Journal of Scientific Research & Reports
11(5): 1-9, 2016; Article no.JSRR.27756
ISSN: 2320-0227

SCIENCEDOMAIN *international*
www.sciencedomain.org

The Finite Sample Performance of Modified Adaptive Kernel Estimators for Probability Density Function

Serpil Cula[1], Serdar Demir[2*] and Oniz Toktamis[3]

[1]*Department of Insurance and Risk Management, Faculty of Commercial Sciences, Baskent University, Ankara, Turkey.*
[2]*Department of Statistics, Faculty of Sciences, Mugla Sitki Kocman University, Mugla, Turkey.*
[3]*Department of Statistics, Faculty of Science, Hacettepe University, Ankara, Turkey.*

Authors' contributions

This work was carried out in collaboration between all authors. These authors contributed equally to this work. All authors read and approved the final manuscript.

Article Information

DOI: 10.9734/JSRR/2016/27756
Editor(s):
(1) Chun-Gang Zhu, School of Mathematical Sciences, Dalian University of Technology, China.
Reviewers:
(1) Yousri Slaoui, University of Poitiers, France.
(2) Fethi Soltani, Higher College of Technology and Informatics, Tunis, Tunisia.
(3) Hasni Neji, University of Carthage, Tunisia.
Complete Peer review History: http://www.sciencedomain.org/review-history/15543

Original Research Article

Received 17th June 2016
Accepted 19th July 2016
Published 27th July 2016

ABSTRACT

It is well-known that the most popular probability density estimator is kernel density estimator in literature. Adaptive kernel density estimators are generally preferred for data with long tailed densities. In this paper, the adaptive kernel estimators for probability density function are studied. A modified adaptive kernel estimator is investigated. For finite sample performance comparisons, the root mean squared errors of the fixed and the adaptive kernel estimations are computed for simulated samples from various density distributions. The simulation results show that the modified adaptive kernel density estimators have better performance than the classical adaptive kernel density estimator.

Keywords: Density estimation; kernel estimator; adaptive kernel estimator; variable bandwidth.

**Corresponding author: E-mail: serdardemir@mu.edu.tr;*

1. INTRODUCTION

The kernel estimation method which is still one of the contemporary issues is a non-parametric estimation method on which many studies have been carried out so far. The kernel estimation methods are widely used in many areas of statistics such as the estimations of probability density, regression function and spectral density function.

Let $x_1, x_2, ..., x_n$ be randomly chosen sample from unknown probability density function *f*, the kernel estimator of a probability density function for any point of *x* is given as

$$\hat{f}(x) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h}K\left(\frac{x-x_i}{h}\right), \quad (1)$$

where *K* is the kernel function and generally a probability density function with a single mode and symmetrical around zero [1,2,3]. Here *h* is the smoothing parameter which is also called as window width or bandwidth. The value of kernel estimation at any point is a weighted mean that considers this point and neighbor observations with certain weights as this point is to be the center. Such weights are obtained by means of the kernel function *K* and the bandwidth *h*. In practice, however, the kernel function *K* and the bandwidth *h* are selected by the user. Epanechnikov [4] has shown that there is an optimal kernel function in one sense, but other kernel functions have results closer to the values obtained by Epanechnikov kernel. Therefore, in practice, the selection of kernel function is not as important as the selection of bandwidth and such selection is made by taking into consideration of the ease of calculation and differentiability features.

In the kernel estimation of probability density function the selection of bandwidth has a significant role [1,5,6,7]. Boneva et al. [8] have shown that minor changes on the bandwidth have caused significant changes on the estimations. The bandwidth *h* controls the smoothing degree. In other words, the selection of the bandwidth is very important in terms of the performance of the estimator. There are many researches in the literature, which have focused on choosing the proper value of the bandwidth.

Various criteria on the difference between the probability density estimation and the true probability density function *f* are used to evaluate the performance of the kernel density estimator. The most widely used one is the mean squared error (MSE) first suggested by Rosenblatt [9]. MSE regarding the kernel density estimator is given as,

$$\text{MSE}\{\hat{f}(x)\} = (nh)^{-1} f(x)\int_{-\infty}^{\infty}K^2(u)du + \frac{1}{4}h^4\left\{f''(x)\int_{-\infty}^{\infty}u^2K(u)du\right\}^2 + o((nh)^{-1}) + o(h^4),$$

where $u = (x - x_i)/h$. It can be written as,

$$\text{MSE}\{\hat{f}(x)\} = (nh)^{-1} f(x)\,R(K) + \frac{1}{4}h^4\{f''(x)\,\mu_2(K)\}^2 + o((nh)^{-1}) + o(h^4), \quad (2)$$

where $R(K) = \int K^2(u)du$ and $\mu_2(K) = \int u^2 K(u)\,du$ [1]. The mean integrated squared error (MISE) of the kernel density estimator is given as

$$\text{MISE}\{\hat{f}(x)\} = (nh)^{-1}R(K) + \frac{1}{4}h^4\{\mu_2(K)\}^2\int_{-\infty}^{\infty}f''(x)^2dx + o\{(nh)^{-1} + h^4\}. \quad (3)$$

MISE cannot be exactly derived from the Equation (3) [7]. Therefore, asymptotic integrated mean squared error (AMISE) is used. AMISE can be written by using the Equation (3) as:

$$\text{AMISE}\{\hat{f}(x)\} = (nh)^{-1}R(K) + \frac{1}{4}h^4\{\mu_2(K)\}^2\int_{-\infty}^{\infty}f''(x)^2dx. \quad (4)$$

The optimal bandwidth value that minimizes the expression of AMISE is given as follows

$$h_{opt} = \mu_2(K)^{-2/5}\{R(K)\}^{1/5}\left\{\int_{-\infty}^{\infty} f''(x)^2\,dx\right\}^{-1/5} n^{-1/5}. \quad (5)$$

As seen in the Equation (5), the optimal bandwidth itself depends on the integral of the square of the second order derivative of the unknown density function being estimated. Therefore, the bandwidth value cannot be obtained from the Equation (5). For this reason, to find the bandwidth, some methods are proposed. Subjective selection method, the least squares cross validation method, the biased cross validation method, the bootstrap method, the plug-in method, and the smoothed cross validation method are some common bandwidth selection methods. A comparison of these methods is performed by Loader [6], Park and Marron [10], Cao et al. [11], Scott and Terrell [12], Sheather and Jones [13], and Horava and Zelinka [14]. As a more specific approach, Slaoui [15] proposed an automatic selection of the bandwidth for recursive kernel density estimator based on a stochastic approximation algorithm. Slaoui [15] showed that the recursive estimator has good performance for small samples.

In this study, the kernel density estimator given in the Equation (1) will be called as the fixed kernel density estimator to distinguish it from the adaptive kernel density estimator that will be given below. In practice, one of the undesirable features of the fixed kernel density estimator is its being inefficient at the tail parts of long tail distributions. While the selected fixed bandwidth is able to perform efficient smoothing around the mode of a distribution, it cannot perform any efficient smoothing at the tail parts. On the other hand, the selected fixed bandwidth values may be efficient at the tail parts of a distribution whereas they may destroy some important characteristics around the mode of a distribution. It is difficult to find a single bandwidth that adequately separates peaks and valleys if the data have a distribution with multi peaks; if the bandwidth is too large, it will result in eliminating significant modes by over smoothing the data; if the bandwidth is too small and then it may result in the appearance of misleading modes.

In the higher dimensional setting, the fixed kernel density estimators may result in untrue conclusion unless sample size is extremely large [16]. In general, the fixed kernel estimators will have difficulties with densities that exhibit large changes in magnitude. Those are the basic motivations for considering the kernel estimator that allows the bandwidth to vary one observation to another. As a result, kernel estimators with variable bandwidth are used in such cases. The variable bandwidth estimator was first introduced by Breiman et al. [17] who had stated that the bandwidth must be large around the parts where density is small and vice versa.

These estimators with variable bandwidth are grouped into two categories. The first group is called balloon estimators or the kth nearest neighbor estimators; the second group is called sample point estimators or adaptive kernel estimators. The first one varies the fixed bandwidth with the estimation point; the second one in which the bandwidth is varied with each sample point not with the estimation point. In this study, the adaptive kernel estimator will be considered.

2. ADAPTIVE KERNEL ESTIMATOR

The difference of the adaptive kernel density estimators and the fixed kernel estimators is the use of different bandwidth at each data point. The adaptive kernel density estimator which is first introduced by Breiman et. al. [17] is given as

$$\tilde{f}(x) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h(x_i)^d}K\left(\frac{x-x_i}{h(x_i)}\right), \quad (6)$$

where $h(x_i)$ is the variable bandwidth at each sample point and d is the number of dimension. Note that the bandwidth, $h(x_i)$, is written explicitly as a function of the sample point. Breiman et al. [17] suggested that $h(x_i)$ must be taken as being proportional to the distance from x_i to its kth nearest neighbor; Abramson [18] suggested that $h(x_i)$ must be proportional to $f^{-1/2}(x_i)$, with f replaced by a pilot estimate, for all dimensions; and Silverman [7] suggested that it must be proportional to $(g/\hat{f}(x_i))^{1/2}$ where g is the geometric mean of $\hat{f}(x_i)$ values.

The main feature of such methods is the use of different bandwidths for each data point. Muller et al. [19] have expressed that the mean squared error of the kernel estimator derived from the use of variable bandwidth is smaller than the mean squared error of kernel estimator derived from the use of the fixed bandwidth. A study carried out by Terrell and Scott [20] and the study by Hall and Marron [21] showed that in certain cases the adaptive kernel estimator has worse results than the fixed kernel estimator [22].

The mean squared error (MSE) of Abramson's estimator is studied by Silverman [7], Hall and Marron [21] and derived MSE expression by Jones [23] for $d = 1$ as follows:

$$\mathrm{MSE}\{\tilde{f}(x)\} \cong \frac{1}{576}\delta_K^2 h^8 A^2(x) + (nh)^{-1} S(K) f^{3/2}(x),$$

where

$$A(x) = \frac{d^4}{dx^4}\left[\frac{1}{f(x)}\right] \quad , \quad \delta_K = \int x^4 K(x)dx \quad ,$$
$$S(K) = \frac{3}{2}R(K) + \frac{1}{4}R(xK').$$

Silverman [7] gives a three-stage algorithm to obtain Abramson type estimator that he calls it first as adaptive kernel estimator.

At the first stage, a pilot estimate that satisfies $\hat{f}(x_i) > 0$ (for each *i*) is obtained. Any method can be applied to obtain a pilot estimate of the probability density function. However, the common method construct estimate is the fixed kernel estimation method.

At the second stage, Silverman defines the factor of local bandwidth as $\lambda_i = \{\hat{f}(x_i)/G\}^{-\alpha}$, where *G* is the geometric mean of the $\hat{f}(x_i)$ values and $\log G = n^{-1}\sum \log \hat{f}(x_i)$. The parameter α that satisfies the condition of $0 \le \alpha \le 1$ is called as sensitivity parameter.

At the third stage, taking $h(x_i) = h\lambda_i$, the adaptive kernel density estimator $\tilde{f}(x)$ can be obtained from the following expression:

$$\tilde{f}_G(x) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{(h\lambda_i)}K(h^{-1}\lambda_i^{-1}(x - x_i)).$$

The use of the local bandwidths depended on α enables flexibility to the method. The bigger α is, the more sensitive the system will be to the changes in pilot estimates and the distance between bandwidths used at different parts of sample will be higher too. If $\alpha = 0$, then the method becomes a kernel estimation with fixed bandwidth as all values of λ_i will be one. If $\alpha = 1$, then adaptive kernel estimator is equivalent to the nearest neighbor approximation [24]. According to the studies carried out by Abramson, taking the sensitive parameter for all dimensions as $\alpha = 0.5$ gives better results [18]. Many researchers have expressed that square root principle has better results for simulation studies with smaller samples. While calculating estimation at the final stage our bandwidth for each x_i point is equal to $h\lambda_i$. This estimate ensures that provided *K* will not be negative; and *K* is a density function, then the obtained estimate is a density function too [12,25].

Furthermore, Hall and Marron [21] also show that the adaptive kernel estimators have higher convergence rate than the fixed kernel estimators.

2.1 Modified Adaptive Kernel Estimator

In this study, having being inspired by the adaptive kernel estimator suggested by Silverman, we searched for the use of arithmetic mean of the pilot estimate of *f*, rather than geometric mean to be employed in the selection of $h(x_i)$; in other words, we searched for the effects and changes on the estimation if we take variable bandwidth which is proportional to $\sqrt{M/\hat{f}(x_i)}$ where *M* is the arithmetic mean of $\hat{f}(x_i)$ which refers to the pilot estimate of *f*. Furthermore, it is compared with both Silverman's adaptive estimator and the fixed kernel estimator. In the adaptive kernel estimation, taking arithmetic mean *M* of $\hat{f}(x_i)$ rather than geometric mean *g* in the equation of λ_i as,

$$\lambda_i' = \left(\hat{f}(x_i)/M\right)^{-1/2},$$

where $M = \sum_{i=1}^{n}\hat{f}(x_i)/n$ and writing the variable bandwidth as $h(x_i) = h\lambda_i'$ in the Equation (6), the modified adaptive kernel density estimation $\tilde{f}_M(x)$ is obtained as

$$\tilde{f}_M(x) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{(h\lambda_i')^d}K\left(\frac{x-x_i}{h\lambda_i'}\right) = \frac{1}{n}\sum_{i=1}^{n}\frac{\hat{f}^{d/2}(x_i)}{h^d m^{d/2}}K\left(\frac{x-x_i}{hm^{1/2}}\hat{f}^{1/2}(x_i)\right) \quad (7)$$

Using the approximated density function and the usual form for the expected value of this estimator can be written as

$$E\left\{\tilde{f}_M(x)\right\} \equiv \frac{1}{h^d}\int \frac{f^{d/2+1}(t)}{m^{d/2}}K\left[\frac{x-t}{hm^{1/2}}f^{1/2}(t)\right]dt, \quad (8)$$

where t is used instead of x_i for continuous case. The expected value of the estimator given in Equation (8) can be written as below;

$$E\left\{\tilde{f}_M(x)\right\} = \int f^{(d/2)+1}(x-hm^{1/2}v)\, K\left(v f(x-hm^{1/2}v)^{1/2}\right)dv,$$

where $v=(x-t)/(hm^{1/2})$ and the variance of the estimator in Equation (7) is obtained as

$$Var\left\{\tilde{f}_M(x)\right\} = (nh^d m^{d/2})^{-1}\int f^{d+1}(x-hm^{1/2}v)\, K^2\left(v f(x-hm^{1/2}v)^{1/2}\right)dv - n^{-1}\left\{E\left(\tilde{f}_2(x)\right)\right\}^2.$$

For $d=1$, the estimator in Equation (7) becomes as

$$\tilde{f}_M(x) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{(h\lambda_i')}K\left(\frac{x-x_i}{h\lambda_i'}\right) = \frac{1}{n}\sum_{i=1}^{n}\frac{f^{1/2}(x_i)}{h\ m^{1/2}}K\left(\frac{x-x_i}{hm^{1/2}}f^{1/2}(x_i)\right) \quad (9)$$

The expected value and the variance of this estimator are respectively given by

$$E\left\{\tilde{f}_M(x)\right\} = \int f^{3/2}(x-hm^{1/2}v)\, K\left(v f(x-hm^{1/2}z)^{1/2}\right)dv, \quad (10)$$

$$Var\left\{\tilde{f}_M(x)\right\} = (nhm^{1/2})^{-1}\int f^2(x-hm^{1/2}v)\, K^2\left(v f(x-hm^{1/2}v)^{1/2}\right)dv - n^{-1}\left\{E\left(\tilde{f}_2(x)\right)\right\}^2. \quad (11)$$

To find this expected value and the variance of the proposed adaptive kernel estimator, the processes similar to the ones performed with Abramson's estimator using the bandwidth $h/f^{1/2}(x)$ by Hall and Marron [21] are done. If $\eta = h(m/f(x))^{1/2}$ and $w = (f(x-w)/f(x))^{1/2}$ are included in Equation (10) and Equation (11) after the transformation of $v = z/f^{1/2}(x)$, then the expected value and variance of this estimator are found respectively as

$$E\left\{\tilde{f}_M(x)\right\} = f(x)\int u^3(\eta z)\, K\left(z u(\eta z)\right)dz, \quad (12)$$

$$Var\left\{\tilde{f}_M(x)\right\} = (nhm^{1/2})^{-1} f^{3/2}(x)\int u^4(\eta z)\frac{dz}{dy}K^2(y)dz - n^{-1}\left\{E\left(\tilde{f}_2(x)\right)\right\}^2. \quad (13)$$

One way of taking the integrals on the right- hand side in Equation (12) and Equation (13) is to change variable from z to $y = z\,u(\eta z)$, which transformation is invertible within a region $|y| \le \varepsilon\eta^{-1}$ for some $\varepsilon > 0$. Then we can rewrite Equation (12) as

$$E\left\{\tilde{f}_M(x)\right\} = f(x)\int u^3(\eta z)\frac{dz}{dy}K(y)dy,$$

and expanding $\varphi(\eta y) = u^3(\eta z)\dfrac{dz}{dy}$ as a series of powers of ηy and after the complex calculations, as $n \to \infty$, $h \to 0$ such that $nh \to \infty$, the expected value of the proposed adaptive kernel density estimator is obtained as

$$E\{\tilde{f}_M(x)\} = f(x) + \frac{1}{24}\frac{h^4 m^2}{f(x)} A(x)\,\delta_K + o(h^4),$$

where $A(x) = \frac{d^4}{dx^4}\left[\frac{1}{f(x)}\right]$, $\delta_K = \int x^4 K(x)dx$.

The bias of $\tilde{f}_2(x)$ is obtained by,

$$Bias\{\tilde{f}_M(x)\} = E(\tilde{f}_M(x)) - f(x) = \frac{1}{24}\frac{h^4 m^2}{f(x)} A(x)\,\delta_K + o(h^4).$$

In the variance in Equation (13), taking $H(\eta y) = u^4(\eta z)\frac{dz}{dy}$ and expanding it series, we obtain as

$$Var\{\tilde{f}_M(x)\} = \frac{1}{nhm^{1/2}} f^{3/2}(x) R(K) + o(nh)^{-1}.$$

The mean squared error will be obtained as follows:

$$MSE\{\tilde{f}_M(x)\} \cong \frac{1}{576}\delta_K^2 \frac{h^8 m^4}{f(x)^2} A^2(x) + (nhm^{1/2})^{-1} R(K) f^{3/2}(x) \cdot$$

We have investigated the rate of convergence of the modified estimator. $MSE\{\tilde{f}_M(x)\}$ is obtained as follows:

$$\inf MSE\{\tilde{f}_M(x)\} \cong \left[\begin{array}{l}\frac{45}{576}(\delta_K^2)^{1/9}(A^2(x))^{1/9} R(K)^{8/9} f^{10/9}(x) + \\ \frac{1}{72^{1/9}}(\delta_K^2)^{1/9}(A^2(x))^{1/9} R(K)^{8/9} f^{13/18}(x)\end{array}\right] n^{-8/9}$$

.

As it can be seen from the above equality, the rate of convergence is $O(n^{-8/9})$ and marginally improved from $O(n^{-4/5})$ *to* $O(n^{-8/9})$.

In next section, a simulation study is performed for the finite sample performance of $\tilde{f}_M(x)$. In addition to $\tilde{f}_M(x)$, we also include an alternative modified kernel estimator which is based on using range of $\hat{f}(x_i)$ by motivating from Aljuhani and Al turk [26]. Let $\tilde{f}_R(x)$ denote this alternative adaptive kernel density estimator.

3. SIMULATION STUDY

A simulation study is conducted to compare the performances of the fixed kernel estimator, the adaptive kernel estimator, and the modified adaptive kernel estimators. For the simulation, 1000 samples of sizes n=25, 100, 250, 500 are generated from some normal mixture type density distributions by following Marron and Wand [27]. The graphics of the normal mixture densities used in the simulation study are given in Fig. 1. For each generated sample, a fixed bandwidth is determined by using the least squares cross-validation method with Gaussian kernel function. Thus the fixed kernel estimations, the adaptive kernel estimations, and the modified adaptive kernel estimations are computed. Then the root mean squared errors (RMSE) are calculated for the finite sample performances for each estimator over 1000 samples. All computations are made by writing R (version 3.2.5) codes. The results are given in Tables 1 and 2.

Table 1. RMSE values for sample sizes of 25 and 100

Distribution	n=25				n=100			
	$\hat{f}$	$\tilde{f}_G$	$\tilde{f}_M$	$\tilde{f}_R$	$\hat{f}$	$\tilde{f}_G$	$\tilde{f}_M$	$\tilde{f}_R$
Standard Normal	0.05837	0.06199	0.06059	0.05842	0.03403	0.03688	0.03596	0.03364
Skewed Unimodal	0.07045	0.07220	0.07057	0.06779	0.04117	0.04283	0.04158	0.03891
Strongly Skewed	0.14016	0.14220	0.13930	0.13424	0.09022	0.08961	0.08725	0.08421
Kurtotic Unimodal	0.15021	0.14645	0.14258	0.13746	0.08899	0.08632	0.08129	0.07570
Outlier	0.17596	0.17124	0.16883	0.16690	0.10206	0.10153	0.09797	0.09399
Bimodal	0.06295	0.07016	0.06940	0.06734	0.03866	0.04188	0.04160	0.04026
Separated Bimodal	0.07866	0.08203	0.08127	0.07860	0.04624	0.04805	0.04734	0.04555
Skewed Bimodal	0.06988	0.07693	0.07595	0.07355	0.04345	0.04711	0.04677	0.04546
Trimodal	0.06625	0.07339	0.07267	0.07044	0.04153	0.04487	0.04461	0.04326

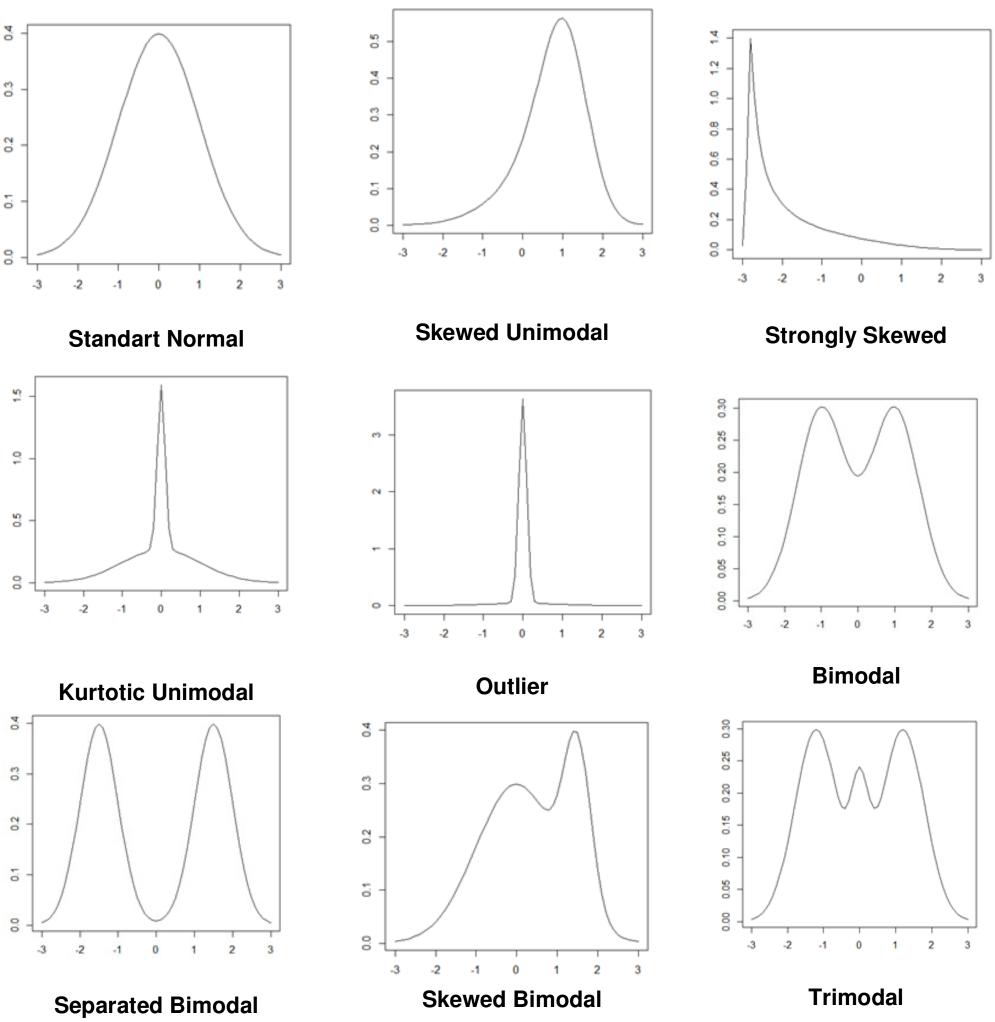

Fig. 1. Graphs of the normal mixture densities used for the simulation study

Table 2. RMSE values for sample Size of 250 and 500

Distribution	n=250				n=500			
	$\hat{f}$	$\tilde{f}_G$	$\tilde{f}_M$	$\tilde{f}_R$	$\hat{f}$	$\tilde{f}_G$	$\tilde{f}_M$	$\tilde{f}_R$
Standard Normal	0.02406	0.02608	0.02541	0.02377	0.01853	0.02028	0.01978	0.01863
Skewed Unimodal	0.02943	0.03006	0.02915	0.02738	0.02264	0.02357	0.02283	0.02141
Strongly Skewed	0.06552	0.06504	0.06242	0.05975	0.04996	0.04927	0.04670	0.04404
Kurtotic Unimodal	0.06302	0.06239	0.05761	0.05148	0.04788	0.04897	0.04471	0.03861
Outlier	0.07072	0.07216	0.06891	0.06622	0.05605	0.05858	0.05502	0.05244
Bimodal	0.02788	0.02940	0.02917	0.02826	0.02083	0.02184	0.02162	0.02088
Separated Bimodal	0.03268	0.03397	0.03341	0.03206	0.02465	0.02598	0.02552	0.02445
Skewed Bimodal	0.03152	0.03319	0.03289	0.03188	0.02446	0.02552	0.02522	0.02428
Trimodal	0.02999	0.03180	0.03160	0.03073	0.02341	0.02466	0.02447	0.02372

From Tables 1 and 2, it is seen that the adaptive kernel density estimators have good performances for long-tailed distributions (strongly skewed, kurtotic unimodal, outlier). For all cases, $\tilde{f}_M(x)$ has better performance than

$\tilde{f}_G$. The performance of $\tilde{f}_R(x)$ is the most attractive result in our study. In all cases, $\tilde{f}_R(x)$ has the best performance among the considered adaptive kernel density estimators.

4. CONCLUSION

In this study, we investigated a modified adaptive kernel density estimator for estimating probability density function. This estimator is based on the using arithmetic mean of pilot kernel density estimations in the Silverman's algorithm. Alternatively, we considered the performance of the third adaptive kernel density estimator based on the using range of pilot kernel density estimations. The simulation results show that the modified adaptive kernel estimators have better performance than the classical adaptive kernel density estimator. Specificially, the adaptive estimator based on using the range has very attractive performance for estimating probability density.

COMPETING INTERESTS

Authors have declared that no competing interests exist.

REFERENCES

1. Hardle W. Smoothing techniques with implementation in S. New York: Springer-Verlag; 1991.
2. Parzen E. On estimation of a probability density and mode. The Annals of Mathematical Statistics. 1962;33:1065-1076.
3. Rudemo M. Empirical choice of histograms and kernel density estimators. Scandinavian Journal of Statistics. 1982;9:65-78.
4. Epanechnikov V. Nonparametric estimates of a multivariate probability density. Theory of Probability and Its Applications. 1969;14:153-158.
5. Duin RPW. On the choice of smoothing parameters of Parzen estimators of Probability density function. Transactions on Computers. 1976;C-25(11):1175–1179.
6. Loader CR. Bandwidth selection: Classical or plug-in. Annals of Statistics. 1999;27: 415-438.
7. Silverman BW. Density estimation for statistics and data analysis. London: Chapman and Hall; 1986.
8. Boneva LI, Kendall DG, Stefanov I. Spline transformations: Three new diagnostics aids for the statistical data-analyst (with discussion). Journal of The Royal Statistical Society B. 1971;33:1-70.
9. Rosenblatt M. Remarks on some nonparametric estimates of a density function. Annals of Mathematical Statistics. 1656;27:832-837.
10. Park BU, Marron JS. Comparison of data-driven bandwidth selectors. Journal of The American Statistical Association. 1990;85: 66-72.
11. Cao R, Cuevas A, Manteiga WG. A comparative study of several smoothing methods in density estimation. Computational Statistics and Data Analysis. 1994;17:153-176.
12. Scott DW, Terrell GR. Biased and unbiased cross-validation in density estimation. Journal of The American Statistical Association. 1987;82(400):1131-1146.
13. Sheather SJ, Jones MC. A reliable data-based bandwidth selection method for kernel density estimation. Journal of The Royal Statistical Society B. 1991;53:683-690.
14. Horova I, Zelinka J. Contribution to the bandwidth choice for kernel density estimates. Computational Statistics. 2007; 22:31-47.
15. Slaoui Y. Bandwidth selection for recursive kernel density estimators defined by stochastic approximation method. Journal of Probability and Statistics. 2014;1-11. DOI: 10.1155/2014/739640
16. Karunamuni RJ, Sriram TN, Wu J. Rates of convergence of an adaptive kernel density estimator for finite mixture models. Statistics and Probability Letters. 2006;76: 221-230.
17. Breiman L, Meisel W, Purcell E. Variable kernel estimates of multivariate densities. Technometrics. 1977;19:135-144.
18. Abramson IS. On bandwidth variation in kernel estimates-a square root law. Annals of Statistics. 1982;10(4):1217-1223.
19. Müller HG, Wang JL, Capra WB. From lifetables to hazard rates: The transformation approach. Biometrika. 1994;84:881-892.
20. Terrell GR, Scott DW. Variable kernel density estimation. Annals of Statistics. 1992;20(3):1236-1265.
21. Hall P, Marron JS. Variable window width kernel estimates of probability densities.

Probability Theory and Related Fields. 1988;80:37-49.
22. Sain SR, Scott DW. On locally adaptive density estimation. Journal of The American Statistical Association. 1996;91: 1525-1534.
23. Jones MC. Variable kernel density estimation and variable kernel density estimation. Australian Journal of Statistics. 1990;32(3):361-371.
24. Sain SR. Adaptive kernel density estimation. Ph D. Thesis. Rice University; 1996.
25. Karunamuni RJ, Alberts T. A locally adaptive transformation method of boundary correction in kernel density estimation. Journal of Statistical Planning and Inference. 2006;136:2936-2960.
26. Aljuhani KH, Al turk LI. Modification of the adaptive Nadaraya-Watson kernel regression estimator. Scientific Research and Essays. 2014;9(22):966-971.
27. Marron JS, Wand MP. Exact mean integrated squared error. Annals of Statistics. 1992;20(2):712-736.

Peer-review history:
The peer review history for this paper can be accessed here:
http://sciencedomain.org/review-history/15543