Journal of the Nigerian Association of Mathematical Physics, Volume 9 (November 2005)

Higher order values of MISE in kernel density estimation

F. O. Oyegue and J. I. Odiase,
Department of Mathematics, University of Benin
Benin City, Nigeria

Abstract

Density estimation is the general approach adopted for the construction of an estimate of the underlying density function for independent and identically distributed random variables. There is need to reduce the error propagation in Kernel Density Estimation. Higher Order values of the exact and asymptotic mean integrated squared error of some kernels are considered. An empirical verification that EMISE is less than AMISE for higher order normal mixture densities is explored through computer algorithms. The effect of small and large values of the window width is examined over equally spaced grid of (**0, 0.5**].

Keywords: Higher order kernels, Exact and Asymptotic MISE, window width, Algorithm

pp 343 - 350

1.0 Introduction

Density estimation is the general approach that is adopted for the construction of an estimate $\hat{f}$ of the underlying density function f for X_1, X_2, …, X_n iid random variables. There abound in the statistical literature, many methods of density estimation, but we shall exploit the properties of Kernel Density estimation in this work. The properties of kernel density estimation are now well understood and moreover, it has a wide range of applications. An important hurdle to the practical use of kernel density estimation is the selection of the smoothing parameter or the window width, *h*. We acknowledge that many data based methods have been proposed for the selection of the window width, *h*. See Scott (1985) [8], Hall and Marron (1988) [2], Shealther and Jones (1991) [9], Hardle (1990) [4] and Wand, Marron and Ruppert (1991) [10]. However, some challenges especially in the area of the measurement of the exact mean integrated squared error (EMISE) still exist.

In recent researches, attention is being focused on the need to reduce error propagation in kernel density estimation. The focus is on the reduction in the bias – term of the mean integrated squared error. While Marron and Wand (1992) [6] showed the exact mean integrated squared error is strictly less than the asymptotic mean integrated squared error (i.e. EMISE < AMISE) at least for the non –negative (i.e. $p = 2$) kernels, we still observed the issues of the behaviour of EMISE and AMISE for higher order (i.e. $p = 4, 6, 8, 10, \ldots$) kernels have not been explored viz–a–viz the value of the window width. In this work, the higher order nature of the values of the EMISE and AMISE for some normal mixture densities in relation to the values of the window width *h* is investigated.

Specifically, the results obtained (through simulation) show that the exact mean integrated squared error (EMISE) is always less than the asymptotic mean integrated squared error (AMISE) for all orders (of the bias). Moreover, for small window width *h*, the EMISE and the AMISE are very close for all orders. But for increasing values of h, the discrepancy between EMISE and AMISE become more pronounced for all higher orders ($p = 4, 6, 8, 10$). These significant results pointed to the conclusion that the mean integrated squared error varies considerably across kernel order $p = 2r$, at different values of window width, h, over an equally spaced grid of (0, 0.5]

The results obtained gave the chance to compare EMISE and AMISE. One way one can view the normal mixture densities is that any density can be approximated arbitrary closely in various senses by a normal mixture. In this paper, three normal mixture densities (i.e. bimodal, skewed bimodal and the separated bimodal) which represent important but simple departure from the unimodal are considered. The reason for this choice of densities is that since they are mildly multimodal, then one might be able to estimate them fairly well with a data set of moderate size.

2.0 Exact and asymptotic mean integrated squared error

Let X_1, X_2, …, X_n be independent observation from an unknown density f on the real line. The traditional nonparametric kernel density estimation is given by

$$\hat{f}(x) = \frac{1}{n}\sum k_h(x - x_i) \tag{2.1}$$

where $k_h(z) = h^{-1}k(h^{-1}z)$ is a kernel function which is taken to be symmetric probability density satisfying $\int k = 1$ and $\int x^j \; k(x)dx = \begin{cases} 0, & j = 1, \dots, p-1 \\ V_p(k) \neq 0, & j = p \end{cases}$

This last expression defines $V(k)$. The integer p is called the order of the kernel and symmetric of k implies that p is an even integer. The integer $\frac{p}{2}$ will be denoted by r (i.e. $r = \frac{p}{2}$)

Rosenblatt (1956) [7] estimation $\hat{f}(x)$ of f (x) using a kernel k and window width h has the mean integrated squared error.

$$MISE(\hat{f}) = E\left[\int \left(\hat{f}(x) - f(x)\right)^2 dx\right] \tag{2.2}$$

and by simple manipulation leads to

$$MISE = h^{-1}h^{-1}\int k^2 + \left(1 - h^{-1}\right)\int (k_h * f)^2 - 2\int (k_h * f)f + \int f^2 \tag{2.3}$$

where (*) stands for convolution and the integrals are from $-\infty$ to ∞ with respect to x.

The result of Theorem (2.1) of Marron and Wand (1992) [6] shows that if f is the normal mixture density expressed as:

$$f(x) = \sum_{i=1}^{m} w_i \; \phi_{\sigma_i}(x - \mu_i) \tag{2.4}$$

where $-\infty < \mu_i < \infty$, $\sigma_i > 0$, $(w_i, \dots, w_m)$ is a vector with positive entries summing to unity, and k is the $(2r)^{th}$ order Gaussian – based kernel defined by

$$G_{2r}(x) = \frac{(-1)^r \; \phi^{(2r-1)}(x)}{2^{r-1}(r-1)!(x)} = \sum_{s=0}^{r-1} \frac{(-1)^s}{2^s \; s!} \phi^{(2s)}(x) \tag{2.5}$$

then

$$EMISE = \frac{L_1(r)}{nh} + \left(1 - \tfrac{1}{n}\right)\sum_{s=0}^{r-1}\sum_{s'=0}^{r-1} \frac{(-1)^{s+s'}}{2^{s+s'} \; s!\,s'!} v(h, \; s+s', 2)$$

$$- 2\sum_{s=0}^{r-1} \frac{(-1)^s}{2^{s+s'} \; s!} \; v(h; \; s, 1) + v(h; 0, 0) \tag{2.6}$$

where

$$v(h;s,q)=\sum_{i=1}^{m}\sum_{i'=1}^{m} w_i w_{i'}\, h^{2s}\, \phi_{\sigma_{ii'q}}\ (\mu_i-\mu_{i'}) \tag{2.7}$$

and
$$\sigma_{ii'q}=\left(\sigma_i^2+\sigma_{i'}^2+qh^2\right)^{1/2} \tag{2.8}$$

equation (2.6) can be written as:

$$EMISE=\frac{L_1(r)}{nh}+\left(1-\tfrac{1}{n}\right)\sum_{s=0}^{r-1}\sum_{s'=0}^{r-1}\sum_{i=1}^{m}\sum_{i'=1}^{m}\frac{(-1)^{s+s'}}{2^{s+s'}\ s!s'!}w_i w_{i'}$$
$$\times h^{2s+2s'}\,\phi^{(2s+2s')}_{\left(2h^2+\sigma_i^2+\sigma_{i'}^2\right)^{1/2}}\ (\mu_i-\mu_{i'})-2\sum_{i=1}^{m}\sum_{i'=1}^{m}\sum_{s=0}^{r-1}\frac{(-1)^s}{2^s\ s!}w_i w_{i'}$$
$$\times h^{2s}\,\phi^{(2s')}_{\left(h^2+2\sigma_i^2\right)^{1/2}}\ (\mu_i-\mu_{i'})+\sum_{i=1}^{m}\sum_{i'=1}^{m}w_i w_{i'}\phi_{\left(\sigma_i^2+\sigma_{i'}^2\right)^{1/2}}(\mu_i-\mu_{i'}) \tag{2.9}$$

where
$$L_1(r)=\frac{1}{\pi^{1/2}}\sum_{s=0}^{r-1}\sum_{s'=0}^{r-1}\frac{(2s+2s')!}{2^{3s+3s'+1}\ s!s'!(s-s')!} \tag{2.10}$$

Equation (2.9) resulted from combing equation (2.5) with the following results from Aldersholf et al. (1991) [1]. That is, for $\sigma,\ \sigma'\ >\ 0$ and $r, r'=0,1,2,\ldots$

(i)
$$\int\phi_\sigma^{(r)}(x-\mu)\ \phi_{\sigma'}^{(r')}(x-\mu')\,dx=(-1)^r\ \phi_{\hat\sigma}^{(r+r')}(\mu-\mu') \tag{2.11}$$

where $\hat\sigma=\left(\sigma^2+\sigma'^2\right)^{1/2}$ and

(ii)
$$\phi_\sigma^{(2r)}(0)=(-1)^r\,2^{-(2r+1)/2}\,\pi^{-1/2}(2r)!(r!)^{-1}\sigma^{-(2r+1)} \tag{2.12}$$

On the other hand, Marron and Wand (1992) showed that the asymptotic mean integrated squared error is

$$\text{AMISE} = n^{-1}h^{-1}L_1(r) + h^{4r}L_2(r) \tag{2.13}$$

where
$$L_{2(r)}=\frac{1}{2^{2r}(r!)^2}\sum_{i=1}^{m}\sum_{i'=1}^{m}w_i w_{i'}\,\phi_{ii'}^{(4r)}(\mu_i-\mu_{i'}) \tag{2.14}$$

and
$$\theta_{ii'}=\left(\sigma_i^2+\sigma_{i'}^2\right)^{1/2} \tag{2.15}$$

$L_1(r)$ is as defined in equation (2.10).

Now, as an example, consider one of the normal mixture test function (i.e. the bimodal) which is given as:
$$\frac{1}{2}N\left(-1,\left(\tfrac{2}{3}\right)^2\right)+\frac{1}{2}N\left(1,\left(\tfrac{2}{3}\right)^2\right)$$

If we considered p = 4, (i.e. the order of the kernel), this implies r =2. We now read from the kernel given: m = 2 (no. of terms in the kernel), the values of ω = ½ , ½ , the values of $\mu=1,1$ and the values of $\sigma=\tfrac{2}{2},\ \tfrac{2}{3}$. All these values are substituted into the equations (2.9) and (2.13) to obtain the exact and asymptotic mean integrated squared error values.

The notation for the derivatives of scaled version of the standard normal distribution, ϕ, is given as

$$\phi_{\sigma}^{(s)}(x) = \frac{d^s}{dx^s}\phi_{\sigma}(x) = \sigma^{-(s+1)}\phi^{(s)}(x/\sigma)$$

The derivatives of the standard normal density, $\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$, which is embedded in (2.9) and (2.13) are as follows:

$$\phi^{(1)}(x) = \frac{-xe^{-\frac{1}{2}x^2}}{\sqrt{2\pi}}$$

$$\phi^{(2)}(x) = \frac{1}{\sqrt{2\pi}}\left(x^2 e^{-\frac{1}{2}x^2} - e^{-\frac{1}{2}x^2}\right),$$

$$\phi^{(3)}(x) = \frac{1}{\sqrt{2\pi}}\left(-x^3 e^{-\frac{1}{2}x^2} + 3xe^{-\frac{1}{2}x^2}\right)$$

$$\vdots$$

$$\phi^{(20)}(x) = \frac{1}{\sqrt{2\pi}}\left(-x^{20}e^{-\frac{1}{2}x^2} - 190x^{18}e^{-\frac{1}{2}x^2} + 14535x^{16}e^{-\frac{1}{2}x^2} - 581400x^{14}e^{-\frac{1}{2}x^2} + 13226850x^{12}e^{-\frac{1}{2}x^2}\right.$$
$$-174594420x^{10}e^{-\frac{1}{2}x^2} + 1309458150x^8 e^{-\frac{1}{2}x^2} - 523783200x^6 e^{-\frac{1}{2}x^2}$$
$$\left.+9820936125x^4 e^{-\frac{1}{2}x^2} - 6547290750x^2 e^{-\frac{1}{2}x^2} + 654729075e^{-\frac{1}{2}x^2}\right)$$

The differentiation of the standard normal density, $\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$, is carried out, up to the twentieth term.

2.1 Algorithm for EMISE and AMISE

The Algorithm (MISE) implements the differentiation of the standard normal density, $\phi(x)$ and subsequently generates the values for EMISE and AMISE for the normal mixture densities considered in this paper.

2.1.1 Algorithm (MISE)

Choose a Normal Mixture Density
Supply Mean, Standard Deviation, Probability, $(\mu_i, \sigma_i, \omega_i)$
(These are the values that determine the type of Normal mixture density)
$h \leftarrow$ *Window width*
$n \leftarrow$ *Sample size*
$m \leftarrow$ *Number of terms in the Normal mixture*
$p \leftarrow 2$ *(Order)*
$r \leftarrow p/2$

For $s \leftarrow 0, r\text{-}1$ ***do***
  Compute s!
  For $s' \leftarrow 0, r\text{-}1$ ***do***
    Compute s'!
    For $l \leftarrow 1, m$ ***do***
      For $l' \leftarrow 1, m$ ***do***

        Compute scaled version of Standard Normal ($\phi_\sigma^{(s)}(x)$ *)*
        Compute derivatives of ϕ *up to the 20th derivative)*
      End do
    End do
  End do
End do
Compute EMISE based on $\phi_\sigma^{(s)}(x)$
For $l \leftarrow 1, m$ ***do***
  For $l' \leftarrow 1, m$ ***do***
    Compute $\phi_\sigma^{(s)}(x)$ *scaled version of Standard Normal*
    Compute derivatives of ϕ *up to the 20th derivative*
  End do
End do
Compute AMISE based on $\phi_\sigma^{(s)}(x)$

3.0 Results of Simulation

Now, in order to obtain higher order values (i.e. for p = 4, 6, 8, 10 …) of the exact and asymptotic mean integrated squared error for the normal mixture densities, we first obtain up to the twentieth (or beyond) derivatives of the standard normal density. Implement the Algorithm (MISE) to realise the values of AMISE and EMISE for the different kernels at the higher orders (p = 4, 6, 8, 10).

A sample of size n = 100 was simulated at random for each of the normal mixture densities while a window width h was chosen subjectively over an equally spaced grid of (0, 0.5]. Thus, for effective comparison, all the conditions of the problem being solved were made to exactly be the same so that any observed differences whatsoever would be attributed to the functional nature of EMISE and AMISE. The normal mixture densities used in this work are: Separated Bimodal, Skewed Bimodal and Bimodal. The results of our simulation exercise are given in the graphs (pictures) of the EMISE and AMISE for the higher order values of p = 2, 4, 6, 8, 10. Figures 1 through 3 show pictorially how EMISE and AMISE depend on the order p = 2, 4, 6, 8, 10 for the normal mixture densities, where the samples size is n = 100 for various values of the window width.

Table 1: Separated Bimodal

	H-value							
	0.05		**0.20**		**0.35**		**0.50**	
Order	EMISE	AMISE	EMISE	AMISE	EMISE	AMISE	EMISE	AMISE
2	0.05361	0.05641	0.01262	0.01547	0.01352	0.02187	0.02564	0.05900
4	0.09237	0.09519	0.02100	0.02382	0.01340	0.01527	0.01010	0.03858
6	0.12195	0.12477	0.02872	0.03119	0.01506	0.01810	0.01035	0.03238
8	0.14672	0.14954	0.03456	0.03739	0.01855	0.02141	0.01288	0.02996
10	0.16844	0.17126	0.03999	0.04282	0.02165	0.02448	0.01434	0.02949

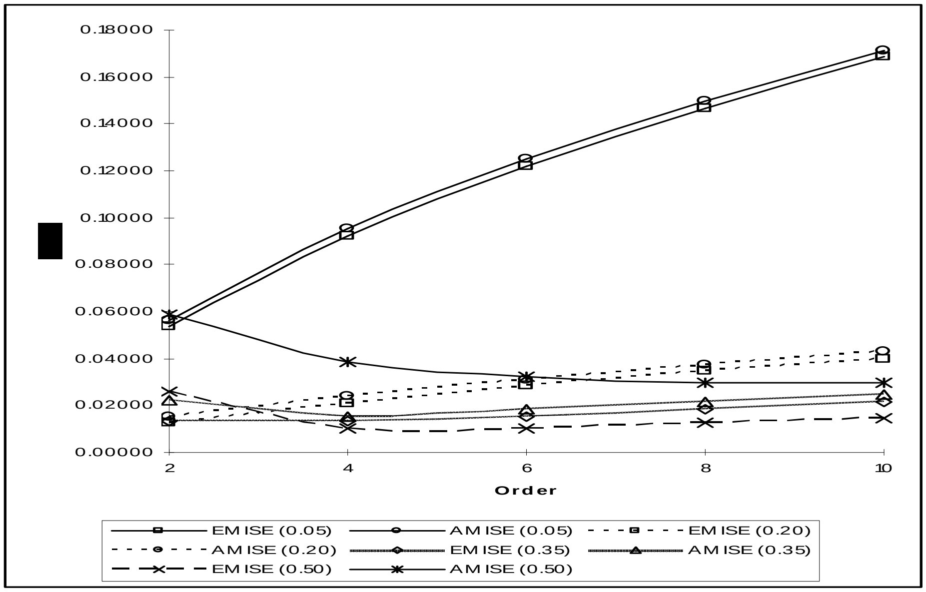

Figure 1: MISE for Separated Bimodal Distribution

Table 2: Skewed Bimodal

	H-value							
	0.05		**0.20**		**0.35**		**0.50**	
Order	EMISE	AMISE	EMISE	AMISE	EMISE	AMISE	EMISE	AMISE
2	0.05379	0.05641	0.01239	0.01535	0.00984	0.01877	0.01238	0.05440
4	0.09256	0.09579	0.02121	0.02389	0.01180	0.02160	0.01015	0.14835
6	0.12214	0.12477	0.02856	0.03120	0.01537	0.02452	0.01112	0.49618
8	0.14691	0.14954	0.03475	0.03739	0.01877	0.02756	0.01284	1.88016
10	0.16863	0.17126	0.04018	0.04282	0.02185	0.03055	0.01471	7.64201

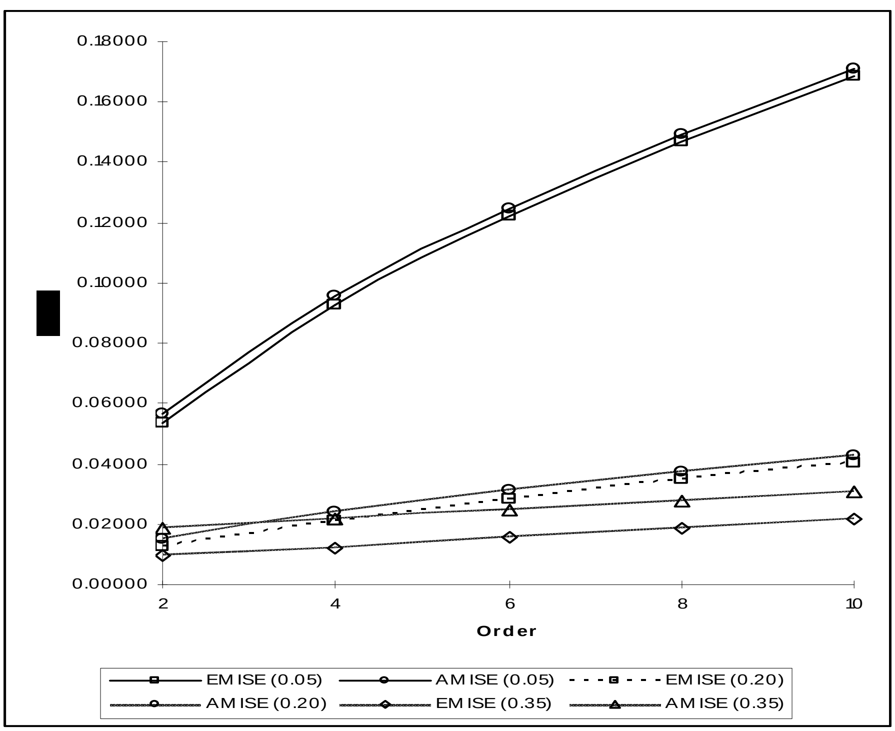

Figure 2: MISE for Skewed Bimodal Distribution

Table 3: Bimodal

	H-value							
	0.05		**0.20**		**0.35**		**0.50**	
Order	EMISE	AMISE	EMISE	AMISE	EMISE	AMISE	EMISE	AMISE
2	0.05407	0.05641	0.01206	0.01438	0.00757	0.01067	0.00838	0.01654
4	0.09285	0.09519	0.02146	0.02380	0.01136	0.01376	0.00791	0.01240
6	0.12243	0.12477	0.02885	0.03119	0.01549	0.01783	0.01023	0.01302
8	0.14720	0.14954	0.03505	0.03739	0.01903	0.02136	0.01263	0.01504
10	0.16892	0.17126	0.04048	0.04282	0.02213	0.02447	0.01479	0.01715

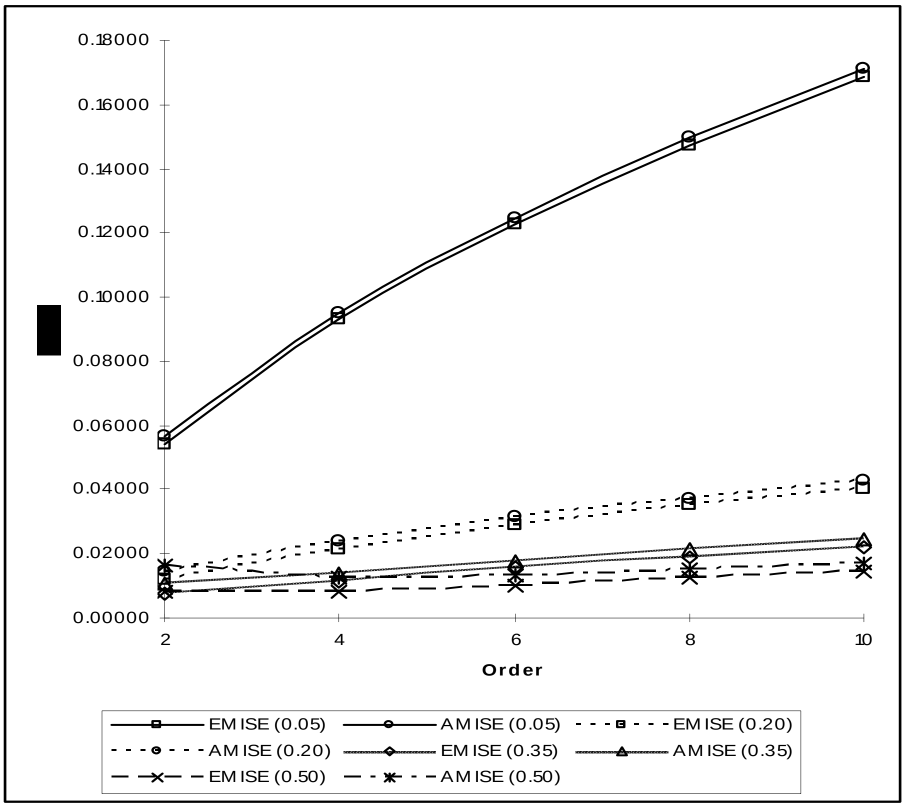

Figure 3: MISE for Bimodal Distribution

Observe that for all orders and all values of the window width h, the values of AMISE are consistently higher than those of EMISE. This is true for all the normal mixture densities considered in this paper. The larger the values of the window width, the higher the discrepancy between EMISE and AMISE.

4.0 Conclusion

In this paper, the higher order values of the EMISE and the AMISE for some normal mixture densities were examined. According to Marron and Wand (1992) [6] the effectiveness of higher order kernels depend on the sample size of the distribution. In fact the work of Marron and Wand (1992) [10], Jones and Signorini (1997) [5] and Hansen (2003) [3] could not recommend the use of higher order kernels in practice due to over dependency on sample size. The issue of sample size was not considered as a problem in this work since it was possible for us to simulate 100 observations for each of the densities used. Thus, in our situation, all conditions of the problem being examined are the same for the calculations involved in the estimation of the EMISE and AMISE. Any observed difference or deviation is taken to have resulted from the functional nature of EMISE or AMISE. This study benefits from the fundamental methodology of the Kernel Density Estimation.

As an illustration, Figures 1 to 3 give the graphs of EMISE and AMISE values against higher orders (p = 2, 4, 6, 8, 10) for different values of window width h for some normal mixture densities. Clearly, the higher order value of the mean integrated squared error is always less than the asymptotic mean integrated squared error. Thus, our simulation results confirms that EMISE < AMISE for all window width h greater than zero (h > 0) within the grid of (0, 0.5]. An important gain arising from this study is that in all situations examined the EMISE is lower in error propagation.

Moreover, the simulation results reveal that for small values of h, EMISE ≈ AMISE. This is true for $h \le 0.35$ for the Skewed Bimodal. The same characteristic was shared for the Separated Bimodal and the Bimodal when $h \le 0.5$. In the normal mixture densities considered, as h moves closer to 0.5, the discrepancy between the EMISE and the AMISE becomes more pronounced. The Separate Bimodal and Skewed Bimodal exhibited this nature when $h \le 0.5$. Nevertheless all the densities tended to have maintained relative closeness when h = 0.05. Hence, the mathematical difficulty in the derivation of EMISE pays off with lower values than the AMISE values. For large values of the window width, it is not advisable to use the asymptotic method.

References

[1] Aldershof, B., Marron, J. S., Park, B. U. and Wand, M. P. (1991). Facts about the Gaussian probability density function. Unpublished manuscript.

[2] Hall, P. and Marron, J. S. (1988). Variable window width kernel estimation of probability densities, *Probability Theory and Related Field*, 80, 37-49.

[3] Hansen, B. E (2003). Exact mean integrated error of higher order kernel estimators, University of Wisconsin, Madison, WI 53706, USA. Online at http//www.ssc.wisc.edu/~bhansen.

[4] Hardle, W. (1990). *Smoothing techniques with implementation*. Springer Verlag, New York.

[5] Jones, M. C. and Signorini, D. F. (1997). A comparison of higher order bias kernel density estimators. *Journal of the American Statistical Association*, 92, 1063-1073.

[6] Marron, J. S. and Wand, M. P. (1992). Exact mean integrated square error. *The Annals of Statistics*, 20, 712-713.

[7] Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function. *Annals of Mathematical Statistics*, 27, 832-837.

[8] Scott, D. W. (1985). Average shifted histograms: Effective nonparametric density estimators in several dimensions. *The Annals of Statistics*, 13, 1024-1040.

[9] Shealter, S. J. and Jones, M. C. (1991). A reliable data-base band with selection method for kernel density estimation. *Journal of the Royal Statistical Society, ser. B*, 53, 683-690.

[10] Wand, M. P., Marron, J. S. and Ruppert, D. (1991). Transformations in density estimation with comments. *Journal of the American Statistical Association*, 86, 343-736.