Journal of Science and Engineering Technology, Vol. 2, No. 1, pp. 61-66 (2006)

Uniform Consistency and Convergence Rate in the Local Polynomial Estimation of Regression

WANN-JYI HORNG
Department of Finance, Ling-Tung University
1 Ling-Tung Rd., Nantun, Taichung, Taiwan

ABSTRACT

One studies the almost sure limiting behavior and convergence rate of the kernel regression estimator. By the local polynomial fit method of Fan and Gijbels [2] to construct a general weight kernel regression estimator. In this paper, the almost sure limiting behavior and the convergence rate of the proposed estimator are given. As the domain of density is compactly supported, the proposed estimator can be improved the problem of boundary effects, this is, it does not also need to adjust the boundary regions. Besides, the proposed estimator can also improve the bias and its convergent rate is achieved at $O(h^{p+1}+\frac{\sqrt{\log(1/h)}}{\sqrt{nh}})$, for all $x\in[a,b]$ or real line and $p\geq 1$.

Key Words: kernel regression estimator, boundary effects, bandwidth, uniform consistency, convergence rate, bias reduction

均勻一致性與其收斂速率在局部多項式迴歸之估計

洪萬吉
嶺東科技大學財金學系
台中市南屯區嶺東路一號

摘　要

本文研究核迴歸估計量之殆必（almost sure）的極限行爲與收斂速率。本文依 Fan 與 Gijbels [2] 的局部多項式擬合法之觀點來建立一個一般化之加權核迴歸估計量，此估計量之極限行爲與收斂速率將被提供。當密度函數之定義域爲有界時，所提之估計量可以改善邊界效果，即不需要在去調整邊界區域。此外，所提之估計量也可以改善估計偏誤，其收斂速率是達到，對所有的 $x\in[a,b]$ 或實數線與 $p\geq 1$。

關鍵詞：核迴歸估計量，邊界效果，帶寬，均勻一致性，收斂速率，偏差減小

I. INTRODUCTION

Let $(X_1,Y_1),\ldots,(X_n,Y_n)$ be a sequence of independently and identically distributed (i.i.d.) random variables with common joint probability density function $f(x,y)=f(x)f(y|x)$. $f(x)$ is the density function of random variable X and $f(y|x)$ the conditional density function of Y, given $X=x$. Let $m(x)=E(Y\mid X=x)$ be the regression curve. In the literature, various estimators for $m(x)$, based on a random sample $(X_1,Y_1),\ldots,(X_n,Y_n)$, have been proposed and their properties are studied, for examples, the books of Wand and Jones [18], Simonoff [14], Fan and Gijbels [3] and Prakasa Rao [12]. A traditional kernel regression estimate for $m(x)$ is defined as follows:

$$\hat{m}(x;h)=\frac{\sum_{j=1}^{n}K\{\frac{x-X_j}{h}\}Y_j}{\sum_{j=1}^{n}K\{\frac{x-X_j}{h}\}}, \qquad (1)$$

where K is a (symmetric) kernel function and $h=h(n)$ is called the bandwidth.

This estimator (1) was first introduction by Nadaraya [11] and Watson [19]. The studies of m(x) can also refer to, for examples, Gasser and Müller [4], Schuster and Yakowitz [15], Mack and Silverman [9] and Stone [17]. When point x near the boundary of their support, the kernel regression estimator (KRE) (1) has been suffered to a serious problem of boundary effects. Some boundary modification methods have been proposed by, for examples, Fan and Gijbels [1], Fan [2], Gasser and Müller [4], Müller [8] and Rice [13].

In this paper, researcher will relax the constraint that the kernel regression estimator (KRE) has to be a bona fide kernel density, and propose a general weighted KRE which it does not encounter the boundary effects. The proposed method are uses the idea of local polynomial fit method (Fan and Gijbels [1]). The method of local polynomial fit is assumed as follows:

Min $Q(\vec{\alpha})$

$$=Min_{\vec{\alpha}}\sum_{i=1}^{n}\left(Y_i-\alpha_0-\alpha_1(x-X_i)-\ldots-\alpha_p(x-X_i)^p\right)^2K\left\{\frac{x-X_i}{h}\right\},$$

where $\vec{\alpha}'=(\alpha_0,\alpha_1,\ldots,\alpha_p)$, $\alpha_0,\ldots,\alpha_p$ are the unknown parameters , for all p≥1.

According to local polynomial fit method, we minimize $Q(\vec{\alpha})$ with respect to $\vec{\alpha}$, denoting the solution of $\vec{\alpha}$ by $\hat{\vec{\alpha}}$. Further to calculate and use the Cramer's rule for $\hat{\alpha}_0$. Denoting $\hat{\alpha}_0=\hat{m}_{p+1}$, we can get a general weighted kernel regression estimator which is given by

$$\hat{m}_{p+1}(x;h)=\frac{1}{nh}\sum_{j=1}^{n}\frac{det(U_{p+1}(\frac{x-X_j}{h}))}{det(A_{p+1})}K\left\{\frac{x-X_j}{h}\right\}Y_j$$

$$=\frac{1}{nh}\sum_{j=1}^{n}W\left\{\frac{x-X_j}{h}\right\}Y_j, \qquad (2)$$

where $W(u)=\dfrac{det(U_{p+1}(u))}{det(A_{p+1})}K(u)$, A_{p+1} and $U_{p+1}(\cdot)$ are $(p+1)\times(p+1)$ matrices, $det(\cdot)$ denotes the determinant. Here A_{p+1} and $U_{p+1}(\cdot)$ are defined as below:

$$A_{p+1}=\begin{bmatrix} S_0 & S_1 & S_2 & \cdots & S_p \\ S_1 & S_2 & S_3 & \cdots & S_{p+1} \\ . & . & . & \cdots & . \\ . & . & . & \cdots & . \\ . & . & . & \cdots & . \\ S_p & S_{p+1} & S_{p+2} & \cdots & S_{2p} \end{bmatrix},$$

and

$$U_{p+1}(\frac{x-X_i}{h})=\begin{bmatrix} 1 & S_1 & \cdots & S_p \\ (x-X_i)/h & S_2 & \cdots & S_{p+1} \\ . & . & \cdots & . \\ . & . & \cdots & . \\ . & . & \cdots & . \\ ((x-X_i)/h)^p & S_{p+1} & \cdots & S_{2p} \end{bmatrix},$$

$$S_j=\frac{1}{nh}\sum_{i=1}^{n}K\left\{\frac{x-X_i}{h}\right\}(\frac{x-X_i}{h})^j, \qquad (3)$$

for all $j=0,1,\ldots,2p$ and $p\geq1$.

Here researcher uses the estimator of $\hat{m}_{p+1}$(x; h) to estimate $m(x)$, for all $p\geq1$. We will study the limiting behavior and convergence rates of estimator (2) such that it does have the uniformly consistency property.

From (2), we known that it does have a nice property which is given by

$$\sum_{i=1}^{n}\frac{1}{h}K\left\{\frac{x-X_i}{h}\right\}\det\left(U_{p+1}(\frac{x-X_i}{h})\right)\ (\frac{x-X_i}{h})^m=0, \qquad (4)$$

for all $m=1,2,\ldots,\ldots,p\geq1$. We know that the above result by the linear algebra theory.

This paper is organized as follows: In section 2, we state the main results of estimator (2) and give the explicit formula for convergence rate. In section 3, we give the proofs of the main results.

II. MAIN RESULTS

In this section, one states the main results of estimator (2). Before one states the main results one will give the following assumptions:

(A1) $(X_1,Y_1),\ldots,(X_n,Y_n)$, as before, is i.i.d. with the joint density function $f(x,y)$, $f(x)$ and $m(x)$ are bounded on its domain, and continuous at the point x, $m^{(l)}(x)$ and $f^{(l)}(x)$ exists, for $l \geq p+1$ and $p\geq 1$. $m^{(l)}(x)$ and $f^{(l)}(x)$ are the l-times derivative of m(x) and f(x), respectively.

(A2) $W(u) = \dfrac{det(U_{p+1}(u))}{det(A_{p+1})}K(u)$ is bounded variation, $K(\cdot)$ is the kernel function and satisfies

a. $\limsup\limits_{|u|\to\infty} \left| u^{k+p+4} K(u) \right| < \infty$, for all $k\geq 0$ and $p\geq 1$,

b. $\int \left| u \log|u| \right|^{1/2} |dW(u)| < \infty$,

c. $W(u) \to 0$, as $|u| \to \infty$.

(A3) a. $\sum\limits_n h^{\lambda}(n) = \sum h^{\lambda}$ converges for every $\lambda > 0$ (Mack and Silverman [9]).

b. $E|Y|^s < \infty, \sup\limits_x \int |y|^s f(x,y)dy < \infty$, for some s>q+1 and q>1.

For conveniences, let

$$
\mathrm{N}_{p+1}(u) = \begin{bmatrix} 1 & \mu_1 & \cdots & \mu_p \\ u & \mu_2 & \cdots & \mu_{p+1} \\ \vdots & \vdots & \ddots & \vdots \\ u^p & \mu_{p+1} & \cdots & \mu_{2p} \end{bmatrix},
$$

$$
D_{p+1} = \begin{bmatrix} \mu_o & \mu_1 & \mu_2 & \cdots & \mu_p \\ \mu_1 & \mu_2 & \mu_3 & \cdots & \mu_{p+1} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ \mu_p & \mu_{p+1} & \mu_{p+2} & \cdots & \mu_{2p} \end{bmatrix},
$$

$$
E_{p+1} = \begin{bmatrix} \mu_{p+1} & \mu_1 & \mu_2 & \cdots & \mu_p \\ \mu_{p+2} & \mu_2 & \mu_3 & \cdots & \mu_{p+1} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ \mu_{2p+1} & \mu_{p+1} & \mu_{p+2} & \cdots & \mu_{2p} \end{bmatrix},
$$

where $\mu_j = \int u^j K\{u\}du$, for j=0,1,2,…, 2p+1 and p≥1.

Let us now give the asymptotic results of the estimator (2) at the point $x\in(-\infty,\infty)$. In the following Theorem 1, one gives the asymptotic bias and variance of the estimator (2), respectively.

Theorem 1:

Assume that conditions (A1) and (A2)-(a) are satisfied. And suppose that $h\to 0$ and $nh\to\infty$, as $n\to\infty$, then the asymptotic bias and variance of the estimator (2) are given by

$$
\text{Bias}[\tilde{m}_{p+1}(\text{x; h})] = \frac{(-1)^{p+1}}{(p+1)!} h^{p+1} \frac{|E_{p+1}|}{|D_{p+1}|} m^{(p+1)}(x) + o(h^{p+1}), \tag{5}
$$

$$
\text{Var}[\tilde{m}_{p+1}(\text{x; h})] = \frac{1}{nh} \frac{v^2(x)}{f(x)} \frac{\int K^2(t)[|N_{p+1}(t)|]^2 dt}{[|D_{p+1}|]^2} + o(\frac{1}{nh}), \tag{6}
$$

where $v^2(x) = E(Y^2 \mid X = x)$, E_{p+1}, D_{p+1} and N_{p+1} are defined as above, for p ≥ 1.

In the following Theorem 2, one gives the strong convergence of the estimator (2) at the point $x\in(-\infty,\infty)$.

Theorem 2:

Assume that conditions (A1)-(A3) are satisfied. And suppose that $h\to 0$ and $nh\to\infty$, as $n\to\infty$, then we have

$$
\sup_{x\in(-\infty,\infty)} \left| \hat{m}_{p+1}(x;h) - m(x) \right| \to 0, \tag{7}
$$

almost surely, as $n\to\infty$.

Let $[a,b]$ be the compact support of $f(x)$. In the following Theorem 3, one gives the strong convergence of the estimator (2) at the point $x\in[a,b]$, for example, a=0 and b=1.

Theorem 3:

Assume that conditions (A1)-(A3) are satisfied. And suppose that $h\to 0$ and $nh\to\infty$, as $n\to\infty$, then we have

$$
\sup_{x\in[a,b]} \left| \hat{m}_{p+1}(x;h) - m(x) \right| \to 0, \tag{8}
$$

almost surely, as $n\to\infty$.

Remark 1:

From Theorem 2, the estimator (2) has the strong convergence property as that of the traditional kernel density estimator as p=1. From Theorem 3, as the domain of $f(x)$ is compactly supported, the strong convergence of $\hat{m}_{p+1}(x;h)$ is also held. The problem of boundary effects, the details can refer to the paper of Fan and Gijbels [1]. In Theorem 1, one provides the general form of asymptotic bias and variance.

If $f^{(p+1)}(x)$ and $m^{(p+1)}(x)$ are uniformly continuous functions on $(-\infty,\infty)$, then, we have the following theorem about the convergence rate of the estimator $\hat{m}_{p+1}(x;h)$.

Theorem 4:

Under the assumptions of Theorem 2 and assume that

$f^{(p+1)}(x)$ and $m^{(p+1)}(x)$ are uniformly continuous, then we have

$$\sup_{x\in(-\infty,\infty)} \left| \hat{m}_{p+1}(x;h) - m(x) \right| = O\left(h^{p+1} + \frac{\sqrt{log(1/h)}}{\sqrt{n}\sqrt{h}} \right), \tag{9}$$

for $p\geq 1$.

If $f^{(p+1)}(x)$ and $m^{(p+1)}(x)$ are uniformly continuous functions on $[a,b]$, then we have the following theorem about the convergence rate of the estimator $\hat{m}_{p+1}(x;h)$.

Theorem 5:

Under the assumptions of Theorem 3 and assume that $f^{(p+1)}(x)$ and $m^{(p+1)}(x)$ are uniformly continuous, then we have

$$\sup_{x\in[a,b]} \left| \hat{m}(x;h) - m(x) \right| = O\left(h^{p+1} + \frac{\sqrt{log(1/h)}}{\sqrt{n}\sqrt{h}} \right), \tag{10}$$

for $p\geq 1$.

Remark 2:

From Theorem 4-5, the convergence rate is constructed for the estimator (2). Choosing the smoother parameter h one can refer to, for examples, the books of Silverman [16] and Härdle [5].

Remark 3:

From Theorem 1 and Theorem 4, when the kernel function $K(\cdot)$ is symmetric, we have

$$\sup_{x\in(-\infty,\infty)} \left| \hat{m}_{p+1}(x;h) - m(x) \right| = O\left(h^{2p} + \frac{\sqrt{log(1/h)}}{\sqrt{n}\sqrt{h}} \right), \tag{11}$$

for $p\geq 1$. The proposed estimator can be also reduced the estimation error.

III. PROOFS

In this section our main purpose is to prove Theorem 1-5. First, we introduce four auxiliary lemmas as follows:

Lemma 1:

Under the conditions (A1)-(A2), A(3)-(a) and assume that $f(x)$ is uniformly continuous, then the estimator (3) the asymptotic results as follows:

$$\sup_{x} \left| S_j - E(S_j) \right| = O\left(\frac{\sqrt{\log(1/h)}}{\sqrt{n}\sqrt{h}} \right), \tag{12}$$

where $E(S_j) = \frac{1}{h}\int K\left\{\frac{x-y}{h}\right\}(\frac{x-y}{h})^j f(y)dy$, for j $\geq$ 0 and x$\in[a,b]$ or real line.

Proof: The proof follows from the definition of estimator (3) and the Theorem B of Silverman [16].

Lemma 2:

Assume that the conditions (A1)-(A2) are satisfied. Then we have

$$E\hat{m}_{p+1}(x;h) - m(x) = O(h^{p+1}), \tag{13}$$

for $p\geq 1$ and $x\in(-\infty,\infty)$ or $x\in[a,b]$.

Proof: By the calculation of expectation, as $x\in(-\infty,\infty)$, we have

$$E\hat{m}_{p+1}(x;h)$$

$$= E\left[\frac{1}{n}\sum_{i=1}^{n}\frac{\det(U_{p+1}(\frac{x-X_i}{h}))}{\det(A_{p+1})}\frac{1}{h}K\left\{\frac{x-X_i}{h}\right\}m(X_i)\right]$$

$$= E\left[\frac{1}{n}\sum_{i=1}^{n}\frac{\det(U_{p+1}(\frac{x-X_i}{h}))}{\det(A_{p+1})}\frac{1}{h}K\left\{\frac{x-X_i}{h}\right\}\sum_{j=0}^{p}\frac{(-1)^j(x-X_i)}{j!}m^{(j)}(x)\right]$$

$$+ E\left[\frac{1}{n}\sum_{i=1}^{n}\frac{\det(U_{p+1}(\frac{x-X_i}{h}))}{\det(A_{p+1})}\frac{1}{h}K\left\{\frac{x-X_i}{h}\right\}\{m(X_i)\right.$$

$$\left.-\sum_{j=0}^{p}\frac{(-1)^j(x-X_i)^j}{j!}m^{(j)}(x)\}\right]$$

$$= m(x)E\{\frac{\det(A_{p+1})}{\det(A_{p+1})}\}$$

$$+ E\{\frac{1}{n}\sum_{i=1}^{n}\frac{\det(U_{p+1}(\frac{x-X_i}{h}))}{\det(A_{p+1})}\frac{1}{h}K\left\{\frac{x-y}{h}\right\}[m(X_i)$$

$$-\sum_{j=1}^{p}\frac{(x-X_i)^j(-1)^j}{j!}m^{(j)}(x)]\}$$

$$= m(x) + \frac{det(E_{p+1})}{det(D_{p+1})}\frac{h^{p+1}}{(p+1)!}(-1)^{p+1}m^{(p+1)}(x) + o(h^{p+1})$$

$$= m(x) + O(h^{p+1}), \tag{14}$$

from the property of (4) and the techniques of calculus underlying the conditions (A1)-(A2) , for $p\geq 1$. Similarly, for $x\in[a,b]$ case, for a and b are finite, or by the paper of Fan and

Gijbels [1], we can also prove that

$$E\hat{m}_{p+1}(x;h) - m(x) = O(h^{p+1}) \tag{15}$$

Therefore, the proof of Lemma 2 is proved.

Lemma 3:

Under the conditions (A1)-(A3), and suppose that $n^{2\eta-1}h \to \infty$, $\eta < 1 - \frac{1}{s}$, then we have

$$\sup_x \left| \hat{m}_{p+1}(x;h) - E\hat{m}_{p+1}(x;h) \right| = O\left(\frac{\sqrt{log(1/h)}}{\sqrt{n}\sqrt{h}}\right), \tag{16}$$

almost surely, for all $x \in [a,b]$or real line.

Proof: The proof follows from the Lemma 1 and the Proposition 4 of Mack and Silverman [9].

Proof of Theorem 4. We know that

$$\sup_x \left| \hat{m}_{p+1}(x;h) - m(x) \right| \le \sup_x \left| \hat{m}_{p+1}(x;h) - E\hat{m}_{p+1}(x;h) \right|$$

$$+ \sup_x \left| E\hat{m}_{p+1}(x;h) - m(x) \right|$$

$$= \mathrm{I}(1) + \mathrm{I}(2) \tag{17}$$

From I(1) , Lemma 1 and Lemma 3, we have

$$\sup_x \left| \hat{m}_{p+1}(x;h) - E\hat{m}_{p+1}(x;h) \right| = O\left(\frac{\sqrt{\log(1/h)}}{\sqrt{n}\sqrt{h}}\right) \tag{18}$$

From I(2) and Lemma 2, we have

$$\sup_x \left| E\hat{m}_{p+1}(x;h) - m(x) \right| = O\left(\frac{\sqrt{\log(1/h)}}{\sqrt{n}\sqrt{h}}\right). \tag{19}$$

From (17)-(19), the proof of Theorem 3 is completed.

Proof of Theorem 5. The proof of Theorem 5 is similar to that of Theorem 4, the details are omitted.

Proof of Theorem 2-3. The proof of Theorem 2-3 is also similar to that of Theorem 3, the details are omitted.

Proof of Theorem 1. The proof of Theorem 1 is also similar to that of Theorem 1 and the evaluation of variance, the details are omitted.

IV. CONCLUSION

In this paper we study the properties of uniform convergence. The proposed estimator can help us to solve the problem of boundary effects underlying the compact support of domain. The rate of convergence of $\hat{m}_{p+1}(x)$ is achieved at $O\left(h^{p+1} + \frac{\sqrt{\log(1/h)}}{\sqrt{nh}}\right)$, for all $x \in [a,b]$ or real line and $p \ge 1$.

REFERENCES

1. Fan, J. (1993) Local linear regression smoothers and their minimax efficiencies. *Annals Statistics*, 21, 196-216.
2. Fan, J. and I. Gijbels (1992) Variable bandwidth and local linear regression smoothers. *Annals Statistics*, 20, 2008-2036.
3. Fan, J. and I. Gijbels (1996) *Local Polynomial Modelling and Its Application*, Chapman & Hall, London.
4. Gasser, T. and H. G. Müller (1984) Estimating regression functions and their derivatives by the kernel method. *Scandinavian Journal of Statistics*, 11, 171-185.
5. Härdle, W. (1991) *Smoothing Techniques with Implementation in S*, Springer-Verlag, New York, NY.
6. Jones, M. C. (1993) Simple boundary correction for kernel density estimation. *Statistics Computing*, 3, 135-146.
7. Jones, M. C. and P. J. Foster (1996) A simple nonnegative boundary correction method for kernel density estimation. *Statistica Sinica*, 6, 1005-1013.
8. Mack,Y. P. and B.W. Silverman (1982) Weak and strong uniform consistency of kernel regression estimates. *Z. Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 61, 405-415.
9. Mack, Y. P. and H. G. Müller (1989) Derivative estimation in nonparametric regression with random predictor. *Sankhya: The Indian Journal of Statistics (Series A)*, 51, 59-72.
10. Müller, H. G. (1993) On the boundary kernel method for nonparametric curve estimation near endpoints. *Scandinavian Journal of Statistics*, 20, 313-328.
11. Nadaraya, E. A. (1964) On estimating regression. *Theory of Probability & Its Applications*, 9, 141-142.
12. Prakasa Rao, B. L. S. (1983) *Nonporametic Functional Estimation*. Academic Press, New York, NY.
13. Rice, J. (1984) Boundary modification for kernel regression. *Communication in Statistics: Theory and Methods*, 13, 893-900.
14. Schuster, E. F. and S. Yakowitz (1979) Contributions to the theory of nonparametric regression, with application to system identification. *Annals Statistics*, 7,139-149.
15. Silverman, B. W. (1978) Weak and strong uniform consistency of the kernel estimate of a density and its derivatives. *Annals Statistics*, 6, 177-184.
16. Simonoff, J. S. (1996) *Smoothing Methods in Statistics*, Springer-Verlag, New York, NY.
17. Stone, C. J. (1977) Consistent nonparametric regression.

Journal of Science and Engineering Technology, Vol. 2, No. 1, 2006

Annals Statistics, 5, 595-620.

18. Wand, M. P. and M. C. Jones (1995) *Kernel Smoothing*, Chapman and Hall, London.
19. Watson, G. S. (1964) Smooth regression analysis. Sankhy*ā*: *The Indian Journal of Statistics, Series A*, 26, 359-372.

Received: Jul. 8, 2004 Revised: Oct. 11, 2004
Accepted: Dec. 12, 2004