Smooth Bootstrap estimate of Mean Integrated Squared Error

Arup Bose [*] Santanu Dutta [†]

Abstract

We propose a smoothed bootstrap estimator M_n^* of the MISE M_n of a kernel density estimator based on i.i.d. observations. Under broad assumptions, we have obtained the rate of convergence of $E\left[\frac{M_n^*}{M_n} - 1\right]^2$ to zero. If the bandwidth sequence $\{h_n\}$ satisfies $\limsup_{n\to\infty} nh_n^{2s} < \infty$, where s is the kernel order, then M_n^* is shown to be more accurate than asymptotic approximation to M_n. M_n^* also compares well with a number of other estimators of M_n. For large values of h_n, M_n^* appears to be the only appropriate estimator of M_n. Simulation reveals that the proposed method works well even for small samples and is not sensitive to the structure of the underlying density. In particular, if h_n satisfies $\log_{10} h_n \leq 0$ and the kernel K is the standard normal density, then the proposed method provides excellent approximation for any sample size and for a wide variety of underlying densities. However, when $\log_{10} h_n > 0$, our estimator is sensitive to the choice of bandwidth λ_n of the bootstrap estimator and, contrary to the traditional belief, the choice of $\lambda_n = h_n$ (proposed by Taylor (1989)) leads to poor estimate of the $MISE$ when the value of h_n is large.

Keywords and Phrases: Kernel density estimator, MISE, resampling, plug-in estimator, unbiased cross validation, biased cross validation.

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

---

[*]Stat-Math Unit, Indian Statistical Institute, 203 B.T. Road, Kolktata 700108, INDIA, e-mail: bosearu@gmail.com. This research has been supported by J.C.Bose Fellowship, Dept. of Science and Technology, Govt. of India.

[†], Mathematical Science Dept.., Tezpur University Napaam:784028, Tezpur, Assam, INDIA, e-mail: tezpur1976@gmail.com

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_n is defined as

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

where $h_n \to 0$ and $nh_n \to \infty$ as $n \to \infty$. The *mean integrated squared error (MISE)* of the kernel density estimator $K_n(\cdot)$ is defined as

$$M_n \equiv M_n(K, h_n) = \int_{-\infty}^{\infty} E\big[K_n(y) - f(y)\big]^2 dy = \int V\left[K_n(y)\right] dy + \int [B_n(y)]^2 dy,$$

where

$$V\left[K_n(y)\right] = E\left[K_n(y) - E(K_n(y))\right]^2 \text{ and } B_n(y) = E[K_n(y)] - f(y)$$

are the pointwise variance and bias of the estimator. The expectation E is computed with respect to the density $f(\cdot)$.

M_n is a global measure of accuracy of $K_n(\cdot)$. It has enjoyed great popularity, especially in the context of optimal bandwidth selection of a kernel estimator. See for instance, Taylor (1989), Jhun and Faraway (1990) and Hall, Marron and Park (1992). The relatively recent approach to kernel density estimation is to produce a family of density estimates based on a number of different choices of the bandwidth, instead of estimating the density with an "optimal" bandwidth (see Marron and Chung (1997), Sheather (2004)) and to learn about different features of the density from the plots at different levels of h_n. This approach can be beneficial if we can plot the MISE against different values of h_n. For instance, one can graphically detect the values of h_n for which MISE is minimum or close to minimum. These bandwidths are useful for estimating those parts of the density which are less wiggly.

It is not possible to evaluate M_n since f is unknown. Among other estimates, bootstrap based estimation of M_n and its use in bandwidth selection has been proposed by several authors such as Taylor (1989), Jhun and Faraway (1990), Hall

(1990) and Hall, Marron and Park (1992). Hall (1990) proposed a bootstrap scheme, where the size of the bootstrap resample is less than the size of the original sample, and proved the theoretical validity of the bootstrap approximation (see his equation (2.10), page 184). He assumed that K is compactly supported, Holder continuous, f is compactly supported, has s bounded uniformly continuous derivatives, $E|X_1|^\epsilon < \infty$ and $\epsilon n_1^{-1/2s+1} < h_n < \epsilon^{-1} n_1^{-1/2s+1}$, for some $\epsilon > 0$ and n_1 is the re-sample size. These assumptions limit the choice of K, h_n and f.

More recently Delaigle and Gijbels (2004, Theorem 4.1, page 27) have proved the validity of the smooth bootstrap approximation to M_n for deconvolution kernel density estimators based on data contaminated by random noise. Their results also imply the validity of smooth bootstrap approximation to M_n for the ordinary kernel density estimator, based on error free data, which we consider here. However they assumed that the characteristic function, $\phi_K(t) = \int e^{itx} K(x)dx$, has a compact support (see their condition B2, page 29) and K is of second order. Again, these assumptions are restrictive in the context of ordinary kernel density estimation based on error free i.i.d. observations.

In this paper we propose to study a smooth bootstrap method of estimating M_n, for any given h_n and for the broadest class of kernels K . Let $K^0$and λ_n be another kernel and bandwidth sequence, called the *auxiliary kernel* and *auxiliary bandwidth* respectively. Let

$$K_n^0(y) = \frac{1}{n\lambda_n} \sum_{i=1}^{n} K^0\Big(\frac{y - X_i}{\lambda_n}\Big).$$

Let $X_1^*, X_2^*, ..., X_n^*$ be i.i.d. *smooth bootstrap resample* of size n with density $K_n^0(\cdot)$. Let

$$K_{nB}(y) = \frac{1}{nh_n} \sum_{i=1}^{n} K\Big(\frac{y - X_i^*}{h_n}\Big)$$

be the *smooth bootstrap* version of $K_n(y)$. Given $X_1, X_2, ..., X_n$, the *smooth bootstrap estimator* of M_n is defined as

$$M_n^* \equiv M_n^*(K, K^0, \lambda_n, h_n) = \int_{-\infty}^{\infty} E_n\left[K_{nB}(y) - K_n^0(y)\right]^2 dy,$$

where the expectation E_n is computed with respect to the density $K_n^0(\cdot)$.

The idea of smooth bootstrap approximation to MISE is not new. For example, Taylor (1989, equation (6), page 707) obtained a closed form expression for the smooth bootstrap approximation to M_n for Gaussian kernel and obtained its asymptotic variance. His estimator (we call it T_n) is actually a special case of our proposed estimator M_n^*, when we choose $\lambda_n = h_n$ and both K and K^0 are standard normal density. Faraway and Jhun (1990) conducted a broad simulation study to investigate the performance of M_n^*.

It appears that there has been no significant theoretical study on the validity and accuracy of M_n^* under reasonably broad assumptions on K or h_n. We focus on this in Section 2. The basic idea of our proposal is to impose conditions only on K^0 and λ_n but not on K and h_n. There are some conditions on K and h_n. But these conditions are quite general and do not limit the choice of either K or h_n. Consequently we do not favour the automatic use of $K = K^0$ or $\lambda_n = h_n$. This approach appears to be new in the theory of smooth bootstrap. Our theoretical results and simulation provide some useful guidelines for choosing K^0 and λ_n and also demonstrate the drawback of some of the traditional ideas, for example, Taylor's (1989) suggestion to choose $\lambda_n = h_n$ for MISE estimation.

In Section 2, we establish the validity of our method and obtain the rates at which $E(M_n^* - M_n)^2$ (the L_2 error or Mean square error) converge to zero, in terms of n and h_n, for a wide class of kernels and underlying densities. In particular $E([\frac{M_n^*}{M_n} - 1]^2) = O(h_n)$ as $n \to \infty$ and this provides insight into the accuracy of M_n^*. This result seems to be new. Our proposed estimator M_n^* compares well with the bootstrap approximation of Delaigle and Gijbels (2004). For second order kernel density estimators, the bias and variance of M_n^* are $o(\frac{1}{n} + h_n^{4+\delta})$ and $o(\frac{1}{n^2} + h_n^{8+\delta'})$, $0 < \delta, \delta' < 1$, respectively, whereas the bias and variance of the smooth bootstrap approximation proposed by Delaigle and Gijbels (2004, Theorem 4.1, page 27)) are $O(\frac{1}{n}) + o(h_n^4)$ and $o(\frac{1}{n^2} + h_n^8) + o(\frac{h_n^4}{n})$ respectively. Moreover, if $\limsup_{n\to\infty} nh_n^{2s+\delta} < \infty$, $0 < \delta < 1$, then $Var(M_n^*)/Var(T_n) \to 0$. M_n^* also compares well with the unbiased cross-validation estimator (we call it UCV_n) and the biased cross-validation estimator (we call it BCV_n), see Scott and Terrell

(1987), of M_n. We find that the bias of M_n^* converges to zero faster than that of BCV_n, whereas UCV_n has a constant bias. Further M_n^* has infinite asymptotic relative efficiency in comparison to both UCV_n and BCV_n for h_n satisfying $\limsup_{n\to\infty} nh_n^{4s+\delta} < \infty$ and $\limsup_{n\to\infty} nh_n^{2s+\delta'} < \infty$ respectively, for some $0 < \delta, \delta' < 1$. So for large n, M_n^* is expected to be a more stable (less sampling fluctuation) than both UCV_n or BCV_n in high variance region, that is, for small values of h_n and less biased as well for any h_n.

In Section 3, we compare the accuracy of M_n^* with that of the plug-in method based asymptotic approximation. If h_n satisfies $\limsup_{n\to\infty} nh_n^{2s} < \infty$, where s is the kernel order, then M_n^* is more accurate asymptotically than the plug-in estimate.

In Section 4, we obtain a closed form expression for M_n^* for the kernel density estimators based on any Gaussian based kernel of order $p = 2r, r \geq 1$ ($p = 2$ yields the standard Gaussian kernel). This formula makes the bootstrap implementation straightforward, bypassing the Monte-Carlo step.

Simulations in Section 5 reveal that M_n^* works well even for small samples and is not sensitive to the structure of f. In particular, if $\log_{10} h_n \leq 0$ and K is the standard normal density, then M_n^* provides excellent approximation to M_n for a wide variety of f and it is not affected much by the choice of λ_n. However, if $\log_{10} h_n > 0$, then M_n^* is sensitive to the choice of λ_n, and Taylor's (1989) choice $\lambda_n = h_n$ can lead to poor estimation (with substantial negative bias) of M_n. Nevertheless the minima of M_n^* closely approximates the minima of M_n even for small samples.

2 Main results

We first 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) The density $f(\cdot)$ is bounded, and for some $s \geq 2$, the sth derivative $f^{(s)}$ is bounded and square integrable.

(ii) There exists M, such that $|f^{(s)}(x) - f^{(s)}(y)| \leq M|x - y|$, for all $x,\ y$.

(iii) There exists $p \geq 1$, such that $(s + p)$ th derivative $f^{(s+p)}(\cdot)$ exists, is integrable and is also square integrable.

Assumption B. (Assumptions on kernel K). The kernel $K(\cdot)$ 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 $\int K(x)x^j dx = 0,\ j = 1, 2, .., s - 1$ and $\int |K(x)x^s|dx < \infty$. Further we assume that $K(-x) = K(x)$ and $\int |K(x)x^{s+1}|dx < \infty$.

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

(i) The auxiliary kernel $K^0(\cdot)$ is a probability density function such that

(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 the sth derivative $K^{0(s)}(\cdot)$ of $K^0(\cdot)$, satisfies the above conditions (a) and (b) and the following assumptions

(c)$\int |K^{0(s)}(x)|dx < \infty$.

(d) $\int K^{0(s)}(x)x^j dx = 0$, where $j = 0, 1, 2, ..., s - 1, s + 1, ..., s + p - 1$,
$\frac{(-1)^s}{s!} \int K^{0(s)}(x)x^s dx = 1$ and $\int |K^{0(s+p)}(x)x^{s+p}|dx < \infty$.

Assumption D. (Assumptions on auxiliary bandwidth λ_n) The sequence λ_n satisfies

(i) $\lambda_n > 0\ \forall\ n \geq 1$ and $\lambda_n \to 0$, as $n \to \infty$.

(ii) $n\lambda_n^{2s+1} \to \infty$ as $n \to \infty$.

(iii) $\limsup_{n\to\infty} n\lambda_n^{2s+2} < \infty$.

(iv) $\liminf_{n\to\infty} n\lambda_n^{2s+1}h_n > 1$.

Assumption E. (Assumptions on given bandwidth h_n) The sequence h_n satisfies

(i) $h_n > 0\ \forall\ n \geq 1$.

(ii) $h_n \to 0$ as $n \to \infty$.

(iii) $nh_n \to \infty$ as $n \to \infty$.

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 are satisfied for any value of s.

(ii) Assumptions $A(i)-(iii)$ 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. In contrast, 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. On the other hand, the assumptions by Hall (1990) or Delaigle and Gijbels (2004) on K prevent the use of a number of popular kernels such as the Gaussian or Gaussian type kernel, as they are neither compactly supported nor do their characteristic functions vanish outside any compact subset of the real line.
We now state our main results. The proofs are given later.

Theorem 1 *Under Assumptions $A-E$, as $n\to\infty$,*
(i) $E\big[M_n^*-M_n\big]^2=o\left(\frac{1}{n^2}\right)+O\left(h_n^{4s+1}\right).$
(ii) $E\left[\frac{M_n^*}{M_n}-1\right]^2=O\left(h_n\right).$

Remark 2 (i) In absence of condition $D(iv)$, M_n^* remains an asymptotically valid estimate and under Assumptions $A-E$,

$$E\big[M_n^*-M_n\big]^2=o\big(\frac{1}{n^2}\big)+O\big(h_n^{4s}r_n\big) \text{ and } E\left[\frac{M_n^*}{M_n}-1\right]^2=O\big(\frac{1}{n\lambda_n^{2s+1}}+h_n\big),$$

where $r_n=\frac{1}{n\lambda_n^{2s+1}}+h_n$.
(ii) Theorem 1(ii) implies that $\frac{M_n^*}{M_n}$ converges to 1 (in L_2 sense) at a rate not slower than $\{h_n\}$. Consequently, 1(ii) ensures the validity of M_n^* and provides an answer to the question: " how large should the sample size be in order that the bootstrap approximation M_n^* is close to M_n? ". Theorem 1(ii) also implies that for any fixed sample size n, M_n^* is expected to be more accurate for smaller values of h_n.

Comparison with other estimators. There are several estimators of MISE available. Here we compare discuss a few of those in relation to our estimator.

(i) **Delaigle and Gijbels' (2004) estimator**: The bias and variance of the Delaigle and Gijbels' estimator for second order deconvolution kernel density estimator are $O(\frac{1}{n}) + o(h_n^4)$ and $o(\frac{1}{n^2} + h_n^8) + o(\frac{h_n^4}{n})$ respectively (see their Theorem 4.1, page 27) which continue to hold good for their smooth bootstrap estimator of M_n in the i.i.d. set up. Let the bias and variance of M_n^* be denoted by B_n and $Var(M_n^*)$ respectively. Since $|B_n| \leq \sqrt{E\left[M_n^* - M_n^2\right]^2}$ and $Var(M_n^*) \leq E\left[M_n^* - M_n\right]^2$, it immediately follows from Theorem 1 (i) that, for any $0 < \delta, \delta' < 1$, $B_n = o(\frac{1}{n} + h_n^{2s+\delta})$ and $Var(M_n^*) = o\left(\frac{1}{n^2} + h_n^{4s+\delta'}\right)$. While the difference of orders, $O(\cdot)$ and $o(\cdot)$ may seem insignificant from a theoretical point of view, it may translate into significant gains in finite samples, just as in the usual bootstrap theory of sample mean type statistics.

(ii) **Taylor's** T_n: For Taylor's estimator T_n, $Var(T_n) = \frac{C}{n^2 h_n} + O(\frac{h_n}{n^2})$, where $C > 0$. Note that if for some $0 < \delta < 1$, $\limsup_{n\to\infty} n h_n^{2s+\delta} < \infty$, then $Var(M_n^*)/Var(T_n) \to 0$ and hence M_n^* has infinite asymptotic relative efficiency in comparison to T_n.

(iii) **Unbiased and biased cross validation estimators** UCV_n **and** BCV_n: These are well known estimators for M_n (see Scott and Terrell (1987)). It is known that (see Theorem 3.1, Scott and Terrell (1987)), $Var(UCV_n) = \frac{C}{n} + O(\frac{1}{n^2 h_n} + \frac{h_n^4}{n})$, where C is a constant. If for some $0 < \delta < 1$, $\limsup_{n\to\infty} n h_n^{4s+\delta} < \infty$ then $Var(M_n^*)/Var(UCV_n) \to 0$ and hence M_n^* has infinite asymptotic relative efficiency in comparison to UCV_n.

Similarly, (see Theorem 3.2, Scott and Terrell (1987)), $Var(BCV_n) = \frac{C'}{n^2 h_n} + O(\frac{h_n}{n^2})$, where C' is a constant. If for some $0 < \delta' < 1$, $\limsup_{n\to\infty} n h_n^{2s+\delta'} < \infty$ then $Var(M_n^*)/Var(BCV_n) \to 0$ and hence M_n^* has infinite asymptotic relative efficiency in comparison to BCV_n.

Further the bias of UCV_n and BCV_n are $-\int f^2(x)dx$ and $O(\frac{1}{n} + h_n^{2s+1})$ respectively. Whereas the bias of M_n^* is $o(\frac{1}{n} + h_n^{2s+\delta})$, is $0 < \delta < 1$. So for large n, M_n^*

has smaller bias than both BCV_n and UCV_n.

(iv) **Plug in estimator**: Park and Marron (1990) compared several data-driven methods for selecting the bandwidth of a kernel density estimator and have recommended the plug-in rule (see Sheather (1986)) to be the best among the methods compared. How does M_n^* compare with the plug-in rule? The basic idea of the plug-in rule is to substitute data based estimates into the asymptotic approximation to M_n. It is well known that if K is of order s then

$M_n = A_n(h_n) + o\big(\frac{1}{nh_n} + h_n^{2s}\big)$ where $A_n(h_n) \equiv A_n$ equals

$$\frac{1}{nh_n}\int K^2(u)du + \frac{a^2}{(s!)^2}h_n^{2s}\int[f^{(s)}(x)]^2dx \;\text{ and }\; a = \int x^s K(x)dx \neq 0.$$

A_n is referred to as the asymptotic mean integrated squared error. In the plug-in method $\int[f^{(s)}(x)]^2dx$ is replaced by a suitable estimator to obtain the following estimator $\hat{A}_n$:

$$\hat{A}_n = \frac{1}{nh_n}\int K^2(u)du + \frac{a^2}{(s!)^2}h_n^{2s}\int\left[K_n^{0(s)}(x)\right]^2dx.$$

Theorem 2 *Under Assumptions A(i) and C(i),*

(i) If $\limsup_{n\to\infty} nh_n^{2s} < \infty$*, then*

$$\liminf_{n\to\infty} nE|\hat{A}_n - M_n| \geq \int f^2 > 0 \;\textit{ and }\; \liminf_{n\to\infty} n^2E(\hat{A}_n - M_n)^2 \geq \Big(\int f^2\Big)^2.$$

(ii) If $\limsup_{n\to\infty} nh_n^{2s} = \infty$*, then* $E(\hat{A}_n - M_n)^2 = o(h_n^{4s})$.

The following result on accuracy of $\hat{A}_n$ and M_n^* follows from Theorems 1 and 2.

Theorem 3 *Under Assumptions* $A - E$*, if* s *is the order of* K*, then*

(i) $\lim_{n\to\infty} nh_n^{2s} < \infty$ *implies,*

$$lim_{n\to\infty}\frac{E(M_n^* - M_n)^2}{E(\hat{A}_n - M_n)^2} = 0 \;\textit{ and }\; \lim_{n\to\infty}\frac{E(\frac{M_n^*}{M_n} - 1)^2}{E(\frac{\hat{A}_n}{M_n} - 1)^2} = 0.$$

(ii) $\lim_{n\to\infty} nh_n^{2s} = \infty$ *implies,*

$$E(M_n^* - M_n)^2 = o(h_n^{4s}) \;\textit{ and }\; E(\hat{A}_n - M_n)^2 = o(h_n^{4s}).$$

Marron and Wand (1992) provides important insight into the effect of h_n on A_n. For fixed n, they plotted M_n and A_n (for Gaussian kernel) against different values of $\log_{10} h_n$. For a wide class of densities, including general normal mixtures, as the value of $\log_{10} h_n$ is increased from -1 to 0 and beyond, A_n diverges to ∞ whereas M_n appears to level off. Moreover, M_n and A_n differ significantly for all values of h_n satisfying $\log_{10} h_n \geq 0$. Consequently, for larger values of h_n, approximation of M_n by $\hat{A}_n$, which in turn is an estimator of A_n, can be poor. Further, A_n is sensitive to the structure of f. In particular, for the double-claw density, which is the density of the distribution $0.49.N(-1,(2/3)^2)+0.49.N(1,(2/3)^2)+0.02.\sum_{l=0}^{6}\frac{1}{350}N((l-3)/2,(0.01)^2)$,[1] A_n is a poor approximation to M_n (see page 725, Figure 4,*loc. cit*). The plug-in estimator $\hat{A}_n$ is not likely to improve these demerits of A_n. So we do not recommend its use to estimate M_n, especially when f is believed to have complicated structures such as multimodality, skewness and existence of spikes.

Fixed Sample performance of M_n^*. What are the effects of different possible choices of h_n and λ_n on the bootstrap estimator M_n^* for *fixed sample sample size n?* The following proposition provides some answers.

Proposition 1 *Suppose f is bounded, continuous and K^0 is a bounded and continuous probability density function. Then for any fixed sample size n, as $h_n \to \infty$,*
(i) for any choice of λ_n, $M_n \to \int f^2(y)dy$ and $M_n^* \to \int (K_n^0(y))^2 dy$ almost surely .
(ii) if $f(x) \to 0$, as $|x| \to \infty$, and $\lambda_n \to \infty$, then , $E(M_n^*) \to 0$. Consequently $E(M_n^*) - M_n \to -\int f^2(y)dy$ and $M_n^* \xrightarrow{P} 0$.
(iii) for any choice of λ_n, $\hat{A}_n \to \infty$ almost surely .

Remark 3 (i) Proposition 1(i) implies that for fixed n, M_n^* and M_n level off and the former succeeds in imitating the behavior of the latter for larger values of h_n. On the other hand, Proposition 1(iii) demonstrates that A_n explodes as h_n is increased, thereby verifying the empirical observation of Marron and Wand (1992).

[1] $N(x, y^2)$ denotes the normal distribution with mean x and variance y^2.

(ii) Proposition 1(ii) implies that, for fixed n if h_n is large, M_n^* is likely to underestimate M_n when λ_n is also large and the structure of $f(\cdot)$ is complicated so that $\int f^2(y)dy$ is large. As a consequence, when h_n is large, Taylor's (1989) choice of $\lambda_n = h_n$, can produce an estimator T_n with substantial negative bias. Proposition 1(ii) also explains the strange behavior of Taylor's (1989) estimator T_n (see Figure 1(a), page 710, *loc. cit.*). As the value of h_n was increased, T_n first decreased and then remained constant, which is very unlikely for an MISE curve.

3 Implementing Smooth Bootstrap

M_n^* does not have a closed form expression in general and hence Monte-Carlo computation is required for its implementation. Marron and Wand (1992) obtained an exact and easily computable expression for MISE, assuming f to be a general normal mixture and Gaussian type kernels are used.

A kernel K is said to be a *Gaussian-based* kernel of order $2r$ if

$$K(x) = \frac{(-1)^r}{2^{r-1}(r-1)!x}\frac{d^{r-1}}{dx^{r-1}}\phi(x)$$

where $\phi(x)$ is standard normal density evaluated at x. Let $\phi_{\sigma^2}(x)$ denote the normal density with mean 0 and variance σ^2, evaluated at x, $\phi^{(s)}_{\sigma^2}(x) = \frac{d^s}{dx^s}\phi_{\sigma^2}(x)$ and

$$C_1(r) = \frac{1}{\sqrt{\pi}}\sum_{i=0}^{r-1}\sum_{j=0}^{r-1}\frac{(2i+2j)!}{2^{3(i+j)+1}i!j!(i+j)!}.$$

Lemma 1 *(Marron and Wand (1992)) If $f(x) = \sum_{i=1}^k w_i\phi_{\sigma_i}(x-\mu_i)$ and if K is a Gaussian-based kernel of order $2r$, then M_n equals*

$$\frac{C_1(r)}{nh_n}+\left(1-\frac{1}{n}\right)\sum_{i=0}^{r-1}\sum_{j=0}^{r-1}\frac{(-1)^{i+j}}{2^{i+j}i!j!}U(h_n;i+j,2)-2\sum_{s=0}^{r-1}\frac{(-1)^s}{2^s\ s!}U(h_n,s,1)+U(h_n,0,0)$$

where $U(h_n,s,q) = \sum_{i=1}^k\sum_{l=1}^k w_iw_lh_n^{2s}\phi^{(2s)}_{\sigma_i^2+\sigma_j^2+qh_n^2}(\mu_i-\mu_l)$.

Now suppose that K is a Gaussian-based kernel of order $2r$ and K_0 is the Gaussian kernel. Since K_0 is Gaussian,

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

This is really a mixed normal distribution. Thus, M_n^* is the MISE of the kernel density estimator $K_{nB}(y)$, where the underlying density $K_n^0(\cdot)$ is a mixed normal density with n components, $w_i = \frac{1}{n}$, $\mu_i = X_i$ and $\sigma_i = \lambda_n$, $i = 1, 2, .., n$. Hence, using the above result, we obtain a closed expression for M_n^*, provided that K is a Gaussian-based kernel of order $2r$.

Theorem 4 *If $K(\cdot)$ is a Gaussian-based kernel of order $2r$ and K_0 is the standard normal density, then M_n^* equals*

$$\frac{C_1(r)}{nh_n}+\left(1-\frac{1}{n}\right)\sum_{i=0}^{r-1}\sum_{j=0}^{r-1}\frac{(-1)^{i+j}}{2^{i+j}i!j!}U(h_n;i+j,2)-2\sum_{s=0}^{r-1}\frac{(-1)^s}{2^s\ s!}U(h_n,s,1)+U(h_n,0,0),$$

where $U(h_n,s,q)=\frac{1}{n^2}\sum_{i=1}^n\sum_{l=1}^n h_n^{2s}\phi^{(2s)}_{2(\lambda_n)^2+qh_n^2}(X_i-X_l)$.

Remark 4 (i) The choice of K^0 as the standard normal density trivially satisfies Assumption C of Section 2 and it also leads to closed form expression for M_n^* when K is a Gaussian or Gaussian type kernel. So these are recommended choices.
(ii) Taylor's (1989) T_n is a special case of M_n^* and is obtained by substituting $r = 1$, $s = 0$ and choosing K as the Gaussian density and $\lambda_n = h_n$ in Theorem 4.

4 Simulation

To what extent does the nature of f and the value of h_n affect the performance of M_n^*, specially when the sample size is small? We probe these questions with simulation. Since any density may be approximated arbitrarily closely in various senses by a normal mixture (see Marron and Wand (1992)), we choose f to be mixed normal. We choose 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 choose K^0 to be Gaussian and hence closed form expression for computing M_n^* and M_n are available from Theorem 4 and Lemma 1.

Since K is standard normal density, $s = 2$. Assumption D in Section 2 provides criterion for choosing λ_n and in our simulation we have used $\lambda_n = \frac{(\log n)^{0.2+\delta}}{n^{0.2}}, 1 >$

$\delta > 0$, which satisfies Assumptions $D\ (i) - (iii)$. Assumption $D(iv)$ is ignored as it is not a necessary condition for the validity of M_n^* and it need not hold for small samples at all.

For $n = 50,\ 500$, we have plotted M_n^* and M_n, against $\log_{10} h_n$. Our choice of $\log_{10}$ scale is motivated by its use by Marron and Wand (1992). For comparison, we have also plotted Taylor's (1989) T_n, which is M_n^* with $\lambda_n = h_n$. We have ignored comparison of M_n^* with UCV_n or BCV_n, since the performance cross-validation method in applications and simulations has been fairly disappointing (see Hall (1992)) and for fixed n, BCV_n converges to zero as $h_n \to \infty$ (see Scott and Terrell (1987)), that is, BCV_n fails to imitate M_n, which converges to $\int f^2(x)dx$, for large values of h_n. The following conclusions may be drawn from these simulations.

(i) Figures 1(a), 2(a), 3(a) reveal excellent performance of M_n^* even for small samples, especially for values of h_n satisfying $\log_{10} h_n \le 0$ and for all types of f. If K is of second order, then natural choices for h_n are $1.06S/n^{1/5}$, where S is the standard deviation (Jones, Marron and Sheather (1996)) and $0.9A/n^{1/5}$, where $A = min\{$sample standard deviation, sample inter quartilerange/1.34$\}$ (Silverman's rule of thumb) etc. All these bandwidths can satisfy $\log_{10} h_n \le 0$ even for moderate sample size. For instance, $\log_{10} h_n < 0$, where $h_n = 1.06S/n^{1/5}$, for the data set on three-month certificate of deposit rates (Simonoff 1996, page 1). Hence for this data set we can expect M_n^* to be accurate for the Gaussian kernel.

(ii) When $\log_{10} h_n > 0$, M_n^* can be sensitive to the choice of λ_n. For instance, from Figures 1(a) to 3(b), we observe that for values h_n satisfying $\log_{10} h_n > 0$, Taylor's T_n (which is M_n^*, with $\lambda_n = h_n$) can be a terrible estimate of M_n.

Acknowledgement.. The authors gratefully acknowledge the detailed comments of Dr. Anil Ghosh, Stat-Math Unit, Indian Statistical Institute. These comments have been of significant help.

5 Proof of Theorems

First we state and prove the following Lemmas which are used in the proof of Theorem 1.

Lemma 2 *Under Assumptions **A**(ii), **C**(ii) and **D**, as* $n \to \infty$,

(i) $E\left[\int\{K_n^{o(s)}(y)\}^2 dy - \int\{f^{(s)}(y)\}^2 dy)^2\right]^2 = O\left(\frac{1}{(n\lambda_n^{2s+1})^2}\right)$.

(ii) $E\left|\int\{K_n^{o(s)}(y)\}^2 dy - \int\{f^{(s)}(y)\}^2 dy\right| = o(1)$.

Let us introduce the following notation. Let $*$ denote convolution and let

$$K_{\lambda_n}^{0(s)}(y) = \frac{1}{\lambda_n^{s+1}} K^{0(s)}\Big(\frac{y}{\lambda_n}\Big), \quad \theta_{sn} = \frac{1}{n(n-1)} \sum\sum_{i \neq j} K_{\lambda_n}^{0(s)} * K_{\lambda_n}^{0(s)}(X_i - X_j),$$

$$\theta = \int (f^{(s)}(y))^2 dy,\ \theta_n = \int \{K_n^{0(s)}(y)\}^2 dy = \frac{1}{n\lambda_n^{2s+1}} \int \{K^{0(s)}(v)\}^2 dv + \frac{(n-1)}{n}\theta_{sn}.$$

Proof of Lemma **2** With the above notation, for suitable constants, A_1, A_2, A_3, $E\big[\int\{K_n^{0(s)}(y)\}^2 dy - \int\{f^{(s)}(y)\}^2 dy\big]^2$ equals

$$\begin{aligned} &E\big[\frac{1}{n\lambda_n^{2s+1}} \int \{K^{0(s)}(v)\}^2 dv + \frac{n-1}{n}\theta_{sn} - \theta\big]^2 \\ &\leq \frac{A_1}{(n\lambda_n^{2s+1})^2} + A_2\big(1 - \frac{1}{n}\big)^2 E\big[\theta_{sn} - \theta\big]^2 + \frac{A_3}{n^2}. \end{aligned}$$

Under Assumptions $A(ii),\ C(ii),\ D(ii)$, it follows from Lemma 3.1 (b) and (d) and Hall and Marron (1987) that as $n \to \infty$,

$$E\,[\theta_{sn} - \theta]^2 = O\big(\frac{1}{n^2\lambda_n^{4s+2}} + \lambda_n^2\big).$$

Therefore under the Assumptions $A(ii),\ C(ii),\ D(ii)$ we get

$$E\big[\int \{K_n^{o(s)}(y)\}^2 dy - \int \{f^{(s)}(y)\}^2 dy\big]^2 = O\big(\frac{1}{(n\lambda_n^{2s+1})^2}\big).$$

This proves (i). To prove (ii), note that

$$E\big|\int \left\{K_n^{0(s)}(y)\right\}^2 dy - \int \left\{f^{(s)}(y)\right\}^2 dy\big| = E\,|\theta_n - \theta|$$

and

$$E\,|\theta_n - \theta| \leq \frac{1}{n\lambda_n^{2s+1}} \int (K_n^{0(s)}(v))^2 dv + E|\theta_{sn} - \theta| + \frac{1}{n} E|\theta_{sn}|.$$

Under Assumptions $A(ii),\ C(ii)$ it follows from Lemma 3.1(d) of Hall and Marron (1987) that $E\,|\theta_{sn}-\theta| = o(1)$. Now, using Assumption D(ii), part (ii) follows, proving the Lemma completely. $\square$

The following Lemma is used in the proof of Lemma 4. Let

$$T_2 = \int \{f^{(s)}(y)\}^2 dy \quad \text{and} \quad C_1 = \frac{1}{s}\int |K(u)u^s| du.$$

Lemma 3 *Under Assumptions **A-C**, as* $n \to \infty$*,*

$$\int \left\{ \int \int_0^1 (1-t)^{s-1} K(u) u^s E K_n^{0(s)}(y - t h_n u) dt du \right\}^2 dy = C_1^2 T_2 + O\left(h_n + \lambda_n^p\right).$$

***Proof of Lemma* 3** Let

$$e_n = \int \left\{ \int \int_0^1 (1-t)^{s-1} K(u) u^s E[K_n^{0(s)}(y - t h_n u)] dt du \right\}^2 dy.$$

It is easy to see that for each y, t and u,

$$E[K_n^{0(s)}(y - t h_n u)] = \frac{1}{(\lambda_n^0)^s} \int K^{0(s)}(z) f(y - t h_n u - \lambda_n^0 z) dz.$$

Under the Assumptions $C(ii)$ (c) and (d) on $K^{0(s)}$ and Assumption A on $f(\cdot)$, using Taylor's expansion with integral remainder,

$$\begin{aligned}
e_n &= \int \Big\{ (-1)^s \int \int_0^1 (1-t)^{s-1} K(u) u^s f^{(s)}(y - t h_n u) dt du \\
&\quad + (-1)^{s+p} \frac{\lambda_n^p}{(s+p-1)!} g_n(y) \Big\}^2 dy \\
&= \int \Big\{ \int \int_0^1 (1-t)^{s-1} K(u) u^s f^{(s)}(y - t h_n u) dt du \Big\}^2 dy \\
&\quad + \frac{\lambda_n^{2p}}{((s+p-1)!)^2} \int g_n^2(y) dy \\
&\quad + 2\frac{(-\lambda_n)^p}{(s+p-1)!} \int \Big\{ \int \int_0^1 (1-t)^{s-1} K(u) u^s f^{(s)}(y - t h_n u) dt du \Big\} g_n(y) dy \\
&= t_{1n} + t_{2n} + t_{3n} \quad \text{(say)}
\end{aligned}$$

where $g_n(y)$ equals

$$\int K(u) u^s \int_0^1 (1-t)^{s-1} \Big\{ \int K^{0(s)}(v) v^{s+p} \int_0^1 (1-t')^{s+p-1} f^{(s+p)}(y^*) dt' dv \Big\} dt du$$

and $y^* = y - tuh_n - t'\lambda_n v$. Under Assumptions $A(i)$ on $f^{(s)}$, for suitable constants C_1', C_2',

$$|t_{3n}| \leq \lambda_n^p C_1' \int |g_n(y)| dy \leq \lambda_n^p C_2' \int |f^{(s+p)}(y)| dy = O(\lambda_n^p).$$

Using Cauchy-Schwartz inequality it is easy to see that, $t_{2n} \leq C_1^2 \lambda_n^{2p} \int \{f^{(s+p)}(y)\}^2 dy$. Using Assumptions $A\ (i), (ii)$ on $f^{(s)}$, it is easy to see that $t_{1n} = C_1^2 T_2 + O(h_n)$. Therefore, combining these estimates,

$$e_n = C_1^2 T_2 + O(h_n) + O(\lambda_n^{2p}) + O(\lambda_n^p) = C_1^2 T_2 + O(h_n + \lambda_n^p).$$

Hence Lemma 3 is proved completely. □

Lemma 4 *Under Assumptions **A-D**, as* $n \to \infty$,

$$2C_1^4 \int E\{K_n^{0(s)}(y)\}^2 dy \int \{f^{(s)}(y)\}^2 dy - 2E(c_n).d_n = O\left(h_n\right)$$

where

$$E(c_n) = \int E[\left\{\int \int_0^1 (1-t)^{s-1} K(u) u^s K_n^{0(s)}(y - th_n u) dt du\right\}^2] dy,$$

and

$$d_n = \int \left\{\int \int_0^1 (1-t)^{s-1} K(u) u^s f^{(s)}(y - th_n u) dt du\right\}^2 dy.$$

***Proof of Lemma* 4** Let

$$e_n = \int \left\{\int \int_0^1 (1-t)^{s-1} K(u) u^s E(K_n^{0(s)}(y - th_n u)) dt du\right\}^2 dy.$$

Using Cauchy-Schwartz inequality it is easy to see that

$$B_n = 2C_1^4 \int E\{K_n^{0(s)}(y)\}^2 dy \int \{f^{(s)}(y)\}^2 dy \geq 2Ec_n.d_n$$

Therefore $\quad 0 \leq B_n - 2E(c_n).d_n \leq B_n - 2e_n.d_n.$

From Lemma 2(i), under the Assumptions $A(ii),\ C(ii)$ and D

$$\begin{aligned} T_{1n} &= \int E\{K_n^{0(s)}(y)\}^2 dy = \int \{f^{(s)}(y)\}^2 dy + O\left(\frac{1}{n\lambda_n^{2s+1}}\right) \\ &= T_2 + O\left(\frac{1}{n\lambda_n^{2s+1}}\right) \ \text{(say)}. \end{aligned}$$

Under Assumption A on $f(\cdot)$, $d_n = C_1^2 T_2 + O(h_n)$. Under the Assumptions $C(ii)$ (c) and (d) on $K^{0(s)}$ and Assumption A on $f(\cdot)$, from Lemma 3, $e_n = C_1^2 T_2 + O(h_n + \lambda_n^p)$. Substituting the above equations in right side of (1) and under Assumptions D on λ_n we get, as $n \to \infty$

$$\begin{aligned} 0 \leq B_n - 2E(c_n).d_n &\leq 2C_1^4 \big[T_2 + O(\frac{1}{n\lambda_n^{2s+1}})\big] T_2 \\ &\quad -2\left[C_1^2 T_2 + O\left(h_n + \lambda_n^p\right)\right]\left[C_1^2 T_2 + O(h_n)\right] \\ &= O\left(\frac{1}{n\lambda_n^{2s+1}}\right) + O\left(h_n + \lambda_n^p\right) + O(h_n) = O\left(h_n\right). \end{aligned}$$

Hence Lemma 4 is proved completely. □

Proof of Theorem 1 Recall that

$$\begin{aligned} M_n &= \int V(K_n(y))dy + \int \left[E(K_n(y)) - f(y)\right]^2 dy, \\ \text{and } M_n^* &= \int V_n(K_{nB}(y))dy + \int \left[E_n(K_{nB}(y)) - K_n^0(y)\right]^2 dy \end{aligned}$$

where

$$V(K_n(y)) = E\big[K_n(y) - E(K_n(y))\big]^2 \text{ and } V_n(K_{nB}(y)) = E_n\big[K_{nB}(y) - E(K_{nB}(y))\big]^2.$$

The expectations E and E_n are computed with respect to $f(\cdot)$ and the bootstrap density $K_n^0(y)$ respectively. So, almost surely,

$$\begin{aligned} \left|M_n^* - M_n\right| &\leq \Big|\int \left[V_n(K_{nB}(y)) - V(K_n(y))\right]dy\Big| \qquad (1) \\ &\quad + \Big|\int \left[E_n(K_{nB}(y)) - K_n^0(y)\right]^2 dy - \int \left[E(K_n(y)) - f(y)\right]^2 dy\Big|. \\ &= L_{1n} + L_{2n} \quad \text{(say)}. \qquad (2) \end{aligned}$$

Squaring and taking expectation on either side of (1),

$$E|M_n^* - M_n|^2 \leq 2E(L_{1n}^2) + 2E(L_{2n}^2). \qquad (3)$$

Now under Assumptions A, B and E, from Rao (1983, pp. 45), we have

$$\int V(K_n(y))dy = \frac{1}{nh_n}\int K^2(v)dv - \frac{1}{n}\int \Big\{\int K(v)f(y - h_n v)dv\Big\}^2 dy. \qquad (4)$$

Repeating those arguments, we have

$$\int V_n(K_{nB}(y))dy = \frac{1}{nh_n}\int K^2(v)dv - \frac{1}{n}\int\Big\{\int K(v)K_n^0(y-h_nv)dv\Big\}^2 dy. \quad \text{(a.s.)} \tag{5}$$

Therefore from Equations (4) and (5) we get, almost surely,

$$\begin{aligned} L_{1n} &= \frac{1}{n}\Big|\int\big\{\int K(v)K_n^0(y-h_nv)dv\big\}^2 dy - \int\big\{\int K(v)f(y-h_nv)dv\big\}^2 dy\Big| \\ &\leq \frac{1}{n}\int\Big|\big\{\int K(v)K_n^0(y-h_nv)dv\big\}^2 - \big\{\int K(v)f(y-h_nv)dv\big\}^2\Big|dy. \end{aligned} \tag{6}$$

Using $|a^2-b^2| = (a+b)|a-b|$, for $a,\ b>0$, the right side of above inequality is dominated by (writing $y^* = y - h_nv$)

$$\frac{1}{n}\int\big[\big\{\int K(v)(K_n^0(y^*)+f(y^*))dv\big\}\big\{\int K(v)|K_n^0(y^*)-f(y^*)|dv\big\}\big]dy$$

$= D_n$, say. We note that, almost surely

$$K_n^0(y-h_nv) + f(y-h_nv) \leq |K_n^0(y-h_nv) - f(y-h_nv)| + 2||f||.$$

Therefore, almost surely

$$\begin{aligned} D_n &\leq \frac{1}{n}\Big[\int\{\int K(v)|K_n^0(y-h_nv) - f(y-h_nv)|dv\}^2 dy \\ &\qquad + 2||f||\int K(v)\int|K_n^0(y-h_nv) - f(y-h_nv)|dydv\Big] \\ &\leq \frac{1}{n}\Big[\int\int K(v)\{K_n^0(y-h_nv) - f(y-h_nv)\}^2 dvdy \\ &\qquad + 2||f||\int|K_n^0(y) - f(y)|dy\Big] \\ &= \frac{1}{n}\Big[\int\{K_n^0(y)-f(y)\}^2 dy + 2||f||\int|K_n^0(y)-f(y)|dy\Big]. \end{aligned}$$

Hence from (6) we have, almost surely

$$L_{1n} \leq \frac{1}{n}\Big[\int\big\{K_n^0(y)-f(y)\big\}^2 dy + 2||f||\int|K_n^0(y)-f(y)|dy\Big].$$

This implies

$$\begin{aligned} E(L_{1n}^2) &\leq \frac{1}{n^2}E\Big[\int\big\{K_n^0(y)-f(y)\big\}^2 dy + 2||f||\int|K_n^0(y)-f(y)|dy\Big]^2 \\ &\leq \frac{2}{n^2}\Big[E\Big[\int\big\{K_n^0(y)-f(y)\big\}^2 dy\Big]^2 + 4||f||^2 E\Big[\int|K_n^0(y)-f(y)|dy\Big]^2\Big]. \end{aligned} \tag{7}$$

Since K^0 is a density function and $nh_n \to \infty$, therefore if follows from Devroye (1983) that, almost surely, $\int |K_n^0(y) - f(y)|dy = o(1)$. Since both f and K^0 are probability density functions, therefore $\int |K_n^0(y) - f(y)|dy \leq 2 \,\forall\, n$. Therefore, by DCT,

$$4||f||^2 E\left[\int |K_n^0(y) - f(y)|dy\right]^2 = o(1).$$

Further under Assumptions $A,\ C$ on f and K^0 $E\big[\int \{K_n^0(y) - f(y)\}^2 dy\big]^2$ equals

$$\begin{aligned} &E\Big[\int \{K_n^0(y)\}^2 dy - \int f^2(y)dy + 2\int f(y)\{f(y) - K_n^0(y)\}dy\Big]^2 \\ \leq\ & 2E\Big[\int \{K_n^0(y)\}^2 dy - \int \{f(y)\}^2 dy\Big]^2 + 8||f|| \int E\{f(y) - K_n^0(y)\}^2 dy. \end{aligned}$$

Under the Assumptions $A(i),\ C(i),\ D(i),(ii)$, from Rao (1983, pp. 45),

$$\int E\{f(y) - K_n^0(y)\}^2 dy = o(1).$$

Under Assumptions $A(i),(ii),C(ii)$ and D , from Hall and Marron (1987) we get

$$E\left[\int \{K_n^0(y)\}^2 dy - \int \{f(y)\}^2 dy\right]^2 = o(1).$$

Therefore under Assumptions $A - E$,

$$E\left[\int \{K_n^0(y) - f(y)\}^2 dy\right]^2 = o(1).$$

Hence from (7) we find that, under Assumptions $A - E$,

$$E(L_{1n}^2) = o\left(\frac{1}{n^2}\right). \tag{8}$$

So in view of equations (3) and (8), to prove Theorem 1(i), it is enough to show that

$$\frac{1}{h_n^{4s+1}} E(L_{2n}^2) = O(1). \tag{9}$$

Now under the smoothness Assumptions $A,\ C$ on f and K^0, using Taylor's expansion with integral remainder we get

$$\begin{aligned} L_{2n} \ =\ & \frac{h_n^{2s}}{((s-1)!)^2}\Big[\int \left\{\int\int_0^1 (1-t)^{s-1} K(u) u^s K_n^{0(s)}(y - th_n u) dt du\right\}^2 dy \\ & - \int \left\{\int\int_0^1 (1-t)^{s-1} K(u) u^s f^{(s)}(y - th_n u) dt du\right\}^2 dy\Big] \end{aligned}$$

$$= \quad \frac{h_n^{2s}}{((s-1)!)^2}\left[c_n - d_n\right] \quad \text{(say)}.$$

Therefore,

$$\begin{aligned}
E(L_{2n}^2) \;&\le\; \frac{h_n^{4s}C_1^4}{((s-1)!)^4}\big[E\big[\int \{K_n^{0(s)}(y)\}^2 dy\big]^2 + \big(\int \{f^{(s)}(y)\}^2 dy\big)^2 - 2Ec_n.d_n\big] \\
&\le\; \frac{h_n^{4s}}{((s-1)!)^4}\big[C_1^4 E\Big[\int \{K_n^{0(s)}(y)\}^2 dy - \int \{f^{(s)}(y)\}^2 dy\Big]^2 \\
&\qquad +2C_1^4 \int E\{K_n^{0(s)}(y)\}^2 dy \int \{f^{(s)}(y)\}^2 dy - 2Ec_n.d_n\big]. \qquad (10)
\end{aligned}$$

Under the Assumptions A, C and D, from Lemma 2(i),

$$E\left[\int \{K_n^{0(s)}(y)\}^2 dy - \int \{f^{(s)}(y)\}^2 dy\right]^2 = O\left(\frac{1}{(n\lambda_n^{2s+1})^2}\right).$$

Under Assumptions A to D, from Lemma 4,

$$2C_1^4 \int E\{K_n^{0(s)}(y)\}^2 dy \int \{f^{(s)}(y)\}^2 dy - 2Ec_n.d_n = O\left(h_n\right).$$

Therefore under Assumptions A to D, substituting the above equations in right side of (10),

$$\frac{1}{h_n^{4s}} E(L_{2n}^2) = O\left(\frac{1}{(n\lambda_n^{2s+1})^2} + h_n\right) = O(h_n).$$

This establishes (9) and recalling (8) part (i) is proved.

We now prove part (ii). From Prakasa Rao (1983, pp. 45, Theorem 2.1.7), if $f(\cdot)$ has bounded continuous derivatives of sth order, $s \ge 2$, the sth derivative $f^{(s)}$ is square integrable and $K(\cdot)$ bounded, symmetric (about zero), square integrable and of sth order then

$$M_n = \frac{1}{nh_n}\int K^2(u)du + \frac{a^2}{(s!)^2}(h_n)^{2s}\int [f^{(s)}(x)]^2 dx + o\left(\frac{1}{nh_n} + h_n^{2s}\right),$$

where $a = \int x^s K(x)dx \neq 0$.

Using part (i) of Theorem 1,

$$E\left[\frac{M_n^*}{M_n} - 1\right]^2 = \frac{E[M_n^* - M_n]^2}{M_n^2} = \frac{o(\frac{1}{n^2}) + O(h_n^{4s+1})}{\frac{D_1}{(nh_n)^2} + D_2 h_n^{4s} + o\left(\frac{1}{(nh_n)^2} + h_n^{4s}\right)}. \qquad (11)$$

where D_1, D_2 are positive constants, independent of n.

If $\limsup_{n\to\infty} nh_n^{2s+1} < \infty$ then, dividing numerator and denominator on the right side of (11) by $(nh_n)^2$ we get $E(\frac{M_n^*}{M_n} - 1)^2 = o(h_n^2) + O(h_n)$ as $n \to \infty$.

If $\lim_{n\to\infty} nh_n^{2s+1} = \infty$ then, dividing numerator and denominator on the right side of (11) by h_n^{4s} we get $E(\frac{M_n^*}{M_n} - 1)^2 = o(h_n^2) + O(h_n)$ as $n \to \infty$.

Hence, $E(\frac{M_n^*}{M_n} - 1)^2 = O(h_n)$ as $n \to \infty$. and Theorem 1 is proved completely.

□

Proof of Theorem 2 Recall that

$$\hat{A}_n = \frac{1}{nh_n}\int K^2(u)du + \frac{a^2}{(s!)^2}h_n^{2s}\int \left\{K_n^{0(s)}(x)\right\}^2 dx \tag{12}$$

where $a = \int x^s K(x)dx$. Using Rao (1983, page 45),

$$M_n = \frac{1}{nh_n}\int K^2 - \frac{1}{n}\int \left\{\int K(v)f(y-h_nv)dv\right\}^2 dy + \int \left\{E(K_n(y) - f(y)\right\}^2 dy. \tag{13}$$

From (12) and (13),

$$\begin{aligned} E\left|\hat{A}_n - M_n\right| \ge E(\hat{A}_n - M_n) = \frac{1}{n}\int \left\{\int K(v)f(y-h_nv)dv\right\}^2 dy \\ + \frac{a^2}{(s!)^2}h_n^{2s}E\left[\int \left\{K_n^{0(s)}(x)\right\}^2 dx\right] - \int \left\{E(K_n(y) - f(y)\right\}^2 dy. \end{aligned} \tag{14}$$

Now from Rao (1983, page 45),

$$\int \left\{E(K_n(y) - f(y)\right\}^2 dy = \frac{a^2}{(s!)^2}h_n^{2s}\int \left\{f^{(s)}(x)\right\}^2 dx + o(h_n^{2s}).$$

and from Lemma 2,

$$\frac{a^2}{(s!)^2}h_n^{2s}E\left[\int \left\{K_n^{0(s)}(x)\right\}^2 dx\right] = \frac{a^2}{(s!)^2}h_n^{2s}\int \left\{f^{(s)}(x)\right\}^2 dx + o(h_n^{2s}).$$

Consequently

$$\frac{a^2}{(s!)^2}h_n^{2s}E\left[\int \left\{K_n^{0(s)}(x)\right\}^2 dx\right] - \int \left\{E(K_n(y) - f(y)\right\}^2 dy = o(h_n^{2s}). \tag{15}$$

So if $\{h_n\}$ satisfies $\lim_{n\to\infty} nh_n^{2s} < \infty$, then from (14) and (15) we get

$$\begin{aligned} \liminf_{n\to\infty} nE|\hat{A}_n - M_n| &\ge \liminf_{n\to\infty} \int \left\{\int K(v)f(y-h_nv)dv\right\}^2 dy \\ &\ge \int \liminf_{n\to\infty} \left\{\int K(v)f(y-h_nv)dv\right\}^2 dy. \end{aligned} \tag{16}$$

Since under Assumptions A and B, f is bounded continuous and $\int K(x)dx = 1$, applying DCT to the right side of (16), the first part of Theorem 2 (i) is proved. Since $n^2 E\big(\hat{A}_n - M_n\big)^2 \geq \big(nE|\hat{A}_n - M_n|\big)^2$, the second part of Theorem 2 (i) is a direct consequence. The proof of part (ii) is easy. $\square$

***Proof of Proposition* 1** We recall that

$$\begin{aligned} E(M_n^*) &= \int E\left[V_n(K_{nB}(y))\right] dy + \int E\left[E_n\left(K_{nB}(y)\right) - K_n^0(y)\right]^2 dy \\ &= a_n + b_n \quad \text{(say)}. \end{aligned}$$

It is easy to see that $V_n(K_{nB}(y)) \leq E_n[K_{nB}^2(y)] = \frac{1}{nh_n}\int K^2(u)K_n^0(y - h_n u)du$. Taking expectation on both sides,

$$E\big[V_n(K_{nB}(y)\big] \leq \frac{1}{nh_n}\int K^2(u)\int K^0(v)f(y - h_n u - \lambda_n v)dvdu.$$

Therefore, integrating (with respect to y) both sides of the above inequality, we get

$$a_n = \int E\big[V_n(K_{nB}(y)\big]dy \leq \frac{1}{nh_n}\int K^2(u)du$$

Further by Cauchy-Schwartz inequality we get

$$\begin{aligned} \big[E_n(K_{nB}(y) - K_n^0(y)\big]^2 &= \left[\int K(u)(K_n^0(y - h_n u) - K_n^0(y))du\right]^2 \\ &\leq \int |K(u)|du. \int |K(u)|\left[K_n^0(y - h_n u) - K_n^0(y)\right]^2 du. \end{aligned}$$

So, writing $T_1 = \int |K(u)|du$,

$$\begin{aligned} b_n &= \int E\left[E_n\left(K_{nB}(y)\right) - K_n^0(y)\right]^2 dy \\ &\leq T_1 \int\int |K(u)| E\left[K_n^0(y - h_n u) - K_n^0(y)\right]^2 dudy \\ &= \frac{2T_1^2}{n\lambda_n}\int \left\{K^0(v)\right\}^2 dv + 2T_1^2\left(1 - \frac{1}{n}\right)\int\left\{\int K^0(u)f(y - \lambda_n u)du\right\}^2 dy \\ &\quad -\frac{2T_1}{n\lambda_n}\int f(y)dy\int |K(u)|\int K^0(x)K^0\left(x - \frac{h_n u}{\lambda_n}\right)dxdu \\ &\quad -2T_1\big(1 - \frac{1}{n}\big)\int\int |K(u)|\int\int K^0(z_1)K^0(z_2)f(y_1^*)f(y_2^*)dz_1dz_2dudy. \end{aligned}$$

where $T_1 = \int |K(u)|du$, $y_1^* = y - \lambda_n u - h_n z_1$ and $y_2^* = y - \lambda_n z_2$.

Letting $h_n \to \infty$ and $\lambda_n \to \infty$, it is easy to see (using the condition that $|f(x)| \to 0$ as $|x| \to \infty$ and DCT) that $a_n, b_n \to 0$ and hence $E(M_n^*) \to 0$, proving first part of (ii). Since M_n^* is a nonnegative random variable, the second part is immediate consequence. Parts (i) and (iii) are easy to prove. $\square$

References

[1] Faraway, J.J. and Jhun, M. (1990). Bootstrap choice of bandwidth for density estimation. *J. Amer. Statist. Assoc.*, 85, No. 412, pp. 1119-1122.

[2] Delaigle, A. and Gijbels, I. (2004). Bootstrap bandwidth selection in kernel density estimation from a contaminated sample. *Ann. Inst., Statist. Math.*, 56, No. 1, pp. 19-47.

[3] Hall, P. and Marron, J.S. (1987). Estimation of integrated squared density derivatives. *Statistics and Probability Letters*, 6, pp. 109-115

[4] 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.

[5] Hall, P., Marron, J.S. and Park, B.U. (1992). Smoothed cross-validation. *Probab Theory Relat Fields*, 92,pp. 1-20.

[6] Jones, M.C., Marron, J.S. and Sheather, S.J. (1996). A brief survey of bandwidth selection for density estimation. *J. Amer. Statist. Assoc.*, 91 401-407.

[7] Marron, J.S. and Park, B.U. (1990). Comparison of data-driven bandwidth selectors *J. Amer. Statist. Assoc.*, 85, No. 409, pp 66-72.

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

[9] Marron, J.S. and Chung, S.S. (1997). Presentation of Smoothers: The Family Approach, unpublished manuscript.

[10] Prakasa Rao, B.L.S. (1983). *Nonparametric Functional Estimation*. Academic Press, Inc, New York.

[11] Scott, D.W. and Terrell, G.R. (1987). Biased and Unbiased Cross-Validation in Density Estimation, *Journal of the American Statistical Association*, 82, No. 400, Theory and Methods.

[12] Sheather, S.J. (2004). Density estimation, *Statistical Science*, 19, No. 4, pp. 588-597.

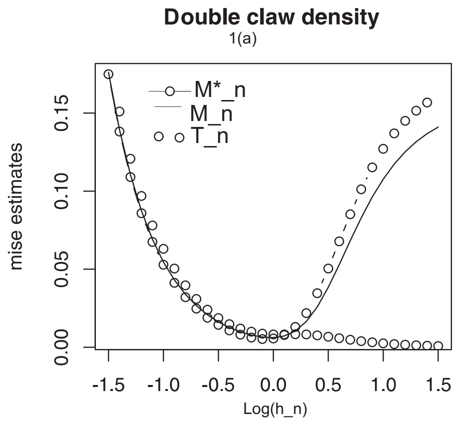
[13] Sheather, S.J. (1986). An improved data-based algorithm for choosing the window width when estimating the density at a point. *Computational Statistics and Data Analysis*, 4, pp. 61-65.

[14] Silverman, B.W. (1986). *Density Estimation for Statistics and Data Analysis*. Monographs on Statistics and Applied Probability. Chapman and Hall, London.

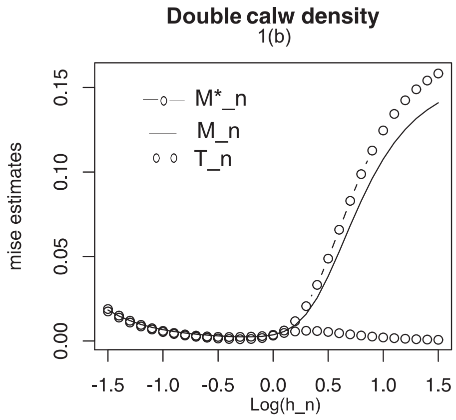
[15] Simonoff, J.S. (1996). *Smoothing Methods in Statistics*, Springer-Verlag, New York.

[16] Taylor, C.C. (1989). Bootstrap choice of the smoothing parameter in kernel density estimation. *Biometrika*, 76, No.4, pp. 705-12.

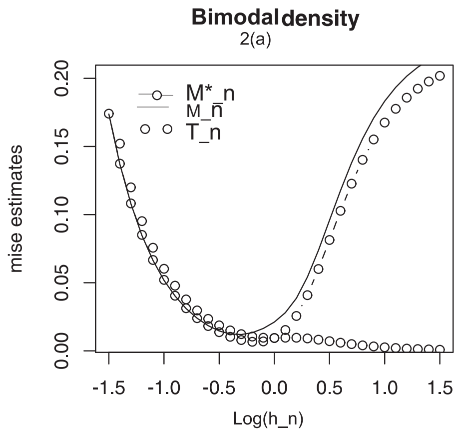
In figures $1(a) - 3(b)$ we plot $T_n \equiv T_n(h_n),\ M_n^* \equiv M_n^*(h_n)$ and $M_n \equiv M_n(h_n)$ against $\text{Log}(h_n) \equiv log_{10}h_n$ for double-claw, bimodal and normal distributions and for sample sizes $n = 50$ and 500. Both K and K^0 are standard normal densities.



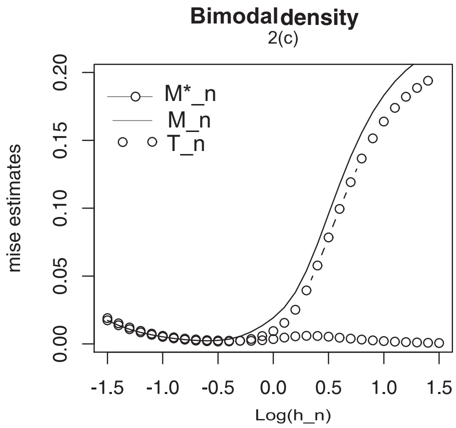
sample size 50



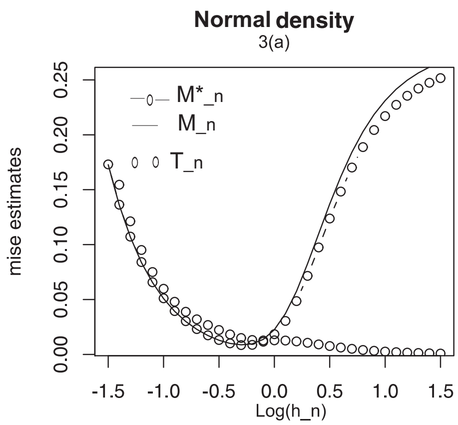
sample size 500



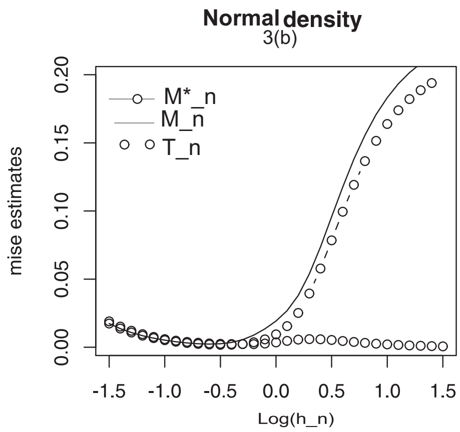
sample size 50



sample size 500



sample size 50



sample size 500