A NONPARAMETRIC ESTIMATOR OF A NONLINEAR FUNCTIONAL OF REGRESSION UNDER A GIVEN PLAN

Yu. I. Pastukhova

UDC 519.2

A lower bound for the estimation quality is found and on a set of sufficiently smooth regression functions one constructs an estimator that attains asymptotically this bound.

1. Formulation of the Problem

Let \( t_i, t_2, \ldots, t_n \) be independent random variables with values in \([0, 1]\) (the plan of the observations), having distribution density \( p(t) \) with respect to the Lebesgue measure on \([0, 1]\), satisfying the inequalities

\[
0 < c < p(t) < C < \infty.
\]  

We denote by \( R(t) = E \{ X_i(t) | t_i = t \} \) the regression function. Then the result of the observations \( X_i = X_i(t_i) \) at the point \( t_i \) can be written in the form

\[
X_i = R(t_i) + \xi_i(t_i).
\]

We shall assume that the "noises" \( \xi_i(t_i) \) are conditionally independent under the given plan \( t^{(n)} \), \( E \{ \xi_i(t_i) | t_i = t \} = 0 \), and the conditional variance

\[
\sigma^2(t) = E \{ \xi_i^2(t_i) | t_i = t \},
\]

is known; moreover \( 0 < \sigma^2(t) < \infty \).

Let \( R(t), \sigma^2(t) \in L^2[0,1] \). We denote by \( \| \cdot \| \) the norm in the space \( L^2 \).

We consider the problem of the nonparametric estimation of a smooth functional \( F(R) : L^2 \rightarrow \mathbb{R}^i \) of regression, under a given plan \( t^{(n)} \), when it is known only that \( R \in K \), \( K \) being a subset of the set of sufficiently smooth functions in \( L^2 \).

Problems of nonparametric estimation of nonlinear functionals with respect to independent, identically distributed observations, have been considered repeatedly in the literature; see, for example, [1]-[3]. In [1] one has described \( \sqrt{n} \)-consistent asymptotically efficient estimators of functionals of the form

\[
\Phi(F) = \int f(x, F(x), F'(x), \ldots, F^{(m)}(x)) dF(x),
\]

based on independent observations \( X_1, \ldots, X_n \) of the random variable \( X \), distributed according to an unknown distribution function \( F \). In [2] one has suggested a general method for the construction of asymptotically efficient nonparametric estimators for a wide class of sufficiently smooth functionals of the distribution density. This method has been used in [3] for the estimation of a smooth functional of the spectral density of a stationary process if a sample \( X_1, \ldots, X_n \) from this process is known.


0090-4104/90/5202-2965$12.50 © 1990 Plenum Publishing Corporation 2965
The purpose of this paper is to apply the concept of asymptotically efficient nonparametric estimation, suggested in [2], to the estimation of smooth functionals of an unknown regression function, observed in an additive noise with a known conditional variance.

In Sec. 2 we obtain lower bounds for a wide class of estimators \( \{\hat{F}_n\} \) of the functional \( F(R) \) under various assumptions regarding the distribution of the "noises" of the observations. In Sec. 3 we solve the auxiliary problem of the order-sharp estimation of a regression function \( R(t) \) from the first \( n_0 < n \) observations. In Sec. 4 we construct an estimator \( \hat{F}_n \) of the functional \( F \), based on the observations \( X_1(t_1), \ldots, X_n(t_n) \), such that the quantity \( \Delta_n = E\{\hat{F}_n - F\}^2 \) turns out to be minimal in some asymptotic sense, under the assumption that the conditional variance of the "noises" of the observations is known.

2. Lower Bounds

We obtain lower bounds for the quality of an arbitrary estimator \( \hat{F}_n \) of the functional \( F(R) \) for a given plan \( t^{(m)} \). Assume that we know the conditional density \( q(x|t) \) of the distribution of the "noises" \( \xi_i(t_i) \) under the condition \( t_i = t \), and, moreover, the function \( q(x|t) \) is absolutely continuous with respect to \( x \) and has a finite Fisher amount of information

\[
I(t) = \int_{\mathbb{R}^d} \frac{(q(x|t))^2}{q(x|t)} \, dx,
\]

satisfying the following regularity conditions

\[
0 < \inf_{[0,1]} I(t) \leq \sup_{[0,1]} I(t) < \infty, \quad \sup_{[0,1]} \int |x| q(x|t) \, dx < \infty.
\] (2.1)

**Theorem 2.1.** Assume that the conditional distribution of \( \xi_i(t_i) \), for a given plan \( t^{(m)} \), satisfies condition (2.1) and also the condition (1.1) is satisfied. Assume that the functional \( F(R) \) is Frechet differentiable on some compactum \( \mathcal{P} \subseteq \mathbb{R}_+ \), and, moreover, \( \sup_{R \in \mathcal{P}} \|F(R,\cdot)\| < \infty \). We assume that the compactum \( \mathcal{P} \) contains some function \( R_0(t) \) and for some \( \varepsilon > 0 \) for \( |s| < \varepsilon \) the parametric family \( R_0(t) + sg(t) \in \mathcal{P} \). Here

\[
q_s(t) = \left( \int_0^1 \frac{(F'(R_0,t))^2}{p(t)I(t)} \, dt \right)^{-1} \frac{F'(R_0,t)}{p(t)I(t)}.
\]

Then for any estimator \( \hat{F}_n \) of the functional \( F(R) \) we have the inequality

\[
\lim_{n \to \infty} \sup_{R \in \mathcal{P}} \left[ n E\{\hat{F}_n - F(R)^2\} \right] \geq \int_0^1 \frac{(F'(R_0,t))^2}{I(t)p(t)} \, dt.
\] (2.2)

**Proof.** Following the general method for the derivation of lower bounds for nonparametric estimates (see, for example, [4]), we consider the parametric family

\[
R_h(t) = R_0(t) + (h - \theta)q_\theta(t), \quad \theta = F(R_0).
\]

Making use of the differentiability of the functional \( F(R) \), we have

\[
F(R_h) = \theta + (h - \theta) \left( \int_0^1 \frac{(F'(R_0,t))^2}{p(t)I(t)} \, dt \right)^{\frac{1}{2}} \left( \int_0^1 \frac{(F'(R_0,t))^2}{p(t)I(t)} \, dt \right)^{\frac{1}{2}} + o(1h - \theta) = h + o(1h - \theta).
\]

Thus, the problem has been reduced to the estimation of the parameter \( h \) from the observations \( t^{(m)} \), \( X^{(m)} \), having a distribution with density