Journal of the Nigerian Association of Mathematical Physics
Volume34,(March, 2016), pp123 – 132

On A Modified Multivariate Cluster Sampling Kernel Approach to Multivariate Density Estimation

[1]Ogbeide E. M,[2]Osemwenkhae J.E. and[3]Oyegue F.O.

[1]Ambrose Alli University, Ekpoma, Edo State, Nigeriaand[2,3]University of Benin, Nigeria.

Abstract

A modified cluster sampling multivariate kernel density estimation (MMCKDE) approach is proposed. This approach is based on the relevant ideas of estimating the population clusters from the data set. Basically, we form empirical clusters samples, which are observations grouped into data cluster either via rows or columns information according to the empirical cluster they belong. The bandwidth parameters derived by these approaches based on the data set clusters are used to smooth the density. The estimates from the proposed approach showed some improvements over the existing methods and the fixed bandwidth method with real life data. This approach compensate for discontinuities in the estimated density curve (pilot plot) by some adjustment/modification to avoid or correct such discontinuities.

1.0 Introduction

Data density estimation provides a nonparametric estimate of the probability function from which a set of data is drawn. In pattern recognition, optimal algorithms often require the knowledge of underlying densities of signal and/or noise. Primarily, it is better to estimate the density from the data. In density estimation, the true density is unknown. Researchers have shown that most real life problems are multivariate in nature. This work is based on the multivariate density estimation which provides estimates of the probability function from which a set of data is drawn. One of the most well-known and popular techniques of density estimation is the kernel density estimation (KDE). It is a nonparametric estimation approach which requires a kernel function and a window size (smoothing parameter *H*). In this study, we proposed adaptive approach that is based on data at hand.

Furthermore, Researchers have shown that estimates based on the varying window sizes in estimating density are superior to estimates based on optimal constant-sized window size –[1-3]. Also, it has been widely regarded that the performance of the kernel methods depends largely on the smoothing parameter (window width) but depends very little on the kernel. We observed most times, analyses of multivariate data are more prevalent in practice than the univariate cases [4-5]. The crucial problem in the multivariate kernel density estimation (MKDE) is to select the window widths (bandwidth parameters) *H*. The window widths control the smoothness of the fitted density curve. In literature, studies considering the problem of window size selection in MKDE exist–[6,7,8,9,1,2,4,10,11], each exploring possible ways of improving on the smoothing effects of the window sizes.

Considering the variable window sizes on the cluster sampling multivariate kernel density estimation approach for estimating densities, points for improvements were identified, so that the methods could be adaptive to the MKDE. In most cases, the above methods could lead to under fitting, an indication that the methods are often less optimal. In this research work, we propose data-driven approach that require only the knowledge of the use of pilot plots and the bandwidth sizes from the data set with a view to correcting the identified problems, while aiming for lower asymptotic mean integrated square error (AMISE) and a faster rate of convergence in the approach. The aim of this study is basically on how to fit density to multivariate data sets. The multivariate kernel density estimator that we are going to study is a direct extension of the univariate estimator. Let $X_1,..., X_n$ denote a d-variate random sample having a density f . We shall use the notation $X_i = (X_{i1},..., X_{in})^T$ to denote the X_i and a generic vector $x \in \Re^d$ has the representation $x = (x_1,..., x_d)^T$. The d-variate random sample $X_1,..., X_n$ drawn from f the kernel estimator evaluated at x is given by;

Corresponding author: Ogbeide, E. M, E-mail:ogbeideoutreach@yahoo.com, Tel.: +2348035910189

$$\hat{f}(X,H)=\frac{1}{n}\sum_{i=1}^{n}K_H(x-X_i) \qquad (1.1)$$

where *n* is the sample size, and H is a symmetric positive definite *d* x *d* matrix called the window widths, the smoothing parameters or the bandwidth matrix, $K_H(x)=|H|^{-\frac{1}{2}}K(H^{-\frac{1}{2}}x)$, $|\ .\ |$ stands for the determinant of H and K is d-variate kernel satisfying $\int k(x)dx=1$, where the integral is over $\Re^d$ unless stated otherwise. The kernel function is often taken to be a *d*-variate probability density function.

The matrix *H* is a smoothing parameter and specifies the 'width' of the kernel around each sample point X_i. A well behaved *K* (that is a kernel bounded compactly) must satisfy the following regularity conditions:

1. $\int_{R^d} K(w)dw=1$
2. $\int_{R^d} wK(w)dw=0$
3. $\int_{R^d} ww^T K(w)dw=I_d$

Where I_d is a *d* dimensional identity matrix.

The first condition accounts for the fact that the sum of the kernel function over the whole region is unity. The second condition imposes the equation constraint that the means of the marginal kernel $K_i(w_i), i=1,...,d$ are all zero. The third condition term states that the marginal kernels are all pairwise uncorrelated and each has unit variance. We shall apply the product kernel in this work.

However, the most important part of the estimator in (1.1) is the bandwidth matrix which contained the window sizes used for smoothing density. The fixed window size method are not sensitive to local peculiarities in the data, such as clustering/sparseness of sample value. Here the smoothing parameter *H* varies, hence the "adaptive" techniques. This new our approach attempt to compensate for the loose of some discontinuities in the estimated density curve (pilot plot) by some adjustment/modification to avoid or correct such discontinuities.

The motivation for this work arose from the works on the nearest neighbour approach by Loftsgaarden and Quesenberry [12], and the cluster approach to density estimation by Wu and Tsai [10,13]. These methods are adaptive to unknown data density. Generally, the true density is unknown, hence, the search for an approach to get the true density. Therefore, we adopt the advantage of varying the smoothing parameter which helps in determining the true density –[7,9,5]. Though these methods are adaptive, one is also tasked with how sensitive these methods are? What are the errors committed using these methods? These questions lead to the need for their improvement.

2.0 Literature Review

The literature for bandwidths choices in the multivariate kernel density estimation is not quite extensive, a number of the methods exist-see [4,9,10,14]. There exist some methods of estimating multivariate kernel density. Some of these methods use a fixed window width. However, the approach that uses varied window widths in the course of density estimation which seems adaptive are few. A review of available variable methods showed basically that the cross-validation, the plug-in bandwidths approaches or any subjective method (which are fixed smoothing approaches). There is the cluster and the average cluster approach by Wu and Tsai [13] and Wu et –al [10] which are more data sensitive are used in the MKDE. The window width controls the smoothness of the fitted density curve. The true density is unknown.

There is stata approach to nonparametric estimation of density functions. Its various versions include; *k*density by Salgado-Ugarte et al [15,16,17] for bandwidth selection and estimation and the Stata module *ak*density approach by Philipe [18]. These approaches are software modification of fixed window size with pre-stated adjustments. Their usage depends on the choice of the users. Their efficiencies are as determined by their developers. They have smooth curve. They do not adapt to small data sizes. Even if users adjust sample sizes, the software uses its self commands. They use large sample size approximation in their evaluation. According to Bowman and Azzalini [1], software approaches serves as guides as stated in the user's manual.

Another adaptive method developed is the ICI approach by Katkovnik [19], and Katkovnik and Shmulevich [12] for univariate kernel density estimation, which may be extended to the multivariate density with appropriate modification. There is bootstrap choice of smoothing parameter.This is a multiple resampling technique used for determining the best choice for *h* or the confidence interval on which it can be chosen, when MISE*{*f**(X;*H*)} is obtained. See [7,9,20,21]. There is also the smoothing by weighted average of rounded points discussed in [7] and the Mean shift approach to KDE by Comaniciu and Meer [22] for univariate data density. It is an adjustment using the fixed smoothing parameter method.

Cluster sampling approach to MKDE in [13,10,3] uses the information matrix rows and columns to form clusters for sampling, where the cluster sizes and bandwidths factors are use to achieve the smoothing parameters. The method is adaptive, but with some points of discontinuities. Wu and Tsai in [13] first proposed the cluster sampling technique to MKDE. There is the **Mean shift-based clustering** approach to MKDE in [3] which utilizes the mean shift (average difference) of the successive bandwidths. Wu et –al in [3] worked on the average cluster sampling approach to density estimation. The multivariate cluster sampling kernel density estimate (MCKDE) adjust the amount of bandwidths using some idea from cluster sampling estimation of the multivariate data set. Its smoothing parameter is an *n* dimensional matrix obtained from forming relevant number of clusters in an information matrix. To correct the problem of discontinuities at some points in the cluster sampling approach to MKDE, Wu et al in [10] suggested the average cluster sampling approach to density estimation. In this case, the bandwidth factors b_i *to* X_i according to the number of clusters formed are calculated. They use the average values of these factor to choose the bandwidths *H*. Wu et –al in [3] mainly use the average cluster method which reflects the average number of clusters formed. These are attempts to reduce the discontinuities experienced in the Cluster sampling approach to MKDE.

When we consider the variable window sizes works on the Cluster sampling approach to MKDE though the methods are adaptive, one is tasked with how sensitive are these methods? What are the errors committed using these methods? These questions led to the reasons for their modifications. We identified points for improvements, so that the methods could be more adaptive.

3.0 Methodology

3.1 The Modified Multivariate Cluster sampling Kernel Density Estimation (MMCSKDE)

The most commonly used optimality criterion for selecting a bandwidth matrix is the mean integrated squared error (MISE)

$$MISE(H) = E\{\int [\hat{f}_h(X) - f(X)]^2 dX \quad (3.1)$$

where $\int$ is a shorthand notation for $\int_{R^n}$ and X is in n Euclidean plane R^n

This equation (3.1) is in general does not have a closed- form expression, so we result to its asymptotic approximation (AMISE). Hence (3.1) could be factored as

$$AMISE(H) \approx n^{-1}|H|^{-\frac{1}{2}} R(K) + \tfrac{1}{4} m_2(K)^2 (vec^T H)\psi_4 (vec^T H) \quad (3.2)$$

where

- $R(K) = \int K(X)^2 dX$, with $R(K) = (4\pi)^{-\frac{d}{2}}$ when K is a normal kernel.
- $\int XX^T K(X)^2 dX = m_2(K) I_d$, with I_d *being the* $d \times d$ *identity matrix* and $m_2 = 1$ for the normal kernel.
- $D^2 f$ *is* $d \times d$ Hessian matrix of second order partial derivatives of f.
- $\psi_4 = \int (vecD^2 f(X))(vec^T D^2) dX$
- D is a diagonal matrix with elements $X_{11}, X_{22}, ..., X_{dd}$
- vec is the vector operator which stacks the columns of a matrix into a single vector. For example, $vec\begin{bmatrix} a & c \\ b & e \end{bmatrix} = [a \quad b \quad c \quad e]^T$, see [25].

We observed that the quality of the AMISE to the MISE is given by

$$MISE(H) = AMISE(H) + o(n^{-1}|H|^{-\frac{1}{2}} + trH^2) \quad (3.3)$$

where o indicates the usual o notation. This implies that AMISE is a ‘good’ approximation of the MISE as $n \to \infty$. It has been shown that optimal bandwidth selector H has $H = O(n^{-\frac{2}{(d+4)}})$. Substituting this into equation (3.3) yield the optimal $MISE(H)$ order as $O(n^{-\frac{4}{(d+4)}})$ - [25]. The big O notation is applied element-wise. Thus when $n \to \infty$, $MISE \to 0$. This implies the kernel density estimate converges in mean squared error and so also in probability to the true density f. According to Wand and Jones and Horova et al, they asserted that it was better to estimate optimal MISE element-wise.[9,11]. They further asserted that the ideal optimal bandwidth selector that is point wise adaptive is given by

$$H_{AMISE} = agr \min_{h \in \mathrm{H}} A\tilde{MISE}(H) \qquad (3.4)$$

Since this ideal bandwidth selector contains the unknown density function f, that cannot be used directly. So some data density based approaches fixed the choice of bandwidth constant. However, we shall adopt point-wise adaptive bandwidth procedures in estimating densities.

The bandwidths used for the cluster approach by Wu et al are optimal for information row/column (one dimensional) bandwidth per time in the multivariate data set [24]. That is, it uses one bandwidth in the row or column during row/column cluster bandwidth selection. It is only row or column adaptive. Our approach is to make bandwidth selection to be data based on the smallest size of the row or column samples selections from the information matrix.

Our procedure is basically a minimization of $AMISE(H)$ with respect to H, where it is equivalent to the selection of optimal $h_{ij}\ in\{H_1, H_2, ..., H_n\}$. This method is a modification of the cluster sampling approach to density estimate. The modified multivariate cluster sampling kernel density estimate (MMCSKDE) is a modification of cluster sampling kernel density estimate by adjusting the amount of bandwidths using some idea from the kernel nearest neighbour estimation of the density to the multivariate data. Its smoothing parameter would be a $n \times d$ dimensional matrix obtained from forming relevant number of clusters in an information matrix. The Euclidean distance would be used to form bandwidths .

Let $h = h^* b$. According to Silverman, we call b the bandwidth factor and h^* the global smoothing parameter [7,23]. The common procedure is to first choose *b* adaptively and then h^*, by regarding *b* as fixed. But Wu et al in [10] used $h = h^* b_i$, where $b = (b_1, ..., b_n)$ are the bandwidths factors reflecting the average local clusters from X_i and adopt the stabilized fixed bandwidths selector of Wu and Tsai in [13] to select the global smoothing parameter. This approach gives a diagonal bandwidth matrix of varying smoothing parameters h_i. In our proposed approach, since we aim at element wise adaptive density estimation for any given data set X_{ij}.

We have more bandwidth factors according to the number of clusters form (starting from step 3 in the proposed algorithm) from the element wise groups from the information data rows. Then we have

$$H = h_i^* b_{i^* j} \qquad (3.5)$$

where $i = 1, ..., n$, $i^* = 1, ..., n_i$ and $j = 1, ..., d$.

with H a finite set of optimal bandwidths $H = H_1, ..., H_n$ and each $H_i = h_{ij}$, we choose our h_i^* via each information data rows' h_{MSE}. That is using the MSE approach to get each h_i^*. This is more data sensitive to any fixed h^*.

$$\text{Where} \quad h_i^* = \begin{pmatrix} h_1^* & . & . & . & h_1^* \\ . & & & & \\ . & & & & \\ . & & & & \\ h_n^* & . & . & . & h_n^* \end{pmatrix}^t \qquad (3.6)$$

Then $b_{i^* j}$ will be small if as n_{i^*} is large (that is a large number of mergers involving X_i). Basically, from the data set, the above scheme clusters are formed from the nearest nested sequence of clusters information data rows' elements

$$\{X_i\} = C_{i0} \subset C_{i1} \subset ... \subset C_{in_{i^*}} = \{X_1, ..., X_n\} \qquad (3.7)$$

This procedure gives a full bandwidth matrix of vary smoothing parameters for possible values of data sizes for *i* rows and *j* columns. $i \le j\ and \quad i > j$.

To correct the problem of discontinuities at some points in the cluster sampling approach to MKDE, points of discontinuity in the estimation are identified using the cluster sampling approach as a pilot guide. In this case, the use of standard techniques from cluster analysis is applied. Here, a modified sampling idea similar to [13] is developed. In this case, when we consider the bandwidth factor $b_i\ to\ X_i$ according to the number of clusters form, and use the idea of density at the boundaries to choose the bandwidths *H*. Wu and Tsai used the average cluster method which reflects the average local clustering form. In this work a proposed scheme to address points of discontinuities is suggested.

We supposed that f is a density function such that $f(x)=0$ for $x<0$ and $f(x)>0$ for $x \geq 0$. We further suppose that f'' is continuous away from $x=0$. Then, we have $\hat{f}(x;h)=\int_{-1}^{\alpha} k(z) f(x-hz)dz,$ where $0 \leq \alpha \leq 1$ -see [9](p.46-47). Then at the boundary they obtained

$$E\hat{f}(0;h)=\tfrac{1}{2} f(0)+O(h) \qquad (3.8)$$

We use this idea base on the intuitive knowledge of kernel estimator having to find a compromise between estimating two distinct values of *f* on either side of discontinuity. We propose the use of semi inter-quartile range at the boundary values. Since the location of the boundary of $\hat{f}(x;H)$ is usually known, we adopted this to achieve better performance in its vicinity. Suppose, we have for S number of row clusters and T number of column clusters, we have;

$$d_{ST}=\sum_{i=1}^{n_S}\sum_{j=1}^{n_T} d_{ij} \qquad (3.9)$$

where $d_{ij}=\sqrt{\sum_{i=1}^{n}\sum_{j=1}^{d}(X_{ij}-X_{i+1\,j+1})^2}$ see Gray for lengths and distances' details [27].

$$H=H_i=\frac{H_i}{v} \text{ and } H_{i+1} \leq H_i. \qquad (3.10)$$

where $H_i=\{h_{ij}\}$. Subjectively we adopt $v=2$, where v is a positive real number.

The bandwidth sizes obtained are substituted into equation (1.1) above to obtain accompanying density estimates. The proposed algorithm is presented below

The modified procedures are:

Step 1: start with n clusters, each containing a single observation and an *n*x*n* symmetric matrix of distances $D=\{d_{ij}\}$.

Step 2: Search the distance matrix for the nearest pair of clusters. Let the distance between the ''nearest'' clusters S and T be $d_{ST}=\sum_{i=1}^{n_S}\sum_{j=1}^{n_T} d_{ij}$ in the case of observation *i* in the cluster S and observation *j* in the cluster T, and n_S *and* n_T are the number of observations in cluster S and cluster T, respectively.

Step 3: Merge (combine) cluster S and T. Label the newly formed cluster (ST). update the entries in the distance matrix by (a) deleting the row's element and column's element corresponding to clusters S and T elements and (b) adding a row's element and a column's element giving the distances between cluster (ST) and the remaining clusters elements.

Step 4: Repeat steps 2 and 3 a total of $n-1$ times so that all observations will be in a single cluster at termination of the algorithm. Record the clusters that are merged and the distance levels at which the mergers take place.

Step 5: Let b_{i^*j} distance level of X_i in the dendrogram. Specifically, if n_{i*} denotes the total number of times that a cluster containing X_i is merged into a larger cluster (that is, total number of mergers that involve X_i), and $\ell_{1i^*},\ldots,\ell_{n_i i^*}$ the distance level at which these n_i mergers take place, then $b_{i^*j} \equiv \ell_{1i^*},\ldots,\ell_{n_i i^*}$.

Step 6: generate $H_i=h_i^* b_{i^*j}$ where h_i^* are determined via the MSE for each information data rows, and let each $H_i=h_{ij}$.

Step 7: In the case of discontinuities, begin by applying (a) $d_{ST(Opt)}=H=\frac{H_i}{2}$ and (b) $H_{i+1} \leq H_i$ in the identified points in H_i from the pilot plot. The window sizes obtained are substituted into equation (1.1) above to obtain accompanying density estimates.

3.2 Statistical Properties of the Proposed Modified Multivariate Cluster Sampling Kernel Density Estimation

1. The estimates of the smoothing parametersare smaller in modified multivariate cluster sampling kernel density estimation approach when compared to the kernel nearest neighbour approach or the Cluster sampling approach. This will contributes significantly to the density estimate by showing more hidden features of the density.
2. The choice of the smoothing parameters at the points of discontinuities follows the step 6 in the algorithm procedure $d_{ST(Opt)}(H) = \frac{(H_i)}{2}$ and generally $d_{ST(Opt)}(H_{i+1}) \le d_{ST(Opt)}(H_i)$ in steps 7 of the proposed algorithm 1. This enables the bandwidth to be controlled such that no new bandwidth would be larger than the preceeding bandwidths in the same co ordinate direction. This ensures that the scheme is adaptive especially at the tails since the tails of any distribution are usually sparse. When this is not the case, each row inverse of the bandwidths matrix is applied.
3 The modified multivariate cluster sampling kernel density estimation utilized the nearest neighbour approach scheme, as well as point 2 above to reduce the problem that could result at the boundaries, particularly when the data are not evenly distributed.

4.0 Application of the Proposed Method

In this section, we apply and compare our method (the modified multivariate cluster sampling kernel density estimation (MMCKDE) with densities of fixed kernel size under the mean squared error criterion, the multivariate cluster sampling kernel density estimation (MCKDE). The error propagation in the proposed MMCKDE with the other approaches listed above would be compared.

Application of the proposed MMCKDE method. We obtained bandwidths and density estimates using Mathematica 6.0 Program. These are given in Table 1 and Table 2 below.

Table 1: Estimated bandwidths for the multivariate cluster sampling kernel density estimation (MCKDE) and the modified multivariate cluster sampling kernel density estimation (MMCKDE) approaches from the adjusted expectation maximization values for data set with missing observation in [28](Pg 310).

Data point	Approaches					
X	Fixed H_{Race}	MCKDE $_{Race}$	MMCKDE $_{Race}$	Fixed H_{Income}	MCKDE $_{Income}$	MMCKDE $_{Income}$
1	0.25	0.25	0.25	5.15	5	5
2	0.25	0.25	0.25	5.15	7.5	7.5
3	0.25	0.25	0.25	5.15	4.5	4.5
4	0.25	0.25	0.25	5.15	13	6.5
5	0.25	0.25	0.25	5.15	5.5	2.75
6	0.25	0.25	0.25	5.15	4.16	3.98
7	0.25	0.25	0.25	5.15	5	5
8	0.25	0.25	0.25	5.15	9	4.5
9	0.25	0.25	0.25	5.15	1	1
10	0.25	0.25	0.25	5.15	5.4	5.4
11	0.25	1	0.5	5.15	5.4	5.4
12	0.25	0.25	0.25	5.15	3.5	3.5
13	0.25	0.25	0.25	5.15	4.5	4.5
14	0.25	0.25	0.25	5.15	5.14	5.14
15	0.25	0.25	0.25	5.15	3.64	3.64

16	0.25	0.25	0.25	5.15	1.71	1.71
17	0.25	1	0.5	5.15	1.71	0.011
18	0.25	0.25	0.25	5.15	4.69	4.69
19	0.25	0.25	0.25	5.15	11.5	5.75
20	0.25	0.25	0.25	5.15	1.06	1.06
Var	**0**	**0.2812**	**0.1875**	**0**	**8.003**	**7.7639**

Table 2: Estimated densities for the multivariate cluster sampling kernel density estimation (MCKDE) and the modified multivariate cluster sampling kernel density estimation (MMCKDE) a approaches from the adjusted expectation maximization values for data set with missing observation in [28](Pg 310).

Data point	Approaches					
X	Fixed H density $_{Race}$	MCKDE $_{Race}$	MMCKDE $_{Race}$	Fixed H density$_{Income}$	MCKDE $_{Income}$	MMCKDE $_{Income}$
1	0.0414	0.0414	0.0414	0.0543	0.051	0.0543
2	0.0414	0.0414	0.0414	0.099	0.0981	0.099
3	0.0414	0.0414	0.0414	0.0674	0.066	0.0674
4	0.0414	0.0414	0.0414	0.0109	0.0109	0.0163
5	0.0414	0.0414	0.0414	0.0348	0.0348	0.037
6	0.0414	0.0414	0.0414	0.0565	0.0565	0.077
7	0.0414	0.0414	0.0414	0.0177	0.0174	0.0174
8	0.0414	0.0414	0.0414	0.0301	0.0331	0.0329
9	0.0414	0.0414	0.0414	0.0042	0.0043	0.0043
10	0.0414	0.0414	0.0414	0.0231	0.0279	0.0279
11	0.0482	0.0488	0.0499	0.0267	0.0312	0.0324
12	0.082	0.082	0.082	0.0431	0.0435	0.0554
13	0.082	0.082	0.082	0.0621	0.063	0.063
14	0.082	0.082	0.082	0.0846	0.0853	0.0867
15	0.082	0.082	0.082	0.0693	0.0695	0.0695
16	0.0414	0.0414	0.0414	0.0414	0.0401	0.0267
17	0.0414	0.0414	0.0418	0.0826	0.0826	0.0826
18	0.0414	0.0414	0.0414	0.0825	0.0825	0.0825
19	0.0414	0.0414	0.0414	0.0341	0.0345	0.0347
20	0.0414	0.0414	0.0414	0.0279	0.0279	0.0257
Density Sum	**0.9972**	**0.9978**	**0.9993**	**0.9523**	**0.9601**	**0.9927**

Density estimates of the data set with missing observation in [28](Pg 310) using MCKDE and MMCKDE methods estimates with the optimal fixed h = 5 approach.

In practise, the smaller the variance of the estimate, the better will its contribution to the overall density estimation, as we do not know the true density $f(x)$ -[7,9,29-31]. We have reduced variances in our proposed approaches. See Table1.

One way of evaluating the method of adaptive window size selection is to compare it to the optimal fixed window size (this is a pilot plot)- [7,4,32-35]. Our new approach behave in quite a similar manner. The other approach is to aim at reducing the AMISE rate in the bandwidth selection method and better convergence rate.

Below are the table of the calculated bandwidth selections errors and convergence rate from the data set with missing observation in [28](Pg 310).

The relative errors, h^* (which is the error in relation to the fixed optimal bandwidth value), $AMISE^*$ and the convergence rates of methods are given below.

Table 3: Table of bandwidth selections errors and convergence rate from the estimated bandwidths for the multivariate cluster sampling kernel density estimation (MCKDE) and the modified multivariate cluster sampling kernel density estimation (MMCKDE) approaches from the adjusted expectation maximization values for data set with missing observation in [28](Pg 310).

Approach	Relative error	h^*	$AMISE^*$	Convergence rate
MCKDE	0.1000	0.1091	1.7552×10^{-3}	0.7411
MMCKDE	0.0080	0.0041	1.5196×10^{-3}	1.9763

Table 3 showed that there are reduced relative errors, h^* (which is the error in relation to the fixed optimal bandwidth value) and $AMISE^*$ in the proposed methods. The proposed methods have faster convergence rates compared to their original versions. That is, the MMCKDE have lower error propagation and faster convergence rates when used to estimates the data in [28](pg310) data with fixed and the MCKDE approaches respectively.

The graphical display of these densities approaches are given in figure 1-3.

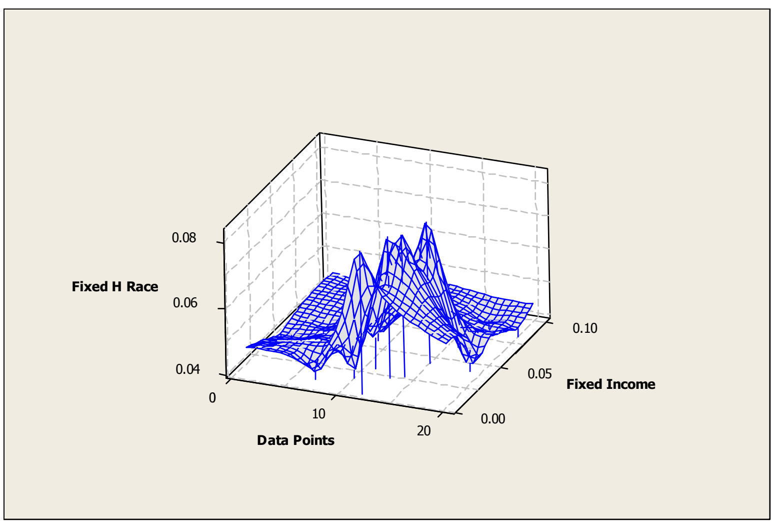

Figure 1: Density for the dataset using the fixed H approach.

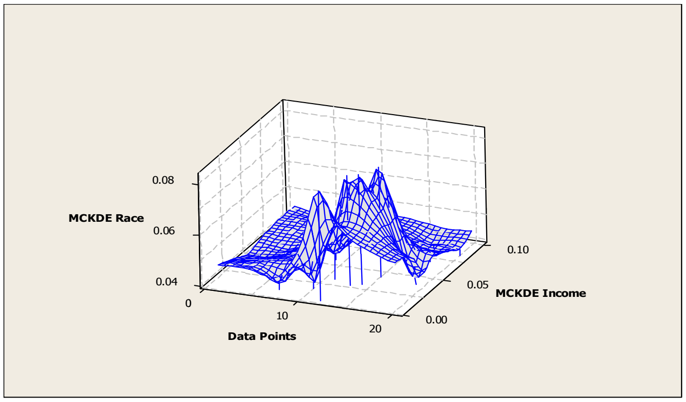

Figure 2: Density for the dataset using the MCKDE approach.

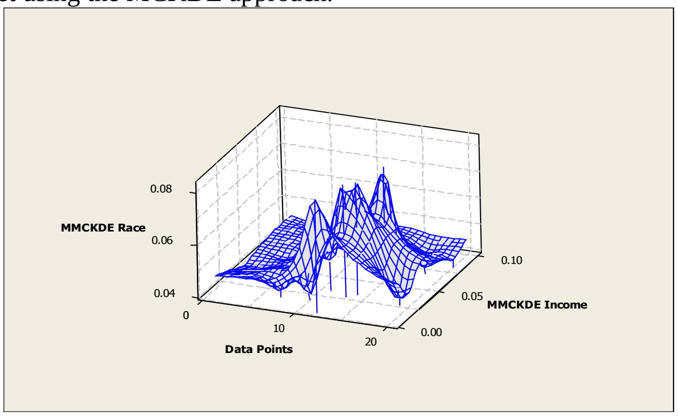

Figure 3: Density for the dataset using the MMCKDE approach.

The various approaches have identifiable differences from Figures 1-3, using the fixed H, MCKDE and MMCKDE for the dataset in Little and Rubin (2002) page 310.

The estimated bandwidth selection errors and convergence rates from the adjusted expectation maximization values for data set with missing observation in [28](Pg 310) data, via the various methods favour the useof the MMCDKE approach over the other approaches. This is because its bandwidth errors are smaller as well as having higher convergence rate. The MMCKDE has some improvement over the MCKDE approach. These can be seen in Table 3. Generally, the AMISE shows the difference between the "true density" and the estimated density. The AMISE for MMCKDE is smaller than that of MCKDE approach.

5.0 Conclusion

We have present a modified multivariate cluster sampling kernel density estimation (MMCKDE) approach. This bandwidth selection approach that is dependent on the data (adaptive). When proper data representations are required, adaptive window sizes should be employed in its density estimation. When the quality of the proposed adaptive density estimates obtained was assessed with some other approaches, we observed some improvements. These are assessed and seen through their AMISE error sizes and rates of convergence. The fixed bandwidth approaches are not adaptive.

6.0 References

[1] Bowman, A.W. and Azzalini, A. (1997): Applied Smoothing Techniques for Data Analysis. Clarendon Press. Oxford.

[2] Katkovnik, V and Shmulevich, I. (2002): Kernel density estimation with adaptive varying window size. Pattern Recognition Letters.Vol. 23, No.14, p.1641-1076.

[3] Wu, D., Tian, Y. and Datt, A. (2008): Analysis of the stochastic interplay between object maintenance and churn. Comput. Comm. 31(2),220-239.

[4] Scott, D. W (1992): Multivariate density estimation. John Wiley, New York.

[5] Osemwenkhae, J. E. (2003): Higher order forms in kernel density estimation. Ph.D thesis, Department of Mathematics, University of Benin, Nigeria.

[6] Abramson I.S. (1982): Arbitrariness of the pilot estimate in adaptive kernel methods. Journal of Multivariate Analysis. Ann. Statist. 12, p.562 -567.

[7] Silverman, B.W. (1986): Density Estimation for Statistics and Data Analysis. Chapman and Hall, London.
[8] Bhattachdrya, P.K. and Gangopadhya, A.K. (1990): Kernel and nearest neighbourhood estimation of conditional Quartile. Ann. of Stat, Vol.18,No.3, p.1400-1415.
[9] Wand, M. P. and Jones, M. C. (1995): Kernel smoothing Chapman and Hall/CRC/London.
[10] Wu, T.J, Chen, C.F and Chen, H.Y, (2006): A variable bandwidths selectors in multivariate kernel density estimation. Stat. and Prob. Letters, 77,462-467.
[11] Horova, I., Kolacek, J., Zelinka, J. and Vopatova, K. (2008): Bandwidth choice for kernel density estimates. IASC, Yokohama, Japan.
[12] Loftsgaarden, D. and Quesenberry, C. (1965): A non parametric estimate of a multivariate density function. Ann. Math. Stat 36, p.1049 -1051.
[13] Wu, T.J and Tsai, M.H, (2004): Root *n* bandwidths selectors in multivariate kernel density estimation. Probab. Theory Related Fields, 129,537-558.
[14] Horova, I., and Zelinka, J.(2007): Contributions to bandwidth choice for kernel density estimates.Comput. Statis. 22,31-47.
[15] Salgado-Ugarte, I.H, and Perez-Hernandez, M.A. (2003): Exploring the use of variable bandwidth kernel density estimators. Stata journal 3 (2).
[16] Salgado-Ugarte, I.H. ,Shimuzu, M. and Taniuchi, T. (1993): snp6; exploring the shape of univariate data using kernel density estimators. Stata Technical bulletin 16, p.8-19.
[17] Salgado-Ugarte, I.H. ,Shimuzu, M. and Taniuchi, T. (1995): snp6.2; Practical rule for bandwidth selection in univariate density estimation. Stata bulletin 27, p.5-19.
[18] Phillippe Van K. (2003): Adaptive kernel density estimation. 9^{th} UK stata users meeting, Royal statistical society, London. May 19-20.
[19] Katkovnik, V. (1999): A new method for varying adaptive band width selection. IEEE Transaction on Signal process. Vol.47, No.9, p.2567 – 2571.
[20] Taylor, C. C. (1989): Bootstrap choice of the smooth parameter in kernel density estimation. Biometrika 76, p.705 – 712.
[21] Efron, B. and Tibshirani, R. J. (1998): Introduction to the Bootstrap. Chapman and Hall, Florida, USA. p. 258 -269.
[22] Comaniciu, D and Meer, P. (2002): Mean Shift: A robust approach towards Feature space Analysis. IEEE Transactions on Pattern Analysis and Mechine Intelligence, V. 24, n. 5, p. 603-619.
[23] Jones, M.C. (1990): Variable kernel density estimates and variable kernel density estimators. Austral J. Statis. 32.p.361-371
[24] Wu, K.L., Shan, K., Yung, K. Miin-Shen, Y. and Yuan, C. (2007): Mean Shift-based clustering to KDE. Journal of Pattern Recognition. Vol 40. Issue 11, 3035-3052.
[25] Duong, T. and Hazelton,M.L.(2005a): Cross-validation bandwidth matrices for multivariate kernel density estimation. Scand. J. Statist.32, 485-507.
[26] Duong, T. and Hazelton ,M.L.(2005b): Convergence rates for unconditional bandwidth matrix selector for multivariate kernel density estimation. J. Multivariate Anal. 93, 417-433.
[27] Gray, A. (1997): The intuitive idea of distance on surfaces in ‘Modern differential geometry of curves and surfaces with Mathematica’. 2^{nd} edition. Boca Ration, FL, CRC press. 341-345
[28] Little R.J.A. and Rubin, D.B.(2002): Statistical analysis with missing data. Second edition. Wiley and Sons Publisher. New Jersey. USA.
[29] Osemwenkhae, J. E. and **Ogbeide, E.M**. (2010a): Asymptotic mean square error for the adaptive kernel density estimation. Adv. Nat. and Appl. Sci. Res.Vol. 8. p 253-258.
[30] Osemwenkhae, J. E. and **Ogbeide, E.M.** (2010b): Adaptive Kernel Density Estimation, A review. Nigerian Annals of Natural Sciences.Vol. 10,No.1, p.88-96.
[31] Osemwenkhae, J. E and Oyegue, F.O. (2004): Optimal window width selection for higher order symmetric kernels. ABACUS (Journal of Mathematical Association of Nigeria). Vol. 32, p.130-139.
[32] **Ogbeide, E.M**., Osemwenkhae, J. E. and Oyegue F.O. (2011):A new adaptive kernel density estimation bandwidth approach. Journal of Engineering Science and Application (JESA), Vol. 7. No3.p 14 – 25.
[33] Breiman, L., Meisel, W., and Purcell, E. (1977): Variable kernel estimates of multivariate density. Technometrics, 19, p.135-144.
[34] Ogbonmwan, S.M. and Osemwenkhae, J. E. (2000): Higher order forms for optimal window width in kernel density estimation. Journal of Nig. Assoc. of Maths. Physics. 4, p 327-334.
[35] Cao, R., Cuevas, A. and Manteiga, W.G. (1994): A comparative study of several smoothing methods in density estimation. Computational Statistics and Data Analysis. 17: 153-176