Ann. Inst. Statist. Math.
Vol. 43, No. 2, 327–346 (1991)

OPTIMAL CONVERGENCE PROPERTIES OF KERNEL DENSITY ESTIMATORS WITHOUT DIFFERENTIABILITY CONDITIONS*

R. J. KARUNAMUNI AND K. L. MEHRA

Department of Statistics and Applied Probability, University of Alberta, Edmonton, Alberta, Canada T6G 2G1

(Received June 14, 1989; revised February 15, 1990)

Abstract. Let $X_1, X_2, \dots, X_n$ be independent observations from an (unknown) absolutely continuous univariate distribution with density f and let $\hat{f}(x) = (nh)^{-1}\sum_{i=1}^{n} K[(x - X_i)/h]$ be a kernel estimator of $f(x)$ at the point x, $-\infty < x < \infty$, with $h = h_n$ ($h_n \to 0$ and $nh_n \to \infty$, as $n \to \infty$) the bandwidth and K a kernel function of order r. "Optimal" rates of convergence to zero for the bias and mean square error of such estimators have been studied and established by several authors under varying conditions on K and f. These conditions, however, have invariably included the assumption of existence of the r-th order derivative for f at the point x. It is shown in this paper that these rates of convergence remain valid without any differentiability assumptions on f at x. Instead some simple regularity conditions are imposed on the density f at the point of interest. Our methods are based on certain results in the theory of semi-groups of linear operators and the notions and relations of calculus of "finite differences".

Key words and phrases: Kernel density estimation, bias, mean square error, finite differences, semi-groups, linear operators.

1. Introduction

Let $X_1, X_2, \dots, X_n$ denote a sample of independent observations from an unknown absolutely continuous univariate distribution with density f. Then a Rosenblatt-Parzen type kernel estimator of f at a given point x, $-\infty < x < \infty$, is given by

$$\hat{f}(x) = (nh)^{-1}\sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right), \tag{1.1}$$

where K is a suitable kernel function and $h = h_n$ the smoothing parameter or bandwidth ($h_n \to 0$ and $nh_n \to \infty$, as $n \to \infty$). There have been numerous

* This research was supported in part by grants from the Natural Sciences and Engineering Research Council of Canada and the University of Alberta Central Research Fund.

papers in literature setting out various estimators of f, with different rates of convergences of f relative to important discrepancy "measures" between f and the proposed estimators (e.g., bias, mean square error, integrated mean square error, etc.) and under a variety of conditions on K, h and the smoothness of f. The literature is too extensive to warrant a complete listing here (see for instance, Tapia and Thompson (1978), Devroye and Györfi (1985) and Silverman (1986) for detailed bibliography). It is also well known that "improved" rates of convergence for bias and mean square error (MSE) of kernel estimators of the type (1.1) at a point x, $-\infty < x < \infty$, can be obtained through the use of "higher-order" kernel functions (defined below) provided higher-order derivatives $f^{(r)}$, $r \geq 1$, of f exist at the point x (see, e.g., Parzen (1962), Bartlett (1963), Rosenblatt (1971), Yamato (1972), Farrell (1972), Singh (1974, 1977), Wahba (1975), Davis (1975), Müller and Gasser (1979) and Müller (1984)). A kernel function K_p ($p \geq 0$, an integer) is said to be of order r (or (p, r) with $r > p$, an integer) if

$$(1.2) \qquad \frac{1}{j!}\int y^j K_p(y)dy = \begin{cases} 1 & \text{if } \ j = p \\ 0 & \text{if } \ j \neq p, \quad j = 0, 1, 2, \ldots, r-1 \\ C_{r,p} & \text{if } \ j = r, \end{cases}$$

where $C_{r,p}$ is a non-zero finite constant depending on r and p alone. The technique for establishing the above referred "improved" rates of convergence is based on the use of such higher-order kernel functions along with an r-th order Taylor expansion of $f(x+hy)$ about x, of the type, say (see Singh (1974), Wahba (1975) and Müller and Gasser (1979)),

$$(1.3) \quad f(x+hy) = \sum_{j=0}^{r-1} \frac{(hy)^j}{j!} f^{(j)}(x) + \frac{1}{(r-1)!}\int_x^{x+hy} (x+hy-u)^{r-1} f^{(r)}(u)du$$

(where $f^0(x)$ stands for $f(x)$), in the expressions for $E[\hat{f}(x)]$ and MSE$[\hat{f}(x)]$ etc. and then deriving appropriate bounds for the discrepancy measures which yield the desired rates of convergence thereof. However, the question regarding "optimum" or "improved" rates of convergence of these measures at points x where $f(x)$ is not differentiable has been virtually ignored in literature.

There are quite a number of important continuous density functions whose domains contain non-differentiable points. A few are given below:

Example 1.1. Let f be the triangular density defined by

$$(1.4) \qquad f(x) = \begin{cases} 1+x & \text{if } \ -1 \leq x \leq 0 \\ 1-x & \text{if } \ 0 < x \leq 1 \\ 0 & \text{otherwise.} \end{cases}$$

At the point $x = 0$, f is not differentiable.

Example 1.2. Let the (triangular) function ϕ be defined by

$$\phi(x) = \begin{cases} x & \text{if } \ 0 \leq x < \frac{1}{2} \\ 1-x & \text{if } \ \frac{1}{2} \leq x \leq 1 \end{cases}$$

and $\phi(x+k) = \phi(x)$ for all integers k; and consider the density f:

$$f(x) = \sum_{i=1}^{\infty} 2^{(2-i)}\phi(2^i x), \qquad -\infty < x < \infty.$$

Then $f(x)$ is continuous on $(-\infty, \infty)$, but fails to have a finite derivative at every point except a dyadic rational.

Example 1.3. For positive integer $k > 1$, let f be the density function defined by

$$f(x) = \begin{cases} k(1-2|x|)^{k-1}, & \text{if } |x| \le \frac{1}{2} \\ 0, & \text{if } |x| > \frac{1}{2}. \end{cases} \tag{1.5}$$

At the point $x = 0$, f is not differentiable. Observe that if the random variables Y_i possess the uniform distribution on the interval $[\theta - 1/2, \theta + 1/2]$, then (1.5) represents the density function of the random variable $T - \theta$, where $T = (\max_{1 \le i \le k} Y_i + \min_{1 \le i \le k} Y_i)/2$. ($T$ is a minimax and sufficient estimator for θ.)

Example 1.4. Let f be the double exponential density defined by $f(x) = (1/2a)e^{-|x-\theta|/a}$ for $-\infty < x < \infty$, $a > 0$. Then f is not differentiable at $x = \theta$.

Example 1.5. Let f be a countable mixture of, say, double exponential densities: for $-\infty < x < \infty$,

$$f(x) = \sum_{i=0}^{\infty} \frac{1}{2^{i+1}} g_i(x), \qquad \text{with} \qquad g_i(x) = \frac{1}{2a} e^{-|x-c_i(\theta)|/a},$$

$a > 0$, $c_i(\theta) \nearrow$ as $i \nearrow \infty$ for each θ, $|\theta| < b$ ($b > 0$), $c_i(\theta) \nearrow$ in $|\theta|$ and $\searrow 0$ as $|\theta| \to 0$. Then $f(x)$ is continuous on $(-\infty, \infty)$ but is not differentiable at $x = c_i(\theta)$, $i = 0, 1, 2, \ldots$ for each given $\theta \in (a, b)$.

Example 1.6. Let f be the density function of the form (an s-parameter exponential family with Lebesgue density $U(x)$):

$$f(x) = \exp\left[\sum_{i=1}^{s} c_i(\theta) T_i(x) - B(\theta)\right] U(x),$$

for $-\infty < x < \infty$. We may take $U(x)$, for example, (i) $e^{-|x|}$ or (ii) $e^{-|x|}/(1+x^2)$. The density f is not differentiable at $x = 0$ in either case; see Lehmann ((1983), p. 62, Exercise 4.1).

More examples can be given, see, e.g., Loh (1984) and Ibiragimov and Has'minski ((1981), Chapters V and VI).

It is worth pointing out that while most densities assumed in applications—and, indeed, the examples quoted above—satisfy varying degrees of strong smoothness conditions, these densities are, in fact, only idealized approximations to what might be the true but unknown underlying state of nature. There is no reason to believe, however, that the unknown densities to be estimated in specific applications actually satisfy the assumed (r-th order, etc.) differentiability and other "smoothness" conditions—the conditions on which the traditional arguments for establishing improved rates of convergence for estimators (1.1) are usually based. It seems worthwhile to go a step further and note that, beyond the assumption of continuity, assuming strict differentiability conditions severely restricts the class of densities and, thereby, may exclude from the statement many situations for which the desired results might still be true. In this context, consider the Banach space $\mathcal{B}$ of continuous functions f defined on a real closed interval $[a, b]$ with the norm $\|f\| = \sup_{a \leq x \leq b} |f(x)|$. Then it is well known (see Hewitt and Stromberg ((1965), p. 260)) that the subclass $\mathcal{D} \subset \mathcal{B}$ of differentiable functions is of category I and, consequently, is nowhere dense in $\mathcal{B}$. In other words, much less to speak about the r-th order differentiability, most functions in $\mathcal{B}$ are, in fact, nowhere differentiable. Thus, for an overwhelmingly large class of continuous densities, the usual differentiability conditions are not satisfied. Besides, in nonparametric estimation who knows at what points the unknown density satisfies the assumed differentiability conditions and, that too, to what order! Accordingly, the weakening of any differentiability conditions on the density f, used in establishing improved rates of convergence for estimators (1.1), should be of considerable interest in applications.

To achieve the above is precisely the object of this paper. We shall show that "optimum" or improved rates of convergence to zero for the bias and mean square error (MSE) of estimator (1.1) can be obtained without any differentiability assumptions on f, provided certain types of kernel functions are employed and certain simple regularity assumptions on f are satisfied. In fact, under these simple restrictions (see (2.5), (2.6), (2.7) and (2.8)) based on notions and relations of "calculous of finite differences", we obtain the best possible rates for bias and MSE for any uniformly continuous density f, with bounds on the bias and MSE improving with appropriate choice of K and conditions on f. We may employ in (1.1) a one-sided kernel K_0, namely, a kernel vanishing off (0,1) or $(-1, 0)$, or more appropriately, a "symmetrized analogue" (see (2.3) below) of the one-sided kernel K_0. Our approach is based on certain results in the theory of semi-groups of linear operators and calculus of finite differences; indeed, the results obtained do not seem to readily follow using traditional arguments.

The main results of the paper along with some corollaries are given in Section 2. In Section 3 are given (above-referred) our main tool results and the proof of results stated in Section 2. Section 4 contains the treatment of two of the above examples, namely 1.1 and 1.4—the object being simply to illustrate the conditions under which the main results of this paper hold—and also some concluding remarks.

2. Main results

In this section, we state our main results followed by a few corollaries and remarks. First let us introduce some notation: Given a sequence $c_0, c_1, \ldots$ of reals, let Δ denote the difference operator $\Delta c_n = c_{n+1} - c_n$. Applying the operator Δ to the sequence $\{\Delta c_n\}$, we get a new sequence $\{\Delta^2 c_n\}$ and so forth. Defining Δ^i, $i = 0, 1, 2, \ldots$ recursively by the relation $\Delta^i = \Delta(\Delta^{i-1})$ ($\Delta^1 = \Delta$, $\Delta^0 c_n = c_n$), an induction argument easily shows that

$$\Delta^i c_\nu = \sum_{k=0}^{i} \binom{i}{k} (-1)^{i-k} c_{\nu+k}. \tag{2.1}$$

Let $\Delta_\eta = \eta^{-1}\Delta$, $\eta \neq 0$, denote the difference ratio operator. Thus, if $c_\nu = u(x+\eta\nu)$ for fixed x and span η, we have $\Delta_\eta u(x) = \eta^{-1}[u(x+\eta) - u(x)]$. Defining the higher-order differences again recursively by $\Delta_\eta^i = \Delta_\eta \Delta_\eta^{i-1}$, with $\Delta_\eta^1 = \Delta_\eta$ and $\Delta_\eta^0 u = u$, the equation (2.1) becomes

$$\Delta_\eta^i u(x) = \eta^{-i} \sum_{k=0}^{i} \binom{i}{k} (-1)^{i-k} u(x + k\eta). \tag{2.2}$$

We can now state the following theorems:

THEOREM 2.1. *Let $\hat{f}$ be defined by*

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^{n} \frac{1}{2} \left\{ K_0\left(\frac{x - X_i}{h}\right) + K_0\left(\frac{X_i - x}{h}\right) \right\}, \tag{2.3}$$

where the kernel function K_0 is measurable, vanishing off (0,1) and of order r ($r \geq 1$). Let f be a uniformly continuous density function defined on $(-\infty, \infty)$. For given $a > 1$, and any reals η and θ, define $A_\eta = \{k\eta\colon\ k = 0, 1, 2, \ldots\}$, denoting A_η by A_η^+ or A_η^- according as $\eta > 0$ and $\eta < 0$, and

$$\alpha(\tau, \theta) = a^{1+\tau} - a^{1+\theta}[1 + (\tau - \theta)\log a] \quad \textit{for} \quad \tau \in A_\eta, \tag{2.4}$$

denoting by $\alpha(\tau, \theta)$ by $\alpha^+(\tau, \theta)$ or $\alpha^-(\tau, \theta)$ according as $\eta > 0$ and $\eta < 0$. Assume further that, for some $\epsilon > 0$ and some $\delta > 0$,

$$\liminf_{\eta \downarrow 0} \sup_{0<\theta<\epsilon} \sup_{\tau \in S_1^+} |\Delta_\eta^r f(x + \tau)| < \infty, \tag{2.5}$$

$$\limsup_{\eta \downarrow 0} \sup_{0<\theta<\epsilon} \sup_{\tau \in S_2^+} \left| \frac{\Delta_\eta^r f(x + \tau)}{\alpha^+(\tau, \theta)} \right| < \infty, \tag{2.6}$$

$$\liminf_{\eta \uparrow 0} \sup_{-\epsilon<\theta<0} \sup_{\tau \in S_1^-} |\Delta_\eta^r f(x + \tau)| < \infty, \tag{2.7}$$

and

$$\limsup_{\eta \uparrow 0} \sup_{-\epsilon<\theta<0} \sup_{\tau \in S_2^-} \left| \frac{\Delta_\eta^r f(x + \tau)}{\alpha^-(\tau, \theta)} \right| < \infty, \tag{2.8}$$

where

$$S_1^+ = S_1^+(\theta) = \{\tau :\ \tau \in A_\eta^+, |\tau - \theta| < \delta\}, \qquad S_2^+ = S_2^+(\theta) = A_\eta^+ - S_1^+(\theta),$$
$$S_1^- = S_1^-(\theta) = \{\tau :\ \tau \in A_\eta^-, |\tau - \theta| < \delta\}, \qquad S_2^- = S_2^-(\theta) = A_\eta^- - S_1^-(\theta).$$

Then for each x, $-\infty < x < \infty$,

$$|E\hat{f}(x) - f(x)| \leq \frac{h^r}{2r!}[C_r^+(x) + C_r^-(x)] \int_0^1 |K_0(t)|t^r dt, \tag{2.9}$$

where $C_r^+(x)$ *and* $C_r^-(x)$ *are defined by the left-hand sides of* (2.5) *and* (2.7), *respectively.*

THEOREM 2.2. *Let the hypotheses be those of Theorem* 2.1 *and assume that the conditions* (2.5) *to* (2.8) *hold. If in addition, the conditions* (2.5) *to* (2.8) *with* Δ_η^r *replaced by* Δ_η *also hold (let these be denoted by* (2.5a) *to* (2.8a) *in the sequel), then for each* x, $-\infty < x < \infty$,

$$\operatorname{Var} \hat{f}(x) \leq (nh)^{-1} \left\{ f(x)a_1 + \frac{1}{2}ha_2(C_1^+(x) + C_1^-(x)) \right\}$$

and

$$E(\hat{f}(x) - f(x))^2 \leq h^{2r}a_3^2 \left[\frac{C_r^+(x) + C_r^-(x)}{2r!}\right]^2 + (nh)^{-1}\left[f(x)a_1 + \frac{1}{2}ha_2(C_1^+(x) + C_1^-(x))\right], \tag{2.10}$$

where

$$a_1 = \int_0^1 K_0^2(t)dt, \qquad a_2 = \int_0^1 tK_0^2(t)dt, \qquad a_3 = \int_0^1 t^r|K_0(t)|dt$$

and $C_r^+(x)$ *and* $C_r^-(x)$ *are as defined in Theorem* 2.1, $r \geq 1$.

Remark 2.1. Theorems 2.1 and 2.2 above are stated for the estimator (2.3), i.e. the estimator (1.1) with K replaced by a "symmetrized" analogue of kernel K_0, vanishing off (0,1). However, similar theorems would also clearly hold if the kernel K_0 used in (2.3) vanishes off $(-1, 0)$, provided (2.5) to (2.8) are replaced by the corresponding conditions with $\eta \uparrow 0$ and $\eta \downarrow 0$. The proofs in this latter case, after appropriate modifications, would be verbatim the same. One-sided kernels vanishing outside (0,1) have been used in literature by several authors in applications, see, e.g., Johns and Van Ryzin (1972), Singh (1977, 1978, 1979) and Karunamuni and Mehra (1990). Such kernels can be easily constructed; see Singh (1978).

Remark 2.2. The estimator (1.1) has been analysed by a number of authors —by Parzen (1962), Bartlett (1963), Rosenblatt (1971), Davis (1975), Müller and Gasser (1979) and Müller (1984) with symmetric kernels and by Johns and Van Ryzin (1972), Singh (1977) with one-sided kernels—for "optimal" convergence rates of its bias and MSE at a point, but under the assumption of existence of derivatives of f at the point. For example, Parzen (1962) proved, with a symmetric kernel under certain conditions on K, that the bias and MSE of the estimator (1.1) are of orders $O(h^r)$ and $O(nh)^{-1}+O(h^{2r})$, respectively, provided f has derivatives of first r orders and the kernel function K used is of order r. In Theorems 2.1 and 2.2, we have shown that these convergence rates can be achieved with the estimator (2.3) without any differentiability assumptions. Instead some weaker regularity conditions on f, namely, (2.5) to (2.8) and (2.5a) to (2.8a) are imposed using notions of finite differences.

In a broad general sense, conditions (2.5) and (2.7) require that certain lower left and right "Dini" type derivatives of f stay "uniformly" bounded in small neighbourhoods of the point x. In other words, under these conditions the results of Theorems 2.1 and 2.2 would hold even if the upper Dini derivatives (left and right) of f at x are infinite. Now, obviously, the relaxation of (r-th order) differentiability conditions to those in (2.5) to (2.8) would enlarge substantially the class of functions for which the above rate results hold. To see this more clearly, the following implication of Theorem 7.8 of Hewitt and Stromberg ((1965), p. 260) is worth noting: Let $D^+f(x)$ and $D_+f(x)$ denote the upper and lower right Dini derivatives, respectively, of f at the point x; then according to this theorem, the space

$$\begin{aligned}&\mathcal{D}^+(-\infty,\infty)\\ &\quad=\{f\in\mathcal{C}(-\infty,\infty): D^+f(x) \text{ and } D_+f(x) \text{ are both finite for some } x\}\end{aligned}$$

(containing the class of differentiable functions in $\mathcal{C}(-\infty,\infty)$) is of first category and consequently, in view of the Baire category theorem, the class $\mathcal{D}^{+c}(-\infty,\infty)=\mathcal{C}(-\infty,\infty)-D^+(-\infty,\infty)$, or more generally the class

$$D^{++}(-\infty,\infty)=\{f\in\mathcal{C}(-\infty,\infty): D^+f(x)=\infty \text{ for some } x\}$$

containing it, is dense in $\mathcal{C}(-\infty,\infty)$ = {the space of continuous functions on $(-\infty,\infty)$ which vanish at $\pm\infty$}. Similar considerations also hold for the analogous classes $\mathcal{D}^-(-\infty,\infty)$ and $D^{--}(-\infty,\infty)$. Now note that our conditions (2.5) and (2.7), which require "uniform" boundedness for the lower Dini derivatives $D_+f(x)$ and $D_-f(x)$ in some neighbourhood of x in question, do allow the upper Dini derivatives $D^+f(x)$ and $D^-f(x)$ to be infinite. Thus the class $\mathcal{G}$ of densities satisfying (2.5) to (2.8) extends into the larger (dense in $\mathcal{C}(-\infty,\infty)$) class $\{D^{++}(-\infty,\infty)\cup\mathcal{D}^{--}(-\infty,\infty)\}$, unlike the class of differentiable densities, which is confined to the nowhere dense (in $\mathcal{C}(-\infty,\infty)$) class $\{D^+(-\infty,\infty)\cup\mathcal{D}^-(-\infty,\infty)\}$. It is to the elements of this preceding larger subclass $\mathcal{G}$ of $\{D^{++}(-\infty,\infty)\cup\mathcal{D}^{--}(-\infty,\infty)\}$ that our results address profitably. However, the analytical question whether the class $\mathcal{G}$ is dense within the class of all continuous density functions is a deeper one and would not be pursued further here.

Thus the achieved improved rates of convergence for the bias and MSE of estimators of the type (2.3) continue to hold under regularity conditions that are quite a bit milder relative to those requiring r-th order differentiability. These regularity conditions reduce to simple restrictions on $f^{(1)}$ and $f^{(r)}$ when f admits derivatives up to first r orders. The following corollary is an immediate consequence of Theorems 2.1 and 2.2 above, and the boundedness condition on K_0.

COROLLARY 2.1. *Let the hypotheses be of Theorem* 2.1 *and suppose that* $|K_0| < M_1$. *Then for each* x, $-\infty < x < \infty$, *we obtain*

(i)
$$E(\hat{f}(x) - f(x))^2 \leq h^{2r} a_3^2 \left(\frac{C_r^+(x) + C_r^-(x)}{2r!} \right)^2 + (2nh^2)^{-1} M_1^2 \int_{x-h}^{x+h} f(t)dt,$$

(ii) *if* $|f| \leq \Lambda_0$, *then*

$$E(\hat{f}(x) - f(x))^2 \leq h^{2r} a_3^2 \left(\frac{C_r^+(x) + C_r^-(x)}{2r!} \right)^2 + (nh)^{-1} M_1^2 \Lambda_0,$$

(iii) *if for a* $p > 1$, $\int (f(t))^p dt < \infty$, *then*

$$E(\hat{f}(x) - f(x))^2 \leq h^{2r} a_3^2 \left(\frac{C_r^+(x) + C_r^-(x)}{2r!} \right)^2 + (2nh^{(p+1)/p})^{-1} M_1^2 \left(\int (f(t))^p dt \right)^{1/p}.$$

Remark 2.3. Suppose now that f has derivatives up to r-th order at the point x. Then conditions (2.5) to (2.8) and (2.5a) to (2.8a) reduce to, say, $(2.5)^*$ to $(2.8)^*$ and $(2.5a)^*$ to $(2.8a)^*$, where the starred conditions are simply the preceding unstarred ones with $\Delta_\eta^r f(x+\tau)$, $\Delta_\eta f(x+\tau)$ replaced, respectively, by $f^{(r)}(\xi)$ and $f^{(1)}(\xi)$, with suitable ξ's lying between $x+\tau$ and $x+\tau+r\eta$. This follows since

$$\Delta_\eta^k f(x+\tau) = \eta^{-k} \sum_{i=0}^{k} \binom{k}{i} (-1)^{k-i} f(x+\tau+i\eta) = f^{(k)}(\xi), \quad \text{for} \quad k = 1, r, \tag{2.11}$$

where ξ is as defined above; see Hardy ((1955), p. 333, Example 5). Accordingly under the above differentiability assumptions, theorems, (say) Theorems 2.1^* and 2.2^*, obtained from Theorems 2.1 and 2.2, respectively, by replacing $\Delta_\eta^k f(x+\tau)$ with $f^{(k)}(\xi)$, $k = 1, r$, continue to hold.

COROLLARY 2.2. *Let the hypotheses be those of Theorems* 2.1^* *and* 2.2^* *and assume additionally that* $f^{(1)}$ *and* $f^{(r)}$ *are bounded. Then* (2.9) *and* (2.10) *hold uniformly over* x, *i.e.*

$$\sup_x |E(\hat{f}(x)) - f(x)| \leq \frac{h^r}{r!} \Lambda_r a_3 \tag{2.12}$$

and

$$\sup_x E[\hat{f}(x) - f(x)]^2 \leq h^{2r} (r!)^{-2} \Lambda_r^2 a_3^2 + (nh)^{-1} \{\Lambda_0 a_1 + h\Lambda_1 a_2\}, \tag{2.13}$$

where $\sup |f^{(k)}(x)| \leq \Lambda_k$, $k = 0, 1, r$ (*with* $f^{(0)} = f$ *and* Λ_k*'s not necessarily independent of* f) *and* a_1, a_2 *and* a_3 *are as defined in Theorem* 2.2.

We now state a Farrell-Wahba type ((1972) and (1975), respectively) result establishing that the preceding convergence rates for the bias and MSE of kernel estimators of the type (2.3) are uniform over a large class of density functions, specifically, the class $\mathcal{T}_{L_0,L_1,L_r}$ which is defined below:

THEOREM 2.3. *Let $\mathcal{T}_{L_0,L_1,L_r}$ be the class of density functions f on $(-\infty, \infty)$, satisfying* (a) $\sup f(x) \le L_0$, (b) $\sup|\Delta_\eta^r f(x)| \le L_r$ *and* (c) $|f(x) - f(y)| \le L_1|x-y|$, *where L_0, L_1 and L_r do not depend on f or η. Then for $\hat{f}(x)$, defined by* (2.3) *and based on a r-th order kernel K_0, measurable and vanishing off* $(0,1)$, *we have for $r \ge 1$*

$$\sup_x \sup_{f\in\mathcal{T}_{L_0,L_1,L_r}} |E\hat{f}(x) - f(x)| \le \frac{h^r}{r!} L_r \left[\int_0^1 |t^r K_0(t)|dt\right] \tag{2.14}$$

and

$$\begin{aligned} \sup_x \sup_{f\in\mathcal{T}_{L_0,L_1,L_r}} & E[\hat{f}(x) - f(x)]^2 \\ & \le h^{2r}\left(\frac{L_r}{r!}\right)^2 \left[\int_0^1 t^r|K_0(t)|dt\right]^2 \\ & \quad + (nh)^{-1}\left[L_0\int_0^1 K_0^2(t)dt + hL_1\int_0^1 K_0^2(t)tdt\right]. \end{aligned} \tag{2.15}$$

If $r = 1$, then the conditions (a) *and* (c) *are enough to obtain* (2.14) *and* (2.15). *Further if $\mathcal{T}^*_{L_0,L_1,L_r}$ denote the class obtained from $\mathcal{T}^*_{L_0,L_1,L_r}$ by simply replacing $\Delta_\eta^r f(x)$ with $f^{(r)}(x)$ (which is assumed to exist at all points x) in the condition* (b), *then the inequalities* (2.14) *and* (2.15), *with $\mathcal{T}$ replaced by $\mathcal{T}^*$ therein (denote these inequalities by* (2.14)* *and* (2.15)*, *respectively) continue to hold.*

Remark 2.4. Rates of convergence to zero for the bias and MSE of estimator (1.1) obtained by Farrell (1972) and Wahba (1975) are uniform, respectively, over Farrell's class $C_{r-1,\phi}$ and the Sobolev (sub)space $W_\nu^{(r)}(M) = \{f : f \in W_\nu^{(r)}, \|f^{(r)}\|_{L_\nu} = \int |f^{(r)}(x)|^p dx \le M\}$ of density functions, where r is the order of a symmetric kernel K, M is a constant and ϕ is a differentiable function defined over $(-\infty, \infty)$. For $\nu \ge 1$, the Sobolev space is the space $W_\nu^{(r)}$ of functions whose first $(r-1)$ derivatives are absolutely continuous and whose p-th power of r-th derivative is integrable. Farrell's class $C_{r,\phi}$ is the class of densities f whose r derivatives are continuous and such that there exists a polynomial s of degree r with $|f(y) - s(y)| \le 2(r!)^{-1} y^r \phi^{(r)}(y)$. Note that $f \in W_\nu^{(r)}(M)$ implies $f \in C_{r-1,\phi}$. The rate obtained in Wahba ((1975), see Theorem 4.1) for the MSE, uniform over the space $W_\nu^{(r)}(M)$, is $n^{-\psi(r,\nu)}$, where $\psi(r, \nu) = [(2r\nu-2)/(2r\nu+\nu-2)]$, $\nu \ge 1$. A similar result is obtained in Theorem 2.3 above with rate $n^{-2r/(2r+1)}$ (taking $h = n^{-1/(2r+1)}$) uniform over the space $\mathcal{T}^*_{L_0,L_1,L_r}$. Observe that the rate $n^{-2r/(2r+1)}$ is slightly better than $n^{-\psi(r,\nu)}$. It can be shown that if $f \in W_\nu^{(r)}(M)$, then f essentially satisfies the conditions (a) and (c) of Theorem 2.3 (see Wahba (1975), p. 16) and, the condition (b) takes the place of the condition $\int |f^{(r)}(x)|^p dx \le M$

for all $f \in W_\nu^{(r)}(M)$. Observe that if Wahba had used a Taylor expansion with the Lagrangian remainder, then in her paper (see the equation (4.3) in Wahba (1975)) she would have (presumably) used the condition (b) as well.

We should point out also that there is no difficulty in proving Theorem 2.3 in a form and under conditions analogous to those of Wahba (1975). In fact, in the proof of Theorem 2.1 (see Section 3 of this paper) if we employ the integral form of the remainder term in the Taylor expansion of (3.13) below, we obtain

$$\begin{aligned}\gamma(y) &= \sum_{j=1}^{r-1} \frac{y^j}{j!}\gamma^{(j)}(0) + \int_0^y \frac{(y-u)^{r-1}}{(r-1)!}\gamma^{(r)}(u)du \\ &= \sum_{j=1}^{r-1} \frac{y^j}{j!}\gamma^{(j)}(0) + \int_0^y \frac{(y-u)^{r-1}}{(r-1)!}\sum_{i=0}^{\infty}\frac{u^i}{i!}\Delta_\eta^{i+r} f(x)du,\end{aligned}$$

so that (3.14) below would now read as

$$\begin{aligned}|E\hat{f}(x) - f(x)| \le &\left| \int_0^1 K_0(t)\left\{ \int_0^{ht} \frac{(ht-u)^{r-1}}{(r-1)!} \lim_{\eta\downarrow 0} \sum_{i=0}^{\infty} \frac{u^i}{i!}\Delta_\eta^{i+r} f(x)du \right\} dt \right| \\ &+ \left| \int_0^1 K_0(t)\left\{ \int_{-ht}^{0} \frac{(ht-u)^{r-1}}{(r-1)!} \lim_{\eta\uparrow 0} \sum_{i=0}^{\infty} \frac{u^i}{i!}\Delta_\eta^{i+r} f(x)du \right\} dt \right|.\end{aligned}$$

Using precisely the same reasoning as in the proof of Theorem 2.1 for (3.15) to (3.21), the proof can be accomplished if we impose suitable Wahba type conditions; but we shall leave these details to the reader.

In view of (2.15), the optimal choice of K_0, no matter what x is, seems clear (cf. Wahba (1975) and Müller and Gasser (1979)). Specifically from the standpoint of minimizing the MSE, one would choose K_0, subject to (1.2), that minimizes $\int_0^1 K_0^2(t)dt$ while noting in this connection that for any K_0 chosen above we shall have $\int_0^1 tK_0^2(t)dt \le \int_0^1 K_0^2(t)dt$ and $\int_0^1 |K_0(t)|tdt \le \left[\int_0^1 K_0^2(t)dt\right]^{1/2}$.

3. Proofs of theorems

First we state a theorem from the theory of semi-groups of linear operators which plays a major role in the construction of proofs below. Let $\mathcal{X}$ be any real or complex Banach space and $\tilde{\mathcal{X}}$ be the Banach algebra of bounded linear operators on $\mathcal{X}$ to itself. Let $\mathcal{N} = \{T(t) : t \ge 0\}$ denote a one-parameter semi-group of bounded linear operators on $\mathcal{X}$ to itself, i.e. with the property that

$$T(t+s)f = T(t)[T(s)f], \quad T(0) = I \tag{3.1}$$

for all non-negative s and t and all $f \in \mathcal{X}$, where I denotes the identity operator. The semi-group is called strongly continuous if in addition, for each $f \in \mathcal{X}$, the map $t \to T(t)f$ is continuous in t on $[0, \infty]$. Define now the operator T_η and an exponential operator $\exp(tT_\eta)$ in $\tilde{\mathcal{X}}$ by

$$T_\eta = \eta^{-1}(T(\eta) - I) \quad \text{and} \quad \exp(tT_\eta) = \sum_{k=0}^{\infty} \frac{t^k}{k!} T_\eta^k, \tag{3.2}$$

where η ($\eta \neq 0$) and t are real, and T_η^k is obtained by applying T_η recursively k times. We can state the following theorem due to Hille (1944), which is also known as the "exponential formula" for semi-groups of linear operators. See Butzer and Berens (1967) for a systematic treatment of semi-groups operators and their relationship to approximation problems.

THEOREM 3.1. (Hille (1944)) *Let $\{T(t) : t \geq 0\}$ be a strongly continuous semi-group of bounded linear operators and let T_η and e^{tT_η} be defined by* (3.2) *with $\eta > 0$. Then, for each $f \in \mathcal{X}$,*

$$T(t)f = \lim_{\eta \downarrow 0} e^{tT_\eta} f = \lim_{\eta \downarrow 0} \sum_{k=0}^{\infty} \frac{t^k}{k!} T_\eta^k f \tag{3.3}$$

uniformly in t over any finite interval.

It should be pointed out here that the semi-group of operators can also be defined with negative parameter $t \leq 0$ and further that Theorem 3.1 remains true when $\eta < 0$ and $t \leq 0$ hold simultaneously and the limit is taken as η increases to zero, i.e.

$$T(t)f = \lim_{\eta \uparrow 0} \sum_{k=0}^{\infty} \frac{t^k}{k!} T_\eta^k f. \tag{3.4}$$

The formulas (3.3) and (3.4) can now be restated using relations of calculus of finite differences. Define the differences $\Delta_\eta^k T(t)$, $k = 1, 2, \ldots$, by (see (2.2))

$$\Delta_\eta^k T(t) = \eta^{-k} \sum_{i=0}^{k} \binom{k}{i} (-1)^{k-i} T(t + i\eta). \tag{3.5}$$

Then it is easy to verify that $[\Delta_\eta^k T(0)]f = T_\eta^k f$, so that the formulas (3.3) and (3.4) yield, for each $f \in \mathcal{X}$

$$T(t)f = \begin{cases} \lim_{\eta \downarrow 0} \sum_{k=0}^{\infty} \frac{t^k}{k!} [\Delta_\eta^k T(0)] f & \text{for } t \geq 0,\ \eta > 0 \\ \lim_{\eta \uparrow 0} \sum_{k=0}^{\infty} \frac{t^k}{k!} [\Delta_\eta^k T(0)] f & \text{for } t \leq 0,\ \eta < 0 \end{cases} \tag{3.6}$$

uniformly in t over any finite interval.

Let us now apply (3.6) to a special class of semi-groups of interest here, namely, the semi-group of translations in $BU(-\infty, \infty)$, the space of uniformly continuous bounded functions on $(-\infty, \infty)$. (This translation semi-group is actually a group of operators, see e.g., Butzer and Berens ((1967), p. 22).) So the basic (Banach) space $\mathcal{X}$ is $BU(-\infty, \infty)$ with the supremum norm. We define the semi-groups $\{T(t) : t \geq 0\}$ and $\{T(t) : t \leq 0\}$ here by

$$[T(t)f](s) = f(t+s), \qquad -\infty < s < \infty, \quad f \in BU(-\infty, \infty), \tag{3.7}$$

where $[T(t)f](s)$ means the value of $T(t)f$ at the point s. It is easy to show that the semi-groups $\{T(t) : t \geq 0\}$ and $\{T(t) : t \leq 0\}$ defined by (3.7) are strongly continuous in view of the uniform continuity of functions in $BU(-\infty, \infty)$. Then, the formulas (3.6) assert that for each $f \in BU(-\infty, \infty)$

$$f(t+s) = \begin{cases} \lim_{\eta \downarrow 0} \sum_{k=0}^{\infty} \frac{t^k}{k!} \Delta_\eta^k f(s), & \text{for} \quad t \geq 0, \ \eta > 0 \\ \lim_{\eta \uparrow 0} \sum_{k=0}^{\infty} \frac{t^k}{k!} \Delta_\eta^k f(s), & \text{for} \quad t \leq 0, \ \eta < 0, \end{cases} \tag{3.8}$$

the limits existing uniformly with respect to s in $(-\infty, \infty)$ and uniformly with respect to t in every finite interval, where $\Delta_\eta^k f(s)$ is obtained by the definition (2.2). It is clear that (3.8) gives a generalization of Taylor's formula for an f which is merely uniformly continuous. Note that if f has a derivative of order k, then $\lim_{\eta \downarrow 0} \Delta_\eta^k f(s) = f^{(k)}(s) = \lim_{\eta \uparrow 0} \Delta_\eta^k f(s)$.

We now give the proofs of theorems:

Proof of Theorem 2.1. From (2.3) it follows immediately that

$$E\hat{f}(x) = \frac{1}{2} \int_0^1 K_0(t)\{f(x+ht) + f(x-ht)\}dt. \tag{3.9}$$

Since f is in the class $BU(-\infty, \infty)$, an application of (3.8) to $f(x+ht)$ and $f(x-ht)$, separately to the two terms with $ht \geq 0$ and $-ht \leq 0$, respectively, yields that

$$\begin{aligned} f(x+ht) &= f(x) + \lim_{\eta \downarrow 0} \sum_{i=1}^{\infty} \frac{(ht)^i}{i!} \Delta_\eta^i f(x), \\ f(x-ht) &= f(x) + \lim_{\eta \uparrow 0} \sum_{i=1}^{\infty} \frac{(-ht)^i}{i!} \Delta_\eta^i f(x), \end{aligned} \tag{3.10}$$

where for fixed x, the preceding convergences are uniform in ht, $0 < t < 1$. Using this fact, which permits interchange of limit and integration, and the orthogonality properties of K_0 (see (1.2)), we obtain from (3.9) and (3.10) that for each x,

$-\infty < x < \infty$,

$$(3.11)\qquad E\hat{f}(x) - f(x) = \frac{1}{2}\left[\lim_{\eta\downarrow 0}\int_0^1 K_0(t)\left\{\sum_{i=r}^{\infty}\frac{(ht)^i}{i!}\Delta_\eta^i f(x)\right\}dt\right] + \frac{1}{2}\left[\lim_{\eta\uparrow 0}\int_0^1 K_0(t)\left\{\sum_{i=r}^{\infty}\frac{(-ht)^i}{i!}\Delta_\eta^i f(x)\right\}dt\right].$$

For fixed r, x and η, define the function γ by

$$(3.12)\qquad \gamma(y) = \sum_{i=r}^{\infty}\frac{y^i}{i!}\Delta_\eta^i f(x), \qquad 0 \le y \le 1;$$

then $\gamma(y)$ is analytic in [0, 1]. Taylor's expansion of $\gamma(y)$ in y at 0, with Lagrange's form of the remainder at the r-th term, gives

$$(3.13)\qquad \gamma(y) = \sum_{j=0}^{r-1}\frac{y^j}{j!}\gamma^{(j)}(0) + \frac{y^r}{r!}\gamma^{(r)}(\xi),$$

where $0 < \xi < y$ and $\gamma^{(j)}$ denotes the j-th derivative of γ. Now applying formula (3.13) to the two terms on the RHS of (3.11) with $y = ht$, $0 < t < 1$, it follows in view of the orthogonality properties of K_0 that

$$(3.14)\qquad E\hat{f}(x) - f(x) = \frac{h^r}{r!}\frac{1}{2}\left\{\lim_{\eta\downarrow 0}\int_0^1 K_0(t)t^r\beta(\eta,\, \xi,\, x)dt + \lim_{\eta\uparrow 0}\int_0^1 K_0(t)(-t)^r\beta(\eta,\, -\xi,\, x)dt\right\}$$

where

$$(3.15)\qquad \beta(\eta,\, \xi,\, x) = \sum_{i=0}^{\infty}\frac{\xi^i}{i!}\Delta_\eta^{i+r}f(x).$$

Now expanding $\Delta_\eta^{i+r}f(x)$ by making use of the definition of Δ_η^i (see (2.2)), we get

$$(3.16)\qquad \beta(\eta,\, \xi,\, x) = \sum_{i=0}^{\infty}\frac{\xi^i}{i!}\eta^{-i}\sum_{k=0}^{i}\binom{i}{k}(-1)^{i-k}\Delta_\eta^r f(x+k\eta) = \sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\left(\frac{\xi}{\eta}\right)^{j+k}\frac{1}{(j+k)!}\binom{j+k}{k}(-1)^j\Delta_\eta^r f(x+k\eta),$$

where the last equality is obtained by putting $j+k=i$ and making j and k as new summation indexes. Now rearranging the terms in RHS of (3.16) one can easily see (since the series is absolutely convergent) that

$$(3.17)\qquad \beta(\eta,\, \xi,\, x) = e^{-\xi/\eta}\sum_{k=0}^{\infty}\left(\frac{\xi}{\eta}\right)^k\frac{1}{k!}\Delta_\eta^r f(x+k\eta).$$

For fixed r, η and ξ, let τ denote a discrete random variable whose distribution function is constant except for jumps of size $(1/k!)(\xi/\eta)^k e^{-\xi/\eta}$ at the points $k\eta$, $k = 0, 1, 2, \ldots$ (which constitute the sets A_η^+ or A_η^- according as $\eta > 0$ and $\eta < 0$ as defined in the statement of the theorem). Then observe that $E_{\xi,\eta}(\tau) = \xi$ and from (3.17)

$$\beta(\eta, \xi, x) = E_{\xi,\eta}\Delta_\eta^r f(x+\tau). \tag{3.18}$$

Since $\alpha(\tau, \xi) > 0$ for $\tau \neq \xi$ (see (2.4) for the definition of $\alpha(\tau, \cdot))$ and letting $I_{S_i^+(\xi)}$ stand for the indicator function of $S_i^+(\xi)$, $i = 1, 2$, (see statement of Theorem 2.1 for the definition of $S_i^+(\xi)$), we obtain

$$\begin{aligned} |\beta(\eta, \xi, x)| &\leq E_{\xi,\eta}|\Delta_\eta^r f(x+\tau)| \\ &\leq \sup_{\tau \in S_1^+(\xi)} |\Delta_\eta^r f(x+\tau)| \\ &\quad + \left[\sup_{\tau\in S_2^+(\xi)} \frac{|\Delta_\eta^r f(x+\tau)|}{\alpha^+(\tau, \xi)}\right] E_{\xi,\eta}\alpha^+(\tau, \xi), \end{aligned} \tag{3.19a}$$

with $\alpha^+(\tau, \xi) = a^{1+\tau} - a^{1+\xi}[1 + (\tau - \xi)\log a]$ (see (2.4)). A similar inequality as in (3.19a) holds with ξ, $S_i^+(\xi)$ and $\alpha^+(\tau, \xi)$ replaced by $-\xi$, $S_i^-(-\xi)$ and $\alpha^-(\tau, -\xi)$, respectively, $i = 1, 2$. Let us call that inequality (3.19b). Observe that

$$\begin{aligned} \lim_{\eta\downarrow 0} E_{\xi,\eta}\alpha^+(\tau, \xi) &= \lim_{\eta\downarrow 0}\left[e^{-\xi/\eta}\sum_{k=0}^{\infty}\left(\frac{\xi}{\eta}\right)^k \frac{1}{k!} a^{1+k\eta} - a^{1+\xi}\right] \\ &= \lim_{\eta\downarrow 0}[ae^{\xi(a^\eta - 1)/\eta} - a^{1+\xi}] \\ &= ae^{\xi\log a} - a^{1+\xi} = 0, \end{aligned} \tag{3.20}$$

uniformly in ξ in $(0, \epsilon)$, $\epsilon > 0$. Similarly one can show that $\lim_{\eta\uparrow 0} E_{-\xi,\eta}\alpha^-(\tau, -\xi) = 0$ uniformly in $-\xi$ in $(-\epsilon, 0)$. Now taking $\sup_{0<\xi<\epsilon}$ and $\sup_{-\epsilon<-\xi<0}$ and then $\liminf_{\eta\downarrow 0}$ and $\liminf_{\eta\uparrow 0}$ on both sides of (3.19a) and (3.19b), respectively, and combining (3.14)–(3.18), (3.19a), (3.19b) and (3.20), we obtain

$$|E\hat{f}(x) - f(x)| \leq \frac{h^r}{r!}\left\{\frac{1}{2}C_r^+(x) + \frac{1}{2}C_r^-(x)\right\}\int_0^1 |K_0(t)|t^r dt, \tag{3.21}$$

provided that (2.5)–(2.8) hold, where

$$C_r^+(x) = \liminf_{\eta\downarrow 0} \sup_{0<\xi<\epsilon} \sup_{\tau\in S_1^+(\xi)} |\Delta_\eta^r f(x+\tau)| \quad \text{and}$$

$$C_r^-(x) = \liminf_{\eta\uparrow 0} \sup_{-\epsilon<-\xi<0} \sup_{\tau\in S_1^-(\xi)} |\Delta_\eta^r f(x+\tau)|,$$

as defined in Theorem 2.1. This completes the proof of Theorem 2.1. □

PROOF OF THEOREM 2.2. Since X_i's and i.i.d. with density f, using a trivial inequality we have for each x, $-\infty < x < \infty$,

$$\operatorname{Var} \hat{f}(x) \le \frac{1}{2}(nh^2)^{-1}\left\{\operatorname{Var} K_0\left(\frac{X-x}{h}\right) + \operatorname{Var} K_0\left(\frac{x-X}{h}\right)\right\}. \tag{3.22}$$

If suffices to show that

$$(nh^2)^{-1}\operatorname{Var} K_0\left(\frac{X-x}{h}\right) \le (nh)^{-1}\left[f(x)\int_0^1 K_0^2(t)dt + hC_1^+(x)\int_0^1 tK_0^2(t)dt\right],$$

since the proof of a similar domination of the other term is the same. With this demonstration, thus, the proof of the theorem would be complete. First note that

$$\begin{aligned}(nh^2)^{-1}\operatorname{Var} K_0\left(\frac{X-x}{h}\right) &\le (nh)^{-1}\int_0^1 K_0^2(t)f(x+ht)dt \\ &\le (nh)^{-1}\left\{f(x)\int_0^1 K_0^2(t)dt\right. \\ &\qquad \left. + \lim_{\eta\downarrow 0}\int_0^1 K_0^2(t)\left(\sum_{i=1}^{\infty}\frac{(ht)^i}{i!}\Delta_\eta^i f(x)\right)dt\right\},\end{aligned} \tag{3.23}$$

the last inequality is obtained using (3.10), since the convergence in there is uniform in ht for fixed x. Now using an argument similar to the one used to obtain (3.13), one can show that

$$\begin{aligned}&\lim_{\eta\downarrow 0}\int_0^1 K_0^2(t)\left(\sum_{i=1}^{\infty}\frac{(ht)^i}{i!}\Delta_\eta^i f(x)\right)dt \\ &\quad = \lim_{\eta\downarrow 0}\int_0^1 K_0^2(t)(ht)\left(\sum_{i=0}^{\infty}\frac{(\xi)^i}{i!}\Delta_\eta^{i+1} f(x)\right)dt,\end{aligned} \tag{3.24}$$

where $0 < \xi < ht$; and using an argument similar to the one used to obtain (3.17), one can show that

$$\begin{aligned}\sum_{i=0}^{\infty}\frac{\xi^i}{i!}\Delta_\eta^{i+1} f(x) &= e^{-(\xi/\eta)}\sum_{k=0}^{\infty}\left(\frac{\xi}{\eta}\right)^k\frac{1}{k!}\Delta_\eta^1 f(x+k\eta) \\ &= E_{\xi,\eta}[\Delta_\eta^1 f(x+\tau)],\end{aligned} \tag{3.25}$$

where τ is a random variable as defined in the proof of Theorem 2.1. We may now use the inequality

$$\begin{aligned}|E_{\xi,\eta}(\Delta_\eta^1 f(x+\tau))| &\le \sup_{\tau\in S_1^+(\xi)}|\Delta_\eta^1 f(x+\tau)| \\ &\quad + \left[\sup_{\tau\in S_2^+(\xi)}\frac{|\Delta_\eta^1 f(x+\tau)|}{\alpha^+(\tau,\xi)}E_{\xi,\eta}[\alpha^+(\tau,\xi)]\right],\end{aligned} \tag{3.26}$$

where $S_1^+(\xi)$, $S_2^+(\xi)$ and $\alpha^+(\tau, \xi)$ are as defined in Theorem 2.1, to obtain that

$$\text{(3.27)} \qquad |\text{LHS of (3.25)}| \leq \sup_{\tau \in S_1^+(\xi)} |\Delta_\eta f(x+\tau)| + \left[\sup_{\tau \in S_2^+(\xi)} \frac{|\Delta_\eta^1 f(x+\tau)|}{\alpha^+(\tau, \xi)} \right] E_{\xi,\eta}[\alpha^+(\tau, \xi)].$$

Again, since $\exp[(\xi/\eta)(a^\eta - 1)] \to a^{1+\xi}$, as $\eta \downarrow 0$, it follows that $\lim_{\eta\downarrow 0} E_{\xi,\eta}[\alpha^+(\tau, \xi)] = 0$ uniformly with respect to ξ in $(0, \epsilon)$. Therefore, from (3.27) one obtains (taking $\liminf_{\eta\downarrow 0}$ on both sides of (3.27))

$$\text{(3.28)} \qquad \liminf_{\eta\downarrow 0} \sup_{0<\xi<\epsilon} \left| \sum_{i=0}^{\infty} \frac{\xi^i}{i!} \Delta_\eta^{i+1} f(x) \right| \leq C_1^+(x),$$

where $C_1^+(x) = \liminf_{\eta\downarrow 0} \sup_{0<\xi<\epsilon} \sup_{\tau\in S_1^+(\xi)} |\Delta_\eta f(x+\tau)|$, provided (2.5a) and (2.6a) hold. Now combining (3.23), (3.24) and (3.28), we have for each x, $-\infty < x < \infty$, that

$$\text{(3.29)} \qquad (nh^2)^{-1} \operatorname{Var} K_0\left(\frac{X-x}{h}\right) \leq (nh)^{-1} \left[f(x) \int_0^1 K_0^2(t)dt + hC_1^+(x) \int_0^1 tK_0^2(t)dt \right],$$

so that, in view of the assertion just after (3.22), the result follows. One obtains (2.10) now by using the standard identity $E[\hat{f}(x) - f(x)]^2 = [\text{LHS of (2.9)}]^2 + V(\hat{f}(x))$. This completes the proof of Theorem 2.2. □

PROOF OF COROLLARY 2.2. If $|f^{(r)}| \leq \Lambda_r$, then (2.5)* implies that $|C_r^+(x)| \leq \Lambda_r$ for all x, and if $|f^{(1)}| \leq \Lambda_1$, (2.5a)* implies $|C_1^+(x)| \leq \Lambda_1$ for all x; similarly for $C_r^-(x)$ using (2.7)* and (2.7a)*. Since $|f(x)| \leq \Lambda_0$, the proof of the corollary would follow from (2.9) and (2.10) if we show that (2.6)*, (2.8)*, (2.6a)* and (2.8a)* are satisfied. In fact, it suffices to establish (2.6)* and (2.6a)*; (2.8)* and (2.8a)* follow similarly. Now since $f^{(r)}$ and $f^{(1)}$ are bounded, it is enough to show that $\sup_{0<\theta<\epsilon} \sup_{\tau\in S_2^+(\theta)} [\alpha^+(\tau, \theta)]^{-1} < \infty$, where $\alpha^+(\tau, \theta) = a^{1+\tau} - a^{1+\theta} - a^{1+\theta}(\tau - \theta)\log a$, $a > 1$. We will show that $\alpha^+(\tau, \theta)$ bounded away from zero on $S_2^+(\theta) \subseteq \{\tau : |\tau - \theta| \geq \delta\}$. To see this, note that using Taylor's expansion in τ around θ, we have for some real s between τ and θ,

$$\text{(3.30)} \qquad \alpha^+(\tau, \theta) = \frac{(\tau-\theta)^2}{2} a^{1+s} (\log a)^2 \geq \frac{\delta^2}{2} a (\log a)^2 (> 0) \quad \text{if} \quad |\tau - \theta| \geq \delta,$$

the last inequality in (3.30) following since τ and θ are positive on $S_2^+(\theta)$ which implies $s > 0$. Thus $(\alpha^+(\tau, \theta))^{-1} \leq (2/\delta^2 a)^{-1} (\log a)^{-2}$, so that $\sup_{0<\theta<\epsilon} \sup_{\tau\in S_2^+(\theta)} [\alpha^+(\tau, \theta)]^{-1} < \infty$. The proof is complete. □

PROOF OF THEOREM 2.3. First note that under the condition (c), $|\Delta_\eta f(x+\tau)| \le L_1$ and, consequently, $C_1^+(x)$, $C_1^-(x) \le L_1$ for all $f \in \mathcal{T}_{L_0,L_1,L_r}$ and all x. Similarly the condition (b) implies that $C_r^+(x)$, $C_r^-(x) \le L_r$ for all $f \in \mathcal{T}_{L_0,L_1,L_r}$ and all x. Also the conditions (b) and (c) imply that [LHS of (2.5)$/L_r$] and [LHS of (2.5a)$/L_1$] do not exceed $\sup_{0<\theta<\epsilon} \sup_{\tau\in S_2^+(\theta)} [\alpha^+(\tau, \theta)]^{-1}$, which is shown to be finite in the proof of Corollary 2.2. Thus all conditions (2.5) to (2.8) and (2.5a) to (2.8a) of Theorems 2.1 and 2.2 are satisfied. Accordingly, in view of the condition (a), the results (2.14) and (2.15) follow from (2.9) and (2.10). The proof of assertions (2.14)* and (2.15)* is verbatim the same. The proof is complete. □

4. Concluding remarks

The methods and arguments presented above for establishing "optimum" rates of convergence to zero of bias and MSE of kernel estimators of density functions, without any differentiability assumptions, can also be applied to kernel and other estimators of regression curves, namely, mean and quantile regression functions. These and other related results would be presented in a subsequent paper. We conclude this paper with the treatment of two examples referred to above and the extensions of the main Theorems 2.1 and 2.2 to cover the estimation of $f^{(p)}$, the p-th derivative of f, $p \ge 1$. Define for a fixed x, $-\infty < x < \infty$,

$$\hat{f}^{(p)}(x) = (nh^{p+1})^{-1} \sum_{i=1}^{n} \frac{1}{2} \left\{ K_p\left(\frac{X_i - x}{h}\right) + K_p\left(\frac{x - X_i}{h}\right) \right\}, \tag{4.1}$$

where K_p is a kernel function of order (r, p), $(r > p)$. In analogy with Theorems 2.1 and 2.2, we can easily prove the following:

THEOREM 4.1. *Let $\hat{f}^{(p)}$ be defined by* (4.1) *with a kernel function K_p, measurable, vanishing off* $(0,1)$ *and of order (r, p), $(p \ge 1)$. Let $f^{(p)}$ be bounded and uniformly continuous on $(-\infty, \infty)$. For given $\theta > 0$, $\eta > 0$ and $\delta > 0$, let S_1^+, S_2^+ and $\alpha^+(\tau, \theta)$, $\tau \in A_\eta^+$, and similarly S_1^-, S_2^- and $\alpha^-(\tau, \theta)$, $\tau \in A_\eta^-$ be as defined in Theorem* 2.1. *Then, if for given $\epsilon > 0$,* (2.5) *to* (2.8) *hold with $f^{(p)}$ in place of f, we have for each x*

$$|E\hat{f}^{(p)}(x) - f^{(p)}(x)| \le \frac{h^{r-p}}{2r!}[C_r^+(x) + C_r^-(x)] \int_0^1 |K_p(t)|t^r dt, \tag{4.2}$$

where $C_r^+(x)$ and $C_r^-(x)$ are as defined in Theorem 2.1 *with $f^{(p)}$ in place of f. Further, if* (2.5a) *to* (2.8a) *and* (4.2) *hold, then*

$$\begin{aligned} &E[\hat{f}^{(p)}(x) - f^{(p)}(x)]^2 \\ &\quad \le h^{2(r-p)} \frac{[C_r^+(x) + C_r^-(x)]^2}{4(r!)^2} b_3^2 \\ &\quad + (nh^{2p+1})^{-1} \left[f^{(p)}(x) b_1 + \frac{1}{2} h [C_1^+(x) + C_1^-(x)] b_2 \right], \end{aligned} \tag{4.3}$$

where $b_1 = \int_0^1 K_p^2(t)dt$, $b_2 = \int_0^1 t^2 K_p^2(t)dt$, $b_3 = \int_0^1 |K_p(t)|t^r dt$ *and* $C_1^+(x)$, $C_1^-(x)$ *are as defined in Theorem* 2.2.

Results analogous to Corollaries 2.1 and 2.2 and Theorem 2.3 can be proved for the estimator (4.1) as well.

We shall now treat the two elementary examples, namely, 1.1 and 1.4 in order to illustrate the nature of the two main conditions (2.5) to (2.8) imposed in Theorems 2.1 and 2.2. The implications of the results of this paper, however, go much further.

Example 1.1. (continued) Consider the uniformly continuous density given by Example 1.1. Let us verify whether the conditions (2.5) to (2.8) and (2.5a) to (2.8a) are satisfied at this point. First note that $\Delta_\eta f(0+\tau) = 1$ and for $r \geq 2$, $\Delta_\eta^r f(0+\tau) = 0$ identically for all $\tau \in S_1^+ U S_2^+$ and positive θ, η and δ, and also for $\tau \in S_1^- U S_2^-$ and any negative η and θ, and positive δ. This means that for $r = 1$, (2.5a), (2.7a) hold and that (2.6a), (2.8a) will also hold if we show that $\sup_{0<\theta<\epsilon} \sup_{\tau \in S_2^+} [\alpha^+(\tau, \theta)]^{-1} < \infty$ and $\sup_{-\epsilon<\theta<0} \sup_{\tau \in S_2^-} [\alpha^-(\tau, \theta)]^{-1} < \infty$; but this was established while proving Corollary 2.2 above. For $r \geq 2$, (2.5) to (2.8) are trivially satisfied since $\Delta_\eta^r f(0+\tau) = 0$ identically. We considered only the point $x = 0$, but it can easily be verified that the required conditions (2.5a) to (2.8a) as well as (2.5) to (2.8) hold for any $x \neq 0$ as well. Note that f is differentiable at $x \neq 0$, and thus, $\Delta_\eta^r f(x+\tau) = f^{(r)}(\xi)$ for some ξ lying between $x+\tau$ and $x+\tau+r\eta$ (see (2.11)). But, for $x \neq 0$, $f^{(r)}(\xi) = 1$ if $r = 1$ and $f^{(r)}(\xi) = 0$ if $r \geq 2$. Hence, for $x \neq 0$, $C_1^+(x) = C_1^-(x) = 1$ and $C_1^+(x) = C_1^-(x) = 0$ if $r \geq 2$, as well. Accordingly, the conclusions of Theorems 2.1, and 2.2 etc. hold for all points x.

Example 1.4. (continued) Let us consider Example 1.4 with $\theta = 0$ and $a = 1$, and consider the point $x = 0$ at which $f(x)$ is not differentiable. Now f is uniformly continuous, and for $r \geq 1$, and $\tau \in S_1^+ U S_2^+$ (see (2.2)),

$$\begin{aligned}
|\Delta_\eta^r f(0+\tau)| &= \eta^{-r} \sum_{k=0}^{r} \binom{r}{k} (-1)^{r-k} \frac{1}{2} e^{-k\eta} \\
&= \frac{1}{2} \eta^{-r} e^{-r\eta} \sum_{k=0}^{r} \binom{r}{r-k} (-1)^{r-k} e^{(r-k)\eta} \\
&= \frac{1}{2} \eta^{-r} e^{-r\eta} \sum_{k'=0}^{r} \binom{r}{k'} (-1)^{k'} e^{k'\eta} \\
&= \frac{1}{2} \eta^{-r} e^{-r\eta} (e^\eta - 1)^r \\
&= \frac{1}{2} e^{-r\eta} \left(\frac{e^\eta - 1}{\eta} \right)^r \leq \frac{1}{2} e^{-r\eta} e^{r\eta} = \frac{1}{2}.
\end{aligned}$$

Therefore (2.5) and (2.5a) hold, and further (2.6) and (2.6a) would also hold if again $\sup_{0<\theta<\epsilon} \sup_{\tau \in S_2^+} [\alpha(\tau, \theta)]^{-1} < \infty$ holds; but this is so as shown in the proof of Corollary 2.2. The same argument applies with η negative also; accordingly it follows similarly that (2.7), (2.7a), (2.8) and (2.8a) also hold. Thus the conclusions of Theorems 2.1 and 2.2 hold at $x = 0$ for $r \geq 1$, and evidently, at all other points as well.

Acknowledgements

The authors wish to thank the referees for a number of useful comments and suggestions. The authors are also grateful to Professor T. Matsunawa.

References

Bartlett, M. S. (1963). Statistical estimation of density functions, *Sankhyā Ser. A*, **25**, 245–254.

Butzer, P. L. and Berens, H. (1967). *Semi-Groups of Operators and Approximation*, Springer, New York.

Davis, K. B. (1975). Mean square error properties of density estimates, *Ann. Statist.*, **3**, 1025–1030.

Devroye, L. and Györfi, L. (1985). *Nonparametric Density Estimation: L_1-view*, Wiley, New York.

Farrell, R. H. (1972). On best obtainable asymptotic rates of convergence in estimation of a density function at a point, *Ann. Math. Statist.*, **43**, 170–180.

Hardy, G. H. (1955). *A Course of Pure Mathematics*, The University Press, Cambridge.

Hewitt, E. and Stromberg, K. (1965). *Real and Abstract Analysis*, Springer, New York.

Hille, E. (1944). *Functional Analysis and Semi-groups*, American Mathematical Society Publication, New York.

Ibiragimov, I. A. and Has'minski, R. Z. (1981). *Statistical Estimation*, Springer, New York.

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.

Karunamuni, R. J. and Mehra, K. L. (1990). Improvements on strong uniform consistency of some known kernel estimates of a density and its derivatives, *Statist. Probab. Lett.*, **9**, 133–140.

Lehmann, E. L. (1983). *Theory of Point Estimation*, Wiley, New York.

Loh, W.-Y. (1984). Bounds on ARE's for restricted classes of distributions defined via tail-ordering, *Ann. Statist.*, **12**, 685–701.

Müller, H. G. (1984). Smooth kernel estimators of densities, regression curves and modes, *Ann. Statist.*, **12**, 766–774.

Müller, H. G. and Gasser, T. (1979). Optimal convergence properties of kernel estimates of derivatives of densities, *Smoothing Techniques for Curve Estimation (Proceedings)* (eds. T. Gasser and M. Rosenblatt), 144–154, Springer, New York.

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

Rosenblatt, M. (1971). Curve estimates, *Ann. Math. Statist.*, **42**, 1815–1842.

Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*, Chapman and Hall, New York.

Singh, R. S. (1974). Estimation of derivatives of average of densities and sequence-compound estimation in exponential families, Research Memo. RM-18, Dept. of Statistics and Probability, Michigan State University.

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

Singh, R. S. (1978). Nonparametric estimation of derivatives of average of μ-densities with covergence rates and applications, *SIAM J. Appl. Math.*, **35**, 637–649.

Singh, R. S. (1979). Empirical Bayes estimation in Lebesgue-exponential families with rates near the best possible rate, *Ann. Statist.*, **7**, 890–902.

Tapia, R. A. and Thompson, J. R. (1978). *Nonparametric Probability Density Estimation*, Johns Hopkins University Press, Baltimore, Maryland.

Wahba, G. (1975). Optimal convergence properties of variable knot, kernel and orthogonal-series methods for density estimation, *Ann. Statist.*, **3**, 15–29.

Yamato, H. (1972). Some statistical properties of estimators of density and distribution functions, *Bull. Math. Statist.*, **19**, 113–131.