Statistics and Computing **3** (1993) 135–146

Simple boundary correction for kernel density estimation

M. C. JONES

Department of Statistics, The Open University, Milton Keynes MK7 6AA, UK

Submitted February 1993 and accepted May 1993

If a probability density function has bounded support, kernel density estimates often overspill the boundaries and are consequently especially biased at and near these edges. In this paper, we consider the alleviation of this boundary problem. A simple unified framework is provided which covers a number of straightforward methods and allows for their comparison: 'generalized jackknifing' generates a variety of simple boundary kernel formulae. A well-known method of Rice (1984) is a special case. A popular linear correction method is another: it has close connections with the boundary properties of local linear fitting (Fan and Gijbels, 1992). Links with the 'optimal' boundary kernels of Müller (1991) are investigated. Novel boundary kernels involving kernel derivatives and generalized reflection arise too. In comparisons, various generalized jackknifing methods perform rather similarly, so this, together with its existing popularity, make linear correction as good a method as any. In an as yet unsuccessful attempt to improve on generalized jackknifing, a variety of alternative approaches is considered. A further contribution is to consider generalized jackknife boundary correction for density derivative estimation. *En route* to all this, a natural analogue of local polynomial regression for density estimation is defined and discussed.

Keywords: Boundary kernels, derivative estimation, generalized jackknifing, local linear regression, mean squared error, optimal kernels, reflection, renormalization

1. Introduction

Kernel convolution smoothers have difficulties at and near the boundaries when curve estimation is attempted over a region with one or more known boundaries. This is illustrated in the case of kernel density estimation for univariate positive data in Fig. 1. The solid curve is the kernel density estimate

$$\hat{f}(x) = n^{-1} \sum_{i=1}^{n} K_h(x - X_i) \qquad (1.1)$$

based on the suicide data of Copas and Fryer (1980) given in Table 2.1 of Silverman (1986). Here, $n = 86$ and we compromise between Figures 2.9(a) and 2.9(b) of Silverman (1986) by taking the bandwidth $h = 40$ when the kernel function K is the standard normal density function. $K_h(.)$ means $h^{-1}K(h^{-1}.)$. Clearly, $\hat{f}(x)$ overspills the known boundary at the origin. This results in considerable increased bias of the estimator at and near the origin compared with the estimator's bias in the interior of the density's support. Simply truncating $\hat{f}(x)$ to $[0, \infty)$ is, therefore, inappropriate, and even truncating and then renormalizing $\hat{f}$ to integrate to 1 makes insufficient difference, as also noted by Silverman (1986, p.29). The dashed curve in Figure 1 is the result of applying one of the simple boundary correction methods discussed below to these data; in fact it uses (3.4) when h and K remain as above. It suggests a continued increase of the density as x approaches zero, as opposed to the artifactual mode at around $x = 50$. We imagine that this curve is a considerably improved estimate of the underlying density at and near the origin.

Our straightforward, and largely unified, treatment of the boundary problem is based on easy manipulations of simple asymptotic approximations to $\hat{f}$'s bias and variance. We first describe how the boundary problem manifests itself in such theoretical terms. Suppose for simplicity that there is a single known boundary to the support of f which we might as well take to be at the origin, so we are dealing with positive data. (Obviously, there is no loss of generality in this choice, nor is it very difficult to extend our work to

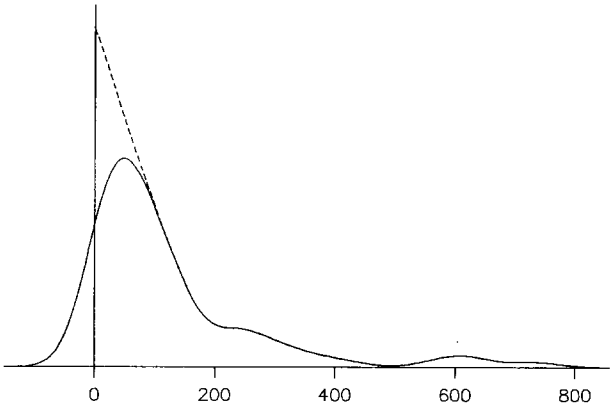

Fig. 1. *Raw kernel density estimate (solid line) and boundary corrected kernel density estimate (using (3.4); dashed line) for the suicide data ($n = 86, h = 40$, normal kernel)*

two boundaries, for example, at 0 and 1.) For convenience, take K to be a symmetric probability density with support $[-1, 1]$ (although our results can be modified to allow, for instance, normal kernels). Away from the boundary, which thus means at any $x \geq h$, there is no overlap of contributing kernels with the boundary, and hence the usual asymptotic mean and variance expressions apply. Suppose f has two continuous derivatives everywhere, and that as $n \to \infty$, $h = h(n) \to 0$ and $nh \to \infty$. Then,

$$E\left\{\hat{f}(x)\right\} \simeq f(x) + \tfrac{1}{2}h^2 s_2 f''(x)$$

and

$$V\left\{\hat{f}(x)\right\} \simeq (nh)^{-1}\kappa f(x),$$

where $s_l = \int_{-1}^{1} u^l K(u)du$ and $\kappa = \int_{-1}^{1} K^2(u)du$.

Near the boundary, however, things are rather different. Write $x = ph$ and bear in mind throughout that p is a function of x. Also, set $a_l(p) = \int_{-1}^{p} u^l K(u)du$ and $b(p) = \int_{-1}^{p} K^2(u)du$. Then,

$$E\left\{\hat{f}(x)\right\} \simeq a_0(p)f(x) - ha_1(p)f'(x) + \tfrac{1}{2}h^2 a_2(p)f''(x)$$

and

$$V\left\{\hat{f}(x)\right\} \simeq (nh)^{-1}b(p)f(x)$$

(e.g. Gasser and Müller, 1979, Härdle, 1990, Eubank and Speckman, 1991). When $p \geq 1$, these expressions reduce to those of the interior. Notice that the basic kernel estimator is not even consistent within h of the boundary (unless $f(x) = 0$ there) with, in particular, $E\{\hat{f}(0)\} = f(0)/2$. Of course, these expressions pertain equally to the naive modification of truncating $\hat{f}(x)$ to $[0, \infty)$ with or without overall renormalization.

Section 2 is devoted to two simple and well-known methods of achieving consistency at and near the boundary, and to a comparison of the two. It has to be stressed that these methods still have a bias of order h which remains an order of magnitude worse than the $O(h^2)$ bias in the interior of the support.

Simple methods which also achieve $O(h^2)$ bias everywhere are the main focus of the paper. The methods described in Section 3 form a unified whole as special cases of applying the simple technique of generalized jackknifing (Schucany *et al.*, 1971) to the boundary bias problem. Rice (1984) made one of these proposals. Another (a linear correction formula, (3.4)) has close links with the boundary properties of local linear fitting (Fan and Gijbels, 1992); these are explored in Section 5 (which also describes a natural adaptation of local polynomial regression to density estimation). Novel boundary kernels involving kernel derivatives and generalized reflection arise too. The methods of Section 3 are compared in Section 4. Our comparison is concerned with the kernel-dependent coefficients of $h^2 f''(x)/2$ (for the bias) and $(nh)^{-1}f(x)$ (for the variance) induced by these boundary corrections. The main message is of little to choose between generalized jackknife alternatives in these terms. The relationship between linear correction and the 'optimal' boundary kernels of Gasser *et al.* (1985) and Müller (1991) is explored in Section 6.

The major role played by the linear boundary correction, which is as good as any generalized jackknife in performance terms, and which is already popular in the literature both as an *ad hoc* device and because of its connections with other viewpoints, becomes clear. We therefore give some examples of its use in Section 7. We are happy to go along with the view that this linear correction is as good a method as any to recommend for practice, although we also note that there seems still to be scope for improvement. A variety of (largely unsuccessful) attempts to do better than linear correction is described in Section 9, together with some further references to more sophisticated methods (of promise).

A further contribution is to consider boundary correction for density derivative estimation in Section 8. While the boundary effect gets worse for raw derivative estimation, appropriate generalized jackknifing achieves boundary derivative estimators almost as good as estimation in the interior of the support. Polynomial correction, linked with local polynomial fitting, retains a high profile. Some final remarks are made in Section 10, including some comments on the otherwise ignored questions of bandwidth selection and adaptation.

2. A first step: ensuring consistency

The consistency requirement is that the leading term in the expectation of a 'boundary-corrected' kernel density estimate is $f(x)$ itself. There are two well-known ways of doing this. Noting that the multiplier of $f(x)$ is $\int_{-1}^{p} K(u)du$, it is clear that the driving force here is indeed

the kernel mass 'lost' beyond the boundary. Forcing integration of each *kernel* to unity by renormalization is one obvious method of accounting for this (e.g. Diggle, 1985; Härdle, 1990). (Note that this 'local' renormalization is quite different from the 'global' one alluded to in the previous section.) That is, use $\hat{f}_N(x) \equiv \hat{f}(x)/a_0(p)$; since $a_0(p) = 1$ for $p \geq 1$, this formula also covers using $\hat{f}$ in the interior. The effect of this is immediately clear:

$$E\left\{\hat{f}_N(x)\right\} \simeq f(x) - h\frac{a_1(p)}{a_0(p)}f'(x)$$

and

$$V\left\{\hat{f}_N(x)\right\} \simeq (nh)^{-1}\frac{b(p)}{a_0^2(p)}f(x).$$

The result is now consistent, but still has a bias of order h near the boundary (except where $f'(x) = 0$): this continues to compare poorly with the $O(h^2)$ bias term which leads in the interior (that is, at boundary points, an optimized MSE of order $n^{-2/3}$ is obtained as opposed to $n^{-4/5}$ elsewhere).

Let us now look at a second way of enforcing consistency that has been quite popular in the literature. This is to reinstate the 'missing mass' by reflecting the estimate in the boundary (e.g. Boneva *et al.*, 1971, Hominal and Deheuvels, 1979, Schuster, 1985, Silverman, 1986, Ghosh and Huang, 1992). That is, utilize $\hat{f}_R(x) \equiv \hat{f}(x) + \hat{f}(-x)$ or, equivalently, replace $K_h(x - X_i)$ by $K_h(x - X_i) + K_h(-x - X_i)$. Since

$$E\left\{\hat{f}(-x)\right\} \simeq \int_{-1}^{-p} K(u)du\, f(x) - h\int_{-1}^{-p}(u + 2p)K(u)du\, f'(x) + \tfrac{1}{2}h^2\int_{-1}^{-p}(u + 2p)^2K(u)du\, f''(x),$$

we have

$$E\left\{\hat{f}_R(x)\right\} \simeq f(x) - h2[a_1(p) + p\{1 - a_0(p)\}]f'(x).$$

This formula at $p = 0$ is given by Cline and Hart (1991), and in general by Marron and Ruppert (1992). The variance of $\hat{f}_R(x)$ is

$$V\left\{\hat{f}_R(x)\right\} \simeq (nh)^{-1}\left(\kappa + 2\int_{-1}^{p} K(u)K(u - 2p)du\right)f(x).$$

Cline and Hart (1991) give the $p = 0$ version of this formula, noting that it corrects the variance expression given by Schuster (1985).

A theoretical comparison of $\hat{f}_N$ and $\hat{f}_R$ can now be obtained by comparing kernel-dependent quantities in biases and variances as functions of p for various choices of K. This is done for the biweight kernel, $15(1 - x^2)^2/16$, $x \in [-1, 1]$, in Fig. 2. In Fig. 2(a), we see that the bias of $\hat{f}_R$ is not greater than that of $\hat{f}_N$ for any $p \in [0, 1]$, but the differences are not great. Diggle and Marron (1988) allude to this result. Likewise, in Fig. 2(b), for variances roles are reversed for p less than about one half, although reflection also 'wins' marginally for larger p. Finally, Fig. 2(c) is an attempt to produce a meaningful combination of squared bias and variance. There, the curves are the kernel-dependent quantities in the pointwise optimized mean squared error, namely $\{B(p)V(p)\}^{2/3}$ where $B(p)$ and $V(p)$ are here the multipliers of $-hf'(x)$ and $(nh)^{-1}f(x)$ in bias and variance, respectively, for $x = ph$ (this can be achieved by canonical scaling; see Section 10). Remember, however, that f' and

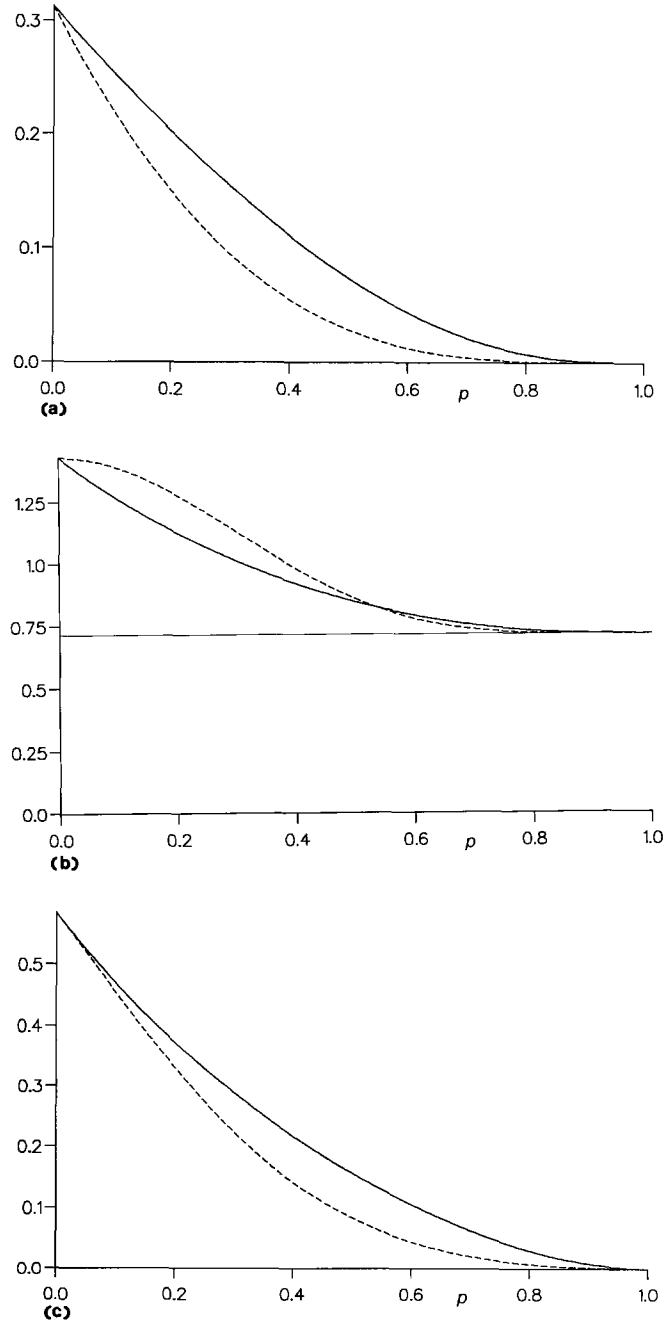

Fig. 2. *(a) $B(p)$, (b) $V(p)$ and (c) $\{B(p)V(p)\}^{2/3}$ for renormalization (solid line) and reflection (dashed line). $B(p)$ refers to $O(h)$ bias. The horizontal line in (b) indicates $V(1)$. Biweight kernel.*

f have a major influence on complete bias and variance. Reflection apparently beats renormalization for all p, but the difference remains minor. Analogous pictures for the uniform kernel and for the normal kernel (for which take the bottom limit as $-\infty$ rather than -1 and allow any $p \geq 0$) are very similar, although for the normal kernel, renormalization is slightly better for p less than about 1/2, and reflection wins otherwise. Note that reflection imposes $\hat{f}'(0) = 0$ (Silverman, 1986) while renormalization does not. But the main points here are that there is really very little to choose between these two approaches (Diggle and Marron, 1988, give an example in which the two edge corrections yield similar results) and, of course, that neither is as good as those to follow.

3. Boundary correction for $O(h^2)$ bias: using generalized jackknifing

Let us now consider obtaining $O(h^2)$ bias near the boundary as well as in the interior. Rather than start from existing methods, of which there are many, we will first take a simple approach to the problem, of which there are many special cases, and relate these to existing methods as we go. The idea is this. Take a linear combination of K and some other function L, closely related to K, in such a way that the resulting kernel has the desired properties $a_0(p) = 1$ and $a_1(p) = 0$. Writing $c_l(p) = \int_{-1}^{p} u^l L(u)du$ (and reserving $a_l(p)$ for the corresponding quantities based on K), we see that the linear combination

$$\{c_1(p)K(x) - a_1(p)L(x)\}/\{c_1(p)a_0(p) - a_1(p)c_0(p)\} \tag{3.1}$$

will have the desired $O(h^2)$ bias property.

This idea is not entirely novel, having been explicitly proposed (in a regression context) for essentially the choice $L(x) = cK(cx)$ by Rice (1984): that is, combine density estimates using the same kernel but different bandwidths h and ch, where $0 < c < 1$ without loss of generality. Immediately, the resulting 'boundary kernel' is

$$K_c(x) = \frac{\{a_1(pc) - a_1(c)\}K(x) - a_1(p)c^2K(cx)}{\{a_1(pc) - a_1(c)\}a_0(p) - a_1(p)c\{a_0(pc) + a_0(c) - 1\}}. \tag{3.2}$$

(Note that Rice's (1984) formula (2.6) is applicable to versions of $K(x)$ and $cK(cx)$ pre-normalized to integrate to 1: this formula and Rice's are then identical). ((3.2) is not useful for the uniform kernel). See also Härdle (1990, Section 4.4).

Notice that (3.2) is actually a family of boundary kernels indexed by $0 < c < 1$. One could think of choosing $c(= c(K))$ to optimize some further measure of effectiveness of the kernel, or by some other criterion such as mentioned by Rice (1984, p.896). But, like Rice, we will not pursue this here; preliminary investigations suggest that there is very little to be gained. Instead, we will mention a related boundary kernel, that obtained by letting $c \to 1$ in (3.2). This is, writing $a_l^{(1)}(p) \equiv \int_{-1}^{p} x^l K'(x)dx$,

$$K_{\mathrm{PD}}(x) = \frac{a_2^{(1)}(p)K(x) - a_1(p)xK'(x)}{a_2^{(1)}(p)a_0(p) - a_1(p)a_1^{(1)}(p)} \tag{3.3}$$

(except for K uniform). An alternative derivation of (3.3) would simply be to seek the appropriate linear combination of $K(x)$ and $xK'(x)$ to use as a boundary kernel.

The development of (3.2) and (3.3) has exactly paralleled the methodology for obtaining fourth-order kernels (ignoring boundary problems) from second-order ones described by Jones and Foster (1993). (While the practicality of higher-order kernels is in doubt, Marron and Wand, 1992, the need for boundary kernels is much clearer). Both procedures are simple instances of the generalized jackknifing idea of Schucany *et al.* (1971). Moreover, the combination of kernels with different bandwidths proposed by Rice (1984) is the exact analogue of the method of Schucany and Sommers (1977) in the higher-order kernel context. Formula (3.3) also has its analogue in Jones and Foster (1993) as do several of the remaining ideas in this section.

A particularly useful boundary kernel formula comes from linearly combining $K(x)$ and $xK(x)$ (in the higher-order kernel context, one needs x^2 rather than x). The kernel is

$$K_{\mathrm{L}}(x) = \frac{(a_2(p) - a_1(p)x)K(x)}{a_0(p)a_2(p) - a_1^2(p)}. \tag{3.4}$$

Relations between this formula, especially, and existing methods in the literature will be made clear in Sections 5 and 6, although here we mention Gasser and Müller (1979) as an early suggestion of it, and Hart and Wehrly (1992) as a recent one.

Other boundary kernels include

$$K_{\mathrm{D}}(x) = \frac{a_1^{(1)}(p)K(x) - a_1(p)K'(x)}{a_1^{(1)}(p)a_0(p) - a_1(p)a_0^{(1)}(p)} \tag{3.5}$$

(the analogue in Jones and Foster, 1993, involves K'' rather than K'). It arises rather naturally as a combination of renormalization for consistency and using $\hat{f}'$ (albeit with the same bandwidth h as used in $\hat{f}$) to estimate bias and remove it. Formula (3.5) is not applicable to the uniform kernel, and is the same as (3.4) for the normal. Analogues of formulae based on convolutions or on powers of K (as in Jones and Foster, 1993) or others could be pursued as well. Relating L to K allows boundary kernels 'derived from K', but it is also reasonable to combine two unrelated kernels.

While renormalization apparently continues to play a role if the above are thought of as extensions of the methods of Section 2 to achieve zero $O(h)$ bias too, can

we also incorporate reflection in methods that achieve bias of order h^2 everywhere? The answer is yes, and the method is more specific to the boundary kernel case than others, having no analogue in higher-order kernel provision. It is to combine $K(x)$ with $K(2p-x)$, so that $\hat{f}_{\mathrm{R1}}$, say, uses the boundary kernel

$$K_{\mathrm{R1}}(x) = \frac{\{2p(1-a_0(p))+a_1(p)\}K(x) - a_1(p)K(2p-x)}{\{2p(1-a_0(p))+a_1(p)\}a_0(p) - a_1(p)(1-a_0(p))}. \tag{3.6}$$

Like formulae (3.3) and (3.5), (3.6) is entirely novel. Comparisons of these suggestions follow.

4. Properties of generalized jackknife boundary corrections

As we did for the $O(h)$ bias reflection and renormalization methods in Section 2, in this section we compare kernel-dependent coefficients in leading bias and variance terms for the $O(h^2)$ bias methods of Section 3. This time, write $B(p)$ for the multiplier of $h^2 f''(x)/2$ and $V(p)$ as before. General formulae in terms of K and L can be given to cover all generalized jackknife boundary kernels. In fact,

$$B(p) = \{c_1(p)a_2(p) - a_1(p)c_2(p)\}/\{c_1(p)a_0(p) - a_1(p)c_0(p)\}.$$

Also, writing $e(p)$ and $g(p)$ for $\int_{-1}^{p} K(x)L(x)dx$ and $\int_{-1}^{p} L^2(x)dx$, respectively, we have

$$V(p) = \{c_1^2(p)b(p) - 2c_1(p)a_1(p)e(p) + a_1^2(p)g(p)\}/\{c_1(p)a_0(p) - a_1(p)c_0(p)\}^2.$$

To make more of these, see Fig. 3. There, $B(p)$ (Fig. 3(a)) and $V(p)$ (Fig. 3 (b)) are shown for each of formulae (3.3) to (3.6) as a function of p for the biweight kernel. All bias curves have much the same shape and range of values. Each has a single point where the bias crosses zero. The four variance curves are also very similar to one another. Optimized mean squared error curves, now $\{B(p)V^2(p)\}^{2/5}$, are shown in Fig. 3(c), but are not very edifying except in as much as all curves are, again, similar, and they will not be displayed in similar figures later on. One might, however, note that the slightly increased variance of $\hat{f}_{\mathrm{L}}$ close to $p = 0$ is balanced by its better bias there. The difficulty is around the points of zero leading bias. A more realistic approximation near these points should also incorporate the next term in the bias expansion (Schucany, 1989), and if this were done, we could expect monotone decreasing curves as for the variance.

Aside from broad equivalence of all generalized jackknives — and this extends to other possibilities mentioned in Section 3 but not dealt with further — the major point

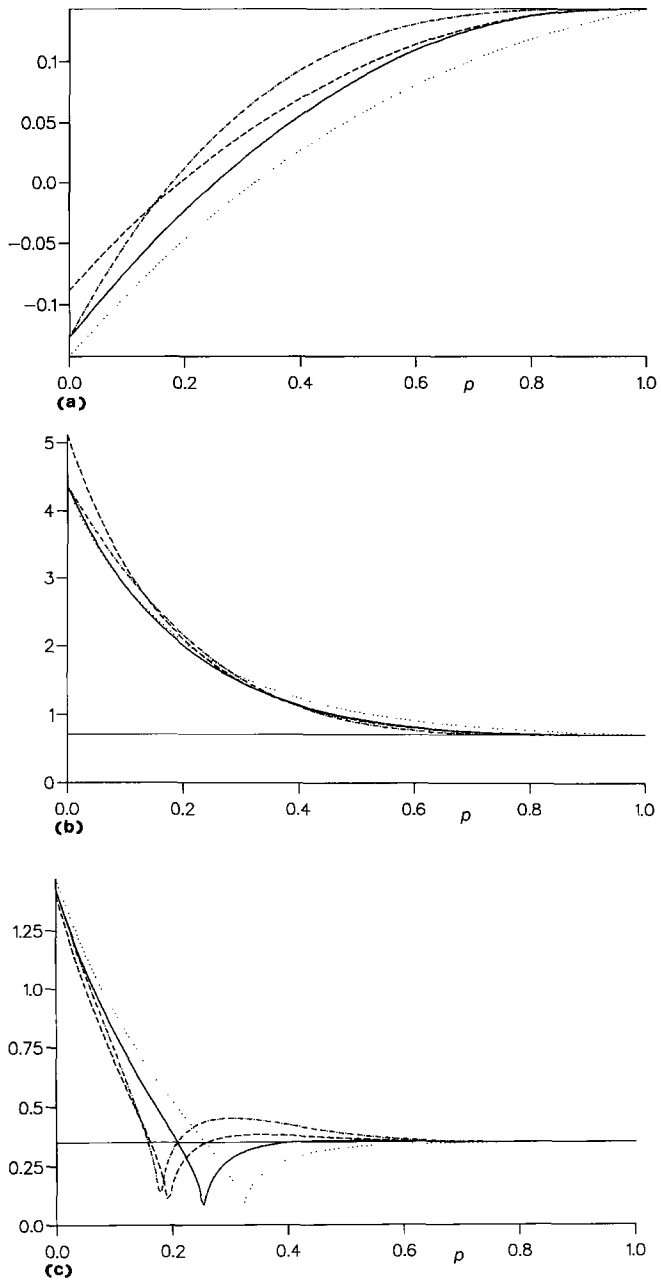

Fig. 3. *(a) $B(p)$, (b) $V(p)$ and (c) $\{B(p)V^2(p)\}^{2/5}$ for generalized jackknives (3.3) (dotted line), (3.4) (dashed line), (3.5) (solid line) and (3.6) (dot–dashed line). $B(\backslash, p)$ refers now to $O(h^2)$ bias. The dotted horizontal line in (a) is at zero, the solid horizontal lines indicate values of quantities at $p = 1$. Biweight kernel*

to make concerns the variance of boundary corrected estimates at (and very close to) $p = 0$. For example, for the biweight kernel, $V\{\hat{f}_{\mathrm{L}}(0)\}/V\{\hat{f}_{\mathrm{L}}(1)\} \simeq 7.16$; meanwhile, $V\{\hat{f}_{\mathrm{N}}(0)\}/V\{\hat{f}_{\mathrm{N}}(1)\} = 2$ (and $V\{\hat{f}_{\mathrm{L}}(1)\} = V\{\hat{f}_{\mathrm{N}}(1)\} = V\{\hat{f}(1)\}$). This considerable increase in variance is disturbing (see also the end of Section 5), and makes one hope that perhaps other boundary correction techniques might appear that beat generalized jackknifing in this regard.

5. Linear multiples, local linear regression, and local linear density estimation

Of all the generalized jackknives so far considered, the one that is already quite popular in the literature is the linear multiple (of K) one, (3.4). Hart and Wehrly (1992) is entirely concerned with it in a regression context. An appealing property of (3.4) is its link with local linear fitting, the subject of this section.

There are various good reasons why kernel weighted local linear regression (the stuff of *lowess*, Cleveland, 1979) is an especially attractive approach to kernel-based regression estimation (Fan 1992, 1993; Hastie and Loader, 1993). One of these reasons is its attractive performance at boundaries (Fan and Gijbels, 1992). The work of Ruppert and Wand (1993) makes it clear that this performance is due to local linear fitting implicitly resulting in the use of the appropriate linear multiple of K at the boundary. See also Section 7 of Müller (1993). The beauty of local linear regression is that this boundary behaviour is not imposed by the user but is an automatic consequence of the algorithm itself.

We now make this explicit in the density estimation context in which the rest of this paper is set. (Also, as we shall see, the density estimation context affords exact 'equivalences', while in the regression setting, these equivalences remain asymptotic; see also Lejeune, 1985; Müller, 1987). To do so, we follow Lejeune and Sarda (1992) in considering the (kernel weighted) local fitting of lines to the empirical distribution function F_n. That is, choose α and β to minimize

$$\int K_h(x-u)\{F_n(u) - \alpha - \beta(x-u)\}^2 du, \qquad (5.1)$$

for each x, and take $\tilde{F}(x)$ as the minimizing value of $\alpha(x)$. To estimate f itself, either differentiate $\tilde{F}(x)$ with respect to x or else obtain precisely the same result by minimizing

$$\int K_h(x-u)\{f_n(u) - \alpha - \beta(x-u)\}^2 du; \qquad (5.2)$$

here, $f_n(u) = n^{-1}\sum_{i=1}^n \delta(x - X_i)$, where δ is the Dirac delta function, is the empirical density function.

In the interior of f's support, this yields precisely the ordinary kernel estimate (1.1) with kernel K. The main interest internally, then, is in fitting higher order polynomials, which turns out to be exactly equivalent to using higher order kernels that are appropriate polynomial multiples of K (the kernel order is $2([l/2]+1)$, where $[.]$ denotes 'integer part of' and l is the degree of the polynomial being fitted) (Sarda, 1991; Lejeune and Sarda, 1992). Jones and Foster (1993) give a generalized jackknifing derivation of such kernels.

To get the boundary behaviour of this approach, simply restrict the integral in (5.2) to the desired support, here $[0, \infty)$, and proceed as before. *The answer is precisely* $\hat{f}$ *in* (1.1) *with kernel given by* (3.4), as is easy to verify. This simple density estimation development (at the boundary) appears to be novel (although in as yet unpublished work, C. Z. Wei and C. K. Chu appear to be on a similar track). (A different, and more complicated, expression arises from (5.1); but one should not really mix differentiation and local polynomial fitting: see Section 8.)

While Fan and Gijbels (1992) stress that convergence rates of local linear regression fitting are maintained right up to the boundary, Ruppert and Wand (1993) introduce a cautionary note by stressing its increased variance in terms of constants. We now see that this is precisely the same point as was being made at the end of Section 4, since $\hat{f}_L$ and $\hat{f}_N$ parallel local linear fitting and the Nadaraya–Watson estimator, respectively.

6. Linear multiples and optimal boundary kernels

A number of papers by H.-G. Müller and colleagues consider optimal kernel theory for the interior; Granovsky and Müller (1991) is a good reference. Various versions of the problem all involve minimization of kernel-dependent quantities appearing in appropriate performance measures. The minimization has to be carried out under additional constraints (see Scott, 1992, Section 6.2.3.2 for a nice explanation). For example, the minimum of asymptotic mean squared error, under the constraint of zero sign changes on the kernel's support, is achieved by the Epanechnikov kernel, $(3/4)(1-x^2), -1 \le x \le 1$.

In some of these papers, e.g. Gasser *et al.* (1985), optimal boundary kernel theory was also considered: Müller (1991) is the main reference on this. (Azari *et al.*, 1992, set up an approach to regression in which such boundary kernels naturally arise). The analogue for minimum (asymptotic) mean squared error is a cubic expression closely related to the Epanechnikov kernel. But it is *not* the result of applying the linear correction (3.4) to the Epanechnikov kernel. This is because of the additional constraint imposed by Müller and colleagues, namely, that $K(p) = 0$ (as well as $K(-1) = 0$). Thus, optimal kernels are examples of what Scott (1992) calls 'zero boundary kernels' while generalized jackknifing results in 'floating boundary kernels'.

The current author finds it difficult to understand why setting $K(p) = 0$ is particularly natural. In Fig. 4, we compare $B(p)$ and $V(p)$ for Müller (1991) and (3.4) extensions of each of the uniform, Epanechnikov, biweight and triweight ($\propto (1-x^2)^3$ on $[-1,1]$) kernels (the other kernels are associated with certain optimization problems involving derivatives). (Correction: formula (4.2) of Müller, 1991, should start its second line with a negative rather than a positive sign.) We see (i) again, that there is not all that much to choose between the two, but (ii) the generalized jackknife seems a little better than the 'optimal' kernel: this is certainly true of variance, and appears

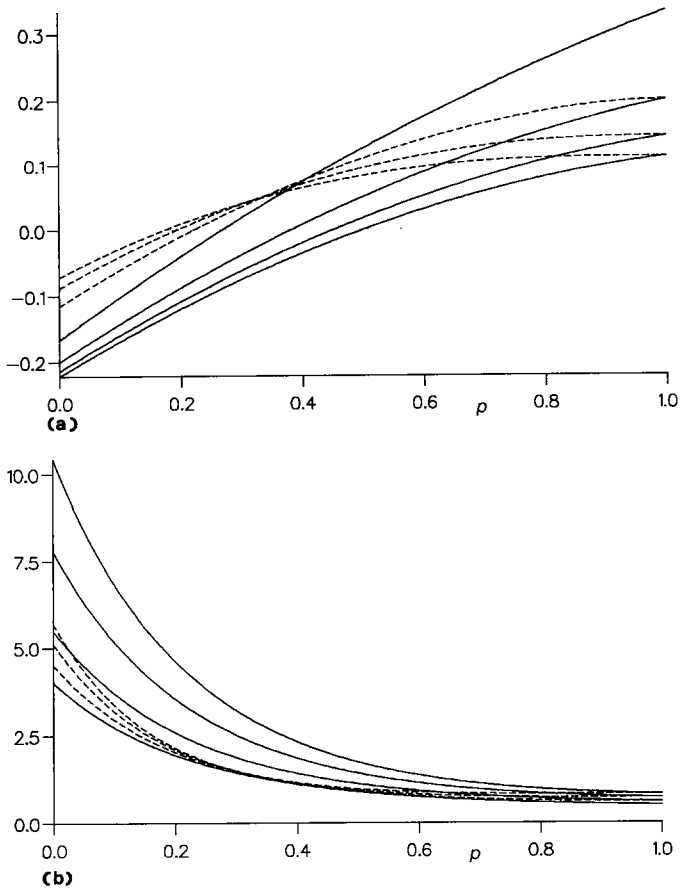

Fig. 4. *(a) B(p) and (b) V(p) for generalized jackknife (3.4) (dashed lines) and optimal zero boundary kernels (solid lines) associated with uniform, Epanechnikov, biweight and triweight kernels. Reading down the vertical axes, in (a) the lines correspond to triweight, biweight, Epanechnikov, uniform lines coincident and then Epanechnikov, biweight, triweight, while in (b) we have triweight, biweight, triweight, Epanechnikov, biweight, Epanechnikov, coincident uniforms. The dotted horizontal line in (a) is at zero*

to carry through to mean squared error (not shown). So, rather than disparaging generalized jackknifing for not obtaining Müller boundary kernels, a minor inappropriateness of Müller kernels might be concluded. Indeed, it can be conjectured that a formulation of optimal kernel theory in which the constraint amounts to truncating K with $K(p) \neq 0$ and $K(1) = 0$ would result in applying (3.4) to interior (second-order) optimal kernels.

Müller (1993) is also relevant to this section.

7. Examples of linear correction

This section is intended to give the reader the briefest of feels for what boundary corrected kernel estimates might look like, using the linear correction generalized jackknife (3.4). We use this because it is reasonably representative of generalized jackknife boundary kernels while already having, as we have just seen, some degree of popularity in the literature. A first example of its application (to the suicide control data) was given in Fig. 1.

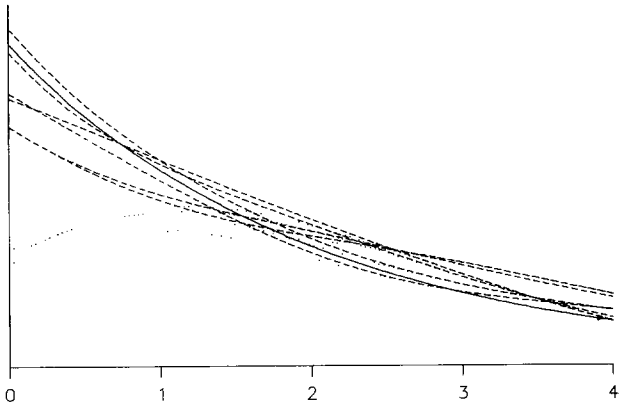

Fig. 5. *Truncated kernel estimates (dotted lines) and linearly corrected estimates (dashed lines) based on six datasets ($n = 100, h = 1$, normal kernel) simulated from an exponential distribution. The true density (solid line) is also shown*

Representatives of some simulations involving χ^2 distributions are given in Figs 5 and 6. In the first of these, six datasets each of 100 points were simulated from the exponential distribution (χ^2_2) and truncated and boundary corrected kernel estimates, each using $h = 1$, are shown. It is clear that the latter are indeed a considerable improvement on the former. The extra variability of the linear correction near $p = 0$ is, however, apparent. Interestingly, the improved bias of the linear correction over, for example, ordinary reflection, is also quite clear. After all, although reflection would give rise to curves shaped like the dashed ones, they would still intercept too low; recall that reflection would only double $\hat{f}(0)$.

Figure 6 repeats the exercise for simulations from a χ^2_6 distribution (again using $n = 100, h = 1$). In these particular samples, the most severely non-zero intercept of the basic kernel estimate results in linear correction erroneously increasing the estimate near the boundary, but in

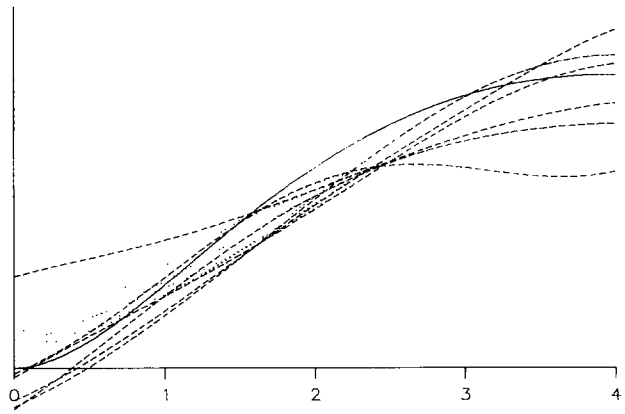

Fig. 6. *Truncated kernel estimates (dotted lines) and linearly corrected estimates (dashed lines) based on six datasets ($n = 100, h = 1$, normal kernel) simulated from a χ^2_6 distribution. The true density (solid line) is also shown*

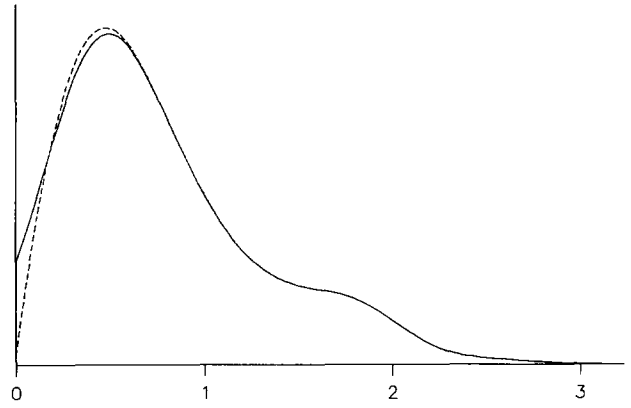

Fig. 7. *Truncated kernel density estimate (solid line) and boundary corrected kernel density estimate (using (3.4); dashed line) for the length biased shrub width data ($n = 46, h = 0.23$, normal kernel)*

the other five samples, estimates hit zero before reaching $x = 0$, and do seem to constitute an improvement. The one potential disadvantage here, of course, and this is common to all generalized jackknives, is that the estimate continues into negative values near the boundary (and this is not unrelated to its increased variance property). An obvious practical remedy is to truncate the estimate to non-negative values. Whatever the reader's reaction to negativity of density estimates, this property is not a problem in regression estimation. Note also that generalized jackknifing does not result in unit integral density estimators either.

Further experiments with other degrees-of-freedom values (not shown) indicated a generally favourable, but by no means infallible, performance of linear correction in terms of increasing or decreasing density estimates near zero, in a manner closely connected with the behaviour of the χ^2 distribution in the boundary region.

One further real data example is shown in Fig. 7. If the smoothing problem is formulated in the most appropriate way, just the same boundary correction methodology is immediately applicable to situations involving indirect

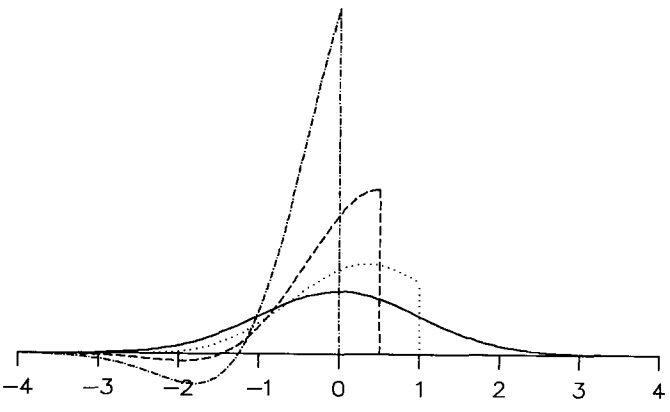

Fig. 8. *Standard normal kernel (solid line) together with linearly corrected boundary versions when $p = 1$ (dotted line), $p = 1/2$ (dashed line) and $p = 0$ (dot–dashed line)*

data. One example is that of length biased data. Here, the data arise from $g(x) = xf(x)/\mu$, where $\mu = \int xf(x)dx$, and interest remains in estimating f. If we use the estimator $n^{-1}(\sum_{j=1}^n X_j^{-1})^{-1} \sum_{i=1}^n X_i^{-1} K_h(x - X_i)$ (Jones, 1991), all generalized jackknifing arguments carry through. Linear boundary correction is used in Fig. 7 on the length-biased shrub width data of Muttlak and McDonald (1990); it is compared there with the truncated estimator used by Jones (1991) ($n = 46, h = 0.23$). It is encouraging to note that linear correction brings the estimate down near zero, while in Fig. 1, linear correction increased $\hat{f}(0)$, although each started from an apparently similar truncated estimator.

Finally in this section, we display a few linearly corrected boundary kernels themselves. In Fig. 8 are shown the standard normal kernel together with associated boundary kernels when $p = 0, 1/2$ and 1. The general formula for these curves is

$$\frac{\{\Phi(p) + (u - p)\phi(p)\}\phi(u)}{\Phi(p)\{\Phi(p) - p\phi(p)\} - \phi^2(p)},$$

for $u < p$.

8. Boundary correction for derivatives

As the reader might imagine, boundary problems are exacerbated when density derivatives are of interest. This shows up in the asymptotics. For instance, the usual estimator of the first derivative given by

$$\hat{f}'(x) = n^{-1} \sum_{i=1}^n K_h'(x - X_i) \tag{8.1}$$

(where $K_h'(.) \equiv h^{-2}K(h^{-1}.)$) has expectation

$$h^{-1}a_0^{(1)}(p)f(x) - a_1^{(1)}(p)f'(x) + O(h)$$

for $x < h$. That is, not only is there a constant multiplying the desired $f'(x)$ near the boundary, but there is also an additional term of order h^{-1}! Likewise, for rth derivatives in general, the leading term in the expectation is of order h^{-r}.

Generalized jackknifing can, however, be utilized to obtain bias of order h^2 for derivative estimation too. Concentrate for the moment on the first derivative of f only. We revert to the general idea at the beginning of Section 3: take a linear combination, J, of K and other functions related to K such that the resulting kernel has the required properties, which are now that $\int_{-1}^p J(u)du = 0 = \int_{-1}^p u^2 J(u)du$ and $\int_{-1}^p uJ(u)du = 1$. That is, we need appropriate K, L *and* M. Division of everything by an extra h is also necessary.

Write $d_l(p) \equiv \int_{-1}^p u^l M(u)du$. Then, in general, a boundary kernel for the first derivative is h^{-1} times

$$\frac{-A_0(p)K(x) + B_0(p)L(x) - C_0(p)M(x)}{a_0(p)A_1(p) - c_0(p)B_1(p) + d_0(p)C_1(p)} \tag{8.2}$$

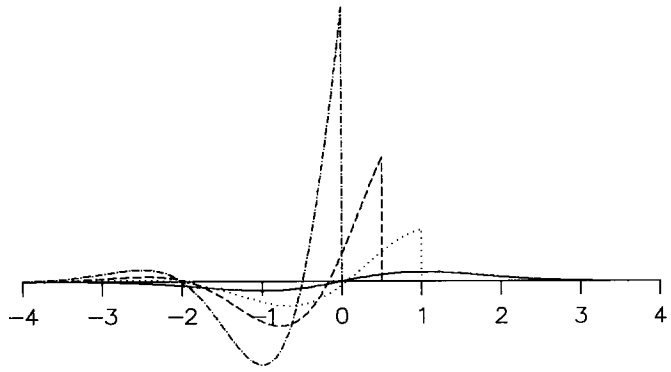

Fig. 9. *Derivative of standard normal kernel (solid line) together with quadratically corrected boundary derivative versions when $p = 1$ (dotted line), $p = 1/2$ (dashed line) and $p = 0$ (dot–dashed line)*

where $A_i(p) = c_i(p)d_2(p) - c_2(p)d_i(p)$, $B_i(p) = a_i(p)d_2(p) - a_2(p)d_i(p)$ and $C_i(p) = a_i(p)c_2(p) - a_2(p)c_i(p), i = 0.1$. We will concentrate on two special cases of this. For the first, take $L = K'$ and $M = K''$. For the second, take $L = xK$ and $M = x^2K$ (so that $c_l(p) = a_{l+1}(p)$ and $d_l(p) = a_{l+2}(p)$). Of course, there is no distinction between these special cases when K is normal, while only the latter is available when K is uniform (or triangular). These boundary kernels for K normal and for $p = 0, 1/2, 1$ and ∞ are shown in Figure 9.

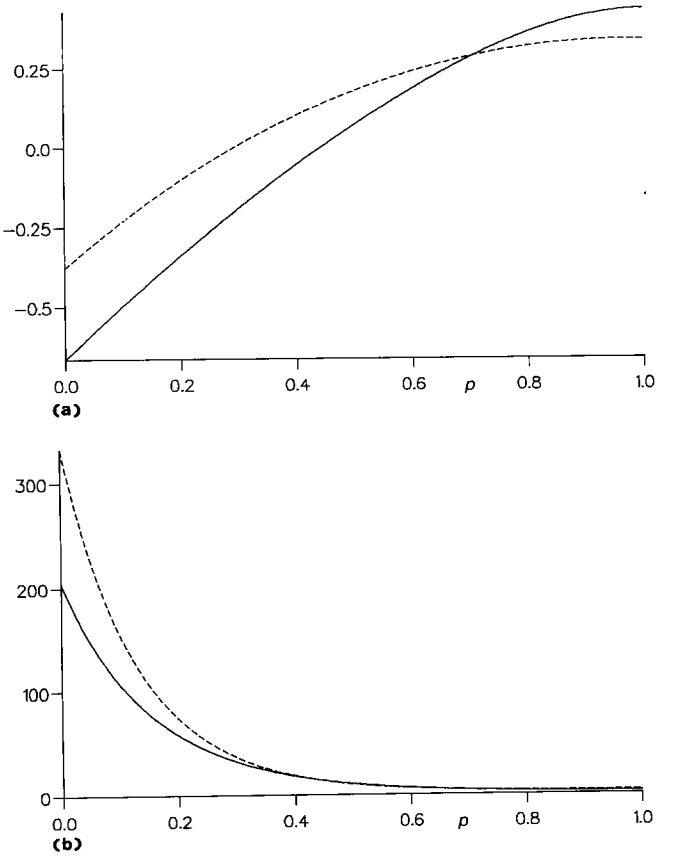

Fig. 10. *(a) $B(p)$ and (b) $V(p)$ for generalized jackknife derivative estimates based on derivatives of K (solid line) and power multiples of K (dashed line). The dotted horizontal line in (a) is at zero. Biweight kernel*

If (8.2) is written as $h^{-1}\{\alpha(p)K(x) + \beta(p)L(x) + \gamma(p)M(x)\}$, the bias coefficient $B(p) = \alpha(p)a_2(p) + \beta(p)c_2(p) + \gamma(p)d_2(p)$ (which multiplies $\frac{1}{2}h^2f'''(x)$) and the variance depends on $V(p) = \alpha^2(p)b(p) + 2\alpha(p)\beta(p)e(p) + \beta^2(p)g(p) + 2\alpha(p)\gamma(p)\int_{-1}^{p} K(x)M(x)dx + 2\beta(p)\gamma(p)\int_{-1}^{p} L(x)M(x)dx + \gamma^2(p)\int_{-1}^{p} M^2(x)dx$, where $V(p)$ is the multiplier of $(nh^3)^{-1}f(x)$. These formulae have been utilized to make comparisons of polynomial multiple and derivative methods for the biweight kernel in Fig. 10. The main messages are that (i) $B(p)$ and especially $V(p)$ are very large near $p = 0$ and that (ii) there is little to choose between the two approaches.

The reason for concentrating on just these two special cases is that they alone reduce to appealingly simple estimates of f' in the support interior. The combination of derivatives leads, unsurprisingly, to (8.1); the other combination yields the alternative derivative estimator

$$\tilde{f}'(x) = -(nh^2s_2)^{-1}\sum_{i=1}^{n}\{h^{-1}(x - X_i)\}K\{h^{-1}(x - X_i)\}, \tag{8.3}$$

i.e. use of the kernel $-uK(u)/s_2$.

The same approach can be generalized to estimating higher-order derivatives at the boundary. For instance, for estimating f'', take h^{-2} times a linear combination of K and *three* related functions (such as K', K'' and K''' or xK, x^2K and x^3K). In the interior, the latter results in the kernel $2(u^2 - s_2^2)K(u)/(s_4 - s_2^2)$. We will not pursue further details here.

These polynomial multiples of K relevant to estimating derivatives also arise from local polynomial fitting, of course. In the regression situation, the work of Ruppert and Wand (1993) shows that the first and second derivative estimates here are obtained as the slope of (either linear or) quadratic local fits, and as the quadratic term of (either quadratic or) cubic fitting, respectively.

9. Alternative boundary corrections for $O(h^2)$ bias

In this section, we consider three other simple boundary correction devices which have met with less success than generalized jackknifing, and end with references to yet other boundary methods proposed in the literature.

9.1 *Simple alternatives*

The first of these methods is based on support transformation. The idea is simply to transform the x-scale by a suitable additive constant (in addition to renormalizing). Define

$$\hat{f}_T(x) = \hat{f}(x + C(p)h)/a_0(p),$$

where $C \equiv C(p)$ solves

$$C = a_1(p + C)/a_0(p + C).$$

Then,

$$E\left\{\hat{f}_T(x)\right\} \simeq f(x) + \tfrac{1}{2}h^2[\{a_2(p + C)a_0(p + C) - a_1^2(p + C)\}/a_0^2(p + C)]f''(x)$$

and

$$V\left\{\hat{f}_T(x)\right\} \simeq (nh)^{-1}(b(p + C)/a_0^2(p + C))f(x).$$

Notice that $\hat{f}_T$ is necessarily positive everywhere whereas generalized jackknifing methods are not. Also, $C(p)$ is necessarily negative, meaning that the excess mass in the negative half-line is moved to a position above zero, and renormalized. If K is uniform on $[-1, 1]$, $C(p) = p - 1$. Unfortunately, $V(x)$ is then $(2p)^{-1}$ which tends to ∞ as $p \to 0$: therein lies the first disappointment of this technique. For non-uniform K, too, simplicity is compromised by the need to solve a non-linear equation for $C(p)$ for each p. Finally, in practice finite support kernels result in a raw density estimate extending only as far as $\min X_i - h$ while, for instance, $C(0) = -1$ and $\hat{f}(-h) = 0$; $\hat{f}_T$ is therefore necessarily zero near the boundary in places where one would often desire a non-zero estimate. Infinite support kernels ameliorate this only at the expense of much greater numerical effort and accuracy.

(As $\hat{f}_T$ is a renormalization version of the x-transformation idea, so there also exists a reflection version which only works for $p \geq 1/2$. This will not be considered further.)

Here is a second alternative approach. Generalized jackknifing (and optimal kernel theory) produces boundary kernels *that vary with point of estimation* x. An alternative is to consider utilizing different kernels *for different datapoints* X_i. The latter is quite natural in the sense that it, and not the earlier approach, results if a single picture displaying all kernels associated with the X_i's which are then averaged out to produce the final estimate is envisaged. Formulawise, the distinction is of the form

$$n^{-1}\sum_{i=1}^{n} K_{x;h}(x - X_i) \text{ versus } n^{-1}\sum_{i=1}^{n} K_{X_i;h}(x - X_i).$$

This dichotomy is the exact analogue of the distinction between local ($h \to h(x)$) and variable ($h \to h(X_i)$) bandwidth variation to improve performance of kernel estimates in the interior (Jones, 1990). Some authors, e.g. Scott (1992, Section 6.2.3.5), consider just X_i-dependent boundary kernels. Scott observes that the restriction $K(p) = 0$ imposed in optimal boundary kernel theory then necessarily results in density estimates which are zero at zero; floating boundary kernels for which $K(p) \neq 0$ are therefore particularly advantageous in this context by not resulting in this unwanted constraint.

While reflection is unaffected by the change from x-dependent kernels to X_i-dependent ones, renormalization and generalized jackknifing techniques very much are. For instance, use of $\alpha(X_i)K(u)$, which we might expect to lead to the renormalization formula $\alpha(X_i) = 1/a_0(X_i/h)$, instead demands that α satisfies $\int_{-1}^{X_i/h} \alpha(X_i/h - u)K(u)du = 1$. We do not follow this up here, but mention it to warn the reader that this approach is quite different from the one we have been mostly considering.

Thirdly, extended forms of generalized jackknifing might be contemplated. For instance, combine three kernels in such a way that the coefficients of $f(x)$, $-hf'(x)$ and $h^2f''(x)/2$ are 1, 0 and s_2, respectively. Unfortunately, the variance appears to be greatly inflated. A further extension in which four quantities are combined in the hope of achieving $B(x) = s_2$ and $V(x) = \kappa$ at all points x immediately runs into intractability problems, even in the case of a uniform kernel.

9.2 *More sophisticated alternatives*

A natural reaction of many to the boundary problem is 'transform!' While this is indeed an approach of much potential, details of how to implement it, as is the case for use of transformations in the interior (Wand *et al.*, 1991, Ruppert and Cline, 1992) are not immediately obvious. (Here, of course, we mean transform the data, estimate the density of the transformed data, and make a final change of variables.) Marron and Ruppert (1992) is a fine initial investigation into the possibilities. Simplicity *may* have to be sacrificed, although estimates based on transformations should retain non-negativity. Marron and Ruppert (1992) are also concerned with densities with poles, which other methods fail to cope with.

Two further interesting papers based on ideas specific to the case of non-parametric regression are Eubank and Speckman (1991) and Hall and Wehrly (1991). Yet another interesting recent contribution to kernel boundary problems is that of Müller (1993) in which polynomial multiples, local bandwidth variation and support transformation are all involved, wrapped up in a general weighted optimality theory.

10. Closing remarks

We have already mentioned that it is not difficult to extend this work to boundaries at points other than zero, or to a closed interval provided the bandwidth is small relative to the length of the interval. See Hart and Wehrly (1992) for interesting work when the bandwidth is relatively large (in finite support regression).

We have not yet considered automatic bandwidth selection taking account of boundary corrections. As a first approximation, one can clearly use a procedure which

ignores boundary considerations, and this we did, using Sheather and Jones' (1991) method, in Fig. 7. The truncated estimators in Fig. 5 might give us pause, however: there, any boundary-unaware estimator is likely to have greater roughness than the true exponential density because of a spurious mode near zero, with the consequence of suggesting too small a bandwidth for the boundary-corrected case.

Another question is local bandwidth choice near the boundary. Many would say it was obvious that larger bandwidths should be used there. Some simple ways of choosing h as a function of x are considered in Section 4 of Müller (1991). We mention a further alternative, which is closely related to Müller's $h_1(x)$. There is a natural 'canonical' scaling which separates bandwidth and density effects (at least asymptotically). Marron and Nolan (1989) describe this for the interior. The immediate extension to the boundary problem is to take $h(x) \propto \{B(x)^{-1}V(x)\}^{1/5}$. Note that this is independent of f, and hence immediately practicable except for $B(x)$ taking a zero value for some p. See Fig. 3 and Müller's (1991) Fig. 3.

Finally, this paper has tried to summarize, and set a framework for, simple approaches to boundary correction for kernel methods. We failed to do better than the generalized jackknifing methods, of which the linear correction (3.4) is proving to be particularly popular. For the moment, then, we agree with various proponents of this method, that it is as good as anything to use in practice. However, we still wonder if the inflated variance of such methods means that there may remain scope for improvement.

Acknowledgement

I am very grateful to Matt Wand for setting me on to the link with local linear fitting, and to Hans-Georg Müller and David Ruppert for sending me their preprints. A referee prompted minor improvements.

References

Azari, A. S., Mack, Y. P. and Müller, H.-G. (1992) Ultraspherical polynomial, kernel and hybrid estimators for nonparametric regression. *Sankhyā* **A, 54,** 80–96.

Boneva, L. I., Kendall, D. G. and Stefanov, I. (1971) Spline transformations: three new diagnostic aids for the statistical data-analyst (with discussion). *Journal of the Royal Statistical Society* B, **33,** 1–70.

Cleveland, W. S. (1979) Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association*, **74,** 829–836.

Cline, D. B. H. and Hart, J. D. (1991) Kernel estimation of densities with discontinuities or discontinuous derivatives. *Statistics*, **22,** 69–84.

Copas, J. B. and Fryer, M. J. (1980) Density estimation and suicide risks in psychiatric treatment. *Journal of the Royal Statistical Society*, A, **143,** 167–176.

Diggle, P. J. (1985) A kernel method for smoothing point process data. *Applied Statistics*, **34,** 138–147.

Diggle, P. J. and Marron, J. S. (1988) Equivalence of smoothing parameter selectors in density and intensity estimation. *Journal of the American Statistical Association*, **83,** 793–800.

Eubank, R. L. and Speckman, P. (1991) A bias reduction theorem with applications in nonparametric regression. *Scandinavian Journal of Statistics*, **18,** 211–222.

Fan, J. (1992) Design-adaptive nonparametric regression. *Journal of the American Statistical Association*, **87,** 998–1004.

Fan, J. (1993) A remedy to regression estimators and nonparametric minimax efficiency. *Annals of Statistics*, **21,** 196–216.

Fan, J. and Gijbels, I. (1992) Variable bandwidth and local linear regression smoothers. *Annals of Statistics*, **20,** 2008–2036.

Gasser, T. and Müller, H.-G. (1979) Kernel estimation of regression functions. In *Smoothing Techniques for Curve Estimation*, T. Gasser and M. Rosenblatt (eds), Springer, Berlin, pp. 23–68.

Gasser, T., Müller, H.-G. and Mammitzsch, V. (1985) Kernels for nonparametric curve estimation. *Journal of the Royal Statistical Society* B, **47,** 238–252.

Ghosh, B. K. and Huang, W.-M. (1992) Optimum bandwidths and kernels for estimating certain discontinuous densities. *Annals of the Institute of Statistical Mathematics*, **44,** 563–577.

Granovsky, B. L. and Müller, H.-G. (1991) Optimizing kernel methods: a unifying variational principle. *International Statistical Review*, **59,** 373–388.

Hall, P. and Wehrly, T. E. (1991) A geometrical method for removing edge effects from kernel-type nonparametric regression estimators. *Journal of the American Statistical Association*, **86,** 665–672.

Härdle, W. (1990) *Applied Nonparametric Regression.* Cambridge University Press.

Hart, J. D. and Wehrly, T. E. (1992) Kernel regression estimation when the boundary is large, with an application to testing the adequacy of polynomial models. *Journal of the American Statistical Association*, **87,** 1018–1024.

Hastie, T. J. and Loader, C. (1993) Local regression: automatic kernel carpentry (with comments). *Statistical Science*, **8,** 120–143.

Hominal, P. and Deheuvels, P. (1979) Estimation non paramétrique de la densité compte-tenu d'informations sur la support. *Revue Statistique Appliqué*, **27,** 47–68.

Jones, M. C. (1990) Variable kernel density estimates and variable kernel density estimates. *Australian Journal of Statistics*, **32,** 361–371. Correction, **33,** 119.

Jones, M. C. (1991) Kernel density estimation for length biased data. *Biometrika*, **78,** 511–519.

Jones, M. C. and Foster, P. J. (1993) Generalized jackknifing and higher order kernels. *Journal of Nonparametric Statistics*, to appear.

Lejeune, M. (1985) Estimation non-paramétrique par noyaux: regression polynomial mobile. *Revue Statistique Appliqué*, **33,** 43–66.

Lejeune, M. and Sarda, P. (1992) Smooth estimators of distri-

bution and density functions. *Computational Statistics and Data Analysis*, **14,** 457–471.

Marron, J. S. and Nolan, D. (1989) Canonical kernels for density estimation. *Statistics and Probability Letters*, **7,** 195–199.

Marron, J. S. and Ruppert, D. (1992) Transformations to reduce boundary bias in kernel density estimation, to appear.

Marron, J. S. and Wand, M. P. (1992) Exact mean integrated squared error. *Annals of Statistics*, **20,** 712–736.

Müller, H.-G. (1987) Weighted local regression and kernel methods for nonparametric curve fitting. *Journal of the American Statistical Association*, **82,** 231–238.

Müller, H.-G. (1991) Smooth optimum kernel estimators near endpoints. *Biometrika*, **78,** 521–530.

Müller, H.-G. (1993) On the boundary kernel method for nonparametric curve estimation near endpoints. *Scandinavian Journal of Statistics*, to appear.

Muttlak, H. A. and McDonald, L. L. (1990) Ranked set sampling with size-biased probability of selection. *Biometrics*, **46,** 435–445.

Rice, J. A. (1984) Boundary modification for kernel regression. *Communications in Statistics — Theory and Methods*, **13,** 893–900.

Ruppert, D. and Cline, D. H. (1992) Transformation-kernel density estimation — bias reduction by empirical transformations. *Annals of Statistics*, to appear.

Ruppert, D. and Wand, M. P. (1993) Multivariate locally weighted least squares regression. *Annals of Statistics*, to appear.

Sarda, P. (1991) Estimating smooth distribution functions, in *Nonparametric Functional Estimation and Related Topics*, G. G. Roussas (ed.), Kluwer, Dordrecht, pp. 261–270.

Schucany, W. R. (1989) On nonparametric regression with higher-order kernels. *Journal of Statistical Planning and Inference*, **23,** 145–151.

Schucany, W. R. and Sommers, J. P. (1977) Improvement of kernel type estimators. *Journal of the American Statistical Association*, **72,** 420–423.

Schucany, W. R., Gray, H. L. and Owen, D. B. (1971) On bias reduction in estimation. *Journal of the American Statistical Association*, **66,** 524–533.

Schuster, E. F. (1985) Incorporating support constraints into nonparametric estimators of densities. *Communications in Statistics — Theory and Methods*, **14,** 1123–1136.

Scott, D. W. (1992) *Multivariate Density Estimation; Theory, Practice and Visualization*. Wiley, New York.

Sheather, S. J. and Jones, M. C. (1991) A reliable data-based bandwidth selection method for kernel density estimation. *Journal of the Royal Statistical Society* B, **53,** 683–690.

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

Wand, M. P., Marron, J. S. and Ruppert, D. (1991) Transformations in density estimation (with comments). *Journal of the American Statistical Association*, **86,** 343–361.