HANDLING THE CURSE OF DIMENSIONALITY IN MULTIVARIATE KERNEL DENSITY ESTIMATION

By

JORDAN JIMMY CRABBE

Bachelor of Science in Mathematics
Kwame Nkrumah University of Science and Technology
Kumasi, Ghana
2001

Master of Science in Statistics
University of Nevada, Las Vegas
Las Vegas, Nevada
2007

Submitted to the Faculty of the
Graduate College of the
Oklahoma State University
in partial fulfillment of
the requirements for
the Degree of
DOCTOR OF PHILOSOPHY
July, 2013

HANDLING THE CURSE OF DIMENSIONALITY IN

MULTIVARIATE KERNEL DENSITY ESTIMATION

Dissertation Approved:

Dr. Ibrahim Abe Ahmad

Dissertation Adviser

Dr. Mark Payton

Dr. Lan Zhu

Dr. David Carter

Outside Committee Member

ACKNOWLEDGEMENTS

This dissertation has come this far not by my singular effort but also with the help, guidance, support and contributions from other people. I will therefore take this opportunity to acknowledge the input of those people.

I would like to express my profound gratitude to my advisor, Dr. Ibrahim Abe Ahmad for his encouragement, direction, guidance, constructive criticisms and patience throughout the duration of this dissertation that help in refining and executing the dissertation to this style. He has been my mentor and his excellent teaching and research skills have been a tremendous help to me. He has instilled in me the qualities of a good scientist and an excellent researcher.

I also wish to express my heartfelt gratitude to my responsible and respectable committee members, Dr. Mark Payton, Dr. Lan Zhu and Dr. David Carter for their time, dedication, guidance and direction. Their contributions were invaluable.

I would like to express my deepest gratitude to all the faculty team at the Department of Statistics at Oklahoma State University for the knowledge imparted on me. The benefits under their outfit were priceless and I really appreciate that.

I will not have said it all if I do not render special thanks to my parents Mr. and Mrs. Samuel H. Crabbe, my wife Abigail Crabbe, my daughters Jodrell and Samuela and my siblings Dora, Victor and Flora for their love, support and encouragement throughout the duration of my studies. They have been part of what has brought me this far.

Acknowledgements reflect the views of the author and are not endorsed by committee members or Oklahoma State University.

Name: JORDAN JIMMY CRABBE

Date of Degree: JULY, 2013

Title of Study: HANDLING THE CURSE OF DIMENSIONALITY IN MULTIVARIATE KERNEL DENSITY ESTIMATION

Major Field: STATISTICS

Kernel density estimation (KDE) is the most widely-used practical method for accurate nonparametric density estimation. Many works had been done on both the univariate and multivariate cases showing the efficacy, practicality and applicability of this method. Despite the fact that multivariate kernel density estimation is an important technique in multivariate data analysis and has a wide range of applications, its performance worsens exponentially with high dimensional data sets, this phenomenon is called "curse of dimensionality", where there is exponential growth in combinatorial optimization as the dimension of the data set increases. Scott and Wand (1991) demonstrated a progressive deterioration of the multivariate kernel density estimation as the dimension p increases by showing that an increase in sample size is required to attain an equivalent amount of accuracy.

This work proposes a new multivariate kernel density estimation approach which is based on the sample means. The method has the characteristic that it works for self-revolving densities or the ellipsoidally symmetric distributions. It also works for spherical distributions since they can be transformed to ellipsoidally symmetric distributions by undergoing an affine transformation. The univariate normal, multivariate normal and the Cauchy distributions, just to mention a few, are some of the distributions that possess this self-revolving or the ellipsoidally symmetric property. In addition, this work also proposes another new multivariate kernel density estimate which handles the curse of dimensionality better.

We applied this new method to the probability density function, the distribution function and nonparametric multivariate regression. In all these cases, our multivariate kernel density estimation approach which is based on the sample means performs better than the regular multivariate kernel density estimation based on the sample data. We also observed that the proposed multivariate kernel density method breaks the "curse of dimensionality" and remedy the deficiency of high dimensional bandwidth selection. Besides, its performance is consistent in most of the bandwidth selection methodologies. The second proposed new multivariate density estimate does not completely breaks the curse of dimensionality but the effect of the curse on it is minimal as compared to the regular multivariate kernel density estimate.

TABLE OF CONTENTS

Chapter Page

I. INTRODUCTION AND LITERATURE REVIEW 1

1.1 Introduction..1
1.2 Multivariate Kernel Density Estimation 2
1.3 Limitations 3
1.4 Multivariate Density estimation Based on Sample Means 4
1.4.1 Ellipsoidal Symmetric Distributions..5
1.4.2 The Proposed Estimate 6

II. BASIC RESULTS..8

2.1 Multivariate Density Estimation 8
2.1.1 Expectation and Bias of the Estimate 9
2.1.2 An Important Special Case 11
2.1.3 Variance and AMISE of Estimate..12
2.2 Curse of Dimensionality ....……..16
2.3 Bivariate Product Kernel Case………………………………………………17
2.4 Multivariate Distribution Function Estimation Based on Sample Means……29

III. ESTIMATION OF NONPARAMETRIC REGRESSION
FUNCTION…………………………………………………………………36

3.1 Introduction..36
3.1.1 Regression Estimation 37
3.2 Properties of the Estimator...41
3.2.1 Derivation of the Expectation of the Estimator 41
3.2.2 Derivation of Variance and *AMSE* of the Estimator...............................47

Chapter Page

IV. BANDWIDTH SELECTION METHODOLOGIES 63

4.1 Cross-Validation for Density Estimation Based on the Sample Means-Univariate Case……………………………………………………………63
4.1.1 Least Squares (Unbiased) Cross-Validation (UCV) ....…65
4.1.2 Biased Cross-Validation (BCV) Based in the Univariate Case.............68
4.2 Cross-Validation of Multivariate Densities Based on the Sample Means…...72
4.2.1 Least Squares (Unbiased) Cross-Validation (UCV)...........................…72
4.2.2 Biased Cross Validation (BCV) 79
4.3 Plug-In Techniques………………………………………………………..82
4.3.1 Plug-In Bandwidth Selection-Univariate Case 82
4.3.2 Plug-In Bandwidth Matrix Selection for Bivariate Kernel Density Estimation……………………………………………………………96

V. MULTIVARIATE KERNEL DENSITY ESTIMATION-NEW APPROACH ...105

5.1 Introduction...105
5.2 Bivariate Case 105
5.2.1 Expectation and Bias of the Estimate 106
5.2.2 Variance and AMISE of the Estimate 108
5.3 Multivariate Product Kernel Case...116
5.3.1 Expectation and Bias of the Estimate 116
5.3.2 Variance and AMISE of the Estimate..118
5.4 Multivariate Case-General 120
5.4.1 Expectation and Bias of the Estimate 121
5.4.2 Variance and AMISE of the Estimate..123
5.4.3 An Important Special Case 125

VI. CONCLUSIONS 129

VII. FUTURE WORK 132

REFERENCES 134

LIST OF TABLES

Table Page

2.3.1 Mean Squared Error (MSE) of the Regular Multivariate Kernel Density Estimate and the Multivariate Kernel Density Estimate Based on the Sample Mean With Dimension *p=2*……………………………………….…...27

2.3.2 Mean Squared Error (MSE) of the Standard Multivariate Kernel Density Estimate and the Multivariate Kernel Density Estimate Based on the Sample Mean With Dimension *p=3*…………………………………………28

2.3.3 Mean Squared Error (MSE) of the Standard Multivariate Kernel Density Estimate and the Multivariate Kernel Density Estimate Based on the Sample Mean With Dimension *p=4*…………………………….……...……....28

LIST OF FIGURES

Figure Page

2.3.1 The Bivariate Normal Density Distribution ……………………………….....24

2.3.2 Contour of the Bivariate Normal Density…………………………..…………24

2.3.3 Regular Kernel Estimate of the Bivariate Normal Distribution………………......25

2.3.4 Contour of the Regular Kernel Estimate of the Bivariate Normal Density……...25

2.3.5 The Kernel Based on the on the Sample Means Estimate of the Bivariate Normal Distribution…………………………………………………………………....26

2.3.6 Contour of the Kernel Based on the Sample Means Estimate of the Bivariate Normal Density...26

4.1.1 Plots Comparing the Least Squares Cross Validation (LSCV) of the Univariate Standard Kernel Estimate Case and Univariate Kernel Estimate Based on the Sample Mean……………………………………………………68

4.1.2 Plots Comparing the Biased Cross Validation (BCV) of the Univariate Standard Kernel Estimate Case and Univariate Kernel Estimate Based on the Sample Mean………………………………………………………………….71

4.2.1 Plots of the Unbiased Cross-Validation of the Multivariate Kernel Estimate Based on the Sample Means With Different Sample Sizes……………………...77

4.2.2 Plots of the Unbiased Cross-Validation of the Multivariate Standard Kernel Estimate at Different Sample Sizes………………………………………………78

4.2.3 Plots Comparing the Biased Cross Validation (BCV) of the Multivariate Standard Kernel Estimate and the Multivariate Kernel Estimate Based on the Sample Means………………………………………………...…..…81

4.3.1 Plots Comparing the Performance of the Plug-In Technique in the Univariate Case for Both Standard Kernel Estimate and the Kernel Estimate Based on the Sample Means…………………………………………………………..95

5.2.1 The Bivariate Normal Density Distribution …………………...........................113

5.2.2 Contour of the Bivariate Normal Density………………………………………113

5.2.3 Regular Kernel Estimate of the Bivariate Normal Distribution………………...114

5.2.4 Contour of the Regular Kernel Estimate of the Bivariate Normal Density…….114

5.2.5 The New Kernel Estimate of the Bivariate Normal Distribution.......................115

5.2.6 Contour of the New Kernel Estimate of the Bivariate Normal Distribution…...115

CHAPTER I

INTRODUCTION AND LITERATURE REVIEW

1.1 Introduction and Historical Background

In the past few decades, one nonparametric density estimation method known as kernel density estimation (KDE) had become a renowned method of density estimation in statistics, economics and other areas where this idea is applicable. KDE is now one of the most popular methods for estimating the underlying probability density function (PDF) based on a data set. It is also the most widely-used practical method for accurate nonparametric density estimation. Since 1951, a lot of work had been done on both the univariate and multivariate cases showing the efficacy, practicality, and applicability of this method.

Kernel density estimation was originated by Fix and Hodges (1951) and Rosenblatt (1956). Fix and Hodges (1951) were concerned with density estimation in connection with nonparametric discrimination. In his fundamental paper, Rosenblatt (1956) gave a full demonstration of the idea of nonparametric estimation of density function. Fix and Hodges (1951) and Rosenblatt (1956) both defined the kernel density estimator as:

$$\hat{f}(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right), \tag{1.1}$$

where $X_1, X_2, ..., X_n$ is a random sample from a continuous density function $f(x)$, K is a known second order symmetric PDF referred to as the kernel and h is the bandwidth(a known sequence of constants) which depends on n such that $h \to 0$ and $nh \to \infty$ as $n \to \infty$. The choice of the bandwidth h, is well-known to be crucial and of great importance since it controls the smoothness of the estimator $\hat{f}(x)$. Choosing the bandwidth, h, involves a trade-off between the variance and the bias of the estimate. Epanechnikov (1969) demonstrated that the choice of the kernel K is not very crucial in density estimation. What both originators proposed was a univariate kernel density estimator. A couple of error criteria were used to assess the performance of this estimator. Scott and Wand (1991) used the mean absolute error (MAE) and mean integrated absolute error (MIAE) to assess the performance of this estimator. Ahmad and Amezziane (2012), among others, used the mean square error (MSE) and mean integrated square error (MISE) to ascertain this assessment.

1.2 Multivariate Kernel Density Estimation

In recent times, there have been improvements which led to the extension of the univariate kernel density estimation to the multivariate case. Specifically, based on the extensive research carried out in the last three decades, multivariate kernel density estimation has reached a level of maturity comparable to their univariate counterparts. Multivariate kernel density is an important technique in multivariate data analysis and has a wide range of applications.

Ahmad and Amezziane (2012) proposed the multivariate kernel density estimator of the PDF $f(\mathbf{x})$ as

$$\hat{f}(\boldsymbol{x}) = n^{-1}\sum_{i=1}^{n} K_{\boldsymbol{H}}(\boldsymbol{x} - \boldsymbol{X}_i), \qquad \boldsymbol{x} \in IR^p \tag{1.2}$$

where $\boldsymbol{H}$, the bandwidth matrix, is a symmetric positive definite $p \times p$ matrix that verifies the following usual two conditions:

$H \underset{n\to\infty}{\rightarrow} \boldsymbol{0}_p$, where $\boldsymbol{0}_p$ is the null $(p \times p)$ matrix and $n^{-1}|\boldsymbol{H}|^{-1/2} \underset{n\to\infty}{\longrightarrow} \boldsymbol{0}_p$,

$K(\cdot)$, the kernel, is a multivariate density function that satisfy the two moment conditions

$$\int_{\Re^p} \mathbf{z}K(\mathbf{z})\,d\mathbf{z} = \mathbf{0}_p$$
$$\int_{\Re^p} \mathbf{z}\mathbf{z}^T K(\mathbf{z})\,d\mathbf{z} = \mathbf{I}_p \tag{1.3}$$

and for every function $\boldsymbol{g}$, they defined:

$$\boldsymbol{g}_{\boldsymbol{H}}(\boldsymbol{x}) = |\boldsymbol{H}|^{-1/2}\,\boldsymbol{g}(\boldsymbol{H}^{-1/2}\boldsymbol{x}).$$

Ahmad and Amezziane (2012) used the MSE and the MISE to measure the performance of the above estimator. They also provided a data-based method to evaluate the bandwidth matrix.

1.3 Limitations

Even though multivariate kernel density estimation is an important technique in multivariate data analysis and has a wide range of applications, its long standing worst-case theoretical results showing that its performance worsens exponentially with the dimension of the data have suppressed its applications to modern high-dimensional data sets for decades. This phenomenon is what Bellman (1961) called the "curse of dimensionality", where he describes the exponential growth in combinatorial optimization as the dimension increases.

Scott and Wand (1991) also demonstrated a progressive deterioration of multivariate kernel density estimation as the dimension p increases by showing that an increase in sample size is required to attain an equivalent amount of accuracy. Ahmad and Amezziane (2012) attempted to address the "curse of dimensionality" by using a weighted version of the MSE or MISE with weights depending on the sample size n.

1.4 Mainstream: Multivariate Density Estimation Based on Sample Means

In all the papers presented on both univariate and multivariate kernel density estimation, a sample from the underlying PDF is used to estimate it. Therefore, the estimator is always a function of the sample from its PDF.

The work at hand develops a new method of kernel density estimation called "kernel density estimation based on the sample means." This work is found as a new modification of the regular kernel density estimation in multivariate case that addresses the curse of dimensionality and it uses average rather than individual data units.

This method has the characteristic that it works for self-revolving densities or the ellipsoidally symmetric distributions. Applying this method of kernel density estimation to the self-revolving densities or the ellipsoidally symmetric distributions, the curse of dimensionality is alleviated and a remedy is found for the deficiency of high dimension bandwidth selection. This is the main thrust of this work.

1.4.1 Ellipsoidally Symmetric Distributions

If $\boldsymbol{Z}$ is a random vector whose distribution is spherical about the origin then a random vector $\boldsymbol{X}$ which is representable as the image of $\boldsymbol{Z}$ under affine transformation is said to have an ellipsoidally symmetric distribution. Affine transformation between two vector spaces consists of a transformation followed by a translation.

Ellipsoidal distributions are characterized by the fact that there exists a linear transformation of the variables that results in a spherically symmetric distribution for the transformed variables. An ellipsoidal distribution is fully specified by (i) its center of symmetry, (ii) its inner product defined by the linear transformation to sphericity and (iii) the distribution on the radii of concentric hyperspheres on which there is uniform probability density. Such multivariate distributions play an important role in the theory of matching because the symmetry allows general results to be obtained.

The model of ellipsoidal symmetry is a useful generalization of multivariate normality. The statistical model most frequently assumed in multivariate analysis is the normal distribution. A notable feature of the $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ density is the property that its constant surfaces are ellipsoids centered at $\boldsymbol{\mu}$ with orientation and shape determined by the matrix $\boldsymbol{\Sigma}$. This ellipsoidal symmetry plays an important role in the geometrical interpretation of normal-model multivariate analysis. Even without normality, ellipsoidal symmetry of the data distribution can provide a rationale for the use of standard multivariate procedures (Dempster (1969)).

Recent interest in robust statistical methods has led to more detailed consideration of statistical models which retain some of the features of the normal model while providing flexibility in data-fitting. In a p-dimensional multivariate setting, a corresponding generalization of the normal model is the model of the ellipsoidal symmetry: each observation has density of the form $\left[\det \mathbf{A}\right]^{-1} h\left[\mathbf{A}^{-1}(x-\boldsymbol{\mu})\right]$, where $\boldsymbol{\mu}$ is a $p\times 1$ vector, $\boldsymbol{A}$ is a $p\times p$ nonsingular matrix, and h is a density on R^p which is a spherically symmetric about the origin.

A special case of the multivariate t distribution when the common denominator has 1 degree of freedom is a multivariate Cauchy distribution. Ferguson (1967) has shown that this multivariate Cauchy is characterized by the fact that any linear function of $Y_1,...,Y_m$ has a (univariate) Cauchy distribution. The multivariate Cauchy and multivariate normal distributions belong to the class of spherical distributions. Therefore under an affine transformation, they become ellipsoidally symmetric distributions.

Lemma 1.1: If f is ellipsoidally symmetric distribution, then $f_{\bar{X}}(\boldsymbol{x}) = n^{p/2} f(\sqrt{n}\boldsymbol{x})$.

1.4.2 The Proposed Estimate

Let $\bar{\boldsymbol{X}} = (\bar{X}_1, \bar{X}_2, ..., \bar{X}_p)'$ be the means of random samples from a population with unknown density, $\boldsymbol{f}(\boldsymbol{x})$ assumed to be ellipsoidally symmetric, then we propose the following estimate of $f(\boldsymbol{x})$ based on the sample means:

$$\hat{f}_{\bar{X}}(\boldsymbol{x}) = \frac{1}{n^{p/2}|\boldsymbol{H}|^{1/2}} K_1\left[\left(\frac{\boldsymbol{x}-\sqrt{n}\bar{\boldsymbol{X}}}{n^{1/2}}\right)(\boldsymbol{H})^{-1/2}\right], \tag{1.4}$$

where $\boldsymbol{H}$ is a positive-definite smoothing matrix and p is the dimension of the $\bar{\boldsymbol{X}}$ and K_1 is the kernel which is assumed to be symmetric. Under the moment condition that $\int_{R^p} K_1(\boldsymbol{u})d\boldsymbol{u} = 1$, then our proposed estimator is a density as shown below:

$$\int \hat{f}_{\overline{X}}(\boldsymbol{x})d\boldsymbol{x} = \frac{1}{n^{p/2}|\boldsymbol{H}|^{1/2}} \int K_1\left[\left(\frac{\boldsymbol{x}-\sqrt{n}\overline{\boldsymbol{X}}}{n^{1/2}}\right)(\boldsymbol{H})^{-1/2}\right]d\boldsymbol{x}.$$

Let $u = \frac{\boldsymbol{x}-\sqrt{n}\overline{\boldsymbol{X}}}{n^{1/2}\boldsymbol{H}^{1/2}}$, then $\boldsymbol{x} = \sqrt{n}\overline{\boldsymbol{X}} + n^{1/2}\boldsymbol{H}^{1/2}\boldsymbol{u}$ and therefore $d\boldsymbol{x} = n^{p/2}|\boldsymbol{H}|^{1/2}d\boldsymbol{u}$.

So, $\int \hat{f}_{\overline{X}}(\boldsymbol{x})d\boldsymbol{x} = \frac{1}{n^{p/2}|\boldsymbol{H}|^{1/2}} \int K_1(\boldsymbol{u})n^{p/2}|\boldsymbol{H}|^{1/2}d\boldsymbol{u} = \int K_1(\boldsymbol{u})d\boldsymbol{u} = 1.$

Since our model works for ellipsoidal symmetric distributions, the above estimator is a very good estimate which yields a trade-off between the bias and the variance to produce a good optimum bandwidth.

Applications of estimate (1.4) are given in the case of estimating the distribution function as well as the regression function in Chapters 2 and 3 to follow.

CHAPTER II

BASIC RESULTS

2.1 Multivariate Kernel Density Estimation

Let $\bar{\boldsymbol{X}} = (\bar{X}_1, \bar{X}_2, \ldots, \bar{X}_p)'$ be the means of random samples from a population with unknown density, $f(\boldsymbol{x})$ assumed to be ellipsoidally symmetric, then the estimate of $f(\boldsymbol{x})$ based on the sample means is given by

$$\hat{f}_{\bar{X}}(\boldsymbol{x}) = \frac{1}{n^{p/2}|\boldsymbol{H}|^{1/2}} K_1\left[\left(\frac{\boldsymbol{x} - \sqrt{n}\bar{\boldsymbol{X}}}{n^{1/2}}\right)(\boldsymbol{H})^{-1/2}\right], \tag{2.1}$$

where $\boldsymbol{H}$ is positive-definite smoothing matrix and K_1 is a symmetric kernel.

In this chapter, we develop some basic properties of the estimate (1.4). We concentrate on the mean square and the mean integrated square errors.

2.1.1 Expectation and Bias of $\hat{f}_{\bar{X}}(\boldsymbol{x})$

$$E\left[\hat{f}_{\bar{X}}(\boldsymbol{x})\right] = \frac{1}{n^{p/2}|\boldsymbol{H}|^{1/2}} \int K_1\left[\left(\frac{\boldsymbol{x}-\sqrt{n}\boldsymbol{y}}{n^{1/2}}\right)(H)^{-1/2}\right] f_{\bar{X}}(\boldsymbol{y})\, d\boldsymbol{y}, \tag{2.2}$$

where $d\boldsymbol{y} = dy_1, ..., dy_p$.

Since $f_{\bar{X}}(\boldsymbol{y}) = n^{p/2} f(\sqrt{n}\boldsymbol{y})$ we get that,

$$E\left[\hat{f}_{\bar{X}}(\boldsymbol{x})\right] = \frac{1}{n^{p/2}|\boldsymbol{H}|^{1/2}} \int K_1\left[\left(\frac{\boldsymbol{x}-\sqrt{n}\boldsymbol{y}}{n^{1/2}}\right)(\boldsymbol{H})^{-1/2}\right] n^{p/2} f(\sqrt{n}\boldsymbol{y})\, d\boldsymbol{y}. \tag{2.3}$$

Let $\sqrt{n}\boldsymbol{y} = \boldsymbol{w},$ then $\frac{d\boldsymbol{w}}{d\boldsymbol{y}} = n^{p/2} \cdot \boldsymbol{1}$. Hence $d\boldsymbol{y} = \frac{1}{n^{p/2}} d\boldsymbol{w}$ and then

$$E\left[\hat{f}_{\bar{X}}(\boldsymbol{x})\right] = \frac{1}{n^{p/2}|\boldsymbol{H}|^{1/2}} \int K_1\left[\left(\frac{\boldsymbol{x}-\sqrt{n}\boldsymbol{y}}{n^{1/2}}\right)(\boldsymbol{H})^{-1/2}\right] f(\boldsymbol{w})\, d\boldsymbol{w}. \tag{2.4}$$

Now, let $\boldsymbol{u} = \frac{(\boldsymbol{x}-\boldsymbol{w})(\boldsymbol{H})^{-1/2}}{n^{1/2}}$, then $n^{1/2}(\boldsymbol{H})^{1/2}\boldsymbol{u} = \boldsymbol{x}-\boldsymbol{w}$, which implies that

$\boldsymbol{w} = \boldsymbol{x} - n^{1/2}(\boldsymbol{H})^{1/2}\boldsymbol{u}$ and $|d\boldsymbol{w}| = n^{p/2}|\boldsymbol{H}|^{1/2} d\boldsymbol{u}$. Thus

$$E\left[\hat{f}_{\bar{X}}(\boldsymbol{x})\right] = \frac{1}{n^{p/2}|\boldsymbol{H}|^{1/2}} \int K_1(\boldsymbol{u}) f(\boldsymbol{x} - n^{1/2}\boldsymbol{H}^{1/2}\boldsymbol{u})(n^{p/2}|\boldsymbol{H}|^{1/2})\, d\boldsymbol{u}$$

$$= \int K_1(\boldsymbol{u}) f(\boldsymbol{x} - n^{1/2}\boldsymbol{H}^{1/2}\boldsymbol{u})\, d\boldsymbol{u}. \tag{2.5}$$

Now applying the Taylor series expansion to second order, we have

$$E\left[\hat{f}_{\bar{X}}(\boldsymbol{x})\right] \approx \int K_1(\boldsymbol{u})\left[f(\boldsymbol{x}) - n^{1/2}\boldsymbol{H}^{1/2}\nabla f(\boldsymbol{x})\boldsymbol{u}^T + \frac{1}{2}n\boldsymbol{u}^{\boldsymbol{T}}\left(\boldsymbol{H}^{1/2}\right)^{\boldsymbol{T}}\nabla^2 f(\boldsymbol{x})\boldsymbol{H}^{1/2}\boldsymbol{u}\right]d\boldsymbol{u}.$$

Hence

$$\begin{aligned} E\left[\hat{f}_{\bar{X}}(\boldsymbol{x})\right] \approx\ & f(\boldsymbol{x})\int K_1(\boldsymbol{u})d\boldsymbol{u} - n^{1/2}\boldsymbol{H}^{1/2}\nabla f(\boldsymbol{x})\int \boldsymbol{u}^T K_1(\boldsymbol{u})d\boldsymbol{u} \\ & + \frac{1}{2}n\left(\boldsymbol{H}^{1/2}\right)^{\boldsymbol{T}}\nabla^2 f(\boldsymbol{x})\boldsymbol{H}^{1/2}\int \boldsymbol{u}\boldsymbol{u}^{\boldsymbol{T}} K_1(\boldsymbol{u})d\boldsymbol{u}. \end{aligned} \tag{2.6}$$

Note that the multivariate kernel customary assumes the three moment's conditions:

(i) $\int_{R^p} K_1(\boldsymbol{u})d\boldsymbol{u} = 1$, a multivariate density

(ii) $\int_{R^p} \boldsymbol{u}K_1(\boldsymbol{u})d\boldsymbol{u} = \boldsymbol{0}$, the means of the marginal kernels are zero

(iii) $\int_{R^p} \boldsymbol{u}\boldsymbol{u}^T K_1(\boldsymbol{u})d\boldsymbol{u} = \boldsymbol{I}_p$, this means the marginal kernels are all pairwise uncorrelated

and that each has a unit variance.

Therefore,

$$E\left[\hat{f}_{\bar{X}}(\boldsymbol{x})\right] \approx f(\boldsymbol{x}) + \frac{n}{2}tr\left[\left(\boldsymbol{H}^{1/2}\right)^{\boldsymbol{T}}\nabla^2 f(\boldsymbol{x})\boldsymbol{H}^{1/2}\int \boldsymbol{u}\boldsymbol{u}^T K_1(\boldsymbol{u})d\boldsymbol{u}\right]. \tag{2.7}$$

Since the quadratic form in the equation (2.7) is a determinant matrix which is equal to its trace, by applying the moment conditions, the covariance matrix of K_1 is $\boldsymbol{I}_d$ and the integral factor within the trace vanishes, so we have

$$E\left[\hat{f}_{\bar{X}}(\boldsymbol{x})\right] \approx f(\boldsymbol{x}) + \frac{n}{2}tr\left[\left(\boldsymbol{H}^{1/2}\right)^{\boldsymbol{T}}\nabla^2 f(\boldsymbol{x})\boldsymbol{H}^{1/2}\right]. \tag{2.8}$$

The expectation of $\hat{f}_{\bar{X}}(\boldsymbol{x})$ will equal $f(\boldsymbol{x})$ to the order $O\left(tr\left((\boldsymbol{H}^{1/2})^T(\boldsymbol{H}^{1/2})\right)\right)$.

From equation (2.8), the bias of $\hat{f}_{\bar{X}}(\boldsymbol{x})$ is given by

$$Bias\left[\hat{f}_{\bar{X}}(\boldsymbol{x})\right] \approx \frac{n}{2} tr\left((\boldsymbol{H}^{1/2})^T \nabla^2 f(\boldsymbol{x})(\boldsymbol{H}^{1/2})\right)$$

$$= \frac{n}{2} tr\left((\boldsymbol{H}^{1/2})(\boldsymbol{H}^{1/2})^T \nabla^2 f(\boldsymbol{x})\right), \tag{2.9}$$

by applying a property of trace. The asymptotic integrated square bias (AISB) is obtained by squaring the right-hand side of (2.9) and integrating out $\boldsymbol{x}$, that is

$$AISB = \frac{1}{4} n^2 \int \left[tr\left\{ (\boldsymbol{H}^{1/2})(\boldsymbol{H}^{1/2})^T \nabla^2 f(\boldsymbol{x}) \right\} \right]^2 d\boldsymbol{x}. \tag{2.10}$$

2.1.2 An Important Special Case

Now, let's define a scalar $h > 0$ and a $p \times p$ matrix $\boldsymbol{Q}$ such that $\boldsymbol{H} = h^2 \boldsymbol{Q}$, where $|\boldsymbol{Q}| = 1$.

The idea behind choosing the matrix $\boldsymbol{Q}$ to have a unit determinant is that the elliptical shape of the kernel is controlled by the matrix $\boldsymbol{QQ}^T$ and the size of this kernel is governed by the scalar h, which is the bandwidth and which controls the amount of smoothness in the estimator for a given sample size n, where $h = (h_1, h_2, ..., h_p)^{1/p}$ is the geometric mean of the smoothing parameters. Hence

$$Bias\left[\hat{f}_{\bar{X}}(\boldsymbol{x})\right] \approx \frac{1}{2} n h^2 tr\left\{ \boldsymbol{Q}^{1/2} (\boldsymbol{Q}^{1/2})^T \nabla^2 f(\boldsymbol{x}) \right\}. \tag{2.11}$$

Therefore, the asymptotic integrated squared bias (AISB) is given by

$$AISB = \frac{1}{4} n^2 h^4 \int \left[tr\left\{ Q^{1/2} \left(Q^{1/2}\right)^T \nabla^2 f(\boldsymbol{x}) \right\} \right]^2 d\boldsymbol{x} \ . \tag{2.12}$$

2.1.3 Variance and AMISE of $\hat{f}_{\bar{\boldsymbol{X}}}(\boldsymbol{x})$

Squaring equation (2.1) yields

$$\hat{f}^2_{\bar{\boldsymbol{X}}}(\boldsymbol{x}) = \frac{1}{n^p|\boldsymbol{H}|} K_1^2\left[\left(\frac{\boldsymbol{x}-\sqrt{n}\bar{\boldsymbol{X}}}{n^{1/2}}\right)\left(\boldsymbol{H}^{-1/2}\right)\right]. \tag{2.13}$$

Again, assuming $f_{\bar{\boldsymbol{X}}}(\boldsymbol{y}) = n^{p/2} f(\sqrt{n}\boldsymbol{y})$,

$$E\left[\hat{f}^2_{\bar{\boldsymbol{X}}}(\boldsymbol{x})\right] = \int \frac{1}{n^p|\boldsymbol{H}|} K_1^2\left[\left(\frac{\boldsymbol{x}-\sqrt{n}\boldsymbol{y}}{n^{1/2}}\right)\left(\boldsymbol{H}^{-1/2}\right)\right] n^{p/2} f(\sqrt{n}\boldsymbol{y})\, d\boldsymbol{y} .$$

Let $\sqrt{n}\boldsymbol{y} = \boldsymbol{w},$ then $\frac{d\boldsymbol{w}}{d\boldsymbol{y}} = n^{p/2}$, $d\boldsymbol{y} = \frac{d\boldsymbol{w}}{n^{p/2}}$

$$E\left[\hat{f}^2_{\bar{\boldsymbol{X}}}(\boldsymbol{x})\right] = \int \frac{1}{n^p|\boldsymbol{H}|} K_1^2\left[\left(\frac{\boldsymbol{x}-\sqrt{n}\boldsymbol{y}}{n^{1/2}}\right)\left(\boldsymbol{H}^{-1/2}\right)\right] f(\boldsymbol{w})\, d\boldsymbol{w} . \tag{2.14}$$

Let $\boldsymbol{u} = \frac{(\boldsymbol{x}-\boldsymbol{w})\left(\boldsymbol{H}\right)^{-1/2}}{n^{1/2}}$, then $\boldsymbol{w} = \boldsymbol{x} - n^{1/2}\left(\boldsymbol{H}\right)^{1/2}\boldsymbol{u}$, which implies $|d\boldsymbol{w}| = n^{p/2}|\boldsymbol{H}|^{1/2} d\boldsymbol{u}$.

$$E\left[\hat{f}^2_{\bar{\boldsymbol{X}}}(\boldsymbol{x})\right] = \frac{1}{n^p|\boldsymbol{H}|} \int K_1^2(\boldsymbol{u}) f\left(\boldsymbol{x} - n^{1/2}\boldsymbol{H}^{1/2}\boldsymbol{u}\right) n^{p/2} |\boldsymbol{H}|^{1/2} d\boldsymbol{u}$$

$$= \int \frac{1}{n^{p/2}|\boldsymbol{H}|^{1/2}} K_1^2(\boldsymbol{u}) f(\boldsymbol{x} - n^{1/2}\boldsymbol{H}^{1/2}\boldsymbol{u}) d\boldsymbol{u}. \tag{2.15}$$

Applying Taylor series expansion to the second order and the moments conditions,

$$E\left[\hat{f}_{\bar{x}}^2(\boldsymbol{x})\right] \approx \frac{1}{n^{p/2}|\boldsymbol{H}|^{1/2}}\left[f(\boldsymbol{x})R(K_1)\right], \tag{2.16}$$

where $R(K_1) = \int K_1^2(\boldsymbol{u}) d\boldsymbol{u}$.

The variance term is dominated by the $E\left[\hat{f}_{\bar{x}}^2\right]$, therefore

$$Var\left[\hat{f}_{\bar{x}}(\boldsymbol{x})\right] \approx \frac{f(\boldsymbol{x})R(K_1)}{n^{p/2}|\boldsymbol{H}|^{1/2}}. \tag{2.17}$$

From equation (2.17), the asymptotic integrated variance (AIV) is given by

$$AIV = \frac{R(K_1)}{n^{p/2}|\boldsymbol{H}|^{1/2}}. \tag{2.18}$$

Therefore, the asymptotic mean squared error *(*AMISE*)* is

$$AMISE = \frac{1}{4} n^2 \int \left[tr\left\{ \boldsymbol{H}^{1/2} (\boldsymbol{H}^{1/2})^T \nabla^2 f(\boldsymbol{x}) \right\} \right]^2 d\boldsymbol{x} + \frac{R(K_1)}{n^{p/2}|\boldsymbol{H}|^{1/2}}. \tag{2.19}$$

In the above special case, where $\boldsymbol{H} = h^2\boldsymbol{Q}$, $|\boldsymbol{Q}| = 1$, we get that

$$AIV = \frac{R(K_1)}{n^{p/2}h^p},$$

since $\boldsymbol{Q}$ is a $p \times p$ matrix with unit determinant.

In this case, for a sufficiently smooth density function $f(\boldsymbol{x})$, the multivariate mean integrated square error (MISE) is asymptotically given by

$$AMISE = AIV + AISB$$

Therefore,

$$AMISE = \frac{R(K_1)}{n^{p/2}h^p} + \frac{1}{4}n^2h^4 \int \left[tr\left\{ \boldsymbol{Q}^{1/2} \left(\boldsymbol{Q}^{1/2}\right)^T \nabla^2 f(\boldsymbol{x}) \right\} \right]^2 d\boldsymbol{x} . \tag{2.20}$$

As it was said in the introduction, the choice of the bandwidth h is well-known to be crucial and of great importance since it controls the smoothness of the estimator $\hat{f}(\boldsymbol{x})$. Choosing the bandwidth h involves a trade-off between the variance and the bias of the estimate as can be seen in the AMISE above. Therefore, an intermediate value of the bandwidth must be chosen to control both the bias and the variance simultaneously and allowing the bandwidth h to slowly decrease as the sample size increases for a better performance of the estimate. In the multivariate standard case parameterized by $\boldsymbol{H} = h\boldsymbol{Q}$, where $|\boldsymbol{Q}| = 1$, the AMISE is given by

$$AMISE = \frac{R(K)}{nh^p} + \frac{1}{4}h^4 \int \left[tr\left(\boldsymbol{Q}\boldsymbol{Q}^T \nabla^2 f(\boldsymbol{x}) \right) \right]^2 d\boldsymbol{x} ,$$

where p is the dimension. Here, the bias is very small and the variance is large. The AMISE in our special case as seen in equation (2.20) to that of the special case of the standard form, we have a smaller variance and a bigger bias which is opposite of that of the standard case.

In this special case, the optimum h can be obtained by differentiating equation (2.20) and equating it to zero as shown below:

$$\frac{\partial (AMISE)}{\partial h} = \frac{-ph^{-p-1}R(K_1)}{n^{p/2}} + n^2h^3 \int \left[tr\left(\boldsymbol{Q}^{1/2} \left(\boldsymbol{Q}^{1/2}\right)^T \nabla^2 f(\boldsymbol{x}) \right) \right]^2 d\boldsymbol{x} .$$

Equating this to zero yields

$$\frac{ph^{-p-1}R(K_1)}{n^{p/2}} = n^2h^3 \int \left[tr\left\{ \boldsymbol{Q}^{1/2} \left(\boldsymbol{Q}^{1/2}\right)^T \nabla^2 f(\boldsymbol{x}) \right\} \right]^2 d\boldsymbol{x}$$

$$\frac{pR(K_1)}{n^{p/2}h^{p+1}} = n^2h^3 \int \left[tr\left\{ \boldsymbol{Q}^{1/2} \left(\boldsymbol{Q}^{1/2}\right)^T \nabla^2 f(\boldsymbol{x}) \right\} \right]^2 d\boldsymbol{x}\,.$$

Thus

$$h^{p+4} = \frac{pR(K_1)}{n^{(p+4)/2} \int \left[tr\left\{ \boldsymbol{Q}^{1/2} \left(\boldsymbol{Q}^{1/2}\right)^T \nabla^2 f(\boldsymbol{x}) \right\} \right]^2}.$$

Hence

$$h_{opt} = n^{-1/2} \left[\frac{pR(K_1)}{\int \left[tr\left\{ \boldsymbol{Q}^{1/2} \left(\boldsymbol{Q}^{1/2}\right)^T \nabla^2 f(\boldsymbol{x}) \right\} \right]^2} \right]^{1/(p+4)}. \tag{2.21}$$

Note that the order of the optimum bandwidth h_{opt} is independent of the dimension p. Despite the tradeoff between the variance and the bias, we obtained an optimum bandwidth which is independent of the dimension and this optimum bandwidth decreases as sample size increases to allow for a better performance of the estimate. Besides for a fixed sample size, the order of the optimum bandwidth is fixed as the dimension increases and it is small enough to guarantee smoothness.

2.2 Curse of Dimensionality

Equation (2.21) is the optimum bandwidth for the multivariate density distribution based on the sample means. This bandwidth preserves the character of the density under averaging. Scott (1992) offers the optimum bandwidth in the usual multivariate special case, where $\boldsymbol{H} = h\boldsymbol{Q}$ as:

$$h_{opt} = n^{-1/(4+p)} \left[\frac{pR(K)}{\int \left[tr\left(Q^T \nabla^2 f(\boldsymbol{x}) \right) \right]^2 d\boldsymbol{x}} \right]^{1/(p+4)} . \qquad (2.22)$$

Here, the order of the optimum bandwidth $h_{opt} \to 1$ as $p \to \infty$, which is the curse of dimensionality. This will give a very rough estimate for large p since the optimum bandwidth is constant irrespective of the sample size n .The optimum bandwidth must depend on the sample size n to avoid too much or too little smoothing.

Since the estimator should be "more local" when more information is added and when the density is rougher, then smaller bandwidths are better for large n. So our optimum bandwidth based on sample means shown in (2.21) is a better bandwidth than the one proposed by Scott (1992) in the multivariate case and Jones, Marron and Sheather (1996) in the usual univariate case.

In practice, multivariate kernel density estimation is often restricted to two dimensions, $p = 2$. The reason is that higher dimensional space (with p large or even of medium size) will be very sparsely populated by data points unless the sample size is very large. This phenomenon is called curse of dimensionality.

The term "curse of dimensionality" was first applied by Bellman (1961) to describe the exponential growth in combinatorial optimization as the dimension increases. In statistics, it reflects the sparsity of data in multiple dimensions and it is the number of bins that grows exponentially as the dimension increases.

Scott and Wand (1991) demonstrated a progressive deterioration of kernel density estimation as the dimension p increases by showing that an increase in sample size is required to attain an equivalent amount of accuracy. Epanechnikov (1969) showed that the growth in sample size is at least exponential as the dimension increases algebraically.

The strategy advocated in this dissertation breaks the curse of dimensionality. For all, $p \geq 1$, $h_{opt} = O(n^{-1/2})$, which leads to an optimum bandwidth which decreases as the sample size increases regardless of the dimensionality. Besides, this will give a bandwidth small enough for large dimensions which guarantee smoothness.

2.3 Bivariate Product Kernel Case

Let $\overline{\boldsymbol{X}} = (\overline{X}_1, \overline{X}_2)'$ be the means of random samples from a population with unknown density, $f(x_1, x_2)$ assumed to be ellipsoidally symmetric, then the estimate of $f(x_1, x_2)$ based on the sample means is given by

$$\hat{f}_{\overline{X}}(x_1, x_2) = \frac{1}{nh_1h_2} K_1\left(\frac{x_1 - \sqrt{n}\overline{X}_1}{n^{1/2}h_1}\right) K_1\left(\frac{x_2 - \sqrt{n}\overline{X}_2}{n^{1/2}h_2}\right), \tag{2.23}$$

where h_1 and h_2 are the bandwidths or the smoothing parameters and K_1 is the symmetric kernel and $\int\int \hat{f}_n^*(x_1, x_2)dx_1 dx_2 = 1$ whenever $\int K_1(u)du = 1$.

Expectation and Bias of $\hat{f}_{\bar{X}}\left(x_1, x_2\right)$

We can obtain the expectation and the bias of this new estimator as follows:

$$E\left[\hat{f}_{\bar{X}}\left(x_1, x_2\right)\right] = \frac{1}{nh_1h_2}\iint K_1\left(\frac{x_1 - \sqrt{n}y_1}{\sqrt{n}h_1}\right)K_1\left(\frac{x_2 - \sqrt{n}y_2}{\sqrt{n}h_2}\right)f_{\bar{X}}\left(y_1, y_2\right)dy_1dy_2 .$$

But $f_{\bar{X}}\left(y_1, y_2\right) = nf\left[\sqrt{n}\left(y_1, y_2\right)\right]$.

Therefore we have

$$\frac{1}{nh_1h_2}\iint K_1\left(\frac{x_1 - \sqrt{n}y_1}{\sqrt{n}h_1}\right)K_1\left(\frac{x_2 - \sqrt{n}y_2}{\sqrt{n}h_2}\right)nf\left[\sqrt{n}\left(y_1, y_2\right)\right]dy_1dy_2 . \quad (2.24)$$

Let $w_1 = \sqrt{n}y_1$ which implies $\frac{dw_1}{dy_1} = n^{1/2}$ and $dy_1 = \frac{dw_1}{n^{1/2}}$.

Similarly, let $w_2 = \sqrt{n}y_2$ which implies $\frac{dw_2}{dy_2} = n^{1/2}$ and $dy_2 = \frac{dw_2}{n^{1/2}}$.

Substituting these into equation (2.24) we have

$$\frac{1}{nh_1h_2}\iint K_1\left(\frac{x_1 - w_1}{\sqrt{n}h_1}\right)K_1\left(\frac{x_2 - w_2}{\sqrt{n}h_2}\right)f\left(w_1, w_2\right)dw_1dw_2 . \quad (2.25)$$

Now, let $u = \frac{x_1 - w_1}{\sqrt{n}h_1}$, then $w_1 = x_1 - u\sqrt{n}h_1$ and $dw_1 = \sqrt{n}h_1du$.

Similarly, let $v = \frac{x_2 - w_2}{\sqrt{n}h_2}$, then $w_2 = x_2 - v\sqrt{n}h_2$ and $dw_2 = \sqrt{n}h_2dv$.

Therefore, substituting these into equation (2.25), we have

$$E\left[\hat{f}_{\bar{X}}\left(x_1, x_2\right)\right] = \iint K_1\left(u\right)K_1\left(v\right)f\left(x_1 - u\sqrt{n}h_1, x_2 - v\sqrt{n}h_2\right)dudv .$$

Now applying the Taylor series expansion to second order, we have

$$\iint K_1(u)K_1(v)[f(x_1,x_2) - \{u\sqrt{n}h_1 f_1(x_1,x_2) + v\sqrt{n}h_2 f_2(x_1,x_2)\}$$
$$+1/2\{nh_1^2u^2 f_{11}(x_1,x_2) + nuvh_1h_2 f_{12}(x_1,x_2) + nvuh_2h_1 f_{21}(x_1,x_2) + nh_2^2v^2 f_{22}(x_1,x_2)\}]dudv$$

$$= \iint K_1(u)K_1(v)f(x_1,x_2)dudv - \iint K_1(u)K_1(v)[u\sqrt{n}h_1 f_1(x_1,x_2) + v\sqrt{n}h_2 f_2(x_1,x_2)]dudv$$
$$+1/2\iint K_1(u)K_1(v)\{nh_1^2u^2 f_{11}(x_1,x_2)\}dudv + 1/2\iint K_1(u)K_1(v)\{nuvh_1h_2 f_{12}(x_1,x_2)\}dudv$$
$$+1/2\iint K_1(u)K_1(v)\{nvuh_2h_1 f_{21}(x_1,x_2)\}dudv + 1/2\iint K_1(u)K_1(v)\{nh_2^2v^2 f_{22}(x_1,x_2)\}dudv.$$

Now applying the moment conditions of kernel, that is $\int uK_1(u)du,\ \int vK_1(v)dv = 0,$

$\int u^2K_1(u)du = \sigma_K^2$, and $\int v^2K_1(v)dv = \sigma_K^2$, we have

$$f(x_1,x_2) + 1/2nh_1^2\sigma_K^2 f_{11}(x_1,x_2) + 1/2nh_2^2\sigma_K^2 f_{22}(x_1,x_2).$$

Therefore

$$E(\hat f_{\bar X}(x_1,x_2)) = f(x_1,x_2) + 1/2n\sigma_K^2[h_1^2 f_{11}(x_1,x_2) + h_2^2 f_{22}(x_1,x_2)]. \qquad (2.26)$$

Therefore, the bias of $\hat f_{\bar X}(x_1,x_2)$ is

$$Bias[\hat f_{\bar X}(x_1,x_2)] = 1/2n\sigma_K^2[h_1^2 f_{11}(x_1,x_2) + h_2^2 f_{22}(x_1,x_2)] \qquad (2.27)$$

The asymptotic square bias (*AISB*) is given by

$$AISB[\hat f_{\bar X}(x_1,x_2)] = \frac{1}{4}n^2\sigma_K^4[h_1^4\iint f_{11}^2(x_1,x_2)dx_1dx_2 + 2h_1^2h_2^2\iint f_{11}(x_1,x_2)f_{22}(x_1,x_2)dx_1dx_2$$
$$+ h_2^4\iint f_{22}^2(x_1,x_2)dx_1dx_2].$$

Therefore, AISB of $\hat f_n^*(x_1,x_2)$ is given by

$$AISB\left[\hat f_{\bar X}(x_1,x_2)\right] = \frac{1}{4}n^2\sigma_K^4[h_1^4R(f_{11}) + 2h_1^2h_2^2\iint f_{11}(x_1,x_2)f_{22}(x_1,x_2)dx_1dx_2 + h_2^4R(f_{22})], \quad (2.28)$$

where $R(f_{11}) = \iint f_{11}^2(x_1,x_2)dx_1dx_2$ and $R(f_{22}) = \iint f_{22}^2(x_1,x_2)dx_1dx_2$.

Variance of $\hat f_{\bar X}(x_1,x_2)$

Squaring equation (2.23) yields

$$\hat f_{\bar X}^2(x_1,x_2) = \frac{1}{n^2h_1^2h_2^2}K_1^2\left(\frac{x_1-\sqrt{n}\bar X_1}{n^{1/2}h_1}\right)K_1^2\left(\frac{x_1-\sqrt{n}\bar X_1}{n^{1/2}h_2}\right).$$

Therefore,

$$E\left[\hat f_{\bar X}^2(x_1,x_2)\right] = \frac{1}{n^2h_1^2h_2^2}\iint K_1^2\left(\frac{x_1-\sqrt{n}y_1}{n^{1/2}h_1}\right)K_1^2\left(\frac{x_1-\sqrt{n}y_2}{n^{1/2}h_2}\right)f_{\bar X}(y_1,y_2)dy_1dy_2 \ .$$

But $f_{\bar X}(y_1,y_2) = nf\left[\sqrt{n}(y_1,y_2)\right]$.

Therefore we have

$$\frac{1}{n^2h_1^2h_2^2}\iint K_1^2\left(\frac{x_1-\sqrt{n}y_1}{\sqrt{n}h_1}\right)K_1^2\left(\frac{x_2-\sqrt{n}y_2}{\sqrt{n}h_2}\right)nf\left[\sqrt{n}(y_1,y_2)\right]dy_1dy_2. \quad (2.29)$$

Let $w_1 = \sqrt{n}y_1$ which implies $\frac{dw_1}{dy_1} = n^{1/2}$ and $dy_1 = \frac{dw_1}{n^{1/2}}$.

Similarly, let $w_2 = \sqrt{n}y_2$ which implies $\frac{dw_2}{dy_2} = n^{1/2}$ and $dy_2 = \frac{dw_2}{n^{1/2}}$.

Therefore, substituting these into equation (2.29), we have

$$\frac{1}{n^2h_1^2h_2^2}\iint K_1^2\left(\frac{x_1-w_1}{\sqrt{n}h_1}\right)K_1^2\left(\frac{x_2-w_2}{\sqrt{n}h_2}\right)f(w_1,w_2)\,dw_1dw_2. \tag{2.30}$$

Now, let $u=\frac{x_1-w_1}{\sqrt{n}h_1}$, then $w_1=x_1-u\sqrt{n}h_1$ and $dw_1=\sqrt{n}h_1du$.

Similarly, let $v=\frac{x_2-w_2}{\sqrt{n}h_2}$, then $w_2=x_2-v\sqrt{n}h_2$ and $dw_2=\sqrt{n}h_2dv$.

Therefore, substituting these into equation (2.30), we have

$$E\left[\hat{f}_{\bar{X}}^2(x_1,x_2)\right]=\frac{1}{nh_1h_2}\iint K_1^2(u)K_1^2(v)f(x_1-u\sqrt{n}h_1,x_2-v\sqrt{n}h_2)\,dudv.$$

Now applying the Taylor series expansion to first order, we have

$$\frac{1}{nh_1h_2}\iint K_1^2(u)K_1^2(v)f(x_1,x_2)\,dudv$$

$$=\frac{f(x_1,x_2)R^2(K_1)}{nh_1h_2},$$

where $\int K_1^2(u)du=R(K_1)$ and $\int K_1^2(v)du=R(K_1)$.

Since the variance term is dominated by the $E\left[\hat{f}_{\bar{X}}^2\right]$, then the variance of $\hat{f}_{\bar{X}}^2(x_1,x_2)$ is given by

$$Var\left[\hat{f}_{\bar{X}}(x_1,x_2)\right]\approx\frac{f(x_1,x_2)R(K)^2}{nh_1h_2}. \tag{2.31}$$

Therefore, the asymptotic integrated variance (*AIV*) is given by

$$AIV\left[\hat{f}_{\bar{X}}(x_1,x_2)\right]=\iint\frac{f(x_1,x_2)R(K)^2}{nh_1h_2}dx_1dx_2=\frac{R(K)^2}{nh_1h_2}. \tag{2.32}$$

Therefore, the asymptotic mean squared integrated error is given by

$$AMISE = AISB + AIV$$
$$= \frac{1}{4} n^2 \sigma_K^4 [h_1^4 R(f_{11}) + 2h_1^2 h_2^2 \iint f_{11}(x_1, x_2) f_{22}(x_1, x_2) dx_1 dx_2 + h_2^4 R(f_{22})] \qquad (2.33)$$
$$+ \frac{R(K)^2}{n h_1 h_2}.$$

We can easily find a solution for the optimum bandwidth if $h_i = h$ for all i.

Therefore, equation (2.33) becomes

$$AMISE = AISB + AIV$$
$$= \frac{1}{4} n^2 \sigma_K^4 [h^4 R(f_{11}) + 2h^4 \iint f_{11}(x_1, x_2) f_{22}(x_1, x_2) dx_1 dx_2 + h^4 R(f_{22})]$$
$$+ \frac{R(K)^2}{n h^2}.$$

Differentiating with respect to h and equating to zero, we have

$$\frac{1}{4} n^2 \sigma_K^4 [4h^3 R(f_{11}) + 8h^3 \iint f_{11}(x_1, x_2) f_{22}(x_1, x_2) dx_1 dx_2 + 4h^3 R(f_{22})] - \frac{2R(K)^2}{n h^3} = 0$$

$$h^6 R(f_{11}) + 2h^6 \iint f_{11}(x_1, x_2) f_{22}(x_1, x_2) dx_1 dx_2 + h^6 R(f_{22}) = \frac{2R(K)^2}{n^3 \sigma_K^4}. \qquad (2.34)$$

From Scott (1992), the general bivariate normal data using a normal kernel gives

$$R(f_{11}) = 3\left[16\pi(1-\rho^2)^{5/2}\sigma_1^5\sigma_2\right]^{-1},$$

$$R(f_{22}) = 3\left[16\pi(1-\rho^2)^{5/2}\sigma_1\sigma_2^5\right]^{-1},$$

and

$$\int_{\Re^2} f_{11} f_{22} dx_1 dx_2 = (1+2\rho^2)\left[16\pi(1-\rho^2)^{5/2}\sigma_1^3\sigma_2^3\right]^{-1},$$

where ρ is the correlation coefficient.

Substituting these into equation (2.34) we have

$$3h^6\left[16\pi\left(1-\rho^2\right)^{5/2}\sigma_1^5\sigma_2\right]^{-1}+2h^6\left(1+2\rho^2\right)\left[16\pi\left(1-\rho^2\right)^{5/2}\sigma_1^3\sigma_2^3\right]^{-1}$$
$$+3h^6\left[16\pi\left(1-\rho^2\right)^{5/2}\sigma_1\sigma_2^5\right]^{-1}=\frac{2}{4\pi n^3}.$$

$$h^6\left[\frac{3}{\left[16\pi\left(1-\rho^2\right)^{5/2}\sigma_1^5\sigma_2\right]}+\frac{2\left(1+2\rho^2\right)}{\left[16\pi\left(1-\rho^2\right)^{5/2}\sigma_1^3\sigma_2^3\right]}+\frac{3}{\left[16\pi\left(1-\rho^2\right)^{5/2}\sigma_1\sigma_2^5\right]}\right]=\frac{1}{2\pi n^3}.$$

Assuming the standard deviation $\sigma_i=\sigma$ for all i, then we have

$$h^6\left[\frac{3\sigma^4+2\left(1+2\rho^2\right)\sigma^4+3\sigma^4}{\left[16\pi\left(1-\rho^2\right)^{5/2}\sigma^{10}\right]}\right]=\frac{1}{2\pi n^3}$$

$$h^6=\frac{16\pi\left(1-\rho^2\right)^{5/2}\sigma^{10}}{8\pi n^3\sigma^4\left(1+\rho^2/2\right)}$$

$$h=\left[\frac{2\left(1-\rho^2\right)^{5/2}\sigma^6}{n^3\left(1+\rho^2/2\right)}\right]^{1/6}.$$

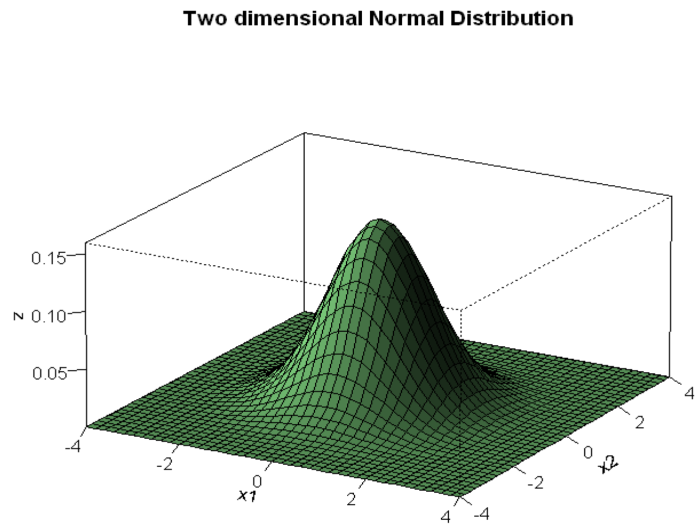
Therefore

$$h_i=n^{-1/2}\left[\frac{2\left(1-\rho^2\right)^{5/2}\sigma_i^6}{\left(1+\rho^2/2\right)}\right]^{1/6},i=1,2. \qquad (2.35)$$

If the variables are independent, then $\rho=0$ and

$$h_i=n^{-1/2}\left[2\sigma_i^6\right]^{1/6},i=1,2. \qquad (2.36)$$

Figure 2.3.1 The Bivariate Normal Density



$$f(\mathbf{x}) = \frac{1}{2\pi\sqrt{\sigma_{11}\,\sigma_{22}\,(1-\rho^2)}} \cdot \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(x_1-\mu_1)^2}{\sigma_{11}} - 2\rho\frac{x_1-\mu_1}{\sqrt{\sigma_{11}}}\frac{x_2-\mu_2}{\sqrt{\sigma_{22}}} + \frac{(x_2-\mu_2)^2}{\sigma_{22}}\right]\right\}$$

Figure 2.3.2 Contour of the Bivariate Normal Density

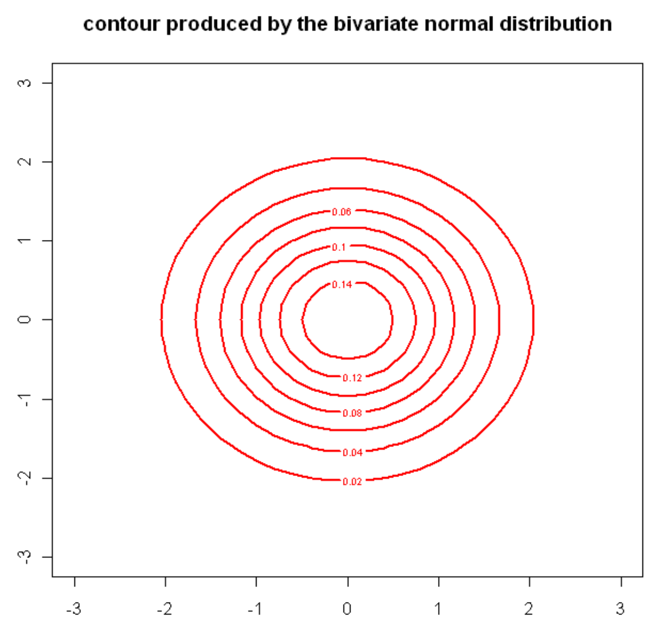


Figure 2.3.3 The Regular Kernel Estimate of the Bivariate Normal Density

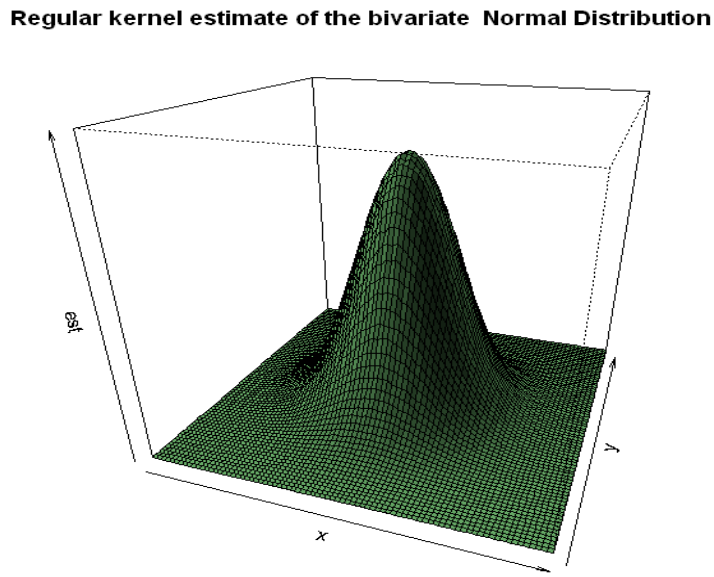


Figure 2.3.4 Contour of the Regular Kernel Estimate of the Bivariate Normal Density

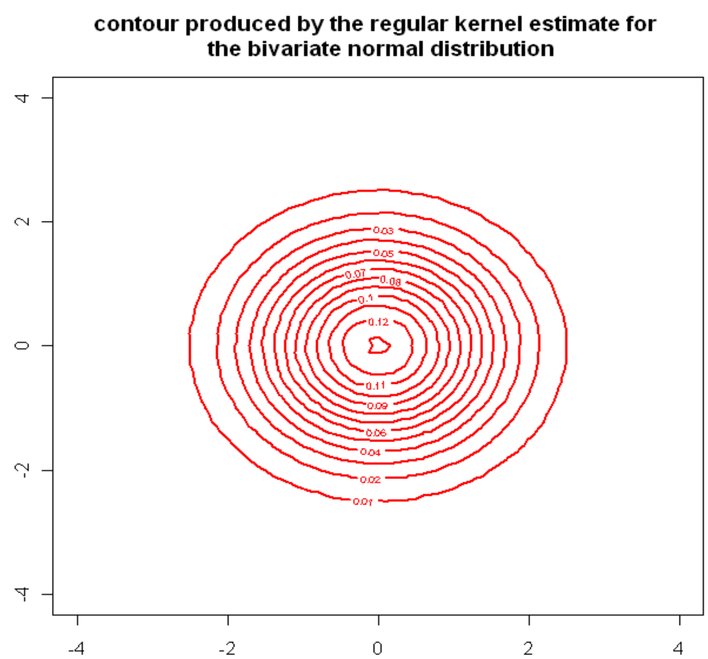


Figure 2.3.5 The Kernel Based on the Sample Means Estimate of the Bivariate Normal Density

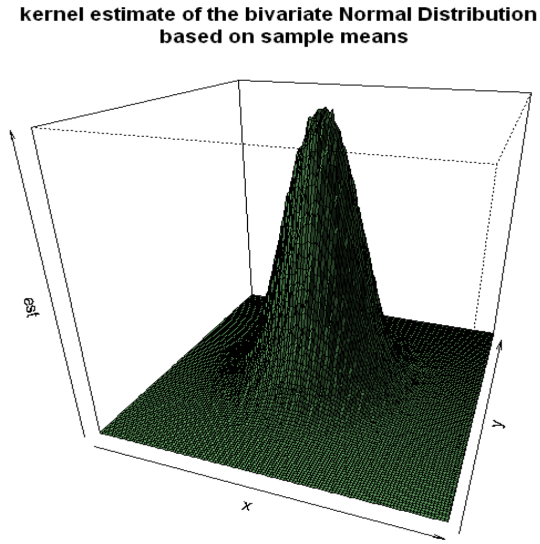
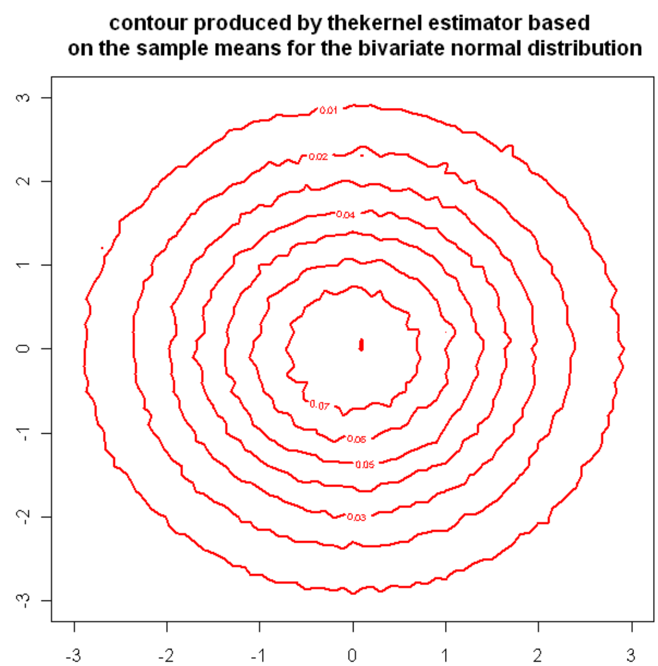


Figure 2.3.6 Contour of the Kernel Based on the Sample Means Estimate of the Bivariate Normal Density



In the simulations below, we used the product kernel which is a special form of the multivariate kernel function. It is the recommended kernel used in practice. In this respect, the individual multivariate product kernel does factor which means that the coordinates are independent but the resultant multivariate kernel density estimate does not factor. In addition, the same kernel is used in each dimension and the univariate standard normal distribution is used.

Table 2.3.1 Mean Squared Error (MSE) of the Regular Multivariate Kernel Density Estimate and the Multivariate Kernel Density Estimate Based on the Sample Mean With Dimension $p=2$

Sample Size, *n*	**MSE for the Regular Multivariate Kernel**	**MSE for the Multivariate Kernel Based on the Sample Means**
10	7.245374×10^{-5}	9.102387×10^{-5}
30	5.086887×10^{-5}	9.097760×10^{-5}
50	3.094227×10^{-5}	9.088346×10^{-5}
100	2.084108×10^{-5}	9.087421×10^{-5}
200	1.391879×10^{-5}	9.086563×10^{-5}
300	1.101187×10^{-5}	9.083921×10^{-5}
500	8.079000×10^{-6}	9.076430×10^{-5}

Table 2.3.2 Mean Squared Error (MSE) of the Regular Multivariate Kernel Density Estimate and the Multivariate Kernel Density Estimate Based on the Sample Mean With Dimension *p=3*

Sample Size, *n*	**MSE for the Regular Multivariate Kernel**	**MSE for the Multivariate Kernel Based on the Sample Means**
10	5.031122×10^{-6}	1.121352×10^{-5}
30	3.205983×10^{-6}	1.121945×10^{-5}
50	2.562826×10^{-6}	1.121509×10^{-5}
100	1.871522×10^{-6}	1.121322×10^{-5}
200	1.349858×10^{-6}	1.120902×10^{-5}
300	1.109322×10^{-6}	1.121277×10^{-5}
500	8.627890×10^{-7}	1.121335×10^{-5}

Table 2.3.3 Mean Squared Error (MSE) of the Regular Multivariate Kernel Density Estimate and the Multivariate Kernel Density Estimate Based on the Sample Mean With Dimension *p=4*

Sample Size, *n*	**MSE for the Regular Multivariate Kernel**	**MSE for the Multivariate Kernel Based on the Sample Means**
10	5.895623×10^{-7}	1.875326×10^{-9}
30	5.205983×10^{-7}	1.875128×10^{-9}
50	4.592657×10^{-7}	1.875112×10^{-9}
100	3.882659×10^{-7}	1.875023×10^{-9}
200	2.358612×10^{-7}	1.875011×10^{-9}
300	2.112615×10^{-7}	1.874933×10^{-9}
500	9.561248×10^{-8}	1.874929×10^{-9}

Simulation results show that the standard or regular multivariate kernel density estimate performs better when the dimension p is less than *4*. When the dimension p is *4* and above, the multivariate kernel density estimate based on the sample means outperforms the regular multivariate kernel density estimate.

2.4 Multivariate Distribution Function Estimation Based on Sample Means

We propose the following estimate of the distribution function $F(\boldsymbol{x})$:

$$\hat{F}(\boldsymbol{x}) = \kappa\left[\left(\frac{\boldsymbol{x}-\sqrt{n}\,\bar{\boldsymbol{X}}}{n^{1/2}}\right)(\boldsymbol{H})^{-1/2}\right], \tag{2.37}$$

where $\kappa(\boldsymbol{x}) = \int_{-\infty}^{x} K_1(\boldsymbol{u})\,d\boldsymbol{u}$ is the kernel corresponding to distribution function and K_1 is the kernel for the density function.

We can compute the expectation of this estimator as follows:

$$E\left[\hat{F}(\boldsymbol{x})\right] = \int \kappa\left[\left(\frac{\boldsymbol{x}-\sqrt{n}\,\bar{\boldsymbol{X}}}{n^{1/2}}\right)(\boldsymbol{H})^{-1/2}\right] f_{\bar{\boldsymbol{X}}}(\boldsymbol{y})\,d\boldsymbol{y}.$$

Since $f_{\bar{\boldsymbol{X}}}(\boldsymbol{y}) = n^{p/2} f(\sqrt{n}\boldsymbol{y})$,

$$E\left[\hat{F}(\boldsymbol{x})\right] = \int \kappa\left[\left(\frac{\boldsymbol{x}-\sqrt{n}\,\bar{\boldsymbol{X}}}{n^{1/2}}\right)(\boldsymbol{H})^{-1/2}\right] n^{p/2} f(\sqrt{n}\boldsymbol{y})\,d\boldsymbol{y}.$$

Letting $\sqrt{n}\boldsymbol{y} = \boldsymbol{w}$, then $\frac{d\boldsymbol{w}}{d\boldsymbol{y}} = n^{p/2}$, $d\boldsymbol{y} = \frac{d\boldsymbol{w}}{n^{p/2}}$, so we have

$$E\left[\hat{F}(\boldsymbol{x})\right] = \int \kappa\left[\left(\frac{\boldsymbol{x}-\boldsymbol{w}}{n^{1/2}}\right)(\boldsymbol{H})^{-1/2}\right] f(\boldsymbol{w})\, d\boldsymbol{w}$$

$$= \int \kappa\left[\left(\frac{\boldsymbol{x}-\boldsymbol{w}}{n^{1/2}}\right)(\boldsymbol{H})^{-1/2}\right] dF(\boldsymbol{w})$$

$$= \kappa\left[\left(\left(\frac{\boldsymbol{x}-\boldsymbol{w}}{n^{1/2}}\right)(\boldsymbol{H})^{-1/2} F(\boldsymbol{w})\right)\right]_{-\infty}^{\infty} + \int F(\boldsymbol{w}) \frac{1}{n^{p/2}|\boldsymbol{H}|^{1/2}} \kappa\left[\left(\frac{\boldsymbol{x}-\boldsymbol{w}}{n^{1/2}}\right)(\boldsymbol{H})^{-1/2}\right] d\boldsymbol{w}\,.$$

Letting $\boldsymbol{u} = \frac{(\boldsymbol{x}-\boldsymbol{w})(\boldsymbol{H})^{-1/2}}{n^{1/2}}$, then $\boldsymbol{w} = \boldsymbol{x} - n^{1/2}(\boldsymbol{H})^{1/2}\boldsymbol{u}$, which implies $|d\boldsymbol{w}| = n^{p/2}|\boldsymbol{H}|^{1/2}\, d\boldsymbol{u}$ so we get that

$$E\left[\hat{F}(\boldsymbol{x})\right] = \int \kappa(\boldsymbol{u}) F(\boldsymbol{x} - n^{1/2}\boldsymbol{H}^{1/2}\boldsymbol{u})\, d\boldsymbol{u}\,.$$

Now, applying the Taylor series expansion to the second order, we have

$$E\left[\hat{F}(\boldsymbol{x})\right] = \int \kappa(\boldsymbol{u})\left[F(\boldsymbol{x}) - n^{1/2}\boldsymbol{H}^{1/2}\nabla F(\boldsymbol{x})\boldsymbol{u}^T + \frac{1}{2} n\boldsymbol{u}^T (\boldsymbol{H}^{1/2})^T \nabla^2 F(\boldsymbol{x}) H^{1/2}\boldsymbol{u}\right] d\boldsymbol{u}$$

$$= F(\boldsymbol{x})\int \kappa(\boldsymbol{u})\, d\boldsymbol{u} - n^{1/2} H^{1/2}\nabla F(\boldsymbol{x})\int \boldsymbol{u}^T \kappa(\boldsymbol{u})\, d\boldsymbol{u} + \frac{1}{2} n (\boldsymbol{H}^{1/2})^T \nabla^2 F(\boldsymbol{x}) \boldsymbol{H}^{1/2} \int \boldsymbol{u}\boldsymbol{u}^T \kappa(\boldsymbol{u})\, d\boldsymbol{u}\,.$$

Applying the moment conditions for kernel, we have

$$\mathrm{E}\left[\hat{F}(\boldsymbol{x})\right] \approx F(\boldsymbol{x}) + \frac{n}{2} tr\left\{(\boldsymbol{H}^{1/2})^T \nabla^2 F(\boldsymbol{x}) \boldsymbol{H}^{1/2}\right\}. \tag{2.38}$$

Therefore, the bias of the estimator is given by

$$Bias\left[\hat{F}(\boldsymbol{x})\right] = \frac{n}{2} tr\left\{(\boldsymbol{H}^{1/2})^T \nabla F(\boldsymbol{x}) (\boldsymbol{H})^{1/2}\right\}\,. \tag{2.39}$$

If *AISB* is the target then we must define a weight function $W(\boldsymbol{x})$ such that

$$AISB = \frac{1}{4} n^2 \int \left[tr\left\{ \left(\boldsymbol{H}^{1/2}\right)^T \nabla F(\boldsymbol{x}) \left(\boldsymbol{H}^{1/2}\right) \right\} \right]^2 dW(\boldsymbol{x}) < \infty \tag{2.40}$$

Next,

$$\hat{F}^2(\boldsymbol{x}) = \kappa^2 \left[\left(\frac{\boldsymbol{x} - \sqrt{n}\,\overline{\boldsymbol{X}}}{n^{1/2}} \right) \left(\boldsymbol{H}\right)^{-1/2} \right] . \tag{2.41}$$

Hence,

$$E\left[\hat{F}^2(\boldsymbol{x})\right] = \int \kappa^2 \left[\left(\frac{\boldsymbol{x} - \sqrt{n}\,\overline{\boldsymbol{X}}}{n^{1/2}} \right) \left(\boldsymbol{H}\right)^{-1/2} \right] f_{\overline{X}}(\boldsymbol{y})\, d\boldsymbol{y} .$$

Since $f_{\overline{X}}(\boldsymbol{y}) = n^{p/2} f\left(\sqrt{n}\boldsymbol{y}\right)$,

$$E\left[\hat{F}^2(\boldsymbol{x})\right] = \int \kappa^2 \left[\left(\frac{\boldsymbol{x} - \sqrt{n}\,\overline{\boldsymbol{X}}}{n^{1/2}} \right) \left(\boldsymbol{H}\right)^{-1/2} \right] n^{p/2} f\left(\sqrt{n}\boldsymbol{y}\right) d\boldsymbol{y} .$$

Letting $\sqrt{n}\boldsymbol{y} = \boldsymbol{w}$, then $\frac{d\boldsymbol{w}}{d\boldsymbol{y}} = n^{p/2}$, $d\boldsymbol{y} = \frac{d\boldsymbol{w}}{n^{p/2}}$, so we have

$$E\left[\hat{F}^2(\boldsymbol{x})\right] = \int \kappa^2 \left[\left(\frac{\boldsymbol{x} - \boldsymbol{w}}{n^{1/2}} \right) \left(\boldsymbol{H}\right)^{-1/2} \right] f(\boldsymbol{w})\, d\boldsymbol{w}$$

$$= \int \kappa^2 \left[\left(\frac{\boldsymbol{x} - \boldsymbol{w}}{n^{1/2}} \right) \left(\boldsymbol{H}\right)^{-1/2} \right] dF(\boldsymbol{w})$$

$$= F(\boldsymbol{w})\kappa^2\left[\left(\left(\frac{\boldsymbol{x}-\boldsymbol{w}}{n^{1/2}}\right)(\boldsymbol{H})^{-1/2}\right)\right]_{-\infty}^{\infty}$$

$$+\int F(\boldsymbol{w})\kappa\left[\left(\frac{\boldsymbol{x}-\boldsymbol{w}}{n^{1/2}}\right)(\boldsymbol{H})^{-1/2}\right]K_1\left[\left(\frac{\boldsymbol{x}-\boldsymbol{w}}{n^{1/2}}\right)(\boldsymbol{H})^{-1/2}\right]\frac{2}{n^{p/2}|\boldsymbol{H}|^{1/2}}d\boldsymbol{w}.$$

Letting $\boldsymbol{u} = \dfrac{(\boldsymbol{x}-\boldsymbol{w})(\boldsymbol{H})^{-1/2}}{n^{p/2}}$, then $\boldsymbol{w} = \boldsymbol{x} - n^{1/2}(\boldsymbol{H})^{1/2}\boldsymbol{u}$, which implies $|d\boldsymbol{w}| = n^{p/2}|\boldsymbol{H}|^{1/2}\, d\boldsymbol{u}$.

So we get

$$E\left[\hat{F}^2(\boldsymbol{x})\right] = 2\int \kappa(\boldsymbol{u})\kappa(\boldsymbol{u})F(\boldsymbol{x} - n^{1/2}\boldsymbol{H}^{1/2}\boldsymbol{u})d\boldsymbol{u}.$$

Thus

$$E\left[\hat{F}^2(\boldsymbol{x})\right] \approx 2F(\boldsymbol{x})\int \kappa(\boldsymbol{u})d\kappa(\boldsymbol{u}) - 2n^{1/2}H^{1/2}\nabla F(\boldsymbol{x})\int \boldsymbol{u}^T\kappa(\boldsymbol{u})\kappa(\boldsymbol{u})d\boldsymbol{u}. \qquad (2.42)$$

Since the variance term is dominated by the $E\left[\hat{F}^2(\boldsymbol{x})\right]$, it follows that

$$Var\left[\hat{F}(\boldsymbol{x})\right] \approx F(\boldsymbol{x}) - 2n^{1/2}\boldsymbol{H}^{1/2}\nabla F(\boldsymbol{x})S(\kappa), \qquad (2.43)$$

where $S(\kappa) = \int \boldsymbol{u}^T\kappa(\boldsymbol{u})\kappa(\boldsymbol{u})d\boldsymbol{u}$.

So, the asymptotic integrated variance is given by

$$AIV = \int F(\boldsymbol{x})dW(\boldsymbol{x}) - 2n^{1/2}\boldsymbol{H}^{1/2}S(\kappa)\int \nabla F(\boldsymbol{x})dW(\boldsymbol{x}). \qquad (2.44)$$

Therefore, the *AMISE* is

$$AMISE = \frac{1}{4}n^2\int\left[tr\left\{(\boldsymbol{H}^{1/2})^T\nabla f(\boldsymbol{x})(\boldsymbol{H}^{1/2})\right\}\right]^2 dW(\boldsymbol{x}) + 1/2 - 2n^{1/2}\boldsymbol{H}^{1/2}S(\kappa)\int \nabla F(\boldsymbol{x})dW(\boldsymbol{x}) \qquad (2.45)$$

Therefore in our special case where $\boldsymbol{H} = h^2\boldsymbol{Q}$, $|\boldsymbol{Q}| = 1$,

$$Bias\left[\hat{F}(x)\right] = \frac{1}{2} n^2 h^2 tr\left((Q^{1/2})^T (Q^{1/2}) \nabla F(x)\right) \tag{2.46}$$

and the asymptotic integrated squared bias (*AISB*) is given by

$$AISB = \frac{1}{4} n^2 h^4 \int \left[tr\left\{ (Q^{1/2})^T \nabla F(x) (Q^{1/2}) \right\} \right]^2 dW(x). \tag{2.47}$$

The asymptotic integrated variance (AIV) is also given by

$$AIV = \int F(x) dW(x) - 2S(x) n^{1/2} h (Q)^{-1/2} \int \nabla F(x) dW(x). \tag{2.48}$$

Therefore $AMISE = AIV + AISB$ leads to,

$$AMISE = \frac{1}{4} n^2 h^4 \int \left[tr\left\{ (Q^{1/2})^T \nabla F(x) (Q^{1/2}) \right\} \right]^2 dW(x) + 1/2 - 2S(x) n^{1/2} h (Q^{1/2}) \int \nabla F(x) dW(x). \tag{2.49}$$

The optimum *h* can be obtained by differentiating equation (2.49) with respect to h and equating it to zero as shown below:

$$n^2 h^3 \int \left[tr\left\{ (Q^{1/2})^T \nabla F(x) Q^{1/2} \right\} \right]^2 dW(x) - 2S(x) n^{1/2} (Q)^{-1/2} \int \nabla F(x) dW(x) = 0$$

$$n^2 h^3 \int \left[tr\left\{ (Q^{1/2})^T \nabla F(x) Q^{1/2} \right\} \right]^2 dW(x) = 2S(x) n^{1/2} (Q)^{-1/2} \int \nabla F(x) dW(x)$$

$$h^3 = \left[\frac{2S(x)(Q)^{-1/2} \int \nabla F(x) dW(x)}{n^{3/2} \int \left[tr\left\{ (Q^{1/2})^T F(x) Q^{1/2} \right\} \right]^2 dW(x)} \right].$$

Hence

$$h_{opt} = n^{-1/2}\left[\frac{2S(x)(Q)^{1/2}\int \nabla F(x)\, dW(x)}{\int\left[tr\left\{(Q^{1/2})^{T}\nabla F(x)\, Q^{1/2}\right\}\right]^{2} dW(x)}\right]^{1/3}. \qquad (2.50)$$

This optimum bandwidth for the multivariate distribution function based on the sample means is also independent of the dimension *p* just like that of its corresponding distribution function.

The optimal point-wise bandwidth for the estimation of multivariate distribution function proposed by Jin and Shao (1999) in the classical kernel case is

$$h^{*}_{opt} = n^{-1/3}\left(\frac{\boldsymbol{M}}{\boldsymbol{L}}\right),$$

where $\boldsymbol{M} = \iint \boldsymbol{w}(\boldsymbol{u})\boldsymbol{w}(\nu)\left(\sum_{\boldsymbol{i}-1}^{\boldsymbol{p}}\frac{\partial \boldsymbol{F}(\boldsymbol{x})}{\partial \boldsymbol{x_i}}\boldsymbol{u_i}\boldsymbol{V}\nu_{\boldsymbol{i}}\right)\boldsymbol{dud}\nu$, where $u_i V \nu_i \equiv \max(u_i, \nu_i)$ and

$\boldsymbol{L} = (\Delta F(\boldsymbol{x}))^{2}$. Note also that, here h is independent of the dimension *p*.

On the other hand, the asymptotic optimal bandwidth for a *p*-dimensional kernel density estimator proposed by Scott (1992) is typically of order $n^{-1/(4+p)}$ and the one in our case for the p-dimensional multivariate kernel density estimator is of order $n^{-1/2}$, which does not depend on the dimension *p* and we realized the bandwidth in our case will yield a good estimator as the dimensions increases. Since a good density estimator might yield a good distribution estimator, it is natural to expect that the optimal bandwidth for the kernel distribution function does not to depend on the dimension *p* and this is what happened in both our case and the one proposed by Jin and Shao (1999). Jin and Shao (1999) established that the optimal bandwidth for a class of kernel estimator of a multivariate distribution function is of order $n^{-1/3}$ for all dimensions.

In this dissertation, the optimal bandwidth for the multivariate distribution function estimator based on sample means is of order $n^{-1/2}$, which does not depend on the dimension p.

Therefore, the optimal bandwidth for the kernel density estimator based on the sample means (equation (2.19)) does not depend on the dimension p and so is the optimal bandwidth for the distribution function based on the sample means.

Also we know those optimum bandwidths which decrease slowly as the sample size increases and the dimensionality increases allow for a better performance of the estimator, which means it guarantees better smoothness. So comparing the optimum bandwidth in our case to that of the one proposed by Shao (1999), the one in our case decrease slowly as sample size increases even though they both do not depend on the dimension p.

CHAPTER III

ESTIMATION OF NONPARAMETRIC MULTIVARIATE REGRESSION FUNCTION BASED ON THE SAMPLE MEANS

3.1 Introduction

In nonparametric multivariate regression, there exists a smooth function $R(\cdot)$ which relates the response variable y and the predictors $\boldsymbol{x}$. The nonparametric multivariate regression is of the form

$$y_i = R(\boldsymbol{x}) + \varepsilon_i \qquad \text{for} \qquad 1 \le i \le n \tag{3.1}$$

where $\varepsilon_i \sim g\left(R(\boldsymbol{x}), \sigma_\varepsilon^2(\boldsymbol{x})\right)$ and g is assumed to be normal and $\sigma_\varepsilon^2(\boldsymbol{x}) = \sigma_\varepsilon^2$, a constant.

3.1.1 Regression Estimation

Based on the random sample means $(\bar{X}_1, \bar{X}_2, ..., \bar{X}_p, \bar{Y})'$, let $\boldsymbol{H} = \begin{bmatrix} \boldsymbol{H_1} & \boldsymbol{0} \\ \boldsymbol{0} & h_2 \end{bmatrix}$, where $\boldsymbol{H_1}$ is the bandwidth matrix for the independent variable $\boldsymbol{X}$ and h_2 is the bandwidth for the response variable Y .Also, let $\boldsymbol{Z} = \begin{pmatrix} \boldsymbol{X} \\ Y \end{pmatrix}$ and K is a (p+1)-dimensional kernel such that $K_2(\boldsymbol{x}) = \int K(\boldsymbol{z})\,dy$. Then, we can estimate $\hat{f}(\cdot)$ as follows:

$$\hat{f}(\boldsymbol{z}) = \hat{f}(\boldsymbol{x}, y) = \frac{1}{n^{(p+1)/2}|\boldsymbol{H}|^{1/2}} K\left[\boldsymbol{H}^{-1/2}\left(\frac{\boldsymbol{z} - \sqrt{n}\bar{\boldsymbol{Z}}}{n^{1/2}}\right)\right]. \tag{3.2}$$

Set the marginal p.d.f of $\boldsymbol{X}$ as:

$$\hat{f}_1(\boldsymbol{x}) = \frac{1}{n^{p/2}|\boldsymbol{H}_1|^{1/2}} K_1\left[\boldsymbol{H}_1^{-1/2}\left(\frac{\boldsymbol{x} - \sqrt{n}\bar{\boldsymbol{X}}}{n^{1/2}}\right)\right], \tag{3.3}$$

where $K_1 \neq K_2 = \int K(\boldsymbol{x}, y)\,dy$.

The nonparametric multivariate kernel regression estimator is given by

$$\hat{R}(\boldsymbol{x}) = E(y \mid \boldsymbol{X} = \boldsymbol{x}) = \int y\hat{f}(y \mid \boldsymbol{x})\,dy = \frac{\int y\hat{f}(\boldsymbol{x}, y)\,dy}{\int \hat{f}(\boldsymbol{x}, y)\,dy}. \tag{3.4}$$

Therefore, from equations (3.3) and (3.4) we have

$$\hat{R}(\boldsymbol{x}) = \frac{\int y\hat{f}(\boldsymbol{x}, y)\,dy}{\hat{f}_1(\boldsymbol{x})}. \tag{3.5}$$

Considering the numerator part of equation (3.4) and substituting equation (3.2) we have

$$\int y\hat{f}\;\;\boldsymbol{x}, y\;\; dy = \int y \frac{1}{n^{\;p+1\;/2}\left|\boldsymbol{H}\right|^{1/2}} K\left[\boldsymbol{H}^{-1/2}\left(\frac{\boldsymbol{z}-\sqrt{n}\bar{\boldsymbol{Z}}}{n^{1/2}}\right)\right]dy$$

$$= \int y \frac{1}{n^{\;p+1\;/2}\left|\boldsymbol{H}\right|^{1/2}} K\left[\boldsymbol{H}^{-1/2}\cdot n^{-1/2}\;\;\boldsymbol{z}-\sqrt{n}\bar{\boldsymbol{Z}}\;\;\right]dy$$

$$= \int y \frac{1}{n^{\;p+1\;/2}\left|\boldsymbol{H}_1\right|^{1/2} h_2^{1/2}} K\left[\begin{bmatrix}\boldsymbol{H}_1^{-1/2}\cdot n^{-1/2} & \boldsymbol{0}\\ \boldsymbol{0} & h_2\cdot n^{-1/2}\end{bmatrix}\begin{pmatrix}\boldsymbol{x}-\sqrt{n}\bar{\boldsymbol{X}}\\ y-\sqrt{n}\bar{Y}\end{pmatrix}\right]dy$$

$$= \frac{1}{n^{\;p+1\;/2}\left|\boldsymbol{H}_1\right|^{1/2} h_2^{1/2}} \int yK\left[\frac{\boldsymbol{x}-\sqrt{n}\bar{\boldsymbol{X}}}{n^{1/2}}\boldsymbol{H}_1^{-1/2}, \frac{y-\sqrt{n}\bar{Y}}{n^{1/2}}h_2^{-1/2}\right]dy\,. \qquad (3.6)$$

Let $\xi = \left(\frac{y-\sqrt{n}\bar{Y}}{n^{1/2}}\right)h_2^{-1/2} \Rightarrow y = h_2^{1/2}n^{1/2}\xi + \sqrt{n}\bar{Y}$ and

$$\frac{dy}{d\xi} = h_2^{1/2}n^{p/2} \Rightarrow dy = h_2^{1/2}n^{p/2}d\xi\,. \qquad (3.7)$$

Substituting these into equation (3.6), we have

$$\int y\hat{f}\;\;\boldsymbol{x}, y\;\; dy$$
$$= \int\;h_2^{1/2}n^{1/2}\xi + \sqrt{n}\bar{Y}\;\; \frac{1}{n^{\;p+1\;/2}\left|\boldsymbol{H}_1\right|^{1/2} h_2^{1/2}} K\left[\boldsymbol{H}_1^{-1/2}\left(\frac{\boldsymbol{x}-\sqrt{n}\bar{\boldsymbol{X}}}{n^{1/2}}\right), \xi\right]h_2^{1/2}n^{p/2}d\xi$$

$$= \int\;h_2^{1/2}n^{1/2}\xi + \sqrt{n}\bar{Y}\;\; \frac{1}{n^{1/2}\left|\boldsymbol{H}_1\right|^{1/2}} K\left[\boldsymbol{H}_1^{-1/2}\left(\frac{\boldsymbol{x}-\sqrt{n}\bar{\boldsymbol{X}}}{n^{1/2}}\right), \xi\right]d\xi$$

$$= \frac{h_2^{1/2}}{|\boldsymbol{H}_1|^{1/2}} \int \xi K \left[\boldsymbol{H}_1^{-1/2} \left(\frac{\boldsymbol{x} - \sqrt{n}\bar{\boldsymbol{X}}}{n^{1/2}} \right), \xi \right] d\xi$$
$$+ \frac{\bar{Y}}{|\boldsymbol{H}_1|^{1/2}} \int K \left[\boldsymbol{H}_1^{-1/2} \left(\frac{\boldsymbol{x} - \sqrt{n}\bar{\boldsymbol{X}}}{n^{1/2}} \right), \xi \right] d\xi$$

$$= \frac{h_2^{1/2}}{|\boldsymbol{H}_1|^{1/2}}$$
$$\times \left[\xi \int K \left[\boldsymbol{H}_1^{-1/2} \left(\frac{\boldsymbol{x} - \sqrt{n}\bar{\boldsymbol{X}}}{n^{1/2}} \right), \xi \right] d\xi - \int \int K \left[\boldsymbol{H}_1^{-1/2} \left(\frac{\boldsymbol{x} - \sqrt{n}\bar{\boldsymbol{X}}}{n^{1/2}} \right), \xi \right] d\xi \right] d\xi$$
$$+ \frac{\bar{Y}}{|\boldsymbol{H}_1|^{1/2}} K_2 \left[\boldsymbol{H}_1^{-1/2} \left(\frac{\boldsymbol{x} - \sqrt{n}\bar{\boldsymbol{X}}}{n^{1/2}} \right) \right]$$

$$= \frac{h_2^{1/2}}{|\boldsymbol{H}_1|^{1/2}} \left[\xi K_2 \left[\boldsymbol{H}_1^{-1/2} \left(\frac{\boldsymbol{x} - \sqrt{n}\bar{\boldsymbol{X}}}{n^{1/2}} \right) \right] - \int K_2 \left[\boldsymbol{H}_1^{-1/2} \left(\frac{\boldsymbol{x} - \sqrt{n}\bar{\boldsymbol{X}}}{n^{1/2}} \right) \right] \right] d\xi$$
$$+ \frac{\bar{Y}}{|\boldsymbol{H}_1|^{1/2}} K_2 \left[\boldsymbol{H}_1^{-1/2} \left(\frac{\boldsymbol{x} - \sqrt{n}\bar{\boldsymbol{X}}}{n^{1/2}} \right) \right], \tag{3.8}$$

by applying integration by parts in the fourth equation above.

Since $\int K \;\; \boldsymbol{x}, \xi \;\; d\xi = K_2 \;\; \boldsymbol{x}$ the right hand side of (3.8) equals

$$\frac{h_2^{1/2}}{|\boldsymbol{H}_1|^{1/2}} \xi K_2 \left[\boldsymbol{H}_1^{-1/2} \left(\frac{\boldsymbol{x} - \sqrt{n}\bar{\boldsymbol{X}}}{n^{1/2}} \right) \right] - \frac{h_2^{1/2}}{|\boldsymbol{H}_1|^{1/2}} \int K_2 \left[\boldsymbol{H}_1^{-1/2} \left(\frac{\boldsymbol{x} - \sqrt{n}\bar{\boldsymbol{X}}}{n^{1/2}} \right) \right] d\xi$$
$$+ \frac{\bar{Y}}{|\boldsymbol{H}_1|^{1/2}} K_2 \left[\boldsymbol{H}_1^{-1/2} \left(\frac{\boldsymbol{x} - \sqrt{n}\bar{\boldsymbol{X}}}{n^{1/2}} \right) \right]$$

$$= \frac{h_2^{1/2}}{|\boldsymbol{H}_1|^{1/2}} \xi K_2 \left[\boldsymbol{H}_1^{-1/2} \left(\frac{\boldsymbol{x} - \sqrt{n}\bar{\boldsymbol{X}}}{n^{1/2}} \right) \right] - \frac{h_2^{1/2}}{|\boldsymbol{H}_1|^{1/2}} \xi K_2 \left[\boldsymbol{H}_1^{-1/2} \left(\frac{\boldsymbol{x} - \sqrt{n}\bar{\boldsymbol{X}}}{n^{1/2}} \right) \right]$$
$$+ \frac{\bar{Y}}{|\boldsymbol{H}_1|^{1/2}} K_2 \left[\boldsymbol{H}_1^{-1/2} \left(\frac{\boldsymbol{x} - \sqrt{n}\bar{\boldsymbol{X}}}{n^{1/2}} \right) \right]$$

$$= \frac{\bar{Y}}{\left|\boldsymbol{H}_1\right|^{1/2}} K_2\left[\boldsymbol{H}_1^{-1/2}\left(\frac{\boldsymbol{x}-\sqrt{n}\bar{\boldsymbol{X}}}{n^{1/2}}\right)\right]. \tag{3.9}$$

Therefore, the nonparametric multivariate kernel regression estimator in equation (3.5) can be written as follows:

$$\hat{R}\left(\boldsymbol{x}\right) = \frac{\dfrac{\bar{Y}}{\left|\boldsymbol{H}_1\right|^{1/2}} K_2\left[\boldsymbol{H}_1^{-1/2}\left(\dfrac{\boldsymbol{x}-\sqrt{n}\bar{\boldsymbol{X}}}{n^{1/2}}\right)\right]}{\dfrac{1}{n^{p/2}\left|\boldsymbol{H}_1\right|^{1/2}} K_1\left[\boldsymbol{H}_1^{-1/2}\left(\dfrac{\boldsymbol{x}-\sqrt{n}\bar{\boldsymbol{X}}}{n^{1/2}}\right)\right]}. \tag{3.10}$$

Or more simply, the equivalent form:

$$\hat{R}\left(\boldsymbol{x}\right) = \frac{\bar{Y} K_2\left[\boldsymbol{H}_1^{-1/2}\left(\dfrac{\boldsymbol{x}-\sqrt{n}\bar{\boldsymbol{X}}}{n^{1/2}}\right)\right]}{K_1\left[\boldsymbol{H}_1^{-1/2}\left(\dfrac{\boldsymbol{x}-\sqrt{n}\bar{\boldsymbol{X}}}{n^{1/2}}\right)\right]}. \tag{3.11}$$

We realize that the nonparametric multivariate kernel regression estimator based on the sample means is linear in the observation of the means $\bar{Y}$ and is therefore a linear smoother, which is a property shared by many other nonparametric regression estimators. Besides, we also realize that this estimator is independent of the particular choice of the smoothing parameter h_2.

3.2 Properties of $\hat{R}\ \ \boldsymbol{x}$

Because the nonparametric multivariate kernel regression estimator based on the sample means is a ratio of two correlated random variables, finding its properties are quite involved. If the numerator and the denominator of the estimator in equation (3.10) each converge to a (positive) constant, then the asymptotic expectation of the ratio is the ratio of the asymptotic expectations of the numerator and the denominator to first order.

3.2.1 Derivation of the Expectation

The properties of the kernel estimator in the denominator were presented in Chapter 2 of this dissertation [see equation (2.8) and (2.16)]; the results are as follows:

$$E\left[\hat{f}_{\bar{X}}(\boldsymbol{x})\right] = f(\boldsymbol{x}) + \frac{n}{2} tr\left(\boldsymbol{H}_1^{1/2\,T} \nabla^2 f(\boldsymbol{x}) \boldsymbol{H}_1^{1/2}\right)$$

and

$$Var\left[\hat{f}_{\bar{X}}\ \ \boldsymbol{x}\ \right] = \frac{f\ \ \boldsymbol{x}\ \ R\ \ K_1}{n^{p/2}\left|\boldsymbol{H}_1\right|^{1/2}} .$$

Now, considering the expectation of the numerator in equation (3.10),

$$E\left[\frac{1}{\left|\boldsymbol{H}_1\right|^{1/2}} K_2\left(\boldsymbol{H}_1^{-1/2}\left(\frac{\boldsymbol{x} - \sqrt{n}\bar{\boldsymbol{X}}}{n^{1/2}}\right)\right)\bar{Y}\right]. \tag{3.12}$$

By letting $\boldsymbol{u} = \sqrt{n}\bar{\boldsymbol{X}}$ and $v = \sqrt{n}\bar{Y}$, and applying the self-revolving property for regression, equation (3.12) becomes

$$= \int\int \frac{1}{n^{p/2}\left|\boldsymbol{H}_1\right|^{1/2}} K_2\left(\boldsymbol{H}_1^{-1/2}\left(\frac{\boldsymbol{x}-\boldsymbol{u}}{n^{1/2}}\right)\right) v f_{\bar{\boldsymbol{X}},y} \;\; \boldsymbol{u}, v \;\; d\boldsymbol{u}dv \, . \qquad (3.13)$$

Let $\boldsymbol{s} = \left(\frac{\boldsymbol{x}-\boldsymbol{u}}{n^{1/2}}\right)\boldsymbol{H}_1^{-1/2}$, then $\boldsymbol{x}-\boldsymbol{u} = n^{1/2}\boldsymbol{H}_1^{1/2}\boldsymbol{s} \Rightarrow \boldsymbol{u} = \boldsymbol{x} - n^{1/2}\boldsymbol{H}_1^{1/2}\boldsymbol{s}$.

Therefore $\left|\frac{d\boldsymbol{u}}{d\boldsymbol{s}}\right| = n^{p/2}\left|\boldsymbol{H}_1\right|^{1/2} \Rightarrow d\boldsymbol{u} = n^{p/2}\left|\boldsymbol{H}_1\right|^{1/2} d\boldsymbol{s}$.

Substituting these into equation (3.13), we have that it is equal to

$$\int\int \frac{1}{n^{p/2}\left|\boldsymbol{H}_1\right|^{1/2}} K_2 \;\; \boldsymbol{s} \;\; v f \;\; \boldsymbol{x} - n^{1/2}\boldsymbol{H}_1^{1/2}\boldsymbol{s}, v \;\; n^{p/2}\left|\boldsymbol{H}_1\right|^{1/2} d\boldsymbol{s}dv$$

$$= \int\int K_2 \;\; \boldsymbol{s} \;\; v f \;\; \boldsymbol{x} - n^{1/2}\boldsymbol{H}_1^{1/2}\boldsymbol{s}, v \;\; d\boldsymbol{s}dv \, . \qquad (3.14)$$

But $f \;\; \boldsymbol{x} - n^{/2}\boldsymbol{H}_1^{1/2}\boldsymbol{s}, v \;\; = f \;\; v \setminus \boldsymbol{x} - n^{1/2}\boldsymbol{H}_1^{1/2}\boldsymbol{s} \;\; f \;\; \boldsymbol{x} - n^{1/2}\boldsymbol{H}_1^{1/2}\boldsymbol{s}$. (3.15)

Substituting (3.15) into equation (3.14) , the integral over v in equation (3.14) is equivalent to [ignoring $K_2 \;\; \boldsymbol{s}$]

$$f \;\; \boldsymbol{x} - n^{1/2}\boldsymbol{H}_1^{1/2}\boldsymbol{s} \;\; \int v f \;\; v \setminus \boldsymbol{x} - n^{1/2}\boldsymbol{H}_1^{1/2}\boldsymbol{s} \;\; dv \qquad (3.16)$$

$$= f \;\; \boldsymbol{x} - n^{1/2}\boldsymbol{H}_1^{1/2}\boldsymbol{s} \;\; R \;\; \boldsymbol{x} - n^{1/2}\boldsymbol{H}_1^{1/2}\boldsymbol{s} \;\; , \qquad (3.17)$$

since the integral is the conditional mean and $R \;\; \cdot$ is the true regression function defined in equation (3.1). Therefore, equation (3.14) becomes

$$\int K_2 \;\; \boldsymbol{s} \;\; f \;\; \boldsymbol{x} - n^{1/2}\boldsymbol{H}_1^{1/2}\boldsymbol{s} \;\; R \;\; \boldsymbol{x} - n^{1/2}\boldsymbol{H}_1^{1/2}\boldsymbol{s} \;\; d\boldsymbol{s} \, . \qquad (3.18)$$

Now expanding $f\left(\boldsymbol{x}-n^{1/2}\boldsymbol{H}_1^{1/2}\boldsymbol{s}\right)$ and $R\left(\boldsymbol{x}-n^{1/2}\boldsymbol{H}_1^{1/2}\boldsymbol{s}\right)$ in the Taylor series to second order, we have

$$f\left(\boldsymbol{x}-n^{1/2}\boldsymbol{H}_1^{1/2}\boldsymbol{s}\right)=f(\boldsymbol{x})-n^{1/2}\boldsymbol{s}^T\left(\boldsymbol{H}_1\right)^{1/2}\nabla f(\boldsymbol{x})+\frac{1}{2}n\boldsymbol{s}^T\left(\boldsymbol{H}_1^{1/2}\right)^T\nabla^2 f(\boldsymbol{x})\left(\boldsymbol{H}_1^{1/2}\right)\boldsymbol{s}$$

and

$$R\left(\boldsymbol{x}-n^{1/2}\boldsymbol{H}_1^{1/2}\boldsymbol{s}\right)=R(\boldsymbol{x})-n^{1/2}\boldsymbol{s}^T\left(\boldsymbol{H}_1\right)^{1/2}\nabla R(\boldsymbol{x})+\frac{1}{2}n\boldsymbol{s}^T\left(\boldsymbol{H}_1^{1/2}\right)^T\nabla^2 R(\boldsymbol{x})\left(\boldsymbol{H}_1^{1/2}\right)\boldsymbol{s}.$$

Therefore, expression (3.18) becomes

$$\int K_2(\boldsymbol{s})\left[f(\boldsymbol{x})-n^{1/2}\boldsymbol{s}^T\left(\boldsymbol{H}_1\right)^{1/2}\nabla f(\boldsymbol{x})+\frac{1}{2}n\boldsymbol{s}^T\left(\boldsymbol{H}_1^{1/2}\right)^T\nabla^2 f(\boldsymbol{x})\left(\boldsymbol{H}_1^{1/2}\right)\boldsymbol{s}\right]\times$$
$$\left[R(\boldsymbol{x})-n^{1/2}\boldsymbol{s}^T\left(\boldsymbol{H}_1\right)^{1/2}\nabla R(\boldsymbol{x})+\frac{1}{2}n\boldsymbol{s}^T\left(\boldsymbol{H}_1^{1/2}\right)^T\nabla^2 R(\boldsymbol{x})\left(\boldsymbol{H}_1^{1/2}\right)\boldsymbol{s}\right]d\boldsymbol{s}$$

$$\begin{aligned}
&=\int K_2(\boldsymbol{s})f(\boldsymbol{x})R(\boldsymbol{x})\,d\boldsymbol{s}-\int K_2(\boldsymbol{s})f(\boldsymbol{x})n^{1/2}\left(\boldsymbol{H}_1\right)^{1/2}\nabla R(\boldsymbol{x})\boldsymbol{s}^T d\boldsymbol{s}\\
&\quad+\int\frac{1}{2}K_2(\boldsymbol{s})n\boldsymbol{s}^T\left(\boldsymbol{H}_1^{1/2}\right)^T\nabla^2 R(\boldsymbol{x})\left(\boldsymbol{H}_1\right)^{1/2}\boldsymbol{s}f(\boldsymbol{x})\,d\boldsymbol{s}\\
&\quad-\int K_2(\boldsymbol{s})n^{1/2}\left(\boldsymbol{H}_1\right)^{1/2}\nabla f(\boldsymbol{x})\boldsymbol{s}^T R(\boldsymbol{x})\,d\boldsymbol{s}\\
&\quad+\int K_2(\boldsymbol{s})n\left(\boldsymbol{H}_1^{1/2}\right)^T\nabla f(\boldsymbol{x})\boldsymbol{s}^T\left(\boldsymbol{H}_1^{1/2}\right)\nabla R(\boldsymbol{x})\boldsymbol{s}^T d\boldsymbol{s}\\
&\quad-\int\frac{1}{2}K_2(\boldsymbol{s})n^{3/2}\boldsymbol{s}^T\left(\boldsymbol{H}_1\right)^{1/2}\nabla f(\boldsymbol{x})\boldsymbol{s}^T\left(\boldsymbol{H}_1^{1/2}\right)^T\nabla^2 R(\boldsymbol{x})\left(\boldsymbol{H}_1^{1/2}\right)\boldsymbol{s}\,d\boldsymbol{s}\\
&\quad+\int K_2(\boldsymbol{s})\frac{1}{2}n\boldsymbol{s}^T\left(\boldsymbol{H}_1^{1/2}\right)^T\nabla^2 f(\boldsymbol{x})\left(\boldsymbol{H}_1^{1/2}\right)\boldsymbol{s}R(\boldsymbol{x})\,d\boldsymbol{s}\\
&\quad-\int\frac{1}{2}K_2(\boldsymbol{s})n^{3/2}\boldsymbol{s}^T\left(\boldsymbol{H}_1^{1/2}\right)^T\nabla^2 f(\boldsymbol{x})\left(\boldsymbol{H}_1^{1/2}\right)\boldsymbol{s}\boldsymbol{s}^T n^{1/2}\left(\boldsymbol{H}_1^{1/2}\right)\nabla R(\boldsymbol{x})\,d\boldsymbol{s}\\
&\quad+\int K_2(\boldsymbol{s})\frac{1}{4}n^{2p}\boldsymbol{s}^T\left(\boldsymbol{H}_1^{1/2}\right)^T\nabla^2 f(\boldsymbol{x})\left(\boldsymbol{H}_1^{1/2}\right)\boldsymbol{s}\boldsymbol{s}^T\left(\boldsymbol{H}_1^{1/2}\right)^T\nabla^2 R(\boldsymbol{x})\left(\boldsymbol{H}_1^{1/2}\right)\boldsymbol{s}\,d\boldsymbol{s}.
\end{aligned}\tag{3.19}$$

Assuming K_2s a second order kernel, equation (3.19) becomes

$$= \int K_2(\boldsymbol{s})\, f(\boldsymbol{x})\, R(\boldsymbol{x})\, d\boldsymbol{s} - \int K_2(\boldsymbol{s})\, f(\boldsymbol{x})\, n^{1/2} (\boldsymbol{H}_1)^{1/2} \nabla R(\boldsymbol{x})\, \boldsymbol{s}^T d\boldsymbol{s}$$
$$+ \int \frac{1}{2} K_2(\boldsymbol{s})\, n\boldsymbol{s}^T (\boldsymbol{H}_1^{1/2})^T \nabla^2 R(\boldsymbol{x}) (\boldsymbol{H}_1)^{1/2} \boldsymbol{s} f(\boldsymbol{x})\, d\boldsymbol{s}$$
$$- \int K_2(\boldsymbol{s})\, n^{1/2} (\boldsymbol{H}_1)^{1/2} \nabla f(\boldsymbol{x})\, \boldsymbol{s}^T R(\boldsymbol{x})\, d\boldsymbol{s}$$
$$+ \int K_2(\boldsymbol{s})\, n (\boldsymbol{H}_1^{1/2})^T \nabla f(\boldsymbol{x})\, \boldsymbol{s}^T (\boldsymbol{H}_1^{1/2}) \nabla R(\boldsymbol{x})\, \boldsymbol{s}^T d\boldsymbol{s}$$
$$+ \int K_2(\boldsymbol{s}) \frac{1}{2} n\boldsymbol{s}^T (\boldsymbol{H}_1^{1/2})^T \nabla^2 f(\boldsymbol{x}) (\boldsymbol{H}_1^{1/2}) \boldsymbol{s} R(\boldsymbol{x})\, d\boldsymbol{s} \qquad (3.20)$$

$$= f(\boldsymbol{x})\, R(\boldsymbol{x}) \int K_2(\boldsymbol{s})\, d\boldsymbol{s} - f(\boldsymbol{x})\, n^{1/2} (\boldsymbol{H}_1)^{1/2} \nabla R(\boldsymbol{x}) \int \boldsymbol{s}^T K_2(\boldsymbol{s})\, d\boldsymbol{s}$$
$$+ \frac{1}{2} n (\boldsymbol{H}_1^{1/2})^T \nabla^2 R(\boldsymbol{x}) (\boldsymbol{H}_1)^{1/2} f(\boldsymbol{x}) \int \boldsymbol{s}^T \boldsymbol{s} K_2(\boldsymbol{s})\, d\boldsymbol{s}$$
$$- n^{1/2} (\boldsymbol{H}_1)^{1/2} \nabla f(\boldsymbol{x})\, R(\boldsymbol{x}) \int \boldsymbol{s}^T K_2(\boldsymbol{s})\, d\boldsymbol{s}$$
$$+ n (\boldsymbol{H}_1^{1/2})^T \nabla f(\boldsymbol{x}) (\boldsymbol{H}_1^{1/2}) \nabla R(\boldsymbol{x}) \int \boldsymbol{s}^T \boldsymbol{s}^T K_2(\boldsymbol{s})\, d\boldsymbol{s}$$
$$+ \frac{1}{2} n (\boldsymbol{H}_1^{1/2})^T \nabla^2 f(\boldsymbol{x}) (\boldsymbol{H}_1^{1/2}) R(\boldsymbol{x}) \int \boldsymbol{s}^T \boldsymbol{s} K_2(\boldsymbol{s})\, d\boldsymbol{s}. \qquad (3.21)$$

Now, applying some of the kernel multivariate moment conditions, equation (3.21) becomes

$$= f(\boldsymbol{x})\, R(\boldsymbol{x}) + \frac{1}{2} n (\boldsymbol{H}_1^{1/2})^T \nabla^2 R(\boldsymbol{x}) (\boldsymbol{H}_1)^{1/2} f(\boldsymbol{x}) \int \boldsymbol{s}^T \boldsymbol{s} K_2(\boldsymbol{s})\, d\boldsymbol{s}$$
$$+ n (\boldsymbol{H}_1^{1/2})^T \nabla f(\boldsymbol{x}) (\boldsymbol{H}_1^{1/2}) \nabla R(\boldsymbol{x}) \int \boldsymbol{s}^T \boldsymbol{s}^T K_2(\boldsymbol{s})\, d\boldsymbol{s}$$
$$+ \frac{1}{2} n (\boldsymbol{H}_1^{1/2})^T \nabla^2 f(\boldsymbol{x}) (\boldsymbol{H}_1^{1/2}) R(\boldsymbol{x}) \int \boldsymbol{s}^T \boldsymbol{s} K_2(\boldsymbol{s})\, d\boldsymbol{s}$$

$$= f(\boldsymbol{x})\, R(\boldsymbol{x}) + \frac{1}{2} tr\left\{ n (\boldsymbol{H}_1^{1/2})^T \nabla^2 R(\boldsymbol{x}) (\boldsymbol{H}_1)^{\frac{1}{2}} f(\boldsymbol{x}) \int \boldsymbol{s}^T \boldsymbol{s} K_2(\boldsymbol{s})\, d\boldsymbol{s} \right\}$$
$$+ tr\left(n (\boldsymbol{H}_1^{1/2})^T \nabla f(\boldsymbol{x}) (\boldsymbol{H}_1^{1/2}) \nabla R(\boldsymbol{x}) \int \boldsymbol{s}^T \boldsymbol{s}^T K_2(\boldsymbol{s})\, d\boldsymbol{s} \right)$$
$$+ \frac{1}{2} tr\left(n (\boldsymbol{H}_1^{1/2})^T \nabla^2 f(\boldsymbol{x}) (\boldsymbol{H}_1^{1/2}) R(\boldsymbol{x}) \int \boldsymbol{s}^T \boldsymbol{s} K_2(\boldsymbol{s})\, d\boldsymbol{s} \right). \qquad (3.22)$$

Again, by applying the moment conditions, the covariance matrix of K_2 is I_d and the integral factor within the trace vanishes, so equation (3.22) becomes

$$
\begin{aligned}
&= f(\boldsymbol{x})R(\boldsymbol{x}) + \frac{1}{2} tr\left(n^p (\boldsymbol{H}_1^{1/2})^T \nabla^2 R(\boldsymbol{x}) (\boldsymbol{H}_1)^{1/2}\right) f(\boldsymbol{x}) \\
&\quad + tr\left(n (\boldsymbol{H}_1^{1/2})^T \nabla f(\boldsymbol{x}) (\boldsymbol{H}_1^{1/2}) \nabla R(\boldsymbol{x})\right) \\
&\quad + \frac{1}{2} tr\left(n (\boldsymbol{H}_1^{1/2})^T \nabla^2 f(\boldsymbol{x}) (\boldsymbol{H}_1^{1/2})\right) R(\boldsymbol{x}) . \qquad (3.23)
\end{aligned}
$$

Therefore, the expectation of the nonparametric multivariate regression kernel estimator based on the sample means is the ratio of the expectations in equations (3.23) and (2.8), that is

$$
\begin{aligned}
E(\hat{R}) = \Bigg[& f(\boldsymbol{x})R(\boldsymbol{x}) + \frac{1}{2} tr\left(n (\boldsymbol{H}_1^{1/2})^T \nabla^2 R(\boldsymbol{x}) (\boldsymbol{H}_1)^{1/2}\right) f(\boldsymbol{x}) \\
& + tr\left(n (\boldsymbol{H}_1^{1/2})^T \nabla f(\boldsymbol{x}) (\boldsymbol{H}_1)^{1/2} \nabla R(\boldsymbol{x})\right) \\
& + \frac{1}{2} tr\left(n (\boldsymbol{H}_1^{1/2})^T \nabla^2 f(\boldsymbol{x}) (\boldsymbol{H}_1^{1/2})\right) R(\boldsymbol{x}) \Bigg] \\
\div \Bigg[& f(\boldsymbol{x}) + \frac{1}{2} tr\left(n (\boldsymbol{H}_1^{1/2})^T \nabla^2 f(\boldsymbol{x}) (\boldsymbol{H}_1^{1/2})\right) \Bigg]. \qquad (3.24)
\end{aligned}
$$

Equation (3.24) can be written as

$$
\begin{aligned}
E(\hat{R}) = f(\boldsymbol{x}) \Bigg[& R(\boldsymbol{x}) + \frac{1}{2} tr\left\{n (\boldsymbol{H}_1^{1/2})^T \nabla^2 R(\boldsymbol{x}) (\boldsymbol{H}_1)^{\frac{1}{2}}\right\} \\
& + tr\left(n (\boldsymbol{H}_1^{1/2})^T \nabla f(\boldsymbol{x}) \nabla R(\boldsymbol{x}) (\boldsymbol{H}_1^{1/2})\right) \Big/ f(\boldsymbol{x}) \\
& + \frac{1}{2} tr\left(n (\boldsymbol{H}_1^{1/2})^T \nabla^2 f(\boldsymbol{x}) R(\boldsymbol{x}) (\boldsymbol{H}_1^{1/2})\right) \Big/ f(\boldsymbol{x}) \Bigg] \\
\div f(\boldsymbol{x}) \Bigg[& 1 + \frac{1}{2 f(\boldsymbol{x})} tr\left(n (\boldsymbol{H}_1^{1/2})^T \nabla^2 f(\boldsymbol{x}) (\boldsymbol{H}_1^{1/2})\right) \Bigg].
\end{aligned}
$$

$$
\begin{aligned}
=\Bigg[& R(\boldsymbol{x}) + \frac{1}{2} tr\left(n \left(\boldsymbol{H}_1^{1/2}\right)^T \nabla^2 R(\boldsymbol{x}) \boldsymbol{H}_1^{1/2}\right) \\
& + tr\left(n \left(\boldsymbol{H}_1^{1/2}\right)^T \nabla f(\boldsymbol{x}) \nabla R(\boldsymbol{x}) \boldsymbol{H}_1^{1/2}\right) \Big/ f(\boldsymbol{x}) \\
& + \frac{1}{2} tr\left(n \left(\boldsymbol{H}_1^{1/2}\right)^T \nabla^2 f(\boldsymbol{x}) R(\boldsymbol{x}) \boldsymbol{H}_1^{1/2}\right) \Big/ f(\boldsymbol{x}) \Bigg] \\
\div \Bigg[& 1 + \frac{1}{2} tr\left(n \left(\boldsymbol{H}_1^{1/2}\right)^T \nabla^2 f(\boldsymbol{x}) \boldsymbol{H}_1^{1/2}\right) \Big/ f(\boldsymbol{x}) \Bigg].
\end{aligned}
\tag{3.25}
$$

Now, using the approximation $(1+t^2c)^{-1} \approx (1-t^2c)$ for small c in the factor in the denominator and t, a scalar, equation (3.25) becomes

$$
\begin{aligned}
=\Bigg[& R(\boldsymbol{x}) + \frac{1}{2} tr\left(n \left(\boldsymbol{H}_1^{1/2}\right)^T \nabla^2 R(\boldsymbol{x}) \boldsymbol{H}_1^{1/2}\right) \\
& + tr\left(n \left(\boldsymbol{H}_1^{1/2}\right)^T \nabla f(\boldsymbol{x}) \nabla R(\boldsymbol{x}) \boldsymbol{H}_1^{1/2}\right) \Big/ f(\boldsymbol{x}) + \frac{1}{2} tr\left(n \left(\boldsymbol{H}_1^{1/2}\right)^T \nabla^2 f(\boldsymbol{x}) R(\boldsymbol{x}) \boldsymbol{H}_1^{1/2}\right) \Big/ f(\boldsymbol{x}) \Bigg] \\
\times \Bigg[& 1 - \frac{1}{2} tr\left(n \left(\boldsymbol{H}_1^{1/2}\right)^T \nabla^2 f(\boldsymbol{x}) \boldsymbol{H}_1^{1/2}\right) \Big/ f(\boldsymbol{x}) \Bigg]
\end{aligned}
$$

$$
\begin{aligned}
=\Bigg[& R(\boldsymbol{x}) + \frac{1}{2} tr\left(n \left(\boldsymbol{H}_1^{1/2}\right)^T \nabla^2 R(\boldsymbol{x}) \boldsymbol{H}_1^{1/2}\right) + tr\left(n \left(\boldsymbol{H}_1^{1/2}\right)^T \nabla f(\boldsymbol{x}) \nabla R(\boldsymbol{x}) \boldsymbol{H}_1^{1/2}\right) \Big/ f(\boldsymbol{x}) \\
& + \frac{1}{2} tr\left(n \left(\boldsymbol{H}_1^{1/2}\right)^T \nabla^2 f(\boldsymbol{x}) R(\boldsymbol{x}) \boldsymbol{H}_1^{1/2}\right) \Big/ f(\boldsymbol{x}) \\
& - \frac{1}{2} R(\boldsymbol{x}) \frac{1}{2} tr\left(n \left(\boldsymbol{H}_1^{1/2}\right)^T \nabla^2 f(\boldsymbol{x}) \boldsymbol{H}_1^{1/2}\right) \Big/ f(\boldsymbol{x}) \\
& - \frac{1}{2} tr\left(n \left(\boldsymbol{H}_1^{1/2}\right)^T \nabla^2 R(\boldsymbol{x}) \boldsymbol{H}_1^{1/2}\right) \frac{1}{2} tr\left(n \left(\boldsymbol{H}_1^{1/2}\right)^T \nabla^2 f(\boldsymbol{x}) \boldsymbol{H}_1^{1/2}\right) \Big/ f(\boldsymbol{x}) \\
& - tr\left(n \left(\boldsymbol{H}_1^{1/2}\right)^T \nabla f(\boldsymbol{x}) \nabla R(\boldsymbol{x}) \boldsymbol{H}_1^{1/2}\right) \Big/ f(\boldsymbol{x}) \frac{1}{2} tr\left(n \left(\boldsymbol{H}_1^{1/2}\right)^T \nabla^2 f(\boldsymbol{x}) \boldsymbol{H}_1^{1/2}\right) \Big/ f(\boldsymbol{x}) \\
& - \frac{1}{2} tr\left(n \left(\boldsymbol{H}_1^{1/2}\right)^T \nabla^2 f(\boldsymbol{x}) R(\boldsymbol{x}) \boldsymbol{H}_1^{1/2}\right) \Big/ f(\boldsymbol{x}) \frac{1}{2} tr\left(n \left(\boldsymbol{H}_1^{1/2}\right)^T \nabla^2 f(\boldsymbol{x}) \boldsymbol{H}_1^{1/2}\right) \Big/ f(\boldsymbol{x}) \Bigg]
\end{aligned}
\tag{3.26}
$$

Keeping $\boldsymbol{H}_1^{\frac{1}{2}}$ to second order, equation (3.26) becomes

$$= \Big[R(\boldsymbol{x}) + \frac{1}{2} tr\{n(\boldsymbol{H}_1^{1/2})^T \nabla^2 R(\boldsymbol{x})(\boldsymbol{H}_1)^{1/2}\} + tr\{n(\boldsymbol{H}_1^{1/2})^T \nabla f(\boldsymbol{x}) \nabla R(\boldsymbol{x})(\boldsymbol{H}_1^{1/2})\} \big/ f(\boldsymbol{x})$$

$$+ \frac{1}{2} tr\{n(\boldsymbol{H}_1^{1/2})^T \nabla^2 f(\boldsymbol{x}) R(\boldsymbol{x})(\boldsymbol{H}_1^{1/2})\} \big/ f(\boldsymbol{x})$$

$$- \frac{1}{2} R(\boldsymbol{x}) \frac{1}{2} tr\{n(\boldsymbol{H}_1^{1/2})^T \nabla^2 f(\boldsymbol{x})(\boldsymbol{H}_1^{1/2})\} \big/ f(\boldsymbol{x}) \Big].$$

Simplifying equation (3.26) we have

$$= R(\boldsymbol{x}) + \frac{1}{2} tr\{n(\boldsymbol{H}_1^{1/2})^T \nabla^2 R(\boldsymbol{x})(\boldsymbol{H}_1)^{1/2}\} + tr\{n(\boldsymbol{H}_1^{1/2})^T \nabla f(\boldsymbol{x}) \nabla R(\boldsymbol{x})(\boldsymbol{H}_1^{1/2})\} \big/ f(\boldsymbol{x}). \tag{3.27}$$

Applying some properties of trace to equation (3.27), we have

$$E\left(\hat{R}(\boldsymbol{x})\right) = R(\boldsymbol{x}) + \frac{n}{2} tr\left((\boldsymbol{H}_1^{1/2})^T \boldsymbol{H}_1^{1/2}\right)\left\{\nabla^2 R(\boldsymbol{x}) + 2\nabla R(\boldsymbol{x}) \frac{\nabla f(\boldsymbol{x})}{f(\boldsymbol{x})}\right\}. \tag{3.28}$$

If we have many data points, then the term $2\nabla R(\boldsymbol{x}) \dfrac{\nabla f(\boldsymbol{x})}{f(\boldsymbol{x})}$ in equation (3.28) will be small.

3.2.2 Variance and AMSE of the Estimate

The variance of the estimator $\hat{R}(\boldsymbol{x})$ can be computed using the approximation of the ratio of two random variables (Stuart and Ord, 1987),

$$Var\left[\frac{\boldsymbol{U}}{V}\right] \approx \left[\frac{EU}{EV}\right]^2 \left[\frac{\operatorname{var} U}{(EU)^2} + \frac{\operatorname{var} V}{(EV)^2} - \frac{2Cov(U,V)}{(EU)(EV)}\right]. \tag{3.29}$$

From equation (3.10), and applying the above approximation, we can write

$$Var(\hat{R}(\boldsymbol{x})) = Var\left\{ \frac{\dfrac{\bar{y}}{|\boldsymbol{H}_1|^{1/2}} K_2\left[\boldsymbol{H}_1^{-1/2}\left(\dfrac{\boldsymbol{x}-\sqrt{n}\bar{\boldsymbol{x}}}{n^{1/2}}\right)\right]}{\dfrac{1}{n^{p/2}|\boldsymbol{H}_1|^{1/2}} K_1\left[\boldsymbol{H}_1^{-1/2}\left(\dfrac{\boldsymbol{x}-\sqrt{n}\bar{\boldsymbol{x}}}{n^{1/2}}\right)\right]} \right\}. \tag{3.30}$$

We already know the expectation of the numerator and we also know both the expectation and the variance of the denominator. Now, we need to find the variance of the numerator and the covariance between the numerator and the denominator.

$$\begin{aligned} &\text{var}\left[\frac{\bar{y}}{|\boldsymbol{H}_1|^{1/2}} K_2\left[\boldsymbol{H}_1^{-1/2}\left(\frac{\boldsymbol{x}-\sqrt{n}\bar{\boldsymbol{x}}}{n^{1/2}}\right)\right]\right] \\ &= \int\int \frac{1}{|\boldsymbol{H}_1|} K_2^2\left[\boldsymbol{H}_1^{-1/2}\left(\frac{\boldsymbol{x}-\sqrt{n}\bar{\boldsymbol{x}}}{n^{1/2}}\right)\right]\bar{y}^2 \, n^{(p+1)/2} f\left(\sqrt{n}\bar{\boldsymbol{x}},\sqrt{n}\bar{y}\right) d\bar{\boldsymbol{x}}d\bar{y} \\ &-\left[E\left(\frac{\bar{y}}{|\boldsymbol{H}_1|^{1/2}} K_2\left[\boldsymbol{H}_1^{-1/2}\left(\frac{\boldsymbol{x}-\sqrt{n}\bar{\boldsymbol{x}}}{n^{1/2}}\right)\right]\right)\right]^2 \end{aligned} \tag{3.31}$$

Now, considering the first term of equation (3.31), by letting $\boldsymbol{u} = \sqrt{n}\bar{\boldsymbol{x}}$ and $v = \sqrt{n}\bar{y}$,

then the first term of equation (3.31) becomes

$$\int\int \frac{1}{n^p |\boldsymbol{H}_1|} K_2^2\left(\boldsymbol{H}_1^{-1/2}\left(\frac{\boldsymbol{x}-\boldsymbol{u}}{n^{1/2}}\right)\right) v^2 f_{\bar{\mathbf{x}},y}\left(\boldsymbol{u},v\right) d\boldsymbol{u}dv. \tag{3.32}$$

Let $\boldsymbol{s} = \left(\dfrac{\boldsymbol{x}-\boldsymbol{u}}{n^{1/2}}\right)\boldsymbol{H}_1^{-1/2}$, then

$$\boldsymbol{x}-\boldsymbol{u} = n^{1/2}\boldsymbol{H}_1^{1/2}\boldsymbol{s} \Rightarrow \boldsymbol{u} = \boldsymbol{x} - n^{1/2}\boldsymbol{H}_1^{1/2}\boldsymbol{s}\,. \tag{3.33}$$

Therefore $\left|\dfrac{d\boldsymbol{u}}{d\boldsymbol{s}}\right| = n^{p/2}|\boldsymbol{H}_1|^{1/2} \Rightarrow d\boldsymbol{u} = n^{p/2}|\boldsymbol{H}_1|^{1/2}\, d\boldsymbol{s}$.

Substituting these into equation (3.32), we have

$$= \int\int \frac{1}{n^{p/2}\left|\boldsymbol{H}_1\right|} K_2^2(\boldsymbol{s}) v^2 f_{\bar{x},y}(\boldsymbol{x} - n^{1/2}\boldsymbol{H}_1^{1/2}\boldsymbol{s}, v) d\boldsymbol{s}dv . \qquad (3.34)$$

$$\text{But } f(\boldsymbol{x} - n^{1/2}H_1^{1/2}\boldsymbol{s}, v) = f(v \mid \boldsymbol{x} - n^{1/2}\boldsymbol{H}_1^{1/2}\boldsymbol{s}) f(\boldsymbol{x} - n^{1/2}\boldsymbol{H}_1^{1/2}\boldsymbol{s}) . \qquad (3.35)$$

Substituting (3.35) into equation (3.34), the integral over v in equation (3.34) is equivalent to [ignoring $K_2^2(s)$]

$$f(\boldsymbol{x} - n^{1/2}\boldsymbol{H}_1^{1/2}\boldsymbol{s}) \int \frac{1}{n^{p/2}\left|\boldsymbol{H}_1\right|^{1/2}} v^2 f(v \mid \boldsymbol{x} - n^{1/2}\boldsymbol{H}_1^{1/2}\boldsymbol{s}) dv . \qquad (3.36)$$

Here, the integral is the conditional second moment, and $R(\cdot)$ is the true regression function defined in equation (3.1). Therefore, equation (3.36) becomes

$$f(\boldsymbol{x} - n^{1/2}\boldsymbol{H}_1^{1/2}\boldsymbol{s}) \frac{1}{n^{p/2}\left|\boldsymbol{H}_1\right|^{1/2}} R^2(\boldsymbol{x} - n^{1/2}\boldsymbol{H}_1^{1/2}\boldsymbol{s}) . \qquad (3.37)$$

Substituting this into equation (3.34) we have

$$\frac{1}{n^{p/2}\left|\boldsymbol{H}_1\right|^{1/2}} \int K_2^2(\boldsymbol{s}) f(\boldsymbol{x} - n^{1/2}\boldsymbol{H}_1^{1/2}\boldsymbol{s}) \left[R(\boldsymbol{x} - n^{1/2}\boldsymbol{H}_1^{1/2}\boldsymbol{s})^2 + \sigma_\varepsilon^2(\boldsymbol{x} - n^{1/2}\boldsymbol{H}_1^{1/2}\boldsymbol{s}) \right] d\boldsymbol{s}$$

$$= \frac{1}{n^{p/2}\left|\boldsymbol{H}_1\right|^{1/2}} \int K_2^2(\boldsymbol{s}) f(\boldsymbol{x} - n^{1/2}\boldsymbol{H}_1^{1/2}\boldsymbol{s}) \sigma_\varepsilon^2(\boldsymbol{x} - n^{1/2}\boldsymbol{H}_1^{1/2}\boldsymbol{s}) d\boldsymbol{s}$$
$$+ \frac{1}{n^{p/2}\left|\boldsymbol{H}_1\right|^{1/2}} \int K_2^2(\boldsymbol{s}) f(\boldsymbol{x} - n^{1/2}\boldsymbol{H}_1^{1/2}\boldsymbol{s}) R(\boldsymbol{x} - n^{1/2}\boldsymbol{H}_1^{1/2}\boldsymbol{s})^2 d\boldsymbol{s}$$

$$= \frac{R(K_2) f(\boldsymbol{x}) \sigma_\varepsilon^2(\boldsymbol{x})}{n^{p/2}\left|\boldsymbol{H}_1\right|^{1/2}} + \frac{R(K_2) f(\boldsymbol{x}) R(\boldsymbol{x})^2}{n^{p/2}\left|\boldsymbol{H}_1\right|^{1/2}} .$$

Therefore equation (3.32) becomes

$$\approx \frac{R(K_2) f(\boldsymbol{x})}{n^{p/2}|\boldsymbol{H}_1|^{1/2}}\left[\sigma_\varepsilon^2 + R(\boldsymbol{x})^2\right], \tag{3.38}$$

where $\sigma_\varepsilon^2(\boldsymbol{x}) = \sigma_\varepsilon^2$ for all $\boldsymbol{x}$.

Also, considering the second term of equation (3.31)

$$\left[E\left(\frac{\bar{y}}{|\boldsymbol{H}_1|^{1/2}} K_2\left[\boldsymbol{H}_1^{-1/2}\left(\frac{\boldsymbol{x}-\sqrt{n}\bar{\boldsymbol{x}}}{n^{1/2}}\right)\right]\right)\right]^2$$

$$= \left[f(\boldsymbol{x})R(\boldsymbol{x}) + \frac{1}{2}tr\left(n^2(\boldsymbol{H}_1^{1/2})^T \nabla^2 R(\boldsymbol{x})(\boldsymbol{H}_1)^{1/2} f(\boldsymbol{x})\right)\right.$$

$$\left. + tr\left(n^2(\boldsymbol{H}_1^{1/2})^T \nabla f(\boldsymbol{x})(\boldsymbol{H}_1)^{1/2} \nabla R(\boldsymbol{x})\right) + \frac{1}{2}tr\left(n^2(\boldsymbol{H}_1^{1/2})^T \nabla^2 f(\boldsymbol{x})(\boldsymbol{H}_1^{1/2})R(\boldsymbol{x})\right)\right]^2$$

$$= O\left(n^2 tr(\boldsymbol{H}_1^T \boldsymbol{H}_1)\right). \tag{3.39}$$

Therefore (3.31) can be written as

$$\operatorname{var}\left[\frac{\bar{y}}{|\boldsymbol{H}_1|^{1/2}} K_2\left[\boldsymbol{H}_1^{-1/2}\left(\frac{\boldsymbol{x}-\sqrt{n}\bar{\boldsymbol{x}}}{n^{1/2}}\right)\right]\right] = \frac{R(K_2) f(\boldsymbol{x})}{n^{p/2}|\boldsymbol{H}_1|^{1/2}}\left[\sigma_\varepsilon^2 + R(\boldsymbol{x})^2\right] - O\left(n^2 tr(\boldsymbol{H}_1^T \boldsymbol{H}_1)\right)$$

$$\approx \frac{R(K_2) f(\boldsymbol{x})}{n^{p/2}|\boldsymbol{H}_1|^{1/2}}\left[\sigma_\varepsilon^2 + R(\boldsymbol{x})^2\right], \tag{3.40}$$

where $\sigma_\varepsilon^2(\boldsymbol{x}) = \sigma_\varepsilon^2$ for all $\boldsymbol{x}$. Next,

$$\operatorname{cov}\left[\frac{\bar{y}}{|\boldsymbol{H}_1|^{p/2}}K_2\left(\boldsymbol{H}_1^{-1/2}\left(\frac{\boldsymbol{x}-\sqrt{n}\bar{\boldsymbol{x}}}{n^{1/2}}\right)\right),\frac{1}{n^{p/2}|\boldsymbol{H}_1|^{1/2}}K_1\left(\boldsymbol{H}_1^{-1/2}\left(\frac{\boldsymbol{x}-\sqrt{n}\bar{\boldsymbol{x}}}{n^{1/2}}\right)\right)\right]$$
$$=E\left[\frac{\bar{y}}{|\boldsymbol{H}_1|^{1/2}}K_2\left(\boldsymbol{H}_1^{-1/2}\left(\frac{\boldsymbol{x}-\sqrt{n}\bar{\boldsymbol{x}}}{n^{1/2}}\right)\right)\frac{1}{n^{p/2}|\boldsymbol{H}_1|^{1/2}}K_1\left(\boldsymbol{H}_1^{-1/2}\left(\frac{\boldsymbol{x}-\sqrt{n}\bar{\boldsymbol{x}}}{n^{1/2}}\right)\right)\right]$$
$$-E\left[\frac{\bar{y}}{|\boldsymbol{H}_1|^{1/2}}K_2\left(\boldsymbol{H}_1^{-1/2}\left(\frac{\boldsymbol{x}-\sqrt{n}\bar{\boldsymbol{x}}}{n^{1/2}}\right)\right)\right]E\left[\frac{1}{n^{p/2}|\boldsymbol{H}_1|^{1/2}}K_1\left(\boldsymbol{H}_1^{-1/2}\left(\frac{\boldsymbol{x}-\sqrt{n}\bar{\boldsymbol{x}}}{n^{1/2}}\right)\right)\right]. \tag{3.41}$$

Now, considering the first term of equation (3.41), we have

$$E\left[\frac{\bar{y}}{|\boldsymbol{H}_1|^{1/2}}K_2\left(\boldsymbol{H}_1^{-1/2}\left(\frac{\boldsymbol{x}-\sqrt{n}\bar{\boldsymbol{x}}}{n^{1/2}}\right)\right)\frac{1}{n^{p/2}|\boldsymbol{H}_1|^{1/2}}K_1\left(\boldsymbol{H}_1^{-1/2}\left(\frac{\boldsymbol{x}-\sqrt{n}\bar{\boldsymbol{x}}}{n^{1/2}}\right)\right)\right] \tag{3.42}$$

$$=\iint\frac{\bar{y}}{|\boldsymbol{H}_1|^{1/2}}K_2\left(\boldsymbol{H}_1^{-1/2}\left(\frac{\boldsymbol{x}-\sqrt{n}\bar{\boldsymbol{x}}}{n^{1/2}}\right)\right)\frac{1}{n^{p/2}|\boldsymbol{H}_1|^{1/2}}K_1\left(\boldsymbol{H}_1^{-1/2}\left(\frac{\boldsymbol{x}-\sqrt{n}\bar{\boldsymbol{x}}}{n^{1/2}}\right)\right)n^{(p+1)/2}$$
$$\times f\left(\sqrt{n}\boldsymbol{x},\sqrt{n}\bar{y}\right)d\bar{\boldsymbol{x}}d\bar{y}$$

$$=\iint\frac{\bar{y}}{n^{p/2}|\boldsymbol{H}_1|}K_2\left(\boldsymbol{H}_1^{-1/2}\left(\frac{\boldsymbol{x}-\sqrt{n}\bar{\boldsymbol{x}}}{n^{1/2}}\right)\right)K_1\left(\boldsymbol{H}_1^{-1/2}\left(\frac{\boldsymbol{x}-\sqrt{n}\bar{\boldsymbol{x}}}{n^{1/2}}\right)\right)n^{(p+1)/2}\,f\left(\sqrt{n}\boldsymbol{x},\sqrt{n}\bar{y}\right)d\bar{\boldsymbol{x}}d\bar{y}\,.$$

Letting $\boldsymbol{u}=\sqrt{n}\bar{\boldsymbol{x}}\Rightarrow\frac{d\boldsymbol{u}}{d\bar{\boldsymbol{x}}}=n^{p/2}$ and $v=\sqrt{n}\bar{y}\Rightarrow\frac{dv}{d\bar{y}}=n^{p/2}$, then equation (3.42) becomes

$$=\iint\frac{v}{n^{p}|\boldsymbol{H}_1|}K_2\left(\boldsymbol{H}_1^{-1/2}\left(\frac{\boldsymbol{x}-\boldsymbol{u}}{n^{1/2}}\right)\right)K_1\left(\boldsymbol{H}_1^{-1/2}\left(\frac{\boldsymbol{x}-\boldsymbol{u}}{n^{1/2}}\right)\right)f\left(\boldsymbol{u},v\right)d\boldsymbol{u}dv\,. \tag{3.43}$$

Now, let $\boldsymbol{s}=\left(\frac{\boldsymbol{x}-\boldsymbol{u}}{n^{1/2}}\right)\boldsymbol{H}_1^{-1/2}$, then $\boldsymbol{x}-\boldsymbol{u}=n^{1/2}\boldsymbol{H}_1^{1/2}\boldsymbol{s}\Rightarrow\boldsymbol{u}=\boldsymbol{x}-n^{1/2}\boldsymbol{H}_1^{1/2}\boldsymbol{s}$.

Therefore $\left|\frac{d\boldsymbol{u}}{d\boldsymbol{s}}\right| = n^{p/2}\left|\boldsymbol{H}\right|^{1/2} \Rightarrow d\boldsymbol{u} = n^{p/2}\left|\boldsymbol{H}\right|^{1/2} d\boldsymbol{s}$

Substituting these into equation (3.43), we have

$$= \iint \frac{v}{n^{p}\left|\boldsymbol{H}_1\right|} K_2(\boldsymbol{s}) K_1(\boldsymbol{s}) n^{p/2}\left|H_1\right|^{1/2} f(\boldsymbol{x} - n^{1/2} H_1^{1/2}\boldsymbol{s}, v)\, d\boldsymbol{s}dv$$

$$= \iint \frac{1}{n^{p/2}\left|\boldsymbol{H}_1\right|^{1/2}} K_2(\boldsymbol{s}) K_1(\boldsymbol{s}) vf(\boldsymbol{x} - n^{1/2} H_1^{1/2}\boldsymbol{s}, v)\, d\boldsymbol{s}dv \,. \qquad (3.44)$$

But $f(\boldsymbol{x} - n^{1/2}\boldsymbol{H}_1^{1/2}\boldsymbol{s}, v) = f(v \mid \boldsymbol{x} - n^{1/2}\boldsymbol{H}_1^{1/2}\boldsymbol{s}) f(\boldsymbol{x} - n^{1/2}\boldsymbol{H}_1^{1/2}\boldsymbol{s})$.

Substituting this into equation (3.44), the integral over v in equation (3.44) is equivalent to [ignoring $K_2(\boldsymbol{s}) K_1(\boldsymbol{s})$]

$$f(\boldsymbol{x} - n^{1/2}\boldsymbol{H}_1^{1/2}\boldsymbol{s}) \int \frac{1}{n^{p/2}\left|\boldsymbol{H}_1\right|^{1/2}} vf(v \mid \boldsymbol{x} - n^{1/2}\boldsymbol{H}_1^{1/2}\boldsymbol{s})\, dv$$

$$= \frac{1}{n^{p/2}\left|\boldsymbol{H}_1\right|^{1/2}} f(\boldsymbol{x} - n^{1/2}\boldsymbol{H}_1^{1/2}\boldsymbol{s}) R(\boldsymbol{x} - n^{1/2}\boldsymbol{H}_1^{1/2}\boldsymbol{s}) ,$$

as the integral is the conditional mean, and $R(\cdot)$ is the true regression function defined in equation (3.1). Therefore, equation (3.44) becomes

$$\int \frac{1}{n^{p/2}\left|\boldsymbol{H}_1\right|^{1/2}} K_2(\boldsymbol{s}) K_1(\boldsymbol{s}) f(\boldsymbol{x} - n^{1/2}\boldsymbol{H}_1^{1/2}\boldsymbol{s}, v) R(\boldsymbol{x} - n^{1/2}\boldsymbol{H}_1^{1/2}\boldsymbol{s}, v)\, d\boldsymbol{s}$$

$$= \frac{R(K) f(\boldsymbol{x}) R(\boldsymbol{x})}{n^{p/2}\left|\boldsymbol{H}_1\right|^{1/2}} \,. \qquad (3.45)$$

Also, considering the second term of equation (3.31)

$$E\left[\frac{\bar{y}}{|\boldsymbol{H}_1|^{1/2}}K_2\left(\boldsymbol{H}_1^{-1/2}\left(\frac{\boldsymbol{x}-\sqrt{n}\bar{\boldsymbol{x}}}{n^{1/2}}\right)\right)\right]E\left[\frac{1}{n^{p/2}|\boldsymbol{H}_1|^{1/2}}K_1\left(\boldsymbol{H}_1^{-1/2}\left(\frac{\boldsymbol{x}-\sqrt{n}\bar{\boldsymbol{x}}}{n^{1/2}}\right)\right)\right]$$

$$=\Big[f\ \boldsymbol{x}\ R\ \boldsymbol{x}\ +\frac{1}{2}tr\ n^2\ \boldsymbol{H}_1^{1/2}\ ^T\nabla^2 R\ \boldsymbol{x}\ \ \boldsymbol{H}_1\ ^{1/2}\ f\ \ \boldsymbol{x}$$

$$+tr\ n^2\ \boldsymbol{H}_1^{1/2}\ ^T\nabla f\ \ \boldsymbol{x}\ \ \boldsymbol{H}_1\ ^{1/2}\nabla R\ \ \boldsymbol{x}\ \ +\frac{1}{2}tr\ n^2\ \boldsymbol{H}_1^{1/2}\ ^T\nabla^2 f\ \ \boldsymbol{x}\ \ \boldsymbol{H}_1^{1/2}\ R\ \boldsymbol{x}\ \Big]$$

$$\times\Big[f\ \ \boldsymbol{x}\ +\frac{1}{2}tr\ n^2\ \boldsymbol{H}_1^{1/2}\ ^T\nabla^2 f\ \ \boldsymbol{x}\ \ \boldsymbol{H}_1^{1/2}\ \Big]$$

$$=O\left(n^4 tr\left(\boldsymbol{H}_1^T\boldsymbol{H}_1\right)\right). \tag{3.46}$$

Now substituting equation (3.45) and (3.46) into equation (3.41), we have

$$\text{cov}\left[\frac{\bar{y}}{|\boldsymbol{H}_1|^{p/2}}K_2\left(\boldsymbol{H}_1^{-1/2}\left(\frac{\boldsymbol{x}-\sqrt{n}\bar{\boldsymbol{x}}}{n^{1/2}}\right)\right),\frac{1}{n^{p/2}|\boldsymbol{H}_1|^{1/2}}K_1\left(\boldsymbol{H}_1^{-1/2}\left(\frac{\boldsymbol{x}-\sqrt{n}\bar{\boldsymbol{x}}}{n^{1/2}}\right)\right)\right]$$

$$=\frac{R\ K\ \ f\ \ \boldsymbol{x}\ R\ \boldsymbol{x}}{n^{p/2}|\boldsymbol{H}_1|^{1/2}}-O\ \ n^2 tr\ \ \boldsymbol{H}_1^{\ T}\boldsymbol{H}_1$$

$$\approx\frac{R\ K\ \ f\ \ \boldsymbol{x}\ R\ \boldsymbol{x}}{n^{p/2}|\boldsymbol{H}_1|^{1/2}}. \tag{3.47}$$

Substituting equations (2.8), (2.16), (3.23), (3.38) and (3.47) into equation (3.29), we have

$$Var\ \ \hat{R}\ \ \boldsymbol{x}\ \ =\boldsymbol{I}\times\boldsymbol{II},$$

where

$$
\boldsymbol{I} = \frac{\left[\begin{array}{l} f(\boldsymbol{x})R(\boldsymbol{x}) + \frac{1}{2}tr\left(n^2\left(\boldsymbol{H}_1^{1/2}\right)^T \nabla^2 R(\boldsymbol{x})\left(\boldsymbol{H}_1\right)^{1/2}\right) f(\boldsymbol{x}) \\ + tr\left(n^2\left(\boldsymbol{H}_1^{1/2}\right)^T \nabla f(\boldsymbol{x})\left(\boldsymbol{H}_1\right)^{1/2} \nabla R(\boldsymbol{x})\right) + \frac{1}{2}tr\left(n^2\left(\boldsymbol{H}_1^{1/2}\right)^T \nabla^2 f(\boldsymbol{x})\left(\boldsymbol{H}_1^{1/2}\right)\right) R(\boldsymbol{x}) \end{array}\right]^2}{\left[f(\boldsymbol{x}) + \frac{1}{2}tr\left(n^2\left(\boldsymbol{H}_1^{1/2}\right)^T \nabla^2 f(\boldsymbol{x})\left(\boldsymbol{H}_1\right)^{1/2}\right)\right]^2}
$$

and

$$
\boldsymbol{II} = \frac{\frac{R(K)f(\boldsymbol{x})}{n^{p/2}\left|\boldsymbol{H}_1\right|^{1/2}}\left[\sigma_\varepsilon^2 + R(\boldsymbol{x})^2\right]}{\left[\begin{array}{l} f(\boldsymbol{x})R(\boldsymbol{x}) + \frac{1}{2}tr\left\{n^2\left(\boldsymbol{H}_1^{\frac{1}{2}}\right)^T \nabla^2 R(\boldsymbol{x})\left(\boldsymbol{H}_1\right)^{1/2} f(\boldsymbol{x})\right\} \\ + tr\left(n^2\left(\boldsymbol{H}_1^{1/2}\right)^T \nabla f(\boldsymbol{x})\left(\boldsymbol{H}_1\right)^{1/2} \nabla R(\boldsymbol{x})\right) + \frac{1}{2}tr\left(n^2\left(\boldsymbol{H}_1^{1/2}\right)^T \nabla^2 f(\boldsymbol{x})\left(\boldsymbol{H}_1^{1/2}\right) R(\boldsymbol{x})\right) \end{array}\right]^2}
$$

$$
+ \frac{\frac{f(\boldsymbol{x})R(K)}{n^{p/2}\left|\boldsymbol{H}_1\right|^{1/2}}}{\left[f(\boldsymbol{x}) + \frac{1}{2}tr\left(n^2\left(\boldsymbol{H}_1^{1/2}\right)^T \nabla^2 f(\boldsymbol{x})\left(\boldsymbol{H}\right)^{1/2}\right)\right]^2}
$$

$$
- \frac{\frac{2R(K)f(\boldsymbol{x})R(\boldsymbol{x})}{n^{p/2}\left|\boldsymbol{H}_1\right|^{1/2}}}{\left[\begin{array}{l} f(\boldsymbol{x})R(\boldsymbol{x}) + \frac{1}{2}tr\left\{n^2\left(\boldsymbol{H}_1^{1/2}\right)^T \nabla^2 R(\boldsymbol{x})\left(\boldsymbol{H}_1\right)^{1/2} f(\boldsymbol{x})\right\} \\ + tr\left\{n^2\left(\boldsymbol{H}_1^{1/2}\right)^T \nabla f(\boldsymbol{x})\left(\boldsymbol{H}_1\right)^{1/2} \nabla R(\boldsymbol{x})\right\} + \frac{1}{2}tr\left\{n^2\left(\boldsymbol{H}_1^{1/2}\right)^T \nabla^2 f(\boldsymbol{x})\left(\boldsymbol{H}_1^{1/2}\right) R(\boldsymbol{x})\right\} \\ \times\left[f(\boldsymbol{x}) + \frac{1}{2}tr\left\{n^2\left(\boldsymbol{H}_1^{1/2}\right)^T \nabla^2 f(\boldsymbol{x})\left(\boldsymbol{H}_1\right)^{1/2}\right\}\right] \end{array}\right]}.
$$

Letting

$$
\boldsymbol{A} = \left[\begin{array}{l} f(\boldsymbol{x})R(\boldsymbol{x}) + \frac{1}{2}tr\left(n^2\left(\boldsymbol{H}_1^{1/2}\right)^T \nabla^2 R(\boldsymbol{x})\left(\boldsymbol{H}_1\right)^{1/2}\right) f(\boldsymbol{x}) \\ + tr\left(n^2\left(\boldsymbol{H}_1^{1/2}\right)^T \nabla f(\boldsymbol{x})\left(\boldsymbol{H}_1\right)^{1/2} \nabla R(\boldsymbol{x})\right) + \frac{1}{2}tr\left(n^2\left(\boldsymbol{H}_1^{1/2}\right)^T \nabla^2 f(\boldsymbol{x})\left(\boldsymbol{H}_1^{1/2}\right)\right) R(\boldsymbol{x}) \end{array}\right]
$$

and

$$\boldsymbol{B} = f(\boldsymbol{x}) + \frac{1}{2} tr\left(n^2 \left(\boldsymbol{H}_1^{1/2}\right)^T \nabla^2 f(\boldsymbol{x}) \left(\boldsymbol{H}_1\right)^{1/2}\right),$$

we get that

$$Var\left(\hat{R}(\boldsymbol{x})\right) = \left[\frac{\boldsymbol{A}}{\boldsymbol{B}}\right]^2 \times \frac{1}{\boldsymbol{A^2B^2}} \left[\frac{R(K) f(\boldsymbol{x})}{n^{(p+1)/2} |\boldsymbol{H}_1|^{1/2}} \left[\sigma_\varepsilon^2 + R(\boldsymbol{x})^2\right] \boldsymbol{B}^2 + \frac{f_{\bar{x}}(\boldsymbol{x}) R(K)}{n^{p/2} |\boldsymbol{H}_1|^{1/2}} A^2 - \frac{2R(K) f(\boldsymbol{x}) R(\boldsymbol{x}) \boldsymbol{AB}}{n^{p/2} |\boldsymbol{H}_1|^{1/2}}\right]$$

$$= \frac{1}{\boldsymbol{B}^4} \left[\begin{array}{l} \frac{R(K) f(\boldsymbol{x})}{n^{p/2} |H_1|^{1/2}} \left[\sigma_\varepsilon^2 + R(\boldsymbol{x})^2\right] \boldsymbol{B}^2 + \\ + \frac{f(\boldsymbol{x}) R(K)}{n^{p/2} |\boldsymbol{H}_1|^{1/2}} A^2 - \frac{2R(K) f(\boldsymbol{x}) R(\boldsymbol{x}) \boldsymbol{AB}}{n^{p/2} |\boldsymbol{H}_1|^{1/2}} \end{array}\right]$$

$$= \frac{\left[\frac{R(K) f(\boldsymbol{x})}{n^{p/2} |\boldsymbol{H}_1|^{1/2}}\right]}{\boldsymbol{B}^4} \left[\left[\sigma_\varepsilon^2 + R(\boldsymbol{x})^2\right] \boldsymbol{B}^2 + \boldsymbol{A}^2 - 2R(\boldsymbol{x}) \boldsymbol{AB}\right]. \tag{3.48}$$

Now, considering the first term of equation (3.48), we have

$$\left[\sigma_\varepsilon^2 + R(\boldsymbol{x})^2\right] \boldsymbol{B}^2 = \left[\sigma_\varepsilon^2 + R(\boldsymbol{x})^2\right] \left[f(\boldsymbol{x}) + \frac{1}{2} tr\left(n^2 \left(\boldsymbol{H}_1^{1/2}\right)^T \nabla^2 f(\boldsymbol{x}) \left(\boldsymbol{H}_1\right)^{1/2}\right)\right]^2$$

$$
\begin{aligned}
&= \left[\sigma_\varepsilon^2 + R(\boldsymbol{x})^2 \right] \\
&\times \left[f^2(\boldsymbol{x}) + f(\boldsymbol{x})\, tr\left(n^2 \left(\boldsymbol{H}_1^{1/2}\right)^T \nabla^2 f(\boldsymbol{x}) \left(\boldsymbol{H}_1\right)^{1/2} \right) + \frac{1}{4} \left(tr\left(n^2 \left(\boldsymbol{H}_1^{1/2}\right)^T \nabla^2 f(\boldsymbol{x}) \left(\boldsymbol{H}_1\right)^{1/2} \right) \right)^2 \right] \\
&= f^2(\boldsymbol{x})\, \sigma_\varepsilon^2 + f(\boldsymbol{x})\, tr\left(n^2 \left(\boldsymbol{H}_1^{1/2}\right)^T \nabla^2 f(\boldsymbol{x}) \left(\boldsymbol{H}_1\right)^{1/2} \right) \sigma_\varepsilon^2 \\
&\quad + \frac{1}{4} \left(tr\left(n^2 \left(\boldsymbol{H}_1^{1/2}\right)^T \nabla^2 f(\boldsymbol{x}) \left(\boldsymbol{H}_1\right)^{1/2} \right) \right)^2 \sigma_\varepsilon^2 \\
&\quad + f^2(\boldsymbol{x})\, R(\boldsymbol{x})^2 + f(\boldsymbol{x})\, tr\left(n^2 \left(\boldsymbol{H}_1^{1/2}\right)^T \nabla^2 f(\boldsymbol{x}) \left(\boldsymbol{H}_1\right)^{1/2} \right) R(\boldsymbol{x})^2 \\
&\quad + \frac{1}{4} \left(tr\left(n^2 \left(\boldsymbol{H}_1^{1/2}\right)^T \nabla^2 f(\boldsymbol{x}) \left(\boldsymbol{H}_1\right)^{1/2} \right) \right)^2 R(\boldsymbol{x})^2 . \qquad (3.49)
\end{aligned}
$$

Also, considering the second term of equation (3.48), we have

$$
\begin{aligned}
\boldsymbol{A}^2 = \Bigg[& f(\boldsymbol{x})\, R(\boldsymbol{x}) + \left\{ \frac{1}{2} tr\left(n^2 \left(\boldsymbol{H}_1^{1/2}\right)^T \nabla^2 R(\boldsymbol{x}) \left(\boldsymbol{H}_1\right)^{1/2} \right) f(\boldsymbol{x}) \right. \\
& \left. + tr\left(n^2 \left(\boldsymbol{H}_1^{1/2}\right)^T \nabla f(\boldsymbol{x}) \left(\boldsymbol{H}_1\right)^{1/2} \nabla R(\boldsymbol{x}) \right) + \frac{1}{2} tr\left(n^2 \left(\boldsymbol{H}_1^{1/2}\right)^T \nabla^2 f(\boldsymbol{x}) \boldsymbol{H}_1^{1/2} \right) R(\boldsymbol{x}) \right\} \Bigg]^2
\end{aligned}
$$

Or,

$$
\begin{aligned}
\boldsymbol{A}^2 = \Bigg[& f^2(\boldsymbol{x})\, R(\boldsymbol{x})^2 + 2 f(\boldsymbol{x})\, R(\boldsymbol{x}) \left[\frac{1}{2} tr\left(n^2 \left(\boldsymbol{H}_1^{1/2}\right)^T \nabla^2 R(\boldsymbol{x}) \left(\boldsymbol{H}_1\right)^{1/2} \right) f(\boldsymbol{x}) \right. \\
& + tr\left(n^2 \left(\boldsymbol{H}_1^{1/2}\right)^T \nabla f(\boldsymbol{x}) \left(\boldsymbol{H}_1\right)^{1/2} \nabla R(\boldsymbol{x}) \right) \\
& \left. + \frac{1}{2} tr\left(n^2 \left(\boldsymbol{H}_1^{1/2}\right)^T \nabla^2 f(\boldsymbol{x}) \boldsymbol{H}_1^{1/2} \right) R(\boldsymbol{x}) \right] \qquad (3.50) \\
& + \left[\frac{1}{2} tr\left(n^2 \left(\boldsymbol{H}_1^{1/2}\right)^T \nabla^2 R(\boldsymbol{x}) \left(\boldsymbol{H}_1\right)^{1/2} \right) f(\boldsymbol{x}) \right. \\
& + tr\left(n^2 \left(\boldsymbol{H}_1^{1/2}\right)^T \nabla f(\boldsymbol{x}) \left(\boldsymbol{H}_1\right)^{1/2} \nabla R(\boldsymbol{x}) \right) \\
& \left. + \frac{1}{2} tr\left(n^2 \left(\boldsymbol{H}_1^{1/2}\right)^T \nabla^2 f(\boldsymbol{x}) \boldsymbol{H}_1^{1/2} \right) R(\boldsymbol{x}) \right]^2 \Bigg].
\end{aligned}
$$

Also, considering the third term of equation (3.48), we have

$$= 2R\ \boldsymbol{x}\left[f\ \ \boldsymbol{x}\ \ R\ \ \boldsymbol{x}\ + \frac{1}{2} tr\ \ n^2\ \ \boldsymbol{H}_1^{1/2}\ \ ^T \nabla^2 R\ \ \boldsymbol{x}\ \ \ \boldsymbol{H}_1\ \ ^{1/2} f\ \ \boldsymbol{x} \right.$$

$$\left. + tr\ \ n^2\ \ \boldsymbol{H}_1^{1/2}\ \ ^T \nabla f\ \ \boldsymbol{x}\ \ \ \boldsymbol{H}_1\ \ ^{1/2} \nabla R\ \ \boldsymbol{x}\ \ + \frac{1}{2} tr\ \ n^2\ \ \boldsymbol{H}_1^{1/2}\ \ ^T \nabla^2 f\ \ \boldsymbol{x}\ \ \ \boldsymbol{H}_1^{1/2}\ \ R\ \ \boldsymbol{x} \right]$$

$$\times \left[f\ \ \boldsymbol{x}\ + \frac{1}{2} tr\ \ n^2\ \ \boldsymbol{H}_1^{1/2}\ \ ^T \nabla^2 f\ \ \boldsymbol{x}\ \ \ \boldsymbol{H}_1\ \ ^{1/2} \right]$$

$$= 2R\ \boldsymbol{x}\ \left[f\ \ \boldsymbol{x}\ \ R\ \ \boldsymbol{x}\ + \frac{1}{2} tr \left\{ n^2 \left(\boldsymbol{H}_1^{\frac{1}{2}} \right)^T \nabla^2 R\ \ \boldsymbol{x}\ \ \ \boldsymbol{H}_1\ \ ^{\frac{1}{2}} f\ \ \boldsymbol{x} \right\} \right.$$

$$\left. + tr \left\{ n^2 \left(\boldsymbol{H}_1^{\frac{1}{2}} \right)^T \nabla f\ \ \boldsymbol{x}\ \ \ \boldsymbol{H}_1\ \ ^{\frac{1}{2}} \nabla R\ \ \boldsymbol{x} \right\} + \frac{1}{2} tr \left\{ n^2 \left(\boldsymbol{H}_1^{\frac{1}{2}} \right)^T \nabla^2 f\ \ \boldsymbol{x}\ \left(\boldsymbol{H}_1^{\frac{1}{2}} \right) R\ \ \boldsymbol{x} \right\} \right]$$

$$\times \left[f\ \ \boldsymbol{x}\ + \frac{1}{2} tr \left\{ n^2 \left(\boldsymbol{H}_1^{\frac{1}{2}} \right)^T \nabla^2 f\ \ \boldsymbol{x}\ \ \ \boldsymbol{H}_1\ \ ^{\frac{1}{2}} \right\} \right]$$

$$= \left[2R\ \ \boldsymbol{x}\ \ f\ \ \boldsymbol{x}\ \ R\ \ \boldsymbol{x}\ + 2R\ \ \boldsymbol{x} \left[\frac{1}{2} tr\ \ n^2\ \ \boldsymbol{H}_1^{1/2}\ \ ^T \nabla^2 R\ \ \boldsymbol{x}\ \ \ \boldsymbol{H}_1\ \ ^{1/2} f\ \ \boldsymbol{x} \right. \right.$$

$$\left. \left. + tr\ \ n^2\ \ \boldsymbol{H}_1^{1/2}\ \ ^T \nabla f\ \ \boldsymbol{x}\ \ \ \boldsymbol{H}_1\ \ ^{1/2} \nabla R\ \ \boldsymbol{x}\ \ + \frac{1}{2} tr\ \ n^2\ \ \boldsymbol{H}_1^{1/2}\ \ ^T \nabla^2 f\ \ \boldsymbol{x}\ \ \ \boldsymbol{H}_1^{1/2}\ \ R\ \ \boldsymbol{x} \right] \right]$$

$$\times \left[f\ \ \boldsymbol{x}\ + \frac{1}{2} tr\ \ n^2\ \ \boldsymbol{H}_1^{1/2}\ \ ^T \nabla^2 f\ \ \boldsymbol{x}\ \ \ \boldsymbol{H}_1\ \ ^{1/2} \right]$$

$$= \left[2R\ \ \boldsymbol{x}\ \ ^2 f\ \ \boldsymbol{x}\ \ ^2 + R\ \ \boldsymbol{x}\ \ ^2 f\ \ \boldsymbol{x} \left[\frac{1}{2} tr\ \ n^2\ \ \boldsymbol{H}_1^{1/2}\ \ ^T \nabla^2 f\ \ \boldsymbol{x}\ \ \ \boldsymbol{H}_1\ \ ^{1/2} \right] \right.$$

$$+ 2R\ \ \boldsymbol{x}\ \ f\ \ \boldsymbol{x} \left[\frac{1}{2} tr\ \ n^2\ \ \boldsymbol{H}_1^{1/2}\ \ ^T \nabla^2 R\ \ \boldsymbol{x}\ \ \ \boldsymbol{H}_1\ \ ^{1/2} f\ \ \boldsymbol{x} \right.$$

$$\left. \left. + tr\ \ n^2\ \ \boldsymbol{H}_1^{1/2}\ \ ^T \nabla f\ \ \boldsymbol{x}\ \ \ \boldsymbol{H}_1\ \ ^{1/2} \nabla R\ \ \boldsymbol{x}\ \ + \frac{1}{2} tr\ \ n^2\ \ \boldsymbol{H}_1^{1/2}\ \ ^T \nabla^2 f\ \ \boldsymbol{x}\ \ \ \boldsymbol{H}_1^{1/2}\ \ R\ \ \boldsymbol{x} \right] \right]. \quad (3.51)$$

Now, substituting equations (3.49), (3.50) and (3.51) into equation (3.48) and carefully simplifying, we have

$$
\begin{aligned}
Var\left(\hat{R}(\boldsymbol{x})\right) &= \frac{R(\boldsymbol{x})\, f(\boldsymbol{x})}{n^{\frac{p}{2}}\left|\boldsymbol{H}_1\right|^{1/2}\left[f(\boldsymbol{x})+\frac{1}{2}tr\left(n^2\left(\boldsymbol{H}_1^{1/2}\right)^T\nabla^2 f(\boldsymbol{x})\left(\boldsymbol{H}_1\right)^{1/2}\right)\right]^4}\times \\
&\Bigg[f^2(\boldsymbol{x})\,\sigma_\varepsilon^2 + f(\boldsymbol{x})\, tr\left(n^2\left(\boldsymbol{H}_1^{1/2}\right)^T\nabla^2 f(\boldsymbol{x})\left(\boldsymbol{H}_1\right)^{1/2}\right)\sigma_\varepsilon^2 + \frac{1}{4}\left\{tr\left(n^2\left(\boldsymbol{H}_1^{1/2}\right)^T\nabla^2 f(\boldsymbol{x})\left(\boldsymbol{H}_1\right)^{1/2}\right)\right\}^2\sigma_\varepsilon^2 \\
&+\frac{1}{4}\left\{tr\left(n^2\left(\boldsymbol{H}_1^{1/2}\right)^T\nabla^2 f(\boldsymbol{x})\left(\boldsymbol{H}_1\right)^{1/2}\right)\right\}^2 R(\boldsymbol{x})^2 \\
&+\Bigg[\frac{1}{2}tr\left(n^2\left(\boldsymbol{H}_1^{1/2}\right)^T\nabla^2 R(\boldsymbol{x})\left(\boldsymbol{H}_1\right)^{1/2}\right) f(\boldsymbol{x}) \\
&+tr\left(n^2\left(\boldsymbol{H}_1^{1/2}\right)^T\nabla f(\boldsymbol{x})\left(\boldsymbol{H}_1\right)^{1/2}\nabla R(\boldsymbol{x})\right)+\frac{1}{2}tr\left\{n^2\left(\boldsymbol{H}_1^{1/2}\right)^T\nabla^2 f(\boldsymbol{x})\left(\boldsymbol{H}_1^{1/2}\right)R(\boldsymbol{x})\right\}\Bigg]^2 \\
&-R(\boldsymbol{x})\left[\frac{1}{2}tr\left\{n^2\left(\boldsymbol{H}_1^{1/2}\right)^T\nabla^2 f(\boldsymbol{x})\left(\boldsymbol{H}_1\right)^{1/2}\right\}\right] \\
&\times\Bigg[\frac{1}{2}tr\left\{n^2\left(\boldsymbol{H}^{1/2}\right)^T\nabla^2 R(\boldsymbol{x})\left(\boldsymbol{H}\right)^{1/2} f(\boldsymbol{x})\right\} \\
&+tr\left\{n^2\left(\boldsymbol{H}_1^{1/2}\right)^T\nabla f(\boldsymbol{x})\left(\boldsymbol{H}_1\right)^{1/2}\nabla R(\boldsymbol{x})\right\}+\frac{1}{2}tr\left\{n^2\left(\boldsymbol{H}_1^{1/2}\right)^T\nabla^2 f(\boldsymbol{x})\left(\boldsymbol{H}_1^{1/2}\right)R(\boldsymbol{x})\right\}\Bigg]\Bigg].
\end{aligned}
\tag{3.52}
$$

Assuming K is a second order kernel, equation (3.52) becomes

$$Var\left(\hat{R}(\boldsymbol{x})\right)$$

$$=\left[\frac{R(K)f(\boldsymbol{x})}{n^{p/2}\left|\boldsymbol{H}_1\right|^{1/2}\left[f(\boldsymbol{x})+\frac{1}{2}tr\left(n^{2}\left(\boldsymbol{H}_1^{1/2}\right)^{T}\nabla^{2}f(\boldsymbol{x})\left(\boldsymbol{H}_1\right)^{1/2}\right)\right]^{4}}\right]$$
$$\times\left[f^{2}(\boldsymbol{x})\sigma_{\varepsilon}^{2}+f(\boldsymbol{x})\,tr\left(n^{2}\left(\boldsymbol{H}_1^{1/2}\right)^{T}\nabla^{2}f(\boldsymbol{x})\left(\boldsymbol{H}_1\right)^{1/2}\right)\sigma_{\varepsilon}^{2}\right]$$

$$=\left[\frac{R(K)f(\boldsymbol{x})}{n^{p/2}\left|\boldsymbol{H}_1\right|^{1/2}f^{2}(\boldsymbol{x})\left[f^{2}(\boldsymbol{x})+2f(\boldsymbol{x})\,tr\left(n^{2}\left(\boldsymbol{H}_1^{1/2}\right)^{T}\nabla^{2}f(\boldsymbol{x})\left(\boldsymbol{H}_1\right)^{1/2}\right)\right]}\right]$$
$$\times f(\boldsymbol{x})\left[f(\boldsymbol{x})\sigma_{\varepsilon}^{2}+tr\left(n^{2}\left(\boldsymbol{H}_1^{1/2}\right)^{T}\nabla^{2}f(\boldsymbol{x})\left(\boldsymbol{H}_1\right)^{1/2}\right)\sigma_{\varepsilon}^{2}\right]$$

$$=\frac{R(K)\left[f(\boldsymbol{x})\sigma_{\varepsilon}^{2}+tr\left(n^{2}\left(\boldsymbol{H}_1^{1/2}\right)^{T}\nabla^{2}f(\boldsymbol{x})\left(\boldsymbol{H}_1\right)^{1/2}\right)\sigma_{\varepsilon}^{2}\right]}{n^{p/2}\left|\boldsymbol{H}_1\right|^{1/2}\left[f^{2}(\boldsymbol{x})+tr\left(n^{2}\left(\boldsymbol{H}_1^{1/2}\right)^{T}\nabla^{2}f(\boldsymbol{x})\left(\boldsymbol{H}_1\right)^{1/2}\right)\right]}$$

$$=\frac{R(K)\left[f(\boldsymbol{x})\sigma_{\varepsilon}^{2}+tr\left(n^{2}\left(\boldsymbol{H}_1^{1/2}\right)^{T}\nabla^{2}f(\boldsymbol{x})\left(\boldsymbol{H}_1\right)^{1/2}\right)\sigma_{\varepsilon}^{2}\right]}{n^{p/2}\left|\boldsymbol{H}_1\right|^{1/2}f^{2}(\boldsymbol{x})\left[1+\frac{tr\left(n^{2}\left(\boldsymbol{H}_1^{1/2}\right)^{T}\nabla^{2}f(\boldsymbol{x})\left(\boldsymbol{H}_1\right)^{1/2}\right)}{f(\boldsymbol{x})}\right]}$$

$$=\frac{R(K)}{n^{p/2}\left|\boldsymbol{H}_1\right|^{1/2}f^{2}(\boldsymbol{x})}\left[f(\boldsymbol{x})\sigma_{\varepsilon}^{2}+tr\left(n^{2}\left(\boldsymbol{H}_1^{1/2}\right)^{T}\nabla^{2}f(\boldsymbol{x})\left(\boldsymbol{H}_1\right)^{1/2}\right)\sigma_{\varepsilon}^{2}\right]\left[1+\frac{tr\left(n^{2}\left(\boldsymbol{H}_1^{1/2}\right)^{T}\nabla^{2}f(\boldsymbol{x})\left(\boldsymbol{H}_1\right)^{1/2}\right)}{f(\boldsymbol{x})}\right]^{-1}.$$

Now, using the approximation $(1+t^2c)^{-1} \approx (1-t^2c)$ for small c, the equation above becomes

$$\frac{R(\boldsymbol{x})}{n^{p/2}|\boldsymbol{H}_1|^{1/2} f^2(\boldsymbol{x})}\left[f(\boldsymbol{x})\sigma_\varepsilon^2 + tr\left(n^2 \left(\boldsymbol{H}_1^{1/2}\right)^T \nabla^2 f(\boldsymbol{x}) \left(\boldsymbol{H}_1\right)^{1/2}\right)\sigma_\varepsilon^2\right]\left[1-\frac{tr\left(n^2 \left(\boldsymbol{H}_1^{1/2}\right)^T \nabla^2 f(\boldsymbol{x}) \left(\boldsymbol{H}_1\right)^{1/2}\right)}{f(\boldsymbol{x})}\right]$$

$$=\frac{R(K)}{n^{p/2}|\boldsymbol{H}_1|^{1/2} f^2(\boldsymbol{x})}\Bigg[f(\boldsymbol{x})\sigma_\varepsilon^2 - \sigma_\varepsilon^2 tr\left(n^2 \left(\boldsymbol{H}_1^{1/2}\right)^T \nabla^2 f(\boldsymbol{x}) \left(\boldsymbol{H}_1\right)^{1/2}\right) + tr\left(n^2 \left(\boldsymbol{H}_1^{1/2}\right)^T \nabla^2 f(\boldsymbol{x}) \left(\boldsymbol{H}_1\right)^{1/2}\right)\sigma_\varepsilon^2$$
$$-\frac{2\left[tr\left(n^2 \left(\boldsymbol{H}_1^{1/2}\right)^T \nabla^2 f(\boldsymbol{x}) \left(\boldsymbol{H}_1\right)^{1/2}\right)\sigma_\varepsilon^2\right]\left[2tr\left(n^2 \left(\boldsymbol{H}_1^{1/2}\right)^T \nabla^2 f(\boldsymbol{x}) \left(\boldsymbol{H}_1\right)^{1/2}\right)\right]}{f(\boldsymbol{x})}\Bigg]. \quad (3.53)$$

Assuming K is a second order kernel, we have that the right hand side of (3.53) is equal to

$$=\frac{R(K)}{n^{p/2}|\boldsymbol{H}_1|^{1/2} f^2(\boldsymbol{x})}\Bigg[f(\boldsymbol{x})\sigma_\varepsilon^2 - \sigma_\varepsilon^2 tr\left(n^2 \left(\boldsymbol{H}_1^{1/2}\right)^T \nabla^2 f(\boldsymbol{x}) \left(\boldsymbol{H}_1\right)^{1/2}\right)$$
$$+ tr\left(n^2 \left(\boldsymbol{H}_1^{1/2}\right)^T \nabla^2 f(\boldsymbol{x}) \left(\boldsymbol{H}_1\right)^{1/2}\right)\sigma_\varepsilon^2\Bigg] = \frac{R(K)}{n^{p/2}|\boldsymbol{H}_1|^{1/2} f^2(\boldsymbol{x})}\left[f(\boldsymbol{x})\sigma_\varepsilon^2\right].$$

Hence

$$Var\left(\hat{R}(\boldsymbol{x})\right) = \frac{R(K)\sigma_\varepsilon^2}{n^{p/2}|\boldsymbol{H}_1|^{1/2} f(\boldsymbol{x})}. \quad (3.54)$$

This variance includes the amount of data through $f(\boldsymbol{x})$ and it also includes the factor relating to the noise variance σ_ε.

The asymptotic mean square error (AMSE) of the estimator is

$$AMSE\left(\hat{R}(\boldsymbol{x})\right) = \frac{R(K)\sigma_\varepsilon^2}{n^{p/2}|\boldsymbol{H}_1|^{1/2} f(\boldsymbol{x})}$$
$$+\frac{1}{4}\left[tr\left\{\left(n\boldsymbol{H}_1^{1/2}\right)^T\left(\boldsymbol{H}_1\right)^{1/2}\right\}\left\{\nabla^2 R(\boldsymbol{x}) + 2\nabla R(\boldsymbol{x})\frac{\nabla f(\boldsymbol{x})}{f(\boldsymbol{x})}\right\}\right]^2. \quad (3.55)$$

This is the AMSE of the multivariate regression estimator based on the sample mean. It minimizes the asymptotic variance as the dimension increases thereby minimizing the asymptotic mean square error. It has the form of the AMSE of the Nadaraya-Watson regression estimator for the classical kernel case when it is generalized from the univariate to the multivariate case.

Now, in our special case where $\boldsymbol{H}_1 = h^2\boldsymbol{Q}_1$ and $|\boldsymbol{Q}_1| = 1$, the $AMSE\left(\hat{R}(\boldsymbol{x})\right)$ is given by

$$AMSE\left(\hat{R}(\boldsymbol{x})\right) = \frac{R(K)\sigma_\varepsilon^2}{n^{p/2}h^p f(\boldsymbol{x})}$$
$$+\frac{1}{4}n^2h^4\left[tr\left\{\left(\boldsymbol{Q}_1^{1/2}\right)^T\boldsymbol{Q}_1^{1/2}\right\}\left\{\nabla^2 R(\boldsymbol{x}) + 2\nabla R(\boldsymbol{x})\frac{\nabla f(\boldsymbol{x})}{f(\boldsymbol{x})}\right\}\right]^2. \quad (3.56)$$

Therefore, the optimum bandwidth h for the regression function can be obtained by differentiating equation (3.56), equating it to zero and solving for h.

So,

$$h_{opt} = n^{-1/2}\left[\frac{pR(K)\sigma_\varepsilon^2}{f(x)\left\{tr\left\{\left(\boldsymbol{Q}_1^{1/2}\right)^T\boldsymbol{Q}_1^{1/2}\right\}\left\{\nabla^2 R(\boldsymbol{x}) + 2\nabla R(\boldsymbol{x})\frac{\nabla f(\boldsymbol{x})}{f(\boldsymbol{x})}\right\}\right\}^2}\right]^{1/(4+p)}. \quad (3.57)$$

Therefore, in this multivariate regression setting just like the special case of the multivariate density estimation, the optimum bandwidth $h_{opt} \to n^{-1/2}$ as $p \to \infty$. This means that the optimum bandwidth does not depend on the dimension as the dimension $p \to \infty$ and this optimum bandwidth decreases as sample size increases to allow for a better performance of the estimator. This breaks the "curse of dimensionality" and remedies the deficiency of high dimensional bandwidth selection as the optimum bandwidth is smaller enough for large dimension to guarantee smoothness.

CHAPTER IV

BANDWIDTH SELECTION METHODOLOGIES BASED ON THE SAMPLE MEANS

4.1 Cross-Validation for Density Estimation Based on the Sample Means-Univariate Case

It follows from the multivariate case with $p=1$ that the estimate of density $f(x)$ based on the sample means is given by

$$\hat{f}(x)=\frac{1}{h\sqrt{n}}K\left(\frac{x-\sqrt{n}\bar{X}}{\sqrt{n}h}\right), \tag{4.1}$$

where h is the window width, also called the smoothing parameter or the bandwidth, and K is the kernel function usually assumed to be symmetric.

Now, the expectation of the density estimate can be obtained as

$$E[\hat{f}(x)]\approx f(x)\int K(u)du-h\sqrt{n}f'(x)\int uK(u)du+\frac{1}{2}h^2nf''(x)\int u^2K(u)du.$$

Applying the moment conditions for a univariate kernel, we have

$$E[\hat{f}(x)]\approx f(x)+\frac{1}{2}h^2nf''(x)\sigma_K^2. \tag{4.2}$$

Therefore, the bias and variance of $\hat{f}(x)$ are given by

$$Bias\left[\hat{f}(x)\right] = \frac{f''(x)}{2} h^2 n \sigma_K^2, \tag{4.3}$$

$$Var\left[\hat{f}(x)\right] \approx \frac{f(x)R(K)}{h\sqrt{n}}. \tag{4.4}$$

Therefore, the asymptotic integrated variance (AIV) is given by

$$AIV = \frac{R(K)}{h\sqrt{n}}. \tag{4.5}$$

The asymptotic integrated squared bias (AISB) is given by

$$AISB = \frac{1}{4} h^4 n^2 \sigma_K^4 \int \left[f''(x)\right]^2 dx$$

$$= \frac{1}{4} h^4 n^2 \sigma_K^4 R(f''), \tag{4.6}$$

where $R(f'') = \int \left[f''(x)\right]^2 dx$.

Therefore, the asymptotic means integrated squared error (AMISE) is given by

$$AMISE = AIV + AISB$$

$$AMISE = \frac{R(K)}{h\sqrt{n}} + \frac{1}{4} h^4 n^2 \sigma_K^4 R(f''). \tag{4.7}$$

Thus the optimum bandwidth is obtained via calculus and is equal to

$$h_{opt} = n^{-1/2} \left\{ \frac{R(K)}{\sigma_K^4 R(f'')} \right\}^{1/5}.$$

Recall that in the standard univariate case, we have

$$AMISE^{*} = \frac{R(K)}{nh} + \frac{1}{4}h^{4}\sigma_K^{4}R(f'')$$

and

$$h^{*}_{opt} = n^{-1/5}\left\{\frac{R(K)}{\sigma_K^{4}R(f'')}\right\}^{1/5}.$$

Cross validation is one of the methods used to estimate how accurate a predictive model will perform in practice. Therefore, to ascertain practical data-based algorithms and to prove beyond theoretical results for optimal bandwidth specification, we perform cross validation. Here, we examined both unbiased (Least Squares) and biased cross-validation.

4.1.1 Least Squares (Unbiased) Cross Validation (LSCV)

The motivation of the least squares cross-validation method of bandwidth selection comes from expanding the mean integrated square error (MISE) of $\hat{f}(x,h)$ and the minimization of it.

$$MISE\left(\hat{f}(x,h)\right) = E\left[\int\left(\hat{f}(x,h) - f(x)\right)^{2}dx\right]$$

$$MISE\left(\hat{f}(x,h)\right) = E\left[\int\left(\hat{f}(x,h)\right)^{2}dx\right] - 2E\left[\int \hat{f}(x,h)f(x)dx\right] + \int\left(f(x)\right)^{2}dx. \quad (4.8)$$

Since the $\int\left(f(x)\right)^{2}dx$ term does not depend on h, minimization of the $MISE\left(\hat{f}(x,h)\right)$ is the same as minimization of $MISE\left(\hat{f}(x,h)\right) - \int f^{2}(x)dx$.

Therefore,

$$MISE\left[\hat{f}(x,h)\right]-\int f^2(x)\,dx = E\left[\int \hat{f}^2(x,h)\,dx - 2\int \hat{f}(x,h)f(x)\,dx\right]. \tag{4.9}$$

The LSCV is obtained by finding the unbiased estimate of the right hand side of equation (4.9) since the second term is unknown due to the fact that it depends on $f(x)$. Using the method of cross validation, it is suggested that we remove one sample mean and use the remaining $n-1$ sample means to construct the estimate (leave-one-out). To determine the quality of the fit, the nth sample mean is then evaluated. The afore-illustrated procedure is then repeated n times, one for each sample mean and the results averaged.

Therefore, the unbiased estimator of the right hand side of equation (4.9) is obtained as follows; for simplicity, $n-1$ has been replaced by n;

$$\int \hat{f}^2(x,h)\,dx - 2\int \hat{f}(x,h)f(x)\,dx$$

$$= \int \hat{f}^2(x,h)\,dx - 2\int \hat{f}(x,h)\,dF_n(x)$$

$$= I_1 - 2I_2 \tag{4.10}$$

where $I_1 = \int \hat{f}^2(x,h)\,dx$ and $I_2 = E\int \hat{f}_{-1}(x,h)\,dF_n(x)\,dx.$

Now, considering $I_1,$ we have

$$I_1 = \int \hat{f}^2(x,h)\,dx = \frac{1}{h^2 n}\int K^2\left(\frac{x-\sqrt{n}\overline{X}}{h\sqrt{n}}\right)dx = \frac{1}{h\sqrt{n}}\int K^2\left(\frac{x}{h\sqrt{n}}-\frac{\overline{X}}{h}\right)d\left(\frac{x}{h\sqrt{n}}\right)$$

Let $u = \frac{x}{\sqrt{n}h} - \frac{\overline{X}}{h} \Rightarrow du = d\left(\frac{x}{h\sqrt{n}}\right).$

Therefore,

$$I_1 = \frac{1}{h\sqrt{n}} \int K^2(u)\,du = \frac{R(K)}{h\sqrt{n}}. \tag{4.11}$$

Also, $$I_2 = \int \hat{f}_{-1}(x,h)\,dF_n(x)\,dx = \frac{1}{n}\sum_{i=1}^{n} \hat{f}_{-i}(x,h) = \left[\frac{1}{h(\sqrt{n})^3} \cdot \sum_{i=1}^{n} K\left(\frac{X_i - \sqrt{n}\,\bar{X}_{-i}}{h\sqrt{n}}\right)\right]. \tag{4.12}$$

Therefore, substituting equations (4.11) and (4.12) into equation (4.10), we have

$$LSCV(h) = \frac{R(K)}{h\sqrt{n}} - \frac{2}{h(\sqrt{n})^3}\sum_{i=1}^{n} K\left(\frac{X_i - \sqrt{n}\,\bar{X}_{-i}}{h\sqrt{n}}\right). \tag{4.13}$$

Thus, it seems reasonable to choose h to minimize $LSCV(h)$.

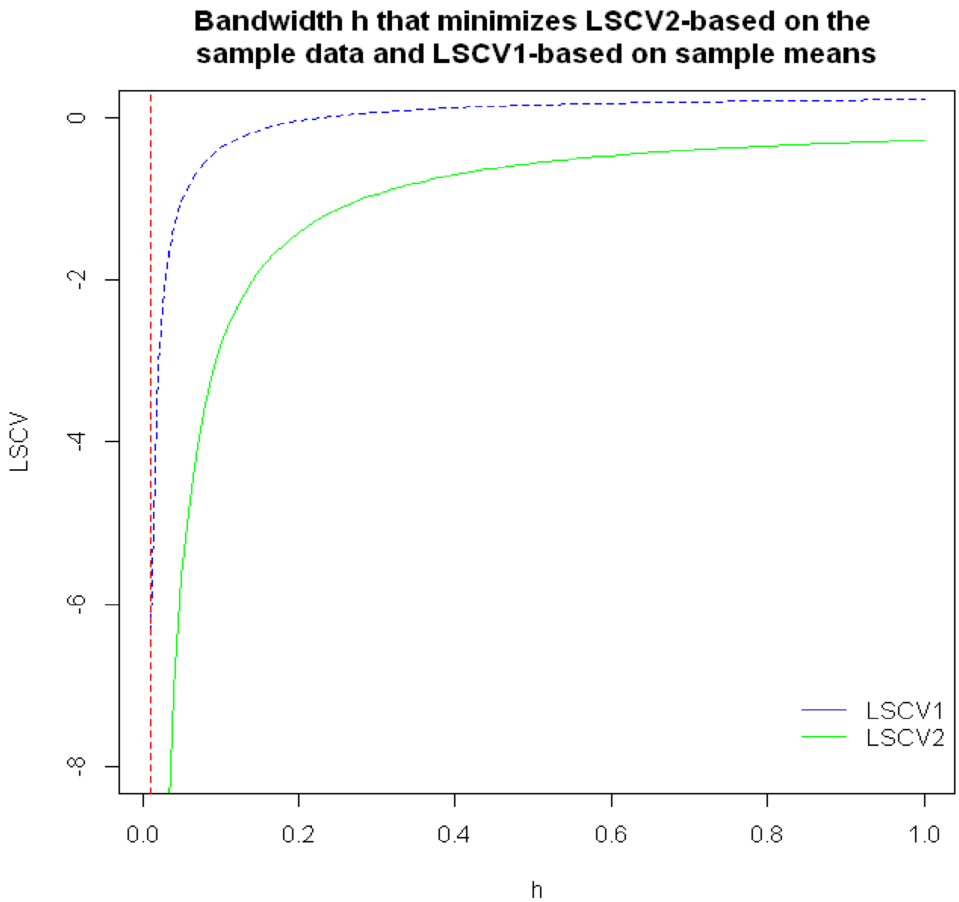
The aforementioned smoothing parameter obtained by the unbiased cross-validation based on the sample means was then compared to that based on the sample data proposed by Wand and Jones (1995); that is

$$LSCV^*(h) = \int \hat{f}^2(x;h)\,dx - 2n^{-1}\hat{f}_{-i}(X_i;h), \tag{4.14}$$

where $\hat{f}_{-i}(x;h) = (n-1)^{-1}\sum_{j \neq i}^{n} K_h(x - X_j)$ is the density estimated on the sample with X_i deleted, often called the “leave-one-out” density estimator.

The plot below shows the results of the comparison. A symmetric kernel function (the standard normal kernel) was used for the comparison.

Figure 4.1.1 Plots Comparing the Least Squares Cross Validation (LSCV) of the Univariate Standard Kernel Estimate Case and Univariate Kernel Estimate Based on the Sample Mean



It was observed from the plot that the smoothing parameter obtained by the unbiased cross-validation based on sample means converges to the optimum bandwidth quicker than the one based on the sample data. Also it was detected that, even for very small LSCV values, the smoothing parameter based on the sample data is quite a bit of a far from the optimum bandwidth.

4.1.2 Biased Cross Validation (BCV) in the Univariate Case

Biased cross-validation (BCV) (Scott and Terrell, 1987) is based on the formula for the asymptotic mean integrated squared error (AMISE) instead of the MISE. Proceeding to our case,

$$AMISE = \frac{R(K)}{h\sqrt{n}} + \frac{1}{4}h^4 n^2 \sigma_K^4 R(f'').$$

In the above asymptotic expansion for the MISE, the only unknown quantity is $R(f'')$. This unknown quantity can be replaced by the estimator $\tilde{R}(f'')$ to get

$$BCV(h) = \frac{R(K)}{h\sqrt{n}} + \frac{1}{4}h^4 n^2 \sigma_K^4 \tilde{R}(f''). \tag{4.15}$$

Now, to estimate $R(f'')$, we know that

$$R(f'') = \int f''^2(x)dx = \int f''(x)f''(x)dx.$$

Applying integration by parts,

$$R(f'') = f'(x)f''(x)\Big|_{-\infty}^{\infty} - \int_{-\infty}^{\infty} f'(x)f'''(x)dx = -\int_{-\infty}^{\infty} f'(x)f'''(x)dx.$$

Applying integration by parts again,

$$R(f'') = -f'''(x)f(x)\Big|_{-\infty}^{\infty} + \int_{-\infty}^{\infty} f(x)f^{(4)}(x)dx = \int_{-\infty}^{\infty} f(x)f^{(4)}(x)dx.$$

Therefore,

$$\tilde{R}(f'') = \int_{-\infty}^{\infty} \hat{f}^{(4)}(x)dF_n(x) = \frac{1}{n}\sum_{i=1}^{n} \hat{f}^{(4)}(x_i). \tag{4.16}$$

But $\hat{f}(x) = \frac{1}{h\sqrt{n}} K\left(\frac{x-\sqrt{n}\bar{X}}{\sqrt{n}h}\right)$, so

$$\hat{f}^{(4)}(x) = \left(\frac{1}{h\sqrt{n}}\right)^5 K^{(4)}\left(\frac{x-\sqrt{n}\bar{X}}{\sqrt{n}h}\right). \tag{4.17}$$

Substituting equation (4.17) into equation (4.16), we have

$$\tilde{R}(f'') = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{h\sqrt{n}}\right)^{5} K^{(4)}\left(\frac{X_i - \sqrt{n}\bar{X}}{\sqrt{n}h}\right). \tag{4.18}$$

Now, substituting equation (4.18) into equation (4.15), we have

$$BCV(h) = \frac{R(K)}{h\sqrt{n}} + \frac{1}{4hn^{\frac{3}{2}}}\sigma_K^4 \sum_{i=1}^{n} K^{(4)}\left(\frac{X_i - \sqrt{n}\bar{X}}{h\sqrt{n}}\right). \tag{4.19}$$

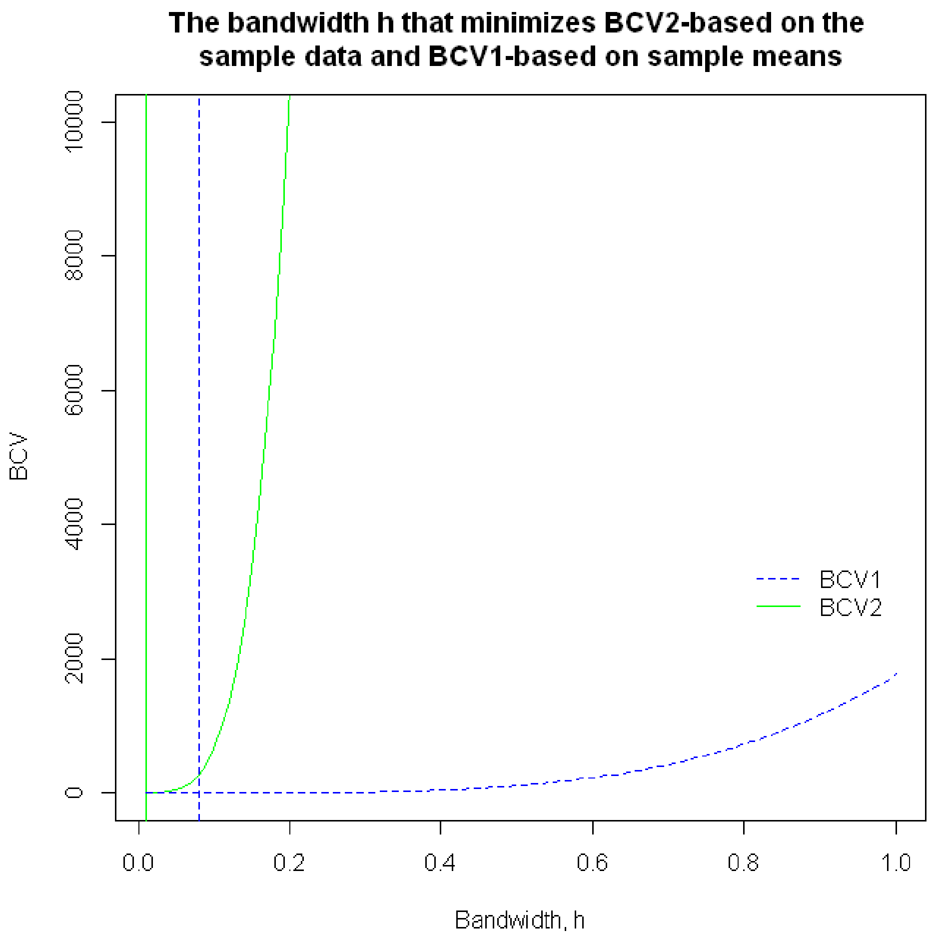
The aforementioned smoothing parameter obtained by the biased cross- validation based on the sample mean was then compared to that based on the sample data proposed by Wand and Jones (1995), that is

$$BCV^{*}(h) = (nh)^{-1} R(K) + \frac{1}{4}h^{4}\mu_2(K)^{2}\hat{R}(f''), \tag{4.20}$$

where $\hat{R}(f'') = n^{-2}\sum\sum_{i\neq j}(K_h'' * K_h'')(X_i - X_j)$.

The plot below shows the results of the comparison. A symmetric kernel function (the standard normal kernel) was used for the comparison.

Figure 4.1.2 Plots Comparing the Biased Cross Validation (BCV) of the Univariate Standard Kernel Estimate Case and Univariate Kernel Estimate Based on the Sample Mean



From the graph, it was observed that, even though the optimum bandwidth for the biased cross-validation based on the sample data $(h_{opt} = 0.01)$ is a little smaller than that based on the sample mean $(h_{opt} = 0.08)$, they both approach their optimums at the same BCV value.

It was observed from the plot that the density of the biased cross-validation based on the sample data increases monotonically as the smoothing parameter increases whilst that of the one based on the sample means increases gradually or steadily as the smoothing parameter increases.

4.2. CROSS-VALIDATION OF MULTIVARIATE DENSITIES BASED ON THE SAMPLE MEANS

Many studies of bandwidth selection or smoothing parameter selection for kernel density estimation have been centered on the univariate case. Not too much has been done on the multivariate due to the complexity of the situation. Here, we will do an explicit multivariate derivation of both least-squares (unbiased) and biased cross-validation based on the sample means using the product kernel estimator.

The multivariate product kernel estimator of $f(\boldsymbol{x})$ based on the sample means is given by

$$\hat{f}(\boldsymbol{x}) = \frac{1}{h_1 \dots h_p n^{p/2}} \prod_{i=1}^{p} K\left(\frac{x_i - \sqrt{n}\bar{X}_j}{\sqrt{n} h_i} \right), \tag{4.21}$$

where $\bar{X}_j$ denote the jth entry $\bar{\boldsymbol{X}}$, and $\boldsymbol{X}$ *is a* $n \times p$ data matrix of random vectors, $\boldsymbol{x} = (x_1, \dots, x_p)'$ is a point in R^p and K is the univariate symmetric kernel with a different smoothing parameter for each dimension.

4.2.1 Least Squares (Unbiased) Cross Validation (UCV)

The unbiased cross-validation is obtained by minimizing the integrated squared error (ISE). The ISE is given by

$$ISE(\boldsymbol{h}) = \int \left[\hat{f}(\boldsymbol{x}, \boldsymbol{h}) - f(\boldsymbol{x}) \right]^2 d\boldsymbol{x}. \tag{4.22}$$

Expanding equation (4.20), we have

$$ISE(\boldsymbol{h}) = \int\left[\hat{f}(\boldsymbol{x},\boldsymbol{h})\right]^2 dx - 2\int \hat{f}(\boldsymbol{x},\boldsymbol{h}) f(\boldsymbol{x})\, d\boldsymbol{x} + \int\left[f(\boldsymbol{x})\right]^2 d\boldsymbol{x}\,.$$

This implies,

$$ISE(\boldsymbol{h}) = R\left(\hat{f}(\boldsymbol{x},\boldsymbol{h})\right) - 2\int \hat{f}(\boldsymbol{x},\boldsymbol{h}) f(\boldsymbol{x})\, d\boldsymbol{x} + R\left(f(\boldsymbol{x})\right), \tag{4.23}$$

where $R\left(\hat{f}(\boldsymbol{x},\boldsymbol{h})\right) = \int\left[\hat{f}(\boldsymbol{x},\boldsymbol{h})\right]^2 d\boldsymbol{x}$, $R\left(f(\boldsymbol{x})\right) = \int\left[f(\boldsymbol{x})\right]^2 d\boldsymbol{x}$, and $R\left(\psi(\boldsymbol{x})\right) = \int\ldots\int \psi(\boldsymbol{x})^2 dx$ and it's the roughness of $\psi(\cdot)$.

The term $R\left(f(\boldsymbol{x})\right) = \int\left[f(\boldsymbol{x})\right]^2 d\boldsymbol{x}$ is independent of $\boldsymbol{h}$ and can therefore be ignored.

Now, applying the cross-validation using the leave-one-out estimator, we have

$$\hat{f}_{-j}(\boldsymbol{x}) = \frac{1}{h_1 \ldots h_p (n-1)^{p/2}} \prod_{i=1}^{p} K\left(\frac{x_i - \sqrt{n-1}\,\overline{X}_{i,-j}}{h_i\sqrt{n-1}}\right), \tag{4.24}$$

where $j = 1,\ldots,n$.

But $E\hat{f}_{-j}(\boldsymbol{X}_j) = \int \hat{f}(\boldsymbol{x},\boldsymbol{h}) f(\boldsymbol{x})\, d\boldsymbol{x}$ hence $\hat{f}_{-j}(\boldsymbol{x})$ estimates the second term of equation (4.23).

Therefore, the UCV estimate is given by

$$UCV(\boldsymbol{h}) = R(\hat{f}) - 2Ef_{-j}(\boldsymbol{X}_j).$$

Therefore,

$$UCV(\boldsymbol{h}) = R(\hat{f}) - 2\int \hat{f}(\boldsymbol{x},\boldsymbol{h}) f(\boldsymbol{x})\, d\boldsymbol{x}\,. \tag{4.25}$$

We can therefore generalize the unbiased cross-validation to arbitrary dimensions with this approach. For a symmetric kernel, the standard normal kernel, $R(\hat{f})$ is obtained as in Sain and Scott (1994), but here, it is based on the sample means and derived as follows:

$$R(\hat{f}(\boldsymbol{x},\boldsymbol{h})) = \int [\hat{f}(\boldsymbol{x},\boldsymbol{h})]^2 d\boldsymbol{x}$$

$$= \int \left[\frac{1}{h_1 \ldots h_p (n)^{p/2}} \prod_{i=1}^{p} K\left(\frac{x_i - \sqrt{n}\bar{X}_i}{\sqrt{n}h_i} \right) \right]^2 dx_i$$

$$= \frac{1}{\left(n^{p/2} \prod_{i=1}^{p} h_i \right)^2} \prod_{i=1}^{p} \int K^2 \left(\frac{x_i - \sqrt{n}\bar{X}_i}{\sqrt{n}h_i} \right) dx_i . \qquad (4.26)$$

Now, let $u = \frac{x_i - \sqrt{n}\bar{X}_i}{\sqrt{n}h_i}$, then $x_i = \sqrt{n}h_i u + \sqrt{n}\bar{X}_i$, which implies $dx_i = \sqrt{n}h_i du$.

Substituting these into equation (4.24) we get

$$R(\hat{f}) = \frac{1}{\left((n)^{p/2} \prod_{i=1}^{p} h_i \right)^2} \left((n)^{p/2} \prod_{i=1}^{p} h_i \right) \prod_{i=1}^{p} \int K^2 (u) du$$

$$= \frac{1}{\left((n)^{p/2} \prod_{i=1}^{p} h_i \right)} (R(K))^p .$$

But for the standard normal kernel, $R(K) = \frac{1}{2\sqrt{\pi}}$. Therefore,

$$R(\hat{f}) = \frac{1}{(2\sqrt{\pi})^p \left(n^{p/2} \prod_{i=1}^{p} h_i \right)} . \qquad (4.27)$$

Also, the second term of equation (4.21) is obtained as follows:

$$\int \hat{f}(\boldsymbol{x},\boldsymbol{h})\hat{f}_{-j}(\boldsymbol{x})d\boldsymbol{x}$$

$$=\frac{1}{n}\sum_{j=1}^{n}\int\frac{1}{h_1...h_p(n)^{p/2}}\prod_{i=1}^{p}K\left(\frac{x_i-\sqrt{n}\overline{X}_i}{\sqrt{n}h_i}\right).\frac{1}{h_1...h_p(n-1)^{p/2}}\prod_{i=1}^{p}K\left(\frac{x_i-\sqrt{n-1}\overline{X}_{i,-j}}{\sqrt{n-1}h_i}\right)dx_i$$

$$=\frac{1}{\left(\prod_{i=1}^{p}h_i\right)^2 nn^{p/2}(n-1)^{p/2}}\sum_{j=1}^{n}\prod_{i=1}^{p}\int K\left(\frac{x_i-\sqrt{n}\overline{X}_i}{\sqrt{n}h_i}\right).K\left(\frac{x_i-\sqrt{n-1}\overline{X}_{i,-j}}{\sqrt{n-1}h_i}\right)dx_i\ .$$

Now, let $u=\frac{x_i-\sqrt{n}\overline{X}_i}{\sqrt{n}h_i}$, then $x_i=\sqrt{n}h_iu+\sqrt{n}\overline{X}_i$, which implies $dx_i=\sqrt{n}h_idu$.

Substituting these into the above equation, we get

$$\frac{1}{\left(\prod_{i=1}^{p}h_i\right)n(n-1)^{p/2}}\sum_{j=1}^{n}\prod_{i=1}^{p}\int K(u)K\left(\frac{\sqrt{n}h_iu+\sqrt{n}\overline{X}_i-\sqrt{n-1}\overline{X}_{i,-j}}{\sqrt{n-1}h_i}\right)du\ .$$

Therefore, the above kernel convolution can be written as

$$\frac{1}{\left(\prod_{i=1}^{p}h_i\right)n(n-1)^{p/2}}\sum_{j=1}^{n}\prod_{i=1}^{p}K*K\left(\frac{\sqrt{n}\overline{X}_i-\sqrt{n-1}\overline{X}_{i,-j}}{\sqrt{n-1}h_i}\right).$$

Now, replacing $n-1$ with n for simplicity, the multivariate leased squared cross validation function based on the sample means is then

$$UCV(h_1,...,h_p)$$

$$=\frac{1}{2\sqrt{\pi}^{\,p}\left(n^{p/2}\prod_{i=1}^{p}h_i\right)}-\frac{2}{\left(\prod_{i=1}^{p}h_i\right)nn^{p/2}}\sum_{j=1}^{n}\prod_{i=1}^{p}K*K\left(\frac{\sqrt{n}\overline{X}_i-\sqrt{n}\overline{X}_{i,-j}}{\sqrt{n}h_i}\right). \qquad (4.28)$$

For a bivariate case, that is when $p = 2$, we have

$$UCV(h_1, h_2) = \frac{1}{4\pi n h_1 h_2} - \frac{2}{h_1 h_2 n^2} \sum_{j=1}^{n} \prod_{i=1}^{2} K * K\left(\frac{\sqrt{n}\bar{X}_i - \sqrt{n}\bar{X}_{i,-j}}{\sqrt{n} h_i} \right). \qquad (4.29)$$

Another way of estimating $\int \hat{f}(\boldsymbol{x}, \boldsymbol{h}) f(\boldsymbol{x}) d\boldsymbol{x}$ is as shown below:

$$\int \hat{f}(\boldsymbol{x}, \boldsymbol{h}) f(\boldsymbol{x}) d\boldsymbol{x} = \frac{1}{n} \sum_{j=1}^{n} \int \hat{f}(\boldsymbol{x}, \boldsymbol{h}) dF_{n,-j}(\boldsymbol{x}),$$

where $F_{n,-j}(x) = \frac{1}{n-1} \sum_{i \neq j} I(X_i \leq x)$.

Therefore

$$\int \hat{f}(\boldsymbol{x}, \boldsymbol{h}) f(\boldsymbol{x}) d\boldsymbol{x} = \frac{1}{n} \sum_{j=1}^{n} \int \left\{ \frac{1}{\prod_{i=1}^{p} h_i n^{p/2}} \prod_{i=1}^{p} K\left(\frac{x_i - \sqrt{n}\bar{X}_i}{\sqrt{n} h_i} \right) \right\} dF_{n,-j}(\boldsymbol{x})$$

$$= \frac{1}{n \prod_{i=1}^{p} h_i n^{p/2}} \sum_{j=1}^{n} \int \left\{ \prod_{i=1}^{p} K\left(\frac{x_i - \sqrt{n}\bar{X}_i}{\sqrt{n} h_i} \right) \right\} dF_{n,-j}(\boldsymbol{x})$$

$$= \frac{1}{n \prod_{i=1}^{p} h_i n^{p/2}} \prod_{i=1}^{p} \sum_{j=1}^{n} \int \left\{ K\left(\frac{x_i - \sqrt{n}\bar{X}_i}{\sqrt{n} h_i} \right) \right\} dF_{n,-j}(\boldsymbol{x}).$$

Thus,

$$\int \hat{f}(\boldsymbol{x}, \boldsymbol{h}) f(\boldsymbol{x}) d\boldsymbol{x} = \frac{1}{n} \sum_{j=1}^{n} \frac{1}{\prod_{i=1}^{p} h_i n^{p/2}} \prod_{i=1}^{p} \frac{1}{n-1} K\left(\frac{X_{ij} - \sqrt{n}\bar{X}_{i,-j}}{\sqrt{n} h_i} \right).$$

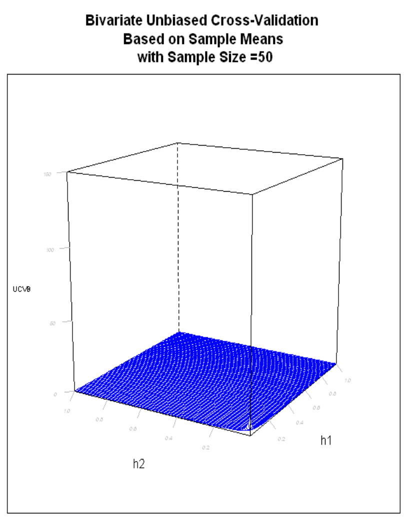
In that case,

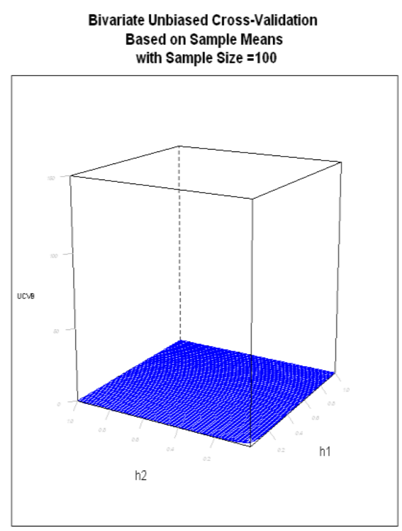
$$UCV(h_1,\ldots,h_p)=\frac{1}{(2\sqrt{\pi})^p\left(n^{p/2}\prod_{l=1}^{p}h_l\right)}-\frac{1}{n}\sum_{j=1}^{n}\frac{1}{\prod_{i=1}^{p}h_i n^{p/2}}\prod_{i=1}^{p}\frac{1}{n-1}K\left(\frac{X_{ij}-\sqrt{n}\bar{X}_{i,-j}}{\sqrt{n}h_i}\right). \quad (4.30)$$

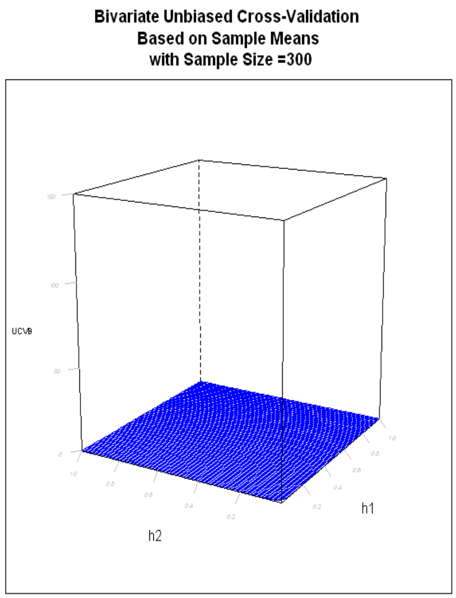
For a bivariate case, that is when $p=2$, we have

$$UCV(h_1,h_2)=\frac{1}{4\pi n h_1 h_2}-\frac{2}{n^2(n-1)h_1.h_2}\sum_{j=1}^{n}\prod_{i=1}^{2}K\left(\frac{X_{ij}-\sqrt{n}\bar{X}_{i,-j}}{\sqrt{n}h_i}\right). \quad (4.31)$$

Figure 4.2.1 Plots of the Unbiased Cross-Validation of the Multivariate Kernel Estimate Based on the Sample Means With Different Sample Sizes







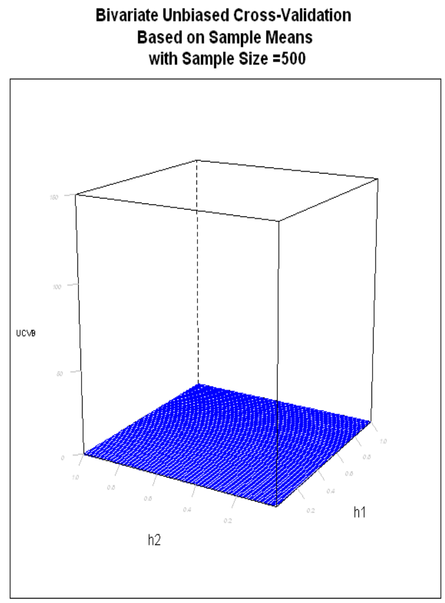
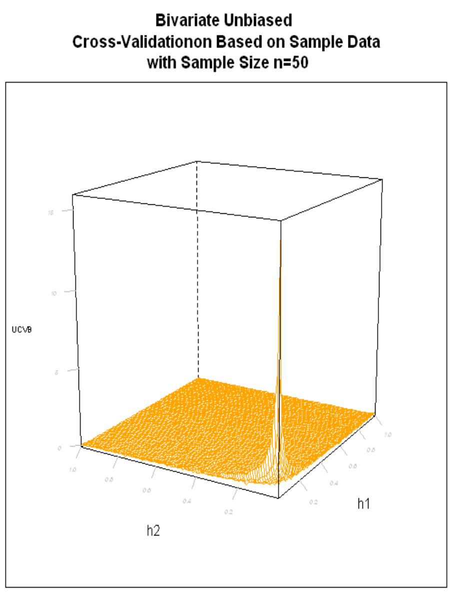
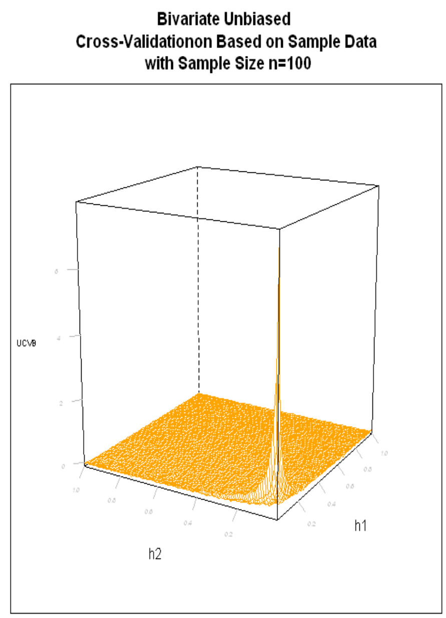
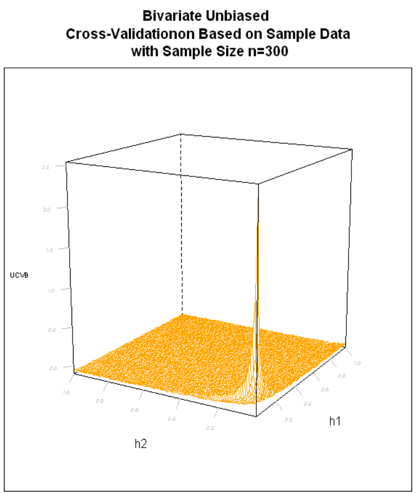
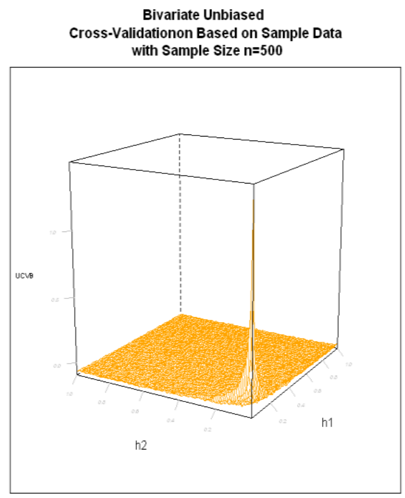


Figure 4.2.2 Plots of the Unbiased Cross-Validation of the Multivariate Standard Kernel Estimate at Different Sample Sizes









Even though both plots have spike due to singularities, the effect is minimal in figure (4.2.1). Also, it can be observed that the performance of the bandwidth selection in the unbiased cross-validation based on the sample means is better than the smoothing based on the sample data. The surfaces of the plots based on the sample data are rougher. The spikes diminish in the plots of unbiased cross-validation based on the sample means as sample size increases but still the same in the case of the sample data even though sample size increases. Besides, comparing the computational time for the unbiased cross-validation bandwidth selection based on the sample means and the sample data, the computational time was significantly reduced in the case of the sample means.

4.2.2 Biased Cross Validation (BCV)

In the univariate biased cross-validation based on the sample means, we derived the $BCV(h)$ based on the AMISE. The AMISE was obtained by

$$AMISE = \frac{R(K)}{h\sqrt{n}} + \frac{1}{4}h^4 n^2 \sigma_K^2 R(f''). \tag{4.32}$$

Here, we will first outline the general derivation of BCV function for the bivariate case, that's when $p = 2$. Now, considering the bivariate form of the AMISE for the product kernel defined in equation (4.21) above,

$$\begin{aligned} AMISE(h_1, h_2) = \frac{R(K)^2}{h_1 h_2 \sqrt{n}} + \frac{1}{4} n^2 \sigma_K^4 \Big\{ h_1^4 \iint f_{x_1x_1}(x_1, x_2)^2 dx_1 dx_2 \\ + h_2^4 \iint f_{x_2x_2}(x_1, x_2)^2 dx_1 dx_2 \\ + 2h_1^2 h_2^2 \iint f_{x_1x_1}(x_1, x_2) f_{x_2x_2}(x_1, x_2) dx_1 dx_2 \Big\}. \end{aligned} \tag{4.33}$$

But $\iint f_{x_1x_1}(x_1, x_2)^2 dx_1 dx_2 = \iint f_{x_1}^{(4)}(x_1, x_2) f(x_1, x_2) dx_1 dx_2 = E\left[f_{x_1}^{(4)}(x_1, x_2)\right]$ and it is estimated by

$$\frac{1}{n}\sum_{i=1}^{n} \hat{f}_{-i,x_1}^{(4)}(x_{i1}, x_{i2}).$$

Similarly, $\iint f_{x_1x_1}(x_1, x_2) f_{x_2x_2}(x_1, x_2) dx_1 dx_2 = E\left[\frac{\partial^4 f(x_1, x_2)}{\partial x_1^2 \partial x_2^2}\right]$ and this is estimated by

$$\frac{1}{n}\sum_{i=1}^{n} \frac{\partial^4 \hat{f}_{-i}^{(4)}(x_{i1}, x_{i2})}{\partial x_1^2 \partial x_2^2}.$$

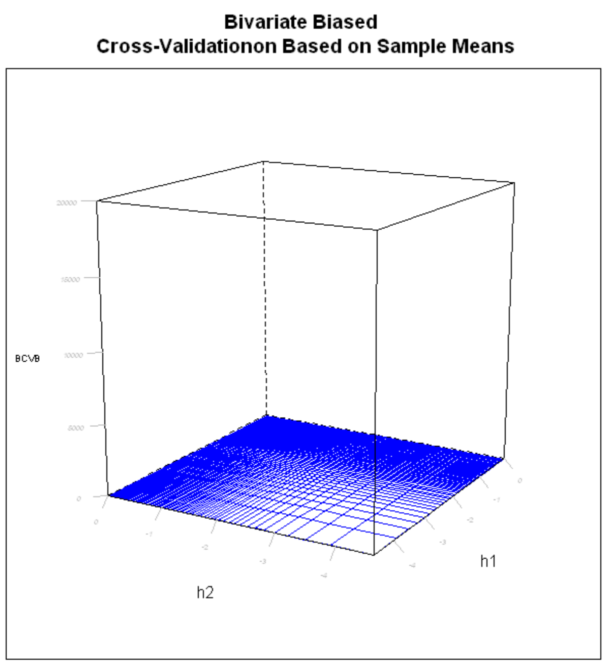
Substituting the estimates into equation (4.33) and simplifying, we have

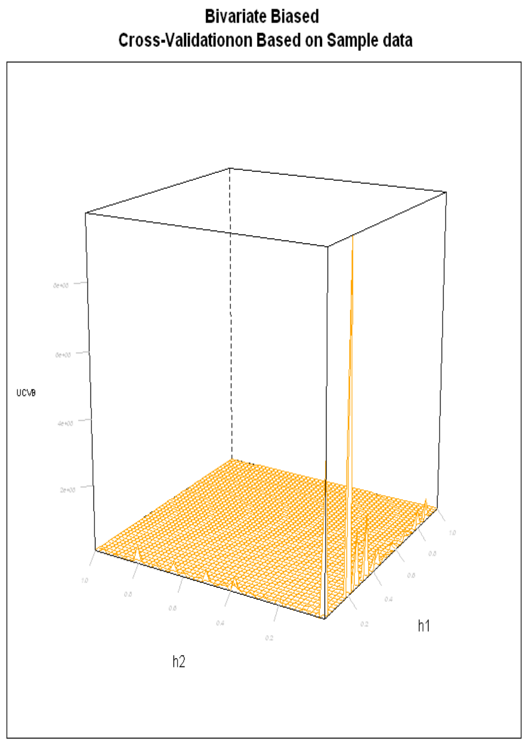
$$\begin{aligned} BCV(h_1, h_2) = \frac{R(K)^2}{h_1 h_2 \sqrt{n}} + \frac{1}{4} n^2 \sigma_K^4 \Big\{ h_1^4 \left(\frac{1}{n}\sum_{i=1}^{n} \hat{f}_{-i,x_1}^{(4)}(x_{i1}, x_{i2}) \right) + h_2^4 \left(\frac{1}{n}\sum_{i=1}^{n} \hat{f}_{-i,x_2}^{(4)}(x_{i1}, x_{i2}) \right) \\ + 2h_1^2 h_2^2 \left(\frac{1}{n}\sum_{i=1}^{n} \frac{\partial^4 \hat{f}_{-i}(x_{i1}, x_{i2})}{\partial x_1^2 \partial x_2^2} \right) \Big\}. \end{aligned}$$

This implies,

$$BCV(h_1,h_2)=\frac{R(K)^2}{h_1h_2\sqrt{n}}+\frac{1}{4(n-1)^{5/2}h_1h_2}n\sigma_K^4\sum_{i=1}^{n}K^{(4)}\sum_{j=1}^{2}\left(\frac{x_i-\sqrt{n}\,\overline{X}}{h_j\sqrt{n}}\right)$$
$$+\frac{1}{4(n-1)^{5/2}h_1h_2}n\sigma_K^4\sum_{i=1}^{n}K^{(4)}\sum_{j=1}^{n}\left(\frac{x_i-\sqrt{n}\,\overline{X}}{h_j\sqrt{n}}\right)$$
$$+\quad\frac{1}{2(n-1)^{5/2}h_1h_2}n\sigma_K^4\sum_{i=1}^{n}K^{(4)}\sum_{j=1}^{n}\left(\frac{x_i-\sqrt{n}\,\overline{X}}{h_j\sqrt{n}}\right). \tag{4.34}$$

Figure 4.2.3 Plots Comparing the Biased Cross Validation (BCV) of the Multivariate Standard Kernel Estimate and the Multivariate Kernel Estimate Based on the Sample Means





From the above plots, the biased cross-validation based on the sample data have some spikes, which might be due to singularities but it not the case in the one based on the sample means.

4.3 Plug-In Technique

4.3.1 Plug-In Bandwidth Selection: Univariate Case

The main idea of the plug-in bandwidth selection is to replace the unknown quantities that appear in the formulae for the asymptotically optimal bandwidth (h_{AIMISE}) by their estimates. From equation (4.7) above, AMISE was given as

$$AMISE = \frac{R(K)}{h\sqrt{n}} + \frac{1}{4} h^4 n^2 \sigma_K^4 R(f'').$$

Therefore, the asymptotically optimal bandwidth can be obtained by differentiating the AMISE, equating it to zero and solving for the h. By so doing, the asymptotically optimal bandwidth

$$h_{AMISE} = n^{-1/2}\left[\frac{R(K)}{\sigma_K^4 R(f'')}\right]^{1/5}. \quad (4.35)$$

The general integrated squared density derivative functional is defined as

$$R\left(f^{(s)}\right) = \int f^{(s)}(x)^2 dx.$$

Using integration by parts, under sufficient smoothness assumption on f, it can be shown that

$$R\left(f^{(s)}\right) = (-1)^s \int f^{(2s)}(x) f(x) dx. \quad (4.36)$$

It is therefore very important to estimates functionals of the form

$$\psi_r = \int f^{(r)}(x) f(x) dx \quad (4.37)$$

for r even. The ψ_r notation is preferred to the usual $R\left(f^{(r)}\right)$ notation since it's easier to extend it to the multivariate settings. From equation (4.37), we know that,

$$\psi_r = E\left\{f^{(r)}(X)\right\}$$

From equation (4.1), the estimate of density $f(x)$ based on the sample means was given by

$$\hat{f}(x) = \frac{1}{h\sqrt{n}} K\left(\frac{x - \sqrt{n}\bar{X}}{\sqrt{n}h}\right).$$

Therefore,

$$\hat{f}^{(r)}(x) = \left(\frac{1}{h\sqrt{n}}\right)^{r+1} K^{(r)}\left(\frac{x-\sqrt{n}\bar{X}}{h\sqrt{n}}\right).$$

This motivates the estimator

$$\hat{\psi}_r = n^{-1}\sum_{i=1}^{n} \hat{f}^{(r)}(X_i; g) = n^{-1}\left(\frac{1}{g\sqrt{n}}\right)^{r+1} \sum_{i=1}^{n} L^{(r)}\left(\frac{X_i-\sqrt{n}\bar{X}}{g\sqrt{n}}\right), \tag{4.38}$$

where g and L are a bandwidth and kernel that are possibly different from h and K respectively.

Now, we know that,

$$MSE(\hat{\psi}_r) = E\{(\hat{\psi}_r(g) - \psi_r)^2\} = Var(\hat{\psi}_r(g)) + \{Bias(\hat{\psi}_r(g))\}^2. \tag{4.39}$$

$$E(\hat{\psi}_r) = n^{-1}(g\sqrt{n})^{-r-1} n \int L^{(r)}\left(\frac{X_i-\sqrt{n}\bar{X}}{g\sqrt{n}}\right)dx$$

$$= (g\sqrt{n})^{-r-1} \int\int L^{r}\left(\frac{x\left(1-\frac{1}{\sqrt{n}}\right) - y\left(\frac{n-1}{n}\right)^{1/2}}{g\sqrt{n}}\right) f(x) f_{\bar{X}_{n-1}}(y)\, dxdy.$$

Assuming $f_{\bar{X}_{n-1}}(y) = \sqrt{n-1}\, f(\sqrt{n-1}\,y)$, we get that

$$E[\hat{\psi}_r(g)] = (g\sqrt{n})^{-r-1} \int\int L^{r}\left(\frac{x\left(1-\frac{1}{\sqrt{n}}\right) - y\left(\frac{n-1}{n}\right)^{1/2}}{g\sqrt{n}}\right) f(x) \sqrt{n-1} f_{\bar{X}_{n-1}}(\sqrt{n-1}\,y)\, dxdy.$$

Now, let $w = \sqrt{n-1}\,y$, then $y = \frac{w}{\sqrt{n-1}}$.

This implies,

$$\frac{\partial y}{\partial w} = \frac{1}{\sqrt{n-1}} \text{ and } \frac{\partial y}{\partial u} = 0.$$

Also, let $u = \dfrac{x\left(1-\dfrac{1}{\sqrt{n}}\right) - y\left(\dfrac{n-1}{n}\right)^{1/2}}{g\sqrt{n}}$, then $x = \left[ug\sqrt{n} + \dfrac{w}{\sqrt{n}}\right]\left(1-\dfrac{1}{\sqrt{n}}\right)^{-1}$.

This implies,

$$\frac{\partial x}{\partial w} = \frac{1}{\sqrt{n}}\left(1-\frac{1}{n}\right)^{-1} \text{ and } \frac{\partial x}{\partial u} = g\sqrt{n}\left(1-\frac{1}{n}\right)^{-1}.$$

So,

$$|J| = \begin{vmatrix} \dfrac{1}{\sqrt{n-1}} & 0 \\ \dfrac{1}{\sqrt{n}}\left(1-\dfrac{1}{n}\right)^{-1} & g\sqrt{n}\left(1-\dfrac{1}{n}\right)^{-1} \end{vmatrix} = \frac{g\sqrt{n}}{\sqrt{n-1}}\left(1-\frac{1}{\sqrt{n}}\right)^{-1}.$$

Therefore,

$$E\left(\hat{\psi}_r(g)\right) = \left(g\sqrt{n}\right)^{-r-1}\int\int L^{(r)}(u) f\left[\left(ug\sqrt{n}+\frac{w}{\sqrt{n}}\right)\left(1-\frac{1}{\sqrt{n}}\right)^{-1}\right]\sqrt{n-1}\, f(w)|J|\,dw\,du$$

$$= \left(g\sqrt{n}\right)^{-r-1}\int\int L^{(r)}(u) f\left[\left(ug\sqrt{n}+\frac{w}{\sqrt{n}}\right)\left(1-\frac{1}{\sqrt{n}}\right)^{-1}\right]\sqrt{n-1}\, f(w)\frac{g\sqrt{n}}{\sqrt{n-1}}\left(1-\frac{1}{\sqrt{n}}\right)^{-1}dw\,du$$

$$= \left(g\sqrt{n}\right)^{-r-1} g\sqrt{n}\left(1-\frac{1}{\sqrt{n}}\right)^{-1}\int\int L^{(r)}(u) f\left[\left(ug\sqrt{n}+\frac{w}{\sqrt{n}}\right)\left(1-\frac{1}{\sqrt{n}}\right)^{-1}\right] f(w)\,dw\,du$$

$$= \left(g\sqrt{n}\right)^{-r}\left(1-\frac{1}{\sqrt{n}}\right)^{-1}\int\int L(u) f\left[\left(ug\sqrt{n}+\frac{w}{\sqrt{n}}\right)\left(1-\frac{1}{\sqrt{n}}\right)^{-1}\right] f^{r}(w)\,dw\,du.$$

Now, by Taylor series expansion and applying the moment conditions for a symmetric kernel

$$= \left(g\sqrt{n}\right)^{-r}\left(1-\frac{1}{\sqrt{n}}\right)^{-1}$$
$$\times \iint L(w)L^{r}(u)\left\{\sum_{l=0}^{k}(l!)^{-1}\left(\frac{1}{\sqrt{n}}\left(1-\frac{1}{\sqrt{n}}\right)^{-1}\right)^{l}\left(ug\sqrt{n}\left(1-\frac{1}{\sqrt{n}}\right)^{-1}\right)\times f^{l}\left[w\frac{1}{\sqrt{n}}\left(1-\frac{1}{\sqrt{n}}\right)^{-1}\right]\right\}dwdu$$

$$= \left(g\sqrt{n}\right)^{-r}\left(1-\frac{1}{\sqrt{n}}\right)^{-1}\left[\psi_r + (k!)^{-1}\mu_k(L)\left(\frac{1}{\sqrt{n}}\left(1-\frac{1}{\sqrt{n}}\right)^{-1}\right)^{k}\left(g\sqrt{n}\left(1-\frac{1}{\sqrt{n}}\right)^{-1}\right)^{k}\psi_{r+k}\right].$$

For large n and with the appropriate choice of g ($g = o(n^{-1/2})$), we have

$$= \left[\psi_r + (k!)^{-1}\mu_k(L)\left(\frac{1}{\sqrt{n}}\left(1-\frac{1}{\sqrt{n}}\right)^{-1}\right)^{k}\left(g\sqrt{n}\left(1-\frac{1}{\sqrt{n}}\right)^{-1}\right)^{k}\psi_{r+k}\right]. \tag{4.40}$$

Therefore, the bias of $\hat{\psi}_r$ is given by

$$bias(\hat{\psi}_r) = E(\hat{\psi}_r(g)) - \psi_r = (k!)^{-1}\mu_k(L)\left(\frac{1}{\sqrt{n}}\left(1-\frac{1}{\sqrt{n}}\right)^{-1}\right)^{k}\left(g\sqrt{n}\left(1-\frac{1}{\sqrt{n}}\right)^{-1}\right)^{k}\psi_{r+k}. \tag{4.41}$$

If $X_1,\ldots,X_n$ be a set of identically and independently distributed random variables and applying the symmetric nature of $L(\cdot)$ for r even, it follows that, the variance of $\hat{\psi}_r$ is given by

$$Var(\hat{\psi}_r) = \left[n^{-1}\left(g\sqrt{n}\right)^{-r-1}\right]^{2}\left\{\begin{array}{l}\frac{n(n-1)}{2}Var\left\{L^{(r)}\left(\frac{X_1-\sqrt{n}\bar{X}}{g\sqrt{n}}\right)\right\}\\ +n(n-1)(n-2)Cov\left\{L^{(r)}\left(\frac{X_1-\sqrt{n}\bar{X}}{g\sqrt{n}}\right),L^{(r)}\left(\frac{X_2-\sqrt{n}\bar{X}}{g\sqrt{n}}\right)\right\}\end{array}\right\}. \tag{4.42}$$

Now, treating each component of both the variance and the covariance in turn, we first consider

$$E\left\{L^{(r)2}\left(\frac{X_1-\sqrt{n}\bar{X}}{g\sqrt{n}}\right)\right\}=\int\int L^{(r)2}\left(\frac{x\left(1-\frac{1}{\sqrt{n}}\right)-y\left(\frac{n-1}{n}\right)^{1/2}}{g\sqrt{n}}\right)f(x)\,f_{\bar{X}_{n-1}}(y)\,dxdy\,. \tag{4.43}$$

Assuming $f_{\bar{X}_{n-1}}(y)=\sqrt{n-1}f\left(\sqrt{n-1}y\right)$, we get that

$$E\left\{L^{(r)2}\left(\frac{X_1-\sqrt{n}\bar{X}}{g\sqrt{n}}\right)\right\}=\int\int L^{(r)2}\left(\frac{x\left(1-\frac{1}{\sqrt{n}}\right)-y\left(\frac{n-1}{n}\right)^{1/2}}{g\sqrt{n}}\right)f(x)\,\sqrt{n-1}f_{\bar{X}_{n-1}}\left(\sqrt{n-1}y\right)dxdy.$$

Now, let $w=\sqrt{n-1}y,$ then $y=\dfrac{w}{\sqrt{n-1}}.$

This implies,

$$\frac{\partial y}{\partial w}=\frac{1}{\sqrt{n-1}}\ \text{ and }\ \frac{\partial y}{\partial u}=0.$$

Also, let $u=\dfrac{x\left(1-\frac{1}{\sqrt{n}}\right)-y\left(\frac{n-1}{n}\right)^{1/2}}{g\sqrt{n}}$, then $x=\left[ug\sqrt{n}+\dfrac{w}{\sqrt{n}}\right]\left(1-\dfrac{1}{\sqrt{n}}\right)^{-1}.$

This implies,

$$\frac{\partial x}{\partial w}=\frac{1}{\sqrt{n}}\left(1-\frac{1}{n}\right)^{-1}\ \text{ and }\ \frac{\partial x}{\partial u}=g\sqrt{n}\left(1-\frac{1}{n}\right)^{-1}.$$

So,

$$|J| = \left| \begin{matrix} \dfrac{1}{\sqrt{n-1}} & 0 \\ \dfrac{1}{\sqrt{n}}\left(1-\dfrac{1}{n}\right)^{-1} & g\sqrt{n}\left(1-\dfrac{1}{n}\right)^{-1} \end{matrix} \right| = \frac{g\sqrt{n}}{\sqrt{n-1}}\left(1-\frac{1}{\sqrt{n}}\right)^{-1}.$$

Therefore,

$$E\left\{L^{(2)}\left(\frac{X_1-\sqrt{n}\bar{X}}{g\sqrt{n}}\right)\right\}$$

$$= \iint L^{(2)}(u)\sqrt{n-1}\,f(w)\,f\left[\left(ug\sqrt{n}+\frac{w}{\sqrt{n}}\right)\left(1-\frac{1}{\sqrt{n}}\right)^{-1}\right]|J|\,dwdu$$

$$= g\sqrt{n}\left(1-\frac{1}{\sqrt{n}}\right)^{-1}\iint L^{(2)}(u)\,f\left[\left(ug\sqrt{n}+\frac{w}{\sqrt{n}}\right)\left(1-\frac{1}{\sqrt{n}}\right)^{-1}\right]f(w)\,dwdu.$$

For large n, we have the above last term equal to

$$g\sqrt{n}\iint L^{2}(u)\,f\left[\left(ug\sqrt{n}+\frac{w}{\sqrt{n}}\right)\left(1-\frac{1}{\sqrt{n}}\right)^{-1}\right]f(w)\,dwdu \approx g\psi_0 R(L). \tag{4.44}$$

$$E\left\{L^{r}\left(\frac{X_1-\sqrt{n}\bar{X}}{g\sqrt{n}}\right)L^{r}\left(\frac{X_1-\sqrt{n}\bar{X}}{g\sqrt{n}}\right)\right\}$$

Now, considering $= \iiint L^{r}\left(\dfrac{x\left(1-\dfrac{1}{\sqrt{n}}\right)-y\left(\dfrac{n-1}{n}\right)^{1/2}}{g\sqrt{n}}\right)L^{r}\left(\dfrac{x\left(1-\dfrac{1}{\sqrt{n}}\right)-y\left(\dfrac{n-1}{n}\right)^{1/2}}{g\sqrt{n}}\right)$

$$\times f(x)\,f_{\bar{X}_{n-1}}(y)\,f(z)\,dxdydz.$$

Assuming $f_{\bar{X}_{n-1}}(y) = \sqrt{n-1}\,f\left(\sqrt{n-1}\,y\right)$, we get

$$= \int\int\int L^r \left(\frac{x\left(1-\frac{1}{\sqrt{n}}\right) - y\left(\frac{n-1}{n}\right)^{1/2}}{g\sqrt{n}} \right) L^r \left(\frac{x\left(1-\frac{1}{\sqrt{n}}\right) - y\left(\frac{n-1}{n}\right)^{1/2}}{g\sqrt{n}} \right)$$

$$\times f \quad x \quad \sqrt{n-1} f_{\bar{X}_{n-1}} \quad \sqrt{n-1} y \quad f \quad z \quad dxdydz.$$

Now, let $w = \sqrt{n-1}y$, then $y = \frac{w}{\sqrt{n-1}}$.

This implies,

$$\frac{\partial y}{\partial w} = \frac{1}{\sqrt{n-1}} \ , \ \frac{\partial y}{\partial u} = 0 \text{ and } \frac{\partial y}{\partial v} = 0.$$

Also, let $u = \frac{x\left(1-\frac{1}{\sqrt{n}}\right) - y\left(\frac{n-1}{n}\right)^{1/2}}{g\sqrt{n}}$, then $x = \left[ug\sqrt{n} + \frac{w}{\sqrt{n}}\right]\left(1-\frac{1}{\sqrt{n}}\right)^{-1}$.

This implies,

$$\frac{\partial x}{\partial w} = \frac{1}{\sqrt{n}}\left(1-\frac{1}{n}\right)^{-1}, \frac{\partial x}{\partial u} = g\sqrt{n}\left(1-\frac{1}{n}\right)^{-1} \text{ and } \frac{\partial x}{\partial v} = 0.$$

Again, let $v = \frac{y\left(1-\frac{1}{\sqrt{n}}\right) - z\left(\frac{n-1}{n}\right)^{1/2}}{g\sqrt{n}}$, then $z = \left[\frac{w}{\sqrt{n}} - vg\sqrt{n}\right]\left(1-\frac{1}{\sqrt{n}}\right)^{-1}$.

This implies,

$$\frac{\partial z}{\partial w} = \frac{1}{\sqrt{n}}\left(1-\frac{1}{n}\right)^{-1}, \frac{\partial z}{\partial u} = 0 \text{ and } \frac{\partial z}{\partial v} = -g\sqrt{n}\left(1-\frac{1}{\sqrt{n}}\right)^{-1}.$$

So,

$$
|J| = \begin{vmatrix} \frac{1}{\sqrt{n-1}} & 0 & 0 \\ \frac{1}{\sqrt{n}}\left(1-\frac{1}{\sqrt{n}}\right)^{-1} & g\sqrt{n}\left(1-\frac{1}{\sqrt{n}}\right)^{-1} & 0 \\ \frac{1}{\sqrt{n}}\left(1-\frac{1}{\sqrt{n}}\right)^{-1} & 0 & -g\sqrt{n}\left(1-\frac{1}{\sqrt{n}}\right)^{-1} \end{vmatrix}
$$

$$
= \frac{1}{\sqrt{n-1}} \begin{vmatrix} g\sqrt{n}\left(1-\frac{1}{\sqrt{n}}\right)^{-1} & 0 \\ 0 & -g\sqrt{n}\left(1-\frac{1}{\sqrt{n}}\right)^{-1} \end{vmatrix} = \frac{-\left(g\sqrt{n}\right)^{2}}{\sqrt{n-1}}\left(1-\frac{1}{\sqrt{n}}\right)^{-2}.
$$

Therefore,

$$
E\left\{L^{r}\left(\frac{X_1-\sqrt{n}\bar{X}}{g\sqrt{n}}\right)L^{r}\left(\frac{X_1-\sqrt{n}\bar{X}}{g\sqrt{n}}\right)\right\}
$$

$$
= \iiint L^{r}(u)\,L^{r}(v)\,f(w)\left(-\left(g\sqrt{n}\right)^{2}\left(1-\frac{1}{\sqrt{n}}\right)^{-2}\right) f\left(\left[ug\sqrt{n}+\frac{w}{\sqrt{n}}\right]\left(1-\frac{1}{\sqrt{n}}\right)^{-1}\right)
$$

$$
\times f\left(\left[\frac{w}{\sqrt{n}}-ug\sqrt{n}\right]\left(1-\frac{1}{\sqrt{n}}\right)^{-1}\right) dwdudv
$$

$$
= -\left(g\sqrt{n}\right)^{2}\left(1-\frac{1}{\sqrt{n}}\right)^{-2} \iiint L(u)\,L(v)\,f(w)\,f^{(r)}\left(\left[ug\sqrt{n}+\frac{w}{\sqrt{n}}\right]\left(1-\frac{1}{\sqrt{n}}\right)^{-1}\right)
$$

$$
\times f^{(r)}\left(\left[\frac{w}{\sqrt{n}}-ug\sqrt{n}\right]\left(1-\frac{1}{\sqrt{n}}\right)^{-1}\right) dwdudv
$$

$$
= \left(g\sqrt{n}\right)^{2+2r} \int f^{(2r)}(w)\,f(w)\,dw + o(1), \tag{4.45}
$$

for n, sufficiently large.

We also know from equation (4.40) that $E\left\{L^{(r)}\left(\frac{X_1-\sqrt{n}\bar{X}}{g\sqrt{n}}\right)\right\} = \psi_r + o(1).$

Substituting these into equation (4.42), we have

$$Var(\hat{\psi}_r)$$

$$\approx \left[n^{-1}(g\sqrt{n})^{-r-1}\right]^2 \left\{ \begin{array}{l} \dfrac{n(n-1)}{2} \times \dfrac{n}{(n-1)} g\psi_0 R(K) + n(n-1)(n-2) \\ \times \dfrac{n^2}{(n-1)(n-2)}\left[(g\sqrt{n})^{2+2r} \int f^{(2)}(w) f(w)\,dw - \psi_r^2 \right] \end{array} \right\}$$

$$Var(\hat{\psi}_r)$$

$$\approx \left[n^{-1}(g\sqrt{n})^{-r-1}\right]^2 \left\{ \frac{n^2}{2} g\psi_0 R(K) + n^2\left[(g\sqrt{n})^{2+2r} \int f^{(2)}(w) f(w)\,dw - \psi_r^2 \right] \right\} \tag{4.46}$$

$$= \frac{n^{-r-1}}{2} g^{-2r-1}\psi_0 R(K) + n^{-2}\left\{\int f^{(2)}(w) f(w)\,dw - \psi_r^2\right\}.$$

The asymptotic MSE is therefore obtained by substituting equations (4.41) and (4.42) into equation (4.39) as shown below:

$$MSE(\hat{\psi}_r) = \left\{ (k!)^{-1}\mu_k(K)\left(\frac{1}{\sqrt{n}}\left(1-\frac{1}{\sqrt{n}}\right)^{-1}\right)^k \left(g\sqrt{n}\left(1-\frac{1}{\sqrt{n}}\right)^{-1}\right)^k \psi_{r+k} \right\}^2$$
$$+ \frac{n^{-r-1}}{2} g^{-2r-1}\psi_0 R(K) + n^{-2}\left\{\int f^{(2)}(w) f(w)\,dw - \psi_r^2\right\}. \tag{4.47}$$

The value of g that minimizes the $MSE(\hat{\psi}_r(g))$ is obtained by differentiating $MSE(\hat{\psi}_r(g))$ with respect to g equating it to zero and solving for g as follows:

$$2\left\{ (k!)^{-1}\mu_k(K)\left(\frac{1}{\sqrt{n}}\left(1-\frac{1}{\sqrt{n}}\right)^{-1}\right)^k g^k n^{k/2}\left(1-\frac{1}{\sqrt{n}}\right)^{-k} \psi_{r+k} \right\}$$
$$\times \left\{ (k!)^{-1}\mu_k(K)\left(\frac{1}{\sqrt{n}}\left(1-\frac{1}{\sqrt{n}}\right)^{-1}\right)^k k g^{k-1} n^{k/2}\left(1-\frac{1}{\sqrt{n}}\right)^{-k} \psi_{r+k} \right\}$$
$$+ \frac{n^{-r-1}}{2}(2r-1)g^{-2r-2}\psi_0 R(K) = 0.$$

Which implies that

$$2\left\{(k!)^{-2}\mu_k^{\,2}(\ldots)\left(1-\frac{1}{\sqrt{n}}\right)^{-2k} k g^{2k-1} n^{k}\left(1-\frac{1}{\sqrt{n}}\right)^{-2k}\psi^{2}{}_{r+k}\right\}+\frac{n^{-r-1}}{2}(-2r-1)g^{-2r-2}\psi_0 R(\ldots)=0.$$

Multiplying through by g^{-2k+1} we have

$$n^{-r-1}(-2r-1)g^{-2r-2k-1}\psi_0 R(\ldots)=4(k!)^{-2}\mu_k^{\,2}(\ldots)\left(1-\frac{1}{\sqrt{n}}\right)^{-2k} k g^{2k-1} n^{k}\left(1-\frac{1}{\sqrt{n}}\right)^{-2k}\psi^{2}{}_{r+k}.$$

For large n, we have

$$n^{-r-1}(-2r-1)g^{-2r-2k-1}\psi_0 R(\ldots)=4(k!)^{-2}\mu_k^{\,2}(\ldots k)g^{2k-1}n^{k}\psi^{2}{}_{r+k}.$$

Therefore,

$$g^{-(2r+2k+1)}=\frac{4(k!)^{-2}\mu_k^{\,2}(\ldots k)g^{2k-1}n^{k}\psi^{2}{}_{r+k}}{(2r+1)n^{-r-1}\psi_0 R(\ldots)}$$

$$g_{AMSE}=\left[\frac{(2r+1)(k!)^{2}\psi_0 R(\ldots)}{4\mu_k^{\,2}(\ldots k)\psi^{2}{}_{r+k}n^{k+r+1}}\right]^{1/2r+2k+1}. \qquad (4.48)$$

Therefore, replacing the unknown quantity that appear in the formulae for the asymptotic optimal bandwidth $R(f'')$ by the functional ψ_r, equation (4.35) becomes

$$h_{AMISE}=n^{-1/2}\left[\frac{R(K)}{\sigma_K^4\psi_4}\right]^{1/5}. \qquad (4.49)$$

Now, replacing ψ_4 by the kernel estimator $\hat{\psi}_4(g)$ leads to what we called the direct plug-in rule:

$$\hat{h}_{DPI} = n^{-1/2}\left[\frac{R(K)}{\sigma_K^4 \hat{\psi}_4(g)}\right]^{1/5}. \tag{4.50}$$

But this is not automatic since $\hat{h}_{DPI}$ depends on the choice of the pilot bandwidth g. The value of g can be chosen by using the formulae for the AMSE-optimal bandwidth estimation of $\hat{\psi}_4(g)$. Using the same second-order kernel,

$$g_{AMSE} = \left[\frac{9\psi_0 R(K^{(4)})}{2\mu_k^{2}(K)\psi_6^2 n^7}\right]^{1/13}. \tag{4.51}$$

But here too, the method for choosing g has the problems as $\hat{h}_{DPI}$ since it also depends on the unknown density functional, ψ_6. From equation (4.48), it is evident that the optimal bandwidth for estimating ψ_r depends on ψ_{r+2}.

This problem can be overcome by estimating the ψ_r using the quick and simple estimate. $l-stage$direct plug-in bandwidth selector $(\hat{h}_{DPI,l})$, the name given to a direct plug-in which involves l successive kernel functional estimations, with the initial bandwidth chosen through a quick and simple method. This approach uses the normal scale rule suggested by Sheather and Jones (1991) as a zero-stage direct plug-in bandwidth.

The normal scale bandwidth selector makes use of the AMISE optimal bandwidth for the normal density having the same scale as that estimated for the underlying density. From equation (4.35), the optimal bandwidth was given as

$$h_{AMISE} = n^{-1/2}\left[\frac{R(K)}{\sigma_K^4 R(f'')}\right]^{1/5}. \tag{4.52}$$

If the unknown distribution f has a normal distribution with mean 0 and variance σ^2, then the AMISE optimal bandwidth becomes (Wand and Jones, 1995, p.60)

$$h_{AMISE} = n^{-1/2}\left[\frac{8\pi^{1/2}R(K)}{3\sigma_K^4}\right]^{1/5}\sigma. \qquad (4.53)$$

To obtain the normal scale bandwidth selector, the standard deviation σ in the above equation replace by its estimate $\hat{\sigma}$ (Silverman 1986). Therefore, the normal scale bandwidth is given by

$$\hat{h}_{NS} = n^{-1/2}\left[\frac{8\pi^{1/2}R(K)}{3\sigma_K^4}\right]^{1/5}\hat{\sigma}. \qquad (4.54)$$

Usually, the sample standard deviation s is used as the estimator of σ or a more robust estimator of σ such as the standardized inter-quartile range $(\hat{\sigma}_{IQR})$ could be used. It's been suggested that, to avoid or lessen the chance of over smoothing, the smaller of s and $\hat{\sigma}_{IQR}$ is used (Silverman, 1986, p.47).

In general, if f is a normal density with variance σ^2 then, for r even (Wand and Jones, 1995, p.72),

$$\psi_r = \frac{(-1)^{r/2}\, r!}{(2\sigma)^{r+1}\,(r/2)!\,\pi^{1/2}}. \qquad (4.55)$$

This is an illustration of the two-stage plug-in bandwidth selector. Assuming $L = K$ where K is a second-order kernel:

Step1: We estimate ψ_8 by the normal scale rule, $\hat{\psi}_8^{NS} = \dfrac{105}{32\pi^{1/2}\hat{\sigma}^9}$

Step2: Now we estimate ψ_6 using the kernel estimate $\hat{\psi}_6(g_1)$, where

$$g_1 = \left[\frac{23\psi_0 R(6)}{4\mu_k^{\,2} (\hat{\psi}_8^{NS^2}) n^9} \right]^{1/17}$$

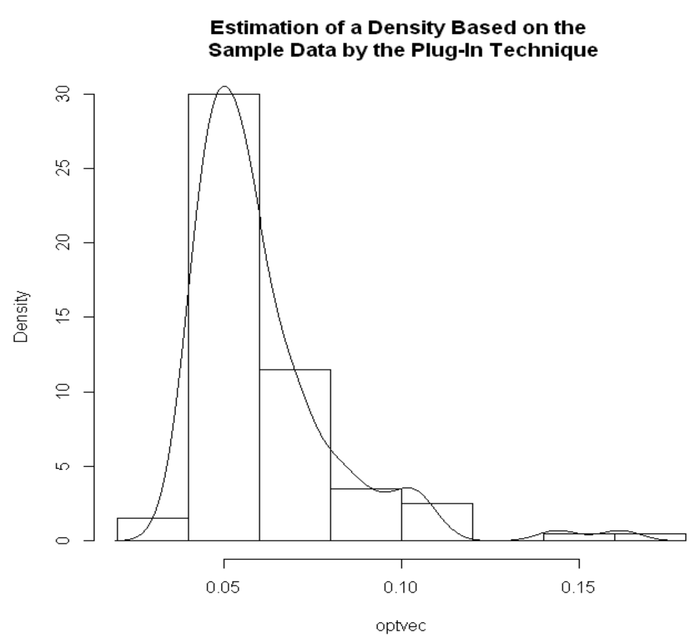
STEP3: We then estimate ψ_4 using the kernel estimate $\hat{\psi}_4(g_2)$,where

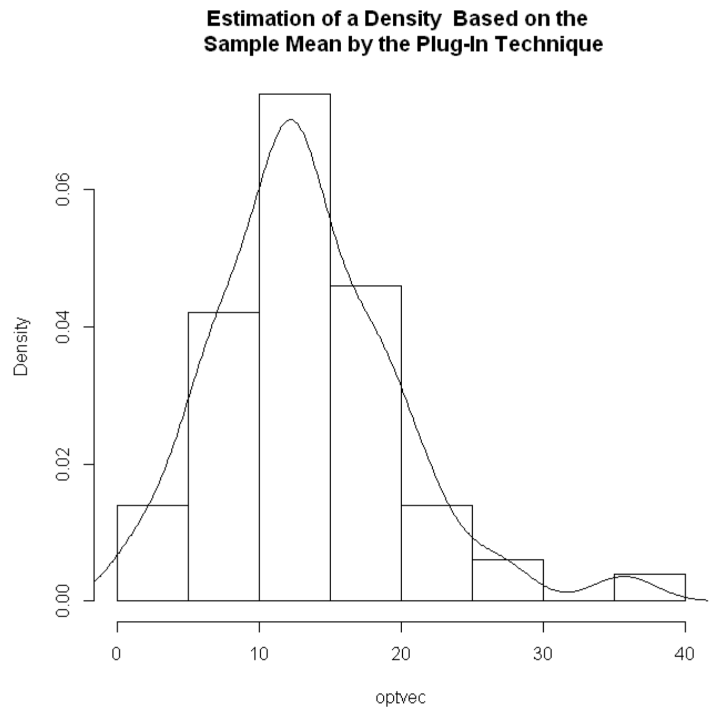
$$g_2 = \left[\frac{9\psi_0 R(4)}{2\mu_k^{\,2} (\hat{\psi}_6^2) n^7} \right]^{1/13}.$$

STEP4: The selected bandwidth is given by

$$\hat{h}_{DPI,2} = n^{-1/2} \left[\frac{R(K)}{\sigma_K^4 \hat{\psi}_4(g_2)} \right]^{1/5}.$$

Figure 4.3.1 Plots Comparing the Performance of the Plug-In Technique in the Univariate Case for Both Standard Kernel Estimate and the Kernel Estimate Based on the Sample Means





In general, both the plug-in technique base on the sample data and the sample means does a good job with the smoothing but the one based on the sample data loses tract of the smoothing at the extreme right of the histogram. The one based on the sample means smoothens evenly throughout the indices.

4.3.2 Plug-In Bandwidth Matrix Selection for Bivariate Kernel Density Estimation

Here, we consider bandwidth matrix selection for the bivariate kernel density estimator based on the sample means. A plug-in selector is developed for full bandwidth matrix.

Bivariate kernel density estimation is very important because it serves as the mediator or the bridge between the univariate and the high-dimensional multivariate cases.

The bivariate kernel density estimate based on the sample means is defined by

$$\hat{f}_{\overline{X}}(\boldsymbol{x}) = \frac{1}{n|\boldsymbol{H}|^{1/2}} K\left[\left(\frac{\boldsymbol{x}-\sqrt{n}\overline{\boldsymbol{X}}}{n}\right)(\boldsymbol{H}^{-1/2})\right], \tag{4.56}$$

where $\boldsymbol{x} = (x_1, x_2)^T$ and $\overline{\boldsymbol{X}} = (\overline{X}_1, \overline{X}_2)^T$, a random sample of means, K is the bivariate kernel density which is usually symmetric and $\boldsymbol{H}$ is bandwidth matrix which is symmetric and positive definite. As we all know, the choice of the bivariate bandwidth matrix $\boldsymbol{H}$ is very crucial in determining the performance of our estimate $\hat{f}_{\overline{X}}(\boldsymbol{x})$. The bandwidth matrix $\boldsymbol{H}$ can be chosen from a class of diagonal (positive definite) matrices or a class of positive multiples of the identity matrix as investigated by Wand and Jones(1993), but here we chose a full (i.e. unconstrained) bandwidth matrix which allows for arbitrary orientation of the kernel function as investigated by Duong and Hazelton (2003).

In order to measure the performance of the our estimate above, we shall make use of the mean integrated squared error (MISE) criterion which we used in Chapter 2 in conformability with majority of the researchers in this field. The MISE of the estimate $\hat{f}_{\overline{X}}(\boldsymbol{x})$ is given by

$$MISE(\hat{f}_{\overline{X}}(\boldsymbol{x})) = E\int_{R^2}[\hat{f}(\boldsymbol{x}) - f(\boldsymbol{x})]^2 d\boldsymbol{x} . \tag{4.57}$$

Therefore, our optimal bandwidth matrix is supposed to be

$$H_{MISE} = \underset{\boldsymbol{H} \in H}{argmin}\, MISE(\hat{f}_{\overline{X}}(\boldsymbol{x})), \tag{4.58}$$

where H is the space of all symmetric, positive definite 2×2 matrices and argmin is the argument of the minimum. But this optimal bandwidth does not have a closed form so the use of asymptotic analysis is employed.

From Chapter 2, the AMISE for our estimate in the multivariate settings (*p*-dimensional) was obtained to be

$$AMISE\left(\hat{f}_{\bar{X}}(\mathbf{x})\right)=\frac{R(K)}{n^{p/2}|\mathbf{H}|^{1/2}}+\frac{1}{4}n^{2p}\int\left[tr\left\{\mathbf{H}^{1/2}\left(\mathbf{H}^{1/2}\right)^{T}\nabla^{2}f(\mathbf{x})\right\}\right]^{2}d\mathbf{x}.$$

But in the bivariate case, $p=2$ and $\mathbf{H}$ is also assumed to be symmetric, therefore our AMISE becomes

$$AMISE\left(\hat{f}_{\bar{X}}(\mathbf{x})\right)=\frac{R(K)}{n|\mathbf{H}|^{1/2}}+\frac{1}{4}n^{4}\int tr^{2}\left\{\mathbf{H}\nabla^{2}f(\mathbf{x})\right\}d\mathbf{x}, \tag{4.59}$$

where $R(K)=\int_{R^2}K^{2}(\mathbf{x})d\mathbf{x}$.

To rewrite the AMISE into the form we need, we will need the following matrix results:

1. If $\mathbf{A}$ is a symmetric matrix $(\mathbf{A}=\mathbf{A}^{T})$, then $\mathbf{D}_d vech\mathbf{A}=vec\mathbf{A}$, where $\mathbf{D}_d$ is a $d^{2}\times\frac{1}{2}d(d+1)$ matrix of zeros and ones and is called the duplication matrix of order d, $vech$ is the vector half operator and vec is the vector.

For example, given a symmetric 2×2 matrix $\mathbf{A}=\begin{bmatrix}a_{11} & a_{21}\\ a_{21} & a_{22}\end{bmatrix}$, and a duplication matrix of order 2, then the above relation can be illustrated as

$$\begin{bmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{bmatrix}\begin{bmatrix}a_{11}\\ a_{21}\\ a_{22}\end{bmatrix}=\begin{bmatrix}a_{11}\\ a_{21}\\ a_{21}\\ a_{22}\end{bmatrix}$$

$$\mathbf{D}_2 vech\mathbf{A}=vec\mathbf{A}$$

2. If $\boldsymbol{A}$ is a square matrix, then

$$\boldsymbol{D}_d^T vec\boldsymbol{A} = vech\left(\boldsymbol{A} + \boldsymbol{A}^T - dg\boldsymbol{A}\right),$$

where $dg\boldsymbol{A}$ is the same as $\boldsymbol{A}$, but with all its off-diagonal entries equal to zero.

3. $tr\left(\boldsymbol{A}^T\boldsymbol{B}\right) = \left(vec^T\boldsymbol{A}\right)\left(vec\boldsymbol{B}\right)$.

Therefore, the second term in equation (4.62) can be simplified by making use of the above matrix results.

Consider

$$\begin{aligned}\int \left[tr\left\{\boldsymbol{H}\nabla^2 f(\boldsymbol{x})\right\}\right]^2 d\boldsymbol{x} &= \int \left(vec^T\boldsymbol{H}\right)\left\{vec\nabla^2 f(\boldsymbol{x})\right\}\left\{vec^T f(\boldsymbol{x})\right\}\left(vec\boldsymbol{H}\right) d\boldsymbol{x} \\ &= \int \left(vech^T\boldsymbol{H}\right)\boldsymbol{D}_d^T\left\{vec\nabla^2 f(\boldsymbol{x})\right\}\left\{vec^T f(\boldsymbol{x})\right\}\boldsymbol{D}_d\left(vech\boldsymbol{H}\right) d\boldsymbol{x}\end{aligned}$$

Therefore,

$$\int tr^2\left\{\boldsymbol{H}\nabla^2 f(\boldsymbol{x})\right\} d\boldsymbol{x} = \left(vech^T\boldsymbol{H}\right)\boldsymbol{\Psi}_4\left(vech\boldsymbol{H}\right), \qquad (4.60)$$

where $\boldsymbol{\Psi}_4$ is a $\frac{1}{2}d(d+1)\times\frac{1}{2}d(d+1) = 3\times 3$ matrix given by

$$\boldsymbol{\Psi}_4 = \int_{R^2} vech\left\{2\boldsymbol{D}^2 f(\boldsymbol{x}) - dg\boldsymbol{D}^2 f(\boldsymbol{x})\right\} vech^T\left\{2\boldsymbol{D}^2 f(\boldsymbol{x}) - dg\boldsymbol{D}^2 f(\boldsymbol{x})\right\} d\boldsymbol{x},$$

where $\boldsymbol{D}^2 f(\boldsymbol{x})$ is the Hessian matrix of f.

The above expansion holds if all the entries $\boldsymbol{D}^2 f(\boldsymbol{x})$ are square integrable and all entries $H \to 0$ and $n^{-1}|H|^{1/2} \to 0$ as $n \to \infty$.

Therefore, substituting equation (4.60) into equation (4.59), we have

$$AMISE\left(\hat{f}_{\bar{X}}(\boldsymbol{x})\right) = \frac{R(K)}{n|\boldsymbol{H}|^{1/2}} + \frac{1}{4}n^4 (vech^T \boldsymbol{H})\boldsymbol{\Psi}_4 (vech\boldsymbol{H}). \tag{4.61}$$

Now, let $\boldsymbol{r} = (r_1, r_2)$, where r_1 and r_2 are non-negative integers. Also let $|\boldsymbol{r}| = r_1 + r_2$, then can write the $\boldsymbol{r}th$ partial derivative of f as

$$f^{(\boldsymbol{r})}(\boldsymbol{x}) = \frac{\partial^{|\boldsymbol{r}|}}{\partial_{x_1}^{r_1}\partial_{x_2}^{r_2}} f(\boldsymbol{x}), \tag{4.62}$$

assuming that this derivative exists.

So the integrated density derivative functional

$$\psi_{\boldsymbol{r}} = E\left\{f^{(\boldsymbol{r})}(\boldsymbol{x})\right\} = \int_{R^2} f^{(\boldsymbol{r})}(\boldsymbol{x}) f(\boldsymbol{x})\, d\boldsymbol{x} \tag{4.63}$$

and

$$\int_{R^2} f^{(\boldsymbol{r})}(\boldsymbol{x}) f^{(\boldsymbol{s})}(\boldsymbol{x})\, d\boldsymbol{x} = (-1)^{|\boldsymbol{r}|} \int_{R^2} f^{(\boldsymbol{r}+\boldsymbol{s})}(\boldsymbol{x}) f(\boldsymbol{x})\, d\boldsymbol{x} \tag{4.64}$$

if $|\boldsymbol{r}+\boldsymbol{s}|$ is even and zero otherwise.

Making use of the equation (4.62), the $\boldsymbol{\Psi}_4$ in our bivariate case becomes

$$\boldsymbol{\Psi}_4 = \begin{pmatrix} \psi_{4,0} & 2\psi_{3,1} & \psi_{2,2} \\ 2\psi_{3,1} & 4\psi_{2,2} & 2\psi_{1,3} \\ \psi_{2,2} & 2\psi_{1,3} & \psi_{0,4} \end{pmatrix}.$$

Where the subscript 4 on the $\boldsymbol{\Psi}$ relates to the order of the derivative involved.

Therefore, the bandwidth matrix for the plug-in method is given by

$$H_{AMISE} = \underset{\boldsymbol{H} \in H}{argmin}\ AMISE\left(f_{\overline{\boldsymbol{X}}}(\boldsymbol{x})\right). \tag{4.65}$$

But here too, just like the situation we faced in the univariate case, the AMISE is a functional of the unknown target density through $\boldsymbol{\Psi}_4$. Therefore, we need to estimate the $\psi_{\boldsymbol{r}}$ functional which would allow us to get an estimated AMISE $\left(A\hat{MISE}\right)$ that can be minimized numerically to give the plug-in bandwidth matrix $\hat{\boldsymbol{H}}$.

From equation (4.63), we know that

$$\psi_{\boldsymbol{r}} = E\left\{f^{(\boldsymbol{r})}(\boldsymbol{x})\right\} = \int_{R^2} f^{(\boldsymbol{r})}(\boldsymbol{x}) f(\boldsymbol{x})\, d\boldsymbol{x}$$

which motivates the estimator

$$\hat{\psi}_{\boldsymbol{r}}(\boldsymbol{G}) = n^{-1}\sum_{i=1}^{n} \hat{f}^{(\boldsymbol{r})}(\boldsymbol{X}_i;\boldsymbol{G}) = n^{-2}\sum_{i=1}^{n} L_{\boldsymbol{G}}^{(\boldsymbol{r})}\left(\frac{x_i - \sqrt{n}\overline{X}}{n}\right), \tag{4.66}$$

where $\boldsymbol{G}$ is the pilot bandwidth matrix usual different from $\boldsymbol{H}$ and L is also a bivariate kernel density usually which is also symmetric but possibly different from *K*.

In section 4.4.1, we derived the bias and the variance for the univariate plug-in method which can be extended to the bivariate bandwidth matrix settings.

Therefore,

$$Bias\,\hat{\psi}_{\boldsymbol{r}}(\boldsymbol{G}) = \frac{1}{2}\mu_2(L)\int_{R^2} tr\left(\boldsymbol{G}\boldsymbol{D}^2 f(\boldsymbol{x})\right) f^{(\boldsymbol{r})}(\boldsymbol{x})\, d\boldsymbol{x} + o(tr\boldsymbol{G}) \tag{4.67}$$

and

$$Var\{\hat{\psi}_{\boldsymbol{r}}(\boldsymbol{G})\} = \psi_0 \int_{R^2} L_{\boldsymbol{G}}^{(\boldsymbol{r})}(\boldsymbol{x}) + n^{-2}\left\{\int_{R^2} f^{(\boldsymbol{r})}(\boldsymbol{x}) f(\boldsymbol{x}) d\boldsymbol{x} - \psi_{\boldsymbol{r}}^2\right\} + o(n^{-1}). \tag{4.68}$$

If we consider the pilot bandwidth matrix $\boldsymbol{G}$ of the form $\boldsymbol{G} = g^2 I$ (where I is a 2×2 identity matrix), then our bias and variance becomes

$$Bias\hat{\psi}_{\boldsymbol{r}}(g) = \frac{1}{2} g^2 \mu_2(L)\psi_{\boldsymbol{r}+2\boldsymbol{e}_i}, \tag{4.69}$$

for large *n*, $\boldsymbol{e}_i$ is the elementary vector(that is a vector of length 2 with 1 in the *ith* position and 0 elsewhere,

and

$$Var\{\hat{\psi}_{\boldsymbol{r}}(g)\} = \frac{n^{-|\boldsymbol{r}|-1}}{2} g^{-2|\boldsymbol{r}|-1}\psi_0 R(L^{(\boldsymbol{r})}) + n^{-2}\left\{\int_{R^2} f^{(\boldsymbol{r})}(\boldsymbol{x}) f(\boldsymbol{x}) d\boldsymbol{x} - \psi_{\boldsymbol{r}}^2\right\}, \tag{4.70}$$

provided the $L^{(\boldsymbol{r})}$ is square integrable and $g \to 0$ and $n^{-|\boldsymbol{r}|-1} g^{-2|\boldsymbol{r}|-1} \to 0$ as $n \to \infty$.

Therefore, the asymptotic mean square error (AMSE) of $\psi_{\boldsymbol{r}}$ is

$$AMSE\{\hat{\psi}_{\boldsymbol{r}}(g)\} = \frac{n^{-|\boldsymbol{r}|-1}}{2} g^{-2|\boldsymbol{r}|-1}\psi_0 R(L^{(\boldsymbol{r})}) + \left[\frac{1}{2} g^2 \mu_2(L)\psi_{\boldsymbol{r}+2\boldsymbol{e}_i}\right]^2 \tag{4.71}$$

since $\boldsymbol{\Psi}_4$ is a positive-definite for any continuous density f, when we assume the kernel L to be a multivariate normal distribution, then it can be shown that $\hat{\boldsymbol{\Psi}}_4$ is the $\boldsymbol{\Psi}_4$ matrix corresponding to $f = \hat{f}(2^{-1} g^2 \boldsymbol{I})$, hence the estimate $\hat{\boldsymbol{\Psi}}_4$ will be positive definite. Therefore, a single or common g is needed for the estimation of all elements of $\boldsymbol{\Psi}_4$. As investigated by (Duong and Hazelton, 2003), we estimate the bandwidth that minimizes the sum of the *AMSE* (*SAMSE*) for $\psi_{\boldsymbol{r}}$, that is

$$g_{4,SAMSE} = \underset{g>0}{\operatorname{argmin}}\, SAMSE(\hat{\boldsymbol{\Psi}}_4), \tag{4.72}$$

where

$$SAMSE(\hat{\boldsymbol{\Psi}}_4) = SAMSE_4(g) = \sum_{\boldsymbol{r}:|\boldsymbol{r}|=4} AMSE(\hat{\psi}_{\boldsymbol{r}}(g)). \tag{4.73}$$

It is obvious from equations (4.69) and (4.71) that our estimate $g_{4,SAMSE}$ will depend on the functionals, $\psi_{\boldsymbol{r}+2\boldsymbol{e}_i}$, for $|\boldsymbol{r}|=4.$ But the functionals, $\psi_{\boldsymbol{r}+2\boldsymbol{e}_i}$ also depends on $\boldsymbol{\Psi}_6$. So we need the bandwidth $g_{4,SAMSE}$, that is the minimize of $SAMSE(\hat{\boldsymbol{\Psi}}_6)$. Therefore, in general, we need *SAMSE* optimal bandwidth $g_{j,SAMSE}$ for $j=|\boldsymbol{r}|=4,6,8,...$, which is available in closed form as shown by (Duong and Hazelton, 2003). So from equation (4.69),

$$SAMSE_j(g) = \sum_{\boldsymbol{r}:|\boldsymbol{r}|=j} AMSE(\hat{\psi}_{\boldsymbol{r}}(g)) = \frac{n^{-j-1}}{2} g^{-2j-1}\psi_0 \sum_{\boldsymbol{r}:|\boldsymbol{r}|=j} R(K^{(\boldsymbol{r})}) + \frac{1}{2} g^4 \mu_2{}^2(K) \sum_{\boldsymbol{r}:|\boldsymbol{r}|=j} \psi^2_{\boldsymbol{r}+2\boldsymbol{e}_i}. \tag{4.74}$$

Now, letting $A_1 = \sum_{\boldsymbol{r}:|\boldsymbol{r}|=j} R(K^{(\boldsymbol{r})})$ and $A_2 = \mu_2^2(K) \sum_{\boldsymbol{r}:|\boldsymbol{r}|=j} \psi^2_{\boldsymbol{r}+2\boldsymbol{e}_i}$, then

$$SAMSE_j(g) = \frac{n^{-j-1}}{2} g^{-2j-1}\psi_0 A_1 + \frac{1}{4} g^4 A_2. \tag{4.75}$$

Now, differentiating the above with respect to g, we get

$$\frac{\partial}{\partial g} SAMSE_j(g) = \frac{n^{-j-1}}{2}(2j-1)g^{-2j-2}\psi_0 A_1 + g^3 A_2. \tag{4.76}$$

Setting equation (4.76) to zero and dividing through by g^3, we have

$$-\frac{n^{-j-1}}{2}\left(j+1\right)g^{-2j-5}\psi_0 A_1 + A_2 = 0.$$

Therefore,

$$g^{-2j-5} = \frac{2A_2}{n^{-j-1}\left(j+1\right)\psi_0 A_1};$$

which implies that,

$$g_{j,SAMSE} = \left[\frac{\left(j+1\right)\psi_0 A_1}{2n^{j+1}A_2}\right]^{1/\left(j+5\right)}. \tag{4.77}$$

CHAPTER V

MULTIVARIATE KERNEL DENSITY ESTIMATION – NEW APPROACH

5.1 Introduction

The estimate we obtained in the previous Chapters which completely break the curse of dimensionality and remedy the deficiency of high dimension bandwidth selection was based on the fact that it works well for self-revolving densities or the ellipsoidally symmetric distributions. We have obtained another solution in the form of an estimate that handles partially the curse of dimensionality. This new estimate is devoid of any restriction and works well for all densities unlike the one we previously obtained. As said earlier, it partially solves the problem of the curse of dimensionality which is better than the case of the classical or regular multivariate kernel density estimate.

5.2 Bivariate Case

Let $X_1,...,X_n$ be random sample of independently distributed observations from a population with unknown bivariate density $f(x_1, x_2)$, then we propose the new bivariate product kernel density estimator for the unknown density $f(x_1, x_2)$ to be

$$\hat{f}_n^*(x_1, x_2) = \frac{2}{n(n+1)h_1h_2} \sum_{i=1}^{n} K\left(\frac{x_1 - X_{1i}}{\sqrt{i}h_1}\right) K\left(\frac{x_2 - X_{2i}}{\sqrt{i}h_2}\right), \tag{5.1}$$

where K is the kernel which is assumed to be symmetric, $\int\int \hat{f}_n^*(x_1, x_2)dx_1dx_2 = 1$ whenever $\int K(u)du = 1$. Therefore $\hat{f}_n^*(x_1, x_2)$ is a density. h_1 and h_2 are the bandwidths or the smoothing parameters.

5.2.1 Expectation and Bias of $\hat{f}_n^*(x_1, x_2)$

We can obtain the expectation and the bias of this new estimator as follows:

$$E(\hat{f}_n^*(x_1, x_2)) = \frac{2}{n(n+1)h_1h_2} E\sum_{i=1}^{n} K\left(\frac{x_1 - X_{1i}}{\sqrt{i}h_1}\right) K\left(\frac{x_2 - X_{2i}}{\sqrt{i}h_2}\right).$$

$$= \frac{2}{n(n+1)h_1h_2} \sum_{i=1}^{n} \int\int K\left(\frac{x_1 - y_1}{\sqrt{i}h_1}\right) K\left(\frac{x_2 - y_2}{\sqrt{i}h_2}\right) f(y_1, y_2)dy_1dy_2. \tag{5.2}$$

Let $u = \frac{x_1 - y_1}{\sqrt{i}h_1}$, then $y_1 = x_1 - u\sqrt{i}h_1$ and $dy_1 = \sqrt{i}h_1du$.

Similarly, let $v = \frac{x_2 - y_2}{\sqrt{i}h_2}$, then $y_2 = x_2 - v\sqrt{i}h_2$ and $dy_2 = \sqrt{i}h_2dv$.

Therefore, substituting these into equation (5.2), we have

$$E(\hat{f}_n^*(x_1, x_2)) = \frac{2}{n(n+1)h_1h_2} \sum_{i=1}^{n} \int\int K(u)K(v)f(x_1 - u\sqrt{i}h_1, x_2 - v\sqrt{i}h_2)\sqrt{i}h_1du\sqrt{i}h_2dv$$

$$= \frac{2}{n(n+1)} \sum_{i=1}^{n} i \int\int K(u)K(v)f(x_1 - u\sqrt{i}h_1, x_2 - v\sqrt{i}h_2)dudv.$$

Now applying the Taylor series expansion to second order, we have

$$\frac{2}{n(n+1)}\sum_{i=1}^{n} i\int\int K(u)K(v)\left[f(x_1,x_2) - u\{\sqrt{i}h_1 f_1(x_1,x_2) + v\sqrt{i}h_2 f_2(x_1,x_2)\}\right.$$
$$\left. +1/2\,\{ih_1^2u^2 f_{11}(x_1,x_2) + iuvh_1h_2 f_{12}(x_1,x_2) + ivuh_2h_1 f_{21}(x_1,x_2) + ih_2^2v^2 f_{22}(x_1,x_2)\}\right]dudv$$

$$= \frac{2}{n(n+1)}\sum_{i=1}^{n} i\Big\{\iint K(u)K(v) f(x_1,x_2)dudv - \iint K(u)K(v)[u\sqrt{i}h_1 f_1(x_1,x_2) + v\sqrt{i}h_2 f_2(x_1,x_2)]dudv$$
$$+1/2\iint K(u)K(v)\{ih_1^2u^2 f_{11}(x_1,x_2)\}dudv + 1/2\iint K(u)K(v)\{iuvh_1h_2 f_{12}(x_1,x_2)\}dudv$$
$$+1/2\iint K(u)K(v)\{ivuh_2h_1 f_{21}(x_1,x_2)\}dudv + 1/2\iint K(u)K(v)\{ih_2^2v^2 f_{22}(x_1,x_2)\}dudv\Big\}.$$

Now applying the moment conditions of kernel, that is $\int uK(u)du,\ \int vK(v)dv = 0,$

$\int u^2K(u)du = \sigma_K^2$, and $\int v^2K(v)dv = \sigma_K^2$, we have

$$\frac{2}{n(n+1)}\sum_{i=1}^{n} i\left\{f(x_1,x_2) + 1/2\,i^2h_1^2\sigma_K^2 f_{11}(x_1,x_2) + 1/2\,i^2h_2^2\sigma_K^2 f_{22}(x_1,x_2)\right\}$$

$$= \frac{2}{n(n+1)}\left\{f(x_1,x_2)\sum_{i=1}^{n} i + 1/2\,h_1^2\sigma_K^2 f_{11}(x_1,x_2)\sum_{i=1}^{n} i^3 + 1/2\,h_2^2\sigma_K^2 f_{22}(x_1,x_2)\sum_{i=1}^{n} i^3\right\}.$$

Applying the power summation, we have

$$\frac{2}{n(n+1)}\left\{f(x_1,x_2)\frac{n(n+1)}{2} + 1/2\,h_1^2\sigma_K^2 f_{11}(x_1,x_2)\frac{n^2(n+1)^2}{4} + 1/2\,h_2^2\sigma_K^2 f_{22}(x_1,x_2)\frac{n^2(n+1)^2}{4}\right\}$$

$$= f(x_1,x_2) + \frac{n(n+1)}{4}h_1^2\sigma_K^2 f_{11}(x_1,x_2) + \frac{n(n+1)}{4}h_2^2\sigma_K^2 f_{22}(x_1,x_2).$$

Therefore

$$E\left(\hat{f}_n^*(x_1,x_2)\right)= f(x_1,x_2)+\frac{n(n+1)}{4}\sigma_K^2\left[h_1^2 f_{11}(x_{1,}x_2)+h_2^2 f_{22}(x_{1,}x_2)\right]. \quad (5.3)$$

Therefore, the bias of $\hat{f}_n^*(x_1,x_2)$ is

$$Bias\left[\hat{f}_n^*(x_1,x_2)\right]=\frac{n(n+1)}{4}\sigma_K^2\left[h_1^2 f_{11}(x_{1,}x_2)+h_2^2 f_{22}(x_{1,}x_2)\right]. \quad (5.4)$$

The asymptotic square bias (AISB) is given by

$$AISB\left[\hat{f}_n^*(x_1,x_2)\right]=\frac{n^2(n+1)^2}{16}\sigma_K^4[h_1^4\int\int f_{11}^2(x_{1,}x_2)dx_1dx_2+2h_1^2h_2^2\int\int f_{11}(x_1,x_2)f_{22}(x_1,x_2)dx_1dx_2$$
$$+h_2^4\iint f_{22}^2(x_{1,}x_2)dx_1dx_2].$$

Therefore, AISB of $\hat{f}_n^*(x_1,x_2)$ is given by

$$AISB\left[\hat{f}_n^*(x_1,x_2)\right]=\frac{n^2(n+1)^2}{16}\sigma_K^4[h_1^4R(f_{11})+2h_1^2h_2^2S(f_{11},f_{22})+h_2^4R(f_{22})]. \quad (5.5)$$

where $R(f_{11})=\iint f_{11}^2(x_1,x_2)dx_1dx_2$, $R(f_{22})=\iint f_{22}^2(x_1,x_2)dx_1dx_2$ and

$S(f_{11},f_{22})=\int\int f_{11}(x_1,x_2)f_{22}(x_1,x_2)dx_1dx_2$.

5.2.2 Variance and AMISE of $\hat{f}_n^*(x_1,x_2)$

Squaring equation (5.1) yields

$$\hat{f}_n^{*2}(x_1,x_2)=\frac{4}{n^2(n+1)^2h_1^2h_2^2}\left\{\sum_{j\neq l}\sum K\left(\frac{x_1-X_{1j}}{\sqrt{i}h_1}\right)K\left(\frac{x_2-X_{2j}}{\sqrt{i}h_2}\right)K\left(\frac{x_1-X_{1l}}{\sqrt{i}h_1}\right)K\left(\frac{x_2-X_{2l}}{\sqrt{i}h_2}\right)\right.$$
$$\left.+\sum_i K^2\left(\frac{x_1-X_{1i}}{\sqrt{i}h_1}\right)K^2\left(\frac{x_2-X_{2i}}{\sqrt{i}h_2}\right)\right\}.$$

Therefore,

$$E\left[\hat{f}_n^{*2}(x_1,x_2)\right]=\frac{4}{n^2(n+1)^2h_1^2h_2^2}E\sum_{i=1}^{n}K^2\left(\frac{x_1-X_{1i}}{\sqrt{i}h_1}\right)K^2\left(\frac{x_2-X_{2i}}{\sqrt{i}h_2}\right)$$

$$=\frac{4}{n^2(n+1)^2h_1^2h_2^2}\sum_{i=1}^{n}\iint K^2\left(\frac{x_1-y_1}{\sqrt{i}h_1}\right)K^2\left(\frac{x_1-y_2}{\sqrt{i}h_2}\right)f(y_1,y_2)\,dy_1dy_2. \quad (5.6)$$

Let $u=\frac{x_1-y_1}{\sqrt{i}h_1}$, then $y_1=x_1-u\sqrt{i}h_1$ and $dy_1=\sqrt{i}h_1du$.

Similarly, let $v=\frac{x_2-y_2}{\sqrt{i}h_2}$, then $y_2=x_2-v\sqrt{i}h_2$ and $dy_2=\sqrt{i}h_2dv$.

Therefore, substituting these into equation (5.6), we have

$$E\left[\hat{f}_n^{*2}(x_1,x_2)\right]=\frac{4}{n^2(n+1)^2h_1^2h_2^2}\sum_{i=1}^{n}\iint K^2(u)K^2(v)f\left(x_1-u\sqrt{i}h_1,x_2-v\sqrt{i}h_2\right)\sqrt{i}h_1du\sqrt{i}h_2dv$$

$$=\frac{4}{n^2(n+1)^2h_1h_2}\sum_{i=1}^{n}i\iint K^2(u)K^2(v)f\left(x_1-u\sqrt{i}h_1,x_2-v\sqrt{i}h_2\right)dudv.$$

.

Now applying the Taylor series expansion to first order, we have

$$\frac{4}{n^2(n+1)^2h_1h_2}\sum_{i=1}^{n}i\iint K^2(u)K^2(v)f(x_1,x_2)\,dudv$$

$$=\frac{4f(x_1,x_2)R^2(K)}{n^2(n+1)^2h_1h_2}\sum_{i=1}^{n}i\,,$$

where $\int K^2(u)du=R(K)$ and $\int K^2(v)du=R(K)$.

Applying the power summation, we have

$$= \frac{4 f(x_1, x_2) R(K)^2}{n^2 (n+1)^2 h_1 h_2} \frac{n(n+1)}{2} = \frac{2 f(x_1, x_2) R(K)^2}{n(n+1) h_1 h_2}.$$

Since the variance term is dominated by the $E\left[\hat{f}_{\bar{x}}^2 \right]$, then the variance of $\hat{f}_n^*(x_1, x_2)$ is given by

$$Var\left[\hat{f}_n^*(x_1, x_2)\right] \approx \frac{2 f(x_1, x_2) R(K)^2}{n(n+1) h_1 h_2}. \tag{5.7}$$

Therefore, the asymptotic integrated variance (AIV) is given by

$$AIV\left[\hat{f}_n^*(x_1, x_2)\right] = \iint \frac{2 f(x_1, x_2) R(K)^2}{n(n+1) h_1 h_2} dx_1 dx_2 = \frac{2R(K)^2}{n(n+1) h_1 h_2}. \tag{5.8}$$

Therefore, the asymptotic mean squared integrated error is given by

$$\begin{aligned} AMISE &= AISB + AIV \\ &= \frac{n^2 (n+1)^2}{16} \sigma_K^4 [h_1^4 R(f_{11}) + 2h_1^2 h_2^2 \iint f_{11}(x_1, x_2) f_{22}(x_1, x_2) dx_1 dx_2 + h_2^4 R(f_{22})] \\ &\quad + \frac{2R(K)^2}{n(n+1) h_1 h_2}, \end{aligned} \tag{5.9}$$

where $R(f_{11}) = \iint f_{11}^2(x_1, x_2) dx_1 dx_2$ and $R(f_{22}) = \iint f_{22}^2(x_1, x_2) dx_1 dx_2$.

We can easily find a solution for the optimum bandwidth if $h_i = h$ for all i.

Therefore, equation (5.9) becomes

$$\begin{aligned} AMISE &= \frac{n^2 (n+1)^2}{16} \sigma_K^4 [h^4 R(f_{11}) + 2h^4 \iint f_{11}(x_1, x_2) f_{22}(x_1, x_2) dx_1 dx_2 + h^4 R(f_{22})] \\ &\quad + \frac{2R(K)^2}{n(n+1) h^2}. \end{aligned}$$

Differentiating with respect to h and equating to zero, we have

$$\frac{n^2(n+1)^2}{16}\sigma_K^4[4h^3R(f_{11})+8h^3\iint f_{11}(x_1,x_2)f_{22}(x_1,x_2)dx_1dx_2+4h^3R(f_{22})]$$
$$-\frac{4R(K)^2}{n(n+1)h^3}=0.$$

$$h^6R(f_{11})+2h^6\iint f_{11}(x_1,x_2)f_{22}(x_1,x_2)dx_1dx_2+h^6R(f_{22})=\frac{16R(K)^2}{n^3(n+1)^3\sigma_K^4}. \qquad (5.10)$$

From Scott (1992), the general bivariate normal data using a normal kernel gives

$$R(f_{11})=3\left[16\pi(1-\rho^2)^{5/2}\sigma_1^5\sigma_2\right]^{-1},$$

$$R(f_{22})=3\left[16\pi(1-\rho^2)^{5/2}\sigma_1\sigma_2^5\right]^{-1},$$

and

$$\int_{\Re^2} f_{11}f_{22}dx_1dx_2=(1+2\rho^2)\left[16\pi(1-\rho^2)^{5/2}\sigma_1^3\sigma_2^3\right]^{-1},$$

where ρ is the correlation coefficient.

Substituting these into equation (5.10) we have

$$3h^6\left[16\pi(1-\rho^2)^{5/2}\sigma_1^5\sigma_2\right]^{-1}+2h^6(1+2\rho^2)\left[16\pi(1-\rho^2)^{5/2}\sigma_1^3\sigma_2^3\right]^{-1}$$
$$+3h^6\left[16\pi(1-\rho^2)^{5/2}\sigma_1\sigma_2^5\right]^{-1}=\frac{16}{4\pi n^3(n+1)^3\sigma_K^4}.$$

$$h^6\left[\frac{3}{\left[16\pi(1-\rho^2)^{5/2}\sigma_1^5\sigma_2\right]}+\frac{2(1+2\rho^2)}{\left[16\pi(1-\rho^2)^{5/2}\sigma_1^3\sigma_2^3\right]}+\frac{3}{\left[16\pi(1-\rho^2)^{5/2}\sigma_1\sigma_2^5\right]}\right]=\frac{16}{4\pi n^3(n+1)^3}.$$

Assuming the standard deviation $\sigma_i = \sigma$ for all i, then we have

$$h^6 \left[\frac{3\sigma^4 + 2(1+2\rho^2)\sigma^4 + 3\sigma^4}{\left[16\pi (1-\rho^2)^{5/2} \sigma^{10}\right]} \right] = \frac{4}{\pi n^3 (n+1)^3}$$

$$h^6 = \frac{64\pi (1-\rho^2)^{5/2} \sigma^{10}}{8\pi n^3 (n+1)^3 \sigma^4 (1+\rho^2/2)}$$

$$h = \left[\frac{8(1-\rho^2)^{5/2} \sigma^6}{n^3 (n+1)^3 (1+\rho^2/2)} \right]^{1/6}.$$

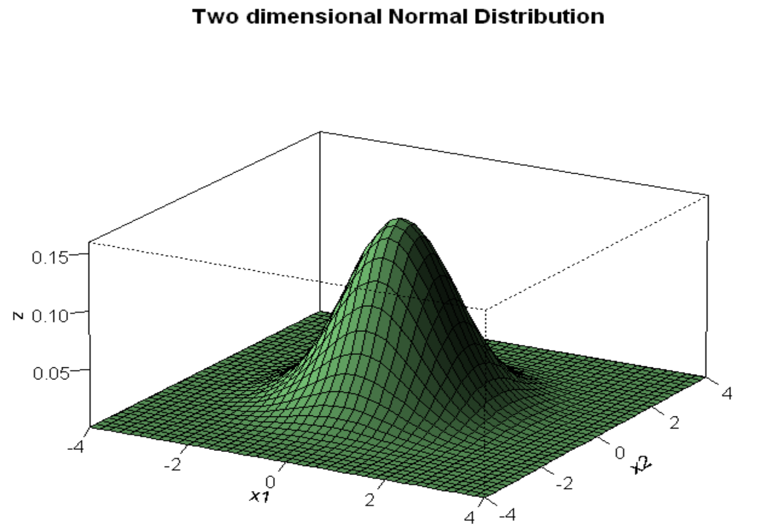
Therefore

$$h_i = n^{-1} \left[\frac{8(1-\rho^2)^{5/2} \sigma_i^6}{(1+1/n)^3 (1+\rho^2/2)} \right]^{1/6}. \tag{5.11}$$

If the variables are independent, then $\rho = 0$ and

$$h_i = n^{-1} \left[\frac{8\sigma_i^6}{(1+1/n)^3} \right]^{1/6}. \tag{5.12}$$

Figure 5.2.1 The Bivariate Normal Density Distribution



$$f(\mathbf{x}) = \frac{1}{2\pi\sqrt{\sigma_{11}\,\sigma_{22}\,(1-\rho^2)}} \cdot \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(x_1-\mu_1)^2}{\sigma_{11}} - 2\rho\,\frac{x_1-\mu_1}{\sqrt{\sigma_{11}}}\,\frac{x_2-\mu_2}{\sqrt{\sigma_{22}}} + \frac{(x_2-\mu_2)^2}{\sigma_{22}}\right]\right\}$$

Figure 5.2.2 Contour of the Bivariate Normal Density

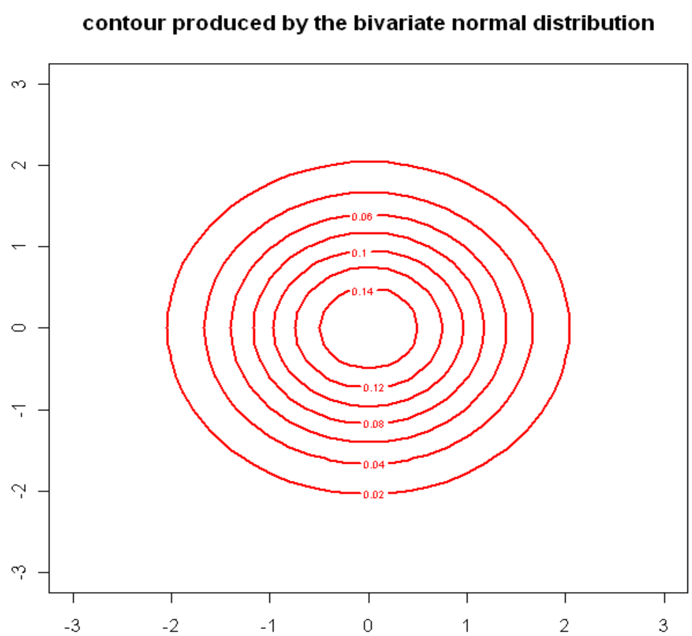


Figure 5.2.3 Regular Kernel Estimate of the Bivariate Normal Distribution

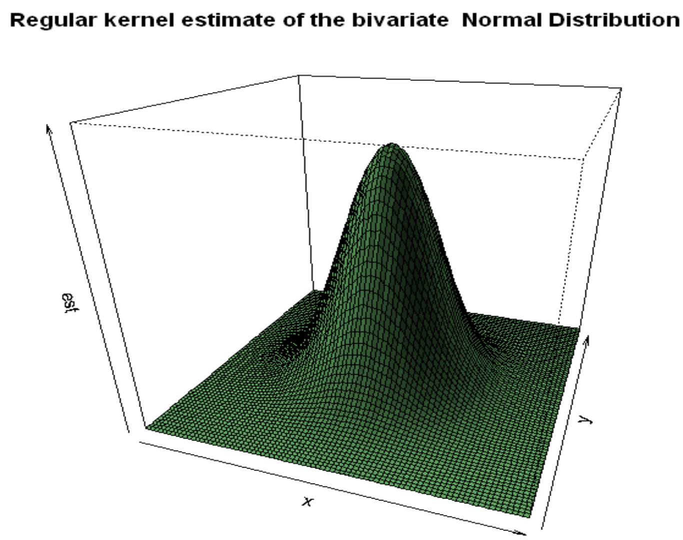


Figure 5.2.4 Contour of the Regular Kernel Estimate of the Bivariate Normal Density

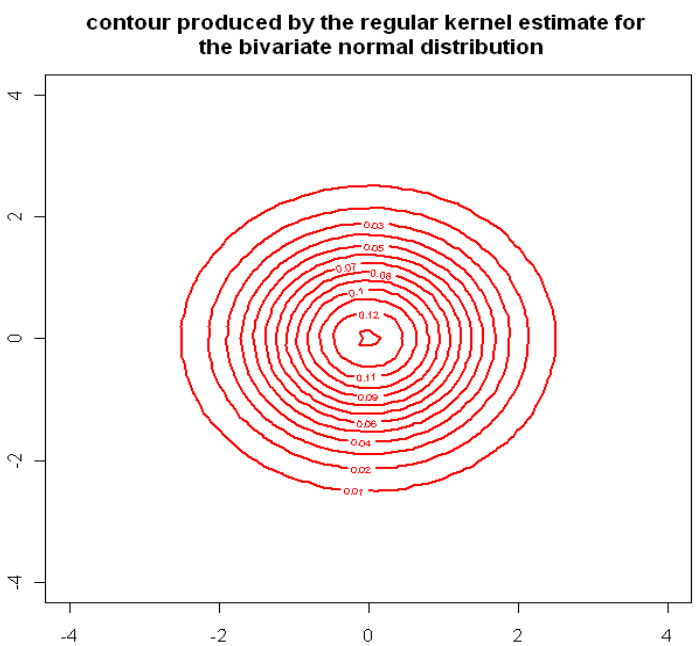


Figure 5.2.5 The New Kernel Estimate of the Bivariate Normal Distribution

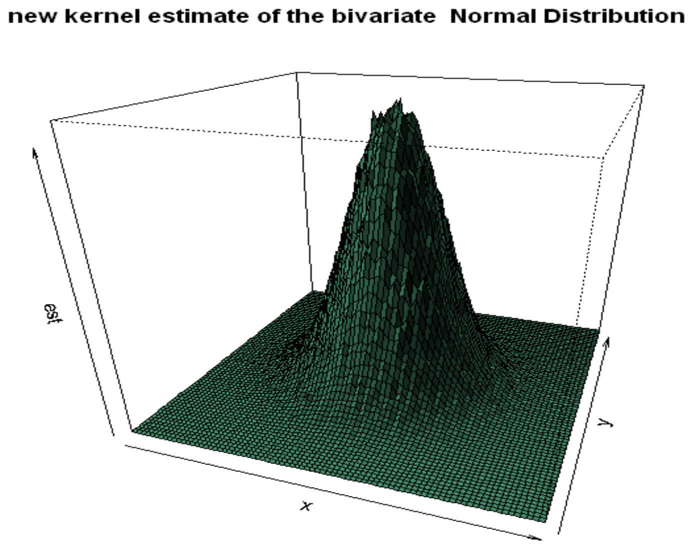
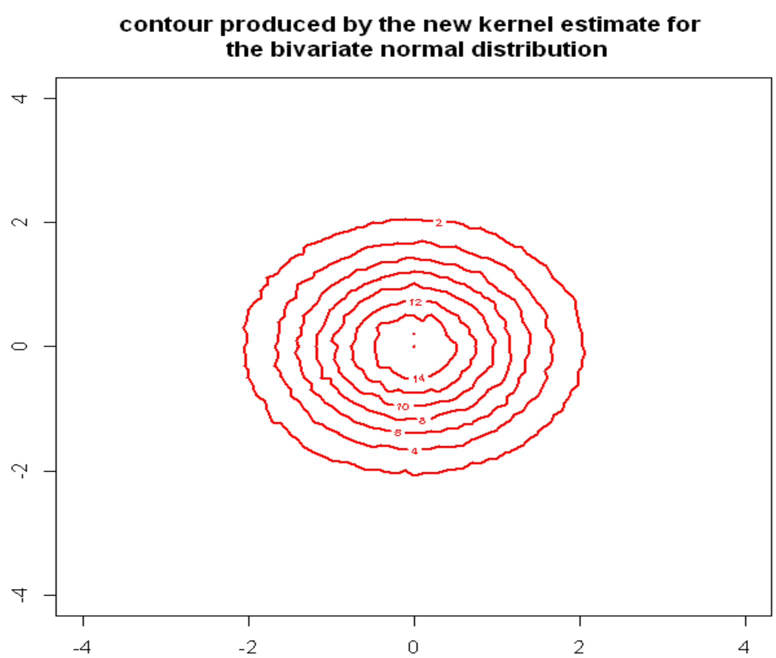


Figure 5.2.6 Contour of the New Kernel Estimate of the Bivariate Normal Distribution



5.3 Multivariate Product Kernel Case

Let $X_1,\ldots,X_n$ be random sample of independently distributed observations from a population with unknown density $f(\boldsymbol{x})$, of dimension p. Let $\boldsymbol{X}$ be an $n\times p$ data matrix of random vectors $\boldsymbol{x}=(x_1,\ldots,x_p)$ and let X_{ij} denote the ij[th] entry of $\boldsymbol{X}$. Then we propose the new multivariate product kernel density estimator for the unknown density $f(\boldsymbol{x})$ to be

$$\hat{f}_n^*(\boldsymbol{x})=\frac{2}{\prod_{j=1}^{p} h_j n(n+1)}\sum_{i=1}^{n}\prod_{j=1}^{p} K\left(\frac{\boldsymbol{x}_i - X_{ij}}{i^{1/p}h_j}\right), \tag{5.13}$$

where K is the kernel which is assumed to be symmetric, h_j are the respective smoothing parameters for each dimension.

5.3.1 Expectation and Bias of $\hat{f}_n^*(\boldsymbol{x})$

The expectation and the bias of this multivariate product kernel estimator as follows:

$$E\left[\hat{f}_n^*(\boldsymbol{x})\right]=E\,\frac{2}{\prod_{j=1}^{p} h_j n(n+1)}\sum_{i=1}^{n}\prod_{j=1}^{p} K\left(\frac{\boldsymbol{x}_i - X_{ij}}{i^{1/p}h_j}\right)$$

$$=\frac{2}{n(n+1)}\sum_{i=1}^{\mathrm{n}}\int\prod_{j=1}^{p}\frac{1}{h_j}K\left(\frac{\boldsymbol{x}_i-\boldsymbol{t}_j}{i^{1/p}h_j}\right)f(\boldsymbol{t})d\boldsymbol{t}\,. \tag{5.14}$$

Let $\boldsymbol{u}_j = \dfrac{\boldsymbol{x}_i - \boldsymbol{t}_j}{i^{1/p} h_j}$, then $\boldsymbol{t}_j = \boldsymbol{x}_i - \boldsymbol{u}_j i^{1/p} h_j$ and $d\boldsymbol{t}_j = i h_j d\boldsymbol{u}$.

Substituting these into equation (5.14), we have

$$\frac{2}{n(n+1)} \sum_{i=1}^{n} \int \prod_{j=1}^{p} \frac{1}{h_j} K(\mathbf{u}_j) f(\mathbf{x}_i - \mathbf{u}_j i^{1/p} h_j) i h_j d\mathbf{u}$$

$$= \frac{2}{n(n+1)} \sum_{i=1}^{n} i \int \prod_{j=1}^{p} K(\mathbf{u}_j) f(\mathbf{x}_i - \mathbf{u}_j i^{1/p} h_j) d\mathbf{u}$$

$$= \frac{2}{n(n+1)} \sum_{i=1}^{n} i \int \prod_{j=1}^{p} K(\mathbf{u}_j) f(x_1 - u_1 i^{1/p} h_1, \ldots, x_n - u_n i^{1/p} h_n) d\mathbf{u}.$$

Now applying the Taylor series expansion to second order, we have

$$\frac{2}{n(n+1)} \sum_{i=1}^{n} i \int \prod_{j=1}^{p} K(\boldsymbol{u}_j) \left[f(\boldsymbol{x}) - \sum_{r=1}^{p} i^{1/p} h_r \boldsymbol{u}_r f_r(\boldsymbol{x}) + \sum_{r,s=1}^{p} (i^{1/p})^2 \frac{h_r h_s}{2} \boldsymbol{u}_r \boldsymbol{u}_s f_{rs}(\boldsymbol{x}) \right] d\boldsymbol{u}.$$

Also applying the moment conditions of kernel, we have

$$\frac{2}{n(n+1)} \sum_{i=1}^{n} i \left[f(\boldsymbol{x}) + \frac{1}{2} \sigma_K^2 i^2 \sum_{r,s=1}^{p} h_j^2 f_{jj}(\boldsymbol{x}) \right]$$

$$= \frac{2}{n(n+1)} \sum_{i=1}^{n} i \left[f(\boldsymbol{x}) + \frac{1}{2} \sigma_K^2 i^2 \sum_{r,s=1}^{p} h_j^2 f_{jj}(\boldsymbol{x}) \right].$$

Applying the power summation, we have

$$= \frac{2}{n(n+1)} \left[f(\boldsymbol{x}) \sum_{i=1}^{n} i + \frac{1}{2} \sigma_K^2 \sum_{i=1}^{n} i^3 \sum_{r,s=1}^{p} h_j^2 f_{jj}(\boldsymbol{x}) \right]$$

$$= \frac{2}{n(n+1)}\left[f(x)\frac{n(n+1)}{2} + \frac{1}{2}\frac{n^2(n+1)^2}{4}\sigma_K^2 \sum_{r,s=1}^{p} h_j^2 f_{jj}(x)\right]$$

$$= f(x) + \frac{n(n+1)}{4}\sigma_K^2 \sum_{j=1}^{p} h_j^2 f_{jj}(x).$$

So,

$$E\left[\hat{f}_n^*(x)\right] = f(x) + \frac{n(n+1)}{4}\sigma_K^2 \sum_{j=1}^{p} h_j^2 f_{jj}(x). \tag{5.15}$$

Therefore, the bias of $\hat{f}_n^*(x)$ is given by

$$Bias\left[\hat{f}_n^*(x)\right] = \frac{n(n+1)}{4}\sigma_K^2 \sum_{j=1}^{p} h_j^2 f_{jj}(x). \tag{5.16}$$

The asymptotic integrated square bias (AISB) is given by

$$AISB\left[\hat{f}_n^*(x)\right] = \frac{n^2(n+1)^2}{16}\sigma_K^4 \left[\sum_{i=1}^{p} h_i^4 R(f_{ii}) + \sum_{i\neq j} h_i^2 h_j^2 S(f_{ii}, f_{jj})\right], \tag{5.17}$$

where $R(f_{ii}) = \iint f_{ii}^2(x)\,dx$ and $S(f_{ii}, f_{jj}) = \int f_{ii} f_{jj} d\mathbf{x}$.

5.3.2 Variance and AMISE of $\hat{f}_n^*(x)$

Squaring equation (5.13) yields

$$\hat{f}_n^{*2}(x) = \frac{4}{n^2(n+1)^2 \prod_{j=1}^{p} h_j^2}\left\{\sum_{j\neq l}\sum\prod_{j=1}^{p} K^2\left(\frac{x_i - X_{ij}}{i^{1/p}h_1}\right)K^2\left(\frac{x_i - X_{ij}}{i^{1/p}h_2}\right) + \sum_{i}^{n}\prod_{j=1}^{p} K^2\left(\frac{x_i - X_{ij}}{i^{1/p}h_1}\right)\right\}.$$

Therefore,

$$E\left[\hat{f}_n^{*2}(x)\right] = \frac{4}{n^2(n+1)^2\prod_{j=1}^{p} h_j^2} E\sum_{i=1}^{n}\prod_{j=1}^{p} K^2\left(\frac{x_i - X_{ij}}{i^{1/p}h_j}\right)$$

$$= \frac{4}{n^2(n+1)^2}\sum_{i=1}^{n}\int\prod_{j=1}^{n}\frac{1}{h_j^2}K^2\left(\frac{\boldsymbol{x}_i - \boldsymbol{t}_j}{i^{1/p}h_j}\right) f(\boldsymbol{t})\,d\boldsymbol{t} \qquad (5.18)$$

Let $\boldsymbol{u}_j = \dfrac{\boldsymbol{x}_i - \boldsymbol{t}_j}{i^{1/p}h_j}$, then $\boldsymbol{t}_j = \boldsymbol{x}_i - \boldsymbol{u}_j i^{1/p}h_j$ and $d\boldsymbol{t}_j = ih_j d\boldsymbol{u}$.

Therefore, substituting these into equation (5.18), we have

$$\mathrm{E}\left[\hat{f}_n^{*2}(x)\right] = \frac{4}{n^2(n+1)^2}\sum_{i=1}^{n}\int\prod_{j=1}^{n}\frac{1}{h_j^2}K^2(\boldsymbol{u}_j) f(\boldsymbol{x} - \boldsymbol{u}_j i^{1/p}h_j)\, ih_j d\boldsymbol{u}$$

$$= \frac{4}{n^2(n+1)^2}\sum_{\mathrm{i}=1}^{\mathrm{n}} i\int\prod_{j=1}^{p}\frac{1}{h_j}K^2(\boldsymbol{u}_j) f(\boldsymbol{x}_i - \boldsymbol{u}_j i^{1/p}h_j)\, d\boldsymbol{u}.$$

Now applying the Taylor series expansion to first order, we have

$$\frac{4}{n^2(n+1)^2}\sum_{i=1}^{n} i\int\prod_{j=1}^{p}\frac{1}{h_j}K^2(\boldsymbol{u}_j) f(\boldsymbol{x})\, d\boldsymbol{u}$$

$$= \frac{4f(\boldsymbol{x})R^d(K)}{n^2(n+1)^2\prod_{j=1}^{p} h_j}\sum_{i=1}^{n} i$$

$$= \frac{4f(\boldsymbol{x})R^p(K)}{n^2(n+1)^2\prod_{j=1}^{p} h_j}\sum_{i=1}^{n} i,$$

where $\int \prod_{j=1}^{p} K^2(u_j) = R^p(K)$.

Applying the power summation, we have

$$\frac{4f(\boldsymbol{x})R^p(K)}{n^2(n+1)^2\prod_{j=1}^{p} h_j}\frac{n(n+1)}{2} = \frac{2f(\boldsymbol{x})R^p(K)}{n(n+1)\prod_{j=1}^{p} h_j}.$$

Since the variance term is dominated by the $E\left[\hat{f}_n^{*2}(\boldsymbol{x})\right]$, then the variance of $\hat{f}_n^{*}(\boldsymbol{x})$ is given by

$$Var\left[\hat{f}_n^{*}(\boldsymbol{x})\right] \approx \frac{2f(\boldsymbol{x})R^p(K)}{n(n+1)\prod_{j=1}^{p} h_j}. \tag{5.19}$$

Therefore, the asymptotic integrated variance (*AIV*) is given by

$$AIV\left[\hat{f}_n^{*}(\boldsymbol{x})\right] = \int \frac{2f(\boldsymbol{x})R^p(K)}{n(n+1)\prod_{j=1}^{p} h_j}d\boldsymbol{x} = \frac{2R(K)^p}{n(n+1)\prod_{j=1}^{p} h_j}. \tag{5.20}$$

Therefore, the asymptotic mean squared integrated error is given by

$$\begin{aligned} AMISE &= AISB + AIV \\ &= \frac{n^2(n+1)^2}{16}\sigma_K^4\left[\sum_{i=1}^{p} h_i^4 R(f_{ii}) + \sum_{i\neq j} h_i^2 h_j^2 S(f_{ii}, f_{jj})\right] + \frac{2R(K)^p}{n(n+1)\prod_{j=1}^{p} h_j}. \end{aligned} \tag{5.21}$$

5.4 Multivariate Case-General

If $\boldsymbol{X_i}$ is a independent random sample from a population with unknown density, $f(\boldsymbol{x})$, and $\boldsymbol{X_i} = (X_{i1},...,X_{ip})^T$, and $\boldsymbol{x} \in \Re^p$ has the representation $\boldsymbol{x} = (x_1,...,x_p)$. Then we propose the p-dimensional multivariate kernel density estimator as

$$\hat{f}_n^*(\boldsymbol{x}) = \frac{2}{n(n+1)|\boldsymbol{H}|}\sum_{i=1}^{n} K\left(\left(\frac{\boldsymbol{x}-\boldsymbol{X_i}}{i^{1/p}}\right)\boldsymbol{H}^{-1}\right), \tag{5.22}$$

where $\boldsymbol{H}$ is a positive definite $p \times p$ matrix called the bandwidth matrix or the smoothing matrix and K is the kernel which is assumed to be symmetric.

5.4.1 Expectation and Bias of $\hat{f}_n^*(\boldsymbol{x})$

$$E\left[\hat{f}_n^*(\boldsymbol{x})\right] = \frac{2}{n(n+1)|\boldsymbol{H}|} E\sum_{i=1}^{n} K\left[\left(\frac{\boldsymbol{x}-\boldsymbol{X}_i}{i^{1/p}}\right)(\boldsymbol{H}^{-1})\right]$$

$$= \frac{2}{n(n+1)|\boldsymbol{H}|}\sum_{i=1}^{n} \int K\left[\left(\frac{\boldsymbol{x}-\boldsymbol{y}}{i^{1/p}}\right)(\boldsymbol{H}^{-1})\right] f(\boldsymbol{y})\, d\boldsymbol{y}, \tag{5.23}$$

where $d\boldsymbol{y} = dy_1,...,dy_p$.

Let $\boldsymbol{u} = \left(\frac{\boldsymbol{x}-\boldsymbol{y}}{i^{1/p}}\right)H^{-1}$, then $\boldsymbol{y} = \boldsymbol{x} - \boldsymbol{u}i^{1/p}\boldsymbol{H}$ and $d\boldsymbol{y} = i|\boldsymbol{H}|d\boldsymbol{u}$.

Substituting the above into equation (5.23) we have

$$E\left[\hat{f}_n^*(\boldsymbol{x})\right] = \frac{2}{n(n+1)|\boldsymbol{H}|}\sum_{i=1}^{n} \int K(\boldsymbol{u}) f(\boldsymbol{x} - i^{1/p}\boldsymbol{uH})\, i|\boldsymbol{H}|d\boldsymbol{u}$$

$$= \frac{2}{n(n+1)}\sum_{i=1}^{n} i\int K(\boldsymbol{u}) f(\boldsymbol{x} - i^{1/p}\boldsymbol{uH})\, d\boldsymbol{u}.$$

Now applying the Taylor series expansion to second order, we have

$$= \frac{2}{n(n+1)}\sum_{i=1}^{n} i\left[\int K(\boldsymbol{u})\{f(\boldsymbol{x}) - \boldsymbol{u}^T i^{1/p}\boldsymbol{H}\nabla f(\boldsymbol{x}) + 1/2\, i^2\boldsymbol{u}^T\boldsymbol{H}^T\nabla^2 f(\boldsymbol{x})\boldsymbol{Hu}\}\, d\boldsymbol{u}\right]$$

$$= \frac{2}{n(n+1)} \sum_{i=1}^{n} i \left[\int K(\boldsymbol{u}) f(\boldsymbol{x}) d\boldsymbol{u} - \int K(\boldsymbol{u}) \boldsymbol{u}^T i^{1/p} \boldsymbol{H} \nabla f(\boldsymbol{x}) d\boldsymbol{u} + \int K(\boldsymbol{u}) 1/2 i^2 \boldsymbol{u}^T \boldsymbol{H}^T \nabla^2 f(\boldsymbol{x}) \boldsymbol{H} \boldsymbol{u} d\boldsymbol{u} \right] \tag{5.24}$$

We notice that the quadratic form in the equation (5.21) above is a 1×1 matrix, which is equal to its trace.

Applying the moment conditions for multivariate kernel and the first two terms and using the trace identity and exchanging the integral operations and the trace, we have

$$\frac{2}{n(n+1)} \sum_{i=1}^{n} i \left[f(\boldsymbol{x}) + 1/2 i^2 tr \left\{ \int \boldsymbol{u}\boldsymbol{u}^T K(\boldsymbol{u}) d\boldsymbol{u} \boldsymbol{H}^T \nabla^2 f(\boldsymbol{x}) \boldsymbol{H} \right\} \right].$$

Now applying the moment conditions for multivariate kernel on the second term we have

$$\frac{2}{n(n+1)} \sum_{i=1}^{n} i \left[f(\boldsymbol{x}) + 1/2 i^2 tr \left\{ \boldsymbol{H}^T \nabla^2 f(\boldsymbol{x}) \boldsymbol{H} \right\} \right],$$

since the covariance of K is assumed to be I_p.

So we have

$$\frac{2}{n(n+1)} \left[f(\boldsymbol{x}) \sum_{i=1}^{n} i + 1/2 tr \left\{ \boldsymbol{H}^T \nabla^2 f(\boldsymbol{x}) \boldsymbol{H} \right\} \sum_{i=1}^{n} i^3 \right]$$

$$= \frac{2}{n(n+1)} f(\boldsymbol{x}) \frac{n(n+1)}{2} + \frac{2}{n(n+1)} 1/2 tr \left\{ \boldsymbol{H}^T \nabla^2 f(\boldsymbol{x}) \boldsymbol{H} \right\} \frac{n^2 (n+1)^2}{4}$$

$$= f(\boldsymbol{x}) + \frac{n(n+1)}{4} tr \left\{ \boldsymbol{H}^T \nabla^2 f(\boldsymbol{x}) \boldsymbol{H} \right\}.$$

Therefore

$$E\left[\hat{f}_n^*(\boldsymbol{x})\right] = f(\boldsymbol{x}) + \frac{n(n+1)}{4} tr\left(\boldsymbol{H}^T \nabla^2 f(\boldsymbol{x}) \boldsymbol{H}\right). \quad (5.25)$$

Therefore, the bias of $\hat{f}_n^*(\boldsymbol{x})$ is

$$Bias\left[\hat{f}_n^*(\boldsymbol{x})\right] = \frac{n(n+1)}{4} tr\left(\boldsymbol{H}^T \nabla^2 f(\boldsymbol{x}) \boldsymbol{H}\right) = \frac{n(n+1)}{4} tr\left(\boldsymbol{H}\boldsymbol{H}^T \nabla^2 f(\boldsymbol{x})\right). \quad (5.26)$$

The asymptotic square bias (AISB) is given by

$$AISB\left[\hat{f}_n^*(\boldsymbol{x})\right] = \frac{n^2(n+1)^2}{16} \int \left[tr\left(\boldsymbol{H}\boldsymbol{H}^T \nabla^2 f(\boldsymbol{x})\right)\right]^2 d\boldsymbol{x}. \quad (5.27)$$

5.4.2 Variance and AMISE of $\hat{f}_n^*(\boldsymbol{x})$

Squaring equation (5.22) yields

$$\hat{f}_n^{*2}(\boldsymbol{x}) = \frac{4}{n^2(n+1)^2|\boldsymbol{H}|^2} \sum_{i=1}^{n} K^2\left[\left(\frac{\boldsymbol{x}-\boldsymbol{X}_i}{i^{1/p}}\right)\boldsymbol{H}^{-1}\right]$$

$$E\left[\hat{f}_n^{*2}(\boldsymbol{x})\right] = \int \frac{4}{n^2(n+1)^2|\boldsymbol{H}|^2} \sum_{i=1}^{n} K^2\left[\left(\frac{\boldsymbol{x}-\boldsymbol{y}}{i^{1/p}}\right)\boldsymbol{H}^{-1}\right] f(\boldsymbol{y})\, d\boldsymbol{y}, \quad (5.28)$$

where $d\boldsymbol{y} = dy_1,...,dy_p$.

Let $\boldsymbol{u} = \left(\frac{\boldsymbol{x}-\boldsymbol{y}}{i^{1/p}}\right)H^{-1}$, then $\boldsymbol{y} = \boldsymbol{x} - \boldsymbol{u}i^{1/p}\boldsymbol{H}$ and $d\boldsymbol{y} = i|\boldsymbol{H}|d\boldsymbol{u}$.

Substituting the above into equation (5.28) we have

$$E\left[\hat{f}_n^{*2}(\boldsymbol{x})\right] = \frac{4}{n^2(n+1)^2|\boldsymbol{H}|^2} \sum_{i=1}^{n} \int K^2(\boldsymbol{u}) f\left(\boldsymbol{x} - i^{1/p}\boldsymbol{u}\boldsymbol{H}\right) i|\boldsymbol{H}|d\boldsymbol{u}$$

$$= \frac{4}{n^2(n+1)^2|H|}\sum_{i=1}^{n} i \int K^2(\boldsymbol{u}) f(\boldsymbol{x} - i^{1/p}\boldsymbol{uH})\, d\boldsymbol{u}.$$

Applying Taylor series expansion to the first order and the moment conditions, we have

$$\frac{4}{n^2(n+1)^2|H|}\sum_{i=1}^{n} i f(\boldsymbol{x}) R(K)$$

$$= \frac{4}{n^2(n+1)^2|H|}\frac{n(n+1)}{2} f(\boldsymbol{x}) R(K).$$

Therefore

$$\mathrm{E}\left[f_{\mathrm{n}}^{*2}(\boldsymbol{x})\right] = \frac{2 f(\boldsymbol{x}) R(K)}{n(n+1)|H|},$$

where $R(K) = \int K^2(\boldsymbol{u})\, d\boldsymbol{u}$.

Since the variance term is dominated by the $E\left[f_n^{*2}\right]$, then

$$Var\left[f_n^{*}(\boldsymbol{x})\right] \approx \frac{2 f(\boldsymbol{x}) R(K)}{n(n+1)|\boldsymbol{H}|}. \qquad (5.29)$$

From equation (5.29), the asymptotic integrated variance (AIV) is given by

$$AIV\left[f_n^{*}(\boldsymbol{x})\right] \approx \frac{2R(K)}{n(n+1)|\boldsymbol{H}|}. \qquad (5.30)$$

Therefore, the asymptotic mean squared error *(*AMISE*)* is

$$AMISE\left[f_n^{*}(\boldsymbol{x})\right] = \frac{n^2(n+1)^2}{16}\int \left[tr\left(\boldsymbol{H}^T \nabla^2 f(\boldsymbol{x})\right)\right]^2 d\boldsymbol{x} + \frac{2R(K)}{n(n+1)|\boldsymbol{H}|}. \qquad (5.31)$$

5.4.3 An Important Special Case

Now, let's define a scalar $h > 0$ and a $p \times p$ matrix $\boldsymbol{Q}$ such that $\boldsymbol{H} = h\boldsymbol{Q}$, where $|\boldsymbol{Q}| = 1$.

The idea behind choosing the matrix $\boldsymbol{Q}$ to have a unit determinant is that the elliptical shape of the kernel is controlled by the matrix $\boldsymbol{QQ}^T$ and the size of this kernel is governed by the scalar h, which is the bandwidth and which controls the amount of smoothness in the estimator for a given sample size n, where $h = (h_1, h_2, ..., h_p)^{1/p}$ is the geometric mean of the smoothing parameters. Hence

$$Bias\left[\hat{f}_n^*(\boldsymbol{x})\right] = \frac{n(n+1)}{4} h^2 tr\left(QQ^T \nabla^2 f(\boldsymbol{x})\right) . \tag{5.32}$$

Therefore, the asymptotic integrated squared bias (AISB) is given by

$$AISB\left[\hat{f}_n^*(\boldsymbol{x})\right] = \frac{n^2(n+1)^2}{16} h^4 \int \left[tr\left(QQ^T \nabla^2 f(\boldsymbol{x})\right)\right]^2 d\boldsymbol{x} . \tag{5.33}$$

In the above special case, where $\boldsymbol{H} = h\boldsymbol{Q}$, $|\boldsymbol{Q}| = 1$, the asymptotic integrated variance will be given by

$$AIV\left[\hat{f}_n^*(\boldsymbol{x})\right] \approx \frac{2R(K)}{n(n+1)h^p} , \tag{5.34}$$

since $\boldsymbol{Q}$ is a $p \times p$ matrix with unit determinant.

In this case, for a sufficiently smooth density function $f(\boldsymbol{x})$, the multivariate mean integrated square error (MISE) is asymptotically given by

$$AMISE = AIV + AISB$$

$$AMISE = \frac{2R(K)}{n(n+1)h^p} + \frac{n^2(n+1)^2}{16} h^4 \int [\mathrm{tr}\{Q^T \nabla^2 f(\mathbf{x})\}]^2 d\mathbf{x} \quad . \tag{5.35}$$

As it was said in the introduction, the choice of the bandwidth h is well-known to be crucial and of great importance since it controls the smoothness of the estimator $\hat{f}(\mathbf{x})$. Choosing the bandwidth, h involves a trade-off between the variance and the bias of the estimate as can be seen in the AMISE above. Therefore, an intermediate value of the bandwidth must be chosen to control both the bias and the variance simultaneously and allowing the bandwidth h to slowly decrease as the sample size increases for a better performance of the estimate.

In the multivariate standard case parameterized by $\boldsymbol{H} = h\boldsymbol{Q}$, where $|\boldsymbol{Q}| = 1$, the AMISE is given by

$$AMISE = \frac{R(K)}{nh^p} + \frac{1}{4} h^4 \int [\mathrm{tr}\{Q^T \nabla^2 f(\mathbf{x})\}]^2 d\mathbf{x} \tag{5.36}$$

where p is the dimension.

In this special case, the optimum h can be obtained by differentiating equation (5.36) and equating it to zero as shown below:

$$\frac{\partial (AMISE)}{\partial h} = \frac{-2pR(K)h^{-p-1}}{n(n+1)} + \frac{4n^2(n+1)^2}{16} h^3 \int [\mathrm{tr}\{Q^T \nabla^2 f(\mathbf{x})\}]^2 d\mathbf{x}.$$

Equating this to zero yields

$$\frac{2ph^{-p-1}R(K)}{n(n+1)} = \frac{n^2(n+1)^2}{4}h^3\int\left[tr\left\{Q^T\nabla^2 f(\mathbf{x})\right\}\right]^2 d\mathbf{x}$$

$$\frac{2ph^{-p-1}R(K)}{n(n+1)} = \frac{n^2(n+1)^2}{4}h^3\int\left[tr\left\{Q^T\nabla^2 f(\mathbf{x})\right\}\right]^2 d\mathbf{x}.$$

Thus

$$h^{p+4} = \frac{8pR(K)}{n^3(n+1)^3\int\left[tr\left\{Q^T\nabla^2 f(\mathbf{x})\right\}\right]^2 d\mathbf{x}} \; .$$

Hence

$$h_{opt} = n^{-6/(4+p)}\left[\frac{8pR(K)}{(1+1/n)^3\int\left[tr\left\{Q^T\nabla^2 f(\mathbf{x})\right\}\right]^2 d\mathbf{x}}\right]^{1/(p+4)} . \qquad (5.37)$$

We know that Scott (1992) gave the optimum bandwidth for the regular or the classical multivariate kernel density estimator as below:

$$h_{opt} = n^{-1/(4+p)}\left[\frac{pR(K)}{\int\left[tr\left\{Q^T\nabla^2 f(\mathbf{x})\right\}\right]^2 d\mathbf{x}}\right]^{1/(p+4)} .$$

Comparing the optimum bandwidth of our new estimate to the one obtained by Scott (1992), we realized that in both cases, the order of the optimum bandwidths depend on the dimension and the optimum bandwidths $h_{opt} \to 1$, a constant as $p \to \infty$. This will give a very rough estimate for large p since as $p \to \infty$ the optimum bandwidth is a constant irrespective of the sample size n. Therefore, they both experience the curse of dimensionality. But our new multivariate kernel density estimate will give a bandwidth smaller enough for large dimensions which will guarantee smoothness. This is not the case for the optimum bandwidth given by the regular or standard multivariate kernel estimate. Therefore, even though they both experience the curse of dimensionality, the effect will be minimal in our new multivariate kernel density estimate case compared to that of the regular multivariate kernel density case.

CHAPTER VI

CONCLUSIONS

We applied the idea of multivariate kernel density estimation to probability density functions which have the self-revolving characteristics or the ellipsoidally symmetric distributions. This time around we allowed the kernel to depend on the sample means rather than the sample data. We observed that the order of the optimum bandwidth that smoothens the density function is $n^{-1/2}$ which is independent of the dimension of the data used. This optimum bandwidth decreases as sample size increases to allow for a better performance of the estimate. The optimum bandwidth becomes small enough for larger dimension which guarantees smoothness. This breaks the "curse of dimensionality" and remedies the deficiency of high dimensional bandwidth selection as the optimum bandwidth gets smaller enough for large dimension to guarantee smoothness.

The simulation results also shows that the regular multivariate kernel density estimate performs better when the dimension of the data is less than *4* but when the dimension is *4* and above, the multivariate kernel density estimate based on the sample means outperforms the regular multivariate kernel density estimate. In addition, the bandwidth selections using Unbiased Cross-Validation and Biased Cross Validation as well as the Plug-in Technique indicate that the multivariate kernel density estimate based on the sample means in general smoothens the function better than the regular multivariate kernel density estimate.

Based on the fact that a good density estimator might yield a good distribution estimator, it is natural to expect the optimal bandwidth for the multivariate kernel distribution function not to depend on the dimension *p* and this is what happened in our case but not in the case of the regular multivariate kernel distribution function. The order of our optimal bandwidth for the multivariate kernel distribution function based on the sample means is $n^{-1/2}$ for all dimensions. This optimum bandwidth decreases as sample size increases to allow for a better performance of the estimate. The optimum bandwidth for the regular multivariate kernel distribution function proposed by Scott (1992) is $n^{-1/(4+p)}$. Jin and Shao (1999) established that the optimal bandwidth for a class of kernel estimator of a multivariate kernel distribution function is of order $n^{-1/3}$ for all dimensions. Both the optimal bandwidth in our case and that of Jin and Shao (1999) are independent of the dimension of the data. But for a given sample size, our estimate will produce a bandwidth small enough for large dimension to guarantee smoothness.

In the multivariate kernel nonparametric regression estimation, we obtained the Asymptotic Mean Square Error (AMSE) similar to that of the Nadaraya-Watson regression estimation in its multivariate form. In our special case, the optimum bandwidth is of order $n^{-1/2}$ which is independent of the dimension of the data set. This optimum bandwidth decreases as sample size increases to allow for a better performance of the estimate. This breaks the "curse of dimensionality" and remedies the deficiency of high dimensional bandwidth selection as the optimum bandwidth gets smaller enough for large dimension to guarantee smoothness.

We also proposed a new multivariate kernel density estimate which does not completely breaks the "curse of dimensionality" but the effect of the curse on it is minimal as compared to the regular multivariate kernel density estimate. The order of the optimal bandwidth for the new multivariate kernel density estimate is $n^{-6/(4+p)}$. This order optimum bandwidth decreases as sample size increases to allow for a better performance of the estimate. The order optimal bandwidth for the regular multivariate kernel density estimate is $n^{-1/(4+p)}$. They both depend on the dimension of the data set and hence both experience the “curse of dimensionality”. The new multivariate kernel density estimate gives a bandwidth small enough for large dimensions which will guarantee smoothness. This is not the case for the optimum bandwidth given by the regular multivariate kernel density estimate. Therefore, even though our new multivariate kernel density estimate and the regular multivariate kernel density estimate both experience the “curse of dimensionality”, the effect is minimal in our new multivariate kernel density estimate case.

CHAPTER VII

FUTURE WORK

In the dissertation, we were able to investigate the bandwidth selection of both the univariate and the multivariate kernel density estimates based on the sample means. We were also able to compare them to their regular univariate and multivariate kernel density estimates counterparts. We would also like to study the bandwidth selection of the multivariate kernel distribution function and the multivariate kernel regression function based on sample means. We would do these investigations via both biased and unbiased cross validations as well as the plug-in technique. We will then compare these to the bandwidth selection of the regular multivariate kernel distribution function and the regular multivariate kernel regression function and take notice of the differences between them. We would then do some simulations to authenticate these differences.

We are also interested in how the different error criteria play a role in the choice of the bandwidth. We used the mean integrated square error $(MISE = E\int[\hat{f}(x) - f(x)]^2 dx)$ as the error criterion to assess the performance of the estimate and to find the optimum bandwidth in our estimate based on the sample means. This error criterion had been used by many but research shows that using the MISE to obtain the optimum bandwidth slightly under smooth the function. The mean supremum is another error criterion that we would like to use to assess the performance of our estimate. The mean supremum $(MSup = E\{\sup_x |\hat{f}(x) - f(x)|\})$ gives a precise upper bound that can also be used to obtain the optimum bandwidth. But research also shows that using *MSup* to obtain the optimum bandwidth slightly over smooth the function. Therefore, in the future, we will find the average of these two optimum bandwidths as the combined optimum bandwidth choice for the estimate which we believe will do a better job in the smoothing of the function than its individual components.

In the future, we would like to use the estimate of the multivariate kernel density function based on the sample means that we obtained to find the non-parametric estimates of the quantiles (percentiles) in the univariate case and then do a direct bandwidth selection on them. The bandwidth selection will be done via biased and unbiased cross-validation as well as plug-in technique. We also like to show their properties, and then generalize it to the multivariate settings.

We are also interested in doing more research on the new multivariate kernel density estimate. We would like to perform a bandwidth selection on this estimate find its asymptotic properties and then compare to that of the regular multivariate kernel density estimate. We will investigate more into why the issue of "the curse of dimensionality" is less intense in the new estimate as compared to the regular multivariate kernel density estimate. We will then compare them by simulations to show how they both handle the issue of "the curse of dimensionality."

The new estimate is designed to work better for positive random variables or random vectors. This major area has a lot of applications on its own. We will pursue how this new estimate works for an aspect of reliability and life testing. We would like to know how the estimate works in relation to hazard rate, multivariate hazard rate, mean residual life and multivariate mean residual life. We would also like to investigate its performance within some life testing properties like New Better than Used (NBU) and Increasing Hazard Rate (IHR).

REFERENCES

(1) Ahmad, I.A. (1995), "On multivariate kernel estimation for samples from weighted distribution." Statistics and Probability Letters, pp. 121-129.

(2) Ahmad, I.A. and Amezziane, M. (2012), "Efficient Bandwidth Selection in Kernel Multivariate Density Estimation." Paper Under Revision.

(3) Ahmad, I.A. and Lin, P. (1976), "Nonparametric Sequential Estimation of a Multiple Regression Function." Bulletin of Mathematical Statistics, **17**, pp. 63-75.

(4) Billingsley, P. (1985), "Probability and measure." Second Edition, John Wiley & Sons (New York; Chichester; Brisbane; Toronto; Singapore).

(5) Bowman, A.W. (1984), "An Alternative Method of Cross-Validation for the smoothing of Density Estimates." Biometrika **71**, 2, pp. 353-60.

(6) Brewer, M.J. (2000), "A Bayesian Model for Local Smoothing in Kernel Density Estimation." Statistics and Computing, **10**, pp. 299-309.

(7) Dempster, A.P. (1969), "Elements of Continuous Multivariate Analysis." Addison-Wesley, Reading, Massachusetts, London, Ontario.

(8) Doung, T. and Hazelton, M.L. (2003), "Plug-in Bandwidth Matrices for Bivariate Kernel Density Estimation." Nonparametric Statistics, **51**, 1, pp. 17-30.

(9) Doung, T. and Hazelton, M. L. (2005), "Cross-Validation Bandwidth Matrices for Multivariate Kernel Density Estimation." The Scandinavian Journal of Statistics, **32**, 2, pp. 485-506.

(10) Epanechnikov, V.A. (1969). "Nonparametric estimation of a multidimensional probability density." Theory of Probability and Its Applications, **14**, pp. 153-58.

(11) Ferguson, T.S. (1967), "Mathematical Statistics-A Decision Theoretic Approach." Academic Press, New York and London.

(12) Folland, G.B. (1999), "Real Analysis-Modern Technique and Their Applications." Second Edition, (New York; Chichester; Brisbane; Toronto; Singapore).

(13) Graybill, F.A. (1983), "Matrices with Applications in Statistics." Second Edition, Duxbury Classic Series, USA.

(14) Hall, Peter and Marron, J.S (1987), "Estimation of Integrated Squared Density Derivatives." Statistics and Probability Letters **6,** pp**.** 109-115.

(15) Hall, P. and Marron, J.S. (1990), "On variance estimation in nonparametric regression." Biometrika, **77**, pp. 415-419.

(16) Härdle, W. (1989), "Applied nonparametric regression." Cambridge University Press, Cambridge (New York; New Rochelle; Melbourne; Sydney).

(17) Jin, Z. and Shao, Y. (1999), "On kernel estimation of a multivariate distribution function." Statistics and Probability Letter **41** pp. 163-168.

(18) Johnson, L.N. and Kotz, S. (1972), "Distributions in Statistics: Continuous Multivariate Distributions." John Wiley & Sons, Inc., New York, London, Sydney, Toronto.

(19) Jones, M.C. (1991), "Kernel density estimation for length biased data." Biometrika, **78**, 3, pp. 511-519.

(20) Jones, M.C, Marron, J.S. and Sheather, S.J. (1996), "A Brief Survey of Bandwidth Selection for Density Estimation." Journal of the American Statistical Association, **91**, 433, pp. 401-407.

(21) Klemela, J. (2009), "Smoothing of Data-Density Estimation and Visualization." John Wiley & Sons, Inc.

(22) Mack, Y.P. and Müller, H. (1989), "Derivative estimation in nonparametric regression with random predictor variable." Sankhya, The Indian Journal of Statistics, Series A, **51**, 1, pp. 59-72.

(23) Marron, J.S. and Wand, M.P. (1992), "Exact Mean Integrated Squared Error." The Annals of Statistics, **20**, 2, pp. 712-739.

(24) Park, B. U. and Marron, J.S. (1990), "Comparison of Data-Driven Bandwidth Selectors." Journal of American Statistical Association, **85**,409, pp. 66-72.

(25) Polansky, A.M. and Baker, E.R. (2000), "Multistage Plug-in Bandwidth Selection for Kernel Distribution Function Estimates." Journal of Statistical Computation and Simulation, **65**, pp. 63-80.

(26) Rao, B.L.S. P. (1983), "Nonparametric Functional Estimation." Academic Press, New York.

(27) Rencher, A.C. (2002), "Method of Multivariate Analysis." John Wiley & Sons, Inc. (New York; Chichester; Brisbane; Toronto; Singapore)

(28) Richardson, G.D., Kazempour, M.K. and Bhattacharyya B.B. (1991), "Length biased density estimation of fibres." Nonparametric Statistics **1**, pp. 127-141.

(29) Rubin, D.B. and Neal, T. (1992), "Affinely Invariant Matching Methods with Ellipsoidal Symmetric Distributions." The Annals of Statistics, **20**, 2, pp. 1079-1093.

(30) Rubin, D.B. and Stuart, E.A. (2006), "Affinely Invariant Matching Methods with Discriminant Mixtures of Proportional Ellipsoidally Symmetric Distributions." The Annals of Statistics, **34**, 4, pp. 1814-1826.

(31) Rudolf, B. (1979), "Testing for Ellipsoidal Symmetry of a Multivariate Density." The Annals Statistics, **7**, 1, pp. 150-162.

(32) Sain, S.R., Baggerly. A. and Scott, D.W. (1994), "Cross-Validation of Multivariate Densities." Journal of the American Statistical Association, **89**, 427, pp. 807-817.

(33) Schott, J.R. (1997), "Matrix Analysis for Statistics." John Wiley & Sons, Inc. (New York; Chichester; Brisbane; Toronto; Singapore).

(34) Scott, D.W. (1992), "Multivariate Density Estimation-theory, Practice and Visualization." John Wiley & Sons, Inc. (New York; Chichester; Brisbane; Toronto; Singapore).

(35) Scott, D.W. and Terrell, G.R. (1987), "Biased and Unbiased Cross-Validation in Density Estimation." Journal of the American Statistical Association, **82**, 400, pp. 1131-1146.

(36) Scott, D.W. and Wand, M.P. (1991), "Feasibility of multivariate density estimates." Biometrika **78**, pp. 197-205.

(37) Sheather S.J. and Jones M.C. (1991), "A Reliable Data-based Bandwidth Selection Method for Kernel Density Estimation." Journal of Royal Statistical Society, **53**, No. 3, pp. 683-690.

(38) Silverman, B.W. (1986), "Density Estimation for Statistics and Data Analysis." Chapman and Hall London.

(39) Silverman, B.W. and Jones, M.C. (1989), "E. Fix and J.L. Hodges (1951): An Important Contribution to Nonparametric Discriminant Analysis and Density Estimation." International Statistical Review, **57**, 3, pp. 233-247.

(40) Wand, M.P. and Jones M.C. (1993), "Comparison of Smoothing Parameterization in Bivariate Kernel Density Estimation." Journal of the American Statistical Association, **88**, 422, pp. 520-528.

(41) Wand, M.P. and Jones M.C. (1995), “Kernel Smoothing.” Chapman & Hall Ltd (London; New York).

VITA

Jordan Jimmy Crabbe

Candidate for the Degree of

Doctor of Philosophy

Dissertation: HANDLING THE CURSE OF DIMENTIONALITY IN MULTIVARIATE KERNEL DENSITY ESTIMATION

Major Field: Statistics

Biographical:

Education:

Completed the requirements for the Doctor of Philosophy in Statistics at Oklahoma State University, Stillwater, Oklahoma in July, 2013.

Completed the requirements for the Master of Science in Statistics at University of Nevada, Las Vegas, in 2007.

Completed the requirements for the Bachelor of Science in Mathematics at Kwame Nkrumah University of Science and Technology, Kumasi, Ghana in 2001.

Experience:

- Data Analyst, Ghana Telecommunication, 2001-2002
- Instructor and Head of Mathematics and Computer Science Department at Ghana-Lebanon High School, 2002-2003
- Auditor, Ghana Audit Service, 2003-2005
- Instructor of Record at University of Nevada, Las Vegas, Department of Mathematical Sciences, January 2005-July, 2007
  Teaching courses: Intermediate College Algebra, College Algebra, Finite Mathematics, Calculus.
- Instructor of Record at Oklahoma State University, Department of Statistics August, 2007-July, 2013
  Teaching courses:
  - Elementary Statistics
  - Elementary Statistics for Business and Economics
  - Intermediate Statistical Analysis
  - Statistical Method for Social Sciences

Professional Memberships: American Statistical Association, AMS