Smooth Bootstrap Estimation of Some Local and Global Measures of Accuracy of Kernel Density Estimators

A

Thesis

Submitted to

Tezpur University

For the degree of

Doctor of Philosophy

in

Mathematical Sciences (Statistics)

by

Santanu Dutta

Registration Number: 027 of 2009

Department of Mathematical Sciences

Under the School of Science and Technology

Tezpur University

Napaam, Tezpur.

Acknowledgement

The author is extremely grateful to his supervisors Prof. Arup Bose, Professor Stat. Math.Unit, Indian Statistical Institute (ISI) Kolkata and Prof. Munindra Borah, Professor Mathematical Science dept., Tezpur University. The author is indebted to his supervisors for their kind guidance, support and suggestions during various stages of the research.

The author is grateful to the Vice-Chancellor, Registrar and Prof. Munindra Borah, Head of Mathematical Science department (during 2007), of Tezpur University for granting one year study leave for undertaking research in Stat. Math. Unit, ISI Kolata. The author expresses his sincere thanks and gratitude to the Dean of Studies, ISI Kolkata, for granting the position of "Visiting Research Fellow" in Stat. Math. Unit ISI Kolkata, during 2007 August to 2008 July.

The author is thankful to the Head and all the esteemed colleagues of Mathematical Science department, Tezpur University, for their kind support.

The author is thankful to Prof. Arup Bose for supporting several visits to Stat Math. Unit, ISI Kolkata, during the various stages of research. Prof. Bose has been a great teacher, motivator and role model as an ideal academician. The author shall always be grateful to Prof. Bose for his patience and the precious time that he devoted to clarify several queries and rectify a number of mistakes during the course of the research.

The author is grateful to NBHM for awarding NBHM Teacher Fellowship for 2009 and 2010 for undertaking this research.

The author will always be indebted to Prof. Arup Bose, Prof. B.V. Rao, Prof. A. Goswami, Prof. P. Chaudhuri, Prof. K. Maulik and Prof. G. K. Basak for allowing him to attend their lectures in ISI Kokata during August 2007 to July 2008. The author is also grateful for having the opportunity to interact with Prof. S. C. Bagchi and Dr. A. K. Ghosh.

The author expresses his thanks and gratitude Prof. R. Cao and Prof. J.S. Marron for sending their papers without which this thesis could not be completed. The author takes this opportunity to thank all research scholars, especially Rajat, Koushik and Subhajit in Stat. Math. Unit ISI Kolkata for their kind help and affection. The author thanks Mr. Dipankar Das for lending his computer, which was used extensively during this research.

The author is grateful to his parents and wife for their continued support and inspiration. Love and affection of the family members have kept the author going during several difficult phases. The contribution of the author's parents and wife is too great to be acknowledged in a few words, and this thesis is dedicated to them.

Santanu Dutta
11/01/2011

Abstract

In this thesis we address two important problems in the context of kernel density estimation, namely estimating some measures of accuracy and automatic bandwidth selection. Accuracy of any density estimator can be measured point-wise or there can be a notion of overall accuracy.

The *mean integrated squared error* (MISE) is a well known measure of overall accuracy of a kernel density estimate. As far as indicators of local performance of a kernel density estimator are concerned, the *bias* and the *mean squared error* (MSE) are the popular measures. An automatic bandwidth selection rule aims to minimize a data based estimate of some accuracy measure, say the MISE. We study the validity and accuracy of the smooth bootstrap estimators of these accuracy measures and the bandwidth minimizing our MISE estimator. The contents in this thesis can be broadly classified into the following four categories.

(i) MISE estimation.

(ii) Estimating some local measures of accuracy.

(iii) Extension of (i) and (ii) to the context of multivariate density estimation.

(iv) MISE optimal bandwidth selection.

The above mentioned problems, (i) to (iv), are discussed in chapters 2 to 5 respectively. Chapter 1 contains a brief introduction to kernel density estimators, followed by motivation and statement of our proposal. Chapter 1 ends with an overview of the results obtained in the subsequent chapters.

To summarize briefly, in chapter 2 and 3 we have obtained the L_1 and L_2 rates of convergence of $\frac{M^*}{M}$, $\frac{B_y^*}{B_y}$ and $\frac{M_y^*}{M_y}$ to 1, where M, B_y and M_y are the MISE, bias and MSE (at y) of a univariate kernel density estimate. B_y^*, M_y^* and M^* are our bootstrap estimators of B_y, M_y and M respectively. We prove some finite sample properties and compare our estimators with a number of well known estimators. In particular we obtain sufficient conditions under which our estimators outperform plug-in estimators asymptotically. A number of well known bootstrap estimators of MSE and MISE are special case of our proposal.

In chapter 4 we introduce multivariate versions of our estimators, defined in chapters 2, 3. We obtain their finite sample and asymptotic properties. We observe that multivariate versions of our estimators exhibit similar finite sample properties as their univariate counterparts. In general it appears that the convergence rates of the univariate versions are special case of the convergence rates of the multivariate versions of our

estimators. But there are more results on various aspects of our univariate estimators than their multivariate extensions. For instance we could obtain only the L_1 rate of convergence for the multivariate version of $\frac{M^*}{M}$ to one. Its L_2 convergence rate seem to depend on the MSE of kernel estimators of integrated squared partial derivatives of a multivariate density. Such results are difficult to obtain.

The asymptotic properties of our univariate estimators hold for any kernel of order $s \geq 2$. But the asymptotic properties of multivariate versions of our estimators have been obtained for second order kernels only. The reason is that if the given kernel is of order s, then most of the accuracy measures, of a multivariate kernel density estimator, are functionals of all the sth order partial derivatives of the underlying density f. Consequently multivariate versions of our bootstrap estimators aim at estimating these functionals of partial derivatives of f. But as kernel order s is increased, the number of sth order partial derivatives of f increase rapidly. So to obtain the asymptotic properties of our estimators, based on multivariate data, we have to estimate functionals of a large number of partial derivatives of f, as s is increased. This makes the theoretical calculations complicated and increases the lengths the proofs. Moreover second order kernels are more intuitively appealing than higher order kernels. So in the context of multivariate density estimation, we restrict to second order kernels.

The univariate kernel density estimator depends on a single bandwidth, whereas a product kernel density estimator, based on d-dimensional data ($d > 1$), depends on d bandwidths $h_1, ..., h_d$. These bandwidths control the amount of smoothing in each coordinate direction. A special case is to have $h_1 = h_2 = ..., h_d = h$. The resulting product kernel density estimator is referred to as the "simple product kernel density estimator". This assumption is not too restrictive (see for instance, Rao (1983), Abraham, Biau and Cadre (2003)).

We propose an automatic bandwidth $\hat{h}^*$ which aims to minimize our MISE estimator M^* for $h \in I$, where I is some compact interval. If M^* equals our MISE estimator defined in chapter 2, $\hat{h}^*$ is an automatic bandwidth for a kernel density estimate based on univariate data. But if M^* is equal to the MISE estimator defined in chapter 4, the corresponding $\hat{h}^*$ is an automatic bandwidth for a simple product kernel density estimate, based on d-dimensional data. In chapter 5 we provide insight into how well $\hat{h}^*$ succeeds in minimizing the MISE M, for $h \in I$. In particular we obtain the L_1 rate of convergence of $\frac{M(\hat{h}^*)}{M(h^*)}$ to one, where h^* is the minimizer of M in I. This rate depends on the data dimension d.

For univariate data, $\hat{h}^*$ compares well with a number of well known automatic bandwidths such as least squared cross validation, smooth validation and plug-in bandwidths. A number of well known smooth bootstrap bandwidths, such as those proposed by Taylor (1989), Faraway and Jhun (1990), Cao (1993) and Jones, Marron and Park (1991), are different versions of our $\hat{h}^*$.

Sain, Baggers and Scott (1994) proposed automatic bandwidth selection rules for multivariate product kernel density estimates. Their cross validation and bootstrap bandwidths are extensions of the well known unbiased cross validation bandwidth h_c and the Taylor's (1989) bootstrap bandwidth for univariate kernel density estimates. Interestingly the multivariate version of h_c (by Sain, Baggers and Scott (1994)) exhibits stronger asymptotic property than its univariate version, as the data dimension is increases. In a simulation study we compare different versions of our $\hat{h}^*$, based on multivariate data, with the cross validation bandwidth for a number of underlying bivariate densities. The bootstrap bandwidth selection rule (by Sain, Baggers and Scott (1994)) for a product kernel density estimator is a special case of $\hat{h}^*$, for $h_1 = h_2 = ... = h_d$.

FORM-XVIA

TEZPUR UNIVERSITY
Format for certificate of the Principal-Supervisor

This is to verify that the thesis titled ``Smooth bootstrap estimation of some local and global measures of accuracy of kernel density estimators" submitted to the Tezpur University in the Department of Mathematical Sciences under the School of Science and Technology in partial fulfillment for the award of the degree of Doctor of Philosophy in Statistics is a record of research work carried out by Mr. Santanu Dutta under my supervision and guidance.

All helps received by him from various sources have been duly acknowledged. No part of the thesis has been reproduced elsewhere for award of any other degree.

Date: 11/1/2011

Place: Tezpur

Signature of the principal supervisor

Designation: Professor.

School: Science and Technology.

Department: Mathematical Sciences.

Professor

Department of Mathematical Sciences

TEZPUR UNIVERSITY

Certificate of the Supervisor

This is to verify that the thesis titled "Smooth bootstrap estimation of some local and global measures of accuracy of kernel density estimators" submitted to the Tezpur University in the Department of Mathematical Sciences under the School of Science and Technology in partial fulfillment for the award of the degree of Doctor of Philosophy in Statistics is a record of research work carried out by Mr. Santanu Dutta under my supervision and guidance.

All helps received by him from various sources have been duly acknowledged. Some necessary corrections/modifications in the light of the comments of the examiners in the thesis have been incorporated.

No part of the thesis has been reproduced elsewhere for award of any other degree.

Date: 26-12-2010

Place: Kolkata

Signature of the supervisor.

Professor, Stat. Math. Unit, INDIAN STATISTICAL INSTITUTE, 203, Barrackpore Trunk Road, Kolkata-700 108

Designation: Professor.

Department: INDIAN STATISTICAL INST.

Institute: STAT-MATH UNIT

Table of Contents

Chapter 1

Introduction

The field of nonparametrics has broadened its appeal in recent years with an array of new tools for statistical analysis These tools offer sophisticated alternatives for traditional models for exploring data without making specific assumptions. As one of these tools, nonparametric density estimation has become a prominent statistical research topic. A density estimate gives the data analyst a graphical overview of the shape of the distribution. Among the several nonparametric density estimation techniques available in the literature, the kernel density estimation is a very useful tool for exploring the distribution structure of unknown populations. Park and Marron (1990) provides an example to exhibit how this method can show structure that can be very difficult to see by classical methods.

Let $X_1, X_2, ..., X_n$ be n i.i.d R^d valued ($d \geq 1$) random variables with joint density f, where $f(\vec{x}) \geq 0$, $\int f(\vec{x})d\vec{x} = 1$ and $\vec{x} = (x_1, x_2, .., x_d)$, the general problem is to estimate f when no formal structure is specified. "Smoothness" conditions are usually imposed on f and its derivatives, although there are applications in which these smoothness assumptions may not be valid. A *product kernel density estimator* with bandwidths $(h_1, ..., h_d) \equiv (h_{1n}, ..., h_{dn})$ and kernel K is defined as

$$K_n(\vec{y}) = \frac{1}{n \prod_{j=1}^{d} h_j} \sum_{i=1}^{n} \prod_{j=1}^{d} K\left(\frac{y_j - X_{ij}}{h_j}\right).$$

where $\vec{y} = (y_1, y_2, ..., y_d) \in R^d$, $h \to 0$ and $n \prod_{j=1}^{d} h_j \to \infty$ as $n \to \infty$.
For univariate data, i.e. when $d = 1$, $K_n(y)$ is simply referred to as the *kernel density*

estimator and it has been widely studied in the literature (see Rao(1983), Silverman (1986), Park and Marron (1990), Scott (1992), Simonoff (1996) and the references therein).

For a density estimator $K_n(\cdot)$, we may be interested in measuring the accuracy of $K_n(\cdot)$ point-wise or we may be interested in the overall accuracy of $K_n(\cdot)$ instead of its accuracy at any particular point. Accordingly there are two notions of accuracy namely "local" and "global" measures of accuracy. *Bias*, *variance* and *mean squared error* (MSE) are popular local measures of accuracy and *mean integrated squared error* (MISE) is an important global measure of accuracy of a density estimator. The performance of a kernel density estimate crucially depends on the bandwidth h and a popular criterion for selecting the bandwidth h is to minimize the MISE or MSE. See for instance, Taylor (1989), Jhun and Faraway (1990), Hall (1990), Hall, Marron and Park (1992) and Falk (1992).

For a product kernel density estimate there are d bandwidths $h_1, ..., h_d$ and the corresponding MISE or MSE are functions of these d bandwidths. A common assumption is that $h_1 = ... = h_d = h$. This simplifies the problem of bandwidth selection for a product kernel density estimate to a great extent, as under this assumption the MISE or MSE are functions of one variable h (see for instance Sain, Baggerly and Scott (1994) and references therein). However it is not possible to evaluate MSE or MISE since they depend on the underlying density f which is unknown.

1.1 Statement of the problem

For any kernel density estimate with kernel K and bandwidth h, the measures of local or global precision (such as the bias, MSE, MISE etc.) are functionals of h, K and f. Let us denote such a functional by $\theta \equiv \theta(f, h, K)$. Often a popular criterion for choosing h is to minimize an appropriate θ, with respect to h. For instance, θ can be the MISE which is a well known criterion for bandwidth selection. But in general f is unknown. So we address the twin problems of estimating θ and h^*, where the latter denotes the value of h that minimizes θ. Let us describe the bootstrap method briefly in this context.

The bootstrap estimator of a functional θ can be defined as $\hat{\theta} \equiv \theta(f_n, h, K)$, where f_n is some data based estimate of the density. A popular choice of f_n is a kernel density estimator defined using some kernel K^0 (call it "pilot" kernel) and some bandwidth λ (call it "pilot bandwidth"). The resulting estimator $\hat{\theta}$ is called the smooth bootstrap estimator of θ (see Shao and Tu (1995)). A smooth bootstrap estimate (call it $\hat{h}^*$) of h^* is defined as the minimizer of $\hat{\theta}$ with respect to h.

There are other versions of the bootstrap estimator as well. For instance, often θ can also be considered as a functional of the underlying distribution function F. A classical bootstrap estimate of θ is obtained by replacing F, in θ, by the empirical distribution function. But the classical bootstrap fails to estimate the bias and hence the mean square error of kernel based estimators (see e.g. Hall (1990, 1992)). Smooth bootstrap is a natural choice for estimating bias, mean square error etc. for kernel based estimators.

While defining a smooth bootstrap estimator $\hat{\theta}$ there is an obvious question, namely "what are appropriate choices for K^0 and λ in f_n?"

Falk (1992) has used $K^0 = K$ for estimating MSE of a kernel density estimator by smooth bootstrap. Cao (1993), Cao et al.(1994) proposed smooth bootstrap estimators of MISE and resulting bootstrap bandwidth selection rules, where $K^0 = K$ and λ is some bandwidth chosen freely, independent of h. Taylor (1989) obtained a bootstrap estimator of MISE, using $K^0 = K$ and $\lambda = h$. Under a number of smoothness conditions on K and f, Jones, Marron and Park (1991) proved that for $\lambda = Cn^{-23/45}h^{-2}$, $\sqrt{n}(\frac{\hat{h}^*}{h^*} - 1)$ is asymptotically normal. Cao (1993), Cao et al. (1994) has proposed some choices for λ as well, especially λ equal to some constant multiple of $n^{-1/7}$. Faraway and Jhun (1990) proposed to use λ equal to some data based automatic bandwidth, such as the cross validation bandwidth, for estimation of MISE and its application in optimal bandwidth selection. Hall (1992) has proposed two smooth bootstrap methods for estimating bias of a kernel density estimator. In the first method Hall (1992) has used $K^0 = K$ and λ to be larger than h, whereas in the latter method λ is another unspecified bandwidth.

It seems that there are a number of proposals for choosing λ. All the above mentioned smooth bootstrap estimators of the bias, MSE or MISE seem to use $K^0 = K$ or λ equal to some function of h. It is natural to question "is there any advantage in choosing both K^0 and λ independent of K and h?" Besides some of the theoretical studies impose a number of smoothness assumptions on K. For instance see assumptions K2 in Cao (1993) and A.2 in Jones, Marron and Park (1991). An important question is that "can we obtain theoretical properties of θ and $\hat{h}^*$ with minimum possible assumptions on K and h?"

Cao-Abad (1990) pointed out that if $\lambda = h$, the resulting MISE estimator (call it T) is not a suitable estimator of MISE, as $T \to 0$ as h is increased. The choice $\lambda = Cn^p h^m$, where $C,\ p,\ m > 0$, (by Jones, Marron and Park (1991)) presents a similar drawback. In fact, in chapter 2, we prove that if λ is an increasing function of h, the resulting MISE estimator goes to zero as h is increased. In general, we believe that the choice $K^0 = K$ or λ equal to any function of h restricts the theoretical framework in the sense that any assumption imposed on K^0 and λ must hold for K or h, as well.

In order to point out some demerits of the proposal $K^0 = K$, let consider a few well known kernels.

$$
\begin{aligned}
(i)\ (\textit{Rectangular kernel})\ \ K(x) &= \frac{1}{2},\ -1 \le x \le 1 \\
&= 0, \quad \text{otherwise.} \\
(ii)\ \ (\textit{Triangular kernel})\ \ K(x) &= 1 - |x|,\ -1 \le x \le 1 \\
&= 0, \quad \text{otherwise.} \\
(iii)\ (\textit{Quadratic kernels})\ (a)\ \ K(x) &= \frac{3}{4\gamma^3}(\gamma^2 - x^2),\ \text{ if }\ x^2 < \gamma^2 \\
&= 0,\ \text{ if }\ x^2 > \gamma^2,\ \ \gamma > 0. \\
(b)\ \ K(x) &= \frac{9}{8}(1 - \frac{5}{3}x^2),\ \text{ if }\ -1 \le x \le 1 \\
&= 0, \quad \text{otherwise.}
\end{aligned}
$$

The kernel K in (iii) (b), is a "higher order kernel" with zero second moment. If $K^0 = K$, $f_n(y) = \frac{1}{n\lambda}\sum_{i=1}^{n} K\left(\frac{y-X_i}{\lambda}\right)$ and if K is any one of the above kernels, then f_n exhibits the following properties.

(i) If K is a *rectangular kernel*, then

$$f_n(y) = \frac{1}{2n\lambda}\sum_{i=1}^{n} I_{[X_i-\lambda, X_i+\lambda]}(y),$$

where $I_{[X_i-\lambda,X_i+\lambda]}(y) = 1$, if $y \in [X_i - \lambda, X_i + \lambda]$ and zero otherwise. Clearly f_n is not continuous at $y = X_i \pm \lambda,\ i = 1, 2, ..., n$.

(ii) If K is a *triangular kernel*, then

$$f_n(y) = \frac{1}{\lambda} - \frac{1}{n\lambda^2}\sum_{i=1}^{n}|y - X_i|.$$

f_n is not differentiable at $y = X_i$, $i = 1, 2, .., n$ and for other values of y, the second and higher order derivatives vanish.

(iii) For *quadratic kernels*, the second order derivatives of f_n is constant multiple of $\frac{1}{\lambda^3}$ and higher order derivatives vanish.

If f_n is not differentiable or derivatives of f_n vanish, then it is extremely difficult to prove the consistency of $\hat{\theta}$. Let us explain this problem in the context of MISE estimation.

MISE estimation: MISE equals $\int \left[V_y + B_y^2\right] dy$, where B_y, V_y are the point wise bias and variance of the density estimate. A smooth bootstrap estimate of the MISE equals $\int \left[V_y^* + (B_y^*)^2\right] dy$, where B_y^*, V_y^* are smooth bootstrap estimates of B_y, V_y. We note that $B_y^* = \int K(u)[f_n(y - hu) - f_n(y)]du$.

For a second order kernel it is easy to verify that

$$\int B_y^2 dy = \frac{h^4 \int \left[f^{(2)}(y)\right]^2 dy}{4}\left[\int K(u)u^2 du\right]^2 + o(h^4),$$

where f is assumed to possess two continuous derivatives and its second derivative ($f^{(2)}$) is assumed to be square integrable.

If we use $K^0 = K$ and K is any one of the above mentioned kernels in $(i)-(iii)$, then we cannot obtain a similar representation (where the leading term is a multiple of h^4) for $\int \left[(B_y^*)^2\right] dy$. This problem arises from the fact that, if K^0 is any one of the above mentioned kernels in $(i) - (iii)$, then we cannot expand the term $f_n(y - hu) - f_n(y)$ as a second or higher order polynomial in h. So for these kernels we cannot prove whether $\int B_y^2 dy$ is consistently estimated by $\int \left[(B_y^*)^2\right] dy$. Consequently, the proof of consistency of the bootstrap estimator of MISE appears to be difficult, for $K^0 = K$ in f_n.

In some of the existing theoretical studies on the validity of bootstrap estimators, authors have bypassed the above mentioned problems by imposing smoothness conditions on K. For instance Cao (1993) has assumed that K is six times differentiable and the derivatives of K are integrable (see his condition (K2), page 141). Reviewing his theorems and their proofs it is obvious that, none the first four derivatives of K

should be equal to zero (For instance see Theorem 1 and its proof, Cao (1993)). But such conditions will preclude all the above mentioned kernels.

We propose to solve the above mentioned problems by taking K^0 to be a smooth kernel, for example K^0 equal to a Gaussian kernel. We note that if K^0 equals a standard normal density, f_n is continuously differentiable and hence $f_n(y - hv) - f_n(y)$ can be expanded as a polynomial (of any order) in h by Taylor's expansion. So irrespective of K, $\hat{\theta}$ can admit an asymptotic representation (where the leading term is a second or higher order polynomial in h) similar to that for θ. Therefore if choose K^0 freely, we can study theoretical properties $\hat{\theta}$ for any K.

Our Proposal. We propose to study the validity and accuracy of $\hat{\theta}$ for the broadest class of kernels and bandwidths. We avoid imposing conditions on K and h as far as possible. Therefore the main tenet of our proposal is to choose both K^0 and λ freely and impose conditions on them, rather than on K or h. Our theoretical study lends insight into appropriate choice of λ and K^0.

An important aspect of our work is that, we get new asymptotic properties of some well known bootstrap estimators as a special case of our theoretical results. For instance, we get L_1, L_2 rates of convergence for Cao's (1993) bootstrap MISE estimator as a special case of our Theorem 2.3.1, in chapter 2. Similarly our results in chapter 3 provide insight into new asymptotic properties of the bootstrap bias and MSE estimators by Falk (1992). The details are discussed in the respective chapters.

1.2 Basic assumptions:

Although the assumptions imposed vary from problem to problem, the basic assumptions in all the problems is that are the following

(i) the given data is realization of n independent and identically distributed (i.i.d) random variables $X_1, X_2, ..., X_n$ and the distribution of X_1 is absolutely continuous with density $f(\cdot)$.

(ii) the density f is assumed to satisfy certain smoothness assumptions. There are no moment assumptions on the underlying distribution and f is not assumed to belong to any specific family of densities.

1.3 Summary of the chapters

Each chapter begins with a statement of the problem which is addressed in that chapter. It is followed by a brief literature review on that specific problem and a description of the notation and assumptions used to state and prove the results in that chapter. We collect all the main results in a one section, referred to as "main results". The proofs are given at the end of each chapter. We describe of the materials of the latter chapters in the following paragraphs.

In chapter 2, we propose a smoothed bootstrap estimator M^*, based on one dimensional data, of the MISE (call it M) and obtain the L_1 and L_2 rates of convergence of $\frac{M^*}{M}$ to 1 for a broad class of kernels and bandwidths. We also investigate some finite sample properties of M^*. If the bandwidth h satisfies $\limsup_{n\to\infty} nh^{2s+1} < \infty$, where s is the kernel order, then M^* is shown to be more accurate than asymptotic approximation to M. M^* compares well with a number of other estimators of M.

M^* depends on a kernel K^0 and bandwidth λ. We suggest some appropriate choices for them. For appropriate choices of λ, M^* works well even for small samples. We find that if λ equals an increasing function of h, then resulting M^* is not a suitable estimate of M. A number of bootstrap MISE estimators, such as those proposed by Taylor (1989), Faraway and Jhun (1990), Cao (1993), Cao et al. (1994), are special cases of M^*. Some new asymptotic properties of Cao's (1993) estimator are obtained as a special case of our results.

In chapter 3, we propose a generalized smooth bootstrap scheme for estimating the bias B_y and mean square error M_y of a kernel density estimator, at y, based on i.i.d one dimensional data. For a fairly broad class of kernel and bandwidth h, we obtain the rates at which $E\left[\frac{B_y^*}{B_y} - 1\right]^2$ and $E\left[\frac{M_y^*}{M_y} - 1\right]^2$ converge to zero as n (sample size) is increased, where B_y^* and M_y^* are the proposed estimators of B_y and M_y respectively. The well known smooth bootstrap estimators (by Falk (1992)) of bias and MSE are special of B_y^* and M_y^*. Our proposed estimators compares well with the plug-in estimators as well. For instance our results imply that, our bias estimator B_y^* is asymptotically more accurate (in L_2 sense) than the plug-in estimator of bias when the bias at y is actually large. If $\limsup_{n\to\infty} n^{2/5}h = \infty$ and $f^{(1)}(y) = 0$, $f^{(2)}(y) \neq 0$ then our bootstrap estimator of variance is asymptotically more accurate (in L_2 sense) than the corresponding the plug-in estimator, where $f^{(1)}$, $f^{(2)}$ are 1st and 2nd derivatives of the density. Simulations reveal that if y is a point in the tail region, MSE , at y, can be minimized for more than one value of the bandwidth h. Our bootstrap estimator M_y^* also exhibits this feature, but the classical plug-in estimate is always uniquely minimized.

In chapter 4, we extend the results of chapter 2 and 3 to the case of a multivariate product kernel density estimator. For instance we have obtained the rates at which $E\left|\frac{M^*}{M} - 1\right|$, $E\left[\frac{M_{\vec{y}}^*}{M_{\vec{y}}} - 1\right]^2$ and $E\left[\frac{B_{\vec{y}}^*}{B_{\vec{y}}} - 1\right]^2$ converge to zero, where M, $M_{\vec{y}}$ and $B_{\vec{y}}$ are MISE, mean-squared error and bias (at $\vec{y} = (y_1, y_2, ..., y_d)$) of a product kernel density estimator. $B_{\vec{y}}^*$, $M_{\vec{y}}^*$ and M^* are the proposed estimators of $B_{\vec{y}}$, $M_{\vec{y}}$ and M respectively. We have used different pilot bandwidths for our bias and variance estimators, say λ and μ. So our MSE and MISE estimators depend on two pilot bandwidths, namely λ, μ. The choice of μ is straightforward. However there seem to be more than one choice of λ and different versions of our estimators correspond to the different choices of λ.

Scott and Wand (1991) introduced a local measure of accuracy which is referred to as the *sample root coefficient of variation* (denoted by $R_{\vec{y}}$). We have obtained the rate of convergence of $\frac{R_{\vec{y}}^*}{R_{\vec{y}}}$ to one, in probability and asymptotic distribution of $\sqrt{n\lambda^d}\left(\frac{R_{\vec{y}}^*}{R_{\vec{y}}} - 1\right)$, where d is the data dimension. $R_{\vec{y}}^*$ is the proposed smooth bootstrap estimator of $R_{\vec{y}}$.

Sain, Baggerly and Scott (1994) proposed multivariate version of the Taylor's MISE estimator (call it T), which is a special case of our M^*. We have obtained some finite sample properties of our estimators based on multivariate data. It is interesting to note that the multivariate versions of our MSE and MISE estimators exhibit similar finite sample properties as our bootstrap estimators based on univariate data. For instance, we prove that $T \to 0$ as $h_i \to \infty$, where h_i, $i = 1, 2, .., d$, are the bandwidths used in the product kernel density estimate. So T is not an appropriate MISE estimator. If λ is not equal to h_i, $i = 1, 2, .., d$, M^* successfully imitates M, for fixed n and large values of h_i, $i = 1, 2, .., d$. If $\vec{y}$ is away from the peaks of f, both $M_{\vec{y}}^*$ and $M_{\vec{y}}$ appear to be minimized for more than one value of h.

In chapter 5, we address the problem of bandwidth selection for univariate kernel and multivariate product kernel density estimates based on i.i.d data. For multivariate product kernel estimators, we assume that all the bandwidths are equal and h is the common bandwidth which controls the amount of smoothing in any direction. We propose an automatic bandwidth $\hat{h}^*$ which is the minimizer of M^*, where M^* denotes our estimator of the MISE of a univariate kernel or multivariate product kernel density estimator. We provide insight into how well $\hat{h}^*$ succeeds in minimizing M for h in $I = \left[\frac{\epsilon_1}{n^{1/(2s+d)}}, \frac{\epsilon_2}{n^{1/(2s+d)}}\right]$, where $\epsilon_1, \epsilon_2 > 0$, s is the kernel order and d is data dimension (in univariate case $d = 1$). For $d > 1$, we restrict to $s = 2$. There a number of different versions of $\hat{h}^*$, depending on the choice of pilot bandwidth λ in M^*. Using simulation, we compare the proposed bandwidth selection rules with a number of well known univariate and multivariate automatic bandwidth selection rules. For fixed sample size and Gaussian kernel, no specific automatic bandwidth selection rule appears to uniformly outperform its peers, in terms of minimizing M. In general for second order kernels and $\lambda = \frac{1}{n^{1/9}}$, the univariate version of $\hat{h}^*$ exhibits lower variance than the cross validation and Sheather-Jones's (1991) plug-in bandwidths.

For second order kernels with finite support and $\lambda = \frac{1}{n^{1/9}}$, the univariate version of $\hat{h}^*$ appears to be asymptotically more accurate, in terms of minimizing M in I, than unbiased cross validation, biased cross validation and Park and Marron's (1990) plug-in bandwidths. This result holds for Cao's (1993) bootstrap bandwidth as well. Smooth bootstrap bandwidths proposed by Taylor (1989), Faraway and Jhun (1990), Cao (1993) and Jones, Marron and Park (1991) are different versions of $\hat{h}^*$, for different choices of K^0 and λ in M^*. The bootstrap bandwidth, proposed by Sain, Baggerly and Scott (1994), is special of our multivariate version of h^*.

Chapter 2

Smooth Bootstrap estimate of Mean Integrated Squared Error

2.1 Introduction

Given $X_1, X_2, ..., X_n$ i.i.d. random variables with density $f(\cdot)$, the *kernel density estimator* (of f) based on the kernel $K(\cdot)$ and bandwidth $h \equiv h_n$ is defined as

$$K_n(y) = \frac{1}{nh} \sum_{i=1}^{n} K\Big(\frac{y - X_i}{h}\Big)$$

where $h \to 0$ and $nh \to \infty$ as $n \to \infty$. The *mean integrated squared error (MISE)* of the kernel density estimator $K_n(\cdot)$ is defined as

$$\begin{aligned}
M \equiv M(K,h) &= \int_{-\infty}^{\infty} E\big[K_n(y) - f(y)\big]^2 dy \\
&= \int V\left[K_n(y)\right] dy + \int [B_n(y)]^2 dy, \text{ where} \qquad (2.1.1) \\
\int V\left[K_n(y)\right] dy &= \frac{1}{nh}\int K^2(v)dv - \frac{1}{n}\int \Big\{ \int K(v) f(y - hv) dv \Big\}^2 dy, \\
\text{and } \int [B_n(y)]^2 dy &= \int \{E(K_n(y) - f(y)\}^2 dy \\
&= \int \Big[\int K(u) f(y - hu) du - f(y) \Big]^2 dy.
\end{aligned}$$

$V\left[K_n(y)\right]$ and $B_n(y)$ are the variance and bias of $K_n(y)$. M is a global measure of accuracy of $K_n(\cdot)$. It has enjoyed great popularity, especially in the context of optimal bandwidth selection of a kernel estimator. See for instance, Taylor(1989), Jhun and

Faraway (1990) and Hall, Marron and Park (1992). It is easy to see that

$$M = E[ISE_n], \quad \text{where } ISE_n = \int_{-\infty}^{\infty} [K_n(y) - f(y)]^2 dy.$$

ISE_n is referred to as *integrated squared error* and it is also a well known measure of discrepancy between K_n and f. Jones (1991) argued that if the goal of data analysis is to estimate f well from every sample, then ISE_n is conceptually more appropriate than M for assessing the performance of density estimates. However Jones (1991) further argued that hoping to be able to estimate f well from every sample (generated by f) is an unrealistic goal and that one can only expect to do well in some average sense. This leads to the conclusion that M serves as a more practical criterion for comparing density estimates.

Note that M is a functional (call it $M(f)$) of the underlying density f. Since f is unknown, M needs to be estimated. A smooth bootstrap MISE estimator, call it M^*, is defined as $M^* = M(K_n^0)$, where K_n^0 is a kernel density estimate with some kernel K^0 and bandwidth λ. In this chapter we obtain some asymptotic and finite sample properties of M^*, where K^0, λ are chosen freely and do not depend on K or h. Different choices of K^0 and λ yield different versions of M^*. Let us review some well known bootstrap estimators of MISE.

2.1.1 A brief literature review

Taylor (1989) defined a smooth bootstrap estimate (we call it T) of M, using K^0 and λ equal to Gaussian kernel and h respectively. He obtained its exact formula and asymptotic variance, see his equation (6) on page 707 (loc. cit.).

However Cao-Abad (1990) argued that T is not a suitable estimator of M, since $T \to 0$, whereas M converges to a positive constant, as h is increased. Jones, Marron and Park (1991) proposed another version of M^*, using $K^0 = K$ and $\lambda = Cn^p h^m$, where C, p, m are constants. Cao (1994) pointed out that the choice $\lambda = Cn^p h^m$, where $m > 0$, presents drawback similar to the choice $\lambda = h$. Cao's (1993) smooth bootstrap based MISE estimator (we call it M_{Cao}) is yet another version of M^*, where $K^0 = K$, K is a second order kernel with six derivatives and λ is independent of h. Bootstrap based estimation of M and its use in bandwidth selection have been proposed by several other authors as well, such as Jhun and Faraway (1990), Hall (1990) and

Hall, Marron and Park (1992).

Hall (1990) proposed a bootstrap scheme, where the size of the bootstrap resample is less than the size of the original sample, and proved the theoretical validity of the bootstrap approximation (see his equation (2.10), page 184). More recently Delaigle and Gijbels (2004, Theorem 4.1, page 27) have proved the validity of the smooth bootstrap approximation to M for deconvolution kernel density estimators based on data contaminated by random noise. Their results also imply the validity of smooth bootstrap approximation to M for the ordinary kernel density estimator, based on error free data, which we consider here. They assumed that the characteristic function, $\phi_K(t) = \int e^{itx}K(x)dx$, has a compact support (see their condition B2, page 29) and K is of order 2. These assumptions are restrictive in the context of ordinary kernel density estimation based on error free i.i.d. observations. Besides Hall (1990) and Delaigle and Gijbels (2004) do not provide insight into how fast the accuracy of their MISE estimators improve, as the sample size is increased.

2.1.2 Our proposal

The main idea of our proposal is that K^0 is some smooth kernel (say a Gaussian kernel) and λ satisfies some conditions, to be specified later. We impose conditions on K^0 and λ to ensure that (a) $K_n^0(\cdot)$ is a smooth function and (b) $\int \left[K_n^{0(s)}(y)\right]^2 dy$ is a consistent estimator of $\int \left[f^{(s)}(y)\right]^2 dy$, where $K_n^{0(s)}$, $f^{(s)}$ are sth derivatives of K_n^0, f respectively and s is the kernel order.

The conditions imposed K^0 and λ need not hold for K or h at all. In chapter 1, we have already mentioned some demerits of the proposal $K^0 = K$, in the context of MISE estimation. In the next paragraph we provide yet another motivation for choosing K^0 and λ freely.

We see that the theoretical properties of a smooth bootstrap estimate M^* depends on the asymptotic properties of $\int \left[K_n^{0(s)}(y)\right]^2 dy$ (see for instance Cao (1993)). The asymptotic properties of $\int \left[K_n^{0(s)}(y)\right]^2 dy$ are obtained under a number of conditions on the kernel K^0 or the bandwidth λ (see for example, the conditions in Cao (1993) and Hall and Marron (1987)). So if we use $K^0 = K$ or λ equal to some function of h, then the assumptions imposed on K^0 or λ must also hold for K or h. This can restrict the class of K or h for which the theoretical properties of M^* hold.

The advantage of introducing the extra parameters K^0 and λ is that we can impose a number of conditions on them, without imposing those conditions on K or h. In this chapter, the conditions on K or h remain quite general. Our results naturally hold for M_{Cao}, as it is a special case of our proposal. So we have also obtained new asymptotic properties of M_{Cao}, with lesser conditions on K than Cao (1993).

2.1.3 Definitions

Let K^0 and $\lambda \equiv \lambda_n$ be another kernel and bandwidth sequence. We define

$$K_n^0(y) = \frac{1}{n\lambda}\sum_{i=1}^{n} K^0\Big(\frac{y - X_i}{\lambda}\Big).$$

Let $X_1^*, X_2^*, \ldots, X_n^*$ be i.i.d. *smooth bootstrap resample* of size n with density $K_n^0(\cdot)$. Then

$$K_{nB}(y) = \frac{1}{nh}\sum_{i=1}^{n} K\Big(\frac{y - X_i^*}{h}\Big)$$

is the *smooth bootstrap* version of $K_n(y)$. Given $X_1, X_2, \ldots, X_n$, the *smooth bootstrap estimator* of M is defined as

$$\begin{aligned}
M^* &= \int_{-\infty}^{\infty} E_n\left[K_{nB}(y) - K_n^0(y)\right]^2 dy \\
&= V^* + B^*(\text{say}), \text{ where} \\
V^* &= \frac{1}{nh}\int K^2(u)du - \frac{1}{n}\int\left[\int K(u)K_n^0(y - hu)du\right]^2 dy \text{ and} \\
B^* &= \int\left[\int K(u)K_n^0(y - hu)du - K_n^0(y)\right]^2 dy.
\end{aligned}$$

The expectation E_n is computed with respect to the density $K_n^0(\cdot)$.

There are several versions of M^*, depending on the choice of λ. For instance, if λ is equal to the cross validation bandwidth, then M^* is the bootstrap estimator proposed by Faraway and Jhun (1990). If $\lambda = h$ and K is the Gaussian kernel, then M^* equals the Taylor's (1989) estimator T.

Next we define a number of other well known estimators of M. Taylor (1989) proposed an estimator which is defined as follows

$$T = \frac{1}{2nh\sqrt{\pi}} + \Big(1 - \frac{1}{n}\Big)\frac{1}{n^2}\sum_{i=1}^{n}\sum_{l=1}^{n}\phi_{4h^2}(X_i - X_l) - \frac{2}{n^2}\sum_{i=1}^{n}\sum_{l=1}^{n}\phi_{3h^2}(X_i - X_l) + \frac{1}{n^2}\sum_{i=1}^{n}\sum_{l=1}^{n}\phi_{2h^2}(X_i - X_l).$$

Another popular estimator of M is the plug-in estimator. The basic idea of the plug-in rule (see Park and Marron (1990); Jones, Marron and Sheather (1996)) is to substitute data based estimates into the asymptotic approximation to M. It is well known that if K is of order s then
$M = A_n(h) + o\left(\frac{1}{nh} + h^{2s}\right)$ where $A_n(h) \equiv A_n$ equals

$$\frac{1}{nh}\int K^2(u)du + \frac{a^2}{(s!)^2}h^{2s}\int [f^{(s)}(x)]^2 dx \quad \text{and} \quad a = \int x^s K(x)dx \neq 0.$$

A_n is referred to as the asymptotic mean integrated squared error. In the plug-in method $\theta_s = \int [f^{(s)}(x)]^2 dx$ is replaced by a suitable estimator (we call it $\hat{\theta}_s$) to obtain the following estimator $\hat{A}_n$:

$$\hat{A}_n = \frac{1}{nh}\int K^2(u)du + \frac{a^2}{(s!)^2}h^{2s}\hat{\theta}_s.$$

There are other popular estimators such as unbiased cross validation (call it UCV_n), biased cross validation (call it BCV_n) and augmented cross validation, which we denote by $AUCV_n$. A detailed discussion on the properties of these estimators can be found in Scott and Terrell (1987) and references therein. UCV_n is a biased estimator of M and it is defined as

$$UCV_n = \int [K_n(y)]^2 \, dy - \frac{2}{n}\sum_{i=1}^{n} K_{-i,n-1}(X_i), \quad \text{where}$$

$K_{-i,n-1}(y) = \frac{1}{(n-1)h}\sum_{j\neq i=1}^{n} K\left(\frac{y-X_j}{h}\right), \; i = 1, 2, .., n$. In fact UCV_n equals

$$\frac{\int K^2(u)du}{nh} + \sum_{i=1}^{n}\sum_{j\neq i}^{n}\left[\frac{1}{n^2h^2}\int K\left(\frac{x-X_i}{h}\right)K\left(\frac{x-X_j}{h}\right)dx - \frac{2}{n(n-1)h}K\left(\frac{X_i-X_j}{h}\right)\right].$$

For a second order kernel K, BCV_n is defined as

$$\begin{aligned} BCV_n &= \frac{1}{nh}\int K^2(u)du + \frac{\left[\int x^2K(x)dx\right]^2}{2n^2h}\sum_{i\neq j=1}^{n}\alpha\left(\frac{X_i-X_j}{h}\right), \quad \text{where} \\ \alpha(y) &= \int K^{(2)}(x)K^{(2)}(x+y)dx, \end{aligned}$$

$K^{(2)}(y)$ is the second derivative $K(y)$. Bowman (1984) gave the following formula for $AUCV_n$, which Hall (1983) argued to be the correct form for theoretical analysis.

$$AUCV_n = \int [K_n(y)]^2 dy - \frac{2}{n}\sum_{i=1}^{n}[K_{-i,n-1}(y) - f(X_i)] - \int f^2(y)dy.$$

2.1.4 Summary of the chapter

There are four theorems and a simulation study in this chapter. In Theorem 2.3.1, we obtain the rates of convergence of the mean squared error of M^* and $\frac{M^*}{M}$. The latter rate provides insight into how fast the accuracy of M^* will improve with increase in sample size. We have also obtained the rates convergence of the mean absolute deviation of M^* and $\frac{M^*}{M}$. The L_1 rates are faster than the square root of the L_2 rates. We have studied how well M^* compares with other estimators. We have obtained conditions under which M^* is asymptotically more accurate than a class of plug-in estimators, see Theorem 2.3.2. An important aspect of our study is to provide insight into some properties of M^* when sample size n is fixed. Theorem 2.3.3 is a result in that direction. We that if λ equals an increasing function in h, the resulting M^* is not an appropriate MISE estimator.

For Gaussian kernel or Gaussian type higher order kernels we have obtained the exact formula of M^*, Theorem 2.4.2. Our theoretical results and simulation provide some useful guidelines for choosing K^0 and λ. For instance we recommend K^0 equal to Gaussian kernel and if K is a second order kernel, then $\frac{1}{n^{1/9}}$ or the least squared cross validation bandwidth are some appropriate values of λ.

2.2 Assumptions

Let us collect below all the assumptions on the two kernels and the bandwidths. Not all of them will be used in all the results. A function H is said to be uniformly bounded if $||H|| = \sup_{-\infty<y<\infty} |H(y)| < \infty$.

Assumption A (on density f).

(i) The density $f(\cdot)$ is uniformly bounded, and for some $s \geq 2$, the sth derivative $f^{(s)}$ is uniformly bounded and square integrable.

(ii) There exists $k > 1$ and M, such that $|f^{(s+k)}(x) - f^{(s+k)}(y)| \leq M|x-y|$, for all x, y.

(iii) There exists $p \geq 1$, such that $(s+p)$ th derivative $f^{(s+p)}(\cdot)$ exists, is integrable and is also square integrable.

Assumption B (on kernel K). The kernel $K(\cdot)$ is square integrable and is of sth order, that is $\int K(x)dx = 1$ and there exists an integer $s \geq 2$ such that $\int K(x)x^j dx =$

0, $j = 1, 2, .., s-1$ and $\int |K(x)x^s| dx < \infty$. Further we assume that $K(-x) = K(x)$ and $\int |K(x)x^{s+1}| dx < \infty$.

Assumption C (on auxiliary kernel K^0).

(i) The auxiliary kernel $K^0(\cdot)$ is a probability density function such that

(a) $K^0(\cdot)$ is continuous and uniformly bounded.

(b) $K^0(x) \to 0$ as $|x| \to \infty$.

(ii) $K^0(\cdot)$ has s continuous derivatives on $(-\infty, \infty)$ and its sth derivative $K^{0(s)}(\cdot)$ satisfies the above conditions (a) and (b) and also the following.

(c) $\int |K^{0(s)}(x)| dx < \infty$.

(d) $\int K^{0(s)}(x)x^j dx = 0$, where $j = 0, 1, 2, ..., s-1, s+1, ..., s+p-1$,
$\frac{(-1)^s}{s!} \int K^{0(s)}(x)x^s dx = 1$ and $\int |K^{0(s+p)}(x)x^{s+p}| dx < \infty$.

Assumption D (on auxiliary bandwidth λ) The sequence $\{\lambda\} \equiv \{\lambda_n\}_{n=1,2,3,}$ satisfies

(i) $\lambda > 0 \ \forall \ n \geq 1$ and $\lambda \to 0$, as $n \to \infty$.

(ii) $n\lambda^{2s+1} \to \infty$ as $n \to \infty$.

Assumption E (on bandwidth h) The sequence $\{h\} \equiv \{h_n\}_{n=1,2,..}$ satisfies $h \to 0$ and $nh \to \infty$ as $n \to \infty$.

Remark 2.2.1. (i) The value of p, in $A(iii)$ depends on K^0. If K^0 is the standard normal density, then we recommend $p = 2$. With this choice of K^0 and p, Assumption C is satisfied for any value of s. Most of our results are obtained assuming $k = p$ and for a second order kernel, we recommend $p = k = 2$.

(ii) Assumptions $A(i) - (iii)$ on f are valid for a wide class of densities which include mixed normal, Cauchy, beta(m,n) (m, n>2) and gamma(n) (n>2) among others. In contrast, the assumption that f has compact support or the assumption $E(|X_1|^\epsilon) < \infty, \epsilon > 0$ (see page 184, Hall (1990)) precludes the mixed normal distributions or the heavy tailed distributions which have no moments.

For a second order kernel K, Cao (1993) obtained asymptotic properties of his MISE estimator assuming that f is six times differentiable, the derivatives are bounded and the 1st four derivatives are integrable. But for a second order kernel and for $p = k = 2$, we require assumptions on first four derivatives and so for a second order kernel, we can obtain the asymptotic properties of M^* imposing fewer smoothness assumptions, on f, than Cao (1993).

(iii) Assumption B on K and Assumption E on h are quite common in density estimation context. On the other hand, the assumptions by Hall (1990) or Delaigle and Gijbels (2004) on K prevent the use of a number of popular kernels such as the Gaussian or Gaussian type kernel, as they are neither compactly supported nor do their characteristic functions vanish outside any compact subset of the real line. Cao (1993) assumed that K is six times differentiable, the first six derivatives are bounded, integrable and satisfy a number of conditions (see condition K2, in page 141 Cao (1993)). We do not impose any smoothness assumption on K at all.

2.3 Main Results

We now state and prove our main results. The proofs are given at the end of the chapter.

Theorem 2.3.1. *Under Assumptions $A - E$ and $k = p$, as $n \to \infty$,*
(i) $E[M^* - M]^2 = o\left(\frac{1}{n^2}\right) + O\left(h^{4s} r_n\right).$
(ii) $E\left[\frac{M^*}{M} - 1\right]^2 = O(r_n),$ *where* $r_n = \frac{1}{n\lambda^{2s+1}} + \lambda^p + h.$
In particular if $\lambda = \frac{1}{n^{1/(2s+2p+1)}}$, *then*
(iii) $E|M^* - M| = o\left(\frac{1}{n}\right) + O\left(\frac{h^{2s}}{n^{p/(2s+2p+1)}}\right).$
(iv) $E\left|\frac{M^*}{M} - 1\right| = o(h) + O\left(\frac{1}{n^{p/(2s+2p+1)}}\right).$

Remark 2.3.1. (i) Theorem 2.3.1(ii),(iv) ensures the validity of M^* and provides an answer to the question: "how large should the sample size be in order that the bootstrap approximation M^* is close to M?".

(ii) The motivation for choosing $\lambda = \frac{1}{n^{1/(2p+2s+1)}}$ is to ensure that the L_1 rates, namely $E|M^* - M|$ and $E\left|\frac{M^*}{M} - 1\right|$ can converge to zero faster than $\sqrt{E(M^* - M)^2}$ and $\sqrt{E\left[\frac{M^*}{M} - 1\right]^2}$ respectively.

(iii) Theorem 2.3.1 (iii) and (iv) are not direct consequences of (i) and (ii). The rate of convergence of $E\left|M^* - M\right|$ and $E\left|\frac{M^*}{M} - 1\right|$ (to zero), in Theorem 2.3.1 (iii) and (iv), can be faster than that of $\sqrt{E(M^* - M)^2}$ and $\sqrt{E\left[\frac{M^*}{M} - 1\right]^2}$ respectively. The detailed arguments are as follows.
We note that for $\lambda = \frac{1}{n^{1/(2p+2s+1)}}$, $r_n = h + \frac{1}{n^{p/(2p+2s+1)}} + o\left(\frac{1}{n^{p/(2p+2s+1)}}\right).$

Therefore for $\lambda = \frac{1}{n^{1/(2p+2s+1)}}$, from Theorem 2.3.1 (i) we see that

$$\sqrt{E(M^* - M)^2} = o\left(\frac{1}{n}\right) + O\left(h^{2s+0.5} + \frac{h^{2s}}{n^{p/2(2p+2s+1)}}\right).$$

Whereas from Theorem 2.3.1 (iii), we see that $E|M^* - M| = o\left(\frac{1}{n}\right) + O\left(\frac{h^{2s}}{n^{p/(2p+2s+1)}}\right)$. Clearly for $s, p \geq 2$, $\frac{h^{2s}}{n^{p/(2p+2s+1)}}$ converges to zero faster than $h^{2s+0.5} + \frac{h^{2s}}{n^{p/2(2p+2s+1)}}$, for any choice of h.

Therefore the rate of convergence of $E\left|M^* - M\right|$ (to zero), in Theorem 2.3.1(iii), can be faster than that of $\sqrt{E(M^* - M)^2}$. Similarly the rate of convergence obtained in Theorem 2.3.1 (iv) can be faster than the square root of the rate obtained in Theorem 2.3.1 (ii).

2.3.1 Comparison with other estimators

We have mentioned several other estimators of MISE. Here we discuss a few of those in relation to our estimator.

Delaigle and Gijbels' (2004) estimator

The bias and variance of the Delaigle and Gijbels' estimator for second order deconvolution kernel density estimator are

$$Bias = O\left(\frac{1}{n}\right) + o\left(h^4\right) \quad \text{and} \quad Var = o\left(\frac{1}{n^2} + h^8\right) + o\left(\frac{h^4}{n}\right) \tag{2.3.1}$$

respectively (see their Theorem 4.1, page 27) which continue to hold good for their smooth bootstrap estimator of M in the usual i.i.d. set up, where the observations are error free. Let the bias and variance of M^* be denoted by B_n and $Var(M^*)$ respectively. Since $|B_n| \leq E\left|M^* - M\right|$ and $Var(M^*) \leq E\left[M^* - M\right]^2$, it immediately follows from Theorem 2.3.1(iii) and (i) that

$$B_n = o\left(\frac{1}{n}\right) + o\left(h^{2s}\right) \quad \text{and} \quad Var(M^*) = o\left(\frac{1}{n^2} + h^{4s}\right) \quad \text{where } s \geq 2. \tag{2.3.2}$$

While the difference of orders, $O(\cdot)$ and $o(\cdot)$ in the rates in (2.3.1) and (2.3.2) at which bias goes to zero may seem insignificant from a theoretical point of view, it may translate into significant gains in finite samples, just as in the usual bootstrap theory of sample mean type statistics. The results of Delaigle and Gijbels' (2004) hold only for K with compactly supported characteristic function (see their Condition B (B2), page 25), whereas we provide the rates of convergence of bias and variance (of M^*) for a wide class of K.

Cao's M_{Cao}

Cao (1993) proposed an estimator M_{Cao} which is a special case of M^* for $K^0 = K$, where K is a second order kernel satisfying a number of smoothness conditions (see his condition $K2$, page 141). Cao (1993) obtained asymptotic representations for M and M_{Cao} (see his Theorem 1, page 142). A direct consequence Cao's (1993) Theorem 1 is that $\frac{M_{Cao}}{M}$ converges to one, *in probability*. Our Theorem 2.3.1 holds for M_{Cao} as well. So we obtain some new asymptotic properties of M_{Cao}, for example the L_1, L_2 rates convergence of $\frac{M_{Cao}}{M}$ to one and also the rate at which $E[M_{Cao} - M]^2$ goes to zero.

We note that rate convergence in probability, of $\frac{M_{Cao}}{M}$ to one, obtained by Cao (1993) can be faster than our corresponding L_1 rate. But comparing our assumptions A, B with the assumptions K2, D (Cao 1993), we see that Cao (1993) obtained his result expense of more smoothness assumptions K and f (see Remark 2.2.1 (ii) and (iii)).

Taylor's T

Taylor's estimator T is a special case of M^* when $\lambda = h$ and K, K^0 are equal to Gaussian kernel. Cao-Abad (1990) observed that $T \to 0$, as h is increased. This result implies that T is not a suitable estimator of M, for large h.

From Taylor (1989) we note that $Var(T) = \frac{C}{n^2 h} + O(\frac{h}{n^2})$, where $C > 0$.

Therefore If $\limsup_{n\to\infty} nh^{2s+0.5} < \infty$, then $Var(M^*)/Var(T) \to 0$.

This result further demonstrates the advantage of choosing λ, satisfying condition D, rather than $\lambda = h$.

Unbiased and biased cross validation estimators UCV_n and BCV_n

These are well known estimators for M. It is known that (see Theorem 3.1, Scott and Terrell (1987)), $Var(UCV_n) = \frac{C}{n} + O(\frac{1}{n^2 h} + \frac{h^4}{n})$,where C is a positive constant.

Therefore

$$\frac{Var(M^*)}{Var(UCV_n)} = \frac{o\left(\frac{1}{n}\right) + O\left(nh^{4s}r_n\right)}{C + O(\frac{1}{nh} + h^4)}. \tag{2.3.3}$$

We note that under Assumptions D and E, $h, r_n = o(1)$ and $nh \to \infty$. Therefore if $\limsup_{n\to\infty} nh^{4s} < \infty$ (which includes the case where h is a multiple of $\frac{1}{n^{1/(2s+1)}}$) then M^* has infinite asymptotic relative efficiency in comparison to UCV_n.

Similarly (see Theorem 3.2, Scott and Terrell (1987))

$$Var(BCV_n) = \frac{C'}{n^2h} + O(\frac{h}{n^2}), \tag{2.3.4}$$

where $C' > 0$. Therefore

$$\frac{Var(M^*)}{Var(BCV_n)} = \frac{o(h) + O\left(n^2h^{4s+1}r_n\right)}{C' + O(h^2)} \tag{2.3.5}$$

and hence if $\limsup_{n\to\infty} nh^{2s+0.5} < \infty$, then M^* has infinite asymptotic relative efficiency in comparison to BCV_n.

Further the bias of UCV_n and BCV_n are respectively (see Theorem 3.1 and 3.2, page 1134 Scott and Terrell (1987))

$$Bias(UCV_n) = -\int f^2(x)dx \quad \text{and} \quad Bias(BCV_n) = O(\frac{1}{n} + h^{2s+1}), \tag{2.3.6}$$

whereas from Theorem 2.3.1(iii),

$$Bias(M^*) = o\left(\frac{1}{n}\right) + O\left(\frac{h^{2s}}{n^{p/(2s+2p+1)}}\right). \tag{2.3.7}$$

So M^* has smaller asymptotic bias than UCV_n, for all h and smaller asymptotic bias than BCV_n, if $\liminf_{n\to\infty} n^{p/(2s+2p+1)}h > 0$ which is satisfied by $h = \frac{c}{n^{1/(2s+1)}}$, $c > 0$ for $p \geq 2$. Besides BCV_n fails imitate M for large values of h (see Scott and Terrell (1987)).

Plug in estimator

Another popular estimator of M is the plug-in estimator. The basic idea of the plug-in rule (see Park and Marron (1990); Sheather and Jones (1991); Jones, Marron and Sheather (1996)) is to substitute data based estimates into the asymptotic approximation to M. It is well known that if K is of order s then $M = A_n(h) + o(\frac{1}{nh} + h^{2s})$ where $A_n(h) \equiv A_n$ equals

$$\frac{1}{nh}\int K^2(u)du + \frac{a^2}{(s!)^2}h^{2s}\int [f^{(s)}(x)]^2dx \quad \text{and} \quad a = \int x^sK(x)dx \neq 0.$$

A_n is referred to as the asymptotic mean integrated squared error. In the plug-in method $\theta_s = \int [f^{(s)}(x)]^2 dx$ is replaced by a suitable estimator (we call it $\hat{\theta}_s$) to obtain the following estimator $\hat{A}_n$:

$$\hat{A}_n = \frac{1}{nh} \int K^2(u)du + \frac{a^2}{(s!)^2} h^{2s} \hat{\theta}_s. \tag{2.3.8}$$

$\hat{A}_n$ is easy to compute. A natural question is "when is M^* more accurate than $\hat{A}_n$?" Hall and Marron (1987) introduced two kernel based estimators of θ_s. Let us denote them commonly by θ_s^H. From their Lemma 3.1 (c) and Theorem 3.2 (a) it follows that $E(\theta_s^H) - \theta_s = O\left(\frac{1}{n^{2k/(4s+2k+1)}}\right) = o\left(\frac{1}{n^{1/(2s+1)}}\right)$ and $n^{1/(2s+1)}(\theta_s^H - \theta_s) = o_P(1)$, for $s, k \geq 2$ and $\lambda_n^* = O\left(\frac{1}{n^{2/(4s+2k+1)}}\right)$, where λ_n^* is bandwidth used in θ_s^H. Our next result is motivated by this.

Theorem 2.3.2. *Under Assumptions $A - E$, $s,\ p \geq 2$ and for $\lambda = \frac{1}{n^{1/(2s+2p+1)}}$,*
(i) $\lim_{n\to\infty} \frac{E\left|\frac{M^*}{M}-1\right|}{E\left|\frac{\hat{A}_n}{M}-1\right|} = 0$ *if either (a) or (b) below holds.*
(a) $h = \frac{C}{n^{1/(2s+1)}}$, *where C is a positive constant, and* $E(\hat{\theta}_s) - \theta_s = o(\frac{1}{n^{1/(2s+1)}})$.
(b) $\limsup_{n\to\infty} nh^{2s} < \infty$ *and* $E(\hat{\theta}_s) - \theta_s = o(1)$.
(ii) $\frac{\left|\frac{M^*}{M}-1\right|}{\left|\frac{\hat{A}_n}{M}-1\right|} = o_P(1)$ *if* $n^{1/(2s+1)}(\hat{\theta}_s - \theta_s) = o_P(1)$ *and* $\limsup_{n\to\infty} nh^{2s+1} < \infty$.

Marron and Wand (1992) provide important insight into the effect of h on A_n. For fixed n, they plotted M and A_n (for Gaussian kernel) against different values of $\log_{10} h$. For a wide class of densities, including general normal mixtures, as the value of $\log_{10} h$ is increased from -1 to 0 and beyond, A_n diverges to ∞ whereas M appears to level off. Moreover, M and A_n differ significantly for all values of h satisfying $\log_{10} h \geq 0$. Consequently, for larger values of h, approximation of M by $\hat{A}_n$, which in turn is an estimator of A_n, can be poor. Further, A_n is sensitive to the structure of f. In particular, for the double-claw density, which is the density of the distribution $0.49.N(-1,(2/3)^2) + 0.49.N(1,(2/3)^2) + 0.02.\sum_{l=0}^{6} \frac{1}{350} N((l-3)/2,(0.01)^2)$,[1] A_n is a poor approximation to M (see page 725, Figure 4, in Marron and Wand (1992)). The plug-in estimator $\hat{A}_n$ is not likely to improve these demerits of A_n. So we do not recommend its use to estimate M, especially when f is believed to have complicated structures such as multi-modality, skewness and spikes.

[1] $N(x, y^2)$ denotes the normal distribution with mean x and variance y^2.

2.3.2 Fixed Sample performance of M^*

. What are the effects of different possible choices of h and λ on the bootstrap estimator M^* for *fixed sample size n?* The following proposition provides some answers.

Theorem 2.3.3. *Suppose f is uniformly bounded, continuous and K^0 satisfies C(i)(a). Then for any fixed sample size n, as $h \to \infty$,*
(i) for any choice of λ, $M \to \int f^2(y)dy$ and $M^* \to \int (K_n^0(y))^2 dy$ almost surely .
(ii) if $f(x) \to 0$, as $|x| \to \infty$, and $\lambda \to \infty$, then $E(M^*) \to 0$. Consequently $E(M^*) - M \to -\int f^2(y)dy$ and $M^* \xrightarrow{P} 0$.
(iii) for any choice of λ, $\hat{A}_n \to \infty$, almost surely.
(iv) if $\lambda \to \infty$ $M^* \to 0$, almost surely.

Remark 2.3.2. (i) Theorem 2.3.3(i) implies that for fixed n, M^* and M level off and the former succeeds in imitating the behavior of the latter for larger values of h. On the other hand, Theorem 2.3.3 (iii) demonstrates that $\hat{A}_n$ explodes as h is increased, thereby verifying the empirical observation of Marron and Wand (1992).
(ii) Theorem 2.3.3 (ii) implies that, for fixed n if h is large, M^* is likely to underestimate M when λ is also large and the structure of $f(\cdot)$ is complicated so that $\int f^2(y)dy$ is large. As a consequence, when h is large and λ is a monotonically increasing function of h, M^* can produce an estimator of M with substantial negative bias. Theorem 2.3.3 (i) and (iv) further prove that for large value of h and λ equal to an increasing function in h, M^* is a poor estimate of M. So Theorem 2.3.3 (i), (ii) and (iv) imply that λ equal to h, $Cn^p h^m$ (where $C,\ p,\ m > 0$) or any increasing function of h is not an appropriate choice of λ.

2.4 Smooth Bootstrap with Gaussian Kernel

M^* does not have a closed form expression in general and hence Monte-Carlo computation is required for its implementation. Marron and Wand (1992) obtained an exact and easily computable expression for MISE, assuming f to be a general normal mixture and Gaussian type kernels are used.

A kernel K is said to be a *Gaussian-based* kernel of order $2r$ if

$$K(x) = \frac{(-1)^r}{2^{r-1}(r-1)!x} \frac{d^{2r-1}}{dx^{2r-1}} \phi(x)$$

where $\phi(x)$ is standard normal density evaluated at x. Let $\phi_{\sigma^2}(x)$ denote the normal density with mean 0 and variance σ^2, evaluated at x, $\phi_{\sigma^2}^{(s)}(x) = \frac{d^s}{dx^s}\phi_{\sigma^2}(x)$ and

$$C_1(r) = \frac{1}{\sqrt{\pi}} \sum_{i=0}^{r-1} \sum_{j=0}^{r-1} \frac{(2i+2j)!}{2^{3(i+j)+1} i! j! (i+j)!}.$$

Lemma 2.4.1. *(Marron and Wand (1992)) If* $f(x) = \sum_{i=1}^{k} w_i \phi_{\sigma_i}(x - \mu_i)$ *and if* K *is a Gaussian-based kernel of order* $2r$*, then* M *equals*

$$\frac{C_1(r)}{nh} + \left(1 - \frac{1}{n}\right) \sum_{i=0}^{r-1} \sum_{j=0}^{r-1} \frac{(-1)^{i+j}}{2^{i+j} i! j!} U(h; i+j, 2) - 2 \sum_{s=0}^{r-1} \frac{(-1)^s}{2^s \; s!} U(h, s, 1) + U(h, 0, 0)$$

where $U(h, s, q) = \sum_{i=1}^{k} \sum_{l=1}^{k} w_i w_l h^{2s} \phi^{(2s)}_{\sigma_i^2 + \sigma_j^2 + qh^2}(\mu_i - \mu_l)$.

Now suppose that K is a Gaussian-based kernel of order $2r$ and K_0 is the Gaussian kernel. Since K_0 is Gaussian,

$$K_n^0(y) = \frac{1}{\sqrt{2\pi} n \lambda} \sum_{i=1}^{n} e^{-\frac{(y - X_i)^2}{2(\lambda)^2}}.$$

This is really a mixed normal distribution. Thus, M^* is the MISE of the kernel density estimator $K_{nB}(y)$, where the underlying density $K_n^0(\cdot)$ is a mixed normal density with n components, $w_i = \frac{1}{n}$, $\mu_i = X_i$ and $\sigma_i = \lambda$, $i = 1, 2, .., n$. Hence, using the above result, we obtain a closed expression for M^*, provided that K is a Gaussian-based kernel of order $2r$.

Theorem 2.4.2. *If* $K(\cdot)$ *is a Gaussian-based kernel of order* $2r$ *and* K_0 *is the standard normal density, then* M^* *equals*

$$\frac{C_1(r)}{nh} + \left(1 - \frac{1}{n}\right) \sum_{i=0}^{r-1} \sum_{j=0}^{r-1} \frac{(-1)^{i+j}}{2^{i+j} i! j!} U(h; i+j, 2) - 2 \sum_{s=0}^{r-1} \frac{(-1)^s}{2^s \; s!} U(h, s, 1) + U(h, 0, 0),$$

where $U(h, s, q) = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{l=1}^{n} h^{2s} \phi^{(2s)}_{2(\lambda)^2 + qh^2}(X_i - X_l)$.

Corollary 2.4.3. *If* K *and* K_0 *are both standard normal densities, then* $r = 1$ *(in Theorem 2.4.2) and hence* M^* *equals*

$$\frac{1}{2nh\sqrt{\pi}} + \left(1 - \frac{1}{n}\right) \frac{1}{n^2} \sum_{i=1}^{n} \sum_{l=1}^{n} \phi_{2(\lambda)^2 + 2h^2}(X_i - X_l) - \frac{2}{n^2} \sum_{i=1}^{n} \sum_{l=1}^{n} \phi_{2(\lambda)^2 + h^2}(X_i - X_l)$$
$$+ \frac{1}{n^2} \sum_{i=1}^{n} \sum_{l=1}^{n} \phi_{2(\lambda)^2}(X_i - X_l).$$

Remark 4 (i) The choice of K^0 as the standard normal density trivially satisfies Assumption C of Section 2 and it also leads to closed form expression for M^* when K is a Gaussian or Gaussian type kernel. So these are recommended choices.
(ii) Taylor's (1989) T is a special case of M^* and is obtained by substituting $r = 1$, $s = 0$ and choosing K as the Gaussian density and $\lambda = h$ in Theorem 2.4.2.

2.5 Simulation

To what extent does the nature of f and the value of h affect the performance of M^*, especially when the sample size is small? What is an appropriate choice of λ? We probe these questions with simulation. Since any density may be approximated arbitrarily closely in various senses by a normal mixture (see Marron and Wand (1992)), we choose f to be mixed normal. We choose K to be Gaussian due to its wide popularity. We also choose K^0 to be Gaussian and hence closed form expression for computing M^* and M are available from Theorem 2.4.2 and Marron and Wand (1992) (see their Theorem 2.1, page 716).

Recall that we already have two choices of λ, namely (i) $\lambda = h$, the proposal of Taylor (1989) and (ii) $\lambda = h_c$, the LCV bandwidth of Faraway and Jhun (1990). From Theorems 2.3.1 (iii), 2.3.2, we have a third choice, $\lambda = \frac{1}{n^{1/(2s+2p+1)}}$. If K is the standard normal density, then $s = p = 2$ and so $\frac{1}{n^{1/9}}$ is a fourth possible choice of λ for the Gaussian kernel. This choice of λ also satisfies Assumptions D.

We use the following notation to plot or results: let T the value of M^* for $\lambda = h$ and the Gaussian kernel and let M_C^*, denote M^* when $\lambda = h_c$.

For $n = 50,\ 500$, we have plotted M^* with $\lambda = \frac{1}{n^{1/9}}$, M_C^*, T and M, against $\log_{10} h$. Our choice of $\log_{10}$ scale is motivated by its use by Marron and Wand (1992). To distinguish the four curves, we number T, M, M^* (with $\lambda = \frac{1}{n^{1/9}}$) and M_c^* as 1, 2, 3 and 4 respectively. Each curve is plotted for four underlying distributions, namely standard normal, bimodal ($\frac{1}{2}N(-1, \frac{2}{3}) + \frac{1}{2}N(1, \frac{2}{3})$), double-claw and claw ($\frac{1}{2}N(0, 1) + \frac{1}{10}\sum_{l=0}^{4} N(l/2-1, (\frac{1}{10})^4)$) distributions. While normal and bimodal densities are simple, double claw and claw densities have rather complicated structures, such as existence of several peaks.

In figures 2.1(a) – 2.4(b) we plot T, M, M^* with $\lambda = \frac{1}{n^{1/9}}$ and M_c^*, numbered as 1,2,3 and 4 respectively, against $\mathrm{Log}(h) \equiv \log_{10} h$ for normal, bimodal, double-claw and claw distributions and for sample sizes $n = 50$ and 500. Both K and K^0 are standard normal densities.

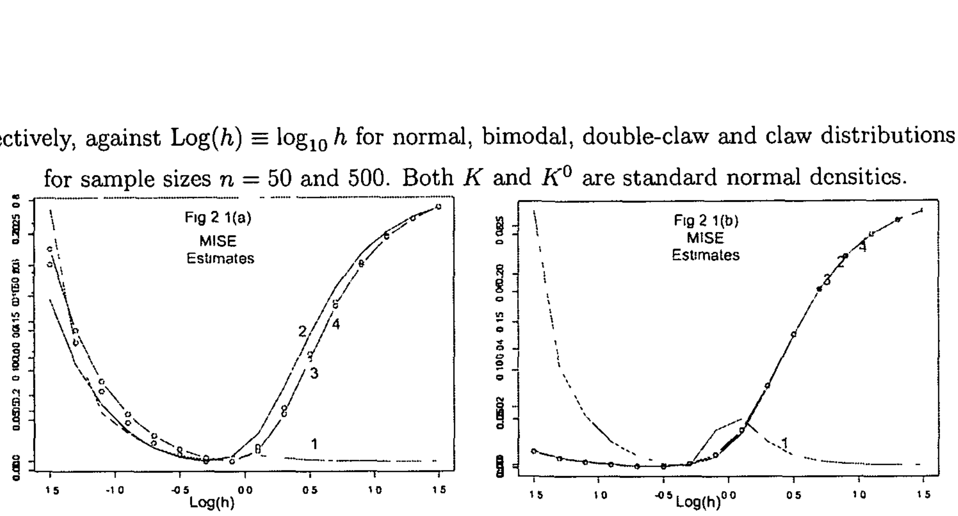


Fig 2.1 (a), (b): underlying density "normal"; $n = 50,\ 500$.

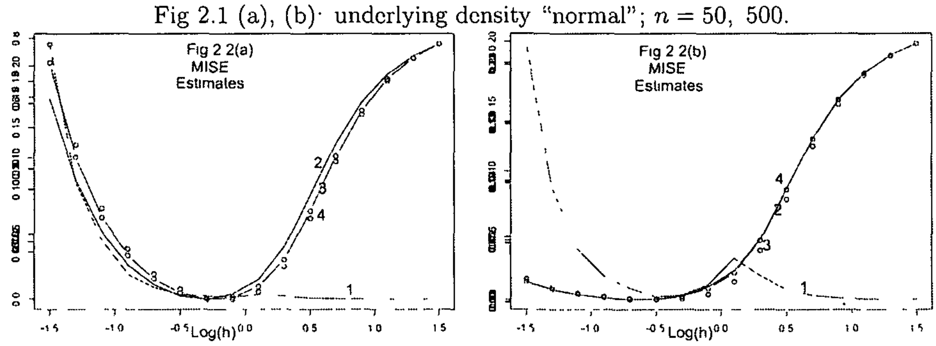


Fig 2.2 (a), (b): underlying density "bimodal", $n = 50,\ 500$.

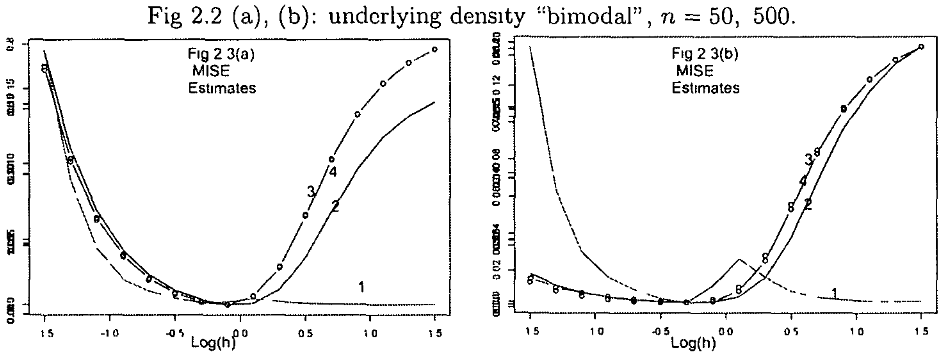


Fig 2.3 (a), (b): underlying density "double claw"; $n = 50,\ 500$.

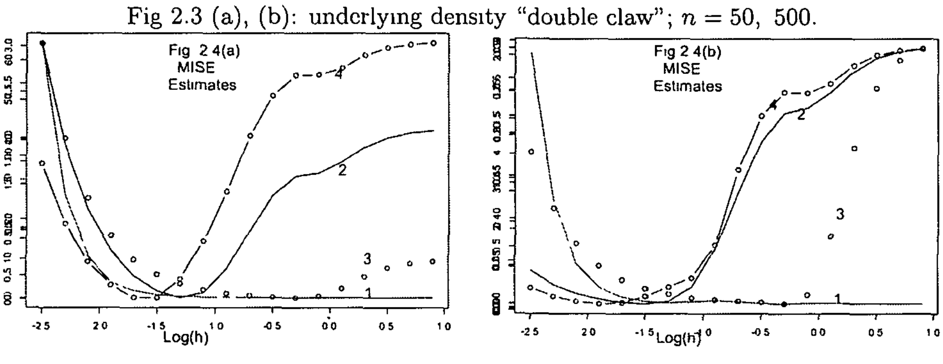


Fig 2.4 (a), (b): underlying density "claw"; $n = 50,\ 500$.

The following conclusions may be drawn from these simulations.

(i) *Small sample behaviour.* From Figures 2.1 (a) to 2.3 (a), for normal, bimodal and double claw densities, both M^* with $\lambda = \frac{1}{n^{1/9}}$ and M_C^* succeed in imitating M, for a wide range values of h and for $n = 50$. So for small samples, M^* can estimate M quite well for a number of different densities.

However comparing Figures 2.1 (a) and 2.2 (a) with 2.3 (a) and 2.4 (a), for $n = 50$, both M^* with $\lambda = \frac{1}{n^{1/9}}$ and M_C^* estimate M more closely for normal and bimodal densities than for double claw and claw densities. In particular if the underlying density is claw and $n = 50$, then M^* differs widely from M (see Figure 2.4 (a)). So for small samples M^* is more accurate if f has less complicated features. However M_C^* appears to imitate the important features of M for all the four underlying densities. For instance, for the claw density, M has two minima and the same is the case for M_C^*.

(ii) *Consistency.* As the sample size is increased from 50 to 500, both M^* with $\lambda = \frac{1}{n^{1/9}}$ and M_C^* show improvement in imitating M, for all the four underlying densities. See Figures 2.1 (a) to 2.4 (b). Even M^* with $\lambda = \frac{1}{n^{1/9}}$ exhibits some improvement as n is increased. Thus M^* appears to be consistent for $\lambda = \frac{1}{n^{1/9}}$ and h_c, irrespective of whether f is simple or complicated in its structure. However it also appears that as n increases, the accuracy of M^* with $\lambda = h_c$ can improve faster than $\lambda = \frac{1}{n^{1/9}}$, especially when f has complicated features.

(iii) *Several minima.* In general, M can have more than one local minima. In Figures 2.4 (a), (b), for the claw density, M possesses a global minima for $log_{10}(h) < -1$ and a local minima for $log_{10}(h)$ close to zero. This feature is imitated by M_C^*. If M attains more than one minima, M_C^*, M^* can differ widely and in such a situation M lies between the two estimates. If M exhibits one minima, M_C^*, M^* are close (see Figures 2.1(a) to 2.3(b)). So we recommend plotting M^* with both $\lambda = h_c$ and $\frac{1}{n^{1/9}}$ before making a final choice.

(iv) *Finite sample behaviour of T.* Cao-Abad (1990) argued that $T \to 0$, as h is increased. This result implies that T should fail to imitate M for large values h. This observation is confirmed by Figures 2.1 (a) to 2.4 (b). Our simulations reveal some more details. For example, from Figures 2.1 (b) to 2.4 (b) we see that, T fails to imitate M for small values of h as well. So in general, T does not seem to be an appropriate estimator of M. However from Figures 2.1 (b) to 2.3 (b) we can conclude that for large samples, T can possess a local minima which appears to be close to the global minima of M, and so T can be used as a criterion to select h, especially for large samples.

2.6 Proofs

We shall use the following notation. Let $*$ denote convolution, C denote some positive constant independent of n, h or λ, $||f|| = \sup_{-\infty<y<\infty} |f(y)|$ and DCT stands for Dominated Convergence Theorem. Recall that $\theta_s = \int \{f^{(s)}(y)\}^2 dy$. Let

$$K_\lambda^{0(s)}(y) = \frac{1}{\lambda^{s+1}} K^{0(s)}\Big(\frac{y}{\lambda}\Big), \quad \theta_{sn} = \frac{1}{n(n-1)} \sum\sum_{i \neq j} K_\lambda^{0(s)} * K_\lambda^{0(s)}(X_i - X_j),$$

$$\theta_n = \int \{K_n^{0(s)}(y)\}^2 dy = \frac{1}{n\lambda^{2s+1}} \int \{K^{0(s)}(v)\}^2 dv + \frac{(n-1)}{n} \theta_{sn}.$$

The following Lemmas will be used in the proof of Theorem 2.3.1.

Lemma 2.6.1. *Under Assumptions* ***A****(ii),* ***C****(ii) and* ***D****, as* $n \to \infty$,
(i) $E[\theta_n - \theta_s]^2 = O\left(\frac{1}{(n\lambda^{2s+1})^2} + \lambda^{2k}\right)$.
(ii) $E|\theta_n - \theta_s| = o(1)$.
(iii) $\int E\left[K_n^{0(s)}(y) - f^{(s)}(y)\right]^2 dy = O\left(\frac{1}{n\lambda^{2s+1}} + \lambda^{2p}\right)$.

Proof of Lemma **2.6.1** For suitable constants A_1, A_2, A_3,

$$\begin{aligned} E[\theta_n - \theta_s]^2 &= E\Big[\frac{1}{n\lambda^{2s+1}} \int \{K^{0(s)}(v)\}^2 dv + \frac{n-1}{n}\theta_{sn} - \theta_s\Big]^2 \\ &\le \frac{A_1}{(n\lambda^{2s+1})^2} + A_2\Big(1 - \frac{1}{n}\Big)^2 E[\theta_{sn} - \theta_s]^2 + \frac{A_3}{n^2}. \end{aligned}$$

Under Assumptions $A(ii)$, $C(ii)$ and $D(ii)$, Lemma 3.1 (b) and (d) and Hall and Marron (1987) implies,

$$E[\theta_{sn} - \theta_s]^2 = O\Big(\frac{1}{n^2\lambda^{4s+2}} + \lambda^{2k}\Big).$$

Hence (i) follows.

To prove (ii), note that under Assumptions $A(ii)$ and $C(ii)$, Lemma 3.1(d) of Hall and Marron (1987) implies that $E|\theta_{sn} - \theta_s| = o(1)$. Now, using Assumption D(ii), part (ii) follows because,

$$E|\theta_n - \theta_s| \le \frac{1}{n\lambda^{2s+1}} \int \{K^{0(s)}(v)\}^2 dv + E|\theta_{sn} - \theta_s| + \frac{1}{n} E|\theta_{sn}| = o(1).$$

To prove part (iii) recall that

$$K_n^0(y) = \frac{1}{n\lambda}\sum_{i=1}^{n} K^0\left(\frac{y-X_i}{\lambda}\right) \Rightarrow K_n^{0(s)}(y) = \frac{1}{n\lambda^{1+s}}\sum_{l=1}^{n} K^{0(s)}\left(\frac{y-X_l}{\lambda}\right).$$

Therefore, $E(K_n^{0(s)}(y)) = \frac{1}{\lambda^s}\int K^{0(s)}(u)f(y-\lambda u)du$. Expanding $f(y-\lambda u)$ under Assumption C on $K^{0(s)}$, $E[K_n^{0(s)}(y)] = f^{(s)}(y) + b(y)$, where

$$b(y) = \frac{(-1)^{s+p}\lambda^p}{(s+p-1)!}\int K^{0(s)}(u)u^{s+p}\int_0^1 (1-t)^{s+p-1} f^{(s+p)}(y-t\lambda u)dtdu.$$

Applying Cauchy-Schwartz inequality it is easy to verify that

$$b^2(y) \;\le\; \frac{C'\lambda^{2p}}{[(s+p-1)!]^2}\int\int_0^1 \left|K^{0(s)}(u)u^{s+p}\right|(1-t)^{s+p-1}\left[f^{(s+p)}(y-t\lambda u)\right]^2 dtdu,$$

where $C' = \frac{\int\left|K^{0(s)}(u)u^{s+p}\right|du}{s+p}$. Consequently, under Assumption A on $f^{(s+p)}(\cdot)$

$$\begin{aligned} \left[E[K_n^{0(s)}(y)] - f^{(s)}(y)\right]^2 = b^2(y) &\le \frac{C'\lambda^{2p}}{[(s+p-1)!]^2}g(y) \\ \Rightarrow \int\left[E[K_n^{0(s)}(y)] - f^{(s)}(y)\right]^2 dy &\le \frac{C'\lambda^{2p}}{[(s+p-1)!]^2}\int g(y)dy. \\ \text{Therefore } \int\left[E[K_n^{0(s)}(y)] - f^{(s)}(y)\right]^2 dy &\le C^2\lambda^{2p}\int[f^{(s+p)}(y)]^2dy, \qquad (2.6.1) \end{aligned}$$

where $C = \frac{C'}{(s+p-1)!}$ and $g(y) = \int\left|K^{0(s)}(u)u^{s+p}\right|\int_0^1(1-t)^{s+p-1}[f^{(s+p)}(y-t\lambda u)]^2dtdu$. It is easy to verify that

$$\begin{aligned} \int Var\left[K_n^{0(s)}(y)\right]dy &= \frac{1}{n\lambda^{2+2s}}\int Var\left[K^{0(s)}\left(\frac{y-X_1}{\lambda}\right)\right]dy \\ &\le \frac{1}{n\lambda^{1+2s}}\int\int\left[K^{0(s)}(u)\right]^2 f(y-u\lambda)dudy = \frac{\int\left[K^{0(s)}(u)\right]^2du}{n\lambda^{1+2s}}. \qquad (2.6.2) \end{aligned}$$

Now
$\int E\left[K_n^{0(s)}(y) - f^{(s)}(y)\right]^2 dy = \int Var\left[K_n^{0(s)}(y)\right]dy + \int\left[E[K_n^{0(s)}(y)] - f^{(s)}(y)\right]^2 dy$. Therefore from (2.6.1) and (2.6.2) we see that

$$\int E\left[K_n^{0(s)}(y) - f^{(s)}(y)\right]^2 dy \le \frac{\int\left[K^{0(s)}(u)\right]^2du}{n\lambda^{1+2s}} + C_1^2\lambda^{2p}\int[f^{(s+p)}(y)]^2dy. \quad (2.6.3)$$

Part (iii) follows from the equation (2.6.3) and so the Lemma 2.6.1 is proved completely. □

The following Lemma is used in the proof of Lemma 2.6.3. Let

$$C_1 = \frac{1}{s}\int|K(u)u^s|du,$$

$$e_n = \int \left\{ \int \int_0^1 (1-t)^{s-1} K(u) u^s E[K_n^{0(s)}(y - thu)] dt du \right\}^2 dy.$$

Lemma 2.6.2. *Under Assumptions **A**-**C**, as $n \to \infty$,*

$$e_n = C_1^2 \theta_s + O\left(h + \lambda^p\right).$$

***Proof of Lemma* 2.6.2** It is easy to see that for each y, t and u,

$$E[K_n^{0(s)}(y - thu)] = \frac{1}{\lambda^s} \int K^{0(s)}(z) f(y - thu - \lambda.z) dz.$$

Under Assumptions $C(ii)$ (c) and (d) on $K^{0(s)}$ and Assumption A on $f(\cdot)$, using Taylor's expansion with integral remainder we see that,

$$\begin{aligned} E[K_n^{0(s)}(y - thu)] &= \frac{1}{\lambda^s} \int K^{0(s)}(z) f(y - thu - \lambda.z) dz, \\ &= f^{(s)}(y - thu) \\ &\quad + \frac{(-1)^{s+p} \lambda^p}{(s+p-1)!} \int K^{0(s)}(z) z^{s+p} \int_0^1 (1-t_1)^{s+p-1} f^{(s+p)}(y^*) dt_1 dz \end{aligned}$$

where $y^* = y - thu - t_1 \lambda.z$ and therefore

$$e_n = \int \Big\{ \int \int_0^1 (1-t)^{s-1} K(u) u^s f^{(s)}(y - thu) dt du + (-1)^{s+p} \frac{\lambda^p}{(s+p-1)!} g_n(y) \Big\}^2 dy.$$

$$\begin{aligned} \text{So } e_n &= \int \Big\{ \int \int_0^1 (1-t)^{s-1} K(u) u^s f^{(s)}(y - thu) dt du \Big\}^2 dy \\ &\quad + \frac{\lambda^{2p}}{((s+p-1)!)^2} \int g_n^2(y) dy \\ &\quad + 2 \frac{(-\lambda)^p}{(s+p-1)!} \int \Big\{ \int \int_0^1 (1-t)^{s-1} K(u) u^s f^{(s)}(y - thu) dt du \Big\} g_n(y) dy \\ &= t_{1n} + t_{2n} + t_{3n} \quad \text{(say), where} \\ g_n(y) &= \int K(u) u^s \int_0^1 (1-t)^{s-1} \Big\{ \int K^{0(s)}(v) v^{s+p} \int_0^1 (1-t_1)^{s+p-1} f^{(s+p)}(y^*) dt_1 dv \Big\} dt du \quad \text{and} \\ y^* &= y - tuh - t_1 \lambda.v. \end{aligned}$$

Under Assumption A (i) on $f^{(s)}$ and for suitable constants C_1', C_2'

$$|t_{3n}| \leq \lambda^p C_1' \int |g_n(y)| dy \leq \lambda^p C_2' \int |f^{(s+p)}(y)| dy = O(\lambda^p).$$

Using Cauchy-Schwartz inequality it is easy to see that, $t_{2n} \leq C^2 \lambda^{2p} \int \{f^{(s+p)}(y)\}^2 dy$. Using Assumptions A $(i), (ii)$ on $f^{(s)}$, it is easy to see that $t_{1n} = C_1^2 \theta_s + O(h)$. Therefore, combining these estimates,

$$e_n = C_1^2 \theta_s + O(h) + O(\lambda^{2p}) + O(\lambda^p) = C_1^2 \theta_s + O(h + \lambda^p).$$

Hence Lemma 2.6.2 is proved completely. □

Let

$$\begin{aligned} c_n &= \int \left\{ \int \int_0^1 (1-t)^{s-1} K(u) u^s K_n^{0(s)}(y - thu) dt du \right\}^2 dy, \\ d_n &= \int \left\{ \int \int_0^1 (1-t)^{s-1} K(u) u^s f^{(s)}(y - thu) dt du \right\}^2 dy. \end{aligned}$$

Lemma 2.6.3. *Suppose $k = p$ and Assumptions A-D hold. Then* $2C_1^4 E(\theta_n)\theta_s - 2E(c_n).d_n = O(r_n)$.

***Proof of Lemma* 2.6.3** Using Cauchy-Schwartz inequality it is easy to see that

$$2Ec_n.d_n \le 2C_1^4 E(\theta_n)\theta_s \quad \text{and} \quad Ec_n \ge e_n$$

Hence

$$0 \le 2C_1^4 E(\theta_n)\theta_s - 2E(c_n).d_n \le 2C_1^4 E(\theta_n)\theta_s - 2e_n.d_n. \tag{2.6.4}$$

From Lemma 2.6.1(i), under the Assumptions $A(ii)$, $C(ii)$, D and using $k = p$,

$$E(\theta_n) = \theta_s + O\left(\frac{1}{n\lambda^{2s+1}} + \lambda^p\right).$$

Under Assumption A on $f(\cdot)$, $d_n = C_1^2\theta_s + O(h)$. Under the Assumptions $C(ii)$ (c) and (d) on $K^{0(s)}$ and Assumption A on $f(\cdot)$, from Lemma 2.6.2, $e_n = C_1^2\theta_s + O(h + \lambda^p)$. Substituting the above equations in the right side of (2.6.4) and under Assumption D on λ we get, as $n \to \infty$.

$$\begin{aligned} 0 &\le 2C_1^4 E(\theta_n)\theta_s - 2E(c_n).d_n \\ &\le 2C_1^4 [\theta_s + O\left(\frac{1}{n\lambda^{2s+1}} + \lambda^p\right)]\theta_s - 2\left[C_1^2\theta_s + O\left(h + \lambda^p\right)\right]\left[C_1^2\theta_s + O(h)\right] \\ &= O\left(\frac{1}{n\lambda^{2s+1}} + \lambda^p + h\right) \quad (\text{since } p \ge 1). \end{aligned}$$

Hence Lemma 2.6.3 is proved completely. □

2.6.1 Proofs of Theorems

Proof of Theorem 2.3.1 Recalling the definitions of M^* and M it is easy to verify that, almost surely,

$$\begin{aligned}
&\left|M^* - M\right| \leq L_{1n} + L_{2n} \quad \text{(say), where} \qquad (2.6.5)\\
&L_{1n} = \frac{1}{n}\Big| \int \Big\{ \int K(v)K_n^0(y-hv)dv\Big\}^2 dy - \int \Big\{ \int K(v)f(y-hv)dv\Big\}^2 dy\Big|,\\
&L_{2n} = \left| \int \left[\int K(u)K_n^0(y-h.u)du - K_n^0(y)\right]^2 dy - \int \left[\int K(u)f(y-h.u)du - f(y)\right]^2 dy \right|.
\end{aligned}$$

Therefore

$$E|M^* - M|^2 \leq 2E(L_{1n}^2) + 2E(L_{2n}^2). \qquad (2.6.6)$$

Using $|a^2 - b^2| = (a+b)|a-b|$, for $a,\ b > 0$, it is easy to see that (writing $y^* = y - hv$)

$$\begin{aligned}
&L_{1n}\\
\leq\ & \frac{1}{n}\int [\{ \int K(v)(K_n^0(y^*) + f(y^*))dv\}\{ \int K(v)|K_n^0(y^*) - f(y^*)|dv\}]dy \qquad (2.6.7)\\
=\ & D_n, \text{ say.}
\end{aligned}$$

We note that, almost surely

$$K_n^0(y^*) + f(y^*) \leq |K_n^0(y^*) - f(y^*)| + 2||f||.$$

Therefore, almost surely

$$\begin{aligned}
D_n &\leq \frac{1}{n}[\int \{ \int K(v)|K_n^0(y^*) - f(y^*)|dv\}^2 dy + 2||f|| \int K(v) \int |K_n^0(y^*) - f(y^*)|dydv]\\
\Rightarrow D_n &\leq \frac{1}{n}[\int \int K(v)\{K_n^0(y-hv) - f(y-hv)\}^2 dvdy\\
&\quad +2||f|| \int |K_n^0(y) - f(y)|dy]\\
&= \frac{1}{n}[\int \{K_n^0(y) - f(y)\}^2 dy + 2||f|| \int |K_n^0(y) - f(y)|dy].
\end{aligned}$$

Hence from (2.6.7) we have, almost surely

$$\begin{aligned}
L_{1n} &\leq \frac{1}{n}[\int \{K_n^0(y) - f(y)\}^2 dy + 2||f|| \int |K_n^0(y) - f(y)|dy] \qquad (2.6.8)\\
\Rightarrow\ E(L_{1n}^2) &\leq \frac{1}{n^2}E[\int \{K_n^0(y) - f(y)\}^2 dy + 2||f|| \int |K_n^0(y) - f(y)|dy]^2\\
&\leq \frac{2}{n^2}[E[\int \{K_n^0(y) - f(y)\}^2 dy]^2\\
&\quad + 4||f||^2 E[\int |K_n^0(y) - f(y)|dy]^2] \qquad (2.6.9)
\end{aligned}$$

Since K^0 is a density function and $nh \to \infty$, therefore if follows from Devroye (1983) that, almost surely, $\int |K_n^0(y) - f(y)|dy = o(1)$. Since both f and K^0 are probability density functions, $\int |K_n^0(y) - f(y)|dy \leq 2 \;\forall\; n$. Therefore, by Dominated Convergence Theorem (DCT),

$$4||f||^2 E\left[\int |K_n^0(y) - f(y)|dy\right]^2 = o(1).$$

Further, under Assumptions A, C on f and K^0, $E\big[\int \{K_n^0(y) - f(y)\}^2 dy\big]^2$ equals

$$E\Big[\int \{K_n^0(y)\}^2 dy - \int f^2(y)dy + 2\int f(y)\{f(y) - K_n^0(y)\}dy\Big]^2$$
$$\leq 2E\Big[\int \{K_n^0(y)\}^2 dy - \int \{f(y)\}^2 dy\Big]^2 + 8||f|| \int E\{f(y) - K_n^0(y)\}^2 dy.$$

Under the Assumptions $A(i)$, $C(i)$, $D(i), (ii)$, from Rao (1983, pp. 45),

$$\int E\{f(y) - K_n^0(y)\}^2 dy = o(1).$$

Under Assumptions $A(i), (ii), C(ii)$ and D, from Hall and Marron (1987) we get

$$E\left[\int \{K_n^0(y)\}^2 dy - \int \{f(y)\}^2 dy\right]^2 = o(1).$$

Therefore under Assumptions $A - E$,

$$E\left[\int \{K_n^0(y) - f(y)\}^2 dy\right]^2 = o(1).$$

Hence using (2.6.9) we find that, under Assumptions $A - E$,

$$E(L_{1n}^2) = o\left(\frac{1}{n^2}\right). \tag{2.6.10}$$

In view of (2.6.6) and (2.6.10), to prove Theorem 2.3.1(i), it is enough to show that

$$\frac{1}{h^{4s} r_n} E(L_{2n}^2) = O(1). \tag{2.6.11}$$

Now under the smoothness Assumptions A, C on f and K^0, using Taylor's expansion with integral remainder we get

$$\begin{aligned} L_{2n} &= \frac{h^{2s}}{((s-1)!)^2}\Big|\int\Big\{\int\int_0^1(1-t)^{s-1}K(u)u^sK_n^{0(s)}(y-thu)dtdu\Big\}^2dy \\ &\quad -\int\Big\{\int\int_0^1(1-t)^{s-1}K(u)u^sf^{(s)}(y-thu)dtdu\Big\}^2dy\Big| \quad (2.6.12) \\ &= \frac{h^{2s}}{((s-1)!)^2}|c_n-d_n| \quad \text{(say)}. \\ \Rightarrow E(L_{2n}^2) &\le \frac{h^{4s}C_1^4}{((s-1)!)^4}\left[E[\theta_n]^2+\theta_s^2-2Ec_n.d_n\right] \\ &\le \frac{h^{4s}}{((s-1)!)^4}\left[C_1^4E\big[\theta_n-\theta_s\big]^2+2C_1^4E(\theta_n)\theta_s-2Ec_n.d_n\right]. \quad (2.6.13) \end{aligned}$$

Now we use Lemmas 2.6.1(i), 2.6.3, under assumption $k=p$, in the right side of (2.6.13) to conclude that

$$\frac{1}{h^{4s}}E(L_{2n}^2)=O\left(\frac{1}{(n\lambda^{2s+1})^2}+\lambda^{2p}+r_n\right)=O(r_n),$$

establishing (2.6.11). Now recalling (2.6.10), part (i) is proved.

We now prove part (ii). Under Assumptions A on $f^{(s)}$ and B on $K(\cdot)$, from Rao (1983, pp. 45, Theorem 2.1.7), we see that

$$\begin{aligned} M &= \frac{1}{nh}\int K^2(u)du+\frac{a^2}{(s!)^2}(h)^{2s}\int[f^{(s)}(x)]^2dx+o\left(\frac{1}{nh}+h^{2s}\right) \quad \text{where} \\ a &= \int x^sK(x)dx\neq 0. \end{aligned}$$

Using part (i) of Theorem 2.3.1,

$$E\left[\frac{M^*}{M}-1\right]^2=\frac{E[M^*-M]^2}{M^2}=\frac{o(\frac{1}{n^2})+O(h^{4s}r_n)}{\frac{D_1}{(nh)^2}+D_2h^{4s}+o\left(\frac{1}{(nh)^2}+h^{4s}\right)}, \quad (2.6.14)$$

where D_1, D_2 are positive constants, independent of n.

If $\limsup_{n\to\infty} nh^{2s+1}<\infty$ then, multiplying numerator and denominator of the right side of (2.6.14) by $(nh)^2$ we get,

$$E\left[\frac{M^*}{M}-1\right]^2=o(h^2)+O(r_n).$$

If $\lim_{n\to\infty} nh^{2s+1}=\infty$ then again, dividing numerator and denominator of the right side of (2.6.14) by h^{4s}, we get again the same estimate. This completes the proof of part (ii). Finally we prove parts (iii) and (iv). From equation (2.6.5) we get

$$E|M^* - M| \leq E[L_{1n}] + E[L_{2n}].$$

Since $[E[L_{1n}]]^2 \leq E[L_{1n}^2]$, therefore from equation (2.6.10) we see that $E[L_{1n}] = o\left(\frac{1}{n}\right)$. Next we show that for $\lambda = \frac{1}{n^{1/(2s+2p+1)}}$, $E[L_{2n}] = O\left(\frac{h^{2s}}{n^{p/(2s+2p+1)}}\right)$.

From equation (2.6.12) it is easy to see that

$$\begin{aligned} E[L_{2n}] &\leq \frac{h^{2s}}{[(s-1)!]^2}\left[\int E\left[f_{2y}^2\right] dy + 2\int E\left[f_{2y}f_{3y}\right] dy\right] \quad \text{where} \\ f_{2y} &= \int |K(u)u^s| \int_0^1 (1-t)^{s-1} \left|f^{(s)}(y-thu) - K_n^{0(s)}(y-thu)\right| dtdu \quad \text{and} \\ f_{3y} &= \int |K(u)u^s| \int_0^1 (1-t)^{s-1} \left|f^{(s)}(y-thu)\right| dtdu. \end{aligned}$$

Using Lemma 2.6.1(iii) we see that

$$\int E\left[f_{2y}^2\right] dy \leq \left[\frac{\int |K(u)u^s|du}{s}\right]^2 \int E\left[K_n^{0(s)}(y) - f^{(s)}(y)\right]^2 dy = O\left(\frac{1}{n^{2p/(2s+2p+1)}}\right) \tag{2.6.15}$$

and consequently

$$\int E\left[f_{2y}f_{3y}\right] dy = \int E\left[f_{2y}\right] f_{3y} dy \leq \sqrt{\int E\left[f_{2y}^2\right] dy. \int f_{3y}^2 dy} = O\left(\frac{\cdot 1}{n^{p/(2s+2p+1)}}\right). \tag{2.6.16}$$

So $E|L_{2n}| = O\left(\frac{h^{2s}}{n^{p/(2s+2p+1)}}\right)$ and therefore

$$E|M^* - M| = o\left(\frac{1}{n}\right) + O\left(\frac{h^{2s}}{n^{p/(2s+2p+1)}}\right).$$

Thus part (iii) is proved.

Part (iv) follows from part (iii) as follows

$$E\left|\frac{M^*}{M} - 1\right| = \frac{E\left|M^* - M\right|}{M} = \frac{o(\frac{1}{n}) + O\left(\frac{h^{2s}}{n^{p/(2p+2s+1)}}\right)}{\sqrt{\frac{D_1}{(nh)^2} + D_2h^{4s} + o\left(\frac{1}{(nh)^2} + h^{4s}\right)}}.$$

If $\limsup nh^{2s+1} < \infty$, multiplying numerator and denominator by nh we see that

$$E\left|\frac{M^*}{M} - 1\right| = o(h) + O\left(\frac{1}{n^{p/(2p+2s+1)}}\right).$$

If $\liminf nh^{2s+1} > 0$, then dividing the numerator and denominator by h^{2s}, using the conditions $\liminf nh^{2s+1} > 0$ and $h = o(1)$, we get the same rate of convergence of

$E\left|\frac{M^*}{M}-1\right|$ to zero.

So Theorem 2.3.1 is proved completely. □

Proof of Theorem 2.3.2 We note that

$$\frac{E\left|\frac{M^*}{M}-1\right|}{E\left|\frac{\hat{A}_n}{M}-1\right|}=\frac{nE|M^*-M|}{nE\left|\hat{A}_n-M\right|}.$$

Recall that, under Assumptions A to E and for $\lambda=\frac{1}{n^{1/(2s+2p+1)}}$, from Theorem 2.3.1 (iii),

$$E|M^*-M|=o\left(\frac{1}{n}\right)+O\left(\frac{h^{2s}}{n^{p/(2s+2p+1)}}\right).$$

Therefore

$$nE|M^*-M|=o(1)\ \text{ when }\ s,p\geq 2\ \text{ and }\ h=O\left(\frac{1}{n^{1/(2s+1)}}\right).$$

So to prove Theorem 2.3.2, it is enough to show that $\liminf_{n\to\infty} nE\left|\hat{A}_n-M\right|>0$. Recalling the formulae of $\hat{A}_n$ and M given in (2.3.8) and (2.1.1) it is easy to verify that

$$\begin{aligned} E\left|\hat{A}_n-M\right| &\geq E(\hat{A}_n-M) \\ &= \frac{a_n}{n}+\frac{a^2h^{2s}}{(s!)^2}E\left[\hat{\theta}_s\right]-\int\{E(K_n(y)-f(y)\}^2dy, \text{ where} \\ a_n &= \int\left\{\int K(v)f(y-hv)dv\right\}^2dy. \end{aligned} \tag{2.6.17}$$

Now from Rao (1983, page 45)

$$\int\{E(K_n(y)-f(y)\}^2dy\leq\frac{a^2}{(s!)^2}h^{2s}\int\left\{f^{(s)}(x)\right\}^2dx=\frac{a^2}{(s!)^2}h^{2s}\theta_s.$$

Therefore from (2.6.17),

$$E\left|\hat{A}_n-M\right|\geq\frac{a_n}{n}+\frac{a^2}{(s!)^2}h^{2s}E\left[\hat{\theta}_s-\theta_s\right].$$

Under Assumptions (a) or (b) of the Theorem on h and $\hat{\theta}_s$, $nh^{2s}E\left[\hat{\theta}_s-\theta_s\right]=o(1)$ and under Assumptions A and B, it also easy to verify that $\liminf_{n\to\infty} a_n\geq\int f^2(y)dy>0$. Therefore $\liminf nE\left|\hat{A}_n-M\right|\geq\int f^2(y)dy>0$ and so part (i) of Theorem 2.3.2 is proved.

To prove part (ii), we see that $\frac{|\frac{M^*}{M}-1|}{|\frac{\hat{A}_n}{M}-1|} = \frac{n|M^*-M|}{n|\hat{A}_n-M|}$.
From Theorem 2.3.1 (iii), under Assumptions $\limsup_{n\to\infty} nh^{2s+1} < \infty$ and $s, p \geq 2$,

$$n|M^* - M| = o_P(1) + O_P\left(\frac{nh^{2s}}{n^{p/(2s+2p+1)}}\right) = o_P(1).$$

So to prove part (ii) it is enough to show that, almost surely, $n|\hat{A}_n - M| \geq C + Y_n$, where C is a positive constant and $Y_n = o_P(1)$, for all large values of n.

It is easy to verify that $n|\hat{A}_n - M| \geq a_n + \frac{a^2}{(s!)^2}nh^{2s}\left[\hat{\theta}_s - \theta_s\right]$, almost surely.
If $n^{1/(2s+1)}(\hat{\theta}_s - \theta_s) = o_P(1)$ and $\limsup_{n\to\infty} nh^{2s+1} < \infty$, then $Y_n = nh^{2s}\left[\hat{\theta}_s - \theta_s\right] = o_P(1)$. We also recall that $\liminf_{n\to\infty} a_n > 0$. Therefore, almost surely, $n|\hat{A}_n - M| \geq C + Y_n$, where $C > 0$ and $Y_n = o_P(1)$, for all large values of n. This proves part (ii) and therefore Theorem 2.3.2 is proved completely. □

***Proof of Theorem* 2.3.3** Recalling the definition of M^* we see that

$$E(M^*) \leq \frac{1}{nh}\int K^2(u)du + E[B^*].$$

We note that $\frac{1}{nh}\int K^2(u)du = o(1)$, as h is increased. So to prove Theorem 2.3.3 it is enough to show that $E[B^*] = o(1)$, as λ, h are increased. Using Cauchy-Schwartz inequality we get

$$\begin{aligned}
0 \leq E[B^*] &\leq \int |K(u)|du \times E\Big[\int\int |K(u)| \left[K_n^0(y - h.u) - K_n^0(y)\right]^2 dudy \\
&\leq T_1 \int\int |K(u)| E\left[K_n^0(y - hu) - K_n^0(y)\right]^2 dudy \quad \text{where } T_1 = \int |K(u)|du, \\
&= \frac{2T_1^2}{n\lambda}\int \left\{K^0(v)\right\}^2 dv + 2T_1^2\left(1 - \frac{1}{n}\right)\int\left\{\int K^0(u)f(y - \lambda.u)du\right\}^2 dy \\
&\quad -\frac{2T_1}{n\lambda}\int f(y)dy\int |K(u)|\int K^0(x)K^0\left(x - \frac{hu}{\lambda}\right)dxdu \\
&\quad -2T_1(1 - \frac{1}{n})\int\int |K(u)|\int\int K^0(z_1)K^0(z_2)f(y_1^*)f(y_2^*)dz_1dz_2dudy. \\
&= b_n \text{ (say)}.
\end{aligned}$$

where $y_1^* = y - \lambda.u - hz_1$ and $y_2^* = y - \lambda.z_2$.

Letting $h \to \infty$ and $\lambda \to \infty$, it is easy to see (using the condition that $|f(x)| \to 0$ as $|x| \to \infty$ and DCT) that $b_n \to 0$ and hence $E(B^*) \to 0$. Consequently $E(M^*) \to 0$, as h, $\lambda \to \infty$, proving the first part of (ii). Since M^* is a nonnegative random variable, the second part follows from Markov's inequality.

(i) and (iii) are easy to prove. Finally we prove part (iv). Recalling the definition of M^*, we note that $M^* \leq \frac{\int K^2(u)du}{nh} + B^*$. Recalling the definition of B^*, it is easy to verify that

$$\begin{aligned} B^* &\leq 2\left[\left(\int |K(u)|du\right)^2 + 1\right] \int \left[K_n^0(y)\right]^2 dy \\ &\leq 2\left[\left(\int |K(u)|du\right)^2 + 1\right] \left[\frac{\int \left[K^0(u)\right]^2 du}{n\lambda} + \frac{2\|K^0\|}{\lambda}\left(1 - \frac{1}{n}\right)\right]. \end{aligned}$$

Consequently

$$M^* \leq \frac{\int K^2(u)du}{nh} + 2C. \left[\frac{\int \left[K^0(u)\right]^2 du}{n\lambda} + \frac{2\|K^0\|}{\lambda}\left(1 - \frac{1}{n}\right)\right], \text{ almost surely,}$$

where $C = \left[\left(\int |K(u)|du\right)^2 + 1\right]$. Therefore if $\lambda \to \infty$, as $h \to \infty$, then $M^* \to 0$, almost surely, as h is increased. So Theorem 2.3.3 is proved completely. □

Chapter 3

Estimating pointwise bias and mean squared error

3.1 Introduction

In this chapter we address the problem of estimating some point-wise measures of accuracy of the kernel density estimator $K_n(\cdot)$, introduced in the previous chapter. The framework remains same, for instance $X_1, X_2, ..., X_n$ be n i.i.d. random variables with density $f(\cdot)$ and $K_n(\cdot)$ is the *kernel density estimator* (of f) based on the kernel $K(\cdot)$ and bandwidth $h \equiv h_n$.

There are two basic assumptions on h, namely $h \to 0$ and $nh \to \infty$ as $n \to \infty$. Let us denote the *bias*, *variance* and *mean squared error (MSE)* of the kernel density estimator $K_n(\cdot)$, at y, by B_y, V_y and M_y respectively. Note that each of the above may be expressed as a functional $T(f)$. These local measures of accuracy of $K_n(\cdot)$ have enjoyed great popularity, especially in the context of locally optimal bandwidth selection of a kernel estimator. See for example Hall (1990), Falk (1992) and references therein.

In general $T(f)$ is unknown. In the smooth bootstrap approach $T(f)$ is estimated by $T(f_n)$ where f_n is an appropriate estimate of f_n (say a kernel density estimate with kernel K^0 and bandwidth λ). Traditionally, such smooth bootstrap estimators use the same kernel, that is $K^0 = K$. See for instance Hall (1992) who described two possible versions of smooth bootstrap bias estimator, where $K^0 = K$ but λ is either larger than h or some unspecified value. The analytic study of such estimators require additional restrictions on the basic kernel K. We have already mentioned some demerits of the proposal $K^0 = K$, especially when K is a *rectangualr*, *triangular* or *quadratic* kernel. For instance the resulting bootstrap bias estimator can be inconsistent.

In this chapter we propose that the kernel K^0 and the bandwidth λ be chosen freely and not tied to the original kernel and bandwidth. We call these estimators B_y^*, V_y^* and M_y^* respectively. In our approach we need additional conditions on K^0 an λ and thus avoid imposing additional conditions on K and h as far as possible. There are other performance based reasons for choosing K^0 and λ over the automatic choice of $K^0 = K$ and $\lambda = h$.

3.1.1 Definitions

The parameters B_y, V_y and M_y are defined as follows

$$B_y \equiv B_y(K,h) = E[K_n(y)] - f(y) = \int K(u)\,[f(y-hu) - f(y)]\,du$$

$$V_y \equiv V_y(K,h) = \frac{1}{nh}\int K^2(u) f(y-hu)du - \frac{1}{n}\left(\int K(u) f(y-hu)du\right)^2$$

$$\text{and } \; M_y \equiv M_y(K,h) = V_y + (B_y)^2 .$$

Let K^0 be another kernel and $\lambda \equiv \lambda_n, \mu \equiv \mu_n$ be two other bandwidth sequences. Let

$$K_n^0(y) = \frac{1}{n\lambda}\sum_{i=1}^{n} K^0\Big(\frac{y - X_i}{\lambda}\Big) \quad \text{and} \quad K_n^*(y) = \frac{1}{n\mu}\sum_{i=1}^{n} K^0\Big(\frac{y - X_i}{\mu}\Big).$$

The proposed *general smooth bootstrap estimator* of B_y and M_y are defined as

$$B_y^* = E_n[K_n(y)] - K_n^0(y) = \int K(u)\,\left[K_n^0(y-hu) - K_n^0(y)\right] du$$

$$V_y^* = \frac{1}{nh}\int K^2(u) K_n^*(y-hu)du - \frac{1}{n}\left(\int K(u) K_n^*(y-hu)du\right)^2 \text{ and } M_y^* = V_y^* + (B_y^*)^2 .$$

So $K_n^0(\cdot)$ is used in bias estimation and $K_n^*(\cdot)$ is used in variance estimation.

Let $r_1 = E\left[\frac{B_y^*}{B_y} - 1\right]^2$, $\; r_2 = E\left[\frac{V_y^*}{V_y} - 1\right]^2$ and $r_3 = E\left[\frac{M_y^*}{M_y} - 1\right]^2$.

Note that r_1, r_2 and r_3 depend on y and n. Plug-in estimators are obtained by substituting data based estimates into the asymptotic approximation of B_y and M_y and are easy to compute. Under Assumption A on f and B on K it is easy to see that

$$B_y = \frac{(-h)^s}{s!} \int K(u) u^s f^{(s)}(y) + o(h^s) \text{ and}$$
$$M_y = \frac{f(y)}{nh} \int K^2(u) du + \left[\frac{(-h)^s}{s!} \int K(u) u^s f^{(s)}(y) \right]^2 + o\left(\frac{1}{nh} + h^{2s} \right).$$

The corresponding plug in estimators may then be defined as

$$B_y^A = \frac{(-h)^s}{s!} \int K(u) u^s K_n^{0(s)}(y),\ V_y^A = \frac{K_n^*(y)}{nh} \int K^2(u) du \text{ and } M_y^A = V_y^A + [B_y^A]^2.$$

Let $r_4 = E\left[\frac{B_y^A}{B_y} - 1\right]^2$, $\quad r_5 = E\left[\frac{V_y^A}{V_y} - 1\right]^2$ and $r_6 = E\left[\frac{M_y^A}{M_y} - 1\right]^2$.

3.1.2 Notation

Let h_y, h_y^* denote the values of h which minimize (globally) M_y and M_y^* respectively. h_y shall be referred to as the *optimal bandwidth*. A bandwidth h will be referred to as *sup-optimal* or *super-optimal* if $n^{1/2s+1}h$ is $o(1)$ or diverges to ∞ respectively.

The point y is said to be a *mode* or an *anti-mode* of f if $f^{(1)}(y) = 0$ and $f^{(2)}(y) < 0$ or $f^{(2)}(y) > 0$ respectively. We note that if f is a smooth function, then our definitions of mode and anti-mode coincide with the peaks and troughs of f.

Let $N(x, y^2)$ denotes the normal distribution with mean x and variance y^2. For any function H, $H^{(i)}$ shall denote its *ith* derivative and $||H|| = \sup_{-\infty<x<\infty} |H(x)|$. For any two positive sequences $\{a_n\}, \{b_n\}$ we write $a_n = Oe(b_n)$ if $0 < \liminf_{n\to\infty} \frac{a_n}{b_n} \leq \limsup_{n\to\infty} \frac{a_n}{b_n} < \infty$.

3.1.3 A brief literature review

There are several estimates of B_y and M_y available in the literature. Hall (1990) has proposed a bootstrap scheme, where the size of the bootstrap resample is less than the size of the original sample and K is compactly supported. Theorems 2.1 (Hall (1990), page 182-183) proves strong consistency of his bootstrap estimator M_y^H of M_y.

Falk (1992) proposed smooth bootstrap estimators B_y^F and M_y^F of B_y and M_y. These are special cases of our B_y^* and M_y^* when we impose $K^0 = K$. Under the assumptions that $h = O(\frac{1}{n^{1/5}})$ and K is a compactly supported second order kernel, Falk (1992) has studied the weak convergence of B_y^F and asymptotic behaviour of $n^{4/5}M_y^F$.

There are also plug-in estimators B_y^A, V_y^A and M_y^A available based on the asymptotic approximations of B_y, V_y and M_y (see Hall (1992)). There does not seem to have been any analytic study on the accuracy of the above three sets of estimators. For instance how fast do $\frac{M_y^F}{M_y}$ converge to one? Besides, no theoretical results on the finite sample properties of the boostrap estimators of B_y or V_y, seem to have been worked out. We investigate both, the asymptotic accuracy as well as finite sample properties of our estimators.

3.1.4 Chapter summary

There are eight Theorems and one simulation study in this chapter. In Theorems 3.3.1 and 3.3.2 we obtain the rates at which $r_1 = E\left[\frac{B_y^*}{B_y} - 1\right]^2$ and $r_3 = E\left[\frac{M_y^*}{M_y} - 1\right]^2$ converge to zero for a reasonably broad class of K, f and for any choice of h. The proposed bootstrap estimators compare well with plug-in estimators B_y^A, V_y^A and M_y^A, see Theorems 3.3.3, 3.3.4 and 3.3.5 respectively. For instance our results imply that for super-optimal h (defined later), B_y^* is asymptotically more accurate (in L_2 sense) than B_y^A, see Theorem 3.3.3. If $\limsup_{n\to\infty} n^{2/5}h = \infty$ and $f^{(1)}(y) = 0$, $f^{(2)}(y) \neq 0$ then V_y^* has infinite asymptotic relative accuracy (in L_2 sense) in comparison to V_y^A. We have proved some properties of the proposed estimators for fixed sample size, see Theorem 3.3.6. In the Theorems 3.3.8 and 3.3.9, we have obtained closed form formulae for B_y, V_y, M_y and their corresponding estimators when K and K^0 are Gaussian kernels. So in this case, all of them can be computed explicitly.

Simulations reveal that when y is a mode or an anti-mode and $\log_{10}(h) \leq -0.5$ both M_y^A and M_y^* estimate M_y accurately. However when y is in the tail region, M_y as a function of h possesses more than one minima, and this feature is successfully imitated by M_y^*. But M_y^A always possesses one minima, say h_y^A. Let h_y and h_y^* be the global minimizers of M_y and M_y^*. If y is in the tail region, then both $\log_{10}(h_y^*)$, $\log_{10}(h_y) \geq 0.5$, whereas $\log_{10}(h_y^A) < 0$. This observation verifies Sain and Scott's (1995) result that when y is in the tail of f, h_y can be rather large. In such a situation h_y^* appears to be the more appropriate estimator of h_y.

3.2 Assumptions

We collect below all the assumptions on the two kernels and the bandwidths. Not all of them will be used in all the results.

Assumption A. (Assumptions on density f).

(i) $f(\cdot)$ is uniformly bounded, continuous and possesses $s \geq 2$ uniformly bounded derivatives. The sth derivative $f^{(s)}$ is continuous and absolutely integrable. Also $f(y) \to 0$, as $|y| \to \infty$.

(ii) There exists $p \geq 1$, such that $f^{(s+p)}(\cdot)$ is uniformly bounded and continuous.

Assumption B. (Assumptions on kernel K). $K(\cdot)$ is square integrable and is of sth order, that is $\int K(x)dx = 1$ and there exists an integer $s \geq 1$ such that $\int K(x)x^j dx = 0$, $j = 1, 2, .., s-1$, $0 < \left|\int K(x)x^s dx\right| \leq \int |K(x)x^s| dx < \infty$. The number s will be called the *order* of the kernel.

Further we assume that $\int |K(x)x^{s+p}| dx < \infty$, where p is the integer for which A(ii) holds.

Assumption C. (Assumption on auxiliary kernel K^0).

(i) $K^0(\cdot)$ is a square integrable probability density function and $\int [K^0(y)]^4 dy < \infty$. Further

(a) $K^0(\cdot)$ is continuous and uniformly bounded.

(b) $K^0(x) \to 0$ as $|x| \to \infty$.

(ii) $K^0(\cdot)$ has s continuous derivatives on $(-\infty, \infty)$ and its sth derivative $K^{0(s)}(\cdot)$, satisfies the above conditions (a) and (b) and the following assumptions

(c) $\int |K^{0(s)}(x)| dx < \infty$ and $\int [K^{0(s)}(y)]^4 dy < \infty$

(d) $\int K^{0(s)}(x)x^j dx = 0$, where $0 \leq j \leq s+p-1, j \neq s$,
$\frac{(-1)^s}{s!}\int K^{0(s)}(x)x^s dx = 1$ and $\int |K^{0(s)}(x)x^{s+p}|dx < \infty$.

For all asymptotic results (as $n \to \infty$), it is understood that λ, μ, $h \to 0$, $n\lambda$, $n\mu$, $nh \to \infty$.

Remark 3.2.1. The number p, in Assumptions A and C, depends on K^0. If K^0 is standard normal density then we recommend $p = 2$. With this choice of K^0 and p, Assumption C is satisfied for any value of s.
(ii) The Assumptions $A(i) - (ii)$ on f, are valid for a wide class of densities which include mixed normal, cauchy, beta(m,n) (m,n>2) and gamma(n) (n>2) among others. Whereas the assumption that f has compact support or the assumption $E(|X_1|^\epsilon) < \infty, \epsilon > 0$ (see page 184, Hall (1990)) precludes the mixed normal distributions or the heavy tailed distributions which have no moments.
(iii) Assumption B on K is quite common in density estimation context and does not limit the choice of K. In contrast the assumptions by Hall (1990) and Falk (1992) prevent the use of a number of popular kernels e.g. the Gaussian or Gaussian type kernel, as they are not compactly supported.

3.3 Main results

We now state our main results. The proofs are given at the end of the chapter.

Theorem 3.3.1. *Suppose Assumptions A–C hold, and* $f^{(s)}(y) \neq 0$ *and* $\lambda = Oe(\frac{1}{n^{1/(2s+2p+1)}})$. *Then*

$$r_1 = E\left[\frac{B_y^*}{B_y} - 1\right]^2 = O\left(\frac{1}{n^{2p/(2s+2p+1)}}\right).$$

Theorem 3.3.2. *Let Assumptions* $A - C$ *hold,* $s \geq 2$, $10 > p \geq 2$, $\mu = Oe(n^{-1/5})$ *and* $\lambda = Oe(\frac{1}{n^{1/(2s+2p+1)}})$.

(i) If $f(y) > 0$, $f^{(s)}(y) \neq 0$ *and* $\liminf_{n\to\infty} nh^{2s+1} > 0$ *then,*

$$r_3 = E\left[\frac{M_y^*}{M_y} - 1\right]^2 = O\left(\frac{1}{n^{2p/(2s+2p+1)}}\right).$$

(ii) If $\limsup_{n\to\infty} nh^{2s+1} = 0$ *and* $f(y) > 0$*, then,*

$$r_3 = E\left[\frac{M_y^*}{M_y} - 1\right]^2 o\left(\frac{1}{n^{2p/(2s+2p+1)}}\right).$$

Remark 3.3.1. (i) If K is a second order kernel (i.e. $s = 2$) then, under Assumptions $A - C$, Theorems 3.3.1 and 3.3.2 hold whenever y is a mode or an anti-mode. Whether a point y is a mode or an anti-mode may be statistically tested using SiZer (Chaudhuri and Marron (1999)) which is a tool for detecting the points of "zero crossings" of $f^{(1)}$.
(ii) If $n\lambda^{2s+1} \to \infty$ and $n\lambda^{2s+2p+1} \to 0$ then under H_0, $\frac{a_n[K_n^{0(s)}(y) - f^{(s)}(y)]}{\sqrt{\int [K^{0(s)}(u)]^2 du . K_n(y)}} \xrightarrow{\mathcal{D}} N(0,\ 1)$ where $a_n = \sqrt{n\lambda^{2s+1}}$.This may be used to test $H_0 : f^{(s)}(y) = 0$ against $H_1 : f^{(s)}(y) \neq 0$.

3.3.1 Comparison with plug in estimator

Computing B_y^* and M_y^* may require Monte-carlo simulation, the plug-in estimators are easier to implement. So a natural question is under what conditions bootstrap estimators are worth the extra computational.effort? The next three theorems provide conditions under which the bootstrap estimators will have infinite asymptotic accuracy (in L_2 sense) compared to their plug-in counterparts.

Theorem 3.3.3. *Suppose Assumptions A–C hold,* $f^{(s+1)}$ *is continuous,* $||f^{(s+1)}|| < \infty$*,* $\lambda = Oe\left(\frac{1}{n^{1/(2s+2p+1)}}\right)$*,* $f^{(s+1)}(y) \neq 0$ *and* $\limsup_{n\to\infty} n^{p/(2s+2p+1)}h = \infty$*. Then*

$$\frac{r_1}{r_4} = \frac{E\left[\frac{B_y^*}{B_y} - 1\right]^2}{E\left[\frac{B_y^A}{B_y} - 1\right]^2} = o(1).$$

Remark 3.3.2. For $s, p \geq 2$, $\limsup_{n\to\infty} n^{1/(2s+1)}h = \infty$ implies $\limsup_{n\to\infty} n^{p/(2s+2p+1)}h = \infty$. So Theorem 3.3.3 holds for second order kernel and super-optimal h for which B_y can be high. Hence B_y^* is expected to be more accurate than B_y^A in the high bias region.

Theorem 3.3.4. *Suppose Assumptions* $A - C$ *hold,* $\limsup_{n\to\infty} n^{2/5}h = \infty$*,* $\mu = Oe\left(\frac{1}{n^{1/5}}\right)$*,* $f(y) > 0$*,* $f^{(2)}(y) \neq 0$ *and* $f^{(1)}(y)\int K^2(u)du \neq f^2(y)$*. Then*

$$\frac{r_2}{r_5} = \frac{E\left[\frac{V_y^*}{V_y} - 1\right]^2}{E\left[\frac{V_y^A}{V_y} - 1\right]^2} = o(1).$$

Remark 3.3.3. The condition $f^{(1)}(y)\int K^2(u)du \neq f^2(y)$ is automatically satisfied whenever y is a mode or anti-mode and $f(y) > 0$.

Theorem 3.3.5. *Suppose $s \geq 2$, $10 > p \geq 2$, $\liminf_{n\to\infty} n^{1/(2s+1)}h > 0$ and $h = O\left(\frac{1}{n^{p/(2s+2p+1)}}\right)$. Suppose further that Assumptions $A - C$ hold, $|\int K(u)u^{s+1}du| < \infty$, $\|f^{(s+1)}\| < \infty$, $\lambda = Oe\left(\frac{1}{n^{1/(2s+2p+1)}}\right)$, $\mu = Oe\left(\frac{1}{n^{1/5}}\right)$, $f^{(2)}(y) \neq 0$ and $f^{(1)}(y)\int K^2(u)du \neq f^2(y)$. Then*

$$\frac{r_3}{r_6} = \frac{E\left[\frac{M_y^*}{M_y} - 1\right]^2}{E\left[\frac{M_y^A}{M_y} - 1\right]^2} = o(1).$$

Remark 3.3.4. (i) In addition to the conditions in Theorem 3.3.5, if we further assume that $f^{(s+1)}(y) = 0$, then condition $h = O\left(\frac{1}{n^{p/(2s+2p+1)}}\right)$ can be replaced by a more general condition $h^2 = O\left(\frac{1}{n^{p/(2s+2p+1)}}\right)$ which is satisfied by $h = Oe\left(\frac{1}{n^{1/(2s+1)}}\right)$ for $p = 2$. So for $p = 2$, Theorem 3.3.5 holds for $h = Oe\left(\frac{1}{n^{1/(2s+1)}}\right)$. The values of h, which are constant multiples of $\frac{1}{n^{1/(2s+1)}}$, have been of great interest in density estimation from the perspective of minimising M_y asymptotically (see Hall (1990), Falk (1992)).

(ii) If y is a mode or anti-mode then the conditions $f^{(2)}(y) \neq 0$ and $f^2(y) \neq f^{(1)}(y)\int K^2(u)du$ are automatically satisfied. So whenever y is mode or an anti-mode, Theorem 3.3.5 ensures that M_y^* has infinite asymptotic accuracy in comparison M_y^A.

3.3.2 Fixed Sample performance of bootstrap estimators:

The following proposition provides some indication of the performance of the bootstrap and plug-in estimators for fixed sample sample size n.

Theorem 3.3.6. *Let the Assumption A on f hold. Also let K^0 be uniformly bounded, continuous and $K^0(y) \to \infty$, as $|y| \to \infty$. Then for fixed sample size n and for any choice of λ, $\mu \neq h$, as $h \to \infty$,*

(i) $B_y \to -f(y)$, $B_y^* \to -K_n^0(y)$ and $B_y^A \to \infty$ almost surely.

(ii) $V_y \to 0$ and $V_y^*, V_y^A \to 0$ almost surely.

(iii) $M_y \to f^2(y)$, $M_y^* \to [K_n^0(y)]^2$ and $M_y^A \to \infty$ almost surely

Remark 6 Thus for any sample size the bootstrap estimators successfully imitate the behaviors of B_y, V_y and M_y, for large value of h. But the asymptotic estimators fail to mimic the behavior of B_y and M_y for large values of h.

3.3.3 Gaussian kernel

B_y^*, V_y^* and M_y^* do not have a closed form expressions in general and hence Monte-Carlo computation is required for its implementation. However we observe that if K is a Gaussian kernel and K^0 is chosen to be the standard normal density then we can obtain closed form expression for the proposed bootstrap estimators. This follows from the following Theorem.

Theorem 3.3.7. *If $g(x) = \sum_{i=1}^k w_i \phi_{\sigma_i^2}(x - \mu_i)$, where $\phi_{\sigma_i^2}(\cdot)$ is the density of $N(0, \sigma_i^2)$ distribution and ϕ is the $N(0,1)$ density then*

$$\int \phi(u) g(x - \sigma u) du = \sum_{i=1}^k w_i \phi_{\sigma_i^2 + \sigma^2}(x - \mu_i).$$

$$\int \phi^2(u) g(x - \sigma u) du = \frac{1}{2\sqrt{\pi}} \sum_{i=1}^k w_i \phi_{\sigma_i^2 + \frac{\sigma^2}{2}}(x - \mu_i).$$

If K^0 is chosen to be $N(0,1)$ density, then

$$K_n^0(y) = \frac{1}{\sqrt{2\pi} n \lambda} \sum_{i=1}^n e^{-\frac{(y - X_i)^2}{2(\lambda)^2}} \text{ and } K_n^*(y) = \frac{1}{\sqrt{2\pi} n \mu} \sum_{i=1}^n e^{-\frac{(y - X_i)^2}{2(\mu)^2}}.$$

K_n^0 and K_n^* are densities of the form $\sum_{i=1}^n w_i \phi_{\sigma_i^2}(x - \mu_i)$, where $w_i = \frac{1}{n}$, $\mu_i = X_i$ and $\sigma_i^2 = (\lambda)^2$ or $(\mu)^2$, $i = 1, 2, .., n$. Therefore if K is a Gaussian kernel then using Theorem 3.3.7 we can easily obtain closed form expression for B_y^*, V_y^* and M_y^*

Theorem 3.3.8. *If both K and K^0 are densities of $N(0,1)$ distribution then*

$$B_y^* = \frac{1}{n} \sum_{i=1}^n \phi_{h_n^2 + \lambda_n^2}(y - X_i) - K_n^0(y),$$

$$V_y^* = \frac{1}{2n^2 h \sqrt{\pi}} \sum_{i=1}^n \phi_{(\mu)^2 + \frac{h_n^2}{2}}(y - X_i) - \frac{1}{n}\left[\frac{1}{n} \sum_{i=1}^n \phi_{h_n^2 + (\mu)^2}(y - X_i)\right]^2$$

and $M_y^* = V_y^* + (B_y^*)^2$.

If the underlying distribution is assumed to be mixed normal distribution, then for Gaussian kernel K we can also obtain a closed form expression for B_y, V_y and M_y.

Theorem 3.3.9. *If $f(x) = \sum_{i=1}^{k} w_i \phi_{\sigma_i^2}(x - \mu_i)$, where $\phi_{\sigma_i^2}(\cdot)$ is the density of $N(0, \sigma_i^2)$ distribution and K is the density of $N(0,1)$ distribution then*

$$B_y = \sum_{i=1}^{k} w_i \phi_{h_n^2 + \sigma_i^2}(y - \mu_i) - f(y),$$

$$V_y = \frac{1}{2nh\sqrt{\pi}} \sum_{i=1}^{k} w_i \phi_{\sigma_i^2 + \frac{h_n^2}{2}}(y - \mu_i) - \frac{1}{n}\left[\sum_{i=1}^{k} w_i \phi_{h_n^2 + \sigma_i^2}(y - \mu_i)\right]^2$$

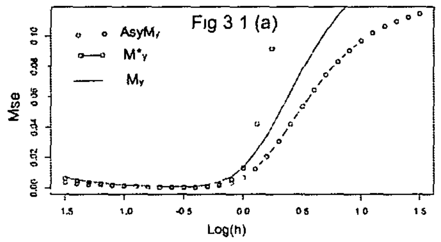
$$\text{and} \quad M_y = V_y + (B_y)^2.$$

3.4 Simulation

We investigated by means of simulations, the effect of y and h on the performance of M_y^* and M_y^A for fixed sample size. Since any density may be approximated arbitrarily closely in various senses by a normal mixture density (see Marron and Wand (1992)), we chose f to be mixed normal. We chose K to be Gaussian due to its wide popularity. Note that a kernel density estimator is not that sensitive to the choice of the kernel. We also chose K^0 to be Gaussian and hence closed form expression for computing M_y^* and M_y are available from Theorems 3.3.8 and 3.3.9. Since K and K^0 are standard normal density, $s = 2$ and $p = 2$. Further we chose $\mu = n^{-1/5}$ and $\lambda = n^{-1/(2s+2p+1)}$.

For $n = 500$, we have plotted M_y^*, M_y^A and M_y against $\log_{10} h$ taking f to be standard normal, bimodal, skewed and kurtic densities. Formulae of these densities are available in Marron and Wand (1992). Our choice of $\log_{10}$ scale is motivated by its use by Marron and Wand (1992). The figures are given in the next page.

In Figures 3.1(a) – 3 4(c) we plot $M_{ny} \equiv M_y(h)$, $M^*_{ny} \equiv M^*_y(h)$ and $AsyM_{ny} \equiv M^A_y(h)$ against $\text{Log}(h) \equiv \log_{10} h$ for normal, bimodal, skewed and kurtic distributions and for sample size $n = 50,\ 200$ Both K and K^0 are standard normal densities



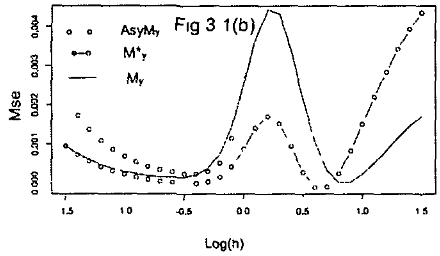
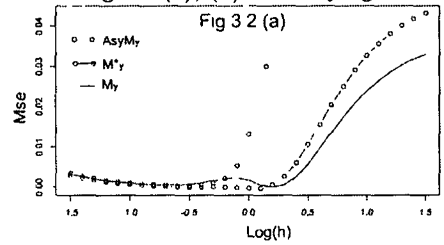


Fig 3 1 (a), (b) underlying distribution"normal"; y=0, -2; n=500.



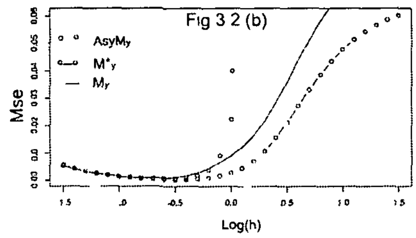
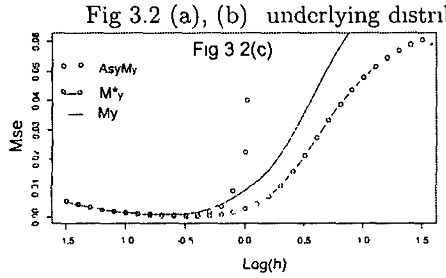


Fig 3.2 (a), (b) underlying distribution"Bimodal"; y=0, 1; n=500.



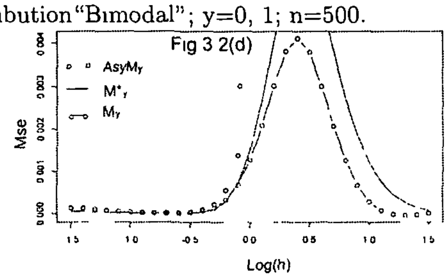
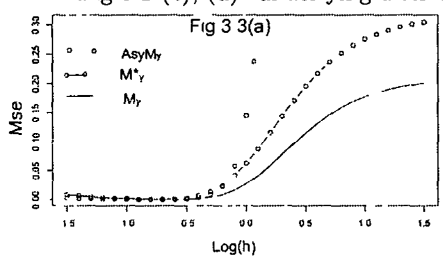


Fig 3.2 (c), (d): underlying distribution"Bimodal"; y=-1, -3, n=500.



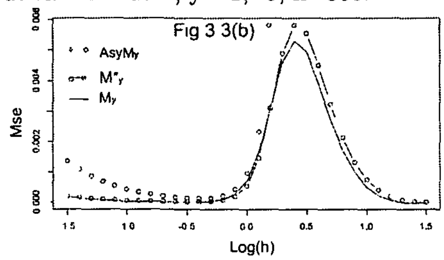


Fig 3.3 (a), (b): underlying distribution"Skewed"; y=1, -2; n=500.

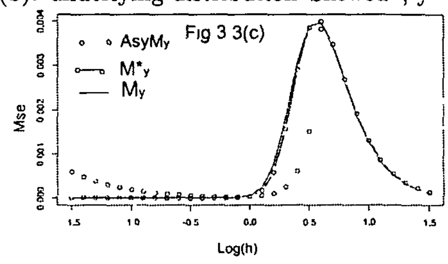


Fig 3.3 (c): underlying distribution "Skewed"; y=-3; n=500.

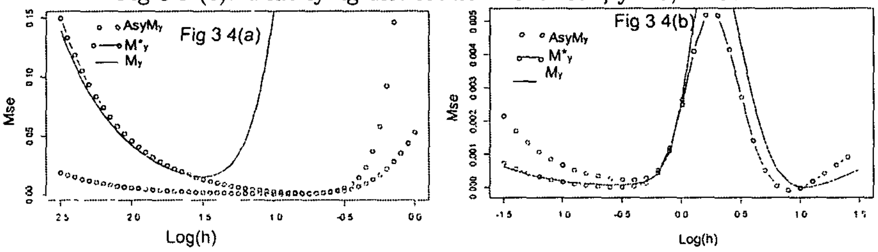


Fig 3.4 (a), (b): underlying distribution"Kurtic"; y=0, -2; n=500.

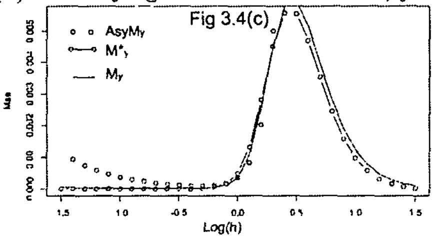


Fig 3.4 (c): underlying distribution"Kurtic"; y=-3; n=500.

The following conclusions may be drawn from these simulations.

(i) Figures 3.1(b), 3.2(d), 3.3(b, c) and 3.4(b, c) reveal that when y is in the tail region, M_y may have more than one minima and M_y^* captures all important features (including multiple minima) of M_y as a function of h. However M_y^A always has one minima irrespective of y and it fails to mimic M_y specially when h is large and y is in the tail region.

(ii) Figures 3.1(a), 3.2(a, b, c) and 3.3(a) consider the case when y is a mode or anti-mode of f. These reveal that both M_y^A and M_y^* successfully imitate M_y when $\log_{10}(h) \leq -0.5$. However when $\log_{10}(h) \geq 0$, M_y^A increases rapidly whereas both M_y^* and M_y first increase and then appear to level-off, as the value of h is increased.

However, in general we noticed that if f is a mixed normal, $f(x) = \sum_{i=1}^{k} w_i \phi_{\sigma_i^2}(x - \mu_i)$ and if a particular σ_i^2 is small and the corresponding w_i is not too small, then both M_y^* and M_y^A are poor estimators of M_y at $y = \mu_i$ for $\log_{10}(h) \geq -1.5$. For example, in Figure 3.4(a) we have plotted the result when f is the density of $\frac{2}{3}N(0,1) + \frac{1}{3}N(0, \frac{1}{10^2})$ and $y = 0$. It is to be noted that M_y^* continues to accurately estimate M_y if $\log_{10}(h) \leq -1.5$.

(iii) From the perspective of estimating h_y we see that both h_y^A and h_y^* perform equally well when y is a mode or an anti-mode.

But if y is in the tail region of f, then from Figures 3.1(b), 3.2(d), 3.3 (b, c) and 3.4(b, c) we see that M_y attains two minima, one in the range $\log_{10}(h) \leq 0$ and the other in the range $\log_{10}(h) \geq 0.5$ and the global minima need not be unique. This feature is successfully imitated by M_y^*. Further, the global minimizer of M_y^* also need not be unique. In any case, the larger minima always turn out to be the global minima and hence without loss h_y and h_y^* are taken to be the largest values of h minimizing M_y and M_y^* respectively. It also turns out that they are close. The above choice of h_y^* and h_y are also supported by Sain and Scott's (1995) observation that the sequence h_y can converge to a positive constant, rather than zero. In contrast, M_y^A has a unique minimizer h_y^A and $\log_{10}(h_y^A) < 0$, irrespective of where y is. In conclusion, if y is in the tail, h_y^A is a poor estimate for h_y whereas h_y^* is close to h_y.

(iv) Hall (1990) showed that the minimizers of M_y^H and M_y, with respect to h over $A = [\frac{\epsilon}{n_1^{1/2s+1}}, \frac{\epsilon^{-1}}{n_1^{1/2s+1}}]$ (n_1 is the resample size, $0 < \epsilon < 1$), are asymptotically equivalent (almost surely). The results of Falk (1992) imply that, for a second order kernel, minimizing M_{ny}^F, as a function of h over $B = \left[\frac{C_1}{n^{1/5}}, \frac{C_2}{n^{1/5}}\right]$, is asymptotically equivalent to minimizing $\frac{f(y)}{cn^{4/5}} \int K^2 + (\frac{c^2}{2n^{2/5}} f^{(2)}(y) \int x^2 K(x)dx)^2$ with respect to c over $[C_1, C_2]$. Let h_y^H denote the minimizer of M_y^H over A and h_y^F be the minimizer of M_y^F over B. For second order kernel, h_y^F and h_y^H are $Oe(\frac{1}{n^{1/5}})$ and $Oe(\frac{1}{n_1^{1/5}})$, where $n_1 < n$ and $n_1 \to \infty$. Consequently for large n, both h_y^F and h_y^H are expected to be close to zero and can be much smaller than h_y when y is in the tail. In fact for such y, both h_y^F, h_y^H can be expected to be closer to the smaller local minima of M_y. Thus h_y^* appears to be a more appropriate estimator of h_y than h_y^F and h_y^H.

3.5 Proofs

3.5.1 Some important Lemmas

We state and prove the lemmas which have been used in the proofs of the Theorems.

Lemma 3.5.1. *Under Assumptions A on f and C(ii) on $K^{0(s)}$ and choosing* $\lambda = Oe\left(\frac{1}{n^{1/(2s+2p+1)}}\right)$,

(i) $\sup_{-\infty<y<\infty} E\left[K_n^{0(s)}(y) - f^{(s)}(y)\right]^2 = O\left(\frac{1}{n^{2p/(2p+2s+1)}}\right)$.

(ii) $\sup_{-\infty<y<\infty} E\left[K_n^{0(s)}(y) - f^{(s)}(y)\right]^4 = O\left(\frac{1}{n^{4p/(2p+2s+1)}}\right)$.

Proof of Lemma 3.5.1 (i) Recall that $K_n^{0(s)}(y) = \frac{1}{n\lambda^{s+1}} \sum_{i=1}^{n} K^{0(s)}\left(\frac{y-X_i}{\lambda}\right)$.

$$E\left[K_n^{0(s)}(y) - f^{(s)}(y)\right]^2 = Var\left(K_n^{0(s)}(y)\right) + \left[E\left(K_n^{0(s)}(y)\right) - f^{(s)}(y)\right]^2.$$

It is easy to verify that

$$Var\left(K_n^{0(s)}(y)\right) \leq \frac{1}{n\lambda^{2s+1}}||f|| \int \left[K^{0(s)}(u)\right]^2 du, \ \forall y.$$

and under Assumption C on $K^{0(s)}$ it is easy to verify that

$$\left[E\left[K_n^{0(s)}(y)\right] - f^{(s)}(y)\right]^2 \leq \lambda^{2p}\left[\frac{||f^{(s+p)}||}{(s+p)!}\int K^{0(s)}(u)u^{s+p}du\right]^2, \quad \forall y.$$

Therefore for $\lambda = Oe\left(\frac{1}{n^{1/(2s+2p+1)}}\right)$, Lemma 3.5.1 (i) is an immediate consequence of the above inequalities. □

Proof of Lemma 3.5.1 (ii) Using $(a+b)^4 \leq 8(a^4+b^4)$ we see that

$$E\left[K_n^{0(s)}(y) - f^{(s)}(y)\right]^4 \leq 8E\left[K_n^{0(s)}(y) - E\left[K_n^{0(s)}(y)\right]\right]^4 + 8\left[E\left[K_n^{0(s)}(y)\right] - f^{(s)}(y)\right]^4 \tag{3.5.1}$$

Let $Y_{ni} = \frac{1}{n\lambda^{s+1}}\left\{K^{0(s)}\left(\frac{y-X_i}{\lambda}\right) - E\left[K^{0(s)}\left(\frac{y-X_i}{\lambda}\right)\right]\right\}, i = 1, 2, ..., n$. Then $Y_{n1}, Y_{n2}, ..., Y_{nn}$ are i.i.d random variables and $E(Y_{n1}) = 0$. Therefore we get the following equation

$$E\left[K_n^{0(s)}(y) - E\left(K_n^{0(s)}(y)\right)\right]^4 = E\left(\sum_{i=1}^{n} Y_{ni}\right)^4 = nE(Y_{n1}^4) + 6n(n-1)\left[E(Y_{n1}^2)\right]^2. \tag{3.5.2}$$

Now

$$\begin{aligned}
E(Y_{n1}^4) &\leq \frac{8}{n^4\lambda^{4s+4}}\left\{E\left[K^{0(s)}\left(\frac{y-X_1}{\lambda}\right)\right]^4 + \left[EK^{0(s)}\left(\frac{y-X_1}{\lambda}\right)\right]^4\right\} \\
&\leq \frac{8}{n^4\lambda^{4s+4}}\left[||f||\lambda\int[K^{0(s)}(v)]^4dv + ||f||^4\lambda^4\left[\int K^{0(s)}(v)dv\right]^4\right] \\
&= \frac{C_1}{n^4\lambda^{4s+3}}\{1 + C_2\lambda^3\}, \text{where } C_1, C_2 \text{ are positive constants}
\end{aligned}$$

and

$$E(Y_{n1}^2) \leq \frac{1}{n^2\lambda^{2s+2}} E\left[K^{0(s)}\left(\frac{y-X_1}{\lambda}\right)\right]^2 \leq \frac{1}{n^2\lambda^{2s+1}} ||f|| \int \left[K^{0(s)}(v)\right]^2 dv$$

Substituting the above inequalities in equation (3.5.2) and using $n\lambda \to \infty$ we get

$$E\left[K_n^{0(s)}(y) - E\left[K_n^{0(s)}(y)\right]\right]^4 \leq \frac{6}{n^2\lambda^{4s+2}}\left[||f|| \int \left[K^{0(s)}(v)\right]^2 dv\right]^2 + o\left(\frac{1}{n^2\lambda^{4s+2}}\right), \ \forall y.$$

Therefore for $\lambda = Oe\left(\frac{1}{n^{1/(2s+2p+1)}}\right)$, we get

$$\sup_{-\infty<y<\infty} E\left[K_n^{0(s)}(y) - E\left[K_n^{0(s)}(y)\right]\right]^4 = O\left(\frac{1}{n^{4p/(2s+2p+1)}}\right). \tag{3.5.3}$$

Further

$$\left[E\left[K_n^{0(s)}(y)\right] - f^{(s)}(y)\right]^4 \leq \left[E\left[K_n^{0(s)}(y) - f^{(s)}(y)\right]^2\right]^2, \ \forall y.$$

Therefore for $\lambda = Oe\left(\frac{1}{n^{1/(2s+2p+1)}}\right)$ and using Lemma 3.5.1 (i), we get

$$\sup_{-\infty<y<\infty} \left[E\left[K_n^{0(s)}(y)\right] - f^{(s)}(y)\right]^4 = O\left(\frac{1}{n^{4p/(2s+2p+1)}}\right). \tag{3.5.4}$$

Substituting the equations (3.5.3) and (3.5.4) in the right side of equation (3.5.1) we get Lemma 3.5.1(ii). So Lemma 3.5.1 is proved completely. □

In order to prove Theorem 3.3.2 we need the following Lemma.

Lemma 3.5.2. *Under Assumptions A – C and choosing* $\lambda = O(\frac{1}{n^{1/(2s+2p+1)}})$ *we get*

$$E\left[(B_y^*)^2 - B_y^2\right]^2 = O\left(\frac{h^{4s}}{n^{2p/(2s+2p+1)}}\right).$$

Proof of Lemma 3.5.2 Recalling the formulae of B_y^* and B_y, from the proof of Theorem 3.3.1, we see that

$$\begin{aligned}\left|\left(B_y^*\right)^2 - \left(B_y\right)^2\right| &= \frac{h^{2s}}{((s-1)!)^2}\Bigg| \left\{\int K(u)u^s \int_0^1 (1-t)^{s-1} K_n^{0(s)}(y-thu)dtdu\right\}^2 \\ &\qquad - \left\{\int K(u)u^s \int_0^1 (1-t)^{s-1} f^{(s)}(y-thu)dtdu\right\}^2 \Bigg| \\ &\leq \frac{h^{2s}}{((s-1)!)^2}\Bigg[\int |K(u)u^s| \int_0^1 (1-t)^{s-1} \left|K_n^{0(s)}(y-thu) + f^{(s)}(y-thu)\right| dtdu. \\ &\qquad \int |K(u)u^s| \int_0^1 (1-t)^{s-1} \left|K_n^{0(s)}(y-thu) - f^{(s)}(y-thu)\right| dtdu\Bigg] \\ &= \frac{h^{2s}}{((s-1)!)^2}\, c_{1n}.c_{2n} \quad \text{(say)}.\end{aligned}$$

It is easy to see that

$$0 \leq c_{1n}.c_{2n} \leq c_{2n}^2 + \frac{||f^{(s)}||C}{s} c_{2n}$$

where $C = \int |K(u)u^s|$. Therefore

$$\frac{1}{h_n^{4s}} E\left[(B_y^*)^2 - (B_y)^2\right]^2 \leq \frac{2}{((s-1)!)^4}\left[E(c_{2n}^4) + (C')^2 E(c_{2n}^2)\right] \tag{3.5.5}$$

where $C' = \frac{||f^{(s)}||C}{s}$. Further using Cauchy–Schwartz inequality for c_{2n}^2 and c_{2n}^4 and taking expectation we get

$$E(c_{2n}^{2j}) \leq \frac{C^{2j}}{s^{2j}} \sup_{-\infty<y<\infty} E\left[K_n^{0(s)}(y) - f^{(s)}(y)\right]^{2j} \quad j = 1, 2.$$

From Lemma 3.5.1 (i) and (ii), we see that for $\lambda = O\left(\frac{1}{n^{1/(2s+2p+1)}}\right)$

$$\sup_{-\infty<y<\infty} E\left[K_n^{0(s)}(y) - f^{(s)}(y)\right]^{2j} = O\left(\frac{1}{n^{2jp/(2s+2p+1)}}\right), \quad j = 1, 2.$$

Therefore from equation (3.5.5), we see that

$$\frac{1}{h_n^{4s}} E\left[(B_y^*)^2 - (B_y)^2\right]^2 = O\left(\frac{1}{n^{2p/(2s+2p+1)}}\right).$$

Hence Lemma 3.5.2 is proved completely. □

To prove Theorem 3.3.5 we need the following Lemma.

Lemma 3.5.3. *Let $s \geq 2$, $10 > p \geq 2$. Under Assumptions A–C and further assuming $\left|\int K(u)u^{s+1}du\right| < \infty$, $||f^{(s+1)}|| < \infty$, $\liminf_{n\to\infty} nh^{2s+1} > 0$, $h = O\left(n^{-p/(2s+2p+1)}\right)$ and for $\mu = Oe(n^{-1/5})$, $\lambda = Oe(n^{-1/(2s+2p+1)})$ we get*

(i) if $f^2(y) \neq f^{(1)}(y)\int K^2(u)du$ *then* $\quad r_7 = \frac{|E[(B_y^A)^2] - B_y^2|}{|E[V_y^A] - V_y|} = o(1).$

(ii) for $f^{(2)}(y) \neq 0 \quad r_8 = \frac{\left[E[(B_y^*)^2] - B_y^2\right]^2}{E\left[V_y^* - V_y\right]^2} = o(1).$

Proof of Lemma 3.5.3(i) Recalling the definition of B_y^A we see that

$$E[(B_y^A)^2] = \frac{h^{2s}}{(s!)^2}\left[\int K(u)u^s du\right]^2 E\left[K_n^{0(s)}(y)\right]^2.$$

Under Assumptions A on f and C on $K^{0(s)}$, it is easy to verify that

$$E\left[K_n^{0(s)}(y)\right]^2 = \frac{1}{n\lambda^{2s+1}}\int [K^{0(s)}(u)]^2 f(y-\lambda u)du$$
$$+\frac{(n-1)}{n}\left[f^{(s)}(y)+\frac{(-1)^{s+p}\lambda^p}{(s+p-1)!}\int K^{0(s)}(u)u^{s+p}\int(1-t)^{s+p-1}f^{(s+p)}(y-t\lambda u)dtdu\right]^2$$
$$= \frac{1}{n\lambda^{2s+1}}\int [K^{0(s)}(u)]^2 f(y-\lambda u)du$$
$$+\frac{(n-1)}{n}\left[f^{(s)}(y)+\frac{(-1)^{s+p}\lambda^p}{(s+p)!}\int K^{0(s)}(u)u^{s+p}du f^{(s+p)}(y)+o\left(\lambda^p\right)\right]^2.$$

Therefore

$$E[(B_y^A)^2] = C_1^2\frac{h^{2s}}{n\lambda^{2s+1}}\int [K^{0(s)}(u)]^2 f(y-\lambda u)du + \frac{(n-1)}{n}C_1^2 h^{2s}[f^{(s)}(y)]^2$$
$$+\frac{(n-1)}{n}C_2 h^{2s}\lambda^p f^{(s)}(y)f^{(s+p)}(y)+o\left(h^{2s}\lambda^p\right) \qquad (3.5.6)$$

where $C_1 = \frac{\int K(u)u^s du}{s!}$ and $C_2 = \frac{\int K^{0(s)}(u)u^{s+p}du}{(s+p)!}C_1$.

Further recall (from proof of Theorem 3.3.1) that

$$B_y = \frac{(-h)^s}{(s-1)!}\int K(u)u^s\int_0^1(1-t)^{s-1}f^{(s)}(y-thu)dtdu.$$

Therefore

$$B_y^2 = \frac{h^{2s}}{[(s-1)!]^2}\Big[\frac{\int K(u)u^s}{s}f^{(s)}(y)$$
$$-h\int K(u)u^{s+1}\int_0^1(1-t)^{(s-1)}t\int_0^1 f^{(s+1)}(y-vthu)dvdtdu\Big]^2$$
$$= \frac{h^{2s}}{[(s-1)!]^2}\Big[\frac{\int K(u)u^s}{s}f^{(s)}(y)-h\frac{\int K(u)u^{s+1}}{s}f^{(s+1)}(y)+o(h)\Big]^2.$$

This implies that with $C_3 = C_1\frac{\int K(u)u^{s+1}}{s!}$,

$$B_y^2 = C_1^2 h^{2s}[f^{(s)}(y)]^2 + C_3 h^{2s+1}f^{(s)}(y)f^{(s+1)}(y)+o\left(h^{2s+1}\right). \qquad (3.5.7)$$

From equations (3.5.6) and (3.5.7) we see that

$$E[(B_y^A)^2]-B_y^2 = C_1^2\frac{h^{2s}}{n\lambda^{2s+1}}\int [K^{0(s)}(u)]^2 f(y-\lambda u)du$$
$$+ C_2 h^{2s}\lambda^p f^{(s)}(y)f^{(s+p)}(y) - C_3 h^{2s+1}f^{(s)}(y)f^{(s+1)}(y)$$
$$+ O\left(h^{2s}\left(\lambda^p+h+\frac{1}{n}\right)\right). \qquad (3.5.8)$$

Therefore under Assumption A that f and its higher order derivatives are uniformly bounded,

$h = O\left(\frac{1}{n^{p/(2s+2p+1)}}\right)$ and choosing $\lambda = Oe\left(\frac{1}{n^{1/(2s+2p+1)}}\right)$ we see (from equation (3.5.8)) that

$$\lim_{n\to\infty} \frac{n^{p/(2s+2p+1)}}{h^{2s}} \left|E[(B_y^A)^2] - B_y^2\right| < \infty. \qquad (3.5.9)$$

Recall that $V_y^A = \frac{K_n^*(y)}{nh} \int K^2(u)du$. Therefore, recalling the definition of V_y, we get the following equation

$$\begin{aligned} E\left[V_y^A\right] - V_y &= \frac{\int K^2(u)du}{nh}\left[E(K_n^*(y)) - f(y)\right] + \frac{1}{nh}\int K^2(u)\left[f(y) - f(y-hu)\right]du \\ &\quad + \frac{1}{n}\left(\int K(u)f(y-hu)du\right)^2. \end{aligned}$$

It is easy to verify that,

$$\begin{aligned} E(K_n^*(y)) - f(y) &= O\left(\mu^2\right), \\ \frac{1}{nh}\int K^2(u)\left[f(y) - f(y-hu)\right]du &= -\frac{1}{n}f^{(1)}(y)\int K^2(u)udu + O\left(\frac{h}{n}\right), \\ \frac{1}{n}\left(\int K(u)f(y-hu)du\right)^2 &= \frac{f^2(y)}{n} + O\left(\frac{h}{n}\right). \end{aligned} \qquad (3.5.10)$$

Therefore for $\mu = Oe\left(\frac{1}{n^{1/5}}\right)$, from the above equations we see that

$$\begin{aligned} \left|E\left[V_y^A\right] - V_y\right| &= \frac{1}{n}\left|f^2(y) - f^{(1)}(y)\int K^2(u)udu\right| + O\left(\frac{1}{n^{1+2/5}h} + \frac{h}{n}\right). \\ &= L_4 + O\left(\frac{1}{n^{1+2/5}h} + \frac{h}{n}\right) \quad \text{(say)}. \end{aligned} \qquad (3.5.11)$$

Under assumptions $\liminf_{n\to\infty} nh^{2s+1} > 0$ and $\left|f^2(y) - f^{(1)}(y)\int K^2(u)udu\right| > 0$, we see that

$$\lim_{n\to\infty} \frac{n^{p/(2s+2p+1)}}{h^{2s}} L_4 = \infty.$$

Further for $s \geq 2$, $10 > p \geq 2$, $\liminf_{n\to\infty} nh^{2s+1} > 0$ and $h = O\left(\frac{1}{n^{p/(2s+2p+1)}}\right)$

$$\frac{1}{n^{1+2/5}h} + \frac{h}{n} = o\left(\frac{h^{2s}}{n^{p/(2s+2p+1)}}\right).$$

Therefore for $s \geq 2$, $10 > p \geq 2$, under assumptions $h = O\left(\frac{1}{n^{p/(2s+2p+1)}}\right)$, $\liminf_{n\to\infty} nh^{2s+1} > 0$ and $f^2(y) \neq f^{(1)}(y)\int K^2(u)udu$, from equation (3.5.11) we see that

$$\lim_{n\to\infty} \frac{n^{p/(2s+2p+1)}}{h^{2s}} \left|E\left[V_y^A\right] - V_y\right| = \infty. \qquad (3.5.12)$$

Lemma 3.5.3(i) is a direct consequence of the equations (3.5.9) and (3.5.12). □

Proof of Lemma 3.5.3(ii) Recalling the definitions of V_y^* and V_y we see that

$$\begin{aligned}
\left|E\left(V_y^*\right) - V_y\right| &= \Bigg|\frac{1}{nh}\int K^2(u)E\left[K_n^*(y-hu) - f(y-hu)\right]du \\
&\quad -\frac{1}{n}\left[E\left(\int K(u)K_n^*(y-hu)du\right)^2 - \left(\int K(u)f(y-hu)du\right)^2\right]\Bigg| \\
&= \left|E[L_1] - E[L_2]\right|,
\end{aligned}$$

where L_1 and L_2 are as defined in the proof of Theorem 2.3.2.

Recalling equation (3.5.23) we see that for $\mu = Oe(n^{-1/5})$

$$E(L_2) = O\left(\frac{1}{n^{1+2/5}}\right).$$

Under the Assumptions A on f, B on K and C on K^0, $s \geq 2$ and choosing $\mu = Oe(n^{-1/5})$ it is easy to verify that

$$E[L_1] = \frac{C' f^{(2)}(y)}{n^{1+2/5}h} + o\left(\frac{1}{n^{1+2/5}h}\right), \text{ where } C' \text{ is a non-zero constant.}$$

Therefore, under the Assumptions A on f, B on K, C on K^0, $s \geq 2$ and $|f^{(2)}(y)| > 0$ and choosing $\mu = Oe(n^{-1/5})$, we see that

$$\liminf_{n\to\infty} n^{1+2/5}h\left|E\left(V_y^*\right) - V_y\right| = |C' f^{(2)}(y)| > 0. \tag{3.5.13}$$

Recalling the definition of r_8 it is easy to see that

$$r_8 = \frac{\left[E[(B_y^*)^2] - B_y^2\right]^2}{E\left[V_y^* - V_y\right]^2} \leq \frac{E\left[(B_y^*)^2 - B_y^2\right]^2}{\left[E[V_y^*] - V_y\right]^2}. \tag{3.5.14}$$

Recalling Lemma 3.5.2 and equations (3.5.13) and (3.5.14), under Assumptions $A - C$ and for $\lambda = O(\frac{1}{n^{1/(2s+2p+1)}})$ and $\mu = Oe(n^{-1/5})$, we get

$$r_8 = O\left(\frac{h^{4s+1}n^{1+2/5}}{n^{2p/(2s+2p+1)}}\right) = O(g_n) \text{ (say)}.$$

We see that for $h = O\left(\frac{1}{n^{p/(2p+2s+1)}}\right)$ and $s, p \geq 2$, $g_n = o(1)$. Therefore Lemma 3.5.3(ii) follows immediately from the above equation. So Lemma 3.5.3 is proved completely.□

3.5.2 Proofs of Theorems

Proof of Theorem 3.3.1 Recall that

$$B_y = \int K(u)\left[f(y-hu) - f(y)\right]du \quad \text{and} \quad B_y^* = \int K(u)\left[K_n^0(y-hu) - K_n^0(y)\right]du.$$

For each fixed y and u, expanding $f(y-hu)$ and $K_n^0(y-hu)$ by Taylor's expansion with integral remainder we get

$$B_y = \frac{(-h)^s}{(s-1)!}\int K(u)u^s \int_0^1 (1-t)^{s-1} f^{(s)}(y-thu)dtdu$$

and almost surely

$$B_y^* = \frac{(-h)^s}{(s-1)!}\int K(u)u^s \int_0^1 (1-t)^{s-1} K_n^{0(s)}(y-thu)dtdu.$$

Therefore, almost surely, we get

$$\frac{1}{h^s}\left|B_y - B_y^*\right| \le \frac{1}{(s-1)!}\int |K(u)u^s| \int_0^1 (1-t)^{s-1}\left|f^{(s)}(y-thu) - K_n^{0(s)}(y-thu)\right|dtdu.$$

Squaring and taking expectation on both sides of the above inequality we get

$$\frac{1}{h^{2s}}E\left[B_y - B_y^*\right]^2 \le \frac{C_1^2}{(s!)^2}\sup_{-\infty<y<\infty} E\left[K_n^{0(s)}(y) - f^{(s)}(y)\right]^2 \tag{3.5.15}$$

where $C_1 = \int |v|^s K(v)dv$.

Under the Assumptions A, C on f and K^0, choosing $\lambda = O(\frac{1}{n^{1/(2s+2p+1)}})$, from Lemma 3.5.1(i) we get

$$\sup_{-\infty<y<\infty} E\left[K_n^{0(s)}(y) - f^{(s)}(y)\right]^2 = O\left(\frac{1}{n^{2p/(2s+2p+1)}}\right).$$

substituting the above equation in right side of (3.5.15) we see that

$$\frac{1}{h^{2s}}E\left[B_y - B_y^*\right]^2 = O\left(\frac{1}{n^{2p/(2s+2p+1)}}\right). \tag{3.5.16}$$

Using the smoothness Assumption A on $f^{(s)}$ it is easy to see that

$$\frac{1}{h^{2s}}B_y^2 = \left(\frac{\int K(u)u^s du}{s!}f^{(s)}(y)\right)^2 + o(1) \tag{3.5.17}$$

Since $r_{1n} = \frac{\frac{1}{h^{2s}}E\left[B_y - B_y^*\right]^2}{\frac{1}{h^{2s}}B_y^2}$, therefore under the Assumptions $|\int K(u)u^s du| > 0$ and $|f^{(s)}(y)| > 0$, Theorem 3.3.1 is a direct consequence equations (3.5.16) and (3.5.17). □

Proof of Theorem 3.3.2 Recalling the definitions of M_y, M_y^* and using $(a+b)^2 \leq 2a^2 + 2b^2$ we find that

$$0 \leq E\left[M_y^* - M_y\right]^2 \leq 2E[B_y^2 - (B_y^*)^2]^2 + 2E[V_y - V_y^*]^2. \qquad (3.5.18)$$

Recalling the definitions of V_y and V_y^* we get the following equation

$$\begin{aligned} V_y^* - V_y &= \frac{1}{nh}\int K^2(u)[K_n^*(y-hu) - f(y-hu)]du \\ &\quad -\frac{1}{n}\left[\left(\int K(u)K_n^*(y-hu)du\right)^2 - \left(\int K(u)f(y-hu)du\right)^2\right] \\ &= L_1 - L_2 \quad \text{(say)} \end{aligned}$$

Hence

$$E[V_y^* - V_y]^2 \leq 2E(L_1^2) + 2E(L_2^2). \qquad (3.5.19)$$

Now

$$E(L_1^2) \leq \frac{\left(\int K^2(u)\right)^2}{(nh)^2} \sup_{-\infty<y<\infty} E[K_n^*(y) - f(y)]^2.$$

Further note that

$$0 \leq E\left[f(y) - K_n^*(y)\right]^2 \leq \frac{||f||}{n\mu}\int (K^0(y))^2 dy + \left[\frac{||f^{(2)}||(\mu)^2}{2!}\int K^0(y)y^2 dy\right]^2, \quad \forall y.$$

So choosing $\mu = Oe(n^{-1/5})$, which minimizes the right side of the above equation, we get

$$\sup_{-\infty<y<\infty} E\left[f(y) - K_n^*(y)\right]^2 = O(n^{-4/5}). \qquad (3.5.20)$$

Therefore

$$E(L_1^2) = O\left(\frac{1}{n^{2+4/5}h^2}\right). \qquad (3.5.21)$$

Now using $a^2 - b^2 = (a-b)(a+b)$ and Cauchy-Schwartz inequality it is easy to see that

$$n^2 EL_2^2 \leq E(c_n.d_n)\left[\int |K(u)|du\right]^2, \quad \text{where}$$

$$c_n = \int |K(u)|\left[f(y-hu) + K_n^*(y-hu)\right]^2 du$$

$$\text{and} \quad d_n = \int |K(u)|\left[f(y-hu) - K_n^*(y-hu)\right]^2 du.$$

Since $|f(y-hu)+K_n^*(y-hu)| \leq |f(y-hu)-K_n^*(y-hu)| + 2||f||$, for all y, therefore it is easy to see that $c_n d_n \leq 2d_n^2 + 8||f||^2 d_n$ and hence

$$n^2 E L_2^2 \leq C^2 E(c_n . d_n) \leq C^2 \left[2E(d_n^2) + 8||f||^2 E(d_n)\right] \tag{3.5.22}$$

where $C = \int |K(u)|du$. Further it is easy verify that

$$0 \leq E(d_n^j) \leq C \sup_{-\infty<y<\infty} E[K_n^*(y) - f(y)]^{2j} \quad j = 1, 2.$$

For $\mu = Oe(n^{-1/5})$ recalling equation (3.5.20) we get

$$E(d_n) = O(n^{-4/5}).$$

By some straight forward algebra it is easy to verify that for $\mu = Oe(n^{-1/5})$

$$\sup_{-\infty<y<\infty} E[K_n^*(y) - f(y)]^4 = O\left(\frac{1}{n^{8/5}}\right).$$

Consequently $E(d_n^2) = O\left(\frac{1}{n^{8/5}}\right)$ and hence recalling equation (3.5.22) we get

$$E(L_2^2) = O\left(\frac{1}{n^{2+4/5}}\right) \tag{3.5.23}$$

From equations (3.5.19), (3.5.21) and (3.5.23) we get

$$E[V_y^* - V_y]^2 = O\left(\frac{1}{n^{2+4/5}h^2}\right). \tag{3.5.24}$$

From equations (3.5.18), (3.5.24) and Lemma 3.5.2 it is easy to see that

$$E\left[M_y^* - M_y\right]^2 = O\left(\frac{h^{4s}}{n^{2p/(2s+2p+1)}} + \frac{1}{n^{2+4/5}h^2}\right). \tag{3.5.25}$$

Using Assumption A(i) on f and Assumption B on K it is easy to prove that

$$M_y^2 = \left[\frac{f(y)}{nh}\int K^2(u)du + \frac{h^{2s}}{(s!)^2}\left[\int K(u)u^s f^{(s)}(y)\right]^2 + o\left(\frac{1}{nh} + h^{2s}\right)\right]^2. \tag{3.5.26}$$

Recall that

$$r_3 = \frac{E\left[M_y^* - M_y\right]^2}{M_y^2}. \tag{3.5.27}$$

If $\liminf_{n\to\infty} nh^{2s+1} > 0$ then, under the Assumption $f(y) > 0$ and $|f^{(s)}(y)| > 0$, dividing numerator and denominator of r_3 by h^{4s} we get, from equations (3.5.25), (3.5.26) and (3.5.27), that

$$r_3 = O\left(\frac{1}{n^{2p/(2s+2p+1)}} + \frac{1}{(nh^{2s+1})^2 n^{4/5}}\right) = O\left(\frac{1}{n^{2p/(2s+2p+1)}}\right)$$

using $s \geq 2$ and $10 > p \geq 2$.

If $\limsup_{n\to\infty} nh^{2s+1} = 0$ then, under the Assumption $f(y) > 0$, dividing numerator and denominator of r_3 by $\frac{1}{(nh)^2}$ we get, from equations (3.5.25) and (3.5.27), that

$$r_3 = O\left(\frac{nh^{2s+1}}{n^{2p/(2s+2p+1)}} + \frac{1}{n^{4/5}}\right) = o\left(\frac{1}{n^{2p/(2s+2p+1)}}\right)$$

using $s \geq 2$, $10 > p \geq 2$ and $\limsup_{n\to\infty} nh^{2s+1} = 0$.

So Theorem 3.3.2 is proved completely. □

Proof of Theorem 3.3.3 Recall that

$$B_y^A = \frac{(-h)^s}{s!} K_n^{0(s)}(y) \int K(u) u^s du.$$

and from the proof of Theorem 2.3.1 we see that

$$B_y = \frac{(-h)^s}{(s-1)!} \int K(u) u^s \int_0^1 (1-t)^{s-1} f^{(s)}(y - thu) dt du$$

Using $|a-b| \geq \big||a-c| - |b-c|\big|$, it is easy to see that

$$\frac{1}{h^s} \left|E[B_y^A] - B_y\right| \geq |d_{1n} - d_{2n}|$$

where

$$d_{1n} = \left|\frac{1}{s!} \int K(u) u^s du . E\left[K_n^{0(s)}(y) - f^{(s)}(y)\right]\right|,$$

$$d_{2n} = \left|\frac{1}{(s-1)!} \int K(u) u^s \int_0^1 (1-t)^{s-1} \left[f^{(s)}(y - thu) - f^{(s)}(y)\right] dt du\right|.$$

Now

$$\frac{1}{h^{2s}} E\left[B_y^A - B_y\right]^2 \geq \frac{1}{h^{2s}} \left[E[B_y^A] - B_y\right]^2 \geq \left[d_{1n} - d_{2n}\right]^2$$

Therefore

$$\frac{r_1}{r_4} = \frac{\frac{1}{h^{2s}} \left[E[B_y^*] - B_y\right]^2}{\frac{1}{h^{2s}} \left[E[B_y^A] - B_y\right]^2} \leq \frac{\frac{1}{h^{2s}} E\left[B_y^* - B_y\right]^2}{\left[d_{1n} - d_{2n}\right]^2} \tag{3.5.28}$$

In view of equations (3.5.16) and (3.5.28), to prove Theorem 3.3.3, it is enough to show that

$$\liminf_{n\to\infty} \frac{|d_{1n} - d_{2n}|}{\lambda^p} = \infty, \quad \text{where } \lambda = Oe\left(\frac{1}{n^{1/(2s+2p+1)}}\right). \tag{3.5.29}$$

Using the Assumption A on $f^{(s)}$, $f^{(s+p)}$ and Assumption C (ii) on $K^{0(s)}$ it is easy to see that

$$E\left[K_n^{0(s)}(y) - f^{(s)}(y)\right] = \frac{(-1)^{(s+p)}\lambda^p}{(s+p)!} f^{(s+p)}(y) \int K^{0(s)}(u)u^{s+p}du + o(\lambda^p). \tag{3.5.30}$$

Substituting the above expression for $E\left[K_n^{0(s)}(y) - f^{(s)}(y)\right]$ in the definition of d_{1n} we get

$$d_{1n} = \frac{\lambda^p}{(s+p)!}\left|f^{(s+p)}(y) \int K^{0(s)}(u)u^{s+p}du\right| + o(\lambda^p).$$

So to prove equation (3.5.29) it is enough to show that $\limsup_{n\to\infty} \frac{d_{2n}}{\lambda^p} = \infty$.
Using the Taylor expansion for $f^{(s)}(y-thu) - f^{(s)}(y)$ and the assumption that $f^{(s+1)}(\cdot)$ is bounded, continuous, it is easy to show that

$$d_{2n} = \frac{h}{(s)!}|f^{(s+1)}(y) \int K(u)u^{s+1}du| + o(h).$$

Choosing $\lambda = Oe\left(\frac{1}{n^{1/(2s+2p+1)}}\right)$, using the assumptions $\limsup_{n\to\infty} n^{p/(2s+2p+1)}h = \infty$ and $f^{(s+1)}(y) \neq 0$, it is easy to see that $\limsup_{n\to\infty} \frac{d_{2n}}{\lambda^p} = \infty$ and consequently

$$\limsup_{n\to\infty} \frac{|d_{1n} - d_{2n}|}{\lambda^p} = \infty.$$

This establishes equation (3.5.29) and hence Theorem 3.3.3 is proved completely. □

Proof of Theorem 3.3.4 Recall that

$$V_y^A = \frac{1}{nh} K_n^*(y) \int K^2(u)du$$

and

$$V_y = \frac{1}{nh}\int K^2(u)f(y-hu)du - \frac{1}{n}\left(\int K(u)f(y-hu)du\right)^2.$$

Using $E(X^2) \geq |E(X)|^2$ and $|a-b| \geq \big||a-c| - |b-c|\big|$, it is easy to see that

$$E\left[V_y^A - V_y\right]^2 \geq \left|E[V_y^A] - V_y\right|^2 \geq |e_{1n} - e_{2n}|^2$$

where

$$e_{1n} = \frac{1}{nh}\int K^2(u)du|E(K_n^*(y)) - f(y)|, \quad (3.5.31)$$

$$e_{2n} = \left|\frac{1}{nh}\int K^2(u)\left[f(y-hu) - f(y)\right]du - \frac{1}{n}\left(\int K(u)f(y-hu)du\right)^2\right| \quad (3.5.32)$$

Under the Assumption C on K^0 and $f^{(2)}(y) \neq 0$, choosing $\mu = Oe(n^{-1/5})$ it is easy to show that

$$e_{1n} = Oe\left(\frac{1}{n^{1+2/5}h}\right) \quad (3.5.33)$$

$$\lim_{n\to\infty} ne_{2n} = \left|f^{(1)}(y)\int K^2(u)du - f^2(y)\right| > 0 \text{ (by assumption).}$$

Since $\lim_{n\to\infty} n^{2/5}h = \infty$,we have $\lim_{n\to\infty}\frac{e_{2n}}{e_{1n}} = \infty$.

$$\frac{r_2}{r_5} = \frac{E\left[V_y^* - V_y\right]^2}{E\left[V_y^A - V_y\right]^2} \leq \frac{E\left[V_y^* - V_y\right]^2}{\left[e_{1n} - e_{2n}\right]^2} = \frac{\frac{1}{e_{1n}^2}E\left[V_y^* - V_y\right]^2}{\left[1 - \frac{e_{2n}}{e_{1n}}\right]^2}.$$

Since $\lim_{n\to\infty}\frac{e_{2n}}{e_{1n}} = \infty$, therefore Theorem 3.3.4 is proved completely if we can show that

$$\frac{E\left[V_y^* - V_y\right]^2}{e_{1n}^2} = O(1). \quad (3.5.34)$$

Equation (3.5.34) is a direct consequence of the equations (3.5.24) and (3.5.33). This completes the proof of Theorem 3.3.4. □

Proof of Theorem 3.3.5 Recalling the definitions of r_3 and r_6 we see that

$$\begin{aligned}
\frac{r_3}{r_6} &= \frac{E\left[M_y^* - M_y\right]^2}{E\left[M_y^A - M_y\right]^2} \leq \frac{2E\left[V_y^* - V_y\right]^2 + 2E\left[(B_y^*)^2 - B_y^2\right]^2}{\left[E(M_y^A) - M_y\right]^2} \\
&\leq \frac{2E\left[V_y^* - V_y\right]^2 + 2E\left[(B_y^*)^2 - B_y^2\right]^2}{\left[|E[(B_y^A)^2] - B_y^2| - |E[V_y^A] - V_y|\right]^2} \\
&= \frac{2E\left[V_y^* - V_y\right]^2}{\left[E[V_y^A] - V_y\right]^2}\frac{(1 + r_8)}{(r_7 - 1)^2} \\
&\leq \frac{2E\left[V_y^* - V_y\right]^2}{\left[e_{1n} - e_{2n}\right]^2}\frac{(1 + r_8)}{(r_7 - 1)^2}
\end{aligned} \quad (3.5.35)$$

where e_{1n}, e_{2n} are as defined in the proof of Theorem 3.3.4.

Recalling the proof of Theorem 3.3.4, under Assumptions $A-C$ and $\lim_{n\to\infty} n^{2/5}h = \infty$, we see that

$$\frac{E\left[V_y^* - V_y\right]^2}{\left[e_{1n} - e_{2n}\right]^2} = o(1) \tag{3.5.36}$$

From Lemma 3.5.3(i) and (ii) we see that, under all the assumptions stated in Lemma 3.5.3

$$r_7 = o(1) \text{ and } r_8 = o(1).$$

Therefore Theorem 3.3.5 is a direct consequence of equation (3.5.35), (3.5.36) and Lemma 3.5.3. This completes the proof. □

Proof of Theorem 3.3.6 Recall that $B_y = \int K(u)f(y-hu)du - f(y)$. So under the assumptions (A) that f is uniformly bounded and $f(y) \to 0$, as $|y| \to \infty$, applying D.C.T we see that $\int K(u)f(y-hu)du \to 0$ as $h \to \infty$. So $B_y \to -f(y)$ as $h \to \infty$.

Recall that $B_y^* = \int K(u)K_n^0(y-hu)du - K_n^0(y)$. Under the stated assumptions on K^0, for each fixed n and λ, $K_n^0(y) = \frac{1}{n\lambda}\sum_{i=1}^n K^0\left(\frac{y-X_i}{\lambda}\right)$ is uniformly bounded and $K_n^0(y) = o(1)$, as $|y| \to \infty$, almost surely. So repeating the previous arguments we see that under stated assumptions on K^0, $B_y^* \to -K_n^0(y)$, almost surely, as h is increased.

The third part follows immediately from the definition of B_y^A. So part (i) is proved. To prove part (ii), we note that

$$V_y \le \frac{1}{nh}\int K^2(u)f(y-hu)du \text{ and } V_y^* \le \frac{1}{nh}\int K^2(u)K_n^*(y-hu)du, \text{almost surely.}$$

So under the stated conditions on f and K^0, V_y, $V_y^* = o(1)$, as $h \to \infty$.

Chapter 4

Estimating measures of accuracy of the multivariate product kernel density estimator

4.1 Introduction

In this chapter we address the problems of estimating some local and global measures of accuracy of a multivariate kernel density estimator. Let $X_1, X_2, ..., X_n$ be n i.i.d R^d valued ($d > 1$) random variables with joint density f. A natural generalization of the ordinary kernel density estimator is the *product kernel density estimator* (see for instance Scott and Wand (1991), Jones and Wand (1993)). A *product kernel density estimator* with bandwidths $h_j \equiv h_{nj}$, $j = 1, 2, .., d$, and kernel K is defined as

$$K_n(\vec{y}) = \frac{1}{n \prod_{j=1}^{d} h_j} \sum_{i=1}^{n} \prod_{j=1}^{d} K\left(\frac{y_j - X_{ij}}{h_j}\right)$$

where $\vec{y} = (y_1, y_2, ..., y_d) \in R^d$, $h_j \to 0$, $j = 1, 2, ., d$ and $n \prod_{j=1}^{d} h_j \to \infty$ as $n \to \infty$. The d bandwidths h_j, $j = 1, 2, .., d$, represent the amount of smoothing along d coordinate directions. This estimator was proposed by Epanechnikov (1969).

Scott and Wand (1991), Wand and Jones (1993) provide evidence in favour of having the flexibility to smooth by different amounts independently in each direction.

In fact, the latter authors have demonstrated good reasons for using full covariance matrix parameterized bandwidth. However there are some implementation problems associated with these proposals. For instance, automatic choice of the full bandwidth matrix is computationally quite expensive, see for instance Duong and Hazelton (2003). Moreover in a recent paper, Dutta (2010) has observed that a product kernel density estimate, based on spherical data, compares well with a density estimate using full bandwidth matrix, in terms of capturing both the location and orientation of the modes. So a computationally efficient alternative seems to implement the product kernel density estimate applied to spherical data $X^* = X\ S^{-1/2}$, where X is the data matrix and S is the the sample covariance matrix. The resulting density estimate is defined as $K_n^{**}(\vec{y}) = |S^{-1/2}| K_n^*(\vec{y}\ S^{-1/2})$, $\vec{y} \in R^2$, $|S^{-1/2}|$ denotes the determinant of $S^{-1/2}$ and K_n^* is the usual product kernel density estimate based on X^* (see Dutta (2010)).

Even the proposal of selecting a number of smoothing parameters along different directions can be undesirable, especially when the bandwidths are chosen subjectively and the data dimension is more than two. A simpler option is to choose $h_1 = h_2 = ... = h_d = h$, which is to have equal amount of smoothing along all direction (Cacoullos (1966)). The resulting simple product kernel estimator is defined as follows

$$K_n(\vec{y}) = \frac{1}{nh^d} \sum_{i=1}^{n} \prod_{j=1}^{d} K\left(\frac{y_j - X_{ij}}{h}\right).$$

It has many applications. For instance, it is used in detecting important features such as peaks and valley of a multivariate density by Significance in Scale-Space method (Ghostliness, Marrow, Chauffeur (2000)). Another interesting application of the simple product kernel estimator is in mode estimation of a multivariate density (Abraham, Bias and Cadre (2003)).

The mean squared error $M_{\vec{y}}$, at $\vec{y}$, and mean integrated squared error M are popular measures of local (point-wise) and global (overall) accuracy of $K_n(\cdot)$. These measures reveal important insight into the effect of dimension on accuracy of $K_n(\cdot)$ (see for instance Scott and Wand (1991) and references therein). $M_{\vec{y}}$ and M are also used as criterion for selecting an optimal $\vec{h} = (h_1, h_2, ..., h_d)$. For instance, Scott and Wand (1991) introduced a local measure of accuracy which is referred to as the *sample root coefficient of variation* (denoted by $R_{\vec{y}}$) and it is defined as follows

$$R_{\vec{y}} = \sqrt{Var(K_n(\vec{y}))}/E(K_n(\vec{y})), \vec{y} \in R^d.$$

Scott and Wand (1991) have stated that a small value of this criterion in high dimensions would suggest that widely separated peaks will be identifiable even if estimates are biased downward with a large window width.

In this chapter we propose smooth bootstrap estimators of bias, mean squared error, sample root coefficient of variation and MISE. We study the asymptotic properties of the proposed estimators and provide insight into their finite sample behaviour.

4.2 Literature review

The amount of research on bootstrap estimators of local and global measures of accuracy of multivariate kernel density estimator appear to be quite less in comparison to the amount of work focused on bootstrap methodology for univariate density estimators. Sain, Baggers and Scott (1994) have proposed a multivariate extension of Taylor's (1989) bootstrap estimator. Apparently asymptotic properties of their bootstrap estimator (we call it T) does not appear to be known. No theoretical results on its finite sample properties seem to have been worked out. We investigate both, the asymptotic accuracy as well as finite sample properties the proposed estimators. T turns out to be a special case of our proposed estimator M^* and so our study also lends some insight into properties of T, especially its finite sample behaviour.

We state and prove all the theoretical results for the case when $h_1 = h_2 = ... = h_d = h$ and K is a second order kernel. This restriction simplifies the theoretical calculations involved in our proofs to a great extent, without compromising the generality significantly. Our results can be easily extended to the case where kernel order exceeds 2 and $h_1, ..., h_d$ may not be equal, at the expense of some more complicated theoretical calculations.

4.2.1 Notation and definitions

To simplify the calculations we introduce some notation which are used in the sequel. Let

$$\prod_{i=1}^{d} K(u_i) = \mathbf{K}(\vec{u}),\ \prod_{i=1}^{d} K^0(u_i) = \mathbf{K}^0(\vec{u}),\ \prod_{i=1}^{d} K^2(u_i) = \mathbf{K}^2(\vec{u}) \text{ and} \prod_{i=1}^{d} [K^0(u_i)]^2 = [\mathbf{K}^0(\vec{u})]^2.$$

Further let

$$\int z(\vec{x})d\vec{x} = \int \ldots \int z(x_1, x_2, ..., x_d)dx_1 dx_2 ... dx_d,$$
$$\vec{h}.\vec{u} = (h_1.u_1, h_2.u_2, ..., h_d.u_d),\ \ \vec{h} = (h_1, ..., h_d) \text{ and } \vec{u} \in R^d.$$

Let DCT stands for Dominated Convergence Theorem. For any two positive sequences $\{a_n\}, \{b_n\}$ we write $a_n = Oe(b_n)$ if $0 < \liminf_{n\to\infty} \frac{a_n}{b_n} \leq \limsup_{n\to\infty} \frac{a_n}{b_n} < \infty$.

We denote the bias, variance and mean squared error of $K_n(\vec{y})$ by $B_{\vec{y}}$, $V_{\vec{y}}$ and $M_{\vec{y}}$ respectively. They are defined as follows

$$\begin{aligned}
B_{\vec{y}} &= E\left[K_n(\vec{y})\right] - f(\vec{y}) = \int \mathbf{K}(\vec{u}) f(\vec{y} - \vec{h}.\vec{u}) d\vec{u} - f(\vec{y}), \\
V_{\vec{y}} &= \frac{1}{n\prod_{j=1}^{d} h_j^2} Var\left[\prod_{j=1}^{d} K\left(\frac{y_j - X_{1j}}{h}\right)\right] \\
&= \frac{1}{n\prod_{j=1}^{d} h_j} \int \mathbf{K}^2(\vec{u}) f(\vec{y} - \vec{h}.\vec{u}) d\vec{u} - \frac{1}{n}\left[\int \mathbf{K}(\vec{u}) f(\vec{y} - \vec{h}.\vec{u}) d\vec{u}\right]^2, \\
\text{and } M_{\vec{y}} &= V_{\vec{y}} + B_{\vec{y}}^2.
\end{aligned}$$

Sample root coefficient of variation and mean integrated squared error (MISE) of the are denoted by $R_{\vec{y}}$ and M respectively. They are defined as follows

$$\begin{aligned}
R_{\vec{y}} &= \frac{[V_{\vec{y}}]^{1/2}}{E_{\vec{y}}}, \text{ where } E_{\vec{y}} = E\left[K_n(\vec{y})\right] = \int \mathbf{K}(\vec{u}) f(\vec{y} - \vec{h}.\vec{u}) d\vec{u} \\
\text{and } M &= \int \left[V_{\vec{y}} + B_{\vec{y}}^2\right] d\vec{y} = V + B \text{ (say).}
\end{aligned}$$

Next we define the proposed smooth bootstrap estimators of these parameters. Let $B_{\vec{y}}^*$, $V_{\vec{y}}^*$, $M_{\vec{y}}^*$, $R_{\vec{y}}^*$ and M^* be the proposed estimators of $B_{\vec{y}}$, $V_{\vec{y}}$, $M_{\vec{y}}$, $R_{\vec{y}}$ and M respectively

They are defined as follows

$$\begin{aligned}
B^*_{\vec{y}} &= \int_{\vec{u}} \mathbf{K}(\vec{u}) g(\vec{y} - h\vec{u}) d\vec{u} - g(\vec{y}), \\
V^*_{\vec{y}} &= \frac{1}{n\prod_{j=1}^d h_j} \int \dots \int \prod_{j=1}^d K^2(u_j) w(\vec{y} - h\vec{u}) d\vec{u} - \frac{1}{n}\left[\int_{\vec{u}} \mathbf{K}(\vec{u}) w(\vec{y} - h\vec{u}) d\vec{u}\right]^2, \\
M^*_{\vec{y}} &= V^*_{\vec{y}} + \left[B^*_{\vec{y}}\right]^2, \\
M^* &= \int \left[V^*_{\vec{y}} + \left[B^*_{\vec{y}}\right]^2\right] d\vec{y} = V^* + B^*, \\
\text{and } R^*(\vec{y}) &= \frac{\left[V^*_{\vec{y}}\right]^{1/2}}{E^*_{\vec{y}}}, \text{ where } E^*_{\vec{y}} = \int \mathbf{K}(\vec{u}) g(\vec{y} - \vec{h}.\vec{u}) d\vec{u}.
\end{aligned}$$

g, w are two *product kernel density estimators* with some other kernel K^0 and bandwidths $\lambda \equiv \lambda_n$ and $\mu \equiv \mu_n$ respectively. They are defined as follows

$$\begin{aligned}
g(\vec{y}) \equiv g_n(\vec{y}) &= \frac{1}{n\lambda^d} \sum_{i=1}^n \prod_{j=1}^d K^0\left(\frac{y_j - X_{ij}}{\lambda}\right) \\
w(\vec{y}) \equiv w_n(\vec{y}) &= \frac{1}{n\mu^d} \sum_{i=1}^n \prod_{j=1}^d K^0\left(\frac{y_j - X_{ij}}{\mu}\right).
\end{aligned}$$

The bandwidths satisfy $h, \lambda, \mu \to 0$ and $nh^d, n\lambda^d, n\mu^d \to \infty$.

Sain, Baggers and Scott (1994) derived a multivariate version of Taylor's (1989) estimator of M. We call this estimator T. The formula for T can be written as

$$T = \frac{1}{n(4\pi)^{d/2} h_1 h_2 \dots h_d} + \vec{w_1}\{(1 - \frac{1}{n})\Omega^*_2 - 2\Omega^*_1 + \Omega^*_0\}(\vec{w_1})'$$

where Ω^*_j is a $n \times n$ matrix having (i_1, i_2) th entry equal to $\phi_{(2+j)D}(\vec{X}_{i_1} - \vec{X}_{i_2})$, $j = 0, 1, 2$ and $\vec{w_1} = \left(\frac{1}{n}, \dots, \frac{1}{n}\right)_{1\times n}$. If $h_1 = h_2 = \dots = h_d$ and K, K^0 are standard normal densities then $M^* = T$.

$M_{\vec{y}}$ and M have the following asymptotic expansions (see Rao (1983))

$$\begin{aligned}
M_{\vec{y}} &= A_{\vec{y}} + o\left(\frac{1}{nh^d} + h^4\right), \text{ where} \\
A_{\vec{y}} &= \frac{f(\vec{y})}{nh^d}\left\{\int K^2(u) du\right\}^d + \frac{h^4}{(2!)^2}\left[\int K(u) u^2 \sum_{i=1}^d f_i^{(2)}(\vec{y})\right]^2 \\
\text{and } M &= A + o\left(\frac{1}{nh^d} + h^4\right), \text{ where} \\
A &= \frac{\left\{\int K^2(u) du\right\}^d}{nh^d} + \frac{h^4}{(2!)^2}\int\left[\int K(u) u^2 \sum_{i=1}^d f_i^{(2)}(\vec{y})\right]^2 d\vec{y}.
\end{aligned}$$

The plug-in estimators of $M_{\vec{y}}$ and M are obtained by replacing unknown quantities in $A_{\vec{y}}$ and A, by corresponding data based estimators. We denote

$$\hat{A} = \frac{\left\{\int K^2(u)du\right\}^d}{nh^d} + \frac{h^4}{(2!)^2}\int\left[\int K(u)u^s\sum_{i=1}^{d}\hat{f}_i^{(2)}(\vec{y})\right]^2 d\vec{y}.$$

$$\hat{A}_{\vec{y}} = \frac{\hat{f}(\vec{y})}{nh^d}\left\{\int K^2(u)du\right\}^d + \frac{h^4}{(2!)^2}\left[\int K(u)u^2\sum_{i=1}^{d}\hat{f}_i^{(2)}(\vec{y})\right]^2.$$

$\hat{f}_i^{(2)}(\vec{y})$ is some data based estimator of $f_i^{(2)}(\vec{y})$, $i = 1, 2, .., d$, which is the 2nd order partial derivative of the function f with respect to x_i.

For a function $H : R^d \to R$, let $||H|| = \sup_{\vec{y}\in R^d} H(\vec{y})$. A function H is said to be uniformly bounded if $||H|| < \infty$.

Let $H_{i_1,..,i_q}(\vec{y}) = \left[\frac{\partial^q}{\partial x_{i_1}\cdots\partial x_{i_q}} H(\vec{x})\right]_{\vec{x}=\vec{y}}$ and $H_i^{(q)}(\vec{y}) = \left[\frac{\partial^q}{\partial x_i^q} H(\vec{x})\right]_{\vec{x}=\vec{y}}$, where $\vec{x} = (x_1, x_2, ..., x_d)$ and $\vec{y} = (y_1, y_2, ..., y_d)$. For $\vec{y} \in R^d$, $||\vec{y}|| = \sqrt{\sum_{i=1}^{d} y_i^2}$.

$$r_1 = E\left[\frac{B^*_{\vec{y}}}{B_{\vec{y}}} - 1\right]^2, \quad r_2 = E\left[\frac{V^*_{\vec{y}}}{V_{\vec{y}}} - 1\right]^2, \quad r_3 = E\left[\frac{M^*_{\vec{y}}}{M_{\vec{y}}} - 1\right]^2 \quad r_4 = E\left|\frac{M^*}{M} - 1\right|$$

and $r = \dfrac{R^*_{\vec{y}}}{R_{\vec{y}}} - 1$.

We note that r_1, r_2, r_3 and r depend on $\vec{y}$ and n. r_4 depends on n. The rates of convergence of r_i, $i = 1, 2, 3, 4$ and r to zero provide insight into how fast the accuracy of the proposed estimators improve.

Let $\phi(\cdot)$ denote the density of $N(0, 1)$ distribution. $\phi_{\vec{m},S}(\cdot)$ be the density of multivariate normal distribution with mean vector $\vec{m} = (m_1, ..., m_d)$ and variance covariance matrix S which is a $d \times d$ positive definite matrix.

I_d denotes the $d \times d$ identity matrix.

4.2.2 Chapter Summary

There are 9 theorems and two simulation studies in this chapter. In the first four theorems we obtain the rates at which r_1, r_3, r and r_4 converge to zero, as n is increased. In Theorem 4.4.5 we obtain some fixed sample properties of the proposed estimators. Theorem 4.4.9 implies that the estimator T, proposed by Sain, Baggerly and Scott

(1994), is not suitable for estimating M when h is large, where h is the smoothing parameter of a simple product kernel density estimator. We have obtained exact formulae for $M^*_{\bar{y}}$, $M_{\bar{y}}$ and M^* for Gaussian kernel, see Theorems 4.4.6, 4.4.7, 4.4.8 respectively. These results reduce the amount of computation involved in the bootstrap estimation.

We note that asymptotic rates of convergence of the proposed estimators depend on two parameters, namely λ and μ. It appears that the choice of λ depends upon the parameter we want to estimate. In particular for estimating M by M^*, there can be two possible choices of λ, namely λ equal to $\frac{1}{n^{1/(d+2)}}$ and $\frac{1}{n^{1/(d+8)}}$. Using a simulation study we investigate how well M^* estimates M for both these choices of λ. In another simulation experiment we study how well $M^*_{\bar{y}}$ imitates $M_{\bar{y}}$ by plotting them against a wide range of values of h. An interesting observation is that, if $M_{\bar{y}}$ attains minima for multiple values of h then the same feature is exhibited by the bootstrap estimate $M^*_{\bar{y}}$. Final section is devoted to proofs of the theorems stated in this chapter. We have addressed the problems of estimating some local and global measures of accuracy of univariate and multivariate kernel density estimates in chapters 2, 3 and in the current chapter.

An important application of M is that it serves a criterion for bandwidth selection for univariate and multivariate kernel density estimates. The problem of automatic bandwidth selection, by minimizing M^*, is addressed in next chapter.

4.3 Assumptions

We now collect some assumptions which will be used in the sequel.

Assumption A (Assumption on f) (i) The density function $f : R^d \rightarrow (0, \infty)$ is uniformly bounded and possesses all uniformly bounded and continuous partial derivatives up to order 4.

(ii) The 2nd and 4th order partial derivatives are square integrable.

Assumption B (Assumptions on K) K is assumed to be a non-negative, square integrable kernel of order 2, i.e. $\int K(u)du = 1$, $\int K(u)udu = 0$, $\int K(u)u^2du < \infty$ and

$\int K^2(u)du < \infty$.

Assumption C (Assumptions on K^0) (i) $K^0(.)$ is a non-negative second order kernel and $\int [K^0(y)]^4 dy < \infty$. Further

(a) $\int K^0(u)|u|^j du < \infty,\ j = 1, .., 4$.

(b) $|u^3|K^0(u) \to 0$, as $|u| \to \infty$.

(ii) $K^0(\cdot)$ has 2 continuous derivatives on $(-\infty, \infty)$ and the two derivatives $K^{0(i)}(\cdot), i = 1, 2$, satisfy the following assumptions

(c) $\int \left[K^{0(1)}(u)\right]^2 du < \infty$, $\int \left[K^{0(1)}(u)\right]^2 u du = 0$ and $\int |K^{0(1)}(u) u^j| du < \infty$, $j = 0, 1, .., 4$.

(d) $u^3 K^{0(1)}(u) \to 0$, as $|u| \to \infty$.

(e) $\int |K^{0(2)}(x)| dx < \infty$, $\int [K^{0(2)}(y)]^4 dy < \infty$ and $0 < \int |K^{0(2)}(x) x^j| dx < \infty$, $j = 0, 1, .., 4$.

Remark 4.3.1. (i) An obvious choice of K^0 is the standard normal density.

(ii) Assumptions $A(i) - (iii)$ on f are valid for a wide class of densities which include mixture of multivariate normal densities.

(iii) We note that the Assumptions B and C are similar to the Assumptions B and C in Chapters 2 and 3 respectively. The Assumption A can be looked at as multivariate extension of the Assumption A in Chapter 2 and 3. This is because the smoothness assumptions, in Chapter 2 and 3, on f and its derivatives are special case of Assumption A (i) and(ii), when $d = 1$ and kernel order is 2.

4.4 Main results

We now state our main results. The proofs are given at the end of the chapter.

Theorem 4.4.1. *Suppose the Assumptions $A - C$, $n\lambda^{d+4} \to \infty$ and $f_{ii}(\bar{y}) \neq 0$, for some $i = 1, 2, ..., d$, hold. Then*

$$r_1 = E\left[\frac{B^*_{\bar{y}}}{B_{\bar{y}}} - 1\right]^2 = O\left(\frac{1}{n\lambda^{d+4}} + \lambda^4\right).$$

Theorem 4.4.2. *Suppose the Assumptions $A - C$, $n\lambda^{d+4}$, $n\mu^d \to \infty$, $f(\vec{y}) > 0$ and $f_{ii}(\vec{y}) \neq 0$, for some $i = 1, 2, .., d$, hold. Then*

$$r_3 = E\left[\frac{M^*_{\vec{y}}}{M_{\vec{y}}} - 1\right]^2 = O\left(\left[\frac{1}{n\lambda^{d+4}} + \lambda^4\right] + \left[\frac{1}{n\mu^d} + \mu^4\right]\right).$$

Theorem 4.4.3. *Let the Assumptions $A - C$ hold and $f(\vec{y}) > 0$. Then*

(i)

$$r = \frac{R^*_{\vec{y}}}{R_{\vec{y}}} - 1 = O_P\left(\frac{1}{\sqrt{n\lambda^d}} + \lambda^2 + \frac{1}{\sqrt{n\mu^d}} + \mu^2\right),$$

(ii) If $h = o(\lambda)$, $n\lambda^{d+4} = o(1)$, $n\lambda^d\mu^4 = o(1)$, $\lambda = o(\mu)$ and K, K^0 are standard normal densities then

$$\sqrt{n\lambda^d}r \to_L N(0, C'), \text{ where } C' = \frac{(2\sqrt{\pi})^d}{f^3(\vec{y})}.$$

Theorem 4.4.4. *Let the Assumptions $A - C$ hold. Also let $n\lambda^{d+4} \to \infty$, $n\mu^d \to \infty$ and the 3rd order partial derivatives of f be integrable. Then*

$$r_4 = E\left|\frac{M^*}{M} - 1\right| = O\left(\frac{1}{\sqrt{n\lambda^{d+4}}} + \lambda^2 + \frac{1}{\sqrt{n\mu^d}} + \mu^2\right).$$

4.4.1 Fixed Sample performance of M^* and $M^*_{\vec{y}}$

What are the effects of different possible choices of h, μ and λ on the bootstrap estimator M^* for *fixed sample size n?* The following proposition provides some answers.

Theorem 4.4.5. *Suppose $f : R^d \to (0, \infty)$ is a uniformly bounded density and $f(\vec{y}) \to 0$, as $||y|| \to \infty$. Let K^0 be a uniformly bounded probability density function and $K^0(y) \to 0$, as $|y| \to \infty$. Then for any fixed sample size n and λ, $\mu \neq h$, as $h \to \infty$*

(i) $M \to \int f^2(\vec{y})d\vec{y}$ and $M^* \to \int g^2(\vec{y})d\vec{y}$ almost surely.

(ii) $A, \hat{A} \to \infty$, almost surely.

(iii) $M_{\vec{y}} \to f^2(\vec{y})$ and $M^*_{\vec{y}} \to g^2(\vec{y})$, almost surely.

(iv) $A_{\vec{y}}$, $\hat{A}_{\vec{y}} \to \infty$, almost surely.

Remark 4.4.1. (i) Theorem 4.4.5(i) implies that for fixed n, M^* and M level off and the former succeeds in imitating the behavior of the latter for larger values of h. On the other hand, Theorem 4.4.5(ii) demonstrates that A and $\hat{A}$ explodes as h is increased. So A and $\hat{A}$ fail to estimate M closely for larger value of h.

(ii) Theorem 4.4.5 (iii) and (iv) imply that the bootstrap estimator $M^*_{\vec{y}}$ successfully imitates $M_{\vec{y}}$, but $A_{\vec{y}}$ and $\hat{A}_{\vec{y}}$ fail to mimic $M_{\vec{y}}$, for large values of h and fixed sample size.

4.4.2 The special case of Gaussian kernel

$B^*_{\vec{y}}$, $V^*_{\vec{y}}$, $M^*_{\vec{y}}$ and M^* do not have a closed form expressions in general and hence Monte-Carlo computation is required for its implementation. However we observe that if K is a Gaussian kernel and K^0 is chosen to be the standard normal density then we can obtain closed form expression for the proposed bootstrap estimators.

Theorem 4.4.6. *If K and K^0 are densities of $N(0,1)$ distribution then*

$$\begin{aligned} B^*_{\vec{y}} &= \frac{1}{n}\sum_{i=1}^{n} \phi_{\vec{X}_i, D+\lambda^2 I_d}(\vec{y}) - g(\vec{y}), \\ V^*_{\vec{y}} &= \frac{1}{n^2(2\sqrt{\pi})^d \prod_{j=1}^{d} h_j} \sum_{i=1}^{n} \phi_{\vec{X}_i, \mu^2 I_d + \frac{1}{2}D}(\vec{y}) - \frac{1}{n}\left[\frac{1}{n}\sum_{i=1}^{n} \phi_{\vec{X}_i, \mu^2 I_d + D}(\vec{y})\right]^2 \\ \text{and } M^*_{\vec{y}} &= V^*_{\vec{y}} + (B^*_{\vec{y}})^2, \quad \text{where } D = diag(h_1^2, \ldots, h_d^2). \end{aligned}$$

If the underlying distribution is assumed to be multivariate-normal distribution or mixtures of normal distributions, then for Gaussian kernel K we can also obtain (repeating similar arguments as in Theorem 4.4.6) closed form expressions for $B_{\vec{y}}$, $V_{\vec{y}}$ and $M_{\vec{y}}$.

Theorem 4.4.7. *Let K be the density of $N(0,1)$ distribution.*
(i) If $f(\vec{x}) = \sum_{i=1}^{k} w_i \phi_{\Sigma_i}(\vec{x} - \vec{m}_i)$, where $\sum_{i=1}^{k} w_i = 1$, then

$$\begin{aligned} B_{\vec{y}} &= \sum_{i=1}^{k} w_i \phi_{D+\Sigma_i}(\vec{y} - \vec{m}_i) - f(\vec{y}), \\ V_{\vec{y}} &= \frac{1}{n(2\sqrt{\pi})^d \prod_{j=1}^{d} h_j} \sum_{i=1}^{k} w_i \phi_{\Sigma_i + \frac{1}{2}D}(\vec{y} - \vec{m}_i) - \frac{1}{n}\left[\sum_{i=1}^{k} w_i \phi_{D+\Sigma_i}(\vec{y} - \vec{m}_i)\right]^2, \\ \text{and } M_{\vec{y}} &= V_{\vec{y}} + (B_{\vec{y}})^2. \end{aligned}$$

(ii) If $f(\vec{x}) = \int_0^\infty \phi_{\frac{1}{v^2}I_d}(\vec{x})dv$, *then*

$$\begin{aligned} B_{\vec{y}} &= \int_0^\infty \phi_{D+\frac{1}{v^2}I_d}(\vec{y})dv - f(\vec{y}), \\ V_{\vec{y}} &= \frac{1}{n(2\sqrt{\pi})^d \prod_{j=1}^d h_j} \int_0^\infty \phi_{\frac{1}{v^2}I_d+\frac{1}{2}D}(\vec{y})dv - \frac{1}{n}\left[\int_0^\infty \phi_{D+\frac{1}{v^2}I_d}(\vec{y})dv\right]^2, \\ \text{and } M_{\vec{y}} &= V_{\vec{y}} + (B_{\vec{y}})^2. \end{aligned}$$

Jones and Wand (1993) obtained closed form expression for M for multivariate kernel density estimator with Gaussian kernel, dispersion matrix H and assuming the underlying density to be mixture of multivariate normal densities (see their Theorem 1, page 524).

Now suppose that K and K^0 are Gaussian. Then we can obtain exact formulae of V^*, B^* and $M^* = V^* + B^*$ by repeating similar arguments as in the proof of Theorem1, Jones and Wand (1993).

Theorem 4.4.8. *If* K, K^0 *are Gaussian kernels, then*

$$M^* = \frac{1}{n(4\pi)^{d/2}h_1h_2...h_d} + \vec{w_1}\{\Omega_2^* - 2\Omega_1^* + \Omega_0^* - \frac{1}{n}\Omega_2^{**}\}(\vec{w_1})'$$

where Ω_2^{**} *is an* $n \times n$ *matrix having* (i_1, i_2) *th entry equal to* $\phi_{\vec{X}_{i_1}-\vec{X}_{i_2},2D+2\mu^2I_d}(\vec{0})$, Ω_j^*, $j = 0,1,2$ *is an* $n \times n$ *matrix having* (i_1, i_2) *th entry equal to* $\phi_{\vec{X}_{i_1}-\vec{X}_{i_2},jD+2\lambda^2I_d}(\vec{0})$, $\vec{w_1} = (\frac{1}{n}, ..., \frac{1}{n})_{1\times n}$, $(\vec{w_1})'$ *is the transpose of* $\vec{w_1}$ *and* $jD = diag(jh_1^2, jh_2^2, .., jh_d^2)$.

Remark 4.4.2. (i) When $h_1 = h_2 = ... = h_d = h$, T (Sain, Baggerly and Scott's (1994) estimator) is in fact a special case of our estimator M^*, where $\lambda = \mu = h$.

(ii) In general T can be looked at as a special case of a generalization of M^* (we call it M^{**}) which is described below.

Let us define $M^{**} = V^{**} + B^{**}$, where V^{**}, B^{**} are as defined below

$$\begin{aligned} B^{**} &= \int \left[\int \mathbf{K}(\vec{u})g(\vec{y} - \vec{h}.\vec{u})d\vec{u} - g(\vec{y})\right]^2 d\vec{y}, \\ V^{**} &= \frac{1}{n\prod_{j=1}^d h_j} \int \prod_{j=1}^d K^2(u_j)w(\vec{y} - \vec{h}.\vec{u})d\vec{u} - \frac{1}{n}\left[\int \mathbf{K}(\vec{u})w(\vec{y} - \vec{h}.\vec{u})d\vec{u}\right]^2, \end{aligned}$$

where $g(\vec{y}) = \frac{1}{n}\sum_{i=1}^n \phi_{\vec{X}_i,D_1}(\vec{y})$, $w(\vec{y}) = \frac{1}{n}\sum_{i=1}^n \phi_{\vec{X}_i,D_2}(\vec{y})$, $D_1 = diag(\lambda_1^2, ..., \lambda_d^2)$ and $D_2 = diag(\mu_1^2, ..., \mu_d^2)$. T is a special case of M^{**}, when $D_1 = D_2 = D = diag(h_1^2, ..., h_d^2)$.

Repeating the arguments in the proof of Theorem 4.4.8, we can easily derive an exact formula of M^{**} which is as given below

$$M^{**} = \frac{1}{n(4\pi)^{d/2}h_1h_2...h_d} + \vec{w_1}\{\Omega_2^* - 2\Omega_1^* + \Omega_0^* - \frac{1}{n}\Omega_2^{**}\}(\vec{w_1})',$$

where Ω_2^{**} is an $n \times n$ matrix having (i_1, i_2) th entry equal to $\phi_{\vec{X}_{i_1}-\vec{X}_{i_2},2D+2D_2}(\vec{0})$ and Ω_j^*, $j = 0, 1, 2$, are $n \times n$ matrix having (i_1, i_2) th entry equal to $\phi_{\vec{X}_{i_1}-\vec{X}_{i_2},jD+2D_1}(\vec{0})$. $\vec{w_1} = (\frac{1}{n}, ..., \frac{1}{n})_{1\times n}$, $(\vec{w_1})'$ is the transpose of $\vec{w_1}$ and $jD = diag(jh_1^2, jh_2^2, .., jh_d^2)$.
Next we obtain a fixed sample property of T.

Theorem 4.4.9. *If $h_i \to \infty$, $i = 1, 2, .., d$, then $T \to 0$, almost surely.*

Remark 4.4.3. From Theorem 4.4.5 (i), we know that for a simple product kernel density estimator $M \to \int f^2(\vec{y})d\vec{y}$, as h is increased, where $h = h_1 = h_2 = ... = h_d$. If $h_1 = h_2 = ... = h_d = h$, Theorem (4.4.9) implies that $T \to 0$ as $h \to \infty$. So T fails to imitate M, when h is large.

4.4.3 Choice of λ and μ

From Theorems 4.4.9 and 4.4.5 we see that the choices of μ and λ are important, especially when any one of the bandwidths $h_1, h_2, ..., h_d$ is large. Theorem 4.4.9 demonstrates the demerit of the choice $\lambda = h_i$, for some $i = 1, 2, .., d$. The Theorems obtained so far, provide some insight into appropriate choice of λ and μ. Following are some important observations.

(1) An appropriate choice of λ can depend on the parameter that we want to estimate. For instance to estimate $R_{\vec{y}}$ by $R_{\vec{y}}^*$, a choice of λ and μ can be obtained by minimizing the rate of convergence (in probability) of $\frac{R_{\vec{y}}^*}{R_{\vec{y}}}$ to one (see Theorem 4.4.3). So we can choose λ and μ to be constant multiples of $\frac{1}{n^{1/(d+4)}}$. For estimating $M_{\vec{y}}$ and M, one can choose λ and μ by minimizing the rates of convergence of $\frac{M_{\vec{y}}^*}{M_{\vec{y}}}$ and $\frac{M^*}{M}$ to one, obtained in Theorems 4.4.2 and 4.4.4 respectively. This leads us to choose λ and μ which are constant multiples of $\frac{1}{n^{1/(d+8)}}$ and $\frac{1}{n^{1/(d+4)}}$ respectively. So the choice of λ can vary depending upon whether we estimate M or $R_{\vec{y}}$.

(2) For estimating M, another criterion for the choice of λ is provided by Theorem 4.4.5. We see that for fixed n and $\lambda \neq h$, $M^* \to \int g^2(\vec{y})d\vec{y}$ and $M \to \int f^2(\vec{y})d\vec{y}$, as $h \to \infty$. Recalling that $g(\cdot)$ depends on λ, a criterion to choose λ can be to ensure that $\int g^2(\vec{y})d\vec{y}$ estimates $\int f^2(\vec{y})d\vec{y}$ (say) as closely as possible.

The problem of estmating $\int f^2(\vec{y})d\vec{y}$ by $\int g^2(\vec{y})d\vec{y}$ has been well studied in the literature for the univariate case (see Hall and Marron (1987)). It is possible to extend the theoretical deductions of (Hall and Marron (1987)) to the multivariate case as well and consequently it is possible to show that

$$E\left(\int g^2(\vec{y})d\vec{y} - \int f^2(\vec{y})d\vec{y}\right)^2 = O\left(\frac{1}{n^2\lambda^{2d}} + \lambda^4\right).$$

So an appropriate choice of λ, for estimating M, can be the minimizer of the right side of the above equation and such a choice of λ is a constant multiple of $\frac{1}{n^{1/(d+2)}}$.

4.5 Simulation

In this section we provide insight into how well the proposed estimators $M^*_{\vec{y}}$ and M^* perform when sample size is fixed. We observe that for M^* there are more than one option for choosing λ. We compare how the different choices of λ can effect M^*.

From Chapter 3 we recall that for a univariate kernel density estimator, the mean squared error M_y and its bootstrap estimator M^*_y can attain local minima for more than one value of h, especially when y is in the tail of the underlying density. A natural question is "does $M_{\vec{y}}$ and $M^*_{\vec{y}}$ exhibit a similar property?" We investigate this in our second simulation experiment. In Table 4.1 we provide the formulae of six bivariate mixed normal densities, namely correlated normal, bimodal, kurtic, skewed, trimodal and quadrimodal density. These distributions are used in the sequel.

For $n = 50,\ 200$, we have plotted M^* with $\lambda = \frac{1}{n^{1/(d+2)}}, \frac{1}{n^{1/(d+8)}}$, T and M, against $\log_{10} h$. Our choice of $\log_{10}$ scale is motivated by its use by Marron and Wand (1992). To distinguish the four curves we number M^* (with $\lambda = \frac{1}{n^{1/(d+2)}}$), M, M^* (with $\lambda = \frac{1}{n^{1/(d+8)}}$) and T as 1, 2, 3 and 4 respectively. Each curve has been plotted for correlated normal, bimodal, kurtic, skewed and quadrimodal densities. The main obsrvations are as follows

Table 4.1: Parameters of 6 bivariate Normal Mixture distributions

Density	$w_1.N(m_{11}, m_{12}, \sigma_{11}^2, \sigma_{12}^2, \rho_1) + ... + w_k.N(m_{k1}, m_{k2}, \sigma_{k1}^2, \sigma_{k2}^2, \rho_k)$
correlated normal	$N(0, 0, 1, 1, -0.5)$
bimodal	$\frac{1}{2}N(-1, 0, (\frac{2}{3}))^2, (\frac{2}{3}))^2, 0) + \frac{1}{2}N(1, 0, (\frac{2}{3}))^2, (\frac{2}{3})^2, 0)$
kurtic	$\frac{2}{3}N(0, 0, 1, 4, \frac{1}{2}) + \frac{1}{3}N(1, 0, (\frac{2}{3}))^2, (\frac{1}{3})^2, -0.5)$
trimodal	$\frac{1}{3}N(-\frac{5}{6}, 0, (\frac{3}{5})^2, (\frac{3}{5})^2, 0\ 7) + \frac{1}{3}N(\frac{5}{6}, 0, (\frac{3}{5})^2, (\frac{3}{5})^2, 0\ 7)$ $+\ \frac{1}{3}N(0, 0, (\frac{3}{5})^2, (\frac{3}{5})^2, -0.7)$
quadrimodal	$\frac{1}{8}N(-1, 1, (\frac{2}{3})^2, (\frac{2}{3})^2, 0.4) + \frac{3}{8}N(-1, -1, (\frac{2}{3})^2, (\frac{2}{3})^2, -0.6)$ $+\frac{1}{8}N(1, -1, (\frac{2}{3})^2, (\frac{2}{3})^2, -0.7) + \frac{3}{8}N(1, 1, (\frac{2}{3})^2, (\frac{2}{3})^2, -0.5)$
skewed	$\frac{1}{3}N(-\frac{5}{6}, 0, (\frac{3}{5})^2, (\frac{3}{5})^2, 0.7) + \frac{1}{3}N(\frac{5}{6}, 0, (\frac{3}{5})^2, (\frac{3}{5})^2, 0.7)$ $+\ \frac{1}{3}N(0, 0, (\frac{3}{5})^2, (\frac{3}{5})^2, -0.7)$

1) *Consistency.* Comparing Figures 4.1(a) with 4.1(b),....,4.5(a) with 4.5(b), we see that the curves 1, 2 and 3 are closer for $n = 200$ than for $n = 50$. So M^* (with $\lambda = \frac{1}{n^{1/(d+2)}}$ and $\frac{1}{n^{1/(d+8)}}$) are consistent estimators of M. However T, curve 4, does not show much improvement in imitating M even as the sample size is increased. This observation and Theorem 4.4.9 demonstrate that $\lambda = h = h_1 = ... = h_d$ is not an appropriate choice for estimating M.

It is interesting to note that $\lambda = \frac{1}{n^{1/(d+2)}}$ does not satisfy the condition $n\lambda^{d+4} \longrightarrow \infty$, which is one of the conditions in Theorem 4.4.4. In view of the consistent behaviour of M^* (with $\lambda = \frac{1}{n^{1/(d+2)}}$) it appears that in general M^* can be consistent even if the condition $n\lambda^{d+4} \longrightarrow \infty$ is relaxed.

2) *Small sample behaviour.* For $n = 50$, the curves 1, 2 and 3 are quite close for correlated normal, bimodal and quadrimodal densities, see Figures 4.1(a), 4.2(a) and 4.5(a). So it appears that M^* (with $\lambda = \frac{1}{n^{1/(d+2)}}$ and $\frac{1}{n^{1/(d+8)}}$) can closely estimate M even for small samples, especially if the underlying density is unimodal, symmetric or posssesses multiple peaks which are not very close to one another. However for skewed density and $n = 50$, the curve 2 is not close to either 1 or 3, especially when h is large (see Figure 4.4(a)). So the small sample performance of M^* can be sensitive to the underlying f, especially if the underlying density possesses features such as skewness or multiple peaks which are close.

In Figures 4.1(a) – 4.5(b) we plot M^* (with $\lambda = \frac{1}{n^{1/(d+2)}}$), M, M^* (with $\lambda = \frac{1}{n^{1/(d+8)}}$) and T numbered as 1,2,3 and 4 respectively, against $\text{Log}(h) \equiv \log_{10} h$ for correlatednormal, bimodal, kurtic, skewed and quadrimodal distributions and for sample sizes $n = 50$ and 500. Both K, K^0 are standard normal densities and $d = 2$.

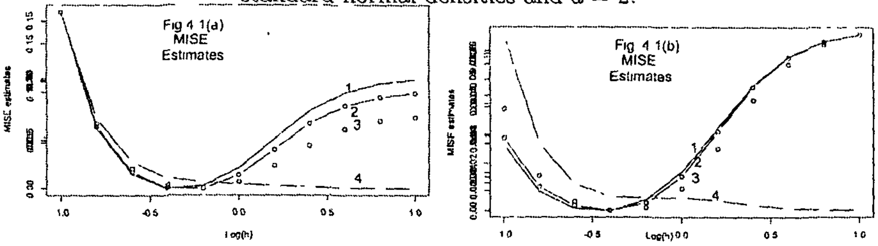


Fig 4.1 (a), (b): underlying density "correlated normal"; n=50, 200; d=2.

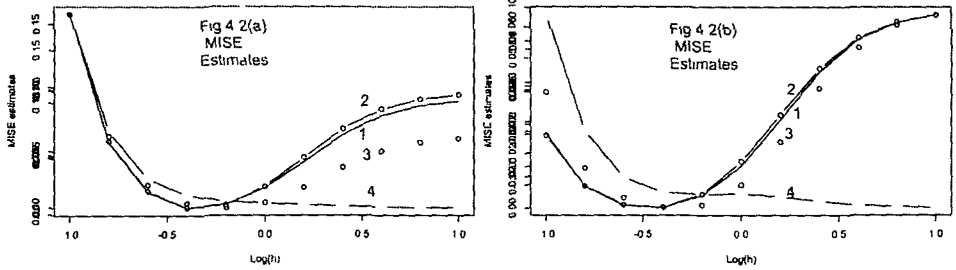


Fig 4 2 (a), (b): underlying density "bimodal", n=50, 200; d=2.

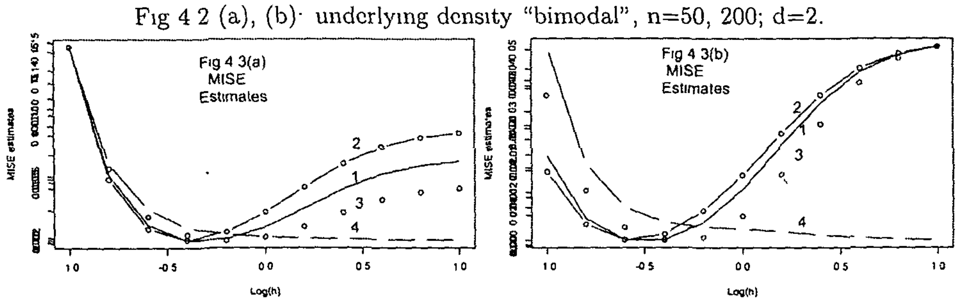


Fig 4.3 (a), (b). underlying density "kurtic"; n=50, 200; d=2.

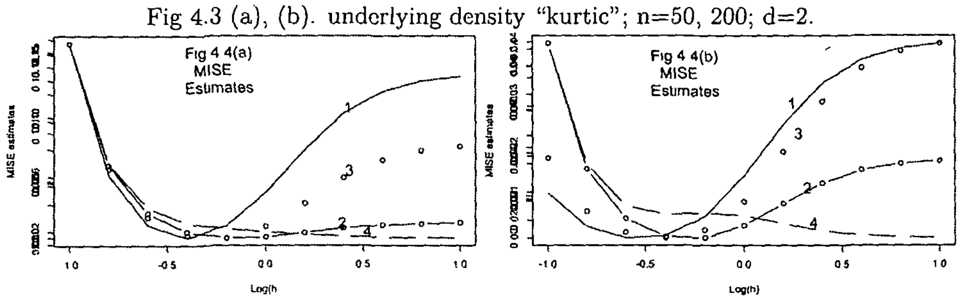


Fig 4.4 (a), (b): underlying density "skewed"; n=50, 200; d=2.

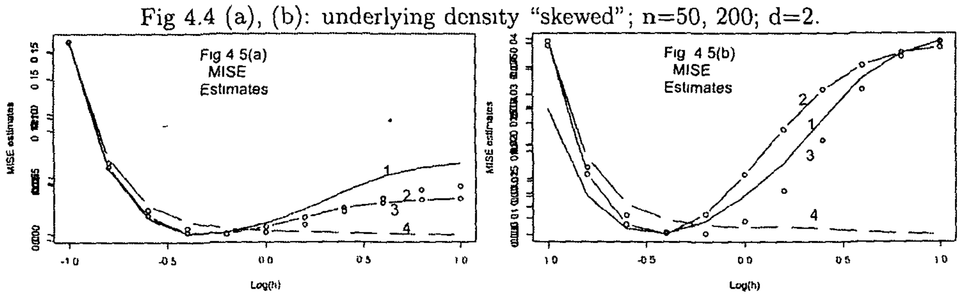


Fig 4 5 (a), (b)· underlying density "quadrimodal", n=50, 200; d=2

In the next experiment we plot the curves $A^*_{\vec{y}}$, $M^*_{\vec{y}}$ and $M_{\vec{y}}$ against a wide range of values of $\log_{10}(h)$, for $n = 50$ and $n = 500$. The three curves are plotted for a number of different values of $\vec{y}$ and for two bivariate densities, namely correlated normal and bimodal density. We have also plotted the three curves for some of the other bivariate densities, mentioned in Table 4.1. But these plots are not shown as the main features appear to be similar. The main observations are as follows

1) *Effect of* $\vec{y}$. From Figures 4.6(a) to 4.13(b), we see that $M_{\vec{y}}$ is sensitive to $\vec{y}$. If $\vec{y}$ is the mode or a point close to the mode of the underlying density then $M_{\vec{y}}$ attains one global minima. But if $\vec{y}$ is a point away from the mode, $M_{\vec{y}}$ can attain minima for more than one value of h. For instance in the Figures 4.11(b) to 4.13(b), we note that $M_{\vec{y}}$ attains two minima, one for $\log_{10}(h) < 0$ and the other for $\log_{10}(h) > 0$. We also note that whenever $M_{\vec{y}}$ attains more than one minima, the global minima appears to be attained for $\log_{10}(h) > 0$, see Figures 4.11(a) to 4.13(b).

$A_{\vec{y}}$ does not appear to be much affected by $\vec{y}$. It always attains one global minimum for a value of h satisfying $\log_{10}(h) \leq 0$. From the Figures 4.6(a) to 4.9(b) we see that $A_{\vec{y}}$ can be poor estimate of $M_{\vec{y}}$ for $\log_{10}(h) > 0$. These observations and Theorem 4.4.5 (iii), (iv) imply that $A_{\vec{y}}$ may not be the appropriate estimate of $M_{\vec{y}}$.

$M^*_{\vec{y}}$ appears to be affected by $\vec{y}$. For correlated normal density and $\vec{y} = (2, 2)$, both $M^*_{\vec{y}}$ and $M_{\vec{y}}$ attains multiple minima, see Figures 4.13(a) and 4.13(b). From Figures 4.7(a) to 4.8(b) and 4.10 (a), (b) we see that, if $\vec{y}$ is the mode of the underlying density then both $M^*_{\vec{y}}$ and $M_{\vec{y}}$ possesses one global minima. In general we observe that as the distance of $\vec{y}$ from the modes of the underlying density increases, $M^*_{\vec{y}}$ appears to be closer to $M_{\vec{y}}$ than $A_{\vec{y}}$.

2) *Small sample behaviour*. From Figures 4.6(a) to 4.13(a) we see that for $n = 50$, $M^*_{\vec{y}}$ imitates $M_{\vec{y}}$ more closely than $A_{\vec{y}}$, for almost all values of $\vec{y}$ and $\log_{10}(h) \geq 0$. In general the Figures 4.6(a) to 4.13(a) reveal that for small samples $M^*_{\vec{y}}$ can estimate $M_{\vec{y}}$ quite well, especially when y is not equal or close to the peaks of the underlying density and $\log_{10}(h) \geq 0$. However from Figures 4.11(a) and 4.11(b) we note that for

correlated normal density and $\vec{y} = (1,1)$, both $A_{\vec{y}}$ and $M^*_{\vec{y}}$ fail to estimate $M_{\vec{y}}$ closely. This only indicates that for any fixed sample size, the problem of accurate estimation of $M_{\vec{y}}$ at every point $\vec{y}$ appears to be very difficult.

In Figures 4.6(a) – 4.13(b) we plot $A^*_{\vec{y}}$, $M^*_{\vec{y}}$ and $M_{\vec{y}}$ against $\text{Log}(h) \equiv \log_{10} h$ for bimodal and correlated normal distributions and for sample size $n = 50,\ 500$. Both K and K^0 are standard normal densitics

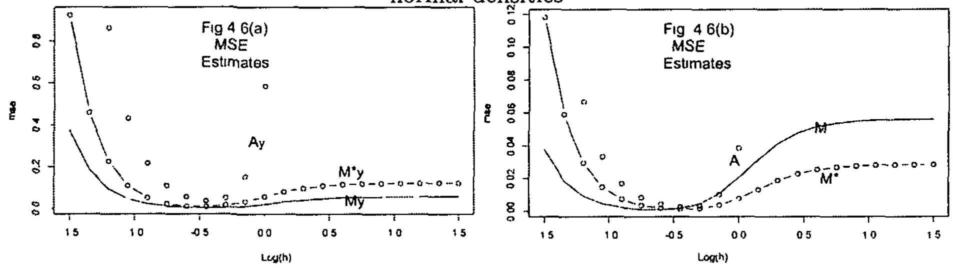


Fig 4.6 (a), (b): underlying density "bimodal"; y=(0,0); n=50, 500.

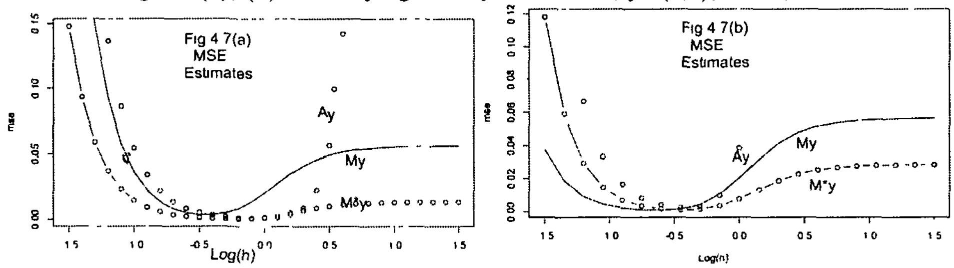


Fig 4.7 (a), (b): underlying density "bimodal"; y=(-1,0); n=50, 500.

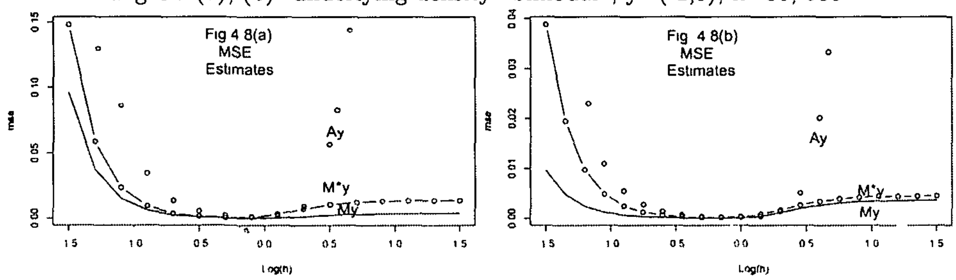


Fig 4.8 (a), (b): underlying density "bimodal"; y=(1,0); n=50, 500.

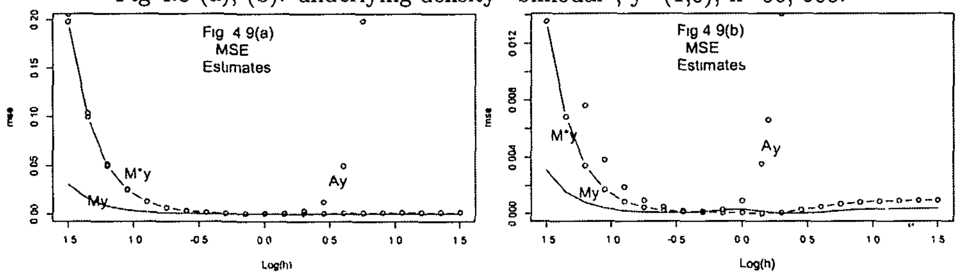


Fig 4.9 (a), (b): underlying density "bimodal"; y=(1,1); n=50, 500.

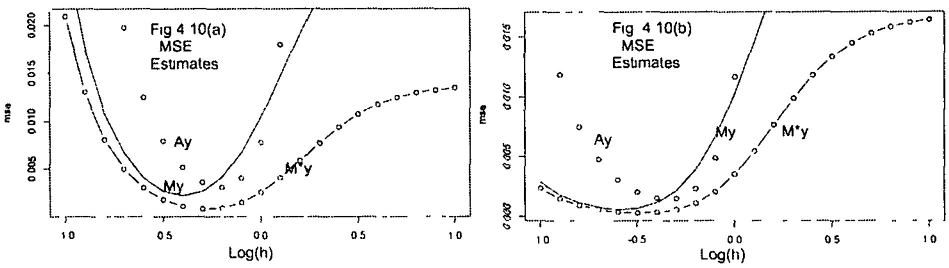


Fig 4.10 (a), (b): underlying density "correlatednormal"; y=(0,0), n=50, 500.

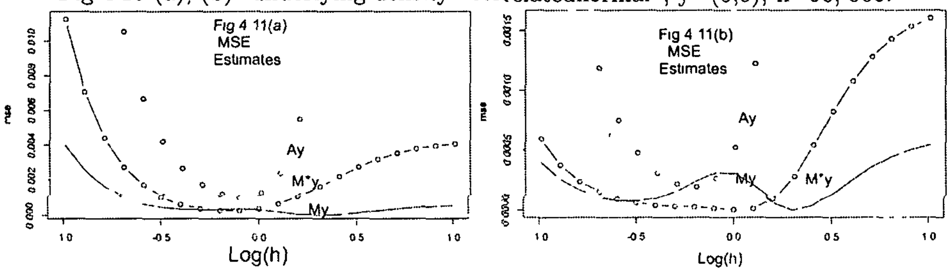


Fig 4.11 (a), (b). underlying density "correlatednormal", y=(1,1), n=50, 500

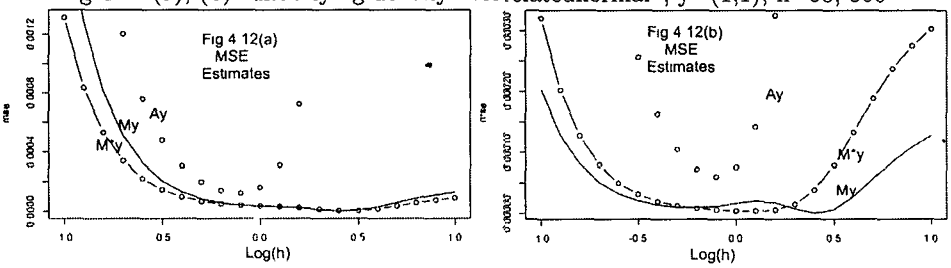


Fig 4.12 (a), (b): underlying density "correlatednormal", y=(-2,2); n=50, 500.

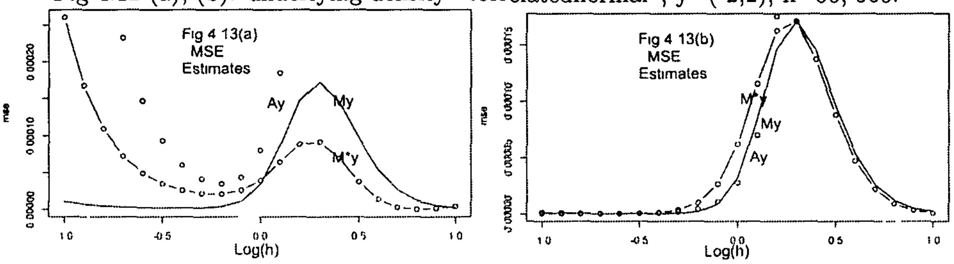


Fig 4.13 (a), (b) underlying density "correlatednormal"; y=(2,2); n=50, 500

4.6 Proofs

First we state and prove a number of Lemmas which are used in the proofs of Theorems.

4.6.1 Some important Lemmata

Lemma 4.6.1. *Under Assumption C, the two derivatives of K^0 possesses the following properties*

(i) $\int K^{0(1)}(u)du = 0$, $\int K^{0(1)}(u)udu = -1$ *and* $\int K^{0(1)}(u)u^2du = 0$.

(ii) $\int K^{0(2)}(x)x^j dx = 0$, *where* $0 \leq j \leq 3, j \neq 2$ *and* $\frac{1}{2!}\int K^{0(2)}(x)x^2dx = 1$.

Proof of Lemma 4.6.1 (i) $\int K^{0(1)}(u)du = 0$ is a direct consequence of Assumption C (b). To prove that $\int K^{0(1)}(u)udu = -1$, let $a(u) = uK^0(u)$. Then

$$\begin{aligned} a^{(1)}(u) &= uK^{0(1)}(u) + K^0(u) \\ \Rightarrow \int_{-\infty}^{\infty} a^{(1)}(u)du &= \int_{-\infty}^{\infty} uK^{0(1)}(u)du + 1 \\ \Rightarrow \int_{-\infty}^{\infty} uK^{0(1)}(u)du &= [a(u)]_{-\infty}^{\infty} - 1. \end{aligned}$$

From Assumption C (b) we note that $a(u) = uK^0(u) = o(1),\ |u| \to \infty$.
So $\int uK^{0(1)}(u)du = -1$. This proves the 2nd part of Lemma 4.6.1 (i).

To prove the third result we take $a(u) = u^2K^0(u)$ and repeat the above arguments. The only difference is that, in this case we use the property $\int uK^0(u)du = 0$ (Assumption C (i) on K^0).

Part (ii) can be proved easily by similar arguments under Assumption C (d). □

Lemma 4.6.2. *Under Assumptions A and C, we get*

(i) $\sup_{\vec{y}\in R^d} E\left[g_{ij}(\vec{y}) - f_{ij}(\vec{y})\right]^2 = O\left(\frac{1}{n\lambda^{d+4}} + \lambda^4\right),\ \forall i,\ j = 1, 2, ..., d.$

(ii) $\sup_{\vec{y}\in R^d} E\left[g_{ij}(\vec{y}) - f_{ij}(\vec{y})\right]^4 = O\left(\frac{1}{n^2\lambda^{8+2d}} + \lambda^8\right),\ \forall i,\ j = 1, 2, ..., d.$

(iii) $\int E\left[g_{ij}(\vec{y}) - f_{ij}(\vec{y})\right]^2 d\vec{y} = O\left(\frac{1}{n\lambda^{d+4}} + \lambda^4\right),\ \forall i, j = 1, 2, ..., d.$

(iv) further assuming that all the 3rd order partial derivatives of f are integrable,
$E\left[\int (w(\vec{y}) - f(\vec{y}))^2\right] = \frac{1}{n\mu^d}\left[\int (K^0(u))^2du\right]^d + \frac{\mu^4\left[\int K^0(z)z^2dz\right]^2 \int\{\sum_{i=1}^d f_{ii}(\vec{y})\}^2 d\vec{y}}{(2!)^2} + o\left(\frac{1}{n\mu^d} + \mu^4\right).$

Proof of Lemma 4.6.2 To prove part (i) we first consider the case where $i \neq j =$

$1, 2, ..., d$. We recall that

$$g(\vec{y}) = \frac{1}{n\lambda^d}\sum_{i_1=1}^{n}\prod_{i_2=1}^{d} K^0\left(\frac{y_{i_2} - X_{i_1 i_2}}{\lambda}\right)$$

$$\Rightarrow g_{ij}(\vec{y}) = \frac{1}{n\lambda^{d+2}}\sum_{l_1=1}^{n}\prod_{l=1,\ l\neq i,j}^{d} K^0\left(\frac{y_l - X_{l_1 l}}{\lambda}\right) K^{0(1)}\left(\frac{y_i - X_{l_1 i}}{\lambda}\right) K^{0(1)}\left(\frac{y_j - X_{l_1 j}}{\lambda}\right).$$

Taking expectation on both sides of the above equation we get

$$E[g_{ij}(\vec{y})] = \frac{1}{\lambda^2}\int_{\vec{u}}\prod_{l=1,\ l\neq i,j}^{d} K^0(u_l)K^{0(1)}(u_i)K^{0(1)}(u_j)f(\vec{y} - \lambda\vec{u})d\vec{u}.$$

Expanding $f(\vec{y} - \lambda\vec{u})$ by Taylor expansion and using Assumption C, and Lemma 4.6.1 we get

$$E[g_{ij}(\vec{y})] = f_{ij}(\vec{y}) + b_{ij}(\vec{y}), \text{ where } b_{ij}(\vec{y}) \text{ is equal to}$$

$$\frac{\lambda^2}{3!}\Big[\sum_{i_1,i_2,i_3,i_4=1}^{d}\int_{\vec{u}}\prod_{l=1,\ l\neq i,j}^{d} K^0(u_l)K^{0(1)}(u_i)K^{0(1)}(u_j)u_{i_1}u_{i_2}u_{i_3}u_{i_4}\int_0^1(1-t)^3 f_{i_1,i_2,i_3,i_4}(\vec{y} - t\lambda\vec{u})d\vec{u}\Big].$$

Under the Assumption A, all the $4th$ order partial derivatives of f are uniformly bounded and under Assumption C, $\int K^0(u)|u^j|du$, $\int |K^{0(1)}(u)u^j|du < \infty$, $j = 0, 1, .., 4$ hold and we get the following equations

$$[E[g_{ij}(\vec{y})] - f_{ij}(\vec{y})]^2 \leq \lambda^4 C,\ \forall\ \vec{y} \text{ and } i \neq j, \tag{4 6.1}$$

where C is a constant independent of n, h, λ. It is easy to verify that

$$\begin{aligned} Var\left[g_{ij}(\vec{y})\right] &= \frac{1}{n\lambda^{2d+4}}Var\left[\prod_{l=1,\ l\neq i,j}^{d} K^0\left(\frac{y_l - X_{1l}}{\lambda}\right)K^{0(1)}\left(\frac{y_i - X_{1i}}{\lambda}\right)K^{0(1)}\left(\frac{y_j - X_{1j}}{\lambda}\right)\right] \\ &\leq \frac{||f||}{n\lambda^{d+4}}\left[\int K^2(u)du\right]^{d-2}\left[\int\left[K^{0(1)}(u)\right]^2 du\right]^2,\ \forall\ \vec{y} \text{ and } i \neq j. \end{aligned} \tag{4.6.2}$$

Now

$$E\left[g_{ij}(\vec{y}) - f_{ij}(\vec{y})\right]^2 = Var\left[g_{ij}(\vec{y})\right] + \left[E[g_{ij}(\vec{y})] - f_{ij}(\vec{y})\right]^2. \tag{4.6.3}$$

From (4.6.1), (4.6.2), (4.6.3) we see that,

$$E\left[g_{ij}(\vec{y}) - f_{ij}(\vec{y})\right]^2 \leq \frac{C_1}{n\lambda^{d+4}} + C_2\lambda^4,\ \forall\vec{y} \text{ and } i \neq j,$$

where C_1, C_2 are constants independent of n, h, λ, μ or f. The first case in Lemma 4.6.2(i) follows immediately from the above equation.

The next case $i = j = 1, 2, ..., d$ can be treated similarly. We see that

$$g_{ii}(\vec{y}) = \frac{1}{n\lambda^{d+2}} \sum_{l_1=1}^{n} \prod_{l=1,\ l\neq i}^{d} K^0\left(\frac{y_l - X_{l_1 l}}{\lambda}\right) K^{0(2)}\left(\frac{y_i - X_{l_1 i}}{\lambda}\right).$$

Taking expectation, using Taylor expansion and Lemma 4.6.1, we get

$$E(g_{ii}(\vec{y})) = f_{ii}(\vec{y}) + b_{ii}(\vec{y}), \text{ where } b_{ii}(\vec{y}) \text{ is equal to}$$

$$\frac{\lambda^2}{3!}\Big[\sum_{i_1,i_2,i_3,i_4=1}^{d} \cdot \int_{\vec{u}} \prod_{l=1,\ l\neq i}^{d} K^0(u_l) K^{0(2)}(u_i) u_{i_1} u_{i_2} u_{i_3} u_{i_4} \int_0^1 (1-t)^3 f_{i_1,i_2,i_3,i_4}(\vec{y} - t\lambda\vec{u}) d\vec{u}\Big].$$

The above equation implies that, under the stated assumptions on f, K^0 and its derivatives, we get

$$[E[g_{ii}(\vec{y})] - f_{ii}(\vec{y})]^2 \leq \lambda^4 C, \ \forall\, \vec{y} \text{ and } i,$$

where C is positive constant. It is easy to verify that

$$Var(g_{ii}(\vec{y})) \leq \frac{||f||}{n\lambda^{d+4}} \left[\int [K^0(u)]^2 du\right]^{d-1} \int [K^{0(2)}(u)]^2 du \ \forall\, \vec{y} \text{ and } i.$$

The above two equations imply that

$$\sup_{\vec{y}\in R^d} E\left[g_{ii}(\vec{y})) - f_{ii}(\vec{y})\right]^2 = O\left(\frac{1}{n\lambda^{d+4}} + \lambda^4\right), \ \forall i.$$

So part (i) is proved completely. □

Now we prove part (ii). First we consider the case where $i \neq j = 1, 2, ..., d$. Using $(a+b)^4 \leq 8(a^4 + b^4)$ we see that

$$\begin{aligned} E\left[g_{ij}(\vec{y}) - f_{ij}(\vec{y})\right]^4 &\leq 8E\left[g_{ij}(\vec{y}) - E\left[g_{ij}(\vec{y})\right]\right]^4 \\ &\quad + 8\left[E\left[g_{ij}(\vec{y})\right] - f_{ij}(\vec{y})\right]^4. \end{aligned} \tag{4.6.4}$$

Let $Y_{nl} = \frac{1}{n\lambda^{2+d}}\left[a_l - E(a_l)\right],\ l = 1, 2, ..., n,$
where $a_l = \prod_{i_1=1,\ i_1\neq i,j}^{d} K^0\left(\frac{y_j - X_{l i_1}}{\lambda}\right) K^{0(1)}\left(\frac{y_i - X_{li}}{\lambda}\right) K^{0(1)}\left(\frac{y_j - X_{lj}}{\lambda}\right)$. Then $Y_{n1}, Y_{n2}, ..., Y_{nn}$

are i.i.d random variables, satisfying $E(Y_{n1}) = 0$ and therefore we get the following equation

$$E\left[g_{ij}(\vec{y}) - E\left[g_{ij}(\vec{y})\right]\right]^4 = E\left(\sum_{l=1}^{n} Y_{nl}\right)^4 = nE(Y_{n1}^4) + 6n(n-1)\left[E(Y_{n1}^2)\right]^2. \quad (4.6.5)$$

Let $c_1 = \int [K^{0(1)}(v)]^4 dv$, $c_2 = \int [K^0(v)]^4 dv$, $c_3 = \int |K^{0(1)}(v)| dv$,
$c_4 = \int \left[K^{0(1)}(v)\right]^2 dv$ and $c_5 = \int [K^0(v)]^2 dv$. Then

$$\begin{aligned} E(Y_{n1}^4) &\leq \frac{8}{n^4\lambda^{8+4d}}\left\{E[a_1^4] + [E(a_1)]^4\right\} \\ &\leq \frac{8}{n^4\lambda^{8+4d}}\left[||f||\lambda^d c_1^2 c_2^{d-2} + ||f||^4\lambda^{4d}c_3^4\right] = Oe\left(\frac{1}{n^4\lambda^{8+3d}} + o(1)\right) \\ \text{and } E(Y_{n1}^2) &\leq \frac{1}{n^2\lambda^{4+2d}} E\left(a_1^2\right) \leq \frac{1}{n^2\lambda^{4+d}}||f||c_4^2 c_5^{d-2}. \end{aligned}$$

Substituting the above inequalities in equation (4.6.5) and using $n\lambda^d \to \infty$ we get

$$E\left[g_{ij}(\vec{y}) - E\left[g_{ij}(\vec{y})\right]\right]^4 \leq \frac{6}{n^2\lambda^{8+2d}}\left[||f|||c_4^2 c_5^{d-2}\right]^2 + o\left(\frac{1}{n^2\lambda^{8+2d}}\right). \quad (4.6.6)$$

Further $\left[E\left[g_{ij}(\vec{y})\right] - f_{ij}(\vec{y})\right]^4 \leq \left[E\left[g_{ij}(\vec{y}) - f_{ij}(\vec{y})\right]^2\right]^2$, $\forall \vec{y}$ and $i \neq j$.
Now using Lemma 4.6.2 (i), we get

$$\left[E\left[g_{ij}(\vec{y})\right] - f_{ij}(\vec{y})\right]^4 = O\left(\left\{\frac{1}{n\lambda^{d+4}} + \lambda^4\right\}^2\right), \ \forall \vec{y} \text{ and } i \neq j. \quad (4.6.7)$$

Substituting the equations (4.6.6) and (4.6.7) in the right side of equation (4.6.4) we get

$$E\left[g_{ij}(\vec{y}) - f_{ij}(\vec{y})\right]^4 = O\left(\lambda^8 + \frac{1}{n^2\lambda^{8+2d}}\right), \forall \vec{y} \text{ and } \ i \neq j.$$

The first case in Lemma 4.6.2(ii) follows immediately from the above equation. The case where $i = j = 1, 2, .., d$, can be proved similarly. The only difference is that in this case $a_l = \prod_{i_1=1,\ i_1\neq i}^{d} K^0\left(\frac{y_j - X_{l i_1}}{\lambda}\right) K^{0(2)}\left(\frac{y_i - X_{li}}{\lambda}\right)$, $l = 1, 2, .., d$. □

To prove (iii) we recall, from the proof of 4.6.2 (i), that for $i \neq j = 1, 2, ..., d$,

$$E[g_{ij}(\vec{y})] = f_{ij}(\vec{y}) + b_{ij}(\vec{y}), \text{ where } b_{ij}(\vec{y}) \text{ is equal to}$$

$$\frac{\lambda^2}{3!}\Big[\sum_{i_1,i_2,i_3,i_4=1}^{d} \int_{\vec{u}} \prod_{l=1,\ l\neq i,j}^{d} K^0(u_l)K^{0(1)}(u_i)K^{0(1)}(u_j)u_{i_1}u_{i_2}u_{i_3}u_{i_4}\int_0^1 (1-t)^3 f_{i_1,i_2,i_3,i_4}(\vec{y} - t\lambda\vec{u})dtd\vec{u}\Big].$$

Applying Cauchy-Schwartz inequality it is easy to verify that

$$\begin{aligned} & b_{ij}^2(\vec{y}) \\ \le\ & \frac{4^d\lambda^4}{(3!)^2}[C \sum_{i_1,i_2,i_3,i_4=1}^{d} \int_{\vec{u}} \prod_{l=1,\ l\neq i,j}^{d} |K^0(u_l)K^{0(1)}(u_i)K^{0(1)}(u_j)u_{i_1}u_{i_2}u_{i_3}u_{i_4}| \\ & \int_0^1 (1-t)^3 f_{i_1,i_2,i_3,i_4}^2(\vec{y}-t\lambda\vec{u})dtd\vec{u}] \end{aligned}$$

where $C = \frac{1}{4}\sum_{i_1,i_2,i_3,i_4=1}^{d} \int_{\vec{u}} \prod_{l=1,\ l\neq i,j}^{d} |K^0(u_l)K^{0(1)}(u_i)K^{0(1)}(u_j)u_{i_1}u_{i_2}u_{i_3}u_{i_4}|d\vec{u}$. Consequently, under the Assumption A that all the 4th order partial derivatives are square integrable,

$$\begin{aligned} & \int [E[g_{ij}(\vec{y})] - f_{ij}(\vec{y})]^2\, d\vec{y} \\ \le\ & \frac{4^d\lambda^4}{(3!)^2 4}[C \sum_{i_1,i_2,i_3,i_4=1}^{d} \int \prod_{l=1,\ l\neq i,j}^{d} |K^0(u_l)K^{0(1)}(u_i)K^{0(1)}(u_j)u_{i_1}u_{i_2}u_{i_3}u_{i_4}| \int f_{i_1,i_2,i_3,i_4}^2(\vec{y})d\vec{y}]. \end{aligned} \tag{4.6.8}$$

Under the assumptions that K^0 and $K^{0(1)}$ are square integrable, we see that

$$\begin{aligned} & \int Var\,[g_{ij}(\vec{y})]\, d\vec{y} \\ =\ & \frac{1}{n\lambda^{2d+4}} \int Var\left[\prod_{l=1,\ l\neq i,j}^{d} K^0\left(\frac{y_l - X_{1l}}{\lambda}\right) K^{0(1)}\left(\frac{y_i - X_{1i}}{\lambda}\right) K^{0(1)}\left(\frac{y_j - X_{1j}}{\lambda}\right)\right] d\vec{y} \\ \le\ & \frac{1}{n\lambda^{d+4}} \int\int \left[\prod_{l\neq i,j=1}^{d} K^0(u_j)K^{0(1)}(u_i)K^{0(1)}(u_j)\right]^2 f(\vec{y}-\vec{u}\lambda)d\vec{u}d\vec{y} \\ =\ & \frac{\left[\int [K^0(u)]^2 du\right]^{d-2}\left[\int \left[K^{0(1)}(u)\right]^2 du\right]^2}{n\lambda^{d+4}} = \frac{C'}{n\lambda^{d+4}} \text{ (say)}. \end{aligned} \tag{4.6.9}$$

Now
$\int E\left[g_{ij}(\vec{y}) - f_{ij}(\vec{y})\right]^2 dy = \int Var\left[g_{ij}(\vec{y})\right] dy + \int \left[E[g_{ij}(\vec{y})] - f_{ij}(\vec{y})\right]^2 dy.$
So from (4.6.8) and (4.6.9) we see that

$$\int E\left[g_{ij}(\vec{y}) - f_{i,j}(\vec{y})\right]^2 d\vec{y} \le \frac{C_1}{n\lambda^{d+4}} + C_2\lambda^4, \tag{4.6.10}$$

where C_1, C_2 are positive constants. So the first case in part (iii) is proved.
To prove the other case, that is i=j=1,....,d, we repeat similar arguments which are as follows. We recall that

$E[g_{ii}(\vec{y})] = f_{ii}(\vec{y}) + b_{ii}(\vec{y})$, where $b_{ii}(\vec{y})$ is equal to

$$\frac{\lambda^2}{3!}\Big[\sum_{i_1,i_2,i_3,i_4=1}^{d}\int_{\vec{u}}\prod_{l=1,\ l\neq i}^{d} K^0(u_l)K^{0(2)}(u_i)u_{i_1}u_{i_2}u_{i_3}u_{i_4}\int_0^1(1-t)^3 f_{i_1,i_2,i_3,i_4}(\vec{y}-t\lambda\vec{u})d\vec{u}\Big].$$

Therefore under the Assumptions A and C, on f and K^0, it is easy to verify that

$$\begin{aligned}&\int\left[E[g_{ii}(\vec{y})]-f_{ii}(\vec{y})\right]^2 d\vec{y}\\ \leq\ &\frac{\lambda^4}{(3!)^2 4}\Big[C\sum_{i_1,i_2,i_3,i_4=1}^{d}\int\prod_{l=1,\ l\neq i,j}^{d}|K^0(u_l)K^{0(2)}(u_i)u_{i_1}u_{i_2}u_{i_3}u_{i_4}|\int f^2_{i_1,i_2,i_3,i_4}(\vec{y})d\vec{y}\Big].\end{aligned}$$

Recalling the definition of $g_{ii}(\vec{y})$ we see that

$$\begin{aligned}\int Var\left[g_{ii}(\vec{y})\right]d\vec{y} &= \frac{1}{n\lambda^{2d+4}}\int Var\left[\prod_{l=1,\ l\neq i}^{d}K^0\left(\frac{y_l-X_{1l}}{\lambda}\right)K^{0(2)}\left(\frac{y_i-X_{1i}}{\lambda}\right)\right]d\vec{y}\\ &\leq \frac{\left(\int\left[K^0(u)\right]^2du\right)^{d-1}\int\left[K^{0(2)}(u)\right]^2du}{n\lambda^{d+4}}.\end{aligned}$$

The above equations imply that

$$\int E\left[g_{ii}(\vec{y})-f_{ii}(\vec{y})\right]^2 d\vec{y}=O\left(\frac{1}{n\lambda^{d+4}}+\lambda^4\right).$$

This completes proof of part (iii). Finally we prove part (iv).
We note that the kernel K^0 is a second order kernel and therefore

$$\begin{aligned}E\left[\int\left(w(\vec{y})-f(\vec{y})\right)^2d\vec{y}\right] &= \frac{1}{n\mu^d}\int[\mathbf{K}^0(\vec{u})]^2d\vec{u}+O\left(\frac{1}{n}\right)\\ &+\ \mu^4\int\{\sum_{i_1,\ i_2=1}^{d}\int\mathbf{K}^0(\vec{z})z_{i_1}z_{i_2}\int_0^1(1-t)f_{i_1,\ i_2}(\vec{y}-t\mu\vec{z})dtd\vec{z}\}^2d\vec{y}.\end{aligned} \tag{4.6.11}$$

Under the Assumptions A, C on $f_{i_1,\ i_2},\ i_1,\ i_2=1,2,...,d$ and K^0

$$\begin{aligned}&\int\{\sum_{i_1,\ i_2=1}^{d}\int\mathbf{K}^0(\vec{z})z_{i_1}z_{i_2}\int_0^1(1-t)f_{i_1,\ i_2}(\vec{y}-t\mu\vec{z})dtd\vec{z}\}^2d\vec{y}\\ &-\frac{\left[\int K^0(u)u^2du\right]^2}{4}\int\{\sum_{i=1}^{d}f_{ii}(\vec{y})\}^2d\vec{y}=\int a_{n\vec{y}}.b_{n\vec{y}}d\vec{y},\end{aligned}$$

where

$$
\begin{aligned}
& a_{n\vec{y}} \\
= & (-\mu)\left[\sum_{i_1,\, i_2,\, i_3=1}^{d} \int \mathbf{K}^0(\vec{z}) z_{i_1} z_{i_2} z_{i_3} \int_0^1 (1-t).t \int_0^1 \left\{ \sum_{j=1}^{d} f_{i_1,i_2,i_3}(\vec{y} - st\mu\vec{z}) z_j \right\} ds dt d\vec{z}\right] \\
b_{n\vec{y}} = & \left[\sum_{i_1,\, i_2=1}^{d} \int \mathbf{K}^0(\vec{z}) z_{i_1} z_{i_2} \int_0^1 (1-t)\left\{ f_{i_1,\, i_2}(\vec{y} - t\mu\vec{z}) + f_{i_1,\, i_2}(\vec{y})\right\} dt d\vec{z}\right].
\end{aligned}
$$

Therefore under the Assumption A, that $||f_{i_1,\, i_2}|| < \infty$, $i_1,\, i_2 = 1, 2, .., d$ and all the $3rd$ order partial derivatives of f are integrable, it is easy to verify that $\int a_{n\vec{y}}.b_{n\vec{y}} d\vec{y} = O(\mu)$. Therefore

$$
\begin{aligned}
& \int \{ \sum_{i_1,\, i_2=1}^{d} \mathbf{K}^0(\vec{z}) z_{i_1} z_{i_2} \int_0^1 (1-t) f_{i_1,\, i_2}(\vec{y} - t\mu\vec{z}) dt d\vec{z} \}^2 d\vec{y} \\
= & \frac{\left[\int K^0(u) u^2 du\right]^2}{4} \int \{ \sum_{i=1}^{d} f_{ii}(\vec{y}) \}^2 d\vec{y} + O(\mu). \qquad (4.6.12)
\end{aligned}
$$

Lemma 4.6.2(iv) follows from the equations (4.6.11) and (4.6.12). So Lemma 4.6.2 is proved completely.

Lemma 4.6.3. *Under Assumptions $A - C$ we get*

$$
E\left[(B^*_{\vec{y}})^2 - B^{2}_{\vec{y}}\right]^2 = O\left(h^8 \left\{ \frac{1}{n\lambda^{d+4}} + \lambda^4 \right\}\right).
$$

Proof of Lemma 4.6.3 Recalling the formulae of $B^*_{\vec{y}}$ and $B_{\vec{y}}$, from the proof of Theorem 4.4.1 and the assumptions on K, we see that

$$
\begin{aligned}
\left|(B^*_{\vec{y}})^2 - (B_{\vec{y}})^2\right| = h^4 \Bigg| & \left\{ \sum_{i,j=1}^{d} \int \mathbf{K}(\vec{u}) u_i u_j \int_0^1 (1-t) g_{ij}(\vec{y} - th\vec{u}) dt d\vec{u} \right\}^2 \\
& - \left\{ \sum_{i,j=1}^{d} \int \mathbf{K}(\vec{u}) u_i u_j \int_0^1 (1-t) f_{ij}(\vec{y} - th\vec{u}) dt d\vec{u} \right\}^2 \Bigg| \\
\leq & \ h^4 \, c_{1n}.c_{2n}, \quad \text{where} \\
c_{1n} = \sum_{i,j=1}^{d} \int \mathbf{K}(\vec{u}) & |u_i u_j| \int_0^1 (1-t) \left| g_{ij}(\vec{y} - th\vec{u}) + f_{ij}(\vec{y} - th\vec{u}) \right| dt d\vec{u} \\
c_{2n} = \sum_{i,j=1}^{d} \int \mathbf{K}(\vec{u}) & |u_i u_j| \int_0^1 (1-t) \left| g_{ij}(\vec{y} - th\vec{u}) - f_{ij}(\vec{y} - th\vec{u}) \right| dt d\vec{u}
\end{aligned}
$$

It is easy to see that

$$0 \leq c_{1n}.c_{2n} \leq c_{2n}^2 + 2\left(\frac{1}{2}\sum_{i,j=1}^{d}\int \mathbf{K}(\vec{u})|u_i.u_j|d\vec{u}||f_{ij}||\right)c_{2n}.$$

Therefore

$$\frac{1}{h^8}E\left[\left(B_{\vec{y}}^*\right)^2 - \left(B_{\vec{y}}\right)^2\right]^2 \leq 2\left[E(c_{2n}^4) + (C')^2E(c_{2n}^2)\right] \qquad (4.6.13)$$

where $C' = \sum_{i,j=1}^{d}\int \mathbf{K}(\vec{u})|u_i.u_j|d\vec{u}||f_{ij}||$. Further using Cauchy–Schwartz inequality for c_{2n}^2 and c_{2n}^4 and taking expectation we get

$$\begin{aligned} E(c_{2n}^2) &\leq \frac{(d^2)}{2^2}\sum_{i,j=1}^{d}\left(\int \mathbf{K}(\vec{u})|u_i.u_j|d\vec{u}\right)^2 \sup_{\vec{y}\in R^d} E\left[g_{ij}(\vec{y}) - f_{ij}(\vec{y})\right]^2 \\ \text{and } E(c_{2n}^4) &\leq \frac{(d^6)}{2^4}\sum_{i,j=1}^{d}\left(\int \mathbf{K}(\vec{u})|u_i.u_j|d\vec{u}\right)^4 \sup_{\vec{y}\in R^d} E\left[g_{ij}(\vec{y}) - f_{ij}(\vec{y})\right]^4. \end{aligned}$$

From Lemma 4.6.2 (i) and (ii), we see that

$$\sup_{\vec{y}\in R^d} E\left[g_{ij}(\vec{y}) - f_{ij}(\vec{y})\right]^{2l} = O\left(\left[\frac{1}{n\lambda^{d+4}} + \lambda^4\right]^l\right), \quad l = 1, 2.$$

Therefore from equation (4.6.13), we see that

$$\frac{1}{h^8}E\left[\left(B_{\vec{y}}^*\right)^2 - \left(B_{\vec{y}}\right)^2\right]^2 = O\left(\frac{1}{n\lambda^{d+4}} + \lambda^4\right).$$

Hence Lemma 4.6.3 is proved completely. □

Lemma 4.6.4. *Let the Assumptions A−C and $f(\vec{y}) > 0$ hold. Further if the conditions $h = o(\lambda)$, $n\lambda^{d+4} = o(1)$, $n\lambda^d\mu^4 = o(1)$, $\lambda = o(\mu)$ hold and K is uniformly bounded, then*

$$\sqrt{n\lambda^d}(E_{\vec{y}}^* - E\left[E_{\vec{y}}^*\right]) \rightarrow_L N(0, C_1), \text{ where } C_1 = \frac{1}{\int\left[\boldsymbol{K}^0(\vec{v})\right]^2 d\vec{v} f(\vec{y})}.$$

Proof of Lemma 4.6.4 we recall the definition of $E_{\vec{y}}^*$ and get the following equation

$$\begin{aligned} \sqrt{n\lambda^d}E_{\vec{y}}^* &= \frac{1}{\sqrt{n\lambda^d}}\sum_{i=1}^{n}\int \mathbf{K}(\vec{u})\mathbf{K}^0\left(\frac{(\vec{y} - h\vec{u} - \vec{X}_i)}{\lambda}\right)d\vec{u} \\ &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}Y_{ni}, \text{ where} \\ Y_{ni} &= \frac{1}{\lambda^{d/2}}\int \mathbf{K}(\vec{u})\mathbf{K}^0\left(\frac{(\vec{y} - h\vec{u} - \vec{X}_i)}{\lambda}\right)d\vec{u}. \end{aligned}$$

We note that for fixed n, Y_{ni}, $i = 1, 2, .., n$, are i.i.d non-negative random variables and

$$E\left(|Y_{ni}|^j\right) = \frac{\lambda^d}{\lambda^{(jd)/2}} \int \left[\int \mathbf{K}(\vec{u})\mathbf{K}^0(\vec{v} - \frac{h}{\lambda}\vec{u})d\vec{u}\right]^j f(\vec{y} - \lambda\vec{v})d\vec{v}.$$

Therefore under the stated conditions on K, K^0, λ, h, f and its partial derivatives , we get the following equations

$$\begin{aligned} E(Y_{n1}) &= o(1),\ E(Y_{n1}^2) = \int \left[\mathbf{K}^0(\vec{v})\right]^2 d\vec{v} f(\vec{y}) + o(1) \\ \text{and } E(Y_{n1}^3) &\le \frac{1}{\lambda^{d/2}} C, \end{aligned} \tag{4.6.14}$$

where C is positive constant. It is easy to see that

$$\sqrt{n\lambda^d}(E_{\vec{y}}^* - E\left[E_{\vec{y}}^*\right]) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} Z_{ni}, \text{ where } Z_{ni} = Y_{ni} - E(Y_{ni}),\ i = 1, 2, .., n.$$

We note that for each value of n, Z_{ni}, $i = 1, 2, , , ., n$, are i.i.d mean zero random variables. Recalling the Lyapounov's condition for C.L.T of row-wise i.i.d triangular array (see Billingsly) we get

$$\begin{aligned} &\limsup_{n\to\infty} \frac{1}{\sqrt{n}} \frac{E|Z_{n1}|^3}{[\sqrt{Var(Z_{n1})}]^3} = 0 \\ &\Rightarrow \frac{1}{\sqrt{nVar(Z_{n1})}} \sum_{i=1}^{n} Z_{ni} \to_L N(0, 1) \end{aligned}$$

Now $Var(Z_{n1}) = Var(Y_{n1})$ and $E|Z_{n1}|^3 \le 4[E(Y_{n1}^3) + [E(Y_{n1})]^3]$. Therefore using the above equations (4.6.14), it is easy to verify that

$$\frac{E|Z_{n1}|^3}{[\sqrt{Var(Z_{n1})}]^3} \le \frac{C}{\lambda^{d/2} \int \left[\mathbf{K}^0(\vec{v})\right]^2 d\vec{v} f(\vec{y}) + o(\lambda^{d/2})}.$$

Therefore under the assumptions $n\lambda^d \to \infty$ and $f(\vec{y}) > 0$, we see that $\frac{1}{\sqrt{n}} \frac{E|Z_{n1}|^3}{[\sqrt{Var(Z_{n1})}]^3} = o(1)$. Hence

$$\begin{aligned} &\frac{1}{\sqrt{nVar(Z_{n1})}} \sum_{i=1}^{n} Z_{ni} \to_L N(0, 1) \\ &\Rightarrow \sqrt{n\lambda^d Var(Z_{n1})}(E_{\vec{y}}^* - E_{\vec{y}}) \to_L N(0, 1). \end{aligned}$$

But $Var(Z_{n1}) = Var(Y_{n1}) = \int \left[\mathbf{K}^0(\vec{v})\right]^2 d\vec{v} f(\vec{y}) + o(1)$. Therefore from the above equation we get

$$\sqrt{n\lambda^d}(E_{\vec{y}}^* - E\left[E_{\vec{y}}^*\right]) \to_L N(0, C_1), \text{ where } C_1 = \frac{1}{\int \left[\mathbf{K}^0(\vec{v})\right]^2 d\vec{v} f(\vec{y})}$$

4.6.2 Proofs of Theorems

Proof of theorem 4.4.1 Recall that

$$\begin{aligned} B_{\vec{y}} &= \int \mathbf{K}(\vec{u}) \left[f(\vec{y} - h\vec{u}) - f(\vec{y}) \right] d\vec{u} \\ \text{and } B^*_{\vec{y}} &= \int \mathbf{K}(\vec{u}) \left[g(\vec{y} - h\vec{u}) - g(\vec{y}) \right] d\vec{u} \end{aligned}$$

where $\vec{y} = (y_1, .., y_d)$ and $\vec{u} = (u_1, .., u_d)$. Under Assumptions A to C, expanding $f(\vec{y} - h\vec{u}) - f(\vec{y})$ and $g(\vec{y} - h\vec{u}) - g(\vec{y})$ by Taylor expansion, for functions of several variables, we get

$$B_{\vec{y}} = h^2 \sum_{i,j=1}^{d} \int \mathbf{K}(\vec{u}) u_i u_j \int_0^1 (1-t) f_{ij}(\vec{y} - th\vec{u}) dt d\vec{u} \tag{4.6.15}$$

$$B^*_{\vec{y}} = h^2 \sum_{i,j=1}^{d} \int \mathbf{K}(\vec{u}) u_i u_j \int_0^1 (1-t) g_{ij}(\vec{y} - th\vec{u}) dt d\vec{u}. \tag{4.6.16}$$

Equation (4.6.16) holds almost surely. Under Assumptions A and B, applying DCT in the right side of (4.6.15), it is easy to show that

$$\begin{aligned} B_{\vec{y}} &= \frac{h^2}{2!} \sum_{i,j=1}^{d} a_{ij} f_{ij}(\vec{y}) + o(h^2), \text{ where } a_{i,j} = \int \mathbf{K}(\vec{u}) u_i u_j d\vec{u}. \\ \Rightarrow \frac{B^2_{\vec{y}}}{h^4} &= \frac{1}{(2!)^2} \left[\sum_{i,j=1}^{d} a_{ij} f_{ij}(\vec{y}) \right]^2 + o(1) \\ &= \frac{1}{(2!)^2} \left[\sum_{i=1}^{d} a_{ii} f_{ii}(\vec{y}) \right]^2 + o(1), \end{aligned} \tag{4.6.17}$$

as $a_{i,j} = \left[\int K(u) du\right]^{d-2} \left[\int K(u) u du\right]^2 = 0,\ i \neq j$. From equations (4.6.15) and (4.6.16) it is easy to see that

$$\left[B^*_{\vec{y}} - B_{\vec{y}}\right]^2 \leq \frac{d^2 h^4}{2} c_n \text{ (almost surely), where } c_n \text{ equals}$$

$$\sum_{i,j=1}^{d} a'_{i,j} \int_{\vec{u}} \left[\mathbf{K}(\vec{u}) |u_i u_j| \int_0^1 (1-t) \left[g_{ij}(\vec{y} - th\vec{u}) - f_{ij}(\vec{y} - th\vec{u}) \right]^2 dt \right] d\vec{u},$$

where $a'_{i,j} = \int \mathbf{K}(\vec{u}) |u_i . u_j| d\vec{u}$. Therefore

$$E\left[B^*_{\vec{y}} - B_{\vec{y}}\right]^2 \leq \frac{d^2 h^4}{4} \sum_{i,j=1}^{d} (a'_{i,j})^2 \sup_{\vec{y} \in R^d} E\left[g_{ij}(\vec{y}) - f_{ij}(\vec{y}) \right]^2 \tag{4.6 18}$$

Consequently applying Lemma 4.6.2 (i) in the right side of equation (4.6.18) we get the following equation

$$\frac{E\left[B_{\vec{y}}^{*}-B_{\vec{y}}\right]^{2}}{h^{4}}=O\left(\frac{1}{n\lambda^{d+4}}+\lambda^{4}\right). \tag{4.6.19}$$

Recall that

$$r_1 = E\left[\frac{B_{\vec{y}}^{*}}{B_{\vec{y}}}-1\right]^{2}=\frac{\frac{E\left[B_{\vec{y}}^{*}-B_{\vec{y}}\right]^{2}}{h^{4}}}{\frac{B_{\vec{y}}^{2}}{h^{4}}}. \tag{4.6.20}$$

Therefore under the assumption that $f_{ii}(\vec{y}) \neq 0$ for some $i = 1, 2, ..., d$, Theorem 4.4.1 is a direct consequence of the equations (4.6.17), (4.6.19) and (4.6.20). □

Proof of Theorem 4.4.2 Recalling the definitions of $M_{\vec{y}}$, $M_{\vec{y}}^{*}$ and using $(a+b)^2 \leq 2a^2 + 2b^2$ we find that

$$0 \leq E\left[M_{\vec{y}}^{*}-M_{\vec{y}}\right]^{2} \leq 2E[B_{\vec{y}}^{2}-(B_{\vec{y}}^{*})^{2}]^{2}+2E[V_{\vec{y}}-V_{\vec{y}}^{*}]^{2}. \tag{4.6.21}$$

Recalling the definitions of $V_{\vec{y}}$ and $V_{\vec{y}}^{*}$ we get the following equation

$$\begin{aligned} V_{\vec{y}}^{*}-V_{\vec{y}} &= \frac{1}{nh^{d}}\int_{\vec{u}}\mathbf{K}^{2}(\vec{u})[w(\vec{y}-h\vec{u})-f(\vec{y}-h\vec{u})]d\vec{u} \\ &\quad -\frac{1}{n}\left[\left(\int_{\vec{u}}\mathbf{K}(\vec{u})w(\vec{y}-h\vec{u})d\vec{u}\right)^{2}-\left(\int_{\vec{u}}\mathbf{K}(\vec{u})f(\vec{y}-h\vec{u})d\vec{u}\right)^{2}\right] \\ &= L_1 - L_2 \quad \text{(say)}. \end{aligned}$$

Hence

$$E[V_{\vec{y}}^{*}-V_{\vec{y}}]^{2} \leq 2E(L_1^2)+2E(L_2^2). \tag{4.6.22}$$

Now

$$E(L_1^2) \leq \frac{\left(\int_{\vec{u}}\mathbf{K}^{2}(\vec{u})d\vec{u}\right)^{2}}{(nh^{d})^{2}} \sup_{\vec{y}\in R^{d}} E[w(\vec{y})-f(\vec{y})]^{2}.$$

It is easy to verify that,

$$E\left[f(\vec{y})-w(\vec{y})\right]^{2} \leq \frac{||f||}{n\mu^{d}}\int_{\vec{u}}\left[\mathbf{K}^{0}(\vec{u})\right]^{2}d\vec{u} + \left[\frac{\sum_{i,j=1}^{d}\left[\int \mathbf{K}^{0}(\vec{u})u_i u_j d\vec{u}\right]^{2}||f_{ij}||\mu^{2}}{2!}\right]^{2}.$$

So

$$\sup_{\vec{y}\in R^{d}} E\left[f(\vec{y})-w(\vec{y})\right]^{2}=O\left(\frac{1}{n\mu^{d}}+\mu^{4}\right). \tag{4.6.23}$$

Therefore

$$E(L_1^2) = O\left(\frac{1}{n^2h^{2d}}\left[\frac{1}{n\mu^d} + \mu^4\right]\right). \tag{4.6.24}$$

Now using $a^2 - b^2 = (a-b)(a+b)$, properties of K and Cauchy-Schwartz inequality

$$n^2EL_2^2 \leq E(c_n.d_n), \quad \text{where}$$

$$c_n = \int \mathbf{K}(\vec{u})\left[f(\vec{y}-h\vec{u}) + w(\vec{y}-h\vec{u})\right]^2 d\vec{u}$$

$$\text{and} \quad d_n = \int \mathbf{K}(\vec{u})\left[f(\vec{y}-h\vec{u}) - w(\vec{y}-h\vec{u})\right]^2 d\vec{u}.$$

Since $|f(\vec{y}-h\vec{u}) + w(\vec{y}-h\vec{u})| \leq |f(\vec{y}-h\vec{u}) - w(\vec{y}-h\vec{u})| + 2||f||$, for all $\vec{y}$, therefore it is easy to see that $c_nd_n \leq 2d_n^2 + 8||f||^2d_n$ and hence

$$n^2EL_2^2 \leq E(c_n.d_n) \leq \left[2E(d_n^2) + 8||f||^2E(d_n)\right]. \tag{4.6.25}$$

Further it is easy verify that

$$0 \leq E(d_n^j) \leq C^{jd}\sup_{\vec{y}\in R^d} E[w(\vec{y}) - f(\vec{y})]^{2j} \quad j = 1,2.$$

Recalling equation (4.6.23) and by some straight forward algebra, it is easy to verify that

$$\sup_{\vec{y}\in R^d} E[w(\vec{y}) - f(\vec{y})]^4 = O\left(\left[\frac{1}{n\mu^d} + \mu^4\right]^2\right).$$

Consequently $E(d_n^2) = O\left(\left[\frac{1}{n\mu^d} + \mu^4\right]^2\right)$ and hence recalling equations (4.6.23) and (4.6.25) we get

$$E(L_2^2) = O\left(\frac{1}{n^2}\left[\frac{1}{n\mu^d} + \mu^4\right]\right) = o\left(\frac{1}{n^2h^{2d}}\left[\frac{1}{n\mu^d} + \mu^4\right]\right). \tag{4.6.26}$$

From equations (4.6.22), (4.6.24) and (4.6.26) we get

$$E[V_{\vec{y}}^* - V_{\vec{y}}]^2 = O\left(\frac{1}{n^2h^{2d}}\left[\frac{1}{n\mu^d} + \mu^4\right]\right). \tag{4.6.27}$$

From equations (4.6.21), (4.6.27) and Lemma 4.6.3 it is easy to see that

$$E\left[M_{\vec{y}}^* - M_{\vec{y}}\right]^2 = O\left(h^8\left[\frac{1}{n\lambda^{d+4}} + \lambda^4\right] + \frac{1}{n^2h^{2d}}\left[\frac{1}{n\mu^d} + \mu^4\right]\right). \tag{4.6.28}$$

Using Assumptions A on f and B on K it is easy to prove that

$$M_{\vec{y}}^2 = \left[\frac{f(\vec{y})}{nh^d}\left\{\int K^2(u)du\right\}^d + \frac{h^4}{(2!)^2}\left[\int K(u)u^2\sum_{i=1}^{d} f_{ii}(\vec{y})\right]^2 + o\left(\frac{1}{nh^d} + h^4\right)\right]^2. \tag{4.6.29}$$

Recall that

$$r_3 = \frac{E\left[M_{\vec{y}}^* - M_{\vec{y}}\right]^2}{M_{\vec{y}}^2}. \tag{4.6.30}$$

If $\liminf_{n\to\infty} nh^{4+d} > 0$ then under the assumption that $|f_{ii}\vec{y})| > 0$, for some i, dividing numerator and denominator of r_3 by h^8 we get, from equations (4.6.28), (4.6.29) and (4.6.30), that

$$r_3 = O\left(\left[\frac{1}{n\lambda^{d+4}} + \lambda^4\right] + \left[\frac{1}{n\mu^d} + \mu^4\right]\right).$$

If $\limsup_{n\to\infty} nh^{4+d} < \infty$ then under the assumption $f(\vec{y}) > 0$, dividing numerator and denominator of r_3 by $\frac{1}{(nh^d)^2}$ we get, from equations (4.6.28), (4.6.29) and (4.6.30), that

$$r_3 = O\left(\left[\frac{1}{n\lambda^{d+4}} + \lambda^4\right] + \left[\frac{1}{n\mu^d} + \mu^4\right]\right).$$

So Theorem 4.4.2 is proved completely. □

Proof of Theorem 4.4.3 Recalling the definition of r we see that

$$\begin{aligned} r &= \frac{E_{\vec{y}}\sqrt{V_{\vec{y}}^*}}{E_{\vec{y}}^*\sqrt{V_{\vec{y}}}} - 1 \\ &= \frac{E_{\vec{y}}}{E_{\vec{y}}^*}\left(\sqrt{\frac{V_{\vec{y}}^*}{V_{\vec{y}}}} - 1\right) + \frac{E_{\vec{y}}}{E_{\vec{y}}^*} - 1 \end{aligned} \tag{4.6.31}$$

Recalling the definition of $V_{\vec{y}}$ and assuming $f(\vec{y}) > 0$ it is easy to verify that,

$$\liminf_{n\to\infty} nh^d V_{\vec{y}} \geq \left[\int K^2(u)du\right]^d f(\vec{y}) > 0.$$

Recalling equation (4.6.27) and using the above inequality it is easy to verify that

$$\begin{aligned} E\left[\frac{V_{\vec{y}}^*}{V_{\vec{y}}} - 1\right]^2 &= O\left(\frac{1}{n\mu^d} + \mu^4\right). \\ \Rightarrow \left|\frac{V_{\vec{y}}^*}{V_{\vec{y}}} - 1\right| &= O_P\left(\sqrt{\frac{1}{n\mu^d} + \mu^4}\right) \end{aligned}$$

Therefore

$$\begin{aligned} \left|\sqrt{\frac{V_{\vec{y}}^*}{V_{\vec{y}}}} - 1\right| &= \left|\frac{V_{\vec{y}}^*}{V_{\vec{y}}} - 1\right| \frac{1}{(\sqrt{V_{\vec{y}}^*/V_{\vec{y}}} + 1)} \\ &= O_P\left(\sqrt{\frac{1}{n\mu^d} + \mu^4}\right). \end{aligned} \tag{4.6.32}$$

Recalling the definitions of $E_{\vec{y}}$, $E^*_{\vec{y}}$ and that K is probability density function it is easy to verify that

$$\begin{aligned} E\left[E^*_{\vec{y}} - E_{\vec{y}}\right]^2 &\leq \int \mathbf{K}(\vec{u}) E\left[g(\vec{y} - h\vec{u}) - f(\vec{u} - h\vec{u})\right]^2 d\vec{u} \\ &\leq \sup_{\vec{y}} E\left[g(\vec{y}) - f(\vec{y})\right]^2. \end{aligned}$$

Under the assumption that the 2nd order partial derivatives of f are uniformly bounded (Assumption A) it is easy to verify that

$$E\left[g(\vec{y}) - f(\vec{y})\right]^2 \leq \frac{C_1}{n\lambda^d} + C_2\lambda^4,$$

where C_1, C_2 are positive constants, independent of n, h, λ, μ.

Therefore from the previous two equations we get the following equation

$$\begin{aligned} E\left[E^*_{\vec{y}} - E_{\vec{y}}\right]^2 &= O\left(\frac{C_1}{n\lambda^d} + C_2\lambda^4\right) \\ \Rightarrow \left|E^*_{\vec{y}} - E_{\vec{y}}\right| &= O_P\left(\sqrt{\frac{C_1}{n\lambda^d} + C_2\lambda^4}\right). \end{aligned} \quad (4.6.33)$$

Recalling the definition of $E_{\vec{y}}$ and the stated assumptions on f it is easy to verify that

$$\lim_{n\to\infty} E_{\vec{y}} = f(\vec{y}).$$

Therefore $E^*_{\vec{y}} = E^*_{\vec{y}} - E_{\vec{y}} + E_{\vec{y}} = f(\vec{y}) + o_P(1)$ and hence under the assumption that $f(\vec{y}) > 0$

$$\left|\frac{E_{\vec{y}}}{E^*_{\vec{y}}} - 1\right| = O_P\left(\sqrt{\frac{C_1}{n\lambda^d} + C_2\lambda^4}\right). \quad (4.6.34)$$

From equations (4.6.31), (4.6.32) and (4.6.34) we see that

$$r = O_P\left(\sqrt{\frac{C_1}{n\lambda^d} + C_2\lambda^4} + \sqrt{\frac{1}{n\mu^d} + \mu^4}\right).$$

So part (i) is proved.

To prove part (ii) we recall Lemma 4.6.4 and get the following equation

$$\sqrt{n\lambda^d}\left(E^*_{\vec{y}} - E\left(E^*_{\vec{y}}\right)\right) \to_L N(0, C_1), \text{ where } C_1 = \frac{1}{\int \left[\mathbf{K}^0(\vec{v})\right]^2 d\vec{v} f(\vec{y})}. \quad (4.6.35)$$

Recalling the definition of $E^*_{\vec{y}}$ it is easy to verify that

$$\left|E(E^*_{\vec{y}}) - E_{\vec{y}}\right| \leq \int \mathbf{K}(\vec{u}) d\vec{u} \sup_{\vec{y}\in R^d} \left|E(g(\vec{y})) - f(\vec{y})\right|.$$

Further recalling that K^0 is a second order kernel we see that

$$E(g(\vec{y})) = f(\vec{y}) - \frac{\lambda^2}{2} \sum_{i_1, i_2 = 1}^{d} \int \mathbf{K}^0(\vec{u}) u_{i_1} u_{i_2} \int_0^1 (1-t) f_{i_1,\, i_2}(\vec{y} - t\lambda\vec{u}) dt d\vec{u}.$$

Therefore under the Assumption A that all the second order partial derivatives are uniformly bounded we see that

$$\sup_{\vec{y} \in R^d} |E(g(\vec{y})) - f(\vec{y})| = O\left(\lambda^2\right).$$

Therefore

$$\sqrt{n\lambda^d} \left| E\left(E^*_{\vec{y}}\right) - E_{\vec{y}} \right| = O\left(\sqrt{n\lambda^{d+4}}\right). \tag{4.6.36}$$

Therefore if $n\lambda^d \to \infty$, $n\lambda^{4+d} = o(1)$ and $f(\vec{y}) > 0$, then from equations (4.6.35) and (4.6.36) we see that

$$\sqrt{n\lambda^d} \left(E^*_{\vec{y}} - E_{\vec{y}}\right) \to_L N\left(0, C_1\right). \tag{4.6.37}$$

The above equation implies that, $E^*_{\vec{y}} - E_{\vec{y}} = o_P(1)$. Under the assumptions that f is continuous and uniformly bounded and recalling the definition of $E_{\vec{y}}$ it is easy to show that $E_{\vec{y}} \to f(\vec{y})$. Therefore

$$E^*_{\vec{y}} - f(\vec{y}) = o_P(1).$$

Hence if $f(\vec{y}) > 0$, under the stated conditions on λ and f, we get (from the above equation and (4.6.37)) the following equation

$$\sqrt{n\lambda^d} \left(\frac{E_{\vec{y}}}{E^*_{\vec{y}}} - 1\right) \to_L N\left(0, \frac{C_1}{f^2(\vec{y})}\right). \tag{4.6.38}$$

Recalling equation (4.6.32), under the assumptions that $\lim_{n\to\infty} n\lambda^d \mu^4 = 0$, $\lambda = o(\mu)$ and $\limsup_{n\to\infty} \frac{h}{\lambda} = 0$, we see that

$$\limsup_{n\to\infty} \sqrt{n\lambda^d} \left| \sqrt{\frac{V^*_{\vec{y}}}{V_{\vec{y}}}} - 1 \right| = 0. \tag{4.6.39}$$

Therefore under the stated assumptions on λ, μ, K and K^0, recalling equations (4.6.31), (4.6.38) and (4.6.39), we see that

$$\sqrt{n\lambda^d} r \to_L N(0, \frac{C_1}{f^2(\vec{y})}),$$

where $C_1 = \frac{1}{\int \left[\mathbf{K}^0(\vec{v})\right]^2 d\vec{v} f(\vec{y})}$. This completes the proof of part (ii) and hence Theorem 4.4.3 is proved completely. □

Proof of Theorem 4.4.4 Recall that

$$\begin{aligned} M &= \int \left[V_{\vec{y}} + B_{\vec{y}}^2\right] d\vec{y} = V + B \text{ (say)} \\ \text{and } M^* &= \int \left[V_{\vec{y}}^* + (B_{\vec{y}}^*)^2\right] d\vec{y} = V^* + B^* \text{ (say)} \\ \text{Therefore } E\left|M^* - M\right| &\leq E\left|V^* - V\right| + E\left|B^* - B\right|. \qquad (4.6.40) \end{aligned}$$

$V^* - V = \int [V_{\vec{y}}^* - V_{\vec{y}}] d\vec{y}$. Integrating the expressions of $V_{\vec{y}}$, $V_{\vec{y}}^*$ we see that

$$\begin{aligned} \int V_{\vec{y}} d\vec{y} &= \frac{1}{nh^d} \int \mathbf{K}^2(\vec{u}) d\vec{u} - \frac{1}{n} \int \left(\int \mathbf{K}(\vec{u}) f(\vec{y} - h\vec{u}) d\vec{u} \right)^2 d\vec{y} \\ \int V_{\vec{y}}^* d\vec{y} &= \frac{1}{nh^d} \int \mathbf{K}^2(\vec{u}) d\vec{u} - \frac{1}{n} \int \left(\int \mathbf{K}(\vec{u}) w(\vec{y} - h\vec{u}) d\vec{u} \right)^2 d\vec{y} \\ \Rightarrow E|V^* - V| &= \frac{1}{n} E \left| \int \left\{ \left(\int_{\vec{u}} \mathbf{K}(\vec{u}) w(\vec{y} - h\vec{u}) d\vec{u} \right)^2 - \left(\int_{\vec{u}} \mathbf{K}(\vec{u}) f(\vec{y} - h\vec{u}) d\vec{u} \right)^2 \right\} d\vec{y} \right|. \end{aligned}$$

Therefore

$$\begin{aligned} E\left|V^* - V\right| &\leq \frac{1}{n} E\left(\int |a_{\vec{y}} b_{\vec{y}}| d\vec{y} \right) \leq \frac{1}{n} \left\{ E(\int a_{\vec{y}}^2 d\vec{y}) + 2E(\int |a_{\vec{y}} c_{\vec{y}}| \, d\vec{y}) \right\} \\ &\leq \frac{1}{n} \left\{ E(\int a_{\vec{y}}^2 d\vec{y}) + 2E \sqrt{\int a_{\vec{y}}^2 d\vec{y} \int c_{\vec{y}}^2 d\vec{y}} \right\} \\ &\leq \frac{1}{n} \left\{ E(\int a_{\vec{y}}^2 d\vec{y}) + 2 \sqrt{E\left(\int a_{\vec{y}}^2 d\vec{y} \right) \int c_{\vec{y}}^2 d\vec{y}} \right\} \qquad (4.6.41) \\ \text{where } a_{\vec{y}} &= \int_{\vec{u}} \mathbf{K}(\vec{u}) \left[w(\vec{y} - h\vec{u}) - f(\vec{y} - h\vec{u})\right] d\vec{u} \\ b_{\vec{y}} &= \int_{\vec{u}} \mathbf{K}(\vec{u}) \left[w(\vec{y} - h\vec{u}) + f(\vec{y} - h\vec{u})\right] d\vec{u} \\ \text{and } c_{\vec{y}} &= \int \mathbf{K}(\vec{u}) f(\vec{y} - h\vec{u}) d\vec{u}. \end{aligned}$$

We note that $E\left(\int a_{\vec{y}}^2 d\vec{y}\right) \leq E\left[\int \left(w(\vec{y}) - f(\vec{y})\right)^2 d\vec{y}\right]$.
From Lemma 4.6.2(iv) we see that, $E\left(\int a_{\vec{y}}^2 d\vec{y}\right) = O\left(\frac{C_1}{n\mu^d} + C_2\mu^4\right)$, where C_1, C_2 are positive constants and substituting this in the right side of (4.6.41) we get

$$E\left|V^* - V\right| = O\left(\frac{1}{n}\sqrt{\frac{1}{n\mu^d} + \mu^4}\right). \qquad (4.6.42)$$

Since $B^* - B = \int \left[\left(B_{\vec{y}}^*\right)^2 - \left(B_{\vec{y}}\right)^2\right] d\vec{y}$, therefore integrating the expressions for $B_{\vec{y}}^*$ and

$B_{\vec{y}}$ (see equations (4.6.15) and (4.6.16)) and taking expectation we get

$$\begin{aligned} E\left|B^{*}-B\right| &\leq h^{4}\int E\left(f_{1\vec{y}}f_{2\vec{y}}\right)d\vec{y} \\ &\leq h^{4}\int\left[E\left(f_{2\vec{y}}^{2}\right)+2E\left(f_{2\vec{y}}f_{3\vec{y}}\right)\right]d\vec{y} \qquad (4.6.43) \end{aligned}$$

$$\begin{aligned} \text{where } f_{1\vec{y}} &= \sum_{i_1,i_2=1}^{d}\int \mathbf{K}(\vec{u})|u_{i_1}u_{i_2}|\int_0^1(1-t)\left|f_{i_1,i_2}(\vec{y}-th\vec{u})+g_{i_1,i_2}(\vec{y}-th\vec{u})\right|dtd\vec{u} \\ f_{2\vec{y}} &= \sum_{i_1,i_2=1}^{d}\int \mathbf{K}(\vec{u})|u_{i_1}u_{i_2}|\int_0^1(1-t)\left|f_{i_1,i_2}(\vec{y}-th\vec{u})-g_{i_1,i_2}(\vec{y}-th\vec{u})\right|dtd\vec{u} \\ \text{and } f_{3\vec{y}} &= \sum_{i_1,i_2=1}^{d}\int \mathbf{K}(\vec{u})|u_{i_1}u_{i_2}|\int_0^1(1-t)\left|f_{i_1,i_2}(\vec{y}-th\vec{u})\right|dtd\vec{u} \qquad (4.6.44) \end{aligned}$$

Using Cauchy- Schewartz inequality it is easy to verify that

$$\int E\left(f_{2\vec{y}}^{2}\right)d\vec{y}\leq d^{2}\sum_{i_1,i_2=1}^{d}\frac{\left[\int \mathbf{K}(\vec{u})|u_{i_1}u_{i_2}|d\vec{u}\right]^{2}}{2^{2}}\int E[f_{i_1,i_2}(\vec{y})-g_{i_1,i_2}(\vec{y})]^{2}d\vec{y}.$$

Using Lemma 4.6.2(iii) in the right side of the above inequality we get

$$\int E\left(f_{2\vec{y}}^{2}\right)d\vec{y}=O\left(\frac{1}{n\lambda^{d+4}}+\lambda^{4}\right). \qquad (4.6.45)$$

Further we note that

$\left[\int E\left(f_{2\vec{y}}f_{3\vec{y}}\right)d\vec{y}\right]^{2}\leq E\left[\int\left(f_{2\vec{y}}f_{3\vec{y}}\right)d\vec{y}\right]^{2}\leq\int E(f_{2\vec{y}}^{2})f_{3\vec{y}}d\vec{y}\int f_{3\vec{y}}d\vec{y}\leq C\int E(f_{2\vec{y}}^{2})d\vec{y}$

where

$C=\frac{1}{2^{2}}\left[\sum_{i_1,i_2=1}^{d}\int \mathbf{K}(\vec{u})|u_{i_1}u_{i_2}|d\vec{u}\|f_{i_1,i_2}\|\right]\left[\sum_{i_1,i_2=1}^{d}\int \mathbf{K}(\vec{u})|u_{i_1}u_{i_2}|d\vec{u}\int|f_{i_1,i_2}(\vec{y})|d\vec{y}\right].$

Therefore it is easy to see that

$$\int E\left(f_{2\vec{y}}f_{3\vec{y}}\right)d\vec{y}=O\left(\sqrt{\frac{1}{n\lambda^{d+4}}+\lambda^{4}}\right). \qquad (4.6.46)$$

From equations (4.6.43), (4.6.45), (4.6.46) we see that

$$E|B^{*}-B|=O\left(h^{4}\sqrt{\frac{1}{n\lambda^{d+4}}+\lambda^{4}}\right). \qquad (4.6.47)$$

From the equations (4.6.40), (4.6.42) and (4.6.47) we see that

$$E\left|M^{*}-M\right|=O\left(\frac{1}{n}\sqrt{\frac{1}{n\mu^{d}}+\mu^{4}}+h^{4}\sqrt{\frac{1}{n\lambda^{d+4}}+\lambda^{4}}\right). \qquad (4.6.48)$$

Under Assumptions A, B on f and K and for $s = 2$ it is easy to show that

$$M = \frac{D_1}{nh^d} + D_2 h^4 + o\left(\frac{1}{nh^d} + h^4\right), \; D_1, D_2 > 0.$$

Therefore from equation (4.6.48) and the above equation we get

$$r_4 = \frac{O\left(\frac{1}{n}\sqrt{\frac{1}{n\mu^d} + \mu^4} + h^4\sqrt{\frac{1}{n\lambda^{d+4}} + \lambda^4}\right)}{\frac{D_1}{nh^d} + D_2 h^4 + o\left(\frac{1}{nh^d} + h^4\right)}.$$

If $\liminf_{n\to\infty} nh^{4+d} > 0$ then dividing numerator and denominator of r_4 by h^4 we get, $r_4 = O\left(h^4\sqrt{\frac{1}{n\mu^d} + \mu^4} + \sqrt{\frac{1}{n\lambda^{d+4}} + \lambda^4}\right)$.
If $\limsup_{n\to\infty} nh^{4+d} = 0$, dividing numerator and denominator of r_3 by $\frac{1}{(nh^d)^2}$ we get, $r_4 = O\left(h^4\sqrt{\frac{1}{n\mu^d} + \mu^4} + \sqrt{\frac{1}{n\lambda^{d+4}} + \lambda^4}\right)$. □

Proof of Theorem 4.4.5 Recalling that $M = V + B$ and $M^* = V^* + B^*$, where V, V^*, B and B^* are as defined earlier. We see that, almost surely

$$0 \le V,\ V^* \le \frac{1}{nh^d}\int \mathbf{K}^2(\vec{u})d\vec{u}.$$

Therefore $V,\ V^* \to 0$, as $h \to \infty$.
Now

$$\begin{aligned} B &= \int\left\{\int \mathbf{K}(\vec{u})f(\vec{y} - h\vec{u})d\vec{u} - f(\vec{y})\right\}^2 d\vec{y} \\ &= \int f^2(\vec{y})d\vec{y} + \int\left[\int \mathbf{K}(\vec{u})f(\vec{y} - h\vec{u})d\vec{u}\right]^2 d\vec{y} \\ &\qquad -2\int f(\vec{y})\int \mathbf{K}(\vec{u})f(\vec{y} - h\vec{u})d\vec{u}d\vec{y} \\ &= \int f^2(\vec{y})d\vec{y} + b_n - c_n \text{ (say)}. \end{aligned}$$

We note that, $0 \le b_n \le \frac{1}{h^d}\int \mathbf{K}^2(\vec{u})d\vec{u} = o(1)$, as $h \to \infty$.
As $h \to \infty$, $||\vec{y} - h\vec{u}|| \to \infty$ for $\vec{u} \ne 0$. Therefore using the stated conditions on f and DCT it is easy to show that $c_n = o(1)$, as $h \to \infty$. Hence under the stated conditions on f, $B \to \int f^2(\vec{y})d\vec{y}$ and consequently $M \to \int f^2(\vec{y})d\vec{y}$.
Further we note that, almost surely,

$$\begin{aligned} B^* &= \int g^2(\vec{y})d\vec{y} + \int \mathbf{K}(\vec{u})\left[\int g(\vec{y} - \vec{h}.\vec{u})d\vec{u}\right]^2 d\vec{y} \\ &\qquad -2\int g(\vec{y})\mathbf{K}(\vec{u})g(\vec{y} - \vec{h}.\vec{u})d\vec{u}d\vec{y} \\ &= \int g^2(\vec{y})d\vec{y} + b_n^* - c_n^* \text{ (say)}, \end{aligned}$$

where $0 \leq b_n^* \leq \frac{1}{\prod_{j=1}^d h_j} \int \mathbf{K}^2(\vec{u})d\vec{u} = o(1)$, as $h \to \infty$.
Recalling the definition of $g(\vec{y})$, we see that $g(\vec{y}-h\vec{u}) = \frac{1}{n\lambda^d}\sum_{i=1}^n \prod_{j=1}^d K^0\left(\frac{y_j-hu_j-X_{ij}}{\lambda}\right)$.
Therefore for fixed n, λ, under the stated condition on K^0, $g(\cdot)$ is uniformly bounded and $g(\vec{y}-h\vec{u}) \to 0$, almost surely, $h \to \infty$. Therefore using DCT it is easy to argue that $c_n^* \to 0$, almost surely, as h is increased.
So under the stated conditions on K^0 and for fixed λ and μ, almost surely, $B^* \to \int g^2(\vec{y})d\vec{y}$ and hence $M^* \to \int g^2(\vec{y})d\vec{y}$, as h is increased. So part (i) is proved.
The proofs of part (ii) and (iv) follow directly from the definitions of A, $\hat{A}$, $A_{\vec{y}}$ and $\hat{A}_{\vec{y}}$ respectively.
The proof of part (iii) follows by repeating similar arguments as used in the proof of part (i). So Theorem 4.4.5 is proved completely. □

Proof of Theorem 4.4.6 Suppose K^0 is the standard normal density. Recalling the definitions of g and w we see that $g(\vec{y})$ and $w(\vec{y})$ are equal to $g(\vec{y}) = \frac{1}{n}\sum_{i=1}^n \phi_{\vec{X}_i,\lambda^2 I_d}(\vec{y})$ and $w(\vec{y}) = \frac{1}{n}\sum_{i=1}^n \phi_{\vec{X}_i,\mu^2 I_d}(\vec{y})$ respectively.
If K is standard normal density, then we see that $K_n(\vec{y}) = \frac{1}{n}\sum_{i=1}^n \phi_{\vec{0},D}(\vec{y}-\vec{X}_i)$.
Consequently if both K, K^0 are both standard normal densities then

$$\begin{aligned} B_{\vec{y}}^* &= \int \phi_{\vec{0},D}(\vec{y}-\vec{u})g(\vec{u})du - g(\vec{y}) \\ &= \frac{1}{n}\sum_{i=1}^n \int \phi_{\vec{0},D}(\vec{y}-\vec{u})\phi_{\vec{X}_i,\lambda^2 I_d}(\vec{u})d\vec{u} - g(\vec{y}) \\ &= \frac{1}{n}\sum_{i=1}^n \phi_{\vec{X}_i,D+\lambda^2 I_d}(\vec{y}) - g(\vec{y}), \end{aligned}$$

where $D = diag(h_1^2, ..., h_d^2)$. So part one is proved.
To prove the second part, we note that if $K(\cdot)$ is standard normal density, then $K^2(u) = \frac{1}{2\sqrt{\pi}}\phi_{0,1/2}(u)$. So $\prod_{j=1}^d K^2(u_j) = \frac{1}{(2\sqrt{\pi})^d}\phi_{\vec{0},\frac{1}{2}I_d}(\vec{u})$. Consequently, if K is standard normal density then recalling the definition of $V_{\vec{y}}^*$ we see that

$$V_{\vec{y}}^* = \frac{1}{n(2\sqrt{\pi})^d \prod_{j=1}^d h_j}\int \phi_{\vec{0},\frac{1}{2}D}(\vec{y}-\vec{u})w(\vec{u})d\vec{u} - \frac{1}{n}\left[\int \phi_{\vec{0},D}(\vec{y}-\vec{u})w(\vec{u})d\vec{u}\right]^2.$$

Now if K^0 is a standard normal density then $w(\vec{u}) = \frac{1}{n}\sum_{i=1}^n \phi_{\vec{X}_i,\mu^2 I_d}(\vec{u})$. Substituting this in the right side of above equation we get the formula for $V_{\vec{y}}$. So Theorem 4.4.6 is proved completely. □

Proof of Theorem 4.4.8 We recall that if K and K^0 are standard normal densities then $w(\vec{y}) = \frac{1}{n}\sum_{i=1}^{n}\phi_{\vec{X}_i,\mu^2 I_d}(\vec{y})$. Further recalling the definition of V^* we see that

$$V^* = \frac{1}{n\prod_{j=1}^{d}h_j}\int\prod_{j=1}^{d}K^2(u_j)d\vec{u} - \frac{1}{n}\int\left[\frac{1}{\prod_{j=1}^{d}h_j}\int\prod_{j=1}^{d}K\left(\frac{y_j-u_j}{h_j}\right)w(\vec{u})d\vec{u}\right]^2 d\vec{y}.$$

Therefore, if K and K^0 are standard normal densities then

$$\begin{aligned}
V^* &= \frac{1}{n\prod_{j=1}^{d}h_j}\int\frac{1}{2^d(\sqrt{\pi})^d}\phi_{\vec{0},\frac{1}{2}I_d}(\vec{u})d\vec{u} - \frac{1}{n}\int\left[\frac{1}{n}\sum_{i=1}^{n}\int\phi_{\vec{0},D}(\vec{y}-\vec{u})\phi_{\vec{X}_i,\mu^2 I_d}(\vec{u})d\vec{u}\right]^2 d\vec{y} \\
&= \frac{1}{n(4\pi)^{d/2}\prod_{j=1}^{d}h_j} - \frac{1}{n}\int\left[\frac{1}{n}\sum_{i=1}^{n}\phi_{\vec{X}_i,D+\mu^2 I_d}(\vec{y})\right]^2 d\vec{y} \\
&= \frac{1}{n(4\pi)^{d/2}\prod_{j=1}^{d}h_j} - \frac{1}{n}\left[\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\int\phi_{\vec{X}_i,D+\mu^2 I_d}(\vec{y})\phi_{\vec{X}_j,D+\mu^2 I_d}(\vec{y})d\vec{y}\right] \\
&= \frac{1}{n(4\pi)^{d/2}\prod_{j=1}^{d}h_j} - \frac{1}{n}\left[\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\phi_{\vec{X}_i-\vec{X}_j,2D+2\mu^2 I_d}(\vec{0})\right] \\
&= \frac{1}{n(4\pi)^{d/2}\prod_{j=1}^{d}h_j} - \frac{1}{n}\vec{w_1}\Omega_2^{**}(\vec{w_1})'
\end{aligned}$$

Similarly recalling that $g(\vec{y}) = \frac{1}{n}\sum_{i=1}^{n}\phi_{\vec{X}_i,\lambda^2 I_d}(\vec{y})$, K^0 is Gaussian kernel and the definition of B^* we see that

$$\begin{aligned}
B^* &= \int\left[\int \phi_{\vec{0},D}(\vec{y}-\vec{u})g(\vec{u})d\vec{u} - g(\vec{y})\right]^2 d\vec{y} \\
&= \int\left[\int \phi_{\vec{0},D}(\vec{y}-\vec{u})g(\vec{u})d\vec{u}\right]^2 d\vec{y} + \int [g(\vec{y})]^2 d\vec{y} - 2\int\left[\int \phi_{\vec{0},D}(\vec{y}-\vec{u})g(\vec{u})d\vec{u}\right]g(\vec{y})d\vec{y} \\
&= \int\left[\frac{1}{n}\sum_{i=1}^{n}\int \phi_{\vec{0},D}(\vec{y}-\vec{u})\phi_{\vec{X}_i,\lambda^2 I_d}(\vec{u})d\vec{u}\right]^2 d\vec{y} + \int\left[\frac{1}{n}\sum_{i=1}^{n}\phi_{\vec{X}_i,\lambda^2 I_d}(\vec{y})\right]^2 d\vec{y} \\
&\quad -\frac{2}{n}\sum_{i=1}^{n}\int\left[\frac{1}{n}\sum_{j=1}^{n}\int \phi_{\vec{0},D}(\vec{y}-\vec{u})\phi_{\vec{X}_j,\lambda^2 I_d}(\vec{u})d\vec{u}\right]\phi_{\vec{X}_i,\lambda^2 I_d}(\vec{y})d\vec{y} \\
&= \int\left[\frac{1}{n}\sum_{i=1}^{n}\phi_{\vec{X}_i,D+\lambda^2 I_d}(\vec{y})\right]^2 d\vec{y} + \int\left[\frac{1}{n}\sum_{i=1}^{n}\phi_{\vec{X}_i,\lambda^2 I_d}(\vec{y})\right]^2 d\vec{y} \\
&\quad -\frac{2}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\int \phi_{\vec{X}_j,D+\lambda^2 I_d}(\vec{y})\phi_{\vec{X}_i,\lambda^2 I_d}(\vec{y})d\vec{y} \\
&= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\{\phi_{\vec{X}_i-\vec{X}_j,2D+2\lambda^2 I_d}(\vec{0}) - 2\phi_{\vec{X}_i-\vec{X}_j,D+2\lambda^2 I_d}(\vec{0}) + \phi_{\vec{X}_i-\vec{X}_j,2\lambda^2 I_d}(\vec{0})\} \\
&= \vec{w_1}\{\Omega_2^* - 2\Omega_1^* + \Omega_0^*\}(\vec{w_1})'
\end{aligned}$$

Recalling $M^* = V^* + B^*$ is proof is complete. □

Proof of Theorem 4.4.9 Recalling the definition of T we see that

$$\begin{aligned}
T &= \frac{1}{n(4\pi)^{d/2}h_1h_2. .h_d} \\
&\quad +\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\{\phi_{\vec{X}_i-\vec{X}_j,4D}(\vec{0}) - 2\phi_{\vec{X}_i-\vec{X}_j,3D}(\vec{0}) + \phi_{\vec{X}_i-\vec{X}_j,2D}(\vec{0})\},
\end{aligned}$$

where $D = diag(h_1^2, ..., h_d^2)$. It is easy to verify that, almost surely,

$$\phi_{\vec{X}_i-\vec{X}_j,jD}(\vec{0}) = \frac{1}{(j2\pi)^(d/2)\prod_{j=1}^{d}h_j}\exp\left(-\sum_{l=1}^{d}\frac{(X_{il}-X_{jl})^2}{2jh_l^2}\right) \to 0,$$

as $h_i \to \infty$, $i = 1, 2, .., d$. Therefore for fixed n, $T \to 0$, almost surely, as $h_i \to \infty$, $i = 1, 2, .., d$. This completes the proof of Theorem 4.4.9. □

Chapter 5

Automatic Bandwidth Selection

5.1 Introduction

An important problem in density estimation is to choose the bandwidth $h \equiv h_n$ by minimizing M, the MISE of a density estimator. In general M is unknown. So several data based schemes for selecting h have been proposed with an aim to minimize some estimate of M. Cross-validation, smoothed cross-validation, smooth bootstrap, plug-in rule are some well known bandwidth selection schemes. In this chapter we propose new automatic bandwidth selection schemes for univariate and multivariate kernel density estimators. In chapters 2 and 4 we have proposed new consistent estimators (we call them M^*) of M for univariate and multivariate density estimators. The basic idea of our proposal is to choose h by minimizing M^*.

In general the multivariate product kernel density estimator (see chapter 4) depends on d bandwidths, say $h_1, .., h_d$, where d is the data dimension. But in order to simplify theoretical calculations we assume that $h_1 = ... = h_d = h$ and the resulting density estimator is referred to as a simple product kernel density estimator. This assumption is not unusual. For instance in chapter 4 we have discussed some applications of the simple product kernel density estimator. In this chapter we address the problem of selecting the common bandwidth h of a simple product kernel density estimator based on data which are realizations of i.i.d, R^d valued random variables.

For a kernel density estimate based on univariate data, let h^* be the minimizer of the MISE M, for $h \in I = \left[\frac{\epsilon_1}{n^{1/(2s+1)}}, \frac{\epsilon_2}{n^{1/(2s+1)}}\right]$ where $0 < \epsilon_1 < \epsilon_2$, s is the kernel order. The assumption of restricting h to constant multiples of $\frac{1}{n^{1/(2s+1)}}$ is not too restrictive.

For example see assumption (2.3) in Park and Marron (1990).

For a simple product kernel density estimate, let h^* be the minimizer of M, for $h \in I = \left[\frac{\epsilon_1}{n^{1/(4+d)}}, \frac{\epsilon_2}{n^{1/(4+d)}}\right]$, where d is the data dimension. Actually the choice of this interval is motivated by asymptotic properties of M. We know that for a simple product kernel density estimator with second order kernel and bandwidth h the following equation holds (see chapter 4)

$$M = \frac{C_1}{nh^d} + h^4 C_2 + o\left(\frac{1}{nh^d} + h^4\right), \text{ where } C_1,\ C_2 \text{ are positive constants.}$$

The leading term in the right side is minimized for a value of h which is a constant multiple of $\frac{1}{n^{1/(4+d)}}$ and minimizing $\frac{C_1}{nh^d} + h^4 C_2$ is asymptotically equivalent to minimizing M in some interval $\left[\frac{\epsilon_1}{n^{1/(4+d)}}, \frac{\epsilon_2}{n^{1/(4+d)}}\right]$, where $\epsilon_1,\ \epsilon_2$ are positive constants.

5.1.1 Definitions

Let us describe our bootstrap bandwidth selection rule. In chapter 2, 4 we have introduced MISE estimators, both denoted by M^*, based on univariate and multivariate data. Consequently there are two definitions of our automatic bandwidth $\hat{h}^*$, corresponding to one- dimensional and d- dimensional ($d > 1$) data.

1. $\hat{h}^*$ *for univariate density estimate.* Given n one-dimensional observations, let us recall our MISE estimator M^*, defined in chapter 2. Then

$$\hat{h}^* = \text{minimizer of } M^*(h),\ h \in I = \left[\frac{\epsilon_1}{n^{1/(2s+1)}}, \frac{\epsilon_2}{n^{1/(2s+1)}}\right].$$

$\hat{h}^*$ is a multiple of $\frac{1}{n^{1/(2s+1)}}$ and it depends on the parameter λ, in M^*. λ can be either some fixed bandwidth or can be a random automatic bandwidth such as LSCV bandwidth. So there are several versions of $\hat{h}^*$, depending on the choice of λ.

Let $\hat{h}^*_c = \hat{h}^*$, for $\lambda = h_c$. This is the bootstrap bandwidth proposed by Faraway and Jhun (1990). If $\lambda = h$ and K is the Gaussian kernel, then $\hat{h}^*$ equals Taylor's (1989) bandwidth h_T.

Cao's (1993) bootstrap bandwidth, call it h_{Cao}, is a special case of $\hat{h}^*$ for $K^0 = K$, where K is a second order kernel possessing six derivatives. Cao (1993) claimed that $\lambda = Cn^{-1/7}$ is an optimal choice for λ, where C is a positive constant and depends on integrated squared derivative of f.

If K, K^0 are second order kernels with eight bounded, continuous derivatives and

$\lambda = Cn^p h^m$, the resulting $\hat{h}^*$ is the bootstrap bandwidth proposed by Jones, Marron and Park (1991).

2. $\hat{h}^*$ *for multivariate kernel esimates.* Given our MISE estimate M^*, based on $d-$ dimensional data (see chapter 4), let us define

$$\hat{h}^* = \text{minimizer of } \; M^*(h), \; h \in I = \left[\frac{\epsilon_1}{n^{1/(4+d)}}, \frac{\epsilon_2}{n^{1/(4+d)}}\right].$$

The multivariate version of M^* depends on two parameters λ, μ. From chapter 4 we recall that an appropriate choice of μ is a constant multiple of $\frac{1}{n^{1/(4+d)}}$. For λ there are two choices, namely constant multiples of $\frac{1}{n^{1/(8+d)}}$ and $\frac{1}{n^{1/(2+d)}}$. Consequently there are two versions of our $\hat{h}^*$, based on multivariate data.

Now let us introduce some of the standard bandwidth selection methods for univariate and multivariate kernel density estimators. For a univariate kernel density estimate $K_n(\cdot)$, with bandwidth h, following are some well known bandwidth selection rules.

1. *Least Squares Cross-Validation* (see Rudemo (1982) and Bowman (1984)). We select h by minimizing UCV_n (defined in chapter 2), with respect to h. Let us denote the minimizer of UCV_n by h_c. In R it is invoked by "bw.ucv" function.

2. *Smooth Cross Validation.* Hall et al. (1992) proposed to select h by minimizing SCV_n which is defined as follows

$$SCV_n = \frac{\int K^2(u)du}{nh} + \frac{1}{n(n-1)} \sum_{i \neq j = 1}^{n} \{K_h * K_h - 2K_h + K^0) * L_g * L_g\} (X_i - X_j),$$

where $K_h(\cdot) = \frac{1}{h}K(\cdot/h)$, $L_g(\cdot) = \frac{1}{g}L(\cdot/g)$ and $*$ denotes convolution. K and h are the given kernel and bandwidth respectively. Whereas L and g are some other kernel and bandwidth. We note that if K and L are standard normal densities then

$$\begin{aligned} SCV_n &= \frac{\int K^2(u)du}{nh} \\ &+ \frac{1}{n(n-1)} \sum_{i \neq j=1}^{n} \{\phi_{2h^2+2g^2}(X_i - X_j) - 2\phi_{h^2+2g^2}(X_i - X_j) + \phi_{2g^2}(X_i - X_j)\}. \end{aligned}$$

Hall et al.(1992) proposed to use $g = s\frac{0.9266}{n^{2/3}}$, where s is the sample standard deviation. We denote the minimizer of SCV_n by h_{sc}.

3. *Plug-in* rule (Silverman (1986), Sheather-Jones (1991)). It is based on the idea of minimizing the asymptotic approximation A (defined in chapter 2) of M. For a second

order kernel, A involves an unknown quantity $R(f^{(2)}) = \int \left[f^{(2)}(x)\right]^2 dx$. Different versions of the plug-in approach depend on the form of the estimate of $R(f^{(2)})$. Two popular versions of the plug-in rule are the *Rules of Thumb* (Silverman (1986)) and *Solve-the-Equation plug-in* method by Sheather-Jones (1991). The Sheather-Jones (1991) plug-in bandwidth (we denote it by h_{sj}) is widely recommended (see Sheather (2004)) and it is available in S-PLUS and R. In R it is invoked by a function "bw.SJ". The other plug-in bandwidth h_a, based on *Rules of Thumb*, is defined as $h_a = 1.06.min(s, Q/1.34).n^{-1/5}$, where s is the sample standard deviation and Q is the difference between the third and first sample quartiles.

4. *Biased cross validation* (Scott and Terrell (1987)). Here we select a bandwidth by minimizing BCV_n defined in chapter 2. We denote the minimizer of BCV_n by h_{bcv}.

For a product kernel density estimator, with bandwidths h_1, $h_2, ...,$ h_d, Sain, Baggerly and Scott (1994) proposed multivariate extensions of the Taylor's (1989) bootstrap bandwidth and the unbiased cross validation bandwidth. They are defined as follows.

1. *Bootstrap bandwidth selection rule.* The bandwidths h_1, $h_2, ...,$ h_d are selected by the minimizing T defined in chapter 4. For a simple product kernel density estimator, with bandwidth $h_1 = h_2 = ... = h_d = h$, T is a function of h and the bootstrap method of selecting h is to minimize T, with respect to h and we denote the minimizer of T by h_T.

2. *Least Squares (Unbiased) Cross-Validation.* The bandwidths h_1, $h_2, ...,$ h_d are selected by minimizing $UCV_n(h_1, ..., h_d)$ which equals

$$\frac{1}{(2\sqrt{\pi})^d n \prod_{j=1}^d h_j} + \frac{1}{(2\sqrt{\pi})^d n^2 \prod_{j=1}^d h_j} \times \sum_{i=1}^{n} \sum_{j \neq i} \left[exp\left\{-\frac{1}{4}\sum_{k=1}^{d} \triangle_{ijk}^2\right\} - 2 \times 2^{d/2} exp\left\{-\frac{1}{2}\sum_{k=1}^{d} \triangle_{ijk}^2\right\}\right],$$

where $\triangle_{ijk} = \frac{X_{ik} - X_{jk}}{h_k}$, $i \neq j = 1, 2, ..., n$, $k = 1, 2, ..., d$, and X_{ik} is the kth component of $\vec{X}_i$. If $h_1 = h_2 = ... = h_d = h$, then the unbiased cross validation rule for selecting h is to minimize $UCV_n(h)$ with respect to h. Let us denote the minimizer of $UCV_n(h)$ by h_c.

5.1.2 A review of some well known automatic bandwidths

Univariate automatic bandwidths. So far we defined a number of well known data based bandwidth selection rules. These include the cross-validation of Rudemo (1982) and Bowman (1984), the smoothed cross-validation of Hall, Marron and Park (1992), and the plug-in rules of Silverman (1986) and Sheather-Jones (1991).

The centerpoint of research on bandwidth selection for kernel density estimators has been the least squares (unbiased) cross-validation method (LSCV). Park and Marron (1990), Hall et al. (1992), Jones, Marron and Sheather (1996) have been strongly critical of LSCV and have advocated other bandwidth selection schemes, such as plug-in methods based on *Rules of Thumb*, Smoothed Cross Validation (SCV) and *Solve-the-equation* plug-in approach which are less variable and possess stronger asymptotic properties than the cross validation method. Taylor (1989) and Faraway and Jhun (1990) proposed to select h by minimizing smooth bootstrap estimates of M.

Taylor (1989) compared his bootstrap bandwidth with cross validation and *Rules of Thumb* plug-in bandwidth. Faraway and Jhun (1990) observed that bootstrap bandwidths are generally larger but less variable than the cross validation bandwidth. For second order kernels with six derivatives, Cao (1993) studied asymptotic properties of his bootstrap bandwidth. Cao et al. (1994) conducted a broad simulation study to compare performances of a number of automatic bandwidths. They observed that LSCV bandwidth exhibits relatively poor behaviour and recommended the plug-in bandwidth by Sheather and Jones (1991) and smooth bootstrap bandwidth (by Cao (1993), Cao et al. (1994)) selection rules for automatic data based choice of h.

Jones, Marron and Park (1991) proposed yet another version of the smooth bootstrap bandwidth (we call it h_{JMP}), where K, K^0 have eight bounded continuous derivatives, f has six bounded derivatives (see their assumptions A.2, A 3) and $\lambda = Cn^p h^m$. For $\lambda = Cn^{-23/45}h^{-2}$, they showed that $\sqrt{n}\left(\frac{h_{JMP}}{h^*} - 1\right)$ is asymptotically normal. This result appears to be very appealing, especially when compared with the rate of convergence of $\frac{h_c}{h^*}$, to one, where h_c is the LSCV bandwidth. It is well known that $\frac{\hat{h}}{h^*} - 1 = O_P\left(\frac{1}{n^{1/10}}\right)$, where $\hat{h}$ equals h_c or similar bandwidth selectors (see Loader (1999)).

Loader (1999) pointed out that the rate of convergence of $\frac{\hat{h}}{h^*}$, to one, is not an appropriate measure of asymptotic performance of any automatic bandwidth $\hat{h}$. He has challenged the superiority of the plug-in methods and pointed out that the plug-in rules can fail if the specification of the pilot bandwidth used in plug-in methods is wrong, and that the often quoted variability and undersmoothing of the LSCV method simply reflects the uncertainty of data based bandwidth selection. He argued that the shortcoming of the less variable bandwidth selection rules such as plug-in methods, manifests in another way: consistently oversmoothing small and difficult to detect features. It appears that there is no unique bandwidth selection scheme that enjoys universal superiority on all fronts.

Automatic bandwidths based on multivariate data. The amount of research on bandwidth selection rules, especially based on bootstrap method, for multivariate kernel estimators appear to be quite less in comparison to the univariate case. Sain, Baggerly and Scott (1994) have proposed multivariate extensions of the LSCV, biased cross validation and Taylor's smooth bootstrap bandwidth selection rule for product kernel density estimator.

Interestingly the rate of convergence of $\frac{h_c}{h^*}$ to 1 improves as d is increased, where h_c is bandwidth selected using cross validation rule and d is the data dimension (Sain, Baggerly and Scott (1994)). The theoretical properties of the bootstrap bandwidth selection rule, proposed by Sain Baggerly and Scott (1994), does not appear to be known.

Motivation for our proposal. We have reviewed a number of well known methods for estimating h^*, especially for univariate kernel density estimates. No automatic bandwidth seems to yield the best possible choice of h, from all perspectives. So a natural question is "why propose yet another bandwidth selection rule ?"

Marron and Chung (1997), Sheather (2004) have argued that a family of kernel density estimates is a more powerful graphical device for capturing various features of the data than a single curve. The various bandwidth selection rules can provide the different values of h to produce a family of density estimates (see Sheather (2004)). We further enrich the existing class of bandwidth selection rules, by proposing a new automatic bandwidth ($\hat{h}^*$) with some desirable asymptotic property.

We note that Cao (1993), Jones, Marron and Park (1991) obtained asymptotic properties of h_{Cao} and h_{JMP}, under a number of smoothness conditions on K (for instance see conditions K2 and A.2 in Cao (1993) and Jones et al. (1991)). h_{Cao} and h_{JMP} are special cases of our $\hat{h}^*$ and we intend to obtain an asymptotic property of $\hat{h}^*$

with no smoothness assumption on K at all. This result can be considered as a new asymptotic property of h_{Cao} and h_{JMP} that holds for a broad class of kernels.

5.1.3 Chapter Summary

There are two Theorems and two simulation studies in this chapter. Theorem 5.2.1 provides insight into how well the proposed automatic bandwidth $\hat{h}^*$ succeeds in minimizing M as the sample size is increased. This result holds for h_{Cao} and h_{JMP}. But unlike Cao (1993) and Jones, Marron and Park (1991), we do not impose any smoothness assumption on K at all. So we obtain a new asymptotic property of h_{Cao} and h_{JMP}, with lesser assumptions on K. Theorem 5.2.3 is the multivariate extension of Theorem 5.2.1. Theorem 5.2 3 provides the rate at which the accuracy (in terms of minimizing M) of $\hat{h}^*$, for a simple product kernel density estimate, improves with increase in sample size. These results seem to be new.

In the first simulation study we compare a wide class of bandwidth selection rules in the context of univariate kernel density estimators. We compare $\hat{h}^*$ with six well known automatic bandwidths, in terms of minimizing Monte-Carlo estimate of $E\left|\frac{M^*(h)}{M(h)} - 1\right|$ based on univariate data. In the next simulation study we compare the proposed automatic bandwidths with LSCV bandwidth, by Sain, Baggerly and Scott (1994), for bivariate product kernel density estimates.

The proofs are given in section 5.5. There are two tables in the first simulation study. These tables, namely Table 5.1 and Table 5.2, are given at the end of the chapter after the proofs.

5.2 Main Results

5.2.1 Univariate case

From the perspective of density estimation, an important question is "how well does $\hat{h}^*$ succeed in minimizing M?" Following Theorem provides some insight. Recall that s is the kernel order, p, k are constants as defined in Assumption A in our chapter 2.

Theorem 5.2.1. *Suppose $s, p \geq 2$, $k = p$ and Assumptions $A - E$ hold. Then*

$$E\left|\frac{M(\hat{h}^*)}{M(h^*)} - 1\right| = o\left(\frac{1}{n^{1/(2s+1)}}\right) + O\left(\frac{1}{\sqrt{n\lambda^{2s+1}}} + \lambda^p\right).$$

Remark 5.2.1. (i) Under the conditions stated in Theorem 5.2.1 and $\lambda = \frac{1}{n^{1/(2s+2p+1)}}$, $E\left|\frac{M(\hat{h}^*)}{M(h^*)} - 1\right| = o\left(\frac{1}{n^{1/(2s+1)}}\right)$.

(ii) From Park and Marron (1990) we see that, for symmetric, second order kernels with finite support, $n^{1/5}\left(\frac{M(\hat{h})}{M(h^*)} - 1\right)$ converges in law to the chi-squared distribution, where $\hat{h}$ is the minimizer of UCV_n, or BCV_n in I. $\hat{h}$ can also be the plug-in bandwidth by Park and Marron (1990) (they call it h_{PI}). Theorem 5.2.1 implies that for any second order kernel (i.e. $s = 2$), $p = 2$ and $\lambda = \frac{1}{n^{1/9}}$, $n^{1/5}\left(\frac{M(\hat{h}^*)}{M(h^*)} - 1\right) = o_P(1)$. Therefore for second order kernels with finite support, assuming the conditions stated in Theorem 5.2.1 hold and $\lambda = \frac{1}{n^{1/9}}$, $\hat{h}^*$ is asymptotically more accurate than the unbiased or biased cross validation bandwidths or the plug-in bandwidth h_{PI}.

(iii) h_{Cao} is a special case of $\hat{h}^*$, for $K^0 = K$, where K is a second order kernel having six derivatives. Cao (1993) has claimed λ equal to a constant multiple $\frac{1}{n^{1/7}}$ to be the optimal choice for λ. Substituting $s = p = 2$ and λ equal to a constant multiple $\frac{1}{n^{1/7}}$ in Theorem 5.2.1 we see that

$$E\left|\frac{M(h_{Cao})}{M(h^*)} - 1\right| = o\left(\frac{1}{n^{1/5}}\right) + O\left(\frac{1}{n^{2/7}}\right) = o\left(\frac{1}{n^{1/5}}\right).$$

Again repeating the arguments in the previous paragraph, we see that for second kernels with finite support, assuming $p = 2$ and the conditions stated in Theorem 5.2.1 hold, h_{Cao} is asymptotically more accurate (in terms of minimizing M) than the LSCV, biased cross validation and the plug-in bandwidth h_{PI}. These are new asymptotic properties of h_{Cao}.

(iv) For a second order kernel K, h_{JMP} is a special case of $\hat{h}^*$, for $\lambda = Cn^p h^m$. In particular if $\lambda = Cn^{-23/45}h^{-2}$ and h equal to a multiple of $\frac{1}{n^{1/5}}$, λ is a multiple of $\frac{1}{n^{1/9}}$. This resembles with our choice of λ in M^*, see Theorem 2.3.1 (iii) in chapter 2. Recall that in chapter 2, we had proposed $\lambda = \frac{1}{n^{1/(2s+2p+1)}}$.

Jones, Marron and Park (1991) proved that for a second order kernel K, satisfying their smoothness assumption (A.2), and $\lambda = Cn^{-23/45}h^{-2}$, $\frac{h_{JMP}}{h^*} - 1 = O_P\left(\frac{1}{\sqrt{n}}\right)$. We obtain the rate of convergence of $\frac{M(h_{JMP})}{M(h^*)}$ to one. For a second order kernel K and $h \in I$, substituting λ equal to a constant multiple of $\frac{1}{n^{1/9}}$, under conditions stated in Theorem 5.2.1 we get

$$E\left|\frac{M(h_{JMP})}{M(h^*)} - 1\right| = o\left(\frac{1}{n^{1/5}}\right).$$

This result seems to be a new asymptotic property of h_{JMP}. But unlike Jones, Marron and Park (1991), we do not impose any smoothness assumption on K at all.

We note that the bootstrap method of optimal bandwidth selection can be computationally expensive. In general, computing $\hat{h}^*$ involves two stages of approximations, namely the Monte-Carlo simulation for computing M^* and numerical optimization for minimizing M^*. For Gaussian type kernels we have obtained closed form expression for M^*, in chapter 2. So an easy way to compute $\hat{h}^*$ is to plot M^*. After tentatively locating $\hat{h}^*$ from the graph, we can compute its exact value using numerical optimization algorithms (for instance using *uniroot* and *optimize* algorithms in R).
Some numerical optimization algorithms (for example *uniroot* in R) require the formulae of derivatives of M^*. Therefore the formulae for derivatives M^* (for Gaussian kernels) are obtained in the next Theorem.

Theorem 5.2.2. *If $K(\cdot)$ is a Gaussian-based kernel of order $2r$ and K^0 is the standard normal density, then*

$$M_n^{*(1)} = -\frac{C_1(r)}{nh_n^2} + \left(1-\frac{1}{n}\right)\sum_{i=0}^{r-1}\sum_{j=0}^{r-1}\frac{(-1)^{i+j}}{2^{i+j}i!j!}U^{(1)}(h;i+j,2) - 2\sum_{s=0}^{r-1}\frac{(-1)^s}{2^s\ s!}U^{(1)}(h,s,1)$$

$$\text{where}\ \ U^{(1)}(h;j,q) = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{l=1}^{n}\{2jh_n^{2j-1}\phi^{(2j)}_{2\lambda_n^2+qh_n^2}(X_i-X_l) + qh_n^{2j+1}\phi^{(2j+2)}_{2\lambda_n^2+qh_n^2}(X_i-X_l)\}.$$

$$M_n^{*(2)} = \frac{2C_1(r)}{nh_n^3} + \left(1-\frac{1}{n}\right)\sum_{i=0}^{r-1}\sum_{j=0}^{r-1}\frac{(-1)^{i+j}}{2^{i+j}i!j!}U^{(2)}(h;i+j,2) - 2\sum_{s=0}^{r-1}\frac{(-1)^s}{2^s\ s!}U^{(2)}(h,s,1)$$

$$\text{where}\ \ U^{(2)}(h;j,q) = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{l=1}^{n}\{2j(2j-1)h_n^{2j-2}\phi^{(2j)}_{2\lambda_n^2+qh_n^2}(X_i-X_l) + q(4j+1)h_n^{2j}\phi^{(2j+2)}_{2\lambda_n^2+qh_n^2}(X_i-X_l) + q^2h_n^{2j+2}\phi^{(2j+4)}_{2\lambda_n^2+qh_n^2}(X_i-X_l)\}.$$

5.2.2 Multivariate case

We now state a Theorem which is an extension of Theorem 5.2.1, to the context of bandwidth selection for a simple product kernel density estimate. This Theorem provides some insight into how well $\hat{h}^*$ succeeds in minimizing the MISE of a simple product kernel density estimator.

Theorem 5.2.3. *Suppose Assumptions $A-C$ in chapter 4 hold and $n\lambda^{d+4}\to\infty$. Then*

$$E\left|\frac{M(\hat{h}^*)}{M(h^*)}-1\right| = O\left(\sqrt{\frac{1}{n\mu^d}+\mu^4}+\sqrt{\frac{1}{n\lambda^{d+4}}+\lambda^4}\right). \tag{5.2.1}$$

5.3 Simulation

We have conducted two simulation experiments. In the first experiment we compare two versions of our proposed bandwidth selection rules , $\hat{h}^*$ (with $\lambda = \frac{1}{n^{1/9}}$) and $\hat{h}_c^*$, with a number of existing bandwidth selection schemes such as LSCV (h_c), the plug-in *Rule of thumb* (h_a) and Sheather-Jones plug-in bandwidth (h_{sj}), SCV bandwidth (h_{sc}) and Taylor's (1989) bootstrap bandwidth (h_T).

In the next experiment we compare LSCV bandwidth (by Sain, Baggerly and Scott (1994)) and $\hat{h}^*$, for λ equal to $\frac{1}{n^{1/(4+d)}}$ and $\frac{1}{n^{1/(2+d)}}$, for simple product kernel density estimates based on bi-variate data.

5.3.1 Comparison of automatic bandwidths for univariate density estimates

We draw 30 samples, of size 50 and 200, from four distributions, namely standard normal, bimodal, claw and double claw.

For each sample we compute the ratios $\frac{\hat{h}}{h^*}$ and $\frac{M(\hat{h})}{M(h^*)}$, where $\hat{h} = h_c$, $\hat{h}^*$, $\hat{h}_c^*$, h_{sc}, h_T, h_a, h_{sj} and h^* denotes the global minima of M. $\hat{h}^*$ denotes the minimizer of M^*, with $\lambda = \frac{1}{n^{1/9}}$. We estimate the bias and variance of $\frac{\hat{h}}{h^*}$ from the 30 sample values of $\frac{\hat{h}}{h^*}$. These are tabulated in Table 5.1. Let $\hat{E}\left|\frac{M(\hat{h})}{M(h^*)} - 1\right|$ denote the average of the values of $\left|\frac{M(\hat{h})}{M(h^*)} - 1\right|$, calculated from the 30 samples, where $\hat{h}$ is any one of the seven data based bandwidths which we compare. $\hat{E}\left|\frac{M(\hat{h})}{M(h^*)} - 1\right|$ is the Monte-Carlo estimate of $E\left|\frac{M(\hat{h})}{M(h^*)} - 1\right|$ for any bandwidth $\hat{h}$ and its values are tabulated in Table 5.2. The main observations are as follows:

(i) *Sampling fluctuation and bias.* From Table 5.1 we find that as expected, there is no unique bandwidth selection rule which exhibits least sampling fluctuation for all the four underlying densities. For instance if $n = 50$ and underlying density is normal, $\hat{h}^*$ has least variance among all the automatic bandwidths. But for bimodal, double claw and claw densities, h_a possesses least variance for $n = 50$. Similarly, for $n = 200$, h_a has least variance for normal and bimodal densities whereas, $\hat{h}^*$ and h_c exhibit least sampling fluctuation for double claw and claw densities.

In general the proposed bandwidth $\hat{h}^*$ possess lower variance than both LSCV and Sheather and Jones (1991) plug-in bandwidths. However $\hat{h}^*$ can have higher bias than

LSCV and Sheather and Jones (1991) plug-in bandwidths.

(ii) *Minimizing M*. From Table 5.2 we may conclude that there is no unique bandwidth selection rule that minimizes $\hat{E}\left|\frac{M(\hat{h})}{M(h^*)} - 1\right|$ uniformly for all the densities and for all sample sizes. For instance h_T, h_a, $\hat{h}^*$, $\hat{h}_c^*$ are the minimizers of $\hat{E}\left|\frac{M(\hat{h})}{M(h^*)} - 1\right|$, for normal, bimodal density, double-claw and claw density respectively and $n = 50$. But $\hat{h}^*$, h_{sj}, h_T and h_c are the minimizers of $\hat{E}\left|\frac{M(\hat{h})}{M(h^*)} - 1\right|$, for normal, bimodal, double-claw and claw density respectively and $n = 200$. So it is recommended to produce a family of density estimates using the above mentioned automatic bandwidths.

5.3.2 Comparison of bandwidths for bivariate density estimates

Sain, Braggerly and Scott (1994) has proved that the theoretical behaviour of the LCV algorithm, for optimal bandwidth selection, improves rapidly as data dimension increases. Besides cross-validation method is a popular bandwidth selection scheme. For instance, it is widely used in home-range estimation in ecology (see Gitzen R.A. and Millspaugh J.J. (2003). In this simulation study we compare the contour and perspective plots of kernel density estimates, using the LSCV bandwidth and the proposed bandwidths $\hat{h}^*$, for $\lambda = \frac{1}{n^{1/(2+d)}}$ and $\frac{1}{n^{1/(8+d)}}$. See Figures 5.1 to 5.30. We compare the plots of the density estimates with the corresponding plots of the underlying density function. The contour and perspective plots for each density estimate is obtained for samples of size 50 and 200. These plots are expected to provide insight into how well the simple product kernel density estimates with $h = \hat{h}^*$ and h_c, imitate the features of the underlying density for small and large samples. We consider four bivariate densities, namely bimodal, kurtic, trimodal and quarimodal densities. These distributions are explained in Table 4.1, chapter 4. The main observations are as follows:

1) The perspective and contour plots of simple product kernel density estimates based on $h = h_c$ and $h = \hat{h}^*$ (with $\lambda = \frac{1}{n^{1/(d+2)}}$) resemble closely. The density estimates, obtained using $h = \hat{h}^*$ (with $\lambda = \frac{1}{n^{1/(d+8)}}$), appear to be oversmoothed.

2) Densities which are unimodal or have well separated multiple modes can be estimated reasonably well using all the three bandwidths based on sample size 200. However if the underlying density has several modes, then the LSCV bandwidth tends to undersmooth, especially for small samples.

3) A density with close multiple peaks are hard to estimate using simple product kernel density estimator. If the underlying density has two or more peaks which are close, then a perspective of the simple product kernel density estimate may not reveal all the peaks even for large samples. For instance we note that for trimodal density, the none of the perspective plots can capture the peaks that are visible in the perspective plot of the density function.

4) There appears to be no unique bandwidth selection rule which produces the best perspective or contour plot for all the five bivariate densities. We note that for $n = 200$ bimodal, skewed and quadrimodal densities, the contour and perspective plots of density estimates based on $h = h_c$ and $h = \hat{h}^*$ (with $\lambda = \frac{1}{n^{1/(d+2)}}$) capture the important features of the underlying density. But for kurtic density, the contour and perspective plots density estimate based $h = \hat{h}^*$ (with $\lambda = \frac{1}{n^{1/(d+8)}}$) resemble the corresponding plots of the underlying density more closely than the other two bandwidths. So for any given set of bivariate data, it is recommended to produce a family of plots using these bandwidths.

The various perspective and contour plots, mentioned in this simulation study, are given in next seven pages.

Fig 5.1: The perspective plot of Bimodal density

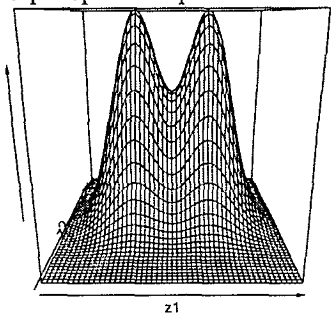


Fig 5.2: Perspective plots of Bimodal density estimates, with $h = h_c, \hat{h}^*(\lambda = \frac{1}{n^{1/(8+d)}}, \frac{1}{n^{1/(2+d)}})$ respectively and $n = 50$.

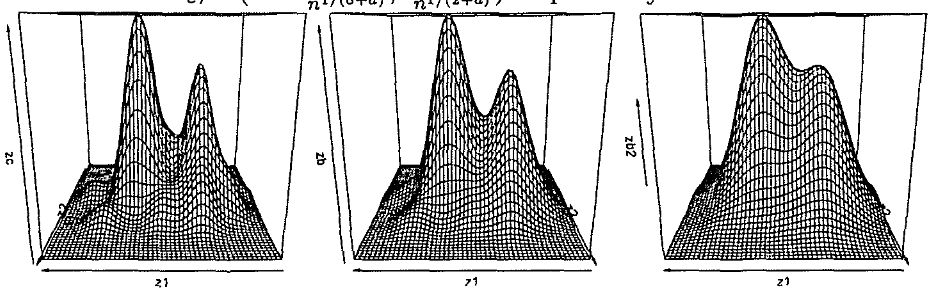


Fig 5.3: Perspective plots of Bimodal density estimates, with $h = h_c, \hat{h}^*(\lambda = \frac{1}{n^{1/9}}, \frac{1}{n^{1/(2+d)}})$ respectively $n = 200$.

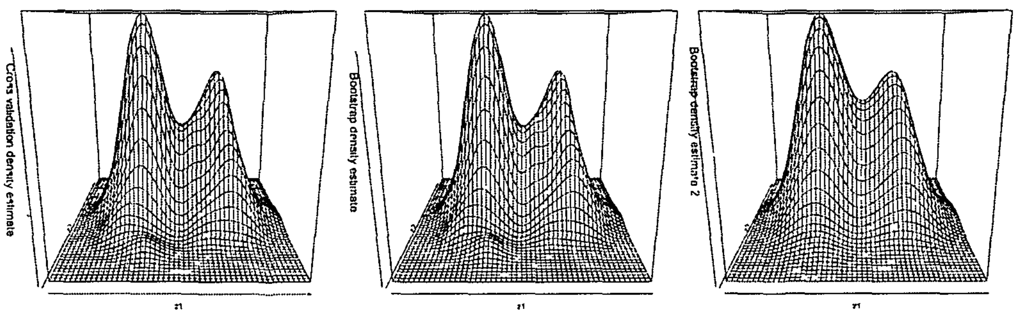


Fig 5.4: The perspective plot of Kurtic density

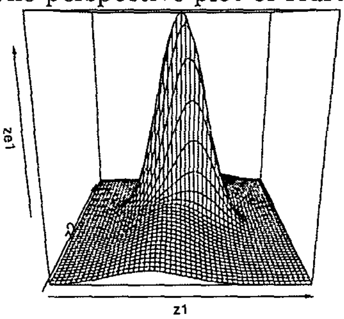


Fig 5.5: Perspective plots of Kurtic density estimates, with

$h = h_c, \hat{h}^*(\lambda = \frac{1}{n^{1/(8+d)}}, \frac{1}{n^{1/(2+d)}})$ respectively and $n = 50$.

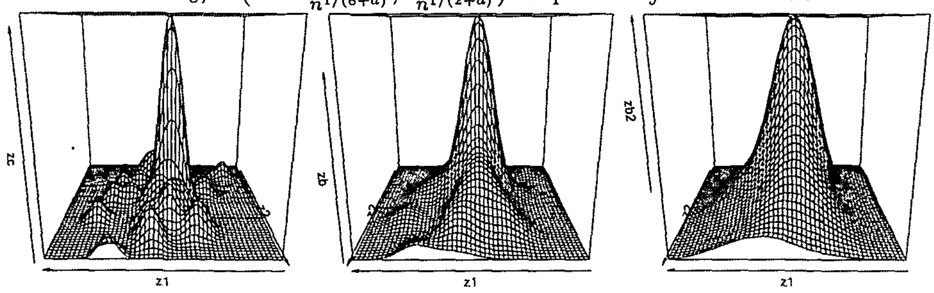


Fig 5 6: Perspective plot of Kurtic density estimates, with

$h = h_c, \hat{h}^*(\lambda = \frac{1}{n^{1/9}}, \frac{1}{n^{1/(2+d)}})$ respectively $n = 200$

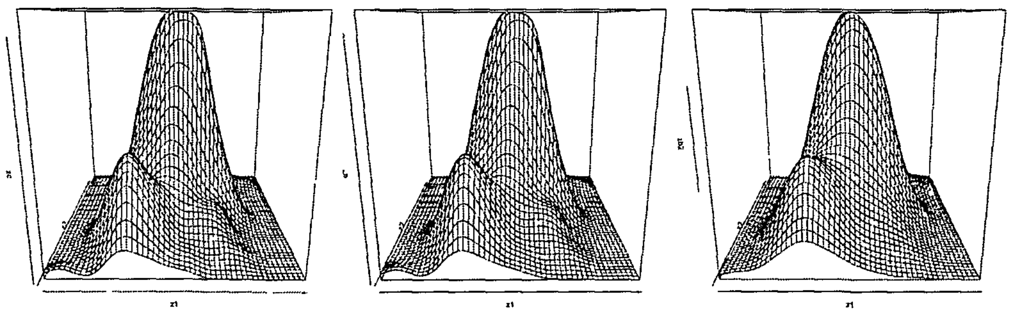


Fig 5.7: Perspective plot of Skewed density.

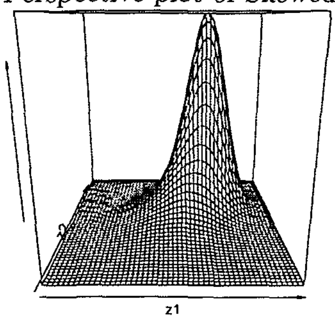


Fig 5.8: Perspective plots of Skewed density estimates, with $h = h_c, \hat{h}^*(\lambda = \frac{1}{n^{1/(8+d)}}, \frac{1}{n^{1/(2+d)}})$ respectively and $n = 50$.

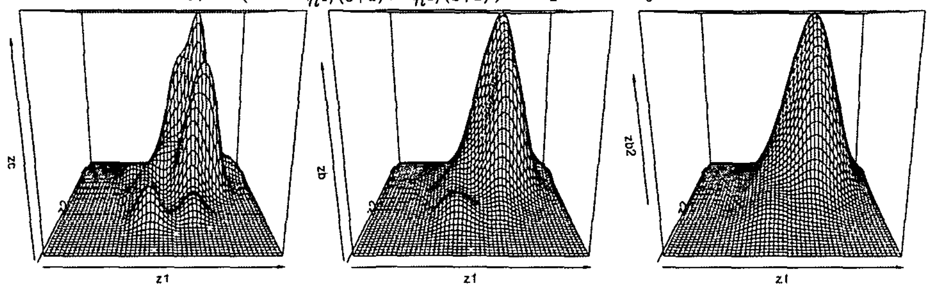


Fig 5.9: Perspective plots of Skewed density estimates, with $h = h_c, \hat{h}^*(\lambda = \frac{1}{n^{1/(8+d)}}, \frac{1}{n^{1/(2+d)}})$ respectively and $n = 200$.

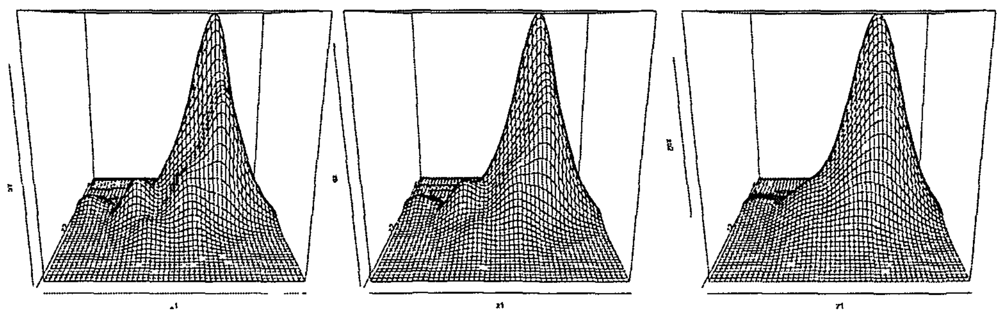


Fig 5.10: The perspective plot of Trimodal density

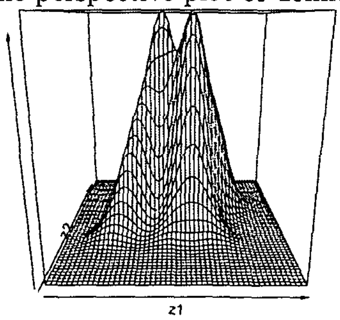


Fig 5.11: Perspective plots of Trimodal density estimates, with $h = h_c, \hat{h}^*(\lambda = \frac{1}{n^{1/9}}, \frac{1}{n^{1/(2+d)}})$ respectively $n = 50$

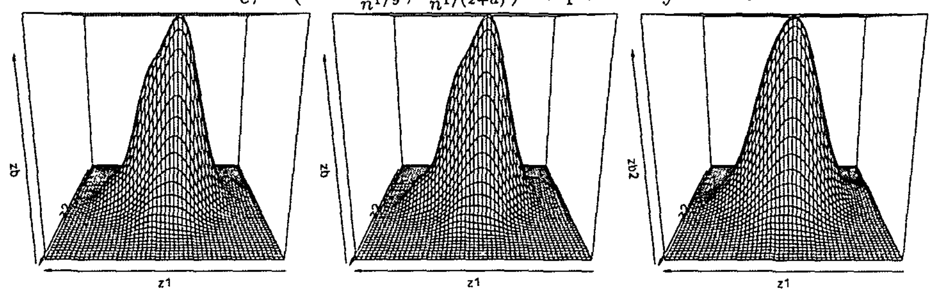


Fig 5.12: Perspective plots of Trimodal density estimates, with $h = h_c, \hat{h}^*(\lambda = \frac{1}{n^{1/9}}, \frac{1}{n^{1/(2+d)}})$ respectively $n = 200$

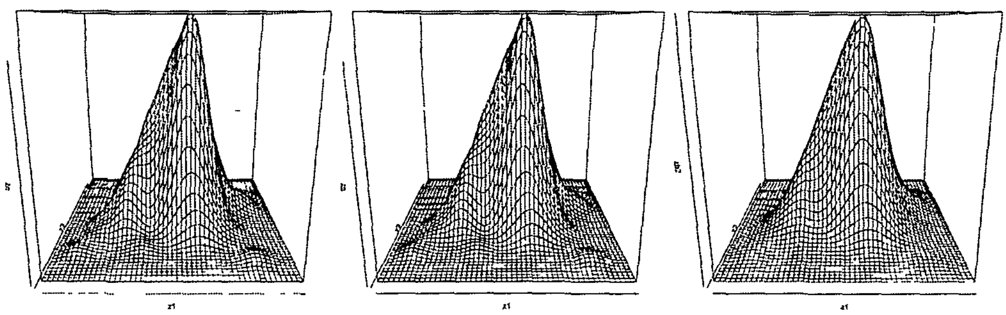


Fig 5.13: The perspective plot of Quadrimodal density

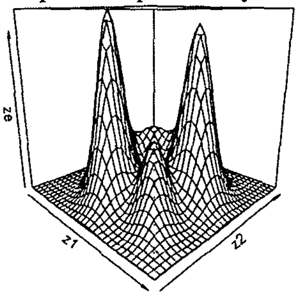


Fig 5.14: Perspective plots of Quadrimodal density estimates, with $h = h_c, \hat{h}^*(\lambda = \frac{1}{n^{1/(8+d)}}, \frac{1}{n^{1/(2+d)}})$ respectively and $n = 50$.

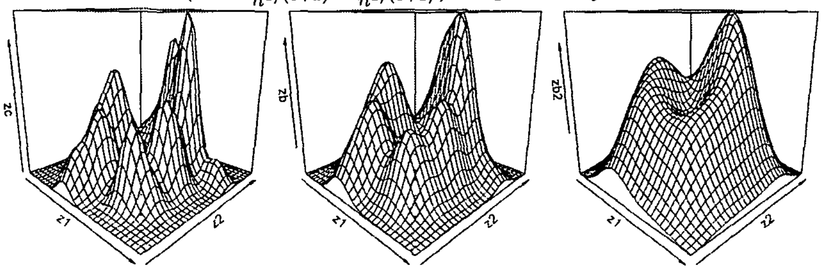


Fig 5.15: Perspective plots of Quadrimodal density estimates, with $h = h_c, \hat{h}^*(\lambda = \frac{1}{n^{1/(8+d)}}, \frac{1}{n^{1/(2+d)}})$ respectively and $n = 200$.

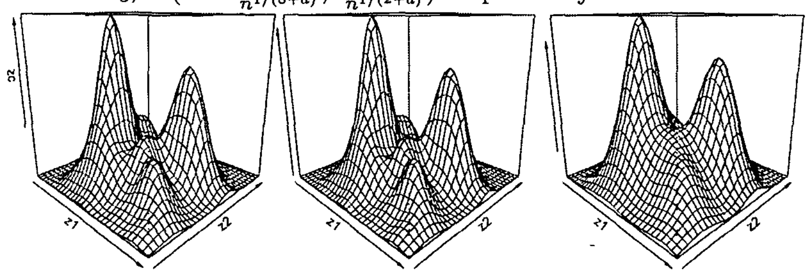


Fig 5.16: The contour plot of Bimodal density

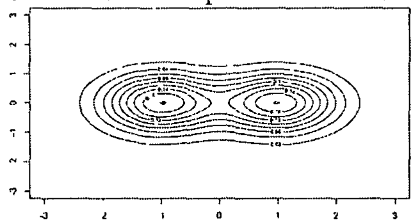


Fig 5.17: Contour plots of bimodal density estimates, with $h = h_c, \hat{h}^*(\lambda = \frac{1}{n^{1/(8+d)}}, \frac{1}{n^{1/(2+d)}})$ respectively and $n = 50$.

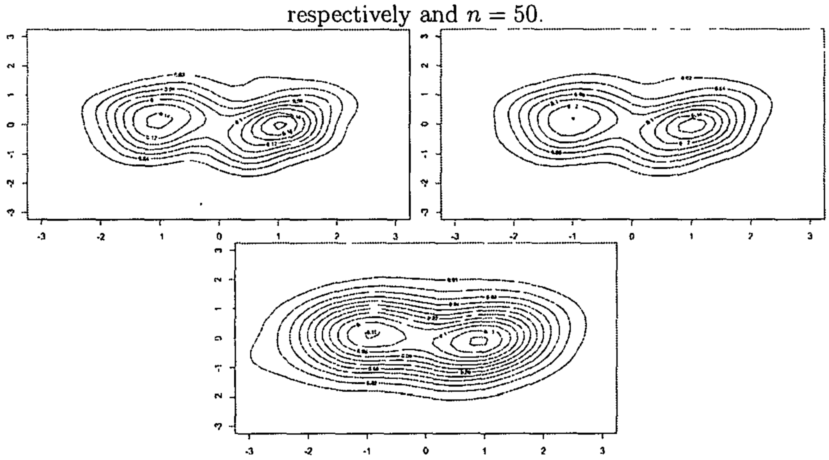


Fig 5.18: Contour plots of bimodal density estimates, with $h = h_c, \hat{h}^*(\lambda = \frac{1}{n^{1/(8+d)}}, \frac{1}{n^{1/(2+d)}})$ respectively and $n = 200$.

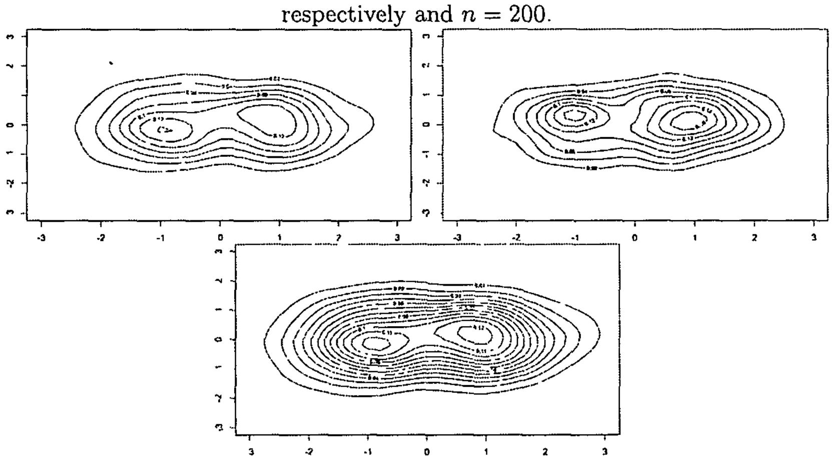


Fig 5.19: Contour plot of Kurtic density

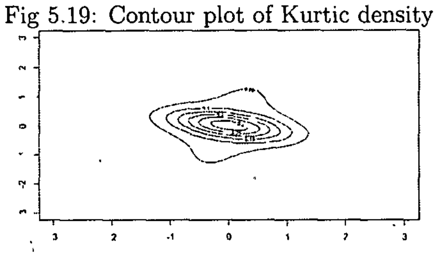


Fig 5.20: Contour plots of Kurtic density estimates, with $h = h_c, \hat{h}^*(\lambda = \frac{1}{n^{1/(8+d)}}, \frac{1}{n^{1/(2+d)}})$ respectively and $n = 50$.

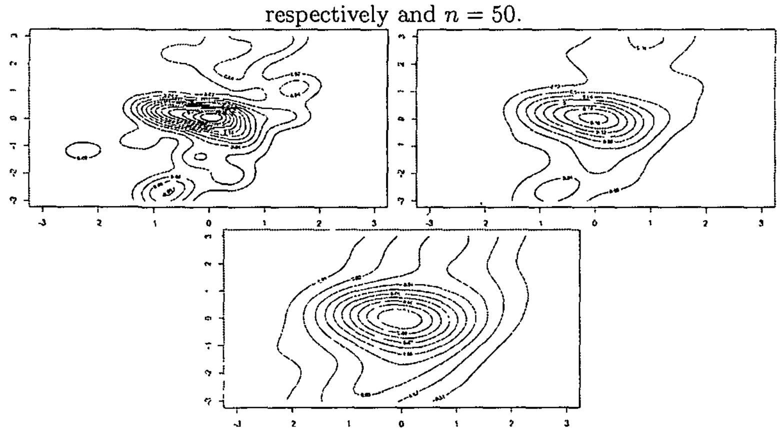


Fig 5.21: Contour plots of Kurtic density estimates, with $h = h_c, \hat{h}^*(\lambda = \frac{1}{n^{1/(8+d)}}, \frac{1}{n^{1/(2+d)}})$ respectively and $n = 200$.

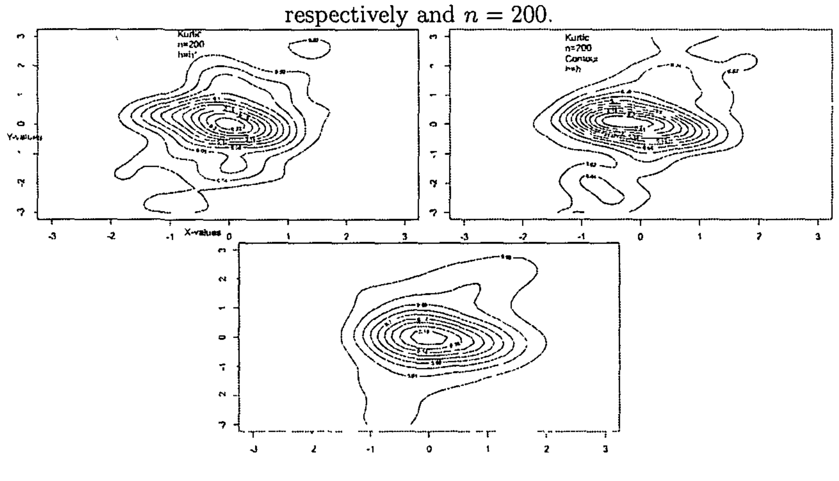


Fig 5 22 Contour plot of Skewed density

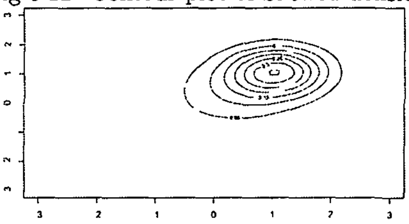


Fig 5 23 Contour plots of Skewed density estimates, with $h = h_c, \hat{h}^*(\lambda = \frac{1}{n^{1/(8+d)}}, \frac{1}{n^{1/(2+d)}})$

respectively and $n = 50$

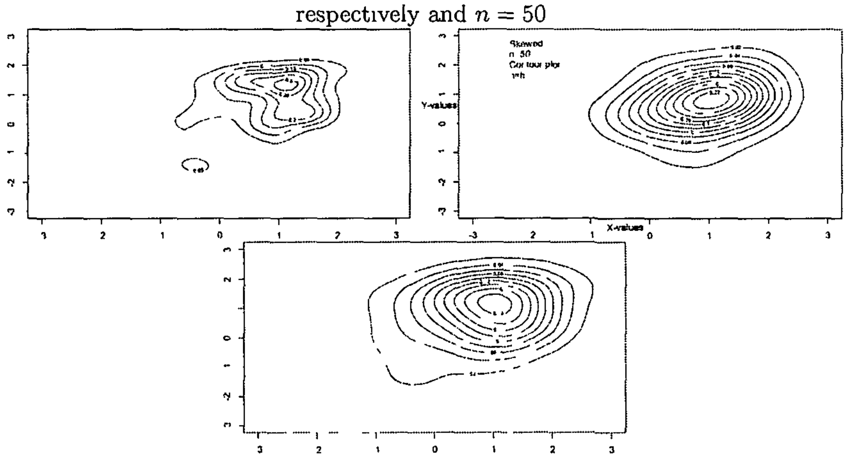


Fig 5 24 Contour plots of Skewed density estimates, with $h = h_c, \hat{h}^*(\lambda = \frac{1}{n^{1/(8+d)}}, \frac{1}{n^{1/(2+d)}})$

respectively and $n = 200$

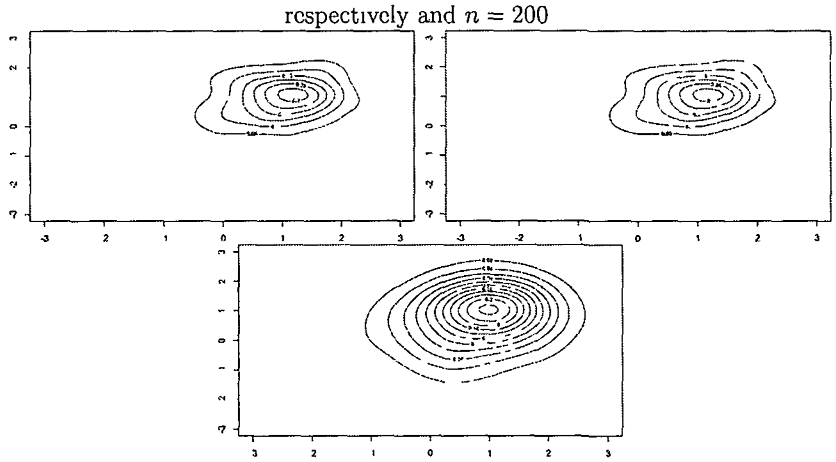


Fig 5.25: Contour plot of Trimodal density

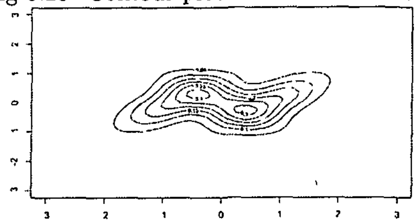


Fig 5.26: Contour plots of Trimodal density estimates, with $h = h_c, \hat{h}^*(\lambda = \frac{1}{n^{1/(8+d)}}, \frac{1}{n^{1/(2+d)}})$ respectively and $n = 50$.

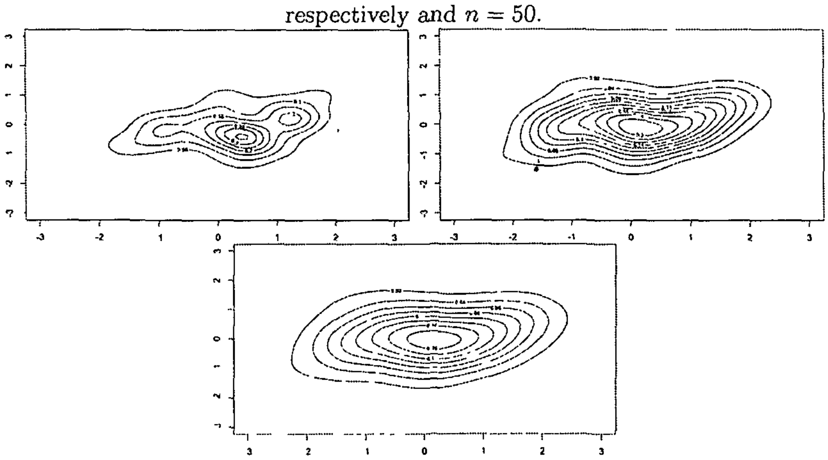


Fig 5.27: Contour plots of Trimodal density estimates, with $h = h_c, \hat{h}^*(\lambda = \frac{1}{n^{1/(8+d)}}, \frac{1}{n^{1/(2+d)}})$ respectively and $n = 200$.

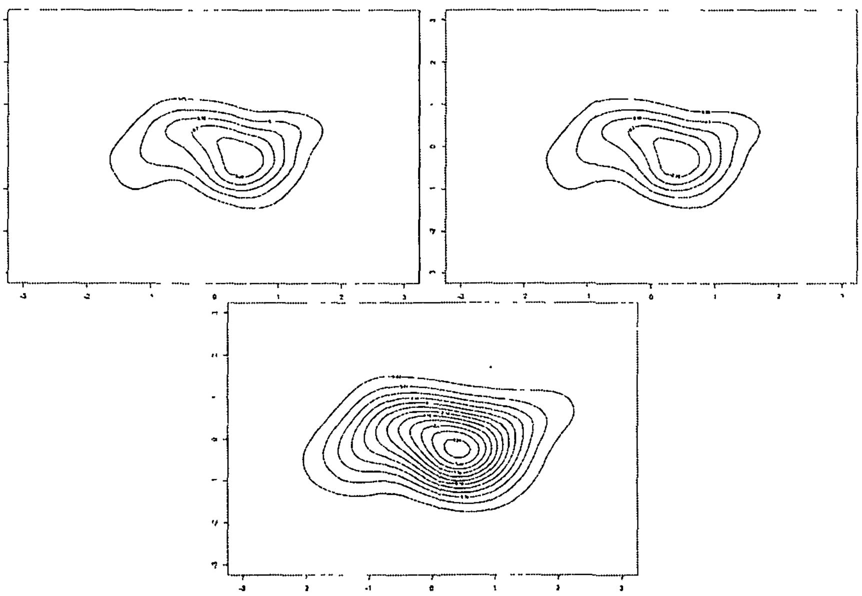


Fig 5.28: Contour plot of Quadrimodal density

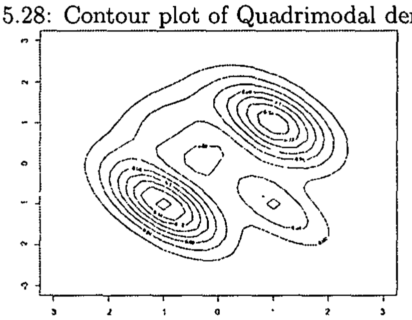


Fig 5.29: Contour plots of Quadrimodal density estimates, with $h = h_c, \hat{h}^*(\lambda = \frac{1}{n^{1/(8+d)}}, \frac{1}{n^{1/(2+d)}})$ respectively and $n = 50$.

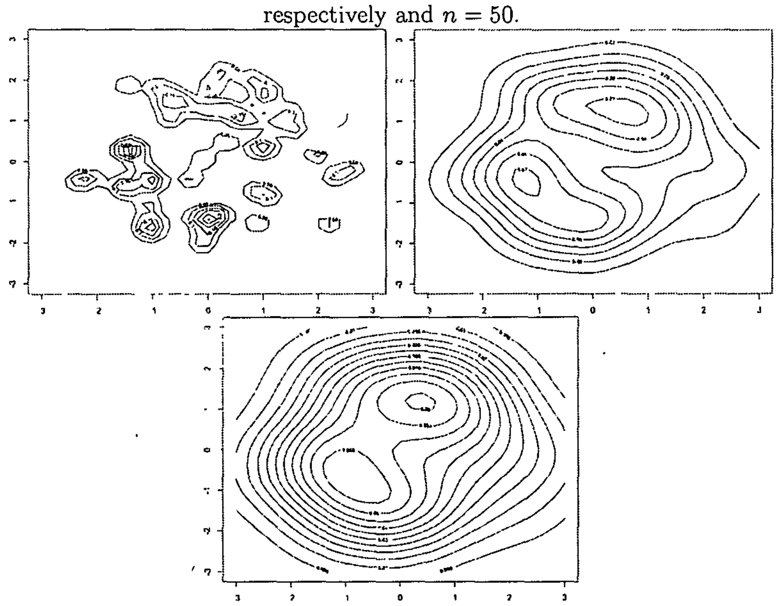
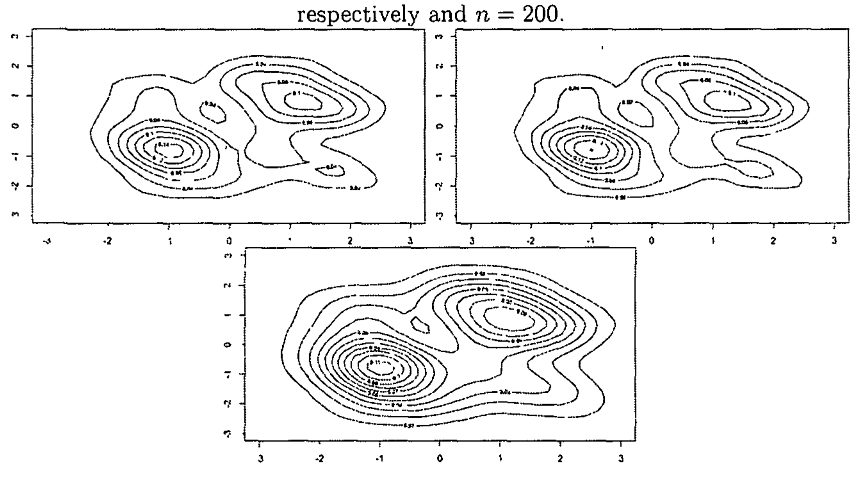


Fig 5.30: Contour plots of Quadrimodal density estimates, with $h = h_c, \hat{h}^*(\lambda = \frac{1}{n^{1/(8+d)}}, \frac{1}{n^{1/(2+d)}})$ respectively and $n = 200$.



5.4 Application in data analysis

We analyze a data set which consists of measurements of two flower parts, namely sepal length and sepal width, of a flower (botanical name is Iris Verginica). There are 50 pairs of sepal length and width measurements. The data is published in page 25, A handbook of small data sets, by Hand et. al.(1994).

Our objective is to visualize various features typical to the data set. We use contour and perspective plots of simple product kernel density estimates using both cross validation and proposed bootstrap bandwidth selection rule. These plots are obtained in R package, version 2.10.1.

Fig 5.31. Contour plots of density estimates for the Iris Verginica data, with $h = h_c$, h^* ($\lambda = \frac{1}{n^{1/(2+d)}}$, $\frac{1}{n^{1/(8+d)}}$) respectively

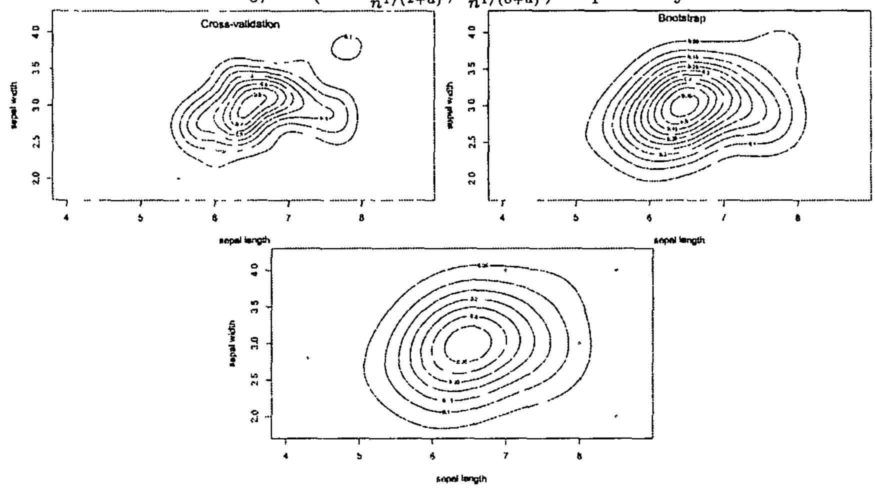
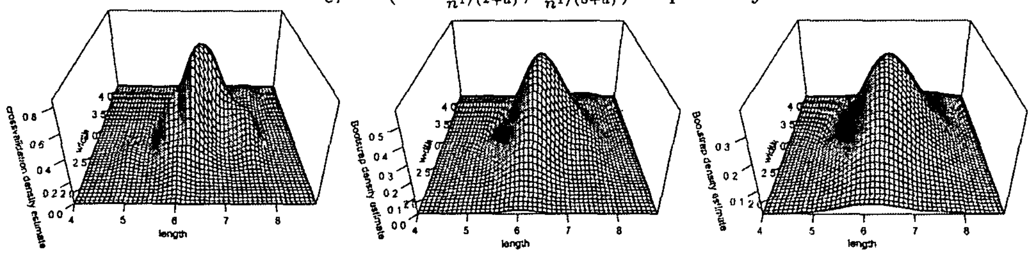


Fig 5.32: Perspective plots of density estimates for the Iris Verginica data, with $h = h_c$, h^* ($\lambda = \frac{1}{n^{1/(2+d)}}$, $\frac{1}{n^{1/(8+d)}}$) respectively.



Remark 5.4.1. The data possesses a global mode at (length, width)=(6.5,3). A weak local mode is visible in the neighborhood of (length, width)=(8,4). Note that the bootstrap bandwidth h^*, with $\lambda = \frac{1}{n^{1/(8+d)}}$, smooths local peaks that are visible in the contour and perspective plots with cross validation bandwidth.

5.5 Proofs

Proof of Theorem 5.2.1 Recall that h^* and $\hat{h}^*$ are minimizers of M and M^*, with respect to h, in $I = \left[\frac{\epsilon_1}{n^{1/(2s+1)}}, \frac{\epsilon_2}{n^{1/(2s+1)}}\right]$. Therefore $\hat{h}^*, h^* \in I$. Recalling (2.6.5), in chapter 2, we see that, almost surely

$$|M - M^*| \leq L_{1n} + L_{2n}, \quad \text{where } L_{1n},\ L_{2n} \text{ are as defined in proof of Theorem 2.3.1.}$$

Recalling (2.6.8), (2.6.12) and the definitions of f_{2y} and f_{3y} (in proof of Theorem 2.3.1 (iii)), it is easy to verify that the following inequalities hold, almost surely.

$$\begin{aligned} L_{1n} &\leq \frac{1}{n}\left[\int \{K_n^0(y) - f(y)\}^2 dy + 2\|f\| \int |K_n^0(y) - f(y)| dy\right] = e_{1n} \text{ (say)}, \\ L_{2n} &\leq \frac{h^{2s}}{[(s-1)!]^2}\left[\int f_{2y}^2 dy + 2\int [f_{2y}f_{3y}]\, dy\right] \\ &\leq \frac{\epsilon_2^{2s}}{n^{2s/(2s+1)}[(s-1)!]^2}\left[\int f_{2y}^2 dy + 2\int [f_{2y}f_{3y}]\, dy\right] = e_{2n} \text{ (say)}, \ \forall h \in I. \end{aligned}$$

We note that e_{1n} and e_{2n} are independent of h. Therefore $\|M - M^*\| = \sup_{h\in I}|M - M^*| \leq e_{1n} + e_{2n}$.
Hence using, $|\inf f - \inf g| \leq \|f - g\|$, we see that

$$\begin{aligned} E|M(\hat{h}^*) - M(h^*)| &\leq E|M(h^*) - M^*(\hat{h}^*)| + E|M(\hat{h}^*) - M^*(\hat{h}^*)| \\ &\leq 2E\|M - M^*\| \leq 2E(e_{1n} + e_{2n}). \end{aligned} \tag{5.5.1}$$

Recalling the arguments in the proof of Theorem 2.3.1 (i) we see that

$$E(e_{1n}) = o\left(\frac{1}{n}\right) \quad \text{for } s, p \geq 2. \tag{5.5.2}$$

Using Lemma 2.6.1(iii) and equation (2.6.15), (2.6.16), in chapter 2, we see that

$$\begin{aligned} \int E\left[f_{2y}^2\right] dy &\leq \left[\frac{\int |K(u)u^s| du}{s}\right]^2 \int E\left[K_n^{0(s)}(y) - f^{(s)}(y)\right]^2 dy \\ &= O\left(\frac{1}{n\lambda^{2s+1}} + \lambda^{2p}\right) \text{ and} \\ \int E\left[f_{2y}f_{3y}\right] dy &\leq \sqrt{\int E\left[f_{2y}^2\right] dy \cdot \int f_{3y}^2 dy} = O\left(\frac{1}{\sqrt{n\lambda^{2s+1}}} + \lambda^p\right). \end{aligned}$$

Substituting these inequalities in the definition of e_{2n} we get

$$E(e_{2n}) = O\left(\frac{1}{n^{2s/(2s+1)}} \cdot \left(\frac{1}{\sqrt{n\lambda^{2s+1}}} + \lambda^p\right)\right). \tag{5.5.3}$$

From equations (5.5.1), (5.5.2) and (5.5.3), we get

$$E\left|M(\hat{h}^*) - M(h^*)\right| = o\left(\frac{1}{n}\right) + O\left(\frac{1}{n^{2s/(2s+1)}}\left(\frac{1}{\sqrt{n\lambda^{2s+1}}} + \lambda^p\right)\right).$$

Further we note that $M \geq \frac{\int K^2}{nh} - \frac{C\int f^2(y)dy}{n}$, $\forall h \in I$ and hence

$$M(h^*) \geq \frac{\int K^2}{\epsilon_2 . n^{(2s)/(2s+1)}} + o\left(\frac{1}{n^{2s/(2s+1)}}\right).$$

Therefore $E\left|\frac{M(\hat{h}^*)}{M(h^*)} - 1\right| = o\left(\frac{1}{n^{1/(2s+1)}}\right) + O\left(\frac{1}{\sqrt{n\lambda^{2s+1}}} + \lambda^p\right)$. So Theorem 5.2.1 is proved completely. □

Proof of Theorem 5.2.3 Recall that h^* and $\hat{h}^*$ are minimizers of M and M^*, with respect to h, in $I = \left[\frac{\epsilon_1}{n^{1/(4+d)}}, \frac{\epsilon_2}{n^{1/(4+d)}}\right]$. Therefore $\hat{h}^*, h^* \in I$ and recalling the definitions of M and M^* and the arguments in proof of Theorem 4.4.4, we see that almost surely

$$\begin{aligned} |M - M^*| &\leq |V^* - V| + |B - B^*|, \text{ where} \\ |V - V^*| &\leq \frac{1}{n}\left\{\int a_{\vec{y}}^2 d\vec{y} + 2\sqrt{\int a_{\vec{y}}^2 d\vec{y} \int c_{\vec{y}}^2 d\vec{y}}\right\}, \\ \text{and } |B - B^*| &\leq h^4\left[\int f_{2\vec{y}}^2 d\vec{y} + 2.\sqrt{\int f_{2\vec{y}}^2 d\vec{y}\int f_{3\vec{y}}^2 d\vec{y}}\right] \end{aligned}$$

where $a_{\vec{y}}$, $c_{\vec{y}}$, $f_{2\vec{y}}$ and $f_{3\vec{y}}$ are as defined in the proof of Theorem 4.4 4 and hence it is easy to verify that

$$\begin{aligned} \int c_{\vec{y}}^2 d\vec{y} &\leq \int f^2(\vec{y})d\vec{y} = e_{1n} \text{ (say)}, \\ \int a_{\vec{y}}^2 d\vec{y} &\leq \int [w(\vec{y}) - f(\vec{y})]^2 d\vec{y} = e_{2n} \text{ (say)} \\ \int f_{2\vec{y}}^2 d\vec{y} &\leq d^2 \sum_{i_1,i_2=1}^{d} \frac{[\int \mathbf{K}(\vec{u})|u_{i_1}u_{i_2}|d\vec{u}]^2}{2^2} \int [f_{i_1,i_2}(\vec{y}) - g_{i_1,i_2}(\vec{y})]^2 d\vec{y} = e_{3n} \text{ (say)}, \\ \text{and } \int f_{3\vec{y}}^2 d\vec{y} &\leq d^2 \sum_{i_1,i_2=1}^{d} \frac{[\int \mathbf{K}(\vec{u})|u_{i_1}u_{i_2}|d\vec{u}]^2}{2^2} \int [f_{i_1,i_2}(\vec{y})]^2 d\vec{y} = e_{4n} \text{ (say)}. \end{aligned}$$

We note that e_{in}, $i = 1, .., 4$ are independent of h. Therefore

$$\begin{aligned} ||M - M^*|| &= \sup_{h\in I}|M - M^*| \\ &\leq \frac{1}{n}\left\{e_{2n} + 2\sqrt{e_{1n}e_{2n}}\right\} + \frac{\epsilon_2^4}{n^{4/(4+d)}}\left\{e_{3n} + 2\sqrt{e_{3n}e_{4n}}\right\} \end{aligned}$$

Since e_{1n}, e_{4n} are constants, therefore taking expectation on either side of the above inequality and using $E(\sqrt{X}) \leq \sqrt{E(X)}$, where X is a nonnegative random variable, we get

$$E||M - M^*|| \leq \frac{1}{n}\left\{E[e_{2n}] + 2\sqrt{e_{1n}E[e_{2n}]}\right\} + \frac{\epsilon_2^4}{n^{4/(4+d)}}\left\{E[e_{3n}] + 2\sqrt{E[e_{3n}]e_{4n}}\right\}.$$

$E[e_{3n}]$ is a linear combination of $\int E\left[f_{i_1,i_2}(\vec{y}) - g_{i_1,i_2}(\vec{y})\right]^2 d\vec{y}$, $i_1, i_2 = 1, 2, .., d$, and $E[e_{2n}] = \int E\left[w(\vec{y}) - f(\vec{y})\right]^2 d\vec{y}$.

We obtained the rates of convergence of $\int E\left[f_{i_1,i_2}(\vec{y}) - g_{i_1,i_2}(\vec{y})\right]^2 d\vec{y}$, $i_1, i_2 = 1, 2, .., d$, and $\int E\left[w(\vec{y}) - f(\vec{y})\right]^2 d\vec{y}$ in Lemma 4.6.2, in chapter 4. So recalling Lemma 4.6.2 (iii) and (iv) we get

$$E||M - M^*|| = O\left(\frac{1}{n}\sqrt{\frac{1}{n\mu^d} + \mu^4} + \frac{1}{n^{4/(4+d)}}\sqrt{\frac{1}{n\lambda^{d+4}} + \lambda^4}\right).$$

Further

$$E|M(\hat{h}^*) - M(h^*)| \leq E|M(h^*) - M^*(\hat{h}^*)| + E|M(\hat{h}^*) - M^*(\hat{h}^*)| \leq 2E||M - M^*||.$$

Therefore

$$E|M(\hat{h}^*) - M(h^*)| = O\left(\frac{1}{n}\sqrt{\frac{1}{n\mu^d} + \mu^4} + \frac{1}{n^{4/(4+d)}}\sqrt{\frac{1}{n\lambda^{d+4}} + \lambda^4}\right).$$

Further we note that there exists $D_1, D_2 > 0$ such that

$$M \geq \frac{D_1}{nh^d} - \frac{D_2}{n}, \ \forall h \in I$$

and hence

$$M(h^*) \geq \frac{\int K^2}{\epsilon_2^d . n^{4/(4+d)}} + o\left(\frac{1}{n^{4/(4+d)}}\right).$$

Therefore

$$E\left|\frac{M(\hat{h}^*)}{M(h^*)} - 1\right| = O\left(\sqrt{\frac{1}{n\mu^d} + \mu^4} + \sqrt{\frac{1}{n\lambda^{d+4}} + \lambda^4}\right).$$

So Theorem 5.2.3 is proved completely. □

Table 5.1: Bias and variances of $\frac{\hat{h}}{h^*}$, where $\hat{h} = h_a,\ h_{sj}, \hat{h}^*, \hat{h}^*_c,\ h_{sc},\ h_T,\ h_c$

n	Distributions				
	$\hat{h}$	Normal	Bimodal	Double-claw	Claw
50	h_a	-0.272 (0.014)	0.028 (0.008)	-0.483 (0.002)	5.146 (0.356)
	h_{sj}	-0.152 (0.034)	-0.065 (0.041)	-0.526 (0.012)	5.868 (1.953)
	$\hat{h}^*$	0.163 (0.011)	0.444 (0.009)	-0.277 (0.003)	8.581 (0.565)
	$\hat{h}^*_c$	0.014 (0.054)	0.216 (0.084)	-0.418 (0.017)	1.357 (15.786)
	h_{sc}	0.064 (0.022)	0.395 (0.018)	-0.295 (0.004)	10.943 (1.125)
	h_T	0.006 (0.012)	0.978 (0.020)	-0.301 (0.004)	8.157 (0.488)
	h_c	-0.099 (0.048)	0.053 (0.075)	-0.521 (0.019)	11.034 (8.937)
200	h_a	-0.215 (0.002)	0.168 (0.002)	-0.469 (0.001)	5.605 (0.181)
	h_{sj}	-0.122 (0.016)	0.009 (0.022)	-0.537 (0.004)	1.353 (2.026)
	$\hat{h}^*$	0.118 (0.003)	0.470 (0.007)	-0.327 (0.001)	15.735 (0.322)
	$\hat{h}^*_c$	-0.102 (0.043)	0.093 (0.055)	-0.505 (0.011)	-0.510 (0.077)
	h_{sc}	0.030 (0.005)	0.379 (0.011)	-0.370 (0.002)	15.735 (0.316)
	h_T	0.259 (0.004)	1.122 (0.031)	-0.168 (0.012)	8.018 (0.355)
	h_c	-0.173 (0.072)	-0.004 (0.080)	-0.557 (0.017)	-0.090 (0.002)

Table 5.2: $\hat{E}\left|\frac{M(\hat{h})}{M(h^*)} - 1\right|$ values for $\hat{h} = \hat{h}^*, h_a,\ h_{sj},\ h_{sc},\ h_T,\ h_c$

n	Distributions				
	$\hat{h}$	Normal	Bimodal	Double-claw	Claw
50	h_a	0.228	0.013	0.636	6.152
	h_{sj}	0.149	0.077	0.887	5.696
	$\hat{h}^*$	0.085	0.072	0.178	6.96
	$\hat{h}^*_c$	0.236	0.143	0.546	2.030
	h_{sc}	0.060	0.193	0.211	7.027
	h_T	0.030	0.656	0.217	6.950
	h_c	0.468	0.363	0.985	6.899
200	h_a	0.121	0.053	0.383	15.740
	h_{sj}	0.042	0.051	0.567	2.664
	$\hat{h}^*$	0.007	0.078	0.168	3.402
	$\hat{h}^*_c$	0.045	0.114	0.541	0.702
	h_{sc}	0.012	1.457	0.223	1 697
	h_T	0.163	0.242	0.062	21.304
	h_c	0.099	0.161	0.793	0.0204

Bibliography

[1] Abraham, C., Biau, G. and Cadre, B. (2003). Simple estimation of the mode of a multivariate density. *The Canadian Journal of Statistics.* 31, No.1, 23-34.

[2] Akaike, H. (1954). An approximation to the density function. *Ann. Inst. Statist. Math.* 6, 127-132.

[3] Billingsley, P. (1995) *Probability and Measure*, Third Edition. John Wiley and Sons, Inc.

[4] Bowman, A.W. (1984). An Alternative Method of Cross-Validation for the Smoothing of Density Estimates, *Biometrika*, 71, 353-360.

[5] Cacoullos, T. (1966). Estimation of a Multivariate Density. *Annals of the institute of Statistical Mathematics.* 18, 179-189.

[6] Cao-Abad, R. (1990). Applicaciones y nuevos resultados del metodo bootstrap en la estimacion no parametrica de curvas, Ph. D. dissertation (University of Santiago de Compostela).

[7] Cao, R. (1993). Bootstrapping the Mean Integrated Squared Error. *Journal of Multivariate Analysis*, No. 45, 137-160.

[8] Cao, R., Cuevas, A. and Gonzlez-Manteiga, W. (1994). A comparative study of several smoothing methods in density estimation. *Computational Statistics and Data Analysis*, 17, 153-176.

[9] Chaudhuri, P. and Marron, J.S. (1999). SiZer for exploration of Structures in curves. *J. Amer. Statist. Assoc.*, 94, No. 447, 807-825.

[10] Delaigle, A. and Gijbels, I. (2004). Bootstrap bandwidth selection in kernel density estimation from a contaminated sample. *Ann. Inst., Statist. Math* , 56, No. 1, 19-47.

[11] Devroye, L. and Gyorfi, L. (1985). *Nonparametric Density Estimation The L_1 View.* John Wiley and Sons, Inc.

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

[13] Dutta, S. (2010). Estimation of the MISE and the optimal bandwidth vector of a product kernel density estimate. *J Statist. Plann. Inference*, doi:10.1016/j.jspi.2010.11.032.

[14] Epanechnikov, V.A. (1969). Non-Parametric Estimation of a Multivariate Probability Density. *Theory of Probability and its Applications*, 14, 153-158.

[15] Falk, M. (1992). Bootstrap optimal bandwidth selection for kernel density estimates. *J. Statist. Plann Inference*, 30, 13-22.

[16] Faraway, J.J. and Jhun, M. (1990). Bootstrap choice of bandwidth for density estimation. *J. Amer. Statist. Assoc.*, 85, No. 412, 1119-1122.

[17] Gitzen, R.A. and Millspaugh, J. J. (2003). Comparison of Least-Squares Cross-Validation Bandwidth Options for Kernel Home-Range Estimation *Wildlife Society Bulletin*, Vol. 31, No.3, Autumn, 823-831

[18] Godtliebsen, F., Marron, J.S., Chaudhuri, P. (2000). Significance in scale space for bivariate density estimation. *Journal of Computational and Graphical Statistics.* March 1, 11(1), 1-21.

[19] Hall, P. (1982). Cross-Validation in Density Estimation. *Biometrika*, Vol. 69, No. 2, 383-390.

[20] Hall, P. and Marron, J.S. (1987). Estimation of integrated squared density derivatives. *Statistics and Probability Letters*, 6, 109-115.

[21] Hall, P. (1990). Using the bootstrap to estimate the mean squared error and select smoothing parameter in nonparametric problems. *Journal of Multivariate Analysis*, 32, Issue 2, 177-203.

[22] Hall, P. (1992). *The Bootstrap and Edgeworth Expansion.* Springer-Verlag New York, Inc.

[23] Hall, P., Marron, J.S. and Park, B.U. (1992). Smoothed cross-validation. *Probab Theory Related Fields*, 92, 1-20.

[24] Jones, M. C., Marron J.S. and Park, B.U. (1991) A simple root n bandwidth selector. *The Annals of Statistics*, 19, No. 4, 1919-1932.

[25] Jones, M.C., Marron, J.S. and Sheather, S.J. (1996). A brief survey of bandwidth selection for density estimation. *J. Amer. Statist. Assoc.*, 91, No. 433, 401-407.

[26] Loader C.R. (1999). Bandwidth Selection: Classical or Plug-in? *The Annals of Statistics.*, Vol. 27, No. 2, 415-438.

[27] Marron, J.S. and Park, B.U. (1990). Comparison of data-driven bandwidth selectors *J. Amer. Statist. Assoc.*, 85, No. 409, 66-72.

[28] Marron, J.S. and Wand, M.P. (1992). Exact mean integrated squared error. *The Annals of Statistics*, 20, No. 2, 712-736.

[29] Marron, J.S. and Chung, S.S. (1997). Presentation of smoothers: the family approach. Unpublished manuscript.

[30] Parzen, E. (1962). On estimation of a probability density function and mode.*Ann. Math. Statist.*, 33, 1065-1076.

[31] Rao, B.L.S.P. (1983). *Nonparametric Functional Estimation.* Academic Press, Inc, New York.

[32] Rudemo, M. (1982). Empirical Choice of Histogram and Kernel Density Estimators. *Scandinavian Journal of Statistics.*, 9, 65-78.

[33] Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function. *Ann. Math. Statist.*, 27, 832-837.

[34] Sain, S.R., Baggerly, K.A. and Scott, D.W. (1994). Cross-Validation of Multivariate Densities, *J. Amer. Statist. Assoc.*, 89, No. 427, 807-817, Theory and Methods.

[35] Sain, S.R. and Scott, D.W. (1996). On Locally Adaptive Density Estimation. *J. Amer. Statist. Assoc.*, 91, No. 436, Theory and Methods, 1525-1534.

[36] Scott, D.W. and Terrell, G.R. (1987). Biased and unbiased cross-validation in density estimation, *J. Amer. Statist. Assoc.*, 82, No. 400, 1131-1146, Theory and Methods

[37] Scott D W. (1992). *Multivariate Density Estimation* Theory, Practice and Visualization, John Wiley and Sons, Inc.

[38] Scott D.W. and Wand M.P. (1991). Feasibility of multivariate density estimates. *Biometrika*, 78, 1, 197-205.

[39] Silverman, B.W. (1986). *Density Estimation for Statistics and Data Analysis* Monographs on Statistics and Applied Probability. Chapman and Hall, London.

[40] Simonoff, S.J. (1996). *Smoothing Methods in Statistics*, Springer-Verlag New York, Inc.

[41] Shao, J. and Tu, D. (1995). *The Jackknife and Bootstrap*, Springer-Verlag New York, Inc.

[42] Sheather, S.J.,and Jones, M.C. (1991). A Reliable Data-Based Bandwidth Selection Method for Kernel Density Estimation, *Journal of the Royal Statistical Society*, Ser.B, 53, 683-690.

[43] Sheather, S.J. (2004). Density estimation, *Statistical Science*, 19, No. 4, 588-597.

[44] Tapia, R.A. and Thompson, J.R. (1978). *Nonparametric Probability Density Estimation.* John Hopkins Univ. Press, Baltimore.

[45] Taylor, C.C. (1989). Bootstrap choice of the smoothing parameter in kernel density estimation. *Biometrika*, 76, No.4, 705-12.

[46] Wand, M.P. and Jones, M.C. (1993). Comparison of Smoothing Parameterizations in Bivariate Kernel Density estimation. *J Amer. Statist. Assoc* , 88, No. 422, Theory and Methods.

[47] Wand, M.P. and Jones, M.C. (1995). *Kernel Smoothing*, Chapmann and Hall.