FARAUTE, Vol. 8, No. 2, 18-22, 2013
ISNN 1698-7418
Deposito Legal PP200402CA1617

The fuzzy set estimation of the density function with Bootstrap method

Jesús A. Fajardo G.
Universidad de Oriente. Nucleo de Sucre. Escuela de Ciencias. Departamento de Matemáticas.
Autor de correspondencia: jfajardogonzalez@gmail.com

Abstract
In this paper we estimate the density function by means of an average fuzzy set estimator based on np_n^* i.i.d random variable, where the size of the samples, the optimal average value, and the number of bootstrap replicas are n, p_n^* and $p_n^* - 1$, respectively. This estimator allows us to obtain a reduction of the integrated mean square error of the average fuzzy set estimator regarding the integrated mean squared error of the fuzzy set estimator and classic kernel estimators. This reduction shows that the average fuzzy set estimator has better performance than the fuzzy set estimator and classic kernel estimators, improving the results obtained by Fajardo. Also, the optimal rate of convergence calculated by Ibragimov and Has'minski, as well as the optimal average value are computed. Moreover, we present the properties of convergence and asymptotic representation of the mean square error of the average fuzzy set estimator, which satisfies the desired properties of a good estimator. Additionally, we obtain the function that minimizes the integrated mean square error of the average fuzzy set estimator. Finally, these theoretical findings are illustrated using a numerical example.

Keywords: Density estimation, fuzzy set estimation, nonparametric estimation.

Estimación difusa de la función de densidad con el método Bootstrap

Resumen
En este artículo estimamos la función de densidad a través de un estimador difuso promedio basado en np_n^* variables aleatorias i.i.d, donde el tamaño de la muestra, el valor del promedio óptimo, y el número de replicas bootstrap son n, p_n^* and $p_n^* - 1$, respectivamente. Este estimador nos permite obtener una reducción del error cuadrático medio integrado del estimador difuso promedio con respecto al error cuadrático medio integrado del estimador difuso y los estimadores clásicos por núcleos. Esta reducción muestra que el estimador difuso promedio tiene mejor rendimiento que el estimador difuso y los estimadores clásicos por núcleos, mejorando los resultados obtenidos por Fajardo. También, se calculan la velocidad óptima de convergencia, obtenida por Ibragimov y Has'minski, y el valor del promedio óptimo. Además, presentamos las propiedades de convergencia y la representación asintótica del error cuadrático medio del estimador difuso promedio, el cual satisface las propiedades deseadas de un buen estimador. Adicionalmente, obtenemos la función que minimiza el error cuadrático medio integrado del estimador difuso promedio. Finalmente, estos resultados teóricos se ilustran con un ejemplo numérico.

Palabras Claves: Estimación de densidad, estimación difusa, estimación no paramétrica.

Introduction

The methods of kernel estimation are among the nonparametric methods commonly used to estimate the density function f of a random variable X, with independent samples. Nevertheless, through the theory of point processes (see e.g Reiss, 1993) we can obtain a new nonparametric estimation method. For example, the method of fuzzy set estimation introduced by Fajardo, Ríos and Rodríguez (Fajardo et al., 2012), which is a particular case of the method introduced by Falk and Liese (Falk & Liese, 1998), is based on defining a fuzzy set estimator of the density function by means of thinned point processes (see e.g Reiss, 1993, Section 2.4).

In this paper we estimate the density function by means of an average fuzzy set estimator based on np_n^* i.i.d random variable, where the size of the samples, the optimal average value, and the number of bootstrap replicas are n, p_n^* and p_n^*-1, respectively. With the implementation of this new estimator, we can obtain a significant reduction of the integrated mean square error (*IMSE*) of the average fuzzy set estimator regarding the fuzzy set estimator and classic kernel estimators, which implies that this estimator has better performance than the fuzzy set estimator and classic kernel estimators. Improving the results obtained by Fajardo (Fajardo, 2014). Also, the optimal rate of convergence calculated by Ibragimov and Has'minski (Ibragimov & Has'minski, 1981), as well as the optimal average value are computed. Moreover, we present the properties of convergence and asymptotic representation of the mean square error (*MSE*) of the average fuzzy set estimator, which satisfies the desired properties of a good estimator. In particular, we obtain the optimal scaling factor order and rate of optimal convergence for the *IMSE* of the average fuzzy set estimator, which are better than those of fuzzy set estimator and classic kernel estimators. Besides, the function that minimizes the *IMSE* of the average fuzzy set estimator is obtained. Finally, these theoretical findings are illustrated using a numerical example estimating a density function with the fuzzy set estimators and classic kernel estimators.

This paper is organized as follows. In Section 2, we introduce the average fuzzy set estimator, obtaining the optimal rate of convergence calculated by Ibragimov and Has'minski (Ibragimov & Has'minski, 1981), as well as the the optimal average value. Moreover, the properties of convergence of the estimator and the asymptotic representation of the *MSE* are presented. Besides, the function that minimizes the *IMSE* of the fuzzy set estimator is obtained. In Section 3, a simulation study was conducted to compare the performances of the average fuzzy set estimator with the fuzzy set estimator and classical kernel estimators.

Average Fuzzy set estimator of the density function and its properties

In this section we introduce the average fuzzy set estimator of the density function, which is defined in terms of the fuzzy set estimator $\hat{\vartheta}_n$ (for more details see Fajardo, 2014). Moreover, the optimal rate of convergence calculated by Ibragimov and Has'minski (Ibragimov & Has'minski, 1981), as well as the optimal average value are computed. Its properties of convergence and the asymptotic representation for the *MSE* are presented. Besides, the function that minimizes the *IMSE* of the average fuzzy set estimator is obtained

Next, we will introduce the average fuzzy set estimator of f. For independent copies $(X_i^{(d)}, V_i^{(d)})$, $1 \leq i \leq n$, $1 \leq d \leq p$, and $x_0 \in \{x_1, ..., x_m\} \subset \mathbb{R}$, we define the average fuzzy set estimator of the density function f as

$$\hat{\beta}_{p,n}(x_0) = \frac{1}{p} \sum_{d=1}^{p} \hat{\vartheta}_n^{(d)}(x_0),$$

where the random variables X_i are generated from the empirical distribution and

$$\hat{\vartheta}_n^{(d)}(x_0) = \frac{1}{na_n} \sum_{i=1}^{n} f_{x_0,b_n}(X_i^{(d)}, V_i^{(d)}) = \frac{\tau_n^{(d)}(x_0)}{na_n}.$$

On the other hand, we observe that $\tau_n^{(d)}(x_0)$ is binomial $\mathscr{B}(n, \alpha_n(x_0))$. Thus,

$$\mathbb{E}\left[\hat{\beta}_{p,n}(x_0)\right] = \mathbb{E}\left[\hat{\vartheta}_n(x_0)\right] = \frac{\alpha_n(x_0)}{a_n}$$

and

$$Var\left[\hat{\beta}_{p,n}(x_0)\right] = \frac{1}{p} Var\left[\hat{\vartheta}_n(x_0)\right],$$

where $\hat{\vartheta}_n$ is the fuzzy set estimator of the density function.

Next, we calculate the value of the optimal average p_n^* to perform the estimation.

Theorem 1 *Under conditions* $(C1)-(C3)$*, we have*

$$p_n^* = \left[\frac{\log(n)\int \varphi(u)du}{\left(\int u^2\varphi(u)du\right)^2 \int [f''(u)]^2 du}\right]^{1/5}.$$

Proof. *By Theorem* 4 *in Fajardo (Fajardo, 2014) and properties of order of magnitude, we obtain*

$$Var\left[\hat{\beta}_{p,n}(x_0)\right] = \frac{f(x)}{np\,b_n}\frac{1}{\int \varphi(u)} + O\left(\frac{1}{np\,b_n}\right). \quad (1)$$

Moreover, the IMSE of $\hat{\vartheta}_n$ *is given by*

$$\begin{aligned} IMSE\left[\hat{\vartheta}_n\right] &= \frac{1}{nb_n\int \varphi(u)du} \\ &+ \frac{b_n^4}{4}\left(\frac{\int u^2\varphi(u)du}{\int \varphi(u)du}\right)^2 \\ &\times \int \left[f''(x)\right]^2 dx. \end{aligned} \quad (2)$$

From (2), we obtain the following formula for the optimal asymptotic scale factor

$$b_n^* = \left[\frac{1}{n}\frac{\int \varphi(u)du}{\left(\int u^2\varphi(u)du\right)^2 \int [f''(u)]^2 du}\right]^{1/5}. \quad (3)$$

To obtain the optimal value calculated by Ibragimov and Has'minski (Ibragimov & Has'minski, 1981), we take in (1) $b_n^* = p\,b_n$*. Thus,*

$$p_n^* = \frac{1}{n^{1/5}b_n}\left[\frac{\int \varphi(u)du}{\left(\int u^2\varphi(u)du\right)^2 \int [f''(u)]^2 du}\right]^{1/5}.$$

Setting $b_n = (n\log(n))^{-1/5}$*, we obtain as optimal average value*

$$p_n^* = \left[\frac{\log(n)\int \varphi(u)du}{\left(\int u^2\varphi(u)du\right)^2 \int [f''(u)]^2 du}\right]^{1/5}.$$

□

Now, we can write the average fuzzy set estimator as

$$\hat{\beta}_{np_n^*}(x_0) = \frac{1}{\left(np_n^*\right)a_n}\sum_{d=1}^{p_n^*}\sum_{i=1}^{n} f_{x_0,b_n}(X_i^{(d)},V_i^{(d)}).$$

Theorem 2 *The estimator* $\hat{\beta}_{np_n^*}$ *satisfies similar theorems to the Theorems* 1, 2, 3 *and* 4 *in Fajardo (Fajardo, 2014), where*

C2 *Sequence* b_n *satisfies:* $b_n \to 0$ *and* $\frac{np_n^* b_n}{\log(n)} \to \infty$ *as* $n \to \infty$

C4 $np_n^* b_n^5 \to 0$ *as* $n \to \infty$.

C6 $b_n \to 0$ *and* $\frac{np_n^* b_n^2}{\log(n)} \to \infty$ *as* $n \to \infty$.

Proof. *In the first place, we can observe that the estimator* $\hat{\beta}_{np_n^*}$ *is proportional to average of* np_n^* *independent random variables* $f_{x_0,b_n}(X_i^{(d)},V_i^{(d)})$*,* $1 \le i \le n$*,* $1 \le d \le p_n^*$*. For the details of the proof of each result see Fajardo, Ríos and Rodríguez (Fajardo et al., 2012) and Fajardo (Fajardo, 2014).* □

Remark 1 *It is important to emphasize that* $\hat{\beta}_{np_n^*}$ *satisfies the desired properties of a good estimator, since it is asymptotically unbiased with respect to* f*, consistent, and more efficient than the estimator* $\hat{\vartheta}_n$*. Moreover, the* $IMSE^*$ *of* $\hat{\beta}_{np_n^*}$ *is given by*

$$IMSE^*\left[\hat{\beta}_{np_n^*}\right] = \left(np_n^*\right)^{-4/5} C_\varphi, \quad (4)$$

where

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

with

$$b_{np_n^*}^* = \left[\frac{1}{np_n^*}\frac{\int \varphi(u)du}{\left(\int u^2\varphi(u)du\right)^2 \int [f''(u)]^2 du}\right]^{1/5}. \quad (5)$$

That is, we obtain a scaling factor of order $\left(np_n^*\right)^{-1/5}$*, which implies a rate of optimal convergence for the* $IMSE^*$ *of* $\hat{\beta}_{np_n^*}$*, of order* $\left(np_n^*\right)^{-4/5}$*. Since* $np_n^* \ge n$ *for each* $p_n^* \ge 1$*, we have that the optimal scaling factor order and rate of optimal convergence for the* $IMSE^*\left[\hat{\beta}_{np_n^*}\right]$ *are better than those of* $\hat{\vartheta}_n$ *and* $\hat{f}_{n_K}$*.*

The following theorem allows us to guarantee that the average fuzzy set estimator has better performance than the fuzzy set estimator and classic kernel estimators

Theorem 3 *Under conditions* $(C1)-(C3)$*, we have*

$$IMSE^*\left[\hat{\beta}_{np_n^*}\right] \le IMSE^*\left[\hat{\vartheta}_n\right] \le IMSE^*\left[\hat{f}_{n_K}\right],$$

where the function that minimizes the $IMSE^*\left[\hat{\beta}_{np_n^*}\right]$ *is*

$$\varphi(x) = \left[1 - \left(\frac{16x}{25}\right)^2\right]_{\left[-\frac{25}{16},\frac{25}{16}\right]}(x). \quad (6)$$

Proof. *The left-hand side of the inequality is immediate, compare (4) with (3,3) in Fajardo (Fajardo, 2014). The right-hand side was obtained by Fajardo (Fajardo, 2014).* □

Simulation results

A simulation study was conducted to compare the performances of the fuzzy set estimators with the classical kernel estimators. For the simulation, we used the density function

$$f(x) = \begin{cases} \frac{15}{32}[x(x+2)]^2 & \text{if} \quad -2 \leq x \leq 0 \\ \\ \frac{15}{32}[x(x-2)]^2 & \text{if} \quad 0 \leq x \leq 2. \end{cases}$$

In this way, we generated samples of sizes 100, 250 and 500 with a number of bootstrap replicas for each sample of $p_n^* - 1$. The bandwidths were computed using (3), (5), and (3,5) in Fajardo (Fajardo, 2014). The fuzzy set estimators and kernel estimations were computed using (6), and Epanechnikov and Gaussian kernel functions. The $IMSE^*$ values of the fuzzy set estimators and kernel estimators are given in Table 1 and Table 2.

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

v	n	p_n^*	$\hat{\beta}_{np_n^*}$	$\hat{\vartheta}_n$	$\hat{f}_{n_{K_E}}$
0,2	100	2	0,0085*	0,0149	0,0178
	250	2	0,0041*	0,0071	0,0085
	500	2	0,0024*	0,0041	0,0049
0,15	100	2	0,0076*	0,0133	0,0178
	250	2	0,0037*	0,0064	0,0085
	500	2	0,0021*	0,0037	0,0049
0,10	100	2	0,0065*	0,0113	0,0178
	250	2	0,0031*	0,0054	0,0085
	500	2	0,0018*	0,0031	0,0049

* Minimum $IMSE^*$ in each row.

Table 2. $IMSE^*$ **values of the estimations for the fuzzy set estimators and Gaussian kernel estimator**

v	n	p_n^*	$\left[\hat{\beta}_{np_n^*}\right]$	$[\hat{\vartheta}_n]$	$\left[\hat{f}_{n_{K_G}}\right]$
0,2	100	2	0,0085*	0,0149	0,0185
	250	2	0,0041*	0,0071	0,0089
	500	2	0,0024*	0,0041	0,0051
0,15	100	2	0,0076*	0,0133	0,0185
	250	2	0,0037*	0,0064	0,0089
	500	2	0,0021*	0,0037	0,0051
0,10	100	2	0,0065*	0,0113	0,0185
	250	2	0,0031*	0,0054	0,0089
	500	2	0,0018*	0,0031	0,0051

* Minimum $IMSE^*$ in each row.

As seen from Table 1 and Table 2, for all sample sizes, the average fuzzy set estimator using varying bandwidths have smaller $IMSE^*$ values than the fuzzy set estimator and kernel estimators with fixed and different bandwidth for each estimator. In each case, it is seen that the average fuzzy set estimator has the best performance, where a bootstrap replica is needed. Moreover, we see that the fuzzy set estimation computed using $\hat{\vartheta}_n$ shows a better performance than the estimations computed using the Epanechnikov and Gaussian kernel functions. Also, 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 density function and the estimations of the density functions are computed over a sample of 500, using 100 points and $v = 0{,}2$. They are illustrated in Figure 1, Figure 2 and Figure 3.

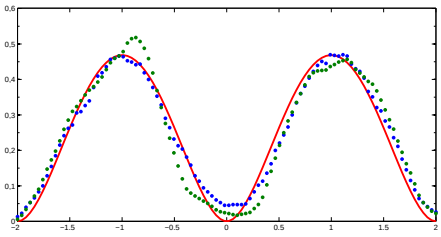

Figura 1. *Estimation of f with* $\hat{\beta}_{np_n^*}$ *and* $\hat{f}_{n_{K_E}}$

f $\hat{\beta}_{np_n^*}$ $\hat{f}_{n_{K_E}}$

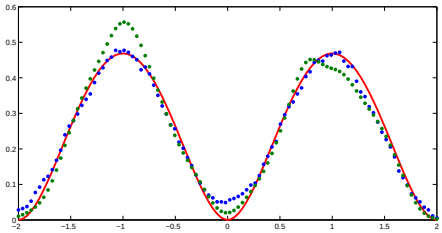

Figura 2. *Estimation of f with* $\hat{\beta}_{np_n^*}$ *and* $\hat{f}_{n_{K_E}}$

f $\hat{\beta}_{np_n^*}$ $\hat{f}_{n_{K_G}}$

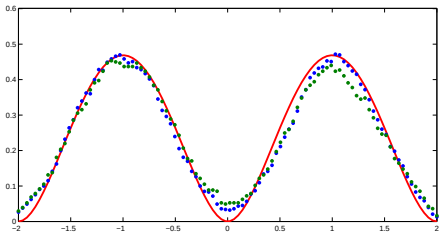

Figura 3. *Estimation of f with* $\hat{\beta}_{np_n^*}$ *and* $\hat{f}_{n_{K_E}}$

f $\hat{\beta}_{np_n^*}$ $\hat{\vartheta}_n$

References

Fajardo, J. (2014). A criterion for the fuzzy set estimation of the density func- tion. *Braz. J. Probab. Stat.*, **28**(3):301-312.

Fajardo, J., R. Ríos & L. Rodríguez. (2012). Properties of convergence of an fuzzy set estimator of the density function. *Braz. J. Probab. Stat.*, **26**(2):208-217.

Falk, M. & F. Liese. (1998). Lan of thinned empirical processes with an application to fuzzy set density estimation. *Extremes*, **1**(3), 323-349.

Ibragimov, I. A. & R. Z. Has'minski. (1981). Statistical Estimation. Asymp- totic Theory. Springer-Verlag, Berlin.

Reiss, R. D. (1993). A Course on Point Processes. Springer Series in Statis- tics, New York.