Ann. Inst. Statist. Math.
Vol. 56, No. 4, 791–818 (2004)

UNIVERSAL CONSISTENCY OF DELTA ESTIMATORS

JOSE M. VIDAL-SANZ[1] AND MIGUEL A. DELGADO[2]

[1] *Department of Business Economics, Universidad Carlos III de Madrid, C/ Madrid, 126, 28903 Getafe, Madrid, Spain*
[2] *Department of Economics, Universidad Carlos III de Madrid, C/ Madrid, 126, 28903 Getafe, Madrid, Spain*

(Received October 22, 2002; revised November 6, 2003)

Abstract. This paper considers delta estimators of the Radon-Nikodym derivative of a probability function with respect to a σ-finite measure. We provide sufficient conditions for universal consistency, which are checked for some wide classes of nonparametric estimators.

Key words and phrases: Nonparametric density estimation, delta estimators, universal consistency.

1. Introduction

Let P be a probability measure in the Borel space $(\mathbb{R}^d, \mathbb{B}^d)$, absolutely continuous with respect to the σ-finite measure μ and $f = dP/d\mu$ be the corresponding Radon-Nikodym derivative, which is assumed to belong to the space $L_p(\mathbb{R}^d, \mathbb{B}^d, \mu)$, with $1 \leq p < \infty$. Usually, the Lebesgue measure λ is considered, and $f = dP/d\lambda$ is the associated probability density function (pdf). Given a random sample $\{X_i\}_{i=1}^n$ 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 called a smoothing sequence, and $\{K_{m_n}\}_{n\in\mathbb{N}}$ a 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 norm, in the usual kernel estimator of a multivariate density; or the order of a polynomial, in the Fourier series estimator. We consider that the smoothing sequence $\{m_n\}_{n\in\mathbb{N}}$ belongs to some *directed set* $\mathbb{I}$. We say that the set $\mathbb{I}$ is directed if it is a non empty set endowed with a partial preorder $\leq$, such that if $\forall m_1, m_2 \in \mathbb{I}$, $\exists m_3 \in \mathbb{I}$ such that $m_1 \leq m_3$ and $m_2 \leq m_3$. We also assume that $\{m_n\}_{n\in\mathbb{N}}$ diverges in $\mathbb{I}$ as $n \to \infty$, i.e., $\forall M \in \mathbb{I}$, $\exists n_M \in \mathbb{N}$ such that $m_n \geq M$, $\forall n \geq n_M$.

The class of delta estimators was introduced by Whittle (1958), encompassing most of the existing nonparametric estimators. 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 interpreted as delta estimators, at least asymptotically.

In case of pdf estimation, $\widehat{f}_n(x)$ is pointwise asymptotically unbiased if,

$$\lim_{n\to\infty} E[\widehat{f}_n(x)] = \int \delta(z-x)f(x)\lambda(dx) = f(x),$$

where δ is the Dirac delta generalized function with a jump at zero. This is why these estimators are known as delta estimators. 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, assuming smoothness conditions on f. Winter (1973, 1975) has studied uniform consistency and consistency of the corresponding smooth integrated distribution function estimator. Watson and Leadbetter (1964) have established asymptotic normality. Basawa and Prakasa Rao ((1980), Chap. 11) have provided results for dependent observations. In this literature, consistency is achieved under restrictive smoothness conditions on the pdf f. The universal consistency of general delta estimators has not been obtained yet.

DEFINITION 1.1. *Universal Consistency.* Let μ be a σ-finite measure in $(\mathbb{R}^d, \mathbb{B}^d)$, and P a probability distribution $P \ll \mu$, with $f = dP/d\mu \in L_p(\mathbb{R}^d, \mathbb{B}^d, \mu)$. Henceforth, $L_p(\mu) := L_p(\mathbb{R}^d, \mathbb{B}^d, \mu)$. We say that a delta estimator $\widehat{f}_n$ is strongly consistent (in $L_p(\mu)$-global sense) when

$$\|\widehat{f}_n - f\|_{L_p(\mu)} \overset{\text{a.e.}}{\to} 0, \tag{1.1}$$

and weakly consistent, when the convergence is in probability. We say that the convergence is universal when (1.1) holds for every probability function $P \ll \mu$ with $f \in L_p(\mu)$. Note that the *degree of universality* depends on p.

Usually, weak universal consistency is defined as,

$$E[\|\widehat{f}_n - f\|^p_{L_p(\mu)}] = \int E[|\widehat{f}_n(x) - f(x)|^p]\mu(dx) \to 0,$$

see Stone (1977). The equivalence with the above definition is a consequence of Markov's inequality and Lebesgue's dominated convergence theorem.

The literature on universal consistency of smooth nonparametric estimators is enormous, and is mainly based on Stone's (1977) seminal paper. Universal consistency of histograms and regressograms has been proved by Abou-Jaoude (1976*a*, *b*, *c*), Devroye and Györfi (1985*a*, 1985*b*), Devroye and Györfi (1983), Györfi *et al.* (2002). Universal consistency of discriminant analysis rules based on partitions have been studied by Devroye *et al.* (1996*b*) and Lugosi and Nobel (1996). Universal consistency of density and regression kernel estimators has been showed by Devroye and Wagner (1979, 1980*a*, 1980*b*), Devroye (1983, 1987), Devroye and Györfi (1985*a*), Devroye and Krzyżak (1989) and Györfi *et al.* (2002). Also discriminant analysis rules have been studied by Devroye and Krzyżak (1989) and Devroye *et al.* (1996*b*). Universal consistency of estimators based on $k-nn$ in density and regression estimation has been considered by Stone (1977), Devroye and Györfi (1985*a*), Györfi (1981), Devroye *et al.* (1996*a*) and Györfi *et al.* (2002). Discriminant analysis rules has been studied by Stone (1977), Devroye and Wagner (1982), Devroye *et al.* (1996*b*). Orthonormal series estimators of density and regression functions, based on sieve estimators theory, have been studied by Devroye and Györfi (1985*a*), Lugosi and Zeger (1995), Györfi *et al.* (2002) and, in the context of pattern recognition by Vapnik (1982) and Devroye *et al.* (1996*b*).

There are also some monographs on universal consistency. Devroye and Györfi (1985*a*) and Devroye (1987) have focused on density estimation, Györfi *et al.* (2002) on regression estimation and Devroye *et al.* (1996*b*) on pattern recognition.

The aim of this paper is to provide fairly primitive conditions which are sufficient for universal consistency, as defined in (1.1). To this end, we use the triangular inequality,

$$\|\widehat{f}_n - f\|_{L_p(\mu)} \leq \|E(\widehat{f}_n) - f\|_{L_p(\mu)} + \|\widehat{f}_n - E(\widehat{f}_n)\|_{L_p(\mu)}. \tag{1.2}$$

The first term on the right-hand side is known as the *bias term*, which is deterministic, and the second term is known as the *variation term*, which is stochastic.

In next section, we provide sufficient conditions for the universal convergence of the bias term to zero, using approximation theory. We also illustrate the results checking the conditions for some wide classes of estimators. In Section 3 we provide sufficient conditions for the universal convergence of the variation term, applying Law of Large Numbers (LLN) for triangular arrays on Banach spaces. Proofs are confined to the last section.

2. Universal asymptotic unbiasedness

First, we introduce the concept of *net*. Let $(B; \|\cdot\|_B)$ be a Banach space. Given a directed set $\mathbb{I}$, a net $\{g_m\}_{m\in\mathbb{I}}$ in $(B; \|\cdot\|_B)$, is an application $g_m = g(m)$ with $g : \mathbb{I} \to B$. We say that the net $\{g_m\}_{m\in\mathbb{I}}$ converges to $f \in B$, denoted by $\lim_{m\in\mathbb{I}} \|g_m - f\|_B = 0$, if $\forall\varepsilon > 0$, $\exists m_\varepsilon \in \mathbb{I}$ such that $\|g_m - f\|_B < \varepsilon$ for all $m \geq m_\varepsilon$. For an introduction to convergence analysis of nets, see Edgar and Sucheston (1992).

Let $\alpha_{m_n}(f)(\cdot) = \int K_{m_n}(\cdot, z)f(z)\mu(dz)$ be the expected value of $\widehat{f}_n(\cdot)$ with respect to the probability distribution P. Notice that α_m is a linear operator in the Banach space $(L_p(\mu), \|\cdot\|_{L_p(\mu)})$,

$$\begin{aligned} \alpha_m : L_p(\mathbb{R}^d, \mathbb{B}^d, \mu) &\to L_p(\mathbb{R}^d, \mathbb{B}^d, \mu) \\ f \mapsto \alpha_m(f)(x) &= \int K_m(x, z)f(z)\mu(dz); \end{aligned}$$

and $\{\alpha_m\}_{m\in\mathbb{I}}$ is a net of linear operators.

Thus, the delta estimator $\widehat{f}_n$ is universally asymptotically unbiased in $L_p(\mathbb{R}^d, \mathbb{B}^d, \mu)$, with $p \in [1, \infty)$, when

$$\lim_{m\in\mathbb{I}} \|\alpha_m(f) - f\|_{L_p(\mu)} = 0, \qquad \forall f \in L_p(\mathbb{R}^d, \mathbb{B}^d, \mu);$$

that is, $\{\alpha_m\}_{m\in\mathbb{I}}$ is a linear approximate identity. Further details can be found in Vidal-Sanz (1999). The next theorem provides sufficient conditions ensuring that $\{\alpha_m\}_{m\in\mathbb{I}}$ is a linear approximate identity. Related results can be found in Kantorovich and Akilov ((1982), Th. 3, p. 203).

THEOREM 2.1. *Central approximation theorem. Let* $\{\alpha_m\}_{m\in\mathbb{I}}$ *be a net of linear operators in a Banach space* $(B, \|\cdot\|_B)$, *such as* $L_p(\mu)$. *Suppose that,* (1) $\sup_{m\in\mathbb{I}} \|\alpha_m\|_B < \infty$, *where* $\|\alpha_m\|_B := \sup_{\|f\|_B\leq 1} \|\alpha_m(f)\|_B$, *and* (2) *there exists a* $\mathcal{G} \subset B$, *which is* *dense in B, such that,* $\lim_{m\in\mathbb{I}} \|\alpha_m(f) - f\|_B = 0$, $\forall f \in \mathcal{G}$. *Then,* $\{\alpha_m\}_{m\in\mathbb{I}}$ *is a linear approximate identity. Moreover, if* $\|\alpha_m\|_B < \infty$ *for each* $m \in \mathbb{I}$, *then conditions* 1) *and* 2) *are necessary.*

From the proof of Theorem 2.1, it follows that if $\|\alpha_m\|_B < \infty$ for each $m \in \mathbb{I}$, but condition 1) is not satisfied, then there exists a dense G_δ set $\mathcal{C} \subset B$, such that

$$\lim_{m\in\mathbb{I}} \|\alpha_m(f) - f\|_B = \infty, \qquad \forall f \in \mathcal{C},$$

(recall that a G_δ set is a countable intersection of open sets). This result is very relevant, since in B spaces without isolated points (such as L_p spaces), every dense G_δ set is non numerable (see e.g., Rudin (1966), Th. 5.3.3). For example, the Dirichlet linear operators in $L_1([-\pi,\pi])$ associated to Fourier series are bounded and satisfy the approximation property for trigonometric polynomials (that is a dense subset). However, the uniform boundedness condition fails. Hence there exists a G_δ dense set $\mathcal{C}$ of divergence. In other words, in $L_1([-\pi,\pi])$ there is a infinite non numerable dense set of densities, such that the bias of their Fourier series estimators tends to infinite in L_1 norm. On the other hand, in $L_p([-\pi,\pi])$, with $1 < p < \infty$, the Dirichlet operators are uniformly bounded and the approximation property holds.

The next corollary is relevant in a nonparametric context. It allows to interchange the limits or take joint limits in $\|\alpha_m(g_r) - g\|_B$.

COROLLARY 2.1. *Let* $(B, \|\cdot\|_B)$ *be a Banach space and* $\{\alpha_m\}_{m\in\mathbb{I}}$ *an approximate identity on such a space. If* $\|\alpha_m\|_B < \infty$ *for each* $m \in \mathbb{I}$, *then for all nets* $\{g_r\}_{r\in\mathbb{M}}$ *in* B, *such that* $\lim_{r\in\mathbb{M}} \|g_r - g\|_B = 0$, *it is satisfied that,*

$$\lim_{r\in\mathbb{M}} \lim_{m\in\mathbb{I}} \|\alpha_m(g_r) - g\|_B = \lim_{m\in\mathbb{I}} \lim_{r\in\mathbb{M}} \|\alpha_m(g_r) - g\|_B = \lim_{(m,r)\in\mathbb{I}\times\mathbb{M}} \|\alpha_m(g_r) - g\|_B = 0.$$

In order to get a smooth estimator, we often apply a linear approximate identity $\{a_m\}$ over some consistent but discontinuous nonparametric estimators $\widehat{g}_{r_n}$, e.g., histograms. For example, we can use some approximation methods, as interpolation techniques, B-splines, or some other linear approximator with continuous images. Notice that the smooth approximator a_{m_n} depends on the sample size n. The histogram $\widehat{g}_{r_n}$ is a sequence of curves that also depends on the sample size, and we take joint limits when $n \to \infty$. Thus, universal consistency of *frequency polygons* derives from the universal consistency of histograms and Corollary 2.1.

The following theorem provides conditions on the generalized kernel net $\{K_m(x,z)\}_{m\in\mathbb{I}}$, which are sufficient for guaranteeing that the net $\{\alpha_m\}_{m\in\mathbb{I}}$ is a linear approximate identity and, therefore, the delta estimator is universally asymptotically unbiased.

Notice that $\alpha_m(1)(x) = \int K_m(x,z)\mu(dz)$. We define the net of majorized operators of $\{\alpha_m\}_{m\in\mathbb{I}}$ as the net $\{|\alpha|_m\}_{m\in\mathbb{I}}$,

$$|\alpha|_m(f)(x) = \int |K_m(x,z)| f(z)\mu(dz).$$

THEOREM 2.2. *Assume that*

A.1. *$\{|\alpha|_m\}_{m\in\mathbb{I}}$ is uniformly bounded in $L_p(\mathbb{R}^d,\mathbb{B}^d,\mu)$, $1\le p<\infty$, i.e.,*

$$\sup_{m\in\mathbb{I}}\||\alpha|_m\|_{L_p(\mu)} := \sup_{m\in\mathbb{I}}\left\{\sup_{\|f\|_{L_p(\mu)}\le 1}\||\alpha|_m(f)\|_{L_p(\mu)}\right\}<\infty. \tag{2.1}$$

A.2. $\lim_{m\in\mathbb{I}}\|\alpha_m(1)-1\|_{L_p(\mu)}=0$.
A.3. *For all compact sets $C\subset\mathbb{R}^d$, $\mu(C)<\infty$.*
A.4. *For all $\delta>0$, and all compact sets $C\subset\mathbb{R}^d$,*

$$\lim_{m\in\mathbb{I}}\left\|\int_{\{z:\|x-z\|>\delta\}}|K_m(x,z)|\mu_C(dz)\right\|_{L_p(\mu_C)}=0,$$

where μ_C, is the restriction of μ to the compact set C.

Then, the net $\{\alpha_m\}_{m\in\mathbb{I}}$ is an approximate identity in $L_p(\mathbb{R}^d,\mathbb{B}^d,\mu)$, i.e., $\widehat{f}_n$ associated to $\{K_m(x,z)\}$ is universally asymptotically unbiased.

Assumption A.1 establishes that the net sequence $\{|\alpha|_m\}_{m\in\mathbb{I}}$ is uniformly bounded in $L_p(\mathbb{R}^d,\mathbb{B}^d,\mu)$. This condition is fairly easy to check when $p=1$ or $p=2$. If $p=1$ and $|K_m(x,z)|$ is continuous for almost all points $x,z\in\mathbb{R}^d$, then (2.1) is equal to

$$\sup_{m\in\mathbb{I}}\left\{\operatorname*{ess\,sup}_{z\in\mathbb{R}^d,\mu}\int|K_m(x,z)|\mu(dx)\right\},$$

where ess sup denotes the "essential supremum", i.e.,

$$\operatorname*{ess\,sup}_{z\in\mathbb{R}^d,\mu}|f(z)| := \inf_{\{B\in\mathbb{B}^d:\mu(B)=0\}}\sup_{z\in B}|f(z)|.$$

If $p=2$ then (2.1) is bounded by

$$\sup_{m\in\mathbb{I}}\left\{\left(\int|K_m(x,z)|^2\mu(dx)\mu(dz)\right)^{1/2}\right\}.$$

See DeVore and Lorentz ((1993), pp. 30–34) and Dunford and Schwartz (1957) for a discussion of these results.

A sufficient condition for A.2 is that there exists an $m_0\in\mathbb{I}$ such that $\forall m\ge m_0$,

$$\alpha_m(1)=1\qquad\text{a.s.}\quad[\mu].$$

If μ is a finite measure, a weaker sufficient condition for A.2 consists of assuming

$$\mu\left(\left\{x\in\mathbb{R}^d:\lim_{m\in\mathbb{I}}|\alpha_m(1)(x)-1|>\delta\right\}\right)=0,\qquad\forall\delta>0,$$
$$\sup_{m\in\mathbb{I}}|\alpha_m(1)(x)|\in L_p(\mathbb{R}^d,\mathbb{B}^d,\mu),$$

which implies A.2 by dominated convergence arguments, see e.g., Chung ((1974), p. 100) and Billingsley ((1986), p. 220).

If μ is a finite measure, condition A.3 holds. The Lebesgue measure also satisfies A.3. A sufficient condition for A.4 is that, when m increases, the support of $|K_m(x,z)|$ shrinks on $\{(x,z) : x = z\}$ and, possibly in other points of null measure.

PROPOSITION 2.1. *Condition* A.2 *in Theorem* 2.2 *can be substituted by,*
A.2′. $\lim_{m\in\mathbb{I}} \|\alpha_m(1) - 1\|_{L_\infty(\mu)} = 0.$

Condition A.4 may be difficult to check. The next proposition provides a sufficient condition.

PROPOSITION 2.2. *A sufficient condition for* A.4 *is*
A.4′. *For some* $s \geq 1$,

$$\lim_{m\in\mathbb{I}} \left\| \int \|x - z\|^s |K_m(x,z)| \mu(dz) \right\|_{L_p(\mu)} = 0.$$

In addition, the next proposition provides sufficient conditions for A.4′.

PROPOSITION 2.3. *The following conditions are sufficient for* A.4′: *for some* $s \geq 1$,
(i) $\int \|x - z\|^s |K_m(x,z)| \mu(dz) \to 0$, *a.s.* $[\mu]$, *(or in measure)*,
(ii) $\int \|x - z\|^s |K_m(x,z)| \mu(dz) < |T(x)|$, $T \in L_p(\mathbb{R}^d, \mathbb{B}^d, \mu)$.

Weaker sufficient conditions in Propositions 2.2 and 2.3 can be obtained, substituting μ by μ_C (i.e., the restriction of μ to C), for every compact set C.

Next, we check approximate identity conditions for some broad classes of nonparametric density estimators.

2.1 *Singular integral estimators*

Consider the class of *singular integral estimators* of a pdf $f \in L_p(\mathbb{R}^d, \mathbb{B}^d, \lambda)$, defined as,

$$\widehat{f}_n(x) = \frac{1}{n} \sum_{i=1}^{n} K_{m_n}(X_i - x). \tag{2.2}$$

Usually, in the nonparametric literature it is assumed that $K_m(u) = K_m(-u)$. These estimators are associated to the singular integral linear approximators in $L_p(\mathbb{R}^d, \mathbb{B}^d, \lambda)$,

$$\alpha_m(f)(x) = \int K_m(z - x) f(z) \lambda(dz).$$

The global unbiasedness of these estimators has been considered by Devroye and Györfi ((1985*a*), Chap. 12, Sec. 8), for the Lebesgue measure restricted to a finite interval.

Singular integral estimators encompass relevant families of nonparametric estimators like,

- *Kernel estimators* in $L_p(\mathbb{R}^d, \mathbb{B}^d, \lambda)$, that take,

$$K_H(u) = \frac{1}{\det(H)} \boldsymbol{K}(H^{-1}u),$$

with H a definite positive matrix, and the matrices are ordered by the relation "to have smaller $\|H\|$". For the multiplicative kernel $\boldsymbol{K}(u)=\prod_{j=1}^{d}\boldsymbol{K}_j(u_j)$, the matrices H are usually diagonal.

- *Fourier series estimators* in $L_p([-\pi,\pi])$, $1<p<\infty$, with Dirichlet's kernel,

$$K_m(u)=\frac{\sin\left(\left(m+\frac{1}{2}\right)u\right)}{2\pi\sin\left(\frac{1}{2}u\right)},\qquad m\in\mathbb{N}.$$

If $p=1$ this is not uniformly bounded, but we can use Fejer's kernel estimators.

- *Fejer estimators* in $L_p([-\pi,\pi])$, $1\leq p<\infty$, with

$$K_m(u)=\frac{1}{2\pi(m+1)}\left(\frac{\sin\left(\left(m+\frac{1}{2}\right)u\right)}{\sin\left(\frac{1}{2}u\right)}\right)^2.$$

There are many other examples, for a review see Butzer and Nessel (1971) and Devroye and Györfi ((1985*a*), Chap. 12). The next result is an application of Theorem 2.2, which provides universal asymptotic unbiasedness for these estimators.

PROPOSITION 2.4. *Assume that,*
S.1. $\{K_m\}_{m\in\mathbb{I}}\subset L_1(\mathbb{R}^d,\mathbb{B}^d,\lambda)$.
S.2. $\int K_m(u)du=1$, $\forall m\in\mathbb{I}$.
S.3. $\lim_{m\in\mathbb{I}}\int\|u\||K_m(u)|du=0$.
Then A.1 *to* A.4 *holds in* $L_p(\mathbb{R}^d,\mathbb{B}^d,\lambda)$, *with* $1\leq p<\infty$, *for the generalized kernels* $\{K_m(z-x)\}_{m\in\mathbb{I}}$.

2.2 *Histogram*

The class of histogram estimators is defined by means of measurable partitions. Let $\mathbb{I}$ be the set of all measurable Borel partitions of $\mathbb{R}^d$ in sets with finite and positive λ-measure. The set $\mathbb{I}$ is ordered by the partial preorder $m_1\leq m_2$ if m_2 is thinner than m_1 almost everywhere, in other words $\forall A_1\in m_1$, $A_2\in m_2$ then $A_2\subset A_1$, or $A_2\cap A_1=\emptyset$, except for a set of λ-measure zero. Then $\mathbb{I}$ is a directed set. Often, we take a regular subset $\mathbb{I}_0\subset\mathbb{I}$ of partitions of finite diameter, 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)).

Define the partitioning approximator by the generalized kernel,

$$K_m(x,z)=\sum_{A\in m}\frac{I_A(x)I_A(z)}{\lambda(A)},\tag{2.3}$$

with corresponding linear approximator,

$$\alpha_m(f)(x)=\sum_{A\in m}\left(\frac{\int_A f(z)\lambda(dz)}{\lambda(A)}\right)I_A(x).$$

The histogram estimator of $f \in L_1(\mathbb{R}^d, \mathbb{B}^d, \lambda)$ is,

$$\widehat{f}_n(x) = \sum_{A \in m_n} \left(\frac{\sum_{i=1}^n I_A(X_i)}{n\lambda(A)} \right) I_A(x).$$

This is the oldest nonparametric estimator (Graunt (1662) is an early reference) which has been studied by Révesz (1971, 1972, 1973, 1974), Tukey (1977), Scott ((1979), (1992, Chap. 3)) and Freedman and Diaconis (1981), among others. Universal consistency of the histogram has been established by Abou-Jaoude (1976*a*, *b*, *c*) and Devroye and Györfi (1985*a*). The following proposition illustrates the use of Theorem 2.2, although it is a well-known result in the literature.

PROPOSITION 2.5. *Consider a regular partitions set*, $\mathbb{I}_0$. *Then* A.1 *to* A.4 *hold for the generalized kernels in* (2.3).

The following result, which is based on Theorem 2.1, avoids the use of regularity condition on the partitions.

PROPOSITION 2.6. *Consider the space* $L_p(\mathbb{R}^d, \mathbb{B}^d, \mu)$, *with* $1 \leq p < \infty$. *If* μ *is absolutely continuous with respect to the Lebesgue measure* λ, *then the net of integral operators* $\{\alpha_m\}_{m \in \mathbb{I}}$, *with partition kernels defined by equation* (2.3), *is an approximate identity.*

2.3 *Estimators based on orthonormal Hilbert space bases*

Here, we consider the particular case where $f \in L_2(\mathbb{R}^d, \mathbb{B}^d, \mu)$. This is a Hilbert space with the inner product

$$\langle f, g \rangle_{L_2(\mu)} = \int f(z) g(z) \mu(dz).$$

The set $\{e_k(z)\}_{k=1}^m \subset L_2(\mathbb{R}^d, \mathbb{B}^d, \mu)$ is said to be orthonormal if $\langle e_k, e_s \rangle_{L_2(\mu)} = I_{\{k=s\}}$.

The orthogonal projection of an arbitrary $f \in L_2(\mathbb{R}^d, \mathbb{B}^d, \mu)$ into the linear subspace spanned by an orthonormal set $\{e_k(z)\}_{k=1}^m$, can be expressed as,

$$\alpha_m(f)(x) = \sum_{k=1}^m \langle f, e_k \rangle_{L_2(\mu)} \cdot e_k(x) = \sum_{k=1}^m \left(\int f(z) e_k(z) \mu(dz) \right) \cdot e_k(x).$$

Note that, if we define,

$$K_m(x, z) = \sum_{k=1}^m e_k(x) e_k(z), \tag{2.4}$$

then, the projection can be expressed as

$$\alpha_m(f)(x) = \int K_m(x, z) f(z) \mu(dz).$$

Thus, we say that a sequence $\{e_k(z)\}_{k=1}^\infty$ of orthonormal functions is an orthonormal Hilbert space basis if the sequence of projections $\{\alpha_m\}_{m=1}^\infty$ is a linear approximate identity on $L_2(\mathbb{R}^d, \mathbb{B}^d, \mu)$; or equivalently, if and only if the span of $\{e_k(z)\}_{k=1}^\infty$ is dense in

$L_2(\mathbb{R}^d, \mathbb{B}^d, \mu)$. Using Zorn's lemma, it can be proved that every Hilbert space has an orthonormal Hilbert basis (see Kreyszig (1978), p. 212).

Notice that the corresponding density estimator is just

$$(2.5)\quad \begin{cases} \widehat{f}_n(x) = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} K_m(x, X_i) = \dfrac{1}{n}\sum_{i=1}^{n}\left(\sum_{k=1}^{m} e_k(x)\cdot e_k(X_i)\right) = \sum_{k=1}^{m}\widehat{f}_{k,n}\cdot e_k(x), \\ \widehat{f}_{k,n} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} e_k(X_i). \end{cases}$$

This estimator was first considered by Čencov (1962) and Bosq (1969). The literature about density estimation by means of orthonormal basis is discussed in Devroye and Györfi ((1985*a*), Chap. 12).

PROPOSITION 2.7. *Assume that,*

O.1. $\{e_k(z)\}_{k=1}^{\infty}$ *is an orthonormal set in* $L_2(\mathbb{R}^d, \mathbb{B}^d, \mu)$, *such that*

$$\sup_{m\in\mathbb{N}}\left\{\sup_{\|f\|_{L_2(\mu)}\le 1}\left\|\int\left|\sum_{k=1}^{m} e_k(x)e_k(z)\right| f(z)\mu(dx)\right\|_{L_2(\mu)}\right\} < \infty.$$

O.2. *There exists a* $k_0 \in \mathbb{N}$ *such that* $e_{k_0}(x) = 1$, *a.s.* $[\mu]$.

O.3. *For all compact sets* C, $\mu(C) < \infty$.

O.4. $\forall\delta > 0$, *and all compact sets* C,

$$\sup_{m\in\mathbb{N}}\left\|\int\left(\sum_{k=1}^{m} e_k(x)e_k(z)\right)\|x - z\|\mu(dz)\right\|_{L_2(\mu_C)} = 0.$$

Then A.1 *to* A.4 *hold for the generalized kernel* (2.4), *and* $\{e_k(z)\}_{k=1}^{\infty}$ *is an orthonormal basis in* $L_2(\mathbb{R}^d, \mathbb{B}^d, \mu)$.

In the particular case that we use Fourier series, this method is equivalent to use the previous result on singular integral estimators.

A useful method for obtaining an orthonormal basis in $L_2(\mathbb{R}^d, \mathbb{B}^d, \mu)$, consists of applying the Graham-Schmidt orthonormalization algorithm to some dense subset of linearly independent functions (see Davis (1975) and Cheney (1982)). For example, if the monomials $\{x^k\}_{k=1}^{\infty}$ belong to $L_2(\mathbb{R}^d, \mathbb{B}^d, \mu)$, the associated orthonormal basis is known as the basis of orthonormal polynomials.

3. Universal convergence of the variation term

Most of the literature on universal consistency is based on Stone's (1977) theorem, whose conditions are usually difficult to check. Empirical process theory has also been applied in order to establish uniform consistency, see e.g., Silverman (1978), Stute (1982) and Pollard ((1984), pp. 35–36). Vapnik (1982), Devroye *et al.* (1996*b*) and Györfi *et al.* (2002) consider universal consistency of series estimators by means of *sieve-estimators theory* (see, e.g., van der Vaart and Wellner (1996), p. 321). Here, we present an alternative approach, providing sufficient conditions which are fairly easy to verify. In order to establish the almost everywhere, or in probability, convergence of the variation term, we use probability theory on Banach spaces. The universal convergence,

$$(3.1)\qquad \|\widehat{f}_n - E[\widehat{f}_n]\|_{L_p(\mu)} = \left\| \frac{1}{n}\sum_{i=1}^{n} Z_{n,i} - E[Z_{n,i}] \right\|_{L_p(\mu)} \overset{\text{a.e.}}{\to} 0,$$

with $Z_{n,i} = K_{m_n}(x; X_i)$, can be established applying a *Law of Large Numbers* (*LLN*) for triangular arrays in separable Banach spaces. If the convergence holds for every probability distribution $P \ll \mu$, with $f = dP/d\mu \in L_p(\mathbb{R}^d, \mathbb{B}^d, \mu)$, the result is universal. The case L_1 requires a separate analysis.

3.1 *Convergence in* $L_p(\mu)$, $1 < p < \infty$

There exists a large literature on probability theory in Banach spaces, see, e.g., Xia (1972), Hoffmann-Jørgensen (1974, 1976), Woyczyński (1978), Vakhania (1981), Schwartz (1981), Araujo and Giné (1980), Linde (1986), Pisier (1986, 1989), Vakhania *et al.* (1987) and Ledoux and Talagrand (1991), among others. Some LLN for Banach spaces have been considered by Taylor and Hu ((1987), Th. 4) and Hu and Chang ((1997), Th. 2.1). The row independence assumption has been weakened by Patterson and Taylor ((1997), Th. 3.2) by assuming weakly negative dependence and that the random elements have a compact and convex support. Patterson and Taylor ((1997), Th. 3.3) provide an additional LLN for triangular arrays in B-spaces with Schauder basis, using a negative dependence assumption. Other results for weighted sums of random elements in B-spaces have been provided by Ordoñez-Cabrera (1994), Hong *et al.* (2000) and Hu *et al.* (2001).

The asymptotic results for sums of random elements depend crucially on the geometric properties of the considered spaces. We say that a Banach space $(B, \|\cdot\|_B)$ is of *type-*γ, if $\exists c_\gamma > 0$ such that, for all finite sets $\{Z_i\}_{i=1}^n$ of independent random elements on B, it is satisfied that,

$$E\left[\left\|\sum_{i=1}^{n} Z_i\right\|_B^{\gamma}\right] \leq c_\gamma \cdot \sum_{i=1}^{n} E[\|Z_i\|_B^{\gamma}].$$

Note that, by the triangular inequality, every Banach space is of type-1. Thus, the only relevant case is $\gamma > 1$. On the other hand, the only spaces of *type-*γ for $\gamma > 2$ is the space $\{0\}$. Hence, the *type-*γ property is useful for $1 < \gamma \leq 2$. There exist many examples of *type-*γ spaces. For instance, all the Hilbert spaces are of *type-*2.

THEOREM 3.1. *Weak LLN for triangular arrays. Let* $(B, \|\cdot\|_B)$ *be a separable Banach space of type* $\gamma \in [1, 2]$ *and* $\{Z_{n,i} : 1 \leq i \leq n\}_{n\in\mathbb{N}}$ *a triangular array. Assume that the row elements are independent and* $E[\|Z_{n,i}\|_B] < \infty$ *(then, the Bochner expectation* $E[Z_{n,i}]$ *exists). Consider the centered sum,*

$$(3.2)\qquad \mathbb{S}_n = \sum_{i=1}^{n} (Z_{n,i} - E[Z_{n,i}]).$$

If,

B.1. $\lim_{n\to\infty} n^{-\gamma} \sum_{i=1}^{n} E[\|Z_{n,i}\|_B^{\gamma}] = 0$,

then $n^{-1}\|\mathbb{S}_n\|_B \overset{L_\gamma}{\to} 0$, *and therefore converges in probability.*

Remark 1. A sufficient condition for B.1 is

$$\max_{1\le i\le n} E[\|Z_{n,i}\|_B^\gamma] = o(n^{\gamma-1}).$$

and if $\{Z_{n,i}\}$ are i.i.d. by rows, $E[\|Z_{n,1}\|_B^\gamma] = o(n^{\gamma-1})$.

Assume that μ is a σ-finite measure on $(\mathbb{R}^d, \mathbb{B}^d)$, then: (a) every $L_p(\mathbb{R}^d, \mathbb{B}^d, \mu)$ space, with $1 \le p < 2$ is of *type-p*; and (b) every $L_p(\mathbb{R}^d, \mathbb{B}^d, \mu)$ space, with $p \ge 2$ is of *type-2*. For a proof see e.g. Araujo and Giné ((1980), Th. 7.2, p. 158). Therefore, taking $Z_{n,i} = K_{m_n}(x, X_i)$, the condition

$$E[\|K_{m_n}(x,X)\|_{L_p(\mu)}^p] = o(n^{p-1}),$$

is sufficient for the weak consistency of the variation term of delta estimators in $L_p(\mathbb{R}^d, \mathbb{B}^d, \mu)$, as $\{X_i\}$ are i.i.d.

Example 1. *Histogram*. Consider the histogram in $L_p(\mathbb{R}^d, \mathbb{B}^d, \lambda)$, $1 < p \le 2$, with kernel defined by equation (2.3). Notice that

$$\left|\sum_{A\in m} \frac{I_A(x)I_A(z)}{\lambda(A)}\right|^p = \sum_{A\in m}\left|\frac{I_A(x)I_A(z)}{\lambda(A)}\right|^p = \sum_{A\in m}\frac{I_A(x)I_A(z)}{\lambda(A)^p}, \qquad \text{a.e.,}$$

since the partitions are disjoints. Define

$$\chi(m) = \inf_{A\in m} \lambda(A).$$

Hence, if $\chi(m_n) \to 0$ with $n \cdot \chi(m_n) \to \infty$,

$$\begin{aligned}\frac{1}{n^{p-1}} E[\|K_{m_n}(x,X)\|_{L_p(\lambda)}^p] &= \frac{1}{n^{p-1}} E\left[\int \left|\sum_{A\in m}\frac{I_A(x)I_A(X)}{\lambda(A)}\right|^p d\lambda\right] \\ &= \frac{1}{n^{p-1}} E\left[\sum_{A\in m}\frac{I_A(X)}{\lambda(A)^{p-1}}\right] \le \frac{\sum_A P(A)}{n^{p-1}\chi(m_n)^{p-1}} \\ &= \frac{1}{[n\cdot\chi(m_n)]^{p-1}} \to 0.\end{aligned}$$

The universal consistency property in L_p follows from Theorem 3.1.

Example 2. *Orthonormal basis estimators*. Consider the orthonormal basis estimator of density functions in $L_2(\mathbb{R}^d, \mathbb{B}^d, \lambda)$, with kernel defined by equation (2.5). Assume that $m_n \to \infty$ with $m_n = o(n)$. Applying Fubini's theorem,

$$\begin{aligned}\frac{1}{n} E[\|K_{m_n}(x,X)\|_{L_2(\lambda)}^2] &= \frac{1}{n} E\left[\int \left|\sum_{k=1}^{m_n} e_k(x)\cdot e_k(X)\right|^2 dx\right] \\ &= \frac{1}{n}\sum_{k_1=1}^{m_n}\sum_{k_2=1}^{m_n}\left(\int e_{k_1}(x)e_{k_2}(x)dx\right)\cdot E[e_{k_1}(X)e_{k_2}(X)] \\ &= \frac{1}{n}\sum_{k=1}^{m_n} E[|e_k(X)|^2] \le \frac{m_n}{n}\max_{1\le k\le m} E[|e_k(X)|^2] \to 0,\end{aligned}$$

whenever $\max_{k\geq 1} E[|e_k(X)|^2] = \max_{k\geq 1} \int |e_k(z)|^2 f(z)dz < \infty$. We get universal consistency when this holds for any $f \geq 0$, $\|f\|_{L_2} \leq 1$.

The following result can be useful in order to check the conditions of the previous theorem.

PROPOSITION 3.1. *Let $\{K_{m_n}(x, X_i)\}$ be a triangular array in $L_p(\mathbb{R}^d, \mathbb{B}^d, \mu)$ with $1 \leq p \leq 2$ and μ a σ-finite measure. Assume that $E[\|K_{m_n}(x, X)\|^p_{L_p(\mu)}] < \infty$, then*

$$E[\|K_{m_n}(x, X)\|^p_{L_p(\mu)}] = \left\| \int |K_{m_n}(x, z)|^p \frac{dP}{d\mu}(z)\mu(dz) \right\|_{L_1(\mu)}.$$

Therefore, if

$$\left\| \int |K_{m_n}(x, z)|^p f(z)\mu(dz) \right\|_{L_1(\mu)} = o(n^{p-1}),$$

for all $f \geq 0$ with $\|f\|_{L_1(\mu)} \leq 1$, the variation term of delta estimators converges universally to zero in probability with respect to $\|\cdot\|_{L_p(\mu)}$ norm. This condition is readily checked for kernel estimators in the following example.

Example 3. *Kernel estimators.* Consider the kernel estimator in $L_p(\mathbb{R}^d, \mathbb{B}^d, \lambda)$, with $1 < p \leq 2$, defined by

$$K_{H_n}(x, z) = \frac{1}{\det(H_n)} \boldsymbol{K}(H_n^{-1}(z - x)).$$

It can be proved that $\forall f$ non negative with $\|f\|_{L_p(\lambda)} \leq 1$,

$$\left\| \int |\boldsymbol{K}(u)|^p f(x + H_n u)\lambda(du) \right\|_{L_1(\lambda)} = O(1).$$

Hence,

$$\begin{aligned}
&\frac{1}{n^{p-1}} \left\| \int |K_{H_n}(x, z)|^p f(z)\lambda(dz) \right\|_{L_1(\lambda)} \\
&\quad = \frac{1}{n^{p-1}\det(H_n)^{p-1}} \left\| \int |\boldsymbol{K}(u)|^p f(x + H_n u)\lambda(du) \right\|_{L_1(\lambda)} \\
&\quad = O\left(\frac{1}{[n \cdot \det(H_n)]^{p-1}}\right).
\end{aligned}$$

The sufficient condition is $n \cdot \det(H_n) \to \infty$, applying Proposition 3.1. Universal consistency of kernel density estimators was originally proved by Devroye (1983).

Note that, the above results are not useful when $p = 1$, because in this case,

$$E[\|Z_{n,i}\|_{L_1(\mu)}] = \left\| \int |K_{m_n}(x, z)| \frac{dP}{d\mu}(z)\mu(dz) \right\|_{L_1(\mu)} = O(1),$$

so that the sufficient condition $\max_{1\leq i\leq n} E[\|Z_{n,i}\|_{L_1(\mu)}] = o(1)$ is not satisfied. The next subsection is devoted to the L_1 convergence of the variation term.

3.2 *Convergence in $L_1(\mu)$*

In this section we provide an alternative approach, which is specially useful in order to deal with L_1-spaces.

THEOREM 3.2. *Let $\{Z_{n,i} = K_{m_n}(x, X_i)\}$ be a triangular array in $L_p(\mathbb{R}^d, \mathbb{B}^d, \mu)$ with $1 \le p \le 2$, and consider $\mathbb{S}_n(x)$ defined as in equation (3.2). Assume that:*

(1) $n^{-1} \max_{1\le i\le n} E[|K_{m_n}(x, X_i)|^2] \to 0$ a.e. $[\mu]$.

(2) $\int \sup_n E[|n^{-1}\mathbb{S}_n(x)|^p]\mu(dx) < \infty$.

Then $\|n^{-1}\mathbb{S}_n\|_{L_p(\mu)} \xrightarrow{p.} 0$.

If conditions (1) and (2) are satisfied for all probability distributions $P \ll \mu$, with $f = dP/d\mu \in L_p(\mu)$, the weak universal L_p-convergence of the variation term follows.

Remark 2. Condition (2), is satisfied when,

$$\sup_{\|f\|_{L_p(\mu)}\le 1, f\ge 0} \int \left(\sup_m \int |K_m(x,z)|^p f(z)\mu(dz)\right) \mu(dx) < \infty.$$

Define the maximal operator in $L_p(\mathbb{R}^d, \mathbb{B}^d, \mu)$,

$$\alpha_p^*(f, x) = \sup_{m\in\mathbb{I}} \int |K_m(x,z)|^p f(z)\mu(dz),$$

which is a sub linear operator (i.e., $\alpha_p^*(f+g) \le \alpha_p^*(f) + \alpha_p^*(g)$ for all $f, g \in L_p$). Then, in order to establish condition (2) universally, we have to check that there exists some $c > 0$ such that,

$$\|\alpha_p^*(f,x)\|_{L_1(\mu)} < c\|f\|_{L_p(\mu)}, \tag{3.3}$$

for any $f \in L_p$, with $f \ge 0$.

This result is particularly useful to prove weak universal consistency on $L_1(\mathbb{R}^d, \mathbb{B}^d, \lambda)$. In this context, expression (3.3) means that for all non negative $f \in L_1$,

$$\int \alpha_1^*(f,x)dx = \int \sup_{m\in\mathbb{I}} |\alpha|_m(f,x)dx < c\|f\|_{L_1(\lambda)}.$$

This kind of properties often can be proved using the Hardy-Littlewood-Paley theory. The Hardy-Littlewood maximal operator on $L_p(\mathbb{R}^d, \mathbb{B}^d, \lambda)$,

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

satisfies for some $c_{p,d} > 0$, that $\|\beta^*(f,x)\|_{L_p(\lambda)} \le c_{p,d}\|f\|_{L_p(\lambda)}$ for all $f \in L_p$. For details, see Stein (1970), de Guzman (1975) and Wheeden and Zygmund (1977), among others.

Example 4. Consider the kernel estimator in $L_1(\mathbb{R}^d, \mathbb{B}^d, \lambda)$, defined by means of,

$$K_{H_n}(x,z) = \frac{1}{\det(H_n)} \boldsymbol{K}(H_n^{-1}(z-x)).$$

Hence, for a.e. $x \in \mathbb{R}^d$, and all densities f,

$$\begin{aligned} n^{-1}E[|K_{H_n}(x,X)|^2] &= \frac{1}{n}\int |K_{H_n}(x,z)|^2 f(z)\lambda(dz) \\ &= \frac{1}{n\det(H_n)^2}\int |K(u)|^2 f(x+H_n u)du = O\left(\frac{1}{n\cdot\det(H_n)}\right). \end{aligned}$$

If $K(\cdot)$ has compact support C, and there exist constants $c_1, c_2 > 0$ such that $c_1 I_C(u) \le |K(u)| \le c_2 I_C(u)$ then,

$$\int \sup_{H>0}\int |K_H(z-x)|f(z)dzdx \le c\|f\|_{L_1(\lambda)},$$

for all integrable non negative f. Assuming $n\cdot\det(H_n)\to\infty$, the L_1-universal consistency property follows from Theorem 3.2.

Example 5. Consider the histogram in $L_1(\mathbb{R}^d, \mathbb{B}^d, \lambda)$, with regular partitions and kernel defined by equation (2.3). For regular partitions, the uniform boundedness condition

$$\begin{aligned} &\int \left(\sup_m \int \left(\sum_{A\in m}\frac{I_A(x)I_A(z)}{\lambda(A)}\right) f(z)dz\right) dx \\ &\quad = \int \sup_m \sum_{A\in m}\frac{I_A(x)P_f(A)}{\lambda(A)}dx \le c\|f\|_{L_1(\lambda)}, \end{aligned}$$

is satisfied. This is because the Hardy-Littlewood maximal function,

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

satisfies $\|\beta^*(f,x)\|_{L_1(\lambda)} \le c_d\|f\|_{L_1(\lambda)}$. Furthermore, by the Lebesgue density theorem,

$$\lim_{m\in\mathbb{I}_0}\sum_{A\in m}\frac{P_f(A)}{\lambda(A)}I_A(x) = f(x), \qquad \text{a.e.}$$

Next, define $\chi(m) = \inf_{A\in m}\lambda(A) > 0$, which depends on the dimension d exponentially, as λ is the Lebesgue measure on $\mathbb{R}^d$. Condition $n\cdot\chi(m)\to\infty$ implies that

$$\begin{aligned} n^{-1}E[|K_m(x,X)|^2] &= \frac{1}{n}E\left[\sum_{A\in m}\left|\frac{I_A(x)I_A(X)}{\lambda(A)}\right|^2\right] \\ &= \frac{1}{n}\sum_{A\in m}\frac{P(A)}{\lambda(A)^2}I_A(x) \le \frac{1}{n\cdot\chi(m)}\sum_{A\in m}\frac{P(A)}{\lambda(A)}I_A(x) \\ &\le \frac{\beta^*(f,x)}{n\cdot\chi(m)} \to 0, \qquad \text{a.e.}, \end{aligned}$$

as $\beta^*(f,x) < \infty$ a.e. by Fubini's theorem. The L_1-universal consistency property follows applying Theorem 3.2.

3.3 *Strong convergence*

In order to obtain strong convergence, we will use boundedness conditions that are usually satisfied in the nonparametric framework. The argument is based on Devroye (1991).

THEOREM 3.3. *Strong consistency. Let* $\{K_{m_n}(x, X_i)\}$ *be a triangular array on* $L_p(\mathbb{R}^d, \mathbb{B}^d, \mu)$, *with* $1 \leq p < \infty$, *and* $\widehat{f}_{m_n} = n^{-1} \sum_{i=1}^{n} K_{m_n}(x, X_i)$. *If there exists an* $M > 0$ *such that,*

$$\|K_{m_n}(x, X_i)\|_{L_p} \leq M, \qquad a.e., \qquad for \quad i \in \{1, \dots, n\}, \quad \forall n \in \mathbb{N},$$

then the following assertions are equivalent,

(i) $\|\widehat{f}_{m_n} - E[\widehat{f}_{m_n}]\|_{L_p} \overset{completely}{\longrightarrow} 0,$

(ii) $\|\widehat{f}_{m_n} - E[\widehat{f}_{m_n}]\|_{L_p} \overset{a.e.}{\longrightarrow} 0,$

(iii) $\|\widehat{f}_{m_n} - E[\widehat{f}_{m_n}]\|_{L_p} \overset{p.}{\longrightarrow} 0.$

The boundedness condition can be weakened to $\|K_{m_n}(x, X_i)\|_{L_p} \leq M_n$, *a.e., with* $\sum_{n=1}^{\infty} \exp\{-n/M_n^2\} < \infty$.

Define,

$$M_n = \operatorname*{ess\,sup}_{z \in \mathbb{R}^d, P} \|K_{m_n}(x, z)\|_{L_p(\mu)} = \operatorname*{ess\,sup}_{z \in \mathbb{R}^d, P} \left(\int |K_{m_n}(x, z)|^p \mu(dx) \right)^{1/p}.$$

(The essential supremum with respect to P). In order to study strong consistency, we have to consider the behavior of this sequence.

Example 6. The kernel estimator trivially satisfies that,

$$M_n = \sup_{z \in \mathbb{R}^d} \int |K_{H_n}(z - x)| dx = \|K\|_{L_1(\lambda)} < \infty.$$

In the case $p > 1$,

$$\begin{aligned} M_n &= \sup_{z \in \mathbb{R}^d} \left(\int |K_{H_n}(z - x)|^p dx \right)^{1/p} = \left(\int |K_{H_n}(x)|^p dx \right)^{1/p} \\ &= \left(\det(H_n^{-1})^p \int |K(H_n^{-1} x)|^p dx \right)^{1/p} = \det(H_n^{-1})^{(p-1)/p} \|K\|_{L_p(\lambda)}, \end{aligned}$$

so that, it is sufficient that $\sum_{n=1}^{\infty} \exp\{-n/\det(H_n^{-1})^{2(p-1)/p}\} < \infty$.

Example 7. The histogram satisfies, for $p = 1$, that,

$$M_n = \operatorname*{ess\,sup}_{z \in \mathbb{R}^d} \int \left| \sum_{A \in m_n} \frac{I_A(x) I_A(z)}{\lambda(A)} \right| dx = \sup_{z \in \mathbb{R}^d} \sum_{A \in m_n} I_A(z) = 1 < \infty,$$

and for $p = 2$,

$$\begin{aligned} M_n^2 &= \operatorname*{ess\,sup}_{z \in \mathbb{R}^d} \int \left| \sum_{A \in m_n} \frac{I_A(x) I_A(z)}{\lambda(A)} \right|^2 dx = \operatorname*{ess\,sup}_{z \in \mathbb{R}^d} \int \sum_{A \in m_n} \frac{I_A(x) I_A(z)}{\lambda(A)^2} dx \\ &= \sup_{z \in \mathbb{R}^d} \sum_{A \in m_n} \frac{I_A(z)}{\lambda(A)} \leq \frac{1}{\chi(m_n)}, \end{aligned}$$

so that, it is sufficient that $\chi(m_n) \to 0$, with $\sum_{n=1}^{\infty} \exp\{-n\chi(m_n)\} < \infty$.

4. Final remarks

In order to establish the universal consistency of a delta estimator in $L_p(\mathbb{R}^d, \mathbb{B}^d, \mu)$ with $1 < p \leq 2$, two conditions should be verified,

(1) The net $\{\alpha_m\}_{m\in\mathbb{I}}$ is an approximate identity in L_p, and $\{m_n\}_{n\in\mathbb{N}}$ is a divergent sequence in $\mathbb{I}$.

(2) $E[\|K_{m_n}(x, X)\|^p_{L_p(\mu)}] = o(n^{p-1})$, universally.

If there exists some function $\varphi(m)$ which tends to zero with m, such that

$$E[\|K_{m_n}(x, X)\|^p_{L_p(\mu)}] = O\left(\frac{1}{\varphi(m_n)^{p-1}}\right),$$

then condition (2) can be replaced by $n \cdot \varphi(m_n) \to \infty$. The first condition establishes the convergence of the bias term, and the second one implies that the variation term tends to zero. Usually there is a *trade-off* between bias and variance. As m_n increases, the bias decreases, but the variance increases. An analogous result holds in $L_1(\mathbb{R}^d, \mathbb{B}^d, \lambda)$, replacing condition (2) by the assumption

$$E[|K_{m_n}(x, X_i)|^2] = o(n),$$

for a.e. $x \in \mathbb{R}^d$, universally, and an appropriate uniform boundedness condition.

The equivalence of weak and strong consistency is proved under boundedness conditions. If $\{K_{m_n}(x, X_i)\}$ are positive and integrate one with respect to x, which is usually the case in density estimation, the boundedness conditions are usually satisfied in the L_1-space.

We have not covered problems associated to the choice of an optimal smoothing parameter, usually defined as

$$m_n^* = \inf_{m\in\mathbb{I}} E\left[\left\|\frac{1}{n}\sum_{i=1}^{n} K_m(x, X_i) - f(x)\right\|^p_{L_p(\mu)}\right].$$

In this general framework, an optimum m_n^* could not exist. For example, if we consider histograms there is not an optimal partition $m_n^* \in \mathbb{I}$ unless we introduce additional restrictions on the shape of the partitions. Nevertheless, there are many families of nonparametric estimators with optimal smoothing parameter. In the main cases it is possible to estimate m_n^* universally, see Devroye and Lugosi (2001) for a monograph of this topic.

5. Proofs

5.1 *Proofs of Section* 2

PROOF OF THEOREM 2.1. First we prove that conditions (1) and (2) are sufficient.

Part I: *Sufficient conditions*. Assume that $\{\alpha_m\}_{m\in\mathbb{I}}$ is uniformly bounded, and there exists a dense set $\mathcal{G} \subset B$, such that

$$\lim_{m\in\mathbb{I}} \|\alpha_m(g) - g\|_B = 0, \qquad \forall g \in \mathcal{G}.$$

By denseness, $\forall \varepsilon > 0$ and $\forall g \in B$, $\exists \widetilde{g} \in \mathcal{G}$ such that $\|g - \widetilde{g}\|_B \leq \varepsilon$. By assumption, $\forall \widetilde{g} \in \mathcal{G}$, $\exists m_0 \in \mathbb{I}$ such that $\forall m \geq m_0$, $\|\alpha_m(\widetilde{g}) - \widetilde{g}\|_B \leq \varepsilon$. Then, using the assumption that $\{\alpha_m\}_{m\in\mathbb{I}}$ is uniformly bounded, the linearity of α_m and the triangular inequality, it is satisfied that,

$$\begin{aligned}\|\alpha_m(g) - g\|_B &\leq \|\alpha_m(g) - \alpha_m(\widetilde{g})\|_B + \|\alpha_m(\widetilde{g}) - \widetilde{g}\|_B + \|\widetilde{g} - g\|_B \\ &\leq \|\alpha_m(g - \widetilde{g})\|_B + \varepsilon + \varepsilon \leq M \cdot \|g - \widetilde{g}\|_B + 2\varepsilon \leq (M+2)\varepsilon.\end{aligned}$$

Since ε is arbitrarily small, the result follows.

Part II: Necessary conditions. Assume that $\{\alpha_m\}_{m\in\mathbb{I}}$ is a linear approximate identity in B. Then, trivially the approximation property holds for dense sets $\mathcal{G} \subset B$.

Assume that $\{\alpha_m\}_{m\in\mathbb{I}}$ are bounded operators, but not uniformly bounded. By the Banach Steinhaus theorem (see, e.g., Rudin (1966)), $\exists \mathcal{C} \subset B$ that is a dense G_δ set, such that

$$\sup_{m\in\mathbb{I}} \|\alpha_m(g)\|_B = \infty, \qquad \forall g \in \mathcal{C}.$$

Furthermore $\sup_{m\in\mathbb{I}} \|\alpha_m(g) - g\|_B = \infty$, for all $g \in \mathcal{C}$, since $\|g\|_B < \infty$ and by the triangular inequality $\|\alpha_m(g) - g\|_B \geq |\|\alpha_m(g)\|_B - \|g\|_B|$, $\forall g \in B$. On the other hand, $\forall m \in \mathbb{I}$

$$\|\alpha_m(g) - g\|_B \leq \|\alpha_m(g)\|_B + \|g\|_B < \infty,$$

hence, the supremum is equal to the limit,

$$\lim_{m\in\mathbb{I}} \|\alpha_m(g) - g\|_B = \sup_{m\in\mathbb{I}} \|\alpha_m(g) - g\|_B = \infty, \qquad \forall g \in \mathcal{C}.$$

This contradicts the assumption that $\{\alpha_m\}_{m\in\mathbb{I}}$ is an approximate identity.

PROOF OF COROLLARY 2.1. It is sufficient to consider that $\sup_{m\in\mathbb{I}} \|\alpha_m\|_B < \infty$. Hence,

$$\begin{aligned}\|\alpha_m(g_r) - g\|_B &\leq \|\alpha_m(g_r) - \alpha_m(g)\|_B + \|\alpha_m(g) - g\|_B \\ &\leq \sup_{m\in\mathbb{I}} \|\alpha_m\|_B \cdot \|g_r - g\|_B + \|\alpha_m(g) - g\|_B.\end{aligned}$$

PROOF O THEOREM 2.2. The theorem follows applying Theorem 2.1 and the following lemmas.

LEMMA 5.1. *Let $\{\alpha_m\}_{m\in\mathbb{I}}$ be a net of linear operators in $L_p(\mathbb{R}^d, \mathbb{B}^d, \mu)$ with $1 \leq p < \infty$. Then for all $m \in \mathbb{I}$, the norm of the operator verifies,*

$$\|\alpha_m\|_{L_p(\mu)} \leq \||\alpha|_m\|_{L_p(\mu)}.$$

Furthermore, the uniform boundedness of $\{|\alpha|_m\}_{m\in\mathbb{I}}$ implies the uniform boundedness of $\{\alpha_m\}_{m\in\mathbb{I}}$.

PROOF. Consider a Riesz space $(B, \leq)$, defined as a linear space B endowed with a partial preorder $\leq$, such that for all pairs $\{f, g\} \subset B$, their supremum and infimum both exist. For any $f \in B$, define $|f| = \sup\{f, -f\}$. We say that $(B, \|\cdot\|_B, \leq)$ is a

Banach lattice if $(B, \|\cdot\|_B)$ is a Banach space, $(B, \leq)$ is a Riesz space and the norm $\|\cdot\|_B$ is lattice (i.e., $|f| \leq |g|$ implies that $\|f\|_B \leq \|g\|_B$ for all $f, g \in B$).

In particular all $L_p(\mathbb{R}^d, \mathbb{B}^d, \mu)$ spaces, endowed with the preorder "$f(x) \leq g(x)$ almost everywhere" are Banach lattices. If $f \in L_p(\mathbb{R}^d, \mathbb{B}^d, \mu)$, then,

$$|\alpha_m(f)(x)| \leq \int |K_m(x,z)||f(z)|\mu(dz) = |\alpha|_m(|f|; x) \qquad \text{a.s.} \quad [\mu].$$

As the norm $\|\cdot\|_{L_p(\mu)}$ is lattice, $\forall f \in L_p$,

$$\|\alpha_m(f)\|_{L_p(\mu)} \leq \||\alpha|_m(|f|)\|_{L_p(\mu)} \leq \||\alpha|_m\|_{L_p} \cdot \|f\|_{L_p(\mu)}.$$

The previous lemma and assumption A.1 imply that $\{\alpha_m\}_{m\in\mathbb{I}}$ is uniformly bounded. Note that the space $C_c(\mathbb{R}^d)$ of continuous functions with compact support is dense in $L_p(\mathbb{R}^d, \mathbb{B}^d, \mu)$, with $1 \leq p < \infty$ (see e.g., Rudin (1966), Th. 3.3.1). By Theorem 2.1, it is sufficient to establish that,

$$\lim_{m\in\mathbb{I}} \|\alpha_m(f) - f\|_{L_p(\mu)} = 0, \qquad \forall f \in C_c(\mathbb{R}^d).$$

LEMMA 5.2. *If the net $\{\alpha_m\}_{m\in\mathbb{I}}$ satisfies the conditions of Theorem 2.2, for all $f \in C_c(\mathbb{R}^d)$,*

$$\lim_{m\in\mathbb{I}} \|\alpha_m(f) - f\|_{L_p(\mu)} = 0.$$

PROOF. For any $f \in C_c(\mathbb{R}^d)$, define $h(x,z) = f(z) - f(x)$. By A.2,

$$\begin{aligned} \|\alpha_m(f)(x) - f(x)\|_{L_p(\mu)} &\leq \left\| \int h(x,z) K_m(x,z)\mu(dz) \right\|_{L_p(\mu)} \\ &\quad + \|f(x)\|_\infty \|\alpha_m(\mathbf{1})(x) - \mathbf{1}\|_{L_p(\mu)} \\ &= \left\| \int h(x,z) K_m(x,z)\mu(dz) \right\|_{L_p(\mu)} + o(1). \end{aligned} \tag{5.1}$$

Since $h(x,z) \in C_c(\mathbb{R}^d \times \mathbb{R}^d)$ we can restrict the measure μ to a compact set C. The restricted measure is denoted by μ_C. Then, the first term in (5.1) is bounded by

$$\begin{aligned} &\left\| \int_{\{z:\|x-z\|\leq\delta\}} |h(x,z)||K_m(x,z)|\mu_C(dz) \right\|_{L_p(\mu_C)} \\ &\quad + \left\| \int_{\{z:\|x-z\|>\delta\}} h(x,z) K_m(x,z)\mu_C(dz) \right\|_{L_p(\mu_C)}. \end{aligned} \tag{5.2}$$

Since f is uniformly continuous, for all $\varepsilon > 0$, $\exists\delta > 0$, such that, if $\|x - z\| \leq \delta$ then $|h(x,z)| = |f(x) - f(z)| \leq \varepsilon$. Thus, (5.2) is bounded by,

$$\varepsilon \cdot \left\| \int |K_m(x,z)|\mu_C(dz) \right\|_{L_p(\mu_C)} + \left\| \int_{\{z:\|x-z\|>\delta\}} h(x,z) K_m(x,z)\mu_C(dz) \right\|_{L_p(\mu_C)}.$$

By A.1 and A.3 the first term is arbitrarily small, since

$$\begin{aligned}\sup_{m\in\mathbb{I}}\left\|\int |K_m(x,z)|\mu_C(dz)\right\|_{L_p(\mu_C)} &\leq \sup_{m\in\mathbb{I}}\left\|\int |K_m(x,z)|\cdot I_C(z)\mu(dz)\right\|_{L_p(\mu)}\\ &\leq \sup_{m\in\mathbb{I}}\||\alpha|_m\|_{L_p(\mu)}\cdot\|I_C\|_{L_p(\mu)}\\ &= \sup_{m\in\mathbb{I}}\||\alpha|_m\|_{L_p(\mu)}\cdot\mu(C)^{1/p}<\infty.\end{aligned}$$

The second term is bounded by

$$\|h\|_\infty\cdot\left\|\int_{\{z:\|x-z\|>\delta\}}|K_m(x,z)|\mu_C(dz)\right\|_{L_p(\mu_C)}\underset{m\in\mathbb{I}}{\rightarrow}0$$

by $\|h\|_\infty<\infty$ and A.4.

PROOF OF PROPOSITION 2.1. The proof is similar to Theorem 2.2, noticing that,

$$\begin{aligned}\|\alpha_m(f)(x)-f(x)\|_{L_p(\mu)} \leq{}& \left\|\int h(x,z)K_m(x,z)\mu(dz)\right\|_{L_p(\mu)}\\ &+\|\alpha_m(\mathbf{1})(x)-1\|_{L_\infty(\mu)}\cdot\|f\|_{L_p(\mu)}.\end{aligned}$$

PROOF OF PROPOSITION 2.2. Since $|\alpha|_m$ is a monotone operator and the norm $\|\cdot\|_{L_p(\mu)}$ is lattice,

$$\left\|\int_{\{z:\|x-z\|>\delta\}}|K_m(x,z)|\mu_C(dz)\right\|_{L_p(\mu_C)}\leq\delta^{-s}\left\|\int\|x-z\|^s|K_m(x,z)|\mu(dz)\right\|_{L_p(\mu)}\to 0,$$

thus A.4 is satisfied.

PROOF OF PROPOSITION 2.3. This is an immediate consequence of Lebesgue's theorem of dominated convergence.

PROOF OF PROPOSITION 2.4. We use Theorem 2.2. First note that assumption S.1 implies A.1, as a consequence of the next result.

LEMMA 5.3. *Generalized Young's inequality. Set $K_m(u)\in L_1(\mathbb{R}^d,\mathbb{B}^d,\lambda)$. Then $\forall f\in L_p(\mathbb{R}^d,\mathbb{B}^d,\lambda)$ with $1\leq p<\infty$,*

$$\||\alpha|_m(f)(x)\|_{L_p(\lambda)}=\left(\int\left|\int|K_m(z-x)|f(z)dz\right|^p dx\right)^{1/p}\leq\|K_m\|_{L_1(\lambda)}\|f\|_{L_p(\lambda)}.$$

PROOF. Using the integral Minkowski's inequality, Fubini's theorem and the invariance of the Lebesgue measure under translation, it is satisfied that,

$$\begin{aligned}
\||\alpha|_m(f)(x)\|_{L_p(\lambda)} &= \left(\int\left|\int |K_m(z-x)|\cdot f(z)dz\right|^p dx\right)^{1/p} \\
&= \left(\int\left(\int |K_m(u)|\cdot|f(x+u)|du\right)^p dx\right)^{1/p} \\
&\leq \int\left(\int (|K_m(u)|\cdot|f(x+u)|)^p dx\right)^{1/p} du \\
&= \int |K_m(u)|\left(\int |f(x+u)|^p dx\right)^{1/p} du = \|K_m\|_{L_1(\lambda)}\cdot\|f\|_{L_p(\lambda)}.
\end{aligned}$$

Note that λ satisfies A.3. Assumption A.4 is a consequence of S.3. For each compact set C,

$$\int |K_m(z-x)|\|x-z\|\lambda_C(dz) = \int_{z\in C} |K_m(z-x)|\|z-x\|dz = \int_{u\in C-x} |K_m(u)|\|u\|du,$$

after a change of variable $u = z - x$. Then,

$$\begin{aligned}
\left\|\int |K_m(x-z)|\|x-z\|\lambda_C(du)\right\|_{L_p(\lambda_C)} &= \left\|\int_{C-x} |K_m(u)|\|u\|du\right\|_{L_p(\lambda_C)} \\
&\leq \int |K_m(u)|\|u\|du\cdot\|1\|_{L_p(\lambda_C)} \\
&= \int |K_m(u)|\|u\|du\cdot\lambda(C)^{1/p}\to 0.
\end{aligned}$$

PROOF OF PROPOSITION 2.5. We apply Theorem 2.2. First, we check condition A.1. Note that α_m is a positive operator, so $\alpha_m = |\alpha|_m$. For all $m \in \mathbb{I}$,

$$\|\alpha_m\|_{L_1(\mu)} = \operatorname*{ess\,sup}_{z\in\mathbb{R}^d,\lambda}\int\left|\sum_{A\in m}\frac{I_A(x)I_A(z)}{\lambda(A)}\right|dx = \operatorname*{ess\,sup}_{z\in\mathbb{R}^d,\lambda}\sum_{A\in m} I_A(z) = 1,$$

then $\{|\alpha|_m\}_{m\in\mathbb{I}}$ is uniformly bounded in $L_1(\mathbb{R}^d, \mathbb{B}^d, \lambda)$.

A.2 is immediate since, $\forall m \in \mathbb{I}$,

$$\alpha_m(1)(x) = \int\left(\sum_{A\in m}\frac{I_A(x)I_A(z)}{\lambda(A)}\right)dz = \sum_{A\in m} I_A(x) = 1 \qquad \text{a.s.} \quad [\lambda].$$

The measure λ satisfies A.3. Now we will check A.4. Let λ_C be the restriction of λ

to any compact set C, then,

$$\begin{aligned}
&\lim_{m\in\mathbb{I}}\left\|\int_{\{z:\|x-z\|>\delta\}}|K_m(x,z)|\lambda_C(dz)\right\|_{L_1(\lambda_C)}\\
&\quad=\lim_{m\in\mathbb{I}}\int\left|\int_{\{z:\|x-z\|>\delta\}}\left(\sum_{A\in m}\frac{I_A(x)I_A(z)}{\lambda(A)}\right)\lambda_C(dz)\right|\lambda_C(dx)\\
&\quad\le\lim_{m\in\mathbb{I}}\int\left|\sum_{A\in m}\left(\frac{\int_{\{\{z:\|x-z\|>\delta\}\cap A\}}\lambda(dz)}{\lambda(A)}\right)I_A(x)\right|\lambda_C(dx).
\end{aligned}$$

But we can take a fine enough partition $m_\delta \in \mathbb{I}_0$ with a maximum diameter arbitrarily small, i.e., such that for every $A \in m_\delta$, $\sup_{x,z\in A}\|x-z\| \le \delta$. The same holds for all $m \ge m_\delta$. Then

$$\sup_{x\in A}\frac{\int_{\{\{z:\|x-z\|>\delta\}\cap A\}}\lambda(dz)}{\lambda(A)}=\frac{\lambda(\emptyset)}{\lambda(A)}=0,\qquad \forall A\in m,\ \forall m\ge m_\delta.$$

Thus, $\forall m \ge m_\delta$,

$$\sum_{A\in m}\left(\frac{\int_{\{\{z:\|x-z\|>\delta\}\cap A\}}\lambda(dz)}{\lambda(A)}\right)I_A(x)=0\qquad \text{a.s.}\quad [\lambda],$$

and by dominated convergence,

$$\lim_{m\in\mathbb{I}}\int\left|\sum_{A\in m}\left(\frac{\int_{\{\{z:\|x-z\|>\delta\}\cap A\}}\lambda(dz)}{\lambda(A)}\right)I_A(x)\right|\lambda_C(dx)=0.$$

PROOF OF PROPOSITION 2.6. First, notice that the set $\mathcal{S}$ of all simple and measurable functions g, such that $\mu(\{x \in \mathbb{R}^d : |g(z)| > 0\}) < \infty$ is dense in $L_p(\mathbb{R}^d, \mathbb{B}, \mu)$, whenever $1 \le p < \infty$ (see e.g., Rudin (1966), Th. 3.2.8).

LEMMA 5.4. *The partition net* $\{\alpha_m\}_{m\in\mathbb{I}}$ *satisfies the approximation theorem on* $L_p(\mathbb{R}^d, \mathbb{B}, \mu)$, *with* $1 \le p < \infty$ *for all* $g \in \mathcal{S}$.

PROOF. Simple functions $g \in \mathcal{S}$, can be expressed as,

$$g(z)=\sum_{r=1}^{s}\beta_r\cdot I_{B_r}(z),$$

for some finite measurable partition $\overline{m} = (B_1, \dots, B_s)$ of $\mathbb{R}^d$, with $\mu(B_r) < \infty$ for $r = 1, \dots, s$. By definition,

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

Thus, using that $\sum_{A\in m} I_A(x) \overset{\text{a.e.}[\mu]}{=} 1$,

$$\begin{aligned}\|\alpha_m(g)(x)-g(x)\|_{L_p(\mu)} &= \left\|\sum_{A\in m}\left(\sum_{r=1}^{s}\beta_r\frac{\mu(A\cap B_r)}{\mu(A)}\right)I_A(x)-\sum_{r=1}^{s}\beta_r I_{B_r}(x)\right\|_{L_p(\mu)}\\ &= \left\|\sum_{A\in m}\frac{1}{\mu(A)}\left(\sum_{r=1}^{s}\beta_r[\mu(A\cap B_r)-\mu(A)I_{B_r}(x)]\right)I_A(x)\right\|_{L_p(\mu)}.\end{aligned}$$

Therefore,

$$\|\alpha_m(g)(x)-g(x)\|_{L_p(\mu)} \le \sum_{A\in m}\frac{1}{\mu(A)}\left\|\sum_{r=1}^{s}\beta_r[\mu(A\cap B_r)I_A(x)-\mu(A)I_{B_r}(x)I_A(x)]\right\|_{L_p(\mu)},$$

which tends to zero when m increases. Namely, if $\overline{m}\le m$, then $\forall B_r\in\overline{m}$ and $\forall A\in m$, we have one of the following cases:

(i) $A\cap B_r=\emptyset$ and therefore $\mu(A\cap B_r)=0$, $I_{\{A\cap B_r\}}(x)=I_\emptyset(x)=0$, or

(ii) $A\subset B_r$ and thus $\mu(A\cap B_r)=\mu(A)$, $I_{\{A\cap B_r\}}(x)=I_A(x)$.

This implies that $\forall B_r\in\overline{m}\le m$,

$$\begin{aligned}&\|\mu(A\cap B_r)I_A(x)-\mu(A)I_{\{A\cap B_r\}}(x)\|_{L_p(\mu)}\\ &=\left(\int|\mu(A\cap B_r)I_A(x)-\mu(A)I_{\{A\cap B_r\}}(x)|^p\mu(dx)\right)^{1/p}\\ &=0.\end{aligned}$$

In other words, $\forall g\in\mathcal{S}$, $\exists\overline{m}$ such that $\forall m\ge\overline{m}$ the approximation error is $\|\alpha_m(g)-g\|_{L_p(\mu)}=0$.

The result follows by Theorem 2.1 and the following lemma.

LEMMA 5.5. *If μ is absolutely continuous with respect to the Lebesgue measure, the positive λ, linear operators $\{\alpha_m\}_{m\in\mathbb{I}}$ are uniformly bounded in $L_p(\mathbb{R}^d,\mathbb{B},\mu)$, for $1\le p\le\infty$.*

PROOF. First notice that in $L_1(\mathbb{R}^d,\mathbb{B}^d,\mu)$, the norm of α_m satisfies,

$$\begin{aligned}\|\alpha_m\|_{L_1} &= \operatorname*{ess\,sup}_{x\in\mathbb{R}^d,\mu}\int|K_m(x,z)|\mu(dz)=\operatorname*{ess\,sup}_{x\in\mathbb{R}^d,\mu}\int\left|\sum_{A\in m}\frac{I_A(x)I_A(z)}{\mu(A)}\right|\mu(dz)\\ &=\operatorname*{ess\,sup}_{x\in\mathbb{R}^d,\mu}\sum_{A\in m}I_A(x)=\sup_{x\in\mathbb{R}^d}1=1.\end{aligned}$$

Also, in $L_\infty(\mathbb{R}^d,\mathbb{B}^d,\mu)$, noticing that $K_m(x,z)$ is continuous a.e. (since $\mu\ll\lambda$ the discontinuity points have measure null), and

$$\begin{aligned}\|\alpha_m\|_{L_\infty} &= \operatorname*{ess\,sup}_{z\in\mathbb{R}^d,\mu}\int|K_m(x,z)|\mu(dx)=\operatorname*{ess\,sup}_{z\in\mathbb{R}^d,\mu}\int\left|\sum_{A\in m}\frac{I_A(x)I_A(z)}{\mu(A)}\right|\mu(dx)\\ &=\operatorname*{ess\,sup}_{z\in\mathbb{R}^d,\mu}\sum_{A\in m}I_A(z)=\sup_{z\in\mathbb{R}^d}1.\end{aligned}$$

For $1 < p < \infty$, the L_p boundedness follows from the Riesz-Thorin interpolation theorem.

PROOF OF PROPOSITION 2.7. Assumption O.1 implies that A.1 holds. Note that the projection operators $\{\alpha_m\}_{m\in\mathbb{I}}$ have norm

$$\|\alpha_m\|_{L_2(\mu)} = \sup_{\|f\|_{L_2(\mu)}\leq 1} \left(\sum_{k=1}^{m} |\langle f, e_k\rangle_{L_2(\mu)}|^2\right)^{1/2} \leq \||\alpha|_m\|_{L_2(\mu)}.$$

We only need to check A.2 to A.4. Note that, since $e_{k_0}(x) = 1$, a.s. $[\mu]$, then $\forall m \geq k_0$,

$$\alpha_m(1)(x) = 1 \quad \text{a.s.} \quad [\mu],$$

because 1 belongs to the space *span* $(\{e_k\}_{k=1}^m)$, and its projection is just equal to itself. Assumptions A.3 and A.4 are a consequence of O.3 and O.4, respectively.

5.2 *Proofs of Section* 3

PFOOF OF THEOREM 3.1. As the space $(B, \|\cdot\|_B)$ is of type $\gamma \in [1, 2]$, there exists $c_\gamma > 0$ such that

$$\begin{aligned} E\left[\left\|\frac{1}{n}\sum_{i=1}^{n}(Z_{n,i} - E[Z_{n,i}])\right\|_B^\gamma\right] &\leq c_\gamma \cdot \frac{1}{n^\gamma}\sum_{i=1}^{n} E[\|Z_{n,i} - E[Z_{n,i}]\|_B^\gamma] \\ &\leq c_\gamma \cdot 2^{\gamma-1}\sum_{i=1}^{n}\frac{E[\|Z_{n,i}\|_B^\gamma] + \|E[Z_{n,i}]\|_B^\gamma}{n^\gamma}, \end{aligned}$$

by the triangular inequality and the C_r inequality. On the other hand, the Bochner integral satisfies $\|E[Z_{n,i}]\|_B^\gamma \leq E[\|Z_{n,i}\|_B]^\gamma \leq E[\|Z_{n,i}\|_B^\gamma]$ (see Araujo and Giné ((1980), Prop. 2.2) and apply Jensen's inequality). Thus,

$$E\left[\left\|\frac{1}{n}\sum_{i=1}^{n}(Z_{n,i} - E[Z_{n,i}])\right\|_B^\gamma\right] \leq c_\gamma 2^\gamma \frac{\sum_{i=1}^n E[\|Z_{n,i}\|_B^\gamma]}{n^\gamma} \to 0.$$

PROOF OF PROPOSITION 3.1. By Fubini's theorem,

$$\begin{aligned} E[\|Z_{n,i}\|_{L_p(\mu)}^p] &= E[\|K_{m_n}(x, X_i)\|_{L_p(\mu)}^p] = E\left[\int |K_{m_n}(x, X_i)|^p \mu(dx)\right] \\ &= \int E[|K_{m_n}(x, X_i)|^p]\mu(dx) = \int \left[\int |K_{m_n}(x, z)|^p \frac{dP}{d\mu}(z)\mu(dz)\right]\mu(dx) \\ &= \left\|\int |K_{m_n}(x, z)|^p \frac{dP}{d\mu}(z)\mu(dz)\right\|_{L_1(\mu)}. \end{aligned}$$

PROOF OF THEOREM 3.2. Notice that for each $x \in \mathbb{R}^d$ the family $\{K_{m_n}(x, X_i)\}$ is an $\mathbb{R}$-valued triangular array. Notice that $\mathbb{R}$ is a Hilbert space and, therefore a type-2 space. Applying Theorem 3.1, condition (1) implies that,

$$n^{-1}\mathbb{S}_n(x) := \frac{1}{n}\sum_{i=1}^{n}(K_{m_n}(x, X_i) - E[K_{m_n}(x, X_i)]) \xrightarrow{L_2} 0,$$

and therefore $E[|n^{-1}\mathbb{S}_n(x)|^p] \to 0$ for any $p \in [1,2]$, that is

$$\phi_n(x) := E[|n^{-1}\mathbb{S}_n(x)|^p] = E\left[\left|\frac{1}{n}\sum_{i=1}^{n}(K_{m_n}(x,X_i) - E[K_{m_n}(x,X_i)])\right|^p\right] \to 0.$$

This result holds for a.e. $x \in \mathbb{R}^d$, with respect to μ (by assumption 1). Notice also that condition (2) states that,

$$\int \sup_{n\geq 1} \phi_n(x)\mu(dx) < \infty.$$

Hence, by Fubini's theorem and Lebesgue's dominated convergence theorem,

$$E[\|n^{-1}\mathbb{S}_n(x)\|^p_{L_p(\mu)}] = \int E[|n^{-1}\mathbb{S}_n(x)|^p]\mu(dx) = \int \phi_n(x)\mu(dx) \to 0.$$

Then, Markov's inequality implies $\|n^{-1}\mathbb{S}_n(x)\|_{L_p(\mu)} \xrightarrow{\text{P}} 0$; that is, the weak L_p-convergence for the variation term.

PROOF OF THEOREM 3.3. It is well known that complete convergence implies almost sure convergence which implies convergence in probability. We will prove that convergence in probability implies complete convergence.

LEMMA 5.6. *Under conditions of Theorem* 3.3, *then, for all* $\delta > 0$,

$$P(|\|\widehat{f}_{m_n} - E[\widehat{f}_{m_n}]\|_{L_p} - E[\|\widehat{f}_{m_n} - E[\widehat{f}_{m_n}]\|_{L_p}]| \geq \delta) \leq \exp\{-n\lambda^2/4M^2\}.$$

PROOF. The result is a consequence of McDiarmid's (1989) inequality. Consider a real function $g(X_1,\dots,X_n)$ where X_i are independent real variables. If g satisfies for each $i_0 \in \{1,\dots,n\}$,

$$\sup_{X'_{i_0}\in B} |g(X_1,\dots,X_{i_0},\dots,X_n) - g(X_1,\dots,X'_{i_0},\dots,X_n)| \leq c_{i_0},$$

with probability one, by McDiarmid's inequality,

$$\Pr[|g(X_1,\dots,X_n) - E[g(X_1,\dots,X_n)]| > \lambda] \leq \exp\left\{\frac{-\lambda^2}{\sum_{i=1}^{n} c_i^2}\right\}.$$

Devroye (1991) and Devroye *et al.* ((1996*b*), p. 136) provide an introduction to McDiarmid's inequality.

Consider the function

$$g(X_1,\dots,X_n) = \left\|n^{-1}\sum_{i=1}^{n} K_{m_n}(x,X_i) - E[K_{m_n}(x,X)]\right\|_{L_p} = \|\widehat{f}_{m_n} - E[\widehat{f}_{m_n}]\|_{L_p}.$$

For any a, b belonging to a normed space, $|\|a\| - \|b\|| \leq \|a-b\|$ is satisfied. Thus, for each $i_0 \in \{1,\dots,n\}$,

$$\begin{aligned}&\sup_{X'_{i_0}\in\mathbb{R}^d} |g(X_1,\dots,X_{i_0},\dots,X_n) - g(X_1,\dots,X'_{i_0},\dots,X_n)| \\ &\quad \leq \frac{1}{n}\sup_{X'_{i_0}\in\mathbb{R}^d} \|K_{m_n}(x,X_{i_0}) - K_{m_n}(x,X'_{i_0})\|_{L_p} \overset{\text{a.e.}}{\leq} \frac{2M}{n}.\end{aligned}$$

Therefore,

$$\Pr[|g(X_1,\dots,X_n) - E[g(X_1,\dots,X_n)]| > \lambda] \le \exp\left\{\frac{-n\lambda^2}{4M^2}\right\},$$

applying McDiarmid's inequality.

The proof of the theorem is now immediate, considering the centered sums of the triangular array. Notice that under the boundedness condition, convergence in probability implies that

$$E[\|\widehat{f}_{m_n} - E[\widehat{f}_{m_n}]\|_{L_p}] = o(1),$$

by the dominated convergence theorem.

Therefore, $\forall\varepsilon \in (0,\delta)$ and $\delta > 0$ there exists an $n(\varepsilon) \in \mathbb{N}$ such that $\forall n > n(\varepsilon)$, $E[\|\widehat{f}_{m_n} - E[\widehat{f}_{m_n}]\|_{L_p}] \le \varepsilon$, and by the triangular inequality,

$$P(\|\widehat{f}_{m_n} - E[\widehat{f}_{m_n}]\|_{L_p} \ge \delta) \le P(|\|\widehat{f}_{m_n} - E[\widehat{f}_{m_n}]\|_{L_p} - E[\|\widehat{f}_{m_n} - E[\widehat{f}_{m_n}]\|_{L_p}]| \ge \delta - \varepsilon),$$

with $\delta - \varepsilon > 0$. By the previous lemma, the right-hand side of last expression is bounded by $\exp\{-n(\delta-\varepsilon)/4M^2\}$. Thus, complete convergence holds. The extension to the case of increasing constants $M_n > 0$ is trivial.

Acknowledgements

We thank Juan A. Cuesta, Gabor Lugosi and two anonymous referees for their helpful comments and suggestions, which have led to an improved version of this paper. This research has been supported by a Marie Curie Fellowship of the European Community programma IHP, under contract number HPMF-CT-2000-00449; and also by the Spanish Ministry of Technology (Dirección General de Enseñanza Superior, DGES), reference number BEC2001-1270.

References

Abou-Jaoude, S. (1976*a*). Sur une condition nécessaire et suffisante de L_1 convergence presque complete de l'estimateur de la partition fixe pour une density, *Comptes Rendus de l'Academie des Sciences de Paris*, *Serie A*, **283**, 1107–1110.

Abou-Jaoude, S. (1976*b*). Sur la convergence L_1 et L_∞ de l'estimateur de la partition aléatorie pour une densité, *Annales de l'Institut Henry Poincairé*, **12**, 299–317.

Abou-Jaoude, S. (1976*c*). Conditions nécessaries et suffisantes de convergence L_1 in probabilité de l'histogram pour une densité, *Annales de l'Institut Henry Poincairé*, **12**, 213–231.

Araujo, A. and Giné, E. (1980). *The Central Limit Theorem for Real and Banach Valued Random Variables*, John Wiley & Sons, New York.

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

Billingsley, P. (1986). *Probability and Measure*, 2nd ed., John Wiley & Sons, New York.

Bosq, D. (1969). Sur l'estimation d'une densité multivariée par une serie de fonctions orthogonales, *Comptes Rendus de l'Academie des Sciences de Paris*, **268**, 555–557.

Butzer, P. L. and Nessel, R. J. (1971). *Fourier Analysis and Approximation*, Vol. 1, Birkhäuser Verlag, Bassel and Stuttgart.

Čencov, N. N. (1962). Evaluation of an unknown density by orthogonal series, *Soviet Mathematics Doklady*, **3**, 1559–1562.

Cheney, E. W. (1982). *Introduction to Approximation Theory*, Chelsea Publishing Company, New York.

Chung, K. L. (1974). *A Course in Probability Theory*, 2nd ed., Academic Press, San Diego, California.

Davis, P. J. (1975). *Interpolation and Approximation*, Dover, New York.

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

DeVore, R. A. and Lorentz, G. G. (1993). *Constructive Approximation. Grundlehren der mathemastischen Wissenschaften*, **303**, Springer Verlag, Berlin.

Devroye, L. (1983). The equivalence of weak, strong and complete convergence in L_1 for kernel density estimates, *Annals of Statistics*, **11**, 896–904.

Devroye, L. (1987). A course in density estimation, *Progress in Probability and Statistics*, Birkhäuser, Boston.

Devroye, L. (1991). Exponential inequalities in nonparametric estimation, *Nonparametric Functional Estimation and Related Topics* (ed. G. Roussas), 31–44, Kluwer Academic Publishers, Dordrecht.

Devroye, L. and Györfi, L. (1983). Distribution-free exponential bound on the L_1 error of partitioning estimates of a regression function, *Proceedings of the Fourth Pannonian Symposium on Mathematical Statistics* (eds. F. Konecny, J. Mogyorodi and W. Wertz), Akadémiai Kiadó, Budapest, Hungary.

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

Devroye, L. and Györfi, L. (1985*b*). Distribution free exponential bound for the L_1 error of the partitioning estimates of a regression function, *Probability and Statistical Decision Theory, Proceedings of the Fourth Pannonian Symposium on Mathematica Statistics* (eds. F. Konecny, J. Mogyorodi and W. Wertz), 67–76, Reidel, Dordrecht.

Devroye, L. and Krzyżak, A. (1989). An equivalence theorem for L_1 convergence of nearest neighbor regression function estimates, *Annals of Statistics*, **22**, 1371–1385.

Devroye, L. and Lugosi, G. (2001). *Combinatorial Methods in Density Estimation*, Springer Verlag, New York.

Devroye, L. and Wagner, T. J. (1979). On the L_1 convergence of kernel density estimators, *Annals of Statistics*, **7**, 1136–1139.

Devroye, L. and Wagner, T. J. (1980*a*). Distribution free consistency results in nonparametric discrimination and regression function estimates, *Annals of Statistics*, **8**, 231–239.

Devroye, L. and Wagner, T. J. (1980*b*). On the L_1 convergence of kernel estimators of regression functions with application in discrimination, *Zeitschrift für Wahrsheinlichkeitstheorie und verwandte Gebiete*, **51**, 15–25.

Devroye, L. and Wagner, T. J. (1982). Nearest neighbor methods in discrimination, *Handbook of Statistics* (eds. P. Krishnaiah and L. Kanal), Vol. 2, 193–197, North Holland, Amsterdam.

Devroye, L., Györfi, L., Krzyżak, A. and Lugosi, G. (1996*a*). On the strong universal consistency of nearest neighbor regression function estimates, *Annals of Statistics*, **22**, 1371–1385.

Devroye, L., Györfi, L. and Lugosi, G. (1996*b*). A probabilistic theory of pattern recognition, *Applications of Mathematics, Stochastic Modelling and Applied Probability*, Springer Verlag, New York.

Dunford, N. and Schwartz, J. T. (1957). *Linear Operators. Part I. General Theory*, Wiley Classics Library Edition, 1988. John Wiley & Sons, New York.

Edgar, G. A. and Sucheston, L. (1992). Stopping times and directed processes, *Encyclopedia of Mathematics and Its Applications* (ed. G. C. Rota), **47**, Cambridge University Press, Cambridge.

Freedman, D. and Diaconis, P. (1981). On the histogram as a density estimator: L_2 theory, *Zeitschrift fur Wahrsheinlichkeitstheorie und verwandte Gebiete*, **58**, 139–157.

Graunt, J. (1662). *Natural and Political Observations Made upon the Bills of Mortality*, Martyn, London.

Györfi, L. (1981). The rate of convergence of $K_n - NN$ regression estimation and classification, *IEEE Transactions on Information Theory*, **IT-27**, 500–509.

Györfi, L., Kohler, M., Krzyżak, A. and Walk, H. (2002). *A Distribution-free Theory of Nonparametric Regression*, Springer Verlag, New York.

Hoffmann-Jørgensen, J. (1974). Sums of independent Banach space valued random variables, *Studia Mathematica*, **52**, 159–186.

Hoffmann-Jørgensen, J. (1976). *Probability in Banach Spaces*, Lecture Notes in Mathematics, **598**, Springer Verlag, New York.

Hong, D. H., Ordoñez-Cabrera, M., Sung, S. H. and Volodin, A. I. (2000). On the weak law for randomly indexed partial sums for arrays of random elements in martingale type p Banach spaces, *Statistics and Probability Letters*, **46**, 177–185.

Hu, T. C., Ordoñez-Cabrera, M. and Volodin, A. I. (2001). Convergence of randomly weighted sums of B-space valued random elements and uniform integrability concerning random weights, *Statistics and Probability Letters*, **51**, 155–164.

Hu, T. H. and Chang, H. H. (1997). Complete convergence and the law of large numbers for arrays of random elements, *Proceedings*, *2nd World Congress of Nonlinear Analysis*, Elsevier Science, U.K. (*Nonlinear Analysis. Methods and Applications*, **30**, 4257–4266).

Kantorovich, L. V. and Akilov, G. P. (1982). *Functional Analysis*, 2nd ed., Pergamon Press, Oxford.

Kreyszig, E. (1978). *Introductory Functional Analysis with Applications*, Wiley Classics Library, John Wiley & Sons, New York.

Ledoux, M. and Talagrand, M. (1991). *Probability in Banach Spaces*: *Isoperimetry and Processes*, Springer Verlag, New York.

Linde, W. (1986). *Probability in Banach Spaces-Stable and Infinitely Divisible Distributions*, John Wiley & Sons, New York.

Lugosi, G. and Nobel, A. (1996). Consistency of data driven histogram methods for density estimation and classification, *Annals of Statistics*, **24**, 687–706.

Lugosi, G. and Zeger, K. (1995). Nonparametric estimation via empirical risk minimization, *IEEE Transactions on Information Theory*, **41**, 677–678.

McDiarmid, C. (1989). On the method of bounded differences, *Surveys in Combinatorics*, 148–188, Cambridge University Press, Cambridge.

Ordoñez-Cabrera, M. (1994). Convergence of weighted sums of random variables and uniform integrability concerning the weights, *Collectanea Mathematica*, **45**, 121–132.

Patterson, R. F. and Taylor, R. L. (1997). Strong Laws of Large Numbers for negatively dependent random elements, *Proceedings*, *2nd World Congress of Nonlinear Analysis*, Elsevier Science, U.K. (*Nonlinear Analysis. Methods and Applications*, **30**, 4229–4235).

Pisier, G. (1986). *Probabilistic Methods in the Geometry of Banach Spaces*, Lecture Notes in Mathematics, **1206**, 167–241, Springer, Berlin.

Pisier, G. (1989). *The Volume of Convex Bodies and Banach Space Geometry*, Cambridge University Press, Cambridge.

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

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

Révesz, P. (1971). Testing of density functions, *Periodica Mathematica Hungarica*, **1**, 35–44.

Révesz, P. (1972). On empirical density function, *Periodica Mathematica Hungarica*, **2**, 85–110.

Révesz, P. (1973). A strong law of the empirical density function, *Transcations of the 6th Prague Conference on Information Theory*, 469–472.

Révesz, P. (1974). On empirical density function, *Probability and Statistical Methods—Summer School*, Bulgarian Academy of Science, Varna, Bulgaria.

Rudin, W. (1966). *Real and Complex Analysis*, 2nd ed., McGraw Hill, New York.

Schwartz, L. (1981). *Geometry and Probability in Banach Spaces*, Lecture Notes in Mathematics, **852**, Springer Verlag, New York.

Scott, D. W. (1979). On optimal data based histograms, *Biometrica*, **66**, 605–610.

Scott, D. W. (1992). *Multivariate Density Estimation. Theory*, *Practice*, *and Visualization*, John Wiley & Sons, New York.

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

Silverman, A. N. (1978). Weak and strong uniform consistency of the kernel estimate of a density and its derivatives, *Annals of Statistics*, **6**, 177–184.

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

Stone, C. J. (1977). Consistent nonparametric regression, *Annals of Statistics*, **5**, 595–645.

Stute, W. (1982). A law of the logarithm for kernel density estimators, *Annals of Probability*, **10**, 414–422.

Taylor, R. L. and Hu, T. H. (1987). Strong Laws of Large Numbers for arrays of rowise independents random elements, *International Journal of Mathematics and Mathematical Sciences*, **10**, 804–814.

Terrell, G. R. (1984). Efficiency of nonparametric density estimators, Tech. Report, Department of Mathematical Sciences, Rice University, Houston, Texas.

Terrell, G. R. and Scott, D. W. (1992). Variable kernel density estimation, *Annals of Statistics*, **20**, 1236–1265.

Tukey, J. W. (1977). *Exploratory Data Analysis*, Addison-Wesley, Reading, Massachusetts.

Vakhania, N. N. (1981). *Probability Distributions on Linear Spaces*, North Holland, New York.

Vakhania, N. N., Tarieladze, V. I. and Chobanyan, S. A. (1987). Probability distributions on Banach spaces, *Mathematics and Its Applications* (Soviet Series), Reidel Publishing Company, Dordrecht (translated from Russian, 1985, Nauka, Moscow).

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes with Applications to Statistics*, Springer Verlag, New York.

Vapnik, V. (1982). *Estimation of Dependencies Based on Empirical Data*, Springer Verlag, New York.

Vidal-Sanz, J. M. (1999). *Universal Consistency of Delta Estimators: An Approximation Theory Based Approach*, Ph.D. Dissertation, Universidad Carlos III de Madrid, Spain (in Spanish).

Walter, G. and Blum, J. R. (1979). Probability density estimation using delta sequences, *Annals of Statistics*, **7**, 328–340.

Watson, G. S. and Leadbetter, M. R. (1963). On the estimation of probability density I, *Annals of Statistics*, **34**, 480–491.

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

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

Whittle, P. (1958). On the smoothing of probability density functions, *Journal of the Royal Statistical Society Series B*, **20**, 334–343.

Winter, B. B. (1973). Strong uniform consistency of integrals of density estimation, *The Canadian Journal of Statistics*, **1**, 247–253.

Winter, B. B. (1975). Rate of strong consistency of two nonparametric density estimators, *Annals of Statistics*, **3**, 759–766.

Woyczyński, W. A. (1978). Geometry and martingales in Banach spaces—Part II: Independent increments, *Probability on Banach Spaces* (ed. J. Kuebs), *Advances in Probability and Related Topics*, Vol. 4, 267–519, Marcel Dekker, New York.

Xia, Dao-Xing (1972). *Measure and Integration on Infinite-dimensional Spaces: Abstract Harmonic Analysis*, Academic Press, New York.