碩士學位 請求論文

# MINIMIZATION OF MEAN ABSOLUTE ERROR OF REGRESSION FUNCTION ESTIMATOR

指導教授 李 世 烈



濟州大學校 教育大學院

數學教育專攻

夫 公 男

1990年度

# MINIMIZATION OF MEAN ABSOLUTE ERROR OF REGRESSION FUNCTION ESTIMATOR

이 論文을 教育學 碩士學位 論文으로 提出함



濟州大學校 教育大學院 數學教育專攻

提出者 夫 公 男

指導教授 李 世 烈

1990年　　月　　日

夫公男의 碩士學位 論文을 認准함

1990年　　月　　日



主審 ________________________ 인

副審 ________________________ 인

副審 ________________________ 인

濟州大學校 敎育大學院

## CONTENTS

〈國文抄錄〉

# 回歸函數推定量에서의 平均絶對誤差의 最小化


夫 公 男

濟州大學校 教育大學院 數學教育專攻

指導教授 李 世 烈



$Y_i = m(x_i)+\varepsilon_i$, $i=1,2,...,n$이고 $\varepsilon_i$의 平均은 0, 分散은 $\sigma^2$으로 주어지는 回歸模型에서 回歸函數 m(x)의 母數的(parametric) 形態를 알 수 없는 경우에는 非母數的(nonparametric)인 方法으로 m(x)을 推定하게 된다. 本 論文에서는 回歸函數 m(x)에 대하여 가서와 뮐러가 1979년에 紹介한 kernel 推定量을 使用하여 m(x)를 推定할 경우 推定量의 平滑度(smoothness)에 아주 敏感하게 反應하는 bandwidth의 選擇問題를 다뤘다. 平均제곱誤差(mean square error:MSE)가 아닌 平均絶對誤差(mean absolute error:MAE)의 觀点에서 MAE를 最小化하는 bandwidth는 唯一하게 存在함을 보였다.

## 1. INTRODUCTION

Let us assume a situation where a random variable $X$ is recorded which depends on a design parameter, e.g. time or age. A parametric approach is very often based on a clever guess, and not on any a priori knowledge in the field of application. In contrast to phisics or engineering, a parametric data analysis is often not appropriate in biomedicine for curve data. One reason is that modeling is more difficult for living organisms ; a second one is that whenever a sample of curves has to be analysed, it may be difficult to find one well fitting parametric family of functions for descriptive purposes.

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 regression model.



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

$$Y_i = m(x_i) + \epsilon_i , \quad i=1,...,n, \qquad (1-1)$$

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

If there are several measurements made at one fixed point $x$, $Y_i$ 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 $(1-1)$.

If the regression function $m$ belongs to a class of functions that are determined by a finite number of parameters,i.e. if $m$ belongs to a parametric family,e.g. the linear functions, the regression model $(1-1)$ is called parametric, if $m$ belongs to a smoothness class ,e.g. $m \in C^k$ for some $k \geq 0$,i.e. the class of $k$ times continuously differentiable functions,it is called nonparametric.

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



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

where $s_j = \frac{x_j + x_{j+1}}{2}$, $s_0 = 0, s_n = 1$. The value $h = h(n)$ is the bandwidth or smoothing parameter, steering the degree of smoothness of the estimated curve $\dot{m}$, variance and bias of $\dot{m}$. 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);

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

where $h$ is a sequence of positive bandwidths depending on $n$ such that $h \to 0, \quad nh \to \infty$ 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 $\dot{m}(x)$ 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 $\dot{m}(x)$ 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 $\int K(x)\,dx = 1.$

$A2.$ $|K(u) - K(v)| \leq |u - v|^{\gamma}$ 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:

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

$$\int x^p K(x)\,dx = B_p(K) \neq 0$$



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:

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

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

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

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

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

$$MSE\left(\overset{\bullet}{m}(x)\right) = \frac{\sigma^2}{nh} \int K^2(x)\, dx + \frac{h^{2p}}{p!^2} B_p(K)^2 m^{(p)}(x)^2 + O(\frac{1}{n^2}) + o(h^{2P})$$



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

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

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

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

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

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

a) $m$ is continuous at $x$.

b) $nh \to \infty, h \to \infty$ as $n \to \infty$.

We quote the following two results by Gasser and Müller (1984)

**Theorem 2.4** Let us assume the conditions of Theorem2.1 and in addition :



i) $E|\epsilon_i|^p < \infty$ for some $p \geq 2, n^{\frac{1}{p}} h \to \infty$

ii) $\sum_{n=1}^{\infty} \exp[-(n^{1-\frac{1}{p}} h)^{\frac{1}{2}}] < \infty$

Then $\overset{\bullet}{m}(x) \to m(x)$ as $n \to \infty$.

**Theorem 2.5** The estimate $(\overset{\bullet}{m}(x_1), \cdots \overset{\bullet}{m}(x_n))$ $(x_i \in (0,1), i = 1, \cdots n)$ are asymptotically normally distributed given the following conditions hold:

i) $k$ is Lipschitz continuous of order $\gamma\,(0 < \gamma \leq 1)$

ii) $\max|s_j - s_{j-1} - \frac{1}{n}| = O(\frac{1}{n^\delta}), \delta > 1$

iii) $h \to 0, nh \to \infty$ as $n \to \infty$.

To obtain the desired result in chapter 3, we need the some lemmas. In the following, $\Phi$ and $\phi$ denote cumulative distribution function and probabilty densty function of standard normal random variable , respectively.

**Lemma 2.6** Let $Z$ be a standard normal random variable. Then

a) $\displaystyle\int_y^\infty z\phi(z)\,dz = \phi(y)$ and $\displaystyle\int_{-\infty}^y z\phi(z)\,dz = -\phi(y)$

b) $\phi'(z)z = -\phi(z)z^2$.



Proof. For a),

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

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

For b),

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

**Lemma 2.7** 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

$$E|Z-y| = 2\Phi(y)y - y + 2\phi(y) \quad \text{from Lemma 2.6.}$$

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

For pratice applications of curve smoothing methods, the choice of a good bandwidth is a very important issue. For kernel and weighted local least squares estimators this is the choice of the bandwidth, which besides the choice of the correct order of the kernel or polynomial has a strong infulence on the quality of the estimate. The badwidth,losely speaking,provides information about the signal-to-noise ratio in the data. In many finite sample situations it is very difficult to make the right decision. Therefore a completely satisfying finite sample solution of the bandwidth choice are problem is not possible. The methods proposed for bandwidth choice are motivated by asymptotic considerations.



Under certain regularity 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 $\overset{\bullet}{m}$, which we shall call the mean absolute error (MAE).

Specifically,

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

which is local analogue of the $L_1$ distance between $m$ and $\overset{\bullet}{m}$.

P.Hall and M.P.Wand (1988) constructed a simple algorithm, which permits asymptotic 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

$$\overset{\bullet}{m}(x,h) - m(x) = \frac{(-1)^p}{p!} h^p m^{(p)}(x) B_p(K) + \frac{\sigma}{(nh)^{\frac{1}{2}}} \left( \int K^2(x)\, dx \right)^{\frac{1}{2}} Z + O(h^p) \tag{3-2}$$

as $n \to \infty$, $h = h(n) \to 0$ 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 $h = u^2 n^{-\frac{1}{2p+1}}$ for some positive constant $u$ not depending on $n$.

Let $b_x$ and $\sigma_k$ 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 $h = u^2 n^{-\frac{1}{2p+1}}$ 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

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

and $\phi$ 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

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

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 2.7, $\delta_x(u)$ is expressed as follows:

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

$$+ 2\sigma_k u^{-1}\phi(u^{2p+1}\frac{b_x}{\sigma_k}) - u^{2p} b_x.$$

Using Lemma 2.6,we obtain

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

$$= u^{-2}\Delta_x(u^{2p+1})$$

where

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



Now, by b) of Lemma 2.6

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

which is positive for all $v > 0$.

So,$\Delta_x(v)$ is an increasing function of $v$.

Also $\lim_{v\to\infty}\Delta_x(v) = \infty$, while for $b_x \neq 0$, and

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

This proves the fact that there exists only one $\overset{\bullet}{v}$ such that $\Delta_x(v) = 0$.

Therefore there exists only one $\overset{\bullet}{u}$ such that $\delta_x(\overset{\bullet}{u}) = 0$.

**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

$(\overset{\bullet}{v})^{\frac{2}{2p+1}} n^{-\frac{1}{2p+1}}$ where $\overset{\bullet}{v}$ is the root of $\Delta_x(v) = 2pvb_x[\Phi(v\frac{b_x}{\sigma_k}) - \frac{1}{2}] - \sigma_k \phi(v\frac{b_x}{\sigma_k})$.

Proof. From Theorem 3.1, we see that minimizing $MAE(\overset{\bullet}{m}(x,h))$ is equivalent to minimize $\delta_x(u)$ in (3-3).

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

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

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

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

Let $H(v) = \dfrac{\Delta_x(v)}{\Delta_x'(v)}$.

Then

$$H(v) = \left[2pvb_x\{\Phi(vb) - \frac{1}{2}\} - \sigma_k\phi(vb)\right] \times$$

$$\left[2pvb_x\{\Phi(vb) - \frac{1}{2}\} + 2(p+1)b^2\sigma^{-1}v\phi(vb)\right]^{-1}.$$

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

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

## REFERENCES


1. Gasser,Th. and Müller, H.G. Kernel estimation of regression functions. Lecture Note in Math. 757,23-68 (1979)
2. Gasser, Th., Müller,H.G. and Mammitzsch,V. Kernels for nonparametric curve estimation. J.Roy. Stat. Soc. B 47,238-252(1985)
3. Müller, H.G. On the asymptotic mean square error of $L_1$ kernel estimates of $C^\infty$ functions. J. Approx. Theory 51, 193-201 (1987)
4. Hall,P. and Wand,M.P. Minimizing $L_1$ distance in nonparametric density estimation. J. Mult. Anal. 26,59-88(1988)
5. Priestly, M.B. and Chao, M.T. Nonparametric function fitting. J.Roy .Stat. Soc. B 34, 384-392(1972)

## 感謝의 글

어려운 逆境 속에서도 教育課程을 履修하게되어 한없이 기쁩니다.

그동안 講座를 擔當하시어 애써 주신 教授님들과 指導教授이신 李世烈 教授님 여러모로 도움을 주신 先輩, 同僚, 後輩 院生들에게 이 紙面을 통하여 고마운 말씀을 전하며, 특히 이 論文을 完成하는데 細心한 指導와 配慮를 해주신 金鐘洙 博士님께 하나님의 恩惠가 充滿하시길 빕니다.

끝으로 어려울 적 마다 激勵를 아끼지 않으신 어머님과 아내, 사랑하는 煥이와 運이에게 이 榮光을 돌립니다.

1990 년. 초여름