Estimating bias and mean squared error of kernel density estimator

Arup Bose [*]
Statistics and Mathematics Unit
Indian Statistical Institute
Kolkata 700108
INDIA
bosearu@gmail.com

Santanu Dutta
Mathematical Sciences Department
Tezpur University
Napaam 784028
INDIA
tezpur1976@gmail.com

July 29, 2009

Abstract

We propose a generalized smooth bootstrap scheme for estimating the bias B_y and mean square error M_y of a kernel density estimator, at y, based on i.i.d data. A number of existing bootstrap schemes are special case of our proposal. For a fairly broad class of kernel and bandwidth h_n, we obtain the rates at which $E\left[\frac{B_y^*}{B_y}-1\right]^2$ and $E\left[\frac{M_y^*}{M_y}-1\right]^2$ converge to zero as n (sample size) is increased, where B_y^* and M_y^* are the proposed estimators of B_y and M_y respectively. We obtain conditions under which the estimators have infinite asymptotic relative accuracy (in L_2 sense) in comparison to corresponding plug-in estimators. Simulations reveal that if y is a mode or an anti-mode and $\log_{10}(h_n) \leq -0.5$, then performance of both M_y^* and M_y^A (plug-in estimator of M_y) are satisfactory. On the other hand, when y is in the tail region, M_y^* is the only estimator which successfully imitates the important features (e.g. existence of multiple minima) of M_y as a function of h_n.

Keywords and Phrases: Kernel density estimator, MSE, optimal-bandwidth, smooth bootstrap, plug-in estimator.

AMS Subject Classification: 62G07, 62G09, 62G20.

---

[*] Research supported by J.C. Bose Fellowship, Govt. of India.

1 INTRODUCTION

Given $X_1, X_2, ..., X_n$ i.i.d. random variables with density $f(\cdot)$, the *kernel density estimator* (of f) based on the kernel $K(\cdot)$ and bandwidth $h \equiv h_n$ is defined as

$$K_n(y) = \frac{1}{nh} \sum_{i=1}^{n} K\Big(\frac{y - X_i}{h}\Big)$$

where $h \to 0$ and $nh \to \infty$ as $n \to \infty$. Let us denote the *bias*, *variance* and *mean squared error (MSE)* of the kernel density estimator $K_n(\cdot)$, at y, by B_y, V_y and M_y respectively. Note that each of the above may be expressed as a functional $T(f)$. These local measures of accuracy of $K_n(\cdot)$ have enjoyed great popularity, especially in the context of locally optimal bandwidth selection of a kernel estimator. See for example Hall (1990), Falk (1992) and references therein.

In general $T(f)$ is unknown. In the smooth bootstrap approach $T(f)$ is estimated by $T(f_n)$ where f_n is an appropriate estimate of f_n (say a kernel density estimate with kernel K^0 and bandwidth λ). Traditionally, such smooth bootstrap estimators use the same kernel density estimate, that is $K^0 = K$ and $\lambda = h$. See for instance Hall (1992) who also described two possible generalizations where $K^0 = K$ but λ is either larger than h or some unspecified value. The analytic study of such estimators require additional restrictions on the basic kernel K.

In this paper we propose that the kernel K^0 and the bandwidth λ be chosen freely and not tied to the original kernel and bandwidth. We call these estimators B_y^*, V_y^* and M_y^* respectively. In our approach we need additional conditions on K^0 an λ and thus avoid imposing additional conditions on K and h as far as possible. There are other performance based reasons for choosing K^0 and λ over the traditional automatic choice of $K^0 = K$ and $\lambda = h$.

There are several other estimates available in the literature. Hall (1990) has proposed a bootstrap scheme, where the size of the bootstrap resample is less than the size of the original sample and K is compactly supported. Theorems 2.1 (Hall (1990), page 182-183) proves strong consistency of his bootstrap estimator M_y^H of M_y. Falk (1992) proposed smooth bootstrap estimators B_y^F and M_y^F of B_y and M_y. These are special cases of our B_y^* and M_y^* when we impose $K^0 = K$. Under the assumptions that $h = O(\frac{1}{n^{1/5}})$ and K is a compactly supported second order kernel, Falk (1992) has studied the weak convergence of B_y^F and asymptotic behaviour of $n^{4/5} M_y^F$. There are also plug-in estimators B_y^A, V_y^A and M_y^A available based on the asymptotic approximations of B_y, V_y and M_y.

There does not seem to have been any analytic study on the accuracy of the above three sets of estimators. In Section 3 we obtain the rates at which $r_1 = E\left[\frac{B_y^*}{B_y} - 1\right]^2$ and $r_3 = E\left[\frac{M_y^*}{M_y} - 1\right]^2$ converge to zero for a reasonably broad class of K, f and for any choice of h, where y is an interior point in the support of f. For instance we find that both r_1 and r_3 are $O\left(\frac{1}{n^{2p/(2s+2p+1)}}\right)$, where p is a known positive constant. The proposed bootstrap estimators compare well with plug-in estimators B_y^A, V_y^A and M_y^A of B_y, V_y and M_y respectively. For instance our results imply that for super-optimal h (defined later), B_y^* is asymptotically more accurate (in L_2 sense) than B_y^A. If $\limsup_{n\to\infty} n^{2/5} h = \infty$ and $f^{(1)}(y) = 0$, $f^{(2)}(y) \neq 0$ then V_y^* has infinite asymptotic relative accuracy (in L_2 sense) in comparison to V_y^A.

In Section 4 we have obtained closed form formulae for B_y, V_y, M_y and their corresponding estimators when K and K^0 are Gaussian kernels. So in this case, all of them can be computed explicitly. Simulations in Section 5 reveal that when y is a mode or an anti-mode and $\log_{10}(h) \leq -0.5$ both M_y^A and M_y^* estimate M_y accurately. However when y is in the tail region, M_y as a function of h

possesses more than one minima, and this feature is successfully imitated by M_y^*. But M_y^A always possesses one minima, say h_y^A. Let h_y and h_y^* be the global minimisers of M_y and M_y^*. If y is in the tail region, then both $\log_{10}(h_y^*)$, $\log_{10}(h_y) \geq 0.5$, whereas $\log_{10}(h_y^A) < 0$. This observation verifies Sain and Scott's (1995) result that when y is in the tail of f, h_y can be rather large. In such a situation h_y^* appears to be the more appropriate estimator of h_y.

2 NOTATION AND ASSUMPTIONS

Note that the parameters B_y, V_y and M_y are defined as follows

$$B_y \equiv B_y(K,h) = E[K_n(y)] - f(y) = \int K(u)\,[f(y-hu) - f(y)]\,du$$

$$V_y \equiv V_y(K,h) = \frac{1}{nh}\int K^2(u) f(y-hu)du - \frac{1}{n}\left(\int K(u) f(y-hu)du\right)^2$$

$$\text{and } \; M_y \equiv M_y(K,h) = V_y + (B_y)^2 .$$

Let K^0 be another kernel and $\lambda \equiv \lambda_n, \lambda^0 \equiv \lambda_n^0$ be two other bandwidth sequences. Let

$$K_n^0(y) = \frac{1}{n\lambda}\sum_{i=1}^{n} K^0\Big(\frac{y-X_i}{\lambda}\Big) \quad \text{and} \quad K_n^*(y) = \frac{1}{n\lambda^0}\sum_{i=1}^{n} K^0\Big(\frac{y-X_i}{\lambda^0}\Big).$$

The proposed *general smooth bootstrap estimator* of B_y and M_y are defined as

$$B_y^* = E_n[K_n(y)] - K_n^0(y) = \int K(u)\left[K_n^0(y-hu) - K_n^0(y)\right]du$$

$$V_y^* = \frac{1}{nh}\int K^2(u) K_n^*(y-hu)du - \frac{1}{n}\left(\int K(u) K_n^*(y-hu)du\right)^2 \quad \text{and} \quad M_y^* = V_y^* + \left(B_y^*\right)^2 .$$

Let h_y, h_y^* denote the values of h which minimize (globally) M_y and M_y^* respectively. h_y shall be referred to as the *optimal bandwidth.* A bandwidth h will be referred to as *sup-optimal* or *super-optimal* if $n^{1/2s+1}h$ is $o(1)$ or diverges to ∞ respectively. The point y is assumed to be an interior point of the support of f. It is said to be a *mode* or an *anti-mode* of f if $f^{(1)}(y) = 0$ and $f^{(2)}(y) < 0$ or $f^{(2)}(y) > 0$ respectively. Let $N(x, y^2)$ denotes the normal distribution with mean x and variance y^2. For any function H, $H^{(i)}$ shall denote its *ith* derivative and $||H|| = \sup_{-\infty<x<\infty}|H(x)|$. Let

$$r_1 = E\left[\frac{B_y^*}{B_y} - 1\right]^2, \quad r_2 = E\left[\frac{V_y^*}{V_y} - 1\right]^2 \quad \text{and} \quad r_3 = E\left[\frac{M_y^*}{M_y} - 1\right]^2 .$$

Note that r_1, r_2 and r_3 depend on y and n. For any two positive sequences $\{a_n\}, \{b_n\}$ we write $a_n = Oe(b_n)$ if $0 < \liminf_{n\to\infty} \frac{a_n}{b_n} \leq \limsup_{n\to\infty} \frac{a_n}{b_n} < \infty$.

We collect below all the assumptions on the two kernels and the bandwidths. Not all of them will be used in all the results.

Assumption A. (Assumptions on density f).

(i) $f(\cdot)$ is bounded, and possesses $s \geq 2$ bounded derivatives. The sth derivative $f^{(s)}$ is continuous and absolutely integrable.

(ii) There exists $p \geq 1$, such that $f^{(s+p)}(\cdot)$ is bounded and continuous.

Assumption B. (Assumptions on kernel K). $K(\cdot)$ is square integrable and is of sth order, that is $\int K(x)dx = 1$ and there exists an integer $s \geq 1$ such that $\int K(x)x^j dx = 0,\ j = 1, 2, .., s-1$, $0 < \left|\int K(x)x^s dx\right| \leq \int |K(x)x^s|dx < \infty$. The number s will be called the *order* of the kernel. Further we assume that $\int |K(x)x^{s+p}|dx < \infty$, where p is the integer for which A(ii) holds.

Assumption C. (Assumption on auxiliary kernel K^0).

(i) $K^0(\cdot)$ is a square integrable probability density function and $\int [K^0(y)]^4 dy < \infty$. Further

(a) $K^0(\cdot)$ is continuous and bounded.

(b) $K^0(x) \to 0$ as $|x| \to \infty$.

(ii) $K^0(\cdot)$ has s continuous derivatives on $(-\infty,\ \infty)$ and its sth derivative $K^{0(s)}(\cdot)$, satisfies the above conditions (a) and (b) and the following assumptions

(c) $\int |K^{0(s)}(x)|dx < \infty$ and $\int [K^{0(s)}(y)]^4 dy < \infty$

(d) $\int K^{0(s)}(x)x^j dx = 0$, where $0 \leq j = 0 \leq s+p-1, j \neq s$,
$\frac{(-1)^s}{s!} \int K^{0(s)}(x)x^s dx = 1$ and $\int |K^{0(s)}(x)x^{s+p}|dx < \infty$.

For all asymptotic results (as $n \to \infty$), it is understood that $\lambda,\ \lambda^0,\ h \to 0,\ n\lambda,\ n\lambda^0,\ nh \to \infty$.

Remark 1(i) The number p, in Assumptions A and C, depends on K^0. If K^0 is standard normal density then we recommend $p = 2$. With this choice of K^0 and p, Assumption C is satisfied for any value of s.
(ii) The Assumptions $A(i)-(ii)$ on f, are valid for a wide class of densities which include mixed normal, Cauchy, beta(m,n) (m,n>2) and gamma(n) (n>2) among others. Whereas the assumption that f has compact support or the assumption $E(|X_1|^\epsilon) < \infty, \epsilon > 0$ (see page 184, Hall (1990)) precludes the mixed normal distributions or the heavy tailed distributions which have no moments.
(iii) Assumption B on K is quite common in density estimation context and does not limit the choice of K. In contrast the assumptions by Hall (1990) and Falk (1992) prevent the use of a number of popular kernels e.g. the Gaussian or Gaussian type kernel, as they are not compactly supported.

3 MAIN RESULTS

We now state our main results. The proofs are given later.

Theorem 1. *Suppose Assumptions $A-C$ hold, and $f^{(s)}(y) \neq 0$ and $\lambda = Oe(\frac{1}{n^{1/(2s+2p+1)}})$. Then $r_1 = O\left(\frac{1}{n^{2p/(2s+2p+1)}}\right)$.*

Theorem 2. *Let Assumptions $A-C$ hold, $s \geq 2$, $10 > p \geq 2$, $\lambda^0 = Oe(n^{-1/5})$ and $\lambda = Oe(\frac{1}{n^{1/(2s+2p+1)}})$.*

(i) If $f(y) > 0$, $f^{(s)}(y) \neq 0$ and $\liminf_{n\to\infty} nh^{2s+1} > 0$ then, $r_3 = O\left(\frac{1}{n^{2p/(2s+2p+1)}}\right)$.

(ii) If $\limsup_{n\to\infty} nh^{2s+1} = 0$ and $f(y) > 0$, then, $r_3 = o\left(\frac{1}{n^{2p/(2s+2p+1)}}\right)$.

Remark 2 (i) If K is a second order kernel (i.e. $s = 2$) then, under Assumptions $A-C$, Theorems 1 and 2 hold whenever y is a mode or an anti-mode. Whether a point y is a mode or an anti-mode may be statistically tested using SiZer (Chaudhuri and Marron (1999)) which is a tool for detecting the points of "zero crossings" of $f^{(1)}$.

(ii) If $n\lambda^{2s+1} \to \infty$ and $n\lambda^{2s+2p+1} \to 0$ then under H_0, $\frac{a_n[K_n^{0(s)}(y)-f^{(s)}(y)]}{\sqrt{\int [K^{0(s)}(u)]^2 du . K_n(y)}} \xrightarrow{\mathcal{D}} N(0,\ 1)$ where $a_n = \sqrt{n\lambda^{2s+1}}$. This may be used to test for testing $H_0 : f^{(s)}(y) = 0$ against $H_1 : f^{(s)}(y) \neq 0$.

Plug in estimator. Plug-in estimators are obtained by substituting data based estimates into the asymptotic approximation of B_y and M_y and are easy to compute. Under Assumption A on f and B on K it is easy to see that

$$B_y = \frac{(-h)^s}{s!} \int K(u)u^s f^{(s)}(y) + o(h^s) \quad \text{and}$$
$$M_y = \frac{f(y)}{nh} \int K^2(u)du + \left[\frac{(-h)^s}{s!} \int K(u)u^s f^{(s)}(y)\right]^2 + o\left(\frac{1}{nh} + h^{2s}\right).$$

The corresponding plug in estimators may then be defined as

$$B_y^A = \frac{(-h)^s}{s!} \int K(u)u^s K_n^{0(s)}(y),\ V_y^A = \frac{K_n^*(y)}{nh} \int K^2(u)du \quad \text{and} \quad M_y^A = V_y^A + [B_y^A]^2.$$

Whereas computing B_y^* and M_y^* may require Monte-carlo simulation, the plug-in estimators are easier to implement. So a natural question is under what conditions bootstrap estimators are worth the extra computational effort ? The next three theorems provide conditions under which the bootstrap estimators will have infinite asymptotic accuracy (in L_2 sense) compared to their plug-in counterparts.

$$\text{Let} \quad r_4 = E\left[\frac{B_y^A}{B_y} - 1\right]^2, \quad r_5 = E\left[\frac{V_y^A}{V_y} - 1\right]^2 \quad \text{and} \quad r_6 = E\left[\frac{M_y^A}{M_y} - 1\right]^2.$$

Theorem 3. *Suppose Assumptions* $A - C$ *hold,* $f^{(s+1)}$ *is continuous,* $||f^{(s+1)}|| < \infty$, $\lambda = Oe\left(\frac{1}{n^{1/(2s+2p+1)}}\right)$, $f^{(s+1)}(y) \neq 0$ *and* $\limsup_{n\to\infty} n^{p/(2s+2p+1)}h = \infty$. *Then* $\lim_{n\to\infty} \frac{r_1}{r_4} = 0$.

Remark 3 For $s, p \geq 2$, $\limsup_{n\to\infty} n^{1/(2s+1)}h = \infty$ implies $\limsup_{n\to\infty} n^{p/(2s+2p+1)}h = \infty$. So Theorem 3 holds for second order kernel and super-optimal h for which B_y can be high. Hence B_y^* is expected to be more accurate than B_y^A in the high bias region.

Theorem 4. *Suppose Assumptions* $A - C$ *hold,* $\limsup_{n\to\infty} n^{2/5}h = \infty$, $\lambda^0 = Oe\left(\frac{1}{n^{1/5}}\right)$, $f(y) > 0$, $f^{(2)}(y) \neq 0$ *and* $f^{(1)}(y)\int K^2(u)du \neq f^2(y)$. *Then* $\lim_{n\to\infty} \frac{r_2}{r_5} = 0$.

Remark 4 The condition $f^{(1)}(y)\int K^2(u)du \neq f^2(y)$ is automatically satisfied whenever y is a mode or anti-mode and $f(y) > 0$.

Theorem 5. *Suppose* $s \geq 2$, $10 > p \geq 2$, $\liminf_{n\to\infty} n^{1/(2s+1)}h > 0$ *and* $h = O\left(\frac{1}{n^{p/(2s+2p+1)}}\right)$. *Suppose further that Assumptions* $A - C$ *hold,* $|\int K(u)u^{s+1}du| < \infty$, $||f^{(s+1)}|| < \infty$, $\lambda = Oe\left(\frac{1}{n^{1/(2s+2p+1)}}\right)$, $\lambda^0 = Oe\left(\frac{1}{n^{1/5}}\right)$, $f^{(2)}(y) \neq 0$ *and* $f^{(1)}(y)\int K^2(u)du \neq f^2(y)$. *Then* $\frac{r_3}{r_6} = o(1)$.

Remark 5: (i) In addition to the conditions in Theorem 5, if we further assume that $f^{(s+1)}(y) = 0$, then condition $h = O\left(\frac{1}{n^{p/(2s+2p+1)}}\right)$ can be replaced by a more general condition $h^2 = O\left(\frac{1}{n^{p/(2s+2p+1)}}\right)$ which is satisfied by $h = Oe\left(\frac{1}{n^{1/(2s+1)}}\right)$ for $p = 2$. So for $p = 2$, Theorem

5 holds for $h = Oe\left(\frac{1}{n^{1/(2s+1)}}\right)$. The values of h, which are constant multiples of $\frac{1}{n^{1/(2s+1)}}$, have been of great interest in density estimation from the perspective of minimising M_y asymptotically (see Hall (1990), Falk (1992)).

(ii) If y is a mode or anti-mode then the conditions $f^{(2)}(y) \neq 0$ and $f^2(y) \neq f^{(1)}(y) \int K^2(u)du$ are automatically satisfied. So whenever y is mode or an anti-mode, Theorem 5 ensures that M_y^* has infinite asymptotic accuracy in comparison M_y^A.

Fixed Sample performance of bootstrap estimators: The following proposition provides some indication of the performance of the bootstrap and plug-in estimators for fixed sample sample size n.

Proposition 1. *Under Assumption A on f and assuming K^0 is bounded, continuous, for any choice of λ and λ^0 and fixed sample size n, as $h \to \infty$,*
(i) $B_y \to -f(y)$, $B_y^* \to -K_n^0(y)$ and $B_y^A \to \infty$ almost surely.

(ii) $V_y \to 0$ and $V_y^*, V_y^A \to 0$ almost surely.

(iii) $M_y \to f^2(y)$, $M_y^* \to [K_n^0(y)]^2$ and $M_y^A \to \infty$ almost surely.

Remark 6 Thus for any sample size the bootstrap estimators successfully imitate the behaviors of B_y, V_y and M_y, for large value of h. But the asymptotic estimators fail to mimic the behavior of B_y and M_y in the high bias region i.e. for large values of h.

4 IMPLEMENTING SMOOTH BOOTSTRAP

B_y^*, V_y^* and M_y^* do not have a closed form expressions in general and hence Monte-Carlo computation is required for its implementation. However we observe that if K is a Gaussian kernel and K^0 is chosen to be the standard normal density then we can obtain closed form expression for the proposed bootstrap estimators. This follows from the following Lemma.

Lemma 1. *If $g(x) = \sum_{i=1}^{k} w_i \phi_{\sigma_i^2}(x-\mu_i)$, where $\phi_{\sigma_i^2}(\cdot)$ is the density of $N(0, \sigma_i^2)$ distribution and ϕ is the $N(0,1)$ density then*

$$\int \phi(u) g(x - \sigma u) du = \sum_{i=1}^{k} w_i \phi_{\sigma_i^2 + \sigma^2}(x - \mu_i).$$

$$\int \phi^2(u) g(x - \sigma u) du = \frac{1}{2\sqrt{\pi}} \sum_{i=1}^{k} w_i \phi_{\sigma_i^2 + \frac{\sigma^2}{2}}(x - \mu_i).$$

If K^0 is chosen to be $N(0,1)$ density, then

$$K_n^0(y) = \frac{1}{\sqrt{2\pi} n \lambda} \sum_{i=1}^{n} e^{-\frac{(y-X_i)^2}{2(\lambda)^2}} \text{ and } K_n^*(y) = \frac{1}{\sqrt{2\pi} n \lambda^0} \sum_{i=1}^{n} e^{-\frac{(y-X_i)^2}{2(\lambda^0)^2}}.$$

K_n^0 and K_n^* are densities of the form $\sum_{i=1}^{n} w_i \phi_{\sigma_i^2}(x - \mu_i)$, where $w_i = \frac{1}{n}$, $\mu_i = X_i$ and $\sigma_i^2 = (\lambda)^2$ or $(\lambda^0)^2$, $i = 1, 2, .., n$. Therefore if K is a Gaussian kernel then using Lemma 1 we can easily obtain closed form expression for B_y^*, V_y^* and M_y^*.

Theorem 6. *If both K and K^0 are densities of $N(0,1)$ distribution then*

$$B_y^* = \frac{1}{n}\sum_{i=1}^{n} \phi_{h_n^2+\lambda_n^2}(y - X_i) - K_n^0(y),$$

$$V_y^* = \frac{1}{2n^2 h\sqrt{\pi}} \sum_{i=1}^{n} \phi_{(\lambda^0)^2+\frac{h_n^2}{2}}(y - X_i) - \frac{1}{n}\left[\frac{1}{n}\sum_{i=1}^{n} \phi_{h_n^2+(\lambda^0)^2}(y - X_i)\right]^2$$

$$\text{and} \;\; M_y^* = V_y^* + (B_y^*)^2.$$

If the underlying distribution is assumed to be mixed normal distribution, then for Gaussian kernel K we can also obtain a closed form expression for B_y, V_y and M_y.

Theorem 7. *If $f(x) = \sum_{i=1}^{k} w_i \phi_{\sigma_i^2}(x - \mu_i)$, where $\phi_{\sigma_i^2}(\cdot)$ is the density of $N(0, \sigma_i^2)$ distribution and K is the density of $N(0,1)$ distribution then*

$$B_y = \sum_{i=1}^{k} w_i \phi_{h_n^2+\sigma_i^2}(y - \mu_i) - f(y),$$

$$V_y = \frac{1}{2nh\sqrt{\pi}} \sum_{i=1}^{k} w_i \phi_{\sigma_i^2+\frac{h_n^2}{2}}(y - \mu_i) - \frac{1}{n}\left[\sum_{i=1}^{k} w_i \phi_{h_n^2+\sigma_i^2}(y - \mu_i)\right]^2$$

$$\text{and} \;\; M_y = V_y + (B_y)^2.$$

5 SIMULATION

We investigated by means of simulations, the effect of y and h on the performance of M_y^* and M_y^A for fixed sample size. Since any density may be approximated arbitrarily closely in various senses by a normal mixture density (see Marron and Wand (1992)), we chose f to be mixed normal. We chose K to be Gaussian due to its wide popularity. Note that a kernel density estimator is not that sensitive to the choice of the kernel. We also chose K^0 to be Gaussian and hence closed form expression for computing M_y^* and M_y are available from Theorems 6 and 7. Since K and K^0 are standard normal density, $s = 2$ and $p = 2$. Further we chose $\lambda^0 = n^{-1/5}$ and $\lambda = n^{-1/(2s+2p+1)}$.

For $n = 500$, we have plotted M_y^*, M_y^A and M_y against $\log_{10} h$ taking f to be standard normal, bimodal, skewed and kurtic densities. Formulae of these densities are available in Marron and Wand (1992). Our choice of $\log_{10}$ scale is motivated by its use by Marron and Wand (1992). The following conclusions may be drawn from these simulations.

(i) Figures 1(b), 2(d), 3(b, c) and 4(b, c) reveal that when y is in the tail region, M_y may have more than one minima and M_y^* captures all important features (including multiple minima) of M_y as a function of h. However M_y^A always has one minima irrespective of y and it fails to mimic M_y specially when h is large and y is in the tail region.

(ii) Figures 1(a), 2(a, b, c) and 3(a) consider the case when y is a mode or anti-mode of f. These reveal that both M_y^A and M_y^* successfully imitate M_y when $\log_{10}(h) \leq -0.5$. However when $\log_{10}(h) \geq 0$, M_y^A increases rapidly whereas both M_y^* and M_y first increase and then appear to level-off, as the value of h is increased.

However, in general we noticed that if f is a mixed normal, $f(x) = \sum_{i=1}^{k} w_i \phi_{\sigma_i^2}(x - \mu_i)$ and if a particular σ_i^2 is small and the corresponding w_i is not too small, then both M_y^* and M_y^A are poor estimators of M_y at $y = \mu_i$ for $\log_{10}(h) \geq -1.5$. For example, in Figure 4(a) we have plotted the result when f is the density of $\frac{2}{3}N(0,1) + \frac{1}{3}N(0, \frac{1}{10^2})$ and $y = 0$. It is to be noted that M_y^* continues to accurately estimate M_y if $\log_{10}(h) \leq -1.5$.

(iii) From the perspective of estimating h_y we see that both h_y^A and h_y^* perform equally well when y is a mode or an anti-mode.

But if y is in the tail region of f, then from Figures 1(b), 2(d), 3 (b, c) and 4(b, c) we see that M_y attains two minima, one in the range $\log_{10}(h) \leq 0$ and the other in the range $\log_{10}(h) \geq 0.5$ and the global minima need not be unique. This feature is successfully imitated by M_y^*. Further, the global minimizer of M_y^* also need not unique. In any case, the larger minima always turn out to be the global minima and hence without loss h_y and h_y^* are taken to be the largest values of h minimizing M_y and M_y^* respectively. It also turns out that they are close. The above choice of h_y^* and h_y are also supported by Sain and Scott's (1995) observation that the sequence h_y can converge to a positive constant, rather than zero. In contrast, M_y^A has a unique minimizer h_y^A and $\log_{10}(h_y^A) < 0$, irrespective of where y is. In conclusion, if y is in the tail, h_y^A is a poor estimate for h_y whereas h_y^* is close to h_y.

(iv) Hall (1990) showed that the minimizers of M_y^H and M_y, with respect to h over $A = [\frac{\epsilon}{n_1^{1/2s+1}}, \frac{\epsilon^{-1}}{n_1^{1/2s+1}}]$ (n_1 is the resample size, $0 < \epsilon < 1$), are asymptotically equivalent (almost surely). The results of Falk (1992) imply that, for a second order kernel, minimizing M_{ny}^F, as a function of h over $B = \left[\frac{C_1}{n^{1/5}}, \frac{C_2}{n^{1/5}}\right]$, is asymptotically equivalent to minimizing $\frac{f(y)}{cn^{4/5}} \int K^2 + (\frac{c^2}{2n^{2/5}} f^{(2)}(y) \int x^2 K(x) dx)^2$ with respect to c over $[C_1, C_2]$. Let h_y^H denote the minimizer of M_y^H over A and h_y^F be the minimizer of M_y^F over B. For second order kernel, h_y^F and h_y^H are $Oe(\frac{1}{n^{1/5}})$ and $Oe(\frac{1}{n_1^{1/5}})$, where $n_1 < n$ and $n_1 \to \infty$. Consequently for large n, both h_y^F and h_y^H are expected to be close to zero and can be much smaller than h_y when y is in the tail. In fact for such y, both h_y^F, h_y^H can be expected to be closer to the smaller local minima of M_y. Thus h_y^* appears to be a more appropriate estimator of h_y than h_y^F and h_y^H.

6 PROOF OF THEOREMS

We start with the following lemma which will be used in the sequel

Lemma 2. *Under Assumptions A on f and C(ii) on $K^{0(s)}$ and choosing* $\lambda = O\left(\frac{1}{n^{1/(2s+2p+1)}}\right)$,

(i) $\sup_{-\infty<y<\infty} E\left[K_n^{0(s)}(y) - f^{(s)}(y)\right]^2 = O\left(\frac{1}{n^{2p/(2p+2s+1)}}\right)$.

(ii) $\sup_{-\infty<y<\infty} E\left[K_n^{0(s)}(y) - f^{(s)}(y)\right]^4 = O\left(\frac{1}{n^{4p/(2p+2s+1)}}\right)$.

Proof of Lemma 2 (i) Recall that $K_n^{0(s)}(y) = \frac{1}{n\lambda_n^{s+1}} \sum_{i=1}^{n} K^{0(s)}\left(\frac{y - X_i}{\lambda}\right)$.

$$E\left[K_n^{0(s)}(y) - f^{(s)}(y)\right]^2 = Var\left(K_n^{0(s)}(y)\right) + \left[E\left(K_n^{0(s)}(y)\right) - f^{(s)}(y)\right]^2.$$

It is easy to verify that we see that

$$Var\left(K_n^{0(s)}(y)\right) \leq \frac{1}{n\lambda^{2s+1}}||f||\int\left[K^{0(s)}(u)\right]^2du,\ \forall y.$$

and under Assumption C on $K^{0(s)}$ it is easy to verify that

$$\left[E\left[K_n^{0(s)}(y)\right]-f^{(s)}(y)\right]^2 \leq \lambda_n^{2p}\left[\frac{||f^{(s+p)}||}{(s+p)!}\int K^{0(s)}(u)u^{s+p}du\right]^2, \quad \forall y.$$

Therefore for $\lambda = O\left(\frac{1}{n^{1/(2s+2p+1)}}\right)$, Lemma 2 (i) is an immediate consequence of the above inequalities. □

Proof of Lemma 2 (ii) Using $(a+b)^4 \leq 8(a^4+b^4)$ we see that

$$E\left[K_n^{0(s)}(y)-f^{(s)}(y)\right]^4 \leq 8E\left[K_n^{0(s)}(y)-E\left[K_n^{(s)}(y)\right]\right]^4+8\left[E\left[K_n^{0(s)}(y)\right]-f^{(s)}(y)\right]^4. \tag{6.1}$$

Let $Y_{ni}=\frac{1}{n\lambda_n^{s+1}}\left\{K^{(s)}\left(\frac{y-X_i}{\lambda}\right)-E\left[K^{(s)}\left(\frac{y-X_i}{\lambda}\right)\right]\right\}, i=1,2,...,n$. Then $Y_{n1},Y_{n2},.....,Y_{nn}$ are i.i.d random variables and $E(Y_{n1})=0$. Therefore we get the following equation

$$E\left[K_n^{0(s)}(y)-E\left(K_n^{(s)}(y)\right)\right]^4 = E\left(\sum_{i=1}^n Y_{ni}\right)^4 = nE(Y_{n1}^4)+6n(n-1)\left[E(Y_{n1}^2)\right]^2. \tag{6.2}$$

Now

$$\begin{aligned} E(Y_{n1}^4) &\leq \frac{8}{n^4\lambda^{4s+4}}\left\{E\left[K^{(s)}\left(\frac{y-X_1}{\lambda}\right)\right]^4+\left[EK^{(s)}\left(\frac{y-X_1}{\lambda}\right)\right]^4\right\} \\ &\leq \frac{8}{n^4\lambda^{4s+4}}\left[||f||\lambda\int[K^{(s)}(v)]^4dv+||f||^4\lambda^4\left[\int K^{(s)}(v)dv\right]^4\right] \\ &= \frac{C_1}{n^4\lambda^{4s+3}}\{1+C_2\lambda^3\}, \text{where } \ C_1,\ C_2 \ \text{ are positive constants} \end{aligned}$$

and

$$E(Y_{n1}^2) \leq \frac{1}{n^2\lambda^{2s+2}}E\left[K^{(s)}\left(\frac{y-X_1}{\lambda}\right)\right]^2 \leq \frac{1}{n^2\lambda^{2s+1}}||f||\int\left[K^{(s)}(v)\right]^2dv$$

Substituting the above inequalities in equation (6.2) and using $n\lambda \to \infty$ we get

$$E\left[K_n^{0(s)}(y)-E\left[K_n^{(s)}(y)\right]\right]^4 \leq \frac{6}{n^2\lambda^{4s+2}}\left[||f||\int\left[K^{(s)}(v)\right]^2dv\right]^2+o\left(\frac{1}{n^2\lambda^{4s+2}}\right), \ \forall y.$$

Therefore for $\lambda = Oe\left(\frac{1}{n^{1/(2s+2p+1)}}\right)$, we get

$$\sup_{-\infty<y<\infty} E\left[K_n^{0(s)}(y)-E\left[K_n^{0(s)}(y)\right]\right]^4 = O\left(\frac{1}{n^{4p/(2s+2p+1)}}\right). \tag{6.3}$$

Further

$$\left[E\left[K_n^{0(s)}(y)\right]-f^{(s)}(y)\right]^4\le\left[E\left[K_n^{0(s)}(y)-f^{(s)}(y)\right]^2\right]^2,\ \forall y.$$

Therefore for $\lambda=Oe\left(\frac{1}{n^{1/(2s+2p+1)}}\right)$ and using Lemma 2 (i), we get

$$\sup_{-\infty<y<\infty}\left[E\left[K_n^{0(s)}(y)\right]-f^{(s)}(y)\right]^4=O\left(\frac{1}{n^{4p/(2s+2p+1)}}\right). \tag{6.4}$$

Substituting the equations (6.3) and (6.4) in the right side of equation (6.1) we get Lemma 2(ii). So Lemma 2 is proved completely. □

Proof of Theorem 1 Recall that

$$B_y=\int K(u)\left[f(y-hu)-f(y)\right]du\ \text{ and }\ B_y^*=\int K(u)\left[K_n^0(y-hu)-K_n^0(y)\right]du.$$

For each fixed y and u, expanding $f(y-hu)$ and $K_n^0(y-hu)$ by Taylor's expansion with integral remainder we get

$$B_y=\frac{(-h)^s}{(s-1)!}\int K(u)u^s\int_0^1(1-t)^{s-1}f^{(s)}(y-thu)dtdu$$

and almost surely

$$B_y^*=\frac{(-h)^s}{(s-1)!}\int K(u)u^s\int_0^1(1-t)^{s-1}K_n^{0(s)}(y-thu)dtdu.$$

Therefore, almost surely, we get

$\frac{1}{h^s}\left|B_y-B_y^*\right|\le\frac{1}{(s-1)!}\int|K(u)u^s|\int_0^1(1-t)^{s-1}\left|f^{(s)}(y-thu)-K_n^{0(s)}(y-thu)\right|dtdu.$

Squaring and taking expectation on both sides of the above inequality we get

$$\frac{1}{h^{2s}}E\left[B_y-B_y^*\right]^2\le\frac{C_1^2}{(s!)^2}\sup_{-\infty<y<\infty}E\left[K_n^{0(s)}(y)-f^{(s)}(y)\right]^2 \tag{6.5}$$

where $C_1=\int|v|^sK(v)dv$.
Under the Assumptions A, C on f and K^0, choosing $\lambda=O(\frac{1}{n^{1/(2s+2p+1)}})$, from Lemma 2(i) we get

$$\sup_{-\infty<y<\infty}E\left[K_n^{0(s)}(y)-f^{(s)}(y)\right]^2=O\left(\frac{1}{n^{2p/(2s+2p+1)}}\right).$$

substituting the above equation in right side of (6.5) we see that

$$\frac{1}{h^{2s}}E\left[B_y-B_y^*\right]^2=O\left(\frac{1}{n^{2p/(2s+2p+1)}}\right). \tag{6.6}$$

Using the smoothness Assumption A on $f^{(s)}$ it is easy to see that

$$\frac{1}{h^{2s}}B_y^2=\left(\frac{\int K(u)u^sdu}{s!}f^{(s)}(y)\right)^2+o(1). \tag{6.7}$$

Since $r_{1n} = \frac{\frac{1}{h^{2s}} E\left[B_y - B_y^*\right]^2}{\frac{1}{h^{2s}} B_y^2}$, therefore under the Assumptions $|\int K(u)u^s du| > 0$ and $|f^{(s)}(y)| > 0$, Theorem 1 is a direct consequence equations (6.6) and (6.7). $\square$

In order to prove Theorem 2 we need the following Lemma.

Lemma 3. *Under Assumptions $A - C$ and choosing $\lambda = O(\frac{1}{n^{1/(2s+2p+1)}})$ we get*

$$E\left[(B_y^*)^2 - B_y^2\right]^2 = O\left(\frac{h_n^{4s}}{n^{2p/(2s+2p+1)}}\right).$$

Proof of Lemma 3 Recalling the formulae of B_y^* and B_y, from the proof of Theorem 1, we see that

$$\begin{aligned}\left|\left(B_y^*\right)^2 - \left(B_y\right)^2\right| &= \frac{h^{2s}}{((s-1)!)^2}\Big|\left\{\int K(u)u^s \int_0^1 (1-t)^{s-1} K_n^{0(s)}(y-thu)dtdu\right\}^2 \\ &\qquad - \left\{\int K(u)u^s \int_0^1 (1-t)^{s-1} f^{(s)}(y-thu)dtdu\right\}^2\Big| \\ &\leq \frac{h^{2s}}{((s-1)!)^2}\Big[\int |K(u)u^s| \int_0^1 (1-t)^{s-1}\left|K_n^{0(s)}(y-thu) + f^{(s)}(y-thu)\right| dtdu. \\ &\qquad \int |K(u)u^s| \int_0^1 (1-t)^{s-1}\left|K_n^{0(s)}(y-thu) - f^{(s)}(y-thu)\right| dtdu\Big] \\ &= \frac{h^{2s}}{((s-1)!)^2}\ c_{1n}.c_{2n} \quad \text{(say)}.\end{aligned}$$

It is easy to see that

$$0 \leq c_{1n}.c_{2n} \leq c_{2n}^2 + \frac{||f^{(s)}||C}{s}\ c_{2n}$$

where $C = \int |K(u)u^s|$. Therefore

$$\frac{1}{h_n^{4s}} E\left[\left(B_y^*\right)^2 - \left(B_y\right)^2\right]^2 \leq \frac{2}{((s-1)!)^4}\left[E(c_{2n}^4) + (C')^2 E(c_{2n}^2)\right] \tag{6.8}$$

where $C' = \frac{||f^{(s)}||C}{s}$. Further using Cauchy–Schwartz inequality for c_{2n}^2 and c_{2n}^4 and taking expectation we get

$$E(c_{2n}^{2j}) \leq \frac{C^{2j}}{s^{2j}} \sup_{-\infty<y<\infty} E\left[K_n^{0(s)}(y) - f^{(s)}(y)\right]^{2j} \quad j = 1, 2.$$

From Lemma 2 (i) and (ii), we see that for $\lambda = O(\frac{1}{n^{1/(2s+2p+1)}})$

$$\sup_{-\infty<y<\infty} E\left[K_n^{0(s)}(y) - f^{(s)}(y)\right]^{2j} = O(\frac{1}{n^{2jp/(2s+2p+1)}}), \quad j = 1, 2.$$

Therefore from equation (6.8), we see that

$$\frac{1}{h_n^{4s}} E\left[\left(B_y^*\right)^2 - \left(B_y\right)^2\right]^2 = O\left(\frac{1}{n^{2p/(2s+2p+1)}}\right).$$

Hence Lemma 3 is proved completely. $\square$

Proof of Theorem 2 Recalling the definitions of M_y, M_y^* and using $(a+b)^2 \leq 2a^2 + 2b^2$ we find that

$$0 \leq E\left[M_y^* - M_y\right]^2 \leq 2E[B_y^2 - (B_y^*)^2]^2 + 2E[V_y - V_y^*]^2. \tag{6.9}$$

Recalling the definitions of V_y and V_y^* we get the following equation

$$\begin{aligned} V_y^* - V_y &= \frac{1}{nh}\int K^2(u)[K_n^*(y-hu) - f(y-hu)]du \\ &\quad - \frac{1}{n}\left[\left(\int K(u)K_n^*(y-hu)du\right)^2 - \left(\int K(u)f(y-hu)du\right)^2\right] \\ &= L_1 - L_2 \quad \text{(say)} \end{aligned}$$

Hence

$$E[V_y^* - V_y]^2 \leq 2E(L_1^2) + 2E(L_2^2). \tag{6.10}$$

Now

$$E(L_1^2) \leq \frac{\left(\int K^2(u)\right)^2}{(nh)^2} \sup_{-\infty<y<\infty} E[K_n^*(y) - f(y)]^2.$$

Further note that

$$0 \leq E\left[f(y) - K_n^*(y)\right]^2 \leq \frac{||f||}{n\lambda^0}\int (K^0(y))^2 dy + \left[\frac{||f^{(2)}||(\lambda^0)^2}{2!}\int K^0(y)y^2 dy\right]^2, \quad \forall y.$$

So choosing $\lambda^0 = Oe(n^{-1/5})$, which minimizes the right side of the above equation, we get

$$\sup_{-\infty<y<\infty} E\left[f(y) - K_n^*(y)\right]^2 = O(n^{-4/5}). \tag{6.11}$$

Therefore

$$E(L_1^2) = O\left(\frac{1}{n^{2+4/5}h^2}\right). \tag{6.12}$$

Now using $a^2 - b^2 = (a-b)(a+b)$ and Cauchy-Schwartz inequality it is easy to see that

$$\begin{aligned} n^2 E L_2^2 &\leq E(c_n.d_n)\left[\int |K(u)|du\right]^2, \quad \text{where} \\ c_n &= \int |K(u)|\left[f(y-hu) + K_n^*(y-hu)\right]^2 du \\ \text{and} \quad d_n &= \int |K(u)|\left[f(y-hu) - K_n^*(y-hu)\right]^2 du. \end{aligned}$$

Since $|f(y-hu) + K_n^*(y-hu)| \leq |f(y-hu) - K_n^*(y-hu)| + 2||f||$, for all y, therefore it is easy to see that $c_n d_n \leq 2d_n^2 + 8||f||^2 d_n$ and hence

$$n^2 E L_2^2 \leq C^2 E(c_n.d_n) \leq C^2\left[2E(d_n^2) + 8||f||^2 E(d_n)\right] \tag{6.13}$$

where $C = \int |K(u)|du$. Further it is easy verify that

$$0 \leq E(d_n^j) \leq C \sup_{-\infty<y<\infty} E[K_n^*(y) - f(y)]^{2j} \quad j = 1, 2.$$

For $\lambda^0 = Oe(n^{-1/5})$ recalling equation (6.11) we get

$$E(d_n) = O(n^{-4/5}).$$

By some straight forward algebra it is easy to verify that for $\lambda^0 = Oe(n^{-1/5})$

$$\sup_{-\infty<y<\infty} E[K_n^*(y) - f(y)]^4 = O\left(\frac{1}{n^{8/5}}\right).$$

Consequently $E(d_n^2) = O\left(\frac{1}{n^{8/5}}\right)$ and hence recalling equation (6.13) we get

$$E(L_2^2) = O\left(\frac{1}{n^{2+4/5}}\right) \tag{6.14}$$

From equations (6.10), (6.12) and (6.14) we get

$$E[V_y^* - V_y]^2 = O\left(\frac{1}{n^{2+4/5}h^2}\right). \tag{6.15}$$

From equations (6.9), (6.15) and Lemma 3 it is easy to see that

$$E\left[M_y^* - M_y\right]^2 = O\left(\frac{h^{4s}}{n^{2p/(2s+2p+1)}} + \frac{1}{n^{2+4/5}h^2}\right). \tag{6.16}$$

Using Assumption A(i) on f and Assumption B on K it is easy to prove that

$$M_y^2 = \left[\frac{f(y)}{nh}\int K^2(u)du + \frac{h^{2s}}{(s!)^2}\left[\int K(u)u^s f^{(s)}(y)\right]^2 + o\left(\frac{1}{nh} + h^{2s}\right)\right]^2. \tag{6.17}$$

Recall that

$$r_3 = \frac{E\left[M_y^* - M_y\right]^2}{M_y^2}. \tag{6.18}$$

If $\liminf_{n\to\infty} nh^{2s+1} > 0$ then, under the Assumption $f(y) > 0$ and $|f^{(s)}(y)| > 0$, dividing numerator and denominator of r_3 by h^{4s} we get, from equations (6.16), (6.17) and (6.18), that

$$r_3 = O\left(\frac{1}{n^{2p/(2s+2p+1)}} + \frac{1}{(nh^{2s+1})^2 n^{4/5}}\right) = O\left(\frac{1}{n^{2p/(2s+2p+1)}}\right)$$

using $s \geq 2$ and $10 > p \geq 2$.

If $\limsup_{n\to\infty} nh^{2s+1} = 0$ then, under the Assumption $f(y) > 0$, dividing numerator and denominator of r_3 by $\frac{1}{(nh)^2}$ we get, from equations (6.16) and (6.18), that

$$r_3 = O\left(\frac{nh^{2s+1}}{n^{2p/(2s+2p+1)}} + \frac{1}{n^{4/5}}\right) = o\left(\frac{1}{n^{2p/(2s+2p+1)}}\right)$$

using $s \geq 2$, $10 > p \geq 2$ and $\limsup_{n\to\infty} nh^{2s+1} = 0$.

So Theorem 2 is proved completely. □

Proof of Theorem 3 Recall that

$$B_y^A = \frac{(-h)^s}{s!} K_n^{0(s)}(y) \int K(u) u^s du.$$

and from the proof of Theorem 1 we see that

$$B_y = \frac{(-h)^s}{(s-1)!} \int K(u) u^s \int_0^1 (1-t)^{s-1} f^{(s)}(y - thu) dt du.$$

Using $|a-b| \geq \big||a-c| - |b-c|\big|$, it is easy to see that

$$\frac{1}{h^s} \left| E[B_y^A] - B_y \right| \geq |d_{1n} - d_{2n}|$$

where

$$d_{1n} = \left| \frac{1}{s!} \int K(u) u^s du . E\left[K_n^{0(s)}(y) - f^{(s)}(y)\right] \right|,$$

$$d_{2n} = \left| \frac{1}{(s-1)!} \int K(u) u^s \int_0^1 (1-t)^{s-1} \left[f^{(s)}(y - thu) - f^{(s)}(y)\right] dt du \right|.$$

Now

$$\frac{1}{h^{2s}} E\left[B_y^A - B_y\right]^2 \geq \frac{1}{h^{2s}} \left[E[B_y^A] - B_y\right]^2 \geq \left[d_{1n} - d_{2n}\right]^2$$

Therefore

$$\frac{r_1}{r_4} = \frac{\frac{1}{h^{2s}} \left[E[B_y^*] - B_y\right]^2}{\frac{1}{h^{2s}} \left[E[B_y^A] - B_y\right]^2} \leq \frac{\frac{1}{h^{2s}} E\left[B_y^* - B_y\right]^2}{\left[d_{1n} - d_{2n}\right]^2} \tag{6.19}$$

In view of equations (6.6) and (6.19), to prove Theorem 3, it is enough to show that

$$\liminf_{n\to\infty} \frac{|d_{1n} - d_{2n}|}{\lambda^p} = \infty, \quad \text{where} \;\; \lambda = Oe\left(\frac{1}{n^{1/(2s+2p+1)}}\right). \tag{6.20}$$

Using the Assumption A on $f^{(s)}$, $f^{(s+p)}$ and Assumption C (ii) on $K^{0(s)}$ it is easy to see that

$$E\left[K_n^{0(s)}(y) - f^{(s)}(y)\right] = \frac{(-1)^{(s+p)} \lambda^p}{(s+p)!} f^{(s+p)}(y) \int K^{0(s)}(u) u^{s+p} du + o(\lambda^p). \tag{6.21}$$

Substituting the above expression for $E\left[K_n^{0(s)}(y) - f^{(s)}(y)\right]$ in the definition of d_{1n} we get

$$d_{1n} = \frac{\lambda^p}{(s+p)!} \left| f^{(s+p)}(y) \int K^{0(s)}(u) u^{s+p} du \right| + o(\lambda^p).$$

So to prove equation (6.20) it is enough to show that $\limsup_{n\to\infty} \frac{d_{2n}}{\lambda^p} = \infty$.
Using the Taylor expansion for $f^{(s)}(y - thu) - f^{(s)}(y)$ and the assumption that $f^{(s+1)}(\cdot)$ is bounded, continuous, it is easy to show that

$$d_{2n} = \frac{h}{(s)!} |f^{(s+1)}(y) \int K(u) u^{s+1} du| + o(h).$$

Choosing $\lambda = Oe\left(\frac{1}{n^{1/(2s+2p+1)}}\right)$, using the assumptions $\limsup_{n\to\infty} n^{p/(2s+2p+1)}h = \infty$ and $f^{(s+1)}(y) \neq 0$, it is easy to see that $\limsup_{n\to\infty} \frac{d_{2n}}{\lambda^p} = \infty$ and consequently

$$\limsup_{n\to\infty} \frac{|d_{1n} - d_{2n}|}{\lambda^p} = \infty.$$

This establishes equation (6.20) and hence Theorem 3 is proved completely. □

Proof of Theorem 4 Recall that

$$V_y^A = \frac{1}{nh} K_n^*(y) \int K^2(u)du$$

and

$$V_y = \frac{1}{nh} \int K^2(u) f(y-hu)du - \frac{1}{n}\left(\int K(u)f(y-hu)du\right)^2.$$

Using $E(X^2) \geq |E(X)|^2$ and $|a-b| \geq \big||a-c| - |b-c|\big|$, it is easy to see that

$$E\left[V_y^A - V_y\right]^2 \geq \left|E[V_y^A] - V_y\right|^2 \geq |e_{1n} - e_{2n}|^2$$

where

$$e_{1n} = \frac{1}{nh}\int K^2(u)du|E(K_n^*(y)) - f(y)|, \tag{6.22}$$

$$e_{2n} = \left|\frac{1}{nh}\int K^2(u)\left[f(y-hu) - f(y)\right]du - \frac{1}{n}\left(\int K(u)f(y-hu)du\right)^2\right|. \tag{6.23}$$

Under the Assumption C on K^0 and $f^{(2)}(y) \neq 0$, choosing $\lambda^0 = Oe(n^{-1/5})$ it is easy to show that

$$e_{1n} = Oe\left(\frac{1}{n^{1+2/5}h}\right) \tag{6.24}$$

$$\lim_{n\to\infty} ne_{2n} = \left|f^{(1)}(y)\int K^2(u)du - f^2(y)\right| > 0 \quad \text{(by assumption)}.$$

Since $\lim_{n\to\infty} n^{2/5}h = \infty$,we have $\lim_{n\to\infty} \frac{e_{2n}}{e_{1n}} = \infty$.

$$\frac{r_2}{r_5} = \frac{E\left[V_y^* - V_y\right]^2}{E\left[V_y^A - V_y\right]^2} \leq \frac{E\left[V_y^* - V_y\right]^2}{\left[e_{1n} - e_{2n}\right]^2} = \frac{\frac{1}{e_{1n}^2}E\left[V_y^* - V_y\right]^2}{\left[1 - \frac{e_{2n}}{e_{1n}}\right]^2}.$$

Since $\lim_{n\to\infty} \frac{e_{2n}}{e_{1n}} = \infty$, therefore Theorem 4 is proved completely if we can show that

$$\frac{E\left[V_y^* - V_y\right]^2}{e_{1n}^2} = O(1). \tag{6.25}$$

Equation (6.25) is a direct consequence of the equations (6.15) and (6.24). This completes the proof of Theorem 4. □

To prove Theorem 5 we need the following Lemma.

Lemma 4. *Let* $s \geq 2$, $10 > p \geq 2$. *Under Assumptions* $A - C$ *and further assuming* $\left|\int K(u)u^{s+1}du\right| < \infty$, $||f^{(s+1)}|| < \infty$, $\liminf_{n\to\infty} nh^{2s+1} > 0$, $h = O\left(n^{-p/(2s+2p+1)}\right)$ *and for* $\lambda^0 = Oe(n^{-1/5})$, $\lambda = Oe(n^{-1/(2s+2p+1)})$ *we get*

$$(i) \text{ if } \; f^2(y) \neq f^{(1)}(y)\int K^2(u)du \text{ then } \qquad r_7 = \frac{\left|E[(B_y^A)^2] - B_y^2\right|}{\left|E[V_y^A] - V_y\right|} = o(1).$$

$$(ii) \text{ for } \; f^{(2)}(y) \neq 0 \qquad r_8 = \frac{\left[E[(B_y^*)^2] - B_y^2\right]^2}{E\left[V_y^* - V_y\right]^2} = o(1).$$

Proof of Lemma 4(i) Recalling the definition of B_y^A we see that

$$E[(B_y^A)^2] = \frac{h^{2s}}{(s!)^2}\left[\int K(u)u^s du\right]^2 E\left[K_n^{0(s)}(y)\right]^2.$$

Under Assumptions A on f and C on $K^{0(s)}$, it is easy to verify that

$$\begin{aligned}
E\left[K_n^{0(s)}(y)\right]^2 &= \frac{1}{n\lambda^{2s+1}}\int [K^{0(s)}(u)]^2 f(y-\lambda u)du \\
&\quad + \frac{2(n-1)}{n}\left[f^{(s)}(y) + \frac{(-1)^{s+p}\lambda^p}{(s+p-1)!}\int K^{0(s)}(u)u^{s+p}\int (1-t)^{s+p-1} f^{(s+p)}(y - t\lambda u)dtdu\right]^2 \\
&= \frac{1}{n\lambda^{2s+1}}\int [K^{0(s)}(u)]^2 f(y-\lambda u)du \\
&\qquad + \frac{2(n-1)}{n}\left[f^{(s)}(y) + \frac{(-1)^{s+p}\lambda^p}{(s+p)!}\int K^{0(s)}(u)u^{s+p}du f^{(s+p)}(y) + o\left(\lambda^p\right)\right]^2.
\end{aligned}$$

Therefore

$$\begin{aligned}
E[(B_y^A)^2] = C_1^2\frac{h^{2s}}{n\lambda^{2s+1}}\int [K^{0(s)}(u)]^2 f(y-\lambda u)du + \frac{(n-1)}{n}C_1^2 h^{2s}[f^{(s)}(y)]^2 \\
+ \frac{(n-1)}{n}C_2 h^{2s}\lambda^p f^{(s)}(y) f^{(s+p)}(y) + o\left(h^{2s}\lambda^p\right) \qquad (6.26)
\end{aligned}$$

where $C_1 = \frac{\int K(u)u^s du}{s!}$ and $C_2 = \frac{\int K^{0(s)}(u)u^{s+p}du}{(s+p)!}C_1$.

Further recall (from proof of Theorem 1) that

$$B_y = \frac{(-h)^s}{(s-1)!}\int K(u)u^s \int_0^1 (1-t)^{s-1} f^{(s)}(y - thu)dtdu.$$

Therefore

$$\begin{aligned}
B_y^2 \;&=\; \frac{h^{2s}}{[(s-1)!]^2}\Big[\frac{\int K(u)u^s}{s} f^{(s)}(y) \\
&\qquad -h\int K(u)u^{s+1}\int_0^1 (1-t)^{(s-1)}t\int_0^1 f^{(s+1)}(y - vthu)dvdtdu\Big]^2 \\
&=\; \frac{h^{2s}}{[(s-1)!]^2}\Big[\frac{\int K(u)u^s}{s} f^{(s)}(y) - h\frac{\int K(u)u^{s+1}}{s} f^{(s+1)}(y) + o(h)\Big]^2.
\end{aligned}$$

This implies that with $C_3 = C_1 \frac{\int K(u)u^{s+1}}{s!}$,

$$B_y^2 = C_1^2 h^{2s}[f^{(s)}(y)]^2 + C_3 h^{2s+1} f^{(s)}(y) f^{(s+1)}(y) + o\left(h^{2s+1}\right). \tag{6.27}$$

From equations (6.26) and (6.27) we see that

$$\begin{aligned} E[(B_y^A)^2] - B_y^2 &= C_1^2 \frac{h^{2s}}{n\lambda^{2s+1}} \int [K^{0(s)}(u)]^2 f(y-\lambda u) du \\ &+ C_2 h^{2s} \lambda^p f^{(s)}(y) f^{(s+p)}(y) - C_3 h^{2s+1} f^{(s)}(y) f^{(s+1)}(y) \\ &+ O\left(h^{2s}\left(\lambda^p + h + \frac{1}{n}\right)\right). \end{aligned} \tag{6.28}$$

Therefore under Assumption A that f and its higher order derivatives are bounded, $h = O\left(\frac{1}{n^{p/(2s+2p+1)}}\right)$ and choosing $\lambda = Oe\left(\frac{1}{n^{1/(2s+2p+1)}}\right)$ we see (from equation (6.28)) that

$$\lim_{n\to\infty} \frac{n^{p/(2s+2p+1)}}{h^{2s}} \left|E[(B_y^A)^2] - B_y^2\right| < \infty. \tag{6.29}$$

Recall that $V_y^A = \frac{K_n^*(y)}{nh} \int K^2(u) du$. Therefore, recalling the definition of V_y, we get the following equation

$$\begin{aligned} E\left[V_y^A\right] - V_y &= \frac{\int K^2(u)du}{nh} \left[E(K_n^*(y)) - f(y)\right] + \frac{1}{nh} \int K^2(u) \left[f(y) - f(y-hu)\right] du \\ &+ \frac{1}{n}\left(\int K(u) f(y-hu) du\right)^2. \end{aligned}$$

It is easy to verify that,

$$\begin{aligned} E(K_n^*(y)) - f(y) &= O\left((\lambda^0)^2\right), \\ \frac{1}{nh}\int K^2(u)\left[f(y) - f(y-hu)\right] du &= -\frac{1}{n} f^{(1)}(y) \int K^2(u) u du + O\left(\frac{h}{n}\right), \\ \frac{1}{n}\left(\int K(u) f(y-hu) du\right)^2 &= \frac{f^2(y)}{n} + O\left(\frac{h}{n}\right). \end{aligned} \tag{6.30}$$

Therefore for $\lambda^0 = Oe\left(\frac{1}{n^{1/5}}\right)$, from the above equations we see that

$$\begin{aligned} \left|E\left[V_y^A\right] - V_y\right| &= \frac{1}{n}\left|f^2(y) - f^{(1)}(y) \int K^2(u) u du\right| + O\left(\frac{1}{n^{1+2/5}h} + \frac{h}{n}\right). \\ &= L_4 + O\left(\frac{1}{n^{1+2/5}h} + \frac{h}{n}\right) \quad \text{(say)}. \end{aligned} \tag{6.31}$$

Under assumptions $\liminf_{n\to\infty} nh^{2s+1} > 0$ and $\left|f^2(y) - f^{(1)}(y)\int K^2(u) u du\right| > 0$, we see that

$$\lim_{n\to\infty} \frac{n^{p/(2s+2p+1)}}{h^{2s}} L_4 = \infty.$$

Further for $s \geq 2$, $10 > p \geq 2$, $\liminf_{n\to\infty} nh^{2s+1} > 0$ and $h = O\left(\frac{1}{n^{p/(2s+2p+1)}}\right)$

$$\frac{1}{n^{1+2/5}h} + \frac{h}{n} = o\left(\frac{h^{2s}}{n^{p/(2s+2p+1)}}\right).$$

Therefore for $s \geq 2$, $10 > p \geq 2$, under assumptions $h = O\left(\frac{1}{n^{p/(2s+2p+1)}}\right)$, $\liminf_{n\to\infty} nh^{2s+1} > 0$ and $f^2(y) \neq f^{(1)}(y) \int K^2(u)udu$, from equation (6.31) we see that

$$\lim_{n\to\infty} \frac{n^{p/(2s+2p+1)}}{h^{2s}} \left| E\left[V_y^A\right] - V_y \right| = \infty. \tag{6.32}$$

Lemma 4(i) is a direct consequence of the equations (6.29) and (6.32). □

Proof of Lemma 4(ii) Recalling the definitions of V_y^* and V_y we see that

$$\begin{aligned} \left|E\left(V_y^*\right) - V_y\right| &= \Big| \frac{1}{nh} \int K^2(u) E\left[K_n^*(y-hu) - f(y-hu)\right] du \\ &\quad -\frac{1}{n}\left[E\left(\int K(u)K_n^*(y-hu)du\right)^2 - \left(\int K(u)f(y-hu)du\right)^2\right] \Big| \\ &= \left|E[L_1] - E[L_2]\right|, \end{aligned}$$

where L_1 and L_2 are as defined in the proof of Theorem 2.

Recalling equation (6.14) we see that for $\lambda^0 = Oe(n^{-1/5})$

$$E(L_2) = O\left(\frac{1}{n^{1+2/5}}\right).$$

Under the Assumptions A on f, B on K and C on K^0, $s \geq 2$ and choosing $\lambda^0 = Oe(n^{-1/5})$ it is easy to verify that

$$E[L_1] = \frac{C' f^{(2)}(y)}{n^{1+2/5}h} + o\left(\frac{1}{n^{1+2/5}h}\right), \text{ where } C' \text{ is a non-zero constant.}$$

Therefore, under the Assumptions A on f, B on K, C on K^0, $s \geq 2$ and $|f^{(2)}(y)| > 0$ and choosing $\lambda^0 = Oe(n^{-1/5})$, we see that

$$\liminf_{n\to\infty} n^{1+2/5} h \left|E\left(V_y^*\right) - V_y\right| = |C' f^{(2)}(y)| > 0. \tag{6.33}$$

Recalling the definition of r_8 it is easy to see that

$$r_8 = \frac{\left[E[(B_y^*)^2] - B_y^2\right]^2}{E\left[V_y^* - V_y\right]^2} \leq \frac{E\left[(B_y^*)^2 - B_y^2\right]^2}{\left[E[V_y^*] - V_y\right]^2}. \tag{6.34}$$

Recalling Lemma 3 and equations (6.33) and (6.34), under Assumptions $A - C$ and for $\lambda = O(\frac{1}{n^{1/(2s+2p+1)}})$ and $\lambda^0 = Oe(n^{-1/5})$, we get

$$r_8 = O\left(\frac{h^{4s+1} n^{1+2/5}}{n^{2p/2s+2p+1}}\right) = O(g_n) \text{ (say)}.$$

We see that for $h = O\left(\frac{1}{n^{p/(2p+2s+1)}}\right)$ and $s, p \geq 2$, $g_n = o(1)$. Therefore Lemma 4(ii) follows immediately from the above equation. So Lemma 4 is proved completely. □

Proof of Theorem 5 Recalling the definitions of r_3 and r_6 we see that

$$\begin{aligned}
\frac{r_3}{r_6} &= \frac{E\left[M_y^* - M_y\right]^2}{E\left[M_y^A - M_y\right]^2} \leq \frac{2E\left[V_y^* - V_y\right]^2 + 2E\left[(B_y^*)^2 - B_y^2\right]^2}{\left[E(M_y^A) - M_y\right]^2} \\
&\leq \frac{2E\left[V_y^* - V_y\right]^2 + 2E\left[(B_y^*)^2 - B_y^2\right]^2}{\left[\left|E[(B_y^A)^2] - B_y^2\right| - \left|E[V_y^A] - V_y\right|\right]^2} \\
&= \frac{2E\left[V_y^* - V_y\right]^2}{\left[E[V_y^A] - V_y\right]^2} \frac{(1 + r_8)}{(r_7 - 1)^2} \\
&\leq \frac{2E\left[V_y^* - V_y\right]^2}{\left[e_{1n} - e_{2n}\right]^2} \frac{(1 + r_8)}{(r_7 - 1)^2} \qquad (6.35)
\end{aligned}$$

where $e_{1n}, \;\; e_{2n}$ are as defined in the proof of Theorem 4.

Recalling the proof of Theorem 4, under Assumptions $A - C$ and $\lim_{n\to\infty} n^{2/5}h = \infty$, we see that

$$\frac{E\left[V_y^* - V_y\right]^2}{\left[e_{1n} - e_{2n}\right]^2} = o(1) \qquad (6.36)$$

From Lemma 4(i) and (ii) we see that, under all the assumptions stated in Lemma 4

$$r_7 = o(1) \text{ and } \; r_8 = o(1).$$

Therefore Theorem 5 is a direct consequence of equation (6.35), (6.36) and Lemma 4. This completes the proof. □

REFERENCES

[1] Chaudhuri, P. and Marron, J.S. (1999). SiZer for exploration of Structures in curves. *J. Amer. Statist. Assoc.*, 94, No. 447, pp 807-825.

[2] Falk, M. (1992). Bootstrap optimal bandwidth selection for kernel density estimates. *J.Stat.Plann.Inf.*, 30, pp 13-22.

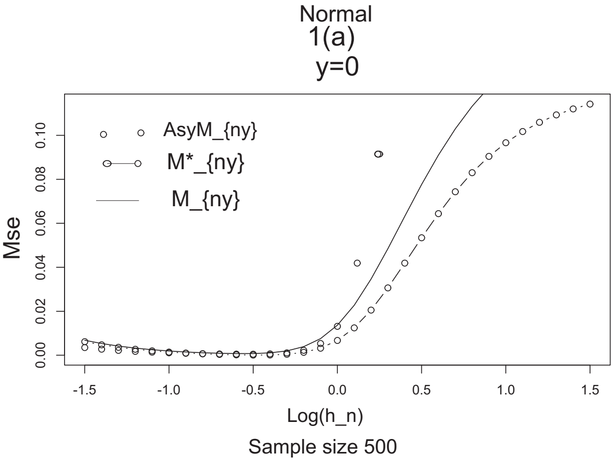
[3] Hall, P. (1990). Using the bootstrap to estimate the mean squared error and select smoothing parameter in nonparametric problems. *Journal of Multivariate Analysis*, 32,Issue 2, pp. 177-203.

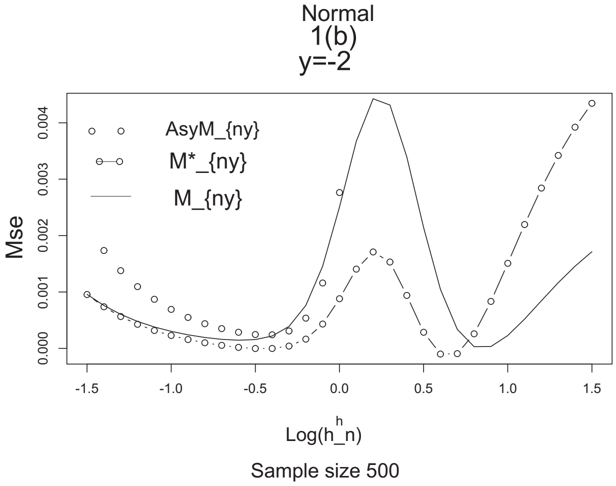
[4] Hall, P. (1992). The Bootstrap and Edgeeorth Expansion. *Springer-Verlar* New York, Inc.

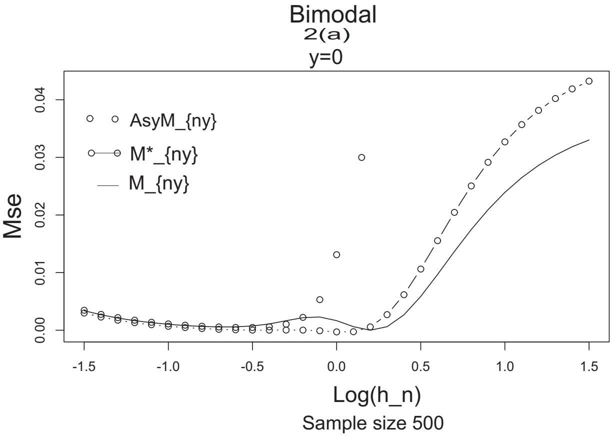
[5] Marron, J.S. and Wand, M.P. (1992). Exact mean integrated squared error. *Ann. Statist*, 20, No. 2, pp. 712-736.

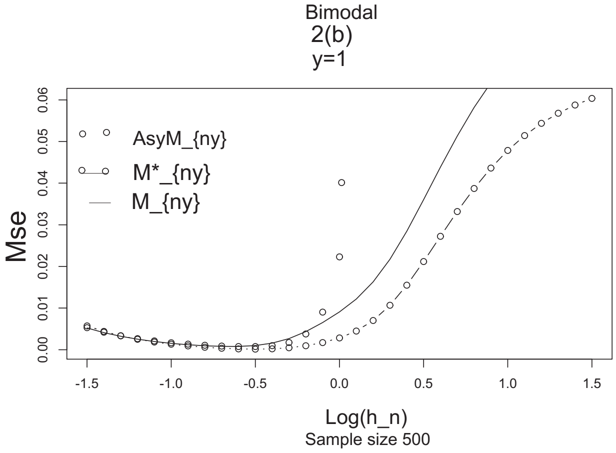
[6] Sain, S.R. and Scott, D.W. (1996). On Locally Adaptive Density Estimation. *Journal of the American Statistical Association*, 91, No. 436, Theory and Methods, pp. 1525-1534.

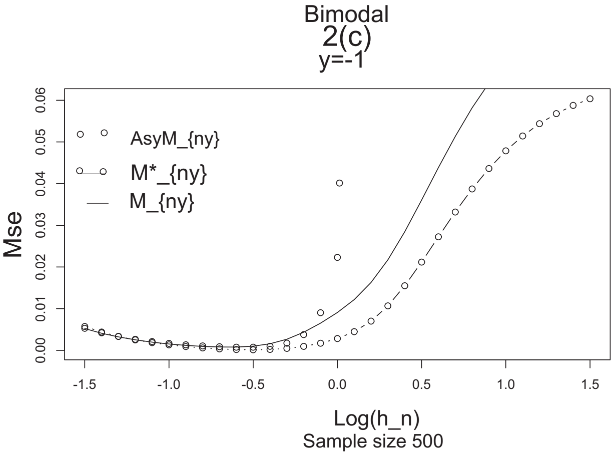
In Figures $1(a) - 4(c)$ we plot $M_{ny} \equiv M_y(h)$, $M^*_{ny} \equiv M^*_y(h)$ and $AsyM_{ny} \equiv M^A_y(h)$ against $\text{Log}(h) \equiv \log_{10} h$ for normal, bimodal, skewed and kurtic distributions and for sample size $n = 500$. Both K and K^0 are standard normal densities.

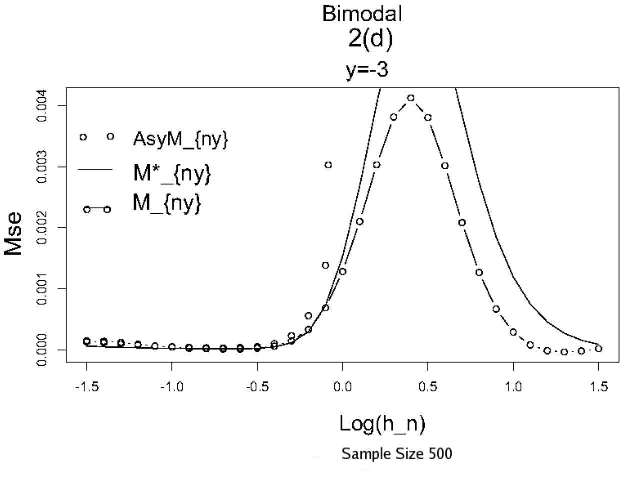


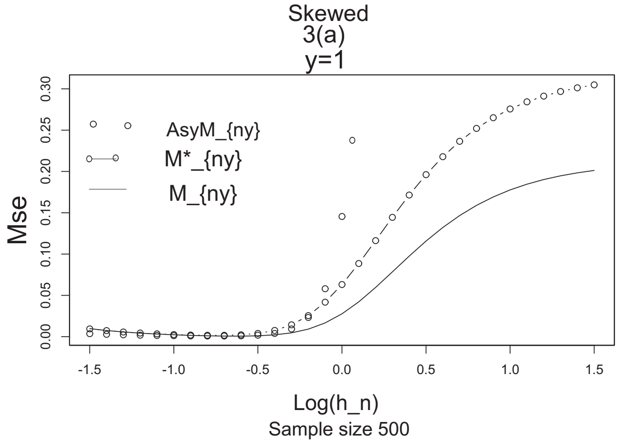


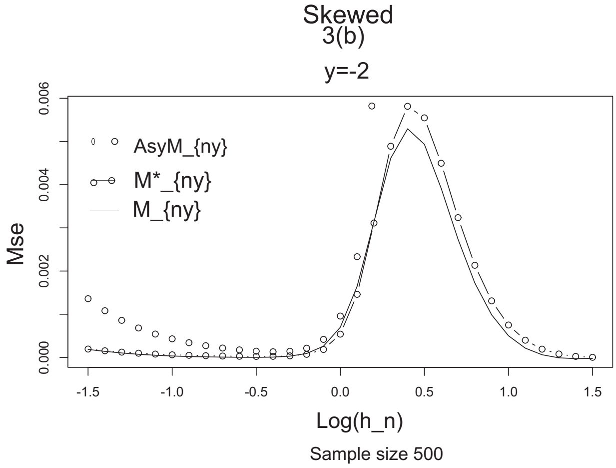


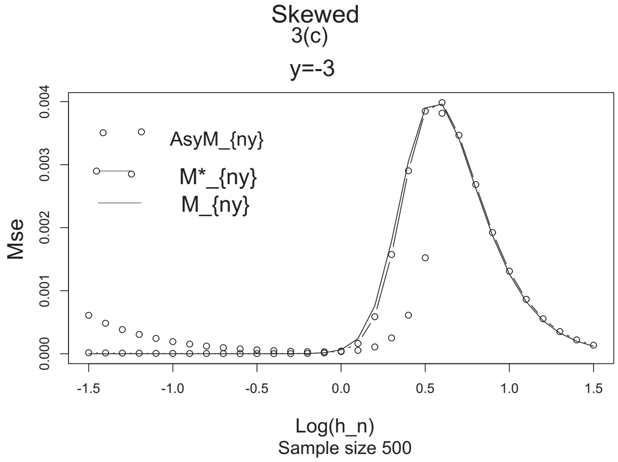


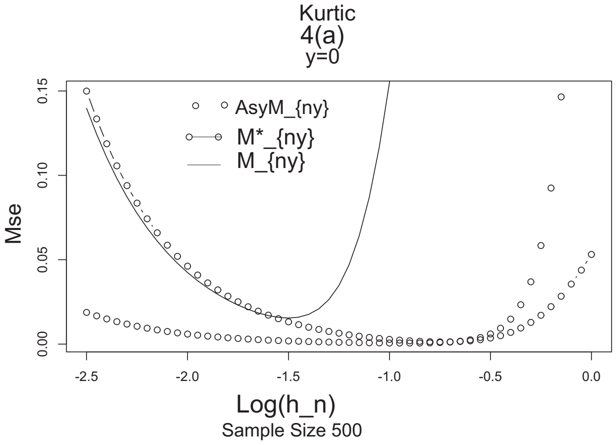


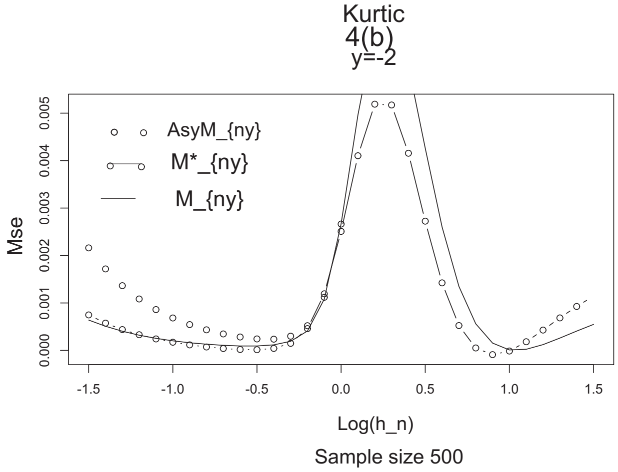


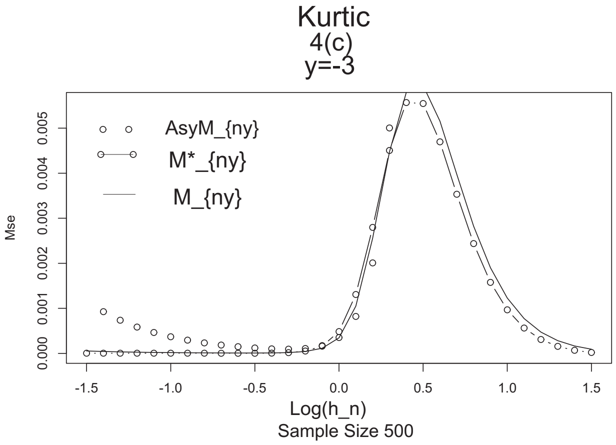

Skewed
3(a)
y=1
AsyM_{ny}
M*_{ny}
M_{ny}
Mse
0.00
0.05
0.10
0.15
0.20
0.25
0.30
-1.5
-1.0
-0.5
0.0
0.5
1.0
1.5
Log(h_n)
Sample size 500


Skewed
3(b)
y=-2
AsyM_{ny}
M*_{ny}
M_{ny}
Mse
0.000
0.002
0.004
0.006
-1.5
-1.0
-0.5
0.0
0.5
1.0
1.5
Log(h_n)
Sample size 500


Skewed
3(c)
y=-3
AsyM_{ny}
M*_{ny}
M_{ny}
Mse
0.000
0.001
0.002
0.003
0.004
-1.5
-1.0
-0.5
0.0
0.5
1.0
1.5
Log(h_n)
Sample size 500


Kurtic
4(a)
y=0
AsyM_{ny}
M*_{ny}
M_{ny}
Mse
0.00
0.05
0.10
0.15
-2.5
-2.0
-1.5
-1.0
-0.5
0.0
Log(h_n)
Sample Size 500


Kurtic
4(b)
y=-2
AsyM_{ny}
M*_{ny}
M_{ny}
Mse
0.000
0.001
0.002
0.003
0.004
0.005
-1.5
-1.0
-0.5
0.0
0.5
1.0
1.5
Log(h_n)
Sample size 500


Kurtic
4(c)
y=-3
AsyM_{ny}
M*_{ny}
M_{ny}
Mse
0.000
0.001
0.002
0.003
0.004
0.005
-1.5
-1.0
-0.5
0.0
0.5
1.0
1.5
Log(h_n)
Sample Size 500