Hindawi Publishing Corporation
Journal of Probability and Statistics
Volume 2012, Article ID 593036, 18 pages
doi:10.1155/2012/593036

Research Article

A Criterion for the Fuzzy Set Estimation of the Regression Function

Jesús A. Fajardo

Departamento de Matemáticas, Universidad de Oriente, Cumaná 6101, Venezuela

Correspondence should be addressed to Jesús A. Fajardo, jfajardogonzalez@gmail.com

Received 1 May 2012; Accepted 30 June 2012

Academic Editor: A. Thavaneswaran

We propose a criterion to estimate the regression function by means of a nonparametric and fuzzy set estimator of the Nadaraya-Watson type, for independent pairs of data, obtaining a reduction of the integrated mean square error of the fuzzy set estimator regarding the integrated mean square error of the classic kernel estimators. This reduction shows that the fuzzy set estimator has better performance than the kernel estimations. Also, the convergence rate of the optimal scaling factor is computed, which coincides with the convergence rate in classic kernel estimation. Finally, these theoretical findings are illustrated using a numerical example.

1. Introduction

The methods of kernel estimation are among the nonparametric methods commonly used to estimate the regression function r, with independent pairs of data. Nevertheless, through the theory of point processes (see e.g, Reiss [1]) we can obtain a new nonparametric estimation method, which is based on defining a nonparametric estimator of the Nadaraya-Watson type regression function, for independent pairs of data, by means of a fuzzy set estimator of the density function. The method of fuzzy set estimation introduced by Falk and Liese [2] is based on defining a fuzzy set estimator of the density function by means of thinned point processes (see e.g, Reiss [1], Section 2.4); a process framed inside the theory of the point processes, which is given by the following:

$$\widehat{\theta}_n = \frac{1}{na_n}\sum_{i=1}^{n} U_i, \tag{1.1}$$

where $a_n > 0$ is a scaling factor (or bandwidth) such that $a_n \to 0$ as $n \to \infty$, and the random variables U_i, $1 \leq i \leq n$, are independent with values in $\{0,1\}$, which decides whether X_i belongs to the neighborhood of x_0 or not. Here x_0 is the point of estimation (for more details, see Falk and Liese [2]). On the other hand, we observe that the random variables that define the estimator $\widehat{\theta}_n$ do not possess, for example, precise functional characteristics in regards to the point of estimation. This absence of functional characteristics complicates the evaluation of the estimator $\widehat{\theta}_n$ using a sample, as well as the evaluation of the fuzzy set estimator of the regression function if it is defined in terms of $\widehat{\theta}_n$.

The method of fuzzy set estimation of the regression function introduced by Fajardo et al. [3] is based on defining a fuzzy set estimator of the Nadaraya-Watson type, for independent pairs of data, in terms of the fuzzy set estimator of the density function introduced in Fajardo et al. [4]. Moreover, the regression function is estimated by means of an average fuzzy set estimator considering pairs of fixed data, which is a particular case if we consider independent pairs of nonfixed data. Note that the statements made in Section 4 in Fajardo et al. [3] are satisfied if independent pairs of nonfixed data are considered. This last observation is omitted in Fajardo et al. [3]. It is important to emphasize that the fuzzy set estimator introduced in Fajardo et al. [4], a particular case of the estimator introduced by Falk and Liese [2], of easy practical implementation, will allow us to overcome the difficulties presented by the estimator $\widehat{\theta}_n$ and satisfy the almost sure, in law, and uniform convergence properties over compact subsets on $\mathbb{R}$.

In this paper we estimate the regression function by means of the nonparametric and fuzzy set estimator of the Nadaraya-Watson type, for independent pairs of data, introduced by Fajardo et al. [3], obtaining a significant reduction of the integrated mean square error of the fuzzy set estimator regarding the integrated mean square error of the classic kernel estimators. This reduction is obtained by the conditions imposed on the thinning function, a function that allows to define the estimator proposed by Fajardo et al. [4], which implies that the fuzzy set estimator has better performance than the kernel estimations. The above reduction is not obtained in Fajardo et al. [3]. Also, the convergence rate of the optimal scaling factor is computed, which coincides with the convergence rate in classic kernel estimation of the regression function. Moreover, the function that minimizes the integrated mean square error of the fuzzy set estimator is obtained. Finally, these theoretical findings are illustrated using a numerical example estimating a regression function with the fuzzy set estimator and the classic kernel estimators.

On the other hand, it is important to emphasize that, along with the reduction of the integrated mean square error, the thinning function, introduced through the thinned point processes, can be used to select points of the sample with different probabilities, in contrast to the kernel estimator, which assigns equal weight to all points of the sample.

This paper is organized as follows. In Section 2, we define the fuzzy set estimator of the regression function and we present its properties of convergence. In Section 3, we obtain the mean square error of the fuzzy set estimator of the regression function, Theorem 3.1, as well as the optimal scale factor and the integrated mean square error. Moreover, we establish the conditions to obtain a reduction of the constants that control the bias and the asymptotic variance regarding the classic kernel estimators; the function that minimizes the integrated mean square error of the fuzzy set estimator is also obtained. In Section 4 a simulation study was conducted to compare the performances of the fuzzy set estimator with the classical Nadaraya-Watson estimators. Section 5 contains the proof of the theorem in the Section 3.

2. Fuzzy Set Estimator of the Regression Function and Its Convergence Properties

In this section we define by means of fuzzy set estimator of the density function introduced in Fajardo et al. [4] a nonparametric and fuzzy set estimator of the regression function of Nadaraya-Watson type for independent pairs of data. Moreover, we present its properties of convergence.

Next, we present the fuzzy set estimator of the density function introduced by Fajardo et al. [4], which is a particular case of the estimator proposed in Falk and Liese [2] and satisfies the almost sure, in law, and uniform convergence properties over compact subset on $\mathbb{R}$.

Definition 2.1. Let $X_1, \ldots, X_n$ be an independent random sample of a real random variable X with density function f. Let $V_1, \ldots, V_n$ be independent random variables uniformly on $[0,1]$ distributed and independent of $X_1, \ldots, X_n$. Let φ be such that $0 < \int \varphi(x)dx < \infty$ and $a_n = b_n \int \varphi(x)dx$, $b_n > 0$. Then the fuzzy set estimator of the density function f at the point $x_0 \in \mathbb{R}$ is defined as follows:

$$\widehat{\vartheta}_n(x_0) = \frac{1}{na_n} \sum_{i=1}^{n} U_{x_0,b_n}(X_i, V_i) = \frac{\tau_n(x_0)}{na_n}, \tag{2.1}$$

where

$$U_{x_0,b_n}(X_i, V_i) = \mathbb{1}_{[0,\varphi((X_i - x_0)/b_n)]}(V_i). \tag{2.2}$$

Remark 2.2. The events $\{X_i = x\}$, $x \in \mathbb{R}$, can be described in a neighborhood of x_0 through the thinned point process

$$N_n^{\varphi_n}(\cdot) = \sum_{i=1}^{n} U_{x_0,b_n}(X_i, V_i)\varepsilon_{X_i}(\cdot), \tag{2.3}$$

where

$$\varphi_n(x) = \varphi\left(\frac{x - x_0}{b_n}\right) = \mathbb{P}(U_{x_0,b_n}(X_i, V_i) = 1 \mid X_i = x), \tag{2.4}$$

and $U_{x_0,b_n}(X_i, V_i)$ decides whether X_i belongs to the neighborhood of x_0 or not. Precisely, $\varphi_n(x)$ is the probability that the observation $X_i = x$ belongs to the neighborhood of x_0. Note that this neighborhood is not explicitly defined, but it is actually a fuzzy set in the sense of Zadeh [5], given its membership function φ_n. The thinned process $N_n^{\varphi_n}$ is therefore a fuzzy set representation of the data (see Falk and Liese [2], Section 2). Moreover, we can observe that $N_n^{\varphi_n}(\mathbb{R}) = \widehat{\vartheta}_n(x_0)$ and the random variable $\tau_n(x_0)$ is binomial $\mathcal{B}(n, \alpha_n(x_0))$ distributed with

$$\alpha_n(x_0) = \mathbb{E}[U_{x_0,b_n}(X_i, V_i)] = \mathbb{P}(U_{x_0,b_n}(X_i, V_i) = 1) = \mathbb{E}\big[\varphi_n(X)\big]. \tag{2.5}$$

In what follows we assume that $\alpha_n(x_0) \in (0,1)$.

Now, we present the fuzzy set estimator of the regression function introduced in Fajardo et al. [3], which is defined in terms of $\widehat{\vartheta}_n(x_0)$.

Definition 2.3. Let $((X_1, Y_1), V_1), \ldots, ((X_n, Y_n), V_n)$ be independent copies of a random vector $((X, Y), V)$, where $V_1, \ldots, V_n$ are independent random variables uniformly on $[0, 1]$ distributed, and independent of $(X_1, Y_1), \ldots, (X_n, Y_n)$. The fuzzy set estimator of the regression function $r(x) = \mathbb{E}[Y \mid X = x]$ at the point $x_0 \in \mathbb{R}$ is defined as follows:

$$\widehat{r}_n(x_0) = \begin{cases} \dfrac{\sum_{i=1}^{n} Y_i U_{x_0,b_n}(X_i, V_i)}{\tau_n(x_0)} & \text{if } \tau_n(x_0) \neq 0, \\ 0 & \text{if } \tau_n(x_0) = 0. \end{cases} \tag{2.6}$$

Remark 2.4. The fact that $U(x, v) = \mathbb{1}_{[0,\varphi(x)]}(v)$, $x \in \mathbb{R}$, $v \in [0, 1]$, is a kernel when $\varphi(x)$ is a density does not guarantee that $\widehat{r}_n(x_0)$ is equivalent to the Nadaraya-Watson kernel estimator. With this observation the statement made in Remark 2 by Fajardo et al. [3] is corrected. Moreover, the fuzzy set representation of the data $(X_i, Y_i) = (x, y)$ is defined over the window $I_{x_0} \times \mathbb{R}$ with thinning function $\varphi_n(x, y) = \varphi((x - x_0)/b_n)\mathbb{1}_{\mathbb{R}}(y)$, where I_{x_0} denotes the neighborhood of x_0. In the particular case $|Y| \leq M$, $M > 0$, the fuzzy set representation of the data $(X_i, Y_i) = (x, y)$ comes given by $\varphi_n(x, y) = \varphi((x - x_0)/b_n)\mathbb{1}_{[-M,M]}(y)$.

Consider the following conditions.

(C1) Functions f and r are at least twice continuously differentiable in a neighborhood of x_0.

(C2) $f(x_0) > 0$.

(C3) Sequence b_n satisfies: $b_n \to 0$, $nb_n/\log(n) \to \infty$, as $n \to \infty$.

(C4) Function φ is symmetrical regarding zero, has compact support on $[-B, B]$, $B > 0$, and it is continuous at $x = 0$ with $\varphi(0) > 0$.

(C5) There exists $M > 0$ such that $|Y| < M$ *a.s.*

(C6) Function $\phi(u) = \mathbb{E}[Y^2 \mid X = u]$ is at least twice continuously differentiable in a neighborhood of x_0.

(C7) $nb_n^5 \to 0$, as $n \to \infty$.

(C8) Function $\varphi(\cdot)$ is monotone on the positives.

(C9) $b_n \to 0$ and $nb_n^2/\log(n) \to \infty$, as $n \to \infty$.

(C10) Functions f and r are at least twice continuously differentiable on the compact set $[-B, B]$.

(C11) There exists $\lambda > 0$ such that $\inf_{x\in[-B,B]} f(x) > \lambda$.

Next, we present the convergence properties obtained in Fajardo et al. [3].

Theorem 2.5. *Under conditions (C1)–(C5), one has*

$$\widehat{r}_n(x_0) \longrightarrow r(x_0) \ \ a.s. \tag{2.7}$$

Theorem 2.6. *Under conditions (C1)–(C7), one has*

$$\sqrt{na_n}(\widehat{r}_n(x_0) - r(x_0)) \xrightarrow{\mathcal{L}} N\left(0, \frac{\mathrm{Var}[Y \mid X = x_0]}{f(x_0)}\right). \tag{2.8}$$

The "$\xrightarrow{\mathcal{L}}$" symbol denotes convergence in law.

Theorem 2.7. *Under conditions (C4)–(C5) and (C8)–(C11), one has*

$$\sup_{x\in[-B,B]} |\widehat{r}_n(x) - r(x)| = o_{\mathbb{P}}(1). \tag{2.9}$$

Remark 2.8. The estimator $\widehat{r}_n$ has a limit distribution whose asymptotic variance depends only on the point of estimation, which does not occur with kernel regression estimators. Moreover, since $a_n = o(n^{-1/5})$ we see that the same restrictions are imposed for the smoothing parameter of kernel regression estimators.

3. Statistical Methodology

In this section we will obtain the mean square error of $\widehat{r}_n$, as well as the optimal scale factor and the integrated mean square error. Moreover, we establish the conditions to obtain a reduction of the constants that control the bias and the asymptotic variance regarding the classic kernel estimators. The function that minimizes the integrated mean square error of $\widehat{r}_n$ is also obtained.

The following theorem provides the asymptotic representation for the mean square error (MSE) of $\widehat{r}_n$. Its proof is deferred to Section 5.

Theorem 3.1. *Under conditions* (C1)–(C6), *one has*

$$\mathbb{E}\left[[\widehat{r}_n(x) - r(x)]^2\right] = \frac{1}{nb_n} V_F(x) + b_n^4 B_F^2(x) + o\left(a_n^4 + \frac{1}{na_n}\right), \tag{3.1}$$

where

$$\begin{aligned} V_F(x) &= \left[\frac{\phi(x) - r^2(x)}{f(x)}\right] \frac{1}{\int \varphi(x)dx} = \frac{c_1(x)}{\int \varphi(x)dx}, \\ B_F(x) &= \frac{1}{2\int \varphi(u)du} \left[\frac{g^{(2)}(x) - f^{(2)}(x)r(x)}{f(x)}\right] \int u^2\varphi(u)du \\ &= \frac{c_2(x)\int u^2\varphi(u)du}{2\int \varphi(u)du}, \end{aligned} \tag{3.2}$$

with

$$a_n = b_n \int \varphi(x)dx. \tag{3.3}$$

Next, we calculate the formula for the optimal asymptotic scale factor b_n^* to perform the estimation. The integrated mean square error (IMSE) of $\widehat{r}_n$ is given by the following:

$$\mathrm{IMSE}[\widehat{r}_n] = \frac{1}{nb_n}\int V_F(x)dx + b_n^4\int B_F^2(x)dx. \tag{3.4}$$

From the above equality, we obtain the following formula for the optimal asymptotic scale factor

$$b_{n_\varphi}^* = \left[\frac{\int \varphi(u)du \int c_1(u)du}{n\left[\int u^2\varphi(u)du\right]^2 \int [c_2(u)]^2 du}\right]^{1/5}. \tag{3.5}$$

We obtain a scaling factor of order $n^{-1/5}$, which implies a rate of optimal convergence for the $\mathrm{IMSE}^*[\widehat{r}_n]$ of order $n^{-4/5}$. We observe that the optimal scaling factor order for the method of fuzzy set estimation coincides with the order of the classic kernel estimate. Moreover,

$$\mathrm{IMSE}^*[\widehat{r}_n] = n^{-4/5}C_\varphi, \tag{3.6}$$

where

$$C_\varphi = \frac{5}{4}\left[\frac{\left[\int c_1(u)du\right]^4\left[\int u^2\psi(u)du\right]^2 \int [c_2(u)]^2 du}{\left[\int \varphi(u)du\right]^4}\right]^{1/5}, \tag{3.7}$$

with

$$\psi(x) = \frac{\varphi(x)}{\int \varphi(u)du}. \tag{3.8}$$

Next, we will establish the conditions to obtain a reduction of the constants that control the bias and the asymptotic variance regarding the classic kernel estimators. For it, we will consider the usual Nadaraya-Watson kernel estimator

$$\widehat{r}_{\mathrm{NW}_K}(x) = \frac{\sum_{i=1}^n Y_i K((X_i - x)/b_n)}{\sum_{i=1}^n K((X_i - x_0)/b_n)}, \tag{3.9}$$

which has the mean squared error (see e.g, Ferraty et al. [6], Theorem 2.4.1)

$$\mathbb{E}\left[\left[\widehat{r}_{\mathrm{NW}_K}(x) - r(x)\right]^2\right] = \frac{1}{nb_n}V_K(x) + b_n^4 B_K^2(x) + o\left(b_n^4 + \frac{1}{nb_n}\right), \tag{3.10}$$

where

$$V_K(x) = c_1(x) \int K^2(u)du,$$
$$B_K(x) = \frac{c_2(x) \int u^2 K(u)du}{2}. \tag{3.11}$$

Moreover, the IMSE of $\widehat{r}_{\mathrm{NW}_K}$ is given by the following:

$$\mathrm{IMSE}[\widehat{r}_{\mathrm{NW}_K}] = \frac{1}{nb_n} \int V_K(x)\ dx + b_n^4 \int B_K^2(x)\ dx. \tag{3.12}$$

From the above equality, we obtain the following formula for the optimal asymptotic scale factor

$$b^*_{n_{\mathrm{NW}_K}} = \left[\frac{\int K^2(u)du \int c_1(u)du}{n\left[\int u^2 K(u)du\right]^2 \int [c_2(u)]^2 du} \right]^{1/5}. \tag{3.13}$$

Moreover,

$$\mathrm{IMSE}^*[\widehat{r}_{\mathrm{NW}_K}] = n^{-4/5} C_K, \tag{3.14}$$

where

$$C_K = \frac{5}{4} \left[\left[\int c_1(u)du \right]^4 \left[\int K^2(u)du \right]^4 \left[\int u^2 K(u)du \right]^2 \int [c_2(u)]^2 du \right]^{1/5}. \tag{3.15}$$

The reduction of the constants that control the bias and the asymptotic variance, regarding the classic kernel estimators, are obtained if for all kernel K

$$\int \varphi(u)du \geq \left[\int K^2(u)du \right]^{-1}, \qquad \int u^2 \varphi(u)du \leq \int u^2 K(u)du. \tag{3.16}$$

Remark 3.2. The conditions on φ allows us to obtain a value of B such that

$$\int_{-B}^{B} \varphi(u)du > \left[\int K^2(u)du \right]^{-1}. \tag{3.17}$$

Moreover, to guarantee that

$$\int u^2 \varphi(u)du \leq \int u^2 K(u)du, \tag{3.18}$$

we define the function

$$\psi(x) = \frac{\varphi(x)}{\int \varphi(u)du}, \tag{3.19}$$

with compact support on $[-B', B'] \subset [B, B]$. Next, we guarantee the existence of B'. As

$$\frac{1}{\int \varphi(u)du} < \int K^2(u)du, \quad \varphi(x) \in [0,1], \tag{3.20}$$

we have

$$x^2\psi(x) \leq x^2\left(\int K^2(u)du\right). \tag{3.21}$$

Observe that for each $C \in (0, \int u^2K(u)du]$ exists

$$B' = \sqrt[3]{\frac{3C}{2\int K^2(u)du}}, \tag{3.22}$$

such that

$$C = \int_{-B'}^{B'} \left(\int K^2(u)du\right)x^2dx \leq \int u^2K(u)du. \tag{3.23}$$

Combining (3.21) and (3.23), we obtain

$$\int_{-B'}^{B'} u^2\psi(u)du \leq \int u^2K(u)du. \tag{3.24}$$

In our case we take $B' \leq B$.

On the other hand, the criterion that we will implement to minimizing (3.6) and obtain a reduction of the constants that control the bias and the asymptotic variance regarding the classic kernel estimation, is the following

$$\text{Maximizing} \int \varphi(u)du, \tag{3.25}$$

subject to the conditions

$$\int \varphi^2(u)du = \frac{5}{3}; \qquad \int u\varphi(u)du = 0; \qquad \int \left(u^2 - v\right)\varphi(u)du = 0, \tag{3.26}$$

with $u \in [-B, B]$, $\varphi(u) \in [0,1]$, $\varphi(0) > 0$ and $v \leq \int u^2 K_E(u)du$, where K_E is the Epanechnikov kernel

$$K_E(x) = \frac{3}{4}\left(1 - x^2\right)\mathbb{1}_{[-1,1]}(x). \tag{3.27}$$

The Euler-Lagrange equation with these constraints is

$$\frac{\partial}{\partial\varphi}\left[\varphi + a\varphi^2 + bx\varphi + c\left(x^2 - v\right)\varphi\right] = 0, \tag{3.28}$$

where a, b, and c the three multipliers corresponding to the three constraints. This yields

$$\varphi(x) = \left[1 - \left(\frac{16x}{25}\right)^2\right]\mathbb{1}_{[-25/16,25/16]}(x). \tag{3.29}$$

The new conditions on φ, allows us to affirm that for all kernel K

$$\mathrm{IMSE}^*[\hat{r}_n] \leq \mathrm{IMSE}^*[\hat{r}_{\mathrm{NW}_K}]. \tag{3.30}$$

Thus, the fuzzy set estimator has the best performance.

4. Simulations

A simulation study was conducted to compare the performances of the fuzzy set estimator with the classical Nadaraya-Watson estimators. For the simulation, we used the regression function given by Härdle [7] as follows:

$$Y_i = 1 - X_i + e^{(-200(X_i-0.5)^2)} + \varepsilon_i, \tag{4.1}$$

where the X_i were drawn from a uniform distribution based on the interval $[0,1]$. Each ε_i has a normal distribution with 0 mean and 0.1 variance. In this way, we generated samples of size 100, 250, and 500. The bandwidths was computed using (3.5) and (3.13). The fuzzy set estimator and the kernel estimations were computed using (3.29), and the Epanechnikov and Gaussian kernel functions. The IMSE* values of the fuzzy set estimator and the kernel estimators are given in Table 1.

As seen from Table 1, for all sample sizes, the fuzzy set estimator using varying bandwidths have smaller IMSE* values than the kernel estimators with fixed and different bandwidth for each estimator. In each case, it is seen that the fuzzy set estimator has the best performance. Moreover, we see that the kernel estimation computed using the Epanechnikov kernel function shows a better performance than the estimations computed using the Gaussian kernel function.

The graphs of the real regression function and the estimations of the regression functions computed over a sample of 500, using 100 points and $v = 0.2$, are illustrated in Figures 1 and 2.

Table 1: IMSE* values of the estimations for the fuzzy set estimator and the kernel estimators.

v	n	IMSE* $[\widehat{r}_n]$	IMSE* $[\widehat{r}_{\mathrm{NW}_{K_E}}]$	IMSE* $[\widehat{r}_{\mathrm{NW}_{K_G}}]$
	100	0.0093*	0.0111	0.0115
0.2	250	0.0045*	0.0053	0.0055
	500	0.0026*	0.0031	0.0032
	100	0.0083*	0.0111	0.0115
0.15	250	0.0040*	0.0053	0.0055
	500	0.0023*	0.0031	0.0032
	100	0.0070*	0.0111	0.0115
0.10	250	0.0034*	0.0053	0.0055
	500	0.0019*	0.0031	0.0032

Minimum IMSE in each row.

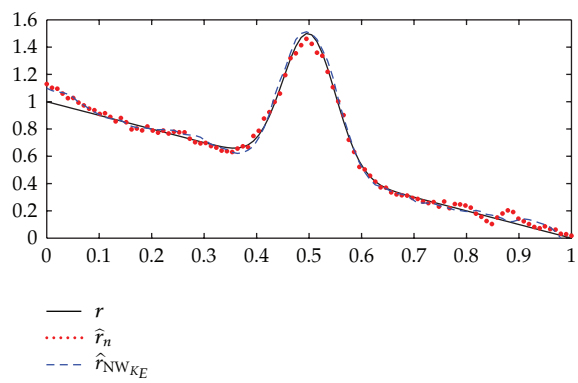

Figure 1: Estimation of r with $\widehat{r}_n$ and $\widehat{r}_{\mathrm{NW}_{K_E}}$.

5. Proof of Theorem 3.1

Proof. Throughout this proof C will represent a positive real constant, which can vary from one line to another, and to simplify the annotation we will write U_i instead of $U_{x,b_n}(X_i, V_i)$. Let us consider the following decomposition

$$\mathbb{E}\left[\left[\widehat{r}_n(x) - r(x)\right]^2\right] = \mathrm{Var}[\widehat{r}_n(x)] + \left(\mathbb{E}[\widehat{r}_n(x) - r(x)]\right)^2. \tag{5.1}$$

Next, we will present two equivalent expressions for the terms to the right in the above decomposition. For it, we will obtain, first of all, an equivalent expression for the expectation. We consider the following decomposition (see e.g, Ferraty et al. [6])

$$\widehat{r}_n(x) = \frac{\widehat{g}_n(x)}{\mathbb{E}\left[\widehat{\vartheta}_n(x)\right]}\left(1 - \frac{\widehat{\vartheta}_n(x) - \mathbb{E}\left[\widehat{\vartheta}_n(x)\right]}{\mathbb{E}\left[\widehat{\vartheta}_n(x)\right]}\right) + \frac{\left[\widehat{\vartheta}_n(x) - \mathbb{E}\left[\widehat{\vartheta}_n(x)\right]\right]^2}{\left[\mathbb{E}\left[\widehat{\vartheta}_n(x)\right]\right]^2}\widehat{r}_n(x). \tag{5.2}$$

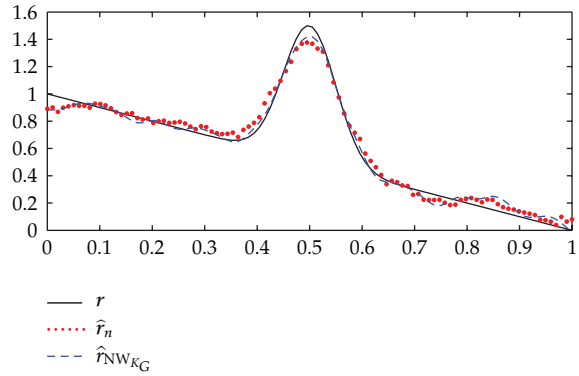

Figure 2: Estimation of r with $\widehat{r}_n$ and $\widehat{r}_{\mathrm{NW}_{K_G}}$.

Taking the expectation, we obtain

$$\mathbb{E}[\widehat{r}_n(x)] = \frac{\mathbb{E}[\widehat{g}_n(x)]}{\mathbb{E}\left[\widehat{\vartheta}_n(x)\right]} - \frac{A_1}{\left[\mathbb{E}\left[\widehat{\vartheta}_n(x)\right]\right]^2} + \frac{A_2}{\left[\mathbb{E}\left[\widehat{\vartheta}_n(x)\right]\right]^2}, \tag{5.3}$$

where

$$\begin{aligned} A_1 &= \mathbb{E}\left[\widehat{g}_n(x)\left(\widehat{\vartheta}_n(x) - \mathbb{E}\left[\widehat{\vartheta}_n(x)\right]\right)\right], \\ A_2 &= \mathbb{E}\left[\left(\widehat{\vartheta}_n(x) - \mathbb{E}\left[\widehat{\vartheta}_n(x)\right]\right)^2 \widehat{r}_n(x)\right]. \end{aligned} \tag{5.4}$$

The hypotheses of Theorem 3.1 allow us to obtain the following particular expressions for $\mathbb{E}[\widehat{g}_n(x)]$ and $\mathbb{E}[\widehat{\vartheta}_n(x)]$, which are calculated in the proof of Theorem 1 in Fajardo et al. [3]. That is

$$\begin{aligned} \mathbb{E}[\widehat{g}_n(x)] &= \mathbb{E}\left[\frac{YU}{a_n}\right] = g(x) + O\left(a_n^2\right), \\ \mathbb{E}\left[\widehat{\vartheta}_n(x)\right] &= \mathbb{E}\left[\frac{U}{a_n}\right] = f(x) + O\left(a_n^2\right). \end{aligned} \tag{5.5}$$

Combining the fact that $((X_i, Y_i), V_i)$, $1 \leq i \leq n$, are identically distributed, with condition (C3), we have

$$A_1 = \mathrm{Cov}\left[\widehat{g}_n(x), \widehat{\vartheta}_n(x)\right] = \frac{1}{na_n}\mathbb{E}\left[\frac{YU}{a_n}\right] - \frac{1}{n}\mathbb{E}\left[\frac{YU}{a_n}\right]\mathbb{E}\left[\frac{U}{a_n}\right]$$

$$= \frac{1}{na_n}\left[g(x)+o(1)\right] - \frac{1}{n}\left[g(x)+o(1)\right]\left[f(x)+o(1)\right]$$
$$= \frac{1}{na_n}g(x) + o\left(\frac{1}{na_n}\right). \tag{5.6}$$

On the other hand, by condition (C5) there exists $C > 0$ such that $|\widehat{r}_n(x)| \leq C$. Thus, we can write

$$|A_2| \leq C\mathbb{E}\left[\left[\widehat{\vartheta}_n(x) - \mathbb{E}\left[\widehat{\vartheta}_n(x)\right]\right]^2\right] = \frac{C}{na_n^2}\left(\mathbb{E}\left[U^2\right] - (\mathbb{E}[U])^2\right)$$
$$= \frac{C}{na_n}\,\frac{\mathbb{E}[U]}{a_n}\{1-\mathbb{E}[U]\}. \tag{5.7}$$

Note that

$$\frac{\alpha_n(x)}{a_n} = \mathbb{E}\left[\widehat{\vartheta}_n(x)\right] = f(x) + O\left(a_n^2\right). \tag{5.8}$$

Thus, we can write

$$|A_2| \leq \frac{C}{na_n}\left[f(x) + O\left(a_n^2\right)\right]\{1-\mathbb{E}[U]\}. \tag{5.9}$$

Note that by condition (C1) the density f is bounded in the neighborhood of x. Moreover, condition (C3) allows us to suppose, without loss of generality, that $b_n < 1$ and by (2.5) we can bound $(1-\mathbb{E}[U])$. Therefore,

$$A_2 = O\left(\frac{1}{na_n}\right). \tag{5.10}$$

Now, we can write

$$\frac{A_1}{\left(\mathbb{E}\left[\widehat{\vartheta}_n(x)\right]\right)^2} = \left(\frac{1}{f^2(x_0)} + o(1)\right)\left(\frac{1}{na_n}g(x_0) + o\left(\frac{1}{na_n}\right)\right) = o(1),$$
$$\frac{A_2}{\left(\mathbb{E}\left[\widehat{\vartheta}_n(x)\right]\right)^2} = \left(\frac{1}{f^2(x)} + o(1)\right)O\left(\frac{1}{na_n}\right) = O\left(\frac{1}{na_n}\right) + o\left(\frac{1}{na_n}\right) \tag{5.11}$$
$$= O\left(\frac{1}{na_n}\right).$$

The above equalities, imply that

$$\mathbb{E}[\widehat{r}_n(x)] = \frac{\mathbb{E}\left[\widehat{g}_n(x)\right]}{\mathbb{E}\left[\widehat{\vartheta}_n(x)\right]} + o(1) + O\left(\frac{1}{na_n}\right) = \frac{\mathbb{E}\left[\widehat{g}_n(x)\right]}{\mathbb{E}\left[\widehat{\vartheta}_n(x)\right]} + O\left(\frac{1}{na_n}\right). \tag{5.12}$$

Once more, the hypotheses of Theorem 3.1 allow us to obtain the following general expressions for $\mathbb{E}[\widehat{\vartheta}_n(x)]$ and $\mathbb{E}[\widehat{g}_n(x)]$, which are calculated in the proofs of Theorem 1 in Fajardo et al. [3, 4], respectively. That is

$$\begin{aligned}\mathbb{E}\left[\widehat{\vartheta}_n(x)\right] &= f(x) + \frac{a_n^2}{2\left[\int \varphi(u)du\right]^3} f''(x) \int u^2\varphi(u)du \\ &\quad + \frac{a_n^2}{2\left[\int \varphi(u)\right]^3} \int u^2\varphi(u)\left[f''(x+\beta u b_n) - f''(x)\right]du,\end{aligned} \tag{5.13}$$

$$\begin{aligned}\mathbb{E}\left[\widehat{g}_n(x)\right] &= g(x) + \frac{a_n^2}{2\left[\int \varphi(u)du\right]^3} g''(x) \int u^2\varphi(u)du \\ &\quad + \frac{a_n^2}{2\left[\int \varphi(u)du\right]^3} \int u^2\varphi(u)\left[g''(x+\beta u b_n) - g''(x)\right]du.\end{aligned} \tag{5.14}$$

By conditions (C1) and (C4), we have that

$$\begin{aligned}&\int u^2\varphi(u)\left[g''(x+\beta u b_n) - g''(x)\right]du = o(1), \\ &\int u^2\varphi(u)\left[f''(x+\beta u b_n) - f''(x)\right]du = o(1).\end{aligned} \tag{5.15}$$

Then

$$\begin{aligned}\mathbb{E}[\widehat{r}_n(x)] &= \frac{g(x) + \left(b_n^2/2\int \varphi(u)du\right)g''(x)\int u^2\varphi(u)du}{f(x) + \left(b_n^2/2\int \varphi(u)du\right)f''(x)\int u^2\varphi(u)du} + O\left(\frac{1}{na_n}\right) \\ &= H_n(x) + O\left(\frac{1}{na_n}\right).\end{aligned} \tag{5.16}$$

Next, we will obtain an equivalent expression for $H_n(x)$. Taking the conjugate, we have

$$\begin{aligned} H_n(x) &= \frac{1}{D_n(x)}\left(g(x)f(x) + \frac{b_n^2\int u^2\varphi(u)du}{2\int\varphi(u)du}\left[g''(x)f(x) - f''(x)g(x)\right]\right. \\ &\qquad \left. + \left(\frac{b_n}{2\int\varphi(u)du}\right)^2 f''(x)g''(x)\left(\int u^2\varphi(u)du\right)^2\right) \\ &= \frac{1}{D_n(x)}\left(g(x)f(x) + \frac{b_n^2\int u^2\varphi(u)du}{2\int\varphi(u)du}\left[g''(x)f(x) - f''(x)g(x)\right]\right) \\ &\quad + o\left(a_n^2\right), \end{aligned} \tag{5.17}$$

where

$$D_n(x) = f^2(x) - \left(\frac{b_n^2 f''(x)\int u^2\varphi(u)du}{2\int\varphi(u)du}\right)^2. \tag{5.18}$$

By condition (C3), we have

$$\frac{1}{D_n(x)} = \frac{1}{f^2(x)} + o(1). \tag{5.19}$$

So that,

$$\begin{aligned} H_n(x) &= \left[\frac{1}{f^2(x)} + o(1)\right]\left(g(x)f(x) + \frac{b_n^2\int u^2\varphi(u)du}{2\int\varphi(u)du}\left[g''(x)f(x) - f''(x)g(x)\right]\right) \\ &\quad + o\left(a_n^2\right) \\ &= r(x) + \frac{b_n^2\int u^2\varphi(u)du}{2\int\varphi(u)du}\left[\frac{g''(x) - f''(x)r(x)}{f(x)}\right] + o\left(a_n^2\right). \end{aligned} \tag{5.20}$$

Now, we can write

$$\begin{aligned} \mathbb{E}[\widehat{r}_n(x) - r(x)] &= \frac{b_n^2\int u^2\varphi(u)du}{2\int\varphi(u)du}\left[\frac{g''(x) - f^{(2)}(x)r(x)}{f(x)}\right] + o\left(a_n^2\right) \\ &\quad + O\left(\frac{1}{na_n}\right). \end{aligned} \tag{5.21}$$

By condition (C3), we have

$$\mathbb{E}[\widehat{r}_n(x) - r(x)] = \frac{b_n^2 \int u^2\varphi(u)du}{2\int \varphi(u)du}\left[\frac{g''(x) - f''(x)r(x)}{f(x)}\right] + o\left(a_n^2\right) + o(1) = b_n^2 B_F(x) + o\left(a_n^2\right), \tag{5.22}$$

where

$$B_F(x) = \left[\frac{g''(x) - f''(x)r(x)}{f(x)}\right]\frac{\int u^2\varphi(u)du}{2\int \varphi(u)du}. \tag{5.23}$$

Therefore,

$$\begin{aligned}(\mathbb{E}[\widehat{r}_n(x) - r(x)])^2 &= b_n^4 B_F^2(x) + 2b_n^2 B_F(x) o\left(a_n^2\right) + o\left(a_n^4\right)\\ &= b_n^4 B_F^2(x) + o\left(a_n^4\right) + o\left(a_n^4\right)\\ &= b_n^4 B_F^2(x) + o\left(a_n^4\right).\end{aligned} \tag{5.24}$$

Next, we will obtain an expression for the variance in (5.1). For it, we will use the following expression (see e.g., Stuart and Ord [8])

$$\begin{aligned}\operatorname{Var}\left[\frac{\widehat{g}_n(x)}{\widehat{\vartheta}_n(x)}\right] &= \frac{\operatorname{Var}[\widehat{g}_n(x)]}{\left(\mathbb{E}\left[\widehat{\vartheta}_n(x)\right]\right)^2} + \frac{\left(\mathbb{E}[\widehat{g}_n(x)]\right)^2}{\left(\mathbb{E}\left[\widehat{\vartheta}_n(x)\right]\right)^4}\operatorname{Var}\left[\widehat{\vartheta}_n(x)\right]\\ &\quad - \frac{2\mathbb{E}[\widehat{g}_n(x)]\operatorname{Cov}\left[\widehat{g}_n(x), \widehat{\vartheta}_n(x)\right]}{\left(\mathbb{E}\left[\widehat{\vartheta}_n(x)\right]\right)^3}.\end{aligned} \tag{5.25}$$

Since that $((X_i, Y_i), V_i)$ are i.i.d and the (X_i, V_i) are i.i.d, $1 \le i \le n$, we have

$$\operatorname{Var}[\widehat{g}_n(x)] = \frac{1}{na_n^2}\operatorname{Var}(YU) = \frac{1}{na_n}\mathbb{E}\left[\frac{1}{a_n}Y^2U\right] - \frac{1}{n}\left(\mathbb{E}\left[\frac{1}{a_n}YU\right]\right)^2, \tag{5.26}$$

$$\operatorname{Var}\left[\widehat{\vartheta}_n(x)\right] = \frac{1}{(na_n)^2}\operatorname{Var}\left[\sum_{i=1}^n U_i\right] = \frac{1}{(na_n)^2} n\alpha_n(x)(1 - \alpha_n(x)), \tag{5.27}$$

the last equality because $\sum_{i=1}^n U_i$ is binomial $\mathcal{B}(n, \alpha_n(x_0))$ distributed. Remember that

$$\mathbb{E}\left[\frac{YU}{a_n}\right] = g(x) + O\left(a_n^2\right). \tag{5.28}$$

Moreover, the hypothesis of Theorem 3.1 allow us to obtain the following expression

$$\mathbb{E}\left[\frac{Y_i^2 U_i}{a_n}\right] = \phi(x) f(x) + O\left(a_n^2\right), \tag{5.29}$$

which is calculated in the proof of Lemma 1 in Fajardo et al. [3]. By condition (C3), we have

$$\begin{aligned} \operatorname{Var}\left[\widehat{g}_n(x)\right] &= \frac{1}{na_n}\left(\phi(x) f(x) + o(1)\right) - \frac{1}{n}\left(g(x) + o(1)\right)^2 \\ &= \frac{1}{na_n}\phi(x) f(x) + o\left(\frac{1}{na_n}\right). \end{aligned} \tag{5.30}$$

Remember that

$$\mathbb{E}\left[\widehat{\vartheta}_n(x)\right] = \frac{1}{a_n}\mathbb{E}[U] = \frac{\alpha_n(x)}{a_n} = f(x_0) + o(1). \tag{5.31}$$

Thus,

$$\begin{aligned} \operatorname{Var}\left[\widehat{\vartheta}_n(x)\right] &= \frac{1}{na_n}\frac{\alpha_n(x)}{a_n} - \frac{1}{n}\left[\frac{\alpha_n(x)}{a_n}\right]^2 \\ &= \frac{1}{na_n}\left(f(x) + o(1)\right) - \frac{1}{n}\left(f(x) + o(1)\right)^2 \\ &= \frac{1}{na_n} f(x) + o\left(\frac{1}{na_n}\right), \\ \frac{1}{\left(\mathbb{E}\left[\widehat{\vartheta}_n(x)\right]\right)^k} &= \frac{1}{f^k(x)} + o(1), \end{aligned} \tag{5.32}$$

for $k = 2, 3, 4$. Finally, we saw that

$$\operatorname{Cov}\left[\widehat{g}_n(x), \widehat{\vartheta}_n(x)\right] = \frac{1}{na_n} g(x) + o\left(\frac{1}{na_n}\right). \tag{5.33}$$

Therefore,

$$\begin{aligned} \frac{\operatorname{Var}\left[\widehat{g}_n(x)\right]}{\left(\mathbb{E}\left[\widehat{\vartheta}_n(x)\right]\right)^2} &= \left[\frac{1}{f^2(x)} + o(1)\right]\left[\frac{1}{na_n}\phi(x) f(x) + o\left(\frac{1}{na_n}\right)\right] \\ &= \frac{1}{na_n}\frac{\phi(x)}{f(x)} + o\left(\frac{1}{na_n}\right), \end{aligned} \tag{5.34}$$

$$\begin{aligned}\frac{\left(\mathbb{E}\left[\widehat{g}_n(x)\right]\right)^2}{\left(\mathbb{E}\left[\widehat{\vartheta}_n(x)\right]\right)^4}\operatorname{Var}\left[\widehat{\vartheta}_n(x)\right] &= \left(\left[\frac{1}{f^4(x)}+o(1)\right]\left[g^2(x)+o(1)\right]\right.\\ &\quad\left.\times\left[\frac{1}{na_n}f(x)+o\left(\frac{1}{na_n}\right)\right]\right)\\ &= \frac{1}{na_n}\frac{g^2(x)}{f^3(x)}+o\left(\frac{1}{na_n}\right),\end{aligned} \tag{5.35}$$

$$\begin{aligned}2\frac{\mathbb{E}\left[\widehat{g}_n(x)\right]}{\left(\mathbb{E}\left[\widehat{\vartheta}_n(x)\right]\right)^3}\operatorname{Cov}\left[\widehat{g}_n(x),\widehat{\vartheta}_n(x)\right] &= \left(2\left[\frac{1}{f^3(x)}+o(1)\right]\left[g(x)+o(1)\right]\right.\\ &\quad\left.\times\left[\frac{1}{na_n}g(x)+o\left(\frac{1}{na_n}\right)\right]\right)\\ &= \frac{2}{na_n}\frac{g^2(x)}{f^3(x)}+o\left(\frac{1}{na_n}\right).\end{aligned} \tag{5.36}$$

Thus,

$$\operatorname{Var}[\widehat{r}_n(x)] = \frac{1}{nb_n}V_F(x)+o\left(\frac{1}{na_n}\right), \tag{5.37}$$

where

$$V_F(x) = \left[\frac{\phi(x)-r^2(x)}{f(x)}\right]\frac{1}{\int\varphi(x)dx}. \tag{5.38}$$

We can conclude that,

$$\begin{aligned}\mathbb{E}\left[\left[\widehat{r}_n(x)-r(x)\right]^2\right] &= \frac{1}{nb_n}V_F(x)+b_n^4B_F^2(x)+o\left(\frac{1}{na_n}\right)+o\left(a_n^4\right)\\ &= \frac{1}{nb_n}V_F(x)+b_n^4B_F^2(x)+o\left(a_n^4+\frac{1}{na_n}\right),\end{aligned} \tag{5.39}$$

where

$$B_F(x) = \frac{\int u^2\varphi(u)du}{2\int\varphi(u)du}\left[\frac{g''(x)-f''(x)r(x)}{f(x)}\right]. \tag{5.40}$$

□

Acknowledgment

The author wants to especially thank the referees for their valuable suggestions and revisions. He also thanks Henrry Lezama for proofreading and editing the English text.

References

[1] R.-D. Reiss, *A Course on Point Processes*, Springer Series in Statistics, Springer, New York, NY, USA, 1993.
[2] M. Falk and F. Liese, "Lan of thinned empirical processes with an application to fuzzy set density estimation," *Extremes*, vol. 1, no. 3, pp. 323–349, 1999.
[3] J. Fajardo, R. Ríos, and L. Rodríguez, "Properties of convergence of an fuzzy set estimator of the regression function," *Journal of Statistic*, vol. 3, no. 2, pp. 79–112, 2010.
[4] J. Fajardo, R. Ríos, and L. Rodríguez, ". Properties of convergence of an fuzzy set estimator of the density function," *Brazilian Journal of Probability and Statistics*, vol. 26, no. 2, pp. 208–217, 2012.
[5] L. A. Zadeh, "Fuzzy sets," *Information and Computation*, vol. 8, pp. 338–353, 1965.
[6] F. Ferraty, V. Núnez Antón, and P. Vieu, *Regresión No Paramétrica: Desde la Dimensión Uno Hasta la Dimensión Infinita*, Servicio Editorial de la Universidad del País Vasco, 2001.
[7] W. Härdle, *Applied Nonparametric Regression.*, New Rochelle, Cambridge, Mass, USA, 1990.
[8] A. Stuart and J. K. Ord, *Kendall's Advanced Theory of Statistics*, vol. 1, Oxford University Press, New York, NY, USA, 1987.