On Efficiencies of Kernel Derivatives of the Beta Polynomial Kernels

I. U. Siloko[1]; O. Ikpotokin[2]; E. A. Siloko[3]

[1]Department of Mathematics and Computer Science,
Edo University, Iyamho, Nigeria.
e-mail: siloko.israel@edouniversity.edu.ng

[2]Department of Mathematics and Statistics,
Ambrose Alli University, Ekpoma, Nigeria.
e-mail: ikpotokinosayomore@yahoo.co.uk

[3]Department of Mathematics,
University of Benin, Benin City, Nigeria.
e-mail: myakpeski@yahoo.com

Professional Statisticians

***Abstract*—The efficiencies of the derivatives of beta polynomial family are the focus of this paper. Kernel density derivatives estimation play a very fundamental role in statistical data analysis especially for exploratory and visualization purposes. As a result of their wide applicability, studying their efficiencies as the order of the derivative increases is of a great importance. In this paper, the efficiencies of the derivatives of beta polynomial kernels are obtained from the efficiency of the classical second order kernels. The efficiencies of the derivatives of beta polynomial kernels for some higher powers were obtained and the results presented showed that the efficiency tends to increase as the powers of the beta polynomial kernels and the derivative order increases.**

Keywords- *Kernel Density Derivatives, Beta Polynomial Kernels, Efficiency, Asymptotic Mean Integrated Squared Error.*

I. Introduction

One of the nonparametric estimation techniques with wider applications is the kernel density estimator. The wide applicability of this estimator is due to the ease of its implementation [1]. Kernel density estimation is the construction of a probability density estimates from a given sample with few assumptions about the kernel density function. As a nonparametric estimator for exploration and visualization of data, its application has been extended to the machine learning community and also forms the building blocks for different semiparametric estimators [2, 3]. The derivatives of the kernel estimator possess vital statistical applications such as locating the local extrema and identification of the point of inflexion of a distribution [4]. Other areas where kernel derivatives can be applied are time series analysis [5], human growth data analysis [6], investigation of data using the submicroscopic nanoparticles property [7], chemical compositions inferences [8], estimation of the optimal smoothing parameter in kernel density estimation and regression estimation [9]. In parameter estimation and hypothesis testing density derivatives has a significant role to play, therefore proper estimation of the density derivatives from the set of observation is very important [10].

The major challenge confronting the implementation of kernel density estimation is the choice of smoothing parameter. In univariate kernel estimation, the problem of smoothing parameter selection is with less complexity when compared with the multivariate setting where there are different forms of smoothing parameterizations [11]. The choice of smoothing parameter is also very important in kernel density derivatives as the order of the derivative to be estimated increases [4]. Different authors have proposed several smoothing parameter selectors for kernel density derivatives in the univariate case [12, 13]

II. The Kernel Density Derivatives And Beta Polynomial Kernels

Kernel estimator is one of the popular nonparametric techniques in density estimation. The univariate kernel estimator is of the form

$$\hat{f}(\mathrm{x}) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{\mathrm{x}-\mathrm{X}_i}{h}\right), \qquad (1)$$

where $K(\cdot)$ is the kernel function, h is the smoothing parameter also called the bandwidth and n is the sample size. In most conditions particularly in scientific computing and data intensive applications, the data set X_i are observations or measurements obtained from real life. The kernel function is a non-negative function that satisfies the following conditions.

$$\begin{cases} \int K(x)dx = 1, \\ \int xK(x)dx = 0 \quad \text{and} \\ \int x^2K(x)dx = k_2(K) > 0 . \end{cases} \tag{2}$$

The first condition in Equation (2) means that any weighting function must integrate to unity, hence most kernel functions are probability density functions; the second condition simply states that the average of the kernel is zero, while the third condition means that the variance of the kernel is not zero [14].

The commonest optimality criterion used in selecting smoothing parameter in Equation (1) is the Asymptotic Mean Integrated Squared Error (AMISE) which is made up of two components. Asymptotic approximation of Equation (1) using Taylor's series expansion will yield the asymptotic integrated variance and the asymptotic integrated squared bias given by

$$AMISE = \frac{R(K)}{nh} + \frac{1}{4}\mu_2(\mathrm{K})^2\, h^4 R(f''), \tag{3}$$

where $R(K)$ is the roughness of kernel, $\mu_2(\mathrm{K})^2$ is the second moment of kernel and $R(f'') = \int f''(x)^2 dx$ is the roughness of the unknown probability density function [14, 15]. The minimum of the AMISE is the solution to the differential equation

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

Therefore, the smoothing parameter that minimizes the AMISE of the kernel estimator is

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

The derivative of the univariate kernel estimator is obtained by taking the derivative of the kernel density estimator in Equation (1). Assuming the kernel K is sufficiently differentiable r times, the rth density derivative of Equation (1) is given by [14, 15, 16]

$$\hat{f}^{(r)}(x) = \frac{d^r}{dx^r}\hat{f}(x) = \frac{1}{nh^{r+1}}\sum_{i=1}^{n} K^{(r)}\left(\frac{x - X_i}{h}\right), \tag{5}$$

where $K^{(r)}$ is the rth derivative of the kernel function K and is taken to be a symmetric probability density. In order for the estimator in Equation (5) to exist, $K^{(r)}$ must exist and not equal to zero. If the rth derivative of the kernel estimator in (5) is not equal to zero and continuously differentiable, it is also required that the $(r+1)th$ derivative is nonzero everywhere. The AMISE of the rth derivative of the kernel function provided the kernel K can be sufficiently differentiated is of the form

$$AMISE\left(\hat{f}^{(r)}(x)\right) = \frac{R\left(K^{(r)}\right)}{nh^{2r+1}} + \frac{1}{4}h^4\mu_2(\mathrm{K})^2\, R\left(f^{(r+2)}\right), \tag{6}$$

where $R\left(K^{(r)}\right)$ is the roughness of rth derivative of kernel, $\mu_2(\mathrm{K})^2$ is the second moment of kernel and $R\left(f^{(r+2)}\right)$ is the roughness of rth unknown probability density function [15]. Again, the smoothing parameter that minimized Equation (6) is given by

$$h^r_{AMISE} \approx \left[\frac{(2r+1)R\left(K^{(r)}\right)}{\mu_2(K)^2\, R\left(f^{(r+2)}\right)}\right]^{\left(\frac{1}{2r+5}\right)} \times n^{-\left(\frac{1}{2r+5}\right)}. \tag{7}$$

The order of the bias term of the rth derivative is the same as $O(h^4)$ but each new derivative order will introduce two additional powers to h in the variance term. The smoothing parameter for a kernel density derivative must be carefully selected because a good density estimator may not necessarily produce density derivative estimators that are good especially when the order of the derivatives increases [10]. The smoothing parameter in Equation (7) will result in the smallest value of the AMISE given by

$$AMISE^r = \left(\frac{2r+5}{4}\right) R\left(K^{(r)}\right)^{4/(2r+5)} \times \left[\frac{\mu_2(\mathrm{K})^2\, R\left(f^{(r+2)}\right)}{2r+1}\right]^{2r+1/2r+5} n^{-4/2r+5}. \tag{8}$$

In kernel density derivative estimation, the smoothing parameters are expected to be larger than kernel density estimation because the derivative of any function tends to be noisier than the function itself. The smoothing parameter that minimizes the AMISE for the first and second derivatives are of orders $O(n^{-1/7})$ and $O(n^{-1/9})$ and the AMISE are of orders $O(n^{-4/7})$ and $O(n^{-4/9})$ respectively [14].

The general pth kernel of the smooth beta polynomial kernel family for $p \geq 0$ with $\{t \in [-1,1]\}$ is of the form

$$K_p(t) = \frac{(2p+1)!!}{2^{p+1}p!}(1-t^2)^p, \tag{9}$$

Professional Statisticians

where $p = 0, 1, 2, \dots, \infty$ and the double factorial can be evaluated as $(2p+1)!! = (2p+1)(2p-1)\dots 5.3.1$. As the value of p increases from 0 to 3, we have the Uniform, Epanechnikov, Biweight and Triweight kernels which belong to the beta polynomial family [17]. The popular normal kernel is not strictly a member of this family but it is the limiting case when $p \to \infty$ [18]. In this class of kernels, the uniform kernel is the simplest kernel while the Epanechnikov kernel is regarded as the optimal kernel with respect to an error criterion, the mean integrated squared error (MISE). The popularity of this class of kernels is due to the desire to study their mathematical properties and the kernels with higher values of p and their estimates are smoother and also possess more derivatives.

III. The Efficiency Of Kernel Derivatives

The efficiency of the univariate symmetric kernel which is measured in comparison with the Epanechnikov kernel is of the form [9].

$$Eff(K) = \left(\frac{C(K_e)}{C(K)}\right)^{1/4} = \left(\frac{R(K_e)^4\mu_2(K_e)^2}{R(K)^4\mu_2(K)^2}\right)^{1/4}, \quad (10)$$

where $C(K) = R(K)^4\mu_2(K)^2$ is a constant of any given kernel and $C(K_e) = R(K_e)^4\mu_2(K_e)^2$ is the constant of Epanechnikov kernel. The Epanechnikov kernel produce smallest AMISE value in the case of the classical second order kernel and therefore, it is regarded as the optimal kernel with respect to the asymptotic mean integrated squared error.

The efficiency of the kernel derivative also requires the determination of the optimal kernel for its computation. The Epanechnikov kernel cannot be the optimal kernel in kernel derivatives because its second derivative is a constant meaning that it is not continuously differentiable. The efficiency of the rth derivative is given by

$$Eff(K^r) = \left(\frac{C(K_{opt}^r)}{C(K_{r+1})}\right)^{(2r+1)/4} = \left(\frac{R(K_{opt}^r)^{4/(2r+1)}\mu_2(K_{opt})^2}{R(K_{r+1}^r)^{4/(2r+1)}\mu_2(K_{r+1})^2}\right)^{(2r+1)/4}, \quad (11)$$

where $C(K_{opt}^r) = R(K_{opt}^r)^{4/(2r+1)}\mu_2(K_{opt})^2$ is the optimal kernel for the rth kernel derivative function and $C(K_{r+1}^r) = R(K_{r+1}^r)^{4/(2r+1)}\mu_2(K_{r+1})^2$ is the constant of any given $(r+1)th$ derivative kernel function. On simplification, Equation (11) can be written as

$$Eff(K^r) = \frac{R(K_{opt}^r)}{R(K_{r+1}^r)}\left(\frac{\mu_2(K_{opt})^2}{\mu_2(K_{r+1})^2}\right)^{(2r+1)/4}. \quad (12)$$

The optimal kernel for estimating the rth derivative has been shown by Muller [19] by solving for the minimum of $R(K^r)$, subject to the conditions $K_0 = 1$, $K_1 = 0$ and $K_2 < \infty$ and the solution obtained is $p = (r+1)th$ kernel from the beta polynomial kernels in Equation (9). This implies that when estimating the first derivative ($r = 1$), the optimal kernel is the Biweight ($p = 2$) but if we desire to estimate the second derivative ($r = 2$), the optimal kernel in this case is the Triweight ($p = 3$), and the optimality of the kernels goes on in that manner. In the computation of the efficiency of kernel derivatives, it is only the derivative of the rth roughness of the kernel function that is needed while the second moment of the kernel function is same irrespective of the derivative order to be estimated.

In computing the efficiency of kernel derivatives, two very important statistical quantities required are the roughness of kernel functions and its second moment as observed in Equation (10) and Equation (12). The rth roughness of a kernel function is given by

$$R(K^r) = \int K^r(t)^2 dt. \quad (13)$$

Also, the second moment of a kernel function is of the form

$$\mu_2(K) = \int t^2 K(t) dt. \quad (14)$$

In computing the statistical properties and efficiencies of the derivatives of smooth beta polynomial kernels, we shall specifically consider the power of p from one to seven only and the limiting case, that is the Gaussian kernel. The quantities in Equation (13) and Equation (14) are the parameters of interest in the determination of the efficiency of any kernel function.

IV. Results And Discussion

We consider the statistical properties of p for which $p = 0, 1, 2, \dots, 7$ and also for $p \to \infty$ which is the Gaussian kernel for the univariate kernel. For $p = 0, 1, 2, 3$, we have the Uniform, Epanechnikov, Biweight and Triweight kernels and they are of wide applications because they form the basis when discussing this class of kernels especially the Epanechnikov kernel in the computation of the efficiencies of other kernel functions of this family.

Professional Statisticians

Table 4.1 is the kernel functions obtained from the general polynomial family stated in Equation (9). Also presented in Table 4.1 are the statistical properties of the kernel functions; the roughnesses and second moments while their efficiencies are corrected to three decimal places and also expressed in percentage. The efficiency of the optimum kernel which is the Epanechnikov kernel is 100 % while the efficiencies of other kernel functions decrease with increase in the values of p.

Tables 4.2; 4.3; 4.4; 4.5; 4.6; are the efficiencies of the first to the fifth derivatives of the beta polynomial kernels while Table 4.7 shows the efficiencies of all the order of the kernel derivatives considered. In all the cases, the efficiencies of the kernel functions decrease as the values of the powers of p increases.

Table 4.1: Kernel Functions with Roughnesses, Moments and Efficiencies.

Kernel Functions	$R(K)$	$\mu_2(K)$	$Eff(K)$	$Eff(K)$ %
$K_0(t) = \frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{3}$	0.930	93.0 %
$K_1(t) = \frac{3}{4}(1-t^2)$	$\frac{3}{5}$	$\frac{1}{5}$	1.000	100 %
$K_2(t) = \frac{15}{16}(1-t^2)^2$	$\frac{5}{7}$	$\frac{1}{7}$	0.994	99.4 %
$K_3(t) = \frac{35}{32}(1-t^2)^3$	$\frac{350}{429}$	$\frac{1}{9}$	0.987	98.7 %
$K_4(t) = \frac{315}{256}(1-t^2)^4$	$\frac{2205}{2431}$	$\frac{1}{11}$	0.981	98.1 %
$K_5(t) = \frac{693}{512}(1-t^2)^5$	$\frac{4158}{4199}$	$\frac{1}{13}$	0.977	97.7 %
$K_6(t) = \frac{3003}{2048}(1-t^2)^6$	$\frac{26679}{25000}$	$\frac{1}{15}$	0.974	97.4 %
$K_7(t) = \frac{6435}{4096}(1-t^2)^7$	$\frac{1139}{1000}$	$\frac{1}{17}$	0.971	97.1 %
$K_\phi(t) = \frac{1}{\sqrt{2\pi}} exp\left(-\frac{t^2}{2}\right)$	$\frac{1}{2\sqrt{\pi}}$	1	0.951	95.1 %

Table 4.2: Roughnesses, Moments and Efficiencies of First Derivative.

Kernel Functions	$R(K)$	$\mu_2(K)$	$Eff(K)$	$Eff(K)$ %
$K_1(t)$	$\frac{3}{2}$	$\frac{1}{5}$	0.862	86.2 %
$K_2(t)$	$\frac{15}{7}$	$\frac{1}{7}$	1.000	100 %
$K_3(t)$	$\frac{35}{11}$	$\frac{1}{9}$	0.982	98.2 %
$K_4(t)$	$\frac{630}{143}$	$\frac{1}{11}$	0.958	95.8 %
$K_5(t)$	$\frac{24255}{4199}$	$\frac{1}{13}$	0.939	93.9 %
$K_6(t)$	$\frac{54054}{7429}$	$\frac{1}{15}$	0.924	92.4 %
$K_7(t)$	$\frac{66066}{7429}$	$\frac{1}{17}$	0.912	91.2 %
$K_\phi(t)$	$\frac{1}{4\sqrt{\pi}}$	1	0.820	82.0 %

Table 4.3: Roughnesses, Moments and Efficiencies of Second Derivative.

Kernel Functions	$R(K)$	$\mu_2(K)$	$Eff(K)$	$Eff(K)$ %
$K_2(t)$	$\frac{45}{2}$	$\frac{1}{7}$	0.830	83.0 %
$K_3(t)$	35	$\frac{1}{9}$	1.000	100 %
$K_4(t)$	$\frac{8505}{143}$	$\frac{1}{11}$	0.972	97.2 %
$K_5(t)$	$\frac{20790}{221}$	$\frac{1}{13}$	0.933	93.3 %
$K_6(t)$	$\frac{45045}{323}$	$\frac{1}{15}$	0.900	90.0 %
$K_7(t)$	$\frac{1459458}{7429}$	$\frac{1}{17}$	0.874	87.4 %
$K_\emptyset(t)$	$\frac{3}{8\sqrt{\pi}}$	1	0.681	68.1 %

Table 4.4: Roughnesses, Moments and Efficiencies of Third Derivative.

Kernel Functions	$R(K)$	$\mu_2(K)$	$Eff(K)$	$Eff(K)$ %
$K_3(t)$	$\frac{1575}{2}$	$\frac{1}{9}$	0.811	81.1 %
$K_4(t)$	$\frac{14175}{11}$	$\frac{1}{11}$	1.000	100 %
$K_5(t)$	$\frac{31185}{13}$	$\frac{1}{13}$	0.964	96.4 %
$K_6(t)$	$\frac{1351350}{323}$	$\frac{1}{15}$	0.912	91.2 %
$K_7(t)$	$\frac{682133}{100}$	$\frac{1}{17}$	0.867	86.7 %
$K_\emptyset(t)$	$\frac{15}{16\sqrt{\pi}}$	1	0.552	55.2 %

Table 4.5: Roughnesses, Moments and Efficiencies of Fourth Derivative.

Kernel Functions	$R(K)$	$\mu_2(K)$	$Eff(K)$	$Eff(K)$ %
$K_4(t)$	$\frac{99225}{2}$	$\frac{1}{11}$	0.798	79.8 %
$K_5(t)$	$\frac{1091475}{13}$	$\frac{1}{13}$	1.000	100 %
$K_6(t)$	$\frac{2837835}{17}$	$\frac{1}{15}$	0.958	95.8 %
$K_7(t)$	$\frac{31378096}{100}$	$\frac{1}{17}$	0.895	89.5 %
$K_\emptyset(t)$	$\frac{105}{32\sqrt{\pi}}$	1	0.440	44.0 %

© 2018, *A Publication of Professional Statisticians Society of Nigeria*

Table 4.6: Roughnesses, Moments and Efficiencies of Fifth Derivative.

Kernel Functions	$R(K)$	$\mu_2(K)$	$Eff(K)$	$Eff(K)$ %
$K_5(t)$	$\frac{9823275}{2}$	$\frac{1}{13}$	0.789	78.9 %
$K_6(t)$	8513505	$\frac{1}{15}$	1.000	100 %
$K_7(t)$	$\frac{88956906}{5}$	$\frac{1}{17}$	0.953	95.3 %
$K_\phi(t)$	$\frac{945}{64\sqrt{\pi}}$	1	0.347	34.7 %

Table 4.7: Efficiencies of Second Order Kernels Derivative.

Kernels	Derivative Orders (r)					
S/N	0	1	2	3	4	5
$K_0(t)$	0.930					
$K_1(t)$	1.000	0.862				
$K_2(t)$	0.994	1.000	0.830			
$K_3(t)$	0.987	0.982	1.000	0.811		
$K_4(t)$	0.981	0.958	0.972	1.000	0.798	
$K_5(t)$	0.977	0.939	0.933	0.964	1.000	0.789
$K_6(t)$	0.974	0.924	0.900	0.912	0.958	1.000
$K_7(t)$	0.971	0.912	0.874	0.867	0.895	0.953
$K_\phi(t)$	0.951	0.820	0.681	0.552	0.440	0.347

V. Conclusion

The focus of this paper is on efficiencies of derivatives of the beta polynomial kernels for higher powers of p and for the limiting case which is the Gaussian kernel. The results presented in Table 4.7 shows that the efficiencies of the kernel functions decreases as the power of p increases and with the Gaussian kernel being less efficient. Also noted in Table 4.7, is that the efficiencies tend to increase as the derivative order of the kernels increases until the optimum point and after that it starts decreasing except for the case of the Gaussian kernel whose efficiencies decrease as the derivative order increases.

However, it should be clearly noted that the choice of a kernel function should not be strictly based on its efficiency but mainly on the degree of its differentiability since kernels with higher powers of p tend to be smoother and possess more derivatives and this also suggests that their estimates will be smoother than those with fewer derivatives.

References

[1] Schauer, K., Duong, T., Gomes-Santos, C. and Goud, B. (2013). Studying Intracellular Trafficking Pathways with Probabilistic Density Maps. Methods in Cell Biology, 118:325–343.

[2] Raykar, V. C., Duraiswami, R. and Zhao, L. H. (2015). Fast Computation of Kernel Estimators. Journal of Computational and Graphical Statistics, 19 (1):205–220.

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

[4] Chacón, J. E. and Duong, T. (2013). Data-Driven Density Derivative Estimation, with Applications to Nonparametric Clustering and Bump Hunting. Electronic Journal of Statistics, 7:499–532.

[5] Rondonotti, V., Marron, J. S. and Park. C. (2007). SiZer for time series: A New Approach to the Analysis of Trends. Electronic Journal of Statistics, 1:268–289.

[6] Ramsay, J. O. and Silverman, B. W. (2002). Applied Functional Data Analysis. Springer-Verlag.

[7] Charnigo, R. and Srinivasan, C. (2011). Self-Consistent Estimation of Mean Response Functions and Their Derivatives. *Canadian Journal of Statistics*, 39(2):280–299.

[8] De Brabanter, K., De Brabanter, J. and De Moor, B. (2011). Nonparametric Derivative Estimation, NAIC, Gent, Belgium.

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

[10] Sasaki, H., Noh, Y. and Sugiyama, M. (2015). Direct Density Derivative Estimation and Its Application in KL-Divergence Approximation. Appearing in Proceedings of the 18th International Conference on Artificial Intelligence and Statistics (AISTATS), San Diego, CA, USA. JMLR: W&CP, Volume 38.

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

[12] Jones, M. C. (1992). Potential for Automatic Bandwidth Choice in Variations on Kernel Density Estimation. Statist. Probab. Letter, 13:351–356.

[13] Dobrovidov, A. V. and Rud'ko, I. M. (2010). Bandwidth Selection in Nonparametric Estimator of Density Derivatives by Smoothed Cross Validation Method. Autom. Remote Control, 71:209–224.

[14] Scott, D. W. (1992). Multivariate Density Estimation. Theory, Practice and Visualisation. Wiley, New York.

[15] Guidoum, A. S. (2015). Kernel Estimator and Bandwidth Selection for Density and its Derivatives. Department of Probabilities and Statistics, University of Science and Technology, Houari Boumediene, Algeria.

[16] Henderson, D. J. and Parmeter, C. F. (2015). Applied Nonparametric Econometrics. Cambridge University Press.

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

[18] Duong, T. (2015). Spherically Symmetric Multivariate Beta Family Kernels. Statistics and Probability Letters, 104, 141–145.

[19] Müller, H. G. (1984). Smooth Optimum Estimators of Densities, Regression Curves and Modes. Annals of Statistics, 12:766-774.

Professional Statisticians

Professional Statisticians

© 2018, *A Publication of Professional Statisticians Society of Nigeria*