Minimization of mean absolute error of nonparametric regression function estimator

金 鐵 洙*

비모수 회귀함수 추정량에서의 평균절대오차의 최소화

Kim, Chul-soo *

Summary

The smoothing parameter or bandwidth for a kernel estimator of a regression function in fixed design model has been specified to minimize either asymptotic mean squre error or other measures. In this article we construct a simple methods, which permits asymptotic minimization of mean absolute error for nonparametric kernel regression function estimator of fixed design model.

적 요

비모수 회귀함수 추정량에서의 평균절대오차의 최소화

Kernel 비모수회귀함수 추정량에서 평균절대오차를 최소화하는 bandwidth의 선택방법과 평균절대오차를 최소화하는 해는 일의적이며 뉴톤방법에 의해 얻어짐을 보였다.

1. INTRODUCTION

In the last decade, nonparametric regression methods have gained considerable interest. Nadaraya (1964) and Watson (1964) introduced kernel estimators in the random design case, where the independent variable is random. From a practical point of view, the fixed design regression model, where the values of the independent variable are fixed in advance, seems to be of broader applicability. Therefore we will concentrate on the fixed design

* 自然科學大學 數學科

regression model.

For fixed designs the design variable is usually assumed to be restricted to some interval say $[0,1]$:

$$Y_i = m(x_i) + \varepsilon_i, \quad i=1,\cdots,n, \qquad (1\text{-}1)$$

where $0 \le x_1 \le x_2 \le \cdots \le \cdots \ x_n \le 1$ and (ε_i) are independently and identically distributed, $E(\varepsilon_i) = 0$ and $Var(\varepsilon_i) = \sigma^2$.

If there are several measurments made at one fixed point ϕ, Y_i can also be sample means or medians, or other location estimators based on the repeated measurement made at the same point. The error structure and the class to which the regression function m belongs have still to be specified for (1-1).

As a nonparametric estimator of the function m, Gasser and Müller (1979) introduced the following kernel estimator.

$$\overset{*}{m}(x) = \frac{1}{h}\sum_{j=1}^{n}\int_{s_{j-1}}^{s_j} K\left(\frac{x-u}{h}\right) du Y_j \qquad (1\text{-}2)$$

where $s_j = \frac{x_j + x_{j+1}}{2}$, $s_0 = 0$, $s_n = 1$. The value $h(n)$ is the bandwidth or smooting parameter, steering the degree of smoothness of the estimated curve $\overset{*}{m}$, variance and bias of m. The kernel K satisfies $\int K(x)dx = 1$ and further conditions to be given in the following section.

2. SOME PROPERTIES OF KERNEL ESTIMATOR

A proposal for estimating m is due to Priestly and Chao (1972) ;

$$m_n(x) = \sum_{j=1}^{n} \frac{x_j - x_{j-1}}{h} K\left(\frac{x - x_j}{h}\right) Y_j \qquad (2\text{-}1)$$

where h is a sequence of positive bandwidths depending on n such that $h \to 0$, $nh \to \infty$ as $n \to \infty$ and where K is a nonnegative kernel function satisfying :

$$\int K(x)dx = 1 \qquad \int K^2(x)dx < \infty.$$

A further kernel estimator proposed in Gasser and Müller (1979) is defined as in (1-2).

The definition of Priestly and Chao (1972) is very close to the definition of Gasser and Müller (1979), since it is a Riemann sum approximation to $\overset{*}{m}(x)$ in (1-2).

A minor advantage of the estimator of Gasser and Müller (1979) is that weights always add to 1. In the rest of the paper, we will concentrate on the kernel regression function estimator $\overset{*}{m}$ in (1-2). In what follows, the kernel K is assumed to be satisfied the following conditions :

A1. K has compact support $[-1,1]$ and $\int K(x)d(x) = 1$.

A2. $|K(u) - K(v)| \le |u-v|^{\gamma}$ for some $\gamma > 0$ and for $u, v \in [-1,1]$.

A hierarchy of kernels may now be defined.

Definition A kernel satisfying A1-A2 is called a kernel of order p if the following holds :

$$\int x^j K(x)dx = 0 \quad j = 1\cdots,\ p-1$$
$$\int x^p K(x)dx = B_p(K) \neq 0$$

Optimal kernels were previously derived in terms of Legendre polynomials (Gasser et al. 1985). Gasser and Müller (1979) derived the following theorems and corollary.

Theorem 2. 1 Let K be a kernel of

order p, and that the regression function $m(x)$ is s times differentiable with a continuous s th derivative on $[0,1]$ $(s \geq p)$. Assume $h \to 0$ and $nh \to \infty$ as $n \to \infty$. Then the bias and variance for all $x \in (0,1)$ can be expressed as follows :

$$\mathrm{Bias}(\overset{*}{m}(x)) = \frac{(-1)^p}{p!} h^p (m^{(p)}(x) B_p(K) + o(1)) + O(\frac{1}{n})$$

$$\mathrm{Var}(\overset{*}{m}(x)) = \frac{\sigma^2}{nh} (\int K^2(x)dx + o(1))$$

Theorem 2.2 If the assumptions of Theorem 2.1 are valid, and if :

$$\max|s_j - s_{j-1} - \frac{1}{n}| = o(\frac{1}{n}),$$

we have for all $x \in (0,1)$ for the mean square error :

$$\mathrm{MSE}(\overset{*}{m}(x)) = \frac{\sigma^2}{nh} \int K^2(x)dx + \frac{h^{2p}}{p!^2} B_p(K)^2 m^{(p)}(x)^2 + o(\frac{1}{n^2}) + o(h^{2p})$$

Corollary The asymptotically optimal bandwith h with respect to MSE is as follows :

$$\overset{*}{h} = \left(\frac{1}{p} \frac{p!^2 \sigma^2 \int K^2(x)dx}{B(K)^2 m^{(p)}(x)^2} \frac{1}{n}\right)^{\frac{1}{2p+1}}$$

where $m^{(p)}(x) \neq 0$.

The above result of the Theorem 2.1 are obtained by approximating sum byintegrals, using Taylor expansion.

By the bias and variance of $m(x)$ the MSE optimal bandwidth sequence is seen to be $h \sim n^{-\frac{1}{2p+1}}$, and this yields the rate convergence $\mathrm{MSE} \sim n^{-\frac{2p}{2p+1}}$. For function $m \in C([0,1])$, this rate is optimal. Consistency in MSE of the estimate $\overset{*}{m}(x)$ is established by the following theorem 2.3

Theroem 2.3. Let m be s times differentible and k bounded. Then $\overset{*}{m}(x)$ is a consistent estimate of

a) m is continuous at x.
b) $nh \to \infty$, $h \to \infty$ as $n \to \infty$.

To obtain the desired result in section 3, we need the some lemmas. In the following, Φ and ϕ denote cumulative distribution function and probabilty densty function of standard normal random variable, respectively.

Lemma 2.1 Let Z be a standard normal random variable. Then

a) $\int_y^\infty z\phi(z)dz = \phi(y)$ and $\int_{-\infty}^y z(\phi)dz = -\phi(y)$
b) $\phi'(z)z = -\phi(z)z^2$.

Proof. For a),

$$\int_y^\infty z\phi(z)dz = \int_y^\infty \frac{z}{\sqrt{2\pi}} \exp(-\frac{z^2}{2})dz = \frac{1}{\sqrt{2\pi}} \exp(-\frac{y^2}{2}) = \phi(y).$$

Similarly, $\int_{-\infty}^y z\phi(z)dz = -\phi(y)$

For b),

$$\phi'(z)z = -\frac{z^2}{\sqrt{2\pi}} \exp(-\frac{z^2}{2}) = -\phi(z)z^2.$$

Lemma 2.2 Let Z be a standard normal random variable and y be a real number. Then

$$E|Z-y| = 2\Phi(y)y - y + 2\phi(y)$$

and

$$E|Z-y|=E|Z+y|.$$

Proof. Note that $E|Z-y|=\int|z-y|\phi(z)dz$. Since $\int|z-y|\phi(z)dz=\int_{-\infty}^{y}(y-z)\phi(z)dz+\int_{y}^{\infty}(z-y)\phi(z)dz$, we obtain

$E|Z-y|=2\Phi(y)y-y+2\phi(y)$ from Lemma 2.1.

Also, $E|Z-y|=E|Z-(-y)|=2(1-\Phi(y))(-y)+y+2\phi(-y)$.

Hence the desired results are obtained.

3. MINIMIZATION OF MAE

Under certain regulatity conditions on $m(x)$ an exact expression for the asymptotically optimal value of h is readily derived (see e.g. Gasser and Müller (1979)). An alternative measure of loss is the mean absolute distance between m and $\overset{*}{m}$, which we shall call the mean absolute error (MAE).

Specifically,

$$MAE(\overset{*}{m}(x,h))=E|\overset{*}{m}(x,h)-m(x)| \qquad (3\text{-}1)$$

which is local analogue of the L_1 distance between m and m.

P.Hall and M.P.Wand(1988) constructed a simple algorithm, which permits asymptootic minimization of L_1 distance for nonparametric density estimators.

In this chapter we apply the results of P. Hall and M. P. Wand(1988) to find for asymptotically optimal bandwidth minimizing MAE in fixed design regression.

From well known Theorem 2.1, we obtain

$$\overset{*}{m}(x,h)-m(x)=\frac{(-1)^p}{P!}h^p m^{(p)}(x)B_p(K)+\frac{\sigma}{(nh)^{\frac{1}{2}}}(\int k^2(x)dx)^{\frac{1}{2}}Z \qquad (3\text{-}2)$$

as $n\to\infty$, $h=h(n)\to0$ and $nh\to\infty$, where $Z=Z(x)$ is a standard normal random variable.

To balance bias and standard deviation we must choose h so that each of these quantities are of the same order of magnitude. This involves taking $h=u^2n^{-\frac{1}{2p+1}}$ for some positive constant u not depending on n.

Let b_x and σ_k stand for $\frac{(-1)^p}{p!}m^{(p)}(x)B_p(K)$ and $\sigma V(K)^{\frac{1}{2}}$ where $V(K)=\int K^2(x)dx$, respectively.

Theorem 3.1 Let the conditions of Theorem 2.1 be satisfied and let $h=u^2n^{-\frac{1}{2p+1}}$ where u is a positive number.

Then the $MAE(\overset{*}{m}(x,h))$ is asymptotic to

$$n^{-\frac{p}{2p+1}}\delta_x(u)$$

where

$$\delta_{(x)}(u)=\int|u^{2p}b_x-u^{-1}\sigma_k z|\phi(z)dz \qquad (3\text{-}3)$$

and ϕ is the standard normal density function.

Proof. Note that $MAE(\overset{*}{m}(x,h))=E|\overset{*}{m}(x,h)-m(x)|$.

using (3-2), we obtain the expression

$$E|\overset{*}{m}(x,u^2n^{-\frac{1}{2p+1}})-m(x)|=n^{-\frac{1}{2p+1}}E|b_xu^{2p}-u^{-1}\sigma_kZ|$$

where Z is a standard normal random variable.

Then

$$MAE((\overset{*}{m}(x,h)) = E|\overset{*}{m}(x, u^2 n^{-\frac{1}{2p+1}}) - m(x)|$$
$$= n^{-\frac{1}{2p+1}} \int |b_x u^{2p} - u^{-1}\sigma_k z| \phi(z)\,dz.$$

Theorem 3.2 Under the conditions of Theorem 3.1, there exists only one u minimizing $\delta_x(u)$ in (3-3).

Proof. From Lemma 2.2, $\delta_x(u)$ is expressed as follows :

$$\delta_x(u) = \int |u^{2p} b_x - u^{-1}\sigma_k z| \phi(z)\,dz$$
$$= 2\sigma_k u^{-1} \Phi(u^{2p+1}\frac{bx}{\sigma_k})u^{2p+1}\frac{bx}{\sigma_k}$$
$$+ 2\sigma_k u^{-1} \phi(u^{2p+1}\frac{b_x}{\sigma_x}) - u^{2p} b_k$$

Using Lemma 2.1, we obtain

$$\frac{1}{2}\delta_x'(u) = u^{-2}[2pu^{2p+1}b_x\{(\Phi(u^{2p-1}\frac{b_x}{\sigma_k}) - \frac{1}{2}\}$$
$$- \sigma_k \phi(u^{2p+1}\frac{b_x}{\sigma_k})]$$
$$= u^{-2}\triangle_x(u^{2p+1})$$

where

$$\triangle_x(v) = 2pvb_x[\Phi(v\frac{b_x}{\sigma_k}) - \frac{1}{2}] - \sigma_k\phi(v\frac{b_x}{\sigma_k}).$$

Now, by b) of Lemma 2.6

$$\triangle_x'(v) = 2pvb_x[\Phi(v\frac{b_x}{\sigma_k}) - \frac{1}{2}] +$$
$$(2p+1)b_x^2 \sigma_k^{-1} v\phi(v\frac{b_x}{\sigma_k}),$$

which is positive for all $v>0$.

So, $\triangle_x(v)$ is an increasing function of v. Also $\lim_{v\to\infty} \triangle_x(v) = \infty$, while for $b_x \neq 0$, and

$$\lim_{v\to 0} \triangle_x(v) = -\sigma_k\phi(0) < 0.$$

This proves the fact that there exists only one $\overset{*}{v}$ such that $\triangle_x(v)=0$. Therefore there exists only one $\overset{*}{u}$ such that $\delta_x(u)=0$.

Theorem 3.3 Let the conditions of theorem 3.1 be satisfied.

Then the value of h which minimizes MAE at x is asymptotic to $(\overset{*}{v})^{\frac{2p}{2p+1}} n^{-\frac{1}{2p+1}}$ where

v is the root of $\triangle_x(v) = 2pvb_x[\Phi(v\frac{b_x}{\sigma_k}) - \frac{1}{2}) - \sigma_k\phi(v\frac{b_x}{\sigma_k})$.

Proof. From Theroem 3.1, we see that minimizaing $MAE(\overset{*}{m}(x,h))$ is equivalent to minimize $\delta_x(v)$ in (3-3).

From the fact that $\frac{1}{2}\delta_x'(u) = u^{-2}\triangle_x(u^{2p+1})$, the value of u which is the zero of $\triangle_x(u^{2p-1}) = 0$ is the value which minimizes $\delta_x(u)$.

Let the value of v for which $\triangle_x(v)=0$ be $\overset{*}{v}$. Then the value of u'for which minimizes $\delta_x(u)$. is $\overset{*}{u} = (\overset{*}{v})^{\frac{1}{2p+1}}$

Therefore the value of $h = u^2 n^{-\frac{1}{2p+1}}$ which minimizes MAE at x is asymptotic to $(\overset{*}{v})^{\frac{2}{2p+1}} n^{-\frac{1}{2p+1}}$

In practice the equation $\triangle_x(v)=0$ may be solved using Newton's method, as follows.

Let $H(v) = \frac{\triangle_x(v)}{\triangle_x'(v)}$.

Then

$$H(v) = [2pvb_x\{\Phi(vb) - \frac{1}{2}\} - \sigma_k\phi(vb)] \times$$
$$[2pvb_x\{\Phi(vb) - \frac{1}{2}\} - 2(p+1)b^2\sigma^{-1}v\phi(vb)]^{-1}$$

If v_1 is an approximation to the solution of $\triangle_x(v)=0$, then $v_2 = v_1 - H(v_1)$.

Continuing this process, we form the sequence $v_1, v_2, \cdots$, where $v_{i+1} = v_i - H(v)$ such that $\lim_{n\to\infty} v_n = \overset{*}{v}$ for $\triangle_x(\overset{*}{v}) = 0$.

REFERENCES

Gasser, Th. and Müller, H.G., 1979, Kernel estimation of regression functions. Lecture Note in Math. 757, 23-68.

Gasser, Th., Müller, H.G. and Mammitzsch, V., 1985, Kernels for nonparametric curve estimation. J. Roy. Stat. Soc. B 47, 238-252.

Müller, H.G., 1987, On the asymptotic mean square error of L_1 kernel estimates of C^∞ functions. J. Approx. Theory 51, 193-201.

Hall, P. and Wand, M.P., 1988, Minimizing L_1 distance in nonparametric density estimation. J. Mult. Anal. 26, 59-88.

Priestly, M. B. and Chao, M. T., 1972, Nonparametric function fitting. J. Roy. Stat. Soc. B 34, 384-392.