J. Nig. Ass. Math. Phys.
Vol. 4 (2000)

HIGHER ORDER FORMS FOR OPTIMAL WINDOW WIDTH IN MULTIVARIATE KERNEL DENSITY ESTIMATION

S. M. OGBONMWAN AND E. J. OSEMWENKHAE
DEPARTMENT OF MATHEMATICS UNIVERSITY OF BENIN
BENIN CITY, NIGERIA

ABSTRACT
The Univariate Kernel Density Estimation (KDE) has received tremendious attention in the literature. But this can not be said of the Multivariate Kernel Density Estimation (KDE) which has received very little attention so far. In this work, a generalized higher order forms for the window width of the Multivariate Kernel Density Estimation is considered. This extends the work of Ogbonmwan and Osemwenkhae (1998). The work provides ways by which the rates of convergence of the Optimal Window Width are increased leading to further reduction of the global error – MISE.

1. INTRODUCTION:

Various methods of estimating density abound and in particular the univariate kernel density method of estimating an unknown function f is well known and studied in the literature [see Silverman (1986), Alan (1991), Ogbonmwan (1993), Titterington (1983) Taylor (1989) Hossjer and Ruppert (1995)]. For the univariate data case, the fixed kernel estimator K is defined by

$$\hat{f}(x,h)=\frac{1}{nh}\sum_{i=1}^{\infty}k\left(\frac{x-X_i}{h}\right) \qquad (1.1)$$

where h is the optimal window width and k (.) is the Kernel.

If k satisfies the symmetric conditions listed in (3.12) of Silverman (1986):

(i) $\int k(t)dt=1$

(ii) $\int tk(t)dt=0$

(iii) $\int t^2k(t)dt=V_2\neq 0$

then the associated h is given as:

$$h_2=n^{-1/5}\left[\int k(t)^2\,dt\right]^{1/5}\left[\int f''(x)^2\,dx\right]^{1/5}V_2 \qquad (1.2)$$

The resultant associated bias and MISE terms are respectively

$$\int(bias_h(x))^2\,dx=\tfrac{1}{4}h^2V_2^2\int\left(f''(x)\right)^2\,dx \qquad (1.3)$$

and

$$MISE_x \hat{f}(x) = \tfrac{5}{4} V_2^{-2/5} \left\{ \int k(t)^2 dt \right\}^{4/5} n^{-4/5} \left\{ \int f''(x)^2 dx \right\}^{\frac{1}{5}} \quad (1.4)$$

[see Ogbonmwan & Osemwenkhae (1997)].
When some more stricter conditions such as:

(i) $\int k(t)dt = 1$

(ii) $\int t^{2m-1} k(t)dt = 0$

(iii) $\int t^{2m} k(t)dt = V_{2m} \neq 0$

are imposed on the symmetric nature of the Kernel K, the optimal window width becomes

$$h_{2m} = \left\{ \frac{((2m)!)^2}{4m} \right\}^{\frac{1}{4m+1}} V_{2m}^{-\frac{2}{4m+1}} \left\{ \int k(t)^2 dt \right\}^{\frac{1}{4m+1}} \left\{ \int f^{(2m)}(x)^2 dx \right\}^{-\frac{1}{4m+1}} n^{-\frac{1}{4m+1}} \quad (1.5)$$

[see Ogbonmwan & Osemwenkhae (1998)]
Under these conditions, the corresponding bias and MISE terms are:

$$\int (bias_h(x))^2 dx = \left(\frac{1}{(2m)!} \right)^2 h^{4m} V_{2m}^2 \int \left(f^{(2m)}(x) \right)^2 dx \quad (1.6)$$

and

$$MISE_x \hat{f}(x) = \tfrac{4m+1}{4m} \left\{ \frac{4m}{((2m)!)^2} \right\}^{\frac{1}{4m+1}} V_{2m}^{\frac{2}{4m+1}} \left\{ \int k(t)^2 dt \right\}^{\frac{4m}{4m+1}} \left\{ \int f^{(2m)}(x)^2 dt \right\}^{\frac{1}{4m+1}} n^{-\frac{4m}{4m+1}} \quad (1.7)$$

where $V_{2m} = \int t^{2m} k(t)dt$ and $m = 1,2,3,---,< \infty$

The essence of all these were to reduce the global error corresponding to this case – the MISE

2. THE MULTIVARIATE KERNEL DENSITY ESTIMATION (MKDE) METHOD

The multivariate kernel density estimation with Kernel K and window width h is defined by:

$$\hat{f}(n) = \frac{1}{nh^d} \sum_{i=1}^{n} k \left\{ \tfrac{1}{h}(\chi - x_i) \right) \quad (2.1)$$

where K (.) is a symmetric d-dimensional density function and where it would be possible to scale the components of x to allow for different variances. Intrinsically, the use of a single h in (2.1) implies that the Kernel version on each data point is scaled equally by either using a vector of smoothing parameters or using a matrix of shrinking coefficient: this is so if the spread of data in one or more co-ordinate direction is much greater than the others – [see Silverman(1986)], [Bowman & Foster (1993)].

Apart from the multivariate kernel density estimation methods, other methods for multivariate density estimation include: The Multivariate Histogram, the scatter plots.

(i) $\int_{R^d} k(t)dt = 1$

(ii) $\int_{R^d} t^{2m-1} k(t)dt = 0$ (4.1)

(iii) $\int_{R^d} t^{2m} k(t)dt = V_{2m} \neq 0 \ for\ m = 1,2,3,---,<\infty$

Then:

(1) $h_{opt} \approx \left\{\frac{((2m)!)^2}{4m}\right\}^{\frac{1}{d+4m}} V_{2m}^{-\frac{2}{d+4m}} \beta^{\frac{1}{d+4m}} \left\{\int\left(\nabla^{2m} f(x)\right)^2 dx\right\}^{-\frac{1}{d+4m}}$

$n^{-\frac{1}{d+4m}} \quad for\ m = 1,2,3,---,<\infty$

and

(2)

$$MISE\ \ f(x) \approx \frac{4m+1}{4m}\left\{\frac{4m}{((2m)!)^2}\right\}^{\frac{d}{d+4m}}\left\{V_{2m}^{\frac{2d}{d+4m}}\beta^{\frac{4m}{d+4m}}\left[\int\left(\nabla^{2m} f(x)\right)^2 dx\right]^{\frac{d}{4m+d}}\right\}$$

$n^{-\frac{4m}{d+4m}}$

where β is as defined in section 2.

Proof:

For the optimal window width,

The bias term is

$$\text{bias}\quad f(x) = E f(x) - f(x)$$

$$= \int k(t)\left\{f(x) - ht\nabla f(x) + \tfrac{1}{2}h^2t^2\nabla^2 f(x)\right\} + \cdots + \tfrac{1}{(2m)!}h^{2m}\nabla^{2m} f(x)V_{2m} + \cdots \}dt - f(x)$$

$$\approx \frac{1}{2m}!\, h^{2m}\nabla^{2m} f(x)\int k(t)dt$$

$$= \frac{1}{(2m)!} h^{2m}\nabla^{2m} f(x)V_{2m} \qquad (4.2)$$

The variance term here is:

$$\int Var\ f = n^{-1}h^{-d}\beta \qquad (4.3)$$

with (4.2) and (4.3) above;

$$MISE\ \ f(x) = \frac{1}{((2m)!)^2} h^{4m} V_{2m}^{2} \int\left(\nabla^{2m} f(x)\right)^2 dx + n^{-1}h^{-d}\beta$$

minimizing over h we get;

$$h^{4m+d} = \frac{((2m)!)^2}{4m} V_{2m}^{-2} \beta \left\{ \int (\nabla^{2m} f(x))^2 \right\}^{-1} n^{-1}$$

$$\therefore h_{opt} \approx \left\{ \frac{((2m)!)^2}{4m} \right\}^{\frac{1}{4m+d}} V_{2m}^{-\frac{2}{d+4m}} \beta^{\frac{1}{d+4m}} \left\{ \int (\nabla^{2m} f(x))^2 dx \right\}^{-\frac{1}{d+4m}} n^{-\frac{1}{d+4m}} \qquad (4.5)$$

for $m = 1,2,3,---,< \infty$

Critical view of equation (4.5) shows that the order of h is thus reduced to $n^{-\frac{1}{d+4m}}$ hence there is an improvement over equations (3.4) and (2.2). This thus speaks for any finite m. for the MISE,

recall that MISE $\hat{f}(x) = \int (Bias)^2 dx + \int \text{var}\ \hat{f}(x) dx$

$$\text{MISE } \hat{f}(x) = \frac{1}{((2m)!)^2} h^{4m} V_{2m}^2 \int (\nabla^{2m} f(x))^2 dx + n^{-1} h^{-d} \beta$$

With the definition of h_{opt} above,

$$MISE\ \hat{f}(x) \approx \frac{1}{((2m)!)^2} \left[\frac{((2m)!)^2}{4m} \right]^{\frac{4m}{d+4m}} V_{2m}^{-\frac{8m}{d+4m}} \beta^{\frac{4m}{d+4m}} .$$

$$\left[\int (\nabla^{2m} f(x))^2 dx \right]^{-\frac{4m}{d+4m}} n^{-\frac{4m}{d+4m}} V_{2m}^2 \int (\nabla^{2m} f(x))^2 dx$$

$$+ n^{-1} \left[\frac{((2m)!)^2}{4m} \right]^{-\frac{d}{d+4m}} V_{2m}^{\frac{2d}{d+4m}} \beta^{-\frac{d}{d+4m}} \left[\int (\nabla^{2m} f(x))^2 dx \right]^{\frac{d}{d+4m}} n^{\frac{d}{d+4m}} \beta$$

$$= \frac{1}{((2m)!)^2} \left\{ \frac{((2m)!)^2}{4m} \right\}^{\frac{4M}{D+4M}} V_{2m}^{\frac{2d}{d+4m}} \beta^{\frac{4m}{d+4m}} \left[\int (\nabla^{2m} f(x))^2 dx \right]^{\frac{d}{d+4m}}$$

$$n^{-\frac{4m}{d+4m}} + \left[((2m)!)^2 \right]^{-\frac{d}{d+4m}} V_{2m}^{\frac{2d}{d+4m}} \beta^{\frac{4m}{d+4m}} \left\{ \int (\nabla^{2m} f(x))^2 \right\}^{\frac{d}{d+4m}} n^{-\frac{4m}{d+4m}}$$

$$= V_{2m}^{\frac{2d}{d+4m}} \beta^{\frac{4m}{d+4m}} \left[\int (\nabla^{2m} f(x))^2 dx \right]^{\frac{d}{d+4m}} n^{-\frac{4m}{d+4m}} \left\{ \frac{1}{((2m)!)^2} \left(\frac{((2m)!)^2}{4m} \right)^{\frac{4m}{d+4m}} + \left[\frac{((2m)!)^2}{4m} \right]^{-\frac{d}{d+4}} \right\}$$

$$= V_{2m}^{\frac{2d}{d+4m}} \beta^{\frac{4m}{d+4m}} \left[\int (\nabla^{2m} f(x))^2 dx \right]^{\frac{d}{d+4m}} n^{-\frac{4m}{d+4m}} \left\{ \frac{\frac{1}{((2m)!)^2} \left[\frac{((2m)!)^2}{4M} \right] + 1}{\left[\frac{((2m)!)^2}{4m} \right]^{\frac{d}{d+4m}}} \right\}$$

$$= V_{2m}^{\frac{2d}{d+4m}} \beta^{\frac{4m}{d+4m}} \left[\int (\nabla^{2m} f(x))^2 dx \right]^{\frac{d}{d+4m}} n^{-\frac{4m}{d+4m}} \left\{ \frac{\frac{1+4m}{4m}}{\left[\frac{((2m)!)^2}{4m} \right]^{\frac{d}{d+4m}}} \right\}$$

the Nearest Neighbourhood Approach, the Variable Kernel Approach, the Adaptive Kernel Estimator, the Orthogonal Series Estimator, the Delta Sequence Density Estimator, the Restricted Maximum Likelihood Estimator and more recently the Projection Pursuit Density approach in which both dimensionality reduction and density estimation can be pursued at the same time.

With assumption that the Kernel defined in (2.1) above is a radially symmetrical probability density function and that the unknown f has bounded and continuous second derivatives, we can easily see that

$$h_{opt} = \left\{ d\beta \propto^{-2} \left[\int (\nabla^2 f)^2 dx \right]^{-1} \right\}^{\frac{1}{d+4}} n^{-\frac{1}{d+4}} \tag{2.2}$$

and

$$MISE \quad f(x) = \tfrac{5}{4} d^{-d/_{d+4}} \propto^{\frac{2d}{d+4}} \beta^{\frac{4}{d+4}} \left\{ \int \left(\nabla^2 f \right)^2 dx \right\}^{-d/_{d+4}} n^{-\frac{4}{d+4}} \tag{2.3}$$

where

$$\propto = \int t^2 k(t)\, dt$$

$$\beta = \int k(t)^2\, dt$$

and ∇ is the del operator acting on the function f (.). Observe that equation (2.2) converges to zero as n increases. However the convergence is very short (at the rate of $n^{-\frac{1}{d+4}}$).

In the next section, our attempt shall be to observe ways by which the rate of convergence of optimal window width is increased and also to further reduce the global error – MISE by considering further bias reduction technique.

3. BIAS REDUCTION TECHNIQUE IN MULTIVARIATE KDE

Case 1: When h is of order 4.

Suppose the symmetric kernel K satisfies

$$\left.\begin{aligned}
&\text{(i)} && \int k(t)\, dt = 1 \\
&\text{(ii)} && \int t\, k(t)\, dt = \int t^2\, k(t) dt = \int t^3\, k(t)\, dt = 0 \\
&\text{(iii)} && \int t^4\, k(t)\, dt = V_4 \neq 0
\end{aligned}\right\} \tag{3.1}$$

Note that the conditions in (3.1) above is so because $K(t)$ is no longer non-negative as the kernel $K(t)$ now takes both positive and negative values – [see Silverman (1986) Osemwenkhae (1994)]

Here, $Bias_h(x) = E\hat f(x) - f(x)$

$$= h^{-d}\int k\left(\frac{x-y}{h}\right)f(y)dy - f(x)$$

$$= \int k(t)\{f(x) - ht\nabla f(x) + \tfrac{1}{2}h^2t^2\nabla^2 f(x)\}$$

$$+ \tfrac{1}{6}h^3t^2\nabla^3 f(x) + \tfrac{1}{24}h^4t^4\nabla^4 f(x) + ---]dt - f(x)$$

$$\approx \tfrac{1}{24}h^4 t\nabla^4 f(x)k(t)dt$$

$$= \tfrac{1}{24}h^4\nabla^4 f(x)V_4 \qquad (3.2)$$

similarly, $\quad var\,\hat f(x) = n^{-1}h^{-d}\beta$

The MISE corresponding to this window width is:

$$\hat f(x) \approx \tfrac{1}{576}h^4 V_4^2 \int\{\nabla^4 f(x)\}^2 dx + n^{-1}h^{-d}\beta \qquad (3.3)$$

Differentiating wrt. h and equating to zero, we get

$$h = (72)^{\frac{1}{d+8}} V_4^{-\frac{2}{d+8}} \beta^{\frac{1}{d+8}}\left[\int(\nabla^4 f(x))^2 dx\right]^{-\frac{1}{d+8}} n^{-\frac{1}{d+8}} \qquad (3.4)$$

the rate of convergence of (3.4) is faster than that of (2.2) - at the rate $n^{-\frac{1}{d+8}}$

Substituting (3.4) into (3.3) we have,

$$MISE\,\hat f(x) \approx \tfrac{1}{576}(72)^{\frac{8}{d+8}} V_4^{-\frac{16}{d+8}} \beta^{\frac{8}{d+8}}\left\{\int(\nabla^4 f(x))^2 dx\right\}^{-\frac{8}{d+8}}$$

$$n^{-\frac{8}{d+8}} V_4^2 \int\{\nabla^4 f(x)\}^2 dx + n^{-1}(72)^{-\frac{d}{d+8}} V_4^{\frac{2d}{d+8}} \beta^{-\frac{d}{d+8}}\left\{\int(\nabla^4 f(x))^2 dx\right\}^{\frac{d}{d+8}} n^{\frac{d}{d+8}}\beta$$

$$= V_4^{\frac{2d}{d+8}} \beta^{\frac{8}{d+8}}\left\{\int(\nabla^4 f(x))^2 dx\right\}^{\frac{d}{d+8}} n^{-\frac{8}{d+8}}\left\{\frac{(72)^{\frac{8}{d+8}}}{576} + \frac{1}{(72)^{\frac{d}{d+8}}}\right\}$$

$$= V_4^{\frac{2d}{d+8}} \beta^{\frac{2}{d+8}}\left\{\int(\nabla^4 f(x))^2 dx\right\}^{\frac{d}{d+8}} n^{\frac{2}{d+8}}\left\{\frac{27}{24}\frac{1}{(72)^{\frac{d}{d+8}}}\right\}$$

$$= \left[V_4^{\frac{2d}{d+8}} \beta^{\frac{8}{d+8}}\left\{\int(\nabla^4 f(x))^2 dx\right\}^{\frac{d}{d+8}} n^{-\frac{8}{d+8}}\left\{9^{\frac{8}{d+8}} 2^{-\frac{6(d+4)}{d+8}}\right\} \qquad (3.5)$$

4. CASE2: WHEN h IS OF ORDER 2m FOR $m = 1,2,3,---,<\infty$

Theorem:

Suppose we insist that the **multivariate kernel density estimation** in (2.1) above satisfies:

$$= \frac{4m+1}{4m}\left[\frac{4m}{((2m)!)^2}\right]^{\frac{d}{d+4m}}\left\{V_{2m}^{\frac{2d}{d+4m}}\ \beta^{\frac{4m}{d+4m}}\left[\int\left(\nabla^{2m} f(x)\right)^2 dx\right]^{\frac{d}{d+4m}}\ n^{-\frac{4m}{d+4m}}\right\} \qquad (4.6)$$

This is the corresponding equation of the MISE for equation (4.1). Clearly, the order of MISE has been reduced to $n^{-4m/4m+1}$.

The inspiring results here is that (3.5), (4.5), and (4.6) have reduced the burden of calculating Bias $\hat{f}(x)$, h_{opt} and MISE for successive even powers of h.

REFERENCES

(1) Alan J. I. (1991): Recent development in Nonparametric Density Estimation. Journal of American Stat. Ass, Vol. 86, No 413 pp 205 – 221.

(2) Bowman A. W. & Foster P. J. (1993): Adaptive Smoothing and Density – Dased Test of Multivariate Normality. Journal of American Stat. Ass Vol. 88 No. 22 pp. 529 – 537

(3) Hossjer O. and Ruppert D. (1995): Asymptotics for the transformation kernel density estimation Annals of Statistics, 23 No. 4, 1198 – 1222

(4) Ogbonmwan S. M. (1993): On Kernel Density Estimation and the likelihood, proceedings of the 17th Annals conference on NSA pp. 93 – 102

(5) Ogbonmwan S. M. & Osemwenkhae E. J. (1997): On the choice of kernel density estimates. Journal of NSA Vol. 11 No. 1

(6) Ogbonmwan S. M. & Osemwenkhae E. J. (1998): Generalized higher order forms for optimal window width in kernel density estimation. Accepted for publication in the Annals of Natural Sciences (Ekpoma).

(7) Osemwenkhae E. J. (1994): On density estimations: A study of the univariate kernel density. M.Sc thesis University of Benin.

(8) Silverman B. W. (1986): Density Estimation for Statistical Data Analysis, Chapman & Hall.

(9) Taylor C. (1989): Bootstrap choice of the smoothing parameter in kernel density estimation. Biometrika, 76, 4 pp. 705 – 712

(10) Titterington D. M. (1983): Smoothing techniques in Statistics. Inter. Stat. Review, Vol. 53 No. 2 pp. 141 – 170.