APPLICATIONES MATHEMATICAE
26,3 (1999), pp. 281–291

W. POPIŃSKI (Warszawa)

A NOTE ON ORTHOGONAL SERIES REGRESSION FUNCTION ESTIMATORS

Abstract. The problem of nonparametric estimation of the regression function $f(x) = E(Y \mid X = x)$ using the orthonormal system of trigonometric functions or Legendre polynomials e_k, $k = 0, 1, 2, \ldots$, is considered in the case where a sample of i.i.d. copies (X_i, Y_i), $i = 1, \ldots, n$, of the random variable (X, Y) is available and the marginal distribution of X has density $\varrho \in L^1[a, b]$. The constructed estimators are of the form $\widehat{f}_n(x) = \sum_{k=0}^{N(n)} \widehat{c}_k e_k(x)$, where the coefficients $\widehat{c}_0, \widehat{c}_1, \ldots, \widehat{c}_N$ are determined by minimizing the empirical risk $n^{-1} \sum_{i=1}^{n} (Y_i - \sum_{k=0}^{N} c_k e_k(X_i))^2$. Sufficient conditions for consistency of the estimators in the sense of the errors $E_X |f(X) - \widehat{f}_n(X)|^2$ and $n^{-1} \sum_{i=1}^{n} E(f(X_i) - \widehat{f}_n(X_i))^2$ are obtained.

1. Introduction. Let X and Y be random variables taking their values in $[a, b]$ and $\mathbb{R}$, respectively, with $EY^2 < \infty$, and let X have a distribution with density ϱ. Let $D_n = ((X_1, Y_1), \ldots, (X_n, Y_n))$ be a sample of independent and identically distributed copies of the random variable (X, Y). In the regression estimation problem the aim is to find a function g with small mean squared error $E(g(X) - Y)^2$ on the basis of the available observations D_n. As is well known, if $E|Y| < \infty$ and g is any measurable function one has

$$\begin{aligned} E(g(X) - Y)^2 &= E(f(X) - Y)^2 + E(f(X) - g(X))^2 \\ &= E(f(X) - Y)^2 + \int_a^b (f(x) - g(x))^2 \varrho(x)\, dx, \end{aligned}$$

where $f(x) = E(Y \mid X = x)$. Clearly the mean squared error for g is close

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

Key words and phrases: orthonormal system, empirical risk minimization, nonparametric regression, consistent estimator.

to its minimum if and only if the excess error

$$J(g) = \int_a^b (f(x) - g(x))^2 \varrho(x)\, dx$$

is close to zero. We will study asymptotic properties of the excess error for certain series type estimators, namely, for estimators of the form

$$f_n(x) = \sum_{k=0}^{N(n)} \widehat{c}_k e_k(x),$$

where the functions $e_k,\ k = 0, 1, 2, \ldots,$ constitute an orthonormal system in $L^2[a, b]$ and the coefficients $\widehat{c}_0, \widehat{c}_1, \ldots, \widehat{c}_{N(n)}$ are chosen according to some rule defined in the sequel. In this work we consider the case when either $a = 0, b = 2\pi$ or $a = -1, b = 1$ and $e_k,\ k = 0, 1, 2, \ldots,$ denotes the well-known complete orthonormal system of trigonometric functions in $L^2[0, 2\pi]$ or Legendre polynomials in $L^2[-1, 1]$ (see [6]), respectively.

Lugosi and Zeger [3] proved the following general theorem for series type regression estimators:

Theorem 1.1 (Lugosi and Zeger). *Let $h_k,\ k = 1, 2, \ldots,$ be a sequence of uniformly bounded functions such that the set of all finite linear combinations*

$$\bigcup_{k=1}^{\infty} \Big\{ \sum_{j=1}^{k} a_j h_j(x) : a_1, \ldots, a_k \in \mathbb{R} \Big\}$$

is dense in $L^2([a, b], \mu)$ for any probability measure μ. Let the coefficients $\widehat{a}_1, \ldots, \widehat{a}_{N(n)}$ minimize the empirical error

$$\frac{1}{n} \sum_{i=1}^{n} \Big(Y_i - \sum_{k=1}^{N(n)} a_k h_k(X_i) \Big)^2$$

under the constraint $\sum_{k=1}^{N(n)} |a_k| \le \beta_n$, and define the empirically optimal estimator f_n (of series type) as

$$f_n(x) = \sum_{k=1}^{N(n)} \widehat{a}_k h_k(x).$$

If $N(n)$ and β_n satisfy

$$N(n) \to \infty, \quad \beta_n \to \infty \quad \text{and} \quad n^{-1} N(n) \beta_n^4 \ln(\beta_n) \to 0,$$

as $n \to \infty$, then $J(f_n) \to 0$ in probability, for all distributions of (X, Y) with $EY^2 < \infty$. If, in addition, $\beta_n^4 = o(n^{1-\delta})$ for some $\delta > 0$, then $J(f_n) \to 0$ almost surely, i.e. the estimator f_n is universally consistent.

However, as remarked in [2], obtaining the empirically optimal estimator f_n is difficult if the minimum is not unique. In Section 2 of the present paper it is shown that if the density ϱ (of the marginal distribution of the predictor variable X) satisfies the condition $\varrho \geq c > 0$ we can obtain weakly consistent series type estimators without the necessity of solving the minimization problem described above. In order to construct such estimators one only has to solve a system of linear equations with unique solution, which may also reduce the computation time. Thus, the aim of this work, similarly to [2], is to offer a remedy, at least in certain cases, for the numerical difficulties which appear in obtaining the estimators described in the above theorem.

Other approaches to nonparametric regression function estimation giving weakly and universally consistent estimators are described and briefly discussed in [2].

In Section 3 we examine the asymptotic mean squared prediction error $n^{-1}\sum_{i=1}^{n} E(f(X_i) - \widehat{f}_n(X_i))^2$ of the series type estimators considered, in the case where $Y_i = f(X_i) + \eta_i,\ i = 1,\ldots,n$, and the observation errors η_i are independent of the predictor variables $X_i,\ i = 1,\ldots,n$. Hence, the present work is also intended to complement and extend the results concerning the consistency of the least squares trigonometric and polynomial regression function estimators, obtained by the author in [4], [5]. A similar approach but restricted to less general regression function classes is presented by Vapnik in the monograph [7].

2. Asymptotic excess error. Consider the vector of coefficients $\widehat{c}^N = (\widehat{c}_0, \widehat{c}_1, \ldots, \widehat{c}_N)^T$ determined, for fixed N, by minimizing the empirical risk:

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

where $e^N(x) = (e_0(x), e_1(x), \ldots, e_N(x))^T$. If the functions $e_k,\ k = 0,1,\ldots,$ are orthogonal in $L^2[a,b]$ and analytic in (a,b), then for $N+1 \leq n$ 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

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

This follows from the author's results (see Lemma 2.2 of [4]) yielding that the matrices G_n are almost surely positive definite for $N+1 \leq n$, when $X_i,\ i = 1,\ldots,n$, form a random sample from a distribution with density $\varrho \in L^1[a,b]$.

All these conditions hold for the observation model considered and systems of orthogonal functions $e_k,\ k = 0, 1, \ldots$

Let λ_n denote the smallest eigenvalue of the normal equations matrix G_n defined in (1). It is easy to see that it is a measurable random variable and (see inequality (7) in [5]) for the orthonormal systems considered and a density ϱ satisfying $\varrho \ge c > 0$,

$$P(0 \le \lambda_n < c/2) \le \frac{4}{nc^2} \int\limits_a^b \|e^N(s)\|^4 \varrho(s)\, ds \le \frac{4}{nc^2} M^2(e^N),$$

where $M(e^N) = \sup_{a \le s \le b} \|e^N(s)\|^2$ and $N + 1 \le n$.

According to Lemma 2.1 of [5] for the trigonometric system in $L^2[0, 2\pi]$ and $N = 2l$ we have $M(e^N) = (N+1)/(2\pi)$, while $M(e^N) \le (N+1)^2/2$ for the Legendre system in $L^2[-1, 1]$. Thus, for $N + 1 \le n$ and $\varrho \ge c > 0$, we have

$$P(0 \le \lambda_n < c/2) \le \frac{(N+1)^{2r}}{nc^2}, \tag{2}$$

where $r = 2$ in the Legendre case and $r = 1,\ N = 2l$ in the trigonometric case, respectively.

To prove the main results of this section we need the following lemma.

Lemma 2.1. *If $EY^2 < \infty$ and the density $\varrho \in L^1[0, 2\pi]$ (resp. $\varrho \in L^1[-1, 1]$) satisfies $\varrho \ge c > 0$, then there exist constants $B, C > 0$ such that the solution of the normal equations* (1) *minimizes the empirical risk*

$$\frac{1}{n} \sum_{i=1}^{n} \Big(Y_i - \sum_{k=0}^{N} c_k e_k(X_i) \Big)^2$$

under the constraint $\sum_{k=0}^{N} |c_k| \le B(N+1)^{(r+1)/2}$ for $N + 1 \le n$ and $D_n \notin A_n$, where $P(D_n \in A_n) \le C(N+1)^{2r}/n$, $r = 1$, $N = 2l$ in the case of trigonometric functions, and $r = 2$ in the case of Legendre polynomials.

Proof. First observe that according to (1),

$$\begin{aligned} \|\widehat{c}^N\| &= \|G_n^{-1} g_n\| \le \|G_n^{-1}\| \cdot \|g_n\| \le \lambda_n^{-1} \|g_n\| \\ &\le \lambda_n^{-1} (\|g_n - g^N\| + \|g^N\|), \end{aligned} \tag{3}$$

where $g^N = Eg_n = (EYe_0(X), EYe_1(X), \ldots, EYe_N(X))^T$, and furthermore

$$\begin{aligned} \|g^N\|^2 &= \sum_{k=0}^{N} (EYe_k(X))^2 \le \sum_{k=0}^{N} EY^2 Ee_k^2(X) \le EY^2 E \sum_{k=0}^{N} e_k^2(X) \\ &\le M(e^N) EY^2 \le \frac{(N+1)^r}{2} EY^2. \end{aligned} \tag{4}$$

Similarly we obtain

$$\begin{aligned} E\|g_n - g^N\|^2 &= \sum_{k=0}^{N} E\bigg[\frac{1}{n}\sum_{i=1}^{n}(Y_i e_k(X_i) - EYe_k(X))\bigg]^2 \\ &= \sum_{k=0}^{N} \frac{1}{n} E(Ye_k(X) - EYe_k(X))^2 \le \frac{1}{n}\sum_{k=0}^{N} E(Ye_k(X))^2 \\ &\le \frac{1}{n} EY^2 \sum_{k=0}^{N} e_k^2(X) \le \frac{1}{n} M(e^N) EY^2 \le \frac{(N+1)^r}{2n} EY^2, \end{aligned}$$

and from the Chebyshev inequality it follows immediately that

$$P(\|g_n - g^N\| > (N+1)^{-r/2}) \le \frac{(N+1)^{2r}}{2n} EY^2. \tag{5}$$

From (2)–(5) we see that for $N+1 \le n$, $\varrho \ge c > 0$, and appropriately chosen constant $B > 0$, the inequality

$$\|\widehat{c}^N\| \le \frac{2}{c}\bigg[\frac{(N+1)^{r/2}}{\sqrt{2}}(EY^2)^{1/2} + \frac{1}{(N+1)^{r/2}}\bigg] \le B(N+1)^{r/2} \tag{6}$$

holds except for D_n belonging to a set $A_n \subset \mathbb{R}^{2n}$, where

$$P(D_n \in A_n) \le \frac{(N+1)^{2r}}{n}\bigg(\frac{1}{c^2} + \frac{EY^2}{2}\bigg),$$

with $r = 1$, $N = 2l$ in the trigonometric case, and $r = 2$ in the Legendre case. It further follows from the Schwarz inequality that then we also have

$$\sum_{k=0}^{N} |\widehat{c}_k| \le (N+1)^{1/2}\|\widehat{c}^N\| \le B(N+1)^{(r+1)/2} \tag{7}$$

except for $D_n \in A_n$, where $P(D_n \in A_n) \le C(N+1)^{2r}/n$, $B, C > 0$.

By their definition the coefficients $\widehat{c}_0, \widehat{c}_1, \ldots, \widehat{c}_N$ minimize the empirical risk $n^{-1}\sum_{i=1}^{n}(Y_i - \sum_{k=0}^{N} c_k e_k(X_i))^2$ over $(c_0, c_1, \ldots, c_N)^T \in \mathbb{R}^{N+1}$ and consequently according to (7) for $D_n \notin A_n$ they also minimize this risk under the constraint $\sum_{k=0}^{N(n)} |c_k| \le B(N+1)^{(r+1)/2}$, which proves the lemma. ■

For any absolutely continuous probability measure μ the set of all finite linear combinations of trigonometric functions or Legendre polynomials is dense in $L^2([0,2\pi],\mu)$ or $L^2([-1,1],\mu)$, respectively, which follows from the fact that the set of continuous functions of compact support is dense in those function spaces [2]. Thus, if we define our regression function estimator by the formula

$$\widehat{f}_n(x) = \sum_{k=0}^{N(n)} \widehat{c}_k e_k(x), \tag{8}$$

then the property of the coefficients $(\widehat{c}_0, \widehat{c}_1, \ldots, \widehat{c}_N)^T$ proved in Lemma 2.1 suggests that upon imposing appropriate conditions on the sequence of integers $N(n)$ we can use Theorem 1.1 to prove the weak consistency of the estimator.

Let us first consider the case of a trigonometric series estimator.

THEOREM 2.1. *If* $EY^2 < \infty$, *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*

$$\lim_{n\to\infty} N(n) = \infty, \qquad \lim_{n\to\infty} \frac{N(n)^5 \ln N(n)}{n} = 0,$$

then the trigonometric series estimator

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

of the regression function with coefficients $\widehat{c}_0, \widehat{c}_1, \ldots, \widehat{c}_{N(n)}$ *minimizing the empirical error*

$$\frac{1}{n} \sum_{i=1}^{n} \Big(Y_i - \sum_{k=0}^{N(n)} c_k e_k(X_i) \Big)^2$$

is weakly consistent, *i.e.* $J(\widehat{f}_n) \xrightarrow{\mathrm{P}} 0$ *as* $n \to \infty$.

Proof. By Lemma 2.1 for $N = 2l$, $N + 1 \le n$,

$$\sum_{k=0}^{N} |\widehat{c}_k| \le B(N+1)$$

except for $D_n \in A_n$, where $P(D_n \in A_n) \le C(N+1)^2/n$.

Putting $\beta_n = B(N(n)+1)$, where the sequence of even integers $N(n)$, $n = 1, 2, \ldots$, satisfies $N(n) \to \infty$, $N(n)^5 \ln N(n)/n \to 0$, we have $\beta_n \to \infty$ and $(N(n)+1)\beta_n^4 \ln \beta_n / n \to 0$ as $n \to \infty$, so for the estimator f_n defined in Theorem 1.1 we have $J_n(f_n) \to 0$ in probability. Since for the sequence $N(n)$ satisfying the above conditions we also have $N(n)^2/n \to 0$ and consequently $P(D_n \in A_n) \to 0$ it is easy to see that $J(\widehat{f}_n) \to 0$ in probability, which completes the proof. ■

Let us remark that we can use bases other than the Legendre polynomials to construct the polynomial series estimator (8). In fact the estimator (8) does not change if we use the vector function $h^N(x) = Ae^N(x)$, where A is a nonsingular matrix, instead of $e^N(x)$ for constructing it.

For polynomial series estimators the following theorem holds.

THEOREM 2.2. *If* $EY^2 < \infty$, *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*

$$\lim_{n\to\infty} N(n) = \infty, \qquad \lim_{n\to\infty} \frac{N(n)^9 \ln N(n)}{n} = 0,$$

then the polynomial series estimator

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

of the regression function with coefficients $\widehat{c}_0, \widehat{c}_1, \ldots, \widehat{c}_{N(n)}$ *minimizing the empirical error*

$$\frac{1}{n}\sum_{i=1}^{n}\Big(Y_i - \sum_{k=0}^{N(n)} c_k e_k(X_i)\Big)^2$$

is weakly consistent, *i.e.* $J(\widehat{f}_n) \xrightarrow{\mathrm{p}} 0$ *as* $n \to \infty$.

Proof. We apply the same technique as for Theorem 2.1. However, since the Legendre polynomials forming an orthonormal system in $L^2[-1,1]$ are not uniformly bounded we have to change the basis used to construct the estimator in order to be able to use Theorem 1.1. We can represent the polynomial series estimator $\widehat{f}_n$ using the basis of polynomials $p_k = (2k+1)^{-1/2} e_k$, $k = 0, 1, 2, \ldots$, which are uniformly bounded [6]; for this basis the coefficients $\widehat{d}_0, \widehat{d}_1, \ldots, \widehat{d}_N$ globally minimizing the empirical risk satisfy $\widehat{d}_k = \sqrt{(2k+1)}\,\widehat{c}_k$, $k = 0, 1, \ldots, N$. Consequently, by Lemma 2.1 for $N + 1 \le n$ we obtain

$$\begin{aligned}\sum_{k=0}^{N} |\widehat{d}_k| = \sum_{k=0}^{N} \sqrt{2k+1}\,|\widehat{c}_k| &\le \sqrt{2N+1} \sum_{k=0}^{N} |\widehat{c}_k| \\ &\le \sqrt{2}\,(N+1)^{1/2} B (N+1)^{3/2} \le \sqrt{2}\, B (N+1)^2,\end{aligned}$$

except for $D_n \in A_n$, where $P(D_n \in A_n) \le C(N+1)^4/n$.

Now, putting $\beta_n = \sqrt{2}\, B(N(n)+1)^2$, where the sequence of integers $N(n), n = 1, 2, \ldots$, satisfies $N(n) \to \infty$, $N(n)^9 \ln N(n)/n \to 0$, we have $\beta_n \to \infty$ and $(N(n)+1)\beta_n^4 \ln \beta_n / n \to 0$ as $n \to \infty$. Since we then also have $P(D_n \in A_n) \to 0$ Theorem 1.1 yields that $J(\widehat{f}_n) \to 0$ in probability. ■

3. Asymptotic mean squared prediction error. In this section we consider the special case of our observation model when $Y_i = f(X_i) + \eta_i$, $i = 1, \ldots, n$, where $f \in L^2([a,b], \mu)$ is an unknown function and η_i, $i = 1, \ldots, n$, are independent identically distributed random variables with zero mean value and finite variance $\sigma_\eta^2 > 0$. We assume that the random variable $\omega = (X_1, \ldots, X_n)$ is independent of the observation errors $\eta = (\eta_1, \ldots, \eta_n)$. As in the previous section we consider series type regression function estimators

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

Define the mean squared prediction error by

$$R_{nN} = \frac{1}{n} E_\omega E_\eta \sum_{i=1}^{n} (f(X_i) - \widehat{f}_n(X_i))^2.$$

We prove the following theorem concerning consistency in the sense of the error R_{nN} of the series type estimators considered and next we show that it has interesting consequences.

THEOREM 3.1. *If the points $X_1, \ldots, X_n$ form a random sample from an absolutely continuous distribution μ with density $\varrho \in L^1[0, 2\pi]$* (*resp. $\varrho \in L^1[-1, 1]$*) *and the sequence of natural numbers $N(n)$, $n = 1, 2, \ldots,$ satisfies*

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

then the trigonometric (*resp. polynomial*) *series estimator $\widehat{f}_n$ of the regression function $f \in L^2([0, 2\pi], \mu)$* (*resp. $f \in L^2([-1, 1], \mu)$*) *is consistent in the sense of the mean squared prediction error*, *i.e.*

$$\lim_{n\to\infty} E_\omega E_\eta \frac{1}{n} \sum_{i=1}^{n} (f(X_i) - \widehat{f}_n(X_i))^2 = 0.$$

Proof. The standard squared bias plus variance decomposition with respect to the η variable yields

$$R_{nN} = \frac{1}{n} E_\omega \sum_{i=1}^{n} (f(X_i) - E_\eta \widehat{f}_n(X_i))^2 + \frac{1}{n} E_\omega \sum_{i=1}^{n} E_\eta (\widehat{f}_n(X_i) - E_\eta \widehat{f}_n(X_i))^2.$$

Taking into account (1) we obtain for $N + 1 \le n$,

$$\begin{aligned}\frac{1}{n} \sum_{i=1}^{n} E_\eta (\widehat{f}_n(X_i) - E_\eta \widehat{f}_n(X_i))^2 &= \frac{1}{n} \sum_{i=1}^{n} E_\eta \left\langle e_N(X_i), G_n^{-1} \frac{1}{n} \sum_{j=1}^{n} \eta_j e^N(X_j) \right\rangle^2 \\ &= \frac{\sigma_\eta^2}{n^3} \sum_{i=1}^{n} \sum_{j=1}^{n} \langle e_N(X_i), G_n^{-1} e_N(X_j) \rangle^2 \\ &= \frac{\sigma_\eta^2}{n^2} \sum_{i=1}^{n} \langle e_N(X_i), G_n^{-1} e_N(X_i) \rangle \\ &= \frac{\sigma_\eta^2}{n} \operatorname{Tr} G_n G_n^{-1} = \sigma_\eta^2 \frac{N+1}{n},\end{aligned}$$

which implies the equality

$$R_{nN} = \frac{1}{n} E_\omega \sum_{i=1}^{n} (f(X_i) - E_\eta \widehat{f}_n(X_i))^2 + \sigma_\eta^2 \frac{N+1}{n}.$$

Now, since for fixed observation points X_i, $i = 1, \ldots, n$, we have

$$\frac{1}{n} \sum_{i=1}^{n} (f(X_i) - E_\eta \widehat{f}_n(X_i))^2 \le \frac{1}{n} \sum_{i=1}^{n} (f(X_i) - f_N(X_i))^2$$

for any linear combination $f_N = \sum_{k=0}^{N} c_k e_k$, we immediately obtain the following bound for the risk R_{nN}:

$$\begin{aligned} R_{nN} &\le \frac{1}{n} \sum_{i=1}^{n} E_\omega (f(X_i) - f_N(X_i))^2 + \sigma_\eta^2 \frac{N+1}{n} \\ &= \int_a^b (f(x) - f_N(x))^2 \, d\mu(x) + \sigma_\eta^2 \frac{N+1}{n}, \end{aligned} \tag{9}$$

where $f_N = \sum_{k=0}^{N} c_k e_k$, $c_0, c_1, \ldots, c_N \in \mathbb{R}$. As already remarked, for any absolutely continuous probability measure μ the set of all trigonometric or algebraic polynomials is dense in $L^2([0, 2\pi], \mu)$ or $L^2([-1, 1], \mu)$, respectively. Hence, in view of inequality (9) the assertion follows. ■

Note that the assertion of Theorem 3.1 can be rewritten in the form

$$\lim_{n \to \infty} E_\omega E_\eta \int_a^b (f - \widehat{f}_n)^2 \, dF_n = 0,$$

where F_n denotes the empirical distribution function of the random sample $X_1, \ldots, X_n$.

Let us now observe that the estimator $\widehat{c}^N$, which is a function of the independent random variables $\eta_1, \ldots, \eta_n$ and $X_1, \ldots, X_n$, has the following symmetry property:

$$\widehat{c}^N(\eta_1, \ldots, \eta_n, X_1, \ldots, X_n) = \widehat{c}^N(\eta_{p(1)}, \ldots, \eta_{p(n)}, X_{p(1)}, \ldots, X_{p(n)})$$

for any permutation p of $\{1, \ldots, n\}$. This implies that the random variables $f(X_i) - \widehat{f}_n(X_i) = f(X_i) - \langle \widehat{c}^N, e^N(X_i) \rangle$, $i = 1, \ldots, n$, have the same distribution and consequently

$$E(f(X_1) - \widehat{f}_n(X_1))^2 = E(f(X_2) - \widehat{f}_n(X_2))^2 = \ldots = E(f(X_n) - \widehat{f}_n(X_n))^2.$$

If the assumptions of Theorem 3.1 hold, then the above equalities imply that for a fixed index i,

$$\lim_{n \to \infty} E_\omega E_\eta (f(X_i) - \widehat{f}_n(X_i))^2 = 0$$

as $n \to \infty$.

4. Conclusions. Originally, Theorem 1.1 was proved for the more general case where the predictor variable X is multivariate [3]. In consequence, using the same technique of proof as above we can obtain a theorem analogous to Theorem 2.1 for the regression function $E(Y \mid X = x)$ using the orthonormal system of trigonometric functions in the space $L^2(Q)$, $Q = [0, 2\pi]^d \subset \mathbb{R}^d$, $d > 1$, when X takes values in the d-dimensional cube Q. The same remark also concerns Theorem 3.1. Moreover, inspection of the proof of Theorem 3.1 reveals that the theorem also holds in the case when the observation errors are zero mean independent random variables with bounded variances, i.e. when $\sup_i E\eta_i^2 \le C < \infty$.

Lugosi and Zeger [3] proved a theorem analogous to Theorem 1.1 also in the case of neural network estimators, i.e. estimators of the form

$$\widehat{r}(z) = \widehat{\xi}_0 + \sum_{i=1}^{M} \widehat{\xi}_i \psi(\langle \widehat{\gamma}_i, z\rangle + \widehat{\gamma}_{i0}), \tag{10}$$

where ψ is the activation function, $z \in \mathbb{R}^d$, and $\widehat{\xi}_0, \widehat{\xi}_j, \widehat{\gamma}_{j0} \in \mathbb{R}$, $\widehat{\gamma}_j \in \mathbb{R}^d$, $j = 1, \ldots, M$.

Our results also contribute to understanding the asymptotic properties of neural network estimators. Namely, as shown by Gallant and White [1], multivariate trigonometric series estimators can be represented as neural network estimators of type (10) with the cosine-squasher activation function and properly chosen weights $\widehat{\xi}_0, \widehat{\xi}_j, \widehat{\gamma}_{j0} \in \mathbb{R}$, $\widehat{\gamma}_j \in \mathbb{R}^d$, $j = 1, \ldots, M$. Thus, the above mentioned multivariate version of Theorem 3.1 assures existence of neural network estimators which are consistent in the sense of the mean squared prediction error for the observation model considered in Section 3.

References

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

[2] L. Györfi and H. Walk, *On the strong universal consistency of a series type regression estimate*, Math. Methods Statist. 5 (1996), 332–342.

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

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

[5] —, *Consistency of trigonometric and polynomial regression estimators*, ibid. 25 (1998), 73–83.

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

[7] V. N. Vapnik, *Estimation of Dependencies Based on Empirical Data*, Springer, New York, 1982.

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

Received on 1.10.1998;
revised version on 4.3.1999