On Evaluation of Smoothing Matrix Performance in Multivariate Kernel Density Estimation

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

[1]Department of Mathematics & Computer Science, Edo University Iyamho, Nigeria.
[2]Department of Mathematics & Statistics, Ambrose Alli University, Ekpoma, Nigeria
[3]Department of Mathematics, University of Benin, Benin City, Nigeria
e-mail: [1]siloko.israel@edouniversity.edu.ng

Abstract **— The multivariate kernel density estimator unlike the univariate case demands more than one smoothing parameter for its axes depending on the smoothing parameterizations employed. The diagonal and the full parameterizations are the common forms of parameterizations of the multivariate kernel density estimator. This paper investigates the performance of these smoothing parameterizations in multivariate density estimation with emphasis on the bivariate kernel density estimator in practical application using the asymptotic mean integrated squared error as the error criterion function. The result of the investigation reveals that the full smoothing parameterization did better than the diagonal parameterization in terms of performance with real data examples.**

Keywords - ***Smoothing matrix, Kernel Density Estimator, Integrated Variance, Integrated Squared Bias, Asymptotic Mean Integration Squared Error (AMISE).***

I. INTRODUCTION

Nonparametric density estimation is of wide applications with the kernel estimator as one of its popular techniques in data smoothing. The wide applicability of this estimator is due to the easy of its implementation (Schauer et al, 2013). Kernel estimation is a data smoothing strategy where inferences and conclusions could be made about the set of random variables under consideration (Duong, 2004). As a nonparametric method for estimating probability density, kernel estimation is a very useful tool for analysis and visualization of the distribution of a data set (Silverman, 1986; Simonoff, 1996).

The general multivariate form of the kernel density estimator introduced by Deheuvels (1977) is of the form

$$\hat{f}(\mathrm{x}) = \frac{1}{n|H|^{1/2}} \sum_{i=1}^{n} K\left(\frac{\mathrm{x} - X_i}{H^{1/2}}\right), \qquad (1)$$

where $\boldsymbol{n}$ is the sample size, $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_d)^{\mathrm{T}}$, $\mathbf{X}_i = (\mathbf{X}_{i1}, \mathbf{X}_{i2}, \dots, \mathbf{X}_{id})^{\mathrm{T}}$, $\boldsymbol{i} = \mathbf{1}, \mathbf{2}, \dots, \boldsymbol{n}$ and $\mathbf{H}$ is the smoothing matrix that is symmetric and positive definite. The data set $\mathbf{X}_i$ are usually observations or measurements obtained from real life. The kernel function $\boldsymbol{K}(\mathbf{x})$ is a multivariate function which is a symmetric probability function (Scott, 1992; Wand and Jones, 1995) that satisfies the conditions

$$\int K(\mathrm{x})\,d\mathrm{x} = 1\,, \int \mathrm{x}K(\mathrm{x})\,d\mathrm{x} = 0 \text{ and}$$

$$\int \mathrm{x}\mathrm{x}^T K(\mathrm{x})\,d\mathrm{x} = \mu_2(K) I_d. \qquad (2)$$

The conditions in Equation (2) are satisfied by all kernels with the kernel function taken to be a d-variate probability density function. The first condition in Equation (2) implies that the sum of the marginal kernels are equal to one, the second condition simply states that the means of the marginal kernels are all zero, and the third condition means that the marginal kernels are all pairwise uncorrelated and with unit variance in each dimension (Scott, 1992; Wand and Jones, 1995).

The kernel estimator in Equation (1) is a useful tool for data exploratory analysis and data visualization especially for bivariate data when $\hat{\boldsymbol{f}}(\mathbf{x})$ can be visualized using the familiar perspectives or contour plots (Silverman, 1986; Scott, 1992; Simonoff, 1996) and also has applications in discriminant analysis and goodness-of-fit testing (Duong and Hazelton, 2003; Duong, 2004). The choice of $\boldsymbol{H}$ is very important to the performance of $\hat{\boldsymbol{f}}(\mathbf{x})$ either in the diagonal matrix form or in the full matrix form.

The purpose of this paper is to compare the performance of the estimator $\hat{\boldsymbol{f}}(\mathbf{x})$ using the diagonal smoothing matrix and the full smoothing matrix in the bivariate case with the asymptotic mean integration square error (AMISE) as the criterion function. The rest of the paper is organized as follows. In section 2, we state the form of the bivariate kernel estimator while section 3 state the asymptotic mean integrated squared error of the multivariate kernel density estimator with brief discussion of forms of parameterizations. Section 4 contains a comparative study of the two forms considered using real data examples and section 5 concludes the paper.

II. THE BIVARIATE KERNEL DENSITY ESTIMATOR

A very natural and important use of the bivariate kernel density estimates is the investigation of the properties of a set of data such as skewness and modality largely because they can be viewed using the surface plots or contour plots (Silverman, 1986; Duong and Hazelton, 2003). The bivariate kernel density estimator is a special case of the multivariate kernel density estimator that deals with random variables taking values in $\mathcal{R}^2$.

The bivariate kernel density estimator is use for the production of two-dimensional diagram in the case of contour plots and three-dimensional diagram in the case of surface plots of the distribution of two variables. In the bivariate kernel density estimator, $\mathbf{x}, \mathbf{y}$ are taken to be the random variables taking values in $\mathcal{R}^2$ and they have a joint density function $\boldsymbol{f}(\mathbf{x}, \mathbf{y}), (\mathbf{x}, \mathbf{y}) \in \mathcal{R}^2$ with $\mathbf{X}_i, \mathbf{Y}_i$, $\boldsymbol{i} = \mathbf{1}, \mathbf{2}, \ldots, \boldsymbol{n}$ being the set of observations of size $\boldsymbol{n}$ drawn from the distribution. The kernel density estimate of $\boldsymbol{f}(\mathbf{x}, \mathbf{y})$ base on this sample is of the form

$$\hat{f}(\mathrm{x}, \mathrm{y}) = \frac{1}{n h_{\mathrm{x}} h_{\mathrm{y}}} \sum_{i=1}^{n} K\left(\frac{\mathrm{x} - \mathrm{X}_i}{h_{\mathrm{x}}}, \frac{\mathrm{y} - \mathrm{Y}_i}{h_{\mathrm{y}}}\right), \qquad (3)$$

where $h_{\mathrm{x}} > 0$ and $h_{\mathrm{y}} > 0$ are the smoothing parameters in the X and Y axes and $K(\mathrm{x}, \mathrm{y})$ is a bivariate kernel function which is usually the product of two univariate kernels. This implies that the bivariate kernel estimator can be written as (Zhang *et al.,* 2011)

$$\hat{f}(\mathrm{x}, \mathrm{y}) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{h_{\mathrm{x}}} K\left(\frac{\mathrm{x} - \mathrm{X}_i}{h_{\mathrm{x}}}\right) \frac{1}{h_{\mathrm{y}}} K\left(\frac{\mathrm{y} - \mathrm{Y}_i}{h_{\mathrm{y}}}\right). \qquad (4)$$

The importance of the bivariate kernel density estimator cannot be overemphasized because it occupied a unique position of bridging the univariate kernel estimator and other higher dimensional kernel estimators. Again, the bivariate kernel density estimates are also very simple to understand and interpret, either as surface plots or contour plots (Silverman, 1986 Duong and Hazelton, 2003).

III. ASYMPTOTIC MEAN INTEGRATED SQUARE ERROR APPROXIMATIONS

The quality of the estimate $\hat{\boldsymbol{f}}(\mathbf{x})$ in Equation (1) is measured by the asymptotic mean integrated squared error (AMISE) defined as

$$\begin{aligned} \mathrm{AMISE}\{\hat{f}(\mathrm{x}; H)\} &= \mathrm{E} \int \left(\hat{f}(\mathrm{x}; H) - f(\mathrm{z})\right)^2 d\mathrm{x} \\ &= \int \left(\mathrm{E}\,\hat{f}(\mathrm{x}; H) - f(\mathrm{x})\right)^2 d\mathrm{x} \\ &\quad + \int \mathrm{Var}\left(\hat{f}(\mathrm{x}; H)\right) d\mathrm{x} \end{aligned}$$

$$= \int \mathrm{Bias}^2\left(\hat{f}(\mathrm{x}; H)\right) d\mathrm{x} + \int \mathrm{Var}\left(\hat{f}(\mathrm{x}; H)\right) d\mathrm{x}. \qquad (5)$$

This gives the asymptotic mean integrated squared error (AMISE) as the sum of the asymptotic integrated squared bias and the asymptotic integrated variance (Silverman, 1986; Wand and Jones, 1995). The asymptotic integrated squared bias and the asymptotic integrated variance can be obtained by using the Multivariate Taylor's series expansion of $\hat{\boldsymbol{f}}(\mathbf{x})$. Therefore the asymptotic integrated squared bias is of the form

$$\left(\int Bias\hat{f}(\mathrm{x}; H) dx\right)^2 \approx \frac{1}{4}\mu_2(K)^2 \int tr^2\left\{\mathrm{H}\mathcal{H}_f(\mathrm{x})\right\} d\mathrm{x}. \qquad (6)$$

Also the asymptotic integrated variance is given by

$$\int Var\hat{f}(\mathrm{x}; H) dx \approx n^{-1}|\mathrm{H}|^{-1/2} R(K). \qquad (7)$$

The estimate of the asymptotic mean integrated squared error (AMISE) is obtain by the combination of the terms in Equation (6) and Equation (7) given as

$$\begin{aligned} \mathbf{A}\mathrm{MISE}\{\hat{f}(\mathrm{x}; H)\} &\approx n^{-1}|\mathrm{H}|^{-1/2} R(K) \\ &\quad + \frac{1}{4}\mu_2(K)^2 \int tr^2\left\{\mathrm{H}\mathcal{H}_f(\mathrm{x})\right\} dx, \end{aligned} \qquad (8)$$

where $|\cdot|$ is the determinant of the smoothing matrix, $R(K)$ 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 (Scott, 1992; Wand and Jones, 1995; Sain, 2002). There is no explicit expression for the AMISE-optimal smoothing matrix in the form of Equation (8) (Scott, 1992; Wand and Jones, 1995; Chacón, 2009).

In choosing the smoothing matrix H, the complexity of the underlying density function and the number of parameters to be estimated must be considered (Sain, 2002). Generally, the commonest parameterizations of the smoothing matrix in the multivariate case are the diagonal parameterization and the full parameterization provided the matrix H is symmetric and positive definite. Assuming K is a multivariate standard d-variate normal kernel, then

$$K(\mathrm{x}) = (2\pi)^{-d/2}\exp\left(-\frac{\mathrm{x}^{\mathrm{T}}\mathrm{x}}{2}\right). \qquad (9)$$

In the case of the multivariate product kernel estimator, Equation (8) can now be written as

$$AMISE = \frac{R(K)^d}{nh_1h_2,\ldots,h_d} + \frac{1}{4}h_j^4\mu_2(\mathrm{K})^2\int tr^2\{\mathcal{H}_f(\mathrm{x})\}d\mathrm{x}. \quad (10)$$

The smoothing parameter that minimizes the $AMISE$ of Equation (10) is given by

$$H_{AMISE} = \left[\frac{dR(K)^d}{\mu_2(\mathrm{K})^2\int tr^2\{\mathcal{H}_f(\mathrm{x})\}d\mathrm{x}}\right]^{\left(\frac{1}{d+4}\right)} \times n^{-\left(\frac{1}{d+4}\right)} (11)$$

This choice of H_{AMISE} will yield an $AMISE = O\left(n^{-\left(\frac{4}{d+4}\right)}\right)$ and the smoothing parameter values are of order $n^{-1/(d+4)}$ where d is the dimension of the kernel (Sain, 2002).

In the case of the bivariate kernel estimator given in Equations (3) and (4) above, the bivariate standard normal function is of the form

$$K(\mathrm{x},\mathrm{y}) = \frac{1}{2\pi}\exp\left(-\frac{\mathrm{x}^2+\mathrm{y}^2}{2}\right). \qquad (12)$$

The smoothing parameterizations of the bivariate case that is consider are of the forms

$$H = \begin{bmatrix} h_x^2 & 0 \\ 0 & h_y^2 \end{bmatrix} \text{ and } H = \begin{bmatrix} h_x^2 & h_{xy} \\ h_{xy} & h_y^2 \end{bmatrix}. \qquad (13)$$

The performance of these forms of parameterizations will be evaluated using the AMISE as the error criterion function.

IV. Results

In this section, we will compare the performance of the diagonal smoothing matrix with the full smoothing matrix using some real data examples. We will represent the smoothing matrix that minimizes the asymptotic mean integrated squared error (AMISE) in the case of the diagonal smoothing matrix by $H_{D-AMISE}$ and that of the full smoothing matrix by $H_{F-AMISE}$. Figures 4.1; 4.2; 4.3; and 4.4 shows the kernel estimates of the two forms of parameterizations considered.

The first data set examined involves the locations of centers of craters of 120 volcanoes in the Bunyaruguru volcanic field in Western Uganda (Bailey and Gatrell, 1995). A map of the distribution shows a broad regional trend in a North-Easterly direction, representing elongation along a major fault. These sets of data were bimodal, indicating the major centres where the volcanic activities occurred. The data were standardized in order to obtain equal variances in each dimension because in most multivariate statistical analysis, the data should be standardized to ensure that the differences among the ranges of variables disappear (Wand and Jones, 1993; Simonoff, 1996 and Cula and Toktamis, 2000; Sain, 2002). The smoothing matrices for the two forms are

$$H_{D-AMISE} = \begin{bmatrix} 0.480183 & 0.000000 \\ 0.000000 & 0.480533 \end{bmatrix} \text{ and } H_{F-AMISE} = \begin{bmatrix} 0.538986995 & -0.00004873 \\ -0.00004873 & 0.539380057 \end{bmatrix}$$

Figure 4.1 and Figure 4.2 below shows the kernel estimates that is the familiar perspectives (surface plots) and the contour plots of the two forms of smoothing parameterizations using the bivariate standard normal product kernel

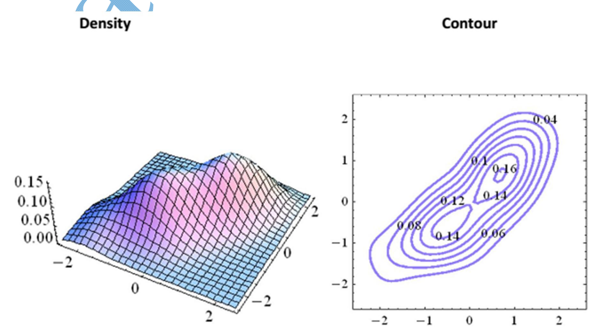

Fig 4.1: Kernel Estimates (Surface and Contour plots) of H_D Smoothing Parameter

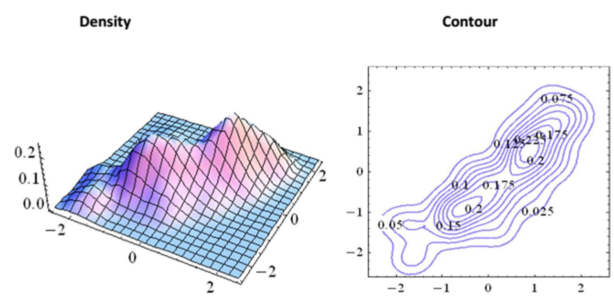

Fig 4.2: Kernel Estimates (Surface and Contour plots) of H_F Smoothing Parameter

Table 4.1 shows the asymptotic integrated variance (AIV), the asymptotic integrated squared bias (AISB) and the asymptotic mean integrated squared error (AMISE) for the first data set

Table 4.1: Variance, Bias^2 and AMISE for First Data Set.

Methods.	***AIV***	***AISB***	***AMISE***

$\mathbf{H}_{D\text{-}AMISE}$	0.002873	0.001436	0.004310
$\mathbf{H}_{F\text{-}AMISE}$	0.002281	0.000184	0.002465

The analysis in Table 4.1 clearly shows that the full smoothing matrix did better than the diagonal smoothing matrix in terms of performance. As generally known, one method is better than the other one when it gives a smaller value of the AMISE (Jarnicka, 2009). However, both parameterizations retained the bimodality of the observed data. Also both parameterizations exemplify the usefulness of the bivariate kernel density estimates for highlighting structures in a data set.

The second data set examined is the blood fat concentration data also known as the lipid data of Scott *et al.* (1978). These data consist of measurements of cholesterol and triglycerides for 320 men diagnosed with coronary artery disease and the original paper showed that the data were bimodal; indicating an increased risk for heart disease is associated with increased cholesterol level (Sain, 2002). The data were standardized to obtain equal variances in each dimension. The smoothing matrices for this data set are

$$\boldsymbol{H}_{D-AMISE} = \begin{bmatrix} 0.403262 & 0.000000 \\ 0.000000 & 0.410678 \end{bmatrix} \text{ and } \boldsymbol{H}_{F-AMISE} = \begin{bmatrix} 0.4526467 & 0.0021664 \\ 0.0021664 & 0.4609705 \end{bmatrix}$$

Figure 4.3 and Figure 4.4 shows the kernel estimates, which are the surface plots and the contour plots of the two forms of smoothing parameterizations considered, using the bivariate standard normal product kernel.

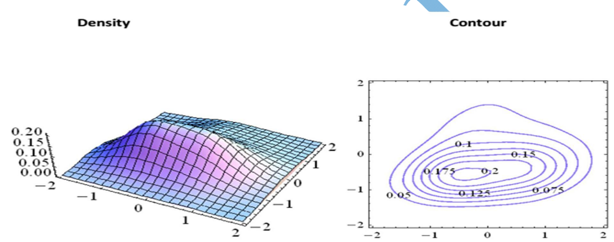

Fig 4.3: Kernel Estimates (Surface and Contour plots) of H_D Smoothing Parameter

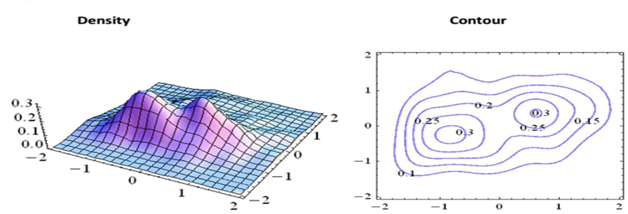

Fig 4.4: Kernel Estimates (Surface and Contour plots) of H_F Smoothing Parameter

Table 4.2 shows the asymptotic integrated variance (AIV), the asymptotic integrated squared bias (AISB) and the asymptotic mean integrated squared error (AMISE) of both forms of parameterizations for the second data set.

Table 4.2: Variance, Bias^2 and AMISE for Second Data Set.

Methods.	*AIV*	*AISB*	*AMISE*
$\mathbf{H}_{D\text{-}AMISE}$	0.001502	0.000751	0.002253
$\mathbf{H}_{F\text{-}AMISE}$	0.001192	0.000241	0.001432

The diagonal smoothing matrix produced an estimate that is considerably oversmoothed and it is difficult to identify the bimodality discussed in Scott *et al.* (1978) as seen in Figure 4.3. The full smoothing matrix produced an estimate with the bimodality being clearly present as shown in Figure 4.4. More clearly noticed from Table 4.2 is that the full smoothing matrix did better in terms of performance than the diagonal smoothing matrix. Another very important issue in kernel density estimation is its usefulness in highlighting structures in the data set and this was achieved with the estimate of the full smoothing matrix unlike the estimate of the diagonal smoothing matrix.

V. Conclusion

This paper examined the performance of the smoothing matrix in multivariate kernel density estimation with emphasis on the bivariate kernel estimator using the diagonal smoothing parameterization and the full smoothing parameterization. The kernel estimates and the AMISE values show that the full smoothing matrices perform better than the diagonal smoothing matrices with respect to the bivariate kernel density estimator. The results show that full smoothing matrices can give markedly better performance when compare to the diagonal smoothing matrices.

References

Chacón, J. E. (2009). Data–Driven Choice of the Smoothing Parametrization for Kernel Density Estimators. The Canadian Journal of Statistics. 37:249-265.

Cula, S. G and Toktamis, O. (2000). Estimation of Multivariate Probability Density Function with Kernel Functions. Journal of the Turkish Statistical Association. 3(2):29–39.

Deheuvels, P. (1977). Estimation Nonparametrique de La Densité Par Histogrammes Generalisés (II). Publ. L´Univ. Paris. 22: 1–23.

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

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

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.

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

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.

Wand, M. P. and Jones, M. C. (1993). Comparison of Smoothing Parameterizations in Bivariate Kernel Density Estimation. Journal of the American Statistical Association. 88:520–528.

Zhang, 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.

Bailey, T. C. and Gatrell, A. C. (1995). Interactive spatial data analysis. Longman, Harlow.

Duong, T. (2004). Bandwidth Selectors for Multivariate Kernel Density Estimation. A Thesis Presented to the School of Mathematics and Statistics for the Degree of Doctor of Philosophy at the University of Western Australia.

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

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.

.