Chiang Mai J. Sci. 2011; 38(1) : 1-12
www.science.cmu.ac.th/journal-science/josci.html
Contributed Paper

Probability Density Estimation Using Two New Kernel Functions

Manachai Rodchuen*[a], and Prachoom Suwattee [b]
[a] Department of Statistics, Faculty of Science, Chiang Mai University, Chiang Mai,50200, Thailand.
[b] School of Applied Statistics, National Institute of Development Administration, Bangkapi, Bangkok 10240, Thailand.
*Author for correspondence; e-mail: manachai@chiangmai.ac.th

Received: 20 March 2010
Accepted: 24 June 2010

ABSTRACT

This paper considers two new kernel estimators of a density function $f(x)$. The errors of the estimators are measured by the mean squared error ($MSE(\hat{f}(x,\underset{\sim}{X}))$) and the mean integrated squared error ($MISE(\hat{f})$). The estimates of these error measures are also given. The estimators of $MSE(\hat{f}(x,\underset{\sim}{X}))$ and $MISE(\hat{f})$ are found to be asymptotically unbiased. Properties of the proposed estimators depend on the corresponding kernel functions used to derive them together with their bandwidths. The bandwidths used for comparison of the properties are the Silverman rule of thump (SRT), two-stage direct plug-in (DPI) and the solve-the-equation (STE) bandwidths. A simulation study is carried out to compare the *AMISE* of the estimates with those of uniform, Epanechnikov and Gaussian kernel functions. For data with outlier and bimodal distributions, the proposed estimates perform better than the uniform and Gaussian estimates. One of the proposed kernel estimates with STE bandwidth performs well when data are with a strongly skewed distribution. This estimates with SRT bandwidth performs well when data are skewed bimodal with small sample size. For data with claw distribution, the estimate with SRT bandwidth is better than the others. The same results hold when the STE bandwidth is used with large sample sizes. For data distributed as discrete comb, one of the proposed estimates with STE bandwidth performs better than the others. Another proposed kernel estimate also performs better than the uniform and Gaussian estimate.

Keywords: density estimation, kernel estimator, mean squared error, mean integrated squared error.

This paper was presented at The 5th IMT-GT International Conference on Mathematics, Statistics, and their Applications 2009 (ICMSA 2009), West Sumatra, June 9 - 11, 2009 Bukittinggi, Indonesia, on Estimators of Errors of Kernel Density Estimator and International Conference on Statistics and Information Analytics (ICSIA-2010), January 11 - 13, 2010 Chennai, India, on Comparison of Kernel Functions and Bandwidth Selection in Kernel Desity Estimation.

1. INTRODUCTION

In statistical inference, both estimation and testing of hypotheses, we need to know the population probability density function $f(x)$ of a random variable X. If we neglect this assumption, then the results will have low reliability. In practice, if we do not know the probability density function of X, an estimate is needed. Density estimation is similar to the estimation of any function. There are two approaches to estimate an unknown density function, the parametric and nonparametric approaches. Choosing an approach depends on prior knowledge of $f(x)$ of the random variable X in a population. If we know the form of $f(x)$ such as normality, then the parametric approach is appropriate. In the contrary case, the nonparametric approach is more appropriate. Applications of density estimation are given in many articles, for example in Sheather [1].

For nonparametric density estimation, one well-known method is the use of a kernel function which is introduced by Rosenblatt [2], and Parzen [3]. Let $\underset{\sim}{X}=(X_1, ..., X_n)$ be a random sample of size n from a population with an unknown probability density function $f(x)$, and let $\underset{\sim}{x}=(x_1,.., x_n)$ be the sample points on $\underset{\sim}{X}$. The kernel density estimator of $f(x)$ at the point x_0 is given by

$$\hat{f}(x_0,\underset{\sim}{X})=\frac{1}{nh}\sum_{i=1}^{n}K(\frac{x_0-X_i}{h}), \qquad (1)$$

where $K(u)$ is a kernel real valued function, $u = (x - X_i)/h$ is in its support and h a positive real number, called *the bandwidth* or *window width of* X_i[2,3]. The properties of the density estimator depend on the properties of the kernel function $K(u)$ and the bandwidth h used. If $K(u)$ is a probability density function, then all estimates $(\hat{f}(x,\underset{\sim}{X})$ of the form (1) are also density functions. Usually, but not always, $K(u)$ will be a symmetric unimodal density function. A kernel is said to be of order p for some $p \geq 2$ if

$$\int u^j K(u)du = \begin{cases} 1, & j=0, \\ 0, & j=1,...,p-1, \\ \mu_p, & j=p. \end{cases} \qquad (2)$$

If the kernel is of order greater than 2, then the density estimator may be negative at some points. The kernel function $K(u)$ should satisfy the properties:

i) $K(u)$ is nonnegative real valued function and continuous on its support,
ii) $\int K(u)du = 1$,
iii) $K(u)$ is symmetric about 0, which implies $\mu_1 = \int uK(u)du = 0$,
iv) $\mu_2 = \int u^2K(u)du < \infty$ i.e. μ_2 is finite.

There are many kernel functions which satisfy the above properties such as uniform, $K_u(u) = 0.5I_{[-1,1]}(u)$, Epanechnikov, $K_E(u) = 0.75(1-u^2)I_{[-1,1]}(u)$,Gaussian, $K_G(u) = e^{-u^2/2}/\sqrt{2\pi}$.

The measure of the errors of the estimators such as the variance or mean-squared error of the estimate depends on the density function $f(x)$. Marron and Wand [4] computed the exact mean integrated squared error of the kernel density estimator using Gaussian kernel to estimate a general normal mixture density $f(x)$. In general, the bias of an estimator is unknown. So it should be estimated from a sample data. Hall [5] stated that to estimate the bias of a density estimator one would typically use a function of the estimate of $f''(x)$. Scott and Terrell [6] suggested an estimator of the squared L_2 norm of $f''(x)$, $R(f'') = \int (f''(x))^2\,dx$, by $\hat{R}(f'') = R(\hat{f}'') - \hat{R}(K'')/nh^5$, where $\hat{f}''(x,\underset{\sim}{X})$ is the second derivative of the kernel density estimator.

In this paper the estimators of some measures of errors of $\hat{f}(x,\underset{\sim}{X})$ are studied. Asymptotically unbiased estimators of $\hat{f}''(x)$ suggested by Jones [7] and of $R(f'')$ suggested by Scott and Terrell [6] are used to construct an estimator of the measures of errors. The $AMISE(\hat{f})$ depends on its bandwidth, $R(f)$ and $R(f'')$. $R(f'')$ is a measure of total curvature which is increasing with its skewness, kurtosis and multimodality [8]. A numerical method for comparing the *AMISE* ($\hat{f}$) in various populations is given later.

2. Estimators of errors of kernel estimators

Assume that $f(x)$ is continuous and squared integrable, having second derivative with respect to x at x_0. The kernel density estimator with $K(u)$ is asymptotically unbiased having bandwidth $h = h(n) \longrightarrow 0$ as $n \longrightarrow \infty$. The bias of $\hat{f}(x_0,\underset{\sim}{X})$ is

$$B(\hat{f}(x_0,\underset{\sim}{X})) = \frac{h^2}{2} f''(x_0)\mu_2 + o(h^2), \quad (3)$$

where $\mu_2 = \int u^2K(u)du$, *the kernel variance*, [9].

The r-th derivative of $\hat{f}(x,\underset{\sim}{X})$ is equivalent to the r-th derivative of $K(u)$, $K^{(r)}(u)$ when $u = (x\text{-}X_i)/h$. The r-th derivative of the estimator of $f(x)$, $\hat{f}^{(r)}(x_0,\underset{\sim}{X})$, is given by

$$\hat{f}^{(r)}(x_0,\underset{\sim}{X}) = \frac{1}{nh^{r+1}}\sum_{i=1}^{n} K^{(r)}(\frac{x_0 - X_i}{h}), \quad (4)$$

[7]. $K^{(r)}(u)$ which gives the best estimator $\hat{f}^{(r)}(x_0,\underset{\sim}{X})$ of $f^{(r)}(x_0)$ is not necessary the r-th derivative of the optimal kernel function [10].

Jones [7] pointed out that not only $K(u)$ is piecewise differentiable on its support but also $K''(u)$ should not be a constant throughout that support. Jones [7] constructed the form of the r-th derivative of $K(u)$ that is a function of $K(u)$ satisfing the required property.

Jones [7] and Muller [11] stated that in general, $K^{(r)}(u)$ should have the property that

$$\mu_{(r),j} = \int u^j K^{(r)}(u)du = \begin{cases} 0, & 0 \le j \le r+1, \\ (-1)^r r!, & j = r, \\ \mu_{(r),j+2} < \infty, & j = r+2. \end{cases} \quad (5)$$

Jones [7] showed that the estimator $\hat{f}^{(r)}(x_0,\underset{\sim}{X})$ obtained from $K^{(r)}(u)$ satisfying the above property is better than the direct derivative of $\hat{f}(x_0,\underset{\sim}{X})$.

The asymptotically unbiased estimator of $f''(x_0)$ proposed by Jones [7] depends on $K^{(2)}(u)$, where $K^{(2)}(u) = 2(u^2\text{-}\mu_2)K(u)/(\mu_4\text{-}\mu_2^2)$, $\mu_4 = \int u^4K(u)du$. Thus, we can estimate $B(\hat{f}(x_0,\underset{\sim}{X}))$ by

$$\hat{B}(\hat{f}(x_0,\underset{\sim}{x})) = \frac{h^2}{2}\hat{f}''(x_0,\underset{\sim}{x})\mu_2. \quad (6)$$

Theorem 2.1 *$\hat{B}(\hat{f}(x_0,\underset{\sim}{X}))$ is asymptotically unbiased estimator of $B(\hat{f}(x_0,\underset{\sim}{X}))$, when $h=h(n)\longrightarrow 0$ as $n\longrightarrow\infty$.*

Proof

$$E[\hat{B}(\hat{f}(x_0,\underset{\sim}{X}))] = E[\frac{h^2}{2}\hat{f}''(x_0,\underset{\sim}{X})\mu_2]. \quad (7)$$

Setting $u = \frac{x_0 - x}{h}$, we have $x = x_0\text{-}hu$ and by Taylor series expansion of $f(x_0\text{-}hu)$ around x_0, (7) becomes

$$E[\hat{B}(\hat{f}(x_0,\underset{\sim}{X}))] = \frac{h^2 f''(x_0)\mu_2}{2} + o(h^2)$$

Thus,

$$B(\hat{B}(\hat{f}(x_0,\underset{\sim}{X}))) = E[\hat{B}(\hat{f}(x_0,\underset{\sim}{X}))] - B(\hat{f}(x_0,\underset{\sim}{X})) = o(h^2).$$

■

The variance of $\hat{f}(x_0,\underset{\sim}{X})$ is

$$V(\hat{f}(x_0,\underset{\sim}{X})) = V[\frac{1}{nh}\sum_{i=1}^{n} K(\frac{x_0 - X_i}{h})] = \frac{1}{nh^2}V[K(\frac{x_0 - X}{h})] = \frac{1}{n}\{\int \frac{1}{h^2}K^2(\frac{x_0 - x}{h})f(x)dx - [\int \frac{1}{h}K(\frac{x_0 - x}{h})f(x)dx]^2\}. \quad (8)$$

By Taylor series expansion of $f(x)$ about x_0 and let $x=x_0\text{-}hu$, $V(\hat{f}(x_0,\underset{\sim}{X}))$ can be expressed in terms of $f''(x)$ as

$$V(\hat{f}(x_0,\underset{\sim}{X})) = \frac{f(x_0)R(K)}{nh} - \frac{(f(x_0))^2}{n} + \frac{hf''(x_0)\int t^2K^2(u)du}{2n} + o(n^{-1}), \quad (9)$$

where $R(K) = \int K^2(u)du$ is *the squared L_2 norm of* $K(u)$ called *"the roughness of K(u)"* [12] (L_2 is the set of all Lebesque measurable real valued functions, $g(t)$, such that $\int g^2(t)dt$ exists and finite.). The variance of $\hat{f}(x_0, \underset{\sim}{X})$ converges to 0 if the bandwidth $h = h(n)$ is such that $nh \longrightarrow \infty$ as $n \longrightarrow \infty$.

An estimator of $B(\hat{f}(x_0, \underset{\sim}{X}))$ obtained from the estimator of $f''(x)$ can be used to create an estimator of $V(\hat{f}(x_0, \underset{\sim}{X}))$.

Thus, we can estimate $V(\hat{f}(x_0, \underset{\sim}{X}))$ by

$$\hat{V}(\hat{f}(x_0, \underset{\sim}{X})) = \frac{\hat{f}(x_0, \underset{\sim}{X})R(K)}{nh} - \frac{(\hat{f}(x_0, \underset{\sim}{X}))^2}{n} + \frac{h\hat{f}''(x_0, \underset{\sim}{X})\int t^2K^2(u)du}{2n}. \tag{10}$$

Theorem 2.2 *$\hat{V}(\hat{f}(x_0, \underset{\sim}{X}))$ as in (10) is asymptotically unbiased for $V(\hat{f}(x_0, \underset{\sim}{X}))$, when $h=h(n) \longrightarrow 0$ as $n \longrightarrow \infty$.*

Proof

Since

$$E(\hat{f}(x_0, \underset{\sim}{X})) = f(x_0) + \frac{h^2}{2} f''(x_0)\mu_2 + o(h^2),$$

$$E((\hat{f}(x_0, \underset{\sim}{X}))^2) = \frac{f(x_0)R(K)}{nh} + (f(x_0))^2 + o(h)$$

$$E(\hat{f}''(x_0, \underset{\sim}{X})) = f''(x_0) + \frac{h^4 f^{(4)}(x_0)}{12(\mu_4 - \mu_2^2)}(\mu_6 - \mu_2\mu_4) + o(h^2)$$

Thus,

$$E[\hat{V}(\hat{f}(x_0, \underset{\sim}{X}))] = \frac{f(x_0)R(K)}{nh} - \frac{(f(x_0))^2}{n} + \frac{hf''(x_0)\int u^2K^2(u)du}{2n} + o(h).$$

Hence,

$$B(\hat{V}(\hat{f}(x_0, \underset{\sim}{X}))) = E[\hat{V}(\hat{f}(x_0, \underset{\sim}{X}))] - V(\hat{f}(x_0, \underset{\sim}{X})) = o(h).$$ ∎

Since the $MSE(\hat{f}(x_0, \underset{\sim}{X}))$ can be written in the from

$$MSE(\hat{f}(x_0, \underset{\sim}{X})) = V(\hat{f}(x_0, \underset{\sim}{X})) + B^2(\hat{f}(x_0, \underset{\sim}{X})) \tag{11}$$

thus

$$\hat{MSE}(\hat{f}(x_0, \underset{\sim}{X})) = \hat{V}(\hat{f}(x_0, \underset{\sim}{X})) + \hat{B}^2(\hat{f}(x_0, \underset{\sim}{X}))$$

$$= \frac{\hat{f}(x_0, \underset{\sim}{X})R(K)}{nh} + \frac{h\hat{f}''(x_0, \underset{\sim}{X})\int t^2K^2(u)du}{2n} - \frac{(\hat{f}(x_0, \underset{\sim}{X}))^2}{n} + \frac{h^4}{4}(\hat{f}''(x_0, \underset{\sim}{X}))^2\mu_2^2. \tag{12}$$

Theorem 2.3 *$\hat{MSE}(\hat{f}(x_0, \underset{\sim}{X}))$ is asymptotically unbiased estimator of $MSE(\hat{f}(x_0, \underset{\sim}{X}))$, when $h=h(n) \longrightarrow 0$ as $n \longrightarrow \infty$.*

Proof

By theorem 2.1 and 2.2

$$E[\hat{MSE}(\hat{f}(x_0, \underset{\sim}{X}))] = E[\hat{V}(\hat{f}(x_0, \underset{\sim}{X}))] + \hat{B}^2(\hat{f}(x_0, \underset{\sim}{X}))]$$

$$= V(\hat{f}(x_0, \underset{\sim}{X})) + o(h) + E[\hat{B}^2(\hat{f}(x_0, \underset{\sim}{X}))]$$

$$= V(\hat{f}(x_0, \underset{\sim}{X})) + o(h) + \frac{h^4\mu_2^2}{4}E[(\hat{f}''(x_0, \underset{\sim}{X}))^2]$$

$$= V(\hat{f}(x_0, \underset{\sim}{X})) + o(h) + \frac{h^4\mu_2^2}{4}\left[\frac{f(x_0)\int (K^{(2)}(u))^2du}{nh} + (f''(x_0))^2 + o(h)\right]$$

$$= V(\hat{f}(x_0, \underset{\sim}{X})) + \frac{h^4\mu_2^2(f''(x_0))^2}{4} + o(h)$$

$$= V(\hat{f}(x_0, \underset{\sim}{X})) + B^2(\hat{f}(x_0, \underset{\sim}{X})) + o(h)$$

$$= MSE(\hat{f}(x_0, \underset{\sim}{X})) + o(h)$$ ∎

The mean integrated squared error (*MISE*) of $\hat{f}(x, \underset{\sim}{X})$, given by integrating the $MSE\ (\hat{f}(x, \underset{\sim}{X}))$, is an error criterion over the real line,

$$MSE(\hat{f}) = \int MSE(\hat{f}(x, \underset{\sim}{X}))dx$$

$$= \int (V(\hat{f}(x_0, \underset{\sim}{X})) + B^2(\hat{f}(x_0, \underset{\sim}{X})))\,dx$$

$$= \frac{R(K)}{nh} - \frac{1}{n}\int f^2(x)dx + \frac{h^4\mu_2^2R(f''(x))}{4} + o((nh)^{-1}) + o(h^4) \tag{13}$$

Scott and Terrell [6] estimated $R(f'')$ under the same bandwidth h as in the estimation of $f(x)$. Using

$$\hat{R}(f^{(d)}) = \hat{R}(\hat{f}^{(d)}) - \frac{\hat{R}(K^{(d)})}{nh^{2d+1}}$$

$$= \int (\hat{f}^{(d)})^2dx - \frac{1}{nh^{2d+1}}\int (K^{(d)}(x))^2dx, \tag{14}$$

$E[\hat{R}(f^{(d)})] = R(f^{(d)}) + o(h)$ ([6]: Lemma 3.2, and [12]). Hence,

$$MISE(\hat{f}) = \frac{R(K)}{nh} - \frac{1}{n}[R(\hat{f}) - \frac{R(K)}{nh}] + \frac{h^4\mu_2^2}{4}(R(\hat{f}'') - \frac{R(K^{(2)})}{nh^5}). \tag{15}$$

Theorem 2.4 *$\hat{MISE}(\hat{f})$ is asymptotically unbiased estimator of $MISE(\hat{f})$, when $h=h(n)\longrightarrow 0$ as $n\longrightarrow\infty$.*

Proof

Consider

$$E[\hat{MISE}(\hat{f})]=\frac{R(K)}{nh}-\frac{1}{n}E[R(\hat{f})-\frac{R(K)}{nh}] +\frac{h^4\mu_2^2}{4}E[(R(\hat{f}'')-\frac{R(K^{(2)})}{nh^5})]$$

$$=\frac{R(K)}{nh}-\frac{1}{n}[R(f)+o(h)]+\frac{h^4\mu_2^2}{4}[(R(f'')+o(h)]$$

$$=\frac{R(K)}{nh}-\frac{R(f)}{n}+\frac{h^4\mu_2^2}{4}(R(f'')+o(h)$$

Thus,

$$B(\hat{MISE}(\hat{f})) = E[\hat{MISE}(\hat{f})] - MISE(\hat{f}) = o(h)$$ ■

The notation *AMISE* of $\hat{f}(x)$ is used to represent asymptotic *MISE* of $\hat{f}(x)$ (ignoring higher order in the expansion of *MISE* of $\hat{f}(x)$), i.e.

$$AMISE(\hat{f})=\frac{R(K)}{nh}-\frac{R(f)}{n}+\frac{h^4\mu_2^2R(f''(x))}{4} \quad (16)$$

An estimate of $AMISE(\hat{f})$ maybe obtained by replacing $\hat{R}(f)$, $\hat{R}(f'')$ as in (15) which is the same as the estimator of $MISE(\hat{f})$.

3. TWO NEW KERNEL ESTIMATORS

In this study we give two kernel functions $K_1(u)$ and $K_2(u)$ that yield "good" estimates of a density function in the sense that the bias and the variance of each density estimator is small. A symmetric kernel function with small variance, μ_2, is presented to decrease the

bias $B(\hat{f}(x_0,\underset{\sim}{X}))$. To decrease the variance of $(\hat{f}(x_0,\underset{\sim}{X}))$, we need to minimize *the roughness of $K(u)$*. To decrease the sum of $|B(f(x_0,\underset{\sim}{X}))|$ and the variance of $\hat{f}(x_0,\underset{\sim}{X}))$ the kernel function will be found to minimize $A_1(K)=R(K)+\mu_2$, *the sum of the squared L_2 norm of $K(u)$ and the kernel variance*. $MSE(\hat{f}(x,\underset{\sim}{X}))$, $MISE(\hat{f})$, and $AMISE(\hat{f})$ depend on the sum of the squared bias and the variance. So a kernel function that minimizes the sum of $R(K)$ and μ_2^2 is chosen, i.e. we choose the kernel function that minimizes *the sum of the squared L_2 norm of $K(u)$ and the squared kernel variance*, $A_2(K)=R(K)+\mu_2^2$.

A kernel function with compact support expressed in the form of polynomials can be found in Gasser, Muller & Mammitzsch [10], Delaigle & Hall [13], Granovsky & Muller [14], Horova [15], Mammitzsch [16], and Muller & Wang [17]. Horova [15] presented the construction of kernel function that minimized the squared L_2 norm of $K(u)$ under the condition that the moments of $K(u)$ were polynomials of certain degrees.

Hence, in this paper we need to find the coefficients of the new kernel functions in the form of second degree polynomial with support [-1, 1];

$$K(u)=\sum_{i=0}^{2}c_iu^i, \quad (17)$$

where c_i are the coefficients to be determined to minimize $A_1(K)$ or $A_2(K)$ subject to the constraint that $K(u)$ is a symmetric density function. To derive the kernel functions $K_1(u)$, $K_2(u)$ we use Lagrange multipliers. To obtain the kernel functions $K_1(u)$ under the constraints that minimize $A_1(K)$ let

$$L_1(K)=A_1(K)+\lambda_1(1-\mu_0)+\lambda_2(\mu_1), \quad (18)$$

where λ_1, λ_2 are Lagrange multipliers. The result is

$$K_1(u)=(\frac{2}{3}-\frac{u^2}{2})I_{[-1,1]}(u). \quad (19)$$

To minimize $A_2(K)$ under the constraints, construct the corresponding Lagrange function

$$L_2(K)=A_2(K)+\lambda_1(1-\mu_0)+\lambda_2(\mu_1), \quad (20)$$

where λ_1, λ_2 are Lagrange multipliers. The result is

$$K_2(u)=(\frac{63}{106}-\frac{15u^2}{53})I_{[-1,1]}(u). \quad (21)$$

$K_1(u)$ and $K_2(u)$ may be used to obtain the density $f(x)$ as in equation (1) once the observations $\underset{\sim}{x}=(x_1, \ldots, x_n)$ are observed.

4. PROPERTIES OF THE KERNEL ESTIMATORS

The notation $\hat{f}_1(x_0, \underset{\sim}{X})$ is used to denote the kernel estimator of $f(x)$ at a point $x = x_0$, and $\hat{f}_2(x_0, \underset{\sim}{X})$ to denote the kernel estimator of $f(x)$ at a point $x = x_0$.

Let the density $f(x)$ be such that its first two derivatives exist and are uniformly continuous. Assume $h = h(n) \longrightarrow 0$ and $nh \longrightarrow \infty$ as $n \longrightarrow \infty$. The biases of $\hat{f}_1(x_0, \underset{\sim}{X})$ and $\hat{f}_2(x_0, \underset{\sim}{X})$ are, respectively

$$B\hat{f}_1(x_0, \underset{\sim}{X})) = \frac{11h^2}{90} f''(x_0) + o(h^2), \quad (22)$$

$$B\hat{f}_2(x_0, \underset{\sim}{X})) = \frac{15h^2}{106} f''(x_0) + o(h^2). \quad (23)$$

The variances of $\hat{f}_1(x_0, \underset{\sim}{X})$, $\hat{f}_2(x_0, \underset{\sim}{X})$ are

$$V(\hat{f}_1(x_0, \underset{\sim}{X})) = \frac{49 f(x_0)}{90nh} + o((nh)^{-1}), \quad (24)$$

$$V(\hat{f}_2(x_0, \underset{\sim}{X})) = \frac{2{,}889 f(x_0)}{5{,}618nh} + o((nh)^{-1}). \quad (25)$$

And the mean square errors of $\hat{f}_1(x_0, \underset{\sim}{X})$, $\hat{f}_2(x_0, \underset{\sim}{X})$ are

$$MSE(\hat{f}_1(x_0, \underset{\sim}{X})) = \frac{49 f(x_0)}{90nh} + \frac{121h^4}{8{,}100}(f''(x_0))^2 + o((nh)^{-1}) + o(h)^4), \quad (26)$$

$$MSE(\hat{f}_2(x_0, \underset{\sim}{X})) = \frac{2{,}889 f(x_0)}{5{,}618nh} + \frac{225h^4}{11{,}236}(f''(x_0))^2 + o((nh)^{-1}) + o(h)^4). \quad (27)$$

The $MISE$ of $\hat{f}_1(x, \underset{\sim}{X})$ and $\hat{f}_2(x, \underset{\sim}{X})$ are, respectively

$$MISE(\hat{f}_1) = \frac{49}{90nh} + \frac{121h^4}{8{,}100}\int (f''(x))^2\,dx + o((nh)^{-1}) + o(h)^4), \quad (28)$$

$$MISE(\hat{f}_2) = \frac{2{,}889}{5{,}618nh} + \frac{225h^4}{11{,}236}\int (f''(x))^2\,dx + o((nh)^{-1}) + o(h)^4). \quad (29)$$

The $AMISE$ of $\hat{f}_1(x, \underset{\sim}{X})$, $\hat{f}_2(x, \underset{\sim}{X})$ are

$$AMISE(\hat{f}_1) = \frac{49}{90nh} + \frac{121h^4}{8{,}100}\int (f''(x))^2\,dx \quad (30)$$

$$AMISE(\hat{f}_2) = \frac{2{,}889}{5{,}618nh} + \frac{225h^4}{11{,}236}\int (f''(x))^2\,dx \quad (31)$$

5. SIMULATION STUDY

In order to study the $AMISE$ of the kernel estimates of a density function using $K_1(u)$ and $K_2(u)$ a simulation study is carried out under various kernel functions and bandwidths of the estimates. The effects of the kernels and bandwidths of the estimates of $f(x)$ at different sample sizes are considered. The simulations are performed using programs written in R. A population of size 50,000 is generated for each of the distribu-tions which are built from normal mixtures [4]. In this study, fifteen different normal mixture distributions are simulated. They are as appeared in Table 1.

Random samples of sizes 50, 100, 200 and 500 are drawn from each population repeatedly 1,000 times. The bandwidths used in the simulation studies are the Silverman rule of thump (SRT) bandwidth (the commonly used quick and simple ideas for selecting the bandwidth [18]), two-stage direct plug-in bandwidth (DPI) [8] and the solve-the-equation method bandwidth (STE) [8]. The $AMISE$ of the kernel estimates are computed. From 1,000 samples with specified size, the mean of $AMISE(\hat{f})$, $\overline{AMISE}(\hat{f})$, of kernel estimates are computed and compared.

6. RESULTS OF THE SIMULATION

The results of the $\overline{AMISE}(\hat{f})$ and the estimated standard deviation (SD) by distributions, bandwidths, and various sample sizes with 1,000 replications are shown in Tables 2-16. The bold number is the smallest $\overline{AMISE}(\hat{f})$ for each bandwidths and various

Table 1. Normal Mixture Distributions of 15 Populations.

Distribution	$w_1N(\mu_1,\sigma_1^2)+\cdots+w_kN(\mu_k,\sigma_k^2)$
1. Gaussian	$N(0,1)$
2. skewed unimodal	$\frac{1}{5}N(0,1)+\frac{1}{5}N(\frac{1}{2},(\frac{2}{3})^2)+\frac{3}{5}N(\frac{13}{12},(\frac{5}{9})^2)$
3. outlier	$\frac{1}{10}N(0,1)+\frac{9}{10}N(0,(\frac{1}{10})^2)$
4. bimodal	$\frac{1}{2}N(-1,(\frac{2}{3})^2)+\frac{1}{2}N(1,(\frac{2}{3})^2)$
5. strongly skewed	$\sum_{i=o}^{7}\frac{1}{8}N(3[(\frac{2}{3})^i-1],(\frac{2}{3})^{2i})$
6. kurtotic unimodal	$\frac{2}{3}N(0,1)+\frac{1}{3}N(0,(\frac{1}{10})^2)$
7. separated bimodal	$\frac{1}{2}N(-\frac{3}{2},(\frac{1}{2})^2)+\frac{1}{2}N(\frac{3}{2},(\frac{1}{2})^2)$
8. skewed bimodal	$\frac{3}{4}N(0,1)+\frac{1}{4}N(\frac{3}{2},(\frac{1}{3})^2)$
9. trimodal	$\frac{9}{20}N(-\frac{6}{5},(\frac{3}{5})^2)+\frac{9}{20}N(\frac{6}{5},(\frac{3}{5})^2)+\frac{1}{10}N(0,(\frac{1}{4})^2)$
10. claw	$\frac{1}{2}N(0,1)+\sum_{i=0}^{4}\frac{1}{10}N(\frac{i}{2}-1,(\frac{1}{10})^2)$
11. double claw	$\frac{49}{100}N(-1,(\frac{2}{3})^2)+\frac{49}{100}N(1,(\frac{2}{3})^2)+\sum_{i=0}^{6}\frac{1}{350}N((i-3)/2,(\frac{1}{100})^2)$
12. asymmetric claw	$\frac{1}{2}N(0,1)+\sum_{i=-2}^{2}\frac{2^{1-i}}{31}N((i+1/2),(\frac{2^{-i}}{10})^2)$
13. asymmetric double claw	$\sum_{i=0}^{1}\frac{46}{100}N(2i-1,(\frac{2}{3})^2)+\sum_{i=1}^{3}\frac{1}{300}N(-\frac{i}{2},(\frac{1}{100})^2)+\sum_{i=1}^{3}\frac{7}{300}N(\frac{i}{2},(\frac{7}{100})^2)$
14. smooth comp	$\sum_{i=0}^{5}\frac{2^{5-i}}{63}N(\frac{65-96*2^{-i}}{21},(\frac{32}{63*2^i})^2)$
15. discrete comb	$\sum_{i=0}^{2}\frac{2}{7}N(\frac{12i-15}{7},(\frac{2}{7})^2)+\sum_{i=8}^{10}\frac{1}{21}N(\frac{2i}{7},(\frac{1}{21})^2)$

sample sizes. For data from Gaussian, skewed unimodal, kurtotic unimodal, separated bimodal, trimodal, double claw, asymmetric double claw and smooth comp distributions, the results of the simulation can be found in [19].

For data from outlier and bimodal distributions, the kernel estimates using the proposed K_1 and K_2 perform better than the uniform and Gaussian estimates as shown in Tables 2-3. For samples of any sizes, the $\overline{AMISE}$ of $\hat{f}(x,\underset{\sim}{x})$ with DPI bandwidth is lower than the $\overline{AMISE}$ of $\hat{f}(x,\underset{\sim}{x})$ with SRT and STE bandwidth for data from outlier distribution. For data from bimodal distribution the $\overline{AMISE}$ of $\hat{f}(x,\underset{\sim}{x})$ with SRT bandwidth is lower than the $\overline{AMISE}$ of $\hat{f}(x,\underset{\sim}{x})$ with DPI and STE bandwidth. The $\overline{AMISE}$ of $\hat{f}(x,\underset{\sim}{x})$ with three bandwidths are not different when the sample size increases. The $\overline{AMISE}$ of $\hat{f}(x,\underset{\sim}{x})$ are closed to zero as the sample size gets larger.

Table 2. $\overline{AMISE}\ (\hat{f})$ of Kernel Estimates for Outlier Distribution.

n	Kernel functions	h_{SRT}		h_{DPI}		h_{STE}	
		$\overline{AMISE}\ (\hat{f})$	SD	$\overline{AMISE}\ (\hat{f})$	SD	$\overline{AMISE}\ (\hat{f})$	SD
50	Uniform	0.106299	0.013783	0.105019	0.013136	0.109251	0.019853
	Epanechnikov	**0.097532**	0.01296	**0.096382**	0.012401	**0.100356**	0.018696
	Gaussian	0.103442	0.013489	0.102245	0.012907	0.106382	0.019459
	K_1	0.09795	0.013033	0.09674	0.012421	0.100741	0.018773
	K_2	0.099841	0.013203	0.098615	0.012583	0.102669	0.019017
100	Uniform	0.062305	0.005	0.061435	0.00464	0.062702	0.00691
	Epanechnikov	**0.057398**	0.004693	**0.056607**	0.004371	**0.057797**	0.0065
	Gaussian	0.060706	0.004885	0.059882	0.00455	0.06112	0.006765
	k1	0.057633	0.004728	0.056809	0.004387	0.058008	0.006534
	k2	0.058691	0.004789	0.057857	0.004444	0.059071	0.006619
200	Uniform	0.036706	0.001736	0.036136	0.001532	0.036509	0.002335
	Epanechnikov	**0.033936**	0.001627	**0.033415**	0.001442	**0.033763**	0.002194
	Gaussian	0.035803	0.001693	0.035261	0.001501	0.035624	0.002283
	K_1	0.034068	0.001641	0.033529	0.001449	0.033881	0.002208
	K_2	0.034665	0.001662	0.034119	0.001468	0.034476	0.002236
500	Uniform	0.018336	0.000461	0.018042	0.000373	0.018118	0.000588
	Epanechnikov	**0.017019**	0.000431	**0.01675**	0.00035	**0.01682**	0.000552
	Gaussian	0.017907	0.000449	0.017626	0.000364	0.0177	0.000575
	K_1	0.017082	0.000436	0.016804	0.000352	0.016876	0.000556
	K_2	0.017366	0.000441	0.017085	0.000357	0.017157	0.000564

Table 3. $\overline{AMISE}\ (\hat{f})$ of Kernel Estimates for Bimodal Distribution.

n	Kernel functions	h_{SRT}		h_{DPI}		h_{STE}	
		$\overline{AMISE}\ (\hat{f})$	SD	$\overline{AMISE}\ (\hat{f})$	SD	$\overline{AMISE}\ (\hat{f})$	SD
50	Uniform	0.010353	0.000308	0.011081	0.001253	0.011804	0.00174
	Epanechnikov	**0.0095**	0.000292	**0.01019**	0.00119	**0.010868**	0.001643
	Gaussian	0.010079	0.000304	0.010797	0.001239	0.011503	0.00171
	K_1	0.009535	0.000292	0.010223	0.001185	0.010908	0.001646
	K_2	0.00972	0.000296	0.010418	0.0012	0.011111	0.001667
100	Uniform	0.006234	7.24E-05	0.00654	0.000473	0.006881	0.000746
	Epanechnikov	**0.005748**	6.94E-05	**0.006037**	0.000448	**0.006356**	0.000703
	Gaussian	0.006078	7.22E-05	0.006379	0.000466	0.006711	0.000731
	k1	0.005767	6.85E-05	0.006057	0.000447	0.006379	0.000705
	k2	0.005873	6.94E-05	0.006167	0.000453	0.006493	0.000715
200	Uniform	0.003742	2.33E-05	0.003861	0.000162	0.004013	0.0003
	Epanechnikov	**0.003463**	2.25E-05	**0.003575**	0.000153	**0.003717**	0.000282
	Gaussian	0.003652	2.34E-05	0.003769	0.000159	0.003916	0.000294
	K_1	0.003474	2.2E-05	0.003588	0.000153	0.003731	0.000284
	K_2	0.003535	2.23E-05	0.00365	0.000155	0.003795	0.000288
500	Uniform	0.001888	6.65E-06	0.001954	6.94E-05	0.002018	0.000117
	Epanechnikov	**0.001755**	6.46E-06	**0.001816**	6.51E-05	**0.001876**	0.00011
	Gaussian	0.001845	6.72E-06	0.001909	6.77E-05	0.001972	0.000115
	K_1	0.00176	6.29E-06	0.001822	6.57E-05	0.001883	0.000111
	K_2	0.001789	6.37E-06	0.001852	6.65E-05	0.001914	0.000112

For data from strongly skewed distribution, from Table 4, the $\hat{f}(x,\underset{\sim}{x})$ using K_1 with STE bandwidth gives lower $\overline{AMISE}\,(\hat{f})$ than the other kernel estimates.

Table 4. $\overline{AMISE}\,(\hat{f})$ of Kernel Estimates for Strongly Skewed Distribution.

n	Kernel functions	h_{SRT}		h_{DPI}		h_{STE}	
		$\overline{AMISE}\,(\hat{f})$	SD	$\overline{AMISE}\,(\hat{f})$	SD	$\overline{AMISE}\,(\hat{f})$	SD
50	Uniform	25.55101	21.60003	6.280713	5.192588	2.257085	2.6098
	Epanechnikov	24.19577	20.45494	5.947045	4.917328	2.136688	2.47147
	Gaussian	25.18433	21.29027	6.190364	5.118141	2.2244	2.572399
	K_1	**24.16042**	20.42498	**5.938436**	4.910111	**2.133693**	2.467827
	K_2	24.47547	20.6912	6.01598	4.974107	2.161648	2.499992
100	Uniform	14.46872	8.929398	2.160192	1.290507	0.662885	0.477496
	Epanechnikov	13.70134	8.456021	2.045306	1.222102	0.627352	0.452196
	Gaussian	14.2611	8.801347	2.12906	1.27201	0.6532	0.470662
	K_1	**13.68132**	8.443638	**2.042374**	1.220303	**0.626521**	0.45152
	K_2	13.85971	8.55369	2.069066	1.236208	0.63476	0.457405
200	Uniform	8.28696	3.833686	0.767395	0.3387	0.219765	0.106315
	Epanechnikov	7.847463	3.63045	0.726521	0.320748	0.207905	0.100686
	Gaussian	8.16805	3.778709	0.756304	0.333847	0.216509	0.104798
	K_1	**7.835997**	3.625133	**0.725498**	0.320274	**0.207658**	0.100532
	K_2	7.938165	3.672382	0.73499	0.324449	0.210401	0.101842
500	Uniform	3.861502	1.134843	0.198502	0.054305	0.05697	0.016142
	Epanechnikov	3.656717	1.074681	0.187895	0.051427	0.053857	0.015289
	Gaussian	3.806095	1.118569	0.195614	0.053527	0.056101	0.015913
	K_1	**3.651375**	1.073107	**0.187643**	0.05135	**0.05381**	0.015264
	K_2	3.69898	1.087094	0.190103	0.05202	0.054526	0.015463

For data with skewed bimodal distribution, rom Table 5, the $\hat{f}(x,\underset{\sim}{x})$ using K_1 and the SRT bandwidth give lower $\overline{AMISE}(\hat{f})$ than the other kernel estimates when the sample size is 50. The $\overline{AMISE}(\hat{f})$ of $\hat{f}(x,\underset{\sim}{x})$ with STE bandwidth is lower than the $\overline{AMISE}$ of the other kernel estimates when the sample sizes are between 100 and 500.

For data with strongly skewed and skewed bimodal distributions, the $\hat{f}(x,\underset{\sim}{x})$ using K_2 perform better than the $\hat{f}(x,\underset{\sim}{x})$ with uniform and Gaussian kernel functions for all sample sizes.

Table 5. $\overline{AMISE}\ (\hat{f})$ of Kernel Estimates for Skewed Bimodal Distribution.

n	Kernel functions	h_{SRT}		h_{DPI}		h_{STE}	
		$\overline{AMISE}\ (\hat{f})$	SD	$\overline{AMISE}\ (\hat{f})$	SD	$\overline{AMISE}\ (\hat{f})$	SD
50	Uniform	0.041101	0.011072	0.049029	0.025845	0.042082	0.028525
	Epanechnikov	0.038582	0.01049	0.04609	0.024484	**0.039503**	0.027026
	Gaussian	0.040372	0.010919	0.048187	0.025483	0.041331	0.02813
	K_1	**0.038579**	0.01047	**0.046075**	0.024439	0.039506	0.026973
	K_2	0.03915	0.010606	0.046744	0.024758	0.04009	0.027325
100	Uniform	0.02369	0.00404	0.024757	0.010134	0.021046	0.010805
	Epanechnikov	0.022259	0.003828	0.023269	0.009601	**0.01975**	0.010239
	Gaussian	0.023276	0.003984	0.024327	0.009993	0.020664	0.010657
	k1	**0.022258**	0.00382	**0.023267**	0.009582	0.019758	0.010218
	k2	0.022582	0.00387	0.023604	0.009707	0.02005	0.010351
200	Uniform	0.01383	0.001645	0.011947	0.003817	0.010158	0.003838
	Epanechnikov	0.013007	0.001559	**0.011222**	0.003617	**0.009525**	0.003638
	Gaussian	0.013592	0.001622	0.011734	0.003765	0.009968	0.003786
	K_1	**0.013006**	0.001556	0.011226	0.00361	0.009534	0.003629
	K_2	0.013193	0.001576	0.011389	0.003657	0.009675	0.003677
500	Uniform	0.006719	0.000524	0.004672	0.000783	0.004103	0.000655
	Epanechnikov	0.006326	0.000496	**0.004385**	0.000742	**0.003845**	0.000621
	Gaussian	0.006605	0.000517	0.004585	0.000773	0.004024	0.000647
	K_1	**0.006325**	0.000495	0.004389	0.00074	0.003851	0.000619
	K_2	0.006414	0.000502	0.004453	0.00075	0.003908	0.000627

For data with claw distribution, from Table 6, the $\hat{f}(x,x)$ using K_1 and the SRT bandwidth gives lower $\overline{AMISE}(\hat{f})$ than the other kernel estimates with sample size between 50 and 200. The same results are true when the STE bandwidth is used with sample size 500.

Table 6. $\overline{AMISE}\ (\hat{f})$ of Kernel Estimates for Claw Distribution.

n	Kernel functions	h_{SRT}		h_{DPI}		h_{STE}	
		$\overline{AMISE}\ (\hat{f})$	SD	$\overline{AMISE}\ (\hat{f})$	SD	$\overline{AMISE}\ (\hat{f})$	SD
50	Uniform	4.52962	2.079192	7.290383	4.736172	6.370666	5.350854
	Epanechnikov	4.289015	1.968973	6.903425	4.485099	6.032459	5.067199
	Gaussian	4.464471	2.049382	7.185649	4.66826	6.279113	5.274133
	K_1	**4.282806**	1.966083	**6.893383**	4.478524	**6.023698**	5.059767
	K_2	4.338723	1.991709	6.983325	4.536896	6.102306	5.125714
100	Uniform	2.589561	0.794629	3.7981	1.794412	3.279423	2.097167
	Epanechnikov	2.452037	0.752506	3.596509	1.699288	3.105326	1.985996
	Gaussian	2.552323	0.783236	3.743534	1.768683	3.232292	2.067099
	k1	**2.448487**	0.751401	**3.591282**	1.696795	**3.100821**	1.983081
	k2	2.480448	0.761194	3.638138	1.718911	3.141284	2.008927
200	Uniform	1.527882	0.322569	1.9605	0.749484	1.615818	0.906283
	Epanechnikov	1.446759	0.30547	1.856444	0.709753	1.530032	0.858242
	Gaussian	1.505917	0.317945	1.932332	0.738738	1.592591	0.89329
	K_1	**1.444664**	0.305022	**1.853748**	0.708712	**1.527817**	0.856981
	K_2	1.463517	0.308997	1.877933	0.717949	1.547754	0.868151
500	Uniform	0.739524	0.095436	0.70984	0.230515	0.382186	0.347998
	Epanechnikov	0.700267	0.090377	0.672157	0.218295	0.361867	0.329555
	Gaussian	0.728895	0.094068	0.699637	0.22721	0.376676	0.343013
	K_1	**0.699253**	0.090244	**0.671185**	0.217975	**0.361355**	0.329067
	K_2	0.708377	0.09142	0.679942	0.220816	0.366074	0.333356

For data with discrete comb distribution, the $\hat{f}(x,\underset{\sim}{x})$ with the STE bandwidth gives lower $\overline{AMISE}(\hat{f})$ than the others $\hat{f}(x,\underset{\sim}{x})$ as in Table 7. The $\hat{f}(x,\underset{\sim}{x})$ using the proposed K_1 performs well. And the $\hat{f}(x,\underset{\sim}{x})$ which used the proposed K_2 perform better than the $\hat{f}(x,\underset{\sim}{x})$ with uniform, Gaussian functions as shown in Table 7.

Table 7. $\overline{AMISE}\ (\hat{f})$ of Kernel Estimates for Discrete Comb Distribution.

n	Kernel functions	h_{SRT}		h_{DPI}		h_{STE}	
		$\overline{AMISE}\ (\hat{f})$	SD	$\overline{AMISE}\ (\hat{f})$	SD	$\overline{AMISE}\ (\hat{f})$	SD
50	Uniform	355.5718	99.41007	247.6135	124.5601	17.11491	49.15704
	Epanechnikov	336.7212	94.13994	234.4861	117.9567	16.20719	46.55103
	Gaussian	350.4724	97.98441	244.0622	122.7737	16.86928	48.45208
	K_1	**336.2284**	94.00215	**234.143**	117.784	**16.18355**	46.48289
	K_2	340.6108	95.22734	237.1948	119.3192	16.39456	47.08873
100	Uniform	206.8157	38.94436	67.37731	25.97716	3.738073	1.200393
	Epanechnikov	195.8514	36.87977	63.80519	24.60001	3.539691	1.13676
	Gaussian	203.8496	38.38586	66.41097	25.60462	3.68436	1.183182
	k1	**195.5647**	36.82579	**63.71183**	24.564	**3.534568**	1.135091
	k2	198.1137	37.30576	64.54226	24.88416	3.580673	1.149886
200	Uniform	119.6137	14.95454	15.30392	3.8735	1.537078	0.303992
	Epanechnikov	113.2724	14.16174	14.4925	3.668152	1.455478	0.287878
	Gaussian	117.8983	14.74008	15.0844	3.817951	1.514974	0.299634
	K_1	**113.1066**	14.14101	**14.47131**	3.662782	**1.453384**	0.287455
	K_2	114.5808	14.32532	14.65994	3.710521	1.472345	0.291201
500	Uniform	57.44589	4.31653	2.376638	0.333644	0.546175	0.067213
	Epanechnikov	54.40041	4.087693	2.250599	0.315957	0.51717	0.063651
	Gaussian	56.62203	4.254626	2.342532	0.32886	0.538314	0.06625
	K_1	**54.3208**	4.08171	**2.247318**	0.315494	**0.516432**	0.063557
	K_2	55.0288	4.13491	2.276616	0.319606	0.523171	0.064385

7. Conclusions and Discussion

Estimators of the bias, the variance, the mean squared error, and the mean integrated squared error of kernel density estimator $\hat{f}(x,\underset{\sim}{X})$ are obtained using the density derivative estimator of Jones [7] and the estimator of the squared L_2 norm of $f''(x)$ of Scott and Terrell [6]. The estimators of $f''(x)$ and the squared L_2 norm of $f''(x)$ use the same bandwidth as in the estimation of $f(x)$. The $\hat{B}(\hat{f}(x_0,\underset{\sim}{X}))$, $\hat{V}(\hat{f}(x,\underset{\sim}{X}))$, $M\hat{S}E(\hat{f}(x,\underset{\sim}{X}))$ and $MI\hat{S}E(\hat{f})$ are asymptotically unbiased.

Results of the simulation study indicate that for data with outlier and bimodal distributions, the $\hat{f}(x,\underset{\sim}{x})$ using K_1 and K_2 performs better than the uniform and Gaussian estimates. For all sample size the $\overline{AMISE}$ of $\hat{f}(x,\underset{\sim}{x})$ with DPI bandwidth is lower than the others for data with symmetry unimodal distribution.

For data with bimodal distribution the $\overline{AMISE}$ of $\hat{f}(x,\underset{\sim}{x})$ with SRT bandwidth is lower than the others. While the data with skewed bimodal, the density estimate using K_1 and SRT bandwidth is better than the other density estimates at sample size 50.

For data with claw distribution, the $\hat{f}(x,\underset{\sim}{x})$ which used K_1 with SRT bandwidth is better than the others $\hat{f}(x,\underset{\sim}{x})$. The same results are true when the STE bandwidth is used with large sample sizes.

For data distributed as discrete comb distribution, the $\overline{AMISE}$ of $\hat{f}(x,\underset{\sim}{x})$ which used

K_1 with STE bandwidth is lower than the $\overline{AMISE}$ of the other kernel estimates.

For these data, the $\hat{f}(x,\underset{\sim}{x})$ using K_2 performs better than the $\hat{f}(x,\underset{\sim}{x})$ with uniform, Gaussian function.

For large sample size, $\overline{AMISE}$ of the estimates $\hat{f}(x,\underset{\sim}{x})$ which use Epanechnikov, K_1 and K_2 functions become closer. The $\overline{AMISE}(\hat{f})$ becomes smaller as the sample sizes increase, meaning that the kernel estimate is becoming more accurate.

The $\overline{AMISE}(\hat{f})$ when the sample data are form highly skewed, kurtosis and multimodal populations give a larger $\overline{AMISE}(\hat{f})$ because the bandwidth is far from the optimal bandwidth which are consistent with the degree of estimation difficulty which increases with skewness, kurtosis and multimodality of the distributions [8].

REFERENCES

[1] Sheather S.J., Density estimation. *Statistical Science*, 2004; **19**: 588-597.

[2] Rosenblatt M., Remark on some nonparametric estimates of a density function. *Annals of Mathematical Statistics*, 1956; **27**: 832-837.

[3] Parzen E., On estimation of a probability density function and mode, *Annals of Mathematical Statistics*, 1962; **33**: 1065-1076.

[4] Marron J.S. and Wand M.P., Exact mean integrated squared error, *The Annals of Statistics*, 1992; **20**: 712-736.

[5] Hall P., Effect of bias estimation on coverage accuracy of bootstrap confidence intervals for a probability density, *The Annals of Statistics*, 1992; **20**: 675-694.

[6] Scott D.W. and Terrell G.R., Biased and unbiased cross-varlidation in density estimation, *Journal of the American Statistical Association*, 1987; **82**: 1131-1145.

[7] Jones M.C., On kernel density derivative estimation, *Communications in Statistics : Theory and Methods*, 1994; **23**: 2133-2139.

[8] Wand M.P. and Jones M.C., *Kernel smoothing*, London: Chapman and Hall, 1995.

[9] Hardle W., *Smoothing techniques, with implementations in S*., Springer, New York, 1991.

[10] Gasser T., Müller H.G. and Mammitzsch V., Kernels for nonparametric curve estimation, *Journal of the Royal statistical Society. Series B*, 1985; **47**: 238-252.

[11] Müller H.G., Smooth optimum kernel estimators of densities, regression curves and modes, *The Annals of Statistics*, 1984; **12**: 766-774.

[12] Scott D.W., *Multivariate density estimation*, New York: Wiley, 1992.

[13] Delaigle A. and Hall P., On optimal kernel choice for deconvolution, *Statistics and Probability Letters*, 2006; **76**: 1594-1602.

[14] Granovsky B.L. and Müller, H.G., Optimizing kernel methods: a unifying variational principle, *International Statistical Review*, 1991; **59**: 373-388.

[15] Horova I., Some remarks on kernels, *Journal of Computational Analysis and Applications*, 2000; **2**: 253-263.

[16] Mammitzsch V., Optimal kernels, *Statistics & Decisions*, 2007; **25**: 153-172.

[17] Müller H.G. and Wang J.L., Hazard rate estimation under random censoring with varying kernels and bandwidths, *Biometrics*, 1994; **50**: 61-76.

[18] Liu K., *Kernel estimates of statistical distribution functions*, Doctoral Dissertation, University of South Florida, 1998.

[19] Rodchuen M. and Suwattee P., Comparison of kernel density estimators, *Thailand Statistician*, 2010; **8(2)**: 167-181.