Journal of the Nigerian Association of Mathematical Physics, Volume 9 (November 2005)

Higher dimensional kernel methods

E. J. Osemwenkhae, S. M. Ogbonmwan and J. I. Odiase
Department of Mathematics, University of Benin, Nigeria.

Abstract

The multivariate kernel density estimator (MKDE) for the analysis of data in more than one dimension is presented. This removes the cumbersome nature associated with the interpretation of multivariate results when compared with most common multivariate schemes. The effect of varying the window width in MKDE with the attendant consequence of distortion in shape especially when the window width is large and when the kernel itself does not fit into the family to which the observations are drawn is also examined.

Keywords: Scatter plot, Histogram, Kernel density, Window width, MKDE

pp 351 - 356

1.0 Introduction

Let $X_1, X_2 \ldots, X_n$ be an independent, identically distributed, real valued random sample from a random variable X, with probability density function *f*. The univariate kernel estimator for X is given as:

$$\hat{f}(x) = \frac{1}{nh}\sum_{i=1}^{n} k\left(\frac{x - \mathrm{X}_i}{h}\right) \qquad (1.1)$$

where $k(x)$ is a symmetric kernel satisfying $\int k(x)dx = 1$ and h is the window width. This estimator has found applications in several fields of human endeavours, see Fadda, et al (1998) [7] and Dinardo and Tobias (2001) [5]. Nonparametric density estimation derives its popularity from a combination of circumstances such as: the growing importance of electronic computer in statistical research, the availability of statistical packages, and the advantages of graphical presentation of information. Literally, the kernel estimator in (1.1) is a sum of 'bumps' placed at the observation X_i. The shape of the bumps is determined by the kernel function k that is used, while the window width h determines the width of the bumps and hence the smoothness of *f*. The choice of window width, *h*, is crucial in KDE, unlike the choice of the kernel *k* which is not too important except for some special cases, see Wand and Jones (1995) [25], Ogbonmwan and Osemwenkhae (1997) [13] and Osemwenkhae (2003) [15]. The univariate kernel as defined in (1.1) has received a lot of attention from statisticians though multivariate data seem to abound more in real life, see Silverman (1986) [21].

Given a multivariate data set (1.2),

$$X_i = (X_{i1}, X_{i2}, \ldots, X_{id}); \quad i = 1(1)n, \qquad (1.2)$$

our interest is to estimate the underlying density corresponding to (1.2). Several methods have been proposed in literature for estimating the density of (1.2). In this paper, the following are examined: (i) methods of estimating multivariate densities with their possible setbacks (ii) the multivariate kernel estimator and (iii) the influence of large *h* on the distribution of some common multivariate kernels.

2.0. Methods of estimating multivariate densities

The oldest well known method of estimating densities is the Histogram method. An excellent

discussion of this method can be found in the work of Louise-Adolphe Bertillon, as presented in Stigler (1986) [22]. Traditionally, the histogram has been used to provide a visual clue to the underlying distribution of *f*, see Izenman (1991) [11]. Suppose *f* has support $\Omega = [a, b]$, partition [a, b] into non-overlapping bin widths given by h_n, where $h_n = (t_{n,i+1} - t_{n,i})$ and $i = 1\ (1)m$. If $\mathrm{I}_i(.)$ is the indicator function for the i^{th} bin, then the histogram estimator is given as

$$\hat{f}(x) = \frac{1}{n h_n} \sum_{i=1}^{m} N_i \mathrm{I}_i(x) \qquad (2.1)$$

where $n = \sum_{i=1}^{m} N_i$ is the sample size, N_i is the size of the ith sample.

Basically, the choice of origin and the length of the bin h_n, affects its smoothing procedure: smaller bin width allows more detailed information about the distribution to be exposed, hence there is the tendency for the occurrence of spurious noise at the tail of the distribution. Also, the larger the bin width the smoother the curve and hence provides less details about the underlying distribution.

The histogram estimator (2.1) lacks accuracy when used in cluster analysis and nonparametric discriminant analysis, see Silverman (1986) [21], and also lacks continuities at cell boundaries when derivatives of estimates are required, see Hand (1982) [10]. Another major pitfall of the histogram estimator is that it does not allow the drawing of contour diagram in the representation of data and so it does not work well in multivariate data, see Tukey and Tukey (1981) [24]. The sensitivity of histograms shapes to the choice of origin is a more serious defect as stated in Silverman (1986) [21] and Devroye and Lugosi (1997 [3], 2001 [4]).

Another method of multivariate density method is the scatter plots. Scott et al (1978) [20] and Silverman (1986) [21] pointed out that other methods of density estimation such as the kernel methods will detect or highlight features that are not obvious in the scatter plot. In most cases, if the data set is very large, the resulting dense picture is difficult to interpret and may also be expensive in time and ink to produce the scatter plot. Scott and Thompson (1983) [19] gave a more foundational argument for using the scatter plot as simply an attempt to discern features for the underlying model of the data. The scatter plot fails in the estimation of multivariate densities.

Other methods of multivariate density estimation include the nearest neighborhood (NN), the maximum penalized likelihood (MPL) and the length biased data approach (LBDA). Fundamentally, these methods among other things, failed to be a proper *pdf*, see Silverman (1986) [21] and Patil et al (1991) [18].

3.0 The Multivariate Kernel Estimator

The mathematical tractability and wide applicability of the univariate kernel estimator are inherited by the multivariate kernel estimator, see Taylor (1989) [23] and Jones, et al (1999) [12]. Our attention is on the multivariate kernel density estimator defined by

$$\hat{f}(x) = \frac{1}{n h^d} \sum k\left\{ \frac{1}{h^d} (x - \mathrm{X}_i) \right\} \qquad (3.1)$$

where k is defined for a d-dim random variable X denoted by X_i, $i = 1(1)n$, $\int_{R^d} k(x)dx = 1$ and h is the smoothing parameter or the window width. In the univariate case, the smoothing parameter h is very crucial. The fixed univariate kernel estimator allows the use of a single window width, this is however not always true in the multivariate case where there are various options of choosing the smoothing parameter h, see Cacoullos (1966) [1], Epanechnikov (1969) [6], Deheuvel (1977) [2], Fukunaga (1972) [8] and Hall, et al (1995) [9] for possible suggestions. The work of Ogbonmwan and Osemwenkhae (2000) [14] and Osemwenkhae (2003) [15] showed that if the kernel, k in (3.1),

satisfies the following regularity conditions:

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

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

(iii) $\int_{R^d} t^{m}k(t)dt = V_m \neq 0$ for m = 1, 2, 3, …, < ∞, then the optimal *h* is given as

$$h_{opt} \approx \left\{\frac{(m!)^2}{2m}\right\}^{\frac{1}{d+4}} V_m^{-\frac{2}{d+2m}} \beta^{\frac{1}{d+4}} \left\{\int \left(\nabla^m f(x)\right)^2 dx\right\}^{-\frac{1}{d+2m}} n^{-\frac{1}{d+2m}} \qquad (3.2)$$

and the corresponding Mean Integrated Square Error (MISE) is given by

$$MISE\,\hat{f}(x) \approx \frac{2m+1}{2m}\left\{\frac{2m}{(m!)^2}\right\}^{\frac{d}{d+2m}} \left\{V_m^{\frac{2d}{d+2m}}\ \beta^{\frac{2m}{d+2m}} \left[\int \left(\nabla^m f(x)\right)^2 dx\right]^{\frac{d}{d+2m}}\right\} n^{-\frac{2m}{d+2m}} \qquad (3.3)$$

where $\beta = \int k(t)^2 dt$, $\nabla^m = \frac{\partial^m}{\partial x^m} + ... + \frac{\partial^m}{\partial x_n^m}$ and *d* is the dimension of *X*.

Equations (3.2) and (3.3) are essentially very important, since they remove the burden of calculating the value of h_{opt} for any even order of the bias when estimating the density of any multivariate kernel. The works of Osemwenkhae (2003) [15] and Osemwenkhae and Ogbonmwan (2003a, [1] b [17]) revealed that the global error resulting from (3.3) is significantly reduced for these successive higher order values of the smoothing parameter *h*.

Two common symmetric multivariate kernels of interest are the d-dim standard normal density given by

$$K_G(x) = (2\pi)^{-d/2} \exp\left(-\frac{1}{2}\mathrm{X^T X}\right), \quad X \in R^d \qquad (3.4)$$

and the d-dim Epanechnikov kernel given by

$$K_e(x) = \begin{cases} \frac{1}{2}C_d^{-1}(d+2)(1-X^T X), & if \quad X^T X \leq 1 \\ 0 \quad otherwise \end{cases} \qquad (3.5)$$

where in (3.5), C_d is the volume of the unit *d*-dim sphere.

We shall examine the benefits that exist in using the MKDE and specifically the kernels in (3.4), (3.5) and similar ones in the next two sections. Precisely, the shape of k(x) and its inherent analyticity is inherited by $\hat{f}(x)$ of (3.1). Furthermore, since each of these kernels in (3.4) and (3.5) are *pdf*s, then the estimate constructed by this method will also be a proper *pdf*.

3.1 **Example**

As an illustration, let us consider the 2-dim kernels in (3.4) and (3.5) as our test kernels which for *d* = 2 respectively become

$$K_G(x_1, x_2) = \frac{1}{2\pi} e^{-\frac{1}{2}\left(x_1^2 + x_2^2\right)} \qquad (3.6)$$

and $$K_e(x_1, x_2) = \begin{cases} \frac{2}{\pi}\left[1 - \left(x_1^2 + x_2^2\right)\right], & if \;\; x_1^2 + x_2^2 \le 1 \\ 0 \quad otherwise \end{cases} \tag{3.7}$$

The value of $\int \left(\nabla^2 f(x)\right)^2 dx$ of (3.2) and (3.3) is $\frac{1}{2\pi}$ if m =2: this is useful in estimating the value of the optimal window width h. If f is the 2-dim normal density, (3.2) becomes

$$h_{opt} \approx \left[4\pi\beta V_2^{-2}\right]^{1/6} n^{-1/6} = 1.77 n^{-1/6} \tag{3.8}$$

However, if the kernel of choice is the d-dim Epanechinkov kernel in (3.7), this kernel is equal to zero in the area $x_1^2 + x_2^2 > 1$ and only observations that fall into the area

$$\left\{(x,\, y) : (x - \mathrm{X}_i)^2 + (y - \mathrm{Y}_i)^2 \le h^2\right\}$$

will influence the probability density function. If we allow the smoothing parameter h to be a single value in both coordinate directions, and apply it on simulated bivariate normal distribution with E(x) = 0 and Var(x) = 2, the density of $\hat{f}(x)$ is obtained. In particular, if k is the kernel in (3.7) and the values of h chosen subjectively as h = 0.3, 0.5, 2.9, and 3.9, the graphs shown in Figures 1a – d are obtained. Similarly, if the kernel of our choice is the 2-dim standard normal in (3.6), then (3.2) reduces to

$$h_{opt} \approx 0.9635 n^{-1/6} \tag{3.9}$$

Applying (3.9) on the simulated bivariate normal distribution, we obtain Figures 2a – d.

Figure 1a – d: 2-dim Epanechnikov kernel for different values of h

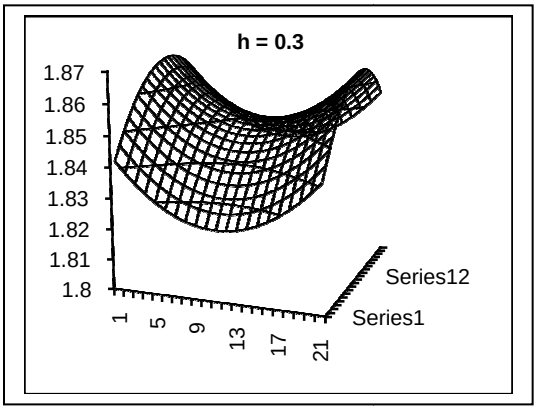

Figure 1a

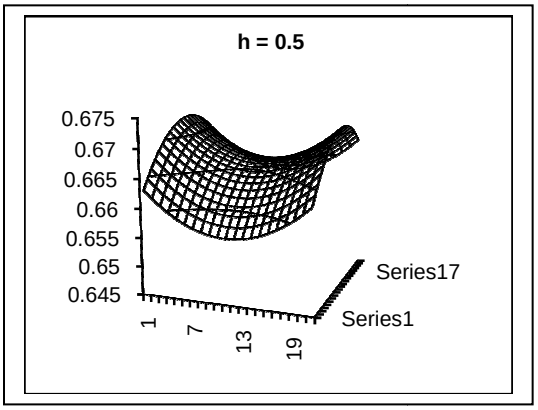

Figure 1b

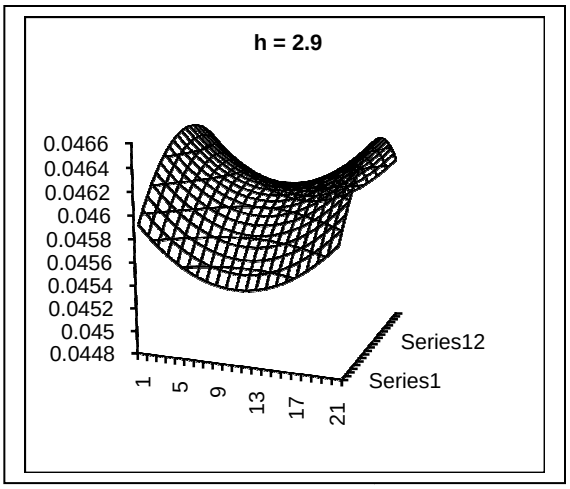

Figure 1c

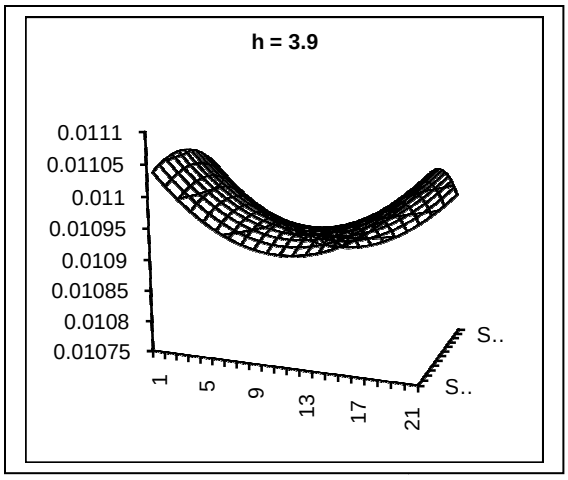

Figure 1d

Figure 2a - d: 2-dim Normal kernel for different values of *h*

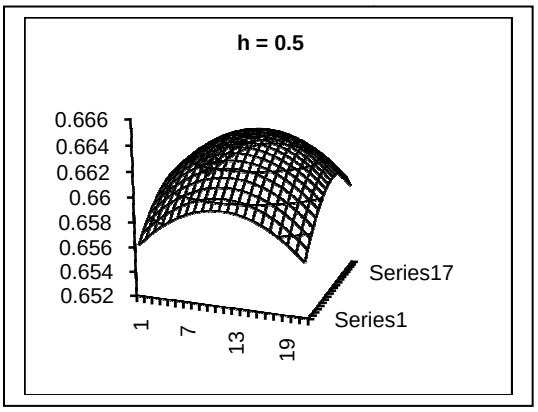

Figure 2a

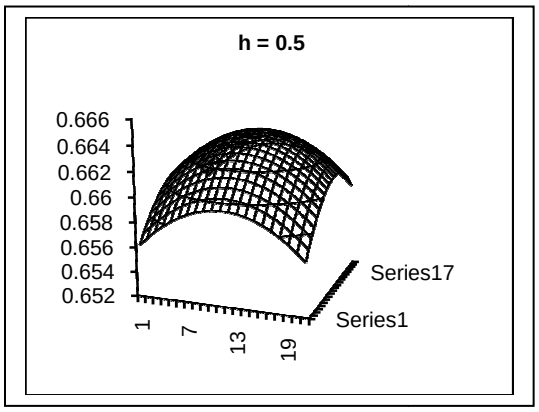

Figure 2b

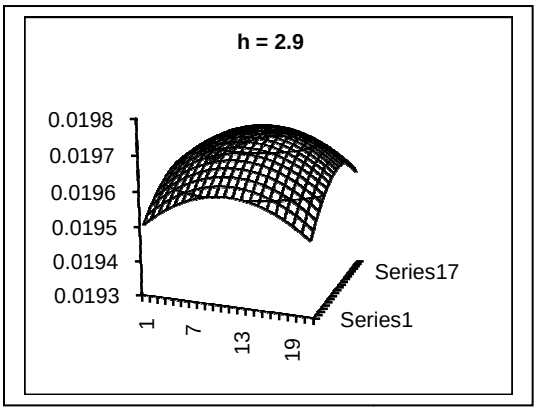

Figure 2c

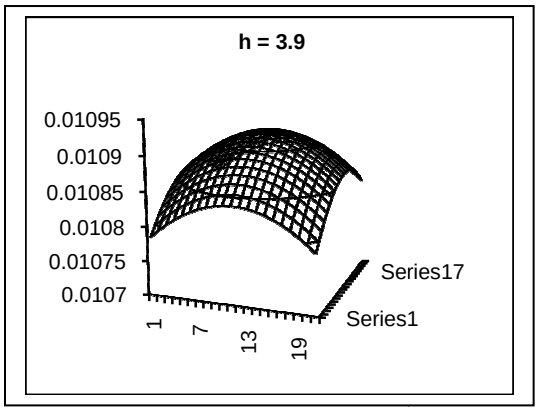

Figure 2d

4.0 Discussion of findings

The multivariate histogram does not allow for the drawing of curves because of discontinuous boundaries but the multivariate kernel density estimation permits the drawing of curves. While the scatter plot is only a pointer to the density of *f*, the multivariate KDE removes the cumbersome nature of the interpretation of results associated with multivariate scatter plots. The fixed value of *h* in the estimation of the density of *f* gives the distribution a ragged nature (see Figures 1c and 1d), although the adaptive schemes tend to handle some of these lapses, with itself failing to be a proper pdf.

Another observation is that when the optimal window width h is small as in the Epanechnikov kernel ($h = 0.3$ and 0.5), the structure of the densities obtained reflects the underlying density of the data set used. When *h* was increased in this kernel, the shape of the distribution was lost. This is however not true if *h* increases when the 2-dim normal density was used. This affirms that the normal density approximates most distributions especially when *n* is large for any dimension of X.

5.0 Conclusion

Much work has been done in univariate KDE, but the multivariate KDE still needs to be exploited. Clearly, the choice of *h* greatly affects the distribution of $\hat{f}$. We have also shown the preference of the multivariate KDE to other familiar methods of multivariate density estimation. In conclusion, there is a great benefit from multivariate KDE considering the way it handles the estimation of densities in more than one dimension.

References

[1] Cacoullos, T. (1966). Estimation of a multivariate density. *Annals of the Inst. of Stat. Maths*., 18, 179-189.

[2] Deheuvel, P. (1977). Estimation nonparametric de le densite par histogrammes generalized II. *Publications de l'institut statistique de l'universite de Paris*, 22, 1-23., p.4

[3] Devroye, L. and Lugosi, G. (1997). Nonasymptotic universal smoothing factors, kernel complexity and yatracos classes. *Annals of Statistics*, 25, 2626-2637.

[4] Devroye, L. and Lugosi, G. (2001). *Combinatorial methods in density estimation*. Springer-Verlag, New York.

[5] Dinardo, J. and Tobias, J. L. (2001). Nonparametric density and regression estimation. *Journal of Economic Perspectives*, 15, 11 – 28.

[6] Epanechnikov, V. A. (1969). Nonparametric estimation of a multivariate probability density. *Theory of Probability and its Application*, 14, 153-158.

[7] Fadda, D. Slezak, E. and Bijaoui, A. (1998). Density estimation with nonparametric methods. *Astronomy and Astrophysics Supplement Series*, 127, 335 – 352.

[8] Fukunaga, K. (1972). *Introduction to statistical pattern recognition*. New York Academic press.

[9] Hall, P., Hu, C. T. and Marron, J. S. (1995). Improval variable window kernel estimates of probability densities. *Annals of Statistics*, 23, 1-10.

[10] Hand, D. (1982).Kernel Discriminant analysis. *Research Studies*, New York.

[11] Izenman, A. J. (1991). Recent developments in nonparametric density estimation. *Journal of Amer. Stat. Ass.*, 86, 205-221.

[12] Jones M. C., Signorini , D. F. and Hjort, N. L. (1999) On multivariate bias correction in kernel density estimation. *Sankya*, A, 61, 422 – 430.

[13] Ogbonmwan S. M. and Osemwenkhae (1997). On the choice of kernel density estimation. *Journal of Nig. Stat. Ass*, 11, 1 – 12.

[14] Ogbonmwan, S. M. and Osemwenkhae, J. E. (2000). Higher order forms for optimal window width in multivariate kernel density estimation. *Journal of Nig. Ass. of Mathematical Physics*, 4, 327 – 333.

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

[16] Osemwenkhae, J. E. and Ogbonmwan, S. M (2003b). Global errors in symmetric kernels. *ABACUS (Journal of Mathematical Ass. of Nig.)*, 30, 2A, 130 – 139.

[17] Osemwenkhae, J. E. and Ogbonmwan, S. M. (2003a). Generalized efficiencies for higher order symmetric univariate kernel. *Journal of the Nig. Ass. of Math. Physics*, 7, 89 – 96.

[18] Patil, P. N., Wells, M. T. and Marron, J. S. (1991). Kernel based estimations of ratio functions. *Journal of Nonparametric Statistics*. 1, 223 – 231.

[19] Scott, D. W. and Thompson, J. R. (1983). Probability density estimation in higher dimensions. In Gentle, J. E. (ed), Comp. Sci. and Stat. Proceedings of the fifteenth Symposium in the Interface. Amsterdam: North – Holland, pp 173 – 179.

[20] Scott, D. W., Gotto, A M., Cole, J. S. and Gorry, G. A. (1978). Plasma lipids as collateral risk factors in coronary heart disease – a study of 371 males with chest pain. *J. Chronic Diseases*, 31, 337 – 345.

[21] Silverman, B. W. (1986). *Density estimation for statistics and data analysis*. London Chapman and Hall.

[22] Stigler, S. M. (1986). *The History of Statistics. The uncertainty before 1900*. Belknap Press of Harvard University Press.

[23] Taylor, C. C. (1989). Bootstrap choice of the smoothing parameter in kernel density estimation. *Biometrika*, 76, 705 – 712.

[24] Tukey, P. A. and Tukey, J. W. (1981). *Graphical display of data sets in 3 or more dimensions. Interpreting Multivariate Data*. Chichester: Wiley 189 – 275.

[25] Wand, M. P. and Jones, M. C. (1995). *Kernel smoothing*. London: Chapman & Hall.