Ann. Inst. Statist. Math.
39 (1987), Part A, 549-562

MODIFIED NONPARAMETRIC KERNEL ESTIMATES OF A REGRESSION FUNCTION AND THEIR CONSISTENCIES WITH RATES*

RADHEY S. SINGH AND MANZOOR AHMAD

(Received June 5, 1985; revised May 1, 1986)

Summary

Two sets of modified kernel estimates of a regression function are proposed: one when a bound on the regression function is known and the other when nothing of this sort is at hand. Explicit bounds on the mean square errors of the estimators are obtained. Pointwise as well as uniform consistency in mean square and consistency in probability of the estimators are proved. Speed of convergence in each case is investigated.

1. Introduction

The theory of regression is concerned with the prediction of the value of a variable, called the response or dependent variable, at a given value of another (correlated) variable, called the predictor or independent variable. Prediction is needed in several practical situations. For example, an agriculturist wants to know the yield of wheat at an amount of a specified fertilizer, a metrologist wants to forecast weather several hours ahead on the basis of previous atmospheric measurements and a physician is interested in determining the weight of a patient in terms of the number of weeks he or she has been on a diet.

Let us denote the response variable by Y and the predictor variable (also known as regressor variable) by X. Then the regression of Y on X evaluated at $X=x$ is given by

$$r(x)=\mathrm{E}(Y|X=x) .$$

* Major work of this research was completed during the first author's two visits (November-December, 1983 and August-September, 1984) to the second author at the Universite du Quebec a Montreal. Part of the work of the second author was supported by the Air Force Office of Scientific Research under contract F49620-85-C-0008 while he was at the University of Pittsburgh during Spring in 1985.

Key words and phrases: Regression curve, retraction, mean square, weak, pointwise, uniform, consistency, rates.

It is well-known that the regression curve $r(X)$ of Y on X is the *best* predictor of Y in terms of x in the sense that if $t(X)$ is any other predictor of Y, then the average square error incurred due to predictor $t(X)$ is not smaller than that incurred due to predictor $r(X)$.

If the joint distribution of the two variables X and Y is known, then the prediction of Y can be made by computing the conditional expectation of Y at the desired value of X. Otherwise, the regression curve $r(X)$ is not directly available to us. In such situations, if observations $(X_1, Y_1), \cdots, (X_n, Y_n)$ on (X, Y) are at hand, then some time the theory of least square methods or that of maximum likelihood methods can be applied to estimation of $r(x)$, but this may be done *only if* the exact model (the functional form) of the regression curve is known, and, further, for the use of m.l. methods, the distribution of the *errors*

$$e_j = Y_j - \mathrm{E}(Y_j | X)$$

must also be known.

However, the population of all suitable functional forms (or of the distributions of errors) is quite often impractically large. Therefore, no matter how carefully chosen a model is adopted, there is always a possibility of misspecification. Moreover, even if the exact functional form of the regression model involving unknown parameters is known (which is extremely rare), the above methods of least squares and/or of the maximum likelihood some time does not work at all. This is especially the case when the model is the mixture of polynomial, exponential, reciprocal, logarithmic, trigonometric and/or likewise functions of the regressor variables, each involving unknown parameters.

The problems of estimation of a regression curve r when nothing is known about the functional form of r but the conditional density of X given $Y=y$ is known to belong to certain class of densities have been treated by Kale [4], Nadaraya [6], Singh and Tracy [14] and Singh [16]. Whereas in the first three of these papers, the conditional density of X given $Y=y$ is normal with mean y and variance one, and the unconditional distribution function of Y possesses a density, in the third and the fourth papers the density of X given $Y=y$ is of the form $C(y)u(x)e^{-yx}$ and $C(y)u(x)e^{-x/y}$ respectively and the distribution of Y need not possess a density. However, the methods cited in these works are too restrictive and may also lead to misspecification of the model, because the conditional density of X given $Y=y$ is rarely known or may incorrectly be specified.

The only way of avoiding misspecification of the functional form of the regression model or of the distributional form of the errors is, in fact, to assume no specific parametric functional form of the model or of the distribution of errors, that is to estimate the regression function completely nonparametrically. In recent years nonparametric estimation of the regression function r by r_n (defined below) using Rosenblatt [9]-Parzen [8] type kernel estimates of a density function has drawn considerable attention. Various asymptotic properties of the estimators r_n, known as kernel estimates of a regression function, have been studied in the literature by a number of authors. For example, Schuster [10] has proved the asymptotic normality of these estimates and Noda [7] has proved the pointwise strong consistency. Devroye and Wagner [2] and Spiegelman and Sacks [21] have proved L^p convergence of r_n in the sense that $\lim_{n\to\infty} \mathrm{E}\int |r_n(x)-r(x)|^p d\mu(x)=0$ where μ is a probability measure generated by the r.v. X. However, strong convergence (pointwise or uniform) and L^p convergence concepts differ from the pointwise and/or uniform mean square consistency concept we shall deal with. Moreover the kernel estimates r_n of a regression function based on a sample $\{(X_1, Y_1), \cdots, (X_n, Y_n)\}$ on (X, Y) considered in the above and other works are defined by $r_n=(h_n/g_n)$ where

$$h_n(x)=(n\delta)^{-1}\sum_{j=1}^{n} Y_j K((X_j-x)/\delta)$$

and

$$g_n(x)=(n\delta)^{-1}\sum_{j=1}^{n} K((X_j-x)/\delta) ,$$

with K and δ being respectively the kernel and the windowwidth functions. Hence with such an estimate, since the kernel function K could assume a zero, negative or positive value, there is always a chance of blowing up the estimate h_n/g_n itself (or of excessively overestimating the regression) in practice for any given set of data whenever g_n is near zero. To avoid this problem, we consider in this paper a modified kernel estimate which is a retraction of the function h_n/g_n to an interval $[-c_n, c_n]$ with c_n converging to infinity with certain rate.

In Section 2 we introduce our modified kernel estimate of the regression function. In Section 3 we prove pointwise mean square consistency and deduce from it the weak consistency of our estimates. In each case the speed of convergence is examined. An explicit bound for the mean square error, lacking to date in the literature for the kernel type regression estimates, is also obtained. In Section 4 uniform mean square and uniform weak consistencies are proved and their speeds of convergence are investigated. In Section 5 remarks are made on the choice of windowwidth function, kernel function and the sequence $\{c_n\}$.

Throughout this paper convergence of a function depending on n is w.r.t. $n\to\infty$. The integrals without showing the limits are over the whole real line.

2. Estimators of regression curves

Let f be the joint density of the regressor variable X and the response variable Y and, let

$$h(x)=\int yf(x, y)dy \qquad \text{and} \qquad g(x)=\int f(x, y)dy .$$

Then the regression curve of Y on X evaluated at $X=x$ is

$$r(x)=\mathrm{E}(Y|X=x)=\frac{h(x)}{g(x)}, \qquad \text{provided} \quad g(x)\neq 0 . \tag{2.1}$$

Our method of estimation of r involves estimation of h and g on the basis of the random sample $\{(X_1, Y_1), \cdots, (X_n, Y_n)\}$ on (X, Y).

Let s be a positive integer and K_s be the class of all real valued Borel measurable bounded functions K such that

$$\begin{aligned} &\int K(y)dy=1 , \qquad \int y^j K(y)dy=0 \qquad \text{for} \quad j=1,\cdots, s-1 , \\ &\int |y|^s |K(y)|dy<\infty \qquad \text{and} \qquad |yK(y)|\to 0 \qquad \text{as} \quad |y|\to\infty . \end{aligned} \tag{2.2}$$

Kernels of the type (2.2) have been used in density estimates by Johns and Van Ryzin [3], and Singh [12], [17], among others. For any given s, the class K_s is quite large. For example, for $s=1$ and 2, $K(y)=(2\pi)^{-1/2}\exp(-y^2/2)I\ (-\infty<y<\infty)$ or $K(y)=(2a)^{-1}I\ (-a<y<a)$ for an $a>0$ belong to K_s. For $s=3$ and 4, the functions $K(y)=(2\pi)^{-1/2}[2\exp(-y^2/2)-(1/2)\exp(-y^2/4)]I\ (-\infty<y<\infty)$ or $K(y)=(2\pi)^{-1/2}(1/2)(3-y^2)\exp(-y^2/2)I\ (-\infty<y<\infty)$ belong to K_s. For any given s, polynomials $K(y)$ in y on a finite interval (a, b) belonging to K_s can be constructed (see e.g., Singh [18]).

Let $\delta=\delta_n$ and $\eta=\eta_n$ be two positive sequences of numbers based on the sample size n so that $\max\{\delta_n, \eta_n\}\to 0$ as $n\to\infty$. Let x be a point at which we wish to estimate $M(x)$. For a fixed s, let K be a fixed member of K_s. Let

$$\hat{h}(x)=(n\delta)^{-1}\sum_{j=1}^{n} Y_j K\left(\frac{X_j-x}{\delta}\right) \tag{2.3}$$

and

$$\hat{g}(x)=(n\eta)^{-1}\sum_{j=1}^{n} K\left(\frac{X_j-x}{\eta}\right) . \tag{2.4}$$

Let

$$r_n(x)=\frac{\hat{h}(x)}{\hat{g}(x)} . \tag{2.5}$$

In the existing literature most of the kernel type estimates of the regression curve are exactly of the type (2.5). However as noted earlier, $\hat{g}(x)$ could be zero or near zero at a number of points x for any given set of data on (X, Y) with a number of symmetric kernels K. In such situations it is hardly advisable to use r_n as an estimate of r. To avoid such problems, we propose a modification of r_n in this paper and study pointwise as well as uniform consistencies of the modified estimator.

For a positive b, let $\{a\}_b$ stand for $-b$, a or b according as $a<-b$, $|a|\leqq b$ or $a>b$. Let $c_n=c_n(x)$ be a positive function of n and x which, for each x, converges to infinity as $n\to\infty$ (see comments in Section 5 on the choice of c_n). Our proposed estimator of $r(x)$ is

$$\hat{r}(x)=\left\{\frac{\hat{h}(x)}{\hat{g}(x)}\right\}_{c_n} .$$

However, if we have the knowledge of some function $c_0(x)$ such that $-c_0(x)\leqq r(x)\leqq c_0(x)$, our proposed estimator of $r(x)$ would be

$$r^*(x)=\left\{\frac{\hat{h}(x)}{\hat{g}(x)}\right\}_{c_0} .$$

A discussion on the choice of c_n, the bandwidth functions δ and η and the kernel function K is made in Section 5.

3. Pointwise consistencies with an upper bound for mean square errors

In this section we prove the pointwise mean square consistency (and hence also the consistency in probability) of our estimators $\hat{r}$ and r^*, and obtain the speed of convergence in each case. In the sequel we also prove the mean square consistency of $\hat{h}$ and $\hat{g}$ as estimators of h and g and establish the speed of convergence. An explicit bound for the mean square errors of $\hat{r}$ and r^* are also obtained.

We denote

$$g_s(x)=\int f^{(s,0)}(x, y)dy ,$$

where $f^{(s,0)}(x, y)=\partial^s f(x, y)/\partial x^s$,

$$h_s(x)=\int y f^{(s,0)}(x, y)dy$$

and

$$p(x)=\int y^2 f(x, y)dy .$$

Under certain regularity conditions, g_s and h_s are the s-th partial derivatives of g and h. We, however, make no such regularity assumptions. Whenever there is no ambiguity, we will not display the argument x in $r(x)$, $\hat{r}(x)$, $r^*(x)$, $c_n(x)$, $h(x)$, $g(x)$, $h_s(x)$, $g_s(x)$ and $p(x)$ throughout this paper.

THEOREM 3.1. *Let h_s, g_s and p be continuous at x and $g(x)>0$. Then*

$$\mathrm{E}\,(\hat{r}(x)-r(x))^2=O(c_n^2\cdot\gamma_n) \tag{3.1}$$

where

$$\gamma_n=\max\{\delta^{2s}, \eta^{2s}, (n\delta)^{-1}, (n\eta)^{-1}\} .$$

To prove the theorem we will need the following lemma due to Singh [13].

LEMMA 3.1. *If g in the definition of r is not zero, then for every $L>0$,*

$$\mathrm{E}\left(\left|\frac{\hat{h}}{\hat{g}}-r\right|\wedge L\right)^2\leqq 8(g)^{-1}\left[\mathrm{E}\,(\hat{h}-h)^2+\left(|r|^2+\frac{L^2}{2}\right)\mathrm{E}\,(\hat{g}-g)^2\right] . \tag{3.2}$$

PROOF. The inequality is a special case of Lemma in the Appendix of Singh [13] and hence it does not need a separate proof.

In the next two lemmas we prove the mean square consistencies of $\hat{h}$ as an estimator of h and of $\hat{g}$ as an estimator of g respectively, and in each case we obtain rates of convergence. With some choices of δ and η these rates are of the order $O(n^{-2s/(1+2s)})$ as noted by Rosenblatt [9] for $s=2$, and Singh [12], [17], among others, and hence can be made arbitrarily close to $O(n^{-1})$ by taking s sufficiently large.

LEMMA 3.2. *Let h_s and p be continuous at x. Then the asymptotic behavior of the mean square error of $\hat{h}$ at x is given by*

$$\begin{aligned}\mathrm{MSE}\,(\hat{h}(x))&=\mathrm{E}\,(\hat{h}(x)-h(x))^2\\&\sim\left[\left(\frac{\delta^s}{s!}h_s(x)\int t^s K(t)\right)^2+(n\delta)^{-1}p(x)\int K^2\right] .\end{aligned} \tag{3.3}$$

PROOF. We first obtain the asymptotic behaviors of $\mathrm{E}\,\hat{h}$ and $\mathrm{var}\,(\hat{h})$. Then we combine these to obtain (3.3).

Since $(X_1, Y_1), \cdots, (X_n, Y_n)$ are i.i.d. with joint density f, from (2.3), we can write

$$\mathrm{E}\,\hat{h}(x)=\int\int yK(t)f(x+\delta t, y)dtdy\,. \tag{3.4}$$

As in Singh [12], expanding $f(x+\delta t, y)$ at (x, y) in δt by Taylor series expansion with the integral form of the remainder, we get

$$\begin{aligned}f(x+\delta t, y)=&\sum_{j=0}^{s-1}\frac{(\delta t)^j}{j!}f^{(j,0)}(x, y)\\ &+\frac{1}{(s-1)!}\int_x^{x+\delta t}(x+\delta t-u)^{s-1}f^{(s,0)}(u, y)du\,.\end{aligned}$$

In view of this expansion and the orthogonality properties (2.2) of K we get from (3.4),

$$\begin{aligned}\mathrm{E}\,\hat{h}(x)=&\int yf(x, y)dy\\ &+\int\int yK(t)\left\{\frac{1}{(s-1)!}\int_x^{x+\delta t}(x+\delta t-u)^{s-1}f^{(s,0)}(u, y)du\right\}dtdy\,.\end{aligned} \tag{3.5}$$

Thus,

$$\begin{aligned}\delta^{-s}\,\mathrm{E}\,(\hat{h}(x)-h(x))=\frac{\delta^{-s}}{(s-1)!}\int\int yK(t)\Big\{\int_x^{x+\delta t}&(x+\delta t-u)^{s-1}\\ &\cdot f^{(s,0)}(u, y)du\Big\}dtdy\,.\end{aligned} \tag{3.6}$$

But since x is a point of continuity of $h_s(x)=\int yf^{(s,0)}(x, y)dy$, K is bounded with $|yK(y)|\rightarrow 0$ as $|y|\rightarrow\infty$, by arguments used in Singh [12] or in Menon, Prasad and Singh [5], the r.h.s. of (3.6) is, as $n\rightarrow\infty$, asymptotically equivalent to

$$\frac{\delta^{-s}}{(s-1)!}\int yf^{(s,0)}(x, y)\int K(t)\int_x^{x+\delta t}(x+\delta t-u)^{s-1}dudtdy=\frac{h_s}{s!}\int t^sK(t)dt\,,$$

and we conclude that, as $n\rightarrow\infty$,

$$(\mathrm{E}\,\hat{h}(x)-h(x))\sim\delta^s\left(\frac{h_s(x)}{s!}\int t^sK(t)dt\right)\,. \tag{3.7}$$

Now we will evaluate the variance of $\hat{h}$. By a change of variable we see that

$$\delta^{-1}\,\mathrm{E}\left[Y_1K\left(\frac{X_1-x}{\delta}\right)\right]^2=\int\int K^2(t)y^2f(x+\delta t, y)dtdy\,. \tag{3.8}$$

Since p is continuous at x, by arguments similar to those given in Lemma 1 of Parzen [8], the r.h.s. of (3.8) is asymptotically equivalent to $p(x)\int K^2$. Further, since

$$\delta^{-1}\left[\mathrm{E}\,Y_1K\left(\frac{X_1-x}{\delta}\right)\right]^2=\delta\left[\int\int yK(t)f(x+\delta t,\,y)dtdy\right]^2=\delta(\mathrm{E}\,\hat{h}(x))^2$$

by (3.4), we have from (3.6), $\delta^{-1}[\mathrm{E}\,Y_1K((X_1-x)/\delta)]^2=o(1)$. Thus since $(X_1, Y_1),\cdots,(X_n, Y_n)$ are i.i.d., we conclude that

$$\operatorname{var}(\hat{h}(x))\sim(n\delta)^{-1}p(x)\int K^2\,. \tag{3.9}$$

Now (3.7) and (3.9) give (3.3). This completes the proof of Lemma 3.2.

LEMMA 3.3. *If g_s is continuous at x, then*

$$\mathrm{MSE}\,(\hat{g}(x))\sim\left[\left(\frac{n^s}{s!}g_s(x)\int t^sK(t)\right)^2+(n\eta)^{-1}g(x)\int K^2\right], \tag{3.10}$$

and if, instead, $g^{(s)}$, the s-th order derivative of g is continuous at x, then (3.10) *holds with g_s replaced by $g^{(s)}$.*

PROOF. Proof of (3.10) follows by arguments given for (3.3).

Remark 3.1. Taking δ and η proportional to $n^{-1/(2s+1)}$, we see from (3.3) and (3.10) that MSE $(\hat{h})$ and MSE $(\hat{g})$ are both of the order $O(n^{-2s/(1+2s)})$. The value of δ that minimizes the r.h.s. of (3.3) and that of η that minimizes the r.h.s. of (3.10), are, nevertheless, given by

$$\delta^*=\left|\frac{n^{-1}p(x)\int K^2}{2s\left(h_s(x)\int t^sK(t)/s!\right)^2}\right|^{1/(1+2s)} \tag{3.11}$$

and

$$\eta^*=\left|\frac{n^{-1}g(x)\int K^2}{2s\left(g_s(x)\int t^sK(t)/s!\right)^2}\right|^{1/(1+2s)} \tag{3.12}$$

respectively. Using these optimal values of δ and η one can easily obtain the asymptotic values of the mean square errors of $\hat{h}$ and $\hat{g}$ which are minimum over the class of all windowwidth functions δ and η. However, since the exact value of the ratio $p(x)/h_s^2(x)$ for δ^* and of the ratio $g(x)/g_s^2(x)$ for η^* are not known, only approximate values of δ^* and η^* (by getting approximate values of these ratios) can be used in practice. The expression for η^* is noted in Rosenblatt [9] for $s=2$ and in Singh [15] for general s, among many others.

PROOF OF THEOREM 3.1. Writing $|\hat{r}-r|=|(\hat{r}-(r)_{c_n})+((r)_{c_n}-r)|$, we have with probability one,

$$|\hat{r}-r|\leqq\left(\left|\frac{\hat{h}}{\hat{g}}-r\right|\wedge c_n\right)+|r|I(|r|>c_n)\ .$$

Hence by Lemma 3.1,

$$\text{(3.13)}\quad \mathrm{E}\,(\hat{r}-r)^2\leqq 16(g)^{-1}\left[\mathrm{E}\,(\hat{h}-h)^2+\frac{3}{2}\max\{|r|^2, c_n^2\}\,\mathrm{E}\,(\hat{g}-g)^2\right] +2|r|^2I(|r|>c_n)\ .$$

Now since $c_n\rightarrow\infty$ as $n\rightarrow\infty$, there exists an $n_0=n_0(x)$ such that for all $n\geqq n_0$, $c_n(x)\geqq|r(x)|$ and the second term on the r.h.s. of (3.13) is equal to zero for all $n\geqq n_0$. The rest of the proof is now an immediate consequence of (3.3) and (3.10).

Remark 3.2. Notice that (3.13) gives an explicit bound for each sample size for the mean square error of the estimator of the regression curve in terms of MSE $(\hat{h})$ and MSE $(\hat{g})$. Exact asymptotic expressions for these latter terms are in turn presented in (3.3) and (3.10) respectively. Hence the exact asymptotic value of the bound (3.13) for MSE $(\hat{r})$ is at hand. To the best of our knowledge an explicit bound with an exact asymptotic value for the MSE of a nonparametric regression curve estimate, of whatsoever nature it may be, is lacking in the existing literature, inspite of a large number of articles on the subject. It will be, however, interesting and challenging to obtain exact expression for $\mathrm{E}\,(r-r)^2$.

Remark 3.3. From Theorem 3.1 it follows that if δ and η are chosen in a way so that

$$\text{(3.14)}\qquad \delta\sim\eta=O(n^{-1/(1+2s)})\ ,$$

then γ_n defined in Theorem 3.1 is of the order,

$$\text{(3.15)}\qquad \gamma_n=O(n^{-2s/(1+2s)})$$

and

$$\text{(3.16)}\qquad \mathrm{MSE}\,(\hat{r}(x))=O(n^{-2s/(1+2s)}c_n^2)\ .$$

Remark 3.4. As pointed out earlier, if there is a known $c_0(x)$ such that $|r(x)|\leqq c_0(x)$, we would instead consider estimating r by r^* defined in Section 2. It follows from the proof of Theorem 3.1 that

$$\mathrm{E}\,(r^*(x)-r(x))^2=O(\gamma_n)\,. \tag{3.17}$$

Thus r^* achieves an MSE rate of convergence better than $\hat{r}$.

The following (3.18) and (3.19) are immediate consequences of (3.16) and (3.17).

COROLLARY 3.1. (*Weak consistency*). *Under the conditions of Theorem* 3.1 *and* (3.14), *for every sequence* $\alpha_n \rightarrow \infty$

$$|\hat{r}(x)-r(x)|=o(n^{-s/(1+2s)}c_n(x)\cdot\alpha_n) \qquad \textit{in prob.} \tag{3.18}$$

and

$$|\hat{r}^*(x)-r(x)|=o(n^{-1/(1+2s)}\alpha_n) \qquad \textit{in prob.} \tag{3.19}$$

Remark 3.5. It is clear from the results in (3.1), (3.15), (3.18) and (3.19) that larger the s the better the rate of convergence. However, choosing a larger value of s means putting more restrictions on h and g. Further, any choice of s more than 4 or 5 makes the computation of $\hat{h}$ and $\hat{g}$ difficult. It is seen quite often in the case of density estimates that the improvement in the rate of convergence with an s being 5 or more is not significant compared to the extra difficulty one incurs in the computation of the estimates. The same is expected in the case of regression estimates.

4. Uniform consistencies

In Section 3 we proved the mean square consistency and deduced the consistency in probability of the estimators $\hat{r}$ and r^* at a point x, and in each case we investigated the speed of convergence. In this section we plan to prove the uniform mean square consistency as well as the uniform in probability consistency of $\hat{r}$ and r^*. The following theorem follows directly from the proof of Theorem 3.1.

THEOREM 4.1. *Let B be any subset of the real line such that* $\inf_{x\in B} g(x) > 0$ *and* $\sup_{x\in B}|r(x)|<\infty$ (*the bounds in respective cases need not be known*); *and* p, h_s *and* g_s *are uniformly continuous on B. Then*

$$\sup_{x\in B}\mathrm{E}\,(\hat{r}(x)-r(x))^2=O(\gamma_n\cdot c_n^{*2}) \tag{4.1}$$

where $c_n^*=\sup_{x\in B} c_n(x)$, *and* γ_n *is as defined in Theorem* 3.1. *Also*

$$\sup_{x\in B}\mathrm{E}\,(r^*(x)-r(x))^2=O(\gamma_n)\,. \tag{4.2}$$

Thus if δ and η are proportional to $n^{-1/(1+2s)}$, then

$$(4.1)' \qquad \sup_{x \in B} \mathrm{MSE}\,(\hat{r}(x)) = O(n^{-2s/(1+2s)} \cdot c_n^{*2})$$

and

$$(4.2)' \qquad \sup_{x \in B} \mathrm{MSE}\,(r^*(x)) = O(n^{-2s/(1+2s)}) .$$

The result (4.1)′ or (4.2)′ does not, however, prove the uniform weak consistency of $\hat{r}$ or r^*. If the characteristic function of K is absolutely integrable and $\mathrm{E}|Y|^2 < \infty$, then it can be shown (see e.g., Parzen [8], Bierens [1] or Singh and Ullah [20]), that

$$(4.3) \qquad \mathrm{E}\left\{\sup_x |\hat{h}(x) - \mathrm{E}\,\hat{h}(x)|\right\} = O((n\delta)^{-1/2}) .$$

Hence it follows from Lemma 3.2 that if h_s and p are uniformly continuous on B, then

$$(4.4) \qquad \mathrm{E}\left\{\sup_{x \in B} |\hat{h}(x) - h(x)|\right\} = O(\max\{\delta^s, (n\delta)^{-1/2}\})$$

which in turn implies that, for every positive sequence $\alpha_n \to \infty$,

$$\sup_{x \in B} |\hat{h}(x) - h(x)| = o(\max\{\delta^s, (n\delta)^{-1/2}\}\alpha_n) \qquad \text{in prob.}$$

Similarly, if the characteristic function of K is absolutely integrable and g_s is uniformly continuous, then

$$(4.5) \qquad \mathrm{E}\left\{\sup_{x \in B} |\hat{g}(x) - g(x)|\right\} = O(\max\{\eta^s, (n\eta)^{-1/2}\})$$

and

$$\sup_{x \in B} |\hat{g}(x) - g(x)| = o(\alpha_n \max\{\eta^s, (n\eta)^{-1/2}\}) \qquad \text{in prob.}$$

for every positive sequence $\alpha_n \to \infty$.

To deduce the uniform weak consistency of $\hat{r}$ and r^* from the above analysis, notice that as in the proof of Theorem 3.1, $|\hat{r} - r|$ is bounded a.s. by $|(\hat{h}/\hat{g}) - (h/g)| \wedge c_n + |r| I(|r| > c_n)$, and the proof of Lemma in the Appendix of Singh [13] gives

$$\mathrm{E} \sup_{x \in B} \left(\left| \frac{\hat{h}(x)}{\hat{g}(x)} - \frac{h(x)}{g(x)} \right| \wedge c_n(x) \right) \leq 2 \left(\inf_{x \in B} g(x) \right)^{-1} \left\{ \mathrm{E} \sup_{x \in B} |\hat{h}(x) - h(x)| + \left(\sup_{x \in B} |r(x)| + c_n^* \right) \mathrm{E} \sup_{x \in B} |\hat{g}(x) - g(x)| \right\} .$$

Further, there exists an n_0 such that for all $n \geq n_0$, $\sup_{x \in B} |r(x)| I(|r(x)| > c_n(x)) \equiv 0$ (this follows because $\sup_{x \in B} |r(x)| < \infty$, though the upper bound need not be known, and $c_n(x) \to \infty$ for each x in B). From these analyses, (4.4) and (4.5) we conclude the following theorem.

THEOREM 4.2. *Let* $\mathrm{E}|Y|^2 < \infty$, *and for a subset* B *of the real line,* *the hypothesis of Theorem* 4.1 *hold. Then*

$$\mathrm{E}\left\{\sup_{x\in B}|\hat{r}(x)-r(x)|\right\}=O(\gamma_n^{1/2}\cdot c_n^*) \tag{4.6}$$

and

$$\mathrm{E}\left\{\sup_{x\in B}|r^*(x)-r(x)|\right\}=O(\gamma_n^{1/2}) . \tag{4.7}$$

Thus for any positive sequence $\alpha_n\to\infty$, it follows from (4.6), that $\sup_{x\in B}|\hat{r}(x)-r(x)|=o(\gamma_n^{1/2}c_n^*\alpha_n)$ in probability, and from (4.7) that $\sup_{x\in B}|r^*(x)-r(x)|=o(\gamma_n^{1/2}\alpha_n)$. Taking δ and η proportional to $n^{-1/(1+2s)}$, γ_n is of the order $n^{-2s/(1+2s)}$.

5. Some concluding remarks

The choice of c_n in the definition of our estimator $\hat{r}$ is completely arbitrary, and it is not possible to give an explicit formula to determine a value of c_n which may fit well in all practical situations. If, however, in a particular situation, we have some knowledge, say A_0, of the range of the possible values of the response variable Y, we may choose $c_n(x)\equiv A_0\alpha_n$ where α_n is a slowly converging to infinity sequence of n, something like $\log n$ or $\log\log n$ (depending on how good is our knowledge about the range of Y). In any case, c_n must be chosen so that $n^{-s/(1+2s)}c_n\to 0$ as $n\to\infty$.

Examining the asymptotic expressions of MSE$(\hat{h})$ and MSE$(\hat{g})$ obtained in Section 3, we remark that one should choose K so that $\left|\int t^sK(t)dt\right|$ and $\int K^2(t)dt$ be as small as possible. This is also the case even if one uses the optimal δ and η given in (3.11) and (3.12) respectively, since with these choices of δ and η, $\min(\mathrm{MSE}(\hat{h}(x)))\sim n^{-2s/(1+2s)}$. $w_1(x)$, where

$$w_1(x)=(1+2s)\left\{\frac{\left|h_s(x)\int t^sK(t)\right|}{s!}\left\{\frac{p(x)\int K^2}{2s}\right\}^s\right\}^{2/(1+2s)}$$

and

$$\min(\mathrm{MSE}(\hat{g}(x)))\sim n^{-2s/(1+2s)}w_2(x)$$

where

$$w_2(x)=(1+2s)\left\{\frac{\left|g_s(x)\int t^sK(t)\right|}{s!}\left\{\frac{g(x)\int K^2}{2s}\right\}^s\right\}^{2/(1+2s)} .$$

Now examining the optimal values of δ and η given in (3.11) and (3.12), we remark that δ and η should be proportional to $n^{-1/(1+2s)}$. (This has been pointed out in a number of articles on density estimates dealing with rates of convergence, e.g. Rosenblatt [9] for $s=2$ and Singh [15] and [17] for $s>0$.) Examining the estimates $\hat{g}$, we see that var$(\hat{g})$ will be large whenever the var$(K((X_j-x)/\eta))$ is large, which in turn will be inflated when var$(X_j)=\sigma_X^2$ (say) is large. To control this (and hence to control var$(\hat{g})$) to some extent, we remark that η should also be proportional to σ_X, that is, if possible, η should be taken to be $\sigma_0\eta'$, where σ_0 is a good guess of σ_X and η' is proportional to $n^{-1/(1+2s)}$. We have the same view with δ as well. This observation is originally made in Singh and Ullah [19] in connection of estimating a multivariate density.

Acknowledgement

Authors are grateful to NSERC of Canada for the research support.

UNIVERSITY OF GUELPH
UNIVERSITE DU QUEBEC A MONTREAL

REFERENCES

[1] Bierens, H. J. (1983). Uniform consistency of kernel estimators of a regression function under generalized conditions, *J. Amer. Statist. Ass.*, 78, 699-707.

[2] Devroye, L. P. and Wagner, T. (1980). Distribution free consistency results in nonparametric discrimination and regression function estimations, *Ann. Statist.*, 8, 231-239.

[3] Johns, M. V. and Van Ryzin, J. (1972). Convergence rates for empirical Bayes two-action problems II. Continuous case, *Ann. Math. Statist.*, **43**, 934-947.

[4] Kale, B. (1962). A note on a problem in estimation, *Biometrika*, **49**, 553-556.

[5] Menon, V. V., Prasad, B. and Singh, R. S. (1984). Nonparametric recursive estimates of a probability density function and its derivations, *J. Statist. Plann. Inference*, **9**, 73-82.

[6] Nadaraya, E. A. (1965). On nonparametric estimates of density function and regression curves, *Theor. Prob. Appl.*, **10**, 186-190.

[7] Noda, K. (1976). Estimation of a regression function by the Parzen kernel-type density estimators, *Ann. Inst. Statist. Math.*, **28**, 221-234.

[8] Parzen, E. (1962). On estimation of a probability density function and mode, *Ann. Math. Statist.*, **33**, 1065-1076.

[9] Rosenblatt, M. (1956). Remarks on some nonparametric estimators of density function, *Ann. Math. Statist.*, **27**, 832-837.

[10] Schuster, E. F. (1972). Joint asymptotic distribution of the estimated regression function at a finite number of distinct points, *Ann. Math. Statist.*, **43**, 84-88.

[11] Schuster, E. F. and Yakowitz, S. (1979). Contributions to the theory of nonparametric regression, with application to system identification, *Ann. Statist.*, **7**, 139-149.

[12] Singh, R. S. (1977a). Improvement on some known nonparametric uniformly consistent estimators of derivatives of a density, *Ann. Statist.*, **5**, 394-399.

[13] Singh, R. S. (1977b). Applications of estimators of a density and its derivatives to

certain statistical problems, *J. R. Statist. Soc.*, **39**, 357-363.

[14] Singh, R. S. and Tracy, D. S. (1977). Strongly consistent estimators of k-th order regression curves and rates of convergence, *Zeit. Wahrscheinlichkeitsth.*, **40**, 339-348.

[15] Singh, R. S. (1979). Mean square errors of estimates of a density and its derivatives, *Biometrika*, **66**, 177-180.

[16] Singh, R. S. (1980). Estimation of regression curves when the conditional density of the predictor variable is in scale exponential family, *Multivar. Statist. Anal.* (ed. R. P. Gupta).

[17] Singh, R. S. (1981a). On the exact asymptotic behaviour of estimators of a density and its derivatives, *Ann. Statist.*, **9**, 453-456.

[18] Singh, R. S. (1981b). Speed of convergence in nonparametric estimation of a multivariate μ-density and its mixed partial derivatives, *J. Statist. Plann. Inference*, **5**, 287-298.

[19] Singh, R. S. and Ullah, A. (1984). Nonparametric recursive estimation of a multivariate, marginal and conditional DGP with an application to specification of econometric models, to appear in *Commun. Statist.*

[20] Singh, R. S. and Ullah, A. (1985). Nonparametric time series estimation of joint DGP, conditional DGP and vector autoregression, *Econ. Theory*, **1**, 27-52.

[21] Spiegelman, C. and Sacks, J. (1980). Consistent window estimation in nonparametric regression, *Ann. Statist.*, **33**, 1065-1076.