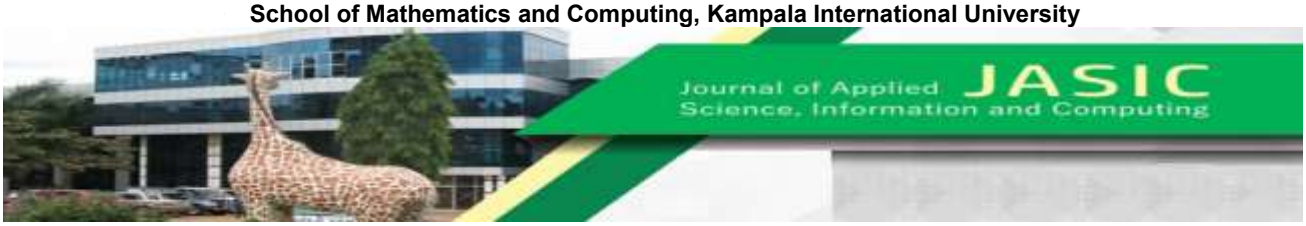
Vol. 1, No. 1 (June 2020)
Journal of Applied Sciences, Information, and Computing
School of Mathematics and Computing, Kampala International University



ISSN: 1813-3509 https://jasic.kiu.ac.ug

A Discourse on Smoothing Parameterizations Using Hypothetical Dataset

Israel U. Siloko[1] , *Wilson Nwankwo[2], Chinecherem Umezuruike[3]
Department of Mathematics and Computer Science, Edo University Iyamho, Nigeria.
nwankwo.wilson@edouniversity.edu.ng
Department of Information Technology, Kampala International University. cumezuruike@kiu.ac.ug

Abstract

The univariate kernel estimator usually requires a smoothing parameter, unlike the multi-dimensional estimators that necessarily require more smoothing parameters. The smoothing parameter(s) of kernels with a higher dimension may be called smoothing matrices. Kernels of higher dimensions have three kinds of parameterizations as estimators viz: constant, diagonal, and full parameterizations. Unlike the full parameterization, the diagonal parameterization exhibit some levels of restrictions. This study attempts to reconnoiter the coherence exhibited by kernel estimators especially where smoothing parameterizations are employed. In this discourse, asymptotic mean-integrated squared error(AMISE) is used as a criterion function and bivariate cases alone are considered. With some hypothetical data, the results show that full smoothing parameterization outperformed the constant and diagonal parameterizations in respect of the asymptotic mean-integrated squared error's value and the kernel estimate's ability to retain the true characteristics of the affected distribution.

Keywords: Smoothing Matrix, Kernel estimation, Variance, AMISE, the Density function

1. Introduction

With recent developments in analytics, machine learning big data, statistical analysis has gained a strong foothold as applied science. Many statistical algorithms have been refined and integrated into many machine models. Modern analytics is almost inseparable from statistical models and reasoning. In statistical terms, data analysis revolves around density estimation, which involves the design of probability estimates with observations drawn from certain or uncertain datasets. Density estimation (DE) is a major statistical inferencing subject and maybe examined from two contexts; the parametric estimation (PE) and the nonparametric estimation (NPE). In PE, the dataset or observations may belong to a certain distribution. In such a scenario, prior knowledge of the distribution may be required to estimate the required parameters. Under NPE, assumptions are not often tenable as to how the observations are distributed but the observations are given the opportunity to "speak for themselves".

Nonparametric density estimation techniques are of wide applications with the kernel density estimator (KDE) playing crucial statistical roles in data analysis. Nonparametric estimation forms the building blocks for different semiparametric estimators where the separability ideology of the independent variables in the semi-parametric model is in line with the devolution of the decision-making process in organizations or stages of production in industries in a real-life situation (Hardle, et. al.,2004). Kernel estimation (KE) is one method in data smoothing. It involves drawing inferences and subsequent conclusions on various observations. KE is a vital tool in analysis, representation, and visualization in relation to a given distribution (Siloko, et. al.,2019, Simonoff, 1996). KE could be applied indirectly to other areas of nonparametric estimation such as discriminant analysis, goodness-of-fit testing, hazard rate estimation, bump-hunting, intensity function estimation, and classification with regression estimation (Raykar, Duraiswami & Zhao, 2015). The kernel estimator is a popular nonparametric technique in density estimation (DE) and its univariate form is stated thus:

$$\hat{f}(\mathrm{x}) = \frac{1}{nh_{\mathrm{x}}}\sum_{i=1}^{n} K\left(\frac{\mathrm{x}-\mathrm{X}_i}{h_{\mathrm{x}}}\right), \qquad (1.1)$$

where $K(\cdot)$ is the kernel function(KF), $h_{\mathrm{x}} > 0$ is the smoothing parameter(SP) i.e. bandwidth, X_i represents the

real observations/measurements, and n represents the sample size. The KF decides the pattern of the estimate generated while the SP controls the extent of smoothing to which the estimate attains. The KF is non-negative and satisfies the following conditions

$$\int K(\mathrm{x})d\mathrm{x} = 1, \int \mathrm{x}K(\mathrm{x})d\mathrm{x} = 0 \quad \text{and} \int \mathrm{x}^2K(\mathrm{x})d\mathrm{x} = \mu_2(K) \neq 0\,. \tag{1.2}$$

Three conditions are associated with equation (1.2):

a. The first implies that KF must integrate into one, which is to the effect that KFs are probability density functions (PDFs);
b. The second is that the average of each kernel is zero.
c. The third indicates that the variance computed for the kernel $\mu_2(K)$ is non-zero (Scott, 1992).

KDE is applied in multivariate scenarios where different sets of observations are analyzed, particularly the bivariate kernel (BK) that presents its estimates in wireframes or contour plots. The BK density estimator is very relevant as it serves as a bridge between univariate KE and higher dimensional kernel estimators. In BK estimation, x,y are considered random variables whose values are in $\mathcal{R}^2$ having the joint density function(JDF) $f(\mathrm{x},\mathrm{y}), (\mathrm{x},\mathrm{y}) \in \mathcal{R}^2$ with $\mathrm{X}_i, \mathrm{Y}_i,\ i = 1, 2, \dots, n$ representing a class of observations of size n drawn from the distribution. The BK density estimator is expressed as:

$$\hat{f}(\mathrm{x},\mathrm{y}) = \frac{1}{nh_\mathrm{x}h_\mathrm{y}}\sum_{i=1}^{n} K\left(\frac{\mathrm{x}-\mathrm{X}_i}{h_\mathrm{x}}, \frac{\mathrm{y}-\mathrm{Y}_i}{h_\mathrm{y}}\right) = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{h_\mathrm{x}} K\left(\frac{\mathrm{x}-\mathrm{X}_i}{h_\mathrm{x}}\right)\frac{1}{h_\mathrm{y}} K\left(\frac{\mathrm{y}-\mathrm{Y}_i}{h_\mathrm{y}}\right) \tag{1.3}$$

where $h_\mathrm{x} > 0$ and $h_\mathrm{y} > 0$ are the SPs in X and Y axes and $K(\mathrm{x},\mathrm{y})$ is a BK function, the product of two univariate kernels. The KEs of Equation (1.3) is simple to understand and interpret whether as surface plots or contour plots. In data analysis and visualization, the bivariate kernel estimator (BKE) is useful for visualization using the familiar contour plots and/or perspectives (Silverman, 1996, Simonoff, 1996, Scott, 1992). The BKE is applied in domains like nonparametric discriminant analysis and goodness-of-fit testing (Duong & Hazelton, 2003). A major problem of KE is in the selection of SP. The SP is relevant in measuring the estimator's performance whether in the univariate or multivariate form (Zhang, Wu, Pitt & Liu 2011, Siloko, Ishiekwene & Oyegue, 2018). The multivariate form of Equation (1.1) with a single bandwidth KE is given as

$$\hat{f}(\mathrm{x}) = \frac{1}{nh_\mathrm{x}^d}\sum_{i=1}^{n} K\left(\frac{\mathrm{x}-\mathrm{X}_i}{h_\mathrm{x}}\right), \tag{1.4}$$

where d is the KF's dimension. The kernel, K, in this case, is a d-variate density function that satisfies the conditions in Equation (1.2) and its contours are assumed to be spherically symmetric. The advantage of this multivariate form of parameterization is that the multivariate asymptotic mean integrated squared error (AMISE) and the optimal smoothing parameter value can be easily computed unlike other complex forms of parameterizations without explicit optimal bandwidth formula. The multivariate product kernel form of Equation (1.1) with each smoothing parameter for the axes and also satisfy the conditions in Equation (1.2) is given as

$$\hat{f}(\mathrm{x}) = n^{-1}\left(\prod_{j=1}^{d} h_j\right)^{-1}\sum_{i=1}^{n} K\left(\frac{\mathrm{x}_1-\mathrm{X}_{i1}}{h_1}, \frac{\mathrm{x}_2-\mathrm{X}_{i2}}{h_2}, \dots, \frac{\mathrm{x}_d-\mathrm{X}_{id}}{h_d}\right) = n^{-1}\left(\prod_{j=1}^{d} h_j\right)^{-1}\sum_{i=1}^{n} K\left(\frac{\mathrm{x}_j-\mathrm{X}_{ij}}{h_j}\right). \tag{1.5}$$

This multivariate product kernel applies the same symmetric univariate kernel and variance in each dimension but different bandwidths for the axes (Sain, 2002).

The aim of this study is to examine the KE's efficiency or performance with real data using the constant smoothing matrix, diagonal smoothing matrix, and full smoothing matrix. In each case, AMISE is used as the error criterion function (ECF).

2. The AMISE Approximations

The mean squared error (MSE) has two components: the bias and standard error (or variance). The MSE of $\hat{f}(\mathrm{x})$ is a function of the argument x defined as:

$$\mathrm{MSE}\left(\hat{f}(\mathrm{x})\right) = E\left(\hat{f}(\mathrm{x}) - f(\mathrm{x})\right)^2 = \left(E\hat{f}(\mathrm{x}) - f(\mathrm{x})\right)^2 + E\left(\hat{f}(\mathrm{x}) - \mathrm{E}\hat{f}(\mathrm{x})\right)^2 = \mathrm{Bias}^2\left(\hat{f}(\mathrm{x})\right) + \mathrm{Var}\left(\hat{f}(\mathrm{x})\right). \tag{2.1}$$

The global measure of the accuracy of $\hat{f}(\mathrm{x})$ is the mean integrated square error given by

$$\mathrm{MISE}\left(\hat{f}(\mathrm{x})\right) = \mathrm{E}\left(\int \left(\hat{f}(\mathrm{x}) - f(\mathrm{x})\right)^2 d\mathrm{x}\right) = \int E\left(\hat{f}(\mathrm{x}) - f(\mathrm{x})\right)^2 d\mathrm{x} = \int \mathrm{MSE}\left(\hat{f}(\mathrm{x})\right) d\mathrm{x} = \int \mathrm{Bias}^2\left(\hat{f}(\mathrm{x})\right) d\mathrm{x} + \int \mathrm{Var}\left(\hat{f}(\mathrm{x})\right) d\mathrm{x}. \tag{2.2}$$

The two components of the mean integrated squared error will be considered individually.

2. 1 The Bias Term

The mathematical expectations of a kernel transformation can be written as integrals which take the form of a convolution of the kernel and the density function as

$$E\left(\frac{1}{h_{\mathrm{x}}}K\left(\frac{\mathrm{x}-\mathrm{X}_i}{h_{\mathrm{x}}}\right)\right) = \frac{1}{h_{\mathrm{x}}}\int K\left(\frac{\mathrm{x}-t}{h_{\mathrm{x}}}\right)f(t)dt \qquad (2.1.1)$$

Since the operator E is linear we have

$$\begin{aligned} E\left(\hat{f}(\mathrm{x})\right) &= E\left(\frac{1}{n}\sum_{i=1}^{n}\frac{1}{h_{\mathrm{x}}}\,\mathrm{K}\left(\frac{\mathrm{x}-\mathrm{X}_i}{h_{\mathrm{x}}}\right)\right) \\ &= \frac{1}{n}\sum_{i=1}^{n}E\left(\frac{1}{h_{\mathrm{x}}}\int \mathrm{K}\left(\frac{\mathrm{x}-t}{h_{\mathrm{x}}}\right)f(t)dt\right) \\ &= \frac{1}{h_{\mathrm{x}}}\int \mathrm{K}\left(\frac{\mathrm{x}-t}{h_{\mathrm{x}}}\right)f(t)dt. \end{aligned} \qquad (2.1.2)$$

The transformation $u = (\mathrm{x}-t)/\,h_{\mathrm{x}},\ t = \mathrm{x} - h_{\mathrm{x}}u\ and\ \left|\frac{du}{dt}\right| = \frac{1}{h_{\mathrm{x}}}$ on Equation (2.1.2) yields

$$\begin{aligned} &E\left(\hat{f}(\mathrm{x})\right) \\ &= \int \mathrm{K}(u)f(\mathrm{x} \\ &- h_{\mathrm{x}}u)du. \end{aligned} \qquad (2.1.3)$$

The integral in Equation (2.1.3) is not analytically solvable, so it will be approximated using Taylor's series expansion on $f(\mathrm{x} - h_{\mathrm{x}}u)$ in the argument $h_{\mathrm{x}}u$, which yields

$$\begin{aligned} f(\mathrm{x} - h_{\mathrm{x}}u) = f(\mathrm{x}) - f^{(1)}(\mathrm{x})\,h_{\mathrm{x}}u + \frac{1}{2!}f^{(2)}(\mathrm{x})h_{\mathrm{x}}^2u^2 + \cdots \\ + \frac{1}{p!}f^{(p)}(\mathrm{x})h_{\mathrm{x}}^pu^p + O(h_{\mathrm{x}}^p). \end{aligned} \qquad (2.1.4)$$

Integrating Equation (2.1.4) term by term and using the conditions in Equation (1.3), we have

$$\begin{aligned} &\int \mathrm{K}(u)f(\mathrm{x} - h_{\mathrm{x}}u)du \\ &= f(\mathrm{x})\int \mathrm{K}(u)du \\ &- f^{(1)}(\mathrm{x})\int h_{\mathrm{x}}u\mathrm{K}(u)du + \\ &f^{(2)}(\mathrm{x})\int \frac{1}{2!}(uh_{\mathrm{x}})^2K(u)du + \cdots \\ &+ f^{(p)}(\mathrm{x})\int \frac{1}{p!}(uh_{\mathrm{x}})^pK(u)du \\ &+ O(h_{\mathrm{x}}^p) \\ &= f(\mathrm{x}) \\ &+ \frac{1}{p!}f^{(p)}(\mathrm{x})f^{(p)}(\mathrm{x})\int (uh_{\mathrm{x}})^p\mathrm{K}(u)du \\ &+ O(h_{\mathrm{x}}^p). \end{aligned} \qquad (2.1.5)$$

The second equality in Equation (2.1.5) uses the assumption that K is a pth order kernel. Hence the bias could be stated thus:

$$\begin{aligned} \mathrm{E}\left(\hat{f}(\mathrm{x})\right) = E\left(\frac{1}{n}\sum_{i=1}^{n}\frac{1}{h_{\mathrm{x}}}K\left(\frac{\mathrm{x}-\mathrm{X}_i}{h_{\mathrm{x}}}\right)\right) \\ = f(\mathrm{x}) \\ + \frac{1}{p!}f^{(p)}(\mathrm{x})f^{(p)}(\mathrm{x})\int (uh_{\mathrm{x}})^pK(u)du \\ + O(h_{\mathrm{x}}^p). \end{aligned}$$

The kernel estimator's bias $\hat{f}(\mathrm{x})$ is of the form

$$\begin{aligned} Bias\left(\hat{f}(\mathrm{x})\right) &= \mathrm{E}\hat{f}(\mathrm{x}) - f(\mathrm{x}) \\ &= \frac{1}{p!}h^pf^{(p)}(\mathrm{x})\int u^pK(u)du + O(h_{\mathrm{x}}^p) \\ &= \frac{1}{p!}h_{\mathrm{x}}^pf^{(p)}(\mathrm{x})\mu_p(K) \\ &+ O(h_{\mathrm{x}}^p), \end{aligned} \qquad (2.1.6)$$

where $\mu_p(K) = \int u^pK(u)du.$

In the case of the second-order kernels, Equation (2.1.6) becomes:

$$\begin{aligned} &Bias\left(\hat{f}(\mathrm{x})\right) \\ &= \frac{1}{2}h_{\mathrm{x}}^2f^{(2)}(\mathrm{x})\mu_2(K) \\ &+ O(h_{\mathrm{x}}^2), \end{aligned} \qquad (2.1.7)$$

where $O(h_{\mathrm{x}}^2)$ represents terms that converge to zero faster than h_{x}^2 as h_{x} approaches zero. Thus the bias is

$$\begin{aligned} &Bias\left(\hat{f}(\mathrm{x})\right) \\ &\approx \frac{1}{2}h_{\mathrm{x}}^2f^{(2)}(\mathrm{x})\mu_2(K). \end{aligned} \qquad (2.1.8)$$

Therefore the integrated squared bias for the mean integrated squared error is of the form

$$\begin{aligned} &\left(\int Bias\hat{f}(\mathrm{x})d\mathrm{x}\right)^2 \\ &\approx \frac{1}{4}h_{\mathrm{x}}^4\mu_2(K)^2\int f^{(2)}(\mathrm{x})^2d\mathrm{x}. \end{aligned} \qquad (2.1.9)$$

This means that the bias depends on the SP h_{x} , the variance of the kernel $\mu_2(K)$ and the curvature of the density $f^{(2)}(\mathrm{x})$ at the point x. Generally, the bias in the square of SP potentially increases in the second-order kernel.

2. 2 Variance Term

The KE is a linear estimator and the X_i are usually independently and identically distributed. Thus the variance is given by

$$
\begin{aligned}
&Var\left(\mathrm{K}\left(\frac{\mathrm{x}-\mathrm{X}_i}{h_{\mathrm{x}}}\right)\right)\\
&= E\left(K\left(\frac{\mathrm{x}-\mathrm{X}_i}{h_{\mathrm{x}}}\right)^2\right)\\
&-\left(E\,K\left(\frac{\mathrm{x}-\mathrm{X}_i}{h_{\mathrm{x}}}\right)\right)^2\\
&= \int K\left(\frac{\mathrm{x}-t}{h_{\mathrm{x}}}\right)^2 f(t)dt\\
&-\left(\int K\left(\frac{\mathrm{x}-t}{h_{\mathrm{x}}}\right)f(t)dt\right)^2. \qquad (2.2.1)
\end{aligned}
$$

Therefore, the variance of $\hat{f}(\mathrm{x})$ is

$$
\begin{aligned}
&\mathrm{Var}\left(\hat{f}(\mathrm{x})\right)\\
&= \frac{1}{n}\int \frac{1}{h_{\mathrm{x}}^2}K\left(\frac{\mathrm{x}-t}{h_{\mathrm{x}}}\right)^2 f(t)dt\\
&-\frac{1}{n}\left(\frac{1}{h_{\mathrm{x}}}\int K\left(\frac{\mathrm{x}-t}{h_{\mathrm{x}}}\right)f(t)dt\right)^2\\
&= \frac{1}{n}\int \frac{1}{h_{\mathrm{x}}^2}K\left(\frac{\mathrm{x}-t}{h_{\mathrm{x}}}\right)^2 f(t)dt\\
&-\frac{1}{n}\left(f(\mathrm{x})+\mathrm{Bias}\left(\hat{f}(\mathrm{x})\right)\right)^2. \qquad (2.2.2)
\end{aligned}
$$

Using the change-of-variables $u = (\mathrm{x}-t)/\,h_{\mathrm{x}}$, $t = \mathrm{x} - h_{\mathrm{x}}u$ and $\left|\frac{dt}{du}\right| = h_{\mathrm{x}}$, we have

$$
\begin{aligned}
&\mathrm{Var}\left(\hat{f}(\mathrm{x})\right)\\
&= \frac{1}{nh_{\mathrm{x}}}\int K\,(u)^2 f(\mathrm{x}-\,h_{\mathrm{x}}u)du\\
&-\frac{1}{n}(f(\mathrm{x})+O(h_{\mathrm{x}}^2))^2. \qquad (2.2.3)
\end{aligned}
$$

Applying Taylor's series expansion on Equation (2.2.3) we have

$$
\begin{aligned}
\mathrm{Var}\left(\hat{f}(\mathrm{x})\right) &= \frac{1}{nh_{\mathrm{x}}}\int K\,(u)^2(f(\mathrm{x})-\,h_{\mathrm{x}}uf^{(1)}+O(h_{\mathrm{x}}))du\\
&-\frac{1}{n}(f(\mathrm{x})+O(h_{\mathrm{x}}^2))^2. \qquad (2.2.6)
\end{aligned}
$$

As n gets large and h_{x} decreases and with conditions in Equation (1.2), we then have Equation (2.2.6) to be approximately given as

$$
\begin{aligned}
&\mathrm{Var}\left(\hat{f}(\mathrm{x})\right)\\
&\approx \frac{1}{nh_{\mathrm{x}}}f(\mathrm{x})\int K(u)^2du. \qquad (2.2.7)
\end{aligned}
$$

The function $f(\mathrm{x})$ in Equation (2.2.7) is a probability density function; therefore integrating it will produce an approximation of the form

$$
\begin{aligned}
&\int \mathrm{Var}\hat{f}(\mathrm{x})d\mathrm{x}\\
&\approx \frac{1}{nh_{\mathrm{x}}}\int K(\mathrm{x})^2d\mathrm{x}. \qquad (2.2.8)
\end{aligned}
$$

Thus, the mean squared error for the second-order kernel is of the form

$$
\begin{aligned}
&\mathrm{MSE}\left(\hat{f}(\mathrm{x})\right)\\
&= \left(E\hat{f}(\mathrm{x})-f(\mathrm{x})\right)^2 + \;\; E\left(\hat{f}(\mathrm{x})-\mathrm{E}\hat{f}(\mathrm{x})\right)^2\\
&= \frac{1}{4}h_{\mathrm{x}}^4 f^{(2)}(\mathrm{x})\mu_2(K)^2\\
&+\frac{1}{nh_{\mathrm{x}}}f(\mathrm{x})\int K(\mathrm{x})^2d\mathrm{x}. \qquad (2.2.9)
\end{aligned}
$$

Again, on integrating Equation (2.2.9) with respect to x yields an estimate of the mean integrated squared error given by

$$
\begin{aligned}
&MISE\left(\hat{f}(\mathrm{x})\right)\\
&= \frac{1}{4}h_{\mathrm{x}}^4\mu_2(K)^2\int f^{(2)}(\mathrm{x})d\mathrm{x}\\
&+\frac{1}{nh_{\mathrm{x}}}\int K(\mathrm{x})^2d\mathrm{x}. \qquad (2.2.10)
\end{aligned}
$$

A global measure of precision is the asymptotic mean integrated squared error given by

$$
\begin{aligned}
&AMISE\left(\hat{f}(\mathrm{x})\right) \approx \int AMSE\left(\hat{f}(\mathrm{x})\right)d\mathrm{x}\\
&= \frac{1}{4}h_{\mathrm{x}}^4\mu_2(\mathrm{K})^2\int f^{(2)}(\mathrm{x})d\mathrm{x}+\frac{1}{nh_{\mathrm{x}}}\int K(\mathrm{x})^2d\mathrm{x}\\
&\approx \frac{1}{4}h_{\mathrm{x}}^4\mu_2(K)^2R\left(f^{(2)}\right)\\
&+\frac{R(K)}{nh_{\mathrm{x}}}, \qquad (2.2.11)
\end{aligned}
$$

where $R\left(f^{(2)}\right) = \int f^{(2)}(\mathrm{x})^2d\mathrm{x}$ is the roughness of the unknown probability function, h_{x} is the smoothing parameter, n is the sample size, $R(K) = \int K(\mathrm{x})^2d\mathrm{x}$ is the roughness of the kernel function and $\mu_2(K)^2$ is the variance of the kernel.

2.3 The Asymptotically Optimal Bandwidth

The formula of AMISE expresses the MSE as a function of the SP denoted by h_{x}. The SP value h_{x} that minimizes the

expression of the AMISE is called the asymptotically optimal bandwidth or asymptotically optimal smoothing parameter. The optimal smoothing parameter is obtained by solving the differential equation

$$\frac{\partial}{\partial h_{\mathrm{x}}} AMISE = \frac{-R(K)}{nh_{\mathrm{x}}^2} + \mu_2(\mathrm{K})^2 h_{\mathrm{x}}^3 R\left(f^{(2)}\right) = 0. \quad (2.3.1)$$

The solution to Equation (2.3.1) will produce the smoothing parameter that minimizes the AMISE of the kernel estimator which is of the form

$$h_{\mathrm{x-AMISE}} = \left[\frac{R(K)}{\mu_2(\mathrm{K})^2 R(f^{(2)})}\right]^{1/5} \times n^{-1/5}. \quad (2.3.2)$$

The smoothing parameter with the minimum AMISE in Equation (2.3.2) can also be expressed in terms of the dimension as

$$h_{\mathrm{x-AMISE}} = \left[\frac{R(K)}{\mu_2(\mathrm{K})^2 R(f^{(2)})}\right]^{\left(\frac{1}{4+d}\right)} \times n^{-1/(4+d)}. \quad (2.3.3)$$

The corresponding AMISE of Equation (1.4) is given as

$$AMISE\left(\hat{f}(\mathrm{x})\right) = \frac{R(K)}{nh_{\mathrm{x}}^d} + \frac{h_{\mathrm{x}}^4}{4}\mu_2(\mathrm{K})^2 \int \{\nabla^2 f(\mathrm{x})\}^2 \, d\mathrm{x}, \quad (2.3.4)$$

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

Also, the smoothing parameter that will minimize the AMISE in Equation (2.3.4) is of the form

$$H_{\mathrm{AMISE}} = \left[\frac{dR(K)}{\mu_2(\mathrm{K})^2 \int \{\nabla^2 f(\mathrm{x})\}^2 \, d\mathrm{x}}\right]^{\left(\frac{1}{4+d}\right)} \times n^{-1/(4+d)}. \quad (2.3.5)$$

The AMISE of the multivariate product kernel estimator in Equation (1.5) is given as

$$AMISE = \frac{R(K)^d}{nh_1 h_2, \dots, h_d} + \frac{1}{4}h_j^4 \mu_2(\mathrm{K})^2 \int tr^2\{\mathcal{H}_f(\mathrm{x})\} d\mathrm{x}, \quad (2.3.6)$$

where $j = 1, 2, \dots, d$, $R(K)$ is the roughness of the kernel, $\mu_2(K)^2$ is the variance of the kernel, $\mathcal{H}_f$ is the Hessian matrix of the density $f(\mathrm{x})$ and tr indicates the trace of a matrix (Wand & Jones, 1995, Sain, Baggely, & Scott, 1994). The popular parameterizations in multivariate kernel estimation are the constant, diagonal, and full parameterizations provided the matrix is symmetric and positive definite. The smoothing parameter that minimizes the *AMISE* of Equation (2.3.6) is

$$H_{AMISE} = \left[\frac{R(K)^d}{\mu_2(\mathrm{K})^2 \int tr^2\{\mathcal{H}_f(\mathrm{x})\} d\mathrm{x}}\right]^{\left(\frac{1}{d+4}\right)} \times n^{-\left(\frac{1}{d+4}\right)}. \quad (2.3.7)$$

The choice of the SPs i.e. the smoothing matrix in the multivariate case is strictly based on the complexity of the underlying density and the number of parameters to be estimated. The choice of a kernel function is not a problem because most kernel functions are probability density function. The kernel function employed in this work is the standard normal kernel that produces smooth density estimates and simplifies the mathematical computations. The standard normal kernel function of the bivariate kernel estimator is

$$K(\mathrm{x}, \mathrm{y}) = \frac{1}{2\pi}\exp\left(-\frac{\mathrm{x}^2 + \mathrm{y}^2}{2}\right). \quad (2.3.8)$$

The matrix form of the diagonal parameterization and the full parameterization of the bivariate kernel estimator is

$$H_{AMISE} = \begin{bmatrix} h_{\mathrm{x}}^2 & 0 \\ 0 & h_{\mathrm{y}}^2 \end{bmatrix} \quad \text{and} \quad H_{AMISE} = \begin{bmatrix} h_{\mathrm{x}}^2 & h_{\mathrm{xy}} \\ h_{\mathrm{xy}} & h_{\mathrm{y}}^2 \end{bmatrix}.$$

The diagonal form of smoothing parameterization considers only the elements of the leading diagonal of the smoothing matrix while the off-diagonal elements are zero whereas the full smoothing matrix takes into consideration all the elements. In constant parameterization, the same smoothing parameter is applied to all axes. The bivariate form of the constant parameterization is based on the assumption that $h_{\mathrm{x}} = h_{\mathrm{y}} = h_{\mathrm{z}}$ and the matrix representation is

$$H_{AMISE} = \begin{bmatrix} h_{\mathrm{z}}^2 & 0 \\ 0 & h_{\mathrm{z}}^2 \end{bmatrix}$$

The performance of these forms of parameterizations will be compared using the asymptotic mean integrated squared error as the error criterion function.

3. Results and Discussion

Here, we discussed the performance of the constant, diagonal, and full smoothing matrices using real data examples. Two bivariate data sets are examined with three classes of smoothing parameterizations. The smoothing matrix that minimizes the AMISE in the constant smoothing

matrix is represented by $H_{C-AMISE}$ while the diagonal and full smoothing matrices are represented by $H_{D-AMISE}$ and $H_{F-AMISE}$ respectively. The results of their performance in terms of the AMISE value are presented in Table 1 and Table 2 respectively and these are statistically relevant. Again kernel performance can also be viewed from the ability of the kernel estimates to retain the true features of the observations and this vital role of retention of features such as bimodality of the distribution is in Figure 6.

The first data set is the ages at marriage for a sample of 100 couples that applied for marriage licenses in Cumberland County, Pennsylvania USA, which is made up of two variables the ages of husbands and their wives at marriage (Sabine & Brain, 2004). The analysis of these data addresses the issue of differences in the ages of husbands and wives at marriage. However; on general observation of the data, wives are younger at marriage. The data were standardized in order to obtain equal variances in each dimension because, in most multivariate statistical analysis, the data are always standardized to ensure that there are no differences in the ranges of variables. The smoothing matrices for the constant, diagonal and full matrices of this data are

$$H_{C-AMISE} = \begin{bmatrix} 0.502031 & 0.000000 \\ 0.000000 & 0.502031 \end{bmatrix} \qquad H_{D-AMISE} = \begin{bmatrix} 0.502031 & 0.000000 \\ 0.000000 & 0.502346 \end{bmatrix}$$

$$H_{F-AMISE} = \begin{bmatrix} 0.5635105 & 0.0285102 \\ 0.0285102 & 0.5638640 \end{bmatrix}$$

The BKEs of the forms of parameterizations are shown in Figure 1, Figure 2, and Figure 3 respectively, and represent the surface and contour plots using the bivariate standard normal kernel(BSNK). In this data set, estimates of the constant and diagonal matrices are similar and it should be noted that the kernel estimates of the forms of parameterization investigated clearly revealed the unimodality of the data which exemplifies the usefulness of bivariate kernel estimates in highlighting structures in a data set. The unimodality of the data set is an indication of the ages at which husbands and their wives were more likely to get married and from the data, these ages are distinctly centered between ages 26 and ages 28. The probability of getting married at these ages is high for both men and women with values between 0.2 and 0.25 respectively as seen in the contour plot with the highest probability value at its mode. The kernel estimates show the tendency for younger men to marry younger women and vice versa where marriages are usually more in the twenty's and the probability of getting married tends to slow downward as there is an increase in ages; hence with lower probability values.

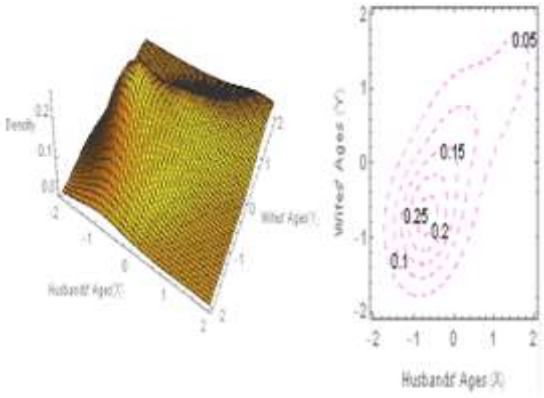


Figure-1. KEs (Surface Contour plots) of Hc SP

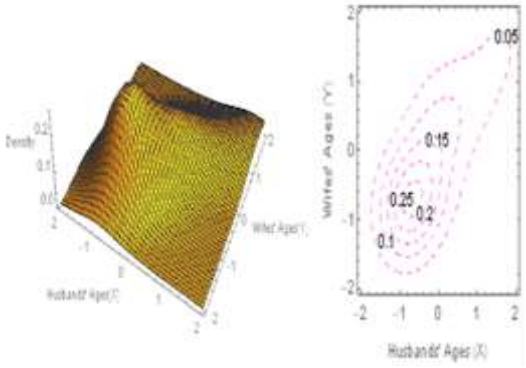


Figure-2. KEs (Surface and Contour plots) of H_D SP

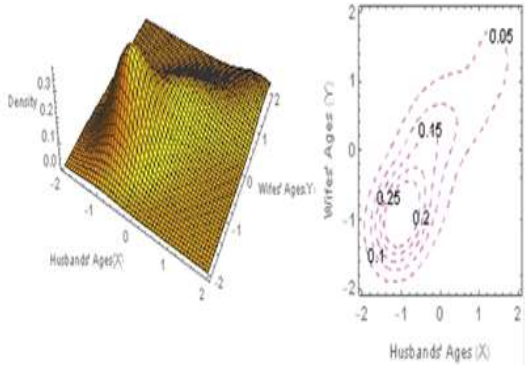


Figure-3. KEs (Surface and Contour plots) of H_p SP

Table 1 shows the asymptotic integrated variance, asymptotic integrated squared bias, and the asymptotic mean integrated squared error (AMISE) for the ages at marriage data.

The analysis presented in Table 1 clearly shows that the full smoothing matrix did better than the constant and diagonal matrices in terms of performance because it produces the smallest value of the AMISE. The superiority of any method in kernel estimation is based on its ability to produce the minimum AMISE value in comparison with other methods (Janicka, 2009).

The second data set examined in the blood fat concentration data also known as the lipid data of (Scott, Golto, Cole & Gorry, 1978). These data consist of measurements of cholesterol and triglycerides for 320 men diagnosed with coronary artery disease and the analysis of the data revealed that they are bimodal. The bimodality of these data is an indication that an increased risk for heart disease is associated with an increased cholesterol level. The data were standardized to obtain equal variances in each dimension and the smoothing matrices for this data set are

$$H_{C-AMISE} = \begin{bmatrix} 0.403262 & 0.000000 \\ 0.000000 & 0.403262 \end{bmatrix} \qquad H_{D-AMISE} = \begin{bmatrix} 0.403262 & 0.000000 \\ 0.000000 & 0.410678 \end{bmatrix}$$

$$H_{F-AMISE} = \begin{bmatrix} 0.4526467 & 0.0021664 \\ 0.0021664 & 0.4609705 \end{bmatrix}$$

Figure 4, Figure 5, and Figure 6 show the kernel estimates, which are the surface plots and the contour plots of the forms of parameterizations using the bivariate standard normal kernel.

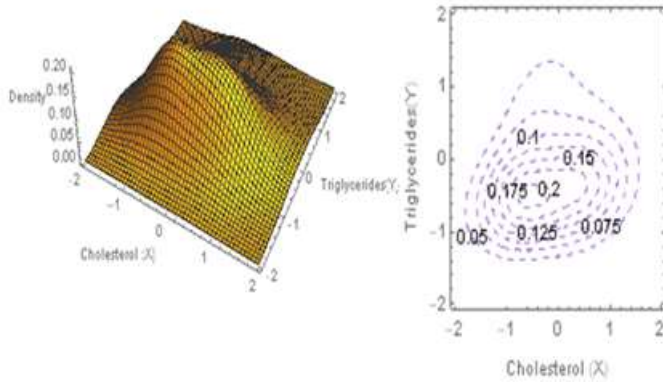


Figure-4. KEs (Surface and Contour plots) of H_c SP

Table-1. Variance, Bias^2, and AMISE of Ages Marriage Data

Methods.	*Variance*	$Bais^2$	*AMISE*
$H_{C\text{-}AMISE}$	0.00315740	0.00157870	0.00473610
$H_{D\text{-}AMISE}$	0.00315542	0.00157771	0.00473313
$H_{F\text{-}AMISE}$	0.00251088	0.00039410	0.00290498

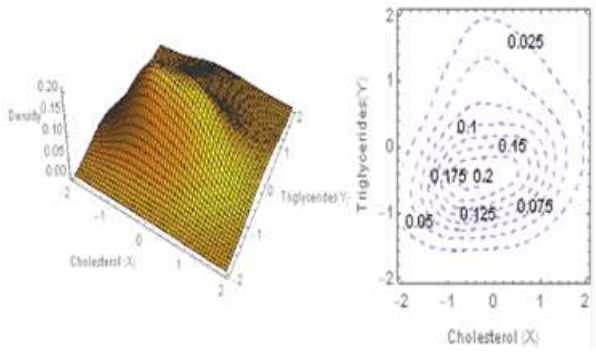


Figure-5. KEs (Surface and Contour plots) of H_D SP.

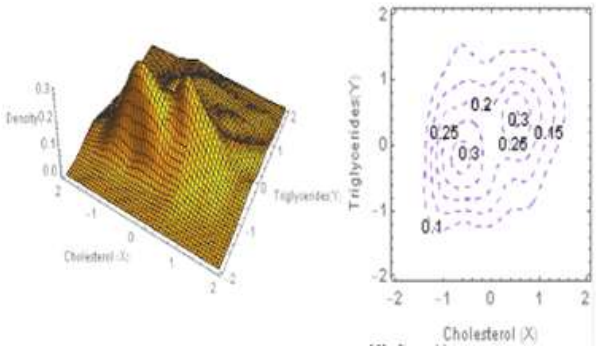


Figure-6. KEs (Surface and Contour plots) of H_p SP.

Table 2 shows the asymptotic integrated variance, asymptotic integrated squared bias, and asymptotic mean integrated squared error (AMISE) of the different forms of parameterizations for the second data set.

Table-2. Variance, Bias^2 and AMISE of Blood Fat Data Set

Methods	*Variance*	$Bais^2$	*AMISE*
$H_{C\text{-}AMISE}$	0.0015292045	0.0007645732	0.0022937777
$H_{D\text{-}AMISE}$	0.0015015908	0.0007509184	0.0022525092
$H_{F\text{-}AMISE}$	0.0011918386	0.0002405931	0.0014324317

The constant smoothing matrix and diagonal smoothing matrix produced similar estimates which are considerably over-smoothed and the bimodality of the data is difficult to identify as presented in Figure 4 and Figure 5. The full smoothing matrix produced an estimate with the bimodality being clearly present as shown in Figure 6. More clearly noticed from Table 2 is that the full smoothing matrix did better in terms of performance that is, it produced the smallest AMISE value when compared with the constant and diagonal smoothing matrices. Another very important issue in kernel

density estimation is its usefulness in highlighting structures of the data set and this vital role of highlighting and retaining structures is achieved in the estimate of the full smoothing matrix unlike the estimates of the constant and diagonal smoothing matrices where the inherent features of the data are smoothed away and presenting the data to be unimodal. The bimodal features of the data are of great importance because important medical decisions or advice can be made only when accurate statistical results are obtained graphically and it also simplifies the interpretation of results via visualization.

4. Conclusion

This study investigates the performance of smoothing matrices in multivariate kernel density estimation with an emphasis on the bivariate case using the constant, diagonal, and full smoothing parameterizations. KDE is primarily for data analysis (Nwankwo & Olayinka, 2019, Nwankwo & Ukhurebor, 2019) visualization. Visualization is becoming an interesting aspect of organizational security architecture and toolset (Nwankwo, 2020, Nwankwo & Ukaoha, 2019, Goldfarb, 2017). While these estimators are very relevant for data presentation purposes especially as it affects a distribution to users for appropriate decision making, the results are best supported by the full smoothing matrices. The AMISE's values show that the full smoothing matrices outperformed the constant and diagonal smoothing matrices for the bivariate KDE. The result shows that full smoothing matrices can give markedly better performance when compared with the constant and diagonal smoothing matrices.

References

Duong, T. and Hazelton,) M. L. (2003. Plug-In Bandwidth Matrices for Bivariate Kernel Density Estimation. *Journal of Nonparametric Statistics*, 15(1):17–30.

Goldfarb, J. (2017). Data Visualization: Keeping an Eye on Security. Available at https://www.darkreading.com/threat-intelligence/data-visualization-keeping-an-eye-on-security/a/d-id/1328493

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

Jarnicka, J. (2009). Multivariate Kernel Density Estimation with a Parametric Support. Opuscula Mathematica. 29(1):41– 45

Nwankwo, W. (2020). A Review of Critical Security Challenges in SQL-based and NoSQL Systems from 2010 to 2019. *International Journal of Advanced Trends in Computer Science and Engineering*, 9(2).

Nwankwo, W., Ukaoha, K.C. (**2019**). Socio-Technical perspectives on Cybersecurity: Nigeria's Cybercrime Legislation in Review*; International Journal of Scientific and Technology Research,* 8(9), 47-58.

Nwankwo, W. and Olayinka, A.S. (2019). Real-time Risk Management and X-ray Cargo Scanning Document Management Prototype for Trade Facilitation. International *Journal of Scientific and Technology Research*. 8(9):93- 105.

Nwankwo, W., and Ukhurebor, K.E. (2019). Small and medium-scale software contracts: From initiation to commissioning*, International Journal of Scientific and Technology Research*, 8(12), pp.1554-1563

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.

Sain, R. S. (2002). Multivariate Locally Adaptive Density Estimation*. Computational Statistics and Data Analysis*. 39:165– 186.

Sain, R. S., Baggerly, A. K. and Scott, D. W. (1994). Cross-validation of Multivariate Densities. *Journal of Statistical Association*. 89:807–817.

Sabine, L. and Brian, S. E. (2004). A Handbook of Statistical Analyses using SPSS. Chapman & Hall/CRC, New York.

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

Scott, D. W., Gotto, A. M., Cole, J. S., and Gorry, G. A.(1978). Plasma Lipids as Collateral Risk Factors in Coronary Heart Disease. A Study of 371 Males with Chest Pain. *Journal of Chronic Diseases*.31:337–345.

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.

Siloko, I. U., Ikpotokin, O., Oyegue, F. O., Ishiekwene, C. C., and Afere, B. A. (2019). A Note on Application of Kernel Derivatives in Density Estimation with the Univariate Case. *Journal of Statistics and Management Systems.* 22(3): 415–423.

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

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

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

Zhange X., Wu X., Pitt D. And Liu Q. (2011). A Bayesian Approach to Parameter Estimation for Kernel Density Estimation via Transformation. *Annals of Actuarial Science*. 5(2): 181–193.