Journal of the Nigerian Association of Mathematical Physics
Volume **22 (November, 2012), pp 185 – 194**

A Modified Kernel Nearest Neighbour Estimate Bandwidth Approach To Density Estimation

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

[1]Department of Mathematics and Statistics,
Ambrose Alli University, Nigeria
[2]Department of Mathematics,
University of Benin, Benin City, Nigeria
[3]Department of Mathematics and Computer,
Federal University of Petroleum Resources, Effurun, Nigeria

Abstract

This research focuses on the empirical illustration of the proposed modified kernel nearest neighbourhood estimate (MKNNE) bandwidth approach in estimating density. This proposed approach is used to estimate the Ambrose Alli University GST 102 Mathematics students' results. The quality of the adaptive density estimates obtained showed some improvements over some existing schemes. This is visible in the reduced error rates and better rate of convergence.

Keywords: Adaptive approaches, bandwidth, error, kernel density estimation.

1.0 Introduction

Kernel density estimation provides a nonparametric estimate of the probability function from which a set of data is drawn. As these densities are usually unknown, unrealistic assumptions are frequently made, thus compromising the performance of the algorithm in question [1]. If a particular form of the density is assumed known, then a parametric estimation is used. If nothing is assumed about the density shape, nonparametric estimation is employed. One of the most well-known and popular techniques of nonparametric density estimation is the kernel or parzen density estimation [2].

Given *a* sample $X_1, X_2, ..., X_n$ drawn from population with density function *f(x)*, the univarite kernel density estimator evaluated at x is given by;

$$\hat{f}_n(x) = \frac{1}{nh}\sum_{i=1}^{n} k\left(\frac{x - X_i}{h}\right) \tag{1.1}$$

where k(.) is a kernel function and *h* is the bandwidth or the smoothing parameter.

Earlier work on the kernel method emphasized asymptotic results, whereas determining an optimal *h* is the main research focus today. A number of other works considering the problem of kernel size selection exist [3 – 5]. It has been shown by many researchers [1, 5 – 8] that estimates based on variable-sized kernel and optimal constant- sized kernel are widely used. However, the performance of the kernel methods depends largely on the smoothing parameter (bandwidth) and very little on the kernel [4, 9, 10].

Recently, a wide variety of sophistication of the basic kernel estimator has been proposed, all pointing to the importance of adaptive kernel estimator [5, 11]. The "adaptive" nature of the density estimate arises from the varying bandwidth used in the estimation process. If *h*, the bandwidth in (1.1) above, is "fixed" during estimation, we have the fixed kernel density estimation approach, but when it is allowed to vary all though the process of the estimation, we have the adaptive kernel method. A number of works considering the problem of kernel size selection exist [1, 3, 4, 6, 12–15]. The motivation for this

Corresponding author: ***Ogbeide E. M.,*** E-mail: -, Tel. +234 8035910189

work arose from the work of Bhattachdrya and Gangopadhya [14] on the kernel with nearest neighbour estimates (KNNE) approach.

Some great pitfalls of the nearest neighbour estimate density approach is that estimates tends to oversmooth data density. Also, it is not a bonafide probability estimates. It has some points of discontinuities. It does not integrate to one. An improvement on it came from the work of Bhattachdrya and Gangopadhya [14]. In their work, the kernel nearest neighbourhood density estimates, a kernel function (which is a probability density function) when added to nearest neighbourhood density estimates improve smoothing of the density of the data. The degree of smoothing is controlled by the distance between two points on the line to be $|x_i - x_{i+1}|$ in which $d_1(x) \le d_2(x) \le ..... \le d_n(x)$ are the distances arranged in ascending order, from *x* to the points of sample. This result is consistent with the intuitive idea of kernel estimator having to find a compromise between estimating two distinct values of *f* on either side of discontinuity. The bandwidths obtained are substituted into equation (1.1) above to obtain density estimate. This approach is a bonafide probability estimates as the estimator integrates to one, but there still some points of discontinuities in the density curve.

Clearly, in practice, one does not have access to the true density function *f(x)* which is proposed to be estimated. Thus, a number of approaches can be taken for finding the bandwidth that will lead to better density estimation via varying the bandwidths [4, 5, 6, 15-17]. To this end, we modified the kernel nearest neighbour estimate method and term it the modified kernel nearest neighbour estimate (MKNNE) approach in estimating densities. The quality of the density estimates are assessed by comparing it to the density, obtained under the mean-square error criterion. The error generated using these approach would be considered.

In this work, we present a novel data-driven method that require the knowledge of pilot plot from optimal fixed window size and the variance of the estimate. This invariably reduces the amount of error at arriving at the "true density". This is an adaptive kernel approach which adapts to the sparseness of the data by using a broader bandwidths over observations located in region of low densities. This is done by varying the bandwidth inversely with the density. An initial (fixed bandwidth) density estimate is computed to get an idea of the density at each of the data points, and this pilot estimate is used to adapt the size of the bandwidth over the data points when computing a new kernel density estimate.

2.0 The Modified Kernel Nearest Neighbour Estimate (MKNNE) Approach

This method of adaptive kernel size selection proposed in this work is based on the kernel nearest neighbour estimate (KNNE). It is a modification of kernel nearest neighbour density estimate by adjusting the amount of bandwidths to the density of the data. The degree of smoothing is controlled by the distance between two points on the line to be $|x_i - x_{i+1}|$ in which $d_1(x) \le d_2(x) \le ..... \le d_n(x)$ are the distances arranged in ascending order, from *x* to the points of sample. To correct the problem of discontinuities in some points in the kernel nearest neighbour density estimate, we have to identify the discontinuous points in the estimation using the kernel nearest neighbour approach as a pilot guide. We use the ideal of density at the boundaries. In this case, we supposed that f is a density such that $f(x) = 0$ for $x < 0$ and $f(x) > 0$ for $x \ge 0$. We further supposed 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 $0 \le \alpha < 1$. See, Wand and Jones (1995). Then at the boundary we obtain

$$E\,\hat{f}(0;h) = \tfrac{1}{2} f(0) + 0(h) \qquad (2.1)$$

This result is consistent with the intuitive idea of kernel estimator having to find a compromise between estimating two distinct values of *f* on either side of discontinuity. We proposed that the inter-quartile range of the boundary values will be used. Since the location of the boundary of $\hat{f}(x;h)$ is usually known, we adopted this to achieve better performance in its vicinity. The window sizes obtained are substituted into equation (1.1) to obtain the corresponding density estimates. The MKNNE procedure can be implemented with the aid of the following Algorithm.

Algorithm : *The algorithm of the proposed modified kernel nearest neighbour approach*

Step 1 `Define y = knn (x,n,t)`

x *as individual observation or random samples,* *output the* n^{th} *nearest neighbour from* t *(scalar)*

Step 2 `Sortx= Sx and individual data =ind`

Step 3 `lx = length (x);`

Step 4 `[sx,ind] = sort (x);`

Step 5 `sort x`

Step 6 `if (t < min (x));then y = sx (n)`

Step 7 `if (t > max (x));then y = sx (lx-n+1)`

Step 8 `for j = 1:n-1,` *do* n^{th} *nn from* $t = 1(1)n$

Step 9 *do* $nn-1$ `=lnn and` *do* $nn+1$ `= Rnn`

Step 10 `while y = sx (lnn, rnn)` > 0 `, continue`

Step 11 `if abs (sx (lnn-1) - t) < abs (sx (rnn+1) -t) continue`

Step 12 `else` $t = 1(1)n-2$

Step 13 `if abs (sx (lnn-1) - t)-abs (sx (rnn+1) - t) <0;` $d_k(t)$

Step 14 `else` *do* $d_k(t) = \dfrac{d_k(t)}{2}$

Step 15 `continue`

Step16 *if* $d_{(k)Opt}(t) \le d_{(k-1)Opt}(t)$ *continue*, *else stop*

Step 17 `write` $d_k(t)$

Step 18 `end while` $j = n$.

3.0 Statistical Properties of the Proposed Modified Kernel Nearest Neighbour Estimates Approach

1. The estimates of the smoothing parameters are smaller in modified kernel nearest neighbour approach scheme when compared to the kernel nearest neighbour approach. This will contributes significantly to the density estimate by showing more hidden features of the density.
2. The choice of the smoothing parameters follows the procedure at the points of discontinuities $d_{(k)Opt}(t) = \dfrac{d_k(t)}{2}$ and $d_{(k)opt}(t) \le d_{(k-1)opt}(t)$ in steps 14 and 16 of the proposed algorithm above. This enables the bandwidth to be controlled such that no new bandwidth would be larger than the preceding bandwidths. This ensures that the scheme is adaptive, since the tails of any distribution are usually sparse.
3 The modified kernel nearest neighbour approach scheme reduces the problem that could result at the boundaries, particularly when the data are not evenly distributed.

4.0 Application

We obtained bandwidths and density estimates using a program in Mathematica 6.0 (see Appendix 2). These are given in Table1 and Table 2.

In Table 1, the optimal fixed bandwidth size of h=5 was found for AAU GST 102 data for every x. The results presented in Table 1 are the bandwidths (smoothing parameters). As expected, (see Appendix I) the bandwidths are larger in regions with few data (see Table 1) compared to the regions with more data, where the bandwidths become smaller in the propose method. This has shown that the propose MKNNE method is sensitive to the data distribution.

The density estimates from the kernel nearest neighbour bandwidths and the modified kernel nearest neighbour bandwidths approaches in Table 1 substituted into (1.1) are presented in Table 2 for AAU GST 102 data. The kernel nearest neighbour estimates approach have lower densities at both ends of the estimates, compared to using the modified kernel nearest neighbour estimates. This leads to heavy tails distributions (which can affects type I and type II errors). A look at the graphical displays of these densities in Figure 1 revealed this.

Table 1: Estimated bandwidths of AAU Students' GST 102 examination result data using kernel nearest neighbour estimates (KNNE) and the Modified Kernel Nearest Neighbour Estimate (MKNNE) with the optimal fixed kernel method approach h=5.

Data points	Bandwidths approach		
X	optimal fixed h=5	kNNE	MkNNE
0	5	5	5
5	5	5	5
10	5	5	4.8
15	5	5	4.62
20	5	4.93	3.91
25	5	4.16	3.98
30	5	3.93	3.93
35	5	3.05	2.5
40	5	2.21	1.25
45	5	3.7	2.5
50	5	3.9	2.97
55	5	4.79	4.48
60	5	4.1	4.34
65	5	4.27	2.5
70	5	4.34	3.96
75	5	5	4.14
80	5	5	4.78
85	5	5	4.93
90	5	5	5
95	5	5	5
100	5	5	5
Mean		4.4319	4.0428
Var		1.91	1.8803

Table 2: Density estimates Values for AAU Students' GST 102 examination result data using the kernel nearest neighbour estimates (KNNE) bandwidths and the Modified Kernel Nearest Neighbour Estimate (MKNNE) bandwidths with the optimal fixed kernel method approach *h*=5.

Data point	Density estimates approach		
X	optimal fixed h=5	*MkNNE*	*kNNE*
0	0	0	0
5	0	0	0
10	0	0.006	0.003
15	0	0.013	0.007
20	0	0.019	0.01
25	0.008	0.028	0.015
30	0.0017	0.057	0.048
35	0.0446	0.064	0.064
40	0.1429	0.144	0.144
45	0.125	0.19	0.17
50	0.1964	0.231	0.22
55	0.125	0.192	0.181
60	0.1607	0.1622	0.1598
65	0.0893	0.0893	0.0748
70	0.0446	0.0611	0.0487
75	0.0357	0.0391	0.0321
80	0.0008	0.0098	0.0087
85	0	0.0008	0.003
90	0	0	0
95	0	0	0
100	0	0	0

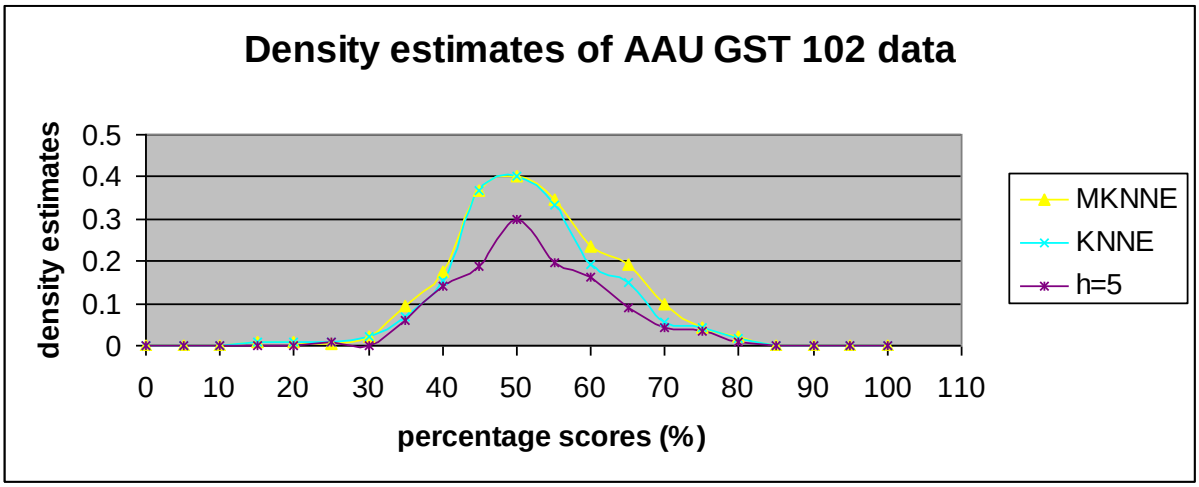

Figure 1: Density estimates of AAU GST 102 data using KNNE and MKNNE methods estimates with fixed *h*=5 approach.

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 $f(x)$ [4 – 6]. We have reduced variance in our new approach.

According to various authors [1, 5, 6, 17 – 19] one way of evaluating the method of adaptive bandwidths selection is to compare it to the optimal fixed bandwidth (this is a pilot plot). Our new approach behaved in quite a similar manner. The other approach is to aim at reduced asymptotic mean integrated squared error (AMISE) rate in the bandwidth selection method and better (faster) convergence rate. $AMISE$ shows the difference between the "true density" and the estimated density.

Table 3 is the table of calculated bandwidth selections errors and convergence rates from the AAU GST 102 data.

Table 3: Table of bandwidth selections scheme errors and convergence rate from the AAU GST 102 data. h^* = bandwidths error rate in relation to the optimal bandwidth value.

Approach	Relative error	h^*	$AMISE^*$	Convergence rate
KNNE	0.7726	0.7879	1.659×10^{-2}	0.0226
MKNNE	0.63352	0.7577	1.630×10^{-2}	0.0503

The simulated bandwidth selection scheme errors and convergence rates from the AAU GST 102 data and eruption data, via the two methods favour the use of the MKNNE approach over the KNNE. This is because MKNNE bandwidth selections scheme errors are smaller and it have higher convergence rate. These can be seen in Table 3.

4.0 Advantages of the Method Proposed

1. The MKNNE method has lower variance in its estimates. It has relative lower errors in the bandwidth selection and better convergence rate than it original version.
2. It should be mentioned that the optimal constant bandwidth density estimate, to which we were comparing our method, can never be achieved in practice as the true density is unknown. At the same time, the proposed method, making no assumptions about the underlying density results is comparatively better. It is sensitive to the data distribution.
3. We have seen from estimated density, that the proposed method can perform comparatively with any constant bandwidth method. The fixed bandwidth methods do undersmooth or oversmooth, but the adaptive approaches are more data sensitive.
4. The MKNNE approaches produce smaller but optimal smoothing parameters. The estimates of the smoothing parameters h_i are smaller in MKNNE than in KNNE. These contribute significantly to the density estimate by showing more hidden features of the density. Like the behaviour and attributes of the data. This includes whether the data is symmetric, skew symmetric, normal, bi-modal, population distribution etc.

5.0 Conclusion

We have proposed a method for varying bandwidths in kernel density estimation. This method is based on the data and a pilot estimate. The MKNNE approach modifies KNNE procedure. The kernel nearest neighbour approach uses the kernel function on the nearest neighbour method. When we compare our new method with the other approaches' distributions in kernel density estimation, their behaviour has some little differences. The fixed approach is the pilot estimates. It is not by any means the best estimate. As expected, the bandwidths are larger in region with few data compared to the region with more data, where the bandwidths become smaller in the new method. The MKNNE approach has a lower AMISE and a faster rate of convergence than it original version. This has shown that the proposed method is more sensitive to the data distribution. This performs significantly better than any constant- bandwidth method.

APPENDIX 1

Table 4 Source: School of General Studies (2008/2009): General Studies results for Department of Mathematics Students (GST 102). Ambrose Alli University, Ekpoma, Nigeria.

Data of 112 GST 102 examination in Ambrose Alli University, Ekpoma, Nigeria

40	35	42	46	55	54	49	58	68	46
50	58	57	47	48	43	49	45	70	78
50	50	43	65	54	67	43	67	49	26
76	57	42	51	53	42	65	64	48	68
32	41	49	55	45	50	58	52	53	52
58	44	53	49	50	37	60	40	40	48
51	48	42	60	40	48	60	65	64	36
63	59	67	55	56	49	59	77	58	60
74	43	43	47	69	53	70	54	35	40
46	48	49	51	41	51	58	57	45	47
61	42	40	48	53	40	48	53	35	51
60	40								

APPENDIX 2

```
function y = knn (x,k,t)

% SYNTAX: function y = knn (x,k,t)
% This function takes as input a vector of random samples x
% and outputs the k-th nearest neighbor from point t (scalar)
% parameter k is for the k-th NN
lx = length (x);
[sx,ind] = sort (x);                                        % sort x

if (t < min (x))
   y = sx (k);
   return;
end
if (t > max (x))
   y = sx (lx-k+1);
   return;
end
nnind = find (sx == x (interp1 (sx,ind,t,' nearest')));     % index
of 1 st NN (in x)

if k == 1                                                   % if we
want only 1 st NN
   y = sx (nnind);
   return;
end

lnn = nnind;                                                % set
left NN pointer
```

```
rnn = nnind;                                                    % and
right NN pointer

for j = 1:k-1,                                                  % find
the k-th NN
   if (lnn == 1)
      rnn = rnn + 1;                                            % move
right pointer right
      y = sx (rnn);
   elseif (rnn == lx)
      lnn = lnn - 1;                                            % move
left pointer left
      y = sx (lnn);
   else
      if abs (sx (lnn-1) - t) < abs (sx (rnn+1) - t)
         lnn = lnn - 1;                                         % move
left pointer left
         y = sx (lnn);
      else
         rnn = rnn + 1;                                         % move
right pointer right
         y = sx (rnn);
      end
   end
end function y = knn (x,k,t)

% SYNTAX: function y = knn (x,k,t)
% This function takes as input a vector of random samples x
% and outputs the k-th nearest neighbor from point t (scalar)
% parameter k is for the k-th NN

lx = length (x);
[sx,ind] = sort (x);                                            % sort x

if (t < min (x))
   y = sx (k);
   return;
end

if (t > max (x))
   y = sx (lx-k+1);
   return;
end

nnind = find (sx == x (interp1 (sx,ind,t,' nearest')));         % index
```



```
of 1 st NN (in x)
```

```
if k == 1                                                    % if we
want only 1 st NN
   y = sx (nnind);
   return;
end

lnn = nnind;                                                 % set
left NN pointer
rnn = nnind;                                                 % and
right NN pointer

for j = 1:k-1,                                               % find
the k-th NN
   if (lnn == 1)
      rnn = rnn + 1;                                         % move
right pointer right
      y = sx (rnn);
   elseif (rnn == lx)
      lnn = lnn - 1;                                         % move
left pointer left
      y = sx (lnn);
   else
      if abs (sx (lnn-1) - t) < abs (sx (rnn+1) - t)
         lnn = lnn - 1;                                      % move
left pointer left
         y = sx (lnn);
      else
         rnn = rnn + 1;                                      % move
right pointer right
         y = sx (rnn);
      end
   end
end.
```

References

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

[2] Rosenblatt, M. (1956): Remarks on some parametric estimates of a density function. Springer-Verlag, Berlin. 181-190.

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

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

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

[6] Wand, M. P. and Jones, M. C. (1995): Kernel smoothing Chapman and Hall/CRC/London.

[7] Goldenshluger A. and Nemirovsky A. (1997): On spatial adaptive estimation of Non parametric Regression. Math. Methods of stat. Vol. 6, No. 2, 135 – 170.

[8] Omar, M.E. (2004): Some improvements in kernel estimation using Line Transect sampling. Journal Modern Applied Statistical Method. Vol.3, No.1,149-157.

[9] 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, 327-334.
[10] Osemwenkhae, J. E. (2003): Higher order forms in kernel density estimation. Ph.D thesis, Department of Mathematics, University of Benin, Nigeria.
[11] Salgado-Ugarte, I.H, and Perez-Hernandez, M.A. (2003): Exploring the use of variable bandwidth kernel density estimators. Stata journal 3 (2).
[12] Breiman, L., Meisel, W., and Purcell, E. (1977): Variable kernel estimates of multivariate density. Technometrics, 19, 135-144.
[13] Simonoff, J.S. (1996): Smoothing methods in Statistics. Springer-Verlag, New York.
[14] Bhattachdrya P.K. and Gangopadhya A.K. (1990): Kernel and nearest neighbourhood estimation of conditional Quartile. Ann. of Stat, Vol.18,No.3, 1400-1415.
[15] Hardle, W. (1990): Smoothing Technique in implementation in Springer –Verlag. New York.
[16] Izenman, A. J. (1991): Recent developments in Non parametric density estimation. Journal of the American statistical Association: Vol. 36. No 413.
[17] Osemwenkhae, J. E. and Ogbeide, E.M. (2010): Adaptive Kernel Density Estimation, A review. Nigerian Annals of Natural Sciences.Vol. 10,No.1, 88-96.
[18] Bowman, A.W. and Azzalini, A. (1997): Applied Smoothing Techniques for Data Analysis. Clarendon Press. Oxford.
[19] Jones, M. C. (1990): Variable kernel density estimates and variable kernel density estimators. Austral J. Statis. 32, 361 -71.
[20] School of General Studies (2009): General Studies results. General Studies Division, Ambrose Alli University, Ekpoma, Nigeria.
[21] Cencov, N. N.(1962): Evaluation of an unknown distribution from observations. Soviet Math., 3. 1559-62.
[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] Hall, P. (1990): On the global properties of variable band width density estimator. Ann. Statist. 20, 762 -78.