APPLICATIONES MATHEMATICAE
22,1 (1993), pp. 91–102

W. POPIŃSKI (Warszawa)

ON LEAST SQUARES ESTIMATION OF FOURIER COEFFICIENTS AND OF THE REGRESSION FUNCTION

Abstract. The problem of nonparametric function fitting with the observation model $y_i = f(x_i) + \eta_i$, $i = 1, \ldots, n$, is considered, where η_i are independent random variables with zero mean value and finite variance, and $x_i \in [a, b] \subset \mathbb{R}^1$, $i = 1, \ldots, n$, form a random sample from a distribution with density $\varrho \in L^1[a, b]$ and are independent of the errors η_i, $i = 1, \ldots, n$. The asymptotic properties of the estimator $\widehat{f}_{N(n)}(x) = \sum_{k=1}^{N(n)} \widehat{c}_k e_k(x)$ for $f \in L^2[a, b]$ and $\widehat{c}^{N(n)} = (\widehat{c}_1, \ldots, \widehat{c}_{N(n)})^T$ obtained by the least squares method as well as the limits in probability of the estimators $\widehat{c}_k$, $k = 1, \ldots, N$, for fixed N, are studied in the case when the functions e_k, $k = 1, 2, \ldots$, forming a complete orthonormal system in $L^2[a, b]$ are analytic.

1. Introduction. Let y_i, $i = 1, \ldots, n$, be observations at points $x_i \in [a, b] \subset \mathbb{R}^1$, according to the model $y_i = f(x_i) + \eta_i$, where $f : [a, b] \to \mathbb{R}^1$ is an unknown square integrable function ($f \in L^2[a, b]$) and η_i, $i = 1, \ldots, n$, are independent identically distributed random variables with zero mean value and finite variance $\sigma_\eta^2 > 0$. Let furthermore the points x_i, $i = 1, \ldots, n$, form a random sample from a distribution with density ϱ ($\varrho \geq 0$, $\int_a^b \varrho(x)\, dx = 1$), independent of the observation errors η_i, $i = 1, \ldots, n$. If the functions e_k, $k = 1, 2, \ldots$, constitute a complete orthonormal system in $L^2[a, b]$, then f has the representation

$$f = \sum_{k=1}^{\infty} c_k e_k, \qquad \text{where } c_k = \frac{1}{b-a} \int_a^b f(x) e_k(x)\, dx,\ k = 1, 2, \ldots$$

We assume that e_k, $k = 1, 2, \ldots$, are analytic in (a, b) and continuous in $[a, b]$. Examples of orthonormal systems satisfying these requirements are [6] the

1991 *Mathematics Subject Classification*: Primary 62G07, 62F12.
Key words and phrases: Fourier series, least squares method, regression, consistent estimator.

trigonometric functions in $L^2[0,2\pi]$ and Legendre polynomials in $L^2[-1,1]$.

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

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

where $\widehat{c}^N = (\widehat{c}_1,\ldots,\widehat{c}_N)^T$, $e^N(x) = (e_1(x),\ldots,e_N(x))^T$.

To such estimators of the Fourier coefficients c_k, $k = 1,\ldots,N$, there corresponds an estimator of the regression function f of the form

$$\widehat{f}_N(x) = \sum_{k=1}^{N} \widehat{c}_k e_k(x)\,,$$

called a *projection type estimator* [4].

The vector $\widehat{c}^N$ can be obtained as a solution of the normal equations

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

where

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

The asymptotic properties of the least squares estimators of the regression function obtained in the same way as described above but for the fixed point design case were examined in [5]. The problem of choosing the regression order for least squares estimators in the case of equidistant observation points was investigated in [4].

In order to investigate the asymptotic properties of the estimators $\widehat{c}_k$, $k = 1,\ldots,N$, we introduce the probability space (Ω, F, P), where

$$\Omega = \bigtimes_{i=1}^{\infty} [a,b], \qquad F = \bigtimes_{i=1}^{\infty} F_i, \qquad P = \bigtimes_{i=1}^{\infty} P_i\,,$$

where each F_i, $i = 1,2,\ldots$, is the σ-field of Borel subsets of $[a,b]$, and P is a probability measure with the property

$$P\Big(A_1\times\ldots\times A_n \times \bigtimes_{i=n+1}^{\infty} [a,b]\Big) = (P_1\times\ldots\times P_n)(A_1\times\ldots\times A_n)$$

for $A_i \in F_i$, $i = 1,\ldots,n$, with P_i, for $i = 1,2,\ldots$, being the probability measure defined on F_i and having density ϱ with respect to the Lebesgue measure μ. The construction and properties of such a probability measure P are described in [2]. The elements of Ω are denoted by $\omega = (x_1, x_2, \ldots)$, $x_i \in [a,b]$, $i = 1,2,\ldots$

If the distribution of the observation errors η_i, $i = 1,2,\ldots$ (defined on a certain probability space (Ψ, Θ, ν)), is known, a similar probability space can be constructed, with elements of the form $\eta = (\eta_1, \eta_2, \ldots)$. From the two above described probability spaces we can of course construct in the usual way the corresponding product space with elements (ω, η) [2].

In the following section we examine the uniqueness of the estimators $\widehat{c}_k(\omega, \eta)$, $k = 1, \ldots, N$, for fixed N, and determine their limits in probability, depending on the density ϱ. In the third section we prove that the estimator $\widehat{f}_{N(n)}$ of the regression function corresponding to the Fourier coefficient estimators $\widehat{c}_k$, $k = 1, \ldots, N(n)$, is consistent in the sense of the mean square prediction error

$$D_{N(n)} = \frac{1}{n} E_\omega E_\eta \sum_{i=1}^{n} (f(x_i) - \widehat{f}_{N(n)}(x_i))^2$$

(i.e. $\lim_{n\to\infty} D_{N(n)} = 0$), on the condition that the density ϱ is bounded and the sequence $N(n)$ is properly chosen.

2. Uniqueness and consistency of Fourier coefficient estimators. First we check whether the Fourier coefficient estimators $\widehat{c}_k$, $k = 1, \ldots, N$, are uniquely determined. In order to do this we need the following two lemmas.

LEMMA 2.1. *Let $v_1, \ldots, v_n \in \mathbb{R}^n$. The matrix $G_n = \sum_{i=1}^n v_i v_i^T$ is singular* ($\det G_n = 0$) *if and only if $v_1, \ldots, v_n$ are linearly dependent.*

Proof. Suppose that G_n is singular and $v_1, \ldots, v_n$ are linearly independent. Then there exists a vector $x \neq 0$ for which $G_n x = 0$ so that

$$\sum_{i=1}^{n} v_i (v_i^T x) = \sum_{i=1}^{n} \langle v_i, x\rangle v_i = 0\,.$$

Since $v_1, \ldots, v_n$ are linearly independent, $\langle v_i, x\rangle = 0$ for $i = 1, \ldots, n$. But $\operatorname{span}\{v_1, \ldots, v_n\} = \mathbb{R}^n$ and consequently x must be zero, contrary to our assumption.

Conversely, if $v_1, \ldots, v_n$ are linearly dependent, then $\dim \operatorname{span}\{v_1, \ldots, v_n\} < n$ and we can choose $x \neq 0$ such that $\langle v_i, x\rangle = 0$ for $i = 1, \ldots, n$. Consequently, $G_n x = \sum_{i=1}^n \langle v_i, x\rangle v_i = 0$, which means that G_n is singular.

By the way, observe that a matrix of the form $G_m = \sum_{i=1}^m v_i v_i^T$, where $m < n$, is always singular since $\dim \operatorname{span}\{v_1, \ldots, v_m\} \le m$ and there exist nonzero vectors orthogonal to $\operatorname{span}\{v_1, \ldots, v_m\}$. ∎

LEMMA 2.2. *If $\varrho \in L^1[a, b]$ is a density* (*i.e.* $\varrho \ge 0$, $\int_a^b \varrho(x)\,dx = 1$), *then for $n \ge N$ the matrices*

$$G_n(\omega) = \frac{1}{n} \sum_{i=1}^{n} e^N(x_i) e^N(x_i)^T, \qquad \omega = (x_1, x_2, \ldots)\,,$$

of the normal equations (1) *are positive-definite with probability one* (*in the probability space* (Ω, F, P)).

Proof. From the definition of G_n it follows that

$$G_{n+1}(\omega) = \frac{n}{n+1} G_n(\omega) + \frac{1}{n+1} e^N(x_{n+1}) e^N(x_{n+1})^T .$$

So for $x \in \mathbb{R}^N$ we have the inequality

$$\begin{aligned}\langle G_{n+1}(\omega)x, x\rangle &= \frac{n}{n+1} \langle G_n(\omega)x, x\rangle + \frac{1}{n+1} \langle e^N(x_{n+1}) e^N(x_{n+1})^T x, x\rangle \\ &= \frac{n}{n+1} \langle G_n(\omega)x, x\rangle + \frac{1}{n+1} \langle e^N(x_{n+1}), x\rangle^2 \geq \frac{n}{n+1} \langle G_n(\omega)x, x\rangle .\end{aligned}$$

Hence $\Omega_{n+1} = \{\omega : \det G_{n+1}(\omega) = 0\} \subset \{\omega : \det G_n(\omega) = 0\} = \Omega_n$ since the matrices $G_n(\omega)$ are nonnegative-definite for $n = 1, 2, \ldots$ Thus in order to prove that $P(\Omega_n) = 0$ for $n \geq N$ it suffices to prove $P(\Omega_N) = 0$. (For $n < N$ we have $P(\Omega_n) = 1$, which is a simple consequence of our remark after the proof of Lemma 2.1.) By Lemma 2.1,

$$\det G_N(\omega) = 0 \;\Leftrightarrow\; e^N(x_1), \ldots, e^N(x_N) \text{ are linearly dependent,}$$

where $\omega = (x_1, x_2, \ldots)$, and consequently,

$$\Omega_N = \bigcup_{j=1}^{N} \{\omega : e^N(x_j) \in \operatorname{span}\{e^N(x_1), \ldots, e^N(x_{j-1}), e^N(x_{j+1}), \ldots, e^N(x_N)\}\} . \tag{2}$$

Moreover,

$$\begin{aligned}&P(\{\omega :\ e^N(x_j) \in \operatorname{span}\{e^N(x_1), \ldots, e^N(x_{j-1}), e^N(x_{j+1}), \ldots, e^N(x_N)\}\}) \\ &\quad = P(\{\omega :\ e^N(x_N) \in \operatorname{span}\{e^N(x_1), \ldots, e^N(x_{N-1})\}\}) \qquad \text{for } j = 1, \ldots, N ,\end{aligned}$$

by the properties of the product measure $P_1 \times \ldots \times P_N$. Further,

$$P(\{\omega : e^N(x_N) \in \operatorname{span}\{e^N(x_1), \ldots, e^N(x_{N-1})\}\}) = \int_a^b \ldots \int_a^b P_N(A_N)\, dP_1 \ldots dP_{N-1} ,$$

where $A_N = (e^N)^{-1}(\operatorname{span}\{e^N(x_1), \ldots, e^N(x_{N-1})\}) \subset [a, b]$, for fixed $x_1, x_2, \ldots, x_{N-1}$, is the counter-image of the closed linear subspace $\operatorname{span}\{e^N(x_1), \ldots, e^N(x_{N-1})\}$ by the continuous mapping $[a, b] \ni x_N \mapsto e^N(x_N) \in \mathbb{R}^N$ (the continuity follows from the continuity of e_k, $k = 1, 2, \ldots$). Assume now that $P_N(A_N) > 0$ for fixed $x_1, \ldots, x_{N-1}$. This means that the Lebesgue measure $\mu(A_N)$ is positive. For $x_N \in A_N$ we have

$$e^N(x_N) \in \operatorname{span}\{e^N(x_1), \ldots, e^N(x_{N-1})\} ,$$

and $\dim \operatorname{span}\{e^N(x_1), \ldots, e^N(x_{N-1})\} \le N-1$. On the other hand,

$$\operatorname{span}\{e^N(x_N) : \; x_N \in A_N\} = \mathbb{R}^N$$

since for any $v = (v_1, \ldots, v_N)^T \in \mathbb{R}^N$ orthogonal to the left-hand side

$$\langle e^N(x), v\rangle = \sum_{k=1}^{N} v_k e_k(x) = 0 \qquad \text{for } x \in A_N\,,$$

and the condition $\mu(A_N) > 0$ and the analyticity of e_k, $k = 1, 2, \ldots$, imply immediately that $v_1 = \ldots = v_N = 0$.

Thus we obtain a contradiction. Consequently, $P_N(A_N) = 0$ for all $x_1, \ldots, x_{N-1}$. This implies that

$$P(\{\omega : \; e^N(x_N) \in \operatorname{span}\{e^N(x_1), \ldots, e^N(x_{N-1})\}\}) = 0$$

and, by (2), $P(\Omega_N) = 0$. ∎

Lemma 2.2 assures that the estimators $\widehat{c}_1, \ldots, \widehat{c}_N$ obtained from the normal equations (1) are uniquely determined with probability one in the probability space (Ω, F, P), provided $n \ge N$.

Observe now that the elements of the matrix $G_n(\omega)$ in (1) have the form

$$g_{nij}(\omega) = \frac{1}{n}\sum_{k=1}^{n} e_i(x_k)e_j(x_k), \qquad \omega = (x_1, x_2, \ldots),\; i, j = 1, \ldots, N\,,$$

and we easily obtain

$$E_\omega g_{nij}(\omega) = \frac{1}{n}\sum_{k=1}^{n} E_\omega e_i(x_k)e_j(x_k) = \int_a^b e_i(x)e_j(x)\varrho(x)\,dx = g_{ij}\,. \tag{3}$$

The expected value exists because e_k, $k = 1, 2, \ldots$, are continuous in $[a, b]$. Further, since $x_1, x_2, \ldots$ are chosen independently,

$$\begin{aligned} E_\omega(g_{nij}(\omega) - g_{ij})^2 &= \frac{1}{n^2}\sum_{k=1}^{n} E_\omega(e_i(x_k)e_j(x_k) - g_{ij})^2 \\ &= \frac{1}{n}\int_a^b (e_i(x)e_j(x) - g_{ij})^2\varrho(x)\,dx \end{aligned}$$

and we see that the elements of $G_n(\omega)$ converge in L^2 to g_{ij} as $n \to \infty$.

Similarly, for the elements of the right-hand side vector of the normal equations, $g_n(\omega, \eta)$, we obtain

$$\begin{aligned} Eg_{ni}(\omega, \eta) &= \frac{1}{n}\sum_{k=1}^{n} Ey_k e_i(x_k) = \frac{1}{n}\sum_{k=1}^{n} E_\omega E_\eta(f(x_k) + \eta_k)e_i(x_k) \\ &= \frac{1}{n}\sum_{k=1}^{n} E_\omega f(x_k)e_i(x_k) = \int_a^b f(x)e_i(x)\varrho(x)\,dx = g_i \end{aligned} \tag{4}$$

for $i = 1, \ldots, N$, because the observation errors η_k, $k = 1, 2, \ldots$, have zero mean values; moreover,

$$E(g_{ni}(\omega,\eta) - g_i)^2 = \frac{1}{n^2}\sum_{k=1}^{n} E_\omega (f(x_k)e_i(x_k) - g_i)^2 + \frac{1}{n^2}\sum_{k=1}^{n} E_\omega E_\eta \eta_k^2 e_i^2(x_k)$$

$$= \frac{1}{n}\int_a^b (f(x)e_i(x) - g_i)^2 \varrho(x)\,dx + \frac{1}{n}\sigma_\eta^2 \int_a^b e_i^2(x)\varrho(x)\,dx.$$

This implies that the elements of $g_n(\omega,\eta)$ converge in L^2 to g_i as $n \to \infty$, provided

$$\int_a^b f^2(x)\varrho(x)\,dx < \infty\,.$$

In that case we can determine the limits in probability of the estimators $\widehat{c}_1, \ldots, \widehat{c}_N$ by applying the following lemma.

Lemma 2.3. *Let (Ω, F, P) be a probability space. Let $A_n(\omega)$, $n = 1, 2, \ldots,$ be a sequence of random matrices of fixed dimension k, nonsingular with probability one, and let $y_n(\omega)$ be a sequence of random vectors of dimension k. If*

1) $\lim_{n\to\infty} A_n(\omega) \overset{\mathrm{P}}{=} A$ *(in probability), where A is a nonsingular matrix,*
2) $\lim_{n\to\infty} y_n(\omega) \overset{\mathrm{P}}{=} y$,

then the sequence of random vectors $x_n(\omega)$ defined with probability one by the equations

$$A_n(\omega)x_n(\omega) = y_n(\omega), \qquad n = 1, 2, \ldots,$$

converges in probability to the vector x which is the unique solution of the equation $Ax = y$.

Proof. Apply the fact that the elements of the inverse matrix A^{-1} are continuous functions of the elements of the matrix A. ■

In order to use Lemma 2.3 in the case of the normal equations (1) it is enough to show that the matrix G with elements g_{ij} defined in (3) is positive-definite. Clearly, for any $v = (v_1, \ldots, v_N)^T \in \mathbb{R}^N$,

$$\langle Gv, v\rangle = \sum_{i=1}^{N}\sum_{j=1}^{N} g_{ij}v_iv_j = \sum_{i=1}^{N}\sum_{j=1}^{N} v_iv_j \int_a^b e_i(x)e_j(x)\varrho(x)\,dx$$

$$= \int_a^b \Big(\sum_{i=1}^{N} v_ie_i(x)\Big)^2 \varrho(x)\,dx \ge 0\,.$$

Suppose that $\langle Gv, v\rangle = 0$. Since ϱ is positive on some set with positive Lebesgue measure, $\sum_{i=1}^N v_ie_i(x) = 0$ for $x \in \Delta$, $\mu(\Delta) > 0$, and then $v_1 = \ldots = v_N = 0$ as already remarked in the proof of Lemma 2.2.

We can now formulate the result concerning the convergence in probability of the estimators $\widehat{c}_1, \ldots, \widehat{c}_N$ for fixed N.

THEOREM 2.1. *If the density* $\varrho \in L^1[a,b]$ *satisfies* $\int_a^b f^2(x)\varrho(x)\,dx < \infty$, *then the estimators* $\widehat{c}_1, \ldots, \widehat{c}_N$, N *being fixed*, *are for* $n \geq N$ *uniquely determined with probability one and*

$$\lim_{n\to\infty} \widehat{c}^N \overset{\mathrm{p}}{=} G^{-1}g\,, \tag{5}$$

where $\widehat{c}^N = (\widehat{c}_1, \ldots, \widehat{c}_N)^T$, G *is the matrix with elements*

$$g_{ij} = \int_a^b e_i(x)e_j(x)\varrho(x)\,dx$$

and $g \in \mathbb{R}^N$ *is the vector with components*

$$g_i = \int_a^b f(x)e_i(x)\varrho(x)\,dx\,,$$

$i, j = 1, \ldots, N$.

P r o o f. The assertion follows from earlier considerations and from Lemmas 2.2 and 2.3. ∎

The vector $G^{-1}g$ can be characterized more precisely. Namely, consider the functional defined for $z \in \mathbb{R}^N$ by the formula

$$J(z) = \int_a^b \Big(f(x) - \sum_{i=1}^N z_i e_i(x)\Big)^2 \varrho(x)\,dx, \qquad z = (z_1, \ldots, z_N)^T\,.$$

In order to find the points of extrema of $J(z)$ we set its partial derivatives with respect to z_i, $i = 1, \ldots, N$, to be zero and we obtain the system of linear equations $Gz = g$, with G positive-definite. So the components of $\widehat{c}^N$ converge in probability to the components of the vector $G^{-1}g$ which minimizes the value of $J(z)$.

In the case of constant density ($\varrho = 1/(b-a)$) we obtain, by (5),

$$\lim_{n\to\infty} \widehat{c}^N \overset{\mathrm{p}}{=} c^N, \qquad c^N = (c_1, \ldots, c_N)^T\,,$$

and so $\widehat{c}_1, \ldots, \widehat{c}_N$ are then consistent estimators of the Fourier coefficients of $f \in L^2[a,b]$.

3. Mean square prediction error and choice of the order of regression. Now we deal with the asymptotic properties of the projection type estimator of the regression function f:

$$\widehat{f}_N(x) = \sum_{k=1}^N \widehat{c}_k e_k(x)\,,$$

where the vector of Fourier coefficient estimators $\widehat{c}^N = (\widehat{c}_1, \ldots, \widehat{c}_N)^T$ is obtained from the normal equations (1),

$$\widehat{c}^N(\omega,\eta) = G_n^{-1}(\omega) g_n(\omega,\eta) = G_n^{-1}(\omega)\left(\frac{1}{n}\sum_{i=1}^n (f(x_i)+\eta_i)e^N(x_i)\right).$$

From the above equality and the decomposition

$$f(x) = \sum_{k=1}^N c_k e_k(x) + r_N(x) = \langle e^N(x), c^N\rangle + r_N(x),$$
$$\text{where} \quad r_N = \sum_{k=N+1}^{\infty} c_k e_k\,,$$

we obtain

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

Set $a^N = (1/n)\sum_{i=1}^n r_N(x_i)e^N(x_i)$. In view of the equalities

$$G_n = \frac{1}{n}\sum_{i=1}^n e^N(x_i)e^N(x_i)^T, \quad E_\eta(\eta_i\eta_j) = \sigma_\eta^2\delta_{ij}, \quad i,j = 1,\ldots,n\,,$$
$$f(x) - \widehat{f}_N(x) = \langle c^N - \widehat{c}^N, e^N(x)\rangle + r_N(x)$$

it is easy to show that

$$\begin{aligned}
E_\eta(f(x) - \widehat{f}_N(x))^2 & \\
&= E_\eta r_N^2(x) + 2r_N(x)E_\eta\langle c^N - \widehat{c}^N, e^N(x)\rangle + E_\eta\langle c^N - \widehat{c}^N, e^N(x)\rangle^2 \\
&= r_N^2(x) - 2r_N(x)\langle G_n^{-1}a^N, e^N(x)\rangle \\
&\quad + \langle G_n^{-1}a^N, e^N(x)\rangle^2 + \frac{1}{n}\sigma_\eta^2\langle e^N(x), G_n^{-1}e^N(x)\rangle\,,
\end{aligned}$$

and further,

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^n E_\eta(f(x_i) - \widehat{f}_N(x_i))^2 & \\
&= \frac{1}{n}\sum_{i=1}^n r_N^2(x_i) - 2\langle G_n^{-1}a^N, a^N\rangle + \langle G_n^{-1}a^N, a^N\rangle + \sigma_\eta^2\frac{N}{n}\,.
\end{aligned}$$

Finally, we obtain the formula

$$\frac{1}{n}\sum_{i=1}^n E_\eta(f(x_i) - \widehat{f}_N(x_i))^2 = \frac{1}{n}\sum_{i=1}^n r_N^2(x_i) - \langle G_n^{-1}a^N, a^N\rangle + \sigma_\eta^2\frac{N}{n}\,. \tag{6}$$

Since G_n is a.s. positive-definite for $n \ge N$,

$$0 \le \frac{1}{n}\sum_{i=1}^{n} E_\eta(f(x_i) - \widehat{f}_N(x_i))^2 \le \frac{1}{n}\sum_{i=1}^{n} r_N^2(x_i) + \sigma_\eta^2 \frac{N}{n}\,. \tag{7}$$

In the case of constant density $\varrho = 1/(b-a)$, this inequality yields

$$E\frac{1}{n}\sum_{i=1}^{n}(f(x_i) - \widehat{f}_N(x_i))^2 \le \frac{1}{n}\sum_{i=1}^{n} E_\omega r_N^2(x_i) + \sigma_\eta^2\frac{N}{n}$$
$$= \frac{1}{b-a}\int_a^b r_N^2(x)\,dx + \sigma_\eta^2\frac{N}{n}\,,$$

and since

$$\frac{1}{b-a}\int_a^b r_N^2(x)\,dx = \frac{1}{b-a}\sum_{k=N+1}^{\infty} c_k^2$$

we can rewrite the last inequality in the form

$$D_N = E\frac{1}{n}\sum_{i=1}^{n}(f(x_i) - \widehat{f}_N(x_i))^2 \le \frac{p_N}{b-a} + \sigma_\eta^2\frac{N}{n},$$

$$\text{where } p_N = \sum_{k=N+1}^{\infty} c_k^2\,.$$

Since the series $\sum_{k=1}^{\infty} c_k^2$ is convergent ($f \in L^2[a,b]$) we conclude from the above inequality that in the case $\varrho = 1/(b-a)$ we have $\lim_{n\to\infty} D_{N(n)} = 0$ provided $\lim_{n\to\infty} N(n) = \infty$ and $\lim_{n\to\infty} N(n)/n = 0$. The estimator $\widehat{f}_{N(n)}$ is then consistent in the sense of the mean square prediction error $D_{N(n)}$. A similar result holds for the case of bounded density ϱ as one can see from inequality (7).

If we define the prediction error by

$$d_{N(n)} = \frac{1}{n}\sum_{i=1}^{n}(f(x_i) - \widehat{f}_{N(n)}(x_i))^2\,,$$

then the condition $\lim_{n\to\infty} D_{N(n)} = \lim_{n\to\infty} E d_{N(n)} = 0$ implies of course $\lim_{n\to\infty} d_{N(n)} \overset{\mathrm{p}}{=} 0$. Consequently, the previously proved facts concerning the convergence of the mean square prediction error $D_{N(n)}$ allow us to formulate the following theorem.

THEOREM 3.1. *If the density $\varrho \in L^1[a,b]$ is bounded and the sequence of natural numbers $N(n)$, $n = 1, 2, \ldots$, satisfies*

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

then the estimator of the regression function

$$\widehat{f}_{N(n)} = \sum_{k=1}^{N(n)} \widehat{c}_k e_k$$

is consistent in the sense of the prediction error $d_{N(n)}$ (*i.e.* $\lim_{n\to\infty} d_{N(n)} \overset{\mathrm{p}}{=} 0$ *in* (Ω, F, P)).

Proof. The assertion follows from Lemma 2.2 and from earlier considerations of Section 3. ∎

Now we consider the problem of choosing the regression order N. If we know the values of p_N, $N = 1, 2, \ldots$, and of σ_η^2, we can choose N according to the criterion

$$N^* = \arg\min_{1\le N\le n} \left(\frac{p_N}{b-a} + \sigma_\eta^2 \frac{N}{n}\right). \tag{8}$$

Then

$$D_{N^*} \le \frac{p_{N^*}}{b-a} + \sigma_\eta^2 \frac{N^*}{n} = \min_{1\le N\le n} \left(\frac{p_N}{b-a} + \sigma_\eta^2 \frac{N}{n}\right).$$

If we only know some estimates $p'_N \ge p_N$ we can replace p_N by p'_N in (8). If the sequence $|c_k|$, $k = 1, 2, \ldots$, is decreasing, then p_N is a convex function (of N) and so is $A_N = p_N/(b-a) + \sigma_\eta^2 N/n$, which cannot then have local minima; we thus have $N^* = \max\{N : c_N^2 \ge (b-a)\sigma_\eta^2/n\}$ [4].

The values of p_N, $N = 1, 2, \ldots$, can of course be unknown, but we can define the statistic

$$s_N = \frac{1}{n}\sum_{i=1}^n (y_i - \widehat{f}_N(x_i))^2$$

for which

$$\begin{aligned}
E_\eta s_N &= \frac{1}{n}\sum_{i=1}^n E_\eta(f(x_i) - \widehat{f}_N(x_i) + \eta_i)^2 \\
&= \frac{1}{n}\sum_{i=1}^n E_\eta(f(x_i) - \widehat{f}_N(x_i))^2 - \frac{2}{n}\sum_{i=1}^n E_\eta \widehat{f}_N(x_i)\eta_i + \sigma_\eta^2 \\
&= \frac{1}{n}\sum_{i=1}^n E_\eta(f(x_i) - \widehat{f}_N(x_i))^2 - \frac{2}{n}\sum_{i=1}^n E_\eta \langle \widehat{c}^N, e^N(x_i)\rangle \eta_i + \sigma_\eta^2 \\
&= \frac{1}{n}\sum_{i=1}^n E_\eta(f(x_i) - \widehat{f}_N(x_i))^2 \\
&\quad - \frac{2}{n}\sum_{i=1}^n E_\eta \left\langle G_n^{-1}\left(\frac{1}{n}\sum_{j=1}^n y_j e^N(x_j)\right), e^N(x_i)\right\rangle \eta_i + \sigma_\eta^2
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{n}\sum_{i=1}^{n} E_\eta(f(x_i) - \widehat{f}_N(x_i))^2 \\
&\quad - \frac{2}{n}\sum_{i=1}^{n} E_\eta\Big\langle G_n^{-1}\Big(\frac{1}{n}\sum_{j=1}^{n}\eta_j e^N(x_j)\Big), e^N(x_i)\Big\rangle\eta_i + \sigma_\eta^2 \\
&= \frac{1}{n}\sum_{i=1}^{n} E_\eta(f(x_i) - \widehat{f}_N(x_i))^2 - \frac{2}{n^2}\sigma_\eta^2\sum_{i=1}^{n}\langle G_n^{-1}e^N(x_i), e^N(x_i)\rangle + \sigma_\eta^2 \\
&= \frac{1}{n}\sum_{i=1}^{n} E_\eta(f(x_i) - \widehat{f}_N(x_i))^2 - 2\sigma_\eta^2\frac{N}{n} + \sigma_\eta^2\,.
\end{aligned}
$$

Hence, remembering the definition of D_N, we obtain

$$
Es_N = E_\omega E_\eta s_N = D_N - 2\sigma_\eta^2\frac{N}{n} + \sigma_\eta^2\,, \tag{9}
$$

which can be rewritten in the form

$$
E\Big(s_N + 2\sigma_\eta^2\frac{N}{n}\Big) = D_N + \sigma_\eta^2\,.
$$

So if we choose N (the order of regression) according to the criterion

$$
N^* = \arg\min_{1\le N\le n}\Big(s_N + 2\sigma_\eta^2\frac{N}{n}\Big)
$$

we can assert that in the mean we obtain those values of N which minimize D_N [4]. This kind of criterion for the choice of N is known in the literature as the Mallows–Akaike criterion [1], [3].

4. Conclusions. It is worth remarking that we can obtain a better lower bound for the mean square prediction error than the obvious one $D_N \ge 0$. We apply the following lemma proved in [5].

LEMMA 4.1. *Let $h = (h_1, \ldots, h_n)^T \in \mathbb{R}^n$. Then*

$$
\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} h_i h_j e^N(x_i)^T G_n^{-1} e^N(x_j) \le \frac{1}{n}\sum_{i=1}^{n} h_i^2\,.
$$

Since $a^N = (1/n)\sum_{i=1}^n r_N(x_i)e^N(x_i)$ and $G_n > 0$ a.s. for $n \ge N$, putting $h_i = r_N(x_i)$, $i = 1, \ldots, n$, by Lemma 4.1 we obtain

$$
0 \le \langle G_n^{-1}a^N, a^N\rangle \le \frac{1}{n}\sum_{i=1}^{n} r_N(x_i)^2
$$

almost surely for $n \ge N$. Now, taking into account (6) we easily obtain the lower and upper bounds for D_N, valid for $n \geq N$:

$$\sigma_\eta^2 \frac{N}{n} \leq D_N \leq M_\varrho p_N + \sigma_\eta^2 \frac{N}{n}, \qquad \text{where } M_\varrho = \sup_{a \leq x \leq b} \varrho(x)\,. \tag{10}$$

From (9) and (10) it follows immediately that in the case when ϱ is bounded and the conditions $\lim_{n\to\infty} N(n) = \infty$ and $\lim_{n\to\infty} N(n)/n = 0$ are satisfied, $s_{N(n)}$ is an asymptotically unbiased estimator of σ_η^2.

The lower and upper bounds for $D_{N(n)}$ also allow us to estimate the bias of $s_{N(n)}$ for $n \geq N(n)$, namely

$$-\sigma_\eta^2 \frac{N(n)}{n} \leq Es_{N(n)} - \sigma_\eta^2 \leq M_\varrho p_{N(n)} - \sigma_\eta^2 \frac{N(n)}{n}\,.$$

The results presented in the two preceding sections can be easily proved in the case of regression functions $f \in L^2(A)$, $A \subset \mathbb{R}^m$, $m > 1$, $\mu(A) < \infty$, and certain complete orthonormal systems of functions (like the functions

$$\exp(ikx + ily)/2\pi, \qquad 0 \leq x, y \leq 2\pi,\ k, l = 0, \pm 1, \pm 2, \ldots,$$

forming a complete orthonormal system in $L^2([0, 2\pi] \times [0, 2\pi])$).

References

[1] H. Akaike, *A new look at the statistical model identification*, IEEE Trans. Automat. Control AC-19 (1974), 716–723.

[2] Y. S. Chow and H. Teicher, *Probability Theory, Independence, Interchangeability, Martingales*, Springer, Heidelberg, 1978.

[3] C. L. Mallows, *Some comments on C_p*, Technometrics 15 (1973), 661–675.

[4] B. T. Polyak and A. B. Tsybakov, *Asymptotic optimality of the C_p criterion in projection type estimation of a regression function*, Teor. Veroyatnost. i Primenen. 35 (1990), 305–317 (in Russian).

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

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

WALDEMAR POPIŃSKI
RESEARCH AND DEVELOPMENT CENTER OF STATISTICS
AL. NIEPODLEGŁOŚCI 208
00-925 WARSZAWA, POLAND

Received on 1.12.1992