NONPARAMETRIC FUNCTION RECOVERING FROM NOISY OBSERVATIONS

Alexander A. GEORGIEV and Włodzimierz GREBLICKI
Institute of Engineering Cybernetics, Technical University of Wrocław,
Wyspiańskiego 27, 50-370 Wrocław, Poland

printed in:
JOURNAL OF STATISTICAL PLANNING AND INFERENCE
Vol. 13, pp. 1–14, 1986

Abstract

Abstract: We consider the nonparametric regression model $Y_i = g(x_i) + \zeta_i$, where g is a bounded function over the interval $[0, 1]$, to be estimated. x_i's are nonrandom and ζ_i's are independent identically distributed random variables with $E\zeta_i = 0$. This paper studies the behavior of the general family of nonparametric estimates $g_n(x) = \sum_{i=1}^{n} Y_i w_{ni}(x)$, where the weight functions $\{w_{ni}\}$ are of the form $w_{ni}(x) = w_{ni}(x; x_1..., x_n)$, $i = 1, \ldots, n$. The family of estimates includes all known estimates proposed by Priestley and Chao (1972), Clark (1977), Gasser and Müller (1979), Cheng and Lin (1981) as well as Georgiev (1984b, 1985). Sufficient conditions for mean square and complete convergence are derived. New results for the Priestley-Chao and Gasser-Müller-Cheng-Lin estimates are obtained. Also proposed is a class of new nearest neighbor estimates of g. Finally, a simulation experiment demonstrates the remarkable success of the nearest neighbor technique with bandwidth depending on the local density of the design points.

AMS Subject Classifications: Primary 62G05, 62J02; Secondary 60F15.

Key words and phrases: Regression function; Nonparametric estimation; Nearest neighbor estimate; Kernel estimate; Consistency; Strong pointwise convergence; Curve fitting.

1. Introduction

Let $Y_1, ..., Y_n$ be observations according to the model

$$Y_i = g(x_i) + \zeta_i,\ i = 1, \ldots, n,$$

taken at fixed points $x_1, \ldots, x_n$, where ζ_i's are independent random variables having zero mean and variance not greater than σ^2. We assume that g is bounded over the interval $[0,1]$ and $0 = x_0 < x_1 < \cdots < x_n \le 1$. Estimates of $g(x)$ examined in this paper are of the following form:

$$g_n(x) = \sum_{i=1}^{n} Y_i w_{ni}(x), \tag{1.1}$$

where $w_{ni}(x) = w_{ni}(x; x_1 ..., x_n)$, $i = 1, \ldots, n$, are weight functions satisfying some conditions. For weights

$$\begin{aligned} &\hat{w}_{ni}(x) = (x_{i+1} - x_i)\frac{1}{h_n} K\left(\frac{x - x_i}{h_n}\right), \; i = 1, \ldots, n-1, \\ &\hat{w}_{nn}(x) = 0, \end{aligned} \tag{1.2}$$

where K is a kernel function and $\{h_n\}$ is a positive number sequence, the estimate becomes that of Priestley and Chao (1972). Putting

$$\begin{aligned} &\bar{w}_{ni}(x) = \frac{1}{h_n}\int_{x_i}^{x_{i+1}} K\left(\frac{x - y}{h_n}\right) dy, \; i = 1, \ldots, n-1, \\ &\bar{w}_{nn}(x) = 0, \end{aligned} \tag{1.3}$$

we obtain an estimate suggested by Gasser and Müller (1979) and Cheng and Lin (1981). In turn, defining

$$\tilde{w}_{ni}(x) = \frac{K\left(\dfrac{x - x_i}{h_n}\right)}{\displaystyle\sum_{i=1}^{n} K\left(\dfrac{x - x_i}{h_n}\right)}, \; i = 1, \ldots, n, \tag{1.4}$$

we get an estimate mentioned by Gasser and Müller (1979). The last estimate has been extensively studied in the literature in the case when observations are taken at random points, see e.g. Stone (1977), Spiegelman and Sacks (1980), Devroye and Wagner (1980) or Greblicki, Krzyżak and Pawlak (1984). For more work and bibliography on nonparametric regression estimation see Collomb (1981, 1985).

For large h_n, it is clear that in all the above estimates even Y_i's taken at points distant from x can play a significant role. On the other hand, for small h_n, the estimates are rough functions of x. Moreover, due to the fact that the bandwidth parameter h_n is controlled only by n, local properties of the estimates can be not satisfactory, especially for g varying rapidly in a wide range and x_i's not distributed uniformly. Thus it seems reasonable to select the bandwidth parameter as a function of not only n, but also of x and x_i's. For x having many close neighbors among the x_i's, the parameter should be small, while for x with a small number of x_i's in its vicinity a large value of the parameter is recommended. This suggests replacing h_n by $r_n = r_n(x; x_1, \ldots, , x_n)$ in (1.2), (1.3) and (1.4), where $r_n = r_n(x; x_1, \ldots, , x_n)$ is the distance from x to the k_n-th nearest neighbor among the x_i's, and where $\{k_n\}$ is an integer sequence. In this way we obtain the next three modified estimates with weight functions

$$\begin{aligned} &\hat{w}'_{ni}(x) = (x_{i+1} - x_i)\frac{1}{r_n} K\left(\frac{x - x_i}{r_n}\right), \; i = 1, \ldots, n-1, \\ &\hat{w}'_{nn}(x) = 0, \end{aligned} \tag{1.5}$$

and

$$
\begin{aligned}
&\bar{w}'_{ni}(x) = \frac{1}{r_n}\int_{x_i}^{x_{i+1}} K\left(\frac{x-y}{r_n}\right) dy,\ i = 1, \ldots, n-1, \\
&\bar{w}'_{nn}(x) = 0,
\end{aligned} \tag{1.6}
$$

and

$$
\tilde{w}'_{ni}(x) = \frac{K\left(\dfrac{x-x_i}{r_n}\right)}{\displaystyle\sum_{i=1}^{n} K\left(\dfrac{x-x_i}{r_n}\right)},\ i = 1, \ldots, n. \tag{1.7}
$$

Recently Silverman (1984) has considered the spline smoothing approach to nonparametric regression and curve estimation. He has showed that, in a certain sense, spline smoothing corresponds approximately to smoothing by a kernel method with bandwidth depending on the local density of design points x_i. This result corresponds with our nearest neighbor (or data adaptive) approach.

In this paper we give two theorems establishing mean square and complete consistency of estimate (1.1). In turn, consistency theorems concerning estimates applying weights (1.2), (1.3) and (1.4) and their modified forms (1.5), (1.6) and (1.7) are more or less direct consequences of those general ones. They, moreover, improve some results known in the literature. Finally in Section 5, a simulation experiment demonstrates how well the estimate (1.1) with new modified weights works in practice.

It will be convenient to introduce the following notations:

$$
\max_i |x_{i+1} - x_i| = \Delta_n \text{ and } \min_i |x_{i+1} - x_i| = \delta_n.
$$

By $C(g)$ we denote a set of all $x \in (0,1)$ at which g is continuous.

2. General consistency theorems

In this section we prove two theorems on mean square and complete consistency of estimate (1.1).

Theorem 1 *Let*

$$
\sup_n \sum_{i=1}^{n} |w_{ni}(x)| \le d < \infty, \tag{2.1}
$$

some d,

$$
\sum_{i=1}^{n} w_{ni}(x) \to 1 \text{ as } n \to \infty, \tag{2.2}
$$

and

$$
\sum_{i=1}^{n} |w_{ni}(x)| I_{\{|x-x_i|>\varepsilon\}} \to 0 \text{ as } n \to \infty, \text{ all } \varepsilon > 0. \tag{2.3}
$$

If, moreover,

$$\sum_{i=1}^{n} w_{ni}^2(x) \to 0 \ \textit{as}\ n \to \infty, \tag{2.4}$$

then

$$E[g_n(x) - g(x)]^2 \to 0 \ \textit{as}\ n \to \infty,$$

provided that $x \in C(g)$.

Remark 1 For nonnegative weights, which is often the case, (2.2) implies (2.1).

Proof of Theorem 1. Let $x \in C(g)$ and let $\varepsilon > 0$. For the bias we get the following upper bound:

$$\begin{aligned}
|Eg_n(x) - g(x)| &\le \sum_{i=1}^{n} |g(x_i) - g(x)||w_{ni}(x)| I_{\{|x-x_i| \le \varepsilon\}} \\
&+ \sum_{i=1}^{n} |g(x_i) - g(x)||w_{ni}(x)| I_{\{|x-x_i| > \varepsilon\}} \\
&+ |g(x)| \left| \sum_{i=1}^{n} w_{ni}(x) - 1 \right| \\
&\le d \sup_{|x-y| \le \varepsilon} |g(x) - g(y)| + 2 \sup_y |g(y)| \sum_{i=1}^{n} |w_{ni}(x)| I_{\{|x-x_i| > \varepsilon\}} \\
&+ \sup_y |g(y)| \left| \sum_{i=1}^{n} w_{ni}(x) - 1 \right|.
\end{aligned}$$

The first term in the above inequality can be made arbitrarily small by choosing ε small enough. Since the other two terms vanish as n tends to infinity, the estimate is asymptotically unbiased.

Since, for variance, we get

$$\operatorname{var} g_n(x) \le \sigma^2 \sum_{i=1}^{n} w_{ni}^2(x),$$

the theorem has been proved. □

The next theorem establishes complete consistency.

Theorem 2 *Let*

$$|\zeta_i| \le \gamma < \infty \ \textit{almost surely},\ i = 1, \ldots, n. \tag{2.5}$$

If, in addition to (2.1), (2.2) and (2.3)

$$\sum_{i=1}^{\infty} \exp\left(-\frac{\alpha}{\sup_i |w_{ni}(x)|} \right) < \infty, \ \textit{all}\ \alpha > 0, \tag{2.6}$$

then

$$g_n(x) \to g(x) \text{ completely as } n \to \infty,$$

provided that $x \in C(g)$.

Remark 2 Restriction (2.6) is satisfied if

$$\sup_i |w_{ni}(x)| \log n \to 0 \text{ as } n \to \infty.$$

Proof of Theorem 2. By virtue of Theorem 1, we now need to show that

$$|g_n(x) - Eg_n(x)| \to 0 \text{ as } n \to \infty \text{ completely.}$$

This convergence is implied by the inequality

$$P\{|g_n(x) - Eg_n(x)| \geq t\} \leq 2 \exp\left(-\frac{M}{\sup_i |w_{ni}(x)|}\right),$$

where $t > 0$ and $M = t^2/2\gamma(\gamma d + t/3)$, which can be obtained by employing Bernstein's inequality, see e.g. Bennett (1962). By this and the definition of complete convergence, see e.g. Stout (1974, p. 225), we get the desired result. □

The following result, an immediate corollary of Theorem 1, Theorem 2 and the Lebesgue dominated convergence theorem, establishe the mean and almost sure integral consistency of the estimate (1.1) under restriction (2.5).

Corollary 1 *Let for some* $c < \infty$, $|g_n(x)| \leq c$ *and* $|g(x)| \leq c$. *Then*

$$E \int_0^1 |g_n(x) - g(x)| dx \to 0, \tag{2.7}$$

or

$$\int_0^1 |g_n(x) - g(x)| dx \to 0, \text{ almost surely as } n \to \infty \tag{2.8}$$

when $g_n(x) \to g(x)$ *as* $n \to \infty$ *in probability or almost surely, respectively.*

3. Priestley-Chao estimate

In this section we examine estimate (1.1) using weights in its original (1.2) and modified (1.5) forms, i.e. estimates

$$\hat{g}_n(x) = \sum_{i=1}^{n-1} Y_i (x_{i+1} - x_i) \frac{1}{h_n} K\left(\frac{x - x_i}{h_n}\right)$$

and

$$\hat{g}'_n(x) = \sum_{i=1}^{n-1} Y_i (x_{i+1} - x_i) \frac{1}{r_n} K\left(\frac{x - x_i}{r_n}\right).$$

We shall need the following lemma the proof of which is deferred to the Appendix.

Lemma 1 *Let K be almost everywhere continuous on R. Let K have a majorant, i.e. let $|K(x)| \leq H(x)$, all $x \in R$, where H is symmetric, nonincreasing on $[0, \infty)$ and $\int H(y)dy < \infty$. Let*

$$\Delta_n \leq \Delta/n, \ \Delta > 0. \tag{3.1}$$

If

$$h_n \to 0 \text{ as } n \to \infty \tag{3.2}$$

and

$$nh_n \to \infty \text{ as } n \to \infty \tag{3.3}$$

then

$$\sum_{i=1}^{n-1}(x_{i+1} - x_i)\frac{1}{h_n}K\left(\frac{x - x_i}{h_n}\right) \to \int K(y)dy \text{ as } n \to \infty$$

for every $x \in C(g)$.

Theorem 3 *Let K be a probability density function and let $\sup_x H(x) < \infty$. If the conditions of the lemma are satisfied, then*

$$E[\hat{g}_n(x) - g(x)]^2 \to 0 \text{ as } n \to \infty,$$

all $x \in C(g)$.

The above theorem improves results in Priestley and Chao (1972) and Cheng and Lin (1981), since the class of kernels applied by us is wider than theirs. Moreover, their condition $nh_n^2 \to \infty$ as $n \to \infty$ is much more restrictive than our (3.3).

Proof of Theorem 3. In the proof we shall verify conditions (2.1)–(2.4) of the general Theorem 1. Since the weights are nonnegative, (2.1) is implied by (2.2), which, in turn, is implied by the lemma. Moreover, by the assumed properties of the kernel

$$\sum_{i=1}^{n-1}(x_{i+1} - x_i)\frac{1}{h_n}K\left(\frac{x - x_i}{h_n}\right)I_{\{|x-x_i|>\varepsilon\}} \leq \frac{n\Delta_n}{\varepsilon}\sup_{u>\varepsilon/h_n}|u|H(u). \tag{3.4}$$

In view of this and (3.2), condition (2.3) has been verified. At last,

$$\begin{aligned}&\sum_{i=1}^{n-1}(x_{i+1} - x_i)^2\frac{1}{h_n^2}K^2\left(\frac{x - x_i}{h_n}\right)\\ &\leq [\sup_u H(u)]\frac{\Delta_n}{h_n}\sum_{i=1}^{n-1}(x_{i+1} - x_i)\frac{1}{h_n}K\left(\frac{x - x_i}{h_n}\right),\end{aligned}$$

which approaches zero as n tends to infinity. Thus, (2.4) is also satisfied and the theorem has been proved. □

Theorem 4 *Let (2.5) hold, let K satisfy appropriate conditions of Theorem 3, and let (3.1) be fulfilled. If, in addition to (3.2),*

$$\sum_{n=1}^{\infty} \exp(-\alpha n h_n) < \infty, \ \textit{all } \alpha > 0, \tag{3.5}$$

then

$$\hat{g}_n(x) \to g(x) \ \textit{as } n \to \infty \ \textit{completely}$$

all $x \in C(g)$.

No result comparable with Theorem 4 is known to the authors, but it seems worth mentioning that strong consistency of the estimate has been examined by Benedetti (1977) as well as Cheng and Lin (1981).

Proof of Theorem 4. The condition (2.6) is fulfilled in view of

$$\sup_i \hat{w}_{ni}(x) \le \frac{\Delta}{n h_n} \sup_u H(u).$$

□

We now present two theorems concerning the estimate (1.1) using modified nearest neighbor weights (1.5). The strong consistency of $\hat{g}'_n$ has been examined by Georgiev (1985).

Theorem 5 *Let K satisfy appropriate conditions of Theorem 3. Let, in addition to (3.1),*

$$\delta/n \le \delta_n, \ \delta > 0, \tag{3.6}$$

$$k_n/n \to 0 \ \textit{as } n \to \infty \tag{3.7}$$

and

$$k_n \to \infty \ \textit{as } n \to \infty. \tag{3.8}$$

Then

$$E[\hat{g}'_n(x) - g(x)]^2 \to 0 \ \textit{as } n \to \infty,$$

all $x \in C(g)$.

Proof Recalling (3.4) we find

$$\sum_{i=1}^{n-1} (x_{i+1} - x_i) \frac{1}{r_n} K\left(\frac{x - x_i}{r_n}\right) I_{\{|x - x_i| > \varepsilon\}} \le \frac{\Delta}{\varepsilon} \sup_{u > \varepsilon n / \Delta k_n} |u| H(u)$$

and then easily verify (2.1) and the other condition of Theorem 1. □

Theorem 6 *Let (2.5) hold, let K satisfy appropriate conditions of Theorem 3, and let (3.1) and (3.6) be satisfied. If, in addition to (3.7),*

$$\sum_{n=1}^{\infty} \exp(-\alpha k_n) < \infty, \text{ all } \alpha > 0, \tag{3.9}$$

then

$$\hat{g}'_n(x) \to g(x) \text{ as } n \to \infty \text{ completely,}$$

all $x \in C(g)$.

Proof It is enough to see that

$$\sup_i \hat{w}'_{ni}(x) \le \frac{\Delta}{\delta} \frac{1}{k_n} \sup_u H(u).$$

□

4. Cheng-Lin and Gasser-Müller estimate

We now examine estimate (1.1) using weights (1.3), i.e.

$$\bar{g}_n(x) = \sum_{i=1}^{n-1} Y_i \frac{1}{h_n} \int_{x_i}^{x_{i+1}} K\left(\frac{x-y}{h_n}\right) dy,$$

and its modified form (1.6), i.e.

$$\bar{g}'_n(x) = \sum_{i=1}^{n-1} Y_i \frac{1}{r_n} \int_{x_i}^{x_{i+1}} K\left(\frac{x-y}{r_n}\right) dy.$$

The speed of convergence of g'_n, to g is treated by Georgiev (1984b).

Theorem 7 *Under the conditions of Theorem 3,*

$$E[\bar{g}_n(x) - g(x)]^2 \to 0 \text{ as } n \to \infty,$$

all $x \in C(g)$. Under the conditions of Theorem 4,

$$\bar{g}_n(x) \to g(x) \text{ as } n \to \infty, \text{ completely}$$

all $x \in C(g)$.

A similar result for weak consistency was achieved by Gasser and Müller (1979) and by Cheng and Lin (1981), but only for kernels with bounded support and for g continuous on the whole interval $[0, 1]$.

Proof of Theorem 7. We easily notice that (2.2) is satisfied and

$$\sum_{i=1}^{n-1} Y_i \frac{1}{h_n} \int_{x_i}^{x_{i+1}} K\left(\frac{x-y}{h_n}\right) dy I_{\{|x-y|>\varepsilon\}} \le \frac{1}{h_n} \int_{|x-y|\ge\varepsilon-\Delta_n} K\left(\frac{x-y}{h_n}\right) dy$$

which, by (3.2), converges to zero as n tends to infinity. Verifying (2.4), we get

$$\sum_{i=1}^{n-1}\left[\frac{1}{h_n}\int_{x_i}^{x_{i+1}} K\left(\frac{x-y}{h_n}\right) dy\right]^2 \le [\sup_u H(u)]\frac{\Delta_n}{h_n}\frac{1}{h_n}\int_0^1 K\left(\frac{x-y}{h_n}\right) dy$$

which also approaches zero. Condition (2.6) is fulfilled in view of

$$\sup_i\left[\frac{1}{h_n}\int_{x_i}^{x_{i+1}} K\left(\frac{x-y}{h_n}\right) dy\right] \le [\sup_u H(u)]\frac{\Delta}{nh_n}.$$

Thus the theorem has been proved. □

The next theorem concerning the modified estimate can be now proved with ease.

Theorem 8 *Under the conditions of Theorem 5 or 6*

$$E[\bar{g}_n'(x) - g(x)]^2 \to 0 \text{ as } n \to \infty,$$

or

$$\bar{g}_n'(x) \to g(x) \text{ as } n \to \infty, \text{ completely,}$$

all $x \in C(g)$.

5. Kernel estimate

In this section we deal with the following two estimates:

$$\tilde{g}_n(x) = \frac{\sum_{i=1}^{n} Y_i K\left(\frac{x-x_i}{h_n}\right)}{\sum_{i=1}^{n} K\left(\frac{x-x_i}{h_n}\right)}$$

and

$$\tilde{g}_n'(x) = \frac{\sum_{i=1}^{n} Y_i K\left(\frac{x-x_i}{r_n}\right)}{\sum_{i=1}^{n} K\left(\frac{x-x_i}{r_n}\right)}.$$

We will treat 0/0 in the above estimates as 0.

Theorem 9 *Let K satisfy appropriate conditions of Theorem 3 and let, moreover,*

$$cI_{\{|x|\le a\}} \le K(x), \text{ all } x \in R, \tag{5.1}$$

with c and a positive. Let (3.1), (3.2) and (3.3) be fulfilled. Then

$$E[\tilde{g}_n(x) - g(x)]^2 \to 0 \text{ as } n \to \infty,$$

all $x \in C(g)$. If, in addition, (2.5) and (3.5) hold, then

$$\tilde{g}_n(x) \to g(x) \text{ as } n \to \infty, \text{ completely,}$$

all $x \in C(g)$.

Proof. Clearly, by the properties of $H(x)$,

$$\sum_{i=1}^{n} K\left(\frac{x - x_i}{h_n}\right) I_{\{|x-y|>\varepsilon\}} \le \frac{nh_n}{\varepsilon} \sup_{u>\varepsilon/h_n} |u|H(u)$$

and by (5.1),

$$\sum_{i=1}^{n} K\left(\frac{x - x_i}{h_n}\right) \ge c \sum_{i=1}^{n} I_{\{|x-y|\le ah_n\}} \ge ac\frac{h_n}{\Delta_n}.$$

Hence

$$\frac{\sum_{i=1}^{n} K\left(\frac{x - x_i}{h_n}\right) I_{\{|x-y|>\varepsilon\}}}{\sum_{i=1}^{n} K\left(\frac{x - x_i}{h_n}\right)} \le \frac{\Delta}{ac\varepsilon} \sup_{u>\varepsilon/h_n} |u|H(u)$$

which implies (2.3). Verification of (2.4) is now easy. We see that (2.6) is satisfied, since

$$\sup_i \frac{K\left(\frac{x - x_i}{h_n}\right)}{\sum_{j=1}^{n} K\left(\frac{x - x_j}{h_n}\right)} \le \frac{\Delta \sup_u H(u)}{ac} \frac{1}{nh_n}.$$

Theorem 9 is now proved. □

The next theorem examines the estimate (1.1) with nearest neighbor weights (1.7).

Theorem 10 *Let K satisfy appropriate conditions of Theorem 9. Let (3.1), (3.7) and (3.8) hold. Then*

$$E[\tilde{g}'_n(x) - g(x)]^2 \to 0 \text{ as } n \to \infty,$$

all $x \in C(g)$. If, in addition, (2.5) and (3.9) hold, then

$$\tilde{g}'_n(x) \to g(x) \text{ as } n \to \infty, \text{ completely,}$$

all $x \in C(g)$.

Proof. In order to check condition (2.3) we see that

$$\frac{\sum_{i=1}^{n} K\left(\frac{x - x_i}{r_n}\right) I_{\{|x-y|>\varepsilon\}}}{\sum_{i=1}^{n} K\left(\frac{x - x_i}{r_n}\right)} \le \frac{\Delta}{ac\varepsilon} \sup_{u>(\varepsilon/\Delta)(n/k_n)} |u|H(u)$$

since $r_n \le k_n \Delta_n$. For (2.4) we have

$$\frac{\sum_{i=1}^{n} K^2\left(\frac{x-x_i}{r_n}\right)}{\left[\sum_{i=1}^{n} K\left(\frac{x-x_i}{r_n}\right)\right]^2} \le \frac{\sup_u H(u)}{ac}\frac{1}{k_n}.$$

For the complete convergence part of the theorem we obtain

$$\exp\left(-a \Big/ \sup_i \frac{K\left(\frac{x-x_i}{h_n}\right)}{\sum_{j=1}^{n} K\left(\frac{x-x_j}{h_n}\right)}\right) \le \exp(-\alpha' k_n),$$

where $\alpha' = \alpha ac/\sup_u H(u) > 0$. Theorem 4 is now obvious. □

6. Monte Carlo study

In this section we present a simulation experiment in which we compare the estimates $\tilde{g}_n$ and $\tilde{g}'_n$ developed in Section 5. The quality of the estimates is measured by (2.7), i.e.

$$V_{in} = E\int_0^1 |g_{in}(x) - g(x)|dx, \tag{6.1}$$

where $g_{1n}(x) = \tilde{g}_n(x)$ and $g_{2n}(x) = \tilde{g}'_n(x)$. In the experiment

$$g(x) = 5 + \sin(15x + 1.5) + 0.25\sin(56x),$$

i.e. the regression is a highly nonlinear function, estimated earlier experimentally by Clark (1977). The points x_i, of total number n, were designed by a pseudo-random generator, uniform on the interval $[-0.5, 1.5]$. Due to the fact that the error (6.1) is evaluated only on the interval $[0, 1]$, it is not sensitive to boundary effects, see e.g. Gasser and Müller (1979). In turn ζ_i's were normal pseudo-random variables with standard deviation 0.1 and generated with the Box and Muller (1958) algorithm. We used the Epanechnikov kernel, i.e.

$$K(u) = \begin{cases} 3(1-u^2)/4, & \text{for } |u| \le 1 \\ 0, & \text{elsewhere.} \end{cases}$$

The errors V_{1n} and V_{2n} were calculated for various n, h_n and k_n, and the results are given in Tables 1 and 2. In each column the errors take, for some values of the smoothing parameters h_n and k_n, respectively, their minimal values. These values are printed boldface.

From the tables it follows that $\tilde{g}'_n$ behaves better than $\tilde{g}_n$, especially for small and moderate n. For estimate $\tilde{g}_n$, optimal values of h_n lie in the interval $[0.02, 0.12]$.

Table 1: Results of simulation with kernel estimate $\tilde{g}_n$

h_n	$n = 25$	$n = 50$	$n = 100$	$n = 200$
0.005	439.1331	405.7623	306.0313	133.2968
0.02	261.2593	124.7182	58.7409	**4.7135**
0.04	169.7729	72.6557	**10.8356**	8.0779
0.06	116.0702	42.3868	14.8583	12.5504
0.08	78.5389	24.0763	18.1068	16.3145
0.10	45.9929	**23.7659**	20.0929	19.3349
0.12	**30.0843**	25.8120	20.8081	21.0777
0.14	34.5163	26.8840	22.6720	23.4748

Table 2: Results of simulation with kernel estimate $\tilde{g}'_n$

k_n	$n = 5$	$n = 50$	$n = 100$	$n = 200$
1	495.1033	495.1033	495.1033	495.1033
2	**26.3269**	17.5842	9.0826	4.6167
3	31.5988	**15.7310**	**8.2890**	4.6153
4	33.3092	20.5010	9.3998	**4.3636**
5	35.4876	22.2012	9.7131	4.6112
6	39.1261	22.9165	12.1065	4.6209
7	43.9114	24.9092	13.5388	5.5246

In this range of the bandwidth parameter the error, for e.g., $n = 100$, varies from 10.8 to 58.7. In turn, for $\tilde{g}'_n$, the optimal values of k_n are in a set $\{2, 3, 4\}$ and the error for $k_n = 2, 3, 4$ varies from 8.3 to 9.4, $n = 100$. Thus it follows that estimate in is more sensitive to the nonoptimal choice of the smoothing parameter.

From the simulation experiment it follows that the nearest neighbor estimate $\tilde{g}'_n$ in is worth recommendation for small and moderate sample sizes. A similar conclusion is true also for the other nearest neighbor estimates studied in this paper.

In practical situations, when the regression $g(x)$ is unknown, the most popular technique for optimal choice of the smoothing parameter is the cross-validation method. For more details the reader should consult recent contributions of Wong (1983), Li (1984) and Rice (1984).

Appendix

Proof of the lemma. Denoting $v_i = (x_i - x)/h_n$, we find $v_1 < v_2 < \cdots < v_n$ and $v_1 = -x/h_n$, $v_n = (1 - x)/h_n$. Clearly

$$\frac{1}{h_n} \sum_{i=1}^{n-1} (x_{i+1} - x_i) K \left(\frac{x - x_i}{h_n} \right) = \frac{1}{h_n} \sum_{i=1}^{n-1} (v_{i+1} - v_i) K \left(-v_i \right).$$

Let d be a positive number and let A, B and C be sets of all $i = \{1, \ldots, n\}$ for which $v_i < -d$, $|v| \leq d$ and $v_i > d$, respectively. For n sufficiently large, we find A,

B and C nonempty and

$$\begin{aligned}
&\sum_{i=1}^{n-1}(v_{i+1}-v_i)K(-v_i)-\int K(y)dy \\
&=\sum_{i\in A}(v_{i+1}-v_i)K(-v_i)+\left[\sum_{i\in B}(v_{i+1}-v_i)K(-v_i)-\int_{|y|\le d}K(y)dy\right] \\
&+\sum_{i\in C}(v_{i+1}-v_i)K(-v_i)+\int_{|y|>d}K(y)dy \\
&=Q_1+Q_2+Q_3+Q_4,
\end{aligned}$$

say. Since H is monotone on $(-\infty,0]$ and $[0,\infty)$,

$$|Q_1|\le\sum_{i\in A}(v_{i+1}-v_i)H(v_i)\le\int_{y<-d}H(y)\,dy,$$

and

$$|Q_3|\le\sum_{i\in C}(v_{i+1}-v_i)H(v_i)\le\int_{y<-d-\Delta_n/h_n}H(y)\,dy.$$

Taking d and n large enough, we can make $|Q_1|$, $|Q_3|$ and $|Q_4|$ arbitrarily small. Recalling that K is Riemann integrable and observing that $\max_i|v_{i+1}-v_i|=\Delta_n/h_n$ approaches zero as n tends to infinity, we find $|Q_2|\to 0$ as $n\to\infty$. Thus the lemma has been proved. □

References

Benedetii, J.K. (1977). On the nonparametric estimation of regression function. *J. Roy. Statist. Soc. Ser. B* 39, 248-253.

Bennett, G. (1962). Probability inequalities for the sum of independent random variables. *J. Amer. Statist. Assoc.* 57, 33-45.

Box, G.E.P. and M.E. Muller (1958). A note on the generation of random normal deviates. *Ann. Math. Statist.* 29, 610-611.

Cheng, K.F. and P.E. Lin (1981). Nonparametric estimation of regression function. *Z. Warsch. Verw. Gebiete* 57, 223-233.

Clark, R.M. (1977). Non-parametric estimation of a smooth regression function. *J. Roy. Statist. Soc. Ser. B* 39, 107-113.

Collomb, G. (1981). Estimation non parametrique de la regression: revue bibliographique. *Internat. Statist. Rev.* 49, 75-93.

Collomb, G. (1985). Nonparametric regression: an up-to-date bibliography. *Statistics* 16, 309-324.

Devroye, L.P. and T.J. Wagner (1980). Distribution-free consistency results in nonparametric discrimination and regression function estimation. *Ann. Statist.* 8, 231-239.

Gasser, T. and H.G. Müller (1979). Kernel estimation of regression function. In: T. Gasser and M. Rosenblatt, Eds., *Smoothing Techniques for Curve Estimation*, Lecture Notes in Math. No. 757. Springer, Berlin-New York, 23-68.

Georgiev, A.A. (1983). Local properties of function fitting estimates with application to system identification. In: *Proc. Fourth Pannonian Symposium on Mathematical Statistics*, Sept. 1983, Bad Tatzmannsdorf, Austria.

Georgiev, A.A. (1984a). Kernel estimates of functions and their derivatives with applications. *Statist. Probab. Lett.* 2, 45-50.

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

Georgiev, A.A. (1985). The strong pointwise convergence of nearest neighbor function fitting algorithm with application to system identification. *Kybernetika* 21, 118-122.

Greblicki, W., A. Krzyżak and M. Pawlak (1984). Distribution-free pointwise consistency of kernel regression estimate. *Ann. Statist.* 12, 1570-1575.

Li, K.C. (1984). Consistency for cross-validated nearest neighbor estimates in nonparametric regression. *Ann. Statist.* 12, 230-240.

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

Rice, J. (1984). Bandwidth choice for nonparametric regression. *Ann. Statist.* 12, 1215-1230.

Schuster, E. and S.J. Yakowitz (1979). Contribution to the theory of nonparametric regression with ap-plication to system identification. *Ann. Statist.* 7, 139-149.

Silverman, B.W. (1984). Spline smoothing: the equivalent variable kernel method. *Ann. Statist.* 12, 898-916.

Spiegelman, C. and J. Sacks (1980). Consistent window estimation in nonparametric regression. *Ann. Statist.* 8, 240-246.

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

Stout, W.F. (1974). *Almost Sure Convergence*, Academic Press, New York.