Journal of the Nigerian Association of Mathematical Physics
***Volume* 24 (July, 2013), pp 193 – 198**
© J. of NAMP

Smoothing Parameter Choices in Multivariate Kernel Density estimates

Siloko I. U. and Ishiekwene C. C.

Mathematics Department,
University of Benin, Nigeria.

Abstract

The choice of an ideal or appropriate smoothing parameter in multivariate kernel density estimation is considered by looking at six different methods. A comparative study of these six methods is done using simulated data. The higher-order smoothing parameter choice is seen to have behaved better than the others with the smallest bias value, mean integrated squared error (MISE) and also gave the best 3-D graph.

Keywords: smoothing parameter, multivariate, kernel density estimation, bias, mean integrated squared error.

1.0 Introduction

Consider a d-dimension random vector $X = (X_1, X_2, ..., X_d)$ where $X_1, X_2, ..., X_d$ are one-dimensional random variables. If we have n observations for each of the d random variables, $X_1, X_2, ..., X_d$, then if we collect the i^{th} observation of each of the d random variables in the vector X_i,

$$X_i = \begin{pmatrix} X_{i1} \\ X_{i2} \\ \vdots \\ X_{id} \end{pmatrix}, \; i = 1, 2, ..., n$$

where X_{ij} is the ith observation of the random variable X_j. The probability density of $X = (X_1, X_2, ..., X_d)^T$ which is the joint pdf *f* of the random variables $X_1, X_2, ..., X_d$ is

$$f(X) = f(x_1, x_2, ..., x_d) \quad (1)$$

From the univariate case, we can see that

$$\hat{f}_h(X) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h^d}K\left(\frac{x - X_i}{h}\right) \quad (2)$$

i.e.

$$\hat{f}_h(X) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h^d}K\left(\frac{x_1 - X_{i1}}{h}, \frac{x_2 - X_{i2}}{h}, ..., \frac{x_d - X_{id}}{h}\right) \quad (3)$$

where K denotes a multivariate kernel function operating on the d arguments [1]. Also, equation (3) assumes that the smoothing parameter *h* is the same for each component. If we relax this assumption then we have a vector of smoothing parameters $h = (h_1, h_2, ..., h_d)^T$ and the multivariate kernel density estimator becomes

$$\hat{f}_h(x) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h_1 h_2 ... h_d}K\left(\frac{x_1 - X_{i1}}{h_1}, \frac{x_2 - X_{i2}}{h_2}, ..., \frac{x_d - X_{id}}{h_d}\right) \quad (4)$$

Corresponding author: ***Ishiekwene C. C.,*** E-mail: cycigar@yahoo.com, Tel.: +2348023348430

The multidimensional kernel $K(U) = K(U_1)K(U_2)\ldots K(U_d)$ (5)

where each $k_i(.)$ for i = 1,2,…,d denotes a univariate kernel function [2].

A general approach is to use a smoothing parameter matrix H (non singular), thus given the multivariate density estimator as

$$\hat{f}_h(X) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{\det(H)}K\left(H^{-1}(x - X_i)\right) = \frac{1}{n}\sum_{i=1}^{n}K_H(x - X_i) \quad (6)$$

The Bivariate case is considered in this work and an attempt is made at proffering a 'best' method of choosing the smoothing parameter – h [3, 4].

2.0 Methods

Six different methods of choosing the smoothing parameter are considered. These includes reference to a standard rule, Silverman's rule of thumb, biased cross validation, unbiased cross validation, data driven and higher-order method [5].

Asymptotic approximations ofequation (4) yield integrated squared bias (ISB) and integrated variance (IV) given by:

$$IV = \frac{(R(K))^d}{nh_1\ldots h_d} \quad (7)$$

and

$$ISB = \frac{1}{4}h_i^{\,4}\sigma_k^4\int tr^2\left[\nabla^2 f(x)\right]dx \quad i = 1,2,\ldots,d \quad (8)$$

Where

$R(K) = \int k^2(\mathbf{x})dx,$ tr indicates the trace of a matrix and $\nabla^2 f(x)$ is the Hessian (Matrix of second partial derivatives) of f . Combining these terms yields an estimate of the mean integrated squared error,

$$MISE = E\int\left\{\hat{f}(x) - f(x)\right\}^2 dx \quad (9)$$

$$= \frac{(R(K))^d}{nh_1\ldots h_d} + \frac{1}{4}h_i^{\,4}\sigma_k^4\int tr^2\left\{\nabla^2 f(x)\right\}dx \quad (10)$$

This MISE is used for all the methods except for the higher order smoothing parameter choice [6, 7].
The optimal bandwidth can then be easily derived and is equal to:

$$h_{AMISE} = \left[\frac{(R(K))^d}{\sigma_k^4\int tr^2\left\{\nabla^2 f(\mathbf{x})\right\}d\mathbf{x}}\right]^{\frac{1}{(d+4)}} n^{-\frac{1}{(d+4)}} \quad (11)$$

This choice of h yields an $MISE\ of\ order\ O\left(n^{-4/(d+4)}\right)$. This is the data-driven choice of the smoothing parameter.
The normal reference rule bandwidth that minimizes the MISE can be approximated by:

$$h_i = \sigma_i\left\{\frac{4}{(d+2)n}\right\}^{1/(d+4)} \qquad for\ i = 1,2,\ldots,d \quad (12)$$

Where σ_i is the standard deviation of the ith variate and d is the dimension of the kernel.

Silverman's rule of thumb [3] is computed as:

$$h = 0.9\min\left[\hat{\sigma}, (Q_s - Q_1)/(1.34)\right]n^{-\frac{1}{5}} \quad (13)$$

Where Q_3 and Q_1 are the third and first sample quartiles respectively.
The multivariate least squares (unbiased) cross validation function $UCV(h_1, h_2, \ldots, h_d)$ is given by:

$$UCV(h_1,h_2,...,h_d)=\frac{1}{\left(2\sqrt{\pi}\right)^d nh_1,...,h_d}+\frac{1}{\left(2\sqrt{\pi}\right)^d n^2h_1,...,d_d}\times$$

$$\sum_{i=1}^{n}\sum_{j\neq i}\exp\left\{-\frac{1}{4}\sum_{k=1}^{d}\Delta_{ijk}^2\right\}-\left(2\times 2^d\right)\exp\left\{-\frac{1}{2}\sum_{k=1}^{d}\Delta_{ijk}^2\right\} \quad (14)$$

Where

$$\Delta_{ijk}=\frac{X_{ik}-X_{jk}}{h_k} \quad (15)$$

The general multivariate form of the biased cross validation is given by:

$$BCV(h_1,h_2,...,h_d)=\frac{1}{\left(2\sqrt{\pi}\right)^d nh_1,...,h_d}+\frac{1}{4n(n-1)h_1,...,h_d}$$

$$\times\sum_{i=1}^{n}\sum_{j\neq i}\left[\left\{\sum_{k=1}^{d}\Delta_{ijk}^2\right\}-(2d+4)\left\{\sum_{k=1}^{d}\Delta_{ijk}^2\right\}+\left(d^2+2d\right)\right]\prod_{k=1}^{d}\phi(\Delta_{ijk}) \quad (16)$$

Where ϕ is the standard normal density and

$$\Delta_{ijk}=\frac{X_{ik}-X_{jk}}{h_k}$$

The generalised expression for the optimal bandwidth of the higher order kernel is of the form:

$$h_{opt}=\left\{\left(\frac{m!}{2m}\right)\times\left(\frac{\pi^{3/2}}{2^{3/2}\Gamma\left(\frac{2m+1}{2}\right)}\right)\right\}^{\left(\frac{1}{2m+1}\right)}\sigma_i n^{-\left(\frac{1}{2m+1}\right)} \quad (17)$$

Where σ_i is the standard deviation of the ithvariate.

The corresponding MISE for the above expression is of the form:

$$MISE=\frac{1}{\left(2\sqrt{\pi}\right)^d n^{h_1},...,n^{d_d}}+\frac{1}{(m!)^2}\left[\frac{(m!)^2}{2}\times\frac{\pi^{\frac{3}{2}}}{2^{\frac{m}{2}}\Gamma\left(\frac{2m+1}{2}\right)\left(\Gamma\left(\frac{m+1}{2}\right)\right)^2}\right]^{\frac{2m}{2m+1}}$$

$$\times\sigma_i^{2m}n^{-\frac{2m}{2m+1}}\left\{\frac{2^{\frac{m}{2}}}{\left(\sqrt{\pi}\right)^d}\Gamma\left(\frac{m+1}{2}\right)\right\}^2\times\frac{1}{2\pi\sigma_i^{(2m+1)}}\Gamma\left(\frac{2m+1}{2}\right) \quad (18)$$

3.0 Simulation, Result And Discussion

A set of Bivariate normal data is simulated using MATLAB. The data has mean $\mu=0$ and variance $\sigma^2=1$. The sample sizes for x and y are equal and sufficiently large, ie, 50 (see Table 1).

Table 1:

x	1.5044	1.5809	0.9267	1.5872	1.3512	0.9019	1.0539	1.2794	1.6243	1.6305
y	90.5599	95.4444	95.1467	89.1876	93.2502	98.8329	91.3387	94.3017	89.9281	96.3103
x	0.9524	1.6353	1.6240	1.2277	1.4922	0.9392	1.1743	1.5892	1.4855	1.6260
y	90.3067	93.3421	95.6788	97.9999	98.8274	93.8413	88.8974	89.0265	90.3358	97.0728
x	1.3708	0.8500	1.5333	1.6046	1.3901	1.4565	1.4442	1.1495	1.3706	0.9638
y	90.1667	98.4641	91.4548	89.5988	90.2581	94.6741	92.9468	91.4751	97.2730	94.3017
x	1.4131	0.8467	1.0526	0.8588	0.9016	1.5117	1.4037	1.0864	1.6182	0.8489
y	93.8717	98.3180	90.6787	96.3821	96.3401	91.8234	94.0906	88.1378	87.8728	93.6427
x	1.1885	1.1405	1.4630	1.4880	0.9770	1.2314	1.1943	1.3629	1.4159	1.4539
y	96.6479	98.5215	88.7919	94.1028	92.8996	87.3640	91.2992	87.0214	90.4642	89.2241

The measure of discrepancy used is the Mean Integrated Squared Error (MISE). It is the sum of the bias squared and the variance (see Table 2).

Also, the 3-D graphs for all the methods are plotted for the different choices of the smoothing parameter (see Figures 1-6).

Table 2: Showing the various methods with their corresponding h, bias, variance and MISE.

METHODS.	H_x	H_y	VARIANCE.	$BIAS^2$	MISE.
DATA DRIVEN.	.572348	.564833	.0154664	.0522734	.0677398
REFERENCE RULE.	.135404	1.17733	.0313646	.480408	.511773
SILVERMAN'S RULE.	.106965	1.40086	.0333683	.962795	.996163
UNBIASED CROSS VALIDATION.	.765399	.539662	.0121049	.107005	.119110
BIASED CROSS VALIDATION.	.890731	.549232	.0102204	.180121	.190341
HIGHER ORDER(FOUR).	.269231	3.52594	.00167656	.0013147	.00299126

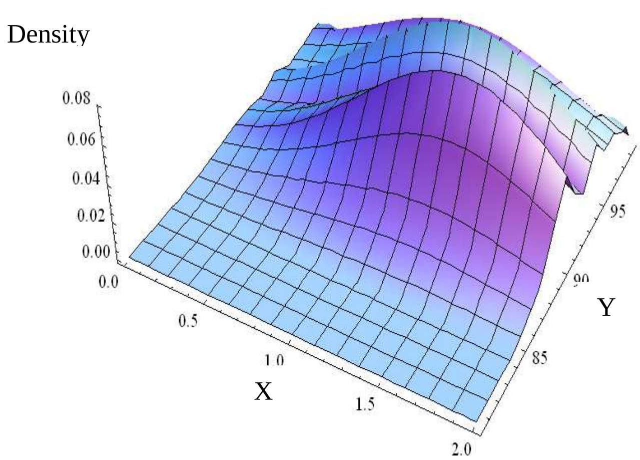

Fig 1: kernel estimate showing the data driven Smoothing parameters.

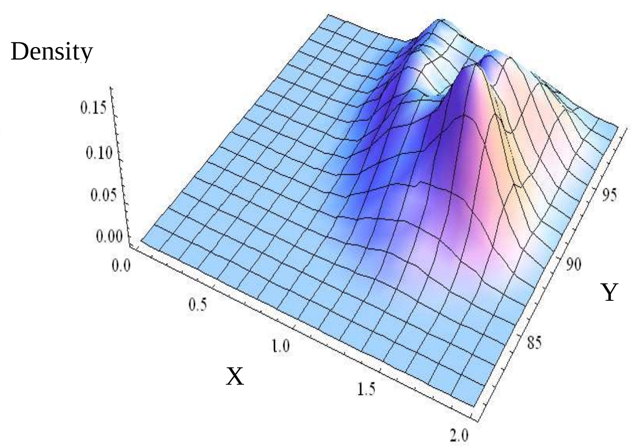

Fig 2: kernel estimate showing reference rule smoothing parameters.

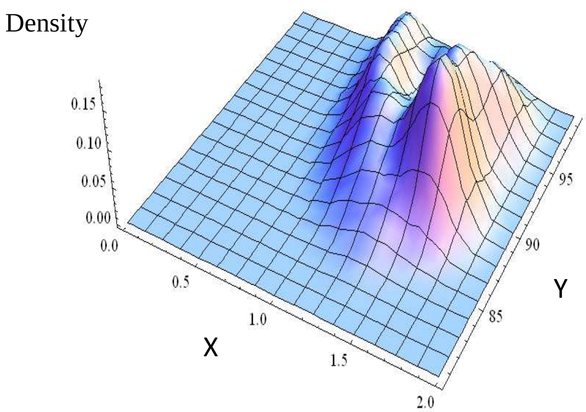

Fig 3: kernel estimate showing Silverman's rule of smoothing parameter.

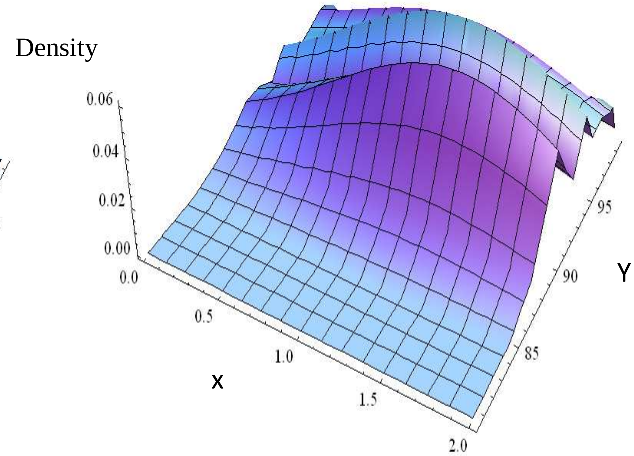

Fig 4: kernel estimate showing the unbiased cross validation smoothing parameter.

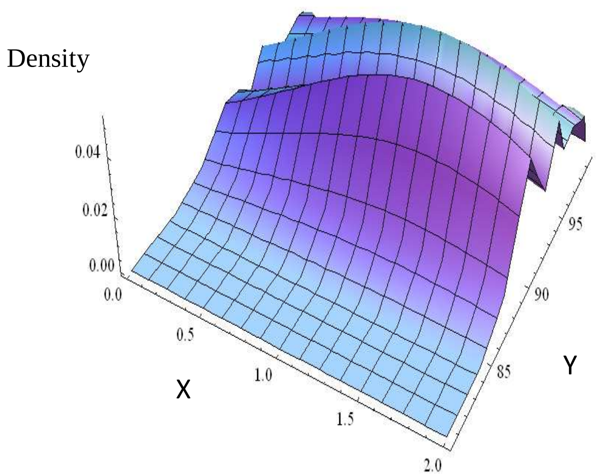

Fig 5: kernel estimate showing the biased cross validation smoothing parameter.

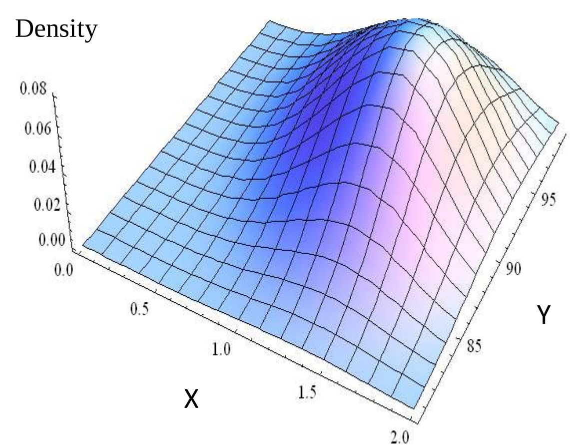

Fig 6: kernel estimate showing the higher order(four) smoothing parameter.

1.0 Conclusion

It is clearly seen from Table 2 that the higher-order choice of the smoothing parameter gave the least MISE and also gave the best 3-D graph. We thus conclude that it should be used in place of all others when dealing with multivariate kernel density estimation.

References

[1] Schucany, W.R. and Sommers, J.P. (1977) – Improvement of kernel-type density estimators. Journal of American Statistical Assoc. **72**, pg 420 -423.

[2] Wand, M.P. and Jones, M.C. (1995) – Kernel Smoothing. Chapman and Hall/CRC, London.

[3] Silverman, B. W. (1986). Density Estimation For Statistics and Data Analysis. London: Chapman and Hall.

[4] Afere, B.A.E (2010). A family of Multivariate Higher Order Hybrid Polynomial Kernels in Kernel Density Estimation. An unpublished Ph.D thesis submitted to the School of Postgraduate Studies, University of Benin, Benin City, Nigeria.

[5] Siloko, I.U. (2012)- Choice of smoothing parameters in multivariate Kernel Density estimation. An unpublished Master's thesis submitted to the School of Postgraduate Studies, University of Benin, Benin City, Nigeria.

[6] Ishiekwene, C.C and Afere, B.A.E. (2001) – Higher Order Window Width Selectors for Empirical Data. Journal of Nigerian Statistical Association (J.N.S.A), **14**, 69 – 82.

[7] Ishiekwene, C. C. and Osemwenkhae, J. E. (2006) – A comparison of fourth order window width selectors in Kernel Density Estimation (A Univarate case), ABACUS, **33**; 14 - 20.

[8] Bartlett, M.S. (1963) – Statistical estimation of density functions. Sankhyā series A, **25**, pg 245 – 254.

[9] Birke, Melanie (2009) – Shape constrained KDE. Journal of Statistical planning & inference, vol **139,** issue 8 , August 2009, pg 2851 – 2862.

[10] Parzen, E. (1962) – On the estimation of a probability function and the nodes. Annals of Mathematical Statistics. **33**,pg 1065 – 1076.