UNIVERSIDAD CARLOS III DE MADRID working papers

Working Paper 04-58
Business Economics Series 21
November 2004

Departamento de Economía de la Empresa
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Spain)
Fax (34-91) 6249608

POINTWISE UNIVERSAL CONSISTENCY OF NONPARAMETRIC LINEAR ESTIMATORS*

Jose M. Vidal-Sanz[1]

Abstract

This paper presents sufficient conditions for pointwise universal consistency of nonparametric delta estimators. We show the applicability of these conditions for some classes of nonparametric estimators.

Keywords: Delta Estimators, Pointwise Universal Consistency, Pointwise approximation.

* I wish to thank Miguel A. Delgado, Juan A. Cuesta for their helpful comments and suggestions. This research has been supported by a Marie Curie Fellowship of the European Community programme IHP, under contract number HPMF-CT-2000-00449.

[1] Dept. of Business Economics, Universidad Carlos III de Madrid, C/ Madrid, 126. 28903 Getafe, Madrid, Spain. E-mail: jvidal@emp.uc3m.es

1. INTRODUCTION

Let P be a probability function in $\left(\mathbb{R}^d, \mathbb{B}^d\right)$ absolutely continuous with respect to the σ-finite measure μ and $f = dP/d\mu$ be the corresponding Radon-Nikodym derivative, which belongs to $L_1\left(\mathbb{R}^d, \mathbb{B}^d, \mu\right)$. Usually, it is considered the Lebesgue's measure λ, with $f = dP/d\lambda$ the corresponding probability density function (pdf), but other possibilities cannot be disregarded. For example, the Lebesgue measure restricted to an interval (e.g., $[-\pi, \pi]^d$ in Fourier series context), or the distribution associated to some control population (e.g., in design of experiments).

Given a random sample of independent observations $\{X_i, 1 = 1, ..., n\}$ from P, a delta estimator of f is defined as,

$$\widehat{f}_n(x) = \frac{1}{n}\sum_{i=1}^{n} K_{m_n}(x; X_i),$$

where $m_n = m(n)$ is known as the smoothing sequence, and $\{K_{m_n}\}_{n\in\mathbb{N}}$ as the generalized kernel sequence. The sequence $\{m_n\}_{n\in\mathbb{N}}$ is not necessarily a sequence of numbers, it may be a sequence of positive definite matrices ordered by decreasing a norm, in the usual kernel estimator of a multivariate density; or the order of a polynomial, in the Fourier series estimator. The smoothing sequence belongs to some directed set $\mathbb{I}$, which is a non empty set endowed with a partial preorder $\leq$, such that if $m_1, m_2 \in \mathbb{I}$, then there exists an $m_3 \in \mathbb{I}$ such that $m_1 \leq m_3$ and $m_2 \leq m_3$. It is assumed that the smoothing sequence $\{m_n\}_{n\in\mathbb{N}}$ diverges in $\mathbb{I}$ as $n \to \infty$, (i.e., for all $M \in \mathbb{I}$ there exists an $n_M \in \mathbb{N}$ such that $m_n \geq M$ for all $n \geq n_M$).

Delta estimators were introduced by Whittle (1958), encompassing all the linear nonparametric estimators of density functions. However, Whittle's (1958) original specification of delta estimators, $\widehat{f}_n(x) = n^{-1}\sum_{i=1}^{n} K_n(x; X_i)$, does not introduce specifically the smoothing parameter m_n which plays a crucial role in consistency arguments. Some examples of delta estimators are,

Estimators	**Generalized Kernel**	**Index set $\mathbb{I}$**
Histograms	$K_m(x,z) = \sum_{A\in m} I_A(x)\, I_A(z)/\lambda(A)$	measurable partitions
Kernels	$K_m(x,z) = \det(m)^{-1}\mathbf{K}\left(m^{-1}(z-x)\right)$	def + matrices,
Biorthonormal Basis	$K_m(x,z) = \sum_{k=1}^{m} a_k(x)\, b_k(z)$	non negative integers

where $I_A(x)$ denotes the characteristic function of the set A (i.e., $I_A(x) = 1$ if $x \in A$, and zero otherwise), $\mathbf{K}$ is integrable and integrates one, and $\{a_k, b_k\}_{k\in\mathbb{N}}$ is a biorthonormal basis on $L_p(\mu) := L_p\left(\mathbb{R}^d, \mathbb{B}^d, \mu\right)$, provided $f \in L_p(\mu)$. Furthermore, many non linear estimators can be approximated, at least asymptotically, by a delta estimator. Terrell (1984) and Terrell and Scott (1992) have shown that all nonparametric density estimators which are continuous and differentiable functionals of the empirical distribution function, can be asymptotically interpreted as delta estimators.

Often, the literature assumes an integrability condition on the pdf (e.g., it belongs to $L_p\left(\mathbb{R}^d, \mathbb{B}^d, \mu\right)$, with $1 < p < \infty$) and a smoothness requirement (e.g., f belongs to a Sobolev space). Watson and Leadbetter (1963), Walter and Blum (1979) and Prakasa Rao (1983) have provided sufficient conditions for global consistency in norm $L_p(\lambda)$ and pointwise consistency of delta estimators. Winter (1973, 1975) has studied uniform consistency and the consistency of the corresponding smooth integrated distribution function estimator. Watson and Leadbetter (1964) have established asymptotic normality. Basawa and Prakasa Rao (1980, Chapter 11) have provided results for dependent observations. In this literature, consistency is achieved under restrictive smoothness conditions on the pdf.

Universal consistency was introduced by Stone (1977), to ensure global L_1-consistency of nonparametric estimators regardless of any smoothness assumption on f. The literature is extensive (for a review, see e.g., Devroye and Györfi (1985) and Devroye (1987) focused on density estimation, Györfi et al (2002) on regression estimation and Devroye et al (1996) on pattern recognition). Usually, universality refers to $L_1(\mu)$ space, but some problems could be confined to other spaces. For example, L_2 is the standard space in nonparametric regression, and L_2 is also the natural framework for density estimation with orthogonal basis. In this context, universality refers to nonsmoothness requirements on the pdf. Universal consistency for delta estimators using L_p norms has been studied in Vidal-Sanz (1999).

The literature on pointwise universal consistency is not so large, and it is focused on the estimation of regression functions, see e.g. Devroye (1981), Greblicki et al (1984) and Walk (2001). In this paper we study the pointwise universal consistency of delta estimators in $L_1(\mu)$.

Definition 1 ***Pointwise Universal Consistency (PUC)****. Let μ be a σ-finite measure in $\left(\mathbb{R}^d, \mathbb{B}^d\right)$, and P a probability function $P << \mu$ (i.e., P absolutely continuous with respect to μ). We say that a delta estimator $\widehat{f}_n$ is strongly (weakly) consistent almost everywhere, if*

$$\left|\widehat{f}_n(x) - f(x)\right| \to 0,$$

almost surely (in probability), for almost every $x \in \mathbb{R}^d$ with respect to the measure μ. We say that the convergence is universal when it holds for all $P << \mu$.

Note that *PUC* property is also relevant for establishing global universal consistency on $L_1(\mu)$, by the Scheffe's Theorem. Some estimators do not satisfy *PUC*, but a weakened version of this property holds; namely, that pointwise consistency is satisfied for all density $f \in L_p(\mu)$, for some $p \in (1, \infty)$. For example, Fourier series estimators do not satisfy *PUC*, but pointwise consistency is satisfied for all density $f \in L_2([-\pi, \pi])$, without smoothness requirements. This weakened form of universality is interesting as pointwise consistency can be used to prove L_p-global consistency using dominated convergence arguments. Though we will not stress this research line, our results can be readily adapted to a $L_p(\mu)$ space.

The aim of this paper is to provide fairly primitive conditions which are sufficient for universal pointwise consistency of delta estimators. To this end, we use the triangular inequality,

$$\left|\widehat{f}_n(x) - f(x)\right| \leq \left|E\left[\widehat{f}_n(x)\right] - f(x)\right| + \left|\widehat{f}_n(x) - E\left[\widehat{f}_n(x)\right]\right|. \qquad (1)$$

The first term on the right hand side is known as the *bias term*, which is deterministic, and the second term as the *variation term*, which is stochastic. In order to study the pointwise universal convergence to zero of the bias term we will consider some functional analysis results related to the approximation theory. In order to study the convergence to zero of the variance term we will use laws of large numbers for triangular arrays.

Section 2 considers pointwise universal unbiasedness. We consider pointwise boundedness of linear operators and provide a characterization for pointwise universal asymptotic unbiasedness. We present some examples that illustrate the application of these results. Section 3 considers sufficient conditions for the weak and strong universal convergence of the variation term. Examples are included to show the application of these conditions.

2. POINTWISE UNIVERSAL UNBIASEDNESS

In this section we study the problem in pointwise sense. Let

$$\alpha_n(f)(x) = \int K_{m_n}(x, z) f(z)\ \mu(dz)$$

be the expected value of $\widehat{f}_n(x)$ with respect to the probability distribution P with pdf f. Notice that α_n is a linear operator, and the estimator $\widehat{f}_n$ is universally asymptotically unbiased in L_1-global sense, for any smoothing number $\{m_n\}_{n\geq 1}$, if and only if $\{\alpha_n\}$ is an approximate identity in $L_1(\mu)$; in other words,

$$\lim_{n\to\infty} \|\alpha_n(f) - f\|_{L_1(\mu)} = 0, \qquad \forall f \in L_1(\mu).$$

Regarding the pointwise convergence, we say that $\alpha_n(f)$ converges almost everywhere (a.e.) to

f, if and only if $|\alpha_n(f)(x) - f(x)| \to 0$ except for sets of μ-null measure; i.e. $\forall\delta > 0$,

$$\lim_{n\to\infty} \mu\left(\left\{x \in \mathbb{R}^d : \sup_{n'\geq n} |\alpha_{n'}(f)(x) - f(x)| > \delta\right\}\right) = 0, \qquad \forall f \in L_1(\mu).$$

To characterize the pointwise approximation property, first we introduce a boundedness condition:

Definition 2 ***Boundedness in measure.*** *Let α_n be a linear operator on $L_1(\mathbb{R}^d, \mathbb{B}^d, \mu)$. We say that α_n is bounded in measure (i.e., it is an operator of weak type-1), if and only if $\forall\varepsilon > 0$, $\exists\delta > 0$ such that,*

$$\sup_{\|f\|_{L_1(\mu)}\leq 1} \mu\left(\left\{x \in \mathbb{R}^d : |\alpha_n(f)(x)| > \delta\right\}\right) \leq \varepsilon.$$

A sequence $\{\alpha_n\}$ of linear operators is uniformly bounded in measure if the maximal operator $\alpha^M(f)(x) = \sup_{n\in\mathbb{N}} |\alpha_n(f)(x)|$ satisfies, $\forall\varepsilon > 0$, $\exists\delta > 0$ such that,

$$\sup_{\|f\|_{L_1(\mu)}\leq 1} \mu\left(\left\{x \in \mathbb{R}^d : \alpha^M(f)(x) > \delta\right\}\right) \leq \varepsilon. \tag{2}$$

If α_n is bounded in norm, then it is bounded in measure, by the Markov's inequality. Notice that the maximal operator is not linear, but a sublinear operator.

Next, we present a Banach-Steinhaus type result, which plays a crucial role for the arguments used in the theory of pointwise approximation. Garsia (1970, Chapter 1) presents some related results. Given a topological space, a G_δ set is a set that can be obtained as a numerable intersection of open sets. Note that in Banach spaces without isolated points, such as $L_1(\mathbb{R}^d, \mathbb{B}^d, \lambda)$, every dense G_δ set is non numerable (see e.g., Rudin 1974, Theorem 5.3.3).

Theorem 1 ***Theorem type Banach-Steinhaus in measure.*** *Let $\{\alpha_n\}$ be a sequence of linear operators in $L_1(\mathbb{R}^d, \mathbb{B}^d, \mu)$, all of them bounded in measure. Then only one of the next statements holds:*

1. *$\{\alpha_n\}_{n\in\mathbb{N}}$ is uniformly bounded in measure,*
2. *$\forall\varepsilon > 0$, $\exists\mathcal{C}_\varepsilon \subset L_1(\mu)$, where $\mathcal{C}_\varepsilon$ is a dense G_δ set, such that,*

$$\mu\left(\left\{x \in \mathbb{R}^d : \alpha^M(f)(x) = \infty\right\}\right) > \varepsilon, \qquad \forall f \in \mathcal{C}_\varepsilon. \tag{3}$$

Proof.

Define the set

$$V_{\overline{\varepsilon}}^{\overline{\delta}} = \left\{f \in L_1(\mu) : \mu\left(\left\{x \in \mathbb{R}^d : \alpha^M(f)(x) > \overline{\delta}\right\}\right) > \overline{\varepsilon}\right\},$$

$\forall\overline{\varepsilon} > 0$ and $\forall\overline{\delta} > 0$. We first prove that this is an open set.

We say that the linear operator α_n is continuous in measure, $n \in \mathbb{N}$, if and only if, for all $\{g_k\}_{k\in\mathbb{N}}, g$ in $L_1(\mu)$ such that $\lim_{k\to\infty} \|g_k - g\|_{L_1(\mu)} = 0$, it is satisfied,

$$\lim_{k\to\infty} \mu\left(\left\{x \in \mathbb{R}^d : |\alpha_n(g_k; x) - \alpha_n(g; x)| > \delta\right\}\right) = 0, \qquad \forall\delta > 0.$$

Since α_n is bounded in measure, it is continuous in measure. Thus, for each $n \in \mathbb{N}$, the sub-linear operator

$$\alpha_n^M(f)(x) = \sup_{n'\leq n} |\alpha_{n'}(f)(x)|$$

is also continuous in measure. Then, $\forall n \in \mathbb{N}$, the sets,

$$\left\{f \in L_1(\mu) : \mu\left(\left\{x \in \mathbb{R}^d : \alpha_n^M(f)(x) > \overline{\delta}\right\}\right) > \overline{\varepsilon}\right\}$$

are open, what implies that $V_{\overline{\varepsilon}}^{\overline{\delta}}$ is open.

Now consider a sequence $\{\delta_k\}_{k\in\mathbb{N}}$ dense in $\mathbb{R}^+$. Thus, $\forall\overline{\varepsilon} > 0$ we have a sequence $\left\{V_{\overline{\varepsilon}}^{\delta_k}\right\}_{k\in\mathbb{N}}$ of open sets. Assume that there exists a $k \in \mathbb{N}$ such that $V_{\overline{\varepsilon}}^{\delta_k}$ is not dense in $L_1(\mu)$. Then $\exists f_0 \in L_1(\mu)$ and $r > 0$ such that $\|f\|_{L_1(\mu)} \leq r$ implies $(f_0 + f) \notin V_{\overline{\varepsilon}}^{\delta_k}$. Thus,

$$\mu\left(\left\{x \in \mathbb{R}^d : \alpha^M(f_0 + f)(x) > \delta_k\right\}\right) \leq \overline{\varepsilon}.$$

$\forall f \in L_1(\mu)$ such that $\|f\|_{L_1(\mu)} \leq r$.

Note that $f = (f_0 + f) - f_0$, so then,

$$\begin{aligned}
\mu\left(\left\{x \in \mathbb{R}^d : \alpha^M(f)(x) > 2\delta_k\right\}\right) &\leq \mu\left(\left\{x \in \mathbb{R}^d : \alpha^M(f_0 + f)(x) > \delta_k\right\}\right) + \\
&\quad + \mu\left(\left\{x \in \mathbb{R}^d : \alpha^M(f_0)(x) > \delta_k\right\}\right) \\
&\leq 2\overline{\varepsilon}.
\end{aligned}$$

Therefore,

$$\sup_{\|f\|_{L_1(\mu)}\leq 1} \mu\left(\left\{x \in \mathbb{R}^d : \alpha^M(f)(x) > 2\delta_k\right\}\right) \leq \frac{2\overline{\varepsilon}}{r},$$

which implies that α^M is bounded in measure, with $\varepsilon = 2\overline{\varepsilon}/r$ and $\delta = 2\delta_k$.

On the other hand, if every $V_{\overline{\varepsilon}}^{\delta_k}$ is dense in $L_1(\mu)$ then

$$\mathcal{C}_{\overline{\varepsilon}} = \bigcap_{k\in\mathbb{N}} V_{\overline{\varepsilon}}^{\delta_k}$$

is a dense G_δ set in $L_1(\mu)$, by the Baire's Theorem (see e.g. Rudin, 1974). Obviously, $\forall f \in \mathcal{C}_{\overline{\varepsilon}}$ we have,

$$\mu\left(\left\{x \in \mathbb{R}^d : \alpha^M(f)(x) > \delta_k\right\}\right) > \varepsilon, \quad \forall\delta_k,$$

and $\{\delta_k\}_{k\in\mathbb{N}}$ is dense in $\mathbb{R}^+$, so that Condition (3) follows.

■

An analogous result to the previous theorem can be established on $L_p\left(\mathbb{R}^d, \mathbb{B}^d, \mu\right)$, with $1 < p < \infty$. For spaces L_p, the uniform boundedness can often be established using an interpolation theorem (see Zygmund (1959,Vol. II, Chapter XII, Section 4), Bergh and Löfström (1976) and Jørsboe and Mejlbro (1982, Theorem 1.9, pp. 8-9)).

The following theorem provides conditions on the generalized kernel sequence $\{K_{m_n}(x,z)\}$, which are sufficient for guaranteeing that the sequence $\{\alpha_n\}$ satisfies a.e. convergence and, therefore, the associated delta estimator is universally asymptotically pointwise unbiased.

Theorem 2 ***Pointwise Approximation Central Theorem.*** *Let $\{\alpha_n\}$ be a sequence of linear operators in $L_1\left(\mathbb{R}^d, \mathbb{B}^d, \mu\right)$. Assume that:*

1. *The sequence $\{\alpha_n\}$ is uniformly bounded in measure.*
2. *$\exists \mathcal{G} \subset L_1(\mu)$ dense, such that, $\alpha_n\left(\widetilde{f}\right) \to \widetilde{f}$ almost everywhere, $\forall \widetilde{f} \in \mathcal{G}$.*

Then, $\{\alpha_n\}$ is an approximate identity in almost everywhere sense, i.e., $\alpha_n(f) \to f$, a.e. $\forall f \in L_1(\mu)$. If the operators $\{\alpha_n\}$ are all bounded in measure on $L_1(\mu)$, then Assumptions 1 and 2 are also necessary.

Proof.

Part I: Sufficient Conditions.

Assume that $\exists \mathcal{G} \subset L_1(\mu)$ dense, such that $\forall \widetilde{f} \in \mathcal{G}$

$$\lim_{n\to\infty} \mu\left(\left\{x \in \mathbb{R}^d : \sup_{n'\geq n}\left|\alpha_{n'}\left(\widetilde{f}\right)(x) - \widetilde{f}(x)\right| > \delta\right\}\right) = 0, \quad \forall\delta > 0.$$

As $\mathcal{G}$ is a dense set, $\forall f \in L_1(\mu)$ and $\forall\varepsilon > 0$, $\exists \widetilde{f} \in \mathcal{G}$ such that $\left\|f - \widetilde{f}\right\|_{L_1(\mu)} \leq \varepsilon$. By the triangular inequality, for each n and each $x \in \mathbb{R}^d$, it is satisfied that

$$\begin{aligned}\sup_{n'\geq n}\left|\alpha_{n'}(f)(x) - f(x)\right| \quad \leq \quad & \sup_{n'\geq n}\left|\alpha_{n'}(f)(x) - \alpha_{n'}\left(\widetilde{f}\right)(x)\right| \\ & + \sup_{n'\geq n}\left|\alpha_{n'}\left(\widetilde{f}\right)(x) - \widetilde{f}(x)\right| + \left|\widetilde{f}(x) - f(x)\right|,\end{aligned}$$

Thus, $\forall f \in L_1(\mu)$ and $\forall \delta > 0$,

$$\begin{aligned}
\mu\left(\left\{x \in \mathbb{R}^d : \sup_{n' \geq n} |\alpha_{n'}(f)(x) - f(x)| > \delta\right\}\right) \leq\ & \mu\left(\left\{x \in \mathbb{R}^d : \sup_{n' \geq n} \left|\alpha_{n'}\left(f - \widetilde{f}\right)(x)\right| > \frac{\delta}{3}\right\}\right) \\
& + \mu\left(\left\{x \in \mathbb{R}^d : \sup_{n' \geq n} \left|\alpha_{n'}\left(\widetilde{f}\right)(x) - \widetilde{f}(x)\right| > \frac{\delta}{3}\right\}\right) \\
& + \mu\left(\left\{x \in \mathbb{R}^d : \left|\widetilde{f}(x) - f(x)\right| > \frac{\delta}{3}\right\}\right).
\end{aligned}$$

The first term is arbitrarily small by uniform boundedness in measure,

$$\begin{aligned}
& \mu\left(\left\{x \in \mathbb{R}^d : \sup_{n' \geq n} \left|\alpha_{n'}\left(f - \widetilde{f}\right)(x)\right| > \frac{\delta}{3}\right\}\right) \leq \mu\left(\left\{x \in \mathbb{R}^d : \alpha^M\left(f - \widetilde{f}\right)(x) > \frac{\delta}{3}\right\}\right) \\
\leq\ & \mu\left(\left\{x \in \mathbb{R}^d : \alpha^M\left(\frac{f - \widetilde{f}}{\left\|f - \widetilde{f}\right\|_{L_1(\mu)}}\right)(x) \cdot \left\|f - \widetilde{f}\right\|_{L_1(\mu)} > \frac{\delta}{3}\right\}\right) \\
\leq\ & \sup_{\|f\|_{L_1(\mu)} \leq 1} \mu\left(\left\{x \in \mathbb{R}^d : \alpha^M(f)(x) > \frac{\delta}{3\varepsilon}\right\}\right) \leq \varepsilon_1.
\end{aligned}$$

Notice that ε_1 can be made arbitrarily small for ε small enough.

Then, $\forall f \in L_1(\mu)$ and $\forall \delta > 0$,

$$\begin{aligned}
& \mu\left(\left\{x \in \mathbb{R}^d : \sup_{n' \geq n} |\alpha_{n'}(f)(x) - f(x)| > \delta\right\}\right) \leq \varepsilon_1 + \\
& + \mu\left(\left\{x \in \mathbb{R}^d : \sup_{n' \geq n} \left|\alpha_{n'}\left(\widetilde{f}\right)(x) - \widetilde{f}(x)\right| > \frac{\delta}{3}\right\}\right) + \frac{\left\|\widetilde{f} - f\right\|_{L_1(\mu)}}{\frac{\delta}{3}} \\
\leq\ & \varepsilon_1 + \mu\left(\left\{x \in \mathbb{R}^d : \sup_{n' \geq n} \left|\alpha_{n'}\left(\widetilde{f}\right)(x) - \widetilde{f}(x)\right| > \frac{\delta}{3}\right\}\right) + \frac{3\varepsilon}{\delta}.
\end{aligned}$$

Since $\varepsilon, \varepsilon_1 > 0$ are arbitrarily small, and

$$\lim_{n \longrightarrow \infty} \mu\left(\left\{x \in \mathbb{R}^d : \sup_{n' \geq n} \left|\alpha_{n'}\left(\widetilde{f}\right)(x) - \widetilde{f}(x)\right| > \frac{\delta}{3}\right\}\right) = 0, \quad \forall \delta > 0,$$

the a.e. approximation follows.

Part II: Necessary Condition.

Assume that $\alpha_n(f) \to f$ a.e. $\forall f \in L_1(\mu)$. Thus, the same property trivially holds for every dense set $\mathcal{G} \subset L_1(\mu)$.

Assume that $\{\alpha_n\}$ is an approximate identity in a pointwise a.e. sense, and that all of the α_n operators are bounded in measure but uniform boundedness in measure is not satisfied. Thus by Theorem 1, $\forall \varepsilon > 0$, $\exists \mathcal{C}_\varepsilon \subset L_1(\mu)$, which is a dense G_δ set, such that

$$\mu\left(\left\{x \in \mathbb{R}^d : \sup_{n \in \mathbb{N}} |\alpha_n(f)(x)| = \infty\right\}\right) > 2\varepsilon, \quad \forall f \in \mathcal{C}_\varepsilon.$$

In other words, $\exists B \subset \mathbb{R}^d$, with $\mu(B) > 2\varepsilon$ such that $\forall x \in B$,

$$\sup_{n\in\mathbb{N}} |\alpha_n(f)(x)| = \infty, \quad \forall f \in \mathcal{C}_\varepsilon.$$

On the other hand, $|f(x)| \overset{a.e.}{<} \infty$ holds $\forall f \in L_1(\mu)$ (in particular, for all $f \in \mathcal{C}_\varepsilon$), because $\exists \delta_\varepsilon > 0$ such that,

$$\mu\left(\left\{x \in \mathbb{R}^d : |f(x)| > \delta_\varepsilon\right\}\right) \leq \frac{\|f\|_{L_1(\mu)}}{\delta_\varepsilon} < \varepsilon.$$

In other words, $\forall \varepsilon > 0$, $\exists A \subset \mathbb{R}^d$ with $\mu(A^c) < \varepsilon$ such that $\sup_{x\in A} |f(x)| < \infty$.

By the triangular inequality,

$$|\alpha_n(f)(x) - f(x)| \geq ||\alpha_n(f)(x)| - |f(x)||.$$

Define $C = A \bigcap B$. Obviously, $\forall x \in C$

$$|\alpha_n(f)(x) - f(x)| \geq ||\alpha_n(f)(x)| - |f(x)|| = \infty, \quad \forall f \in \mathcal{C}_\varepsilon.$$

Notice that $\mu^*(C) > \varepsilon$ since,

$$\mu(B) = \mu\left(A \bigcap B\right) + \mu\left(A^c \bigcap B\right) \leq \mu\left(A \bigcap B\right) + \mu(A^c) = \mu(C) + \mu(A^c),$$

so then,

$$\mu(C) \geq \mu(B) - \mu(A^c) > 2\varepsilon - \varepsilon = \varepsilon.$$

Thus, $\forall \varepsilon > 0$, $\exists \mathcal{C}_\varepsilon \subset L_1(\mu)$, which is a dense G_δ set, such that,

$$\mu\left(\left\{x \in \mathbb{R}^d : \sup_{n\in\mathbb{N}} |\alpha_n(f)(x) - f(x)| = \infty\right\}\right) > \varepsilon, \quad \forall f \in \mathcal{C}_\varepsilon. \tag{4}$$

Since all elements of the sequece $\{\alpha_n\}$ are bounded in measure, the triangular inequality implies that

$$|\alpha_n(f)(x) - f(x)| \leq |\alpha_n(f)(x)| + |f(x)| \overset{a.e.}{<} \infty, \quad \forall n \in \mathbb{N}.$$

Thus $\forall f \in \mathcal{C}_\varepsilon$,

$$\left\{x \in \mathbb{R}^d : \sup_{n\in\mathbb{N}} |\alpha_n(f)(x) - f(x)| = \infty\right\} = \left\{x \in \mathbb{R}^d : \lim_{n\in\mathbb{N}} |\alpha_n(f)(x) - f(x)| = \infty\right\}.$$

Therefore, (4) implies,

$$\mu\left(\left\{x \in \mathbb{R}^d : \lim_{n\in\mathbb{N}} |\alpha_n(f)(x) - f(x)| = \infty\right\}\right) > \varepsilon, \quad \forall f \in \mathcal{C}_\varepsilon,$$

that contradicts the a.e. approximation property.

■

Assume that $\{\alpha_n\}$ satisfies a.e. universal approximation property in $L_1(\mu)$. Then, for all $\{f_r\}_{r\in\mathbb{N}}$, $f \subset L_1(\mu)$ such that $\lim_{r\to\infty} \|f_r - f\|_{L_1(\mu)} = 0$, it is satisfied

$$\lim_{r\to\infty} \lim_{n\to\infty} |\alpha_n(f_r)(x) - f(x)| \xrightarrow{a.e.} 0.$$

The proof is a slight modification of the above result.

Next we present sufficient conditions for the pointwise approximation property. First we define the positive majorized operator of $\alpha_n(f)(x) = \int K_{m_n}(x,z) f(z)\, \mu(dz)$, as the operator

$$|\alpha|_n(f)(x) = \int |K_{m_n}(x,z)| f(z)\, \mu(dz).$$

Theorem 3 ***Sufficient Conditions for Pointwise Approximation.*** *Let $\{\alpha_n\}$ be a sequence of linear operators on $L_1(\mathbb{R}^d, \mathbb{B}^d, \mu)$. Assume that:*

1. *The sequence $\{|\alpha|_n\}$ is uniform bounded in measure,*
2. *$\int K_{m_n}(x,z)\,\mu(dz) \to 1$, a.e.*
3. *$\forall\delta > 0$, $\exists M_\delta > 0$ such that $\sup_{n\in\mathbb{N}} \int_{\|x-z\|<\delta} |K_{m_n}(x,z)|\,\mu(dz) < M_\delta$, a.e.,*
4. *$\int_{\|x-z\|>\delta} |K_{m_n}(x,z)|\,\mu(dz) \to_{a.e.} 0$, $\forall\delta > 0$.*

Then $\alpha_n(f) \to f$ a.e. for all $f \in L_1(\mu)$.

Proof.

First, we prove that if $\{|\alpha|_n\}$ is uniformly bounded in measure, then $\{\alpha_n\}$ also is uniformly bounded in measure. As the maximal operators satisfy,

$$\alpha^M(f)(x) = \sup_{n\in\mathbb{N}} |\alpha_n(f)(x)| \le \sup_{n\in\mathbb{N}} \int |K_{m_n}(x,z)|\,|f(z)|\,\mu(dz) = |\alpha|^M(|f|)(x),$$

with $|\alpha|^M = \sup_{n\in\mathbb{I}} |\alpha|_n$. Then, $\forall\delta > 0$,

$$\mu\left(\left\{x \in \mathbb{R}^d : \alpha^M(f)(x) > \delta\right\}\right) \le \mu\left(\left\{x \in \mathbb{R}^d : |\alpha|^M(|f|)(x) > \delta\right\}\right).$$

Taking the supremum in the unit ball $\|f\|_{L_1(\mu)} \le 1$ the result follows.

Let $C_c(\mathbb{R}^d)$ be the set of continuous and compactly supported functions. Next, we prove the approximation property for any $f \in L_1(\mu)$ with some version in $C_c(\mathbb{R}^d)$. As $C_c(\mathbb{R}^d)$ is a dense set in $L_p(\mu)$, $1 \le p < \infty$, the result follows from Theorem 2. We proceed in 2 steps.

Step 1) For all $\delta > 0$ and all $h(x,z) \in C_c\left(\mathbb{R}^d \times \mathbb{R}^d\right)$, it is satisfied that

$$\left|\int_{\{z:\|x-z\|>\delta\}} h(x,z) K_{m_n}(x,z)\,\mu(dz)\right| \leq \|h\|_\infty \cdot \int_{\{z:\|x-z\|>\delta\}} |K_{m_n}(x,z)|\,\mu(dz) \xrightarrow{a.e.} 0,$$

using Assumption 4, and $\|h\|_\infty < \infty$.

Step 2) We prove that for all $f \in L_1(\mu)$ with some version in $C_c\left(\mathbb{R}^d\right)$, the sequence $\alpha_n(f) \to f$ a.e. By the triangular inequality,

$$\begin{aligned} \sup_{n'\geq n} |\alpha_{n'}(f)(x) - f(x)| \quad \leq \quad & \sup_{n'\geq n} \left|\int (f(z) - f(x))\, K_{m_{n'}}(x,z)\,\mu(dz)\right| \\ & + \sup_{n'\geq n} \left|\int K_{m_{n'}}(x,z)\,\mu(dz) f(x) - f(x)\right|. \end{aligned}$$

By Assumption 2,

$$\begin{aligned} & \sup_{n'\geq n} |\alpha_{n'}(f)(x) - f(x)| \leq \sup_{n'\geq n} \left|\int (f(z) - f(x))\, K_{m_{n'}}(x,z)\,\mu(dz)\right| \\ & + \|f\|_\infty \sup_{n'\geq n} \left|\int K_{m_{n'}}(x,z)\,\mu(dz) - 1\right| \\ = \quad & \sup_{n'\geq n} \left|\int (f(z) - f(x))\, K_{m_{n'}}(x,z)\,\mu(dz)\right| + o(1), \end{aligned}$$

where the $o(1)$ convergence holds in a.e. sense. Then,

$$\begin{aligned} \sup_{n'\geq n} |\alpha_{n'}(f)(x) - f(z)| \quad \leq \quad & \sup_{n'\geq n} \left|\int_{\{z:\|x-z\|\leq\delta\}} (f(z) - f(x))\, K_{m_{n'}}(x,z)\,\mu(dz)\right| + \\ & + \sup_{n'\geq n} \left|\int_{\{z:\|x-z\|>\delta\}} (f(z) - f(x))\, K_{m_{n'}}(x,z)\,\mu(dz)\right| + o(1). \end{aligned}$$

As f is uniformly continuous, $\forall\varepsilon > 0$, $\exists\delta > 0$ such that $\|x - z\| \leq \delta$ implies that $|f(x) - f(z)| \leq \varepsilon$, applying Assumption 3 we obtain,

$$\sup_{n'\geq n} |\alpha_{n'}(f)(x) - f(x)| \overset{a.e.}{\leq} \varepsilon \cdot M_\delta + \sup_{n'\geq n} \left|\int_{\{z:\|x-z\|>\delta\}} h(x,z)\, K_{m_{n'}}(x,z)\,\mu(dz)\right| + o(1)$$

with $h(x,z) = (f(z) - f(x))$. The first term on the right hand side is arbitrarily small, whilst the second term tends to zero a.e. by Step 1, and the result follows.

■

A sufficient condition for Assumption 4 in Theorem 3 is: for some $s \geq 1$, it is satisfied that

$$\lim_{n\to\infty} \mu\left(\left\{x \in \mathbb{R}^d : \sup_{n'\geq n} \int \left|K_{m_{n'}}(x,z)\right| \|x - z\|^s \mu(dz) > \delta\right\}\right) = 0,$$

for all $\delta > 0$. This is a consequence of $I_{\{\|x-z\|>\delta\}}(z) < \|x-z\|^s \cdot \delta^{-s}$, and since $|\alpha|_n$ is a monotone operator, then $\forall\delta > 0$

$$\sup_{n'\geq n} |\alpha|_{n'} \left(I_{\{\|x-z\|>\delta\}}(z)\right)(x) < \delta^{-s} \sup_{n'\geq n} |\alpha|_{n'} \left(\|x-z\|^s\right)(x).$$

Theorems 2 and 3 can be applied to the most popular nonparametric estimators, using the Hardy-Littlewood-Paley theory. The Hardy-Littlewood maximal operator on $L_1\left(\mathbb{R}^d, \mathbb{B}^d, \lambda\right)$ defined as

$$\beta^*(f,x) = \sup_{\varepsilon>0} \frac{1}{\lambda(B(x,\varepsilon))} \int_{B(x,\varepsilon)} f(z)\,dz,$$

with $B(x,\varepsilon)$ the ε-ball, satisfies for some $c_d > 0$, $\|\beta^*(f,x)\|_{L_p(\lambda)} \leq c_d \|f\|_{L_1(\lambda)}$ for all $f \in L_1$; and therefore $\beta^\varepsilon(f,x) = f(z)\, I(B(x,\varepsilon)) / \lambda(B(x,\varepsilon))$ is uniformly bounded in measure. For details, see Stein (1970), de Guzman (1975) and Wheeden and Zygmund (1977).

Example 1 *Consider the kernel estimator in* $L_1\left(\mathbb{R}^d, \mathbb{B}^d, \lambda\right)$*, defined by means of,*

$$K_m(x,z) = \frac{1}{\det(m)} \mathbf{K}\left(m^{-1}(z-x)\right), \tag{5}$$

and a smoothing sequence $\{m_n\} \subset \mathbb{I}$*, where* $\mathbb{I}$ *is the set of positive definite matrices. If there exists a closed interval* $C \subset \mathbb{R}^d$ *such that* $c_1 I_C(u) \leq |\mathbf{K}(u)| \leq c_2 I_C(u)$ *for some* $c_1, c_2 > 0$ *then,*

$$\int \sup_{m\in\mathbb{I}} \int |K_m(x,z)|\, f(z)\,dzdx \leq c\,\|f\|_{L_1(\lambda)},$$

by the Hardy-Littlewood argument, so that kernel operators are uniformly bounded in measure. The pointwise universal unbiasedness readily follows from Theorem 3.

Example 2 *Define the set* $\mathbb{I}_0$ *of regular partitions of* $\mathbb{R}^d$ *as the set of Borel measurable partitions* m *of finite diameter, satisfying* $\inf_{A\in m} \lambda(A) > 0$*, such that the maximum diameter of the partition tends to zero as partitions become thinner, and all subsets form a* Vitali system *(the definition can be found in, e.g., Shilov and Gurevich, 1997). Consider the histogram in* $L_1\left(\mathbb{R}^d, \mathbb{B}^d, \lambda\right)$*, with kernel*

$$K_m(x,z) = \sum_{A\in m} \frac{I_A(x)\, I_A(z)}{\lambda(A)},$$

defined for $\{m_n\} \subset \mathbb{I}_0$*. Using that*

$$\beta^*(f,x) = \sup_{\varepsilon>0} \frac{P_f(B(x,\varepsilon))}{\lambda(B(x,\varepsilon))}$$

satisfies $\|\beta^*(f,x)\|_{L_1(\lambda)} \leq c_d \|f\|_{L_1(\lambda)}$*, then*

$$\begin{aligned} &\int \left(\sup_{n\in\mathbb{N}} \int \left(\sum_{A\in m_n} \frac{I_A(x)\, I_A(z)}{\lambda(A)}\right) f(z)\,dz\right) dx \\ = &\int \sup_{n\in\mathbb{N}} \sum_{A\in m_n} \frac{I_A(x)\, P_f(A)}{\lambda(A)} dx \leq c\,\|f\|_{L_1(\lambda)}, \end{aligned}$$

and the operators are uniformly bounded in measure. The pointwise universal unbiasedness follows from an argument analogous to Györfi et al (2002, Lemma 24.5), which is related to the Lebesgue density theorem,

$$\lim_{n\to\infty}\sum_{A\in m_n}\frac{P_f(A)}{\lambda(A)}I_A(x)=f(x), \qquad a.e.$$

Alternatively we can apply Theorem 2 to prove that the approximation theory is satisfied for all simple functions $\mathcal{S}\subset L_1\left(\mathbb{R}^d,\mathbb{B}^d,\lambda\right)$, which is a dense class in L_1. If $g\in\mathcal{S}$, then $g(z)=\sum_{r=1}^{s}\beta_r\cdot I_{B_r}(z)$, for some finite measurable partition $\overline{m}=(B_1,...,B_s)$ of $\mathbb{R}^d$, with $\lambda(B_r)<\infty$ for $r=1,...,s$. By definition,

$$\begin{aligned}\alpha_n(g)(x) &= \sum_{A\in m_n}\left(\frac{1}{\lambda(A)}\int_A g(z)\lambda(dz)\right)I_A(x)\\ &= \sum_{A\in m_n}\left(\sum_{r=1}^{s}\beta_r\frac{1}{\lambda(A)}\int_A I_{B_r}(z)\lambda(dz)\right)I_A(x)\\ &= \sum_{A\in m_n}\left(\sum_{r=1}^{s}\beta_r\frac{\lambda(A\cap B_r)}{\lambda(A)}\right)I_A(x).\end{aligned}$$

Thus, using that $\sum_{A\in m_n}I_A(x)=1$, a.e.,

$$\begin{aligned}&\lambda\left(\{|\alpha_n(g)(x)-g(x)|>\delta\}\right)\\ =\ &\lambda\left(\sup_{n'\geq n}\left|\sum_{A\in m_{n'}}\sum_{r=1}^{s}\beta_r\frac{\lambda(A\cap B_r)}{\lambda(A)}I_A(x)-\sum_{r=1}^{s}\beta_r I_{B_r}(x)\right|>\delta\right)\\ \leq\ &\lambda\left(\sup_{n'\geq n}\sum_{A\in m_{n'}}\frac{1}{\lambda(A)}\left|\sum_{r=1}^{s}\beta_r\left(\lambda(A\cap B_r)-\lambda(A)I_{B_r}(x)\right)I_A(x)\right|>\delta\right).\end{aligned}$$

Next we prove that this measure tends to zero. If $m_n\geq\overline{m}$, i.e. m_n is thinner than $\overline{m}$, then $\forall B_r\in\overline{m}$ and $\forall A\in m$, and therefore we have one of the following cases: (i) either $A\cap B_r=\emptyset$ and therefore $\lambda(A\cap B_r)=0$, $I_{\{A\cap B_r\}}(x)=0$, or (ii) $A\subset B_r$ and thus $\lambda(A\cap B_r)=\lambda(A)$, $I_{A\cap B_r}(x)=I_A(x)$ so that

$$|\lambda(A\cap B_r)I_A(x)-\lambda(A)I_{A\cap B_r}(x)|=0.$$

Thus, $\forall g\in\mathcal{S}$, $\exists\overline{m}$ such that $\sup_{m\geq\overline{m}}|\alpha_m(g)(x)-g(x)|=0$, except for sets of null measure, and the result follows.

Example 3 *We also consider the almost everywhere convergence of the Dirichlet's approximate identity $\{\alpha_n\}$ related to the Fourier sums in $L_p([-\pi,\pi])$, with $1\leq p<\infty$. This operator can be* *expressed by*

$$\alpha_n(f)(x) = \frac{1}{2\pi}\int_{-\pi}^{\pi}\left(\frac{\sin\left(\left(m_n+\frac{1}{2}\right)(z-x)\right)}{\sin\left(\frac{1}{2}(z-x)\right)}\right)f(z)\,dz.$$

with $\{m_n\} \subset \mathbb{N}$. *A detailed exposition about Fourier sums can be seen, e.g. in Bary (1964), Zygmund (1959) and Edwards (1979). Using Theorem 2, we only need to establish a.e. convergence for a dense set of functions and uniform boundedness in measure.*

- *The trigonometric polynomials are a dense subspace in* $L_p([-\pi,\pi])$ *with* $1 \leq p < \infty$ *and the Fourier sums of trigonometric polynomials converges a.e. to the respective polynomials. See, e.g. Mozzochi (1970, pp. 9), Jørsboe and Mejlbro (1982, pp. 17-20), and Arias de Reyna (2002, Part II).*

- *The Carleson-Hunt Theorem establishes that the Fourier sums are uniform bounded in measure in the space* $L_p([-\pi,\pi])$, *with* $1 < p < \infty$. *This result was first proved by Carleson (1966) for* $p = 2$, *and extended to the case* $1 < p < \infty$ *by Hunt (1968). The original Carleson-Hunt Theorem proves that,*

$$\sup_{\|f\|_{L_p([-\pi,\pi])}\leq 1}\left\|\alpha^M(f)\right\|_{L_p([-\pi,\pi])} < \infty. \tag{6}$$

  Then, by Markov's inequality, (6) implies the result.

Thus, Theorem 2 implies that the Fourier sums satisfies the almost everywhere approximation property for every curve in $L_p([-\pi,\pi])$ with $1 < p < \infty$. The proof of (6) presents great technical difficulties. Monographs of Mozzochi (1971), Jørsboe and Mejlbro (1982) and Arias de Reyna (2002) are devoted to self-contained proofs. Garsia (1970) studies a simplification of Carleson's result. In Fefferman (1971) and Sjölin (1971) the Carleson Hunt theorem is extended to dimensions $d > 1$.

However, in $L_1([-\pi,\pi])$ the Fourier sums are bounded in measure, but they are not uniformly bounded in measure. As a consequence of Theorem 2, the almost everywhere approximation fails. This is a well known problem. A very famous counter-example due to Kolmogorov (1926), shows that for some function in $L_1([-\pi,\pi])$ the Fourier sum diverges almost everywhere. Some additional results on pointwise divergence can be seen in Körner (1981), Edwards (1979, pp. 80) and the monograph of Zygmund (1959, sec. 8.4). As we can see in the proof of Theorem 2, there is a dense G_δ set of functions in $L_1([-\pi,\pi])$ that $\alpha_n(f)(x)$ diverges almost everywhere. Since any dense G_δ set in $L_1([-\pi,\pi])$ is non numerable, the curve considered by Kolmogorov is just one in the dense and uncountable set of functions with divergence problems.

3. POINTWISE CONVERGENCE OF THE VARIATION TERM

In this section, our aim is to prove that

$$\left|\widehat{f}_n(x) - E\left[\widehat{f}_n(x)\right]\right| = n^{-1}\sum_{i=1}^{n}\left(K_{m_n}(x, X_i) - E\left[K_{m_n}(x, X_i)\right]\right) \to 0,$$

almost surely (in probability) for almost every $x \in \mathbb{R}^d$ with respect to μ, which is immediate by using a simple LLN for triangular arrays. As usual, a condition on the smoothing number $\{m_n\}$ is necessary in order to prove consistency.

Proposition 1 ***Universal Pointwise Weak Consistency of Variation Term.*** *Assume that for all probability P with $f = dP/d\mu \in L_1(\mu)$, the triangular array $\{K_{m_n}(x, X_i) : 1 \le i \le n\}_{n\in\mathbb{N}}$ satisfies that for some $r > 1$,*

$$E\left[\left|K_{m_n}(x, X)\right|^r\right] = o\left(n^{(r-1)}\right), \tag{7}$$

almost everywhere $[\mu]$. Then,

$$\begin{aligned} E\left[\left|\widehat{f}_n(x) - E\left[\widehat{f}_n(x)\right]\right|^r\right] &\to 0, \\ \left|\widehat{f}_n(x) - E\left[\widehat{f}_n(x)\right]\right| &\to {}_p 0, \end{aligned}$$

almost everywhere $[\mu]$, with $f = dP/d\mu$, and the result holds universally in P.

Proof.

Define $Z_{n,i} = K_{m_n}(x; X_i)$, then by Markov's, c_r and Jensen inequalities,

$$E\left[\left|\widehat{f}_n(x) - E\left[\widehat{f}_n(x)\right]\right|^r\right] \le 2^{r-1}\frac{\sum_{i=1}^{n} E\left[\left|Z_{n,i} - E\left[Z_{n,i}\right]\right|^r\right]}{n^r} \le \frac{2^r\sum_{i=1}^{n} E\left[\left|Z_{n,i}\right|^r\right]}{n^r} \to 0.$$

The result is immediate.

■

The following examples illustrate the application of the previous result.

Example 4 *Consider the kernel estimator (5), with $\mathbf{K} \in L_r\left(\mathbb{R}^d, \mathbb{B}^d, \lambda\right)$ for some $r > 1$. Then, for all integrable density f,*

$$n^{-(r-1)}E\left[\left|K_{m_n}(x, X)\right|^r\right] = \frac{1}{n^{(r-1)}\det(m_n)^r}\int\left|\mathbf{K}\left(m_n^{-1}(z - x)\right)\right|^r f(z)\,\lambda(dz)$$

$$= \frac{1}{\left[n\cdot\det(m_n)\right]^{(r-1)}}\int\left|\mathbf{K}(u)\right|^r f(x + m_n u)\,du = O\left(\frac{f(x)\int\left|\mathbf{K}(u)\right|^r du}{\left[n\cdot\det(m_n)\right]^{(r-1)}}\right),$$

for a.e. $x \in \mathbb{R}^d$, by the dominated convergence Theorem. It tends to zero when $n\cdot\det(m_n) \to \infty$.

Example 5 *Consider the Histogram in* $L_1\left(\mathbb{R}^d, \mathbb{B}^d, \lambda\right)$, *for regular partitions. Notice that for any partition* $m \in \mathbb{I}_0$, *it is satisfied that* $|K_m(x,z)|^2 = \sum_{A\in m} |I_A(x) I_A(z)/\lambda(A)|^2$ *a.e., since the sets in the partition* m *are disjoint. Define,*

$$\gamma(m) = \inf_{A\in m} \lambda(A) > 0.$$

The condition $n\cdot\gamma(m_n) \to \infty$ *implies that*

$$\begin{aligned} n^{-1}E\left[|K_{m_n}(x,X)|^2\right] &= \frac{1}{n}E\left[\sum_{A\in m_n}\left|\frac{I_A(x)\,I_A(X)}{\lambda(A)}\right|^2\right] \\ = \frac{1}{n}\sum_{A\in m_n}\frac{P(A)}{\lambda(A)^2}I_A(x) &\leq \frac{1}{n\cdot\gamma(m_n)}\sum_{A\in m_n}\frac{P(A)}{\lambda(A)}I_A(x) \\ = \frac{1}{n\cdot\gamma(m_n)}E\left[\widehat{f}_n(x)\right] &\to 0, \end{aligned}$$

a.e, as $\widehat{f}_n$ *is pointwise universally unbiased.*

Example 6 *Consider the Dirichlet kernel in* $L_p([-\pi,\pi])$, *with real* $p>1$. *Let*

$$K_{m_n}(u) = \frac{\sin((2m_n+1)\,u/2)}{2\pi\sin(u/2)}.$$

Observe that,

$$\begin{aligned} 2\pi K_m(u) &= \cot\left(\frac{u}{2}\right)\sin(mu) + \cos(mu) \\ &= \frac{2}{u}\sin(mu) + \left(\cot\left(\frac{u}{2}\right) - \frac{2}{u}\right)\sin(mu) + \cos(mu) \end{aligned}$$

and $\cot(t) - t^{-1}$ *is bounded on* $(-\pi/2, \pi/2)$, *hence*

$$K_m(u) = \frac{1}{\pi}\frac{\sin(mu)}{u} + O(1).$$

Since $|\sin(mu)| \leq |mu|$,

$$\begin{aligned} n^{-1}E\left[|K_{m_n}(x,X)|^2\right] &= \frac{1}{n}\int_{-\pi}^{\pi}|K_{m_n}(u)|^2 f(u-x)\,du \\ &\leq \frac{1}{n}\int_{-\pi}^{\pi}\left|\frac{m_n|u|}{\pi u}\right|^2 f(u-x)\,du + O\left(\frac{1}{n}\right) \\ &= \frac{m_n^2}{\pi^2 n} + O\left(\frac{1}{n}\right) \end{aligned}$$

a.e., and weak universal consistency follows from condition $m_n^{-1} + m_n^2/n \to 0$.

The next result establishes strong consistency using a logarithmic growth rate on the smoothing numbers. We illustrate its application with some examples.

Theorem 4 ***Universal Pointwise Strong Consistency of Variation Term****. Assume that for any probability function P with $f = dP/d\mu \in L_1(\mu)$,*

$$\sum_{n=1}^{\infty} \exp\left\{ \frac{-n}{M_n(x)^2} \right\} < \infty, \qquad a.e.\,[\mu], \tag{8}$$

where $M_n(x) = ess\sup_z |K_{m_n}(x,z)|$. Then, universal complete pointwise convergence is satisfied a.e. $[\mu]$, and

$$\Pr\left\{ \lim_{n\to\infty} \left| \widehat{f}_{m_n}(x) - E\left[\widehat{f}_{m_n}(x)\right] \right| > \varepsilon \right\} = 0, \qquad \forall \varepsilon > 0,$$

a.e. $[\mu]$ universally in P.

Proof.

The result is a consequence of Hoeffding's inequality (see, e.g. Györfi et al 2002). Let consider $Z_{n,i} = K_{m_n}(x,z)$. By assumption, $Z_{n,i} \in [-M_n(x), M_n(x)]$ for $i = 1,..,n$ with probability one. Therefore,

$$\Pr\left[\left| \frac{1}{n} \sum_{i=1}^{n} (Z_{n,i} - E[Z_{n,i}]) \right| > \lambda \right] \leq \exp\left\{ \frac{-2n\lambda^2}{\frac{1}{n}\sum_{i=1}^{n} (2M_n(x))^2} \right\} = \exp\left\{ \frac{-n\lambda^2}{2M_n(x)^2} \right\},$$

and the result follows from the Borel-Cantelli Lemma. ■

Example 7 *Consider the kernel estimator. If $\mathbf{K}(u)$ has a global maximum at $u = 0$, then*

$$M_n(x) = \sup_{z\in\mathbb{R}^d} |K_{m_n}(z-x)| = K_{m_n}(0) = \frac{\mathbf{K}(0)}{\det(m_n)},$$

and the condition in expression (8) is satisfied if $\sum_{n=1}^{\infty} \exp\left\{-n\det(m_n)^2\right\} < \infty$, for which it suffices that $n\det(m_n)^2/\log n \to \infty$.

Example 8 *The histogram satisfies,*

$$M_n(x) = \sup_{z\in\mathbb{R}^d} \left| \sum_{A\in m_n} \frac{I_A(x)\, I_A(z)}{\lambda(A)} \right| = \sum_{A\in m_n} \frac{I_A(x)}{\lambda(A)} \leq \frac{\sum_{A\in m_n} I_A(z)}{\gamma(m_n)} = \frac{1}{\gamma(m_n)},$$

and the condition in expression (8) is satisfied if $\sum_{n=1}^{\infty} \exp\left\{-n\gamma(m_n)^2\right\} < \infty$, for which it suffices that $n\gamma(m_n)^2/\log n \to \infty$.

Example 9 *Consider the Dirichlet kernel in* $L_p([-\pi,\pi])$, *with real* $p>1$. *Let*

$$M_n(x)=\sup_{u\in[-\pi,\pi]}\left|\frac{\sin((2m_n+1)u/2)}{2\pi\sin(u/2)}\right|\leq\frac{1}{\pi}\sup_{u\in[-\pi,\pi]}\left|\frac{\sin(m_n u)}{u}\right|\leq\frac{m_n}{\pi},$$

and the condition in expression (8) is satisfied if $\sum_{n=1}^{\infty}\exp\left\{-n/m_n^2\right\}<\infty$, *for which it suffices that* $m_n^2(\log n)/n\to 0$.

Acknowledgments

I wish to thank Miguel A. Delgado, Juan A. Cuesta and two anonymous reviewers for their helpful comments and suggestions. This research has been supported by a Marie Curie Fellowship of the European Community programme IHP, under contract number HPMF-CT-2000-00449.

REFERENCES

[1] Arias de Reyna, J. (2002): *Pointwise Convergence of Fourier Series.* Springer Verlag, New York.

[2] Basawa, I. V. and B. L. S. Prakasa Rao (1980): *Statistical Inference for Stochastic Processes.* New York, Academic Press.

[3] Bary, N. K. (1964): *A Treatise on Trigonometric Series.* Vols **1** and **2**. Oxford, Pergamon Press.

[4] Bergh, J. and J. Löfström (1976): *Interpolation Spaces. An Introduction.* In: *Grundlehren der Mathemastischen Wissenschaften,* **223**. Berlin, Springer Verlag.

[5] Carleson, L. (1966): "On convergence and Growth of Partial Sums of Fourier Series". *Acta Mathematica*, **116**, 135-157.

[6] de Guzman, M. (1975): *Differentiation of Integrals in* $\mathbb{R}^n$. Lecture Notes in Mathematics, **481**, Springer Verlag, Berlin.

[7] Devroye, L. (1981): "On the Almost Everywhere Convergence of Nonparametric Regression Function Estimates". *Annals of Statistics*, **9**, 1310-1319.

[8] Devroye, L. (1987): *A Course in Density Estimation.* Progress in Probability and Statistics. Boston, Birkhauser.

[9] Devroye, L. and Györfi, L. (1985): *Nonparametric Density Estimation. The* L_1 *View.* New York, John Wiley & Sons.

[10] Devroye, L.; L. Györfi and G. Lugosi (1996): *A Probabilistic Theory of Pattern Recognition.* Applications of Mathematics, Stochastic Modelling and Applied Probability. Springer Verlag, New York.

[11] Edwards, R. E. (1979): *Fourier Series. A modern Introduction.* Vol **1**. Berlin, Springer Verlag.

[12] Fefferman, C. (1971): "On the Convergence of Multiple Fourier Series". *Bulletin of the American Mathematical Society*, **77**, 744-755.

[13] Garsia, A. M. (1970): *Topics in Almost Everywhere Convergence.* Chicago, Markham Publishing Company.

[14] Greblicki, W.; A. Krzyzak and M. Pawlak (1984): "Distribution-Free Pointwise Consistency of Kernel Regression Estimate". In: Short Communications, *Annals of Statistics*, **12**, 1570-1575.

[15] Györfi, L.; M. Kohler; A. Kryżak and H. Walk (2002): *A Distribution Free Theory of Nonparametric Regression.* Springer Verlag, New York.

[16] Hunt, R. A. (1968): "On the Convergence of Fourier Series". *Orthogonal Expansions and Their Continuous Analogues*, Proceedings of a conference held at Edwardsville, Illinois. Southern Illinois University Press, Carbondale, pp. 235-255.

[17] Jørsboe, O. G. and L. Mejlbro (1982): *The Carleson Hunt Theorem of Fourier Series.* Berlin, Springer Verlag.

[18] Kolmogorov, A. N. (1926): "Une série de Fourier-Lebesgue divergente partout". *Comptes Rendes de lácadémie des Sciences de Paris*, **183**, 1237-1328.

[19] Körner, T. W. (1981): "Everywhere Divergent Fourier Series". *Colloquium in Mathematics*, **45**, 103-118.

[20] Mozzochi, C. J. (1971): *On the Pointwise Convergence of Fourier Series.* Springer Verlag, Berlin.7.

[21] Prakasa Rao, B. L. S. (1983): *Nonparametric Functional Estimation.* Londres, Academic Press.

[22] Rudin, W. (1974): *Real and Complex Analysis.* (2nd ed.). New York, Mc Graw Hill.

[23] Shilov, G. E. and B. L. Gurevich (1997): *Integral, Measure & Derivative: A Unified Approach.* Dover, New York.

[24] Sjölin, P. (1971): "Convergence Almost Everywhere of Certain Singular Integrals and Multiple Fourier Series". *Arkiv fü Mathematik*, 8, 65-90.

[25] Stein, E. M. (1970): *Singular Integrals and Differentiability Properties of Functions.* Princeton University Press, Princeton, New Jersey.

[26] Stone, C. J. (1977): "Consistent Nonparametric Regression". *Annals of Statistics*, **5**, 595-645.

[27] Terrell, G. R. (1984): "Eficiency of Nonparametric Density Estimators". *Technical Report.* Department of Math. Sciences, Rice University.

[28] Terrell, G. R. and D. W. Scott (1992): "Variable Kernel Density Estimation". *Annals of Statistics*, **20**, 1236-1265.

[29] Vidal-Sanz, J.M. (1999): *Universal consistency of delta estimators.* PhD Thesis. Universidad Carlos III de Madrid.

[30] Walk, H. (2001): "Strong universal pointwise consistency of recursive regression estimates". *Ann. Inst. Statist. Math.,* **53**, 691-707.

[31] Walter, G. and J. R. Blum (1979): "Probability Density Estimation using Delta Sequences". *Annals of Statistics*, **7**, 328-340.

[32] Watson, G. S. y M. R. Leadbetter (1963): "On the Estimation of Probability Density I". *Annals of Statistics*, 34, 480-491.

[33] Watson, G. S. and M. R. Leadbetter (1964): "Hazard analysis II". *Shankhyā, ser. A*, **26**, 101-116.

[34] Wheeden, R. and A. Zygmund (1977): *Measure and Integral.* Marcel Dekker, New York.

[35] Whittle, P. (1958): "On Smoothing a Probability Density Function". *J. of the Royal Statistical Soc., ser. B*, **20**, 334-343.

[36] Winter, B. B. (1973): "Strong uniform Consistency of Integrals of Density Estimation". *Canad. J. Statist.*, **1**, 247-253.

[37] Winter, B. B. (1975): "Rate of Strong Consistency of Two Nonparametric Density Estimators". *Annals of Statistics*, **3**, 759-766.

[38] Zygmund, A (1959): *Trigonometric Series.* Vols **I, II**. Cambridge, Cambridge University Press.