International Journal of Computational and Theoretical Statistics
ISSN (2210-1519)
Int. J. Comp. Theo. Stat. 6, No. 1 (May-2019)
http://dx.doi.org/10.12785/ijcts/060110

On Asymptotic Mean Integrated Squared Error's Reduction Techniques in Kernel Density Estimation

Siloko, I. U.[1], Siloko, E. A.[2], Ikpotokin, O.[3], Ishiekwene, C. C.[4] and Afere, B.A.[5]

[1]Department of Mathematics and Computer Science, Edo University Iyamho, Nigeria.
[2,4]Department of Mathematics, University of Benin, Benin City, Nigeria.
[3]Department of Mathematics and Statistics, Ambrose Alli University, Ekpoma, Nigeria.
[5]Department of Mathematics and Statistics, Federal Polytechnic Idah, Nigeria.

Received July 24, 2018, Revised December 11, 2018, Accepted January, 14 2019, Published May 1, 2019

Abstract: The techniques of asymptotic mean integrated squared error's reduction in kernel density estimation is the focus of this paper. The asymptotic mean integrated squared error (AMISE) is an optimality criterion function that measures the performance of a kernel density estimator. This criterion function is made up of two components, and the contributions of both components to the AMISE are mainly regulated by the smoothing parameter. Kernel density estimation are of vitally importance in statistical data analysis especially for exploratory and visualization purposes. In performance evaluation, a method is better when it produces a smaller value of the AMISE; hence effort is being made to develop techniques that reduce the AMISE while ensuring that in practical implementation using real data, the statistical properties of the given observations are retained. We consider the kernel density derivative and kernel boosting as the AMISE reduction techniques. In kernel boosting, we introduce the optimal smoothing parameter selector for each boosting steps as the number of iteration increases. The presented results show that the AMISE decreases with higher kernel derivatives and also as the number of boosting steps increases.

Key words: Kernel, Derivatives, Boosting, Bandwidths, AMISE.

1. INTRODUCTION

Density estimation is the construction of a probability density estimates from a given sample using the sample values and few assumptions about the density estimator. Kernel density estimators are widely used nonparametric estimation techniques in statistics due to their simple forms and smoothness. Kernel estimation is an important statistical data analytic tool whose ideas can be extended to other fields of studies that requires data analysis.

Kernel density derivatives are of wider applications in statistics and other related fields of studies. The first and second derivatives of any density function are fundamental in estimation because some statistical properties of the distribution like local extrema and point of inflexion can be identified [1]. The derivatives of a probability density function are also applicable in clustering analysis [2], time series analysis [3], estimation of the optimal smoothing parameter in kernel density estimation and the location of modes and bumps of a density estimate [4].

Boosting in density estimation was introduced by Freund and Shapire [5] and applied basically in regression and classification problems. The idea was extended to kernel density estimation by Marzio and Taylor [6] as bias reduction techniques. The practical successful applications of boosting in many fields of statistics has accounted for its popularity while effort is been made to develop the statistical theory which explains the principles of its mode of operations. The boosting method involves the systematic reweighting of data base on a kernel function that depends on the smoothing

E-mail address: siloko.israel@edouniversity.edu.ng, suzuazor@yahoo.com, myakpeski@yahoo.com, ikpotokinosayomore@yahoo.co.uk, cycigar@yahoo.co.uk, baafere@gmail.com.

parameter. In kernel density estimation, boosting is the process of updating the weights of the estimator so that the product of the aggregate integrates to unity.

Kernel density estimation is mainly confronted with the problem of smoothing parameter choices. In univariate kernel estimation, the problem of smoothing parameter selection is with less complexity unlike the multidimensional case where there are different forms of smoothing parameterizations [7]. The choice of smoothing parameter is also very important in kernel density derivatives particularly as the order of the derivative to be estimated increases. The determination of the smoothing parameter is mainly data related for a better objective use of kernel estimator. In kernel boosting, smoothing parameter plays a vital role in its implementation; hence the smoothing parameter is regarded and interpreted as a resolution factor when viewing observations and giving better interpretation of the structures of the observations.

2. KERNEL DENSITY ESTIMATOR AND ITS DERIVATIVES

The univariate kernel estimator is a nonparametric technique in density estimation. In density estimation, the univariate kernel estimator provides an excellent platform for displaying features in given observations due to easy implementation unlike other complex estimators. The univariate kernel estimator has its compact form as

$$\hat{f}(\mathrm{x}) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{\mathrm{x}-\mathrm{X}_i}{h}\right), \tag{1}$$

where K is the kernel function, $h > 0$ is the smoothing parameter, X_i are the set of observations and n is the sample size. The kernel function is a non-negative function that satisfies the following conditions

$$\begin{cases} \int K(\mathrm{x})d\mathrm{x} = 1, \\ \int \mathrm{x}K(\mathrm{x})d\mathrm{x} = 0 \quad \text{and} \\ \int \mathrm{x}^2K(\mathrm{x})d\mathrm{x} = k_2(K) > 0\,. \end{cases} \tag{2}$$

The first condition in (2) implies that any weighting function must integrate to unity, hence most kernel functions are probability density functions; the second condition simply states that the average of the kernel is zero, while the third condition means that the variance of the kernel is not zero [4].

The derivative of the univariate kernel density function is obtained by taking the derivative of the kernel density estimator in (1). Assuming the kernel K is sufficiently differentiable r times, the rth density derivative of (1) is given by

$$\hat{f}^{(r)}(\mathrm{x}) = \frac{1}{nh^{r+1}}\sum_{i=1}^{n} K^{(r)}\left(\frac{\mathrm{x}-\mathrm{X}_i}{h}\right), \tag{3}$$

where $K^{(r)}$ is the rth derivative of the kernel function and the kernel K is usually a symmetric probability density function [8]. In kernel density derivative estimation, the smoothing parameters are expected to be larger than the estimation without derivative because the derivative of any function tends to be noisier than when the function is not differentiated. Hence kernel density derivatives are associated with larger smoothing parameter for their estimation.

We use the Gaussian kernel function that has zero mean and a unit variance because it produces smooth density estimates and simplifies the required mathematical computations. Again, the Gaussian kernel possesses derivatives of all orders and that has supported its wide spread uses in kernel density estimation and kernel density derivative estimation. The Gaussian rth density derivative is usually estimated from the Gaussian kernel and is denoted by $K^{(r)}(\mathrm{x}) = (-1)^r H_r(\mathrm{x})K(\mathrm{x})$, where $H_r(\mathrm{x})$ is the rth Hermite polynomial. The first five values of the Hermite polynomials are $H_0(\mathrm{x}) = 1$, $H_1(\mathrm{x}) = \mathrm{x}$, $H_2(\mathrm{x}) = \mathrm{x}^2 - 1$, $H_3(\mathrm{x}) = \mathrm{x}^3 - 3\mathrm{x}$ and $H_4(\mathrm{x}) = \mathrm{x}^4 - 6\mathrm{x}^2 + 3$. The Gaussian kernel density derivative estimate from (3) is of the form

$$\hat{f}^{(r)}(\mathrm{x}) = \frac{(-1)^r}{\sqrt{2\pi}nh^{r+1}} \sum_{i=1}^{n} H_r\left(\frac{\mathrm{x}-\mathrm{X}_i}{h}\right) exp^{-\frac{1}{2}\left(\frac{\mathrm{x}-\mathrm{X}_i}{h}\right)^2}. \quad (4)$$

Equation (4) above can be used to estimate the rth derivative of the kernel function but it should be noted that when $r = 0$, it will result in the usual kernel density estimator.

3. THE ASYMPTOTIC MEAN INTEGRATED SQUARED ERROR APPROXIMATIONS

The estimates of $\hat{f}(\mathrm{x})$ and $\hat{f}^{(r)}(\mathrm{x})$ in (1) and (3) are measured mainly by the asymptotic mean integrated squared error (AMISE). An asymptotic approximation of (1) using Taylor's series expansion will yield the integrated variance and the integrated squared bias given by

$$AMISE\left(\hat{f}(\mathrm{x})\right) = \frac{R(K)}{nh} + \frac{1}{4}\mu_2(\mathrm{K})^2\, h^4 R(f''), \quad (5)$$

where $R(K)$ is the roughness of the kernel, $\mu_2(\mathrm{K})^2$ is the second moment of the kernel and $R(f'') = \int f''(\mathrm{x})^2 d\mathrm{x}$ is the roughness of the unknown probability density function [4]. The value with the minimum of the AMISE is the solution to the differential equation

$$\frac{\partial}{\partial h} AMISE(h) = \frac{-R(K)}{nh^2} + \mu_2(\mathrm{K})^2 h^3 R(f'') = 0.$$

Therefore, the smoothing parameter that minimizes the AMISE of the kernel density estimator in (1) is of the form

$$h_{\mathrm{AMISE}} = \left[\frac{R(K)}{\mu_2(\mathrm{K})^2 R(f'')}\right]^{1/5} \times n^{-1/5}. \quad (6)$$

In terms of dimension, (6) can be written as

$$h_{\mathrm{AMISE}} = \left[\frac{R(K)}{\mu_2(\mathrm{K})^2 R(f'')}\right]^{1/(4+d)} \times n^{-1/(4+d)}, \quad (7)$$

where d is the dimension of the kernel function.

The AMISE of the rth derivative of the kernel function provided the kernel K can be sufficiently differentiated is of the form

$$AMISE\left(\hat{f}^{(r)}(\mathrm{x})\right) = \frac{R\left(K^{(r)}\right)}{nh^{2r+1}} + \frac{1}{4}h^4\mu_2(\mathrm{K})^2\, R\left(f^{(r+2)}\right), \quad (8)$$

where $R\left(K^{(r)}\right)$ is the roughness of the rth derivative of the kernel, $\mu_2(\mathrm{K})^2$ is the second moment of the kernel and $R\left(f^{(r+2)}\right)$ is the roughness of the rth unknown probability density function [9]. The order of the bias term of the rth derivative of the AMISE is the same as $O(h^4)$ but any new derivative order will introduce two additional powers to the smoothing parameter h of the variance term. The rth roughness of the Gaussian kernel function denoted by $R\left(K_{\emptyset}^{(r)}\right)$ can be calculated from the relation

$$R\left(K_{\emptyset}^{(r)}\right) = \frac{(2r-1)!!}{2^{r+1}\sqrt{\pi}}. \quad (9)$$

The smoothing parameter that minimized the rth AMISE in (8) is given by

$$h_{AMISE}^{r} \approx \left[\frac{(2r+1)R\left(K^{(r)}\right)}{\mu_2(K)^2\, R(f^{(r+2)})}\right]^{\left(\frac{1}{2r+5}\right)} \times n^{-\left(\frac{1}{2r+5}\right)}. \tag{10}$$

The smoothing parameter that minimizes the AMISE for the first and second derivatives are of orders $O(n^{-1/7})$ and $O(n^{-1/9})$ respectively [4].

4. BOOSTING IN KERNEL DENSITY ESTIMATION

Boosting in kernel density estimation is a multiplicative aggregation model that was introduced into kernel estimation by Marzio and Taylor [10] and is considered as a bias reduction method. If K is the kernel function and $h > 0$ is the smoothing parameter, then [11]

$$\hat{f}_m(\mathrm{x}) = \int \frac{1}{h} W_m(t) K\left(\frac{\mathrm{x}-t}{h}\right) f(t)dt, \tag{11}$$

where $W_1(t)$ is taken to be 1. The standard normal kernel will be used with the transformation $u = (\mathrm{x}-t)/h$, $t = \mathrm{x} - hu$ and $\left|\frac{du}{dt}\right| = \frac{1}{h}$ with $dt = hdu$. Using Taylor's series expansion on (11) with the transformation, $t = \mathrm{x} - hu$ and $\left|\frac{du}{dt}\right| = \frac{1}{h}$ with $dt = hdu$ we have

$$\hat{f}_1(\mathrm{x}) = f(\mathrm{x}) + \frac{h^2 f''(\mathrm{x})}{2} + \; O(h^2). \tag{12}$$

The bias of $\hat{f}_1(\mathrm{x})$ is of order $O(h^2)$. Again using the change-of-variables $u = (t-\mathrm{x})/h$, $t = \mathrm{x} + hu$ and $\frac{du}{dt} = \frac{1}{h}$ with $W_2 = \left(\hat{f}_1(\mathrm{x})\right)^{-1}$, we have the boosted estimate of $\hat{f}(\mathrm{x})$ at the second step which is of the form

$$\begin{aligned}\hat{f}_2(\mathrm{x}) &= \int K(u)\left\{f(\mathrm{x}+hu) + h^2\frac{f''(\mathrm{x}+hu)}{2} + O(h^2)\right\}^{-1} f(\mathrm{x}+hu)du \\ &= 1 + \frac{h^2 f''(\mathrm{x})}{2f(\mathrm{x})} + \; O(h^2).\end{aligned} \tag{13}$$

The overall estimate of $\hat{f}(\mathrm{x})$ at the second step is of the form

$$\begin{aligned}\hat{f}_1(\mathrm{x}) \times \hat{f}_2(\mathrm{x}) &= f(\mathrm{x})\left\{1 + \frac{h^2 f''(\mathrm{x})}{2f(\mathrm{x})} + \; O(h^2)\right\}\left\{1 + \frac{h^2 f''(\mathrm{x})}{2f(\mathrm{x})} + \; O(h^2)\right\} \\ &= f(\mathrm{x}) + O(h^4).\end{aligned} \tag{14}$$

Equation (14) is of order four, $O(h^4)$ and there is a bias reduction from order two, $O(h^2)$ to order four showing the bias reduction in kernel estimation via boosting. The multiplicative aggregation also known as boosting algorithm is a systematized algorithm where each step of m is computed by

$$\hat{f}_m(\mathrm{x}) = \frac{1}{h}\sum_{i=1}^{n} \mathrm{W}_m(i) K\left(\frac{\mathrm{x}-\mathrm{X}_i}{h}\right), \tag{15}$$

where K is the kernel function, h is the smoothing parameter and $\mathrm{W}_m(i)$ is the weight of observation i at step m. The weight of each observation is then updated using a log-odds ratio as [10]

$$\mathrm{W}_{m+1}(i) = \mathrm{W}_m(i) + \log\left(\frac{\hat{f}_m(\mathrm{x})}{\hat{f}_m^{(-i)}(\mathrm{x})}\right), \tag{16}$$

where $\hat{f}_m^{(-i)}(\mathrm{x}_i)$ is the leave-one-out estimator. The leave-one-out estimator is given by

$$\hat{f}_m^{(-i)}(\mathrm{x}) = \frac{1}{(n-1)h}\sum_{i=1}^{n} K\left(\frac{\mathrm{x}-\mathrm{X}_i}{h}\right). \tag{17}$$

Boosting in kernel density estimation involves the weights of the observations being updated at each step and with the final estimator being a product of all the density estimates that will integrate to unity. The kernel boosting algorithm is given below.

STEP1. Given that $i = 1, 2, \ldots, n$, initialise the weights of the observations

$$W_1(i) = 1/n$$

STEP2. Select h which is the smoothing parameter.

STEP3. For $m = 1, \ldots, M$.

(i) Obtain a weighted kernel estimate

$$\hat{f}_m(\mathrm{x}) = \frac{1}{h}\sum_{i=1}^{n} W_m(i) K\left(\frac{\mathrm{x}-\mathrm{X}_i}{h}\right)$$

(ii) Update the weights according to

$$\mathrm{W}_{m+1}(i) = \mathrm{W}_m(i) + \log\left(\frac{\hat{f}_m(\mathrm{x})}{\hat{f}_m^{(-i)}(\mathrm{x})}\right)$$

STEP4. Provide as output

$$C\prod_{m=1}^{M}\hat{f}_m(\mathrm{x}),$$

where C is normalization constant such that $\hat{f}_m(\mathrm{x})$ integrates to unity.

5. THE PROPOSED KERNEL BOOSTING BANDWIDTH SELECTOR

Kernel density estimation as a nonparametric estimation technique is mainly confronted with bandwidth selection problem. However, smoothing parameter selectors for kernel density estimation and kernel density derivatives have been proposed by many authors even though there is no universally acceptable rule in bandwidth selection but there is no bandwidth selector for kernel boosting. Smoothing parameter selection in kernel boosting since its introduction in kernel density estimation has been subjective; that is the selection is at the user's discretion. The subjective idea may not be objective; that is, may not be efficient because it does not consider the statistical properties of the distribution and the kernel function; hence cannot be applied in all circumstances. In kernel boosting, the smoothing parameter for its implementation must be larger than the classical second order kernel smoothing parameter because of the multiplication of the estimates and the principle of over smoothing.

Generally, kernel boosting in density estimation is a bias reducing approach that requires larger smoothing parameter in its implementation [12]. Boosting in kernel density estimation will result in reduction in the bias component of the AMISE and that translates to a reduction in the AMISE but with its major problem being the smoothing parameter require for each of the boosting steps. Kernel boosting in density estimation is comparable with the gradient methods in optimization theory where each of the iteration known as the boosting steps requires a smoothing parameter for its computation.

In solving the problem of smoothing parameter selection in kernel boosting with respect to the number of boosting steps, we introduce a multiplier known as the bandwidth multiplier which is denoted by β. The bandwidth multiplier regulates the selection of the smoothing parameter require for each boosting step since the boosting idea involves the multiplication of the different estimates to produce the overall estimate. Recall the general smoothing parameter that minimizes the AMISE in (5) given as

$$h_{\text{AMISE}} = \left[\frac{R(K)}{\mu_2(\text{K})^2 R(f^{''})}\right]^{1/5} \times n^{-1/5.}$$

Since kernel boosting is a higher order bias reduction method, the propose smoothing parameter selector require for boosting in density estimation is of the form

$$h_m = \beta^{1/2} \times h_{\text{AMISE}}\,, \tag{18}$$

where

$$\begin{cases} m = 2, 3, \ldots, M \\ \\ \beta = 2m\,. \end{cases}$$

In (18) above, h_{AMISE} is the smoothing parameter obtains from (6), m represents the number of boosting steps and $2m$ denotes the order of the kernel. The bandwidth multiplier is for selection of smoothing parameter for boosting in kernel density estimation. Boosting in kernel estimation is a higher order bias reduction method, hence when $m = 1$ is excluded from (18) because it produces the classical second order kernel.

6. RESULTS AND DISCUSSION

This section is about the implementation of kernel density derivative and kernel boosting using real data example. Kernel density derivatives and kernel boosting are both AMISE reduction methods that require larger smoothing parameter. While kernel density derivatives select its smoothing parameter for each derivative order using (10), kernel boosting will make use of (18) for the smoothing parameter require for each boosting step. We compute the first and second derivative orders and also obtain the first and second boosting steps only using the Gaussian kernel function. We obtain the results for the kernel derivatives and kernel boosting using Mathematica version 9 platforms. The procedure for obtaining the estimates of the kernel derivatives and kernel boosting for the Gaussian kernel is in the appendix. The result in Table 6.1 and Table 6.2 respectively shows the smoothing parameter and the value of the asymptotic mean integrated squared error for each derivative order and each boosting step. The kernel estimates with the various derivatives order are in Figure 6.1 while Figure 6.2 is the kernel estimates of the various boosting steps.

The data set examine is the eruption lengths of 107 eruptions of Old Faithful Geyser [13]. The usual kernel estimates (i.e. when $r = 0$) of this data show that the data set is bimodal and this provides an evidence in favour of eruption times exhibiting a bimodal distribution. However, the kernel estimate of the first derivative presents the data to be trimodal but at the second derivative order, the kernel estimate is oversmoothed and this is due to larger bandwidth as the order of the derivative increases. Generally, the first and second derivatives of a kernel function are for the estimation of modes and bumps of a distribution in density estimation.

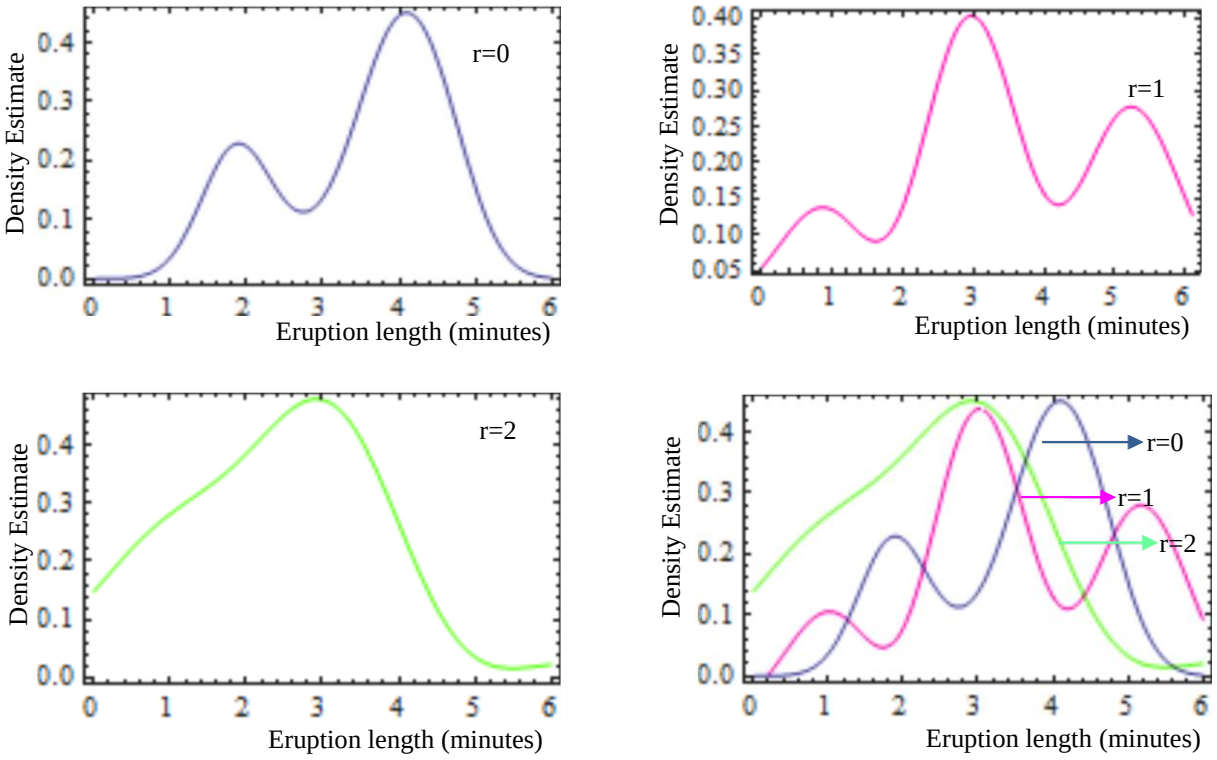

Figure 6.1. Kernel Estimates of the Old Faithful Data and Derivatives

Table 6.1. Bandwidths and AMISE of Kernel Derivative

Derivative Order	Bandwidths	AMISE
0	0.437301	0.00755559
1	0.759298	0.00526965
2	1.091920	0.00286615

In Table 6.1, as the derivative order increases, the smoothing parameters also increase while the AMISE decreases as well and this is a characteristic of kernel density derivative. Higher derivative order of kernel density estimation is an AMISE reduction technique but often times it may smooth out some significant features of the distribution if the distribution is not unimodality.

In kernel boosting, the density estimates are often times oversmoothed especially for multimodal distribution. Generally, kernel boosting is always associated with oversmoothing due to larger smoothing parameter requirement for its implementation and the multiplication of the estimates involve. A method is better than the other when it produces a smaller value of the asymptotic mean integrated squared error [14, 15]. The bandwidths and the AMISE value of the boosted kernel are presented in Table 6.2 and the results show that as the number of boosting steps increases, the asymptotic mean integrated squared error (AMISE) reduces.

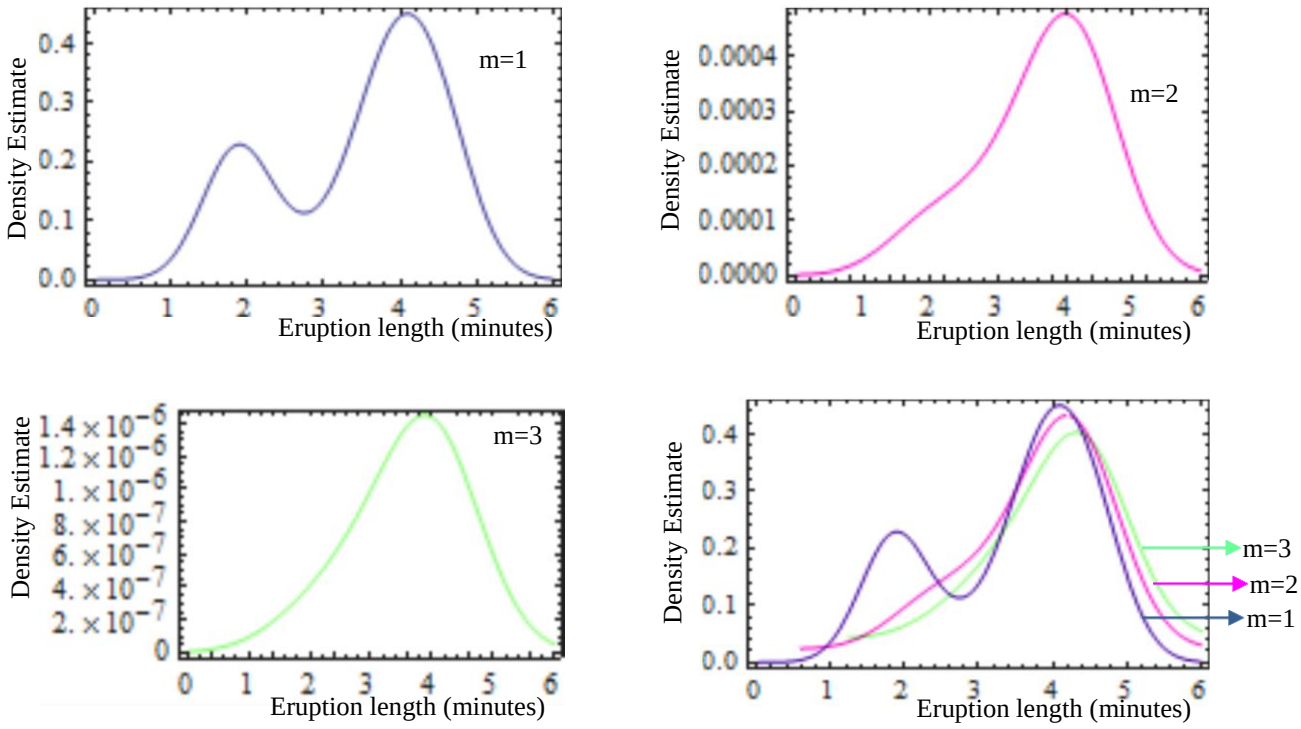

Figure 6.2. Kernel Estimates of the Old Faithful Data and Boosted Estimates

Table 6.2. Bandwidths and AMISE of Kernel Boosting

Boosting Step	Bandwidths	AMISE
1	0.437301	0.00755559
2	0.874602	0.00366535
3	1.071164	0.00255186

Boosting in kernel density estimation is an AMISE reduction technique but the inherent features in the data set at times might disappear due to the multiplication of the estimates and the use of large smoothing parameter.

7. CONCLUSION

The study is on the techniques of reducing the asymptotic mean integrated squared error using the kernel density derivative and kernel boosting approaches. Both methods depend on the smoothing parameter which must be larger than the classical second order smoothing parameter. Kernel density derivatives and kernel boosting may smooth away some desirable features of a data set such as multimodality but retained the characteristics of reducing the AMISE. While kernel boosting and kernel density derivative tends to produce smaller value of the AMISE, the proposed kernel boosting bandwidth selector produce AMISE values that are smaller than the AMISE values of the kernel density derivative method in kernel density estimation.

ACKNOWLEDGEMENT

The authors appreciate the anonymous reviewer for painstakingly going through the manuscript and for his valuable comments.

REFERENCES

[1] K.. De Brabanter, J. De Brabanter, and B. De Moor, Nonparametric derivative estimation, NAIC, Gent, Belgium, 2011.

[2] D. Comaniciu and P. Meer, Mean Shift: A robust approach toward feature space analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence, 24(5), 603–619, 2002.

[3] V. Rondonotti, J. S. Marron, and C. Park, SiZer for time series: A new approach to the analysis of trends. Electronic Journal of Statistics, 1, 268–289, 2007.

[4] D. W. Scott, Multivariate density estimation. Theory, Practice and Visualisation. Wiley, New York, 1992.

[5] Y. Freund and R. Schapire, A decision-theoretic generalization of on-line learning and an application to boosting. Journal of Comp. System Sci. 55(1), 119–139, 1997.

[6] D. M. Marzio and C. C. Taylor, "Boosting kernel density estimates: A bias reduction technique?" Biometrika, 91, 226–233, 2004.

[7] I. U. Siloko, C. C. Ishiekwene, and F. O. Oyegue, New gradient methods for bandwidth selection in bivariate kernel density estimation. Mathematics and Statistics, 6(1), 1–8, 2018.

[8] V. C. Raykar, R. Duraiswami, and L. H. Zhao, Fast computation of kernel estimators. Journal of Computational and Graphical Statistics, 19 (1), 205–220, 2015.

[9] A. S. Guidoum, Kernel estimator and bandwidth selection for density and its derivatives. Department of Probabilities and Statistics, University of Science and Technology, Houari Boumediene, Algeria, 2015.

[10] D. M. Marzio and C. C. Taylor , "On boosting kernel density methods for multivariate data: Density estimation and classification". Statistical Methods and Applications, 14, 163–178, 2005.

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

[12] I. U. Siloko and C. C. Ishiekwene, Boosting and bagging in kernel density estimation. The Nigerian Journal of Science and Environment, 14(1), 32–37, 2016.

[13] B. W. Silverman, Density estimation for statistics and data analysis, Chapman and Hall, London, 1986.

[14] J. Jarnicka, Multivariate kernel density estimation with a parametric support, Opuscula Mathematica, 29(1), 41–45, 2009.

[15] I. U. Siloko, O. Ikpotokin, F. O. Oyegue, C. C. Ishiekwene, and B. A. E. Afere, A note on application of kernel derivatives in density estimation with the univariate case. Journal of Statistics and Management Systems, 22(3), 415–423, 2019.

APPENDIX

Kernel Derivatives

$$n = 107$$
$$h = 0.437301$$
$$X_i = \{Data\}$$
$$F = \left(\left(1/\left((2*\pi)^{1/2}*n*h\right)\right)*Exp^{-((((x-X_i)/h)^2)/2)}\right)$$
$$\hat{f}_0 = \text{Total}[F]$$
$$\text{Plot}\left[\hat{f}_0,\ \{x,\ 0,\ 6\},\ \text{Frame} \to \text{True},\ \text{FrameStyle} \to \text{GrayLevel}[0.125]\right]$$

First Derivative

$$h_1 = 0.759298$$
$$F_1 = -\left(\left((-1)/\left((2*\pi)^{1/2}*n*h_1^2\right)\right)*Exp^{-\left((((x-X_i)/h_1)^2)/2\right)((x-X_i)/h_1)^2}\right)$$
$$\hat{f}_1 = \text{Total}[F_1]$$
$$\text{Plot}\left[\hat{f}_1,\ \{x,\ 0,\ 6\},\ \text{Frame} \to \text{True}, \text{FrameStyle} \to \text{GrayLevel}[0.125]\right]$$

Second Derivative

$$h_2 = 1.09192$$
$$\mathrm{F}_2 = \left(\left(1/\left((2*\pi)^{1/2}*n*h_2^3\right)\right)*Exp^{-\left(\left(\left(\left(\mathrm{x}-X_i\right)/h_2\right)^2\right)/2\right)}*\left(\left(\left(\mathrm{x}-X_i\right)/h_2\right)-1\right)^2\right)$$
$$\hat{f}_2 = \mathrm{Total}[\mathrm{F}_2]$$
$$\mathrm{Plot}\left[\hat{f}_2,\quad \{\mathrm{x},\quad 0,\quad 6\},\quad \mathrm{Frame}\rightarrow \mathrm{True}, \mathrm{FrameStyle}\rightarrow \mathrm{GrayLevel}[0.125]\right]$$

Kernel Boosting

$$n = 107$$
$$\mathrm{w}_1 = 1/n$$
$$h = 0.437301$$
$$X_i = \{Data\}$$
$$F = \left(\left(1/\left((2*\pi)^{1/2}*n*h\right)\right)*Exp^{-\left(\left(\left(\mathrm{x}-X_i\right)/h\right)^2\right)/2\right)}\right)$$
$$\hat{f} = \mathrm{Total}[F]$$
$$\mathrm{Plot}\left[\hat{f},\quad \{\mathrm{x},\ 0,\ 6\},\quad \mathrm{Frame}\rightarrow \mathrm{True}, \mathrm{FrameStyle}\rightarrow \mathrm{GrayLevel}[0.125]\right]$$

To obtain the next boosting step, update the weights according to

$$\mathrm{w}_2(i) = \mathrm{w}_1(i) + \log\left(\frac{\hat{f}_m(\mathrm{x})}{\hat{f}_m^{(-i)}(\mathrm{x})}\right)$$

by using the next smoothing parameter value for m_2 which is the next boosting step.