Chin. Ann. of Math.
6B (1) 1985

ALMOST SURE CONVERGENCE OF NONPARAMETRIC REGRESSION ESTIMATES

CHEN XIRU(陈希孺)*

Abstract

Let (X, Y), (X_i, Y_i), $i=1, \cdots, n$, be iid. $R^d\times R^1$-ralued vandom vectors with $E(|Y|)<\infty$ and $m(x)=E(Y|X=x)$ be the regression function. Select the weight functions $W_{ni}(x)=W_{ni}(x; X_1, \cdots, X_n)$, and use $m_n(x)=\sum_{i=1}^n W_{ni}(x)Y_i$ as an estimator of $m(x)$. This paper shows that $\lim_{n\to\infty} m_n(X)=m(X)$, a. s., under weaker conditions.

§ 1. The main result

Let (X, Y), (X_i,Y_i) $i=1, \cdots, n$, be iid. $R^d\times R^1$-valued random vectors with $E(|Y|)<\infty$. A much-studied method (see [1]) for estimating the regression function

$$m(x)=E(Y|X=x)$$

is as follows: Select the weight functions $W_{ni}(x)=W_{ni}(x; X_1, \cdots, X_n)$, $i=1, \cdots, n$, and use

$$m_n(x)=\sum_{i=1}^n W_{ni}(x)Y_i$$

as an estimator of $m(x)$.

The purpose of this paper is to prove a result concerning the a. s. convergence of m_n. To begin with, for each fixed $x\in R^d$, define the ranks $R_1, \cdots, R_n$ by

$$\|X_{R_1}-x\|\leqslant\|X_{R_2}-x\|\leqslant\cdots\leqslant\|X_{R_n}-x\|$$

with ties broken by comparing indices. Choose an integer k_n for each n such that

$$1\leqslant k\leqslant n,\ k/n\to 0,\ \log n/k\to 0, \text{ as } n\to\infty \tag{1}$$

and define

$$C(\varepsilon)=\sup_n\{\max(\sum_i{}' W_{ni}\text{: the number of terms contained in } \sum{}' \text{ does not exceed } k\varepsilon)\}. \tag{2}$$

Following is the main result of this paper.

Theorem 1. *Suppos that $k=k_n$ satisfies (1), Y is bounded, and*

Manuscript received January 3, 1983.

* Department of Mathematics, China University of Science and Technology, Hefei, China.

$$W_{ni}(X)\geqslant 0,\ \sum_{i=1}^{n} W_{ni}(X)=1,\ \text{a. s.} \tag{3}$$

$$\sum_{i=1}^{k} W_{nR_i}^2(X)=o\left(\frac{1}{\log n}\right),\ \text{a. s.} \tag{4}$$

$$\lim_{n\to\infty}\sum_{i=k+1}^{n} W_{nR_i}(X)=0,\ \text{a. s.} \tag{5}$$

$$\lim_{\varepsilon\to 0} C(\varepsilon)=0,\ \text{a. s.} \tag{6}$$

Then we have

$$\lim_{n\to\infty} m_n(X)=m(X),\ \text{a. s.} \tag{7}$$

This result gives an improvement of a result by Devroye (See [2], Th. 4). See also the Remark 2 at the end of this paper).

Remark 1. The proof, to be given below, applies to the case where (3) is weakened to

$$\lim_{n\to\infty}\sum_{i=1}^{n} W_{ni}(X)=1,\ \limsup_{n\to\infty}\sum_{i=1}^{n} |W_{ni}(X)|<\infty,\ \text{a. s.}, \tag{3'}$$

but, in (5) and the definition of $C(\varepsilon)$, $W_{nR_i}(x)$ should be replaced by $|W_{nR_i}(x)|$.

§ 2. First part of the proof

Write

$$m_n(X)-m(X)=\sum_{i=1}^{n} W_{ni}(X)(Y_i-m(X_i))+\sum_{i=1}^{n} W_{ni}(X)(m(X_i)-m(X))$$
$$\triangleq J_{1n}(X)+J_{2n}(X)+J_{3n}(X), \tag{8}$$

where

$$J_{1n}(X)=\sum_{i=1}^{k} W_{nR_i}(X)(Y_{R_i}-m(X_{R_i})), \tag{9}$$

$$J_{2n}(X)=\sum_{i=1}^{k} W_{nR_i}(X)(m(X_{R_i})-m(X)), \tag{10}$$

$$J_{3n}(X)=\sum_{i=k+1}^{n} W_{nR_i}(X)(Y_{R_i}-m(X)). \tag{11}$$

Further,

$$|Y|\leqslant M<\infty,\ M \text{ is a constant.} \tag{12}$$

By (5), (11), (12), we have

$$\lim_{n\to\infty} J_{3n}(X)=0,\ \text{a. s.} \tag{13}$$

To deal with J_{1n}, we need a result proved in [3]:

Lemma 0. *Let $z_1, \cdots, z_n$ be independent variables with mean zero, and $a_1, \cdots, a_n$ be constants such that $\sum_{i=1}^{n} a_i^2=1$. Then*

$$E\left(\sum_{i=1}^{n} a_i z_i\right)^{2s}\leqslant 3^s(2s-1)!!\max_{1\leqslant i\leqslant n} Ez_i^{2s},\ s=1,2,\cdots. \tag{14}$$

For simplicity, we assume that (3)—(6) hold for all values of$(X, X_1, X_2, \cdots)$. For given $X=x$, $X_i=x_i$, $i=1, 2, \cdots$, the conditional distribution of $J_{1n}(X)$ is the same as that of $\sum_{i=1}^{k} c_i z_i$, where $z_1, \cdots, z_k$ are independent variables each bounded by $2M$ and with mean zero, $c_1, \cdots, c_k$ are constants which, according to (4), satisfy

$$d_n \triangleq \sum_{i=1}^{k} c_i^2 = \sum_{i=1}^{k} W_{nR_i}(x) = o\left(\frac{1}{\log n}\right). \tag{15}$$

Writing $T_n = \sum_{i=1}^{k} c_i z_i / \sqrt{d_n}$, by (14) we have, for fixed $\varepsilon > 0$,

$$\begin{aligned} P\left(\left|\sum_{i=1}^{k} c_i z_i\right| \geqslant \varepsilon\right) &= P(|T_n| \geqslant \varepsilon/\sqrt{d_n}) = \exp(-\varepsilon^2/d_n) E(e^{T_n^2}) \\ &= \exp(-\varepsilon^2/d_n) \sum_{s=0}^{\infty} \frac{1}{s!} E T_n^{2s} \\ &\leqslant \exp(-\varepsilon^2/d_n)\left[1 + \sum_{s=1}^{\infty} \frac{1}{s!} 3^s (2s-1)!!(2M)^{2s}\right] \\ &\leqslant \exp(-\varepsilon^2/d_n)\left[1 + \sum_{s=1}^{\infty} (24M^2)^s\right]. \end{aligned} \tag{16}$$

Suppose first that $M \leqslant 1/5$, then by (15), (16), we have

$$\sum_{n=1}^{\infty} P\left(\left|\sum_{i=1}^{k} c_i z_i\right| \geqslant \varepsilon\right) \leqslant 25 \sum_{n=1}^{\infty} \exp(-\varepsilon^2/d_n) < \infty.$$

Hence we have proved that for any fixed $X=x$, $X_i=x_i$, $i=1, 2, \cdots$, the assertion (note that J_{1n} depends also on X_i, Y_i, $i=1, \cdots, n$)

$$\lim_{n\to\infty} J_{1n}(x; x_1, Y_1, \cdots, x_n, Y_n) = 0, \text{ a. s.}$$

is true. This in turn proves that

$$\lim_{n\to\infty} J_{1n}(X; X_1, Y_1, \cdots, X_n, Y_n) = 0, \text{ a. s.} \tag{17}$$

Replacing by $Y/(5M)$(and Y_i by $Y_i/(5M)$ accordingly) in case $M > 1/5$, we see that (17) is true for any M.

§ 3. Second part of the proof

We shall put $C_{ki} = W_{nR_i}(x)$ for simplicity of writing.

Lemma 1. *Let ξ be binomial variable with parameters n and p. Then for $\varepsilon > 0$,*

$$P(|\xi/n - p| \geqslant \varepsilon) \leqslant 2\exp(-n\varepsilon^2/(2p+\varepsilon)).$$

For the proof, see Hoeffding [4].

Lemma 2. *Let F be a probability measure on the class $\mathfrak{B}^r$ of all Borel sets in R^r, $A \in \mathfrak{B}^r$ and $F(A) > 0$. Denote by $S(x, \rho)$ the open or closed sphere centered at x and with radius ρ, and*

$$\tilde{A} = \{x : x \in A, \lim_{\rho\to 0} F(S(x, \rho) \cap A)/F(S(x, \rho)) = 1\}.$$

Then $F(A - \tilde{A}) = 0$.

For the Proof, see [5], p. 189.

Now turn to the proof of $J_{2n}(x)\to 0$, a. s. For any $\rho>0$, denote by S_ρ, $\bar{S}_\rho$ and $\bar{S}_\rho^*$ the open, closed and the surface of the sphere with radius ρ and centered at x, respectively. Let N be a natural number, and

$$B_{Ni}=\left(x:\frac{i-1}{N}\leqslant m(x)<\frac{i}{N}\right),\ i=0,\ \pm1,\ \pm2,\ \cdots,$$

$$\tilde{B}_{Ni}=\{x:x\in B_{Ni},\ \lim_{\rho\to0}F(S_\rho\cap B_{Ni})/F(S_\rho)=1\}.$$

According to Lemma 2, $F(B_{Ni}-\tilde{B}_{Ni})=0$, $i=0,\ \pm1,\ \pm2,\ \cdots$. Hence, on writing

$$\tilde{B}_N=\bigcup_{i=-\infty}^{\infty}\tilde{B}_{Ni},\ \tilde{B}=\bigcap_{N=1}^{\infty}\tilde{B}_N,$$

we have $F(\tilde{B}_N)=1$ for $N=1,\ 2,\ \cdots$, thus $F(\tilde{B})=1$. Further, by Lemma 2, we also have $\lim_{\rho\to0}F(\bar{S}_\rho\cap B_{Ni})/F(\bar{S}_\rho)=1$.

Further, if $\rho'\to0$ along a sequence such that $F(S_{\rho'}^*)/F(S_{\rho'})\geqslant\alpha>0$ for some fixed $\alpha>0$, then $\lim_{\rho'\to0}F(S_{\rho'}^*\cap B_{Ni})/F(S_{\rho'}^*)=1$ for $x\in\tilde{B}_{Ni}$.

Define $\tilde{C}=\{x:F(S_\rho)>0$ for any $\rho>0\}$, which is the support of F. We have $F(\tilde{C})=1$. Finally, putting $A=\tilde{B}\cap\tilde{C}$, we have $F(A)=1$.

Now fix arbitrarily $x\in A$ and proceed to show that $J_{2n}(x)\to0$, a. s. First consider the case $F(\{x\})>0$. In this case, by the Law of Large Numbers and the definition of $J_{2n}(x)$, it is easily seen that

$$P(J_{2n}(x)=0 \text{ for } n \text{ sufficiently large})=1.$$

Therefore, for such x, $J_{2n}(x)\to0$ a. s. is trivially true.

Next assume that $F(\{x\})=0$. For fixed n, find $\rho_n>0$ (in the following ρ_n will be simplified to ρ) such that

$$F(S_\rho)\leqslant k/n,\ F(\bar{S}_\rho)\geqslant k/n.$$

From $F(\{x\})=0$, $x\in\tilde{C}$ and $k/n\to0$, we see that $\lim_{n\to\infty}\rho=0$. Further, Since $x\in\tilde{B}$, for any positive integer N there exists i_N such that

$$x\in\tilde{B}_{Ni_N},\ N=1,\ 2,\ \cdots.$$

We shall fix N in the following discussion. Given $0<\eta<1$, consider separately three cases for n:

1. $$k/n\geqslant F(S_\rho)\geqslant(1-\eta)k/n.$$

From $x\in\tilde{B}_{Ni_N}$ and $\lim_{n\to\infty}\rho=0$, we see that for n sufficienty large

$$\begin{aligned}p_n&=F(S_\rho\cap B_{Ni_N})\geqslant(1-\eta^2)k/n,\\ p_n^*&=F(S_\rho\cap B_{Ni_N}^c)\leqslant\eta k/n.\end{aligned}\tag{18}$$

Define

$$\begin{aligned}D_n&=\{X_1,\ \cdots,\ X_n\}\cap S_\rho\cap B_{Ni_N},\ Q_n=\#(D_n),\\ D_n^*&=\{X_1,\ \cdots,\ X_n\}\cap S_\rho\cap B_{Ni_N}^c,\ Q_n^*=\#(D_n^*).\end{aligned}\tag{19}$$

According to Lemma 1, and noticing (18), one gets

$$\begin{aligned}P(|Q_n/n-p_n|\geqslant\eta p_n)&\leqslant2\exp[-n\eta^2p_n^2/(2p+\eta p_n)]\\ &\leqslant2\exp[-k(1-\eta)^2\eta^2/(2+\eta)],\end{aligned}\tag{20}$$

$$P(|Q_n^*/n-p_n^*|\eta p_n)\leqslant 2\exp[-k(1-\eta)^2\eta^2/(2+\eta)]. \tag{21}$$

Since $k/\log n\to\infty$. for any $\eta\in(0, 1)$ the right hand sides of (20), (21) are terms of convergent series. Hence by the lemma of Borel-Cantell, the following assertion is true with probability one: For n large enough

$$Q_n\geqslant n(1-\eta)p_n\geqslant k(1-\eta)^3\geqslant k(1-3\eta), \tag{22}$$

$$Q_n^*\leqslant n(p_n^*+\eta p_n)\leqslant k\eta+k\eta(1-\eta)^2\leqslant 2k\eta. \tag{23}$$

By the definition of $R_1, \cdots, R_n$, in case that (22) and (23) hold, among $X_{R_1}, \cdots, X_{R_k}$ there will be at least $(1-3\eta)k$ members lying in D_n. By the meaning of B_{Ni_N}, for such X_{R_i} we have $|m(x)-m(X_{R_i})|\leqslant 1/N$. Hence, noticing the definition of $c(\varepsilon)$ mentioned earlier, we get

$$|J_{2n}(x)|\leqslant 1/N+2Mc(3\eta). \tag{24}$$

That is to say, for those n under discussion, (24) holds for large n with probaility one.

2. $k/n\leqslant F(\bar{S}_\rho)\leqslant(1+\eta)k/n$.

Modifying previous definitions of D_n, Q_n, p_n, D_n^*, Q_n^*, p_n^* by changing S_ρ to $\bar{S}_\rho$, we get for n large enough

$$p_n\geqslant(1-\eta)k/n,\ p_n^*\leqslant\eta(1+\eta)k/n\leqslant 2\eta k/n.$$

Again by Lemma 1

$$P(|Q_n|n-p_n|\geqslant\eta p_n)\leqslant 2\exp[-k(1-\eta)^2\eta^2/(2+\eta)],$$

$$P(|Q_n^*/n-p_n^*|\geqslant\eta p_n)\leqslant 2\exp[-k(1-\eta)^2\eta^2/(2+\eta)].$$

An argument similar to the previous case shows that the following assertion holds with probability one: For n large enough

$$Q_n\geqslant(1-\eta)np_n\geqslant k(1-\eta^2)\geqslant k(1-2\eta), \tag{25}$$

$$Q_n^*\leqslant n(p_n^*+\eta p_n)\leqslant 2\eta k+\eta(1+\eta)k\leqslant 4\eta k. \tag{26}$$

In the same vein as the previous case, in case that (25) and (26) hold, among $X_{R_1}, \cdots, X_{R_k}$ there will be at least $(1-4\eta)k$ members lying in D_n, and we get

$$|J_{2n}(x)|\leqslant 1/N+2Mc(4\eta). \tag{27}$$

3. $F(S_\rho)<(1-\eta)k/n$, $F(\bar{S}_\rho)>(1+\eta)k/n$.

Restore the definitions of D_n, Q_n, p_n, Q_n^*, Q_n^*, p_n^* to its original form in case 1, and let

$$Z_n=\{\text{the number of } X_1, \cdots, X_n'\text{s contained in } \bar{S}_\rho\}$$

and $H_{n1}=\{Z_n\geqslant k\}$. Using Lemma 1, one can show that

$$P(H_{n1}\text{ occurs for } n \text{ sufficiently large})=1. \tag{28}$$

Indeed, putting $q=F(\bar{S}_\rho)\geqslant(1+\eta)k/n$, by Lemma 1, we have

$$P(H_{n1}^c)\leqslant P(|Z_n/n-q|\geqslant q-k/n)\leqslant 2\exp[-n(q-k/n)^2/3q].$$

The last expression is strictly decreasing in q. Hence

$$P(H_{n1}^c)\leqslant 2\exp[-n(\eta^2k^2/n^2)/(3(1+\eta)k/n)]=2\exp[-k\eta^2/(3(1+\eta))].$$

This proves (28). Similarily we prove that

$$P(Q_n^*\leqslant 2\eta k\text{ for } n \text{ sufficiently large})=1, \tag{29}$$

$$P(Q_n+Q_n^*\leqslant(1-\eta/2)k \text{ for } n \text{ sufficiently large})=1. \tag{30}$$

Define

$$H_{n2}=\{Q_n+Q_n^*\leqslant(1-\eta/2)k,\ Z_n\geqslant k\},$$

then by (28), (30), it follows that

$$P(H_{n2} \text{ occurs for } n \text{ sufficiently large})=1. \tag{31}$$

A few words on (31): It means that for n sufficiently large, it is almost certain that there will be at most $k(1-\eta/2)$ members of $\{X_{R_1}, \cdots, X_{R_k}\}$ lying in S_ρ, but the remaining part can all be found in S_ρ^*. Further, by (29) we see that

$$\#(\{X_{R_1}, \cdots, X_{R_k}\}\cap S_\rho\cap B^c_{Ni_N}\}\leqslant 2\eta k.$$

Since $F(S_\rho^*)=F(\bar{S}_\rho)-F(S_\rho)\geqslant 2\eta k/n$ and $F(S_\rho^*)/F(S_\rho)\geqslant 2\eta(1-\eta)$, if we define $w_n=F(S_n^*\cap B_{Ni_N})/F(S_\rho^*)$ and $w_n'=F(S_\rho^*\cap B^c_{Ni_N})/F(S_\rho^*)$, then $w_n\geqslant 1-\eta$, $w_n'\leqslant\eta$ for n large enough. It is not difficult to prove that if a certain part of $X_1, \cdots, X_n$, for example $X_{i_1}, \cdots, X_{i_j}$, fall into S_ρ^*, then each of them will appear in B_{Ni_N} or $B^c_{Ni_N}$ with probability w_n or w_n' respectively, and these events occur independently. Suppose that we need to choose l members of $\{X_1, \cdots, X_n\}\cap S_\rho^*$ to cover the deficit resulting from $Q_n+Q_n'<k$, then as stated earlier, with probability one, for n sufficiently large we have $l\geqslant k\eta/2$, and these l members should be chosen from $\{X_1, \cdots, X_n\}\cap S_\rho^*$ with possibly smallest indices. Hence by Lemma 1, for $l\geqslant k\eta/2$ we have

$$\begin{aligned}&P(\text{The number of elements falling into } S_\rho^*\cap B^c_{Ni_N} \text{ among the } l \text{ elements}\\ &\quad\text{mentioned above } \geqslant 2\eta l)\leqslant 2\exp(-l\eta^2/3\eta)=2\exp(-l\eta/3)\\ &\leqslant 2\exp(-k\eta^2/6),\end{aligned} \tag{32}$$

and this estimation does not depend on l for $l\geqslant k\eta^2/2$. Summing up (28), (29), (31), and (32), one sees that with probability one the following assertion holds: For n sufficiently large $\#(\{X_{R_1}, \cdots, X_{Q_k}\}\cap B^c_{Ni_N})\leqslant 2\eta k+2\eta k=4\eta k$, and when this occurs we have

$$|J_{2n}(x)|\leqslant 1/N+2Mc(4\eta). \tag{33}$$

From (24), (27) and (33). we see that, with probability one,

$$J_{2n}(x)|\leqslant 1/N+2Mc(4\eta) \tag{34}$$

for n sufficiently large. Since the positive integer N and $\eta\in(0, 1)$ can be arbitaraily chosen and $\lim\limits_{s\uparrow 0} c(s)=0$,one gets from (34)

$$\lim_{n\to\infty} J_{2n}(x)=0, \text{ a. s.} \tag{35}$$

From (35) it follows that $J_{2n}(X)\to 0$, a. s. This, combining with (8)—(11), (13) and (17), finally gives (7). The proof of Theorem 1 is concluded.

References

[1] Stone, C. J., *Ann. Statist.*, **5** (1977), 595—645.
[2] Devroye, L., *Ann. Statist.*, **9** (1981), 1310—1319.
[3] Tao Bo & Cheng Ping, Chin. Ann. of Math., **2**: 4 (1981), 451—461, (in Chinese).
[4] Hoeffding, W., *J. Amer. Statist. Assoc.*, **58** (1963), 13—30.
[5] Wheeden, R. L. & Zygmund, A., Measure and Integral, Marcel Dekker, New York, 1977.