Journal of the Nigerian Association of Mathematical Physics,
***Volume 8* (*November 2004*)**

Higher Order Bootstrap likelihood

S. M. Ogbonmwan
Department of Mathematics, University of Benin
Benin City, Nigeria

Abstract

In this work, higher order optimal window width is used to generate bootstrap kernel density likelihood. A simulated study is conducted to compare the distributions of the higher order bootstrap likelihoods with the exact (empirical) bootstrap likelihood. Our results indicate that the optimal window width of orders 2 and 4 perform better than those of higher orders. The higher order kernels (≥ 6***) provided window widths, which obscured the details of the distribution when the exact bootstrap likelihood was taken to be the true density.***

Keywords: Higher order kernels, exact bootstrap empirical likelihood, Bootstrap kernel likelihood, optimal window width.

pp 87 - 92

1.0 Introduction

A variety of kernel functions have been proposed. Some references include [3], [14], [17], [18] and [19]. The choice of the optimal window width in kernel density estimation has been crucial. The optimal window width determines the form of the distribution since it controls the degree of smoothing applied to a data set. Of importance is the kernel order, which can have a major impact on finite sample Mean Integrated Square Error (MISE) even in small samples. Hansen in [10] proposed selecting the kernel order by the criterion of minimize regret, where the regret is maximized over a set of Marron–Wand in [9] test density function. The work of [6] among others has drawn attention to higher order kernels. Their work placed emphasis on the fourth order kernels for the Gaussian, skewed unimodal, outlier and the separated bimodal densities. However, [9] showed earlier that the fourth order kernels are of little use for small samples. The concept of higher orders in density estimation is being proposed with a view of enhancing the performance of the estimate of the density (see [2003]). Thus, the justification for higher order optimal window width as used in this work stems from the fact that the rate of convergence of the density estimator to the actual density depends on the amount of smoothness typically quantified in terms of the number of bounded derivatives of the underlying density. For a random sample $X_1, X_2, X_3, \ldots, X_n$ from a random variable X with unknown density function $f(x)$, let us consider the function

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

where $k(x)$ is a non-negative function which normalizes to unity. The function $\hat{f}(x)$ is a kernel density estimator (of the unknown density $f(x)$) with $k(x)$ as the kernel function. The scaled version of the estimate $\hat{f}(x)$ in (1.1) is of the form

$$\hat{f}(x) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h}k\left(\frac{x - X_i}{h}\right) \quad (1.2)$$

where h is the smoothing parameter. Equation (1.2) is generally referred to as the kernel density function. Small values of h in (1.2) lead to spurious noise at the tails of the distribution while large values of h obscures the details of the tails of the distribution (see [4], [19]). A balanced choice of the smoothing parameter h in (1.2) would produce the best kernel density estimator $\hat{f}$. The choice of h has bothered many researchers and several optimal mathematical expressions for h can be found in the literature. (see

[6], [9], [10], and [12].). Mathematically, a global accuracy of the estimate $\hat{f}(x)$ is evaluated via the Mean integrated square error (MISE), which is expressed, as

$$MISE\left(\hat{f}(x)\right) = \mathrm{E}\left[\int\left(\hat{f}(x) - f(x)\right)^2 dx\right] = \int Bias^2(x)\,dx + \int Var\left(\hat{f}(x)\right)dx \qquad (1.3)$$

A mathematical derivation of (1.3) would show that the integral bias and the integral variance are proportional to h^2 and $(nh)^{-1}$ respectively. Hence, a reduction in the variance produces an increase of the bias, while a smaller h immediately reduces the bias but increases the variance. Methods of dealing with this imbalance in

the density of f led to other density estimators such as adaptive kernel estimators of the form:

$$\hat{f}(x) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{\lambda_i h}k\left(\frac{x - \mathrm{X}_i}{h}\right) \qquad (1.4)$$

where λ_i are quantities related to the local density at X_i. Our approach of dealing with the optimal choice of the smoothing parameter h is not through the adaptive kernel method. Rather, we seek to work through higher order kernels, get an appropriate higher order optimal window width, which balances the interplay between the bias and the variance terms of the MISE. Thus through empirical results of an example of a bootstrap generated likelihood we seek:

i. to provide higher order bootstrap generated likelihood.
ii. to examine the similarities between the various higher order optimal window width (indeed higher order kernels).
iii. to make comparisons between the higher order kernels using the results for the various optimal window widths.
iv. make a recommendation of the "best" higher order optimal window width. Best here means in terms of fit.

The rest of the paper is organized as follows: Section 2 deals with a brief review of higher order kernels, while section 3 specifically introduces higher order forms of optimal window width (i.e. smoothing parameter) for bootstrap generated likelihood. Through theoretical and simulation comparisons, higher order kernels are examined in this work. Generally, the works of [6], [8], [9], have given a lead in this direction. Results of simulation are given in section 4 and in section 5; we give our recommendations and conclusion.

2.0 Higher order kernels

When the following conditions: (i). $\int_{-\infty}^{\infty} k(t)\,dt = 1$ (ii) $\int_{-\infty}^{\infty} k(t)\,dt = \int t^2 k(t)dt = \Lambda = \int_{-\infty}^{\infty} t^{m-1}k(t)dt = 0$ (iii) $\int_{-\infty}^{\infty} t^m k(t)\,dt = V_m \neq 0$ were imposed on symmetric kernels, [12] obtained the optimal window width as:

$$h_m \approx \left\{\frac{\left((m)!\right)^2}{2m}\right\}^{\frac{1}{2m+1}} V_m^{-\frac{2}{2m+1}}\left\{k(t)^2\,dt\right\}^{\frac{1}{2m+1}}\left\{\int_{-\infty}^{\infty} f^{(m)}(x)^2\,dx\right\}^{-\frac{1}{2m+1}} .\ n^{-\frac{1}{2m+1}} \qquad (2.1)$$

for m even (i.e. for m = 2, 4, 6, 8, ….). Equation (2.1) is equivalent to:

$$h_{2m} = \left\{\frac{\left((2m)!\right)^2}{4m}\right\}^{\frac{1}{4m+1}} V_{2m}^{-\frac{2}{4m+1}}\left\{\int k(t)^2\,dt\right\}^{\frac{1}{4m+1}}\left\{\int f^{(2m)}(x)^2\,dx\right\}^{-\frac{1}{4m+1}} n^{-\frac{1}{4m+1}} \qquad (2.2)$$

for m = 1, 2, 3, …., $< \infty$ obtained earlier in [13] with milder conditions. Jones and Signorini in [6] used similar conditions to obtain the fourth order kernel of the form:

$$K_{(4)}^{p}(u) = \frac{\left(S_4 - S_2 u^2\right)K(u)}{\left(S_4 - S_2^2\right)} \qquad (2.3)$$

where $S_j = \int_{-\infty}^{\infty} u^j k(u)\,du$, and $k(u)$ is a kernel function. Specifically, if f is the normal distribution i.e

$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \tag{2.4}$$

and k is the Gaussian kernel i.e $k(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}$ then by Taylor series expansion, [12] obtained the optimal values for any order of h_m to be:

$$h_{m,\,opt}\approx\left\{\frac{(m!)^2}{2m}\cdot\frac{\pi^{3/2}}{2^m\Gamma\left(\frac{2m+1}{2}\right)\left(\Gamma\left(\frac{m+1}{2}\right)\right)^2}\right\}^{\frac{1}{2m+1}}.\ \sigma.n^{-\frac{1}{2m+1}} \tag{2.5}$$

We now apply (2.5) in generating higher order form of the bootstrap kernel likelihood, and examine the results to be obtained for each order m = 2, 4, 6, 8,… . This is done in sections 3 and 4. The practical problem posed by the expressions in (2.1) and (2.2) is that the ideal optimal window width of any order depends on the unknown density *f*. In order to provide a practical solution to the problem of kernel order selection, we adopt a bootstrap, approach

to generate likelihoods whose densities are compared with those of the "true" density. Our approach can easily be implemented in practice.

3.0 Higher order Bootstrap likelihood

Suppose the transformed data set $y_\theta=(y_1,y_2,...,y_n)_\theta$ based on a realized data set x_1, x_2, …,x_n are *iid* with cdf F, not depending on θ. Suppose the statistic of interest estimated on the transformed data set is $T_\theta=t(y_\theta)$. This could be any contrast of parameters. Bootstrap the values of y_θ and for each bootstrap configuration, calculate $T^*_\theta=t(y^*_\theta)$ and list them in some order, $T^*_1,T^*_2,...,T^*_B$ then, the empirical cdf will be of the form:

$$P_T(t\mid\theta)=\frac{1}{B}\#\{T^*_i\le t\} \tag{3.1}$$

where # stands for cardinality. The empirical cdf in (3.1) is smoothed by using the kernel density estimator to obtain a continuous density $\hat{f}_T(t\mid\theta)$ (which is the bootstrap likelihood). Simply put, the bootstrap kernel likelihood is the density of the bootstrap value T^*_θ at t_θ, where $T^*_\theta=t(y^*_\theta)$ sampled randomly from the transformed data set $(y^*_1,y^*_2,....,y^*_n)\theta$. Thus, at the point where the cdf in (3.1) is smoothed (by the kernel density estimator), we get the kernel bootstrap likelihood to be:

$$L^*(\theta)=\hat{f}_{T^*_\theta}(t)=\frac{1}{Bh}\sum_{i=1}^{B}K\left(\frac{t-T^*_i}{h}\right) \tag{3.2}$$

where $T^*_i=T_i(y^*_\theta)$, i = 1(1)B, are the bootstrap values of any contrast of parameters, $y^*_\theta=(y^*_1,y^*_2,y^*_3....,y^*_n)_\theta$ are the bootstrap values of the transformed data set indexed by θ. t is the actual value of the contrast, B is the number of bootstrap configurations, while K(**.**) and h have their usual meanings. The m^{th} (m even) order bootstrap kernel likelihood is hence given as:

$$L^*_m(\theta)=\frac{1}{B\,h_{m,opt}}\sum_{i=1}^{B}K\left(\frac{t-T^*_i}{h_{m,opt}}\right) \tag{3.3}$$

where $h_{m,opt}$ is as defined in equation (2.5). By adopting equation (3.3), one can now generate higher order bootstrap kernel likelihood. This idea is used in section 4 to generate the densities of bootstrap likelihoods for a simple contrast of difference of means using higher order window widths. The exact bootstrap likelihood is rather simple to generate, especially with the advent of high-speed computers. The exact bootstrap likelihood is generated by simply doing a count of the proportion of T_j values, which lie within some intervals around j. For some ε > 0, the exact empirical bootstrap likelihood or simply the bootstrap likelihood is defined as:

$$L(\theta)=\frac{1}{B}\left(\mathrm{T}_j^*\,\middle|\,\tilde{t}-\varepsilon\le \mathrm{T}_j^*\le \tilde{t}+\varepsilon\right) \qquad (3.4)$$

We remark that the value of ε is also very crucial and care must be exercised in order to make a good subjective choice.

4.0 Results of simulation

Our example consists of a simple two-sample problem for which the contrast of interest is the difference of mean. Each of the two samples consists of 12 observations of results of laboratory experiment concerning rat's immune resistance to new vaccine. By following the procedure for generating bootstrap likelihood, equally spaced values of θ were chosen for the transformation of the original data sets. The transformed data sets were bootstrapped and the bootstrap kernel likelihoods were eventually generated. This was done for each parameter θ and for each higher order optimal window width. In the application of the higher order optimal window width, we took $f(x)$ to be a normal distribution and $k(x)$ to be the Gaussian kernel. The application of equation (3.3) yielded the graphs of the distribution of the generated higher order bootstrap likelihoods, which are shown in Figure 1. Also shown in figure 1 is the distribution of the exact bootstrap likelihood. Thus, the graphs in figure 1, give the bootstrap distribution functions for the "differences of means" at the various higher order optimal window width.

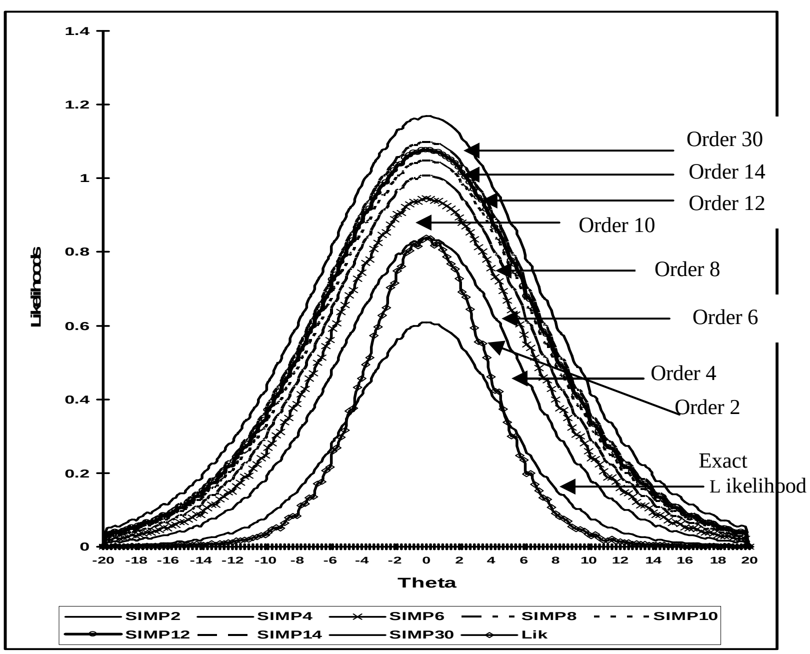


Figure 1: Densities of the Exact Likelihood and those of the Higher Order Kernel Bootstrap Likelihoods

Generally, the picture of the higher order bootstrap likelihood densities for orders 2, 4, 6, 8, 10, 12, 14 and 30 presented in figure 1 are of the same form. This is obviously so since they all stemmed from the same generating seeds. To generate each of the bootstrap kernel likelihood that gave rise to the densities in figure 1, we assumed $f(x)$ to be a normal density $k(x)$ to be the Gaussian kernel and did a bootstrap of

3,000 for each parameter θ. From looking at the pictures of the nine densities in figure 1, we believe that it is fair to say that in situations where the "true" density is the one provided by the exact bootstrap likelihood, the densities of orders 2 and 4 are more dominant in terms of fit. The shapes and general features of orders 2 and 4 densities are closest to that of the exact bootstrap likelihood. The pictures show that the densities of higher order kernels (i.e. ≥ order 6) have no advantage over those of orders 2 and 4. This is seen in the sense that the pictures of the densities for greater or equal to order 6 are all above that of the exact bootstrap likelihood. We observe that for some points, their values are above 1. This makes them unreliable as true densities. The densities for order 6 and beyond are clearly much farther from that of the exact bootstrap likelihood. Moreover, as from order 6, the densities are only marginally better than each other. In the foregoing, if it becomes relevant to make a choice of higher order kernel via the bootstrap likelihood approach, then the choice should be order 2 or 4. In this study, they (i.e. orders 2 and 4) exhibit better pictures – i.e. shapes and features – "closest" to that of the "true" densities (i.e. the exact bootstrap likelihood). Closest here means more comparable.

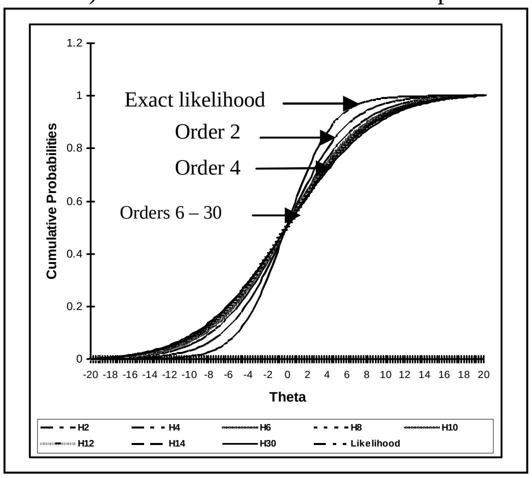


Figure 2: CDFs of the Exact Likelihood and those of the Higher Order Kernel Bootstrap Likelihoods

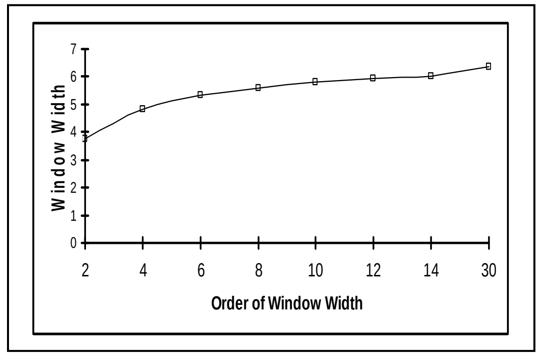


Figure 3: Distribution of Optimal Window Width as a Function of the Order (*m*)

Figure 2 presents the graphs of the cumulative density functions (cdfs) for the higher order bootstrap kernel likelihoods and also that of the exact bootstrap likelihood. These cdfs were generated by using the following expression:

$$F(\theta) = \frac{\sum_{i=1}^{k} P(X_i \le \theta)}{\sum_{i=1}^{n} P(X_i \le \theta)}, \quad k = 1(1)n, \text{ where } X_i = L^*_{m,opt}(\theta_i)$$

The pictures of the cdfs compliment those of the densities in Figure 1. Figure 3, shows the behaviour of $h_{m,opt}$ as a function of the order. Specifically, by using the results of our example, $h_{m,opt}$ is a linear function of the order (*m*) having a slope of 0.0732. A more mathematical approach for measuring the closeness between the higher order bootstrap kernel density functions and the exact (empirical) likelihood is in the form of a discrepancy such as the Kolmogorov–Smirnov (KS) or the Wald–Wolfowitz (WW) distance, $d_{ks}(F(\theta), F_m(\theta)) = \underset{\theta}{Sup} \left| F_e(\theta) - F_m(\theta) \right|$, where $F_e(\theta)$ is the cdf for the exact bootstrap likelihood and $F_m(\theta)$ is the cdf for order *m* bootstrap kernel likelihood. (See Boos, 1981). Tables 1 and 2 give the p–values and the most extreme absolute difference (distance) of the nonparametric goodness of fit tests of significance between the exact bootstrap likelihood and the higher order bootstrap likelihoods. The KS test gave a p–values of 0.000 for all tests of goodness of fit between the exact bootstrap likelihood and the higher order bootstrap likelihoods while the WW gave a p–values of 0.003 for order 2, 0.171 for order 4 and 0.000 for orders 6 – 30. Thus, while the KS test shows that there is significance difference between the exact bootstrap likelihood and all (higher) order bootstrap likelihoods including orders 2 and 4, the WW shows that there is no significant difference between the exact bootstrap likelihood and order 4 bootstrap likelihood but that there is significant difference between the exact bootstrap likelihood and orders 2, 6, 8, 10, 12, 14 and 30 bootstrap likelihoods.

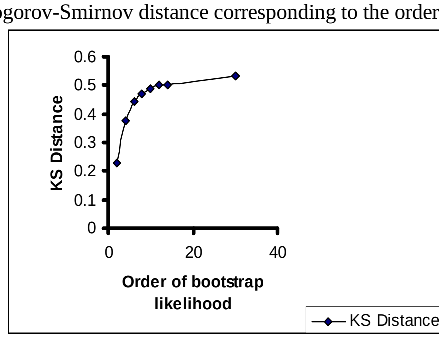
Table 1: *p*–values of the nonparametric goodness of fit test of significance between the exact bootstrap likelihood and the higher order bootstrap likelihoods.

Type of Nonparametric Test	***Order of the Bootstrap Likelihood***							
	2	4	6	8	10	12	14	30
Kolmogorov–Smirnov	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Wald–Wolfowitz	0.003	0.171	0.000	0.000	0.000	0.000	0.000	0.000

Table 2: Kolmogorov–Smirnov most extreme absolute difference between the exact bootstrap likelihood and the higher order bootstrap likelihoods.

	Order of the Bootstrap Likelihood							
	2	4	6	8	10	12	14	30
Most extreme absolute difference	0.228	0.374	0.443	0.468	0.488	0.500	0.500	0.532

Figure 4: A plot of Kolmogorov-Smirnov distance corresponding to the order of the bootstrap likelihood.



In table 2 we observe that the most extreme absolute difference (distance) given by the KS test shows that as the order of the bootstrap likelihood increases the distance also increases. This shows that the higher the order the larger the discrepancy between the exact bootstrap likelihood and the higher order bootstrap likelihood. However, it is observed that orders 2 and 4 provide the least “Kolmogorov–Smirnov distance”.

Figure 4 gives the distribution of the values of the KS distance against the order of the bootstrap likelihood. Again, we observe that the jumps between the densities of the higher order likelihoods (≥ 6) are quite small

5.0 Discussion

In the application of kernel density estimator to the generation of bootstrap likelihood, the smoothing parameter (i.e. the window width) usually comes into focus. In the earlier works, Ogbonmwan and Wynn (1988, 1992) and [11], the smoothing parameter of order 2 was utilized in generating bootstrap likelihood. This work introduces and examines the effect of higher order kernels in the generation of the bootstrap likelihood. According to [9] the effectiveness of higher order kernels depends on the sample size of the distribution, and queried how large the sample size should be. The works of [7] on the effectiveness of higher order kernels are also sample size dependent. The works of [5], [6], [9], etc. could not recommend the use of higher order kernels in practice due to over dependency on sample sizes. Our ideas circumvent the problem of sample size dependency in the effectiveness of higher order kernels. Our approach depends entirely on the bootstrap and therefore can have the desired number of bootstrap configurations for all orders of the kernels being considered. Thus in our situation, all conditions of the problem being solved are the same except for the values of the higher order window widths. Hence any observed differences whatsoever in the densities must have emanated from the values of the higher order window widths, $h_{m,opt}$ themselves. Our study benefits from the methodological aspects of the bootstrap, which continues to grow as computing, power increases (see [2]). The densities for the higher order kernels in figure 1 reveal the trade-off involved in the choice of the best order of the optimal window width. The smaller values of the window widths i.e. $h_{2,opt} = 3.76$ and $h_{4,opt} = 4.81$ in our example, yielded densities which are closed to that of the exact bootstrap likelihood.. On the other hand, the larger values of $h_{m,opt}(i.e. \geq h_{6,opt})$ produced densities, which are again seen to move father off from the exact bootstrap likelihood density. The densities for orders 2 and 4 are approximately close comparatively and competitively to the density of the exact bootstrap likelihood. Thus, orders 2 and 4 kernels are quite good and uniform in the sense that their vertical distances (comparing their densities with that of the "true" density) are small and fairly constant. Generally, by taking the exact bootstrap likelihood as the true density, the approximations of these densities are good for $h_{2,opt}$ $h_{4,opt}$ but poor for $h_{6,opt}$ $h_{8,opt}$ $h_{10,opt}$ $h_{12,opt}$ $h_{14,opt}$ and $h_{30,opt}$.

6.0 Conclusion

Practical applications of kernel density estimation are always dependent on the choice of the smoothing parameter (see [15]). In determining the choice of the window width many data driven approaches have been proposed and studied dating back to the work of [18] on the Least Squares Cross–Validation and going across to the work of [6]. All through the literature, the optimal window width (or the smoothing parameter) has mainly been of order 2. In this work, we have produced higher order bootstrap likelihoods. In the simulation example considered, we found that the densities (realized through the application of the higher order optimal window width in equation (3.3)) for orders 2 and 4 were better than those of orders 6, 8, 10, 12, 14 and 30 in comparison with the exact bootstrap likelihood. With this finding we argue that choice of the best optimal window widths are those of orders 2 and 4.

References

[1] D. D. Boos (1981), Minimum distance estimators for location and goodness of fit, Journal of American Statistical Association 76, 663 – 670.

[2] B. Efron (2003), Second thoughts on the bootstrap, Statistical Science, 18 (2), 135 – 140.

[3] V. I Epanechnikov (1969), Nonparametric estimation of a multivariate probability density, Theory of probability and its application, 14, 153 – 158.

[4] D. Fadda, E. Slexak and A. Bijaoui (1998) Density Estimation With Non-parametric Methods, Astron, Astrophys. Suppl. Ser 127, pp 335 – 352.

[5] Hanssen and Pauls (2003) How do Bootstrap and Permutation Tests Work? Annals of Statistic. Vol.31, 768 – 806.

[6] M C. Jones and D. E. Signorini (1997), A Comparison of High Order Bias Kernel Density Estimators JASA Vol 92, pp 1063 – 1078.

[7] M. C. Jones, D. F. Signorini, and N. L. Hjort (1999). On Multiplicative Bias Correction In Kernel Density Estimation. Sanklya, The Indian Journal of Statistics 61 (ser. A), 422 – 430.

[8] P. Hall and B. U. Park (2002), New Methods for Bias Correction at Endpoints and Boundaries, Annals of Statistics, Vol. 30 (#5) pp. 1460 – 1479.

[9] J. S. Marron and M. P. Wand (1992) Exact Mean Integrated Square Error – Annals of Statistics, Vol 20, pp 712 – 736

[10] Bruce E. Hansen (2003), Exact Mean Integrated Error of Higher – Order Kernel Estimators, University of Wisconsin, Madison, WI 53706, USA. – Sourced through the Internet (18.04.2004). Online at:http//www.ssc.wisc.edu/˜bhansen

[11] S. M. Ogbonmwan (2003), "Empirical Bootstrap Likelihood For Contrasts of Parameters". Journal of the Nigerian Association Mathematical Physics., Vol 7, pp. 97 – 104

[12] J. E. Osemwenkhae (2003). Higher Order Forms in Kernel Density Estimation, Ph.D Thesis, Department of Mathematics, University of Benin, Nigeria.

[13] J. E. Osemwenkhae and S. M Ogbonmwan, (2003), "Global Errors In Symmetric Kernels".:ABACUS (Journal of mathematical association of Nigeria) , Vol. 30 (# 2^A), pp. 130 - 139

[14] Hans–Goerg MÜller (1984), Smooth optimal kernel estimators of densities, regression curves and modes, Annals of Statistics 12, 766 – 774.

[15] B. U. Park and J. S Marron (1990). Comparison of Data – Drive B and Width Selectors, Journal of American Statistical Association, 85, 66 – 72.

[16] E. Parzen (1962), On estimation of a probability density and mode, Annals of Mathematical Statistics, 35, 1065 – 1076.

[17] D. N. Politis and J. P. Romano (1999), Multivariate density estimation with general flat – top kernels of infinite order, Journal of Multivariate Analysis, 68, 1 – 25.

[18] M. Rudemo (1982). Empirical Choice of Density Estimators, Scandinavian Journal of Statistics, 9, 65 – 78.

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