The Canadian Journal of Statistics
Vol.6 (1978), No.2, 179-191.
La Revue Canadienne de Statistique

THE UNIFORM CONVERGENCE OF THE NADARAYA-WATSON REGRESSION FUNCTION ESTIMATE

by

Luc P. Devroye*
McGill University

Key words and phrases: Uniform consistency, regression function estimators, kernel estimates, strong uniform convergence. *AMS 1970 subject classifications:* Primary 62G05; secondary 60F15.

ABSTRACT

If $(X_1,Y_1), \ldots, (X_n,Y_n)$ is a sequence of independent identically distributed $\mathbb{R}^d \times \mathbb{R}$-valued random vectors then Nadaraya (1964) and Watson (1964) proposed to estimate the regression function $m(x) = \mathcal{E}\{Y_1 | X_1 = x\}$ by

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

where K is a known density and $\{h_n\}$ is a sequence of positive numbers satisfying certain properties. In this paper a variety of conditions are given for the strong convergence to 0 of $\operatorname{ess}_X \sup |m_n(X)-m(X)|$ (here X is independent of the data and distributed as X_1). The theorems are valid for all distributions of X_1 and for all sequences $\{h_n\}$ satisfying $h_n \to 0$ and $nh_n^d/\log n \to \infty$.

1. INTRODUCTION

In regression function estimation we are concerned with approximations of the regression function $m(x) = \mathcal{E}\{Y|X = x\}$ that are constructed from a sample of independent identically distributed random vectors $(X_1,Y_1),\ldots,$ (X_n,Y_n) with the same distribution as the $\mathbb{R}^d \times \mathbb{R}$-valued random vector (X,Y). In this paper some new properties of the kernel estimate

* Part of this work was supported by USAF Grant 77-3385 while the author was at the University of Texas, Austin.

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

proposed by Nadaraya (1964) and Watson (1964), are highlighted. In (1) K is a nonnegative integrable function and $\{h_n\}$ is a sequence of positive numbers satisfying

$$h_n \to 0 \tag{2}$$

and

$$nh_n^d \to \infty \ . \tag{3}$$

The main theme pursued in this paper is the strong uniform convergence of m_n to m, that is, when does

$$\operatorname{ess}_A \sup |m_n(X)-m(X)| \to 0 \ a.s. \ ? \tag{4}$$

Throughout this paper, A is a fixed subset of $\mathbb{R}^d$ and the essential supremum is with respect to the restriction of the distribution of X to A. We are particularly interested in finding weak conditions on the distribution of (X,Y) that imply (4). We do not want to assume that X or Y have densities. Results regarding (4) that are based on the existence of these densities can be found in Nadaraya (1964, 1965) and Greblicki (1974) for the case $d=1$. More recently, Nadaraya (1970) showed that (4) holds for $d=1$ if $A=[a,b]$ and (2), (3) are satisfied, and if the following is true:

$$X \text{ has a continuous density } f \ , \tag{5}$$

$$\inf_A f(x) > 0 \ , \tag{6}$$

$$m \text{ is continuous on } \mathbb{R} \tag{7}$$

$$|Y| \le c < \infty \ a.s., \tag{8}$$

$$K \text{ is a bounded density on } \mathbb{R} \text{ satisfying } |x|K(x) \to 0 \text{ as } |x| \to \infty, \tag{9}$$

$$K \text{ is of bounded variation }, \tag{10}$$

$$\sum_{n=1}^{\infty} \exp(-\alpha n\, h_n^2) < \infty \text{ for all } \alpha > 0 \ . \tag{11}$$

The main result of this paper is that the conditions (5-6), (8), (9), and (11) can be considerably weakened and that (10) can be omitted in order for (4) to hold. Some of the techniques we develop can also be used to obtain pointwise consistency results, cf. Nadaraya (1964), Watson (1964), Schuster (1972), Greblicki (1974), Noda (1976) and Devroye and Wagner (1978).

2. MAIN RESULTS

We will treat the simplest cases first and gradually obtain more general results. Throughout it is assumed that K satisfies

CONDITION 1. *(i)* $0 \le K(x) \le K^* $ *for some* $K^* < \infty$,

(ii) $K(x) = L(\| x \|)$ *for some nonincreasing function* L; $\| \cdot \|$ *is one of the standard* ℓ_p *norms on* $\mathcal{R}^d$ $(p \ge 1;$ *p integer or* $p = \infty)$,

(iii) $u^d L(u) \to 0$ *as* $u \to \infty$,

(iv) $L(u^*) > 0$ *for some* $u^* > 0$.

The restriction *(ii)* that K be of the radial type is not unreasonable; what is interesting is that K can be non-integrable. For example, for d=1,

$$L(u) = 1/(1+u \log(1+u))$$

is not integrable but satisfies Condition 1.

No continuity restrictions are put on K. Some of the results presented below can be generalized towards slightly broader classes of functions K but the cost of doing so outweighs the advantages.

The probability measure of X is denoted by μ. The only restriction put on μ is the following:

CONDITION 2. *There exist* a, $b > 0$ *such that*

$$\inf_A \mu(S(x,r)) \ge a r^d, \quad \text{all} \quad r \in [0,b] \ ,$$

where $S(x,r)$ *is the closed sphere (under the* ℓ_p *norm) with centre* x *and radius* r.

Condition 2 insures that each neighbourhood of every x in A has sufficient probability. It imposes a restriction on the shape of A and on μ . If A is a sphere and μ has density f, then Condition 2 holds iff

$$\operatorname*{ess\,inf}_A f(x) > 0 \ ,$$

cf. (6). Condition 2 also implies that A must be contained in $\mathrm{supp}(\mu)$, the support of μ, that is, the set of all x such that $\mu(S(x,r)) > 0$ for all $r > 0$.

The existence of m follows if $\mathcal{E}\{|Y|\} < \infty$, and this will always be assumed. Of course, m is only defined up to μ - null sets.

The conditions of convergence depend heavily on the joint distribution of X and Y. We say that the observations (the Y_i) are

deterministic if $Y = m(X)$ *a.s.*,

uniformly bounded if $|Y-m(X)| \le c$ *a.s.* for some $c < \infty$, *uniformly generalized Gaussian* if

$$\operatorname{ess}_\mu \sup \mathcal{E}\{e^{\lambda(Y-m(X))}|X\} \le e^{\sigma^2\lambda^2/2(1-|\lambda|c)}, \quad \text{all} \quad |\lambda| \le 1/c ,$$

for some $\sigma \ge 0$ and $c \ge 0$,

uniformly L_q ($q > 0$ is fixed) if

$$\operatorname{ess}_\mu \sup \mathcal{E}\{|Y-m(X)|^q|X\} \le c \quad \text{for some} \quad c < \infty .$$

If the observations are uniformly bounded or uniformly Gaussian (i.e., given X, Y is Gaussian with variance $\sigma^2(X)$, and $\operatorname{ess\,sup} \sigma^2(X) < \infty$), then they are uniformly generalized Gaussian which in turn implies that they are uniformly L_q for all $q > 0$.

LEMMA 1. The deterministic case. *If the observations are deterministic, if Conditions 1 and 2 hold, if* A *is a compact set, if some version of* m *is bounded and continuous on* $\operatorname{supp}(\mu)$ *and if (2) holds and*

$$nh_n^d/\log n \to \infty , \tag{12}$$

then (4) follows.

We notice here that Lemma 1 remains true if Condition 1 (i) is dropped (the boundedness of K). The fact that K need not even be integrable seems to indicate that the shortest route to the study of (1) is not via Parzen-density estimation where the integrability of K is required.

THEOREM 1. *Let the observations be uniformly generalized Gaussian, let conditions 1, 2, (2) and (12) hold, let* A *be a compact set and let some version of* m *be bounded and continuous on* $\operatorname{supp}(\mu)$. *If* K *takes only finitely many* $(k+1)$ *values, then (4) obtains. If the observations are just uniformly in* L_q *for some* $q \ge 2$, *and if*

$$nh_n^d/n^{t(p,d)k/(q-1)} \to \infty , \tag{13}$$

then

$$\operatorname{ess}\sup_A |m_n(X)-m(X)| \to 0 \quad \text{in probability.} \tag{14}$$

If furthermore

$$\sum_{n=1}^{\infty} n^{t(p,d)k}/(n\, h_n^d)^{q-1} < \infty , \tag{15}$$

then (4) is true.

Notice that Theorem 1 applies to the kernel K that is uniform on the unit sphere. While condition (12) is considerably weaker than Nadaraya's condition (11) on the sequence $\{h_n\}$, the generalization of Theorem 1 towards all radial kernels satisfying Condition 1 (not just the ones taking finitely many values) is only possible if we also require

$$nh_n^{2d}/\log n \to \infty . \tag{16}$$

Lemma 2 below deals with a geometrical-probability-theoretical inequality which will be used in the proofs of Theorems 1 and 2.

LEMMA 2. *Let* $s(x_1,\dots,x_n)$ *be the number of different subsets of* $\{x_1, \dots, x_n\}$ *that can be obtained through intersections of* $\{x_1, x_2, \dots, x_n\}$ *with spheres (under the* ℓ_p*-norm) from* $\mathcal{R}^d$. *Then*

$$\max_{(x_1,\dots,x_n)\in\mathcal{R}^{dn}} s(x_1,\dots,x_n) \le (2n)^{t(p,d)} , \tag{17}$$

where

$$t(p,d) = \begin{cases} 1+d(p-1) & p\ge 1,\ p \text{ even}, \\ 2^d\,(1+d(p-1)), & p\ge 1,\ p \text{ odd}, \\ 2d, & p=\infty . \end{cases}$$

THEOREM 2. *Let the observations be uniformly generalized Gaussian, let conditions 1, 2, (2) and (16) hold, let* A *be a compact set and let some version of* m *be bounded and continuous on* $\operatorname{supp}(\mu)$. *Then (4) is true.*

Note. A quick inspection of the proofs reveals that under the conditions of strong consistency stated in Theorems 1 and 2, we have complete convergence as well, that is

$$\sum_{n=1}^{\infty} P\{\operatorname{ess}_A\sup|m_n(X)-m(X)| > \varepsilon\} < \infty \quad \text{for all} \quad \varepsilon > 0 .$$

3. PROOFS

Proof of Lemma 1. Since $Y_1 = m(X_1),\dots, Y_n = m(X_n)$ *a.s.*, it follows that

$$\left.\begin{aligned} &\operatorname{ess}_A\sup|m_n(x)-m(x)| \\ &\quad\le \operatorname{ess}\sup_A \sum_{i=1}^{n} |m(X_i)-m(x)|K((X_i-x)/h_n)I_S((X_i-x)/h_n)/g_n(x) \\ &\quad+ \operatorname{ess}\sup_A \sum_{i=1}^{n} |m(X_i)-m(x)|K((X_i-x)/h_n)I_{S^c}((X_i-x)/h_n)/g_n(x), \end{aligned}\right\} \tag{18}$$

where

$$g_n(x) = \sum_{i=1}^{n} K((x-X_i)/h_n)$$

and S is an arbitrary set from $\mathcal{R}^d$. The first term on the right-hand side of (18) is not greater than

$$\sup_{\substack{\|y-x\|\le r \\ x\in A}} |m(y)-m(x)|$$

if we take for S the closed sphere centered at 0 with radius r/h_n. With $c_m = \sup_{\mathrm{supp}(\mu)} |m(x)|$, we notice that the last term in (18) can be estimated by

$$2c_m L^*(r/h_n) nh_n^d / r^d \inf_A g_n(x) \ , \tag{19}$$

where $L^*(u) \to 0$ as $u \to \infty$. For r so small that $\sup_{\substack{\|y-x\| \le r \\ x \in A}} |m(y)-m(x)| < \epsilon/2$, we have,

$$P\{\operatorname{ess\,sup}_A |m_n(x)-m(x)| > \epsilon\} \le P\{\inf_A g_n(x) < \gamma nh_n^d\} \ , \tag{20}$$

where $\gamma = 4c_m c_L / r^d \epsilon$, and $c_L = L^*(r/h_n) \to 0$. In the remainder of the proof we show that for sufficiently small $\gamma > 0$ there exist constants k_1, k_2 and k_3, all positive, such that

$$P\{\inf_A g_n(x) < \gamma nh_n^d\} \ \le \ k_1 h_n^{-d} \exp(-k_2 nh_n^d), \quad \text{if } h_n < k_3 \tag{21}$$

Lemma 1 will then follow by the Borel-Cantelli lemma.

If μ_n denotes the empirical measure for $X_1, X_2, \ldots, X_n$, then it follows from Condition 1 that

$$P\{\inf_A g_n(x) < \gamma nh_n^d\} \ \le \ P\{\inf_A \mu_n(S(x,u^*h_n)) < \gamma h_n^d / L(u^*)\} \ . \tag{22}$$

Now, one can find $N = N_1 + N_2/h_n^d$ points $y_1, y_2, \ldots, y_N$ in the cube containing A such that spheres $S(y_i, u^*h_n/4)$ cover A. Since for each $x \in A$, there exists an i with $S(y_i, u^*h_n/2) \subseteq S(x, u^*h_n)$, and since for each y_i we can find a y_i^* in A such that $S(y_i^*, u^*h_n/4) \subseteq S(y_i, u^*h_n/2)$, we can write

$$\left.\begin{aligned} & P\{\inf_A \mu_n(S(x,u^*h_n)) < \gamma h_n^d/L(u^*)\} \\ & \le \sum_{i=1}^N P\{\mu_n(S(y_i,u^*h_n/2)) < \gamma h_n^d/L(u^*)\} \le \sum_{i=1}^N P\{\mu_n(S(y_i^*,u^*h_n/4)) < \gamma h_n^d/L(u^*)\} \\ & \le N \sup_A P\{\mu_n(S(x,u^*h_n/4)) - \mu(S(x,u^*h_n/4)) < \gamma h_n^d/L(u^*) - \mu(S(x,u^*h_n/4))\} \\ & \le N \sup_A P\{\mu_n(S(x,u^*h_n/4)) - \mu(S(x,u^*h_n/4)) < -\mu(S(x,u^*h_n/4))/2\} \end{aligned}\right\} \tag{23}$$

provided that $\inf_A \mu(S(x,u^*h_n/4)) \ge 2\gamma h_n^d / L(u^*)$. Condition 2 insures that we can find a small constant $\gamma > 0$ such that this is true whenever $\frac{1}{4}u^*h_n < b$. It suffices to pick $\gamma = L(u^*)au^{*d}/(2\times 4^d)$. For any set D, Bennett's inequality (Bennett, 1962; Hoeffding, 1963) implies that

$P\{\mu_n(D)-\mu(D) < -\mu(D)/2\} \le \exp(-n\mu(D)/2)^2/(2\mu(D)+\mu(D)/2) = \exp(-n\mu(D)/10)$.

Thus, (23) can be bounded above by $N\exp[-n\inf_A \mu(S(x,u^*h_n/4))/10]$ which by our choice of γ is smaller than $N\exp(-nh_n^d\gamma/5L(u^*))$ when $h_n < 4b/u^*$. Q.E.D.

Proof of Lemma 2. If only intersections with linear halfspaces from $\mathcal{R}^d$ are considered, then (17) is true with $t(p,d)=d$ in view of a result of Cover (1965) (see also Vapnik and Chervonenkis, 1971) who showed that $s(x_1,\ldots,x_n) \le 1+n^d$ for all $x_1\in\mathcal{R}^d,\ldots,x_n\in\mathcal{R}^d$. Now, a sphere under an ℓ_p norm ($p<\infty$) can be viewed as the set of $(y_1^1,\ldots,y^d)\in\mathcal{R}^d$ with $\Sigma a_i|y^i-y_0^i|^p \le r^p$, where $a_1,\ldots,a_d$ are scaling constants, $(y_0^1,\ldots,y_0^d)$ is the centre of the sphere and r is the radius. Equivalently, it is the intersection of all possible 2^d sets defined by

$$\sum_{i=1}^d a_i\lambda_i(y^i-y_0^i)^p \le r^p,\ (\lambda_1,\ldots,\lambda_d)\in\{-1,+1\}^d .$$

(If p is even, just consider the set corresponding to $\lambda_1=\ldots=\lambda_d=1$.) This sum can be regarded as a linear combination of a constant, all d functions y^i, all d functions $(y^i)^2,\ldots,$ all d functions $(y^i)^{p-1}$, and the function $a_1\lambda_1(y^1)^p+\ldots+a_d\lambda_d(y^d)^p$, that is, a linear halfspace in a Euclidean space with dimension $1+d(p-1)$. This covers (17) for the case p even. For p odd, we obtain by an obvious argument $t(p,d)=2^d(1+d(p+1))$. For $p=\infty$, a sphere is an intersection of $2d$ cylinder sets, each one of them generated by an interval on the real line. Hence we deduce: $s(x_1,\ldots,x_n)\le(1+n)^{2d}\le(2n)^{2d}$. Q.E.D.

LEMMA 3. *Assume that $Z_1,\ldots,Z_n$ are independent zero mean random variables satisfying*

$$\sup_i \mathcal{E}\{|Z_i|^q\} \le c<\infty \quad (\textit{where } q\ge 2) . \tag{24}$$

Then

$$P\{|n^{-1}\sum_{i=1}^n z_iZ_i| \ge \varepsilon\} \le k_1\sum_{i=1}^n z_i^q/(n\varepsilon)^q + 2\exp(-k_2n^2\varepsilon^2/\sum_{i=1}^n z_i^2) ,$$

where $\varepsilon>0$, and $z_1,\ldots,z_n$ is a sequence of nonnegative constants. The constants k_1 and k_2 only depend upon c and q.

Assume that instead of (24) all Z_i satisfy

$$\mathcal{E}\{\exp(\lambda Z_i)\} \le \exp(-\lambda^2\sigma^2/2(1-|\lambda|c)) ,\quad \textit{all } |\lambda|<1/c , \tag{25}$$

for some $\sigma>0$, $c\ge 0$. Then

$$P\{|n^{-1}\sum_{i=1}^{n} z_i Z_i| \geq \varepsilon\} \geq 2\exp(-n\varepsilon^2/2(\sigma^2 \sum_{i=1}^{n} z_i^2/n + c\,\varepsilon \max_i z_i)) \ .$$

Proof. The first inequality follows from a result due to Fuk and Nagaev (1971):

$$P\{|n^{-1}\sum_{i=1}^{n} z_i Z_i| \geq \varepsilon\} \leq 2(1+2/q)^q \sum_{i=1}^{n} \mathcal{E}\{|z_i Z_i|^q\}/(n\varepsilon)^q$$

$$+ 2\exp(-2e^{-q} n^2 \varepsilon^2 (q+2)^{-1} / \sum_{i=1}^{n} \mathcal{E}\{(z_i Z_i)^2\}) \ .$$

The second inequality is proved as follows. If $z_m = \max(z_1,\ldots,z_n)$, then for all λ with $|\lambda| z_m c < 1$ we have

$$P\{|n^{-1}\sum_{i=1}^{n} z_i Z_i| \geq \varepsilon\} \leq e^{-\lambda n\varepsilon} \prod_{i=1}^{n} \exp(\sigma^2 z_i^2 \lambda^2 / 2(1 - z_m|\lambda| c))$$

$$\leq \exp(-n\varepsilon^2/2(\sigma^2 \sum_{i=1}^{n} z_i^2/n + c z_m \varepsilon))$$

upon taking $\lambda c z_m = \varepsilon c z_m/(\sigma^2 \sum_{i=1}^{n} z_i^2/n + c z_m \varepsilon)$. Lemma 3 follows by symmetry since the same bound is valid if the Z_i are replaced by $-Z_i$. Q.E.D.

Proof of Theorem 1. Since

$$\operatorname*{ess\,sup}_A |m_n(x) - m(x)| \leq \operatorname*{ess\,sup}_A |\sum_{i=1}^{n} (Y_i - m(X_i)) K((x-X_i)/h_n)/g_n(x)| \tag{26}$$

$$+ \operatorname*{ess\,sup}_A |\sum_{i=1}^{n} (m(X_i) - m(x)) K((x-X_i)/h_n)/g_n(x)| \ ,$$

it follows from Lemma 1 that we need only take care of the first term on the right-hand side of (26). Since K can only take $k+1$ values, K must have compact support. Let us define these values as $a_1 > a_2 > \ldots > a_k > a_{k+1} = 0$, and let $u^* > 0$ be any point on the real line with $L(u^*) > 0$. Looking at the vector $[K((x-X_1)/h_n),\ldots,K((x-X_n)/h_n)]$ as a function of x, we notice that it can take at most

$$(2n)^{t(p,d)k}$$

values. For $k=1$, this is a consequence of Lemma 2. For $k > 1$, use Lemma 2 and the fact that the set of y in $\mathbb{R}^d$ on which $K(y) = a_i$ is the intersection of at most $k-1$ sets, each defined by an inequality of the form $K(y) \geq a_j$ (or, equivalently, by $\|y\| \leq b_j$ for some $b_j > 0$). This is in contrast to the exponential bound k^n that one would suspect. If $A = A(X_1,\ldots,X_n)$ is the partition induced by the said vector (notice that A can have at most $(2n)^{tk}$ member sets) and if the value of the vector on the i-th member set

A_i is $(a_{i1},\ldots,a_{in})$, then we have the following inequality:

$$P\{\operatorname*{ess\,sup}_A \Big|\sum_{i=1}^{n}(Y_i-m(X_i))K((x-X_i)/h_n)/g_n(x)\Big| \geq \varepsilon\} \tag{27}$$

$$= \mathcal{E}\{P\{\operatorname*{ess\,sup}_A \Big|\sum_{i=1}^{n}(Y_i-m(X_i))K((x-X_i)/h_n)/g_n(x)\Big| \geq \varepsilon : X_1,\ldots,X_n\}\}$$

$$\leq P\{(X_1,\ldots,X_n)\notin B\} + \mathcal{E}\{I_B(X_1,\ldots,X_n)\operatorname*{ess\,sup}_{(X_1,\ldots,X_n)} \sup_{A_j\in A(X_1,\ldots,X_n)} (2n)^{tk} \times$$

$$\times P\{\Big|\sum_{i=1}^{n}(Y_i-m(X_i))a_{ji}/\sum_{i=1}^{n}a_{ji}\Big| \geq \ |X_1,\ldots,X_n\ \}\}$$

for any set B. Now given $X_1,\ldots,X_n$ the random variables $Y_1-m(X_1)$ ..., $Y_n-m(X_n)$ are independent and have zero mean. For any $A_j\in A(X_1,\ldots,X_n)$ with corresponding vector $(a_{j1},\ldots,a_{jn})$ we have by Lemma 3:

$$P\{\Big|\sum_{i=1}^{n}(Y_i-m(X_i))a_{ji}/n\Big| \geq (\sum_{i=1}^{n}a_{ji}/n)\varepsilon\,|X_1,\ldots,X_n\}$$

$$\leq \begin{cases} 2\exp(-n\varepsilon^2 b_1^2/2(\sigma^2 b_2 + c\varepsilon b_1 b_\infty)), & \text{(28a)}\\ k_1 b_q/n^{q-1}\ \varepsilon^q b_1^q + 2\exp(-k_2 n^2\varepsilon^2 b_1^2/nb_2), & \text{(28b)}\end{cases}$$

according as the observations are (a) uniformly generalized Gaussian, or (b) uniformly in L_q $(q\geq 2)$. Here $b_r = \sum_{i=1}^{n}a_{ji}^r/n$ for any $r>0$, and $b_\infty = \max(a_{j1},\ldots,a_{jn})$. As usual, k_1, k_2, σ and c are used to denote nonnegative constants. The inequalities can be further reduced to the point where the right-hand sides are of the forms

$$2\exp(-k_3 nb_1/b_\infty)$$

and

$$k_4(b_\infty/nb_1)^{q-1} + 2\exp(-k_5 nb_1/b_\infty) < k_6(b_\infty/nb_1)^{q-1}$$

in which k_3,k_4,k_5,k_6 are positive numbers not depending upon n or $(a_{j1},\ldots,a_{jn})$. We use $b_2\leq b_1 b_\infty$ and $b_q \leq b_1 b_\infty^{q-1}$ in (28).

Let us take for B the set of all $(x_1,\ldots,x_n)\in\mathcal{R}^{dn}$ such that

$$\inf_A \mu_n(S(x,u^*h_n)) \geq \gamma h_n^d\ ,$$

where a small constant $\gamma > 0$ is to be picked later on. If $(X_1,\ldots,X_n) \in B$ then it is clear that for any $A_j \in \mathcal{A}(X_1,\ldots,X_n)$ we must have

$$nb_1/b_\infty \geq (\gamma L(u^*)/a_1)nh_n^d \ .$$

Consequently, (27) can be bounded above by

$$\left.\begin{array}{l} P\{\inf_A \mu_n(S(x,u^*h_n))<\gamma h_n^d\} + (2n)^{tk_2} \exp(-k_7 nh_n^d) \\ \\ \text{or} \\ \\ P\{\inf_A \mu_n(S(x,u^*h_n))<\gamma h_n^d\} + (2n)^{tk}k_8/(nh_n^d)^{q-1} \end{array}\right\} \quad (29)$$

depending on whether the observations are uniformly generalized Gaussian or uniformly in L_q. Again, k_7 and k_8 are positive constants not depending upon n. Proceeding as in the proof of Lemma 1 we notice that for some k_9, $k_{10},k_{11} > 0$ and some $\gamma > 0$ sufficiently small,

$$P\{\inf_A \mu_n(S(x,u^*h_n)) < \gamma h_n^d\} \leq k_9 h_n^{-d} \exp(-k_{10} nh_n^d) \ , \text{ if } \ h_n < k_{11} \ . \qquad (30)$$

Theorem 1 follows from (26-30) and the Borel-Cantelli lemma. Q.E.D.

Proof of Theorem 2. If $K_{\max} = \sup K(x)$ and $\{s_n\}$ is a sequence of positive integers with $s_n \overset{n}{\uparrow} \infty$, then we define the functions K_n on $\mathbb{R}^d$ by

$$K_n(x) = \inf\{K_{\max} j/s_n \mid K_{\max} j/s_n \geq K(x),\ j = 0,1,\ldots,s_n\} \ .$$

Using a small modification of the proof of Theorem 1 we see that there exist constants $k_1,k_2,k_3,k_4 \geq 0$ depending upon ε such that

$$P\{\operatorname{ess}\sup_A \Big| \sum_{i=1}^n (Y_i - m(X_i))K_n((x-X_i)/h_n) \Big| \Big/ \sum_{i=1}^n K_n((x-X_i)/h_n) \geq \varepsilon\} \qquad (31)$$

$$\leq k_1 h_n^{-d} \exp(-k_2 nh_n^d) + k_3 (2n)^{ts_n} \exp(-k_4 nh_n^d) \ .$$

Indeed, K_n can take at most $1+s_n$ values. Now (31) follows if we can show that for s_n large enough, the bounds in (29) hold uniformly over all K_n. If K takes more than 2 values, then one can find $u^{**} > u^* > 0$ and s^* such that $K_n(x) \geq L(u^{**})$ whenever $\|x\| \leq u^*$ and $s_n > s^*$. As in the previous proof, this would give us a lower bound on nb_1/b_∞ that looks like

$$(\gamma L(u^{**})/K_{\max})nh_n^d \ .$$

The same bound is valid for all K_n with $s_n > s^*$. Since the first term in (29) does not depend upon K_n, (31) follows.

Using (26) and (31) it is easily seen that we need only find an upper bound for the following probability:

$$\left.\begin{aligned} &P\{\operatorname{ess\,sup}_A \sum_{i=1}^n |Y_i-m(X_i)|\cdot|K_n((x-X_i)/h_n)-K((x-X_i)/h_n)|/g_n(x) \ge \varepsilon\} \\ &\qquad \le \{P \inf_A \sum_{i=1}^n K((x-X_i)/h_n) \le \sum_{i=1}^n |Y_i-m(X_i)|/(\varepsilon s_n)\} \\ &\le P\{\inf_A \sum_{i=1}^n K((x-X_i)/h_n) \le \delta n/\varepsilon s_n\} + P\{n^{-1}\sum_{i=1}^n |Y_i-m(X_i)| > \delta\} \end{aligned}\right\} \quad (32)$$

because

$$\left|\sum_{i=1}^n (Y_i-m(X_i))K_n((x-X_i)/h_n)/\sum_{i=1}^n K_n((x-X_i)/h_n)\right| - \left|\sum_{i=1}^n (Y_i-m(X_i))K((x-X_i)/h_n)/g_n(x)\right|$$
$$\le \sum_{i=1}^n |Y_i-m(X_i)|\,|K_n((x-X_i)/h_n)-K((x-X_i)/h_n)|/g_n(x).$$

The random variables $|Y_1|,|Y_2|,\ldots,|Y_n|$ are independent identically distributed and if m is bounded, they satisfy an inequality as in (25). If we pick $\delta > 4\mathcal{E}\{|Y_1|\}$ and $\delta > 2\operatorname{ess\,sup}_A |m(X_1)|$ then

$$\left.\begin{aligned} &P\{n^{-1}\sum_{i=1}^n |Y_i-m(X_i)| > \delta\} \le P\{n^{-1}\sum_{i=1}^n (|Y_i| + |m(X_i)|) > \delta\} \\ &\qquad \le P\{n^{-1}\sum_{i=1}^n |Y_i| > \delta/2\} + P\{n^{-1}\sum_{i=1}^n |m(X_i)| > \delta/2\} \\ &\qquad \le P\{n^{-1}\sum_{i=1}^n (|Y_i|-\mathcal{E}\{|Y_i|\}) > \delta/2 - \mathcal{E}\{|Y_1|\}\} \\ &\qquad \le P\{n^{-1}\sum_{i=1}^n (|Y_i|-\mathcal{E}\{|Y_i|\}) > \delta/4\} \le k_5 \exp(-k_6 n), \end{aligned}\right\} \quad (33)$$

where k_5,k_6 are positive constants depending upon δ. Let us now pick s_n as the nearest integer larger than θ/h_n^d and let $\theta > 0$ be chosen large enough to make $\delta/\varepsilon\theta<\gamma$. The first term in (32) is then bounded above as in (21). Collecting bounds, noting that

$$k_3(2n)^{t\theta/h_n^d} \exp(-k_4 n h_n^d)$$

is summable with respect to n for all k_3, t, θ, and invoking the Borel-Cantelli lemma yields Theorem 2. Q.E.D.

RÉSUMÉ

Soit $(X_1,Y_1), \ldots, (X_n,Y_n)$ une suite de vecteurs aléatoires indépendants, identiquement distribués à valeurs dans $\mathcal{R}^d \times \mathcal{R}$, Nadaraya (1964) et Watson (1964) ont proposé d'estimer la fonction de régression $m(x) = \mathcal{E}(Y_1|X_1 = x)$ par

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

où K est une densité connue et $\{h_n\}$ est une suite de nombres positifs satisfaisant certaines hypothèses. Dans cet article, on présente plusieurs ensembles de conditions sous lesquelles $\mathrm{ess}_X\sup|m_n(X)-m(X)|$ converge fortement vers 0 (X est un vecteur aléatoire, indépendent des données ayant la même distribution que X_1). Les théorèmes présentés sont vrais quelle que soit la distribution de X, et pour toutes les suites $\{h_n\}$ satisfaisant $h_n \to 0$ et $nh_n^d/\log n \to \infty$.

REFERENCES

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

Cover, T. (1965). Geometrical and statistical properties of systems of linear inequalities with applications in pattern recognition. *IEEE Trans. Computers*, 10, 326-334.

Devroye, L.P., and Wagner, T.J. (1978). On the L_1 convergence of kernel estimators of regression functions with applications in discrimination. Submitted for publication.

Fuk, D.K., and Nagaev, S.V. (1971). Probability inequalities for sums of independent random variables. *Theor. Probability Appl.*, 16, 643-660.

Greblicki, W. (1974). Asymptotically optimal probabilistic algorithms for pattern recognition and identification. *Seria Monografie*, 3, Prace Naukowe Instytutu Cybernetyki Technicznej Politechniki Wroclawskiej 18, Wroclaw.

Hoeffding, W. (1963). Probability inequalities for the sums of bounded random variables. *J. Amer. Statist. Assoc.*, 58, 13-30.

Nadaraya, E.A. (1964). On estimating regression. *Theor. Probability Appl.*, 9, 141-142.

Nadaraya, E.A. (1965). On nonparametric estimates of density functions and regression curves. *Theor. Probability Appl.*, 10, 186-190.

Nadaraya, E.A. (1970). Remarks on non-parametric estimates for density functions and regression curves. *Theor. Probability Appl.*, 15, 134-137.

Noda, K. (1976). Estimation of a regression function by the Parzen kernel-type density estimators. *Ann. Inst. Statist. Math.*, 28, 221-234.

Révész, P. (1977). How to apply the method of stochastic approximation in the nonparametric estimation of a regression function. *Math. Operationsforsch. Statist.*, 8, 119-126.

Schuster, E.F. (1972). Joint asymptotic distribution of the estimated regression function at a finite number of distinct points. *Ann. Math. Statist.*, 43, 84-88.

Vapnik, V.N., and Chervonenkis, A.Ya. (1971). On the uniform convergence of the relative frequencies of events to their probabilities. *Theor. Probability Appl.*, 16, 264-280.

Watson, G.S. (1964). Smooth regression analysis. *Sankhyā Ser. A*, 26, 359-378.

Received 26 March 1978
Revised 28 September 1978

School of Computer Science
McGill University
805 Sherbrooke Street West
Montréal, Québec
H3A 2K6