Palestine Journal of Mathematics

Vol. 5(Special Issue: 1) (2016) , 259–268

On the Construction of the Family of d-Dimensional Spherically Symmetric Polynomial Kernels

B. A. Afere and F. O. Oyegue

Communicated by Ayman Badawi

MSC 2010 Classifications: 62G07

Keywords and phrases: Kernel density estimator, spherically symmetric kernel,generalised optimal bandwidth.

Abstract. In this paper, the family of generalised d-dimensional spherically symmetric polynomial kernels is presented. Based upon the general convention on deriving the asymptotic mean integrated squared error (AMISE) and its corresponding optimal bandwidth (h_{opt}); the generalised scheme for AMISE and h_{opt} as well as its generalised efficiency are derived. The advantage of this is that, it completely removes the rigour of first calculating both the second moment and L_2-norm of any kernel in the family before obtaining the h_{opt} as well as getting the AMISE and their corresponding efficiency of the kernel. In addition, it is useful for visual comparison of kernel estimates in a data-driven method.

1 Introduction

Density estimation methods are robust and elegant smoothing methods used for solving statistical (and its related) problems. To most statisticians, they are methods used for constructing an estimate of the true probability density function (pdf) from the observed data. That is, a random variable $\mathbf{X} \in \mathbb{R}^d$ has a density f if, for all Borel sets $\mathrm{A} \in \mathbb{R}^d$, $\int_{\mathrm{A}} f(\mathbf{x})d\mathbf{x} = P\{\mathbf{x} \in \mathrm{A}\}$. The major problem here is to estimate $f(\mathbf{x})$ from an independent identically distributed (i.i.d.) d-variate random sample $\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_n$ drawn from f [5].

There are basically two approaches to density estimation: the parametric and nonparametric approaches. The nonparametric methods consist of sophisticated alternatives to the conventional parametric models for studying multivariate data [11]. This is because the methods eliminate the need for model specification. That is, they can be used with arbitrary distributions and without the assumptions that the forms of the underlying densities are known [15].

In this paper, the focus is on one class of nonparametric density estimators which is kernel density estimators. The kernel density estimator is a more reliable statistical technique that deals with some of the problems associated with histogram which is discussed in Bowman and Azzalini [2] , Härdle, *et al* [10], Silverman [16]. Though, the procedure was first introduced in 1951 in an unpublished paper by Fix and Hodges [9], the first published work in this area of mathematical statistics is due to Rosenblatt [14], and the scope widened by Parzen [13]. Since then, and with the advent of computers, this area has been expanded greatly.

Given a random sample $\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_n$ drawn from a common density f, the d-dimensional multivariate kernel density estimator is given by:

$$\hat{f}_h(\mathbf{x}) = \frac{1}{nh^d}\sum_{i=1}^{n} K(\frac{\mathbf{x}-\mathbf{X}_i}{h}) \tag{1.1}$$

where $\mathbf{x} = (x_1, x_2, \cdots, x_d)^{\mathrm{T}}$ and $\mathbf{X}_i = (X_{i1}, X_{i2}, \cdots, X_{id})^{\mathrm{T}}$,$i = 1, 2, \cdots, n$. $\mathbf{H}$ is a symmetric positive definite $d \times d$ nonsingular matrix called the bandwidth matrix which generalizes the bandwidth h and K is a d-variate kernel function which satisfies $\int_{\mathbb{R}^d} K(\mathbf{x})d\mathbf{x} = 1$ [4]. We make use of the parameterisation $(\mathbf{H} = h\mathbf{I}_d)$ given by Cacoullos [3] rather than the range of bandwidths suggested by Epanechnikov [7] and Deheuvels [4], and thus the multivariate kernel density estimator in (1.1) becomes:

$$\hat{f}_h(\mathbf{x}) = \frac{1}{nh^d}\sum_{i=1}^{n} K(\frac{\mathbf{x}-\mathbf{X}_i}{h}) \tag{1.2}$$

where $\mathbf{x}$, $\mathbf{X}_i$ and K are as given in (1.1) and h is the univariate smoothing parameter [3]. To use the parameterisation $\mathbf{H} = h\mathbf{I}_d$ effectively, the components of the data vector should be commensurate. This can be achieved by using appropriate transformation in the data set [6, 16, 19]. This transformation involves either pre-scaling each axis or pre-whitening the data. A detailed study of this can be found in [8].

Many of the studies in density estimation have centred on the univariate kernel density estimation [16]. This work, however, focuses on the multivariate kernels with emphasis on the classical spherically symmetric d-dimensional polynomial kernels. The basic motivation for considering (1.2) is that it enables one to obtain a closed form expression for the asymptotic mean integrated squared error (AMISE) as well as its corresponding optimal bandwidths which enable us to derive the generalised AMISE, optimal bandwidth and the generalised efficiency.

The global accuracy used in measuring (1.2) is the mean integrated squared error (MISE) which is given as:

$$\mathrm{MISE}\hat{f}_{\mathbf{H}}(\mathbf{x}) = \mathbb{E}\int_{\mathbb{R}} \left(\hat{f}_{\mathbf{H}}(\mathbf{x}) - f(\mathbf{x})\right)^2 d\mathbf{x} \tag{1.3}$$

Thus, from [17], the expression (1.3) can be put into the form:

$$\mathrm{MISE}\hat{f}_{\mathbf{H}}(\mathbf{x}) = \int_{\mathbb{R}} \mathbb{E}[\hat{f}_{\mathbf{H}}(\mathbf{x}) - f_{\mathbf{H}}(\mathbf{x})]^2 d\mathbf{x} + \int_{\mathbb{R}} \mathrm{Var}\hat{f}_{\mathbf{H}}(\mathbf{x})d\mathbf{x} \tag{1.4}$$

With the introduction above, the remaining part of this paper is organized as follows: Section 2 centres on the construction of the family of the d-dimensional spherically symmetric polynomial kernels. Section 3 covers the derivation of the generalised global error schemes. Section 4 deals with how the generalised efficiency scheme for any d-dimensional spherically symmetric polynomial kernel is obtained. While Section 5 focuses on numerical verification of results, Section 6 sheds light on the discussion of results and equally gives the concluding remarks.

2 The family of d-dimensional spherically symmetric polynomial kernels

The family of d-dimensional spherically symmetric polynomial kernels is heuristically deduced as follows: The d-dimensional spherically symmetric kernel function of the classical Epanechnikov kernel is:

$$K^S_{e,d}(\mathbf{t}) = \left\{ \begin{array}{ll} \frac{d(d+2)\Gamma(\frac{d}{2})}{4\pi^{\frac{d}{2}}}(1-\mathbf{t}^{\mathrm{T}}\mathbf{t}) & , \ |\mathbf{t}^{\mathrm{T}}\mathbf{t}| \le 1 \\ 0 & , \ \text{elsewhere} \end{array} \right\}. \tag{2.1a}$$

or

$$K^S_{1,\,d}(\mathbf{t}) = \begin{cases} \frac{(d+2)\Gamma(\frac{d}{2}+1)}{2\pi^{\frac{d}{2}}}(1-\mathbf{t}^{\mathrm{T}}\mathbf{t}) & , \ |\mathbf{t}^{\mathrm{T}}\mathbf{t}| \le 1 \\ 0 & , \ \text{elsewhere} \end{cases}. \tag{2.1b}$$

where $d \ge 1$ and $\Gamma(\cdot)$ is a gamma function (see Afere and Ishiekwene [1] for the construction of (2.1a) and (2.1b). When $d = 1$, we obtain the classical 1-dimensional Epanechnikov kernel density function. When $d = 2$, we obtain the 2 - dimensional spherically symmetric Epanechnikov kernel function ([17]; pp 105). Also, the d-dimensional spherically symmetric kernel function of the classical biweight kernel is:

$$K^S_{b,d}(\mathbf{t}) = \begin{cases} \frac{(d+2)(d+4)\Gamma(\frac{d}{2}+1)}{8\pi^{\frac{d}{2}}}(1-\mathbf{t}^{\mathrm{T}}\mathbf{t})^2 & , \ |\mathbf{t}^{\mathrm{T}}\mathbf{t}| \le 1 \\ 0 & , \ \text{elsewhere} \end{cases} \tag{2.2a}$$

or

$$K^S_{2,d}(\mathbf{t}) = \begin{cases} \frac{(d+4)\Gamma(\frac{d}{2}+2)}{4\pi^{\frac{d}{2}}}(1-\mathbf{t}^{\mathrm{T}}\mathbf{t})^2 & , \ |\mathbf{t}^{\mathrm{T}}\mathbf{t}| \le 1 \\ 0 & , \ \text{elsewhere} \end{cases} \tag{2.2b}$$

where $d \ge 1$ and $\Gamma(\cdot)$ is a gamma function (see Afere and Ishiekwene [1] for the construction of (2.2a) and (2.2b)). When $d = 1$, we obtain the classical 1-dimensional biweight kernel density function. When $d = 2$, we obtain the 2-dimensional spherically symmetric biweight kernel function. Based on (2.1b) and (2.2b), the proposed spherically symmetric kernel functions of the classical triweight kernel and classical quadriweight kernels are respectively:

$$K^S_{3,d}(\mathbf{t}) = \begin{cases} \frac{(d+6)\Gamma(\frac{d}{2}+3)}{12\pi^{\frac{d}{2}}}(1-\mathbf{t}^{\mathrm{T}}\mathbf{t})^3 & , \ |\mathbf{t}^{\mathrm{T}}\mathbf{t}| \le 1 \\ 0 & , \ \text{elsewhere} \end{cases} \tag{2.3}$$

$$K^S_{4,d}(\mathbf{t}) = \begin{cases} \frac{(d+8)\Gamma(\frac{d}{2}+4)}{48\pi^{\frac{d}{2}}}(1-\mathbf{t}^{\mathrm{T}}\mathbf{t})^4 & , \ |\mathbf{t}^{\mathrm{T}}\mathbf{t}| \le 1 \\ 0 & , \ \text{elsewhere} \end{cases} \tag{2.4}$$

where $d \ge 1$ and $\Gamma(\cdot)$ is a gamma function. Thus, the proposed generalised family of d-dimensional spherically symmetric polynomial kernels is:

$$K^S_{d,p}(\mathbf{t}) = \begin{cases} \frac{(d+2p)\Gamma(\frac{d}{2}+p)}{2(p)!\pi^{\frac{d}{2}}}(1-\mathbf{t}^{\mathrm{T}}\mathbf{t})^p & , \ |\mathbf{t}^{\mathrm{T}}\mathbf{t}| \le 1 \\ 0 & , \ \text{elsewhere} \end{cases} \tag{2.5}$$

When $p = 0, 1, 2, 3$, we have respectively the d-dimensional spherically symmetric uniform, Epanechnikov, biweight and triweight kernels. As $p \to \infty$, (2.5) becomes a d - dimensional spherically symmetric Gaussian kernel. The constants of versions of (2.1) [both a and b] and (2.2) [both a and b] are presented in Table 1.

Table 1. Constants of versions of Epanechnikov and Biweight kernels

Kernel / Dimension	Epanechnikov		Biweight	
	$\frac{d(d+2)\Gamma(\frac{d}{2})}{4\pi^{\frac{d}{2}}}$	$\frac{(d+2)\Gamma(\frac{d}{2}+1)}{2\pi^{\frac{d}{2}}}$	$\frac{(d+2)(d+4)\Gamma(\frac{d}{2}+1)}{8\pi^{\frac{d}{2}}}$	$\frac{(d+4)\Gamma(\frac{d}{2}+2)}{4\pi^{\frac{d}{2}}}$
1	$\frac{3}{4}$	$\frac{3}{4}$	$\frac{15}{16}$	$\frac{15}{16}$
2	$\frac{2}{\pi}$	$\frac{2}{\pi}$	$\frac{3}{\pi}$	$\frac{3}{\pi}$
3	$\frac{15}{8\pi}$	$\frac{15}{8\pi}$	$\frac{105}{32\pi}$	$\frac{105}{32\pi}$
4	$\frac{6}{\pi^2}$	$\frac{6}{\pi^2}$	$\frac{12}{\pi^2}$	$\frac{12}{\pi^2}$
5	$\frac{105}{16\pi^2}$	$\frac{105}{16\pi^2}$	$\frac{945}{64\pi^2}$	$\frac{945}{64\pi^2}$

3 The generalised global error schemes

The AMISE is one of the most significant segments in bandwidth selection. By using symmetric kernels, the AMISE and its corresponding optimal bandwidth for the multivariate kernel density estimator are derived. In this section, we state the global error schemes of the family of d-dimensional spherically symmetric polynomial kernels in (2.5). We first state the basic definitions before stating the theorem, which will play a significant role in the rest of this paper.

3.1 Basic definitions

Definition 3.1: If $f(\mathbf{x})$ is the $d\times d$ normal (Gaussian) distribution which is pairwise uncorrelated (i.e. $\mathbf{x}\sim N(0\,,\,\Sigma))where\Sigma$ is the covariance matrix, then the Hessian matrix of $f(\mathbf{x})$ denoted by $\mathcal{H}$ is:

$$\mathcal{H}=\begin{pmatrix} \frac{\partial^2 f(\mathbf{x})}{\partial x_1^2} & 0 & \cdots & 0 & 0\\ 0 & & & & 0\\ \vdots & & \ddots & & \vdots\\ 0 & & & & 0\\ 0 & 0 & \cdots & 0 & \frac{\partial^2 f(\mathbf{x})}{\partial x_d^2}\end{pmatrix} \tag{3.1}$$

Definition 3.2: The trace of the Hessian matrix $\mathcal{H}$ in (10) denoted by ∇^2 is given by:

$$\nabla^2 f(\mathbf{x})=\frac{\partial^2 f(\mathbf{x})}{\partial x_1^2}+\frac{\partial^2 f(\mathbf{x})}{\partial x_2^2}+\frac{\partial^2 f(\mathbf{x})}{\partial x_3^2}+\cdots+\frac{\partial^2 f(\mathbf{x})}{\partial x_d^2} \tag{3.2}$$

Definition 3.3: Let f be a d-variate function and $\mathbf{t}$ be a sequence of $d\times 1$ vectors with all components tending to zero. Also, let $\mathcal{D}_f(\mathbf{x})$ be the vector of first-order partial derivatives of f and $\mathcal{H}$ be the Hessian matrix in (3.1), the $d\times d$ matrix having $(\ i\ ,\ j\ )$ entry equal to $\frac{\partial^2 f(\mathbf{x})}{\partial x_i \partial x_j}=0\ ;\ i\neq j$. Then, assuming that all entries of $\mathcal{H}$ are continuous in a neighbourhood of $\mathbf{x}$, we have:

$$f(\mathbf{x}\,+\mathbf{t})=f(\mathbf{x})+\sum_{n=1}^{m}\{\frac{1}{(2n-1)!}(\mathbf{t}^{\mathrm{T}}\mathcal{D}_f)(\mathbf{t}^{\mathrm{T}}\mathcal{H}_f\mathbf{t})^{n-1}+\frac{1}{(2n)!}(\mathbf{t}^{\mathrm{T}}\mathcal{H}_f\mathbf{t})^{n}\}+o(\mathbf{t}^{\mathrm{T}}\mathbf{t}) \tag{3.3}$$

Theorem 3.4: *Suppose f is continuously differentiable and $K_m(\mathbf{t})$ is any differentiable multivariate spherically symmetric polynomial kernel that is parameterised by $\mathbf{H}=h\mathbf{I}_d$, which satisfies*:

$$U_{im}=\int_{\mathbb{R}^d}(\mathbf{t}^{\mathrm{T}}\mathbf{t})^{i}K_m(\mathbf{t})d\mathbf{t},\ \ 0\leq i\leq m,\ \ m=1,2,\cdots,<\infty \tag{3.4}$$

such that:

$$\left.\begin{array}{ll} i. \ \ U_{im}=1_d & ,\ i=0\\ ii. \ \ U_{im}=0_d & ,\ 0<i\leq m-1\\ iii. \ \ U_{im}=K_{2m}\cdot \mathrm{I}_d & ,\ i=m \end{array}\right\} \tag{3.5}$$

If in addition, $\int_{\mathbb{R}^d}(\nabla^{2m}f(\mathbf{x}))^2 d\mathbf{x}\ =\ ||\nabla^{2m}f||_2^2$ is the generalised constant of squared integrable multivariate Gaussian distribution with means $\mu\ =\ (\mu_1,\mu_2,\cdots,\mu_d)^T$ equals zero and pair-wise uncorrelated covariance matrix $\sum\ =$ diag$(\sigma_1^2,\sigma_2^2,\cdots,\sigma_d^2)$, then the optimal asymptotic bandwidth parameter is given by:

$$h_{opt}\cong(\frac{((2m)!)^2}{(4m)}\,||K||_2^2\,\frac{d}{[K_{2m}\cdot \mathrm{I}_d]^2\,||\nabla^{2m}f||_2^2})^{\frac{1}{d+4m}}\,n^{-\frac{1}{d+4m}} \tag{3.6}$$

and the corresponding optimal asymptotic mean integrated squared error is:

$$\mathrm{AMISE}\hat{f}_h(\mathbf{x})\cong(\frac{d+4m}{d(4m)})[\frac{(4m)}{((2m)!)^2}]^{\frac{d}{d+4m}}(d\cdot||K||_2^2)^{\frac{4m}{d+4m}}((K_{2m}\cdot \mathrm{I}_d)^2)^{\frac{d}{d+4m}}(\left\|\nabla^{2m}f\right\|_2^2)^{\frac{d}{d+4m}}n^{-\frac{4m}{d+4m}} \tag{3.7}$$

where K_{2m}^S and $||K^S||_2^2$ are given in Proposition 3.5.
Proof: Now, the AMISE for (1.4) is given by:

$$\mathrm{AMISE}\hat{f}_{\mathbf{H}}(\mathbf{x})=\mathrm{AISB}\hat{f}_{\mathbf{H}}(\mathbf{x})+\mathrm{AIV}\hat{f}_{\mathbf{H}}(\mathbf{x}) \tag{3.8}$$

where

$$\mathrm{AISB}\hat{f}_{\mathbf{H}}(\mathbf{x})=\int_{\mathbb{R}^d}\mathrm{Bias}^2\hat{f}_{\mathbf{H}}(\mathbf{x})d\mathbf{x} \tag{3.9}$$

and

$$\mathrm{AIV}\hat{f}_{\mathbf{H}}(\mathbf{x})=\int_{\mathbb{R}^d}\mathrm{Var}\hat{f}_{\mathbf{H}}(\mathbf{x})d\mathbf{x} \tag{3.10}$$

The bias in (3.9) is given by:

$$\mathrm{Bias}\hat{f}_{\mathbf{H}}(\mathbf{x})=\mathbb{E}\hat{f}_{\mathbf{H}}(\mathbf{x})-f(\mathbf{x}) \tag{3.11}$$

If we substitute (1.2) into (3.11) and simplify, we have:

$$\mathrm{Bias}\hat{f}_{\mathbf{H}}(\mathbf{x})=\int_{\mathbb{R}^d}K(\mathbf{t})f(\mathbf{x}-\mathbf{t}\mathbf{H}^{\frac{1}{2}})d\mathbf{t}-f(\mathbf{x}) \tag{3.12}$$

Thus, on using (3.1) and (3.3) and on using the moment conditions of (3.5), the bias in (3.12) reduces to:

$$\mathrm{Bias}\hat{f}_{\mathbf{H}}(\mathbf{x})=\frac{1}{(2m)!}\mathrm{tr}\{(\int_{\mathbb{R}^d}(\mathbf{t}^{\mathrm{T}}\mathbf{t})^{m}K(\mathbf{t})d\mathbf{t})(\mathbf{H}^{\frac{1}{2}}\mathcal{H}_f\mathbf{H}^{\frac{1}{2}})^{m}\}+o((\mathbf{t}^{\mathrm{T}}\mathbf{t})\mathbf{H}^{\frac{1}{2}})^{m} \tag{3.13}$$

Using Definition (1.2), the bias becomes:

$$\text{Bias}\hat{f}_{\mathbf{H}}(\mathbf{x}) = \frac{1}{(2m)!}[\int_{\mathbb{R}^d}(\mathbf{t}^{\mathrm{T}}\mathbf{t})^m K(\mathbf{t})d\mathbf{t}][\nabla^2 f(\mathbf{x})]^m[\text{tr}(\mathbf{H})^m] + o((\mathbf{t}^{\mathrm{T}}\mathbf{t})\mathbf{H})^m \tag{3.14}$$

Squaring both sides of (3.14) and using the bandwidth matrix $\mathbf{H} = h\mathrm{I}_d$ and Definition 3.2 and then substituting into (3.9), the AISB is obtained as:

$$\text{AISB}\hat{f}_h(\mathbf{x}) = \frac{1}{((2m)!)^2}[K_{2m}\cdot \mathrm{I}_d]^2[\left\|\nabla^{2m}f\right\|_2^2]^2[h^{4m}] \tag{3.15}$$

where

$$K_{2m} = \int_{\mathbb{R}^d}(\mathbf{t}^{\mathrm{T}}\mathbf{t})^m K(\mathbf{t})d\mathbf{t} \tag{3.16}$$

The variance term is given by:

$$Var\hat{f}_{\mathbf{H}}(\mathbf{x}) = \mathbb{E}\hat{f}^2_{\mathbf{H}}(\mathbf{x}) - \mathbb{E}^2\hat{f}_{\mathbf{H}}(\mathbf{x}) \tag{3.17}$$

On substituting (1.2) into (3.17) and using the necessary assumptions as in the case of bias, we have:

$$Var\hat{f}_{\mathbf{H}}(\mathbf{x}) = (n\mathbf{H}^{\frac{1}{2}})^{-1}\int_{\mathbb{R}^d}K^2(\mathbf{t})f(\mathbf{x}-\mathbf{t}\mathbf{H}^{\frac{1}{2}})d\mathbf{t} - [(n\mathbf{H}^{\frac{1}{2}})^{-1}\int_{\mathbb{R}^d}K^2(\mathbf{t})f(\mathbf{x}-\mathbf{t}\mathbf{H}^{\frac{1}{2}})d\mathbf{t}]^2 \tag{3.18}$$

Hence, following similar algebraic substitution and Taylor's series expansion argument as in the case of bias and using all the assumptions, (3.18) becomes:

$$Var\hat{f}_{\mathbf{H}}(\mathbf{x}) = (n\mathbf{H}^{\frac{1}{2}})^{-1}\|K\|_2^2 f(\mathbf{x}) \tag{3.19}$$

where

$$\|K\|_2^2 = \int_{\mathbb{R}^d}K^2(\mathbf{t})d\mathbf{t} \tag{3.20}$$

is the d-dimensional squared L_2-norm of $K(\mathbf{t})$. On substituting (3.19) into (3.11), the resulting equation becomes:

$$\text{AIV}\hat{f}_{\mathbf{H}}(\mathbf{x}) = \int_{\mathbb{R}^d}Var\hat{f}_{\mathbf{H}}(\mathbf{x})d\mathbf{x} = (n\mathbf{H}^{\frac{1}{2}})^{-1}\|K\|_2^2 \tag{3.21}$$

Also, using the parameterisation $\mathbf{H} = h^2\mathrm{I}_d$ and Definition 3.2, (3.21) becomes:

$$\text{AIV}\hat{f}_h(\mathbf{x}) = (nh^d)^{-1}\|K\|_2^2 \tag{3.22}$$

Plugging back (3.15) and (3.22) into (3.8), we have:

$$\text{AMISE}\hat{f}_h(\mathbf{x}) = \frac{h^{4m}}{((2m)!)^2}[K_{2m}\cdot \mathrm{I}_d]^2\left\|\nabla^{2m}f\right\|_2^2 + (nh^d)^{-1}\|K\|_2^2 \tag{3.23}$$

Thus, on optimizing (3.23), the generalised asymptotic optimal bandwidth is:

$$h_{opt} \cong (\frac{((2m)!)^2}{(4m)}\|K\|_2^2\frac{d}{[K_{2m}\cdot \mathrm{I}_d]^2\|\nabla^{2m}f\|_2^2})^{\frac{1}{d+4m}} n^{-\frac{1}{d+4m}} \tag{3.24}$$

On substituting (3.24) into (3.23), then AMISE is:

$$\text{AMISE}\hat{f}_h(\mathbf{x}) \cong (\frac{d+4m}{d(4m)})[\frac{(4m)}{((2m)!)^2}]^{\frac{d}{d+4m}}(d\cdot\|K\|_2^2)^{\frac{4m}{d+4m}}((K_{2m}\cdot \mathrm{I}_d)^2)^{\frac{d}{d+4m}}(\left\|\nabla^{2m}f\right\|_2^2)^{\frac{d}{d+4m}}n^{-\frac{4m}{d+4m}} \tag{3.25}$$

The proposition below deals with the second moment and squared L_2-norm of any d-dimensional spherically symmetric polynomial kernel.

Proposition 3.5: *Under the same conditions in Theorem 3.4, if $K^S_{d,p}$ is the proposed family of d-dimensional spherically symmetric polynomial kernels in (2.5), then the second moment is given by*:

$$K^S_{2m} = \begin{cases} \dfrac{d}{2m+d}\displaystyle\prod_{i=1}^{p}\frac{(d+2i)}{(2m+d+2i)} & , \ p<\infty \\ \dfrac{2^{m+\frac{d}{2}-\frac{1}{2}}}{\Gamma(\frac{d}{2})}\Gamma(m+\dfrac{d}{2}) & , \ p\to\infty \end{cases} \tag{3.26}$$

and the squared L_2-norm is:

$$||K^S||_2^2 = \begin{cases} (2p)!\dfrac{\pi^{\frac{d}{2}}}{\Gamma(\frac{d}{2}+2p+1)}\left(\dfrac{(d+2p)\Gamma(\frac{d}{2}+p)}{2(p)!\pi^{\frac{d}{2}}}\right)^2 & , \ p<\infty \\ \dfrac{1}{2^d\pi^{\frac{d}{2}}} & , \ p\to\infty \end{cases} \tag{3.27}$$

Proof: Substituting (2.5) into (3.16), we have:

$$\begin{aligned} K_{2m} &= \int_{\mathbb{R}^d}\sum_{i=1}^d t_i^{2m}K(\mathbf{t})d\mathbf{t} = \int_{\mathbb{R}^d}\sum_{i=1}^d t_i^{2m}\frac{(d+2p)\Gamma(\frac{d}{2}+p)}{2(p)!\pi^{\frac{d}{2}}}(\mathbf{t}^T\mathbf{t})^p d\mathbf{t} \\ &= \frac{(d+2p)\Gamma(\frac{d}{2}+p)}{2(p)!\pi^{\frac{d}{2}}}\int_0^1 r^{2m+d-1}(1-r^2)^p dr\int_0^{\pi}\sin^{d-2}\theta_1 d\theta_1\times \\ &\quad \int_0^{\pi}\sin^{d-3}\theta_2 d\theta_2\times\cdots\times\int_0^{\pi}\sin\theta_{d-2}d\theta_{d-2}\int_0^{2\pi}d\theta_{d-1} \end{aligned} \tag{3.28}$$

Therefore:

$$K_{2m} = \frac{d}{2m+d}\prod_{i=1}^{p}\frac{(d+2i)}{(2m+d+2i)}$$

This proves the first part of (3.26). However, as $p \to \infty$, the second part of (3.26) is obtained. Hence:

$$K_{2m} = \begin{cases} \dfrac{d}{2m+d}\displaystyle\prod_{i=1}^{p}\frac{(d+2i)}{(2m+d+2i)} & , \; p < \infty \\ 2^{m+\frac{d}{2}-\frac{1}{2}}\pi^{-\frac{d}{2}}\Gamma(m+\dfrac{d}{2}) & , \; p \to \infty \end{cases}$$

Also, on substituting (2.5) into (3.20), we have:

$$\begin{aligned} ||K||_2^2 &= \int_{\mathbb{R}^d}(K(\mathbf{t}))^2 d\mathbf{t} \\ &= \left(\frac{(d+2p)\Gamma(\frac{d}{2}+p)}{2(p)!\pi^{\frac{d}{2}}}\right)^2 \int_0^1 r^{d-1}\left((1-r^2)^p\right)^2 dr \int_0^{\pi}\sin^{d-2}\theta_1 d\theta_1 \times \\ &\qquad \int_0^{\pi}\sin^{d-3}\theta_2 d\theta_2 \times \cdots \times \int_0^{\pi}\sin\theta_{d-2}d\theta_{d-2}\int_0^{2\pi}d\theta_{d-1} \end{aligned}$$

Therefore:

$$||K||_2^2 = (2p)!\frac{\pi^{\frac{d}{2}}}{\Gamma(\frac{d}{2}+3)}\left(\frac{(d+2p)\Gamma(\frac{d}{2}+p)}{2(p)!\pi^{\frac{d}{2}}}\right)$$

This proves the first part of (3.27). However, as $p \to \infty$, the second part of (3.27) is obtained. Therefore:

$$||K||_2^2 = \begin{cases} (2p)!\dfrac{\pi^{\frac{d}{2}}}{\Gamma(\frac{d}{2}+3)}\left(\dfrac{(d+2p)\Gamma(\frac{d}{2}+p)}{2(p)!\pi^{\frac{d}{2}}}\right) & , \; p < \infty \\ \dfrac{1}{2^d\pi^{\frac{d}{2}}} & , \; p \to \infty \end{cases}$$

The generalised second moment and the squared L_2-norm of the family of d-dimensional spherically symmetric polynomial kernels above have completely removed the rigour of first calculating the second moment and L_2-norm of any kernel in this family before calculating its optimal asymptotic bandwidth and the corresponding optimal AMISE.

4 The generalised efficiency scheme

Extensive work has been done on univariate kernels [16]. However, Wand and Jones [17] did an insight into the efficiency of the second-order multivariate kernels. Their method is based on taking the ratio of the spherically symmetric kernel relative to the product kernel. Hence, we develop a method that is different from the technique adopted by Wand and Jones [17], even though our method is motivated by the work of Silverman [16].

We focus our attention on the derivation of the general scheme in which the efficiency of any d-dimensional spherically symmetric polynomial kernel can be calculated. Kernel efficiency is measured in comparison to Epanechnikov kernel. That is, it is the ratio of the asymptotic mean integrated squared error of Epanechnikov kernel to the asymptotic mean integrated squared error of any other symmetric kernels. The choice of the Epanechnikov kernel is that it yields the minimum AMISE [16, 17].

The efficiency of any symmetric kernel in the sense of equation (1.2) is defined by:

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

[16] where $C(K_{sk}) = (K_2)^{\frac{2}{5}}(\int K(t)^2 dt)^{\frac{4}{5}}$ is a constant of any given kernel under discussion and $C(K_e)$ is the Epanechnikov kernel constant. Thus, by drawing inspiration from (4.1), the general expression for the efficiency of multivariate kernels based on the spherically symmetric kernel approach is defined by:

$$\mathrm{Eff}^{2m}(K_{sk}^S) = \left(\frac{C^{2m}(K_e^S)}{C^{2m}(K_{sk}^S)}\right)^{\frac{d+4m}{4m}} \tag{4.2}$$

where $C^{2m}(K_{sk}^S) = (K_{2m}^S)^{\frac{2}{d+4m}}(\|K^S\|_2^2)^{\frac{4m}{d+4m}}$ is the generalised higher-order constants of any d-dimensional multivariate spherically symmetric kernel. K_{sk}^S and $\|K^S\|_2^2$ are as defined in (3.16) and (3.20) respectively. In addition, $C^{2m}(K_e^S)$ in equation (4.2) is the generalised higher-order constant of the d-dimensional spherically symmetric Epanechnikov kernel. In what follows, we state the generalised efficiency scheme, which is presented in Theorem 4.1.

Theorem 4.1: *Under the same conditions in Theorem 3.4, if $K_{d,p}^S$ is the proposed family of d-dimensional spherically symmetric polynomial kernels in (2.3), then the generalised $d-$dimensional spherically symmetric polynomial kernel in the sense* *of (1.2) is given by*:

$$Eff^{2m}(K^S_{d,p}) = \begin{cases} \left(\dfrac{d+2}{2m+d+2} \left(\prod_{i=1}^{p} \dfrac{d+2i}{2m+d=2i} \right)^{-1} \right)^{\frac{1}{2m}} \times \\ \qquad \left(\dfrac{2\Gamma(\frac{d}{2}+3)}{(2p)!\Gamma(\frac{d}{2}+2p+1)} \left(\dfrac{p!(d(d+2))\Gamma(\frac{d}{2})}{2(d+2p)\Gamma(\frac{d}{2}+p)} \right)^2 \right), \ p < \infty \\ \left(\dfrac{\Gamma(\frac{d}{2})(d+2)}{2^{m+\frac{d}{2}-\frac{1}{2}}(2m+d+2)\Gamma(\frac{d}{2}+m)} \right)^{\frac{1}{2m}} \times \\ \qquad \left(\dfrac{2^{d+1}\pi^d}{\Gamma(\frac{d}{2}+3)} \left(\dfrac{(d(d+2))\Gamma(\frac{d}{2})}{4\pi^{\frac{d}{2}}} \right)^2 \right) \qquad , \ p \to \infty \end{cases} \tag{4.3}$$

Proof If $K^S_{de}(\mathbf{t})$ is the d-dimensional spherically symmetric classical Epanechnikov kernel in (2.1b), then with (3.16), we have:

$$\begin{aligned} K_{2m} &= \int_{\Re^d} \sum\nolimits_{i=1}^{d} t_i^{2m} K(\mathbf{t})d\mathbf{t} = \int_{\Re^d} \sum\nolimits_{i=1}^{d} t_i^{2m} \frac{(d+2)\Gamma(\frac{d}{2}+1)}{2\pi^{\frac{d}{2}}} (\mathbf{t}^T\mathbf{t})d\mathbf{t} \\ &= \frac{(d+2)\Gamma(\frac{d}{2}+1)}{2\pi^{\frac{d}{2}}} \int_0^1 r^{2m+d-1}(1-r^2)dr \int_0^\pi \sin^{d-2}\theta_1 d\theta_1 \times \\ &\qquad \int_0^\pi \sin^{d-3}\theta_2 d\theta_2 \times \cdots \times \int_0^\pi \sin\theta_{d-2} d\theta_{d-2} \int_0^{2\pi} d\theta_{d-1} \end{aligned}$$

Therefore:

$$K_{2m} = \frac{d(d+2)}{(2m+d)(2m+d+2)} \tag{4.4}$$

Also, putting (2.1b) in (3.20), we have:

$$\begin{aligned} ||K||_2^2 &= \int_{\mathbb{R}^d} (K(\mathbf{t}))^2 d\mathbf{t} \\ &= \left(\frac{(d+2)\Gamma(\frac{d}{2}+1)}{2\pi^{\frac{d}{2}}} \right)^2 \int_0^1 r^{d-1}(1-r^2)^2 dr \int_0^\pi \sin^{d-2}\theta_1 d\theta_1 \times \\ &\qquad \int_0^\pi \sin^{d-3}\theta_2 d\theta_2 \times \cdots \times \int_0^\pi \sin\theta_{d-2} d\theta_{d-2} \int_0^{2\pi} d\theta_{d-1} \end{aligned}$$

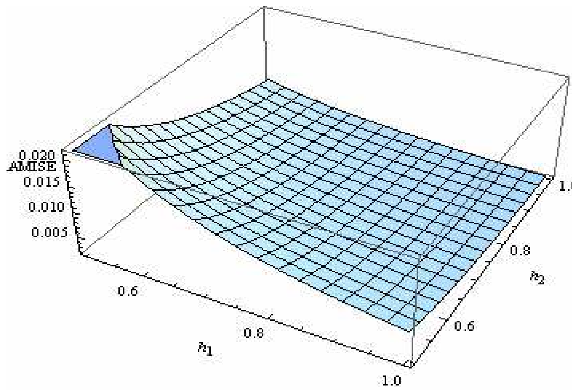
Therefore:

$$||K||_2^2 = 2\frac{\pi^{\frac{d}{2}}}{\Gamma(\frac{d}{2}+3)} \left(\frac{(d+2)\Gamma(\frac{d}{2}+1)}{2\pi^{\frac{d}{2}}} \right)^2 \tag{4.5}$$

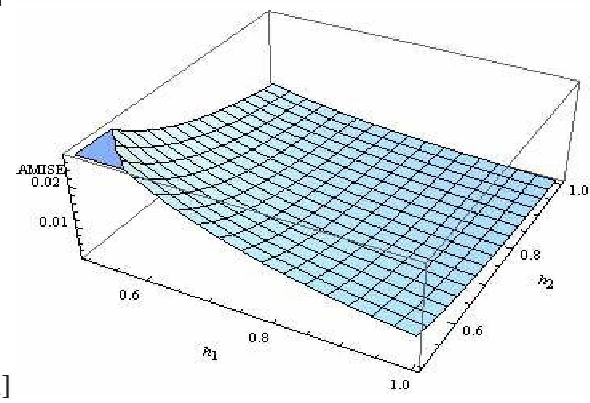
On substituting (4.4) and (4.5) and using Proposition 3.5 with necessary simplifications yield (4.3) which completes the proof.

5 Numerical verification of results

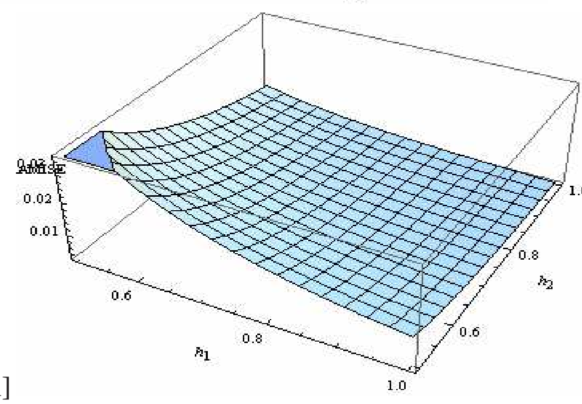
In this section, the results in Sections 3 and 4 shall be verified by means of numerical simulations. In doing this, we assumed that the size of the samples is large. Throughout this section, $f(\mathbf{x})$ is assumed to be the unit d-variate Gaussian distributions (see [16], pp 86). The AMISE of some 2-dimensional (biweight, triweight and quadriweight) kernels considered were obtained using a two dimensional Blood fat concentration data of size $n = 320$ (see [15]), and the efficiencies of some d-dimensional (biweight, triweight and quadriweight) kernels (for $d = 1, 2, 3, 4, 5, 6$) are obtained. These are achieved using the platform of Mathematica 6.0 and the graphs plotted using MINITAB 17 [12] and Mathematica 6.0 [18].



[1a. AMISE of Bivariate Biweight kernel]



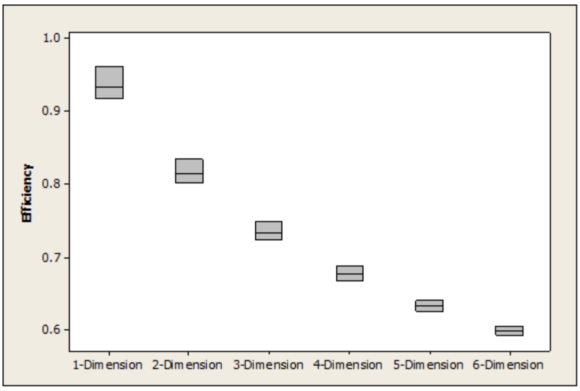
[1b. AMISE of Bivariate Triweight kernel]



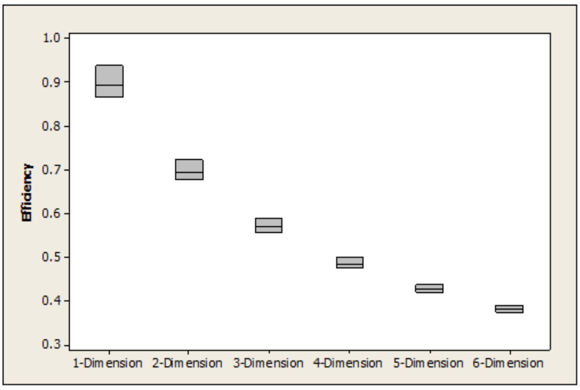
[1c. AMISE of Bivariate Quadriweight kernel]

Figure 1. AMISE of Bivariate kernels using Blood fat concentration data $n = 320$.

[2a. Boxplot of efficiency of d-dimensional spherically symmetric biweight kernel at different dimension]



[2b. Boxplot of efficiency of d-dimensional spherically symmetric triweight kernel at different dimension]



[2c. Boxplot of efficiency of d-dimensional spherically symmetric quadriweight kernel at different dimension]

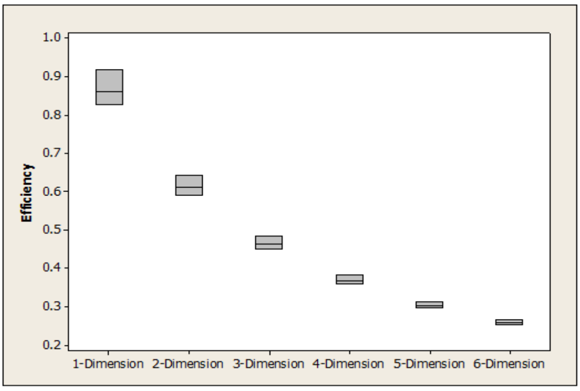
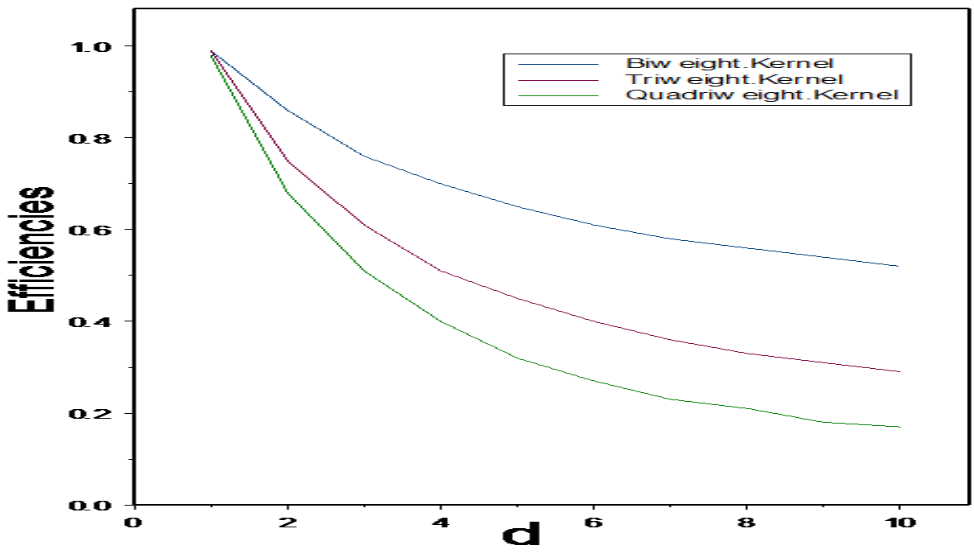


Figure 2. Efficiencies of kernels



Comparison of efficiencies of some spherically symmetric kernels for different dimensions

6 Discussion of results and conclusion

This work exhibits the convention in literature - see [16] and [17]. That is, as the order of the efficiency increases, the efficiency decreases. However, it is evident in the boxplot in Figures 2a through 2c that the efficiency of these kernels decreases as the dimension of the kernels increases. This leads to the observation that relative to the Epanechnikov kernel, the biweight kernel (when $p = 2$ in equation (2.5)), the triweight kernel (when $p = 3$ in equation (2.5)) and the quadriweight kernel (when $p = 4$ in equation (2.5)) shed about 1%, 1% and 2% respectively in efficiency for $d = 1$. Examining these figures for higher dimensional kernel for $d = 2$, it is observed that relatives to the Epanechnikov kernel, there is 14% loss in efficiency for biweight kernel, the triweight kernel lost 25% in efficiency and the quadriweight kernel lost about 32% in efficiency. For $d = 3, 4, 5, 6$; the biweight kernel lost $24\%, 30\%, 35\%, 39\%$ respectively in efficiency; the triweight kernel lost about $38\%, 49\%, 53\%, 60\%$ respectively in efficiency; and the quadriweight kernel lost about $49\%, 60\%, 68\%, 73\%$ respectively in efficiency. These may have been attributed to the curse of multidimensionality.

A further comparison of these three beta densities (biweight, triweight and quadriweight) shows that the biweight kernel gives relatively better efficiency than the triweight and quadriweight kernels. This is visible in Figure 4 where there is a sharp drop in the efficiencies of both the triweight and quadriweight kernels as their dimension increases. This clearly shows that the triweight and quadriweight kernel are highly inefficient with the increase efficiency loss as the dimension inceases. These remarks are evidently buttressed with the smallness of AMISE of the bivariate biweight kernel and the smoothness of the AMISE plot of the bivariate biweight kernel using the blood fat concentration data in Figure 1a in comparison with the AMISE plots of bivariate triweight and bivariate quadriweight in Figure 1b and 1c respectively.

In this work, we are able to propose the family of d-dimensional spherically symmetric polynomial kernels. Also, the rigour of first computing the second moment and its corresponding L_2-norm of any spherically symmetric polynomial kernel for each dimension before calculating their AMISE, the optimal bandwidth and their efficiencies have been simplified. Thus, a new computational approach has been developed for the AMISE, optimal bandwidth and efficiency of the family of d-dimensional spherically symmetric polynomial kernels. The Epanechnikov kernel was used as a theoretical underpinning for the derivation of the efficiency formula. The constants of the new generalised family of kernels were compared with existing one in the literature and were seen to be in consonance with those in the literature (see Table 1).

In addition, the new generalised efficiency formula was experimented with three of the beta kernels - biweight, triweight and quadriweight kernel; our findings reveal that the biweight kernel has relatively high efficiency values at both the higher dimension and higher-order. However, we cannot just jump into concluding that the biweight kernel supersedes all other kernels in the beta family in terms of efficiency. More investigation is needed in this regard.

References

[1] B. A. Afere and C. C. Ishiekwene, On the efficiency of multivariate spherically symmetric polynomial kernels: A biweight kernel perspective, *Journal of the Nigerian Association of Mathematical Physics*, Volume 28, No. 2, (2014) 497 - 504.

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

[3] T. Cacoullos, Estimation of a multivariate density, *Annals of the Institute of Statistical Mathematics* 18, (1966) 179 - 189.

[4] P. Deheuvels, Estimation non parametriani de la densite par histogrammes generalises II, Publications de l'insitute de statistique de l'universite de Paris 22, (1977) 1 - 25.

[5] L. Devroye and G. Lugosi, *Combinatorial Method in Density Estimation*, Springer-Verlag, New York, 2001.

[6] T Duong and M. L. Hazelton, Convergence rates for unconstrained bandwidth matrix selectors in multivariate kernel density estimator, *Journal of Multivariate Analysis*, Volume 93, (2005) 417 - 433.

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

[8] K. Fakunaga, *Introduction to Statistical Pattern Recognition*, Academic Press, New York, 1990.

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

[10] W. Härdle, M. Müller, S. Sperlich and A. Werwatz, *Nonparametric and Semiparametric Models*. Springer-Verlag, Berlin, 2004.

[11] A. J. Izenman, *Modern Multivariate Statistical Techniques: Regression, Classification and Manifold Learning*, Springer Science, New York, 2008.

[12] Minitab 17 Statistical Software. [Computer software]. State College, PA: Minitab, Inc. (www.minitab.com), 2010.

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

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

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

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

[17] M. P. Wand and M. C. Jones, *Kernel Smoothing*, Chapman and Hall, London, 1995. Wolfram Research, Inc.,

[18] Mathematica Edition: Version 6.0, Wolfram Research, Inc., Champaign, Illinois, 2007.

[19] T. J. Wu, C. F. Chen and H. Y. Chen, A variable bandwidth selector in multivariate kernel density estimation, *Statistics and Probability Letters*, 77 (2007), 462 - 467.

Author information

B. A. Afere, Department of Mathematics and Statistics,Federal Polytechnic,Idah, Kogi State, Nigeria.
E-mail: baafere@gmail.com

F. O. Oyegue, Department of Mathematics,University of Benin, Benin City, Edo State, Nigeria.
E-mail: fooyegue@yahoo.com

Received: July 26, 2015.

Accepted: December 4, 2015