Ann. Inst. Statist. Math.
37 (1985), Part A, 443-454

FOURIER AND HERMITE SERIES ESTIMATES OF REGRESSION FUNCTIONS

WLODZIMIERZ GREBLICKI AND MIROSLAW PAWLAK

(Received May 7, 1984; revised May 13, 1985)

Summary

In the paper we estimate a regression $m(x)=\mathrm{E}\{Y|X=x\}$ from a sequence of independent observations $(X_1, Y_1), \cdots, (X_n, Y_n)$ of a pair (X, Y) of random variables. We examine an estimate of a type $\hat{m}(x)=\sum_{j=1}^{n} Y_j\varphi_N(x, X_j)\Big/\sum_{j=1}^{n} \varphi_N(x, X_j)$, where N depends on n and φ_N is Dirichlet kernel and the kernel associated with the Hermite series. Assuming, that $\mathrm{E}|Y|<\infty$ and $|Y|\leqq\gamma\leqq\infty$, we give condition for $\hat{m}(x)$ to converge to $m(x)$ at almost all x, provided that X has a density. If the regression has s derivatives, then $\hat{m}(x)$ converges to $m(x)$ as rapidly as $O(n^{-(2s-1)/4s})$ in probability and $O(n^{-(2s-1)/4s}\log n)$ almost completely.

1. Introduction

In this paper we estimate a regression function $m(x)=\mathrm{E}\{Y|X=x\}$ from a sequence of independent observations $(X_1, Y_1), \cdots, (X_n, Y_n)$ of a pair (X, Y) of random variables. Assuming that X has a density f, we use the following estimator:

$$\hat{m}(x)=\hat{p}(x)/\hat{q}(x)$$

where
$$\hat{p}(x)=\sum_{j=1}^{n} Y_j\varphi_N(x, X_j)$$

and
$$\hat{q}(x)=\sum_{j=1}^{n} \varphi_N(x, X_j)$$

and where N depends on n and $\{\varphi_n\}$ is a sequence of some kernel functions. For X taking values in an interval $[-\pi, \pi]$, φ_N is the Dirichlet kernel, whereas, for X valued in the whole real line, φ_N is a kernel associated with the Hermite series. We will treat 0/0 in the

Key words and phrases: Regression function, Fourier series, Hermite series, nonparametric estimate.

above definition as 0 and moreover we define $m(x)=0$ for such points where $f(x)=0$. The estimate is closely related with the orthogonal series estimate $\hat{q}(x)/n$ of the density $f(x)$ studied in the literature. Density estimates derived from the trigonometric series were studied by Kronmal and Tarter [10], Foldes and Révész [3], Wahba [19], Bleuez and Bosq [1] as well as Tarter and Kronmal [18]. In turn, the Hermite series was employed by Schwartz [14], Foldes and Révész [3], Bleuez and Bosq [1], Walter [20], as well as Greblicki and Pawlak [6], [7].

The estimate examined in this paper resembles the kernel one

$$\sum_{j=1}^{n} Y_j K\left(\frac{x-X_j}{h}\right) \Big/ \sum_{j=1}^{n} K\left(\frac{x-X_j}{h}\right)$$

where h depends on n and K is a Borel kernel, deeply studied by many authors, see e.q. Watson [21], Nadaraya [12] and more recent works of Krzyzak and Pawlak [11] as well as Greblicki, Krzyzak and Pawlak [5]. Additional references about regression estimates can be found in the survey paper of Collomb [2].

We show that the estimates using both the trigonometric and the Hermite series converge to the true regression in probability and almost completely for almost all points. Moreover the rates of the convergence equals $O(n^{-(2s-1)/4s})$ in probability and $O(n^{-(2s-1)/4s}\log n)$ almost completely.

Pointwise rates of convergence for a class nonparametric regression estimators have been investigated by Stone [15]. The rate given by him is slightly better than that obtained by us, but his assumptions are not comparable with ours.

2. Trigonometric series estimate

Let (X, Y) be a pair of random variables taking values in Q and R, respectively, where $Q=[-\pi, \pi]$. We estimate $m(x)=\mathrm{E}\{Y|X=x\}$ from a sequence $(X_1, Y_1), \cdots, (X_n, Y_n)$ of independent observations of (X, Y). We assume that there exists a density of X denoted by f and that, for some $p>1$, $f \in L_p$ and $g \in L_p$, where $g(x)=m(x)f(x)$.

Let

$$g(x) \sim \sum_k a_k e^{ikx}$$

and

$$f(x) \sim \sum_k b_k e^{ikx}$$

i.e. let

$$a_k = (2\pi)^{-1}\,\mathrm{E}\{Ye^{-ikX}\}$$

and

$$b_k = (2\pi)^{-1}\,\mathrm{E}\{e^{-ikX}\}$$

be Fourier coefficients of g and f, respectively.

Since $m(x)=g(x)/f(x)$, we estimate $m(x)$ with the following formula:

$$\hat{m}(x)=\frac{\sum_{|k|\leq N}\hat{a}_k e^{ikx}}{\sum_{|k|\leq N}\hat{b}_k e^{ikx}}$$

where N depends on n and

$$\hat{a}_k=(2\pi)^{-1}n^{-1}\sum_{j=1}^{n}Y_j e^{-ikX_j}$$

and

$$\hat{b}_k=(2\pi)^{-1}n^{-1}\sum_{j=1}^{n}e^{-ikX_j}$$

are unbiased estimates of a_k and b_k, respectively.

One can rewrite the estimate with the following form:

$$\hat{m}(x)=\frac{\sum_{j=1}^{n}Y_j D_N(x-X_j)}{\sum_{j=1}^{n}D_N(x-X_j)}$$

where πD_N is the Dirichlet kernel of the Nth order i.e. where

$$2\pi D_N(y)=\sum_{m=-N}^{N}e^{-imy}=(\sin[(N+1/2)x])/\sin(x/2)\ .$$

3. Consistency of trigonometric series estimate

Let us denote

$$\hat{g}(x)=n^{-1}\sum_{j=1}^{n}Y_j D_N(x-X_j)$$

and

$$\hat{f}(x)=n^{-1}\sum_{j=1}^{n}D_N(x-X_j)\ .$$

Clearly $\hat{m}(x)=\hat{g}(x)/\hat{f}(x)$.

THEOREM 1. *Let* $f, g\in L_p$, $p>1$. *Let* $\mathrm{E}|Y|<\infty$. *If*

(1) $N(n)\xrightarrow{n}\infty$

(2) *and* $N(n)/n\xrightarrow{n}0$

then $\hat{m}(x)\xrightarrow{n}m(x)$ *in probability for almost all* $x\in Q$.

PROOF. From $g\in L_p$, $p>1$, it follows that $\int D_N(x-y)g(y)dy\rightarrow g(x)$ as $N\rightarrow\infty$ for almost all $x\in Q$, see Hunt [9]. Thus

(3) $$\mathrm{E}\,\hat{g}(x)\xrightarrow{n}g(x)$$

for almost all $x \in Q$.

For Y such that $\mathrm{E}\, Y^2 < \infty$ we have

$$\operatorname{var} \hat{g}(x) \leqq n^{-1} \int D_N^2(x-y) w(y) dy ,$$

where $w(x) = \mathrm{E}\{Y^2 | X = x\}$. Hence

$$\operatorname{var} \hat{g}(x) \leqq (2\pi)^{-1} n^{-1} (2N+1) \int F_{2N}(x-y) w(y) dy ,$$

where πF_N is the Fejér kernel of the Nth order i.e. $2\pi F_N(x) = (\sin^2 [(N+1)x/2])/(N+1) \sin^2 (x/2)$. Since $w \in L_1$, by virtue of the Lebesque theorem on the Cezaro summability, see Zygmund [22], $\int F_N(x-y) w(y) dy \to w(x)$ as $N \to \infty$ for almost all $x \in Q$. Thus,

$$\operatorname{var} \hat{g}(x) = O(N/n) . \tag{4}$$

Therefore

$$\operatorname{var} \hat{g}(x) \xrightarrow{n} 0 \tag{5}$$

for almost all $x \in Q$, whenever $\mathrm{E}\, Y^2 < \infty$.

Let now $\mathrm{E}\, |Y| < \infty$. We shall now use (5) and apply truncation arguments. For $t > 0$, let Y_j' equal Y_j or zero depending on Y_j is not greater or greater than t, respectively. Let, moreover,

$$Y_j'' = Y - Y_j' , \quad m'(x) = \mathrm{E}\{Y_1' | X_1 = x\} \quad \text{and} \quad m''(x) = \mathrm{E}\{Y_1'' | X_1 = x\} .$$

Thus

$$\begin{aligned} \mathrm{E}\{|\hat{g}(x) - \mathrm{E}\, \hat{g}(x)|\} \leqq n^{-1} \mathrm{E} \left\{ \left| \sum_{j=1}^{n} (Y_j' - m'(X_j)) D_N(x - X_j) \right| \right\} \\ + n^{-1} \mathrm{E} \left\{ \left| \sum_{j=1}^{n} (Y_j'' - m''(X_j)) D_N(x - X_j) \right| \right\} . \end{aligned} \tag{6}$$

The second term in (6) does not exceed $2 \int g_t(y) |D_N(x-y)| dy$, where $g_t(x) = \mathrm{E}\{|Y_1''| | X_1 = x\}$. The fact that g_t is monotone in t and that $\mathrm{E}\, g_t(X) \to 0$ as $t \to \infty$ imply $g_t(x) \to 0$ as $t \to \infty$ for almost all $x \in Q$. Hence, $\int g_t(y) |D_N(x-y)| dy$ can be made arbitrarily small by taking t large enough.

On the other hand, by (5), for every fixed t, the first term in (6) converges to zero as n tends to infinity for almost all $x \in Q$.

Finally, the quantity in (6) converges to zero as n tends to infinity for almost all $x \in Q$. From this and (3) it follows that

$$\hat{g}(x) \xrightarrow{n} g(x) \text{ in probability for almost all } x \in Q .$$

Using similar arguments one can easily verify that $\hat{f}(x) \to f(x)$ as $n \to \infty$ for almost all $x \in Q$. The theorem has been proved.

THEOREM 2. *Let* $|Y| \leqq \gamma < \infty$ *almost surely. Let* $f, g \in L_p$, $p>1$. *If, in addition to* (1),

$$\sum_{n=1}^{\infty} \exp\{-\alpha n/N(n)\} < \infty \tag{7}$$

all $\alpha > 0$, *then*

$$\hat{m}(x) \xrightarrow{n} m(x) \text{ almost completely for almost all } x \in Q .$$

PROOF. Using Hoeffding [8] inequality one gets

$$\begin{aligned} \mathrm{P}\{|\hat{g}(x) - \mathrm{E}\,\hat{g}(x)| > t\} &= \mathrm{P}\left\{n^{-1}\left|\sum_{j=1}^{n} [Y_j D_N(x-X_j) - \mathrm{E}\{Y_j D_N(x-X_j)\}]\right| > t\right\} \\ &\leqq 2\exp\{-(2\pi)t^2 n/2(N+1)\} . \end{aligned}$$

By this, (3), conditions (1) and (7),

$$\hat{g}(x) \xrightarrow{n} g(x) \text{ almost completely for almost all } x \in Q .$$

In the same way one can verify that

$$\hat{f}(x) \xrightarrow{n} f(x) \text{ almost completely}$$

for almost all $x \in Q$. Thus, the proof has been completed.

It is well known that there exist functions integrable over Q such that their Fourier series do not converge to these functions for almost all $x \in Q$, see Zygmund ([22], Section 8.4). Thus, there exist densities for which $\mathrm{E}\,\hat{f}(x)$ does not converge to $f(x)$ as N tends to infinity for almost all $x \in Q$. In the consequence, for $p=1$, one should not expect the consistency of the estimate for almost all $x \in Q$.

4. Convergence rate for Fourier series estimate

By multiple integrating $\int e^{-ikx} g(x)dx$ and $\int e^{-ikx} f(x)dx$ by parts we get.

LEMMA 1. *Let* $k \neq 0$. *If, for* $j = 0, 1, \cdots, s-1$,

$$f^{(j)}(-\pi) = f^{(j)}(\pi) = 0 ,$$

then

$$b_k=(-i)^s\beta_k/k^s$$

where β_k *is the Fourier coefficient of* $f^{(s)}$. *If, moreover, for* $j=0,1,\cdots,s-1$,

$$m^{(j)}(-\pi)=m^{(j)}(\pi)=0\ ,$$

then

$$a_k=(-1)^s\alpha_k/k^s$$

where α_k *is the kth Fourier coefficient of* $g^{(s)}$.

By virtue of (3) and Lemma 1,

$$\begin{aligned} (8)\qquad |\mathrm{E}\,\hat{g}(x)-g(x)| &= \Big|\sum_{|k|>N} a_k\Big| \\ &\leqq \Big(\sum_{|k|>N}|\alpha_k|^2\Big)^{1/2}\Big(\sum_{j=N+1}^{\infty} j^{-2s}\Big)^{1/2} \\ &\leqq (2\pi)^{-1/2}\|g^{(s)}\|\,N^{-(s-1/2)}=O(N^{-(s-1/2)}) \end{aligned}$$

for almost all $x\in Q$, where $\|\cdot\|$ is the L_2 norm. By this and (4),

$$\mathrm{E}\,(\hat{g}(x)-g(x))^2=O(N/n)+O(N^{-2(s-1/2)})\ ,$$

for almost all $x\in Q$. Hence, for

$$(9)\qquad N(n)\sim n^{1/2s}\ ,$$

$$(10)\qquad \mathrm{E}\,(\hat{g}(x)-g(x))^2=O(n^{-(2s-1)/2s})$$

for almost all $x\in Q$. Similarily, for $\{N(n)\}$ picked up according to (9)

$$(11)\qquad \mathrm{E}\,(\hat{f}(x)-f(x))^2=O(n^{-(2s-1)/2s})$$

for almost all $x\in Q$.

In order to get the convergence rate for $\hat{m}(x)$ we shall use.

LEMMA 2. *If*

$$\mathrm{E}\,(\hat{g}(x)-g(x))^2=O(\alpha_n)\qquad \textit{and}\qquad \mathrm{E}\,(\hat{f}(x)-f(x))^2=O(\beta_n)\ ,$$

then $\mathrm{P}\,\{|\hat{m}(x)-m(x)|>\varepsilon|m(x)|\}=O(\gamma_n)\ ,$ *where* $\gamma_n=\max(\alpha_n,\beta_n)$.

PROOF. We begin with the following inequality

$$|\hat{m}(x)-m(x)|\leqq\left|\frac{\hat{g}(x)}{f(x)}\right|\left|\frac{\hat{f}(x)-f(x)}{f(x)}\right|+\left|\frac{\hat{g}(x)-g(x)}{f(x)}\right|\ .$$

Making use of the above inequality one can easily verify that, for $\varepsilon>0$,

$$|\hat{g}(x)-g(x)|<|g(x)|\varepsilon/(2+\varepsilon)\quad\text{and}\quad|\hat{f}(x)-f(x)|<f(x)\varepsilon/(2+\varepsilon)$$

imply $|\hat{m}(x)-m(x)|<\varepsilon|m(x)|$.

Thus

$$P\{|\hat{m}(x)-m(x)|>\varepsilon|m(x)|\}\leqq P\{|\hat{g}(x)-g(x)|>|g(x)|\varepsilon/(2+\varepsilon)\} + P\{|\hat{f}(x)-f(x)|>f(x)\varepsilon/(2+\varepsilon)\} .$$

Using Chebyshev's inequality we complete the proof.

THEOREM 3. *Let* $E\,Y^2<\infty$. *Let* $f, g\in L_p$, $p>1$. *For some* s, *let* m *and* f *satisfy condition of Lemma* 1. *Then, for* $\{N(n)\}$ *picked up according to* (9)

$$P\{|\hat{m}(x)-m(x)|>\varepsilon|m(x)|\}=O(n^{-(2s-1)/2s})$$

and $$|\hat{m}(x)-m(x)|=O(n^{-(2s-1)/4s}) \text{ in probability}$$

for almost all $x\in Q$.

PROOF. The first part of the assertion is a consequence of Lemma 2, (10) and (11). The second part is easy to verify and its proof is omitted.

THEOREM 4. *Let* $|Y|\leqq\gamma<\infty$ *almost surely. Let* $f, g\in L_p$, $p>1$. *For some* s, *let* m *and* f *satisfy condition of Lemma* 1. *Then, for* $\{N(n)\}$ *selected as in Theorem* 3

$$|\hat{m}(x)-m(x)|=O(n^{-(2s-1)/4s}\log n) \text{ almost completely}$$

for almost all $x\in Q$.

The proof is easy and is omitted.

5. Hermite series estimate

Now we employ the Hermite system $\{h_k,\ k=0, 1,\cdots\}$, where

$$h_k(x)=(2^k k!\pi^{1/2})^{-1/2}e^{-x^2/2}H_k(x) ,$$

and where

$$H_k(x)=e^{x^2}(d^k/dx^k)e^{-x^2}$$

is the kth Hermite polynomial.

Let now

$$g(x)\sim\sum_{k=0}^{\infty}a_k h_k(x) \quad\text{and}\quad f(x)\sim\sum_{k=0}^{\infty}b_k h_k(x)$$

i.e. let $$a_k=\int g(x)h_k(x)dx \quad\text{and}\quad b_k=\int f(x)h_k(x)dx .$$

Now

$$\hat{m}(x)=\frac{\sum_{k=0}^{N}\hat{a}_k h_k(x)}{\sum_{k=0}^{N}\hat{b}_k h_k(x)}$$

where

$$\hat{a}_k=n^{-1}\sum_{j=1}^{n} Y_j h_k(X_j)$$

and

$$\hat{b}_k=n^{-1}\sum_{j=1}^{n} h_k(X_j)$$

estimate a_k and b_k, respectively. We shall also use the following representation of the estimate:

$$\hat{m}(x)=\frac{\sum_{j=1}^{n} Y_j d_N(x, X_j)}{\sum_{j=1}^{n} d_N(x, X_j)}$$

where $d_N(x, y)=\sum_{k=0}^{N} h_k(x)h_k(y)$.

Applying Christoffel's formula one gets

$$d_N(x, y)=(N+1)^{1/2}\frac{h_{N+1}(x)h_N(y)-h_{N+1}(y)h_N(x)}{2^{1/2}(y-x)}\,. \tag{12}$$

We shall need the next two Lemmas:

LEMMA 3.

$$\max_y |d_N(x, y)|\leqq c(x)(N+1)^{1/2}\,.$$

The proof of Lemma 3 may be found in Greblicki and Pawlak [7].

LEMMA 4. *If* $\mathrm{E}\,Y^2<\infty$, *then*

$$\operatorname{var}\{Yd_N(x, X)\}\leqq d(x)(N+1)^{1/2} \textit{ for almost all } x\in Q\,.$$

PROOF. Clearly

$$\operatorname{var}\{Yd_N(x, X)\}\leqq\int d_N^2(x, y)w(y)f(y)dy\,, \tag{13}$$

where $w(x)=\mathrm{E}\{Y^2|X=x\}$.

Let us fix $x\in R$. Let $\delta>0$. From the inequalities

$$\max_{|x|\leqq a}|h_k(x)|\leqq c(a)(k+1)^{-1/4}\,,$$

$$\max_x |h_k(x)|\leqq c(k+1)^{-1/12}\,,$$

see Szegö ([16], p. 242) and Christoffel's formula (12), we get

$$\text{(14)}\qquad \int_{|x-y|\geqq\delta} d_N^2(x, y)w(y)f(y)dy \leqq \max_{|x-y|\geqq\delta} d_N^2(x, y)\,\mathrm{E}\,Y^2 \leqq c^2c(|x|)(N+1)^{1/2}/\delta^2 = O(N^{1/2}) .$$

In Sansone [13] we find

$$\pi d_N(x, y) = [\sin \mu(x-y)]/(x-y) + T_N(x, y)/\mu ,$$

where $2\mu=(2N+1)^{1/2}+(2N+3)^{1/2}$, and where T_N is bounded by a constant independent of N, for x and y varying in finite intervals. Thus

$$\int_{|x-y|<\delta} T_N^2(x, y)w(y)f(y)dy \leqq \lambda\,\mathrm{E}\,Y^2 ,$$

where λ is independent of N. Moreover,

$$\text{(15)}\qquad (\mu\pi)^{-1}\int_{|x-y|\leqq\delta} \{[\sin^2 \mu(x-y)]/(x-y)^2\}\, w(y)f(y)dy \leqq (\mu\pi)^{-1}\int_{|x-y|\leqq\delta} \{[\sin^2 \mu(x-y)]/\sin^2(x-y)\}\, w(y)f(y)dy .$$

Let us observe that the function in brackets is the Fejer kernel. Therefore, by virtue of the Lebesque theorem on the Cesaro summability of the trigonometric series, the quantity in (15) converges to $w(x)f(x)$ as n tends to infinity for almost all $x\in R$. Finally

$$\int_{|x-y|<\delta} d_N^2(x, y)w(y)f(y)dy = O(N^{1/2}) ,$$

for almost all $x \in R$.

This and (14) yield the desired assertion. The proof has been completed.

We are now in a position to give.

THEOREM 5. *Let* $f, g \in L_p$, $p>1$. *Let* $\mathrm{E}\,|Y|<\infty$. *If, in addition to* (1),

$$\text{(16)}\qquad N^{1/2}(n)/n \xrightarrow{n} 0 ,$$

then $\hat{m}(x) \xrightarrow{n} m(x)$ *in probability for almost all* $x\in R$.

PROOF. Let us take $\hat{g}(x)$ into account. Clearly

$$\text{(17)}\qquad \mathrm{E}\,\hat{g}(x) = \sum_{k=0}^{N} a_k h_k(x) .$$

The equiconvergence theorem in Szegö ([16], p. 247) says that at every point $x\in R$ both the Hermite expansion of any integrable function and the Fourier expansion of the function taken at arbitrary interval containing x in its interior have the same limit. By virtue of this, Hunt's [9] result and (17),

$$\mathrm{E}\,\hat{g}(x) \xrightarrow{n} g(x) \tag{18}$$

for almost all $x \in R$.

From this and Lemma 4 it follows that

$$\mathrm{E}\,(\hat{g}(x)-g(x))^2 \xrightarrow{n} 0$$

for almost all $x \in R$, whenever $\mathrm{E}\,Y^2 < \infty$. Now, using truncation arguments similar to those in the proof of Theorem 1 one can easily show that if $\mathrm{E}\,|Y| < \infty$,

$$\hat{g}(x) \xrightarrow{n} g(x) \text{ in probability for almost all } x \in R\,.$$

Since also, see Greblicki and Pawlak [7]

$$\hat{f}(x) \xrightarrow{n} f(x) \text{ in probability} \tag{19}$$

for almost all $x \in R$, the proof has been completed.

The next theorem can be verified by using Hoeffding's [8] inequality, Lemma 3 and Lemma 4.

THEOREM 6. *Let* $|Y| \leqq \gamma < \infty$ *almost surely. Let* $f, g \in L_p$, $p > 1$. *If, in addition to* (1)

$$\sum_{n=1}^{\infty} \exp\,(-\alpha n/N^{1/2}(n)) < \infty$$

all $\alpha > 0$, *then*

$$\hat{m}(x) \xrightarrow{n} m(x) \textit{ almost completely for almost all } x \in R\,.$$

The Hermite series density estimate seems to be useful if it is suspected that the unknown density to be estimated is close to a normal density, see Tapia and Thompson [17]. As far as the regression estimate is concerned the Hermite series estimate seems to be adequate if density f is not too far from normal density and moreover the regression function is close to polynomial one with few nonzero coefficients.

6. Convergence rate

For an integer s, let us denote

$$\tau_s(\cdot\,; f) = (x - d/dx)^s f(x)\,.$$

Walter [20] has shown that if $\tau_s(\cdot\,; g)$ and g are square integrable,

$$|a_k| \leqq |\alpha_{k+s}|/(k+1)^{s/2}\,, \tag{20}$$

where α_k is the kth coefficient of the expansion of $\tau_s(\cdot\,;g)$ in the Hermite series. For similar reasons

$$|b_k| \leqq |\beta_{k+s}|/(k+1)^{s/2} \tag{21}$$

where β_k is the kth coefficient of the expansion of $\tau_s(\cdot\,;f)$, provided that f and $\tau_s(\cdot\,;f)$ are square integrable. By Lemmas 2, 3, 4 and (20) and (21) we have

THEOREM 7. *Let* $\mathrm{E}\,Y^2 < \infty$. *Let* $f, g, \tau_s(\cdot\,;f)$ *and* $\tau_s(\cdot\,;g) \in L_2$. *If*

$$N(n) \sim n^{1/s}\,,$$

then

$$\mathrm{P}\,\{|\hat{m}(x) - m(x)| > \varepsilon\,|m(x)|\} = O(n^{-(2s-1)/2s})$$

and

$$|\hat{m}(x) - m(x)| = O(n^{-(2s-1)/4s}) \textit{ in probability}\,.$$

THEOREM 8. *Let* $|Y| \leqq \gamma < \infty$ *almost surely. Let* $f, g, \tau_s(\cdot\,;f)$ *and* $\tau_s(\cdot\,;g) \in L_2$. *If* $\{N(n)\}$ *is selected as in Theorem* 7, *then*

$$|\hat{m}(x) - m(x)| = O(n^{-(2s-1)/4s} \log n) \textit{ almost completely}\,.$$

7. Conclusion

In the paper we have discussed the pointwise properties of the orthogonal series regression estimates. On the other hand by virtue of the Lebesque dominated convergence theorem on product spaces, see Glick [4], we have the following global result:

COROLLARY. *Let* $|Y| \leqq \gamma < \infty$. *Then, with the conditions of Theorem* 1 *or* 2 *and Theorem* 5 *or* 6

$$\int (\hat{m}(x) - m(x))^2 f(x) w(x) dx \xrightarrow{n} 0$$

in probability or almost surely, respectively. Here w *is positive and bounded weight function.*

Acknowledgement

The authors wish to express their thanks to the referee for his suggestions concerning some improvements.

INSTITUTE OF ENGINEERING CYBERNETICS, TECHNICAL UNIVERSITY OF WROCLAW
CONCORDIA UNIVERSITY

REFERENCES

[1] Bluez, J. and Bosq, D. (1976). Conditions nécessaires et suffisantes de convergence pour une class d'estimateurs de la densité, *C. R. Acad. Sci.*, Paris, **282**, 636-666.

[2] Collomb, G. (1981). Estimation non paramétrique de la regression: revue bibliographique, *Internat. Statist. Rev.*, **49**, 75-93.

[3] Földes, A. and Révész, R. (1974). A general method for density estimation, *Studia Sci. Math. Hungar.*, **9**, 81-92.

[4] Glick, N. (1974). Consistency conditions for probability estimators and integrals for density estimators, *Utilitas Math.*, **5**, 61-74.

[5] Greblicki, W., Krzyzak, A. and Pawlak, M. (1984). Distribution-free pointwise consistency of kernel regression estimates, *Ann. Statist.*, **12**, 1570-1575.

[6] Greblicki, W. and Pawlak, M. (1984). Hermite series estimate of a probability density and its derivatives, *J. Multivariate Anal.*, **15**, 174-182.

[7] Greblicki, W. and Pawlak, M. (1985). Pointwise consistency of the Hermite series density estimate, *Probability and Statistics Letters*, to appear.

[8] Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables, *J. Amer. Statist. Ass.*, **58**, 13-30.

[9] Hunt, R. A. (1968). On the convergence of Fourier series, Orthogonal Expansion and their Continuous Analogues, *Proc. Conf. Edwardsville*, 1967, 235-255, Southern Illinois Univ. Press, Carbondale, I11.

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

[11] Krzyzak, A. and Pawlak, M. (1984). Distribution-free consistency of nonparametric kernel regression estimate and classification, *IEEE Trans. Inform. Theory*, IT-30, 78-81.

[12] Nadaraya, E. A. (1964). On estimating regression, *Theory of Probability and its Applications*, **9**, 141-142.

[13] Sansone, G. (1959). *Orthogonal Functions*, Interscience Publishers Inc.

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

[15] Stone, C. J. (1980). Optimal rates of convergence for non-parametric estimators, *Ann. Statist.*, 8, 1348-1360.

[16] Szegö, G. (1959). Orthogonal Polynomials, *Amer. Math. Soc. Coll. Publ.*

[17] Tapia, R. A. and Thompson, J. R. (1978). *Nonparametric Probability Density Estimation*, The John Hopkins University Press, Baltimore.

[18] Tarter, M. and Kronmal, R. (1970). On multivariate density estimates based on orthogonal expansions, *Ann. Math. Statist.*, **41**, 718-722.

[19] Wahba, G. (1975). Optimal convergence properties of variable knot, kernel and orthogonal series methods for density estimation, *Ann. Statist.*, **3**, 15-29.

[20] Walter, G. G. (1977). Properties of Hermite series estimation of probability density, *Ann. Statist.*, **5**, 1258-1264.

[21] Watson, G. S. (1964). Smooth regression analysis, *Sankhya*, A, **26**, 359-372.

[22] Zygmund, A. (1959). *Trigonometric series* II, Cambridge University Press.