Jordan Journal of Mathematics and Statistics (JJMS) 13(3), 2020, pp 327 - 340

A MINI REVIEW OF DIMENSIONAL EFFECTS ON ASYMPTOTIC MEAN INTEGRATED SQUARED ERROR AND EFFICIENCIES OF SELECTED BETA KERNELS

I. U. SILOKO[(1)], E. A. SILOKO[(2)] AND O. IKPOTOKIN[(3)]

Abstract. The asymptotic mean integrated squared error (AMISE) is one of the popular performance measures in density estimation. The popularity of the AMISE in kernel estimation is because of its consideration of dimensions while other performance measures are dimensionless. This error criterion comprises of two components whose contributions are determine by the bandwidth. This paper briefly discusses the effects of dimension on the performances and efficiencies of some kernel functions of the beta polynomial family using the asymptotic mean integrated squared error. The results of the study show that as the power of the kernel function increases, the AMISE increases and with decrease in the efficiency as the power and dimensions increases. Also an increase in dimensions resulted in increase in AMISE but decreases with increase in sample sizes

1. Introduction

Data smoothing which involves analysis and virtualizations of data is a fundamental and important aspect of statistics. Data smoothing technique considers inferences and conclusions about the set of distributions. One of the widely used nonparametric data analytic tools is the kernel estimator. Kernel density estimators are nonparametric data smoothing techniques in density estimation. The popularity of the kernel density

2010 *Mathematics Subject Classification.* 62G05, 62G08, 62G20.

Key words and phrases. AMISE, Bandwidth, Efficiency, Kernels, Polynomial.

Received: Jan. 10, 2019 Accepted: Nov. 21, 2019 .

estimator is due to its simplicity of implementation and interpretation of results. The kernel estimators are of wide applicability in nonparametric density estimation techniques and even form the bedrock in semiparametric estimation [5]. One of the advantages of the kernel estimators as nonparametric estimators is their flexibility in modeling a given set of observations unlike the parametric estimators that are rigid and are affected by bias specification. However, the flexibility of nonparametric estimators has resulted in high computational cost which limited their uses. Due to the statistical importance of the kernel estimators in data analysis and visualization, they have become the most widely studied amongst the classes of estimators in this family [1, 12]. The univariate kernel estimator has its compact form as

$$\hat{f}(x) = \frac{1}{nh_x} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h_x}\right) \tag{1.1}$$

where K(.) and $h_x > 0$ are the kernel function and bandwidth or smoothing parameter while X_i are the data to be analysed with n representing the total number of the observations. The kernel estimator in Equation (1.1) has been of wide applicability in many fields such as archaeology, banking, climatology, economics, genetics, hydrology, and physiology [14]. The shape of the resulting estimates of the data is determined by the kernel function while the level of smoothness of $\hat{\text{f}}$(x) is regulated by the magnitude of the smoothing parameter. A small value of h_x will lead to undersmoothing while large value of h_x yields a smoother estimate but might not reveal the important features that are present in the set of observations. A major disadvantage of the kernel estimator is its biasness especially with long tail density but this problem can be avoided by adopting the adaptive smoothing parameter methods. The estimator in Equation (1.1) possesses the nonnegativity property and every kernel function must satisfy these assumptions

$$\int K(x)dx = 1$$

$$\int xK(x)dx = 0 \tag{1.2}$$

$$\int x^2K(x)dx = \mu_2(K)^2 < \infty$$

In the assumptions in Equation (1.2), it is expected that every kernel function must be a probability density function which must integrate to one with a zero mean while its variance denoted by $\mu_2(K)^2 < \infty$ must not be zero [9].

Generally, in Equation (1.1), proper implementation of the estimator is dependent on accurate selection of the smoothing parameter. Bandwidths selection is critical in kernel methods and has attracted the attention of researchers over the years; however there is no generally acceptable selection method. Recent data based smoothing parameter selectors can be found in the following articles [2, 6, 10]. The smoothing parameter is the major determinant when measuring the performance of any kernel function. In kernel density estimation, performance simply means the closeness of a kernel density estimate to its target density.

The aim of this paper is to investigate the effect of dimension on the performance and efficiencies of some selected kernels from the beta polynomial family using the asymptotic mean integrated squared error as the criterion function which shall henceforth be referred to as AMISE. There are other global measures of discrepancies but the AMISE has an advantage of being the best mathematically explained measure that considers the dimensions of the kernel function.

2. Asymptotic MISE Approximations and Multivariate Kernel Estimator.

The mean integrated squared error with its two components is the commonest optimality criterion function and is given by

$$MISE\left(\hat{f}(x)\right) = \int Var\left(\hat{f}(x)\right)dx + \int Bias^2\left(\hat{f}(x)\right)dx \tag{2.1}$$

There is usually the bias-variance trade-off between the terms of the AMISE, the bias can be reduced while the variance increases and vice versa, by varying the magnitude of the smoothing parameter. If the smoothing parameter is small, it gives a smaller value of the bias and the variance increases resulting in an undersmoothed curve and with large value of the smoothing parameter, it gives a smaller variance while the bias increases resulting in an oversmoothed density estimate [11]. The exact form of the mean integrated squared error in Equation (2.1) is obtained by kernel convolution while its approximate form can easily be obtained using the Taylor's series expansion. The approximate form of Equation (2.1) known as the asymptotic mean integrated squared error will produce the integrated variance and the integrated squared bias given by

$$AMISE\left(\hat{f}(x)\right) = \frac{R(K)}{nh_x} + \frac{h_x^4}{4}\mu_2(K)^2 R(f'') \tag{2.2}$$

where R(K) and $\mu_2(K)^2$ are the roughness and second moment of the kernel, $f^{"}$ is the second derivative of the unknown function while $R(f^{"}) = \int f^{"}(x)^2 dx$ represents the roughness of the unknown density function [9]. The AMISE in Equation (2.2) will produce its smallest value when the solution to the differential equation is obtain

$$\frac{\partial}{\partial h_x} AMISE\left(\hat{f}(x)\right) = \frac{-R(K)}{nh_x^2} + \mu^2(K)^2 h_x^3 R(f'') = 0$$

On solving the differential equation, we have the smoothing parameter that minimizes the AMISE of the kernel estimator as

$$h_{x-AMISE} = \left[\frac{R(K)}{\mu_2(K)^2 R(f'')}\right]^{\frac{1}{5}} \times n^{-\frac{1}{5}} \tag{2.3}$$

The smoothing parameter that produces the minimum AMISE in Equation (2.3) can be expressed in dimensional form as

$$h_{x-AMISE} = \left[\frac{R(K)}{\mu_2(K)^2 R(f'')}\right]^{\frac{1}{4+d}} \times n^{-\frac{1}{(4+d)}} \tag{2.4}$$

The multivariate form of Equation (1.1) with a single bandwidth kernel estimator is given as

$$\hat{f}(x) = \frac{1}{nh_x^d}\sum_{i=1}^{n} K\left(\frac{x - X_i}{h_x}\right) \tag{2.5}$$

where d is the dimension of the kernel function. The kernel, K, in this case is a d-variate density function that satisfy the conditions in Equation (1.2) and its contours are assumed to be spherically symmetric. In this multivariate form, the same smoothing parameter that controls the size of the kernel is applied to each dimension. An advantage of this multivariate estimator is that the AMISE and its optimal smoothing parameter can be easily computed unlike other complex forms without explicit optimal smoothing parameter formula. The corresponding AMISE of Equation (2.5) is given as

$$AMISE\left(\hat{f}(x)\right) = \frac{R(K)}{nh_x^d} + \frac{h_x^4}{4}\mu_2(K)^2\int(\nabla^2 f(x))^2 dx \tag{2.6}$$

where $\nabla^2 f(x) = \sum_{i=1}^{d} \frac{\partial^2 f(x)}{\partial x_i}$

Also, the smoothing parameter that will minimize the AMISE in Equation (2.6) is of the form

$$h_{x-AMISE} = \left[\frac{dR(K)}{\mu_2(K)^2\int(\nabla^2 f(x))^2 dx}\right]^{\frac{1}{(4+d)}} \times n^{\frac{-1}{4+d}} \tag{2.7}$$

The statistical quantities required for the computation of smoothing parameter in Equation (2.4) and Equation (2.7) are the roughness of the kernel function, moment of the kernel and roughness of the unknown probability function. The smoothing parameter in Equation (2.7) will result in the smallest value of the AMISE given by

$$AMISE\left(\hat{f}(x)\right) = \left(\frac{d+4}{4d}\right)\left(\mu_2(K)^{2d}(dR(K))^4\left(\int(\nabla^2 f(x))^2 dx\right)^d n^{-4}\right)^{\frac{1}{d+4}} \tag{2.8}$$

The rate of convergence of the AMISE in Equation (2.8) is of order$n^{\frac{-4}{4+d}}$, and this rate tends to be slower as the dimension of the kernel function increases [13]. The smoothing parameter with the minimum AMISE value is of order $n^{\frac{-1}{(4+d)}}$ and the expression contains the second derivatives of the unknown density f being estimated.

3. The Beta Polynomial Kernel Functions.

The beta polynomial kernel family is a class of kernel functions in which the function with higher degree has smoother estimates and also possesses more derivatives. The general pth kernel of the beta polynomial family for $p \geq 0$ with $t \in [-1, 1]$ is of the form

$$K_p(t) = \frac{(2p+1)!}{2^{p+1}(p!)^2}(1-t^2)^p \tag{3.1}$$

where p=0,1,2,$\cdots$,∞ is the power of the polynomial function [3]. As the value of p increases from 0 to 3, the resulting kernels are the Uniform, Epanechnikov, Biweight and Triweight kernels which are members of this polynomial family. In this class of kernels, the uniform kernel is the simplest kernel while the popular normal kernel is the limiting case when p tends to infinity [4, 7]. These classes of kernels are popular due to the desire to study their mathematical properties. This study examines the effects of dimension on the AMISE and efficiencies of some polynomial kernels for p=1,2,3,4 which are the Epanechnikov, Biweight, Triweight and Quadriweight kernel functions and also their efficiencies as the dimension increases. The resulting kernel functions for p=1,2,3,4 from the general form in Equation (3.1) are as follows. When p=1, we have the Epanechnikov kernel also known as the quadratic kernel which is of the form

$$K_1(t) = \frac{3}{4}(1-t^2) \tag{3.2}$$

Again when p=2,3 and 4 we have the Biweight also known as the quartic kernel, Triweight kernel and Quadriweight kernel which are as follows

(3.3)
$$K_2(t) = \frac{15}{16}(1 - t^2)^2$$

(3.4)
$$K_3(t) = \frac{35}{32}(1 - t^2)^3$$

(3.5)
$$K_4(t) = \frac{315}{256}(1 - t^2)^4$$

The Epanechnikov, Biweight, Triweight and Quadriweight kernel functions are of wider applications because they form the basis when discussing this class of kernel functions. Often times, the results obtain from these first members of this family of kernels can be easily generalized to other higher powers of this family. The multivariate form of Equation (3.1) using the product approach that employs the product of the univariate kernels is of the form

(3.6)
$$K_p^{product}(t) = A^d \prod_{i=1}^{d} (1 - t_i^2)^p$$

where $A = \frac{(2p+1)!}{2^{p+1}(p!)^2}$ is the normalization constant and d is the dimension of the kernel. The order of the smoothing parameter that minimizes the AMISE of the product kernel is same as that of the multivariate fixed kernel in Equation (2.7) and the AMISE is also of the same order as that of the multivariate fixed kernel. The advantage of the multivariate product kernel over other forms is that the product approach is beneficial especially when the scales of the variables to be considered differ. Also, in the case of unimodal densities, the product kernel has been suggested by many authors [8].

4. The Efficiency of Kernel Density Estimator.

The efficiency of the univariate symmetric kernel which is measured in comparison with the Epanechnikov kernel is of the form

$$Eff(K) = \left(\frac{C(K_e)}{C(K)}\right)^{\frac{5}{4}} \tag{4.1}$$

where $C(K) = (R(K)^4 \mu_2(K)^2)^{\frac{1}{5}}$ is a constant of any given kernel and $C(K_e)$ is the constant of the Epanechnikov kernel [11]. The efficiency of the multivariate kernel using the product method is given by

$$Eff(K^p) = \left(\frac{C(K_e^p)}{C(K^p)}\right)^{\frac{d+4}{4}} \tag{4.2}$$

where d is the dimension of the kernel; $C(K_e^p)$ is the higher dimensional product form of the Epanechnikov kernel constant and $C(K^p)$ is the higher dimensional product form of any other given kernel. The value of the constant C(K) for the higher dimensional product kernel can be obtained from the relation

$$C(K) = \left(R(K)^4 \mu_2(K)^2\right)^{\frac{d}{d+4}} \tag{4.3}$$

where R(K) is the roughness of the kernel function and $\mu_2(K)$ is the second moment of the kernel function [13]. The minimizer of Equation (4.3) over K is exactly the minimizer of C(K), which is the same as the Epanechnikov kernel. This simply implies that the AMISE optimal product kernel of this family of kernels is

$$K^p(t) = \left(\frac{3}{4}\right)^d \prod_{i=1}^{d} (1 - t_i^2) \tag{4.4}$$

The computation of the smoothing parameter as stated in Equation (2.7) and the d-dimensional efficiency in Equation (4.2) require the roughness of the kernel, moment of the kernel and the roughness of the unknown probability density function. The roughness of the kernel function is usually of the form

$$R(K) = \int K(t)^2 dt \tag{4.5}$$

Also, the second moment of any kernel function also known as the variance is of the form

$$\mu_2(K)^2 = \int t^2 K(t) dt \tag{4.6}$$

The roughness of the unknown probability distribution function given in Equation (2.2) is of the form

$$R(f'') = \int f''(x)^2 dx \tag{4.7}$$

The roughnesses of the Epanechnikov, Biweight, Triweight and Quadriweight kernels will be obtained by numerically integrating their functions within the support interval [-1,1].

5. Discussion of Results

The contribution of the bias and variance to the AMISE is determined by the magnitude of the smoothing parameter. The results for the efficiencies and AMISE of the Epanechnikov, Biweight, Triweight and Quadriweight kernels are obtain using Mathematica version 9. We examine the AMISE value for the stated kernel functions with respect to their dimension using various sample sizes. The Epanechnikov, Biweight, Triweight and Quadriweight kernel functions are of wide applications because they form the basis when discussing this class of kernel functions. The efficiencies of these kernels are in Table 5.1 and from the results, the efficiencies decrease as the power and dimension increases except for the Epanechnikov kernel which is the optimum kernel. The decrease in the efficiencies is as a result of the curse of dimensionality effect which is associated with nonparametric estimators. However, the loss of efficiencies with respect to power and dimension is minimal since the results in Table

5.1 are all above 90% when express in percentage while the optimum kernel, which is the Epanechnikov kernel, is 100%.

Table 5.1: Efficiencies of Beta Polynomial Kernel Functions for d=1,2,3,4 and 5

kernels	p	d=1	d=2	d=3	d=4	d=5
Epanechnikov	1	1.000	1.000	1.000	1.000	1.000
Biweight	2	0.994	0.988	0.982	0.976	0.970
Triweight	3	0.987	0.974	0.961	0.948	0.935
Quadriweight	4	0.981	0.963	0.945	0.927	0.909

Table 5.2: AMISE of Beta Polynomial Kernel Functions for the First Sample n=2500.

kernels	d=1	d=2	d=3	d=4	d=5
Epanechnikov	0.00155571	0.00304896	0.00454710	0.00583200	0.00681662
Biweight	0.00230713	0.00898822	0.03150200	0.10542500	0.34475400
Triweight	0.00266570	0.01218483	0.05054756	0.20132725	0.78641901
Quadriweight	0.00309917	0.01748168	0.09259572	0.48102589	2.4857679

Table 5.3: AMISE of Beta Polynomial Kernel Functions for the Second Sample n=5000.

kernels	d=1	d=2	d=3	d=4	d=5
Epanechnikov	0.00089352	0.00192072	0.00305997	0.00412385	0.00500931
Biweight	0.00132510	0.00566222	0.02119930	0.07454661	0.25334842
Triweight	0.00153104	0.00767596	0.03401599	0.14235987	0.57791357
Quadriweight	0.00178001	0.01101277	0.06231230	0.34013667	1.82670942

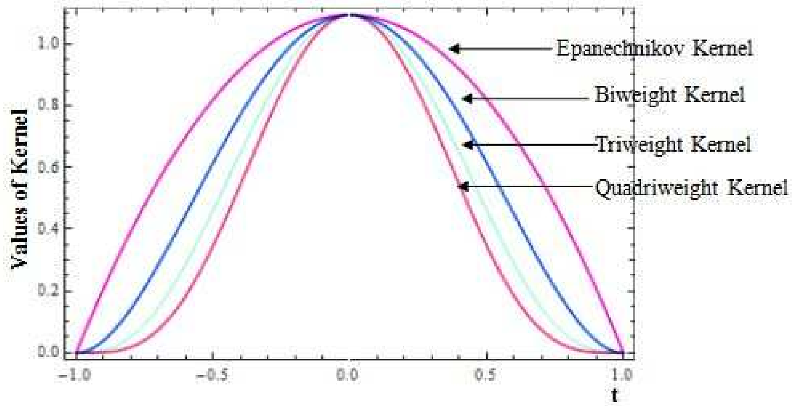

FIGURE 1. A graph of the kernel functions for the values of p.

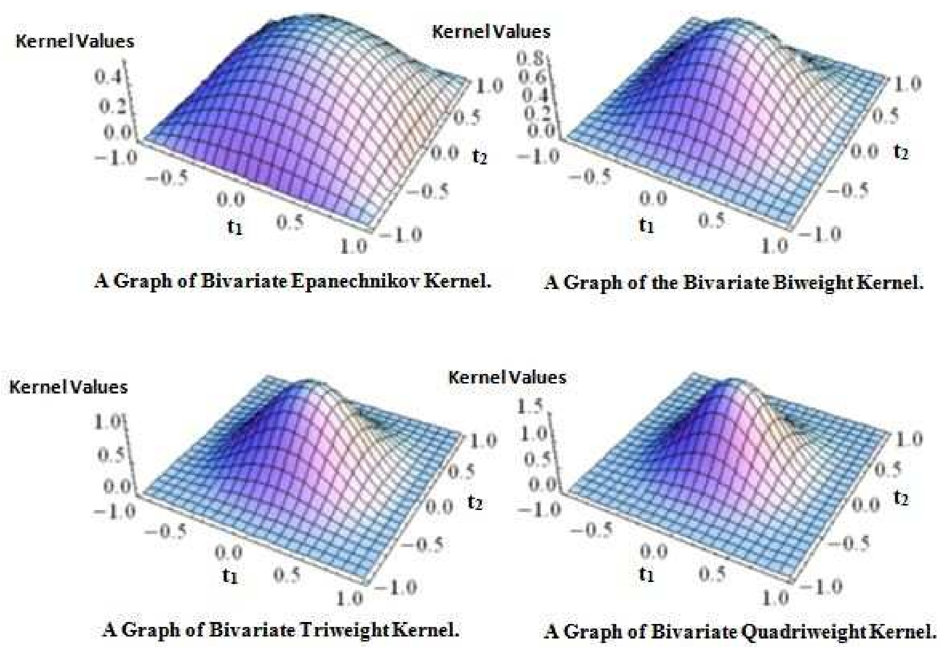

FIGURE 2. Bivariate Epanechnikov, Biweight, Triweight and Quadriweight Kernels.

The estimates of the univariate and bivariate kernels of the Epanechnikov, Biweight, Triweight and Quadriweight kernels are in Figure 1 and Figure 2 respectively.

The kernel functions with higher power tends to be smoother and the loops of the bivariate kernels of Triweight and Quadriweight moves closer to the origin than the bivariate kernels of the Epanechnikov and Biweight kernel functions in Figure 2, this is due to their degree of differentiability and this implies that the Triweight and Quadriweight kernel estimates will be better than those of Epanechnikov and Biweight because they possess more derivatives. The effect of dimension on the AMISE of the studied kernel functions is illustrated using two different sample sizes. The first sample size is 2500 while the second sample size is 5000 and this is because nonparametric estimation is mainly beneficial when the sample size is reasonably large. Table 5.2 and Table 5.3 show the kernel functions and the AMISE values of the different sample sizes for the various dimensions. The results of Table 5.2 and Table 5.3 show that the value of the AMISE increases as the power of the polynomial function increases and the AMISE increases as the dimension of the kernel function increases. The Epanechnikov kernel is the optimum kernel with respect to the AMISE, where optimality in kernel density estimation implies the kernel function with the minimum AMISE value

The results in Table 5.2 and Table 5.3 are for the first and second sample sizes, and from the results, the value of the AMISE increases with increase in the power of the polynomial. Also, as the dimension of the kernel function increases, the value of the AMISE increases as well. This simply means that an increase in dimension will result in an increase in the value of the asymptotic mean integrated squared error and vice versa. The reduction in the AMISE as a result of the increase in the sample size implies that for nonparametric estimation to be beneficial; it requires large sample sizes for its implementation. Hence the performances of the kernel estimators also depend on the sample size.

6. Conclusion.

This study illustrates the effects of dimension on the efficiencies and performance of some kernels of the beta polynomial family using the AMISE as the error criterion with different sample sizes. The results of the investigation reveal that as the dimension of the kernel increases, there is decrease in the efficiencies and increase in AMISE. The decrease in the efficiencies values of these kernel functions is as a result of the curse of dimensionality effect that is associated with nonparametric estimators. Again, as the sample size increases, there is decrease in the AMISE value and this shows that sample sizes and dimensions affect the contribution of the bias and variance to the AMISE in kernel density estimation.

References

[1] Z. I. Botev, J. F. Grotowski and D. P. Kroese *Kernel Density Estimation via Diffusion. Annals of Statistics,*,**28**(2010), 2916-2957.

[2] J. E. Chacón and T. Duong *Data-driven density derivative estimation, with applications to nonparametric clustering and bump hunting. Electronic Journal Statistics*, **7**(2013), 499-532.

[3] T. Duong,*Spherically symmetric multivariate beta family kernels. Statistics and Probability Letters* **104** (2015), 141-145.

[4] B. E. Hansen, *Exact Mean Integrated Squared Error of Higher Order Kernel Estimators. Econometric Theory*, **21**((2005)), 1031-1057.

[5] W. Hárdle, M. Müller, S. Sperlich and A. Werwatz, *Nonparametric and Semiparametric Models.* Springer-Verlag Berlin Heidelberg New York, (2004).

[6] M. Jiang and S. B. Provost, *A hybrid bandwidth selection methodology for kernel density estimation. Journal of Statistical Computation and Simulation*, **84**(2014), 614-625.

[7] J. Marron and D. Nolan,*kernels for density estimation. Statistics and Probability Letters*,**7**(1988),195-199.

[8] R. S. Sain,*Multivariate Locally Adaptive Density Estimation. Computational Statistics and Data Analysis*,textbf39(2002),165-186.

[9] D. W. Scott,*Multivariate density estimation. Theory, practice and visualisation*,Wiley, New York (1992).

[10] I. U. Siloko, C. C. Ishiekwene and F. O. Oyegue, *New Gradient Methods for Bandwidth Selection in Bivariate Kernel Density Estimation*, *Mathematics and Statistics* **6**(1)(2018),1-8 .

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

[12] J. S. Simonoff,. *Smoothing Methods in Statistics*,Springer-Verlag, New York, (1996).

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

[14] S. J. Sheather,*Density Estimation*,Statistical Science, textbf19(4)(2004), 588-597.

(1) Department of Mathematics and Computer Science, Edo University Iyamho, Nigeria

Email address: siloko.israel@@edouniversity.edu.ng; suzuazor@@yahoo.com

(2) Department of Mathematics, University of Benin, Benin City, Nigeria

Email address: myakpeski@@yahoo.com

(3) Department of Mathematics and Statistics, Ambrose Alli University, Ekpoma, Nigeria.

Email address: ikpotokinosayomore@@yahoo.co.uk