Strong Consistency of Kernel Regression Estimate

Wenquan Cui, Meng Wei

Department of Statistics and Finance, University of Science and Technology of China, Hefei, China
Email: wqcui@ustc.edu.cn

Received February 20, 2013; revised March 23, 2013; accepted April 1, 2013

Copyright © 2013 Wenquan Cui, Meng Wei. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

ABSTRACT

In this paper, regression function estimation from independent and identically distributed data is considered. We establish strong pointwise consistency of the famous Nadaraya-Watson estimator under weaker conditions which permit to apply kernels with unbounded support and even not integrable ones and provide a general approach for constructing strongly consistent kernel estimates of regression functions.

Keywords: Kernel Regression Estimator; Bandwidth; Strong Pointwise Consistency

1. Introduction

Let \( (X_1, Y_1), \ldots, (X_n, Y_n) \) be independent observations of a \( \mathbb{R}^d \times \mathbb{R} \) valued random vector \((X, Y)\) with \( E|Y| < \infty \). We estimate the regression function \( m(x) = E(Y|X = x) \) by the following form of kernel estimates

\[
m_n(x) = \sum_{i=1}^{n} Y_i K((X_i - x)/h)/\sum_{j=1}^{n} Y_j K((X_j - x)/h) \quad (1.1)
\]

where \( h = h(n) \) is called the bandwidth and \( K \) is a given nonnegative Borel kernel. The estimator (1.1) was first introduced by Nadaraya ([1]) and Watson ([2]). The studies of \( m(x) \) can also refer to, for examples, Stone ([3]), Schuster and Yakowitz ([4]), Gasser and Muller ([5]), Mack and Muller ([6]), Greblicki and Pawlak ([7]), Kohler, Krzyżak and Walk ([8,9]), and Walk ([10]). When point \( x \) is near the boundary of their support, the kernel regression estimator (1.1) has suffered from a serious problem of boundary effects. Hereafter 0/0 is treated as 0. For the kernel function we assume that

\[
c_1 H(\|x\|) \leq K(x) \leq c_2 H(\|x\|), \quad (1.2)
\]

and

\[
c_1 I(\|x\| \leq r) \leq K(x), \quad (1.3)
\]

where \( c_1, c_2, c \) and \( r \) are positive constants, \( \|x\| \) is either always \( l_2 \) or always \( l_{\infty} \) norm, \( I(\cdot) \) denotes the indicator function of a set, and \( K \) is a bounded decreasing Borel function in \([0, \infty)\) such that

\[
t^p H(t) \to 0, \text{as } t \to \infty, \quad (1.4)
\]

Through this paper we assume that

\[
h = h(n) \to 0, \text{as } n \to \infty, \quad (1.5)
\]

One of the fundamental problems of asymptotic study on nonparametric regression is to find the conditions under which \( m_n(x) \) is a strongly consistent estimate of \( m(x) \) for almost all \((\mu)x \in \mathbb{R}^d \) \((\mu \text{ probability distribution of } X)\). The first general result in this direction belongs to Devroye ([11]), who established strong pointwise consistency of \( m_n(x) \) for bounded \( Y \). Zhao and Fang ([12]) establish its strong consistency under the weaker condition that \( E|Y| < \infty \) \( \text{for some } p > 1 \). However, the dominating function \( H(\|x\|) \) of (1.3) in the above literature is confined as \( I(\|x\| \leq r) \) for some \( r > 0 \). Greblicki, Krzyżak and Pawlak ([13]) establish the complete convergence of \( m_n(x) \) for bounded \( Y \) and rather general dominating function \( H(\|x\|) \) of (1.3) for almost all \((\mu)x \in \mathbb{R}^d \). This permits to apply kernels with unbounded support and even not integrable ones. In this paper, we establish the strong consistency of \( m_n(x) \) under the conditions of GKP ([13]) on the kernel and various moment conditions on \( Y \), which provides a general approach for constructing strongly consistent kernel estimates of regression functions. We have

**Theorem 1.1** Assume that \( E|Y| < \infty \) \( \text{for some } p > 1 \), and (1.2)-(1.5) are satisfied, and that

\[
nh^d/(n^{p/d} \log n) \to 0, \text{as } n \to \infty. \quad (1.6)
\]

Then

\[
m_n(x) \to m(x), \text{a.s. for almost all } (\mu)x \in \mathbb{R}^d, \quad (1.7)
\]

as \( n \to \infty \).
Theorem 1.2 Assume that $E \exp\left(\frac{|y|^4}{4}\right) < \infty$ for some $\lambda > 0$ and $t > 0$, and (1.2)-(1.5) are met, and that

$$nh^d/\log n\to \infty, \text{ as } n\to \infty. \quad (1.8)$$

Then (1.7) is true.

It is worthwhile to point out that in the above theorems we do not impose any restriction on the probability distribution $\mu$ of $X$.

2. Proof of the Theorems

For simplicity, denote by $c$ a positive constant, by $c(x)$ a positive constant depending on $x$. These constants may assume different values in different places, even within the same expression. We denote by $S_r$ as a sphere of the radius $r$ centered at $x$, $x \in \mathbb{R}^d$.

**Lemma 2.1** Assume that $\lim_{n \to \infty} h(n) = 0$. For all $c > 0$, there exists a nonnegative function $q(x)$ with $q(x) < \infty$ such that for almost all $(\mu)x \in \mathbb{R}^d$,

$$h^\delta/\mu(S_n) \to q(x), \text{ as } n \to \infty$$

Refer to Devroye ([11]).

**Lemma 2.2** Assume that (1.2)-(1.5) are satisfied. Let $f/\mu$ be integrable for some $r > 0$. Then

$$\int K \left(\frac{y-x}{h}\right)f(y) - f(x) \mu(dy) / \int K \left(\frac{y-x}{h}\right) \mu(dy) \to 0$$

as $h \to 0$ for almost all $(\mu)x \in \mathbb{R}^d$.

It is easily proved by using Lemma 1 of GKP ([13]).

**Lemma 2.3** Assume that (1.2)-(1.5) are met, and that $nh^d/\log n \to \infty$, as $n \to \infty$.

Then for almost all $(\mu)x \in \mathbb{R}^d$

$$B_n(x) = \left(nE\left(\|X-x\|/h\right)\right)^{-1} \sum_{i=1}^n K((X_i-x)/h) \to 1$$

a.s. as $n \to \infty$

Refer to GKP ([13]).

Now we are in a position to prove Theorems 1.1 and 1.2.

**Proof**. For simplicity, we write “for a.e. $x$” instead of the longer phrase “for almost all $(\mu)x \in \mathbb{R}^d$”. Write

$$N = nE\left(\|X-x\|/h\right),$$

$$B^*_n(x) = \left(N^{-1} \sum_{i=1}^n (Y_i - m(x))K((X_i-x)/h)\right)$$

$$U_n(x) = \left(N^{-1} \sum_{i=1}^n m(X_i) - m(x)K((X_i-x)/h)\right),$$

$$T_n(x) = \left(N^{-1} \sum_{i=1}^n (Y_i - m(X_i))K((X_i-x)/h)\right).$$

Since

$$m_n(x) = \left(B_{2n}(x) + m(x)B_n(x)\right)/(1 + B_n(x)),$$

and by Lemma 2.3, $B_n(x) \to 1$ a.s. for a.e. $x$, it suffices to verify that $B_{2n}(x) \to 0$ a.s. for a.e. $x$, or, to prove $U_n(x) \to 0$ a.s. and $T_n(x) \to 0$ a.s. for a.e. $x$.

Since $|y|^p$ is convex in $y$ for $p > 1$, and for fixed $t > 0$ and $\lambda > 0$, $\exp(|y|^p)/I(y > a)$ is convex in $y \in (a, +\infty)$ for large $a$, it follows from Jensen’s inequality that $E\left|m(X) - m(x)\right|^p < \infty$ and $E\left|Y - m(X)\right|^p < \infty$ when $E\left|Y\right|^p < \infty$, and that

$$E\exp\left(s\left|m(X) - m(x)\right|^p\right) < \infty$$

and

$$E\exp\left(s\left|Y - m(X)\right|^p\right) < \infty$$

for some $s > 0$ and $\lambda > 0$ when $E\left|Y\right|^p < \infty$.

Write $g_j = G\left(X_j\right) = \left|m(X_j) - m(x)\right|$, $c_j = j^{1/p}$ (in Theorem 1.1) or $c_j = \left((1/s)\log j\right)^{1/p}$ (in Theorem 1.2). It follows that

$$\sum_j P(g_j > c_j) < \infty,$$

and

$$P(g_j > c_j, i.o.) = 0$$

by Borel-Cantelli’s lemma, and

$$\sum_j g_j I(g_j > c_j) < \infty, \text{ a.s.} \quad (2.1)$$

Write $a_n \gg b_n$, if $a_n/b_n \to \infty$ as $n \to \infty$. By (1.6) or (1.8), $nh^d/c_n \gg \log n$, we can take $d_n$ such that

$$nh^d/c_n \gg d_n \gg \log n, \quad (2.2)$$

Put

$$g^*_j = g_j I(g_j > c_j), U^*_n(x) = N^{-1} \sum_{j=1}^n g^*_j K((X_j-x)/h),$$

$$g^*_n = g_j I(g_j \leq c_j), U^*_n(x) = N^{-1} \sum_{j=1}^n g^*_j K((X_j-x)/h),$$

$$\tilde{g}_n = g_j I\left(N^{-1}K((X_j-x)/h)\right) g_j \leq a_n^{-1},$$

$$\tilde{U}_n(x) = N^{-1} \sum_{j=1}^n \tilde{g}_j K((X_j-x)/h).$$

By (1.3) and Lemma 2.1, for a.e. $x$,

$$N^{-1} = \left(nE\left(\|X-x\|/h\right)\right)^{-1} \leq \left(cn\mu(S_n)\right)^{-1} \leq c(x)\left(nh^d\right)^{-1} \to 0. \quad (2.3)$$

By Lemma 2.3,
By Schwarz’s inequality, (2.1), (2.3) and (2.4),
\[
(U_n(x))^2 \leq N^{-2} \sum_{j=1}^{N} K((X_j - x)/h) \Rightarrow 0 \quad \text{a.s. for a.e. } x
\]
(2.5)

Write
\[
Z_j = d_n N^{-1/2} (\hat{g}_n K((X_j - x)/h) - E \hat{g}_n K((X_j - x)/h))
\]
and for \( n \) large,
\[
E \exp\left(\sum_{j=1}^{N} |Z_j|^2\right) \leq \exp\left\{c d_n^{-1} N^{-b} n EK((X_1 - x)/h)\right\}.
\]
By Lemma 2.2,
\[
EK((X_i - x)/h) |g_i|^2 / EK((X_i - x)/h) = \left[K \left(\frac{y-x}{h}\right) m(y) - m(x)\right] \mu(dy) / \int K \left(\frac{y-x}{h}\right) \mu(dy),
\]
\[\Rightarrow 0 \quad \text{for a.e. } x, \text{ as } n \to \infty
\]
for (2.7)

By (2.2) and (2.3),
\[
(d_n/N)^{b-1} \ll c(x) (d_n/(nh^d))^{b-1} \ll (1/c_n)^{b-1} \to 0
\]
for a.e. x.

Given \( d > 0 \), it follows that for a.e. x and for \( n \) large,
\[
c d_n^{b} N^{-b} n EK((X_i - x)/h) |g_i|^2 < d e/2, \quad (2.6)
\]
and
\[
P\{U_n(x) - E \hat{U}_n(x) \geq \varepsilon\} \leq e^{-d \varepsilon} e^{d \varepsilon/2} \varepsilon^{-d/e}.
\]
By Borel-Cantelli’s lemma and for a.e. x,
\[
\sum_n P\{U_n(x) - E \hat{U}_n(x) \geq \varepsilon\} < \infty \quad \text{for any } \varepsilon > 0,
\]
we have
\[
\hat{U}_n(x) - E \hat{U}_n(x) \to 0 \quad \text{a.s. for a.e. } x
\]
Since, by Lemma 2.2, for a.e. x
\[
E \hat{U}_n(x) \leq E m(x) - m(x) K((X_i - x)/h) / EK((X_i - x)/h) \to 0,
\]
we have
\[
\hat{U}_n(x) \to 0 \quad \text{a.s. for a.e. } x, \text{ as } n \to \infty.
\]
By (2.2) and (2.3), when \( g_j \leq c_j \), for a.e. x,
\[
N^{-1} K((X_j - x)/h) g_j \leq c(x) (nh^d)^{-1} c_n \ll d_n^{-1},
\]
and for \( n \) large, \( g_j^2 \ll \hat{g}_n \), \( 1 \leq j \leq n \), and
\[
0 \leq U_n(x) \leq \hat{U}_n(x) \to 0 \quad \text{a.s. for a.e. } x. \quad (2.8)
\]
By (2.5) and (2.8), noticing that
\[
U_n(x) = U_n(x) + U_n^*(x),
\]
we have
\[
U_n(x) \to 0 \quad \text{a.s. for a.e. } x. \quad (2.9)
\]
To prove \( T_n(x) \to 0 \quad \text{a.s. for a.e. } x \), we write
\[
e_j = Y_j - m(X_j),
\]
and put
\[
e_j = e_j I(e_j > c_j),
\]
\[
T_n(x) = N^{-2} \sum_{j=1}^{N} e_j^2 K((X_j - x)/h)
\]
\[
e_j = e_j I(e_j < c_j),
\]
\[
\hat{c}_n = e_j I(N^{-1} K((X_j - x)/h) e_j \ll d_n^{-1}),
\]
\[
\hat{F}(x) = N^{-1} \sum_{j=1}^{N} \hat{c}_n K((X_j - x)/h).
\]
By using the same argument as above,
\[
\sum_j P\{|e_j| > c_j\} < \infty
\]
\[
P(e_j > c_j), a.s.
\]
\[
\sum_j e_j I(e_j > c_j) < \infty
\]
and for a.e. x,
\[
(T_n^*(x))^2 \leq N^{-2} \sum_{j=1}^{N} K^2((X_j - x)/h) \sum_{j=1}^{N} e_j^2 I(e_j > c_j)
\]
\[
0 \quad \text{a.s.}
\]
Also, for a.e. x and for \( n \) large,
\[
e_j^2 \ll \hat{c}_n \leq \hat{c}_n \quad 1 \leq j \leq n \quad \text{and} \quad T_n^*(x) \leq \hat{F}(x).
\]
Write
\[
Z_j = d_n N^{-1} (\hat{g}_n K((X_j - x)/h) - E \hat{g}_n K((X_j - x)/h)).
\]
Copyright © 2013 SciRes.
$Z_j \leq 1, E[Z_j] \leq 0, j = 1, 2, \ldots, n$. Take $1 < b \leq 2, b < p$. Since $e^z \leq 1 + z + |z|^p$ for $z \leq 1$, we have

$$Ee^{Z_j} \leq 1 + E|Z_j|^p \leq \exp\left(E|Z_j|^p\right),$$

and

$$E \exp\left(d_n \hat{T}_n(x)\right) \leq \exp\left(\sum_{j=1}^n |Z_j|^p\right) \leq \exp\left(cd_n^p N^{-b} n E(K((X-x)/h)|Y|^p)\right).$$

By Lemma 2.2, for a.e. $x$,

$$E(K((X-x)/h)|Y|^p) = \lim_{n \to \infty} \frac{E(K((X-x)/h)|Y|^p)}{E(K((X-x)/h))} = \frac{E(K((X-x)/h)|Y|^p)}{E(K((X-x)/h))} \to h_j(x)$$

Given $c > 0$, similar to (2.6), for a.e. $x$ and $n$ large, $cd_n^p N^{-b} n E(K((X-x)/h)|Y|^p) \leq d_n c/2$, and

$$P(\hat{T}_n(x) \geq c) \leq e^{-dc} E \exp\left(d_n \hat{T}_n(x)\right) \leq e^{-dc} \cdot e^{-dc/2} = e^{-3dc/2},$$

and it follows that

$$\limsup_{n \to \infty} \hat{T}_n(x) \leq 0 \quad \text{a.s. for a.e. } x$$

(2.12)

from

$$\sum_{n} P(\hat{T}_n(x) \geq c) < \infty \quad \text{for a.e. } x \text{ and } \forall c > 0$$

and Borel-Cantelli’s lemma.

By (2.10)-(2.12),

$$\limsup_{n \to \infty} T_n(x) \leq 0 \quad \text{a.s. for a.e. } x$$

and

$$\liminf_{n \to \infty} T_n(x) \geq 0 \quad \text{a.s. for a.e. } x$$

(2.13)

Replacing $e_j$ by $(-e_j)$, it implies that

$$\liminf_{n \to \infty} T_n(x) \geq 0 \quad \text{a.s. for a.e. } x$$

(2.14)

(2.13) and (2.14) give

$$T_n(x) \to 0 \quad \text{a.s. for a.e. } x$$

(2.15)

The theorems follow from (2.9) and (2.15).

3. Acknowledgements

Cui’s research was supported by the Natural Science Foundation of Anhui Province (Grant No.1308085MA02), the National Natural Science Foundation of China (Grant No. 10971210), and the Knowledge Innovation Program of Chinese Academy of Sciences (KJCX3-SYW-S02).

REFERENCES


