Proceedings of the Pakistan Academy of Sciences: A Physical and Computational Sciences 63(1): 61-70 (2026)

ISSN (Print): 2518-4245; ISSN (Online): 2518-4253
http://doi.org/10.53560/PPASA(63-1)706

Pakistan Academy of Sciences

Research Article

Density Estimation and Efficiency Analysis of a New Beta Polynomial Family

Edith Akpevwe Siloko[1]*, Rita Nneka Nwaka[2], Israel Uzuazor Siloko[3], Esosa Enoyoze[3], Oyaimare Osemengbe Uddin[4], Simon Ejokema Imoisi[5], Cyril Chukwuka Ishiekwene[6], and Osayomore Ikpotokin[7]

[1]Department of Mathematics, Delta State College of Education, Mosogar, Nigeria
[2]Department of Mathematics and Statistics, University of Delta, Agbor, Nigeria
[3]Department of Mathematics and Statistics, Edo State University, Iyamho, Nigeria
[4]Department of Computer Science, Edo State University, Iyamho, Nigeria
[5]Department of Public Law, Edo State University, Iyamho, Nigeria
[6]Department of Statistics, University of Benin, Benin City, Nigeria
[7]Department of Statistics, Ambrose Alli University, Ekpoma, Nigeria

Abstract: A popular non-parametric approach constantly employed in the estimation of probability density function (PDF) of data is kernel density estimation (KDE) and the technique is vital in several statistical methodologies because of its functionality in data analysis. This study assesses the efficacy and efficiency of new kernel family known as the new beta polynomial family (NBPF) kernels which is generated by additional power to the polynomial functional form. The numerical efficiency of the newly introduced kernel family improves substantially due to the integration of the functional modification. The efficiencies of the classical polynomial family decreases with increase in the polynomial power while the reverse is the case with the new family which demonstrates an exceptional improvement with increase in its power and also sustaining computational flexibility. A comparative assessment of the efficiencies of NBPF shows that it exhibits superiority over the existing beta kernel family (BKF), demonstrating their adaptability to modern techniques in probability density functions without rigid parametric assumptions. Again, a real data application of the NBPF reveals that the family possesses retention capacity of intrinsic characteristics of the dataset. The numerical improvement of the efficiencies of NBPF as well as the ability of NBPF to retain essentials statistical characteristics emphasise the importance of the NBPF in data analysis and data visualization.

Keywords: Bandwidth, Beta, Density, Estimation, Efficiency, Kernel.

1. INTRODUCTION

Kernel density estimation (KDE) is one of the vital non-parametric methods in statistics primarily for the estimation of probability density functions (PDFs) of random variables [1-3]. The classical parametric methods usually assume that data been evaluated belong to distributional group, whereas, KDE is void of distributional assumption but allow the observations to speak for themselves. The lack of distributional assumptions in KDE enhances more flexibility in the detection of complex features, particularly in real-life observations. Kernel estimation is greatly important in the analysis of data with no prior knowledge on the fundamental distribution by generating a smooth continuous density estimate over the entire datasets [4]. KDE usually generate a smooth curve that provides the information embedded in the data and also display statistical features such as peaks known as the

Received: December 2025; Revised: February 2026; Accepted: March 2026
* Corresponding Author: Edith Akpevwe Siloko <siloko.akpevwe@descoem.edu.ng>

modes of the data and troughs of the distributions. The analysis of data using the KDE strategy is vital in most data analytic techniques and of wide applications as well as a strong foundation in different fields. One of the importance of the KDE approach is that the knowledge derived is generic and can be extended to other smoothing setting. Again, the effectiveness of the KDE approach in several disciplines is due to the simplicity of extending the technique to other complex estimation methodologies.

The KDE methodology is a prominent tool in exploratory data analysis (EDA) and data visualization (DV) especially amid investigation of the statistical composition of dataset in reference to the underlying probability distribution. One of the essential benefits of KDE is the idea of non-imposition of distributional assumptions on the observations which has resulted in widespread use of the approach on different datasets [5-7]. The occurrence of vehicular road accidents between 2010 and 2020 which accounted for 30% of road traffic crashes as published by the European Union was analysed by Baranyai and Sipos [8] using kernel method by implementing the black-spot technique. Additionally, owing to the extensive applicability of KDE, the technique is of great utilization in identification of region having elevated data density in clustering analysis as well as bump hunting with significant uses in pattern identification and image editing [9]. More so, KDE has gained popularity in anomalies detection or outliers especially in estimation of underlying data distribution, affirming KDE as a veritable tool in the detection of fraud as well as network security [10, 11]. The kernel approach is of great significance in estimation of income distribution, analysis of income inequality, and suitable policies formulation with recommendations [12]. The volatility of the stock market can be determined using the kernel technique in examining the frequent changes usually associated with the trading volumes in order to acknowledge the behaviour of the stock market. Fan and Cheng [13] investigated stock trading volumes with respect to market share and their probability distribution using KDE with the aid of long-term and short-term integration memory network strategies in modelling the dynamics of trading volume. The prevention of traffic accidents is critical to effective road management systems particularly, the identification of hotspots. In the study of Srikanth and Srikanth [14], hotspots in the city of Des Moines, United States, were identified with KDE for five consecutive years' data from 2008 to 2012. The identified hotspots were provided by a visual display of traffic crashes locations based on the statistical significance of their density estimates.

Heavy rainfall that is characterized by massive destructive winds and accompanied by snow is frequently experienced in Bangladesh which is located in Southern Asia due to the climatic irregularity and distinctive geographical position. The analysis of cyclone data was investigated by Tuya *et al.* [15] from 1960 to 2023 in Bangladesh using the KDE method by identifying areas that are prone to cyclones. The study furnished the government with invaluable knowledge to improve effective planning that mitigates the consequence of cyclones in the country. Lasocki [16] examined the application of KDE in seismology in which the magnitude was estimated as seismic variables with continuous distributions. Similarly, Pollard *et al.* [17] applied KDE in archaeological investigation of spatial analysis and the findings provided valuable insights into spatial archaeological results especially in Geographic Information Systems (GIS). Additionally, in clustering analysis, KDE approach is of great importance especially in the identification of areas having high density of data points. The KDE methodology has gain popularity in several fields of studies, such as road network, image editing, and pattern identification [18-20].

The extensive applications of the kernel approach are accompanied with numerous difficulties for effective implementation. The main problem associated with the KDE approach is the accurate selection of the bandwidth as well as the kernel function. The kernel estimator (KE) and kernel function (KF) are impacted by the choice of the bandwidth greatly however, with minimal impact on KF since most KFs are probability density function [21]. Despite the difficulties connected with accurate bandwidth selection, researchers are regularly introducing novel bandwidth selectors to improve KDE's performance [22, 23]. Moreover, there is a decrease of the applicability of kernel technique in multivariate environment due to the phenomenon known as "curse of dimensionality effect" usually connected with non-parametric statistical estimation [24, 25]. This condition has given rise to the application of sophisticated techniques capable of successfully addressing these challenges.

The introduction of modern methodologies into kernel approaches have largely contributed effectively to performance improvement especially in bandwidth selection, new kernel estimators, and extension of the kernel method in numerous disciplines [26]. Again, the introduction of modern computational techniques has addressed the problem of scalability in large datasets [27, 28]. The performance and efficiency of a newly introduced beta polynomial kernel is examined in this study. The justification for the introduction of new kernel family is hinge on the need for improvement in kernel performance as well as efficient estimation techniques. The new kernel family promises improved performance, that is reduction of estimation error using the asymptotic mean integrated squared error (AMISE) as criterion function, and better efficiency in density estimation. The improvement of performance and efficiency of the novel beta kernel family is ascertained in comparison with the classical beta family.

2. MATERIALS AND METHODS

This study employed strategy geared towards the exploration of the efficacy and effectiveness of a novel kernel family in statistical density estimation. The kernel techniques are versatile and widely recognized especially in modelling probability density functions that do not require stringent parametric assumptions. This study assesses the efficacy and efficiency of NBPF in KDE, demonstrating their adaptability to modern methodologies in probability density functions without the rigid parametric assumptions. The proposed kernel family demonstrate a propitious behaviour and adaptability feature when evaluated within the interval support of [−1, 1]. Typically, the KDE technique is of widespread uses in non-parametric estimation methods in modelling several events. Rosenblatt [29] proposed the kernel estimator, while Parzen [30] popularized its applications across multiple disciplines, particularly in exploratory data analysis and data visualization [2, 31]. Generally, the kernel's closed form is:

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

In this context, $K(\cdot)$ represent a kernel function, h_{y} denotes the bandwidth that controls the estimates' smoothness, n is the total observations, y refers to data variations and Y_i denotes variables to be estimated [12]. Certain conditions are usually satisfied by this estimator, which are:

$$\int K(\mathrm{y})d\mathrm{y} = 1, \quad \int \mathrm{y}K(\mathrm{y})d\mathrm{y} = 0 \quad \text{and} \quad \int \mathrm{y}^2K(\mathrm{y})d\mathrm{y} = \mu_2(K) \neq 0 \tag{2}$$

A kernel function is typically symmetric, integrates to one with zero mean, non-zero moment, and is a PDF (probability density function). The two fundamental components in KDE are the selection of bandwidth and the choice of the kernel function. In terms of bandwidth selection, numerous researchers have introduced innovative selectors, whereas there has been limited investigation into the kernel function [1, 32]. The selection of bandwidth significantly influences the efficacy of the KDE, and a globally utilized evaluation performance metric in KDE is the mean integrated squared error, which is:

$$MISE\left(\hat{f}(\mathrm{y})\right) = \frac{R(K)}{nh_{\mathrm{y}}} + \frac{1}{4}\mu_2(K)^2h_{\mathrm{y}}^4R(f'') + O\left(\frac{1}{nh_{\mathrm{y}}} + h_{\mathrm{y}}^4\right) \tag{3}$$

where $R(K)$ denotes kernel roughness, $\mu_2(K)^2$ depicts the kernel moment and $R(f'') = \int f''(\mathrm{y})^2 d\mathrm{y}$ represents unknown function roughness [33]. The MISE has an asymptotic version, which is:

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

The bandwidth with the minimum AMISE, as presented in Equation (4), is referred to as the optimal bandwidth and is given by:

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

The bandwidth value derived from Equation (5) yields the minimum AMISE. Accurate selection of the bandwidth is of great significance in KDE methodologies due to its smoothing role. Innovative techniques have been recently introduced to address the difficulties linked with accurate bandwidth selection [21, 34].

2.1. The Existing Beta Polynomial Kernel Functions

There are different kernel families in density estimation but the beta kernel family is frequently used due to its unique mathematical tractability. The compact form of the beta kernel is:

$$K_{[p,r]}(t) = \frac{(2r+1)!}{2^{2r+1}(r!)^2}(1-t^2)^r \tag{6}$$

Here $r = 0, 1, 2, \ldots, \infty$ is the input of the kernels, and *p* is the kernel order, which is two in this study [35]. The kernel orders typically refer to kernel moments, which are non-zero values, while the odd moments of kernels are equal to zero. Kernels that display negativity within their interval of support are typically not classified as second-order. The beta kernels are probability density functions that are assessed within the closed interval $-1 \leq t \leq 1$. Certain kernel functions, including Uniform, Epanechnikov, Biweight, and Triweight kernels, have corresponding values of r ranging from 0 to 3. As r approaches infinity, the resulting kernel converges to that of the Gaussian kernel [36]. The Epanechnikov kernel, characterized by $r = 1$, is considered optimal for the AMISE because it produced the least value which is:

$$K_{[2,1]}(t) = \frac{3}{4}(1-t^2) \tag{7}$$

Other members corresponding to $r = 2, 3, 4$ called Biweight, Triweight, and Quadriweight kernels, are:

$$K_{[2,2]}(t) = \frac{15}{16}(1-t^2)^2 \tag{8}$$

$$K_{[2,3]}(t) = \frac{35}{32}(1-t^2)^3 \tag{9}$$

$$K_{[2,4]}(t) = \frac{315}{256}(1-t^2)^4 \tag{10}$$

However; as r approaches infinity, the resultant is the normal kernel function that is not exactly a member of this family, and it is expressed as:

$$K_{[2,\mathbf{Normal}]}(t) = \frac{1}{\sqrt{2\pi}} exp\left(-\frac{t^2}{2}\right) \tag{11}$$

The commonly used kernels from the beta family in KDE include Equations (7) to (9), along with the normal distribution function, classified not firmly belonging this group. Other members derived from this family are rarely used because those with higher powers tend to be less efficient. The efficiency of these kernels is evaluated using the Epanechnikov kernel (EK), the first member of this family, which is regarded as optimal in reference to AMISE. Generally, the efficiency of kernel functions besides Equations (7) to (9) decline as the index of the generated function increases. The optimum efficiency for the function is typically equal to one, whereas other functions usually exhibit efficiencies that are less than one, with numerical values progressively decreasing.

2.2. The New Beta Polynomial Family

The technique for the derivation of the new beta polynomial family (NBPF) is based on the modification of the additive construction approach. The modified additive construction rule is:

$$K_{[p,r]}(t) = \frac{3}{2}K_{[p,r]}(t) + \frac{1}{2}tK'_{[p,r+1]}(t) \tag{12}$$

where p is the kernel order and r denotes the argument of the kernel. On letting $r = r+1$ in Equation (6), we have Equation (13) given as:

$$K_{[p,r+1]}(t) = \frac{(2r+3)!}{2^{2r+3}[(r+1)!]^2}(1-t^2)^{r+1} \tag{13}$$

Again, taking the derivative of Equation (13) with respect to t, yields Equation (14), which is:

$$K'_{[p,r+1]}(t) = \frac{(2r+3)!}{2^{2r+3}[(r+1)!]^2}(r+1)(1-t^2)^r(-2t) \tag{14}$$

Furthermore, substituting Equation (6) and Equation (14) into Equation (12), we have:

$$K_{[p,r]}(t) = \frac{3}{2}\left[\frac{(2r+1)!}{2^{2r+1}(r!)^2}(1-t^2)^r\right] + \frac{1}{2}t\left[\frac{(2r+3)!}{2^{2r+3}[(r+1)!]^2}(r+1)(1-t^2)^r(-2t)\right] \tag{15}$$

Further simplification of Equation (15) results in the new second-order kernel, which is:

$$K_{[2,r]}(t) = \frac{1}{2}\left\{\left(\frac{(2r+1)!}{2^{2r+1}(r!)^2}(1-t^2)^r\right)\{3-(3+2r)t^2\}\right\} \tag{16}$$

The initial members of the newly proposed second-order kernels are:

$$K_{[2,1]}(t) = \frac{3}{8}(3-5t^2)(1-t^2) \tag{17}$$

$$K_{[2,2]}(t) = \frac{15}{32}(3-7t^2)(1-t^2)^2 \tag{18}$$

$$K_{[2,3]}(t) = \frac{35}{64}(3-9t^2)(1-t^2)^3 \tag{19}$$

$$K_{[2,4]}(t) = \frac{315}{512}\,(3-11t^2)(1-t^2)^4 \quad (20)$$

Similarly, when r tends to infinity, the resulting normal kernel is:

$$K_{[2,\mathrm{Normal}]}(t) = \frac{1}{2\sqrt{2\pi}}\,(1-t^2)(3-t^2)exp\left(-\frac{t^2}{2}\right) \quad (21)$$

The bandwidth of both the traditional and proposed kernels, as well as the AMISE, demonstrates comparable orders. However, the proposed kernels exhibit remarkable qualities that distinguish them in performance. The proposed kernels meet the criteria for second-order kernels in Equation (2). Again, the advantage of the proposed method over its classical counterpart is the possession of additional powers that resulted in a high degree of differentiability, which significantly enhances their performance in terms of AMISE and efficiency. Polynomial kernels are of broader uses due to their ability of handling data with nonlinearity characteristics in data analysis [37]. In classification and regression methodologies, the kernel approached has been widely applied as noted by Liu and Xie [38], while Ngu *et al*. [39] presented a detailed assessment of the applications of kernel functions particularly with emphasis on regression and classification in the framework of sensors.

2.3. The Efficiency of Kernel Function

Generally, kernel efficiency is often assessed in comparison with the Epanechnikov function. The kernel efficiency is derived from the established relationship defined as follows:

$$Eff(K) = \left(\frac{C(K_e)}{C(K)}\right)^{1/4} = \left(\frac{R(K_e)^4\mu_2(K_e)^2}{R(K)^4\mu_2(K)^2}\right)^{1/4} \quad (22)$$

Here $C(K) = R(K)^4\mu_2(K)^2$ denotes other kernel constants while $C(K_e) = R(K_e)^4\mu_2(K_e)^2$ is Epanechnikov kernel constant [36]. Equation (22) shows that determining the efficiency of a kernel function requires two key statistics: kernel roughness and kernel moment. The kernel roughness is defined by:

$$R(K) = \int K(t)^2 dt \quad (23)$$

Again, the kernel moment, commonly known as variance, is:

$$\mu_2(K) = \int t^2 K(t)dt \quad (24)$$

Typically, the AMISE optimum kernel of the proposed family is also Epanechnikov kernel (EK), as it yields a minimal value as the conventional kernels.

3. RESULTS AND DISCUSSION

To evaluate the effectiveness of the new beta kernel family, the study employed both simulated and real-world data. Application of the NBPF to real-world data of the old faithful geyser is done for practical implementation [2]. The achievement of NBPF was compared to that of BKF using AMISE as a performance metric. This metric provides a quantitative basis for measuring the accuracy and efficiency of the new kernel function. The numerical computations and the graphical analysis were performed using Mathematica version 12.3 software. This study examines the statistical properties of Epanechnikov, Biweight, Triweight, and Quadriweight kernels because they serve as foundational kernel members for a comprehensive discussion, particularly the optimum kernel function, in the evaluation of efficiencies of other functions within this family of kernels. The efficiency of kernels can be used as an assessment of their performance.

The efficiency values were calculated for both the NBPF and BKF kernels. The efficiencies of the NBPF and BKF kernels are presented in Table 1 with the results obtained being compared with EK as the optimum kernel. NBPF displayed superior efficiencies over BKF kernel family comparatively, demonstrating their dominance and effectiveness over the traditional BKF family. The noticeable improvement in efficiencies of the proposed NBPF kernels is ascribed to the degree of the constant of normalization as well as the functional modification incorporated into the NBPF family. The coefficient of any beta kernel is typically the normalization constant which contributes significantly to the performance and numerical values of the efficiency. The optimum kernel is the Epanechnikov function, hence its efficiency value is always one while other members of the traditional BKF kernels exhibited numerical values that are less than one as their efficiencies. Table 1 illustrates that as the power of the kernel increases, the efficiency of the proposed NBPF kernels also rises, in contrast to the decreasing efficiency observed in conventional BKF kernels. The enhanced efficiency of the NBPF functions is linked to their larger normalization constants and the functional modifications, hence more effective than traditional BKF functions. Figures 1 and 2 are the graphical representations of the conventional BKF and newly introduced NBPF kernels, respectively. The plots of both NBPF and BKF were evaluated in a specified interval of $-1 \leq t \leq 1$. All beta polynomial kernels are typically analyzed within this interval, and statistical components such as kernel roughness and moment are numerically computed within this interval in the determination of bandwidths and kernel efficiency. The efficiency of the existing BKF is denoted by $Eff\,(K_{BKF})$ while the efficiency of NBPF is denoted by $Eff\,(K_{NBPK})$ in Table 1 respectively.

The AMISE of the simulated data of 6000 as the sample size for both BKF and NBPF is presented in Table 2. The results in Table 2, in terms of AMISE as a metric of performance, suggested that the NBPF kernels are more effective than BKF. An increase in kernels' power leads to graphs with narrower peaks, as shown in Figures 1 and 2, and both plots illustrate the characteristics of the conventional BKF and proposed NBPF kernel functions regarding peakedness. The estimates derived from the traditional BKF Epanechnikov function and Biweight kernel are classified as platykurtic, as they exhibit a wider peak. In contrast, the newly introduced NBPF exhibit leptokurtic behaviour, especially at higher powers with more acute peaks. The level of differentiability of the kernel serves as the primary factor in determining the peakedness of the kernel, as illustrated in graphs featuring higher-degree polynomials in Figure 2 of the NBPF. Kernel functions characterized by higher degrees possess more derivatives and substantial normalization constants, which typically result in graphs that exhibit acute peaks with enhanced quality kernel estimates.

One of the main applications of kernel density estimation is the exploration of obscured or hidden features in a dataset [40]. Statistical properties such as skewness and modes in a dataset are easily revealed by KDE while in some cases, KDE do provide invaluable insights that would give essential information for additional investigation. The identification of the underlying statistical properties of a dataset through a smooth curve of the data is a fundamental role of KDE, and this is vital in understanding the intrinsic nature of the observations, thereby providing information for making decisions. This study investigated the Old Faithful Geyser data which comprises 107 observations with emphasis on the durations of

Table 1. Traditional BKF efficiencies and proposed NBPF efficiencies.

Kernel Functions $K(t)$	**Traditional BKF**			**Proposed NBPK**		
	$R(K)$	$\mu_2(K)$	$Eff(K_{BKF})$	$R(K)$	$\mu_2(K)$	$Eff(K_{NBPK})$
Epanechnikov	$\frac{3}{5}$	$\frac{1}{5}$	1.000	$\frac{31}{35}$	$\frac{3}{35}$	1.000
Biweight	$\frac{5}{7}$	$\frac{1}{7}$	0.994	$\frac{1110}{1001}$	$\frac{1}{21}$	1.072
Triweight	$\frac{350}{429}$	$\frac{1}{9}$	0.987	$\frac{3150}{2450}$	$\frac{1}{33}$	1.159
Quadriweight	$\frac{2205}{2431}$	$\frac{1}{11}$	0.981	$\frac{67410}{46189}$	$\frac{3}{143}$	1.227

Table 2. AMISE and efficiencies of BKF with NBPF using a sample size of 6000.

Kernel Functions	**AMISE of BKF**	$Eff(K_{BKF})$	**AMISE of NBPF**	$Eff(K_{NBPK})$
Epanechnikov	0.000559711	1.000	0.000264222	1.000
Biweight	0.000776042	0.994	0.000294581	1.072
Triweight	0.000852696	0.987	0.000323289	1.159
Quadriweight	0.000952354	0.981	0.000353209	1.227

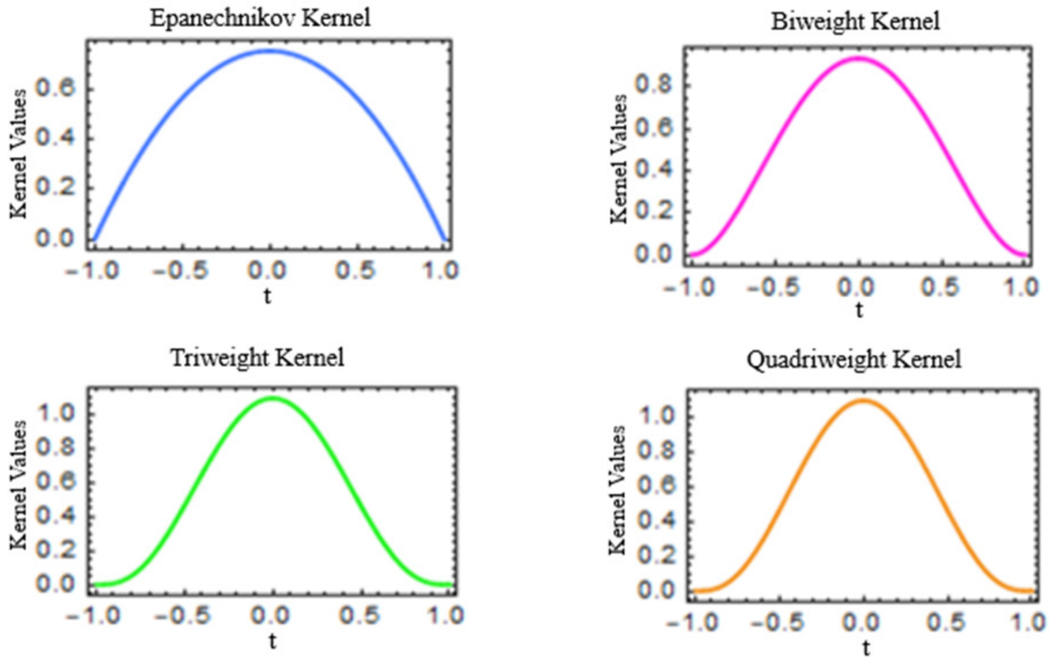

Fig. 1. The plots of traditional (BKF) second order beta kernel family.

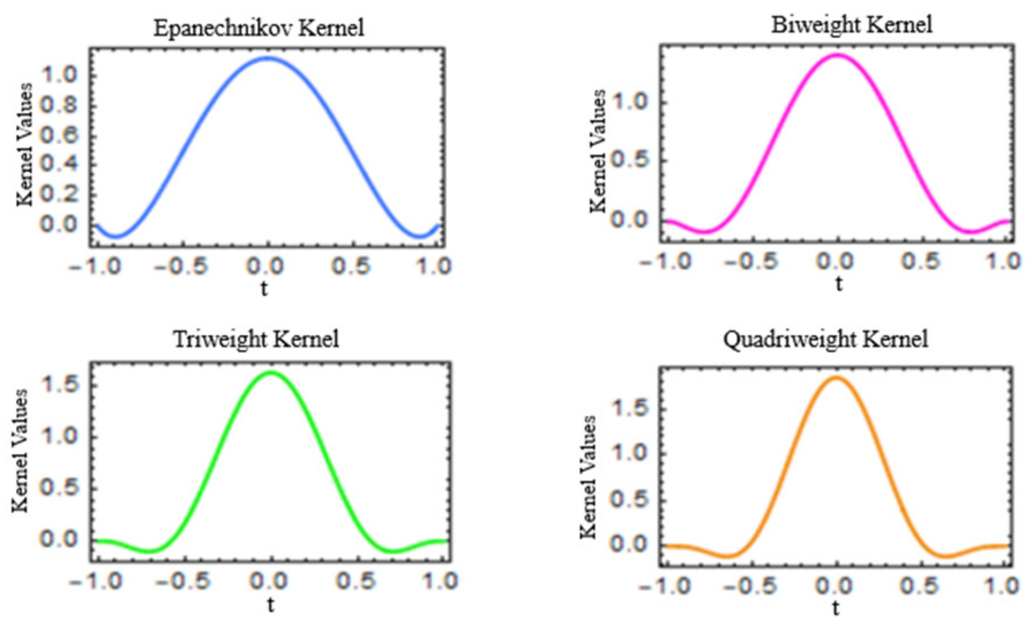

Fig. 2. The plots of proposed (NBPF) second order beta kernel family.

eruption in minutes [2]. The kernel estimates of the real data as seen in both the BKF and NBPF demonstrate bimodality as presented in Figures 3 and 4 by Epanechnikov kernel (EK) and Biweight kernel (BK), respectively. The results reveal that there are two unique clusters in times of eruption with focus at 1.9 minutes and 4.5 minutes, respectively. This bimodal characteristic implies that the duration of the eruption can be influenced by several indispensable factors.

The visualization of the data using the KDE approach has established its usefulness in capturing features in the data with the bimodality being evidenced. The bandwidth employed in constructing the kernel estimates of the BKF and NBPF in Figures 3 and 4 is obtained by standardization of the data, and the bandwidth is 0.437. The bimodality of the real data as evidenced in Figures 3 and 4 which are estimates of BKF and NBPF is a validation of the proposed NBPF functions in preservation of

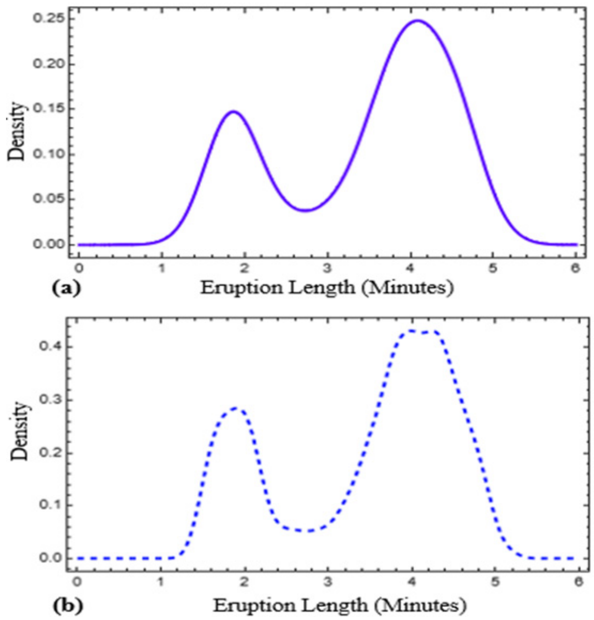

Fig. 3. The plots of real data of EK **(a)** BKF and **(b)** NBPF.

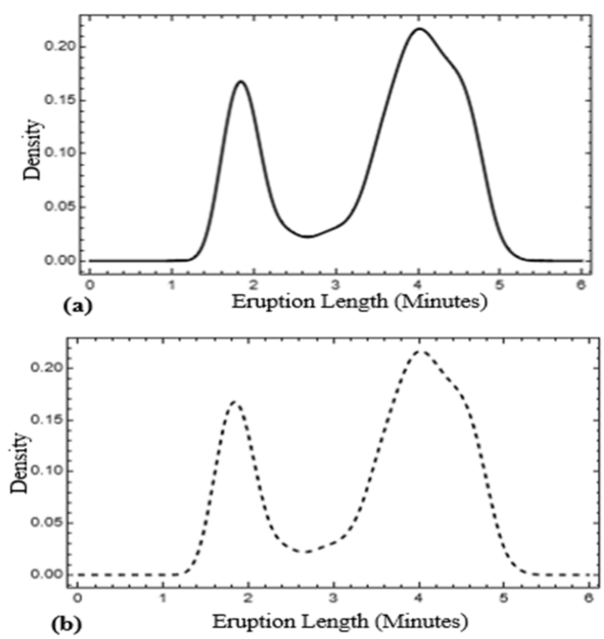

Fig. 4. The plots of real data of BK **(a)** and BKF **(b)** NBPF.

vital statistical characteristics. The estimates of the proposed NBPF of Epanechnikov and Biweight kernels exhibit similarity but the traditional BKF Biweight estimate display a sharp presentation of the modes. The retention of inherent features of a dataset is essentially based on the appropriate selection of the bandwidth as well as the kernel function. The vital responsibility of revealing inherent statistical features in a dataset has been corroborated by the proposed NBPF functions.

4. CONCLUSIONS

The proposed kernel family are versatile in density estimation, effective in data analysis as authenticated by the AMISE, and superior efficiency values. The versatility of the proposed kernels is particularly noticeable in the AMISE as well as the efficiencies. The incorporation of the functional form gives high level of smoothness that guaranteed their wide applicability across the spectrum of many statistical fields. Selection of kernel function typically depends on level of smoothness of its estimates, and the degree of differentiability of the function. Kernel functions with higher degree usually generate smoother estimates because they can be differentiated continuously, while noisier estimates are often produced by kernels with lower degrees of differentiability. The outcomes of the research show that NBPFs outplay BKFs with reference to AMISE and efficiency. The improvement in performance of both AMISE and efficiency of the new kernel family is attributed to the magnitude of the coefficients of the kernel functions and the structural modification of the functional form.

5. CONFLICTS OF INTEREST

The authors declare no conflict of interest.

6. ETHICAL STATEMENT

This work does not include any studies involving human or animal subjects.

7. FUNDING

No funding was received for this work.

8. AUTHORSHIP CONTRIBUTION

All authors equally contributed in this study.

9. DECLARATION

All authors declare the originality of results and copyright are assigned to the Pakistan Academy of Sciences.

10. REFERENCES

1. S.J. Sheather. Density estimation. *Statistical Science* 19(4): 588-597 (2004). https://www.jstor.org/stable/4144429

2. B.W. Silverman (Ed.). Density estimation for statistics and data analysis. *Routledge, NY, USA* (2018). https://doi.org/10.1201/9781315140919
3. I.U. Siloko, K.E. Ukhurebor, E.A. Siloko, E. Enoyoze, A.O. Bobadoye, C.C. Ishiekwene, O.O. Uddin, and W. Nwankwo. Effects of some meteorological variables on cassava production in Edo State, Nigeria via density estimation. *Scientific Africa* 13: e00852 (2019). doi:10.1016/j.sciaf.2021.e00852
4. P. Janssen, J. Swanepoel, and N. Veraverbeke. A note on the behaviour of a kernel-smoothed kernel density estimator. *Statistics and Probability Letters* 158: 108663 (2019). doi:10.1016/j.spl.2019.108663
5. B. Bèranger, T. Duong, S.E. Perkins-Kirkpatrick, and S.A. Sisson. Tail density estimation for exploratory data analysis using kernel methods. *Journal of Nonparametric Statistics* 31(1): 144-174 (2018). doi:10.1080/10485252.2018.1537442
6. I.U. Siloko and O.O. Uddin. A statistical study of wind speed and its connectivity with relative humidity and temperature in Ughelli, Delta State, Nigeria. *Science World Journal* 18(3): 404-413 (2023). https://dx.doi.org/10.4314/swj.v18i3.13
7. J.C.Y. Ngu, W.S. Yeo, T.F. Thien, and J. Nandong. A comprehensive overview of the applications of kernel functions and data-driven models in regression and classification tasks in the context of software sensors. *Applied Soft Computing Journal* 164: 111975 (2024). doi:10.1016/j.asoc.2024.111975
8. D. Baranyai and T. Sipos. Black-spot analysis in Hungary based on kernel density estimation. *Sustainability* 14(14): 8335 (2022). doi:10.3390/su14148335
9. J.E. Chacón and T. Duong. Data-driven density derivative estimation, with applications to nonparametric clustering and bump hunting. *Electronic Journal of Statistics* 7: 499-532 (2013). doi:10.1214/13-EJS781
10. J. Rosenberger, J. Müller, A. Selig, M. Bühren, and D. Schramm. Extended kernel density estimation for anomaly detection in streaming data. *15th CIRP Conference on Intelligent Computation in Manufacturing Engineering, Gulf of Naples, Italy, Procedia CIRP* 112: 156-161 (2022). doi:10.1016/j.procir.2022.09.065
11. G. Aarthi, S. Sharon-Priya, and A.W. Banu. KRF-AD: Innovating anomaly detection with KDE-KL and random forest fusion. *Intelligent Decision Technologies* 18(3): 2275-2287 (2024). doi:10.3233/IDT-240628
12. S. Wang, A. Li, K. Wen, and X. Wu. Robust kernels for kernel density estimation. *Economics Letters* 191: 109138 (2020). doi:10.1016/j.econlet.2020.109138
13. Y. Fan and X. Cheng. Application of kernel density estimation in stock trading volume distribution. *Proceedings of Artificial Intelligence of Things and Computing, Hangzhou Jiangsu, China* pp. 173-180 (2024). doi:10.1145/3708282.3708314
14. L. Srikanth and I. Srikanth. A case study on kernel density estimation and hotspot analysis methods in traffic safety management. 2020 International Conference on Communication Systems & Networks (COMSNETS). *Bengaluru, India* pp. 99-104 (2020). doi:10.1109/COMSNETS48256.2020.9027448
15. J.H. Tuya, U.M. Coalee, and R. Ahmed. Potentiality of kernel density estimation tool for tropical cyclone frequency analysis in Bangladesh. *The Jahangirnagar Review: Part II: Social Science* 48(1): 87-105 (2024).
16. S. Lasocki. Kernel density estimation in seismology. In: Statistical Methods and Modeling of Seismogenesis. 1st Edition. N. Limnios, E. Papadimitriou, and G. Tsaklidis (Eds.). *Wiley Online Library, New Jersey, United States* pp. 1-25 (2021). doi:10.1002/9781119825050.ch1
17. A.M. Pollard, Q. Ma, A.I. Bidegaray, and R. Liu. The use of kernel density estimates on chemical and isotopic data in Archaeology. In: Handbook of Archaeological Sciences. 2nd Edition. A.M. Pollard, R.A. Armitage, and C.A. Makarewicz (Eds.). *John Wiley & Sons, New Jersey, United States* pp. 1227 (2023). doi:10.1002/9781119592112.ch61
18. S. Hashimoto, S. Yoshiki, R. Saeki, Y. Mimura, R. Ando, and S. Nanba. Development and application of traffic accident density estimation models using kernel density estimation. *Journal of Traffic and Transportation Engineering* 3(3): 262-270 (2016). https://doi.org/10.1016/j.jtte.2016.01.005
19. M. Govorov, G. Beconyte, and G. Gienko. Trivariate kernel density estimation of spatiotemporal crime events with case study for Lithuania. *Sustainability* 15(11): 8524 (2023). doi:10.3390/su15118524
20. A.C. Neto, A.L.M. Levada, and M.F.C. Haddad. A new kernel density estimation-based entropic isometric feature mapping for unsupervised metric learning. *Annals of Data Science* 12(3): 929-945 (2024). doi:10.1007/s40745-024-00548-x
21. Y. Tsuruta and M. Sagae. Theoretical properties of bandwidth selectors for kernel density estimation

on the circle. *Annals of the Institute of Statistical Mathematics* 72: 511-530 (2020). doi:10.1007/s10463-018-0701-x

22. I.U. Siloko, O. Ikpotokin, F.O. Oyegue, C.C. Ishiekwene, 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). doi.org/10.1080/09720510.2018.1524956
23. I.U. Siloko, E.A. Siloko, S.A. Ojobor, S.A. Wasan, E. Enoyoze, C.C. Ishiekwene, O. Ikpotokin, and E.M. Ogbeide. A new multivariate product kernel function of the beta polynomial family. *Journal of Statistics Applications and Probability* 12(3): 1385-1398 (2023). http://dx.doi.org/10.18576/jsap/120340
24. M.P. Wand and M.C. Jones (Eds.). Kernel Smoothing. *Chapman and Hall, London, UK* (1995). http://dx.doi.org/10.1007/978-1-4899-4493-1
25. D.W. Scott (Ed.). Multivariate density estimation. Theory, practice, and visualization. 2nd Edition. *Wiley, New Jersey*, *USA* (2015). https://doi.org/10.1002/9781118575574
26. S.M. Somé and C.C. Kokonendji. Bayesian selector of adaptive bandwidth for multivariate gamma kernel estimator on $[0, \infty)^d$. *Journal of Applied Statistics* 49(7): 1692-1713 (2022). doi:10.1080/02664763.2021.1881456
27. S. Zámečník, I. Horová, S. Katina, and K. Hasilová. An adaptive method for bandwidth selection in circular kernel density estimation. *Computational Statistics* 39: 1709-1728 (2023). https://doi.org/10.1007/s00180-023-01401-0
28. I. Fuentes-Santos, W. González-Manteiga, and J. Mateu. Bootstrap bandwidth selection for the pair correlation function of inhomogeneous spatial point processes. *Journal of Statistical Computation and Simulation* 93(18): 3329-3361 (2023). https://doi.org/10.1080/00949655.2023.2220860
29. M. Rosenblatt. Remarks on some nonparametric estimates of a density function. *Annals of Mathematical Statistics* 27: 832-837 (1956). http://dx.doi.org/10.1214/aoms/1177728190
30. E. Parzen. On the estimation of a probability density function and the mode. *Annals of Mathematical Statistics* 33: 1065-1076 (1962). http://dx.doi.org/10.1214/aoms/1177704472
31. I.U. Siloko and E.A. Siloko. An investigation on interdependence between rainfall and temperature in Ekpoma, Edo State, Nigeria. *Iranian (Iranica) Journal of Energy and Environment* 14 (3): 197-204 (2023). https://doi.org/10.5829/ijee.2023.14.03.01
32. I.U. Siloko, C.C. Ishiekwene, and F.O. Oyegue. New gradient methods for bandwidth selection in bivariate kernel density estimation. *Journal of Mathematics and Statistics* 6(1): 1-8 (2018). http://dx.doi.org/10.13189/ms.2018.060101
33. I.U. Siloko, O. Ikpotokin, and 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). doi:10.18187/pjsor.v15i3.2625
34. N. Gündüz and S. Karakoç. Optimal bandwidth selection methods with application to wind speed distribution. *Mathematics* 11(21): 4478 (2023). doi:10.3390/math11214478
35. T. Duong. Spherically symmetric multivariate beta family kernels. *Statistics and Probability Letters* 104: 141-145 (2015). http://dx.doi.org/10.1016/j.spl.2015.05.012
36. I.U. Siloko, O. Ikpotokin, F.O. Oyegue, E.A. Siloko, and C.C. Ishiekwene. Numerical computation of efficiency of beta polynomial kernels using product method. *Journal of Applied Science and Technology* 23 (1&2): 32-38 (2019). https://www.ajol.info/index.php/jast/article/view/192584
37. K. Haddad and A. Rahman. Regional flood frequency analysis: evaluation of regions in cluster space using support vector regression. *Natural Hazards* 102: 489-517 (2020). https://doi.org/10.1007/s11069-020-03935-8
38. Y. Liu and M. Xie. Rebooting data-driven soft-sensors in process industries: a review of kernel methods. *Journal of Process Control* 89: 58-73 (2020). https://doi.org/10.1016/j.jprocont.2020.03.012
39. J.C.Y. Ngu, W.S. Yeo, T.F. Thien, and J. Nandong. A comprehensive overview of the applications of kernel functions and data-driven models in regression and classification tasks in the context of software sensors. *Applied Soft Computing Journal* 164: 111975 (2024). https://doi.org/10.1016/j.asoc.2024.111975
40. I.U. Siloko, K.E. Ukhurebor, and A.I. Otsupius. Non-parametric statistical approach for assessing the environmental impacts of petroleum spillages in the Niger Delta region of Nigeria. *Discover Environment* 3: 173 (2025). doi.org/10.1007/s44274-025-00273-z