Thailand Statistician
July 2010; 8(2) : 167-181
http://statassoc.or.th
Contributed paper

Comparison of Kernel Density Estimators

Manachai Rodchuen *[a], 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 March2010
Accepted: 23 June 2010

Abstract

Two kernel estimators of a density function $f(x)$ are considered. The measures of errors of the estimates depend on the corresponding kernel functions used to derive them together with the bandwidths of the kernels. Simulation study is carried out to compare the *AMISE* of the estimates with those of uniform, Epanechnikov and Gaussian kernel functions. The bandwidths used for comparison of the errors of the estimates are the Silverman rule of thump (SRT), two-stage direct plug-in (DPI) and the solve-the-equation (STE) method bandwidths. For data from Gaussian, skewed unimodal, and separated bimodal distributions, the proposed kernel estimates perform better than the uniform and Gaussian estimates. One of the proposed kernel estimates with STE bandwidth performs well when the sample data are from a kurtotic unimodal and trimodal distributions and with samples of sizes 50 and 100. This kernel estimate also performs better than the others for data from multimodal distribution. Another proposed kernel estimate also performs better than the uniform and Gaussian estimates.

Keywords: Density estimation, error criteria, kernel estimator, mean squared error, mean integrated squared error.

1. Introduction

Density estimation is an interesting problem in statistical inference for a long time. One well-known method of density estimation is the use of kernel functions introduced by Rosenblatt [1], and Parzen [2]. Let $\underset{\sim}{X} = (X_1, \ldots , X_n)$ be a random sample of size n from a population with an unknown probability density function $f(x)$, and $\underset{\sim}{x} = (x_1, \ldots , x_n)$ the sample observations on $\underset{\sim}{X}$. The kernel density estimate of $f(x)$ at the point x_0 is given by

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

where $K(u)$ is a real valued kernel 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 [1, 2]. The properties of a kernel density estimate depend on the properties of the kernel function $K(u)$ and the bandwidth *h* used. If $K(u)$ is a probability density function, then the 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

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

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

i) $K(u)$ is a 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) $\mu_2 = \int u^2 K(u)du < \infty$ 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, $K_G(u) = e^{-u^2/2}/\sqrt{2\pi}$. Also, there are many criteria of measuring the errors of the kernel estimates. These include the mean squared error ($MSE(\hat{f}(x,\underset{\sim}{X}))$) and the mean integrated squared error ($MISE(\hat{f})$).The mean squared error, MSE of $\hat{f}(x_0,\underset{\sim}{X})$, is the error at a point x_0 of the density.

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

or

$$MSE(\hat{f}(x_0,\underset{\sim}{X})) = V(\hat{f}(x_0,\underset{\sim}{X})) + B^2(\hat{f}(x_0,\underset{\sim}{X})), \qquad (3)$$

where $V(\hat{f}(x_0,\underset{\sim}{X}))$ is the variance of $\hat{f}(x_0,\underset{\sim}{X})$ and $B(\hat{f}(x_0,\underset{\sim}{X}))$ is its bias. 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 $h = h(n) \to 0$ as $n \to \infty$. In such a case, the bias of $\hat{f}(x_0,\underset{\sim}{X})$ is

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

where μ_2 is *the kernel variance*, (Hardle [3]). Also, the variance of $\hat{f}(x_0,\underset{\sim}{X})$ is

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

$$= \frac{1}{nh^2} V[K(\frac{x_0 - X}{h})]$$

$$= \frac{1}{n}\{\int \frac{1}{h^2} K^2(\frac{x_0 - x}{h}) f(x)dx - [\int \frac{1}{h} K(\frac{x_0 - x}{h}) f(x)dx]^2\}. \qquad (5)$$

By Taylor series expansion of $f(x)$ about x_0 and let $x = x_0 - hu$, we have

$$V(\hat{f}(x_0,\underset{\sim}{X})) = \frac{f(x_0)}{nh} R(K) - \frac{(f(x_0))^2}{n} + o(n^{-1}), \qquad (6)$$

where $R(K)=\int K^2(u)du$ is *the squared* L_2 *norm of* $K(u)$, L_2 is the set of all Lebesque measurable real valued functions, $g(t)$, such that $\int g^2(t)dt$ exists and is finite, or *"the roughness of* $K(u)$" [4]. So

$$MSE(\hat{f}(x_0,\underset{\sim}{X}))=\frac{f(x_0)}{nh}R(K)-\frac{(f(x_0))^2}{n}+\frac{h^4}{4}[f''(x_0)\mu_2]^2 +o(n^{-1})+o(h^4). \quad (7)$$

The mean integrated squared error (*MISE*) of $\hat{f}(x,\underset{\sim}{X})$, obtained by integrating the *MISE* of $\hat{f}(x,\underset{\sim}{X})$, is an error measure over the real line, i.e.

$$MISE(\hat{f})=\int MSE(\hat{f}(x,\underset{\sim}{X}))dx$$
$$=\frac{R(K)}{nh}-\frac{1}{n}\int f^2(x)dx+\frac{h^4\mu_2^2R(f'')}{4}+o((nh)^{-1})+o(h^4). \quad (8)$$

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

$$AMISE(\hat{f})=\frac{R(K)}{nh}-\frac{R(f)}{n}+\frac{h^4\mu_2^2R(f'')}{4}. \quad (9)$$

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 [5]. A numerical method for comparing the $AMISE(\hat{f})$ in various populations are needed.

2. New Kernel Functions

In this study we give two new 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 are small. A symmetric kernel function with small variance, μ_2, is used 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 $\left|B(\hat{f}(x_0,\underset{\sim}{X}))\right|$ and the variance of $\hat{f}(x_0, \underset{\sim}{X}))$ the kernel function is 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.* $MSE(\hat{f}(x_0, \underset{\sim}{X}))$, $MISE(\hat{f})$ and $AMISE(\hat{f})$ depend on the sum of the squared bias and the variance. So the kernel function that minimizes the sum of $R(K)$ and μ_2^2 are 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 Muller [6], Gasser, Muller & Mammitzsch [7], Granovsky & Muller [8], Muller & Wang [9], Horova [10], Delaigle & Hall [11], and Mammitzsch [12]. Horova [10] presented the construction of kernel functions that minimize the squared L_2 norm of $K(u)$ under the condition that the moments of $K(u)$ are 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];

$$K(u) = \sum_{i=0}^{2} c_i u^i \ , \tag{10}$$

where c_i are the coefficients to be determined in order 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 function $K_1(u)$ under the constraints that minimizes $A_1(K)$ let

$$L_1(K) = A_1(K) + \lambda_1(1-\mu_0) + \lambda_2(\mu_1) \tag{11}$$

where λ_1, λ_2 are Lagrange multipliers. The result is

$$K_1(u) = (\frac{2}{3} - \frac{u^2}{2}) I_{[-1,1]}(u) . \tag{12}$$

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

$$L_2(K) = A_2(K) + \lambda_1(1-\mu_0) + \lambda_2(\mu_1) \tag{13}$$

where λ_1, λ_2 are Lagrange multipliers. The result is

$$K_2(u) = (\frac{63}{106} - \frac{15u^2}{53}) I_{[-1,1]}(u)\,. \tag{14}$$

The kernel density estimates using the proposed two kernel functions $K_1(u)$ and $K_2(u)$ are then obtained from equation (1) at the given sample observations $\underset{\sim}{x} = (x_1, \ldots, x_n)$.

3. Simulation Study

In order to study the $AMISE$ of the kernel estimates of a density function, a simulation study is carried out under various kernel functions and bandwidths. The effects of the kernels and bandwidths of the estimates of $f(x)$ for different sample sizes are considered. The simulations are performed using programs written in R. Symmetric and asymmetric multimodal populations of size 50,000 are generated for each of the populations which are built from normal mixtures [13]. In this study, nine different normal mixture distributions are simulated, namely Gaussian $N(0,1)$, skewed unimodal $\frac{1}{5}N(0,1)+\frac{1}{5}N(\frac{1}{2},(\frac{2}{3})^2)+\frac{3}{5}N(\frac{13}{12},(\frac{5}{9})^2)$, kurtotic unimodal $\frac{2}{3}N(0,1)+\frac{1}{3}N(0,(\frac{1}{10})^2)$, separated bimodal $\frac{1}{2}N(-\frac{3}{2},(\frac{1}{2})^2)+\frac{1}{2}N(\frac{3}{2},(\frac{1}{2})^2)$, trimodal $\frac{9}{20}N(-\frac{6}{5},(\frac{3}{5})^2)+\frac{9}{20}N(\frac{6}{5},(\frac{3}{5})^2)+\frac{1}{10}N(0,(\frac{1}{4})^2)$, double claw $\frac{49}{100}N(-1,(\frac{2}{3})^2)+\frac{49}{100}N(1,(\frac{2}{3})^2)+\sum_{i=0}^{6}\frac{1}{350}N(\frac{(i-3)}{2},(\frac{1}{100})^2)$, asymmetric claw $\frac{1}{2}N(0,1)+\sum_{i=-2}^{2}\frac{2^{1-i}}{31}N((i+1/2),(\frac{2^{-i}}{10})^2)$, asymmetric double claw $\sum_{i=0}^{1}\frac{46}{100}N(2i-1,(\frac{2}{3})^2)+\sum_{i=1}^{3}\frac{1}{300}N(-\frac{i}{2},(\frac{1}{100})^2)+\sum_{i=1}^{3}\frac{7}{300}N(\frac{i}{2},(\frac{7}{100})^2)$, and smooth comp $\sum_{i=0}^{5}\frac{2^{5-i}}{63}N(\frac{65-96*2^{-i}}{21},(\frac{32}{63*2^i})^2)$. Random samples of sizes 50, 100, 200 and 500 are drawn from the generated population, each repeats 1,000 times. The bandwidths used in the simulation studies are the Silverman rule of thump (SRT) bandwidth (the commonly used quick and simple idea for selecting the bandwidth, most popular and easy to implement [14], two-stage direct plug-in bandwidth (DPI) [5] (it provides a good estimates for a data-based bandwidth [15], and the solve-the-equation method bandwidth (STE) ([5]) (the solve-the-equation method bandwidth is the best in term of overall performance [14,16]. The $AMISE$ of the kernel estimates are computed. From 1,000 samples with specified sizes, the mean of $AMISE(\hat{f})$, $\overline{AMISE}(\hat{f})$, of kernel estimates are computed and compared.

4. Results of the Simulation

The values of $\overline{AMISE}(\hat{f})$ and the estimated standard deviation (SD) by distributions, bandwidths, and various sample sizes with 1,000 replications are shown in Tables 1-7. The bold number is the smallest $\overline{AMISE}(\hat{f})$ for each bandwidths and various sample sizes.

For data from Gaussian or skewed unimodal distribution, the kernel estimates using the proposed K_1 and K_2 perform better than the uniform and Gaussian estimates as shown in Tables 1, 2. 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 with Gaussian distribution. For data from skewed unimodal 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 or STE bandwidth. The $\overline{AMISE}$ of $\hat{f}(x,\underset{\sim}{x})$ is close to zero as the sample size gets larger.

Table 1. $\overline{AMISE}(\hat{f})$ of kernel estimates for Gaussian distribution.

n	Kernel functions	h_{SRT} $\overline{AMISE}(\hat{f})$	h_{SRT} SD	h_{DPI} $\overline{AMISE}(\hat{f})$	h_{DPI} SD	h_{STE} $\overline{AMISE}(\hat{f})$	h_{STE} SD
50	Uniform	0.010929	0.001674	0.010241	0.001624	0.010878	0.002721
	Epanechnikov	**0.009978**	0.001575	**0.009336**	0.001529	**0.009936**	0.002561
	Gaussian	0.010616	0.00164	0.009948	0.001591	0.010572	0.002665
	K_1	0.010027	0.001583	0.009377	0.001536	0.009979	0.002573
	K_2	0.010231	0.001604	0.009573	0.001556	0.010183	0.002606
100	Uniform	0.006476	0.000627	0.006076	0.000556	0.006304	0.00101
	Epanechnikov	**0.005943**	0.000589	**0.005569**	0.000522	**0.005784**	0.000949
	Gaussian	0.006301	0.000613	0.005912	0.000544	0.006135	0.000988
	k1	0.00597	0.000593	0.005592	0.000526	0.005807	0.000955
	k2	0.006085	0.0006	0.005701	0.000533	0.00592	0.000967
200	Uniform	0.003834	0.000223	0.003603	0.000202	0.003666	0.000342
	Epanechnikov	**0.003533**	0.000209	**0.003317**	0.000189	**0.003377**	0.000322
	Gaussian	0.003735	0.000218	0.003511	0.000197	0.003573	0.000335
	K_1	0.003549	0.000211	0.00333	0.000191	0.00339	0.000324
	K_2	0.003613	0.000213	0.003392	0.000193	0.003453	0.000328
500	Uniform	0.001925	6.17E-05	0.001814	4.24E-05	0.001825	6.66E-05
	Epanechnikov	**0.001783**	5.79E-05	**0.001679**	3.97E-05	**0.001689**	6.24E-05
	Gaussian	0.001878	6.03E-05	0.00177	4.13E-05	0.001781	6.49E-05
	K_1	0.00179	5.83E-05	0.001685	4.01E-05	0.001695	6.3E-05
	K_2	0.00182	5.91E-05	0.001714	4.06E-05	0.001724	6.38E-05

Table 2. $\overline{AMISE}(\hat{f})$ of kernel estimates for skewed unimodal distribution.

n	Kernel functions	h_{SRT} $\overline{AMISE}(\hat{f})$	h_{SRT} SD	h_{DPI} $\overline{AMISE}(\hat{f})$	h_{DPI} SD	h_{STE} $\overline{AMISE}(\hat{f})$	h_{STE} SD
50	Uniform	0.016285	0.002614	0.019509	0.006874	0.019624	0.007043
	Epanechnikov	**0.014942**	0.00248	**0.018003**	0.006519	**0.018108**	0.006678
	Gaussian	0.01586	0.002582	0.019046	0.006785	0.019156	0.00695
	K_1	0.014989	0.002472	0.018037	0.0065	0.018147	0.00666
	K_2	0.015283	0.002504	0.018371	0.006585	0.018482	0.006747
100	Uniform	0.009645	0.00096	0.011203	0.002628	0.011129	0.00267
	Epanechnikov	**0.008889**	0.000913	**0.010369**	0.002494	**0.010297**	0.002533
	Gaussian	0.009406	0.00095	0.010946	0.002596	0.010872	0.002637
	k1	0.008915	0.000908	0.010388	0.002486	0.010318	0.002525
	k2	0.00908	0.00092	0.010573	0.002518	0.010502	0.002558
200	Uniform	0.005712	0.000368	0.006431	0.000935	0.006355	0.000943
	Epanechnikov	**0.005284**	0.00035	**0.005967**	0.000887	**0.005895**	0.000895
	Gaussian	0.005577	0.000364	0.006288	0.000924	0.006212	0.000932
	K_1	0.005299	0.000348	0.005979	0.000884	0.005907	0.000892
	K_2	0.005392	0.000352	0.006081	0.000895	0.006008	0.000903
500	Uniform	0.002861	0.000112	0.003138	0.000294	0.003109	0.000299
	Epanechnikov	**0.002657**	0.000107	**0.00292**	0.000279	**0.002893**	0.000284
	Gaussian	0.002797	0.000111	0.00307	0.000291	0.003042	0.000295
	K_1	0.002664	0.000106	0.002926	0.000278	0.002899	0.000283
	K_2	0.002709	0.000107	0.002974	0.000282	0.002947	0.000286

For sample data from a population distributed as kurtotic unimodal, the $\overline{AMISE}$ of $\hat{f}(x,\underset{\sim}{x})$ with STE bandwidth is lower than the $\overline{AMISE}$ of the other kernel estimates and $\hat{f}(x,\underset{\sim}{x})$ using K_1 performs well when sample sizes are 50 and 100 as in Table 3. For data from separated bimodal distribution, from Table 4, 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. From Table 4, the $\overline{AMISE}$ of $\hat{f}(x,\underset{\sim}{x})$ with STE bandwidth is lower than the $\overline{AMISE}$ of $\hat{f}(x,\underset{\sim}{x})$ with SRT or DPI bandwidth. For data from kurtotic unimodal and separated 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.

Table 3. $\overline{AMISE}(\hat{f})$ of kernel estimates for kurtotic unimodal 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	1.708632	2.436767	1.221782	3.575027	0.675117	3.409776
	Epanechnikov	1.617245	2.307618	1.156175	3.385528	0.638445	3.229033
	Gaussian	1.683792	2.401856	1.203893	3.523785	0.66502	3.360899
	K_1	**1.615013**	2.304207	**1.154648**	3.380546	**0.637721**	3.224283
	K_2	1.636223	2.334239	1.169857	3.424607	0.646194	3.266308
100	Uniform	0.76216	0.742883	0.261676	0.296474	0.108328	0.091476
	Epanechnikov	0.721346	0.703517	0.247366	0.280776	0.102119	0.086647
	Gaussian	0.751055	0.732247	0.257719	0.292242	0.106541	0.090186
	k1	**0.720364**	0.70247	**0.247106**	0.280345	**0.102101**	0.0865
	k2	0.729833	0.711626	0.250407	0.283999	0.103512	0.087627
200	Uniform	0.369239	0.280356	0.080956	0.055355	0.042642	0.01631
	Epanechnikov	0.349454	0.265501	0.07643	0.05243	**0.040131**	0.015455
	Gaussian	0.36385	0.276343	0.079677	0.054571	0.041896	0.016086
	K_1	**0.348985**	0.265105	**0.076385**	0.052343	0.040155	0.015422
	K_2	0.353573	0.26856	0.07742	0.053025	0.040719	0.015623
500	Uniform	0.155967	0.078679	0.023782	0.006618	0.016641	0.002169
	Epanechnikov	0.147611	0.074511	**0.022418**	0.00627	**0.015648**	0.002057
	Gaussian	0.153689	0.077554	0.023383	0.006527	0.016337	0.002141
	K_1	**0.147415**	0.074399	0.022421	0.006258	0.015669	0.002051
	K_2	0.149353	0.075369	0.02273	0.006339	0.015889	0.002077

Table 4. $\overline{AMISE}(\hat{f})$ of kernel estimates for separated 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	0.160904	0.029481	0.030682	0.005448	0.022796	0.002676
	Epanechnikov	0.152033	0.027921	**0.028687**	0.005164	**0.021205**	0.002541
	Gaussian	0.158472	0.029061	0.030089	0.005375	0.022301	0.002645
	K_1	**0.151844**	0.027878	0.028706	0.005152	0.021249	0.00253
	K_2	0.153897	0.028241	0.029154	0.005219	0.021599	0.002563
100	Uniform	0.092554	0.01118	0.015327	0.001445	0.012937	0.000843
	Epanechnikov	0.087474	0.010588	**0.014322**	0.001371	**0.012052**	0.000802
	Gaussian	0.091162	0.01102	0.015022	0.001427	0.01266	0.000834
	k1	**0.087366**	0.010572	0.01434	0.001367	0.01208	0.000797
	k2	0.088541	0.010709	0.014563	0.001385	0.012274	0.000808
200	Uniform	0.053486	0.004379	0.008215	0.000449	0.007457	0.000287
	Epanechnikov	0.050561	0.004147	**0.007679**	0.000426	**0.006958**	0.000274
	Gaussian	0.052684	0.004316	0.00805	0.000443	0.0073	0.000285
	K_1	**0.050499**	0.00414	0.007691	0.000424	0.006975	0.000272
	K_2	0.051176	0.004194	0.00781	0.00043	0.007084	0.000275
500	Uniform	0.025703	0.001405	0.003792	0.000105	0.003622	7.04E-05
	Epanechnikov	0.024304	0.001331	**0.003547**	0.0001	**0.003386**	6.73E-05
	Gaussian	0.02532	0.001385	0.003715	0.000104	0.003547	7E-05
	K_1	**0.024274**	0.001329	0.003555	9.96E-05	0.003394	6.66E-05
	K_2	0.024598	0.001346	0.003608	0.000101	0.003446	6.75E-05

From Table 5, the data from trimodal distribution, $\hat{f}(x,\underset{\sim}{x})$ using K_1 with the STE bandwidth gives a smaller $\overline{AMISE}(\hat{f})$ than the others $\hat{f}(x,\underset{\sim}{x})$ with the SRT or DPI bandwidth for samples of sizes 50 and 100. The $\overline{AMISE}$ of $\hat{f}(x,\underset{\sim}{x})$ are close to zero as the sample size gets large.

For data from double claw distribution, the $\overline{AMISE}$ of $\hat{f}(x,\underset{\sim}{x})$ using K_1 with the SRT bandwidth is lower than the $\overline{AMISE}(\hat{f})$ of the others $\hat{f}(x,\underset{\sim}{x})$ with the STE or DPI bandwidth for sample of size 50. For samples of sizes between 100 and 500, the $\overline{AMISE}$ of $\hat{f}(x,\underset{\sim}{x})$ using K_1 with STE bandwidth is lower than the $\overline{AMISE}(\hat{f})$ of the others $\hat{f}(x,\underset{\sim}{x})$ with the SRT or DPI bandwidth. The $\overline{AMISE}$ of all $\hat{f}(x,\underset{\sim}{x})$ decrease as the sample size increase.

Table 5. $\overline{AMISE}(\hat{f})$ of kernel estimates for trimodal 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	0.06103	0.01464	0.062204	0.03317	0.050482	0.036083
	Epanechnikov	0.057503	0.013867	0.058613	0.031419	0.047505	0.034181
	Gaussian	0.060036	0.014433	0.061191	0.032702	0.049629	0.035577
	K_1	**0.057464**	0.013843	**0.058575**	0.031365	**0.047491**	0.03412
	K_2	0.058272	0.014024	0.059397	0.031774	0.048168	0.034564
100	Uniform	0.035266	0.005868	0.029997	0.012964	0.02421	0.013325
	Epanechnikov	0.033247	0.005558	0.028255	0.01228	0.022771	0.012624
	Gaussian	0.034697	0.005785	0.029501	0.012782	0.023793	0.01314
	k1	**0.033225**	0.005549	**0.028243**	0.012258	**0.022771**	0.0126
	k2	0.033687	0.005621	0.02864	0.012418	0.023097	0.012765
200	Uniform	0.020294	0.002268	0.014262	0.004484	0.011583	0.004342
	Epanechnikov	0.019142	0.002148	0.013427	0.004248	**0.010887**	0.004114
	Gaussian	0.019969	0.002236	0.014021	0.004422	0.011378	0.004282
	K_1	**0.019129**	0.002145	**0.013425**	0.00424	0.010891	0.004106
	K_2	0.019393	0.002173	0.013615	0.004296	0.011048	0.004159
500	Uniform	0.009795	0.000696	0.005635	0.001005	0.004759	0.000926
	Epanechnikov	0.009244	0.000659	**0.005302**	0.000953	**0.004472**	0.000877
	Gaussian	0.00964	0.000686	0.005537	0.000992	0.004673	0.000913
	K_1	**0.009238**	0.000658	0.005304	0.000951	0.004476	0.000875
	K_2	0.009364	0.000667	0.005379	0.000963	0.00454	0.000887

Table 6. $\overline{AMISE}(\hat{f})$ of kernel estimates for double 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	1864.231	561.3389	2316.902	1541.768	1877.222	1757.202
	Epanechnikov	1765.401	531.5801	2194.074	1460.032	1777.702	1664.046
	Gaussian	1837.496	553.2886	2283.675	1519.657	1850.3	1732.002
	K_1	**1762.817**	530.802	**2190.862**	1457.895	**1775.1**	1661.61
	K_2	1785.793	537.7203	2219.417	1476.897	1798.236	1683.267
100	Uniform	1062.739	217.1984	1116.853	660.1903	902.1075	715.8835
	Epanechnikov	1006.399	205.6838	1057.644	625.1909	854.283	677.9317
	Gaussian	1047.498	214.0835	1100.836	650.7224	889.1701	705.6169
	k1	**1004.926**	205.3828	**1056.096**	624.2759	**853.0326**	676.9394
	k2	1018.024	208.0597	1069.861	632.4125	864.1508	685.7624
200	Uniform	609.9593	88.11598	501.7934	246.2819	397.1231	256.8389
	Epanechnikov	577.6228	83.4446	475.1912	233.2255	376.0699	243.2228
	Gaussian	601.2117	86.85229	494.5971	242.7499	391.4278	253.1555
	K_1	**576.7774**	83.32246	**474.4957**	232.8842	**375.5195**	242.8668
	K_2	584.2949	84.40846	480.6802	235.9195	380.4139	246.0323
500	Uniform	293.7504	27.15246	184.725	62.10136	151.6257	61.42863
	Epanechnikov	278.1775	25.713	174.932	58.80911	143.5873	58.17205
	Gaussian	289.5377	26.76306	182.0758	61.21075	149.4511	60.54767
	K_1	**277.7704**	25.67537	**174.6759**	58.72304	**143.3772**	58.08691
	K_2	281.3907	26.01001	176.9526	59.48841	145.2459	58.84399

For data from asymmetric claw distribution and samples of sizes 50 and 100, the $\overline{AMISE}$ of $\hat{f}(x,\underset{\sim}{x})$ with the SRT bandwidth is lower than the $\overline{AMISE}(\hat{f})$ of the others $\hat{f}(x,\underset{\sim}{x})$ with the STE or DPI bandwidth. For samples of sizes 200 and 500, the $\overline{AMISE}$ of $\hat{f}(x,\underset{\sim}{x})$ with the STE bandwidth is lower than the $\overline{AMISE}$ of the others $\hat{f}(x,\underset{\sim}{x})$ with the SRT or DPI bandwidth.

Table 7. $\overline{AMISE}(\hat{f})$ of kernel estimates for asymmetric 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	25.274	10.56433	41.96607	26.00978	36.83363	29.79249
	Epanechnikov	23.93366	10.00428	39.74082	24.6309	34.88047	28.21308
	Gaussian	24.91137	10.41283	41.36406	25.63678	36.30522	29.36524
	K_1	**23.89868**	9.989625	**39.6827**	24.59484	**34.82947**	28.17177
	K_2	24.21027	10.11983	40.20001	24.9154	35.28352	28.53895
100	Uniform	14.60458	3.988819	22.67927	10.76772	19.32426	13.21801
	Epanechnikov	13.8301	3.777358	21.47672	10.19688	18.29957	12.51728
	Gaussian	14.39504	3.931617	22.35393	10.6133	19.04704	13.02846
	k1	**13.80988**	3.771827	**21.44531**	10.18196	**18.27282**	12.49895
	k2	13.98993	3.820988	21.72487	10.31466	18.51103	12.66186
200	Uniform	8.517721	1.56352	11.23	4.42696	8.163123	6.054571
	Epanechnikov	8.066043	1.480633	10.63453	4.19227	7.730241	5.733599
	Gaussian	8.395519	1.541098	11.0689	4.363474	8.046004	5.967746
	K_1	**8.054254**	1.478465	**10.61898**	4.186133	**7.718946**	5.725202
	K_2	8.159255	1.497734	10.75741	4.240693	7.819577	5.799822
500	Uniform	4.101475	0.476412	3.211195	1.168112	1.024393	0.978196
	Epanechnikov	3.88399	0.451156	3.040906	1.106186	0.970031	0.92634
	Gaussian	4.042634	0.46958	3.165121	1.151361	1.009676	0.96417
	K_1	**3.878313**	0.450495	**3.036464**	1.104566	**0.968625**	0.924982
	K_2	3.928872	0.456367	3.07605	1.118963	0.981259	0.937038

For data from a population distributed as asymmetric double claw and smooth comp, the $\hat{f}(x,\underset{\sim}{x})$ with the STE bandwidth gives lower $\overline{AMISE}(\hat{f})$ than $\hat{f}(x,\underset{\sim}{x})$ with the SRT or DPI bandwidth as shown in Tables 8 and 9, respectively.

For data from asymmetric claw, asymmetric double claw and smooth comp distributions, the $\hat{f}(x,\underset{\sim}{x})$ using the proposed K_1 performs well. The $\hat{f}(x,\underset{\sim}{x})$ with the proposed K_2 performs better than the $\hat{f}(x,\underset{\sim}{x})$ with uniform, Gaussian functions as shown in Tables 7-9.

Table 8. $\overline{AMISE}(\hat{f})$ of kernel estimates for asymmetric double 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	1064.488	329.0727	1207.742	809.28	963.3038	893.6315
	Epanechnikov	1008.055	311.6272	1143.715	766.3768	912.2349	846.2565
	Gaussian	1049.222	324.3534	1190.421	797.6739	949.4887	880.8158
	K_1	**1006.579**	311.1711	**1142.041**	765.2551	**910.8997**	845.0179
	K_2	1019.699	315.2268	1156.926	775.2292	922.7722	856.0316
100	Uniform	605.3458	127.3865	551.4556	328.0383	428.9463	350.8718
	Epanechnikov	573.2538	120.6332	522.2206	310.6477	406.206	332.2707
	Gaussian	596.6643	125.5596	543.547	323.3339	422.7946	345.8399
	k1	**572.4148**	120.4566	**521.4562**	310.193	**405.6115**	331.7844
	k2	579.8755	122.0266	528.2528	314.236	410.8981	336.1088
200	Uniform	348.5102	51.31864	247.881	121.2693	193.1326	122.6449
	Epanechnikov	330.0342	48.59803	234.7398	114.8404	182.8938	116.143
	Gaussian	343.5121	50.58267	244.3261	119.5302	190.3628	120.886
	K_1	**329.5512**	48.5269	**234.3962**	114.6723	**182.6261**	115.973
	K_2	333.8464	49.15938	237.4512	116.1669	185.0064	117.4845
500	Uniform	167.9083	15.01381	87.99997	27.26422	69.85181	25.67868
	Epanechnikov	159.0067	14.21787	83.3347	25.81883	66.14865	24.31735
	Gaussian	165.5003	14.7985	86.73793	26.87322	68.85003	25.31042
	K_1	**158.774**	14.19706	**83.21274**	25.78104	**66.05184**	24.28176
	K_2	160.8434	14.3821	84.29731	26.11706	66.91274	24.59824

Table 9. $\overline{AMISE}(\hat{f})$ of kernel estimates for smooth comp 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	3287.642	796.9182	853.4582	405.0623	225.7792	149.1765
	Epanechnikov	3113.35	754.6704	808.2126	383.5883	213.8093	141.2681
	Gaussian	3240.493	785.4894	841.2184	399.2532	222.5411	147.0372
	K_1	**3108.794**	753.5658	**807.0297**	383.0269	**213.4965**	141.0613
	K_2	3149.313	763.3875	817.5483	388.0191	216.2792	142.8999
100	Uniform	1863.589	318.6248	278.9935	99.16704	76.31698	33.59005
	Epanechnikov	1764.793	301.7332	264.2027	93.9098	72.27091	31.80931
	Gaussian	1836.863	314.0554	274.9923	97.74487	75.22241	33.10834
	k1	**1762.21**	301.2916	**263.8161**	93.77234	**72.16518**	31.76275
	k2	1785.178	305.2185	267.2546	94.99454	73.10579	32.17674
200	Uniform	1070.634	129.5253	97.80193	22.71941	27.78746	9.245328
	Epanechnikov	1013.875	122.6586	92.61695	21.51497	26.31423	8.755199
	Gaussian	1055.28	127.6677	96.39928	22.39359	27.3889	9.112742
	K_1	10**12.391**	122.4791	**92.48142**	21.48347	**26.27575**	8.742382
	K_2	1025.586	124.0754	93.68681	21.76348	26.61823	8.856327
500	Uniform	513.0715	37.13692	26.1958	3.752915	7.347895	1.504265
	Epanechnikov	485.8715	35.16814	24.80702	3.553958	6.958305	1.424519
	Gaussian	505.7135	36.60433	25.8201	3.699094	7.24249	1.482693
	K_1	**485.1604**	35.11667	**24.77072**	3.548756	**6.948138**	1.422433
	K_2	491.4838	35.57437	25.09358	3.595009	7.038705	1.440973

5. Conclusions

When sample sizes are 50 and 100, the $\hat{f}(x,\underset{\sim}{x})$ using K_1 with STE bandwidth performs well for sample data from populations distributed as kurtotic unimodal or trimodal. For data from multimodal distributions, the $\hat{f}(x,\underset{\sim}{x})$ using the proposed K_1 performs well, and the $\hat{f}(x,\underset{\sim}{x})$ using the proposed K_2 perform better than the $\hat{f}(x,\underset{\sim}{x})$ with uniform, Gaussian functions.

The estimates with DPI bandwidth perform well when the sample data are from Gaussian distribution. For data withfrom skewed unimodal distributions, the estimates with SRT bandwidth perform well which is the same as data from asymmetric claw population with samples of sizes 50 and 100. For sample from populations distributed as kurtotic unimodal, separated bimodal or multimodal, the $\overline{AMISE}$ of $\hat{f}(x,\underset{\sim}{x})$ with STE bandwidth is lower than the $\overline{AMISE}$ of the other kernel estimates except for data with double claw distribution and the sample sizes are small.

For sample from unimodal (Gaussian, skewed, kurtotic), separated bimodal or trimodal distributions, the $\overline{AMISE}$ of $\hat{f}(x,\underset{\sim}{x})$ are closed to zero as the sample size gets larger. For large sample size, $\overline{AMISE}$ of the estimates $\hat{f}(x,\underset{\sim}{x})$ using Epanechnikov, K_1 or K_2 functions become closer. The $\overline{AMISE}(\hat{f})$ becomes smaller as the sample sizes increase which implies 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, is large because the bandwidth is far from the optimal bandwidth which is consistent with the degree of estimation difficulty which increases with skewness, kurtosis and multimodality of the distributions [5].

References

[1] Rosenblatt, M. Remark on some Nonparametric Estimates of a Density Function. *Annals of Mathematical Statistics*, 1956; 27: 832-837.

[2] Parzen, E. On Estimation of a Probability Density Function and Mode, *Annals of Mathematical Statistics*, 1962; 33: 1065-1076.

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

[4] Scott, D. W. *Multivariate Density Estimation*, New York: Wiley,1992.

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

[6] Muller, H. G. Smooth Optimum Kernel Estimators of Densities, Regression Curves and Modes, *The Annals of Statistics*, 1984; 12: 766-774.

[7] Gasser, T., Muller, H-G., and Mammitzsch, V. Kernels for Nonparametric Curve Estimation, *Journal of the Royal statistical Society. Series B*, 1985; 47: 238-252.

[8] Granovsky, B. L. and Muller, H. G. Optimizing Kernel Methods: A Unifying Variational Principle, *International Statistical Review*, 1991; 59: 373-388.

[9] Muller, H. G. and Wang, J. L. Hazard Rate Estimation Under Random Cencoring with Varying Kernels and Bandwidths, *Biometrics*, 1994; 50: 61-76.

[10] Horova, I. Some Remarks on Kernels, *Journal of Computational Analysis and Applications*, 2000; 2: 253-263.

[11] Delaigle, A. and Hall, P. On Optimal Kernel Choice for Deconvolution, *Statistics and Probability Letters*, 2006; 76: 1594-1602.

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

[13] Marron, J. S. and Wand, M. P. Exact Mean Integrated Squared Error, *The Annals of Statistics*, 1992; 20: 712-736.

[14] Liu, K. *Kernel Estimates of Statistical Distribution Functions*, Doctoral Dissertation, University of South Florida, 1998.

[15] Jeter, J. *Comparison Between the Exact and the Approximate Bandwidths in the Kernel Density Estimation*, Master's thesis, Southern Illinois University, 2005.

[16] Jones, M. C., Marron, J. S. and Sheather S. J. A Brief Survey of Bandwidth Selection for Density Estimation, *Journal of the American Statistical Association*, 1996; 91; 401 -407.