Chin. Ann. of Math.
7B (2) 1986

THE OPTIMAL RATE OF CONVERGENCE OF ERROR FOR kNN MEDIAN REGRESSION ESTIMATES

CHEN XIRU (陈希孺)* ZHAO LINCHENG (赵林城)*

Abstract

Let (X, Y), (X_1, Y_1), $\cdots$, (X_n, Y_n) be iid. random vectors, where Y is one-dimensional. It is desired to estimate the conditional median $\xi(X)$ of Y, by use of $Z_n=\{(X_i, Y_i), i=1, \cdots, n\}$ and X. Denote by $\xi_{nk}(X, Z_n)$ the kNN estimate of $\xi(X)$, and put $H_{nk}(Z_n)=E\{|\xi_{nk}(X, Z_n)-\xi(X)||Z_n\}$, the conditional mean absolute error. This artical establishes the optimal convergence rate of $P(H_{nk_n}(Z_n)\geqslant\varepsilon)$, under fairly general assumptions on (X, Y) and k_n, which tends to ∞ in some suitable way.

§ 1. Introduction and Main Result

Let (X, Y), (X_1, Y_1), $\cdots$, (X_n, Y_n) be $R^d\times R^1$-valued iid. random vectors. Denote

$$Z_n=\{(X_1, Y_1), \cdots, (X_n, Y_n)\},\ X^n=\{X_1, \cdots, X_n\}. \tag{1}$$

Z_n is the sample of (X, Y). The conditional distribution function of Y given $X=x$ is denoted by $F(y|x)=P(Y<y|X=x)$.

Suppose that $R(x)$ is a quantity determined by $F(y|x)$, and we wish to estimate it, basing on the sample Z_n. The case that $R(x)=E(Y|X=x)$, the mean-value regression function, has been much studied in the literature. On considering the robustness, Zheng Zhongguo proposed in [1] the conditional median $\xi(x)$—— the median of $F(y|x)$, as the object of estimation. He introduced the k-Nearest Neighbor (kNN) estimate $\xi_{nk}(x, Z_n)$ of $\xi(x)$ as follows: Introduce a suitable distance $\|x-x'\|$ in R^d. Rearrange $X_1, \cdots, X_n$ according to their distances from x:

$$\|X_{n1}(x)-x\|\leqslant\|X_{n2}(x)-x\|\leqslant\cdots\leqslant\|X_{nn}(x)-x\|.$$

Call $\{X_{n1}(x), \cdots, X_{nk}(x)\}$ the kNN of x. Denote by $Y_{ni}(x)$ the Y-value corresponding to $X_{ni}(x)$, that is, $Y_{ni}(x)=Y_j$ if $X_{ni}(x)=X_j$. $\xi_{nk}(x, Z_n)$ is defined as the sample-median of $\{Y_{n1}(x), \cdots, Y_{nk}(x)\}$. Zheng studied in [1] the strong consistency and asymptotic normality of this estimate.

Manucript received January 25, 1984.

* The University of Science and Technology of China, Anhui, China.

The mean absolute error, and the conditional mean absolute error for given Z_n, of this estimate, are

$$H_{nk}=E|\xi_{nk}(X, Z_n)-\xi(X)|, \text{ and } H_{nk}(Z_n)=E\{|\xi_{nk}(X, Z_n)-\xi(X)||Z_n\}$$

respectively. From a practical point of view the latter is more sensible than the former. As pointed out by Wagner in [2], in practice Z_n is not always easilly available. On the contrary, Z_n is gathered during a period of time, and will be used repeatedly in problems with the same nature.

The purpose of this paper is to study the behavior of $H_{nk_n}(X, Z_n)$, as $n\to\infty$ and $k=k_n$ varies with n. A remarkable work of this kind of study is due to Beck [3]. In his work Beck used $\sum_{i=1}^{k} Y_{ni}(X)/k$ to estimate $E(Y|x)$. Denote by $G_{nk_n}(Z_n)$ the conditional mean absolute error $E\left\{\left|\sum_{i=1}^{k} Y_{ni}(x)/k-E(Y|X)\right| \,\middle|\, Z_n\right\}$. Beck established the following exponential rate of convergence

$$P(G_{nk_n}(Z_n)\geqslant\varepsilon)=O(e^{-cn})$$

under a number of conditions, among which the crucial one is the boundedness of Y. From a theoratical point of view the boundedness condition on Y is too restrictive. We shall tackle the problem under more reasonable condition imposed on Y.

Define $m(x, p)$ as the largest possible absolute value of the p-percentile of $F(y|x)$, $0<p<1$, and

$$G(x, \delta)=\sup\left\{m\left(x', \frac{1}{2}\pm\delta\right): \|x'\|\leqslant 2\|x\|\right\}, \quad 0<\delta<\frac{1}{2}.$$

The main result of this paper is as follows:

Theorem 1. *Suppose that* 1°. $F(y|x)$ *has a unique median* $\xi(x)$ *for any* x. 2°. $F(y|x')\xrightarrow{\mathscr{L}}F(y|x)$ *for* $x'\to x$. 3°. $E[G(X, \delta_0)]<\infty$ *for some* $\delta_0\in\left(0, \frac{1}{2}\right)$. 4°. *The distribution* Q *of* X *possesses a density* f, *and for* $a>0$ *small enough the set* $\{x: f(x)>a\}$ *differs from some open set by only a Lebesgue null-set. Also, suppose that*

$$k_n/n\to 0, \ \log n/k_n\to 0, \text{ for } n\to\infty. \tag{2}$$

Then for any $\varepsilon>0$ *there exists* $C>0$ *(depending on* ε *but not* n*) such that*

$$P(H_{nk_n}(Z_n)\geqslant\varepsilon)=O(e^{-Ck_n}). \tag{3}$$

This rate cannot be improved further: For any given $\{k_n\}$ with $k_n\to\infty$ and $k_n/n\to 0$, one can find (X, Y) whose distribution satisfies the four conditions of Theorem 1, but for any $\varepsilon>0$ and $C>\log 2$ it is true that

$$\limsup_{n\to\infty}\{e^{Ck_n}P(H_{nk_n}(Z_n)\geqslant\varepsilon)\}\geqslant 1. \tag{4}$$

Condition 3° is a mild restriction on Y. In order that $H_{nk_n}(Z_n)$ is meaningful, we must have $E|\xi(X)|<\infty$. Condition 3° is stronger than this. Hence there is a question: Can condition 3° be replaced by the weaker one $E|\xi(X)|<\infty$? We shall

show by an example that this is impossible: The condition $E|\xi(X)|<\infty$ is not enough even for the far-weaker assertion that

$$\lim_{n\to\infty} P(H_{nk_n}(Z_n)\geqslant\varepsilon)=0.$$

Since in (3) k_n may tend to ∞ with any rate slower than n, the rate $O(e^{-ck_n})$ given by Theorem 1 can be made arbitrarily near $O(e^{-cn})$ but cannot reach it. One naturally asks: Whether or not there exists (X, Y) with Y unbounded, such that $P(H_{nk_n}(Z_n)\geqslant\varepsilon)=O(e^{-cn})$ for some $\{k_n\}$, with $k_n\to\infty$ and $k_n/n\to 0$? The question is interesting but the probable answer is not apparent to guess.

We shall give the proof of Theorem 1 in § 2, and then make some remarks related to this theorem.

§ 2. Proof of Theorem 1

(I) Some preliminary facts.

1. Hoeffding inequality ([4]). Suppose that the random variable ξ obeys the Binomial law $B(n, p)$, then for any $\varepsilon>0$ we have

$$P\left(\left|\frac{\xi}{n}-p\right|\geqslant\varepsilon\right)\leqslant 2\exp(-n\varepsilon^2/(2p+\varepsilon)).$$

From this the following corollary follows easily: Let $A_1, \cdots, A_n$ be n independent events,

$$\min_{1\leqslant i\leqslant n} P(A_i)=p^*>p\geqslant 0.$$

Put

$$\xi=\sum_{i=1}^{n} I_{A_i},$$

then

$$P(\xi\leqslant np)\leqslant 2\exp\{-n(p^*-p)^2/(3p^*-p)\}. \tag{5}$$

2. Under conditions 1°, 2° of Theorem 1, the median $\xi(x)$ of $F(y|x)$ is a continuous function of x.

The proof is easy and therefore omitted.

3. For $\varepsilon>0$, define

$$f_{1\varepsilon}(x)=P(Y\leqslant\xi(x)-\varepsilon|x),\ f_{2\varepsilon}(x)=P(Y\geqslant\xi(x)+\varepsilon|x). \tag{6}$$

Suppose that conditions 1°, 2° of Theorem 1 hold. Then for any constant b, the sets

$$\{x: f_{1\varepsilon}(x)<b\},\ \{x: f_{2\varepsilon}(x)<b\}$$

are both open.

Proof Evidently one only has to show that

$$x_n\to x\Rightarrow\limsup_{n\to\infty} f_{i\varepsilon}(x_n)\leqslant f_{i\varepsilon}(x),\ i=1, 2. \tag{7}$$

Given $\varepsilon>0$, find $\delta>0$ sufficiently small, such that $\xi(x)\mp\varepsilon\pm\delta$ are continuity points of the distribution $F(y|x)$. Since $\xi(x)$ is continuous, we have

$$|\xi(x_n)-\xi(x)|<\delta$$

for large n. Hence for large n we have

$$\xi(x_n)-\varepsilon<\xi(x)-\varepsilon+\delta,\ \xi(x_n)+\varepsilon>\xi(x)+\varepsilon-\delta.$$

Therefore we have for large n,

$$f_{1\varepsilon}(x_n)\leqslant F(\xi(x)-\varepsilon+\delta|x_n),\ f_{2\varepsilon}(x_n)\leqslant 1-F(\xi(x)+\varepsilon-\delta|x_n). \tag{8}$$

Since $\xi(x)\mp\varepsilon\pm\delta$ are continuity points, from condition 2° and (8), we have

$$\limsup_{n\to\infty} f_{1\varepsilon}(x_n)\leqslant\lim_{n\to\infty} F(\xi(x)-\varepsilon+\delta|x_n)=F(\xi(x)-\varepsilon+\delta|x).$$

Setting $\delta\downarrow 0$, we get $\limsup_{n\to\infty} f_{1\varepsilon}(x_n)\leqslant f_{1\varepsilon}(x)$. The case $i=2$ in (7) can be handled in a similar fashion.

4. Under condition 2° of Theorem 1, we have

$$\sup\{|m(x,p)|:\|x\|\leqslant r\}<\infty$$

for any $p\in(0,1)$ and $r<\infty$.

5. There exist $\eta>0$ and r sufficiently large such that $Q(S_{x,\|x\|})\geqslant\eta$ for any x belonging to the support of Q and $\|x\|\geqslant r$, where $S_{a,\rho}$ is the sphere with radius ρ and centered at a.

These two facts are easily proved and details omitted.

(II) Since $\xi(x)$ is unique, we have

$$f_{i,\varepsilon/6}(x)<\frac{1}{2}$$

for all x and $i=1, 2$. Choose $b<\frac{1}{2}$ so that $Q(B_1)>1-\varepsilon_1\varepsilon$, where $B_1=\{x: f_{i,\varepsilon/6}(x)<b,\ i=1,2\}$, ε_1 is a positive number to be specified. It follows from (I) 3 that B_1 is open. By condition 4° of Theorem 1 one can find $a>0$ such that $\{x: f(x)>a\}$ differs from some open set B_2 by only a Lebesgue null-set, and $Q(B_2)>1-\varepsilon_1\varepsilon$. So there is an open set $B_3\subset B_1\cap B_2$ such that $Q(B_3)>1-3\varepsilon_1\varepsilon$, and $\xi(x)$ is continuous uniformly on B_3.

For simplicity we shall call the set

$$\{x=(x_1,\cdots,x_d): a_i\leqslant x_i\leqslant a_i+h,\ i=1,\cdots,d\}$$

a regular supercube with size h. Find regular supercubes $V_1,\cdots V_N$ with the same size such that $V_i\subset B_3$, $V_i\cap V_j=\emptyset$ for $i, j=1,\cdots,N$, $i\neq j$, and

$$\sum_{i=1}^{N} Q(V_i)>1-4\varepsilon_1\varepsilon,$$

$$\sup\{|\xi(x)-\xi(x')|: x\in V_i,\ x'\in V_i\}<\varepsilon/6,\ i=1,\cdots,N.$$

Denote by $\rho(S_1, S_2)$ the distance of two sets S_1 and S_2 in R^d, and by $D(S)$ the diameter of set S. Find regular supercubes $V_1^*,\cdots, V_N^*$ with the same size q^* such that $V_i^*\subset V_i^0$ (the inner of V_i), $i=1,\cdots,N$, and

$$\sum_{i=1}^{N} Q(V_1^*)>1-5\varepsilon_1\varepsilon. \tag{9}$$

We have $t \triangleq \min_{1\leqslant i\leqslant N} \rho(V_i^*, V_i - V_i^0) > 0$. Denote by $[a]$ the largest integer not exceeding a, and

$$c_n = [q^*(an/(2k_n))^{1/d}],\ b_n = q^*/c_n.$$

From $k_n/n \to 0$ it follows that $b_n \to 0$.

For each i ($i=1, \cdots, N$) there exists a family of regular supercubes $H_i = \{H_{i1}, H_{i2}, \cdots\}$ such that 1°. Each H_{iu} has a size b_n. 2°. $H_{iu}^0 \cap H_{iv}^0 = \emptyset$ for $u \neq v$. 3°.

$$R^d = \bigcup_{j=1}^{\infty} H_{ij}.$$

4°. There is a subset of H_i, to be denoted by $\mathscr{G}_{ni}^* = \{G_{i1}^*, \cdots, G_{il_n}^*\}$ ($l_n = c_n^d$), such that

$$V_i^* = \bigcup_{j=1}^{l_n} G_{ij}^*.$$

Write $\mathscr{G}_n^* = \bigcup_{i=1}^{N} \mathscr{G}_{ni}^*$. Similarly, the set $\{H_{iu}: H_{iu} \subset V_i,\ u=1, 2, \cdots\}$ will be denoted by $\mathscr{G}_{ni} = \{G_{i1}, G_{i2}, \cdots\}$, and $\mathscr{G}_n = \bigcup_{i=1}^{N} \mathscr{G}_{ni}$. Evidently one has $\mathscr{G}_{ni} \supset \mathscr{G}_{ni}^*$ and $\mathscr{G}_n \supset \mathscr{G}_{nj}^*$ Also, if the size of V_i is q, then the number of elements in $\mathscr{G}_{ni}$ does not exceed

$$(q/b_n)^d = (q/q^*)^d c_n^d \leqslant \frac{1}{2} aq^d n/k_n.$$

Hence, the number of elements in $\mathscr{G}_n$ does not exceed $\frac{a}{2} Nq^d n/k_n$. Note that $\frac{1}{2} aNq^d$ does not depend on n.

Let $\mathscr{F}$ be a family of point sets in R^d. For simplicity, we shall denote by $(\mathscr{F})$ the union of all sets contained in $\mathscr{F}$. Also, for a set $B \subset R^d$, the number of X_i's ($i=1, \cdots, n$) contained in B will be denoted by $\#(B)$. Now define

$$\hat{\mathscr{G}}_{ni}^* = \{G: G \in \hat{\mathscr{G}}_{ni}^*,\ \#(G) < k_n\},\ i=1, \cdots, N;$$

$$\hat{\mathscr{G}}_n^* = \bigcup_{i=1}^{N} \hat{\mathscr{G}}_{ni}^*.$$

For $G \in \mathscr{G}_{ni}^* - \hat{\mathscr{G}}_{ni}^*$, define

$$W(G) = \cup \{H: H \in H_i,\ \rho(G, H) \leqslant D(G)\}.$$

From the definition of $\hat{\mathscr{G}}_{ni}^*$ it follows that if $x \in G$ ($G \in \mathscr{G}_{ni}^* - \hat{\mathscr{G}}_{ni}^*$), then $X_{ni}(x) \in W(G)$ for $i=1, \cdots, k_n$. Also, when n is sufficiently large such that $b_n < t/d$, then $W(G) \subset (\mathscr{G}_{ni})$. It is obvious that there exists constant m_d depending only upon d, such that for $G \in \mathscr{G}_{ni}^* - \hat{\mathscr{G}}_{ni}^*$, $W(G)$ can be expressed as the union of less than m_d regular supercubes in H_i.

Since the volume of G_{ij} is $b_n^d \geqslant 2k_n/(an)$ and $f(x) > a$ for $x \in B_3$, a.e., we have

$$Q(G_{ij}) \geqslant 2k_n/n. \tag{10}$$

By Hoeffding inequality one finds easily that

$$P(\#(G_{ij}) < k_n) \leqslant 2\exp(-k_n/5). \tag{11}$$

Put

$$S_n = \{\#(G_{ij}) \geqslant k_n, \text{ for all } G_{ij} \in \mathscr{G}_n\}.$$

By (11), and noticing that the number of elements in $\mathscr{G}_n$ does not exceed $\frac{1}{2}aNq^d n/k_n$, one finds that

$$P(S_n)\geqslant 1-aNq^d nk_n^{-1}\exp(-k_n/5). \tag{12}$$

Since $\log n/k_n\to 0$, we have $n=o(\exp(k_n/15))$. Therefore it follows from (12) that there exists constant $C>0$ not depending on n, such that.

$$P(S_n)\geqslant 1-\exp(-Ck_n) \tag{13}$$

For simplicity the symbol C will be employed to denote any positive constant not depending on n. C can assume diffrent values in each of its appearance, even within the same expression.

(III) Now we proceed to prove the first half of Theorem 1. we have

$$\begin{aligned}&P(H_{nk_n}(Z_n)\geqslant\varepsilon)\\&\leqslant P\left\{E[|\xi_{nk_n}(X,Z_n)-\xi(X)|I_{(\mathscr{G}_n^*-\hat{\mathscr{G}}_n^*)}(X)|Z_n]\geqslant\frac{1}{2}\varepsilon\right\}\\&\quad+P\left\{E[|\xi_{nk_n}(X,Z_n)-\xi(X)|I_{(\mathscr{G}_n^*-\hat{\mathscr{G}}_n^*)^c}(X)|Z_n]\geqslant\frac{1}{2}\varepsilon\right\}\\&\triangleq J_{1n}+J_{2n}.\end{aligned} \tag{14}$$

Take J_{2n} first. We have

$$\begin{aligned}J_{2n}&\leqslant P\left\{E[|\xi_{nkn}(X,Z_n)|I_{(\mathscr{G}_n^*-\hat{\mathscr{G}}_n^*)^c}(X)|Z_n]\geqslant\frac{1}{4}\varepsilon\right\}\\&\quad+P\left\{E[|\xi(X)|I_{(\mathscr{G}_n^*-\hat{\mathscr{G}}_n^*)^c}(X)|Z_n]\geqslant\frac{1}{4}\varepsilon\right\}\triangleq J'_{2n}+J''_{2n}.\end{aligned} \tag{15}$$

It follows easily from condition 3° of Theorem 1 that $E|\xi(X)|<\infty$. Find M sufficiently large such that

$$\int_{(x:|\xi(x)|>M)}|\xi(x)|dQ(x)<\frac{\varepsilon}{8}.$$

Then we have

$$J''_{2n}\leqslant P\left\{Q((\mathscr{G}_n^*-\hat{\mathscr{G}}_n^*)^c)\geqslant\frac{\varepsilon}{8M}\right\}. \tag{16}$$

Choose $\varepsilon_1\in\left(0,\ \frac{1}{40M}\right)$(see(9)). Since from (9) we have

$$Q((\mathscr{G}_n^*-\hat{\mathscr{G}}_n^*)^c)=Q((\hat{\mathscr{G}}_n^*))+1-Q((\mathscr{G}_n^*))\leqslant Q((\hat{\mathscr{G}}_n^*))+5\varepsilon_1\varepsilon,$$

it follows from (16) and the choice of ε_1 that

$$J''_{2n}\leqslant P(Q((\hat{\mathscr{G}}_n^*))>0)\leqslant P(\hat{\mathscr{G}}_n^*\neq\phi). \tag{17}$$

Since $\{\hat{\mathscr{G}}_n^*\neq\phi\}\subset S_n^c$, by (13), (17), we have

$$J''_{2n}\leqslant P(R_n^c)\leqslant\exp(-Ck_n). \tag{18}$$

To deal with J'_{2n}, define a set T_n containing all points Z_n satisfying the follewing condition: "For each sphere $S\subset R^d$ such that $\#(S)\geqslant k_n$, if

$$S\cap\{X_1,\cdots,X_n\}=\{X_{i_1},\cdots,X_{i_l}\},$$

then the number of elements in $\{Y_{i_1},\cdots,Y_{i_l}\}$ satisfying the inequality

$$Y_{i_j}\leqslant m\left(X_{i_j},\frac{1}{2}+\delta_0\right)$$

is not smaller than $\frac{1}{2}l(1+\delta_0)$, the number of elements in $\{Y_{i_1}, \cdots, Y_{i_l}\}$ satisfying the inequality $Y_{i_j} \geqslant -m\left(X_{i_j}, \frac{1}{2}-\delta_0\right)$ is also not smaller than $\frac{1}{2}l(1+\delta_0)$", where δ_0 is the number appearing in condition 3° of Theorem 1. Note that the number of diffrent sets formed by $S \cap \{X_1, \cdots, X_n\}$ with all possible sphere S, does not exceed n^{2d}.

By the definition of $m(x, p)$ and use (5), it is not difficult to show that there exists constant $C>0$ not depending on $X^n=\{X_1, \cdots, X_n\}$, such that under the condition of given X_n, if a sphere S satisfies $\#(S) \geqslant k_n$, then the conditional probability that S fulfils the requirement specified in the definition of T_n, is not smaller than $1-\exp(-Ck_n)$. Considering this and the remark made in the end of the last paragragh, we get

$$P(T_n | X^n) \geqslant (1-e^{-Ck_n})^{n^{2d}} \geqslant 1-n^{2d}e^{-Ck_n}.$$

Hence $P(T_n) \geqslant 1-n^{2d}e^{-Ck_n}$. Since $\log n/k_n \to 0$, we get

$$P(T_n) \geqslant 1-\exp(-Ck_n). \tag{19}$$

(Notice the meaning of the symbol C explained earlier).

According to (I) 5 and $k_n/n \to 0$, it is easily seen from the Hoeffding inequality that

$$P(T_{nr}^c) < \exp(-Cn), \text{ for } n \text{ sufficiently large,} \tag{20}$$

where T_{nr} denotes the event "$\#(S_{x, \|x\|}) \geqslant k_n$, for any x belonging to the support of Q with $\|x\| \geqslant r$".

Now we have

$$J'_{2n} \leqslant P(T_n^c)+P(T_{nr}^c)+P\left(T_n \cap T_{nr} \cap \left\{E[|\xi_{nk_n}(X, Z_n)| I_{(\mathscr{G}_n^* - \hat{\mathscr{G}}_n^*)^c}(X) | Z_n] \geqslant \frac{\varepsilon}{4}\right\}\right). \tag{21}$$

By the definition of T_n, T_{nr}, $m(x, p)$ and $G(x, \delta)$, one has for $Z_n \in T_n \cap T_{nr}$ that

$$|\xi_{nk_n}(x, Z_n)| \leqslant G(x, \delta_0), \ \|x\| > r \text{ and } n \text{ large.} \tag{22}$$

Further, from (I) 4, it follows that there exists constant M_1 such that

$$|\xi_{nk_n}(x, Z_n)| \leqslant M_1, \ \|x\| \leqslant r, \ Z_n \in T_n \cap T_{nr} \text{ and } n \text{ large.} \tag{23}$$

From (22), (23) and noticing that

$$P((\mathscr{G}_n^* - \hat{\mathscr{G}}_n^*)^c) \leqslant 5\varepsilon_1\varepsilon + P((\hat{\mathscr{G}}_n^*)),$$

we get

$$E[|\xi_{nk_n}(X, Z_n)| I_{(\mathscr{G}_n^* - \hat{\mathscr{G}}_n^*)^c}(X) | Z_n] \leqslant \int_{\|x\|>r} G(x, \delta_0) dQ(x) + 5M_1\varepsilon_1\varepsilon$$

when $\hat{\mathscr{G}}_n^* = \phi$, $Z_n \in T_n \cap T_{nr}$ and n large. By condition 3° of Theorem 1, choose r sufficiently large, we can make the integral on the right hand side not exceed $\varepsilon/10$. Fix this r and hence M_1 is also fixed, take $\varepsilon_1 \in \left(0, \frac{1}{50M_1}\right)$, we get under the above conditions ($\hat{\mathscr{G}}_n^* = \phi$, $Z_n \in T_n \cap T_{nr}$ and n large) that

$$E[|\xi_{nk_n}(X, Z_n)|I_{(\mathscr{G}_n^*-\hat{\mathscr{G}}_n^*)^c}(X)|Z_n] \leqslant \varepsilon/5 < \varepsilon/4.$$

Therefore, on noticing that $\{\hat{\mathscr{G}}_n^* \neq \phi\} \subset S_n^c$, and (13), (19)—(21), we obtain

$$J'_{2n} \leqslant P(T_n^c) + P(T_{nr}^c) + P(S_n^c) \leqslant \exp(-Ck_n). \tag{24}$$

Now turn to J_{1n}. Define a set $T_n^{(1)}$ formed by all Z_n satisfying the following condition: "For each sphere $S \subset R^d$ such that $\#(S) \geqslant k_n$, if

$$S \cap \{X_1, \cdots, X_n\} = \{X_{i_1}, \cdots, X_{i_l}\},$$

then the number of elements in $\{Y_{i_1}, \cdots, Y_{i_l}\|$ satisfying the inequality $Y_{i_j} \geqslant \xi(X_{i_j}) - \varepsilon/6$ is larger than $l/2$, and the number of elements in $\{Y_{i_1}, \cdots, Y_{i_l}\}$ satisfying the inequality $Y_{i_j} \leqslant \xi(X_{i_j}) + \varepsilon/6$ is also large than $l/2$". Considering the choice of B_1 in (II), the definition of $f_{i,\varepsilon/6}(x)$, the fact that in the definition of B_1 we haue $b < \frac{1}{2}$, the definition of V_i, V_i^*, $W(G)$, $\mathscr{G}_n^*$, and the fact that

$$\{X_{n1}(x), \cdots, X_{nk}(x)\} \subset W(G_{ij}^*) \subset V_i$$

for $x \in G_{ij}^* \in \mathscr{G}_n^* - \hat{\mathscr{G}}_n^*$, we can show that

$$P(T_n^{(1)}) \geqslant 1 - \exp(-Ck_n)$$

by an argument similar to that employed in dealing with T_n. Hence, if we put

$$T_n^{(2)} = \{Z_n\text{: for each } x \in \mathscr{G}_n^* - \hat{\mathscr{G}}_n^*,\ x \in V_i \text{ for some } i = 1, \cdots, n,$$

we have $\inf_{x' \in V_i} \xi(x') - \frac{\varepsilon}{6} \leqslant \xi_{nk_n}(x, Z_n) \leqslant \sup_{x' \in V_i} \xi(x') + \frac{\varepsilon}{6}\}$,

we shall have $T_n^{(2)} \supset T_n^{(1)}$, and therefore

$$P(T_n^{(2)}) \geqslant P(T_n^{(1)}) \geqslant 1 - \exp(-Ck_n). \tag{25}$$

Since $\sup\{|\xi(x) - \xi(x')|\text{: } x \in V_i,\ x' \in V_i\} < \varepsilon/6$, for $Z_n \in T_n^{(2)}$ and $x \in \mathscr{G}_n^* - \hat{\mathscr{G}}_n^*$, we have

$$|\xi(x) - \xi_{nk_n}(x, Z_n)| \leqslant \varepsilon/3.$$

Hence for $Z_n \in T_n^{(2)}$ we have

$$E[|\xi_{nk_n}(X, Z_n) - \xi(X)|I_{(\mathscr{G}_n^* - \hat{\mathscr{G}}_n^*)}(X)|Z_n] \leqslant \varepsilon/3.$$

From this and (25) we get

$$J_{1n} \leqslant 1 - P(T_n^{(2)}) \leqslant \exp(-Ck_n). \tag{26}$$

Finally, summing up (14), (18), (24) and (26), we get (3). This concludes the proof of the first half of the theorem.

(IV). To prove the remaining part of Theorem 1, suppose that a sequence of positive integers $\{k_n\}$ is given to satisfy $k_n \to \infty$, $k_n/n \to 0$. Take a sequence of positive integers $n_1 < n_2 < \cdots$ such that $\sum_{i=1}^{\infty} k_{n_i}/n_i < 1$. Choose a such that

$$\sum_{i=1}^{\infty} k_{n_i}/n_i < \frac{1}{a} < 1,$$

$a < \sqrt{2}$. Define a one-dimensional density function f satisfying the following conditions:

$$1.\ \int_{2i}^{2i+1} f(x)dx = ak_{n_i}/n_i,\ i = 1, 2, \cdots, \tag{27}$$

2. $\int_{-\infty}^{0} f(x)dx = 1 - a\sum_{i=1}^{\infty} k_{n_i}/n_i$,

3. f is everywhere continuous on $(-\infty, \infty)$.

Now take $d=1$, and define a two-dimensional random vector (X, Y) such that X has a marginal density f, and the conditional distribution $F(y|x)$ of Y given $X=x$ satisfies the following four conditions: (a). $F(y|x)$, as a function of (x, y), is continuous everywhere on R^2. (b). $F(y|x)$, as a function of y for fixed x, is strictlyincreasing. (c). For all x we have

$$F(1|x)=\frac{a}{2},\ F(0|x)=\frac{1}{2},\ F(-1|x)=1-\frac{a}{2}.$$

(d).

$$F\left(\frac{in_i}{k_{n_i}}|x\right)=\frac{a^2}{2}$$

for $x\in(2i,\ 2i+1)$, $i=1,\ 2,\ \cdots$. It is easily seen that such a $(X,\ Y)$ satisfies all conditions $1^\circ\sim4^\circ$ of Theorem 1.

Denote again by $Z_n=\{(X_i, Y_i), i=1, \cdots, n\}$ the random sample drawn from (X, Y), $X^n=\{X_1, \cdots, X_n\}$. Define the event

$$B_i=\{a^2k_{n_i}\leqslant\#\{(2i,\ 2i+1)\}\leqslant a^2k_{n_i}\},\ i=1,\ 2,\ \cdots$$

(Remember that $\#(G)$ is the number of elements of the set $G\cap X^n$). Noticing (27) and employing the Hoeffding inequality, one sees easily that

$$\varliminf_{i\to\infty} P(B_i)=1. \tag{28}$$

Denote the elements of the set $X^{n_i}\cap(2i,\ 2i+1)$ by $X_{i1}, \cdots, X_{iN_i}$, and their corresponding Y-values are denoted by $Y_{i1}, \cdots, Y_{iN_i}$. Put

$$K_i=\{Y_{ij}\geqslant in_i/k_{n_i},\ j=1, \cdots, N_i\},$$
$$G_i=B_i\cap K_i,\ i=1,\ 2,\ \cdots.$$

Then by condition d of $F(y|x)$, we see that for $X^{n_i}\in B_i$:

$$P(K_i|X^{n_i})=\left(\frac{1}{2}a^2\right)^{n_i}\geqslant\left(\frac{1}{2}a^2\right)^{a^2k_{n_i}}\triangleq C_a^{-k_{n_i}}, \tag{29}$$

where $C_a=\left(\frac{1}{2}a^2\right)^{-a^2}$.

On the other hand, by the definition of kNN and the three conditions satisfied by the deneity f, it is easily seen that $\{X_{n_ik_{n_i}}(x), \cdots, X_{n_ik_{n_i}}(x)\}\subset(2i,\ 2i+1)$ and $\xi_{n_ik_{n_i}}(x, Z_{n_i})\geqslant in_i/k_{n_i}$ for $x\in(2i,\ 2i+1)$ and $Z_{n_i}\in G_i$. Hence

$$H_{n_ik_{n_i}}(Z_{n_i})\geqslant\int_{2i}^{2i+1}|\xi_{n_ik_{n_i}}(x,\ Z_{n_i})-0|f(x)dx\geqslant ia>i$$

for $Z_{n_i}\in G_i$. From this and (28), (29), also noticing that $P(G_i)\geqslant C_a^{-k_{n_i}}P(B_i)$, we get

$$\varlimsup_{i\to\infty}\{C_a^{k_{n_i}}P(H_{n_ik_{n_i}}(Z_{n_i})\geqslant\varepsilon)\}\geqslant\varlimsup_{i\to\infty}\{C_a^{k_{n_i}}P(G_i)\}\geqslant\varliminf_{i\to\infty}P(B_i)=1. \tag{30}$$

Since a can be chosen arbitrarily near $\sqrt{2}$, (4) follows from (30). This proves the latter conclusion of Theorem 1. The proof is completed.

§ 3. Some Remarks

1. If the condition 3° of Theorem 1 is replaced by the weaker one

$$E|\xi(X)|<\infty \tag{31}$$

and maintain the other conditions, then the probability $P(H_{nk_n}(Z_n)\geqslant\varepsilon)$ may not tend to zero.

Choose the (X, Y) as in § 2 (IV), but modify the conditions c and d imposed on $F(y|x)$ to: "$F(0|x)=\frac{1}{2}$, $F(in_i/k_{n_i}|x)-F(in_i/(2k_{n_i})|x)+F(-in_i/(4k_{n_i})|x)-F(-in_i/(8k_{n_i})|x)=1-k_{n_i}^{-2}$ for $x\in(2i, 2i+1)$, $i=1, 2, \cdots$." Then the condition (31) and the conditions 1°, 2°, 4° of Theorem 1 are satisfied. But an argument similar to those employed in § 2 (IV) yields

$$\lim_{i\to\infty} P(H_{n_ik_{n_i}}(Z_{n_i})\geqslant\varepsilon)=1, \text{ for any } \varepsilon>0.$$

2. In case Y is bounded, we have the following

Theorem 2. *Maintain the conditions* 1°, 2° *of Theorem* 1, *and modify the conditions* 3°, 4° *as follows*: 3′. *Y is bounded.* 4′. *For sufficiently small $a>0$ and sufficiently large A, the set $\{x: a<f(x)<A\}$ differs from an open set by only a Lebesgue nullset. Suppose that $k_n\to\infty$, $k_n/n\to 0$. Then for any given $\varepsilon>0$ there exists $C>0$ (depending on ε) such that*

$$P(H_{nk_n}(Z_n)\geqslant\varepsilon)=0(e^{-cn}).$$

This can be proved by combining the methods of Beck [3] and the present paper. Details are not presented here.

3. Even if Y is bounded, one cannot establish in general a rate faster than the type of $O(e^{-cn})$.

Example Take $d=1$. Define (X, Y), whose distribution $F(x, y)$ is as follows: X has a marginal density $(1-|x|)I_{(-1,1)}(x)$, and the conditional distribution $F(y|x)$ is $R(x-1, x+1)$. Here we have $|Y|\leqslant 2$. Noticing that

$$P(Y>3/2)=\int_{\frac{1}{2}}^{2}\left(x-\frac{1}{2}\right)(1-x)dx=\frac{1}{48}.$$

One gets $P(Y_i>3/2, i=1, \cdots, n)=48^{-n}$. But $|\xi(x)|\leqslant 1$. Hence

$$P\left(H_{nk_n}(Z_n)\geqslant\frac{1}{2}\right)\geqslant 48^{-n}=e^{-cn}, \quad C=\log 48.$$

References

[1] Zheng Zhongguo, Asymptotic properties of NN estimates of conditional median (in Chinese, to appear).
[2] Wagner, T. J., *IEEE Trans. Inform. Theory*, IT 17, 1971, 566.
[3] Beck, J., Problems of Control and Information Theory, Vol. 8, 1979, 303.
[4] Hoeffding, W., *JASA*, 1963, 13.