Ann. Inst. Statist. Math.
Vol. 48, No. 4, 675–686 (1996)

MASS SHIFTING ROLES OF NEGATIVE KERNEL MASS IN DENSITY ESTIMATION

MICHAEL STURGEON

Mathematics Department, Salem State College, 45 Blaisdell Drive, Carlisle, MA 01741, U.S.A.

(Received March 13, 1995; revised October 9, 1995)

Abstract. $f(x)$ is a univariate density in C^4 with bounded support. For any n and sufficiently small kernel bandwidths, the symmetric appendage of any negative mass, $-U$, to any smooth unimodal symmetric kernel of order $p = 2$ shifts expected estimator mass from regions where $f''(x) > 0$ to regions where $f''(x) < 0$. For large n, the mean automatic kernel adaptation induced by $-U$ is analyzed in the simplest MISE reduction scenario: The symmetric appendage of $-U$ to the uniform kernel $K(x, X)$ over MISE-optimal bandwidths reduces MISE by shifting $K(x, X)$ mass asymmetrically across the observation X in the direction of decreasing $|f''(x)|$.

Key words and phrases: Kernel adaptation, higher order kernels, induced kernel asymmetry.

1. Introduction

The kernel $K(x, X)$ is a symmetric measurable function of x, centered on an observation X, that integrates to 1 over an interval $(X - H/2, X + H/2)$. $\{X_i\}_1^n$ is a set of n random observations, each with density f. E_n denotes expectation with respect to the joint observation density f^n on R^n.

For a univariate density estimator $\hat{f}(x) = (1/n)\sum_{i=1}^n K(x, X_i)$, it is well known that bias, $B(x) = E_n\{\hat{f}(x) - f(x)\}$, inherently accompanies the use of Parzen kernels. For a density f in C^∞ and a kernel bandwidth H, bias is reduced to $O(H^p)$ by using a negative-valued kernel of order $p \geq 3$, cf. Gasser and Müller (1979). There exist optimally shaped kernels of order $p \geq 3$, Berlinet (1993), which can either globally minimize VAR(x) or minimize MISE $= \int_R \text{MSE}(x)dx$, MSE$(x) = \text{VAR}(x) + B^2(x)$, VAR$(x) = E_n\{[\hat{f}(x) - E_n\{\hat{f}(x)\}]^2\}$, for an appropriate choice of $(p, H(n))$. Such kernels can reduce MISE to $O(n^{-2p/(2p+1)})$, Schucany and Sommers (1977), and are viable since Gajek (1986) provides a truncation algorithm for eliminating the estimator's negative values while further reducing MISE.

It is well known that higher order kernels offset increased variance with reduced bias, for large n. Also, for any $n \geq 1$, the bias at x for a kernel of order

p is proportional to the p-th derivative of f evaluated near x. Thus a positive symmetric kernel (of order $p = 2$) tends to overestimate f where $f''(x) > 0$ and tends to underestimate f where $f''(x) < 0$, while the bias is zero on intervals where $f''(x) = 0$.

Aside from the fact that higher order kernels can reduce bias and MISE, it is not known how symmetrically appended negative kernel mass $-U$ mechanically acts to accomplish bias and MISE reduction. In this regard, one can pose three basic questions: (1) For what values of (U, n) is the direction of bias reversed? (2) Where and why is the direction of bias reversed? (3) What is the automatic kernel adaptation process by which negative kernel mass achieves bias and MISE reduction?

Questions (1) and (2) are addressed for symmetric rectangular and continuous unimodal kernels, respectively, in Sections 2 and 3, where the discussion implicitly includes kernels of order $p = \{2, 4\}$. It will be shown, for f in C^4 with bounded support, that the bias is reversed for any $U > 0$ and any $n \geq 1$, at any x where $f''(x) \neq 0$. This reversal occurs simply because the sign of the symmetric negative kernel mass opposes that of the symmetric positive mass in $K(x, X)$, suggesting that kernel symmetry is somehow inappropriate whenever $f''(x) \neq 0$. For f in C^4, the bias reversal is tantamount to a shift in expected estimator mass from the region where $f''(x) > 0$ to the region where $f''(x) < 0$.

Question (3) is addressed, at least by example, in Section 4. It is clear that the negative kernel masses in the kernels $\{K(x, X_j)\}_1^{n-1}$, for X_j sufficiently near X_n, diminish the positive mass in $K(x, X_n)$ in some adaptive manner. Beginning with a uniform kernel on MISE-optimal bandwidth H and symmetrically appending appropriate $-U$ mass to produce a rectangular kernel of order $p = 4$ over MISE-optimal bandwidth $H' > H$, one might expect that the U-induced adaptation would create, on average, a kernel akin to the Epanechnikov kernel. However, as shown in Section 4, this simplest adaptation actually creates an asymmetric kernel. The illustration involves uniform kernels $K_r(x, X)$ of order $p = 2$ and rectangular kernels $K_w(x, X)$ when (a) K_w is of order $p = 4$, (b) K_r and K_w are both on MISE-optimal bandwidths, (c) MISE(K_w) < MISE(K_r) and (d) the positive part of K_w, denoted K_w^+, differs from K_r only by a constant. Then, the U-induced mean adaptation of K_w^+ is shown to be equivalent to a mean adaptation of $K_r(x, X)$, whereby K_r mass is shifted asymmetrically across X in the direction of decreasing $|f''(x)|$. In this case, the efficacy of higher order kernels for sufficiently large finite n derives from the capacity of $-U$ to induce a certain f''-related kernel asymmetry. In more general circumstances, when item (d) cannot be satisfied, as when considering MISE-optimally shaped kernels of order $p = \{2, 4\}$, it is conjectured that a similar adaptation process is at work creating an asymmetric shape or an asymmetric bandwidth.

2. Preliminaries

The observations $\{X_i\}_1^n$ are i.i.d. and real-valued, with unknown but strictly curvilinear density f with bounded support $S(f)$. It is assumed that f is in C^4, so that f is bounded.

DEFINITION 2.1. For any fixed $h > 0$, with $h = h_1 + h_2$, $h[x]$ is the open interval of length h centered on x, $h_1[x]$ is the open interval of length h_1 centered on x and $h_2[x]$ is the disjoint set $h[x] - h_1[x]$ with length $h_2 = h - h_1$. Integral averages of f over the sets h_1, h_2 and h are defined for any x in R as:

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

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

$$CU(h) = \{x \in S(f) : f''(\cdot) > 0 \text{ on } h[x]\}$$
$$CD(h) = \{x \in S(f) : f''(\cdot) < 0 \text{ on } h[x]\}$$

define closed proper subsets of $S(f)$ on which f is, respectively, concave up and concave down. Denote $CU(h) \cup CD(h) = C_h$.

Subsequently, h is taken sufficiently small that $CU(h)$ and $CD(h)$ both contain intervals. Note that $C_h \uparrow S(f)$ as $h \downarrow 0$, for strictly curvilinear f.

PROPOSITION 2.1.

$$\textit{On } CU(h) : 0 < f(x) < \bar{f}_1(x) < \bar{f}_2(x)$$
$$\textit{On } CD(h) : 0 < \bar{f}_2(x) < \bar{f}_1(x) < f(x).$$

COROLLARY 2.1. *For symmetrical* $h_{2'}[x] \subsetneqq h_2[x]$ *and symmetrical* $h_{1'}[x] \subsetneqq h_1[x]$,

$$\textit{On } CU(h) : 0 < \bar{f}_{1'}(x) < \bar{f}_{2'}(x)$$
$$\textit{On } CD(h) : 0 < \bar{f}_{2'}(x) < \bar{f}_{1'}(x).$$

DEFINITION 2.3. For $h_1 < h$ and $U > 0$,

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

where $k_1(U) = (1 + U)/h_1$ and $k_2(U) = -U/h_2$.

DEFINITION 2.4. $\hat{f}_r(x)$ and $\hat{f}_w(x)$ denote estimators using kernels $K_r(x, X)$ and $K_w(x, X)$, respectively.

PROPOSITION 2.2. *For any $U > 0$ and any $n \geq 1$,*

$$E_n\{\hat{f}_w(x)\} = \begin{cases} E_n\{\hat{f}_r(x)\} + U|\bar{f}_1(x) - \bar{f}_2(x)| & on\ CD(h) \\ E_n\{\hat{f}_r(x)\} - U|\bar{f}_1(x) - \bar{f}_2(x)| & on\ CU(h) \end{cases}$$

where

$$E_n\{\hat{f}_r(x)\} = \bar{f}_1(x) \begin{cases} < f(x) & on\ CD(h) \\ > f(x) & on\ CU(h). \end{cases}$$

PROOF.

$$E_n\{\hat{f}_r(x)\} = E_X\{K_r(x,X)\} = \int_R K_r(x,X)f(X)dX = \bar{f}_1(x), \tag{2.1}$$

where $K_r(x,X)$ is reconfigured as a function of X for fixed x. Similarly, $E_n\{\hat{f}_w(x)\} = (1+U)\bar{f}_1(x) - U\bar{f}_2(x)$. Proposition 2.1 completes the proof. □

COROLLARY 2.2. *Proposition 2.2 extends to estimators using kernels based on symmetric intervals $h_{1'}[X] \subsetneqq h_1[X]$ and $h_{2'}[X] \subsetneqq h_2[X]$.*

PROOF. Use Corollary 2.1. □

So, for any (U,n), the symmetric appendage of $-U$ to K_r reverses the direction of bias on C_h. The pattern of bias and bias reversal also suggests that kernel symmetry is inappropriate when $f'' \neq 0$, a notion that will be pursued further in Section 4. For example, considering any x in $CU(h)$: For the kernel K_r, the bias is expressed $B_r(x) = \bar{f}_1(x) - f(x) > 0$. For the kernel K_w, defining $B_u(x) = \bar{f}_2(x) - f(x)$, the bias is $B_w(x) = B_r(x)(1+U) - B_u(x)U$, which is just a sum of two biases $B_r(x)$ and $B_u(x)$, each weighted and signed by the K_w kernel masses evenly distributed over the respective sets $h_1[x]$ and $h_2[x]$. The biases B_r and B_u are each directly the result of kernel symmetry, when $f''(x) \neq 0$. It so happens that B_u reverses B_r, since $-U < 0$.

3. Expected estimator mass shift for continuous kernels

Proposition 2.2 is next extended to continuous analogs of K_r and K_w. $K_c(x,X)$ is any nonnegative smooth symmetric unimodal kernel with support $h_1[X] \subset h[X]$. For $U > 0$, $K_{cw}(x,X)$ is defined as $(1+U)K_c(x,X)$ on $h_1[X]$, with unimodal smooth symmetric negative side lobes of total mass $-U$ supported over $h_2[X]$, so that K_{cw} is continuous and symmetric over $h[X]$. Denote the estimators using K_c and K_{cw} as $\hat{f}_c(x)$ and $\hat{f}_{cw}(x)$, respectively.

For $f(x)$ in C^4, the bias can be written, cf. Berlinet (1993), for any smooth symmetric kernel K of order $p = \{2,4\}$ as

$$B(x) = (1/p!)(h/2)^p f^{(p)}(x_0)M_p(K), \tag{3.1}$$

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

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

So, quite generally, if $p = 2$, the bias is positive over $CU(h)$ and negative over $CD(h)$.

PROPOSITION 3.1. *For any $U > 0$ and any $n \geq 1$,*

$$E_n\{\hat{f}_{cw}(x)\} = \begin{cases} E_n\{\hat{f}_c(x)\} + UD(x,n) \text{ on } CD(h) \\ E_n\{\hat{f}_c(x)\} - UD(x,n) \text{ on } CU(h), \end{cases}$$

where $D(x,n) > 0$ and

$$E_n\{\hat{f}_c(x)\} \begin{cases} < f(x) \text{ on } CD(h) \\ > f(x) \text{ on } CU(h). \end{cases}$$

PROOF. For given n and $\{X_j\}_1^n$, approximate each $K_{cw}(x,X_j)$ by step function kernels

$$K_m(x,X_j) = \sum_{i=1}^{m} a_{i,m} K_{i,m}(x,X_j)$$

so that on R $K_m(x,X_j) \overset{m}{\to} K_{cw}(x,X_j)$ for each j. Each $K_{i,m}$ is a symmetric kernel with positive mass over $h_{1;i,m}[X_j] \subsetneq h_1[X_j]$ and negative mass $-U$ over $h_{2;i,m}[X_j] \subsetneq h_2[X_j]$, with $\sum_{i=1}^m a_{i,m} = 1$. Let $\hat{f}_m(x) = (1/n)\sum_{j=1}^n K_m(x,X_j)$, so that $\hat{f}_m(x) \overset{m}{\to} \hat{f}_{cw}(x)$ on R. By Corollary 2.2, $E_n\{\hat{f}_m(x)\} = E_X\{K_m(x,X)\} = \sum_{i=1}^m a_{i,m}\bar{f}_{1;i,m}(x) + U\sum_{i=1}^m a_{i,m}[\bar{f}_{1;i,m}(x) - \bar{f}_{2;i,m}(x)]$. Also, by construction, $(1/n)\sum_{j=1}^n \sum_{i=1}^m a_{i,m}(1/h_{1;i,m})1_{h_{1;i,m}[X_j]}(x) \overset{m}{\to} (1/n)\sum_{j=1}^n K_c(x,X_j) = \hat{f}_c(x)$. Since f is bounded, the Lebesgue Dominated Convergence Theorem, Royden (1963), can be used twice to obtain

$$E_n\{\hat{f}_{cw}(x)\} = E_n\{\hat{f}_c(x)\} + U \lim_m \left\{ \sum_{i=1}^m a_{i,m}[\bar{f}_{1;i,m}(x) - \bar{f}_{2;i,m}(x)] \right\}. \tag{3.2}$$

The limit in Equation (3.2) exists and is denoted $d(x,n)$, for any $U > 0$ and any $n \geq 1$. The sum in Equation (3.2) is a weighted average of differences that, by Corollary 2.1, are positive on $CD(h)$ and negative on $CU(h)$. Take $D(x,n) = |d(x,n)|$. □

Thus, symmetrically appended negative kernel mass rather generally reverses the direction of bias on C_h, for any (U,n), simply because the sign of $-U$ opposes the sign of the symmetric positive mass in K_{cw}. For f in C^4 and nontrivial CU and CD, as has been assumed here, the bias reversal process on C_h amounts to an expected estimator mass shift from CU to CD, since $\int_R E_n\{\hat{f}(x)\}dx = 1$, when $C_h \cong S(f)$.

4. Induced mean kernel asymmetry in the (K_r, K_w) case

When the use of negative kernel mass reduces MISE, the bias reversal due to $-U$ is regarded as the result of some U-induced automatic kernel adaptation. To examine this fundamental bias reversing mechanism, for n observations $\{X, Y_1, Y_2, \ldots, Y_{n-1}\}$, the mean impact of the negative masses in $\{K_w(x, Y_j)\}_1^{n-1}$ on the shape of the positive part of $K_w(x, X)$ will be analyzed when:

(a) U has a value such that K_w is of order $p = 4$.
(b) Both K_w and K_r are on MISE-optimal bandwidths.
(c) MISE$(K_w) <$ MISE(K_r).
(d) The positive part of $K_w(x, X)$ is equal to $(1 + U)K_r(x, X)$.

Let the positive part of K_w be denoted K_w^+. Item (d) above requires that the adaptation process be viewed when the modification of K_w^+ is equivalent (as will be shown) to a modification of K_r. So, item (d) ensures that U adapts K_r shape in the simplest adaptation scenario.

The arguments in this section are confined to rectangular kernels of order $p = \{2, 4\}$ because they are relatively easy to analyze. Regarding more practical kernels, such as the Epanechnikov kernel $K_e(x, X)$ and the MISE-optimally shaped kernel of order $p = 4$, denoted $K_4(x, X)$, cf. Gasser and Müller (1979), there is a serious complication: K_4^+ cannot be superimposed on $(1 + U)K_e$ since K_4^+ is a quartic in $x - X$ while K_e is a quadratic in $x - X$. For the case (K_e, K_4), one cannot attain the analog of item (d). We proceed, then, with the case (K_r, K_w).

DEFINITION 4.1. For n observations $\{X, Y_1, Y_2, \ldots, Y_{n-1}\}$, $K_v(x, X)$ is defined at x in $h_1[X]$ as the positive part of $K_w(x, X)$ modified by the negative parts of $K_w(x, Y_j)$. Of the $n-1$ Y_j values, N_1 lie in $h_1[x]$ and are assumed for sufficiently small h_1 to equally absorb the negative contributions of the N_2 $K_w(x, Y_j)$ kernels for Y_j that lie in $h_2[x]$. $K_v(x, X)$ is written as

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

$K_v(x, X)$ is a diminished version of $K_w^+(x, X)$. Given n and fixed but arbitrary X, $N_2(x)$ and $N_1(x)$ are random variables with mean values $(n - 1)\bar{f}_2(x)h_2$ and $(n - 1)\bar{f}_1(x)h_1$, respectively, by the binomial law. The expected or mean shape of $K_v(x, X)$, considering the random Y_j, is

$$\begin{aligned} \bar{K}_v(x, X) &= E_{n-1}\{K_v(x, X)\} \\ &= 1_{h_1[X]}(x)\left[k_1(U) + k_2(U)E_{n-1}\left\{\frac{N_2(x)}{N_1(x) + 1}\right\}\right]. \end{aligned} \tag{4.1}$$

For small h and (as will become the case after Proposition 4.3) for $h_1(n)$ such that $nh_1(n) \xrightarrow{n} \infty$, a crude large n approximation of the expectation of the ratio $N_2(x)/(N_1(x) + 1)$ is h_2/h_1. Subsequently, using Definition 2.3, $\bar{K}_v(x, X) \cong K_r(x, X)$, so that prior to Gajek truncation, $\bar{K}_v$ is approximately a kernel.

Following Definition 4.1, the mean shape of $K_v(x, X)$, accounting probabilistically for the variety of Y_j observation locations about X, is taken to measure the automatic kernel adaptation performed by $-U$ upon $K_w^+(x,X)$. So, we next compute $\bar{K}_v(x,X)$ in Equation (4.1) exactly.

PROPOSITION 4.1. *For sufficiently small h_1 and for any X in C_h, the mean value of $K_v(x,X)$ at x in $h_1[X]$ is*

$$\bar{K}_v(x,X) = 1_{h_1[X]}(x)[k_1(U) + k_2(U)H(x,n)], \tag{4.2}$$

where

$$\begin{aligned} H(x,n) &= \bar{f}_2 h_2 S_1(x) \\ S_1(x) &= 1 + (1-\bar{f}_1 h_1) + (1-\bar{f}_1 h_1)^2 + \cdots + (1-\bar{f}_1 h_1)^{n-2}. \end{aligned} \tag{4.3}$$

PROOF. Referring to Definition 4.1,

$$0 < H(x,n) = \sum_{(N_1,N_2)} (N_2/(N_1+1))C(N_1,N_2)\Pr(x,N_1,N_2),$$

where $C(N_1,N_2)$ counts observation combinations and permutations and

$$\Pr(x,N_1,N_2) = (\bar{f}_1(x)h_1)^{N_1}(\bar{f}_2(x)h_2)^{N_2}(1-\bar{f}_h h)^{n-1-(N_1+N_2)}$$

is the probability that N_1 observations lie in $h_1[x]$ and N_2 observations lie in $h_2[x]$, using $\bar{f}_h h = \bar{f}_1 h_1 + \bar{f}_2 h_2$. As an example, for $n=5$, the values of $C(N_1,N_2)$ used to construct $H(x,n)$ are presented in Table 1.

Table 1. Values of $C(N_1,N_2)$ for constructing $H(x,n)$ with $n=5$.

N_1	N_2	$n-1-(N_1+N_2)$	$N_2/(N_1+1)$	$C(N_1,N_2)$
0	0	4	0	1
0	4	0	4	1
4	0	0	0	1
0	1	3	1	$4=\binom{4}{3}$
1	0	3	0	4
1	3	0	3/2	4
0	3	1	3	4
3	1	0	1/4	4
3	0	1	0	4
0	2	2	2	$6=\binom{4}{2}$
2	0	2	0	6
2	2	0	2/3	6
1	1	2	1/2	$12=\binom{4}{2}\binom{2}{1}$
1	2	1	1	12
2	1	1	1/3	12

For $2 \leqslant n \leqslant 5$, $H(x,n)$ is written as follows:

$$\begin{aligned} n=2 &: H(x,n) = \bar{f}_2 h_2[1] \\ n=3 &: H(x,n) = \bar{f}_2 h_2[2 - \bar{f}_1 h_1] \\ n=4 &: H(x,n) = \bar{f}_2 h_2[3 - 3\bar{f}_1 h_1 + (\bar{f}_1 h_1)^2] \\ n=5 &: H(x,n) = \bar{f}_2 h_2[4 - 6\bar{f}_1 h_1 + 4(\bar{f}_1 h_1)^2 - (\bar{f}_1 h_1)^3]. \end{aligned}$$

Equation (4.3) follows by induction, with

$$0 < (1 - \bar{f}_1 h_1)^{n-2} < \cdots < (1 - \bar{f}_1 h_1)^2 < (1 - \bar{f}_1 h_1) < 1. \quad (4.4)$$ □

DEFINITION 4.2. Let $h = \beta h_1$, where $\beta > 1$ is a parameter.

PROPOSITION 4.2. *$K_w(x, X)$ is of order $p = 4$ if and only if $U = U^*(\beta) = 1/(\beta(\beta+1)) < 0.5$.*

PROOF. Set

$$\int_{-1}^{1} \tilde{K}_w(v) v^2 dv = 0.$$ □

DEFINITION 4.3. For given (n, β) and using $U = U^*(\beta)$, $h^*(n, \beta)$ denotes the bandwidth for K_w that minimizes MISE for f, yielding MISE(K_w^*). For given n, $h_1^*(n)$ denotes the bandwidth for K_r that minimizes MISE for f, yielding MISE(K_r^*).

Next, it's shown that items (c) and (d) above can be accomplished simultaneously, making appropriate use of the parameter β.

PROPOSITION 4.3. *For any fixed large $n = n_0$, there exists fixed large $\beta = \beta_0$ such that $h^*(n_0, \beta_0) = \beta_0 h_1^*(n_0)$ while MISE$(K_w^*) = O(n^{-8/9})$ and MISE$(K_r^*) = O(n^{-4/5})$.*

PROOF. For the kernels K_r and K_w, VAR(x) is computed as

$$\begin{aligned} \mathrm{VAR}_r(x) &= (1/n)[\bar{f}_1/h_1 - \bar{f}_1^2] \\ \mathrm{VAR}_w(x) &= (1/n)[k_1^2 \bar{f}_1 h_1 + k_2^2 \bar{f}_2 h_2 - (\bar{f}_1 h_1 k_1 + \bar{f}_2 h_2 k_2)^2]. \end{aligned}$$

Since $h_1^*(n) \xrightarrow{n} 0$ and $h^*(n,\beta) \xrightarrow{n} 0$, Parzen (1962), and since $U^*(\beta) \to 0$ as $\beta \to \infty$ by Proposition 4.2, large (n, β) approximations of VAR(x), using Definition 2.3, are $\mathrm{VAR}_r(x) \cong f(x)/nh_1$ and $\mathrm{VAR}_w(x) \cong f(x)\beta/nh$. So, using equation (3.1) for large (n, β), for $p(K_w) = 4$,

$$\begin{aligned} \mathrm{MISE}(K_r) &\cong 1/nh_1 + [M_2(K_r)(h_1/2)^2 1/2!]^2 \int_R [f^{(2)}(x)]^2 dx \\ \mathrm{MISE}(K_w) &\cong \beta/nh + [M_4(K_w)(h/2)^4 1/4!]^2 \int_R [f^{(4)}(x)]^2 dx. \end{aligned}$$

Then, using Lemma 4a of Parzen (1962), large (n, β) approximations of the MISE-optimal bandwidths for K_w (with $U = U^*(\beta)$) and K_r are

$$h^*(n,\beta) \cong \left\{(\beta/n) \div \left[8(M_4(K_w)/4!)^2(1/2)^8 \int_R [f^{(4)}(x)]^2 dx\right]\right\}^{1/9}$$

$$h_1^*(n) \cong \left\{(1/n) \div \left[4(M_2(K_r)/2!)^2(1/2)^4 \int_R [f^{(2)}(x)]^2 dx\right]\right\}^{1/5},$$

where $M_2(K_r) = 1/3$ and $M_4(K_w) \cong -1/5\beta^2$. Thus, for large (n, β),

$$h^*(n,\beta)/h_1^*(n) \cong c_1\beta^{5/9}n^{4/45} \xrightarrow{n} \infty. \tag{4.5}$$

For a fixed large n, denoted n_0, the objective is to determine a particular large β, denoted β_0, such that

$$h^*(n_0,\beta_0)/h_1^*(n_0) = \beta_0. \tag{4.6}$$

For β satisfying $c_1 < \beta^{4/9}$, the objective is achieved by taking

$$\beta_0 = (c_2c_1n_0^{4/45})^{9/4}, \tag{4.7}$$

where $c_2 \cong 1$ makes equation (4.5) exact. The parameter β would then be the fixed value β_0 for all n. For the particular $n = n_0$, equation (4.6) is satisfied. Since β is thus fixed as a function of n_0, $h^*(n,\beta_0) = O(n^{-1/9})$, $h_1^*(n) = O(n^{-1/5})$, $\text{MISE}(K_w^*) = O(n^{-8/9})$ and $\text{MISE}(K_r^*) = O(n^{-4/5})$. □

PROPOSITION 4.4. *The (n_0, β_0) in Proposition 4.3 can be taken large enough so that* $\text{MISE}(K_w^*) < \text{MISE}(K_r^*)$.

PROOF. From the equations for MISE in the prior proof, when $(n, \beta) = (n_0, \beta_0)$, the integrated $\text{VAR}(x)$ terms for estimators based on K_w^* and K_r^* are asymptotically identical by equation (4.6). Using Equation (4.7) and the equations for h^* and h_1^* in the prior proof, the integrated $B^2(x)$ terms for estimators based on K_w^* and K_r^*, denoted respectively as $IB^2(K_w^*)$ and $IB^2(K_r^*)$, compare as $IB^2(K_w^*)/IB^2(K_r^*) \cong c_0n_0^{-26/45} \xrightarrow{n_0} 0$. □

DEFINITION 4.4. Denote the (n_0, β_0) in Proposition 4.4 as (n_0^*, β_0^*).

The use of (n_0^*, β_0^*) ensures items (c) and (d) above since equation (4.6) implies that $(K_w^*)^+ = (1+U)K_r^*$. The use of (n_0^*, β_0^*) will provide convenient glimpses of the two optimal rectangular kernels K_w^* and K_r^*. Next, it is shown that the mean U-induced adaptation of K_r^* involves an asymmetric shape with a particular asymmetric shift in K_r^* mass.

PROPOSITION 4.5. *Let X be in C_h, with h sufficiently small so that the averages $\bar{f}_1(x)$ and $\bar{f}_2(x)$ differ approximately by a constant over x in $h_1[X]$, with*

$$\bar{f}_2(x) \cong \bar{f}_1(x) + c_3 \quad \textit{on} \quad h_1[X],$$

where c_3 is an (X, n, β)-dependent constant. Then,

(i) *For any X in $CD(h)$, $\bar{K}_v'(x, X)$ is opposite in sign from $f'(x)$, for any $h_1 < h$.*

(ii) *For any X in $CU(h)$, $\bar{K}_v'(x, X)$ has the same sign as $f'(x)$ for $h_1 = h_1^*(n)$ and $(n, \beta) = (n_0^*, \beta_0^*)$, for sufficiently large n_0^*.*

PROOF. From Proposition 2.1, $c_3 > 0$ on CU and $c_3 < 0$ on CD. From Proposition 4.1,

$$\bar{K}_v'(x, X) = -U1_{h_1[X]}(x)[H'(x, n)/h_2],$$

where $H(x, n)/h_2 \cong [\bar{f}_1(x) + c_3]S_1(x)$ and

$$d/dx\{\bar{f}_1(x)S_1(x)\} = \bar{f}_1'(x)(1 - \bar{f}_1 h_1)^{n-2}(n - 1) \tag{4.8}$$

$$c_3 d/dx\{S_1(x)\} = -c_3 h_1 \bar{f}_1'(x) S_2(x), \tag{4.9}$$

with $S_2(x) = 1 + 2(1 - \bar{f}_1 h_1) + \cdots + (n - 2)(1 - \bar{f}_1 h_1)^{n-3}$. Aside from $\bar{f}_1'(x)$, the terms in equations (4.8) and (4.9) are positive for X in CD, proving part (i). For X in CU, let

$$\begin{aligned} S_2(x) &= [1 + 2 + 3 + \cdots + (n - 2)]G(x, n) \\ &= [(n - 2)(n - 1)/2]G(x, n); \qquad G(x, n) > 0. \end{aligned}$$

Then for X in CU,

$$H'(x, n)/h_2 \cong \bar{f}_1'(x)G(x, n)(n - 1)\left[\frac{(1 - \bar{f}_1 h_1)^{n-2}}{G(x, n)} - c_3 h_1 \left(\frac{n - 2}{2}\right)\right]. \tag{4.10}$$

$G(x, n)$ is a weighted average of terms in inequality (4.4) such that $(1 - \bar{f}_1 h_1)^{n-3} < G(x, n) < 1$. Thus,

$$0 \leq \lim_{n \to \infty} (1 - \bar{f}_1 h_1)^{n-2}/G(x, n) \leq 1. \tag{4.11}$$

Considering the second term in brackets in equation (4.10),

$$\begin{aligned} 0 < c_3 h_1 &\cong h_1[\bar{f}_2(x) - \bar{f}_1(x)] = h_1[\bar{f}_h(x) - \bar{f}_1(x)](\beta/\beta - 1) \\ &= (\beta/\beta - 1)\left[(1/\beta)\int_{h[x]} f(y)dy - \int_{h_1[x]} f(y)dy\right]. \end{aligned}$$

Taylor's formula for some x_0 in $h[x]$ yields

$$c_3 h_1 \cong (1/24)\beta(\beta - 1)f''(x_0)h_1^3.$$

Using $h_1 = h_1^*(n)$ and $(n, \beta) = (n_0^*, \beta_0^*)$,

$$\begin{aligned} c^3 h_1^* n_0^* &\cong (\beta_0^*)^2 f''(x_0)(n_0^*)^{2/5} c_4 \\ &= c_5(x_0)(n_0^*)^{4/5} \stackrel{n_0^*}{\to} \infty. \end{aligned} \tag{4.12}$$

Equations (4.9)–(4.12) yield part (ii). □

COROLLARY 4.1. *For X in C_h and (n, β) sufficiently large, let K_w and K_r both lie on* MISE-*optimal bandwidths, with $h^*(n,\beta) = \beta h_1^*(n)$, $U = U^*(\beta)$ and* $\text{MISE}(K_w^*) < \text{MISE}(K_r^*)$. *Then the mean U-induced adaptation in the shape of the kernel $K_r^*(x, X)$ shifts K_r^* mass across X in the direction of decreasing $|f''(x)|$.*

PROOF. U acts twice to adapt $K_r^*(x, X)$, with $U = U^*$. U first adapts $K_r^*(x, X)$ by its mere appendage, elevating $K_r^*(x, X)$ uniformly over $h_1^*[X]$ to become $[K_w^*(x, X)]^+ = (1 + U^*)K_r^*(x, X)$. Secondly, U diminishes $[K_w^*(x, X)]^+$ to become $\bar{K}_v(x, X)$, by the mean action of random observations near X. Analytically, these two adaptations are written using Definition 2.3 and equation (4.2) as

$$\begin{aligned} \bar{K}_v(x, X) &= K_r^*(x, X) + U^* 1_{h_1^*[X]}(x)[1/h_1^* - H(x, n)/h_2] \\ &= [K_w^*(x, X)]^+ - U^* 1_{h_1[X]}(x)[H(x, n)/h_2]. \end{aligned}$$

Gajek truncation is accomplished by forming the estimator

$$\tilde{f}(x) = \text{Max}\{0, \hat{f}_w(x) - a\}$$

for some constant $a > 0$, and so does not affect mean adapted shape. The result in Proposition 4.5 then geometrically says that the mean U-adapted shape of $K_r^*(x, X)$, for X in C_h, is asymmetric and slopes toward the nearest inflection point of f that one would encounter without passing through a point where $f'(\cdot) = 0$. □

In light of the behavior in equations (4.11) and (4.12), it is emphasized that the large (n, β) conditions in Corollary 4.1 may not be necessary to induce the particular mean asymmetrical mass shift in K_r^*. For example, large n is also required for reduced $B(x)$ to offset increased $\text{VAR}(x)$ on C_h since

$$\text{VAR}_w(x) - \text{VAR}_r(x) = (1/h_1 n)[A(x)U^2 + 2I(x)\bar{f}_1(x)U],$$

where for large $n : A(x) \cong \bar{f}_1(x)$ and $I(x) \cong 1$.

It has been assumed that f is strictly curvilinear since bias and bias correction would both be zero on intervals where $f''(\cdot) = 0$. Roughly speaking, then, the bias is zero at inflection points of f and there is no adaptation at inflection points since the U-induced asymmetry vanishes at such points.

So, in the case of rectangular kernels, under conditions (a)–(d) and for sufficiently large n, the use of negative kernel mass destroys the original uniform kernel symmetry in order to achieve reduced bias and MISE, for X in $C_h \cong S(f)$. In this case, the U-induced asymmetry affects the mean shape of the adapted kernel. The U also increases variance. One must wonder if it is possible, especially for smaller n, to improve estimator performance by incorporating estimates of f'' in a recursive scheme using properly skewed rather simple nonnegative kernels.

$U^*(n)$ and the ratio $h_1^*(n)/h^*(n)$ both asymptotically go to zero in the (K_r^*, K_w^*) case, while their analogs in the (K_e, K_4) case are fixed. Thus, the adaptation mechanism by which $-U$ reduces bias and MISE may be different in the two cases. Still, it is conjectured that the efficacy of higher order kernels for sufficiently large finite n is based on the ability of negative kernel mass to induce an appropriate kernel asymmetry.

REFERENCES

Berlinet, Alain. (1993). Hierarchies of higher order kernels, *Probab. Theory Related Fields*, **94**, 489–504.

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

Gasser, T. and Müller, H-G. (1979). Kernel estimation of regression functions, *Lecture Notes in Mathematics*, **757**, 23–68, Springer, New York.

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

Royden, H. L. (1963). *Real Analysis*, Macmillan, New York.

Schucany, W. R. and Sommers, J. P. (1977). Improvement of kernel type density estimators, *J. Amer. Statist. Assoc.*, **72**, 420–423.