APPLICATIONES MATHEMATICAE
25,1 (1998), pp. 73–83

W. POPIŃSKI (Warszawa)

CONSISTENCY OF TRIGONOMETRIC AND POLYNOMIAL REGRESSION ESTIMATORS

Abstract. The problem of nonparametric regression function estimation is considered using the complete orthonormal system of trigonometric functions or Legendre polynomials e_k, $k = 0, 1, \ldots$, for the observation model $y_i = f(x_i) + \eta_i$, $i = 1, \ldots, n$, where the η_i are independent random variables with zero mean value and finite variance, and the observation points $x_i \in [a, b]$, $i = 1, \ldots, n$, form a random sample from a distribution with density $\varrho \in L^1[a, b]$. Sufficient and necessary conditions are obtained for consistency in the sense of the errors $\|f - \widehat{f}_N\|$, $|f(x) - \widehat{f}_N(x)|$, $x \in [a, b]$, and $E\|f - \widehat{f}_N\|^2$ of the projection estimator $\widehat{f}_N(x) = \sum_{k=0}^{N} \widehat{c}_k e_k(x)$ for $\widehat{c}_0, \widehat{c}_1, \ldots, \widehat{c}_N$ determined by the least squares method and $f \in L^2[a, b]$.

1. Introduction. Let y_i, $i = 1, \ldots, n$, be observations at points $x_i \in [a, b]$ according to the model $y_i = f(x_i) + \eta_i$, where $f \in L^2[a, b]$ is an unknown function, η_i, $i = 1, \ldots, n$, are independent identically distributed random variables with zero mean value and finite variance $\sigma_\eta^2 > 0$, and x_i, $i = 1, \ldots, n$, form a random sample from an absolutely continuous distribution with density $\varrho \in L^1[a, b]$. It is also assumed that the random variable $\omega = (x_1, \ldots, x_n)$ is independent of the observation error vector $\eta = (\eta_1, \ldots, \eta_n)$.

Let the functions e_k, $k = 0, 1, \ldots$, constitute a complete orthonormal system in the space $L^2[a, b]$. Any function $f \in L^2[a, b]$ can then be represented by its Fourier series

$$f = \sum_{k=0}^{\infty} c_k e_k, \quad \text{where} \quad c_k = \int_a^b f e_k, \ k = 0, 1, \ldots$$

We assume that the functions e_k, $k = 0, 1, 2, \ldots$, are analytic in (a, b). As

1991 *Mathematics Subject Classification*: 62G07, 62F12.

Key words and phrases: orthonormal system, least squares method, regression, consistent estimator.

an estimator of the vector $c^N = (c_0, c_1, \ldots, c_N)^T$ of coefficients, for a fixed N, we take the vector $\widehat{c}^N = (\widehat{c}_0, \widehat{c}_1, \ldots, \widehat{c}_N)^T$ obtained by the least squares method:

$$\widehat{c}^N = \arg \min_{a \in \mathbb{R}^{N+1}} \frac{1}{n} \sum_{i=1}^{n} (y_i - \langle a, e^N(x_i) \rangle)^2,$$

where $e^N(x) = (e_0(x), e_1(x), \ldots, e_N(x))^T$.

The vector $\widehat{c}^N$ can be uniquely determined with probability one as the solution of the normal equations

$$\widehat{c}^N = G_n^{-1} g_n, \tag{1}$$

where

$$G_n = \frac{1}{n} \sum_{i=1}^{n} e^N(x_i) e^N(x_i)^T, \qquad g_n = \frac{1}{n} \sum_{i=1}^{n} y_i e^N(x_i),$$

since according to the results presented in the author's earlier work (see Lemma 2.2 of [7]) the matrices G_n are almost surely positive definite for $N+1 \le n$, when x_i, $i = 1, \ldots, n$, form a random sample from a distribution with density $\varrho \in L^1[a,b]$.

Thus, we can study asymptotic properties of the projection estimator of the regression function f defined by the formula

$$\widehat{f}_N(x) = \sum_{k=0}^{N} \widehat{c}_k e_k(x).$$

In Sections 2 and 3 we will consider the case when either $a = 0, b = 2\pi$ or $a = -1, b = 1$ and e_k, $k = 0, 1, 2, \ldots$, are the well-known complete orthonormal system of trigonometric functions in $L^2[0, 2\pi]$ or Legendre polynomials in $L^2[-1, 1]$, respectively (see [10]). The results obtained give sufficient conditions for the consistency in the sense of L^2-norm and uniform pointwise consistency of the estimators.

In Section 4 a necessary condition for consistency in the sense of the integrated mean-square error is given in the case where the projection estimators are obtained using any orthonormal system of analytic functions.

In [5] Lugosi and Zeger proved general results concerning universal consistency of trigonometric and polynomial estimators of the regression function $E(Y \mid X = x)$ in the case where pairs of random variables (X_i, Y_i), $i = 1, \ldots, n$, are observed. The estimators considered in [5] are, however, determined by minimizing the empirical error

$$\frac{1}{n} \sum_{i=1}^{n} \Big| Y_i - \sum_{k=0}^{N(n)} a_k e_k(X_i) \Big|^p, \qquad p \ge 1,$$

under the constraint $\sum_{k=0}^{N(n)} |a_k| \le \beta_n$, where, to obtain consistency, $N(n)$ and β_n have to grow, but not too rapidly, as the sample size n grows.

Certain results concerning asymptotic properties of the polynomial regression function estimators in the case of a fixed-point design are presented in [1].

Another approach to nonparametric regression function estimation using polynomials, together with recent results concerning its consistency and comparison with other estimation methods are presented in the monograph [2].

According to the Jackson theorem [4] for any continuous 2π-periodic function (i.e. for $f \in C[0, 2\pi]$ satisfying $f(0) = f(2\pi)$) the following inequality is valid:

$$d_N^{\mathrm{T}}(f) = \inf_{T \in T_N} \sup_{0 \le s \le 2\pi} |f(s) - T(s)| \le 12\omega(1/l, f), \tag{2}$$

where $N = 2l$, $l = 1, 2, \ldots$, $T_N = \operatorname{span}\{1, \sin(s), \cos(s) \ldots, \sin(ls), \cos(ls)\}$ and $\omega(\delta, f)$ for $\delta > 0$ denotes the modulus of continuity of the function f (see [4]). A similar theorem on uniform polynomial approximation (e.g. Theorem 3.11 of [6]) implies that for $f \in C[-1, 1]$,

$$d_N^{\mathrm{P}}(f) = \inf_{P \in P_N} \sup_{-1 \le s \le 1} |f(s) - P(s)| \le 6\omega(1/N, f), \tag{3}$$

where $N = 1, 2, \ldots$, and P_N denotes the set of algebraic polynomials of degree N.

2. L^2-norm consistency for square-integrable regression functions.

First, we prove the following two lemmas.

LEMMA 2.1. (a) *If e_k, $k = 0, 1, \ldots$, denote the trigonometric functions forming a complete orthonormal system in $L^2[0, 2\pi]$, then for $N = 2l$, $l = 0, 1, \ldots$,*

$$\sup_{0 \le s \le 2\pi} \|e^N(s)\|^2 = \frac{N+1}{2\pi}.$$

(b) *If e_k, $k = 0, 1, \ldots$, denote the Legendre polynomials forming a complete orthonormal system in $L^2[-1, 1]$, then for $N = 0, 1, \ldots$,*

$$\sup_{-1 \le s \le 1} \|e^N(s)\|^2 \le \frac{(N+1)^2}{2}.$$

Proof. For the trigonometric system we have $e_0(s) = 1/\sqrt{2\pi}$, $e_{2l-1}(s) = \sin(ls)/\sqrt{\pi}$, $e_{2l}(s) = \cos(ls)/\sqrt{\pi}$, $l = 1, 2, \ldots$; accordingly for $N = 2l$ and $s \in [0, 2\pi]$,

$$\|e^N(s)\|^2 = \sum_{k=0}^{N} e_k^2(s) = \frac{1}{2\pi} + \frac{1}{\pi} \sum_{j=1}^{l} [\sin^2(js) + \cos^2(js)] = \frac{N+1}{2\pi}.$$

For the Legendre polynomials we have the inequalities $|e_k(s)| \le \sqrt{(2k+1)/2}$ for $k = 0, 1, \ldots$ and $s \in [-1, 1]$ (see [10]). Hence,

$$\|e^N(s)\|^2 = \sum_{k=0}^{N} e_k^2(s) \le \frac{1}{2}\sum_{k=0}^{N}(2k+1) = \frac{(N+1)^2}{2}. \ \blacksquare$$

LEMMA 2.2. *If the vector* $\widehat{c}^N = (\widehat{c}_0, \widehat{c}_1, \ldots, \widehat{c}_N)$ *of Fourier coefficient estimators is obtained from* (1), *then*

$$E_\eta \int_a^b (f - \widehat{f}_N)^2 = \sigma_\eta^2 \frac{\operatorname{Tr} G_n^{-1}}{n} + p_N + \|G_n^{-1} a_N\|^2,$$

where

$$a_N = \frac{1}{n}\sum_{i=1}^{n} r_N(x_i) e^N(x_i), \qquad r_N = \sum_{k=N+1}^{\infty} c_k e_k, \qquad p_N = \sum_{k=N+1}^{\infty} c_k^2.$$

Proof. It is easy to see that

$$E_\eta \int_a^b (f - \widehat{f}_N)^2 = E_\eta \|c^N - \widehat{c}^N\|^2 + p_N$$

so we must calculate the first term in the above formula. Since $f(x) = \langle e^N(x), c^N\rangle + r_N(x)$ we have in view of the definitions in (1),

$$\widehat{c}^N = c^N + G_n^{-1}\Big(\frac{1}{n}\sum_{i=1}^{n} r_N(x_i) e^N(x_i)\Big) + G_n^{-1}\Big(\frac{1}{n}\sum_{i=1}^{n} \eta_i e^N(x_i)\Big).$$

Now, we easily obtain the equalities

$$\begin{aligned} E_\eta \|c^N - \widehat{c}^N\|^2 &= \|G_n^{-1} a_N\|^2 + \frac{\sigma_\eta^2}{n^2}\sum_{i=1}^{n} e^N(x_i)^T G_n^{-1} G_n^{-1} e^N(x_i) \\ &= \|G_n^{-1} a_N\|^2 + \frac{\sigma_\eta^2}{n^2}\operatorname{Tr}\Big(\sum_{i=1}^{n} e^N(x_i) e^N(x_i)^T G_n^{-1} G_n^{-1}\Big) \\ &= \|G_n^{-1} a_N\|^2 + \frac{\sigma_\eta^2}{n}\operatorname{Tr} G_n^{-1}. \ \blacksquare \end{aligned}$$

Let $\lambda_n(\omega)$ denote the smallest eigenvalue of the matrix $G_n(\omega)$ defined in (1). It is clearly a measurable random variable since it can be defined as $\lambda_n(\omega) = \inf_{m=1,2,\ldots}\langle G_n(\omega) z_m, z_m\rangle$, where the points z_m, $m = 1, 2, \ldots$, form a dense subset of the unit sphere $S_{N+1} = \{x \in \mathbb{R}^{N+1} : \|x\| = 1\}$.

Further, the (norm-one) eigenvector $y_n(\omega)$ of the matrix $G_n(\omega)$ corresponding to the eigenvalue $\lambda_n(\omega)$ is also a measurable random variable since it is defined as a solution of the linear equation $(G_n(\omega) - \lambda_n(\omega) I) y_n(\omega) = 0$.

It is also easy to see that the elements of the matrix G_n converge in the mean-square sense to the quantities

$$g_{kl} = \frac{1}{n} E_\omega \sum_{i=1}^{n} e_k(x_i) e_l(x_i) = \int_a^b e_k e_l \varrho \tag{4}$$

since

$$E_\omega \Big(\frac{1}{n} \sum_{i=1}^{n} e_k(x_i) e_l(x_i) - g_{kl} \Big)^2 \le \frac{1}{n} \int_a^b e_k^2 e_l^2 \varrho$$

for $k, l = 0, 1, \ldots, N$. Putting $G^N = E_\omega G_n(\omega)$ we immediately obtain

$$E_\omega \|G_n - G^N\|^2 \le \frac{1}{n} \sum_{k=0}^{N} \sum_{l=0}^{N} \int_a^b e_k^2 e_l^2 \varrho = \frac{1}{n} \int_a^b \Big(\sum_{k=0}^{N} e_k^2 \Big)^2 \varrho$$

and consequently since $\|y_n\| = 1,\ n = 1, 2, \ldots,$ we have

$$\begin{aligned} E_\omega(\lambda_n - \langle G^N y_n, y_n \rangle)^2 &= E_\omega(\langle G_n y_n, y_n \rangle - \langle G^N y_n, y_n \rangle)^2 \\ &\le E_\omega \|G_n - G^N\|^2 \|y_n\|^4 \le \frac{1}{n} \int_a^b \|e^N\|^4 \varrho. \end{aligned} \tag{5}$$

Furthermore, if the density ϱ satisfies the condition $\varrho \ge c > 0$, then taking into account (4) and putting $y_n = (y_{n0}, y_{n1}, \ldots, y_{nN})^T$ we obtain

$$\begin{aligned} \langle G^N y_n, y_n \rangle &= \sum_{k=0}^{N} \sum_{l=0}^{N} g_{kl} y_{nk} y_{nl} = \int_a^b \Big(\sum_{k=0}^{N} y_{nk} e_k \Big)^2 \varrho \\ &\ge c \int_a^b \Big(\sum_{k=0}^{N} y_{nk} e_k \Big)^2 = c \|y_n\|^2 = c > 0. \end{aligned} \tag{6}$$

Now, applying the Chebyshev inequality we conclude in view of (5) that in that case

$$P_\omega(|\lambda_n - \langle G^N y_n, y_n \rangle| > c/2) \le \frac{4}{nc^2} \int_a^b \|e^N\|^4 \varrho$$

and from (6) since $\lambda_n \ge 0$ we finally have for $\varrho \ge c > 0$,

$$P_\omega(0 \le \lambda_n < c/2) \le \frac{4}{nc^2} \int_a^b \|e^N\|^4 \varrho \le \frac{4}{nc^2} \sup_{a \le s \le b} \|e^N(s)\|^4. \tag{7}$$

Since the matrix G_n is symmetric and almost surely positive definite for $N + 1 \le n$, by Lemma 2 of [9] we have $\|G_n^{-1} a_N\|^2 \le \lambda_n^{-1} \langle G_n^{-1} a_N, a_N \rangle$ and using Lemma 3 of [9] we also have $\langle G_n^{-1} a_N, a_N \rangle \le n^{-1} \sum_{i=1}^{n} r_N(x_i)^2$.

Consequently, according to Lemma 2.2 the following inequality is almost surely true for $N+1 \le n$:

$$E_\eta \|f - \widehat{f}_N\|_2^2 \le \sigma_\eta^2 \lambda_n^{-1} \frac{N+1}{n} + p_N + \lambda_n^{-1} \frac{1}{n} \sum_{i=1}^{n} r_N(x_i)^2, \tag{8}$$

where $\| * \|_2$ denotes the norm in $L^2[a,b]$. Furthermore, taking into account (7) and (8) we see that for $N+1 \le n$ the inequality

$$E_\eta \|f - \widehat{f}_N\|_2^2 \le \frac{2}{c} \left[\sigma_\eta^2 \frac{N+1}{n} + \frac{1}{n} \sum_{i=1}^{n} r_N(x_i)^2 \right] + p_N \tag{9}$$

holds except for $\omega \in A_n \subset [a,b]^n$, where

$$P_\omega(A_n) \le \frac{4}{nc^2} \sup_{a \le s \le b} \|e^N(s)\|^4.$$

Since

$$\frac{1}{n} E_\omega \sum_{i=1}^{n} r_N(x_i)^2 = \int_a^b r_N^2 \varrho \le D p_N \quad \text{for } \varrho \le D,$$

we also see that for a bounded density ϱ,

$$P_\omega \left(\frac{1}{n} \sum_{i=1}^{n} r_N(x_i)^2 > p_N^{1/2} \right) \le D p_N^{1/2}.$$

The last inequality together with (9) implies that for $N+1 \le n$, $\varepsilon > 0$ and a density ϱ satisfying $D \ge \varrho \ge c > 0$,

$$P(\|f - \widehat{f}_N\|_2 > \varepsilon) \le \frac{4M(e^N)}{nc^2} + D p_N^{1/2} + \frac{2}{\varepsilon^2 c} \left[\sigma_\eta^2 \frac{N+1}{n} + p_N^{1/2} + \frac{c}{2} p_N \right],$$

where $M(e^N) = \sup_{a \le s \le b} \|e^N(s)\|^4$. According to Lemma 2.1 for the trigonometric system in $L^2[0, 2\pi]$ and $N = 2l$ we have $M(e^N) = (N+1)^2/4$, and for the system of Legendre polynomials in $L^2[-1,1]$ we have $M(e^N) \le (N+1)^4/4$. Thus, for $N+1 \le n$, $\varepsilon > 0$ and $D \ge \varrho \ge c > 0$,

$$P(\|f - \widehat{f}_N\|_2 > \varepsilon) \le \frac{(N+1)^r}{nc^2} + D p_N^{1/2} + \frac{2}{\varepsilon^2 c} \left[\sigma_\eta^2 \frac{N+1}{n} + p_N^{1/2} + \frac{c}{2} p_N \right],$$

where $r = 4$ in the case of polynomial regression and $r = 2$, $N = 2l$ for trigonometric regression. The above conclusions allow us to formulate the following theorems.

THEOREM 2.1. *If the density $\varrho \in L^1[0, 2\pi]$ satisfies $D \ge \varrho \ge c > 0$ and the sequence of even natural numbers $N(n)$, $n = 1, 2, \ldots$, satisfies*

$$\lim_{n\to\infty} N(n) = \infty, \qquad \lim_{n\to\infty} N(n)^2/n = 0,$$

then the trigonometric projection estimator $\widehat{f}_{N(n)}$ *of the regression function* $f \in L^2[0, 2\pi]$ *is consistent in the* L^2*-norm, i.e.*

$$\lim_{n\to\infty} \|f - \widehat{f}_{N(n)}\|_2 \overset{\mathrm{p}}{=} 0.$$

THEOREM 2.2. *If the density* $\varrho \in L^1[-1, 1]$ *satisfies* $D \geq \varrho \geq c > 0$ *and the sequence of natural numbers* $N(n)$, $n = 1, 2, \ldots,$ *satisfies*

$$\lim_{n\to\infty} N(n) = \infty, \qquad \lim_{n\to\infty} N(n)^4/n = 0,$$

then the polynomial projection estimator $\widehat{f}_{N(n)}$ *of the regression function* $f \in L^2[-1, 1]$ *is consistent in the* L^2*-norm, i.e.*

$$\lim_{n\to\infty} \|f - \widehat{f}_{N(n)}\|_2 \overset{\mathrm{p}}{=} 0.$$

3. Uniform pointwise consistency of the projection estimator. In order to obtain the results concerning uniform pointwise consistency of the projection estimators considered we shall make use of an inequality proved in [9] in the case $\widehat{c}^N = G_n^{-1} g_n$ and $f \in C[a, b]$,

$$E_\eta(f(x) - \widehat{f}_N(x))^2 \leq \sigma_\eta^2 B_N^2 \lambda_n^{-1} \frac{N+1}{n} + [2(N+1)B_N^2 \lambda_n^{-1} + 2] d_N(f)^2$$

for $x \in [a, b]$, where $d_N(f) = d_N^{\mathrm{T}}(f)$ or $d_N(f) = d_N^{\mathrm{P}}(f)$, and B_k, $k = 0, 1, 2, \ldots,$ form a non-decreasing sequence of bounds with $B_k \geq \sup_{a\leq s\leq b} |e_k(s)|$. Hence, in view of (7) and Lemma 2.1 we then have for $N + 1 \leq n$, $\varrho \geq c > 0$ and $x \in [a, b]$,

$$E_\eta(f(x) - \widehat{f}_N(x))^2 \leq \sigma_\eta^2 \frac{2(N+1)B_N^2}{nc} + \frac{4(N+1)B_N^2 + 2c}{c} d_N(f)^2 \tag{10}$$

except for $\omega \in A_n \subset [a, b]^n$, where

$$P_\omega(A_n) \leq \frac{(N+1)^r}{nc^2}, \qquad r = 2, 4.$$

Let us now consider the case of trigonometric regression. Then we have $B_k^2 = 1/\pi$, $k = 0, 1, \ldots$ If the regression function is 2π-periodic and satisfies the Lipschitz condition with exponent $0 < \alpha \leq 1$, then $\omega(\delta, f) \leq L\delta^\alpha$, where $L > 0$, and in view of (10) and (2) we obtain for $\varrho \geq c > 0$, $N = 2l$, $N + 1 \leq n$ and $\varepsilon > 0$ the following estimate:

$$P(|f(x) - \widehat{f}_N(x)| > \varepsilon) \leq \frac{(N+1)^2}{nc^2} + \frac{2}{c\pi\varepsilon^2}\left[\sigma_\eta^2 \frac{N+1}{n} + \frac{[2(N+1) + c\pi](12L2^\alpha)^2}{N^{2\alpha}}\right],$$

valid for $x \in [0, 2\pi]$. Hence, we can formulate the following theorem.

Theorem 3.1. *If the density* $\varrho \in L^1[0,2\pi]$ *satisfies* $\varrho \ge c > 0$ *and the sequence of even natural numbers*, $N(n)$, $n = 1,2,\ldots$, *satisfies*

$$\lim_{n\to\infty} N(n) = \infty, \qquad \lim_{n\to\infty} N(n)^2/n = 0,$$

then the trigonometric projection estimator $\widehat{f}_{N(n)}$ *of the* 2π*-periodic regression function* f *satisfying the Lipschitz condition with exponent* $1/2 < \alpha \le 1$ *is uniformly pointwise consistent in* $[0,2\pi]$, *i.e.*

$$\lim_{n\to\infty} \widehat{f}_{N(n)}(x) \overset{\mathrm{P}}{=} f(x) \quad \text{uniformly in } [0,2\pi].$$

In the case of polynomial regression we have $B_k^2 = (2k+1)/2$, $k = 0,1,\ldots$, (see [10]) so the sequence of bounds B_k is non-decreasing and in view of (10) we obtain for $\varrho \ge c > 0$, $N+1 \le n$ and $\varepsilon > 0$ the following estimate:

$$P(|f(x) - \widehat{f}_N(x)| > \varepsilon) \le \frac{(N+1)^4}{nc^2} + \frac{1}{c\varepsilon^2}\bigg[\sigma_\eta^2 \frac{(N+1)(2N+1)}{n} + [2(N+1)(2N+1) + 2c]d_N^2(f)\bigg],$$

valid for $x \in [-1,1]$. In this case we can formulate the following theorem.

Theorem 3.2. *If the density* $\varrho \in L^1[-1,1]$ *satisfies* $\varrho \ge c > 0$ *and the sequence of natural numbers* $N(n)$, $n = 1,2,\ldots$, *satisfies*

$$\lim_{n\to\infty} N(n) = \infty, \qquad \lim_{n\to\infty} N(n)^4/n = 0,$$

then the polynomial projection estimator $\widehat{f}_{N(n)}$ *of the regression function* $f \in C[a,b]$ *satisfying the condition* $d_N^{\mathrm{P}}(f) = o(N^{-1})$ *is uniformly pointwise consistent in* $[-1,1]$, *i.e.*

$$\lim_{n\to\infty} \widehat{f}_{N(n)}(x) \overset{\mathrm{P}}{=} f(x) \quad \text{uniformly in } [-1,1].$$

4. Necessary condition for convergence of the integrated mean-square error. In this section a theorem giving certain necessary conditions for consistency of the projection estimators of the regression function $f \in L^2[a,b]$ in the sense of the integrated mean-square error $E_\eta E_\omega \|f - \widehat{f}_N\|_2^2$ is proved. It should be noted that this theorem is proved under the assumption (see Section 1) that the functions e_k, $k = 0,1,2,\ldots$, forming a complete orthonormal system in $L^2[a,b]$ are analytic in (a,b).

Theorem 4.1. *Assume that the density* $\varrho \in L^1[a,b]$ *is bounded and* $f \in L^2[a,b]$ *is such that* $c_k \ne 0$ *for infinitely many* k. *Then the projection estimator* $\widehat{f}_{N(n)}$ *is consistent in the sense of the integrated mean-square error* $E_\eta E_\omega \|f - \widehat{f}_{N(n)}\|_2^2$ *only if the sequence of natural numbers*

$N(n)$, $n = 1, 2, \ldots,$ *satisfies*

$$\lim_{n\to\infty} N(n) = \infty, \quad \lim_{n\to\infty} N(n)/n = 0.$$

P r o o f. If the Fourier coefficient estimators are obtained as the solution of the normal equations $\widehat{c}^N = G_n^{-1}(\omega) g_n(\omega, \eta)$, then necessarily $N + 1 \le n$ since for $N + 1 > n$ we have $\det G_n(\omega) = 0$ almost surely (see Lemma 2.1 of [7]).

From Lemma 2.2 it follows immediately that the inequality

$$E_\eta \|f - \widehat{f}_N\|_2^2 \ge \frac{\sigma_\eta^2}{n} \operatorname{Tr} G_n^{-1} + p_N$$

is valid almost surely for $N + 1 \le n$, which implies also that

$$E_\omega E_\eta \|f - \widehat{f}_N\|_2^2 \ge \frac{\sigma_\eta^2}{n} E_\omega \operatorname{Tr} G_n^{-1} + p_N \ge p_N. \tag{11}$$

Hence, since $p_N = \sum_{k=N+1}^\infty c_k^2$ it is easy to see that $N(n) \to \infty$ if we have $E_\omega E_\eta \|f - \widehat{f}_{N(n)}\|_2^2 \to 0$ as $n \to \infty$.

Further, the inequality between the arithmetic and harmonic means implies that

$$\sum_{i=0}^N \lambda_i^{-1} \ge \frac{(N+1)^2}{\sum_{i=0}^N \lambda_i}$$

for $\lambda_i > 0$, $i = 0, 1, \ldots, N$. Consequently, if λ_i, $i = 0, 1, \ldots, N$, denote the eigenvalues of the matrix G_n (which is almost surely positive definite for $N + 1 \le n$) we obtain $\operatorname{Tr} G_n^{-1} \ge (N+1)^2 / \operatorname{Tr} G_n$, which together with Jensen's inequality immediately implies

$$E_\omega \operatorname{Tr} G_n^{-1} \ge E_\omega \frac{(N+1)^2}{\operatorname{Tr} G_n} \ge \frac{(N+1)^2}{E_\omega \operatorname{Tr} G_n}.$$

However, since the density ϱ satisfies $\varrho \le D$, $D > 0$ and $\|e_k\|_2 = 1$, $k = 0, 1, 2, \ldots,$

$$E_\omega \operatorname{Tr} G_n = E_\omega \sum_{k=0}^N \frac{1}{n} \sum_{i=1}^n e_k^2(x_i) = \sum_{k=0}^N \int_a^b e_k^2 \varrho \le D(N+1)$$

and we finally have $E_\omega \operatorname{Tr} G_n^{-1} \ge (N+1)/D$. Together with (11) this implies that for $N + 1 \le n$,

$$E_\omega E_\eta \|f - \widehat{f}_N\|_2^2 \ge \frac{\sigma_\eta^2}{n} E_\omega \operatorname{Tr} G_n^{-1} + p_N \ge \sigma_\eta^2 \frac{(N+1)}{nD}.$$

Hence if $E_\omega E_\eta \|f - \widehat{f}_{N(n)}\|_2^2 \to 0$ as $n \to \infty$, we must have $\lim_{n\to\infty} N(n)/n = 0$. ∎

It is worth remarking that in the case of uniform distribution of the observation points on $[0, 2\pi]$ ($\varrho = 1/(2\pi)$) there exist Fourier coefficient estimators such that the conditions $\lim_{n\to\infty} N(n) = \infty$, $\lim_{n\to\infty} N(n)/n = 0$ are also sufficient for consistency in the sense of the integrated mean-square error of the corresponding trigonometric projection estimator for $f \in L^2[0, 2\pi]$ (see [8]).

5. Conclusions. Theorems similar to 2.1 and 3.1 can also be easily proved in the case of regression functions defined on the d-dimensional cube $Q = [0, 2\pi]^d \subset \mathbb{R}^d$, $d > 1$, and the orthonormal system of trigonometric functions in the space $L^2(Q)$. A certain class of multivariate regression functions for which a theorem analogous to 3.1 holds is characterized in [9]. Moreover, according to Gallant and White [3] the functions of the form

$$s(x) = a_0 + \sum_{|k_\alpha| \le K} a_\alpha \cos(\langle k_\alpha, x\rangle) + b_\alpha \sin(\langle k_\alpha, x\rangle),$$

where $x = (x_1, \ldots, x_d) \in Q$, $k_\alpha = (k_{1\alpha}, \ldots, k_{d\alpha})$, $|k_\alpha| = |k_{1\alpha}| + \ldots + |k_{d\alpha}|$, $k_{i\alpha} = 0, \pm 1, \pm 2, \ldots$, $i = 1, \ldots, d$, $K > 0$, can be represented as a single hidden layer feedforward neural network

$$r(x) = \beta_0 + \sum_{i=1}^{m} \beta_i \psi(\langle \gamma_i, x\rangle + \gamma_{i0})$$

with the cosine-squasher activation function

$$\psi(t) = \begin{cases} 0, & -\infty < t \le -\pi/2, \\ [\cos(t + 3\pi/2) + 1]/2, & -\pi/2 \le t \le \pi/2, \\ 1, & \pi/2 \le t < \infty, \end{cases}$$

and properly chosen vector $(\beta_0, \beta_1, \gamma_1, \gamma_{10}, \ldots, \beta_m, \gamma_m, \gamma_{m0})$ of weights, where $\beta_0, \beta_j, \gamma_{j0} \in \mathbb{R}$, $\gamma_j \in \mathbb{R}^d$, $j = 1, \ldots, m$, $m = m(K)$. Thus, the above mentioned multivariate versions of Theorems 2.1 and 3.1 assure existence of neural network estimators [3] consistent in the L^2-norm or uniformly pointwise consistent for appropriate regression function classes.

References

[1] R. L. Eubank and B. R. Jayasuriya, *The asymptotic average square error for polynomial regression*, Statistics 24 (1993), 311–319.

[2] J. Fan and I. Gijbels, *Local Polynomial Modelling and Its Applications*, Chapman and Hall, London, 1996.

[3] A. R. Gallant and H. White, *There exists a neural network that does not make avoidable mistakes*, in: Proc. Second Annual IEEE Conf. on Neural Networks, San Diego, Calif., IEEE Press, New York, 1988, 657–664.

[4] G. G. Lorentz, *Approximation of Functions*, Holt, Reinehart and Winston, New York, 1966.

[5] G. Lugosi and K. Zeger, *Nonparametric estimation via empirical risk minimization*, IEEE Trans. Inform. Theory IT-41 (3) (1995), 677–687.

[6] G. V. Milovanovič, D. S. Mitrinovič and T. M. Rassias, *Topics on Polynomials: Extremal Problems, Inequalities, Zeros*, World Scientific, Singapore, 1994.

[7] W. Popiński, *On least squares estimation of Fourier coefficients and of the regression function*, Appl. Math. (Warsaw) 22 (1993), 91–102.

[8] —, *On Fourier coefficient estimators consistent in the mean-square sense*, ibid., 275–284.

[9] E. Rafajłowicz, *Nonparametric least-squares estimation of a regression function*, Statistics 19 (1988), 349–358.

[10] G. Sansone, *Orthogonal Functions*, Interscience Publ., New York, 1959.

Waldemar Popiński
Department of Standards
Central Statistical Office
Al. Niepodległości 208
00-925 Warszawa, Poland
E-mail: popinskw@gus.stat.gov.pl

Received on 22.1.1997;
revised version on 23.5.1997