Ann. Inst. Statist. Math.
Vol. 42, No. 4, 699–708 (1990)

NONPARAMETRIC ESTIMATION OF A REGRESSION FUNCTION BY DELTA SEQUENCES

EIICHI ISOGAI

Department of Mathematics, Niigata University, 8050 Ikarashi 2-cho, Niigata 950-21, Japan

(Received August 18, 1989; revised January 23, 1990)

Abstract. A somewhat more general class of nonparametric estimators for estimating an unknown regression function g from noisy data is proposed. The regressor is assumed to be defined on the closed interval [0, 1]. This class of estimators is shown to be pointwisely consistent in the mean square sense and with probability one. Further, it turns out that these estimators can be applied to a wide class of noises.

Key words and phrases: Nonparametric regression, mean square convergence, strong consistency, delta sequences.

1. Introduction

Recently, there have been many papers on the problem of nonparametric estimation of a regression function g under a situation where a random variable Y is recorded which depends on a design parameter $x \in R^p$ ("fixed design regression"). In this paper, for $p = 1$ we consider the problem of estimating the following nonparametric regression function g which is a completely unknown bounded real-valued function on the closed interval [0, 1]. Let $(x_1, Y_1), (x_2, Y_2), \dots, (x_n, Y_n)$ be observations according to the nonparametric regression model

$$Y_i = g(x_i) + Z_i \qquad i = 1, \dots, n ,$$

where the errors Z_i are independently and identically distributed (i.i.d.) random variables such that

$$(1.1) \qquad EZ_i = 0 \quad \text{and} \quad E|Z_i|^{\alpha} < \infty \qquad \text{for some } \alpha > 1 .$$

Without loss of generality, we assume that the design points x_i satisfy $0 = x_0 \leq x_1 < x_2 < \cdots < x_n \leq x_{n+1} = 1$, where all x_i are known without error. For these measurements a somewhat more general class of nonparametric estimators by delta sequences $\{g_{nm}(x)\}$ is proposed. We show that these $g_{nm}(x)$ are pointwisely consistent in the mean square sense and with probability one.

For fixed design regression with $p = 1$, many classes of estimators have been proposed, including the kernel method (Priestley and Chao (1972), Chen and Lin (1981), Gasser and Müller (1984), Georgiev (1984*a*, 1985), etc.), the nearest neighbor method (Georgiev (1984*b*), etc.), the orthogonal series method (Rutkowski and Rafajlowicz (1989), etc.) and the spline method (Rice and Rosenblatt (1983), Eubank (1988), etc.). Georgiev and Greblicki (1986) proposed the general family of estimators including the estimators given by Priestley and Chao (1972), Gasser and Müller (1979), Chen and Lin (1981) and Georgiev (1984*b*). Ahmad and Lin (1984), Galkowski and Rutkowski (1986) and Georgiev (1988) discussed the fixed design regression with $p > 1$. Our method by delta sequences is a somewhat more general method which includes the kernel method, the polynomial approximation method and the characteristic function approach.

Stone (1977) has discussed the same class of nonparametric regression estimators as that of Georgiev (1988) for the stochastic design model with $p > 1$ in which it is assumed that the $X_1,\ldots, X_n$ are random variables.

The paper consists of four sections. In Section 2 we shall define a class of estimators by delta sequences. Section 3 contains the results about weak and strong pointwise consistency. In Section 4 examples of delta sequences are given.

2. Estimators by delta sequences

In this section we propose a general class of estimators by delta sequences.

DEFINITION 2.1. Let J be an interval of the real line R. A sequence of bounded measurable functions $\{\delta_m, m = 1,2,\ldots\}$ on $J \times J$ is said to be a delta sequence on J if each $\phi \in C_0^\infty(J)$ and $x \in J$

$$\int_J \delta_m(x,y)\phi(y)dy \to \phi(x) \qquad \text{as} \qquad m \to \infty ,$$

where J° is the interior of J, $C_0^\infty(J)$ denotes the space of continuous functions on J which are infinitely differentiable on J° and have compact support in J, and all integrals are taken with respect to Lebesgue measure throughout the paper.

We consider the closed interval $J = [0, 1]$ throughout the paper. Set $A_i = (x_{i-1}, x_i]$ for $i = 1,\ldots, n + 1$. Our estimators of g are proposed by

$$g_{nm}(x) = \sum_{i=1}^{n} Y_i \int_{A_i} \delta_m(x, u)\,du \qquad n \text{ and } m \geq 1$$

for each $x \in J$, where $\{\delta_m, m \geq 1\}$ is a delta sequence on J satisfying the following Condition A.

CONDITION A. For each $x \in J$

(i) $\sup_{m \geq 1} \int_J |\delta_m(x, y)|\,dy < \infty$,

(ii) $\int_J \delta_m(x, y)\,dy \to 1$ as $m \to \infty$,

(iii) $\int_J |\delta_m(x, y)| I(|x - y| > \eta)\,dy \to 0$ as $m \to \infty$ for each $\eta > 0$ and

(iv) $\sup_{y \in J} |\delta_m(x, y)| = O(m)$ as $m \to \infty$,

where $I(B)$ denotes the indicator function of B.

3. Consistency

In this section we shall show that the estimator $g_{nm}(x)$ is pointwisely consistent in the mean square sense and with probability one (w.p.1). Let

$$\gamma(n) = \max\{x_i - x_{i-1} \mid 1 \leq i \leq n + 1\} \tag{3.1}$$

and $m(n)$ be a positive integer depending on n. Throughout the remainder of this paper $C_1, C_2, \ldots$ denote appropriate positive constants.

LEMMA 3.1. (Asymptotic unbiasedness) *Suppose $g(x)$ is bounded on $[0, 1]$. If*

$$\gamma(n) \to 0, \quad m(n) \to \infty \quad \textit{and} \quad m(n)\gamma(n) \to 0 \quad \textit{as} \quad n \to \infty, \tag{3.2}$$

then

$$Eg_{nm(n)}(x) \to g(x) \qquad \textit{as} \qquad n \to \infty$$

at every continuity point $x \in [0, 1]$ of the function g.

PROOF. Fix any continuity point $x \in [0, 1]$ of g. For notational simplicity let $m = m(n)$. It is easy to see that

$$(3.3)\qquad \begin{aligned} &Eg_{nm}(x) - g(x) \\ &= \sum_{i=1}^{n} (g(x_i) - g(x)) \int_{A_i} \delta_m(x,u)\,du \\ &\quad + \left(\sum_{i=1}^{n} \int_{A_i} \delta_m(x,u)\,du - 1 \right) g(x) \\ &= I_n^1 + I_n^2, \qquad \text{say}\,. \end{aligned}$$

By (3.1), (3.2) and Condition A(iv) we get

$$\left| \int_{A_{n+1}} \delta_m(x,u)\,du \right| \le C_1 m\gamma(n) \to 0 \qquad \text{as} \qquad n \to \infty\,,$$

which, together with Condition A(ii), yields that

$$(3.4)\qquad I_n^2 = \left[\left(\int_J \delta_m(x,u)\,du - 1 \right) - \int_{A_{n+1}} \delta_m(x,u)\,du \right] g(x) \to 0 \qquad \text{as} \qquad n \to \infty\,.$$

Fix any $\varepsilon > 0$. Since g is continuous at x there exists a positive constant $\xi = \xi(\varepsilon, x)$ such that

$$|x-u| < \xi \quad \text{and} \quad u \in J \quad \text{imply} \quad |g(x) - g(u)| < \varepsilon\,.$$

Hence

$$(3.5)\qquad \begin{aligned} |I_n^1| &\le \sum_{i=1}^{n} |g(x_i) - g(x)|\, I(|x_i - x| < \xi) \int_{A_i} |\delta_m(x,u)|\,du \\ &\quad + \sum_{i=1}^{n} |g(x_i) - g(x)|\, I(|x_i - x| \ge \xi) \int_{A_i} |\delta_m(x,u)|\,du \\ &\le \varepsilon \int_J |\delta_m(x,u)|\,du + C_2 \sum_{i=1}^{n} I(|x_i - x| \ge \xi) \int_{A_i} |\delta_m(x,u)|\,du\,, \end{aligned}$$

where $C_2 = 2 \sup_{y \in J} |g(y)|$. By (3.2) there exists a positive integer $N = N(\xi)$ such that $\gamma(n) < \xi/2$ for all $n \ge N$. Fix any $n \ge N$. Assume that $|x_i - x| \ge \xi$ and $u \in A_i$ for each $i = 1,\dots,n$. Then we get

$$|x-u| \ge |x - x_i| - |u - x_i| \ge \xi - \gamma(n) > \frac{\xi}{2}\,,$$

which yields that

$$I(|x_i - x| \geq \xi)\int_{A_i} |\delta_m(x,u)|\,du \leq \int_{A_i} |\delta_m(x,u)|\, I\left(|x-u| > \frac{\xi}{2} \right) du\,.$$

Thus

$$\sum_{i=1}^{n} I(|x_i - x| \geq \xi)\int_{A_i} |\delta_m(x,u)|\,du \leq \int_{J} |\delta_m(x,u)|\, I\left(|x-u| > \frac{\xi}{2} \right) du \tag{3.6}$$

for all $n \geq N$. Hence it follows from (3.5), (3.6), Conditions A(i) and A(iii) that

$$\limsup_{n\to\infty} |I_n^1| \leq C_3\varepsilon\,,$$

which yields that

$$I_n^1 \to 0 \quad \text{as} \quad n \to \infty\,. \tag{3.7}$$

Therefore, the relations (3.3), (3.4) and (3.7) conclude Lemma 3.1. This completes the proof.

THEOREM 3.1. (Mean square convergence) *Suppose $g(x)$ is bounded on* $[0,1]$. *Assume the condition* (1.1) *holds for* $\alpha = 2$. *Then, under the conditions of Lemma* 3.1

$$E(g_{nm(n)}(x) - g(x))^2 \to 0 \quad \textit{as} \quad n \to \infty$$

at every continuity point $x \in [0,1]$ of the function g.

PROOF. Fix any continuity point $x \in [0,1]$ of g. Since

$$E(g_{nm(n)}(x) - g(x))^2 = \mathrm{Var}\,(g_{nm(n)}(x)) + (Eg_{nm(n)}(x) - g(x))^2\,,$$

by Lemma 3.1 it is sufficient to show that

$$\mathrm{Var}\,(g_{nm(n)}(x)) \to 0 \quad \text{as} \quad n \to \infty\,. \tag{3.8}$$

It is easy to show that

$$\mathrm{Var}\,(g_{nm(n)}(x)) = \mathrm{Var}\,(Z_1) \sum_{i=1}^{n} a_{m(n)i}^2\,, \tag{3.9}$$

where

$$a_{m(n)i} = \int_{A_i} \delta_{m(n)}(x, u)\, du \qquad \text{for} \quad i = 1, \ldots, n\ .$$

By Conditions A(i) and A(iv) we get that

$$|a_{m(n)i}| \le C_1 m(n)\gamma(n) \qquad \text{for} \quad i = 1, \ldots, n$$

and

$$\sum_{i=1}^{n} |a_{m(n)i}| \le C_2 \qquad \text{for} \quad n \ge 1\ ,$$

which implies that

$$(3.10) \qquad \sum_{i=1}^{n} a_{m(n)i}^2 \le C_3 m(n)\gamma(n)\ .$$

Thus, from (3.2), (3.9) and (3.10) we obtain (3.8). This completes the proof.

To obtain the result about strong pointwise consistency we need the following lemma due to Teicher (1985).

LEMMA 3.2. *Let* $\{X, X_n, n \ge 1]$ *be i.i.d. random variables and* $\{a_{ni}, 1 \le i \le k_n < \infty, k_n \uparrow \infty, n \ge 1\}$ *an array of constants satisfying*

$$\max_{1 \le i \le k_n} |a_{ni}| d_i = O(1/\log n)\ ,$$

where $0 < d_n \uparrow$, $d_n = O(n^{1/\alpha})$ *for some* α *in* $(0, 2]$ *and*

$$\sum_{n=1}^{\infty} P\{|X| > d_n\} < \infty\ .$$

If $1 \le \alpha \le 2$, $d_n/n \downarrow$, $EX = 0$ *and*

$$\sum_{i=1}^{k_n} a_{ni}^2 d_i^{2-\alpha} = o(1/\log n), \qquad \sum_{i=1}^{k_n} a_{ni}^2 d_i^{2-\alpha} = O(1/\log k_n)\ ,$$

then

$$\sum_{i=1}^{k_n} a_{ni} X_i \to 0 \quad w.p.1. \quad \text{as} \quad n \to \infty\ .$$

THEOREM 3.2. (Strong consistency) *Suppose* $g(x)$ *is bounded on* $[0, 1]$. *Assume*

(3.11) $$\gamma(n) = O(n^{-1}) \quad \textit{and} \quad m(n) = [n^q] \quad \textit{for each fixed} \quad q \in (0, \min(1 - \alpha^{-1}, 2^{-1})) ,$$

where $[a]$ *denotes the largest integer not greater than* a *and* α *is as defined in* (1.1). *Then*

$$g_{nm(n)}(x) \to g(x) \quad \textit{w.p.1.} \quad \textit{as} \quad n \to \infty$$

at every continuity point $x \in [0, 1]$ *of the function* g.

PROOF. Fix any continuity point $x \in [0, 1]$ of g. Let $m = m(n)$. (3.11) implies (3.2). Since

$$g_{nm}(x) - g(x) = (g_{nm}(x) - Eg_{nm}(x)) + (Eg_{nm}(x) - g(x)) ,$$

by Lemma 3.1 it suffices to show that

(3.12) $$g_{nm}(x) - Eg_{nm}(x) = \sum_{i=1}^{n} a_{ni} Z_i \to 0 \quad \text{w.p.1.} \quad \text{as} \quad n \to \infty ,$$

where

$$a_{ni} = \int_{A_i} \delta_m(x, u)\, du \quad \text{for} \quad i = 1, \ldots, n .$$

First we consider the case $1 < \alpha \leq 2$. Set $k_n = n$ and $d_n = n^{1/\alpha}$. It follows from Conditions A(i) and A(iv) and (3.11) that

(3.13) $$|a_{ni}| \leq C_1 n^{q-1} \quad \text{for} \quad i = 1, \ldots, n$$

and

(3.14) $$\sum_{i=1}^{n} |a_{ni}| \leq C_2 \quad \text{for} \quad n \geq 1 ,$$

which yields that

$$\max_{1 \leq i \leq n} |a_{ni}| d_i \leq C_1 n^{q-1+1/\alpha} .$$

Hence, by $q < 1 - \alpha^{-1}$ we get

$$\max_{1 \leq i \leq n} |a_{ni}| d_i = O(1/\log n) .$$

Let Z be a random variable whose distribution is identical with that of Z_1. It is easy to see that

$$\sum_{n=1}^{\infty} P\{|Z| > d_n\} = \sum_{n=1}^{\infty} d_n^{\alpha} P\{d_n^{\alpha} < |Z|^{\alpha} \leq d_{n+1}^{\alpha}\} \leq E|Z|^{\alpha} < \infty .$$

By (3.13) and (3.14) we get

$$\sum_{i=1}^{n} a_{ni}^2 d_i^{2-\alpha} \leq C_2 n^{q-1} d_n^{2-\alpha} \sum_{i=1}^{n} |a_{ni}| \leq C_3 n^{q-2(1-1/\alpha)} ,$$

which, together with $q < 2(1 - 1/\alpha)$, implies that

$$\sum_{i=1}^{n} a_{ni}^2 d_i^{2-\alpha} = o(1/\log n) .$$

Thus, since all the conditions of Lemma 3.2 are fulfilled, Lemma 3.2 concludes (3.12). When $\alpha > 2$, by Hölder's inequality we have that $E|Z_1|^2 < \infty$. Hence Theorem 3.2 holds by the previous result with $\alpha = 2$. This completes the proof.

Remark. By the use of the kernel method, Georgiev (1985) obtained the strong consistency under the condition (1.1) with some $\alpha > 2$. Thus Theorem 3.2 includes the result of Georgiev (1985).

4. Examples

In this section we give several methods of estimating the regression function g by the use of delta sequences $\{\delta_m, m \geq 1\}$ on J with $J = [0, 1]$ satisfying Condition A.

(1) The kernel method: Let $K_i(x)$, $i = 0, 1, 2$, be bounded integrable Borel measurable functions on R satisfying

$$\int_{-\infty}^{\infty} K_i(x)\,dx = 1 \quad \text{for} \quad i = 0, 1, 2 ,$$

$$K_0(x) = 0 \quad \text{on} \quad (0, \infty) \quad \text{and} \quad K_1(x) = 0 \quad \text{on} \quad (-\infty, 0) .$$

Let $\{\lambda_m, m \geq 1\}$ is a sequence of positive numbers satisfying that $\lambda_m \to 0$ as $m \to \infty$ and $\lambda_m^{-1} = O(m)$. λ_m is so-called smoothing parameter. Set

$$\delta_m(x, y) = \begin{cases} \lambda_m^{-1} K_2(\lambda_m^{-1}(x - y)) & \text{for} \quad (x, y) \in J^{\circ} \times J , \\ \lambda_m^{-1} K_0(\lambda_m^{-1}(x - y)) & \text{for} \quad x = 0 \quad \text{and} \quad y \in J , \\ \lambda_m^{-1} K_1(\lambda_m^{-1}(x - y)) & \text{for} \quad x = 1 \quad \text{and} \quad y \in J . \end{cases}$$

Then $\{\delta_m,\ m \geq 1\}$ is a delta sequence on J satisfying Condition A.

(2) The Histogram method: Let

$$I_j(x) = I((j-1)/m \leq x < j/m) \quad \text{for} \quad j = 1, \ldots, m-1$$

and

$$I_m(x) = I((m-1)/m \leq x \leq 1)\,,$$

where $I(B)$ denotes the indicator function of B. Set

$$\delta_m(x,y) = m \sum_{j=1}^{m} I_j(x) I_j(y) \quad \text{for} \quad (x,y) \in J \times J\,.$$

This is a delta sequence on J with Condition A.

(3) The polynomial approximation method: Let

$$\bar{\delta}_m(x,y) = \{1-(y-x)^2\}^m \Big/ \int_{-1}^{1} (1-t^2)^m dt \quad \text{for} \quad (x,y) \in J \times J\,.$$

Set

$$\delta_m(x,y) = \begin{cases} \bar{\delta}_m(x,y) & \text{for} \quad (x,y) \in J^\circ \times J \\ 2\bar{\delta}_m(x,y) & \text{for} \quad x = 0, 1 \quad \text{and} \quad y \in J\,. \end{cases}$$

This $\{\delta_m,\ m \geq 1\}$ becomes a delta sequence on J satisfying Condition A.

(4) The characteristic function approach: Let

$$\bar{\delta}_m(x,y) = \sin^2 m(x-y)/\{\pi m(x-y)^2\} \quad \text{for} \quad (x,y) \in J \times J\,.$$

Put

$$\delta_m(x,y) = \begin{cases} \bar{\delta}_m(x,y) & \text{for} \quad (x,y) \in J^\circ \times J \\ 2\bar{\delta}_m(x,y) & \text{for} \quad x = 0, 1 \quad \text{and} \quad y \in J\,. \end{cases}$$

Then $\{\delta_m,\ m \geq 1\}$ is a delta sequence on J with Condition A.

Remark. $\delta_m(x,y)$ in (2), and $\bar{\delta}_m(x,y)$ in (3) and (4) are found in Walter and Blum (1979).

Acknowledgements

The author is grateful to the referees for helpful comments and for pointing out references overlooked in the original draft of the paper.

REFERENCES

Ahmad, I. A. and Lin, P. E. (1984). Fitting a multiple regression function, *J. Statist. Plann. Inference*, **9**, 163–176.

Chen, K. F. and Lin, P. E. (1981). Nonparametric estimation of a regression function, *Z. Wahrsch. Verw. Gebiete*, **57**, 223–233.

Eubank, R. L. (1988). *Spline Smoothing and Nonparametric Regression*, Marcel Dekker, New York and Basel.

Galkowski, T. and Rutkowski, L. (1986). Nonparametric fitting of multivariate functions, *IEEE Trans. Automat. Control*, **31**, 785–787.

Gasser, T. and Müller, H. G. (1979). Kernel estimation of regression functions, *Smoothing Techniques for Curve Estimation*, (eds. T. Gasser and M. Rosenblatt), *Lecture Notes in Math.*, **757**, 26–68, Springer, Berlin-New York.

Gasser, T. and Müller, H. G. (1984). Estimating regression functions and their derivatives by the kernel method, *Scand. J. Statist.*, **11**, 171–185.

Georgiev, A. A. (1984*a*). Speed of convergence in nonparametric kernel estimation of a regression function and its derivatives, *Ann. Inst. Statist. Math.*, **36**, 455–462.

Georgiev, A. A. (1984*b*). On the recovery of functions and their derivatives from imperfect measurements, *IEEE Trans. Systems Man Cybernet.*, **14**, 900–903.

Georgiev, A. A. (1985). Nonparametric kernel algorithm for recovery of functions from noisy measurements with applications, *IEEE Trans. Automat. Control*, **30**, 782–784.

Georgiev, A. A. (1988). Consistent nonparametric multiple regression: The fixed design case, *J. Multivariate Anal.*, **25**, 100–110.

Georgiev, A. A. and Greblicki, W. (1986). Nonparametric function recovering from noisy observations, *J. Statist. Plann. Inference*, **13**, 1–14.

Priestley, M. B. and Chao, M. T. (1972). Nonparametric function fitting, *J. Roy. Statist. Soc. Ser. B*, **34**, 385–392.

Rice, J. and Rosenblatt, M. (1983). Smoothing splines: Regression, derivatives and deconvolution, *Ann. Statist.*, **11**, 141–156.

Rutkowski, L. and Rafajlowicz, E. (1989). On optimal global rate of convergence of some nonparametric identification procedures, *IEEE Trans. Automat. Control*, **34**, 1089–1091.

Stone, C. J. (1977). Consistent nonparametric regression, *Ann. Statist.*, **5**, 595–645.

Teicher, H. (1985). Almost certain convergence in double arrays, *Z. Wahrsch. Verw. Gebiete*, **69**, 331–345.

Walter, G. and Blum, J. (1979). Probability density estimation using delta sequences, *Ann. Statist.*, **7**, 328–340.