Palestine Journal of Mathematics

Vol. 8(2)(2019) , 286–294

On the Reduction of Global Error of Multivariate Higher-Order Product Polynomial Kernels

Benson Ade Afere and Ekele Alih

Communicated by Jose Luis Lopez-Bonilla

MSC 2010 Classifications: 62G07

Keywords and phrases: Density estimation, Higher-order kernels, Multivariate product polynomial kernel, Generalized global error, Convergence rate.

Abstract. A higher-order kernel has the features of both negative and positive kernels. The advantage of this over the lower-order kernel is that it leads to faster rate of convergence. Thus, in this paper, we presented the reduction of global error of multivariate higher-order product polynomial kernels. The family of product polynomial multivariate higher-order kernels is constructed. A generalized scheme for determining the global error of any kernel in this family is proposed. A Monte Carlo experiment is performed using six different data sets and it was observed that our scheme is efficient even if the data set departs from the standard normal distribution; and thus have higher rate of convergence

1 Introduction

Density estimation methods are considered as the construction of an estimate of an underlying probability density function (*pdf*) based on an observed data. Let $\mathbf{X}_1, \mathbf{X}_2, ..., \mathbf{X}_n$ be independent, identically distributed random samples with bounded continuously density $f(\mathbf{x})$. The multivariate kernel density estimator is given by:

$$\hat{f}_h(\mathbf{X}) = \frac{1}{n|\mathbf{H}|}\sum_{i=1}^{n} K(\mathbf{H}(\mathbf{x}-\mathbf{X}_i)) \tag{1.1}$$

where $\mathbf{x} = (x_1, x_2, \cdots, x_d)^T$ and $\mathbf{X}_i = (X_{i1}, X_{i2}, \cdots, X_{id})^T,\ i = 1, \cdots, n$ $\mathbf{H}$ is a symmetric positive definite $d \times d$ nonsingular bandwidth matrix, K is a $d-$ variate kernel function which integrates to 1. We shall adopt the parameterization given by [1], instead of those suggested by [2] and [3]. Hence, the multivariate kernel density estimator in (1.1) is modified as:

$$\hat{f}_\lambda(\mathbf{X}) = \frac{1}{n\lambda^d}\sum_{i=1}^{n} K(\frac{\mathbf{x}-\mathbf{X}_i}{\lambda}) \tag{1.2}$$

where $\mathbf{x},\ \mathbf{X}_i$ and K are as defined in (1.1) and λ is the univariate bandwidth such that as $n \to \infty$, $\lambda \to 0$ and $n\lambda \to \infty$ [13].

Quite a good numbers of articles have been published in kernel density estimation that is centered on bias reduction and bandwidth selection (for instance, see [4], [6] and [10]). The main objective of this paper is to develop a higher-order generalized global error scheme that is data driven for any higher-order multivariate product polynomial kernel. Our idea is however based on the work of [5] and [7].

Theorem 1.1. - *Review of Silverman (1986) [12] and Orava (2011) [7]: Let f be the bounded density function. The kernel function $K(t)$ is assumed to be bounded with finite second moment. Furthermore, let x be a point with $f(x) > 0$ and f be continuously differentiable up to the second order in a neighbourhood of x. Then, the asymptotic variance and asymptotic bias of $\hat{f}_\lambda(x)$ estimate can be expressed respectively as*

$$Var(\hat{f}_\lambda(x)) \approx \frac{1}{n\lambda}\|K\|_2^2 f(x) \tag{1.3}$$

$$Bias(\hat{f}_h(x)) \approx \frac{1}{2}\lambda^2 K_2 f''(x) \tag{1.4}$$

where $\|K\|_2^2$ *and* K_2 *in (1.3) and (1.4) are respectively:*

$$\|K\|_2^2 = \int K^2(t)dt \tag{1.5}$$

and

$$K_2 = \int t^2 K(t)dt \tag{1.6}$$

Proof. The proof of this theorem is contained in [12]. □

The global error which is measured by asymptotic mean integrated square error (AMISE) as contained in [13] is given by:

$$\text{AMISE}(\hat{f}_\lambda(x)) = \int \text{Var}(\hat{f}_\lambda(x))dx + \int \text{Bias}^2(\hat{f}_\lambda(x))dx \tag{1.7}$$

Substituting (1.5) and (1.6) into (1.7) gives:

$$\text{AMISE}(\hat{f}_\lambda(x)) = \frac{1}{4}\lambda^4 (K_2)^2 \int f''(x)^2 dx + \frac{1}{n\lambda}\|K\|_2^2 \tag{1.8}$$

Theorem 1.2. *Let f be the bounded density function. The multivariate kernel function $K(\mathbf{t})$ is assumed to be bounded with finite second moment and satisfying*

$$\|K\|_2^2 = \int_{\mathbb{R}^d} K^2(\mathbf{t})d\mathbf{t} \tag{1.9}$$

and

$$K_2 = \int_{\mathbb{R}^d} (\mathbf{t}^T\mathbf{t})K(\mathbf{t})d\mathbf{t} \tag{1.10}$$

where $K_2 = \int_{\mathbb{R}^d} t_i^2 K(\mathbf{t})d\mathbf{t}$ is independent of i [12, 13]. In addition, let $\mathbf{x}$ be a vector with $f(\mathbf{x}) > 0$ and f be continuously differentiable up to the second order in a neighbourhood of $\mathbf{x}$. Then, the asymptotic mean integrated square error of $\hat{f}_\lambda(x)$ estimate can be expressed as:

$$AMISE\hat{f}_\lambda(x) = \frac{1}{4}\lambda^4 (K_2)^2 \int_{\mathbb{R}^d} (\nabla^2 f(\mathbf{x}))^2 d\mathbf{x} + \frac{\|K\|_2^2}{n\lambda^d} \tag{1.11}$$

Proof. Substituting (1.2) in (1.7) and using the $d-$ variate Taylor′s Theorem. Then apply conditions (1.9) and (1.10) in (1.7) yields (1.11). □

Theorems 1.1 and 1.2 above hold for univariate lower-order kernels and multivariate lower-order kernels respectively. The lower-order kernels are basically non-negative kernels which are symmetric about 0. However, [9] proposed the use of both negative and non-negative (higher-order) kernels. The advantage of this is that, it leads to faster rates of convergence over the lower-order kernels. In view of this, we shall relax the conditions in Theorem 1.2.

The remaining part of this paper shall be structured as follows: Section 2 is on the multivariate higher-order product polynomial kernels. In Section 3, the generalized higher-order global error scheme is developed. A Monte Carlo experiment is performed in Section 4 and the concluding remarks is given in Section 5.

2 Multivariate higher-order product polynomial kernels

The family of classical polynomial kernels as given in [13] is of the form:

$$K_1(t,p) = \{2^{2p+1}\text{B}(p+1,p+1)\}^{-1}(1-t^2)^p,\ |t| \le 1;\ p = 0,1,2,... \tag{2.1}$$

where B is the beta density function and p is the power of K_1. The specific results yield a Uniform kernel, Epanenchnikov kernel, Biweight kernel, and Triweight kernel when $p = 0$, $p = 1$, $p = 2$ and $p = 3$ respectively. However, as $p \to \infty$, (2.1) tends to normality and thus we have a Gaussian kernel.

To construct higher-order kernel, we adopt the method suggested by [5]. Suppose $K_{[\ell]}(t)$ denote an ℓ^{th}-order Gaussian kernel given by:

$$K_{[\ell]}(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} \tag{2.2}$$

Expanding (2.2) by using Taylor's Theorem about 0 up to order two, we have

$$\begin{aligned}
\hat{K}_{[\ell]}(t) &= K_{[\ell]}(0) - \frac{1}{2}t^2 K''_{[\ell]}(0), \;\; K'_{[\ell]}(0) = 0 \\
&= \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} - \frac{1}{2}(t^2 - 1)\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} \\
&= \frac{3}{2}\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} + \frac{1}{2}t(-t)\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} \\
&= \frac{3}{2}K_{[\ell]}(t) + \frac{1}{2}t(-tK_{[\ell]}(t))
\end{aligned}$$

But $K'_{[\ell]}(t) = -tK_{[\ell]}(t)$ and let $\hat{K}_{[\ell]}(t) = K_{[\ell+2]}(t)$. Hence we have,

$$K_{[\ell+2]}(t) = \frac{3}{2}K_{[\ell]}(t) + \frac{1}{2}tK'_{[\ell]}(t) \tag{2.3}$$

Thus, (2.3) can be used to generate any higher-order kernels. Using (2.2) and its derivative in (2.3) results in the family of higher-order polynomial kernel given as:

$$K_2(t,p) = \{2^{2p+2}\mathrm{B}(p+1,p+1)\}^{-1}(1-t^2)^{p-1}(3-(3+2p)t^2), \; |t| \le 1; \; p = 0,1,2,... \tag{2.4}$$

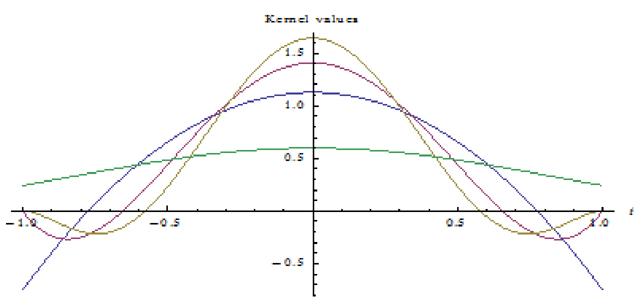


Figure 1. Plots of family of higher-order polynomial kernels. The blue, red, yellow and green lines are respectively higher-order Epanechnikov, Biweight, Triweight and Gaussian kernel.

There are two methods for obtaining the multivariate version of any univariate kernel. These are: the product kernel approach and the spherically symmetric kernel approach [12]. But, in this paper, we shall consider the former. This is given by [13] as:

$$k^P(\mathbf{t}) = \prod_{i=1}^{d} K(t_i) \tag{2.5}$$

where $K(t)$ in (2.5) is the univariate symmetric kernel [11]. Now substituting (2.4) in (2.5) gives:

$$K_d^P(\mathbf{t}) = \{2^{2p+2}\mathrm{B}(p+1,p+1)\}^{-d}\mu(\mathbf{t}), \; |\mathbf{t}| \le 1; \; p = 0,1,2,... \tag{2.6}$$

where $\mu(\mathbf{t}) = \prod_{i=1}^{d}(1-t_i^2)^{p-1}(3-(3+2p)t_i^2)$.

Theorem 2.1. *Let f be bounded density function. The kernel function $K(\mathbf{t})$ is assumed to be bounded with finite fourth moment and satisfying*

$$\|K\|_2^2 = \int_{\mathbb{R}^d} K^2(\mathbf{t})d\mathbf{t} \tag{2.7}$$

and

$$K_4 = \int_{\mathbb{R}^d} |\mathbf{t}^T\mathbf{t}|^2 \, |K(\mathbf{t})|d\mathbf{t} \tag{2.8}$$

In addition, let $\mathbf{x}$ be a vector with $f(\mathbf{x}) > 0$ and f be continuously differentiable up to the fourth order in a neighbourhood of $\mathbf{x}$. Then, the asymptotic mean integrated square error of the $\hat{f}_\lambda$ estimate can be expressed as:

$$AMISE\hat{f}_\lambda(\mathbf{x}) = (\frac{d+8}{8d})(\frac{8}{(4!)^2})^{\frac{d}{d+8}}(d(K_4.\mathbf{I}_d)^2 \|K\|_2^2 \left\|\nabla^4 f\right\|_2^2)^{\frac{8}{d+8}} n^{-\frac{8}{d+8}} \tag{2.9}$$

See [8].

Proof. The result above can be proved by using (1.2) in (1.7). Taking $d-$variate Taylor's Theorem up to the fourth order and applying (2.7) and (2.8) yields (2.9). See [8]. □

3 The Generalized Global Error Schemes

This result in Theorem 2.1 is only for the fourth moments about zero. In this section, we shall state our theorem which extends and generalizes the result in Theorem 2.1 to the $(2m+2)^{th}$ moments.

Theorem 3.1. *In addition to the conditions in Theorem 2.1, let K be any kernel in the family of multivariate $d-$ dimensional product polynomial kernels defined in (2.4) which is bounded with finite $(2m+2)^{th}$ moment and satisfying*

$$K_{2m+2} = \int_{\mathbb{R}^d} t_1^{2m+2} K(\mathbf{t})d\mathbf{t} \tag{3.1}$$

Also, let f be a bounded reference density function which is continuously differentiable up to the $(2m+2)^{th}$ order with infinite support $(-\infty, \infty)$, then the $AMISE\hat{f}_\lambda(\mathbf{x})$ is given by:

$$AMISE\hat{f}_\lambda(\mathbf{x}) = (\frac{d+4m+4}{d(4m+4)})\alpha^{\frac{d}{d+4m+4}}(d(\frac{8}{[2^{2p+2}B(p+1,p+1)]^2})^d)^{\frac{4m+4}{d+4m+4}} n^{-\frac{4m+4}{d+4m+4}} \tag{3.2}$$

where

$$\alpha = \left(\left(\frac{8}{2^{2p+2}B(p+1,p+1)}\right)^d \frac{m(4m+4)^{\frac{1}{2}}}{3^{d-1}((2m+2)!)(2m+3)(2m+5)}\right)^2 \|\nabla^{2m+2}\|_2^2 \tag{3.3}$$

Proof. The proof is of two parts.
The bias term is given by

$$\text{Bias}\hat{f}_{\mathbf{H}}(\mathbf{x}) = \mathbb{E}\hat{f}_{\mathbf{H}}(\mathbf{x}) - f(\mathbf{x}) \tag{3.4}$$

If we substitute (1.2) into (3.4) and simplify, we have,

$$\text{Bias}\hat{f}_{\mathbf{H}}(\mathbf{x}) = \int_{\mathbb{R}^d} K(\mathbf{t})f(\mathbf{x}-\mathbf{t}\mathbf{H}^{\frac{1}{2}}))d\mathbf{t} - f(\mathbf{x}) \tag{3.5}$$

Applying $d-$ variate Taylor's Theorem in (3.5) and simplify using conditions (2.5) and (2.8), we have approximate Bias to be:

$$\text{Bias}\hat{f}_{\mathbf{H}}(\mathbf{x}) = \frac{1}{(2m+2)!}[\int_{\mathbb{R}^d} (\mathbf{t}^{\mathrm{T}}\mathbf{t})^{m+1} K(\mathbf{t})d\mathbf{t}][\nabla^2 f(\mathbf{x})]^{m+1}[\text{tr}(\mathbf{H})^{m+1}] \tag{3.6}$$

Hence, as in the case of Bias, the variance term is:

$$Var\hat{f}_{\mathbf{H}}(\mathbf{x}) = (n\mathbf{H}^{\frac{1}{2}})^{-1} \|K\|_2^2 f(\mathbf{x}) \tag{3.7}$$

Substituting (3.6) and (3.7) into (1.7) and simplify we have:

$$\text{AMISE}\hat{f}_{\lambda}(\mathbf{x}) = \frac{\lambda^{4m+4}}{((2m+2)!)^2}[K_{2m+2}\cdot \text{I}_d]^2 \left\|\nabla^{2m+2}f\right\|_2^2 + (n\lambda^d)^{-1}\|K\|_2^2 \tag{3.8}$$

The value of λ that minimizes $\text{AMISE}\hat{f}_{\lambda}(\mathbf{x})$ can be achieved by differentiating (3.8) and setting the resultant expression to 0 and thus we have:

$$\lambda \cong (\frac{((2m+2)!)^2}{(4m+4)}\|K\|_2^2 \frac{d}{[K_{2m+2}\cdot \text{I}_d]^2 \|\nabla^{2m+2}f\|_2^2})^{\frac{1}{d+4m+4}} n^{-\frac{1}{d+4m+4}} \tag{3.9}$$

Substituting (3.9) into (3.8) gives:

$$\begin{aligned}\text{AMISE}\hat{f}_{\lambda}(\mathbf{x}) \cong (\frac{d+4m+4}{d(4m+4)})[\frac{(4m+4)}{((2m+2)!)^2}]^{\frac{d}{d+4m+4}}(d\|K\|_2^2)^{\frac{4m+4}{d+4m+4}}\times \\ ((K_{2m+2}\cdot \text{I}_d)^2)^{\frac{d}{d+4m+4}}(\left\|\nabla^{2m+2}f\right\|_2^2)^{\frac{d}{d+4m+4}} n^{-\frac{4m+4}{d+4m+4}}\end{aligned} \tag{3.10}$$

Equation (3.10) is completely free of the value of λ. Hence, we have been able to completely remove the rigour of first specifying the optimal bandwidth λ before obtaining the $\text{AMISE}\hat{f}_{\lambda}(\mathbf{x})$ of any multivariate higher-order polynomial kernel. This completes the first part.
In the second part, we compute the functional values K^P_{2m+2} and $||K^P||_2^2$ for the family of multivariate product polynomial kernels. Thus, we have the following:

$$\begin{aligned}K^P_{2m+2} &= \int_{\mathbb{R}^d} t_1^{2m+2}\frac{1}{[2^{2p+2}\text{B}(p+1,p+1)]^d}\prod_{i=1}^{d}(1-t_i^2)^{p-1}(3-(3+2p)t_i^2)dt_i \\ &= \frac{1}{[2^{2p+2}\text{B}(p+1,p+1)]^d}\int_{-1}^{1}.\int_{-1}^{1}.\cdots. \\ &\qquad \int_{-1}^{1} t_1^{2m+2}\prod_{i=1}^{d}(1-t_i^2)^{p-1}(3-(3+2p)t_i^2)dt_1.dt_2.\cdots.dt_d\end{aligned}$$

And therefore, on simplification we have,

$$K^P_{2m+2} = \left(\frac{8}{2^{2p+2}\text{B}(p+1,p+1)}\right)^d \frac{m}{3^{d-1}(2m+3)(2m+5)} \tag{3.11}$$

and

$$\begin{aligned}\left\|K^P\right\|_2^2 &= \int_{\mathbb{R}^d}\left(\frac{1}{[2^{2p+2}\text{B}(p+1,p+1)]^d}\prod_{i=1}^{d}(1-t_i^2)^{p-1}(3-(3+2p)t_i^2)\right)^2 dt_i \\ &= \frac{1}{[2^{2p+2}\text{B}(p+1,p+1)]^{2d}}\int_{-1}^{1}.\int_{-1}^{1}.\cdots. \\ &\qquad \int_{-1}^{1}\left[\prod_{i=1}^{d}(1-t_i^2)^{p-1}(3-(3+2p)t_i^2)\right]^2 dt_1.dt_2.\cdots.dt_d\end{aligned}$$

Thus

$$\left\|K^P\right\|_2^2 = \left(\frac{8}{[2^{2p+2}\text{B}(p+1,p+1)]^2}\right)^d \tag{3.12}$$

On substituting (3.11) and (3.12) into (3.10), we have:

$$\text{AMISE}^* \hat{f}_\lambda(\mathbf{x}) = \frac{(d+4m+4)}{d(4m+4)} \alpha^{\frac{d}{d+4m+4}} (d(\frac{8}{[2^{2p+2}\mathrm{B}(p+1,p+1)]^2})^d)^{\frac{4m+4}{d+4m+4}} n^{-\frac{4m+4}{d+4m+4}}.$$

where α is as given in (3.3). This completes the proof. □

Throughout the remaining part of this work, the value of AMISE$^* \hat{f}_\lambda(\mathbf{x})$ shall be called the global error. Since the expression of the global error still depends on an unknown density function, it cannot be used in practice. Thus, we herein in this work substitute the unknown density f with a reference density that will enable us to find the estimated value of the global error. This result has completely removed the rigour of first specifying the dimension (d) of any higher-order multivariate product kernel before obtaining its global error.

4 Simulation Experiment

To study the performance of the higher-order multivariate product polynomial kernels, the bivariate distribution $\mathbf{X} = (X_1, X_2)$ was considered and hence we conducted a Monte Carlo Simulation experiment for six bivariate normal mixture densities listed below:

(i) standard normal $f_1 \sim N(\{0,0\},\{(1,0),(0,1)\})$

(ii) skewed data $f_2 \sim \frac{1}{5}N(\{0,0\},\{(1,0),(0,1)\}) + \frac{1}{5}N(\{\frac{1}{4},\frac{1}{4}\},\{(\frac{4}{9},0),(0,\frac{4}{9})\}) + \frac{3}{5}N(\{\frac{13}{12},\frac{13}{12}\},\{(\frac{25}{81},0),(0,\frac{25}{81})\})$

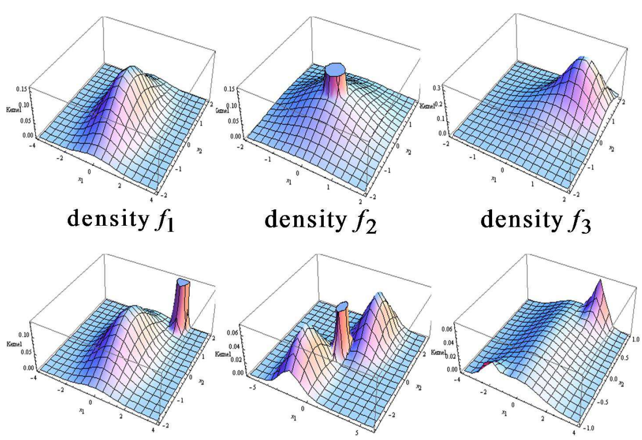
(iii) kurtosis unimodal $f_3 \sim \frac{2}{3}N(\{0,0\},\{(1,0),(0,1)\}) + \frac{1}{3}N(\{0,0\},\{(\frac{1}{100},0),(0,\frac{1}{100})\})$

(iv) asymmetric bimodal $f_4 \sim \frac{4}{5}N(\{0,0\},\{(1,0),(0,1)\}) + \frac{1}{5}N(\{2,2\},\{(\frac{1}{25},0),(0,\frac{1}{25})\})$

(v) symmetric trimodal $f_5 \sim \frac{9}{20}N(\{-\frac{7}{4},-\frac{7}{4}\},\{(1,0),(0,1)\}) + \frac{9}{20}N(\{\frac{7}{4},\frac{7}{4}\},\{(1,0),(0,1)\}) + \frac{1}{10}N(\{0,0\},\{(\frac{1}{25},0),(0,\frac{1}{25})\})$

(vi) asymmetric trimodal $f_6 \sim \frac{3}{10}N(\{-2,-2\},\{(\frac{1}{4},0),(0,\frac{1}{4})\}) + \frac{3}{10}N(\{\frac{7}{4},\frac{7}{4}\},\{(\frac{1}{5},0),(0,\frac{1}{5})\}) + \frac{3}{5}N(\{0,0\},\{(2,0),(0,2)\})$

All the bivariate densities are continuous with infinite support. However the chosen different bivariate unimodal, bimodal and trimodal densities have low and high peaks.



density f_1 density f_2 density f_3

density f_4 density f_5 density f_6

Figure 2. Graphs of bivariate densities used in the simulation experiment.

For each of the mixture densities, the vector of random variable $\mathbf{X}$ was generated and the parameters estimated from it. Thereafter, the global error as given in (3.2) was considered. The simulation was performed for $r = 1000$ runs such that the average of the global error is given as:

$$\text{Global error} = \frac{1}{r}\sum_{j=1}^{r} \text{AMISE}_j^{2m}, \; m = 1, 2, ... \tag{4.1}$$

Equation (3.2) was computed for Epanechnikov, Biweight, Triweight and Gaussian kernels and the results were presented in Figures 3 through 6 based on the kernel functions considered. Also (4.1) was computed for standard normal (f_1), skewed data (f_2), kurtosis (f_3), aymmetric bimodal (f_4), symmetric trimodal (f_5) and asymmetric trimodal (f_6) and the results were presented in Figure 7 based on the density functions considered .

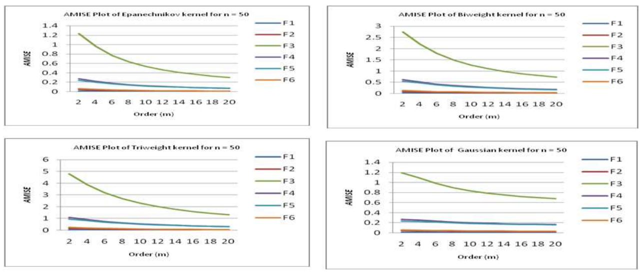


Figure 3. The Line Graphs of AMISE by Kernels n=50.

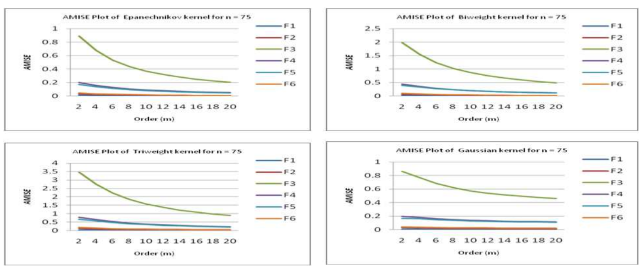


Figure 4. The Line Graphs of AMISE by Kernels n=75.

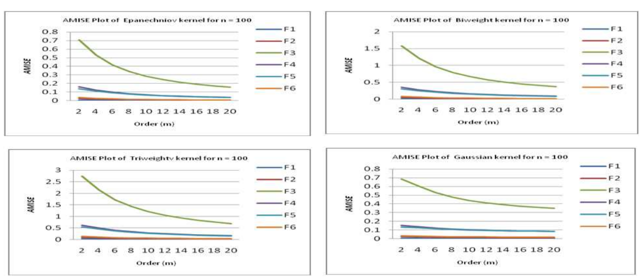


Figure 5. The Line Graphs of AMISE by Kernels n=100.

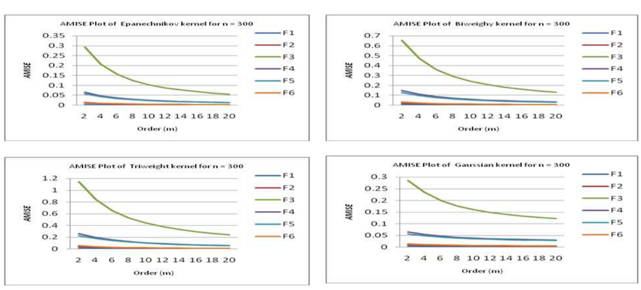


Figure 6. The Line Graphs of AMISE by Kernels n=300.

The results obtained from the simulation experiment shows the followings: In the first instance, for the six (6) bivariate mixture normal densities considered, the Epanechnikov, biweight, triweight and Gaussian kernels obeyed the large sample theorem in a manner that as n increases

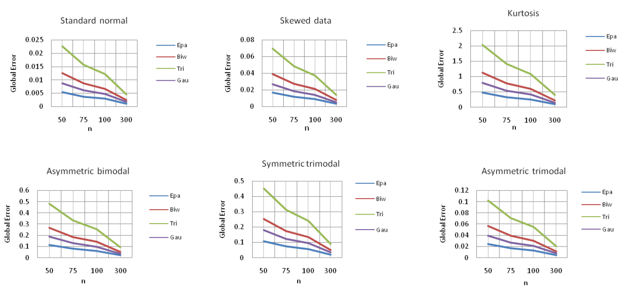


Figure 7. The Line Graphs of Global Error by Densities.

from 50, 75, 100 to 300, the global error decreases or rather tends to zero (see Figure 7). This follows one of the properties of a good estimator which states that as the sample size increases, the estimator tends to the true value and variance viz a viz global error tends to zero.

Also, the Gaussian kernel performed better than the Epanechnikov, biweight and triweightht kernels for all the mixture normal bivariate distribution. (see Figures 3, 4, 5 and 6). For all six bivariate mixture normal densities considered, the kurtosis unimodal data, f_3 has the worst performance (see Figures 3, 4, 5 and 6).

Furthermore, for all the estimation methods considered, their efficiency, improves (i.e its global error reduces) as the order of the polynomials, (m) increases, indicating that higher - order kernels have higher rate of convergence. (see Figures 3, 4, 5 and 6). And thus in general, the estimators were resistant to shifts in the mean vector especially at higher - order except in situations where that data follow a kurtosis unimodal density as can be found in f_3.

5 Concluding Remarks

In this paper, a generalized higher - order global error scheme is proposed and tested in a Monte Carlo experiment. It is also evident in this work that apart from obtaining a closed form solution for the bandwidth which minimizes the expression of the global error of higher - order kernels, we can still draw other conclusions as well. In the first instance, the introduction of non-positive kernels has led to an improvement in the global error. Secondly, a faster rate of convergence is obtained for the global error of the higher - order kernels when compared with the second and fourth order kernels in [12, 13]. However, we still observed that this new rate of convergence is slower than the univariate $(d = 1)$ case. This however can be attributed to the curse of dimensionality discussed in [11].

This paper has also simplified the rigourous work of first calculating the functional values K^P_{2m+2} and $||K^P||_2^2$ at any dimension for any kernel in the family of multivariate product polynomial kernels before obtaining its global error.

References

[1] T. Cacoullos, Estimation of a multivariate density, *Annals of the Institute of Statistical Mathematics* **18**, 179 - 189 (1966).

[2] P. Deheuvels, Estimation non parametriani de la densite par histogrammes generalises II, Publications de l'insitute de statistique de l'universite de Paris **22**, 1 - 25 (1977).

[3] V. A. Epanechnikov, Nonparametric estimation of a multivariate probability density, *Theory of Probabilty and its Applications* **14**, 153 - 158 (1969).

[4] M. Hirukawa and M.Sakudo, Nonnegative bias reduction methods for density estimation using asymmetric kernels, *Computational Statistics and Data Analysis* **75**, 112 - 123 (2014).

[5] M. C. Jones and P. J. Foster, Generalized Jackknifing and higher-order kernels, *Journal of Nonparametric Statistics* **3**, 81 - 94 (1993).

[6] K. Mynbaev, C. Martins-Filho, Reducing bias in nonparametric density estimation via bandwidth dependent kernels: L_1 view, *Statistics and Probability Letters*, 1 - 6 (2016).

[7] J. Orava, k-Nearest neigbour kernrl density estimation the choice of optimal k, *Tatra Mountains Mathematical Publications* **50**, 39 - 50 (2011).

[8] J. E. Osemwenkhae, Higher-order forms in kernel density estimation, Ph.D Thesis, University of Benin, Benin City, Nigeria, (2003).

[9] E. Parzen, On the estimation of a probability density function and the mode, *Annals of Mathematical Statistics*, **33**, 1065 - 1076 (1962).

[10] D. Ruppert, S. J. Sheather and M. P. Wand, An effective bandwidth selector for local least squares regression, *Journal of the American Statistical Association* **90** (432), 1257 - 1270 (1995).

[11] D. W. Scott, *Multivariate Density Estimation: Theory, practice and visualization*, John Wiley Sons Inc. New York, (1992).

[12] B. W. Silverman, *Density Estimation for Statistics and Data Analysis*, Chapman and Hall, London, (1986).

[13] M. P. Wand and M. C. Jones, *Kernel Smoothing*, Chapman and Hall, London, (1995).

Author information

Benson Ade Afere and Ekele Alih, Department of Mathematics and Statistics, Federal Polytechnic, Idah, Kogi State, Nigeria.
E-mail: baafere@gmail.com

Received: June 15, 2017.
Accepted: January 26, 2018.