Journal of the Nigerian Association of Mathematical Physics
Volume **18 (May, 2011), pp 353 – 356**
© J. of NAMP

Rates of Convergence of Higher order forms of Multivariate Kernel Density Estimation

F. O. Oyegue And J. E. Osemwenkhae
Department Of Mathematics,
University Of Benin,
Benin City, Nigeria

Abstract

The efficacy of the multivariate kernel density estimator is measured by the closeness of kernel density estimates $\hat{f}$ to its true density f. One of the most used measures on the global accuracy of the multivariate kernel density estimator is the mean Integrated Squared Error (MISE). In this paper, we obtain the generalized higher order optimal bandwidth and the asymptotic mean integrated square error (AMISE) for the multivariate kernel density estimator with faster rates of convergence than some of the ones in the literature. The advantage is that it enables one to know the speed at which the estimated density approaches the true density.

Keywords: Multivariate kernels, Mean Integrated Squared Error, Optimal Bandwidth, Rates of Convergence

1. Introduction:

Density estimation is the construction of an estimate $\hat{f}$ of an underlying density function f for a random variable X from observed data. There are two approaches to density estimation: parametric and nonparametric. The parametric ones, e.g. the maximum likelihood method, requires assumptions on the form of the unknown density. Though imposing the functional form is quite subjective, this does simplify the problem. The only problem is to estimate the parameters of the distributions. Sometimes, however, having no additional information about the distribution we use nonparametric methods such as the kernel density estimation.

Several other types of nonparametric density estimations have been discussed in the monographs of [1, 8, 14]. However, concentration will be on the kernel density estimator due to its simple mathematical framework and also its being an important technique in exploratory data and confirmatory analyses, and presentation of data-[14]. Only recently , for example, [9] looked at kernel density estimation for grouped data with application to line transect sampling.

The kernel density estimator, introduced by [12] (for the univariate case), is characterized by two components, the bandwidth h and the kernel K. We consider its multivariate version,where the dimension is $d \geq 2$.

The general form of the d-dimensional kernel density estimator given in [3] is

$$\hat{f}(\mathbf{x}; H) = n^{-1} \sum_{i=1}^{n} K_H(\mathbf{x} - X_i) \tag{1.1}$$

where $\mathbf{x} = (x_1, ..., x_d)^T$ and $X_i = (X_{i1}, X_{i2}, ..., X_{id})^T, i = 1, ..., n$. H is a symmetric positive definite $d \times d$ matrix called the bandwidth. The scaled and the unscaled kernels are related by

$K_H(\mathbf{x}) = |H|^{-\frac{1}{2}} K(H^{-1/2}\mathbf{x})$ and K is a symmetric probability density function.

A bandwidth matrix includes all simpler cases as special cases. Epanechnikov [5] proposed a multivariate kernel density estimate in which different bandwidths were suggested for each of the coordinate directions. An equal bandwidth in all direction, suggested by [2], as in (1.1) corresponds to $H = h^2 I_d$, where I_d denotes the $d \times d$ identity matrix.

Using the parameterization ($H = h^2 I_d$), [2] gave the multivariate kernel density estimator as

$$\hat{f}(\mathbf{x}; H) = \frac{1}{nh^d} \sum_{i=1}^{n} K(\mathbf{x} - X_i) \tag{1.2}$$

To use the parameterization ($H = h^2 I_d$) effectively the components of the data vector should be commensurate. This, according to [4, 14, 16], can be achieved if the data are transformed. If this is done, there will be no need to use a more complicated form of the kernel density estimator rather than the form involving a single bandwidth as in (1.2).

Corresponding authors: Oyegue, Tel. +2348036799995

Since the works of [6, 10, 11] have shown that better rates of convergence can be obtained when higher order forms of the univariate kernel density estimator are considered, then the issue of higher order forms (as it affects the rates of convergence) in multivariate kernel density estimator becomes the issues we tackled in this paper. Section 2 describes the optimal bandwidth (h_{AMISE}) and asymptotic mean integrated squared error (AMISE) for the second order multivariate case. Section 3 is dedicated to the higher order generalized optimal bandwidth and the asymptotic mean integrated squared error, while section 4 contains the conclusion.

1. The optimal bandwidth and the generalized AMISE for the second order multivariate kernels.

One of the most used measures on the global accuracy of (1.2) is the mean integrated squared error (MISE). The derivation of the multivariate MISE is analogous to the one-dimensional case. The MISE is given as

$$MISE\ \hat{f}(\mathbf{x};H)=E\left[\int\left(\hat{f}(\mathbf{x};H)-f(\mathbf{x};H)\right)^2 d\mathbf{x}\right] \tag{2.1}$$

$$=\int bias^2\left(\hat{f}(\mathbf{x};H)\right)d\mathbf{x}+\int \operatorname{var}\left(\hat{f}(\mathbf{x};H)\right)d\mathbf{x}$$

Since the MISE does not have a closed form, except if the target density f is a normal mixture and the choice of K is also normal, see [15], finding H_{AMISE} is generally very difficult. However, an easy way out is to find a tractable approximation to the MISE. The Asymptotic Mean Integrated Squared Error (AMISE) can be calculated as

$$AMISE\ \hat{f}(\mathbf{x};H)=AISB\hat{f}(\mathbf{x};H)+AIV\left(\hat{f}(\mathbf{x};H)\right) \tag{2.2}$$

Another problem of MISE and AMISE is that they depend on the bandwidth in a very complicated way thereby making it difficult to interpret the influence of the bandwidth on the performance of the kernel density estimation. These approximations have very simple expressions that allow a deeper appreciation of the role of the bandwidth, this is visible in the variance-bias trade-off. In addition to the aforementioned is that they can also be used to obtain the rate of convergence in kernel density estimation and MISE-Optimal bandwidth.

An asymptotic approximation of MISE through the multivariate Taylor series expansion of f around $\mathbf{x}$ and using the assumptions

i. $\int K(\mathbf{w})d\mathbf{w}=1$

ii. $\int w_i K(\mathbf{w})d\mathbf{w}=0$ and

iii. $$\int w_i w_j K(\mathbf{w})d\mathbf{w}=\begin{cases}\mu_2(K), & i=j\\ 0, & i\neq j\end{cases}\qquad \mu_2<\infty \tag{2.3}$$

$i,\ j=1,2,...,\ d$

yields a simpler version of MISE for the multivariate kernel as

$$AMISE\ \left(\hat{f}(\mathbf{x};H)\right)=\frac{R(K)}{n|H|^{1/2}}+\frac{1}{4}\mu_2(K)^2\int tr^2\left(HD^2 f(\mathbf{x})\right)d\mathbf{x} \tag{2.4}$$

where $R(K)=\int K^2(\mathbf{w})d\mathbf{w}$ and $D^2 f(\mathbf{x})$ is the Hessian matrix with respect to $\mathbf{x}$.

Now, in the case $H=h^2 I_d$, equation (2.4) becomes

$$AMISE\ \left(\hat{f}(\mathbf{x};H)\right)=\frac{R(K)}{nh^4}+\frac{1}{4}\mu_2(K)^2 h^4\int\left(\nabla^2 f(\mathbf{x})\right)d\mathbf{x} \tag{2.5}$$

The equation (2.5) is called the asymptotic MISE since it provides a useful large sample approximation to the MISE.

By minimizing (2.5) over h, and solving for h, [15] obtained the asymptotic optimal bandwidth and the minimum AMISE of order *2* respectively as:

$$h_{opt}=\left[\frac{dn^{-1}R(K)}{\mu_2(K)^2\int\left(\nabla^2 f(x)\right)^2 d\mathbf{x}}\right]^{\frac{1}{d+4}}, \tag{2.6}$$

and

$$AMISE\ \left(\hat{f}(\mathbf{x};H)\right)=\left(\frac{d+4}{4d}\right)\left[\mu_2(K)^{2d}\left(dR(K)^4\right)\left(\int\left(\nabla^2 f(\mathbf{x})\right)^2 d\mathbf{x}\right)n^{-4}\right]^{\frac{1}{d+4}} \tag{2.7}$$

Observe that the rates of convergence is of order $n^{-\frac{1}{d+4}}$ for the optimal bandwidth and $n^{-\frac{d}{d+4}}$ for the minimum AMISE.

In the next section , we shall generalize (2.6) and (2.7) which is an improvement to the theorem given by [15].

2. Theorem: Assuming that the multivariate kernels satisfy the following moment conditions

i. $\int_{R^d} K(\mathbf{w})d\mathbf{w} = 1$

ii. $\int_{R^d} \mathbf{w}K(\mathbf{w})d\mathbf{w} = \int_{R^d} \mathbf{w}\mathbf{w}^T K(\mathbf{w})d\mathbf{w} = \int_{R^d} \left(\mathbf{w}\mathbf{w}^T\right)^{m-1}\mathbf{w}^T K(\mathbf{w})d\mathbf{w} = O_d$ and (3.0)

iii. $\int_{R^d} \left(\mathbf{w}\mathbf{w}^T\right)^m K(\mathbf{w})d\mathbf{w} = \mu_{2m} I_d$, for $m = 1,2,... < \infty$.

where $\mu_{2m}(K)$ is the $(2m)^{th}$ central moment of *k* (that is $\mu_{2m}(K) = \int w_i^{2m} K(\mathbf{w})d\mathbf{w}$), and suppose f is differentiable and that the multivariate kernel is of order 2m, then for a general multivariate kernel estimator (1.1) parameterized by $H = h^2 I_d$, the optimal bandwidth and the global error is given respectively as;

(a) $$h_{AMISE} = \left[\left(\frac{((2m)!)^2}{4m}\right)^{\frac{1}{4m+d}} d^{\frac{1}{4m+d}} R(K)^{\frac{1}{4m+d}} \mu_{2m}(K)^{-\frac{2}{4m+d}} \left(\int \left(\nabla^2 f(\mathbf{x})\right)^{2m} d\mathbf{x}\right) n^{-1}\right]^{\frac{1}{4m+d}}$$

(b) $$AMISE\left(\hat{f}(\mathbf{x};h)\right) = \left[\frac{d+4m}{\left(((2m)!)^d (4m)^{2m}\right)^{\frac{2m}{4m+d}}}\right]\left(dR(K)^{4m} \mu_{2m}(K)^{2d} \left(\int \left(\nabla^2 f(\mathbf{x})\right)^{2m} d\mathbf{x}\right)^d n^{-4m}\right)^{\frac{1}{4m+d}}$$

where $m = 1,2,..., < \infty$.

Proof:

Recall that;

$$E\left(\hat{f}(\mathbf{x};H)\right) = \int_{R^d} K_H(\mathbf{x}-\mathbf{y}) f(\mathbf{y})d\mathbf{y}$$
$$= \int_{R^d} K(\mathbf{w}) f\left(\mathbf{x} - H^{\frac{1}{2}}\mathbf{w}\right)d\mathbf{w} \qquad (3.1)$$

By conditions (i) – (iii) in (3.0) and taking the higher Taylor expansion of $f\left(\mathbf{x} - H^{1/2}\right)$ in (3.1) and then substituting into (2.2), we obtained the asymptotic integrated squared bias (AISB) as:

$$AISB\left(\hat{f}(\mathbf{x};H)\right) = \frac{1}{((2m)!)^2} \mu_{2m}(K)^2 \int tr \left(\left(H^{1/2} D^2 f(\mathbf{x}) H^{1/2}\right)^{2m}\right) d\mathbf{x} \qquad (3.2)$$

Also, the asymptotic integrated variance (AIV) for this case is

$$AIV\left(\hat{f}(\mathbf{x};H)\right) = \frac{R(K)}{n|H|^{1/2}} \qquad (3.3)$$

The combination of (3.2) and (3.3) gives

$$AISB\left(\hat{f}(\mathbf{x};H)\right) = \frac{R(K)}{n|H|^{1/2}} + \frac{1}{((2m)!)^2} \mu_{2m}(K)^2 \int_{R^d} tr\left(H^{1/2} D^2 f(\mathbf{x}) H^{1/2}\right)^{2m} d\mathbf{x}$$, and since

$H = h^2 I_d$, we obtain,

$$AMISE\left(\hat{f}(\mathbf{x};H)\right) = \frac{R(K)}{nh^d} + \frac{1}{((2m)!)^2} \mu_{2m}(K)^2 h^{4m} \int_{R^d} \left(\nabla^2 f(\mathbf{x})\right)^{2m} d\mathbf{x} \qquad (3.4)$$

Hence the approximately higher-order optimal bandwidth, in the sense of minimizing equation (3.4), becomes

$$\frac{4m}{((2m)!)^2} h^{4m-1} h^{d+1} \mu_{2m}(K)^2 \int_{R^d} \left(\nabla^2 f(\mathbf{x})\right)^{2m} d\mathbf{x} = dn^{-1} R(K)$$

or

$$h_{opt} = \left(\frac{((2m)!)^2}{4m}\right)^{\frac{1}{4m+d}} d^{\frac{1}{4m+d}} R(K)^{\frac{1}{4m+d}} \mu_{2m}(K)^{\frac{2}{4m+d}} \left(\int_{R^d} \left(\nabla^2 f(\mathbf{x})\right)^{2m} d\mathbf{x}\right)^{-\frac{1}{4m+d}} n^{-\frac{1}{4m+d}} \qquad (3.5)$$

Substituting (3.5) into (3.4) and with some algebraic manipulation, we obtained

$$AMISE\left(\hat{f}(\mathbf{x};h)\right)=\left[\frac{d+4m}{\left(\left(((2m)!)^2\right)^d(4m)^{2m}\right)^{\frac{2}{4m+d}}}\right]\left(\begin{array}{l}(dR(K))^{4m}\mu_{2m}(K)^{2d}\times\\ \left(\int_{R^d}\left(\nabla^2 f(\mathbf{x})\right)^{2m}d\mathbf{x}\right)^d n^{-4m}\end{array}\right)^{\frac{1}{4m+d}} \quad (3.6)$$

Equation (3.5) and (3.6) are the generalized asymptotic optimal bandwidth and asymptotic mean integrated square error respectively of order 2m, $m=1,2,...<\infty$.

4. Conclusion.

Equations (3.5) and (3.6) provide theoretical results for the generalized optimal and asymptotic mean integrated square error (AMISE) of order 2m. It is clear that these results (i.e. (3.5) and (3.6) enable one to find the optimal bandwidth and AMISE of symmetric multivariate kernels of any order. We also observe that the optimal bandwidth as shown in (3.5) converges to zero at the rate $n^{-\frac{1}{4m+d}}$ which is faster than that of order $n^{-\frac{1}{(d+4)}}$ of [15] (see, page 99 of cited reference).

Moreover, the convergence rate of $n^{-\frac{4m}{4m+d}}$ was obtained for the generalized AMISE which is faster than that of order $n^{-\frac{4}{4+d}}$ of [15] (see page 100 of cited reference). However, these rates of convergence are still very slow compared to the univariate case, see [14]. These rates become slower as the dimension increases which is as a result of the curse of dimensionality discussed in [8, 13, 15].

References.

[1] Bowman A. W. and Azzalini A. (1997). Applied Smoothing Techniques for Data Analysis: The Kernel Approach with S-plus illustration, Oxford University Press, Oxford, UK.

[2] Cacoullos, T. (1966). Estimation of a Multivariate density. Annals of Institute of Statistical Mathematics **18**, 178-189.

[3] Deheuvel, P. (1977). Estimation non parametrique de la den site par histogrammes generalises II. Publications de I institute de statisque de I' Universite de Paris **22**, 1-25.

[4] Duong, T. and Hazelton, M. L. (2003). Plug-in bandwidth matrices for bivariate kernel density estimation. Journal of Nonparametric Statistics **15**, 17-30.

[5] Epanechnikov, V. K. (1969). Nonparametric estimation of a Multivariate Probability Density. Theory of Probability and its Application **14**, 154-158.

[6] Hall, P., Sheather, S, J., Jones, M. C., and Marron, J. S. (1991). On optimal data-based bandwidth selection in kernel density estimation. Biometrika **78**, 263-269.

[8] Hardle W., Muller, M., Sperlich, S., and Werwatz, A. (2004). Nonparametric and Semi parametric models. Springer-Verlag, Berlin.

[9] Jang, W. and Loh, J.I. (2010). Density estimation for grouped data with application to line transect sampling. Annals of Applied Statistics **4**, No**2,** 893-915.

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

[11] Osemwenkhae J. E. (2003). Higher-order forms in kernel density estimation. A. Ph.D Thesis, University of Benin, Benin City Nigeria.

[12] Rosenblatt, M. (1956). Remarks on Some nonparametric estimates of a density function. Annals of Mathematical Statistics **27**, 832-837.

[13] Scott, D. W. (1992). Multivariable Density Estimation: Theory, Practice and Visualization. Wiley, New York.

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

[15] Wand, M. P. and Jones, M. C. (1995). Kernel Smoothing. Chapman and Hall, London.

[16] Wu T. J., Chen C. F. and Chen H. Y. (2007). A variable bandwidth selector in multivariate kernel density estimation. Statistics and Probability Letters **77**, 462-467.