Zhang Shuang Lin

Strong consistency of regression function estimates

Kybernetika, Vol. 31 (1995), No. 4, 375--384

Persistent URL: http://dml.cz/dmlcz/124674

KYBERNETIKA — VOLUME 31 (1995), NUMBER 4, PAGES 375–384

STRONG CONSISTENCY OF REGRESSION FUNCTION ESTIMATES[1]

ZHANG SHUANG LIN

Let $m_n(x)$ and $M_n(x)$ be a partitioning estimate and the kernel estimate, respectively, of a regression function $m(x) = E(Y|X = x)$ for the i.i.d. sample $(X_1, Y_1), \ldots, (X_n, Y_n)$. Under the condition $E|Y|^p < \infty$, where $p > 1$, and some conditions on the partition and the kernel function, the strong L_1-consistency is proved.

INTRODUCTION

Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be independent observations of an $\mathcal{R}^d \times \mathcal{R}$-valued random vector (X, Y). Denote the probability measure of X by μ, and the empirical measure for $X_1, X_2, \ldots, X_n$ by μ_n.

The regression function

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

can be estimated by the *kernel estimate*

$$M'_n(x) = \frac{\sum_{i=1}^{n} Y_i K_h(x - X_i)}{\sum_{i=1}^{n} K_h(x - X_i)} \tag{1}$$

where $h > 0$ is a smoothing factor depending upon n, K is an absolutely integrable function (the kernel), and $K_h(x) = K(x/h)$ ([11,14]), and the *k-nearest neighbor estimate,*

$$m_n(x) = \sum_{i=1}^{n} W_{ni}(x; X_1, \ldots, X_n)\, Y_i \,,$$

where $W_{ni}(x; X_1, \ldots, X_n)$ is $1/k$ if X_i is one of the k nearest neighbors of x among $X_1, \ldots, X_n$, and W_{ni} is zero otherwise. Note in particular that $\sum_{i=1}^{n} W_{ni} = 1$. The k-nearest neighbor estimate was studied by [13] among others. Alternatively, one can use the partitioning estimate, which is based on a finite or countable infinite

[1]The author is partly supported by The Fund of Natural Science of Heilongjiang Province, China.

Borel measurable partition $\mathcal{P}_n$ of $\mathcal{R}^{\rm d}$ ($\mathcal{P}_n = \{A_{n1}, A_{n2}, \ldots\}$). If $A_n(x)$ is the set from $\mathcal{P}_n$ to which x belongs, then the *partitioning estimate* is defined as

$$m_n(x) = \begin{cases} \dfrac{\sum_{i=1}^{n} I_{\{X_i \in A_n(x)\}} Y_i}{n\mu_n(A_n(x))} & \text{if } \mu_n(A_n(x)) > 0 \\ \frac{1}{n}\sum_{i=1}^{n} Y_i & \text{otherwise.} \end{cases} \tag{2}$$

We are concerned with the L_1 convergence of m_n to m as measured by

$$J_n = \int |m_n(x) - m(x)|\,\mu(\mathrm{d}x).$$

This quantity is particularly important in discrimination (see [8] or [13]). Stone [13] first pointed out that there exist estimators for which $J_n \to 0$ in probability for all distributions of (X, Y) with $E|Y| < \infty$. This included the nearest neighbor and partitioning estimates. In 1980, Devroye and Wagner, and independently Spiegelman and Sacks, showed that this is also the case for the kernel estimate with smoothing factor h provided that K is a bounded nonnegative function with compact support such that for a small fixed sphere S centered at the origin, $\inf_{x\in S} K(x) > 0$, and that

$$\lim_{n\to\infty} h = 0, \quad \lim_{n\to\infty} nh^{\rm d} = \infty.$$

Since then, many authors have considered the strong consistency of the three estimates. We summarize what is known in this respect:

- For the *k-nearest neighbor estimates*, $J_n \to 0$ almost surely under the conditions $k \to \infty$, $k/n \to 0$ whenever X has a density and Y is bounded (chapter X of Devroye and Györfi [5] and Zhao [15]). Beck [1] showed this result earlier under the additional constraint that m has a continuous version. The asymptotic normality is proved by Falk and Reiss [9]. Moreover, $J_n \to 0$ almost surely for all distributions of (X, Y) with Y bounded (the existence of the density of X is not required), provided that $k/n \to 0$ and $k/\log\log n \to \infty$ (Devroye [4]). $J_n \to 0$ almost surely for all distributions of (X, Y), under the conditions that $E|Y| < \infty$, and $k/n \to 0$ and $k/\log n \to \infty$ (Devroye, Györfi, Krzyżak and Lugosi [6]).

- Assuming that Y is bounded, the *kernel estimate* is strongly consistent if $\lim_{n\to\infty} h = 0$, $\lim_{n\to\infty} nh^{\rm d} = \infty$ holds, K is a Riemann integrable kernel and $K \geq aI_S$, where $a > 0$ is a constant, and S is a ball centered at the origin (Devroye and Krzyżak [7]).

- For the partitioning estimates, Devroye and Györfi [5] gave an exponential bound for the tail distribution of J_n for all (X, Y) with Y being bounded, and Györfi [10] has pointed out that $J_n \to 0$ almost surely for a modification of *partitioning estimates*

$$m_n(x) = \begin{cases} \frac{\sum_{i=1}^{n} I_{\{X_i \in A_n(x)\}} Y_i}{n\mu_n(A_n(x))} & \text{if } \mu_n(A_n(x)) > \frac{\log n}{n} \\ \frac{1}{n}\sum_{i=1}^{n} Y_i & \text{otherwise,} \end{cases}$$

whenever $E|Y| < \infty$, provided that $\lim_{n\to\infty} h = 0$, $\lim_{n\to\infty} nh^{\mathrm{d}}/\log n = \infty$ is satisfied.

In this note, we study the strong consistency of the partitioning estimate (2) and the modified kernel estimate

$$M_n(x) = \begin{cases} \frac{\sum_{i=1}^{n} Y_i K_h(x-X_i)}{\sum_{i=1}^{n} K_h(x-X_i)} & \text{if } \|x\| \le t_n \\ \frac{1}{n}\sum_{i=1}^{n} Y_i & \text{otherwise,} \end{cases} \tag{3}$$

(where $t_n^{\mathrm{d}} = \frac{nh^{\mathrm{d}}}{(\log n)^{q+1}}$, $q = \frac{p}{p-1}$, and $0/0$ is defined as 0) when Y is unbounded.

In the following we use $m_n(x)$, $M_n(x)$, $M_n'(x)$ to represent the partitioning estimate (2), the modified kernel estimate (3) and the kernel estimate (1), respectively.

1. MAIN RESULTS

Theorem 1. Suppose that $E|Y|^p < \infty$ for some $p > 1$, and assume that the finite partition $\mathcal{P}_n = \{A_{n1}, \ldots, A_{nk_n}\}$ with

$$k_n \le \frac{n}{(\log n)^q} \tag{4}$$

and $q = \frac{p}{p-1}$ satisfies that for each sphere S centered at the origin

$$\limsup_{n\to\infty} \max_{i: A_{ni}\cap S \ne \phi} \left(\sup_{x,y\in A_{ni}} \| x - y \| \right) = 0. \tag{5}$$

Then partitioning estimate (2) is strongly consistent, i. e. $J_n \to 0$, a. s.

Theorem 2. Suppose that $E|Y|^p < \infty$ for some $p > 1$, $K(x)$ is a kernel function such that there are constants $C_1 > 0$, $C_2 > 0$, $r > 0$ with $C_1 I_{S_r}(x) \le K(x) \le C_2 I_{S_r}(x)$, where S_r is a sphere centered at origin of radius r. If $\lim_{n\to\infty} h = 0$ and $\lim_{n\to\infty} \frac{nh^{\mathrm{d}}}{(\log n)^{q+1}} = \infty$, then

$$\int |M_n(x) - m(x)| \mu(\mathrm{d}x) \to 0 \quad \text{a. s.}$$

2. PROOFS

In order to prove the theorems, we need the following lemmas:

Lemma 1. (Devroye and Györfi [5]) If Y is bounded, that is there is a constant L such that $|Y| \le L < \infty$, then under conditions (4) and (5), for each $\varepsilon > 0$, there exists n_0 such that

$$P\left(\int |m_n(x) - m(x)|\mu(\mathrm{d}x) \ge \varepsilon\right) \le \exp(-c(\varepsilon/L^2)n); \quad n \ge n_0.$$

for some constant c.

Lemma 2. (Chernoff [3]) Let B be a binomial random variable with parameters n and p. Then

$$P(B < \varepsilon) \le \exp(\varepsilon - np - \varepsilon\log(\varepsilon/np)) \quad (\varepsilon < np).$$

Proof of the Theorem 1. For an arbitrary L, let

$$Y_{iL} = \begin{cases} Y_i & \text{if } |Y_i| \le L \\ L\,\mathrm{sign}(Y_i) & \text{otherwise} \end{cases}$$

and let m_L, m_{nL} be the functions m, and m_n when Y_i are replaced by Y_{iL}. Then

$$J_n = \int |m_n(x) - m(x)|\,\mu(\mathrm{d}x)$$
$$\le \int |m_n(x) - m_{nL}(x)|\,\mu(\mathrm{d}x) + \int |m_{nL}(x) - m_L(x)|\,\mu(\mathrm{d}x) + \int |m_L(x) - m(x)|\,\mu(\mathrm{d}x).$$

From Lemma 1, we know that

$$\int |m_{nL}(x) - m_L(x)|\,\mu(\mathrm{d}x) \to 0 \quad \text{a.s.} \tag{6}$$

and it is easy to see that for any $\varepsilon > 0$, if L is large enough, then

$$\int |m_L(x) - m(x)|\,\mu(\mathrm{d}x) = E(|E(Y_L|X) - E(Y|X)|) \le E(|Y - Y_L|) \le \varepsilon \quad \text{a.s.} \tag{7}$$

So, in order to prove $J_n \to 0$ a.s., we only need to prove that there is a large constant L such that

$$\limsup \int |m_n(x) - m_{nL}(x)|\,\mu(\mathrm{d}x) \le \varepsilon \quad \text{a.s.} \tag{8}$$

Introduce the notations

$$\overline{J_n} = \left\{i : \mu(A_{ni}) > 8\tfrac{\log n}{n}\right\}, \qquad B_n = \textstyle\bigcup_{i\in\overline{J_n}} A_{ni},$$

$$J_n = \left\{i : \mu_n(A_{ni}) > \tfrac{\log n}{n}\right\}, \qquad D_n = \textstyle\bigcup_{i\in J_n} A_{ni},$$
$$J_n^0 = \{i : \mu_n(A_{ni}) = 0\}, \qquad D_n^0 = \textstyle\bigcup_{i\in J_n^0} A_{ni},$$
$$J_n^1 = \left\{i : \tfrac{1}{n} \le \mu_n(A_{ni}) \le \tfrac{\log n}{n}\right\}, \qquad D_n^1 = \textstyle\bigcup_{i\in J_n^1} A_{ni},$$

$$W_{ni}(x) = \begin{cases} \frac{I_{\{X_i \in A_n(x)\}}}{n\mu_n(A_n(x))} & \text{if } \mu_n(A_n(x)) > 0 \\ \frac{1}{n} & \text{otherwise.} \end{cases}$$

With these notations,

$$\begin{aligned} \int |m_n(x) - m_{nL}(x)|\,\mu(dx) &\le \sum_{i=1}^{n} \int W_{ni}(x)\,\mu(dx)|Y_i - Y_{iL}| \\ &\le \sum_{i=1}^{n} \int_{B_n^c \cap D_n^1} W_{ni}(x)\,\mu(dx)|Y_i - Y_{iL}| \\ &+ \sum_{i=1}^{n} \int_{B_n \cup D_n} W_{ni}(x)\,\mu(dx)|Y_i - Y_{iL}| \\ &+ \sum_{i=1}^{n} \int_{D_n^0} W_{ni}(x)\,\mu(dx)|Y_i - Y_{iL}| \\ &\triangleq I_{n1} + I_{n2} + I_{n3}. \end{aligned}$$

Since

$$I_{n3} \le \frac{1}{n}\sum_{i=1}^{n} |Y_i - Y_{iL}| \to E|Y - Y_L| \quad \text{a.s.}$$

so, for sufficiently large L,

$$\limsup I_{n3} \le \frac{\varepsilon}{3}\,. \tag{9}$$

Note that

$$n\int_{B_n^c \cap D_n^1} W_{ni}(x)\,\mu(dx) = \frac{\mu(A_n(X_i))}{\mu_n(A_n(X_i))} I_{\{X_i \in B_n^c \cap D_n^1\}}\,.$$

For $q = \frac{p}{p-1}$ we get

$$\begin{aligned} I_{n1} &= \sum_{i=1}^{n} \frac{\mu(A_n(X_i))}{n\mu_n(A_n(X_i))} I_{\{X_i \in B_n^c \cap D_n^1\}} |Y_i - Y_{iL}| \\ &\le \left[\sum_{i=1}^{n} \frac{\mu^q(A_n(X_i))}{n^q \mu_n^q(A_n(X_i))} I_{\{X_i \in B_n^c \cap D_n^1\}}\right]^{1/q} \left[\sum_{i=1}^{n} |Y_i - Y_{iL}|^p\right]^{1/p} \\ &= \left[\sum_{i=1}^{n}\sum_{j=1}^{k_n} \frac{\mu^q(A_{nj})}{n^q \mu_n^q(A_{nj})} I_{\{X_i \in A_{nj}\}} I_{\{A_{nj} \subset B_n^c \cap D_n^1\}}\right]^{1/q} \left[\sum_{i=1}^{n} |Y_i - Y_{iL}|^p\right]^{1/p} \\ &= \left[\sum_{j \in \overline{J_n}^c \cap J_n^1} \frac{\mu^q(A_{nj})}{n^{q-1}\mu_n^{q-1}(A_{nj})}\right]^{1/q} \left[\sum_{i=1}^{n} |Y_i - Y_{iL}|^p\right]^{1/p} \\ &\le 8k_{n^{1/q}} \frac{\log n}{n^{1/q}} \left[\frac{1}{n}\sum_{i=1}^{n} |Y_i - Y_{iL}|^p\right]^{1/p} \end{aligned}$$

$$\leq \ 8\left[\frac{1}{n}\sum_{i=1}^{n}|Y_i - Y_{iL}|^p\right]^{1/p},$$

so when L is large enough

$$\limsup I_{n1} \leq \frac{\varepsilon}{3} \quad \text{a.s.} \tag{10}$$

Moreover

$$I_{n2} = \frac{1}{n}\sum_{i=1}^{n}\frac{\mu(A_n(X_i))}{\mu_n(A_n(X_i))}\, I_{\{X_i\in B_n\cup D_n\}}|Y_i - Y_{iL}|.$$

So for sufficiently large L

$$\limsup I_{n2} \leq \limsup \frac{8}{n}\sum_{i=1}^{n}|Y_i - Y_{iL}| = 8E|Y - Y_L| \leq \frac{\varepsilon}{3} \quad \text{a.s.}, \tag{11}$$

if we show that

$$\limsup_{n\to\infty}\left\{\max_{1\leq i\leq n}\frac{\mu(A_n(X_i))}{\mu_n(A_n(X_i))}\, I_{\{X_i\in B_n\cup D_n\}}\right\} \leq 8 \quad \text{a.s.} \tag{12}$$

It follows

$$\begin{aligned}
& P\left(\limsup_{n\to\infty}\left\{\max_{1\leq i\leq n}\frac{\mu(A_n(X_i))}{\mu_n(A_n(X_i))}\, I_{\{X_i\in B_n\cup D_n\}}\right\} > 8\right)\\
= \ & \lim_{k\to\infty} P\left(\bigcup_{n=k}^{\infty}\left(\max_{1\leq i\leq n}\frac{\mu(A_n(X_i))}{\mu_n(A_n(X_i))}\, I_{\{X_i\in B_n\cup D_n\}} > 8\right)\right)\\
\leq \ & \lim_{k\to\infty}\sum_{n=k}^{\infty} P\left(\max_{1\leq i\leq n}\frac{\mu(A_n(X_i))}{\mu_n(A_n(X_i))}\, I_{\{X_i\in B_n\cup D_n\}} > 8\right)\\
\leq \ & \lim_{k\to\infty}\sum_{n=k}^{\infty} nP\left(X_1\in D_n\cup B_n, \frac{\mu(A_n(X_1))}{\mu_n(A_n(X_1))} > 8\right)\\
= \ & \lim_{k\to\infty}\sum_{n=k}^{\infty} nP\left(X_1\in B_n, \frac{\mu(A_n(X_1))}{\mu_n(A_n(X_1))} > 8\right)
\end{aligned}$$

and

$$\begin{aligned}
& P\left(X_1\in B_n, \frac{\mu(A_n(X_1))}{\mu_n(A_n(X_1))} > 8\right)\\
= \ & \sum_{j\in\overline{J_n}} P\left(X_1\in A_{nj}, \frac{\mu(A_{nj})}{\mu_n(A_{nj})} > 8\right)\\
= \ & \sum_{j\in\overline{J_n}} P\left(\mu_n(A_{nj}) < \frac{\mu(A_{nj})}{8}\,\middle|\, X_1\in A_{nj}\right)\mu(A_{nj})
\end{aligned}$$

$$= \sum_{j \in \overline{J_n}} P\left(\sum_{i=2}^{n} I_{\{X_i \in A_{nj}\}} < \frac{n\mu(A_{nj})}{8} - 1 \right) \mu(A_{nj})$$
$$\leq \sum_{j \in \overline{J_n}} \mu(A_{nj}) \exp(-n\mu(A_{nj})) \leq n^{-2}$$

from Lemma 2 for sufficently large n. Then we get

$$P\left(\limsup_{n\to\infty} \left\{ \max_{1 \le i \le n} \frac{\mu(A_n(X_i))}{\mu_n(A_n(X_i))} I_{\{X_i \in B_n \cup D_n\}} \right\} > 8 \right) = 0\,,$$

so we completed the proof of the theorem. □

Lemma 3. (Devoye and Krzyzak [7]) Let $M_n'(x)$ be the kernel estimate with kernel K. If K satisfies the condition of Theorem 2, and $\lim_{n\to\infty} h^{\mathrm{d}} = 0$, $\lim_{n\to\infty} nh^{\mathrm{d}} = \infty$ then for every distribution of (X, Y) with bounded Y,

$$\int |M_n'(x) - m(x)|\, \mu(\mathrm{d}x) \to 0 \quad \text{a.s.}$$

P r o o f of the Theorem 2. For bounded Y, from Lemma 3, it follows that

$$\int |M_n'(x) - m(x)|\, \mu(\mathrm{d}x) \to 0 \quad \text{a.s.}$$

and

$$\int |M_n(x) - m(x)|\, \mu(\mathrm{d}x)$$
$$\leq \int |M_n(x) - M_n'(x)|\, \mu(\mathrm{d}x) + \int |M_n'(x) - m(x)|\, \mu(\mathrm{d}x)$$
$$\leq \int |M_n'(x) - m(x)|\, \mu(\mathrm{d}x) + L\mu(C_n) \to 0 \quad \text{a.s.},$$

where $C_n = \{x : \|x\| > t_n\}$. For unbouded Y, from now on, without loss of generality, assume $r = 1$, $h \leq 1$, $t_n \geq 2$. For an arbitrary L, let Y_{iL} be as in the proof of Theorem 1 and let m_L, M_{nL} be the functions m, and M_n when Y_i are replaced by Y_{iL}. Then

$$\int |M_n(x) - m(x)|\, \mu(\mathrm{d}x)$$
$$\leq \int |M_n(x) - M_{nL}(x)|\, \mu(\mathrm{d}x) + \int |M_{nL}(x) - m_L(x)|\, \mu(\mathrm{d}x) + \int |m_L(x) - m(x)|\, \mu(\mathrm{d}x).$$

We know that

$$\int |M_{nL}(x) - m_L(x)|\, \mu(\mathrm{d}x) \to 0 \quad \text{a.s.} \tag{13}$$

and, as in (7), if L is large enough, then

$$\int |m_L(x) - m(x)|\, \mu(\mathrm{d}x) \le \varepsilon. \tag{14}$$

So, in order to prove the theorem, we only need to prove that there is a constant L, such that

$$\limsup \int |M_n(x) - M_{nL}(x)|\, \mu(\mathrm{d}x) \le \varepsilon \quad \text{a.s.} \tag{15}$$

Thus

$$\int |M_n(x) - M_{nL}(x)|\, \mu(\mathrm{d}x) \le \sum_{i=1}^{n} V_{ni}|Y_i - Y_{iL}| + \frac{1}{n}\sum_{i=1}^{n} |Y_i - Y_{iL}|,$$

where

$$\begin{aligned} V_{ni} &= \int_{\|x\|\le t_n} \frac{K_h(x - X_i)}{\sum_{j=1}^{n} K_h(x - X_j)}\, \mu(\mathrm{d}x) \\ &\le \frac{C_2}{C_1}\int_{\|x\|\le t_n} \frac{I_{\{x\in X_i+S_h\}}}{\sum_{j=1}^{n} I_{\{X_j\in x+S_h\}}}\, \mu(\mathrm{d}x) \\ &= \frac{C_2}{C_1}\int_{\|x\|\le t_n} \frac{I_{\{x\in X_i+S_h\}}}{\sum_{j=1}^{n} I_{\{X_j\in x+S_h\}}} I_{\{\|X_i\|\le t_n+1\}}\, \mu(\mathrm{d}x) \\ &\le 2\frac{C_2}{C_1}\int_{\|x\|\le t_n} \frac{I_{\{x\in X_i+S_h\}}}{\sum_{j=1}^{n} I_{\{X_j\in x+S_h\}}+1} I_{\{\|X_i\|\le t_n+1\}}\, \mu(\mathrm{d}x). \end{aligned}$$

We split the cube $[-2t_n, 2t_n]^{\mathrm{d}}$ into small cubes with side length $\frac{h}{2}$, we get the set of small disjoint cubes $\mathcal{P}_n = \{A_{n1}, \dots, A_{nk_n}\}$, $k_n = [\frac{8t_n}{h}]^{\mathrm{d}} = \frac{8^{\mathrm{d}} n}{(\log n)^{q+1}}$. For any X_i with $\|X_i\| \le t_n + 1$, $A_n(X_i)$ is the small cube from $\mathcal{P}_n$ to which X_i belongs, so, there is a constant N ($N \le 5^{\mathrm{d}}$), such that $A_n^1(X_i), \dots, A_n^N(X_i) \in \mathcal{P}_n$, and $X_i + S_h \subset \bigcup_{m=1}^{N} A_n^m(X_i)$. Using $C' \triangleq 2\frac{C_2}{C_1}$, now we get

$$\begin{aligned} V_{ni} &\le C'\sum_{m=1}^{N}\int_{\|x\|\le t_n} \frac{I_{\{x\in A_n^m(X_i)\}}}{\sum_{j=1}^{n} I_{\{X_j\in x+S_h\}}+1} I_{\{\|X_i\|\le t_n+1\}}\, \mu(\mathrm{d}x) \\ &\le C'\sum_{m=1}^{N}\int_{\|x\|\le t_n} \frac{I_{\{x\in A_n^m(X_i)\}}}{\sum_{j=1}^{n} I_{\{X_j\in A_n^m(X_i)\}}+1} I_{\{\|X_i\|\le t_n+1\}}\, \mu(\mathrm{d}x) \\ &= C'\sum_{m=1}^{N} \frac{\mu(A_n^m(X_i)}{n\mu_n(A_n^m(X_i))+1} I_{\{\|X_i\|\le t_n+1\}}. \end{aligned}$$

Now let us introduce some notations

$$\begin{aligned} J &= \{i : \|X_i\| \le t_n + 1\}, \\ \overline{J}_{nm} &= \left\{i : i \in J, \mu(A_n^m(X_i) > C\frac{\log n}{n}\right\}, \\ J_{nm} &= \left\{i : i \in J, \mu_n(A_n^m(X_i) > \frac{\log n}{n}\right\}. \end{aligned}$$

For a set A $|A|$ is the number of the elements of set A. With these notations we get

$$\begin{aligned}
\sum_{i=1}^{n} V_{ni}|Y_i - Y_{iL}| &\leq \sum_{i\in J} C' \sum_{m=1}^{N} \frac{\mu(A_n^m(X_i)}{n\mu_n(A_n^m(X_i))+1} |Y_i - Y_{iL}| \\
&\leq C' \sum_{m=1}^{N} \sum_{i\in J_{nm}\cup \overline{J}_{nm}} \frac{\mu(A_n^m(X_i)}{n\mu_n(A_n^m(X_i))+1} |Y_i - Y_{iL}| \\
&+ C' \sum_{m=1}^{N} \sum_{i\in J_{nm}^c \cap \overline{J}_{nm}^c} \frac{\mu(A_n^m(X_i)}{n\mu_n(A_n^m(X_i))+1} |Y_i - Y_{iL}| \\
&\triangleq I_{n1} + I_{n2}.
\end{aligned}$$

For I_{n2}, note that $|J_{nm}^c| \leq k_n \log n$, similarly to the proof of (10), for sufficiently large L, we get

$$\limsup I_{n2} \leq \frac{\varepsilon}{3},$$

and as in the proof of (12) we can show

$$\limsup_{n\to\infty} \left\{ \max_{i\in J_{nm}\cup \overline{J}_{nm}} \frac{\mu(A_n^m(X_i))}{\mu_n(A_n^m(X_i))+1/n} \leq C \right\} \quad \text{a.s.}$$

so for sufficiently large L

$$\limsup I_{n1} \leq \limsup \frac{CC'N}{n} \sum_{i=1}^{n} |Y_i - Y_{iL}| = CC'NE|Y_i - Y_{iL}| \leq \frac{\varepsilon}{3} \quad \text{a.s.}$$

so we completed the proof of the theorem. □

ACKNOWLEDGEMENT

We want to thank Prof. László Györfi for his stimulating proposal and fruitful discussions leading to the present version of this note.

(Received August 22, 1994.)

REFERENCES

[1] J. Beck: The exponential rate of convergence of error for k_n NN nonparametric regression and decision. Problems Control Inform. Theory *8* (1979), 303–311.

[2] G. Collomb: Nonparametric regression: an up-to-date bibliography. Statistics *16* (1985), 300–324.

[3] H. Chernoff: A measure of asymptotic efficiency of tests of a hypothesis based on the sum of observations. Ann. Math. Statist. *23* (1952), 493–507.

[4] L. Devroye: Necessary and sufficient conditions for the almost everywhere convergence of nearest neighbor regression function estimates. Z. Wahrsch. verw. Geb. *61* (1982), 467–481.

[5] L. Devroye and L. Györfi: Distribution-free exponential bound on the L_1 error of partitioning estimates of a regression function. In: Proceedings of the Fourth Pannonian Symposium on Mathematical Statistics (F. Konecny, J. Mogyorodi, W. Wertz, eds.), Akademiai Kiado, Budapest 1983, pp. 67–76.
[6] L. Devroye, L. Györfi, G. Lugosi and A. Krzyżak: On strong universal consistency of nearest neighbor regression function estimates. Ann. Statist. To appear.
[7] L. Devroye and A. Krzyżak: An equivalence theorem for L_1 convergence of the kernel regression estimate. J. Statist. Plann. Inference *23* (1989), 71–82.
[8] L. Devroye and T. J. Wagner: Distribution-free consistency results in nonparametric discrimination and regression function estimation. Ann. Statist. *8* (1980), 231–239.
[9] M. Falk and R. D. Reiss: A Hellinger distance bound for the nearest neighbor approach in conditional curve estimation. Statist. Decisions, Supplement Issue *3* (1993), 55–68.
[10] L. Györfi: Universal consistencies of a regression estimate for unbounded regression functions. In: Nonparametric Functional Estimation (G. Roussas, ed.), NATO ASI Series, Springer–Verlag, Berlin 1991, pp. 329–338.
[11] E. A. Nadaraya: On estimating regression. Theory Probab. Appl. *9* (1964), 141–142.
[12] C. Spiegelman and J. Sacks: Consistent window estimation in nonparametric regression. Ann. Statist. *8* (1980), 240–246.
[13] C. J. Stone: Consistent nonparametric regression. Ann. Statist. *8* (1977), 1348–1360.
[14] G. S. Watson: Smooth regression analysis. Sankhyã Ser. A *26* (1964), 359–372.
[15] L. C. Zhao: Exponential bounds of mean error for the nearest neighbor estimates of regression functions. J. Multivariate Anal. *21* (1987), 168–178.

Zhang Shuang Lin, Department of Mathematics, Heilongjiang University, 150080, Harbin, China and Department of Mathematics, Technical University of Budapest, 1521 Stoczek u. 2, Budapest, Hungary.