ESTIMATION OF A REGRESSION FUNCTION BY THE PARZEN KERNEL-TYPE DENSITY ESTIMATORS

KAZUO NODA

(Received Nov. 1, 1975; revised June 28, 1976)

Summary

In this paper a theory of estimation of a regression function by the Parzen kernel-type density estimators is developed in the following points: 1) convergence of the estimators to the regression function at a continuous point, 2) convergence of the mean square error at a continuous point, and 3) the speed of the convergence in 2).

1. Introduction

Let $(X_1, Y_1), \cdots, (X_n, Y_n)$ be independent and identically distributed random variables with a probability density function f on the two dimensional real space $(\mathcal{R}^2, \mathcal{B}^2)$. Denote the marginal density function of the random variable X by g and the regression of Y with respect to X by η. Of course we assume the existence of $\eta(x)$. Let μ be the Lebesgue measure on $(\mathcal{R}, \mathcal{B})$.

To estimate the regression $\eta(x)$, Nadaraya [4] introduced the class of statistics

$$(1.1) \qquad \tilde{\eta}_n(x)=\sum_{i=1}^{n} Y_i K\left(\frac{x-X_i}{h(n)}\right)\Big/\sum_{i=1}^{n} K\left(\frac{x-X_i}{h(n)}\right),$$

where K is a kernel which is a $\mathcal{B}$-measurable function satisfying the conditions,

1° $\sup_{-\infty<x<\infty} |K(x)|<\infty$,

2° $\lim_{x\to\pm\infty} |xK(x)|=0$,

3° $\int_{-\infty}^{\infty} |K(x)|\mu(dx)<\infty$,

4° $\int_{-\infty}^{\infty} K(x)\mu(dx)=1$,

and h is a function of natural numbers n satisfying the conditions,

1′ $h(n)\to 0$,

2′ $nh(n)\to\infty$

with increasing n.

In addition, these K and h were given by Parzen [1] for the first time to estimate the density function g.

Further, Nadaraya showed that

$$\sup_{a\leqq x\leqq b} |\tilde{\eta}_n(x)-\eta(x)|\to 0 \qquad (n\to\infty)$$

with probability one, under the conditions; a) g and η are continuous over the entire real line, b) $\min_{-\infty<a\leqq x\leqq b<\infty} g(x)=\alpha>0$, c) $-\infty<A\leqq Y\leqq B<\infty$ with probability one, d) K is a function of bounded variation, and e) the series $\sum_{n=1}^{\infty}\exp\{-\nu nh^2(n)\}$ is convergent for any $\nu>0$.

In this paper we show the following results as the main theorems under the conditions given by Parzen without those a)–e) added by Nadaraya;

$$\text{(1.2)} \qquad P\{\lim_{n\to\infty} |\tilde{\eta}_n(x)-\eta(x)|=0\}=1 ,$$

$$\text{(1.3)} \qquad \lim_{n\to\infty} E[\tilde{\eta}_n(x)-\eta(x)]^2=0 ,$$

if $x\in\mathcal{R}$ with $g(x)\neq 0$ is a continuous point of η, g, and f for each fixed $y\in\mathcal{R}$. Further we will show the speed of the convergence of (1.3) under the above-mentioned conditions and those added by Wahba [7].

When the marginal density function g is known, we can consider the estimator

$$\text{(1.4)} \qquad \hat{\eta}_n(x)=\frac{1}{nh(n)g(x)}\sum_{i=1}^{n} Y_i K\left(\frac{x-X_i}{h(n)}\right)$$

instead of $\tilde{\eta}_n(x)$. For this estimator we will show the same result as stated above.

It should be noted that our results will easily be extended to the case where the variable X is multivariate, using Cacoullos' extension [2] of Parzen's results.

2. Convergence of the estimator $\tilde{\eta}_n(x)$ to $\eta(x)$ with probability one

Let

$$\text{(2.1)} \qquad k(x)=g(x)\eta(x) ,$$

$$\text{(2.2)} \qquad \tilde{g}_n(x)=\frac{1}{nh(n)}\sum_{i=1}^{n} K\left(\frac{x-X_i}{h(n)}\right) ,$$

$$\text{(2.3)} \qquad \tilde{k}_n(x)=\frac{1}{nh(n)}\sum_{i=1}^{n} Y_i K\left(\frac{x-X_i}{h(n)}\right)$$

for $x \in \mathcal{R}$ and for the sample $(X_1, Y_1), \cdots, (X_n, Y_n)$. Then we have

$$\tilde{\eta}_n(x) = \tilde{k}_n(x)/\tilde{g}_n(x) . \tag{2.4}$$

As the condition of h, we also impose the following one which implies the condition $2'$;

$2''$ There are a constant β and a monotone increasing function H of natural numbers m such that

$$H(0)=0 , \quad H(m)\uparrow \infty \quad (m \to \infty) , \qquad \text{and} \qquad Dm^{1+\alpha} \leqq H(m) \tag{2.5}$$

for sufficiently large n where D is some positive constant and $0<\alpha<1$, and

$$\sum_{n=m}^{\infty} \frac{1}{n^2 h(n)} \leqq \frac{\beta}{H(m)} . \tag{2.6}$$

We assume that $h(n)$ satisfies the conditions $1'$ and $2''$ so that it satisfies $2'$ as well.

Remark 2.1. We think that the following properties are well-known.

If $h(n)=O(n^{-\lambda})$ or $h(n)=O(\exp(-n^{\lambda}))$ for $0<\lambda<1$, the conditions $2'$ and $2''$ are equivalent. Therefore the condition $2'$ can be replaced by $2''$ in those cases.

Existence of $\eta(x)$ is equivalent to the following assumption:

ASSUMPTION (A)

$$\left| \int_{-\infty}^{\infty} y f(x, y) \mu(dy) \right| < \infty$$

for each fixed $x \in \mathcal{R}$.

We omit to mention Assumption (A) in each proposition throughout this paper.

First we obtain the following lemma by a closer examination of the proof of the Kolmogorov strong law of large numbers (see Theorem 7.2.2 in [6], for instance).

LEMMA 2.1. *Let $(X_1, Y_1), (X_2, Y_2), \cdots$ be independent random variables satisfying* $\mathrm{E}(X_n Y_n) < \infty$ *and* $\mathrm{E}[X_n Y_n - \mathrm{E}(X_n Y_n)]^2 < \infty$ *for each n, and let $\{b_n\}$ be an increasing sequence of positive real numbers with $b_n \to \infty$. If*

$$\sum_{n=1}^{\infty} \frac{\mathrm{E}[X_n Y_n - \mathrm{E}(X_n Y_n)]^2}{b_n^2} < \infty ,$$

then (with $S_n = X_1 Y_1 + \cdots + X_n Y_n$)

$$\frac{S_n - \mathrm{E}\, S_n}{b_n} \to 0 \qquad (n \to \infty)$$

with probability 1.

LEMMA 2.2. *Suppose that the kernel K satisfies the conditions* 1°–4°. *If $x \in \mathcal{R}$ is a continuous point of g, then*

$$|\tilde{g}_n(x) - \mathrm{E}\, \tilde{g}_n(x)| \to 0 \qquad (n \to \infty)$$

holds with probability 1.

PROOF. It suffices to show that

$$V \equiv \sum_{n=1}^{\infty} \frac{1}{[nh(n)]^2} \mathrm{Var}\left[K\left(\frac{x - X_n}{h(n)}\right)\right] < \infty \tag{2.7}$$

holds by the Kolmogorov strong law of large numbers. In fact, we have

$$\begin{aligned} V &\leqq \sum_{n=1}^{\infty} \frac{1}{[nh(n)]^2} \mathrm{E}\left[K\left(\frac{x - X_n}{h(n)}\right)\right]^2 \\ &= \sum_{n=1}^{\infty} \frac{1}{n^2 h(n)} \frac{1}{h(n)} \int_{-\infty}^{\infty} K^2\left(\frac{x - u}{h(n)}\right) g(u) \mu(du) . \end{aligned} \tag{2.8}$$

However, by Parzen's result (Theorem 1A in [1]), for any $\varepsilon > 0$ there exists a natural number $n_0(\varepsilon)$ such that if $n \geqq n_0(\varepsilon)$, then

$$\frac{1}{h(n)} \int_{-\infty}^{\infty} K^2\left(\frac{x - u}{h(n)}\right) g(u) \mu(du) \leqq g(x) \int_{-\infty}^{\infty} K^2(u) \mu(du) + \varepsilon \tag{2.9}$$

holds*. Consequently, by (2.8) and (2.9) we have

$$\begin{aligned} V \leqq & \sum_{n=1}^{n_0(\varepsilon)-1} \frac{1}{[nh(n)]^2} \int_{-\infty}^{\infty} K^2\left(\frac{x - u}{h(n)}\right) g(u) \mu(du) \\ & + \left[g(x) \int_{-\infty}^{\infty} K^2(u) \mu(du) + \varepsilon\right] \sum_{n=n_0(\varepsilon)}^{\infty} \frac{1}{n^2 h(n)} . \end{aligned} \tag{2.10}$$

The result of this lemma follows from this (2.10) and the condition 2″ of the function h.

LEMMA 2.3. *Suppose that the kernel K satisfies the conditions* 1°–4°. *If $x \in \mathcal{R}$ is a continuous point of f for each fixed $y \in \mathcal{R}$, then*

$$|\tilde{k}_n(x) - \mathrm{E}\, \tilde{k}_n(x)| \to 0 \qquad (n \to \infty)$$

holds with probability 1.

PROOF. Since the Y_n have the same distribution, by (2.5) there exists a natural number n_0 such that if $n \geqq n_0$, then

* For any $r \geqq 1$ $\int_{-\infty}^{\infty} |K^r(u)| \mu(du) < \infty$, since K has the conditions 1° and 3°.

$$\begin{aligned}\sum_{r=1}^{\infty} \mathrm{P}\{|Y_r|\geqq H(r)\} &= \sum_{r=1}^{\infty}\sum_{m=r}^{\infty} \mathrm{P}\{H(m)\leqq |Y_1|<H(m+1)\}\\ &= \sum_{m=1}^{\infty} m\,\mathrm{P}\{H(m)\leqq |Y_1|<H(m+1)\}\\ &\leqq \sum_{m=1}^{n_0-1} m\,\mathrm{P}\{H(m)\leqq |Y_1|<H(m+1)\}\\ &\quad +\frac{1}{D}\sum_{m=n_0}^{\infty} H(m)\,\mathrm{P}\{H(m)\leqq |Y_1|<H(m+1)\}\\ &\leqq \sum_{m=1}^{n_0-1} m\,\mathrm{P}\{H(m)\leqq |Y_1|<H(m+1)\}\\ &\quad +\frac{1}{C_1}\mathrm{E}\{|Y_1|\}\ .\end{aligned}$$

Since $\mathrm{E}\{|Y_1|\}<\infty$ by Assumption (A), the above inequality gives

$$\sum_{n=1}^{\infty} \mathrm{P}\{|Y_n|\geqq H(n)\}<\infty\ .$$

Hence by the Borel-Cantelli lemma we have

$$\mathrm{P}\{|Y_n|\geqq H(n) \text{ for infinitely many } n\}=0\ .$$

Let

$$Y_n' \Leftarrow \begin{cases} Y_n & \text{if } Y_n<H(n)\,,\\ 0 & \text{if } Y_n\geqq H(n)\,. \end{cases} \tag{2.11}$$

Then by the above statement $Y_n'=Y_n$ for sufficiently large n with probability 1. By Lemma 2.1, it suffices for the proof to show that

$$V_0 \Leftarrow \sum_{n=1}^{\infty} \frac{1}{[nh(n)]^2}\,\mathrm{E}\left[Y_n' K\left(\frac{x-X_n}{h(n)}\right)\right]^2<\infty \tag{2.12}$$

holds. Let $\bar{f}(x|y)$ be the conditional probability density function of X given Y. Then we have

$$\begin{aligned} V_0 &= \sum_{n=1}^{\infty} \frac{1}{[nh(n)]^2}\,\mathrm{E}\left[Y_n'^2\,\mathrm{E}\left\{K^2\left(\frac{x-X_n}{h(n)}\right)\Big|\,Y_n\right\}\right]\\ &= \sum_{n=1}^{\infty} \frac{1}{[nh(n)]^2}\,\mathrm{E}\left[Y_n'^2 \int_{-\infty}^{\infty} K^2\left(\frac{x-u}{h(n)}\right)\bar{f}(u\,|\,Y_n)\mu(du)\right]\ . \end{aligned} \tag{2.13}$$

However, since x is a continuous point of $\bar{f}$ for any fixed $y\in\mathcal{R}$, for any $\varepsilon>0$ there exists a natural number $n_0(\varepsilon)$ such that if $n\geqq n_0(\varepsilon)$, then

$$\frac{1}{h(n)}\int_{-\infty}^{\infty} K^2\left(\frac{x-u}{h(n)}\right)\bar{f}(u\,|\,y_n)\mu(du)\leqq \bar{f}(x\,|\,y_n)\int_{-\infty}^{\infty} K^2(u)\mu(du)+\varepsilon \tag{2.14}$$

holds for a.e. y_n with respect to μ by Parzen's result [1].

Consequently, it follows from (2.13) and (2.14) that

$$V_0 \leqq \sum_{n=1}^{n_0(\varepsilon)-1} \frac{1}{[nh(n)]^2} \mathrm{E}\left[Y_n'^2 K^2\left(\frac{x-X_n}{h(n)}\right)\right] + \sum_{n=n_0(\varepsilon)}^{\infty} \frac{1}{n^2h(n)} \mathrm{E}\left[Y_n'^2 \left\{ \bar{f}(x\,|\,Y_n) \int_{-\infty}^{\infty} K^2(u)\mu(du) + \varepsilon \right\}\right] . \tag{2.15}$$

Then it suffices for the proof of this lemma to show that the second term of the right-hand side of (2.15) is bounded. To do so we shall show

$$\begin{aligned} V_1 &\overset{\leftarrow}{=} \int_{-\infty}^{\infty} K^2(u)\mu(du) \sum_{n=n_0(\varepsilon)}^{\infty} \frac{1}{n^2h(n)} \mathrm{E}\,[Y_n'^2 \bar{f}(x\,|\,Y_n)] \\ &= \int_{-\infty}^{\infty} K^2(u)\mu(du) \sum_{n=n_0(\varepsilon)}^{\infty} \frac{1}{n^2h(n)} \int_{-\infty}^{\infty} y_n'^2 f(x, y_n')\mu(dy_n') < \infty , \end{aligned} \tag{2.16}$$

and

$$V_2 \overset{\leftarrow}{=} \varepsilon \sum_{n=n_0(\varepsilon)}^{\infty} \frac{1}{n^2h(n)} \mathrm{E}\, Y_n'^2 < \infty . \tag{2.17}$$

Let

$$\mathrm{E}_x'\, l(Y) \overset{\leftarrow}{=} \int_{-\infty}^{\infty} l(y) f(x, y)\mu(dy) . \tag{2.18}$$

Then it suffices for the proof of (2.16) to show that

$$V_x' \overset{\leftarrow}{=} \sum_{n=1}^{\infty} \frac{1}{n^2h(n)} \mathrm{E}_x'\, Y_n'^2 < \infty \tag{2.19}$$

for the fixed value x. In fact we have

$$\begin{aligned} V_x' &= \sum_{n=1}^{\infty} \frac{1}{n^2h(n)} \mathrm{E}_x'\,[Y_1^2 \cdot I_{\{|Y_1|<H(n)\}}] \\ &= \sum_{n=1}^{\infty} \frac{1}{n^2h(n)} \sum_{m=1}^{n} \mathrm{E}_x'\,[Y_1^2 \cdot I_{\{H(m-1)\leq |Y_1|<H(m)\}}] \\ &= \sum_{m=1}^{\infty} \mathrm{E}_x'\,[Y_1^2 \cdot I_{\{H(m-1)\leq |Y_1|<H(m)\}}] \sum_{n=m}^{\infty} \frac{1}{n^2h(n)} \\ &\leqq \sum_{m=1}^{\infty} \frac{\beta}{H(m)} \mathrm{E}_x'\,[Y_1^2 \cdot I_{\{H(m-1)\leq |Y_1|<H(m)\}}] , \end{aligned} \tag{2.20}$$

where I_B is the indicator function of a set $B \in \mathcal{B}$. Here, if $H(m-1) \leqq |Y_1| < H(m)$, then $Y_1^2 \leqq H(m)\cdot|Y_1|$. Thus by Assumption (A) we have

$$V_x' \leqq \beta \sum_{m=1}^{\infty} \mathrm{E}_x'\,[|Y_1|\cdot I_{\{H(m-1)\leq |Y_1|<H(m)\}}] = \beta\, \mathrm{E}_x'\, |Y_1| < \infty .$$

Further (2.17) can be shown by the same method as stated above.

We have the following result under the above preparations.

THEOREM 2.1. *Suppose that the kernel K satisfies the conditions* 1°–4°. *If $x \in \mathcal{R}$ with $g(x) \neq 0$ is a continuous point of g, η, and f for each fixed $y \in \mathcal{R}$, then*

$$P\{\lim_{n\to\infty} |\tilde{\eta}_n(x)-\eta(x)|=0\}=1 .$$

PROOF. It suffices for the proof of this theorem to show that

$$\tilde{g}_n(x) \to g(x) \qquad (n\to\infty) , \tag{2.21}$$

$$\tilde{k}_n(x) \to k(x) \qquad (n\to\infty) \tag{2.22}$$

with probability 1.

However by Parzen's result [1] we have

$$E\,\tilde{g}_n(x) \to g(x) , \quad E\,\tilde{k}_n(x) \to k(x) \qquad (n\to\infty)$$

under the above-mentioned conditions. Hence (2.21) and (2.22) follow from Lemma 2.2 and Lemma 2.3, respectively.

Remark 2.2. Continuity of f, g, and η as the condition in each theorem in this paper can be replaced by the following conditions of f.

1) f is continuous in the variable x at x_0 for almost all $y \in \mathcal{R}$.
2) There is a neighbourhood $V(x_0)$ of x_0 such that

$$|f(x, y)| < G(y)$$

holds on $V(x_0) \times \mathcal{R}$ for a positive and Lebesque integrable function G of y.

Under these conditions of f the other functions g, η, and $\bar{f}$ are continuous at x_0 (see Theorem 130 in [3]).

3. Convergence of the mean square error $E[\tilde{\eta}_n(x)-\eta(x)]^2$

THEOREM 3.1. *Suppose that the kernel K satisfies the conditions* 1°–4°. *Suppose that the conditional variance of Y given X has a finite value for each $x \in \mathcal{R}$. If $x \in \mathcal{R}$ with $g(x) \neq 0$ is a continuous point of g, η, and f for each fixed $y \in \mathcal{R}$, then*

$$\lim_{n\to\infty} E[\tilde{\eta}_n(x)-\eta(x)]^2=0 .$$

PROOF. We have

$$E[\tilde{\eta}_n(x)-\eta(x)]^2 = E\left[\frac{\tilde{k}_n(x)}{\tilde{g}_n(x)} - \frac{E\,\tilde{k}_n(x)}{E\,\tilde{g}_n(x)}\right]^2 + 2\left[\frac{E\,\tilde{k}_n(x)}{E\,\tilde{g}_n(x)} - \frac{k(x)}{g(x)}\right] E\left[\frac{\tilde{k}_n(x)}{\tilde{g}_n(x)} - \frac{E\,\tilde{k}_n(x)}{E\,\tilde{g}_n(x)}\right] \tag{3.1}$$

$$+\left[\frac{\mathrm{E}\,\tilde{k}_n(x)}{\mathrm{E}\,\tilde{g}_n(x)}-\frac{k(x)}{g(x)}\right]^2 .$$

We can easily see the convergence of each term of the right-hand side of (3.1) by Theorem 2.1 and Parzen's result, since we have the fact that if we take a sufficiently small positive number ε, then there exists $n_0(\varepsilon)$ such that for $n \geqq n_0(\varepsilon)$

$$\frac{1}{g^2(x)+\varepsilon}<\frac{1}{\tilde{g}_n(x)\cdot \mathrm{E}\,\tilde{g}_n(x)}<\frac{1}{g^2(x)-\varepsilon}$$

and

$$\frac{1}{g^2(x)+\varepsilon}<\frac{1}{g(x)\cdot \mathrm{E}\,\tilde{g}_n(x)}<\frac{1}{g^2(x)-\varepsilon}$$

hold.

4. The speed of the convergence of the mean square error $\mathrm{E}\,[\tilde{\eta}_n(x)-\eta(x)]^2$

In this section we examine the speed of the convergence of the mean square error $\mathrm{E}\,[\tilde{\eta}_n(x)-\eta(x)]^2$ stated in Section 3.

According to Wahba [7], we set up the following formation. Let $W_r^m(M)$ be the Sobolev space of functions whose first $m-1$ derivatives are absolutely continuous, and whose mth derivative is in $L_r(\mathcal{R}, \mathcal{B}, \mu)$. Let

$$(4.1)\qquad \|f^{(m)}\|_r \overset{\leftarrow}{\equiv} \begin{cases} \left[\int_{-\infty}^{\infty} |f^{(m)}(x)|^r \mu(dx)\right]^{1/r}, & \text{if } r \geqq 1, \\ \mu\text{-ess sup}\, |f^{(m)}(x)|, & \text{if } r=\infty, \end{cases}$$

and let

$$(4.2)\qquad W_r^{(m)}(M) \overset{\leftarrow}{\equiv} \{f;\ f \in W_r^{(m)},\ \|f^{(m)}\|_r \leqq M\},$$

where M is a positive constant. We assume $mr>1$.

In this section we assume that g and k are members of $W_r^{(m)}(M)$.

Further for the kernel K we add the following conditions to those stated in Section 2;

5° $\int_{-\infty}^{\infty} x^i K(x)\mu(dx)=0 \ (i=1, 2, \cdots, m-1)$,

6° $\int_{-\infty}^{\infty} |x|^{(m-1)/r} |K(x)|\mu(dx)<\infty$.

Now we examine the speed of the convergence of each term of the right-hand side of (3.1) in Section 3, under the above-mentioned assumptions and those stated in Theorem 3.1.

I For the term $Q_n(x) \rightleftarrows \left[\frac{\mathrm{E}\,\tilde{k}_n(x)}{\mathrm{E}\,\tilde{g}_n(x)} - \frac{k(x)}{g(x)}\right]$

Let

$$\Delta\,\mathrm{E}\,\tilde{g}_n(x) \rightleftarrows [\mathrm{E}\,\tilde{g}_n(x) - g(x)]/g(x)\ . \tag{4.3}$$

Then, we have

$$Q_n(x) = \frac{1}{g(x)}\left[\frac{\mathrm{E}\,\tilde{k}_n(x)}{1+\Delta\,\mathrm{E}\,\tilde{g}_n(x)} - k(x)\right]\ . \tag{4.4}$$

Since $\mathrm{E}\,\tilde{g}_n(x) \to g(x)$ as $n \to \infty$, there exists a natural number n_0 such that if $n \geqq n_0$, then

$$|\Delta\,\mathrm{E}\,\tilde{g}_n(x)| < 1\ . \tag{4.5}$$

Consequently by the Taylor theorem there exists a constant θ_0 such that $0<\theta_0<1$ and

$$\frac{1}{1+\Delta\,\mathrm{E}\,\tilde{g}_n(x)} = 1 - \Delta\,\mathrm{E}\,\tilde{g}_n(x) + \frac{[\Delta\,\mathrm{E}\,\tilde{g}_n(x)]^2}{[1+\theta_0\Delta\,\mathrm{E}\,\tilde{g}_n(x)]^3} \tag{4.6}$$

hold, if $n \geqq n_0$.

Hence by (4.4) we have

$$Q_n(x) = \frac{1}{g(x)}\left[\{\mathrm{E}\,\tilde{k}_n(x) - k(x)\} - \mathrm{E}\,\tilde{k}_n(x)\cdot\Delta\,\mathrm{E}\,\tilde{g}_n(x) + \frac{\mathrm{E}\,\tilde{k}_n(x)[\Delta\,\mathrm{E}\,\tilde{g}_n(x)]^2}{[1+\theta_0\Delta\,\mathrm{E}\,\tilde{g}_n(x)]^3}\right]\ . \tag{4.7}$$

Here, by Wahba's result [7], we have

$$|\mathrm{E}\,\tilde{k}_n(x) - k(x)| \leqq MA[h(n)]^{m-1/r}\ , \tag{4.8}$$

$$|\mathrm{E}\,\tilde{g}_n(x) - g(x)| \leqq MA[h(n)]^{m-1/r}\ , \tag{4.9}$$

and hence

$$|\mathrm{E}\,\tilde{k}_n(x)| \leqq |k(x)| + MA[h(n)]^{m-1/r}\ , \tag{4.10}$$

$$|\mathrm{E}\,\tilde{g}_n(x)| \leqq |g(x)| + MA[h(n)]^{m-1/r}\ , \tag{4.11}$$

where

$$A \rightleftarrows \frac{1}{(m-1)!\,[(m-1)q+1]^{1/q}} \int_{-\infty}^{\infty} |K(x)|\,|x|^{m-1/r} \mu(dx) \tag{4.12}$$

with $1/r + 1/q = 1$.

Consequently, as the evaluation of $Q_n(x)$, we have

$$(4.13)\qquad |Q_n(x)| \leqq \frac{MA\{g(x)+|k(x)|\}}{g^2(x)}[h(n)]^{m-1/r}[1+O([h(n)]^{m-1/r})]$$

for $n \geqq n_0$.

II For the term $R_n(x) \rightleftharpoons \mathrm{E}\left[\frac{\tilde{k}_n(x)}{\tilde{g}_n(x)} - \frac{\mathrm{E}\,\tilde{k}_n(x)}{\mathrm{E}\,\tilde{g}_n(x)}\right]$

Since $\mathrm{E}\,\tilde{g}_n(x) \to g(x)$ as $n \to \infty$, there exists a natural number n_1' such that $\mathrm{E}\,\tilde{g}_n(x) > 0$ for $n \geqq n_1'$. Hence let

$$(4.14)\qquad \Delta\,\tilde{g}_n(x) \rightleftharpoons [\tilde{g}_n(x) - \mathrm{E}\,\tilde{g}_n(x)]/\mathrm{E}\,\tilde{g}_n(x)$$

for $n \geqq n_1'$. Then we have

$$(4.15)\qquad R_n(x) = \frac{1}{\mathrm{E}\,\tilde{g}_n(x)}\,\mathrm{E}\left[\frac{\tilde{k}_n(x)}{1+\Delta\tilde{g}_n(x)} - \mathrm{E}\,\tilde{k}_n(x)\right].$$

Since $\Delta\tilde{g}_n(x) \to 0$ as $n \to \infty$ (by Lemma 2.2 and the fact that $\mathrm{E}\,\tilde{g}_n(x) \to g(x)$ as $n \to \infty$), there exists a natural number n_1'' such that

$$(4.16)\qquad |\Delta\tilde{g}_n(x)| < 1$$

holds for $n \geqq n_1''$ with probability 1. So let

$$(4.17)\qquad n_1 \rightleftharpoons \max(n_1', n_1'').$$

By the Taylor theorem, for some θ_1 satisfying $0 < \theta_1 < 1$

$$(4.18)\qquad \frac{1}{1+\Delta\tilde{g}_n(x)} = 1 - \frac{\Delta\tilde{g}_n(x)}{[1+\theta_1\Delta\tilde{g}_n(x)]^2}$$

holds with probability 1, if $n \geqq n_1$.

Hence, in view of (4.15) and (4.18) we have

$$(4.19)\qquad R_n(x) = \frac{-1}{\mathrm{E}\,\tilde{g}_n(x)}\,\mathrm{E}\left[\frac{\tilde{k}_n(x)\Delta\tilde{g}_n(x)}{\{1+\theta_1\Delta\tilde{g}_n(x)\}^2}\right]$$

for $n \geqq n_1$. Here (4.14) gives that there exists a positive number C such that

$$(4.20)\qquad \frac{1}{\{1+\theta_1\Delta\tilde{g}_n(x)\}^2} \leqq C$$

for $n \geqq n_1$ with probability 1. Since $\Delta\tilde{g}_n(x) \to 0$ as $n \to \infty$, this C satisfies

$$(4.21)\qquad C = 1 + o(1)$$

for sufficiently large n with probability 1. Hence, under Assumption (A) and the assumption that the conditional variance of Y given X has a finite value for each fixed $x \in \mathcal{R}$, we can apply the Schwarz inequality to the right-hand side of (4.19) in the following way;

$$\text{(4.22)}\quad \left|\mathrm{E}\left[\frac{\tilde{k}_n(x)\Delta\tilde{g}_n(x)}{\{1+\theta_1\Delta\tilde{g}_n(x)\}^2}\right]\right| \leqq \sqrt{\mathrm{E}\,\frac{1}{\{1+\theta_1\Delta\tilde{g}_n(x)\}^4}}\sqrt{\mathrm{E}\,\{\tilde{k}_n(x)\Delta\tilde{g}_n(x)\}^2} \leqq C\sqrt{\mathrm{E}\,\{\tilde{k}_n(x)\Delta\tilde{g}_n(x)\}^2}$$

for $n \geqq n_1$. However Wahba's result (4.4a) in [7] gives

$$\text{(4.23)}\quad \left|\frac{1}{h(n)}\int_{-\infty}^{\infty} K^2\left(\frac{x-u}{h(n)}\right)k(u)\mu(du) - Bk(x)\right| \leqq MA_2[h(n)]^{m-1/r},$$

where

$$\text{(4.24)}\quad A_2 \overset{\leftarrow}{=} \frac{1}{(m-1)![(m-1)q+1]^{1/q}}\int_{-\infty}^{\infty} K^2(u)\,|u|^{m-1/r}\mu(du),$$

$$\text{(4.25)}\quad B \overset{\leftarrow}{=} \int_{-\infty}^{\infty} K^2(u)\mu(du).$$

Further let

$$\text{(4.26)}\quad k^*(x) \overset{\leftarrow}{=} g(x)\int_{-\infty}^{\infty} y^2 f(y\,|\,x)\mu(dy),$$

where $f(y\,|\,x)$ is the conditional density function of Y given X. This function k^* has a finite value for each fixed $x \in \mathcal{R}$ under Assumption (A) and the above assumption of the conditional variance of Y given X. Since

$$(nh(n))\operatorname{Var}(\tilde{k}_n(x)) = h(n)\operatorname{Var}\left[\{h(n)\}^{-1}YK\left(\frac{x-X}{h(n)}\right)\right] = \{h(n)\}^{-1}\int_{-\infty}^{\infty} K^2\left(\frac{x-u}{h(n)}\right)k^*(u)\mu(du) - h(n)\cdot\left[\{h(n)\}^{-1}\int_{-\infty}^{\infty} K\left(\frac{u}{h(n)}\right)k(x-u)\mu(du)\right]^2,$$

$$\text{(4.27)}\quad \mathrm{E}\,\{\tilde{k}_n(x) - \mathrm{E}\,\tilde{k}_n(x)\}^2 \leqq \frac{B\,|k^*(x)|}{nh(n)}[1+O\{[h(n)]^{m-1/r} + O(h(n))\}],$$

applying (4.20) with k replaced by k^* to the first term and Parzen's result (Theorem 1A) to the second term, and similarly

$$\text{(4.28)}\quad \mathrm{E}\,[\tilde{k}_n(x)\tilde{g}_n(x)]^2 \leqq \frac{B[k^*(x)]^2}{g(x)nh(n)}[1+O\{[h(n)]^{m-1/r}\}]^2 \cdot \left\{[1+O\{[h(n)]^{m-1/r}\}]^3 + O\left(\frac{1}{n}\right) + O\left[\frac{1}{nh(n)}\right] + O[h(n)]\right\}.$$

Applying (4.9), (4.22) and (4.28) to (4.19), we have

$$(4.29)\quad |R_n(x)| \leqq \frac{B^{1/2}C|k^*(x)|}{[g(x)]^{3/2}} \frac{1}{[nh(n)]^{1/2}} [1+O\{h(n)\}^{m-1/r}] \cdot \sqrt{[1+O\{[h(n)]^{m-1/r}\}]^2 \left\{[1+O\{[h(n)]^{m-1/r}\}]^3 + O\left(\frac{1}{n}\right) + O\left[\frac{1}{nh(n)}\right] + O[h(n)]\right\}}$$

for $n \geqq n_1$.

III For the term $S_n(x) \rightleftarrows \mathrm{E}\left[\frac{\tilde{k}_n(x)}{\tilde{g}_n(x)} - \frac{\mathrm{E}\,\tilde{k}_n(x)}{\mathrm{E}\,\tilde{g}_n(x)}\right]^2$

By (4.18) we have

$$(4.30)\quad S_n(x) = \frac{1}{[\mathrm{E}\,\tilde{g}_n(x)]^2} \mathrm{E}\left[\{\tilde{k}_n(x) - \mathrm{E}\,\tilde{k}_n(x)\} - \frac{\tilde{k}_n(x)\Delta\tilde{g}_n(x)}{\{1+\theta_1\Delta\tilde{g}_n(x)\}^2}\right]^2 .$$

The same method as (4.22) has been obtained gives

$$(4.31)\quad \left|\frac{\mathrm{E}\,[\{\tilde{k}_n(x) - \mathrm{E}\,\tilde{k}_n(x)\}\{\tilde{k}_n(x)\Delta\tilde{g}_n(x)\}]}{\{1+\theta_1\Delta\tilde{g}_n(x)\}^2}\right| \leqq C\sqrt{\mathrm{E}\,[\tilde{k}_n(x) - \mathrm{E}\,\tilde{k}_n(x)]^2}\sqrt{\mathrm{E}\,[\tilde{k}_n(x)\Delta\tilde{g}_n(x)]^2} .$$

Consequently, applying (4.27), (4.28) and (4.31) to (4.30), we have for $n \geqq n_1$

$$(4.32)\quad S_n(x) \leqq \frac{B|k^*(x)|g(x) + BC[k^*(x)]^{3/2}[g(x)]^{1/2} + BC^2[k^*(x)]^2}{g^3(x)} \cdot \frac{1}{nh(n)}(1+o(1)) ,$$

bounding the factors $[1+O\{h(n)\}^{m-1/r} + O\{h(n)\} + O(1/n) + O(1/(nh(n)))]$ by $[1+o(1)]$.

In view of (3.1), I, II, and III, we can obtain the following result, bounding the factors $[1+O([h(n)]^{m-1/r}) + O(h(n)) + O(1/n) + O(1/(nh(n)))]$ by $[1+o(1)]$.

THEOREM 4.1. *Suppose that the kernel K satisfies the conditions* 1°–6°. *Suppose that the conditional variance of Y given X has a finite value for each $x \in \mathfrak{R}$. Furthermore suppose that g, k and k^* are members of $W_r^{(m)}(M)$ defined by* (4.1), (4.2) *and* (4.26). *If $x \in \mathfrak{R}$ with $g(x) \neq 0$ is a continuous point of g, η, and f for each fixed $y \in \mathfrak{R}$, then the following inequality holds for sufficiently large n* ($n \geqq \max(n_0, n_1)$);

$$\begin{aligned}\mathrm{E}\,[\tilde{\eta}_n(x)-\eta(x)]^2 \leqq Q^2(x)\,[h(n)]^{2m-2/r}[1+o(1)] \\ +2Q(x)R(x)\frac{[h(n)]^{m-1/r}}{[nh(n)]^{1/2}}[1+o(1)] \\ +S(x)\frac{1}{nh(n)}[1+o(1)]\,,\end{aligned}$$

where

$Q(x) \rightleftarrows MA\{g(x)+|k(x)|\}/g^2(x)$,

$R(x) \rightleftarrows \sqrt{B}\,C|k^*(x)|/[g(x)]^{3/2}$,

$S(x) \rightleftarrows \{B|k^*(x)|g(x)+BC|k^*(x)|^{3/2}[g(x)]^{1/2}+BC^2[k^*(x)]^2\}/g^3(x)$.

Constants M, A, B, and C are given by (4.2), (4.11), (4.25), *and* (4.20), *respectively.*

Remark 4.1. In Theorem 4.1 the natural number n_0 (given by (4.5)) can be determined explicitly by (4.9). For $n_1=\max(n_1', n_1'')$, n_1' (satisfying $\mathrm{E}\,\tilde{g}_n(x)>0$) is also determined by (4.9). Determination of n_1'' may be obtained by the result of Ghosh and Sen (Corollary 2.2 in [5]), if n_1'' is sufficiently large.

5. Consistency of the estimator $\hat{\eta}_n(x)$

In this section we consider the case where the marginal density function g is known. Namely we take the statistic

$$\hat{\eta}_n(x)=\frac{1}{g(x)nh(n)}\sum_{i=1}^{n} Y_i K\left(\frac{x-X_i}{h(n)}\right) \tag{5.1}$$

instead of $\tilde{\eta}_n(x)$ as an estimator of $\eta(x)$ in this case.

For this estimator we can obtain the analogous results to those in the previous sections.

THEOREM 5.1. *Suppose that the kernel K satisfies the conditions* 1°–4°. *If $x \in \mathcal{R}$ with $g(x)\neq 0$ is a continuous point of g, η, and f for each fixed $y \in \mathcal{R}$, then*

$$\mathrm{P}\,\{\lim_{n\to\infty}|\hat{\eta}_n(x)-\eta(x)|=0\}=1\,.$$

THEOREM 5.2. *Suppose that K satisfies the conditions* 1°–4°. *Suppose that the conditional variance of Y given X has a finite value for each $x \in \mathcal{R}$. If x with $g(x)\neq 0$ is a continuous point of g, η, and f for each fixed $y \in \mathcal{R}$, then*

$$\lim_{n\to\infty}\mathrm{E}\,[\hat{\eta}_n(x)-\eta(x)]^2=0\,. \tag{5.2}$$

The speed of the convergence of (5.2) can be obtained directly by Wahba's result and those obtained in Section 4. Namely, we have

$$\text{(5.3)} \qquad \mathrm{E}\,[\hat{\eta}_n(x)-\eta(x)]^2 \leqq \frac{1}{g^2(x)}\left[M^2A^2\{h(n)\}^{2m-2/r}+\frac{B\,|k^*(x)|}{nh(n)} \cdot \{1+O(h(n))+o(1)\}\right].$$

Therefore, in this case, we can seek $h(n)$ explicitly which minimizes the right-hand side of (5.3). Namely we have

$$\text{(5.4)} \qquad nh(n)=\left[\frac{1}{2(m-1/r)}\,\frac{B\,|k^*(x)|}{M^2A^2}\right]^{1/(2m+1-2/r)} n^{2(m-1/r)/(2m+1-2/r)}\,.$$

Thus we have the following result.

THEOREM 5.3. *Suppose that K satisfies the conditions $1°$–$6°$. Suppose that the conditional variance of Y given X has a finite value for each $x \in \mathcal{R}$. Furthermore suppose that k^* is a member of $W_r^{(m)}(M)$.*

If $x \in \mathcal{R}$ with $g(x)\neq 0$ is a continuous point of g, η and f for each fixed $y \in \mathcal{R}$, then

$$\mathrm{E}\,[\hat{\eta}_n(x)-\eta(x)]^2 \leqq T(x)n^{-(2m-2/r)/(2m+1-2/r)}(1+o(1))\,,$$

where

$$T(x)=\frac{2m+1-2/r}{(2m-2/r)^{(2m-2/r)}g^2(x)}\{M^2A^2(B\,|k^*(x)|)^{2m-2/r}\}^{1/(2m+1-2/r)}\,.$$

Constants M, A, and B are given by (4.2), (4.11), *and* (4.25), *respectively.*

THE INSTITUTE OF STATISTICAL MATHEMATICS

REFERENCES

[1] Parzen, E. (1962). On estimation of a probability density function and mode, *Ann. Math. Statist.*, **33**, 1065-1076.
[2] Cacoullos, T. (1966). Estimation of a multivariate density, *Ann. Inst. Statist. Math.*, 18, 179-189.
[3] Schwartz, L. (1967). *Cours d'analyse*, I, Hermann, Paris.
[4] Nadaraya, É. A. (1970). Remarks on non-parametric estimates for density functions and regression curves, *Theory Prob. Appl.*, **15**, 134-137.
[5] Ghosh, M. and Sen, P. K. (1970). On the almost sure convergence Von Mises' differentiable statistical functions, *Calcutta Statist. Ass. Bull.*, **19**, 41-44.
[6] Ash, R. B. (1972). *Real Analysis and Probability*, Academic Press, New York.
[7] Wahba, G. (1975). Optimal convergence properties of variable knot, kernel, and orthogonal series methods for density estimation, *Ann. Statist.*, **3**, 15-29.