Journal of Statistics Applications & Probability
An International Journal

http://dx.doi.org/10.18576/jsap/120340

A New Multivariate Product Kernel Functions of the Beta Polynomial Family

I. U. Siloko[1,], E. A. Siloko[2], S. A. Ojobor[3], Wasan S. Awad[4], E. Enoyoze[1], C. C. Ishiekwene[2], O. Ikpotokin[5] and E. M. Ogbeide[5]*

[1]Department of Mathematics, Edo State University Uzairue, Nigeria
[2]Department of Statistics, University of Benin, Benin City, Edo State, Nigeria
[3]Department of Mathematics, Delta State University, Abraka, Delta Stata, Nigeria
[4]Information Technology Department, College of Information Technology, Ahlia University, Manama, Bahrain
[5]Department of Mathematics and Statistics, Ambrose Alli University, Ekpoma, Edo State, Nigeria

Received: 6 Jan. 2023, Revised: 3 Mar. 2023, Accepted: 8 Apr. 2023.
Published online: 1 Sep. 2023.

Abstract: Multivariate analysis of data is of wide applicability in data science especially in big data analytic due to the volume of concealed information to be analyzed. Accurate analysis of multivariate variables is pertinent because predictions from analyzed data are good statistical indicators for making helpful decisions economically and industrially. One of the statistical analytic tools for analyzing multidimensional observations is the kernel density estimator in data exploration and visualization. The functionality of the kernel depends on the kernel function and bandwidth which influences smoothness of estimates. Several kernel functions and bandwidth selectors exist in literature; however novel estimators are being introduced to handle complex circumstances. This paper introduces a new multivariate beta kernel functions whose derivation is contingent on the product techniques. The performances of the newly introduced and existing kernels are evaluated with a known objective function and the numerical results distinctly indicating that the introduced family transcended the traditional beta family.

Keywords: Bandwidth; Beta; Density; Estimation; Kernel; Multivariate; Polynomial.

1. Introduction

One of the pivotal concepts in mathematical statistics and probability theory is density estimation which involves analysis of data. Data that are not properly interpreted often generate beguiling knowledge whose implementation may have a negative effect. Analysis of data begins with the assessment of the frequency of the data and establishment of probability estimate of the observations for visualization purposes [1]. Data smoothing techniques often times cogitate results whose conclusion can reflect the features of the observations being investigated. Density estimation in relation to data analysis has two main methods which are parametric and nonparametric techniques with the semiparametric estimation method as the hybrid of the two forms.

The parametric method of density estimation presupposes that the observations to be analyzed originates from a known distribution such that the parameter of the distribution will be the details to be estimated. The maximum likelihood estimator is one of the methods that apply the parametric estimation method. Conversely, the nonparametric density method advocates that the observations to be analyzed must not be knotted to a family of distribution and the assumption of observations being distributional is negated. As a result of the prior knowledge assumption in parametric estimation, the method has a determinate structure while the nonparametric method that depends on the observation in its estimation is flexible [2]. The resilience connected with nonparametric density estimation created a computationally expensive process that accounted for the popularity of the parametric estimation before the advent of fast computing machines. The analysis of large volume of data with statistical models that are complex do attract huge cost of computation due to the time required [3].

In studying nonparametric estimators, the kernel method has received more attention despite the numerous methods of estimation in literature. The well-received attention of kernel is ascribed to the clarity of its analytical formulation along with elucidation of outcomes wherein vital estimation issues are vividly ascertained. The kernel method is a vital nonparametric technique in determination of the intuitive probability distribution of observations. The estimator is of great industrial value in detection of anomalies and departure from the natural distribution family of data [4]. In data exploration and other numerous inferential estimation techniques, the kernel estimator has been widely applied. As a nonparametric method, the kernel estimator is of great advantage because the estimator is capable of handling large databases and obtaining smooth approximation for the observations being analyzed since nonparametric estimation is most beneficial with large sample sizes [5]. Kernel density estimation is the bedrock of most data analytic methods because the knowledge of kernel is applicable in

*Corresponding author e-mail: siloko.israel@edouniversity.edu.ng

other data smoothing environments and its ideas can be transferred to other complex estimators. The usefulness of kernel method is mainly on exploratory data analysis and data visualization with other ancillary applications in hazard rate estimation, intensity function estimation, goodness-of-fit testing, regression estimation and archaeological studies [6, 7, 8, 9, 10, 11].

This paper presents a novel multivariate kernel function of beta polynomial kernel using the product techniques with the asymptotic mean integrated squared error (AMISE) as performance metric. The structure of the other part of the article is presented thus. Section 2 discusses kernel method with the performance metric while section 3 talks about the beta kernel group with its proposed multivariate form. In section 4, results of proposed beta family and the classical beta kernels were compared using the bivariate product kernel estimator. Section 5 is the conclusion of the paper.

2 .The Kernel Density Estimator

One prominent method of investigating the statistical properties of data along with their distributional behaviour is the kernel techniques which employ the kernel estimator. Several researchers have recently employed the kernel estimator in examination of statistical details of observations by virtue of the importance of embedded information in data especially in situations where there is no historical knowledge of the data being examined [12, 13]. The kernel method of data estimation was initiated by Rosenblatt [14] and further developed by Parzen [15] for the purpose of exploratory and visualization of data in statistics. However; the estimator has gained popularity in several fields of studies due to its uniqueness in data analysis. In practice, the kernel estimator is strongly contingent on two basic statistical factors which are the smoothing parameter and the kernel formula. The compact form of the estimator as weighting function has its one-dimensional form as

$$\hat{f}(\mathrm{x}) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{\mathrm{x} - X_i}{h}\right). \quad (1)$$

Here $K(\cdot)$ is kernel formula while n is size of observations, $h > 0$ is bandwidth (smoothing parameter), x is spectrum to be considered and X_i are the observations being analyzed. The bandwidth regulates the intensity of evenness of the estimate and therefore determines the performance of the estimator. The bandwidth relies on the total number of the observations, with large sample sizes, the smoothing parameter will be smaller and vice-versa. Hence; obtaining its optimal value is critical in data examination. The kernel functions, $K(\cdot)$ must satisfied some basic axioms in addition to being symmetrical and unimodal and these axioms are given as

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

The implication of the first condition is that integral of any kernel function is one; hence several kernel functions are probability density functions. The other two conditions imply that all kernel functions must have a mean of zero and a variance greater than zero [16, 17, 18].

Irrespective of the incontrovertible usefulness of the kernel estimator in data analysis, the kernel formula alongside the bandwidth are predominant factors in kernel density estimation with much emphasis been placed on the bandwidth. In addition to the problem of accurate bandwidths' selection in kernel method, another challenge of the kernel estimator is its boundaries bias. However; several techniques have been introduced by researchers to circumvent the boundaries bias problem [13, 19, 20]. On bandwidths selection researchers are constantly introducing novel selectors since no single method can be considered generally acceptable and applicable in all circumstances.

The achievement of kernel estimator depends greatly on bandwidths and several data driven bandwidths selectors have been discussed in literature. A survey on bandwidths selectors is done by Heidenreich et al [21] on cross-validation and plug-in algorithms with the conclusion that cross validation selectors usually produce under smooth estimates and with the challenge of extreme sample variability. On the other hand, the estimates of the plug-in algorithms oftentimes display more stability unlike the cross-validation selectors but with the tendency of producing over-smoothed estimates. Smoothing parameter selectors such as the cross-validation algorithms and plug-in selectors have displayed some pitfalls; hence researchers have suggested the application of the Fast Fourier Transformation approach in the selection of suitable bandwidths. The efficiency and effectiveness of the Fourier transform technique in kernel density estimation had been demonstrated by authors such as Raykar et al. [22] and Suhre et al. [23] were the Fast Fourier Transform (FFT) resulted in bandwidths with improved performance in contrast with the over-smoothness and under-smoothness of estimates of cross validation algorithms and plug-in algorithms.

2.1 The Multivariate Kernel Density Estimator

Generally, notably applications of kernel method are basically in multivariate environment where different variables and their effects are statistically investigated. In multivariate kernel estimation, the joint probability distribution of at least two continuous random observations is estimated [11, 24, 25]. The generalized form of the kernel estimator is

$$\hat{f}(\mathrm{x}) = \frac{1}{n|H|^{1/2}} \sum_{i=1}^{n} K\left(\frac{\mathrm{x} - \mathrm{X}_i}{H^{1/2}}\right), \tag{3}$$

where H is the bandwidth matrix which is symmetric and $K(\cdot)$ denotes the multivariate kernel function. Again, the achievement of the multivariate estimator also relies on the bandwidth matrix and several techniques for its selection have been elaborately discussed in literature. One of the commonly used techniques in selection of bandwidth matrix is the minimization of a performance metric as an objective function such as AMISE of some plug-in algorithms.

Selection of bandwidth matrices in multivariate kernel method is subject to the kernel orientation and regulated by the type of parameterization. One of the parameterization methods in multivariate kernel estimation is the diagonal parameterization where the smoothing matrix is denoted by $H = \mathrm{diag}\,(h_1^2, \dots, h_d^2); h_1, \dots, h_d > 0$. The diagonal parameterization is appropriate for the product kernel estimator that employs distinct smoothing parameter in the various coordinates respectively. The multivariate product kernel is

$$\hat{f}(\mathrm{x}) = n^{-1} \left(\prod_{j=1}^{d} h_j \right)^{-1} \sum_{i=1}^{n} K\left(\frac{\mathrm{x}_1 - \mathrm{X}_{i1}}{h_1}, \frac{\mathrm{x}_2 - \mathrm{X}_{i2}}{h_2}, \dots, \frac{\mathrm{x}_d - \mathrm{X}_{id}}{h_d} \right), \tag{4}$$

with $K(\cdot)$ as the multi-dimensional kernel function, $h_j > 0,\ j = 1, 2, \dots, d$ are bandwidths of the respective coordinates and d is dimension of kernel. The advantage of the product kernel is the closed form presentation of its performance metric (AMISE) and smoothing parameter formulas contrary to other intricate forms whose optimal smoothing parameter formulas may not be easily expressed explicitly [25, 26].

A special case of the multi-dimensional kernel estimator is the bivariate case with two smoothing parameters. The bivariate product kernel applies the product of two univariate kernel estimators in analyzing its observations with a single probability function. Assuming the two observations denoted by $\mathrm{X}_i, \mathrm{Y}_i,\ i = 1, 2, \dots, n$ and n is sample size while $f(\mathrm{x}, \mathrm{y})$ is joint probability density function, the bivariate product estimator is

$$\hat{f}(\mathrm{x}, \mathrm{y}) = \frac{1}{n h_{\mathrm{x}} h_{\mathrm{y}}} \sum_{i=1}^{n} K\left(\frac{\mathrm{x} - \mathrm{X}_i}{h_{\mathrm{x}}}, \frac{\mathrm{y} - \mathrm{Y}_i}{h_{\mathrm{y}}} \right) = \frac{1}{n\, h_{\mathrm{x}} h_{\mathrm{y}}} \sum_{i=1}^{n} K\left(\frac{\mathrm{x} - \mathrm{X}_i}{h_{\mathrm{x}}} \right) K\left(\frac{\mathrm{y} - \mathrm{Y}_i}{h_{\mathrm{y}}} \right). \tag{5}$$

Again, $h_{\mathrm{x}} > 0$ and $h_{\mathrm{y}} > 0$ are bandwidths of the two coordinates while $K(\mathrm{x}, \mathrm{y})$ is bivariate kernel estimator. The popularity of bivariate product kernel estimator is hinged on the variations that usually exist in the data in their respective coordinates. One advantage of the bivariate kernel method is the power of presenting its estimates in a simple and understandable manner such as wire frames for exploration and graphical demonstration purposes. Another significant characteristic of the bivariate estimator is the determination of the kernel orientation by the bandwidth which could be generalized to higher dimensions but this unique feature is completely absent with the univariate case [1, 27].

2.2 The Performance Metric of Kernel Density Estimator

Performance in nonparametric estimator is typically evaluated by some known performance metrics such as the kullback-liebler distance, mean integrated absolute error and Hellinger distance. However; these error criteria functions are dimensionless and dimensionality is an imperative attribute in kernel estimation because its application is substantially important in higher dimensions. On account of the potential benefits of dimensionality with reference to kernel density estimation, the asymptotic mean integrated squared error (AMISE) is appropriate. The AMISE is obtained by Taylor's series expansion and its popularity over other criteria is hinged on its incorporation of dimension and the mathematical tractability. This performance metric has two components that are both influence by the bandwidth and they are the integrated variance and integrated squared bias. The AMISE has its compact form as

$$AMISE\left(\hat{f}(\mathrm{x})\right) = \int \mathrm{Variance}\left(\hat{f}(\mathrm{x})\right) d\mathrm{x} + \int \mathrm{Bias}^2\left(\hat{f}(\mathrm{x})\right) d\mathrm{x}. \tag{6}$$

On further simplification of Equation (6) using Taylor's expansion produces the AMISE given as

$$AMISE\left(\hat{f}(\cdot)\right) = \frac{R(K)}{nh_{\mathrm{x}}} + \frac{1}{4}\mu_2(K)^2 h_{\mathrm{x}}^4 R(f''), \tag{7}$$

with $R(K) = \int K(\mathrm{x})^2 dx$ as kernel roughness, $\mu_2(K)^2$ is its moment whilst $R(f'') = \int f''(\mathrm{x})^2 dx$ is roughness of unknown distribution for the estimation process. The AMISE rely completely on the bandwidth which determines its value. The bandwidth that yields the optimal numerical value is

$$h_{\mathrm{AMISE}} = \left[\frac{R(K)}{\mu_2(K)^2 R(f'')}\right]^{1/(d+4)} \times n^{-1/(d+4)}. \tag{8}$$

The multivariate version of the AMISE in Equation (7) using the product technique is

$$\begin{aligned} \mathrm{AMISE}\left(\hat{f}(\mathrm{x})\right) &= \frac{R(K)^d}{nh_1 h_2, \dots, h_d} + \frac{1}{4} h_j^4 \mu_2(K)^2 \int tr^2\left(\nabla^2 f(\mathrm{x})\right) d\mathrm{x} \\ &= \frac{R(K)^d}{nh_1 h_2, \dots, h_d} + \frac{1}{4} h_j^4 \mu_2(K)^2 R\left(\nabla^2 f(\mathrm{x})\right). \end{aligned} \tag{9}$$

The closed form of the optimal bandwidth for Equation (9) is

$$H_{AMISE} = \left[\frac{dR(K)^d}{\mu_2(K)^2 \int tr^2\left(\nabla^2 f(\mathrm{x})\right) dx}\right]^{\left(\frac{1}{d+4}\right)} \times n^{-\left(\frac{1}{d+4}\right)}, \tag{10}$$

with $R\left(\nabla^2 f(\mathrm{x})\right) = \int tr^2\left(\nabla^2 f(\mathrm{x})\right) dx$ is the roughness of the distribution and tr represents the trace of matrix, $h_1, h_2, \dots, h_d$ are bandwidths for the different dimension while $\nabla^2 f(\mathrm{x})$ is the distribution Hessian matrix [28].

3. Methodology

3.1 The beta polynomial kernel function

Several families of kernel estimators exist in literature and one prominent family is the beta polynomial family whose form is given as

$$K_p(t) = \frac{(2p+1)!}{2^{2p+1}(p!)^2}(1-t^2)^p, \tag{11}$$

where $p = 0, 1, 2, \dots, \infty$ is the determinant of the resulting kernel and t is evaluated within the interval $[-1, 1]$. The popularity of the beta kernel is due to their fascinating mathematical attributes. The values of p determines the produced kernel such that the simplest kernel called the Uniform kernel is when $p = 0$ but when $p = 1$, we have the Epanechnikov function called the optimal kernel with respect to the AMISE. Other higher values such as $p = 2, 3$ and 4 will produce Biweight kernel, Triweight kernel and Quadriweight kernels. The Gaussian kernel that is of great application in mathematical statistics did not belong to this family, however; it is obtained when p moves to infinity [29, 30]. The Epanechnikov, Biweight, Triweight and Quadriweight kernels are mathematically expressed in the forms

$$\begin{cases} K_1(t) = \frac{3}{4}(1-t^2). \\ K_2(t) = \frac{15}{16}(1-t^2)^2. \\ K_3(t) = \frac{35}{32}(1-t^2)^3. \\ K_4(t) = \frac{315}{256}(1-t^2)^4. \end{cases} \tag{12}$$

Similarly, as p approaches infinity, the resulting function is the Gaussian function given as

$$K_{\emptyset}(t) = \frac{1}{\sqrt{2\pi}} exp\left(-\frac{t^2}{2}\right). \tag{13}$$

Apart from the Uniform kernel, the other member that is Epanechnikov (quadratic), Biweight (quartic), Triweight and Quadriweight kernels are widely used in nonparametric estimation. The computations of efficiency of the beta polynomial kernels do employ the Quadratic function because of its optimality quality of the AMISE.

3.2 The Multivariate Beta Kernel Function.

Constructing multivariate kernel functions using the product techniques usually involved the multiplication of several univariate functions. The product multivariate version of the beta kernels is given by

$$K_p(t) = M^d \prod_{i=1}^{d} (1 - t_i^2)^p, \tag{14}$$

where $M = \frac{(2p+1)!}{2^{2p+1}(p!)^2}$ is the constant of normalization of the function while d is its dimension. The bivariate product kernel of the beta family is frequently used in density estimation for visualization of two observations and its compact form is

$$K_p(t) = M^2 \prod_{i=1}^{2} (1 - t_i^2)^p, \tag{15}$$

The compact form of two-dimensional product kernel in Equation (15) is express as

$$K_p(t) = M^2 (1 - t_1^2)^p (1 - t_2^2)^p. \tag{16}$$

Hence, for kernel functions with values of $p = 1, 2, 3$ and 4, the corresponding bivariate forms are as follows

$$\begin{cases} K_1(t) = \left(\frac{3}{4}\right)^2 (1 - t_1^2)(1 - t_2^2) \\ K_2(t) = \left(\frac{15}{16}\right)^2 (1 - t_1^2)^2 (1 - t_2^2)^2 \\ K_3(t) = \left(\frac{35}{32}\right)^2 (1 - t_1^2)^3 (1 - t_2^2)^3 \\ K_4(t) = \left(\frac{315}{256}\right)^2 (1 - t_1^2)^4 (1 - t_2^2)^4 \end{cases} \tag{17}$$

Again, as in the univariate case, the bivariate product Gaussian kernel is given by

$$K_\emptyset(t) = \frac{1}{2\pi} exp\left(-\frac{t_1^2 + t_2^2}{2}\right). \tag{18}$$

The bivariate Gaussian kernel estimator is widely applied in mathematical statistics and other statistical related fields owing to its simplicity with its continuous differentiability attribute.

3.3 The Proposed Multivariate Beta Polynomial Kernel Functions.

The proposed beta kernel employed the principle of exponential progression in its formulation with a constant common factor to members of the family. The members of the beta kernel are regarded as the consecutive terms of the sequence for which $p = 1, 2, 3$ and 4 stands for Epanechnikov, Biweight, Triweight, and Quadriweight functions.
Assuming the first member is denoted by a while the common factor is represented by r with $K_p(t)$ representing current term while $K_{p-1}(t)$ representing the immediate past term, then the common factor of the sequence is

$$r = \frac{K_p(t)}{K_{p-1}(t)}. \tag{19}$$

Recall Equation (12), the first member of the beta kernel when $p = 1$ is

$$a = K_1(t) = \frac{3}{4}(1 - t^2). \tag{20}$$

The factor common to the members of the kernel is obtain by

$$r = \frac{K_2(t)}{K_1(t)} = \frac{\frac{15}{16}\,(1-t^2)^2}{\frac{3}{4}\,(1-t^2)} = \left(\frac{15\,(1-t^2)^2}{16}\right)\left(\frac{4}{3(1-t^2)}\right) = \frac{5}{4}\,(1-t^2). \tag{21}$$

Similarly, this common factor or ratio is

$$r = \frac{K_3(t)}{K_2(t)} = \frac{\frac{35}{32}\,(1-t^2)^3}{\frac{15}{16}\,(1-t^2)^2} = \left(\frac{35\,(1-t^2)^3}{32}\right)\left(\frac{16}{15(1-t^2)^2}\right) = \frac{7}{6}\,(1-t^2). \tag{22}$$

It is also possible to obtain the common factor as

$$r = \frac{K_4(t)}{K_3(t)} = \frac{\frac{315}{256}\,(1-t^2)^4}{\frac{35}{32}\,(1-t^2)^3} = \left(\frac{315\,(1-t^2)^4}{256}\right)\left(\frac{32}{35(1-t^2)^3}\right) = \frac{9}{8}\,(1-t^2). \tag{23}$$

As depicted in Equations (21–23) the common factor is expressed as

$$r = \begin{cases} \frac{5}{4}\,(1-t^2) \\ \frac{7}{6}\,(1-t^2) \\ \frac{9}{8}\,(1-t^2) \end{cases} \tag{24}$$

Therefore, the generalized common factor of any two successive terms of Equation (24) is

$$r = \frac{3+2p}{2+2p}(1-t^2). \tag{25}$$

Here $p = 1, 2, 3\,, \ldots$ determines the generated kernel.

3.4 The P^{th} Term of the Proposed Kernel Functions

If $K_p(t)$ is the P^{th} beta member with a denoting the first member and r representing the common factor, then

$$K_1(t) = a$$

$$\frac{K_2(t)}{K_1(t)} = r, \quad \therefore K_2(t) = K_1(t) \times r = ar$$

$$\frac{K_3(t)}{K_2(t)} = r, \quad \therefore K_3(t) = K_2(t) \times r = ar \times r = ar^2$$

$$\frac{K_4(t)}{K_3(t)} = r, \quad \therefore K_4(t) = K_3(t) \times r = ar^2 \times r = ar^3$$

The generalized form of the P^{th} proposed member with the first term as denoted in Equation (20) and common factor in Equation (25) is given by

$$K_p(t) = ar^{p-1}, \qquad p = 1, 2, 3, \ldots \tag{26}$$

On substituting Equation (20) and Equation (25) into Equation (26), we obtain the generalized form of the proposed beta kernel as

$$K_p(t) = \left(\frac{3}{4}\,(1-t^2)\right)\left(\frac{3+2p}{2+2p}(1-t^2)\right)^{p-1}, \qquad p = 1, 2, 3, \ldots \tag{27}$$

The proposed kernel is same as the classical kernel when $p = 0$ and 1 giving rise to the Uniform and Epanechnikov kernels respectively. As in the classical case, the optimum kernel is the Epanechnikov with reference to the AMISE. Now, for values of p such as $p = 1, 2, 3, 4$, the new kernels are as follows

$$\begin{cases} K_1(t) = \frac{3}{4}(1-t^2) \\ K_2(t) = \frac{7}{8}(1-t^2)^2 \\ K_3(t) = \frac{243}{256}(1-t^2)^3 \\ K_4(t) = \frac{3993}{4000}(1-t^2)^4 \end{cases} \tag{28}$$

The multivariate product version of Equation (28) can simply be constructed using Equation (14). The bivariate product form of the first four members of the proposed kernel is

$$\begin{cases} K_1(t) = \left(\frac{3}{4}\right)^2 (1-t_1^2)(1-t_2^2) \\ K_2(t) = \left(\frac{7}{8}\right)^2 (1-t_1^2)^2(1-t_2^2)^2 \\ K_3(t) = \left(\frac{243}{256}\right)^2 (1-t_1^2)^3(1-t_2^2)^3 \\ K_4(t) = \left(\frac{3993}{4000}\right)^2 (1-t_1^2)^4(1-t_2^2)^4 \end{cases} \tag{29}$$

The choices of kernel functions are usually base on the computational efficiency of the function and its performance with reference to a particular error criterion function. A method of kernel density estimation is deemed the best when it produces the least AMISE value assuming the error criterion function is the AMISE [31, 32].

4. Results and Discussion.

The performance of the introduced and traditional kernels is compared numerically using the AMISE and graphically for picturization. Mathematica version 12 is the software employed for all the graphical demonstration and numerical analysis. The statistical attributes of four members will be investigated with different sample sizes and real data. The sizes of 5000 and 10000 is apply in investigating the performance of the introduced kernels and their classical counterpart on account of the potential gains associated with large sample sizes in nonparametric estimation particularly in kernel estimation.

Table 1: Univariate AMISE with Different Sizes

Kernel Functions	**Classical Kernels AMISE**		**Proposed Kernels AMISE**	
Types of Kernels	N=5000	N=10000	N=5000	N=10000
Epanechnikov	0.000893518	0.000513192	0.000893518	0.000513192
Biweight	0.001325099	0.000761069	0.001138490	0.000653891
Triweight	0.001531044	0.000879354	0.001139209	0.000654379
Quadriweight	0.001600333	0.000919150	0.001143551	0.000654531

Table 2: Bivariate AMISE with Different Sizes

Kernel Functions	**Classical Kernels AMISE**		**Proposed Kernels AMISE**	
Types of Kernels	N=5000	N=10000	N=5000	N=10000
Epanechnikov	0.001551193	0.000977190	0.001551193	0.000977190
Biweight	0.002674226	0.001684657	0.002029295	0.001278376
Triweight	0.003128830	0.001971040	0.001774904	0.001311812
Quadriweight	0.003759469	0.002368317	0.001829537	0.001526544

Table 1 and Table 2 are results of the different sizes and from the results, the AMISE values of the proposed kernels are smaller in comparison with the AMISE values of the traditional kernels indicating they outperformed the traditional counterpart. The superiority of any kernel method over existing methods is contingent on its capacity of producing the minimal value when investigated with a known performance metrics. Figure 1 and Figure 2 are graphical displays of the novel and conventional functions of the univariate case with both graphs displaying similarity in appearance but in performance the proposed kernels outperformed the existing kernels as vividly shown in Table 1 and Table 2 for the univariate and bivariate cases.

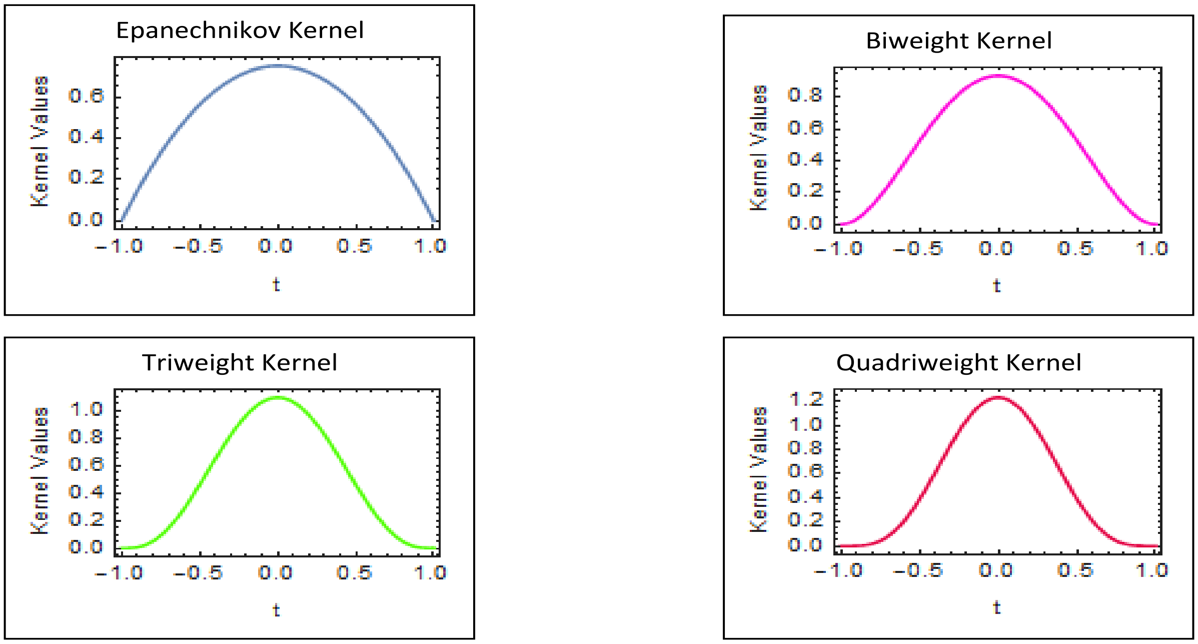

Fig. 1: Graphs of Classical Beta Polynomial Kernel Functions

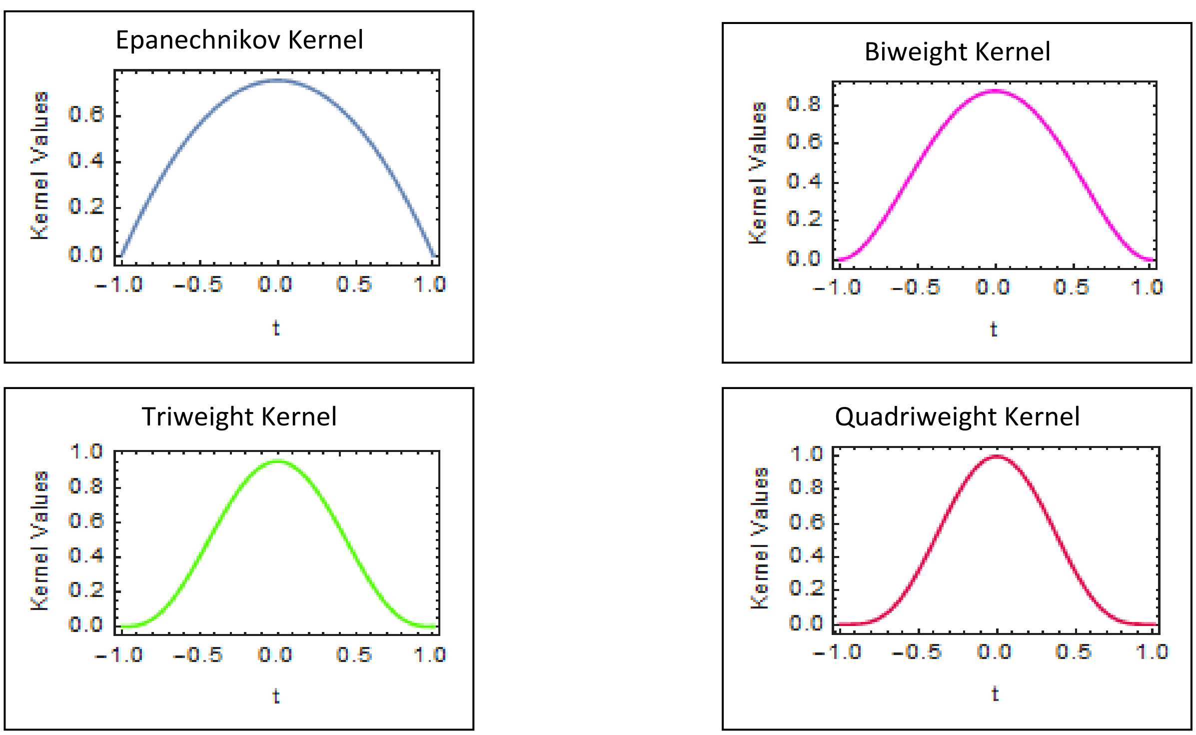

Fig. 2: Graphs of Proposed Beta Polynomial Kernel Functions

Figures (3–9) are bivariate graphs of the conventional and novel product kernel functions. As observed from the estimates the conventional and newly introduced family, the loops of the graphs moved nearer to the center for the Triweight and Quadriweight kernels when compared with the loops of the Epanechnikov and Biweight kernels. The movement of the loop to the center of the graph in cases of higher values of p is occasioned by the degrees of differentiability. Kernel functions that have higher p tends to produce graphs whose loops are towards the center due to the possession of more derivatives and by implication will produce better kernel estimates with reference to level of smoothness.

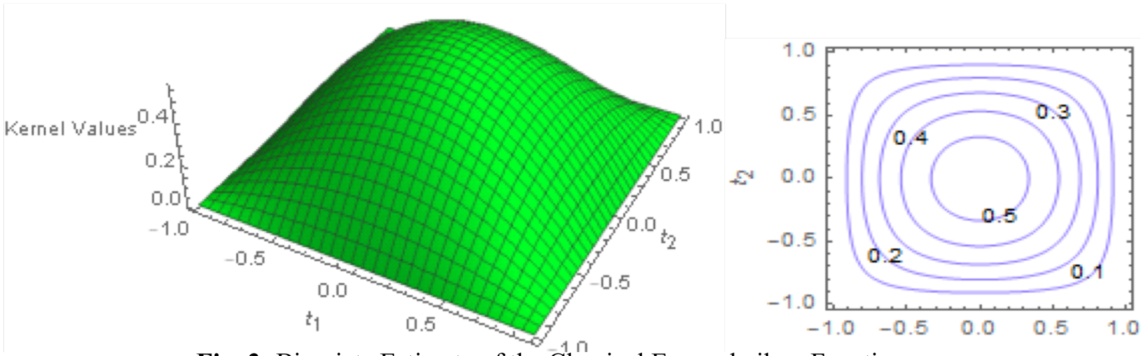

Fig. 3: Bivariate Estimate of the Classical Epanechnikov Function

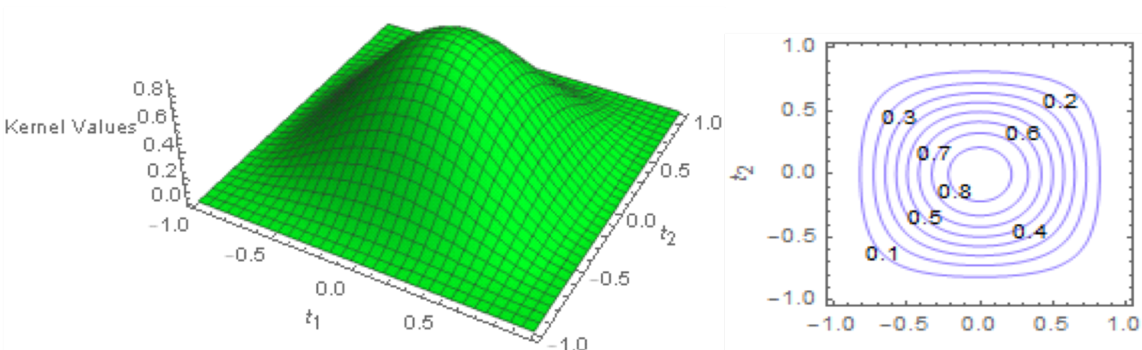

Fig. 4: Bivariate Estimate of the Classical Biweight Function

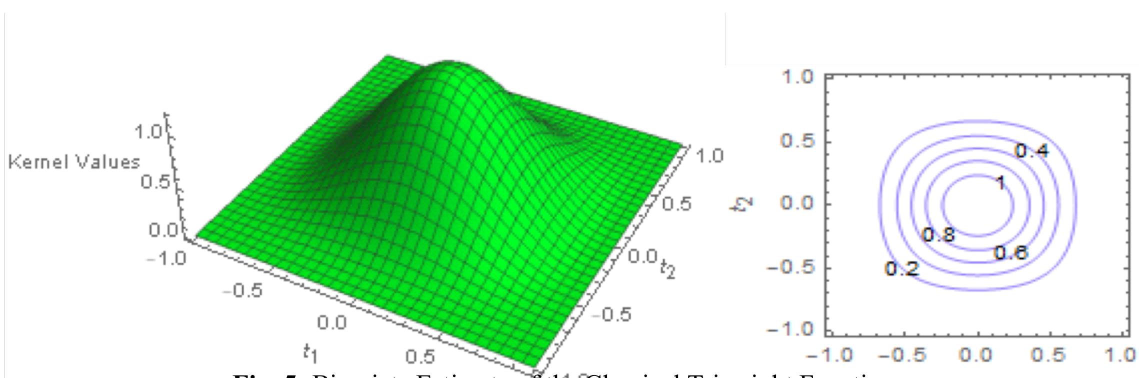

Fig. 5: Bivariate Estimate of the Classical Triweight Function

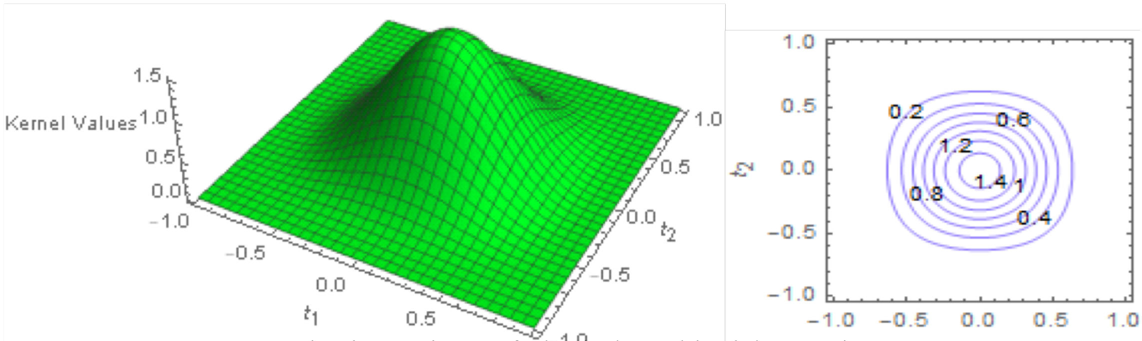

Fig. 6: Bivariate Estimate of Classical Quadriweight Function

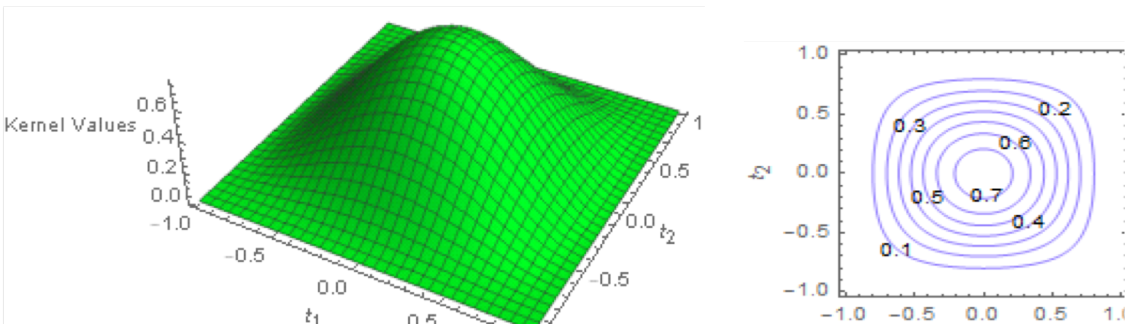

Fig. 7: Bivariate Estimate of Proposed Biweight Function

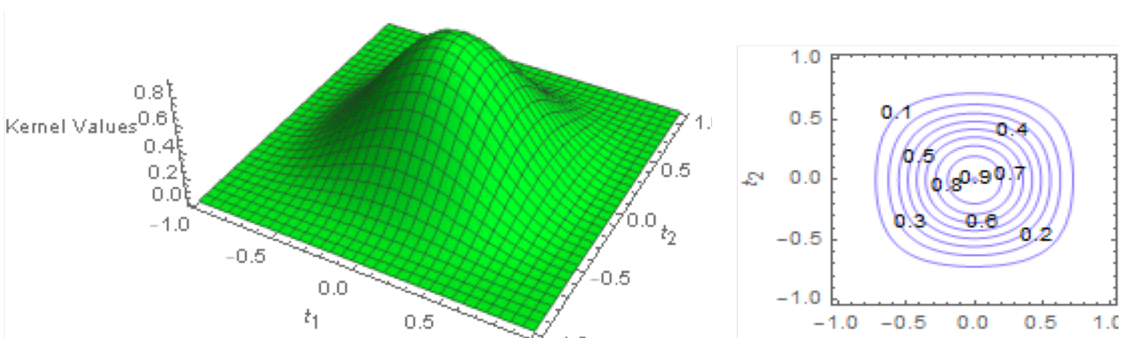

Fig. 8: Bivariate Estimate of the Proposed Triweight Function

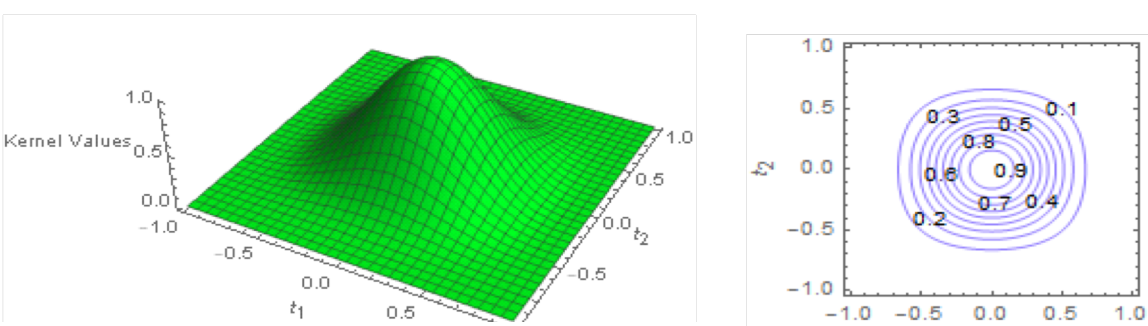

Fig. 9: Bivariate Estimate of the Proposed Quadriweight Function

Despite the similarity depicted by the graphs in Figures (3–9), the empirical results in Table 2 have demonstrated that the proposed kernels outperformed the classical version with the AMISE and this is due to the variation in their normalization constant. The effect of dimension was also displayed in Table 1 and Table 2 as noticed in the AMISE of the univariate and bivariate case. As the dimension increases, the AMISE values increases which is curse of dimensionality effect of nonparametric density estimation [33].

The old faithful data is use to access the performances of newly introduced kernels [34]. The bivariate data comprises of two hundred and seventy-two (272) observations with duration and waiting time in minutes representing the two axes respectively. The examination of bivariate observations oftentimes begins with the consideration of the scatterplot of the data because they are the mostly employed tools for displaying bivariate data graphically. However, the scatterplot as observed in most situations will not highlight the fundamental features in the observation but draws the attention of the observer to the peripheries of the clouded density of the data point. The hiccups of the obscured nature of the features of the data cloud generated by the scatterplot are surmounted by the kernel estimate especially in the preservation and presentation of vital inherent information of the observations. The pictorial demonstration of the observation portrays two modes but this bimodality is not apparent in the scatterplot which reveals the applications of bivariate estimates in structures identification. The bimodal nature of the data indicated that duration of eruption and waiting period do exhibit bimodal characteristics [27]. The scatterplot of the observations is in Figure 10 and with vivid evidence of positive correlation. The value of the correlation coefficient is 0.90087.

Table 3 is the smoothing parameters and AMISE values of the data with the Epanechnikov kernel maintaining its optimality quality. In this real data application of the proposed kernel, the investigation will be centered on the Biweight kernel because other members will exhibit similar characteristics. As earlier stated, for $p = 0$ and 1, the resulting proposed kernels which are Uniform and Epanechnikov kernels are the same as the classical kernels. Again, as p increases, there is decrease in magnitude of bandwidths which ultimately affects the AMISE value. It is obviously shown in Table 3 that the Biweight AMISE of the conventional family is larger than that of the new family demonstrating that the latter outperformed the former kernel functions.

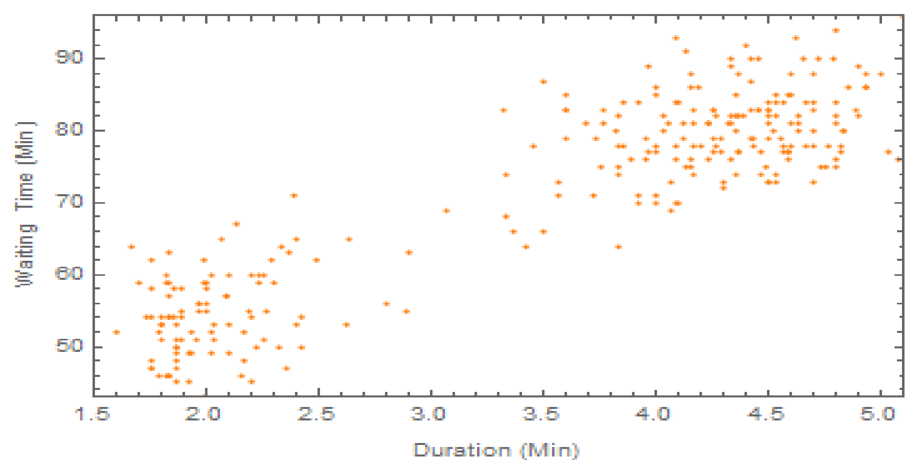

Fig. 10: Scatter Diagram of the Old Faithful Data

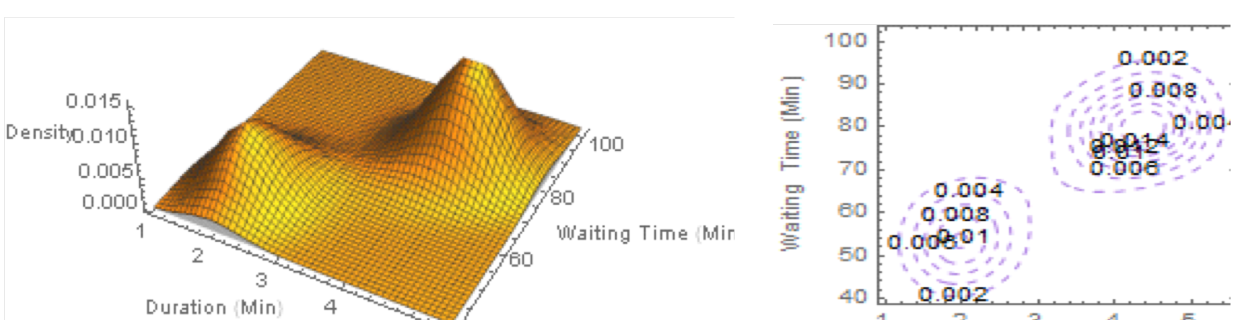

Fig. 11: Bivariate Estimate of Epanechnikov Kernel of Old Faithful Data

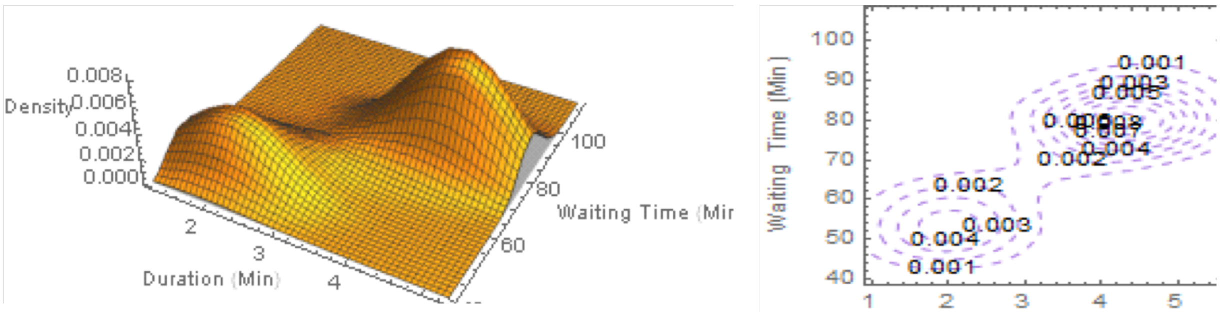

Fig. 12: Bivariate Estimate of Classical Biweight Kernel of Old Faithful Data

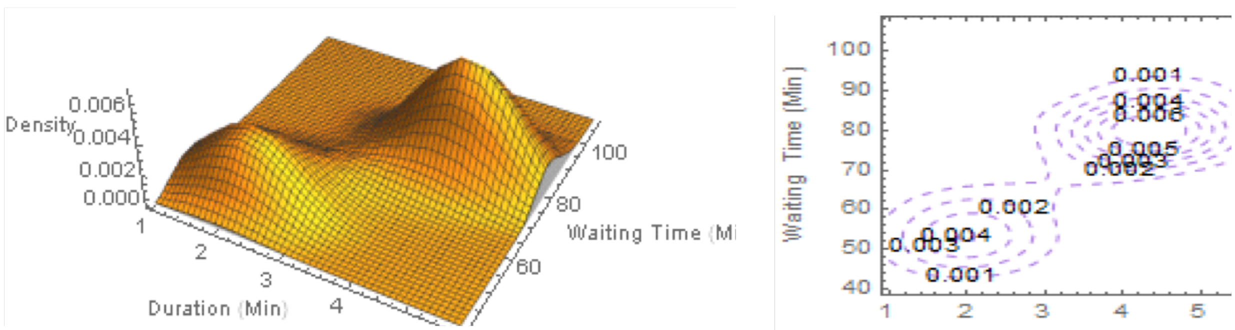

Fig. 13: Bivariate Estimate of Proposed Biweight Kernel of Old Faithful Data

Table 3: AMISE of Conventional and New Kernel Functions for Old Faithful Data

Kernel Function	**AMISE of Classical Kernel Functions**			**AMISE of Proposed Kernel Functions**		
Kernel Types	h_x	h_y	AMISE	h_x	h_y	AMISE
Epanechnikov	0.565089	6.729150	0.010804381	0.565089	6.729150	0.010804381
Biweight	0.512356	6.101194	0.018626539	0.522356	6.121191	0.014134461

One of the indispensable roles of kernel techniques is the exploration of data and pictorial presentation for visualization. Data visualization could either be in two-dimensional or three-dimensional forms such as the scatterplot, contour plots or the surface plots. Accurate analysis of data is vital in taking decisions that are data reliant, hence efforts should be geared towards avoidance of inaccurate results during data analysis. The data estimates delineate the immanent attribute of bimodality which is a unique characteristic of the data. The bimodality of the observations supports the claim that such data often exhibit bimodal distribution. As clearly displayed in Figures (11–13), the traditional and proposed estimates retained that unique characteristic of bimodality of the observations. The estimate of the proposed Biweight kernel compete favourably well with the estimate of the traditional Biweight kernel in terms of retention of inherent statistical quality of the data. Hence the estimates of the proposed kernel family will retain inherent characteristics of observations as vividly demonstrated.

5. Conclusion.

This paper introduces a new multivariate beta family from its classical kernels with the aid of the multivariate product construction techniques. On numerical evaluation with the AMISE as performance measure, the introduced kernel has established its superiority over the classical kernels in empirical verification and data application with emphasis on the univariate and bivariate cases. Features retention is a vital aspect of multivariate kernel density estimation especially in bivariate estimation primarily for extraction of information in decision making and future prediction. Multivariate kernel estimate provides enormous information with reference to essential features of data, hence the proposed kernel functions like their classical counterparts also retained the statistical properties of the data investigated.

Conflicts of Interest Statement

The authors declare that there is no conflict of interest.

Acknowledgement

The authors sincerely appreciate the anonymous reviewers and the editorial members for painstakingly going through the manuscript and making positive suggestions that improved the paper.

References

[1] B. W. Silverman, Density estimation for statistics and data analysis, *Routledge, Abingdon*, UK, 3–10, (2018).
[2] I. U. Siloko, O. Ikpotokin, F. O. Oyegue, C. C. Ishiekwene, B. A. 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).
[3] P. J. Green, K. Latuszyski, M. Pereyra, C. P. Robert, Bayesian computation: A summary of the current state, and samples backwards and forwards, *Statistics and Computing,* **25**, 835–862 (2015).
[4] S, Wang, J, Wang, F. L. Chung, Kernel density estimation, kernel methods, and fast learning in large data sets, *IEEE Transactions on Cybernetics,* **44(1)***,* 1–20 (2014).
[5] K. Jiwani. Optimal kernel density estimation using FFT based cost function, IEEE International Conference on Data Mining, Sorrento, Italy, (2020).
[6] I. U. Siloko, K. E. Ukhurebor, E. A. Siloko, E. Enoyoze, A. O. Bobadoye, C. C. Ishiekwene, O. O. Uddin, W. Nwankwo, Effects of some meteorological variables on cassava production in Edo State, Nigeria via density estimation. *Scientific African,* e00852, https://doi.org/10.1016/j.sciaf.2021.e00852, (2021).
[7] I. U. Siloko, K. E. Ukhurebor, C. C. Ishiekwene, E. A. Siloko, O. O. Uddin, E. Enoyoze, Statistical estimation of some meteorological variables using the beta kernel function. *Ethiopian Journal of Environmental Studies and Management*, **14(4)**, 474–486 (2021).
[8] P. Yin, Kernels and Density Estimation, *The Geographic Information Science & Technology Body of Knowledge*. DOI: 10.22224/gistbok/2020.1.12, (2020).
[9] A. Bonnier, M. Finné, E. Weiberg, Examining land-use through GIS-Based kernel density estimation: A Re-Evaluation of legacy data from the Berbati-Limnes Survey, *Journal of Field Archaeology,* **44(2)**, 70–83 (2019).
[10] G. S. Mariani, F. Brandolini, M. Pelfini, A. Zerboni, Matilda's Castles, Northern Apennines: Geological and geomorphological constrains, *Journal of Maps,* **15(2)**, 521–529 (2019).
[11] J. E. Chacón and T. Duong. Multivariate kernel smoothing and its applications. Boca Raton, FL: Chapman & Hall/CRC Press, 705–77 (2018).
[12] Y. Ziane, N. Zougab, S. Adjabi, Body tail adaptive kernel density estimation for nonnegative heavy-tailed data, *Monte Carlo Methods and Applications,* **27(1)**, 57–69 (2021).
[13] C. Bolancé, C, C. A. Acuña, A New Kernel Estimator of Copulas Based on Beta Quantile Transformations, *Mathematics*, 9, 1078. https://doi.org/10.3390/math9101078, (2021).
[14] M. Rosenblatt, Remarks on some nonparametric estimates of a density function, *Annals of Mathematical Sciences*, **27**, 832–837 (1956).
[15] E. Parzen, On the estimation of a probability density function and the mode, *Annals of Mathematical Statistics,* **33**, 1065–1076 (1962).
[16] I. U. Siloko, E. A. Siloko, O. Ikpotokin, A mini review of dimensional effects on asymptotic mean integrated squared error and efficiencies of selected beta kernels, *Jordan Journal of Mathematics and Statistics*, **13(3)**, 327–340 (2020).
[17] D. W. Scott, Multivariate density estimation, Theory, practice and visualization. Second edition, Wiley, New Jersey, 15–23 (2015).
[18] M. P. Wand, M. C. Jones. Kernel Smoothing; Chapman & Hall: London, UK, 105–120 (1995).
[19] S. Chen, T. Huang, Nonparametric estimation of copula functions for dependence modeling, *Canadian Journal Statistics*, **35**, 265–282 (2007).
[20] M. Omelka, I. Gijbels, N. Veraverbeke, Improved kernel estimation of copulas: Weak convergence and goodness-of-fit testing, *Annals of Statistics*, **37**, 3023–3058 (2009).
[21] H. N. Heidenreich, A. Schindler, S. Sperlich, Bandwidth selection for kernel density estimation: A review of fully automatic selectors, *German Statistical Society*, **97(4)**, 403–433 (2013).
[22] V. C. Raykar, D. Ramani, H. Z. Linda, Fast computation of kernel estimators, *Journal of Computational and Graphical Statistics*, **19(1)**, 205–220 (2010).
[23] A. Suhre, A. Orhan, E. C. Ahmed, Bandwidth selection for kernel density estimation using Fourier domain constraints, *IET signal processing,* **10(3)**, 280–283 (2016).
[24] M. Groß, U. Rendtel, T. Schmid, S. Schmon, N. Tzavidis, Estimating the density of ethnic minorities and aged people in

Berlin: Multivariate kernel density estimation applied to sensitive geo-referenced administrative data protected via measurement error, http://hdl.handle.net/10419/107684, (2015).
[25] V. A. Epanechnikov, Nonparametric estimation of a multivariate probability density, *Theory Probab. Appl.,* **14**, 153–158 (1969).
[26] T. Duong, M. L. Hazelton, Plug-in bandwidth matrices for bivariate kernel density estimation, *Journal of Nonparametric Statistics,* **159 (1)**, 17–30 (2003).
[27] I. U. Siloko, C. C. Ishiekwene, F. O. Oyegue, New gradient methods for bandwidth selection in bivariate kernel density estimation, *Journal of Mathematics and Statistics,* **6(1)**,1–8 (2018).
[28] R. S. Sain, Multivariate Locally Adaptive Density Estimation, *Computational Statistics and Data Analysis*, **39**,165–186 (2002).
[29] I. U. Siloko, O. Ikpotokin, E. A. Siloko, On hybridizations of fourth order kernel of the beta polynomial family, *Pakistan Journal of Statistics and Operations Research*, **15 (3)**, 819–829 (2019).
[30] J. Marron, D. Nolan, Canonical kernels for density estimation, *Statistics and Probability Letter*, **7**, 195–199 (1988).
[31] I. U. Siloko, W. Nwankwo, E. A. Siloko, A new family of kernels from the beta polynomial kernels with applications in density estimation, *International Journal of Advances in Intelligent Informatics*, **6(3)**, 235–245 (2020).
[32] J. Jarnicka, Multivariate kernel density estimation with a parametric support, *Opuscula Mathematica,* **29(1)**, 41–45 (2009).
[33] I. U. Siloko, O. Ikpotokin, F. O. Oyegue, E. A. Siloko, C. C. Ishiekwene, Numerical computation of efficiency of beta polynomial kernels using product method, *Journal of Applied Science and Technology,* **23 (1&2)**, 33–39 (2019).
[34] A. Azzalini, A. W. Bowman, A Look at some data on the Old Faithful Geyser. *Applied Statistics,* **39**, 357–365 (1990).