

A new family of kernels from the beta polynomial kernels with applications in density estimation



Israel Uzuazor Siloko ^{a,1,*}, Wilson Nwankwo ^{b,2}, Edith Akpevwe Siloko ^{c,3}

^a Department of Mathematics and Computer Science, Edo University, Iyamho, P.M.B. 04 Auchi, Nigeria

^b Department of Mathematics and Computer Science, Edo University, Iyamho, P.M.B. 04 Auchi, Nigeria

^c Department of Statistics, University of Benin, P.M.B. 1154 Benin City, Nigeria

¹ siloko.israel@edouniversity.edu.ng; ² drswilson@live.com; ³ myakpeski@yahoo.com

* corresponding author

ARTICLE INFO

Article history

Received January 21, 2020

Revised June 22, 2020

Accepted July 22, 2020

Available online November 12, 2020

Keywords

Smoothing parameter

Kernel function

Beta polynomial

AMISE

ABSTRACT

One of the fundamental data analytics tools in statistical estimation is the non-parametric kernel method that involves probability estimates production. The method uses the observations to obtain useful statistical information to aid the practicing statistician in decision making and further statistical investigations. The kernel techniques primarily examine essential characteristics in a data set, and this research aims to introduce new kernel functions that can easily detect inherent properties in any given observations. However, accurate application of kernel estimator as data analytics apparatus requires the kernel function and smoothing parameter that regulates the level of smoothness applied to the estimates. A plethora of kernel functions of different families and smoothing parameter selectors exist in the literature, but no one method is universally acceptable in all situations. Hence, more kernel functions with smoothing parameter selectors have been propounded customarily in density estimation. This article proposes a distinct kernel family from the beta polynomial kernel family using the exponential progression in its derivation. The newly proposed kernel family was evaluated with simulated and life data. The outcomes clearly indicated that this kernel family could compete favorably well with other kernel families in density estimation. A further comparison of numerical results of the new family and the existing beta family revealed that the new family outperformed the classical beta kernel family with simulation and real data examples with the aid of asymptotic mean integrated squared error (AMISE) as criterion function. The information obtained from the data analysis of this research could be used for decision making in an organization, especially when human and material resources are to be considered. In addition, Kernel functions are vital tools for data analysis and data visualization; hence the newly proposed functions are vital exploratory tools.



This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



1. Introduction

The probability density estimation concept in statistics has far-reaching effects in other fields of studies because most actions can be expressed in numerical form, which must be analyzed to avoid misleading information. Density estimation, which is the foundation of data analysis, involves constructing a probability density estimate from given observations. Density estimation is a fundamental concept in statistics primarily for data smoothing; analysis and virtualizations of observations [1] [2]. Data smoothing techniques usually consider findings such that inferences and conclusions regarding the observations regarding a particular estimation method can be made [3] [4]. Generally, density estimation

is viewed from two main perspectives; the parametric and non-parametric perspectives, but the semiparametric approach combines the two known approaches. Parametric density estimation assumes the observations to be estimated are from a known family and the only information required are the parameters of such distribution. The estimated parameters with the family of the distribution will produce a parametric estimator, and one of such estimators is the maximum likelihood estimators. The non-parametric estimation does not require prior knowledge of the observations' distribution, but they are subjected to "self-explanation" using some known statistical tools. Unlike the parametric approach that is with fixed structure, the non-parametric density estimation are flexible, however, their flexibility has resulted in high computational cost, which restricted their widespread applications. High computational costs are mainly encountered in analysing a large volume of data, especially with complex statistical models [5].

The non-parametric data estimation techniques as analytics tools will be employed in this paper. Non-parametric estimation techniques are of numerous uses and are gaining popularity in data analysis, particularly in statistics and other related fields of studies, because accurate information about data such as historical data is not readily available. The statistical information obtained using non-parametric estimation helps provide a complete understanding of the observations' underlying properties and features. There are varieties of non-parametric methods in density estimation, but one of the most widely employed non-parametric estimators is the kernel estimator [6]. The kernel estimators are non-parametric techniques in density estimation for data smoothing using the kernel function and a regulating factor known as the smoothing parameter. The kernel estimators are popular in density estimation than other non-parametric estimators due to the simplicity of its implementation and presentation of results using graphical approaches. In semiparametric estimation, the kernel estimators are regarded as the bedrock in their estimation process; hence, kernel estimators are known as the building blocks of semiparametric density estimation [7].

The kernel estimators are elegant density estimation tools for exploring and visualizing observations, which are often presented graphically. As a result of kernel estimators' statistical importance in data analysis and visualization, the kernel estimators have been the most studied estimators amongst the various non-parametric estimators [6][8]. The kernel method directly explores and virtualizes data with indirect applications in classification and other estimation processes. Some recent kernel estimation applications are in progressive censoring, which is fundamental in industry-related research for estimation of hazard rate [9][10]. One of the major advantages of the kernel estimators over other non-parametric estimators is their flexibility in modeling observations, and kernels are not affected by bias specification [11]. In machine learning, kernel estimators are also fundamental and have been studied extensively. The application of kernel estimators in supervised learning has improved learning knowledge, especially in nonlinear hypothesis testing. The knowledge of kernel density estimation has also be extended to deep learning and with different applications. The kernel estimator application to spectral distribution was recently investigated through implicit kernel learning models using deep neural networks in which the training and inferences were illustrated with Fourier features through random sampling [12].

Non-parametric estimators, mostly the kernel estimators, are of excellent suitability for unsupervised learning. Kernel estimators have addressed decentralized classification and clustering analysis problems and also extended to the distributed estimation method, which is the bedrock for numerous distributed systems [13][14]. The knowledge of the underlying model in distributed data is fundamental when studying distributed systems. A proposed method is known as the "gossip-based distributed kernel density estimation technique" was introduced by Li *et al.* [15], and its convergence properties also analyzed with the results revealing accurate underlying density distribution of the distributed observations.

In statistics and machine learning, kernel estimators are vital and versatile tools in the estimation of observations. Despite the modern methods of data estimation and numerous kernel functions in literature, new kernel functions are still introduced due to the great influence of kernel function when

evaluating its performance empirically [16]-[20]. This paper introduces a new kernel family of the beta polynomial kernel family, and the results of both families revealed that the modified kernel functions outperformed the current beta polynomial family with AMISE as a performance measure.

2. Method

2.1. The kernel density estimator

Rosenblatt [21] and Parzen [22] initiated the data analysis method mainly for exploratory and visualization of observations. The kernel estimator is popular amongst non-parametric density functions due to its simplicity and computational advantages. There are two basic fundamental concepts in density estimation using the kernel estimator, and the abstractions are the kernel function and the smoothing factor, also called the smoothness coefficient. As a standardized weighting function, its one-dimensional form is

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

with $K(\cdot)$ being kernel function, n represent the sample size, $h > 0$ is bandwidth (smoothing parameter), x is the range of the observations and X_i is a set of observations. The kernel function $K(\cdot)$ is usually symmetrical and unimodal with the axioms in Equation (2). The Equation shows three conditions: unity integrant in every probability density function, the kernel functions average is zero, and the variance of any kernel is not equal to zero [2][23].

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

2.2. The beta polynomial kernel function

The beta polynomial kernel family is amongst the popular classes of estimators in data estimation. This class of estimator is given as

$$K_p(t) = \frac{(2p+1)!}{2^{2p+1}(p!)^2} (1-t^2)^p \quad (3)$$

where $p = 0, 1, 2, \dots, \infty$ is called the polynomial power while the variable t assumes value in the interval $-1 \leq t \leq 1$. The beta polynomial kernel functions are probability density functions since all kernel functions are probability density function and are usually supported within an interval [4][24]. The range of evaluating this class of kernel functions is $[-1, 1]$. Different values of p will produce different kernel functions, but as p tends to infinity, the resulting kernel is the popular Gaussian kernel function whose support is unbounded and is not strictly a member of this family of kernels [25][26]. The uniform kernel is the primordial of the class, that is when $p = 0$ whereas for $p = 1$ will produce the Epanechnikov kernel known to be the optimal kernel with regards to the AMISE. The Epanechnikov kernel, which is regarded as the optimal kernel, is expressed as

$$K_1(t) = \frac{3}{4} (1-t^2) \quad (4)$$

The kernel functions that correspond to $p = 2, 3$ and 4 from the general form in Equation (3) called the Biweight, Triweight, and Quadriweight kernels are given as

$$K_2(t) = \frac{15}{16} (1-t^2)^2 \quad (5)$$

$$K_3(t) = \frac{35}{32} (1-t^2)^3 \quad (6)$$

$$K_4(t) = \frac{315}{256} (1-t^2)^4 \quad (7)$$

Again, when the value of p goes to infinity, the kernel obtained is the popular Normal kernel or Gaussian kernel given by

$$K_{\theta}(t) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right) \quad (8)$$

The Epanechnikov, Biweight, Triweight, and Quadriweight kernel functions, which are the first four beta polynomial kernels excluding the Uniform kernel, have broad statistical applications. The efficiency of other kernel functions of this family is computed using the Epanechnikov kernel due to its optimality property regarding the AMISE.

2.3. Evaluation of kernel performance

In non-parametric density estimation, there is usually an error criterion function that measures its productivity. There is numerous performance evaluation in kernel estimation, but efforts will be targeted at the mean integrated squared error (MISE). Other performance evaluation measures such as the integrated absolute error, hellinger distance, the likelihood criterion function, and kullback-liebler distance exist. However, the mean integrated squared error has gained popularity over other measures due to its inclusion of dimensionality in the expression while other measures are dimensionless. There is the exact and approximate MISE obtained either by convolution or Taylors' series approximation. The asymptotic MISE has the variance and the bias components which is

$$AMISE(\hat{f}(x)) = \frac{R(K)}{nh} + \frac{1}{4}\mu_2(K)^2 h^4 R(f'') \quad (9)$$

with $R(K)$ representing the roughness of the kernel, $\mu_2(K)^2$ is the variance of kernel while $R(f'') = \int f''(x)^2 dx$ is the roughness of unknown probability distribution. The size of the smoothing parameter usually regulates the contributions of the two components to the AMISE. The bias can be reduced while the variance increases and vice versa due to the variation in the size of the smoothing parameter [4][27]. The smoothing parameter with the minimum AMISE called the optimal smoothing parameter is the solution to the differential equation

$$\frac{\partial}{\partial h} AMISE(h) = \frac{-R(K)}{nh^2} + \mu_2(K)^2 h^3 R(f'') = 0 \quad (10)$$

The solution of Equation (10) will give the smoothing parameter with the minimum AMISE value known as the optimal smoothing parameter given as

$$h_{AMISE} = \left[\frac{R(K)}{\mu_2(K)^2 R(f'')} \right]^{1/5} \times n^{-1/5} \quad (11)$$

The smoothing parameter with the minimum AMISE value is of order $O(n^{-1/(d+4)})$ while the AMISE is of order $O(n^{-4/(d+4)})$ with d representing the dimension of the kernel. In probability distributions that are unimodal and slightly skewed, the unanimity in terms of the asymptotic mean integrated squared error is more evident [28].

2.4. The proposed beta polynomial kernel functions

The fundamental concepts in kernel density estimation are kernel function and the smoothness factor or bandwidth. Some research has been geared towards these two concepts, but there is no universally accepted method in all situations; hence new methods are usually initiated [29]. The proposed kernel functions from the polynomial beta family use an exponential progression, where there is a constant common ratio to all the polynomial functions. The ratio of two consecutive terms in an exponential progression is always a constant, and the constant is called the common ratio.

Recall the general form of the beta polynomial kernels given in Equation (3) and the terms of the polynomial functions for $p = 1, 2, 3$ and 4 , which are Epanechnikov, Biweight, Triweight, and Quadriweight kernels, respectively. Let the first term of the sequence and the common ratio be denoted

by a and r , respectively. If $K_p(t)$ is the p th term of the sequence, then the constant common ratio designated by r can be written as

$$r = \frac{K_p(t)}{K_{p-1}(t)} \quad (12)$$

where $K_{p-1}(t)$ is the $(p - 1)$ th term. From Equation (4), the first term of the sequence is

$$a = K_1(t) = \frac{3}{4} (1 - t^2) \quad (13)$$

The common ratio in Equation (12) can be obtained as

$$r = \frac{K_2(t)}{K_1(t)} = \frac{\frac{15}{16}(1-t^2)^2}{\frac{3}{4}(1-t^2)} = \frac{15(1-t^2)^2}{16} \times \frac{4}{3(1-t^2)} = \frac{5}{4} (1 - t^2) \quad (14)$$

Alternatively, the common ratio can be expressed as

$$r = \frac{K_3(t)}{K_2(t)} = \frac{\frac{35}{32}(1-t^2)^3}{\frac{15}{16}(1-t^2)^2} = \frac{35(1-t^2)^3}{32} \times \frac{16}{15(1-t^2)^2} = \frac{7}{6} (1 - t^2) \quad (15)$$

Again, the common ratio can be obtained as

$$r = \frac{K_4(t)}{K_3(t)} = \frac{\frac{315}{256}(1-t^2)^4}{\frac{35}{32}(1-t^2)^3} = \frac{315(1-t^2)^4}{256} \times \frac{32}{35(1-t^2)^3} = \frac{9}{8} (1 - t^2) \quad (16)$$

The common ratios of Equations (14–16) can simply be presented as

$$r = \begin{cases} \frac{5}{4} (1 - t^2) \\ \frac{7}{6} (1 - t^2) \\ \frac{9}{8} (1 - t^2) \end{cases} \quad (17)$$

Hence; the generalised common ratio of two consecutive terms of Equation (17) is given as

$$r = \frac{3+2p}{2+2p} (1 - t^2) \quad (18)$$

where $p = 1, 2, 3, \dots$ is the power of the polynomial kernel.

2.5. The p th term of the proposed beta polynomial kernel functions

Let the first term of the exponential progression be a and the common ratio be r , respectively. Given that $K_p(t)$ is the p th term of the sequence then

$$K_1(t) = a$$

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

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

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

Therefore, the generalized p th kernel of the proposed beta polynomial kernels whose first term is Equation (13) and constant common ratio in Equation (18) is of the form

$$K_p(t) = ar^{p-1}, \quad p = 1, 2, 3, \dots \tag{19}$$

Hence the generalized *p*th kernels from Equation (19) of the proposed kernels can be written as

$$K_p(t) = \left(\frac{3}{4} (1 - t^2)\right) \left(\frac{3+2p}{2+2p} (1 - t^2)\right)^{p-1}, \quad p = 1, 2, 3, \dots \tag{20}$$

As in the classical beta polynomial kernels, when $p = 1$ the resulting kernel from Equation (20) is the Epanechnikov kernel in Equation (4), which is the optimum kernel of this family. However, when $p = 2, 3, 4, \dots$, the resulting new kernels of Biweight, Triweight, and Quadriweight kernel functions of the proposed kernels are as follows

$$K_2(t) = \frac{7}{8} (1 - t^2)^2 \tag{21}$$

$$K_3(t) = \frac{243}{256} (1 - t^2)^3 \tag{22}$$

$$K_4(t) = \frac{3993}{4000} (1 - t^2)^4 \tag{23}$$

The difference between the new kernel family and the classical polynomial functions is the value of the normalization constant while the two families' powers are the same. A change in the normalization constant produced corresponding changes in the value of the AMISE as the performance measure. Choices of kernel functions are based on their achievements, and a method or kernel function is better than the other when it produces a smaller value of the AMISE [30].

3. Results and Discussion

Investigation of the new polynomial kernels' statistical properties will be considered for the first four members of the family. All graphical and data analysis were implemented with Mathematica version 12 software platforms. The AMISE is the performance measure of the proposed kernel family and the classical polynomial kernels. The results on comparison reveal that the new family outperformed the classical beta kernel functions. The investigation was done for the first three family members since the Epanechnikov kernel maintained its optimality regarding the AMISE in both kernel families. Sample sizes of 2500 and 5000 were employed to illustrate the new beta polynomial kernels' performance since large sample sizes are most beneficial in non-parametric density estimation. The simulated results from the different sample sizes are presented in Table 1. The results revealed that the proposed beta kernel functions have smaller AMISE values than the classical kernel functions that show their superiority over existing kernel functions [31]. Fig. 1 and Fig. 2 are the graphs of the classical beta polynomial kernels and the proposed beta kernel functions.

Table 1. AMISE of Classical Kernels and Proposed Kernels of Sample sizes 2500 and 5000

Kernel Functions <i>Types of Kernels</i>	Classical Kernels AMISE		Proposed Kernels AMISE	
	<i>N=2500</i>	<i>N=5000</i>	<i>N=2500</i>	<i>N=5000</i>
Epanechnikov	0.00155571	0.00089352	0.00155571	0.00089352
Biweight	0.00230713	0.00132510	0.00198223	0.00113849
Triweight	0.00266570	0.00153104	0.00195163	0.00112092
Quadriweight	0.00309917	0.00178001	0.00195622	0.00112355

The real data set examined comprises 272 observations of the old faithful data [32]. The observations revealed bimodal features with the kernel estimates of the classical beta kernels and proposed kernel functions, and it supports the assertion that the duration of eruption often exhibits a bimodal distribution. Fig. 3 and Fig. 4 are the kernel estimates of the classical beta polynomial kernels, and the proposed beta kernels with the bimodality have been evident in both kernel functions. Table 2 shows

the performance of the various kernel functions using real data. The results also vividly show that the proposed kernel functions outperformed the classical beta kernel functions.

Table 2. Bandwidths and AMISE of classical and proposed kernels of old faithful data

Kernel Functions	Classical Kernel Functions		Proposed Kernel Functions	
Types of Kernels	Bandwidths	AMISE	Bandwidths	AMISE
Epanechnikov	0.300516	0.00917539	0.300516	0.00917539
Biweight	0.241237	0.01360722	0.244589	0.01169112
Triweight	0.238475	0.01572241	0.245332	0.01151051
Quadriweight	0.228046	0.01827862	0.237787	0.01153760

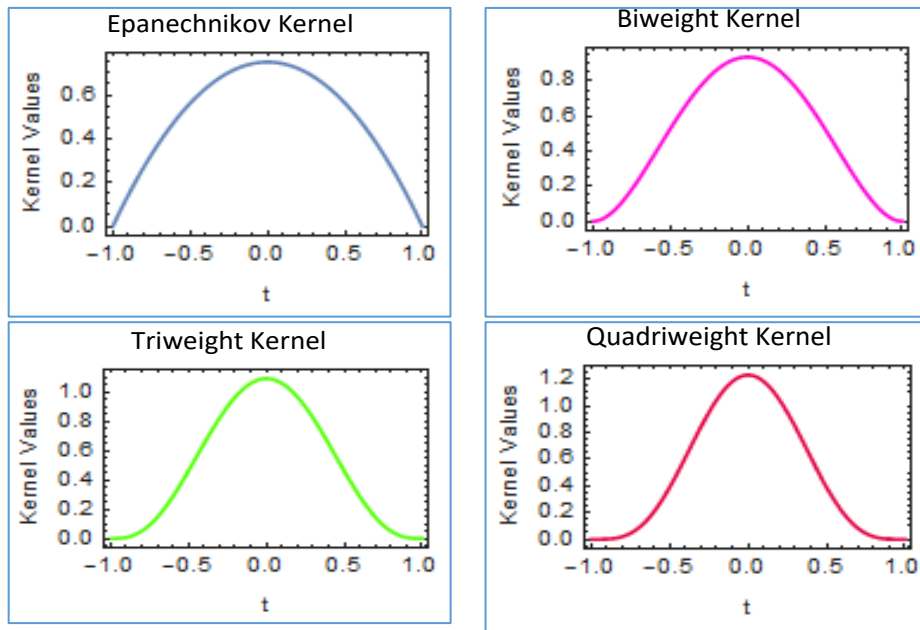


Fig. 1. Graphs of classical beta polynomial kernel functions

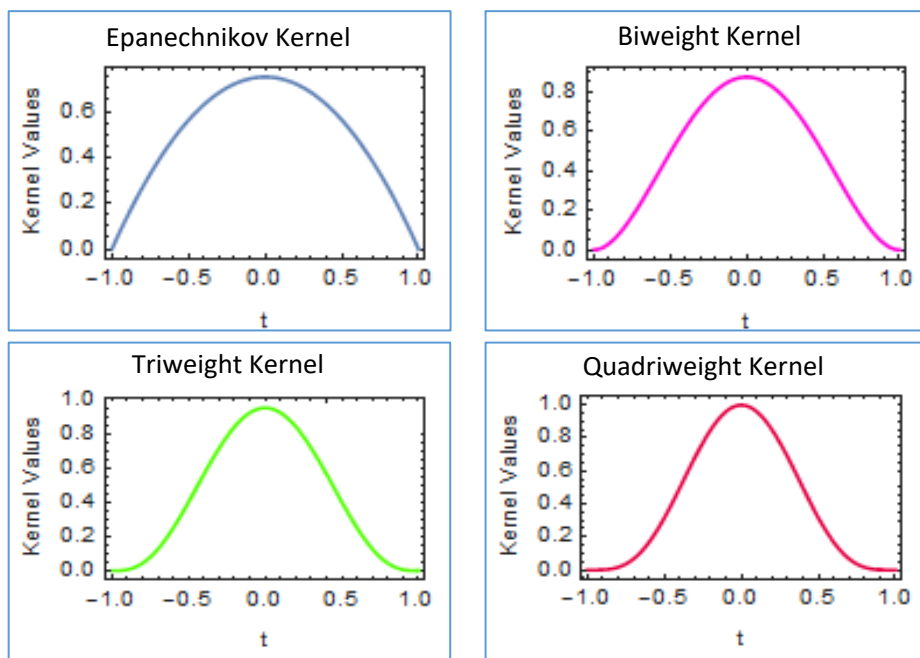


Fig. 2. Graphs of proposed beta polynomial kernel functions

The graphs of the classical beta polynomial and the proposed beta polynomial kernels are similar because they have the same powers but different normalization constants. Again, as earlier stated, the Epanechnikov kernel is the same for both kernels, which implies that its optimality position in the second-order kernel is maintained. However, in performance, as seen in Table 1, the proposed kernels did better than the classical implying that their kernel estimates will compete favorably well with the estimates of the classical kernels with real data examples. The value of the AMISE for the sample size of 5000 in Table 1 is smaller than the sample size of 2500, and this simply implies that larger sample sizes are more beneficial in the non-parametric estimation, particularly in kernel density estimation. It should be clearly pointed out that our comparison of performance started with the Biweight kernel since the Epanechnikov kernel is the same for the proposed and the classical kernels. However, the Epanechnikov kernel is included in the graphs and Tables in the analysis with simulated and real data.

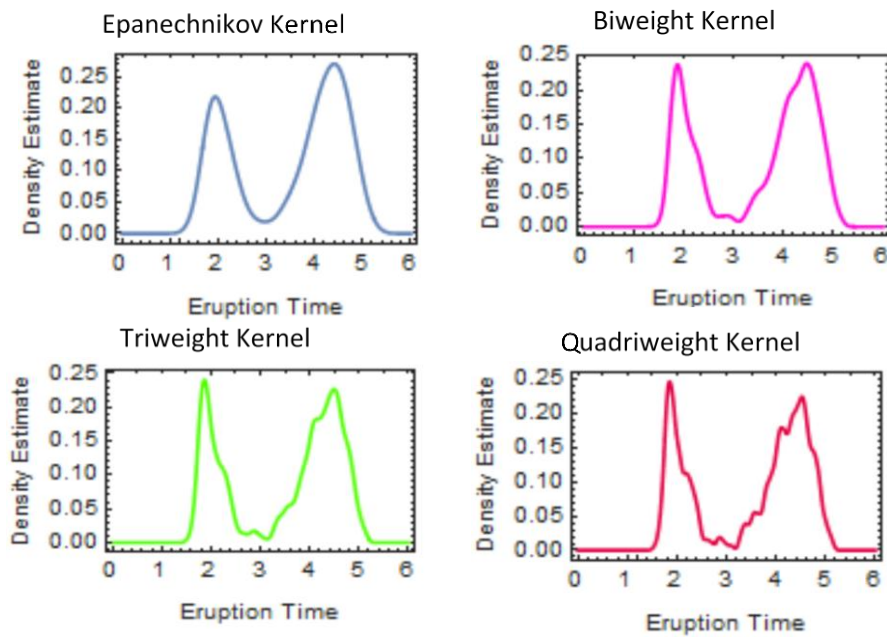


Fig. 3. Kernel estimated of classical beta polynomial functions using old faithful data

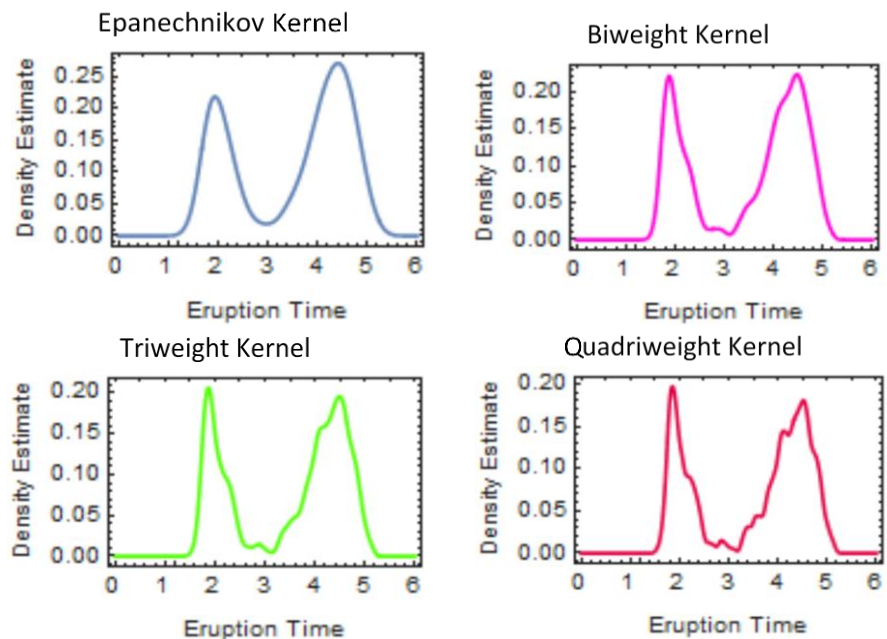


Fig. 4. Kernel estimates of proposed beta polynomial functions using old faithful data

The kernel estimates of the proposed beta polynomial with the old faithful data are similar to the classical beta polynomial kernel estimates, as seen in Fig. 3 and Fig. 4, respectively. In terms of retaining the data's inherent and essential features, the proposed kernel functions retained the data's bimodal feature just as the classical kernels. However, in performance evaluation using the AMISE criterion, the proposed beta polynomial kernels outperformed their classical counterpart because they produce a minimum value of the AMISE in all the cases considered, whether in simulations or real data applications.

4. Conclusion

This paper propounds new polynomial kernels from the classical beta polynomial family using the exponential progression approach. On evaluating the performance of the proposed beta polynomial kernels with the AMISE, the outcomes show that the proposed kernels performed better than the classical beta polynomial kernels. In terms of visualization of data, due to the huge information that kernel estimates provide in features highlighting for decision making, the proposed kernels compete favorably with the existing kernels to retain and preserve inherent statistical features in the observations examined. An extension of the proposed univariate kernel functions to the multi-dimensional kernel is an area of future research work since many kernel estimation applications are higher dimensional.

Acknowledgment

The authors appreciate the anonymous reviewers for painstakingly going through the manuscript and valuable comments and suggestions that have greatly improved the manuscripts' quality.

Declarations

Author contribution. All authors contributed equally to the main contributor to this paper. All authors read and approved the final paper.

Funding statement. None of the authors have received any funding or grants from any institution or funding body for the research.

Conflict of interest. The authors declare no conflict of interest.

Additional information. No additional information is available for this paper.

References

- [1] S. Węglarczyk, "Kernel density estimation and its application," *ITM Web Conf.*, vol. 23, p. 00037, Nov. 2018, doi: [10.1051/itmconf/20182300037](https://doi.org/10.1051/itmconf/20182300037).
- [2] B. W. Silverman, *Density Estimation for Statistics and Data Analysis*. Boston, MA: Springer US, 1986, doi: [10.1007/978-1-4899-3324-9](https://doi.org/10.1007/978-1-4899-3324-9).
- [3] I. U. Siloko, O. Ikpotokin, F. O. Oyegue, C. C. Ishiekwene, and B. A. E. Afere, "A note on application of kernel derivatives in density estimation with the univariate case," *J. Stat. Manag. Syst.*, vol. 22, no. 3, pp. 415–423, Apr. 2019, doi: [10.1080/09720510.2018.1524956](https://doi.org/10.1080/09720510.2018.1524956).
- [4] I. U. Siloko, O. Ikpotokin, and E. A. Siloko, "On Hybridizations of Fourth Order Kernel of the Beta Polynomial Family," *Pakistan J. Stat. Oper. Res.*, pp. 819–829, Sep. 2019, doi: [10.18187/pjsor.v15i3.2625](https://doi.org/10.18187/pjsor.v15i3.2625).
- [5] P. J. Green, K. Łatuszyński, M. Pereyra, and C. P. Robert, "Bayesian computation: a summary of the current state, and samples backwards and forwards," *Stat. Comput.*, vol. 25, no. 4, pp. 835–862, Jul. 2015, doi: [10.1007/s11222-015-9574-5](https://doi.org/10.1007/s11222-015-9574-5).
- [6] W. J. Lee, G. P. Mendis, M. J. Triebe, and J. W. Sutherland, "Monitoring of a machining process using kernel principal component analysis and kernel density estimation," *J. Intell. Manuf.*, vol. 31, no. 5, pp. 1175–1189, Jun. 2020, doi: [10.1007/s10845-019-01504-w](https://doi.org/10.1007/s10845-019-01504-w).
- [7] W. Härdle, A. Werwatz, M. Müller, and S. Sperlich, *Nonparametric and Semiparametric Models*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2004, doi: [10.1007/978-3-642-17146-8](https://doi.org/10.1007/978-3-642-17146-8).
- [8] J. S. Simonoff, *Smoothing Methods in Statistics*. New York, NY: Springer New York, 1996, doi: [10.1007/978-1-4612-4026-6](https://doi.org/10.1007/978-1-4612-4026-6).

- [9] O. Ikpotokin and I. U. Siloko, "A Comparative Analysis of Bootstrap Multivariate Exponentially-Weighted Moving Average (BMEWMA) Control Limits," *Ind. Eng. Manag. Syst.*, vol. 18, no. 3, pp. 315–329, Sep. 2019, doi: [10.7232/iems.2019.18.3.315](https://doi.org/10.7232/iems.2019.18.3.315).
- [10] A. Helu, H. Samawi, H. Rochani, J. Yin, and R. Vogel, "Kernel density estimation based on progressive type-II censoring," *J. Korean Stat. Soc.*, vol. 49, no. 2, pp. 475–498, Jun. 2020, doi: [10.1007/s42952-019-00022-y](https://doi.org/10.1007/s42952-019-00022-y).
- [11] I. U. Siloko, O. Ikpotokin, F. O. Oyegue, E. A. Siloko, and C. C. Ishiekwene, "Numerical computation of efficiency of beta polynomial kernel using product approach.," *J. Appl. Sci. Technol.*, vol. 23, 2019, Available at: [Google Scholar](https://scholar.google.com/).
- [12] C.-L. Li, W.-C. Chang, Y. Mroueh, Y. Yang, and B. Póczos, "Implicit kernel learning," *arXiv Prepr. arXiv1902.10214*, 2019, Available at: [Google Scholar](https://scholar.google.com/)
- [13] M. Klusch, S. Lodi, and G. Moro, "Distributed clustering based on sampling local density estimates," in *IJCAI*, 2003, vol. 3, pp. 485–490, Available at: [Google Scholar](https://scholar.google.com/)
- [14] X. Nguyen, M. J. Wainwright, and M. I. Jordan, "Decentralized detection and classification using kernel methods," in *Twenty-first international conference on Machine learning - ICML '04*, 2004, p. 80, doi: [10.1145/1015330.1015438](https://doi.org/10.1145/1015330.1015438).
- [15] Y. Hu, H. Chen, J. Lou, and J. Li, "Distributed Density Estimation Using Non-parametric Statistics," in *27th International Conference on Distributed Computing Systems (ICDCS '07)*, 2007, pp. 28–28, doi: [10.1109/ICDCS.2007.100](https://doi.org/10.1109/ICDCS.2007.100).
- [16] B. Albadareen and N. Ismail, "Several new kernel estimators for population abundance," 2017, p. 080018, doi: [10.1063/1.4981002](https://doi.org/10.1063/1.4981002).
- [17] A. Gramacki, *Nonparametric Kernel Density Estimation and Its Computational Aspects*, vol. 37. Cham: Springer International Publishing, 2018, doi: [10.1007/978-3-319-71688-6](https://doi.org/10.1007/978-3-319-71688-6)
- [18] H. Samawi, H. Rochani, J. Yin, D. Linder, and R. Vogel, "Notes on kernel density based mode estimation using more efficient sampling designs," *Comput. Stat.*, vol. 33, no. 2, pp. 1071–1090, Jun. 2018, doi: [10.1007/s00180-017-0787-2](https://doi.org/10.1007/s00180-017-0787-2).
- [19] J. Jiang, Y.-L. He, D.-X. Dai, and J. Z. Huang, "A new kernel density estimator based on the minimum entropy of data set," *Inf. Sci. (Ny.)*, vol. 491, pp. 223–231, Jul. 2019, doi: [10.1016/j.ins.2019.04.010](https://doi.org/10.1016/j.ins.2019.04.010).
- [20] S. Wang, A. Li, K. Wen, and X. Wu, "Robust kernels for kernel density estimation," *Econ. Lett.*, vol. 191, p. 109138, Jun. 2020, doi: [10.1016/j.econlet.2020.109138](https://doi.org/10.1016/j.econlet.2020.109138).
- [21] M. Rosenblatt, "Remarks on a Multivariate Transformation," *Ann. Math. Stat.*, vol. 23, no. 3, pp. 470–472, Sep. 1952, doi: [10.1214/aoms/1177729394](https://doi.org/10.1214/aoms/1177729394).
- [22] E. Parzen, "On Estimation of a Probability Density Function and Mode," *Ann. Math. Stat.*, vol. 33, no. 3, pp. 1065–1076, Sep. 1962, doi: [10.1214/aoms/1177704472](https://doi.org/10.1214/aoms/1177704472).
- [23] D. W. Scott, *Multivariate Density Estimation*. Wiley, 2015. doi: [10.1002/9781118575574](https://doi.org/10.1002/9781118575574)
- [24] B. E. Hansen, "Exact Mean Integrated Squared Error of Higher Order Kernel Estimators," *Econom. Theory*, vol. 21, no. 06, Dec. 2005, doi: [10.1017/S0266466605050528](https://doi.org/10.1017/S0266466605050528).
- [25] J. S. Marron and D. Nolan, "Canonical kernels for density estimation," *Stat. Probab. Lett.*, vol. 7, no. 3, pp. 195–199, Dec. 1988, doi: [10.1016/0167-7152\(88\)90050-8](https://doi.org/10.1016/0167-7152(88)90050-8).
- [26] T. Duong, "Spherically symmetric multivariate beta family kernels," *Stat. Probab. Lett.*, vol. 104, pp. 141–145, Sep. 2015, doi: [10.1016/j.spl.2015.05.012](https://doi.org/10.1016/j.spl.2015.05.012).
- [27] C. Lacour, P. Massart, and V. Rivoirard, "Estimator Selection: a New Method with Applications to Kernel Density Estimation," *Sankhya A*, vol. 79, no. 2, pp. 298–335, Aug. 2017, doi: [10.1007/s13171-017-0107-5](https://doi.org/10.1007/s13171-017-0107-5).
- [28] S. R. Sain, "Multivariate locally adaptive density estimation," *Comput. Stat. Data Anal.*, vol. 39, no. 2, pp. 165–186, Apr. 2002, doi: [10.1016/S0167-9473\(01\)00053-6](https://doi.org/10.1016/S0167-9473(01)00053-6).

-
- [29] I. Barbeito and R. Cao, "Nonparametric curve estimation and bootstrap bandwidth selection," *WIREs Comput. Stat.*, vol. 12, no. 3, May 2020, doi: [10.1002/wics.1488](https://doi.org/10.1002/wics.1488).
- [30] J. Jarnicka, "Multivariate kernel density estimation with a parametric support," *Opusc. Math.*, vol. 29, no. 1, p. 41, 2009, doi: [10.7494/OpMath.2009.29.1.41](https://doi.org/10.7494/OpMath.2009.29.1.41).
- [31] I. U. Siloko, E. A. Siloko, O. Ikpotokin, C. C. Ishiekwene, and B. A. Afere, "On Asymptotic Mean Integrated Squared Error's Reduction Techniques in Kernel Density Estimation," *Int. J. Comput. Theor. Stat.*, vol. 06, no. 01, pp. 90–98, May 2019, doi: [10.12785/ijcts/060110](https://doi.org/10.12785/ijcts/060110).
- [32] A. Azzalini and A. W. Bowman, "A Look at Some Data on the Old Faithful Geyser," *Appl. Stat.*, vol. 39, no. 3, p. 357, 1990, doi: [10.2307/2347385](https://doi.org/10.2307/2347385).