Journal of Statistical Planning and Inference 66 (1998) 289–304

journal of statistical planning and inference

Nonparametric kernel regression estimation near endpoints

Cha Kyung-Joon[a,*], William R. Schucany[b]

[a]Department of Mathematics, Hanyang University, Seongdong-Ku, Haengdang-Dong, Seoul, 133-791, South Korea

[b]Department of Statistical Science, Southern Methodist University, Dallas TX 75275, USA

Received 15 November 1995

Abstract

When kernel regression is used to produce a smooth estimate of a curve over a finite interval, boundary problems detract from the global performance of the estimator. A new kernel is derived to reduce this boundary problem. A generalized jackknife combination of two unsatisfactory kernels produces the desired result. One motivation for adopting a jackknife combination is that they are simple to construct and evaluate. Furthermore, as in other settings, the bias reduction property need not cause an inordinate increase in variability. The convergence rate with the new boundary kernel is the same as for the non-boundary. To illustrate the general approach, a new second-order boundary kernel, which is continuously linked to the Epanechnikov (1969, Theory Probab. Appl. 14, 153–158) kernel, is produced. The asymptotic mean square efficiencies relative to smooth optimal kernels due to Gasser and Müller (1984, Scand. J. Statist. 11, 171–185), Müller (1991, Biometrika 78, 521–530) and Müller and Wang (1994, Biometrics 50, 61–76) indicate that the new kernel is also competitive in this sense. © 1998 Elsevier Science B.V. All rights reserved.

AMS classification: Primary 62G07; secondary 62G20

Keywords: Bandwidth; Boundary; Convergence rate; Mean square error

1. Introduction

Consider the nonparametric regression model,

$$y_i = m(t_i) + \varepsilon_i, \quad i = 1, 2, \ldots, n,$$

where the ε_i are independent, identically distributed random variables with zero mean and common variance σ^2, m is an unknown regression function in $C^p[0, 1]$ for some integer $p \geqslant 2$ and the t_i are the nonstochastic design points satisfying

* Corresponding author.

PII S0378-3758(97)00082-7

$0 \leqslant t_1 < t_2 < \cdots < t_n \leqslant 1$. Without having to assume more about m than that it satisfies certain smoothness conditions, we may want to estimate $m(t)$ at some fixed argument t.

There are many interesting nonparametric estimators for $m(t)$. Examples of these can be found in Eubank (1988) and Gasser and Müller (1979). For simplicity consider the class of kernel estimators of $m(t)$ defined by Priestley and Chao (1972) and examined by Müller and Stadtmüller (1987),

$$\hat{m}(t; h) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{h} K\left(\frac{t - t_i}{h}\right) y_i, \tag{1.1}$$

where the design points, t_i, are equally spaced and $h > 0$ is the bandwidth or window width. The function K is called a kernel function. It is supported and symmetric on $[-1, 1]$. When it satisfies

$$\int_{-1}^{+1} z^j K(z)\,\mathrm{d}z = \begin{cases} 1 & j = 0 \\ 0 & j = 1, 2, \ldots, p-1 \\ k_p \neq 0 & j = p, \end{cases} \tag{1.2}$$

it is called a kernel of order p. The various other classes of kernel estimators are not treated in this paper. Some discussion of parallel developments is in Section 6.

If the bandwidth depends on t (or t_i), the estimator is called a local-bandwidth kernel estimator, otherwise it is called a global-bandwidth kernel estimator. Next, suppose that for a specific point t, the value of h is fixed in (1.1). If some part of the support of the kernel, $[t - h, t + h]$, is not contained in $[0, 1]$, then t is said to be in the boundary. Gasser and Müller (1979) identified the unsatisfactory behavior of (1.1) for points in the boundary.

The goal of this paper is to find a kernel that is suitable for local estimation in boundary cases. In Section 2, we review the asymptotic optimality considerations which lead to balancing variance and bias squared. In Section 3, we describe a general combination that produces a boundary kernel with the desired characteristics. This is illustrated in Section 4, compared with other boundary kernels such as Gasser and Müller (1979), Müller (1991) and Müller and Wang (1994) in Section 5, and some concluding remarks are made in the final section regarding local linear fitting as an alternative boundary adaptation.

2. Nonboundary kernel regression estimators

Let us examine the first moment of the kernel estimator (1.1) for the nonboundary case. Using a standard Taylor series argument (see Gasser and Müller 1979), it can be shown that if $m \in C^p[0, 1]$, then the expected value of $\hat{m}$ at a fixed t is

$$\begin{aligned} E[\hat{m}(t; h)] = \int_{-1}^{+1} K(z) \Big\{ & m(t) - zhm^{(1)}(t) + \frac{(zh)^2}{2!} m^{(2)}(t) \\ & + \cdots + \frac{(-1)^p}{p!} (zh)^p m^{(p)}(t) \Big\} \mathrm{d}z + \mathrm{o}(h^p) + \mathrm{O}\left(\frac{1}{n}\right), \end{aligned}$$

where $m^{(j)}(t)$ is the jth derivative of $m(t)$ and h is sufficiently small that $h < t$ and $h < 1 - t$. Hence, as a result of (1.2) the asymptotic bias of $\hat{m}(t; h)$ is

$$E[\hat{m}(t; h)] - m(t) = \frac{(-1)^p}{p!} h^p m^{(p)}(t) k_p + o(h^p). \tag{2.1}$$

Next, the asymptotic variance of $\hat{m}(t; h)$, by methods similar to those leading to (2.1), is

$$\operatorname{Var}[\hat{m}(t; h)] = \frac{\sigma^2}{nh} \int_{-1}^{+1} K^2(z)\, dz + o\left(\frac{1}{nh}\right). \tag{2.2}$$

Therefore, the mean square error can be expressed as

$$\operatorname{mse}[\hat{m}(t; h)] = \frac{\sigma^2}{nh} Q + \left[\frac{h^p}{p!} m^{(p)}(t) k_p\right]^2 + o\left(\frac{1}{nh}\right) + o(h^{2p}), \tag{2.3}$$

where $Q = \int_{-1}^{+1} K^2(z)\, dz$. Hence, the optimal bandwidth which minimizes the asymptotic $\operatorname{mse}[\hat{m}(t; h)]$ is

$$h_{\mathrm{opt}} = \left\{\frac{\sigma^2 Q}{2pn(k_p m^{(p)}(t)/p!)^2}\right\}^{1/(2p+1)}. \tag{2.4}$$

3. Boundary kernel regression estimators

Nonparametric regression function estimators usually show a sharp increase in variance and bias when estimating $m(\cdot)$ at points near the boundary of the support of the function (e.g., $t < h$). Gasser and Müller (1979, 1984) identified the crucial nature of these effects. They proposed optimal boundary kernels but did not give any formulas. However, Gasser and Müller (1979) and Müller (1988) suggested multiplying the truncated kernel at the boundary by a linear function. Rice (1984) proposed another approach using a generalized jackknife, also known as Richardson extrapolation which linearly combines the two bandwidths. Schuster (1985) introduced a reflection technique for density estimation. Eubank and Speckman (1991) have given a method for removing boundary effects using a 'bias reduction theorem'. The fundamental idea of their work is to use a biased estimator to improve another estimator in some sense. Müller (1991) proposed an explicit construction for a boundary kernel which is the solution of a variational problem under asymmetric support. He tables many polynomials that are optimal in a specified sense. Moreover, Müller (1993a) introduced a general method of constructing a boundary kernel which is the solution of a variational problem involving a certain weight function. More recently, Müller and Wang (1994) gave explicit formulas for a new class of polynomial boundary kernels and showed that these new boundary kernels have some advantages over the smooth optimum boundary kernels in Müller (1991), i.e., these new kernels have higher mse efficiency. Some of the methods discussed above are investigated further and compared with the proposed boundary kernel in Section 5.

In the context of density estimation, Wand and Schucany (1990) and Berlinet (1993) worked with the Gaussian kernel which exhibits a first-order boundary effect because the Gaussian kernel has noncompact support. In fact, Berlinet (1993) proposed a framework for building kernels of increasing order apart from some specific methods based on moment relationships.

Now, let us assume that conditions (1.2) hold for the class of kernels with $p = 2$. In this section, a new boundary kernel is derived by combining two kernels with inferior bias properties. The bias of the proposed boundary kernel estimator has the same convergence rate as in the interior. Therefore, the best rate of convergence for mean square error of this boundary kernel is the same as that of the non-boundary kernel. To do this, a combination of 'cut-and-normalize' and the generalized jackknife methods is used so that the bias is improved to $O(h^2)$ throughout the boundary.

A method of cut-and-normalize was first introduced by Gasser and Müller (1979). Schucany and Sommers (1977) used a generalized jackknife method for the construction of a higher-order kernel. Wand and Schucany (1990) adapted this to a class of kernels based on the Gaussian kernel.

For a fixed t and bandwidth h, define an index of how much of the window remains within the set of design points

$$q = \min\left\{\frac{t}{h}, 1\right\} \quad \text{if } t \in [0, \tfrac{1}{2})$$

and

$$q = \min\left\{\frac{1-t}{h}, 1\right\} \quad \text{if } t \in (\tfrac{1}{2}, 1],$$

where q is a real number such that $q \in [0, 1]$. For $q = t/h < 1$, the support of the kernel estimator is less than $t \pm h$. The effective domain of the kernel is $[-1, q)$ instead of $[-1, 1]$ as for an interior point. Hence, the 'cut-and-normalize' modification omits that part of kernel lying between q and 1 and renormalizes the kernel between -1 and q. The result is a boundary kernel. Then, the specific linear combination of two different such boundary kernels, that is the required generalized jackknife estimator, gives a bias that has the same order as in the interior.

3.1. Cut and normalize

For simplicity, only the left boundary effects, i.e., $q = t/h < 1$, will be discussed here. The right boundary effects proceed in the same manner. Since Gasser and Müller (1979) investigated the cut-and-normalize method, we briefly explain the general approach described above. Let $\hat{m}_1(t; h)$ is a kernel estimator with second-order kernel

$K_1(z)$ on $[-1, 1]$. If t is a fixed value in $[0, h)$, then it is in the left boundary. Thus, the expectation of $\hat{m}_1(t; h)$ becomes

$$E[\hat{m}_1(t; h)] = \int_{-1}^{q} K_1(z)\left\{m(t) - zhm^{(1)}(t) + \frac{(zh)^2}{2!} m^{(2)}(t)\right\} \mathrm{d}z + \mathrm{o}(h^2), \tag{3.1}$$

where $q = t/h$. Since $t \in [0, h)$, the symmetry of the kernel is lost and $\int_{-1}^{q} K_1(z)\,\mathrm{d}z \neq 1$ and $\int_{-1}^{q} zK_1(z)\,\mathrm{d}z \neq 0$. Hence, no terms vanish and (3.1) cannot be reduced further.

Therefore, a boundary kernel modification of $\hat{m}_1(t; h)$ is

$$\hat{m}_{1q}(t; h) = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{h} K_{1q}\left(\frac{t - t_i}{h}\right) y_i,$$

where

$$K_{1q}(z) = \frac{K_1(z)}{\int_{-1}^{q} K_1(u)\,\mathrm{d}u}, \qquad -1 \leqslant z \leqslant q,$$

by 'cutting'.

Further, this is 'normalized' in the sense that it is rescaled to integrate to 1.0. Then, the corresponding expectation is

$$E[\hat{m}_{1q}(t; h)] = m(t) - hm^{(1)}(t)k_{1q}^{(1)} + \frac{h^2 m^{(2)}(t)}{2!} k_{1q}^{(2)} + \mathrm{o}(h^2), \tag{3.2}$$

where

$$k_{1q}^{(1)} = \int_{-1}^{q} zK_{1q}(z)\,\mathrm{d}z \neq 0 \quad \text{and} \quad k_{1q}^{(2)} = \int_{-1}^{q} z^2 K_{1q}(z)\,\mathrm{d}z.$$

From (2.1), the dominant part of bias$[\hat{m}_{1q}(t; h)]$ for the nonboundary is of order h^2 with $p = 2$. However, the dominant part of bias$[\hat{m}(t; h)]$ in (3.2) is of order h, so $\hat{m}_{1q}(t; h)$ is still subject to more boundary bias.

The asymptotic variance of $\hat{m}_{1q}(t; h)$ can be obtained by the same method as (2.2) for the nonboundary, i.e.,

$$\mathrm{var}[\hat{m}_{1q}(t; h)] = \frac{\sigma^2}{nh}\int_{-1}^{q} K_{1q}^2(z)\,\mathrm{d}z + \mathrm{o}\left(\frac{1}{nh}\right).$$

Hence, the asymptotic mean square error has the form

$$\mathrm{amse}[\hat{m}_{1q}(t; h)] = \frac{\sigma^2 Q_1}{nh} + [hm^{(1)}(t)k_{1q}^{(1)}]^2,$$

where $Q_1 = \int_{-1}^{q} K_{1q}^2(z)\,\mathrm{d}z$. Therefore, the best rate of convergence to be anticipated for the amse$[\hat{m}_{1q}(t; h)]$ is $n^{-2/3}$. This can be compared to $n^{-4/5}$ that would be obtained with $p = 2$ if there were no boundary effect. Next, to obtain the same local asymptotic behavior, the generalized jackknife method is applied to reduce the order of the bias.

3.2. Generalized jackknife

To apply this method, first define another kernel estimator as

$$\hat{m}_2(t;h) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h}K_2\left(\frac{t-t_i}{h}\right)y_i,$$

where $K_2(z)$ is a second-order kernel function supported on $[-1,1]$ and not identically equal to $K_1(z)$. Now define another boundary kernel estimator $\hat{m}_{2q}(t;h)$ with a similarly renormalized kernel, $K_{2q}(z)$. Hence, similar to (3.2),

$$E[\hat{m}_{2q}(t;h)] = m(t) - hm^{(1)}(t)k_{2q}^{(1)} + \frac{h^2 m^{(2)}(t)}{2}k_{2q}^{(2)} + \mathrm{o}(h^2). \tag{3.3}$$

Then, obviously, (3.2) and (3.3) can be rewritten as

$$E[\hat{m}_{1q}(t;h)] = m(t) - hm^{(1)}(t)k_{1q}^{(1)} + \mathrm{O}(h^2)$$

$$E[\hat{m}_{2q}(t;h)] = m(t) - hm^{(1)}(t)k_{2q}^{(1)} + \mathrm{O}(h^2).$$

The generalized jackknife principle formally expresses a solution of a related linear system of equations

$$K_q^*(z) = \frac{\begin{vmatrix} K_{1q}(z) & K_{2q}(z) \\ k_{1q}^{(1)} & k_{2q}^{(1)} \end{vmatrix}}{\begin{vmatrix} 1 & 1 \\ k_{1q}^{(1)} & k_{2q}^{(1)} \end{vmatrix}} = \frac{K_{1q}(z) - rK_{2q}(z)}{1-r}, \qquad -1 \leqslant z \leqslant q,$$

where $r = k_{1q}^{(1)}/k_{2q}^{(1)} \neq 1$. Notice that $K_q^*(z)$ is a linear combination of $K_{1q}(z)$ and $K_{2q}(z)$, see Gray and Schucany (1972).

Finally, define the boundary kernel estimator, $\hat{m}_q^*(t;h)$, whose kernel function is $K_q^*(z)$ as

$$\hat{m}_q^*(t;h) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h}K_q^*\left(\frac{t-t_i}{h}\right)y_i. \tag{3.4}$$

Using (3.2) and (3.3), the leading terms cancel and expansions similar to those leading to (2.1) yield

$$E[\hat{m}_q^*(t;h)] = m(t) + \frac{h^2 m^{(2)}(t)}{2}k_q^{2*} + \mathrm{o}(h^2),$$

where $k_q^{2*} = \int_{-1}^{q} z^2 K_q^*(z)\,\mathrm{d}z$. Hence, the dominant part of the asymptotic bias of the estimator, $\hat{m}_q^*(t;h)$, is of order h^2, the same order as in the interior. Similarly, the asymptotic variance of $\hat{m}_q^*(t;h)$ is

$$\mathrm{var}[\hat{m}_q^*(t;h)] = \frac{\sigma^2 Q^*}{nh} + \mathrm{o}\left(\frac{1}{nh}\right),$$

where $Q^* = \int_{-1}^{q} [K_q^*(z)]^2 \, dz$. Hence, the asymptotic mean square error of $\hat{m}_q^*(t; h)$ is

$$\operatorname{mse}[\hat{m}_q^*(t; h)] = \frac{\sigma^2 Q^*}{nh} + \left\{\frac{k_q^{2*} m^{(2)}(t)}{2}\right\}^2 h^4 + o\left(\frac{1}{nh}\right) + o(h^4).$$

Therefore, comparing Eq. (2.3) for the nonboundary with $p = 2$, the asymptotic mean square error of $m_q^*(t; h)$ has the same optimal rate of convergence as in the interior, namely $O(n^{-4/5})$.

As we mentioned, Rice (1984) used this generalized jackknife principle to achieve the same order. From (3.1), even consistency is a problem. Thus, Rice (1984) defined a kernel function for asymptotic unbiasedness, namely $K(z;q) = K(z)/k_q^{(0)}$, where $k_b^{(i)} = \int_{-1}^{b} z^i K(z) \, dz$ which, in fact, is normalization. Asymptotic unbiasedness is achieved, but the bias is still of lower order at the boundary. Thus, he used a generalized jackknife combination, also known as Richardson extrapolation, to eliminate this lower order bias term. He defined the new kernel estimator

$$\hat{m}_q^{\mathrm{R}}(t; h) = (1 - \alpha)\hat{m}_q(t; h) + \alpha \hat{m}_q(t; \alpha h),$$

where

$$\alpha = -\frac{(k_q^{(1)}/k_q^{(0)})}{\{(\beta k_{q/\beta}^{(1)}/k_{q/\beta}^{(0)}) - (k_q^{(1)}/k_q^{(0)})\}} \quad \text{and} \quad \hat{m}_q(t; h)$$

is the kernel estimator with kernel $K(z;q)$. Also, he recommended using $\beta = 2 - q$. In fact, $\hat{m}_q^{\mathrm{R}}(t; h)$ is the jackknife estimator with kernel function $K_{\mathrm{R}}(z;q) = (1 - \alpha)K(z) - (\alpha/\beta)K(z/\alpha)$.

4. A Boundary kernel of order two matching the Epanechnikov kernel

In this section, the $p = 2$ case is demonstrated with two specific kernels. First, let $K_1(z)$ be the second-order uniform kernel, $K_1(z) = \frac{1}{2}$, $|z| \leqslant 1$ and $K_2(z)$ be the (also second-order) Epanechnikov kernel, $K_2(z) = \frac{3}{4}(1 - z^2)$, $|z| \leqslant 1$. Then, cut and normalize yields

$$K_{1_q}(z) = \frac{K_1(z)}{\int_{-1}^{q} K_1(u) \, du} = \frac{1}{1 + q}$$

and

$$K_{2_q}(z) = \frac{K_2(z)}{\int_{-1}^{q} K_2(u) \, du} = \frac{3}{(3q - q^3 + 2)}(1 - z^2), \quad -1 \leqslant z \leqslant q$$

for the boundary kernels, respectively. Then, from (3.2) and (3.3),

$$E[\hat{m}_{1_q}(t; h)] = m(t) - \frac{1}{2}(q - 1)hm^{(1)}(t) + \frac{1}{6}(q^2 - q + 1)h^2 m^{(2)}(t) + o(h^2).$$

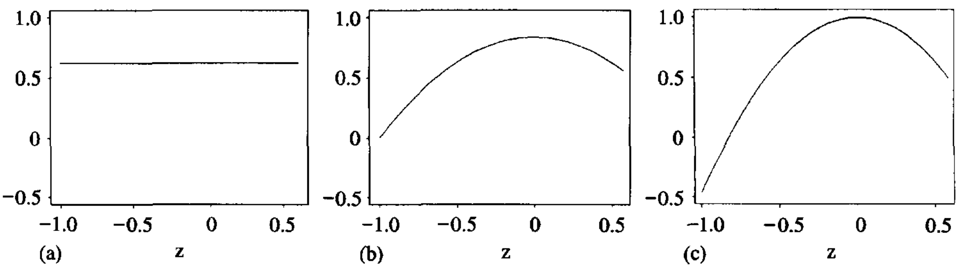
and

$$E[\hat{m}_{2q}(t;h)] = m(t) - \frac{3}{4}\frac{(q-1)^2}{(q-2)} hm^{(1)}(t) + \frac{1}{10}\frac{(3q^5-5q^3-2)}{(q-1)^2(q-2)} h^2 m^{(2)}(t) + o(h^2).$$

Hence, as in Section 3 we use the leading terms of these two bias expansions. After some algebra, the boundary kernel, $K_q^*(z)$, simplifies to

$$K_q^*(z) = \frac{\begin{vmatrix} \dfrac{1}{1+q} & \dfrac{3(1-z^2)}{(3q-q^3+2)} \\ -\dfrac{1}{2}(q-1) & -\dfrac{3}{4}\dfrac{(q-1)^2}{(q-2)} \end{vmatrix}}{\begin{vmatrix} 1 & 1 \\ -\dfrac{1}{2}(q-1) & -\dfrac{3}{4}\dfrac{(q-1)^2}{(q-2)} \end{vmatrix}} = 3\frac{(q^2+1)}{(q+1)^3} - \frac{6z^2}{(q+1)^3}, \qquad -1 \leqslant z \leqslant q, \tag{4.1}$$

for $0 \leqslant q \leqslant 1$. Figs. 1(a) and (b) show the 'cut-and-normalized' boundary kernels when $q = 0.6$. Fig. 1(c) shows the boundary kernel that is the generalized jackknife combination of Figs. 1(a) and (b). It should be noticed that $K_q^*(z)$ is a second-order boundary kernel that has the desirable property that it depends continuously upon q. Furthermore, it clearly converges to the Epanechnikov kernel as $q \to 1$. This is an attractive feature that $K_q^*(z)$ converges to the optimal kernel of order 2, as the boundary problem recedes. Furthermore, (4.1) has a simple form and achieves the best attainable rate. The remaining question is whether the constant is unacceptably large. The calculations reported in the next section confirm that it does not have serious problems in this regard.



(a) Cut-and-Normalized Kernel from Uniform Kernel
(b) Cut-and-Normalized Kernel from Epanechnikov Kernel
(c) Jackknife Boundary Kernel from Both Kernels

Fig. 1. Proposed boundary kernel when $q = 0.6$.

5. Efficiency comparison

As we discussed in Section 3, there are a few other boundary kernels. In this section, we compare these to the proposed boundary kernel.

First, Gasser and Müller (1979) and Müller (1988) proposed multiplying the truncated kernel at the boundary by a linear function. That is, consider

$$K_{\mathrm{GM}}(z;q) = (a + bz)\tfrac{3}{4}(1 - z^2), \tag{5.1}$$

where a and b depend on q such that $\int_{-1}^{q} K_{\mathrm{GM}}(z;q)\,\mathrm{d}z = 1$ and $\int_{-1}^{q} zK_{\mathrm{GM}}(z;q)\,\mathrm{d}z = 0$. Then, the explicit solutions of this for uniform and Epanechnikov kernels can be obtained. See the Appendix for details for $p = 2$.

Second, a boundary kernel of second-order with a certain smoothness characteristic is derived by Müller (1991). It also agrees with the Epanechnikov kernel in the interior. From Table 1 of Müller (1991) and Müller (1993b),

$$K_{\mathrm{M}}(z;q) = \frac{6(1+z)(q-z)}{(1+q)^3}\left\{1 + 5\left(\frac{1-q}{1+q}\right)^2 + 10\frac{(1-q)}{(1+q)^2}z\right\}. \tag{5.2}$$

Also, Müller (1993a) introduced a method which is the solution of a variational problem involving a certain weight function. If we restrict our attention to compact support with a second-order boundary kernel, we can get the same boundary kernel as (5.2) by using Theorem 4.1 and the procedure for finding boundary kernels in Section 5 of Müller (1993a).

Other boundary kernels are introduced by Müller and Wang (1994). In fact, they showed that these new boundary kernels have smaller asymptotic variance than Müller's (1991) so that these new boundary kernels have greater efficiency in terms of asymptotic mse. For second order, their boundary kernel is given by

$$K_{\mathrm{MW}}(z;q) = \frac{12}{(1+q)^4}(z+1)\left[z(1-2q) + \frac{(3q^2-2q+1)}{2}\right] \tag{5.3}$$

and this explicit formula can be found in Müller (1993b) and Müller and Wang (1994).

Table 1
ARE for $K_{\mathrm{GM}}(z;q)$, $K_{\mathrm{M}}(z;q)$, $K_{\mathrm{MW}}(z;q)$ with respect to $K_q^*(z)$

q	$\mathrm{ARE}_{\mathrm{GM}}$	$\mathrm{ARE}_{\mathrm{M}}$	$\mathrm{ARE}_{\mathrm{MW}}$
0.0	1.227	0.765	1.237
0.1	1.519	0.788	1.699
0.2	3.119	0.815	3.316
0.3	1.132	0.847	0.900
0.4	0.366	0.881	0.329
0.5	0.769	0.914	0.727
0.6	0.902	0.945	0.879
0.7	0.963	0.969	0.954
0.8	0.990	0.987	0.987
0.9	0.998	0.997	0.998
1.0	1.000	1.000	1.000

The asymptotic relative efficiency (ARE) of two different estimators does not depend on σ^2 or $m^{(2)}(t)$. To see this, note that substituting the optimal bandwidth from (2.4) with $p = 2$ into the asymptotic mean square error (amse) yields for each fixed value of q

$$\operatorname{amse}[\hat{m}(t; h_{\text{opt}})] = \frac{\sigma^2 Q}{n h_{\text{opt}}} + \frac{1}{4} h_{\text{opt}}^4 [m^{(2)}(t) k_2]^2$$

$$= \frac{5}{4} [\sigma^2 Q]^{4/5} [k_2 m^{(2)}(t)]^{2/5} n^{-4/5}.$$

Hence, in comparing two kernels by the ratio of their respective amse, the only factors that do not cancel are $Q^{4/5}(k_2)^{2/5}$. It has been argued that the sample size interpretation of ARE requires one to use $[\text{amse}]^{5/4}$, so that each is proportional to $1/n$. Therefore, the relevant ratio involves $Q|k_2|^{1/2}$ for each boundary kernel. For $K_q^*(z)$ in (4.1), direct evaluation of the relevant integrals yields

$$Q^* = \frac{36(1+q^5) - 60(1+q^2)(1+q^3) + 45(1+q)(1+q^2)^2}{5(1+q)^6}$$

and after some simplification

$$k_q^{2*} = \frac{-q^2 + 3q - 1}{5}.$$

Table 1 reports the ARE of (5.1)–(5.3) relative to the boundary kernel derived here (4.1) for eleven equally spaced values of q.

For example, ARE_M is evaluated by

$$\text{ARE}_\text{M} = \left[\frac{\text{amse of new boundary kernel}}{\text{amse of Müller (1991) boundary kernel}}\right]^{5/4}$$

$$= \frac{Q^* |k_q^{2*}|^{1/2}}{Q_\text{M} |k_\text{M}|^{1/2}},$$

where

$$Q_\text{M} = \int_{-1}^{q} [K_\text{M}(z;q)]^2 \, \mathrm{d}z \quad \text{and} \quad k_\text{M} = \int_{-1}^{q} z^2 K_\text{M}(z;q) \, \mathrm{d}z.$$

(See the Appendix for explicit formulas.)

Fig. 2 shows overlay plots of 'efficacy' $= (\text{amse})^{5/4}$ for each of the four boundary kernel. It can be seen that $K_\text{M}(z;q)$ and $K_q^*(z)$ have generally the same large sample behavior. Likewise, $K_\text{GM}(z;q)$ and $K_\text{MW}(z;q)$ have similar large sample behavior. For $K_\text{M}(z;q)$ and $K_q^*(z)$, the efficacies have minima for q near 0.4 and increase slowly to the same value as q goes to 1.0. However, that the performance of $K_q^*(q)$ is better in this specific sense when q values are close to 0.0, is visible here. Whereas, the efficacies of $K_\text{GM}(z;q)$ and $K_\text{MW}(z;q)$ have minima for q near 0.2 and increase slowly to the same value as q goes to 1.0. In fact, the performance of $K_\text{MW}(z;q)$ is better in this specific

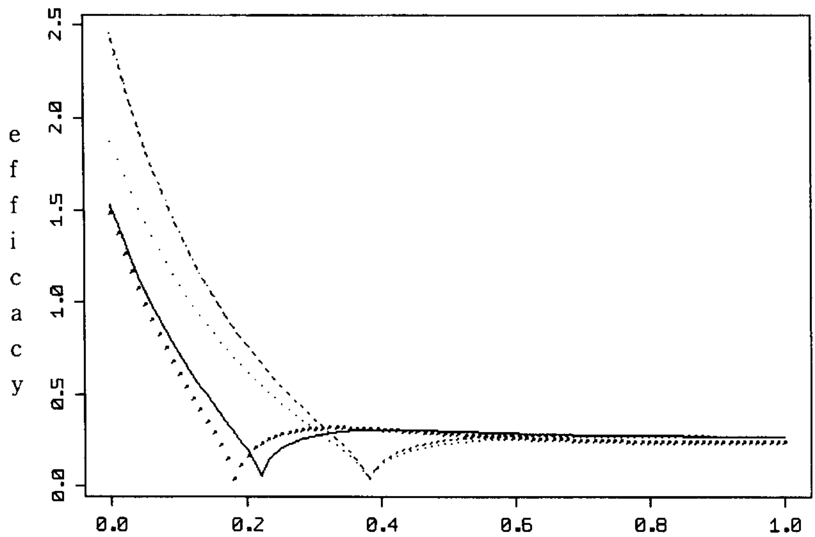


Fig. 2. Efficacies of boundary kernels: ($\cdots$) $K_q^*(\cdot)$, (—) $K_{\mathrm{GM}}(\cdot\,;q)$, (- - -) $K_{\mathrm{M}}(\cdot\,;q)$, (,,,) $K_{\mathrm{MW}}(\cdot\,;q)$.

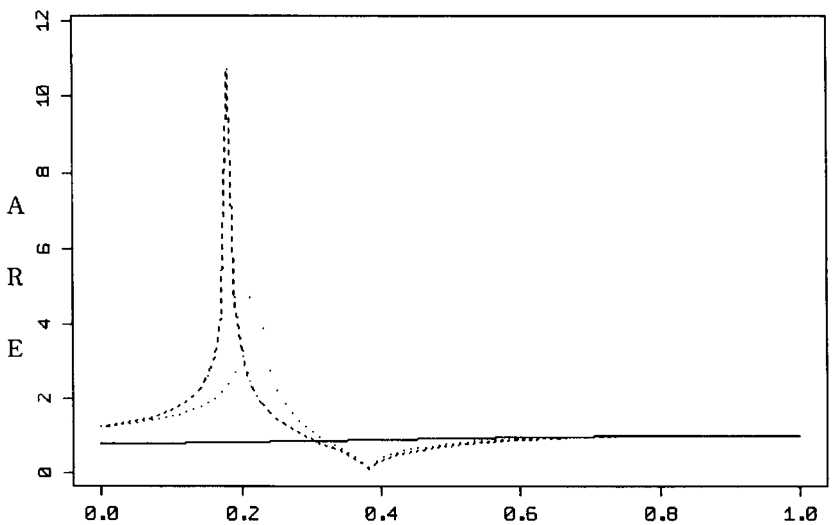


Fig. 3. Asymptotic relative efficiency: ($\cdots$) ARE of $K_{\mathrm{GM}}(\cdot\,;q)$ with respect to $K_q^*(q\,\cdot)$, (—) ARE of $K_{\mathrm{M}}(\cdot\,;q)$ with respect to $K_q^*(\cdot)$, (- - -) ARE of $K_{\mathrm{MW}}(\cdot\,;q)$ with respect to $K_q^*(\cdot)$.

sense than $K_{\mathrm{GM}}(z;q)$. Also $K_{\mathrm{GM}}(z;q)$ and $K_{\mathrm{MW}}(z;q)$ have the better performance when q values are between 0.0 and 0.3, whereas $K_{\mathrm{M}}(z;q)$ and $K_q^*(z)$ are better when q values are between 0.3 and 1.0.

Fig. 3 is a plot of the ARE as a function of q. Since the range of values is large for certain values of q, the logarithm of ARE is plotted in Fig. 4. This more clearly

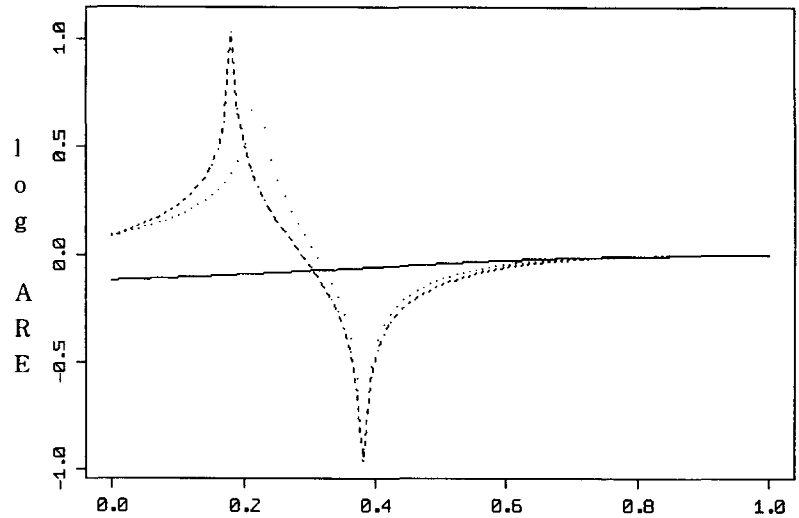


Fig. 4. Log of asymptotic relative efficiency: (···) log(ARE) of $K_{GM}(\cdot;q)$ with respect to $K_q^*(\cdot)$, (—) log(ARE) of $K_M(\cdot;q)$ with respect to $K_q^*(\cdot)$, (– – –) log(ARE) of $K_{MW}(\cdot;q)$ with respect to $K_q^*(\cdot)$.

displays the character of the comparisons than the limited number of values in Table 1. The new kernel, $K_q^*(z)$, has uniformly smaller amse than $K_M(z;q)$ for every $q \in [0, 1)$. This is no contradiction of the optimality found by Müller (1991), because (4.1) does not satisfy the 'smoothness' constraints. The relative efficiency of $K_M(z;q)$ is down to 95% at $q = 0.6$, 85% at $q = 0.3$, and achieves its minimum of 76.5% at the extreme boundary. This advantage to $K_q^*(z)$ has its price, namely, it is discountinuous at each end of its support. Müller (1991) observes that this may 'lead to relatively unsmooth curve estimates'. Also, we can see the same close agreement of $K_{GM}(z;q)$ and $K_{MW}(z;q)$. Clearly, either of these are far more efficient than $K_q^*(z)$ when q is near 0.2.

6. Conclusions

The general method examined here has some merits. Even though it would be tedious to calculate the resulting determinants, it can be extended to higher-order boundary kernels, although this is not recommended. For the specific case of $p = 2$ as the point of estimations enters the interior, the proposed kernel becomes the optimal Epanechnikov (1969) kernel. The bias of the new boundary kernel estimator has the same convergence rate as in the interior. Therefore, it has the same rate of mean square convergence as that of the nonboundary case. The new simple boundary kernel estimator depends continuously upon q and converges to the Epanechnikov kernel estimator as $q \to 1$.

For a fixed value of t in the boundary, it is natural to seek the optimal bandwidth $h(t)$. However, local bandwidth selection is not as straightforward as it is in the interior, since a different kernel is involved for each value of $q = t/h(t)$. See Müller (1991), Section 4. The practical difficulties of adaptive bandwidth selection should be essentially the same for the boundary kernels introduced here. Finite sample simulations would be necessary to determine which of the kernels would be better able to tolerate the under- and over-smoothing that occurs with data-based bandwidth choice.

We have used a general method to produce a specific boundary kernel, that has a smaller amse than Müller's (1991) 'smooth optimum' boundary kernels. Despite this favorable large sample relative efficiency, not matching the 'endpoint continuity' characteristic of the Epanechnikov may be a disadvantage with small samples. Our simulation experience with $K_q^*(z)$ in the boundary has not been disappointing with regard to curve estimates with too much wiggliness. However, we have made no small sample comparisons with other boundary kernels. Perhaps, these are less relevant in practice in that locally weighted polynomial regression adapts to the boundary automatically. See Hastie and Loader (1993) and the discussion for detailed illustrations of this important feature.

Acknowledgements

The authors are grateful to a referee for helpful comments and suggestions that improved the paper.

Appendix

(1) Explicit solutions and Q_{GM} and k_{GM} for the Gasser and Müller (1979) and Müller (1988) boundary kernel (5.1) require that we solve

$$\int_{-1}^{q} K_{\mathrm{GM}}(z;q)\,\mathrm{d}z = \frac{1}{2}a - \frac{3}{16}b + \frac{3}{4}aq + \frac{3}{8}bq^2 - \frac{1}{4}aq^3 - \frac{3}{16}bq^4 = 1$$

and

$$\int_{-1}^{q} zK_{\mathrm{GM}}(z;q)\,\mathrm{d}z = -\frac{3}{16}a + \frac{1}{10}b + \frac{3}{8}aq^2 + \frac{1}{4}bq^3 - \frac{3}{16}aq^4 - \frac{3}{20}bq^5 = 0.$$

In order to find a and b, let

$$\left(\frac{1}{2} + \frac{3}{4}q - \frac{1}{4}q^3\right)a + \left(-\frac{3}{16} + \frac{3}{8}q^2 - \frac{3}{16}q^4\right)b = 1,$$

$$\left(-\frac{3}{16} + \frac{3}{8}q^2 - \frac{3}{16}q^4\right)a + \left(\frac{1}{10} + \frac{1}{4}q^3 - \frac{3}{20}q^5\right)b = 0.$$

Then, using Cramer's rule,

$$a=\left(\frac{1}{10}+\frac{1}{4}q^3-\frac{3}{20}q^5\right)\Big/\left(\left(\frac{1}{2}+\frac{3}{4}q-\frac{1}{4}q^3\right)\left(\frac{1}{10}+\frac{1}{4}q^3-\frac{3}{20}q^5\right)\right.$$
$$\left.-\left(-\frac{3}{16}+\frac{3}{8}q^2-\frac{3}{16}q^4\right)^2\right),$$

$$b=-\left(-\frac{3}{16}+\frac{3}{8}q^2-\frac{3}{16}q^4\right)\Big/\left(\left(\frac{1}{2}+\frac{3}{4}q-\frac{1}{4}q^3\right)\left(\frac{1}{10}+\frac{1}{4}q^3-\frac{3}{20}q^5\right)\right.$$
$$\left.-\left(-\frac{3}{16}+\frac{3}{8}q^2-\frac{3}{16}q^4\right)^2\right).$$

Thus,

$$Q_{\mathrm{GM}}=\int_{-1}^{q}[K_{\mathrm{GM}}(z;q)]^2\,\mathrm{d}z$$
$$=\frac{9}{16}a^2-\frac{3}{16}ab+\frac{9}{112}b^2+\frac{9}{80}(a^2-2b^2)+\frac{3}{16}(-2a^2+b^2)$$
$$+\frac{9}{16}a^2q+\frac{9}{16}abq^2+\frac{3}{16}(-2a^2+b^2)q^3-\frac{9}{16}abq^4+\frac{9}{80}(a^2-2b^2)q^5$$
$$+\frac{3}{16}abq^6+\frac{9}{112}b^2q^7$$

and

$$k_{\mathrm{GM}}=\int_{-1}^{q}z^2K_{\mathrm{GM}}(z;q)\,\mathrm{d}z$$
$$=\frac{1}{10}a-\frac{1}{16}b+\frac{1}{4}aq^3+\frac{3}{16}bq^4-\frac{3}{20}aq^5-\frac{1}{8}bq^6.$$

(2) The values of Q_{M} and k_{M} for the Müller (1991) boundary kernel (5.2) are

$$Q_{\mathrm{M}}=\int_{-1}^{q}[K_{\mathrm{M}}(z;q)]^2\,\mathrm{d}z$$
$$=\frac{36}{(1+q)^6}\left\{\frac{100(-1+q)^2}{7(1+q)^4}+\frac{100(-1+q)^2q^7}{7(1+q)^4}-\frac{40(4-11q+11q^2-4q^3)}{3(1+q)^4}\right.$$
$$+\frac{40q^6(4-11q+11q^2-4q^3)}{3(1+q)^4}+\frac{4(3q-4q^2+3q^3)^2}{(1+q)^4}$$
$$+\frac{4q(3q-4q^2+3q^3)^2}{(1+q)^4}+\frac{8(47-187q+282q^2-187q^3+47q^4)}{5(1+q)^4}$$

$$
\begin{aligned}
&+\frac{8q^5(47-187q+282q^2-187q^3+47q^4)}{5(1+q)^4}\\
&-\frac{2(24-153q+323q^2-323q^3+153q^4-24q^5)}{(1+q)^4}\\
&+\frac{2q^4(24-153q+323q^2-323q^3+153q^4-24q^5)}{(1+q)^4}\\
&-\frac{12(-3q+16q^2-31q^3+31q^4-16q^5+3q^6)}{(1+q)^4}\\
&+\frac{12q^2(-3q+16q^2-31q^3+31q^4-16q^5+3q^6)}{(1+q)^4}\\
&+\frac{4(9-120q+364q^2-514q^3+364q^4-120q^5+9q^6)}{3(1+q)^4}\\
&\left.+\frac{4q^3(9-120q+364q^2-514q^3+364q^4-120q^5+9q^6)}{3(1+q)^4}\right\}
\end{aligned}
$$

and

$$
\begin{aligned}
k_{\mathrm{M}} &= \int_{-1}^{q} z^2 K_{\mathrm{M}}(z;q)\,\mathrm{d}z\\
&= \frac{6}{(1+q)^3}\left\{\frac{-5(-1+q)}{3(1+q)^2}+\frac{-5(-1+q)q^6}{3(1+q)^2}\right.\\
&\quad+\frac{4(-4+7q-4q^2)}{5(1+q)^2}+\frac{4q^5(-4+7q-4q^2)}{5(1+q)^2}-\frac{3(-1+4q-4q^2+q^3)}{2(1+q)^2}\\
&\quad+\frac{3q^4(-1+4q-4q^2+q^3)}{2(1+q)^2}+\frac{2(3q-4q^2+3q^3)}{3(1+q)^2}\\
&\quad\left.+\frac{2q^3(3q-4q^2+3q^3)}{3(1+q)^2}\right\}.
\end{aligned}
$$

(3) The values of Q_{MW} and k_{MW} for the Müller and Wang (1994) boundary kernel (5.3) are

$$
\begin{aligned}
Q_{\mathrm{MW}} &= \int_{-1}^{q} [K_{\mathrm{MW}}(z;q)]^2\,\mathrm{d}z\\
&= \frac{-108(1-2q)(-1+q)^2}{(1+q)^8}+\frac{108(1-2q)(-1+q)^2q^4}{(1+q)^8}+\frac{144(-1+2q)^2}{5(1+q)^8}\\
&\quad+\frac{144q^5(-1+2q)^2}{5(1+q)^8}-\frac{108(-1+q)^2(1-2q+3q^2)}{(1+q)^8}
\end{aligned}
$$

$$+\frac{108(-1+q)^2q^2(1-2q+3q^2)}{(1+q)^8}+\frac{36(1-2q+3q^2)^2}{(1+q)^8}$$

$$+\frac{36q(1-2q+3q^2)^2}{(1+q)^8}+\frac{12(13-52q+82q^2-60q^3+9q^4)}{(1+q)^8}$$

$$+\frac{12q^3(13-52q+82q^2-60q^3+9q^4)}{(1+q)^8}$$

and

$$k_{\mathrm{MW}}=\int_{-1}^{q} z^2K_{\mathrm{MW}}(z;q)\,\mathrm{d}z$$

$$=\frac{12(1-2q)}{5(1+q)^4}-\frac{9(-1+q)^2}{2(1+q)^4}+\frac{9(-1+q)^2q^4}{2(1+q)^4}+\frac{12(1-2q)q^5}{5(1+q)^4}$$

$$+\frac{2(1-2q+3q^2)}{(1+q)^4}+\frac{2q^3(1-2q+3q^2)}{(1+q)^4}.$$

References

Berlinet, A., 1993. Hierarchies of higher order kernels. Probab. Theory Related Fields 94, 489–504.

Epanechnikov, V.A., 1969. Nonparametric estimation of a multivariate probability density. Theory Probab. Appl. 14, 153–158.

Eubank, R.L., 1988. Spline Smoothing and Nonparametric Regression. Marcel Dekker, New York.

Eubank, R.L., Speackman, P.L., 1991. A bias reduction theorem with application in nonparametric regression, Scan. J. Statist. 18, 211–222.

Gasser, Th., Müller, H.G., 1979. In: Gasser, Th., Rosenblatt, M. (Eds.) Kernel estimation of regression functions: in Smoothing Techniques for Curve Estimation. Springer, Heidelberg, pp. 23–68.

Gasser, Th., Müller, H.G., 1984. Estimating regression functions and their derivatives by the kernel method. Scand. J. Statist. 11, 171–185.

Gray, H.L., Schucany, W.R., 1972. The Generalized Jackknife Statistic. New York. Marcel Dekker.

Hastie, T., Loader, C., 1993. Local regression: automatic kernel carpentry (with discussion), Statist. Sci. 8, 120–143.

Müller, H.G., 1988. Nonparametric Regression Analysis of Longitudinal Data, Springer, New York.

Müller, H.G., 1991. Smooth optimum kernel estimator near endpoints, Biometrika 78, 521–30.

Müller, H.G., 1993a. On the boundary kernel method for non-parametric curve estimation near endpoints, Scand. J. Statist. 20, 313–328.

Müller, H.G., 1993b. Comment to "Local regression: automatic kernel carpentry" by Hastie and Loader, Statist. Sci. 8, 134–139.

Müller, H.G., Stadtmüller, U., 1987. Variable bandwidth kernel estimators of regression curves. The Ann. Statist. 15, 182–210.

Müller, H.G., Wang, J.L., 1994. Hazard rate estimation under random censoring with varying kernels and Bandwidths. Biometrics 50, 61–76.

Priestley, M.B., Chao, M.T., 1972. Nonparametric function fitting J. Roy. Statist. Soc. B 34, 385–392.

Rice, J., 1984. Boundary modification for kernel regression. Commun. Statist. A 13, 893–900.

Schucany, W.R., Sommers, J.P., 1977. Improvement of kernel type density estimators. J. Amer. Statist. Assoc. 72 (358), 420–423.

Schuster, E.F., 1985. Incorporating support into nonparametric estimators of densities, Commun. Statist. A 14, 1123-36.

Wand, M.P., Schucany, W.R., 1990. Gaussian-based kernels. Canad. J. Statist. 18, 197–204.