Chin. Ann. of Math.
6B (3) 1985

THE UNIFORM CONVERGENCE RATE OF KERNEL DENSITY ESTIMATE

YANG ZHENHAI(杨振海)*

Abstract

In this paper, we study the uniform convergence rate of kernel density estimate $\hat{f}_n$ and get optimal uniform rate of convergence without the assumption of compact support for kernel function. It is proved that if the density function f satisfies λ-condition and the kernel function K is λ-good (see section 1), then we have

$$\limsup_{n\to\infty}\left(\frac{n}{\log n}\right)^{\lambda/(1+2\lambda)}\sup_{x\in R^1}|\hat{f}_n(x)-f(x)|\leqslant\text{const. } a.s.$$

§1. Introduction

Let $X_1, \cdots, X_n$ be a random sample drawn from a population with distribution function F and probability density function f (p. d. f.). Denote the empirical distribution function of $X_1, X_2, \cdots, X_n$ by F_n. The kernel estimate of f is difined by

$$\hat{f}_n(x)=\frac{1}{nh_n}\sum_{v=1}^{n}K\left(\frac{x-X_v}{h_n}\right)=\frac{1}{h_n}\int_{-\infty}^{\infty}K\left(\frac{x-y}{h_n}\right)dF_n(y), \tag{1}$$

where $h_n>0$ is a constant depending on n and the kernel K is a p. d. f.

This method was proposed by Rosenblatt [1956] and has aroused the interest of many authors. Since then statisticians have been concerned with the problem of uniform convergence of $\hat{f}_n$ to f. The best result was given by Devroye and Wagner [1980]. In the recent years, the problem about uniform strong convergence rate has attracted much attention of statisticians. Under certain conditions imposed on the density f and kernel K, Schuster [1969] proved that

$$\sup_{x\in R^1}|\hat{f}_n(x)-f(x)|=O(n^{-\frac{1}{4}+\varepsilon})\qquad \text{a. s.}$$

for sufficiently large n. Under the condition that the r-th order derivative of f is uniform bounded on R^1, Singh [1976] proved that, with a suitable choice of K and h_n, we have

$$\sup_{x\in R^1}|\hat{f}_n(x)-f(x)|=O(n^{-r/(2+2r)}\sqrt{\log\log n})\,\text{a.s.}$$

for sufficiently large n. For an m-variate p. d. f. f, Susarla [1981] proved that if

Manuscript received July 12, 1983.
* Department of Mathematics, Beijing Polytechnic University, Beijing, China.

all the first order partial derivatives of f are uniformly bounded, then

$$\sup_{x\in R^m}|\hat{f}_n(x)-f(x)|=O(n^{-1/(2m+2)}\log\log n) \qquad \text{a. s.}$$

for sufficiently large n.

If we suppose that f satisfies δ-order Lipschitz condition $0<\delta\leqslant 1$ (for $\delta=1$ is the case above), the convergence rate can reach

$$\sup_{x\in R^m}|\hat{f}_n(x)-f(x)|=O(n^{-\delta/(2m+2\delta)}\sqrt{\log\log n}) \qquad \text{a. s.}$$

Lately, Chen Guijing and Zhao Lincheng have studied the same problem. They improved the convergence rate to

$$\limsup_{n\to\infty}\left(\frac{n}{\log n}\right)^{\lambda/(1+2\lambda)}\sup_{x\in R}|\hat{f}_n(x)-f(x)|\leqslant \text{const.} \qquad \text{a. s.} \tag{2}$$

for sufficiently large n under the condition that f satisfies λ-condition (refer to the following definition) and kernel has compact support. In this paper, using a new method, we get the same convergence rate as (2) under weaker condition. Precisely, we omit the assumption that K possesses a compact support.

First, we give two definitions.

We say that p. d. f. $f(x)$ satisfies λ-condition if there exists a real number λ, $0<\lambda\leqslant 2$ such that

$$\sup_x\left|\int_{x-a}^{x+a}f(t)dt-2af(x)\right|\leqslant Ra^{\lambda+1}$$

holds for all $a>0$, where R is a constant.

It is not difficult to verify that if this satisfies δ^{th} Lipschitz condition, then $f(x)$ satisfies the λ-condition with $\lambda=\delta$; if $f'(x)$ satisfies δ^{th} Lipschitz condition, then $f(x)$ satisfies the λ-condition with $\lambda=1+\delta$. We say that the kernel function $K(x)$ is λ-good if it satisfies the following conditions:

(i) $K(x)$ is a symmetric p. d. f. with

$$\int_{-\infty}^{\infty}|x|^{\lambda}K(x)dx<+\infty.$$

(ii) $K(x)$ is bounded and strictly decreasing on $[0, \infty)$ or when $K(x)$ possesses compact support S, it is strictly decreasing on $S\cap[0, \infty)$.

We state the theorems whose proofs are given in section 2 and 3, respectively.

Theorem 1. *Suppose that $K(x)$ is λ-good and $f(x)$ satisfies λ-condition and*

$$h_n/(n^{-1}\log n)^{\frac{1}{1+2\lambda}}$$

has a finite positive limit as $n\to\infty$. Then we have

$$\sup_{x\in R^1}|\hat{f}_n(x)-f(x)|\leqslant A\left(\frac{\log n}{n}\right)^{\frac{\lambda}{1+2\lambda}} \qquad \text{a. s.}$$

for sufficiently large n, where A is a constant independent of x and n.

Corollary 1. *If $K(x)$ is λ-good with $\lambda=\delta$ and $f(x)$ satisfies δ^{th} order Lipschitz condition, then, with a suitable choice of h_n, we have*

$$\sup_{x\in R_1}|\hat{f}_n(x)-f(x)|\leqslant A(n^{-1}\log n)^{\frac{\delta}{1+2\delta}} \quad \text{a. s.} \tag{4}$$

for sufficiently large n.

Corollary 2. *If $K(x)$ is λ-good with $\lambda=\delta+1$ and $f'(x)$, the derivative of $f(x)$, satisfies δ^{th} order Lipschitz condition, then, with a suitable choice of h_n, we have*

$$\sup_{x\in R_1}|\hat{f}_n(x)-f(x)|\leqslant A(n^{-1}\log n)^{\frac{1+\delta}{3+2\delta}} \quad \text{a. s.} \tag{5}$$

for sufficiently large n.

In order to prove those corollaries, it is enough to verify that f satisfies λ-condition with $\lambda=\delta$, $\lambda=\delta+1$ respectively. Especially, if $f(x)$ possesses bounded second order derivative, then $f'(x)$ satisfies Lipschitz condition. From Corollary 2, it follows that

$$\sup_{x\in R_1}|\hat{f}_n(x)-f(x)|\leqslant A(n^{-1}\log n)^{2/5} \quad \text{a. s.}$$

for sufficiently large n.

But even if $f(x)$ is k-times ($k>2$) differentiable, we could not improve the rate of convergence of $\sup_{x\in R_1}|\hat{f}_n(x)-f(x)|$, where $\hat{f}_n(x)$ is kernel estimate with nonnegative kernel. In other words $2/5$ is the highest convergence rate for kernel estimate with nonnegative kernel.

Theorem 2. *For any fixed sequence $\{h_n\}$ and $\{c_n\}$ such that*

$$\lim_{n\to\infty} h_n=0,\quad \lim_{n\to 8} c_n=\infty$$

and

$$\lim_{n\to\infty}\frac{\log n}{nh_n}=0$$

there exists uniform continuous density function f such that

$$\sup_{x\in R_1}|\hat{f}_n(x)-f(x)|\geqslant\frac{1}{c_n} \quad \text{a. s.} \tag{6}$$

holds for sufficiently large n, where $\hat{f}_n(x)$ is a kernel estimate with a suitable choice of kernel.

This theorem means that it is impossible to establish any convergence rate of $\sup_x|\hat{f}_n(x)-f(x)|$ without some further conditions imposed on f besides of being uniformly continuous.

In the definition of kernel estimate, h_n does not depend on random sample. It is natural to use a function $H_n(X_1, \cdots, X_n)$ of the sample instead of h_n. Therefore (1) can be replaced by

$$\overset{*}{f}_n(x)=\frac{1}{n}\sum_{i=1}^{n}\frac{1}{H_n}K\left(\frac{x-X_i}{H_n}\right). \tag{7}$$

Form Theorem 1, it is easy to get the following theorem.

Theorem 3. *If $K(x)$ is λ-good and p. d. f. $f(x)$ satisfies λ-condition and*

$$|H_n-(n^{-1}\log n)^{\lambda/(1+2\lambda)}|\leqslant a(n^{-1}\log n)^{\beta\lambda(1+2\lambda)} \quad \text{a. s.}$$

holds for sufficiently large n, then we have, with probability one,

$$|\overset{*}{f}_n(x)-f(x)|\leqslant A(n^{-1}\log n)^{\frac{\lambda}{1+2\lambda}} \text{ for large } n,$$

where β is a positive number, $0<\beta\leqslant 1$.

In this paper we shall not discuss the estimator defined by (7) in detail.

§ 2. The Proof of Theorem 1.

In this section, A, A_1, A_2 $\cdots$ are all constants. We shall use Devroye's probability inequality in proving theorem and state it in a form that is suitable for our use.

Lemma 1. *Suppose that X_1, X_2, $\cdots$, X_n is a random sample drawn from a one-dimensional population with probability distribution F. Denote the empirical distribution of X_1, $\cdots$, X_n by F_n. Suppose $T\subset R^1$, $\mathscr{A}_l=\{[x-l', x+l']: x\in T, l'\leqslant l\}$, $\mathscr{A}_{2l}=\{[x-2l, x+2l]: x\in R^1\}$, and*

$$\sup_{A\in\mathscr{A}_{2l}} F(A)\leqslant B\leqslant 1/4.$$

Then for $\varepsilon>0$ and $n\geqslant\max(1/B, 8B/\varepsilon^2)$, we have

$$\begin{aligned}&P\{\sup_{A\in\mathscr{A}_l}|F_n(A)-F(A)|\geqslant\varepsilon\}\\&\leqslant 16n^2\exp\{-n\varepsilon^2/(64B+4\varepsilon)\}+8n\exp\{-nB/10\}.\end{aligned}$$

Proof See [2].

Lemma 2. *Let $\{h_n\}$ be decreating sequence with*

$$\lim_{n\to\infty} h_n=0$$

and

$$\lim_{n\to\infty}\frac{\log n}{n h_n}=0 \tag{8}$$

If

$$\sup_{A\in\mathscr{A}_{2l_n}} F(A)\leqslant M_0 h_n, \tag{9}$$

then we have

$$\limsup_{n\to\infty}\left(\frac{n}{h_n\log n}\right)^{\frac{1}{2}}\sup_{\mathscr{A}_{l_n}}|F_n(A)-F(A)|\leqslant C \text{ a. s.} \tag{10}$$

where M_0 is any fixed constant and C is a constant.

Proof Choose C such that $C^2>8M_0$, and take $B=M_0h_n$ and $\varepsilon=A\sqrt{h_n\log n/n}$ in Lemma 1. Then for large n,

$$\begin{aligned}&P\left(\left(\frac{n}{h_n\log n}\right)^{1/2}\sup_{\mathscr{A}_{l_n}}|F_n(A)-F(A)|>C\right)\\&\leqslant 16n^2\exp\{-n(A^2h_n\log n/n)/(M_0h_n+A\sqrt{h_n\log n/n})\}\\&\quad+8n\exp\{-nA\sqrt{h_n\log n/n}/10\}\end{aligned}$$

$$\leqslant 16n^2 \exp\left(-\frac{A^2}{2M_0}\log n\right)+8n\exp\{-3\log n\}\leqslant 24n^2.$$

Hence

$$\sum_{n=1}^{\infty} P\left(\left(\frac{n}{h_n \log n}\right)^{1/2} \sup_{\mathscr{A}_{l_n}} |F_n(A)-F(A)|>C\right)<\infty,$$

and (10) follows from the Borel–Cantelli's lemma.

Lemma 3. *Let $R(x)$ be a strictly decreasing continuous function defined on $[0, \infty)$ and $r(x)$ is its inverse function, then we have*

$$\int_0^\infty R(x)dx=\int_0^{R_0} r(t)dt$$

where $R_0=R(0)$.

Now we prove Theorem 1.

Let $g(x)$ be the inverse function of $K(x)$ on $[0, \infty)$. It is easy to verify that $g^{1+\lambda}(y)$ is the inverse function of $K(x^{\frac{1}{1+\lambda}})$. Since

$$\int_0^\infty K(x^{\frac{1}{1+\lambda}})dx=\int_0^\infty K(y)y^\lambda(1+\lambda)dy\leqslant\int_{-\infty}^\infty(1+\lambda)|y|^\lambda K(y)dy<+\infty,$$

we have

$$\int_0^{K_0} g^{1+\lambda}(y)dy=(1+\lambda)\int_0^\infty y^\lambda K(y)dy<+\infty,\ K_0=K(0). \tag{11}$$

In particular, for $\lambda=0$, we have

$$\int_0^{K_0} g(t)dt=\int_0^\infty K(y)dy. \tag{12}$$

From (12), for any n, we can choose an integer N_n such that

$$\left|\sum_{i=1}^{N_n}\frac{2a_{ni}}{N_n}K_0-1\right|\leqslant h_n^\lambda, \tag{13}$$

where

$$a_{ni}=g\left(\frac{iK_0}{N_n}\right),\quad i=1, 2, \cdots N_n. \tag{14}$$

From (11), we also have

$$\lim_{n\to\infty}\sum_{i=1}^{N_n}\frac{K_0}{N_n}[a_{ni}]^{1+\lambda}=\int_0^\infty[g(x)]^{1+\lambda}dx<+\infty.$$

Hence

$$\sup_n\sum_{i=1}^{N_n}\frac{K_0}{N_n}[a_{ni}]^{1+\lambda}<A_1. \tag{15}$$

Let

$$K_n(x)=\sum_{i=1}^{N_n}\frac{K_0}{N_n}I_{[-a_{ni},a_{ni}]}(x)=\sum_{i=1}^{N_n}\frac{2K_0a_{ni}}{N_n}K_{ni}(x), \tag{16}$$

where

$$K_{ni}(x)=\frac{1}{2a_{ni}}I_{[-a_{ni},a_{ni}]}(x)$$

and $I_{[B]}(x)$ denotes the indicator function of B. It is obvious that

$$|K_n(x)-K(x)|\leqslant K_0/N_n. \tag{17}$$

Let

$$\overset{*}{f}_n(x)=\frac{1}{nh_n}\sum_{i=1}^{n}K_n\left(\frac{(x-x_i)}{h_n}\right). \tag{18}$$

From (18), (17) and (1), we have

$$|f_n(x)-\overset{*}{f}_n(x)|\leqslant K_0/N_nh_n. \tag{19}$$

Certainly, we can choose N_n such that it satisfies the further condition $K_0/N_nh_n\leqslant h_n^\lambda$.

We can write (18) as

$$\overset{*}{f}_n(x)=\frac{1}{nh_n}\sum_{i=1}^{n}\sum_{j=1}^{N_n}\frac{2K_0a_{nj}}{N_n}K_{nj}\left(\frac{(x-x_i)}{h_n}\right)=\sum_{j=1}^{N_n}\frac{2K_0a_{nj}}{N_n}\overset{*}{f}_{nj}(x),$$

where

$$\overset{*}{f}_{nj}(x)=\frac{1}{nh_n}\sum_{i=1}^{n}K_{nj}\left(\frac{(x-x_i)}{h_n}\right).$$

Hence

$$\begin{aligned}|\overset{*}{f}_n(x)-f(x)|&=\left|\sum_{j=1}^{N_n}\frac{2K_0a_{nj}}{N_n}\overset{*}{f}_{nj}(x)-f(x)\right|\\&\leqslant\sum_{j=1}^{N_n}\frac{2K_0a_{nj}}{N_n}|\overset{*}{f}_{nj}(x)-f(x)|+\left|\sum_{j=1}^{N_n}\frac{2K_0a_{nj}}{N_n}-1\right|f(x).\end{aligned}$$

From (13), the second term on the right side is less than $A_1h_n^\lambda$. Now we estimate the first term. For any j, $1\leqslant j\leqslant N_n$, we have

$$\begin{aligned}|\overset{*}{f}_{nj}(x)-f(x)|&=\left|\frac{N_n(x-a_{nj}h_n,\ x+a_{nj}h_n)}{2na_{nj}h_n}-f(x)\right|\\&\leqslant\frac{1}{2a_{nj}h_n}\left|\frac{N_n(x-a_{nj}h_n,\ x+a_{nj}h_n)}{n}-P_{nj}(x)\right.\\&\quad+\frac{1}{2a_{nj}h_n}|P_{nj}(x)-2a_{nj}h_nf(x)|\triangleq I_{nj}(x)+J_{nj}(x),\end{aligned} \tag{20}$$

where

$$P_{nj}(x)=\int_{x-a_{nj}h_n}^{x+a_{nj}h_n}f(t)dt=F(x+a_{nj}h_n)-F(x-a_{nj}h_n)$$

and $N_n(a, b)$ denotes the number of X_i's among $X_1,\cdots,X_n$ which lie in the interval $[a, b]$.

It follows from λ-condition that

$$J_{nj}(x)\leqslant Ra_{nj}^\lambda h_n^\lambda,\ \forall x,\ j,\ n. \tag{21}$$

From Lemma 2, we have

$$\limsup_{n\to\infty}\left(\frac{nh_n}{\log n}\right)^{\frac{1}{2}}\sup_{1\leqslant j\leqslant N_n}\sup_{x\in R}I_{nj}(x)\leqslant C\quad\text{a. s.} \tag{22}$$

From (15) and (19)—(22), we have

$$\begin{aligned}&\limsup_{n\to\infty}\min\left(\sqrt{\frac{nh_n}{\log n}},\ h_n^{-\lambda}\right)\sup_x|\hat{f}_n(x)-f(x)|\\&\leqslant\limsup_{n\to\infty}\frac{K_0}{N_n}\sum_{j=1}^{N_n}\left(\frac{nh_n}{\log n}\right)^{\frac{1}{2}}\sup_x I_{nj}(x)\\&\quad+\limsup_{n\to\infty}h_n^{-\lambda}\left[\sum_{j=1}^{N_n}\frac{2a_{nj}K_0}{N_n}R(a_{nj})^\lambda h_n^\lambda+A_2h_n^\lambda\right]\leqslant A_3\quad\text{a. s.},\end{aligned} \tag{23}$$

where both A_2 and A_3 are constants.

We choose h_n such that

$$\sqrt{(n^{-1}\log n)h_n^{-1}}=h_n^{\lambda},$$

that is $h_n=(n^{-1}\log n)^{\frac{1}{1+2\lambda}}$, then (23) becomes (3). This completes the proof of Theorem 1.

Remarks.

1. In the process of this proof, we can see that Theorem 1 still holds when $K(x)$ is uniform distribution on $[-a, a]$, where a is a positive constant. Therefore, it is enough to suppose that kernel function K is decreasing on $[0, \infty)$ for Theorem 1.

2. About λ-condition and λ-good. We suppose that the kernel function $K(x)$ is symmetric only because it makes the proof simple and the idea clear. In fact, we can change them as follows:

The p. d. f. $f(x)$ is called satisfying λ-condition if

$$\left|\int_{x-a}^{x+b} f(t)dt-(a+b)f(x)\right|\leqslant R(b+a)^{\lambda+1}$$

holds for all $a>0$, $b>0$ and $x\in R_1$.

The kernel function $K(x)$ is called λ-good if

$$\int_{-\infty}^{\infty}|x|^{\lambda}K(x)dx<+\infty$$

and $K(x)$ decreases on $[0, \infty)$ and increases on $(-\infty, 0]$ or when $K(x)$ possesses a compact support S, $K(x)$ decreases on $[0, \infty)\cap S$ and increases on $(-\infty, 0]\cap S$.

3. It is not difficult to see that the sequence $\{\hat{f}_n\}$ defined by (1) is asymptotically optimal in C. J. Stone's sense[8].

4. We can replace λ-condition by

$$\sup_{x\in R_1}\left|\frac{1}{2a}\int_{x-a}^{x+a} f(t)dt-f(x)\right|=O(a^{\lambda}) \text{ as } a\to 0+. \tag{25}$$

At the same time, we have to suppose that

$$K(x)x^{1+\lambda}\to 0 \text{ as } x\to\infty \tag{26}$$

in order to ensure that (21) holds. In fact, we choose N_n so large that (13) and $K_0/N_nh_n\leqslant h_n^{\lambda}$ hold, that is

$$N_n\geqslant\frac{K_0}{h_n^{1+\lambda}}.$$

For any $1\leqslant j\leqslant N_n$, we have

$$a_{nj}h_n\leqslant a_{n1}h_n.$$

Set

$$M_n=g(h_n^{1+\lambda}). \tag{27}$$

Hence

$$h_n=[K(M_n)]^{\frac{1}{1+\lambda}}. \tag{28}$$

It follows from (27), (28) and (14) that

$$a_{n1}h_n=g\left(\frac{K_0}{N_n}\right)h_n\leqslant g(h_n^{1+\lambda})h_n\leqslant M_n[K(M_n)]^{\frac{1}{1+\lambda}}$$
$$\leqslant[K(M_n)M_n^{1+\lambda}]^{\frac{1}{1+\lambda}}0 \qquad \text{a.s. } n\to\infty.$$

§ 3. The Proof of Theorem 2.

Since Theorem 2 can be proved in a way similar to [6], we only give an outline.

For fixed constants $a>0$, $b>0$ and d with $0<d\leqslant 2ab$, we choose a function $g(a, b, d; x)$ defined on $|x|\leqslant b$ satisfying the following conditions:

1) $g(a, b, d; 0)=a$, $g(a, b, d: \pm b)=0$;
2) $0\leqslant g(a, b, d; x)\leqslant a$ for $|x|\leqslant b$;
3) $g(a, b, d; x)$ is continuous on $[-b, b]$;
4) $\int_{-b}^{b} g(a, b, d; x)dx=d$.

It is obvious that there exists such function. Let $\{c_n\}$ is any sequnce such that $c_n\to\infty$. First, we suppose that $\{C_n\}$ satisfies

$$C_n\leqslant C_n^0=\sqrt{\frac{nh_n}{8\log n}}. \tag{29}$$

Let

$$a_n=\frac{3}{C_n},\ b_n=C_nh_n,\ d_n=\frac{1}{2n},\ n\geqslant 1$$

and

$$e_1=C_1h_1,\ e_n=2\sum_{i=1}^{n-1}C_ih_i+C_nh_n,\ n\geqslant 2.$$

Let

$$f(x)=\begin{cases} g(a_n, b_n, d_n; x-e_n) \text{ for } |x-e_n|\leqslant b_n,\ n=1, 2, \cdots, \\ 0 \text{ for } x<0. \end{cases}$$

It is not difficult to see that f is uniformly continuous on R_1 and is a p. d. f.

Let

$$\xi_n=N_n(e_n-b_n,\ e_n+b_n).$$

It follows from Lemma 1 or Theorem 3 of [7] that

$$P(|\xi_n/n-d_n|\geqslant h_n/C_n)\leqslant 2\exp\{-nh_n^2/C_n^2/(2d_n+h_n/C_n)\}$$
$$\leqslant 2\exp\left(-\frac{nh_n}{4C_n^2}\right)\leqslant 2n^{-2} \text{ (for sufficiently large } n). \tag{30}$$

Let

$$T_n=\{w: w=(X_1\cdots X_n\cdots) \text{ such that } |\xi_n/n-d_n|\geqslant h_n/C_n\}.$$

It follows from (30) and Borel–Cantelli Lemma that

$$P(\limsup_{n\to\infty} T_n)=0.$$

We choose the uniform distribution on $[-1, 1]$ or a symmetric p. d. f. possessing compact support as the kernel function, then for any

$$w=(X_1, X_2\cdots)\overline{\in}\limsup_{n\to\infty} T_n,$$

we have, for large n,

$$\xi_n\leqslant 2h_n/C_n \text{ and } \hat{f}_n(e_n)\leqslant 2/C_n. \tag{31}$$

It follows from (31) and the definition of f that

$$|\hat{f}_n(e_n)-f(e_n)|\geqslant\frac{3}{C_n}-\frac{2}{C_n}=\frac{1}{C_n}. \tag{32}$$

It is obvious that (32) holds for C_n not satisfying (29). This completes the proof.

References

[1] Parzen, E., On the estimation of a probability density function and the mode, *Ann. Math. Statist.*, **33** (1962), 1065—1076.

[2] Devroye, L. P., and Wagner, T. J., The Strong Uniform Consistency of Kernel Density Estimation, *Multivariate Analysis V.*, (1980), 55—77.

[3] Schuster, E. F., *Ann. Math. Statist.*, **40** (1969), 1187.

[4] Singh, R. S., *Ann. Statist.*, (1976), 431.

[5] Chen Xiru, Uniform convergence rate of kernel density estimate of a probability density function. *Scientia Sinica*, to appear.

[6] Chen Xiru, Convergence rate of the nearest Neibor density estimate, *Scientia Sinica*, (1981), 12.

[7] Hoeffbing, W. J., *Amer. Statist. Assoc.*, **58** (1963), 13.

[8] Stone, C. J., Optimal uniform rate of convergence for nonparametric estimates of a density function or its derivatives,Recet Advances in Statistics: Papers Presented in Honor of Herman Chernoff's Sixtieth Birthday,Rizvi, Rustagi, & Siegmund (eds.), Academic Press, New York.