Working Paper 99-37
Statistics and Econometrics Series 10
May 1999

Departamento de Estadística y Econometría
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Spain)
Fax (34-91) 624-9849

ON UNIVERSAL UNBIASEDNESS OF DELTA ESTIMATORS.

Miguel A. Delgado and José M. Vidal*

Abstract

This paper considers delta estimators of a Radon-Nicodym derivative of a probability function with respect to a measure. Sufficient conditions for asymptotic unbiasedness and global rates of convergence, which can be improved by imposing differentiability conditions on the estimated curves, are provided. A bias reduction technique is proposed, and the application of the results to regression estimation is discussed. The sufficient conditions for asymptotic unbiasedness are checked for some broad classes of nonparametric estimators.

Keywords: Bias of delta estimators; Universal unbiasedness; Approximation of the bias; Global rates of convergence for the bias; Bias reduction techniques.

*Delgado, Departamento de Estadística y Econometría, Universidad Carlos III de Madrid. C/ Madrid, 126 28903 Getafe -Madrid-. Spain. Ph: 34-91-6249804. Fax: 34-91-624.98.49, e-mail: delgado@est-econ; Vidal, Departamento de Estadística y Econometría, Universidad Carlos III de Madrid. Research funded by Spanish "Dirección General de Enseñanza Superior" (DGES), reference number PB95-0292.

1. INTRODUCTION

Let P be a probability function in $(\mathbb{R}^d, \mathbb{B}^d)$ absolutely continuous with respect to the σ-finite measure μ. Let $f = dP/d\mu$ 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, it is considered the Lebesgue measure λ and then, $f = dP/d\lambda$ is the corresponding probability density function (pdf). Given a random sample $\{X_i, 1 = 1, ..., n\}$ from P, a delta estimator of f is defined as

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

where $m_n = m(n)$ is known as smoothing sequence, and it is assumed that m_n diverges as $n \to \infty$. The sequence $\{m_n\}_{n\in\mathbb{N}}$ is not necessarily a sequence of numbers, it could be a sequence of matrices, partitions, functions, etc. The sequence $\{K_{m_n}\}_{n\in\mathbb{N}}$ is known as generalized kernel sequence.

This class of estimators was introduced by Whittle (1958), encompassing most of the existing nonparametric estimators. Terrell (1984) and Terell and Scott (1992) show 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 the case of pdf estimation $\widehat{f}_n(x)$ is pointwise asymptotically unbiased if

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

where δ is the Dirac delta generalized function with 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) provide sufficient condition for global consistency in norm $L_p(\lambda)$ and pointwise consistency, assuming smoothness conditions on f. Winter (1973, 1975) studies uniform consistency and the consistency of the corresponding smooth distribution function estimator. Watson and Leadbetter (1964) establish asymptotic normality. Basawa and Prakasa Rao (1980, Chap. 11) provide results for dependent observations. In this literature unbiasedness is achieved under restrictive smoothness conditions on the pdf f. Universal asymptotic unbiasedness results and global rates of convergence for the bias have not been obtained yet.

One of the main objectives of this paper is to provide fairly primitive conditions on K_{m_n} which are sufficient for the universally asymptotic unbiasedness of $\widehat{f}_n$. The expected value of $\widehat{f}_n$ is given by

$$\alpha_{m_n}(f;x) = \int K_{m_n}(x;z) f(z) \ \mu(dz).$$

which is a sequence of linear operators in $L_p(\mathbb{R}^d, \mathbb{B}^d, \mu)$. We say that the delta estimator is asymptotically unbiased (in global sense) when

$$\lim_{n\to\infty} \|\alpha_{m_n}(f;x) - f(x)\|_{L_p(\mu)} = 0. \tag{1}$$

Notice that the smoothing sequence $\{m_n\}_{n\in\mathbb{N}}$ is a divergent sequence in a directed set[1] $\mathbb{I}$. The elements of such a sequence can be positive definite matrices ordered by decreasing norm, in the usual kernel estimator of a multivariate density; the order of a polynomial, in the orthogonal series estimators; or measurable partitions, in the histogram. Thus $\{\alpha_m(f;x)\}_{m\in\mathbb{I}}$ is a net[2] of curves in $L_p(\mathbb{R}^d, \mathbb{B}^d, \mu)$, which allows to define universally asymptotic unbiasedness in a very general framework.

Definition 1 *We say that the delta estimator $\widehat{f}_n$ is universally asymptotically unbiased in $L_p(\mathbb{R}^d, \mathbb{B}^d, \mu)$, with $1 \leq p < \infty$, iff the net of linear operators $\{\alpha_m\}_{m\in\mathbb{I}}$, with*

$$\begin{array}{rccl} \alpha_m : & L_p(\mathbb{R}^d, \mathbb{B}^d, \mu) & \longrightarrow & L_p(\mathbb{R}^d, \mathbb{B}^d, \mu) \\ & f & \longmapsto & \alpha_m(f) = \alpha_m(f;x) = \int K_m(x,z) f(z) \ \mu(dz); \end{array}$$

is such that $\forall f \in L_p(\mathbb{R}^d, \mathbb{B}^d, \mu)$,

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

This is the same as saying that the net of linear operators $\{\alpha_m\}_{m\in\mathbb{I}}$ is a linear approximator of the identity or a linear approximate identity[3].

Notice that when $\widehat{f}_n$ is asymptotically unbiased, according to definition 1, (1) holds for any sequence $\{m_n\}_{n\in\mathbb{N}}$ which is an increasing sequence that diverges in $\mathbb{I}$, for all probability functions P absolutely continuous with respect to μ, with $f = dP/d\mu \in L_p(\mathbb{R}^d, \mathbb{B}^d, \mu)$.

[1] A directed set $\mathbb{I}$, is a non empty set endowed with a partial preorder $\leq$, such that if $m_1, m_2 \in \mathbb{I}$, then, there exists an $m_3 \in \mathbb{I}$ such that $m_1 \leq m_3$ and $m_2 \leq m_3$.

[2] A net $\{a_m\}_{m\in\mathbb{I}}$ in a Banach space $(B, \|\cdot\|_B)$, is such that $a_m = a(m)$ with $a : \mathbb{I} \to B$, where $\mathbb{I}$ is a directed set. So we say that $\lim_{m\in\mathbb{I}} \|a_m - a\|_B = 0$, $a \in B$ iff $\forall \varepsilon > 0$, $\exists m(\varepsilon) \in \mathbb{I}$ such that $\|a_m - a\|_B < \varepsilon$ for al $m \geq m(\varepsilon)$. See e.g. Edgar and Sucheston (1992, pp 4). In our case $B = L_p(\mathbb{R}^d, \mathbb{B}^d, \mu)$, $a_m = \alpha_m(f;x)$ and $a = f(x)$.

[3] Let $(B; \|\cdot\|_B)$ a Banach space. The net $\{a_m\}_{m\in\mathbb{I}}$ of bounded linear operators $a_m : B \to B$ is a linear approximator of the identity if

$$\lim_{m\in\mathbb{I}} \|a_m(f) - f\|_B = 0, \quad \forall f \in B.$$

See e.g. Davis (1975, pp 346).

The rest of the paper is organized as follows. In section 2 we provide sufficient conditions on the kernel function $\{K_m(x,z)\}_{m\in\mathbb{I}}$, for universally asymptotically unbiasedness. Section 3 provides a rate of convergence for the bias, which can be improved under differentiability conditions on f. Section 4 presents a bias reduction technique. In section 5 we check the universal asymptotic unbiasedness sufficient conditions for for some broad classes of nonparametric estimators. In section 6 we discuss the bias of regression delta estimators. Proofs can be found in section 7.

2. SUFFICIENT CONDITIONS FOR UNIVERSAL ASYMPTOTICALLY UNBIASEDNESS

Define the net of majorized operators of $\{\alpha_m\}_{m\in\mathbb{I}}$

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

The following theorem provides conditions on the generalized kernel net $\{K_m(x,z)\}_{m\in\mathbb{I}}$ which are sufficient to guarantee that the net $\{\alpha_m\}_{m\in\mathbb{I}}$ is a linear approximator of the identity and, therefore, the delta estimator is universally asymptotically unbiased. We say that $\{\lfloor\alpha\rfloor_m\}_{m\in\mathbb{I}}$ is uniformly bounded[4] in $L_p(\mathbb{R}^d,\mathbb{B}^d,\mu)$ if

$$\sup_{m\in\mathbb{I}} \|\lfloor\alpha\rfloor_m\|_{L_p(\mu)} \overset{def.}{=} \sup_{m\in\mathbb{I}}\left\{\sup_{\|f\|_{L_p(\mu)}\leq 1} \|\lfloor\alpha\rfloor_m(f;x)\|_{L_p(\mu)}\right\} < \infty. \tag{2}$$

Theorem 1 *Assume that*

A.1. $\{\lfloor\alpha\rfloor_m\}_{m\in\mathbb{I}}$ *is uniformly bounded in* $L_p(\mathbb{R}^d,\mathbb{B}^d,\mu)$, $1\leq p<\infty$.

A.2.

$$\lim_{m\in\mathbb{I}} \|\alpha_m(1;x)-1\|_{L_p(\mu)} = \lim_{m\in\mathbb{I}} \left\|\int K_m(x,z)\,\mu(dz)-1\right\|_{L_p(\mu)} = 0.$$

A.3. *For all compact set* $C\subset\mathbb{R}^d$, *the measure* $\mu(C)<\infty$.

[4] Let $(B_1,\|\cdot\|_{B_1})$, $(B_2,\|\cdot\|_{B_2})$ be two Banach spaces. We say that a linear operator $a: B_1 \to B_1$ is bounded (equivalently continuous), if

$$\|a\|_{B_1,B_2} \overset{def.}{=} \sup_{\|f\|_{B_1}\leq 1} \|a(f)\|_{B_2} < \infty.$$

A net $\{a_m\}_{m\in\mathbf{I}}$ is uniformly bounded iff

$$\sup_{m\in\mathbf{I}} \|a_m\|_{B_1,B_2} = \sup_{m\in\mathbf{I}}\left\{\sup_{\|f\|_{B_1}\leq 1} \|a(f)\|_{B_2}\right\} < \infty.$$

See e.g. Kantorovich and Akilov (1982).

A.4. *For all* $\delta > 0$, *and all compact set* $C \subset \mathbb{R}^d$,

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

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

Then the net sequence $\{\alpha_m\}_{m \in \mathbb{I}}$ *is a linear approximator of the identity in* $L_p(\mathbb{R}^d, \mathbb{B}^d, \mu)$.

The following remarks are in order

Remark 1 *Assumption A.1. establishes that the net sequence* $\{\lfloor \alpha \rfloor_m\}_{m \in \mathbb{I}}$ *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, then the left hand side of (2) is equal to*

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

If $p = 2$ *the left hand side of (2) is bounded by*

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

See DeVore and Lorentz (1993, pp 30-34) and Dundford and Schwartz (1956) for a discussion of these results.

Remark 2 *A sufficient condition, but not necessary, for A.2. is that there exist an* $m_0 \in \mathbb{I}$ *such that* $\forall m \geq m_0$,

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

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

$$\begin{aligned} \mu\left(\left\{x \in \mathbb{R}^d : \lim_{m \in \mathbb{I}} |\alpha_m(1;x) - 1| > 0\right\}\right) &= 0 \\ \sup_{m \in \mathbb{I}} |\alpha_m(1;x)| &\in L_p(\mathbb{R}^d, \mathbb{B}^d, \mu), \end{aligned}$$

Observe that if μ *is a finite measure, those conditions implies A.2. (see e.g. Chung (1974, pp 100) and Billingsley (1986, pp 220)).*

Remark 3 *If* μ *is a finite measure, condition A.3. holds. The Lebesgue's measure also satisfies A.3..*

Remark 4 *A sufficient condition for A.4. is that for all* $\delta > 0$,

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

Assumption A.4. says that when m *increases the support of* $|K_m(x,z)|$ *concentrates on* $\{(x,z) : x = z\}$ *and, perhaps, in other points of null measure.*

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

Proposition 1 *The assumption:*

A.4'. *For some* $s \geq 1$

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

is a sufficient condition for A.4.

Proposition 2 *The assumptions:*

1. $\int \|z - x\|^s \, |K_m(x, z)| \, \mu(dz) \to 0, \quad a.s. [\mu],$

2. $\int \|z - x\|^s \, |K_m(x, z)| \, \mu(dz) < |T(x)|, \quad T \in L_p(\mathbb{R}^d, \mathbb{B}^d, \mu),$

are sufficient conditions for A.4'.

Weaker sufficient conditions in propositions 1 and 2 can be obtained substituting μ by μ_C, for every compact set C.

In order to obtain rates of convergence for the bias of delta pdf estimators, we need to assume differentiability of the density function.

3. GLOBAL RATES OF CONVERGENCE FOR THE BIAS

When f belongs to the Sobolev's[5] space $W_p(\mathbb{R}^d, \mathbb{B}^d, \lambda)$, it is fairly straightforward to obtain global rates of convergence for the bias. The next proposition provides a bounding condition, which is useful in order to obtain rates of convergence.

Proposition 3 *If* $\{\alpha_m\}_{m \in \mathbf{I}}$ *is a linear approximator of the identity in* $L_p(\mathbb{R}^d, \mathbb{B}^d, \lambda)$ *with* $1 \leq p < \infty$, *and* $f \in W_p(\mathbb{R}^d, \mathbb{B}^d, \lambda)$, *then*

$$\|\alpha_m(f; x) - f(x)\|_{L_p(\lambda)} \leq \|f(x)(1 - \alpha_m(\mathbf{1}; x))\|_{L_p(\lambda)} + O(\zeta(m)),$$

where

$$\zeta(m) = \|\lfloor\alpha\rfloor_m(\|x - z\|; x)\|_{L_p(\lambda)} = \left\| \int |K_m(x, z)| \, \|x - z\| \, \lambda(dz) \right\|_{L_p(\lambda)}.$$

[5] For an introduction to Sobolev spaces see, e.g. Adams (1978) and Maz'ja (1985).

Remark 5 *This result is useful when* $\lim_{m\in\mathbb{I}} \zeta(m) = 0$ *holds, e.g. under A.4'. with* $s = 1$. *In some cases this property may not be satisfied. However, we can use the proposition 3, substituting the measure* λ *by the measure* λ_C, *where* $C \subset \mathbb{R}^d$ *is a compact set and* λ_C *is the Lebesgue's measure restricted to* C. *Also* $\zeta(m)$ *must be substituted by* $\zeta_C(m)$, *such that,*

$$\lim_{m\in\mathbb{I}} \zeta_C(m) \stackrel{def}{=} \lim_{m\in\mathbb{I}} \left\| \int |K_m(x,z)| \, \|x-z\| \, \lambda_C(dz) \right\|_{L_p(\lambda_C)} = 0.$$

Note that $\|\alpha_m(f;x) - f(x)\|_{L_p(\lambda)} = O(\zeta(m))$ when,

$$\|f(x)(1-\alpha_m(\mathbf{1};x))\|_{L_p(\lambda)} = o(\zeta(m)), \qquad (3)$$

which may be achieved ensuring that $\exists m_0 \in \mathbb{I}$ such that $\forall m \geq m_0$,

$$\alpha_m(\mathbf{1};x) = 1 \quad a.s.[\lambda],$$

this condition is satisfied when $\{\alpha_m\}_{m\in\mathbb{I}}$ is a net of normalized operators.

Define the normalized generalized kernel,

$$\widetilde{K}_m(x,z) = \frac{K_m(x,z)}{\int K_m(x,z)\,\lambda(dz)},$$

and the corresponding normalized operator,

$$\widetilde{\alpha}_m(f;x) = \frac{\alpha_m(f;x)}{\alpha_m(1;x)} \cdot I_{\{|\alpha_m(1;x)|>0\}}(x) = \int \widetilde{K}_m(x,z) f(z)\,\lambda(dz).$$

The next proposition shows that (3) holds for normalized operators.

Proposition 4 *If* $\{\alpha_m\}_{m\in\mathbb{I}}$ *is net such that,*

A.5.

$$\inf_{m\in\mathbb{I}} |\alpha_m(\mathbf{1};x)| \geq c > 0 \quad a.s.[\lambda],$$

then,

$$\widetilde{\alpha}_m(\mathbf{1};x) = 1 \quad a.s.[\lambda],$$

and $\forall f \in L_p(\mathbb{R}^d, \mathbb{B}^d, \lambda)$,

$$\|\widetilde{\alpha}_m(f;x) - f(x)\|_{L_p(\lambda)} \leq c^{-1} \cdot \left\| \int K_m(x,z)(f(z)-f(x))\,\lambda(dz) \right\|_{L_p(\lambda)}.$$

Thus, when A.5. holds, the resulting normalized operators satisfy condition A.2.. Notice also, that Proposition 3 establishes that $\{\widetilde{\alpha}_m\}_{m\in\mathbb{I}}$ is a linear approximator of the identity when,

$$\lim_{m\in\mathbb{I}} \left\| \int K_m(x,z)(f(z)-f(x))\,\lambda(dz) \right\|_{L_p(\lambda)} = 0, \quad \forall f \in L_p(\mathbb{R}^d, \mathbb{B}^d, \lambda),$$

which follows when $\{\alpha_m\}_{m\in\mathbb{I}}$ satisfies, despite of A.5., conditions A.1, A.3. and A.4. in Theorem 1.

The next corollary states the rate of convergence for the standardized operators $\{\widetilde{\alpha}_m\}_{m\in\mathbb{I}}$.

Corollary 1 *Suppose A.5. holds. If $\{\alpha_m\}_{m\in\mathbb{I}}$ is a linear approximator of the identity in $L_p(\mathbb{R}^d,\mathbb{B}^d,\lambda)$ with $1\le p<\infty$, and $f\in W_p(\mathbb{R}^d,\mathbb{B}^d,\lambda)$, then,*

$$\|\widetilde{\alpha}_m(f;x)-f(x)\|_{L_p(\lambda)}=O(\zeta(m)).$$

Therefore, pdf delta estimators of the form,

$$\widetilde{f}_n(x)=\frac{1}{n}\sum_{i=1}^{n}\widetilde{K}_{m_n}(x,X_i),$$

are universally unbiased in $L_p(\mathbb{R}^d,\mathbb{B}^d,\lambda)$, with a bias $O(\zeta(m))$ for differentiable density functions.

In general, the rate of convergence is not faster when f belongs to the Sobolev's spaces $W_p^s(\mathbb{R}^d,\mathbb{B}^d,\lambda)$, with higher derivatives of order $s>1$. But, under certain moment conditions on the linear operator, it is possible to obtain rates of convergence $o(\zeta(m))$ imposing the nullity of some moments for the linear approximator $\{\alpha_m\}_{m\in\mathbb{I}}$, as stated in the following theorem.

Theorem 2 *Let $\{\alpha_m\}_{m\in\mathbb{I}}$ be a linear approximator of the identity in $L_p(\mathbb{R}^d,\mathbb{B}^d,\lambda)$ with $1\le p<\infty$. Assume that there exist an $m_0\in\mathbb{I}$ such that for all $m\ge m_0$,*

$$\begin{aligned}\alpha_m(\mathbf{1};x)&=1\quad a.s.[\lambda],\\ \alpha_m((z-x)^{\nu};x)&=0\quad a.s.[\lambda],\end{aligned}\tag{4}$$

for all $\nu\in\mathbb{N}^d$ such that $0\le\|\nu\|_1<\tau$, $\tau\in\mathbb{N}$, with $(z-x)^{\nu}=\prod_{j=1}^{d}(z_j-x_j)^{\nu_j}$, $\nu!=\prod_{j=1}^{d}\nu_j!$. Then:

(i) *If $f\in W_p^s(\mathbb{R}^d,\mathbb{B}^d,\lambda)$ with $s\in\mathbb{N}$, $s\ge\tau$, then,*

$$\|\alpha_m(f;x)-f(x)\|_{L_p(\lambda)}=O(\zeta_\tau(m)),$$

where,

$$\zeta_\tau(m)=\|\lfloor\alpha\rfloor_m(\|z-x\|^{\tau})\|_{L_p(\lambda)}.$$

(ii) *If $f\in W_p^s(\mathbb{R}^d,\mathbb{B}^d,\lambda)$ with $s\in\mathbb{N}$, $s\ge\tau$, then,*

$$|\alpha_m(f;x)-f(x)|=O(\zeta_\tau(m;x))\quad a.s.[\lambda],$$

where,

$$\zeta_\tau(m;x)=|\lfloor\alpha\rfloor_m(\|z-x\|^{\tau})|.$$

(iii) *If* $f \in C_c^s(\mathbb{R}^d)$ *with* $s \in \mathbb{N}$, $s \geq \tau$, *then,*

$$\|\alpha_m(f;x) - f(x)\|_{L_\infty(\lambda)} = O(\zeta_\tau^\infty(m)),$$

where,

$$\zeta_\tau^\infty(m) = \|\lfloor\alpha\rfloor_m(\|z-x\|^\tau)\|_{L_\infty(\lambda)}.$$

So under the conditions of the above propositions, it is possible to obtain faster rates of convergence, since $O(\zeta_\tau(m)) = o(\zeta(m))$ for all $\tau > 1$, with $\zeta(m) = \zeta_1(m)$. This fact is formally stated in the following lemma.

Lemma 1 *Define for all* $\tau \geq 1$,

$$\zeta_\tau(m) = \|\lfloor\alpha\rfloor_m(\|z-x\|^\tau)\|_{L_p(\lambda)}.$$

and $\zeta(m) = \zeta_1(m)$.

- *If* $\lim_{m\in\mathbb{I}} \zeta(m) = 0$, *then,*

$$\lim_{m\in\mathbb{I}} \zeta_\tau(m) = 0, \quad \forall\tau > 1.$$

- *If* $1 \leq \tau_1, \tau_2 < \infty$, *then,*

$$\zeta_{\tau_2}(m) = o(\zeta_{\tau_1}(m)).$$

The same results follow for the rates $\zeta_\tau(m;x)$, $\zeta_\tau^\infty(m)$ *in (ii) (iii) of Theorem 2.*

A leading example of higher order delta estimators are the higher order kernels, studied by Singht (1979), Gasser and Müller (1984), Gasser et al (1985), Devroye (1987), Hall and Marron (1987) and Berlinet (1991) among others. These bias reduction techniques have been proven very useful in semiparametric estimation problems, in order to make compatible the convergence of the variance and bias terms in statistics which are weighted averages of nonparametric estimators, see e.g. Robinson (1988) and Powell et al (1989); also Delgado and Gonzalez-Manteiga (1998) for testing restrictions on nonparametric curves.

In next section we propose a bias reduction technique for obtaining higher order delta estimators based on local polynomials.

4. A BIAS REDUCTION TECHNIQUE.

Let $\{\theta_m\}_{m\in\mathbb{I}}$ a linear approximator of the identity in $L_p(\mathbb{R}^d, \mathbb{B}^d, \lambda)$, with $1 \leq p < \infty$, given by

$$\theta_m(f;x) = \int G_m(x,z) f(z) \lambda(dz),$$

with $\{\lfloor\theta\rfloor_m\}_{m\in\mathbb{I}}$ uniformly bounded. Assuming that $f \in W_p^s\left(\mathbb{R}^d, \mathbb{B}^d, \lambda\right)$, the objective is to obtain a new higher order bias delta estimator, whose corresponding net of expected values $\{\alpha_m\}_{m\in\mathbb{I}}$ achieve rates of convergence $O\left(\zeta_\tau\left(m\right)\right)$. De net $\{\alpha_m\}_{m\in\mathbb{I}}$ has generalized kernels of the form

$$K_m\left(x,z\right) = \left[1 - \sum_{l=1}^{\tau-1} \sum_{\|\gamma\|_1=l} \beta_\gamma\left(x\right)\left(z-x\right)^\gamma\right] G_m\left(x,z\right),$$

with $1 < \tau < s$. The polynomial coefficients β_γ are locally constant for each point x. These coefficients are obtained by means of que moment conditions

$$\alpha_m\left(\left(z-x\right)^\nu ; x\right) = \int \left(z-x\right)^\nu K_m\left(x,z\right)\lambda\left(dz\right) = 0,$$

with $1 \leq \|\nu\|_1 \leq \tau - 1$. So $K_m\left(x,z\right)$ is a higher order generalized kernel. The coefficients $\left\{\beta_\gamma\left(x\right)\right\}$ are deterministics. The pdf estimator is,

$$\widehat{f}_n\left(x\right) = \frac{1}{n}\sum_{i=1}^{n}\left[1 - \sum_{l=1}^{\tau-1} \sum_{\|\gamma\|_1=l} \beta_\gamma\left(x\right)\left(X_i - x\right)^\gamma\right] G_m\left(x, X_i\right).$$

The computation of the coefficient $\beta_\gamma\left(x\right)$ is fairly easy. For each x, $\beta_\gamma\left(x\right)$ is obtained by solving the linear system

$$\int \left(z-x\right)^\nu \left\{\left[1 - \sum_{l=1}^{\tau-1} \sum_{\|\gamma\|_1=l} \beta_\gamma\left(x\right)\left(z-x\right)^\gamma\right] G_m\left(x,z\right)\right\} \lambda\left(dz\right) = 0, \tag{5}$$

with $1 \leq \|\nu\|_1 \leq \tau - 1$. Notice that (5) can be written equivalently, as,

$$\int \left(z-x\right)^\nu G_m\left(x,z\right)\lambda\left(dz\right) = \sum_{l=1}^{\tau-1} \sum_{\|\gamma\|_1=l} \beta_\gamma\left(x\right) \int \left(z-x\right)^{\gamma+\nu} G_m\left(x,z\right)\lambda\left(dz\right),$$

or,

$$\theta_m\left(\prod_{j=1}^{d}\left(z_j - x_j\right)^{\nu_j} ; x\right) = \sum_{l=1}^{\tau-1} \sum_{\|\gamma\|_1=l} \beta_\gamma\left(x\right)\theta_m\left(\prod_{j=1}^{d}\left(z_j - x_j\right)^{\gamma_j+\nu_j} ; x\right).$$

In order to guarantee a solution for the system, we need that $\forall m \in \mathbb{I}$, the matrix

$$S\left(x\right) = \left(\left\{\theta_m\left(\prod_{j=1}^{d}\left(z_j - x_j\right)^{\gamma_j+\nu_j} ; x\right)\right\}_{1\leq\|\gamma_j\|_1\leq\tau-1, 1\leq\|\nu\|_1\leq\tau-1}\right)$$

must be non singular in almost every point $x \in \mathbb{R}^d$. In principle, this is the only restriction that has to be taken into account for choosing τ. It may be the case that such solution does not exist.

5. SOME EXAMPLES

Next we check conditions A.1.-A.4. for some broad classes of nonparametric density estimators in the next examples.

5.1. SINGULAR INTEGRAL ESTIMATORS

A relevant case are the singular integral kernels,

$$K_m(x,z) = K_m(x-z), \tag{6}$$

with correponding linear approximator

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

The pdf singular integral estimators singular is

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

The global unbiasedness of this estimators has been considered by Devroye and Györfi (1985, chap 12, sec. 8), for the measure λ restricted to a finite interval. They encompass relevant families of nonparametric estimators like:

- Kernels in $L_p\left(\mathbb{R}^d, \mathbb{B}^d, \lambda\right)$, that take

  $$K_H(u) = \frac{1}{\det(H)} \int \mathbf{K}\left(H^{-1}u\right)$$

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

- Fourier series estimators in $L_p([-\pi,\pi])$, $1 < p < \infty$, with

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

  $K_m(u)$ is known as Dirichlet's kernel. If $p = 1$ this is not uniformly bounded, but we use:

- Fejer series estimators in $L_1([-\pi,\pi])$,

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

There are many other examples, for a review see Butzer and Nessel (1971), Devroye and Györfi (1985, Chap. 12, sec. 8). The next proposition provides sufficient conditions which satisfy theorem 1 for these estimators:

Proposition 5 *Assume that:*

S.1. $\{K_m\}_{m\in\mathbb{I}} \subset L_1\left(\mathbb{R}^d, \mathbb{B}^d, \lambda\right)$.

S.2. $\exists m_0 \in \mathbb{I}$ *such that* $\int K_m(u)\,du = 1$, $\forall m \geq m_0$.

S.3. $\lim_{m\in\mathbb{I}} \int \|u\|\,|K_m(u)|\,du = 0$.

Then A.1.to A.4. holds in $L_p\left(\mathbb{R}^d, \mathbb{B}^d, \lambda\right)$, *with* $1 \leq p < \infty$, *for the generalized kernels in (6).*

5.2. HISTOGRAM

A broad class of density estimators is de histogram. It is defined by means of measurable partitions. Denote by $m = \{A_1, A_2, ...\}$ a measurable Borel partition of $\mathbb{R}^d$, which is formed by non null and finite λ-measure sets. Denote the set of such partitions by $\mathbb{I}$, which is ordered by the partial preorder $m_1 \leq m_2$ iff m_2 is more thin [6] than m_1. Then $\mathbb{I}$ is a directed set. Sometimes we take a subset $\mathbb{I}_0 \subset \mathbb{I}$ of nested partitions. Define the partitioning approximator by the generalized kernel,

$$K_m(x,z) = \sum_{A\in m} \frac{I_A(x)\,I_A(z)}{\lambda(A)}, \tag{7}$$

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 $f \in L_1\left(\mathbb{R}^d, \mathbb{B}^d, \lambda\right)$ 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 estimators that is known, Graunt (1662) is an early reference. It is studied by Rèvesz ((1971), (1972), (1973), (1974)), Tukey (1977), Scott ((1979), (1992, Chap. 3)), Freedman and Diaconis (1981), among others. Universal consistency of the histogram has been established by Abou-Jaude (1976a, 1976b, 1976c) and Devroye and Györfi (1985).

Proposition 6 *A.1.to A.4. holds for the generalized kernels in (7).*

[6] We say π_2 is thinner than π_1, iff every set in π_2 belongs to some set of π_1; that is, $\forall A_1 \in \pi_1$, $A_2 \in \pi_2$ then $A_2 \subset A_1$, or $A_2 \cap A_1 = \emptyset$.

5.3. ESTIMATORS BASED ON ORTHONORMAL HILBERT SPACE BASIS

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

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

We say that the set $\{e_k(z)\}_{k=1}^m \subset L_2(\mathbb{R}^d, \mathbb{B}^d, \mu)$ is an orthonormal set if $\langle e_k, e_s \rangle_{L_2(\mu)} = I_{\{k=s\}}$.

The use of an orthonormal set is relevant, because if we have an orthonormal set $\{e_k(z)\}_{k=1}^m$, the orthogonal projection of an arbitrary $f \in L_2(\mathbb{R}^d, \mathbb{B}^d, \mu)$, onto the linear subspace spanned by this set, 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,

$$K_m(x, z) = \sum_{k=1}^m e_k(x)\, e_k(z), \tag{8}$$

then, the projection can be expressed as,

$$\begin{aligned} \alpha_m(f; x) &= \int K_m(x, z)\, f(z)\, \mu(dz) \\ &= \int \left(\sum_{k=1}^m e_k(x)\, e_k(z) \right) f(z)\, \mu(dz). \end{aligned}$$

Thus, we say that a sequence $\{e_k(z)\}_{k=1}^\infty$ is an orthonormal Hilbert space basis if the sequence of projections $\{\alpha_m\}_{m=1}^\infty$, is a linear approximator of the identity in $L_2(\mathbb{R}^d, \mathbb{B}^d, \mu)$; or equivalently, iff the set $\{e_k(z)\}_{k=1}^\infty$ has a span which is dense in the $L_2(\mathbb{R}^d, \mathbb{B}^d, \mu)$ space. Using the Zorn's Lemma it can be proved that every Hilbert space has, at least, an orthonormal Hilbert basis (see Kreyszig (1978, pp 212)).

Notice that the corresponding density estimation is just

$$\begin{aligned} \widehat{f}_n(x) &= \frac{1}{n} \sum_{i=1}^n K_m(x, X_i) = \sum_{k=1}^m \widehat{f}_{k,n} \cdot e_k(x), \\ \widehat{f}_{k,n} &= \frac{1}{n} \sum_{i=1}^n e_k(X_i). \end{aligned}$$

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

Establishing the approximation property is far from be obvious. A possibility is to use Theorem 1, however even in this case the conditions are not easy to check.

Proposition 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 is a* k_0 *such that* $e_{k_0}(x)=1$, *a.s.*$[\mu]$.

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

O.4. $\forall\delta>0$, *and all compact set* 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. holds for the generalized kernel (8), 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 result about singular integral estimators.

A useful technique in $L_2(\mathbb{R}^d, \mathbb{B}^d, \mu)$ consists of taking orthonormal polynomials. The corresponding base is obtained by means of the Graham-Schmidt algorithm of orthonormalization (See Davis (1975) and Cheney (1981)), applied to a previous set of functions that are dense in $L_2(\mathbb{R}^d, \mathbb{B}^d, \mu)$, usually $\{x^k\}_{k=1}^{\infty}$, provided that such functions belongs to $L_2(\mathbb{R}^d, \mathbb{B}^d, \mu)$. This orthonormal basis contains polynomials like,

$$e_k(x)=a_{k,k}x^k+a_{k,k-1}x^{k-1}+\ldots..+a_{k,0},$$

where the coefficients are obtained from the Graham Schmidt technique. Establishing the approximation property is not obvious, but in the case of orthonormal polynomials in $L_2(\mathbb{R}, \mathbb{B}, \mu)$, we can use the Christoffel-Darboux formulae,

$$K_m(x,z)=\sum_{k=1}^{m} e_k(x)\,e_k(z)=\frac{a_{m,m}}{a_{m+1,m+1}}\left(\frac{e_{m+1}(x)\,e_m(z)-e_m(x)\,e_{m+1}(z)}{x-z}\right).$$

(See Davis (1975) and Cheney (1981)). This result can be used to check assumption 4 because,

$$\int |K_m(x,z)|\,|x-z|\,\mu(dz)\le\frac{a_{m,m}}{a_{m+1,m+1}}\int |e_{m+1}(x)\,e_m(z)-e_m(x)\,e_{m+1}(z)|\,\mu(dz).$$

6. ON THE BIAS OF REGRESSION CURVES

Let $P^{X,Y}$ a probability function in $(\mathbb{R}^{d+1}, \mathbb{B}^{d+1})$ and P^X the marginal probability function of X. Assume that $E[|Y|^p] < \infty$, with $1 \leq p < \infty$, so,

$$g_{P^{X,Y}}(x) = E_{P^{X,Y}}[Y|X = x] \in L_p(\mathbb{R}^d, \mathbb{B}^d, P^X).$$

Given a random sample $\{(X_i, Y_i), i = 1, ..., n\}$ from $P^{X,Y}$, if we have an linear approximator of the identity $\{\beta_m\}_{m\in\mathbb{I}}$ in $L_p(\mathbb{R}^d, \mathbb{B}^d, P^X)$ with,

$$\beta_m(f; x) = \int W(x, z) f(z) \, P^X(dz),$$

then se can get a delta sequence estimator of the regression function as,

$$\widehat{g}_n(x) = \frac{1}{n}\sum_{i=1}^{n} W_{m_n}(x; X_i) Y_i.$$

where $m_n = m(n)$ is a divergent sequence in n. By the law of iterated expectation,

$$E[\widehat{g}_n(x)] = E[W_{m_n}(x; X_i) Y_i] = \int W(x, z) g_{P^{X,Y}}(z) \, P^X(dz) = \beta_{m_n}(g_{P^{X,Y}}; x),$$

so $\widehat{g}_n(x)$ is asymptotically unbiased if,

$$\lim_{n\in\mathbb{I}} \left\|\beta_{m_n}(g_{P^{X,Y}}; x) - g_{P^{X,Y}}(x)\right\|_{L_p(P^X)} = 0.$$

This condition universally holds if $\{\beta_m\}_{m\in\mathbb{I}}$ is an approximator of the identity in $L_p(\mathbb{R}^d, \mathbb{B}^d, P^X)$. However, in practice, this kind of approximator $\{\beta_m\}_{m\in\mathbb{I}}$ can be obtained when the probability P^X of the regressors is known. It happens, for instance, when we have deterministic regressors. In this case, the techniques developed in this work are also useful. Higher order rates of convergence for the bias are also applicable when P^X is the uniform distribution in $\prod_{j=1}^{d}[a_j, b_j]$.

For instance, in the orthonormal series regression estimator in $L_2(\mathbb{R}^d, \mathbb{B}^d, P^X)$ take $W_{m_n}(x; z) = \sum_{k=1}^{m_n} e_k(x) e_k(z)$, and

$$\begin{aligned}
\widehat{g}_n(x) &= \frac{1}{n}\sum_{i=1}^{n}\left(\sum_{k=1}^{m_n} e_k(x) e_k(X_i)\right) \cdot Y_i = \sum_{k=1}^{m_n} \widehat{g}_{k,n} \cdot e_k(x). \\
\widehat{g}_{k,n} &= \frac{1}{n}\sum_{i=1}^{n} e_k(X_i) \cdot Y_i.
\end{aligned}$$

Notice that if P^X is unknown, then is also unknown the orthonormal Hilbert basis $\{e_k(x)\}_{k=1}^{\infty}$.

7. PROOFS

7.1. PROOFS OF SECTION 2.

PROOF OF THEOREM 1.

The following theorem provides high level assumptions which ensures that net of linear operators is a linear approximator of the identity. This is obtained from the Banach-Steinhaus theorem (see e.g. Rudin (1966)).

Theorem 3 *Let* $(B, \|\cdot\|_B)$ *a Banach space and* $\{\alpha_m\}_{m\in\mathbb{I}}$ *a net of linear operators such that,*

$$\begin{array}{rccc} \alpha_m : & B & \longrightarrow & B \\ & f & \longmapsto & \alpha_m(f). \end{array}$$

If,

1. $\{\alpha_m\}_{m\in\mathbb{I}}$ *is uniform bounded in* $(B, \|\cdot\|_B)$*, that is*

$$\sup_{m\in\mathbb{I}} \|\alpha_m\| = \sup_{m\in\mathbb{I}} \left\{ \sup_{\|f\|_B\leq 1} \|\alpha_m(f)\|_B \right\} < \infty.$$

2. *There exists a subset* $\mathcal{G} \subset B$*, dense en* B*, such that,*

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

Then $\{\alpha_m\}_{m\in\mathbb{I}}$ *is a linear approximator of the identity in* B*. If also* $\|\alpha_m\| < \infty$ *for each* $m \in \mathbb{I}$*, then the conditions are necessary.*

Proof.

See e.g. Kantorovich and Akilov (1982, Th.. 3, pp.. 203) and Davis (1975, pp.. 351).

■

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 (1974, Th. 3.3.1.)). The theorem follows applying theorem 3 and the following lemmas.

Lemma 2 *Let* $\{\alpha_m\}_{m\in\mathbb{I}}$ *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 operator norm verify,*

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

Furthermore, the uniform boundness of $\{\lfloor\alpha\rfloor_m\}_{m\in\mathbb{I}}$ *implies the uniform boundness of* $\{\alpha_m\}_{m\in\mathbb{I}}$.

Proof.

If $f \in L_p\left(\mathbb{R}^d, \mathbb{B}^d, \mu\right)$, then,

$$|\alpha_m(f;x)| \leq \int |K_m(x,z)| \, |f(z)| \, \mu(dz) = \lfloor \alpha \rfloor_m (|f|;x) \quad a.s. [\mu].$$

Because the norm $\|\cdot\|_{L_p(\mu)}$ is lattice,

$$\|\alpha_m(f;x)\|_{L_p(\mu)} \leq \|\lfloor \alpha \rfloor_m (|f|;x)\|_{L_p(\mu)}.$$

then $\forall f \in L_p\left(\mathbb{R}^d, \mathbb{B}^d, \mu\right)$,

$$\|\alpha_m(f;x)\|_{L_p(\mu)} \leq \|\lfloor \alpha \rfloor_m (|f|;x)\|_{L_p(\mu)} \leq \|\lfloor \alpha \rfloor_m\|_{L_p} \cdot \|f\|_{L_p(\mu)}.$$

and so $\|\alpha_m\|_{L_p} \leq \|\lfloor \alpha \rfloor_m\|_{L_p}$.

■

The previous lemma and assumption A.1. implies that $\{\alpha_m\}_{m\in\mathbb{I}}$ is uniformly bounded. By theorem 3, it is sufficient to establish that,

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

Lemma 3 *If the net $\{\alpha_m\}_{m\in\mathbb{I}}$ satisfies the conditions of the theorem, for all $f \in C_c\left(\mathbb{R}^d\right)$,*

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

Proof.

First, notice that,

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

By (A.2.), for all $f \in C_c\left(\mathbb{R}^d\right)$,

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

Function $(f(z) - f(x)) \in C_c\left(\mathbb{R}^d \times \mathbb{R}^d\right)$ so then we can restrict the measure μ to a compact set C. Denote such restricted measure by μ_C. Then,

$$\|\alpha_m(f;x) - f(x)\|_{L_p(\mu)} = \left\| \int (f(z) - f(x)) \, K_m(x,z) \, \mu_C(dz) \right\|_{L_p(\mu_C)} + o(1).$$

Since f is uniform continuous, for all $\varepsilon > 0$, $\exists \delta > 0$, such that, if $\|x - z\| \leq \delta$ then $|f(x) - f(z)| \leq \varepsilon$. Thus,

$$\begin{aligned}
\|\alpha_m(f;x) - f(x)\|_{L_p(\mu)} &\leq \left\| \int_{\{z:\|x-z\|\leq\delta\}} |f(z) - f(x)|\, |K_m(x,z)|\, \mu_C(dz) \right\|_{L_p(\mu_C)} + \\
&\quad + \left\| \int_{\{z:\|x-z\|>\delta\}} (f(z) - f(x))\, K_m(x,z)\, \mu_C(dz) \right\|_{L_p(\mu_C)} + o(1) \\
&\leq \varepsilon \cdot \left\| \int |K_m(x,z)|\, \mu_C(dz) \right\|_{L_p(\mu_C)} + \\
&\quad + \left\| \int_{\{z:\|x-z\|>\delta\}} h(x,z)\, K_m(x,z)\, \mu_C(dz) \right\|_{L_p(\mu_C)} + o(1),
\end{aligned}$$

where $h(x,z) = (f(z) - f(x))$.

The first term is arbitrarily small by A.1. and A.3.,

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

The second term is $o(1)$ because for all $h(x,z) \in C_c(\mathbb{R}^d \times \mathbb{R}^d)$ we have $\|h\|_\infty < \infty$ and then,

$$\begin{aligned}
&\left\| \int_{\{z:\|x-z\|>\delta\}} K_m(x,z)\, h(x,z)\, \mu_C(dz) \right\|_{L_p(\mu_C)} \\
\leq\ & \|h\|_\infty \cdot \left\| \int_{\{z:\|x-z\|>\delta\}} |K_m(x,z)|\, \mu_C(dz) \right\|_{L_p(\mu_C)} \underset{m\in\mathbb{I}}{\longrightarrow} 0.
\end{aligned}$$

by A.4..

PROOF OF PROPOSITION 1.

Since $\lfloor\alpha\rfloor_m$ is a monotone operator and the norm $\|\cdot\|_{L_p(\mu)}$ is lattice,

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

thus A.4. is satisfied.

PROOF OF PROPOSITION 2.

This is an inmediate consequence of the Lebesgue's Theorem of dominated convergence.

7.2. PROOFS OF SECTION 3

PROOF OF PROPOSITION 3.

By the triangle inequality,

$$\begin{aligned}\|f(x)-\alpha_m(f;x)\|_{L_p(\lambda)} &\leq \|f(x)-f(x)\cdot\alpha_m(\mathbf{1};x)\|_{L_p(\lambda)} \\ &\quad+\|f(x)\cdot\alpha_m(\mathbf{1};x)-\alpha_m(f;x)\|_{L_p(\lambda)} \\ &= \|f(x)(1-\alpha_m(\mathbf{1};x))\|_{L_p(\lambda)} \\ &\quad+\|f(x)\cdot\alpha_m(\mathbf{1};x)-\alpha_m(f;x)\|_{L_p(\lambda)},\end{aligned}$$

since $f\in W_p\left(\mathbb{R}^d,\mathbb{B}^d,\lambda\right)$. Using the mean value theorem argument,

$$\begin{aligned}f(z) &= f(x)+R(x,(z-x)) \\ R(z,(z-x)) &= O(\|x-z\|).\end{aligned}$$

almost everywhere. Thus,

$$\alpha_m(f;x)=f(x)\,\alpha_m(\mathbf{1};x)+\alpha_m(R(z,(z-x));x),$$

and

$$|\alpha_m(f;x)-f(x)\,\alpha_m(\mathbf{1};x)|\leq\lfloor\alpha\rfloor_m(|R(z,(z-x))|;x),$$

which implies that,

$$\|\alpha(f;x)-f(x)\,\alpha_m(\mathbf{1};x)\|_{L_p(\lambda)}\leq\|\lfloor\alpha\rfloor_m(|R(z,(z-x))|;x)\|_{L_p(\lambda)}.$$

Since $\lfloor\alpha\rfloor_m$ is monotone,

$$\lfloor\alpha\rfloor_m(|R(z,(z-x))|;x)=O(\lfloor\alpha\rfloor_m(\|x-z\|;x)),$$

and taking into account that the norm $\|\cdot\|_{L_p(\lambda)}$ is lattice,

$$\|\alpha_m(R(z,(z-x));x)\|_{L_p(\lambda)}=O\left(\|\lfloor\alpha\rfloor_m(\|x-z\|;x)\|_{L_p(\lambda)}\right)=O(\zeta(m)).$$

Therefore,

$$\|\alpha(f;x)-f(x)\,\alpha_m(\mathbf{1};x)\|_{L_p(\lambda)}=O(\zeta(m)).$$

■

PROOF OF PROPOSITION 4.

Obviously,

$$\widetilde{\alpha}_m(\mathbf{1};x) = \frac{\alpha_m(\mathbf{1};x)}{\alpha_m(\mathbf{1};x)} \cdot I_{\{|\alpha_m(1;x)|>0\}}(x) = 1 \quad a.s.\,[\lambda].$$

Then,

$$\begin{aligned} |\widetilde{\alpha}_m(f;x) - f(x)| &= \left|\frac{\alpha_m(f;x)}{\alpha_m(\mathbf{1};x)} - f(x)\right| = \left|\frac{\alpha_m(f;x) - f(x)\,\alpha_m(\mathbf{1};x)}{\alpha_m(\mathbf{1};x)}\right| \quad a.s.\,[\lambda], \\ &\leq c^{-1} \cdot |\alpha_m(f;x) - f(x)\,\alpha_m(\mathbf{1};x)| \\ &= c^{-1} \cdot \left|\int K_m(x,z)\,(f(z) - f(x))\,\lambda(dz)\right|, \end{aligned}$$

using the regularity condition.

PROOF OF COROLLARY 1.

Is an immediate consequence of Proposition 4, and 3.

PROOF OF THEOREM 2.

The proof of the theorem 2 is based on the following proposition, which permits to apply higher order Taylor expansions to the linear operators.

Proposition 8 *Let $\{\alpha_m\}_{m\in\mathbb{I}}$ be a linear approximator of the identity in $L_p(\mathbb{R}^d, \mathbb{B}^d, \lambda)$ with $1 \leq p < \infty$. Then,*

(i) *Assuming that $f \in W_p^s(\mathbb{R}^d, \mathbb{B}^d, \lambda)$ with $s \in \mathbb{N}$. Thus,*

$$\left\|\alpha_m(f;x) - f(x)\,\alpha_m(\mathbf{1};x) - \sum_{l=1}^{s-1} \sum_{\|\nu\|_1 = l} \frac{1}{\nu!}\alpha_m\left((z-x)^\nu;x\right) D^\nu f(x)\right\|_{L_p(\lambda)} = O\left(\zeta_s(m)\right),$$

where $(z-x)^\nu = \prod_{j=1}^d (z_j - x_j)^{\nu_j}$, $\nu! = \prod_{j=1}^d \nu_j!$ and

$$\alpha_m\left((z-x)^\nu;x\right) = \alpha_m\left(\prod_{j=1}^d (z_j - x_j)^{\nu_j};x\right)$$

$$\zeta_s(m) = \left\|\lfloor\alpha\rfloor_m\left(\|z-x\|^s;x\right)\right\|_{L_p(\lambda)}.$$

(ii) *Assume that $f \in W_p^s(\mathbb{R}^d, \mathbb{B}^d, \lambda)$ with $s \in \mathbb{N}$. Thus,*

$$\left|\alpha_m(f;x) - f(x)\,\alpha_m(\mathbf{1};x) - \sum_{l=1}^{s-1} \sum_{\|\nu\|_1 = l} \frac{1}{\nu!}\alpha_m\left((z-x)^\nu;x\right)\right| = O\left(\zeta_s(m;x)\right) \quad a.s.\,[\lambda]$$

where

$$\zeta_s(m;x) = \lfloor\alpha\rfloor_m\left(\|z-x_0\|^s;x_0\right).$$

(iii) *Assume that $f \in C_c^s\left(\mathbb{R}^d\right)$ with $s \in \mathbb{N}$. Thus,*

$$\left\|\alpha_m(f;x) - f(x)\,\alpha_m(\mathbf{1};x) - \sum_{l=1}^{s-1}\sum_{\|\nu\|_1=l}\frac{1}{\nu!}\alpha_m\left((z-x)^\nu;x\right)D^\nu f(x)\right\|_{L_\infty(\lambda)} = O\left(\zeta_s^\infty(m)\right),$$

where

$$\zeta_s^\infty(m) = \left\|\lfloor\alpha\rfloor_m\left(\|z-x\|^s;x\right)\right\|_{L_\infty(\lambda)}.$$

Proof.

If $f \in W_p^s\left(\mathbb{R}^d, \mathbb{B}^d, \lambda\right)$, then in almost every point we have a Taylor expansion,

$$f(z) = f(x) + \sum_{l=1}^{s-1}\sum_{\|\nu\|_1=l}\frac{1}{\nu!}(z-x)^\nu D^\nu f(x) + R_s(z,(z-x))$$

where,

$$\lim_{z\to x}|R_s(z,(z-x))| = o\left(\|z-x\|^s\right).$$

Therefore,

$$\begin{aligned}\alpha_m(f;x) &= f(x)\,\alpha_m(\mathbf{1};x) + \sum_{l=1}^{s-1}\sum_{\|\nu\|_1=l}\frac{1}{\nu!}D^\nu f(x)\,\alpha_m\left((z-x)^\nu;x\right)\\ &\quad+\alpha_m\left(R_s(z,(z-x));x\right).\end{aligned}$$

Notice that $\lfloor\alpha\rfloor_m$ is a monotone operator, $|R_s(z,(z-x))| = O\left(\|z-x\|^s\right)$, so for almost every point,

$$|\alpha_m(R_s(z,(z-x));x)| \le \lfloor\alpha\rfloor_m\left(|R_s(x,(z-x))|;x\right) = O\left(\lfloor\alpha\rfloor_m\left(\|z-x\|^s;x\right)\right).$$

Since the norm $\|\cdot\|_{L_p(\lambda)}$ is lattice,

$$\begin{aligned}\|\alpha_m(R_s(x,(z-x));x)\|_{L_p(\lambda)} &\le \left\|\lfloor\alpha\rfloor_m\left(|R_s(x,(z-x))|;x\right)\right\|_{L_p(\lambda)}\\ &= O\left(\left\|\lfloor\alpha\rfloor_m\left(\|z-x\|^s;x\right)\right\|_{L_p(\lambda)}\right)\\ &= O\left(\zeta_s(m)\right).\end{aligned}$$

Hence,

$$\left\|\alpha_m(f;x) - f(x)\,\alpha_m(\mathbf{1};x) - \sum_{l=1}^{s-1}\sum_{\|\nu\|_1=l}\frac{1}{\nu!}D^\nu f(x)\,\alpha_m\left((z-x)^\nu;x\right)\right\|_{L_p(\lambda)} = O\left(\zeta_s(m)\right).$$

Proof for the other convergence criterions is similar.

■

The proof of the theorem follows fairly straightforwardly from the above proposition. There exist an m_0 such that for all $m > m_0$,

$$\left\|D^\nu f(x)\,\alpha_m\left((z-x)^\nu;x\right)\right\|_{L_p(\lambda)} = 0$$

for all ν such that $1 \leq \|\nu\|_1 \leq \tau - 1$.

By proposition 8, for $\tau \leq s$ is satisfied that,

$$\begin{aligned}
\|\alpha_m(f;x) - f(x)\alpha_m(\mathbf{1};x)\|_{L_p(\lambda)} &\leq \left\| \sum_{\|l\|_1=1}^{\tau-1} \sum_{\|\nu\|_1=l} \frac{1}{\nu!} D^\nu f(x)\, \alpha_m\left((z-x)^\nu;x\right) \right\|_{L_p(\lambda)} + \\
&\quad + O\left(\left\| \lfloor \alpha \rfloor_m \left(\|z-x\|^\tau ;x\right) \right\|_{L_p(\lambda)} \right).
\end{aligned}$$

then,

$$\|\alpha_m(f;x) - f(x)\alpha_m(\mathbf{1};x)\|_{L_p(\lambda)} = O\left(\left\| \lfloor \alpha \rfloor_m \left(\|z-x\|_1^\tau ;x\right) \right\|_{L_p(\lambda)} \right).$$

Since $\alpha_m(\mathbf{1};x) = 1 \quad a.s.[\lambda]$, for all $m \geq m_0$, the result follows. For the rest of the convergence criterion is similar.

PROOF OF LEMMA 1.

Assume that $\{\lfloor\alpha\rfloor_m\}_{m\in\mathbb{I}}$ is a linear approximator of the identity. If $\lim_{m\in\mathbb{I}} \zeta(m) = 0$, by proposition 1,

$$\lim_{m\in\mathbb{I}} \left\| \alpha_m\left(I_{\{\|z-x\|>\delta\}}(z);x\right) \right\|_{L_p(\lambda)} = 0, \quad \forall\delta > 0,$$

Then, for m enough large, in almost every point the support of the kernel verify,

$$Supp\,(K_m(x,z)) \subset \left\{(x,z) \in \mathbb{R}^{2d} : \|z-x\| \leq 1\right\}.$$

If $1 \leq \tau_1 < \tau_2$, in the $Supp\,(K_m(x,z))$,

$$\|z-x\|^{\tau_2} \leq \|z-x\|^{\tau_1},$$

with strict inequality in almost every point (x,z). For those points such that $x = z$ we have the equality.

Using that $\lfloor\alpha\rfloor_m$ is a monotone operator,

$$\lfloor\alpha\rfloor_m\left(\|z-x\|^{\tau_2};x\right) < \lfloor\alpha\rfloor_m\left(\|z-x\|^{\tau_1};x\right)$$

and noting that $\|\cdot\|_{L_p(\lambda)}$ is lattice,

$$\begin{aligned}
0 &\leq \zeta_{\tau_2}(m) = \left\|\lfloor\alpha\rfloor_m\left(\|z-x\|^{\tau_2};x\right)\right\|_{L_p(\lambda)} \\
&< \left\|\lfloor\alpha\rfloor_m\left(\|z-x\|^{\tau_1};x\right)\right\|_{L_p(\lambda)} = \zeta_{\tau_1}(m).
\end{aligned}$$

By an analogous argument,

$$0 \leq \zeta_{\tau_1}(m) < \zeta(m).$$

so then $\lim_{m\in\mathbb{I}} \zeta(m) = 0$ implies $\lim_{m\in\mathbb{I}} \zeta_\tau(m) = 0$, $\forall\tau > 1$. If $1 \leq \tau_1 < \tau_2 < \infty$ then

$$\zeta_{\tau_2}(m) = o\left(\zeta_{\tau_1}(m)\right).$$

7.3. PROOFS OF SECTION 5

PROOF OF PROPOSITION 5.

We use theorem (1). First note that assumption S.1. implies the A.1, as a consequence of next result.

Lemma 4 ***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* $1 \leq p < \infty$,

$$\|\lfloor\alpha\rfloor_m(f;x)\|_{L_p(\lambda)} \leq \|K_m\|_{L_1(\lambda)} \cdot \|f\|_{L_p(\lambda)}.$$

Proof.

Using the integral's Minkowsky inequality and Fubini theorem and the invariance of Lebesgue's measure is invariant to translations, then,

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

■

Note than λ satisfies A.3.. The assumption A.4. is a consequence of S.3. For each compact set C,

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

changing the variable $u = (x-z)$,

$$= \int_{u\in x-C} |K_m(u)|\,\|u\|\,du.$$

Then,

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

PROOF OF PROPOSITION 6

We use the Theorem 1. First we check condition A.1.. Note that α_m is a positive operator, so $\alpha_m = \lfloor\alpha\rfloor_m$. For all $m \in \mathbb{I}$,

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

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

A.2. is immediate because, for all $m \in \mathbb{I}$,

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

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,

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

But for a partition $m_\delta \in \mathbb{I}$ thin enough, we have that for every $A \in m_\delta$,

$$\sup_{x,z\in A} \|x - z\| \leq \delta.$$

Then for all $m \geq m_\delta$,

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

Thus,

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

and by dominated convergence,

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

PROOF OF PROPOSITION 7.

Assumption O.1. ensures that condition A.1. holds. Note that the projection operators $\{\alpha_m\}_{m\in\mathbb{I}}$ has norm

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

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

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

because 1 belongs to the space $span(\{e_k\}_{k=1}^m)$, ant its projection is just equal to itself.

Assumptions A.3. and A.4. are a consequence of O.3. and O.4., respectively.

8. REFERENCES

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

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

Adams, R. A. (1978): *Sobolev Spaces.* Academic Press, San Diego.

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

Berlinet, A. (1991): "Reproducing kernels and finite orden kernels". In: *Nonparametric Functional Estimation and Related Topics.* (Proceedings of the NATO Advances Study Institute on Nonparametric Functional Estimation and Related Topics, Spetses, Grece, July 29-August 10, 1990). Kluwer Academic Publishers, Dordrecht.

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.

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

Cencov, N. N. (1962): "Evaluation of an unknown density by orthogonal series". *Soviet Math. Dokl.*, **3**, 1559-1562.

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

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

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

Delgado M. A. and W. Gonzalez-Manteiga (1998): "Significance testin in nonparametric regression based on the Bootstrap". *W.P. 96-98, Statistics and Econometrics series,* University Carlos III de Madrid, Spain.

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

Devroye, L. (1987): *A Course in Density Estimation.* Birkhauser, Boston.

Devroye, L. and L. Györfi (1985): *Nonparametric Density Estimation: The L_1 view.* Wiley series in Probability and Matehmatical statistics. Hohn Wiley & Sons, New York.

Dundford, N. and J. T. Schwartz (1957): *Linear operators. Part I. General theory.* Wiley Classics Library Edition, pub. 1988. New York, John Wiley & Sons.

Edgar, G. A. and L. Sucheston (1992): *Stopping Times and Directed Processes.* Enciclopedia of Mathematics and its Applications, (edit. por G.-C. Rota), **vol 47**. Cambridge, Cambridge University Press.

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

Gasser, T. and H. G. Müller (1984): "Optimal convergence properties of kernel estimates of derivatives of a density". En: *Smoothing Techniques for Curve Estimation*, (T. Gasser y M. Rosenblatt eds.), *Lecture Notes in Mathematics*, **757**, Springer Verlag, 144-154.

Gasser, T.; H. G. Muller and V. Mammitzsch (1985): "Kernels for nonparametric curve estimation". *Journal of the Royal Statistical Society, Ser. B*, **47**, 238-252.

Graunt, J. (1662): *Natural and political Observations Made upon the Bills of Mortality.* Martyn, Londres.

Hall, P. and J. S. Marron (1987): "Choice of the kernel order in density estimation". *Annals of Statistics*, **16**, 161-173.

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

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

Maz'ja, V. G. (1985): *Sobolev Spaces.* Springer Verlag, Berlin.

Powell, J. L.; J. H. Stock and T. M. Stoker (1989): "Semiparametric estimation of index coefficients". *Econometrica*, **57**, 1403-1430.

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

Rèvesz, P. (1971): "Testing of density functions". *Periodica Mathematica Hungarica*, **1**, 35-44.

Rèvesz, P. (1972): "On empirical density function". *Periodica Mathematica Hungarica*, **2**, 85-110.

Rèvesz, P. (1973): "A strong law of the empirical density funcion". *Trans 6th.Prague Confer.*, 469-472.

Rèvesz, P. (1974): "On empirical density function". En: *Probability and Statistical mehods -summer school.* Bulgarian Academy of Science, Varna.

Robinson, P. (1988): "Root-n consistent semiparametric regression". *Econometrica*, **56**, 931-954.

Rudin, W. (1974): *Real and Complex Ansalysis.* (2nd ed.). Mc Graw Hill, 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. ISBN 0-471-54770-0.

Singht, R. S. (1979): "Mean squared errors of estimates of a density and its derivatives". *Biometrika*, **66**, 177-180.

Terrell, G. R. (1984): "Eficiency of nonparametric density estimators". *Technical Report.* Department of Math. Sciences, Rice University.

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

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

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

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

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

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

Whittle, P. (1958): "On smoothing a probability density function". *Journal of the Royal Statistical Society, ser. B*, **20**, 334-343.

Winter, B. B. (1973): "Strong uniform consistency of integrals of density estimation". *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.