Divulgaciones Matemáticas Vol. 16 No. 1(2008), pp. 1–27

Estimación Paramétrica de la Función de Regresión en un Modelo No Lineal

Parametric Estimation of the Regression Function in a nonlinear model

María Margarita Olivares (molivar@euler.ciens.ucv.ve)
Escuela de Matemática. Facultad de Ciencias. U.C.V

Harú Martinez (martinezh@agr.ucv.ve)
Instituto de Ingeniería Agrícola
Facultad de Agronomía U.C.V.

Abstract

Se construye un estimador paramétrico de la función de regresión en un modelo no lineal basándonos en una técnica de mínima distancia utilizada por Beran para estimación de densidades (1977). Obtenemos consistencia y un teorema central del límite para estos estimadores.
Palabras Claves: Modelo de Regresión no lineal, mínima distancia, estimación.

Abstract

In this paper we construct estimators of the parameter that identifies a regression function in the nonlinear regression model using Beran Technique for densities estimation (1977). We give consistency and a central limit theorem for these estimators.
Key words and phrases: Regresion nonlinear model, minimum distance, estimation.

1 Introducción

Consideremos un modelo de uso frecuente en áreas aplicadas, como es el modelo de regresión no lineal, el cual consiste en suponer que un conjunto de datos $(t_i, x_i), i = 1, 2, \cdots, n$ verifican la ecuación:

$$x_i = h(t_i) + \varepsilon_i$$

Recibido 2006/02/25. Revisado 2006/06/21. Aceptado 2006/07/15.
MSC (2000): Primary 62J02; Secondary 62F12.

Los puntos t_i se consideran determinísticos y los ε_i son variables aleatorias independientes y equidistribuidas. Si se supone que la función h depende de una familia paramétrica, es decir:

$$h(t) = h(t, \theta) \text{ con } \theta \in \mathbb{R}^p$$

un problema importante se refiere a obtener estimadores del parámetro θ a partir de las observaciones. Para obtener tales estimadores y demostrar su comportamiento asintótico aplicamos una técnica de mínima distancia utilizada por Beran (1977) para presentar los estimadores de distancia de Hellinger mínima en la estimación de densidades. Este método de mínima distancia fue desarrollado por J. Wolfowitz (1957) y ha sido ampliamente usado para obtener estimadores fuertemente consistentes de funciones de distribución, densidades y de regresión.

En regresión, Pak (1996) usa la técnica de Beran (1977), considerando un modelo de regresión lineal y la distancia de Hellinger. Construye estimadores de densidad basados en los errores estandarizados y sobre estos usa la distancia de Hellinger.

En este trabajo se aplica de una forma diferente el procedimiento de Beran (1977) en la función regresora y considerando que ésta no es una densidad, se utiliza la distancia usual de $L^2[0,1]$. Para el estimador planteado se demuestra su consistencia y se establece un Teorema Central del Límite.

Preliminares y Resultados.

En lo que sigue se considera el modelo de regresión:

$$x_i^n = h(t_i^n) + \varepsilon_i, \quad i = 1, ..., n \;\; y \;\; t_i^n \in (0,1)$$

con ε_i variables aleatorias independientes e idénticamente distribuidas tales que $E(\varepsilon_i) = 0$ y $E\left(\varepsilon_i^2\right) = \sigma^2$.

La verdadera función regresora $h \in L^2[0,1]$ es desconocida, y asumimos que pertenece a una familia

$$\mathcal{F} = \{kf(\theta, k) : \|f(\theta, k)\|_2 = 1\,, \;\; \theta \in \Theta, k \in I\}$$

donde Θ es un conjunto compacto de $\mathbb{R}^p, I$ es un intervalo compacto de $\mathbb{R}$, $\| \; \|_2$ representa la norma $L^2[0,1]$.

Siguiendo un enfoque análogo al utilizado por Beran (1977) para presentar los funcionales de distancia de Hellinger mínima, estimamos el parámetro $\theta \in \Theta$ y estimamos $k = \| h\|_2$ valiéndonos de una estimación no paramétrica $\hat{h}_n$ de la función de regresión h definida en Gasser y Müller (1979).

Para estimar θ definiremos los siguientes operadores para $g \in L^2[0,1]$

$$\begin{aligned} T(g) &= \underset{\mathbf{t}\in\boldsymbol{\Theta}}{\arg\,\text{mín}} \left\| kf(\mathbf{t},k) - g \right\|_2 \\ \hat{T}_n(g) &= \underset{\mathbf{t}\in\boldsymbol{\Theta}}{\arg\,\text{mín}} \left\| \hat{k}_n f(\mathbf{t},\hat{k}_n) - g \right\|_2 \end{aligned}$$

donde $\hat{k}_n$ es el estimador de $k = \|g\|_2$. Bajo hipótesis de regularidad de la familia $\mathcal{F}$, demostramos que para $g \in L^2[0,1], \hat{g}_n$ el estimador de g no paramétrico:

1. $T(\hat{g}_n) \to T(g)$ en probabilidad

2. $\sqrt{n}\,(T(\hat{g}_n) - T(g)) \overset{\mathcal{D}}{\to} \mathcal{N}\left(0, \sigma^2 \int_0^1 \rho_g\,(t)\,\rho_g^T\,(t)\,dt\right)$ donde $\overset{\mathcal{D}}{\to}$, significa convergencia en distribución, ρ_g es una función vectorial definida en $[0,1]$ que depende de g y ρ_g^T es su traspuesta , $\mathcal{N}$ es la distribución normal.

3. $\hat{k}_n^2 \to k^2$ en probabilidad.

4. $\sqrt{n}\left(\hat{k}_n^2 - k^2\right) \overset{\mathcal{D}}{\to} \mathcal{N}(0, 4\sigma^2 \int\limits_0^1 g^2(t)dt)$

5. $\hat{k}_n \to k$ en probabilidad.

6. $\sqrt{n}\left(\hat{k}_n - k\right) \overset{\mathcal{D}}{\to} \mathcal{N}(0,\sigma^2)$ note que la distribución no depende de $\|g\|_2$

7. $\sqrt{n}\left(\hat{T}(\hat{g}_n) - T(g)\right) \overset{\mathcal{D}}{\to} \mathcal{N}\left[0, V(\sigma^2, g)\right]$, donde

$$\begin{gathered} V(\sigma^2,g) = \sigma^2 \int\limits_0^1 \rho_g(t)\rho_g^T(t)dt + \\ \frac{\sigma^2}{\|g\|_2} \int\limits_0^1 \rho_g(t)\mathbf{h}^T(f,\theta_0,k,g)g(t)dt + \frac{\sigma^2}{\|g\|_2} \int\limits_0^1 \mathbf{h}(f,\theta_0,k,g)\rho_g^T(t)g(t)dt + \\ \sigma^2\mathbf{h}(f,\theta_0,k,g)\mathbf{h}^T(f,\theta_0,k,g)\,\|g\|_2 \end{gathered}$$

donde $\mathbf{h}(f,\theta_0,k,g)$ es un vector en $\mathbb{R}^p$ y $\mathbf{h}^T$ es su traspuesta.

Se supondrán ciertas las siguientes hipótesis e introducimos las definiciones y notaciones dadas a continuación.

Hipótesis. Definiciones. Notaciones.

Se considera el modelo de regresión:

$$x_i^n = h(t_i^n) + \varepsilon_i, \quad i = 1, ..., n \ \ y \ \ t_i^n \in (0,1)$$

con ε_i variables aleatorias independientes e idénticamente distribuidas tales que $E(\varepsilon_i) = 0$ y $E\left(\varepsilon_i^2\right) = \sigma^2$.

H1 La verdadera función regresora $h \in L^2[0,1]$ es desconocida, y asumimos que pertenece a una familia

$$\mathcal{F} = \{kf(\theta, k) : \|f(\theta, k)\|_2 = 1\ , \ \ \theta \in \Theta, k \in I\}$$

donde Θ es un conjunto compacto de $\mathbb{R}^p, I$ es un intervalo compacto de $\mathbb{R}$, $\| \ \|_2$ representa la norma dos de $L^2[0,1]$

H2 $f(\mathbf{t}, k)$ es continua en cada uno de los parámetros, $\mathbf{t} \in \Theta$, $k \in I$, bajo la norma de $L^2\,[0,1]$

H3 $\overset{\bullet}{f}(\mathbf{t}, k)$ y $\overset{\bullet\bullet}{f}(\mathbf{t}, k)$ representan la primera y segunda derivada respecto a $\mathbf{t} \in \Theta$ de f, $\frac{\partial \overset{\bullet}{f}(\mathbf{t},k)}{\partial k}$ continua en $[0,1]\,,\, k \in I$.

H4 $\mathcal{F}$ satisface la siguiente condición: si $\theta_1, \theta_2 \in \Theta\,,\ \theta_1 \neq \theta_2$,entonces $f(\theta_1, k) \neq f(\theta_2, k)$, para cada $k \in I$.

H5 $A_g(\mathbf{t}, k) = \int\limits_0^1 \overset{\bullet\bullet}{f}(x; \mathbf{t}, k) g(x) dx$ es una matriz $p \times p$ no singular, $\mathbf{t} \in Int(\Theta)$.

H6 $\rho_g\,(x) = -A_g^{-1}(\theta_0, k)\overset{\bullet}{f}\,(x; \theta_0, k)\,, x \in [0,1]\,, \rho_g \in C^2\,[0,1]\,.$

H7 $\mathcal{N}(\mu, V)$ representa la distribución Normal de media μ y varianza V.

H8 $f(\mathbf{t}, k)$ y $\overset{\bullet}{f}(\mathbf{t}, k)$ admiten el siguiente desarrollo de Taylor de primer orden, para $\mathbf{t} \in int\,(\Theta)$:

$$\begin{aligned} f(x; \mathbf{t} + \alpha e, k) &= f(x; \mathbf{t}, k) + \alpha e^T \overset{\bullet}{f}(x; \mathbf{t}, k) + \alpha e^T U(\alpha; x); x \in [0,1] \\ \overset{\bullet}{f}(x; \mathbf{t} + \alpha e, k) &= \overset{\bullet}{f}(x; \mathbf{t}, k) + \alpha \overset{\bullet\bullet}{f}\,(x; \mathbf{t}, k) e + \alpha V(\alpha; x) e; x \in [0,1] \end{aligned}$$

donde

- e es un vector columna unitario en la norma Euclidiana de $\mathbb{R}^p$
- e^T es el vector fila, traspuesto del vector e.

- $\dot{f}$ es un vector $p \times 1$ que representa las derivadas parciales de 1er orden de $f(\mathbf{t}, k)$ con respecto a las p variables del vector $\mathbf{t} \in \mathbb{R}^p$.

- $\ddot{f}(x; \mathbf{t}, k)$ es una matríz $p \times p$ que representa las derivadas parciales de segundo orden de la función $f(x; \mathbf{t}, k)$ con respecto a las p variables del vector $\mathbf{t} \in \mathbb{R}^p$.

- $U(\alpha; x)$ (vector $p \times 1$) y $V(\alpha; x)$ (matriz $p \times p$) son tales que cada una de sus componentes tienden a cero en $L^2[0,1]$ cuando $\alpha \to 0$.

Definición 1. *Denotaremos por* $\hat{g}_n(z)$, *un estimador no paramétrico de* $g \in L^2[0,1]$, *definido en Gasser y Müller (1979), como:*

$$\hat{g}_n(z) = \frac{1}{b(n)} \sum_{j=1}^{n} \int_{s_{j-1}^n}^{s_j^n} w\left(\frac{z-s}{b(n)}\right) x\left(t_j^n\right) ds \tag{1}$$

donde

- $\left\{s_j^n\right\}_{j=0,\dots,n}$ *es una sucesión creciente tal que:*

$$\begin{array}{l} s_0^n = 0,\ s_{j-1}^n \le t_j^n \le s_j^n,\ j = 1, ..., n-1, s_n^n = 1 \\ \max\limits_{1 \le j \le n} \left|s_j^n - s_{j-1}^n - \frac{1}{n}\right| = O\left(\frac{1}{n^a}\right),\ a > 1. \end{array}$$

- *w es un núcleo de orden $d = 2$, es decir, w tiene soporte compacto en* $[-c, c]$,

$$\int_{-c}^{c} w(x)dx = 1, \int_{-c}^{c} xw(x)dx = 0, \int_{-c}^{c} x^2 w(x)dx \neq 0.$$

- *w es Lipschitz de orden $\nu_w = 1$.*

- $\{b(n)\}_{n \in N}$ *satisface:* $\lim\limits_{n\to\infty} b(n) = 0$ *y* $\lim\limits_{n\to\infty} \sqrt{n}\, b(n) = \infty$.

- *g es dos veces diferenciable con segunda derivada continua en* $(0,1)$.

Gasser y Müller(1979) demuestran que este estimador posee buenas propiedades, tales como:

- *Consistencia en error medio cuadrático:*

$$\text{para todo } x \in (0,1) \lim_{n\to\infty} E\left[\hat{g}_n(x) - g(x)\right]^2 = 0$$

- *Consistencia en error medio cuadrático integrado:*

$$\lim_{n\to\infty} \int_0^1 E\left[\hat{g}_n(x) - g\left(x\right)\right]^2 dx = 0 \tag{2}$$

- *La velocidad de convergencia del error medio cuadrático integrado:*

$$\begin{gathered}
\textstyle\int_0^1 E\left[\hat{g}_n(x) - g\left(x\right)\right]^2 dx = \frac{\sigma^2}{nb(n)} \int\limits_{-c}^{c} w^2(x)dx + \\
\frac{1}{d!^2} b^{2d}(n) \left(\int\limits_{-c}^{c} w(x)x^d dx\right)^2 \cdot \int\limits_0^1 \left(g^{(d)}(x)\right)^2 dx + \\
\circ\left(\frac{1}{n^a b(n)} + \frac{1}{n^2 b^2(n)}\right) + \circ\left(\frac{b^d(n)}{n} + \circ(b^d(n))\right), a > 1,\ d = 2
\end{gathered} \tag{3}$$

Antes de enunciar los teoremas y hacer las demostraciones vamos a demostrar los siguientes lemas para obtener el comportamiento asintótico del vector aleatorio $Z_n(\rho)$, $\rho : \mathbb{R} \to \mathbb{R}^m, m \geq 1$

$$Z_n(\rho) = \frac{\sqrt{n}}{b\left(n\right)} \sum_{j=1}^{n} \varepsilon_j \left\{\int_0^1 \rho\left(t\right) \left[\int_{s_{j-1}^n}^{s_j^n} w\left(\frac{t-s}{b\left(n\right)}\right) ds\right] dt\right\} \tag{4}$$

el cual es centrado pues $\mathbb{E}\left(\varepsilon_j\right) = 0, j = 1, 2, \cdots, n$; donde, en particular, si $\rho : \mathbb{R} \to \mathbb{R}^p$ es

$$\rho\left(t\right) = \rho_g\left(t\right) = -A_g^{-1}(\theta_0, k) \overset{\bullet}{f}\left(t; \theta_0, k\right)$$

se obtiene la distribución asintótica de

$$\sqrt{n}\left(T(\hat{g}_n) - T(g)\right)$$

y también obtendremos la distribución asintótica de

$$\sqrt{n}\left(\hat{T}(\hat{g}_n) - T(g)\right) \text{ y } \sqrt{n}\left(\hat{k}_n - k\right)$$

a partir de $Z_n(\rho)$.

Lema 1. *Sea $\rho : \mathbb{R} \to \mathbb{R}^m, m \geq 1$, con segunda derivada continua en $[0, 1]$, w es un núcleo de orden $d = 2$, continuo en $\mathbb{R}$ y tiene soporte compacto en $[-c, c]$, $Z_n(\rho)$ es el vector aleatorio definida en (4) entonces*

1. $Z_n(\rho) = Y_n(\rho) + H_n(\rho)$,

   donde

$$\begin{gathered} Y_n(\rho) = \sqrt{n}\sum_{j=1}^{n} \varepsilon_j \, \rho(\xi_j^n) \left[s_j^n - s_{j-1}^n\right] \\ H_n(\rho) = \sqrt{n} b\,(n) \sum_{j=1}^{n} \varepsilon_j \left[s_j^n - s_{j-1}^n\right] \left\{\int_{-c}^{c} \rho'(\eta_j^n)\, z\ w\,(z)\ dz\right\} \end{gathered}$$

con $\xi_j^n \in \left[s_{j-1}^n, s_j^n\right]$, $\eta_j^n \in \left[\xi_j^n, zb\,(n)\right]$ *ó* $\eta_j^n \in \left[zb\,(n)\,, \xi_j^n\right]$ *para* $n \geq N_0$, *y* $H_n(\rho) \to 0$ *en probabilidad.*

2. $Var(Y_n(\rho)) \underset{n\to\infty}{\to} \sigma^2 \int\limits_0^1 \rho(t)\rho^T(t)dt$, $Var(Y_n(\rho))$ *es la matriz de varianzas y covarianzas de* $Y_n(\rho)$. :

Demostración: Usando Fubini, cambio de variable, el hecho que w es a soporte compacto y teorema de valor medio para integrales, obtenemos que para $n \geq N_0$

$$Z_n(\rho) = \sqrt{n}\sum_{j=1}^{n} \varepsilon_j \left\{\int_{-c}^{c} \rho\left(zb(n) + \xi_j^n\right)\ w\,(z) \left[s_j^n - s_{j-1}^n\right] dz\right\}$$

con $\xi_j^n \in \left[s_{j-1}^n, s_j^n\right]$, mediante el desarrollo de Taylor de ρ alrededor de ξ_j^n, existe $\eta_j^n \in \left[\xi_j^n, zb\,(n)\right]$ ó $\eta_j^n \in \left[zb\,(n)\,, \xi_j^n\right]$ tal que

$$Z_n(\rho) = \sqrt{n}\sum_{j=1}^{n} \varepsilon_j \left\{\int_{-c}^{c} \left[\rho(\xi_j^n) + \rho'(\eta_j^n)\, z\, b(n)\right]\ w\,(z) \left[s_j^n - s_{j-1}^n\right] dz\right\}$$

por ser w un núcleo obtenemos la descomposición

$$\begin{gathered} Z_n(\rho) = \sqrt{n}\sum_{j=1}^{n} \varepsilon_j \, \rho(\xi_j^n) \left[s_j^n - s_{j-1}^n\right] + \\ \sqrt{n} b\,(n) \sum_{j=1}^{n} \varepsilon_j \left[s_j^n - s_{j-1}^n\right] \left\{\int_{-c}^{c} \rho'(\eta_j^n)\, z\ w\,(z)\ dz\right\} \end{gathered}$$

para demostrar que $H_n(\rho)$ tiende a cero en probabilidad hacemos el desarrollo de Taylor de $\rho'(\eta_j^n)$ alrededor de s_j^n y puesto que $\int\limits_0^1 zw(z)dz = 0$ obtenemos que existe $\nu_j^n \in \left[\eta_j^n, s_j^n\right]$, tal que

$$H_n(\rho) = b(n)\sqrt{n}\sum_{j=1}^{n} \varepsilon_j \left(s_j^n - s_{j-1}^n\right) \int_{-c}^{c} \rho''(\nu_j^n)(s_j^n - \eta_j^n)\, w\,(z)\, z\, dz$$

$$\begin{gathered} ||H_n(\rho)|| \leq b\,(n)\,\sqrt{n} \left|\left|\int_{-c}^{c}\ \rho''(\nu_j^n) w\,(z)\, z\, dz\right|\right| \sum_{j=1}^{n} |\varepsilon_j| \left[s_j^n - s_{j-1}^n\right]^2 \\ ||H_n(\rho)||^2 \leq n^2 \cdot b^2(n) K_{\rho,w} \left(\frac{1}{n}\sum_{j=1}^{n} \varepsilon_j^2\right) \sum_{j=1}^{n} \left[s_j^n - s_{j-1}^n\right]^4 \to 0 \end{gathered}$$

casi siempre, pues por la Ley fuerte de los grandes números,

$$\begin{aligned} \frac{1}{n}\sum_{j=1}^{n}\varepsilon_j^2 &\rightarrow \sigma^2 \text{ casi siempre,} \sum_{j=1}^{n}\left[s_j^n - s_{j-1}^n\right]^4 \leq \frac{1}{n^3} \\ b(n) &\rightarrow 0 \text{ y } \left(\int_{-c}^{c}\left\|\rho''(\nu_j^n)w\left(z\right)z\right\| dz\right)^2 = K_{\rho,w} < \infty \end{aligned}$$

por lo que:

$$\mathbb{P}\left(||H_n(\rho)|| \geq \varepsilon\right) \rightarrow 0 \text{ si } n \rightarrow \infty.$$

donde $\|H_n(\rho)\|$ es la norma euclídea del vector aleatorio $H_n(\rho)$. De este hecho se desprende que $Y_n(\rho)$ y $Z_n(\rho)$ son asintóticamente equivalentes, es decir, tienen la misma distribución asintótica.

$Y_n(\rho)$ es centrado, $Var((Y_n(\rho))) = \mathbb{E}\left((Y_n(\rho))(Y_n^T(\rho))\right) = \mathbb{E}\left(Y_n^{ih}(\rho)\right)_{m\times m}$ donde

$$Y_n^{ih}(\rho) = \left[\sqrt{n}\sum_{j=1}^{n}\varepsilon_j\,\rho^i(\xi_j^n)\left[s_j^n - s_{j-1}^n\right]\right]\left[\sqrt{n}\sum_{k=1}^{n}\varepsilon_k\,\rho^h(\xi_k^n)\left[s_k^n - s_{k-1}^n\right]\right]$$

ρ^i es la i-ésima componente del vector ρ , $Y_n^T(\rho)$ es el vector traspuesto del vector $Y_n(\rho)$.

$$\mathbb{E}\left(Y_n^{ih}(\rho)\right) = n\sum_{j=1}^{n}\sigma^2\,\rho^i(\xi_j^n)\rho^h(\xi_j^n)\left[s_j^n - s_{j-1}^n\right]^2$$

pues $\mathbb{E}\left(\varepsilon_j^2\right) = \sigma^2$ y $\mathbb{E}\left(\varepsilon_j\varepsilon_k\right) = 0, j \neq k$,como

$$n\sum_{j=1}^{n}\sigma^2\,\rho^i(\xi_j^n)\rho^h(\xi_j^n)\left[s_j^n - s_{j-1}^n\right]^2 \cong \left\{n\sigma^2\left[\frac{K}{n^a} + \frac{1}{n}\right]\sum_{j=1}^{n}\rho^i(\xi_j^n)\rho^h(\xi_j^n)\left[s_j^n - s_{j-1}^n\right]\right\}$$

donde K es constante, usando la continuidad de ρ y considerando que $a > 1$, tenemos que

$$\mathbb{E}\left(Y_n^{ih}(\rho)\right) \underset{n\rightarrow\infty}{\rightarrow} \sigma^2\int_0^1 \rho^i\left(t\right)\rho^h\left(t\right)\, dt$$

□

Lema 2. *Bajos las hipótesis del Lema 1*

$$Z_n(\rho) \rightarrow N\left[0, \sigma_2\int_0^1 \rho(t)\rho^T(t)dt\right]$$

Demostración: $Z_n(\rho)$ es asintóticamente equivalente a $Y_n(\rho)$, pues $H_n(\rho) \to 0$ en probabilidad; si demostramos que

$$Y_n(\rho) \to N\left[0, \sigma_2 \int_0^1 \rho(t)\rho^T(t)dt\right]$$

obtendremos el resultado para $Z_n(\rho)$.

Puesto que

$$Y_n(\rho) = \sqrt{n}\sum_{j=1}^{n} \varepsilon_j\, \rho(\xi_j^n)\left[s_j^n - s_{j-1}^n\right]$$

para obtener la normalidad asintótica de $Y_n(\rho)$, demostremos que para todo vector $y \in \mathbb{R}^m$ no nulo

$$\langle Y_n(\rho), y\rangle \to N\left[0, \sigma^2\,(y)\right]$$

por las propiedades del producto escalar

$$\langle Y_n(\rho), y\rangle = \sum_{j=1}^{n} \varepsilon_j\, a_{nj}(y)$$

donde

$$a_{nj}(y) = \sqrt{n}\,\left\langle \rho(\xi_j^n), y\right\rangle\left[s_j^n - s_{j-1}^n\right]; \xi_j^n \in \left(s_j^n - s_{j-1}^n\right)$$

Si

$$X_{nj}(y) = \varepsilon_j\, a_{nj}(y) \text{ y } S_n(y) = Var(\langle Y_n(\rho), y\rangle) = \sigma^2 \sum_{j=1}^{n} a_{nj}^2(y)$$

$$\langle Y_n(\rho), y\rangle = \sum_{j=1}^{n} X_{nj}(y)$$

probaremos a continuación que $\langle Y_n(\rho), y\rangle$ satisface la condición de Lindeberg (Billingsley, 1968), esto es :

$$\frac{1}{S_n(y)}\sum_{j=1}^{n}\int_{\left\{X_{nj}(y)\geq\varepsilon\sqrt{S_n(y)}\right\}} X_{nj}^2(y)d\mathbb{P} \to 0 \text{ , para cada } \varepsilon > 0$$

Sea $\varepsilon > 0$, como$\int \varepsilon_j^2 d\mathbb{P} < \infty$,

$$\text{existe } R > 0 \text{ tal que } \int_{\left\{\varepsilon_j^2\geq\varepsilon\sigma^2 R\right\}} \varepsilon_j^2\, d\mathbb{P} \leq \sigma^2\varepsilon$$

$$\text{existe } N_1 \text{ tal que para cada } n \geq N_1 \text{ se tiene } \frac{\sum\limits_{k=1}^{n} a_{nk}^2(y)}{a_{nj}^2(y)} \geq R$$

pues

$$|a_{nj}(y)| \leq (\|\rho\|_\infty \|y\|) K \left(\frac{\sqrt{n}}{n^a} + \frac{\sqrt{n}}{n} \right) \underset{n\to\infty}{\longrightarrow} 0, a > 1 \text{ donde}$$

$$\|\rho\|_\infty = \sup_{t\in[0,1]} \|\rho(t)\|, \|y\| \text{ es la norma Euclídea en } \mathbb{R}^m, K \text{ constante.}$$

Por otro lado:
$\sum\limits_{j=1}^{n} a_{nj}^2(y) \leq K (\|\rho\|_\infty \|y\|)^2 \left[\frac{n^2}{n^{2a}} + \frac{2n^2}{n^{a+1}} + \frac{n^2}{n^2} \right] \underset{n\to\infty}{\longrightarrow} K (\|\rho\|_\infty \|y\|)^2$ no nulo. Así tenemos que:

$\frac{a_{nj}^2(y)}{\sum\limits_{k=1}^{n} a_{kj}^2(y)} \underset{n\to\infty}{\longrightarrow} 0^+ \Rightarrow \frac{\sum\limits_{k=1}^{n} a_{nk}^2(y)}{a_{nj}^2(y)} \underset{n\to\infty}{\longrightarrow} \infty$ luego, para cada $n \geq N_1$

$$A_{n,j} = \left\{ \varepsilon_j^2 \geq \varepsilon^2 \sigma^2 \frac{\sum\limits_{j=1}^{n} a_{nj}^2(y)}{a_{nj}^2(y)} \right\} \subset \left\{ \varepsilon_j^2 \geq \varepsilon^2 \sigma^2 R \right\}$$

concluyendo que para $n \geq N_1$

$$\begin{aligned} \frac{1}{S_n(y)} \sum_{r=1}^{n} \int_{\left\{X_{nr}(y) \geq \varepsilon\sqrt{S_n(y)}\right\}} X_{nr}^2(y) d\mathbb{P} &\leq \\ \frac{1}{\sigma^2 \sum\limits_{j=1}^{n} a_{nj}^2(y)} \sum_{r=1}^{n} a_{nr}^2(y) \int_{\{\varepsilon_r^2 \geq \varepsilon^2\sigma^2 R\}} \varepsilon_r^2 \, d\mathbb{P} &\leq \frac{1}{\sigma^2 \sum\limits_{j=1}^{n} a_{nj}^2(y)} \sum_{r=1}^{n} a_{nr}^2(y) \varepsilon \sigma^2 = \varepsilon. \end{aligned}$$

□

Corolario 1. *Si* $\rho_g(t) = -A_g^{-1}(\theta_0, k) f(t; \theta_0, k) \in C^2[0,1]$, *donde* $A_g^{-1}(\theta_0, k)$ *está definido en (H6),* $g \in C^2[0,1]$, w *es un núcleo de orden 2 y* $\rho = \rho_g$, *entonces*

$$Z_n(\rho_g) \to N\left[0, \sigma_2 \int_0^1 \rho_g(t) \rho_g^T(t) dt\right]$$

Corolario 2. *Si* $\rho = g \in C^2[0,1]$ *y* w *es un núcleo de orden 2, entonces*

$$Z_n(g) \to N\left[0, \sigma_2 \int_0^1 g^2(t) dt\right]$$

Lema 3. *Si ρ y $g \in C^2[0,1]$ y w es un núcleo de orden 2, entonces $Z_n(\rho)$ es asintóticamente equivalente a:*

$$\sqrt{n}\int_0^1 \rho(t)\left[\hat{g}_n(t) - g(t)\right] dt$$

(es decir, sus distribuciones límites son las mismas)

Demostración: Sea $I_1(n) = \sqrt{n}\int_0^1 \rho(t)\left[\hat{g}_n(t) - g(t)\right] dt$, usando la definición de $\hat{g}_n$

$$I_1(n) = Z_n(\rho) + \sqrt{n}\int_0^1 \rho(t)\left[\left\{\frac{1}{b(n)}\sum_{j=1}^{n}\int_{s_{j-1}^n}^{s_j^n} w\left(\frac{t-s}{b(n)}\right) g\left(t_j^n\right) ds\right\} - g(t)\right] dt$$

para $n \geq N_0$, usando Fubini, cambio de variable y que w es un núcleo

$$\sqrt{n}\int_0^1 \rho(t)\left[\left\{\frac{1}{b(n)}\sum_{j=1}^{n}\int_{s_{j-1}^n}^{s_j^n} w\left(\frac{t-s}{b(n)}\right) g\left(t_j^n\right) ds\right\} - g(t)\right] dt = I_{11} + I_{12}$$

donde

$$I_{11} = \sqrt{n}\sum_{j=1}^{n}\int_{s_{j-1}^n}^{s_j^n}\int_{-c}^{c} w(z)\left[\rho(b(n)z + s) - \rho(s)\right] g\left(t_j^n\right) dz\, ds$$
$$I_{12} = \sqrt{n}\sum_{j=1}^{n}\int_{s_{j-1}^n}^{s_j^n} \rho(s)\left[g\left(t_j^n\right) - g(s)\right] ds$$

Si en I_{11}, aplicamos el Teorema del valor medio, existe $\lambda_{jn} \in \left[s_{j-1}^n, s_j^n\right]$ tal que:

$$\begin{aligned} I_{11} &= \sqrt{n}\sum_{j=1}^{n}\int_{-c}^{c} w(z)\left[\rho(b(n)z + \lambda_{jn}) - \rho(\lambda_{jn})\right]\left[s_j^n - s_{j-1}^n\right] g\left(t_j^n\right) dz \\ &= \sqrt{n}\sum_{j=1}^{n}\int_{-c}^{c} \rho'\left(\xi_{z,s_j}\right) b(n)\, w(z)\, z\left[s_j^n - s_{j-1}^n\right] g\left(t_j^n\right) dz \end{aligned}$$

donde $\xi_{z,s_j} \in I_*$ con I_* intervalo de extremos λ_{jn} , $\lambda_{jn} + b(n)z$ o viceversa. Observe que para n suficientemente grande, $I_* \subset \left[s_{j-1}^n, s_j^n\right]$, aplicando el desarrollo de Taylor a la función ρ' alrededor de s_j^n obtenemos:

$$I_{11} = \sqrt{n}\sum_{j=1}^{n} b(n)\left[s_j^n - s_{j-1}^n\right] g\left(t_j^n\right)\rho'\left(s_j^n\right)\int_{-c}^{c} w(z)\, z\, dz +$$
$$\sqrt{n}\sum_{j=1}^{n} b(n)\left[s_j^n - s_{j-1}^n\right] g\left(t_j^n\right)\int_{-c}^{c}\rho''\left(\nu_j^n\right)\left(s_j^n - \xi_{z,s_j}\right) w(z)\, z\, dz$$

donde $\nu_j^n \in \left(\xi_{z,s_j}, s_j^n\right)$, ya que $\int_{-c}^{c} w(z)\, z\, dz = 0$,

$$
\begin{aligned}
\|I_{11}\| &\leq \sqrt{n} b(n) \sum_{j=1}^{n} \left(s_j^n - s_{j-1}^n\right) \left|g\left(t_j^n\right)\right| \frac{1}{n} \int_{-c}^{c} \left\|\rho''\left(\nu_j^n\right)\right\| |w(z)\, z|\, dz \\
&\leq \frac{1}{\sqrt{n}} b(n) \sum_{j=1}^{n} \left(s_j^n - s_{j-1}^n\right) \left|g\left(t_j^n\right)\right| \int_{-c}^{c} \left\|\rho''\left(\nu_j^n\right)\right\| |w(z)\, z|\, dz \to 0
\end{aligned}
$$

si $n \to \infty$, pues $\frac{1}{\sqrt{n}} b(n) \to 0$,

$$
\sum_{j=1}^{n} \left(s_j^n - s_{j-1}^n\right) \left|g\left(t_j^n\right)\right| \to \int_0^1 |g(t)|\, dt, \text{ y } \int_{-c}^{c} \left\|\rho''\left(\nu_j^n\right)\right\| |w(z)\, z|\, dz < \infty.
$$

Además, por el teorema del valor medio para integrales y por el desarrollo de Taylor, existen $\varpi_{jn} \in \left(s_j^n, s_{j-1}^n\right)$ y α_{jn} entre t_j^n y ϖ_{jn} tales que la expresión I_{12} satisface:

$$
\begin{aligned}
\|I_{12}\| &\leq \sqrt{n} \sum_{j=1}^{n} \left\| \int_{s_{j-1}^n}^{s_j^n} \rho(s) \left[g(t_j^n) - g(s)\right] ds \right\| \\
&= \sqrt{n} \sum_{j=1}^{n} \left\| \rho(\varpi_{jn}) \left[g(t_j^n) - g(\varpi_{jn})\right] \left[s_j^n - s_{j-1}^n\right] \right\| \\
&\leq \sqrt{n} \|\rho\|_\infty \sum_{j=1}^{n} \left| g'(\alpha_{jn}) \left[t_j^n - \varpi_{jn}\right] \left[s_j^n - s_{j-1}^n\right] \right| \\
&\leq \sqrt{n} \|\rho\|_\infty \|g'\|_\infty \sum_{j=1}^{n} \left[t_j^n - \varpi_{jn}\right] \left[s_j^n - s_{j-1}^n\right] \\
&\leq \sqrt{n} \|\rho\|_\infty \|g'\|_\infty \sum_{j=1}^{n} \left[s_j^n - s_{j-1}^n\right]^2 \\
&\leq K \left(\frac{1}{n^{a-\frac{1}{2}}} + \frac{1}{n} \right) \|\rho\|_\infty \|g'\|_\infty \underset{n\to\infty}{\rightarrow} 0, K \text{ es constante y } a > 1
\end{aligned}
$$

□

Distribución Límite de $T(\overset{\wedge}{g_n})$

El enfoque que seguiremos será análogo al utilizado por Beran (1977) para presentar los funcionales de distancia de Hellinger mínima.

Es decir, $T(g)$ es un elemento de Θ que minimiza la distancia en norma dos de las funciones $kf(\theta, k)$ pertenecientes al conjunto $\mathcal{F}$ y $g \in L^2[0,1]$. Aunque

puede que existan varios $\theta \in \Theta$ donde se alcance el mínimo, tomaremos uno cualquiera de éstos.
Bajo las hipótesis generales impuestas a la familia de funciones $\mathcal{F}$, se puede garantizar que el funcional T además de estar bien definido es continuo en la métrica de $L^2\left[0,1\right]$, como se verá en los siguientes teoremas:

Teorema 1. *Bajo la hipótesis (H2), existe $T\left(g\right)$ para todo $g \in L^2\left[0,1\right]$.*

Demostración: Sean $p(\mathbf{t},k) = \|kf(\mathbf{t},k) - g\|_2$ y $\{\mathbf{t}_n\} \subset \Theta$, tal que $\mathbf{t}_n \to \mathbf{t}$ de la desigualdad

$$|p(\mathbf{t}_n,k) - p\left(\mathbf{t},k\right)| \leq |k| \, \| \, f(\mathbf{t}_n,k) - f(\mathbf{t},k)\|_2 \to 0 \text{ cuando } n \to \infty$$

concluimos que $p(\mathbf{t},k)$ es continua en $\mathbf{t}$, para k fijo y para cada $\mathbf{t} \in \Theta$, como este conjunto es compacto en $\mathbb{R}^p$, entonces tenemos que existe θ_0 tal que $p\left(\theta_0,k\right)$ es mínimo, con lo cual se garantiza la existencia de $T(g)$ $\square$

Teorema 2. *Supongamos que $T(g)$ es única y que se cumple la hipótesis (H4), entonces*

1. *T es continua*
2. *$T(kf(\theta,k)\ ) = \theta$ únicamente para $\theta \in \Theta$, k fijo.*

Demostración: 1) Sean g_n y $g \in L^2\left[0,1\right]$ tales que $\|g_n - g\|_2 \to 0$, definimos $p_n(\mathbf{t},k) = \|kf(\mathbf{t},k) - g_n\|_2$ y $p\left(\mathbf{t},k\right) = \|kf(\mathbf{t},k) - g\|_2$ con k fijo,

$$\left|\underset{\mathbf{t}\in\Theta}{Min} p_n\left(\mathbf{t},k\right) - \underset{\mathbf{t}\in\Theta}{Min} p\left(\mathbf{t},k\right)\right| \leq \underset{\mathbf{t}\in\Theta}{Sup} |p_n(\mathbf{t},k) - p\left(\mathbf{t},k\right)| \leq \underset{\mathbf{t}\in\Theta}{Sup} \|g_n - g\|_2 \to 0$$

por lo tanto $p_n\left(\theta_n,k\right) \to p\left(\theta_0,k\right)$ donde

$$\theta_n = T\left(g_n\right) = \underset{\mathbf{t}\in\Theta}{\arg\text{mín}} \|kf(\mathbf{t},k) - g_n\|_2$$
$$\theta_0 = T\left(g\right) = \underset{\mathbf{t}\in\Theta}{\arg\text{mín}} \|kf(\mathbf{t},k) - g\|_2$$

Falta demostrar que la sucesión $\{\theta_n\}$ converge a θ_0. Supongamos que θ_n no converge a θ_0, como Θ es compacto , existe una subsucesión $\left\{\theta_{n_j}\right\} \subset \{\theta_n\}$ tal que $\theta_{n_j} \to \theta_1$ con θ_0 distinto a θ_1, luego por ser $p(\mathbf{t},k)$ continua en $\mathbf{t}$, $p\left(\theta_{n_j},k\right) \to p\left(\theta_1,k\right)$ y por la unicidad del límite tenemos que $p\left(\theta_1,k\right) = p\left(\theta_0,k\right) = \lim\limits_{n\to\infty} p(\theta_n,k)$, lo cual es una contradicción, ya que hemos supuesto que $T\left(g\right)$ es único (el mínimo se alcanza en un único θ_0).
2) $T\left(kf(\theta_0,k)\right) = \theta_0$ es inmediato por la hipótesis de identificación sobre los parámetros. $\square$

Teorema 3. *Bajo las siguientes hipótesis:*

- $f(\mathbf{t},k)$ *satisface las hipótesis (H2,H3,H8)*
- *El parámetro* k *es fijo (conocido)*
- $T(g)$ *existe, es única y pertenece al* $Int(\Theta)$.
- T *es continua en la métrica de* $\mathbb{L}^2\left[0,1\right]$.
- $\int_0^1 \overset{\bullet\bullet}{f}(x;\mathbf{t},k)g(x)dx$ *satisface la hipótesis (H5).*
- g_n , $g \in L^2\left[0,1\right]$,$\left\|g_n - g\right\|_2 \to 0$

Entonces

$$\theta_n - \theta_0 =$$
$$-\left[\int_0^1 \overset{\bullet\bullet}{f}\left(x;\theta_0,k\right)g\left(x\right)\,dx\right]^{-1}\int_0^1 \overset{\bullet}{f}\left(x;\theta_0,k\right)\left[g_n\left(x\right)-g\left(x\right)\right]\,dx$$
$$+a_n\int_0^1 \overset{\bullet}{f}\left(x;\theta_0,k\right)\left[g_n\left(x\right)-g\left(x\right)\right]\,dx$$

donde a_n *es una matriz* $p\times p$ *tal que sus componentes tienden a cero cuando* $n\to\infty$.

Demostración: Por definición:

$$\begin{aligned}\left\|kf(\mathbf{t},k)-g\right\|_2^2 &= \int_0^1\left[k^2f^2(x;\mathbf{t},k)-2kf(x;\mathbf{t},k)g(x)+g^2(x)\right]\,dx\\ &= k^2-2k\int_0^1 f\left(x;\mathbf{t},k\right)g\left(x\right)\,dx+\int_0^1 g^2\left(x\right)\,dx\end{aligned}$$

como para g fija y k fija, las expresiones

$$\left\|kf(\mathbf{t},k)-g\right\|_2^2 \text{ y } -2k\int_0^1 f\left(x;\mathbf{t},k\right)g\left(x\right)\,dx$$

alcanzan el mínimo en el mismo punto $\mathbf{t}$, entonces trabajaremos con la segunda expresión.
Definimos $H\left(k,\mathbf{t}\right)=-2k\int_0^1 f\left(x;\mathbf{t},k\right)g\left(x\right)\,dx$, luego usando (H8)

$$\lim_{\alpha\to 0}\frac{H\left(k,\mathbf{t}+\alpha e\right)-H\left(k,\mathbf{t}\right)}{\alpha}=-2k\int_0^1 e^T\overset{\bullet}{f}\left(x;\mathbf{t},k\right)g\left(x\right)\,dx$$

para cualquiere vector e y $\mathbf{t} \in int(\Theta)$, como$\| kf(\mathbf{t},k) - g\|_2^2$ alcanza un mínimo en $\mathbf{t} = \theta_0$ obtenemos:

$$\int_0^1 \overset{\bullet}{f}(x;\theta_0,k)\, g(x)\ dx = \mathbf{0}$$

de igual manera,

$$\int_0^1 \overset{\bullet}{f}(x;\theta_n,k)\, g_n(x)\ dx = \mathbf{0}$$

de esta última expresión y (H8):

$$\begin{gathered}\textstyle\int_0^1 \overset{\bullet}{f}(x;\theta_0,k)\, g_n(x)\ dx + \alpha \int_0^1 \overset{\bullet\bullet}{f}(x;\theta_0,k)\ e\, g_n(x)\ dx + \\ \textstyle\alpha \int_0^1 V(\alpha;x)\, e\, g_n(x)\ dx = \mathbf{0}\end{gathered}$$

si $\alpha e = \theta_n - \theta_0$ y n es suficientemente grande

$$\begin{aligned}(\theta_n - \theta_0) \quad = & \\ & -\left\{\int_0^1 \overset{\bullet\bullet}{f}(x;\theta_0,k)\ \ g(x)\ dx\right\}^{-1} \int_0^1 \overset{\bullet}{f}(x;\theta_0,k)\ [\,g_n(x) - g(x)]\ dx \\ & + a_n \int_0^1 \overset{\bullet}{f}(x;\theta_0,k)\ [\,g_n(x) - g(x)]\ dx\end{aligned}$$

donde:

$$\begin{aligned}& a_n = -\left\{\textstyle\int_0^1 \overset{\bullet\bullet}{f}(x;\theta_0,k)\ \ g(x)\ dx\right\}^{-1} A_n \left\{\textstyle\int_0^1 \overset{\bullet\bullet}{f}(x;\theta_0,k)\ \ g(x)\ dx\right\}^{-1} + \\ & R_n \left\{\textstyle\int_0^1 \overset{\bullet\bullet}{f}(x;\theta_0,k)\ \ g(x)\ dx\right\}^{-1}, \\ & \text{con } A_n = \left[\textstyle\int_0^1 \overset{\bullet\bullet}{f}(x;\theta_0,k)\ \ [g_n(x) - g(x)]\ dx + \int_0^1 V(\alpha;x)\ g_n(x)\ dx\right]\end{aligned}$$

Ahora bien, el primer sumando en a_n y R_n dependen de :

$$\int_0^1 \overset{\bullet\bullet}{f}(x;\theta_0,k)\ \ [g_n(x) - g(x)]\ dx + \int_0^1 V(\alpha;x)\ g_n(x)\ dx$$

es por ello que estudiaremos a continuación el comportamiento de cada sumando en la última expresión.

Usando Cauchy-Schwart y considerando que $(g_n - g) \stackrel{L^2}{\to} 0$ tenemos que el primer sumando tiende a cero y por otro lado, por hipótesis cada componente de $V(\alpha;\cdot) \stackrel{L^2}{\to} 0$, de aquí el segundo sumando también converge a cero, por lo que $a_n \to 0$ y $R_n \to 0$. $\square$

Teorema 4. *Sea $\hat{g}_n$ definida en (1),entonces:*

1) $\hat{g}_n \stackrel{\mathbb{L}^2}{\to} g$ (en probabilidad)

2) $T(\hat{g}_n) \to T(g)$ (en probabilidad).

Demostración: 1) Sea

$$\|\hat{g}_n - g\|_2^2 = \int_0^1 [\hat{g}_n(t) - g(t)]^2 \, dt$$

la desigualdad de Chebichev, Fubbini y (2) aseguran que

$$\begin{aligned} \mathbb{P}\left[\|\hat{g}_n - g\|_2 \geq \varepsilon\right] &\leq \frac{1}{\varepsilon^2} E\left[\|\hat{g}_n - g\|_2^2\right] = \\ \frac{1}{\varepsilon^2} E\left[\int_0^1 (\hat{g}_n(t) - g(t))^2 \, dt\right] &= \frac{1}{\varepsilon^2}\int_0^1 E\left[(\hat{g}_n(t) - g(t))^2\right] dt \to 0 \end{aligned}$$

2) por el resultado anterior y por ser T continua $T(\hat{g}_n) \to T(g)$ en probabilidad. $\square$

Teorema 5. *Si ρ_g , $g \in C^2[0,1]$ y w es un núcleo de orden 2 entonces*

$$\sqrt{n}\left[T(\hat{g}_n) - T(g)\right] \stackrel{\mathcal{D}}{\to} \mathcal{N}\left[0, \sigma^2 \int_0^1 \rho_g(t)\,\rho_g^T(t)\,dt\right]$$

Demostración:

$$\begin{gathered} \sqrt{n}\left[T(\hat{g}_n) - T(g)\right] = \sqrt{n}\textstyle\int_0^1 \rho_g(t)\left[\hat{g}_n(t) - g(t)\right] dt + \\ \sqrt{n}a_n \textstyle\int_0^1 \dot{f}(t;\theta_0,k)\left[\hat{g}_n(t) - g(t)\right] dt = I_1(n) + I_2(n) \end{gathered}$$

por el Teorema 3

El primer sumando, que denotamos por $I_1(n)$ es asintóticamente equivalente a $Z_n(\rho_g)$ por el Lema 3, donde $Z_n(\rho_g)$ converge a

$$\mathcal{N}\left[0, \sigma^2 \int_0^1 \rho_g(t)\,\rho_g^T(t)\,dt\right]$$

Resta ver que el segundo sumando $I_2(n)$ converge a cero en probabilidad:

$$I_2(n) = \sqrt{n} a_n \int_0^1 \overset{\bullet}{f}(t; \theta_0, k) \left[\hat{g}_n(t) - g(t)\right] dt = a_n W_n$$

donde W_n por el Lema 3, es asintóticamente equivalente a $Z_n(\overset{\bullet}{f}(\theta_0, k))$ que converge en distribución a

$$\mathcal{N}\left[0, \sigma^2 \int_0^1 \rho(t)\, \rho^T(t)\, dt\right]$$

$\rho = \overset{\bullet}{f}(\theta_0, k)\,, \theta_0 = T(g)$ y a_n converge a cero, por lo que $I_n(2)$ converge a cero en probabilidad. □

Estimación de $\|g\|_2$

Sea $\hat{g}_n$ el estimador de la función g dado por Gasser y Muller, definido en (1)

Usaremos el siguiente estimador de la norma de g :

$$\hat{k}_n^2 = \int_0^1 \hat{g}_n^2(t)dt$$

y sea

$$k^2 = \int_0^1 g^2(t)dt$$

Los siguientes teoremas nos aseguran la consistencia de este estimador y la distribución límite.

Teorema 6. *Bajo las hipótesis consideradas en la Definición 1*

- $\hat{k}_n^2 \to k^2$ *en probabilidad.*
- $\sqrt{n}\left(\hat{k}_n^2 - k^2\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, 4\sigma^2 \int_0^1 g^2(t)dt)$.

Demostración:

$$\begin{aligned}
\left|\hat{k}_n^2 - k^2\right| &= \left|\int_0^1 \hat{g}_n^2(t)dt - \int_0^1 g^2(t)dt\right| = \\
&\left|\int_0^1 (\hat{g}_n(t) - g(t))^2 dt + 2\int_0^1 g(t)(\hat{g}_n(t) - g(t))dt\right| \leq \\
&\int_0^1 \left|\hat{g}_n(t) - g(t)\right|^2 dt + 2\int_0^1 \left|g(t)\right| \left|\hat{g}_n(t) - g(t)\right| dt
\end{aligned}$$

Usando el Teorema de Fubini,

$$\mathbb{E}\left(\int_0^1 \left|\hat{g}_n(t)-g(t)\right|^2 dt\right)=\int_0^1 \mathbb{E}(\left|\hat{g}_n(t)-g(t)\right|^2)\to 0 \text{ por } (2)$$

de aquí se deduce la convergencia en probabilidad a cero de

$$\int_0^1 \left|\hat{g}_n(t)-g(t)\right|^2 dt$$

Por otro lado,

$$\int_0^1 |g(t)|\,|\hat{g}_n(t)-g(t)|\,dt \le \|g\|_2 \left[\int_0^1 \left|\hat{g}_n(t)-g(t)\right|^2\right]^{\frac{1}{2}} \to 0$$

en probabilidad, se concluye que

$$\mathbb{P}\left(\left|\hat{k}_n^2-k^2\right|\ge\varepsilon\right)\to 0$$

con lo cual queda demostrada la consistencia del estimador.

Para obtener la convergencia débil del estadístico, lo expresamos como:

$$\sqrt{n}\left(\hat{k}_n^2-k^2\right)=\sqrt{n}\left(\int_0^1 (\hat{g}_n(t)-g(t))^2dt+2\int_0^1 g(t)(\hat{g}_n(t)-g(t))dt\right)$$

sea

$$I_1(n)=2\sqrt{n}\int_0^1 g(t)(\hat{g}_n(t)-g(t))dt$$
$$I_2(n)=\sqrt{n}\int_0^1 (\hat{g}_n(t)-g(t))^2dt$$

El término

$$I_1(n)=2Z_n(g)+2\sqrt{n}\int_0^1 g(t)\left(\frac{1}{b(n)}\sum_{j=1}^n w(\frac{t-s}{b(n)})g(t_j^n)ds-g(t)\right)dt$$

usando la definición de $\hat{g}_n$, donde $Z_n(g)$ está definido en (4), con $\rho(t)=g(t)$, en este caso, del Lema 2, se obtiene que

$$Z_n(g)\xrightarrow{\mathcal{D}}\mathcal{N}(0,\sigma^2\int_0^1 g^2(t)dt),$$

y de manera análoga a la demostración del Lema 2 se obtiene que el otro sumando que define $I_1(n)$ tiende a cero si $n \to \infty$.

Falta estudiar únicamente el término

$$I_2(n) = \sqrt{n} \int_0^1 (\hat{g}_n(t) - g(t))^2 dt$$

Por (3):

$$\int_0^1 E\left[\hat{g}_n - g\left(t\right)\right]^2 dt = \frac{\sigma^2}{nb(n)} \int_{-c}^{c} w^2(x)dx +$$
$$\frac{1}{4} b(n)^4 \left(\int_{-c}^{c} w(x)x^2 dx \right)^2 \cdot \int_0^1 \left(g^{(2)}(t)\right)^2 dt +$$
$$o\left(\frac{1}{n^a b(n)} + \frac{1}{n^2 b(n)^2}\right) + \circ\left(\frac{b^2(n)}{n} + \circ(b(n)^2)\right), a > 1$$

de aquí se obtiene que

$$\sqrt{n} \int_0^1 E\left[\hat{g}_n - g\left(t\right)\right]^2 \to 0.$$

El $b(n)$ optimal (Gasser y Müller,1979) es un $o((1/n)^{\frac{1}{2d+1}})$, con $d = 2$ en nuestro caso. □

Corolario 3. *Bajo las hipótesis consideradas en la Definición 1 se obtiene que:*

- $\hat{k}_n \to k$ *en probabilidad*
- $\sqrt{n}\left(\hat{k}_n - k\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2)$ *, el límite no depende de* $\|g\|_2$ *.*

Demostración: $\hat{k}_n - k = \dfrac{\hat{k}_n^2 - k^2}{\hat{k}_n + k} \to 0$ en probabilidad y

$$\sqrt{n}\left(\hat{k}_n - k\right) = \sqrt{n}\frac{\hat{k}_n^2 - k^2}{\hat{k}_n + k} \to N(0, \frac{4\sigma^2 \int_0^1 g^2(t)dt}{4k^2}) = N(0, \sigma^2). \qquad \square$$

Distribución límite de $\hat{T}_n\left(\hat{g}_n\right)$

Definimos

$$\hat{T}_n(g) = \arg\min_{\theta \in \Theta} \left\| \hat{k}_n f(\theta, \hat{k}_n) - g \right\|_2$$
$$\hat{\theta}_n = \hat{T}_n(\hat{g}_n) = \arg\min_{\theta \in \Theta} \left\| \hat{k}_n f(\theta, \hat{k}_n) - \hat{g}_n \right\|_2$$
$$\theta_0 = T(g) = \arg\min_{\theta \in \Theta} \left\| k f(\theta, k) - g \right\|_2$$

donde

$$\int_0^1 f^2(x;\theta,k)dx = 1$$

para todo $k \in I$, para todo $\theta \in \Theta$.

Teorema 7. *Bajo la hipótesis (H2), $\hat{T}_n(g)$ existe casi siempre para toda $g \in L^2[0,1]$.*

Demostración: La demostración es idéntica a la del Teorema 1 □

Teorema 8. *Supongamos que $\hat{T}_n(\hat{g}_n)$ es única casi siempre, T es única, que la familia $\mathcal{F}$ satisface las hipótesis (H2) y (H4), entonces*

$$\hat{\theta}_n \to \theta_0 \textit{ en probabilidad.}$$

Demostración: Para hacer una demostración análoga a la del Teorema 2 , demostraremos que

$$\lim_{n\to\infty} p(\hat{k}_n,\theta) = p(k,\theta) \text{ en probabilidad,}$$

donde

$$p(\hat{k}_n,\theta) = \left\| \hat{k}_n f(\theta,\hat{k}_n) - \hat{g}_n \right\|_2$$
$$p(k,\theta) = \| k f(\theta,k) - g \|_2$$

En efecto, por propiedades del valor absoluto y de la norma,

$$\left| p(\hat{k}_n,\theta) - p(k,\theta) \right| \leq \left| \hat{k}_n \right| \left\| f(\theta,\hat{k}_n) - f(\theta,k) \right\|_2 + \left| \hat{k}_n - k \right| + \| \hat{g}_n - g \|_2 \to 0$$

en probabilidad.

Usando el argumento de unicidad de $\hat{T}_n$ y T de manera análoga a la demostración del Teorema 2, obtenemos la consistencia del estimador $\hat{T}_n(\hat{g}_n)$. □

Distribución Límite de $\sqrt{n}\left(\hat{T}_n(\hat{g}_n) - T(g)\right)$.

Obtendremos una expresión para $\hat{\theta}_n - \theta_0$ que nos permita estudiar la distribución asintótica del estadístico:

$$\sqrt{n}\left(\hat{T}_n(\hat{g}_n) - T(g)\right).$$

donde $\hat{\theta}_n - \theta_0 = \hat{T}_n(\hat{g}_n) - T(g)$

Teorema 9. *Supondremos que $f(\mathbf{t},k)$ verifica (H2,H3,H8),bajo la hipótesis (H5), $T(g)$ existe, es única y pertenece al $int(\Theta), \hat{T}_n\left(\hat{g}_n\right)$ existe y es única casi siempre, entonces*

$$\begin{gathered}\hat{\theta}_n-\theta_0=-\left\{\int_0^1 \overset{\bullet\bullet}{f}(x;\theta_0,\hat{k}_n)\ g\left(x\right)\ dx\right\}^{-1}\int_0^1 \overset{\bullet}{f}(x;\theta_0,\hat{k}_n)\left[\hat{g}_n\left(x\right)-g(x)\right]\ dx\\ +\left\{\int_0^1 \overset{\bullet\bullet}{f}\left(x;\theta_0,\hat{k}_n\right)\ g\left(x\right)\ dx\right\}^{-1}\int_0^1\left(\overset{\bullet}{f}(x;\theta_0,\hat{k}_n)-\overset{\bullet}{f}(x;\theta_0,k)\right)g(x)dx+\\ a_n\int_0^1 \overset{\bullet}{f}(x;\theta_0,\hat{k}_n)\left[\,\hat{g}_n\left(x\right)-g(x)\right]dx+\\ a_n\int_0^1\left(\overset{\bullet}{f}(x;\theta_0,\hat{k}_n)-\overset{\bullet}{f}(x;\theta_0,k)\right)g(x)dx\end{gathered}$$

donde

$$\begin{aligned}a_n=-\left\{\int_0^1 \overset{\bullet\bullet}{f}(x;\theta_0,\hat{k}_n)\ g\left(x\right)\ dx\right\}^{-1}A_n\left\{\int_0^1 \overset{\bullet\bullet}{f}(x;\theta_0,\hat{k}_n)\ g\left(x\right)\ dx\right\}^{-1}\\ +R_n\left\{\int_0^1 \overset{\bullet\bullet}{f}(x;\theta_0,\hat{k}_n)\ g\left(x\right)\ dx\right\}^{-1}\end{aligned}$$

$$A_n=\int_0^1 \overset{\bullet\bullet}{f}(x;\theta_0,\hat{k}_n)(\hat{g}_n\left(x\right)-g(x))\ dx+\int_0^1 V(\alpha,x)\hat{g}_n(x)dx\to 0,$$

casi siempre, si $n\to\infty$ y el término R_n depende de A_n.

Demostración: Por el mismo argumento utilizado en la demostración del Teorema 3, se obtienen las siguientes ecuaciones:

$$\int_0^1 \overset{\bullet}{f}\left(x;\theta_0,k\right)g\left(x\right)\ dx=\mathbf{0}$$

$$\int_0^1 \overset{\bullet}{f}\left(x;\hat{\theta}_n,\hat{k}_n\right)\hat{g}_n\left(x\right)\ dx=\mathbf{0},\ \text{casi siempre}$$

de la hipótesis H8 y esta última expresión

$$\begin{gathered}\int_0^1 \overset{\bullet}{f}(x;\theta_0,\hat{k}_n)\hat{g}_n\left(x\right)\ dx+\alpha\int_0^1 \overset{\bullet\bullet}{f}(x;\theta_0,\hat{k}_n)\,e\,\hat{g}_n\left(x\right)\ dx+\\ \alpha\int_0^1 V(\alpha;x)\,e\,\hat{g}_n\left(x\right)\ dx=\mathbf{0}\end{gathered}$$

si $\alpha e=\hat{\theta}_n-\theta_0$

$$
\begin{aligned}
&\left\{ \int_0^1 \overset{\bullet\bullet}{f}(x;\theta_0,\hat{k}_n)\hat{g}_n(x)\ dx + \int_0^1 V(\alpha;x)\ \hat{g}_n(x)\ dx \right\} \left(\hat{\theta}_n - \theta_0\right) \\
= \quad & - \int_0^1 \overset{\bullet}{f}(x;\theta_0,\hat{k}_n)\hat{g}_n(x)\ dx
\end{aligned}
$$

puesto que $\int_0^1 \overset{\bullet\bullet}{f}(x;\theta_0,\hat{k}_n)\, g(x)\ dx$ es no singular, casi siempre

$$
\begin{gathered}
\hat{\theta}_n - \theta_0 = -\left\{ \int_0^1 \overset{\bullet\bullet}{f}(x;\theta_0,\hat{k}_n)\ g(x)\ dx \right\}^{-1} \int_0^1 \overset{\bullet}{f}(x;\theta_0,\hat{k}_n)\left[\hat{g}_n(x) - g(x)\right]\ dx \\
+ \left\{ \int_0^1 \overset{\bullet\bullet}{f}\left(x;\theta_0,\hat{k}_n\right)\ g(x)\ dx \right\}^{-1} \int_0^1 \left(\overset{\bullet}{f}(x;\theta_0,\hat{k}_n) - \overset{\bullet}{f}(x;\theta_0,k) \right) g(x)dx + \\
a_n \int_0^1 \overset{\bullet}{f}(x;\theta_0,\hat{k}_n)\left[\hat{g}_n(x) - g(x)\right] dx + \\
a_n \int_0^1 \left(\overset{\bullet}{f}(x;\theta_0,\hat{k}_n) - \overset{\bullet}{f}(x;\theta_0,k) \right) g(x)dx
\end{gathered}
$$

donde a_n, A_n y R_n verifican las expresiones dadas en el enunciado del teorema. □

Teorema 10. *Bajo las hipótesis (H2,H3,H5), si $f(\mathbf{t},k)$ verifica (H8), $T(g)$ existe, es única y pertenece al int$(\Theta), \hat{T}_n(\hat{g}_n)$ existe y es única casi siempre, entonces el estadístico:*

$$\sqrt{n}\left(\hat{\theta}_n - \theta_0\right)$$

es asintóticamente equivalente a

$$Z_n(\rho_g) + \mathbf{h}(f,\theta_0,k,g) \cdot 2\sqrt{n} \int_0^1 g(t)\left(\hat{g}_n(t) - g(t)\right) dt$$

donde $Z_n(\rho)$ está definido en (4), $\rho_g(t)$ en (H6) y

$$\mathbf{h}(f,\theta_0,k,g) = -A_g^{-1}(\theta_0,k) \int_0^1 \frac{\partial}{\partial k} \overset{\bullet}{f}(x;\theta_0,k)\, g(x)\ dx$$

es un vector $p \times 1$.

Demostración: Por el Teorema 9

$$\sqrt{n}\left(\hat{\theta}_n - \theta_0\right) = I_1(n) + I_2(n) + I_3(n)$$

donde

$$I_1(n) = -\left\{\int_0^1 \overset{\bullet\bullet}{f}(x;\theta_0,\hat{k}_n)\ g(x)\ dx\right\}^{-1}\sqrt{n}\int_0^1 \overset{\bullet}{f}(x;\theta_0,\hat{k}_n)\left[\hat{g}_n(x)-g(x)\right]dx$$

$$I_2(n) = \left\{\int_0^1 \overset{\bullet\bullet}{f}\left(x;\theta_0,\hat{k}_n\right)\ g(x)\ dx\right\}^{-1}\sqrt{n}\int_0^1\left(\overset{\bullet}{f}(x;\theta_0,\hat{k}_n)-\overset{\bullet}{f}(x;\theta_0,k)\right)g(x)dx$$

$$\begin{aligned} I_3(n) = a_n\int_0^1 \overset{\bullet}{f}(x;\theta_0,\hat{k}_n)\sqrt{n}\left[\hat{g}_n(x)-g(x)\right]dx+ \\ a_n\sqrt{n}\int_0^1\left(\overset{\bullet}{f}(x;\theta_0,\hat{k}_n)-\overset{\bullet}{f}(x;\theta_0,k)\right)g(x)dx\end{aligned}$$

Estudiaremos por separado cada uno de los sumandos:

$I_1(n)$ es asintóticamente equivalente a $Z_n(\rho_g)$, por Lema 3, corolario 3 e hipótesis.

Si en $I_2(n)$, aplicamos el teorema del valor medio a

$$\overset{\bullet}{f}(x;\theta_0,\hat{k}_n)-\overset{\bullet}{f}(x;\theta_0,k)$$

y expresamos $\hat{k}_n - k = \dfrac{\hat{k}_n^2-k^2}{\hat{k}_n+k}$ obtenemos:

$$I_2(n) = \left\{\int_0^1 \overset{\bullet\bullet}{f}\left(x;\theta_0,\hat{k}_n\right)\ g(x)\ dx\right\}^{-1}\sqrt{n}\frac{\hat{k}_n^2-k^2}{\hat{k}_n+k}\int_0^1\frac{\partial}{\partial k}\overset{\bullet}{f}(x;\theta_0,\xi_k^n)g(x)dx$$

donde ξ_k^n es un punto entre $\hat{k}_n$ y k, por hipótesis, por el Lema 3, Teorema 6 y Corolario 3 $I_2(n)$ es asintóticamente equivalente a

$$\begin{aligned}\left\{\int_0^1 \overset{\bullet\bullet}{f}\ (x;\theta_0,k)\ \ g(x)\ dx\right\}^{-1}\int_0^1\tfrac{\partial}{\partial k}\overset{\bullet}{f}(x;\theta_0,k)g(x)dx\cdot\tfrac{Z_n(g)}{k} = \\ \mathbf{h}(f,\theta_0,k,g)\tfrac{Z_n(g)}{k}\end{aligned}$$

Puesto que $I_3(n)$ es asintóticamente equivalente a:

$$a_nZ_n(\overset{\bullet}{f}(\theta_0,k))+a_n\left(\int_0^1\frac{\partial}{\partial k}\overset{\bullet}{f}(x;\theta_0,k)g(x)dx\right)\frac{Z_n(g)}{k}$$

donde $a_n\to 0$ en probabilidad $\left|\int_0^1\frac{\partial}{\partial k}\overset{\bullet}{f}(x;\theta_0,k)g(x)dx\right|<\infty$, $Z_n(\overset{\bullet}{f}(\theta_0,k))$ y $Z_n(g)$ tienen distribución asintótica normal, se concluye que $I_3(n)\to 0$ en probabilidad □

Corolario 4. *Bajo las hipótesis del teorema anterior, el estadístico*

$$\sqrt{n}\left(\hat{\theta}_n-\theta_0\right)$$

es asintóticamente equivalente a

$$L_n = \sqrt{n} \sum_{j=1}^{n} \varepsilon_j \left(\rho_g(\xi_j^n) + 2\mathbf{h}(f, \theta_0, k, g) \frac{g(\eta_j^n)}{k} \right) \left(s_j^n - s_{j-1}^n \right)$$

donde $\xi_j^n, \eta_j^n \in \left[s_j^n - s_{j-1}^n \right]$

Demostración: Por el Teorema 10,

$$\sqrt{n} \left(\hat{\theta}_n - \theta_0 \right)$$

es asintóticamente equivalente a

$$Z_n(\rho_g) + \mathbf{h}(f, \theta_0, k, g) \cdot 2\sqrt{n} \int_0^1 g(t) \left(\hat{g}_n(t) - g(t) \right) dt$$

donde

$$\mathbf{h}(f, \theta_0, k, g) = -\left[\int_0^1 \ddot{f}(x; \theta_0, k)\, g(x)\, dx \right]^{-1} \int_0^1 \frac{\partial}{\partial k} \dot{f}(x; \theta_0, k)\, g(x)\, dx$$

y por el Lema 1, $Z_n(\rho_g)$ es asintóticamente equivalente a

$$Y_n(\rho) = \sqrt{n} \sum_{j=1}^{n} \varepsilon_j \rho(\xi_j^n) \left(s_j^n - s_{j-1}^n \right), \xi_j^n \in \left[s_j^n - s_{j-1}^n \right], \rho = \rho_g$$

y por otro lado, para $\rho = g$, $\mathbf{h}(f, \theta_0, k, g) \cdot 2\sqrt{n} \int_0^1 g(t) \left(\hat{g}_n(t) - g(t) \right) dt$ es asintóticamente equivalente a

$$2\mathbf{h}(f, \theta_0, k, g) Y_n(g) = 2\mathbf{h}(f, \theta_0, k, g) \sqrt{n} \sum_{j=1}^{n} \varepsilon_j \frac{g(\eta_j^n)}{k} \left(s_j^n - s_{j-1}^n \right), \eta_j^n \in \left[s_j^n - s_{j-1}^n \right]$$

por Lema 3 y Lema 1, de donde se obtiene el resultado □

Teorema 11. *Si ρ_g y g son continuas y denotamos por $Var(L_n)$ la matriz de varianzas y covarianza de L_n*

$$Var(L_n) = Var(L_n L_n^T) = \left(\mathbb{E}\left(L_{im}(n) \right) \right)_{p \times p}$$

donde

$$L_{im}(n) = A_i(n) \cdot A_m(n)$$
$$A_i(n) = \left[\sqrt{n} \sum_{j=1}^{n} \varepsilon_j \left(\rho_g^i(\xi_j^n) + h^i(f, \theta_0, k, g) \frac{g(\eta_j^n)}{k} \right) \left(s_j^n - s_{j-1}^n \right) \right],$$
$$\xi_j^n, \eta_j^n \in \left(s_j^n - s_{j-1}^n \right)$$

entonces:

$$\mathbb{E}\left(L_{im}(n) \right) \underset{n \to \infty}{\longrightarrow} \sigma^2 \int_0^1 \rho_g^i(t) \rho_g^m(t) dt +$$
$$\frac{\sigma^2}{\|g\|_2} \int_0^1 \rho_g^i(t) h^m(f, \theta_0, k, g) g(t) dt + \frac{\sigma^2}{\|g\|_2} \int_0^1 h^i(f, \theta_0, k, g) \rho_g^m(t) g(t) dt +$$
$$\sigma^2 h^i(f, \theta_0, k, g) h^m(f, \theta_0, k, g) \|g\|_2$$

donde $\rho_g^i(t)$ es la componente i del vector $\rho_g(t)$ y $h^i(f, \theta_0, k, g)$ es la componente i del vector $\mathbf{h}(f, \theta_0, k, g)$.

Demostración: Puesto que $\max_j \left\| s_j^n - s_{j-1}^n - \frac{1}{n} \right\| = O\left(\frac{1}{n^a} \right)$ para $a > 1$, usando la continuidad de ρ_g y de g, y el hecho que las variables aleatorias $\varepsilon_j, j = 1, 2, \cdots, n$ son independientes, centradas y de varianza σ^2:

$$\mathbb{E}\left(L_{im}(n) \right) = n\sigma^2 \sum_{j=1}^{n} B_i(n) B_m(n) \left(s_j^n - s_{j-1}^n \right)^2$$
$$B_i(n) = \left(\rho_g^i(\xi_j^n) + h^i(f, \theta_0, k, g) \frac{g(\eta_j^n)}{k} \right)$$

$$\mathbb{E}\left(L_{im}(n) \right) \underset{n \to \infty}{\longrightarrow} \sigma^2 \int_0^1 \rho_g^i(t) \rho_g^m(t) dt +$$
$$\frac{\sigma^2}{\|g\|_2} \int_0^1 \rho_g^i(t) h^m(f, \theta_0, k, g) g(t) dt + \frac{\sigma^2}{\|g\|_2} \int_0^1 h^i(f, \theta_0, k, g) \rho_g^m(t) g(t) dt +$$
$$\sigma^2 h^i(f, \theta_0, k, g) h^m(f, \theta_0, k, g) \|g\|_2$$

$\square$

Teorema 12. *Si ρ_g y g son continuas*

$$L_n \xrightarrow{\mathcal{D}} \mathcal{N}\left[0, V(\sigma^2, g) \right]$$

donde

$$V(\sigma^2,g)=\sigma^2\int\limits_0^1\rho_g(t)\rho_g^T(t)dt+$$
$$\frac{\sigma^2}{\|g\|_2}\int\limits_0^1\rho_g(t)\mathbf{h}^T(f,\theta_0,k,g)g(t)dt+\frac{\sigma^2}{\|g\|_2}\int\limits_0^1\mathbf{h}(f,\theta_0,k,g)\rho_g^T(t)g(t)dt+$$
$$\sigma^2\mathbf{h}(f,\theta_0,k,g)\mathbf{h}^T(f,\theta_0,k,g)\left\|g\right\|_2$$

Demostración: Sea $y\in\mathbb{R}^p$ un vector no nulo, probaremos que

$$<L_n,y>\xrightarrow{\mathcal{D}}\mathcal{N}\left[0,V(\sigma^2,g,y)\right]$$

$$<L_n,y>=\sum_{j=1}^n\varepsilon_j a_{nj}(y)$$

usando las propiedades del producto escalar, donde

$$a_n(j)=\sqrt{n}\left[<\rho_g(\xi_j^n),y>+<\mathbf{h}(f,\theta_0,k,g),y>\frac{g(\eta_j^n)}{k}\right]\left(s_j^n-s_{j-1}^n\right).$$

Si

$$c_{nj}(y)=\varepsilon_j a_{nj}(y),\ S_n(y)=Var(<L_n,y>),$$

entonces tenemos que

$$<L_n,y>=\sum_{j=1}^n c_{nj}(y),$$

probaremos a continuación que $<L_n,y>$ satisface la condición de Lindeberg, esto es:

$$\frac{1}{S_n\left(y\right)}\sum_{j=1}^n\int_{\left\{c_{nj}(y)\geq\varepsilon\sqrt{S_n(y)}\right\}}c_{nj}\left(y\right)d\mathbb{P}\rightarrow 0\ ,\text{ para cada }\varepsilon>0,$$

la demostración es análoga a la del Lema 2, para concluir que

$$\frac{\langle L_n,y\rangle}{\sqrt{V\left\langle L_n,y\right\rangle}}\rightarrow N\left[0,1\right]$$

esto es, L_n converge en ley a una distribución normal. □

Corolario 5. *Bajo las hipótesis del Teorema 10 y si ρ_g y g son continuas*

$$\sqrt{n}\left(\hat{\theta}_n-\theta_0\right)\xrightarrow{\mathcal{D}}\mathcal{N}\left[0,V(\sigma^2,g)\right]$$

donde

$$V(\sigma^2, g) = \sigma^2 \int_0^1 \rho_g(t)\rho_g^T(t)dt +$$
$$\frac{\sigma^2}{\|g\|_2} \int_0^1 \rho_g(t)\mathbf{h}^T(f, \theta_0, k, g)g(t)dt + \frac{\sigma^2}{\|g\|_2} \int_0^1 \mathbf{h}(f, \theta_0, k, g)\rho_g^T(t)g(t)dt +$$
$$\sigma^2\mathbf{h}(f, \theta_0, k, g)h^T(f, \theta_0, k, g)\|g\|_2$$

Referencias

[1] Beran, Rudolf. (1977), Minumun Hellinguer distance estimates for parametric models, The Annals of Statistics, **5**, 445–463.

[2] Billingsley, Patric. (1968), Convergence of probability measures. John Wiley.

[3] Gasser, Theo. Muller, Hars-Georg. (1979), Kernel estimation of regression functions. Lectures Notes in Mathematics. **757.** 23–68

[4] Pak, J. (1996). Minimun Hellinguer distance estimation in simple linear regression model; distribution and efficiency. Statistics & Probability Letters. **26.** 263–269.

[5] Wolfowitz, J. (1957). The minimum distance method. The Annals of Mathematical Statistics. 28**.** 58–75.