A new Family of Hybrid Classical Polynomial Kernels in Density Estimation

Benson Ade Afere

Department of Mathematics and Statistics, Federal Polytechnic, Idah, Nigeria

***Abstract* - Kernel density estimation over the years has been placing more emphasis on the problem of the choice of optimal bandwidth. Nonetheless, the kernel function still has some roles to perform in the curve smoothing settings. Thus, in this paper, a new family of hybrid polynomial kernels is proposed. A generalized error scheme of the proposed family of kernels is constructed. A Monte Carlo experiment is performed using three univariate densities and it was discovered that the proposed family of hybrid polynomial kernels have significant low asymptotic mean integrated square error as compared with the existing family of polynomial kernels in the literature especially as the order of the kernels increases. Four real life data sets were equally used to show the performance of the proposed new family. It was observed that the proposed hybrid kernels perform well for the data sets considered.**

***Keywords* - Kernel density estimation, polynomial kernels, hybrid kernels, generalized global error, Monte Carlo experiment.**

MSC 2010 Classification: 62G07

I. INTRODUCTION

One class of smoothing techniques used for data analysis and visualizations is the nonparametric kernel density estimator (NKDE). This method is essentially the construction of an estimate of an underlying probability density function from an observed data set. This method has been used vastly in many areas such as in random differential equations problems [12], insurance [14], archeology, banking, climatology, economics, genetics, hydrology and physiology [17]. Its vast usability is based mainly on the simpleness of its implementation and interpretation of results [18]. It has been widely asserted that NKDE method is essentially marred majorly by the difficulties that centered around the choice of optimal bandwidth and minorly by the kernel functions [19]. Nevertheless, this does not undermine the choice and even the development of kernel functions [1]. In view of this, a family of hybrid polynomial kernels is proposed in this work.

Now, suppose $x_1, x_2, \ldots, x_n$ is an independent and identically distributed random data set with the probability density function f on a bounded interval $[-1,1]$, then the univariate NKDE method is given by

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

where $K(t)$ and h in (1) are respectively the kernel function and bandwidth. The kernel function $K(t)$ is required to satisfy the following properties:

$$\left.\begin{array}{ll} i. & \int_{\mathbb{R}} K(t)dt = 1 \\ ii. & \int_{\mathbb{R}} tK(t)dt = 0 \\ iii. & \int_{\mathbb{R}} K^2(t)dt = \|K\|_2^2 = \alpha < \infty \end{array}\right\} \tag{2}$$

Equation (1) surfaced into the mathematical statistics community in a technical report of Fix and Hodges [8]. The unpublished work of Akaike [4] also contains some of the basic ideas presented by [8]. The first notable published paper in NKDE is credited to Rosenblatt [15]. However, the major landmark of NKDE was achieved through Parzen [13]. Hence, the method often referred to as Rosenblatt-Parzen estimators or simply Parzen estimators [19]. Ever since this major landmark, several teams of tireless academia have contributed immensely to the advancement of NKDE. See, for instance, the work of [2, 3, 5, 7, 9, 11, 16, 18, 19] and the references therein.

The family of classical polynomial kernels is given as:

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

where $\mathrm{B}(\cdot,\cdot)$ and p in (3) are respectively the beta density and power of the family [20]. The specific results of K_c yields Epanechnikove kernel, Biweight kernel, Triweight kernel and Quadriweight kernel when $p = 1,2,3$ and 4 respectively. Several works have been done on the construction of kernels. See for instance, [2, 3].

II. FAMILY OF HYBRID CLASSICAL POLYNOMIAL KERNELS

In this section, a proposed family of hybrid classical polynomial kernels is constructed. Consider the simple formula

$$K_H(t) = \rho_1 K_{[p-1]}(t) + \rho_2 K_{[p]}(t) \tag{4}$$

where $K_{[p-1]}$ and $K_{[p]}$ are respectively the families of classical polynomial kernels of order $p-1$ and p as presented in (3) and $\rho_1 + \rho_2 = 1$.

Now, let $\rho_1 = 1/10$ and then substitute (3) into (4) and apply necessary algebraic rules, then the proposed family of hybrid polynomial kernels is given by

$$K_H(t) = \{5p2^{2p+1}\mathrm{B}(p,p)\}^{-1}[5(2p+1) - 10p(t^2-1)^{-1}] \times (1-t^2)^p, |t| \le 1, \ p = 1,2,3,... < \infty \tag{5}$$

The kernel functions of the proposed family of hybrid kernels in (5) are presented in Table 1 below. Figure 1 shows the probability density function (pdf) of the kernel functions (see Table 1). The biweight, triweight and quadriweight kernels have short support than the Epanechnikov kernel.

Table 1: Hybrid classical polynomial kernels

Kernel	Definition
Epanechnikov	$\frac{1}{8}(5-3x^2)$
Biweight	$\frac{3}{32}(1-x^2)(9-5x^2)$
Triweight	$\frac{5}{64}(1-x^2)^2(13-7x^2)$
Quadriweight	$\frac{35}{512}(1-x^2)^3(17-9x^2)$

The shapes of the hybrid symmetric kernels defined in Table 1 above are presented in Figure 1.

III. THE ERROR SCHEME OF THE PROPOSED KERNELS

It is a well-known fact in the NKDE setting that the quality of the kernels is determined by the relatively low global error and relatively high efficiency [19]. Thus, in this section, the global error scheme of the kernel estimators $\hat{f}_h(x)$ of the proposed family of hybrid polynomial kernels with respect to the mean integrated squared error is herein derived. According to [20], the mean integrated squared error (MISE) relative to $\hat{f}_h(x)$ is given by

$$\mathrm{MISE}\hat{f}_h(x) = \int \mathrm{Var}\hat{f}_h(x)dx + \int \mathrm{Bias}^2 \hat{f}_h(x)dx \tag{6}$$

where $\mathrm{Var}\hat{f}_h(x)$ and $\mathrm{Bias}\hat{f}_h(x)$ in (6) are respectively given by

$$\mathrm{Var}\hat{f}_h(x) = E\hat{f}_h^2(x) - E^2\hat{f}_h(x) \tag{7}$$

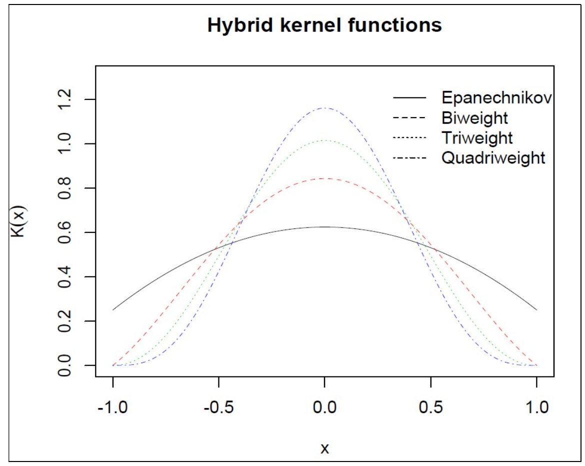

Figure 1: Shapes of hybrid symmetric kernels in Table 1

and

$$\mathrm{Bias}\hat{f}_h(x) = E\hat{f}_h(x) - f(x) \tag{8}$$

Theorem 1: Review of Silverman [19] and Afere and Alih [2]

Suppose f is a bounded density function. Suppose also that the kernel function $K(t)$ is bounded with the second moment $\int t^2K(t)dt$. If x is a point with $f(x) > 0$ such that f is continuously differentiable up to the 2^{nd} -order in a neighbourhood of x, then the variance and bias^2 *of (6) are given respectively as:*

$$\mathrm{Var}\hat{f}_h(x) = \frac{1}{nh}[\int K^2(t)dt]f(x) + o(n^{-1}h^{-1}) \tag{9}$$

and

$$\mathrm{Bias}^2\hat{f}_h(x) = \frac{1}{4}h^4\left(\int t^2K(t)dt\right)^2(f''(x))^2 + o(h^4) \tag{10}$$

Proof: The proof is contained in [2].

Remarks

1. Upon substituting (9) and (10) into (6), neglecting higher-order terms and integrated accordingly, the result yields the asymptotic mean integrated squared error (AMISE) given as:

$$\mathrm{AMISE}\hat{f}_h(x) = \frac{1}{4}h^4\left(\int t^2K(t)dt\right)^2\int(f''(x))^2dx + \frac{1}{nh}\int K^2(t)dt \tag{11}$$

2. Equation (11) provides the basis for determining the global error of any kernel.

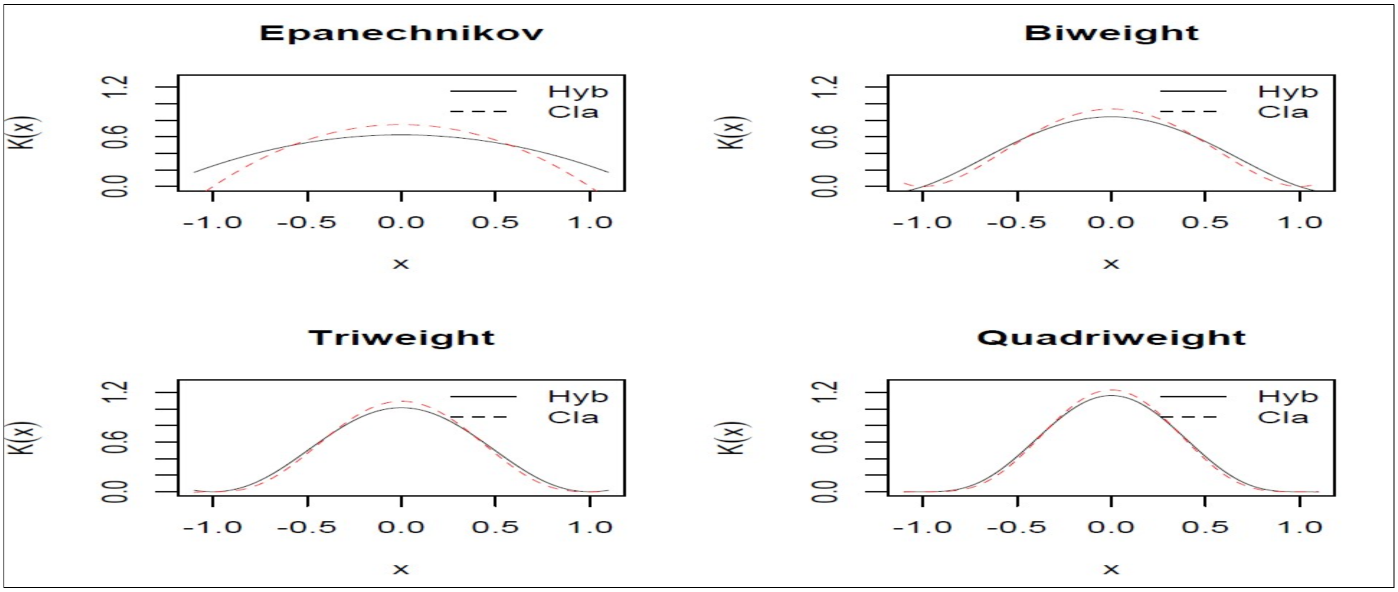

Figure 2: Plot showing the relationship between the proposed hybrid polynomial kernels and the classical polynomial kernels

Theorem 2

In addition to the conditions given in Theorem 1, let $K_H(t)$ *be any member of the family of classical polynomial kernels defined in (5), bounded with* $2m^{\text{th}}$ *moment and satisfies the property*

$$\int t^{2m} K_H(t)dt < \infty \tag{12}$$

Suppose $\int (f^{(2m)}(x))^2 dx = \frac{1}{\sigma^{4m+1} 2\pi} \Gamma(\frac{4m+1}{2})$, $m = 1, 2, 3, \cdots, < \infty$ *is the constant of* $2m^{\text{th}}$ *continuously differentiable Gaussian distribution with the support* $(-\infty, \infty)$, *then the global error scheme of the proposed family of hybrid polynomial kernel is given by*

$$\text{AMISE}\hat{f}(x) \cong \frac{8^{-\frac{8m+1}{4m+1}}}{25} m^{-\frac{4m}{4m+1}} (4m+1) \pi^{\frac{2(m-1)}{4m+1}} \gamma^{\frac{1}{4m+1}} \lambda^{\frac{4m}{4m+1}} n^{-\frac{4m}{4m+1}} \tag{13}$$

where

$$\gamma = \frac{(5+10p+5m)^2 \sigma^{-4m-1} \Gamma(m+\frac{1}{2})^2 \Gamma(2m+\frac{1}{2}) \Gamma(p+\frac{1}{2})^2}{\Gamma(p+m+\frac{3}{2})^2} \tag{14}$$

and

$$\lambda = \frac{(-20(4p^2-1)(5+20p) - 75p)\,\text{Csc}(2p\pi) \Gamma(p+\frac{1}{2})^2}{p\Gamma(2-2p)\Gamma(p)^2 \Gamma(2p+\frac{3}{2})} \tag{15}$$

Proof: The proof of this theorem is of two parts. To prove the first part, we substitute (1) in (8) and simplify, then (8) becomes

$$\text{Bias}\hat{f}_h(x) = \int K(t)\{f(x-ht) - f(x)\}dt \tag{16}$$

On applying the univariate Taylor series expansion in (16), and simplifying using the properties in (2), (16) becomes

$$\text{Bias}\hat{f}_h(x) = \frac{1}{(2m)!} h^{2m} f^{(2m)}(x) \int t^{2m} K(t)dt + o(h^{2m}) \tag{17}$$

As in Bias term, plugging (1) into (7), then applying the univariate Taylors series expansion with the necessary simplifications using the properties in (2), the variance term becomes

$$\text{Var}\hat{f}_h(x) = \frac{1}{nh} f(x) \int K^2(t)dt + 0(nh)^{-1} \tag{18}$$

Now, substituting (16) and (17) into (6), and simplify, the expression of $\text{MISE}\hat{f}_h(x)$ becomes

$$\text{MISE}\hat{f}_h(x) = \frac{1}{nh}\int K^2(t)dt + \frac{1}{((2m)!)^2} \times h^{4m} \int (f^{(2m)}(x))^2 dx (\int t^{2m} K(t)dt)^2 + o(h^{4m}) + 0(nh)^{-1} \tag{19}$$

Upon neglecting the higher-order terms, then the asymptotic mean integrated squared error (AMISE) of the kernel estimators $\hat{f}_h(x)$ is given by

$$\text{AMISE}\hat{f}_h(x) = \frac{1}{nh}\int K^2(t)dt + \frac{1}{((2m)!)^2}\times h^{4m}\int (f^{(2m)}(x))^2 dx(\int t^{2m}K(t)dt)^2 \quad (20)$$

On optimizing (20) with respect to h, the optimal bandwidth h can be obtained as

$$h_{opt} = \left(\frac{((2m)!)^2\int K^2(t)dt}{n(4m)\int (f^{(2m)}(x))^2 dx(\int t^{2m}K(t)dt)^2}\right)^{\frac{1}{4m+1}} \quad (21)$$

Thus, substituting (21) into (20), the expression of the global error ($\text{AMISE}\hat{f}_h(x)$) that is completely free from the value of h is obtained as:

$$\text{AMISE}\hat{f}_h(x) = (4m+1)(4m)^{-\frac{4m}{4m+1}}\left(\int K^2(t)dt\right)^{\frac{4m}{4m+1}}\times \left(\left(\frac{\int t^{2m}K(t)dt}{((2m)!)}\right)^2\int (f^{(2m)}(x))^2 dx\right)^{\frac{1}{4m+1}} n^{-\frac{4m}{4m+1}} \quad (22)$$

This completes the first part of the proof.

The second part shall be achieved by first computing the value of $\int (f^{(2m)}(x))^2 dx$ for the constant of $2m^{\text{th}}$ continuously differentiable Gaussian distribution with the support $(-\infty,\infty)$, $\int K^2(t)dt$ and $\int t^{2m}K(t)dt$ for the family of hybrid polynomial kernels and substitute in (22).

Now, consider the univariate normal distribution given in the equation below with mean (μ) zero and variance (σ) (i.e. $x \sim N(0\ ,\sigma^2)$).

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{x^2}{2\sigma^2}} \quad (23)$$

Differentiating equation (23) severally, and using the properties in (2), the odd derivatives will fizzle out and thus, on integrating the square of the even derivatives, we obtain the generalized form as:

$$\int (f^{(2m)}(x))^2 dx = \frac{1}{\sigma^{4m+1}2\pi}\times \Gamma(\tfrac{4m+1}{2})\ \ , m = 1,2,3,\cdots,<\infty \quad (24)$$

Also, the L_2 - norm and $2m^{th}$ moment of Equation (5) yields respectively

$$\int K_H^2(t)dt = \frac{\sqrt{\pi}(-20(4p^2-1)(5+20p) - 75p)\text{Csc}(2p\pi)\Gamma(p+\frac{1}{2})^2}{400p\Gamma(2-2p)\Gamma(p)^2\Gamma(2p+\frac{3}{2})} \quad (25)$$

and

$$\int t^{2m}K_H(t)dt = \frac{(5+10p+5m)^2\times \Gamma(m+\frac{1}{2})^2\Gamma(2m+\frac{1}{2})\Gamma(p+\frac{1}{2})^2}{\Gamma(p+m+\frac{3}{2})^2} \quad (26)$$

Thus, substituting (24), (25) and, (26) into (23) gives

$$\text{AMISE}\hat{f}(x) \cong \frac{8^{-\frac{8m+1}{4m+1}}}{25}m^{-\frac{4m}{4m+1}}(4m+1)\pi^{\frac{2(m-1)}{4m+1}}\times \gamma^{\frac{1}{4m+1}}\lambda^{\frac{4m}{4m+1}}n^{-\frac{4m}{4m+1}}$$ where

γ and λ are as given in equations (14) and (15) respectively. This completes the proof.

IV. NUMERICAL EXPERIMENT

In this section, two simulation experiments were conducted to show the performance of the proposed hybrid kernels over their classical kernels' counterparts. Four real life data were also used to show the behaviour of the proposed hybrid kernels that is akin to their classical kernels' counterparts.

A. MONTE CARLO EXPERIMENT

To investigate the performance of the proposed hybrid kernels, some simulation studies are conducted. A Monte Carlo experiment is conducted for sample sizes $n=10$, $n=25$, $n=200$ and $n=1000$ for standard normal distribution $N(0,1)$. In the second scenario, a Monte Carl experiment of sample sizes $n=10$, $n=25$, $n=125$ and $n=300$ is conducted for three univariate densities,

(i) Standard normal $f_1 \sim N(0,1)$

(ii) Symmetric trimodal

$$f_2 \sim \frac{9}{20}N(-\frac{7}{4},1) + \frac{9}{20}N(\frac{7}{4},1) + \frac{1}{10}N(0,\frac{1}{25})$$

(iii) Asymmetric trimodal

$$f_3 \sim \frac{3}{10}N(-2,\frac{1}{4}) + \frac{3}{10}N(\frac{7}{4},\frac{1}{5}) + \frac{2}{5}N(0,2)$$

Figure 3 (a), (b), (c) and (d) shows the AMISE – (which shall be henceforth called the global error) of the proposed hybrid kernels and their respective classical polynomial kernels' counterparts for $N(0,1)$ of sample sizes $n=10$, $n=25$, $n=200$ and $n=1000$ respectively. Examination of Figure 3 reveals that all the proposed hybrid kernels have a reduced global error as compared to their classical kernels'

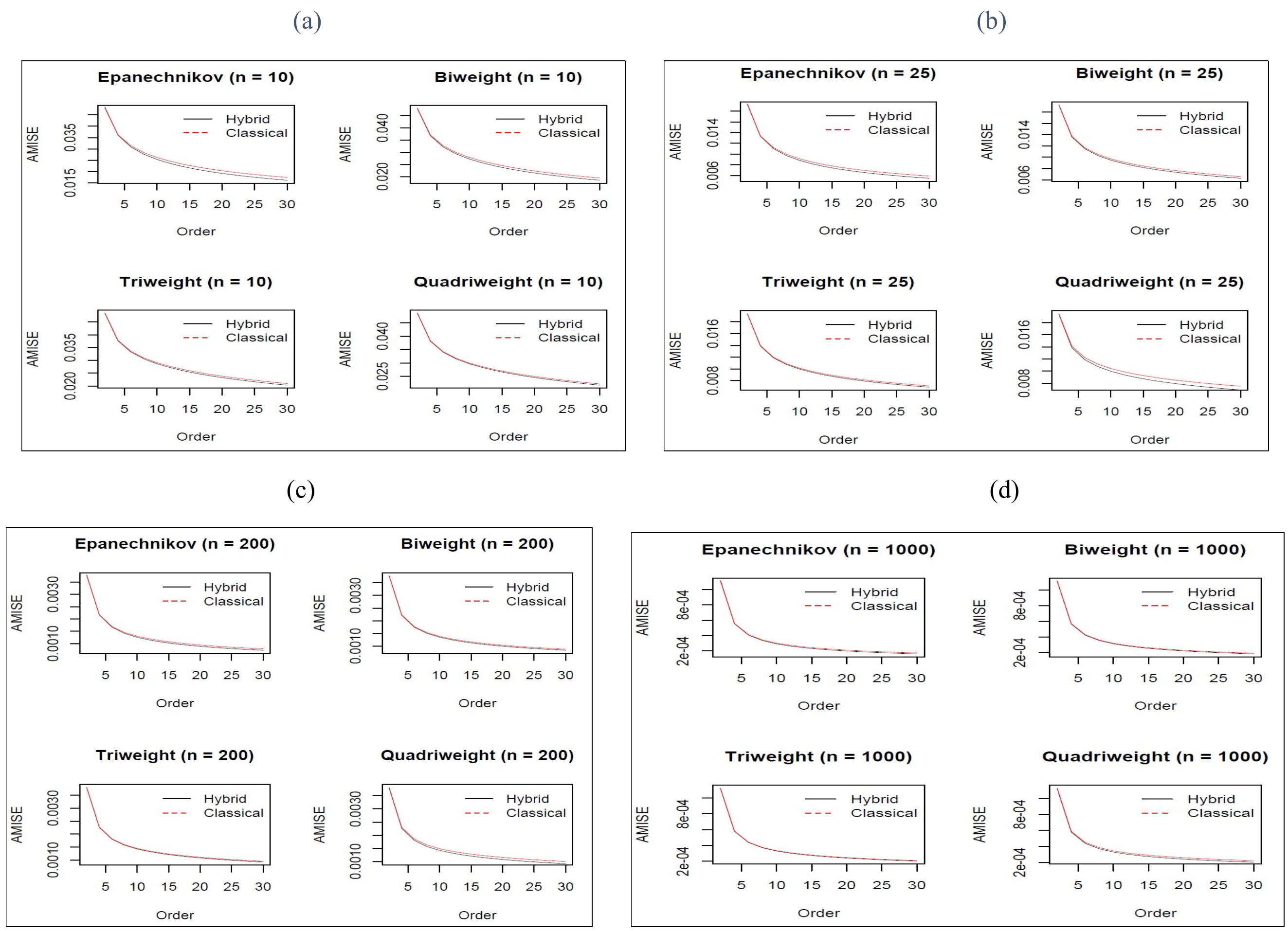

Figure 3: Plot showing the comparison of AMISE for the proposed hybrid polynomial kernels and the classical polynomial kernels

counterparts. It was equally revealed that as sample size n increases, the global error become smaller which obeys one of the properties of a good and robust estimator.

Figure 4 (a), (b), (c) and (d) shows the result of the simulation experiment conducted for three univariate mixture densities. For each of the mixture densities, the random variable X was generated and the standard deviation parameters estimated from it. After this, the global error in (13) is used by performing the simulation for $r = 1000$ runs such that the average of the global error is given as

$$\text{Error} = \frac{1}{r}\sum_{j=1}^{r} \text{AMISE}_j^{2m},\ m = 1, 2, ... \tag{27}$$

Equation (27) was computed for hybrid – Epanechnikov, Biweight, Triweight and, Quadriweight kernels for all the univariate mixture densities as presented in Figure 4. The kernels considered have a reduced global error for normal density when $n = 10, 25, 125$ and 300. However, when $n = 10$, the kernels have a reduced global error for symmetric trimodal than in the case of asymmetric trimodal. But, when $n = 25, 125$ and 300; the kernels perform better for asymmetric trimodal density than the symmetric trimodal density.

B. REAL LIFE DATA

In this subsection, the real-life applications of the proposed family of hybrid polynomial kernels is demonstrated. Equation (27) is used for the four data sets. The first – three data sets are: (1) the lifespan of car batteries in years, (2) the number of written words without mistakes in every 100 words by a set of students in a written essay and (3) the scar length of patients randomly selected in millimeters. These data sets are from [10]. The fourth data set is the Waiting time between eruptions and the duration of the eruption for the Old Faithful geyser in Yellowstone National Park, Wyoming, USA [6]. The data vizualisations and graphics were done in Mathematica 10.3 and R Studio software. The result as presented in Figure 5 below shows that as n increases from 40, 64, 107 to 272, the global error decreases. or rather tends to zero for all the proposed hybrid kernels considered. This is inherent in the properties of a good estimator which states that as the sample sizes increases, the estimator tends to the true value.

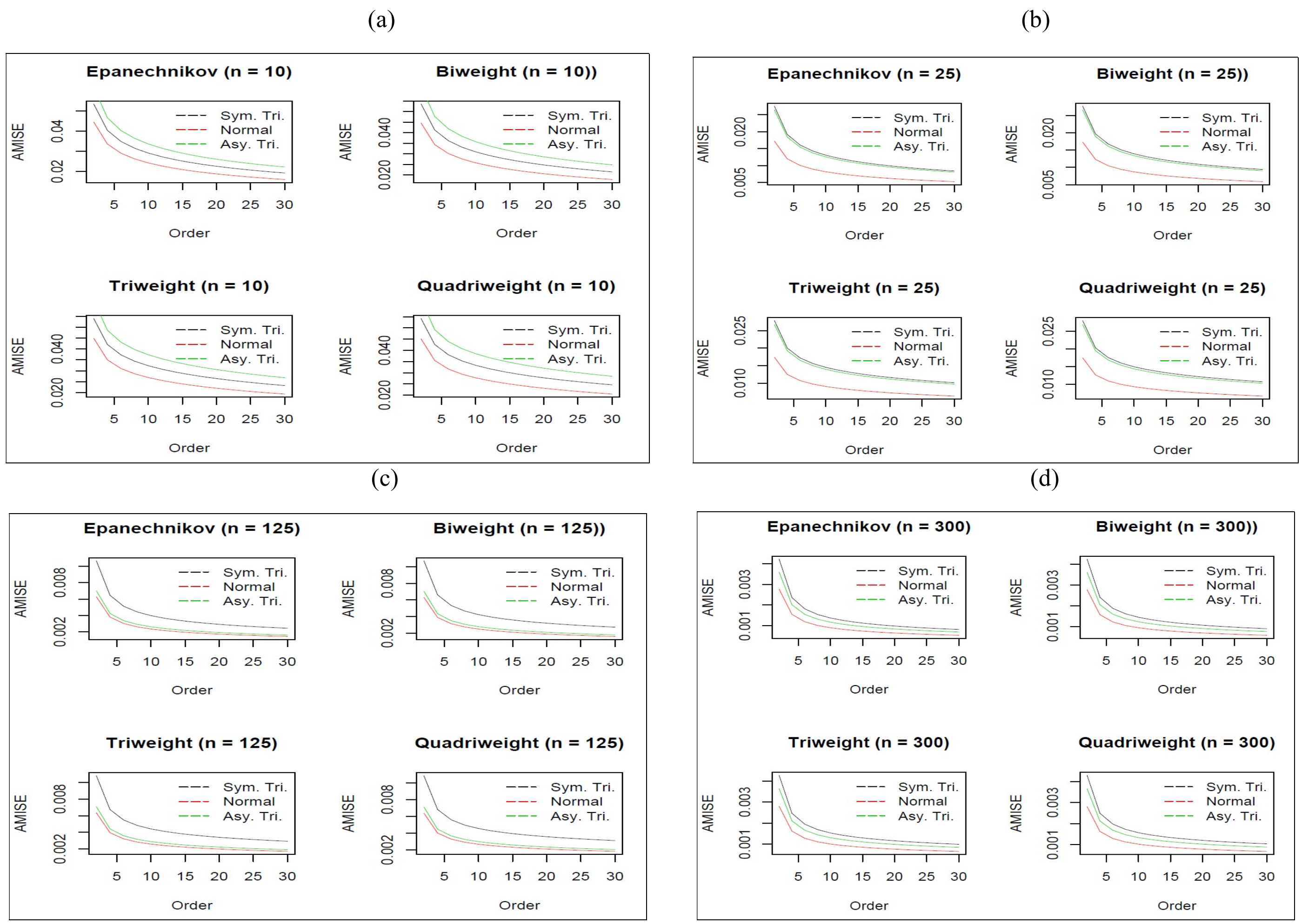

Figure 4: Plot showing the line graphs of global error by densities of the proposed hybrid polynomial kernels

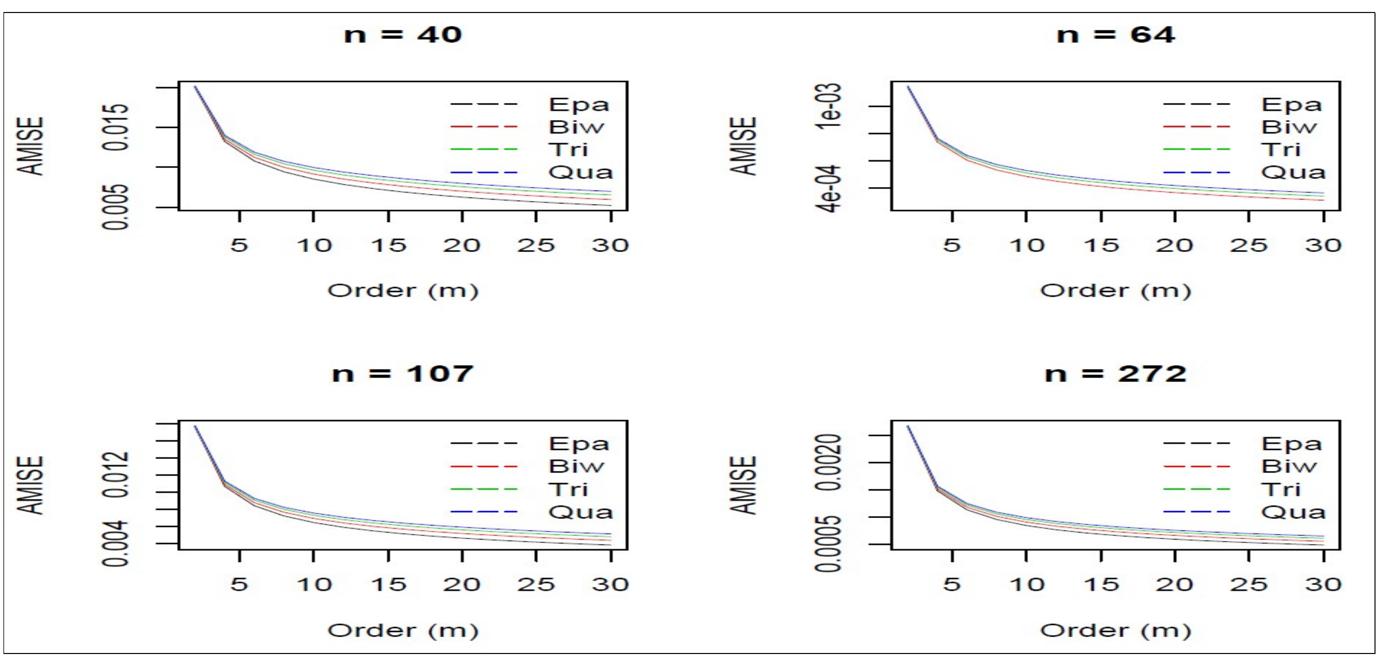

Figure 5: Plot showing the relationship between the proposed hybrid polynomial kernels and the classical polynomial kernels

V. SUMMARY AND CONCLUSION

In this paper, a new family of classical polynomial kernels of the hybrid type is developed. The generalized asymptotic mean integrated squared error and its corresponding generalized optimal bandwidth for this proposed family of kernels were equally obtained. The simulation results show that the proposed family of kernels have reduced AMISE as compared with the existing kernels. Real life applications via four data sets were also examined, the results also reveal that the proposed new family of kernels outperform the existing kernels.

REFERENCES

[1]. Afere, B. A. (2010). A family of multivariate higher-order hybrid polynomial kernels in kernel density estimation, Ph.D Thesis, University of Benin, Benin City, Nigeria.

[2]. Afere, B. A. and Alih, E. (2019). On the reduction of global error of multivariate higher-order product polynomial kernels, *Palestine Journal of Mathematics,* Vol. 8(2), 286 – 294.

[3]. Afere, B. A. and Oyegue, F. O. (2016). On the construction of the family of *d* – dimensional spherically symmetric polynomial kernels, *Palestine Journal of Mathematics,* Vol. 5(Special Issue: 1), 259 – 268.

[4]. Akaike, H. (1954). An approximation to the density functions, *Annals of Institute of Statistical Mathematics* **6**, 127 – 132.

[5]. Backurs, A., Indyk, P. and Wagner, T. (2019). Space and time efficient kernel density estimation in high dimensions, 33rd Conference on Neural Information Processing Syatems (NeurIPS 2019), Vancouver, Canada.

[6]. Bowman, A. W. and Azzalini, A. (1997). *Applied Smoothing Techniques for Data Analysis*: *The Kernel Approach with S-Plus Illustration*, Oxford University Press, Oxford, U.K.

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

[8]. Fix, E. and Hodges, J. L. (1951). Discriminatory nonpar-ametric discrimination: consistency properties, Report No.4, Project No. 21 – 29 – 004, USAF School of Aviation, Medicine, Randolph Field, Texas.

[9]. Humbert, P., Bars, L. B., Minvielle, L. and Vayatis, N. (2020). Robust kernel density estimation with mean-of-means principle, hal-02882092, https://hal.archives-ouvertes.fr/hal-02882092.

[10]. Ishiekwene, C. C. and Afere, B. A. E. (2001). Higher-order window width selectors for emphirical data, *Journal of Nigerian Statistical Association* **14**, 69 – 82.

[11]. Lee, J., Tran, D., Firat, O. and Cho, K. (2020). On the discrepancy between density estimation and sequence generation. https://arxiv.org/pdf/2002.07233.pdf.

[12]. López, J. C. C. and Sanz, M. J. (2020). Improving kernels methods for density estimation in random differential equations problems, *Mathematical and Computational Applications,* 25 – 33.

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

[14]. Ramsey, A. F. (2014). An application of kernel density estimation via diffusion to group yield insurance. Paper prepared for presentation at the Agricultural & Applied Economics Association's 2014 AAEA Annual Meetings, Minneapolis, MN, July 27 – 29, 2014, https/www.ageconsearch.umn.edu/bitstream/170173/2/diffusion_aaea.pdf.

[15]. Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function, *Annals of Mathematical Statistics* **27**, 832 – 837.

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

[17]. Shealter, S. J. (2004). *Density estimation,* Statistical Science, textbf19(4), 588 – 597.

[18]. Siloko, I. U., Siloko, E. A. and Ikpotokin, O. (2020). A mini review of dimensional effects on asymptotic mean integrated squared error and efficiencies of selected beta densities, *Jordan Journal of Mathematics and Statistics* (JJMS), **13**(3), 327 – 340.

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

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