

Divulgaciones Matemáticas Vol. 22, No. 2 (2021), pp. 48–65
<https://produccioncientificaluz.org/index.php/divulgaciones/>
DOI: <https://doi.org/10.5281/zenodo.7487484>
 (CC BY-NC-SA 4.0)

©Autor(es)
e-ISSN 2731-2437
p-ISSN 1315-2068

Notas sobre el desempeño de los estimadores fronteras de densidad con núcleo localmente adaptable y con conjunto difuso

*Notes on the performance of the boundary locally adaptive kernel and boundary
fuzzy set density estimators*

Jesús A. Fajardo
(jfajardogonzalez@gmail.com; jfajardo@udo.edu.ve)
ORCID: <https://orcid.org/0000-0002-3762-4824>
Departamento de Matemática Núcleo Sucre Universidad de Oriente
Cumaná 6101, República Bolivariana de Venezuela

Resumen

Estas notas proporcionan un nuevo resultado relacionado con el problema de estimación no paramétrica de la función de densidad, no basado en núcleos, el cual permite extender el alcance del método de estimación de la densidad con conjunto difuso. Para ello y bajo la presencia del problema frontera en los estimadores de la densidad con núcleo y con conjunto difuso, se considerarán los estimadores fronteras de la densidad con núcleo localmente adaptable y con conjunto difuso, con el propósito de comparar sus desempeños. Cada rendimiento se obtiene tomando en cuenta cuatro formas de densidades específicas y dos conjuntos de datos reales. Los resultados de las extensas simulaciones muestran que el estimador frontera con conjunto difuso tiene mejor rendimiento en los puntos cercanos a 0, en una dispersión de la vecindad del parámetro b_n , cuando se comparó con el rendimiento del estimador frontera localmente adaptable, para las cuatro formas de densidades consideradas. Aquí, b_n es el ancho de banda del estimador con conjunto difuso.

Palabras y frases clave: Estimador de densidad con conjunto difuso, estimación frontera, estimador de la densidad con núcleo adaptable.

Abstract

These notes provide a new result related to the nonparametric density function estimation problem, not based on kernels, which allows to extend the range of the fuzzy set density estimation method. For this, and under the presence of the boundary problem in the density estimators with kernel and with fuzzy set, the boundary density estimators with locally adaptable kernel and with fuzzy set will be considered, with the purpose of comparing their performances. Each performance is obtained by taking into account four forms of specific densities and two sets of real data. The results of the extensive simulations show that the boundary fuzzy set estimator performs best at points close to 0, at a spread from the neighborhood of the parameter b_n , when compared with the performance of the boundary locally

Recibido 20-09-2021. Revisado 28-11-2021. Aceptado 30/08/2022.

MSC (2010): Primary 62G07; Secondary 62G05.

Autor de correspondencia: Jesús Fajardo.

Esta investigación ha sido apoyada por una subvención de la Academia de Ciencias de América Latina-ACAL.

adaptive estimator, for the four density forms considered. Here, b_n is the bandwidth of the fuzzy set estimator.

Key words and phrases: Fuzzy set density estimator, boundary estimation, adaptive kernel density estimator.

1 Introducción

Las notas del presente artículo están enmarcadas dentro del contexto general de la teoría de estimación no paramétrica de la función de densidad f , con muestras independientes. No obstante, estas se desarrollan considerando un aspecto particular de la teoría señalada previamente: la estimación frontera de f con estimadores no basados en núcleos.

En cada soporte o dominio $[0, \infty)$ ó $[0, 1]$ de una densidad f , los puntos fronteras e interiores se definen a través de un parámetro de suavizado o ancho de banda p_n , $p_n \rightarrow 0$ cuando $n \rightarrow \infty$, donde ambos tienen forma general $x = s p_n$, $s \in [0, 1)$, y $z = k p_n$, $k \geq 1$, respectivamente. Observe que, cada punto frontera e interior satisface $x \in [0, p_n)$ y $z \geq p_n$. Los intervalos $[0, p_n)$ y $[p_n, \infty)$ se llaman región frontera e interior, respectivamente. Cabe destacar que los términos n y p_n forman parte de la expresión que define al estimador de f , donde n es el tamaño de la muestra independiente que se considera para estimar f . La clasificación anterior, para los puntos del soporte de f , se debe a la presencia del fenómeno o problema “efectos fronteras” en el estimador de f . En estas notas, el fenómeno anterior será reseñado como problema frontera y no se tratará teóricamente. No obstante, se subraya que el problema frontera afecta el desempeño general del estimador de f , ya que este es diferente en los puntos fronteras e interiores. Teóricamente ocurre lo siguiente, en los puntos fronteras el sesgo del estimador de f tiene una tasa o velocidad de convergencia más lenta que en los puntos interiores. Técnicamente se tiene que, en los puntos fronteras el sesgo del estimador de f tiene una tasa de convergencia de orden $O(p_n)$ en lugar de $O(p_n^2)$, donde el orden óptimo para la tasa de convergencia del sesgo del estimador de f es $O(p_n^2)$ (ver Stone [17]).

La teoría de estimación no paramétrica de f , con muestras independientes, está formada por dos clases o grupos de estimadores: los tipo núcleo y los no basados en núcleos. No obstante, entre los estimadores de ambos grupos sólo se considerarán los estimadores clásico con núcleo y con conjunto difuso, los cuales se definen a través de una función de núcleo K , y una función atenuante φ (ver Reiss [14], Sección 2.4) de la siguiente manera

$$\hat{f}_K(t) = \frac{1}{nh_n} \sum_{i=1}^n K\left(\frac{X_i - t}{h_n}\right) \quad (1.1)$$

y

$$\hat{v}_n(t) = \frac{1}{n(b_n \int \varphi(u) du)} \sum_{i=1}^n \mathbb{I}_{[0, \varphi(\frac{X_i - t}{b_n})]}(V_i), \quad (1.2)$$

donde X_1, \dots, X_n es una muestra aleatoria independiente de la variable aleatoria X con densidad f , V_1, \dots, V_n son variables aleatorias independientes uniformemente distribuidas en $[0, 1]$ e independientes de X_1, \dots, X_n , h_n y b_n son los parámetros de suavizado de cada estimador con $h_n \rightarrow 0$ y $b_n \rightarrow 0$ cuando $n \rightarrow \infty$, y las funciones $K : \mathbb{R} \rightarrow [0, \infty)$ y $\varphi : \mathbb{R} \rightarrow [0, 1]$ satisfacen las

siguientes condiciones:

$$K(-u) = K(u), \int K(u) du = 1, \int u K(u) du = 0 \text{ y } 0 \neq \int u^2 K(u) du < \infty, \quad (1.3)$$

y

$$0 < \int \varphi(u) du < \infty, \text{ y } \int \varphi(u) du \neq 1 \text{ en general.}$$

En cuanto a los orígenes de ambos estimadores, es oportuno puntualizar que el estimador (1.1) fue introducido hace más de medio siglo y de forma independiente por Rosenblatt [15] y Parzen [13]. En cambio, el estimador (1.2) fue presentado en una fecha más reciente, menos de una década, por Fajardo, Ríos y Rodríguez [6]. Por otro lado, la existencia de una amplia referencia bibliográfica donde se discuten en profundidad las características teóricas de (1.1), justifica el hecho de presentar, en lo que sigue, sólo algunas características teóricas puntuales de (1.2) con función atenuante (c.f.a.) φ :

- Es una versión o caso particular del estimador introducido por Falk y Liese [7].
- No está basado en núcleos, ya que la función $\mathbb{I}_{[0, \varphi(x)]}(v)$ no es una función de núcleo.
- El término “conjunto difuso” fue justificado por Fajardo [4] en la Observación 2, siendo esta una consecuencia directa de la Observación 1 en Fajardo, Ríos y Rodríguez [6].

No obstante, entre las características comunes que comparten los estimadores (1.1) y (1.2) resaltan:

- En la práctica ambos estimadores dependen de parámetros de suavizados particulares y de funciones específicas, lo que fue caracterizado por Fajardo como un paralelismo entre (1.1) y (1.2). Es oportuno destacar que las siguientes funciones minimizan el error cuadrático medio integrado óptimo ($ECMI^*$) de (1.1) y (1.2), respectivamente:

$$K_E(x) = \frac{3}{4}(1-x^2)\mathbb{I}_{[-1,1]}(x), \quad \varphi(x) = \left[1 - \left(\frac{16x}{25}\right)^2\right]\mathbb{I}_{[-\frac{25}{16}, \frac{25}{16}]}(x).$$

(Núcleo de Epanechnikov) (Función Atenuante)

Para más detalles, ver Epanechnikov [3] y Fajardo [4].

- En Fajardo [4] se demostró que (1.2) presenta el mismo comportamiento asintótico de (1.1), donde $n^{-1/5}$ y $n^{-4/5}$ son los valores correspondientes para el orden de las tasas de convergencia óptimas de ambos parámetros de suavizados y ambos $ECMI^*$'s.
- El rol de la función atenuante o función de pertenencia φ (ver [6], Observación 1) fue determinante en los resultados obtenidos por Fajardo, ya que su adecuada definición permitió establecer el siguiente e importante resultado

$$ECMI^* \left[\hat{\vartheta}_n \right] \leq ECMI^* \left[\hat{f}_K \right],$$

el cual garantiza que el estimador $\hat{\vartheta}_n$ proporciona mejores estimaciones que el estimador \hat{f}_K , para todo núcleo K . Destacándose como una función que le permite al estimador (1.2) seleccionar puntos de la muestra con diferentes probabilidades a diferencia de los estimadores clásicos con núcleos, los cuales asignan pesos iguales a todos los puntos de la muestra.

Otra característica no deseada, enmarcada en el contexto de estas notas, es la existencia del problema frontera en (1.1) y (1.2). Los antecedentes bibliográficos señalan que Hominal y Deheuvels [9] describieron el problema frontera en (1.1) para densidades con soporte compacto. Mientras que, en un reciente trabajo, Fajardo y Harmath [5] demostraron la presencia del problema frontera en (1.2) para densidades con soporte $[0, \infty)$. Es oportuno señalar que sólo para el caso del estimador (1.1), el problema frontera han sido estudiado por muchos autores y en la literatura existe una extensa variedad de métodos desarrollados para eliminarlo. Un excelente resumen de algunos métodos muy conocidos son reseñados por Karunamuni y Alberts [12].

Estas notas comparan el desempeño del estimador frontera de f con conjunto difuso, que fue introducido recientemente por Fajardo y Harmath, con el desempeño del estimador propuesto por Karunamuni y Alberts, estimador frontera de f con núcleo localmente adaptable, en puntos cercanos a 0 en una dispersión de la vecindad de b_n . Para ello, se consideraron cuatro formas de densidades específicas, y en los puntos anteriores se realizaron extensas simulaciones para comparar el error cuadrático medio (*ECM*) local de los estimadores definidos por Karunamuni-Alberts y Fajardo-Harmath, observándose que el *ECM* local del estimador propuesto por Fajardo y Harmath es menor. La reducción anterior, muestra que el desempeño del estimador propuesto por Fajardo y Harmath supera el desempeño del estimador definido por Karunamuni y Alberts. Además, es oportuno destacar que el resultado obtenido extiende las propiedades del estimador de f con conjunto difuso, proporcionando un nuevo resultado relacionado con los problemas de estimación no paramétrica de la densidad no basado en núcleo. Cabe resaltar que todas las simulaciones se desarrollaron a través de la plataforma de programación y cálculo numérico conocida como *MATLAB*.

La anterior y particular elección se fundamentó, principalmente, en los resultados de las simulaciones obtenidas por Karunamuni y Alberts para las cuatro formas de densidades consideradas en estas notas. Tales simulaciones mostraron que el estimador de Karunamuni y Alberts funcionó bastante bien para la mayoría de las densidades consideradas, cuando fue comparado con los estimadores propuestos por Cowling y Hall [2], Jones y Foster [10], Zhang y Karunamuni [18] y su sencilla modificación que permite obtener el estimador de ajuste lineal local, Zhang, Karunamuni y Jones [19], y Hall y Park [8]. Entre otras razones que sustentaron la particular elección, se destacan los resultados de las recientes simulaciones presentadas por Fajardo y Harmath [5] para cuatro formas de densidades distintas en su mayoría a las consideradas en estas notas, pero con compartimiento análogo en 0. Tales simulaciones mostraron que el desempeño de su estimador frontera fue superior al desempeño del estimador frontera introducido por Karunamuni y Alberts [11]. Cabe destacar que los resultados de las simulaciones realizadas por Karunamuni y Alberts [11] con las cuatro formas de densidades consideradas por Fajardo y Harmath [5], mostraron que el desempeño de su estimador fue superior cuando se comparó con los estimadores definidos por Jones y Foster [10], Zhang y Karunamuni [18] y su sencilla modificación que permite obtener el estimador de ajuste lineal local, Zhang, Karunamuni y Jones [19], y Hall y Park [8]. Además, otra de las razones que justifican la elección anterior son las propiedades teóricas que comparten los estimadores propuestos por Karunamuni y Alberts [12], y Fajardo y Harmath [5]: no negatividad, “continuación frontera natural”, y en cuanto al desempeño ambos mejoran el sesgo mientras que sus varianzas son pequeñas. Finalmente, es necesario señalar que una revisión de la literatura, sobre el tema propuesto, reveló que no hay evidencia de publicaciones con respecto a comparaciones del desempeño de otros estimadores con el desempeño del estimador propuesto por Karunamuni y Alberts [12].

Las notas están organizadas de la siguiente manera. La Sección 2 incluye los estimadores propuestos por Karunamuni y Alberts [12], y Fajardo y Harmath [5]. Las Secciones 3 y 4 presentan

los análisis de los datos simulados y los datos reales, respectivamente. Las conclusiones en la Sección 5.

2 Estimadores Fronteras de la Densidad

En esta sección se presentan los estimadores propuestos por Karunamuni y Alberts [12], y Fajardo y Harmath [5], considerando sólo los detalles teóricos que permitieron su construcción como estimadores sin problema frontera. Además, se resaltan algunas observaciones únicas, así como las características particulares comunes entre tales estimadores. Es oportuno destacar la importancia de realizar un estudio teórico formal para detectar la presencia o no del problema frontera en el estimador de cualquier función, ya que no es obvio que el comportamiento de un estimador sea el mismo en los puntos fronteras e interiores.

3 Referencias cruzadas, numeración y entorno de teorema

Todos los teoremas, proposiciones, lemas, corolarios, definiciones, etc. deben tener sus propias etiquetas para hacer referencias cruzadas internas; esto se hace usando las instrucciones `\label` y `\ref`. La numeración de todos estos elementos se refiere a cada apartado del artículo, que ya ha sido establecido en el entorno teorema. Cada uno numerado se debe hacer referencia a la ecuación usando la instrucción `\eqref`.

3.1 Estimador Frontera con Núcleo Localmente Adaptable

Sea X una variable aleatoria con función de densidad f , tal que f tiene soporte $[0, \infty)$. Para una muestra aleatoria independiente X_1, \dots, X_n de la variable aleatoria X y para cada $c \in [0, \infty)$, Karunamuni y Alberts inician la construcción de su estimador frontera considerando la siguiente transformación de la muestra original $g(X_1), \dots, g(X_n)$, donde $g : [0, \infty) \rightarrow [0, \infty)$ es una función continua y monótona creciente. A partir de los datos transformados y para $r = ch_n$, con $c \geq 0$, definen el siguiente estimador adaptable

$$\hat{f}_n(r) = \frac{1}{nh_n} \sum_{i=1}^n K\left(\frac{r - g(X_i)}{h_n}\right) / \int_{-1}^c K(u) du, \quad (3.1)$$

donde h_n satisface $h_n \rightarrow 0$ cuando $n \rightarrow \infty$, y las condiciones sobre K se dan en (1.3). Además, calcularon las expresiones para el sesgo y la varianza del estimador (3.1), las cuales se describen en el siguiente teorema. A lo largo de estas notas, se denotará por $q^{(i)}$ la i -ésima derivada de toda función arbitraria q .

Teorema 3.1 (Lema 2.1 en [12]). *Sean f y g funciones cuyas segundas derivadas existen y son continuas en $[0, \infty)$. Además, suponga que $g(0) = 0$ y $g^{(1)}(0) = 1$. Entonces las expresiones para*

el sesgo y la varianza del estimador (3.1) en los puntos $r = ch_n$ con $c \in [0, 1]$, son las siguientes

$$\begin{aligned} \mathbb{E} [\hat{f}_n(r) - f(r)] &= \frac{-h_n}{\int_{-1}^c K(u) du} \left\{ f(0)g^{(2)}(0) \int_{-1}^c (c-u)K(u) du + f^{(1)}(0) \int_{-1}^c uK(u) du \right\} \\ &+ \frac{h_n^2}{2 \int_{-1}^c K(u) du} \left\{ -f^{(2)}(0) c^2 \int_{-1}^c K(u) du + \int_{-1}^c (c-u)^2 K(u) du \left[f^{(2)}(0) \right. \right. \\ &\left. \left. - f(0)g^{(3)}(0) - 3g^{(2)}(0) \left(f^{(1)}(0) - f(0)g^{(2)}(0) \right) \right] \right\} + o(h_n^2), \end{aligned} \tag{3.2}$$

y

$$\begin{aligned} Var [\hat{f}_n(r)] &= \frac{f(0)}{n h_n \left(\int_{-1}^c K(u) du \right)^2} \int_{-1}^c K^2(u) du + o\left(\frac{1}{n h_n}\right) \\ &= \frac{f(r)}{n h_n \left(\int_{-1}^c K(u) du \right)^2} \int_{-1}^c K^2(u) du + o\left(\frac{1}{n h_n}\right). \end{aligned} \tag{3.3}$$

Además, en su trabajo señalaron las siguientes propiedades:

- El término principal de (3.3) no está afectado por la transformación g .
- Cuando $c = 1$, $Var [\hat{f}_n(r)] = Var [\hat{f}_K(r)]$. Es decir, en el punto interior $r = h_n$ la varianza del estimador (3.1) se reduce a la varianza del estimador (1.1).

No obstante, es oportuno señalar que el estimador (3.1) presenta el problema frontera. En efecto, de la expresión (3.2) se desprende que la tasa de convergencia del sesgo del estimador (3.1) es de orden $O(h_n)$, en cada punto de la región frontera $[0, h_n)$ y en el punto interior h_n .

El siguiente paso que Karunamuni y Alberts tomaron para definir su estimador frontera, se baso en construir una función de transformación g con la siguiente propiedad

$$g^{(2)}(0) = -f^{(1)}(0) \int_{-1}^c u K(u) du / f(0) \int_{-1}^c (c-u) K(u) du, \tag{3.4}$$

siempre que $f(0) \neq 0$. Observe que tomando adecuadamente una función g con la propiedad anterior, desaparecerá el problema frontera en (3.1). No obstante, el hecho que (3.4) dependa de c permitió que Karunamuni y Alberts señalaran que tal función g dependería del punto de estimación dentro de la región frontera $[0, h_n)$, y a esa propiedad la llamaron adaptabilidad local. Además, en su trabajo modificaron la notación reemplazando g por g_c , $0 \leq c \leq 1$, para resaltar tal dependencia local. A continuación, se sintetizan las condiciones que Karunamuni y Alberts imponen a la función g_c para cada c , $0 \leq c \leq 1$:

- (i) $g_c : [0, \infty) \rightarrow [0, \infty)$, g_c es continua y monótona creciente, y existe $g_c^{(i)}$ para cada $i = 1, 2, 3$,
- (ii) $g_c(0) = 0$ y $g_c^{(1)}(0) = 1$,

(iii) la segunda derivada de g_c es

$$g_c^{(2)}(0) = -f^{(1)}(0) \int_{-1}^c u K(u) du / f(0) \int_{-1}^c (c-u) K(u) du.$$

Bajo las condiciones sobre g_c , (i), (ii) y (iii), Karunamuni y Alberts construyen e implementan la siguiente función de transformación para definir su estimador frontera

$$g_c(y) = y + \frac{d}{2} l_c y^2 + d^2 l_c^2 y^3, \quad (3.5)$$

donde

$$l_c = - \int_{-1}^c u K(u) du / \int_{-1}^c (c-u) K(u) du \quad (3.6)$$

y

$$d = f^{(1)}(0)/f(0). \quad (3.7)$$

Para la aplicación práctica de la transformación (3.5), Karunamuni y Alberts implementan la idea de Zhang, Karunamuni y Jones [19], y reemplazan (3.7) por la estimación piloto de tipo núcleo definida por

$$\hat{d} = \left(\log(f_n^*(h_1)) - \log(f_n^*(0)) \right) / h_1, \quad (3.8)$$

donde

$$f_n^*(h_1) = \frac{1}{n h_1} \sum_{i=1}^n K\left(\frac{h_1 - X_i}{h_1}\right) + \frac{1}{n^2}$$

y

$$f_n^*(0) = \max \left\{ \frac{1}{n h_0} \sum_{i=1}^n K_{(0)}\left(\frac{-X_i}{h_0}\right), \frac{1}{n^2} \right\},$$

con $h_1 = o(h_n)$, h_n y K son dadas en (3.1), y $K_{(0)}$ es el llamado núcleo de orden dos con punto extremo inferior, el cual satisface las siguientes condiciones

$$\int_{-1}^0 K_{(0)}(u) du = 1, \quad \int_{-1}^0 u K_{(0)}(u) du = 0, \quad y \quad 0 < \int_{-1}^0 u^2 K_{(0)}(u) du < \infty,$$

y $h_0 = b(0) h_1$, con $b(0)$ dado por

$$b(0) = \left[\frac{\left(\int_{-1}^1 u^2 K(u) du \right)^2 \left(\int_{-1}^0 K_{(0)}^2(u) du \right)}{\left(\int_{-1}^0 u^2 K_{(0)}(u) du \right)^2 \left(\int_{-1}^1 K^2(u) du \right)} \right]^{1/5}. \quad (3.9)$$

Para cada $c \in [0, 1]$, Karunamuni y Alberts definen el estimador de la función de transformación (3.5) como

$$\hat{g}_c(y) = y + \frac{\hat{d}}{2} l_c y^2 + \hat{d}^2 l_c^2 y^3, \quad (3.10)$$

donde \hat{d} viene dado por (3.8). Ellos destacaron que \hat{g}_c y \hat{d} dependen de n , tal dependencia no la resaltaron para simplificar la notación. Finalmente, Karunamuni y Alberts definen el estimador frontera de f con núcleo localmente adaptable de la siguiente manera

$$\tilde{f}_n(r) = \frac{1}{n h_n} \sum_{i=1}^n K\left(\frac{r - \hat{g}_c(X_i)}{h_n}\right) \Big/ \int_{-1}^c K(u) du, \tag{3.11}$$

donde $r \in [0, \infty)$, $0 \leq c \leq 1$, \hat{g}_c viene dada por (3.10), h_n y K como en (3.1). Observe que en cada punto interior $r \in [h_n, \infty)$, $c = 1$, se tiene que $\hat{g}_1(X_i) = X_i$, $i = 1, \dots, n$. Es decir, en cada punto interior el estimador (3.11) se reduce al estimador (1.1). Como la propiedad anterior se cumple en particular para h_n y tomando en cuenta que la región frontera del estimador (3.11) es $[0, h_n)$, tal comportamiento lo interpretaron de la siguiente manera: el estimador (3.11) es una continuación frontera natural del estimador (1.1). Además, las expresiones para el sesgo y la varianza del estimador (3.11) que obtuvieron, se describen en el siguiente teorema.

Teorema 3.2 (Teorema 2.1 en [12]). *Sea $\tilde{f}_n(r)$ definida por (3.11) con función de núcleo K dada en (3.1) y con parámetro de suavizado $h_n = O(n^{-1/5})$. Suponga que h_1 en (3.8) es de la forma $h_1 = O(n^{-1/4})$. Además, asuma que $K^{(1)}$ existe y es continua en $[-1, 1]$, $f(0) > 0$, y que existe $f^{(2)}$ y es continua en una vecindad de 0. Entonces para $r = c h_n$, $0 \leq c \leq 1$, se tiene que*

$$\begin{aligned} \mathbb{E}[\tilde{f}_n(r) - f(r)] &= \frac{h_n^2}{2 \int_{-1}^c K(u) du} \left\{ f^{(2)}(0) \int_{-1}^c (u^2 - 2uc) K(u) du \right. \\ &\quad \left. + 6 \frac{(f^{(1)}(0))^2}{f(0)} (l_c^2 + l_c) \int_{-1}^c (u - c)^2 K(u) du \right\} + o(h_n^2), \end{aligned} \tag{3.12}$$

donde l_c es dada en (3.6), y

$$\text{Var}[\tilde{f}_n(r)] = \frac{f(0)}{n h_n \left(\int_{-1}^c K(u) du\right)^2} \int_{-1}^c K^2(u) du + o\left(\frac{1}{n h_n}\right)$$

De la expresión (3.12) se desprende que el estimador (3.11) no presenta el problema frontera, ya que su sesgo tiene una tasa de convergencia de orden $O(h_n^2)$ en los puntos fronteras $r \in [0, h_n)$ y en el punto interior h_n . Los autores Karunamuni y Alberts atribuyen ese importante ajuste en el estimador (3.11) a la transformación adaptativa local (3.10), ya que transforma los datos dependiendo del punto de estimación.

3.2 Estimador Frontera con Conjunto Difuso

Sean X_1, \dots, X_n una muestra aleatoria independiente de la variable aleatoria X con densidad f , y V_1, \dots, V_n una muestra aleatoria independiente uniformemente distribuida en $[0, 1]$ e independiente de X_1, \dots, X_n . Para ambas muestras, Fajardo y Harmath [5] inician la construcción de su estimador frontera imponiendo las siguientes condiciones:

- C1** La función de densidad f tiene soporte $[0, \infty)$, y es al menos dos veces continuamente diferenciable en una vecindad de $z \in [0, \infty)$.

C2 El parámetro b_n satisface: $b_n \rightarrow 0$ y $nb_n \rightarrow \infty$, cuando $n \rightarrow \infty$.

C3 La función atenuante φ es simétrica con respecto al cero, tiene soporte compacto $[-B, B]$, $B > 0$, $\varphi(u) \in [0, 1]$ y es continua en 0 con $\varphi(0) > 0$.

Seguidamente los autores establecen los siguientes resultados, en los cuales describen el comportamiento del estimador (1.2) en los puntos 0 y $x \in (0, b_n]$. Además, para simplificar la notación definen

$$\psi(u) = \frac{\varphi(u)}{\int \varphi(u) du} \quad \text{y} \quad x = s b_n, \quad 0 < s \leq 1.$$

Teorema 3.3 (Teoremas 1 y 2 en [5]). *Bajo las condiciones (C1)-(C3), se tiene que*

$$\mathbb{E} \left[\hat{\vartheta}_n(0) - f(0) \right] = \frac{b_n^2}{2} f^{(2)}(0) \int u^2 \psi(u) du + o(b_n^2).$$

y

$$\begin{aligned} \mathbb{E} \left[\hat{\vartheta}_n(x) - f(0) \right] &= -f(0) + f(0) \int_{-s}^B \psi(u) du + b_n f^{(1)}(0) \int_{-s}^B u \psi(u) du \\ &+ \frac{b_n^2}{2} f^{(2)}(0) \int_{-s}^B u^2 \psi(u) du + o(b_n^2), \end{aligned} \quad (3.13)$$

donde $0 < s \leq 1$.

A través de los resultados anteriores, Fajardo y Harmath garantizaron que el estimador (1.2) no presenta el problema frontera en el punto interior 0, ya que su sesgo tiene una tasa de convergencia de orden $O(b_n^2)$ en 0. En cambio, con la sutil modificación introducida en (3.13) con respecto a la versión original, el lector podrá apreciar, con mayor facilidad, que el estimador (1.2) presenta el problema frontera en cada $x \in (0, b_n]$. En efecto, su sesgo tiene una tasa de convergencia de orden $O(b_n)$ en cada x .

El siguiente paso que Fajardo y Harmath tomaron para definir su estimador frontera, fue construir una función atenuante φ con la siguiente propiedad:

$$\int_{-s}^B u \varphi(u) du = 0, \quad \text{para cada } s \in (0, 1]. \quad (3.14)$$

Para tal fin, formalizan la Observación 4 introducida por Fajardo [4] reescribiéndola como el próximo teorema, con el cual controlarán adecuadamente las constantes que definen el sesgo del estimador (1.2) y justificarán una condición en el criterio que les permitirá obtener una función atenuante φ que satisface la condición (3.14).

Teorema 3.4 (Teorema 3 en [4]). *Bajo la condición (C3), se tiene que para $\mathbf{M} > 0$ existe $B' > 0$ tal que*

$$v = \int_{-B'}^{B'} u^2 \psi(u) du \leq \mathbf{M}.$$

La combinación de (C3) y el Teorema 3.4, les permitió garantizar que

$$v = \int_{-B'}^{B'} u^2 \psi(u) du = \int u^2 \psi(u) du \leq \mathbf{M}, \text{ con } B' > B,$$

y así redefinir ψ como

$$\psi(u) = \frac{\varphi(u)}{\int \varphi(u) du} \mathbb{I}_{[-B', B']}(u), \text{ con } B' \leq B. \tag{3.15}$$

El criterio propuesto por Fajardo y Harmath para eliminar el término con coeficiente b_n en (3.13), haciendo que $\int_{-s}^B u \varphi(u) du = 0$ para cada $s \in (0, 1]$, se basó en deducir una función atenuante φ como solución del siguiente problema variacional:

$$\begin{aligned} &\text{Maximizar : } \int \varphi(u) du. \\ &\text{Sujeto a : } \int \varphi^2(u) du = k, \int u \varphi(u) du = 0, \int (u^2 - v) \varphi(u) du = 0, k > 0, \text{ (PV)} \\ &\varphi(u) = 0 \text{ para } u \in (-B, B)^c, \varphi(0) > 0, \varphi(u) \in [0, 1], \text{ y } v \in (0, \mathbf{M}]. \end{aligned}$$

Es importante puntualizar que el criterio anterior generaliza el criterio propuesto por Fajardo [4], el cual implementó para obtener una función φ que minimiza el $ECMI^*$ del estimador (1.2) (ver Fajardo [4], página 307). El siguiente teorema garantiza la solución de (PV).

Teorema 3.5 (Teorema 4 en [5]). *La solución de (PV) viene dado por*

$$\varphi_k(u) = \left[1 - \left(\frac{16}{15k} \right)^2 u^2 \right] \mathbb{I}_{[-\frac{15}{16}k, \frac{15}{16}k]}(u), \quad k > 0. \tag{20}$$

En particular, para $s \in (0, 1]$ se tiene que

$$\varphi_s(u) = \left[1 - \left(\frac{16}{15s} \right)^2 u^2 \right] \mathbb{I}_{[-\frac{15}{16}s, \frac{15}{16}s]}(u). \tag{21}$$

Las siguientes observaciones fueron establecidas por Fajardo y Harmath:

- A partir de (1.2) c.f.a. φ_s y (3.13), se tiene que

$$\mathbb{E} \left[\hat{\vartheta}_n(x) - f(0) \right] = \frac{b_n^2}{2} f^{(2)}(0) \int_{-B'}^{B'} u^2 \psi_s(u) du + o(b_n^2), \tag{22}$$

donde $0 < s \leq 1$, $B' \leq \frac{15}{16} s$, ψ_s viene dada por (3.15) c.f.a. φ_s y φ_s viene dada por (21). Así, el estimador (1.2) no presenta el problema frontera en x cuando la función atenuante es φ_s .

- Combinando el Teorema 3.3 con el Teorema 4 en Fajardo [4], se tiene que

$$\mathbb{E} \left[\hat{\vartheta}_n(z) - f(z) \right] = \frac{b_n^2}{2} f^{(2)}(z) \int_{-B'}^{B'} u^2 \psi_k(u) du + o(b_n^2), \tag{23}$$

para cada $k > 1$ y $z \in \{0\} \cup (b_n, \infty)$, donde $B' \leq \frac{15}{16} k$, ψ_k viene dada por (3.15) c.f.a. φ_k , y φ_k viene dada por (20). Así, el estimador (1.2) no presenta el problema frontera en $z \in \{0\} \cup (b_n, \infty)$ cuando la función atenuante es φ_k , para cada $k > 1$.

- Denotando el núcleo Epanechnikov por K_E , y sustituyendo en (PV) k por $\frac{5}{3}$ y \mathbf{M} por $\mathbf{M}_E = \int u^2 K_E(u) du$, se tiene que el estimador (1.2) c.f.a. $\varphi_{\frac{5}{3}}$ es el estimador propuesto por Fajardo [4]. Además, el resultado alcanzado en Fajardo [4] permite garantizar que la función atenuante φ_k minimiza $ECM[\hat{\vartheta}_n(z)]$, para cada $k > 1$ y cada $z \in \{0\} \cup (b_n, \infty)$.

Apoyados en (22) y (23), Fajardo y Harmath definen su estimador frontera de f con conjunto difuso de la siguiente manera

$$\tilde{\vartheta}_n(x) = \frac{1}{n a_n} \sum_{i=1}^n \mathbb{I}_{[0, \varphi_s(\frac{x_i - x}{b_n})]}(V_i), \quad (24)$$

donde $0 < s \leq 1$, $a_n = b_n \int \varphi_s(u) du$, y φ_s viene dada por (21). Ellos señalaron que, para $z \geq b_n$, $s = 1$, el estimador (24) se reduce al estimador (1.2) c.f.a. φ_1 . Así, el estimador (24) es la continuación frontera natural del estimador (1.2) c.f.a. φ_1 . Además, resaltaron que los resultados obtenidos por Fajardo [4] permiten garantizar que la función atenuante φ_s minimiza localmente a $ECM[\tilde{\vartheta}_n(x)]$ para cada $s \in (0, 1]$. Por otro lado, las expresiones para el sesgo y la varianza del estimador (24) que obtuvieron, se describen en el siguiente teorema.

Teorema 3.6 (Lema 1 en [5]). *Bajo las condiciones (C1)-(C3), se tiene que*

$$\mathbb{E}[\tilde{\vartheta}_n(x) - f(0)] = \frac{b_n^2}{2} f^{(2)}(0) \int_{-B'}^{B'} u^2 \psi_s(u) du + o(b_n^2) \quad (25)$$

y

$$Var[\tilde{\vartheta}_n(x)] = \frac{f(0)}{n b_n} \left(\int \varphi_s(u) du \right)^{-1} + o\left(\frac{1}{n b_n}\right),$$

donde $0 < s \leq 1$, $0 < B' \leq \frac{15}{16}s$, ψ_s viene dada por (3.15) c.f.a. φ_s y φ_s viene dada por (21).

Finalmente señalaron que (25) garantiza que el estimador (3.11) no presenta el problema frontera, ya que su sesgo tiene una tasa de convergencia de orden $O(b_n^2)$, en los puntos fronteras $x \in (0, b_n)$ y en el punto interior b_n .

4 Análisis de Datos Simulados

En esta sección se presentan los resultados sobre los datos simulados, los cuales fueron diseñados para comparar el desempeño entre los estimadores $\tilde{f}_n(t)$ y $\tilde{\vartheta}_n(x)$ en puntos cercanos a 0 en una dispersión de la vecindad de b_n .

Cada estimador se probó usando cuatro densidades específicas, con comportamiento variado en 0. La densidad 1 trata el caso $f(0) = 0$, mientras que las densidades 2, 3 y 4 describen lo que sucede cuando $f(0) > 0$ pero $f'(0) = 0$, $f'(0) > 0$ y $f'(0) < 0$, respectivamente. Además, se implementaron los siguientes anchos de banda globales óptimos como los parámetros de suavizados de $\tilde{f}_n(t)$ y $\tilde{\vartheta}_n(x)$, respectivamente:

$$h_n = \left[\frac{\int K_E^2(u) du}{[\int u^2 K_E(u) du]^2 \int [f^{(2)}(u)]^2 du} \right]^{1/5} n^{-1/5} \quad (1)$$

y

$$b_n = \left[\frac{1}{\int \varphi_1(u) du [\int u^2 \psi_1(u) du]^2 \int [f^{(2)}(u)]^2 du} \right]^{1/5} n^{-1/5}. \tag{2}$$

Para más detalles sobre (1) y (2), ver Silverman [16], páginas 39 – 40, y Fajardo [4], respectivamente. Es oportuno destacar que la razón que motivó la implementación de (1) y (2) como los parámetros de suavizados se debió al hecho que, las comparaciones basadas en los parámetros de suavizados óptimos son más convincentes que las comparaciones basadas en los parámetros de suavizados aproximados, ya que debido a la calidad o no del método de selección del parámetro de suavizado estas podrían ser engañosas. Además, se hace una elección del parámetro de suavizado global en lugar de uno local, porque es mucho más probable que se utilicen en aplicaciones.

En todas las simulaciones, se utilizó una muestra de tamaño $n = 200$. Además, para cada densidad se calculó el sesgo (valor estimado menos el valor verdadero), la varianza y *ECM* de ambos estimadores en los puntos $r = ch_n$, $0 \leq c \leq 1$, y $x = sb_n$, $0 < s \leq 1$, donde $c \in \{\frac{k}{25} : k = 1, 4, 7, 9, 12\}$, $s = (h_n/b_n)c$, y los parámetros h_n y b_n son dados por (1) y (2), respectivamente. Asimismo, para estimar d se eligió, como en Karunamuni y Alberts [12], $h_1 = h_n n^{-1/20}$ y $K_{(0)}(t) = 12(1+t)(t + \frac{1}{2}) \mathbb{I}_{[-1,0]}(t)$.

Todos los resultados se promediaron a través de 1000 pruebas y se muestran en las Tablas 1 a 4. Simultáneamente, basado en los resultados obtenidos, también se representan gráficamente el desempeño de cada estimador, junto con cada densidad f sobre sus respectivas regiones fronteras. Tales desempeños se muestran en las Figuras 1 a 4. Una minuciosa revisión de las Tablas 1 a 4 permite afirmar que, para todas las formas de densidades, el estimador $\tilde{\vartheta}_n$ tiene un mejor desempeño que el estimador \tilde{f}_n , ya que $ECM[\tilde{\vartheta}_n(x)] \leq ECM[\tilde{f}_n(x)]$ para cada $x \in (0, h_n)$. Además, se observa que ambos estimadores mejoran el sesgo mientras que sus varianzas son pequeñas. Sin embargo, la comparación del desempeño de los estimadores \tilde{f}_n y $\tilde{\vartheta}_n$ a través del error cuadrático medio integrado (*ECMI*) no es conveniente, ya que para todas las formas de densidades se tiene que $(0, h_n] \subset (0, b_n]$. Finalmente, es oportuno enfatizar que a pesar de existir una significativa diferencia entre los *ECM* de ambos estimadores en cada una de las Tablas 1 a 4, esta se manifiesta en la Tabla 2 con menor magnitud cuando tales diferencias se comparan

con las r... observar
 en la Fig... erver que
 para \tilde{f}_n 1

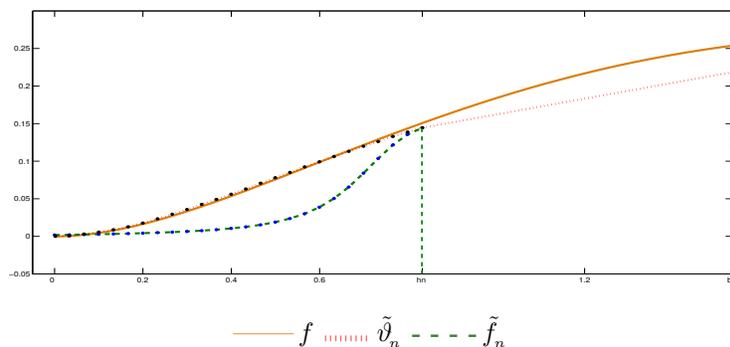


Figura 1: Estimaciones de $f(z) = \frac{1}{5} z^2 e^{-z}$, densidad 1, en la regiones fronteras $(0, h_n]$ y $(0, b_n]$, donde $v = \frac{1}{4} M_E$, $n = 200$, $h_n = 0,832551$ y $b_n = 1,535404$.

Tabla 1: Sesgo, varianza y ECM de las estimaciones indicadas, a través de un promedio de 1000 pruebas en los puntos $r = ch_n$ y $x = sb_n$, $0 \leq c \leq 1$ y $0 < s \leq 1$, para $f(z) = \frac{1}{2} z^2 e^{-z}$, densidad 1, donde $v = \frac{1}{4} \mathbf{M}_E$, $n = 200$, y óptimos globales $h_n = 0,832551$ y $b_n = 1,535404$.

$\tilde{f}_n(r)$				$\tilde{\vartheta}_n(x)$			
c	Sesgo	Var	$ECM \times 10^{-5}$	s	Sesgo	Var	$ECM \times 10^{-5}$
0,0400	0,0015	0,0000	0,2258	0,0217	0,0001	0,0000	0,00031
0,1600	-0,0045	0,0000	2,0659	0,0868	0,0010	0,0000	0,09458
0,2800	-0,0167	0,0000	30,000	0,1518	0,0017	0,0000	0,30495
0,3600	-0,0268	0,0000	70,000	0,1952	0,0024	0,0000	0,59692
0,4800	-0,0430	0,0000	190,00	0,2603	0,0025	0,0000	0,61875

Tabla 2: Sesgo, varianza y ECM de las estimaciones indicadas, a través de un promedio de 1000 pruebas en los puntos $r = ch_n$ y $x = sb_n$, $0 \leq c \leq 1$ y $0 < s \leq 1$, para $f(z) = \sqrt{\frac{2}{\pi}} e^{-\frac{z^2}{2}}$, densidad 2, donde $v = \frac{1}{6} \mathbf{M}_E$, $n = 200$, y óptimos globales $h_n = 0,707481$ y $b_n = 1,534486$.

$\tilde{f}_n(r)$				$\tilde{\vartheta}_n(x)$			
c	Sesgo	Var	$ECM \times 10^{-4}$	s	Sesgo	Var	$ECM \times 10^{-4}$
0,0400	-0,0189	0,0000	4	0,0184	0,0049	0,0000	0,23612
0,1600	-0,0247	0,0000	6	0,0738	-0,0027	0,0000	0,07275
0,2800	-0,0287	0,0000	8	0,1291	-0,0026	0,0000	0,06517
0,3600	-0,0304	0,0000	9	0,1660	-0,0040	0,0000	0,15860
0,4800	-0,0316	0,0000	10	0,2213	-0,0089	0,0000	0,79334

5 Análisis de Datos Reales

En esta sección se pondrán a prueba los estimadores \tilde{f}_n y $\tilde{\vartheta}_n$ en dos conjuntos de datos reales conocidos, con el propósito de mostrar su utilidad práctica. Sólo para la estimación de la densidad con $\tilde{\vartheta}_n$ se realizaron 1000 pruebas, donde para cada muestra fija X_1, \dots, X_n se tomó una muestra independiente $V_1^{(d)}, \dots, V_n^{(d)}$, $1 \leq d \leq 1000$, distribuida uniformemente en $[0, 1]$. Por otro lado, el parámetro h_n utilizado fue implementado por otros autores en sus simulaciones, para cada conjunto de datos propuesto. No obstante, cada b_n se obtendrá combinando (2) con el valor aproximado de $\int [f''(u)]^2 du$, valor que se calculará a través de (1) para cada h_n considerado. Es oportuno señalar que, h_n y b_n se reflejarán en cada gráfica y haciendo uso de la propiedad de “continuación frontera natural” de los estimadores \tilde{f}_n y $\tilde{\vartheta}_n$, se “abusará” de la notación para mostrar sólo dos curvas en la representación gráfica asociada a cada conjunto de datos, teniendo en cuenta que a la derecha de h_n y b_n las gráficas mostradas representan las gráficas de (1.1), y (1.2) c.f.a. φ_1 , respectivamente. En concreto, se identificará la estimación con núcleo y con conjunto difuso con líneas discontinuas - - - - y -.-.-.-., respectivamente.

Tabla 3: Sesgo, varianza y *ECM* de las estimaciones indicadas, a través de un promedio de 1000 pruebas en los puntos $r = ch_n$ y $x = sb_n$, $0 \leq c \leq 1$ y $0 < s \leq 1$, para $f(z) = (z^2 + 2z + \frac{1}{2}) e^{-2z}$, densidad 3, donde $v = \frac{1}{4} \mathbf{M}_E$, $n = 200$, y óptimos globales $h_n = 0,467044$ y $b_n = 0,861329$.

$\tilde{f}_n(r)$				$\tilde{\vartheta}_n(x)$			
c	Sesgo	Var	$ECM \times 10^{-5}$	s	Sesgo	Var	$ECM \times 10^{-5}$
0,0400	0,0393	0,0000	154,2479	0,0217	0,0014	0,0000	0,2088
0,1600	0,0065	0,0000	4,1782	0,0868	0,0003	0,0000	0,0088
0,2800	-0,0159	0,0000	25,2148	0,1518	-0,0061	0,0000	3,6991
0,3600	-0,0259	0,0000	67,1897	0,1952	-0,0079	0,0000	6,2911
0,4800	-0,0351	0,0000	123,0843	0,2603	-0,0118	0,0000	13,9596

Tabla 4: Sesgo, varianza y *ECM* de las estimaciones indicadas, a través de un promedio de 1000 pruebas en los puntos $r = ch_n$ y $x = sb_n$, $0 \leq c \leq 1$ y $0 < s \leq 1$, para $f(z) = 2e^{-2z}$, densidad 4, donde $v = \frac{1}{2} \mathbf{M}_E$, $n = 200$, y óptimos globales $h_n = 0,342128$ y $b_n = 0,478176$.

$\tilde{f}_n(r)$				$\tilde{\vartheta}_n(x)$			
c	Sesgo	Var	$ECM \times 10^{-5}$	s	Sesgo	Var	$ECM \times 10^{-5}$
0,0400	-0,3717	0,0000	13810	0,0286	0,0132	0,0000	17,4451
0,1600	-0,2494	0,0000	6220	0,1145	-0,0016	0,0000	0,2562
0,2800	-0,1522	0,0000	2320	0,2003	0,0005	0,0000	0,0227
0,3600	-0,1003	0,0000	1010	0,2576	0,0050	0,0000	2,4650
0,4800	-0,0409	0,0000	170	0,3434	0,0093	0,0000	8,6437

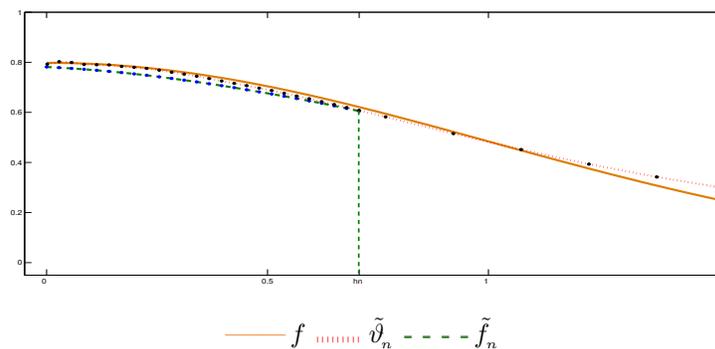


Figura 2: Estimaciones de $f(z) = \sqrt{\frac{2}{\pi}} e^{-\frac{z^2}{2}}$, densidad 2, en la regiones fronteras $(0, h_n]$ y $(0, b_n]$, donde $v = \frac{1}{6} \mathbf{M}_E$, $n = 200$, $h_n = 0,707481$ y $b_n = 1,534486$.

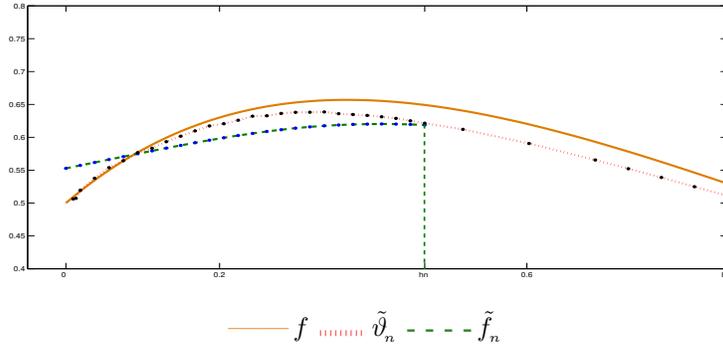


Figura 3: Estimaciones de $f(z) = (z^2 + 2z + \frac{1}{2}) e^{-2x}$, densidad 3, en la regiones fronteras $(0, h_n]$ y $(0, b_n]$, donde $v = \frac{1}{4} \mathbf{M}_E$, $n = 200$, $h_n = 0,467044$ y $b_n = 0,861329$.

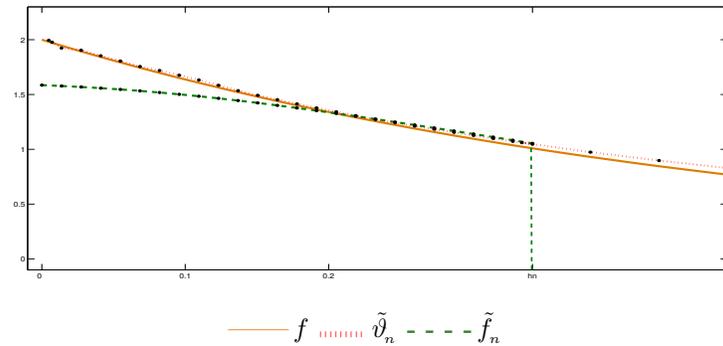


Figura 4: Estimaciones de $f(z) = 2e^{-2x}$, densidad 4, en la regiones fronteras $(0, h_n]$ y $(0, b_n]$, donde $v = \frac{1}{2} \mathbf{M}_E$, $n = 200$, $h_n = 0,342128$ y $b_n = 0,478176$.

El primer conjunto de datos reales consta de 35 mediciones correspondientes a la precipitación promedio durante el mes de diciembre en Des Moines-Iowa-EUA, desde 1961 hasta 1995. Para este conjunto de datos, Karunamuni y Alberts [12] usan $h_n = 0,4250$, obtenido a través de validación cruzada (ver Bowman y Azzalini [1]), el cual permite obtener un valor para $b_n = 0,4502$. En la Figura 5 se puede apreciar una marcada similitud entre la estimación con núcleo y con conjunto difuso en los puntos interiores y en la mayor parte de los puntos fronteras. En cambio, \tilde{f}_n no reconoce que la densidad es cero para valores muy cercanos a 0 por su derecha, tal cual lo señala el histograma de los datos al pie de la gráfica. Mientras que el estimador \tilde{v}_n si lo reconoce sin importarle la curvatura.

El segundo conjunto de datos reales son los famosos datos sobre suicidios que se encuentran en Silverman [16], Tabla 2.1. Estos corresponden a los períodos de duración de 86 tratamientos psiquiátricos aplicado a los pacientes utilizados como controles en un estudio sobre riesgos de suicidio, y es un ejemplo clásico de un conjunto de datos donde el estimador (1.1) falla en la región frontera. También se mostró, en el reciente trabajo de Fajardo y Harmath, que el estimador (1.2) c.f.a. φ_1 falla en la región frontera. Para este conjunto de datos, Karunamuni y Alberts [11] eligen subjetivamente $h_n = 60$, el cual permite obtener un valor para $b_n = 83,8593$. La Figura 6

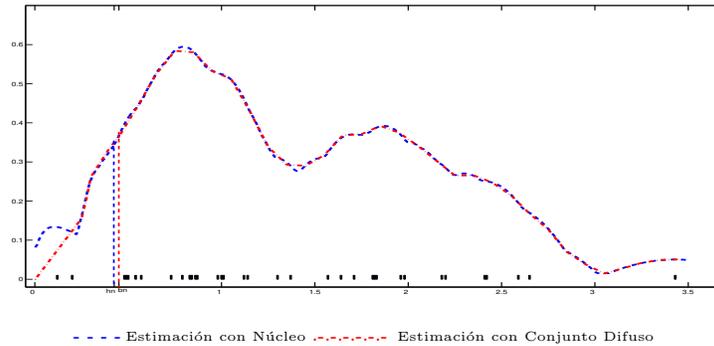


Figura 5: Estimaciones de las densidades para 35 mediciones correspondientes a la precipitación promedio en el mes de diciembre en Des Moines-Iowa-EUA desde 1961 hasta 1965, mostradas en el piso de la gráfica, donde $v = \mathbf{M}_E$, $h_n = 0,4250$, $b_n = 0,4502$ y $h_1 = h_n n^{-1/20}$.

muestra el comportamiento de las estimaciones con núcleo y con conjunto difuso. Se aprecia que el valor de la densidad en 0 es capturado de forma efectiva por el estimador $\tilde{\vartheta}_n$ con una apreciable curvatura, tomando en cuenta que la densidad es 0 como lo indica el histograma al pie de la gráfica. No obstante, el estimador \tilde{f}_n no reconoce el comportamiento anterior de la densidad y decide truncarla en 0 asignándole el valor 0,008.

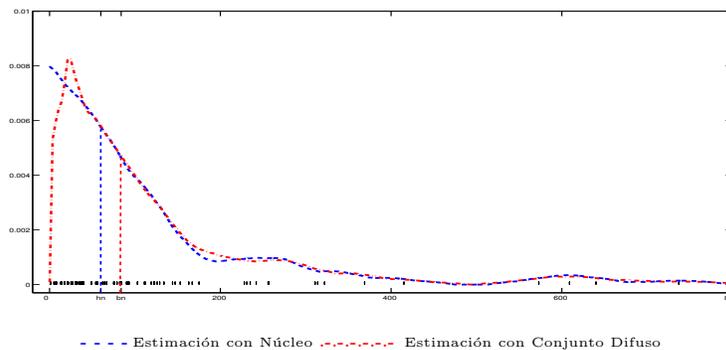


Figura 6: Estimaciones de las densidades para 86 mediciones correspondientes a los tratamientos psiquiátricos aplicado a los pacientes utilizados como controles en un estudio sobre riesgos de suicidio, mostradas en el piso de la gráfica, donde $v = \frac{1}{2} \mathbf{M}_E$, $h_n = 60$, $b_n = 83,8593$ y $h_1 = h_n n^{-1/20}$.

6 Conclusiones

Es claro que ningún estimador existente en la literatura domina al resto de los estimadores para todas las formas de densidades. Además, cada estimador tiene ciertas ventajas y funciona bien en determinados momentos. Sin embargo, para las cuatro formas de densidades consideradas en estas notas el estimador frontera de la densidad con conjunto difuso presentó un rendimiento superior con respecto al rendimiento del estimador frontera de la densidad con núcleo localmente

adaptable, en puntos cercanos a 0 en una dispersión de la vecindad del parámetro de suavizado b_n . El resultado anterior fue producto de las extensas simulaciones realizadas para comparar los errores cuadráticos medios locales de los estimadores fronteras anteriores y forma parte de una nueva contribución en el área de estimación de la función de densidad no basada en núcleo. Un resultado similar fue presentado en Fajardo y Harmath, donde bajo condiciones análogas a las anteriores el estimador frontera con conjunto difuso presentó un rendimiento superior con respecto al rendimiento del particular estimador frontera con núcleo introducido por Karunamuni y Alberts [11]. Cabe destacar que el estimador frontera con conjunto difuso y los estimadores fronteras con núcleos considerados en estas notas como en Fajardo y Harmath, presentan estructuras totalmente diferentes pero comparten propiedades comunes como: no negatividad, “continuación frontera natural”, y un desempeño robusto con respecto a diversas formas de densidades, ya que permiten importantes reducciones del sesgo mientras que sus varianzas son pequeñas. Simultáneamente, es oportuno resaltar que tales propiedades se obtienen como consecuencia de la influencia de las funciones atenuante, φ_k , y de transformación, \hat{g}_c , en cada estimador. Finalmente, es importante recalcar que a través del proceso puntual atenuado que describe el método de estimación de densidad con conjunto difuso, el conjunto de observaciones considerada pueden describirse en una vecindad del punto de estimación, donde las funciones indicadoras que definen al estimador de densidad con conjunto difuso deciden si la observación pertenece o no a la vecindad del punto de estimación, y las funciones atenuantes que definen a cada función indicadora se utilizan para seleccionar los puntos de la muestra con diferentes probabilidades, en contraste con los estimadores con núcleo que asignan el mismo peso a todos los puntos de la muestra.

7 Agradecimientos

El autor agradece a los árbitros anónimos por sus oportunos comentarios y útiles sugerencias, que mejoraron en gran medida la presentación de los resultados. El autor también desea agradecer al Licenciado Pedro Luis Bossio, Jefe del Departamento de Admisión y Control de Estudios del Núcleo Sucre de la Universidad de Oriente, por su apoyo con el equipo de computación, el cual fue utilizado para la transcripción de estas notas y la realización de las simulaciones presentadas.

Referencias

- [1] Bowman, A.W., Azzalini, A. *Applied smoothing techniques for data analysis: the kernel approach with S-plus illustrations*, Oxford University Press, (1997)
- [2] Cowling, A., Hall, P. *On pseudodata methods for removing boundary effects in kernel density estimation*, J. Roy. Statist. Soc. Ser. B, **58** (1996), 551 – 563.
- [3] Epanechnikov, V. A. *Nonparametric estimates of a multivariate probability density*, Theory Prob. Appl., **14** (1969), 153 – 158.
- [4] Fajardo, J. *A criterion for the fuzzy set estimation of the density function*, Braz. J. Prob. Stat., **28**(3) (2014), 301 – 312.
- [5] Fajardo, J., Harmath, P. *Boundary estimation with the fuzzy set density estimator*, METRON, **79**(3) (2021), 285 – 302.

- [6] Fajardo, J., Ríos, R., Rodríguez, L. *Properties of convergence of a fuzzy set estimator of the density function*, Braz. J. Prob. Stat., **26**(2) (2012), 208 – 217.
- [7] Falk, M., Liese, F. *Lan of thinned empirical processes with an application to fuzzy set density estimation*, Extremes, **1**(3) (1998), 323 – 349.
- [8] Hall, P., Park, B. U. *New methods for bias correction at endpoints and boundaries*, The Ann. Statist. **30** (2002), 1460 – 1479.
- [9] Hominal, P., Deheuvels, P. *Estimation non paramétrique de la densité compte-tenu d'informations sur le support*, Rev. Stat. Appl., **27**(3) (1979), 47 – 68.
- [10] Jones, M.C., Foster, P. J. *A simple nonnegative boundary correction method for kernel density estimation*, Stat. Sin., **6** (1996), 1005 – 1013.
- [11] Karunamuni, R. J., Alberts, T. *On boundary correction in kernel density estimation*, Stat. Methodol., **2** (3) (2005), 191 – 212.
- [12] Karunamuni, R. J., Alberts, T. *A locally adaptive transformation method of boundary correction in kernel density estimation*, J. Stat. Plan. Infer., **136** (2006), 2936 – 2960.
- [13] Parzen, E. *On estimation of a probability density function and mode*, Ann. Math. Statist., **33**(3) (1962), 1065 – 1076.
- [14] Reiss, R. D. *A course on point processes*, Springer Series in Statistics. New York, 1993.
- [15] Rosenblatt, M. *Remarks on some nonparametric estimates of a density function*, Anna. Math. Statist., **27**(3) (1956), 832 – 837.
- [16] Silverman, B. W. *Density estimation for statistics and data analysis*, Chapman and Hall. New York, 1986.
- [17] Stone, C. J. *Optimal rates of convergence for nonparametric estimators*, Ann. Statist., **8** (1980), 1348 – 1360.
- [18] Zhang, S., Karunamuni, R.J. *On kernel density estimation near endpoints*, J. Stat. Plan. Infer., **70** (1998), 301 – 316.
- [19] Zhang, S., Karunamuni, R. J., Jones, M. C. *An improved estimator of the density function at the boundary*, J. Amer. Statist. Assoc. **448** (1999), 1231 – 1241.