The Annals of Statistics
1984, Vol. 12, No. 4, 1570–1575

DISTRIBUTION-FREE POINTWISE CONSISTENCY OF KERNEL REGRESSION ESTIMATE

BY WŁODZIMIERZ GREBLICKI, ADAM KRZYŻAK AND MIROSŁAW PAWLAK

Technical University of Wrocław

An estimate $\sum_{i=1}^n Y_i K((x - X_i)/h)/\sum_{j=1}^n K((x - X_j)/h)$, calculated from a sequence $(X_1, Y_1), \cdots, (X_n, Y_n)$ of independent pairs of random variables distributed as a pair (X, Y), converges to the regression $E\{Y \mid X = x\}$ as n tends to infinity in probability for almost all (μ) $x \in R^d$, provided that $E \mid Y \mid < \infty$, $h \to 0$ and $nh^d \to \infty$ as $n \to \infty$. The result is true for all distributions μ of X. If, moreover, $\mid Y \mid \le \gamma < \infty$ and $nh^d/\log n \to \infty$ as $n \to \infty$, a complete convergence holds. The class of applicable kernels includes those having unbounded support.

1. Introduction. We estimate $m(x) = E\{Y \mid X = x\}$ from a sequence $(X_1, Y_1), \cdots, (X_n, Y_n)$ of independent observations of a pair (X, Y) of random variables. X and Y take their values in R^d and R, respectively. Throughout the paper we do not impose any restrictions on the probability distribution μ of X. Hence, all the results presented here are distribution-free in the sense that they are true for all μ. The estimate is of the following form:

$$m_n(x) = \sum_{i=1}^n Y_i K((x - X_i)/h)/\sum_{j=1}^n K((x - X_j)/h),$$

where h depends on n and K is a Borel kernel. In the above definition and in the paper 0/0 is treated as 0.

Assuming that

$$h(n) \to 0 \quad \text{as} \quad n \to \infty, \tag{1}$$

$$nh^d(n) \to \infty \quad \text{as} \quad n \to \infty, \tag{2}$$

and

$$c_1 H(\| x \|) \le K(x) \le c_2 H(\| x \|), \tag{3}$$

c_1, c_2 being positive, Devroye [1] has shown that $E \mid m_n(x) - m(x) \mid^p$, $p \ge 1$, converges to zero as n tends to infinity for almost all (μ) $x \in R^d$, whenever $E \mid Y \mid^p < \infty$. H is a function defined over the nonnegative half real line. In [1] it equals 1 for $\| x \| \le r$, r positive, and 0 otherwise. Let us observe that the class of kernels satisfying the above requirement is practically confined to the window kernel i.e. the kernel which equals 1 for $\| x \| \le 1$ and 0 otherwise.

We study the weak and complete convergence on $m_n(x)$ to $m(x)$ for almost all (μ) $x \in R^d$, and we get as a simple consequence some results concerning the convergence of $\int \mid m_n(x) - m(x) \mid \mu(dx)$ to zero. We show that it is possible to apply kernels with unbounded support and even not integrable ones.

Received February 1983; revised April 1984.
AMS 1980 *subject classifications.* Primary 62G05.
Key words and phrases. Nonlinear regression, kernel estimate, universal consistency.

We assume that

(4) $$cI_{\{\|x\|\le r\}}(x) \le K(x),$$

c and r positive. Moreover, the kernel satisfies (3). H is a bounded decreasing Borel function and

(5) $$t^d H(t) \to 0 \quad \text{as} \quad t \to \infty.$$

As far as the convergence in probability is concerned, we impose on the sequence $\{h(n)\}$ conditions (1) and (2), while the complete convergence is achieved under an additional restriction

(6) $$\textstyle\sum_{n=1}^{\infty} \exp(-\alpha n h^d(n)) < \infty,$$

for all positive α. Condition (6) is satisfied if

(7) $$nh^d(n)/\log n \to \infty \quad \text{as} \quad n \to \infty.$$

In the paper the norms are either all l_2 or all l_∞.

2. Preliminaries and lemmas. The crucial point of this paper is the asymptotic behaviour of the following expression:

$$U_h(x) = \int K\left(\frac{x-y}{h}\right) f(y)\mu(dy) \bigg/ \int K\left(\frac{x-y}{h}\right)\mu(dy)$$

as h tends to zero, where f is a μ integrable function. In Wheeden and Zygmund [8] we find $U_h(x) \to f(x)$ as $h \to 0$ for almost all (μ) $x \in R^d$, provided that K is the window kernel. In the next lemma we extend the class of applicable kernels.

LEMMA 1. *Let a nonnegative Borel kernel K satisfy* (3) *and* (4). *Let a bounded Borel function H be decreasing in the interval* $[0, \infty)$ *and satisfy* (5). *Let f be μ integrable. Then*

$$U_h(x) \to f(x)$$

as $h \to 0$ for almost all (μ) $x \in R^d$.

In the proof of Lemma 1 as well as in the sequel, we shall need the following result due to Devroye [1]:

LEMMA 2. *For almost all (μ) $x \in R^d$,*

$$a_h(x) = h^d/\mu(S_h)$$

has a finite limit as h tends to zero.

In Lemma 2 and throughout the paper S_r is a sphere of the radius r centered at x, $x \in R^d$.

PROOF OF LEMMA 1. Clearly,

$$| U_h(x) - f(x) | \le \frac{c_2}{c_1} \int H\left(\frac{\| x - y \|}{h}\right) | f(x) - f(y) | \mu(dy) \Big/ \int H\left(\frac{\| x - y \|}{h}\right) \mu(dy).$$

Let us observe

$$H(t) = \int_0^\infty I_{\{H(t)>s\}}(s)\, ds.$$

Thus,

$$\int H\left(\frac{\| x - y \|}{h}\right) \mu(dy) = \int_0^\infty \mu(A_{t,h})\, dt \tag{8}$$

and

$$\int H\left(\frac{\| x - y \|}{h}\right) | f(x) - f(y) | \mu(dy) = \int_0^\infty \left[\int_{A_{t,h}} | f(x) - f(y) | \mu(dy)\right] dt, \tag{9}$$

where $A_{t,h} = \{y\colon H(\| x - y \| / h) > t\}$.

Let $\delta = \varepsilon h^d$, $\varepsilon > 0$. Obviously,

$$\begin{aligned} &\int_\delta^\infty \left[\int_{A_{t,h}} | f(x) - f(y) | \mu(dy)\right] dt \Big/ \int_0^\infty \mu(A_{t,h})\, dt \\ &\qquad \le \sup_{t\ge\delta}\left[\int_{A_{t,h}} | f(x) - f(y) | \mu(dy) \Big/ \mu(A_{t,h})\right]. \end{aligned} \tag{10}$$

It is clear that the radii of sets $A_{t,h}$, $t \ge \delta$, are not greater than the radius of the set $A_{\delta,h}$. The radius of $A_{\delta,h}$ is in turn h times greater than that of the set $A_{\delta,1}$. We shall now estimate the radius of $A_{\delta,1}$. It does not exceed $H^{-1}(\delta)$, H^{-1} being the inverse of H. Thus the radius of $A_{\delta,h}$ is majorized by $hH^{-1}(\delta)$. Now, by virtue of (5) and by the definition of δ, $hH^{-1}(\delta) = hH^{-1}(\varepsilon h^d)$ converges to zero as h tends to zero. Since $A_{t,h}$ is either a cube or a ball, then by Wheeden and Zygmund [8, page 189], the quantity in (10) converges to zero as h tends to zero for almost all (μ) $x \in R^d$.

On the other hand,

$$\int_0^\delta \left[\int_{A_{t,h}} | f(x) - f(y) | \mu(dy)\right] dt \le (c_3 + | f(x) |)\delta, \tag{11}$$

where $c_3 = \int | f(x) | \mu(dx)$. Using (4), we get

$$\int H\left(\frac{\| x - y \|}{h}\right) \mu(dy) \ge c\mu(S_{rh}) \ge \frac{c(rh)^d}{a_{rh}(x)}, \tag{12}$$

where $a_{rh}(x)$ is as in Lemma 2. From (11), (12), and by the definition of δ, we have in turn

$$\int_0^{\delta}\left[\int_{A_{t,h}} |f(x) - f(y)|\,\mu(dy)\right] dt \Big/ \int_0^{\infty} \mu(A_{t,h})\, dt \le \varepsilon\left[\frac{c_3 + |f(x)|}{cr^d}\right] a_{rh}(x).$$

Finally, using Lemma 2, the above quantity can be made arbitrarily small for almost all (μ) $x \in R^d$ when ε is small enough. The proof has been completed.

3. Consistency.

We are now in a position to show:

THEOREM 1. *Let $E|Y| < \infty$. Let K and H satisfy the conditions of Lemma* 1. *Let* (1) *and* (2) *hold. Then*

$$m_n(x) \to m(x) \quad \text{as} \quad n \to \infty \quad \text{in probability}$$

for almost all (μ) $x \in R^d$.

PROOF. Let us denote

$$A_n = E\{YK((x - X)/h)\}/EK((x - X)/h),$$

$$B_{1n} = n^{-1} \textstyle\sum_{i=1}^n (V_{in} - EV_{in}),$$

where

$$V_{in} = Y_i K((x - X_i)/h)/EK((x - X)/h).$$

Let, moreover,

$$B_{2n} = n^{-1} \textstyle\sum_{i=1}^n (Z_{in} - EZ_{in}),$$

where

$$Z_{in} = K((x - X_i)/h)/EK((x - X)/h).$$

Now, the estimate can be rewritten in the following form:

$$m_n(x) = (A_n + B_{1n})/(1 + B_{2n}). \tag{13}$$

Since, by Lemma 1 and (1), $A_n \to m(x)$ as $n \to \infty$ for almost all (μ) $x \in R^d$, it suffices to verify that both B_{1n} and B_{2n} converge to zero in probability as n tends to infinity for almost all (μ) $x \in R^d$.

Let us take B_{1n} into account. For $N > 0$, let $Y' = YI_{\{|Y| \le N\}}$, and $Y'' = Y - Y'$. Let, moreover, $g_N(x) = E\{|Y''| \mid X = x\}$. Let B_{1n}' and B_{1n}'' be the expressions obtained from B_{1n} by replacing Y_i with Y_i' and Y_i'', respectively. Now, it suffices to verify that both B_{1n}' and B_{1n}'' converge to zero in probability as n tends to infinity for almost all (μ) $x \in R^d$. From Chebyshev's inequality and (4), we have

$$\begin{aligned} P\{|B_{1n}'| > t\} &\le (nt^2)^{-1} kN g_{N,h}(x)/EK((x - X)/h) \\ &\le kN g_{N,h}(x) a_{rh}(x)/t^2 cr^d nh^d, \end{aligned}$$

where $k = \sup_x K(x)$ and $g_{N,h}(x) = E\{|Y'|\, K((x - X)/h)\}/EK((x - X)/h)$. By virtue of Lemma 1 and (1), $g_{N,h}(x) \to E\{|Y'| \mid X = x\}$ as $n \to \infty$ for almost all (μ) $x \in R_d$. By this, Lemmas 1 and 2, and from (2), for each fixed N, the above expression converges to zero as n tends to infinity for almost all (μ) $x \in R^d$. Then we apply Markov's inequality and get

$$P\{|B''_{1n}| > t\} \le 2t^{-1}E\{g_N(X)K((x - X)/h)\}/EK((x - X)/h).$$

By virtue of Lemma 1, the last expression converges to $g_N(x)$ as n tends to infinity for almost all (μ) $x \in R^d$. Since $E|Y| < \infty$, $Eg_N(X)$ converges to zero as N tends to infinity. Since, moreover, g_N is monotone in N, by the Lebesgue monotone convergence theorem $g_N(x)$ converges to zero as N tends to infinity for almost all (μ) $x \in R^d$. Thus, let us first choose N large enough so that $g_N(x)$ is small, and then let n grow large.

As the convergence of B_{2n} can be verified in the same way, the proof has been completed.

In the next theorem we show a complete convergence.

THEOREM 2. *Let* $|Y| \le \gamma < \infty$. *Let K and H satisfy the conditions of Lemma* 1. *Let* (1) *and* (6) *hold. Then*

$$m_n(x) \to m(x) \quad \text{as} \quad n \to \infty \quad \text{completely}$$

for almost all (μ) $x \in R^d$.

Devroye's result [1] says that the assertion of Theorem 2 holds, provided that H is the window kernel and (7) is satisfied.

PROOF OF THEOREM 2. Clearly, it suffices to show that B_{1n} and B_{2n} in (13) converge to zero completely as n tends to infinity for almost all (μ) $x \in R^d$. Taking into account

$$|V_{in}| \le \gamma k a_{rh}(x)/cr^d h^d,$$

and the fact the variance of V_{in} is bounded by $\gamma^2 k a_{rh}(x)/cr^d h^d$, the application of Bernstein's inequality, see e.g. Hoeffding [5], yields

$$P\{|B_{1n}| > t\} \le 2\exp(-cr^d t^2 nh^d/2\gamma k a_{rh}(x)(\gamma + t)).$$

This, Lemma 1 and (6) yield convergence of B_{1n}.

Since the convergence of B_{2n} can be verified by using similar arguments, the proof has been completed.

4. Conclusion. The class of applicable kernels includes those having unbounded support and the following ones, in particular: $e^{-|x|}$, e^{-x^2}, $1/(1 + |x|^{1+\delta})$, $\delta > 0$, and

$$K(x) = \begin{cases} 1/e & \text{for } |x| \le e \\ 1/|x| \ln |x| & \text{otherwise.} \end{cases}$$

The last kernel is even not integrable.

By virtue of the Lebesgue dominated convergence theorem on product spaces, see Glick [3], we have:

COROLLARY. *Let $|Y| \leq \gamma < \infty$. Then, with the conditions of Theorem* 1 *or* 2,

$$\int |m_n(x) - m(x)| \mu(dx) \to 0 \quad as \quad n \to \infty \tag{17}$$

in the mean or almost surely, respectively.

The convergence in the mean of the integrated absolute error in (17) has been studied by Spiegelman and Sacks [6] as well as by Devroye and Wagner [2]. These authors, however, assumed that $E|Y| < \infty$, but considered only kernels with bounded support.

Finally, we would like to mention that distribution-free results concerning regression estimation were first obtained by Stone [7]. For a review paper we refer to Györfi [4].

Acknowledgment. The authors wish to express their thanks to the referee for his suggestions concerning some improvements.

REFERENCES

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

[2] DEVROYE, L. and WAGNER, T. J. (1980). Distribution-free consistency results in nonparametric discrimination and regression function estimation. *Ann. Statist.* **8** 231–239.

[3] GLICK, N. (1974). Consistency conditions for probability estimators and integrals for density estimators. *Utilitas Math.* **5** 61–74.

[4] GYÖRFI, L. (1981). Recent results on nonparametric regression estimate and multiple classification. *Problems Control Inform. Theory* **10** 43–52.

[5] HOEFFDING, W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58** 13–30.

[6] SPIEGELMAN, C. and SACKS, J. (1980). Consistent window estimation in nonparametric regression. *Ann. Statist.* **8** 240–246.

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

[8] WHEEDEN, R. L. and ZYGMUND, A. (1977). *Measure and Integral.* Dekker, New York.

INSTITUTE OF ENGINEERING CYBERNETICS
TECHNICAL UNIVERSITY OF WROCŁAW
50-370 WROCŁAW, POLAND