Choice of the bandwidth ratio in Rice's boundary modification

Ming-Yen Cheng

Department of Mathematics

National Taiwan University

Taipei 106, Taiwan

July 11, 2006

Abstract

Rice (1984) proposed a boundary modified kernel regression method which linearly combines two kernel regression estimators based on different bandwidths. In the context of density estimation, advantages of this method over two other popular approaches, local linear fitting and the boundary kernels of Müller (1991), are discussed. Selection of the ratio of the two bandwidths is studied. Asymptotic and exact mean squared errors are provided as tools to analyze the problem. In the case of Normal kernels, keeping the bandwidth ratio fixed, for ease and speed of implementation, and a specific bandwidth ratio are suggested.

KEYWORDS: boundary effect, boundary kernel, exact mean squared error, kernel smoothing, local linear smoothing.

JEL subject classification: C14, C13.

1 Introduction

Most nonparametric smooth curve estimators have difficulties caused by boundary effects. If the support of the curve has important boundaries, then they are seriously biased in regions near the endpoints. In practice, such effects are visually disturbing and often become misleading in modeling the data. For small or moderate sample sizes, boundary area can be a substantial portion of the support. Boundary effects can also cause a slower rate of convergence in the usual asymptotic analysis of global performance.

Gasser and Müller (1979), Gasser, Müller and Mammitzsch (1985), Granovsky and Müller (1991) and Müller (1991) discussed boundary kernel approach to correct this problem. Schuster (1985) proposed a mirror image estimator which folds back the conventional kernel density estimator beyond the support of the density. This approach does not fully correct the boundary problems, see Marron and Ruppert (1994), and will not be included in our discussion. Local linear regression techniques achieve automatic boundary corrections (Fan and Gijbels, 1992) and enjoy some optimal properties (Fan, 1993). Lejeune and Sarda (1992) and Jones (1993) applied the local linear techniques to density estimation. Another approach is to bin the data and then apply local linear fitting to the bin counts. These two methods produce approximately the same estimators, see Cheng (1997). Rice (1984) suggested linearly combining two kernel regression estimators which uses different bandwidths such that the bias at the boundary area has the same order of magnitude as in the interior.

The context of density estimation is considered and the support of the density function is assumed to be known throughout. It is shown in Section 3 that, compared to local linear fitting or the approaches of Müller (1991), Rice's boundary modified estimators are typically much more stabilized; i.e. having smaller variances, while having competitive mean squared errors near the boundary. The intuition is clear: it makes use of more nearby data points. Variance inflation near the boundary is a common and serious problem for many boundary modified methods. Rice's boundary modification is advantageous in the direction of controlling the variance at the boundary area. Another advantage of Rice's boundary modified estimators is that, in comparison with the other two, they are easier and faster to compute if the ratio of the two bandwidth is kept fixed. Details are given in Section 3.

In Rice's boundary method, the bandwidth ratio decides which boundary kernel to use. Rice (1984) suggested a bandwidth ratio which depends on the location in such a way that the range of moving average has the same length everywhere. But that makes implementation of the estimator over a range difficult and slow. This choice is compared with another possibility: keeping the bandwidth ratio unchanged over the range which allows the fast binned implementation of kernel estimators, see Fan and Marron (1994). It is found that, in the sense of asymptotic mean squared error, there is essentially no loss in using the latter instead of the former. Asymptotic analysis of the mean squared error becomes noninformative when the asymptotic bias is equal to zero. Another approach is to calculate the exact mean squared error. For the Normal kernel, taking the limiting case of the bandwidth ratio being equal to one is suggested as a general rule.

Section 2 briefly discusses the conventional kernel density estimator and its boundary effects. In Section 3, Rice's boundary modified kernel regression is adopted to the setting of density estimation. Asymptotic properties and some merits of this method are also discussed. Section 4 and Section 5 study choices of the bandwidth ratio by asymptotic and exact analyses of the mean squared error.

2 Kernel Density Estimation

Suppose $X_1, ..., X_n$ is an i.i.d. sample observed from a population following a density f. A kernel estimator of f is

$$\widehat{f}_h(x) = n^{-1}h^{-1}\sum_{i=1}^{n} K(\frac{x - X_i}{h}),$$

where K is the kernel function, assumed to be a symmetric density function, and the bandwidth $h > 0$ controls the amount of smoothing. To discuss boundary effects of kernel density estimators, without loss of generality, consider that the support of the density is known and assumed to be $[0, \infty)$; i.e. $f(x) = 0$ if $x < 0$ and $f(x) > 0$ if $x \geq 0$.

As $n \to \infty$, the bandwidth h tends to zero and therefore the boundary region shrinks to the set $\{0\}$. And any fixed point other than zero is eventually not in the boundary region. Therefore we analyze the kernel estimator at points $x = ch$, $c \geq 0$, which is a sequence of points with a fixed distance, with respect to the bandwidth, from the boundary point zero. Quantification of boundary effects of $\widehat{f}_h$ is given below. Some conditions are convenient for analyses of the asymptotic mean squared error.

Condition 1.

- f has two derivatives and f'' is bounded and uniformly continuous in a neighborhood of zero or x when x is a boundary or interior point, respectively.
- K satisfies $\int K^2 < \infty$ and $\int |u^2 K| < \infty$.

Here, the first two derivatives of f at 0 are taken as limits from the right. For any real-valued function χ on R, $c \in R$ and $l = 0, 1, 2$, define $\mu_{l,c}(\chi) = \int_{-\infty}^{c} u^l \chi(u)du$ and $\mu_l(\chi) = \int_{-\infty}^{\infty} u^l \chi(u)du$. Suppose that Condition 1 holds, then for $x = ch, c \geq 0$, as $n \to \infty$, $h \to 0$ and $nh \to \infty$, $\widehat{f}_h(x)$ has expected value

$$E\{\widehat{f}_h(x)\} = \mu_{0,c}(K)f(x) - h\mu_{1,c}(K)f'(x) + \frac{h^2}{2}\mu_{2,c}(K)f''(x) + o(h^2), \tag{1}$$

and variance

$$Var\{\widehat{f}_h(x)\} = n^{-1}h^{-1}f(x)\mu_{0,c}\left(K^2\right) + o(n^{-1}h^{-1}). \tag{2}$$

See Marron and Ruppert (1994) for details.

Suppose that K is supported on $[-1, 1]$. Then for any $c \in [0, 1)$, $\mu_{0,c}(K) < 1$ and $\widehat{f}_h(x), x = ch$, as an estimator of $f(x)$, has a nonzero constant bias unless $f(0+)$ is equal to zero. And for $c \geq 1$, (1) and (2) become

$$E\{\widehat{f}_h(x)\} = f(x) + \frac{h^2}{2}\mu_2(K)f''(x) + o(h^2), \tag{3}$$

and

$$Var\{\widehat{f}_h(x)\} = n^{-1}h^{-1}f(x)\mu_0\left(K^2\right) + o(n^{-1}h^{-1}), \tag{4}$$

respectively, which are the same as the usual interior bias and variance expansions.

3 Rice's Boundary Modification

Rice (1984) proposed a boundary modification of kernel regression estimators. In the boundary area, the method takes a linear combination of two kernel regression estimators based on different bandwidths such that the bias is of the same order of magnitude as in the interior. The idea is similar to the bias reduction technique discussed in Schucany and Sommers (1977). We adapt the method to the context of density estimation.

Given $\alpha > 0$, the boundary modified kernel estimator of $f(x)$, $x = ch$, $c \geq 0$ is

$$\overline{f}_{\alpha,h}(x) = a\widehat{f}_h(x) - b\widehat{f}_{\alpha h}(x) = n^{-1}\sum_{i=1}^{n}(aK_h - bK_{\alpha h})(x - X_i), \tag{5}$$

where

$$a = \frac{\alpha\mu_{1,c/\alpha}(K)}{\alpha\mu_{0,c}(K)\,\mu_{1,c/\alpha}(K) - \mu_{0,c/\alpha}(K)\,\mu_{1,c}(K)}, b = \frac{\mu_{1,c}(K)}{\alpha\mu_{1,c/\alpha}(K)}a. \tag{6}$$

Here, a and b depend on c and are obtained by requiring $\overline{f}_{\alpha,h}(x)$ to have a bias of order h^2, see Rice (1984). With a and b as in (6), define

$$\overline{K}_\alpha(\cdot) = a\ K(\cdot) - \frac{b}{\alpha}K(\frac{\cdot}{\alpha}). \tag{7}$$

Asymptotic bias and variance of $\overline{f}_{\alpha,h}(x)$ are given in the following theorem. The proof is similar to analyses of bias and variance of the usual kernel density estimators and hence is omitted.

Theorem 1 *Under Condition 1, for $x = ch$, $c \geq 0$, as $n \to \infty, h \to \infty$ and $nh \to \infty$,*

$$Bias\{\overline{f}_{\alpha,h}(x)\} = \frac{h^2}{2}f''(0+)\mu_{2,c}\left(\overline{K}_\alpha\right) + o(h^2), \tag{8}$$

and

$$Var\{\overline{f}_{\alpha,h}(x)\} = \frac{f(0+)}{nh}\mu_{0,c}\left(\overline{K}_\alpha^2\right) + o(n^{-1}h^{-1}). \tag{9}$$

Hence $\overline{f}_{\alpha,h}$ retains the same rate of convergence in mean squared error everywhere. This method introduces an extra parameter α , the ratio of the two bandwidths. Rice (1984) recognized that it is difficult to find the best solution for each c and suggested taking

$\alpha = 2 - c$, when K is supported on $[-1, 1]$. It ensures that the estimator is a weighted average of observations falling into the interval $[0, 2h]$ which has the same length as $[x - h, x + h]$, the region of smoothing for x in the interior. However, by taking $\alpha = 2 - c$, $\overline{f}_{\alpha,h}(x)$ involves K_h and $K_{(2-c)h}$ with the latter vary as the location $x = ch$ changes. Then, a fair amount of computation is required when implementing the estimator over a range. On the other hand, if α is fixed, then for every x, $\overline{f}_{\alpha,h}(x)$ is a linear combination of two kernel estimators based on K_h and $K_{\alpha h}$. Applying the fast binned calculation of kernel estimates (Fan and Marron, 1994) to compute the two kernel estimates and then forming the linear combination we obtain the values of $\overline{f}_{\alpha,h}$ over a range. Sections 4 and 5 investigate on issues concerning choice of the bandwidth ratio.

Next we compare Rice's boundary modification with two other popular boundary correction methods: local linear fitting and Müller's boundary kernel approach. For brevity, assume the kernels for those two estimators to be K. Then, their asymptotic biases and variances are the respective versions of (8) and (9) with $\overline{K}_\alpha$ replaced by some equivalent kernels $K_c^*(u)$ and $K_+(c, u)$. Formulae for those equivalent kernels are given in Jones (1993) and Müller (1991). Each of the estimators is a conventional kernel density estimator based on its corresponding equivalent kernel. Figure 1 depicts the three equivalent kernels for four different values of c (0, 0.2, 0.4 and 0.6). The kernel K was taken as the Epanechnikov kernel $\frac{3}{4}(1 - u^2)\mathrm{I}_{(0,1)}(u)$. The bandwidth ratio α was fixed at $2 - c$, as suggested by Rice (1984), and for Müller's method we took $k = 2$ and $\mu = 1$. Figure 1 suggests use of kernels with smooth tails for Rice boundary modification. Otherwise; e.g. with the Epanechnikov kernel, the equivalent kernel $\overline{K}_\alpha$ has two sharp corners which would appear in the curve estimates. It is also suggested that, for the values of c, the norm of $\overline{K}_\alpha$ is the smallest (which means that $\overline{f}_{\alpha,h}$ has the smallest asymptotic variance) among the three equivalent kernels.

[Put Figure 1 about here please]

Figure 2 depicts the kernel parts of the asymptotic biases, variances and mean squared errors (after the canonical rescales, see (10) for the version for $\overline{K}_\alpha$) as functions of c.

The setups are the same as in Figure 2 For all $0 \leq c \leq 1$, Müller's boundary kernels yield the worst and comparatively large asymptotic bias, variance and mean squared error. Comparing the other two, they have asymptotic mean squared errors close to each other. But for $0 \leq c \leq 0.2$, Rice's boundary modified estimator has an asymptotic variance much smaller than that of the local linear estimator; i.e. the former minimizes the instability in estimating the function near the boundary. This is an important merit for Rice's boundary modification. Estimation near the boundary is difficult since there is less information given by the data; i.e. less data points available in the smoothing. Therefore boundary corrected methods, though achieve the same rate of convergence everywhere, are subject to increase in variance in the boundary region. For example, it is well known that although local linear smoothing has many appealing properties it has the problem of large variance inflation near the boundary.

[Put Figure 2 about here please]

Consider implementation of the three boundary corrected methods. Boundary kernels of Müller (1991) vary over the range of estimation, so it is comparatively hard and slow to compute the estimates. Local linear estimates can be fast computed using binned calculation. That involves five convolutions, see Fan and Marron (1994). As discussed earlier, keeping the bandwidth ratio fixed allows the Rice's boundary modified estimates to be computed with two convolutions.

Each bandwidth ratio α decides a unique shape of the boundary kernel function $\overline{K}_\alpha$. Thus choosing the best kernel from the family $\{\overline{K}_\alpha, \alpha > 0\}$ is equivalent to deciding the best bandwidth ratio among $\{\alpha > 0\}$. For any $x = ch, c \geq 0$, it can be verified that

$$\overline{f}_{\alpha,h}(x) = \overline{f}_{\alpha^{-1},\alpha h}(x)$$

for any $\alpha > 0$. Hence it suffices to choose the best bandwidth ratio among $\{\alpha \geq 1\}$. Notice that for $\alpha = 1$, $\overline{K}_{\alpha=1} \neq K$ and the both the denominators and numerators in (6) are zero. Instead $\overline{K}_{\alpha=1}$ is taken as the limiting case of (7):

$$\overline{K}_{\alpha=1}(\cdot) = \lim_{\alpha \to 1} \overline{K}_\alpha(\cdot).$$

And it is interpreted as the kernel which yields bias of order h^2, see Rice (1984), by linearly combining two kernel estimators with bandwidths that are very close to each other.

4 Asymptotic MSE Study on Bandwidth Ratio

Equations (8) and (9) indicate that the asymptotic bias and variance depend on K and α only through the quantities $\mu_{2,c}\left(\overline{K}_\alpha\right)$ and $\mu_{0,c}\left(\overline{K}_\alpha^2\right)$ respectively. The effects of $\overline{K}_\alpha$, f and h on the asymptotic mean squared error

$$AMSE\{\overline{f}_{\alpha,h}(x)\} \equiv \left\{\frac{h^2}{2} f''(0+)\mu_{2,c}\left(\overline{K}_\alpha\right)\right\}^2 + \frac{f\,(0+)}{nh}\mu_{0,c}\left(\overline{K}_\alpha^2\right),$$

can be separated from each other by applying the idea of canonical kernels introduced by Marron and Nolan (1989). Let $\delta_\alpha = \mu_{0,c}\left(\overline{K}_\alpha^2\right)^{1/5}\mu_{2,c}\left(\overline{K}_\alpha\right)^{-2/5}$, then

$$AMSE\{\overline{f}_{\alpha,\delta_\alpha h}(x)\} = \mu_{0,c}\left(\overline{K}_\alpha^2\right)^{4/5}\mu_{2,c}\left(\overline{K}_\alpha\right)^{2/5}\left\{\frac{f(0+)}{nh} + \frac{h^4}{4}f''(0+)^2\right\}.$$

Define

$$S_{K,c}(\alpha) = \mu_{0,c}\left(\overline{K}_\alpha^2\right)^2\left|\mu_{2,c}\left(\overline{K}_\alpha\right)\right|. \tag{10}$$

Then AMSE$\left(\overline{f}_{\alpha,\delta_\alpha h}(x)\right)$ is factorized into two parts: $S_{K,c}^{2/5}(\alpha)$ which depends on $\overline{K}_\alpha$, and $\{n^{-1}h^{-1}f\,(0+) + h^4 f''(0+)^2/4\}$ which involves only h and f. Thus comparison among kernels $\overline{K}_\alpha$ can be done through investigating the quantity $S_{K,c}(\alpha)$. We call δ_α, which depends on $\overline{K}_\alpha$, the canonical bandwidth for the kernel $\overline{K}_\alpha$.

Given any $c \geq 0$, the best α, in the sense of optimizing the asymptotic mean squared error, minimizes $S_{K,c}(\alpha)$ over all $\alpha \geq 1$. If there exists an α_{opt} which uniformly minimizes $S_{K,c}(\alpha)$ for every $c \geq 0$, then it would be the best constant bandwidth ratio. Typically the minimizer of $S_{K,c}(\alpha)$ depends on c. The following is an attempt to seek for some feasible solutions.

Consider $K \equiv \varphi$, the standard normal density. Note that the best α depends on the kernel. The following analyses and those in the next section can be done for kernels other than the Normal kernel. We choose to work with Normal kernels since they are popular in practice for they produce nice smooth curve estimates. In that case,

$$a = \left\{ \Phi\left(c\right) - \frac{1}{\alpha}\varphi\left(c\right)\varphi\left(\frac{c}{\alpha}\right)^{-1}\Phi\left(\frac{c}{\alpha}\right) \right\}^{-1}, b = \frac{\varphi\left(c\right)}{\alpha\varphi\left(\frac{c}{\alpha}\right)}a. \tag{11}$$

Hence

$$\mu_{2,c}\left(\overline{\varphi}_{\alpha}\right) = a\left(-c\varphi(c) + \Phi(c)\right) - b\alpha^2\left\{-\frac{c}{\alpha}\varphi(\frac{c}{\alpha}) + \Phi(\frac{c}{\alpha})\right\}, \tag{12}$$

and

$$\mu_{0,c}\left(\overline{\varphi}_{\alpha}^2\right) = \frac{1}{\sqrt{2\pi}}\left\{\frac{a^2}{\sqrt{2}}\Phi\left(\sqrt{2}c\right) - \frac{2ab}{\sqrt{1+\alpha^2}}\Phi\left(\frac{\sqrt{1+\alpha^2}c}{\alpha}\right) + \frac{b^2}{\sqrt{2}\alpha}\Phi\left(\frac{\sqrt{2}c}{\alpha}\right)\right\}, \tag{13}$$

where Φ is the c.d.f. of the standard normal density. These kernel parts, for various choices of α, of the asymptotic bias and variance are given as functions of c in Figure 3 The $\mu_{2,c}\left(\overline{\varphi}_{\alpha}\right)$ and $\mu_{0,c}\left(\overline{\varphi}_{\alpha}^2\right)$ curves with $\alpha \equiv 2$ are very close to those with $\alpha = 2 - c/4$ respectively. Hence there is essentially no gains in using the complicated bandwidth ratio $\alpha = 2 - c/4$ rather than $a \equiv 2$.

[Put Figure 3 about here please]

Figure 4 depicts $S_{\varphi,c}(\alpha)^{2/5}$, the kernel part of the asymptotic mean squared error (see (10)), as a function of α for several values of c. It is shown that $S_{\varphi,0}(\alpha)$ is minimized at $\alpha = 1$. Thus for $c = 0$ the best kernel function among $\{\overline{\varphi}_{\alpha} : \alpha \geq 1\}$ is

$$\lim_{\alpha \to 1} \overline{\varphi}_{\alpha}(u) = 2(2 - u^2)\varphi(u), u \in R.$$

Numerical study suggests that $S_{\varphi,c}(\alpha)$ is minimized at $\alpha = 1$ for any fixed $c \in [0, 0.84]$. Figure 4 provides visual evidence for some chosen values of c.

[Put Figure 4 about here please]

As for $c \in (0.84, \infty)$, the minimal value of $S_{\varphi,c}(\alpha)$ is equal to zero at some $\alpha_{opt}(c) > 1$, see Figure 3 in case of some illustrative values of c. But $S_{\varphi,c}(\alpha)$ is zero when $\mu_{2,c}\left(\overline{\varphi}_{\alpha}\right) = 0$, which implies that the asymptotic bias is of some order smaller than h^2. In that case, analysis based on $S_{\varphi,c}(\alpha)$ is meaningless and some higher order bias terms need to be included while studying the asymptotic mean squared error. We do not pursuit in that direction since the situation becomes much more difficult and the density function and its higher derivatives play important roles.

In conclusion, for the Normal kernel, we recommend taking $\alpha \equiv 1$ for the following reasons. First, Figure 3 suggests fixing α at a constant over the estimation range for simplicity and computational speed considerations. Second, as indicated in Figure 4, $\alpha = 1$ is the best choice for $c \in [0, 0.84]$, a range where the boundary effect is worse than it is in $(0.84, \infty)$. Finally, it can be shown that for any α, $\lim_{c\to\infty} \overline{\varphi}_\alpha \equiv \varphi$; i.e. there is not much difference among $\{\overline{\varphi}_\alpha : \alpha \geq 1\}$ for large values of c, and hence choice of α is a minor issue for $c \in (0.84, \infty)$. Notice that for any $c \geq 0$, $\alpha = 1$ is the limiting case

$$\lim_{\alpha\to 1} \overline{\varphi}_\alpha(u) = \frac{(2 + c^2 - u^2)\varphi(u)}{(c^2 + 1)\Phi(c) + c\varphi(c)},$$

by L'Hopital's rule. Taking $\alpha \equiv 1$ does allow the fast binned calculation of the estimates since, for any $c \geq 0$, $\overline{\varphi}_{\alpha=1}$ is a linear combination of two kernel estimates based on the kernels $\varphi(u)$ and $u^2\varphi(u)$ with coefficients $(2 + c^2)\left\{(c^2 + 1)\Phi(c) + c\varphi(c)\right\}^{-1}$ and $\left\{-(c^2 + 1)\Phi(c) - c\varphi(c)\right\}^{-1}$ respectively.

We have seen that minimizing $S_{\varphi,c}(\alpha)$ for $c \geq 0.84$ is irrelevant to the issue of choosing α, and instead higher order bias terms need to be considered. This has not been carefully investigated because there is another approach by calculating the exact mean squared error. We discuss the details in the next section.

5 Exact Mean Squared Error

When analyzing effects of the bandwidth ratio on performance of the Rice boundary modified estimator, asymptotic mean squared error analysis becomes noninformative in some cases. An alternative approach is to study the exact mean squared error. Small to moderate sample size properties are emphasized, but the results interact with knowledge learned from asymptotic studies. Mean squared error of the estimator depends on the density f. We take members from the family of truncated normal mixture densities which contains a rich variety of shapes. In addition, the kernel is chosen as the normal density. Then the bias and variance of kernel estimators are easy to compute. Four densities from the family of truncated normal mixture family are studied. Formulae for these densities are listed Table 5.1.

Table 5.1. Truncated normal mixture densities.

Density #1	$\Big\{\frac{1}{2}\Phi(-\frac{1}{5})^{-1}N(-\frac{1}{5},1)+\frac{1}{2}\Phi(-\frac{7\sqrt{2}}{10})^{-1}N(-\frac{7}{10},\frac{1}{2})\Big\}\mathrm{I}_{(0,\infty)}(x)$
Density #2	$\Big\{\frac{3}{5}\Phi(-\sqrt{\frac{3}{10}})^{-1}N(\frac{3}{10},\frac{3}{10})+\frac{2}{5}\Phi(\frac{5}{2})^{-1}N(\frac{5}{2},1)\Big\}\mathrm{I}_{(0,\infty)}(x)$
Density #3	$\Big\{\frac{4}{5}\Phi(0)^{-1}N(0,\frac{4}{9})+\frac{1}{5}\Phi(9)^{-1}N(3,\frac{1}{9})\Big\}\mathrm{I}_{(0,\infty)}(x)$
Density #4	$\Big\{\sum_{l=0}^{7}\frac{1}{8}\Phi(\frac{\mu_l}{\sigma_l})^{-1}N(\mu_l,\sigma_l^2)\Big\}\mathrm{I}_{(0,\infty)}(x),\mu_l=3\Big\{(\frac{2}{3})^l-1\Big\},\sigma_l=(\frac{2}{3})^l$

The expected value and variance of the Rice boundary modified estimator are respectively

$$E\{\overline{f}_{\alpha,h}(x)\}=\frac{1}{h}\int_0^\infty\overline{K}_\alpha(\frac{x-s}{h})f(s)ds, \tag{14}$$

and

$$Var\{\overline{f}_{\alpha,h}(x)\}=\frac{1}{nh^2}\int_0^\infty\overline{K}_\alpha(\frac{x-s}{h})^2f(s)ds-\frac{1}{nh^2}\Big\{\int_0^\infty\overline{K}_\alpha(\frac{x-s}{h})f(s)ds\Big\}^2. \tag{15}$$

Note that

$$h^{-1}\overline{K}_\alpha\left(\frac{\cdot}{h}\right)=a\varphi_h(\cdot)-b\varphi_{\alpha h}(\cdot),$$

and

$$f(x)=\sum_{l=1}^m w_l\Phi\left(\frac{\mu_l}{\sigma_l}\right)^{-1}\varphi(x-\mu_l)\,I_{(x\geq 0)},$$

for some $m\in Z^+$, $\mu_l\in R,\sigma_l>0,0\leq w_l\leq 1,l=1,...,m$ and $\sum_{l=1}^m w_l=1$. It can be shown that

$$\int_0^\infty\frac{1}{h}\overline{K}_\alpha(\frac{x-s}{h})f(s)ds=\sum_{l=1}^m w_l\Phi\left(\frac{\mu_l}{\sigma_l}\right)^{-1}\left\{aA_h^l(x)-bA_{\alpha h}^l(x)\right\}, \tag{16}$$

and

$$\int_0^\infty h^{-2}\overline{K}_\alpha(\frac{x-s}{h})^2f(s)ds=\sum_{l=1}^m\frac{w_l}{\sqrt{2\pi}h}\Phi\left(\frac{\mu_l}{\sigma_l}\right)^{-1}\left\{\frac{a^2}{\sqrt{2}}A^l_{\frac{h}{\sqrt{2}}}(x)-\right.$$

$$\frac{2ab}{\sqrt{1+\alpha^2}} A^l_{\sqrt{\frac{\alpha^2 h^2}{1+\alpha^2}}}(x) + \frac{b^2}{\sqrt{2}\alpha} A^l_{\frac{\alpha h}{\sqrt{2}}}(x) \Bigg\}, \tag{17}$$

where

$$A^l_t(x) = \frac{1}{\sqrt{t^2+\sigma_l^2}} \varphi\left(\frac{x-\mu_l}{\sqrt{t^2+\sigma_l^2}}\right) \Phi\left(\frac{t\sigma_l^{-1}\mu_l + \sigma_l t^{-1} x}{\sqrt{t^2+\sigma_l^2}}\right), t > 0, l = 1, ..., m.$$

The above results can be derived using formulae given in Aldershof, et al. (1992). Combining (14) - (17), we obtain an explicit formula for the mean squared error of the Rice boundary modified estimators.

Next, exact mean squared error is used to investigate selection of the bandwidth ratio. We consider an equally spaced grid of values of α, ranging from 1 to 5, and $x = 0$. (For boundary points other than $x = 0$, one can consider $x = \eta n^{-1/5}$ for some $\eta > 0$.) The bandwidth is taken as $h_{f,\alpha} = C_f n^{-1/5} \delta_\alpha$, where C_f is some constant, depending on the density, chosen to be visually appropriate for some simulated data sets. For the densities #1, #2, #3 and #4, $C_f = 0.7, 1.2, 0.5$, and 1 respectively. Notice that these bandwidths are multiples of the canonical bandwidths. Hence comparison among kernels does not confound with the effects of the density and the choice of bandwidth on the mean squared error.

The mean squared error is increasing in α for densities #1 and #3. It implies that $\alpha = 1$ is best for these two densities under all of the five sample sizes. On the other hand, the mean squared error is always improved with larger α when the density is either density #2 or #4. The following provides some insights. For any $1 \le \alpha_1 < \alpha_2$, kernel $\overline{\varphi}_{\alpha_1}$ has a deeper negative right-hand-side tail than $\overline{\varphi}_{\alpha_2}$ does. Since density #1 achieves its maximum at $x = 0$ and then decreases, a kernel estimator of $f(0)$ has a negative bias. The value of density #1 is large, which means more data points, in the region where $h_{f,1}^{-1}\overline{\varphi}_1(\cdot/h_{f,1})$ is larger than $h_{f,2}^{-1}\overline{\varphi}_2(\cdot/h_{f,2})$. This implies that $\overline{f}_{1,h_{f,1}}(0)$ is larger and therefore better than $\overline{f}_{2,h_{f,2}}(0)$. Similarly, density #1 is relatively low at places where $h_{f,1}^{-1}\overline{\varphi}_1(\cdot/h_{f,1})$ is more negative than $h_{f,2}^{-1}\overline{\varphi}_2(\cdot/h_{f,2})$, which implies that there are rarely any observations that lower the estimate. The situation is similar when estimating density #3 at $x = 0$.

In the case of estimating density #2 at $x = 0$, the negative tails of the kernels happen at a region where the value of the density is large and hence makes the estimates too small.

So a larger α, which results in a lighter negative tail of the kernel $\overline{K}_\alpha$, is preferred.

Finally, the negative tail and higher positive part of $h_{f,1}^{-1}\overline{\varphi}_1\left(\cdot/h_{f,1}\right)$ are both less important for estimating density #4 at $x=0$. Since the density is negligible in those regions. On the other hand, $h_{f,2}^{-1}\overline{\varphi}_2\left(\cdot/h_{f,2}\right)$ is relatively higher at a very small region right next to $x=0$ and the density is extremely large there. Hence $\alpha=2$ is superior to $\alpha=1$ in this case.

When the kernel is Gaussian, our asymptotic studies recommend taking $\alpha\equiv 1$. And the studies on the exact mean squared errors suggest that the "improved" positive part of $\overline{\varphi}_1$and its relatively small support are preferred while estimating at locations where the density has a moderate peak. Hence $\alpha\equiv 1$ is recommended as a general choice.

In conclusion, we discussed advantages of Rice's boundary modification. For that method, best choice of the bandwidth ratio α depends on the density, the sample size, the kernel and the location in a complicated way. We provided both asymptotic and exact formulae of the mean squared errors to analyze the problem. We also performed some analyses in the case of Normal kernel and made some useful suggestions.

References

[1] Aldershof, B., Marron, J. S., Park, B. U. and Wand, M. P. (1992). Facts about the Gaussian probability density function. Manuscript.

[2] Cheng, M.-Y. (1997). Boundary aware estimators of integrated density derivative products. *J. Roy. Statist. Soc. Ser. B.*, **59**, 191 – 203.

[3] Fan, J. (1993). Local linear regression smoothers and their minimax efficiency. *Ann. Statist.*, **21**, 196 – 216.

[4] Fan, J. and Gijbels, I. (1992). Variable bandwidth and local linear regression smoothers. *Ann. Statist.*, **20**, 2008 – 2036.

[5] Fan, J. and Marron, J. S. (1994). Fast implementation of nonparametric curve estimators. *J. Computat. Graph. Statist.*, **3**, 35 – 56.

[6] Gasser, Th. and Müller, H. G. (1979). Kernel estimation of regression functions. In *Lecture Notes in Mathematics*, **757**, 23 – 68. New York: Springer.

[7] Gasser, Th. , Müller, H. G. and Mammitzsch, V. (1985). Kernels for nonparametric curve estimation. *J. Roy. Statist. Soc. Ser. B*, **47**, 238 – 252.

[8] Granovsky, B. L. and Müller, H. G. (1991). Optimizing kernel methods: A unifying variational principle. *Internat. Statist. Rev.*, **59**, 373 – 388.

[9] Jones, M. C. (1993). Simple boundary correction for kernel density estimation. *Statist. Comput.*, **3**, 135 – 146.

[10] Lejeune, M. and Sarda, P. (1992). Smooth estimators of distribution and density functions. *Comput. Statist. Data Anal.*, **14**, 457 – 471.

[11] Marron, J. S. and Nolan, D. (1989). Canonical kernels for density estimation. *Statist. Probab. Lett.*, **7**, 195 – 199.

[12] Marron, J. S. and Ruppert, D. (1994). Transformations to reduce boundary bias in kernel density estimation. *J. Roy. Statist. Soc. Ser. B*, **56**, 653 – 671.

[13] Marron, J. S. and Wand, M. P. (1992). Exact mean squared error. *Ann. Statist.*, **20**, 712 – 736.

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

[15] Rice, J. (1984). Boundary modification for kernel regression. *Comm. Statist. A – Theory Methods*, **13**, 893 – 900.

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

[17] Schuster, S. J. (1985). Incorporating support constraints into nonparametric estimators of densities. *Comm. Statist. A – Theory Methods*, **14**, 1123 – 1136.

[18] Silverman, B. W. (1986). *Density estimation for statistics and data analysis.* Chapman and Hall.

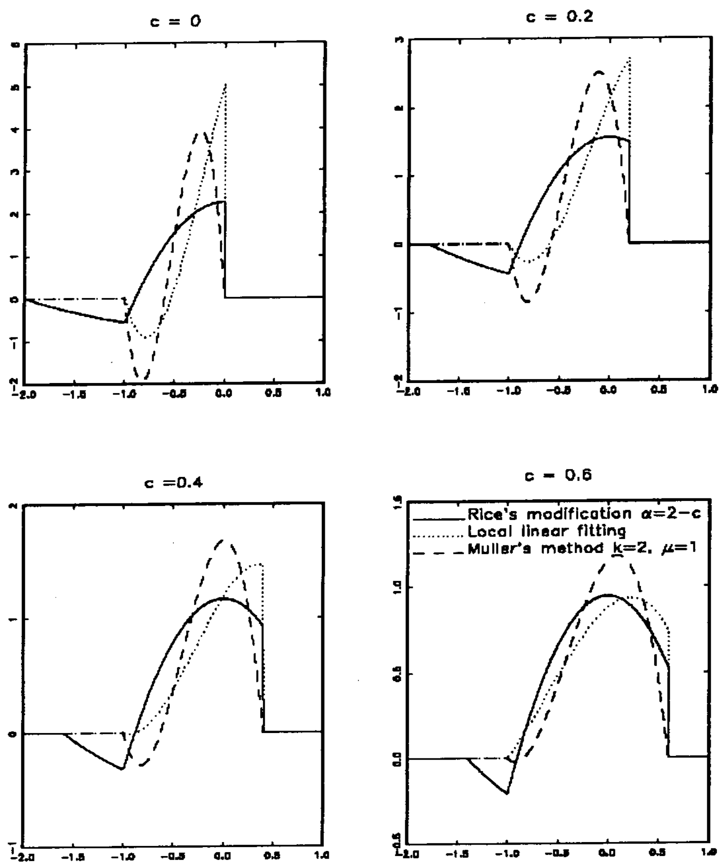

Figure 1: Equivalent kernels from Rice's boundary modification with $\alpha = 2 - c$ (solid), local linear fitting (dotted) and Müller's method with $k = 2$, $\mu = 1$ (dashed).

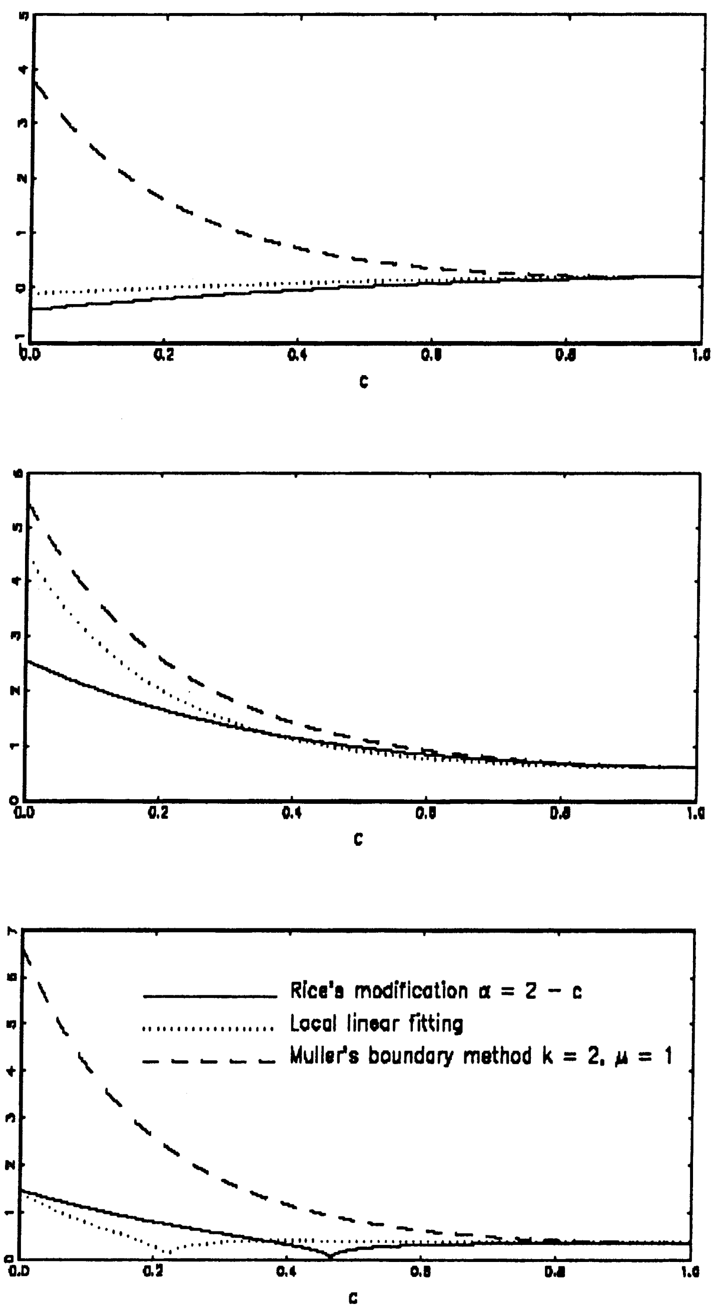

Figure 2: Kernel parts in the asymptotic biases (top), asymptotic variances (middle) and asymptotic mean squared errors (bottom) of the Rice's boundary estimator with $\alpha = 2 - c$ (solid), local linear estimator (dotted) and Müller's estimator with $k = 2$, $\mu = 1$ (dashed).

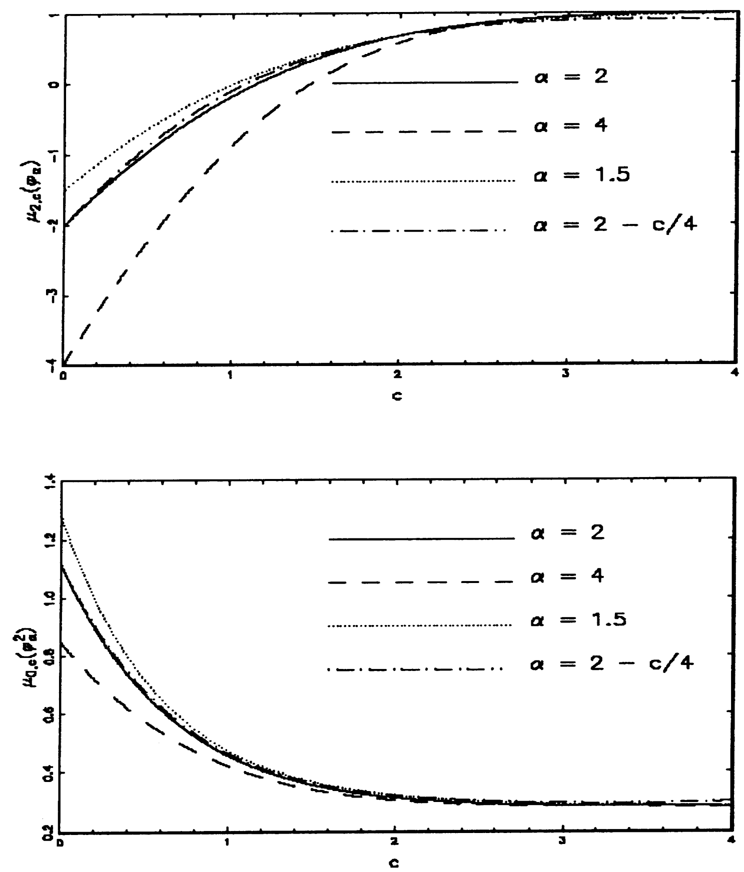

Figure 3: Kernel parts $\mu_{2,c}\left(\overline{\varphi}_{\alpha}\right)$ and $\mu_{0,c}\left(\overline{\varphi}_{\alpha}^{2}\right)$ in the asymptotic bias and variance of Rice's boundary modified density estimator at $x = ch$, given as functions of c for $\alpha \equiv 2$ (solid line), 4 (dashed), 1.5 (dotted) or $2 - c/4$ (dotted-dashed).

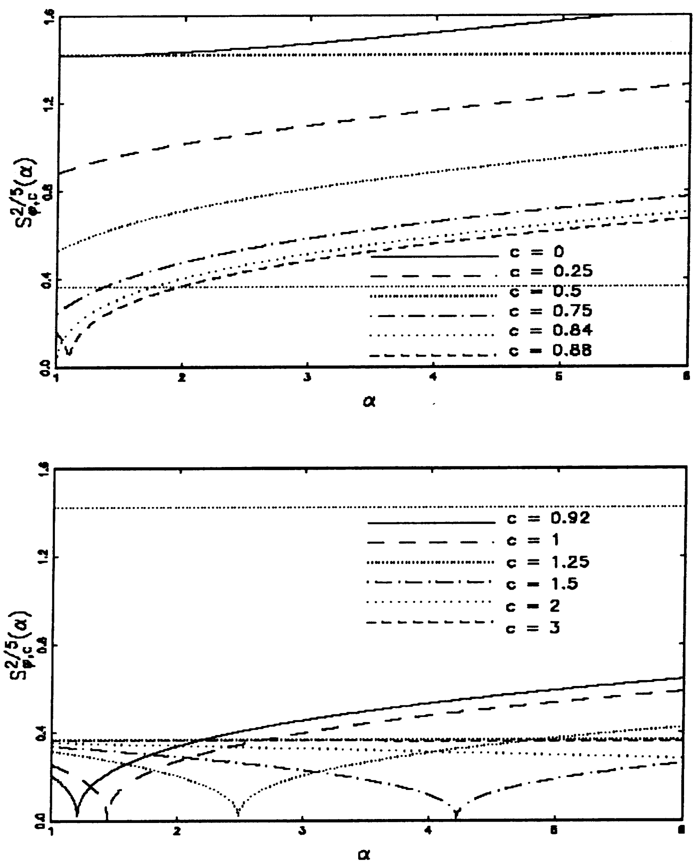

Figure 4: Kernel part, i.e. $S_{\varphi,c}^{2/5}(\alpha)$, in the mean squared error of Rice's boundary modified density estimator at $x = 0$ is plotted as a function of α for various c.