Open Journal of Statistics, 2013, 3, 179-182
http://dx.doi.org/10.4236/ojs.2013.33020 Published Online June 2013 (http://www.scirp.org/journal/ojs)

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

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 $(X_1,Y_1),\cdots,(X_n,Y_n)$ be independent observations of a $R^d\times R$ 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

$$m_n(x)=\sum_{i=1}^{n}Y_iK\left((X_i-x)/h\right)\Big/\sum_{j=1}^{n}Y_iK\left((X_j-x)/h\right) \quad (1.1)$$

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 Müller ([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

$$c_1H(\|x\|)\le K(x)\le c_2H(\|x\|), \quad (1.2)$$

and

$$cI(\|x\|\le r)\le K(x), \quad (1.3)$$

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

$$t^dH(t)\to 0, \text{as } t\to\infty, \quad (1.4)$$

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 $(\mu)x\in R^d$ (μ 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 $E|Y|^p<\infty$ for some $p>1$. However, the dominating function $H(\|x\|)$ of (1.3) in the above literature is confined as $I(\|x\|\le r)$ 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 of (1.3) for almost all $(\mu)x\in R^d$. 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* $E|Y|^p<\infty$ *for some* $p>1$, *and* (1.2)-(1.5) *are satisfied, and that*

$$nh^d\Big/\left(n^{1/p}\log n\right)\to\infty, \text{as } n\to\infty. \quad (1.6)$$

Then

$$m_n(x)\to m(x)\ a.s. \text{ for almost all } (\mu)x\in R^d, \text{ as } n\to\infty. \quad (1.7)$$

Theorem 1.2 *Assume that* $E\exp\left(t|Y|^{\lambda}\right)<\infty$ *for some* $\lambda>0$ *and* $t>0$, *and* (1.2)-(1.5) *are met, and that*

$$nh^d\Big/(\log n)^{1+\frac{1}{\lambda}}\to\infty,\text{as } n\to\infty. \quad (1.8)$$

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 μ 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*, $x\in R^d$.

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* $(\mu)\,x\in R^d$,

$$h^d/\mu(S_{ch})\to q(x),\text{as } n\to\infty$$

Refer to Devroye ([11]).

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

$$\int K\left(\frac{y-x}{h}\right)\left|f(y)-f(x)\right|^r\mu(\mathrm{d}y)\Big/\int K\left(\frac{y-x}{h}\right)\mu(\mathrm{d}y)\to 0$$

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

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

Lemma 2.3 *Assume that* (1.2)-(1.5) *are met*, *and that*

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

Then for almost all $(\mu)\,x\in R^d$

$$B_{1n}(x)\equiv\left(nEK\left((X-x)/h\right)\right)^{-1}\sum_{i=1}^{n}K\left((X_i-x)/h\right)\to 1$$

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 $(\mu)\,x\in R^d$". Write

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

$$B_{2n}(x)=N^{-1}\sum_{i=1}^{n}\left(Y_i-m(x)\right)K\left((X_i-x)/h\right),$$

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

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

Since

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

and by Lemma 2.3, $B_{1n}(x)\to 1$ a.s. for a.e. *x*, it suffices to vertify 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 $\lambda>0$, $\exp\left(ty^{\lambda}\right)I(y>a)$ 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 $E|Y|^p<\infty$, and that $E\exp\left(s\left|m(X)-m(x)\right|^{\lambda}\right)<\infty$ and $E\exp\left(s\left|Y-m(X)\right|^{\lambda}\right)<\infty$ for some $s>0$ and $\lambda>0$ when $E\exp\left(|Y|^{\lambda}\right)<\infty$.

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

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

and

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

by Borel-Cantelli's lemma, and

$$\sum_j g_j^2 I\left(g_j>c_j\right)<\infty,\quad \text{a.s.} \quad (2.1)$$

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

$$g_j'=g_jI\left(g_j>c_j\right),U_n'(x)=N^{-1}\sum_{j=1}^{n}g_j'K\left((X_j-x)/h\right)$$

$$g_j''=g_jI\left(g_j\le c_j\right),U_n''(x)=N^{-1}\sum_{j=1}^{n}g_j''K\left((X_j-x)/h\right),$$

$$\tilde g_{nj}=g_jI\left(N^{-1}K\left((X_j-x)/h\right)g_j\le d_n^{-1}\right),$$

$$\tilde U_n(x)=N^{-1}\sum_{j=1}^{n}\tilde g_{nj}K\left((X_j-x)/h\right).$$

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

$$N^{-1}=\left(nEK\left((X-x)/h\right)\right)^{-1}\le\left(cn\mu(S_{rh})\right)^{-1}\le c(x)\left(nh^d\right)^{-1}\to 0. \quad (2.3)$$

By Lemma 2.3,

$$N^{-1}\sum_{j=1}^{n}K^2\left(\left(X_j-x\right)/h\right)$$
$$\le cN^{-1}\sum_{j=1}^{n}K\left(\left(X_j-x\right)/h\right)\to c \qquad (2.4)$$

a.s. for a.e. x

By Schwarz's inequality, (2.1), (2.3) and (2.4),

$$\left(U_n'(x)\right)^2$$
$$\le N^{-2}\sum_{j=1}^{n}K^2\left(\left(X_j-x\right)/h\right)\sum_{j=1}^{n}g_j^2 I\left(g_j>c_j\right) \qquad (2.5)$$
$$\to 0$$

a.s. for a.e. x

Write

$$Z_j=d_nN^{-1}\left(\tilde{g}_{nj}K\left(\left(X_j-x\right)/h\right)-E\tilde{g}_{nj}K\left(\left(X_j-x\right)/h\right)\right).$$

We have $EZ_j=0, Z_j\le 1, j=1,2,\cdots,n.$ Take $b<p, 1<b\le 2$. Since $e^z\le 1+z+|z|^b$ for $z\le 1$, we have

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

and

$$E\exp\left(d_n\left(\tilde{U}_n(x)-E\tilde{U}_n(x)\right)\right)$$
$$\le\exp\left(\sum_{j=1}^{n}\left|Z_j\right|^b\right)\le\exp\left\{cd_n^bN^{-b}nEK\left(\left(X_1-x\right)/h\right)\left|g_1\right|^b\right\}.$$

By Lemma 2.2,

$$EK\left(\left(X_1-x\right)/h\right)\left|g_1\right|^b/EK\left(\left(X-x\right)/h\right)$$
$$=\int K\left(\frac{y-x}{h}\right)\left|m(y)-m(x)\right|^b\mu(\mathrm{d}y)\Big/\int K\left(\frac{y-x}{h}\right)\mu(\mathrm{d}y).$$
$$\to 0$$

for a.e. x, as $n\to\infty$

By (2.2) and (2.3),

$$\left(d_n/N\right)^{b-1}\ll c(x)\left(d_n/\left(nh^d\right)\right)^{b-1}\ll\left(1/c_n\right)^{b-1}\to 0$$

for a.e. x.

Given $\varepsilon>0$, it follows that for a.e. x and for n large,

$$cd_n^bN^{-b}nEK\left(\left(X_1-x\right)/h\right)\left|g_1\right|^b<d_n\varepsilon/2, \qquad (2.6)$$

and

$$P\left\{\tilde{U}_n(x)-E\tilde{U}_n(x)\ge\varepsilon\right\}$$
$$\le e^{-d_n\varepsilon}E\exp\left\{d_n\left(\tilde{U}_n(x)-E\tilde{U}_n(x)\right)\right\}$$
$$\le e^{-d_n\varepsilon}\cdot e^{-d_n\varepsilon/2}=e^{-3d_n\varepsilon/2}.$$

By Borel-Cantelli's lemma and for a.e. x,

$$\sum_n P\left(\tilde{U}_n(x)-E\tilde{U}_n(x)\ge\varepsilon\right)<\infty \quad \text{for any } \varepsilon>0,$$

we have

$$\tilde{U}_n(x)-E\tilde{U}_n(x)\to 0 \quad \text{a.s for a.e. } x$$

Since, by Lemma 2.2, for a.e. x

$$E\tilde{U}_n(x)$$
$$\le E\left|m(X)-m(x)\right|K\left(\left(X-x\right)/h\right)/EK\left(\left(X-x\right)/h\right)\to 0,$$

we have

$$\tilde{U}_n(x)\to 0 \quad \text{a.s for a.e. } x, \text{ as } n\to\infty. \qquad (2.7)$$

By (2.2) and (2.3), when $g_j\le c_j$, for a.e. x,

$$N^{-1}K\left(\left(X_j-x\right)/h\right)g_j\le c(x)\left(nh^d\right)^{-1}\cdot c\cdot c_n\ll d_n^{-1},$$

and for n large, $g_j''\le\tilde{g}_{nj}, 1\le j\le n,$ and

$$0\le U_n''(x)\le\tilde{U}_n(x)\to 0 \quad \text{a.s. for a.e. } x. \qquad (2.8)$$

By (2.5) and (2.8), noticing that $U_n(x)=U_n'(x)+U_n''(x)$, we have

$$U_n(x)\to 0 \quad \text{a.s for a.e. } x. \qquad (2.9)$$

To prove $T_n(x)\to 0$ a.s for a.e. x, we write $e_j=Y_j-m\left(X_j\right)$, and put

$$e_j'=e_jI\left(e_j>c_j\right), T_n'(x)=N^{-1}\sum_{j=1}^{n}e_j'K\left(\left(X_j-x\right)/h\right)$$
$$e_j''=e_jI\left(e_j\le c_j\right), T_n''(x)=N^{-1}\sum_{j=1}^{n}e_j''K\left(\left(X_j-x\right)/h\right),$$
$$\tilde{e}_{nj}=e_jI\left(N^{-1}K\left(\left(X_j-x\right)/h\right)e_j\le d_n^{-1}\right),$$
$$\tilde{T}_n(x)=N^{-1}\sum_{j=1}^{n}\tilde{e}_{nj}K\left(\left(X_j-x\right)/h\right).$$

By using the same argument as above,

$$\sum_j P\left(\left|e_j\right|>c_j\right)<\infty$$
$$P\left(e_j>c_j, i.o.\right)=0 \quad \text{a.s.}$$
$$\sum_j e_j^2 I\left(e_j>c_j\right)<\infty$$

and for a.e. x,

$$\left(T_n'(x)\right)^2$$
$$\le N^{-2}\sum_{j=1}^{n}K^2\left(\left(X_j-x\right)/h\right)\sum_{j=1}^{n}e_j^2I\left(e_j>c_j\right) \qquad (2.10)$$
$$\to 0$$

a.s.

Also, for a.e. x and for n large,

$$e_j''\le\tilde{e}_{nj}, 1\le j\le n \quad \text{and} \quad T_n''(x)\le\tilde{T}_n(x). \qquad (2.11)$$

Write $Z_j=d_nN^{-1}K\left(\left(X_j-x\right)/h\right)\tilde{e}_{nj}$, then

$Z_j \le 1, EZ_j \le 0, j = 1,2,\cdots,n$. Take $1 < b \le 2, b < p$. Since $e^z \le 1 + z + |z|^b$ for $z \le 1$, we have

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

and

$$E\exp\left(d_n\tilde{T}_n(x)\right) \le \exp\left(\sum_{j=1}^{n}|Z_j|^b\right) \le \exp\left(cd_n^bN^{-b}nEK\left((X-x)/h\right)|Y|^b\right).$$

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

$$EK\left((X-x)/h\right)|Y|^b / EK\left((X-x)/h\right) = EK\left((X-x)/h\right)h_b(x) / EK\left((X-x)/h\right) \to h_b(x)$$

Given $\varepsilon > 0$, similar to (2.6), for a.e. *x* and *n* large, $cd_n^bN^{-b}nEK\left((X-x)/h\right)|Y|^b < d_n\varepsilon/2$, and

$$P\left(\tilde{T}_n(x) \ge \varepsilon\right) \le e^{-d_n\varepsilon}E\exp\left(d_n\tilde{T}_n(x)\right) \le e^{-d_n\varepsilon} \cdot e^{-d_n\varepsilon/2} = e^{-3d_n\varepsilon/2},$$

and it follows that

$$\limsup_{n\to\infty}\tilde{T}_n(x) \le 0 \quad \text{a.s. for a.e. } x \qquad (2.12)$$

from

$$\sum_n P\left(\tilde{T}_n(x) \ge \varepsilon\right) < \infty \quad \text{for a.e. } x \text{ and } \forall\varepsilon > 0$$

and Borel-Cantelli's lemma.

By (2.10)-(2.12),

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

and

$$\limsup_{n\to\infty} T_n(x) \le 0 \quad \text{a.s. for a.e. } x \qquad (2.13)$$

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

$$\liminf_{n\to\infty} T_n(x) \ge 0 \text{ a.s. for a.e. } x \qquad (2.14)$$

(2.13) and (2.14) give

$$T_n(x) \to 0 \quad \text{a.s. for a.e. } x \qquad (2.15)$$

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

[1] E. A. Nadaraya, "On Estimating Regression," *Theory of Probability and Its Applications*, Vol. 9, No. 1, 1964, pp. 141-142. doi:10.1137/1109020

[2] G. S. Watson, "Smooth Regression Analysis," *Sankhya*: *The Indian Journal of Statistics*, *Series A*, Vol. 26, No. 4, 1964, pp. 359-372.

[3] C. J. Stone, "Consistent Nonparametric Regression," *Annals of Statistics*, Vol. 5, No. 4, 1977, pp. 595-620. doi:10.1214/aos/1176343886

[4] E. F. Schuster and S. Yakowitz, "Contributions to the Theory of Nonparametric Regression, with Application to System Identification," *Annals Statistics*, Vol. 7, No. 1, 1979, pp. 139-149. doi:10.1214/aos/1176344560

[5] T. Gasser and H. G. Müller, "Estimating Regression Functions and Their Derivatives by the Kernel Method," *Scandinavian Journal of Statistics*, Vol. 11, 1984, pp. 171-185.

[6] Y. P. Mack and H. G. Müller, "Derivative Estimation in Nonparametric Regression with Random Predictor," *Sankhya*: *The Indian Journal of Statistics* (*Series A*), Vol. 51, No. 1, 1989, pp. 59-72.

[7] W. Greblicki and M. Pawlak, "Cascade Non-Linear System Identification by a Non-Parametric Method," *International Journal of Systems Science*, Vol. 25, No. 1, 1994, 129-153. doi:10.1080/00207729408928949

[8] M. Kohler, A. Krzyżak and H. Walk, "Strong Consistency of Automatic Kernel Regression Estimates," *Annals of the Institute of Statistical Mathematics*, Vol. 55, No. 2, 2003, pp. 287-308.

[9] M. Kohler, A. Krzyżak and H. Walk, "Rates of Convergence for Partition-Zing and Nearest Neighbor Regression Estimates with Unbounded Data," *Journal of Multivariate Analysis*, Vol. 97, No. 2, 2006, pp. 311-323. doi:10.1016/j.jmva.2005.03.006

[10] H. Walk, "Strong Universal Consistency of Smooth Kernel Regression Estimates," *Annals of the Institute of Statistical Mathematics*, Vol. 57, No. 4, 2005, pp. 665-685. doi:10.1007/BF02915432

[11] L. Devroye, "On the Almost Everywhere Convergence of Nonparametric Regression Function Estimates," *Annals Statistics*, Vol. 9, No. 6, 1981, pp. 1310-1319. doi:10.1214/aos/1176345647

[12] L. C. Zhao and Z. B. Fang, "Strong Convergence of Kernel Estimates of Nonparametric Regression Functions," *Chinese Annals of Mathematics*, *Series B*, Vol. 6, No. 2, 1985, pp. 147-155.

[13] W. Greblicki, A. Krzyżak and M. Pawlak, "Distribution-Free Pointwise Consistency of Kernel Regression Estimate," *Annals Statistics*, Vol. 12, No. 4, 1984, pp. 1570-1575. doi:10.1214/aos/1176346815

