Ann. Inst. Statist. Math.
34 (1982), Part A, 339-350

INTEGRATED MEAN SQUARE PROPERTIES OF DENSITY ESTIMATION BY ORTHOGONAL SERIES METHODS FOR DEPENDENT VARIABLES

IBRAHIM A. AHMAD*

(Received Dec. 14, 1979; revised Nov. 4, 1981)

Summary

The rates at which integrated mean square and mean squre errors of nonparametric density estimation by orthogonal series method for sequences of strictly stationary strong mixing random variables are obtained. These rates are better than those known to hold for the independent case and they are shown to hold for Markov processes. In fact our results when specialized to the independent case are improvements over previously known results of Schwartz (1967, *Ann. Math. Statist.*, 38, 1262–1265). An extension of the results to estimation of the bivariate density is also given.

1. Introduction

Density estimation by orthogonal series method was first discussed by Čencov [2] and later by Schwartz [6], Kronmal and Tarter [5], and Watson [8]. Let f be a probability density function (p.d.f.) and assume that it is square integrable. Thus $f(x)$ can be expanded by orthogonal series, viz.,

$$f(x)=\sum_{j=0}^{\infty} \theta_j\phi_j(x) , \tag{1.1}$$

where $\theta_j=\int f(x)\phi_j(x)dx$, $j=1, 2, \cdots$ and $\{\phi_j(\cdot)\}$ is an orthonormal basis of $f(x)$. A special choice of $\{\phi_j(\cdot)\}$ that is popular is the normalized Hermite functions;

* Research supported by a faculty summer research grant MS-STAT-42 from the University of Petroleum and Minerals.
AMS 1970 subject classification: Primary G2G05; Secondary 60G10.
Key words and phrases: Density estimation, strong mixing, mean square, rates of convergence, strictly stationary sequences.

$$\phi_j(x)=(2^j j!\pi^{1/2})^{-1/2}e^{-x^2/2}H_j(x)\,, \tag{1.2}$$

where $H_j(x)$ is the jth Hermite polynomial defined by:

$$H_j(x)=(-1)^j e^{x^2}\left(\frac{d^j}{dx^j}e^{-x^2}\right), \qquad j=0, 1, 2, \cdots . \tag{1.3}$$

For other choices of $\phi_j(x)$, $j\geqq 0$, see Kronmal and Tarter [5]. Note that for $\phi_j(x)$ defined in (1.2) we have (see Szegö [7], p. 242) that $|\phi_j(x)| \leqq C_1(j+1)^{-1/12}$ on $(-\infty, \infty)$ and $|\phi_j(x)|\leqq C_1(j+1)^{-1/4}$ on a closed bounded interval $[-M, M]$ such that $f(x)=0$ for all $x\notin[-M, M]$. Thus it is reasonable to assume throughout this paper that for any orthonormal basis $\{\phi_j(\cdot)\}$ used in (1.1),

(1.4) $|\phi_j(x)|<C(j+1)^{-\gamma}$ for some $\frac{1}{2}>\gamma\geqq 0$ and $C>0$ a constant.

Note also that the special case $\gamma=0$ is a customary assumption.

Let $\{X_n\}$ be a sequence of random variables with common marginal p.d.f. $f(x)$ and common bivariate p.d.f. $g(x_1, x_2)$ and assume that $f(x)$ and $g(x_1, x_2)$ are square integrable, so that (1.1) holds for $f(x)$ and that

$$g(x_1, x_2)=\sum_{j=0}^{\infty}\sum_{j^*=0}^{\infty}\delta_{jj^*}\psi_{jj^*}(x_1, x_2)\,, \tag{1.5}$$

where $\delta_{jj^*}=\iint g(x_1, x_2)\psi_{jj^*}(x_1, x_2)dx_1dx_2$ and $\{\psi_{jj^*}(x_1, x_2)\}$ an orthonormal basis of $g(x_1, x_2)$.

Note that since $\{\phi_j(x_1)\cdot\phi_{j^*}(x_2)\}$ from an orthonormal basis over R^2 for $g(x_1, x_2)$ then an example of $\psi_{jj^*}(x_1, x_2)$ is $\phi_j(x_1)\cdot\phi_{j^*}(x_2)$, $j, j^*=0, 1, 2, \cdots$.

An unbiased estimate of θ_j is given by

$$\hat{\theta}_j=n^{-1}\sum_{i=1}^{n}\phi_j(X_i)\,, \qquad j=0, 1, 2, \cdots , \tag{1.6}$$

and thus an estimate of $f(x)$ may be given by:

$$\hat{f}(x)=\sum_{j=0}^{q(n)}\hat{\theta}_j\phi_j(x)\,, \tag{1.7}$$

where $q(n)$ is an integer-valued function of n such that $q(n)\rightarrow\infty$ as $n\rightarrow\infty$. In a similar fashion an unbiased estimate of δ_{jj^*} is given by:

$$\hat{\delta}_{jj^*}=n^{-1}\sum_{l=1}^{n}\psi_{jj^*}(X_l, X_{l+1})\,, \tag{1.8}$$

and hence an estimate of $g(x_1, x_2)$ may be given by:

$$\hat{g}(x_1, x_2)=\sum_{j=0}^{q_1(n)}\sum_{j^*=0}^{q_2(n)}\hat{\delta}_{jj^*}\psi_{jj^*}(x_1, x_2)\,, \tag{1.9}$$

where $q_i(n)\rightarrow\infty$ as $n\rightarrow\infty$, $i=1, 2$, are integer-valued functions of n.

Throughout this paper we shall assume that $q(n)/n \to 0$ and $q_i(n)/n \to 0$, $i=1, 2$, as $n \to \infty$. We shall assume that

$$(1.10) \qquad |\psi_{jj^*}(x_1, x_2)| \leqq C(j+1)^{-\gamma_1}(j^*+1)^{-\gamma_2} \text{ for some } \frac{1}{2} > \gamma_i \geqq 0, \ i=1, 2, \ j, j^*=0, 1, 2, \cdots .$$

The above assumption is motivated by the special choice $\psi_{jj^*}(x_1, x_2) = \phi_j(x_1)\phi_{j^*}(x_2)$, since in this case (in view of (1.4)) $|\psi_{jj^*}(x_1, x_2)| \leqq C(j+1)^{-\gamma} \cdot (j^*+1)^{-\gamma}$, $\frac{1}{2} > \gamma \geqq 0$.

When $\{X_n\}$ are independent and identically distributed (iid) random variables with p.d.f. $f(x)$, Schwartz [6] proved that under certain conditions (see his Theorems 1 and 2),

$$(1.11) \qquad \mathrm{E}\int [\hat{f}(x) - f(x)]^2 dx = O(n^{-(r-1)/r}) ,$$

whenever $q(n) = O(n^{-1/r})$, $r \geqq 3$, and also that

$$(1.12) \qquad \mathrm{E}\,[\hat{f}(x) - f(x)]^2 = O(n^{-(r-2)/r}) ,$$

whenever $q^2(n)/n \to 0$, and $q(n) = O(n^{-1/r})$, $r \geqq 3$.

The purpose of the present investigation is to derive better rates than those given in (1.10) and (1.11) not only when the observations are independent but also when they are taken from a strictly stationary strong mixing process. We also obtain the corresponding rates for $\hat{g}(x_1, x_2)$. This improvement is due basically to the better assumption (1.4) than that used by Schwartz [6]. The bivariate case we tackle gives us a great insight into the performance of the orthogonal series method of density estimation in higher dimensions, a problem that is not heavily investigated in the literature.

Assume that $\{X_n\}$ is a strictly stationary strong mixing sequence of random variables having square integrable marginal p.d.f. $f(x)$ and bivariate p.d.f. $g(x_1, x_2)$ and consider estimating these p.d.f. by $\hat{f}(x)$ and $\hat{g}(x_1, x_2)$ given in (1.7) and (1.9) respectively. Recall that (cf. Ahmad [1] or Ibragimov [4]) a sequence of random variables $\{X_n\}$ is said to be strictly stationary if the joint distribution of $(X_{n_1+k}, \cdots, X_{n_k+k})$ is independent of k, for every $k \geqq 1$ and also recall that a sequence $\{X_n\}$ of random variables is said to be strong mixing with mixing numbers $\{\alpha(n)\}$ if

$$(1.13) \qquad |\mathrm{P}(AB) - \mathrm{P}(A)\,\mathrm{P}(B)| \leqq \alpha(n) ,$$

for any events $A \in \mathscr{F}(1, m)$ and $B \in \mathscr{F}(m+n, \infty)$ with $\mathscr{F}(a, b)$ denoting the σ-field generated by $X_a, \cdots, X_b$ for all integers $a<b$. Strong mix-

ing sequence includes many special cases, e.g. Markov dependence and m-dependence.

In Section 2, bounds on $\mathrm{E}\int[\hat{f}(x)-f(x)]^2dx$ and $\mathrm{E}[\hat{f}(x)-f(x)]^2$ are developed when $\{X_n\}$ is a strictly stationary strong mixing sequence and in Section 3 the corresponding results for $\mathrm{E}\iint[\hat{g}(x_1, x_2)-g(x_1, x_2)]^2dx_1\,dx_2$ and $\mathrm{E}[\hat{g}(x_1, x_2)-g(x_1, x_2)]^2$ are developed.

Note that conditions under which $\hat{f}(x)$ is strongly consistent have been recently given by Ahmad [1]. Analogous results for $\hat{g}(x_1, x_2)$ may be developed similarly. Demonstrating the asymptotic normality of $\hat{f}(x)$ and $\hat{g}(x_1, x_2)$ and other concluding remarks, are embodied in Section 4.

2. Mean square properties of the marginal density

THEOREM 2.1. *Assume that $f(x)$ is square integrable, $q(n)/n\to 0$ as $n\to\infty$ and $\sum_{n=1}^{\infty}\alpha(n)<\infty$, then $\mathrm{E}\int[\hat{f}(x)-f(x)]^2dx\to 0$ as $n\to\infty$. If further $\sum_{j=q+1}^{\infty}\theta_j^2=O(q^{-r+1-2\gamma})$, for some integer $r>0$ then*

$$\mathrm{E}\int[\hat{f}(x)-f(x)]^2dx=O(n^{-1+(1-2\gamma)/r})\,, \tag{2.1}$$

whenever $q(n)=O(n^{1/r})$ and $\gamma\geqq 0$ is such that $|\phi_j(x)|\leqq C(j+1)^{-\gamma}$ for all $j=0, 1, 2, \cdots$.

Remark 2.1. The condition $\sum_{j=q+1}^{\infty}\theta_j^2=O(q^{-r+1-2\gamma})$ for some integer $r>0$ and some $\frac{1}{2}>\gamma\geqq 0$ is satisfied when the coefficients in the expansion of $f(x)$ in (1.1) are the Hermite polynomials, provided that $G^{(r)}(x)=\exp(x^2/2)\frac{d^r}{dx^r}[\exp(-x^2/2)\cdot f(x)]$ is square integrable. To see this; note that $\beta_j=\int G^{(r)}(x)\cdot\phi_j(x)dx$ is such that $\sum_j\beta_j^2<\infty$, and also that by integration by parts

$$\beta_j=(2j)^{1/2}(2j-2)^{1/2}\cdots(2j-2r+2)^{1/2}\theta_{j-r}\,. \tag{2.2}$$

Thus $\sum_{j=q+1}^{\infty}\theta_j^2\leqq\sum_{j=q+1}^{\infty}\beta_{j+r}^2(2j)^{-r}\leqq(2q+1)^{-r}\sum_{j=q+r+1}^{\infty}\beta_j^2=O(q^{-r})$, which is more than we need.

In general, a sufficient condition for $\sum_{j=q+1}^{\infty}\theta_j^2=O(q^{-r+1-2\gamma})$ is that $|\theta_j|=O((j+1)^{-r/2-\gamma})$ for all $j=0, 1, \cdots$ since in this case $\sum_{j=q+1}^{\infty}\theta_j^2\leqq C\sum_{j=q+1}^{\infty}(j+1)^{-r-2\gamma}$

$$\leqq C\int_{q+1}^{\infty} x^{-r-2\gamma}dx = \frac{c}{(r+2\gamma-1)}(q+1)^{-r-2\gamma+1} = O(q^{-r+(1-2\gamma)}).$$

PROOF OF THEOREM 2.1. Write $q(n)=q$, then

$$\text{(2.3)}\quad \mathrm{E}\int[\hat{f}(x)-f(x)]^2dx = \mathrm{E}\int\Big[\sum_{j=0}^{q}(\hat{\theta}_j-\theta_j)\phi_j(x)\Big]^2dx$$
$$+2\,\mathrm{E}\int\Big[\sum_{j=0}^{q}(\hat{\theta}_j-\theta_j)\phi_j(x)\sum_{k=q+1}^{\infty}\theta_k\phi_k(x)\Big]dx$$
$$+\int\Big[\sum_{j=q+1}^{\infty}\theta_j\phi_j(x)\Big]^2dx$$
$$=\sum_{j=0}^{q}\mathrm{E}\,(\hat{\theta}_j-\theta_j)^2\int\phi_j^2(x)dx+\sum_{j=q+1}^{\infty}\theta_j^2\int\phi_j^2(x)dx$$
$$=\sum_{j=0}^{q}\mathrm{E}\,(\hat{\theta}_j-\theta_j)^2+\sum_{j=q+1}^{\infty}\theta_j^2\,,$$

since $\{\phi_j\}$ is an orthonormal basis. But since $\{X_n\}$ are strictly stationary

$$\text{(2.4)}\quad \mathrm{E}\,(\hat{\theta}_j-\theta_j)^2=\mathrm{Var}\,(\hat{\theta}_j)$$
$$=\Big(\frac{1}{n^2}\Big)\sum_{k=1}^{n}\mathrm{Var}\,\phi_j(X_k)+\Big(\frac{2}{n^2}\Big)\sum_{1\leqq k<l\leqq n}\mathrm{Cov}\,(\phi_j(X_k),\,\phi_j(X_l))$$
$$=\Big(\frac{1}{n}\Big)\mathrm{Var}\,\phi_j(X_1)+\Big(\frac{2}{n^2}\Big)\sum_{k=1}^{n-1}(n-k)\,\mathrm{Cov}\,(\phi_j(X_1),\,\phi_j(X_{k+1}))\,.$$

But since $\{\phi_j(X_n)\}$ are bounded and stationary strong mixing, cf. Remark 1.1 of Ahmad [1], then it follows from Lemma 1.2 of Ibragimov [4] that $|\mathrm{Cov}\,(\phi_j(X_1),\,\phi_j(X_{k+1}))|\leqq\dfrac{4C^2\alpha(k)}{(j+1)^{2\gamma}}$, for all $j=0,1,2,\cdots$, and some $\gamma\geqq 0$. Hence for all $j=0,1,2,\cdots$,

$$\text{(2.5)}\quad \mathrm{E}\,(\hat{\theta}_j-\theta_j)^2\leqq\frac{C_1}{n(j+1)^{2\gamma}}+\frac{2C_2}{n^2(j+1)^{2\gamma}}\sum_{k=1}^{n-1}(n-k)\alpha(k)$$
$$\leqq\frac{C_3}{n(j+1)^{2\gamma}}\Big[1+\sum_{n=1}^{\infty}\alpha(n)\Big]\leqq\frac{C_4}{n(j+1)^{2\gamma}}\,.$$

Hence

$$\text{(2.6)}\quad \sum_{j=0}^{q}\mathrm{E}\,(\hat{\theta}_j-\theta_j)^2\leqq\frac{C_5}{n}\sum_{j=0}^{q}(j+1)^{-2\gamma}\leqq\frac{C_5}{n}\Big[1+\int_1^{q+1}x^{-2\gamma}dx\Big].$$

But the integral part of the last upper bound is equal to:

$$[(q+1)^{1-2\gamma}/(1-2\gamma)]-[2\gamma/(1-2\gamma)]=O(q^{1-2\gamma})\,,$$

since $\frac{1}{2}>\gamma\geqq 0$. Therefore,

$$\sum_{j=0}^{q}\mathrm{E}\,(\hat{\theta}_j-\theta_j)^2=O(n^{-1}q^{1-2\gamma})\,.$$

But since $\sum_{j=q+1}^{\infty} \theta_j^2 = O(^{-r+1-2\gamma})$ for some $r>0$, then we have

$$\text{(2.7)} \qquad \mathrm{E}\int [\hat{f}(x)-f(x)]^2 dx \leqq C_6\left\{\frac{q^{1-2\gamma}}{n}+q^{-r+1-2\gamma}\right\},$$

which leads to the desired conclusion upon choosing $q=O(n^{1/r})$. QED.

Remark 2.2. If we assume only that $\alpha(n)=O(n^{-1})$, the bound in (2.1) is not attainable but a weaker bound is possible, viz., in this case

$$\text{(2.8)} \qquad \mathrm{E}\int [\hat{f}(x)-f(x)]^2 dx = O(n^{-1+(1-2\gamma)/r}\ln n).$$

PROOF OF REMARK 2.2. Proceed as in the proof of Theorem 2.1 but instead of (2.5) we get

$$\text{(2.9)} \qquad \mathrm{E}\,(\hat{\theta}_j-\theta_j)^2 \leqq \frac{C_3}{n(j+1)^{2\gamma}}\left[1+\sum_{j=0}^{n}\alpha(j)\right] \leqq \frac{C_3(1+\ln n)}{n(j+1)^{2\gamma}},$$

provided that $\alpha(n)=O(n^{-1})$ and n sufficiently large. Hence the term $\ln n$ must be subsequently added in (2.7) and the remark is proved. QED.

In the next theorem we present the rate at which the mean square error diminishes to zero and again the result refines that of Schwartz [6]. The assumptions imposed are somewhat stronger than those needed for Theorem 2.1.

THEOREM 2.2. *Assume that $f(x)$ is continuous, of bounded variation, and square integrable. If $q^2(n)/n \to 0$ as $n\to\infty$, and if $\sum_{n=0}^{\infty}\alpha(n) < \infty$, then $\mathrm{E}\,[\hat{f}(x)-f(x)]^2 \to 0$ as $n\to\infty$ uniformly in x. If further $\left[\sum_{j=q+1}^{\infty}\theta_j\right]^2 = O(q^{-r+2})$ for some integer $r>2$ then*

$$\text{(2.10)} \qquad \mathrm{E}\,[\hat{f}(x)-f(x)]^2 = O(n^{-1+2(1-\gamma)/r}),$$

whenever $q(n)=O(n^{1/r})$ and $\frac{1}{2}>\gamma\geqq 0$ is such that $|\phi_j(x)|\leqq C(j+1)^{-\gamma}$ for all $j=0, 1, 2, \cdots$.

PROOF. Note that

$$\text{(2.11)} \qquad \mathrm{E}\,[\hat{f}(x)-f(x)]^2 = \mathrm{Var}\,(\hat{f}(x)) + [\mathrm{E}\,\hat{f}(x)-f(x)]^2.$$

We shall evaluate each of these terms separately. Since $\{X_n\}$ are strictly stationary, then

$$\text{(2.12)} \quad \mathrm{Var}\,(\hat{f}(x)) = \mathrm{Var}\left(\sum_{j=0}^{q}\hat{\theta}_j\phi_j(x)\right)$$

$$=\sum_{j=0}^{q}\phi_j^2(x)\operatorname{Var}(\hat{\theta}_j)+2\sum_{j<k}\sum\phi_j(x)\,\phi_k(x)\operatorname{Cov}(\hat{\theta}_j,\hat{\theta}_k)$$
$$=A_n+B_n\,,\quad\text{say.}$$

Now, as in (2.5), above,

$$A_n\leqq C\sum_{j=0}^{q}\phi_j^2(x)(n(j+1)^{2\gamma})^{-1}\leqq\frac{C^*}{n}\sum_{j=0}^{q}(j+1)^{-4\gamma}\tag{2.13}$$
$$\leqq\frac{C}{n(q+1)^{2\gamma}}\sum_{j=0}^{q}(j+1)^{-2\gamma}$$
$$=O(n^{-1}q^{1-4\gamma})\,,$$

where the last bound follows from (2.6) since $1-2\gamma>0$. But choosing $q=O(n^{1/\gamma})$ we conclude that for sufficiently large n,

$$A_n=O(n^{-1+(1-4\gamma)/\gamma})=O(n^{-1+(2-2\gamma)/\gamma})\,.\tag{2.14}$$

Next,

$$B_n\leqq C\sum_{j<k}\sum(j+1)^{-\gamma}(k+1)^{-\gamma}|\operatorname{Cov}(\hat{\theta}_j,\hat{\theta}_k)|\,.\tag{2.15}$$

Now,

$$|\operatorname{Cov}(\hat{\theta}_j,\hat{\theta}_k)|=n^{-2}\left|\operatorname{Cov}\left(\sum_{l=1}^{n}\phi_j(X_l),\ \sum_{m=1}^{n}\phi_k(X_m)\right)\right|\tag{2.16}$$
$$\leqq n^{-2}\sum_{l=1}^{n}\sum_{m=1}^{n}|\operatorname{Cov}(\phi_j(X_l),\phi_k(X_m))|$$
$$=n^{-2}\left\{\sum_{l=1}^{n}|\operatorname{Cov}(\phi_j(X_l),\phi_k(X_l))|\right.$$
$$+\sum_{l<m}\sum|\operatorname{Cov}(\phi_j(X_l),\phi_k(X_m))|$$
$$\left.+\sum_{l>m}\sum|\operatorname{Cov}(\phi_j(X_l),\phi_k(X_m))|\right\}$$
$$\leqq C_1n^{-2}\left\{n(j+1)^{-\gamma}(k+1)^{-\gamma}\right.$$
$$\left.+2\left[\sum_{l=q}^{n}(n-l)\alpha(l)\right][(j+1)^{-\gamma}(k+1)^{-\gamma}]\right\}$$
$$\leqq C_1n^{-1}\left\{(j+1)^{-\gamma}(k+1)^{-\gamma}\left[1+\sum_{n=1}^{\infty}\alpha(n)\right]\right\}$$
$$\leqq\frac{C_2}{n(j+1)^{\gamma}(k+1)^{\gamma}}\,,$$

where C_1 and C_2 are positive constants. Hence from (2.6) it follows that

$$B_n\leqq\frac{C}{n}\left[\sum_{j=0}^{q}(j+1)^{-2\gamma}\right]^2=O(n^{-1}q^{2-4\gamma})=O(n^{-1}q^{2-2\gamma})=O(n^{-1+(2-2\gamma)/\gamma})\,,\tag{2.17}$$

upon choosing $q=O(n^{1/r})$. Hence we obtain that

$$\text{Var}(\hat{f}(x))=O(n^{-1+2(1-\gamma)/r}), \tag{2.18}$$

for some $\gamma\geqq 0$ and $r>0$ an integer.

Finally,

$$[\text{E}\,\hat{f}(x)-f(x)]^2=\left[\sum_{j=q+1}^{\infty}\theta_j\phi_j(x)\right]^2\leqq C(q+1)^{-2\gamma}\left[\sum_{j=q+1}^{\infty}\theta_j\right]^2 = O(q^{-r+2(1-\gamma)}). \tag{2.19}$$

The desired conclusion now follows by taking $q=O(n^{1/r})$. QED.

Remark 2.3. The condition $\left[\sum_{j=q+1}^{\infty}\theta_j\right]^2=O(q^{-r+2})$, in Theorem 2.2 is satisfied whenever the θ_j's are the Hermite polynomials. To see this, note that

$$\left|\sum_{j=q+1}^{\infty}\theta_j\right|\leqq\sum_{j=q+1}^{\infty}|\theta_j|\leqq C_r q^{-r/2+1}, \tag{2.20}$$

as shown in Theorem 2 of Schwartz [6].

Remark 2.4. If we assume the weaker condition that $\alpha(n)=O(n^{-1})$ then the bound in (2.10) is not attainable but we get a weaker bound, viz.,

$$\text{E}\,[\hat{f}(x)-f(x)]^2=O(n^{-1+2(1-\gamma)/r}\ln n). \tag{2.21}$$

The proof of this result follows from the fact that if $\alpha(n)=O(n^{-1})$ then in (2.16), we obtain

$$|\text{Cov}(\hat{\theta}_j,\hat{\theta}_k)|\leqq n^{-1}\left\{(j+1)^{-\gamma}(k+1)^{-\gamma}\left[1+\sum_{p=1}^{\infty}\alpha(p)\right]\right\} \leqq\frac{C\ln n}{n(j+1)^{\gamma}(k+1)^{\gamma}}, \tag{2.22}$$

since for n sufficiently large we have that $\sum_{p=1}^{n}\alpha(p)=O(\ln n)$. Thus, $B_n=O(n^{-1+2(1-2\gamma)/r}\ln n)$. Thus (2.20) follows from (2.21) and (2.19).

Remark 2.5. Consider the case when $\{X_n\}$ is a Markov process defined on some probability space $(R,\mathscr{B},P)$, where R is the real line, $\mathscr{B}$ is the σ-field of Borel subsets of R and P is a probability measure, with stationary transition measure $p(\xi,A)=\text{P}\,[X_{n+1}\in A\,|\,X_n=\xi]$ such that $p(\cdot,A)$ is measurable for a fixed A and $p(\xi,\cdot)$ a probability measure on $\mathscr{B}$ for fixed ξ. Assume that $\{X_n\}$ satisfies Doeblin's condition D_0, viz., there is a finite measure τ on $\mathscr{B}$ with $\tau(R)>0$, an integer $N\geqq 0$ and $\varepsilon>0$ such that $p^{(N)}(\xi,A)\leqq 1-\varepsilon$, if $\tau(A)\leqq\varepsilon$ and there is only one ergodic set $E\subset R$ with $\tau(E)>0$ and this set contains no cyclically moving subsets, where $p^{(n)}(\cdot,\cdot)$ is the n-step transition measure. Under D_0 there exist positive constants $r\geqq 1$ and $0<\rho<1$ and a unique stationary distribution $\pi(\cdot)$ such that $|p^{(n)}(\xi, A)-\pi(A)|<r\rho^n$, $n\geqq 1$. Note that $\pi(\cdot)$ taken as initial distribution and $p^{(n)}(\cdot,\cdot)$ together determine a stationary Markov process (cf. Doob [3], p. 221). Suppose that $p(\xi,\cdot)$ and $\pi(\cdot)$ are absolutely continuous with respect to Lebesgue measure on $(R, \mathscr{B})$ with corresponding densities $g(\cdot,\cdot)$ and $f(\cdot)$. If $f(\cdot)$ is square integrable then it can be estimated by $\hat{f}(x)$ as (1.7). Note that it follows that (cf. Doob [3], p. 222) the following is true; $|\text{Cov}\,(\phi_i(X_1), \phi_j(X_{k+1}))|\leqq 2r^{1/2}\rho^{k/2}[\text{E}\,\phi_i^2(X_1)\,\text{E}\,\phi_j^2(X_1)]^{1/2}$ for all $i, j=0, 1, \cdots, q$. Thus it is obvious that in this case $\sum_{n=0}^{\infty}\alpha(n)<\infty$ and Theorems 2.1 and 2.2 apply to $\hat{f}(x)$. If $\{X_n\}$ are iid, then $\sum_{n=0}^{\infty}\alpha(n)<\infty$, since $\alpha(n)=0$ for all $n\geqq 1$. Thus Theorems 2.1 and 2.2 apply in this case as well, in which case the bounds in (2.1) and (2.10) are improvements over those of Schwartz [6], Theorems 1 and 2 where he has $\gamma=0$.

3. Mean square properties of the bivariate density

The following two theorems are analogous results for $\hat{g}(x_1, x_2)$ to Theorems 2.1 and 2.2.

THEOREM 3.1. *Assume that $g(x_1, x_2)$ is square integrable, $q_i(n)/n\to 0$, $i=1, 2$ as $n\to\infty$, and $\sum_{n=1}^{\infty}\alpha(n)<\infty$, then* $\text{E}\iint[\hat{g}(x_1, x_2)-g(x_1, x_2)]^2dx_1\,dx_2\to 0$ *as $n\to\infty$. If further $\sum_{j=q_1+1}^{\infty}\sum_{j^*=q_2+1}^{\infty}\vartheta_{jj^*}^2=O(q_1^{-(r_1/2)+1-2\gamma_1}\cdot q_2^{-(r_2/2)+1-2\gamma_2})$ for some positive integers r_1 and r_2 then*

$$\text{(3.1)}\qquad \text{E}\iint[\hat{g}(x_1, x_2)-g(x_1, x_2)]^2dx_1\,dx_2=O(n^{-1+(1-2\gamma_1)/r_1+(1-2\gamma_2)/r_2}),$$

whenever $q_i(n)=O(n^{1/r_i})$ and $\gamma_i\geqq 0$ are such that $|\psi_{jj^*}(x_1, x_2)|\leqq C(j+1)^{-\gamma_1}\cdot(j^*+1)^{-\gamma_2}$ for all $j, j^*=0, 1, 2, \cdots$, $i=1, 2$.

Remark 3.1. Choosing $\psi_{jj^*}(x_1, x_2)=\phi_j(x_1)\,\phi_{j^*}(x_2)$, it follows that from (1.4), $|\psi_{jj^*}(x_1, x_2)|\leqq C(j+1)^{-\gamma}(j^*+1)^{-\gamma}$, $\frac{1}{2}>\gamma\geqq 0$, a special case of the assumptions of Theorem 3.1.

SKETCH OF PROOF. Write $q_i(n)=q_i$, $i=1, 2$, then

$$\text{(3.2)}\qquad \text{E}\iint[\hat{g}(x_1, x_2)-g(x_1, x_2)]^2dx_1\,dx_2=\sum_{j=0}^{q_1}\sum_{j^*=0}^{q_2}\text{E}\,(\hat{\vartheta}_{jj^*}-\vartheta_{jj^*})^2 + \sum_{k=q_1+1}^{\infty}\sum_{k^*=q_2+1}^{\infty}\vartheta_{kk^*}^2\,.$$

But as in Theorem 2.1 we have

$$\mathrm{E}\,(\hat{\delta}_{jj^*}-\delta_{jj^*})^2 \leqq \frac{C}{n(j+1)^{2r_1}(j^*+1)^{2r_2}}\,. \tag{3.3}$$

Thus

$$\sum_{j=0}^{q_1}\sum_{j^*=0}^{q_2}\mathrm{E}\,(\hat{\delta}_{jj^*}-\delta_{jj^*})^2 \leqq \frac{C}{n}\Big[\sum_{j=0}^{q_1}(j+1)^{-2r_1}\Big]\Big[\sum_{j^*=0}^{q_2}(j^*+1)^{-2r_2}\Big]$$
$$=O(n^{-1}q_1^{1-2r_1}q_2^{1-2r_2})\,.$$

But also since $\sum_{j=q_1+1}^{\infty}\sum_{j^*=q_2+1}^{\infty}\delta_{jj^*}^2=O(q_1^{-(r_1/2)+1-2r_1}\cdot q_2^{-(r_2/2)+1-2r_2})$, the desired conclusion follows by taking $q_i=O(n^{1/r_i})$, $i=1, 2$. QED.

Remark 3.2. The condition $\sum_{j=q_1+1}^{\infty}\sum_{j^*=q_2+1}^{\infty}\delta_{jj^*}^2=O(q_1^{-(r_1/2)+1-2r_1}\cdot q_2^{-(r_2/2)+1-2r_2})$ is satisfied if δ_{jj^*} are the bivariate Hermite polynomials while a general sufficient condition is that $|\delta_{jj^*}|=O((j+1)^{(r_1/4)-r_1}\cdot(j^*+1)^{(r_2/4)-r_2})$.

THEOREM 3.2. *Assume that $g(x_1, x_2)$ is continuous, of bounded variation, and square integrable. If $q_i^2(n)/n \to 0$ as $n\to\infty$, $i=1, 2$ and if $\sum_{n=1}^{\infty}\alpha(n)<\infty$, then $\mathrm{E}\,[\hat{g}(x_1, x_2)-g(x_1, x_2)]^2\to 0$ as $n\to\infty$ uniformly in (x_1, x_2). If further $\sum_{j=q_1+1}^{\infty}\sum_{j^*=q_2+1}^{\infty}|\delta_{jj^*}|\leqq O(q_1^{-(r_1/2)+2}\cdot q_2^{-(r_2/2)+2})$ for some positive integers r_1 and r_2 then*

$$\mathrm{E}\,[\hat{g}(x_1, x_2)-g(x_1, x_2)]^2=O(n^{-1+2(1-r_1)/r_1+2(1-r_2)/r_2})\,, \tag{3.4}$$

whenever $q_i(n)$, $i=1, 2$ and $\psi_{jj^*}(x_1, x_2)$, $j, j^*=1, 2, \cdots$ as in Theorem 3.1.

We note here that Remarks 2.2 and 2.3 have their counterparts for $\hat{g}(x_1, x_2)$ thus if $\alpha(n)=O(n^{-1})$ we arrive at the estimates in (3.1) and (3.4) with the add factor $(\ln n)$.

Remark 3.3. The condition $\sum_{j=q_1+1}^{\infty}\sum_{j^*=q_2+1}^{\infty}|\delta_{jj^*}|\leqq O(q_1^{-(r_1/2)+2}\cdot q_2^{-(r_2/2)+2})$ holds for the case when δ_{jj^*}'s are the Hermite polynomials.

Remark 3.4. Recall the Markov process defined in Remark 2.4 and let $g(\cdot, \cdot)$ denote the joint p.d.f. Again in this case, $\sum_n \alpha(n)<\infty$, and then Theorems 3.1 and 3.2 apply.

We conclude this section by noticing that we can extend Theorems 3.1 and 3.2 to the case of multivariate p.d.f. $g(x_1, \cdots, x_k)$ (the pdf of $X_1, \cdots, X_k$). This extension is straightforward after one defines the orthonormal basis for $g(x_1, \cdots, x_k)$ and modifies the conditions of these theorems accordingly.

4. Concluding remarks

(i) Writing $Y_{n,k}=\sum_{j=0}^{q(n)}\phi_j(X_k)\phi_j(x)$, $k=1,\cdots,n$, then $\hat{f}(x)=n^{-1}\sum_{k=1}^{n}Y_{n,k}$, $\{Y_{n,k},\ 1\leqq k\leqq n,\ n=1,2,\cdots\}$ is an array of stationary strong mixing bounded (with bound $(1+q(n))C$) random variables. Further let $W_{n,k}=(Y_{n,k}-\mathrm{E}\,Y_{n,k})/(1+q(n))$, then $\mathrm{E}\,W_{n,k}=0$ and $|W_{n,k}|<C$ for some positive constant C. If $\sum_{n=0}^{\infty}\alpha(n)<\infty$, then it follows that (Theorem 1.6, Ibragimov [4])

$$\lim_{n\to\infty}\mathrm{E}\left\{\frac{\sqrt{n}}{1+q(n)}\sum_{k=1}^{n}W_{n,k}\right\}^2=\sigma_1^2,$$

exists. Moreover, if $\sigma_1>0$, then $\mathrm{P}\left[\frac{\sqrt{n}}{1+q(n)}\{\hat{f}(x)-\mathrm{E}\,\hat{f}(x)\}\leqq\sigma_1 x\right]\to\Phi(x)$ as $n\to\infty$ where $\Phi(\cdot)$ is the standard normal df. Note that the condition $\alpha(n)<m/n.\ \log n$ in Theorem 1.6 of Ibragimov [4] is not needed.

Similarly we can argue the asymptotic normality of $\hat{g}(x_1, x_2)$. Let $U_{n,k}=\sum_{j=0}^{q_1(n)}\sum_{j^*=0}^{q_2(n)}\psi_{jj^*}(X_k, X_{k+1})\psi_{jj^*}(x_1, x_2)$, then $\hat{g}(x_1, x_2)=n^{-1}\sum_{k=1}^{n}U_{n,k}$ where $\{U_{n,k},\ 1\leqq k\leqq n,\ n=1,2,\cdots\}$ is an array of stationary strong mixing bounded (with bound $(1+q_1(n))(1+q_2(n))C$) random variables. Further, let $V_{n,k}==(U_{n,k}-\mathrm{E}\,U_{n,k})/(1+q_1(n))(1+q_2(n))$, then $\mathrm{E}\,V_{n,k}=0$ and $|V_{n,k}|<C$ for some $C>0$. If $\sum_{n=0}^{\infty}\alpha(n)<\infty$, then again we have

$$\lim_{n\to\infty}\mathrm{E}\left\{\frac{\sqrt{n}}{(1+q_1(n))(1+q_2(n))}\sum_{k=1}^{n}V_{n,k}\right\}^2=\sigma_2^2,$$

exists. Thus if $\sigma_2>0$, then

$$\mathrm{P}\left[\frac{\sqrt{n}}{(1+q_1(n))(1+q_2(n))}(\hat{g}(x_1, x_2)-\mathrm{E}\,\hat{g}(x_1, x_2))\leqq\sigma_2 x\right]\to\Phi(x)\qquad\text{as } n\to\infty\,.$$

(ii) An application cited in Ahmad [1] for the estimate $\hat{f}(x)$ is to use it in estimating the functional $\Delta(f)=\int f^2(x)dx$. Two estimates are proposed, viz., $\hat{\Delta}(f)=\int\hat{f}(x)dF_n(x)$, where $F_n(x)=n^{-1}\sum_{k=1}^{n}I(X_k\leqq x)$ is the empirical distribution function and $\tilde{\Delta}(f)=\int\hat{f}^2(x)dx$. It is shown in Ahmad [1] that $\hat{\Delta}(f)$ and $\tilde{\Delta}(f)$ are strongly consistent (under certain conditions), here we can evaluate the rate at which $\mathrm{E}\,|\hat{\Delta}(f)-\Delta(f)|$ and $\mathrm{E}\,|\tilde{\Delta}(f)-\Delta(f)|$ converge to 0 as $n\to\infty$ using Theorems 2.1 and 2.2. To see this note that

$$\begin{aligned}\mathrm{E}\,|\hat{\Delta}(f)-\Delta(f)| &\leqq C\Big\{\mathrm{E}^{1/2}\,[\sup_x |\hat{f}(x)-\mathrm{E}\,\hat{f}(x)|]^2+\sup_x |\mathrm{E}\,\hat{f}(x)-f(x)| \\ &\qquad +\mathrm{E}^{1/2}\Big[\Big[\int f(x)dF_n(x)-\int f(x)dF(x)\Big]^2\Big\} \\ &=O(n^{-1/2+(1-\gamma)/r})+O(n^{-1/2}) \\ &=O(n^{-1/2+(1-\gamma)/r})\,,\end{aligned}$$

by Theorem 2.2, assuming its conditions hold. Next,

$$\begin{aligned}\mathrm{E}\,|\tilde{\Delta}(f)-\Delta(f)| &\leqq C\Big\{\mathrm{E}\int [\hat{f}(x)-f(x)]^2dx+\mathrm{E}^{1/2}\,[\sup_x |\hat{f}(x)-f(x)|]^2\Big\} \\ &=O(n^{-1+(1-2\gamma)/r})+O(n^{-1/2+(1-\gamma)/r}) \\ &=O(n^{-1/2+(1-\gamma)/r})\,,\end{aligned}$$

again by application of Theorems 2.1 and 2.2. Hence the two estimates achieve the same rate of L_1-consistency.

Acknowledgement

The author is indebted to the referee for his detailed comments which resulted in correcting and improving the paper over earlier versions.

UNIVERSITY OF PETROLEUM AND MINERALS

REFERENCES

[1] Ahmad, I. A. (1979). Strong consistency of density estimation by orthogonal series methods for dependent variables with applications, *Ann. Inst. Statist. Math.*, A, **31**, 279-288.

[2] Čencov, N. N. (1962). Estimation of an unknown density function from observations, *Dokl. Akad. Nauk, SSSR*, **147**, 45-48 (in Russian).

[3] Doob, J. L. (1953). *Stochastic Processes*, Wiley and Sons, New York.

[4] Ibragimov, I. A. (1962). Some limit theorems for stationary sequences, *Theory Prob. Appl.*, **7**, 349-382.

[5] Kronmal, R. and Tarter, M. (1968). The estimation of probability densities and cumulative by Fourier series methods, *J. Amer. Statist. Ass.*, **63**, 925-952.

[6] Schwartz, S. C. (1967). Estimation of a probability density by an orthogonal series, *Ann. Math. Statist.*, **38**, 1262-1265.

[7] Szegö, G. (1959). *Orthogonal Polynomials*, Amer. Math. Soc. Coll. Publ., 32, Providence, R.I.

[8] Watson, G. S. (1969). Density estimation by orthogonal series, *Ann. Math. Statist.*, **40**, 1496-1498.