Brazilian Journal of Probability and Statistics
2012, Vol. 26, No. 2, 208–217
DOI: 10.1214/10-BJPS130

Properties of convergence of a fuzzy set estimator of the density function

Jesús A. Fajardo[a], Ricardo R. Ríos[b] and Luis A. Rodríguez[c]

[a]*Universidad de Oriente*
[b]*Universidad Central de Venezuela*
[c]*Universidad de Carabobo*

Abstract. In this paper we establish the almost sure, in law, and uniform convergence over compact subsets on $\mathbb{R}$ of a fuzzy set estimator of the density function, based on n i.i.d. random variable.

1 Introduction

The theory of density estimation has developed rapidly since the second half of the 1960s. There is an extensive reference to different techniques for nonparametric density estimation. For example, the works of Castellan (2003), Donoho et al. (1996), Gray and Moore (2003), Katkovnik and Shmulevich (2002), López et al. (2008), Miller and Horn (1998) and Miyoshi et al. (1999) present new nonparametric density estimation methods, in which the density is estimated by means of exponential model selection, wavelet thresholding algorithms for adaptive window size, soft clustering, entropy maximization, and the use of kernel methods in regression. The first use of wavelet bases for density estimation appears in papers by Doukhan and León (1990), Kerkyacharian and Picard (1993) and Walter (1992). Furthermore, Fiori and Bucciarelli (2001) addressed the problem of estimating the density function of a quasi-stationary random process by means of neurons with adaptive activation function, and Pelletier (2005) discussed the estimation of a probability density on a Riemannian manifold.

In this paper we establish the almost sure, in law, and uniform convergence over compact subsets on $\mathbb{R}$ of a fuzzy set estimator of the density function, based on n i.i.d. random variables. We used Bernstein type inequalities, properties of local asymptotic normality thinned point processes, as well as Talagrand's inequalities for empirical processes, symmetrization techniques, Rademacher average, and Vapnik–Chervonenkis dimensions (VC dimensions) to obtain such convergence properties.

The proposed fuzzy set estimator is a particular case of the estimator introduced by Falk and Liese (1998), which is defined through thinned point processes

Key words and phrases. Density estimation, fuzzy set, nonparametric estimation, thinned point process, Vapnick–Chervonenkis dimensions.

Received August 2010; accepted September 2010.

(see, e.g., Reiss, 1993, Section 2.4). In Falk and Liese (1998) only the asymptotic efficiency within the class of all estimators that are based on randomly selected points from the sample $X_1, \ldots, X_n$ was proved. Efficiency was established using LeCam's LAN theory. On the other hand, it is important to emphasize that the almost sure and in law convergence of the proposed estimator in Falk and Liese (1998) are obtained by applying the technique implemented in this paper. However, the technique utilized to obtain the uniform convergence is nonviable since the random variables that define it do not possess, for example, precise functional characteristics in regards to the sample. Therefore, some part of the results obtained in this paper are a consequence of the functional characteristics of the random variables that define the proposed estimator. For example, this estimator will facilitate the study of VC dimensions, which will allow us to find an upper bound on the expectation for the supremum of an empirical process using symmetrization techniques, and Rademacher averages as an important step to establish the uniform convergence of the estimator.

The thinned point processes allow us to introduce a thinning function which 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 density function. In Section 3, we present the conditions under which the three main results are true, Theorems 1, 2, and 3. The Appendix contains the proofs of the theorems in Section 3.

2 Fuzzy set estimator of the density function

In this section we define through thinned point processes a nonparametric and fuzzy set estimator of the density function, obtaining a particular case of estimator introduced by Falk and Liese (1998). Moreover, we postulate the pointwise convergence in law.

For each measurable Borel function $\varphi : \mathbb{R} \to [0, 1]$ and each random variable V, uniformly on $[0, 1]$ distributed and independent of X, the random variable $U = \mathbb{1}_{[0,\varphi(X)]}(V)$ satisfies $\varphi(x) = \mathbb{P}(U = 1|X = x)$. This simple observation allows us to construct a fuzzy set estimator of the density function f of a random variable X that satisfies the conditions required in Falk and Liese (1998).

Let $X_1, \ldots, X_n$ be an independent random sample of a real random variable X whose distribution $\mathcal{L}(X)$ has density f with respect to the Lebesgue measure. Let $V_1, \ldots, V_n$ be independent random variables uniformly on $[0, 1]$ distributed and independent of $X_1, \ldots, X_n$. Let $f_{x_0,b_n}(X_i, V_i) = \mathbb{1}_{I_i}(V_i)$ be random variables where $I_i = [0, \varphi(\frac{X_i - x_0}{b_n})]$ and $b_n > 0$ is a scaling factor (or bandwidth) such that $b_n \to 0$ as $n \to \infty$. For each $x \in \mathbb{R}$, we have

$$\varphi\left(\frac{x - x_0}{b_n}\right) = \mathbb{P}\big(f_{x_0,b_n}(X_i, V_i) = 1|X_i = x\big),$$

then $\varphi_n(x) = \varphi(\frac{x-x_0}{b_n})$ is a Markov kernel (see Reiss, 1993, Section 1.4); thus, for independent copies (X_i, V_i), $1 \le i \le n$, of (X, V), we can define the thinned point process

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

with underlying point process $N_n(\cdot) = \sum_{i=1}^{n} \varepsilon_{X_i}(\cdot)$ and a thinning function φ_n (see Reiss, 1993, Section 2.4), where ε_X is the random Dirac measure.

Remark 1. The events $\{X_i = x\}$, $x \in \mathbb{R}$, can be described in a neighborhood of x_0 through the thinned point process $N_n^{\varphi_n}$, where $f_{x_0,b_n}(X_i, V_i)$ 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 (1965), given by 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, 1998, Section 2).

Next, we present the fuzzy set estimator of the density function, which is a particular case of the estimator proposed by Falk and Liese (1998).

Definition 1. Let φ be such that $0 < \int \varphi(x)\,dx < \infty$ and $a_n = b_n \int \varphi(x)\,dx$. Then the fuzzy set estimator of the density function f at the point $x_0 \in \mathbb{R}$ is defined as

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

Remark 2. We observe that the random variable $\tau_n(x_0)$ is binomial $\mathcal{B}(n, \alpha_n(x_0))$ distributed with

$$\alpha_n(x_0) = \mathbb{E}[f_{x_0,b_n}(X_i, V_i)] = \mathbb{P}\big(f_{x_0,b_n}(X_i, V_i) = 1\big) = \mathbb{E}[\varphi_n(X)]. \tag{2.1}$$

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

3 Main results

In this section we state our main results. Next, we shall give sufficient conditions to assure the almost sure convergence of $\hat{\vartheta}_n(x_0)$ in a neighborhood of x_0.

(C1) The density function f is at least twice continuously differentiable in a neighborhood of x_0.

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

(C3) Function φ is symmetrical with respect to zero, has compact support on $[-B, B]$, $B > 0$, and it is continuous at $x = 0$ with $\varphi(0) > 0$.

To obtain the pointwise convergence in law of $\hat{\vartheta}_n(x_0)$, we need the following condition.

(C4) $nb_n^5 \to 0$ as $n \to \infty$.

To obtain the uniform convergence of $\hat{\vartheta}_n$ over compact subsets on $\mathbb{R}$, we introduce a new condition for both function φ and the sequence b_n, as well as the uniform version of condition (C1).

(C5) Function $\varphi(\cdot)$ is monotone on the positives.
(C6) $b_n \to 0$ and $\frac{nb_n^2}{\log(n)} \to \infty$ as $n \to \infty$.
(C7) Density function f is at least twice continuously differentiable on the compact set $[-B, B]$.

Next, we present the three main results of our work:

Theorem 1. *Under conditions* (C1)–(C3), *we have*

$$\hat{\vartheta}_n(x_0) \to f(x_0) \qquad a.s.$$

Theorem 2. *Under conditions* (C1)–(C4), *we have*

$$\sqrt{na_n}\big(\hat{\vartheta}_n(x_0) - f(x_0)\big) \xrightarrow{\mathcal{L}} N(0, f(x_0)).$$

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

Remark 3. The estimator $\hat{\vartheta}_n$ has a limit distribution whose asymptotic variance depends only at the point of estimation, this does not hold to the kernel estimator. However, since $a_n = o(n^{-1/5})$ we see that the same restrictions are imposed for the smoothing parameter of the kernel estimators.

Theorem 3. *Under conditions* (C3) *and* (C5)–(C7), *we have*

$$\sup_{a\in[-B,B]} |\hat{\vartheta}_n(a) - f(a)| = o_{\mathbb{P}}(1).$$

Appendix: Proofs

Throughout this section C represents a positive real constant, which can vary from one line to another and $W_i = (X_i, V_i)$, $1 \le i \le n$.

Proof of Theorem 1. Let us consider the sequence of i.i.d. random variables

$$H_i = \frac{f_{x_0,b_n}(W_i) - \mathbb{E}[f_{x_0,b_n}(W_i)]}{a_n},$$

$1 \le i \le n$. Next, we will obtain upper bounds for $|H_i|$ and $\mathbb{E}[H_i^2]$. From (2.1) and the fact that $\alpha_n(x_0) \in (0, 1)$, we have that $|H_i| \le \frac{2}{a_n}$ and $\mathbb{E}[H_i^2] \le \frac{\alpha_n(x_0)}{a_n^2}$. Now, we calculate an upper bound for $\mathbb{E}[H_i^2]$, independent of x_0, by upper bounding $\alpha_n(x_0)/a_n$. For it, if we combine condition (C1), which allows us to make a Taylor expansion of the density function f on the neighborhood of x_0, with condition (C3), we can write (2.1) as follows

$$\alpha_n(x_0) = a_n\left\{f(x_0) + \frac{b_n^2}{2\int \varphi(x)\,dx}\int u^2\varphi(u) f''(x_0 + \beta u b_n)\,du\right\}, \tag{A.1}$$

where $\beta \in (0, 1)$. Moreover, condition (C2) allows us to suppose, without loss of generality, that $b_n < 1$; and conditions (C3) and (C1), imply that φ, f and f'' are bounded in a neighborhood of x_0. Therefore, there exists $C > 0$ such that $[\alpha_n(x_0)/a_n] < C$. Then, $\mathbb{E}[H_i^2] \le \frac{C}{a_n}$. On the other hand, using Bernstein's inequality (see, e.g., Ferraty et al., 2001, Lemma 2.3.1) for random variables $H_1, \dots, H_n$, we have

$$\mathbb{P}\left(\frac{1}{n}\left|\sum_{i=1}^n H_i\right| > \epsilon\right) = \mathbb{P}\big(|\hat{\vartheta}_n(x_0) - \mathbb{E}[\hat{\vartheta}_n(x_0)]| > \epsilon\big) \le 2e^{-\epsilon^2 n a_n/(4C)}$$

for each $\epsilon \in (0, C)$. For all sufficiently large n condition (C2), implies that $\epsilon = \epsilon_0\sqrt{\frac{\log(n)}{n a_n}} < C$, for each $\epsilon_0 > 0$. Therefore,

$$\mathbb{P}\left(|\hat{\vartheta}_n(x_0) - \mathbb{E}[\hat{\vartheta}_n(x_0)]| > \epsilon_0\sqrt{\frac{\log(n)}{n a_n}}\right) \le 2e^{-C\epsilon_0^2 \log(n)}. \tag{A.2}$$

Cauchy's integral criterion for positive term series, implies that there exists $\epsilon_0 > 0$ such that the term to the right of (A.2) is the general term of a sequence whose sum is convergent. That is, there exists $\epsilon_0 > 0$ such that

$$\sum_{n=1}^{\infty}\mathbb{P}\left(|\hat{\vartheta}_n(x_0) - \mathbb{E}[\hat{\vartheta}_n(x_0)]| > \epsilon_0\sqrt{\frac{\log(n)}{n a_n}}\right) < \infty. \tag{A.3}$$

Since φ and f'' are bounded in a neighborhood of x_0, by conditions (C3) and (C1), we can write (A.1) as

$$\mathbb{E}[\hat{\vartheta}_n(x_0)] - f(x_0) = \frac{\alpha_n(x_0)}{a_n} - f(x_0) = O(b_n^2). \tag{A.4}$$

Again, for all sufficiently large n condition (C2) together with (A.4) and (A.3), imply that

$$\sum_{n=1}^{\infty}\mathbb{P}\big(|\hat{\vartheta}_n(x_0) - f(x_0)| > \epsilon\big) < \infty$$

for a given $\epsilon > 0$. The demonstration now ends by applying the Borel–Cantelli lemma and Theorem 4.2.2 in Chung (2001). □

Proof of Theorem 2. Let us consider the following decomposition

$$(na_n)^{1/2}\big(\hat{\vartheta}_n(x_0) - f(x_0)\big) = Z_n(x_0)\left[\frac{n\alpha_n(x_0)(1-\alpha_n(x_0))}{na_n}\right]^{1/2} + (na_n)^{1/2}\left(\frac{\alpha_n(x_0)}{a_n} - f(x_0)\right),$$

where

$$Z_n(x_0) = \frac{\tau_n(x_0) - n\alpha_n(x_0)}{[n\alpha_n(x_0)(1-\alpha_n(x_0))]^{1/2}}$$

with

$$\tau_n(x_0) = \sum_{i=1}^{n} U_{x_0,b_n}(W_i).$$

Next, we will present equivalent expressions for the different terms to $Z_n(x_0)$ in the above decomposition. The conditions (C1) and (C3) allows us to obtain (A.1), and from (A.1) we can write

$$(na_n)^{1/2}\left(\frac{\alpha_n(x_0)}{a_n} - f(x_0)\right) = \frac{(na_n^5)^{1/2}\int u\varphi(u) f''(x_0 - \gamma u b_n)\,du}{(\int \varphi(u)\,du)^2}, \tag{A.5}$$

where $\gamma \in (0, 1)$. The combination of (C1) and (C3), which allows us to guarantee that f'' and φ are bounded in neighborhood of x_0, with (C4) implies

$$(na_n)^{1/2}\left(\frac{\alpha_n(x_0)}{a_n} - f(x_0)\right) = o(1).$$

On the other hand, combining (A.5) and (C1)–(C3) we obtain $\frac{\alpha_n(x_0)}{a_n} \to f(x_0)$ and $\alpha_n(x_0) \to 0$ as $n \to \infty$. Thus,

$$\left[\frac{n\alpha_n(x_0)(1-\alpha_n(x_0))}{na_n}\right]^{1/2} = [f(x_0)]^{1/2} + o(1).$$

Therefore,

$$(na_n)^{1/2}\big(\hat{\vartheta}_n(x_0) - f(x_0)\big) = [f(x_0)]^{1/2} Z_n(x_0) + Z_n(x_0)o(1) + o(1).$$

Since $\tau_n(x_0)$ is binomial $\mathcal{B}(n, \alpha_n(x_0))$ distributed, we have $\mathbb{E}[Z_n^2(x_0)] = 1$. Thus,

$$(na_n)^{1/2}\big(\hat{\vartheta}_n(x_0) - f(x_0)\big) = [f(x_0)]^{1/2} Z_n(x_0) + o_{\mathbb{P}}(1).$$

The proof now ends by combining the Moivre–Laplace theorem and Slutsky's theorem. □

We will need the following lemma. This result, which holds true in spaces of greater dimension, establishes that the class of functions $\mathcal{F} = \{f_{a,b} : Z \to \{0,1\} : (a,b) \in Z\}$, $Z = \mathbb{R} \times [0,1]$ and $b \neq 0$, has finite VC dimension, $V(\mathcal{F}) < \infty$. This will allow us to upper bound $\mathbb{E}[Q]$ (see (A.6)) in terms of the combinatorial quantity $V(\mathcal{F})$ when we use symmetrization techniques and Rademacher average as an important step to establish the uniform convergence of $\hat{\vartheta}_n$.

Lemma 1. *Under conditions* (C3) *and* (C5), *we have* $\mathcal{F}$ *is VC class.*

Proof. Let $T = \{((a,b),(x,v)) \in Z^2 : \varphi(\frac{x-a}{b}) \geq v\}$, where $Z = \mathbb{R} \times [0,1]$ and $b \neq 0$. The structures of incidence on $Z \times Z$, also called dual classes relative to T (see Assouad, 1983, Section 2.7), are expanded next: $T(Z) = \{T_{(a,b)} : (a,b) \in Z\}$ and $T^{-1}(Z) = \{T^{(x,v)} : (x,v) \in Z\}$, where $T_{(a,b)} = \{(x,v) \in Z : \varphi(\frac{x-a}{b}) \geq v\}$ and $T^{(x,v)} = \{(a,b) \in Z : \varphi(\frac{x-a}{b}) \geq v\}$. Furthermore, conditions (C3) and (C5) allow us to write $T^{(x,v)}$ as $\{(a,b) \in Z : (x-a)^2 - b^2 C \geq 0\}$, and for fixed (x,v), we find that the set $T^{(x,v)}$ represents the positiveness of a quadratic form in variables a and b. Thus, class $T^{-1}(Z)$ represents the class of the quadratic forms in variables a and b that are positive. Lema 18 in Pollard (1984) establishes that $T^{-1}(Z)$ is VC class and Proposition 2.12 in Assouad (1983) allows us to assert that $T(Z)$ is VC class. Moreover, as class $\mathcal{F}$ is a set of indicators, it can be written as $T(Z)$. Thus, $\mathcal{F}$ is VC class. □

Proof of Theorem 3. Let A be a countable subset of $[-B, B]$. Next, we will obtain a concentration inequality for

$$Q = \sup_{a \in A} \left| \sum_{i=1}^{n} \{f_{a,b_n}(W_i) - \mathbb{E}[f_{a,b_n}(W_i)]\} \right|, \tag{A.6}$$

by means of Talagrand's inequalities for empirical processes (see, e.g., Massart, 2000, Section 2.2). We observe that $\|f_{a,b_n} - \mathbb{E}[f_{a,b_n}]\|_\infty \leq 2 = b$. According to Theorem 2.3 in Massart (2000), we have

$$\mathbb{P}\big(Q > \mathbb{E}[Q] + c_1\sqrt{Lx} + c_2 bx\big) \leq K e^{-x}$$

for each $x > 0$, where K, c_1 and c_2 are universal positive constants, and

$$L = \mathbb{E}\left[\sup_{a \in A} \sum_{i=1}^{n} \big(f_{a,b_n}(W_i) - \mathbb{E}[f_{a,b_n}(W_i)]\big)^2\right].$$

Since $|f_{a,b_n} - \mathbb{E}[f_{a,b_n}]| \leq 2$, we have $L \leq 4n$. Therefore,

$$\begin{aligned} \mathbb{P}\big(Q > \mathbb{E}[Q] + c_1\sqrt{nx} + c_2 bx\big) &\leq \mathbb{P}\big(Q > \mathbb{E}[Q] + c_1\sqrt{Lx} + c_2 bx\big) \\ &\leq K e^{-x}. \end{aligned}$$

Setting $x = \log(n)$, we obtain

$$\mathbb{P}\Big(\sup_{a\in A}|\hat{\vartheta}_n(a) - \mathbb{E}[\hat{\vartheta}_n(a)]| > S_n(Q)\Big) \le Ke^{-\log(n)}, \tag{A.7}$$

where $S_n(Q) = \frac{\mathbb{E}[Q]}{na_n} + c_1\sqrt{C\frac{\log(n)}{na_n^2}} + \frac{c_2 b\log(n)}{na_n}$. Now, we show that

$$\mathbb{P}\Big(\sup_{a\in A}|\hat{\vartheta}_n(a) - \mathbb{E}[\hat{\vartheta}_n(a)]| > \epsilon\Big) \le \mathbb{P}\Big(\sup_{a\in A}|\hat{\vartheta}_n(a) - \mathbb{E}[\hat{\vartheta}_n(a)]| > S_n(Q)\Big)$$

for each $\epsilon > 0$. In order to do that, we use symmetrization techniques (see, e.g., Bousquet et al., 2005, Section 3) to obtain a combinatorial quantity as an upper bound of $\mathbb{E}[Q]$. Let $W_1', \dots, W_n'$ be independent copies of random variables $W_1, \dots, W_n$, and $\eta_1, \dots, \eta_n \in \{-1, 1\}$ independent random variables (Rademacher) and independent of the first two with $\mathbb{P}(\eta_i = -1) = \mathbb{P}(\eta_i = 1) = 1/2$. The properties of the conditional expectation, together with the independence of W and W', and Jensen's inequality (over supremum), allow us to obtain

$$\mathbb{E}[Q] \le \mathbb{E}\left[\sup_{a\in A}\left|\sum_{i=1}^{n}\big(f_{a,b_n}(W_i) - f_{a,b_n}(W_i')\big)\right|\right].$$

On the other hand, since $[f_{a,b_n}(W_i) - f_{a,b_n}(W_i')]$ is symmetric and independent of η_i, we have $[f_{a,b_n}(W_i) - f_{a,b_n}(W_i')] \overset{\mathcal{L}}{=} \eta_i[f_{a,b_n}(W_i) - f_{a,b_n}(W_i')]$, where the symbol "$\overset{\mathcal{L}}{=}$" denotes equality in law. Then,

$$\mathbb{E}[Q] \le 2n\mathbb{E}\left[\sup_{a\in A}\frac{1}{n}\left|\sum_{i=1}^{n}\eta_i f_{a,b_n}(W_i)\right|\right].$$

To bound the right side of the above inequality, Lemma 5.2 in Massart (2000) is applied to the finite set $\mathcal{F}(w_1^n) = \{(f_{a,b_n}(w_1), \dots, f_{a,b_n}(w_n)) : f_{a,b_n} \in \mathcal{F}\} \subset \mathbb{R}^n$, where $w_1^n = (w_1, \dots, w_n)$. Let $\mathbb{S}_{\mathcal{F}}(w_1^n)$ be cardinality or VC shatter coefficient of $\mathcal{F}(w_1^n)$. We observe that, $\mathbb{S}_{\mathcal{F}}(w_1^n) \le 2^n$. Thus, we can bound the Rademacher average $R_n(\mathcal{F}(w_1^n))$ associated with $\mathcal{F}(w_1^n)$ for all w_1^n as follows:

$$R_n(\mathcal{F}(w_1^n)) = \mathbb{E}\left[\sup_{a\in A}\frac{1}{n}\left|\sum_{i=1}^{n}\eta_i f_{a,b_n}(w_i)\right|\right] \le \sqrt{\frac{2\log(\mathbb{S}_{\mathcal{F}}(w_1^n))}{n}},$$

since $[\sum_{i=i}^{n}[f_{a,b_n}(w_i)]^2]^{1/2} \le \sqrt{n}$. Then,

$$\mathbb{E}\left[\mathbb{E}\left[\sup_{a\in A}\frac{1}{n}\left|\sum_{i=1}^{n}\eta_i f_{a,b_n}(W_i)\right|\middle|W_1^n = w_1^n\right]\right] \le \mathbb{E}\left[\sqrt{\frac{2\log(\mathbb{S}_{\mathcal{F}}(W_1^n))}{n}}\right].$$

According to the Cauchy–Schwarz inequality, we have

$$2n\mathbb{E}\left[\sup_{a\in A}\frac{1}{n}\left|\sum_{i=1}^{n}\eta_i f_{a,b_n}(W_i)\right|\right] \le 2\sqrt{2n}\sqrt{\mathbb{E}[\log(\mathbb{S}_{\mathcal{F}}(W_1^n))]}.$$

Now, combining Lemma 1 and Sauer's lemma (see, e.g., Ludeña and Ríos, 2003, Lemma 11-(2)), we can upper bound the logarithm of the VC shatter coefficient with the combinatorial quantity $V(\mathcal{F})$, called the VC dimension of the $\mathcal{F}$ class. That is,

$$\log(\mathbb{S}_{\mathcal{F}}(W_1^n)) \leq V(\mathcal{F})\left(1+\log\left(\frac{n}{V(\mathcal{F})}\right)\right).$$

The above inequality, together with condition (C6), imply that $S_n(Q) \to 0$ as $n \to \infty$. Thus, from inequality (A.7) we have that for all sufficiently large n there exist $\epsilon > 0$ and $\delta > 0$ such that

$$\mathbb{P}\Big(\sup_{a\in A} |\hat{\vartheta}_n(a) - \mathbb{E}\hat{\vartheta}_n(a)| > \epsilon\Big) < \delta. \tag{A.8}$$

On the other hand, the approaches with $O(\cdot)$ carried out to obtain (A.4), which holds true under the hypotheses of Theorem 3, are really uniform approaches for $x \in [-B, B]$ since the density f satisfies the regularity condition (C7). Then

$$\sup_{a\in A} |\mathbb{E}[\hat{\vartheta}_n(a)] - f(a)| = O(a_n^2). \tag{A.9}$$

Combining (A.8) and (A.9) allow us to obtain the desired result for an arbitrary countable set $A \subset [-B, B]$. In particular these exists a countable dense set $D \subset [-B, B]$ such that the result holds. On the other hand, to extend this result to the compact set $[-B, B]$ it will be enough to observe that, the family $\{\mathbb{1}_{[0,\varphi(x)]}(v) : (x, v) \in \mathbb{R} \times [0, 1]\}$ is continuous at x under conditions that guarantee that $[0, \varphi(x)] \supseteq [0, \varphi(x_0)]$ or $[0, \varphi(x)] \subseteq [0, \varphi(x_0)]$ when $| x - x_0 | < \delta$. For example: continuity of φ, or increasing or decreasing of φ. □

Acknowledgments

The authors want to especially thank the referees for their valuable suggestions and revisions. We also thank Carenne Ludeña, Adolfo Quíroz and Henrry Lezama for their very helpful comments and suggestions.

References

Assouad, P. (1983). Densité et dimension. *Annales de l'Institut Fourier (Grenoble)* **33**, 233–282. MR0723955

Bousquet, O., Boucheron, S. and Lugosi, G. (2005). Theory of classification: A survery of some recent advancement. *ESAIM: Probability and Statistics* **9**, 323–375. MR2182250

Castellan, G. (2003). Density estimation via exponential model selection. *IEEE Transactions on Information Theory* **49**, 2052–2068. MR2004713

Chung, K. L. (2001). *A Course in Probability Theory*, 3rd ed. New York: Academic Press. MR1796326

Donoho, D. L., Johnstone, I. M., Kerkyacharian, G. and Picard, D. (1996). Density estimation by wavelet thresholding. *The Annals of Statistics* **24**, 508–539. MR1394974

Doukhan, P. and León, J. (1990). Deviation quadratique d'estimateur de densité par projection orthogonale. Note aux. *Comptes Rendus Acad. Sciences Paris* **310**, 424–430. MR1046526

Falk, M. and Liese, F. (1998). Lan of thinned empirical processes with an application to fuzzy set density estimation. *Extremes* **1**, 323–349. MR1814708

Ferraty, F., Núñez-Antón, V. and Vieu, P. (2001). Regresión No Paramétrica: Desde la Dimensión Uno Hasta la Dimensión Infinita. Servicio Editorial de la Universidad del País Vasco.

Fiori, S. and Bucciarelli, P. (2001). Probability density estimation using adaptive activation function neurons. *Neural Processing Letters* **13** 31–42.

Gray, A. and Moore, A. (2003). Nonparametric density estimation: Toward computational tractability. In *Proceedings at SIAM International Conference on Data Mining*, San Francisco, USA, 2003, 9 p.

Katkovnik, V. and Shmulevich, I. (2002). Kernel density estimation with adaptive varying window size. *Pattern Recognition Letters* **23**, 1641–1648.

Kerkyacharian, G. and Picard, D. (1993). Density estimation by kernel and wavelets methods: Optimality of Besov space. *Statistics and Probability Letters* **18**, 327–336. MR1245704

López, E., Ortiz, J., de Lazcano, D. L. and Vargas, M. (2008). Soft clustering for nonparametric probability density function estimation. *Pattern Recognition Letters* **29** 2085–2091.

Ludeña, C. and Ríos, R. (2003). *Teoría de aprendizaje estadístico y selección de modelos*. Mérida-Venezuela: Decimosexta Escuela Venezolana de Matemáticas.

Massart, P. (2000). Some applications of concentration inequalities to statistics. Probability theory. *Annales de la Faculté des Sciences de Toulouse* **9**, 245–303. MR1813803

Miller, G. and Horn, D. (1998). Probability density estimation using entropy maximization. *Neural Computation* **10**, 1925–1938.

Miyoshi, T., Ichihashi, H. and Nagasaka, K. (1999). Fuzzy projection pursuit density estimation by eigenvalue method. *International Journal of Approximate Reasoningr* **20**, 237–248. MR1685081

Pelletier, B. (2005). Kernel density estimation on Riemannian manifolds. *Statistics and Probability Letters* **73**, 297–304. MR2179289

Pollard, D. (1984). *Convergence of Stochastic Processes*. New York: Springer-Verlag. MR0762984

Reiss, R. D. (1993). *A Course on Point Processes*. *Springer Series in Statistics*. New York: Springer. MR1199815

Walter, G. G. (1992). Approximation of the Delta function by wavelets. *Journal of Approximation Theory* **71**, 328–343. MR1191579

Zadeh, L. A. (1965). Fuzzy sets. *Information and Control* **8**, 338–353. MR0219427

J. A. Fajardo
Núcleo de Sucre
Departamento de Matemáticas
Universidad de Oriente
Escuela de Ciencias
Cumaná 6101
Venezuela
E-mail: jfajardo@udo.edu.ve

R. R. Ríos
Escuela de Matemáticas
Facultad de Ciencias
Universidad Central de Venezuela
Caracas 1040
Venezuela
E-mail: ricardo.rios@ciens.ucv.ve

L. A. Rodríguez
Departamento de Matemáticas
Facultad de Ciencias y Tecnología
Universidad de Carabobo
Valencia
Venezuela
E-mail: larodri@uc.edu.ve