J. Math. Res. & Exposition, 1986, No. 2

On the Pojntwisise L_p Convergence Rates of Nearest Neighbor Estimate of Nonparametric Regression Function*

Wei Lai-sheng (韦来生) and *Su Chun* (苏淳)

(University of Science and Technology of China)

Abstract

Let (X, Y) be a $R^d \times R^1$- valued random vector with $E(|Y|)<\infty$, $m(x) = E(Y|X=x)$ be the regression function of Y with respect to X. Suppose that (X_i, Y_i), $i=1, \cdots, n$, are iid. samples drawn from (X, Y). It is desired to estimate $m(x)$ based on these samples. Deveroye discussed in 1981 (see [2]) the pointwise L_p- convergence of the nearest neighbor estimate $m_n(x)$ (see (5) of the present paper). Ih this article we further study the rate of this convergence. It is shown that if there exists $p \geqslant 2$ such that $E|Y|^p<\infty$, then $E|m_n(x)-m(x)|^p = O(n^{-\frac{p}{d+2}})$ a. s. for suitable choice of the weights C_{ni} (see (4) of the present paper).

1. Introduction and the main result

In order to estimate $m(x)$, we introduce a metric $\|x-y\|$ in R^d, and arrange $\|X_i - x\|$ in increasing order, i. e.

$$\|X_{R_1} - x\| \leqslant \|X_{R_2} - x\| \leqslant \cdots \leqslant \|X_{R_n} - x\| \tag{1}$$

(ties are breaken by comparing indices), Two common choice of $\|x-y\|$ are as follows

$$\|x-y\|^2 = \sum_{i=1}^{d} (x_i - y_i)^2 \tag{2}$$

$$\|x-y\| = \max_{1 \leqslant i \leqslant d} |x_i - y_i|, \tag{3}$$

The result of this paper is valid for both specifications.

Now choose a probability weight vector $\{C_{ni}, i=1, \cdots, n\}$ (i. e., $C_{ni} \geqslant 0$, $\sum_{i=1}^{n} C_{ni} = 1$), and define

$$W_{nR_i}(x) = C_{ni} \qquad i = 1, 2, \cdots, n, \tag{4}$$

The estimator of $m(x)$ which will be studied in the following is

$$m_n(x) = \sum_{i=1}^{n} W_{ni}(x) Y_i \tag{5}$$

this estimate is often called "nearest neighbor estimate". Since the appearance of the foundational work [1] of Stone (1977), further progress have been

* Received Jan. 31, 1983.

made in the research of this direction. Devroye[2] (1981) conisdered the point-wise consistency of this estimate, in this paper we shall conside the convergence rates of it.

Let $k=k_n$ be a natural number depend solely on n. Suppose that the weight vector $\{C_{ni}\}$ satisfies

$$(\mathrm{I})\begin{cases}(\mathrm{i})\ \text{there exist constants } c_1>0 \text{ and } c_2<\infty, \text{ such that } c_1\leqslant k_n n^{-\frac{2}{d+2}}\leqslant c_2, & (6)\\ (\mathrm{ii})\ \sup_n\{k_n \max_{1<i<k_n} C_{ni}\}<\infty,\ C_{ni}=O(n^{-\frac{1}{d+2}-1}),\ i=k_n+1,\cdots,n. & (7)\end{cases}$$

An obvious choice satisfying these conditions is $k_n=[n^{2/(d+2)}]$, $C_{ni}=\frac{1}{k_n}$ for $i=1,\cdots,k_n$ and $=0$ for $i=k_n+1,\cdots,n$.

Denote by μ the probability distribution of X, then the main result of this paper can be formulated in the following

Theorem If (i) $E|Y|^p<\infty$ for some $p\geqslant 2$.

(ii) $m(x)$ satisfies the Lipschitz condition of order 1,

(iii) $\{C_{ni}\}$ satisfies condition (I),

then

$$E|m_n(x)-m(x)|^p=O(n^{-\frac{p}{d+2}}),\ a.s.\ x(\mu) \tag{8}$$

2. Proof of the theorem.

We prove first some lemmas. In the following, c_0, c and c_p denote positive constants, $M_0(x)$, $M_1(x)$ and $M(x)$ denote constant depending upon x, (these constants can assume different values in each of their appearance, even within the same expression). S_ρ denote the open sphere of radius ρ centered at x.

Lemma 1. Let $h_n=n^{-\frac{1}{d+2}}a_n$, where $\{a_n\}$ is positive sequence of real number such that

$$\lim_{n\to\infty}a_n=\infty\ ,\quad \lim_{n\to\infty}h_n=\lim_{n\to\infty}a_n\cdot n^{-\frac{1}{d+2}}=0, \tag{9}$$

then

$$\lim_{n\to\infty}n^{\frac{d}{d+2}}\mu(S_{h_n})=\infty,\quad a.s.\ x(\mu) \tag{10}$$

Proof From the proof of Lemma 2.2 of [2], we know that the Lebesgue measure λ on R^d can be split up into two parts λ_1, λ_2, such that $\lambda=\lambda_1+\lambda_2$, $\lambda_1\ll\mu$, $\lambda_2\perp\mu$, and

$$\lim_{n\to\infty}\frac{\lambda(S_{h_n})}{\mu(S_{h_n})}=\lim_{n\to\infty}\frac{Mh_n^d}{\mu(S_{h_n})}=g(x)\quad,\ a.s.\ x(\mu), \tag{11}$$

where M is a positive constant and $g(x)=\frac{d\lambda_1}{d\mu}(x)$ is a nonnegative finite function.

By (11) we have

$$\frac{h_n^d}{\mu(S_{h_n})}=\frac{n^{\frac{d}{d+2}}h_n^d}{n^{\frac{d}{d+2}}\mu(S_{h_n})}=\frac{a_n^d}{n^{\frac{d}{d+2}}\mu(S_{h_n})}\to\frac{g(x)}{MM}\triangleq g_1(x)\ ,\ a.s.x(\mu),\ (n\to\infty)$$

since $0\leqslant g_1(x)<\infty$ and $\lim\limits_{n\to\infty}a_n^d=\infty$, we have

$$\lim_{n\to\infty}n^{\frac{d}{d+2}}\mu(S_{h_n})=\infty,\quad a.s.x(\mu)\ .$$

Lemma 2 . If (i) $E|f(X)|^p<\infty$, $p\geqslant 1$,

(ii) $f(x)$ satisfies the Lipschitz condition of order 1,

(iii) $\{W_{ni}(x)\}$ satisfies condition (I),

then

$$E[\sum_{i=1}^n W_{ni}(x)|f(X_i)-f(x)|^p]=O(n^{-\frac{p}{d+2}}\cdot a_n^p),\ a.s.x(\mu), \tag{12}$$

where $\{a_n\}$ satisfies (9).

Proof $I\triangleq E[\sum\limits_{i=1}^n W_{ni}(x)|f(X_i)-f(x)|^p]=E[\sum\limits_{i=1}^k C_{ni}|f(X_{R_i})-f(x)|^p]+$

$$+E[\sum_{i+k_n+1}^n C_{ni}|f(X_{R_i})-f(x)|^p]\triangleq I_1+I_2. \tag{13}$$

We consider I_2 first. Since

$$C_{ni}\leqslant c\cdot n^{-(\frac{p}{d+2}+1)}\leqslant cn^{-\frac{p}{d+2}}\cdot\frac{1}{n-k_n},\ i=k_n+1,\ \cdots,\ n,$$

we get

$$I_2\leqslant cn^{-\frac{p}{d+2}}\cdot E[\frac{1}{n-k_n}\sum_{i=k_n+1}^n|f(X_{R_i})-f(x)|^p\triangleq c\cdot n^{-\frac{p}{d+2}}E(J_0)\ , \tag{14}$$

$$E(J_0\mid\|X_{R_{k_n}}-x\|)=\int_{S^c_{\|X_{R_{k_n}}-x\|}}|f(y)-f(x)|^p\mu(dy)/\mu(S^c_{\|X_{R_{k_n}}-x\|}),$$

where $S^c_{\|\cdot\|}$ denote the set $R^d-S_{\|\cdot\|}$.

Write $G(\rho)=\mu(S_\rho)$, then the distribution of $\|X_{R_k}-x\|$ is

$$n\binom{n-1}{k-1}[G(\rho)]^{k-1}[1-G(\rho)]^{n-k}dG(\rho) \tag{15}$$

Put $A=\int|f(y)-f(x)|^p\mu(dy)$, then $A\leqslant 2^{p-1}[E|f(X)|^p+|f(x)|^p]\triangleq M_0(x)$,
Therefore

$$E(J_0)=E\left\{E(J_0\mid\|X_{R_k}-x\|)\right\}\leqslant An\binom{n-1}{k-1}\int_0^\infty\frac{1}{1-G(\rho)}[G(\rho)]^{k-1}\cdot$$

$$[1-G(\rho)]^{n-k}dG(\rho)=An\binom{n-1}{k-1}\int_0^1 x^{k-1}(1-x)^{n-k-1}dx=An\binom{n-1}{k-1}\beta(k,n-k)$$

$$=An\binom{n-1}{k-1}\frac{(k-1)!\ (n-k-1)!}{(n-1)!}=A\cdot\frac{n}{n-k}, \tag{16}$$

Since $\lim\limits_{n\to\infty}\frac{n}{n-k_n}=1$, we have

$$E(J_0)\leqslant 2A\leqslant 2M_0(x)$$

for n large enough. By (14), we get

$$I_2\leqslant c\cdot M_0(x)n^{-\frac{p}{d+2}}=O(n^{-\frac{p}{d+2}}) \tag{17}$$

Next consider I_1. By (7), we have $C_{ni}\leqslant c/k_n$, $i=1, 2, \cdots, k_n$. Hence

$$I_1\leqslant c\cdot E\Big[\frac{1}{k_n}\sum_{i=1}^{k_n}|f(X_{R_i})-f(x)|^p\Big]\triangleq c\cdot E(\tilde{J}_0),$$

$$E(\tilde{J}_0\,|\,\|X_{R_{k_n+1}}-x\|)=\int_{S_{\|X_{R_{k+1}}-1\|}}|f(y)-f(x)|^p\mu(dy)/\mu(S_{\|X_{R_{k_n}}-x\|})$$

Since $f(x)$ satisfies the Lipschitz condition of order 1, i.e., $|f(x)-f(y)|\leqslant L\|x-y\|$, (where L is a constant), we have

$$E(\tilde{J}_0)=E\{E(\tilde{J}_0\,|\,\|X_{R_{k_n+1}}-x\|)\}\leqslant ch_n^p+2^{p-1}(f^*(x)+|f(x)|^p)\cdot P(\|X_{R_{k_n+1}}-x\|\geqslant h_n)$$
$$\leqslant h_n) \tag{18}$$

where $f^*(x)\triangleq\sup\limits_{r>0}[\int_{S_r}|f(y)|^p\mu(dy)/\mu(S_r)]$. By [3] (page 188) we get

$$0<f^*(x)<\infty,\quad a\cdot s\cdot x(\mu) \tag{19}$$

We get (18) by means of

$$\int_{S_{\|X_{R_{k_n+1}}-x\|}}|f(y)-f(x)|^p\mu(dy)/\mu(S_{\|X_{R_{k_n+1}}-x\|})\leqslant 2^{p-1}(f^*(x)+|f(x)|^p).$$

Let Z_n be the number of $X_1, X_2, \cdots X_n$ falling into S_{h_n}, then Z_n obeys the binomial distribution, i.e., $Z_n\sim B(n, t_n)$, $t_n=\mu(S_{h_n})>0$, If $\|X_{R_{k_n+1}}-x\|\geqslant h_n$, then $Z_n\leqslant k_n$.

From lemma 1 we obtain, for $h_n=n^{-\frac{1}{d+1}}\cdot a_n$ and n large enough

$$n^{\frac{d}{d+2}}\mu(S_{h_n})\geqslant 2c_2$$

Thus

$$\frac{1}{2}\mu(S_{h_n})\geqslant\frac{c_2n^{\frac{2}{d+2}}}{n}\geqslant\frac{k_n}{n}\geqslant\frac{Z_n}{n} \tag{20}$$

Therefore by Hoeffding's inequality (see [4]), we get

$$P(\|X_{R_{k_n+1}}-x\|\geqslant h_n)\leqslant P(|\frac{Z_n}{n}-t_n|\geqslant\frac{1}{2}t_n)\leqslant 2\exp\{-\frac{nt_n}{10}\}\leqslant 2\exp\{-\frac{k_n}{5}\}\leqslant c\cdot n^{-\nu} \tag{21}$$

for n large enough. Where ν may be any positive real number.

Take $\nu=\frac{p}{d+2}$, by (18) and (21), we get

$$E(\tilde{J}_0)\leqslant ch_n^p+M_1(x)n^{-\frac{p}{d+2}}\leqslant M_1(x)\cdot n^{-\frac{p}{d+2}}\cdot a_n^p$$

when n is large enough. Then

$$I_1 \leqslant c \cdot M_1(x) \cdot n^{-\frac{p}{d+2}} \cdot a_n^p = O(n^{-\frac{p}{d+2}} \cdot a_n^p), \quad a.s.\, x(\mu). \tag{22}$$

By (22), (17) and (13), we get

$$I = O(n^{-\frac{p}{d+2}} \cdot a_n^p), \quad a.s.\, x(\mu),$$

Lemma 3 Let $\{X_n\}$ be a $r.v.$ sequence such that for any constant sequence with $\lim_{n\to\infty} c_n = 0$, we have $\limsup_{n\to\infty} |c_n X_n| < \infty$, $a.s.$. Then

$$\limsup_{n\to\infty} |X_n| < \infty, \quad a.s. \tag{23}$$

Proof See [5].

Proof of the theorem By Minkowski's inequality

$$J \triangleq \{E|m_n(x) - m(x)|^p\}^{\frac{1}{p}} \quad \{E|\sum_{i=1}^{k_n} C_{ni}[Y_{R_i} - m(X_{R_i})]|^p\}^{\frac{1}{p}} + \{E|\sum_{i+k_n+1}^{n} C_{ni}[Y_{R_i}$$

$$- m(X_{R_i})]|^p\}^{\frac{1}{p}} + \{E|\sum_{i=1}^{n} C_{ni}[m(X_{R_i}) - m(x)]|^p\}^{\frac{1}{p}} \triangleq J_1^{\frac{1}{p}} + J_2^{\frac{1}{p}} + J_3^{\frac{1}{p}}. \tag{24}$$

Since $\{C_{n_i}\}$ satisfies condition (I), by lemma 2 we get

$$J_3 = O(n^{-\frac{p}{d+2}} \cdot a_n^p), \quad a.s.\, x(\mu). \tag{25}$$

Now consider J_1. write

$$\tilde{C}_{ni} = \begin{cases} C_{ni}, & 1 \leqslant i \leqslant k_n, \\ 0, & k_n + 1 \leqslant i \leqslant n, \end{cases} \qquad \tilde{W}_{nR_i}(x) = \tilde{C}_{ni}, \quad i = 1, 2, \cdots, n.$$

It is obvious that $\sum_{i=1}^{n} \tilde{C}_{ni} \leqslant 1$. Write $Z_i = Y_i - m(X_i)$, $\Delta_n = (X_1, \cdots, X_n)$, $h(X_i) = E(|Z_i|^p | \Delta_n) = E(|Z_i|^p | X_i)$, then $Eh(X_i) = E|Z_i|^p \leqslant 2^p E|Y|^p < \infty$. Since $Z_1, Z_1, \cdots, Z_n$ are iid and $EZ_1 = 0$, by Marcinkiwicz's inequality[2], we get

$$E\{|\sum_{i=1}^{n} \tilde{C}_{ni}[Y_{R_i} - m(X_{R_i})]|^p | \Delta_n\} = E\{|\sum_{i=1}^{n} \tilde{W}_{ni}(X_i)[Y_i - m(X_i)]|^p | \Delta_n\}$$

$$\leqslant C_p \cdot E\{[\sum_{i=1}^{n} \tilde{W}_{ni}^2(x) Z_i^2]^{\frac{p}{2}} | \Delta_n\} \leqslant C_p \cdot \max_{1 \leqslant i \leqslant n} \{\tilde{W}_{ni}(x)\}^{\frac{p}{2}} \cdot E\{\sum_{i=1}^{n} \tilde{W}_{ni}(x)|Z_i|^p |$$

$$\Delta_n\} = C_p \cdot k_n^{-\frac{p}{2}} [\sum_{i=1}^{n} \tilde{W}_{ni}(x) h(X_i)] = C_p \cdot k_n^{-\frac{p}{2}} [\sum_{i=1}^{k_n} C_{ni} h(X_{R_i})]. \tag{26}$$

$$J_1 \leqslant C_p \cdot k_n^{-\frac{p}{2}} \cdot E\{\frac{1}{k_n} \sum_{i=1}^{k_n} h(X_{R_i})\} = C_p \cdot k_n^{-\frac{p}{2}} \cdot E\{E[\frac{1}{k_n} \sum_{i=1}^{k_n} h(X_{R_i}) \mid \|X_{R_{k_n+1}} - x\|]\}, \tag{27}$$

$$E\{\frac{1}{k_n} \sum_{i=1}^{k_n} h(X_{R_i}) \mid \|X_{R_{k_n+1}} - x\|\} = \int_{S_{\|X_{R_{k_n+1}} - x\|}} h(y)\mu(dy)/\mu(S_{\|X_{R_{k_n+1}} - x\|}) \leqslant h^*(x), \tag{28}$$

where $h^*(x) = \sup_{r>0} [\int_{S_r} |h(y)| \mu(dy)/\mu(S_r)]$. Similarily to (19) we obtain

$$0 < h^*(x) < \infty \qquad a.s.\, x(\mu) \tag{29}$$

Consequently,

$$J_1 \leqslant C_p \cdot k_n^{-\frac{p}{2}} \cdot h^*(x) = M_2(x) \cdot n^{-\frac{p}{d+2}} = O(n^{-\frac{p}{d+2}}), \quad a.s.\, x(\mu) \tag{30}$$

For the J_2 part, write

$$C_{ni}^* = \begin{cases} 0, & 1 \leqslant i \leqslant k_n \\ C_{ni}, & k_n + 1 \leqslant i \leqslant n \end{cases} \qquad W_{nR_i}^*(x) = C_{ni}^*, \quad i = 1, 2, \cdots, n,$$

We have $\sum_{i=1}^{n} C_{ni}^* \leqslant 1$. Introuduce Z_i, Δ_n and $h(X_i)$ as earlier. By an argument similar to those leading to (26), we get

$$E\left\{ \left| \sum_{i=1}^{n} W_{ni}^*(x) Z_i \right|^p \Big| \Delta_n \right\} \leqslant C_p \cdot n^{-\frac{p}{d+2}} \sum_{i=1}^{n} W_{ni}^*(x) h(X_i),$$

Ther efore

$$J_2 \leqslant c_p \cdot n^{-\frac{p}{d+2}} \cdot E\left\{ \sum_{i=1}^{n} W_{ni}^*(x) h(X_i) \right\} = c_p \cdot n^{-\frac{p}{d+2}} \cdot E\left\{ \sum_{i=k_n+1}^{n} C_{ni} h(X_{R_i}) \right\}$$

$$\leqslant c_p \cdot n^{-\frac{p}{d+2}} \cdot E\left\{ \frac{1}{n-k_n} \sum_{i=k_n+1}^{n} h(X_{R_i}) \right\} \triangleq c_p \cdot n^{-\frac{p}{d+2}} \cdot E(Q) \tag{31}$$

Similarily to the proof of (16), we write

$$A_x = \int h(y) \mu(dy) = E|Z|^p,$$

then $A_x \leqslant 2^p E|Y|^p < \infty$. Let $G(\rho) = \mu(S_\rho)$. Then the distribution of $\|X_{R_{k_n}} - x\|$ is (15). Therefore.

$$E(Q) = E\left\{ E\left[\frac{1}{n-k_n} \sum_{i=k_n+1}^{n} h(X_{R_i}) \,\Big|\, \|X_{R_{k_n}} - x\| \right] \right\} \leqslant \frac{nA_2}{n-k_n},$$

Thus, when n is large enough, we have

$$E(Q) \leqslant 2A_2 \leqslant c < \infty,$$

hence

$$J_2 \leqslant c \cdot n^{-\frac{p}{d+2}} = O(n^{-\frac{p}{d+2}}) \tag{32}$$

By (25)、(30) and (32), we get for n large enough

$$E|M_n(x) - m(x)|^p \leqslant (J_1^{\frac{1}{p}} + J_2^{\frac{1}{p}} + J_3^{\frac{1}{p}})^p \leqslant c_p(J_1 + J_2 + J_3) \leqslant M(x) \cdot n^{-\frac{p}{d+2}} a_n^p.$$

This can be written as

$$\limsup_{n \to \infty} \left| a_n^{-p} n^{\frac{d}{p+2}} E|m_n(x) - m(x)|^p \right| < \infty \quad a.s. \tag{33}$$

Here $\{a_n\}$ can be chosen as any constant sequence tending to ∞, evidently this is equivalent to say that $\{a_n^{-n}\}$ can be chosen as any constant sequence tending to zero. Hence by Lemma 3 we finally get

$$E|m_n(x) - m(x)|^p = O(n^{-\frac{p}{d+2}}). \quad a.s.x(\mu),$$

which ends the proof of The Theorem.

Acknowl edgment . The authors wish to express their deep gratitude to Professor Chen Xi-ru for his guidance.

References

〔1〕Stone, C·J·, Consistent nonparamtric regression, Ann.Statist., 5 (1977). No. 4, 595—645.

〔2〕Devroye, B. L., On the almost everywhere convergence of nonparametric regression function estimates, Ann. Statist., 9 (1981), No.6, 1310—1319.

〔3〕Wheeden, R.L. and Zygmund, A., Measure and Integral. Marcel Dekker, New York, 1977.

〔4〕Hoeffding, W., Probability inequalities for sums of bounded random variables, J. Amer. Statist. Assoc., **58**(1963), 13—30.

〔5〕Chen Xi-ru (陈希孺), The best convergence rates of kernel density function estimates, to appear in Chinese Annals of Mathematics.

（接124页）

征根 $\lambda_j(t)\leqslant-\delta(t)$，$\int_{t_0}^{+\infty}\delta(t)\mathrm{d}t=+\infty$，则系统（4）的零解渐近稳定。

注 若$D=E$，则得文（3）中相应定理。

本文是在斯力更付教授热情指导下完成的，深致谢意。

参 考 文 献

〔1〕秦元勋、王联、王慕秋，缓变系数动力系统的运动稳定性，中国科学，专辑（Ⅰ）、（1979），242—253；

〔2〕尤秉礼、邱元庆，关于驻定系统零解的稳定性问题，数学学报，14(1964)，6，769—780；

〔3〕许淞庆，常微分方程稳定性理论，上海科学技术出版社（1984），332—335。