View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Research Papers in Economics

Bernoulli **11**(6), 2005, 971–985

Pointwise universal consistency of nonparametric density estimators

JOSE M. VIDAL-SANZ

Department of Business Economics, Universidad Carlos III de Madrid, Calle Madrid 126, 28903 Madrid, Spain. E mail: jvidal@emp.uc3m.es

This paper presents sufficient conditions for pointwise universal consistency of nonparametric delta estimators and shows the application of these conditions for some classes of nonparametric estimators.

Keywords: delta estimators; pointwise approximation; pointwise universal consistency

1. Introduction

Let P be a probability function in $(\mathbb{R}^d, \mathbb{B}^d)$ which is absolutely continuous with respect to the σ-finite measure μ, and let $f = \mathrm{d}P/\mathrm{d}\mu \in L_1(\mathbb{R}^d, \mathbb{B}^d, \mu)$ be the corresponding Radon–Nikodym derivative. Usually the Lebesgue measure λ is considered and f is the associated probability density function (pdf), but other measures cannot be disregarded – for example, the restriction of λ to some interval (such as $[-\pi, \pi]^d$ in Fourier series framework), or the distribution associated with some control population in the design of experiments context. Given a random sample of independent observations $\{X_i, i = 1, \ldots, n\}$ from P, a delta estimator of f is defined as

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

where $\{m_n\}_{n\in\mathbb{N}}$ is known as a smoothing sequence and $\{K_{m_n}\}_{n\in\mathbb{N}}$ as a generalized kernel sequence. The smoothing sequence belongs to some directed set $\mathbb{I}$, which is a non-empty set endowed with a partial preorder $\leqslant$, such that if $m_1, m_2 \in \mathbb{I}$, then there exists an $m_3 \in \mathbb{I}$ such that $m_1 \leqslant m_3$ and $m_2 \leqslant m_3$. 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 $m_n \geqslant M$ for all $n \geqslant n_M$). For example, $\{m_n\}$ is a sequence of positive definite matrices ordered by decreasing a norm in kernel estimation of multivariate densities, whilst m_n is the order of a polynomial in Fourier series estimators.

Delta estimators were introduced by Whittle (1958), encompassing all the linear nonparametric estimators of density functions. The asymptotic unbiasedness of these estimators requires that the limit of $\mathrm{E}[\hat{f}_n(x)]$ can be thought of as $\int \delta(z-x)f(z)\lambda(\mathrm{d}z)$, where δ is the Dirac delta generalized function with a jump at zero, and this is the reason for the name 'delta estimator'. Some examples of delta estimators are given in Table 1, where $I_A(x)$ denotes the characteristic function of the set A (i.e., $I_A(x) = 1$ if $x \in A$, and

1350 7265 © 2005 ISI/BS

Table 1. Examples of delta estimates

Estimators	Generalized kernel	Index set $\mathbb{I}$
Histograms	$K_m(x, z) = \sum_{A \in m} I_A(x) I_A(z)/\lambda(A)$	Countable measurable partitions
Kernels	$K_m(x, z) = \det(m)^{-1}\mathbf{K}(m^{-1}(z - x))$	Positive definite matrices
Biorthonormal basis	$K_m(x, z) = \sum_{k=1}^{m} a_k(x) b_k(z)$	Non negative integers

zero otherwise), $\mathbf{K}$ is integrable and integrates to unity, and $\{a_k, b_k\}_{k \in \mathbb{N}}$ is a biorthonormal basis on $L_p(\mu) := L_p(\mathbb{R}^d, \mathbb{B}^d, \mu)$, provided $f \in L_p(\mu)$. Furthermore, many nonlinear 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.

Watson and Leadbetter (1963), Walter and Blum (1979) and Prakasa Rao (1983) have provided sufficient conditions for global L_p-consistency 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, some integrability conditions on the pdf are often assumed (e.g., $f \in L_p(\mathbb{R}^d, \mathbb{B}^d, \mu)$, with $1 < p < \infty$), and consistency is achieved under smoothness conditions on the pdf (e.g., f belongs to a Sobolev space).

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 Devroye and Györfi (1985) and Devroye (1987) who focused on density estimation, Györfi *et al.* (2002) on regression estimation and Devroye *et al.* (1996) on pattern recognition. Universality usually refers to $L_1(\mu)$, but sometimes other L_p spaces are considered. For example, L_2 is the standard space in nonparametric regression, and L_2 is also the natural framework for density estimation with an orthogonal basis. In this context, universality refers to non-smoothness requirements on the pdf. Universal consistency for delta estimators using L_p norms has been studied in Vidal-Sanz (1999) and Vidal-Sanz and Delgado (2004).

The literature on pointwise universal consistency is not so extensive and focuses on the estimation of regression functions; see 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. Let μ be a σ-finite measure in $(\mathbb{R}^d, \mathbb{B}^d)$ and P a probability function satisfying $P \ll \mu$ (i.e., P is absolutely continuous with respect to μ). We say that a delta estimator $\hat{f}_n$ is strongly (or weakly) consistent almost everywhere (a.e.) if $|\hat{f}_n(x) - f(x)| \to 0$ 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 $P \ll \mu$.*

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

$$|\hat{f}_n(x) - f(x)| \leqslant |\mathrm{E}[\hat{f}_n(x)] - f(x)| + |\hat{f}_n(x) - \mathrm{E}[\hat{f}_n(x)]|. \tag{1}$$

The first term on the right-hand side is known as a *bias term*, which is deterministic, and the second term as a *variation term*, which is stochastic. In order to study the pointwise universal convergence to zero of the bias term we will consider 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 bias problem in a pointwise sense. Let $\alpha_n(f)(x) = \int K_{m_n}(x, z)f(z)\mu(\mathrm{d}z)$ be the expected value of $\hat{f}_n(x)$ with respect to the probability distribution P with pdf f. For any smoothing number $\{m_n\}_{n\geqslant 1}$, the estimator $\bar{f}_n$ is universally asymptotically unbiased in the L_1-global sense if and only if the sequence of linear operators $\{\alpha_n\}$ is an approximate identity in $L_1(\mu)$; in other words, for all $f \in L_1(\mu)$ we have that $\lim_{n\to\infty}\|\alpha_n(f) - f\|_{L_1(\mu)} = 0$.

Regarding the pointwise convergence, we say that $\alpha_n(f)$ converges a.e. to f if and only if $|\alpha_n(f)(x) - f(x)| \to 0$ except for sets of μ-null measure; that is, for all $f \in L_1(\mu)$ and all $\delta > 0$, $\lim_{n\to\infty}\mu(\{x \in \mathbb{R}^d : \sup_{n'\geqslant n}|\alpha_{n'}(f)(x) - f(x)| > \delta\}) = 0$. To characterize the pointwise approximation property, we first 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 for all $\varepsilon > 0$, there exists a $\delta > 0$ such that*

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

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

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

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

Next, we present a Banach–Steinhaus type result which plays a crucial role in 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 ($\mathbb{R}^d$, $\mathbb{B}^d$, λ), every dense G_δ set is non-numerable (see Rudin 1974, Theorem 5.3.3).

Theorem 1 (Pointwise uniform boundedness theorem). *Let* $\{\alpha_n\}$ *be a sequence of linear operators in* L_1 ($\mathbb{R}^d$, $\mathbb{B}^d$, μ)*, all of them bounded in measure. Then only one of the following statements holds:*

(i) $\{\alpha_n\}_{n\in\mathbb{N}}$ *is uniformly bounded in measure.*
(ii) *For all* $\varepsilon > 0$*, there exists a* $\mathcal{C}_\varepsilon \subset L_1(\mu)$*, where* $\mathcal{C}_\varepsilon$ *is a dense* G_δ *set, such that for all* $f \in \mathcal{C}_\varepsilon$*,*

$$\mu(\{x \in \mathbb{R}^d : \alpha^M(f)(x) = \infty\}) > \varepsilon. \tag{3}$$

Proof. Define the set $V_{\bar\varepsilon}^{\bar\delta} = \{f \in L_1(\mu) : \mu(\{x \in \mathbb{R}^d : \alpha^M(f)(x) > \bar\delta\}) > \bar\varepsilon\}$, for all $\bar\varepsilon > 0$ and all $\bar\delta > 0$. 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 $\lim_{k\to\infty}\|g_k - g\|_{L_1(\mu)} = 0$,

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

Since α_n is bounded in measure, it is continuous in measure. Thus, for each $n \in \mathbb{N}$, the sublinear operator $\alpha_n^M(f)(x) = \sup_{n'\leqslant n}|\alpha_{n'}(f)(x)|$ is also continuous in measure. Then, for all $n \in \mathbb{N}$, the sets

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

are open, implying that $V_{\bar\varepsilon}^{\bar\delta}$ is open.

Now consider a sequence $\{\delta_k\}_{k\in\mathbb{N}}$ dense in $\mathbb{R}^+$. Thus, for all $\bar\varepsilon > 0$, we have a sequence $\{V_{\bar\varepsilon}^{\delta_k}\}_{k\in\mathbb{N}}$ of open sets. Assume that there exists a $k \in \mathbb{N}$ such that $V_{\bar\varepsilon}^{\delta_k}$ is not dense in $L_1(\mu)$. Then there exists an $f_0 \in L_1(\mu)$ and $r > 0$ such that $\|f\|_{L_1(\mu)} \leqslant r$ implies $(f_0 + f) \notin V_{\bar\varepsilon}^{\delta_k}$. Thus, $\mu(\{x \in \mathbb{R}^d : \alpha^M(f_0 + f)(x) > \delta_k\}) \leqslant \bar\varepsilon$ for all $f \in L_1(\mu)$ such that $\|f\|_{L_1(\mu)} \leqslant r$. Note that $f = (f_0 + f) - f_0$, so then

$$\mu(\{x \in \mathbb{R}^d : \alpha^M(f)(x) > 2\delta_k\}) \leqslant \mu(\{x \in \mathbb{R}^d : \alpha^M(f_0 + f)(x) > \delta_k\}) + \mu(\{x \in \mathbb{R}^d : \alpha^M(f_0)(x) > \delta_k\}) \leqslant 2\bar{\varepsilon}.$$

Therefore,

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

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

On the other hand, if every $V_{\bar{\varepsilon}}^{\delta_k}$ is dense in $L_1(\mu)$ then $\mathcal{C}_{\bar{\varepsilon}} = \cap_{k \in \mathbb{N}} V_{\bar{\varepsilon}}^{\delta_k}$ is a dense G_δ set in $L_1(\mu)$, applying Baire's theorem (see Rudin 1974). Obviously, for all $f \in \mathcal{C}_{\bar{\varepsilon}}$ we have, for all δ_k, $\mu(\{x \in \mathbb{R}^d : \alpha^M(f)(x) > \delta_k\}) > \varepsilon$, and $\{\delta_k\}_{k \in \mathbb{N}}$ is dense in $\mathbb{R}^+$, so that condition (3) follows. □

A result analogous to the previous theorem can be established on $L_p(\mathbb{R}^d, \mathbb{B}^d, \mu)$, 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; 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 to guarantee that the sequence $\{\alpha_n\}$ satisfies a.e. convergence and, therefore, that 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 $L_1(\mathbb{R}^d, \mathbb{B}^d, \mu)$. Assume that:*

(i) *the sequence $\{\alpha_n\}$ is uniformly bounded in measure;*
(ii) *there exists a set $\mathcal{G} \subset L_1(\mu)$ dense such that for all $\tilde{f} \in \mathcal{G}$, $\alpha_n(\tilde{f}) \to \tilde{f}$ a.e.*

Then $\{\alpha_n\}$ is an approximate identity in the a.e. sense, that is, $\alpha_n(f) \to f$ a.e. for all $f \in L_1$ (μ). If the operators $\{\alpha_n\}$ are all bounded in measure on $L_1(\mu)$, then assumptions (i) and (ii) are also necessary.

Proof. We divide the proof into two parts. *Part I. Sufficient conditions.* Assume that there exists a dense set $\mathcal{G} \subset L_1(\mu)$, such that for all $\tilde{f} \in \mathcal{G}$ and all $\delta > 0$,

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

As $\mathcal{G}$ is a dense set, for all $f \in L_1(\mu)$ and for all $\varepsilon > 0$, there exists an $\tilde{f} \in \mathcal{G}$ such that $\|f - \tilde{f}\|_{L_1(\mu)} \leqslant \varepsilon$ By the triangular inequality, for each n and each $x \in \mathbb{R}^d$,

$$\sup_{n' \geqslant n} |\alpha_{n'}(f)(x) - f(x)| \leqslant \sup_{n' \geqslant n} |\alpha_{n'}(f)(x) - \alpha_{n'}(\tilde{f})(x)| + \sup_{n' \geqslant n} |\alpha_{n'}(\tilde{f})(x) - \tilde{f}(x)| + |\tilde{f}(x) - f(x)|,$$

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

$$\mu\left(\left\{x \in \mathbb{R}^d : \sup_{n' \geqslant n} |\alpha_{n'}(f)(x) - f(x)| > \delta\right\}\right) \leqslant \mu\left(\left\{x \in \mathbb{R}^d : \sup_{n' \geqslant n} |\alpha_{n'}(f - \tilde{f})(x)| > \frac{\delta}{3}\right\}\right)$$
$$+ \mu\left(\left\{x \in \mathbb{R}^d : \sup_{n' \geqslant n} |\alpha_{n'}(\tilde{f})(x) - \tilde{f}(x)| > \frac{\delta}{3}\right\}\right)$$
$$+ \mu\left(\left\{x \in \mathbb{R}^d : |\tilde{f}(x) - f(x)| > \frac{\delta}{3}\right\}\right).$$

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

$$\mu\left(\left\{x \in \mathbb{R}^d : \sup_{n' \geqslant n} |\alpha_{n'}(f - \tilde{f})(x)| > \frac{\delta}{3}\right\}\right) \leqslant \mu\left(\left\{x \in \mathbb{R}^d : \alpha^M(f - \tilde{f})(x) > \frac{\delta}{3}\right\}\right)$$
$$\leqslant \mu\left(\left\{x \in \mathbb{R}^d : \alpha^M\left(\frac{f - \tilde{f}}{\|f - \tilde{f}\|_{L_1(\mu)}}\right)(x) \cdot \|f - \hat{f}\|_{L_1(\mu)} > \frac{\delta}{3}\right\}\right)$$
$$\leqslant \sup_{\|f\|_{L_1(\mu)} \leqslant 1} \mu\left(\left\{x \in \mathbb{R}^d : \alpha^M(f)(x) > \frac{\delta}{3\varepsilon}\right\}\right) \leqslant \varepsilon_1.$$

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

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

$$\mu\left(\left\{x \in \mathbb{R}^d : \sup_{n' \geqslant n} |\alpha_{n'}(f)(x) - f(x)| > \delta\right\}\right)$$
$$\leqslant \varepsilon_1 + \mu\left(\left\{x \in \mathbb{R}^d : \sup_{n' \geqslant n} |\alpha_{n'}(\tilde{f})(x) - \tilde{f}(x)| > \frac{\delta}{3}\right\}\right) + \frac{\|\tilde{f} - f\|_{L_1(\mu)}}{\delta/3}$$
$$\leqslant \varepsilon_1 + \mu\left(\left\{x \in \mathbb{R}^d : \sup_{n' \geqslant n} |\alpha_{n'}(\tilde{f})(x) - \tilde{f}(x)| > \frac{\delta}{3}\right\}\right) + \frac{3\varepsilon}{\delta}.$$

Since $\varepsilon, \varepsilon_1 > 0$ are arbitrarily small, and the second term on the right-hand side of last inequality tends to zero for all $\delta > 0$, the a.e. approximation follows.

Part II. Necessary condition. Assume that $\alpha_n(f) \to f$ a.e. for all $f \in L_1(\mu)$. 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, for all $\varepsilon > 0$, there exists a $\mathcal{C}_\varepsilon \subset L_1(\mu)$, which is a dense G_δ set, such that for all $f \in \mathcal{C}_\varepsilon$, $\mu(\{x \in \mathbb{R}^d : \sup_{n \in \mathbb{N}} |\alpha_n(f)(x)| = \infty\}) > 2\varepsilon$. In other words, there exists a $B \subset \mathbb{R}^d$, with $\mu(B) > 2\varepsilon$, such that for all $x \in B$, $\sup_{n \in \mathbb{N}} |\alpha_n(f)(x)| = \infty$. On the other hand, $|f(x)| < \infty$ holds a.e. for all $f \in L_1(\mu)$ (in particular, for all $f \in \mathcal{C}_\varepsilon$), because there exists a $\delta_\varepsilon > 0$, such that $\mu(\{x \in \mathbb{R}^d : |f(x)| > \delta_\varepsilon\}) \leqslant \delta_\varepsilon^{\ 1} \|f\|_{L_1(\mu)} < \varepsilon$. In other words, for all $\varepsilon > 0$, there exists an $A \subset \mathbb{R}^d$ with $\mu(A^c) < \varepsilon$, such that $\sup_{x \in A} |f(x)| < \infty$.

Applying the triangular inequality, $|\alpha_n(f)(x) - f(x)| \geqslant | \, |\alpha_n(f)(x)| - |f(x)| \, |$. Define $C = A \cap B$. Obviously, for all $x \in C$ and all $f \in \mathcal{C}_\varepsilon$,

$$|\alpha_n(f)(x) - f(x)| \geqslant \big|\,|\alpha_n(f)(x)| - |f(x)|\,\big| = \infty.$$

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

$$\mu(B) = \mu(A \cap B) + \mu(A^c \cap B) \leqslant \mu(A \cap B) + \mu(A^c) = \mu(C) + \mu(A^c),$$

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

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

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

Since all elements of the sequence $\{\alpha_n\}$ are bounded in measure, the triangular inequality implies that for all $n \in \mathbb{N}$, $|\alpha_n(f)(x) - f(x)| \leqslant |\alpha_n(f)(x)| + |f(x)| < \infty$ a.e. Thus,

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

Therefore, (4) implies that for all $f \in \mathcal{C}_\varepsilon$,

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

which, contradicts the a.e. approximation property. □

Assume that $\{\alpha_n\}$ satisfies the 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 $\lim_{r\to\infty} \|f_r - f\|_{L_1(\mu)} = 0$, we have that

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

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 $\alpha_n(f)(x) = \int K_{m_n}(x, z) f(z)\, \mu(\mathrm{d}z)$ as the operator

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

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:*

(i) *The sequence $\{|\alpha|_n\}$ is uniform bounded in measure.*
(ii) $\int K_{m_n}(x, z)\mu(\mathrm{d}z) \to 1$, *a.e.*
(iii) *For all $\delta > 0$, there is some $M_\delta > 0$, such that* $\sup_{n\in\mathbb{N}} \int_{\|x\ \ z\|<\delta} |K_{m_n}(x, z)|\mu(\mathrm{d}z) < M_\delta$ *a.e.*
(iv) $\int_{\|x\ \ z\|>\delta} |K_{m_n}(x, z)|\mu(\mathrm{d}z) \to 0$ *a.e., for all $\delta > 0$.*

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

Proof. First, it will be proved that if $\{|\alpha|_n\}$ is uniformly bounded in measure, then $\{\alpha_n\}$ is also uniformly bounded in measure. Since the maximal operators satisfy

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

with $|\alpha|^M = \sup_{n\in\mathbb{I}}|\alpha|_n$, then, for all $\delta > 0$,

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

Taking the supremum in the unit ball $\|f\|_{L_1(\mu)} \leqslant 1$, the aforementioned result is proved.

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 an a.e. identical element in $C_c(\mathbb{R}^d)$. As $C_c(\mathbb{R}^d)$ is a dense set in $L_p(\mu)$, $1 \leqslant p < \infty$, the result follows from Theorem 2. We proceed in two steps.

Step 1. For all $\delta > 0$, and all $h(x, z) \in C_c(\mathbb{R}^d \times \mathbb{R}^d)$,

$$\left|\int_{\{z\,:\|x\quad z\|>\delta\}} h(x, z)K_{m_n}(x, z)\,\mu(\mathrm{d}z)\right| \leqslant \|h\|_\infty \cdot \int_{\{z:\|x\quad z\|>\delta\}} |K_{m_n}(x, z)|\mu(\mathrm{d}z) \to 0 \text{ a.e.},$$

using assumption (iv), and $\|h\|_\infty < \infty$.

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

$$\sup_{n'\geqslant n}|\alpha_{n'}(f)(x) - f(x)| \leqslant \sup_{n'\geqslant n}\left|\int(f(z) - f(x))\,K_{m_{n'}}(x, z)\,\mu(\mathrm{d}z)\right| + \sup_{n'\geqslant n}\left|\int K_{m_{n'}}(x, z)\mu(\mathrm{d}z)f(x) - f(x)\right|.$$

By assumption (ii),

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

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

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

As f is uniformly continuous, for all $\varepsilon > 0$, there exists a $\delta > 0$ such that $\|x - z\| \leqslant \delta$ implies that $|f(x) - f(z)| \leqslant \varepsilon$. Applying assumption (iii) we obtain

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

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 is proved. □

A sufficient condition for assumption (iv) in Theorem 3 is that, for some $s \geq 1$,

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

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 for all $\delta > 0$,

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

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 $L_1(\mathbb{R}^d, \mathbb{B}^d, \lambda)$, defined as

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

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

Example 1. Consider the kernel estimator in $L_1(\mathbb{R}^d, \mathbb{B}^d, \lambda)$. If there exists a closed interval $C\subset\mathbb{R}^d$ such that $c_1I_C(u) \leq |\mathbf{K}(u)| \leq c_2I_C(u)$, for some $c_1, c_2 > 0$, then

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

applying the Hardy–Littlewood argument, so that kernel operators are uniformly bounded in measure. If $\|m_n\| \to 0$, the pointwise universal unbiasedness readily follows from Theorem 3, as

$$\int_{\|x-z\|>\delta}|K_{m_n}(x, z)|\mathrm{d}z \leq \delta^{-1}\frac{1}{\det(m_n)}\int\|x-z\|\mathbf{K}(m^{-1}(z-x))\mathrm{d}z$$

$$= \delta^{-1}\int\|m_nu\|\mathbf{K}(u)\mathrm{d}u \leq \|m_n\|\delta^{-1}\int_C\|u\|\mathbf{K}(u)\mathrm{d}u \to 0,$$

for any $\delta > 0$ and any matrix norm such that $\|AB\| \leq \|A\|\,\|B\|$.

Example 2. Define the set $\mathbb{I}_0$ of regular partitions of $\mathbb{R}^d$ as the set of Borel measurable countable 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 $\mathbb{I}_0$ covers $\mathbb{R}^d$ in the sense of Vitali (i.e., for each $x \in \mathbb{R}^d$ and every sequence $\{m_n\} \subset \mathbb{I}_0$ ordered with respect to the thinner relation, there exist $A_n \in m_n$ such that $x \in \cap_n A_n$ and the diameter of A_n tends to zero). Consider the histogram in $L_1(\mathbb{R}^d, \mathbb{B}^d, \lambda)$, defined for $\{m_n\} \subset \mathbb{I}_0$. Using the fact that $\beta^*(f)(x) = \sup_{\varepsilon>0} P_f(B(x, \varepsilon))/\lambda(B(x, \varepsilon))$ satisfies $\|\beta(f)\|_{L_1(\lambda)} \leqslant c_d \|f\|_{L_1(\lambda)}$, then

$$\int\left(\sup_{n\in\mathbb{N}}\int\left(\sum_{A\in m_n}\frac{I_A(x)I_A(z)}{\lambda(A)}\right)f(z)\mathrm{d}z\right)\mathrm{d}x = \int\sup_{n\in\mathbb{N}}\sum_{A\in m_n}\frac{I_A(x)P_f(A)}{\lambda(A)}\,\mathrm{d}x \leqslant c\|f\|_{L_1(\lambda)},$$

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

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

$$\alpha_n(g)(x) = \sum_{A\in m_n}\left(\frac{1}{\lambda(A)}\int_A g(z)\lambda(\mathrm{d}z)\right)I_A(x) = \sum_{A\in m_n}\left(\sum_{r=1}^{\mathcal{S}}\beta_r\frac{1}{\lambda(A)}\int_A I_{B_r}(z)\lambda(\mathrm{d}z)\right)I_A(x)$$

$$= \sum_{A\in m_n}\left(\sum_{r=1}^{\mathcal{S}}\beta_r\frac{\lambda(A\cap B_r)}{\lambda(A)}\right)I_A(x).$$

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

$$\lambda(\{|\alpha_n(g)(x) - g(x)| > \delta\}) = \lambda\left(\sup_{n'\geqslant n}\left|\sum_{A\in m_{n'}}\sum_{r=1}^{\mathcal{S}}\beta_r\frac{\lambda(A\cap B_r)}{\lambda(A)}I_A(x) - \sum_{r=1}^{\mathcal{S}}\beta_r I_{B_r}(x)\right| > \delta\right)$$

$$\leqslant \lambda\left(\sup_{n'\geqslant n}\sum_{A\in m_{n'}}\frac{1}{\lambda(A)}\left|\sum_{r=1}^{\mathcal{S}}\beta_r(\lambda(A\cap B_r) - \lambda(A)I_{B_r}(x))I_A(x)\right| > \delta\right).$$

Next, we prove that this measure tends to zero. If $m_n \geqslant \overline{m}$, that is, m_n is thinner than $\overline{m}$, then for all $B_r \in \overline{m}$ and for all $A \in m_n$ we have one of the following cases: either (i) $A \cap B_r = \varnothing$ 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, for all $g \in \mathcal{S}$, there exists an $\overline{m}$ such that $\sup_{m\geqslant\overline{m}}|\alpha_m(g)(x) - g(x)| = 0$, except for sets of null measure, and the aforementioned result is proved.

Example 3. We also consider the a.e. convergence of Dirichlet's approximate identity $\{\alpha_n\}$, where $\alpha_n(f)(x) = \int_{\pi}^{\pi} K_{m_n}(z - x)f(z)\mathrm{d}z$,

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

is the Dirichlet kernel, and $\{m_n\} \subset \mathbb{N}$, which is related to the Fourier sums in $L_p([-\pi, \pi])$, with $1 \leqslant p < \infty$; see 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. First, trigonometric polynomials are a dense subspace in $L_p([-\pi, \pi])$ with $1 \leqslant p < \infty$, and the Fourier sums of trigonometric polynomials converge a.e. to the respective polynomials; see Mozzochi (1970, p. 9), Jørsboe and Mejlbro (1982, pp. 17–20), and Arias de Reyna (2002, Part II). Secondly, 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 $\sup_{\|f\|_{L_p([-\pi,\pi])} \leqslant 1} \|\alpha^M(f)\|_{L_p([-\pi,\pi])} < \infty$, which implies the result using Markov's inequality.

Thus, Theorem 2 implies that the Fourier sums satisfy the a.e. approximation property for every curve in $L_p([-\pi, \pi])$ with $1 < p < \infty$. The proof of the Carleson–Hunt theorem presents great technical difficulties. The 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 a.e. approximation fails. This is a well-known problem. A very famous counterexample, due to Kolmogorov (1926), shows that for some function in $L_1([-\pi, \pi])$ the Fourier sum diverges a.e. Some additional results on pointwise divergence can be seen in Körner (1981), Edwards (1979, p. 80) and Zygmund (1959, Section 8.4). As we can see in the proof of Theorem 2, there is a dense G_δ set of functions in $L_1([-\pi, \pi])$ on which $\alpha_n(f)(x)$ diverges a.e. 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

The aim of this section is to prove that

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

almost surely (in probability) for almost every $x \in \mathbb{R}^d$ with respect to μ, which is immediate by using a simple law of large numbers 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 $f = \mathrm{d}P/\mathrm{d}\mu \in L_1(\mu)$, the triangular array $\{K_{m_n}(x, X_i) : 1 \leqslant i \leqslant n\}_{n \in \mathbb{N}}$ is such that for some $r > 1$, $\mathrm{E}[|K_{m_n}(x, X)|^r] = o(n^{(r-1)})$, a.e. $[\mu]$. Then*

$$\mathrm{E}[|\hat{f}_n(x) - \mathrm{E}[\hat{f}_n(x)]|^r] \to 0, \qquad |\hat{f}_n(x) - \mathrm{E}[\hat{f}_n(x)]| \to_p 0,$$

a.e. $[\mu]$*, with* $f = \mathrm{d}P/\mathrm{d}\mu$*, and the result holds universally in* P*.*

Proof. We define $Z_{n,i} = K_{m_n}(x; X_i)$. Then

$$\mathrm{E}[|\hat{f}_n(x) - \mathrm{E}[\hat{f}_n(x)]|^r] \leqslant 2^{r-1} \frac{\sum_{i=1}^n \mathrm{E}[|Z_{n,i} - \mathrm{E}[Z_{n,i}]|^r]}{n^r} \leqslant \frac{2^r \sum_{i=1}^n \mathrm{E}[|Z_{n,i}|^r]}{n^r} \to 0,$$

where we have applied the c_r and Jensen inequalities. The consistency follows applying Markov's inequality. □

The following examples illustrate the application of the previous result.

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

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

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

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 $L_1(\mathbb{R}^d, \mathbb{B}^d, \lambda)$, for regular partitions. Notice that for any partition $m \in \mathbb{I}_0$, $|K_m(x, z)|^2 = \sum_{A \in m} |I_A(x)I_A(z)/\lambda(A)|^2$ 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

$$n^{-1}\mathrm{E}[|K_{m_n}(x, X)|^2] = \frac{1}{n}\mathrm{E}\left[\sum_{A \in m_n} \left|\frac{I_A(x)I_A(X)}{\lambda(A)}\right|^2\right] = \frac{1}{n} \sum_{A \in m_n} \frac{P(A)}{\lambda(A)^2} I_A(x)$$

$$\leqslant \frac{1}{n \cdot \gamma(m_n)} \sum_{A \in m_n} \frac{P(A)}{\lambda(A)} I_A(x) = \frac{1}{n \cdot \gamma(m_n)} \mathrm{E}[\hat{f}_n(x)] \to 0,$$

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

Example 6. Consider the Dirichlet kernel in $L_p([-\pi, \pi])$, with $p > 1$. Using the fact that

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

and $\cot(t) - t^{-1}$ is bounded on $(-\pi/2, \pi/2)$, then $K_m(u) = \pi^{-1}\sin(mu)/u + O(1)$. Thus,

$$n^{-1}\mathrm{E}[|K_{m_n}(x, X)|^2] = \frac{1}{n}\int_{-\pi}^{\pi} |K_{m_n}(u)|^2 f(u-x)\mathrm{d}u$$

$$\leqslant \frac{1}{n}\int_{-\pi}^{\pi} \left|\frac{m_n|u|}{\pi u}\right|^2 f(u-x)\mathrm{d}u + O\left(\frac{1}{n}\right) = \frac{m_n^2}{\pi^2 n} + O\left(\frac{1}{n}\right), \text{ a.e.,}$$

as $|\sin(mu)| \leqslant |mu|$. Weak universal consistency follows from the 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. Its application is illustrated with some examples.

Theorem 4 (Universal pointwise strong consistency of variation term). *Assume that for any probability function P with $f = \mathrm{d}P/\mathrm{d}\mu \in L_1(\mu)$,*

$$\sum_{n=1}^{\infty} \exp\left\{\frac{-n}{M_n(x)^2}\right\} < \infty \text{ a.e. } [\mu], \tag{5}$$

where $M_n(x) = \operatorname{ess\,sup}_z |K_{m_n}(x, z)|$. Then universal pointwise convergence is satisfied a.e.$[\mu]$, universally in P.

Proof. The result is a consequence of Hoeffding's inequality (see Györfi *et al.* 2002). Let us consider $Z_{n,i} = K_{m_n}(x, z)$. By assumption, $Z_{n,i} \in [-M_n(x), M_n(x)]$ for $i = 1, \ldots, n$ with probability one. Therefore, for all $\varepsilon > 0$,

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

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 (5) is satisfied if $\sum_{n=1}^{\infty} \exp\{-n\det(m_n)^2\} < \infty$, for which it suffices that $n\det(m_n)^2/\log n \to \infty$.

Example 8. The histogram satisfies

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

and the condition in (5) is satisfied if $\sum_{n=1}^{\infty} \exp\{-n\gamma(m_n)^2\} < \infty$, 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

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

The condition in (5) is satisfied if $\sum_{n=1}^{\infty}\exp\{-n/m_n^2\} < \infty$, for which it suffices that $m_n^2(\log n)/n \to 0$.

Note that in Theorem 4, if Bernstein's inequality is used instead of Hoeffding's,

$$Pr\left[\left|\frac{1}{n}\sum_{i=1}^{n}(Z_{n,i} - \mathrm{E}[Z_{n,i}])\right| > \varepsilon\right] \leqslant \exp\left\{\frac{-n\varepsilon^2}{2\,\mathrm{var}(Z_{n,i}) + 2M_n(x)\varepsilon/3}\right\},$$

then condition (5) could be replaced by

$$\sum_{n=1}^{\infty}\exp\left\{\frac{-n}{max\{\mathrm{E}[|K_{m_n}(x, X)|^2],\, M_n(x)\}}\right\} < \infty \text{ a.e. } [\mu],$$

so that the required rates in the kernel and histogram examples can be reduced to $n\det(m_n)/\log n \to \infty$ and $n\gamma(m_n)/\log n \to \infty$, respectively.

Acknowledgements

I wish to thank Miguel A. Delgado, Juan A. Cuesta, the editor 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

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

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

Bary, N.K. (1964) *A Treatise on Trigonometric Series*, 2 vols. Oxford: Pergamon Press.

Bergh, J. and Löfström, J. (1976) *Interpolation Spaces. An Introduction*, Grundlehren Math. Wiss. 223. Berlin: Springer Verlag.

Carleson, L. (1966) On convergence and growth of partial sums of Fourier series. *Acta Math.*, **116**, 135–157.

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

Devroye, L. (1981) On the almost everywhere convergence of nonparametric regression function estimates. *Ann. Statist.*, **9**, 1310–1319.

Devroye, L. (1987) *A Course in Density Estimation*, Progr. Probab Statist. 14. Boston: Birkhäuser.

Devroye, L. and Györfi, L. (1985) *Nonparametric Density Estimation. The* L*1* *View*. New York: Wiley.

Devroye, L., Györfi, L. and Lugosi, G. (1996) *A Probabilistic Theory of Pattern Recognition*. New York: Springer Verlag.

Edwards, R.E. (1979) *Fourier Series. A Modern Introduction. Vol. 1*. Berlin: Springer Verlag.

Fefferman, C. (1971) On the convergence of multiple Fourier series. *Bull. Amer. Math. Soc.*, **77**, 744 755.

Garsia, A.M. (1970) *Topics in Almost Everywhere Convergence*. Chicago: Markham.

Greblicki, W., Krzyżak, A. and Pawlak, M. (1984) Distribution free pointwise consistency of kernel regression estimate. *Ann. Statist.*, **12**, 1570 1575.

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

Hunt, R.A. (1968) On the convergence of Fourier series. In D.T. Haimo (ed.), *Orthogonal Expansions and Their Continuous Analogues*, pp. 235 255. Carbondale: Southern Illinois University Press.

Jørsboe, O.G. and Mejlbro, L. (1982) *The Carleson Hunt Theorem on Fourier Series*, Lecture Notes in Math. 911. Berlin: Springer Verlag.

Kolmogorov, A.N. (1926) Une série de Fourier Lebesgue divergente partout. *C. R. Acad. Sci. Paris*, **183**, 1327 1329.

Körner, T.W. (1981) Everywhere divergent Fourier series. *Colloq. Math.*, **45**, 103 118.

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

Prakasa Rao, B.L.S. (1983) *Nonparametric Functional Estimation*. Orlando, FL: Academic Press.

Rudin, W. (1974) *Real and Complex Analysis*. (2nd edn). New York: McGraw Hill.

Sjölin, P. (1971) Convergence almost everywhere of certain singular integrals and multiple Fourier series. *Ark. Math.*, **9**, 65 90.

Stein, E.M. (1970) *Singular Integrals and Differentiability Properties of Functions*. Princeton, NJ: Princeton University Press.

Stone, C.J. (1977) Consistent nonparametric regression. *Ann. Statist.*, **5**, 595 645.

Terrell, G.R. (1984) Efficiency of nonparametric density estimators. Technical report, Department of Mathematical Sciences, Rice University.

Terrell, G.R. and Scott, D.W. (1992) Variable kernel density estimation. *Ann. Statist.*, **20**, 1236 1265.

Vidal Sanz, J.M. (1999) Universal consistency of delta estimators. PhD thesis (in Spanish), Universidad Carlos III de Madrid.

Vidal Sanz, J.M. and Delgado, M.A. (2004) Universal consistency of delta estimators. *Ann. Inst. Statist. Math.,* **56**, 791 818.

Walk, H. (2001) Strong universal pointwise consistency of recursive regression estimates. *Ann. Inst. Statist. Math.,* **53**, 691 707.

Walter, G. and Blum, J.R. (1979) Probability density estimation using delta sequences. *Ann. Statist.*, **7**, 328 340.

Watson, G.S. and Leadbetter, M.R. (1963) On the estimation of the probability density I. *Ann. Math. Statist.*, **34**, 480 491.

Watson, G.S. and Leadbetter, M.R. (1964) Hazard analysis II. *Sankhyā Ser. A*, **26**, 101 116.

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

Whittle, P. (1958) On smoothing a probability density function. *J. Roy. Statist. Soc. Ser. B*, **20**, 334 343.

Winter, B.B. (1973) Strong uniform consistency of integrals of density estimation. *Canad. J. Statist.*, **1**, 247 253.

Winter, B.B. (1975) Rate of strong consistency of two nonparametric density estimators. *Ann. Statist.*, **3**, 759 766.

Zygmund, A. (1959) *Trigonometric Series*, 2 vols. Cambridge: Cambridge University Press.

Received October 2003 and revised February 2005