IMPROVING THE CHOICE OF HIGHER ORDER UNIVARIATE KERNELS THROUGH BIAS REDUCTION TECHNIQUE

J. E. Osemwenkhae and J. I. Odiase
Department of Mathematics
University of Benin
Benin City, Nigeria

ABSTRACT

Within the last two decades, higher order univariate kernels have been under focus with respect to its importance in examining the concept of curve fitting. This paper has taken this direction by examining some basic properties of the univariate kernels in assessing and improving the choice of kernels. The minimum efficiency of the selected kernels is 82% at order 6. The global error diminishes as the order of h increases, and it is highest between orders 2 and 6, and beyond order 12 the global error seems to level off. Depending on the tolerance limit specified for the MISE and the percentage efficiency permitted, the extent of bias reduction required, can be monitored.

Keywords: *Higher order kernels, bias reduction, efficiency, global error.*

INTRODUCTION

One popular nonparametric method for estimating a probability function f, is the kernel density estimation, Parrzen (1962), Rosenblatt (1956) and Wand and Jones (1995). Let $X_1, X_2, \ldots, X_n$ be a random sample from a distribution with unknown density f, for which an estimator $\hat{f}$ is to be constructed. The univariate kernel estimate at x is

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

where k is usually a symmetric probability density function and the window width, h, is a smoothing parameter. The choice of the smoothing parameter is crucial because it is a trade off between the bias and the variance terms in the mean integrated square error (MISE). The kernel density estimation method has been studied extensively in the last two decades and has been found to perform very well under many circumstances- see for example Izenman (1991) and Minnottee (1998).

There are several ways of improving performance of the basic kernel estimator of (1), see Rosenblatt (1956), Rudemo (1982) and Silverman (1986). There has been a wide variety of applications of the basic kernel estimator in (1) with many pointing to the possibility of harmonizing higher order kernels and bias corrections and reduction approaches, see Scott (1992), Jones and Signorini (1997) and Osemwenkhae (2003).

This paper has the following aims:

i) to show how bias reduction techniques can be achieved theoretically,
ii) to show the consequence of (i) above on the mean integrated square error (MISE).
iii) to show how (i) and (ii) would enhance our choice of kernel via their efficiency and
iv) to show empirically how (i)-(iii) can improve our choice of kernels at higher order.

If the kernel estimator defined in (1) satisfies the second order symmetric regularity conditions:

$$\text{i) } \int k(t)dt = 1 \qquad \text{ii) } \int tk(t)dt = 0$$

$$\text{iii) } \int t^2 k(t)dt = V_2 \neq 0 \qquad (2)$$

Silverman (1986) obtained (asymptotically) for (2) the optimal window width, h, the MISE of $\hat{f}$ and the efficiency of k respectively as:

$$h_{opt} \approx n^{\frac{-1}{5}} \left\{\int k(t)dt\right\}^{\frac{1}{5}} V_2^{-\frac{2}{5}} \left\{\int f''(x)^2 dx\right\}^{\frac{-1}{5}} \qquad (3)$$

$$MISE\ \hat{f}(x) \approx \tfrac{5}{4} V_2^{2/5} \left\{\int k(t)^2 dt\right\}^{\frac{4}{5}} \left\{\int f''(x)^2 dx . n^{-\frac{4}{5}}\right\} \qquad (4)$$

and

$$Eff(k) \approx \frac{3}{5\sqrt{5}} \left\{\int t^2 k(t)dt\right\}^{-\frac{1}{2}} \left\{\int k(t)^2 dt\right\}^{-1} \qquad (5)$$

where n is the sample size. These fundamental properties (3) – (5) have attracted a lot of researches and modifications in recent times, see Polansky and Bakar (2000), Marzio and Taylor (2004).

Generalized bias reduction in kernel density estimation

The second order symmetric regularity conditions (2) only permits the use of kernels that take non-negative values. An earlier argument put forward by Parzen (1962) and Bartlett (1963) and later introduced into higher order kernels by Scott and Wand (1991), Jones et al (1995), Jones and Signorini (1997), Osemwenkhae and Ogbonmwan (2003) and Osemwenkhae (2003) allows the use of kernels which take both negative and positive values. The basic regularity conditions in (2) are modified thus:

$$\text{i) } \int k(t)dt = 1$$

$$\text{ii) } \int tk(t)dt = \ldots = \int t^{m-1} k(t)dt = 0$$

$$\text{iii) } \int t^2 k(t)dt = V_2 \neq 0 \qquad (6)$$

where m is even. But

$$MISE\ \hat{f}(x) = \int\left\{E\hat{f}(x) - f(x)\right\}^2 dx + \int \text{var}\ \hat{f}(x)dx$$

$$\equiv \int\left(Bias\ \hat{f}(x)^2\right)dx + \int \text{var}\ \hat{f}(x)dx \qquad (7)$$

To obtain the bias term corresponding to (6) take the terms of the Taylor series expansion of $f(x)$ up to m, that is, h is of order m, and obtain

$$Bias\hat{f}(x) = E\hat{f}(x) - f(x) \approx \frac{1}{m!} h^m f^{(m)}(x) V_m \qquad (8)$$

$$\text{var}\ \hat{f}(x) \approx n^{-1} h^{-1} k(t)^2 \qquad (9)$$

The generalized asymptotic bias term (8) for any given even order of the smoothing parameter h and the variance term (9) are substituted into (10) and minimized over h, thus

$$h_{opt} \approx \left\{\frac{(m!)^2}{2m} V_m^{-2} \left[\int k(t)^2 dt\right] \left[\int f^{(m)}(x)^2 dx\right]^{-1} n^{-1}\right\}^{\frac{1}{2m+1}} \qquad (10)$$

The burden of obtaining the value of the optimal window width when (6) is satisfied is greatly simplified by (10).

$$MISE\,\hat{f}(x) \approx \frac{2m+1}{2m}\left\{\left[\frac{2m}{(m!)^2}\right]V_m^2\left[\int k(t)^2 dt\right]^{2m}\left[f^{(m)}(x)^2 dx\right]n^{-2m}\right\}^{\frac{1}{2m+1}} \quad (11)$$

To obtain the efficiency of any symmetric kernel corresponding to (6), adopt the ratio defined in Silverman (1986). For the Epanechnikov kernel $K_e(t) = \frac{3}{4\sqrt{5}}\left(1 - \frac{1}{5}t^2\right),\ |t| \le \sqrt{5}$. (6iii) becomes

$$\int_{-\sqrt{5}}^{\sqrt{5}} t^m k_e(t)dt = \frac{3(\sqrt{5})^m}{(m+1)(m+3)} \quad (12)$$

and $$\int_{-\sqrt{5}}^{\sqrt{5}} k_e(t)^2 dt = \frac{3(\sqrt{5})}{25} \quad (13)$$

therefore (12) and (13), the efficiencies of any m^{th} order symmetric kernel is

$$Eff(k_m) \approx \frac{\left\{\left[\frac{3(\sqrt{5})^m}{(m+1)(m+3)}\right]^{\frac{2}{2m+1}}\left[\frac{3}{5\sqrt{5}}\right]^{\frac{2m}{2m+1}}\right\}^{\frac{2m+1}{2m}}}{V_m^{\frac{1}{m}}\int k(t)^2 dt}$$

$$= \frac{1}{5}\left\{\frac{3^{(1+m)}}{(m+1)(m+3)}\right\}^{\frac{1}{m}}\left\{\int t^m k(t)dt\right\}^{-\frac{1}{m}} \cdot \left\{\int k(t)^2 dt\right\} \quad (14)$$

$$Eff(k) \approx \frac{1}{5}\left\{\frac{3^{(1+m)}}{(m+1)(m+3)}\right\}^{\frac{1}{m}}\left\{\int t^m k(t)dt\right\}^{-\frac{1}{m}}\left\{\int k(t)^2 dt\right\}^{-1} \quad (15)$$

A close look at (3) and (10) reveal that the order of smoothing parameter h has reduced from $n^{-1/5}$ to $n^{-\frac{1}{2m+1}}$ Also, (4) and (11) reveal that the order of the MISE reduced from $n^{-4/5}$ to $n^{-\frac{2m}{2m+1}}$

In (5) and (14), a definite improvement and generalization of the efficiency is revealed. The results of (8) – (14) have generalized the values of h_{opt}, MISE for any even order h and the efficiency of any symmetric kernel. The equation for the generalized bias (7) reduces the size of the global error (MISE) at any given even order of the smoothing parameter *h*. For proofs of (8) – (14) see Osemwenkhae (2003).

Empirical implementation of (11) and (14) to depict the relevance of higher order kernels is achieved by examining the Rectangular, Bi-

Table 1: Values of MISE and Efficiencies at higher order of *h*

	$MISE\ \hat{f}(x)$			Efficiencies		
	Selected Kernels			Selected Kernels		
Order (*h*)	Rectangular	Biweight	Gaussian	Rectangular	Biweight	Gaussian
2	6.022770 E-03	5.708567 E-03	5.911735 E-03	0.9295	0.9939	0.9512
4	3.336485 E-03	3.330274 E-03	3.661083 E-03	0.9709	0.9730	0.8745
6	2.519964 E-03	2.619974 E-03	3.039569 E-03	0.9992	0.9580	0.8174
8	2.108022 E-03	2.261830 E-03	2.753537 E-03	1.0201	0.9466	0.7679
10	1.851571 E-03	2.036930 E-03	2.589845 E-03	1.0362	0.9315	0.7284
12	1.672666 E-03	1.877929 E-03	2.484019 E-03	1.0492	0.9251	0.6949
14	1.538710 E-03	1.757027 E-03	2.410043 E-03	1.0601	0.9198	0.6660

weight and Gaussian kernels. The data generating the results of Table 1 is the call duration in seconds of 100 independent GSM calls from a business centre in an urban city in Nigeria.

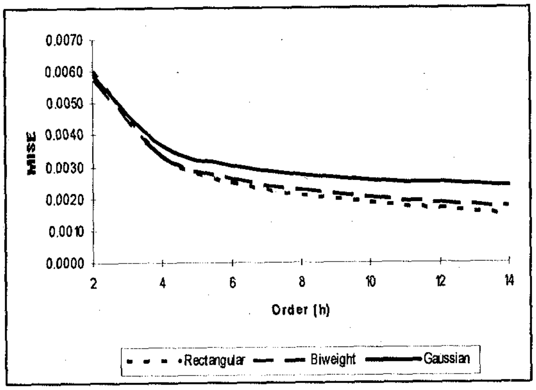

Fig. 1: Graph of the MISE of the 3 selected kernels for higher h.

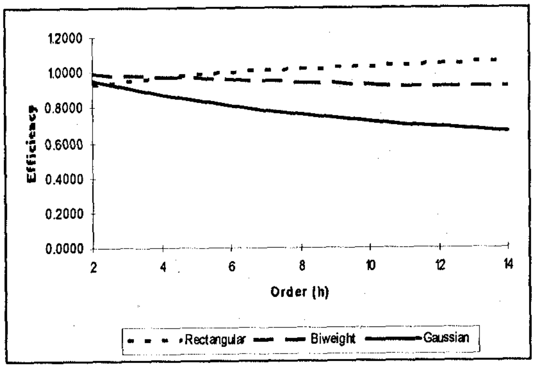

Fig. 2: Graph of the Efficiencies of the 3 selected kernels for higher h.

CONCLUSION

The MISE and the efficiencies of three selected kernels were obtained using the call duration of 100 independent calls from a commercial phone booth up to the order of the smoothing parameters, $h = 14$. The graphs in Figure 1 fall sharply when h increases from 2 to 6 and beyond this there is a gradual leveling off with no appreciable decrease after order 12. This reduction is caused by the reduction in the bias term (11). Bias measures the rapidity of fluctuations of the fitted curve from one caller to another and one way of reducing this fluctuation is to fit the curve using the order of the smoothing parameter greater than 4. However, beyond order 8, the decrease in MISE is negligible.

In the same vein, from Table 1, the efficiency of the selected kernels showed a minimum of 82% at order 6 with the efficiency of the Gaussian kernel becoming poor at $h > 6$. Modelling call time with the Gaussian kernel may not be the best, especially at higher order values of h, as its efficiency cannot be guaranteed. Nevertheless, the fact that the Gaussian kernel is differentiable to any order is an advantage. So, depending on the tolerance limit specified for the MISE and the percentage efficiency admitted, the extent of bias reduction can be predetermined.

Higher order univariate Kernel Density Estimator is important in the continuous reduction in the size of the global error propagated as the order of h increases. However, the extent of the relevance of higher order univariate kernel density estimation is dependent upon a specified allowable tolerance limit and the percentage efficiency so desired.

REFERENCES

Bartlett, M. S. (1963). Statistical estimation of density functions. *Sankhya*, 25, 245-254.

Izenman, A. J. (1991). Recent developments in nonparametric density estimation. *Journal of Amer. Stat.*, 1217-1223.

Jones, M. C., and Signorini, D. F. (1997). A comparison of higher order bias kernel density estimators. *Journal of Amer. Stat. Ass.*, 92, 1063-1073.

Jones, M. C., Linton, O., and Nielson, J. P. (1995). A simple bias reduction method for kernel density estimation. *Biometrika* 82, 327-338.

Marzio, M. D., and Taylor C. C. (2004). Boosting kernel density estimates: A bias reduction technique. *Biometrika*, 71, 226-233.

Minnotee, M. C. (1998). Achieving higher order convergence rate in density estimation with binned data. *Journal of Amer. Stat. Ass.*, 93, 663-672.

Osemwenkhae, J. E. (2003). Higher order forms in kernel density estimation method: Ph.D, Thesis, Department of Mathematics, Univ. of Benin, Nigeria.

Osemwenkhae, J. E., and Ogbonmwan, S. M. (2003). Generalized efficiencies for higher order symmetric univariate kernels. *Journal of the Nigeria Ass. of Math. Physics*, 7, 89-96.

Parzen, E. (1962). On estimation of a probability density function and mode. *Ann. of Math. Stat.* 33, 1065-1076.

Polansky, A. M., and Barker, E. R. (2000). Multistage plug-on bandwidth selection for kernel density formation estimate. *Journal of Stat. Comp. Simul.* 65, 63-80.

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

Rudemo, M. (1982). Empirical choice of histograms and kernel density estimators. *Scand. Journal of Stat.* 9, 65-78.

Scott, D. W. (1992). Multivariate density estimation: Theory, practice and visualization. New York: John Wiley and Sons.

Scott, D. W., and Wand, M. P. (1991). Feasibility of multivariate density estimates. *Biometrika*, 78, 197-206.

Silverman, B. W., (1986) Density estimation for statistics and data analysis. London: Chapman and Hall.

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