

ESTIMADORES DE VARIANZA EN REGRESIÓN NO PARAMÉTRICA BASADOS EN SUCESIÓN DE DIFERENCIAS

María C. Paz Sabogal

*Profesor Auxiliar. Universidad del Valle, Escuela de Ingeniería Industrial y Estadística, Cali.
karo.paz@gmail.com*

Luz A. Pereira Hoyos

*Profesor Auxiliar. Universidad del Valle, Escuela de Ingeniería Industrial y Estadística, Cali.
luzadrianaph@yahoo.com*

Javier Olaya Ochoa

*Profesor Titular. Universidad del Valle, Escuela de Ingeniería Industrial y Estadística, Cali.
olaya@univalle.edu.co*

Resumen

En este trabajo se pretende mostrar los tres principales estimadores de varianza que se han propuesto en regresión no paramétrica cuando se tiene una observación de la variable respuesta Y por cada punto de la variable independiente X , los cuales han sido una herramienta fundamental para obtener un estimado del parámetro de la varianza, ellos son el estimador de Rice (1984), el estimador de Gasser – Sroka y Jennen Steinmetz (1986) y el estimador de Hall – Kay y Tittetington (1990), todo esto con el objetivo de mostrar que tanto en regresión paramétrica como en regresión no paramétrica es de importancia además de la estimación de la función de regresión, la estimación de la varianza σ^2 , con el propósito de estimar y probar características de la función de regresión.

Palabras clave: Regresión no paramétrica, Estimadores, Varianza.

Abstract

This paper aims to show the three main estimators variance that have been proposed in nonparametric regression when you have a comment on the response variable And for every point on the independent variable X , which have been an essential tool to obtain an estimate parameter variance, they are the estimator Rice (1984), the estimator Gasser – Sroka y Jennen Steinmetz (1986) and the estimator Hall - Kay and Tittetington (1990), all this with the aim of showing that both Parametric regression as a nonparametric regression is important addition to estimating the regression function, the estimate of variance, with the purpose of estimating and test characteristics of the role of regression.

Keywords: Nonparametric regression, Estimators, Variance.

1. Introducción

En regresión lineal se requiere una estimación de la varianza (σ^2) para probar hipótesis y formar estimados de intervalos pertenecientes al modelo de regresión. Bajo la misma filosofía, en regresión no paramétrica también resulta de importancia el cálculo de la estimación de varianza, siendo este el único parámetro a estimar en un modelo de regresión de este tipo. Por lo tanto, a través de varios años han surgido tres principales estimadores de varianza en regresión no paramétrica, estimador de Rice (1984), estimador de Gasser, Sroka y Jennen-Steinmetz (1986) y estimador de Hall, Kay y Titterington (1990), que han sido una herramienta fundamental para la obtención de una estimación del parámetro de la varianza.

2. Análisis de regresión

En análisis de regresión, una pregunta clave es cuál será el modelo más adecuado para estudiar el comportamiento de la variable respuesta con respecto al conjunto de variables conocidas como variables explicativas. En el contexto paramétrico, la respuesta se busca dentro de un conjunto de funciones cuya forma es conocida, pero de la cual se desconocen algunas partes específicas llamadas *parámetros* (Olaya 2002).

En Eubank (1999) se describe el modelo de regresión en la siguiente forma general. Se supone, que se desea producir información sobre $\mathbf{Y}$ a partir del conocimiento de $\mathbf{X}$. Se dispone de observaciones de la variable continua $\mathbf{Y}$ para n valores predeterminados de p variables continuas $x_1, x_2, \dots, x_p$, las cuales se representan como $\mathbf{X}$. Sean $(x_i, y_i) = (x_{1i}, x_{2i}, \dots, x_{pi}, y_i)$, $i = 1, \dots, n$ las n observaciones de $(\mathbf{X}, \mathbf{Y})$. Se asume que estas observaciones se relacionan a través del modelo muestral:

$$y_i = m(x_i) + \varepsilon_i \quad i = 1, \dots, n \quad (1)$$

Donde:

$$E(\varepsilon_i) = 0 \text{ y } \text{var}(\varepsilon_i) = \sigma^2 < \infty.$$

Los $m(x_i)$ son valores de una función m en los puntos $x_1, x_2, \dots, x_p$. $\mathbf{Y}$ se llama comúnmente la *respuesta*, y $\mathbf{X}$ la *covariable* (o *variable de predicción*, o *variable independiente*, o *variable explicativa*).

3. Análisis de Regresión no Paramétrica

El nombre de regresión no paramétrica, hace idea de que se trata de la contraparte de lo que se conoce como regresión Paramétrica. “Los objetivos del *Análisis de Regresión No-Paramétrica* son los mismos de su contraparte paramétrica, vale decir, estimar y probar las características de la *función de regresión*.” En Eubank (1999) al hacer introducción a la regresión no paramétrica se menciona principalmente que en este tipo de regresión los datos son utilizados para determinar el espacio de la función a la cual pertenece, siendo esta función una curva de regresión desconocida. En regresión paramétrica, en contraste a lo que ocurre en el modelo no paramétrico, el investigador escoge una posible familia de curvas m , de un conjunto global de curvas, y con la familia de curvas elegidas desarrolla el proceso inferencial, la información que puede suministrar los datos concerniente al modelo es restringida bajo la forma del modelo paramétrico supuesto. Por lo tanto, las técnicas de regresión no paramétrica confían más en el peso de la información de los datos acerca de m que su contraparte paramétrica.

Cabe precisar que en regresión no paramétrica la forma de la función de regresión por la cual se relacionan la variable explicativa ($\mathbf{X}$) y respuesta ($\mathbf{Y}$) es desconocida, pero con una característica particular denominada “suavidad”, de ahí que el procedimiento utilizado para estimar la función de regresión m en regresión no paramétrica se denomina “Suavización”, los datos se dejan hablar y elegir una función sobre la cual solo sabemos que es suave, de tal modo que, si se quiere estimar m en x entonces se espera que las observaciones y_i asociadas a los x_i más cercanos a x , contengan alguna información de m en ese punto, por lo tanto es posible

promediar de alguna forma las respuestas y_i mas cercanas al punto donde se desea obtener una estimación de $m(x)$ (Eubank, 1999). m es una función perteneciente a algún espacio de funciones W . La selección de W depende de qué tan suave es posible asumir m ; usualmente W se asume como el conjunto de todas las funciones m que son continuas en un intervalo cerrado $[a, b]$ o quizás como el conjunto de todas las funciones m que tienen k derivadas continuas en $[a, b]$, para algún entero positivo k . Los límites los fija el investigador, pero se pueden asumir, sin pérdida de generalidad $a=0$ y $b=1$ (Olaya 2002).

El procedimiento de “Suavización” utilizado para la estimación de la curva requiere del uso de suavizadores, los cuales son una herramienta útil para resumir el comportamiento de una respuesta Y como función de una variable explicativa X .

4. Estimación de la varianza en regresión no paramétrica

A fin de completar la descripción de un modelo de regresión no paramétrico $y_i = m(x_i) + \varepsilon_i$, donde $m(x_i)$ es una función desconocida, $x_i \in [0, 1]$, $0 \leq x_1 \leq \dots \leq x_n \leq 1$, y los ε_i son variables aleatorias independientes idénticamente distribuidas, con media cero ($\mu = 0$) y varianza constante (σ^2), es necesario suministrar una estimación del parámetro de la varianza. Durante las dos últimas décadas varios estimadores de σ^2 han sido propuestos, todas estas clases de estimadores son aproximaciones de formas cuadráticas del vector de observaciones $Y = (Y_1, Y_2, \dots, Y_n)'$ (Dette, Munk, Wagner, 1998).

Bowman y Azzalini (1997), presentan la estimación de σ^2 de la siguiente forma:

En un modelo lineal estándar la suma de cuadrados de los residuales suministra una base

para el cálculo de la varianza en un escenario no paramétrico,

$$SCR = \sum \{y_i - \hat{m}(x_i)\}^2 \quad (2)$$

Usando la notación matricial, los valores ajustados $\hat{m}$ son dados por $\hat{m} = \mathbf{S}y$, donde los puntos de evaluación son los puntos de diseño x_i . La matriz $\mathbf{S}$ es análoga a la matriz hat, $\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ en un modelo lineal estándar de la forma $y = X\theta + \varepsilon$. Los grados de libertad para el error en un modelo lineal puede ser expresado como $tr(\mathbf{I} - \mathbf{H})$, si se realiza una definición análoga se obtiene una aproximación de los grados de libertad en un escenario no paramétrico de la siguiente forma:

$df_{error} = tr(\mathbf{I} - \mathbf{S})$, donde tr denota la traza de la matriz. La cantidad:

$$\hat{\sigma}^2 = SCR / df_{error} \quad (3)$$

es una estimación natural de σ^2 . La analogía con un modelo lineal es complicada por la presencia del sesgo en un estimador $\hat{m}$, lo que tiene un efecto de inflación en el tamaño de la suma de cuadrados de los residuales. Una posible solución es reducir el sesgo, empleando un pequeño parámetro de suavización para propósitos de estimar σ^2 .

Los estimadores que han sido propuestos para σ^2 basados en las formas cuadráticas del vector de observaciones $Y = (Y_1, Y_2, \dots, Y_n)'$ siguen la forma general:

$$\hat{\sigma}_D^2 = Y'DY / tr(D) \quad (4)$$

Y se encuentran divididos en dos clases (Tiejun Tong y Yuedong Wang (2005); Dette, Munk, Wagner, 1998):

1. En la primera clase, los estimadores dependen explícitamente del parámetro de suavización y son aquellos estimadores basados en la suma de cuadrados de los

residuales resultantes de algún ajuste no paramétrico de m por métodos de suavización tales como Spline o Kernel. En general, este es un ajuste lineal de la forma $\hat{Y} = AY$, donde A es una matriz de suavización, que conduce a un estimador correspondiente a la forma general que presenta la ecuación (2) con $D = (I - A)^T(I - A)$. Estos estimadores de primera clase se conocen como estimadores basados en residuales, dependen críticamente de la cantidad de suavización. Algunos métodos requieren el conocimiento de cantidades desconocidas como:

$$\int_0^1 m'(x)^2 dx \quad \text{o} \quad \int_0^1 m''(x)^2 dx$$

- En la segunda clase, los estimadores no dependen explícitamente de la constante de suavización, usan diferencias para remover la tendencia en la función media, una idea originada del análisis de series de tiempo. Esta clase de método no requiere una estimación de la función media y son llamados estimadores basados en diferencias. Aquí D es una matriz simétrica $n \times n$, definida positiva que depende de los puntos de diseño y no de las observaciones. A primera vista estos estimadores parecen ser independientes de una constante de suavización, pero se debe observar que el orden de las diferencias tiene una influencia, la cual es comparable con el parámetro de suavización de la primera clase. El orden de un estimador basado en diferencias es el número de observaciones involucradas para calcular un residual local. En la literatura, casi todos los métodos basados en diferencias han sido desarrollados de forma univariada en x , con excepción de Kulasekera y Gallagher (2002) quienes extendieron los métodos basados en diferencias al espacio multivariado $x \in [0,1]^d$. Sin embargo, los métodos de Kulasekera y Gallagher (2002) requieren de un orden artificial de los puntos de diseño en $x \in [0,1]^d$.

Entre los estimadores de varianza más conocidos de la segunda clase que han sido propuestos, de forma univariada en x , se encuentra el estimador de Hall, Kay y Titterton (1990), Gasser, Sroka y Jennen-Steinmetz (1986) y Rice (1984).

4.1 Estimador de Rice

El estimador propuesto por Rice (1984), es de primer orden y conocido como un estimador simple de residual local, está basado en las diferencias ordinarias de la variable respuesta (Dette, Munk, Wagner, 1998). Específicamente, la forma del estimador es:

$$\hat{\sigma}_R^2 = \frac{1}{2(n-1)} \sum_{i=2}^n (y_i - y_{i-1})^2 \quad (5)$$

Donde, por simple notación, se asume que las observaciones (x_i, y_i) han sido ordenadas por x_i , es decir $0 \leq x_1 \leq \dots \leq x_n \leq 1$ (Bowman y Azzalini, 1997). Una desventaja de esta aproximación es que las diferencias $y_i - y_{i-1}$ resultan influenciadas por la fluctuación de la función de regresión subyacente a las observaciones y_i y por lo tanto se producirá una sobreestimación de σ^2 . Bowman y Azzalini (1997) hacen alusión al comportamiento de este estimador considerando que:

$$E(y_i - y_{i-1})^2 = 2\sigma^2 + \{m(x_i) - m(x_{i-1})\}^2$$

De ahí que:

$$E(\hat{\sigma}_R^2) = E\left(\frac{1}{2(n-1)} \sum_{i=2}^n (y_i - y_{i-1})^2\right)$$

$$E(\hat{\sigma}_R^2) = \frac{1}{2(n-1)} \left[\sum_{i=2}^n 2\sigma^2 + \sum_{i=2}^n \{m(x_i) - m(x_{i-1})\}^2 \right]$$

$$E(\hat{\sigma}_R^2) = \frac{2(n-1)\sigma^2}{2(n-1)} + \left[\frac{1}{2(n-1)} \sum_{i=2}^n \{m(x_i) - m(x_{i-1})\}^2 \right]$$

$$E(\hat{\sigma}_R^2) = \sigma^2 + \text{sesgo}$$

De tal manera que si el tamaño de muestra es muy grande, n tiende a infinito ($n \rightarrow \infty$), el sesgo tiende a cero (Bowman y Azzalini, 1997).

4.2 Estimador de Gasser – Sroka y Jennen Steinmetz

Gasser, Sroka y Jennen-Steinmetz (1986) proponen un estimador de varianza de segundo orden (Dette, Munk, Wagner, 1998), que trabaja el problema de la influencia de las fluctuaciones del estimador de Rice, usando una interpolación lineal para la estimación de la observación i -ésima, lo cual involucra la creación de pseudo-residuales definidos como:

$$\tilde{\varepsilon}_i = \frac{x_{i+1} - x_i}{x_{i+1} - x_{i-1}} y_{i-1} + \frac{x_i - x_{i-1}}{x_{i+1} - x_{i-1}} y_{i+1} - y_i \quad (6)$$

Estos miden la diferencia entre los y_i y la línea que asocia sus dos vecinos más cercanos (Bowman y Azzalini, 1997), en otras palabras, mide la diferencia entre y_i y el valor x_i en el ajuste lineal resultante de (x_{i-1}, y_{i-1}) y (x_{i+1}, y_{i+1}) (Dette, Munk, Wagner, 1998).

Si $\tilde{\varepsilon}_i$ es escrito en la forma

$$a_i y_{i-1} + b_i y_{i+1} - y_i, \quad \text{con} \quad a = \frac{x_{i+1} - x_i}{x_{i+1} - x_{i-1}} \quad \text{y}$$

$$b = \frac{x_i - x_{i-1}}{x_{i+1} - x_{i-1}},$$

Entonces el estimador de la varianza es:

$$\hat{\sigma}_{GSJ}^2 = \frac{1}{n-2} \sum_{i=2}^{n-1} c_i^2 \tilde{\varepsilon}_i^2 \quad (7)$$

Donde $c_i^2 = (a_i^2 + b_i^2 + 1)^{-1}$ corresponde a la expresión del coeficiente inverso de la varianza de los $\tilde{\varepsilon}_i$ (Bowman y Azzalini, 1997), puesto que:

$$\text{var}(\tilde{\varepsilon}_i) = \text{var}(a_i y_{i-1} + b_i y_{i+1} - y_i)$$

$$\text{var}(\tilde{\varepsilon}_i) = a_i^2 \text{var}(y_{i-1}) + b_i^2 \text{var}(y_{i+1}) + \text{var}(y_i)$$

$$\text{var}(\tilde{\varepsilon}_i) = a_i^2 \sigma^2 + b_i^2 \sigma^2 + \sigma^2$$

$$\text{var}(\tilde{\varepsilon}_i) = \sigma^2 (a_i^2 + b_i^2 + 1)$$

$$\text{Por lo tanto: } \text{var}(\tilde{\varepsilon}_i) = \sigma^2 c_i^2.$$

De igual forma es posible obtener la esperanza de los pseudo-residuales como:

$$E(\tilde{\varepsilon}_i^2) = (a_i^2 + b_i^2 + 1) \sigma^2 + O(n^{-2}) \quad (\text{Gasser, Sroka y Jennen-Steinmetz, 1986}).$$

Donde $O(n^{-\delta})$ es la tasa de convergencia que hace que el riesgo decaiga a cero para los estimadores no paramétricos, para algún $\delta > 0$ que depende de la suavidad de m (Eubank, 1999).

Para puntos de diseños equidistantes, $\hat{\sigma}_{GSJ}^2$ se reduce a:

$$\hat{\sigma}_{GSJ}^2 = \frac{2}{3(n-2)} \sum_{i=2}^{n-1} \left(\frac{1}{2} y_{i-1} - y_i + \frac{1}{2} y_{i+1} \right)^2 \quad (8)$$

Gasser, Sroka y Jennen-Steinmetz (1986), hacen mención a los siguientes supuestos, los cuales son necesarios para obtener resultados asintóticos:

- Supuesto 1: No existen medidas múltiples en algún punto de diseño, esto es $a = x_1 < x_2 < \dots < x_n = b$; $a = 0$, $b = 1$ sin pérdida de generalidad.

- Supuesto 2: Se requiere que:

$$\max |x_i - x_{i-1}| = O(1/n)$$

- Supuesto 3: Los errores $\{\varepsilon_i\}$ son independientes e idénticamente distribuidos con $E(\varepsilon_i) = 0$, $\text{var}(\varepsilon_i) = \sigma^2$, $E(\varepsilon_i^4) < \infty$.

- Supuesto 4: La función m es continua.

Si $\mathbf{C}$ denota una matriz diagonal $(n-2) \times (n-2)$ con $\mathbf{C}_{ii} = c_{i+1}$, $\mathbf{A}$ denota una

matriz tridiagonal $(n-2) \times n$ con elementos

$$\mathbf{A}_{ii} = a_{i+1}, \mathbf{A}_{i,i+1} = -1, \mathbf{A}_{i,i+2} = b_{i+1} \\ \text{y } \mathbf{D} = \mathbf{A}'\mathbf{C}^2\mathbf{A}.$$

El vector de pseudo residuales $\tilde{\boldsymbol{\varepsilon}} = (\tilde{\varepsilon}_2, \dots, \tilde{\varepsilon}_{n-1})$ y $\boldsymbol{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_n)$ están relacionados por $\tilde{\boldsymbol{\varepsilon}} = \mathbf{A}\boldsymbol{\varepsilon} + \mathbf{B}$, donde $\mathbf{B}$ representa el sesgo, entonces $\hat{\sigma}_{GSJ}^2 = (n-2)^{-1} \|\mathbf{C}\tilde{\boldsymbol{\varepsilon}}\|^2$, lo que conduce a (Gasser, Sroka y Jennen-Steinmetz 1986):

$$\text{var}(\hat{\sigma}_{GSJ}^2) = \frac{2\sigma^4}{(n-2)^2} \text{tr}(\mathbf{D}^2) \\ + \frac{\sigma^4}{(n-2)^2} (\gamma_4 - 3) \sum_{i=1}^n \mathbf{D}_{ii}^2 + \gamma_3 O\left(\frac{1}{n^3}\right) + O\left(\frac{1}{n^5}\right) \quad (9)$$

Donde:

$$\gamma_i = E\left[\left(\frac{\varepsilon}{\sigma}\right)^i\right], i = 3, 4$$

$$E(\varepsilon_i^3) = \gamma_3 \sigma^3, E(\varepsilon_i^4) = \gamma_4 \sigma^4$$

$O(n^{-\delta})$ Es la tasa de convergencia que hace que el riesgo decaiga a cero para los estimadores no paramétricos, para algún $\delta > 0$ que depende de la suavidad de $\mathbf{m}$ (Eubank, 1999).

El primer término restado γ_4 se anula si la distribución de los errores es simétrica. Para un diseño equidistante con errores normalmente distribuidos, la expresión anterior se reduce a:

$$\sigma^4 \left(2 + \frac{17}{9}\right) / (n-2).$$

4.3. Estimador de Hall – Kay y Titterton

Hall, Kay y Titterton en 1990, introducen otro estimador basado en diferencias sucesivas $\{d_i\}_{i=-m_1, \dots, m_2}$, este estimador es una generalización de los dos anteriormente mencionados, con:

$$\sum_{k=0}^r d_{ik} = 0 \quad (10)$$

$$\sum_{k=0}^r d_{ik}^2 = 1 \quad (11)$$

Donde $d_{-m_1}, d_{m_2} \neq 0$, m_1 y m_2 son enteros positivos y $r = m_1 + m_2$ denota el orden del estimador de varianza:

$$\hat{\sigma}_{D,r}^2 = \frac{1}{n-r} \mathbf{Y}'\mathbf{D}\mathbf{Y} = \frac{1}{n-r} \sum_{i=m_1+1}^{n-m_2} \left(\sum_{k=-m_1}^{m_2} d_{ik} Y_{i+k} \right)^2 \quad (12)$$

Note que si $r=1$ se obtiene el estimador de Rice (1984), donde $(d_0, d_1) = (2^{-1/2}, -2^{-1/2})$. Ahora si $r=2$ se logra el estimador de Gasser, Sroka y Jennen-Steinmetz, donde:

$$(d_0, d_1, d_2) = (6^{-1/2}, \left(\frac{2}{3}\right)^{1/2}, -6^{-1/2})$$

para puntos de diseño equidistantes (Dette, Munk, Wagner, 1998).

Hall, Kay y Titterton (1990), en su artículo titulado “estimación de la varianza en regresión no paramétrica basada en diferencias óptimas asintóticas”, exponen el cálculo de las sucesiones de diferencias que resultan óptimas asintóticamente al utilizar este tipo de estimadores, puesto que por ejemplo para un estimador de segundo orden existen diferentes sucesiones que se podrían utilizar en su cálculo y cumplen las condiciones de la ecuación (10) y ecuación (11), tal es el caso de la sucesión de diferencias más comúnmente usada

$$(d_0, d_1, d_2) = \left(-6^{-\frac{1}{2}}, \left(\frac{2}{3}\right)^{\frac{1}{2}}, -6^{-\frac{1}{2}} \right), \text{ aunque como}$$

mencionan estos autores resulta sustancialmente mejor utilizar

$$(d_0, d_1, d_2) = \left(\frac{1}{4} \left(5^{\frac{1}{2}} + 1 \right), -\frac{1}{2}, -\frac{1}{4} \left(5^{\frac{1}{2}} - 1 \right) \right) \quad \text{para}$$

llevar a cabo la estimación, las diferencias de primer orden son únicas. Para los estimadores basados en sucesiones de diferencias óptimas la

varianza asintótica y el error cuadrático medio son ambos iguales a $n^{-1}\tau^2$, donde:

$$\begin{aligned}\tau^2 &= \text{var}(\varepsilon^2) + 2\sigma^4 \sum_{k \neq 0} \left(\sum_j d_j d_{j+k} \right)^2 \\ &= \sigma^4 \left\{ \kappa + 2 \sum_k \left(\sum_j d_j d_{j+k} \right)^2 \right\}\end{aligned}\quad (13)$$

Donde κ denota la curtosis de ε/σ , $\kappa=0$ en caso de que los errores sean normales y $\kappa>0$ en caso de que la distribución de los errores tenga un comportamiento cercano a la normal.

Como ya se mencionó las sucesiones de diferencias de primer orden son únicas, de ahí que el problema se centre en la escogencia de las diferencias de un orden mayor. Asumiendo por simplicidad que $m_1=0$, $m_2=r$. La selección óptima de la sucesión $d_0, \dots, d_r$ implica

$\sum_{j=1}^r d_j d_{j+k} = -(2r)^{-1} \quad (1 \leq |k| \leq r)$. Además de ello la varianza mínima asintótica al utilizar un r -ésimo orden de sucesiones está dada por $n^{-1}\tau_1^2$, donde: $\tau_1^2 = \text{var}(\varepsilon^2) + r^{-1}\sigma^4$.

5. Conclusiones

Los estimadores basados en diferencias son atractivos desde el punto de vista práctico, porque ellos frecuentemente tienen pequeños sesgos para tamaños de muestra pequeños. En aplicaciones prácticas la elección del orden r y una apropiada sucesión de las diferencias, es delicado, puesto que estas dos componentes son aquellas que minimizan el Error Cuadrático Medio (*ECM*)

Tabla 1. Sucesiones de diferencias óptimas para $0 \leq r \leq 10$. Cifras redondeadas a cuatro decimales.

r	$(d_0, \dots, d_r)$
1	(0.7071, -0.7071)
2	(0.8090, -0.5, -0.3090)
3	(0.1942, 0.2809, 0.3832, -0.8582)
4	(0.2708, -0.0142, 0.6909, -0.4858, -0.4617)
5	(0.9064, -0.2600, -0.2167, -0.1774, -0.1420, -0.1103)
6	(0.2400, 0.0300, -0.0342, 0.7738, -0.3587, -0.3038, -0.3472)
7	(0.9302, -0.1965, -0.1728, -0.1506, -0.1299, -0.1107, -0.0930, -0.0768)
8	(0.2171, 0.0467, -0.0046, -0.0348, 0.8207, -0.2860, -0.2453, -0.2260, -0.2879)
9	(0.9443, -0.1578, -0.1429, -0.1287, -0.1152, -0.1025, -0.0905, -0.0792, -0.0687, -0.0588)
10	(0.1995, 0.0539, 0.0104, -0.0140, -0.0325, 0.8510, -0.2384, -0.2079, -0.1882, -0.1830, -0.2507)

minimizar: $\zeta = \sum_{k \neq 0} \left(\sum_j d_j d_{j+k} \right)^2$ sujeta a las condiciones de las ecuaciones (10) y (11).

La Tabla 1. lista las sucesiones de orden r de diferencias óptimas para $1 \leq r \leq 10$. Estas sucesiones de diferencias óptimas son únicas en orden y signo (Hall, Kay y Titterington, 1990). Para el orden r -ésimo de las sucesiones de diferencias óptimas ζ es equivalente a $(2r)^{-1}$ y

en muestras finitas. Tiejun, Tong y Yuedong Wang (2005) al realizar repetidas simulaciones encontraron que el estimador de Hall – Kay y Tittetington (1990) (1990) se desempeña mejor cuando la función m es suave y/o el tamaño de muestra n es grande. Con respecto al estimador de Gasser – Sroka y Jennen Steinmetz (1986) se encontró que este se desempeña mejor cuando m es poco suave y/o n es pequeño.

6. Referencias bibliográficas

1. Bowman, A. W. and Azzalini, A. (1997) Applied smoothing techniques for data analysis: the kernel approach with S-Plus illustrations. Oxford University Press, Oxford.
2. Dette, H., Munk, A. and Wagner, T. (1998) Estimating the variance in nonparametric regression – What is a reasonable choice? Journal of the royal statistical society. Series B (Statistical methodology), Vol. 60, No. 4, 751 – 764.
3. Eubank, R. L. (1999) Nonparametric regression and spline smoothing. Marcel Dekker, New York, second edition.
4. Gasser, T., Sroka, L., Jennen-Steinmetz, C. (Dec., 1986) Residual variance and residual pattern in nonlinear regression. Biometrika, Vol.73, No.3, 625-633.
5. Hall, P., Kay, W.J., Titterington, D.M. (Sep., 1990) Asymptotically Optimal Difference-Based Estimation of Variance in Nonparametric Regression. Biometrika, Vol.77, No.3, 521-528. Hall, P. and Marron, J.S. (1990) On variance estimation in nonparametric regression. Biometrika, 77, 415-419.
6. Kulasekera, K. and Gallagher, C. (2002) “Variance estimation in nonparametric multiple regression”, Communications in Statistics 15, 610-635.
7. Olaya, J. (2002) Suavización y regresión no paramétrica. Universidad del Valle.
8. Tong, T., Wang, Y. (2005) Estimating residual variance in nonparametric regression using least squares. Department of statistics and applied probability, University of California, Santa Barbara.