REVISTA MATEMATICA de la
Universidad Complutense de Madrid
Volumen 2, número 1; 1989.
http://dx.doi.org/10.5209/rev_REMA.1989.v2.n1.18146

Nonparametric estimation of probability density functions based on orthogonal expansions

ALDO JOSÉ VIOLLAZ

ABSTRACT. Let $(X_j)_{j=1}^n$ be i.i.d. r.v.'s each with density function f, and let $(k_n(x, t))$ be a sequence (a so-called kernel sequence) of Borel measurable functions defined on $\mathbb{R}\times\mathbb{R}$. Let $f_n(x)$ be the density function estimate defined by

$$f_n(x)=n^{-1}\sum_{j=1}^n k_n(x, X_j).$$

We prove that under general conditions on f and (k_n), $(f_n(x))$ is consistent in the mean square sense. We find an asymptotic expression for the variance of the estimate and prove that its asymptotic distribution is Gaussian. These results apply to a large class of density estimates which includes the estimates considered by Parzen (1962), Leadbetter (1963) with kernels with compact support and also those estimates derived fron orthogonal expansions.

Density estimates derived from trigonometric and Jacobi orthogonal expansions are studied in detail. For f belonging to classes of functions defined in terms of the derivatives of f, we find explicit bounds for the mean square error of the estimates, holding uniformly over these classes. We compare the rates of mean square consistency obtained with the best possible rates found by Farrell and Wahba.

0. INTRODUCTION AND SUMMARY

0.1. Introduction

In all this work we only consider the estimation of one dimensional probability density functions f with respect to Lebesgue measure on $\mathbb{R}$.

There are at least two well known nonparametric methods for estimating a probability density function (p.d. function), namely, the kernel method and the orthogonal series method. Let $(X_j)_{j=1}^n$ be n independent observations of a random variable X with p.d. function f. The kernel method consists of choosing as estimator of the p.d. function at x the following

$$f_n(x)=\frac{1}{n}\sum_{j=1}^{n} k_m(x-X_j) \qquad [0.1]$$

1980 Mathematics Subject Classification (1985 revision): 62G05.
Editorial de la Universidad Complutense. Madrid, 1989.

where k_m is a sequence of functions satisfying certain conditions and $m = m(n)$ is a sequence of integers depending on n. Essentially this method is a generalization of the intuitive procedure which consists of choosing a narrow interval around the point x and estimating $f(x)$ by the number of observations X_j belonging to that interval divided by n times the length of the interval. There are several papers about the local properties of kernel p.d. functions estimators: Bartlett (1963), Rosenblatt (1956, 1971), Parzen (1962), Leadbetter (1963), Woodroofe (1957), etc. See Wegman (1972) for more detailed references. In the orthogonal series method we have to choose a convenient orthogonal system of functions and write the p.d. function f as the corresponding orthogonal series expansion. In order to estimate f we first have to cut the series keeping only a finite number of terms, and then estimate the coefficients of this finite series. This method was studied among others by Schwartz (1967) using the system of Hermite functions, by Kronmal and Tarter (1968) using the trigonometric system of functions and by Cencov (1962) using a general orthogonal system. See also Rosenblatt (1971), Watson (1969), Crain (1974), Wegman (1972), Hall (1982) and Viollaz (1980).

The research in the area of nonparametric estimation of probability density functions has grown exponentialy. In the last twenty years much research work has been done, dealing with both the kernel and the orthogonal series methods as well as with others nonparametric methods like the penalized maximum likelihood of Good and Gaskins (1971, 1980), the near neighbor estimators of Loftsgaarden and Quesenbarry (1965), the spline methods of Wahba (1971) or the histogram type estimator of Van Ryzin (1969).

The questions studied also covers a wide spectrum, running from problems of consistency in several senses to problems of asymptotic distribution of some functionals of the density estimator as in Bickel and Rosenblatt (1973) or Viollaz (1976, 1980).

The important problem of the choice of the bandwith in the kernel estimator or equivalently the choice of the number of terms in the orthogonal series estimator has been studied by several authors. See Woodroofe (1970), Kronmal and Tarter (1968), Duin (1976), Hermans and Habbema (1976), and Viollaz and Cardozo (1980). See also Hall (1982) and Marron (1985) for more recent works on this problem.

The present paper intends to throught light over the question: how much different or how much similar are the kernel and the orthogonal series density estimates?

For this, we introduce a class of density estimators of variable kernel type which includes both the kernel and the orthogonal series density estimators. For this class we study pointwise and uniform consistency in the Mean

Square Error sense, we find an asymptotic expression for the variance of the estimator and prove its asymptotic normality. The results apply to the algebraic estimators of Parzen with compact support, to the Leabbetter estimators with compact support and to estimators based on orthogonal expansions in the trigonometric system and in the system of Jacobi polynomials.

The contain and organization of the paper is as follow:

In Section 1 we state some results from Alexits (1961) which are used in the sequel. Some elementary properties of kernel sequences are proved here.

In Section 2, some local properties of the estimators are studied. It is proved that the sequence of estimators is consistent in the mean square sense. We find an asymptotic expression for the variance of the estimator and we show that its asymptotic distribution is Gaussian. The results of this section apply for a general class of estimators which includes the so-called algebraic estimators of Parzen with compact support and also classes of estimators constructed using the orthogonal series method. The estimators derived from the trigonometric and Jacobi system are included here. We hope the results of this chapter will contribute to understanding the similarities and differences between the algebraic estimators and the orthogonal series estimators.

In Section 3 estimators derived from Jacobi orthogonal expansions are studied in detail. The results of Section 2 are applied to prove the mean square consistency, to find an asymtotic expression for the variance of the estimator and to show that its asymptotic distribution is Gaussian. For f belonging to classes of functions defined in terms of the derivatives of f, we find explicit bounds for the mean square error for the estimators, holding uniformly over these classes. We compare these rates of mean square consistency with the best possible rates found by Farrell (1972) and Wahba (1975). Explicit bounds for the Mean Square Error are obtained for estimators derived from Legendre Series. It is shown that the estimator derived from the Cesaro summation of the Legendre series is saturated with rate $n^{-2/3}$ in the sense that the best possible rate of convergence to zero of the Mean Square Error is $n^{-2/3}$, independently of the smoothness of the function f.

0.2. The Estimator

Let $(\phi_v)_{v=0}^{\infty}$ be a complete orthonormal system with respect to a weight function ρ. It is known that if f is a square integrable function with respect to ρ, i.e. $\int f^2\rho$ is finite, then the orthonormal expansion of f,

$$\sum_{v=0}^{\infty} c_v \phi_v(x) \qquad [0.2]$$

where

$$c_v = \int f(x)\bar{\phi}_v(s)\rho(s)ds \qquad [0.3]$$

converges to f in the L^2_ρ sense. Here $\bar{\phi}_v$ stands for the complex conjugate of ϕ_v which in general is assumed to be a complex valued function of a real variable.

Let $(X_j)_{j=1}^n$ be a sequence of independent observations of a random variable X with p.d. function f and distribution function F, and let us assume that its orthonormal expansion converges pointwise to $f(x)$ at the point x, so we can write

$$f(x) = \sum_{v=0}^{\infty} c_v \phi_v(x) \qquad [0.4]$$

where c_v is given by [0.3]. The infinite sequence of coefficients $(c_v)_{v=0}^\infty$ is unknown since f is assumed to be unknown, so it is not of much help to use this formula for estimating $f(x)$. But we can do the following: Cut the expansion [0.4] keeping only a finite number of terms, say $m = m(n)$ expecting that this finite expansion will be a good approximation for $f(x)$, and then estimate the finite sequence of coefficients $(c_v)_{v=0}^m$. Writing c_v as

$$c_v = \int \bar{\phi}_v(s)\rho(s)dF(s) \qquad [0.5]$$

a natural estimate for c_v appears to be

$$\hat{c}_v = \int \bar{\phi}_v(s)\rho(s)dF_n(s) \qquad [0.6]$$

where F_n is the empirical distribution function corresponding to the finite sequence $(X_j)_{j=1}^n$. Therefore in a natural way we are led to consider p.d. function estimators of the form

$$f_n(x) = \sum_{v=0}^{m(n)} \hat{c}_v \phi_v(x) \qquad [0.7]$$

where

$$\hat{c}_v = \frac{1}{n} \sum_{j=1}^{n} \bar{\phi}_v(X_j)\rho(X_j). \qquad [0.8]$$

In the same way we can consider more general estimators of the form

$$f_n(x) = \sum_{v=0}^{m(n)} a_v(m)\hat{c}_v \phi_v(x) \qquad [0.9]$$

where the $a_v(m)$ are known numbers and the c_v are defined as above. This kind of motivation can be found in the existing literature on the subject; see Cencov (1962), Krommal and Tarter (1962), Rosemblatt (1971).

After replacing $\hat{c}_v$ by its expression [0.9] we obtain

$$f_n(x)=\sum_{j=1}^{n} k_m(x, X_j)\,\rho(X_j) \qquad [0.10]$$

where

$$k_m(x, s)=\sum_{v=0}^{m} a_v(m)\,\phi_v(x)\,\bar{\phi}_v(s). \qquad [0.11]$$

Parzen (1962), Rosenblatt (1956, 1971), Bickel and Rosemblatt (1973) among others consider density estimators of the from [0.10] with kernel

$$k_m(x, t)=\frac{1}{b(n)}\,K\left(\frac{x-t}{b(n)}\right). \qquad [0.12]$$

Following Leadbetter (1963) and Wegman (1972) we call these estimators *algebraic estimators.* Rosenblatt (1956), Whittle (1958), Watson and Leadbetter (1963) and Leadbetter (1963) consider estimators of the form [0.10] with kernel $k_m(x, t)=\delta_n(x-t)$. Woodroofe (1957) considers estimators of the form [0.10] with

$$k_m(x, t)=G\left(x, \frac{x-t}{b(n)}\right). \qquad [0.13]$$

We propose to study estimators of the form [0.10] independently of whether they originated from an orthogonal expansion or not. This, moreover, has two advantages.

(1) We get results which apply to the estimators derived from orthonormal expansions in trigonometric functions and Jacobi polynomials, and to the estimators considered by Parzen, Rosenblatt and Woodroofe introduced above, under the restriction that their corresponding kernel have compact support.

(2) We can better understand the similarities and differences between density estimators constructed using different functions k_m.

1. SOME RESULTS ABOUT ORTHOGONAL SERIES

1.1. The Singular Integrals

As stated before we will study p.d. function estimators of the form

$$f_n(x)=\frac{1}{n}\sum_{j=1}^{n} k_m(x, X_j)\,\rho(X_j)$$

where $m=m(n)$, independently of whether they originated from orthogonal expansions or not. Of course, we could absorb the factor $\rho(X_j)$ into $k_m(x, X_j)$ but we prefer the above form since the conditions we will impose on the kernel take a simpler form in this case.

Consistency of f_n is equivalent to the following: Under what conditions on k_m, ρ, and f, does the Lebesgue integral

$$I_m(f,x)=\int k_m(x,t)\rho(t)f(t)\,dt \qquad [1.1]$$

exist and converge to $f(x)$? So we will start studying this kind of integrals which were called by Lebesgue (1909) ~~singular~~ *integrals*. In this Section we will state some theorems about convergence of singular integrals, we will define orthogonal polynomial-like systems and will particularize these theorems to them. Also we will present some results about the kernels of the singular integrals.

Definition 1.1.1. *Let $(k_m(x,t))$ be a sequence of measurable real valued functions defined on the finite square $[a,b]\times[a,b]$ and let ρ be an integrable and a.e. strictly positive function on $[a,b]$. The pair $((k_m(x,t), \rho(t))$ will be called a kernel sequence with singular point x if for every $\lambda>0$ and every subinterval $[\alpha,\beta]$ of $[a,b]$, $a\leq\alpha<\beta\leq b$, the following conditions hold:*

a) $\lim_{m\to\infty}\int_I k_m(x,t)\rho(t)\,dt=1$ *and* $\lim_{m\to\infty}\int_J k_m(x,t)\rho(t)\,dt=0$

where $I=[a,b]\cap[x-\lambda,x+\lambda]$, $J=[\alpha,\beta]-[x-\lambda,x+\lambda]$,

b) $\sup\{\,|k_m(x,t)|: t\in[a,b]-[x-\lambda,x+\lambda]\}\leq L(x,\lambda)$

where $L(x,\lambda)$ is a finite function of x and λ but independent of m,

c) $\sup_t\{\,|k_m(x,t)|: t\in[a,b]\}=L_m(x)<\infty.$

The sequence of integrals $(I_m(f,x))$ constructed using a kernel sequence as above defined is known as a *singular sequence of integrals* (Lebesgue (1909)).

Theorem 1.1.1 (Lebesgue). *Let $(I_m(f,t))$ be a singular sequence of integrals in the sense of Definition 1.1.1. A necessary and sufficient condition for $I_m(f,x)$ to converge to $f(x)$ for every function f which is continuous at x and of bounded variation on $[a,b]$ is*

Condition 1.1.1. *There exists a constant K independent of m but eventually depending on x such that for every subinterval $[A,B]$ of $[a,b]$, $a\leq A\leq B\leq b$*

$$\left|\int_A^B k_m(x,t)\rho(t)\,dt\right|\leq K,$$

holds uniformly with respect to the interval $[A, B]$. *Moreover, if Condition 1.1.1 is uniformly satisfied for* $x \in [c, d] \subset [a, b]$, *the function of finite variation* f *is continuous in* $[c, d]$ *(It is understood that* f *is continuous at* c *from the left and at* d *from the right) and conditions a) and b) are uniformly satisfied for* $x \in [c, d]$ *then* $I_m(f, x)$ *converges uniformly to* f *for* $x \in [c, d]$.

Lebesgue (1909) proved this theorem and a proof is also given by Alexits (1961).

Theorem 1.1.2. *Let* $(I_m(f, x))$ *be a singular sequence of integrals in the subinterval* $[c, d]$ *of* $[a, b]$. *Then* $I_m(f, x)$ *converges at the point* $x \in [c, d]$ *to* $f(x)$, *for every* f *which is* L_ρ*-integrable and continuous in* $[c, d]$, *if and only if the following condition holds:*

Condition 1.1.2. *There exists a constant K independent of m, but eventually depending on x, such that*

$$\int_a^b | k_m(x, t) | \rho(t) dt \leq K.$$

Here and in what follows f L_ρ-integrable means that $\int_a^b |f(t)| \rho(t) dt$ is finite. For a proof of this theorem, see Alexits (1961), pp. 257-260.

Remark. Condition c) of Definition 1.1.1 in Alexits (1961) is stated under the folowing form: «..., assuming, that they (the singular integrals) exist for every L_ρ-integrable function». Both forms are essentially equivalent since condition c) implies the existence of $I_m(f, x)$ for every L_ρ-integrable function, and the existence of $I_m(f, x)$ for every L_ρ-integrable function f implies that ess sup $\{ | k_m(x, t) | : a \leq t \leq b\}$ is finite. See Alexits (1961), p. 247. In what follows we shall refer to Alexits (1961) only by his name and corresponding page numbers.

Theorem 1.1.3. *If the function* $f \in L_\rho$ *is uniformly continuous in a subset* E *of* $[a, b]$ *and conditions a) and b) of Definition 1.1.1 and Condition 1.1.2 are uniformly satisfied for* $x \in E$, *then* $I_m(f, x)$ *converges to* f *uniformly in* E.

For a proof of this theorem see Alexits.

Theorem 1.1.4 *Let the integrals* $I_m(f, x)$ *be singular at the point* $x \in [a, b]$. *If the function* f *is* ρ*-integrable, continuous at* x *and of bounded variation in some interval* $[x-\epsilon, x+\epsilon]$ *around the point* x, *and if Condition 1.1.1 holds, then* $I_m(f, x) \to f(x)$ *as* $n \to \infty$.

This theorem follows by combining the arguments of the sufficiency part of Theorems 1.1.1 and 1.1.2.

1.2. The Christoffel-Darboux Formula

In this paragraph we state some definitions and results about orthogonal systems of functions that will be needed later.

Let ρ be an almost everywhere positive and integrable function on the interval $[a, b]$ of $\mathbb{R}$, and let $(p_v)_{v=0}^{\infty}$ be the (unique) complete system of polynomials in x, orthonormal on the interval $[a, b]$ with respect to the weight function ρ, that is

$$\int_a^b p_i(x) p_j(x) \rho(x) dx = \delta_{ij}$$

where δ_{ij} is equal to one if $i = j$ and equal to 0 otherwise. If f is a ρ-integrable function (on $[a, b]$) then the m^{th} sum of its orthogonal expansion can be written as

$$s_m(f, x) = \int_a^b f(t) \sum_{v=0}^{m} p_v(x) p_v(t) \rho(t) dt. \qquad [1.2]$$

The function

$$K_m(x, t) = \sum_{v=0}^{m} p_v(x) p_v(t) \qquad [1.3]$$

is called the *m^{th} kernel of the system (p_v)*. We have used the notation K_m for this kernel instead of k_m because we want to reserve the latter for kernels which satisfy the conditions of Definition 1.1.1 and it is not known yet if K_m satisfies these conditions or not.

Lemma 1.2.1 (Christoffel-Darboux formula). *Let (p_v) be the system of orthonormal polynomials with respect to the weight function ρ, then*

$$\sum_{v=0}^{m} p_v(x) p_v(t) = \frac{\alpha_m}{\alpha_{m+1}} \frac{p_m(x) p_{m+1}(t) - p_{m+1}(x) p_m(t)}{t - x} \qquad [1.4]$$

where α_m and α_{m+l} denote the leading coefficients of p_m and p_{m+l}, respectively.

For a proof of this lemma see for example Alexits, pp. 25-26 or Szegö (1939), pp. 41-42.

The Cristoffel-Darboux formula holds for orthogonal polynomials. Alexits, motivated by the Christoffel-Darboux formula, has defined orthogonal systems that the calls polynomials-like. We now give Alexits' definition.

Definition 1.2.1. *An orthonormal system (ϕ_v) is called polynomial-like for $x \in E$, if its m^{th} kernel has the following structure*

$$K_m(x, t) = \sum_{k=1}^{r} F_k(x, t) \sum_{i,j=-p}^{p} \gamma_{ijk}^{(m)} \phi_{m+i}(t) \phi_{m+j}(x) \qquad [1.5]$$

where p and r are natural numbers, independent of m, and the constants $\gamma_{ijk}^{(m)}$ have a common bound, independent of m while the measurable functions $F_k(x, t)$ satisfy the condition

$$F_k(x, t) = O\left(\frac{1}{|t-x|}\right), \qquad [1.6]$$

where $O(\cdot)$ is uniform in k and $x \in E$.

We agree here that ϕ_{n+i} with eventually negative indices are defined to be identically equal to zero.

Since α_m/α_{m+1} is bounded uniformly in m (see Alexits, p. 28), it is clear that the polynomials (p_v) which are orthonormal with respect to a weight function ρ, form a polynomial-like system with $p=1$, $r=1$, $\gamma_{101}^{(m)} = \alpha_m/\alpha_{m+1}$, $\gamma_{011}^{(m)} = -\alpha_m/\alpha_{m+1}$, $\gamma_{001}^{(m)} = 0$ and $\gamma_{ijk}^{(m)} = 0$ for all other indices, $F_k(x, t) = (t-x)^{-1}$.

The trigonometric system is also polynomial-like for $x \in [-\pi+\delta, \pi-\delta]$, $\delta > 0$ (Alexits, pp. 178-179).

Theorems 1.1.1, 1.1.2, 1.1.3, 1.1.4 stated above can be applied to study questions of convergence of the expansions in terms of the functions of polynomial-like orthonormal systems. Following Alexits we will say that the orthonormal system (ϕ_v) is *constant-preserving* if ϕ_0 is constant.

Theorem 1.2.1 *Let $(\phi_v(x))$ be a complete orthonormal, constant-preserving, polynomial-like system with respect to a weight function ρ on a finite interval $[a, b]$. Let the functions $F_k(x, t)$ be continuous on the square $[a, b] \times [a, b]$, except possibly on the diagonal $t = x$. Assume that the sequence (ϕ_v) is dominated by a function w continuous on (a, b), a.e. positive on $[a, b]$ and such that ρw is integrable on $[a, b]$. Then the sequence of kernels (k_m, ρ^*) where*

$$k_m(x, t) = K_m(x, t)/w(t) = \sum_{v=0}^{m} \phi_v(x) \phi_v(t)/w(t)$$

$$\rho^*(t) = \rho(t) w(t),$$

satisfies the requirements of Definition 1.1.1 uniformly for $x \in [a+\epsilon, b-\epsilon]$, $\epsilon > 0$.

Proof. This theorem is a generalization of Theorem 4.3.1 in Alexits, pp. 264-266 and we follow closely Alexits' proof.

We have to show that the conditions of Definition 1.1.1 hold. Since (ϕ_v) is polynomial-like the m^{th} kernel has the form

$$K_m(x,t)=\sum_{k=1}^{r} F_k(x,t)\sum_{i,j=-p}^{p} \gamma_{ijk}^{(m)}\phi_{m+j}(x)\phi_{m+i}(t)$$

Since by Definition 1.2.1 the $\gamma_{ijk}^{(m)}$ are uniformly bounded, $F_k(x,t)$ is continuous on $[a,b]-[x-\lambda,x+\lambda]$, $\lambda>0$, and the ϕ_v are dominated by w, which is continuous on (a,b), it follows that

$$\sup\{\,|k_m(x,t)|:t\in[a,b]-[x-\lambda,x+\lambda]\,\}=O(1)$$

uniformly for $x\in[a+\epsilon,b-\epsilon]$, $\epsilon>0$, so that condition b) of Definition 1.1.1 holds. The rest of the proof goes exactly like Alexits' and it is omitted.

Definition 1.2.2 *Let (ϕ_v) be an orthonormal system of functions on some interval $[a,b]$ with respect to a weight function ρ. The Cesàro summation kernel of the system (ϕ_v) is defined by*

$$\tilde{k}_m(x,t)=\sum_{v=0}^{m}\left(1-\frac{v}{m}\right)\phi_v(x)\phi_v(t). \qquad [1.7]$$

The following theorem is useful to deal with the Cesàro kernels.

Theorem 1.2.2. *Let (ϕ_v) be a complete constant-preserving orthonormal system on an interval $[a,b]$ with respect to a weight function ρ and suppose that there exists a function w continuous on (a,b), positive a.e. on $[a,b]$ and such that $\sup_v \sup\{\,|\phi_v(t)|/w(t):t\in[a,b]\}<\infty$. Let us assume that (ϕ_v) is polynomial-like for $x\in[c,d]\subset[a,b]$, the functions $F_k(x,t)$ are continuous in the rectangle $[c,d]\times[a,b]$, except possibly on the line $t=x$. Let us assume also $0\le\rho(x)\le$ constant and $\sum_{v=0}^{m}\phi_v^2(x)=O(m)$ hold uniformly for $x\in[c,d]$. Let $\tilde{k}_m$ be the Cesàro kernel defined in [1.7]. Define*

$$k_m^*(x,t)=\tilde{k}_m(x,t)/w(t), \qquad \rho^*(t)=\rho(t)w(t). \qquad [1.8]$$

Then the sequence of kernels (k_m^*,ρ^*) satisfies the requirements of Definition 1.1.1 and Condition 1.1.2 uniformly for $x\in[c+\epsilon,d-\epsilon]$.

Proof. The kernel $\tilde{k}_m$ can be written as

$$k_m^*(x,t)=\frac{1}{m}\sum_{v=0}^{m-1}K_v(x,t) \qquad [1.9]$$

where K_υ is given by [1.5]. Therefore

$$k_m^*(x,t)=\frac{1}{m}\sum_{\upsilon=0}^{m-1} K_\upsilon(x,t)/w(t). \qquad [1.10]$$

Using [1.10] and arguing essentially as in Theorem 1.2.1 it follows that (k_m^*, ρ^*) is a kernel sequence in the sense of Definition 1.1.1. (See also Alexits, Theorem 4.3.2, pp. 267-268). From Alexits, p. 210 it follows that the kernel also satisfies Condition 1.1.2.

1.3. Some Properties of Kernels

In this paragraph we prove some elementary properties of kernels that will be needed later for the discussion of density estimators. Leadbetter (1963) proved analogous properties for the type of kernels he considers there.

Proposition 1.3.1. *Let ρ and w be a.e. positive functions on $[a,b]$ such that $\int_a^b \rho<\infty, \int_a^b \rho w<\infty$. Let $(k_m(x,t))$ be a sequence of Borel measurable functions from $[a,b]\times[a,b]$ to $\mathbb{R}$ such that for every $x\in[c,d]\subset[a,b]$, $\int_a^b k_m^2(x,t)\rho(t)dt<\infty$. Assume that $(k_m(x,t)/w(t), w(t)\rho(t))$ is a kernel sequence in the sense of Definition 1.1.1 satisfying a) and b) uniformly for $x\in[c,d]$. Define.*

$$\alpha_m(x)=\int_a^b k_m^2(x,t)\rho(t)dt. \qquad [1.11]$$

Then

$$\lim_{m\to\infty} \inf\{\alpha_m(x): a\leq x\leq b\}=\infty.$$

Proof. for notational convenience we set $\rho(t)dt=d\mu(t)$. Since k_m^2 is μ-integrable by Schwarz's inequality we have that for every $\lambda>0$

$$\Big[\int_{x-\lambda}^{x+\lambda}\frac{k_m(x,t)}{w(t)}w(t)d\mu(t)\Big]^2=\Big[\int_{x-\lambda}^{x+\lambda}k_m(x,t)d\mu(t)\Big]^2$$
$$\leq\Big[\int_{x-\lambda}^{x+\lambda}k_m^2(x,t)d\mu(t)\Big]\Big[\int_{x-\lambda}^{x+\lambda}d\mu(t)\Big].$$

Taking the infimum for $x \in [c, d]$ and then limit inferior, by condition a) of Definition 1.1.1 we have that the left-hand side becomes one, so that

$$1 \leq \int_{x-\lambda}^{x+\lambda} d\mu(t) \lim_{m\to\infty} \inf \inf \{ \int_a^b k_m^2(x, t) d\mu(t) : c \leq x \leq d \}.$$

Since ρ is an integrable function; this inequality is true for all $\lambda > 0$ if and only if $\lim \inf \inf \{ \int_a^b k_m^2(x, t) d\mu(t) : c \leq x \leq d \} = \infty$.

Proposition 1.3.2. *Let ρ and w be a.e. positive functions on $[a, b]$ $\int_a^b \rho < \infty$, $\int_a^b \rho w^2 < \infty$. Let $(k_m(x, t))$ be a sequence of Borel measurable functions from $[a, b] \times [a, b]$ to $\mathbb{R}$ such that for every $x \in [c, d] \subset [a, b]$, $\int_a^b k_m^2(x, t) \rho(t) dt < \infty$. Assume that $(k_m(x, t)/w(t), w(t), w(t)\rho(t))$ is a kernel sequence in the sense of Definition 1.1.1 satisfying a) and b) uniformly for $x \in [c, d]$. Let $\alpha_m(x)$ be defined by [1.11]. Define*

$$k_m^*(x, t) = \alpha_m^{-1}(x) k_m^2(x, t)/w^2(t) \qquad [1.12]$$

$$\rho^*(t) = w^2(t)\rho(t).$$

Then (k_m^*, ρ^*) is a non-negative kernel sequence satisfying uniformly a) and b) for $x \in [c, d]$, and

$$\int_a^b |k_m^*(x, t)| \rho^*(t) dt = \frac{1}{\alpha_m(x)} \int_a^b k_m^2(x, t) \rho(t) dt = 1. \qquad [1.13]$$

Proof. Let $J = [\alpha, \beta] - [x - \lambda, x + \lambda]$, $a \leq \alpha < \beta \leq b$. Then

$$|\int_J k_m^*(x, t) \rho^*(t) dt| = |\int_J \frac{k_m^2(x, t)}{\alpha_m(x) w^2(t)} \rho^*(t) dt|$$

$$\leq \alpha_m^{-1}(x) \sup \{ k_m^2(x, t)/w^2(t) : t \in J \} \int_a^b \rho^*(t) dt.$$

By Proposition 1.3.1 $\inf \{ \alpha_m(x) : c \leq x \leq d \} \to \infty$. By condition b) of Definition 1.1.1 $\sup \{ \sup \{ k_m^2(x, t)/w^2(t) : t \in J \} : c \leq x \leq d \}$ is finite. By hypothesis ρw^2 is integrable. Hence the right-hand side, and a fortiori, the left-hand side tends to zero uniformly for $x \in [c, d]$ as $m \to \infty$. That is, the kernel sequence

(k_m^*, ρ^*) satisfies the first part of condition a). It is obvious that [1.13] holds, and since the sequence (k_m^*, ρ^*) was just proved to satisfy the first part of condition a) it follows that it also satisfies its second part.

Condition b) follows from condition b) of Definition 1.1.1 and $\inf\{\alpha_m(x): c \leq x \leq d\} \to \infty$. Condition c) follows from condition c) of Definition 1.1.1 and $\inf\{\alpha_m(x): c \leq x \leq d\} \to \infty$.

Proposition 1.3.3. *Assume that all the hypotheses of Proposition 1.3.2 hold. Then as* $m \to \infty$

$$\int_a^b |k_m(x,t)|\,\rho(t)\,dt = o[(\alpha_m(x))^{1/2}]$$

where $o[\cdot]$ *is uniform for* $x \in [c, d]$.

Proof. Let $\lambda > 0$ arbitrary, $H = [a,b] - [x-\lambda, x+\lambda]$. Then,

$$\left|\frac{\int_a^b |k_m(x,t)|\,\rho(t)\,dt}{(\alpha_m(x))^{1/2}}\right| = \frac{\int_H [\,|k_m(x,t)|/w(t)]\,\rho(t)\,w(t)\,dt}{(\alpha_m(x))^{1/2}} + \frac{\int_{x-\lambda}^{x+\lambda} |k_m(x,t)|\,\rho(t)\,dt}{(\alpha_m(x))^{1/2}}$$

$$\leq \frac{\sup\{\,|k_m(x,t)|/w(t): t \in \mathrm{H}\}\int_a^b \rho w}{(\alpha_m(x))^{1/2}} + \frac{\int_{x-\lambda}^{x+\lambda} |k_m(x,t)|\,\rho(t)\,dt}{(\alpha_m(x))^{1/2}}. \qquad [1.14]$$

But, by Schwarz inequality we have

$$\int_{x-\lambda}^{x+\lambda} |k_m(x,t)|\,\rho(t)\,dt \leq \left[\int_{x-\lambda}^{x+\lambda} \rho(t)\,dt\right]^{1/2}\left[\int_a^b k_m^2(x,t)\,\rho(t)\,dt\right]^{1/2},$$

so that the second term in [1.14] can be made uniformly less than ϵ by choosing λ such that $\sup\{\int_{x-\lambda}^{x+\lambda} \rho : c \leq x \leq d\}/\inf\{\alpha_m(x) : c \leq x \leq d\} < \epsilon$, which is possible because ρ is integrable. The first term of [1.14] tends to zero uniformly for $x \in [c, d]$ because of condition b) of Definition 1.1.1 and Proposition 1.3.1.

Remark on Theorem 1.1.1. Since Condition 1.1.1 is necessary and sufficient for the converge of every function which is continuous at x and of bounded variation on $[a, b]$, it follows that Condition 1.1.1 is the minimum requirement that every kernel must satisfy if we want the corresponding density estimator $f_n(x)$ to be asymptotically unbiased for every p.d. function which is continuous at x.

1.4 Examples of Kernels

Several families of kernels have been proposed for estimating a p.d. function. Here we present three examples.

(1) Kernels given by [0.12]. Parzen (1962) requires that the even Borel-measurable function K satisfies

$$\int_{-\infty}^{\infty} K(u)\,du = 1\,,$$

$$\sup\{\,|K(u)| : -\infty < u < \infty\} < \infty\,,$$

$$\lim_{|u| \to \infty} |uK(u)| = 0\,,$$

$$\int_{-\infty}^{\infty} |K(u)\,du| < \infty\,.$$

(2) Kernels of the form $k_m(x, t) = \delta_m(x - t)$. Leadbetter (1963) requires that

$$\int_{-\infty}^{\infty} \delta_m(u)\,du = 1\,,$$

$$\text{for every } \lambda > 0,\ \sup\{\,|\delta_m(u)| : u \in \mathbb{R} - [-\lambda, \lambda]\} \to 0,$$

$$\text{for every } \lambda > 0, \int_{\mathbb{R} - [-\lambda, \lambda]} |\delta_m(u)|\,du \to 0\,,$$

$$\sup_m \int_{-\infty}^{\infty} |\delta_m(u)|\,du < \infty\,.$$

(3) Woodroofe's kernels given by [0.13] where G is a non-negative Borel-measurable and bounded function defined on $\mathbb{R}^2$ such that

$$\text{for every } x, \int_{-\infty}^{\infty} G(x, y)\,dy = 1,$$

$$\text{as } t \to \infty, \sup \{\int_{|y| \geq t} |y|\, G(x, y)\,dy : x \in \mathbb{R}\} \to 0.$$

The kernels defined by Definition 1.1.1 are defined on a finite square while those of Examples (1), (2) and (3) are defined on the whole $\mathbb{R}^2$ and therefore from this point of view the first kernels are less general than the other ones. But if we only consider kernels with compact support, the kernels of Examples (1), (2) and (3) are particular cases of kernels in the sense of Definition 1.1.1 which satisfy Condition 1.1.2.

Typical examples of kernels which do not satisfy Condition 1.1.2 are the m^{th} kernels of polynomial-like orthonormal systems (see Alexits, p. 179). Therefore, if we want to study a class of density estimators large enough to include the kernels of the trigonometric and Jacobi orthonormal systems we have to weaken Condition 1.1.2. This is the reason for considering kernels satisfying Condition 1.1.1.

2. LOCAL PROPERTIES OF THE DENSITY ESTIMATOR

2.1. Introduction

In what follows we will study density estimators defined by:

Definition 2.1.1. *Let ρ and w be functions defined on a closed interval $[a, b]$, such that ρ and w are strictly positive and continuous on (a, b), $\int_a^b \rho < \infty$ and $\int_a^b \rho w^2 < \infty$. Let $(k_m(x, t))$ be a sequence of Borel-measurable functions from $[a, b] \times [a, b]$ to $\mathbb{R}$ such that at $x \in [a, b]$, $\int_a^b k_m^2(x, t)\,\rho(t)\,dt < \infty$ and $(k_m(x, t)/w(t), w(t)\rho(t))$ is a kernel sequence in the sense of Definition 1.1.1. Let $(X_j)_{j=1}^n$ be i.i.d. r.v.'s with p.d. function f and distribution function F. Define*

$$f_n(x) = \frac{1}{n} \sum_{j=1}^{n} k_m(x, X_j)\,\rho(X_j)\,I_{[a,b]}(X_j) \qquad [2.1]$$

Where $I_{[a,b]}(t) = 1$ if $a \leq t \leq b$ and equal to zero otherwise.

In this Section we discuss the consistency of the estimator [2.1], we find its asymptotic variance and prove that its asymptotic distribution is Gaussian. Since the conditions imposed on the kernel sequence are very general it is not possible to obtain more informative and useful results than those of the present Section. For example at this stage, without further assumptions on k_m, it is not possible to obtain the rate at which the bias tends to zero. This kind of question is discussed in Section 3 for estimators derived from Jacobi series in general and for those derived from Legendre series in more detailed form.

2.2. Consistency and Asymptotic Variance

Theorem 2.2.1. *Let $f_n(x)$ be the estimator of $f(x)$ given by Definition 2.1.1. If either: (1) the p.d. function f is continuous at x, $w\rho$-integrable and of bounded variation in some interval around x, the kernel sequence satisfies Condition 1.1.1 of Theorem 1.1.1 and $m(n)\to\infty$ as $n\to\infty$; or (2) f is continuous at x, $w\rho$-integrable, the kernel sequence satisfies Condition 1.1.2 of Theorem 1.1.2 and $m(n)\to\infty$ as $n\to\infty$, then*

$$E[f_n(x)]\to f(x) \text{ as } n\to\infty. \tag{2.2}$$

Proof. We have

$$\begin{aligned} E[f_n(x)] &= \int_a^b k_m(x,t)f(t)\rho(t)\,dt \\ &= \int_a^b \frac{k_m(x,t)}{w(t)}(wf)(t)\rho(t)\,dt. \end{aligned}$$

From the hypotheses it follows that either Theeorem 1.1.4 or 1.1.2 apply and hence the right-hand side of [2.3] converges to $f(x)$.

Theorem 2.2.2. *Let $f_n(x)$ be the estimator of $f(x)$ given by Definition 2.1.1. If either: (1) the function f is continuous at x $w^2\rho^2$-integrable and of bounded variation in some interval around x, (k_m) satisfies the Condition 1.1.1 and $m(n)\to\infty$ as $n\to\infty$; or (2) f is continuous at x, $w^2\rho^2$-integrable, (k_m) satisfies the Condition 1.1.2 and $m(n)\to\infty$ as $n\to\infty$, then as $n\to\infty$*

$$\frac{n\,\mathrm{Var}[f_n(x)]}{\alpha_m(x)\rho(x)}\to f(x) \tag{2.4}$$

where $\alpha_m(x)=\int_a^b k_m^2(x,t)\rho(t)\,dt.$

Proof.

$$\operatorname{Var} f_n(x)=\frac{1}{n}\left[\int_a^b k_m^2(x,t)f(t)\rho^2(t)\,dt-\left(\int_a^b k_m(x,t)f(t)\rho(t)\,dt\right)^2\right]$$

so that

$$\frac{n}{\alpha_m(x)}\operatorname{Var} f_n(x)=\frac{1}{\alpha_m(x)}\int_a^b \frac{k_m^2(x,t)}{w^2(t)}(f\rho)(t)(w^2\rho)(t)\,dt \qquad [2.5]$$
$$-\frac{1}{\alpha_m(x)}\left[\int_a^b k_m(x,t)f(t)\rho(t)\,dt\right]^2.$$

Now $f\rho$ being $w^2\rho$-integrable applying Proposition 1.3.2 and Theorem 1.1.2 it follows that the first term converges to $f(x)\rho(x)$. The integral $\int_a^b k_m(x,t)\ f(t)\rho(t)\,dt$ converges to $f(x)$ because of Theorem 2.2.1, and therefore the second term in [2.5] tends to zero.

Indeed [2.4] holds under weaker conditions on the kernel than that of Theorem 2.2.2 as is shown in the following theorem where neither Condition 1.1.2 nor Condition 1.1.2 is required. The conditions on f are a little stronger since now the boundedness of f is required.

Theorem 2.2.3. *Let $f_n(x)$ be given by Definition 2.1.1. If f is bounded, continuous at x and $\int_a^b f\rho^2 w^2<\infty$, then as $n, m\to\infty$*

$$\frac{n\operatorname{Var}[f_n(x)]}{\alpha_m(x)\rho(x)}\to f(x).$$

Moreover for all n and m

$$\operatorname{Var}[f_n(x)]\leq\frac{\rho(x)\alpha_m(x)\,\|f|_{[a,b]}\|}{n}$$

where $\|f|_{[a,b]}\|=\sup\{|f(t)|:a\leq t\leq b\}$.

Proof. By Proposition 1.1.2 and Theorem 1.1.2 the first term in the right-hand side of [2.5] tends to $f(x)\rho(x)$, and the second term is bounded by

$$\frac{1}{\alpha_m(x)}[\|f|_{[a,b]}\|\int_a^b |k_m(x,t)|\rho(t)dt]^2$$

which tends to zero because f is bounded and Proposition 1.3.3 holds.

Theorem 2.2.4. *Let the hypotheses of Theorem 2.2.2 hold. If $m(n)$ is chosen such that $\alpha_m(x)=o_x(n)$, then $f_n(x)$ is a mean square consistent estimator of $f(x)$, i.e.,*

$$E[f_n(x)-f(x)]^2 \to 0 \text{ as } n \to \infty.$$

Proof. We have

$$E[f_n(x)-f(x)]^2 = \mathrm{Var}[f_n(x)]+[b(f_n(x))]^2$$

where $b(f_n(x)) \equiv Ef_n(x)-f(x)$ is the bias of $f_n(x)$.

Consequently from the fact that $\alpha_m(x)=o_x(n)$ and $n\,\mathrm{Var}\,f_n(x)/(\alpha_m(x)\,\rho(x)) \to f(x)$ it follows that $\mathrm{Var}\,f_n(x)$ converges to zero, and since by Theorem 2.2.1 $Ef_n(x)$ converges to $f(x)$ the bias $b(f_n(x))$ also converges to zero.

Remark 2.2.1. If the kernel sequence satisfies the conditions of Definition 1.1.1 uniformly on $[c,d]\subset[a,b]$, and either: (1) f is of bounded variation and continuous on $[c,d]$ and Condition 1.1.1 hols uniformly on $[c,d]$ or (2) f is continuous on $[c,d]$ and ρ-integrable and Condition 1.1.2 holds uniformly on $[c,d]$, then Theorems 2.2.1-2.2.3 hold uniformly for $x\in[c+\delta, d-\delta]$, $\delta>0$.

2.3. The Asymptotic Distribution of $f_n(x)$

Theorem 2.3.1. *Let the hypotheses of Theorem 2.2.2 hold. Assume that $f(x)>0$ and that there exists some function $q(x)$ and some $\gamma>0$ such that*

$$\sup\{|k_m(x,t)|\rho(t): a\leq t\leq b\}\leq q(x)[\alpha_m(x)]^{\gamma}. \qquad [2.6]$$

Then

$$\sqrt{n}[f_n(x)-Ef_n(x)]/\sqrt{\alpha_m(x)\rho(x)f(x)} \qquad [2.7]$$

is asymptotically normal with mean zero and variance one, for $n\to\infty$ and $m(n)\to\infty$ such that $\alpha_m(x)=o_x[n^{1/(2\gamma-1)}]$.

Proof. Since we can write [2.7] as

$$\frac{f_n(x)-Ef_n(x)}{\sqrt{n^{-1}\alpha_m(x)\rho(x)f(x)}}=\frac{f_n(x)-Ef_n(x)}{\sqrt{\mathrm{Var} f_n(x)}}\sqrt{\frac{\mathrm{Var} f_n(x)}{n^{-1}\alpha_m(x)\rho(x)f(x)}}, \quad [2.8]$$

where the factor $[n \ \mathrm{Var}\ f_n(x)/\alpha_m(x)\rho(x)f(x)]^{1/2}$ converges to one by Theorem 2.2.2, it will suffice to find the asymptotic distribution of $[f_n(x)-Ef_n(x)]/[\mathrm{Var}\ f_n(x)]^{1/2}$.

$$\frac{f_n(x)-Ef_n(x)}{\sqrt{\mathrm{Var}\ f_n(x)}}=\frac{\sum_{j=1}^{n} X_{nj}}{\sqrt{\mathrm{Var}\sum_{j=1}^{n} X_{nj}}} \quad [2.9]$$

Where

$$X_{nj}=[k_m(x,X_j)\rho(X_j)-E(k_m(x,X_j)\rho(X_j))] \quad [2.10]$$

Since by hypothesis [2.6] holds it follows that $|X_{nj}|\leq 2q(x)[\alpha_m(x)]^{\gamma}$. On the other hand

$$\sigma_n^2\equiv \mathrm{Var}\sum_{j=1}^{n} X_{nj}=n\,\mathrm{Var}[k_m(x,X)\rho(X)]\sim n\alpha_m(x)\rho(x)f(x).$$

Therefore choosing $m(n)$ such that $\alpha_m(x)=o(n^{1/(2\gamma-1)})$ we have that $|X_{nj}|=o(1)\sigma_n$. Hence given $\epsilon>0$ there exists $N(\epsilon)$ such that $|X_{nj}|\leq\epsilon\sigma_n$ for all $n\geq N(\epsilon)$ and the Lindeberg condition

$$\sum_{j=1}^{n}\int_{|t|\geq\epsilon\sigma_n} t^2\,dF_{nj}(t)\to 0 \text{ as } n\to\infty$$

holds.

Asymptotic normality was proven by Parzen (1962) for density estimators with kernels of the form [0.12], by Leadbetter (1963) for estimators with kernels of the form $\delta_m(x-t)$ and by Rosenblatt (1971) for estimators with kernels satisfying Condition 1.1.2 but under different conditions than those used here.

2.4. Examples

In order to illustrate the applications of the results of the present Section and show that the conditions imposed to the estimator are not too stringent we present here some estimators which will be studied in more detail in Section 3.

Let $X_1, ..., X_n$ be n independent observations of a random variable X with p.d. function f. Let (ϕ_v) be a complete system of orthonormal functions on an interval $[a, b]$ with respect to a weight function ρ. Define the m^{th} *kernel* and the *Cesàro kernel*, respectively, by

$$\hat{k}_m(x, t) = \sum_{v=0}^{m} \phi_v(x)\phi_v(t)\rho(t)I_{[a,b]}(t) \qquad [2.11]$$

$$\tilde{k}_m(x, t) = \sum_{v=0}^{m-1} \left(1 - \frac{v}{m}\right)\phi_v(x)\phi_v(t)\rho(t)I_{[a,b]}(t) \qquad [2.12]$$

and define

$$\hat{f}_n(x) = \frac{1}{n}\sum_{j=1}^{n} \hat{k}_m(x, X_j), \qquad \tilde{f}_n(x) = \frac{1}{n}\sum_{j=1}^{n} \tilde{k}_m(x, X_j).$$

In the case we take $\phi_0(x) = \dfrac{1}{\sqrt{2\pi}}$, $\phi_v(x) = \dfrac{1}{\sqrt{\pi}} \cos vx$, $v = 1, 2, ...,$

$\phi_v(x) = \dfrac{1}{\sqrt{\pi}} \sin vx$, $v = -1, -2, ...,$ $[-\pi \leq x \leq \pi]$, we obtain

$$\hat{k}_m(x, t) = \frac{1}{2\pi} \frac{\sin[(m+1/2)(x-t)]}{\sin[(x-t)/2]}$$

$$\tilde{k}_m(x, t) = \frac{1}{2\pi m}\left[\frac{\sin[m(x-t)/2]}{\sin[(x-t)/2]}\right]^2$$

where the sums in [2.11] and [2.12] were calculated for v running from $-m$ to m and from $-m+1$ to $m-1$ instead of from 0 to m and from 0 to $m-1$, respectively. It is easy to check that the system of trigonometric functions satisfies all the conditions of Theorems 1.2.1 and 1.2.2 and therefore $\tilde{k}_m$ satisfies Condition 1.1.2. The kernel $\hat{k}_m$ does not satisfy (1.1.2) but satisfies (1.1.1) (see Bèla Sz-Nagy [1965], pp. 402-404). On the other hand, as it is easy to show $\hat{\alpha}_m(x) = \Omega(m)$, $\tilde{\alpha}_m(x) = \Omega(m)$ where the symbol $b_m = \Omega(m)$ means that b_m/m is bounded away from 0 and ∞. It follows at once that $\sup\{|k_m(x, t)| : -\pi \leq t < \pi\} = O(m)$ holds for both $\hat{k}_m$ and $\tilde{k}_m$; therefore [2.6] holds with $q(x)$=constant and $\gamma = 1$. In conclusion, from the above discussion it follows that all the theorems proved in this Section apply for both estimators: $\hat{f}_n(x)$ and $\tilde{f}_n(x)$.

Let us consider now estimators derived from the system of Jacobi polynomials. Denote $p_v^{(\alpha,\beta)}$ the v^{th} normed Jacobi polynomial (see Paragraph 3.1 for a definition of these polynomials). The weight function of the system is $\rho(x)=(1-x)^\alpha(1+x)^\beta$. Let $\alpha'=\max\{\alpha, -1/2\}$, $\beta'=\max\{\beta,-1/2\}$. In Section 3 it is proved that the polynomials $p_v^{(\alpha,\beta)}(x)$ are dominated by a constant times $w(x)$, where

$$w(x)=(1-x)^{-\alpha'/2-1/4}(1+x)^{-\beta'/2-1/4},$$

and that the kernels $(\hat{k}_m(x,t)/w(t), (w\rho)(t))$ and $(\tilde{k}_m(x,t)/w(t), (w\rho)(t))$ are kernels in the sense of Definition 1.1.1. Theorem 1.2.2 guarantees Condition 1.1.2 for the kernel $(\tilde{k}_m/w, w\rho)$. In Section 3 we prove that Condition 1.1.1 holds for $(\hat{k}_m/w, w\rho)$ and that $\hat{\alpha}_m(x)\sim\pi^{-1}(1-x^2)^{-1/2}m$, $\tilde{\alpha}_m(x)\sim(3\pi)^{-1}(1-x^2)^{-1/2}m$. Moreover the following bound holds (see Szegö [1939], p. 164)

$$\max\{|p_v^{(\alpha,\beta)}(x)|:-1\leq x\leq 1\}\leq O(1)\,v^{c'}$$

where $c'=1/2+\max\{\alpha,\beta,-1/2\}$.

From the results just cited it follows at once that Theorem 2.2.1, Theorem 2.2.2 and Theorem 2.2.3 apply to both estimators $\hat{f}_n(x)$ and $\tilde{f}_n(x)$. Theorem 2.2.4 also applies to them and here the condition $\alpha_m(x)=o_x(n)$ means $m=o_x(n)$ because $\hat{\alpha}_m(x)=0_x(m)$ and $\tilde{\alpha}_m(x)=0_x(m)$.

From $\hat{\alpha}_m(x)=\Omega_x(m)$, $\tilde{\alpha}_m(x)=\Omega_x(m)$ and the bound for $p^{(\alpha,\beta)}$ given above it follows that condition [2.6] of Theorem 2.3.1 is satisfied with $q(x)=0(1)(1-x^2)^{-1/2}$ and $\gamma=3/2$, provided that $\min\{\alpha,\beta\}>0$, and therefore Theorem 3.2.1 applies to both estimators under the restriction $\min\{\alpha,\beta\}>0$.

Unfortunately condition [2.6] is too stringent for Jacobi polynomials in general, since it is not satisfied for $\min\{\alpha,\beta\}<0$, ruling out some important cases as for example that of Tchebychef polynomials of the first kind. Condition [2.6] is far from being the best possible and it would be desirable to have a better one but we do not pursue in this direction.

Finally, let us consider the algebraic estimators of Parzen defined in Paragraph 1.4 and assume that the kernel satisfies the conditions stated there and has a compact support. Then the kernel $b^{-1}(n)K\left(\frac{x-t}{b(n)}\right)$ is a kernel in the sense of Definition 1.1.1 which satisfies Condition (1.1.2) and [2.6] with $q=$constant and $\gamma=1$.

3. DENSITY ESTIMATORS DERIVED FROM EXPANSIONS IN JACOBI POLYNOMIALS

3.1. Introduction

In this Section we study the local properties of p.d. function estimators derived from orthonormal expansions of the p.d. function in terms of the Jacobi polynomials.

The Jacobi polynomials are orthogonal on $[-1, 1]$ with respect to the weight function

$$\rho(x) = (1-x)^{\alpha}(1+x)^{\beta},\ \alpha > -1,\ \beta > -1\,.$$

The *unnormed Jacobi polynomials* $P_m^{(\alpha,\beta)}$ are defined by the equation

$$(1-x)^{\alpha}(1+x)^{\beta}P_m^{(\alpha,\beta)}(x) = \frac{(-1)^m}{2^m m!}\left(\frac{d}{dx}\right)^m[(1-x)^{m+\alpha}(1+x)^{m+\beta}].$$

Some special cases of Jacobi polynomials are: (1) the Tchebychef polynomials of the first kind obtained for $\alpha=\beta=-1/2$, (2) the Tchebychef polynomials of the second kind obtained for $\alpha=\beta=1/2$, and (3) the Legendre polynomials obtained for $\alpha=\beta=0$.

We denote by $p_m^{(\alpha,\beta)}$ the *normed Jacobi polynomials*, that is those that satisfy the equations

$$\int_{-1}^{1} p_j^{(\alpha,\beta)}(x)p_k^{(\alpha,\beta)}(x)(1-x)^{\alpha}(1+x)^{\beta}dx = \delta_{jk}\,, \qquad [3.2]$$

where δ_{jk} is equal to one if $j=k$ and equal to zero otherwise. The relation between the normed and unnormed Jacobi polynomials is

$$p_m^{(\alpha,\beta)} = \left[\frac{2m+\alpha+\beta+1}{2^{\alpha+\beta+1}}\ \frac{\Gamma(m+1)\Gamma(m+\alpha+\beta+1)}{\Gamma(m+\alpha+1)\Gamma(m+\beta+1)}\right]^{1/2} P_m^{(\alpha,\beta)}, \qquad [3.3]$$

therefore

$$p_m^{(\alpha,\beta)}(x) = O(m^{1/2})\,P_m^{(\alpha,\beta)}(x). \qquad [3.4]$$

Define

$$\hat{k}_m(x,t) = \sum_{v=0}^{m} p_v^{(\alpha,\beta)}(x)\,p_v^{(\alpha,\beta)}(t) \qquad [3.5]$$

$$\tilde{k}_m(x,t)=\sum_{v=0}^{m-1}\left(1-\frac{v}{m}\right)p_v^{(\alpha,\beta)}(x)\,p_v^{(\alpha,\beta)}(t). \qquad [3.6]$$

From Szegö (1939), p. 70 it follows that

$$\hat{k}_m(x,t)=\frac{2^{-\alpha-\beta}}{2^{m+\alpha+\beta+2}}\;\frac{\Gamma(m+2)\,\Gamma(m+\alpha+\beta+2)}{\Gamma(m+\alpha+1)\,\Gamma(m+\beta+1)}\;\frac{P_m(x)\,P_{m+1}(t)-P_{m+1}(x)\,P_m(t)}{t-x}$$

where P_m means $P_m^{(\alpha,\beta)}$.

3.2. Technical Lemmas

Lemma 3.2.1.

$$P_m^{(\alpha,\beta)}(x)=(1-x)^{-\alpha'/2-1/4}(1+x)^{-\beta'/2-1/4}\,O(m^{-1/2}) \qquad [3.7]$$

$$-1<x<1,\ \alpha>-1,\ \beta>-1,$$

where $\alpha'=\max\{\alpha,-1/2\}$, $\beta'=\max\{\beta,-1/2\}$.

Proof. From Szegö (1939), p. 164 we have that

$$P_m^{(\alpha,\beta)}(\cos\theta)=\theta^{-\alpha-1/2}\,O(m^{-1/2}),\quad 0<\theta\leq\frac{\pi}{2},\ \alpha\geq-1/2 \qquad [3.8]$$

$$=O(m^{-1/2}),\qquad 0<\theta\leq\frac{\pi}{2},\ -1<\alpha<-1/2$$

Since $P_m^{(\alpha,\beta)}(x)=(-1)^m P_m^{(\alpha,\beta)}(-x)$ we also have

$$P_m^{(\alpha,\beta)}(\cos\theta)=(\pi-\theta)^{-\beta-1/2}\,O(m^{-1/2}),\ \frac{\pi}{2}\leq\theta<\pi,\ \beta\geq-1/2 \qquad [3.9]$$

$$=O(m^{-1/2}),\qquad \frac{\pi}{2}\leq\theta<\pi,\ -1<\beta<-1/2$$

It is clear that we can replace θ by $2\sin(\theta/2)$ in [3.8] and $(\pi-\theta)$ by $2\cos(\theta/2)$ in [3.9]. With these substitutions the lemma follows from [3.8] and [3.9].

Lemma 3.2.2. *Let $\hat{k}_m$ be defined by [3.5]. Then*

$$\rho(x)\int_{-1}^{1} \hat{k}_m^2(x,t)\rho(t)\,dt \sim \frac{1}{\pi\sqrt{1-x^2}}\,m,$$

where as usual $\sim$ means that the ratio of both quantities tends to one.

Proof. To keep the notation simple, we write p_m and P_m instead of $p_m^{(\alpha,\beta)}$ and $P_m^{(\alpha,\beta)}$, respectively. By Parseval's relation we have

$$\int_{-1}^{1} \hat{k}_m^2(x,t)\rho(t)\,dt = \sum_{\nu=0}^{m} p_\nu^2(x) = \hat{k}_m(x,x), \tag{3.10}$$

Calculating lim $\hat{k}_m(x,t)$ as $t \to x$ we obtain,

$$\hat{k}_m(x,x) = \frac{2^{-\alpha-\beta}}{2^{m+\alpha+\beta+2}}\,\frac{\Gamma(m+2)\,\Gamma(m+\alpha+\beta+2)}{\Gamma(m+\alpha+1)\,\Gamma(m+\beta+1)} \cdot [P'_{m+1}(x)\,P_m(x) - P'_m(x)\,P_{m+1}(x)]. \tag{3.11}$$

On the other hand (see Szegö, pp. 192, 230)

$$P_m(\cos\theta) = m^{-1/2}\,k(\theta)\cos(M\theta+\gamma) + 0(m^{-3/2})$$
$$\frac{dP_m(\cos\theta)}{d\theta} = m^{1/2}\,k(\theta)[-\sin(M\theta+\gamma) + 0(m^{-1})] \tag{3.12}$$

where $M = m+(\alpha+\beta+1)/2$, $\gamma = -\pi(\alpha+1/2)/2$ and

$$k(\theta) = \pi^{-1/2}(\sin\frac{\theta}{2})^{-\alpha-1/2}(\cos\frac{\theta}{2})^{-\beta-1/2}.$$

From [3.10], [3.11] and [3.12] the lemma follows.

Lemma 3.2.3. *Let $\tilde{k}_m$ be defined by [3.6]. Then*

$$\rho(x)\int_{-1}^{1} \tilde{k}_m^2(x,t)\rho(t)\,dt \sim \frac{1}{3}\,\frac{1}{\pi\sqrt{1-x^2}}\,m.$$

Proof. Define $\hat{\alpha}_m(x) = \int_{-1}^{1} \hat{k}_m^2(x,t)\rho(t)\,dt$. Then using [1.9] we can write

$$\int_{-1}^{1} \tilde{k}_m^2(x,t)\rho(t)\,dt = \int_{-1}^{1} \Big[\frac{1}{m}\sum_{v=0}^{m-1} \hat{k}_v(x,t)\Big]^2 \rho(t)\,dt$$

$$= \frac{1}{m^2}\Big[\sum_{v=0}^{m-1}\int_{-1}^{1} k_v^2(x,t)\rho(t)\,dt + 2\sum_{\mu=1}^{m-1}\sum_{v<\mu}\sum_{i=0}^{v} p_i^2(x)\Big\}$$

$$= \frac{1}{m^2}\Big[\sum_{v=0}^{m-1}\hat{\alpha}_v(x) + 2\sum_{\mu=1}^{m-1}\sum_{v<\mu}\hat{\alpha}_v(x)\Big\}. \qquad [3.13]$$

The lemma follows using Lemma 3.2.2 and Toeplitz lemma (Loève [1963], p. 238) in [3.13].

3.3. Convergence of Jacobi Series

In this section we show that the Jacobi polynomials satisfy the hypotheses of Theorems 1.2.1 and 1.2.2.

Define

$$w(x) = (1-x)^{-\alpha'/2-1/4}(1+x)^{-\beta'/2-1/4} \qquad [3.14]$$

where $\alpha' = \max\{\alpha, -1/2\}$, $\beta' = \max\{\beta, -1/2\}$. From Lemma 3.2.1 it follows that $p_m^{(\alpha,\beta)}$ is dominated by a constant times w. Without loss of generality we can take w as being the function w of Theorems 1.2.1 and 1.2.2. Since w is positive and continuous on (a, b) and the function.

$$\rho(x)w(x) = (1-x)^{\alpha-\alpha'/2-1/4}(1+x)^{\beta-\beta'/2-1/4} \qquad [3.15]$$

is integrable, the function w satisfies all the requirements of the theorems just cited. From Christoffel-Darboux formula it follows that the functions $F_k(x,t)$ of Definition 1.2.4 of $K_m(x,t)$ $(\equiv \hat{k}_m(x,t))$ are equal to $(x-t)^{-1}$ and therefore they satisfy the required conditions. The condition $\sum_{v=0}^{m} p_v^2(x) = O(m)$ for $x \in [-1+\epsilon, 1-\epsilon]$ follows from Alexits, p. 39. Therefore the kernels $(\hat{k}_m/w, w\rho)$ and $(k_m/w, w\rho)$ satisfy the conditions of Definition 1.1.1 uniformly in every inner subinterval of $[-1, 1]$. Moreover $(\tilde{k}_m/w, w\rho)$ satisfies uniformly (1.1.3). In order to guarantee convergence of the singular integrals with the kernel $\hat{k}_m$ we have to show that Condition (1.1.1) holds.

Lemma 3.3.1. *Let $(\hat{k}_m(x,t)/w(t),\ w(t)\rho(t))$ be the Jacobi kernel defined by [3.5]. Then Condition 1.1.1 holds uniformly for $x \in [-1+\epsilon, 1-\epsilon]$, $\epsilon > 0$.*

The present lemma is apparently a standard result, but no explicit proof could be found in the literature. It seems that most authors study the convergence of orthonormal expansions in series of Jacobi polynomials by reducing the problem to one of convergence of a Fourier series via a equiconvergence argument. In Appendix B we give a proof of the lema.

Remark. Since the convergence and Cesàro summability of orthonormal expansions in Jacobi polynomials is a well known matter (see for example Szegö) we could restrict ourselves to these results. The reason for adopting the above approach is intended to give a unified treatment of the problem of convergence of $f_n(x)$ to $f(x)$ in the mean square sense, which we hope, will help to better understand the behaviour of different density functions estimators.

It should be noted that the normalization in order to pass from the kernel (k_m, ρ) to the kernel $(k_m/w, w\rho)$ does not change the corresponding singular integral or the density estimator.

3.4. Density Estimators Derived from Jacobi Polynomials

Let f be a p.d. function defined on $\mathbb{R}$ whose support is not necessarily a finite interval. In order to be able to construct an estimator of the type discussed in this work we first have to decide in which interval we want to estimate f. Call $[a, b]$ to this interval. Without loss of generality we can take $a=-1$ and $b=1$ since we always can map $[a, b]$ onto $[-1, 1]$ by means of a linear transformation applied to r.v. X.

Let $\hat{k}_m$ and $\tilde{k}_m$ be the kernels defined by [3.5] and [3.6], respectively. Define

$$\hat{f}_n(x)=\frac{1}{n}\sum_{j=1}^{n}\hat{k}_m(x, X_j)\rho(X_j)I_{[-1,1]}(X_j), \qquad [3.16]$$

$$\tilde{f}_n(x)=\frac{1}{n}\sum_{j=1}^{n}\tilde{k}_m(x, X_j)\rho(X_j)I_{[-1,1]}(X_j), \qquad [3.17]$$

That $\hat{f}_n$ and $\tilde{f}_n$ satisfy the requirements of Definition 2.1.1 follows from the previous section and from $\int_a^b \rho w^2<\infty$. Therefore the following theorems are corollaries of the corresponding theorems of Section 2.

Theorem 3.4.1. *Let $\hat{f}_n(x)$ be defined by [3.16] and ρw be defined by [3.15]. If f is continuous at x, of bounded variation in a neighborhood of x and $\int_{-1}^{1} fw\rho<\infty$, then as n, $m\to\infty$ $\quad E\hat{f}_n(x)\to f(x)$.*

Proof. The conclusion follows from Theorem 2.2.1.

Theorem 3.4.2. *Let $\tilde{f}_n(x)$ be defined by [3.17] and ρw be defined by [3.15]. If f is continuous at x, and $\int_{-1}^{1} fw\rho<\infty$, then as $n, m\to\infty$, $E\tilde{f}_n(x)\to f(x)$.*

Proof. The conclusion follows from Theorem 2.2.1.

Theorem 3.4.3. *Let $\hat{f}_n(x)$ and $\tilde{f}_n(x)$ be defined by [3.16] and [3.17] respectively and ρw be defined by [3.15]. If f is bounded, continuous at x, and $\int_{-1}^{1} fw^2\rho^2<\infty$, then as $n, m\to\infty$*

$$\operatorname{Var}\hat{f}_n(x)\sim\frac{1}{\pi}\,\frac{f(x)}{\sqrt{1-x^2}}\,\frac{m(n)}{n}$$

$$\operatorname{Var}\tilde{f}_n(x)\sim\frac{1}{3\pi}\,\frac{f(x)}{\sqrt{1-x^2}}\,\frac{m(n)}{n}$$

where $a_n\sim b_n$ means $\lim(a_n/b_n)=1$.

Proof. The conclusion follows from Theorem 2.2.3 and lemmas 3.2.2 and 3.2.3.

Theorem 3.4.4. *Assume that the hypotheses of Theorem 3.4.1 holds and that $m(n)=o(n)$. Then $\hat{f}_n(x)$ and $\tilde{f}_n(x)$ are mean square consistent estimators of $f(x)$, i.e.*

$$E[f_n(x)-f(x)]^2\to 0$$

hold for both estimators.

Proof. The conclusion follows from Theorem 2.2.4, Lemma 3.2.2 and Lemma 3.2.3.

Theorem 3.4.5. *Assume that the hypotheses of Theorem 3.4.1 hold. Assume that $f(x)>0$ and $\min\{\alpha,\beta\}>0$. Define $c=3/2+\max\{\alpha,\beta,-1/2\}$. Choose $m(n)=O(n^{1/(2c-1)})$. Then as $n, m\to\infty$*

$$\sqrt{\pi}\,\sqrt{\frac{n}{m}}\,\frac{\sqrt[4]{1-x^2}}{\sqrt{f(n)}}[\hat{f}_n(x)-E\hat{f}_n(x)]\xrightarrow{L}N(0,1)$$

$$\sqrt{3\pi}\,\sqrt{\frac{n}{m}}\,\frac{\sqrt[4]{1-x^2}}{\sqrt{f(n)}}[\tilde{f}_n(x)-E\tilde{f}_n(x)]\xrightarrow{L}N(0,1).$$

Proof. The conclusion follows from Theorem 2.3.1, Lemma 3.2.2, Lemma 3.2.3 and the bound (see Szegö [1939], p. 164)

$$\max\{\,|p_m^{(\alpha,\beta)}(x)|: -1\leq x\leq 1\,\}\leq m^{c-1}.$$

Several remarks are in order.

Remark 3.4.1. The condition $\int_{-1}^{1} fw^2\rho^2<\infty$ looks rather restrictive at least for the case $\min\{\alpha,\beta\}<-1/2$, because in this case this condition implies that $f(x)$ is $o(1)$ as $x\to\pm 1$.

Remark 3.4.2. Theorem 3.4.3 shows that the asymptotic variance of $\hat{f}_n(x)$ and $\tilde{f}_n(x)$ does not depend on α and β provided, of course, that the hypotheses of this theorem hold. Therefore if we were only interested in density estimators derived from Jacobi orthonormal expansions, we should look at their bias to decide which is better since their variances are equal (asymptotically).

Remark 3.4.3. Theorem 3.4.3 shows that the variance of both estimators tends to be large for x near to the end points ± 1. Algebraic estimators do not have this unpleasant feature. Assume that f is continuous at x and let $f_n(x)$ be the algebraic estimator with kernel $b^{-1}(n)K[(x-t)/b(n)]$. If we assume that K has compact support then from Theorem 2.2.3 it follows that the asymptotic variance of $f_n(x)$ is $f(x)b(n)n\int K^2(t)dt$ and there is no instability of the variance at any point.

A way of meeting this difficulty with polynomials estimators is to choose the original interval $[a,b]$ larger than the interval in which we want to estimate f.

3.5. Rate of Convergence of the Bias and the Mean Square Error to Zero

Let $s_m(f,x)$ and $\sigma_m(f,x)$ be the m^{th} partial sum and the m^{th} Cesàro sum corresponding to the orthonormal expansion of the function f in Jacóbi polynomials. Then the biases of $\hat{f}_n(x)$ and $\tilde{f}_n(x)$ can be written as

$$b(\hat{f}_n(x))=s_m(f,x)-f(x)$$

$$b(\tilde{f}_n(x))=\sigma_m(f,x)-f(x)$$

Therefore we can use known results about orthogonal expansions to estimate the bias of $\hat{f}_n(x)$ and $\tilde{f}_n(x)$.

For the Jacobi expansions we state the following result.

Proposition 3.5.1. *Assume that the p.d. function f has a continuous k^{th} derivative $f^{(k)}$. Let $\omega(\delta, f^{(k)}) = \sup\{ |f^{(k)}(x) - f^{(k)}(y)| : |x-y| \leq \delta, x, y \in [-1,1]\}$ be the modulus of continuity of $f^{(k)}$. Then*

$$|s_m(f,x) - f(x)| \leq K \frac{\log m}{m^k} \omega\left(\frac{1}{m}, f^{(k)}\right)$$

For a proof of this statement see for instance Alexits, pp. 287, 308, Lorentz (1965), Timan (1963).

We do not try to find a bound for the bias of $\tilde{f}_n(x)$. In the next paragraph we prove a negative result about $\tilde{f}_n(x)$.

The rest of this paragraph is devoted to the special case of Legendre polynomials. In what follows we find a more precise bound than that of Proposition 3.5.1, for the estimator $\hat{f}_n(x)$ derived from Legendre polynomials.

Let $\hat{f}_n(x)$ be the p.d. function estimator defined in [3.16] for the case $\alpha = \beta = 0$, that is,

$$\hat{f}_n(x) = \frac{1}{n} \sum_{j=1}^{n} \sum_{v=0}^{m} p_v(x) p_v(X_j) I_{[-1,1]}(X_j) \qquad [3.18]$$

where $p_v, v = 0, 1, \ldots$ are the normed Legendre polynomials.

Theorem 3.5.1. *If on $[-1,1]$ the p.d. function is absolutely continuous with an absolutely continuous derivative whose derivative f'' is of bounded variation, then,*

$$|E\hat{f}_n(x) - f(x)| < \frac{8}{3\sqrt{\pi}} V'' \frac{1}{m^{3/2}} \quad \text{for } |x| \leq 1,\ m \geq 2$$

$$|E\hat{f}_n(x) - f(x)| < \frac{8\sqrt{2}}{\pi} V'' \frac{1}{\sqrt[4]{1-\delta^2}} \frac{1}{m^2} \quad \text{for } |x| \leq \delta < 1, m \geq 2$$

where $V''[-1,1] \equiv V''$ is the total variation of $f''|_{[-1,1]}$, where $f''|_{[-1,1]}$, is the second derivative of the restriction of f to $[-1,1]$.

Proof.

$$f(x) - E\hat{f}_n(x) = \sum_{v=m+1}^{\infty} c_v p_v(x) = \sum_{v=m+1}^{\infty} a_v P_v(x)$$

where $a_v = [(2v+1)/2] \int_{-1}^{1} f(t) P_v dt$, and $P_v = 0, 1, \ldots$ are the unnormed Legendre polynomials. Since $(2v+1)\,P_v=P'_{v+1}-P'_{v-1}$ (see Sansone [1959], p. 178) and since $P_v(1)=1$, $P_v(-1)=(-1)^v$ (see Sansone [1959], p. 180), integrating by parts we obtain

$$a_v=\frac{1}{2}\int_{-1}^{1} f(x)(2v+1)\,P_v(x)\,dx$$

$$=\frac{1}{2}[f(x)(P_{v+1}(x)-P_{v-1}(x))\,\Big|_{-1}^{1}-\int_{-1}^{1}(P_{v+1}(x)-P_{v-1}(x))f'(x)\,dx]$$

$$=\frac{1}{2}\int_{-1}^{1} P_{v-1}(x)f'(x)\,dx-\frac{1}{2}\int_{-1}^{1} P_{v+1}(x)f'(x)\,dx. \qquad [3.20]$$

Note that we have obtained the recurrence formula

$$\int_{-1}^{1} f(x)(2v+1)\,P_v(x)\,dx \qquad [3.21]$$

$$=\int_{-1}^{1} P_{v-1}(x)\;f'(x)\,dx-\int_{-1}^{1} P_{v+1}\;f'(x)\,dx.$$

Applying [3.21] to [3.20] we have

$$a_v=\frac{1}{2}\Big[\frac{1}{2v-1}\Big(\int_{-1}^{1} P_{v-2}\;f''-\int_{-1}^{1} P_v\;f''\Big) \qquad [3.22]$$

$$-\frac{1}{2v+3}\Big(\int_{-1}^{1} P_v\;f''-\int_{-1}^{1} P_{v+2}\;f''\Big).$$

Now, by the second theorem of the mean value for $m\geq 1$

$$\int_{-1}^{1} P_v(x)\;f''(x)\,dx=V''\int_{c}^{d} P_v(x)\,dx \qquad [3.23]$$

and from Sansone (1959), p. 200,

$$\left| \int_{-1}^{t} P_v(x)\,dx \right| < \frac{4}{\sqrt{\pi}\;\sqrt{v+1}\,(2v+1)}. \qquad [3.24]$$

Hence from [3.22], [3.23] and [3.24] it follows that

$$a_v < V'' \frac{4}{\sqrt{\pi}} \left\{ \frac{1}{2v-1} \left[\frac{1}{\sqrt{v-1}\;(2v-3)} + \frac{1}{\sqrt{v+1}\;(2v+1)} \right] \right.$$

$$\left. + \frac{1}{2v+3} \left[\frac{1}{\sqrt{v+1}\;(2v+1)} + \frac{1}{\sqrt{v+3}\;(2v+5)} \right] \right\}.$$

Since for $v \geq 2$ the quantity inside the braces is less or equal to $v^{-5/2}$ we have that

$$a_v < \frac{4\,V''}{\sqrt{\pi}\;v^{5/2}}. \qquad [3.25]$$

Since $|P_v(x)| \leq 1$ for $|x| \leq 1$ from [3.25] it follows that

$$\left| \sum_{v=m+1}^{\infty} a_v P_v \right| \leq \sum_{v=m+1}^{\infty} |a_v P_v| < \frac{4}{\sqrt{\pi}} V'' \frac{2}{3} \frac{1}{m^{3/2}}$$

for $|x| \leq 1$, $m \geq 2$. Since $|P_v(x)| \leq 4\sqrt{2/\pi}\;\dfrac{1}{\sqrt{v}\;\sqrt[4]{1-x^2}}$, for

$|x| \leq \delta < 1$ (see Sansone, p. 198) it follows that

$$\left| \sum_{v=m+1}^{\infty} a_v P_v(x) \right| < \sum_{v=m+1}^{\infty} |a_v P_v(x)| < \frac{4}{\sqrt{\pi}} V'' \frac{4\sqrt{2}}{\sqrt{\pi}} \frac{1}{\sqrt[4]{1-\delta^2}} \frac{1}{2m^2}$$

$$\left| \sum_{v=m+1}^{\infty} a_v P_v(x) \right| < \frac{8\sqrt{2}}{\pi} V'' \frac{1}{\sqrt[4]{1-\delta^2}} \frac{1}{m^2},$$

for $|x| \leq \delta < 1$, $m \geq 2$, and this completes the proof of the theorem.

For the proof of this theorem we have followed the proof of a similar theorem due to Jackson (see Sansone, p. 205). Using the same approach it is easy to prove that if f has a $(r-1)^{th}$ derivative which is absolutely continuous and its r^{th} derivative is of bounded variation, then

$$|E\hat{f}_n(x)-f(x)| = \frac{V^{(k)}}{\sqrt[4]{1-\delta^2}}\, O(m^{-r}) \text{ for } |x|\leq\delta<1$$

where $O(\cdot)$ is uniformly for $|x|\leq\delta<1$ and $V^{(k)}$ is the total variation of $f^{(k)}|_{[-1,1]}$.

We want to point out that in the formula for the asymptotic variance and in the square of the second bound of Theorem 3.5.1 the same factor $(1-x^2)^{-1/2}$ appears. Of course we can't conclude that the bias behaves as a function of x in the same way as the corresponding bounds. However, we feel that where their order in m is concerned the bounds are the best possible.

Let us assume that the hypotheses of Theorem 3.5.1 hold, i.e. on $[-1,1]$ the p.d. function f is absolutely continuous with an absolutely continuous derivative whose derivative f'' is of bounded variation. Then from Lemma 3.4.2 and Theorem 3.5.1 it follows that $E(\hat{f}_n(x)-f(x))^2$ is asymptotically dominated by

$$A\frac{m}{n} + B'\frac{1}{m^4} \qquad [3.25]$$

where

$$A=\frac{f(x)}{\pi\sqrt{1-x^2}}, \qquad B=\frac{128}{\pi^2}\;\frac{V''^2}{\sqrt{1-x^2}}. \qquad [3.27]$$

Then it follows from straightforward calculations (see Parzen [1962]) that [3.26] is minimized by taking

$$m=m(n)=\left[\frac{4B}{A}\,n\right]^{1/5}, \qquad [3.28]$$

and the corresponding minimum is

$$\frac{5}{4}\sqrt[5]{4}\;A^{4/5}B^{1/5}n^{-4/5}. \qquad [3.29]$$

From [3.26], [3.27] and [3.28] it follows that the mean square error of $\hat{f}_n(x)$ is asymptotically dominated by

$$\frac{5}{4}\ [f(x)]^{4/5}(V'')^{1/5}\ \frac{1}{\sqrt{1-x^2}}n^{-4/5}. \qquad [3.30]$$

The expressions [3.28] and [3.30] are useful for choosing a right m. Since we are trying to estimate $f(x)$, V'' and $f(x)$ as a rule will be unknown. Given a sample $(X_j)_{j=1}^n$ of the r.v. X we have to choose m. If we have a rough idea of the order of magnitude of $f(x)$ and V'' we should use this information to choose m. Observe that the expression [3.30] is rather insensible to variation in V'' since it appears to the power 1/5. This fact makes the choice of m a little easier.

3.6. Uniform Rate of Mean Square Consistency over Classes of Density Functions

Definition 3.6.1. *For $k=1,2,...$ and $\gamma>0$ define the class $A_{k\gamma}$ as being the class of p.d. functions f such that their restrictions to $[-1,1]$ satisfy the following conditions a) and b):*

a) The $(k-1)^{st}$ derivative $f^{(k-1)}|_{[-1,1]}$ is absolutely continuous.

b) The k^{th} derivative $f^{(k)}|_{[-1,1]}$ exists everywhere and satisfies the condition $V^{(k)}[-1,1]\leq\gamma$ where $V^{(k)}[-1,1]$ is the total variation of $f^{(k)}|_{[-1,1]}$.

Theorem 3.6.1. *Let $f\in A_{k\gamma}$, $k=1,2,\ldots$. Let $m=m(n)=Kn^{1/(2k+1)}$, K constant and let $\hat{f}_n(x)$ be the estimator given by [3.16] constructed using the Legendre polynomials. Then for $x\in(-1,1)$,*

$$\sup_{f\in A_{k\gamma}} E_f[\hat{f}_n(x)-f(x)]^2=0(n^{-2k/(2k+1)}).$$

Proof. Arguing as in Theorem 2.2.2 it follows that Var $\hat{f}_n(x)\leq \|f|_{[-1,1]}\|\ \rho(x)\hat{\alpha}_m(x)n^{-1}$. Since $\hat{\alpha}_m(x)=0_x(m)$ where $0_x(m)$ does not depend on f we have that

$$\text{Var}\ \hat{f}_n(x)\leq\|f|_{[-1,1]}\|\ \rho(x)n^{-1}\ O_x(m). \qquad [3.31]$$

From the remark after Theorem 3.5.1 it follows that

$$|\ E\hat{f}_n(x)-f(x)\,|\leq O_x\ V^{(k)}[-1,1]\,O(m^{-k})\leq O_x\,\gamma\ O(m^{-k}). \qquad [3.32]$$

Since by the Lemma of Appendix C $\sup\{\| f|_{[-1,1]}\| : f\in A_{k\gamma}\}$ is finite, the conclusion of the theorem follows from [3.31] and [3.32].

Theorem 3.6.2. *Let $f\in A_{k\gamma}$, $k=1,2,\ldots$. Assume that $m=m(n)=$ $=Kn^{1/(2+1)}$, K constant and let $\hat{f}_n(x)$ be the estimator given by [3.16] constructed using the Jacobi polynomials. Then for $x\in(-1,1)$,*

$$\sup_{f\in A_{k\gamma}} E_f[\hat{f}_n(x)-f(x)]^2 = O(n^{-2k/(2k+1)}\log n).$$

The proof is argued like that of Theorem 3.6.1.

Farrell (1972) has found the best possible rate of consistency in probability that can be attained uniformly over certain classes of p.d. functions. Wahba (1975) using Farrell's approach has found the best possible rate of mean square consistency that can be attained uniformly over the classes $C_{k\gamma}^{(p)}$ defined below.

Definition 3.6.2. *Let $1\leq p\leq\infty$. Define $\|f|_{[-1,1]}\|_p=[\int_{-1}^{1}(f)^p]^{1/p}$ if $1\leq p<\infty$, and $\|f|_{[-1,1]}\|_\infty\equiv\|f|_{[-1,1]}\|=\sup\{|f(t)|:-1\leq t\leq 1\}$. For $k=1$, $2,\ldots$ and $\gamma>0$ define the class $C_{k\gamma}^{(p)}$ as being the class of p.d. functions f satisfying the conditions:*

a) For $r=0,1,\ldots,k-1$ the derivative $f^{(r)}$ is absolutely continuous on $[-1,1]$ and $\|f^{(r)}|_{[-1,1]}\|_p$ is finite.

b) The derivative $f^{(k)}$ exists everywhere on $[-1,1]$ and satisfies the condition $\|f^{(k)}|_{[-1,1]}\|_p\leq\gamma$.

Theorem 3.6.3. (Wahba). *Let $(\delta_n(x))$ be the class of all density estimators of $f(x)$ at the point x. Define $\lambda(m,p)=(2m-2/p)\,(2m+1-2/p)$. Then, for every $\epsilon>0$.*

$$\sup_{f\in C_{k\gamma}^{(p)}} E_f[\delta_n(x)-f(x)]^2 = b_n n^{-\lambda(m,p+\epsilon)}$$

implies that there exists D_o such that $b_n>D_o$ for infinitely many n.

We do not know the best possible rate for the classes $A_{k\gamma}$. However the following inclusion relation gives a rather precise bound for such best rates.

Proposition 3.6.1. *For every class $A_{k\gamma}$ there exists $\alpha<\infty$ such that*

$$C_{k+1,\frac{\gamma}{2}}^{(\infty)} \subset A_{k\gamma} \subset C_{k\alpha}^{(\infty)}.$$

Proof. The first inclusion follows from $V^{(k)} = \int_{-1}^{1} |f^{(k+1)}| \leq 2 \|f^{(k+1)}\|_\infty$.

Using the same argument as in Appendix C it follows that there exists α such that $\sup \{ \|f^{(k)}\|_\infty : f \in A_{k\gamma}\} \leq \alpha$ and then the second inclusion holds.

Assume that $f \in A_{k\gamma}$. Then Theorem 3.6.1 guarantees for the Legendre estimator $\hat{f}_n(x)$ the rate $n^{-2k/(2k+1)}$ which is the best possible rate for the class $C_{k\alpha}^{(\infty)}$. Therefore in view of Proposition 3.6.1 the rate of $\hat{f}_n(x)$ if not the best it is near to it. We mention here that with algebraic estimators we can attain the best possible rate over $C_{k\alpha}^{(\infty)}$. For the Cesàro estimator $\tilde{f}_n(x)$ we prove the following negative result.

Let f be the p.d. function defined by

$$f(x) = \begin{cases} \dfrac{1}{2} + \dfrac{1}{\sqrt{6}}\, p_1(x), & -1 \leq x \leq 1 \\ 0 & \text{, otherwise} \end{cases}$$

where $p_1(x) = \sqrt{3/2}\, x$ is the Legendre polynomial of degree one. Then

$$E[\tilde{f}_n(x)] = f(x) - \frac{1}{m}\, p_1(x).$$

Therefore the mean square error tends to zero no faster than $n^{-2/3}$ despite $f \in A_{k,\gamma(k)}$ for some $\gamma(k)$, all k. Viollaz (1976) have proved that the same happens for the Cesàro estimators constructed from trigonometric orthonormal expansions. We think that for the estimator $\tilde{f}_n(x)$ in the Jacobi case, it is possible to prove, following the same approach as that used in Viollaz that under appropriate conditions on f the mean square error tends to zero at the rate $n^{-2/3}$. We have not checked this guess.

We want to emphasize a result of the previous discussion. For the estimator $\hat{f}_n(x)$ the smoother the function f the faster the corresponding mean square error tends to zero, while for the estimator $\tilde{f}_n(x)$ the rate of convergence is not faster than $n^{-2/3}$ no matter how smooth the function f is.

APPENDIX A

PROOF OF CONDITION 1.1.2 FOR JACOBI POLYNOMIALS

Let $p_m^{(\alpha,\beta)}$ and $P_m^{(\alpha,\beta)}$ be the normed and unnormed Jacobi polynomials respectively, which we will denote also by p_m and P_m respectively. The following results about Jacobi polynomials are needed in the sequel (see Szegö):

(i) $\sup_v \sqrt{v} \sup\{ | P_v^{(\alpha,\beta)}(x) | : -1+\delta \leq x \leq 1-\delta\} < \infty$

(ii) $p_m^{(\alpha,\beta)}(x) = O(m^{1/2}) P_m(x)$

(iii) $P_m^{(\alpha,\beta)}(\cos\theta) = P_m(\cos\theta) = m^{-1/2} k(\theta) \cos(M_m\theta + \gamma) + 0(m^{-3/2})$

where

$$k(\theta) = \pi^{-\frac{1}{2}} (\sin\frac{\theta}{2})^{-\alpha-\frac{1}{2}} (\cos\frac{\theta}{2})^{-\beta-\frac{1}{2}}$$

$$M_m = m + \frac{\alpha+\beta+1}{2}, \qquad \gamma = -\frac{(\alpha+\frac{1}{2})\pi}{2}, \quad 0<\theta<\pi$$

and the bound $O(m^{-3/2})$ for the error term holds uniformly in the interval $[\epsilon, \pi-\epsilon]$.

Lemma B. *Let $\hat{k}_m(x, t)$ be the Jacobi kernel given by [3.1.5]. Then Condition 1.1.2 holds uniformly for $x \in [-1+\delta, 1-\delta]$, $\delta > 0$.*

Proof. Let $-1 \leq A < B \leq 1$. Since

$$\int_A^B \hat{k}_m(x, t)\rho(t)\,dt = \int_A^x \hat{k}_m(x, t)\rho(t)\,dt + \int_x^B \hat{k}_m(x, t)\rho(t)\,dt$$

we will consider only $\int_x^B$ because the same arguments apply for $\int_A^x$.

Let ϵ be a number which will be fixed later. Then

$$\int_x^B \hat{k}_m(x, t)\rho(t)\,dt = [\int_x^{x+\frac{1}{m}} + \int_{x+\frac{1}{m}}^{x+\epsilon} + \int_{x+\epsilon}^B] \hat{k}_m(x, t)\rho(t)\,dt$$

$$= I_1 + I_2 + I_3. \qquad [1]$$

In what follows we assume that $-1+\delta \leq x \leq 1-\delta$ and without loss of generality we assume that $-1+\delta \leq x + \frac{1}{m} \leq 1 - \frac{\delta}{2}$. Because of conditions (i) and (ii) it follows that

$$|I_1| \leq \int_x^{x+\frac{1}{m}} \sum_{v=0}^{m} |p_v(x)p_v(t)|\, \rho(t)dt = O(1) \qquad [2]$$

where $O(1)$ is uniform for $-1+\delta \leq x \leq 1-\delta$. From the expression for $\hat{k}_m(x,t)$ given in Section 3.1 it follows that

$$\hat{k}_m(x,t) = O(m)\,\frac{P_m(x)\,P_{m+1}(t) - P_{m+1}(x)\,P_m(t)}{t-x}. \qquad [3]$$

Define $x = \cos\theta$, $t = \cos\phi$. Using (i), (ii) and (iii) it follows that

$$\hat{k}_m(x,t) = \frac{O(1)\,k(\phi)\cos[M_{m+1}\phi+\gamma] - O(1)\,k(\phi)\cos[M_m\phi+\gamma] + O(m^{-1})}{\cos\phi - \cos\theta}.$$

Define $\cos\theta_m = x + \frac{1}{m}$, $\cos\eta = x+\epsilon$. Then

$$I_2 = \int_{x+\frac{1}{m}}^{x+\epsilon} \hat{k}_m(x,t)\rho(t)dt = O(1)\int_\eta^{\theta_m} \frac{(\sin\frac{\phi}{2})^{\alpha+\frac{1}{2}}(\cos\frac{\phi}{2})^{\beta+\frac{1}{2}}(\cos[M_{m+1}\phi+\gamma]}{\cos\phi - \cos\theta}\,d\phi$$

$$-\,O(1)\int_\eta^{\theta_m} \frac{(\sin\frac{\phi}{2})^{\alpha+\frac{1}{2}}(\cos\frac{\phi}{2})^{\beta+\frac{1}{2}}(\cos[M_m\phi+\gamma]}{\cos\phi - \cos\theta}\,d\phi$$

$$+\,O(m^{-1})\int_{x+\frac{1}{m}}^{x+\epsilon} \frac{dt}{t-x}$$

$$= I_{21} + I_{22} + I_{23}. \qquad [4]$$

Define

$$g(\phi) = \frac{(\sin\frac{\phi}{2})^{\alpha+\frac{1}{2}}(\cos\frac{\phi}{2})^{\beta+\frac{1}{2}}}{\cos\phi - \cos\theta}.$$

For ϵ small enough we have that g is monotone in the interval $[\eta, \theta]$ and the maximum of $|g|$ in the interval $[\eta, \theta_m]$ is attained for $\phi = \theta_m$. Note that since g does not depend on m we can assume that $\eta < \theta_m < \theta$.

Let $[\frac{\theta_m-\eta}{\pi} M_{m+1}]$ be the integer part of $\frac{\theta_m-\eta}{\pi} M_{m+1}$. The function $\cos(M_{m+1}\phi+\gamma)$ has $N=[\frac{\theta_m-\eta}{\pi} M_{m+1}]+1$ zeros in the interval $[\eta,\theta_m]$. Let us denote them by $\phi_1<\phi_2<\ldots<\phi_N$. Then

$$I_{21}=\int_\eta^{\theta_m} g(\phi)\cos(M_{m+1}\phi+\gamma)\,d\phi$$

$$=[\int_{\phi_N}^{\phi_m}+\int_{\phi_{N-1}}^{\phi_N}+\ldots+\int_\eta^{\theta_1}]g(\phi)\cos(M_{m+1}\phi+\gamma)\,d\phi.$$

Except eventually for the first term this sum has terms with alternating signs and monotone decreasing absolute values. Hence the whole sum is less than or equal to the sum of the absolute values of the first two terms, i.e.

$$|I_{21}|\leq\int_{\phi_{N-1}}^{\theta_m}|g(\phi)\cos[M_{m+1}\phi+\gamma]|\,d\phi, \qquad [5]$$

Define $t_{N-1}=\cos\phi_{N-1}$. Since

$$-1+\delta\leq x<x+\frac{1}{m}\leq t_{N-1}\leq x+\epsilon<1-\frac{\delta}{2}$$

it follows that

$$\text{Arccos}(-1+\delta)\geq\theta>\theta_m\geq\phi_{N-1}\geq\eta>\text{Arccos}(1-\frac{\delta}{2}).$$

Therefore

$$\sup\{\left|(\sin\frac{\phi}{2})^{\alpha+\frac{1}{2}}(\cos\frac{\phi}{2})^{\beta+\frac{1}{2}}\right|:\phi_{N-1}\leq\phi\leq\theta_m\}\leq K_1 \qquad [6]$$

where $K_1<\infty$ is independent of x and m.

On ther other hand

$$\min_x\{\min\{|\cos\theta-\cos\phi|:\phi_{N-1}\leq\phi\leq\theta_m\}\}=\min_x\{\cos\theta-\cos\theta_m\}$$

$$\geq 1-\cos(1-\frac{1}{m}) \qquad [7]$$

$$\geq\frac{K_2}{m}$$

where $K_2>0$ is independent of m. From [5]−[7] it follows that

$$|I_{21}|\leq\frac{K_1}{K_2}\ m\ \frac{2\pi}{M_{m+1}}.$$

Since $M_{m+1}=m+(\alpha+\beta+1)/2$ it follows that I_{21} is uniformly bounded with respect to $x\in[-1+\delta,1-\delta]$ and m. Since the same arguments apply to I_{22} it follows that I_{22} is uniformly bounded with respect to $x\in[-1+\delta,1-\delta]$ and m.

For I_{23} we have

$$|I_{23}|=O(m^{-1})\left|\int_{x+\frac{1}{m}}^{x+\epsilon}\frac{dt}{t-x}\right|=O(m^{-1})\,|\log\epsilon+\log m\,|=O(1).$$

Therefore I_2 is uniformly bounded with respect to $x\in[-1+\delta,\ 1-\delta]$ and m.

For I_3 we have

$$|I_3|\leq\int_{x+\epsilon}^{B}|\hat{k}_m(x,t)|\,\rho(t)dt\,|\leq\int_{x+\epsilon}^{1}\frac{O(1)\,w(t)}{|t-x|}\ \rho(t)dt$$

$$\leq\frac{1}{\epsilon}\,O(1)\int_{x+\epsilon}^{1}w(t)\rho(t)dt$$

where $w(t)=(1-x)^{-\alpha'/2-1/4}(1+x)^{-\beta'/2-1/4}$. Since for $x\in[-1+\delta,\ 1-\delta]$ ϵ can chosen independent of x (see the definition of ϵ after formula [4]) we have that the above bound is $O(1)$ uniformly for x in that interval. This completes the proof of the lemma.

APPENDIX B

A bound for the classes $C_{k\gamma}^{(\infty)}$

Proposition C. *Let $C_{k\gamma}\equiv C_{k\gamma}^{(\infty)}$ be the classes defined by Definition 3.6.2. Then, for every $C_{k\gamma}$ there exists $L(\gamma,k)<\infty$ such that $\sup\{f: f\in C_{k\gamma}\}\leq$ $\leq L(\gamma,k)$.*

Proof. Without loss of generality we assume that $[a,b]=[0,1]$. Integrating k times we obtain.

$$f(x)=\sum_{r=0}^{k-1} x^r \frac{f^{(r)}(0)}{r!} + \frac{x}{(k-1)!}\int_0^1 (1-\lambda)^{k-1} f^{(k)}(\lambda x)\, d\lambda . \qquad [1]$$

Integrating again we obtain

$$F(x)=\sum_{r=0}^{k-1} x^{r+1} \frac{f^{(r)}(0)}{(r+1)!} + \frac{1}{(k-1)!}\int_0^x x_1 \int_0^1 (1-\lambda)^{k-1} f^{(k)}(\lambda x_1)\, d\lambda dx_1 \qquad [2]$$

Since the integral in the right-hand side of [1] is bounded by γ the proposition will follow if we prove that $\sum_{r=0}^{k-1} x^r f^{(r)}(0)(r!)^{-1}$ is uniformly bounded for $f\in C_{k\gamma}$. To prove this let us note that since F is a d.f. there exists $M>0$ such that

$$\sup_{f\in C_{k\gamma}} \sup_{0\le x\le 1} \left| \sum_{r=1}^{k} a_r x^r \right| \le M \qquad [3]$$

where $a_r = f^{(r-1)}(0)/(r-1)!$, $r=1,2,\dots,k$. Define

$$A=\{(a_1,\dots,a_k): \sup_{0\le x\le 1} \left| \sum_{r=1}^{k} a_r x^r \right| \le M\}$$

The proposition will be proved if we show that there exists $A>0$ such that $(a_1,\dots,a_k)\in A$ implies $|a_r|\le A$ for $r=1,2,\dots,k$.

Let $(a_1,\dots,a_k)\in A$ and let a_r be an arbitrary component of $(a_1,\dots,a_k)$. If $a_r=0$ certainly $|a_r|\le A$. If $a_r\neq 0$ we can write

$$\sup_{0\le x\le 1} \left| \sum_{v=1}^{k} x^v a_v/a_r \right| \le \frac{M}{|a_r|}. \qquad [4]$$

Now the fundamental theorem of linear approximation of functions (see Lorentz [1965] or Davis [1973]) guarantees the existence of a polynomial of degree k, say, $T_k^{(r)}$ which minimizes the left-hand side of [4] among all the polynomials with coefficient 1 in x^r. Since $T_k^{(r)}$ is not constant, $\| T_k^{(r)} \| > 0$. Therefore $\| T_k^{(r)} \| \le M/|a_r|$ and then $|a_r| \le M/\| T_k^{(r)} \|$. Taking $A = M/\min_r \| T_k^{(r)} \|$ it follows that $|a_r|\le A$ and the proposition is proved.

Remark. By using the same proof as above, it can be shown that for every class $C_{k\gamma}$ there exist numbers $L^{(r)}(\gamma, k)$ such that $\sup\{ |f^{(r)}| : f\in C_{k\gamma}\} \le L^{(r)}(\gamma, k)$.

References

ALEXITS, G. (1961). *Convergence problems of orthogonal series.* Pergamon Press, New York.

BARTLETT, M. S. (1963). *Statistical estimation of density functions.* Sankhya, Ser. A 25, 245-254.

BEAN, S. J. and TSOKOS, C. P. (1980). *Developments in nonparametric density estimation.* International Statist. Review 28, 267-287.

BICKEL, P. and ROSENBLATT, M. (1973). *On some global measures of the deviations of density function estimates.* Ann. Statist. 1. 1071-1095.

CENCOV, N. N. (1962). *Evaluation of an unknown distribution density from observations.* Soviet Math. 3, 1559-1562.

CRAIN, B. R. (1974). *Estimation of distributions using orthogonal expansions.* Ann. of Statist. 2, 454-463.

DUIN, R. P. W. (1976). *On the choice of the smoothing parameters for Parzen estimators of probability density functions.* IEEE Transactions on Computers, C-25, 1175-1179.

FARRELL, R. H. (1972). *On best obtainable asymptotic rates of convergence in estimation of a density function at a point.* Ann. Math. Statist. 43, 170-180.

GOOD, I. J. and GASKINS, R. A. (1980). *Density estimation and bumphunting by the penalized likelihood method exemplified by scatering and meteorit date.* J. Amer. Statist. Assoc. 75, 42-73.

HALL, P. (1982). *Comparison of two orthogonal series methods of estimating a density and its derivatives on an interval.* J. Mult. Analysis 12, 432-449.

HERMANS, J. and HABBEMA, J. D. F. (1976). *Manual for the LLOC_Biscrimin and analysis programs,* Univ. of Leiden, Depart. of Medical Statist.

JACKSON, D. (1930). *The theory of approximation.* Amer. Math. Soc. Vol. XI.

KRONMAL, R. and TARTER, M. (1968). *The estimation of probability densities and cumulatives by Fourier series methods.* J. Amer. Statist. Assoc. 63, 925-952.

LEADBETIER, M. (1963). *On the nonparametric estimation of probability densities,* Ph. D. thesis, University of North Carolina.

LEBESGUE, H. (1909). *Sur les integrales singulieres.* Annales de Toulouse, 25-117.

LOÈVE, M. (1963). *Probability theory.* Van Nostrand Company, New York.

LOFTSGAARDEN, D. D. and QUESENBERRY, C. P. (1965). *A nonparametric estimate of a multivariate density function.* Ann. Math. Statist. 36, 1049-1051.

MARRON, J. S. (1987). *A comparison of cross-validation techniques in density estimation.* Ann. Statist. 15, 152-162.

PARZEN, E. (1962). *On the estimation of a probability density and mode.* Ann. Math. Statist. 33, 1065-1076.

ROSENBLATT, M. (1956). *Remarks on some nonparametric estimates of a density function.* Ann. Math. Statist. 27, 832-837.

ROSENBLATT, M. (1971). *Curve estimates.* Ann. Math. Statist. 42, 1895-1842.

SANSONE, G. (1959). *Orthogonal functions.* Interscience Publishers Inc., New York.

VIOLLAZ, A. J. (1976). *Nonparametric estimation of probability density functions using orthogonal expansions.* Ph. D. Thesis, University of California, Berkeley.

VIOLLAZ, A. J. (1980). *Asymptotic distribution of L_2 norms of the deviations of density functions estimates.* Ann. of Statist. 8, 322-346.

VIOLLAZ, A. J. and CARDOSO, Z. N. (1981). *Calibration of nonparametric probability density estimates.* Contributed Papers 2, 219-222.

VAN RYZIN, I. (1969). *On strong consistency of density estimates.* Ann. Math. Statist., 40, 1765-1772.

SCHWARTZ, S. C. (1967). *Estimation of probability density by an orthogonal series.* Ann. Math. Statist. 38, 1261-1265.

SZEGO, G. (1939). *Orthogonal polynomials.* Amer. Math. Soc. Colloquium publications, Vol. XXIII.

TIMAN, A. (1963). *Theory of approximation of functions of a real variable.* Pergamon Press, New York.

WAHBA, G. (1975). *Optimal convergence propierties of variable knot, kernel, orthonormal series methods for density estimation.* Ann. Statist. 3, 15-29.

WATSON, G. (1959). *Density estimation by orthogonal series.* Ann. Math. Statist. 40, 1496-1498.

WATSON, G. and LEADBETTER, M. (1963). *On the estimation of the probability density, II.* Ann. Math. Statist. 34, 480-491.

WEGMAN, E. (1972). *Non parametric probability density estimation: I. A sumary of available methods.* Technometrics 14, 533-546.

WOODROOFE, M. (1967). *On the maximum deviation of the sample density.* Ann. Math. Statist. 33, 475-481.

WOODROOFE, M. (1970). *On choosing a delta sequence.* Ann. Math. Statist. 41, 1165-1171.

WHITTLE, P. (1958). *On the smoothing of probability density functions.* Roy. Statist. Soc. Serv. B·20, 334-343.

Universidad Nacional de Tucumán
Facultad de Ciencias Exactas y Tecnología
Instituto de Matemática
4000 Tucumán
Argentina

Recibido: 6 de junio de 1988