PROBABILITY
AND
MATHEMATICAL STATISTICS
Vol. 23, Fasc. 1 (2003), pp. 189–207

THE USE OF VARIABLE KERNEL MASS IN DENSITY ESTIMATION

BY

MICHAEL L. STURGEON (PEABODY, MASSACHUSETTS)

Abstract. A new theoretical approach is developed for estimating a univariate density $f(x)$, for large n, using a nonnegative symmetric kernel with variable mass. Compared to kernels of order 4, kernels of variable mass asymptotically achieve (i) smaller variance, (ii) essentially the same bias, and so (iii) a reduced MISE of order $O(n^{-8/9})$. The analysis uses a common MISE-optimal bandwidth h and locally adapted kernel mass $M(X) = 1-(f''(X)h^2)/(24\bar{f}(X))$, to be estimated at the kernel center, where $\bar{f}(X)$ is the average f value over an interval of length h centered on X. Mass adaptation derives from considering the expected effect of negative mass, in kernels of order 4, upon the positive part of such kernels. Unlike the Abramson procedure for varying local bandwidth, this procedure does not require any special accommodation for small values of $f(X)$, for f in C^4.

AMS Subject Classification: 62G07, 60E99.

Key words and phrases: Kernel mass adaptation, higher order kernels.

1. INTRODUCTION

The function $f(x)$ is a univariate density in C^4, with bounded support, that is estimated by $\hat{f}(x) = (1/n)\sum K(x, X_i)$, using a kernel K and a set of independent identically distributed observations $X_1, \ldots, X_n$. For an observation X, the kernel $K(x, X)$ is a measurable function of x supported on the interval $[X-h/2, X+h/2]$, denoted by $h[X]$, with a bandwidth h and a symmetric shape that are invariant with X.

Estimator bias $B(x)$ and variance $\mathrm{VAR}(x)$ are written as

$$(1.1)\qquad B(x) = E_n\{\hat{f}(x)-f(x)\}, \qquad \mathrm{VAR}(x) = E_n\{(\hat{f}(x))^2\}-E_n^2\{\hat{f}(x)\},$$

where $E_n\{\cdot\}$ denotes expectation with respect to the joint observation density f^n on R^n. The mean integral squared error, MISE, is

$$(1.2)\qquad \mathrm{MISE} = \int_R \mathrm{MSE}(x)\,dx; \qquad \mathrm{MSE}(x) = B^2(x)+\mathrm{VAR}(x).$$

For a symmetric kernel $K(x, X)$, the order $p(K)$ of the kernel, by definition, is the first even integer $p \geqslant 2$ such that the integral $\int_{X-h/2}^{X+h/2}(x-X)^p K(x, X)\,dx$ is nonzero. For $p(K) \geqslant 4$, the kernel K has negative values. For large n, it is well known that kernels of order p achieve MISE $= O(n^{-2p/(2p+1)})$ (cf. Parzen [6]) and bias $B(x) = O(h^p)$ (cf. Berlinet [1]). Thus, for kernels of order $p = 4$ which are symmetric with some negative values, MISE $= O(n^{-8/9})$, while for kernels of order $p = 2$ which are symmetric with $K \geqslant 0$, MISE $= O(n^{-4/5})$. Despite their negative values, kernels of higher order, with $p(K) \geqslant 4$, are viable, since Gajek in [3] provides a truncation algorithm for eliminating estimator negative values while further reducing MISE.

But, for large n, it is also known that higher order kernels achieve reduced bias at the expense of increased variance. As will be illustrated in Section 2, this increased variance is directly due to the negative mass in such kernels. Also, negative kernel values would seem unnatural or at least superficial. So, the attitude taken here is that negative kernel mass is a problematic surrogate for some other kernel modification(s), to be determined.

It is incumbent, then, to address two questions regarding higher order kernels:

(1) How do such kernels mechanically function to reduce bias?

(2) Is it possible to attain the reduced bias associated with higher order kernels without incurring increased variance, especially by eliminating negative kernel values and making appropriate compensatory kernel modifications?

A partial answer to query (1) is provided by Sturgeon [8], who shows that the negative mass in kernels of order $p = 4$ induces a systematic asymmetry in the positive parts of such kernels, shifting kernel mass across the observation X in the direction of increasing $|f''(x)|$. This has the effect of reducing bias everywhere that f is curvilinear. In particular, expected estimator mass is thereby shifted from regions where $f''(x) > 0$, i.e., where f is concave up, to regions where $f''(x) < 0$.

Regarding query (2), methods have been proposed in efforts to yield the MISE of kernels of order $p = 4$, without resorting to negative kernel values. These methods involve local bandwidth variation for kernels of order $p = 2$.

The Abramson procedure $h(X_i) = h\cdot\big(f(X_i)\big)^{-0.5}$ can yield $B(x) = O(h^4)$ (Silverman [7], p. 104) and MISE $= O(n^{-8/9})$ (Jones [4]) like kernels of order $p = 4$. Terrell and Scott [10] show, for a normal density and large n, that the Abramson procedure is subject to a "nonlocality" phenomenon that degrades bias to $O\big((h/\log h)^2\big)$, because the varied bandwidth becomes too large at observations where the normal density is small. McKay [5] uses a "smooth clipping" process to bound $\big(f(X_i)\big)^{-0.5}$ and shows that the Abramson procedure then attains $O(h^4)$ bias uniformly on any set where the estimated density can be bounded away from zero.

Fan et al. in [2] use bandwidth variation with $h(x) = h\cdot g(x)$, requiring a pilot estimate of the function $g(x) = \big[f(x)/\big(4\,(f''(x))^2\big)\big]^{1/5}$. However, in [4], analysis demonstrates that the Abramson $h(X)$ method is superior to the Fan $h(x)$ method, in terms of asymptotic bias and MISE. It is notable, too, that simulations of density estimation in [2] using $h(x)$ tend to show $\hat{f}$ underestimating f where $f'' < 0$ and overestimating f where $f'' > 0$. Such behavior has been shown (cf. [8]) to be typical, in the expected sense, for kernels of order $p = 2$.

Here, a new kernel variation technique is developed that theoretically attains $B(x) = O(h^4)$ and $\text{MISE} = O(n^{-8/9})$, using an X-invariant bandwidth and kernels $K \geqslant 0$. The method analytically simulates the bias reducing behavior of kernels of order $p = 4$ by varying kernel mass. It virtually eliminates the variance increment associated with kernels of order $p = 4$. Unlike the Abramson procedure, this new mass varying procedure does not require any "clipping", since the need for kernel adaptation fades as $f''(x) \to 0$.

In keeping with query (2), this new kernel variation method is obtained by analyzing the impact of negative kernel mass on estimator shape and mass. In particular, for observations Y_j sufficiently close to an observation X, the effect of the negative kernel values in $K(x, Y_j)$ upon the positive portion of $K(x, X)$ will be determined when forming the estimator $\hat{f}$ with kernels of order $p = 4$. So, the kernel adaptations described here are induced by negative kernel mass.

There are actually two such kernel adaptations. One, mentioned above, involves asymmetry. The other, reported here, involves kernel mass variation. This mass variation produces a kernel $K \geqslant 0$ whose mass theoretically deviates from 1 at the observation X by an amount equal to $-f''(X)h^2/(24\bar{f}(X))$, where $\bar{f}(X)$ is the average f value over an interval of length h centered on X. For such an adapted kernel, $\int K(x, X)\,dx < 1$ where $f''(X) > 0$, i.e., where f is concave up, and $\int K(x, X)\,dx > 1$ where $f''(X) < 0$. For the kernels studied here, it will be shown that the mass variation effect dominates the asymmetry effect in terms of reducing bias, for large n. As well, it will be shown that the mass variation $|f''(X)h^2/(24\bar{f}(X))|$ is small compared with unity for all observations X, for large n.

The kernels K rigorously dealt with here are rectangular, for analytical simplicity. Section 7 provides the basis for an extension of results to continuous analogs of K.

This appears to be the first explicit proposal to vary kernel mass to achieve reduced bias. Kernel mass variation is used, but only implicitly and along with bandwidth variation, in Terrell and Scott [9], where the estimator is nonnegative but does not integrate to 1.

2. PRELIMINARIES

It is assumed that the unknown density f is strictly curvilinear and that f is in C^4, with bounded support $S(f)$ on the real line R.

DEFINITION 2.1. For any $h > 0$ with $h = h_1 + h_2$, $h[x]$ denotes the open interval of length h centered on x. The open interval of length h_1 centered on

x is $h_1[x]$, while $h_2[x]$ is the disjoint symmetric set $h[x]-h_1[x]$ with length $h_2 = h-h_1 > 0$. Integral averages of $f(\cdot)$ over the intervals $h_1[x]$ and $h_2[x]$ are written as

$$\bar{f}_j(x) = (1/h_j) \int_{h_j[x]} f(t)\,dt, \quad j = 1, 2.$$

DEFINITION 2.2. For any sufficiently small $h > 0$, the sets

$$\mathrm{CU}(h) = \{x \in S(f)\colon\ f''(\cdot) > 0 \text{ on } h[x]\},$$
$$\mathrm{CD}(h) = \{x \in S(f)\colon\ f''(\cdot) < 0 \text{ on } h[x]\} \quad -$$

define nontrivial proper subsets of $S(f)$ on which f is concave up and concave down, respectively. Their union is denoted by

$$C_h = \mathrm{CU}(h) \cup \mathrm{CD}(h).$$

The sets $\mathrm{CU}(h)$, $\mathrm{CD}(h)$ and C_h are closed subsets of $S(f)$. To see this, consider any maximally contiguous open portion P of $S(f)$ on which $f''(\cdot) > 0$. If the length of P exceeds h, P contains a subset A of $\mathrm{CU}(h)$ which is maximally contiguous within P. The endpoints of A, generically denoted by x_0, are precisely a distance $h/2$ from either an endpoint of $S(f)$, which is closed and bounded, or an inflection point of $f(x)$. Since $f''(x) > 0$ on $h[x_0]$, x_0 lies in A. Thus A is closed. If the length of P equals h, P contains an isolated point p_0 belonging to $\mathrm{CU}(h)$. If the length of P is less than h, P contains no part of $\mathrm{CU}(h)$. Since $S(f)$ is bounded and h is a fixed width, $\mathrm{CU}(h)$ is the union of a finite number of disjoint closed intervals and possibly a finite number of separated points. So $\mathrm{CU}(h)$ is closed. A similar argument holds for $\mathrm{CD}(h)$.

From Proposition 2.1 in [8] we obtain

$$\text{(2.1)} \qquad \begin{aligned} &\text{On } \mathrm{CU}(h)\colon\ 0 < f(x) < \bar{f}_1(x) < \bar{f}_2(x).\\ &\text{On } \mathrm{CD}(h)\colon\ 0 < \bar{f}_2(x) < \bar{f}_1(x) < f(x).\end{aligned}$$

The kernels K_r and K_w analyzed here are symmetric and rectangular. K_r is nonnegative and K_w assumes negative values in its extremities. These kernels are defined as follows.

DEFINITION 2.3. For $U > 0$,

$$K_r(x, X) = (1/h_1)\,1_{h_1[X]}(x), \qquad K_w(x, X) = k_1(U)\,1_{h_1[X]}(x) + k_2(U)\,1_{h_2[X]}(x),$$

where $k_1(U) = (1+U)/h_1$ and $k_2(U) = -U/h_2$ and where, referring to Definition 2.1,

$$1_{h_1[X]}(x) = \begin{cases} 1 & \text{for } x \text{ in } h_1[X],\\ 0 & \text{otherwise},\end{cases}$$

$$1_{h_2[X]}(x) = \begin{cases} 1 & \text{for } x \text{ in } h_2[X],\\ 0 & \text{otherwise}.\end{cases}$$

DEFINITION 2.4. The parameter β is defined as $\beta = h/h_1 > 1$. Note that $h_2 = h_1(\beta - 1)$.

The kernel K_r is symmetric, rectangular or uniform, and has order $p(K_r) = 2$. The kernel K_w is symmetric and rectangular, formed by symmetrically appending negative mass $-U/2$ to each extremity of K_r and rescaling to ensure that $\int K_w(x, X)\,dx = 1$. By Proposition 4.4 in [8], the kernel K_w has order $p(K_w) = 4$ if and only if its negative mass is specified by

$$U = 1/(\beta(\beta+1)). \tag{2.2}$$

Subsequently, $p(K_w) = 4$.

DEFINITION 2.5. The positive part of a kernel K is denoted by K^+. Note that $K_w^+ = K_r(1+U)$.

DEFINITION 2.6. For n observations $\{X, Y_1, \ldots, Y_{n-1}\}$, giving rise to the kernels $\{K_w(x, X), K_w(x, Y_1), \ldots, K_w(x, Y_{n-1})\}$, $K_v(x, X)$ is defined at x in $h_1[X]$ as $K_w^+(x, X)$ modified by the negative parts of the $K_w(x, Y_j)$. Of the $n-1$ Y_j values, $N_1(x)$ lie in $h_1[x]$ and, for small h_1, their kernels are assumed to equally absorb the negative contributions of the $N_2(x)$ kernels $K_w(x, Y_j)$ for Y_j that lie in $h_2[x]$. The function $K_v(x, X)$ is written in the form

$$K_v(x, X) = 1_{h_1[X]}(x)\left[k_1(U) + \left[N_2(x)/(N_1(x)+1)\right] k_2(U)\right].$$

The expected shape of $K_v(x, X)$, considering the random Y_j, is denoted by $\bar{K}_v(x, X)$.

By Proposition 4.1 in [8], $\bar{K}_v(x, X)$ is written for any X in C_h as

$$\bar{K}_v(x, X) = 1_{h_1[X]}(x)\,[k_1(U) + k_2(U)\,H(x, n)], \tag{2.3}$$

$$H(x, n) = \bar{f}_2(x)\,h_2\,S_1(x),$$

$$S_1(x) = 1 + (1-\bar{f}_1(x)h_1) + (1-\bar{f}_1(x)h_1)^2 + \ldots + (1-\bar{f}_1(x)h_1)^{n-2}.$$

$\bar{K}_v(x, X)$ is considered to be an adaptation of $K_w^+(x, X) = K_r(x, X)(1+U)$ induced by the mean impact of $-U$ masses in $K_w(x, Y)$ for Y near X, and can also be considered as an adaptation of K_r due to the use of the negative kernel mass $-U$.

DEFINITION 2.7. Referring to equation (1.2), MISE $= \mathrm{IB}^2 + \mathrm{IVAR}$ defines the integral squared bias and the integral variance.

Regarding an evaluation of IB^2, for f in C^4, the bias can be written (cf. [1]) for any symmetric kernel K of even order p on support $h[X]$ as

$$B(x) = (1/p!)(h/2)^p f^{(p)}(x_0)\,M_p(K), \tag{2.4}$$

where $f^{(p)}(x_0)$ is the p^{th} derivative of f evaluated at a point x_0 in $h[x]$ and

$$M_p(K) = \int_{-1}^{1} K(v)\,v^p\,dv, \qquad K(v) = (h/2)\,K(x, X), \qquad v = (x-X)/(h/2).$$

By equation (2.4), the bias, using a kernel of order $p = 2$, would be zero on any interval where $f''(x) = 0$; hence f has been taken to be strictly curvilinear.

It can be shown that

$$M_4(K_w) = -1/(5\beta^2). \tag{2.5}$$

Also, by Definition 2.4, equations (2.4) and (2.5), the absolute bias $|B_w(x)|$ for K_w, although of order $O(h^4)$, increases in proportion to β^2. So, there is reason to keep β "small".

CONDITION 2.0. Referring to Definition 2.4, it is assumed that h_1 depends on n so that $h_1 \to 0$ and $nh_1 \to \infty$ as $n \to \infty$. This behavior will become explicit in Sections 3 and 4, where h_1 will be MISE-optimal for a particular kernel.

DEFINITION 2.8. For positive n-independent quantities A and B, $A \ll B$ or $B \gg A$ means that $A/B < 0.1$. For positive quantities A and B, one or both of which can be n-dependent, $A \ll B$ means that A is negligible compared with B for the n being considered and that $A/B \xrightarrow{n} 0$.

For unequal quantities A and B, $A \approx B$ means that A and B are approximately equal. When one or both of the quantities are n-dependent, $A \approx B$ further means that $\lim_{n\to\infty} A = \lim_{n\to\infty} B$. For Lebesgue measurable sets A and B, with $A \subsetneqq B$, where one or both sets can be n-dependent, $A \approx B$ means that $L(B-A) \ll L(B)$, where $L(E)$ is the Lebesgue measure of the set E.

CONDITION 2.1. Referring to Definition 2.4 and equation (2.2), β is considered "large" when $U^2 \ll U \ll 1$, a strategy that will simplify the analysis. As remarked above, bias increases with β. Hence there is a need for a minimally large β. For this purpose, a working value of "large" β is $\beta = 3$, with $U = 1/12$ by equation (2.2). Subsequently, large β will refer to $\beta = 3$. It follows for large β that $h_1 \beta \xrightarrow{n} 0$.

LEMMA 2.0. *$C_h \uparrow S(f)$ as $h \downarrow 0$, where a finite number N of endpoints of $S(f)$ and inflection points of f are the final accumulation points of a growing C_h. Further, for large n, $L(S(f)-C_h) = O(h)$.*

Proof. Referring to Definition 2.2, for large n, the set difference $S(f)-C_h$ consists of tail portions of f and neighborhoods of inflection points of f, each of length $h/2$ and h_1, respectively. Since $S(f)$ is compact and $f \in C^4$, there are a finite number of such intervals in $S(f)-C_h$, since for appropriate $\delta > 0$ and $\varepsilon > 0$, $\{\bar{f}_1^{-1}(\delta, \varepsilon)\}$ and $\{f''^{-1}(-\varepsilon, \varepsilon)\}$ are open covers of tail and inflection point portions of $S(f)-C_h$, respectively. It follows, for a minimally small N, that $L(S(f)-C_h) < Nh$. ∎

DEFINITION 2.9. The supports of the functions $\bar{f}_1(x)$ and $\bar{f}_2(x)$ are denoted by $S(\bar{f}_1)$ and $S(\bar{f}_2)$, respectively. The interior of the support of a function $g(x)$ is written as $S_0(g) = \{x\colon\ g(x) \neq 0\}$.

LEMMA 2.1. *$R \supset S(\bar{f}_2) \supset S(\bar{f}_1) \supset S(f) \supset C_h$.*

Proof. Referring to Definition 2.1, sets $S(\bar{f}_2)$ and $S(\bar{f}_1)$ each extend beyond $S(f)$ by distances $h/2$ and $h_1/2$, respectively, at each point of $S(f)-S_0(f)$. ∎

CONDITION 2.2. Subsequently, large n means that $h = \beta h_1$ is sufficiently small so that

(i) $C_h \approx S(f)$;

(ii) $hf(x) \ll 1$, where $hf(x) \approx \int_{h[x]} f(t)\,dt$ for any x in R;

(iii) $\bar{f}_2(x) \approx \bar{f}_1(x)$ and $\bar{f}_1(x) \approx f(x)$ for any x in R;

(iv) $N/n \approx 0$ and

(v) the usual asymptotic approximations for $B(x)$, $\mathrm{VAR}(x)$, $\overline{\mathrm{MISE}}$ and MISE-optimal h_1 can be employed.

DEFINITION 2.10. Large (n, β) means simultaneous compliance with Conditions 2.1 and 2.2.

DEFINITION 2.11. A kernel is called a *subkernel*, a *true kernel*, or a *suprakernel* if $\int K(x, X)\,dx < 1$, $\int K(x, X)\,dx = 1$, or $\int K(x, X)\,dx > 1$, respectively.

DEFINITION 2.12. The functions $\hat{f}_r(x)$ and $\hat{f}_w(x)$ denote the estimators of f using the kernels K_r and K_w, respectively. Bias, variance, etc., will be subscripted in a similar fashion. $\mathrm{MISE}(K_w)$, e.g., will denote the mean integral squared error for the estimator using the kernel K_w.

DEFINITION 2.13. $E_X\{\cdot\}$ denotes expectation with respect to $f(X)$.

PROPOSITION 2.1. *For estimators based on the kernels K_r and K_w we have*:

$$E_n\{\hat{f}_r(x)\} = E_X\{K_r(x, X)\} = \bar{f}_1(x)$$

and

$$E_n\{\hat{f}_w(x)\} = E_X\{K_w(x, X)\} = (1+U)\,\bar{f}_1(x) - U\bar{f}_2(x).$$

Proof. For an estimator based on a kernel K, $E_n\{\hat{f}\} = E_X\{K\}$ follows since the X_i are independent and identically distributed. To evaluate $E_X\{K(x, X)\}$, reconfigure the kernel as a function of X for fixed x. ∎

COROLLARY 2.1. *For any kernel $K(x, X)$, whether it be a subkernel, a true kernel or a suprakernel, or if its shape depends on X,*

$$E_n\{\hat{f}(x)\} = E_X\{K(x, X)\}. \quad ∎$$

For x in R, it can be shown (cf. Section 4 in [8]) that

$$\begin{aligned} \mathrm{VAR}_r(x) &= (1/n)\left[(\bar{f}_1(x)/h_1) - \bar{f}_1^2(x)\right], \\ \mathrm{VAR}_w(x) - \mathrm{VAR}_r(x) &= (nh_1)^{-1}\left[A(x)\,U^2 + 2I(x)\bar{f}_1(x)\,U\right], \end{aligned} \tag{2.6}$$

where

$$A(x) = \bar{f}_1(x) + (h_1/h_2)\,\bar{f}_2(x) - h_1\left(\bar{f}_2(x) - \bar{f}_1(x)\right)^2, \quad I(x) = 1 + h_1\left(\bar{f}_2(x) - \bar{f}_1(x)\right).$$

PROPOSITION 2.2. *For x in R and large (n, β),*

$$A(x) \approx 1.5\bar{f}_1(x), \qquad I(x) \approx 1,$$

(2.7) $\mathrm{VAR}_w(x) \approx \mathrm{VAR}_r(x) + 2U\bar{f}_1(x)(nh_1)^{-1} > \mathrm{VAR}_r(x),$

(2.8) $\mathrm{VAR}_r(x) \approx \bar{f}_1(x)(nh_1)^{-1},$

(2.9) $\mathrm{MSE}_r(x) \approx \left(\bar{f}_1(x) - f(x)\right)^2 + \bar{f}_1(x)(nh_1)^{-1},$

(2.10) $\mathrm{MSE}_w(x) \approx \left(\bar{f}_1(x) + U\left(\bar{f}_1(x) - \bar{f}_2(x)\right) - f(x)\right)^2 + \bar{f}_1(x)(1+2U)(nh_1)^{-1}.$

Proof. Use equation (2.6) with $h_2 = h_1(\beta - 1)$, $U^2 \ll \bar{U}$ and, e.g., $h_1 \bar{f}_2^2(x) \ll \bar{f}_2(x)$ since $h_1 \bar{f}_2(x) \ll 1$. For example, let us show that $A(x) \approx 1.5\bar{f}_1(x)$. In this case we have $\beta = 3$ and, by Definition 2.4,

$$A(x) = \bar{f}_1(x) + \bar{f}_2(x)/2 - h_1 \bar{f}_2^2(x) - h_1 \bar{f}_1^2(x) + 2h_1 \bar{f}_2(x)\bar{f}_1(x).$$

Then, using Condition 2.2, we obtain

$$A(x) \approx \bar{f}_1(x) + \bar{f}_1(x)/2 - h_1 \bar{f}_1^2(x) - h_1 \bar{f}_1^2(x) + 2h_1 \bar{f}_1(x)\bar{f}_1(x) = 1.5\bar{f}_1(x),$$

where, for x in R,

$$\lim_{n\to\infty} A(x) = \lim_{n\to\infty} 1.5\bar{f}_1(x) = 1.5 f(x). \quad \blacksquare$$

Equation (2.7) illustrates the increase in variance due to the use of negative kernel mass. For large (n, β),

(2.11) $$\mathrm{VAR}_w(x) \approx \mathrm{VAR}_r(x)(1+2U).$$

LEMMA 2.2. *For large (n, β),*

$$L\left(S(\bar{f}_1) - S(f)\right) < L\left(S(\bar{f}_2) - S(f)\right) < L\left(S(f) - C_h\right)$$

and

$$L\left(S(\bar{f}_2) - S(\bar{f}_1)\right) < L\left(S(f) - C_h\right).$$

Proof. For every tail portion of f of length $h/2$ in $S(f) - C_h$, there is a subset of $S(\bar{f}_2) - S(f)$ of length $h/2$, a subset of $S(\bar{f}_1) - S(f)$ of length $h_1/2$ and a subset of $S(\bar{f}_2) - S(\bar{f}_1)$ of length $h_2/2$. $S(f) - C_h$ contains neighborhoods of inflection points of f, while the sets $S(\bar{f}_1) - S(f)$ and $S(\bar{f}_2) - S(f)$ do not. $\blacksquare$

LEMMA 2.3. *For large n,*

$$L\left(S(\bar{f}_1) - S(f)\right) = O(h) \quad \textit{and} \quad L\left(S(\bar{f}_2) - S(f)\right) = O(h).$$

Proof. Referring to Lemmas 2.0 and 2.2, for some minimal N and large n, we have

$$L\left(S(\bar{f}_1) - S(f)\right) < L\left(S(f) - C_h\right) < Nh,$$

$$L\left(S(\bar{f}_2) - S(f)\right) < L\left(S(f) - C_h\right) < Nh. \quad \blacksquare$$

PROPOSITION 2.3. $S(\mathrm{MSE}_w) = S(\bar{f}_2)$ *and* $S(\mathrm{MSE}_r) = S(\bar{f}_1)$.

Proof. From Proposition 2.1 and Corollary 2.1 we obtain $S(B_w) = S(\bar{f}_2)$, $S(\mathrm{VAR}_w) = S(\bar{f}_2)$, $S(B_r) = S(\bar{f}_1)$ and $S(\mathrm{VAR}_r) = S(\bar{f}_1)$. ∎

PROPOSITION 2.4. *For large* (n, β),

$$\mathrm{MISE}(K_r) = \int_{S(\bar{f}_1)} \mathrm{MSE}_r(x)\,dx \approx \int_{S(f)} \mathrm{MSE}_r(x)\,dx$$

and

$$\mathrm{MISE}(K_w) = \int_{S(\bar{f}_2)} \mathrm{MSE}_w(x)\,dx \approx \int_{S(f)} \mathrm{MSE}_w(x)\,dx.$$

Proof. Use Proposition 2.3. Referring to equations (2.9) and (2.10) as well as Condition 2.0 we see that $\mathrm{MSE}_r(x)$ and $\mathrm{MSE}_w(x)$ are bounded on the sets $S(\bar{f}_1)-S(f)$ and $S(\bar{f}_2)-S(f)$, respectively. Then use Lemma 2.3. For example,

$$\begin{aligned}\mathrm{MSE}_r(x) &\approx \bar{f}_1^2(x) - 2f(x)\bar{f}_1(x) + f^2(x) + \bar{f}_1(x)/(nh_1)\\ &\approx \bar{f}_1(x)\left[\bar{f}_1(x) - 2\bar{f}_1(x) + \bar{f}_1(x) + 1/(n/h_1)\right] \approx \bar{f}_1(x)/(nh_1).\end{aligned}$$

From Lemma 2.3, for some x_0 in $S(\bar{f}_1)-S(f)$, we obtain

$$\int_{S(\bar{f}_1)-S(f)} \mathrm{MSE}_r(x)\,dx \approx \bar{f}_1(x_0)Nh/(nh_1) = (N/n)(h/h_1)\bar{f}_1(x_0),$$

where $N/n \approx 0$, $h/h_1 = \beta = 3$ and $\bar{f}_1(x_0) \approx 0$. ∎

Note that a boundary effect due to an X-independent bandwidth h, whereby the support of $\hat{f}$ strictly includes $S(f)$, has been accommodated in Proposition 2.4 for the MISE associated with kernels K_r and K_w.

3. INDUCED KERNEL MASS VARIATION

The formula for $\bar{K}_v(x, X)$ in equation (2.3) will now be refined to show that it contains an f''-related pattern of U-induced mass variation over X in C_h.

PROPOSITION 3.1. *For large* n *and* X *in* C_h,

$$\bar{K}_v(x, X) = K_r(x, X)\left\{1 + U\left[1 - \frac{\bar{f}_2(x)}{\bar{f}_1(x)}\left[1 - (1 - \bar{f}_1(x)h_1)^{n-1}\right]\right]\right\}. \tag{3.1}$$

Proof. Begin with equation (2.3) and Definition 2.3. Convergence of the geometric series $\sum_{m\geq 0} z^m = (1-z)^{-1}$ for $0 < z = 1 - \bar{f}_1(x)h_1 < 1$ implies that

$$1 + z + \ldots + z^{n-2} = (1-z)^{-1}(1 - z^{n-1}).\ ∎$$

DEFINITION 3.1. The sets $\{X \in \mathrm{CU}\colon h[X] \subset \mathrm{CU}\}$ and $\{X \in \mathrm{CD}\colon h[X] \subset \mathrm{CD}\}$ are denoted by CU* and CD*, respectively. The union $\mathrm{CD}^* \cup \mathrm{CU}^*$ is denoted by C_h^*.

LEMMA 3.1. *For large (n, β), $C_h^* \approx C_h$.*

Proof. C_h^* is formed from C_h by deducting a finite number of intervals of length $h/2$ from the edges of C_h. Let this number be m. Then

$$L(C_h) - L(C_h^*) = mh/2 = m\beta h_1/2 \xrightarrow{n} 0.$$

Consequently, for large (n, β), $L(C_h^*) \approx L(C_h)$ and

$$\begin{aligned} L(C_h - C_h^*) &< L(S(f) - C_h^*) \\ &= L(S(f)) - L(C_h^*) \approx L(S(f)) - L(C_h) \ll L(C_h). \ \blacksquare \end{aligned}$$

PROPOSITION 3.2. *For large n, $\bar{K}_v(x, X)$ is a suprakernel for X in CD*.*

Proof. By definition, $X \in \mathrm{CD}^* \Rightarrow h[X] \subset \mathrm{CD}$. Then, for $x \in h[X]$, $0 < \bar{f}_2(x)/\bar{f}_1(x) < 1$. For large n, $0 < \bar{f}_1(x)h_1 \ll 1$. Thus, in equation (3.1), U is multiplied by a positive quantity. $\blacksquare$

LEMMA 3.2. *For large n and x in R,*

$$\begin{aligned} \bar{f}_1(x) - \bar{f}_2(x) &\approx -f''(x)(\beta+1)\beta h_1^2/24, \\ \bar{f}_1(x) - f(x) &\approx f''(x) h_1^2/24. \end{aligned} \tag{3.2}$$

Proof. By Taylor's theorem, for $y = t - x \in [-h/2, h/2]$, fixed x and variable t, $f(t) = f(x+y) \approx f(x) + f'(x)y + f''(x)y^2/2$. Then, e.g.,

$$\bar{f}_1(x) - \bar{f}_2(x) \approx f''(x)\Big[(1/h_1)\int_{h_1[0]} (y^2/2)\,dy - (1/h_2)\int_{h_2[0]} (y^2/2)\,dy\Big],$$

where the integrals are evaluated by using Definitions 2.1 and 2.4 to incorporate the parameter β. $\blacksquare$

DEFINITION 3.2. Subsequently, it is assumed that h_1 is MISE-optimal for an appropriate kernel so that, for large n, $h_1 = O(n^{-\gamma})$, where $0 < \gamma \leqslant 1/5$. The appropriate kernel and the value of γ are to be determined. This particular dependency of h_1 on n is denoted by $h_1(n)$.

LEMMA 3.3. *For large n, $h_1(n) = c(n) M n^{-\gamma}$, where $M > 0$ is a constant and $c(n) \approx 1$ is bounded away from zero. Considering n to be a continuous variable we have*

$$dh_1(n)/dn \approx c(n) M d(n^{-\gamma})/dn.$$

Proof. If, e.g., h_1 is MISE-optimal for the kernel K_r, then, as shown in [8], $h_1 = O(n^{-1/5})$ and for a constant $M_r > 0$ and a function of n, $c_r(n)$, which is approximately unity for large n, $h_1(n) = c_r(n) M_r n^{-1/5}$. $\blacksquare$

PROPOSITION 3.3. *For large n, $\bar{K}_v(x, X)$ is a subkernel for X in CU*.*

Proof. By Definition 3.1, $X \in \mathrm{CU}^* \Rightarrow h[X] \subset \mathrm{CU}$. By Condition 2.2, for x in $h[X]$, $\bar{f}_2(x)/\bar{f}_1(x) \xrightarrow{n} 1$. Using Lemma 3.2 for x in CU, for large n, we have

$$1 < \bar{f}_2(x)/\bar{f}_1(x) \approx 1 + c(x)h_1^2,$$

where, by definition, $c(x) = f''(x)(\beta+1)\beta/(24f(x))$. Also, $0 < c(x)h_1^2 \ll 1$. For equation (3.1), define $0 < T_1(x, n) = (1-\bar{f}_1(x)h_1)^{n-1} < 1$. Referring to Definition 3.2 and Lemma 3.3, take $h_1 = h_1(n)$ and define $d(x) = c(n)M\bar{f}_1(x)$. Since $d(x) > 0$ and $\gamma \leqslant 1/5$, it follows that

$$T_1(x, n) = (1-d(x)n^{-\gamma})^{n-1} \leqslant (1-d(x)n^{-1})^{n-1} \xrightarrow{n} \exp(-Mf(x)) < 1.$$

Then for large n the coefficient of U in equation (3.1) is approximately

$$\begin{aligned} 1-&(1+c(x)h_1^2(n))(1-T_1(x, n)) \\ &= T_1(x, n)(1+c(x)h_1^2(n)) - c(x)h_1^2(n) \approx T_1(x, n) - c(x)h_1^2(n). \end{aligned}$$

It will now be shown that $c(x)h_1^2(n)$ dominates $T_1(x, n)$ for large n so that the coefficient of U in equation (3.1) is negative for X in CU*:

$$\ln T_1(x, n) = (n-1)\ln(1-c(n)M\bar{f}_1(x)n^{-\gamma}).$$

Using l'Hôspital's rule to evaluate $\lim_{n\to\infty} \ln T_1(x, n)$ and, by Lemma 3.3, treating $c(n) \approx 1$ as a constant relative to n, we obtain

$$\lim_{n\to\infty} \ln T_1(x, n) = \lim_{n\to\infty} (-\gamma n^{1-\gamma} d(x)),$$

so that

$$\lim_{n\to\infty} T_1(x, n) = \lim_{n\to\infty} \exp\{-\gamma n^{1-\gamma} d(x)\}.$$

Next, let $T_1(x, n) = b(x, n)\exp\{-\gamma n^{1-\gamma} d(x)\}$, where for large n and x in CU, by definition, $b(x, n) \approx 1$. Then, since $e^y \gg y$ for large positive y, we have

$$\begin{aligned} c(x)h_1^2(n)/T_1(x, n) &\approx c(x)c^2(n)M^2 n^{-2\gamma}\exp\{\gamma n^{1-\gamma} d(x)\}/b(x, n) \\ &\gg c(x)c^2(n)M^2 n^{-2\gamma}(\gamma n^{1-\gamma} d(x))/b(x, n) \\ &= c(x)c^2(n)M^2\gamma d(x)n^{1-3\gamma}/b(x, n) = Q(x, n). \end{aligned}$$

By Definition 3.2 and Lemma 3.3, for $0 < \gamma \leqslant 1/5$ and $c^2(n) > \delta > 0$, we have $Q(x, n) \xrightarrow{n} \infty$. ∎

COROLLARY 3.1. *For large n and X in C_h^*,*

$$\bar{K}_v(x, X) \approx K_r(x, X)(1+U(1-\bar{f}_2(x)/\bar{f}_1(x))) \geqslant 0.$$

Proof. In Proposition 3.3, the coefficient of U in equation (3.1) is approximately $T_1(x, n) - c(x)h_1^2(n) \approx -c(x)h_1^2(n) \approx 1-\bar{f}_2(x)/\bar{f}_1(x)$ for X in CU*. A similar analysis holds for X in CD*. ∎

DEFINITION 3.3. For large n and X in C_h^*, a *mass-adapted kernel that approximates $\bar{K}_v$ and is particularly amenable to the evaluation of estimator performance using mass-adapted kernels* is

$$K_m(x, X) = K_r(x, X)\big(1+U\big(1-\bar{f}_2(x)/\bar{f}_1(x)\big)\big) \geqslant 0. \tag{3.3}$$

PROPOSITION 3.4. *For large (n, β) and X in C_h^*, $K_m(x, X) \approx K_r(x, X)$ and K_m emulates $\bar{K}_v$ in terms of kernel mass variation and asymmetry.*

Proof. For large (n, β), $K_m \approx K_r$ since $U \ll 1$ and the coefficient of U in equation (3.3) is a small fraction. Also, $K_m \geqslant 0$ and $S(K_m) = S(K_r) = h_1[X]$. Referring to equation (2.1), K_m is a reduced version of K_r, and so is a subkernel for X in CU*. K_m is an elevated version of K_r, and so is a suprakernel for X in CD*. The term $-\bar{f}_2(x)/\bar{f}_1(x) \approx -1$ in equation (3.3) gives K_m a distinctive asymmetry, since

$$d/dx\big(-\bar{f}_2(x)/\bar{f}_1(x)\big) \approx -f'(x)\big(\bar{f}_1(x)-\bar{f}_2(x)\big)/\bar{f}_1^2(x). \tag{3.4}$$

Thus the sign of $d/dx\big(K_m(x, X)\big)$ opposes $f'(x)$ on CD and is the same as $f'(x)$ on CU. This same kind of asymmetry is shown for $\bar{K}_v$ in Proposition 4.5 in [8]. ∎

PROPOSITION 3.5. *For large n and X in C_h^*,*

$$K_m(x, X) \approx K_r(x, X)\big(1-f''(x)h_1^2/(24f(x))\big) \geqslant 0. \tag{3.5}$$

Proof. Use equation (3.3), Lemma 3.2 and equation (2.2). ∎

DEFINITION 3.4. For large (n, β) and X in C_h^*, a *mass-adapted kernel that approximates $\bar{K}_v$ and is amenable to pilot estimation of the factor that modifies the slope and mass of K_r* is

$$K_{p1}(x, X) = K_r(x, X)\big(1-f''(x)h_1^2(n)/(24f(x))\big) \geqslant 0. \tag{3.6}$$

PROPOSITION 3.6. *For large (n, β), the definition of the kernel K_m in equation* (3.3) *can be extended from X in C_h^* and x in C_h to X in $S_0(f)$ and x in $S(f)$ without incurring negative or infinite kernel values.*

Proof. Referring to Definition 3.1 and Lemma 3.1, the set $S(f)-C_h^*$ consists of (i) small tail portions of f and (ii) small neighborhoods of inflection points of f.

In case (i), $S(\bar{f}_2) \supsetneqq S(\bar{f}_1) \supsetneqq S(f)$ and $0 < \delta_1 < \bar{f}_1(x) < \bar{f}_2(x)$ with $\bar{f}_2(x) \approx \approx \bar{f}_1(x)$ and $U \ll 1$.

In case (ii), $\bar{f}_1(x) > \delta_2$, $\bar{f}_2(x) \approx \bar{f}_1(x)$ and $U \ll 1$. ∎

The proof of Proposition 3.6 states that $K_m(x, X) \approx K_r(x, X)$ for large (n, β) and for X in $S_0(f)$ and x in $S(f)-C_h^*$. We can now write a definition of $K_m(x, X)$ that extends its domain to x in R, i.e., to x in $S(\bar{f}_1)-S(f)$, by appealing to the notion that kernel mass adaptation is not required to reduce bias "where" $f''(x) = 0$. Recall that Lemma 2.0 states that $S(f)-S_0(f)$ consists of a finite number of points. Also, the assumption $f \in C^4$ implies that f, f' and f'' all smoothly go to zero on $S(f)-S_0(f)$.

DEFINITION 3.5. For x in $R-S(f)$ and any X in $S_0(f)$, define

$$K_m(x, X) = K_r(x, X), \tag{3.7}$$

so that for X in $S_0(f)$:

$$K_m(x, X) = \begin{cases} K_r(x, X)\left[1+U\left(1-\bar{f}_2(x)/\bar{f}_1(x)\right)\right] & \text{for } x \text{ in } S(f), \\ K_r(x, X) & \text{for } x \text{ in } R-S(f). \end{cases} \tag{3.8}$$

Definition 3.5 avoids boundary effects that would otherwise evolve from K_m as written in equation (3.3) because, by Lemma 2.1, the domain of $K_r(x, X)$ reaches the edge of $S(\bar{f}_1)$ and the ratio $\bar{f}_2(x)/\bar{f}_1(x)$ approaches ∞ as x in $S_0(\bar{f}_1)$ approaches an edge of $S(\bar{f}_1)$.

In practice, equation (3.8) is tantamount to using K_r in lieu of K_m at values of X_i near an edge X_0 of $S_0(f)$ where $f''(X_i)$ is deemed to be sufficiently small. In theory, such X_i are within a distance $h_1/2$ of X_0.

The kernels $\bar{K}_v$, K_{p1} and K_m in equations (3.1), (3.6) and (3.8), respectively, all involve the unknown density f. In practice, a pilot estimate of f would be required to formulate the kernels. In this theoretical discussion, however, it will suffice to treat f as an unknown but differentiable and integrable function.

4. ESTIMATOR PERFORMANCE USING VARIABLE KERNEL MASS

The discussion now focuses on the kernel $K_m(x, X)$ in equation (3.8), which adapts kernel mass in a manner analogous to $\bar{K}_v$ for X in C_h^* and x in C_h and otherwise behaves like $K_r(x, X)$ for large (n, β).

The asymptotic properties of $\hat{f}_m(x) = \sum K_m(x, X_i)$ will now be developed. In particular, it will be shown that $\mathrm{MISE}(K_m) = O(n^{-8/9})$ when $h_1(n)$ is MISE-optimized for the kernel K_m, and that $\gamma = 1/9$, so that indeed $\gamma \leqslant 1/5$, as in Definition 3.2.

PROPOSITION 4.1. *For large (n, β) and x in $S(f)$,*
(a) $B_m(x) = B_w(x) = O\left(h_1^4(n)\right)$,
(b) $\mathrm{VAR}_m(x) \approx \mathrm{VAR}_r(x)$.

Proof. Note that $x \in S(f) \Rightarrow x \in S_0(\bar{f}_1)$. To prove (a), from Corollary 2.1 and equation (3.8), for any (n, β), we infer that

$$\begin{aligned} E_n\{\hat{f}_m(x)\} &= E_X\{K_m(x, X)\} = \left[1+U\left(1-\bar{f}_2(x)/\bar{f}_1(x)\right)\right] E_X\{K_r(x, X)\} \\ &= \bar{f}_1(x)+U\left(\bar{f}_1(x)-\bar{f}_2(x)\right) = E_n\{\hat{f}_w(x)\}. \end{aligned} \tag{4.1}$$

So, referring to equation (2.4), we have $B_m(x) = B_w(x) = O(h^4) = \beta^4 O\left(h_1^4(n)\right)$, since $p(K_w) = 4$.

To prove (b), from equations (1.1), (3.3) and (4.1), for any (n, β), we see that

$$\hat{f}_m(x) = \left(1+U\left(1-\bar{f}_2(x)/\bar{f}_1(x)\right)\right)\hat{f}_r(x)$$

implies

$$\mathrm{VAR}_m(x) = \left(1+U\left(1-\bar{f}_2(x)/\bar{f}_1(x)\right)\right)^2 \mathrm{VAR}_r(x). \tag{4.2}$$

Using equations (2.2) and (3.2), with $\bar{f}_1(x) \approx \bar{f}_2(x)$ for large n and $U \ll 1$ for large β, we infer that equation (4.2) takes the form $\mathrm{VAR}_m(x) \approx \mathrm{VAR}_r(x)$. ∎

In fact, $\mathrm{VAR}_m(x)$ is slightly larger than $\mathrm{VAR}_r(x)$ on CD and slightly smaller than $\mathrm{VAR}_r(x)$ on CU.

COROLLARY 4.1. *For large (n, β) and X in $S_0(f)$, the mass-adapted kernel K_m essentially eliminates the variance increment associated with the negative mass $-U$ in K_w.*

Proof. From equation (2.7) we get

$$\mathrm{VAR}_w(x) \approx \mathrm{VAR}_r(x) + 2U\bar{f}_1(x)/(nh_1).$$

Squaring the term on the right-hand side of equation (4.2) and implementing equation (2.8) with $U^2 \ll U$, we obtain

$$\begin{aligned} \mathrm{VAR}_m(x) &\approx \mathrm{VAR}_r(x) + 2U\left(\bar{f}_1(x) - \bar{f}_2(x)\right)/(nh_1) \\ &\approx \mathrm{VAR}_w(x) - 2U\bar{f}_2(x)/(nh_1). \end{aligned} \tag{4.3}$$

So, the variance increment $2U\bar{f}_1(x)/(nh_1)$ in $\mathrm{VAR}_w(x)$ is offset by the term $-2U\bar{f}_2(x)/(nh_1)$ in $\mathrm{VAR}_m(x)$, with $\bar{f}_1(x) \approx \bar{f}_2(x)$ for large (n, β). ∎

LEMMA 4.1. *For x in $R-S(f)$,*

$$\mathrm{VAR}_m(x) = \mathrm{VAR}_r(x), \qquad \mathrm{MSE}_m(x) = \mathrm{MSE}_r(x).$$

Proof. For x in $R-S(f)$, from equation (3.7) we get $\hat{f}_m(x) = \hat{f}_r(x)$, $B_m(x) = B_r(x)$ and $\mathrm{VAR}_m(x) = \mathrm{VAR}_r(x)$, so that $\mathrm{MSE}_m(x) = \mathrm{MSE}_r(x)$. ∎

PROPOSITION 4.2. *For large (n, β), $h_1(n) = O(n^{-1/9})$ and $\mathrm{MISE}(K_m) = O(n^{-8/9})$.*

Proof. By definition,

$$\mathrm{MISE}(K_m) = \int_{R-S(f)} \mathrm{MSE}_m(x)\,dx + \int_{S(f)} \mathrm{MSE}_m(x)\,dx.$$

Invoke Lemma 4.1. Also, from Proposition 2.3 we obtain $S(\mathrm{MSE}_r) = S(\bar{f}_1)$, so that

$$\mathrm{MISE}(K_m) = \int_{S(\bar{f}_1)-S(f)} \mathrm{MSE}_r(x)\,dx + \int_{S(f)} \mathrm{MSE}_m(x)\,dx.$$

Referring to the proof of Proposition 2.4, $\mathrm{MSE}_r(x)$ is bounded on the set $S(\bar{f}_1) - S(f)$ and $L\left(S(\bar{f}_1) - S(f)\right) = O(h)$. So, for large (n, β),

$$\mathrm{MISE}(K_m) \approx \int_{S(f)} \mathrm{MSE}_m(x)\,dx.$$

From Proposition 4.1 we get

$$\mathrm{MISE}(K_m) \approx \int_{S(f)} B_w^2(x) + \mathrm{VAR}_r(x)\, dx.$$

Then, using equations (2.4), (2.5) and (2.8), we have

$$\text{(4.4)} \qquad \mathrm{MISE}(K_m) \approx \left[\left(\frac{-1}{5\beta^2}\right)\left(\frac{\beta h_1}{2}\right)^4\left(\frac{1}{4!}\right)\right]^2 \int_R [f^{(4)}(x)]^2\, dx + \frac{1}{nh_1}.$$

By Lemma 4a in [6], the optimal $h_1(n)$ to minimize $\mathrm{MISE}(K_m)$ is approximated by

$$\text{(4.5)} \qquad h_1(n) = 4.259\, \beta^{-4/9} \Big[\int_R [f^{(4)}(x)]^2\, dx\Big]^{-1/9} n^{-1/9}. \quad ■$$

Since the support for f is bounded and the bandwidth h (or h_1) is independent of X, the support of the estimator $\hat{f}(x)$ can include the support of f. This has a boundary effect on $B(x)$, $\mathrm{VAR}(x)$ and $\mathrm{MSE}(x)$, as indicated in equations (2.7)–(2.10). But, as demonstrated in the proof of Proposition 4.2, the integral boundary effect on the MISE associated with the kernel K_m is negligible.

The estimates of $h_1(n)$ in equation (4.5), of $f''(x)/f(x)$ in equation (3.6), and of $\bar{f}_2(x)/\bar{f}_1(x)$ in equation (3.8) would be data-based and would degrade attained MISE, an important issue beyond the scope of this study.

5. COMPARING THE KERNEL ASYMMETRY AND KERNEL MASS ADAPTATIONS

There are two distinct U-induced adaptations in K_m, involving (i) variable kernel mass and (ii) kernel shape asymmetry. Here, these two adaptations will be compared in terms of their asymptotic contributions to bias reduction. To do this, referring to Proposition 3.4, the shape asymmetry in equation (3.4) will be isolated and imposed upon the true kernel K_r. The kernel K_m in Definition 3.5 already incorporates both adaptations.

PROPOSITION 5.1. *For large* (n, β) *and* x *in* C_h:

(a) *The bias reduction achieved by* $\hat{f}_m$ *relative to* $\hat{f}_r$ *due to the combination of U-induced kernel mass and kernel shape variations is*

$$\Delta B_M(x) = U\,|\bar{f}_1(x) - \bar{f}_2(x)|.$$

(b) *The bias reduction achieved by* $\hat{f}_m$ *relative to* $\hat{f}_r$ *due only to the U-induced kernel asymmetry is*

$$\Delta B_A(x) \approx U\,|\bar{f}_1(x) - \bar{f}_2(x)| \cdot |x - \bar{X}| \cdot |f'(x)|/\bar{f}_1(x),$$

where $\bar{X}$ *is the expected value of* X *over the interval* $h_1[x]$.

Proof. First we show (a) for, e.g., x in CD. From Definition 3.5, the expected rise at x in the estimator $\hat{f}_m(x)$ relative to $\hat{f}_r(x)$ due to the suprakernel status and asymmetry of K_m is

$$E_n\{\hat{f}_m(x)\}-E_n\{\hat{f}_r(x)\} = E_X\{K_m(x, X)-K_r(x, X)\}$$
$$= E_X\{K_r(x, X)\,U(1-\bar{f}_2(x)/\bar{f}_1(x))\} = \bar{f}_1(x)\,U(1-\bar{f}_2(x)/\bar{f}_1(x)).$$

Now we show (b) for, e.g., x in CD where $f'(x) > 0$. Referring to Proposition 3.4, define an asymmetry-adapted true kernel variant of K_r as

$$K_{ra}(x, X) = p(x)K_r(x, X)(1+b(x)(x-X)),$$

where, from equations (3.4) and (3.8), for x in CD we get

$$b(x) = -Uf'(x)(\bar{f}_1(x)-\bar{f}_2(x))/(\bar{f}_1(x))^2 < 0$$

and $p(x) \approx 1$ ensures that K_{ra} integrates to 1. Then for fixed but arbitrary x we get

$$E_X\{K_{ra}(x, X)\} \approx E_X\{K_r(x, X)\}+b(x)(x-\bar{X})\,\bar{f}_1(x),$$

where

$$\bar{X} = (h_1\bar{f}_1(x))^{-1}\cdot \int_{h_1[x]} Xf(X)\,dX > x. \quad ■ \tag{5.1}$$

Proposition 5.2. *For large (n, β) and x in C_h, the bias reducing effect of U-induced K_m mass variability dominates the bias reducing effect of U-induced K_m asymmetry.*

Proof. Referring to Proposition 5.1, we get

$$\Delta B_A(x)/\Delta B_M(x) \approx |x-\bar{X}|\,|f'(x)|/\bar{f}_1(x).$$

At a fixed x in C_h, for large n, the factor $f'(x)/\bar{f}_1(x)$ is close to $f'(x)/f(x)$. Referring to Lemma 3.2 and equation (5.1),

$$\bar{X}-x = \bar{y} = (h_1\bar{f}_1(x))^{-1}\cdot \int_{-h_1/2}^{h_1/2} f(x+y)\,y\,dy$$
$$\approx f'(x)\,h_1^2(n)/(12\bar{f}_1(x)) \xrightarrow{n} 0. \quad ■$$

As a consequence of Proposition 5.2, for large (n, β) and X in C_h^*, equation (3.5) can be rewritten to focus on kernel mass variation:

$$K_m(x, X) \approx K_r(x, X)\left[1-f''(X)\,h_1^2(n)/(24f(X))\right] \geqslant 0. \tag{5.2}$$

Definition 5.1. For large (n, β), X in C_h^* and x in C_h, a *mass-adapted kernel that approximates $\bar{K}_v$ and is amenable to pilot estimation of the factor that modifies the mass of* K_r is

$$K_{p2}(x, X) = K_r(x, X)\left(1-f''(X)\,h_1^2(n)/(24f(X))\right) \geqslant 0. \tag{5.3}$$

By equation (5.2), the mass change in the kernel K_m relative to K_r, for X in C_h^*, is

$$\Delta I(X) \approx -f''(X)\, h_1^2(n)/(24 f(X)). \tag{5.4}$$

PROPOSITION 5.3. *A variation of K_m in equation* (3.8), *defined for large* (n, β) *and X in $S_0(f)$, that focuses on kernel mass adaptation is*

$$K_{m1}(x, X) = \begin{cases} K_r(x, X)\left[1+U\left(1-\bar{f}_2(X)/\bar{f}_1(X)\right)\right] & \textit{for } x \textit{ in } S(f), \\ K_r(x, X) & \textit{for } x \textit{ in } R-S(f). \end{cases} \tag{5.5}$$

Proof. The right-hand side of equation (5.2) can be approximated by using $\bar{f}_1(X) \approx f(X)$, while avoiding boundary problems associated with $f(X)$ approaching 0 as X approaches a point in $S(f)-S_0(f)$. Then the domain of the right-hand side of equation (5.2) can be extended from X in C_h^* and x in C_h to X in $S_0(f)$ and x in $S(f)$, using Lemma 3.2 and equation (2.2), whereby

$$1-[f''(X)\, h_1^2(n)/24 \bar{f}_1(X)] \approx 1+U[1-\bar{f}_2(X)/\bar{f}_1(X)]. \;\blacksquare$$

The kernel K_{m1} is the simplest version of $\bar{K}_v$, incorporating only variable kernel mass, the dominant source of U-induced bias reduction for large (n, β). It can be shown that the estimator $\hat{f}_{m1} = (1/n)\sum K_{m1}(x, X_i)$ has the same asymptotic properties as $\hat{f}_m$. Also, the mass of either $\hat{f}_m$ or $\hat{f}_{m1}$ can be corrected by using Gajek's P-algorithm in [3].

By virtue of the definition of K_m in equation (3.8) and the definition of K_{m1} in equation (5.5), there is no "clipping" problem associated with $\hat{f}_m$ or $\hat{f}_{m1}$, in the sense that either $\bar{f}_2(x)/\bar{f}_1(x) \approx 1$ for x in $S(f)$ near the edges of $S(f)$ or $\bar{f}_2(X)/\bar{f}_1(X) \approx 1$ for X in $S_0(f)$ near the edges of $S_0(f)$, so that neither ratio requires artificial bounding or "clipping".

It has been assumed that the effect of the Gajek truncation on the variable shape or mass of the kernel $\bar{K}_v$ is relatively insignificant.

6. KERNEL MASS VARIATION

It will now be shown that the mass variation induced by $-U$ in $K_{m1}(x, X)$ is small with respect to unity for all observations X, for large (n, β).

PROPOSITION 6.1. *For large (n, β), the change in kernel mass, $\Delta_M(X)$, for $K_{m1}(x, X)$ relative to $K_r(x, X)$ in equation* (5.5) *satisfies the condition* $\Delta_M(X) \ll 1$.

Proof. Referring to equation (5.5),

$$\Delta_M(X) = U\left[1-\left(\bar{f}_2(X)/\bar{f}_1(X)\right)\right]. \tag{6.1}$$

By Condition 2.2, for X in $S_0(f)$, we have

$$\bar{f}_2(X)/\bar{f}_1(X) \xrightarrow{n} f(X)/f(X) = 1.$$

For large (n, β), from Lemma 3.2, for X in $S_0(f)$, we obtain

$$\bar{f}_2(X)/\bar{f}_1(X) \approx 1 + c_1(X) h_1^2(n),$$

where $c_1(X) = f''(X)(\beta+1)(\beta)/(24\bar{f}_1(X))$ and $|c_1(X)| h_1^2(n) \ll 1$. Thus, for X in $S_0(f)$,

$$|\Delta_M(X)| \approx U \cdot |c_1(X)| \cdot h_1^2(n) \ll U = 1/12 \ll 1. \blacksquare$$

COROLLARY 6.1. *Referring to equations* (5.2)–(5.4), *for* X *in* $S_0(f)$,

$$[|f''(X)| \cdot h_1^2(n)]/(24\bar{f}_1(X)) \ll 1. \tag{6.2}$$

Proof. From equation (2.2) we obtain $U = 1/\beta(\beta+1)$. $\blacksquare$

7. AN EXTENSION TO CONTINUOUS KERNELS

Continuous analogs of the kernels K_r and K_w are denoted by K_c and K_{wc}, respectively, with bandwidths h_{1c} and $h_c = \beta_c h_{1c}$ and negative mass $-U_c$ in K_{wc} over the split interval $h_{2c}[X] = h_c[X] - h_{1c}[X]$. By Proposition 3.1 in [8], K_{wc} (like K_w) shifts expected estimator mass from CU to CD. For large (n, β_c), it is anticipated that this mass shift is accomplished largely by kernel mass variation akin to that in equation (5.4).

A generalization of Definition 2.6 must account for the fact that $K_{wc}^+(x, X)$ goes to zero at the edges of $h_{1c}[X]$ and is limited, in terms of absorbing negative value, by its value at x. An extended definition of K_v can be written as

$$K_{vc}(x, X) = K_{wc}^+(x, X)\left[1 + \frac{\sum_{Y_j \in h_{2c}[x]} K_{wc}(x, Y_j)}{K_{wc}^+(x, X) + \sum_{Y_j \in h_{1c}[x]} K_{wc}^+(x, Y_j)}\right].$$

Subsequently, analysis of the continuous kernel case should parallel that for rectangular kernels.

REFERENCES

[1] A. Berlinet, *Hierarchies of higher order kernels*, Probab. Theory Related Fields 94 (1993), pp. 489–504.

[2] J. Fan, P. Hall, M. Martin and P. Patil, *On local smoothing of nonparametric curve estimators*, JASA 91 (1996), pp. 258–266.

[3] L. Gajek, *On improving density estimators which are not bona fide functions*, Ann. Statist. 14 (1986), pp. 1612–1618.

[4] M. C. Jones, *Variable kernel density estimators and variable kernel density estimators*, Austral. J. Statist. 32 (1990), pp. 361–371.

[5] I. McKay, *A note on bias reduction in variable-kernel estimates*, Canad. J. Statist. 21 (1993), pp. 367–375.

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

[7] B. W. Silverman, *Density estimation for statistics and data analysis*, Chapman and Hall, New York 1990.

[8] M. Sturgeon, *Mass shifting roles of negative kernel mass in density estimation*, Ann. Inst. Statist. Math. 48 (1996), pp. 675–686.

[9] G. R. Terrell and D. W. Scott, *On improving convergence rates for nonnegative kernel density estimators*, Ann. Statist. 8 (1980), pp. 1160–1163.

[10] G. R. Terrell and D. W. Scott, *Variable kernel density estimation*, Ann. Statist. 20 (1992), pp. 1236–1265.

Department of Mathematics
Salem State College
13 Castle Circle
Peabody, MA 01960, U.S.A.

Received on 7.10.1998;
revised version on 24.3.2003