = Q'(\lambda^*) \text{ con } \lambda^* \in [0, 1].$$

Donde hemos considerado $Q(\lambda) = \phi(P^{-1} + \lambda(P_N^{-1} - P^{-1}))$ tomando $H_1 = P^{-1}$, $H_2 = P_N^{-1}$.

Como $Q'(\lambda) = -\frac{(H_2 - H_1)}{H_1' + \lambda(H_2' - H_1')} \circ (H_1 + \lambda(H_2 - H_1))^{-1}(\theta)$ en nuestro caso tenemos que

$$Q'(\lambda^*) = -\frac{(P_N^{-1} - P^{-1})}{P^{-1'} + \lambda^*(P_N^{-1} - P^{-1})'} \circ (P^{-1} + \lambda^*(P_N^{-1} - P^{-1}))^{-1}(\theta)$$

Por el lema 4.2.6 $P^{-1} + \lambda^*(P_N^{-1} - P^{-1}) \rightarrow P^{-1}$, entonces, $\theta_N = (P^{-1} + \lambda^*(P_N^{-1} - P^{-1}))^{-1}(\theta) \rightarrow P(\theta)$

Aplicando el teorema del valor medio

$$(P_N^{-1} - P^{-1})(\theta_N) - (P_N^{-1} - P^{-1})(P(\theta)) = (P_N^{-1} - P^{-1})'(\eta)(\theta_N - P(\theta))$$

con $\eta - P(\theta) \leq \theta_N - P(\theta)$. Ahora

$$\begin{aligned} (P_N^{-1} - P^{-1})'(\eta) &= (P_N^{-1})'(\eta) - (P^{-1})'(\eta) = \left[\int_{-\infty}^{\eta} \int_0^1 \frac{1}{h_d} K_d\left(\frac{P(x)-u}{h_d}\right) dx du + O\left(\frac{1}{Nh_d}\right) \right]'(\eta) - P^{-1'}(\eta) = \\ &= \int_0^1 \frac{1}{h_d} K_d\left(\frac{P(x)-\eta}{h_d}\right) dx - P^{-1'}(\eta) + O\left(\frac{1}{Nh_d}\right) \end{aligned}$$

A partir de la igualdad

$$\int_0^1 \int_{-\infty}^{\eta} K_d\left(\frac{P(x)-u}{h_d}\right) \frac{du}{h_d} dx = P^{-1}(\eta) + P^{-1''}(\theta)h_d^2\kappa_2(K_d) + o(h_d^2)$$

obtenemos $\int_0^1 K_d\left(\frac{P(x)-\eta}{h_d}\right) \frac{1}{h_d} dx = (P^{-1})'(\eta) + O(h_d^2)$ reemplazando queda

$$Q'(\lambda^*) = -\frac{(P_N^{-1} - P^{-1})(P(\theta)) + [O(h_d^2) + O(\frac{1}{Nh_d})](\theta_N - P(\theta))}{(P^{-1} + \lambda^*(P_N^{-1} - P^{-1}))'(\theta_N)}. \text{ Por el lema anterior tenemos que,}$$

$$P_N^{-1} - P^{-1} = \kappa_2(K_d)h_d^2(P^{-1}(\theta))'' + o(h_d^2) + O\left(\frac{1}{Nh_d}\right) \text{ entonces}$$

$$Q'(\lambda) = -\frac{\kappa_2(K_d)h_d^2(P^{-1}(\theta))'' + o(h_d^2) + O(\frac{1}{Nh_d}) + [O(h_d^2) + O(\frac{1}{Nh_d})](\theta_N - P(\theta))}{(P^{-1} + \lambda^*(P_N^{-1} - P^{-1}))'(\theta_N)} \text{ de donde}$$

$$Q'(\lambda) = \frac{\kappa_2(K_d)h_d^2 \frac{P''(\theta)}{P'(\theta)^3} + o(h_d^2) + O(\frac{1}{Nh_d}) + [O(h_d^2) + O(\frac{1}{Nh_d})](\theta_N - P(\theta))}{\frac{1}{P'(\theta)}}$$

donde se ha utilizado que $(P^{-1}(P(\theta)))'' = -\frac{P''(\theta)}{(P'(\theta))^3}$.

Como $\lim_N \frac{O(h_d^2)(\theta_N - P(\theta))}{h_d^2} = 0$ y $\lim_N \frac{O(\frac{1}{Nh_d})(\theta_N - P(\theta))}{\frac{1}{Nh_d}} = 0$, ya que $\theta_N \rightarrow P(\theta)$ obtenemos el resultado.

Asumiremos ahora que $h_r \rightarrow 0$, $h_d \rightarrow 0$, $Nh_r \rightarrow \infty$, $Nh_d \rightarrow \infty$ y adicionalmente que

$$Nh_r^5 = O(1), \quad n = O(T), \quad \frac{\log h_r^{-1}}{Nh_r h_d^3} = o(1)$$

Teorema 4.2.12. Si las condiciones anteriores son satisfechas, $\lim_{n \rightarrow \infty} \frac{h_r}{h_d} = c \in [0, \infty)$ existe y P es estrictamente creciente; entonces, para todo $\theta \in (P(0), P(1))$ con $P'(P^{-1}(\theta)) > 0$,

$$(4.16) \quad \sqrt{Nh_d}(\widehat{P}_I^{-1}(\theta) - P_T^{-1}(\theta) + \kappa_2(K_r)h_r^2 \left(\frac{P''f + 2P'f'}{Pf'} \right)(P^{-1}(\theta))) \Rightarrow N(0, r^2(\theta))$$

donde la constante $\kappa_2(K_r)$ está definida en el lema 4.2.6 y la varianza asintótica está dada por

$$(4.17) \quad r^2(\theta) = \frac{\sigma^2(P^{-1}(\theta))}{P'(P^{-1}(\theta))f(P^{-1}(\theta))} \times \int \int \int K_d(w + cP'(P^{-1}(\theta))(v - u))K_d(w)K_r(v)dw dv du$$

Si $\lim_{n \rightarrow \infty} \frac{h_r}{h_d} = \infty$, entonces tenemos, para todo $\theta \in (P(0), P(1))$ con $P'(P^{-1}(\theta)) > 0$,

$$\sqrt{N h_d}(\hat{P}_I^{-1}(\theta) - P_T^{-1}(\theta) + \kappa_2(K_r)h_r^2(\frac{P''f+2P'f'}{P f'})(P^{-1}(\theta))) \Rightarrow N(0, \bar{r}^2(\theta))$$

donde la varianza asintótica está dada por

$$\bar{r}^2(\theta) = \frac{\sigma^2(P^{-1}(\theta))}{P'(P^{-1}(\theta))f(P^{-1}(\theta))} \int K_r^2(u)du$$

Demostración. Usando la descomposición

$$(4.18) \quad \hat{P}_I^{-1}(\theta) = \frac{1}{N h_d} \int_{\infty}^{\theta} \sum_{i=1}^N K_d(\frac{\hat{P}(\frac{i}{N})-u}{h_d})du = P_N^{-1}(\theta) + \Delta_N(\theta)$$

donde hemos sumado y restado $P_T^{-1}(\theta) = \frac{1}{N h_d} \int_{\infty}^{\theta} \sum_{i=1}^T K_d(\frac{P(\frac{i}{N})-u}{h_d})du$ y tomado

$$(4.19) \quad \Delta_N(\theta) = \frac{1}{N h_d} \int_{\infty}^{\theta} \sum_{i=1}^N \left\{ K_d(\frac{\hat{P}(\frac{i}{N})-u}{h_d}) - K_d(\frac{P(\frac{i}{N})-u}{h_d}) \right\} du$$

Ahora descomponemos Δ_N en

$$(4.20) \quad \Delta_N(\theta) = \Delta_N^1(\theta) + \frac{1}{2} \Delta_N^2(\theta)$$

donde,

$$\Delta_N^1(\theta) = \frac{1}{N h_d^2} \int_{\infty}^{\theta} \sum_{i=1}^N K'_d(\frac{P(\frac{i}{N})-u}{h_d}) \{ \hat{P}(\frac{i}{N}) - P(\frac{i}{N}) \} du$$

$$\Delta_N^2(\theta) = \frac{1}{N h_d^3} \int_{\infty}^{\theta} \sum_{i=1}^N K''_d(\frac{\xi_i-u}{h_d}) \{ \hat{P}(\frac{i}{N}) - P(\frac{i}{N}) \}^2 du$$

con $|\xi_i - P(\frac{i}{N})| < |\hat{P}(\frac{i}{N}) - P(\frac{i}{N})|$ ($i = 1, \dots, n$) donde hemos usado que $f(x) = f(c) + f'(c)(x - c) + f''(\xi)(x - c)^2$. Luego

$$\Delta_N^2(\theta) = \frac{1}{N h_d^3} \sum_{i=1}^N \int_{\infty}^{\theta} K''_d(\frac{\xi_i-u}{h_d}) \{ \hat{P}(\frac{i}{N}) - P(\frac{i}{N}) \}^2 du = \frac{1}{N h_d^3} \sum_{i=1}^N \int_{\infty}^{\theta} K''_d(\frac{\xi_i-u}{h_d}) \{ \hat{P}(\frac{i}{N}) - P(\frac{i}{N}) \}^2 \frac{du}{h_d} =$$

$$\frac{1}{N h_d^2} \left| \sum_{i=1}^N K'_d(\frac{\xi_i-u}{h_d}) \right|_{\infty}^{\theta} \{ \hat{P}(\frac{i}{N}) - P(\frac{i}{N}) \}^2 = \frac{1}{h_d^2} \left| \frac{1}{N} \sum_{i=1}^N K'_d(\frac{\xi_i-\theta}{h_d}) \{ \hat{P}(\frac{i}{N}) - P(\frac{i}{N}) \}^2 \right| =$$

$$= \frac{1}{h_d^2} \left| \int_0^1 K'_d(\frac{P(x)-\theta}{h_d}) \{ \hat{P}(x) - P(x) \}^2 dx \right| \cdot (1 + o_p(1))$$

Si asumimos que el núcleo K_r ha sido modificado cerca de los bordes entonces tenemos que este término es de orden $O((\frac{1}{N h_r} + h_r^4)/h_d)$. Esto implica que

$$(4.21) \quad \sqrt{N h_d} \Delta_N^2(\theta) = o_p(1),$$

ya que,

$$\lim_{n \rightarrow \infty} \sqrt{N h_d} \left(\frac{1}{N h_r} + h_r^4 \right) (1 + o_p(1)) = \lim_{n \rightarrow \infty} \sqrt{N h_d} \left(\frac{1 + N h_r^5}{N h_r h_d} \right) O_p(1) = \lim_{n \rightarrow \infty} \frac{O_p(1)}{\sqrt{N h_d h_r}} =$$

$$\lim_{n \rightarrow \infty} \frac{O_p(1)}{n^{\frac{1}{2}} h_r^{\frac{3}{2}}} = \frac{O_p(1) h_r}{(N h_r^5)^{\frac{1}{2}}} = o_p(1).$$

Combinando 4.18, 4.20 y 4.21 muestra que la afirmación del teorema puede ser probada demostrando la convergencia débil

$$(4.22) \quad \sqrt{N h_d} \left(\Delta_N^1(\theta) + \kappa_2(K_r) h_r^2 \left(\frac{P'' f + 2P' f'}{P f'} \right) (P^{-1}(\theta)) \right) \xrightarrow{D} N(0, r^2(\theta))$$

Para esto usamos la descomposición

$$\Delta_N^1(\theta) = (\Delta_N^{(1,1)}(\theta) + \Delta_N^{(1,2)}(\theta))(1 + o_p(1))$$

con

$$\Delta_N^{(1,1)}(\theta) = \frac{-1}{n^2 h_d h_r} \sum_{i,j=1}^N K_d\left(\frac{P(\frac{i}{N}) - \theta}{h_d}\right) K_r\left(\frac{\Theta_j - \frac{i}{N}}{h_r}\right) \frac{P(\Theta_j) - P(\frac{i}{N})}{f(\frac{i}{N})}$$

$$\Delta_N^{(1,2)}(\theta) = \frac{-1}{n^2 h_d h_r} \sum_{i,j=1}^N K_d\left(\frac{P(\frac{i}{N}) - \theta}{h_d}\right) K_r\left(\frac{\Theta_j - \frac{i}{N}}{h_r}\right) \frac{\sigma(\Theta_j) \epsilon_j}{f(\frac{i}{N})}$$

ya que, $X_i = P(\Theta_i) + \sigma(\Theta_i) \epsilon_i$, $\sum_{j=1}^N K_r\left(\frac{\Theta_j - \frac{i}{N}}{h_r}\right) \rightarrow f$ y $\hat{P}(\Theta) = \frac{\sum_{j=1}^N K_r\left(\frac{\Theta_j - \Theta}{h_r}\right) X_j}{\sum_{j=1}^N K_r\left(\frac{\Theta_j - \Theta}{h_r}\right)}$ luego

$$\begin{aligned} (\Delta_N^{(1)}(\theta)) &= \frac{1}{n h_d^2} \sum_{i=1}^N \int_0^1 K_d'\left(\frac{P(\frac{i}{N}) - u}{h_d}\right) [(\hat{P}(\frac{i}{N}) - (P(\frac{i}{N}))) du = \\ &= \frac{1}{n h_d} \sum_{i=1}^N (-K_d\left(\frac{P(\frac{i}{N}) - u}{h_d}\right)|_{\infty}^{\theta}) [(\hat{P}(\frac{i}{N}) - (P(\frac{i}{N}))) \\ &= \frac{-1}{n h_d} \sum_{i=1}^N (K_d\left(\frac{P(\frac{i}{N}) - \theta}{h_d}\right)) [(\hat{P}(\frac{i}{N}) - (P(\frac{i}{N}))) \end{aligned}$$

Como

$$\hat{P}(\frac{i}{N}) = \frac{\sum_{j=1}^N K_r\left(\frac{\frac{i}{N} - \Theta_j}{h_r}\right) X_j}{\sum_{j=1}^N K_r\left(\frac{\frac{i}{N} - \Theta_j}{h_r}\right)}$$

Entonces,

$$\hat{P}(\frac{i}{N}) - P(\frac{i}{N}) = \frac{\sum_{j=1}^N K_r\left(\frac{\Theta_j - \frac{i}{N}}{h_r}\right) (X_j - P(\frac{i}{N}))}{\sum_{j=1}^N K_r\left(\frac{\Theta_j - \frac{i}{N}}{h_r}\right)} = \frac{1}{N h_r} \frac{\sum_{j=1}^N K_r\left(\frac{\Theta_j - \frac{i}{N}}{h_r}\right) (P(\Theta_j) + \sigma(\Theta_j) \epsilon_j - P(\frac{i}{N}))}{\frac{1}{N h_r} \sum_{j=1}^N K_r\left(\frac{\Theta_j - \frac{i}{N}}{h_r}\right)}$$

Entonces tenemos que

$$\begin{aligned} \Delta_N^1(\theta) &= -\frac{1}{n^2 h_r h_d} \frac{\sum_{i=1}^N \sum_{j=1}^N K_d\left(\frac{P(\frac{i}{N}) - \theta}{h_d}\right) K_r\left(\frac{\Theta_j - \frac{i}{N}}{h_r}\right) (P(\Theta_j) + \sigma(\Theta_j) \epsilon_j - P(\frac{i}{N}))}{\hat{f}(\frac{i}{N})} \\ \Delta_N^1(\theta) &= -\frac{1}{n^2 h_r h_d} \frac{\sum_{i=1}^N \sum_{j=1}^N K_d\left(\frac{P(\frac{i}{N}) - \theta}{h_d}\right) K_r\left(\frac{\Theta_j - \frac{i}{N}}{h_r}\right) (P(\Theta_j) + \sigma(\Theta_j) \epsilon_j - P(\frac{i}{N}))}{f(\frac{i}{N})} (1 + o_p(1)) \end{aligned}$$

Entonces,

$$\begin{aligned} \Delta_N^1 &= (\Delta_N^{(1,1)}(\theta) + \Delta_N^{(1,2)}(\theta))(1 + o_p(1)) \\ E(\Delta_N^{(1,1)}(\theta)) &= \frac{-1}{n h_d h_r} \sum_{i=1}^N \int_0^1 K_d\left(\frac{P(\frac{i}{N}) - \theta}{h_d}\right) K_r\left(\frac{y - \frac{i}{N}}{h_r}\right) \frac{P(y) - P(\frac{i}{N})}{f(\frac{i}{N})} f(y) dy (1 + o(1)) = \\ &= \frac{-1}{h_d h_r} \int_0^1 \int_0^1 K_d\left(\frac{P(x) - \theta}{h_d}\right) K_r\left(\frac{y - x}{h_r}\right) \frac{P(y) - P(x)}{f(x)} f(y) dy dx (1 + o(1)) \end{aligned}$$

Como $|\frac{y-x}{h_r}| < 1$ tenemos que $x - h_r < y < x + h_r$

$$\int_{x-h_r}^{x+h_r} K_r\left(\frac{y-x}{h_r}\right) \frac{P(y) - P(x)}{f(x)} f(y) dy = \int_{-1}^1 K_r(v) h_r \frac{P(x+h_r v) - P(x)}{f(x)} f(x+h_r v) dv =$$

Donde en la última integral hemos hecho el cambio de variables $v = \frac{y-x}{h_r}$.

Ahora como $f(x + h_r v) = f(x) + f'(x) h_r v + o(h_r v)$, $P(x + h_r v) = P(x) + P'(x) h_r v + \frac{1}{2} P''(x) h_r^2 v^2 + o(h_r^2 v^2)$

$$= \int_{-1}^1 K_r(v) h_r \frac{P'(x) h_r v + \frac{1}{2} P''(x) h_r^2 v^2 + o(h_r^2 v^2)}{f(x)} (f(x) + f'(x) h_r v + o(h_r v)) dv =$$

$$\begin{aligned}
&= h_r^2 P'(x) \int_{-1}^1 K_r(v) v dv + \frac{1}{2} P''(x) h_r^3 \int_{-1}^1 v^2 K_r(v) dv + h_r^3 \frac{P'(x) f'(x)}{f(x)} \int_{-1}^1 K(v) v^2 dv + \\
&\quad + \frac{h_r^2}{2} \frac{P''(x) f'(x)}{f(x)} \int_{-1}^1 v^3 K_r(v) dv + o(h_r^3)
\end{aligned}$$

Entonces,

$$E(\Delta_N^{(1,1)}(\theta)) = -h_r^2 \kappa_2(K_r) \int_0^1 \frac{1}{h_d} K_d\left(\frac{P(x)-\theta}{h_d}\right) \left\{ P''(x) + \frac{2P'(x)f'(x)}{f(x)} \right\} \cdot (1 + o(1)) =$$

Haciendo el cambio de variable $z = P(x)$ y aplicando Bochner

$$= -h_r^2 \kappa_2(K_r) \left(\frac{P''f + 2P'f'}{fP'}(P^{-1}(\theta)) \right) \cdot (1 + o(1))$$

Calculemos la varianza

$$Var(\Delta_N^{(1,1)}(\theta)) = \frac{N}{n^4 h_d^2 h_r^2} Var(\sum_{i=1}^N K_d\left(\frac{P(\frac{i}{N})-\theta}{h_d}\right) K_r\left(\frac{\Theta_j - \frac{i}{N}}{h_r}\right) \frac{P(\Theta_j) - P(\frac{i}{N})}{f(\frac{i}{N})}) \leq$$

usando el hecho que $E(\Theta^2) - E(\Theta)^2 \leq E(\Theta^2)$

$$\leq \frac{1}{n h_d^2 h_r^2} E\left(\left(\int_0^1 K_d\left(\frac{P(x)-\theta}{h_d}\right) K_r\left(\frac{\Theta_j - x}{h_r}\right) \frac{P(\Theta_j) - P(x)}{f(x)} dx\right)^2 (1 + o(1))\right) = o\left(\frac{1}{N h_d}\right).$$

Esto pues $|\int_0^1 K_d\left(\frac{P(x)-\theta}{h_d}\right) K_r\left(\frac{\Theta_j - x}{h_r}\right) \frac{P(\Theta_j) - P(x)}{f(x)} dx| \leq C$, luego,

$$\begin{aligned}
&\frac{1}{N h_d^2 h_r^2} E\left(\left(\int_0^1 K_d\left(\frac{P(x)-\theta}{h_d}\right) K_r\left(\frac{\Theta_j - x}{h_r}\right) \frac{P(\Theta_j) - P(x)}{f(x)} dx\right)^2 (1 + o(1))\right) \leq \\
&\quad C \frac{1}{n h_d^2 h_r^2} E\left(\int_0^1 K_d\left(\frac{P(x)-\theta}{h_d}\right) K_r\left(\frac{\Theta_j - x}{h_r}\right) \frac{P(\Theta_j) - P(x)}{f(x)} dx\right) (1 + o(1)) \\
&\leq \frac{1}{N h_d^2 h_r^2} C h_d h_r (h_r^2) \kappa_2(K_r) \left(\frac{P''f + 2P'f'}{fP'}(P^{-1}(\theta)) \right) \\
&= \frac{h_r}{N h_d} C \kappa_2(K_r) \left(\frac{P''f + 2P'f'}{fP'}(P^{-1}(\theta)) \right) = o\left(\frac{1}{N h_d}\right)
\end{aligned}$$

Esto implica operando y usando las condiciones asumidas, que

$$\Delta_N^{(1,1)}(\theta) + \kappa_2(K_r) h_r^2 \left(\frac{P''f + 2P'f'}{fP'} \right)(P^{-1}(\theta)) = o_p\left(\frac{1}{\sqrt{N h_d}}\right)$$

y por lo tanto 4.22 se deduce de

$$\sqrt{N h_d} \Delta_N^{(1,2)}(\theta) \xrightarrow{D} N(0, r^2(\theta))$$

Para probar esta afirmación notemos que $E[\Delta_N^{(1,2)}(\theta)] = 0$ y calculamos la varianza

$$\begin{aligned}
var(\sqrt{N h_d} \Delta_N^{(1,2)}(\theta)) &= \frac{1}{n^3 h_d h_r^2} \sum_{j=1}^N var \sum_{i=1}^N \left(K_d\left(\frac{P(\frac{i}{N})-\theta}{h_d}\right) K_r\left(\frac{\Theta_j - \frac{i}{N}}{h_r}\right) \frac{\sigma(\Theta_j) \epsilon_j}{f(\frac{i}{N})} \right) = \\
&= \frac{1}{h_d h_r^2} \int_0^1 \sigma^2(x) \int_0^1 \left(K_d\left(\frac{P(y)-\theta}{h_d}\right) K_r\left(\frac{x-y}{h_r}\right) \frac{dy}{f(y)} \right)^2 f(x) dx \cdot (1 + o(1)) = \\
&= \frac{1}{h_d h_r^2} \int_0^1 K_d\left(\frac{P(z)-\theta}{h_d}\right) \int_0^1 K_d\left(\frac{P(y)-\theta}{h_d}\right) \frac{1}{f(y)f(z)} \times \\
&\quad \times \int_0^1 \sigma^2(x) K_r\left(\frac{x-y}{h_r}\right) K_r\left(\frac{x-z}{h_r}\right) f(x) dx dy dz \cdot (1 + o(1)) =
\end{aligned}$$

Ahora hacemos el cambio de variables $w = \frac{P(z)-\theta}{h_d}$, entonces $z = P^{-1}(\theta + h_d w)$, $v = \frac{P(y)-\theta}{h_d}$, entonces $y = P^{-1}(\theta + h_d v)$, por último, $u = \frac{x-y}{h_r}$, entonces, $x = y + h_r u$ luego, $x = P^{-1}(\theta + h_d v) + h_r u$.

$$\frac{x-z}{h_r} = \frac{P^{-1}(\theta + h_d v) + h_r u - P^{-1}(\theta + h_d w)}{h_r} = \frac{P^{-1}(\theta + h_d v) - P^{-1}(\theta + h_d w)}{h_r} + u.$$

El Jacobiano nos da $|J| = h_r \left(\frac{h_d^2}{P(P^{-1}(\theta))^2} + o(h_d) \right)$ tenemos también

$P^{-1}(\theta + h_d w) = P^{-1}(\theta) + P^{-1'}(\theta) h_d w + P^{-1''} h_d^2 w^2$ entonces lo anterior queda:

$$\begin{aligned}
&= \frac{1}{h_d h_r^2} \int_0^1 K_d(w) \int_0^1 K_d(v) \frac{1}{f(P^{-1}(\theta+h_d v))f(P^{-1}(\theta+h_d w))} \int_0^1 \sigma^2(P^{-1}(\theta+h_d v)+h_r u) K_r(u) \times \\
&\quad \times K_r\left(\frac{P^{-1}(\theta+h_d v)-P^{-1}(\theta+h_d w)}{h_r} + u\right) \frac{h_d^2 h_r}{P'(P^{-1}(\theta))^2} f(P^{-1}(\theta)+h_d+h_r u) du dv dw \\
&= \frac{h_d}{h_r} \int_0^1 K_d(w) \int_0^1 K_d(v) \frac{1}{f(P^{-1}(\theta))f(P^{-1}(\theta))} \int_0^1 \sigma^2(P^{-1}(\theta)) K_r(u) \times \\
&\quad \times K_r\left(\frac{P^{-1}(\theta+h_d v)-P^{-1}(\theta+h_d w)}{h_r} + u\right) \frac{h_d^2 h_r}{P'(P^{-1}(\theta))^2} f(P^{-1}(\theta)) du dv dw = \\
&= \frac{h_d \sigma^2(P^{-1}(\theta))}{h_r P'(P^{-1}(\theta))^2 f(P^{-1}(\theta))} \int_0^1 K_d(w) \int_0^1 K_d(v) \frac{1}{f(P^{-1}(\theta))f(P^{-1}(\theta))} \int_0^1 K_r(u) \times \\
&\quad \times K_r\left(\frac{P^{-1}(\theta+h_d v)-P^{-1}(\theta+h_d w)}{h_r} + u\right) du dv dw =
\end{aligned}$$

Haciendo el cambio de variables $y = \frac{P^{-1}(\theta+h_d v)-P^{-1}(\theta+h_d w)}{h_r} + u$, $w = w$, $u = u$, $v = \frac{P(h_r(y-u))+P^{-1}(\theta+h_d u)-\theta}{h_d}$ cuyo Jacobiano $|J| = \frac{h_r}{h_d} P'(P^{-1}(\theta))(1+o(1))$

$$= \frac{h_d \sigma^2(P^{-1}(\theta))}{h_r P'(P^{-1}(\theta))f(P^{-1}(\theta))} \int \int \int K_d(w) K_d\left(\frac{P(h_r(y-u))+P^{-1}(\theta+h_d u)-\theta}{h_d}\right) K_r(u) \times K_r(y) du dw dy =$$

Aplicando Taylor tomando $x = P^{-1}(\theta+h_d w)$ tenemos que

$$\lim_{h_r \rightarrow 0, h_d \rightarrow 0} \frac{h_r}{h_d} K_d\left(\frac{P(P^{-1}(\theta+h_d v)+h_r(v-u))-\theta}{h_d}\right) = K_d(w + cP'(P^{-1}(\theta))(v-u))$$

, entonces,

$$= \frac{h_d \sigma^2(P^{-1}(\theta))}{h_r P'(P^{-1}(\theta))f(P^{-1}(\theta))} \int \int \int K_d(w) K_d(w + cP'(P^{-1}(\theta))(v-u)) K_r(u) \times K_r(y) du dw dy =$$

Con esto hemos probado la varianza asintótica en 4.17. Para la normalidad asintótica calculamos de manera análoga

$$\begin{aligned}
E((\sqrt{N} h_d \Delta_N^{(1,2)}(\theta))^4) &= \sum_{j=1}^N E \left[\left\{ \frac{\sigma(\Theta_j)}{n^{3/2} h_d^{1/2} h_r} \epsilon_j \sum_{i=1}^N K_d\left(\frac{P(\frac{i}{N})-\theta}{h_d}\right) K_r\left(\frac{\Theta_j - \frac{i}{N}}{h_r}\right) \frac{1}{f(\frac{i}{N})} \right\}^4 \right] = \\
&= \sum_{j=1}^N E \left[\int_0^1 \frac{(\sigma(\Theta) \epsilon_j)^4}{n^6 h_d^2 h_r^4} \left\{ n \int_0^1 K_d\left(\frac{P(y)-\theta}{h_d}\right) K_r\left(\frac{\Theta_j - y}{h_r}\right) \frac{1}{f(y)} dy \right\}^4 f(x) \right] = \\
&= \frac{n n^4}{n^6 h_d^2 h_r^4} E(\epsilon_j^4) \left[\int_0^1 \sigma^4(x) \left\{ \int_0^1 K_d\left(\frac{P(\Theta_1)-\theta}{h_d}\right) K_r\left(\frac{x-\Theta_1}{h_r}\right) \frac{1}{f(\Theta_1)} dx_1 \right\} f(x_1) \dots \right. \\
&\quad \left. \left\{ \int_0^1 K_d\left(\frac{P(\Theta_4)-\theta}{h_d}\right) K_r\left(\frac{x-\Theta_4}{h_r}\right) \frac{1}{f(\Theta_4)} dx_4 \right\} f(x_4) \right] =
\end{aligned}$$

Ahora hacemos el cambio de variable $y_1 = \frac{P(x_1)-\theta}{h_d}$ entonces $x_1 = P^{-1}(\theta+y_1 h_d)$, $y_2 = \frac{P(x_2)-\theta}{h_d}$
 $y_3 = \frac{P(x_3)-\theta}{h_d}$ $y_4 = \frac{P(x_4)-\theta}{h_d}$ $\tilde{x} = \frac{x-x_1}{h_r}$. El Jacobiano $|J| = \frac{h_d^4}{(P'(P^{-1}(\theta)))^5} h_r$

$$= \frac{N}{h_d^2 h_r^4} E(\epsilon_1)^4 \int_0^1 \sigma^4(P^{-1}(\theta)) \frac{h_d^4 h_r}{(f(P^{-1}(\theta)))^4 [P'(P^{-1}(\theta))]^5} \int_0^1 K_d(y_1) K_r(\tilde{x}) \cdot$$

$$\begin{aligned}
&\int K_d(y_2) K_r(\tilde{x} + \frac{P^{-1}(\theta+h_d y_1)-P^{-1}(\theta+h_d y_1)}{h_r}) \int_0^1 K_d(y_3) K_r(\tilde{x} + \frac{P^{-1}(\theta+h_d y_3)-P^{-1}(\theta+h_d y_3)}{h_r}) \\
&\int_0^1 K_d(y_4) K_r(\tilde{x} + \frac{P^{-1}(\theta+h_d y_4)-P^{-1}(\theta+h_d y_4)}{h_r}) =
\end{aligned}$$

Ahora $x = x_1 + h_r \tilde{x}$ entonces $x = P^{-1}(\theta+h_d y_1) + h_r \tilde{x}$

$$= \frac{\sigma^4(P^{-1}(\theta)) h_d^2}{N h_r^3} \frac{E(\epsilon_1)^4}{f(P^{-1}(\theta))^3 P'(P^{-1}(\theta))^5} \int \int K_d(y_1) K_r(\tilde{x}) \left[\prod_{j=2}^4 \int K_d(y_j) K_r(\tilde{x} + \frac{P^{-1}(\theta+h_d y_3)-P^{-1}(\theta+h_d y_3)}{h_r}) dy_j \right] dy_1 d\tilde{x}$$

Hacemos el cambio de variables $\tilde{x} = \tilde{x}$, $y_1 = y_1$ $y_j = \frac{P((P^{-1}(\theta+h_d \tilde{x})+h_r(y_j-y_1-z))}{h_d}$

calculemos $P^{-1}(\theta + h_d y_j) = \frac{P^{-1}(\theta + h_d P(P^{-1}(\theta + h_d \tilde{x}) + h_r(y_j - y_1) - \theta))}{h_d} = P^{-1}(\theta + h_d \tilde{x}) + h_r(y_j - y_1)$.

Calculemos el Jacobiano $|J| = \frac{h_r^3 P'(P^{-1}(\theta))^3}{h_d^3}$

Entonces,

$$= \frac{\sigma^4(P^{-1}(\theta))h_d^2}{N h_r^3} \frac{E(\epsilon_1)^4}{f(P^{-1}(\theta))^3 P'(P^{-1})^5} \int \int K_d(y_1) K_r(\tilde{x}).$$

$$\left[\prod_{j=2}^4 \int K_r(y_j) K_d(y_1 + \frac{h_r}{h_d} P'(P^{-1}(\theta)(y_j - \tilde{x})) dy_j \right] d\tilde{x} dy_1 \cdot \frac{h_r^3 P'(P^{-1}(\theta))^3}{h_d^3} = O(\frac{1}{N h_d}) = o(1)$$

Teorema 4.2.13. Asuma las hipótesis del teorema anterior y sean \hat{P}_I y P_T denotan las inversas de las funciones \hat{P}_I^{-1} y P_T^{-1} definidas por 4.7 y 4.8 respectivamente. Si $\lim_{n \rightarrow \infty} \frac{h_r}{h_d} = c \in [0, \infty)$ existe, entonces tenemos, que para $\theta \in (0, 1)$ con $P'(\theta) > 0$,

$$\sqrt{N h_d} \left(\hat{P}_I(\theta) - P_T(\theta) - \kappa_2(K_r) h_r^2 \left(\frac{P'' f + 2P' f'}{f} \right) (\theta) \right) \Rightarrow N(0, s^2(\theta))$$

donde la varianza asintótica está dada por

$$s^2(\theta) = \frac{\sigma^2(\theta) P'(\theta)}{f(\theta)} \int \int \int K_d(w + c P'(\theta)(v - u)) K_d(w) K_r(u) K_r(v) dw du dv$$

Si $\lim_{n \rightarrow \infty} \frac{h_r}{h_d} = \infty$ tenemos que para todo $\theta \in (0, 1)$ con $P'(\theta) > 0$,

$$\sqrt{N h_d} \left(\hat{P}_I(\theta) - P_T(\theta) - \kappa_2(K_r) h_r^2 \left(\frac{P'' f + 2P' f'}{f} \right) (\theta) \right) \Rightarrow N(0, \tilde{s}^2(\theta))$$

donde la varianza asintótica está dada por $\tilde{s}^2(\theta) = \frac{\sigma^2(\theta)}{f(\theta)} \int K_r^2(u) du$.

Demostración. Del Lema 4.2.10 obtenemos del desarrollo de Taylor

$$H_2^{-1}(\theta) - H_1^{-1} = \Phi(H_2)(\theta) - \Phi(H_1)(\theta) = \Phi(H_1 + H_2 - H_1) - \Phi(H_1) = Q(1) - Q(0) = Q'(0) + \frac{1}{2} Q''(\lambda^*)$$

$$\text{con } H_2 = \hat{P}_I^{-1} \quad H_1 = P_N^{-1}.$$

Definamos además $A_N = Q'(0) = -\frac{\hat{P}_I^{-1} - P_N^{-1}}{P_N^{-1}} \circ P_N(\theta)$ y

$$B_N = Q''(\lambda^*) = \frac{2(\hat{P}_I^{-1} - P_N^{-1})(\hat{P}_I^{-1} - P_N^{-1})'}{P_N^{-1} + \lambda^*(\hat{P}_I^{-1} - P_N^{-1})} \circ (P_N^{-1} + \lambda^*(\hat{P}_I^{-1} - P_N^{-1}))(\theta) - \\ - \frac{(\hat{P}_I^{-1} - P_N^{-1})^2 (P_N^{-1} + \lambda^*(\hat{P}_I^{-1} - P_N^{-1}))''}{(P_N^{-1} + \lambda^*(\hat{P}_I^{-1} - P_N^{-1}))^3} \circ (P_N^{-1} + \lambda^*(\hat{P}_I^{-1} - P_N^{-1}))(\theta)$$

o sea $P_I(\theta) - P_N(\theta) = A_N + \frac{1}{2} B_N$. Mostraremos que

$$(4.23) \quad A_N = -\frac{(\hat{P}_I^{-1} - P_N^{-1})}{(P^{-1})'} \circ P(\theta) + o_p\left(\frac{1}{\sqrt{N h_d}}\right)$$

y que,

$$(4.24) \quad B_N = o_p\left(\frac{1}{\sqrt{N h_d}}\right)$$

Para probar 4.23 consideramos

$$D_N = (\hat{P}_I^{-1}(\theta) - P_N^{-1}(\theta)) \circ P_N(\theta) - (\hat{P}_I^{-1}(\theta) - P_N^{-1}(\theta)) \circ P(\theta) = (\hat{P}_I^{-1}(\theta) - P_N^{-1}(\theta))'(\zeta_N)(P_N(\theta) - P(\theta))$$

donde $|\zeta_N - P(\theta)| \leq |P_N(\theta) - P(\theta)|$. Como $(\hat{P}_I^{-1}(\theta) - P_N^{-1}(\theta)) = \Delta_N$, entonces,

$$(\hat{P}_I^{-1}(\theta) - P_N^{-1}(\theta))'(\zeta_N) = \Delta'_N(\zeta_N) = \frac{1}{N h_d} \sum_{i=1}^N (K_d(\frac{\zeta_N - \hat{P}(\frac{i}{N})}{h_d}) - K_d(\frac{\zeta_N - P(\frac{i}{N})}{h_d})) =$$

$$= \frac{1}{Nh^2} \sum_{i=1}^N (K'_d(\frac{\zeta_{i,n}-\varphi_N}{h_d})(\hat{P}(\frac{i}{N}) - P(\frac{i}{N})))$$

$$\text{con } |\zeta_{i,n} - P(\frac{i}{N})| \leq |\hat{P}(\frac{i}{N}) - P(\frac{i}{N})| = O(R_N) \text{ c.s. siendo } R_N = \left(\frac{\log h_r^{-1}}{Nh_r}\right)^{1/2}$$

$$\Delta'_N(\zeta_N) = \frac{1}{Nh_d^2} \sum_{i=1}^N (K'_d(\frac{\zeta_N - P(\frac{i}{N})}{h_d})(\hat{P}(\frac{i}{N}) - P(\frac{i}{N}))) + \frac{1}{Nh_d^2} \sum_{i=1}^N (K'_d(\frac{\eta - \varphi_N}{h_d}) - K'_d(\frac{P(\frac{i}{N}) - \varphi_N}{h_d})(\hat{P}(\frac{i}{N}) - P(\frac{i}{N})))$$

donde para el segundo término de la última igualdad tenemos que

$$L_N = \frac{N}{Nh_d^2} K''_d(\frac{u - \varphi_N}{h_d})(\frac{\eta - \varphi_N}{h_d} - \frac{P(\frac{i}{N}) - \varphi_N}{h_d}) O(R_N) = \frac{O(R_N^2)}{h_d^3}$$

Entonces tenemos que:

$$\begin{aligned} \Delta'_N(\zeta_N) &= \frac{1}{Nh_d^2} \sum_{i=1}^N (K'_d(\frac{\zeta_N - P(\frac{i}{N})}{h_d})(\hat{P}(\frac{i}{N}) - P(\frac{i}{N}))) + O(\frac{R_N^2}{h_d^3}) \text{ c.s.} \\ &= \frac{1}{h_d^2} \int (K'_d(\frac{P(x) - P(\theta)}{h_d})(\hat{P}(x) - P(x)) dx + O(R_N + \frac{R_N^2}{h_d^3} + \frac{1}{Nh_d}) \text{ c.s.} \\ &= O(\frac{R_N}{h_d} + \frac{R_N^2}{h_d^2} + \frac{1}{Nh_d}) \text{ c.s.} \end{aligned}$$

Tenemos que: $D_N = (\hat{P}_I^{-1} - P_N^{-1})(\varphi_N)(P_N(\theta) - P(\theta)) = O(R_N h_d + \frac{R_N^2}{h_d} + \frac{h_d}{N}) = o(\frac{1}{\sqrt{Nh_d}})$ c.s. luego como $(P_N^{-1})'(\theta) = (P^{-1})'(\theta) + o(1)$ sale la igualdad

$$A_N = -\frac{\hat{P}_I^{-1} - P_N^{-1}}{P_N^{-1}} \circ P_N(\theta) = -\frac{\hat{P}_I^{-1} - P_N^{-1}}{P_N^{-1}} \circ P(\theta) + D_N = -\frac{\hat{P}_I^{-1} - P_N^{-1}}{P_N^{-1}} \circ P_N(\theta) + o(\frac{1}{\sqrt{Nh_d}}) \text{ c.s.}$$

Para B_N notemos que $B_N = 2B_{N1} - B_{N2}$ con

$$B_{N1} = \frac{(\hat{P}_I^{-1} - P_N^{-1})(\hat{P}_I^{-1} - P_N^{-1})'(\theta_N)}{(\hat{P}_I^{-1} + \lambda^*(\hat{P}_I^{-1} - P_N^{-1}))^2(\theta_N)} \text{ y } B_{N2} = \frac{(\hat{P}_I^{-1} - P_N^{-1})^2(\hat{P}_I^{-1} + \lambda^*(\hat{P}_I^{-1} - P_N^{-1}))''}{(\hat{P}_I^{-1} + \lambda^*(\hat{P}_I^{-1} - P_N^{-1}))^3(\theta_N)}$$

donde $\theta_N = (P_N^{-1} + \lambda^*(\hat{P}_I^{-1} - P_N^{-1}))^{-1}(\theta)$. Como $\hat{P}_I^{-1} - P_N^{-1} \rightarrow 0$ y $P_N^{-1} \rightarrow P^{-1}$ entonces $\theta_N \rightarrow (P^{-1})^{-1}(\theta) = P(\theta)$ de donde el denominador de B_{N1} tiende a θ^2 .

Recordemos que si $\sqrt{Nh_d}Z \rightarrow N(0, \sigma^2)$ entonces $Z = O_p(\frac{1}{\sqrt{Nh_d}})$ y si $\Theta_N + a_N = O_p(b_N)$ entonces $\Theta_N = O_p(b_N)$ luego $P_I^{-1} - P_N^{-1} = O_p(\frac{1}{\sqrt{Nh_d}})$ además $(P_I^{-1} - P_N^{-1})' = O_p(\frac{R_N}{h_d} + \frac{R_N^2}{h_d^3} + \frac{1}{Nh_d}) = O_p(\frac{1}{\sqrt{Nh_d}}(\frac{R_N}{h_d} + \frac{R_N^2}{h_d^3})) = o_p(\frac{1}{\sqrt{Nh_d}})$ pues $\frac{R_N}{h_d} \rightarrow 0$ ya que $\frac{(\log h_r^{-1})^{1/2}}{n^{1/2}h_r^{1/2}h_d} = \frac{(\log h_r^{-1})^{1/2}h_r h_r}{n^{1/2}h_r^{5/2}h_d} \rightarrow 0$ y $\frac{R_N^2}{h_d^3} = \frac{\log h_r^{-1}}{Nh_r h_d^3} = o(1)$ por hipótesis o sea $B_{n1} = o_p(\frac{1}{\sqrt{Nh_d}})$

Igualmente $B_{n2} = O_p(\frac{1}{Nh_d})$ lo que prueba la segunda igualdad.

Tenemos que:

$$\begin{aligned} &\sqrt{Nh_d}(\hat{P}_I(\theta) - P_N(\theta) - \kappa_2(K_r)h_r^2(\frac{P''f+2P'f'}{f})(\theta)) = \\ &= \sqrt{Nh_d}(\frac{\hat{P}_I^{-1}(\theta) - P_N^{-1}(\theta)}{(P_N^{-1})'} \circ P_N(\theta) + \frac{1}{2}B_N - \kappa_2(K_r)h_r^2(\frac{P''f+2P'f'}{f})(\theta)) = \\ &= -\sqrt{Nh_d}(\frac{\hat{P}_I^{-1}(\theta) - P_N^{-1}(\theta)}{(P_N^{-1})'} \circ P_N(\theta) + o_p(\frac{1}{\sqrt{Nh_d}}) + \frac{1}{2}o_p(\frac{1}{\sqrt{Nh_d}}) - \kappa_2(K_r)h_r^2(\frac{P''f+2P'f'}{f})(\theta)) = \\ &= -\frac{\sqrt{Nh_d}(\hat{P}_I^{-1}(\theta) - P_N^{-1}(\theta)) \circ P_N(\theta) - \kappa_2(K_r)h_r^2(\frac{P''f+2P'f'}{f})(\theta)(P_N^{-1})' \circ P_N(\theta)}{(P^{-1})' \circ P(\theta)} + o_p(1) = \\ &= -P'(\theta)\sqrt{Nh_d}(\hat{P}_I^{-1}(\theta) - P_N^{-1}(\theta)) \circ P_N(\theta) - \kappa_2(K_r)h_r^2(\frac{P''f+2P'f'}{P'f})(\theta)) + o_p(1) = Z_N \end{aligned}$$

Usando el teorema anterior en $P(\theta)$ $Z_N \Rightarrow N(0, s^2(\theta))$ con

$$\begin{aligned} s^2(\theta) &= \frac{\sigma^2(P^{-1}(\theta))}{P'(P^{-1}(\theta))f(P^{-1}(\theta))} \times \int \int \int K_d(w + cP'(P^{-1}(\theta)))(v - u)K_d(w)K_r(u)K_r(v)dwdudv = \\ &= \frac{\sigma^2(\theta)P'(\theta)}{f(\theta)} \times \int \int \int K_d(w + cP'(P^{-1}(\theta)))(v - u)K_d(w)K_r(u)K_r(v)dwdudv \end{aligned}$$

lo que demuestra el teorema.