THE INSTITUTE OF STATISTICS

THE CONSOLIDATED UNIVERSITY
OF NORTH CAROLINA

ON BOUNDARY EFFECTS OF SMOOTH CURVE ESTIMATORS

(Dissertation)

by

Ming-Yen Cheng

April 1994

Mimeo Series #2319

DEPARTMENT OF STATISTICS
Chapel Hill, North Carolina

Mimeo Series # 2319
Ming-Yen Cheng
On Boundary Effects of Smooth Curve Estimators (Dissertation)

NAME	DATE

On Boundary Effects of Smooth Curve Estimators

Ming-Yen Cheng

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Statistics.

Chapel Hill, 1994

Approved by:

Advisor

Reader

Reader

Ming-Yen Cheng, On Boundary Effects of Smooth Curve Estimators
(under the supervision of Professor James S. Marron)

Abstract

Many nonparametric smooth curve estimators have a problem with boundary effects. Roughly speaking, the discontinuity of the curves under investigation at their endpoints causes difficulties for this kind of estimators. These estimators are visually disturbing at boundary regions and can become misleading in modeling the data because they are seriously biased there. In applications, boundary regions can be a substantial portion of the entire support. This has been recognized as an important problem and there are many adjustments suggested in the literature. We investigate properties of Shuster's boundary fold method and Rice's modification. Noticing the automatic boundary adaptive property of the local linear smoother recently highlighted by Fan, we further find out it is 100% efficient; i.e. best out of all possible estimators, for estimation at endpoints in a typical minimax sense. This result is important since it shows in one step that the local linear approach is as good as or better than all of the many other approaches proposed in the literature. The problem of choosing an appropriate smoothing parameter is crucial for the local linear estimators. We develop a simple and efficient bandwidth selector based on the plug-in idea and show that it works both in the boundary and non-boundary cases. A simulation study on the practical performance of the bandwidth selector is conducted. Finally, it is applied to analyzing two real data sets.

Acknowledgements

I am very grateful to Dr.s J. Stephen Marron and Jianqing Fan for their inspiring guidance and enthusiastic support in this research.

I would like to thank the other members of the committee (Professors E. Carlstein, P. K. Sen, and Y. Truong) for their careful reading of the manuscript and helpful suggestions. I would also like to acknowledge the people who supported me throughout my graduate studies.

Contents

List of Figures

Chapter 1

Introduction

Nonparametric curve estimation methods make no assumptions on the functional form of the curves of interest and hence allow flexible modeling of the data. Most such curve estimators have difficulties caused by boundary effects; i.e. if the support of the true curve has important boundaries then they produce estimates that are severely biased in regions near the endpoints. Visual evidence of this for the conventional kernel density estimator is provided in Section 2.1. These effects are visually very disturbing in practice and affect the global performance in terms of a slower rate of convergence in the usual asymptotic analysis. Also these estimators often become misleading in modeling the data in the boundary regions. In applications, for moderate sample sizes, boundary areas can be a substantial portion of the entire support.

This has been recognized as a serious problem and much research has been devoted to reduce its effect. Gasser and Müller (1979), Gasser, Müller and Mammitzsch (1985), Granovsky and Müller (1991), Müller (1991a), and Müller (1991b) discuss applying boundary kernels to correct this problem; i.e. kernels that retain the same rate of convergence as in the interior, for the conventional kernel estimators with efforts in finding "optimal boundary kernels". Rice (1984) suggests linearly combining two usual kernel estimators with different bandwidths such that the boundary bias is reduced to the same order as in the interior. Schuster's (1985) mirror image estimator in density estimation "folds back" the estimated probability mass that extends beyond the support. The estimator introduced in Hall and Wehrly (1991) is essentially a more sophisticated regression version of Schuster's approach.

Djojosugito and Speckman (1992) approach boundary bias reduction based on a finite-dimensional projection in a Hilbert space. Marron and Ruppert (1993) propose to transform the data to a density that can be estimated by conventional kernel estimators with essentially no boundary effects, and estimate the original density by a change-of-variable. Boundary effects for smoothing splines are discussed in Rice and Rosenblatt (1981). Eubank and Speckman (1991) also provide some boundary correction methods. Fan and Gijbels (1992) point out that the local linear regression smoothers correctly handle boundary effects automatically, without the complicated adjustments required by other smoothers. Although, as we have seen, there is a variety of methods for tackling the boundary problem, questions like which is the best are still open for discussion.

We investigate boundary effects in the context of density estimation. Here the locations of the boundaries are assumed to be known. First, we look into some technical problems, and analytical and practical behaviors for several of the estimators mentioned above. Given in Section 2.2, the boundary folding technique of Schuster (1985), achieves some improvement in bias, yet is still inferior to the performance of the usual kernel density estimator in the interior. Also, if the kernel function is even and differentiable, this kind of adjusted estimates always have zero first derivative at the endpoints. Such a property is undesired since the estimates can not model boundary features well but simply assume the curve is flat there.

Section 2.3 studies the Rice (1984) boundary modification. It linearly combines two kernel estimators with different bandwidths so that it successfully retains the rate of convergence in mean squared error. But the question of how to choose the ratio of the two bandwidths remains. It is suggested there to let the ratio depend on the location in such a way that the estimator is a weighted average of observations falling into an interval of length two times the bandwidth. But, this makes the implementation very difficult. We study the possibility of using one bandwidth ratio everywhere. The kernel decides the coefficients in the linear combination. Hence the optimal bandwidth ratio depends on the kernel function and the location and we choose to work with the Gaussian kernel. Asymptotic study becomes noninformative when the asymptotic bias equals zero. This motivates calculation of the exact mean squared errors. Some suggestion on the bandwidth ratio is made based on these studies.

Local linear regression smoothers are known, see Fan and Gijbels (1992), to give very good visual performance at boundary regions and have an asymptotic rate of convergence which is as good as at interior points. We implement this idea to density estimation in Section 2.4. Binning of the data is essential for this purpose; it provides intuitive interpretations of the bin centers as the design points and the bin counts as the responses in the regression setting. Also each of the bin counts reflects the value of the density at the bin center. Hence, it is natural to import the local linear regression idea here by using locally linear fits to the bin counts. The resulting estimator has the same boundary adaptive property as the local linear regression smoother. Besides, data binning has the advantage of very large savings in computation time over direct-evaluation kernel estimators, see Fan and Marron (1994). Wei and Chu (1993) discuss density estimators based on the same motivations but pay more attention to asymptotic homogeneity of variation errors.

It is then shown in Section 3.3 that a local linear estimator is asymptotically efficient even in the deep sense of best possible constant coefficient for estimating density functions at endpoints by minimax lower bound results. A similar result for regression function estimation is also available, see Cheng, Fan and Marron (1993). This result settles the important question of how the local linear estimators compare with the much more complicated "optimal boundary kernels", by showing the former must be at least as efficient. The results are extended to estimation of derivatives in Section 3.6, since this is vital to applications in plug-in bandwidth selection. In Fan (1993), a local linear regression smoother is shown to be best among all linear estimators in a minimax sense at an interior point. The analog of this property for the local linear density estimator holds. Hence, local linear estimators are very efficient both at interior and boundary points. We feel this gives the local linear estimators an important advantage, because they are also easier to interpret, much easier to implement, appear far faster to compute (to factors of 100, see Fan &

Marron (1994)), and, unlike many other nonparametric smoothers, they do not require any adjustments to achieve boundary correction.

The bandwidth used for the local least squares fits controls the neighborhood of the smoothing and hence is crucial. Fan and Gijbels (1992a) and (1995) discuss location varying bandwidth selection procedures while Ruppert, Sheather and Wand (1993) propose some global bandwidth selectors for local linear regression smoothers. In Chapter 4, we discuss this issue in the setting of estimating a density function through applying a local linear smoother to the bin counts. The goal of this research is to find a simple and effective global bandwidth selection procedure for such estimators.

Inspired by the outstanding performance of the Sheather and Jones (1991) bandwidth selector for the conventional kernel density estimators, as reported in Jones, Marron and Sheather (1994), we attempt to bring in the plug-in idea. The Sheather and Jones bandwidth selector is motivated from an explicit expression of the asymptotic optimal bandwidth for kernel estimators. In Section 4.2, we examine the integrated mean squared error of the local linear estimators and show that such an expression is valid for them also. Hence, it is natural to consider this style of plug-in approach for the local linear estimators as well.

In the plug-in bandwidth procedures, estimating the integrated squared density second derivative is very important. Kernel type estimators of those quantities are discussed in, for example, Hall and Marron (1987), Jones and Sheather (1991), and Aldershof (1991). But, these estimators are relatively inefficient when the density or its derivative is discontinuous at the boundaries. The main reason is that the non-smoothness of the density at the boundaries introduces an extra bias term which dominates the mean squared error, see Van Es and Hoogstrate (1993a) for detailed discussion. A consequence of such inefficiency is that the plug-in bandwidth does not behave like the asymptotic optimal bandwidth and instead is of some different order as the sample size increases, see Van Es and Hoogstrate (1993b). Therefore, we need to explore some estimators of the integrated squared density derivatives which are more robust to boundary problems so the resulting bandwidth selector will work in both boundary and non-boundary cases.

In Section 4.3, the boundary problem for estimating integrated squared density derivatives is overcome by introducing and utilizing density derivative estimators obtained from locally weighted polynomial fits. The motivation is that such density derivative estimators retain the same rate of convergence at boundary regions, as given in Section 3.5. Indeed, they are shown in Section 3.6 to be as efficient as any other linear estimator in a typical minimax sense. The estimators of integrated squared density derivatives are basically Riemann sums of the squared local polynomial derivative estimators. Asymptotic properties of these estimators are further investigated. They retain the rate of convergence in mean squared error familiar from non-boundary cases, and the constant coefficients have similar forms. Their asymptotic optimal bandwidths depend on integrated products of density derivatives which are included in this study.

Bickel and Ritov (1988) give information bounds, assuming certain smoothness of the density function, for nonparametric estimation of the quantities $\theta_{\gamma,\gamma}$ and provides estimators which attain the best $\sqrt{n}$–convergence when it is possible. The estimators we present here can achieve the same best $\sqrt{n}$–consistency if the density has some more smoothness in its support. On the other hand, notice that our estimators need smoothness of the density only in its support but those of Bickel and Ritov (1988) require that over the real line. However, nothing has been said about information bounds for classes of densities that are smooth except for possible jumps of the density or its derivatives. We conjecture that similar information bounds are available for these more general classes of densities and the estimators presented here will be the corresponding best estimators.

Once we have efficient estimators of the integrated squared density second derivative and their asymptotically optimal bandwidths, we apply them to the plug-in bandwidth procedure in Section 4.4. Weak convergence and asymptotic normality of the resulting bandwidth are established. Interestingly, the rate of convergence to the optimal bandwidth depends on the sign of the integrated product of the second and fourth derivatives of the density: a rate of $n^{-5/14}$, same as Sheather and Jones

(1991) in non-boundary cases, is achieved if the quantity is negative and otherwise we have a slower rate of $n^{-2/7}$. Note that if the underlying density essentially has no boundary features, the above mentioned functional is less than zero. For example, suppose its second and third derivatives both vanish at the endpoints. Then applying an integration by parts shows the quantity is minus the integrated squared density third derivative. Such a difference in rate of convergence already appears while estimating the integrated squared density second derivative and is carried over to the bandwidth selector. However, the bandwidth selector is always consistent to the optimal bandwidth no matter whether there is a nonsmooth boundary or not. Ruppert, Sheather and Wand (1993) develop a bandwidth selector for local linear regression estimator based on similar ideas. Theory as deep as that of this paper has not been developed for their selector, but it is conjectured that analogous behavior will be observed.

The motivation for this bandwidth selector is entirely analogous to that of the plug-in rules for kernel density estimation, for example those discussed in Park and Marron (1990) and Sheather and Jones (1991). But we show that such an approach is appropriate for the local linear estimators which are appealing in overcoming boundary effects without any extra adjustment. Moreover, the strength of our approach is that it applies to both cases either the density derivatives at boundaries are smooth or not. This is well contrast with the method of Sheather and Jones (1991) which is only applicable to the case that the derivatives at boundaries are continuous.

In Section 4.5, simulation on the behavior of our proposed bandwidth selector is conducted for both boundary and non-boundary densities. It is very close to the Sheather and Jones (1991) bandwidth selector in the non-boundary case. It is seen to be performing nicely in boundary cases as well. Finally, it is applied to two real data sets and seems to be satisfactory.

Chapter 2

Boundary Effects

It is common to many smooth curve estimators that they are less efficient or even inconsistent when estimating at points near the boundaries. The range at which such a problem presents itself is called the boundary region. It, and hence how to define the boundary points, depends on the estimator under investigation. In this chapter, we start by discussing the motivation for our current study; boundary effects for smooth nonparametric curve estimators. Kernel density estimation setting is chosen for this purpose and the subsequent developments will be devoted to density estimation solely. One benefit of considering the density estimation setting is that ideas and intuition can be conveyed clearer, meanwhile results obtained from this context can be extended to others such as regression without much difficulty. Following that, the boundary fold and Rice's modified estimators are reviewed and their boundary behaviors are examined. Local linear regression estimators are boundary adaptive, c.f. Fan (1992). Data binning produces statistics parallel to responses so that local linear fit technique can be used in estimating densities. The resulting estimators are shown to have the same property of automatic boundary adjustment. Finally, a simulation is conducted to compare the four estimators.

2.1 Conventional Kernel Estimators

Kernel density estimators have been popular since they are conceptually natural and simple. Such an estimator is constructed by assigning equal probability mass to a neighborhood of each data point. A weight function, usually called the kernel function, determines how the probability mass is distributed around one observation. For example, in Figure 2.1a, we have a sample of size five taken from a density having support $[0, \infty)$, shown as the dotted and dashed line. The arrows locate the data on the horizontal axis. Shifting the kernel function around by centering it at the observations distributes the total probability of one into intervals around the data. Adding up those individual kernels gives the dashed line, a kernel estimate of the density. Formally, 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

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

where K is the kernel function, usually a symmetric density, and h is called the bandwidth which decides the scale of K and hence controls the amount of smoothing in $\hat{f}_h$. Intuitively the bandwidth h should be small because only observations nearby a point contain information about the value of f there. Indeed bandwidths which are too big will smooth out any local structure and result in less informative estimates. On the other hand, if h is too small, the estimator inherits too much variability from the data and looks too wiggly. Further discussion of bandwidth selection for kernel estimators is deferred to Chapter 4. Papers discussing this issue include Jones, Marron and Sheather (1994) and the references therein.

Figure 2.1a also provides some insights into boundary effects of kernel estimators. First, it is very unpleasant that some estimated probability mass is spread outside the support of the density. This is easy to understand because those observations that are close to the endpoint have neighborhoods below zero which receive positive probability mass. Moreover, the estimate does not reflect the shape of the density well; it fails to detect the discontinuity at $x = 0$ and suggests the density is much lower near the endpoint. This can be misleading in modeling the population where the data came from. These phenomena are common for kernel estimators in boundary cases. Such disturbing problems are quantified by the unusually large amount of bias from the true curve in the boundary regions. But, before such justifications are given, binned approximation to kernel estimators is addressed because it demands much less computing time, which is very desirable in application.

2.1.1 Fast Binned Approximation to Kernel Estimators

Obviously to calculate the kernel estimator at a point x we need only to sum up evaluations of the kernel function at the distances from x to the data points. But, in order to provide full information of the curve, we usually calculate the density estimates on a dense set of grid points. Suppose there are g grid points then there are $n \cdot g$ kernel evaluations. The binning approach requires much less calculation and gives results that are close to the exact kernel estimates. Its idea is to evaluate the estimate at a set of equally spaced bin centers and replace each observation by its nearest bin center. Then big savings in kernel evaluations are achieved.

Suppose $\{x_1, x_{2,...,}x_g\}$ is a collection of equally spaced bin centers and $b > 0$ is the bin width, then the set $\{c_1, c_2, ..., c_g\}$ with $c_i = \sum_{j=1}^{n} I_{[x_i - b/2, x_i + b/2)}(X_j)$ contains the counts of observations near the bin centers. The amount of computation needed for this data binning is of the order n, see Fan and Marron (1994). After binning, the data is transformed into the set $\{\underbrace{x_1, ..., x_1}_{c_1}, ..., \underbrace{x_g, ..., x_g}_{c_g}\}$. Then, the kernel estimate at x_j based on these transformed data is

$$\hat{f}_{h,binned}(x_j) = n^{-1} \sum_{i=1}^{g} c_i K_h (x_j - x_i). \tag{2.1}$$

Here, and throughout this article, the notation $K_h(\cdot)$ is referred to the rescale of the kernel function K, $\frac{1}{h}K(\frac{\cdot}{h})$. Furthermore, since the bin centers are equally spaced, if K is an even function, then

$$\hat{f}_{h,binned}(x_j) = n^{-1} \sum_{i=1}^{g} c_i K_h (|i - j| \, b).$$

Observe that i and j both range over $\{1, 2, .., g\}$ hence $|i - j|$ takes value in the set $\{0, 1, ..., g-1\}$. Therefore there are only g kernel evaluations needed and the estimates at the bin centers $\{x_1, x_{2,...,}x_g\}$ are simply convolutions between $n^{-1}\{c_1, c_2, ..., c_g\}$ and $\{K_h(0), K_h(b), ..., K_h((g-1)b)\}$. Estimator $\hat{f}_{h,binned}$ is different from $\hat{f}_h$ since the former is based on the binned data and the latter on the original data. But, they are very close to each other visually. We refer to Fan and Marron (1994) for more discussion. Theoretical study on the binned version of kernel density estimators can be found in, for example, Jones (1989) and Hall and Wand (1993).

2.1.2 Boundary Behaviors of Kernel Estimators

The boundaries in the support of the density are assumed to be known. We consider that the underlying density f has support $[0,\infty)$, i.e. $f(x) = 0$ if $x < 0$ and $f(x) > 0$ if $x \geq 0$, while discussing boundary effects for estimators of f. There is no loss of generality from such an assumption since the right-hand-side and left-hand-side boundary adjustments are symmetric and the behaviors of kernel estimators at a point only have to do with its distance to the endpoints. Then the boundary region is

$$\{x : x \geq 0, \int_{-\infty}^{\frac{x}{h}} K(u)du < 1\}, \tag{2.2}$$

c.f. (2.3). For example, if the kernel K is positive only on $[-1,1]$, the boundary region is the interval $[0,h)$. Since the bandwidth h tends to zero as the sample size grows, the boundary region shrinks to the set $\{0\}$ as $n \to \infty$. Hence any fixed point besides zero is eventually not in the scope of the boundary region. In order to understand how estimators perform near zero, the endpoint, we will analyze the kernel estimator at points $x = ch$, $c \geq 0$. This is a series of points tending to zero when n becomes large and are close to zero with distance c relative to the shrinking bandwidth. The motivation for studying points $x = ch$ is then to analyze estimators at a "fixed distance with respect to the bandwidth" from the endpoint.

Some assumptions on f and K are necessary for asymptotic studies.

(A1) f has two derivatives and f'' is bounded and uniformly continuous on a neighborhood of zero (x) when estimating at a boundary (an interior) point.

(A2) $\int K^2 < \infty$ and $\int |u^2 K| < \infty$.

(A3) $\int K = 1$ and $\int uK = 0$.

Here, since $f(x) = 0$ if $x < 0$, the derivatives of f at 0 is taken from the right. For example the first derivative at zero is defined as

$$f'(0) \equiv \lim_{\delta \downarrow 0} \frac{f(\delta) - f(0)}{\delta}.$$

If f and K satisfy conditions (A1)-(A2) and $nh \to \infty$, $h \to 0$, then the expected value of the estimator $\hat{f}_h(x)$, $x = ch$, is

$$E\hat{f}_h(x) = \mu_{0,c}(K)f(x) - h\mu_{1,c}(K)f'(x) + \frac{h^2}{2}\mu_{2,c}(K)f''(x) + o(h^2), \qquad (2.3)$$

and the variance is

$$Var\left(\hat{f}_h(x)\right) = n^{-1}h^{-1}f(x)\int_{-\infty}^{c} K^2(u)\,du + o(n^{-1}h^{-1}), \qquad (2.4)$$

where $\mu_{l,c}(K) = \int_{-\infty}^{c} u^l K(u)du$, $l = 0, 1, \ldots$. In particular, if x belongs to the interior of the support and condition (A3) holds, then the above expressions reduce

$$E\hat{f}_h(x) = f(x) + \frac{h^2}{2}\mu_2(K)f''(x) + o(h^2), \; x \text{ in the interior}, \qquad (2.5)$$

and

$$Var\left(\hat{f}_h(x)\right) = n^{-1}h^{-1}f(x)\int K^2 + o(n^{-1}h^{-1}), \; x \text{ in the interior}, \qquad (2.6)$$

where $\mu_l(\zeta) = \int_{-\infty}^{\infty} u^l \zeta(u)du$ for any function ζ and $l = 0, 1, \ldots$ Notice that, if K is nonnegative, $\mu_{0,c} < 1$ for any c in the support of K. Therefore, from (2.3), the usual kernel density estimator has a nonzero constant bias term, i.e. 0-order bias, as long as $f(x) \neq 0$. This means that, no matter how carefully the bandwidth is chosen, the mean squared error always has a dominant constant term and therefore never converges to zero.

One can immediately see from (2.5) and (2.6) that the asymptotically optimal bandwidth is of the order $n^{-1/5}$ and bandwidths of any other order will result in a slower rate of convergence. Combining this fact with (2.2), if one chooses to formulate the boundary points as a sequence of points which tend to zero at a certain rate, then it needs to be $x = dn^{-1/5}$ for some nonnegative constant d. Otherwise, if x

tends to zero faster or slower, the problem degenerates to estimating at the endpoint or becomes estimating in the interior, respectively. Then one can replace $x = ch$ by $x = dn^{-1/5}$ in the asymptotic study with same lessons.

2.2 Boundary Folding

The boundary fold modification of the conventional kernel estimators in Schuster (1985) is developed to reduce the boundary bias. One way of understanding it is the following, c. f. Silverman (1986). If the sample $X_1, ..., X_n$ have density $f(x) > 0$ for $x \geq 0$ and $f(x) = 0$ otherwise, then $\{X_1, -X_1, X_2, -X_2, ..., X_n, -X_n\}$ is a , not independent, sample of size $2n$ from a population following density g with

$$g(x) = \frac{1}{2}f(|x|), x \in R. \tag{2.7}$$

Now since g has no boundaries in its support, it can be estimated by the ordinary kernel estimators well,

$$\widehat{g}_h(x) = 2^{-1}n^{-1}\sum_{i=1}^{n}\left[K_h(x - X_i) + K_h(x + X_i)\right], x \in R.$$

Then apply $\widehat{g}_h(x)$ back to (2.7), f can be estimated by

$$\widetilde{f}_h(x) = \widehat{g}_h(x) + \widehat{g}_h(-x) = 2\widehat{g}_h(x) = \widehat{f}_h(x) + \widehat{f}_h(-x), x \in [0, \infty).$$

Another viewpoint of this estimator is that since the ordinary kernel estimator distributes some estimated probability mass outside the support of the density, see Figure 1a, it folds this mass back to the original kernel estimate. Or equivalently, add the mirror images of the observations reflected at the boundary to the sample and do usual kernel estimation. We construct this estimator from the same sample used for Figure 2.1a and show it in Figure 2.1b. The dotted arrows point to the reflected data and contribute those dotted kernels to the estimate, only effective in the support $[0, \infty)$ certainly.

If (A1)-(A2) hold and K is a symmetric density function, then for $x = ch, h \to 0, nh \to \infty$, and $n \to \infty$,

$$E\widetilde{f}_h(x) = f(x) + \frac{h^2}{2}f''(x)\mu_2(K) - 2hf'(x)\left[\mu_{1,-c}(K) + c\mu_{0,-c}(K)\right]$$

$$+ 2h^2 f''(x) \left[c^2 \mu_{0,-c}(K) + c\mu_{1,-c}(K)\right] + o(h^2), \tag{2.8}$$

see Marron and Ruppert (1993). At the boundary, in addition to the familiar h^2 order asymptotic bias of the conventional kernel estimator in the interior, c.f. (2.5), this estimator has bias terms of order h unless $f'(x) = 0$. Indeed, this estimator seriously under estimates at the boundary if the density has a negative slope there, and vice versa. Also it is flat at the boundary no matter what the slope of the density is there. In fact this undesired property is inherited from the folding-back nature of this estimator and agrees with the following theorem, which is also mentioned under slightly more restrictive conditions in Silverman (1986).

Theorem 1 *Suppose K is symmetric about zero and differentiable except on a finite set and $X_1, ..., X_n$ is an i.i.d. sample following a continuous distribution. Then for all n, h, the boundary fold estimator $\tilde{f}_h$ has a zero derivative at the endpoint; i.e.*

$$\frac{d}{dx}\tilde{f}_h(x)\bigg|_{x=0} \equiv \lim_{\delta \downarrow 0} \frac{\tilde{f}_h(\delta) - \tilde{f}_h(0)}{\delta} = 0 \ a.e.$$

The proof is deferred to the end of the chapter. A further analysis of this boundary fold estimator also reveals that

$$Var\left(\tilde{f}_h(x)\right) = \frac{f(x)}{nh}\left\{\int K^2 + 2\int_{-\infty}^{c} K(u)K(u-2c)du\right\} + o(n^{-1}h^{-1}). \tag{2.9}$$

Comparing (2.4) and (2.9), we observe that the boundary folded estimator has more variability than the conventional kernel estimator if K is positive. This is explained by the fact that each observation is used twice in this reflection technique. Especially when $x = 0$ this estimator has four times the variance since the folding at $x = 0$ is exactly doubling the estimator. This circumstance commonly happens to more advanced boundary adjusting estimators as well. The reason is that boundary estimation uses fewer data points and the estimators are "magnified" to achieve the same small order of bias as in the interior.

The larger rate of bias from the true density also causes a slower rate of convergence for $\tilde{f}_h(x)$. The asymptotic mean squared error of this estimator is

$$AMSE\left(\tilde{f}_h(x)\right) = 4h^2 \left(f'(x)\right)^2 \left(\mu_{1,-c}(K) + c\mu_{0,-c}(K)\right)^2$$

$$+\frac{f(x)}{nh}\left\{\int_{-\infty}^{\infty} K^2 + 2\int_{-\infty}^{c} K(u)K(u-2c)du\right\}.$$

Denote the minimizer of $AMSE\left(\tilde{f}_h(x)\right)$ by $h_{aopt(bf)}$. Then $AMSE\left(\tilde{f}_{h_{aopt(bf)}}(x)\right)$ converges to zero at the rate of $n^{-2/3}$, but the optimal asymptotic mean squared error of the usual kernel estimator tends to zero at the rate $n^{-4/5}$ in the interior. Thus the boundary folded estimator has slower pointwise rate of convergence at boundary regions. This inefficiency also effects its global performance. We choose to measure global performance of estimators by their asymptotic integrated mean squared errors. Assume K has support $[-1,1]$ and the first two derivatives of f are squared integrable, then since the boundary region is $[0,h]$,

$$AIMSE\left(\tilde{f}_h\right) \equiv \int_0^{\infty} AMSE\left(\tilde{f}_h(x)\right) dx$$

$$= \int_0^h \left(h^2 C_1 f'(x)^2 + \frac{C_2}{nh} f(x)\right) dx + \int_h^{\infty} \left(h^4 C_3 f''(x)^2 + \frac{C_4}{nh} f(x)\right) dx,$$

for some fixed numbers C_i , $i = 1, \ldots, 4$ depending on K. Hence,

$$AIMSE\left(\tilde{f}_h\right) \sim C_1 h^3 \left(f'(0)\right)^2 + C_4 n^{-1} h^{-1},$$

Optimizing h for $AIMSE\left(\tilde{f}_h\right)$ with $h_{aimse(bf)} \sim n^{-1/4}$, we have $AIMSE(\tilde{f}_{h_{aimse(bf)}})$ $\sim n^{-3/4}$. But the ordinary kernel estimator achieves the rate $n^{-4/5}$ in non-boundary case. So this estimator still does not fully conquer the difficulty of the boundary situation.

So far we have seen that neither the ordinary kernel estimator nor the boundary folded estimator is satisfactory. And one might ask whether it is possible to do any better than that or even as well as at the interior points. We will see two methods that do achieve the familiar $n^{-4/5}$ rate of convergence in the following sections and provide some detailed investigation of them.

2.3 Rice's Boundary Modification

In the setting of kernel estimation of regression functions, Rice (1984) suggests a boundary modification which can be directly translated to the framework of density estimation. Rice's boundary modification uses a kernel function which is a linear combination of two kernels with bandwidths h and αh at boundary points such that the bias is of the same order of magnitude as at interior points. This is similar to the idea of bias reduction in Schucany and Sommers (1977). Keeping consistent with our discussed setting , we adapt this to kernel density estimation. The analogous Rice boundary estimator of density f at $x = ch$, $c \geq 0$ is

$$\bar{f}_{h,\alpha}(x) = a\hat{f}_h(x) - b\hat{f}_{\alpha h}(x) = n^{-1}\sum_{i=1}^{n}(aK_h - bK_{\alpha h})(x - X_i), \tag{2.10}$$

where

$$a = \frac{\alpha\mu_{1,c/\alpha}}{\alpha\mu_{0,c}\mu_{1,c/\alpha} - \mu_{0,c/\alpha}\mu_{1,c}}, b = \frac{\mu_{1,c}}{\alpha\mu_{1,c/\alpha}}a, \text{ and } \alpha > 0. \tag{2.11}$$

This estimator is the same as the usual kernel estimator with kernel function $aK_h - bK_{\alpha h}$. Hence the bias of this estimator is the same as in equation (2.3) with K replaced by $aK_h - bK_{\alpha h}$. But now, with the careful choice of a and b as shown in (2.11), the coefficients of $f(x)$ and $f'(x)$ in (2.3) are zero. Therefore the asymptotic bias is of the order h^2. Recall all the troubles with the conventional kernel estimator comes from its large bias at the boundary area. Hence the key to the success of Rice's estimator is this bias reduction.

However, before implementing this modification one has to decide on the choice of α, the ratio between the bandwidths for the two kernel estimators in the linear combination. In the following sections, tools for studying this issue are provided and some answers for certain special cases are given.

2.3.1 Asymptotics for Studying Bandwidth Ratio

Recall that the Rice estimator is a linear combination of two usual kernel density estimators with bandwidths h and αh. To choose α, Rice (1984) recognizes the difficulty in finding possible solutions for each c. But it is suggested there to take $\alpha = 2 - c$ so that the estimator is a weighted average of observations fall in an interval of length $2h$, the same as in the interior. Notice that with $\alpha = 2-c$ different combinations of kernel functions are used at different points, namely K_h and $K_{(2-c)h}$ for $x = ch$. Hence a fair amount of computation is involved when implementing the estimate with this selection of α. On the other hand, if α is fixed for every x then one can apply the fast binned calculation of kernel estimators discussed in Section 2.1.1. In the sequel we compare this to $\alpha = 2 - c$ and investigate what would be the best fixed bandwidth ratio.

First, some simple algebra shows that $\hat{f}_{h,\alpha}(x) = \hat{f}_{\alpha h, \frac{1}{\alpha}}(x)$. Therefore it is enough to consider only the set of bandwidth ratios $\{\alpha, \alpha \geq 1\}$. Write

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

where a and b are the coefficients given in (2.11). Then Rice's boundary kernel at $x = ch$ is

$$\bar{K}_{h,\alpha}(\cdot) = \frac{a}{h} K(\frac{\cdot}{h}) - \frac{b}{\alpha h} K(\frac{\cdot}{\alpha h}) = \frac{1}{h} \overline{K}_\alpha(\frac{\cdot}{h}). \tag{2.12}$$

The bias and variance of this estimator are analyzed in the following theorem as information for choice of α. The proof is left to the end of chapter.

Theorem 2 *Under assumptions (A1), (A2), $h \to 0, n \to \infty$, and $nh \to \infty$,*

$$Bias^2\left(\overline{f}_{h,\alpha}(x)\right) = \frac{h^4}{4} (f''(x))^2 \left(\int_{-\infty}^{c} u^2 \overline{K}_\alpha(u) du\right)^2 + o(h^4). \tag{2.13}$$

and

$$Var\left(\overline{f}_{h,\alpha}(x)\right) = n^{-1} h^{-1} f(x) \int_{-\infty}^{c} \overline{K}_\alpha^2(u) du + o(n^{-1} h^{-1}). \tag{2.14}$$

From the asymptotic expressions (2.13) and (2.14) we see that the asymptotic bias and variance depend on α through parts that are independent of the true underlying density f. This implies that study of asymptotics in the direction of choice of α will rely only on those "kernel parts". So we call $\int_{-\infty}^{c} u^2 \overline{K}_\alpha(u) du$ and $\int_{-\infty}^{c} \overline{K}_\alpha^2(u) du$ "kernel part of asymptotic bias" and "kernel part of asymptotic variance", respectively.

Next, some further investigations are intended to answer the question of which fixed α is the best for this modification. Obviously, each α decides a unique shape of the combined kernel function $\overline{K}_\alpha$, up to a rescaling factor controlled by h. Thus deciding the best α among $\{\alpha \geq 1\}$ is equivalent to choosing the best kernel function from the family $\{\overline{K}_\alpha(\cdot), \alpha \geq 1\}$. Applying the idea of canonical kernels of Marron and Nolan (1989), if $\delta_0 = \left(\int_{-\infty}^c \overline{K}_\alpha^2\right)^{\frac{1}{5}} \left(\int_{-\infty}^c u^2 \overline{K}_\alpha\right)^{-\frac{2}{5}}$ then

$$AMSE\left(\overline{f}_{\delta_0 h,\alpha}(x)\right) = \left(\int_{-\infty}^c \overline{K}_\alpha^2\right)^{\frac{4}{5}} \left(\int_{-\infty}^c u^2 \overline{K}_\alpha\right)^{\frac{2}{5}} \left(\frac{f(0+)}{nh} + \frac{h^4}{4}\left(f''(0+)\right)^2\right).$$

Denote $S_c^{\frac{2}{5}}(\alpha) \equiv \left(\int_{-\infty}^c \overline{K}_\alpha^2\right)^{\frac{4}{5}} \left(\int_{-\infty}^c u^2 \overline{K}_\alpha\right)^{\frac{2}{5}}$. Notice that $AMSE\left(\overline{f}_{\delta_0 h,\alpha}(x)\right)$ is factorized into two parts: $S_c^{\frac{2}{5}}(\alpha)$ depends solely on $\overline{K}_\alpha$ and the remaining part depends only on h and f. Thus the problem of deciding the kernel $\overline{K}_\alpha$ is independent of the underlying density and the choice of h. Applying the bandwidth $\delta_0 h$ enables a fair comparison among kernels. We shall call δ_0, which depends on $\overline{K}_\alpha$, the canonical bandwidth for the kernel $\overline{K}_\alpha$.

Clearly the asymptotic bias and variance of the Rice's estimator depend on the kernel function as well as α. We consider the Gaussian kernel $K \equiv \varphi$ and investigate the best bandwidth ratio from this family. Now, with φ in the place of K for equations (2.13) and (2.14), the squared bias is

$$Bias^2\left(\overline{f}_{h,\alpha}(x)\right) = \frac{h^4}{4}\left(f''(x)\right)^2 \left(\int_{-\infty}^c u^2 \left[a\varphi(u) - \frac{b}{\alpha}\varphi(\frac{u}{\alpha})\right] du\right)^2 + o(h^4).$$

By some algebra and the fact that $x \to 0$,

$$Bias^2\left(\overline{f}_{h,\alpha}(x)\right) = \frac{h^4}{4}\left(f''(x)\right)^2 \left(a\int_{-\infty}^c u^2\varphi(u)du - b\alpha^2 \int_{-\infty}^{\frac{c}{\alpha}} u^2\varphi(u)du\right)^2 + o(h^4)$$

$$= \frac{h^4}{4}\left(f''(0+)\right)^2 \left(a\left\{-c\varphi(c) + \Phi(c)\right\} - b\alpha^2\left\{-\frac{c}{\alpha}\varphi(\frac{c}{\alpha}) + \Phi(\frac{c}{\alpha})\right\}\right)^2 + o(h^4).$$

Also some direct calculation shows that

$$Var\left(\overline{f}_{h,\alpha}(x)\right) = n^{-1}h^{-1}f(x)\int_{-\infty}^c \left(a\varphi(u) - \frac{b}{\alpha}\varphi\left(\frac{u}{\alpha}\right)\right)^2 du + o\left(n^{-1}h^{-1}\right)$$

$$= \frac{f(x)}{nh}\left\{a^2\int_{-\infty}^c \varphi^2(u)du - 2ab\int_{-\infty}^c \varphi(u)\varphi\left(\frac{u}{\alpha}\right)\frac{du}{\alpha} + \frac{b^2}{\alpha}\int_{-\infty}^c \varphi^2\left(\frac{u}{\alpha}\right)\frac{du}{\alpha}\right\} + o\left(n^{-1}h^{-1}\right)$$

$$= \frac{f(x)}{nh}\left\{a^2\int_{-\infty}^c \frac{1}{2\sqrt{\pi}}\varphi_{\frac{1}{\sqrt{2}}}(u)du - 2ab\int_{-\infty}^c \frac{1}{\sqrt{2\pi(1+\alpha^2)}}\varphi_{\frac{\alpha}{\sqrt{1+\alpha^2}}}(u)du\right.$$

$$+b^2 \int_{-\infty}^{c} \frac{\alpha}{2\sqrt{\pi}} \varphi_{\frac{\alpha}{\sqrt{2}}}(u)du \Big\} + o\left(n^{-1}h^{-1}\right).$$

Since $x = ch \to 0$ and f is continuous around 0 we can write

$$Var\left(\overline{f}_{h,\alpha}(x)\right) = \frac{f(0+)}{nh} \frac{1}{\sqrt{2\pi}} \left\{ \frac{a^2}{\sqrt{2}} \Phi\left(\sqrt{2}c\right) - \frac{2ab}{\sqrt{1+\alpha^2}} \Phi\left(\frac{\sqrt{1+\alpha^2}c}{\alpha}\right) \right.$$
$$\left. + \frac{b^2}{\sqrt{2}\alpha} \Phi\left(\frac{\sqrt{2}c}{\alpha}\right) \right\} + o\left(n^{-1}h^{-1}\right),$$

where Φ is the c.d.f. of the standard normal density.

Plots of the kernel parts of asymptotic bias and variance with Gaussian kernel versus c for various choices of α show that the estimate using $\alpha \equiv 2$ is performing very much the same as that with $\alpha = 2 - c/4$ in terms of asymptotic bias and variance, see Figures 2.2a and 2.2b. This suggests favor of fixed α for simplicity and computational speed considerations.

The best α for a fixed c is obtained via minimizing $S_c(\alpha)$ over all $\alpha \geq 1$. Ideally if there is an α_{opt} that uniformly minimizes $S_c(\alpha)$ for every $c \geq 0$, then the best kernel would be $\overline{\varphi}_{\alpha_{opt}}$. But since $S_c(\alpha)$ is a function of α indexed by c, this is not necessarily the case. Numerical work shows α_{opt} typically depends on c. None the less, since it is the estimation at the endpoint $x = 0$ that suffers from the largest amount of mean squared error; i.e. where the density estimation is most difficult, we will study $S_c(\alpha)$ at this point first. When $x = 0$ and K is Gaussian,

$$Bias^2\left(\overline{f}_{h,\alpha}(0)\right) = h^4 \left(\frac{f''(0)}{2}\right)^2 (-\alpha)^2 + o(h^4), \tag{2.15}$$

and

$$Var\left(\overline{f}_{h,\alpha}(0)\right) = \frac{f(0)}{nh} \frac{1}{\sqrt{\pi}(\alpha-1)^2} \left(\frac{\alpha^3+1}{\alpha} - \frac{2\sqrt{2}\alpha}{\sqrt{1+\alpha^2}}\right) + o(n^{-1}h^{-1}). \tag{2.16}$$

If α is very large, the kernel function used for $c = 0$ almost looks like $2\varphi(u), u \geq 0$, except for having a negative right-hand-side tail. This negative tail leads to a large negative bias. But since the variability due to this small negative tail is negligible compared to that from the positive part of the kernel, the variance tends to $\frac{1}{\sqrt{\pi}}$ which is equal to $\int_0^\infty (2\varphi(u))^2 du$. Now, from (2.15) and (2.16),

$$S_\alpha(0) = \frac{\alpha}{\pi(\alpha-1)^4} \left(\frac{\alpha^3+1}{\alpha} - \frac{2\sqrt{2}\alpha}{\sqrt{1+\alpha^2}}\right)^2.$$

The solid line in Figure 2.2c shows that $S_\alpha(0)$ is increasing for $\alpha \geq 1$; i.e. $S_\alpha(0)$ is minimized at $\alpha = 1$. Thus for $c = 0$ the best kernel function among the Gaussian family considered here is $\lim_{\alpha\to 1} \overline{\varphi}_\alpha(u)$ which equals $2(2-u^2)\varphi(u)$. Numerical study also suggests that $S_\alpha(c)$ is always minimized at $\alpha = 1$ for any fixed $c \in [0, 0.84]$, Figures 2.2c and 2.2d provide visual evidence of this for some chosen values of c.

As for those c belonging to $(0.84, \infty)$, $S_\alpha(c)$ is minimized and at the same time equal to zero at some $\alpha_{opt}(c) > 1$, see Figures 2.2d and 2.2e. But $S_\alpha(c)$ is zero when $\int_{-\infty}^{c} u^2\overline{\varphi}_\alpha = 0$ which means that the asymptotic bias is of some other order higher than h^4. Then, analysis of $S_\alpha(c)$ and even the $AMSE$ does not reflect behavior of MSE in this case. Hence $\alpha_{opt}(c)$ is not the right choice of bandwidth ratio and instead higher order bias terms need to be studied. But that situation becomes more difficult and the density function and its higher derivatives play important roles. Also, for any α, $\overline{\varphi}_\alpha$ is essentially the same as φ when c is large. Thus it becomes less meaningful to optimize asymptotic mean squared error among $\overline{\varphi}_\alpha$ for large c. These concerns and the fact that $\alpha = 1$ is the best for $c \in [0, 0.84]$, a range that requires more careful choice of α than $(0.84, \infty)$ does, recommend the kernel $\overline{\varphi}_{\alpha=1}$. Note that this is a degenerate 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.

Notice that $S_{\alpha=1}(c) \leq \left(\int \varphi^2\right)^2 \left(\int u^2\varphi\right) = \frac{1}{4\pi}$ for $c \geq 0.85$, hence $\lim_{\alpha\to 1} \overline{\varphi}_\alpha$ is as good as the usual normal kernel at the points where we didn't try to optimize α. Also, $\lim_{\alpha\to 1} \overline{\varphi}_\alpha$ is approximately the same as φ, the kernel used for interior point, for large c. Another nice feature of this kernel function is that it allows convenient implementation using the binning method discussed in Section 2.1.1. Namely the estimate is simply a linear combination of two binned estimates, whose kernels are $\varphi(u)$ and $u^2\varphi(u)$, with coefficients $\frac{2+c^2}{(c^2+1)\Phi(c)+c\varphi(c)}$ and $\frac{1}{(c^2+1)\Phi(c)+c\varphi(c)}$, respectively.

Figure 2.2f shows the discrepancy between the optimized $S_c^{\frac{2}{5}}$ and the best kernel part of the asymptotic mean squared error. From this we can see that minimizing $S_c(\alpha)$ for $c \geq 0.84$ is irrelevant to the choice of α and instead higher order bias terms of the asymptotic mean squared error need to be considered. This has not been carefully investigated because the difference among $\overline{\varphi}_\alpha$'s becomes much smaller for larger c and because there is another approach by calculating the exact mean squared errors, which is discussed in the next section. Again there should be awareness that especially when $S_{\alpha=1} < \frac{1}{4\pi}$ we are not actually having an optimistically small mean squared error but that is the effect of not accounting for terms of bias other than the h^4 term.

2.3.2 Exact Mean Squared Error

We saw in the previous section that, when analyzing behaviors of the mean squared error of Rice's modification as a function of α, the asymptotic analysis sometimes becomes totally noninformative. An alternative approach is to study the mean squared error directly, c.f. Marron and Wand (1992). Then, there is no longer a problem of not enough terms in the asymptotics to provide information about the mean squared error.

Here, small to moderate sample size properties are emphasized, but the results interact with knowledge learnt from asymptotic study. Mean squared error of the estimator also depends on the unknown density f, but we can choose some representative ones. The normal mixture family contains a rich variety of shapes. Since now we are investigating boundary effects, we shall consider the family of truncated normal mixture densities. Also, we choose to work with the Gaussian kernel function. One advantage of this is, along with the normal mixture densities, the convolutions in the bias and variance become very easy to compute.

For compactness of this discussion, we will only treat four density functions from this family. Figure 2.3 shows how they look. They will also serve as illustrative examples in later studies on boundary effects. Density #1 decides whether an estimator can detect concavity of a density function. Density #2 will be able to distinguish between estimators that can model curvatures near the boundaries and those that can not. Density #3 is considered since its moderately shaped mode near zero brings in the boundary effect issue. If a density is flat and low near the endpoints the usual kernel estimator will do well without any boundary modification. Density #4 has a very sharp spike extremely close to the endpoint zero, it represents challenging densities that have sharp features right at the boundaries. The formulas for these densities are listed in the following table.

	Truncated Normal Mixture Density
#1	$\left\{\frac{1}{2}\Phi\left(-\frac{1}{5}\right)^{-1}N\left(-\frac{1}{5},1\right)+\frac{1}{2}\Phi\left(-\frac{7\sqrt{2}}{10}\right)^{-1}N\left(-\frac{7}{10},\frac{1}{2}\right)\right\}I_{(x\geq 0)}$
#2	$\left\{\frac{3}{5}\Phi\left(\sqrt{\frac{3}{10}}\right)^{-1}N\left(\frac{3}{10},\frac{3}{10}\right)+\frac{2}{5}\Phi\left(\frac{5}{2}\right)^{-1}N\left(\frac{5}{2},1\right)\right\}I_{(x\geq 0)}$
#3	$\left\{\frac{4}{5}\Phi(0)^{-1}N\left(0,\frac{4}{9}\right)+\frac{1}{5}\Phi(9)^{-1}N\left(3,\frac{1}{9}\right)\right\}I_{(x\geq 0)}$
#4	$\left\{\sum_{l=0}^{7}\frac{1}{8}\Phi\left(\frac{\mu_l}{\sigma_l}\right)^{-1}N\left(\mu_l,\sigma_l^2\right)\right\}I_{(x\geq 0)}, \mu_l=3\left\{\left(\frac{2}{3}\right)^l-1\right\}, \sigma_l=\left(\frac{2}{3}\right)^l.$

Remember that bias and variance of Rice's estimator are

$$Bias\left(\overline{f}_{\alpha,h}(x)\right)=\int_0^\infty \bar{K}_{h,\alpha}(x-s)f(s)ds-f(x),$$

and

$$Var\left(\overline{f}_{\alpha,h}(x)\right)=n^{-1}\int_0^\infty \bar{K}_{h,\alpha}^2(x-s)f(s)ds-n^{-1}\left(\int_0^\infty \bar{K}_{h,\alpha}(x-s)f(s)ds\right)^2.$$

Now $\bar{K}_{h,\alpha}=a\varphi_h-b\varphi_{\alpha h}$ and $f(x)=\sum_{l=1}^{m}w_l\Phi\left(\frac{\mu_l}{\sigma_l}\right)^{-1}\varphi\left(x-\mu_l\right)I_{(x\geq 0)}$. If we define

$$A_t^l(x)=\frac{1}{\sqrt{t^2+\sigma_l^2}}\varphi\left(\frac{x-\mu_l}{\sqrt{t^2+\sigma_l^2}}\right)\Phi\left(\frac{t\sigma_l^{-1}\mu_l+\sigma_l t^{-1}x}{\sqrt{t^2+\sigma_l^2}}\right),$$

then it is simple to show that

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

and

$$\int_0^\infty \bar{K}_{h,\alpha}^2(x-s)f(s)ds=a^2\int_0^\infty\varphi_h^2(x-s)f(s)ds-2ab\int_0^\infty\varphi_h(x-s)\varphi_{\alpha h}(x-s)f(s)ds$$
$$+b^2\int_0^\infty\varphi_{\alpha h}^2(x-s)f(s)ds$$

$$=\sum_{l=1}^{m} \frac{w_l}{\sqrt{2\pi}h}\Phi\left(\frac{\mu_l}{\sigma_l}\right)^{-1}\left\{\frac{a^2}{\sqrt{2}}A^l_{\frac{h}{\sqrt{2}}}(x)-\frac{2ab}{\sqrt{1+\alpha^2}}A^l_{\sqrt{\frac{\alpha^2h^2}{1+\alpha^2}}}(x)+\frac{b^2}{\sqrt{2}\alpha}A^l_{\frac{\alpha h}{\sqrt{2}}}(x)\right\}.$$

Utilizing these expressions we obtain very compact formula for the mean squared errors of Rice's modified estimators. We consider α ranges from one to five. The bandwidths we use in this calculation are $C_f n^{-1/5}\delta_0$. Here C_f is some constant, depending on the density, chosen to be visually appropriate for the simulated data sets. In the order of the above list of densities, $C_f = 0.7, 1.2, 0.5$, and 1, respectively. Notice that these bandwidths are multiples of the canonical bandwidths, hence comparison among kernels does not confound with what the density is and the choice of bandwidth.

This exact calculation study is conducted only for estimating at $x = 0$ since it is most important there. The exact mean squared errors for estimating $f(0)$ are displayed in Figure 2.4. There are several interesting observations from these pictures. First, there is a clear increasing trend of the mean squared error in α for both density #1 and density #3. Hence in these two cases the asymptotic result that the best α is $\alpha = 1$ holds even for small sample sizes. On the other hand, the mean squared error is improved with larger α when the density is either density #2 or density #4. Figure 2.5 provides some answers to the curiosity of what is behind these different performances.

The key machinery is the following: for $1 \le \alpha_1 < \alpha_2$, $\overline{\varphi}_{\alpha_1}$ has a deeper negative right tail than that of $\overline{\varphi}_{\alpha_2}$. Now we examine how this affects their performances when applied to our example densities by taking $\alpha_1 = 1$ and $\alpha_2 = 2$. First, density #1 is highest at $x = 0$ and then decreases. The kernel estimators are always too low at $x = 0$ since they are weighted moving averages of the density. But, estimating density #1 at zero is better with $\overline{\varphi}_{\alpha=1}$ since the density is high at the region where $\overline{\varphi}_{\alpha=1}$ is greater than $\overline{\varphi}_{\alpha=2}$, meaning that a lot of observations in the range contribute more to the estimate. Also the density is relatively low at places where $\overline{\varphi}_{\alpha=1}$ is more negative than $\overline{\varphi}_{\alpha=2}$; which implies that only rarely are there observations that lower the estimate. The same situation occurs when estimating density #3 at $x = 0$. Now, as for estimating density #2 at $x = 0$, we see that the negative tail of $\overline{\varphi}_{\alpha=1}$

happens where the density is still high and makes the estimator worse. So larger α is preferred in this case. Finally, we see that the negative tail and the higher positive part of $\overline{\varphi}_{\alpha=1}$ are both irrelevant in estimating density #4 at $x = 0$ since the density is negligibly small in those regions. On the other hand, $\overline{\varphi}_{\alpha=2}$ is higher at a small region next to $x = 0$ and the density is extremely large there. This makes $\overline{\varphi}_{\alpha=2}$ superior to $\overline{\varphi}_{\alpha=1}$ for in this case.

Besides what we learned about these small sample properties through this exact calculation, we would like to remark that asymptotics favors $\overline{\varphi}_{\alpha=1}$ because its "improved" positive part and the shrinking domain of kernels makes this useful. In conclusion, "best α" depends on the particular density and sample size. However, $\alpha = 1$ is recommended as a useful general purpose choice.

2.4 Local Linear Fit in Density Estimation

Stone (1977) and Cleveland (1979) discuss locally least squares regression. Fan (1992) refines this method by introducing smoothed kernel weights to the least squares and credited its success in efficiency and adaptive properties. The motivation of the local polynomial regression technique is that if the regression function is smooth then it can be approximated by a polynomial very well locally. Then it is reasonable to estimate the regression curve by fitting some polynomial through least squares. Fan (1993) shows particularly that local linear fit attains some minimax efficiency within the class of linear estimators in the interior. And such a estimator has a very appealing feature of adapting to estimation at the boundaries; i.e. without any extra effort of modification such as boundary folding or Rice's linear combination, it has the same rate of convergence as in the interior, see Fan (1992). The estimator itself and these nice properties carry over to the setting of density estimation nicely. The purpose of this section is to implement the local linear fit idea to the density estimation context and derive some asymptotic properties.

In density estimation, the density function under investigation is parallel to the response curve of regression analysis. Hence, the idea is to fit some sort of "response" which is equal to the density plus some random error. An apparent way of producing such responses is through data binning, as discussed in Section 2.1.1, of the data. The bin centers can be thought of as the design points in regression. The bin counts reflect the heights of the density at the bin centers and hence can be viewed as the responses there. With these statistics available, the local linear regression technique is applicable to density estimation by linear fitting to the bin counts. Also, from the computational point of view, the estimator obtained this way can be fast implemented since it needs only to handle the bin counts instead of the original data.

The notations follow those in Section 2.1.1 where binning is discussed. Since the density has support $[0, \infty)$, the bin centers are taken as $x_i = (i - \frac{1}{2})b, i = 1, ..., g$ for some fixed positive b which is referred to as the bin width. The bin counts provide information of heights of the density at the bin centers in the following sense. Since $X_1, ..., X_n$ are i.i.d., the Strong Law of Large Numbers implies that

$$n^{-1}b^{-1}c_i \overset{a.s.}{\to} b^{-1}\int_{x_i-b/2}^{x_i+b/2} f(u)du \approx f(x_i),$$

for each $i = 1, ..., g$. Furthermore, if the second derivative of f at x exists, the Taylor function approximation is

$$f(z) \approx f(x) + (z - x)f'(x),$$

for any z belonging to a neighborhood of x. Hence we would fit f at x by a local linear least squares solution with smooth weights decided by K_h; i.e.

$$\min_{b_0,b_1} \sum_{i=1}^{g} K(\frac{x_i - x}{h}) \left(n^{-1}b^{-1}c_i - b_0 - b_1(x_i - x)\right)^2. \qquad (2.17)$$

Suppose $\widehat{b_0}(x)$ and $\widehat{b_1}(x)$ together are the minimizing solution to the least squares problem (2.17), then $\widehat{b_0}(x)$ would be a good estimate of $f(x)$. Hence we define

$$\widehat{f_L}(x) \equiv \widehat{b_0}(x) = \frac{S_{n,2}(x)T_{n,0}(x) - S_{n,1}(x)T_{n,1}(x)}{S_{n,2}(x)S_{n,0}(x) - S_{n,1}(x)S_{n,1}(x)},$$

where

$$S_{n,j}(x) = \sum_{i=1}^{g} K(\frac{x_i - x}{h})(x_i - x)^j, j = 0, 1, 2,$$

and

$$T_{n,j}(x) = n^{-1}b^{-1}\sum_{i=1}^{g} K(\frac{x_i - x}{h})(x - x_i)^j c_i, j = 0, 1.$$

Here the bandwidth h controls the amount of smoothing: the fitted curve becomes a straight line if h is large enough and has a lot of spikes if h is too small. This estimator can be rewritten as

$$\hat{f}_L(x) = n^{-1}\sum_{i=1}^{g} \frac{S_{n,2}(x) - S_{n,1}(x)(x_i - x)}{b\left[S_{n,2}(x)S_{n,0}(x) - S_{n,1}(x)S_{n,1}(x)\right]} K(\frac{x_i - x}{h})c_i. \tag{2.18}$$

Note that in (2.18) only c_i's are random and $\hat{f}_L(x)$ is a linear function of these pre binned data. Comparing (2.18) more closely to the binned version of the kernel estimators, see (2.1) , it is seen that $\hat{f}_L(x)$ is also a binned kernel estimator and

$$K_{n,x}(u) \equiv \frac{hS_{n,2}(x) - h^2 S_{n,1}(x)u}{b\left[S_{n,2}(x)S_{n,0}(x) - S_{n,1}(x)S_{n,1}(x)\right]} K(u)$$

is the kernel for estimation at x. The effective kernel $K_{n,x}$ depends on the location of x. If assumption (A4) stated below holds and x is in the interior then $K_{n,x}$ is approximately the same as K. The following condition is convenient for investigating asymptotic properties of $\hat{f}_L(x)$.

(A4) For $l = 0, 1, 2$, $K^{(l)}$ is bounded and absolutely integrable with finite second moments.

Next we state the asymptotic properties.

Theorem 3 *Suppose conditions (A1)-(A4) hold and $h \to 0, nh \to \infty, b = o(h)$. Then, if x is away from the boundary region, $\hat{f}_L(x)$ has bias*

$$Bias\left(\hat{f}_L(x)\right) = \frac{h^2}{2} f''(x) \int_{-\infty}^{\infty} u^2 K(u)du + o(h^2), \tag{2.19}$$

and variance

$$Var\left(\hat{f}_L(x)\right) = n^{-1}h^{-1}f(x)\int_{-\infty}^{\infty} K^2(u)du + o(n^{-1}h^{-1}). \tag{2.20}$$

Otherwise, if x is a boundary point $x = ch, c \geq 0$,

$$Bias\left(\hat{f}_L(x)\right) = \frac{h^2}{2} f''(0+)\int_{-c}^{\infty} u^2 K_c^*(u)du + o(h^2), \tag{2.21}$$

$$Var\left(\widehat{f}_L(x)\right) = n^{-1}h^{-1}f(0+)\int_{-c}^{\infty} K_c^*(u)^2 du + o(n^{-1}h^{-1}), \tag{2.22}$$

where

$$K_c^*(u) = \frac{S_{2,c} - S_{1,c}u}{S_{2,c}S_{0,c} - S_{1,c}^2} K(u),$$

with $S_{j,c} = \int_{-c}^{\infty} u^j K(u)du, j = 0, 1, 2.$

Knowing that $\widehat{f}_L(x)$ is essentially the same as the conventional kernel estimator in the interior, it is not surprising that they have the same asymptotic bias and variance. But, unlike the ordinary kernel estimator having difficulty at boundary regions, the local linear fit technique itself adjusts to have the same rate of convergence at the boundary as in the interior. The reason for this adaptive property is that the effective kernels $K_{n,x}$ for boundary points are automatically modified such that condition (A3) hold.

The second derivative of f is assumed to exist in analyzing the mean squared error for $\widehat{f}_L(x), x = ch$. If this condition does not hold, then the asymptotic results change according to the behavior of f near zero. First, if the density has a pole at zero, say $f(x) \sim px^{\alpha}$ as $x \to 0$ for some $p > 0$ and known $\alpha < 0$, then the estimator has an infinite bias of the order h^{α}. Secondly, if the density exists at zero but its derivative tends to infinity there, which is roughly $f(x) \sim px^{\alpha} + q$ as $x \to 0$ for some $p > 0, q \geq 0$ and known α, $0 < \alpha < 1$, then the bias is of order h^{α}, the variance is of order $n^{-1}h^{-1}$ and the optimal rate for the mean squared error is $n^{-\frac{2\alpha}{1+2\alpha}}$ which is slower than the usual $n^{-\frac{4}{5}}$ rate. Finally, if f has first derivative at zero but its second derivative has a pole there, say $f(x) \sim px^{\alpha} + q$ as $x \to 0$ for some $p > 0, q \geq 0$ and known α, $1 < \alpha < 2$, then the results are analogous to the previous case.

Since both the Rice's modified estimators and the local linear estimators preserve the rate of convergence everywhere, an interesting question is how they compare to each other. We will look make this comparison by studying the best constant for the estimation problem in the next chapter. In fact, our result there answers much more than how these two estimators compare by considering a class of general estimators. For the next section, we show that our asymptotic results are very informative by a simulation study.

2.5 A Simulation Study

In this section we will compare the boundary behaviors of the four density estimators in the previous sections by simulation. The truncated normal mixture densities of Section 2.3.2 will be the subjects. Ten independent random samples each of size 100 are drawn from each of the densities and then the four estimators are constructed from those samples. Simulating ten realizations of the estimators will help obtaining some insight into their variability. The bandwidths are the best ones chosen by visual judgment.

Figures 2.6 show the four types of estimators from the samples coming from density #1. First, the conventional kernel estimators are unacceptably biased from the density in a very wide region, about from zero to point six. This agrees with the asymptotic bias given in (2.3). Especially the estimates at zero are only about half the height of $f(0)$, which is quantified in (2.3) by $E\hat{f}_h(0) \sim \mu_{0,0}f(0) = 1/2f(0)$. Also, note that approximately 7% of the estimated probability mass is spread to the negative half line. This is an inevitable property of this estimator when boundaries are present and there are observations near the boundaries. Such a property is disagreeable since positive estimated probability is assigned to where it should not be. Moreover, this estimator is not able to model the concavity and the sharp spike of the density function; it misinterprets the true curve as having only one single slightly skewed peak. All the above observations strengthen and clarify the need for boundary modifications.

Next we examine the boundary fold estimators. Of course now the probability mass below $x = 0$ is removed and used to improve the height of the estimate near zero. But this is still somewhat low. This is explained by the order h asymptotic bias seen in equation (2.8). Further, notice that the estimates all suggest a flat slope at zero and were not able to indicate the concavity there. This agrees with the result of Theorem 1.

The benefit of the Rice's bias reduction, see (2.13), is visually clear. The estimators are closer to the density than the boundary fold estimators. Even more importantly, they present the concavity of the density nicely; this never happened for the previous two estimators. Figure 2.6d shows that the local linear estimators are also very satisfactory. Most of the estimates are concave, as is the true density, and the heights at the boundary region are improved.

Studying density #2, the ordinary kernel estimates in Figure 2.7 give the same lessons as those we learned from the simulation on the density #1. They smooth out features of the density near the boundary and tend to give over simplified structures there. Now we are convinced that the uncorrected kernel estimator is not performing well in both theoretical and practical senses near the boundary. Rice's modification and the local linear fit estimates are again both better than the boundary fold estimates. Most of the time they catch the first peak. The Rice estimations seem to be closer to the density there while being a little too wiggly at the second peak.

As for estimating density #3, see Figures 2.8, the lessons for the ordinary kernel estimator and the boundary fold modification are mainly the same as what was learned before. The Rice estimator is clearly better in the sense of variance. But the local linear fit estimators are on the average as close to the curve as the Rice's estimations. It is interesting that these two estimators both work well and are very close to each other.

Now it is very convincing, from both the asymptotic and simulation studies, that the Rice's modification and the local linear fit technique are competitive. They both preserve the rate of convergence everywhere. It will be interesting to know which one is optimizing the mean squared error. The next chapter is devoted to give some answer to this question.

2.6 Proofs

I. Proof of Theorem 1:

For any fixed n and h,

$$\frac{d}{dx}\tilde{f}_h(x)\bigg|_{x=0} = \lim_{\delta\downarrow 0}\frac{\tilde{f}_h(\delta)-\tilde{f}_h(0)}{\delta}$$

$$= n^{-1}h^{-1}\lim_{\delta\downarrow 0}\frac{1}{\delta}\sum_{i=1}^{n}\left\{K\left(\frac{\delta-X_i}{h}\right)+K\left(\frac{\delta+X_i}{h}\right)-K\left(\frac{0-X_i}{h}\right)-K\left(\frac{0+X_i}{h}\right)\right\}$$

$$= n^{-1}h^{-2}\sum_{i=1}^{n}\left\{\lim_{\delta\downarrow 0}\frac{h}{\delta}\left[K\left(\frac{X_i}{h}+\frac{\delta}{h}\right)-K\left(\frac{X_i}{h}\right)\right]-\lim_{\delta\downarrow 0}\frac{h}{\delta}\left[K\left(\frac{X_i}{h}\right)-K\left(\frac{X_i}{h}-\frac{\delta}{h}\right)\right]\right\},$$

since K is symmetric about zero. Now, K is differentiable except on a finite set and $X_i's$ are continuous random variables imply

$$\frac{d}{dx}\tilde{f}_h(x)\bigg|_{x=0} \overset{a.s.}{=} n^{-1}h^{-2}\sum_{i=1}^{n}\left\{K'\left(\frac{X_i}{h}\right)-K'\left(\frac{X_i}{h}\right)\right\}=0.$$

II. Proof of Theorem 2:

To analyze the bias of the Rice's estimator, from (2.10) and (2.12),

$$Bias^2\left(\overline{f}_{h,\alpha}(x)\right)=\left(\int_{-\infty}^{+\infty}\bar{K}_{h,\alpha}\left(x-s\right)f(s)ds-f(x)\right)^2$$

$$=\left(\int_{-\infty}^{c}\overline{K}_\alpha(u)f(x-uh)du-f(x)\right)^2.$$

By a standard Taylor expansion of f around x and condition (A1),

$$Bias^2\left(\overline{f}_{h,\alpha}(x)\right)=\left\{\int_{-\infty}^{c}\overline{K}_\alpha(u)\left(f(x)-uhf'(x)+\frac{u^2h^2}{2}f''(x)+o(h^2)\right)du-f(x)\right\}^2$$

$$=\left\{\left(\int_{-\infty}^{c}\overline{K}_\alpha-1\right)f(x)-hf'(x)\left(\int_{-\infty}^{c}u\overline{K}_\alpha\right)+\frac{h^2}{2}f''(x)\left(\int_{-\infty}^{c}u^2\overline{K}_\alpha\right)+o(h^2)\right\}^2.$$

Note that a and b as given in (2.11) were chosen such that $\int_{-\infty}^{c}\overline{K}_\alpha=1$ and $\int_{-\infty}^{c}u\overline{K}_\alpha$. Hence the above expression is reduced to

$$Bias^2\left(\overline{f}_{h,\alpha}(x)\right)=\frac{h^4}{4}\left(f''(x)\right)^2\left(\int_{-\infty}^{c}u^2\overline{K}_\alpha(u)du\right)^2+o(h^4).$$

As for the variance,

$$Var\left(\overline{f}_{h,\alpha}(x)\right)=n^{-1}Var\left(\bar{K}_{h,\alpha}\left(x-X_1\right)\right)$$

$$= n^{-1} \left(\int_{-\infty}^{+\infty} \bar{K}_{h,\alpha}^2 \, (x-s) f(s) ds \right) - n^{-1} \left(\int_{-\infty}^{+\infty} \bar{K}_{h,\alpha} \, (x-s) f(s) ds \right)^2 .$$

The second term is the squared expected value of $\overline{f}_{h,\alpha}(x)$ divided by n and from (2.13),

$$Var\left(\overline{f}_{h,\alpha}(x)\right) = n^{-1}h^{-1} \left(\int_{-\infty}^{+\infty} \overline{K}_{\alpha}^2(u) f(x-uh) du \right) - n^{-1} \left(f(x) + O(h^2) \right)^2$$

$$= n^{-1}h^{-1} \int_{-\infty}^{c} \overline{K}_{\alpha}^2(u) \left(f(x) + o(1) \right) du + O(n^{-1}).$$

Hence,

$$Var\left(\overline{f}_{h,\alpha}(x)\right) = n^{-1}h^{-1} f(x) \int_{-\infty}^{c} \overline{K}_{\alpha}^2(u) du + o(n^{-1}h^{-1}).$$

We give the following lemma before proving Theorem 3.

Lemma 1 *Let G be a real valued function with domain $D(G)$, and $\{t_i\}$ is a set of equally spaced points on $D(G)$ with grid width Δ. If G is twice differentiable and its second derivative is integrable, then*

$$\left| \sum_i G(t_i)\Delta - \int G(s) ds \right| \le \frac{\Delta^2}{8} \int \left| G^{(2)}(t) \right| dt.$$

Therefore, if the grid width Δ is small enough, $\sum_i G(t_i)\Delta$ is approximately $\int G(s)ds$. This is simply the usual Riemann sum approximation, but with more precise bounds of the difference.

Proof of Lemma 1: We can write the difference between $\sum_{i=1}^{g} G(t_i)\Delta$ and $\int G(s)ds$ as a sum of integrals,

$$\left| \sum_{i=1}^{g} G(t_i)\Delta - \int G(s) ds \right| = \left| \sum_{i=1}^{g} G(t_i)\Delta - \sum_{i=1}^{g} \int_{t_i-\Delta/2}^{t_i+\Delta/2} G(s) ds \right|$$

$$= \left| \sum_{i=1}^{g} \int_{t_i-\Delta/2}^{t_i+\Delta/2} \left[G(t_i) - G(s) \right] ds \right| .$$

Taking a second order Taylor expansion around each x_i, the above absolute difference equals

$$\left| \sum_{i=1}^{g} \int_{t_i-\Delta/2}^{t_i+\Delta/2} \left[G^{(1)}(t_i)(t_i - s) + \frac{1}{2} \int_{s}^{t_i} (s-t) G^{(2)}(t) dt \right] ds \right|$$

$$= \frac{1}{2} \left| \sum_{i=1}^{g} \left[\int_{t_i-\Delta/2}^{t_i} \int_{t_i-\Delta/2}^{t} - \int_{t_i}^{t_i+\Delta/2} \int_{t}^{t_i+\Delta/2} \right] (s-t) G^{(2)}(t) ds dt \right|$$

$$\leq \frac{\Delta^2}{8}\sum_{i=1}^{g}\int_{t_i-\Delta/2}^{t_i+\Delta/2}\left|G^{(2)}(t)\right|dt = \frac{\Delta^2}{8}\int\left|G^{(2)}(t)\right|dt.$$

III. Proof of Theorem 3:

Only the boundary case is handled here and the results for interior points hold by letting $c \to \infty$ and by condition (A3). First, for $x = ch, c > 0$ and any $l = 0, 1, 2$, take $t_i = \frac{x_i - x}{h}, \Delta = \frac{b}{h}$, and $G(u) = u^l K(u) I_{(-c,\infty)}(u)$. Then Lemma 1 and assumption (A4) imply

$$S_{n,l}(x) = \frac{b}{h^{l+1}} S_{l,c}\left(1 + o(1)\right), l = 0, 1, 2.$$

Hence

$$K_{n,x}(u) = K_c^*(u)\left(1 + o(1)\right).$$

Now,

$$E\left(\hat{f}_L(x)\right) = n^{-1}h^{-1}\sum_{i=1}^{g} K_{n,x}\left(\frac{x_i - x}{h}\right) E\left(c_i\right)$$

$$= h^{-1}\sum_{i=1}^{g} K_c^*\left(\frac{x_i - x}{h}\right) f(x_i) b\left(1 + o(1)\right).$$

Once again, apply Lemma 1 with $t_i = x_i, \Delta = b$, and $G(u) = K_0^*(\frac{u-ch}{h}) f(u) I_{[0,\infty)}(u)$,

$$E\left(\hat{f}_L(x)\right) = h^{-1}\int_0^\infty K_c^*(\frac{u-x}{h}) f(u) du\left(1 + o(1)\right)$$

$$= \int_{-c}^{\infty} K_c^*(t) f(x + ht) dt\left(1 + o(1)\right).$$

Then, since f'' exists and is uniformly continuous at x,

$$E\left(\hat{f}_L(x)\right) = f(x) + hf'(x)\int_{-c}^{\infty} tK_c^*(t)dt + \frac{h^2}{2}f''(x)\int_{-c}^{\infty} t^2 K_c^*(t)dt + o(h^2)$$

$$= f(x) + \frac{h^2}{2}f''(x)\int_{-c}^{\infty} t^2 K_c^*(t)dt + o(h^2).$$

Similarly,

$$E\left(\hat{f}_L^2(x)\right) = n^{-2}h^{-2}\sum_{i=1}^{g}\sum_{j=1}^{g} K_{n,x}\left(\frac{x_i - x}{h}\right) K_{n,x}\left(\frac{x_j - x}{h}\right) E\left(c_i c_j\right)$$

$$= n^{-2}h^{-2}\sum_{i=1}^{g} K_c^*\left(\frac{x_i - x}{h}\right) K_c^*\left(\frac{x_i - x}{h}\right) nbf(x_i)\,(1 + o(1))$$

$$+n^{-2}h^{-2}\sum_{i=1}^{g}\sum_{j=1}^{g} K_c^*\left(\frac{x_i - x}{h}\right) K_c^*\left(\frac{x_j - x}{h}\right) n(n-1)b^2 f(x_i)f(x_j)\,(1 + o(1))$$

$$= \left\{\frac{1}{nh^2}\int_0^\infty K_c^*\left(\frac{u - x}{h}\right)^2 f(u)du + \frac{1}{h^2}\left[\int_0^\infty K_c^*(\frac{u-x}{h})f(u)du\right]^2\right\}(1 + o(1))\,.$$

Hence,

$$Var\left(\hat{f}_L(x)\right) = \frac{1}{nh^2}\int_0^\infty K_c^*\left(\frac{u - x}{h}\right)^2 f(u)du\,(1 + o(1))$$

$$= \frac{1}{nh}\int_{-c}^\infty K_c^*(t)^2 f(x + ht)dt\,(1 + o(1)) = \frac{f(x)}{nh}\int_{-c}^\infty K_c^*(t)^2 dt + o(\frac{1}{nh}).$$

Figure 2.1a: Constructing kernel density estimates

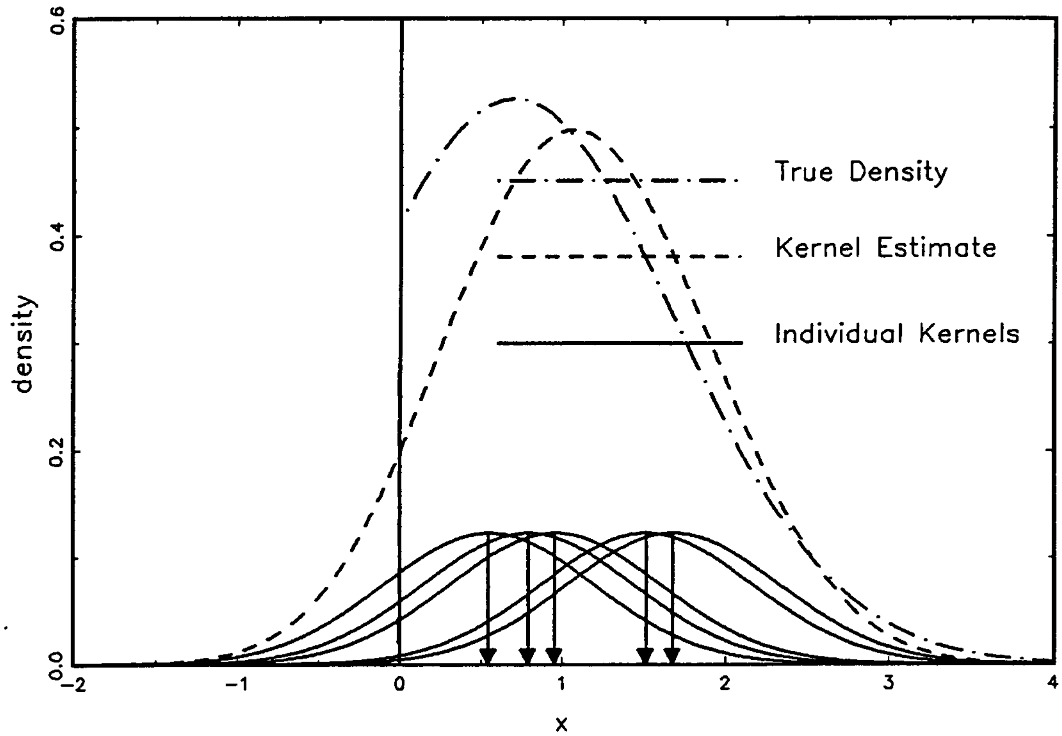


Figure 2.1b: Constructing boundary fold estimates

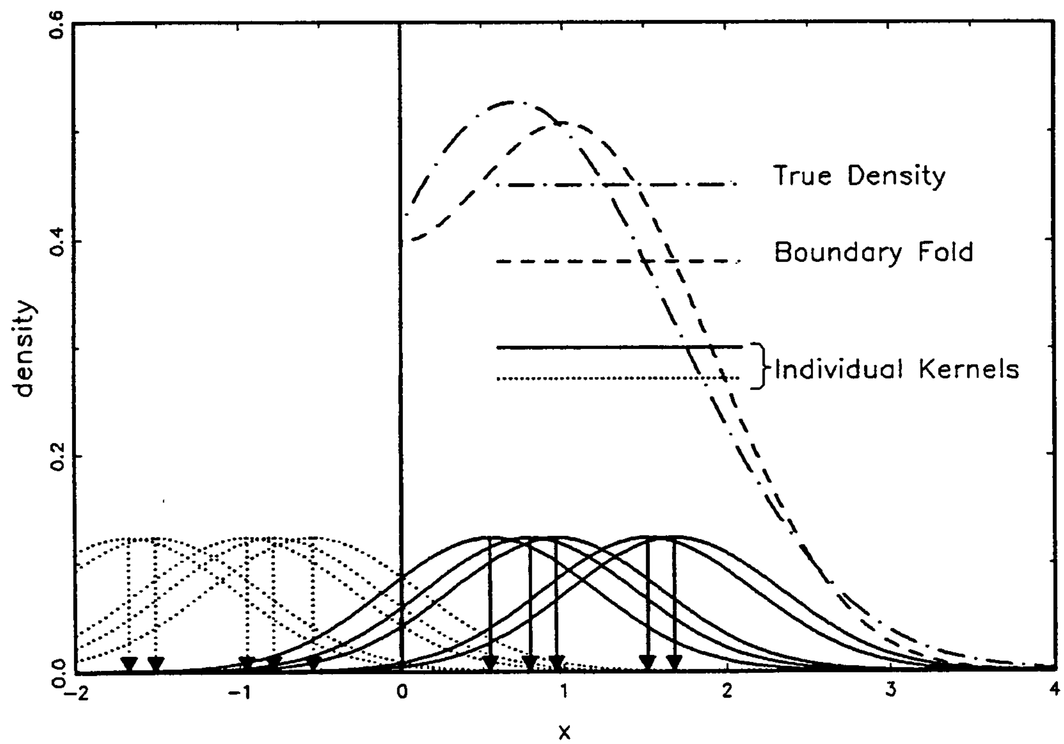


Figure 2.2a: Asymtotic bias of Rice modification

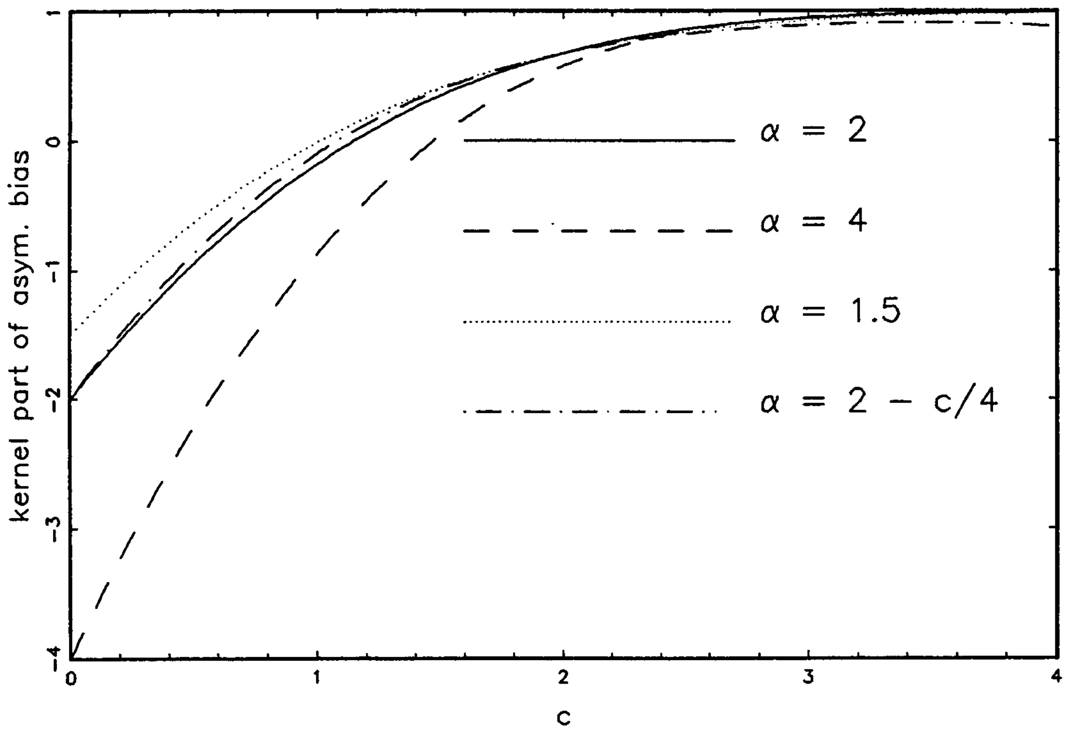


Figure 2.2b: Asymptotic variance of Rice modification

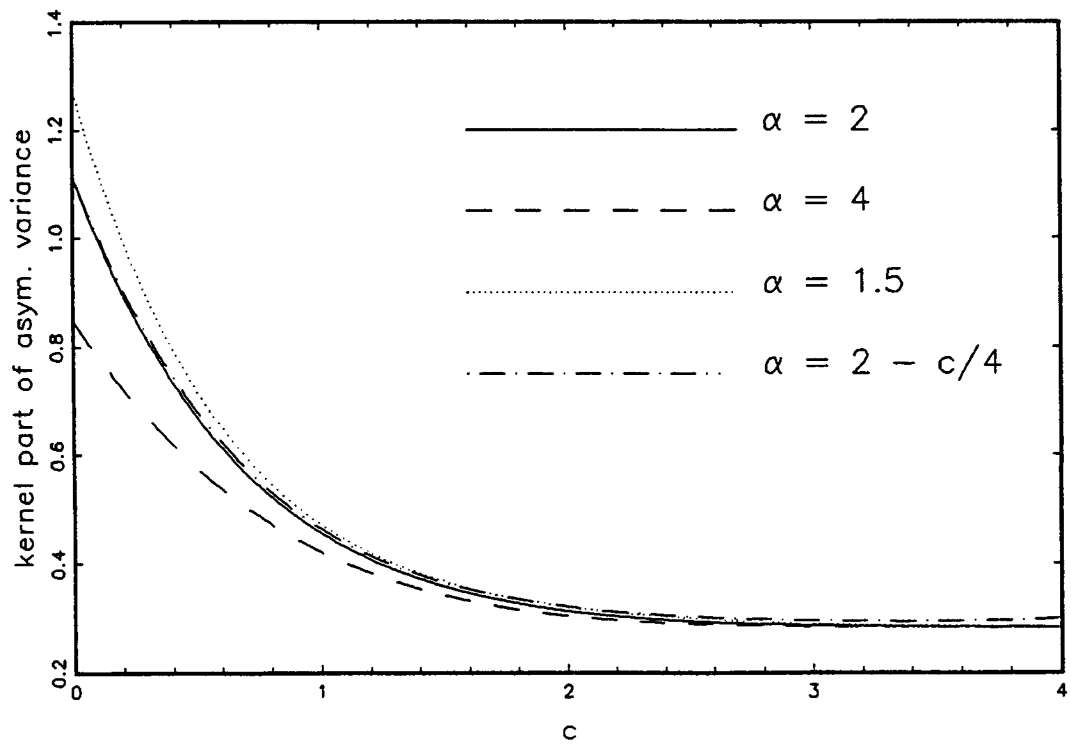


Figure 2.2c: Rice modification close to boundary

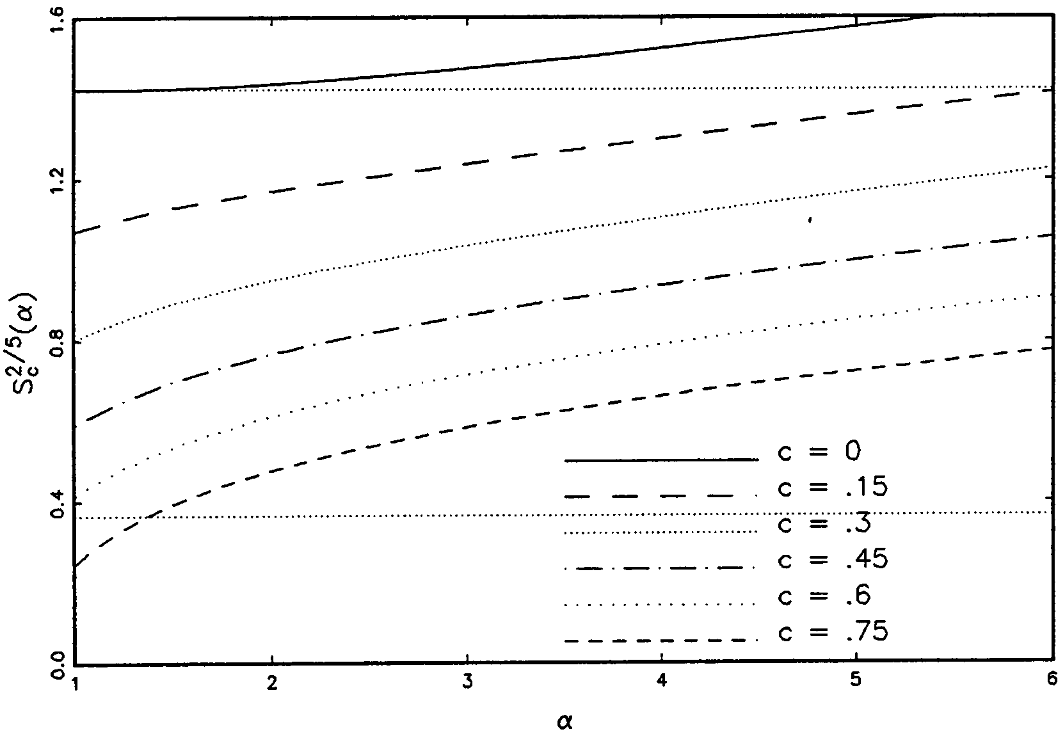


Figure 2.2d: Rice modification around 0.85

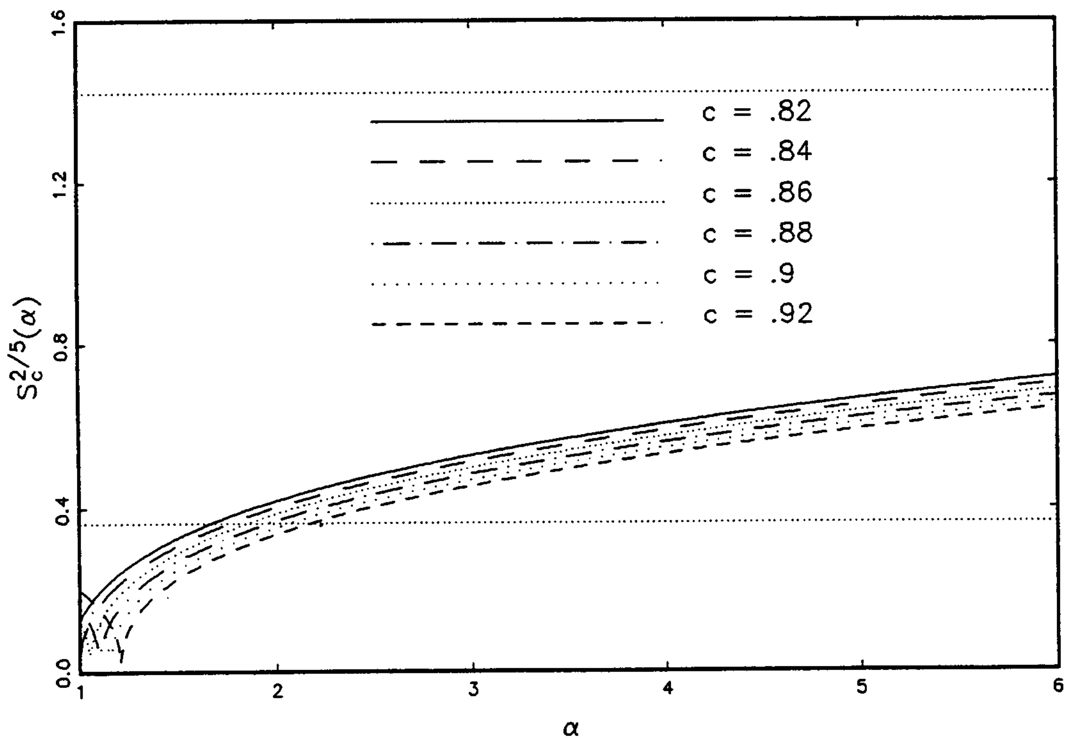


Figure 2.2e: Rice modification beyond 1

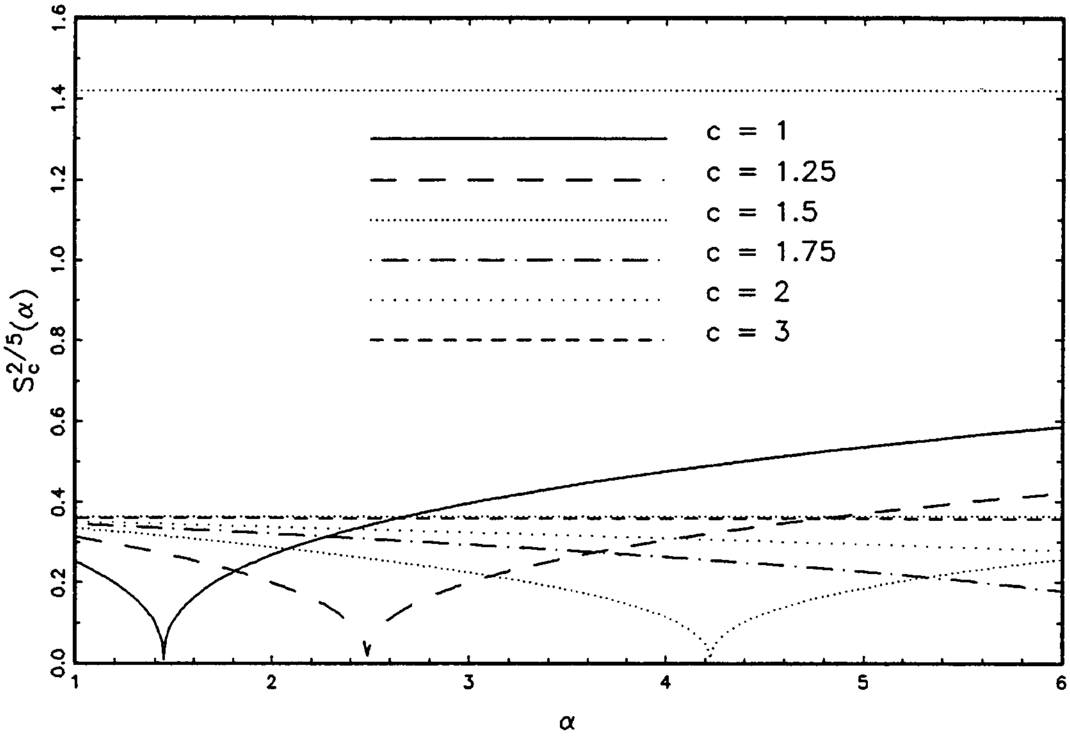


Figure 2.2f: Different asymptotic studies on Rice estimators

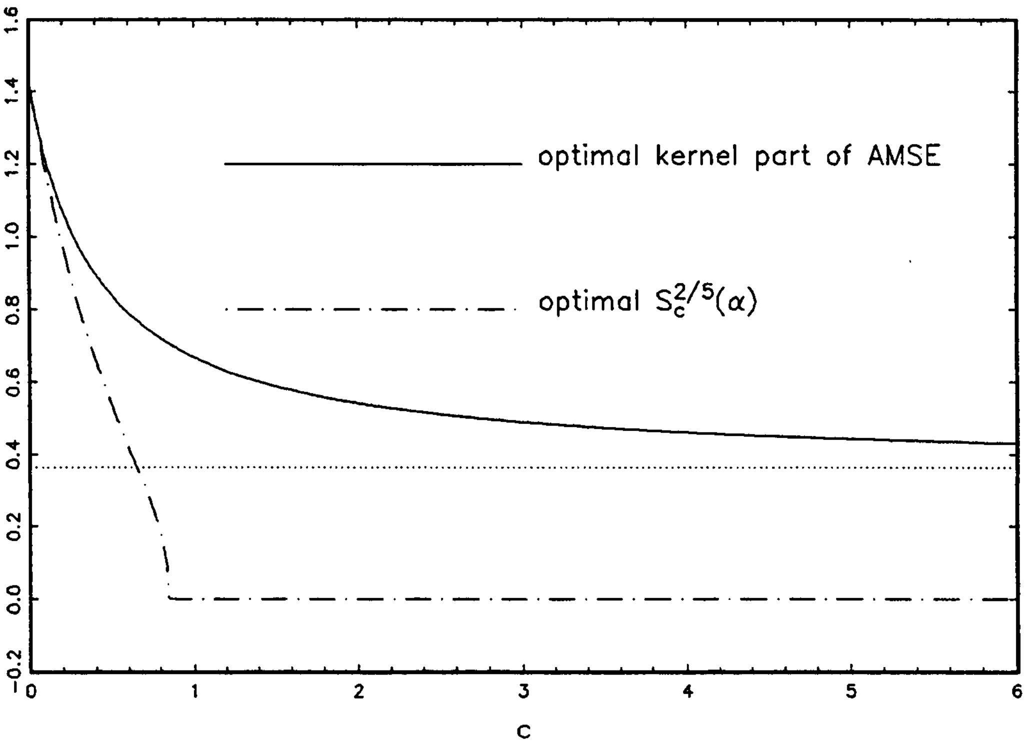


Figure 2.3: Truncated normal mixture densities

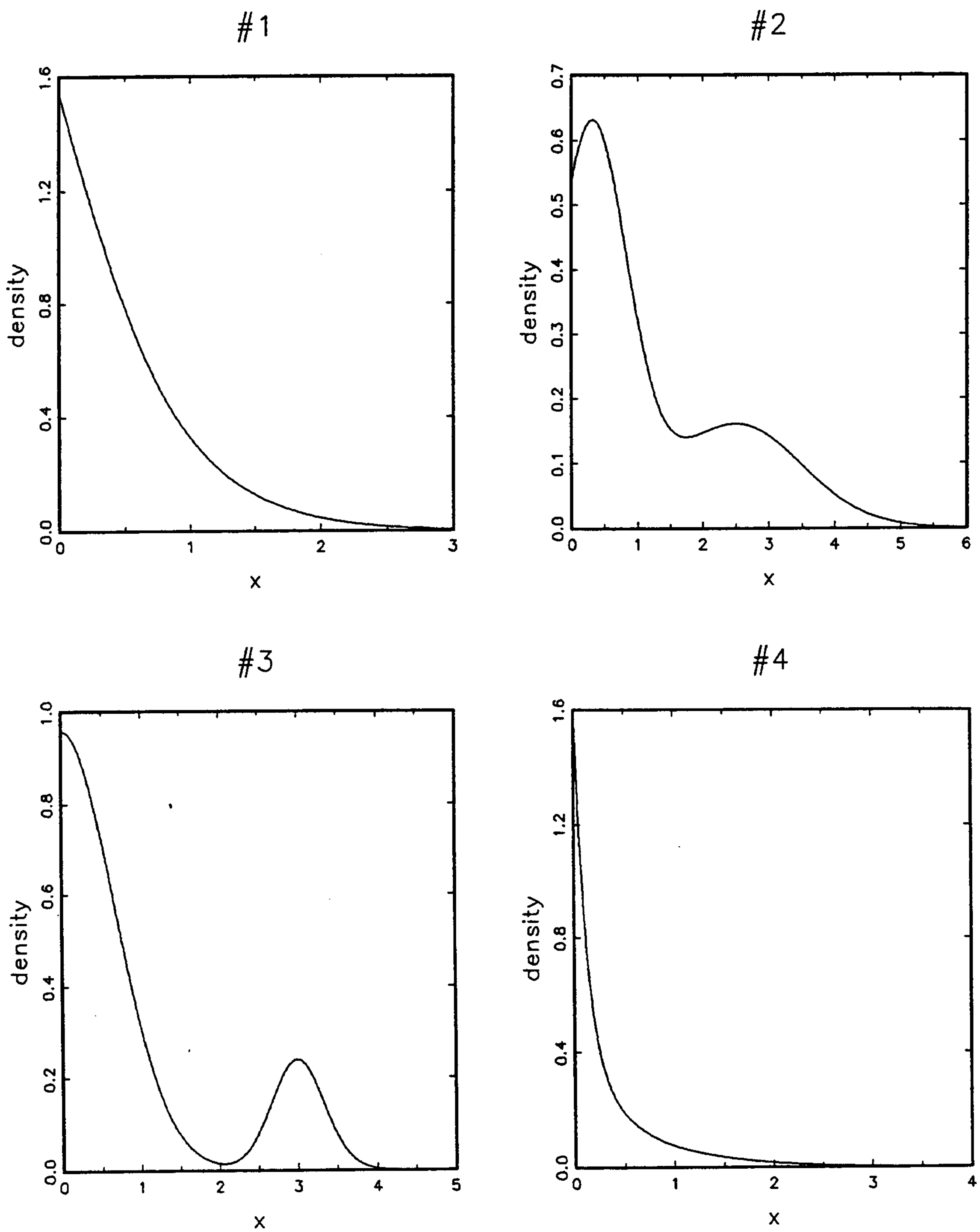


Figure 2.4: MSE of Rice's estimator at zero

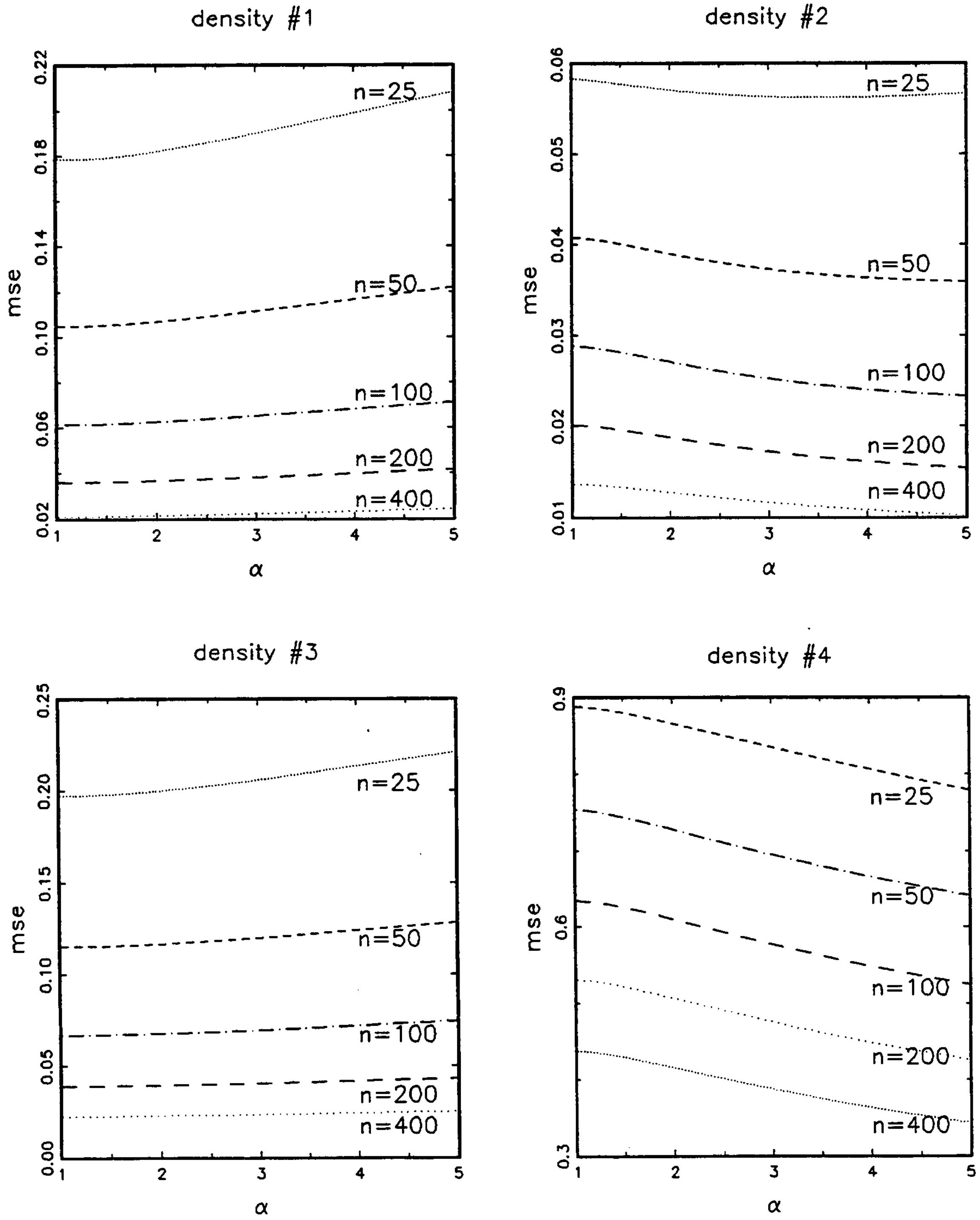


Figure 2.5: Rice boundary kernels

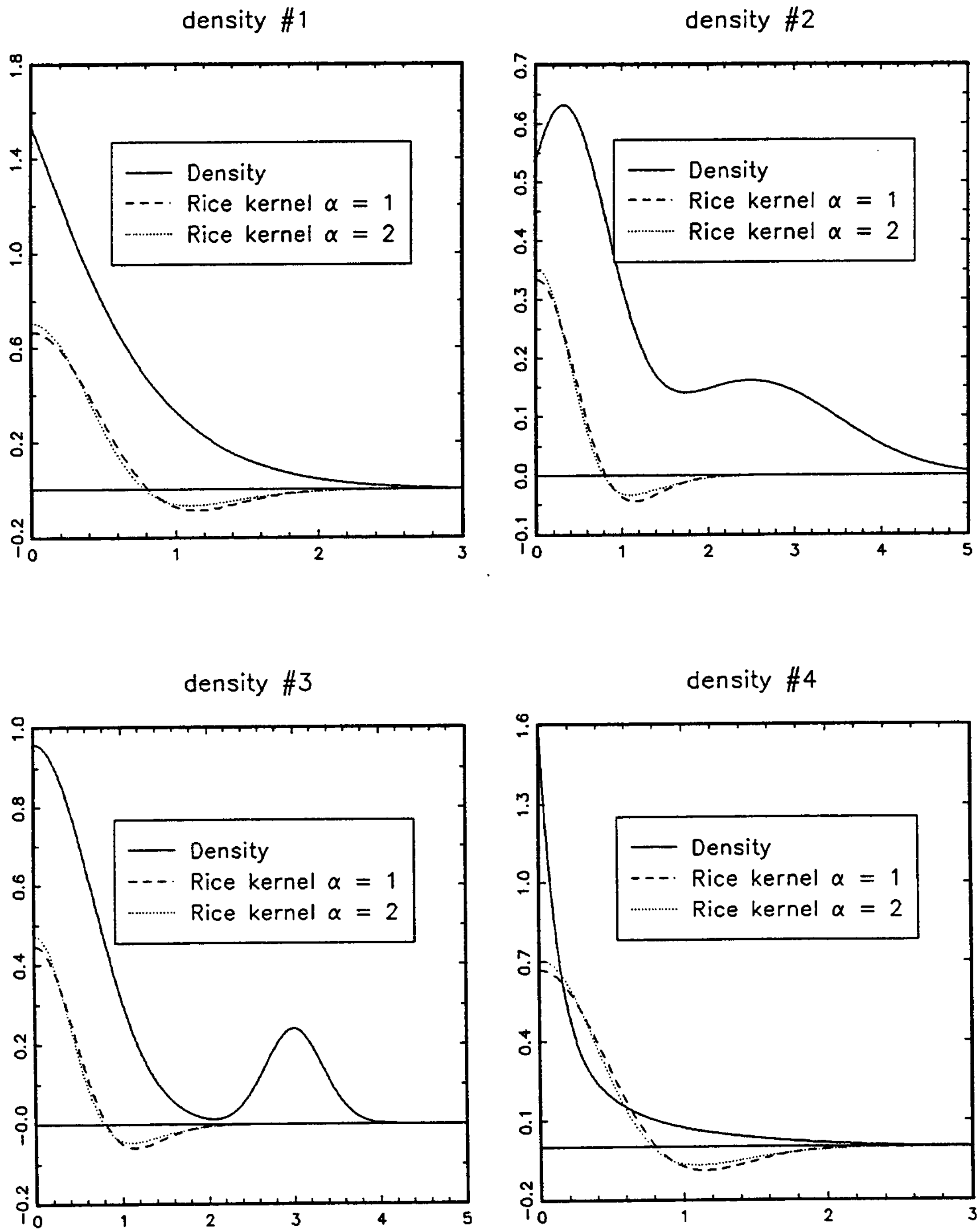


Figure 2.6: Estimates for density #1

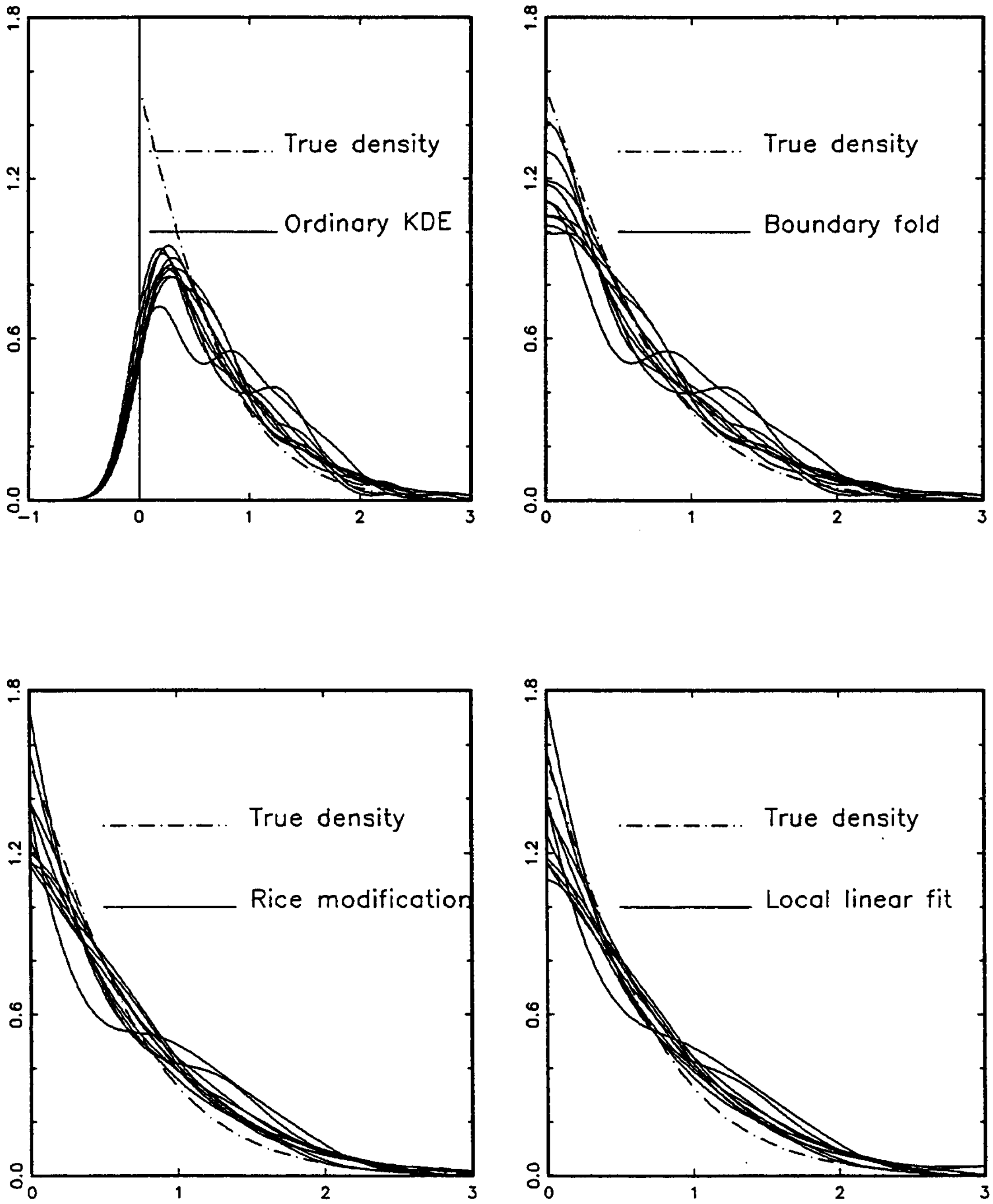


Figure 2.7: Estimates for density #2

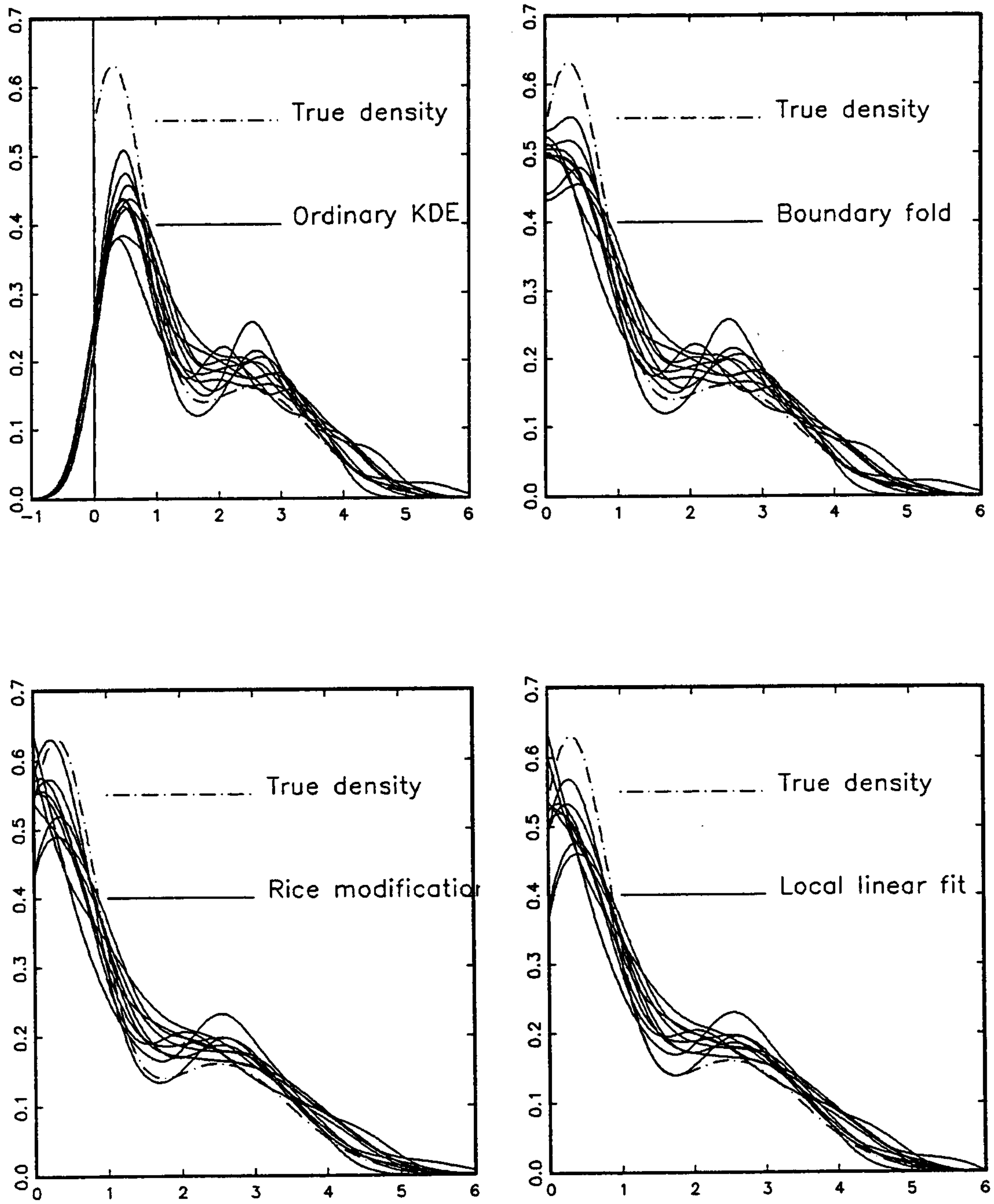
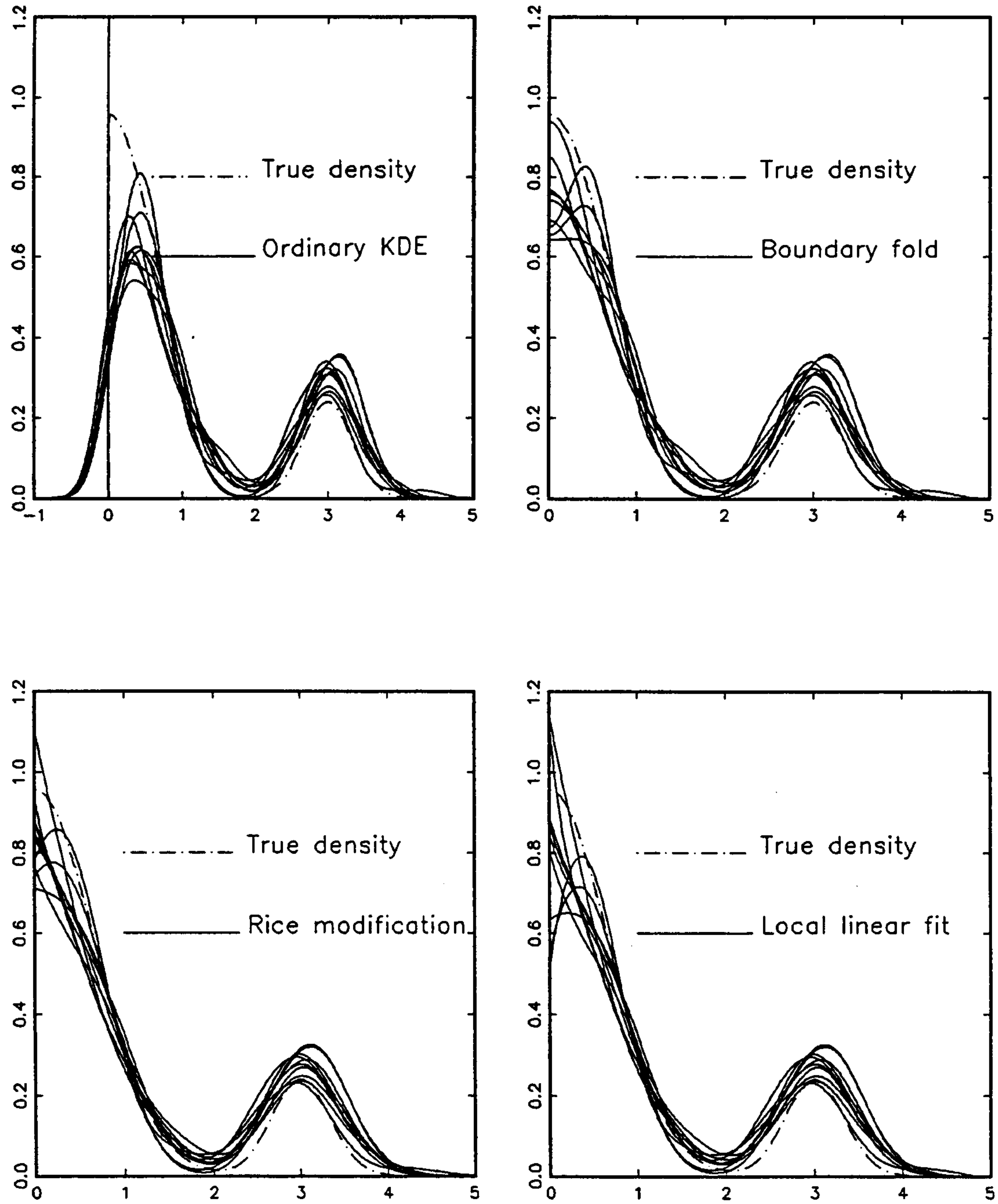


Figure 2.8: Estimates for density #3



Chapter 3

Best Constants

The main purpose of this chapter is to show that a local linear estimator is asymptotically efficient in the deep sense of constant coefficient for estimating density functions at endpoints from a minimax point of view. This result settles the important question of how local linear estimators compare with "optimal boundary kernels", by showing the former must be at least as efficient. We feel this gives local linear estimators an important advantage, because they are also easier to interpret, much easier to implement, and appear far faster to compute, see Fan and Marron (1994). We calculate the relative efficiency of the local linear estimator with the Gaussian kernel, and also that of Rice's boundary adjusted estimator. These results are extended to estimating function derivatives, since this is vital to applications in plug-in bandwidth selection.

Nonparametric minimax problems are interesting and challenging. Recent advances in this area can be found in, for example, Nussbaum (1985), Donoho and Liu (1991), Fan and Hall (1994), Donoho and Johnstone (1992), Brown and Low (1993), Efroimovich (1993), Fan (1993), and references therein. Most articles focus either on the minimax risk of estimating a whole function or on that of estimating a function at interior points. However, the minimax problem at a boundary point has not been studied yet. Here, we handle the effective higher order kernels through a representation in terms of the Legendre polynomials on an asymmetric interval.

3.1 Linear Estimators

Note from Section 2.4 that the local linear estimator $\hat{f}_L(x)$ is a weighted average of the bin counts,

$$\hat{f}_L(x) = \frac{1}{nh}\sum_{i=1}^{g} K_{n,x}\left(\frac{x_i - x}{h}\right) c_i,$$

where the bin counts are the c_i's which are linear functions of the data,

$$c_i = \sum_{k=1}^{n} I_{[x_i - \frac{b}{2}, x_i + \frac{b}{2})}(X_k).$$

Hence,

$$\begin{aligned}\hat{f}_L(x) &= \frac{1}{nh}\sum_{i=1}^{g}\sum_{k=1}^{n} K_{n,x}\left(\frac{x_i - x}{h}\right) I_{[x_i - \frac{b}{2}, x_i + \frac{b}{2})}(X_k) \\ &= \frac{1}{n}\sum_{k=1}^{n}\left\{\frac{1}{h}\sum_{i=1}^{g} K_{n,x}\left(\frac{x_i - x}{h}\right) I_{[x_i - \frac{b}{2}, x_i + \frac{b}{2})}(X_k)\right\}.\end{aligned}$$

Therefore, the estimator has the form

$$\hat{f}_L(x) = \frac{1}{n}\sum_{i=1}^{n} \psi(X_i, x),$$

for some function ψ. The conventional kernel density estimator and the Rice's modified version can also be written in such a form. Estimators of this type are called linear and may be expressed as

$$\hat{f}_\psi(x) = \int \psi(t, x) d\hat{F}_n(t),$$

where $\hat{F}_n$ is the empirical distribution function of the sample. Since the class of linear estimators cover many interesting density estimators, we will pay attention only to them while studying the best asymptotic mean squared error. The optimal criterion, among linear estimators for boundary estimation, we choose is discussed in the next section.

3.2 Minimax Sense

Apparently, mean squared error of linear estimators depends on the unknown density function. We can rule out this factor by examining an estimator's worst m.s.e over a certain class of densities and then find the minimum of them. This is the familiar minimax approach. There are some benefits from taking such an approach. First, even for a nonsense estimator, e. g. one that takes value of one everywhere no matter what the underlying distribution is, there always exists a density that makes it most favorable. But, such estimators will have large maximum m.s.e. over a family of densities and hence can not be an optimal one. Secondly, the true density has completely no effect on the optimization. This is an advantage over the optimal kernel approach which will make no sense if the underlying density has zero second derivative.

A class of smooth densities that reflects the idea of "twice differentiable (from the right) at $x = 0$" is the following.

$$C_{M,2} = \{f : |f| \leq M, |f(x) - f(0) - f'(0)x| \leq \frac{C}{2}x^2,\ x \geq 0\}, \tag{3.1}$$

where C and M are some fixed positive constants. A parallel class of functions which contains infinitely supported densities is considered in Donoho and Liu (1991). Then we denote the minimax mean squared error over linear estimators of $f(0)$ for this class as

$$R_{0,L}(n, C_{M,2}) = \inf_{\psi} \sup_{f \in C_{M,2}} E\left(\hat{f}_{\psi}(0) - f(0)\right)^2. \tag{3.2}$$

This is called the linear minimax lower bound for the boundary estimation problem. The next section is devoted to finding out what it is and what linear estimator, if any, achieves it.

3.3 Best Linear Estimator

We first give the following lemma.

Lemma 2 *Suppose $\xi \geq 0$ and η are functions on R^1, and b is a constant then*

$$\min_{\psi} \left(\int \psi\eta - b\right)^2 + \int \psi^2\xi = \frac{b^2}{1 + \int \frac{\eta^2}{\xi}},$$

and the minimum is attained when $\psi = \frac{b}{1+\int \frac{\eta^2}{\xi}} \frac{\eta}{\xi}$.

Now we are ready to give the minimax lower bound for linear estimators of $f(0)$. Proof of the following theorem is at the end of chapter.

Theorem 4 *With the definitions of $C_{M,2}$ and $R_{0,L}(n, C_{M,2})$ given in (3.1) and (3.2),*

$$R_{0,L}(n, C_{M,2}) = 3 \cdot 15^{-1/5} \left(\frac{\sqrt{C}M}{n}\right)^{4/5} (1 + o(1)), \tag{3.3}$$

and the best linear estimator is the local linear fit estimator with kernel weight function $(1-u)I_{[0,1]}(u)$ and $h = \left(\frac{480M}{C^2 n}\right)^{1/5}$.

Remark 3.1. Fan (1993) derives the linear minimax lower bound for estimating a regression function at an interior point. It is also shown there that a local linear smoother is fully efficient in that case. These results can be easily adapted to density estimation. Hence, local linear estimators are efficient both in the interior and boundaries.

Remark 3.2. If the kernel function is symmetric about zero then there are on average only half as many observations as at the interior used in estimating $f(0)$. In order for the bias to be comparable to that in the interior, the weights of the bin counts are roughly doubled. Hence the variances at the boundary are about four times larger. This reflects in the best possible risks; the constant multiplier in (3.3) is exactly four times of that for the minimax risk when estimating at interior points, c.f. (4.5) of Fan (1993) which is given there for interior points in the regression setting but is also true for density estimation.

Remark 3.3. Suppose there is a right-hand-side boundary point, say f has support $[0, 1]$, a similar argument leads to a best linear smoother for estimating $f(1)$ which is $\hat{f}_L(1)$ with kernel $(1+u)I_{[-1,0]}(u)$ and $h = \left(\frac{480M}{C^2 n}\right)^{1/5}$.

Remark 3.4. An analogous minimax risk for the regression context is available. Moreover, it is possible to extend the results to derivative estimation and in that case the estimator obtained from fitting local least squares polynomials with the same kernel will achieve the minimax lower bound. See Cheng, Fan and Marron (1993) for more details.

3.4 Relative Efficiency

Now that the best possible risk for linear estimators is known, it can be used as a base line for measuring the performance of other boundary estimators. Define the efficiency of a linear estimator $\hat{f}(0)$ of $f(0)$ as

$$\text{Eff}\left(\hat{f}(0)\right) = \lim_{n\to\infty} \left(\frac{R_{0,L}(n, C_{M,2})}{\sup\limits_{f\in C_{M,2}} E\left(\hat{f}(0) - f(0)\right)^2} \right)^{5/4}.$$

The power $5/4$ puts efficiency on the traditional and interpretable "sample size scale" since both numerator and denominator have asymptotic rate of convergence $n^{-4/5}$.

In the previous sections we have shown that the local linear estimator is 100% efficient with the kernel weight function $K_0 = (1-u)I_{[0,1]}(u)$. The Gaussian density function is often used in kernel smoothing methods since it make the estimators visually more pleasant; i.e. smooth and without undesired angles. The local linear estimator with Gaussian kernel is easily shown to have efficiency

$$\left(\frac{3 \cdot 15^{-1/5}}{\frac{5}{4}\left(\frac{\pi-4}{\pi-2}\right)^{2/5} \left(\frac{\sqrt{\pi}(\pi+1-\sqrt{8})}{(\pi-2)^2}\right)^{4/5}} \right)^{5/4} \approx 0.9802$$

at end points. Hence there is very small lost of efficiency while gaining the visual benefit of the Gaussian kernel in local linear fitting.

Another important boundary corrected method whose efficiency is considered here is Rice's modification. For Rice's estimator, it is not known which kernel function combined with what bandwidth ratio, i.e. the ratio of the two bandwidths in the combination, will give the best performance. Here the relative efficiencies for the well known Epanechnikov kernel and the popular Gaussian kernel are considered. The best bandwidth ratio for the Gaussian kernel at $x = 0$ is shown to be 1, in Figure 2c, which results in the effective kernel

$$\varphi_1(u) = 2(2 - u^2)\varphi(u),$$

and this estimator has efficiency

$$\left(\frac{3 \cdot 15^{-1/5}}{\frac{5}{4} \left(\frac{11}{4\sqrt{\pi}} \right)^{4/5}} \right)^{5/4} \approx 0.9783.$$

Using the Epanechnikov kernel for Rice's estimator, the best bandwidth ratio for $x = 0$ can be shown to be $1 + \sqrt{5/2}$. Then the efficiency is

$$\left(\frac{3 \cdot 15^{-1/5}}{\frac{5}{4} \left(\frac{18(8\sqrt{10}+24)^2}{125(2+\sqrt{10})^3} \right)^{4/5}} \right)^{5/4} \approx 0.9517.$$

From these numbers we can conclude that the Rice's modification is an excellent method of boundary adjustment in terms of efficiency. However, the local linear fit estimators are preferred. First, they are highly interpretable while the Rice's modification is motivated by purely technical considerations. Moreover, since they benefit from linear binning, they can be fast implemented.

Although there are many advantages of the local linear estimators, the choice of bandwidth dominates their performance. Later, we will attempt to develop a plug-in bandwidth selector. In that situation, estimating density derivatives are crucial. In the remaining sections of this chapter, the results of Theorem 4 are extended to that case for the purpose of effective density derivative estimation.

3.5 Estimating Density Derivatives

Cleveland (1979) discusses local least squares regression methods. Smooth weighted local polynomial approximation for estimating the regression function and its derivatives is discussed by Fan and Gijbels (1992b). This idea can be implemented to estimating density derivatives. The motivation is that the density at a bin center is equal to its normalized bin count plus an error term and is well approximated by a polynomial locally if it is smooth there. Since now we aim at estimating the derivatives, we need more smoothness of f inside its support. Then it is reasonable to fit a higher degree polynomial locally. It is convenient to investigate the more general situation of estimating the ν^{th} derivative of the density.

If f has $(p+1)-$th derivative, the Taylor's function approximation is

$$f(z) \approx \sum_{j=0}^{p} \frac{f^{(j)}(x)}{j!}(z-x)^j, \tag{3.4}$$

in a neighborhood of x which is any point in the support of f. This suggests a locally least squares polynomial fit of $f(x)$ to the normalized bin counts $\{n^{-1}b^{-1}c_i, i = 1, ..., g\}$:

$$\min_{\underset{\sim}{\beta}} \sum_{i=1}^{g} \left(n^{-1}b^{-1}c_i - \sum_{j=0}^{p} \beta_j (x_i - x)^j \right)^2 K\left(\frac{x_i - x}{h}\right), \tag{3.5}$$

where K is a weight function and h is a smoothing parameter controlling the range of the polynomial fitting. Denote the solution of the least squares problem (3.5) as $\widehat{b}_j(x), j = 0, ..., p$. From (3.4), an estimator of $f^{(\nu)}(x)$ is $\widehat{f}^{(\nu)}(x) = \nu!\widehat{b}_\nu(x)$. To obtain explicit formula for these estimators, write

$$X = \left((x_i - x)^{j-1}\right)_{1\le i\le g, 1\le j\le p+1}, Y = n^{-1}b^{-1}\left(c_1, ..., c_g\right)^T,$$

$$\widehat{b}(x) = \left(\widehat{b}_0(x), ..., \widehat{b}_p(x)\right), \text{ and } W = \operatorname{diag}\left(K\left(\frac{x_i - x}{h}\right)\right).$$

Then, by standard weighted least squares theory,

$$\widehat{b}(x) = \left(X^T W X\right)^{-1} X^T W Y. \tag{3.6}$$

Observe from (3.6) that each entry of $\widehat{b}(x)$, and therefore each $\widehat{f}^{(\nu)}(x)$, is a weighted sum of the bin counts. In order to obtain some insights into how the weights are distributed to the bin counts, consider defining

$$S_n = \left(S_{n,i+j-2}(x)\right)_{0\le i,j\le p+1},$$

where

$$S_{n,j}(x) = \sum_{i=1}^{g} K\left(\frac{x_i - x}{h}\right)(x_i - x)^j, j = 0, 1, ..., 2p.$$

Then, simple algebra shows that

$$\widehat{b}_\nu(x) = \sum_{i=1}^{g} W_\nu^n\left(\frac{x_i - x}{h}\right) n^{-1}b^{-1}c_i, \tag{3.7}$$

where

$$W_\nu^n(t) = e_{\nu+1}^T S_n^{-1}(1, ht, ..., h^p t^p)^T K(t),$$

with $e_{\nu+1}^{\mathrm{T}}$ being the $(\nu+1)^{\text{th}}$ unit vector whose $(\nu+1)^{\text{th}}$ element equals one and all the other entries are zero. See Fan et al. (1993) for details of the above calculation. In connection to higher order kernel approaches such as those discussed in Gasser, Müller, and Mammitzsch (1985), if $b = o(h)$ then

$$S_{n,j}(x) = b^{-1}h^{j+1}S_j(1 + o(1)),$$

where $S_j = \int_{-\infty}^{+\infty} t^j K(t)dt$. Hence, if we write $S = (S_{i+j-2})_{0\le i,j\le p+1}$, then

$$W_\nu^n(t) \approx \frac{b}{h^{\nu+1}} e_{\nu+1}^{\mathrm{T}} S^{-1}(1, t, ..., t^p)^{\mathrm{T}} K(t). \tag{3.8}$$

Therefore, we have the "equivalent kernel" representation

$$\widehat{b}_\nu(x) \approx \frac{1}{nh^{\nu+1}} \sum_{i=1}^{g} K_\nu^* \left(\frac{x_i - x}{h}\right) c_i, \tag{3.9}$$

where

$$K_\nu^*(t) = e_{\nu+1}^{\mathrm{T}} S^{-1}(1, t, ..., t^p)^{\mathrm{T}} K(t). \tag{3.10}$$

We remark that (3.9) is provided here only to illustrate relations between the local polynomial fit approach and the higher order kernel methods. Papers giving such connections include Müller (1987) and Lejeune (1985). Although it suggests similar properties of the two approaches, advantages of the local polynomial fits include much better interpretability, automatic boundary corrections, ease in implementation and requiring much less computation time.

Next, we briefly discuss some asymptotic properties of the derivative estimators. If there is any edge point, without loss of generality, consider that f is supported on $[0, \infty)$. To study what happens on the appropriate neighborhood of the boundary, let $x = ch$, $c \ge 0$, then $S_j = \int_{-c}^{+\infty} t^j K(t)dt$. We write the equivalent kernel as $K_{\nu,c}^*$ in order to mark its dependence on x. In that case $K_\nu^*(t) = \lim_{c\to\infty} K_{\nu,c}^*(t)$. The equivalent kernel $K_{\nu(,c)}^*$ satisfies the moment conditions

$$\int_{-\infty(-c)}^{\infty} t^q K_{\nu(,c)}^*(t)dt = e_{\nu+1}^T S^{-1} S e_{q+1} = \delta_{\nu,q} \text{ for } 0 \le \nu, q \le p. \tag{3.11}$$

The following proposition is important to understanding the behavior of $\widehat{b}_m$. It can be proved by arguments similar to those in the proof of Theorem 3 with (3.11). So we state it without going into the proof.

Proposition 1 *Suppose $h \to 0, nh^{2\nu+1} \to \infty$, and $b = o(h)$. Then as $n \to \infty$, the asymptotic mean squared error of $\widehat{b}_\nu(x)$ is*

$$E\left(\widehat{b}_\nu(x) - \frac{f^{(\nu)}(x)}{\nu!}\right)^2 = \left(\int_{-\infty(-c)}^{+\infty} t^{p+1} K^*_{\nu(,c)}(t)dt\right)^2 \left(\frac{f^{(p+1)}(x)}{(p+1)!}\right)^2 h^{2(p+1-\nu)}$$

$$+\frac{f(x)}{nh^{2\nu+1}} \int_{-\infty(-c)}^{+\infty} K^{*2}_{\nu(,c)}(t)dt + o\left(h^{2(p+1-\nu)} + \frac{1}{nh^{2\nu+1}}\right).$$

Therefore $\widehat{f}^{(\nu)} = \nu!\widehat{b}_\nu$ retains the same rate of convergence everywhere including boundary points. The conventional kernel method uses kernels for derivative estimation which are very hard to interpret. This problem becomes even worse at boundary regions. Comparatively, the local polynomial fitting naturally provides interpretable and effective estimators, e.g. in the sense of rate of convergence, and requires no additional complicated boundary modifications.

3.6 Best Constants for Derivative Estimation

Now some minimax theory for general derivative estimation at endpoints is developed. We shall focus on optimizing over the class of linear estimators as was introduced in Section 3.1. We discuss this problem for estimating at the right endpoint $x = 0$. The class of smooth density functions parallel to $C_{M,2}$ is

$$C_{M,p+1} = \left\{ f : |f| \le M, \left| f(z) - \sum_{j=0}^{p} \frac{f^{(j)}(0)}{j!} z^j \right| \le C \frac{|z|^{p+1}}{(p+1)!}, z \ge 0 \right\}. \quad (3.12)$$

Define

$$R_{0,L}(\nu) = \inf_{\widehat{T}_\nu \text{linear}} \sup_{f \in C_{M,p+1}} E\left\{ \left(\widehat{T}_\nu - f^{(\nu)}(0)\right)^2 \middle| X_1, ..., X_n \right\}$$

as the linear minimax risk for estimating $f^{(\nu)}(0)$. Let

$$r = \frac{2(p+1-\nu)}{2p+3}, s = \frac{2\nu+1}{2p+3},$$

and

$$\theta_{\nu,p} = \left(\frac{2p+3}{2\nu+1}\right) \left(\frac{(p+\nu+2)!}{(p-\nu+1)!\nu!}\right)^2 \left(\frac{rM}{2(p+\nu+2)n}\right)^r \left(\frac{(p+1)!C}{(2p+3)!}\right)^{2s} \quad (3.13)$$

Theorem 5 *The linear minimax risk for estimating the ν^{th} derivative of the density function at its right endpoint is*

$$R_{0,L}(\nu) = \theta_{\nu,p}(1 + o(1)).$$

Theorem 6 *Let $\widehat{f}^{(\nu)}(0)$ be the ν^{th} derivative estimator resulting from a local polynomial fit of order p with the kernel function $K_0(u) = (1-u)I_{[0,1]}(u)$ and bandwidth $h = \Lambda n^{-\frac{1}{2p+3}}$, where $\Lambda = \left(\frac{(p-\nu+1)(2p+3)!^2 M}{(p+\nu+2)(p+1)!^2(2p+3)C^2}\right)^{\frac{1}{2p+3}}$. Then it is the best linear estimator of $f^{(\nu)}(0)$ in the sense that*

$$\frac{R_{0,L}(\nu)}{\sup\limits_{f\in C_{M,p+1}} E\left\{\left(\widehat{f}^{(\nu)}(0) - f^{(\nu)}(0)\right)^2\right\}} \to 1, \text{ as } n \to \infty.$$

Moreover, its equivalent kernel is

$$K^{opt}_{\nu,p+1}(t) = \sum_{j=0}^{p+1} \lambda_j t^j I_{[0,1]}(t),$$

where

$$\lambda_j = \frac{(-1)^{j+\nu}(p+j+1)!(p+\nu+2)!}{j!^2\nu!(p-\nu)!(p-j+1)!(j+\nu+1)}, j = 0, 1, ..., p+1.$$

In Fan et al. (1993) the local polynomial fit estimators of regression function derivatives are shown to be minimax efficient in the interior. Similar results hold for estimators $\widehat{f}^{(\nu)}$. Although it was noticed earlier that the Rice's modified kernel estimators of the density are highly efficient, we don't know whether a similar bias reduction technique is applicable to derivative estimation. Again the "optimal boundary kernels" for estimating derivatives are extremely hard to interpret. Since the local polynomial estimators follow a simple and intuitive rule with many nice properties, alternatives are not investigated here.

3.7 Proofs

I. Proof of Lemma 2.

First note

$$\left(\int \psi\eta - b\right)^2 + \int \psi^2\xi \geq \min_t \left(\min_{\int \psi\eta = t} (t-b)^2 + \int \psi^2 \xi\right).$$

But under the constraint $\int \psi\eta = t$,

$$\left(\int \psi^2\xi\right)\left(\int \frac{\eta^2}{\xi^2}\cdot\xi\right) \geq \left(\int \psi\cdot\frac{\eta}{\xi}\cdot\xi\right)^2 = t^2,$$

where equality holds when $\psi = \frac{t}{\int \frac{\eta^2}{\xi}}\frac{\eta}{\xi}$. Hence,

$$\left(\int \psi\eta - b\right)^2 + \int \psi^2\xi \geq \min_t \left((t-b)^2 + \frac{t^2}{\int \frac{\eta^2}{\xi}}\right) = \frac{b^2}{1+\int \frac{\eta^2}{\xi}},$$

where the minimum is attained by $t = \frac{b}{1+\left(\int \frac{\eta^2}{\xi}\right)^{-1}}$.

II. Proof of Theorem 4.

The estimator $\widehat{f}_\psi(x)$ has mean squared error

$$\mathrm{E}\left(\widehat{f}_\psi(0) - f(0)\right)^2 = \left(\int \psi(t,0)f(t)dt - f(0)\right)^2 + \frac{1}{n}Var\left(\psi(X_1,0)\right)$$

For any $f_1, f_2 \in C_{M,2}$

$$\sup_{f\in C_{M,2}} \mathrm{E}\left(\widehat{f}_\psi(0) - f(0)\right)^2 \geq \frac{1}{4}\left\{\left[\int \psi(f_1-f_2) - (f_1(0)-f_2(0))\right]^2 + \frac{2}{n}\int \psi^2(f_1+f_2)\right\}$$
$$-\frac{1}{n}\left\{\left(\int \psi f_1\right)^2 + \left(\int \psi f_2\right)^2\right\}.$$

By Lemma 2, minimizing the first term in the right-hand-side of the above inequality over all ψ yields

$$\sup_{f\in C_{M,2}} \mathrm{E}\left(\widehat{f}_\psi(0) - f(0)\right)^2 \geq \frac{1}{4}\frac{(f_1(0)-f_2(0))^2}{1+\frac{n}{2}\int\frac{(f_1-f_2)^2}{f_1+f_2}}$$

$$-\frac{n}{4}\left(\frac{f_1(0)-f_2(0)}{1+\frac{n}{2}\int\frac{(f_1-f_2)^2}{f_1+f_2}}\right)^2\left\{\left(\int\frac{f_1(f_1-f_2)}{f_1+f_2}\right)^2 + \left(\int\frac{f_2(f_1-f_2)}{f_1+f_2}\right)^2\right\}. \qquad (3.14)$$

Let $f_1(x) = g_0(x) + g_n(x) - c_n$ and $f_2(x) = g_0(x) - g_n(x) + c_n$, where

$$g_0(x) = \left[-\frac{(M-\delta)^2}{2}x + (M-\delta)\right] I_{[0,\frac{2}{M-\delta}]}(x),$$

$$g_n(x) = \frac{b_n^2}{2}\left(\frac{1}{2} - \frac{3\sqrt{C}x}{2b_n} + \frac{Cx^2}{b_n^2}\right) I_{[0,\frac{b_n}{\sqrt{C}}]}(x),$$

with $b_n = \left(\frac{480\sqrt{C}M}{n}\right)^{1/5}$, $c_n = \int g_n = \frac{b_n^3}{24\sqrt{C}}$, and

$$\max\left(\frac{b_n^2}{4}, \frac{b_n^2}{32} + c_n\right) \le \delta \le \min\left(\frac{M^2 b_n}{2\sqrt{C}}, M - \frac{M^2 b_n}{\sqrt{C}}\right).$$

Then f_1, $f_2 \in C_{2,M}$ and from (3.14) we have

$$\begin{aligned}
&\inf_{\psi} \sup_{f\in C_{M,2}} \mathrm{E}\left(\hat{f}_\psi(0) - f(0)\right)^2 \\
&\ge \left\{\frac{1}{4}\frac{4g_n(0)^2}{1+\frac{n}{2}\int\frac{4g_n^2}{2g_0}} - 2n\left(\frac{g_n(0)}{1+n\int\frac{g_n^2}{g_0}}\right)^2 c_n^2\right\}(1+o(1)) \\
&= \frac{g_n(0)^2}{1+\frac{n}{g_0(0)}\int g_n^2}(1+o(1)) \\
&= \frac{b_n^4/16}{1+\frac{n}{M}\frac{b_n^5}{120\sqrt{C}}}(1+o(1)) \\
&= 3\cdot 15^{-1/5}\left(\frac{\sqrt{C}M}{n}\right)^{4/5}(1+o(1)). \qquad (3.15)
\end{aligned}$$

On the other hand, let $\hat{f}_{\psi_0}$ be the local linear fit estimator $\hat{f}_L(0)$ with kernel $K_0(u) = (1-u)I_{[0,1]}(u)$ and bandwidth h. Then by Theorem 3,

$$\begin{aligned}
&\inf_{\psi} \sup_{f\in C_{M,2}} \mathrm{E}\left(\hat{f}_\psi(0) - f(0)\right)^2 \le \sup_{f\in C_{M,2}} \mathrm{E}\left(\hat{f}_{\psi_0}(0) - f(0)\right)^2 \\
&\le \frac{h^4}{4}\left(\int_0^1 u^2 K_0^*(u)du\right)^2 C^2 + \frac{1}{nh}\left(\int_0^1 K_0^*(u)^2 du\right) M + o\left(h^4 + \frac{1}{nh}\right) \\
&= 3\cdot 15^{-1/5}\left(\frac{\sqrt{C}M}{n}\right)^{4/5}(1+o(1)). \qquad (3.16)
\end{aligned}$$

The last equality is valid with $h = \left(\frac{480M}{C^2 n}\right)^{1/5}$. Combining (3.15) and (3.16) finishes the proof of the theorem.

III. Proofs of Theorems 5 and 6:

Denote $K^{opt}_{\nu,p+1}$ as the equivalent kernel of $\widehat{f}^{(\nu)}(0) = \nu!\widehat{b}_\nu(0)$, given by (3.10), with $K(u) = K_0(u) = (1-u)I_{[0,1]}(u)$. We prove the theorems using the norm $\left\|K^{opt}_{\nu,p+1}\right\|$ and the $(p+1)^{th}$ moment of $K^{opt}_{\nu,p+1}$. The calculation of these quantities and the function $K^{opt}_{\nu,p+1}$ is very technically involved and appears in Subsection **IV**. First we construct an upper bound for the linear minimax risk $R_{0,L}(\nu)$,

$$\sup_{f\in C_{M,p+1}} E\left\{\left(\widehat{f}^{(\nu)}(0) - f^{(\nu)}(0)\right)^2\right\}$$

$$\leq \left(\frac{C}{(p+1)!}\int_0^1 t^{p+1}K^{opt}_{\nu,p+1}(t)dt\right)^2 h^{2(p+1-\nu)} + \int_0^1 K^{opt\,2}_{\nu,p+1}(t)dt\frac{M}{nh^{2\nu+1}}$$

$$\equiv A_1 h^{2(p+1-\nu)} + A_2 h^{-(2\nu+1)}.$$

Take $h = \left(\frac{(2\nu+1)A_2}{2(p+1-\nu)A_1}\right)^{1/(2p+3)} n^{-1/(2p+3)}$ which minimizes the above quantity, then

$$\sup_{f\in C_{M,p+1}} E\left\{\left(\widehat{f}^{(\nu)}(0) - f^{(\nu)}(0)\right)^2\right\}$$

$$\leq A_1^s A_2^r (2p+3)(2\nu+1)^{-s}\left[2(p+1-\nu)\right]^{-r}(1+o(1))$$

$$= r^{-r}s^{-s}\left(\frac{C}{(p+1)!}\right)^{2s}\left(\frac{M}{n}\right)^r\left(\int_0^1 t^{p+1}K^{opt}_{\nu,p+1}(t)dt\right)^{2s}\left\|K^{opt}_{\nu,p+1}\right\|^{2r}(1+o_p(1)).$$

From (3.26) and (3.28) the above expression equals

$$\left(\frac{2(p+1-\nu)}{2p+3}\right)^{-r}\left(\frac{2\nu+1}{2p+3}\right)^{-s}\left(\frac{C}{(p+1)!}\right)^{2s}\left(\frac{M}{n}\right)^r$$

$$\times\left(\frac{(p+\nu+2)!(p+1)!^2}{\nu!(2p+3)!(p-\nu+1)!}\right)^{2s}\left(\frac{2(p+\nu+2)(p+\nu+1)!^2}{(2\nu+1)(2p+3)\nu!^2(p-\nu)!^2}\right)^r$$

$$= \theta_{\nu,p}\ , \tag{3.17}$$

as defined in (3.13). Hence

$$R_{0,L}(\nu) \leq \theta_{\nu,p}(1+o(1)).$$

A quantity closely related to the $R_{0,L}(\nu)$ is the modulus of continuity

$$\omega_{p+1,\nu}(\epsilon) = \sup\left\{\left|f_1^{(\nu)}(0) - f_0^{(\nu)}(0)\right| : f_0, f_1 \in C_{M,p+1}, \|f_1 - f_0\| = \epsilon\right\}. \tag{3.18}$$

To establish a lower bound for $R_{0,L}(\nu)$, let

$$a = \left(\frac{C}{(p+1)!\,|\lambda_{p+1}|}\right)^{1/(p+1)},$$

and

$$f(x) = \begin{cases} K^{opt}_{\nu,p+1}(ax) & , \text{ if } 0 \le x \le 1. \\ o & , \text{ otherwise.} \end{cases}$$

Then

$$\|f\|^2 = \frac{\left\|K^{opt}_{\nu,p+1}\right\|^2}{a}, f^{(\nu)}(0) = a^\nu \nu! \lambda_\nu. \tag{3.19}$$

Take

$$f_0(x) = g_0(x) + g_1(x), f_1(x) = g_0(x) - g_1(x),$$

where

$$g_0(x) = \left[-\frac{(M-\sigma)^2}{2}x + (M-\sigma)\right] I_{[0,\frac{2}{M-\sigma}]}(x),$$

$$g_1(x) = \delta^{p+1} f(x/\delta),$$

with $\delta = \left(\frac{\epsilon^2}{4\|f\|^2}\right)^{\frac{1}{2p+3}}$. For some suitable σ, $f_0, f_1 \in C_{M,p+1}$, we have

$$\|f_1 - f_0\|^2 = 4\delta^{2p+3} \|f\|^2 = \epsilon^2.$$

Therefore the modulus of continuity defined in (3.18) satisfies

$$\omega_{p+1,\nu}(\epsilon) \ge \left|f_1^{(\nu)}(0) - f_0^{(\nu)}(0)\right| = 2\left|f^{(\nu)}(0)\right| \left(\frac{\epsilon^2}{4\|f\|^2}\right)^{\frac{p+1-\nu}{2p+3}}. \tag{3.20}$$

From (3.19), (3.20) becomes

$$\omega_{p+1,\nu}(\epsilon) \ge 2\nu! \left|\lambda_\nu\right| \left(\frac{C}{(p+1)! \left|\lambda_{p+1}\right|}\right)^s \left\|K^{opt}_{\nu,p+1}\right\|^{-r} \left(\frac{\epsilon^2}{4}\right)^{r/2}$$

$$= 2\nu! \left(\frac{(p+\nu+1)!^2(p+\nu+2)}{\nu!^3(p-\nu)!^2(p-\nu+1)(2\nu+1)}\right) \left(\frac{C(p+1)!^2\nu!(p-\nu)!(p+\nu+2)}{(p+1)!(2p+2)!(p+\nu+1)!}\right)^s$$

$$\times \left(\frac{(2\nu+1)(2p+3)}{2(p+\nu+2)}\right)^{r/2} \left(\frac{\nu!(p-\nu)!}{(p+\nu+1)!}\right)^r \left(\frac{\epsilon}{2}\right)^r$$

$$= \epsilon^r \left(\frac{(p+\nu+2)!}{\nu!(p-\nu+1)!(2\nu+1)}\right) \left(\frac{2C(p+1)!}{(2p+2)!}\right) \left(\frac{(2\nu+1)(2p+3)}{2(p+\nu+2)}\right)^{r/2}.$$

Applying Theorem 6 of Fan (1993), we have

$$R_{0,L}(\nu) \ge r^r s^s \left[\nu! \lambda_\nu \left(\frac{C}{(p+1)! \left|\lambda_{p+1}\right|}\right)^s \left\|K^{opt}_{\nu,p+1}\right\|^{-r} \left(\frac{M}{n}\right)^{r/2}\right]^2$$

$$= \left(\frac{C}{(p+1)!}\right)^{2s} \left(\frac{M}{n}\right)^{r} \frac{r^r s^s \nu!^2 \lambda_\nu^2}{\lambda_{p+1}^{2s} \left\|K_{\nu,p+1}^{opt}\right\|^{2r}}.$$

Equations (3.27) and (3.28) give

$$R_{0,L}(\nu) \geq r^r s^s \nu!^2 \left(\frac{C}{(p+1)!}\right)^{2s} \left(\frac{M}{n}\right)^{r} \left(\frac{\nu!(p-\nu)!(p+1)!^2}{(p+\nu+1)!(2p+2)!}\right)^{2s}$$

$$\times \left(\frac{(p+\nu+1)!^2(p+\nu+2)}{\nu!^3(p-\nu)!^2(2\nu+1)(p-\nu+1)}\right)^2 \left(\frac{(2\nu+1)(2p+3)\nu!^2(p-\nu)!^2}{2(p+\nu+2)(p+\nu+1)!^2}\right)^r$$

$$= \theta_{r,\nu}$$

given in (3.13).

In summary, (i) since the upper and lower bounds are the same, we have finished the proof of Theorem 5, (ii) the maximum risk of $\widehat{f}^{(\nu)}(0)$ is given in (3.17) and the first part of Theorem 6 follows immediately, the second part is shown in (3.27). Thus we complete the proof.

IV. Calculation of the function $K_{\nu,p+1}^{opt}$, its norm and *(p+1)*th moment.

The Legendre polynomials on the interval [-1,+1] are defined as

$$P_n(x) = \frac{d^n}{dx^n}\left((1+x)(1-x)\right)^n, -1 \leq x \leq 1,\ n = 0,1,2,...$$

The linear transformation $y = (x+1)/2$ yields an orthogonal system with respect to the Lebesgue measure on $[0,1]$. Write

$$Q_n(y) = \frac{d^n}{dy^n}\left(y(1-y)\right)^n \equiv \sum_{j=0}^{n} q_{n,j} y^j \ n = 0,1,2,...$$

Then

$$\|Q_n\|^2 = \int_0^1 Q_n^2(y)dy = (-1)^n \int_0^1 y^n(1-y)^n \frac{d^{2n}}{dy^{2n}}\left(y(1-y)\right)^n dy$$

$$= \int_0^1 y^n(1-y)^n(2n)!dy = (2n)!\frac{n!^2}{(2n+1)!} = \frac{n!^2}{2n+1}. \qquad (3.21)$$

Explicitly,

$$Q_n(x) = \frac{d^n}{dy^n}\sum_{j=0}^{n}\binom{n}{j}(-y)^j y^n = \sum_{j=0}^{n}\binom{n}{j}(-1)^j\frac{(n+j)!}{j!}y^j.$$

So,

$$q_{n,j} = \begin{pmatrix} n \\ j \end{pmatrix} (-1)^j \frac{(n+j)!}{j!}, n = 0, 1, ..., p+1, j = 0, 1, ..., n.$$

Let $K^{opt}_{\nu,p+1}$ denote the equivalent kernel of $\widehat{f}^{(\nu)}(0) = \nu!\widehat{b}_\nu(0)$, given by (3.10), with $K(u) = K_0(u) = (1-u)I_{[0,1]}(u)$. Since $K^{opt}_{\nu,p+1}$ is a polynomial of order $(p+1)$, we can write

$$K^{opt}_{\nu,p+1}(x) = \sum_{i=0}^{p+1} a_i Q_i(x).$$

The coefficients a_i can be determined by the moment properties in (3.11). Let $\beta = \int_0^1 x^{p+1} K^{opt}_{\nu,p+1}(x)dx$. Then

$$a_i \|Q_i\|^2 = \int_0^1 Q_i(x) K^{opt}_{\nu,p+1}(x)dx = \begin{cases} 0 & , \text{ if } 0 \le i < \nu. \\ \nu! q_{i,\nu} & , \text{ if } \nu \le i \le p. \\ \nu! q_{p+1,\nu} + q_{p+1,p+1}\beta & , \text{ if } i = p+1. \end{cases} \tag{3.22}$$

Therefore, from (3.21) and (3.22),

$$\frac{1}{\nu!} K^{opt}_{\nu,p+1}(x) = \sum_{i=0}^{p} q_{i,\nu} \frac{(2i+1)}{i!^2} Q_i(x) + \frac{(2p+3)}{(p+1)!^2} (q_{p+1,\nu} + \frac{q_{p+1,p+1}}{\nu!}\beta) Q_{p+1}(x)$$

$$= \sum_{i=0}^{p} q_{i,\nu} \frac{(2i+1)}{i!^2} \sum_{j=0}^{i} q_{i,j} x^j + \frac{(2p+3)}{(p+1)!^2} (q_{p+1,\nu} + \frac{q_{p+1,p+1}}{\nu!}\beta) Q_{p+1}(x)$$

$$= \sum_{j=0}^{p} \left(\sum_{i=j \vee \nu}^{p} q_{i,\nu} \frac{(2i+1)}{i!^2} q_{i,j} \right) x^j + \frac{(2p+3)}{(p+1)!^2} (q_{p+1,\nu} + \frac{q_{p+1,p+1}}{\nu!}\beta) Q_{p+1}(x). \tag{3.23}$$

Here,

$$\sum_{i=j \vee \nu}^{p} q_{i,\nu} \frac{(2i+1)}{i!^2} q_{i,j} = \frac{(-1)^{j+\nu}}{j!^2 \nu!^2} \sum_{i=j \vee \nu}^{p} \frac{(i+\nu)!(2i+1)(j+i)!}{(i-\nu)!(i-j)!}$$

(Note: $(2i+1) = \{(i+j+1)(i+\nu+1) - (i-j)(i-\nu)\} / (j+\nu+1)$.)

$$= \frac{(-1)^{j+\nu}}{j!^2 \nu!^2 (j+\nu+1)} \sum_{i=j \vee \nu+1}^{p} \left(\frac{(i+\nu+1)!(j+i+1)!}{(i-\nu)!(i-j)!} - \frac{(i+\nu)!(j+i)!}{(i-\nu-1)!(i-j-1)!} \right)$$

$$+ \frac{(-1)^{j+\nu}}{j!^2 \nu!^2} \frac{((j \vee \nu)+\nu)!(2(j \vee \nu)+1)(j+(j \vee \nu))!}{((j \vee \nu)-\nu)!((j \vee \nu)-j)!}$$

$$= \frac{(-1)^{j+\nu}}{j!^2 \nu!^2 (j+\nu+1)} \frac{(p+\nu+1)!(j+p+1)!}{(p-\nu)!(p-j)!}. \tag{3.24}$$

Also, since $Q_i(1) = \frac{d^i}{dy^i}\left(y(1-y)\right)^i\Big|_{y=1} = (-1)^i i!$ and $K^{opt}_{\nu,p+1}(1) = 0$ (see (3.10))

$$K^{opt}_{\nu,p+1}(1) = \sum_{i=0}^{p+1} a_i Q_i(1) = \sum_{i=0}^{p+1} a_i (-1)^i i! = 0.$$

This is the same as

$$\sum_{i=\nu}^{p} \frac{(2i+1)}{i!^2}\nu! q_{i,\nu}(-1)^i i! + \frac{(2p+3)}{(p+1)!^2}(\nu! q_{p+1,\nu} + q_{p+1,p+1}\beta)(-1)^{p+1}(p+1)! = 0. \quad (3.25)$$

The first term is

$$\sum_{i=\nu}^{p} \frac{(2i+1)}{i!^2}\nu! \begin{pmatrix} i \\ \nu \end{pmatrix} (-1)^\nu \frac{(i+\nu)!}{\nu!}(-1)^i i! = \frac{(-1)^\nu}{\nu!^2} \sum_{i=\nu}^{p} \frac{(-1)^i (i+\nu)!(2i+1)}{(i-\nu)!}$$

(Note: $(2i+1) = (i+\nu+1) - (i-\nu)$.)

$$= \frac{(-1)^\nu}{\nu!}\left[\sum_{i=\nu}^{p} \frac{(-1)^i(i+\nu+1)!}{(i-\nu)!} + \sum_{i=\nu+1}^{p} \frac{(-1)^i(i+\nu)!}{(i-\nu-1)!}\right] = \frac{(-1)^{\nu+p}(p+\nu+1)!}{\nu!(p-\nu)!}.$$

Thus equation (3.25) yields

$$\beta = \frac{(-1)^{\nu+p}(p+\nu+2)!(p+1)!^2}{\nu!(2p+3)!(p-\nu+1)!}. \quad (3.26)$$

Combining this with equations (3.23) and (3.24) we have

$$K^{opt}_{\nu,p+1}(x) = \sum_{j=0}^{p+1} \lambda_j x^j,$$

where

$$\lambda_j = \frac{(-1)^{j+\nu}(p+j+1)!(p+\nu+2)!}{j!^2\nu!(p-\nu)!(p-j+1)!(j+\nu+1)}, j = 0, 1, ..., p+1. \quad (3.27)$$

Since the polynomials $\{Q_i\}$ are orthogonal,

$$\left\|K^{opt}_{\nu,p+1}\right\|^2 = \sum_{i=0}^{p+1} a_i^2 \|Q_i\|^2 = \sum_{i=\nu}^{p} \nu!^2 q_{i,\nu}^2 \frac{2i+1}{i!^2} + \frac{(2p+3)}{(p+1)!^2}\left(\nu! q_{p+1,\nu} + q_{p+1,p+1}\beta\right)^2$$

$$= \sum_{i=\nu}^{p} \frac{(2i+1)(i+\nu)!^2}{\nu!^2(i-\nu)!^2} + \frac{(2p+3)}{(p+1)!^2}\left(\nu! q_{p+1,\nu} + q_{p+1,p+1}\beta\right)^2$$

From (3.26), (Note: $(2i+1) = \{(i+\nu+1)^2 - (i-\nu)^2\}/(2\nu+1)$.)

$$= \frac{1}{\nu!^2(2\nu+1)} \sum_{i=\nu+1}^{p} \left\{\frac{(i+\nu+1)!^2}{(i-\nu)!^2} - \frac{(i+\nu)!^2}{(i-\nu-1)!^2}\right\} + \frac{(2\nu)!^2(2\nu+1)}{\nu!^2}$$

$$+ \frac{(p+\nu+1)!^2}{(2p+3)\nu!^2(p-\nu)!^2}$$

$$= \frac{2(p+\nu+2)(p+\nu+1)!^2}{(2\nu+1)(2p+3)\nu!^2(p-\nu)!^2}. \quad (3.28)$$

Chapter 4

Bandwidth Choice

Local least squares technique are highly interpretable with clear intuition. And they perform excellently in both theoretical and practical senses. Particularly, they adapt to boundary estimation automatically and do not suffer from boundary effects. They also enjoy some important optimal properties as shown in the last chapter. In applications, they are easy to compute and fast implementations exist, see Fan and Marron (1994). The bandwidth controls the neighborhood of the local smoothing and hence is crucial. The goal of this chapter is to find a global data-based bandwidth selector for the local linear density estimators.

Inspired by the outstanding performances of the Sheather and Jones (1991) bandwidth selector as reported in Jones, Marron and Sheather (1994), we attempt to bring in the root finding plug-in idea. The Sheather and Jones bandwidth selector is motivated from an explicit expression of the asymptotic optimal bandwidth for kernel estimators. An examination on the integrated mean squared error shows that such an expression holds for the local linear estimators as well. Therefore it is natural to consider this style of plug-in idea here.

Conventional estimators of the integrated square density second derivative are relative inefficient due to boundary effects, see Van Es and Hoogstrate (1993a). As a consequence, the Sheather-Jones plug-in bandwidth tends to zero at a rate different from $n^{-1/5}$, which is the rate of the asymptotically optimal bandwidth. See Van Es and Hoogstrate (1993b). Therefore, we construct estimators that are more robust to boundary problems and then incorporate them and their asymptotically optimal bandwidths in the bandwidth selection procedure with the goal that it will work in both boundary and non-boundary cases.

Weak convergence and asymptotic normality of the resulting bandwidth are established. Interestingly, the rate of convergence to the asymptotically optimal bandwidth depends on the sign of the integrated product of the second and fourth derivatives of the density: a rate of $n^{-5/14}$, same as Sheather and Jones (1991) in non-boundary cases, is achieved if the quantity is negative and otherwise we have a slower rate of $n^{-2/7}$. Note that if the underlying density essentially has no boundary features; i.e. its second and third derivatives are zero at boundaries, the above mentioned functional is less than zero by virtue of an integration by parts. Such a difference already appears in estimating the integrated squared density second derivative and is carried over to the bandwidth selector. However, the bandwidth selector is always consistent to the optimal bandwidth no matter whether there is a nonsmooth boundary or not. Ruppert, Sheather and Wand (1993) develop a bandwidth selector for local linear regression estimator which uses similar ideas. Theory as deep as that of this study has not yet been developed for their selector, but it is conjectured that analogous asymptotic behavior will be observed.

4.1 Plug-in Rules

First, we review the Sheather and Jones plug-in bandwidth selector for the conventional kernel estimators. Now the kernel function K is fixed and the integrated mean squared error of the kernel estimator $\hat{f}_h$ is defined as

$$MISE(h) = E\int \{f(x) - \hat{f}_h(x)\}^2 dx.$$

Under conditions (B1) and (B2) below, as $n \to \infty, h \to 0$, and $nh \to \infty$, this quantity is well approximated by the asymptotic integrated mean squared error

$$AMISE(h) = \frac{h^4}{4}\mu_2^2(K)R(f'') + \frac{1}{nh}R(K), \tag{4.1}$$

where $R(\xi) \equiv \int \xi^2$ for any function ξ, in the sense that

$$MISE(h) = AMISE(h) + o\left(h^4 + \frac{1}{nh}\right).$$

(B1) K satisfies $\int uK(u)du = 0, |u^2K(u)|$ is bounded, $\int |u^2K(u)|\, du < \infty$ and $\int K^2 < \infty$.

(B2) f is twice differentiable with f'' uniformly continuous and for some $\alpha \in (0,1]$, $|f''(x-t) - f''(x)| \le M(x)\,|t|^\alpha$ where $\int M < \infty$.

The optimal bandwidth is the minimizer of $MISE(h)$, denoted by h_0. The best bandwidth h_0 which optimizes $MISE(h)$ is asymptotically equal to h_*, the minimizer of $AMISE(h)$. Therefore one way to approach h_0 is through finding h_*. It is easy to derive from (4.1) that

$$h_* = \left\{ \frac{R(K)}{\mu_2^2(K)R(f'')} \right\}^{1/5} n^{-1/5}. \tag{4.2}$$

This expression makes it clear that the best bandwidth depends on the true density f. In the formula for h_*, only the quantity $R(f'')$ is unknown. Hence, the plug-in bandwidth approach is to substitute some estimate of $R(f'')$ into (4.2). One possibility is to estimate $R(f'')$ by $R\left(\hat{f}''(\cdot;g)\right)$ since $\hat{f}''(\cdot;g)$ is a natural estimator of f''. Note that here a bandwidth g different from that for estimating the density is used. This is an appealing idea since estimating a functional of the density is a different problem from estimating the density itself, see discussions in Jones, Marron and Sheather (1994).

Jones and Sheather (1991) show that the optimal g for $R\left(\hat{f}''(\cdot;g)\right)$ is

$$g_* = C_1(K)R^{-1/7}\left(f^{(3)}\right)n^{-1/7}, \tag{4.3}$$

where $C_1(K) = \left(\frac{2K^{(4)}(0)}{\mu_2(K)}\right)^{1/7}$. From (4.2) and (4.3),

$$g_* = C_2(K)R^{-1/7}\left(f^{(3)}\right)R^{1/7}\left(f^{(2)}\right)h_*^{5/7}, \tag{4.4}$$

for some appropriate $C_2(K)$. Sheather and Jones (1991) apply this relation back to (4.2) and their bandwidth selector is the solution in H of the equation

$$H = \left\{ \frac{R(K)}{\mu_2^2(K)R(\hat{f}''(\cdot;g_1(H)))} \right\}^{1/5} n^{-1/5},$$

where g_1 is the same function as in (4.4) with H replacing h_*. Note that in solving the equation (4.4) $R\left(f^{(2)}\right)$and $R\left(f^{(3)}\right)$ are unknown and have to be estimated. Here some accuracy is necessary; it requires some estimate $\widehat{T}$ such that $\widehat{T} = R\left(f^{(3)}\right) + O_p\left(n^{-1/14}\right)$. Any of the consistent estimators discussed in Hall and Marron (1987) or Jones and Sheather (1991) is enough at least asymptotically. Consistency and small sample properties of this bandwidth selector are discussed in Sheather and Jones (1991), Sheather (1992) and Jones, Marron and Sheather (1994).

4.2 Asymptotically Optimal Bandwidth

Local linear estimators are different in form from the conventional kernel estimators. Essentially, they are linear combinations of the bin counts with coefficients depending on the location of x. In attempting to develop a plug-in bandwidth selector for them, the first and the most natural question is whether their asymptotic optimal bandwidths have the simple representation of (4.2). Hence, we start with finding the asymptotic optimal bandwidth. Results from Theorem 3 are: when x is an interior point,

$$E\left(\hat{f}_L(x) - f(x)\right)^2 = \frac{h^4}{4}\left(f''(x)\right)^2 \mu_2(K) + \frac{1}{nh} f(x) R(K) + o\left(h^4 + \frac{1}{nh}\right), \quad (4.5)$$

or when x is a boundary point, $x = ch, c \geq 0$,

$$E\left(\hat{f}_L(x) - f(x)\right)^2 = \frac{h^4}{4}\left(f''(0+)\right)^2 \mu_2(K_c^*) + \frac{f(0+)}{nh} R(K_c^*) + o\left(h^4 + \frac{1}{nh}\right), \quad (4.6)$$

where

$$K_c^*(u) = \frac{S_{2,c} - uS_{1,c}}{S_{2,c}S_{0,c} - S_{1,c}S_{1,c}} K(u) I_{[-c,\infty)}(u),$$

with $S_{j,c} = \int_{-c}^{+\infty} t^j K(t)dt, j = 0, 1, 2$. With conditions (B1), (B2), by the Fubini Theorem

$$\begin{aligned} MISE\left(\hat{f}_L; h\right) &= \int E\left(\hat{f}_L(x) - f(x)\right)^2 dx \\ &= \int AMSE\left(\hat{f}_L(x); h\right) dx + o\left(h^4 + \frac{1}{nh}\right), \end{aligned}$$

where $AMSE\left(\widehat{f}_L(x); h\right)$ is the dominating term of the mean squared error. When f is supported on the entire real line, from (4.5),

$$\int AMSE\left(\widehat{f}_L(x); h\right) dx = \frac{h^4}{4}\left(\mu_2(K)\right)^2 R\left(f''\right) + \frac{R(K)}{nh}.$$

Otherwise, $AMSE\left(\widehat{f}_L(x); h\right)$ depends on x in a more complicated way, see (4.6). Nonetheless,

$$\int_0^h AMSE\left(\widehat{f}_L(x); h\right) dx = h\int_0^1 AMSE\left(\widehat{f}_L(ch); h\right) dc$$

$$= h\int_0^1 \left\{\frac{h^4}{4}\left(f''(0+)\right)^2 \int_{-c}^{\infty} u^2 K_c^*(u)du + \frac{f(0+)}{nh}\int_{-c}^{\infty}\left(K_c^*(u)\right)^2 du\right\} dc \qquad (4.7)$$

$$\approx \frac{h^5}{4}\left(f''(0+)\right)^2\left(\int_0^{\infty} u^2 K_0^*\right) + \frac{f(0+)}{n}\int_0^{\infty}\left(K_0^*\right)^2. \qquad (4.8)$$

If K is assumed to be zero outside the interval $[-1, 1]$, then the interval $[0, h)$ is the only boundary region and

$$\int AMSE\left(\widehat{f}_L(x); h\right) dx = \int_0^h AMSE\left(\widehat{f}_L(x); h\right) dx + \int_h^{\infty} AMSE\left(\widehat{f}_L(x); h\right) dx$$

$$\approx \frac{h^4}{4}\left(\mu_2(K)\right)^2 \int_h^{\infty}\left(f''(x)\right)^2 dx + \frac{1}{nh}R(K)\int_h^{\infty} f(x)dx \qquad (4.9)$$

$$\approx \frac{h^4}{4}\left(\mu_2(K)\right)^2 \int_0^{\infty}\left(f''(x)\right)^2 dx + \frac{R(K)}{nh}\int_0^{\infty} f(x)dx$$

$$= \frac{h^4}{4}\left(\mu_2(K)\right)^2 R\left(f''\right) + \frac{R(K)}{nh},$$

where the first approximation follows from (4.8) being of a smaller order than (4.9), and the second from the fact that the Lebesgue measure of $[0, h)$ tends to zero and both f and f'' are integrable. Hence, define the asymptotic mean integrated squared error of the local linear estimator $\widehat{f}_L$ as

$$AMISE\left(\widehat{f}_L; h\right) = \frac{h^4}{4}\left(\mu_2(K)\right)^2 R\left(f''\right) + \frac{R(K)}{nh},$$

and

$$MISE\left(\widehat{f}_L; h\right) = AMISE\left(\widehat{f}_L; h\right) + o\left(h^4 + \frac{1}{nh}\right).$$

The minimizer, with respect to h, of $AMISE\left(\widehat{f}_L; h\right)$ is

$$h_* = \left\{\frac{R(K)}{\left(\mu_2(K)\right)^2 R\left(f''\right)}\right\}^{1/5} n^{-1/5}. \qquad (4.10)$$

Therefore, it is appropriate to consider a plug-in type bandwidth selector for the local linear fit estimators. Then, parallel to the Sheather and Jones (1991) procedure, an estimator of $R(f'')$ and its asymptotically optimal bandwidth are necessary.

4.3 Integrated Squared Density Derivative Estimates

Consider estimating the following functional of the density f,

$$\theta_{\gamma,\nu} = \int f^{(\gamma)}(x) f^{(\nu)}(x) dx, \ \ \gamma, \nu \geq 0, \gamma + \nu \text{ even.}$$

It is shown in the last section that the special case of $\gamma = \nu = 2$ is very important to a plug-in bandwidth selector for the local linear estimators. It will become clear later that general cases, where γ and ν are not necessarily the same, are relevant to bandwidth choice for the estimators we propose. Hence this form of parameters and their estimators are studied.

Local polynomial fit estimators of the density derivatives were studied in Section 3.5. They retain the same rate of convergence everywhere including boundary points. Therefore it is proposed to plug such estimators into $\theta_{\gamma,\nu}$ with the goal of producing efficient estimators that overcome support constraints of the density. Precisely, define our estimators of $\theta_{\gamma,\nu}$ as

$$\begin{aligned}
\hat{\theta}_{\gamma,\nu} &= b \sum_{i=1}^{g} \hat{f}^{(\gamma)}(x_i) \hat{f}^{(\nu)}(x_i) \\
&= \frac{\gamma!\nu!}{n^2 b} \sum_{i=1}^{g} \left(\sum_{j=1}^{g} W_{\gamma}^{n} \left(\frac{x_j - x_i}{h} \right) c_j \right) \left(\sum_{k=1}^{g} W_{\nu}^{n} \left(\frac{x_k - x_i}{h} \right) c_k \right) \\
&= \frac{\gamma!\nu!}{n^2 b} \sum_{i=1}^{g} \left(\sum_{j=1}^{g} W_{\gamma}^{n} \left(\frac{x_j - x_i}{h} \right) c_j \right) \left(\sum_{k=1}^{g} W_{\nu}^{n} \left(\frac{x_k - x_i}{h} \right) c_k \right) \\
&= \frac{\gamma!\nu!}{n^2 b} \sum_{j=1}^{g} \sum_{k=1}^{g} \sum_{i=1}^{g} W_{\gamma}^{n} \left(\frac{x_j - x_i}{h} \right) W_{\nu}^{n} \left(\frac{x_k - x_i}{h} \right) c_j c_k. \qquad (4.11)
\end{aligned}$$

Then, one would ask whether this estimator works, and if so, how well. The following is devoted to answering this question by examining its mean squared error.

First we need some regularity condition. The density function f is said to be in the class F_{p+1}, where p is a nonnegative integer, if there exists a constant $M > 0$ so that, for any x and y belonging to the support of f,

$$\left|f^{(p+1)}(x) - f^{(p+1)}(y)\right| \leq M\,|x-y|. \tag{4.12}$$

For a more precise version of (4.12), see Bickel and Ritov (1988). Comparatively, we are assuming more smoothness. However, note that this smoothness condition on f is assumed only over its support here, while it is required over the real line in Bickel and Ritov (1988). We are now ready to state the following theorem.

Theorem 7 *Suppose $f \in F_{p+1}$ with $p+2 > \gamma+\nu$, $\widehat{f}^{(\gamma)}$and $\widehat{f}^{(\nu)}$ are the derivative estimators with a local polynomial fit of order p, and the weight function K is compactly supported with two derivatives. Since $\widehat{\theta}_{\gamma,\nu}$ is symmetric in γ and ν, we assume $\gamma \leq \nu$ for clean presentation. Then $\widehat{\theta}_{\gamma,\nu}$ has bias*

$$E(\widehat{\theta}_{\gamma,\nu} - \theta_{\gamma,\nu}) = \frac{\gamma!\nu!}{nh^{\gamma+\nu+1}}\int K_\gamma^* K_\nu^* + \frac{(1+\delta_{\gamma\nu})\nu!}{(p+1)!}h^{p-\nu+1}\theta_{\gamma,p+1}\left(\int u^{p+1}K_\nu^*\right)$$
$$+\,O(n^{-1}h^{-\gamma-\nu}) + O(h^{p-\gamma+1}), \tag{4.13}$$

and variance

$$Var\left(\widehat{\theta}_{\gamma,\nu}\right) = \frac{2(\gamma!\nu!)^2}{n^2h^{2(\gamma+\nu)+1}}\left(\int f^2\right)\left(\int \left(K_\gamma^* * K_\nu^*\right)^2\right)$$
$$+\frac{4}{n}\left(\int f\left(f^{(\gamma+\nu)}\right)^2 - \theta_{\gamma,\nu}^2\right) + o(n^{-2}h^{-2(\gamma+\nu)-1}) + o(n^{-1}), \tag{4.14}$$

provided that $n \to \infty, h \to 0, nh^{\gamma+\nu+1} \to \infty$, and $b = o(h)$.

Remark 4.1. Let $k = p-\nu+1$ and $\gamma = \nu$, then when f is supported and smooth on the entire real line, the expected value and variance of $\widehat{\theta}_{\gamma,\gamma}$ have the same rate of convergence as the estimator introduced in Jones and Sheather (1991) using a kernel of order k. Indeed, if K is the standard normal density, the equivalent kernel K_ν^* in (3.9) is exactly $\frac{1}{\nu!}K^{(\nu)}$ and hence $\widehat{\theta}_{\nu,\nu}$ is approximately the same as their estimators. Furthermore, our estimator is so effective that even the constant coefficients depending on f remain the same, whether there is a smooth boundary or not.

Corollary 1 *Suppose that* $\int K^*_\gamma K^*_\nu$ *and* $\int u^{p+1}K^*_\nu$ *have the same signs. Then, if* $\theta_{\gamma,p+1} > 0$, *the optimal bandwidth for* $\widehat{\theta}_{\gamma,\nu}$ *minimizing its asymptotic mean squared error is*

$$h_{AMSE} = \left(\frac{\gamma!(p+1)!(\gamma+\nu+1)\int K^*_\gamma K^*_\nu}{n(1+\delta_{\gamma\nu})(p-\nu+1)\theta_{\gamma,p+1}\int u^{p+1}K^*_\nu}\right)^{\frac{1}{p+\gamma+2}},$$

and the minimized asymptotic mean squared error equals

$$((p+\gamma+2)\nu!)^2\left(\frac{\gamma!\int K^*_\gamma K^*_\nu}{(p-\nu+1)n}\right)^{\frac{2(p-\nu+1)}{p+\gamma+2}}\left(\frac{(1+\delta_{\gamma\nu})\theta_{\gamma,p+1}\int u^{p+1}K^*_\nu}{(\gamma+\nu+1)(p+1)!}\right)^{\frac{2(\gamma+\nu+1)}{p+\gamma+2}}$$
$$+\frac{4}{n}\left(\int f\left(f^{(\gamma+\nu)}\right)^2-\theta^2_{\gamma,\nu}\right), \tag{4.15}$$

and if $\theta_{\gamma,p+1} < 0$,

$$h_{AMSE} = \left(\frac{\gamma!(p+1)!\int K^*_\gamma K^*_\nu}{n(1+\delta_{\gamma\nu})\,|\theta_{\gamma,p+1}|\int u^{p+1}K^*_\nu}\right)^{\frac{1}{p+\gamma+2}},$$

and the minimized asymptotic mean squared error equals

$$2(\gamma!\nu!)^2\left(\int f^2\right)\int\left(K^*_\gamma * K^*_\nu\right)^2\left(\frac{(1+\delta_{\gamma\nu})\,|\theta_{\gamma,p+1}|\int u^{p+1}K^*_\nu}{\gamma!(p+1)!\int K^*_\gamma K^*_\nu}\right)^{\frac{2(\gamma+\nu)+1}{p+\gamma+2}} n^{-\frac{2(p-\nu)+3}{p+\gamma+2}}$$
$$+\frac{4}{n}\left(\int f\left(f^{(\gamma+\nu)}\right)^2-\theta^2_{\gamma,\nu}\right). \tag{4.16}$$

Remark 4.2. If $\int K^*_\gamma K^*_\nu$ and $\int u^{p+1}K^*_\nu$ have different signs, then we need only switch the conditions on $\theta_{\gamma,p+1}$ for (4.15) and (4.16), and the results hold as above. Particularly, if $K(x) = (1-x^2)I_{[-1,+1]}(x)$, then choosing p such that $\frac{p+1-\gamma}{2}$ is an odd integer is equivalent to $\int K^*_\gamma K^*_\nu$ having the same sign as $\int u^{p+1}K^*_\nu$. Note that (4.16); i.e. the Jones-Sheather mean cancellation technique, has a better rate of convergence than (4.15). One interesting observation here is, in the special case of $\gamma = \nu$ and if f has no important feature at the boundaries, $\theta_{\gamma,p+1} = (-1)^{\frac{p+1-\gamma}{2}}\int\left(f^{\left(\frac{p+\gamma+1}{2}\right)}\right)^2$. So we can always have (4.16) which coincides with the results in Jones and Sheather (1991). However, if we take into account a more general class of densities, there is a distinction in terms of rate of convergence between "nice" densities and "not nice" densities.

Remark 4.3. From (4.15) and (4.16), $\widehat{\theta}_{\gamma,\nu}$ is $\sqrt{n}$–consistent if $p > \gamma + 2\nu$ or $p+1 > \gamma + 2\nu$ when $\theta_{\gamma,p+1} > 0$ or $\theta_{\gamma,p+1} < 0$, respectively. And the constant coefficient is $\sigma^2_{\gamma,\nu} \equiv 4\left(\int f\left(f^{(\gamma+\nu)}\right)^2 - \theta^2_{\gamma,\nu}\right)$. Note that $\left(\sigma^2_{\gamma,\gamma}\right)^{-1}$ equals the information bound for nonparametric estimation of $\theta_{\gamma,\gamma}$ given in Bickel and Ritov (1988). If f has support $(-\infty,+\infty)$, the requirement p (or $p+1$) $> 3\gamma$ for $\widehat{\theta}_{\gamma,\gamma}$ to achieve this optimal convergence is more than what is needed for the estimator there. However, if $f \in F_{p+1}$ and has nonsmooth boundaries, the estimator of Bickel and Ritov (1988) is not consistent but $\widehat{\theta}_{\gamma,\gamma}$ still possesses the same convergence. We feel two possible generalizations of the information bound results in Bickel and Ritov (1988) are implied from our results. One is to include the cases of $\gamma \neq \nu$, and the other is to extend the class of densities by considering those that are smooth except possible discontinuities on a finite set of points.

4.4 A Bandwidth Selector

Now we combine the results of the previous sections to propose a data-based bandwidth selector for the local linear estimators. It has been shown in Section 4.2 that the plug-in bandwidth selection idea is appropriate here. And a proper estimator of the integrated squared density second derivative is desired for such a procedure. We make use of the estimators of this quantity introduced in the last section, since it is shown there that the estimators which utilize local polynomial fit estimators have nice adaptive properties to boundary effects. The estimator is

$$\widehat{\theta}_{2,2}(a) = b\sum_{i=1}^{g}\left(\widehat{f}_a^{(2)}(x_i)\right)^2$$

$$= \frac{4}{n^2 b}\sum_{j=1}^{g}\sum_{k=1}^{g}\sum_{i=1}^{g} W_2^n\left(\frac{x_j - x_i}{a}\right) W_2^n\left(\frac{x_k - x_i}{a}\right) c_j c_k, \tag{4.17}$$

where $\widehat{f}_a^{(2)}$ is obtained from a local cubic fit of the bin counts with weight function K and bandwidth a , and

$$W_2^n(t) = e_3^T S_n^{-1}(1, at, a^2t^2, a^3t^3)^T K(t),$$

with $S_n = \left(\sum_{k=1}^{g} K\left(\frac{x_k - x}{a}\right)(x_k - x)^{i+j-2}\right)_{1\le i,j\le 4}$. Write

$$K_2^*(t) = e_3^T S^{-1}(1, at, a^2t^2, a^3t^3)^T K(t),$$

where $S = (\int t^{i+j-2}K(t)dt)_{1\le i,j\le 4}$. Consistency of $\hat{\theta}_{2,2}(a)$ is guaranteed by proper choice of the bandwidth a and the following conditions on the density f and the weight function K,

(C1) There exists a constant $M > 0$ so that, for any x and y belonging to the support of f

$$\left|f^{(4)}(x) - f^{(4)}(y)\right| \le M\,|x-y|\,.$$

(C2) The weight function K has compact support.

(C3) The first two derivatives of K exist.

Then, according to Corollary 1, the asymptotic optimal bandwidth for $\hat{\theta}_{2,2}$ is

$$a_* = \left(\frac{24\chi R(K_2^*)}{n\int f^{(2)}f^{(4)}\int u^4K_2^*}\right)^{1/7}, \tag{4.18}$$

where

$$\chi = \begin{cases} -1 & ,\text{if } \int f^{(2)}f^{(4)} < 0. \\ \frac{5}{2} & ,\text{if } \int f^{(2)}f^{(4)} > 0. \end{cases}$$

From (4.10) and (4.18),

$$a_* = C(K)D(f)h_*^{5/7}, \tag{4.19}$$

where

$$C(K) = \left(\frac{24R(K_2^*)\left(\int u^2K\right)^2}{R(K)\int u^4K_2^*}\right)^{1/7}, D(f) = \left(\frac{\chi R(f^{(2)})}{\int f^{(2)}f^{(4)}}\right)^{1/7}.$$

Then we apply this relation in (4.10) and find the solution in H of the following equation

$$H = \left\{\frac{R(K)}{n\left(\int u^2K\right)^2\hat{\theta}_{2,2}\left(a(H)\right)}\right\}^{1/5}, \tag{4.20}$$

where

$$a(h) = C(K)D(f)h^{5/7}. \tag{4.21}$$

Here $D(f)$ involves $\int f^{(2)}f^{(4)}$ and $\int \left(f^{(2)}\right)^2$ which can be estimated by some reference estimator through a scale parameter model of f. Let g_1 be a fixed density function, e.g. the standard normal, that has been normalized so that some measure of scale such as the standard deviation is equal to one. Then $D(g_\lambda) = \lambda^{2/7}D(g_1)$, where $g_\lambda(x) = \frac{1}{\lambda}g_1(\frac{x}{\lambda})$. Then, from (4.21), set

$$a_\lambda(h) = C(K)D(g_1)\lambda^{2/7}h^{5/7}.$$

The plug-in bandwidth $\widehat{h}$ is defined as the solution of the analogous equation of (4.1) with $a(H)$ replaced by $a_{\widehat{\lambda}}(H)$, where $\widehat{\lambda}$ is a $\sqrt{n}$–consistent estimator of λ.

The following theorem states that this bandwidth selector works regardless whether there are boundaries in the support of the density or not. In addition to the conditions (C1) - (C3), we need two more assumptions on K:

(C4) The weight function K vanishes at the endpoints of its support.

(C5) The function K is symmetric about the origin.

Theorem 8 *Under conditions (C1) - (C5), as $n \to \infty$,*

$$\frac{\widehat{h}}{h_*} = 1 + O_p\left(n^{-\alpha}\right),$$

where

$$\alpha = \begin{cases} 5/14 & , \text{ if } \int f^{(2)}f^{(4)} < 0 \\ 2/7 & , \text{ if } \int f^{(2)}f^{(4)} > 0. \end{cases}$$

Furthermore,

$$n^\alpha\left(\frac{\widehat{h}}{h_*} - 1\right) \xrightarrow{D} N\left(\mu_{PI}, \sigma_{PI}^2\right), \tag{4.22}$$

where

$$\mu_{PI} = \frac{-7}{150}\sigma_K^{-8/7}R(K)^{2/7}R(f^{(2)})^{-9/7}C(K)^2D(g_\lambda)^2\int f^{(2)}f^{(4)}\int u^4K_2^*I\left(\int f^{(2)}f^{(4)} > 0\right),$$

and

$$\sigma_{PI}^2 = \frac{32}{25}n^{2\alpha-5/7}\sigma_K^{118/35}R(K_2^* * K_2^*)R(K)^{-38/35}R(f^{(2)})^{-5/7}R(f)C(K)^{-9}D(g_\lambda)^{-9},$$

with $\sigma_K^2 = \int u^2K$.

Remark 4.4. The convergence rate of bandwidth $\widehat{h}$ depends on the functional $\int f^{(2)} f^{(4)}$, see (4.22). In non-boundary cases, the faster rate $n^{-5/14}$, same as Sheather and Jones (1991), always holds since $\int f^{(2)} f^{(4)} = -\int \left(f^{(3)}\right)^2 < 0$ under such circumstance. When boundaries present and the density is nonsmooth there, the Sheather and Jones (1991) bandwidth will not be consistent, see Van Es and Hoogstrate (1993b), but $\widehat{h}$ still is, only with possible slightly slower rate of convergence. And if the density is well behaved at the endpoints in the sense that it has zero second and third derivatives there then, taking an integration by parts,

$$\int_0^\infty f^{(2)}(x) f^{(4)}(x) dx = -f^{(2)}(0) f^{(3)}(0) - \int_0^\infty \left(f^{(3)}(x)\right)^2 dx = -\int \left(f^{(3)}\right)^2 < 0,$$

and the rate is $n^{-4/15}$.

Remark 4.5. Since the asymptotically optimal bandwidth (4.10) is exactly the same as that for the conventional kernel density estimator, one might ask why not simply use the Sheather and Jones (1991) bandwidth selection procedure? This won't work because the estimators in Jones and Sheather (1991) are influenced by boundary effects and the bandwidth selector will not be even consistent as a consequence. This is observed in the study of Van Es and Hoogstrate (1993b), who show that the bandwidth selector tends to zero at some rate other than $n^{-1/5}$ in nonsmooth cases.

Remark 4.6. The motivation for this bandwidth selector is entirely analogous to that for the plug-in rules for kernel density estimators; e.g. those discussed in Park and Marron (1990) and Sheather and Jones (1991). Yet, the contributions of this work include, first, showing that the same approach is applicable to the local linear estimators which have the appealing property of automatic boundary adaptivity, and secondly, providing a bandwidth selector that is as effective in the boundary case as in the non-boundary case.

Remark 4.7. In the regression setting, Ruppert, Sheather and Wand (1993) propose a plug-in bandwidth selector for local linear regression called $\widehat{h}_{STE}$ which makes use of parallel estimators from local polynomial fits. Performance of $\widehat{h}_{STE}$ is confirmed there by simulation studies. It is conjectured that $\widehat{h}_{STE}$ and the bandwidth selector introduced here have analogous asymptotic behaviors.

Remark 4.8. This research is about the situation of known boundary locations in density estimation problem. If the boundary locations are unknown, say $f(x) > 0$ if $x \geq a$ and $f(x) = 0$ otherwise. Then one can apply the local linear (polynomial) method to estimate the density (derivatives) on $[X_{(1)}, \infty)$, where $X_{(1)}$ is the sample minimum, and set the estimator to zero below $X_{(1)}$. For $f(a) > 0$, $X_{(1)}$ converges almost surely to a at the rate n^{-1}. This rate of convergence is much faster than the bandwidths for estimating density and its derivatives, the asymptotic mean square error results in Chapters 2 and 3 will still hold. Furthermore, the n^{-1} rate is faster than the error rates in estimating the density and its derivatives. Hence the theorems in the present chapter will be also valid.

4.5 Simulation

A simulation study on the plug-in bandwidth selector introduced in the previous section is conducted over the densities given in Section 2.3.2. Two sample sizes, 100 and 1000, are chosen and there are 500 replicates for each simulation. Before going into the outcomes, some technical notes are provided for estimating the integrated squared density second derivatives. They appear to be very important, and sometimes actually dominate the performance of the bandwidth selector.

4.5.1 Binning Range

Recall that our estimator of the integrated squared density second derivative is based on local polynomial fitting of the density function to some bin counts. In practice, the bin centers span a finite interval. Under the situation that the underlying density has a known compact support, there is no question about what range the data should be binned to. Otherwise, for simplicity of argument the support is assumed to be the entire real line. In that case, it is suggested to make the binning range so wide that all most all observations will fall in the interior.

The reason for this suggestion is to avoid the boundary adjustments imposed by local polynomial fits at places where the data are sparse. It is very noticeable that the local polynomial fit estimators automatically adapt to estimation at boundaries. But, typically the mean squared error of the estimators at boundaries regions are much larger, although only by a constant factor but not the rate of convergence, than that in the interior. To better understand this idea, we consider estimating the second derivative of the standard normal density at -2.5 from four independent samples each with 100 i.i.d. observations. In Figure 4.1, the four data sets are binned to the range [-3,3] which we chose because the data points will fall outside this interval only very rarely. The data sets are represented by, or transformed to, the bin counts which are shown as the squares in the plots. The dashed line shows 1/3 of the equivalent kernel, with a Gaussian weight function, for estimating the density second derivative at -2.5 which is about one bandwidth, 0.454, away from -3, the left end of the binning interval. And the dashed line in Figure 4.2 shows that for estimating at a point, -2.5, which is about two bandwidths away from the end, -4. Note that Figures 4.1 and 4.2 are using the same sets of data. Obviously, the estimator with the equivalent kernel in Figure 4.1 has more variability since a little shift in the horizontal direction causes a huge change in the function's value. This, combined with instability of the bin counts around -2.5, which is in the tail area of the standard normal density, makes the estimates very noisy.

In all the four cases, there are very few observations that appear within the area where the weight function for estimating at -2.5 is significantly positive, and the bin counts vary a lot from one data set to another. Hence, according to the equivalent kernel, the estimates which are located by the circles are different from each other by a large amount. Also these estimates are far away from the true value indicated by the plus sign. But, the situation changes if the binning range is extended to [-4,4], see Figure 4.2. Now, even though the bin counts are still noisy, the estimates are more stable since the equivalent kernel is stabilized.

In conclusion, with the local polynomial fits, boundary corrections are sensitive to the range where the techniques are applied. This issue is important for the rate of convergence at actual boundary regions. A boundary estimator is not preferred when estimating at tail regions where (i) every point is still in the interior and (ii) data are sparse, but this problem can be easily avoided by extending the binning range.

4.5.2 Boundary Bandwidths

Another problem that occurs while estimating integrated squared density derivatives is that optimal bandwidths; e.g. those minimize the asymptotic mean squared error, tend to be too small in boundary regions. Then the estimated derivatives are under smoothed there and the functionals are seriously over estimated. In this case the equation solving procedure (4.20) will have problems; $\widehat{h}$ becomes too small since $\widehat{\theta}_{2,2}$ is too large.

To give an idea of how serious this situation can be, five estimates of f'' with f being density #1 are plotted in Figure 4.3. For both sample sizes 100 and 1000, the estimates are all extremely large in absolute value compared to the true curve in a boundary region $[0, 2h_{amse}]$, where h_{amse} is the asymptotically optimal bandwidth for estimating $\widehat{\theta}_{2,2}$ given in corollary 1. It is seen that the 1000 observations case is slightly better indicating the asymptotics are taking effect, but still the estimates of $\widehat{\theta}_{2,2}$ are too large for our purpose of plug-in rules.

One way to solve this problem is to use a larger bandwidth for the boundary area. But, what is a reasonable one? The approach taken here is to find the asymptotic optimal bandwidth for estimating the quantity $\int_0^{2h_{amse}} (f''(x))^2\, dx$ by

$$\frac{4}{n^2 b} \sum_{j=1}^{g} \sum_{k=1}^{g} \sum_{i=1}^{d} W_2^n\left(\frac{x_j - x_i}{a}\right) W_2^n\left(\frac{x_k - x_i}{a}\right) c_j c_k,$$

where

$$d = \left[\frac{h_{amse}}{b}\right],$$

the integer part of h_{amse}/b. By entirely similar arguments to those in the proof of Theorem 7, the mean squared error of the above estimator is

$$\frac{a^2}{6} \int_0^{2h_{amse}} \int_{-x/a}^{\infty} s^4 K_{2,x/a}^*(s) f^{(2)}(x) f^{(4)}(x)\, ds\, dx$$

$$+\frac{4}{na^5}\int_0^{2h_{amse}}\int_{-x/a}^{\infty}\left(K^*_{2,x/a}(s)\right)^2 f(x)dsdx + o\left(a^2+\frac{1}{na^5}\right)$$

The dominating terms are very complicated functionals of the weight function K and the density f. However, since both h_{amse} and a are shrinking to zero, it is approximately

$$\frac{a^3 f^{(2)}(0)f^{(4)}(0)}{6}\int_0^2\int_{-c}^{\infty} s^4 K^*_{2,c}(s)dsdc + \frac{4f(0)}{na^4}\int_0^2\int_{-c}^{\infty}\left(K^*_{2,c}(s)\right)^2 dsdc.$$

Therefore the asymptotic optimal bandwidth is

$$\left(\frac{24\tau f(0)\int_0^2\int_{-c}^{\infty}\left(K^*_{2,c}(s)\right)^2 dsdc}{f^{(2)}(0)f^{(4)}(0)\int_0^2\int_{-c}^{\infty} s^4 K^*_{2,c}(s)ds}\right)^{1/7} n^{-1/7}, \tag{4.23}$$

where

$$\tau = \begin{cases} 1 & \text{, if } f^{(2)}(0)f^{(4)}(0) < 0. \\ 4/3 & \text{, if } f^{(2)}(0)f^{(4)}(0) > 0. \end{cases}$$

Then, estimating $R(f^{(2)})$ in the plug-in procedure is separated into two parts; bandwidth using relation (4.19) is used for the interior portion and bandwidth from its analog which combines (4.10) and (4.23) is implemented at the boundary region. Here, the reference value of $f(0)/(f^{(2)}(0)f^{(4)}(0))$ is obtained from an ordinary least squares fourth degree polynomial fit of the density to the bin counts.

4.5.3 Simulation Results

Since bandwidth works multiplicatively, the 500 simulated bandwidths are studied by the logarithm of their ratio to the optimal bandwidth; i.e. $\log_3\left(\frac{\hat{h}}{h_{mise}}\right)$. Ordinary kernel density estimates for the population density of this quantity are presented in Figure 4.4. The Sheather-Jones bandwidths are given to provide evidence of boundary effects and as a baseline for the performance of the proposed bandwidths in non-boundary case. Solutions to the equation (4.20) were searched within the range $[\frac{1}{3}h_{mise}, 9h_{mise}]$, if the right hand side is always greater than the left hand side then $\frac{1}{3}h_{mise}$ is assigned.

The proposed bandwidth is quite satisfactory, see Figure 4.4. It becomes more accurate when the sample size change from 100 to 1000, indicating its consistency.

And it is close to the optimal bandwidth most of the time. Density #5 is the standard normal which has no boundary at all. The two studied bandwidths behave close to each other. This is not surprising since it is known that $\widehat{\theta}_{2,2}(a)$ is very similar to the estimators of $R(f'')$ in Jones and Sheather (1991) as pointed out in remark 4.1. But, it shows that our bandwidth selector works as well as the Sheather-Jones rule in non-boundary cases. Hence, beside theoretical justification for $\widehat{h}$, we have also support it by simulation as a efficient bandwidth which is at the same time robust to boundary problems.

Curiously, the Sheather-Jones bandwidth is very close to the optimal bandwidth for density #4 but the proposed one is "far away from optimal", see Figure 4.4. Local linear estimates using the sample medians of the two bandwidths are given in Figure 4.5. The same size is 100 and 10 independent estimates are drawn. It is visually clear that the proposed bandwidth gives better estimates of the density; they indicate the concavity well and the variability is smaller. On the other hand, estimates using the Sheather-Jones bandwidth, which is close to h_{mise}, are too noisy and apparently under smoothed. This suggests the proposed bandwidth is more reliable. Also note "L_2 optimal" is pretty rotten, as discussed in Marron and Tsybakov (1993).

4.6 Real Data Analysis

There are two real data analyses used to test the proposed bandwidth in practical application. The first one is patients' compliance in a medical study. See Lipids Research Clinic Program (1984). The two data sets, "Col 5" and "Col 6", each consists i.i.d observations of compliance; i.e. the portion of pills actually taken. One interest is to know how the compliance is distributed. The ordinary kernel estimate, see Figure 4.6, uses a bandwidth chosen by visual judgment and shows suspicious modes near the boundaries. As indicated above it is improper for boundary estimation and so we concentrate our attention on the local linear estimates.

As a first attempt, the whole data set was used to decide the plug-in bandwidth and the resulting one was very small. Plots of the local linear estimates using these bandwidths reveal that there are point mass at the two endpoints. As a matter of fact, $\hat{\theta}_{2,2}$ feels the nonexistence of the density function there and thus become extremely large and hence $\hat{h}$ is very small. This however yields the recognition of point masses which smooth density estimators could not indicate. The circles in the plots are the percentages of zeros and ones in the data sets. Next, the 0's and 1's were deleted from the data sets and the corresponding automatic bandwidths were used to construct the estimates. It is seen that the interior part is under smoothed. But, as a bandwidth for global measure purpose, the bandwidth gives good insights of the density structure. It can be viewed as a first attempt in the data analysis and indeed is a good one. One might next implement a variable bandwidth technique, as suggested by these estimates, in a later stage but this is beyond the scope of our discussion. Interestingly, there isn't a mode around zero for "Col 5" and it could have been easily mistaken without this bandwidth selection procedure.

The other data, provided by Dr. Les Roberts of the Centers for Disease Control, is from the study of choliform counts in water taken from some public water source in Malawi. The purpose of the study is to determine whether there is any difference in the two populations, "Unimproved" and "Improved". Again, the two data sets were separately used to decide the automatic bandwidth. Unfortunately, the procedure again picked up the smallest bandwidth allowed in our search. Here the largest (smallest) bandwidth given in the search was decided by visual examination that it over (under) smooths the data. Bandwidth 1, respectively 3, is the smallest bandwidth that we can handle for the unimproved data , improved data respectively, otherwise numeric underflow occurs when calculating the local linear estimates. Figure 4.7 shows the estimates. The lesson is the same as the compliance data: there are point masses at 300 and 600. The circles (squares) are the portions of 300's and 600's in the unimproved (improved) data divided by four. The reason for these point masses is that the counting sometimes stops at 300 and otherwise it stops at 600. Again, after removing these observations the bandwidth procedure does well, see Figure 4.7. In conclusion, the proposed bandwidth is very helpful in the beginning stage of density analysis. It provides good insight into the structure of the population and suggests directions in a later stage in the study.

4.7 Proofs

(I) Proof of Theorem 7:

The first two moments of $\widehat{\theta}_{\gamma,\nu}$ depend on the expected values in the following lemma. And their approximations stated therein show some links between the moments and the density f.

Lemma 3 *If $f \in F_{p+1}$ then as the sample size n grows*

(i) $E(c_i) = nbf(x_i)(1+o(1)), i = 1, ..., g.$

(ii) $E(c_i^2) = (nbf(x_i) + n(n-1)b^2 f(x_i)^2)(1+o(1)), i = 1, ..., g.$

(iii) $E(c_i^4) = \left(nbf(x_i) + \frac{7n!b^2}{(n-2)!} f(x_i)^2 + \frac{6n!b^3}{(n-3)!} f(x_i)^3 + \frac{n!b^4}{(n-4)!} f(x_i)^4\right)(1+o(1)),$

$i = 1, ..., g.$

(iv) $E(c_i c_j) = (n(n-1)b^2 f(x_i) f(x_j))(1+o(1)),$ *for* $i \neq j, 1 \leq i, j \leq g.$

(v) $E\left(c_i^2 c_j^2\right) = \left(\frac{n!b^2}{(n-2)!} + \frac{n!b^3}{(n-3)!} f(x_i) + \frac{n!b^3}{(n-3)!} f(x_j) + \frac{n!b^4}{(n-4)!} f(x_j) f(x_j)\right)$

$\times f(x_i) f(x_j)(1+o(1)),$ *for* $i \neq j, 1 \leq i, j \leq g.$

(vi) $E(c_i^3 c_j) = \left(\frac{n!b^2}{(n-2)!} + \frac{3n!b^3}{(n-3)!} f(x_i) + \frac{n!b^4}{(n-4)!} f(x_j)^2\right) f(x_i) f(x_j)(1+o(1)),$

for $i \neq j, 1 \leq i, j \leq g.$

(vii) $E\left(c_i^2 c_j c_k\right) = \left(\frac{n!b^3}{(n-3)!} + \frac{n!b^4}{(n-4)!} f(x_i)\right) f(x_i) f(x_j) f(x_k) \left(1 + o(1)\right)$,

for $i \neq j \neq k \neq i, 1 \leq i, j, k \leq g$.

(viii) $E\left(c_i c_j c_k c_l\right) = \frac{n!b^4}{(n-4)!} f(x_i) f(x_j) f(x_k) f(x_l) \left(1 + o(1)\right)$, *for* i, j, k, l

all different and between 1 *and* g.

It will become clear in the proof of Theorem 7 that the mean and variance of $\widehat{\theta}_{\gamma,\nu}$ are asymptotically linear combinations of those complicated functionals in the next lemma. The results further simplify the asymptotic expressions.

Lemma 4 *Suppose the function* $\omega(\cdot,\cdot) : R \times R \to R$ *is defined as*

$$\omega(t, u) = \int_D K_\gamma^* \left(\frac{t - x}{h}\right) K_\nu^* \left(\frac{u - x}{h}\right) dx,$$

where D *is the support of* f. *Then the following asymptotics hold under the assumptions* $b = o(h)$, $f \in F_{p+1}$ *and* K *has finite support.*

(i) $\int f(t)\omega(t,t)dt = h \int K_\gamma^* K_\nu^* + O(h^2)$.

(ii) $\int f(t)\omega(t,t)^2 dt = h^2 \left(\int K_\gamma^* K_\nu^*\right)^2 + o(h^2)$.

(iii) $\int \int f(t) f(u) \omega(t,u) dt du = \frac{h^{\gamma+\nu+2}}{\gamma!\nu!} \theta_{\gamma,\nu} + \frac{(1+\delta_{\gamma\nu}) h^{p+(\gamma\wedge\nu)+3}}{(\gamma\wedge\nu)!(p+1)!} \left(\int u^{p+1} K_{\gamma\vee\nu}^*\right) \theta_{\gamma\wedge\nu,p+1}$

$+ O(h^{p+(\gamma\vee\nu)+3})$.

(iv) $\int \int f(t) f(u) \omega(t,u)^2 dt du = h^3 \left(\int f^2\right) \left(\int \left(K_\gamma^* * K_\nu^*\right)^2\right) + o(h^3)$.

(v) $\int \int f(t) f(u) \omega(t,t) \omega(t,u) dt du = \frac{h^{\gamma+\nu+3}}{\gamma!\nu!} \theta_{0,\gamma+\nu} \left(\int K_\gamma^* K_\nu^*\right) + o(h^{\gamma+\nu+3})$.

(vi) $\int \int \int f(t) f(u) f(v) \omega(t,v) \omega(t,u) dt du dv = \frac{h^{2\gamma+2\nu+4}}{(\gamma!\nu!)^2} \left(\int f \left(f^{(\gamma+\nu)}\right)^2\right)$

$+ o(h^{2\gamma+2\nu+4})$

Proof of Lemma 3:

First define $I_i(X_j) = I_{[-\frac{1}{2},\frac{1}{2})}\left(\frac{X_j - x_i}{b}\right)$ for any $1 \le j \le n$ and $1 \le i \le g$. Then

$$|E(I_i(X_j)) - bf(x_i)| = \left|\int f(s) I_{[-\frac{1}{2},\frac{1}{2})}\left(\frac{s - x_i}{b}\right) ds - bf(x_i)\right|$$

$$= \left|\int_{x_i - b/2}^{x_i + b/2} f(s)ds - bf(x_i)\right| = \left|\int_{x_i - b/2}^{x_i + b/2} [f(s) - f(x_i)]\, ds\right|$$

$$= \left|\int_{x_i - b/2}^{x_i + b/2} \left[f^{(1)}(x_i)(x_i - s) + \frac{1}{2}\int_s^{x_i} (s-t) f^{(2)}(t)\right] dtds\right|$$

$$= \frac{1}{2}\left|\int_{x_i - b/2}^{x_i + b/2} \int_s^{x_i} (s-t) f^{(2)}(t) dtds\right| \le \frac{b^2}{8}\int_{x_i - b/2}^{x_i + b/2} \left|f^{(2)}(t)\right| dt.$$

Therefore,

$$E(I_i(X_j)) = bf(x_i) + o(b), 1 \le i \le g, 1 \le j \le n \qquad (4.24)$$

Also, since $I_i(\cdot)$ is an indicator function, for any positive integer k,

$$E(I_i(X_j))^k = E(I_i(X_j)) = bf(x_i) + o(b) \qquad (4.25)$$

The approximations (4.24) and (4.25) are repeatedly used in the proofs of (i) - (vii).

(i) Since $X_1, ..., X_n$ are i.i.d random variables,

$$E(c_i) = E\left(\sum_{j=1}^{n} I_i(X_j)\right) = nE(I_i(X_1)) = nbf(x_i)(1 + o(1)).$$

(ii)

$$E\left(c_i^2\right) = E\left(\sum_{j=1}^{n} I_i(X_j)\right)^2$$

$$= \sum_{j=1}^{n} E(I_i(X_j))^2 + \sum\sum_{j \ne k} E(I_i(X_j) I_i(X_k))$$

$$= nE(I_i(X_1)) + n(n-1)E(I_i(X_1) I_i(X_2))$$

$$= nE(I_i(X_1)) + n(n-1)[E(I_i(X_1))]^2.$$

The last two equalities follow from $X_1, ..., X_n$ being i.i.d. Applying (4.24) to the above expression yields

$$E\left(c_i^2\right) = \left(nbf(x_i) + n(n-1)b^2 f(x_i)^2\right)(1 + o(1)).$$

(iii)

$$E\left(c_i^4\right) = E\left(\sum_{j=1}^{n} I_i(X_j)\right)^4$$

$$= \sum_{k=1}^{n}\sum_{l=1}^{n}\sum_{m=1}^{n}\sum_{n=1}^{n} E(I_i(X_k) I_i(X_l) I_i(X_m) I_i(X_n))$$

$$= \sum_{k=1}^{n} E(I_i(X_k))^4 + 3 \times \sum\sum_{k \neq l} E(I_i(X_k) I_i(X_l))^2$$

$$+ 4 \times \sum\sum_{k \neq l} E\left(I_i(X_k)^3 I_i(X_l)\right) + 6 \times \sum\sum\sum_{k \neq l \neq m \neq k} E\left(I_i(X_k)^2 I_i(X_l) I_i(X_m)\right)$$

$$+ \sum\sum\sum\sum_{k,l,m,j \text{ all different}} E(I_i(X_k) I_i(X_l) I_i(X_m) I_i(X_j))$$

$$= nE(I_i(X_1)) + \frac{7n!}{(n-2)!}(EI_i(X_1))^2 + \frac{6n!}{(n-3)!}(EI_i(X_1))^3$$

$$+ \frac{n!}{(n-4)!}(EI_i(X_1))^4.$$

Combining the above and (4.24) proves the result.

(iv) For $1 \leq i \neq j \leq g$,

$$E(c_i c_j) = E\left(\sum_{k=1}^{n}\sum_{l=1}^{n} I_i(X_k) I_j(X_l)\right)$$

$$= \sum_{k=1}^{n} E(I_i(X_k) I_j(X_k)) + \sum\sum_{k \neq l} E(I_i(X_k) I_j(X_l))$$

Notice that $I_i(X_k) I_j(X_k) \equiv 0$ because X_k can not fall into two disjoint intervals, therefore

$$E(c_i c_j) = n(n-1) E(I_i(X_1)) E(I_j(X_2))$$

$$= \frac{n!}{(n-2)!}\left(bf(x_i)+o(b)\right)\left(bf(x_j)+o(b)\right) = \frac{n!}{(n-2)!}\left(b^2 f(x_i) f(x_j)\right)(1+o(1)).$$

(v) For $1 \le i \ne j \le g$,

$$E\left(c_i^2 c_j^2\right) = E\left(\sum_{k=1}^{n}\sum_{l=1}^{n}\sum_{r=1}^{n}\sum_{s=1}^{n} I_i(X_k) I_i(X_l) I_j(X_r) I_j(X_s)\right)$$

$$= \sum_{k=1}^{n} E\left(I_i(X_k) I_j(X_k)\right)^2 + 4\times \sum\sum_{k\ne l} E\left(I_i(X_k)^2 I_j(X_k) I_j(X_l)\right)$$

$$+2\times \sum\sum_{k\ne l} E\left(I_i(X_k) I_i(X_l) I_j(X_k) I_j(X_l)\right) + \sum\sum_{k\ne l} E\left(I_i(X_k)^2 I_j(X_l)^2\right)$$

$$+4\times \sum\sum\sum_{k\ne l\ne r\ne k} E\left(I_i(X_k) I_i(X_l) I_j(X_k) I_j(X_r)\right)$$

$$+ \sum\sum\sum_{k\ne l\ne r\ne k} E\left(I_i(X_k)^2 I_j(X_l) I_j(X_r) + I_i(X_k) I_i(X_l) I_j(X_r)^2\right)$$

$$+ \sum\sum\sum\sum_{k,l,r,s \text{ all different}} E\left(I_i(X_k) I_i(X_l) I_j(X_r) I_j(X_s)\right)$$

$$= 0+0+0+n(n-1)E\left(I_i(X_1) I_j(X_2)\right)+0+$$

$$+\frac{n!}{(n-3)!}E\left\{I_i(X_1) I_j(X_2) I_j(X_3) + I_i(X_1) I_j(X_2) I_i(X_3)\right\}$$

$$+\frac{n!}{(n-4)!}E\left\{I_i(X_1) I_i(X_2) I_j(X_3) I_j(X_4)\right\}.$$

Again use (4.24) and the fact that $X_1, ..., X_n$ are i.i.d. to obtain the result.

(vi) For $1 \le i \ne j \le g$,

$$E\left(c_i^3 c_j\right) = E\left(\sum_{k=1}^{n}\sum_{l=1}^{n}\sum_{r=1}^{n}\sum_{s=1}^{n} I_i(X_k) I_i(X_l) I_i(X_r) I_j(X_s)\right)$$

$$= \sum_{k=1}^{n} E\left(I_i(X_k)^3 I_j(X_k)\right) + 3\times \sum\sum_{k\ne l} E\left(I_i(X_k)^2 I_i(X_l) I_j(X_k)\right)$$

$$+ \sum\sum_{k\ne l} E\left(I_i(X_k)^3 I_j(X_l)\right) + 3\times \sum\sum_{k\ne l} E\left(I_i(X_k)^2 I_i(X_l) I_j(X_l)\right)$$

$$+3\times \sum\sum\sum_{k\ne l\ne r\ne k} E\left\{I_i(X_k)^2 I_i(X_l) I_j(X_r) + I_i(X_k) I_i(X_l) I_i(X_r) I_j(X_k)\right\}$$

$$+ \sum\sum\sum\sum_{k,l,r,s \text{ all different}} E\left(I_i(X_k) I_i(X_l) I_i(X_r) I_j(X_s)\right)$$

$$= 0 + 0 + \frac{n!b^2}{(n-2)!} f(x_i) f(x_j) (1 + o(1)) + 0 + \frac{3n!b^3}{(n-3)!} f(x_i)^2 f(x_j) (1 + o(1))$$

$$+ 0 + \frac{n!b^4}{(n-4)!} f(x_i)^3 f(x_j) (1 + o(1)).$$

(vii) For $1 \leq i \neq j \neq k \leq g$,

$$E\left(c_i^2 c_j c_k\right) = E\left(\sum_{l=1}^{n}\sum_{m=1}^{n}\sum_{r=1}^{n}\sum_{s=1}^{n} I_i(X_l) I_i(X_m) I_j(X_r) I_k(X_s)\right)$$

$$= \sum_{l=1}^{n} E\left(I_i(X_l)^2 I_j(X_l) I_k(X_l)\right) + 2 \times \sum\sum_{l \neq m} E\left(I_i(X_l) I_i(X_m) I_j(X_l) I_k(X_l)\right)$$

$$+ 2 \times \sum\sum_{l \neq m} E\left(I_i(X_l)^2 I_j(X_l) I_k(X_m)\right) + \sum\sum_{l \neq m} E\left(I_i(X_l)^2 I_j(X_m) I_k(X_m)\right)$$

$$+ \sum\sum_{l \neq m} E\left\{I_i(X_l) I_i(X_m) I_j(X_l) I_k(X_m) + I_i(X_l) I_i(X_m) I_j(X_m) I_k(X_l)\right\}$$

$$+ \sum\sum\sum_{l \neq m \neq r \neq l} E\left\{I_i(X_l)^2 I_j(X_m) I_k(X_r) + I_i(X_m) I_i(X_r) I_j(X_l) I_k(X_l)\right\}$$

$$+ \sum\sum\sum_{l \neq m \neq r \neq l} E\left\{I_i(X_l) I_i(X_m) I_j(X_l) I_k(X_r) + I_i(X_l) I_i(X_m) I_j(X_r) I_k(X_l)\right\}$$

$$+ \sum\sum\sum_{l \neq m \neq r \neq l} E\left\{I_i(X_m) I_i(X_l) I_j(X_l) I_k(X_r) + I_i(X_m) I_i(X_l) I_j(X_r) I_k(X_l)\right\}$$

$$+ \sum\sum\sum\sum_{l,m,r,s \text{ all different}} E\left(I_i(X_l) I_i(X_m) I_j(X_r) I_k(X_s)\right)$$

$$= 0 + 0 + 0 + 0 + 0 + 0 + \frac{n!b^3}{(n-3)!} f(x_i) f(x_j) f(x_k) (1 + o(1)) +$$

$$+ 0 + 0 + 0 + 0 + 0 + \frac{n!b^4}{(n-4)!} f(x_i)^2 f(x_j) f(x_k) (1 + o(1)).$$

(viii) For $1 \leq i \neq j \neq k \neq l \leq g$,

$$E\left(c_i c_j c_k c_l\right) = E\left(\sum_{p=1}^{n}\sum_{q=1}^{n}\sum_{r=1}^{n}\sum_{s=1}^{n} I_i(X_p) I_j(X_q) I_k(X_r) I_l(X_s)\right)$$

$$= \sum\sum\sum\sum_{p,q,r,s \text{ not all different}} E\left(I_i(X_p) I_j(X_q) I_k(X_r) I_l(X_s)\right)$$

$$+ \sum\sum\sum\sum_{p,q,r,s \text{ all different}} E\left(I_i(X_p)\, I_j(X_q)\, I_k(X_r)\, I_l(X_s)\right)$$

$$= 0 + \frac{n!b^4}{(n-4)!} f(x_i)f(x_j)f(x_k)f(x_l)\,(1+o(1))\,.$$

Proof of Lemma 4:

Without any loss of generality, assume D, the support of f, to be $[0,\infty)$ in demonstrating the results hold in boundary case. Non-boundary situation is proved in the same way except arguments taking care of boundary regions are removed. Since K has finite support, we assume it, and hence $K^*_{m,c}$, has support $[-1,1]$ without loss of generality. Therefore $K^*_{m,c} \equiv K^*_m$ for every $c \geq 1$.

(i)

$$\int_0^\infty f(t)\omega(t,t)dt = \int_0^\infty \int_0^\infty K^*_{\gamma,\frac{x}{h}}\left(\frac{t-x}{h}\right) K^*_{\nu,\frac{x}{h}}\left(\frac{t-x}{h}\right) f(t)dxdt$$

Since the integrand is non negative, by the Fubini Theorem, the integral equals

$$\int_0^\infty \int_0^\infty K^*_{\gamma,\frac{x}{h}}\left(\frac{t-x}{h}\right) K^*_{\nu,\frac{x}{h}}\left(\frac{t-x}{h}\right) f(t)dtdx$$

$$= \int_0^\infty \int_{-\frac{x}{h}}^1 K^*_{\gamma,\frac{x}{h}}(s)\, K^*_{\nu,\frac{x}{h}}(s)\, f(x+sh)hdsdx$$

$$= h\int_0^\infty f(x) \int_{-\frac{x}{h}}^1 K^*_{\gamma,\frac{x}{h}}(s)\, K^*_{\nu,\frac{x}{h}}(s)\, dsdx(1+O(h)).$$

The last equality holds from $f \in F_{p+1}$. Break the above integral into two parts

$$\int_0^h f(x)\left[\int_{-\frac{x}{h}}^1 K^*_{\gamma,\frac{x}{h}}(s)\, K^*_{\nu,\frac{x}{h}}(s)ds - \int_{-1}^1 K^*_\gamma(s)\, K^*_\nu(s)ds\right]dx$$

and

$$\int_0^\infty f(x)\int_{-1}^1 K^*_\gamma(s)\, K^*_\nu(s)dsdx.$$

The first part integrates a integrable function over a region which has Lebesgue measure shrinking to zero and thus tends to zero. Now it is clear that

$$\int_0^\infty f(t)\omega(t,t)dt = h\left(\int K^*_\gamma K^*_\nu\right)\int_0^\infty f(x)dx + O(h^2) = h\int K^*_\gamma K^*_\nu + O(h^2).$$

Proofs of (ii) - (vi) depend on the Fubini Theorem, the smoothness of f, and K's being finitely supported as well. And we will skip the details.

(ii) It is proved exactly in the same way as part (i).

(iii)

$$\begin{aligned}
&\int_0^\infty \int_0^\infty f(t)f(u)\omega(t,u)dtdu \\
&= \int_0^\infty \int_0^\infty \int_0^\infty f(t)f(u)K^*_{\gamma,\frac{x}{h}}\left(\frac{t-x}{h}\right) K^*_{\nu,\frac{x}{h}}\left(\frac{u-x}{h}\right) dtdudx \\
&= \int_0^\infty \left[\int_0^\infty f(t)K^*_{\gamma,\frac{x}{h}}\left(\frac{t-x}{h}\right) dt\right]\left[\int_0^\infty f(u)K^*_{\nu,\frac{x}{h}}\left(\frac{u-x}{h}\right) du\right] dx \\
&= h^2 \int_0^\infty \left[\int_{-\frac{x}{h}}^1 f(x+uh)K^*_{\gamma,\frac{x}{h}}(u)\,du\right]\left[\int_{-\frac{x}{h}}^1 f(x+uh)K^*_{\nu,\frac{x}{h}}(u)\,du\right] dx
\end{aligned}$$

With a Taylor expansion around x,

$$f(x+uh) = \sum_{i=0}^{p} \frac{u^i h^i}{i!} f^{(i)}(x) + \frac{u^{p+1}h^{p+1}}{(p+1)!}\int_0^1 (1-v)^p f^{(p+1)}(x+vuh)dv,$$

and $f \in F_{p+1}$,

$$\begin{aligned}
&\int_0^\infty \int_0^\infty f(t)f(u)\omega(t,u)dtdu \\
&= h^2 \int_0^\infty \left[\sum_{i=0}^{p+1} \frac{h^i}{i!} f^{(i)}(x) \int_{-\frac{x}{h}}^1 u^i K^*_{\gamma,\frac{x}{h}}(u)\,du\right] \\
&\times \left[\sum_{i=0}^{p+1} \frac{h^i}{i!} f^{(i)}(x) \int_{-\frac{x}{h}}^1 u^i K^*_{\nu,\frac{x}{h}}(u)\,du\right] dx\,(1+o(1)).
\end{aligned}$$

Hence,

$$\begin{aligned}
&\int_0^\infty \int_0^\infty f(t)f(u)\omega(t,u)dtdu \\
&= \frac{h^{\gamma+\nu+2}}{\gamma!\nu!}\theta_{\gamma,\nu} + \frac{(1+\delta_{\gamma\nu})h^{p+(\gamma\wedge\nu)+3}}{(\gamma\wedge\nu)!(p+1)!}\theta_{\gamma\wedge\nu,p+1}\left(\int u^{p+1}K^*_{\gamma\vee\nu}\right) + O(h^{p+(\gamma\vee\nu)+3}).
\end{aligned}$$

(iv)

$$\begin{aligned}
&\int_0^\infty \int_0^\infty f(t)f(u)\omega(t,u)^2 dudt \\
&= \int_{2h}^\infty \int_0^\infty f(t)f(u)\left[\int_0^\infty K^*_{\gamma,\frac{x}{h}}\left(\frac{t-x}{h}\right) K^*_{\nu,\frac{x}{h}}\left(\frac{u-x}{h}\right) dx\right]^2 dudt\,(1+o(1)).
\end{aligned}$$

Since K is supported on $[-1,1]$, for $t \geq 2h$, $K^*_{\gamma,\frac{x}{h}}\left(\frac{t-x}{h}\right) K^*_{\nu,\frac{x}{h}}\left(\frac{u-x}{h}\right)$ is non zero only when $x \geq h$, $u \in [x-h, x+h]$, and $x \in [t-h, t+h]$. And, in that case $K^*_{\gamma,\frac{x}{h}} \equiv K^*_\gamma, K^*_{\nu,\frac{x}{h}} \equiv K^*_\nu$. Therefore the above quantity can be written as

$$\int_{2h}^{\infty} \int_{t-2h}^{t+2h} f(t)f(u) \left[\int_{t-h}^{t+h} K^*_\gamma \left(\frac{t-x}{h}\right) K^*_\nu \left(\frac{u-x}{h}\right) dx\right]^2 dudt\,(1+o(1))$$

$$= \int_{2h}^{\infty} \int_{t-2h}^{t+2h} f(t)f(u) \left[h \int_{-1}^{1} K^*_\gamma(s) K^*_\nu \left(s - \frac{t-u}{h}\right) ds\right]^2 dudt\,(1+o(1))$$

$$= h^2 \int_{2h}^{\infty} \int_{t-2h}^{t+2h} f(t)f(u) \left(K^*_\gamma * K^*_\nu\right)^2 \left(\frac{t-u}{h}\right) dudt\,(1+o(1))$$

$$= h^3 \int_{2h}^{\infty} f(t) \int_{-2}^{2} f(t-vh) \left(K^*_\gamma * K^*_\nu\right)^2 (v)\, dvdt\,(1+o(1))$$

$$= h^3 \left(\int f^2\right) \left(\int \left(K^*_\gamma * K^*_\nu\right)^2\right) + o(h^3).$$

(v) & (vi) First we show that for $t \geq 2h$,

$$\int_{t-2h}^{t+2h} \int_{t-h}^{t+h} f(u) K^*_\gamma \left(\frac{t-x}{h}\right) K^*_\nu \left(\frac{u-x}{h}\right) dxdu$$

$$= \int_{t-h}^{t+h} K^*_\gamma \left(\frac{t-x}{h}\right) \int_{x-h}^{x+h} f(u) K^*_\nu \left(\frac{u-x}{h}\right) dudx$$

$$= h \int_{t-h}^{t+h} K^*_\gamma \left(\frac{t-x}{h}\right) \int_{-1}^{1} f(x+sh) K^*_\nu(s)\, dsdx$$

$$= \frac{h^{\nu+1}}{\nu!} \int_{t-h}^{t+h} K^*_\gamma \left(\frac{t-x}{h}\right) f^{(\nu)}(x) dx\,(1+o(1)),$$

by a ν-th order Taylor expansion of $f(x+sh)$ around x and then applying the moment conditions (3.11). Once again with a change of variable and similar arguments,

$$\int_{t-2h}^{t+2h} \int_{t-h}^{t+h} f(u) K^*_\gamma \left(\frac{t-x}{h}\right) K^*_\nu \left(\frac{u-x}{h}\right) dxdu$$

$$= \frac{h^{\nu+2}}{\nu!} \int_{-1}^{1} K^*_\gamma(s)\, f^{(\nu)}(t-sh) ds\,(1+o(1))$$

$$= \frac{h^{\gamma+\nu+2}}{\gamma!\nu!} f^{(\gamma+\nu)}(t)\,(1+o(1)) \tag{4.26}$$

Now to validate (v),

$$\int_{2h}^{\infty} \int_{0}^{\infty} f(t)f(u)\omega(t,t)\omega(t,u) dtdu$$

$$= \int_{2h}^{\infty}\int_0^{\infty} f(t)f(u)\left[\int_0^{\infty} K^*_{\gamma,\frac{x}{h}}\left(\frac{t-x}{h}\right)K^*_{\nu,\frac{x}{h}}\left(\frac{t-x}{h}\right)dx\right]\times$$

$$\left[\int_0^{\infty} K^*_{\gamma,\frac{y}{h}}\left(\frac{t-y}{h}\right)K^*_{\nu,\frac{y}{h}}\left(\frac{u-y}{h}\right)dy\right]dudt$$

$$= \int_{2h}^{\infty} f(t)\left[\int_{t-h}^{t+h} K^*_{\gamma,\frac{x}{h}}\left(\frac{t-x}{h}\right)K^*_{\nu,\frac{x}{h}}\left(\frac{t-x}{h}\right)dx\right]\times$$

$$\left[\int_{t-2h}^{t+2h} f(u)\int_{t-h}^{t+h} K^*_{\gamma}\left(\frac{t-y}{h}\right)K^*_{\nu}\left(\frac{u-y}{h}\right)dydu\right]dt$$

$$= \frac{h^{\gamma+\nu+3}}{\gamma!\nu!}\left(\int K^*_\gamma K^*_\nu\right)\left(\int_{2h}^{\infty} f^{(\gamma+\nu)}(u)f(u)du\right)(1+o(1)),$$

by (4.26). Therefore,

$$\int_0^{\infty}\int_0^{\infty} f(t)f(u)\omega(t,t)\omega(t,u)dtdu = \frac{h^{\gamma+\nu+3}}{\gamma!\nu!}\left(\int K^*_\gamma K^*_\nu\right)\theta_{\gamma+\nu,0} + o(h^{\gamma+\nu+3}).$$

(vi)

$$\int_0^{\infty}\int_0^{\infty}\int_0^{\infty} f(t)f(u)f(v)\omega(t,v)\omega(t,u)dvdudt$$

$$= \int_0^{\infty} f(t)\left[\int_0^{\infty} f(u)\omega(t,u)du\right]^2 dt$$

$$= \int_{2h}^{\infty} f(t)\left[\int_0^{\infty} f(u)\omega(t,u)du\right]^2 dt\,(1+o(1))$$

$$= \int_{2h}^{\infty} f(t)\left[\int_{t-2h}^{t+2h}\int_{t-h}^{t+h} f(u)K^*_\gamma\left(\frac{t-x}{h}\right)K^*_\nu\left(\frac{u-x}{h}\right)dxdu\right]^2 dt\,(1+o(1))$$

$$= \frac{h^{2\gamma+2\nu+4}}{(\gamma!\nu!)^2}\int_{2h}^{\infty} f(t)\left(f^{(\gamma+\nu)}(t)\right)^2 dt\,(1+o(1)),$$

applying (4.26). So,

$$\int_0^{\infty}\int_0^{\infty}\int_0^{\infty} f(t)f(u)f(v)\omega(t,v)\omega(t,u)dtdudv = \frac{h^{2\gamma2+\nu+4}}{(\gamma!\nu!)^2}\left(\int f\left(f^{(2m)}\right)^2\right)+o(h^{2\gamma+2\nu+4}).$$

This finishes the proof.

Proof of Theorem 7 :

Since the first two moments of $\widehat{\theta}_{\gamma,\nu}$ are linear combinations of the expected values in Lemma 3 and the weight function W^n_m is approximately the same as $\frac{b}{h^{m+1}}K^*_m$, it is straightforward to prove the theorem using the asymptotic tools provided in Lemma 4. First to calculate the expected value of $\widehat{\theta}_{\gamma,\nu}$,

$$E\widehat{\theta}_{\gamma,\nu} = b\gamma!\nu! \sum_{i=1}^{g}\sum_{j=1}^{g}\sum_{k=1}^{g} W_\gamma^n\left(\frac{x_j - x_i}{h}\right) W_\nu^n\left(\frac{x_k - x_i}{h}\right) E\left(\frac{c_j c_k}{n^2 b^2}\right)$$

$$= b\gamma!\nu! \sum_{i=1}^{g}\sum_{j=1}^{g} W_\gamma^n\left(\frac{x_j - x_i}{h}\right) W_\nu^n\left(\frac{x_j - x_i}{h}\right) E\left(\frac{c_j^2}{n^2 b^2}\right)$$

$$+ b\gamma!\nu! \sum_{i=1}^{g}\sum\sum_{j\neq k} W_\gamma^n\left(\frac{x_j - x_i}{h}\right) W_\nu^n\left(\frac{x_k - x_i}{h}\right) E\left(\frac{c_j c_k}{n^2 b^2}\right). \qquad (4.27)$$

From (ii) and (iv) of Lemma 3 and (3.8),

$$E\widehat{\theta}_{\gamma,\nu} = \frac{b^2\gamma!\nu!}{h^{\gamma+\nu+2}} \sum_{i=1}^{g}\sum_{j=1}^{g} K_\gamma^*\left(\frac{x_j - x_i}{h}\right) K_\nu^*\left(\frac{x_j - x_i}{h}\right) \frac{f(x_j) + nbf(x_j)^2}{n} (1 + \circ(1))$$

$$+\frac{b^3\gamma!\nu!}{h^{\gamma+\nu+2}} \sum_{i=1}^{g}\sum\sum_{j\neq k} K_\gamma^*\left(\frac{x_j - x_i}{h}\right) K_\nu^*\left(\frac{x_k - x_i}{h}\right) f(x_j)f(x_k) (1 + \circ(1))$$

$$= \frac{b\gamma!\nu!}{nh^{\gamma+\nu+2}} \sum_{j=1}^{g} f(x_j) \left[\sum_{i=1}^{g} K_\gamma^*\left(\frac{x_j - x_i}{h}\right) K_\nu^*\left(\frac{x_j - x_i}{h}\right) b\right] (1 + \circ(1))$$

$$+ \frac{b^2\gamma!\nu!}{h^{\gamma+\nu+2}} \sum_{j=1}^{g}\sum_{k=1}^{g} f(x_j)f(x_k) \left[\sum_{i=1}^{g} K_\gamma^*\left(\frac{x_j - x_i}{h}\right) K_\nu^*\left(\frac{x_k - x_i}{h}\right) b\right] (1 + \circ(1)). \quad (4.28)$$

Now use Lemma 1 to approximate the first term of (4.28) by some integral. Take

$$G(s) = K_\gamma^*\left(\frac{x_j - s}{h}\right) K_\nu^*\left(\frac{x_j - s}{h}\right),$$

and the grid width $\triangle$ to be b/h, then

$$b\sum_{j=1}^{g} f(x_j) \left[\sum_{i=1}^{g} K_\gamma^*\left(\frac{x_j - x_i}{h}\right) K_\nu^*\left(\frac{x_j - x_i}{h}\right) \frac{b}{h}\right]$$

$$= b\sum_{j=1}^{g} f(x_j) \left[\int K_\gamma^*\left(\frac{x_j - u}{h}\right) K_\nu^*\left(\frac{x_j - u}{h}\right) \frac{du}{h} + R_j\right]$$

$$= b\sum_{j=1}^{g} f(x_j) \left[\int K_\gamma^*\left(\frac{x_j - u}{h}\right) K_\nu^*\left(\frac{x_j - u}{h}\right) \frac{du}{h}\right] (1 + \circ(1)),$$

since, for each j, $|R_j| \leq \frac{b^2}{4h^2} \int \left|(K_\gamma^* K_\nu^*)^{(2)}(s)\right| ds$ and $b \sum_{j=1}^{g} f(x_j)$ is finite. Once more applying the same method, the above quantity approximately equal to $\int\int f(t)\omega(t,t)dt$.

Similar arguments show that the second term of (4.28) can be approximated by $\frac{\gamma!\nu!}{h^{\gamma+\nu+2}}\int\int f(t)f(u)\omega(t,u)dudt$. Hence

$$E\widehat{\theta}_{\gamma,\nu} =$$

$$\frac{\gamma!\nu!}{nh^{\gamma+\nu+2}}\left[\int f(t)\omega(t,t)dt + n\int\int f(t)f(u)\omega(t,u)dudt\right](1+o(1)). \qquad (4.29)$$

Combining (4.29), (i) and (iii) of Lemma 4 yields (4.13), the asymptotic bias of $\widehat{\theta}_{\gamma,\nu}$.

Next, we calculate the second moment of $\widehat{\theta}_{\gamma,\nu}$. Define

$$\omega_n(t,u) = \sum_{i=1}^{g} W_\gamma^n\left(\frac{t-x_i}{h}\right) W_\nu^n\left(\frac{u-x_i}{h}\right) b,$$

for any real numbers t and u. Then

$$E\widehat{\theta}^2_{\gamma,\nu} = \frac{(\gamma!\nu!)^2}{n^4b^4} E\left(\sum_{i=1}^{g}\sum_{j=1}^{g}\omega_n(x_i,x_j)c_ic_j\right)^2$$

$$= \frac{(\gamma!\nu!)^2}{n^4b^4}\sum_{i=1}^{g}\sum_{j=1}^{g}\sum_{k=1}^{g}\sum_{l=1}^{g}\omega_n(x_i,x_j)\omega_n(x_k,x_l)E\left(c_ic_jc_kc_l\right)$$

By Lemma 3,

$$\frac{n^4b^4}{(\gamma!\nu!)^2}E\widehat{\theta}^2_{\gamma,\nu}$$

$$\approx \sum\sum\sum\sum_{i,j,k,l \text{ all different}} \frac{n!b^4}{(n-4)!} f(x_i)f(x_j)f(x_k)f(x_l)\omega_n(x_i,x_j)\omega_n(x_k,x_l)$$

$$+\sum\sum\sum_{i\neq j\neq k\neq i}\frac{2n!b^3}{(n-3)!}f(x_i)f(x_j)f(x_k)\left[(n-3)bf(x_i)+1\right]\times$$

$$\left[\omega_n(x_i,x_i)\omega_n(x_j,x_k)+2\omega_n(x_i,x_j)\omega_n(x_i,x_k)\right]$$

$$+\sum\sum_{i\neq j}\frac{n!b^2}{(n-2)!}f(x_i)f(x_j)\left[1+(n-2)b\left(f(x_i)+f(x_j)\right)+\frac{b^2(n-2)!}{(n-4)!}f(x_i)f(x_j)\right]\times$$

$$\left[\omega_n(x_i,x_i)\omega_n(x_j,x_j)+2\omega_n(x_i,x_j)\omega_n(x_i,x_j)\right]$$

$$+\sum\sum_{i\neq j}\frac{4n!b^2}{(n-2)!}f(x_i)f(x_j)\left[1+3(n-2)bf(x_i)+\frac{b^2(n-2)!}{(n-4)!}f(x_i)^2\right]\omega_n(x_i,x_i)\omega_n(x_i,x_j)$$

$$+\sum_{i=1}^{g} nbf(x_i)\left[1+7(n-1)bf(x_i)+\frac{6(n-1)!b^2}{(n-3)!}f(x_i)^2+\frac{b^3(n-1)!}{(n-4)!}f(x_i)^3\right]\omega_n(x_i,x_i)^2.$$

The last expression can be rewritten as

$$\frac{n!b^4}{(n-4)!}\sum_{i=1}^{g}\sum_{j=1}^{g}\sum_{k=1}^{g}\sum_{l=1}^{g} f(x_i)f(x_j)f(x_k)f(x_l)\omega_n(x_i,x_j)\omega_n(x_k,x_l)$$

$$+\frac{2n!b^3}{(n-3)!}\sum_{i=1}^{g}\sum_{j=1}^{g}\sum_{k=1}^{g} f(x_i)f(x_j)f(x_k)\left[\omega_n(x_i,x_i)\omega_n(x_j,x_k)+2\omega_n(x_i,x_j)\omega_n(x_i,x_k)\right]$$

$$+\frac{n!b^2}{(n-2)!}\sum_{i=1}^{g}\sum_{j=1}^{g} f(x_i)f(x_j)\left[\omega_n(x_i,x_i)\omega_n(x_j,x_j)+2\omega_n(x_i,x_j)^2+4\omega_n(x_i,x_i)\omega_n(x_i,x_j)\right]$$

$$+\frac{n!b^2}{(n-2)!}\sum_{i=1}^{g}\sum_{j=1}^{g} f(x_i)f(x_j)\left[\omega_n(x_i,x_i)\omega_n(x_j,x_j)+2\omega_n(x_i,x_j)^2+4\omega_n(x_i,x_i)\omega_n(x_i,x_j)\right]$$

$$+nb\sum_{i=1}^{g} f(x_i)\omega_n(x_i,x_i)^2.$$

Similar to the approximation (4.29) of (4.27), $\frac{n^4h^{2\gamma+2\nu+4}}{(\gamma!\nu!)^2}E\widehat{\theta}^2_{\gamma,\nu}$ is approximately

$$\frac{n!}{(n-4)!}\int\int\int\int f(r)f(s)f(t)f(u)\omega(r,s)\omega(t,u)drdsdtdu$$

$$+\frac{2n!}{(n-3)!}\int\int\int f(t)f(u)f(v)\left[\omega(t,t)\omega(u,v)+2\omega(t,u)\omega(t,v)\right]dtdudv$$

$$+\frac{n!}{(n-2)!}\int\int f(t)f(u)\left[\omega(t,t)\omega(u,u)+2\omega(t,u)^2+4\omega(t,t)\omega(t,u)\right]dudt$$

$$+n\int f(t)\omega(t,t)^2dt.$$

Subtracting the above quantity by the squared asymptotic bias of $\frac{n^2h^{\gamma+\nu+2}}{\gamma!\nu!}\widehat{\theta}_{\gamma,\nu}$ yields

$$\frac{n^4h^{2\gamma+2\nu+4}}{(\gamma!\nu!)^2}Var\left(\widehat{\theta}_{\gamma,\nu}\right)$$

$$\approx\left[\frac{n!}{(n-4)!}-n^2(n-1)^2\right]\left[\int\int f(t)f(u)\omega(t,u)dtdu\right]^2$$

$$+2\left[\frac{n!}{(n-3)!}-n^2(n-1)\right]\left[\int\int f(t)f(u)\omega(t,u)dtdu\right]\left[\int f(t)\omega(t,t)dt\right]$$

$$+\frac{4n!}{(n-3)!}\int\int\int f(t)f(u)f(v)\omega(t,u)\omega(t,v)drdsdt$$

$$+\left[n(n-1)-n^2\right]\left[\int f(t)\omega(t,t)dt\right]^2dudt+2n(n-1)\int\int f(t)f(u)\omega(t,u)^2dtdu$$

$$+4n(n-1)\int\int f(t)f(u)\omega(t,t)\omega(t,u)dtdu+n\int f(t)\omega(t,t)^2dt.$$

Applying the results in Lemma 4, we have $\frac{n^4 h^{2\gamma+2\nu+4}}{(\gamma!\nu!)^2} Var\left(\widehat{\theta}_{\gamma,\nu}\right)$ being

$$-(4n-6)n(n-1)\left[\frac{h^{\gamma+\nu+2}}{(\gamma!\nu!)^2}\int f^{(\gamma)}f^{(\nu)} + o(h^{\gamma+\nu+2})\right]^2$$

$$-4n(n-1)\left[\frac{h^{\gamma+\nu+2}}{(\gamma!\nu!)^2}\int f^{(\gamma)}f^{(\nu)} + o(h^{\gamma+\nu+2})\right]\left[h\int K_\gamma^* K_\nu^* + o(h)\right]$$

$$+\frac{4n!}{(n-3)!}\left[\frac{h^{2\gamma+2\nu+4}}{(\gamma!\nu!)^2}\int f\left(f^{(\gamma+\nu)}\right)^2 + o(h^{2\gamma+2\nu+4})\right]$$

$$-n\left[h\int K_\gamma^* K_\nu^* + o(h)\right]^2 + 2n(n-1)\left[h^3\left(\int f^2\right)\int\left(K_\gamma^* * K_\nu^*\right)^2 + o(h^3)\right]$$

$$+4n(n-1)\left[\frac{h^{\gamma+\nu+3}}{\gamma!\nu!}\left(\int K_\gamma^* K_\nu^*\right)\int f^{(\gamma+\nu)}f + o(h^{\gamma+\nu+3})\right]+n\left[h^2\left(\int K_\gamma^* K_\nu^*\right)^2 + o(h^2)\right].$$

Therefore

$$Var\left(\widehat{\theta}_{\gamma,\nu}\right) = \frac{2(\gamma!\nu!)^2}{n^2 h^{2\gamma+2\nu+1}}\left(\int f^2\right)\left(\int\left(K_\gamma^* * K_\nu^*\right)^2\right) +$$

$$\frac{4}{n}\left(\int f\left(f^{(\gamma+\nu)}\right)^2 - \theta_{\gamma,\nu}^2\right) + \left(o(n^{-2}h^{-2(\gamma+\nu)-1}) + o(n^{-1})\right).$$

This finishes the proof.

(II) Proof of Theorem 8: First define

$$A_n(a) = \frac{-24b}{n^2 a^7}\sum_{i=1}^{g}\sum_{j=1}^{g}\sum_{k=1}^{g} K_2^*\left(\frac{x_j - x_k}{a}\right)K_2^*\left(\frac{x_i - x_k}{a}\right)c_i c_j,$$

$$B_n(a) = \frac{-8b}{n^2 a^7}\sum_{i=1}^{g}\sum_{j=1}^{g}\sum_{k=1}^{g} \frac{x_j - x_k}{a}\left(K_2^*\right)'\left(\frac{x_j - x_k}{a}\right)K_2^*\left(\frac{x_i - x_k}{a}\right)c_i c_j.$$

And some necessary lemmas for the proof are given below.

Lemma 5 *Under condition (C3), if $b = o(a)$,*

$$\frac{d}{da}\widehat{\theta}_{2,2}(a) = \left(A_n(a) + B_n(a)\right)\left(1 + o_p(1)\right),$$

as $n \to \infty$.

Lemma 6 *Under conditions (C1) - (C5), as* $n \to \infty, a \to 0, na^5 \to \infty$,

$$E\left(B_n(a)\right) = \frac{6}{a} R\left(f^{(2)}\right) + O\left(\frac{1}{na^6}\right) + O\left(a\right),$$

and

$$Var\left(B_n(a)\right) = \frac{32}{n^2 a^{11}} R(f)\left(R(G * K_2^*)\right)^2 + O\left(\frac{1}{na^2}\right),$$

where $G(x) = x\left(K_2^*\right)'(x)$.

Lemma 7 *Under conditions (C1) - (C3), as* $n \to \infty, a \to 0, na^5 \to \infty$, *and* $b = o(a)$,

$$na^5\left(\widehat{\theta}_{2,2}(a) - R\left(f^{(2)}\right)\right) \xrightarrow{D} N\left(\mu_*, \sigma_*^2\right),$$

where

$$\mu_* = 4R\left(K_2^*\right) + 4a\int u\left(K_2^*\right)^2 \int f^{(1)} + \frac{na^7}{6}\left(\int f^{(2)} f^{(4)}\right)\left(\int u^4 K_2^*\right),$$

and

$$\sigma_*^2 = 32aR(f)R\left(K_2^* * K_2^*\right).$$

Lemma 8 *Under conditions (C1) - (C3), if* $h \sim n^{-1/5}$,

$$\widehat{\theta}_{2,2}\left(a_{\widehat{\lambda}}(h)\right) - \widehat{\theta}_{2,2}\left(a_{\lambda}(h)\right) = o_p(n^{-1/2}).$$

Proof of Lemma 5:

Note from (4.17) that S_n^{-1} is itself a function of a and hence it is necessary to calculate $\frac{d}{da} e_3^T S_n^{-1}$ for taking derivative of $\widehat{\theta}_{2,2}(a)$ with respect to a. Write

$$S_{n,j} = \sum_{k=1}^{g} K\left(\frac{x_k - x}{a}\right)(x_k - x)^j, j = 0, 1, ..., 6.$$

Then explicitly,

$$e_3^T S_n^{-1} = \frac{1}{\det\left(S_n\right)}\left(A_{S_n}^{(1)}, A_{S_n}^{(2)}, A_{S_n}^{(3)}, A_{S_n}^{(4)}\right),$$

here , for any matrix $M = \left(M_{i+j-2}\right)_{1 \le i,j \le 4}$,

$$A_M^{(1)} = M_1\left(M_3 M_6 - M_4 M_5\right) - M_2\left(M_2 M_6 - M_3 M_5\right) + M_4\left(M_2 M_4 - M_3 M_3\right),$$

$$A_M^{(2)} = -M_0\left(M_3M_6 - M_4M_5\right) + M_1\left(M_2M_6 - M_3M_5\right) - M_3\left(M_2M_4 - M_3M_3\right),$$
$$A_M^{(3)} = M_6\left(M_0M_2 - M_1M_1\right) - M_4\left(M_0M_4 - M_1M_3\right) + M_3\left(M_1M_4 - M_2M_3\right),$$
$$A_M^{(4)} = -M_5\left(M_0M_2 - M_1M_1\right) + M_3\left(M_0M_4 - M_1M_3\right) - M_2\left(M_1M_4 - M_2M_3\right). \tag{4.30}$$

Also,

$$\det(M) = \sum_{i=1}^{4} M_{i+1} A_M^{(i)}. \tag{4.31}$$

Let $S = \left(S_{i+j-2}\right)_{1 \le i,j \le 4}$, where $S_j = \int u^j K, j = 0, ..., 6$. Then

$$\begin{aligned}
\frac{b}{a^{j+1}} S_{n,j} - S_j =& \sum_{i=1}^{g} K\left(\frac{x_i - x}{a}\right)\left(\frac{x_i - x}{a}\right)^j \frac{b}{a} - \int u^j K(u)\, du \\
=& \sum_{i=1}^{g} \int_{x_i^* - \frac{b}{2a}}^{x_i^* + \frac{b}{2a}} K(x_i^*)(x_i^*)^j du - \sum_{i=1}^{g} \int_{x_i^* - \frac{b}{2a}}^{x_i^* + \frac{b}{2a}} u^j K(u)\, du \\
=& \sum_{i=1}^{g} \int_{x_i^* - \frac{b}{2a}}^{x_i^* + \frac{b}{2a}} \left[G(x_i^*) - G(u)\right] du,
\end{aligned}$$

where $x_i^* = \frac{x_i - x}{a}$ and $G(u) = u^j K(u)$. Taking a second order Taylor expansion at each x_i^* and writing $R_j = \frac{b}{a^{j+1}} S_{n,j} - S_j$,

$$\begin{aligned}
R_j =& \sum_{i=1}^{g} \int_{x_i^* - \frac{b}{2a}}^{x_i^* + \frac{b}{2a}} \left[G^{(1)}(x_i^*)(x_i^* - u) + \frac{1}{2}\int_u^{x_i^*} (u - t)\, G^{(2)}(t) dt\right] du \\
=& \frac{1}{2} \sum_{i=1}^{g} \int_{x_i^* - \frac{b}{2a}}^{x_i^* + \frac{b}{2a}} \int_u^{x_i^*} (u - t)\, G^{(2)}(t) dt du \\
=& \frac{1}{2} \sum_{i=1}^{g} \left[\int_{x_i^* - \frac{b}{2a}}^{x_i^*} \int_{x_i^* - \frac{b}{2a}}^{t} - \int_{x_i^*}^{x_i^* + \frac{b}{2a}} \int_t^{x_i^* + \frac{b}{2a}}\right] (u - t)\, G^{(2)}(t) du dt \\
=& \frac{-b^2}{8a^2} \int_{-\infty}^{+\infty} G^{(2)}(t) dt.
\end{aligned}$$

There are two implications from this result. First,

$$\frac{b}{a^{j+1}} S_{n,j} = S_j + o(1), \tag{4.32}$$

and

$$W_2^n(t) = \frac{b}{a^3} K_2^*(t) + o\left(\frac{b}{a^3}\right), \tag{4.33}$$

since $b = o(a)$. Secondly, observe that

$$\frac{d}{da}\frac{b}{a^j} S_{n,j} = \frac{d}{da}\left(aS_j + aR_j\right) = \frac{d}{da}\left(aS_j - \frac{b^2}{8a}\int_{-\infty}^{+\infty} G^{(2)}(t) dt\right)$$

$$= S_j + \frac{b^2}{8a^2}\int_{-\infty}^{+\infty} G^{(2)}(t)dt,$$

and since $b = o(a)$,

$$\frac{d}{da}\frac{b}{a^j}S_{n,j} = S_j + o(1). \tag{4.34}$$

From (4.31),

$$\frac{b^4}{a^{12}}\det(S_n) = \sum_{\underset{\sim}{l}\in L}(-1)^{p(\underset{\sim}{l})}S_{n,l_1}S_{n,l_2}S_{n,l_3}S_{n,l_4}\frac{b^4}{a^{12}},$$

where $\underset{\sim}{l}= (l_1, ..., l_4), L$ denotes the set of all $\underset{\sim}{l}$ in the expression of $\det(S_n)$, and $p(\underset{\sim}{l})$ equals to +1 or -1 according to $\underset{\sim}{l}$. Noticing $l_1 + l_2 + l_3 + l_4 \equiv 12$ for every $\underset{\sim}{l}\in L$,

$$\frac{d}{da}\left(\frac{b^4}{a^{12}}\det(S_n)\right)$$

$$=\sum_{\underset{\sim}{l}\in L}(-1)^{p(\underset{\sim}{l})}\frac{d}{da}\left\{\left(\frac{b}{a^{l_1}}S_{n,l_1}\right)\left(\frac{b}{a^{l_2}}S_{n,l_2}\right)\left(\frac{b}{a^{l_3}}S_{n,l_3}\right)\left(\frac{b}{a^{l_4}}S_{n,l_4}\right)\right\}$$

$$=\sum_{\underset{\sim}{l}\in L}(-1)^{p(\underset{\sim}{l})}\left\{\left(\frac{d}{da}\frac{b}{a^{l_1}}S_{n,l_1}\right)\left(\frac{b}{a^{l_2}}S_{n,l_2}\right)\left(\frac{b}{a^{l_3}}S_{n,l_3}\right)\left(\frac{b}{a^{l_4}}S_{n,l_4}\right)\right.$$

$$+\left(\frac{b}{a^{l_1}}S_{n,l_1}\right)\left(\frac{d}{da}\frac{b}{a^{l_2}}S_{n,l_2}\right)\left(\frac{b}{a^{l_3}}S_{n,l_3}\right)\left(\frac{b}{a^{l_4}}S_{n,l_4}\right)$$

$$+\left(\frac{b}{a^{l_1}}S_{n,l_1}\right)\left(\frac{b}{a^{l_2}}S_{n,l_2}\right)\left(\frac{d}{da}\frac{b}{a^{l_3}}S_{n,l_3}\right)\left(\frac{b}{a^{l_4}}S_{n,l_4}\right)$$

$$\left.+\left(\frac{b}{a^{l_1}}S_{n,l_1}\right)\left(\frac{b}{a^{l_2}}S_{n,l_2}\right)\left(\frac{b}{a^{l_3}}S_{n,l_3}\right)\left(\frac{d}{da}\frac{b}{a^{l_4}}S_{n,l_4}\right)\right\}.$$

From (4.32) and (4.34),

$$\frac{d}{da}\left(\frac{b^4}{a^{12}}\det(S_n)\right) = 4a^3\sum_{\underset{\sim}{l}\in L}(-1)^{p(\underset{\sim}{l})}S_{l_1}S_{l_2}S_{l_3}S_{l_4}(1+o(1)),$$

or equivalently,

$$\frac{d}{da}\left(\frac{b^4}{a^{12}}\det(S_n)\right) = 4a^3\det(S)(1+o(1)). \tag{4.35}$$

With exactly analogous arguments, it can be shown that

$$\frac{d}{da}\left(\frac{b^3}{a^{11-l}}A_{S_n}^{(l)}\right) = 3a^2A_S^{(l)}(1+o(1)), l = 1, ..., 4. \tag{4.36}$$

Also,

$$\frac{b^4}{a^{12}}\det(S_n) = a^4\det(S)(1+o(1)), \tag{4.37}$$

and

$$\frac{b^3}{a^{11-l}} A_{S_n}^{(l)} = a^3 A_S^{(l)} \left(1 + o(1)\right), l = 1, ..., 4. \tag{4.38}$$

Combining (4.35) - (4.38), for any $l \in \{1, 2, 3, 4\}$,

$$\frac{d}{da}\left(\frac{a^{1+l} A_{S_n}^{(l)}}{b \det\left(S_n\right)}\right) = \frac{d}{da}\left(\frac{\frac{b^3}{a^{11-l}} A_{S_n}^{(l)}}{\frac{b^4}{a^{12}} \det\left(S_n\right)}\right)$$

$$= \frac{\left(\frac{d}{da}\frac{b^3}{a^{11-l}} A_{S_n}^{(l)}\right)\left(\frac{b^4}{a^{12}} \det\left(S_n\right)\right) - \left(\frac{b^3}{a^{11-l}} A_{S_n}^{(l)}\right)\left(\frac{d}{da}\frac{b^4}{a^{12}} \det\left(S_n\right)\right)}{\left(\frac{b^4}{a^{12}} \det\left(S_n\right)\right)^2}$$

$$= \frac{\left(3a^2 A_S^{(l)}\right)\left(a^4 \det\left(S\right)\right) - \left(a^3 A_S^{(l)}\right)\left(4a^3 \det\left(S\right)\right)}{\left(a^4 \det\left(S\right)\right)^2}\left(1 + o(1)\right).$$

That is

$$\frac{d}{da}\left(\frac{a^{1+l} A_{S_n}^{(l)}}{b \det\left(S_n\right)}\right) = \frac{-1}{a^2}\frac{A_S^{(l)}}{\det\left(S\right)}\left(1 + o(1)\right), l = 1, ..., 4. \tag{4.39}$$

Making use of (4.37) - (4.39),

$$\frac{d}{da}\left(\frac{A_{S_n}^{(l)}}{\det\left(S_n\right)} K\left(\frac{x_i - x_k}{a}\right)\left(x_i - x_k\right)^{l-1}\right)$$

$$= \left(\frac{d}{da}\frac{a^{1+l} A_{S_n}^{(l)}}{b \det\left(S_n\right)}\right)\left(\frac{b}{a^2} K\left(\frac{x_i - x_k}{a}\right)\left(\frac{x_i - x_k}{a}\right)^{l-1}\right)$$

$$+ \left(\frac{a^{1+l} A_{S_n}^{(l)}}{b \det\left(S_n\right)}\right)\left(\frac{d}{da}\frac{b}{a^2} K\left(\frac{x_i - x_k}{a}\right)\left(\frac{x_i - x_k}{a}\right)^{l-1}\right)$$

$$= \left(\frac{-1}{a^2}\frac{A_S^{(l)}}{\det\left(S\right)}\left(1 + o(1)\right)\right)\left(\frac{b}{a^2} K\left(\frac{x_i - x_k}{a}\right)\left(\frac{x_i - x_k}{a}\right)^{l-1}\right) + \left(\frac{1}{a}\frac{A_S^{(l)}}{\det\left(S\right)}\left(1 + o(1)\right)\right)$$

$$\times\left(\frac{-2}{a^3} K\left(\frac{x_i - x_k}{a}\right)\left(\frac{x_i - x_k}{a}\right)^{l-1} + \frac{b}{a^2}\left\{\frac{d}{da} K\left(\frac{x_i - x_k}{a}\right)\left(\frac{x_i - x_k}{a}\right)^{l-1}\right\}\right)$$

$$= \frac{A_S^{(l)}}{\det\left(S\right)}\left(\frac{-3b}{a^4} K\left(\frac{x_i - x_k}{a}\right)\left(\frac{x_i - x_k}{a}\right)^{l-1} + \frac{b}{a^3}\left\{\frac{d}{da} K\left(\frac{x_i - x_k}{a}\right)\left(\frac{x_i - x_k}{a}\right)^{l-1}\right\}\right).$$

Applying this result and (4.33) to the first derivative of $\widehat{\theta}_{2,2}(a)$ with respect to a,

$$\frac{d}{da}\widehat{\theta}_{2,2}(a) = \frac{d}{da}\frac{4}{n^2 b}\sum_{i=1}^{g}\sum_{j=1}^{g}\sum_{k=1}^{g} W_2^n\left(\frac{x_i - x_k}{a}\right) W_2^n\left(\frac{x_j - x_k}{a}\right) c_i c_j$$

$$= \frac{8}{n^2 b}\sum_{i=1}^{g}\sum_{j=1}^{g}\sum_{k=1}^{g}\left(\frac{d}{da} W_2^n\left(\frac{x_i - x_k}{a}\right)\right) W_2^n\left(\frac{x_j - x_k}{a}\right) c_i c_j$$

$$= \frac{8}{n^2 b} \sum_{i=1}^{g}\sum_{j=1}^{g}\sum_{k=1}^{g} W_2^n\left(\frac{x_j - x_k}{a}\right)\left(\sum_{l=1}^{4} \frac{d}{da}\frac{A_{S_n}^{(l)}}{\det(S_n)} K\left(\frac{x_i - x_k}{a}\right)(x_i - x_k)^{l-1}\right) c_i c_j$$

$$= \frac{8}{n^2 b} \sum_{i=1}^{g}\sum_{j=1}^{g}\sum_{k=1}^{g} \left(\frac{b}{a^3} K_2^*\left(\frac{x_j - x_k}{a}\right)\right)\left(\frac{-3b}{a^4}\sum_{l=1}^{4} \frac{A_S^{(l)}}{\det(S)} K\left(\frac{x_i - x_k}{a}\right)\left(\frac{x_i - x_k}{a}\right)^{l-1}\right.$$

$$\left. + \frac{b}{a^3}\sum_{l=1}^{4} \frac{A_S^{(l)}}{\det(S)}\left\{\frac{d}{da} K\left(\frac{x_i - x_k}{a}\right)\left(\frac{x_i - x_k}{a}\right)^{l-1}\right\}\right) c_i c_j (1 + o_p(1)),$$

i.e.

$$\frac{d}{da}\hat{\theta}_{2,2}(a) = \frac{-24b}{n^2 a^7} \sum_{i=1}^{g}\sum_{j=1}^{g}\sum_{k=1}^{g} K_2^*\left(\frac{x_i - x_k}{a}\right) K_2^*\left(\frac{x_j - x_k}{a}\right) c_i c_j (1 + o_p(1))$$

$$-\frac{8b}{n^2 a^7} \sum_{i=1}^{g}\sum_{j=1}^{g}\sum_{k=1}^{g} K_2^*\left(\frac{x_i - x_k}{a}\right)\left(\frac{x_j - x_k}{a}\right)(K_2^*)'\left(\frac{x_j - x_k}{a}\right)(1 + o_p(1)).$$

Proof of Lemma 6:

The function K_2^* satisfies the following moment conditions

$$\int u^i K_2^*(u) du = \delta_{i,2}, i = 0, 1, 2, 3. \tag{4.40}$$

Together with condition (C4) this implies

$$\int u\,(K_2^*)'(u) = uK_2^*(u)|_{-1}^{1} - \int_{-1}^{1} K_2^*(u) du = 0, \tag{4.41}$$

$$\int u^2\,(K_2^*)'(u) = u^2 K_2^*(u)\Big|_{-1}^{1} - 2\int_{-1}^{1} uK_2^*(u) du = 0, \tag{4.42}$$

and

$$\int u^3\,(K_2^*)'(u) = u^3 K_2^*(u)\Big|_{-1}^{1} - 3\int_{-1}^{1} u^2 K_2^*(u) du = -3. \tag{4.43}$$

Some primary asymptotic results needed for the proof are provided in the following series of items (i) - (vii). First, define the real valued function ϖ on R^2 as

$$\varpi(t, u) = \int \frac{(u - s)}{a}(K_2^*)'\left(\frac{u - s}{a}\right) K_2^*\left(\frac{t - s}{a}\right) ds.$$

Since K is symmetric about zero,

$$\varpi(u, t) = \int \frac{(t - s)}{a}(K_2^*)'\left(\frac{t - s}{a}\right) K_2^*\left(\frac{u - s}{a}\right) ds$$

$$= a \int r \, (K_2^*)' \, (r) K_2^* (r - \frac{t-u}{a}) dr$$

$$= a \int s \, (K_2^*)' \, (s) K_2^* (s + \frac{t-u}{a}) ds.$$

Therefore

$$\varpi(u,t) = \varpi(t,u) \tag{4.44}$$

(i) Using conditions (C1) - (C4), equations (4.40) through (4.43), and Taylor expansions of f ,

$$\int \varpi(t,u) f(t) dt = \int \int \frac{(u-s)}{a} (K_2^*)' \left(\frac{u-s}{a}\right) K_2^* \left(\frac{t-s}{a}\right) f(t) ds dt$$

$$= a \int \int \frac{(u-s)}{a} (K_2^*)' \left(\frac{u-s}{a}\right) K_2^*(r) f(s + ra) dr ds$$

$$= \frac{a^3}{2} \int \frac{(u-s)}{a} (K_2^*)' \left(\frac{u-s}{a}\right) f(s) ds \left(1 + O(a^2)\right)$$

$$= \frac{a^4}{2} \int r \, (K_2^*)' \, (r) f^{(2)}(u - ra) dr \left(1 + O(a^2)\right).$$

Hence,

$$\int \varpi(t,u) f(t) dt = -\frac{3a^6}{4} f^{(4)}(u) + O(a^7). \tag{4.45}$$

(ii)

$$\int \varpi(t,t) f(t) dt = \int \int \frac{(t-s)}{a} (K_2^*)' \left(\frac{t-s}{a}\right) K_2^* \left(\frac{t-s}{a}\right) f(t) dt ds$$

$$= a \int \int u \, (K_2^*)' \, (u) \, K_2^*(u) f(s + ua) du ds$$

$$= a \int u \, (K_2^*)' \, (u) \, K_2^*(u) du \, (1 + O(a)), \tag{4.46}$$

by a Taylor expansion of $f(y + ua)$ at y and condition (C1).

(iii)

$$\int \int \varpi(t,u) f(t) f(u) dt du$$

$$= \int \int \int \frac{(u-z)}{a} (K_2^*)' \left(\frac{u-z}{a}\right) K_2^* \left(\frac{t-z}{a}\right) f(t) f(u) dt du dz$$

$$= a \int \int \int \frac{(u-z)}{a} (K_2^*)' \left(\frac{u-z}{a}\right) K_2^* \, (s) \, f(u) f(z + sa) ds du dz$$

$$= \frac{a^3}{2} \int \int \frac{(u-z)}{a} (K_2^*)' \left(\frac{u-z}{a}\right) f(u) f''(z) du dz \, (1 + o(1))$$

$$= \frac{a^4}{2} \int\int s\,(K_2^*)'(s) f(z+sa) f''(z)dsdz\,(1+o(1))$$

$$= \frac{a^4}{2} \int\int s\,(K_2^*)'(s) \left[\sum_{j=0}^{2} \frac{s^j a^j}{j!} f^{(j)}(z)\right] f''(z)dsdz\,(1+o(1)).$$

Utilizing the moment properties of $(K_2^*)'$ in (4.41) - (4.43),

$$\int\int \varpi(t,u) f(t) f(u) dtdu = -\frac{3a^6}{4} R(f'') \left(1 + O(a^2)\right). \tag{4.47}$$

(iv)

$$\varpi(t,t) = \int \frac{(t-s)}{a} (K_2^*)' \left(\frac{t-s}{a}\right) K_2^* \left(\frac{t-s}{a}\right) ds$$

$$= a \int r\,(K_2^*)'(r) K_2^*(r) dr. \tag{4.48}$$

Hence,

$$\int f(t)\varpi(t,t)^2 dt = \int f(t) \left[a \int r\,(K_2^*)'(r)K_2^*(r)dr\right]^2 dt\,(1+o(1))$$

$$= a^2 \left[\int r\,(K_2^*)'(r)K_2^*(r)dr\right]^2 (1+o(1)). \tag{4.49}$$

(v) From (4.45) and (4.48),

$$\int\int \varpi(t,t)\varpi(t,u) f(t) f(u) dtdu$$

$$= \frac{-3a^7}{4} \left[a \int r\,(K_2^*)'(r)K_2^*(r)dr\right] \left[\int f f^{(4)}\right] \left(1 + O(a^2)\right). \tag{4.50}$$

(vi)

$$\int\int \varpi(t,u)^2 f(t) f(u) dtdu$$

$$= \int\int f(t)f(u) \left[\int \frac{(u-s)}{a} (K_2^*)' \left(\frac{u-s}{a}\right) K_2^* \left(\frac{t-s}{a}\right) ds\right]^2 dtdu$$

$$= a^2 \int\int f(t)f(u)\,(G * K_2^*)^2 \left(\frac{u-t}{a}\right) dudt$$

$$= a^3 \int f(t) \int f(t-va)\,(G * K_2^*)^2\,(v) dvdt.$$

Thus,

$$\int\int \varpi(t,u)^2 f(t) f(u) dtdu = a^3 R(f) \left(\int (G * K_2^*)^2\right) + O(a^4). \tag{4.51}$$

(vii) From (4.44) and (4.45),

$$\int\int\int \varpi(t,v)\varpi(t,u)f(t)f(u)f(v)dtdudv$$
$$= \int f(t)\left[\int f(u)\varpi(u,t)du\right]^2 dt$$
$$= \frac{9a^{12}}{16}\int f(t)\left(f^{(4)}(t)\right)^2 dt\left(1+O(a^2)\right). \qquad (4.52)$$

To calculate the expected value of B_n, notice from the results of Lemma 3,

$$E\left\{\sum_{i=1}^{g}\sum_{j=1}^{g}\sum_{k=1}^{g}(x_j-x_k)(K_2^*)'\left(\frac{x_j-x_k}{a}\right)K_2^*\left(\frac{x_i-x_k}{a}\right)c_ic_j\right\}$$
$$=\sum_{i=1}^{g}\sum_{k=1}^{g}(x_i-x_k)(K_2^*)'\left(\frac{x_i-x_k}{a}\right)K_2^*\left(\frac{x_i-x_k}{a}\right)nbf(x_i)(1+o(1))$$
$$+\sum_{i=1}^{g}\sum_{j=1}^{g}\sum_{k=1}^{g}(x_j-x_k)(K_2^*)'\left(\frac{x_j-x_k}{a}\right)K_2^*\left(\frac{x_i-x_k}{a}\right)n(n-1)b^2f(x_i)f(x_j)(1+o(1)).$$

Therefore,

$$-\frac{n^2a^8}{8b}E(B_n)$$
$$\approx \frac{n}{b}\int\int(x-y)(K_2^*)'\left(\frac{x-y}{a}\right)K_2^*\left(\frac{x-y}{a}\right)f(x)dxdy$$
$$+\frac{n(n-1)}{b}\int\int\int(x-z)(K_2^*)'\left(\frac{x-z}{a}\right)K_2^*\left(\frac{y-z}{a}\right)f(x)f(y)dxdydz. \qquad (4.53)$$

Combining this with (4.46) and (4.47),

$$E(B_n)=\frac{-8b}{n^2a^8}\left[\frac{-3n^2a^7}{4b}R(f'')\left(1+O(a^2)\right)+O\left(\frac{na^2}{b}\right)\right]$$
$$=\frac{6}{a}R(f'')+O(a)+O\left(\frac{1}{na^6}\right).$$

As for the variance of B_n, write

$$\pi(x_i,x_j)=\sum_{k=1}^{g}\frac{(x_j-s)}{a}(K_2^*)'\left(\frac{x_j-s}{a}\right)K_2^*\left(\frac{x_i-s}{a}\right)b,$$

for every $1\le i,j\le g$. Then

$$E\left(B_n^2\right)=\frac{64}{n^4a^{14}}E\left(\sum_{i=1}^{g}\sum_{j=1}^{g}\pi(x_i,x_j)c_ic_j\right)^2$$
$$=\frac{64}{n^4a^{14}}\sum_{i=1}^{g}\sum_{j=1}^{g}\sum_{k=1}^{g}\sum_{l=1}^{g}\pi(x_i,x_j)\pi(x_k,x_l)E\left(c_ic_jc_kc_l\right)$$

$$= \frac{64}{n^4 a^{14}} \sum_{i=1}^{g}\sum_{j=1}^{g}\sum_{k=1}^{g}\sum_{l=1}^{g} \varpi(x_i, x_j)\varpi(x_k, x_l) E\left(c_i c_j c_k c_l\right)\left(1 + o(1)\right),$$

applying Lemma 1 to $\pi(x_i, x_j), 1 \leq i, j \leq g$. The terms $E(c_i c_j c_k c_l)$ in the summands can be approximated case by case as in Lemma 3 by appropriate functionals of f. Hence,

$$\frac{n^4 a^{14}}{64} E\left(B_n^2\right)$$

$$\approx \sum\sum\sum\sum_{i,j,k,l \text{ all different}} \frac{n! b^4}{(n-4)!} f(x_i) f(x_j) f(x_k) f(x_l) \varpi(x_i, x_j)\varpi(x_k, x_l)$$

$$+ \sum\sum\sum_{i \neq j \neq k \neq i} \frac{n! b^3}{(n-3)!} f(x_i) f(x_j) f(x_k) \left[1 + b(n-3) f(x_i)\right] \times \left[2\varpi(x_i, x_i)\varpi(x_j, x_k) + \right.$$

$$\left. + \varpi(x_i, x_j)\varpi(x_i, x_k) + 2\varpi(x_i, x_j)\varpi(x_k, x_i) + \varpi(x_j, x_i)\varpi(x_k, x_i)\right]$$

$$+ \sum\sum_{i \neq j} \frac{n! b^2}{(n-2)!} f(x_i) f(x_j) \left[1 + b(n-2)\left\{f(x_i) + f(x_j)\right\} + \frac{b^2 (n-2)!}{(n-4)!} f(x_i) f(x_j)\right] \times$$

$$\left[\varpi(x_i, x_i)\varpi(x_j, x_j) + \varpi(x_i, x_j)\varpi(x_i, x_j) + \varpi(x_i, x_j)\varpi(x_j, x_i)\right]$$

$$+ \sum\sum_{i \neq j} \frac{n! b^2}{(n-2)!} f(x_i) f(x_j) \left[1 + b(n-2)\left\{f(x_i) + f(x_j)\right\} + \frac{b^2 (n-2)!}{(n-4)!} f(x_i) f(x_j)\right] \times$$

$$\left[2\varpi(x_i, x_i)\varpi(x_i, x_j) + 2\varpi(x_i, x_i)\varpi(x_j, x_i)\right]$$

$$+ \sum_{i=1}^{g} n b f(x_i) \left[1 + 7b(n-1) f(x_i) + \frac{6 b^2 (n-1)!}{(n-3)!} f(x_i)^2 + \frac{b^3 (n-1)!}{(n-4)!} f(x_i)^3\right] \varpi(x_i, x_i)^2.$$

Therefore we can write

$$\frac{n^4 a^{14}}{64} E\left(B_n^2\right) \approx \frac{n! b^4}{(n-4)!} \sum_{i=1}^{g}\sum_{j=1}^{g}\sum_{k=1}^{g}\sum_{l=1}^{g} f(x_i) f(x_j) f(x_k) f(x_l) \varpi(x_i, x_j)\varpi(x_k, x_l)$$

$$+ \frac{n! b^3}{(n-3)!} \sum_{i=1}^{g}\sum_{j=1}^{g}\sum_{k=1}^{g} f(x_i) f(x_j) f(x_k) \left[2\varpi(x_i, x_i)\varpi(x_j, x_k) + \varpi(x_i, x_j)\varpi(x_i, x_k)\right.$$

$$\left. + 2\varpi(x_i, x_j)\varpi(x_k, x_i) + \varpi(x_j, x_i)\varpi(x_k, x_i)\right]$$

$$+ \frac{n! b^2}{(n-2)!} \sum_{i=1}^{g}\sum_{j=1}^{g} f(x_i) f(x_j) \left[\varpi(x_i, x_i)\varpi(x_j, x_j) + \varpi(x_i, x_j)^2 + \varpi(x_i, x_j)\varpi(x_j, x_i)\right.$$

$$\left. + 2\varpi(x_i, x_i)\varpi(x_i, x_j) + 2\varpi(x_i, x_i)\varpi(x_j, x_i)\right]$$

$$+ \sum_{i=1}^{g} n b f(x_i) \varpi(x_i, x_i)^2.$$

From the above and equations (4.44) - (4.53),

$$\frac{n^4 a^{14}}{64} Var(B_n) = \frac{n^4 a^{14}}{64} \left\{E\left(B_n^2\right) - \left(E B_n\right)^2\right\}$$

$$\approx \left[\frac{n!}{(n-4)!} - n^2(n-1)^2\right]\left[\int\int f(t)f(u)\varpi(t,u)dtdu\right]^2$$

$$+\left[\frac{2n!}{(n-3)!} - 2n^2(n-1)\right]\left[\int f(t)\varpi(t,t)dt\right]\left[\int\int f(u)f(v)\varpi(u,v)dudv\right]$$

$$+\frac{4n!}{(n-3)!}\int\int\int f(t)f(u)f(v)\varpi(t,u)\varpi(t,v)dtdudv - n\left[\int f(t)\varpi(t,t)dt\right]^2$$

$$+\frac{2n!}{(n-2)!}\int\int\int f(t)f(u)\left[\varpi(t,u)^2 + 2\varpi(t,t)\varpi(t,u)\right]dtdu + n\int f(t)\varpi(t,t)^2dt$$

$$= -\frac{9n^3a^{12}}{4}\left[R(f^{(2)}\right]^2 + O(n^2a^7) + \frac{9n^3a^{12}}{4}\left[\int f\left(f^{(4)}\right)^2\right] + O(na^2)$$

$$+2n^2a^3R(f)\left(\int(G*K_2^*)^2\right) + O(n^2a^7) + O(na^2).$$

Hence, because $a \sim n^{-1/7}$,

$$Var(B_n) = \frac{32}{n^2a^{11}}R(f)\left(\int(G*K_2^*)^2\right) + O\left(\frac{1}{na^2}\right).$$

Proof of Lemma 7:

For every $1 \le l, k \le g$, define

$$\psi_{lk} = \sum_{i=1}^{g} W_2^n\left(\frac{x_i - x_k}{a}\right)I\left(X_l \in A_i\right),$$

with $A_i = [x_i - \frac{b}{2}, x_i + \frac{b}{2}), i = 1, ..., g$, and

$$\mu_k = E\left(\psi_{1k}\right).$$

Then,

$$\widehat{\theta}_{2,2}(a) = \frac{4}{n^2b^2}\sum_{i=1}^{g}\sum_{j=1}^{g}\sum_{k=1}^{g} W_2^n\left(\frac{x_i - x_k}{a}\right)W_2^n\left(\frac{x_j - x_k}{a}\right)c_ic_jb$$

$$= \frac{4}{n^2b^2}\sum_{i=1}^{g}\sum_{j=1}^{g}\sum_{k=1}^{g}\sum_{l=1}^{n}\sum_{m=1}^{n} I\left(X_l \in A_i\right)I\left(X_m \in A_j\right)W_2^n\left(\frac{x_i - x_k}{a}\right)W_2^n\left(\frac{x_j - x_k}{a}\right)c_ic_jb$$

$$= \frac{4}{n^2b^2}\sum_{l=1}^{n}\sum_{m=1}^{n}\left[b\sum_{k=1}^{g}\psi_{lk}\psi_{mk}\right]$$

$$= \frac{4}{n^2b^2}\sum_{l=1}^{n}\sum_{m=1}^{n} b\left[\sum_{k=1}^{g}\left(\psi_{lk} - \mu_k\right)\left(\psi_{mk} - \mu_k\right) + \sum_{k=1}^{g}\mu_k\left(\psi_{lk} + \psi_{mk}\right) - \sum_{k=1}^{g}\mu_k^2\right].$$

The asymptotic normality is obtained by showing

(a)

$$U_n = \sum\sum_{1\le l<m\le n} \left[\sum_{k=1}^{g} b\,(\psi_{lk}-\mu_k)\,(\psi_{mk}-\mu_k)\right] \xrightarrow{D} N\left(0, \frac{n^2b^4}{2a^9}R(f)R(K_2^* * K_2^*)\right),$$

(b)

$$\frac{1}{n^2b}\sum_{l=1}^{n}\sum_{k=1}^{g}(\psi_{lk}-\mu_k)^2 \xrightarrow{p} \frac{1}{na^5}\left(R(K_2^*) + a\int u\,(K_2^*)^2\int f^{(1)}\right)$$

(c)

$$\frac{1}{nb}\sum_{l=1}^{n}\sum_{k=1}^{g}\mu_k\psi_{lk} \xrightarrow{p} \frac{1}{4}R(f^{(2)}) + \frac{a^2}{4!}\left(\int f^{(2)}f^{(4)}\right)\left(\int u^4K_2^*\right),$$

(d)

$$\frac{1}{b}\sum_{k=1}^{g}\mu_k^2 \to \frac{1}{4}R(f^{(2)}) + \frac{a^2}{4!}\left(\int f^{(2)}f^{(4)}\right)\left(\int u^4K_2^*\right),$$

and by the Slutsky Theorem.

(a) Theorem 1 of Hall (1984) is applied to U_n for its asymptotic normality and the notations follow those therein. Therefore,

$$H_n(X_1,X_2) = b\sum_{k=1}^{g}(\psi_{1k}-\mu_k)\,(\psi_{2k}-\mu_k),$$

and

$$G_n(x,y) = E\left\{H_n(X_1,x)H_n(X_1,y)\right\}$$
$$= b^2\sum_{k=1}^{g}\sum_{l=1}^{g}(\psi_{xk}-\mu_k)\,(\psi_{yl}-\mu_l)\,E\left\{(\psi_{1k}-\mu_k)\,(\psi_{1l}-\mu_l)\right\}, \qquad (4.54)$$

where $\psi_{xk} = \sum_{i=1}^{g} W_2^n\left(\frac{x_i-x_k}{a}\right)I\,(x\in A_i)$ for $x \in R$. Hence H_n is symmetric and $E\,(H_n(X_1,X_2)|\,X_2) \equiv 0$ and $U_n = \sum\sum_{1\le l<m\le n} H_n\,(X_l,X_m)$ is a degenerate U-statistics. It remains only to check (2.1) of Hall (1984). This involves first two moments of ψ_{1k}'s:

$$E\psi_{1k} = \mu_k = \sum_{i=1}^{g} W_2^n\left(\frac{x_i-x_k}{a}\right)\int_{A_i} f(t)dt$$
$$\approx \sum_{i=1}^{g} W_2^n\left(\frac{x_i-x_k}{a}\right)bf(x_i) \approx \int W_2^n\left(\frac{t-x_k}{a}\right)f(t)dt. \qquad (4.55)$$

Similarly,

$$E\,(\psi_{1k}\psi_{1l}) \approx \int W_2^n\left(\frac{t-x_k}{a}\right)W_2^n\left(\frac{t-x_l}{a}\right)f(t)dt, \qquad (4.56)$$

and

$$E\left(\psi_{1k}^2\psi_{1l}^2\right) \approx \int \left[W_2^n\left(\frac{t-x_k}{a}\right) W_2^n\left(\frac{t-x_l}{a}\right)\right]^2 f(t)dt. \tag{4.57}$$

From (4.55) - (4.56),

$$E\left(H_n(X_1,X_2)\right)^2 = b^2 E\left(\sum_{k=1}^{g}\left(\psi_{1k}-\mu_k\right)\left(\psi_{2k}-\mu_k\right)\right)^2$$

$$= b^2 \sum_{k=1}^{g}\sum_{l=1}^{g} E\left(\left(\psi_{1k}-\mu_k\right)\left(\psi_{2k}-\mu_k\right)\left(\psi_{1l}-\mu_l\right)\left(\psi_{2l}-\mu_l\right)\right)$$

$$= b^2 \sum_{k=1}^{g}\sum_{l=1}^{g} \left\{E\left(\psi_{1k}-\mu_k\right)\left(\psi_{1l}-\mu_l\right)\right\}^2$$

$$\approx b^2 \sum_{k=1}^{g}\sum_{l=1}^{g} \left\{\int W_2^n\left(\frac{t-x_k}{a}\right) W_2^n\left(\frac{t-x_l}{a}\right) f(t)dt\right.$$

$$\left. - \int W_2^n\left(\frac{t-x_k}{a}\right) f(t)dt \int W_2^n\left(\frac{t-x_l}{a}\right) f(t)dt\right\}^2$$

$$\approx \int\cdots\int W_2^n\left(\frac{t-r}{a}\right) W_2^n\left(\frac{t-s}{a}\right) W_2^n\left(\frac{u-r}{a}\right) W_2^n\left(\frac{u-s}{a}\right) f(t)f(u)dtdudrds$$

$$-2\int\cdots\int W_2^n\left(\frac{t-r}{a}\right) W_2^n\left(\frac{t-s}{a}\right) W_2^n\left(\frac{u-r}{a}\right) W_2^n\left(\frac{v-s}{a}\right) f(t)f(u)f(v)dtdudvdrds$$

$$+\left[\int\int\int W_2^n\left(\frac{t-r}{a}\right) W_2^n\left(\frac{u-r}{a}\right) f(t)f(u)dtdudr\right]^2$$

Define $\omega : R^2 \to R$ as

$$\omega(t,u) = \int K_2^*\left(\frac{t-x}{a}\right) K_2^*\left(\frac{u-x}{a}\right) dx. \tag{4.58}$$

Then, by (4.33),

$$E\left(H_n(X_1,X_2)\right)^2 \approx \frac{b^4}{a^{12}} \int\int \omega(t,u)^2 f(t)f(u)dtdu$$

$$-\frac{2b^4}{a^{12}}\int\int\int \omega(t,u)\omega(t,v)f(t)f(u)f(v)dtdu + \frac{b^4}{a^{12}}\left[\int\int \omega(t,u)f(t)f(u)dtdu\right]^2.$$

Then, from Lemma 4,

$$E\left(H_n(X_1,X_2)\right)^2 = \frac{b^4}{a^9} R(f) R\left(K_2^* * K_2^*\right)\left(1+o(1)\right). \tag{4.59}$$

Next,

$$E\left(H_n^4(X_1,X_2)\right) = b^4 E\left(\sum_{k=1}^{g}\left(\psi_{1k}-\mu_k\right)\left(\psi_{2k}-\mu_k\right)\right)^4$$

$$= b^4 \sum_{i=1}^{g}\sum_{j=1}^{g}\sum_{k=1}^{g}\sum_{l=1}^{g} E\left((\psi_{1i}-\mu_i)(\psi_{1j}-\mu_j)(\psi_{1k}-\mu_k)(\psi_{1l}-\mu_l)\right) \times$$

$$E\left((\psi_{2i}-\mu_i)(\psi_{2j}-\mu_j)(\psi_{2k}-\mu_k)(\psi_{2l}-\mu_l)\right)$$

$$= b^4 \sum_{i=1}^{g}\sum_{j=1}^{g}\sum_{k=1}^{g}\sum_{l=1}^{g} E\left((\psi_{1i}-\mu_i)(\psi_{1j}-\mu_j)(\psi_{1k}-\mu_k)(\psi_{1l}-\mu_l)\right)^2.$$

By the Cauchy-Schwartz inequality,

$$E\left(H_n^4(X_1,X_2)\right)$$

$$\le b^4 \sum_{i=1}^{g}\sum_{j=1}^{g}\sum_{k=1}^{g}\sum_{l=1}^{g} E\left((\psi_{1i}-\mu_i)(\psi_{1j}-\mu_j)\right)^2 E\left((\psi_{1k}-\mu_k)(\psi_{1l}-\mu_l)\right)^2$$

$$= b^4 \left[\sum_{k=1}^{g}\sum_{l=1}^{g} E\left((\psi_{1k}-\mu_k)(\psi_{1l}-\mu_l)\right)^2\right]^2. \tag{4.60}$$

But,

$$E\left((\psi_{1k}-\mu_k)(\psi_{1l}-\mu_l)\right)^2 = E\left(\psi_{1k}^2\psi_{1l}^2\right) - \mu_l^2 E\left(\psi_{1k}^2\right) - \mu_k^2 E\left(\psi_{1l}^2\right) + \mu_k^2\mu_l^2. \tag{4.61}$$

Combining (4.60) and (4.61),

$$\left\{E\left(H_n^4(X_1,X_2)\right)\right\}^{1/2}$$

$$\le b^2 \sum_{k=1}^{g}\sum_{l=1}^{g} E\left(\psi_{1k}^2\psi_{1l}^2\right) - 2b^2\left(\sum_{k=1}^{g}\mu_k^2\right)\left(\sum_{l=1}^{g} E\left(\psi_{1l}^2\right)\right) + b^2\left(\sum_{k=1}^{g}\mu_k^2\right)^2.$$

Using (4.55) - (4.57), the above quantity approximately equals

$$b^2 \sum_{k=1}^{g}\sum_{l=1}^{g} \int \left[W_2^n\left(\frac{t-x_k}{a}\right) W_2^n\left(\frac{t-x_l}{a}\right)\right]^2 f(t)dt$$

$$-2b^2\left(\sum_{k=1}^{g} \int\int W_2^n\left(\frac{t-x_k}{a}\right) W_2^n\left(\frac{u-x_k}{a}\right) f(t)f(u)dtdu\right)\left(\sum_{l=1}^{g}\int\left[W_2^n\left(\frac{t-x_l}{a}\right)\right]^2 f(t)dt\right)$$

$$+b^2\left(\sum_{k=1}^{g}\int\int W_2^n\left(\frac{t-x_k}{a}\right) W_2^n\left(\frac{u-x_k}{a}\right) f(t)f(u)dtdu\right)^2$$

$$\approx \int\int\int\left[W_2^n\left(\frac{t-r}{a}\right) W_2^n\left(\frac{t-s}{a}\right)\right]^2 f(t)dtdrds$$

$$-2\left(\int\int\int W_2^n\left(\frac{t-r}{a}\right) W_2^n\left(\frac{u-r}{a}\right) f(t)f(u)dtdudr\right)\left(\int\int\left[W_2^n\left(\frac{t-r}{a}\right)\right]^2 f(t)dtdr\right)$$

$$+\left(\int\int W_2^n\left(\frac{t-r}{a}\right) W_2^n\left(\frac{u-r}{a}\right) f(t)f(u)dtdudr\right)^2$$

$$= \frac{b^4}{a^{12}} \left\{ \int \omega(t,t)^2 f(t) dt - 2 \int\int \omega(t,u) f(t) f(u) dt du \times \int \omega(t,t) f(t) dt \right.$$

$$\left. \left(\int\int \omega(t,u) f(t) f(u) dt du \right)^2 \right\} (1 + o(1)),$$

from (4.33) and with the notation (4.58). Hence, by Lemma 4,

$$E\left(H_n^4(X_1, X_2)\right) \leq \frac{b^8}{a^{20}} \left(R(K_2^*)\right)^4 (1 + o(1)). \tag{4.62}$$

Next, observing from (4.54),

$$G_n(X_1, X_2) = b^2 \sum_{k=1}^{g}\sum_{l=1}^{g} (\psi_{1k} - \mu_k)(\psi_{2l} - \mu_l) E\{(\psi_{1k} - \mu_k)(\psi_{1l} - \mu_l)\}.$$

Therefore,

$$E\left\{G_n^2(X_1, X_2)\right\} = b^4 \sum_{i=1}^{g}\sum_{j=1}^{g}\sum_{k=1}^{g}\sum_{l=1}^{g} E\{(\psi_{1i} - \mu_i)(\psi_{2j} - \mu_j)(\psi_{1k} - \mu_k)(\psi_{2l} - \mu_l)\}$$

$$\times E\{(\psi_{1i} - \mu_i)(\psi_{1j} - \mu_j)\} E\{(\psi_{1k} - \mu_k)(\psi_{1l} - \mu_l)\}$$

$$= b^4 \sum_{i=1}^{g}\sum_{j=1}^{g}\sum_{k=1}^{g}\sum_{l=1}^{g} E(\psi_{1i}\psi_{1k}) E(\psi_{1j}\psi_{1l}) E(\psi_{1i}\psi_{1j}) E(\psi_{1k}\psi_{1l})$$

$$-4b^4 \sum_{i=1}^{g}\sum_{j=1}^{g}\sum_{k=1}^{g}\sum_{l=1}^{g} \mu_i\mu_k E(\psi_{1j}\psi_{1l}) E(\psi_{1i}\psi_{1j}) E(\psi_{1k}\psi_{1l})$$

$$+2b^4 \left(\sum_{i=1}^{g}\sum_{j=1}^{g} \mu_i\mu_j E(\psi_{1i}\psi_{1j})\right)^2 + 4b^4 \sum_{i=1}^{g}\sum_{j=1}^{g}\sum_{k=1}^{g}\sum_{l=1}^{g} \mu_i^2\mu_j\mu_k E(\psi_{1j}\psi_{1l}) E(\psi_{1k}\psi_{1l})$$

$$- 4b^4 \left(\sum_{i=1}^{g}\sum_{j=1}^{g} \mu_i\mu_j E(\psi_{1i}\psi_{1j})\right) \left(\sum_{i=1}^{g} \mu_i^2\right)^2 + b^4 \left(\sum_{i=1}^{g} \mu_i^2\right)^4. \tag{4.63}$$

(i) The first term of (4.63) is approximately

$$\int \cdots \int W_2^n\left(\frac{t_1 - u_1}{a}\right) W_2^n\left(\frac{t_1 - u_3}{a}\right) W_2^n\left(\frac{t_2 - u_2}{a}\right) W_2^n\left(\frac{t_2 - u_4}{a}\right) W_2^n\left(\frac{t_3 - u_1}{a}\right)$$

$$\times W_2^n\left(\frac{t_3 - u_2}{a}\right) W_2^n\left(\frac{t_4 - u_3}{a}\right) W_2^n\left(\frac{t_4 - u_4}{a}\right) \prod_{d=1}^{4} f(u_d) dt_1 \cdots dt_4 du_1 \cdots du_4$$

$$\approx \frac{b^8}{a^{20}} \int \cdots \int (K_2^* * K_2^*)\left(\frac{u_3 - u_1}{a}\right) (K_2^* * K_2^*)\left(\frac{u_2 - u_4}{a}\right) (K_2^* * K_2^*)\left(\frac{u_1 - u_2}{a}\right)$$

$$\times (K_2^* * K_2^*)\left(\frac{u_4 - u_3}{a}\right) \prod_{d=1}^{4} f(u_d) du_1 \cdots du_4.$$

By a change of variables: $\frac{u_3-u_1}{a}=t$, $\frac{u_2-u_4}{a}=u$, $\frac{u_1-u_2}{a}=v$, $\frac{u_4}{a}=s$, the above quantity is

$$\frac{b^8}{a^{16}}\int\cdots\int (K_2^* * K_2^*)(t)\,(K_2^* * K_2^*)(u)\,(K_2^* * K_2^*)(v)\,(K_2^* * K_2^*)(-t-u-v)$$

$$\times f(sa)\,f((s+u)a)\,f((s+u+v)a)\,f((s+u+v+t)a)\,dtdudvds$$

$$=O\left(\frac{b^8}{a^{16}}\right).$$

(ii) Similar to (i) and using the moment conditions (4.40) with Taylor expansions of f, the second term of (4.63) is approximately

$$\frac{b^8}{a^{24}}\int\cdots\int K_2^*\left(\frac{t_1-u_1}{a}\right)K_2^*\left(\frac{t_2-u_3}{a}\right)K_2^*\left(\frac{t_3-u_2}{a}\right)K_2^*\left(\frac{t_3-u_4}{a}\right)K_2^*\left(\frac{t_4-u_1}{a}\right)$$

$$\times K_2^*\left(\frac{t_4-u_2}{a}\right)K_2^*\left(\frac{t_5-u_3}{a}\right)K_2^*\left(\frac{t_5-u_4}{a}\right)\prod_{d=1}^{4} f(u_d)dt_1\cdots dt_5du_1\cdots du_4$$

$$\approx \frac{b^8}{4a^{15}}\int\cdots\int f^{(2)}(u_1)f^{(2)}(u_3)\,(K_2^* * K_2^*)\left(\frac{u_2-u_4}{a}\right)(K_2^* * K_2^*)\left(\frac{u_1-u_2}{a}\right)$$

$$\times (K_2^* * K_2^*)\left(\frac{u_4-u_3}{a}\right)\prod_{d=1}^{4} f(u_d)du_1\cdots du_4$$

$$=O\left(\frac{b^8}{a^{12}}\right).$$

(iii) Similar to (ii), the third term of (4.63) is approximately

$$\frac{b^8}{16a^{10}}\int\cdots\int (K_2^* * K_2^*)\left(\frac{u_1-u_2}{a}\right)(K_2^* * K_2^*)\left(\frac{u_4-u_3}{a}\right)\prod_{d=1}^{4}\left(f(u_d)f^{(2)}(u_d)\right)\prod_{d=1}^{4} du_d$$

$$=O\left(\frac{b^8}{a^8}\right).$$

(iv) Similar to (ii), the fourth term of (4.63) is approximately

$$\frac{b^8}{16a^{10}}\int\cdots\int \left(f^{(2)}(u_1)\right)^2 f^{(2)}(u_2)f^{(2)}(u_3)\,(K_2^* * K_2^*)\left(\frac{u_1-u_2}{a}\right)$$

$$\times (K_2^* * K_2^*)\left(\frac{u_4-u_3}{a}\right)\prod_{d=1}^{4}\left(f(u_d)\right)\prod_{d=1}^{4} du_d$$

$$=O\left(\frac{b^8}{a^8}\right).$$

(v) Similar to (ii), the fifth term of (4.63) is approximately

$$\frac{b^8}{2^6a^4}\left(\int\int f^{(2)}(u)f^{(2)}(v)\,(K_2^* * K_2^*)\left(\frac{u-v}{a}\right)dudv\right)R\left(f^{(2)}\right)^2=O\left(\frac{b^8}{a^3}\right),$$

and the sixth term of (4.63) is approximately

$$\frac{b^8}{a^8} R\left(f^{(2)}\right)^4 .$$

Combining (4.63) and (i) - (v),

$$E\left\{G_n^2(X_1, X_2)\right\} = O\left(\frac{b^8}{a^{16}}\right). \tag{4.64}$$

Concluding from (4.59), (4.62), and (4.64), (2.1) of Hall (1984) holds and hence U_n is asymptotic normal with mean zero and variance $\frac{n^2 b^4}{2a^9} R(f) R\left(K_2^* * K_2^*\right)$.

(b) According to the Strong Law of Large Numbers,

$$\frac{1}{n^2 b} \sum_{l=1}^{n} \sum_{k=1}^{g} (\psi_{lk} - \mu_k)^2 \xrightarrow{p} \frac{1}{nb} E\left\{\sum_{k=1}^{g} (\psi_{lk} - \mu_k)^2\right\}.$$

But,

$$E\left\{b \sum_{k=1}^{g} (\psi_{lk} - \mu_k)^2\right\}$$

$$\approx \int\int \left[W_2^n\left(\frac{t-r}{a}\right)\right]^2 f(t)dtdr - \int\left[\int W_2^n\left(\frac{t-r}{a}\right) f(t)dt\right]^2 dr$$

$$\approx \frac{b^2}{a^6}\left\{\int K_2^*\left(\frac{t-r}{a}\right)^2 f(t)dtdr - \int\left[\int K_2^*\left(\frac{t-r}{a}\right) f(t)dt\right]^2 dr\right\}$$

$$= \frac{b^2}{a^5}\left(R(K_2^*) + a\int u (K_2^*)^2 \int f^{(1)}\right)(1 + o(1)).$$

Hence,

$$\frac{1}{n^2 b} \sum_{l=1}^{n} \sum_{k=1}^{g} (\psi_{lk} - \mu_k)^2 \xrightarrow{p} \frac{1}{na^5} R\left(K_2^* + a\int u (K_2^*)^2 \int f^{(1)}\right),$$

(c) Similar to (b),

$$E\left\{b \sum_{k=1}^{g} \mu_k \psi_{lk}\right\} = b \sum_{k=1}^{g} \mu_k^2$$

$$\approx \frac{b^2}{a^6} \int\int\int K_2^*\left(\frac{t-x}{a}\right) f(t) K_2^*\left(\frac{u-x}{a}\right) f(u)dtdudx$$

$$= \frac{b^2}{a^6}\left\{\frac{a^6}{4} R(f^{(2)}) + \frac{a^8}{4!}\left(\int f^{(2)} f^{(4)}\right)\left(\int u^4 K_2^*\right)\right\}.$$

(d) Already shown in (c).

Proof of Lemma 8:

Observe from the asymptotic mean and variance of B_n in Lemma 5,

$$B_n(a) = \frac{6}{a} R(f^{(2)}) + o_p\left(\frac{1}{a}\right). \tag{4.65}$$

Also note that $A_n(a) = \frac{-6}{a}\widehat{\theta}_{2,2}(a)\,(1 + o_p(1))$ and from Lemma 7,

$$A_n(a) = \frac{-6}{a} R(f^{(2)}) + o_p\left(\frac{1}{a}\right). \tag{4.66}$$

Now,

$$\widehat{\theta}_{2,2}(a_{\widehat{\lambda}}(h)) - \widehat{\theta}_{2,2}(a_\lambda(h)) = \frac{d}{da}\widehat{\theta}_{2,2}(a)\bigg|_{a=a^*} \left(a_{\widehat{\lambda}}(h) - a_\lambda(h)\right),$$

where a^* lies between $a_{\widehat{\lambda}}(h)$ and $a_\lambda(h)$. By virtue of Lemma 5,

$$\widehat{\theta}_{2,2}(a_{\widehat{\lambda}}(h)) - \widehat{\theta}_{2,2}(a_\lambda(h)) = (A_n(a^*) + B_n(a^*)) \left(a_{\widehat{\lambda}}(h) - a_\lambda(h)\right)(1 + o_p(1)).$$

Using (4.66) and (4.65),

$$\widehat{\theta}_{2,2}\left(a_{\widehat{\lambda}}(h)\right) - \widehat{\theta}_{2,2}\left(a_\lambda(h)\right) = o_p\left(\frac{1}{a^*}\right) \times \left(a_{\widehat{\lambda}}(h) - a_\lambda(h)\right)$$

$$= o_p\left(\frac{1}{h^{5/7}}\right) C(K)D(g_1)h^{5/7}\left(\widehat{\lambda}^{2/7} - \lambda^{2/7}\right) = o_p(1)\left[\left(\lambda + O_p(\frac{1}{\sqrt{n}})\right)^{2/7} - \lambda^{2/7}\right]$$

$$= o_p(1)\left[\lambda^{2/7} + \begin{pmatrix} \frac{2}{7} \\ 1 \end{pmatrix} \lambda^{-5/7} O_p(\frac{1}{\sqrt{n}}) - \lambda^{2/7}\right] = o_p(\frac{1}{\sqrt{n}}).$$

Proof of Theorem 8:

Let $\sigma_K^2 = \int u^2 K$ and define the function L_λ as

$$L_\lambda(h) = h\left[\sigma_K^4 \widehat{\theta}_{2,2}(a_\lambda(h))\right]^{1/5} - n^{-1/5} R(K)^{1/5}.$$

Assume K is positive only on $[-1, 1]$. Then it can be shown that, for and fixed sample $x_1, ..., x_n$ from f, $L_{\widehat{\lambda}}(h) \to \infty$ as $h \to \infty$ and $L_{\widehat{\lambda}}(h) < 0$ as $a_{\widehat{\lambda}}(h) \downarrow b$ and

$0 < b \downarrow 0$. Hence $L_{\widehat{\lambda}}(h)$ has roots on the positive real line. Note that $\widehat{h}$ is a root of $L_{\widehat{\lambda}}(h)$. Naturally $\widehat{h} \sim n^{-1/5}$ and

$$0 = L_{\widehat{\lambda}}(\widehat{h}) = h\left[\sigma_K^4 \widehat{\theta}_{2,2}(a_{\widehat{\lambda}}(h))\right]^{1/5} - n^{-1/5} R(K)^{1/5}.$$

And from Lemma 7 and Lemma 8,

$$\left[\widehat{\theta}_{2,2}(a_{\widehat{\lambda}}(h))\right]^{1/5} = \left[\widehat{\theta}_{2,2}(a_\lambda(h)) + o_p(n^{-1/2})\right]^{1/5}$$
$$= \left[\widehat{\theta}_{2,2}(a_\lambda(h))\right]^{1/5} + \frac{1}{5}\left[\widehat{\theta}_{2,2}(a_\lambda(h))\right]^{-4/5} o_p\left(n^{-1/2}\right)$$
$$= \left[\widehat{\theta}_{2,2}(a_\lambda(h))\right]^{1/5} + o_p(n^{-1/2}).$$

Hence

$$L_\lambda(\widehat{h}) = L_{\widehat{\lambda}}(\widehat{h}) + O_p(n^{-7/10}) = O_p(n^{-7/10}). \tag{4.67}$$

Next, by Lemma 7 and the Delta method,

$$n^{\alpha_1} L_\lambda(h_*) \xrightarrow{D} N\left(\mu_1, \sigma_1^2\right), \tag{4.68}$$

where

$$\alpha_1 = \begin{cases} \frac{39}{70} & \text{, if } \int f^{(2)} f^{(4)} < 0 \\ \frac{17}{35} & \text{, if } \int f^{(2)} f^{(4)} > 0, \end{cases}$$

$$\mu_1 = n^{\alpha_1}\left\{h_* \sigma_K^{4/5}\left[R(f^{(2)}) + \frac{4R(K_2^*)}{n a_\lambda(h_*)^5} + \frac{a_\lambda(h_*)^2 \int f^{(2)} f^{(4)} \int u^4 K_2^*}{6} + O\left(\frac{1}{n a_\lambda(h_*)^4}\right)\right]^{1/5} - \left(\frac{R(K)}{n}\right)^{1/5}\right\}$$

$$= n^{\alpha_1}\left\{h_*\left[\sigma_K^4 R(f^{(2)})\right]^{1/5} + h_* \sigma_K^{4/5} R(f^{(2)})^{-4/5} \frac{a_\lambda(h_*)^2 \int f^{(2)} f^{(4)} \int u^4 K_2^*}{30}\left[1 + \frac{1}{\chi}\right] + O\left(\frac{h_*}{n a_\lambda(h_*)^4}\right) - \left(\frac{R(K)}{n}\right)^{1/5}\right\}$$

$$= \begin{cases} O\left(n^{-1/14}\right) & \text{, if } \int f^{(2)} f^{(4)} < 0 \\ \frac{7}{150} \sigma_K^{-8/7} R(K)^{17/35} R(f^{(2)})^{-9/7} C(K)^2 D(g_\lambda)^2 \int f^{(2)} f^{(4)} \int u^4 K_2^* & \text{, if } \int f^{(2)} f^{(4)} > 0, \end{cases}$$

and

$$\sigma_1^2 = n^{2\alpha_1} h_*^2 \sigma_K^{8/5}\left[\frac{1}{5} R(f^{(2)})^{-4/5}\right]^2 \frac{32 R(f) R(K_2^* * K_2^*)}{n^2 a_\lambda(h_*)^9}$$

$$= \frac{32}{25} n^{2\alpha_1 - 35/39} \sigma_K^{118/35} R(K_2^* * K_2^*) R(K)^{-31/35} R(f^{(2)})^{-5/7} R(f) C(K)^{-9} D(g_\lambda)^{-9}.$$

Furthermore,

$$\frac{d}{dh} L_\lambda(h) = \left[\sigma_K^4 \widehat{\theta}_{2,2}(a_\lambda(h))\right]^{1/5}$$

$$+ h\sigma_K^{4/5} \frac{1}{5} \left[A_n(a_\lambda(h)) + B_n(a_\lambda(h))\right] C(K) D(g_\lambda) \frac{5}{7} h^{-2/7}$$

$$= \left[\sigma_K^4 R(f^{(2)})\right]^{1/5} + o_p(1), \tag{4.69}$$

by Lemma 7, (4.65) and (4.66). Now,

$$L_\lambda(\widehat{h}) = L_\lambda(h_*) + \frac{d}{dh} L_\lambda(h^{**})(\widehat{h} - h_*), \tag{4.70}$$

where h^{**} lies in between $\widehat{h}$ and h_*. Combining (4.67) - (4.70),

$$n^\alpha \left(\frac{\widehat{h} - h_*}{h_*}\right) = n^\alpha \left(\frac{L_\lambda(\widehat{h}) - L_\lambda(h_*)}{h_* \frac{d}{dh} L_\lambda(h^{**})}\right)$$

$$= n^\alpha \left(\frac{O_p(n^{-7/10}) - L_\lambda(h_*)}{h_* \left(\left[\sigma_K^4 R(f^{(2)})\right]^{1/5} + o_p(1)\right)}\right)$$

$$= n^\alpha \left(\frac{n}{R(K)}\right)^{1/5} L_\lambda(h_*) \left(-1 + o_p(1)\right) \xrightarrow{D} N(\mu_{PI}, \sigma_{PI}^2).$$

This completes the proof.

Figure 4.1: Small binng range

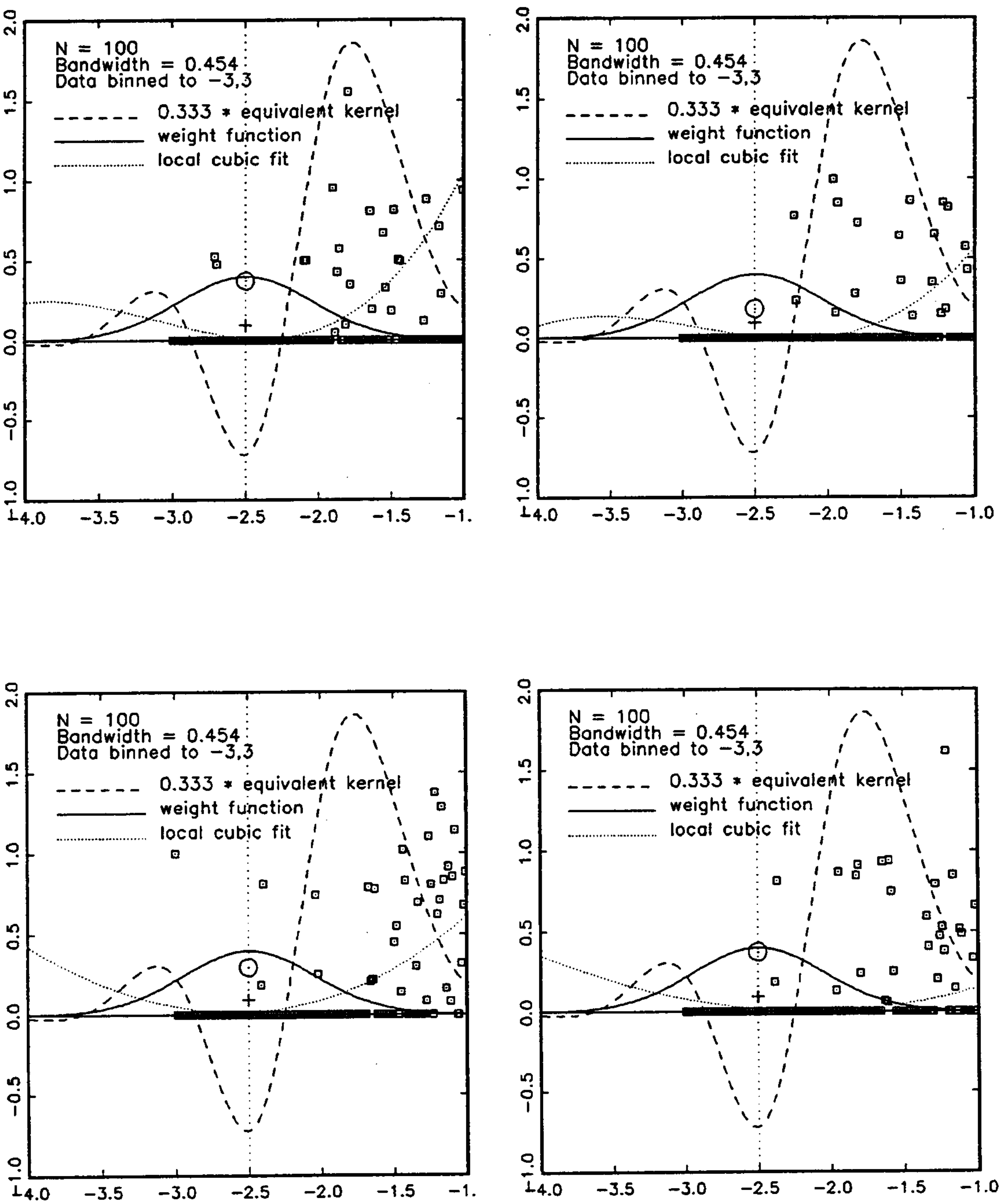


Figure 4.2: Big binning range

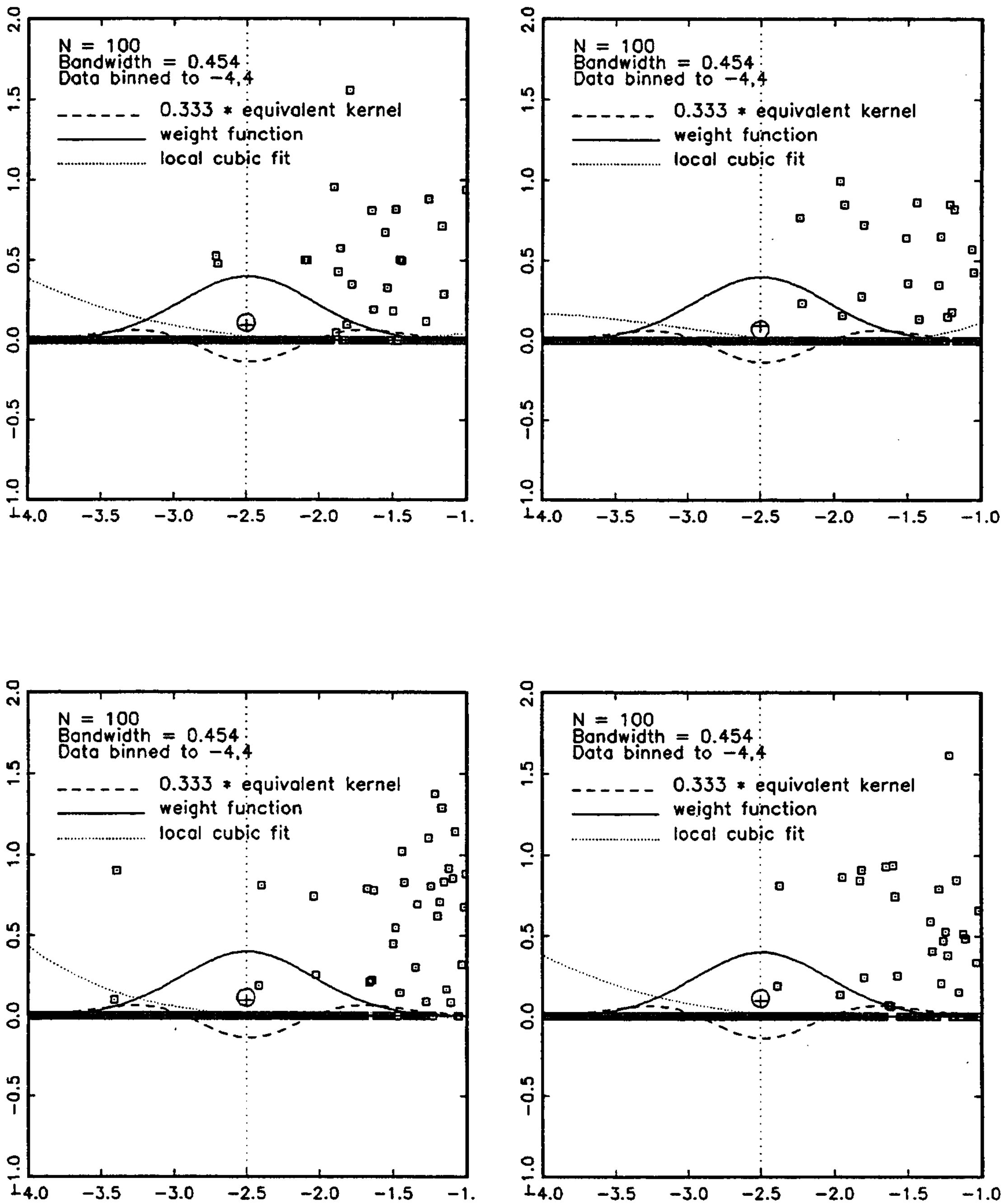
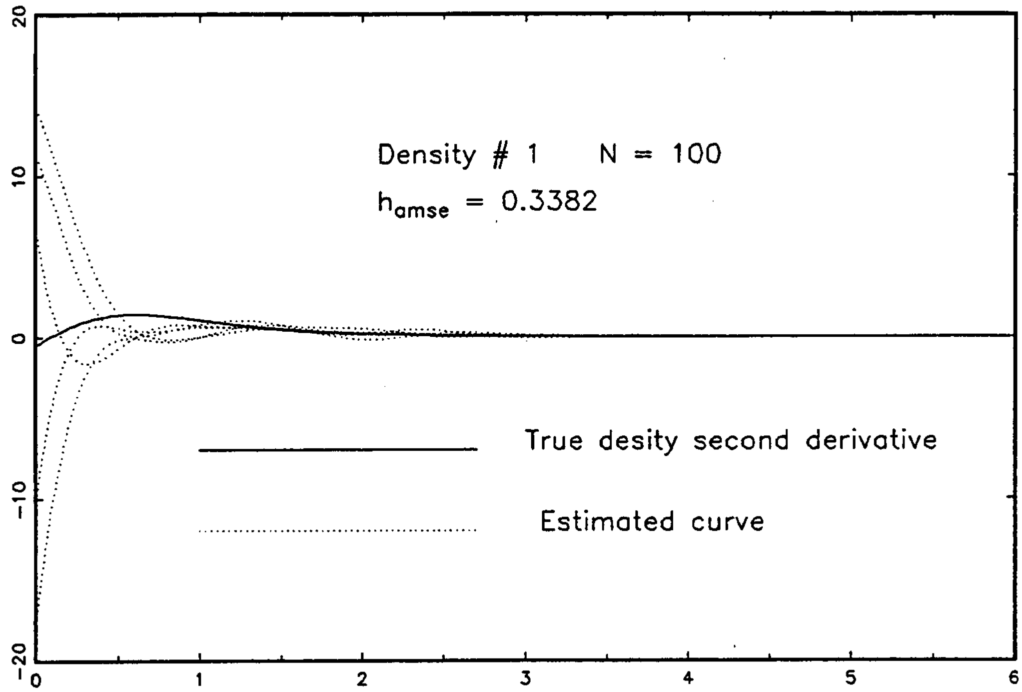


Figure 4.3: Density second derivative estimates

4.3a



4.3b

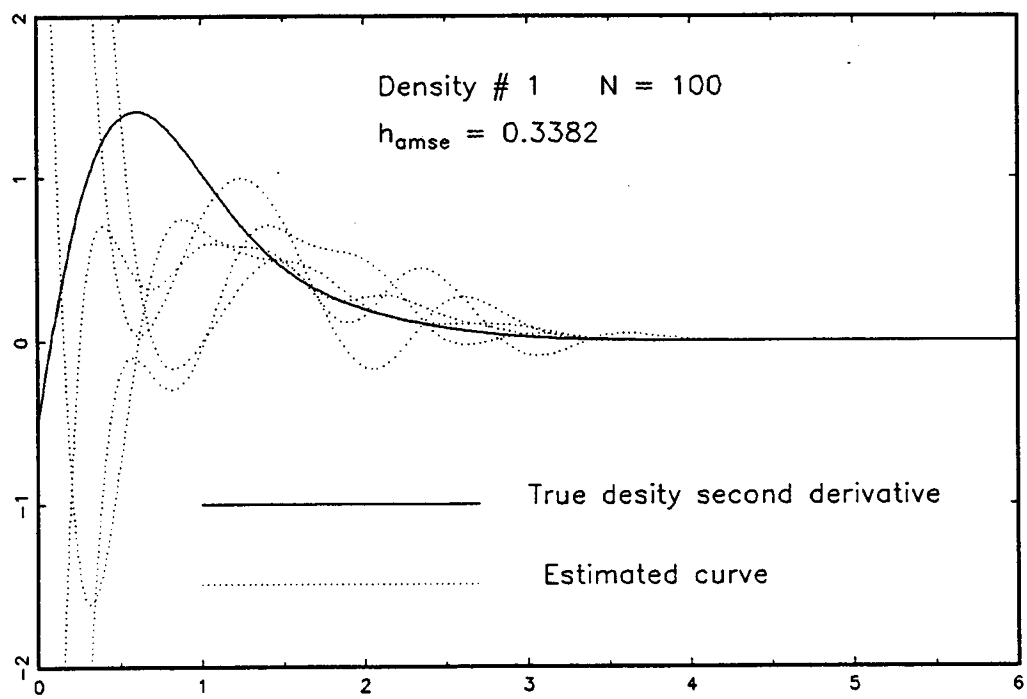
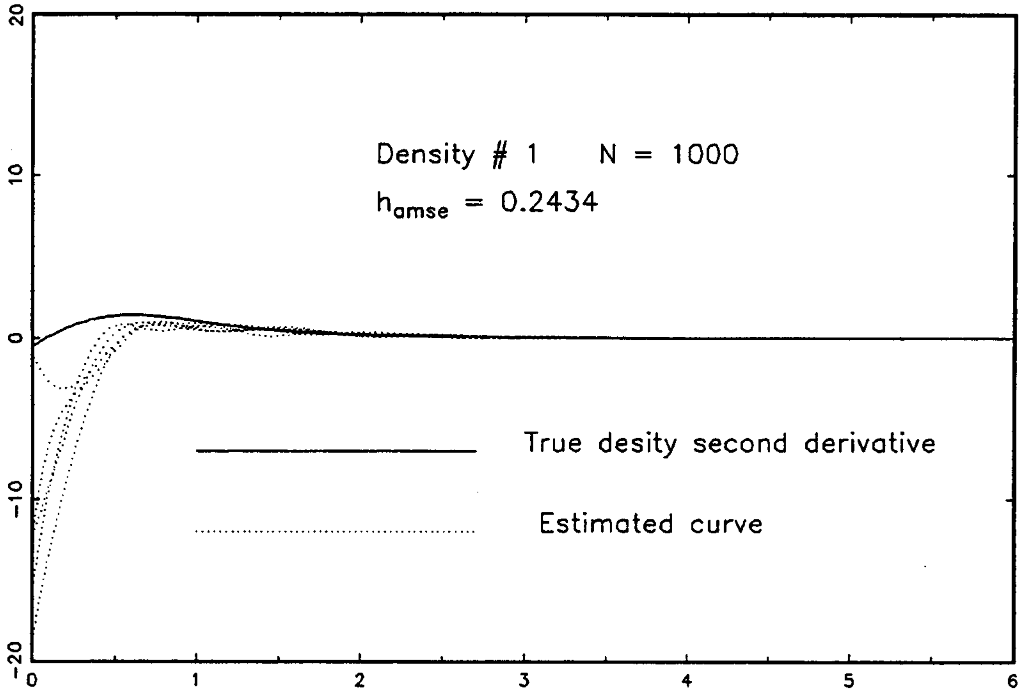


Figure 4.3: Density second derivative estimates

4.3c



4.3d

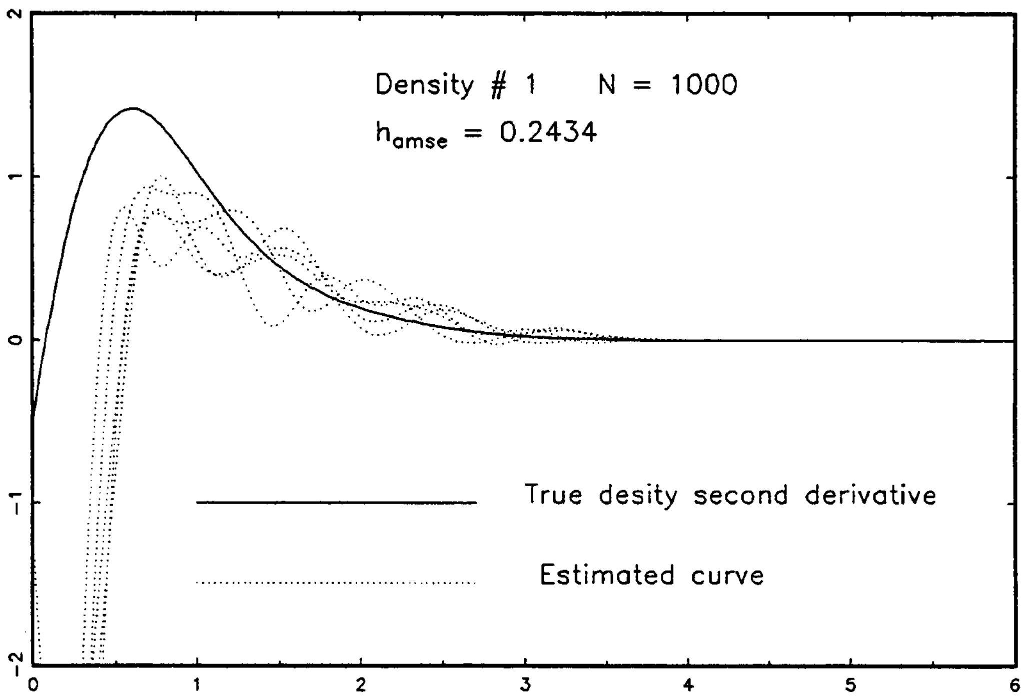
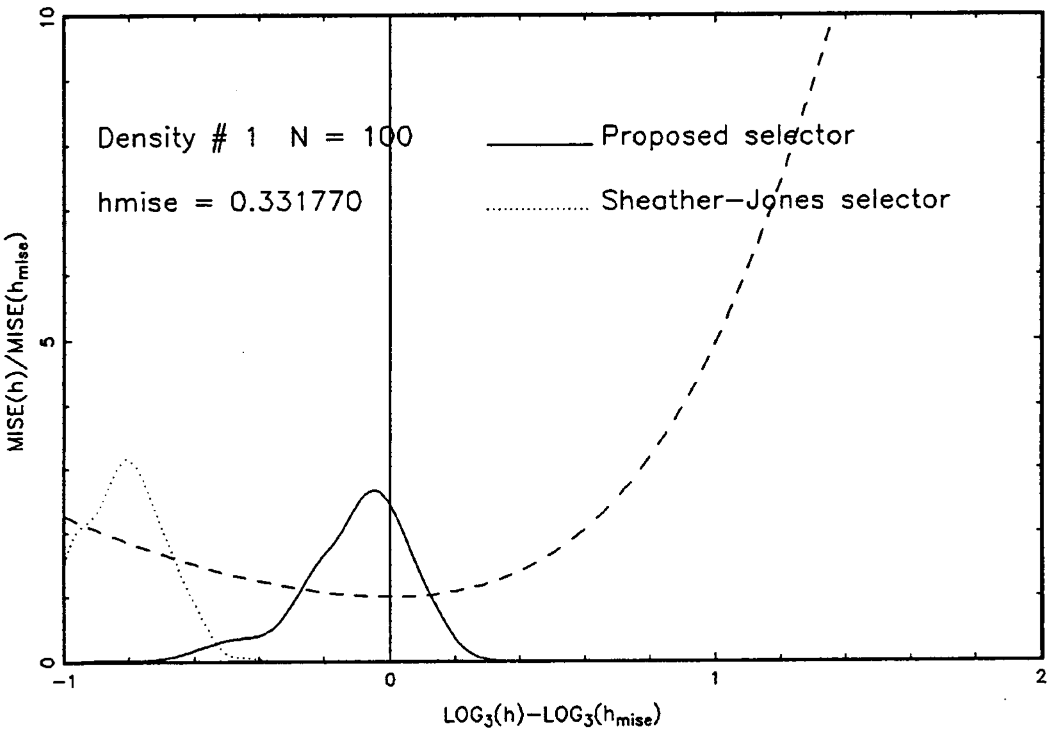


Figure 4.4: Bandwidth seclector populations



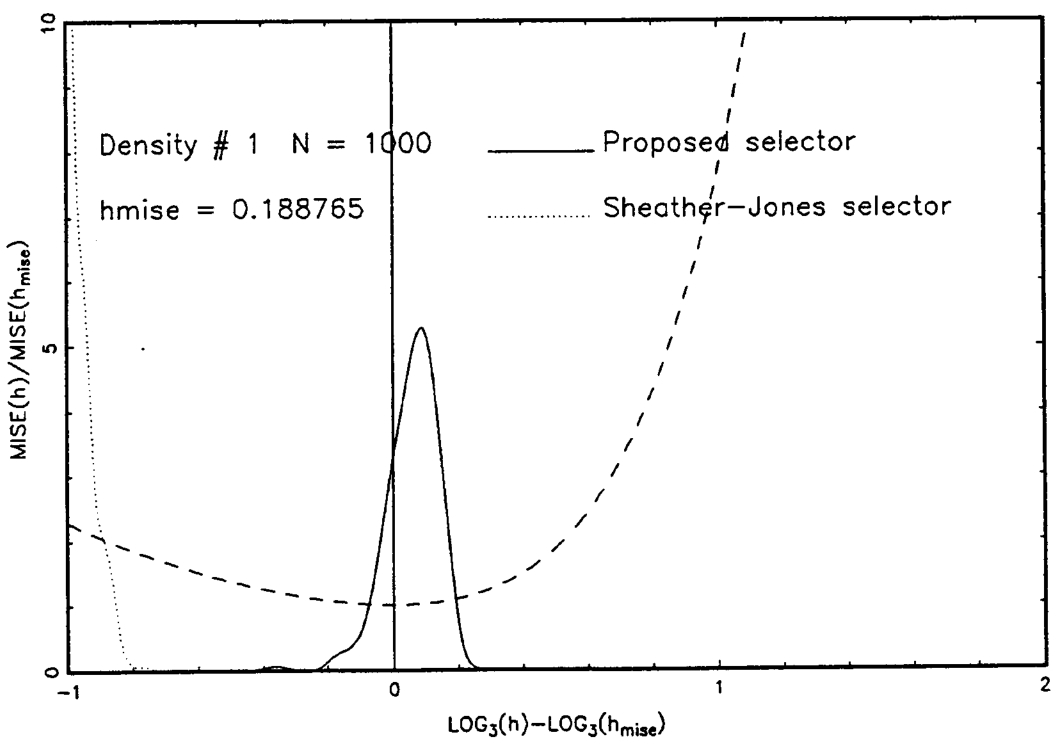
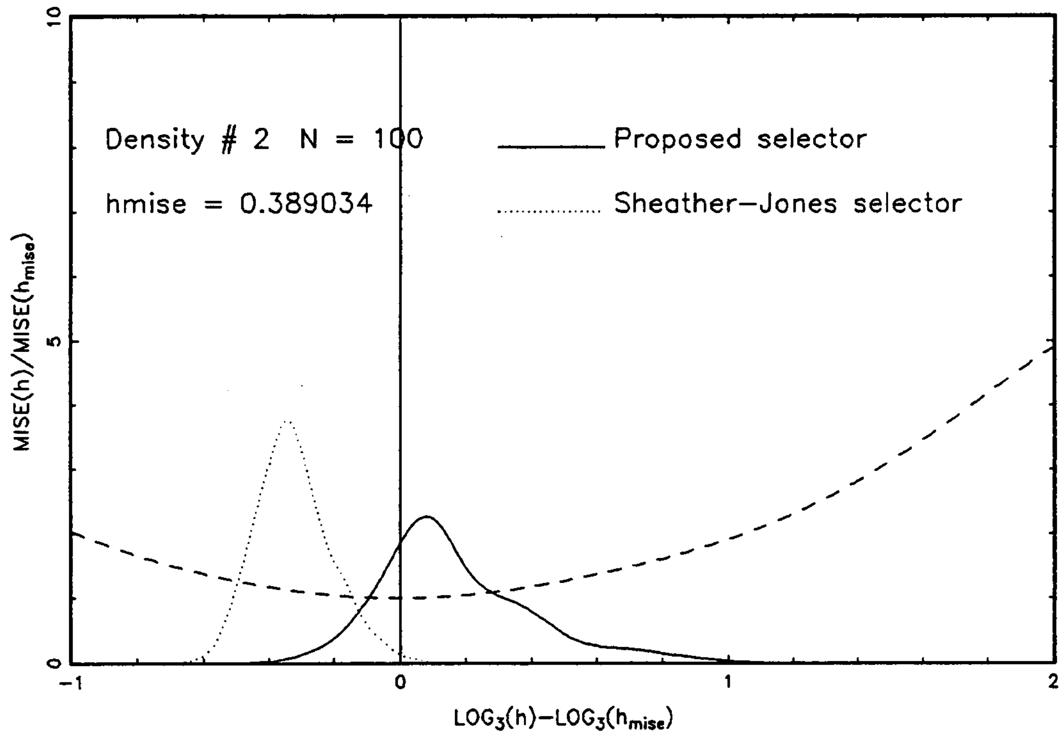


Figure 4.4: Bandwidth seclector populations



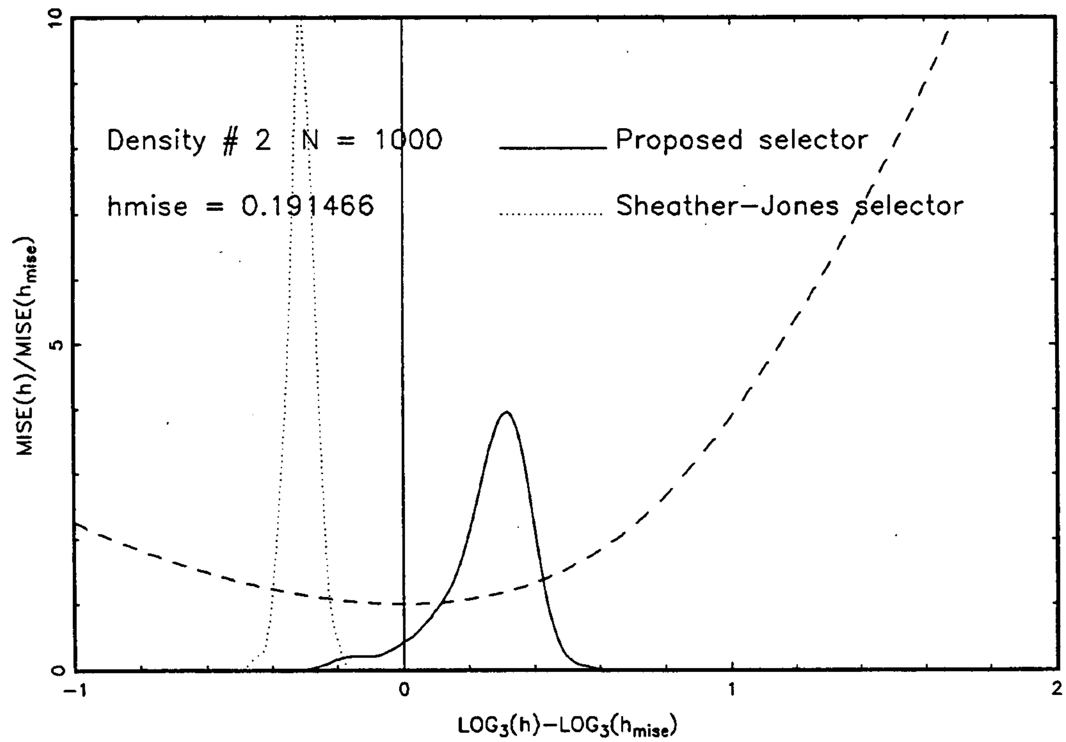
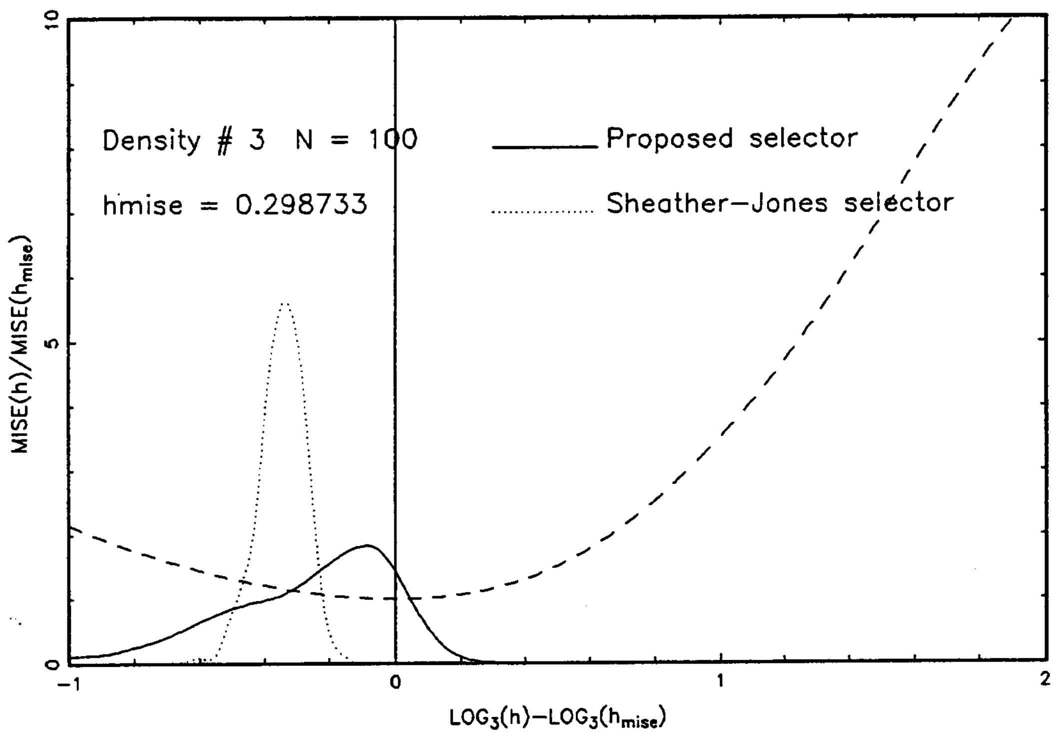


Figure 4.4: Bandwidth seclector populations



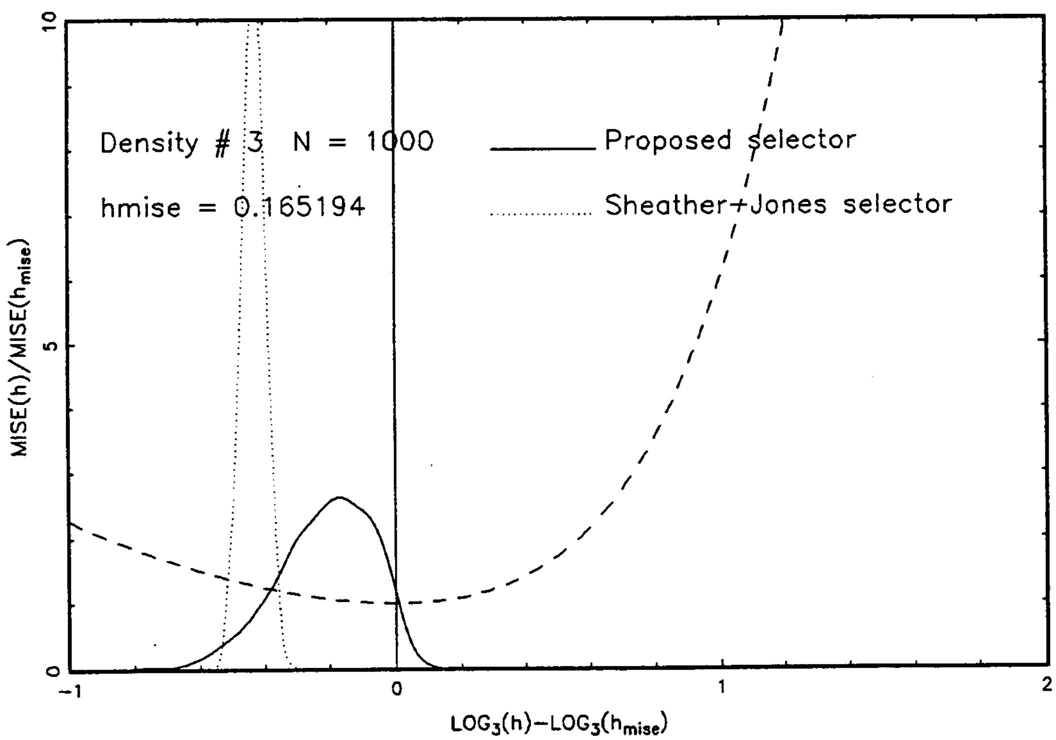
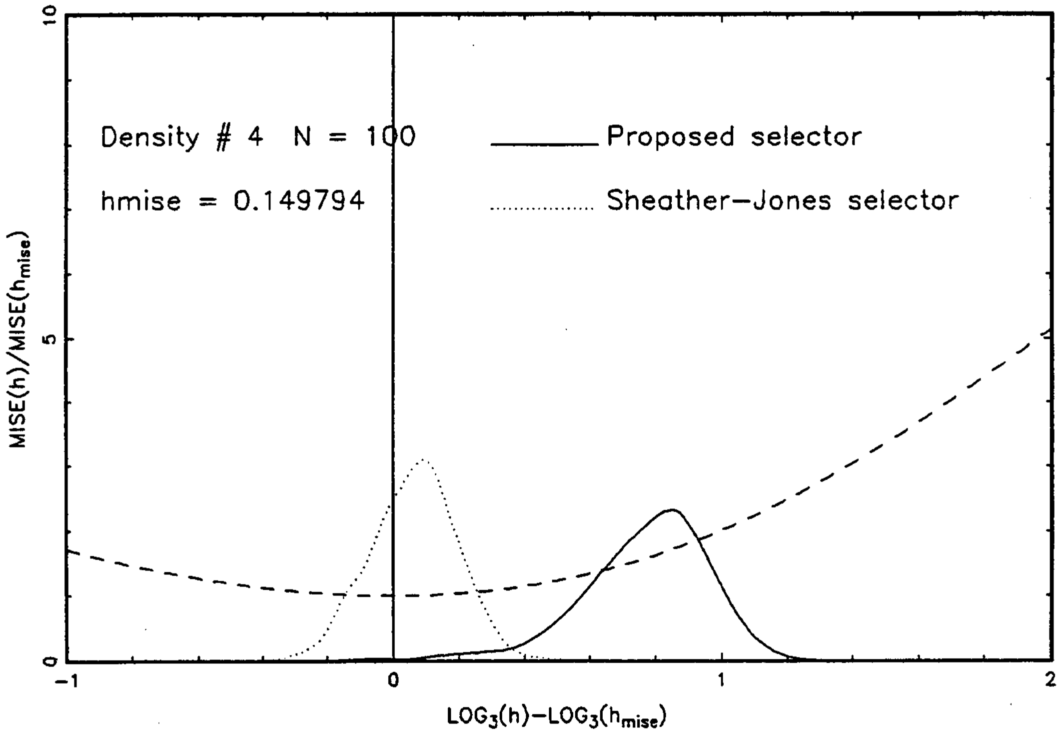


Figure 4.4: Bandwidth seclector populations



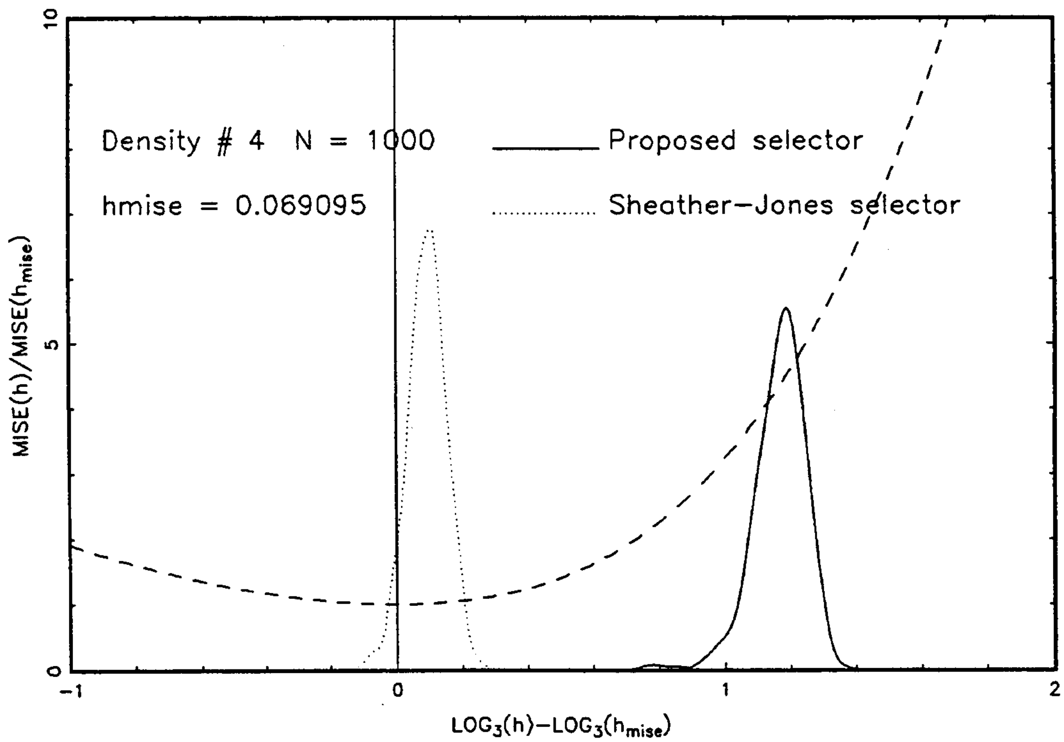
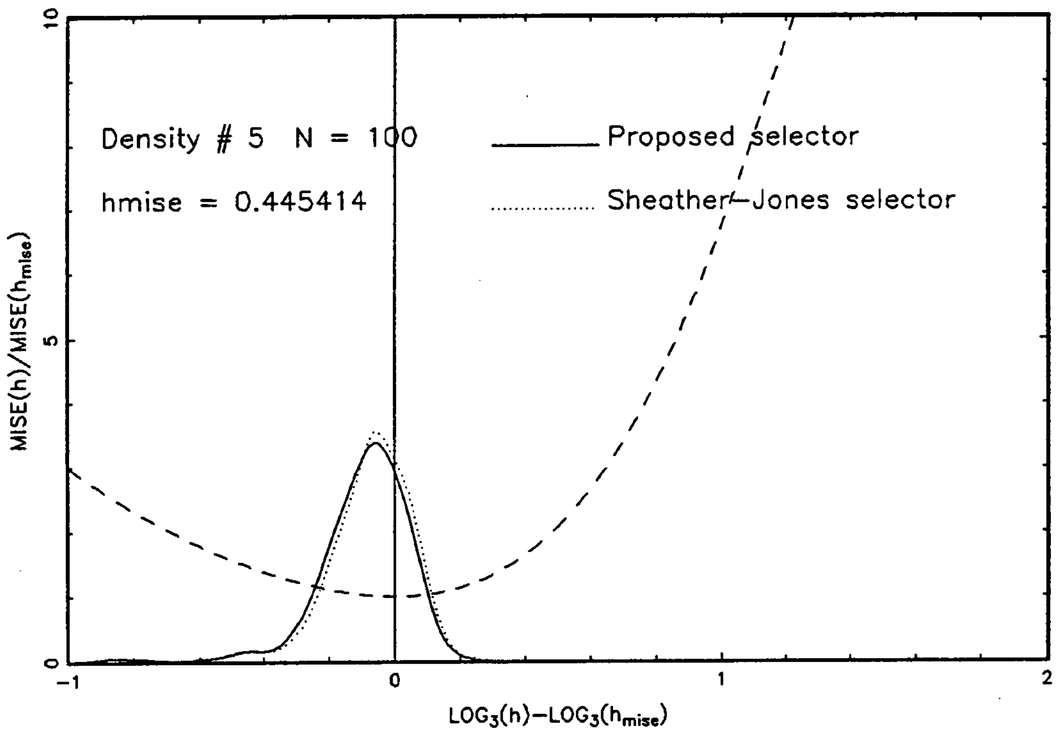


Figure 4.4: Bandwidth seclector populations



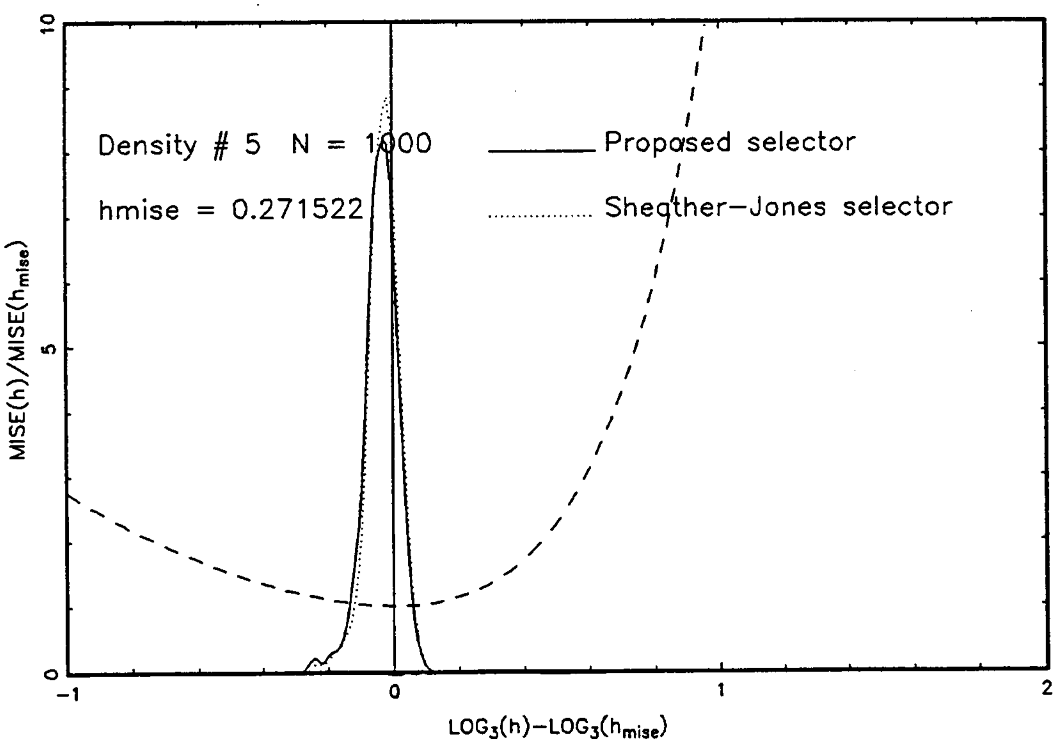


Figure 4.5: Local linear estimators with selected bandwidth

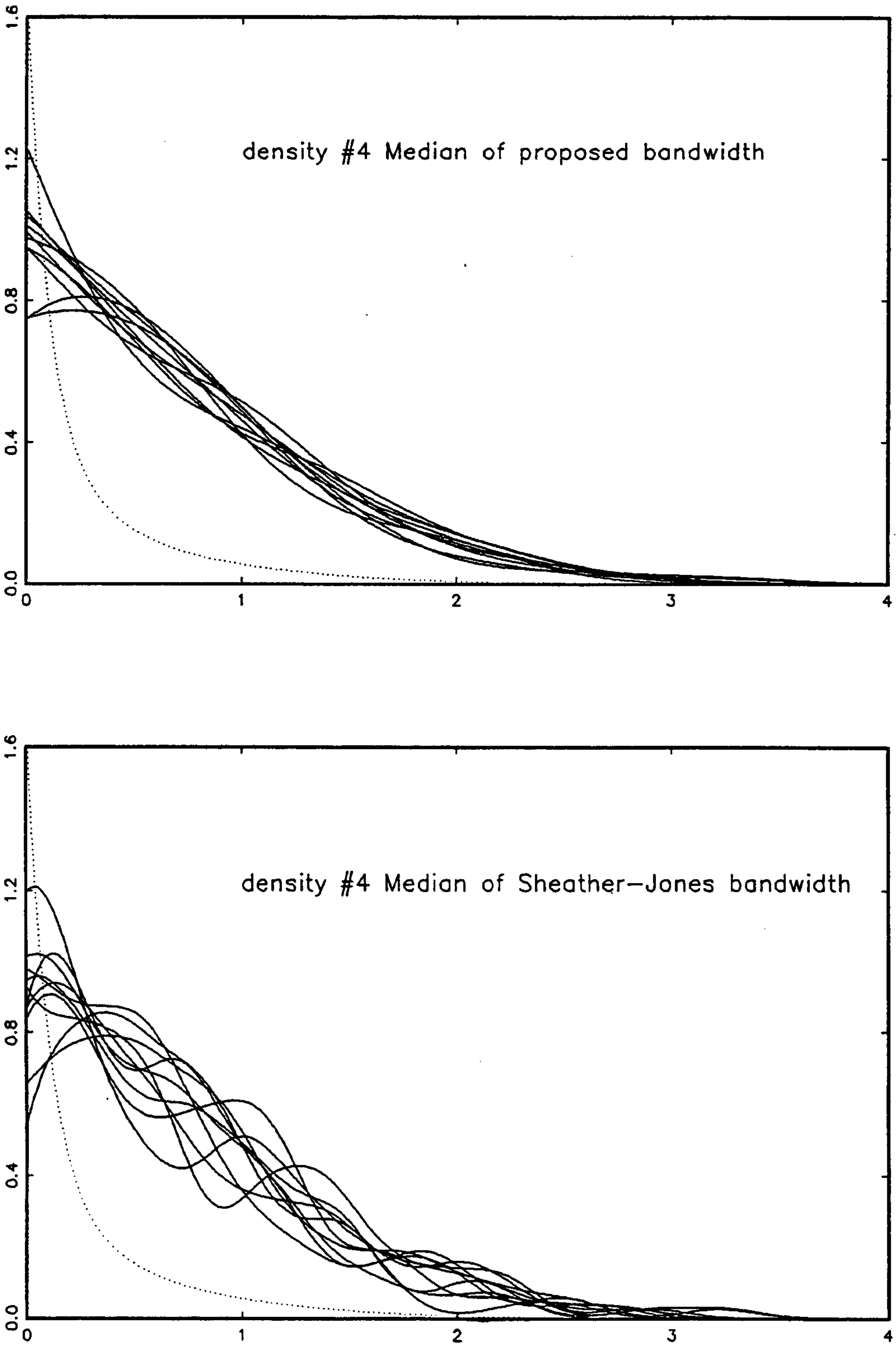
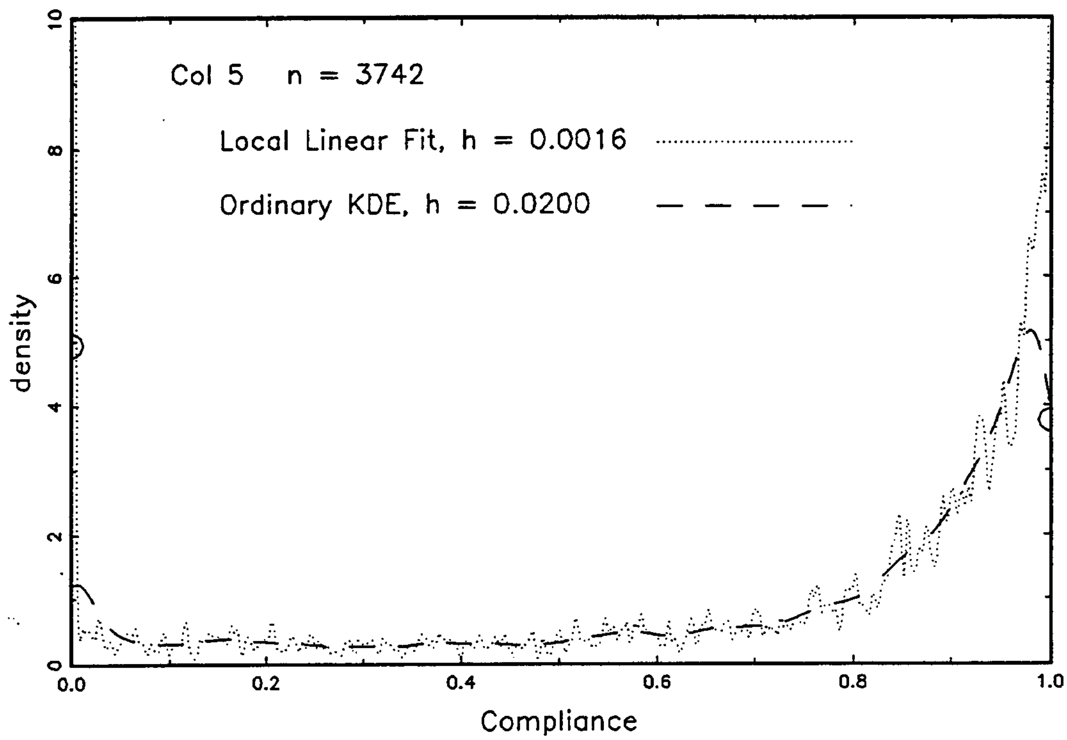


Figure 4.6: Density estimators for compliance data



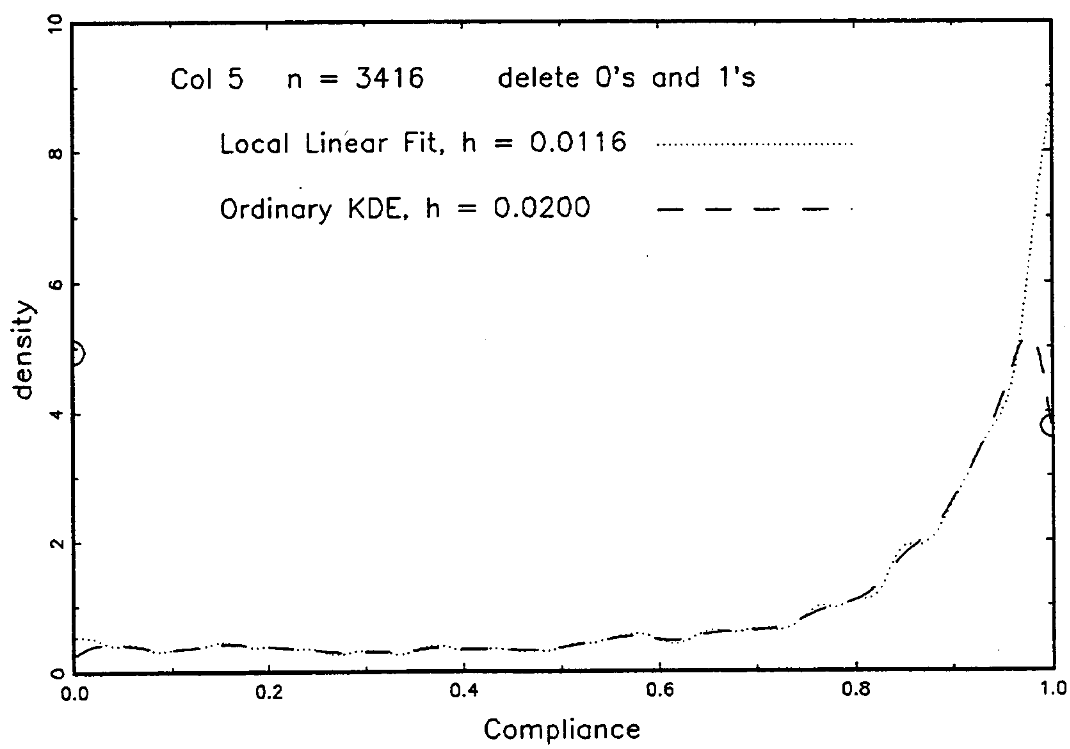
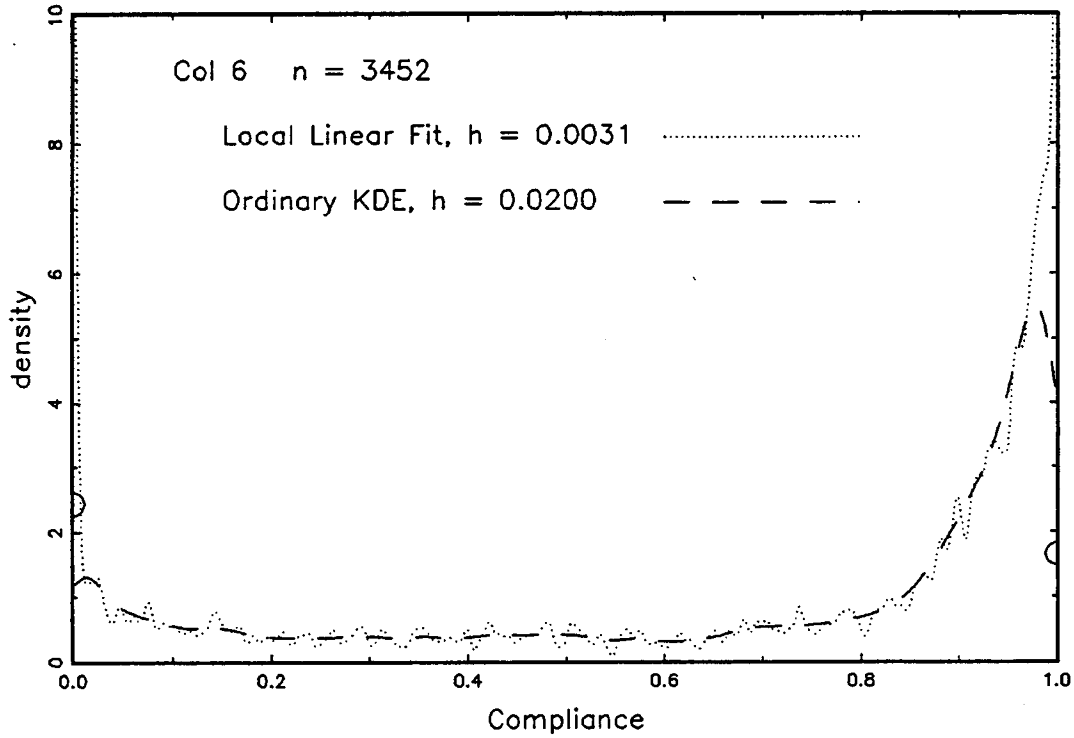


Figure 4.6: Density estimators for compliance data



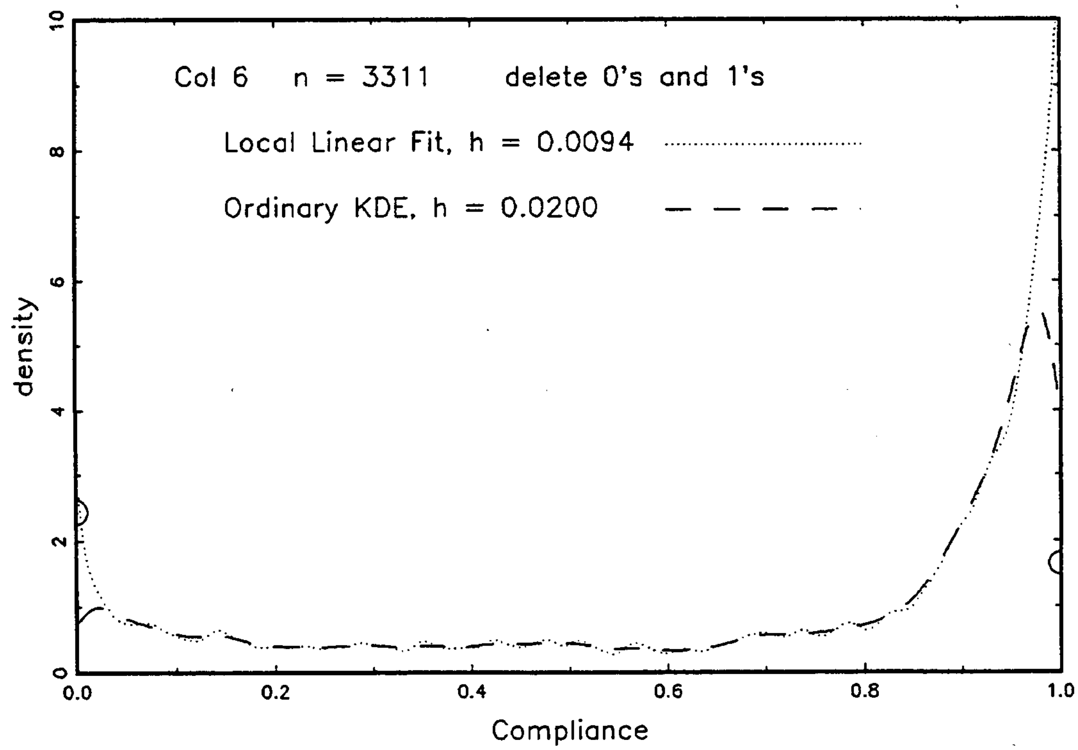
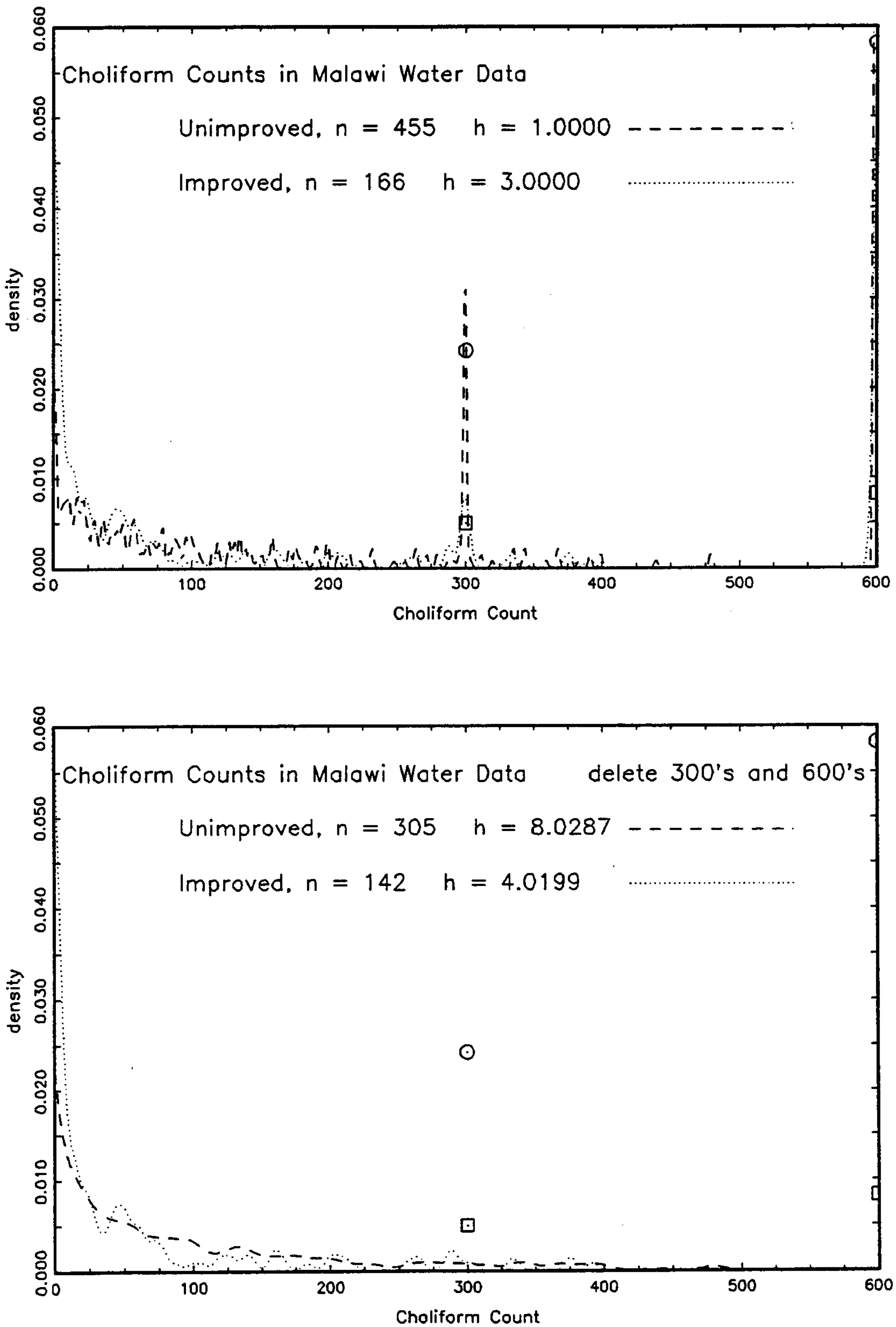


Figure 4.7: Local linear esitmators for choliform counts



Bibliography

[1] Aldershof, B. (1991) Estimation of integrated squared density derivatives. Mimeo Series #2053, Institute of Statistics, University of North Carolina.

[2] Aldershof, B., Marron, J. S., Park, B. U. and Wand, M. P. (1991) List of facts about the normal density. Manuscript.

[3] Bickel, P. J. and Ritov, Y. (1988) Estimating integrated squared density derivatives: sharp best convergence estimates. Sankhyā, Series A, 50, 381-393.

[4] Brown, L. D. and Low, M. G. (1993) Asymptotic equivalence of nonparametric regression and white noise. Ann. Statist., to appear.

[5] Chu, C. K. and Marron, J. S. (1991) Choosing a kernel regression estimator. Statistical Science, 6, No. 4, 404-436.

[6] Cheng, M. Y., Fan, J., and Marron, J. S. (1993) Minimax efficiency of local polynomial fit estimators at boundaries. Mimeo Series #2098, Institute of Statistics, University of North Carolina.

[7] Cleveland, W. S. (1979) Robust locally weighted regression and smoothing scatterplots. J. Amer. Statist. Assoc., 74, 829-836.

[8] Copas, J. B. and Fryer, M. J. (1980) Density estimation and suicide risks in psychiatric treatment. J. Roy. Statist. Soc. A, 143, 167-176.

[9] Djojosugito, R. A. and Speckman, P. L. (1992) Boundary bias correction in nonparametric density estimation. Commun. Statist. - Theory Meth., 21, 69-88.

[10] Donoho, D. L. and Johnstone, I. (1992) Ideal spatial adaptation by wavelet shrinkage. Technical report 400, Stanford University.

[11] Donoho, D. L. and Liu, R. C. (1991) Geometrizing rates of convergence, II. Ann. Statist., 19, No. 2, 633-667.

[12] Donoho, D. L., Liu, R. C., and MacGibbon, B. (1990) Minimax risk over hyper rectangles, implications. Ann. Statist., 18, 1416-1437.

[13] Eubank, R. L. and Speckman, P. (1991) A bias reduction theorem with applications in nonparametric regression. Scand. J. Statist., 211-222.

[14] Efroimovich, S. (1993) On nonparametric regression for i.i.d. observations in general setting. Ann. Statist., to appear.

[15] Fan, J. (1992) Design-adaptive Nonparametric Regression. J. Amer. Statist. Assoc., 87, 998-1004.

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

[17] Fan, J., Gasser, T., Gijbels, I., Brockmann, M., and Engels, J. (1993) Local polynomial fitting: a standard for nonparametric regression. Mimeo Series #2302, Institute of Statistics, University of North Carolina.

[18] Fan, J. and Gijbels, I. (1992a) Variable bandwidth and local linear regression smoothers. Ann. Statist., 20, 2008-2036.

[19] Fan, J. and Gijbels, I. (1992b) Spatial and design adaptation: adaptive order polynomial approximation in function estimation. Mimeo Series #2080, Institute of Statistics, University of North Carolina.

[20] Fan, J. and Gijbels, I. (1995) Data-driven bandwidth selection in local polynomial fitting: variable bandwidth and spatial adaptation. To appear in J. Royal Statist. Soc. B.

[21] Fan, J. and Hall, P. (1994) On curve estimation by minimizing mean absolute deviation and its implications. Ann. Statist., to appear.

[22] Fan, J. and Marron, J. S. (1994) Fast implementation of nonparametric curve estimators. J. Comput. Graphical Statist., to appear.

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

[24] Gasser, T. , Müller, H. G., & Mammitzsch, V. (1985) Kernels for nonparametric curve estimation. J. Royal Statist. Soc. B, 47, No.2, 238-252.

[25] Granovsky, B. L. and Müller, H. G. (1991) Optimizing kernel methods: A unifying variational principle. International Statistical Review, 59, No. 3, 373-388.

[26] Hall, P. (1984) Central limit theorem for integrated squared error of multivariate nonparametric density estimators, J. Multivariate Analysis, 14, 1-16.

[27] Hall, P. and Marron, J. S. (1987) Estimation of integrated squared density derivatives. Stat. Prob. Lett. 6, 109-115.

[28] Hall, P. and Wand, M. P. (1993) On the accuracy of binned kernel density estimators. Manuscript.

[29] Hall, P. and Wehrly, T. E. (1991) A geometrical method for removing edge effects from kernel-type nonparametric regression estimators. J. Amer. Statist. Assoc., 86, 665-672.

[30] Jones, M.C. (1989) Discretized and interpolated kernel density estimates. J. Amer. Stat. Assoc. 84, 733-741.

[31] Jones, M. C., Marron, J. S., and Sheather, S. J. (1994) Progress in data-based bandwidth selection for kernel density estimation. Mimeo Series #2088, Inst. of Statist., University of North Carolina at Chapel Hill.

[32] Jones, M. C. and Sheather, S. J. (1991) Using non-stochastic terms to advantage in kernel-based estimation of integrated squared density derivatives. Statist. & Prob. Lett. 11,511-514.

[33] Lejeune, M. (1985) Estimation non-parametrique par noyaux: regression polynomiale mobile. Revue de Statistiques Appliqueès, 33, No. 3, 43-67.

[34] Lipids Research Clinic Program (1984). "The Lipid Research Clinics Coronary Primary Prevention Trial Results, Parts I and II," Journal of the American Medical Association, 251, 351-374.

[35] Marron, J. S. and Nolan, D. (1989) Canonical kernels for density estimation. Statistics & Probability Letters 7, 195-199.

[36] Marron, J. S. and Ruppert, D. (1993) Transformations to reduce boundary bias in kernel density estimation. To appear J. Roy. Statist. Soc.

[37] Marron, J. S. and Tsybakov, A. B. (1993) Visual error criteria for qualitative smoothing. Manuscript.

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

[39] Marron, J. S. (1992) Graphical understanding of higher order kernels. Mimeo Series #2082, Inst. of Statist., University of North Carolina at Chapel Hill.

[40] Müller, H. G. (1987) Weighted local regression and kernel methods for nonparametric curve fitting. J. Amer. Statist. Assoc., 82, No. 397, 231-238.

[41] Müller, H. G. (1991a) Smooth optimum kernel estimators near endpoints. Biometrika, 78, No. 3, 521-530.

[42] Müller, H. G. (1991b) On the boundary kernel method for nonparametric curve estimation near endpoints. Unpublished manuscript.

[43] Nussbaum, M. (1985) Spline smoothing in regression models and asymptotic efficiency in L_2. Ann. Statist., 13, 984-997.

[44] Park, B. U. and Marron, J. S. (1990) Comparison of data-driven bandwidth selectors. J. Amer. Stat. Assoc., 85, 66-72.

[45] Rice, J. (1984) Boundary modification for kernel regression. Commun. in Statist. A,13, 893-900.

[46] Rice, J. & Rosenblatt, M. (1981) Integrated mean squared error of a smoothing spline. J. Approx. Theory, 33, 353-365.

[47] Ruppert, D., Sheather S. J., and Wand, M. P. (1993) An effective bandwidth selector for local least squares regression. Manuscript.

[48] Ruppert, D. and Wand, M. P. (1993) Multivariate locally weighted least squares regression. To appear in Ann. Statist.

[49] Sheather, S. J. and Jones, M. C. (1991) A reliable data-based bandwidth selection method for kernel density estimation. J. Roy. Statist. Soc. B, 53, 683-690.

[50] Sheather, S. J. (1992) The performance of six popular bandwidth selection methods on some real data sets. Comput. Statist., 7, 225-250.

[51] Schucany, W. R. and Sommers, J. P. (1977) Improvement of kernel type density estimators. J. Amer. Stat. Assoc., 72, 420-423.

[52] Schuster, E. F. (1985) Incorporating support constraints into nonparametric estimators of Densities. Commun. in Statist. A, 14, 1123-1136.

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

[54] Stone, C. J. (1977) Consistent nonparametric regression. Ann. Statist., 5, No. 4, 595-645.

[55] Van Es, A. J. and Hoogstrate, A. J. (1993a) Kernel estimators of integrated squared density derivatives in non-smooth cases. Manuscript.

[56] Van Es, A. J. and Hoogstrate, A. J. (1993b) Properties of plug-in bandwidth selection procedures in non-smooth cases. Manuscript.

[57] Wand, M. P. and Schucany, W. R. (1990) Gaussian-based kernels. Canadian J. of Stat., 18, No.3, 197-204.

[58] Wei, C. Z. and Chu, C. K. (1993) A regression point of view towards density estimation. Manuscript.