Reprinted from JOURNAL OF MULTIVARIATE ANALYSIS
All Rights Reserved by Academic Press, New York and London

Vol. 23, No. 1, October 1987
Printed in Belgium

Necessary and Sufficient Consistency Conditions for a Recursive Kernel Regression Estimate

WŁODZIMIERZ GREBLICKI

Institute of Engineering Cybernetics, Technical University of Wrocław, Wrocław, Poland

AND

MIROSŁAW PAWLAK

Department of Electrical Engineering, The University of Manitoba, Winnipeg, Manitoba, Canada R3T 2N2

Communicated by P. Hall

A recursive kernel estimate $\sum_{i=1}^n Y_i K((x-X_i)/h_i)/\sum_{j=1}^n K((x-X_j)/h_j)$ of a regression $m(x)=E\{Y|X=x\}$ calculated from independent observations $(X_1, Y_1),..., (X_n, Y_n)$ of a pair (X, Y) of random variables is examined. For $E|Y|<\infty$, the estimate is weakly pointwise consistent for almost all $(\mu)\, x \in R^d$, μ is the probability measure of X, if and only if $\sum_{i=1}^n h_i^d I_{\{h_i>\varepsilon\}}/\sum_{j=1}^n h_j^d \to 0$ as $n\to\infty$, all $\varepsilon>0$, and $\sum_{i=1}^\infty h_i^d=\infty$, d is the dimension of X. For $E|Y|^{1+\delta}<\infty$, $\delta>0$, the estimate is strongly pointwise consistent for almost all $(\mu)\, x\in R^d$, if and only if the same conditions hold. For $E|Y|^{1+\delta}<\infty$, $\delta>0$, weak and strong consistency are equivalent. Similar results are given for complete convergence.

1. INTRODUCTION

Let $(X_1, Y_1),..., (X_n, Y_n)$ be a sample of independent observations of a pair (X, Y) of random variables of which X takes values in R^d, while Y in R. Let μ be the probability measure of X. We examine the kernel estimate

$$\hat{m}(x)=\frac{\sum_{i=1}^n Y_i K(x-X_i)/h_i)}{\sum_{i=1}^n K((x-X_i)/h_i)}$$

Received November 21, 1984; revised November 20, 1986.

AMS 1980 subject classification: primary 62G05.

Key words and phrases: regression function, nonparametric estimation, kernel estimate, recursive estimate, consistency.

0047-259X/87 $3.00

of the regression $m(x)=E\{Y|X=x\}$, where $\{h_n\}$ is a sequence of positive numbers and K is a non-negative Borel kernel. The estimate is a recursive version of

$$\bar{m}(x)=\frac{\sum_{i=1}^{n} Y_i K((x-X_i)/h_n)}{\sum_{i=1}^{n} K((x-X_i)/h_n)}.$$

Devroye [3] showed that, for the window kernel, i.e., for a kernel which equals 1 or zero according as $\|x\|$ is smaller or greater than 1, respectively, $E|\bar{m}(x)-m(x)|^p \to 0$ as $n\to\infty$, whenever $E|Y|^p<\infty$ and $p\geqslant 1$. Krzyżak and Pawlak [6] proved that the estimate is weakly and strongly consistent for kernels having bounded support. In turn, kernels with unbounded support were applied by Greblicki *et al.* [5]. All of these authors examined pointwise convergence of the estimate at almost all (μ) $x\in R^d$.

For the recursive estimate studied in this article, its weak and strong consistency assuming the existence of a density of the measure μ have been investigated by Devroye and Wagner [4]. Distribution-free consistency of the estimate $\hat{m}$ for kernels with bounded support and under some restrictive assumptions on sequence $\{h_n\}$ was shown by Krzyżak and Pawlak [7]. In this article we apply kernels with bounded as well as unbounded support and give conditions on $\{h_n\}$ which are both necessary and sufficient in probability, almost sure and complete pointwise convergence of the estimate to $m(x)$ at almost all (μ) $x\in R^d$. For the distribution-free results concerning other regression estimates we refer to the paper of Stone [9].

The estimate examined in this article is a function defined over $R^d\times(R^d\times R)^n$, i.e., is of the following form: $\hat{m}(x, X_1, Y_1,..., X_n, Y_n)$, and for convenience is denoted by $\hat{m}(x)$. We examine its weak and strong pointwise consistency at almost all $(\mu) x\in R^d$. So, e.g., weak pointwise consistency at almost all (μ) $x\in R^d$, in brief weak consistency a.e. (μ), means that $\hat{m}(x)$ converges to $m(x)$ in probability at almost all (μ) $x\in R^d$, i.e., on a subset of R^d which μ measure equals 1.

Assuming $E|Y|<\infty$, we show that the estimate is weakly consistent a.e. (μ) if and only if

$$\lim_{n\to\infty}\frac{\sum_{i=1}^{n} h_i^d I_{\{h_i>\varepsilon\}}}{\sum_{i=1}^{n} h_i^d}=0 \qquad \text{all } \varepsilon>0 \tag{1}$$

and

$$\sum_{n=1}^{\infty} h_n^d=\infty. \tag{2}$$

Let us observe that if a sequence $\{h_n\}$ satisfying (1) has a limit, it equals zero. For $E|Y|^{1+\delta}<\infty$, $\delta>0$, (1) and (2) constitute a condition which is both necessary and sufficient for strong consistency a.e. (μ). Thus, for $E|Y|^{1+\delta} < \infty$, $\delta > 0$, weak and strong consistency are equivalent. In turn, for bounded Y, i.e., for $|Y| \leqslant \gamma < \infty$ almost surely, the estimate converges to $m(x)$ completely a.e. (μ) if and only if, in addition to (1),

$$\sum_{n=1}^{\infty} \exp\left(-\alpha \sum_{i=1}^{n} h_i^d\right) < \infty \qquad \text{all} \quad \alpha > 0. \tag{3}$$

2. Lemmas

Throughout the article norms are either all l_∞ or all l_2. By $S_h(x)$ we denote an open sphere with radius h centered at $x \in R^d$. For convenience, we denote

$$\sup_x K(x) = k. \tag{4}$$

We shall apply the following two lemmas:

Lemma 1. *Let kernel K satisfy the condition*

$$c_1 H(\|x\|) \leqslant K(x) \leqslant c_2 H(\|x\|),$$

$c_1 > 0$, *where H is a non-negative and non-increasing Borel function defined on the real half-line* $(0, \infty)$ *with* $0 < H(0+) < \infty$. *Let, moreover,*

$$\lim_{t \to \infty} t^d H(t) = 0.$$

Then

$$\lim_{h \to 0} \frac{E\{g(X)\, K((x-X)/h)\}}{EK((x-X)/h)} = g(x) \qquad a.e. \quad (\mu),$$

for any Borel function g such that $E|g(X)| < \infty$.

The lemma can be proved as Lemma 1 in Greblicki *et al.* [5]. The only difference is that one should use $H^+(\delta)$ instead of $H^{-1}(\delta)$, where $H^+(\delta)$ is the length of the interval $\{t \colon H(t) > \delta\}$.

Let us observe that conditions of Lemma 1 concerning the kernel imply

$$cI_{\{\|x\| < r\}} \leqslant K(x), \tag{5}$$

for some positive c and r.

LEMMA 2. *Under conditions of Lemma* 1 *and* (1),

$$\lim_{n\to\infty} \frac{\sum_{i=1}^n E\{g(X)\,K((x-X)/h_i)\}}{\sum_{i=1}^n EK((x-X)/h_i)} = g(x) \qquad a.e.\ (\mu).$$

The proof of the lemma is deferred to Appendix.

We shall also use Corollary 10.50 in Wheeden and Zygmund [10] which says that

$$a_h(x) = h^d/\mu(S_h(x)) \tag{6}$$

has a finite limit as h tends to zero a.e. (μ).

3. WEAK CONSISTENCY

In this section we prove

THEOREM 1. *Let* $E\,|Y| < \infty$. *Let* K *satisfy all the conditions of Lemma* 1. *If* (1) *and* (2) *hold, then*

$$\hat{m}(x) \to m(x) \text{ as } n\to\infty \text{ in probability a.e. } (\mu), \text{ for all } \mu \text{ and all } m. \tag{7}$$

Let, moreover,

$$\int K(x)\,dx < \infty. \tag{8}$$

If (7) *holds, then* (1) *and* (2) *are satisfied.*

Proof of Theorem 1. We first show that (1) and (2) imply (7). It is clear that $\hat{m}(x) = A_n(x)/B_n(x)$, where

$$A_n(x) = \sum_{i=1}^n Y_i K\left(\frac{x-X_i}{h_i}\right) \Bigg/ \sum_{j=1}^n EK\left(\frac{x-X_j}{h_j}\right) \tag{9}$$

and

$$B_n(x) = \sum_{i=1}^n K\left(\frac{x-X_i}{h_i}\right) \Bigg/ \sum_{j=1}^n EK\left(\frac{x-X_j}{h_j}\right).$$

By virtue of Lemma 2,

$$\lim_{n\to\infty} EA_n(x) = m(x) \qquad \text{a.e. } (\mu). \tag{10}$$

In turn, we shall now verify that (2) implies

$$\sum_{i=1}^{\infty} EK\left(\frac{x-X_i}{h_i}\right)=\infty \qquad \text{a.e. } (\mu). \tag{11}$$

Using (5) and (6), we get

$$\sum_{i=1}^{n} EK\left(\frac{x-X}{h_i}\right) \geqslant cr^d \sum_{i=1}^{n} (h_i^d/a_{rh_i}(x))\, I_{\{h_i<\varepsilon\}} + c\mu(S_{r\varepsilon}(x)) \sum_{i=1}^{n} I_{\{h_i \geqslant \varepsilon\}}, \tag{12}$$

all $\varepsilon>0$. On the other hand, by virtue of (6), for almost every $(\mu)\, x\in R^d$, there exist $\delta>0$ and $\varepsilon>0$ such that $a_h(x)<\delta$, for $0<h<\varepsilon$. Thus, for such δ and ε, the quantity in (12) is underbounded by

$$(cr^d/\delta) \sum_{i=1}^{n} h_i^d I_{\{h_i<\varepsilon\}} + c\mu(S_{r\varepsilon}(x)) \sum_{i=1}^{n} I_{\{h_i\geqslant\varepsilon\}},$$

which, by virtue of (2), increases to infinity as n tends to infinity. Thus, (11) holds.

In order to show that

$$A_n(x)-EA_n(x)\to 0 \qquad \text{as} \quad n\to\infty \text{ in probability a.e. } (\mu), \tag{13}$$

it suffices to use (11), Chebyshev's inequality, and classical truncation argument.

From (10) and (13) it follows that $A_n(x)\to m(x)$ as $n\to\infty$ in probability a.e. (μ). Since convergence of $B_n(x)$ to 1 can be verified in the same way, the first part of the theorem has been proved.

We shall now show that (7) implies (2) and (1). In order to prove the first implication we assume that $Y-m(X)$ is independent of X and bounded with variance 1. Let μ have a density f which is constant on a sphere $S_1(0)$ and zero outside. Let moreover, $m(x)\equiv 0$. Thus,

$$\begin{aligned}\operatorname{var} \hat{m}(x) &= E\left\{\sum_{i=1}^{n} K^2\left(\frac{x-X_i}{h_i}\right)\Big/\left[\sum_{j=1}^{n} K\left(\frac{x-X_j}{h_j}\right)\right]^2\right\}\\ &\geqslant E^2\left\{K\left(\frac{x-X_1}{h_1}\right)\Big/\sum_{j=1}^{n} K\left(\frac{x-X_j}{h_j}\right)\right\},\end{aligned}$$

which converges to zero as n tends to infinity a.e. (μ). By (4) and Jensen's inequality, the expectation in the above inequality is not smaller than

$$EK\left(\frac{x-X}{h_1}\right)\Big/\left[k+\sum_{i=2}^{n} EK\left(\frac{x-X}{h_i}\right)\right].$$

Hence (11) is satisfied. Now (2) can be verified by a contradiction. If (2) is not satisfied, $h_n \to 0$ as $n \to \infty$, In turn, one can easily verify that, for the assumed density,

$$\lim_{h \to 0} h^{-d} \int K\left(\frac{x-y}{h}\right) f(y)\, dy = f(x) \int K(y)\, dy, \tag{14}$$

for all $x \in S_1(0)$. In this way we obtained a contradiction since (11) cannot be satisfied. Therefore we have shown that (7) implies (2).

In order to complete the proof it suffices to prove that (7) and (2) imply (1). Verifying this implication is, however, more arduous. Let us assume that $Y = m(X)$, μ has a density which is constant on a sphere $S_a(0)$, $a > 0$, and zero outside and $m(x)$ equals 1 on a sphere $S_{a/2}(0)$ and 2 outside. We have

$$\hat{m}(x) - m(x) = \frac{\sum_{i=1}^{n} (m(X_i) - m(x))\, K((x - X_i)/h_i)}{\sum_{i=1}^{n} K((x - X_i)/h_i)} = \frac{g_n(x)}{f'_n(x)},$$

which converges to zero in probability as n tends to infinity a.e. (μ), where

$$g_n(x) = \sum_{i=1}^{n} (m(X_i) - m(x))\, K\left(\frac{x - X_i}{h_i}\right) \Big/ \sum_{j=1}^{n} EK\left(\frac{x - X}{h_j}\right)$$

and

$$f'_n(x) = \sum_{i=1}^{n} K\left(\frac{x - X_i}{h_i}\right) \Big/ \sum_{j=1}^{n} EK\left(\frac{x - X}{h_j}\right) \tag{15}$$

Clearly $Ef'_n(x) = 1$ and, by virtue of (11), $\operatorname{var} f'_n(x) \to 0$ as $n \to \infty$. One can verify that $\sup_n Eg_n^2(x) < \infty$. From this and a result in Loève [8, p. 166, Corollary 2], it follows that $Eg_n(x) = V_1(x) + V_2(x)$ converges to zero, where

$$V_1(x) = \frac{\sum_{i=1}^{n} \int [m(y) f(y) - m(x) f(x)]\, K((x - y)/h_i)\, dy}{\sum_{i=1}^{n} \int K((x - y)/h_i)\, f(y)\, dy}$$

and

$$V_2(x) = m(x) \left[\frac{f(x) \int K(y)\, dy}{f_n(x)} - 1\right]$$

with

$$f_n(x) = \sum_{i=1}^{n} \int K\left(\frac{x - y}{h_i}\right) f(y)\, dy \Big/ \sum_{j=1}^{n} h_j^d.$$

Since both $V_1(x)$ and $V_2(x)$ are non-negative, it follows that

$$\lim_{n\to\infty} f_n(x) = f(x) \int K(y)\, dy \qquad \text{a.e. } (\mu).$$

We shall now show that this convergence implies (1). It is clear that there exists δ such that

$$0 < \int_{\|y\| \leqslant \delta} K(y)\, dy < \int K(y)\, dy.$$

Moreover, let us observe that, for the assumed density and all $x \in S_{a/4}(0)$,

$$\sup_{h > 2a/\delta} h^{-d} \int K\left(\frac{x-y}{h}\right) f(y)\, dy \leqslant f(x) \int_{\|y\| \leqslant \delta} K(y)\, dy$$

and

$$\sup_{h \leqslant 2a/\delta} \int h^{-d} K\left(\frac{x-y}{h}\right) f(y)\, dy < f(x) \int K(y)\, dy.$$

Hence,

$$f_n(x) \leqslant f(x) \left[\int K(y)\, dy - \frac{\sum_{i=1}^n h_i^d I_{\{h_i > 2a/\delta\}}}{\sum_{i=1}^n h_i^d} \int_{\|y\| > \delta} K(y)\, dy\right],$$

for all $x \in S_{a/4}(0)$. The above inequality and convergence of $f_n(x)$ to $f(x) \int K(y)\, dy$, holding for all positive a, yield (1). The proof has been completed.

4. Strong Consistency

In this section we examine almost sure and complete convergence of the estimate to $m(x)$.

THEOREM 2. *Let $E|Y|^{1+\delta} < \infty$, $\delta > 0$. Let K satisfy all the conditions of Lemma* 1. *If* (1) *and* (2) *hold, then*

$$\hat{m}(x) \to m(x) \text{ as } n \to \infty \text{ almost surely a.e. } (\mu), \text{ for all } \mu \text{ and all } m. \quad (16)$$

Let moreover, (8) *be satisfied. If* (16) *holds, then* (1) *and* (2) *are satisfied.*

From Theorems 1 and 2 we obtain the corollary which establishes the equivalence of weak and strong consistency.

COROLLARY. *Let* $E|Y|^{1+\delta}<\infty, \delta>0$. *Let* K *satisfy appropriate conditions of Lemma* 1 *and* (8). *Then*

(a) *weak consistency* (7),

(b) *strong consistency* (16),

(c) *conditions* (1) *and* (2)

are equivalent.

Proof of Theorem 2. In view of Theorem 1 it suffices to show that (1) and (2) imply (16). Let $A_n(x)$ and $B_n(x)$ be as in the proof of Theorem 1. By virtue of theorem in Loève [8, p. 253], $A_n(x)$ converges to zero as n tends to infinity almost surely if

$$\sum_{n=1}^{\infty} \frac{E\{|Y|^{1+\delta} K^{1+\delta}((x-X)/h_n)\}}{[\sum_{i=1}^{n} EK((x-X)/h_i)]^{1+\delta}} < \infty. \tag{17}$$

In turn, by virtue of Lemma 1, for almost every $(\mu)\, x \in R^d$ and for every $\varepsilon>0$, there exists $\eta>0$ such that

$$E\left\{|Y|^{1+\delta} K\left(\frac{x-X}{h}\right)\right\} \Big/ EK\left(\frac{x-X}{h}\right)$$

is not greater than $E\{|Y|^{1+\delta}\,|X=x\}+\varepsilon$, for $0<h<\eta$. On the other hand, for $h \geqslant \eta$, the quantity is uniformly bounded for all positive h, a.e. (μ). Hence, (17) is satisfied, if

$$\sum_{n=1}^{\infty} \frac{a_n}{[\sum_{i=1}^{n} a_i]^{1+\delta}} < \infty,$$

where $a_n = EK((x-X)/h_n)$. This is, in turn, implied by (11); see, e.g., Barry [1, p. 470]. Thus, $A_n(x) - EA_n(x)$ converges to zero almost surely a.e. (μ). Since similar arguments can be applied to $B_n(x)$, the proof has been completed.

In the next theorem without proof we give conditions for complete convergence. Sufficient part of the theorem easily results from Bernstein's inequality, see Bennett [2], while necessity of conditions (1) and (3) from Lemma 2 and Kolmogorov exponential inequality, see Loève [8, p. 266].

THEOREM 3. *Let* $|Y| \leqslant \gamma < \infty$ *almost surely. Let* K *satisfy all the conditions of Lemma* 1. *If* (1) *and* (3) *hold, then*

$$\hat{m}(x) \to m(x) \text{ as } n \to \infty \text{ completely a.e. } (\mu) \text{ for all } \mu \text{ and all } m. \tag{18}$$

Let, moreover, (8) *be satisfied. If* (18) *holds, then* (1) *and* (3) *are satisfied.*

APPENDIX

Proof of Lemma 2. Clearly, for any $\varepsilon > 0$,

$$\frac{\sum_{i=1}^{n} E\{(g(X)-g(x))\, K((x-X)/h_i)\}}{\sum_{i=1}^{n} EK((x-X)/h_i)} = W_{1n}(x) + W_{2n}(x),$$

where

$$W_{1n}(x) = \frac{\sum_{i=1}^{n} E\{(g(X)-g(x))\, K((x-X)/h_i)\}\, I_{\{h_i > \varepsilon\}}}{\sum_{i=1}^{n} EK((x-X)/h_i)}$$

and

$$W_{2n}(x) = \frac{\sum_{i=1}^{n} E\{(g(X)-g(x))\, K((x-X)/h_i)\}\, I_{\{h_i \leqslant \varepsilon\}}}{\sum_{i=1}^{n} EK((x-X)/h_i)}.$$

Obviously,

$$|W_{1n}(x)| \leqslant (E\,|g(X)| + |g(x)|)\, k\alpha_n \beta_n(x),$$

where

$$\alpha_n = \sum_{i=1}^{n} I_{\{h_i > \varepsilon\}} \bigg/ \sum_{j=1}^{n} h_j^d,$$

which, by virtue of (1), converges to zero as n tends to infinity, all $\varepsilon > 0$, and

$$\beta_n(x) = \sum_{i=1}^{n} h_i^d \bigg/ \sum_{j=1}^{n} EK\left(\frac{x-X}{h_j}\right).$$

Using (5), we get

$$\beta_n(x) \leqslant \sum_{i=1}^{n} h_i^d \bigg/ cr^d \sum_{j=1}^{n} [h_j^d / a r_{h_j}(x)].$$

In turn, by virtue of (6), for almost every (μ) $x \in R^d$, there exist $\gamma > 0$ and ε_1 such that $1/a_h(x) > \gamma$ for $0 < h < \varepsilon_1$. Thus, for $0 < \varepsilon < \varepsilon_1$,

$$\beta_n(x) < \sum_{i=1}^{n} h_i^d \bigg/ \left[r^d c\gamma \sum_{j=1}^{n} h_j^d I_{\{h_j \leqslant \varepsilon\}} + c\mu(S_{r\varepsilon}(x)) \sum_{j=1}^{n} I_{\{h_j > \varepsilon\}} \right]$$

a.e. (μ), which, by virtue of (1), approaches $1/c\gamma$ as n tends to infinity. Finally, for $0 < \varepsilon < \varepsilon_1$,

$$W_{1n}(x) \to 0 \quad \text{as} \quad n \to \infty \quad \text{a.e. } (\mu).$$

On the other hand, by virtue of Lemma 1, for almost every $(\mu)\, x \in R^d$ and every $\delta > 0$ there exists ε_2 such that

$$\left| \frac{E\{(g(X)-g(x))\, K((x-X)/h)\}}{EK((x-X)/h)} \right| < \delta,$$

for $0 < h < \varepsilon_2$. Hence,

$$|W_{2n}(x)| < \delta,$$

for $0 < h < \varepsilon_2$. Since δ can be arbitrarily small,

$$W_{2n}(x) \to 0 \quad \text{as} \quad n \to \infty \quad \text{a.e.} \quad (\mu).$$

The lemma has been proved.

Acknowledgment

The authors wish to express their thanks to the referee for his comments.

References

[1] Barry, N. K. (1964). *A Treatise on Trigonometric Series*, Vol. 2. Pergamon, Elmsford, NY.

[2] Bennett, G. (1962). Probability inequalities for sum of independent random variables. *J. Amer. Statist. Assoc.* **57** 33–45.

[3] Devroye, L. (1981). On the almost everywhere convergence of nonparametric regression function estimates. *Ann. Statist.* **9** 1310–1319.

[4] Devroye, L., and Wagner, T. J. (1980). On the L_1 convergence of kernel estimators of regression function with application in discrimination. *Z. Wahrsch. Verw. Gebiete* **51** 15–25.

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

[6] Krzyżak, A., and Pawlak, M. (1984). Distribution-free consistency of nonparametric kernel regression estimate and classification. *IEEE Trans. Inform. Theory* **30** 78–81.

[7] Krzyżak, A., and Pawlak, M. (1984). Almost everywhere convergence of recursive regression function estimate and classification. *IEEE Trans. Inform. Theory* **30** 91–93.

[8] Loève, M. (1977). *Probability Theory I*. Springer-Verlag, Berlin.

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

[10] Wheeden, R. L., and Zygmund, A. (1977), *Measure and Integral*. Dekker, New York.

Printed by Catherine Press, Ltd., Tempelhof 41, B-8000 Brugge, Belgium