응용통계연구 제 6권 2호
1993년 9월 pp.329-339.

Modification of Boundary Bias in Nonparametric Regression[1)]

Kyung-Joon, Cha[2)]

Abstract

Kernel regression is a nonparametric regression technique which requires only differentiability of the true function. If one wants to use the kernel regression technique to produce smooth estimates of a curve over a finite interval, one can realize that there exist distinct boundary problems that detract from the global performance of the estimator. This paper develops a kernel to handle boundary problem. In order to develop the boundary kernel, a generalized jackknife method by Gray and Schucany (1972) is adapted. Also, it will be shown that the boundary kernel has the same order of convergence rate as non-boundary.

1. Introduction

One simple but interesting version of the nonparametric regression problem is to consider modelling the independent variable, y_i, by

$$y_i = m(t_i) + \varepsilon_i, \qquad i = 1,2,\ldots,n,$$

where the t_i are the non-stochastic design points and satisfying $0 \le t_1 < t_2 < \cdots < t_n \le 1$, the ε_i are independent, identically distributed random variables with zero mean and common variance σ^2, m is an unknown regression function in $C^p[0,1]$ for some integer $p \ge 2$ and $C^p[0,1]$ is the collection of all p times continuously differentiable functions on the unit interval. Without having assume more about m than it satisfies such smoothness conditions, we may want to estimate $m(t)$ at some fixed argument t

There are many interesting nonparametric estimators for $m(t)$. Examples of these can be found in Eubank (1988) and Gasser and Müller (1979). In particular, the class of kernel estimators of $m(t)$ proposed by Priestley and Chao (1972) and examined by Müller and Stadtmüller (1987) is defined by

$$\hat{m}(t;\ h) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h}K\left(\frac{t-t_i}{h}\right)y_i \tag{1.1}$$

where non-stochastic design points, t_i, are equally spaced and $h > 0$ is the bandwidth or

1) Research was supported by the Daeyang Foundation Grant of Sejong University
2) Department of Mathematics, Hanyang University, Haengdang-Dong 17, Seongdong-Ku, Seoul, 133-791

window width that determines which observations are included in $\hat{m}(t;\, h)$. The function K is called a kernel function. It is supported and symmetric on [-1,1]. When it satisfies

$$\int_{-1}^{+1} z^j K(z)dz = \begin{cases} 1 & j = 0 \\ 0 & j = 1,2,\cdots,p-1 \\ k_p \neq 0 & j = p, \end{cases} \qquad (1.2)$$

it is called a kernel of order p.

The bandwidth is considered to be an element of sequence indexed by n. The sequence may be different for each fixed $0 \le t \le 1$. That is, $h = h(n)$ and the corresponding kernel estimator of $m(t)$ is consistent if $h \to 0$ and $nh \to \infty$ as $n \to \infty$ (Gasser and Müller, 1979). In general, there are two types of kernel estimators. If the bandwidth depends on t, the estmator is called a local-bandwidth kernel estimator, otherwise it is called a global-bandwidth kernel estimator. Figure 1 shows the overlay of kernel weight at $t=0.4$ using a second order kernel, Epanechnikov kernel. By moving a kernel over a finite interval and averaging weighted observations, a kernel estimate of the curve over that interval is obtained.

The goal of this paper is to find an optimal local bandwidth kernel that is suitable for boundary cases where the optimal implies the local bandwidth that minimizes the asymptotic mean square error at a fixed t. Before explaining further, formal definition for non-boundary and boundary situations is given below.

<u>Definition</u>

Suppose that t is a fixed design point and h is a fixed bandwidth for a kernel estimator (1.1). Whenever the support of the kernel, $[t-h, t+h]$ is contained in the interval [0,1], it is called a non-boundary or interior situation. Whereas, if some part of the support of the kernel, $[t-h, t+h]$, is no longer contained in [0,1], it is said to be in the boundary.

Figure 1 above shows non-boundary situation. However, we easily can see the boundary situation from Figure 1. When the kernel estimator moves to estimate a point of the curve close the both ends, some part of the kernel estimator stays outside of interval [0,1]. Thus, some part of the kernel estimator attempts to smooth the curve with no observation.

2. Kernel Regression Estimators

2.1 Non-Boundary Cases

First, let us investigate the non-boundary case of the kernel estimator (1.1). Using the results from Müller (1988) and the Taylor's expansion of $m(t)$ about t, it can be shown that if $m \in C^p[0,1]$, then the expected value of $\hat{m}$ at a fixed t is

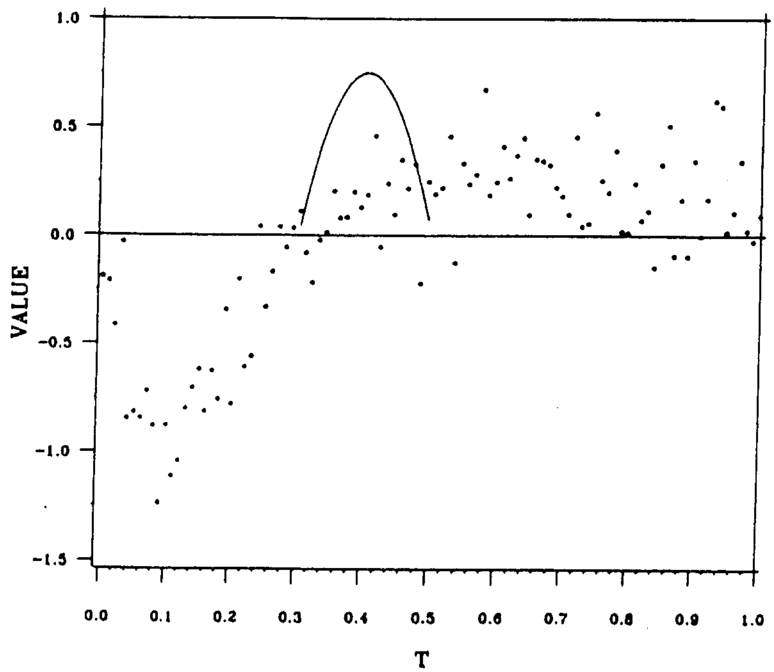


Figure 1. Overlay of Kernel Weight at t=0.4
Solid Line: Kernel Function Dots: Observations

$$E[\hat{m}(t;h)] = m(t) + \frac{(-1)^p}{p!} h^p m^{(p)}(t) k_p + o(h^p).$$

Hence, the asymptotic bias of $\hat{m}(t;h)$ is

$$E[\hat{m}(t;h)] - m(t) = \frac{(-1)^p}{p!} h^p m^{(p)}(t) k_p + o(h^p). \tag{2.1}$$

Now, let us find the asymptotic variance of $\hat{m}(t;h)$. By the method similar to those leading to (2.1)

$$\begin{aligned} \mathrm{var}[\hat{m}(t;h)] &= \frac{1}{n^2h^2} \sum_{i=1}^{n} K^2\left(\frac{t-t_i}{h}\right) \mathrm{var}(y_i) \\ &= \frac{\sigma^2}{nh} \int_{-1}^{+1} K^2(z)dz + o\left(\frac{1}{nh}\right) \end{aligned} \tag{2.2}$$

with sufficiently small h and $z = (t-s)/h$.

Therefore, the mean squre error of the estimator (1.1) is obtained as

$$mse[\hat{m}(t;h)] = \frac{\sigma^2}{nh} Q + \left[\frac{h^p}{p!} m^{(p)}(t) k_p\right]^2 + o\left(\frac{1}{nh}\right) + o(h^{2p}) \tag{2.3}$$

where $Q = \int_{-1}^{+1} K^2(z)dz$. Hence, by ignoring the vanishing terms and using Lemma 4a of

Parzen (1962), the optimal bandwidth which minimizes the asymptotic $mse[\hat{m}(t; h)]$ is

$$h_{opt} = \left\{ \frac{\sigma^2 Q}{2pn(k_p m^{(p)}(t)/p!)^2} \right\}^{\frac{1}{(2p+1)}} . \tag{2.4}$$

2.2 Boundary Cases

2.2.1 Introduction

It is known when points near the boundary of the support of the function are estimated, nonparametric regression function estimators usually show a sharp increase in variance and bias. Gasser and Müller (1979) proposed a so-called minimum variance boundary kernel, not an optimal boundary kernel. Also, Rice (1984) proposed another approach which uses the technique as Richardson extrapolation. More recently, Eubank and Speckman (1989) have given a method for removing boundary effects using a "bias reduction theorem". The fundamental idea of their work is to use a biased estimator to improve another estimator in some sense.

Let us consider the same kernel estimator as (1.1) with $p = 2$ and assume the same conditions as (1.2) hold for the kernel estimator. In this section, a new boundary kernel is proposed. Hence, it will be shown that the bias of a proposed boundary kernel estimator has the same order of convergence rate as in the interior. Therefore, the best rate of convergence to be anticipated for mean square error of this boundary kernel is the same as that of the non-boundary kernel. To do this, a combination of "cut-and-normalize" and a generalized jackknife method is used so that the bias will be $O(h^2)$.

For a fixed t and bandwidth h, define

$$q = \min\{\frac{t}{h}, 1\} \qquad \text{if } t \in [0, \frac{1}{2})$$

$$q = \min\{\frac{1-t}{h}, 1\} \qquad \text{if } t \in (\frac{1}{2}, 1]$$

where q is a real number such that $q \in [0,1]$. The two q values above represent the left and right boundary effects, respectively. For $t = q/h$, the support of the kernel estimator, $\hat{m}(t; h)$, that is mapped into the interval [0,1] is [−1,q) instead of [−1,1] as for an interior point. Hence, the "cut-and-normalize" modification omits that part of kernel lying between q and 1 and normalizes the kernel between −1 and q. The result is a boundary kernel. Then, the linear combination of two different such boundary kernels, which is a generalized jackknife estimator, gives a bias that has the same order as in the interior. In general, a generalized jackknife estimator is used for reduction of bias in an estimator. Thus, replacing an estimator by a kernel estimator, the same order as in the interior can be achieved at the boundary. (See p36 of Gray and Schucany, 1972)

2.2.2 Modification of Boundary Bias

In this section, only the left boundary effets, i.e., $q=t/h<1$, will be considered for simplicity, but the right boundary effects can be developed in the same manner.

Now, to examine the details of the general approach described above, let

$$\hat{m}_1(t;\ h) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h}K_1\left(\frac{t-t_i}{h}\right)y_i$$

where $K_1(z)$ is a second order kernel function supported on [-1,1]. Then, by the same steps as in (2.1)

$$E[\hat{m}_1(t;\ h)] = \int_{(t-1)/h}^{t/h} K_1(z)\left\{m(t) - zhm^{(1)}(t) + \frac{(zh)^2}{2!}m^{(2)}(t)\right\}dz + o(h^2)$$

where $z=(t-s)/h$. If t is a fixed value in $[0, h)$, then the left boundary problem occurs. Thus, this expression for the expectation becomes

$$E[\hat{m}_1(t;\ h)] = \int_{-1}^{q} K_1(z)\left\{m(t) - zhm^{(1)}(t) + \frac{(zh)^2}{2!}m^{(2)}(t)\right\}dz + o(h^2) \quad (2.5)$$

where $q=t/h$. Since $t\in[0,h)$, the symmetry of the kernel is lost by the left boundary effect and $\int_{-1}^{q}K_1(z)dz \neq 1$ and $\int_{-1}^{q}zK_1(z)dz \neq 0$. Hence, no terms vanish and (2.5) can not be reduced further.

Therefore, let us define a boundary kernel estimator of $\hat{m}_1(t;\ h)$ as

$$\hat{m}_{1q}(t;\ h) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{n}K_{1q}\left(\frac{t-t_i}{h}\right)y_i$$

where $K_{1q}(z)$ is a boundary kernel function defined by "cutting" ,i.e.,

$$K_{1q}(z) = \frac{K_1(z)}{\int_{-1}^{q}K_1(u)du} \qquad -1\leq z\leq q.$$

Obviously, this is "normalized" in the sense that it satisfies $\int_{-1}^{q}K_{1q}(z)dz=1$ Then, the corresponding estimator has expectation

$$E[\hat{m}_{1q}(t;\ h)] = m(t) - hm^{(1)}(t)k_{1q}^{(1)} + \frac{h^2m^{(2)}(t)}{2!k_{1q}^{(2)}} + o(h^2) \quad (2.6)$$

where $k_{1q}^{(1)} = \int_{-1}^{q}zK_{1q}(z)dz \neq 0$ and $k_{1q}^{(2)} = \int_{-1}^{q}z^2K_{1q}(z)dz$. Therefore, the bias of $\hat{m}_{1q}(t;\ h)$ would be

$$\text{bias}\left[\hat{m}_{1q}(t;\ h)\right] = -hm^{(1)}(t)k_{1q}^{(1)} + \frac{h^2m^{(2)}(t)}{2!}k_{1q}^{(2)} + o(h^2).$$

Recall that for the non-boundary the dominant part of bias $[\hat{m}(t;\ h)]$ is of order h^2. However, the dominant part of bias $[\hat{m}_{1q}(t;\ h)]$ is of order h, so in this sense this normalization is still subject to a boundary bias.

The asymptotic variance of $\hat{m}_{1q}(t;\ h)$ can be obtained by the same method as for the non-boundary, (2.2), i.e.,

$$\mathrm{var}\ [\ \hat{m}_{1q}(t;\ h)\ ] = \frac{\sigma^2}{nh} \int_{-1}^{q} K_{1q}^2\ (z)\ dz + o\left(\frac{1}{nh}\right).$$

Hence, the asymptotic mean square error has the form

$$mse[\ \hat{m}_{1q}(t;\ h)] = \frac{\sigma^2 Q_1}{nh} + [hm^{(1)}(t)k_{1q}^{(1)}]^2 + o\left(\frac{1}{nh}\right) + o(h^2)$$

where $Q_1 = \int_{-1}^{q} K_{1q}^2(z)dz$. Therefore, the best rate of convergence to be anticipated for the $mse[\hat{m}_{1q}(t;\ h)]$ of $\hat{m}_{1q}(t;\ h)$ is $n^{-3/4}$. This can be compared to $n^{-4/5}$ that would be obtained from (2.3) and (2.4) with $p = 2$ if there was no boundary effect. Thus, in order to have the same local asymptotic behavior, the generalized jackknife method can be applied to reduce the order of the bias.

To apply this method, let us first define another kernel estimator as

$$\hat{m}_2(t;\ h) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{h} K_2\left(\frac{t-t_i}{h}\right) y_i$$

where $K_2(\,z)$ is the second order kernel function supported on [-1,1] and not identically equal to $K_1(\,z)$.

Now, define another boundary kernel estimator

$$\hat{m}_{2q}(t;\ h) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{h} K_{2q}\left(\frac{t-t_i}{h}\right) y_i$$

where $K_{2q}(\,z)$ is a similarly renormalized version defined by

$$K_{2q}(z) = \frac{K_2(z)}{\int_{-1}^{q} K_2(u)du}, \qquad -1 \le\ z\ \le q.$$

Hence, similar to (2.6),

$$\mathrm{E}[\hat{m}_{2q}(t;\ h)] = m(t) - hm^{(1)}(t)k_{2q}^{(1)} + \frac{h^2 m^{(2)}(t)}{2!} k_{2q}^{(2)} + o(h^2) \tag{2.7}$$

Then, obviously, (2.6) and (2.7) can be rewritten as

$$\mathrm{E}[\hat{m}_{1q}(t;h)] = m(t) - hm^{(1)}(t)k_{1q}^{(1)} + O(h^2)$$
$$\mathrm{E}[\hat{m}_{2q}(t;h)] = m(t) - hm^{(1)}(t)k_{2q}^{(1)} + O(h^2)$$

Now, let $K_q^*(\,z)$ be a kernel which is a linear combination of $K_{1q}(\,z)$ and $K_{2q}(\,z)$ such that

$$K_q^*(z) = \frac{\begin{vmatrix} K_{1q}(\,z) & K_{2q}(\,z) \\ k_{1q}^{(1)} & k_{2q}^{(1)} \end{vmatrix}}{\begin{vmatrix} 1 & 1 \\ k_{1q}^{(1)} & k_{2q}^{(1)} \end{vmatrix}} = \frac{K_{1q}(z) - rK_{2q}(\,z)}{1-r}, \qquad -1 \le\ z \le\ q$$

where $|\ \mathrm{M}\ |$ denotes the determinant of a matrix M and $r = k_{1q}^{(1)}/k_{2q}^{(1)} \neq 1$. Notice that

$K_q^*(z)$ is a linear combination of $K_{1q}(z)$ and $K_{2q}(z)$ which is a generalized jackknife estimator by Gray and Schucany (1972).

Now, define the boundary kernel estimator, $\hat{m}_q^*(t;h)$, whose kernel function is $K_q^*(z)$ as

$$\hat{m}_q^*(t;h) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h}K_q^*\left(\frac{t-t_i}{h}\right)y_i . \tag{2.8}$$

Using the same technique as for $\hat{m}_{1q}(t;h)$ and $\hat{m}_{2q}(t;h)$, bias $[\hat{m}_q^*(t;h)]$ is expressed as

$$\text{bias}\,[\hat{m}_q^*(t;h)] = \frac{h^2 m^{(2)}(t)}{2}k_q^{2*} + o(h^2),$$

where $k_q^{2*} = \int_{-1}^{q} z^2 K_q^*(z)dz$. Hence, the dominant part of the asymptotic bias of the estimator, $\hat{m}_q^*(t;h)$, is of order h^2, the same order as in interior. Also, it is easy to obtain the asymptotic variance of $\hat{m}_q^*(t;h)$ as

$$\text{var}\,[\hat{m}_q^*(t;h)] = \frac{\sigma^2 Q^*}{nh} + o\left(\frac{1}{nh}\right)$$

where $Q^* = \int_{-1}^{q}[K_q^*(z)]^2 dz$. Hence, the asymptotic mean square error of $\hat{m}_q^*(t;h)$ is

$$mse[\hat{m}_q^*(t;h)] = \frac{\sigma^2 Q^*}{nh} + \left\{\frac{k_q^{2*}m^{(2)}(t)}{2}\right\}^2 h^4 + o\left(\frac{1}{nh}\right) + o(h^4) .$$

Therefore, it is clear, by comparing equation (2.3) for the non-boundary with $p=2$, that the asymptotic mean square error of $\hat{m}_q^*(t;h)$ has the same order of convergence rate as in the interior. Thus, the best rate of convergence to be anticipated for the $mse[\hat{m}_q^*(t;h)]$ is $n^{-4/5}$.

2.2.3. A Second Order Boundary Kernel

We demonstrate the $p = 2$ case with specific kernels in this section. It would be noticed that creating specific high-order boundary kernel may involve tedious evaluation of a determinant. For higher-order boundary kernels, more than two estimators would need to be combined. One should refer to the generalized jackknife method of higher-order by Gray and Schucany (1972).

First, let $K_1(z)$ be the second order uniform kernel, $K_1(z) = \frac{1}{2}$, $|z| \le 1$. Then, by "cutting", the boundary kernel function is,

$$K_{1q}(z) = \frac{K_1(z)}{\int_{-1}^{q} K_1(u)du} = \frac{1}{1+q}, \quad -1 \le z \le q.$$

Then, from (2.6), the expectation of the kernel estimator for $K_{1q}(z)$ is

$$E[\hat{m}_{1q}(t;h)] = m(t) - \frac{1}{2}(q-1)hm^{(1)}(t) + \frac{1}{6}(q^2+q+1)h^2 m^{(2)}(t) + o(h^2).$$

Next, let $K_2(z)$ be the Epanechnikov kernel, i.e., $K_2(z) = \frac{3}{4}(1-z^2)$, $|z| \le 1$. Then, by "cutting", the boundary kernel function, $K_{2q}(z)$, is

$$K_{2q}(z) = \frac{K_2(z)}{\int_{-1}^{q} K_2(u)du} = \frac{3}{(3q-q^3+2)}(1-z^2), \qquad -1 \le z \le q.$$

Then, from (2.7), the expectation of the kernel estimator for $K_{2q}(z)$ is

$$E[\hat{m}_{2q}(t;h)] = m(t) - \frac{3}{4}\frac{(q-1)^2}{(q-2)} hm^{(1)}(t) + \frac{1}{10}\frac{(3q^5-5q^3-2)}{(q-1)^2(q-2)} h^2 m^{(2)}(t) + o(h^2).$$

Hence, $E[\hat{m}_{1q}(t;h)]$ and $E[\hat{m}_{2q}(t;h)]$ can be written as

$$E[\hat{m}_{1q}(t;h)] = m(t) - \frac{1}{2}(q-1)hm^{(1)}(t) + O(h^2)$$

$$E[\hat{m}_{2q}(t;h)] = m(t) - \frac{3(q-1)^2}{4(q-2)} hm^{(1)}(t) + O(h^2)$$

It follows that the boundary kernel, $K_q^*(z)$, can be written as

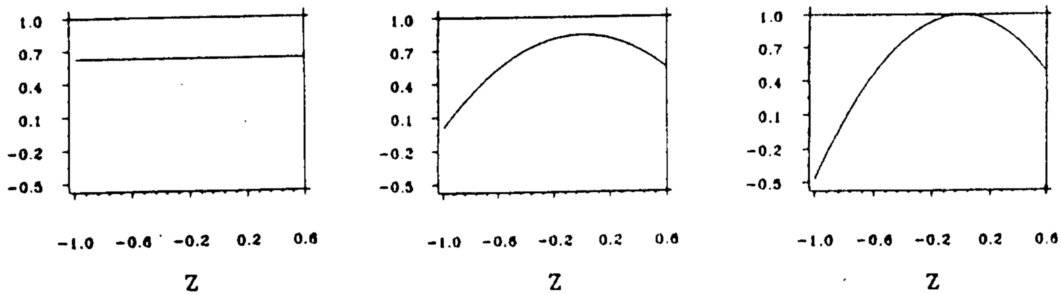
$$K_q^*(z) = \frac{\begin{vmatrix} \frac{1}{1+q} & \frac{3(1-z^2)}{(3q-q^3+2)} \\ -\frac{1}{2}(q-1) & -\frac{3(q-1)^2}{4(q-2)} \end{vmatrix}}{\begin{vmatrix} 1 & 1 \\ -\frac{1}{2}(q-1) & -\frac{3(q-1)^2}{4(q-2)} \end{vmatrix}} = 3\frac{(q^2+1)}{(q+1)^3}, \quad -1 \le z \le q.$$

Therefore, we can define a specific boundary kernel estimator as (2.8). Figure 2(a) and 2(b) show the "cut-and normalized" boundary kernel when q=0.6, also Figure 2(c) shows that boundary kernel that is the generalized jackknife combination of Figure 2(a) and 2(b). It should be noticed that $K_q^*(z)$ is the second order boundary kernel that converges to the Epanechnikov kernel as $q \to 1$. This result shows a good behavior of $K_q^*(z)$. That is, $K_q^*(z)$ converges to the optimal kernel of order 2, Epanechnikov kernel, as the boundary problem goes away.

3. Conclusions

It would be noticed that the proposed method has some merits. That is, this method easily can be extended to the higher-order boundary kernels with aids of computer, even if it is tedious to calculate determinants. Comparing to a minimum variance boundary kernel by Gasser and Müller (1979), the proposed method gives the optimal kernel estimator that minimizes both variance and bias of a kernel estimator. Hence, with the method proposed in this paper, we can obtain the same order of convergence rate as that of non-boundary for $p \ge 2$ whereas it may not be obtained from a minimum variance boundary kernel.

The purpose of this paper is to develop a boundary kernel estimator that has the same rate of convergence anticipating mean squre error as that of non-boundary. Therefore, it would be possible to estimate a whole curve by combining both non-boundary and boundary kernels without any assumption at the boundary.



(a) (b) (c)

Figure 2(a) Cut-and-Normalized Kernel Function from the Uniform Kernel.
2(b) Cut-and-Normalized Kernel Function from the Epanechnikov Kernel
2(c) Jackknife Boundary Kernel from Both Kernels

Appendix

Proof of equation (2.1):

$$
\begin{aligned}
E[\hat{m}(t;h)] &= \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h}K\left(\frac{t-t_i}{h}\right)E(y_i) \\
&= \frac{1}{n}\int_0^1 K\left(\frac{t-s}{h}\right)m(s)\,ds + O\left(\frac{1}{n}\right) \\
&= \int_{(t-1)/h}^{t/h} K(z)\Big\{ m(t) - zhm^{(1)}(t) + \frac{(zh)^2}{2!}m^{(2)}(t) + \cdots \\
&\quad + \frac{(-1)^p}{p!}(zh)^p m^{(p)}(t) + o(h^p)\Big\}dz + O\left(\frac{1}{n}\right).
\end{aligned}
$$

For h sufficiently small, so that $[-1,1] \subset [\frac{t-1}{h}, \frac{t}{h}]$, the above expansion reduces to

$$
\begin{aligned}
E[\hat{m}(t;h)] &= \int_{-1}^{+1} K(z)\Big\{ m(t) - zm^{(1)}(t) + \frac{(zh)^2}{2!}m^{(2)}(t) + \cdots \\
&\quad + \frac{(1)^p}{p!}(zh)^p m^{(P)}(t)\Big\}dz + o(h^p) + O\left(\frac{1}{n}\right)
\end{aligned}
$$

where $m^{(j)}(t)$ is the j^{th} derivative of $m(t)$ and $z = \frac{t-s}{h}$.

Thus, the asymptotic bias of $\hat{m}(t;h)$ is

$$E[\hat{m}(t;h)] - m(t) = \frac{(-1)^p}{p!}h^p m^{(p)}(t)k_p + o(h^p).$$

The above equation can be obtained by results of (1.2). Also, the error terms, $o(h^p) + O(\frac{1}{n}) = o(n^{-p/(2p+1)}) + O(\frac{1}{n})$ when the best rate of convergence, $h = O(n^{-1/(2p+1)})$

that can be obtained from (2.4), is used. However $p/(2p+1)-1<0$ which implies that $o(n^{-p/(2p+1)})$ is slower than $O(\frac{1}{n})$. Thus $o(n^{-p/(2p+1)})+O(\frac{1}{n}) = o(n^{-p/(2p+1)}) = o(h^p)$.

References

[1] Epanechnikov, V. A. (1969), "Nonparametric Estimation of a Multivariate Probability Density," *Theory of Probability and Its Application,* Vol. 14, 153-158.

[2] Eubank, R. L. (1988), *Spline Smoothing and Nonparametric Regression,* Marcel Dekker, New York.

[3] Eubank, R. L. and Speckman, P. L. (1989), "A Bias Reduction Theorem with Application in Nonparametric Regression," Technical Report #53, Dept. of Stat., Texas A&M, College Station, Tx.

[4] Gasser, Th. and Müller, H. G. (1979), *Kernel Estimation of Regression Functions: In Smoothing Techniques for Curve Estimation* (Gasser, Th. and Rosenblatt, M. editions), 23-68, Heidelberg: Springer-Verlag.

[5] Gray, H. L. and Schucany, W. R. (1972), *Generalized Jackknife Statistics,* Marcel Dekker, New York.

[6] Müller, H. G. (1988), *Nonparametric Regression Analysis of Longitudinal Data,* Springer-Verlag.

[7] Müller, H. G. and Stadtmüller, U.(1987), "Variable Bandwidth Kernel Estimators of Regression Curves," *The Annals of Statistics,* Vol. 15, 182-210.

[8] Parzen, E. (1962), "On Estimation of a Probability Density and Mode," *Annals of Mathematical Statistics,* Vol. 33, 1065-1076.

[9] Priestley, M. B. and Chao, M. T. (1972), "Nonparametric Function Fitting," *Journal of the Royal Statistical Society, B* **34**, 385-392.

[10] Rice, J.(1984), "Boundary Modification for Kernel Regression," *Communications in Statistics, A,* Vol. 13, 893-900.

비모수적 회귀선추정의 바운더리 편의 수정[1)]

차 경 준[2)]

요 약

커널을 이용한 회귀선의 추정은 단지 참 함수의 미분성만을 요구하는 비모수적인 회귀선의 추정방법이다. 유한구간에서 어떤 곡선의 완만한 추정곡선을 커널을 이용하여 추정할때 추정량의 전체적인 성능을 감소시키는 바운더리 문제가 발생하게 된다. 본논문에서는 바운더리 문제를 다룰수 있는 커널을 개발하였다. Gray와 Schcany(1972)의 일반화된 jackknife 방법을 이용하여 바운더리 커널을 개발하였고 또한 이 바운더리 커널이 비 바운더리 커널과 같은 수렴속도를 갖는것을 보였다.

1) 이 논문은 1992년도 세종대학교 대양학술연구비에 의하여 연구 되었음.
2) (133-791) 서울특별시 성동구 행당동 17, 한양대학교 수학과.