Chin. Ann. of Math.
21B: 4(2000),489-498

DISTRIBUTION FREE LAWS OF THE ITERATED LOGARITHM FOR KERNEL ESTIMATOR OF REGRESSION FUNCTION BASED ON DIRECTIONAL DATA***

WANG Xiaoming*,** ZHAO Lincheng** WU Yaohua**

Abstract

The authors derive laws of the iterated logarithm for kernel estimator of regression function based on directional data. The results are distribution free in the sense that they are true for all distributions of design variable.

Keywords Directional data, Laws of the iterated logrithm, Regression function kernel estimator, Strong convergence rates
1991 MR Subject Classification 62G05
Chinese Library Classification O212.7 **Document code** A
Article ID 0252-9599(2000)04-0489-10

§1. Introduction and Main Result

When statistical data consists of direction only, it can be represented as point of Ω, the surface of the unit sphere in d-dimensional Euclidean space, defined as $\Omega = \{x \in R^d : \|x\| = 1\}, d \geq 2$. We call it directional (or spherical) data.

The study of directional data is of practical interest. There are many situations where observed data are in the form of direction cosines or in the form of vectors but with an unknown positive scalar so that only the direction is known. Up to now, works based on such data all concentrated upon density problems, such as statistical inference for p.d.f. (probability density function) on Ω under various parametric models (see [7, 9, 10]) and nonparametric density estimation (see [1, 3]). In this paper, we consider another important subject, i.e. regression problems based on directional data, precisely speaking, regression problems with design variable taking values on Ω.

Given data $(x_i, Y_i) \in \Omega \times R$, $i = 1, \cdots, n$, consider regression model on Ω,

$$Y_i = m(x_i) + e_i, \quad 1 \leq i \leq n, \tag{1.1}$$

Manuscript received April 29, 1999. Revised January 18, 2000.

*Department of statistics, Shanghai University of Finance and Economics, Shanghai 200433, China.
**Department of statistics and Finance, University of Science and Technology of China, Hefei 230026, China.

***Project supported by the National Natural Science Foundation of China (Nos. 19631040; 19971085), the Doctoral Program Foundation of the Ministry of Education of China and the Special Foundation of Academia Sinica.

where the pairs (x_i, Y_i) are observed, the $x_i's$ are fixed vectors on Ω, $m(x) : \Omega \to R$ is a smooth function which is to be estimated and the errors e_i are independent and identically distributed (i.i.d.) random variables satisfying

$$E(e_i) = 0, \quad E(e_i^2) = 1. \tag{1.2}$$

To estimate the regression function $m(x) = E(Y|X = x)$, we introduce the following estimator

$$\hat{m}(x) = \sum_{i=1}^{n} K\Big(\frac{1 - x'x_i}{h_i^2}\Big)Y_i \Big/ \sum_{i=1}^{n} K\Big(\frac{1 - x'x_i}{h_i^2}\Big), \quad x \in \Omega, \tag{1.3}$$

where K is a non-negative kernel function defined on $R^+ = [0, +\infty)$ and $h_n \to 0$ are a sequence of positive numbers which are called bandwidth or window size. By the law of cosines, $\|x - y\|^2 = 2(1 - x'y)$ for $x, y \in \Omega$, one can easily find the similarity between $\hat{m}(x)$ and the definition of the general kernel estimator with data in R^d.

In the present article, we give a precise description of the exact rate of strong convergence of $\hat{m}(x) - E\hat{m}(x)$ by deriving laws of the iterated logarithm (LIL) of $\hat{m}(x) - E\hat{m}(x)$, in the most common setting for regression problems: where the design variables x_i are fixed, or conditioned upon, and are representation of a random sequence drawn from X (with an unknown distribution).

In our main theorem, we assume that the kernel function K is a non-negative function of bounded variation defined on R^+, vanishing outside $[0, \rho)$ for some $\rho > 0$ and bounded away from zero in $[0, \rho)$. For the bandwidth, we assume that

$$h_n \downarrow 0, \qquad n^\alpha \ll \sum_{i=1}^{n} h_i^{d-1} \tag{1.4}$$

for some $\alpha > 0$, where "$\ll$" denotes "O".

Next, we will state our main theorem. We shall say that a result holds for a class of realizations of $x_1, x_2, \cdots$ having $\mathcal{X}$ probability 1 if that class has probability 1 in the distribution of random sequence $X_1, X_2, \cdots$ drawn from the design population X.

Put

$$\lambda(K) \triangleq \frac{(2\pi)^{(d-1)/2}}{\Gamma((d-1)/2)} \int_0^\infty K(t)t^{(d-3)/2}dt$$

and μ denotes the probability measure of X on Ω.

Theorem. *For a class of realizations $x_1, x_2, \cdots$ having $\mathcal{X}$ probabiliaty* 1 *and for* a.e. $(\mu)x \in \Omega$, *we have*

$$\limsup_{n\to\infty} \Big(\sum_{i=1}^{n} h_i^{d-1} / \log\log n\Big)^{1/2} \{\hat{m}(x) - E\hat{m}(x)\} = \Big\{2\frac{\lambda(K^2)}{\lambda^2(K)}g(x)\Big\}^{1/2} \quad \text{a.s.}, \tag{1.5}$$

where $g(x)$ will be defined in Lemma 2.1 *given below.*

Replacing e_i by $-e_i$ in the regression model, the theorem implies

$$\liminf_{n\to\infty} \Big(\sum_{i=1}^{n} h_i^{d-1} / \log\log n\Big)^{1/2} \{\hat{m}(x) - E\hat{m}(x)\} = -\Big\{2\frac{\lambda(K^2)}{\lambda^2(K)}g(x)\Big\}^{1/2} \quad \text{a.s.}$$

So, we have the following corollary.

Corollary. *For a class of realizations $x_1, x_2, \cdots$ having $\mathcal{X}$ probabiliaty* 1 *and for* a.e. $(\mu)x \in \Omega$, *we have*

$$\limsup_{n\to\infty} \Big(\sum_{i=1}^{n} h_i^{d-1} / \log\log n\Big)^{1/2} |\hat{m}(x) - E\hat{m}(x)| = \Big\{2\frac{\lambda(K^2)}{\lambda^2(K)}g(x)\Big\}^{1/2} \quad \text{a.s.} \tag{1.6}$$

As a footnote, we would like to mention some related works on LIL for regression function estimator with data in the Euclidean space. These include, among others, [5] (LIL for kernel estimator in the case where the designs $x_i = X_i$ are univariate and are regarded as random), [4] (LIL for kernel estimator of regression function where the designs $x_i, i = 1, 2, \cdots$ are representation of a random sequence drawn from X) and [11] (LIL for kernel estimator of the ν-th derivative of regression function in fixed design case). Each of those works was discussed under some assumptions on the p.d.f. of the design variable X. In this paper, our results hold without any assumption on the distribution of the design variable, thus, they are distribution free. From the procedure of our proof given below, one can see that similar results also hold for the general kernel estimator with data in R^d.

§2. Preparatory Lemmas

Our proof of the main theorem is based on the following lemmas. Consider

$$\hat{m}(x) = \sum_{i=1}^{n} K\Big(\frac{1 - x'x_i}{h_i^2}\Big)Y_i \Big/ \sum_{i=1}^{n} K\Big(\frac{1 - x'x_i}{h_i^2}\Big).$$

Then $\hat{m}(x) - E\hat{m}(x) = \frac{U_n(x)}{V_n(x)}$, where

$$U_n(x) = \sum_{i=1}^{n} K\Big(\frac{1 - x'x_i}{h_i^2}\Big)e_i \Big/ \sum_{i=1}^{n} EK\Big(\frac{1 - x'X}{h_i^2}\Big),$$

$$V_n(x) = \sum_{i=1}^{n} K\Big(\frac{1 - x'x_i}{h_i^2}\Big) \Big/ \sum_{i=1}^{n} EK\Big(\frac{1 - x'X}{h_i^2}\Big).$$

Let ω be the Lebesgue measure on Ω and put $S_x(r) = \{t \in \Omega : \|t - x\| < r\}$.

Lemma 2.1. *Let $h = h_n$ be a sequence of positive numbers with $h_n \to 0$. Then there exists a non-negative function g with $g(x) < \infty$, such that*

$$\frac{\omega(S_x(h))}{\mu(S_x(h))} \to g(x) \quad \textit{as} \quad n \to \infty \quad \textit{for} \quad \text{a.e.} \ \ (\mu)x \in \Omega. \tag{2.1}$$

For a proof, refer to the proof of [2, Lemma 2.2].

Lemma 2.2. *Suppose that h_n satisfy*

$$h_n \to 0, \quad \frac{\sum\limits_{i=1}^{n} h_i^{d-1}}{\log n} \to \infty, \quad \textit{as} \quad n \to \infty \tag{2.2}$$

and the kernel $K(x)$ satisfies $K(x) \ge BI_{[0,r)}$ for some positive constants B and r, and $\sup K(x) \le M$. *Then, for a class of realizations of $x_1, x_2, \cdots$ having $\mathcal{X}$ probability* 1, *we have*

$$V_n(x) \to 1 \quad \text{a.e.} \ \ (\mu)x \in \Omega. \tag{2.3}$$

Proof. Let $X_1, X_2, \cdots$ denote the independent and identically distributed random variables of which $x_1, x_2, \cdots$ represents a realization. Put

$$V_n^*(x) = \sum_{i=1}^{n} K\Big(\frac{1 - x'X_i}{h_i^2}\Big) \Big/ \sum_{i=1}^{n} EK\Big(\frac{1 - x'X}{h_i^2}\Big).$$

Then

$$V_n^*(x) - 1 = \sum_{i=1}^{n} \Big[K\Big(\frac{1 - x'X_i}{h_i^2}\Big) - EK\Big(\frac{1 - x'X}{h_i^2}\Big)\Big] \Big/ \sum_{i=1}^{n} EK\Big(\frac{1 - x'X}{h_i^2}\Big) \triangleq \sum_{i=1}^{n} \eta_{ni}.$$

To prove Lemma 2.2, it is enough to prove that

$$\sum_{i=1}^{n} \eta_{ni} \to 0 \quad \text{a.s.} \quad \text{for} \quad \text{a.e.} \ \ (\mu)x \in \Omega. \tag{2.4}$$

Obviously

$$|\eta_{ni}| \le M\Big[\sum_{i=1}^{n} EK\Big(\frac{1-x'X}{h_i^2}\Big)\Big]^{-1}, \quad \sum_{i=1}^{n} \mathrm{Var}(\eta_{ni}) \le M\Big[\sum_{i=1}^{n} EK\Big(\frac{1-x'X}{h_i^2}\Big)\Big]^{-1}.$$

Thus, by Bernstain inequality, we have

$$P(|V_n^*(x)-1|>\varepsilon) \le 2\exp\Big\{-\varepsilon^2 \sum_{i=1}^{n} EK\Big(\frac{1-x'X}{h_i^2}\Big)\Big/\Big[(2+\frac{2}{3}\varepsilon)M\Big]\Big\} \tag{2.5}$$

holds for each $\varepsilon > 0$.

By Borel-Cantelli's lemma, (2.4) follows from (2.5) if we show that the right-hand side of (2.5) is summable. For this, by Lemma 2.1 and the fact (refer to the result (1.7) given in [1]) that

$$h^{-(d-1)} \int_\Omega K\Big(\frac{1-x'y}{h^2}\Big)\omega(dy) \to \lambda(K), \quad \text{as} \quad h \to 0,$$

we have

$$\begin{aligned} h_i^{d-1}/EK\Big(\frac{1-x'X}{h_i^2}\Big) &\le h_i^{d-1}\Big/\Big[BEI_{[0,r)}\Big(\frac{1-x'X}{h_i^2}\Big)\Big] \\ &\le \Big[h_i^{d-1}\Big/\Big(B\int_\Omega I_{[0,r)}\Big(\frac{1-x'y}{h_i^2}\Big)\omega(dy)\Big)\Big]\frac{\omega(S_x(\sqrt{2r}h_i))}{\mu(S_x(\sqrt{2r}h_i))} \\ &\to (B\lambda(K))^{-1}g(x) < \infty \quad \text{as} \quad i \to \infty \quad \text{for a.e.} \ \ (\mu)x \in \Omega. \end{aligned}$$

So there exist a constant $c(x) > 0$, such that

$$h_i^{d-1}/EK\Big(\frac{1-x'X}{h_i^2}\Big) \le c(x), \quad i = 1, 2, \cdots, \quad \text{for a.e.} \ \ (\mu)x \in \Omega.$$

From (2.2) and (2.6), it is easy to see that the right-hand side of (2.5) is summable.

The next two lemmas will play an important role in our theorem's proof.

Lemma 2.3. *Suppose that K is a non-negative function of bounded variation, vanishing on $[\rho, \infty)$, for some positive constant ρ, and further, $K \ge BI_{[0,r)}$ for some $B, r > 0$. Then*

$$\sum_{i=1}^{n} h_i^{d-1}\Big/\sum_{i=1}^{n} EK\Big(\frac{1-x'X}{h_i^2}\Big) \to \lambda^{-1}(K)g(x) \quad \textit{as} \ \ n \to \infty, \quad \textit{for} \ \text{a.e.} \ \ (\mu)x \in \Omega. \tag{2.7}$$

Proof. We need only to show that for each $x \in \{t : o \le g(t) < \infty\}$, (2.7) holds.

(i) Case of $g(x) = 0$

In this case, we need only to show that the left-hand side of (2.7) tends to zero. In fact, from Lemma 2.1, it is easy to see that

$$h_i^{d-1}/EK\Big(\frac{1-x'X}{h_i^2}\Big) \le h_i^{d-1}/B\mu(S_x(\sqrt{2r}h_i)) \to 0 \quad \text{as } i \to \infty.$$

So, for each $\varepsilon > 0$, there exists a positive integer $N(\varepsilon)$, such that for all $i > N(\varepsilon)$,

$$EK\Big(\frac{1-x'X}{h_i^2}\Big) \ge \varepsilon^{-1}h_i^{d-1}.$$

Then

$$\lim_{n\to\infty}\sum_{i=1}^{n} h_i^{d-1}\Big/\sum_{i=1}^{n} EK\Big(\frac{1-x'X}{h_i^2}\Big)$$
$$\le \lim_{n\to\infty}\sum_{i=1}^{n} h_i^{d-1}\Big/\Big[\sum_{i=1}^{N(\varepsilon)} EK\Big(\frac{1-x'X}{h_i^2}\Big)+\varepsilon^{-1}\sum_{i=N(\varepsilon)+1}^{n} h_i^{d-1}\Big]=\varepsilon$$

holds for any $\varepsilon > 0$. Therefore, (2.7) is true.

(ii) case of $g(x) > 0$.

Firstly, we assume that K is a step function, i.e. $K(x) = \sum_{j=1}^{l} \alpha_j I_{[a_{j-1},a_j)}(x)$, where $\alpha_j \ge 0$, $j = 1, 2, \cdots, l$ (say), and $0 = a_0 < a_1 < \cdots < a_l = \rho$. It is easy to see that K can be represented as

$$K(x)=\sum_{j=1}^{l}\beta_j I_{[0,a_j)}(x)=\sum_{\{j:\beta_j\ge 0\}}\beta_j I_{[0,a_j)}(x)+\sum_{\{j:\beta_j<0\}}\beta_j I_{[0,a_j)}(x)\triangleq K_1(x)+K_2(x)$$

where β_j is a linear combination of $\alpha_1, \cdots, \alpha_l$, for $j = 1, 2, \cdots, l$.

For each $\varepsilon > 0$, from Lemma 2.1, there exists a positive integer $N(\varepsilon)$, such that for all $j = 1, 2, \cdots, l$ and all $i > N(\varepsilon)$,

$$\Big|\frac{\mu(S_x(\sqrt{2a_j}h_i))}{\omega(S_x(\sqrt{2a_j}h_i))}-g^{-1}(x)\Big|<\varepsilon.$$

Thus

$$\sum_{i=1}^{n} EK_1\Big(\frac{1-x'X}{h_i^2}\Big)=\sum_{i=1}^{N(\varepsilon)} EK_1\Big(\frac{1-x'X}{h_i^2}\Big)+\sum_{N(\varepsilon)+1}^{n}\sum_{j:\beta_j\ge 0}\beta_j\mu(S_x(\sqrt{2a_j}h_i))$$
$$\le\sum_{i=1}^{N(\varepsilon)} EK_1\Big(\frac{1-x'X}{h_i^2}\Big)+(g^{-1}(x)+\varepsilon)\sum_{N(\varepsilon)+1}^{n}\sum_{j:\beta_j\ge 0}\beta_j\omega(S_x(\sqrt{2a_j}h_i))$$
$$=\sum_{i=1}^{N(\varepsilon)} EK_1\Big(\frac{1-x'X}{h_i^2}\Big)+(g^{-1}(x)+\varepsilon)\sum_{N(\varepsilon)+1}^{n}\int_\Omega K_1\Big(\frac{1-x'y}{h_i^2}\Big)\omega(dy)$$
$$\ll (g^{-1}(x)+\varepsilon)\sum_{i=1}^{n}\int_\Omega K_1\Big(\frac{1-x'y}{h_i^2}\Big)\omega(dy). \tag{2.8}$$

Similarly, we get

$$\sum_{i=1}^{n} EK_1\Big(\frac{1-x'X}{h_i^2}\Big)\gg (g^{-1}(x)-\varepsilon)\sum_{i=1}^{n}\int_\Omega K_1\Big(\frac{1-x'y}{h_i^2}\Big)\omega(dy). \tag{2.9}$$

Combining (2.8) and (2.9), we get

$$\lim_{n\to\infty}\sum_{i=1}^{n} EK_1\Big(\frac{1-x'X}{h_i^2}\Big)\Big/\sum_{i=1}^{n}\int_\Omega K_1\Big(\frac{1-x'y}{h_i^2}\Big)\omega(dy)=g^{-1}(x). \tag{2.10}$$

Similar to K_1 , we can prove that (2.10) also holds for K_2, and those imply

$$\lim_{n\to\infty}\sum_{i=1}^{n} EK\Big(\frac{1-x'y}{h_i^2}\Big)\Big/\sum_{i=1}^{n}\int_\Omega K\Big(\frac{1-x'X}{h_i^2}\Big)\omega(dy)=g^{-1}(x). \tag{2.11}$$

On the other hand, notice the fact that

$$\int_\Omega K\Big(\frac{1-x'y}{h_i^2}\Big)\omega(dy)/h_i^{d-1} \to \lambda(K) \quad \text{as} \quad i\to\infty.$$

In a way similar to the proof of (2.11), one finds easily that

$$\lim_{n\to\infty}\sum_{i=1}^n \int_\Omega K\Big(\frac{1-x'y}{h_i^2}\Big)\omega(dy)\Big/\sum_{i=1}^n h_i^{d-1} = \lambda(K). \tag{2.12}$$

Combining (2.11) and (2.12), we see that (2.7) follows.

Secondly, we will prove that (2.7) also holds if K satisfies the condition described in Lemma 2.3. In fact, for any fixed $\varepsilon>0$, there exists a partition $0=a_0<a_1<\cdots<a_{l(\varepsilon)}=\rho$ on $[0,\rho)$ and a sequence of non-negative constants $\alpha_j, j=1,2,\cdots,l(\varepsilon)$ (say), such that $K^*(x)=\sum\limits_{i=1}^{l(\varepsilon)}\alpha_j I_[a_{j-1},a_j)(x)$ satisfies $0\le K(x)-K^*(x)\le\varepsilon$ and $\alpha_1>0$.

Then, for the step function K^*, from (2.7), we have

$$\begin{aligned}\lambda(K^*)g^{-1}(x) &\le \liminf_{n\to\infty}\sum_{i=1}^n EK\Big(\frac{1-x'y}{h_i^2}\Big)\Big/\sum_{i=1}^n h_i^{d-1} \le \limsup_{n\to\infty}\sum_{i=1}^n EK\Big(\frac{1-x'y}{h_i^2}\Big)\Big/\sum_{i=1}^n h_i^{d-1}\\ &\le [\lambda(K^*)g^{-1}(x)+\varepsilon\lambda(I_{[0,\rho)}]g^{-1}(x).\end{aligned}$$

Let $\varepsilon\to0$, we get $\lambda(K^*)\to\lambda(K)$. So (2.7) holds for K.

Lemma 2.4. *Let $c_1,\cdots,c_N$ be real numbers and $I_1,\cdots,I_N$ be numbers taking only the values 0 and 1, such that $|c_i|I_i=|c_i|$ for $1\le i\le N$. There exists an absolute constant A such that for all $u>0$ and all $C_1,C_2>0$,*

$$\begin{aligned}&P\Big\{\Big|\sum_{i=1}^N e_ic_i\Big|>(C_1+C_2)\Big(\sum_{i=1}^N c_i^2\Big)^{1/2}u\Big\}\le A\Big[1-\Phi(C_2u)+C_2^{-3}\Big(\sum_{i=1}^N c_i^2\Big)^{-1/2}\\ &\times\Big(\sup_{1\le i\le N}|c_i|\Big)u^{-3}E\{|e|^3I_{(|e|\le\lambda)}\}+\Big(\sum_{i=1}^N I_i\Big)P(|e|>\lambda)\Big],\end{aligned}$$

where $\lambda=(\sum I_i)^{1/2}/C_1$ and Φ denotes the standard normal distribution function.

For a proof, see [4, Lemma 3.1].

§3. Proof of the Theorem

Put $H_n=\sum\limits_{i=1}^n h_i^{d-1}$, $S_n=\sum\limits_{i=1}^n e_iK\Big(\frac{1-x'x_i}{h_i^2}\Big)$, and $\sigma_n^2=\mathrm{Var}(S_n)=\sum\limits_{i=1}^n K^2\Big(\frac{1-x'x_i}{h_i^2}\Big)$. From Lemma 2.2, Lemma 2.3 and the condition (1.4), noticing that K^2 also satisfies all the conditions imposed on K, we have $n^\alpha\ll H_n\ll\sigma_n^2\ll n$. Thus

$$\log\log\sigma_n^2\sim\log\log n\sim\log\log H_n, \tag{3.1}$$

where $a_n\sim b_n$ means that $a_n/b_n\to1$.

To prove (1.5), from Lemma 2.2, we need only to prove that for all $x\in\Omega$ satisfying $0\le g(x)<\infty$,

$$\limsup_{n\to\infty}(H_n/\log\log H_n)^{1/2}U_n(x)=\Big\{2\frac{\lambda(K^2)}{\lambda^2(K)}g(x)\Big\}^{1/2} \quad \text{a.s.} \tag{3.2}$$

holds for a class of realizations having $\mathcal{X}$ probability 1.

If $g(x)>0$, one can check, by Lemma 2.2 and Lemma 2.3, that with $\mathcal{X}$ probability 1,

$$(2\sigma_n^2H_n)^{\frac12}\Big/\sum_{i=1}^n EK\Big(\frac{1-x'X}{h_i^2}\Big)\to\Big\{\frac{2\lambda(K^2)}{\lambda^2(K)}g(x)\Big\}^{\frac12}.$$

Thus, to prove (3.2) for $g(x)>0$, we need only to prove that with $\mathcal{X}$ probability 1,

$$\limsup_{n\to\infty}(2\sigma_n^2\log\log\sigma_n^2)^{-1/2}S_n=1,\quad \text{a.s.} \tag{3.3}$$

If $g(x)=0$, we have that with $\mathcal{X}$ probability 1,

$$(\sigma_n^2H_n)^{\frac{1}{2}}\Big/\sum_{i=1}^n EK\Big(\frac{1-x'X}{h_i^2}\Big)\to 0\quad \text{a.s.}$$

Therefore, to prove (3.2) for $g(x)=0$, we need only to prove that with $\mathcal{X}$ probability 1,

$$\limsup_{n\to\infty}(2\sigma_n^2\log\log\sigma_n^2)S_n<\infty\quad \text{a.s.} \tag{3.4}$$

For those, we do in two steps, which give upper bound for the case $g(x)\ge 0$ and lower bound for the case $g(x)>0$.

Step (i) Upper bound for $g(x)\ge 0$

In this step, we will show (3.4) and the upper bound of (3.3), i.e.

$$\limsup_{n\to\infty}(2\sigma_n^2\log\log\sigma_n^2)^{-1/2}S_n\le 1\quad \text{a.s.} \tag{3.5}$$

Let $c>1$, write m_k for the integer part of c^k and write f.a.s.l.k as an abbreviation for 'for all sufficiently large k'. Put $t_n=(2\sigma_n^2\log\log\sigma_n^2)^{1/2}$. Without lose of generality we assume that t_n is increasing in n. Obviously, (3.4) and (3.5) will follow if we prove that:

(i) in the case $g(x)\ge 0$, for any $\varepsilon>0$

$$S_{m_k}/t_{m_k}<1+\varepsilon\quad \text{f.a.s.l.k.} \tag{3.6}$$

(ii) in the case $g(x)>0$, for any $\varepsilon>0$ there exists $c>0$, chosen sufficiently close to 1, such that

$$t_{m_k}^{-1}\sup_{m_k<n<m_{k+1}}|S_n-S_{m_k}|<\varepsilon\quad \text{f.a.s.l.k.} \tag{3.7}$$

(iii) in the case $g(x)=0$, there exists a constant C, such that

$$t_{m_k}^{-1}\sup_{m_k<n<m_{k+1}}|S_n-S_{m_k}|<C\quad \text{f.a.s.l.k.} \tag{3.8}$$

By the Borel-Cantelli lemma, (i) will follows if we prove that

$$\sum_{k=1}^\infty P\{S_{m_k}>(1+\varepsilon)t_{m_k}\}<\infty. \tag{3.9}$$

Applying Lemma 2.4 with $c_i=K(\frac{1-x'x_i}{h_i^2})$, $I_i=I_{(1-x'x_i<\rho h_i^2)}$, $C_1=\varepsilon/2$, $C_2=1+\varepsilon/2$, and $u=u_k=(2\log\log\sigma_{m_k}^2)^{1/2}$, we have

$$\begin{aligned}P\{S_{m_k}>(1+\varepsilon)t_{m_k}\}&\ll\Big(1-\Phi\Big\{\Big(1+\frac{\varepsilon}{2}\Big)u_k\Big\}\Big)+\lambda_k^{-1}E\{|e|^3I_{(|e|\le\lambda_k)}\}+\lambda_k^2P(|e|>\lambda_k)\\ &\triangleq \mathrm{I}+\mathrm{II}+\mathrm{III},\end{aligned}$$

where $\lambda_k=2\Big(\sum\limits_{i=1}^{m_k}I_i\Big)^{\frac{1}{2}}/\varepsilon$. Noticing that, from Lemma 2.2 and Lemma 2.3, with $\mathcal{X}$ probability 1,

$$\begin{aligned}\sum_{i=1}^{m_k}h_i^{d-1}/\lambda_k^2&=\frac{\varepsilon}{2}\sum_{i=1}^{m_k}h_i^{d-1}\Big/\sum_{i=1}^{m_k}I_{[0,\rho)}\Big(\frac{1-x'x_i}{h_i^2}\Big)\sim\frac{\varepsilon}{2}\sum_{i=1}^{m_k}h_i^{d-1}\Big/\sum_{i=1}^{m_k}EI_{[0,\rho)}\Big(\frac{1-x'X}{h_i^2}\Big)\\ &\to\begin{cases}\frac{\varepsilon}{2}\lambda^{-1}(I_{[0,\rho)})g(x) & \text{as}\quad g(x)>0,\\ 0 & \text{as}\quad g(x)=0,\end{cases}\end{aligned}$$

from (1.4), we have

$$m_k^\alpha \ll \sum_i^{m_k} h_i^{d-1} \ll \lambda_k^2. \tag{3.10}$$

So, we can find a positive sequence $\{\lambda_l^*\}$ including $\{\lambda_k\}$ as its subsequence such that $\lambda_l^* \asymp m_l^\alpha$, where $a_n \asymp b_n$ means that a_n/b_n and b_n/a_n are both bounded. Hence

$$\sum_{k=1}^\infty \lambda_k^{-1} E\{|e|^3 I(|e| \le \lambda_k)\} \le \sum_{l=1}^\infty (\lambda_l^*)^{-1} E\{|e|^3 I_{(|e|\le\lambda_l^*)}\} < \infty.$$

Here the last inequality is proved by making integral approximation to the series. Now, we know that II is summable. Similarly we can prove that III is also summable.

Applying the famous exponential inequality

$$\frac{1}{\sqrt{2\pi}}\Big(\frac{1}{x}-\frac{1}{x^3}\Big)e^{-x^2/2} \le 1-\Phi(x) \le \frac{1}{\sqrt{2\pi}x}e^{-x^2/2} \tag{$*$}$$

we can easily varify that I is summable. Up to now, we have proved (i).

Now, we proceed to prove (ii) and (iii). From Kolmogorov inequality (refer to [6, p.260] or [4, Lemma 3.2]), for any constant $C > 0$, we have

$$\begin{aligned} &P\Big\{t_{m_k}^{-1} \sup_{m_k<n<m_{k+1}} \Big|\sum_{i=m_k+1}^{n} e_i K\Big(\frac{1-x'x_i}{h_i^2}\Big)\Big| > C\Big\} \\ &\le P\Big\{\Big|\sum_{i=m_k+1}^{m_{k+1}} e_i K\Big(\frac{1-x'x_i}{h_i^2}\Big)\Big| > \frac{C}{2}A^{1/2}(k) \\ &\quad\times\Big(2\sum_{i=m_k+1}^{m_{k+1}} K^2\Big(\frac{1-x'x_i}{h_i^2}\Big)\log\log\sigma_{m_k}^2\Big)^{1/2}\Big\} \end{aligned} \tag{3.11}$$

where

$$A(k) = \sum_{i=1}^{m_k} K^2\Big(\frac{1-x'x_i}{h_i^2}\Big)\Big/\sum_{i=m_k+1}^{m_k} K^2\Big(\frac{1-x'x_i}{h_i^2}\Big).$$

If $g(x) > 0$, taking $C = \varepsilon$, from Lemma 2.2, we have that with $\mathcal{X}$ probability 1,

$$A^{-1}(k) = \frac{\sigma_{m_{k+1}}^2}{\sigma_{m_k}^2} - 1 \sim \frac{\sum_{i=1}^{m_{k+1}} EK^2\Big(\frac{1-x'x_i}{h_i^2}\Big)}{\sum_{i=1}^{m_k} EK^2\Big(\frac{1-x'x_i}{h_i^2}\Big)} - 1 \sim \frac{H_{m_{k+1}}}{H_{m_k}} - 1.$$

Then, by taking c sufficiently small and k sufficiently large, we have $A^{1/2}(k) > \frac{2(1+\varepsilon)}{\varepsilon}$. So, similar to (i), we can prove that the right-hand side of (3.11) is summable.

If $g(x) = 0$, taking $c < 2$ such that $m_{k+1} = [c^{k+1}] \le 2[c^k] = 2m_k$, then by the conditions imposed on K, we have

$$A^{-1}(k) = \frac{\sum_{i=m_k+1}^{m_{k+1}} K^2\Big(\frac{1-x'x_i}{h_i^2}\Big)}{\sum_{i=1}^{m_k} K^2\Big(\frac{1-x'x_i}{h_i^2}\Big)} \le \frac{M^2}{m^2}\frac{\sum_{i=1}^{m_{k+1}-m_k} I(1-x'x_i \le \rho h_i^2)}{\sum_{i=1}^{m_k} I(1-x'x_i \le \rho h_i^2)} \le \frac{M^2}{m^2}.$$

So, taking $C = \frac{6M}{m}$, similar to (i), we can prove that the right-hand side of (3.11) is summable. Hence (ii) and (iii) hold.

Step (ii) Lower bound for $g(x) > 0$

Write i.o. for infinitely often. In this step, we shall prove that for any $\varepsilon \in (0,1)$, with $\mathcal{X}$ probabilily 1,

$$S_n/(2\sigma_n^2 \log\log \sigma_n^2)^{1/2} > 1-\varepsilon \quad \text{i.o.} \tag{3.12}$$

Define m_k and t_k as before. Result (3.12) will hold if we prove that for each $\varepsilon > 0$, there exists a sufficiently large c such that

$$|S_{m_k}|/t_{m_{k+1}} < \varepsilon \quad \text{f.a.s.l.k.}, \tag{3.13}$$

$$(S_{m_{k+1}} - S_{m_k})/t_{m_{k+1}} > 1-\varepsilon \quad \text{i.o.} \tag{3.14}$$

Result (3.13) can be deduced immediately from Step (i). In fact, Step (i) implies that for all sufficiently large n, $S_n/t_n < 1+\varepsilon$ and $-S_n/t_n < 1+\varepsilon$ as may be seen on replacing e_i by $-e_i$ in the regression model. Therefore $|S_n|/t_n < 1+\varepsilon$ holds for all sufficiently large n. Choose $c > 1$ so large that for sufficiently large k, $t_{m_k}/t_{m_{k+1}} < \varepsilon/2$. Therefore $|S_{m_k}|/t_{m_{k+1}} < (1+\varepsilon)\varepsilon/2 < \varepsilon$, for all sufficiently large k.

Next, we will prove (3.14). Observing that

$$S_{m_{k+1}} - S_{m_k} = \sum_{i=m_k+1}^{m_{k+1}} e_i K\Big(\frac{1-x'x_i}{h_i^2}\Big)$$

by the Borel-Cantelli's lemma, we need only to prove that

$$\sum_{k=1}^{\infty} P\{S_{m_{k+1}} - S_{m_k} > (1-\varepsilon)t_{m_{k+1}}\} = \infty \tag{3.15}$$

holds for $c > 1$ sufficiently large. Write

$$e_{ik} = e_i I(|e_i|^2 \le H_{m_{k+1}}), \quad T_k = \sum_{i=m_k+1}^{m_{k+1}} e_{ik} K\Big(\frac{1-x'x_i}{h_i^2}\Big),$$

$$\mu_k = E(T_k) = E(e_{ik}) \sum_{i=m_k+1}^{m_{k+1}} K\Big(\frac{1-x'x_i}{h_i^2}\Big),$$

$$v_k = \mathrm{Var}(T_k) = \mathrm{Var}(e_{ik}) \sum_{i=m_k+1}^{m_{k+1}} K^2\Big(\frac{1-x'x_i}{h_i^2}\Big).$$

Then, putting $y_k = ((1-\varepsilon)t_{m_{k+1}} - \mu_k)/v_k^{1/2}$, we have

$$\begin{aligned}
&P\{S_{m_{k+1}} - S_{m_k} > (1-\varepsilon)t_{m_{k+1}}\}\\
&\ge P(T_k > (1-\varepsilon)t_{m_{k+1}}) - \Big\{\sum_{i=m_k+1}^{m_{k+1}} I_{(1-x'x_i \le \rho h_i^2)}\Big\} P(|e|^2 > H_{m_{k+1}})\\
&= P\Big(\frac{T_k-\mu_k}{v_k^{1/2}} > y_k\Big) - P(|e|^2 > H_{m_{k+1}}) \sum_{i=m_k+1}^{m_{k+1}} I_{(1-x'x_i \le \rho h_i^2)}\\
&\ge 1 - \Phi(y_k) - \sup_{-\infty<y<\infty} \Big|P\Big(\frac{T_k-\mu_k}{v_k^{1/2}} \le y_k\Big) - \Phi(y_k)\Big| - P(|e|^2 > H_{m_{k+1}})\\
&\quad \times \sum_{i=m_k+1}^{m_{k+1}} I_{(1-x'x_i \le \rho h_i^2)}.
\end{aligned}$$

Therefore, (3.15) will follow if we prove that for sufficiently large c,

$$\sum_{k=1}^{\infty} 1-\Phi(y_k)=\infty, \tag{3.16}$$

$$\sum_{k=1}^{\infty} \sup_{-\infty<y<\infty} |P([T_k-\mu_k]/v_k^{1/2}\le y_k)-\Phi(y_k)|<\infty, \tag{3.17}$$

$$\sum_{k=1}^{\infty} P(|e|^2>H_{m_{k+1}}) \sum_{i=m_k+1}^{m_{k+1}} I(1-x'x_i\le \rho h_i^2)<\infty \tag{3.18}$$

all hold with $\mathcal{X}$ probability 1. Noticing the fact that $E(e_{1k})\to 0$ and $\mathrm{Var}(e_{1k})\to 1$ as $k\to\infty$, it is easy to see, from Lemma 2.2, that for large k, $y_k\le(1-\varepsilon/2)(2\log\log H_{m_{k+1}})$, by using the exponential inequality given in $(*)$, we can verify (3.16) without difficulty. The proof of (3.18) is very similar to the derivation of (III) in step (i).

The remainder of our proof is dedicated to deriving (3.17). By the Esseen's inequality (see [6, p.111]), and Lemma 2.2, there exists an absolute constant $A>0$ such that

$$\sup_{-\infty<y<\infty}\Big|P\Big(\frac{T_k-\mu_k}{v_k^{1/2}}\le y\Big)-\Phi(y)\Big|\le Av_k^{-3/2}(E|e_{ik}|^3)\sum_{i=m_k+1}^{m_{k=1}}K^3\Big(\frac{1-x'x_i}{h_i^2}\Big)$$

$$\sim A\frac{\lambda(K^3)}{\lambda(K^2)^{3/2}}g(x)\Big(\sum_{i=m_k+1}^{m_{k+1}}h_i^{k-1}\Big)^{-\frac12}E\{|e|^3 I_{(|e|^2\le H_{m_{k+1}})}\}$$

$$\sim A\frac{\lambda(K^3)}{\lambda(K^2)^{3/2}}g(x)(1-c^{-1})H_{m_{k+1}}^{-\frac12}E\{|e|^3 I_{(|e|^2\le H_{m_{k+1}})}\}.$$

In a way similar to the proof of (II) in step (i), we can prove that

$$\sum_{k=1}^{\infty} H_{m_{k+1}}^{-\frac12}E\{|e|^3 I_{(|e|\le H_{m_{k+1}}^{1/2})}\}<\infty.$$

So (3.17) holds. By then, we have completed the proof of (3.12).

References

[1] Bai, Z. D., Radhakrishna Rao, C. & Zhao, L. C. , Kernel estimators of density function of directional data [J], *J.Multivariate Anal.*, **27**(1988), 24–39.

[2] Devroye, L., On the almost everywhere convergence of nonparametric regression estimates [J], *Ann. Statist.*, (1981), 1310–1319.

[3] Hall, P., Watson, G. S. & Cabrera, J., Kernel density estimation with spherical data [J], *Biometrika*, **74**(1987), 751–762.

[4] Hall, P., On iterated logrithm laws for linear arrays and nonparametric regression estimators [J], *Ann. probab.*, **19**(1991), 740–757.

[5] Hardle, W., A law of the iterated logarithm for nonparametric regression function estimators [J], *Ann. Statist.*, **12**(1984), 625–635.

[6] Loeve, M., Probablity theory 1 [M], 4th ed., Springer, Berlin, 1977.

[7] Mardia, K. V., Statistics of directional date [M], Academic Press, New York, 1972.

[8] Petrov, V. V., Sums of independent random variables [M], Springer, Berlin 1975.

[9] Pukkila, L. & Rao, C. R., Pattern recognition based on scale imariant discriminant functions [A], Technical Report 86-09 [R], Center for Multicariate Analysis, University of Pittsburgh, 1986.

[10] Watson, G. S., Statistics on spheres [M], University of Arkansas , Lecture Notes in Mathemtical Scieuces 6, New York: Wiley, 1983.

[11] Chen, Z. G. & Gasser, T., A law of the iterated logarithm for kernel estimators of regression function [J], *Probab. Theory Relat. Fields*, **93**(1992), 185–296.