A comparative study of non-parametric distribution function estimators

H.J. Viljoen (Hons. B.Sc.)

Mini dissertation submitted in partial fulfilment of the requirements for the degree Magister Scientiae in Statistics at the North-West University

Supervisor: Prof F.C. van Graan

Co-Supervisor: Prof J.W.H. Swanepoel

2007

Potchefstroom

Abstract

The purpose of this study is to investigate a procedure for estimating the bandwidth in kernel distribution function estimation based on the bootstrap. To investigate the performance of the estimator, simulation studies were performed comparing the above mentioned estimator to estimators suggested by Altman and Léger (1995) and Van Graan (1983). The findings and conclusion of this study are reported.

Uittreksel

Die doel van hierdie studie is om 'n prosedure wat die bandwydte in kernfunksie-beraming vir verdelingsfunksies te beraam, te ondersoek. Om die kwaliteit van die beramer te ondersoek sal dit in 'n simulasiestudie met twee beramers voorgestel deur Altman and Léger (1995) en Van Graan (1983) vergelyk word. Die bevindinge en gevolgtrekkings van die simulasiestudie word bespreek.

Summary

The purpose of this study is to investigate a procedure for estimating the bandwidth in kernel distribution function estimation based on the bootstrap. To investigate the performance of the estimator, simulation studies were performed comparing the above mentioned estimator to estimators suggested by Altman and Léger (1995) and Van Graan (1983).

Chapter 1 gives an overview of kernel distribution function estimation, as well as the discrepancy measures commonly used in this estimation. Chapter 2 deals with the classical bootstrap procedure with some applications. Chapter 3 provides an overview of the existing methods for estimating the bandwidth. Chapter 4 introduces a new procedure based on the bootstrap for estimating the bandwidth and some of its properties. Chapter 5 describes the methodology followed in the simulation study as well as the conclusions that can be made.

In Chapter 1 the reader is given an introduction to kernel distribution function estimators where the origin, application and properties of these estimators are described. Several discrepancy measures commonly used in distribution function estimation are also discussed.

Chapter 2 describes the non-parametric classical bootstrap procedure. It explains various important aspects of the bootstrap including the plug-in principle and a bootstrap sample and gives several applications in statistical inference.

Chapter 3 provides an overview of the existing methods for estimating the bandwidth, which can be classified in two categories, namely those based on estimating an appropriate discrepancy measure and those based on estimating the asymptotical optimal bandwidth.

Chapter 4 introduces a new procedure based on the bootstrap for estimating the bandwidth. This procedure is investigated in a special case where the distribution is known and its asymptotical properties are investigated.

Chapter 5 deals with the simulation study used in this study. It explains the outputs, which can be found in Appendixes A and B, as well as the inputs and the algorithm used by

Opsomming

Die hoofdoel van die studie is om 'n prosedure wat die strookwydte beraam in kern verdelingsfunksie beraming, te ondersoek. Die prosedure, gebaseer op die skoenlusmetode, sal ook vergelyk word met twee ander prosedures om die strookwydte te kies, naamlik die prosedures soos voorgestel deur Altman en Léger (1995) en Van Graan (1983).

Hoofstuk 1 bied 'n oorsig oor kernfunksie-beraming sowel as die verlies-funksies wat algemeen daarmee gepaard gaan. Hoofstuk 2 handel oor die klassieke skoenlusmetode en sommige van sy toepassings. Hoofstuk 3 gee 'n oorsig van die bestaande metodes om die strookwydte te beraam. Hoofstuk 4 stel 'n nuwe prosedure voor, gebaseer op die skoenlusmetode, om die strookwydte te beraam en bespreek sommige kenmerke van die beramer. In hoofstuk 5 word die simulasiestudie beskryf sowel as die resultate daarvan.

Hoofstuk 1 bied aan die leser 'n oorsig van kernfunksie-beraming, insluitende die oorsprong, toepassing en kenmerke van die beramers. Daar word ook gekyk na verliesfunksies en hul gebruik in kernfunksie-beraming.

Hoofstuk 2 beskryf die nie-parametriese skoenlusmetode en verduidelik verskeie aspekte van die metode asook toepassings daarvan.

Hoofstuk 3 bied 'n oorsig oor die bestaande prosedures om die strookwydte in kernfunksie-beraming te beraam. Die prosedures kan in twee kategorie verdeel word, naamlik die wat 'n toepaslike verliesfunksie beraam en die wat die asimptoties optimale strookwydte beraam.

Hoofstuk 5 beskryf die simulasiestudie van die ondersoek. Die algoritme van die program, wat in die binneblad van die skripsie gevind kan word, word bespreek. 'n Bespreking van die resultate van die simulasiestudie wat in Bylae A en B gekry kan word, word ook hier gegee.

Dankbetuigings

Die skrywer wil hiermee graag die volgende bedankings doen:

- Prof. F. C. van Graan, die studieleier, vir die voorstel van die onderwerp, sy leiding, geduld en ondersteuning wat nodig was vir die voltooiing van die studie.
- Prof. J. W. H. Swanepoel, die medestudieleier vir sy waardevolle insette.
- My vrou, Carol, vir al haar liefde, geduld en ondersteuning
- My ouers en my suster vir my opvoeding en hul volgehoue ondersteuning
- Vir die geleentheid om die studie aan te pak en die voorreg om dit te voltooi bedank ek God.

Notation

Symbol	Description
$F(x)$	The distribution function of a random variable x
$\mathbf{X}_n$ or $(X_1, X_2, \ldots, X_n)$	A sample vector of n independent identically distributed random variables
$F_n(x)$	The empirical distribution function (EDF)
$\mathrm{I}(A)$	The indicator function of the event A
$f_{n,h}$	The kernel density function estimator
$k(u)$	The kernel density function
$K(u)$	The kernel distribution function
$F_{n,h}(x)$	The kernel distribution function estimator
h	The bandwidth of the kernel estimator
$X_{(1)}, X_{(2)}, \ldots, X_{(n)}$	The order statistics of $\mathbf{X}_n$
$\mu_j(k)$	$\int_{-\infty}^{\infty} z^j k(z) dz$
$D_j(k)$	$\int_{-\infty}^{\infty} z^j k(z) K(z) dz$
$\mathrm{MSE}(h)$	The mean squared error when using bandwidth h
$\mathrm{ISE}(h)$	The integrated squared error when using bandwidth h
$\mathrm{MISE}(h)$	The mean integrated squared error when using bandwidth h
$\mathrm{ASE}(h)$	The average squared error when using bandwidth h
$\mathrm{MASE}(h)$	The mean average squared error when using bandwidth h
$\mathbf{X}_n^*$	The bootstrap sample vector of size n
$\mathrm{E}_* [t(F_n)]$	The bootstrap expected value of functional $t(F_n)$
V_2	$2D_1(K)A_0(F)$
B_3	$\frac{1}{4}\mu_2^2(k)\Psi_{1,1}(F)$
C	$D_2(K)A_1(F) - E_1(F)\mu_2(k)$

Symbol	Description
$A_j(F)$	$\int_{-\infty}^{\infty} f^{(j)}(x)W(x)dF(x)$
$V_1(F)$	$\int_{-\infty}^{\infty} F(x)\left[1-F(x)\right]W(x)dF(x)$
$\Psi_{1,1}(F)$	$\int_{-\infty}^{\infty} \left[f'(x)\right]^2 W(x)dF(x)$
$E_j(F)$	$\int_{-\infty}^{\infty} F(x)f^{(j)}(x)dF(x)$
h_{opt}	$\left[\frac{V_2}{4B_3}\right]^{1/3} n^{-1/3}$
LNO(h)	"Leave-non-out" estimator of ASE
CV(h)	"Cross-validation" criterion
$\beta_{n,h}(x,F)$	The bias of $F_{n,h}(x)$
$b_{n,h}(x,F_n)$	The bootstrap estimator of $\beta_{n,h}(x,F)$
$\hat{h}$	The data-driven bandwidth obtained from a procedure for a random sample $X_1, X_2, \ldots, X_n$ from a distribution F
d_{ASE}	$\text{E}\left[\text{ASE}\left(\hat{h}\right)\right]$
$\hat{d}_{\text{ASE}}$	$\frac{1}{\text{MC}}\sum_{i=1}^{\text{MC}} \text{ASE}\left(\hat{h}_i\right)$
$\overline{h}$	$\frac{1}{\text{MC}}\sum_{i=1}^{\text{MC}} \hat{h}_i$
SE	Standard error of $\overline{h}$ in simulation study

Contents

Chapter 1

Introduction

1.1 Distribution function estimators

In the area of statistical inference, estimation of the unknown distribution function $F(x)$, of a population is important. Other statistical methods that are dependent on knowledge of the distribution function include hypothesis testing and confidence interval estimation. Also, if one is interested in estimating the proportion of elements in the population that are less than or equal to x, then knowledge of the distribution function is essential.

Existing methods to estimate an unknown distribution function from data can be classified into two groups, namely: parametric and non-parametric methods. Parametric methods are dependent on the assumption that the functional form of the distribution function is specified. If it is known that the data are normally distributed with unknown mean and variance, the unknown parameters can be estimated from the data and the distribution function is then completely determined. If the assumption of normality cannot be made, then the parametric method of estimation cannot be used and a non-parametric method of estimation must be implemented.

Non-parametric methods have the advantage that they are independent of distributional assumptions. The best known and simplest non-parametric estimator of the distribution function is the empirical distribution function (EDF). Let $X_1, X_2, \ldots, X_n$ be i.i.d. random variables. Then the EDF is defined by

$$F_n(x) = \frac{1}{n}\sum_{k=1}^{n} I(X_k \leq x),$$

where

$$\mathrm{I}(A) \;=\; \begin{cases} 1, & \text{if } A \text{ occurs;} \\ 0, & \text{if } A^c \text{ occurs..} \end{cases}$$

Note that $I(X_k \leq x)$ are independent Bernoulli random variables such that:

$$\mathrm{I}(X_k \leq x) = \begin{cases} 1, & \text{with probability } F(x); \\ 0, & \text{with probability } 1 - F(x). \end{cases}$$

Thus $nF_n(x)$, is a binomial random variable (n trials, probability $F(x)$ of success) and so

$$\begin{aligned} \mathrm{E}\left[F_n(x)\right] &= F(x) \\ \mathrm{Var}\left[F_n(x)\right] &= \frac{1}{n}F(x)\left[1 - F(x)\right]. \end{aligned}$$

$$\begin{aligned} \text{Furthermore, } \mathrm{E}\left[F_n(x) - F(x)\right]^2 &= \mathrm{Var}\left[F_n(x)\right] + \left[\mathrm{E}\left\{F_n(x)\right\} - F(x)\right]^2 \\ &= \frac{1}{n}F(x)\left[1 - F(x)\right] \to 0 \text{ as } n \to \infty. \end{aligned}$$

Thus as an estimator of $F(x)$, $F_n(x)$ is unbiased and its variance tends to 0 as $n \to \infty$. Also, the Glivenko-Cantelli theorem states that F can be approximated by F_n in an uniform manner for large sample sizes such that

$$\lim_{n\to\infty} \sup_{-\infty<x<\infty} |F_n(x) - F(x)| = 0 \text{ almost surely.} \tag{1.1}$$

It was also shown by Jacod and Protter (2000) that the order of convergence is $O(n^{-1/2}(\log\log n)^{1/2})$ almost surely.

Despite its good theoretical properties, the empirical distribution function is a step function. In many applications, a smooth estimate of F is desired. Examples include estimating the tails of F and in survival analysis, where it would be advantageous to have a smooth estimate of F. A non-parametric alternative to F_n has been introduced in Nadaraya (1964). This estimator can be derived from the Rosenblatt-Parzen density estimate, see Rosenblat (1956) and Parzen (1962); namely

$$f_{n,h}(x) = \frac{1}{nh}\sum_{i=1}^{n} k\left(\frac{x - X_i}{h}\right),$$

where $h = h(n)$ is a sequence of smoothing parameters (also called the bandwidth), for which it is required that $h \to 0$ and $nh \to \infty$ as $n \to \infty$. The kernel function k usually satisfies the requirements:

1. $k(u) \geq 0, \forall\, u \in \mathbb{R}$.

2. $\int_{-\infty}^{+\infty} k(u)du = 1$, hence k is a density function.

3. $k(-u) = k(u)$, hence k is symmetric function.

4. $\int_{-\infty}^{+\infty} uk(u)dx = 0$.

The non-parametric kernel distribution function estimator is then defined as

$$\begin{aligned} F_{n,h}(x) &= \int_{-\infty}^{x} f_{n,h}(t)dt \\ &= \frac{1}{n}\sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right). \end{aligned} \tag{1.2}$$

where $K(u) = \int_{-\infty}^{u} k(t)dt$, i.e. K is the distribution function corresponding to k.

Note that the kernel distribution function estimator includes the EDF as an extreme case. To see this, let $h \longrightarrow 0$, then

$$\mathrm{K}\left(\frac{x - X_i}{h}\right) = \begin{cases} 0, & \text{if } X_i > x; \\ \frac{1}{2}, & \text{if } X_i = x; \\ 1, & \text{if } X_i < x. \end{cases}$$

Hence, as $h \longrightarrow 0$, it can be shown that $F_{n,h}(x) \longrightarrow F_{n,0}(x)$ for all x, where

$$F_{n,0}(x) = \frac{1}{n}\sum_{i=1}^{n}\left[I(X_i < x) + \frac{1}{2}I(X_i = x)\right].$$

Van Graan (1983) demonstrated that $F_{n,0}(x)$ differs from $F_n(x)$ only when x is equal to one of the order statistics $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$, i.e.,

$$F_{n,0}(x) = \begin{cases} \frac{i}{n} + \frac{1}{2n} = \frac{i - \frac{1}{2}}{n}, & \text{if } x = X_{(i)}; \\ \frac{i}{n} = F_n(x), & \text{if } x \neq X_{(i)}. \end{cases}$$

Thus kernel distribution function estimators also includes the empirical distribution function as an extreme case.

In order to compare the kernel distribution function estimator to the EDF, expressions for the aforementioned estimator will now be derived, see Van Graan (1983). To obtain $\mathrm{Var}\,[F_{n,h}(x)]$ note that (under certain conditions on F and K)

$$\begin{aligned} \mathrm{Var}\,[F_{n,h}(x)] &= \frac{1}{n}\mathrm{Var}\left[K\left(\frac{x - X}{h}\right)\right] \\ &= \frac{1}{n}\left\{\mathrm{E}\left\{\left[K\left(\frac{x - X}{h}\right)\right]^2\right\} - \left[\mathrm{E}\left\{K\left(\frac{x - X}{h}\right)\right\}\right]^2\right\}. \end{aligned} \tag{1.3}$$

Now

$$\mathrm{E}\left[K\left(\frac{x-X}{h}\right)\right] = \int_{-\infty}^{\infty} K\left(\frac{x-y}{h}\right) dF(y).$$

Using partial integration, the substitution $\frac{x-y}{h} = z$ and a Taylor series expansion it follows that

$$\begin{aligned}
&\mathrm{E}\left[K\left(\frac{x-X}{h}\right)\right] \\
&= \frac{1}{h}\int_{-\infty}^{\infty} F(y)k\left(\frac{x-y}{h}\right) dy \\
&= \int_{-\infty}^{\infty} F(x-hz)k(z)dz \\
&= \int_{-\infty}^{\infty} \left[F(x) - hzf(x) + \frac{1}{2}h^2z^2f'(x) + h^3R_1(x,z)\right] k(z)dz \\
&= F(x) + \frac{1}{2}h^2f'(x)\mu_2(k) + O(h^3), \qquad (1.4)
\end{aligned}$$

where

$$\mu_j(k) = \int_{-\infty}^{\infty} z^j k(z)dz \text{ for } j \geq 1 \qquad (1.5)$$

and $R_1(x,z)$ represents a remainder term. From (1.4) it is clear that $\mathrm{E}\left[F_{n,h}(x)\right]$ and consequently the bias of $F_{n,h}(x)$ is $O(h^3)$, which approaches 0 as h approaches 0.

Using a similar approach as above an expression for $\mathrm{E}\left\{\left[K\left(\frac{x-X}{h}\right)\right]^2\right\}$ can be obtained:

$$\begin{aligned}
\mathrm{E}\left\{\left[K\left(\frac{x-X}{h}\right)\right]^2\right\} &= \int_{-\infty}^{\infty}\left[K\left(\frac{x-y}{h}\right)\right]^2 dF(y) \\
&= \frac{2}{h}\int_{-\infty}^{\infty} F(y)K\left(\frac{x-y}{h}\right)k\left(\frac{x-y}{h}\right) dy \\
&= 2\int_{-\infty}^{\infty} F(x-hz)K(z)k(z)dz \\
&= 2\int_{-\infty}^{\infty} K(z)k(z)\left[F(x) - hzf(x) + \frac{1}{2}h^2z^2f'(x) + h^3R_1(x,z)\right] dz \\
&= F(x) - 2hf(x)\int_{-\infty}^{\infty} zK(z)k(z)dz + O(h^2), \\
&= F(x) - 2hf(x)D_1(K) + O(h^2) \qquad (1.6)
\end{aligned}$$

where

$$D_j(K) = \int_{-\infty}^{\infty} z^j k(z)K(z)dz \text{ for } j \geq 0. \qquad (1.7)$$

Using (1.3), (1.4) and (1.6) an expression for $\mathrm{Var}\left[F_{n,h}(x)\right]$ is given by

$$\begin{aligned}
\mathrm{Var}\left[F_{n,h}(x)\right] &= \frac{1}{n}\left[F(x) - 2hf(x)D_1(K) + O(h^2) - \left\{F(x) + \frac{1}{2}h^2f'(x)\mu_2(k) + O(h^3)\right\}^2\right] \\
&= \frac{F(x)\left\{1-F(x)\right\}}{n} - 2\frac{h}{n}f(x)D_1(K) + O\left(\frac{h^2}{n}\right),
\end{aligned}$$

where $2f(x)D_1(K) > 0$. The previous result shows that the asymptotic variance of $F_{n,h}$ is smaller than the variance of the EDF. It is evident that for larger values of h, the quantity $2hf(x)D_1(K)$ increases, resulting in a smaller variance expression but larger bias. This observation has important implications for choosing the bandwidth.

Several other properties of the estimator $F_{n,h}$ have been investigated. Nadaraya (1964), Winter (1973) and Yamato (1973) proved almost uniform convergence of $F_{n,h}$ to F; Watson and Leadbetter (1964) established asymptotic normality for $F_{n,h}$, and Winter (1979) showed that $F_{n,h}$ has the Chung-Smirnov property. Reiss (1981) pointed out that the loss in bias with respect to F_n is compensated by a gain in variance. This result is referred to as the deficiency of F_n with respect to $F_{n,h}$. Falk (1983) provided a complete solution to the question as to which of F_n or $F_{n,h}$ is the better estimator of F. Using the concept of relative deficiency, conditions (as $n \to \infty$) on K and $h = h(n)$ are derived, which enables the user to decide exactly whether a given kernel distribution function estimator should be preferred to the EDF. It is generally accepted that the choice of bandwidth is more important than the choice of the kernel function K. This study will focus on methods for choosing the bandwidth.

1.2 Discrepancy measures

In order to propose methods for estimating the bandwidth, discrepancy measures that quantify the quality of $F_{n,h}$ as an estimator for F must be introduced. One such measure is the mean squared error, which in the case of the kernel distribution function estimator is defined as

$$\mathrm{MSE}(h) = \mathrm{E}\left\{[F_{n,h}(x) - F(x)]^2\right\}. \tag{1.8}$$

However, minimising (1.8) with respect to h will yield a value that depends on x. To overcome this problem, it is necessary to use a discrepancy measure that measures the difference between $F_{n,h}$ and F over all possible values. A global discrepancy measure is the integrated squared error given by

$$\mathrm{ISE}(h) = \int_{-\infty}^{\infty} [F_{n,h}(x) - F(x)]^2 \, dx, \tag{1.9}$$

or the mean integrated squared error, defined by

$$\mathrm{MISE}(h) = \mathrm{E}\left\{\int_{-\infty}^{\infty} [F_{n,h}(x) - F(x)]^2 \, dx\right\}. \tag{1.10}$$

Since each x-value is weighed equally in the measures above, a more general choice would be to weigh x-values according to the probability of a particular x-value, i.e., define MISE(h) as

$$\text{MISE}(h) = \text{E}\left\{\int_{-\infty}^{\infty} [F_{n,h}(x) - F(x)]^2 \, dF(x)\right\}. \tag{1.11}$$

By introducing a nonnegative weight function $W(x)$, the measure in (1.11) can be generalised to

$$\text{MISE}(h) = \text{E}\left\{\int_{-\infty}^{\infty} [F_{n,h}(x) - F(x)]^2 \, W(x) dF(x)\right\}. \tag{1.12}$$

The advantage of the discrepancy measure in (1.12) is that it is always bounded. Using the results of the previous section the discrepancy measure in (1.11) can be written in one of two alternative forms, see Van Graan (1983):

$$\begin{aligned}\text{MISE}(h) &= \frac{1}{n}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left[K\left(\frac{x-y}{h}\right) - F(x)\right]^2 dF(y)dF(x) \\ &\quad + \left(1-\frac{1}{n}\right)\int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty}\left\{K\left(\frac{x-y}{h}\right) - F(x)\right\}dF(y)\right]^2 dF(x),\end{aligned} \tag{1.13}$$

or

$$\begin{aligned}\text{MISE}(h) &= \frac{1}{n}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} K^2\left(\frac{x-y}{h}\right) dF(y)dF(x) \\ &\quad -\frac{1}{n}\int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty} K\left(\frac{x-y}{h}\right) dF(y)\right]^2 dF(x) \\ &\quad + \int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty} K\left(\frac{x-y}{h}\right) - F(x)\right]^2 dF(x).\end{aligned} \tag{1.14}$$

Sarda (1993) proposed a discrete approximation of the measure in (1.12), known as the average squared error

$$\text{ASE}(h) = \frac{1}{n}\sum_{i=1}^{n} [F_{n,h}(X_i) - F(X_i)]^2 \, W(X_i). \tag{1.15}$$

Finally, another discrepancy measure based on the random variable in (1.15), called the mean average squared error is defined by

$$\begin{aligned}\text{MASE}(h) &= \text{E}\left\{\frac{1}{n}\sum_{i=1}^{n} [F_{n,h}(X_i) - F(X_i)]^2 \, W(X_i)\right\} \\ &= \frac{1}{n}\sum_{i=1}^{n}\text{E}\left\{[F_{n,h}(X_i) - F(X_i)]^2 \, W(X_i)\right\}.\end{aligned} \tag{1.16}$$

In literature one of two methods are usually followed for obtaining an estimate of the bandwidth. In the first method an estimate of one of the discrepancy measures defined above

is obtained, and then minimised with respect to h. The second method consists of finding an asymptotic expression for (1.11) or (1.12), obtaining an optimal value of h and then estimating the unknown quantities in this expression.

This study will concentrate on the discrepancy measure defined in (1.16). The main aim of this study will be to propose an estimator for this measure, minimising the expression with respect to h, obtaining a data based choice of the bandwidth. The case when both F and K are known will be considered in Section 4.3. In order to do this, the expression in (1.16) must be written in a more manageable form.

We will now derive the following results:

Let $W(X_i) = 1$. Since K is a symmetric distribution function

$$\begin{aligned} F_{n,h}(X_j) &= \frac{1}{n}\sum_{i=1}^{n} K\left(\frac{X_j - X_i}{h}\right) \\ &= \frac{1}{n}K(0) + \frac{1}{n}\sum_{i\neq j}^{n} K\left(\frac{X_j - X_i}{h}\right) \\ &= \frac{1}{2n} + \frac{1}{n}\sum_{i\neq j}^{n} K\left(\frac{X_j - X_i}{h}\right). \end{aligned} \tag{1.17}$$

Using the fact that $X_1, X_2, \ldots, X_n$ are i.i.d. and the result in (1.17)

$$\begin{aligned} &\text{MASE}(h) \\ &= \mathrm{E}\left\{\frac{1}{n}\sum_{i=1}^{n}[F_{n,h}(X_i) - F(X_i)]^2\right\} \\ &= \frac{1}{n}\sum_{i=1}^{n}\mathrm{E}\left\{[F_{n,h}(X_i) - F(X_i)]^2\right\} \\ &= \mathrm{E}\left\{[F_{n,h}(X_1) - F(X_1)]^2\right\} \\ &= \mathrm{E}\left\{\left[\frac{1}{2n} + \frac{1}{n}\sum_{i\neq 1}^{n} K\left(\frac{X_1 - X_i}{h}\right) - F(X_1)\right]^2\right\} \\ &= \mathrm{E}\left[\frac{1}{4n^2} + \frac{1}{n}\left\{\frac{1}{n}\sum_{i\neq 1}^{n} K\left(\frac{X_1 - X_i}{h}\right) - F(X_1)\right\}\right. \\ &\qquad \left. + \left\{\frac{1}{n}\sum_{i\neq 1}^{n} K\left(\frac{X_1 - X_i}{h}\right) - F(X_1)\right\}^2\right]. \end{aligned} \tag{1.18}$$

The last two terms in (1.18) will be treated separately. Note that

$$\begin{aligned} &\frac{1}{n}\sum_{i\neq 1}^{n} K\left(\frac{X_1 - X_i}{h}\right) - F(X_1) \\ &= \frac{1}{n}\sum_{i\neq 1}^{n}\left[K\left(\frac{X_1 - X_i}{h}\right) - \frac{n}{n-1}F(X_1)\right]. \end{aligned} \tag{1.19}$$

and by taking the expected value it follows that

$$= \mathrm{E}\left[\frac{1}{n}\left\{\frac{1}{n}\sum_{i\neq 1}^{n} K\left(\frac{X_1 - X_i}{h}\right) - F(X_1)\right\}\right]$$
$$= \frac{n-1}{n^2}\,\mathrm{E}\,K\left(\frac{X_1 - X_2}{h}\right) - \frac{1}{n}\,\mathrm{E}\,F(X_1). \tag{1.20}$$

Using the result in (1.19), the last term in (1.18) can be written as

$$\left\{\frac{1}{n}\sum_{i\neq 1}^{n}\left[K\left(\frac{X_1 - X_i}{h}\right) - \frac{n}{n-1}F(X_1)\right]\right\}^2$$
$$= \frac{1}{n^2}\sum_{i\neq 1}^{n}\left[K\left(\frac{X_1 - X_i}{h}\right) - \frac{n}{n-1}F(X_1)\right]^2$$
$$+\frac{1}{n^2}\sum_{j\neq 1}^{n}\sum_{\substack{i\neq 1\\ j\neq i}}^{n}\left\{\left[K\left(\frac{X_1 - X_j}{h}\right) - \frac{n}{n-1}F(X_1)\right]\times\right.$$
$$\left.\left[K\left(\frac{X_1 - X_i}{h}\right) - \frac{n}{n-1}F(X_1)\right]\right\}. \tag{1.21}$$

Expanding the square and taking the expectation of the first term in the previous expression it follows that

$$\mathrm{E}\left\{\frac{1}{n^2}\sum_{i\neq 1}^{n}\left[K\left(\frac{X_1 - X_i}{h}\right) - \frac{n}{n-1}F(X_1)\right]^2\right\}$$
$$= \frac{n-1}{n^2}\,\mathrm{E}\left\{K^2\left(\frac{X_1 - X_2}{h}\right)\right\} - \frac{2}{n}\,\mathrm{E}\left\{F(X_1)\,K\left(\frac{X_1 - X_2}{h}\right)\right\}$$
$$+\frac{1}{n-1}\,\mathrm{E}\left\{F^2(X_1)\right\}. \tag{1.22}$$

Similarly it follows that

$$\mathrm{E}\left\{\frac{1}{n^2}\sum_{j\neq 1}^{n}\sum_{\substack{i\neq 1\\ j\neq i}}^{n}\left[K\left(\frac{X_1 - X_j}{h}\right) - \frac{n}{n-1}F(X_1)\right]\left[K\left(\frac{X_1 - X_i}{h}\right) - \frac{n}{n-1}F(X_1)\right]\right\}$$
$$= \frac{(n-1)(n-2)}{n^2}\,\mathrm{E}\left[K\left(\frac{X_1 - X_2}{h}\right)K\left(\frac{X_1 - X_3}{h}\right)\right]$$
$$-\frac{2(n-2)}{n}\,\mathrm{E}\left[F(X_1)\,K\left(\frac{X_1 - X_2}{h}\right)\right] + \frac{n-2}{n-1}\,\mathrm{E}\left[F^2(X_1)\right]. \tag{1.23}$$

Using the results obtained in (1.18), (1.20), (1.22) and (1.23) and noting that $\mathrm{E}\left[F(X_1) = \frac{1}{2}\right]$

and $\mathrm{E}\left[F^2\left(X_1\right)=\frac{1}{3}\right]$, MASE$(h)$ can be written as

$$\begin{aligned}\mathrm{MASE}\left(h\right)=\frac{n-1}{n^2}\left[\mathrm{E}\left\{K\left(\frac{X_1-X_2}{h}\right)\right\}+\mathrm{E}\left\{K^2\left(\frac{X_1-X_2}{h}\right)\right\}\right]\\ +\frac{\left(n-1\right)\left(n-2\right)}{n^2}\mathrm{E}\left\{K\left(\frac{X_1-X_2}{h}\right)K\left(\frac{X_1-X_3}{h}\right)\right\}\\ -\frac{2\left(n-1\right)}{n}\mathrm{E}\left[F\left(X_1\right)K\left(\frac{X_1-X_2}{h}\right)\right]+\frac{1}{3}-\frac{1}{2n}+\frac{1}{4n^2}.\end{aligned}\tag{1.24}$$

Remarks

1. In Section 4.2 an exact calculation of the MASE discrepancy measure will be done in a special case, using the expression in (1.24).

2. In Section 4.3 asymptotic theory based on this measure will be developed, using the expression in (1.24).

Chapter 2

Bootstrap methodology

2.1 Introduction

The classical bootstrap, introduced by Efron in 1979, is described as a non-parametric, computer intensive statistical methodology used in a wide range of applications. A popular use thereof is to describe the variance of the data, or to estimate the standard error. The advantage in using the bootstrap instead of classical statistical theory is that it does not require any model assumptions, such as assumptions about the statistical distribution of the population. In practice, one does not always have this information at hand and has to resort to other methods such as making assumptions that are not always correct. One can then resort to large sampling theory, but then the statistician has to decide if the sample is large enough and if his sample too small, what to do then.

The bootstrap, however, does not have these problems, and can be used in cases when there are small sample sizes involved; in these cases large sample theory will fail. Another advantage of the bootstrap is that finding the variance or standard error of complex statistics are relatively easy with the bootstrap. It does, however, require a lot of processing power and as such its widespread use was only made possible by the power of modern computers. This chapter will introduce the classical bootstrap procedure as described by Efron and Tibshirani (1993).

In short, the bootstrap is implemented by sampling from the original sample of size n with replacement. The size of this sample is usually taken to be n with the classical bootstrap, but can also be less. Then we refer to the procedure as the modified bootstrap. Interested readers are referred to (Swanepoel, 1986). This sampling is repeated a large number of times

and these are called bootstrap replications. The distribution of these replications serves as an approximation to the sampling distribution of the statistic under consideration.

The accuracy of the estimator $\hat{\theta} = t(F_n)$, where t is some functional of the population distribution function F, depends on how well F_n approximates F. The Glivenko-Cantelli theorem given in (1.1) gives an indication of how well this approximation is.

2.2 The classical bootstrap procedure

Let $\mathbf{X}_n = (X_1, X_2, \ldots, X_n)$ be an independent, identically distributed sample of data from the unknown population distribution F. Let F_n be the EDF as defined in Section 1.1. As noted in Section 1.1 $nF_n(x)$ is a binomial random variable (n trials and probability $F(x)$ of a success), with expected value and variance as calculated in Section 1.1. As previously mentioned one can estimate $F(x)$ using the EDF, which places mass $\frac{1}{n}$ on each element of $\mathbf{X}_n$.

The empirical distribution can then be used to generate a bootstrap sample denoted by $\mathbf{X}_n^* = (X_1^*, X_2^*, \ldots, X_n^*)$, which is the same as sampling with replacement from the data $X_1, X_2, \ldots, X_n$. Stated in terms of probabilities

$$P^*(X_j^* = X_i) = \frac{1}{n}, \forall\ i, j = 1, \ldots, n,$$

where P^* is the probability calculated under F_n.

Now, let $T_n(\mathbf{X}_n; F)$ be some specified random variable that we are interested in. In the classical bootstrap method the sampling distribution of $T_n(\mathbf{X}_n; F)$ under F, will be estimated by the bootstrap distribution of $T_n(\mathbf{X}_n^*; F_n)$ under F_n.

The estimate $T_n(\mathbf{X}_n^*; F_n)$ depends on how well F_n approximates F (see section 1.1). One method of calculation of the bootstrap distribution is with Monte-Carlo approximation methods. The algorithm is as follows:

1. Generate a bootstrap sample $\mathbf{X}_n^*(1) = (X_{11}^*, X_{12}^*, \ldots, X_{1n}^*)$, from the EDF, F_n.
2. From this sample calculate $\hat{\theta}^*(1) = T_n(\mathbf{X}_n^*(1); F_n)$.
3. Independently repeat steps 1 and 2 a large number of times, (say B), to obtain bootstrap samples $\mathbf{X}_n^*(1), \mathbf{X}_n^*(2), \ldots, \mathbf{X}_n^*(B)$ and the statistics $\hat{\theta}^*(1), \hat{\theta}^*(2), \ldots, \hat{\theta}^*(B)$.

4. The distribution of $T_n(\mathbf{X}_n^*; F_n)$ is then estimated by the empirical distribution of $\hat{\theta}^*(i),\ i = 1, 2, \ldots, B$.

2.3 The plug-in principle

Consider a parameter of the form $\theta = t(F)$, where t is some functional of the unknown population distribution function F. The plug-in principle asserts that the bootstrap estimator of the parameter θ is:

$$\hat{\theta} = t(F_n), \tag{2.1}$$

with F_n the EDF.

Let X denote any of the random variables in $\mathbf{X}_n$. It follows that

$$\mathrm{E}_F\left[t\left(X\right)\right] = \int_{-\infty}^{\infty} t(x) dF(x).$$

According to the plug-in principle, the bootstrap expected value of $t\left(X^*\right)$ can be calculated by

$$\begin{aligned} \mathrm{E}_{F_n}\left[t\left(X^*\right)\right] = \mathrm{E}_*\left[t\left(X^*\right)\right] &= \int_{-\infty}^{\infty} t(x) dF_n(x) \\ &= \frac{1}{n}\sum_{i=1}^{n} t\left(x_i\right), \end{aligned}$$

where x_i is the observed value of X_i and $\frac{1}{n}$ implies the probability under the empirical distribution function

$$P^*\left(X_j^* = X_i\right) = \frac{1}{n},\ \forall\ i,\ j = 1, 2, ..., n.$$

2.4 The bootstrap estimate of standard error

Consider an unknown parameter θ with corresponding estimate $\hat{\theta}_n = \hat{\theta}_n(X_1, X_2, \ldots, X_n)$. The idea is then to obtain an estimate of the standard deviation of $\hat{\theta}_n$ using the bootstrap. Denote by $\sigma(F)$ the standard deviation of $\hat{\theta}_n$. Now $\sigma(F) = \sqrt{\mathrm{Var}_F\,\hat{\theta}_n}$, and by using the plug-in principle it follows that an estimate of $\sigma(F)$ is

$$\begin{aligned} \hat{\sigma}_n &= \sigma(F_n) \\ &= \sqrt{\mathrm{Var}_{F_n}\left(\hat{\theta}_n^*\right)} \\ &= \sqrt{\mathrm{Var}_*\left(\hat{\theta}_n^*\right)}. \end{aligned}$$

Note that $\hat{\theta}_n^* = \hat{\theta}_n\left(X_1^*, X_2^*, \ldots, X_n^*\right)$ is calculated using a bootstrap sample. A Monte-Carlo algorithm for approximating the standard error proceeds as follows:

1. Generate a sample $(X_1^*, X_2^*, \ldots, X_n^*)$ with replacement from F_n.
2. Calculate $\hat{\theta}^*(1) = \hat{\theta}_n(X_1^*, X_2^*, \ldots, X_n^*)$.
3. Repeat steps 1 and 2 B times to obtain $\hat{\theta}^*(1), \hat{\theta}^*(2), \ldots, \hat{\theta}^*(B)$.
4. Calculate

$$\hat{\sigma}_B = \sqrt{\frac{1}{B-1}\sum_{j=1}^{B}\left\{\hat{\theta}^*(j) - \hat{\theta}_{n\cdot}^*\right\}^2},$$

where

$$\hat{\theta}_{n\cdot}^* = \frac{1}{B}\sum_{j=1}^{B}\hat{\theta}^*(j).$$

It follows that $\hat{\sigma}_B \longrightarrow \hat{\sigma}_n$ as $B \longrightarrow \infty$. According to Efron and Tibshirani (1993) a value of B between 50 and 200 is usually sufficient.

2.5 The bootstrap estimate of bias

Let θ be a parameter with corresponding estimate $\hat{\theta}_n = \hat{\theta}_n(\mathbf{X}_n)$. The bias of the estimator is given by $\beta(F) = \mathrm{E}_F\left[\hat{\theta}_n(\mathbf{X}_n)\right] - \theta$. Using the plug-in principle, the bias can be estimated with the bootstrap. The bootstrap estimate of $\beta(F)$ is

$$\begin{aligned}\hat{\beta}_n &= \beta(F_n)\\ &= \mathrm{E}_{F_n}\left[\hat{\theta}_n(\mathbf{X}_n^*)\right] - \hat{\theta}_n\\ &= \mathrm{E}_*(\hat{\theta}_n^*) - \hat{\theta}_n,\end{aligned}$$

where $\hat{\theta}_n^* = \hat{\theta}_n(\mathbf{X}_n^*) = \hat{\theta}_n(X_1^*, X_2^*, \ldots, X_n^*)$. The Monte Carlo algorithm for approximating the bootstrap estimate of bias follows:

1. Generate a sample (with replacement) $X_1^*, X_2^*, \ldots, X_n^*$ from F_n.
2. Calculate $\hat{\theta}^*(1) = \hat{\theta}_n(X_1^*, X_2^*, \ldots, X_n^*)$.
3. Repeat steps 1 and 2 B times to obtain $\hat{\theta}^*(1), \hat{\theta}^*(2), \ldots, \hat{\theta}^*(B)$.
4. Calculate

$$\hat{\beta}_B = \frac{1}{B}\sum_{j=1}^{B}\hat{\theta}^*(j) - \hat{\theta}_n.$$

2.6 Properties of bootstrap kernel distribution function estimators

Let $\mathbf{X}_n = (X_1, X_2, \ldots, X_n)$ be a sample of i.i.d. random variables from the unknown population distribution F. As before let $\mathbf{X}_n^* = (X_1^*, X_2^*, \ldots, X_n^*)$ denote a bootstrap sample from the EDF, F_n. Considering the bootstrap kernel distribution function estimator, we will now derive the following results:

$$\hat{F}_{n,h}^*(x) = \frac{1}{n}\sum_{i=1}^{n} K\left(\frac{x - X_i^*}{h}\right).$$

It follows from Section 1.1 that

$$\begin{aligned} \mathrm{E}_F\left[F_{n,h}(x)\right] &= \mathrm{E}\left[\frac{1}{n}\sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right)\right] \\ &= \mathrm{E}\left[K\left(\frac{x - X}{h}\right)\right] \\ &= \int_{-\infty}^{\infty} K\left(\frac{x - y}{h}\right) dF(y). \end{aligned}$$

Using the plug-in principle, it follows that

$$\begin{aligned} \mathrm{E}_{F_n}\left[\hat{F}_{n,h}^*(x)\right] &= \mathrm{E}_*\left[\hat{F}_{n,h}^*(x)\right] \\ &= \int_{-\infty}^{\infty} K\left(\frac{x - y}{h}\right) dF_n(y) \\ &= \frac{1}{n}\sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right) \\ &= F_{n,h}(x). \end{aligned}$$

The bias of the ordinary kernel distribution function estimator, $F_{n,h}(x)$, may be expressed as

$$\beta_{n,h}(x, F) = \mathrm{E}_F\left[F_{n,h}(x)\right] - F(x).$$

A bootstrap estimator of the bias is then, using the plug-in principle

$$\begin{aligned} b_{n,h}(x, F_n) &= \mathrm{E}_{F_n}\left[F_{n,h}^*(x)\right] - F_n(x) \\ &= F_{n,h}(x) - F_n(x). \end{aligned}$$

Furthermore, it was shown in Section 1.1 that

$$\begin{aligned}
\mathrm{Var}_F\left[F_{n,h}(x)\right] &= \frac{1}{n}\mathrm{Var}\left[K\left(\frac{x-X}{h}\right)\right] \\
&= \frac{1}{n}\left\{\mathrm{E}\left[K^2\left(\frac{x-X}{h}\right)\right] - \left[\mathrm{E}\left[K\left(\frac{x-X}{h}\right)\right]\right]^2\right\} \\
&= \frac{1}{n}\int_{-\infty}^{\infty} K^2\left(\frac{x-y}{h}\right) dF(y) - \frac{1}{n}\left[\int_{-\infty}^{\infty} K\left(\frac{x-y}{h}\right) dF(y)\right]^2.
\end{aligned}$$

Using the plug-in principle once more, it follows that

$$\begin{aligned}
\mathrm{Var}_{F_n}\left[F^*_{n,h}(x)\right] &= \mathrm{Var}_*\left[F^*_{n,h}(x)\right] \\
&= \frac{1}{n^2}\sum_{i=1}^{n} K^2\left(\frac{x-X_i}{h}\right) - \frac{1}{n}F^2_{n,h}(x).
\end{aligned}$$

If $\mathrm{E}_*\left[\left\{F^*_{n,h}(x) - F_n(x)\right\}^2\right]$ denotes the bootstrap version of the mean squared error of the estimator $F_{n,h}(x)$, then

$$\mathrm{E}\left[\left\{F^*_{n,h}(x) - F_n(x)\right\}^2\right] = \frac{1}{n^2}\sum_{i=1}^{n} K^2\left(\frac{x-X_i}{h}\right) - \frac{1}{n}F^2_{n,h}(x) + \left[F_{n,h}(x) - F_n(x)\right]^2.$$

Remark

It can be shown that the bootstrap bias of the kernel density function estimator is zero. The reason for this is that there is no proxy available for the population density function. The proxy for F is the EDF.

Chapter 3

Bandwidth selection procedures

3.1 Introduction

As opposed to non-parametric distribution function estimation, there is a large volume of literature available on the choice of the bandwidth in non-parametric density function estimation. It is generally accepted that an optimal choice of the bandwidth in density function estimation will not necessarily be optimal for distribution function estimation. In the latter case the existing procedures can be divided in two groups: procedures based on estimating the asymptotical bandwidth and procedures that estimate an appropriate discrepancy measure.

3.2 Procedures based on estimating the mean integrated square error (MISE)

3.2.1 The procedure of Van Graan (1983)

In this procedure a bootstrap estimator for the mean integrated squared error is proposed. Consider the measure defined in (1.11), i.e.,

$$\begin{aligned} J(F,h) &= \text{MISE}(h) \\ &= \text{E}\left[\int_{-\infty}^{\infty}\{F_{n,h}(x)-F(x)\}^2\,dF(x)\right], \end{aligned} \tag{3.1}$$

where the notation $J(F,h)$ indicates that the expression on the right hand side is dependent on the unknown distribution function F. In order to obtain a choice of the bandwidth

parameter, the expression in (3.1) can be minimised with respect to h. Unfortunately this cannot be done since F is unknown. One possible solution is to estimate $J(F,h)$ by a bootstrap estimator. Using the plug-in-estimate described in (2.1) and the i.i.d. assumption, it follows that an estimate (see also (1.12)) for $J(F,h)$ is given by

$$\begin{aligned}
J(F_n,h) &= \frac{1}{n}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left[K(\frac{x-y}{h})-F_n(x)\right]^2 dF_n(y)dF_n(x) \\
&\quad +\frac{n-1}{n}\int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty}\left\{K(\frac{x-y}{h})-F_n(x)\right\}dF_n(y)\right]^2 dF_n(x) \\
&= \frac{1}{n^3}\sum_{j=1}^{n}\sum_{i=1}^{n}\left[K\left(\frac{X_j-X_i}{h}\right)-F_n\left(X_j\right)\right]^2 \\
&\quad +\frac{n-1}{n^4}\sum_{j=1}^{n}\left[\sum_{i=1}^{n}\left\{K\left(\frac{X_j-X_i}{h}\right)-F_n\left(X_j\right)\right\}\right]^2 \\
&= \frac{1}{n^3}\sum_{j=1}^{n}\sum_{i=1}^{n}\left[K(\frac{X_{(j)}-X_{(i)}}{h})-\frac{j}{n}\right]^2 \\
&\quad +\frac{n-1}{n^4}\sum_{j=1}^{n}\left[\sum_{i=1}^{n}\left\{K(\frac{X_{(j)}-X_{(i)}}{h})-\frac{j}{n}\right\}\right]^2, \qquad (3.2)
\end{aligned}$$

where $X_{(1)}, X_{(2)}, ..., X_{(n)}$ denotes the order statistics and

$$F_n\left(X_{(j)}\right)=\frac{j}{n}. \qquad (3.3)$$

If the kernel function K is specified, then $J(F_n,h)$ can be evaluated as a function of h. It is then possible to determine that value of h which minimises $J(F_n,h)$.

Remarks

1. The properties of the procedure described above were not investigated in an empirical study

2. The estimator proposed in (3.2) was actually considered as a first step to determine the bandwidth. In a next step the idea is to estimate $J(F,h)$ by $J(F_{n,\hat{g}},h)$, where $\hat{g}$ is the bandwidth determined in the first step. In this way a recursive procedure is obtained for estimating the bandwidth. This procedure will not be considered in this study. The interested reader is referred to Van Graan (1983)

3. The estimator in (3.2) will be included in the empirical study of Chapter 5.

3.2.2 The procedure of Sarda (1993)

Sarda (1993) considered estimating the mean integrated squared error defined in (1.12), i.e.,

$$\mathrm{MISE}(h) = \mathrm{E}\left[\int_{-\infty}^{\infty} \{F_{n,h}(x) - F(x)\}^2 W(x) dF(x)\right].$$

For the discussion below the following conditions are imposed:

1. W is bounded and supported on a compact set.

2. The set of bandwidths considered is

$$H_n = [\gamma_1 n^{-a}, \gamma_2 n^{-b}], \quad \frac{1}{4} < b \leq a < \frac{1}{2},$$

   where γ_1 and γ_2 are constants.

3. The function K is absolutely continuous and satisfies

$$\lim_{x \to -\infty} K(x) = 0 \quad \text{and} \quad \lim_{x \to \infty} K(x) = 1.$$

4. For $k = K'$,

$$\int_{-\infty}^{\infty} xk(x)dx = 0 \text{ and } \int_{-\infty}^{\infty} x^2 k(x)dx < \infty.$$

5. The function F is twice differentiable and F and $|f'|$ are bounded from below on the support of W.

Reiss (1981) and Lejuene and Sarda (1992) showed that

$$\mathrm{MISE}(h) = V_1(F)n^{-1} - V_2 h n^{-1} + B_3 h^4 + Ch^2/n + \text{smaller order terms}, \tag{3.4}$$

where $V_1(F)$ and

$$V_2 = 2D_1(K)A_0(F), \tag{3.5}$$

are the variance terms and

$$B_3 = \frac{1}{4}\mu_2^2(k)\Psi_{1,1}(F), \tag{3.6}$$

is the squared bias term. Also, $C = D_2(K)A_1(F) - E_1(F)\mu_2(k)$ and

$$V_1(F) = \int_{-\infty}^{\infty} F(x)\left[1 - F(x)\right] W(x) dF(x), \tag{3.7}$$

$$\Psi_{1,1}(F) = \int_{-\infty}^{\infty} \left[f'(x)\right]^2 W(x) dF(x), \tag{3.8}$$

$$A_j(F) = \int_{-\infty}^{\infty} f^{(j)}(x) W(x) dF(x) \text{ for } j \geq 0, \tag{3.9}$$

$$E_j(F) = \int_{-\infty}^{\infty} F(x) f^{(j)}(x) dF(x) \text{ for } j \geq 1 \text{ and}, \tag{3.10}$$

$\mu_j(k)$ and $D_j(K)$ were defined in (1.5) and (1.7).

Sarda noted that the balance between the variance and bias term occur at the "theoretical optimal bandwidth", i.e., the bandwidth minimising the two central terms in (3.4), namely

$$h_{\text{opt}} = \left[\frac{V_2}{4B_3}\right]^{1/3} n^{-1/3}. \tag{3.11}$$

This bandwidth cannot be computed since it depends on the true distribution function F.

As mentioned in Section 1.2, Sarda proposed a discrete approximation (see (1.15)) to the mean integrated squared error, i.e.,

$$\text{ASE}(h) = \frac{1}{n}\sum_{i=1}^{n}\left[F_{n,h}\left(X_i\right) - F(X_i)\right]^2 W(X_i).$$

In order to estimate ASE(h), the unknown distribution function, $F\left(X_i\right)$, can be replaced by the EDF, giving the so-called "leave-none-out" estimator of ASE(h),

$$\text{LNO}(h) = \frac{1}{n}\sum_{i=1}^{n}\left[F_{n,h}\left(X_i\right) - F_n(X_i)\right]^2 W(X_i). \tag{3.12}$$

Sarda argued that this measure will produce a very small bandwidth. To overcome this problem, he introduced a so-called "cross-validation" criterion,

$$\text{CV}(h) = \frac{1}{n}\sum_{i=1}^{n}\left[F_{n,h:-i}(X_i) - F_n(X_i)\right]^2 W(X_i), \tag{3.13}$$

where $F_{n,h:-i}$ is the kernel estimator computed by leaving out X_i. The bandwidth minimising the criterion is then selected. For an account of cross-validation procedures in kernel density estimation, the interested reader is referred to Bowman (1984) and Marron (1987).

The main result of Sarda (1993) concerns the asymptotic optimality of the bandwidth derived from (3.13). Define $d(F_{n,h}, F)$ as one of the measures of accuracy MISE(h), ASE(h) or ISE(h), where $\text{ISE}(h) = \int_{-\infty}^{\infty}\left[F_{n,h}(x) - F(x)\right]^2 W(x)dF(x)$. It should be noted that the term $V_1(F)n^{-1}$ in (3.4) does not depend on h. Now define $d'(F_{n,h}, F) = d(F_{n,h}, F) - V_1(F)n^{-1}$ which will yield the same minimising bandwidth h as $d(F_{n,h}, F)$. Using these definitions Sarda proved that the bandwidth $\hat{h}$ that minimises (3.13) is asymptotically optimal in the sense that

$$\lim_{n\to\infty}\frac{d'(F_{n,\hat{h}}, F)}{\inf_{h\in H_n} d'(F_{n,h}, F)} = 1 \quad \text{a.s.},$$

i.e., the bandwidth minimising (3.13) yields a discrepancy that is equivalent to the smallest discrepancy when the set of bandwidths H_n is considered.

Remarks

1. Altman and Léger (1995) showed that under conditions 1 to 5 that the leave-none-out criterion, LNO(h), and the cross-validation criterion, CV(h), are asymptotically equivalent over the set of bandwidths H_n considered.

2. Assuming that:

   - F has five derivatives and that the fifth derivative is bounded,
   - the kernel K has a symmetric density with respect to 0 and its fifth moment exist and,
   - the weight function has a finite first moment with respect to F,

   Altman and Léger (1995) proved that the expected values of the leave-none-out criterion, LNO(h), and the cross-validation criterion, CV(h), are asymptotically identical up to smaller order terms.

3. Using the assumptions that the kernel k is symmetric about 0 and that the density has a continuous first derivative on the support of the function W, it was shown by Altman and Léger (1995) that the expected value of the derivative of LNO(h) is positive when $n \to \infty$ and $nh^2 \to \infty$, making LNO(h) a strictly increasing function of h.

4. Using the results mentioned in remarks 1, 2 and 3 the conclusion can be made that both the leave-none-out criterion LNO(h) and the cross-validation criterion CV(h) are likely to perform poorly. This conclusion is in sharp contradiction to the optimality result for cross-validation derived by Sarda (1993). This result was supported by the numerical results of Altman and Léger (1995).

3.2.3 The procedure of Bowman, Hall and Prvan (1998)

Bowman, Hall and Prvan (1998) proposed a procedure for obtaining the bandwidth, which is derived more directly from the general paradigm of cross-validation. In the case of non-parametric kernel density estimation, the method of cross-validation can be viewed as representing each observation by a Dirac delta function $\delta(x - X_i)$ whose expectation is $f(x)$, and contrasting this with a density estimate based on the remainder of the data. In the current

context of non-parametric kernel distribution function estimation, a natural characterisation of each observation is by the indicator function $I(x - X_i)$ defined by

$$I(x - X_i) = \begin{cases} 1, & x - X_i > 0 \\ 0, & \text{otherwise.} \end{cases}$$

It follows that $\mathrm{E}\left[I(x - X_i)\right] = F(x)$, as shown in Section 1. Bowman et al. (1998) considered the discrepancy measure defined in (1.10), i.e.,

$$\mathrm{MISE}(h) = \mathrm{E}\left[\int_{-\infty}^{\infty} \left\{F_{n,h}(x) - F(x)\right\}^2 dx\right].$$

A cross-validation function based on a discrete approximation of $\mathrm{MISE}(h)$ is then defined as

$$\mathrm{CV}(h) = \frac{1}{n}\sum_{i=1}^{n} \int_{-\infty}^{\infty} \left[I(x - x_i) - F_{n,h:-i}(x)\right]^2 dx, \tag{3.14}$$

where $F_{n,h:-i}$ is the kernel estimator constructed from the data with observation X_i omitted. The bandwidth minimising this criterion is then selected.

To demonstrate heuristically why this procedure can have attractive properties, consider the expression

$$\mathrm{H}(h) = \mathrm{CV}(h) - \frac{1}{n}\sum_{i=1}^{n} \int_{-\infty}^{\infty} \left[I(x - x_i) - F(x)\right]^2 dx. \tag{3.15}$$

Note that the second term subtracted on the right hand side characterises the performance of the true F when compared to the expression in (3.14). This term does not involve h and so the cross-validation procedure is unaffected. Taking the expected value of $H(h)$, we have

$$\begin{aligned}
&\mathrm{E}\left[H(h)\right] \\
&= \mathrm{E}\left[\frac{1}{n}\sum_{i=1}^{n} \int_{-\infty}^{\infty} \left\{\left[I(x - X_i) - F_{n,h:-i}(x)\right]^2 - \left[I(x - X_i) - F(x)\right]^2\right\} dx\right] \\
&= \mathrm{E}\left[\frac{1}{n}\sum_{i=1}^{n} \int_{-\infty}^{\infty} \left\{F^2_{n,h:-i}(x) - 2I(x - X_i)F_{n,h:-i}(x) + 2I(x - X_i)F_{n,h:-i}(x) - F^2(x)\right\}\right] \\
&= \frac{1}{n}\sum_{i=1}^{n} \int_{-\infty}^{\infty} \left[\mathrm{E}\left\{F^2_{n,h:-i}(x)\right\} - 2F(x)\,\mathrm{E}\left\{F_{n,h:-i}(x)\right\} + 2F^2(x) - F^2(x)\right] dx \\
&= \mathrm{E}\left[\int_{-\infty}^{\infty} \left\{F_{n-1,h}(x) - F(x)\right\}^2 dx\right],
\end{aligned} \tag{3.16}$$

where $F_{n-1,h}(x)$ denotes a kernel estimator based on a sample size of $n - 1$. This shows that the cross-validation function $\mathrm{CV}(h)$ produces an unbiased estimator of the discrepancy measure in (1.10), for a sample size of $n - 1$ shifted vertically by an unknown constant.

Remarks

1. It can be shown that under mild regularity conditions

$$\text{MISE}(h) = n^{-1}v_1(F) - 2hn^{-1}D_1(K) + b_3h^4 + o(\frac{h}{n} + h^4), \tag{3.17}$$

where

$$v_1(F) = \int_{-\infty}^{\infty} F(x)\left[1 - F(x)\right] dx, \tag{3.18}$$

$$r(f') = \int_{-\infty}^{\infty} \left[f'(x)\right]^2 dx, \tag{3.19}$$

$$b_3 = \frac{1}{4}\mu_2^2(k)r(f'), \tag{3.20}$$

and $D_1(K)$ were defined in (1.7). It follows from (3.16) that $H(h)$ might be a good approximation to MISE(h). However, if $H(h)$ is adjusted by adding the quantity

$$J_n = \int_{-\infty}^{\infty} \left[\{F_{n,h}(x) - F(x)\}^2 - \text{E}\left[\{F_{n,h}(x) - F(x)\}^2\right]\right] dx,$$

which has zero mean and does not depend on h, one of the results in the paper states that a particular good approximation of MISE(h) is obtained.

2. The result in the previous remark means that cross-validation yields a bandwidth that is asymptotically equivalent to the one that produces second-order optimality.

3. In the numerical study the method of Sarda was shown to produce very small values of h, indicating that the resulting estimator is essentially the EDF. The method of Altman and Léger (1995) (which is to be discussed in the next section) and cross-validation behaved satisfactorily in the study.

4. Since the method of cross-validation is based on the unweighted criterion (1.10) it will not be included in the numerical study of Chapter 5.

3.3 Procedures based on estimating the asymptotical optimal bandwidth

3.3.1 The procedure of Altman and Léger (1995)

Using the assumptions introduced by Sarda (1993) in Section 3.2.2 and the expression for MISE(h) in (3.4) we have

$$\text{MISE}\,(h) = V_1(F)n^{-1} - V_2hn^{-1} + B_3h^4 + Ch^2/n + \text{smaller order terms},$$

where $C > 0$.

The bandwidth that minimises MISE(h) is the weighted asymptotic mean integrated squared error (WAMISE) optimal bandwidth h_{opt} from (3.11) where

$$h_{\text{opt}} = \left[\frac{V_2}{4B_3}\right]^{1/3} n^{-1/3},$$

using the same notation as in Section 3.2.2.

To obtain a plug-in estimate for h_{opt}, the variance term V_2 and the bias term B_3 in (3.11) must be estimated first. Altman and Léger used the kernel estimator for V_2 as suggested by Hall and Marron (1987), namely:

$$\begin{aligned} \hat{V}_2 &= 2D_1(K)\hat{A}_0(F) \\ &= 2D_1(K)\left[\frac{1}{n^2 g_1}\sum_{i\neq j} w_1\left(\frac{X_i - X_j}{g_1}\right) W(X_i)\right], \end{aligned} \tag{3.21}$$

where w_1 is a kernel function with bandwidth g_1. Hall and Marron (1987) proved that the estimator in (3.21) is consistent when g_1 satisfies $ng_1 \to \infty$ and $n^{\frac{1}{4}} g_1 \to 0$. The estimator of $\Psi_{1,1}(F)$ is :

$$\hat{\Psi}_{1,1}(F) = \frac{1}{n^3 {g_2}^4}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n} w_2'\left(\frac{X_i - X_j}{g_2}\right) w_2'\left(\frac{X_i - X_k}{g_2}\right) W(X_i), \tag{3.22}$$

where w_2' is the derivative of a kernel function w_2 and g_2 is the associated bandwidth. These estimators are then used to calculate a plug-in estimate, $\hat{h}_{\text{AL}}$, of the asymptotical optimal bandwidth

$$\hat{h}_{\text{AL}} = \left[\frac{\hat{V}_2}{4\hat{B}_3}\right]^{\frac{1}{3}} n^{-\frac{1}{3}}, \tag{3.23}$$

where $\hat{B}_3 = \frac{1}{4}\mu_2^2(k)\hat{\Psi}_{1,1}(F)$.

Remarks

1. The procedure proposed by Altman and Léger (1995) required two bandwidths to obtain the estimator for the asymptotical optimal bandwidth. In their simulation study a pilot bandwidth of the form $g_1 = g_2 = n^{-0.3}$ was used. Koekemoer (2004, Chapter 2) has proposed a data driven approach for estimating these bandwidths, which will be presented in the next section.

2. The simulation study of Altman and Léger (1995) included both the leave-none-out and cross-validation criteria proposed by Sarda (1993). The results obtained by Altman and Léger (1995) were similar to those obtained by Bowman et al. (1998), discussed in Section 3.2.3. It was found that for larger sample sizes both these criteria consistently choose the smallest possible bandwidth as the optimal bandwidth, resulting in the EDF as an estimator.

3. The proposed plug-in estimator performed satisfactorily in the simulation study for the sample sizes studied.

3.3.2 The improved procedure of Altman and Léger (2004)

In the previous section the quantity B_3 in the expression for h_{opt}, was estimated by $\hat{B}_3 = \frac{1}{4}\mu_2^2(k)\hat{\Psi}_{1,1}(F)$, where

$$\hat{\Psi}_{1,1}(F) = \frac{1}{n^3 {g_2}^4} \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n} w_2'\left(\frac{X_i - X_j}{g_2}\right) w_2'\left(\frac{X_i - X_k}{g_2}\right) W(X_i).$$

Altman and Léger (1995) calculated the asymptotic mean squared error of $\hat{\Psi}_{1,1}(F)$ under the assumptions:

1. The kernel function w_2 has mean 0, finite variance and $w_2'(0) = 0$.

2. The density f has a bounded fourth derivative.

3. The bandwidth g_2 is a non-random sequence of positive numbers. Also assume that g_2 satisfies

$$\lim_{n\to\infty} g_2 = 0 \quad \text{and} \quad \lim_{n\to\infty} n g_2 = \infty.$$

   Hence g_2 is of the form $g_2 = cn^{-t}$ where $0 < t < 1$ and c is a constant.

Then the asymptotic mean squared error (AMSE) of $\hat{\Psi}_{1,1}(F)$ is given by

$$\begin{aligned}\text{AMSE}\left[\hat{\Psi}_{1,1}(F)\right] &= \frac{2C(w_2)}{n^2 {g_2}^5}\int_{-\infty}^{\infty}[f(x)]^4 W(x)\,dx \\ &\quad + \left[g_2^2\mu_2(w_2)\int_{-\infty}^{\infty} f'(x) f'''(x) f(x) W(x)\,dx \right. \\ &\quad \left. + \frac{r(w_2')}{n^2 {g_2}^3}\int_{-\infty}^{\infty}[f(x)]^2 W(x)\,dx\right]^2,\end{aligned}$$

where

$$C(w_2) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} w_2'(s)\, w_2'(t)\, w_2'(u)\, w_2'(t+u-s)\,dsdtdu,$$

and $r(f')$ was defined in (3.19). Minimising the AMSE with respect to g_2 gives the asymptotically optimal bandwidth

$$g_2 = \left[\frac{5C(w_2) \int_{-\infty}^{\infty} \{f(x)\}^4 W(x) dx}{2\mu_2^2(w_2) \Psi_{1,3}^2(F)} \right]^{1/9} n^{-2/9}, \tag{3.24}$$

where

$$\Psi_{m,p}(F) = \int_{-\infty}^{\infty} f^{(m)}(x) f^{(p)}(x) f(x) W(x) dx.$$

The following discussion is due to Koekemoer (2004) and will be repeated here for the sake of completeness. If $W(x) = 1$, a number of simplifications and improvements are possible:

- Define $R\left(f^{(\frac{m}{2})}\right) = \psi_m = \int_{-\infty}^{\infty} f^{(m)}(x) f(x) dx = \mathrm{E}\left[f^{(m)}(X)\right]$, $m = 2s$ and $s = 0, 1, 2, \ldots$. It follows that $A_0(F) = \int_{-\infty}^{\infty} [f(x)]^2 dx = \psi_0$ can be estimated by the l-stage estimator suggested by Sheather and Jones (1991), given by

$$\hat{\psi}_0(g_1) = \frac{1}{n^2 g_1} \sum_{i=1}^{n} \sum_{j=1}^{n} w_1 \left(\frac{X_i - X_j}{g_1} \right).$$

- The unknown quantity $\Psi_{1,1}(F) = \int_{-\infty}^{\infty} [f'(x)]^2 f(x) dx$ may also be estimated using a l-stage estimator, with a normal reference utilised at stage l. A general estimator for the quantity $\Psi_{m,p}(F) = \int_{-\infty}^{\infty} f^{(m)}(x) f^{(p)}(x) f(x) dx$ is given by

$$\hat{\Psi}_{m,p}(F) = \frac{1}{n^3 g_2^{m+1} g_3^{p+1}} \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{n} w_2^{(m)} \left(\frac{X_i - X_j}{g_2} \right) w_3^{(p)} \left(\frac{X_i - X_k}{g_3} \right),$$

  where $w_2(\cdot)$ and $w_3(\cdot)$ are possibly different kernel functions with different associated bandwidths g_2 and g_3 respectively.

Now the procedure for $l = 1$, $w_1(\cdot) = w_2(\cdot) = w_3(\cdot) = \phi(\cdot)$, $K(\cdot) = \Phi(\cdot)$, $W(x) = 1$ and a normal reference for the unknown density f will be described:

Step 1 Define the $N(\mu, \sigma^2)$ normal density by

$$\phi_\sigma(x - \mu) = \frac{1}{\sigma} \phi \left(\frac{x - \mu}{\sigma} \right) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2},$$

and let $\phi_\sigma^{(m)}$ denote the m^{th} derivative. Now

$$\psi_m = \int_{-\infty}^{\infty} \phi_\sigma^{(m)}(x - \mu) \phi_\sigma(x - \mu) dx = \frac{(-1)^{\frac{m}{2}} m!}{(2\sigma)^{m+1} (m/2)! \sqrt{\pi}}.$$

Estimate ψ_2 with

$$\hat{\psi}_2 = -\frac{1}{4\hat{\sigma}^3\sqrt{\pi}}. \tag{3.25}$$

where $\hat{\sigma}$ is defined as the robust scale estimator

$$\hat{\sigma} = \min\left\{S, \frac{\hat{q}_3 - \hat{q}_1}{\Phi^{-1}\left(\frac{3}{4}\right) - \Phi^{-1}\left(\frac{1}{4}\right)}\right\}, \tag{3.26}$$

see Silverman (1986, p. 47), S is the sample standard deviation, q_1 and q_3 are the first and third sample quartiles and $\Phi(x)$ the standard normal distribution function.

Step 2 Use $\hat{\psi}_2$ to estimate $\psi_0 = A_0(F)$ with the estimator

$$\hat{\psi}_0 = \hat{A}_0(F) = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\phi_{\hat{g}_1}(X_i - X_j),$$

where

$$\hat{g}_1 = \left[\frac{-2\phi(0)}{\mu_2(\phi)\hat{\psi}_2}\right]^{\frac{1}{3}} n^{-\frac{1}{3}}, \tag{3.27}$$

and where

$$\mu_2(\phi) = 1 \quad \text{and} \quad \phi(0) = \frac{1}{\sqrt{2\pi}}. \tag{3.28}$$

See also remark 3 at the end of the section. Substituting (3.25) and (3.28) into (3.27) we obtain

$$\hat{g}_1 = \left[4\sqrt{2}\hat{\sigma}^3\right]^{\frac{1}{3}} n^{-\frac{1}{3}}.$$

Step 3 Calculating (3.24) using a normal reference gives

$$\hat{g}_2 = 0.9743\hat{\sigma}n^{-2/9}.$$

Now, estimate $\Psi_{1,1}(F) = \int_{-\infty}^{\infty}[f'(x)]^2 f(x)dx$ with

$$\hat{\Psi}_{1,1}(F) = \frac{1}{n^3\hat{g}_2^4}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\phi'\left(\frac{X_i - X_j}{\hat{g}_2}\right)\phi'\left(\frac{X_i - X_k}{\hat{g}_2}\right). \tag{3.29}$$

Step 4 Use $\hat{A}_0(F)$ and $\hat{\Psi}_{1,1}(F)$ to estimate the WAMISE optimal bandwidth (direct plug-in), thus:

$$\hat{h}_{DPI,1} = \left[\frac{2D_1(K)\hat{A}_0(F)}{\mu_2^2(\phi)\hat{\Psi}_{1,1}(F)}\right]^{1/3} n^{-1/3}, \tag{3.30}$$

where $\mu_2(\phi) = 1$ and $D_1(K) = \frac{1}{2\sqrt{\pi}}$.

Remarks

1. Note that the only difference between this procedure and the procedure originally proposed by Altman and Léger (1995) is that the choices of g_1 and g_2 are now no longer arbitrary.

2. The estimator in (3.30) will be included in the empirical study of Chapter 5.

3. Note that the estimated value of $\hat{g}_2$ in (3.29) is obtained from direct application of the expression

$$g_{\text{AMSE},m} = \left[\frac{-r!w^{(m)}(0)}{\mu_r(w)\psi_{m+r}}\right]^{1/(m+r+1)} n^{-1/(m+r+1)},$$

which represent the asymptotic mean squared error (AMSE) optimal bandwidth, derived from the estimator

$$\hat{\psi}_m(g) = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} w_g^{(m)}(X_i - X_j),$$

of ψ_m. Here r denotes the order of the kernel used.

3.3.3 The procedure of Polansky (1997)

Using the assumptions that

- $h = h_n$ is a sequence of bandwidths such that $h \to 0$ and $nh \to \infty$ as $n \to \infty$ and
- f is continuous and differentiable with finite mean and has a square integrable derivative,

Jones (1990) showed that the mean integrated squared error in (1.10) can be written as

$$\text{MISE}(h) = v_1(F)n^{-1} - 2D_1(K)hn^{-1} + \tfrac{1}{4}h^4\mu_2^2(k)r(f') + o\left(\tfrac{h}{n} + h^4\right), \qquad (3.31)$$

where $v_1(F)$, $D_1(K)$ and $r(f')$ were defined in (3.18), (1.7) and (3.19) respectively.

Minimising (3.31) with respect to h gives the optimal value as

$$h_0 = \left[\frac{2D_1(K)}{n\mu_2^2(k)r(f')}\right]^{1/3}. \qquad (3.32)$$

The value of MISE(h) when using $h = h_0$ is

$$\text{MISE}(h_0) = v_1(F)n^{-1} - \frac{3}{4}\left[\frac{2D_1(K)}{n}\right]^{4/3}\left\{\mu_2^2(k)r(f')\right\}^{-1/3} + o\left(n^{-4/3}\right), \qquad (3.33)$$

so that $\text{MISE}(h_0) = O(n^{-1})$, which is the same rate achieved by the empirical distribution function, thus smoothing has only a second-order effect on the MISE of distribution functions. However, as was noted in Section 1.1 there exist applications where a smoothed estimate of the distribution function is required.

One should note that in (3.32), h_0 still depends on the unknown density function f. This can be solved by assuming that f follows a certain parametric form and using the corresponding bandwidth. This is also known as a reference bandwidth. However it can be argued that if f is known, or even if the form of f is known that it will not be necessary to estimate the distribution function. For example if f is assumed normal with mean μ and variance σ^2 then

$$r(f') = \frac{1}{4\sigma^3\pi^{1/2}},$$

and the corresponding bandwidth is estimated by

$$\hat{h}_N = \left[\frac{8\hat{\sigma}^3 D_1(K)\pi^{1/2}}{n\mu_2^2(k)}\right],$$

where $\hat{\sigma}$ is defined in (3.26).

It follows from (3.32) that in order to estimate h_0, an estimate for $r(f') = -\psi_2$ (using the results in 3.3.2) must be obtained. From the remarks in section 3.3.2 it is clear that in order to estimate ψ_2 an estimate for ψ_4 will be required if a second order kernel function, such as the standard normal, is used. In general, an estimate of ψ_{m+2} is required for the estimation of ψ_m. The procedure is therefore recursive. Note that if a normal reference for f is used then

$$\hat{\psi}_m^{NR} = \frac{(-1)^{\frac{m}{2}} m!}{(2\hat{\sigma})^{m+1}(m/2)!\sqrt{\pi}}, \tag{3.34}$$

where $\hat{\sigma}$ is defined in (3.26). Note that in general an estimator for ψ_m is given by

$$\hat{\psi}_m(g) = \frac{1}{n^2}\frac{1}{g^{m+1}}\sum_{i=1}^{n}\sum_{j=1}^{n} w^{(m)}\left(\frac{X_i - X_j}{g}\right),$$

where g can be determined from the expression

$$g_{\text{AMSE},m} = \left[\frac{2w^{(m)}(0)}{-n\mu_2(w)\psi_{m+2}}\right]^{1/(m+3)},$$

in the case of a second order kernel function.

Let $w(\cdot) = \phi(\cdot)$ and $K(\cdot) = \Phi(\cdot)$. An algorithm for obtaining a l-stage estimator for h_0 is described below. Let $l > 0$ be an integer:

Step 1 Calculate $\hat{\psi}_{2l+2}^{NR}$ using (3.34),

Step 2 Starting with $j = l$ and iterating until $j = 1$, calculate $\hat{\psi}_{2j}(\hat{g}_{2j})$ where

$$\hat{g}_{2j} = \left[\frac{2\phi^{(2j)}(0)}{-n\mu_2(\phi)\psi_{2j+2}}\right]^{1/(2j+3)},$$

where

$$\hat{\psi}_{2j+2} = \begin{cases} \hat{\psi}_{2l+2}^{NR}, & \text{when } j = l \\ \hat{\psi}_{2j+2}(\hat{g}_{2j+2}), & \text{when } j < l, \end{cases}$$

Step 3 Calculate

$$\hat{h}_b = \left[\frac{2D_1(K)}{-n\mu_2^2(\phi)\hat{\psi}_2(\hat{g}_2)}\right]^{1/3}.$$

Remarks

1. The method proposed by Polansky (1997) is based on the unweighted criterion (1.10) and will not be included in the numerical study of Chapter 5.

2. Making $l > 2$ in the l-stage estimator does not increase the rate of convergence or decrease the variance. See Polansky (1997) for more details.

3. Since the assumption of normality is made at stage l, this procedure works best for data that is normal or that can be transformed to normality. The interested reader is referred to Koekemoer (2004) for more details on transforming the data to normality using kernel distribution function estimation.

4. The procedure of Polansky behaved satisfactorily in the simulation study with the two-stage estimator having the lowest estimated mean integrated squared error in most of the cases.

Chapter 4

A Procedure based on the bootstrap

4.1 Introduction

The discussion in Chapter 3 revealed that there are mainly two approaches available for choosing the bandwidth parameter in distribution function estimation, i.e., estimating a suitable version of a discrepancy measure and estimating the asymptotical optimal bandwidth. In this chapter the focus will be on the first approach.

The discrepancy measure that will be considered is the mean integrated squared error (MISE) defined in (1.10) and the mean average squared error (MASE) defined in (1.16). In Section 4.2 MASE will be calculated in a special case where both the kernel and density functions are uniformly distributed. In Section 4.3 asymptotic theory will be developed which shows that MASE is approximately equal to MISE in an asymptotic sense. In Section 4.4 an estimator for MASE based on bootstrap methodology will be derived.

4.2 Calculation of the MASE discrepancy measure in a special case

In this section the MASE discrepancy measure defined in (1.24) will be calculated for the distribution function F and kernel function K given by

$$F(x) = \begin{cases} 0, & \text{if } x < 0 \\ x, & \text{if } 0 \leq x \leq 1 \\ 1, & \text{if } x > 1, \end{cases} \tag{4.1}$$

and

$$K(x) = \begin{cases} 0, & \text{if } x < -\sqrt{3} \\ \frac{x+\sqrt{3}}{2\sqrt{3}}, & \text{if } -\sqrt{3} \leq x \leq \sqrt{3} \\ 1, & \text{if } x < \sqrt{3}. \end{cases} \tag{4.2}$$

Each of the terms appearing in the expression for MASE(h) in (1.24), i.e. $\mathrm{E}\left[K\left(\frac{X_1-X_2}{h}\right)\right]$, $\mathrm{E}\left[K^2\left(\frac{X_1-X_2}{h}\right)\right]$, $\mathrm{E}\left[K\left(\frac{X_1-X_2}{h}\right)K\left(\frac{X_1-X_3}{h}\right)\right]$ and $\mathrm{E}\left[K\left(\frac{X_1-X_2}{h}\right)F(X_1)\right]$ will be evaluated according to whether (see Figures 4.1, 4.2 and 4.3):

$$\begin{aligned} &\text{Case 1} \quad 0 < \sqrt{3}h < \tfrac{1}{2}, \\ \text{or} \quad &\text{Case 2} \quad \tfrac{1}{2} < \sqrt{3}h < 1, \\ \text{or} \quad &\text{Case 3} \quad \sqrt{3}h > 1, \end{aligned}$$

is applicable. Consider evaluating

$$\begin{aligned} \mathrm{E}\left[K\left(\frac{X_1 - X_2}{h}\right)\right] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} K\left(\frac{x-y}{h}\right) f(x,y)dydx \\ &= \int_0^1 \left\{\int_0^1 K\left(\frac{x-y}{h}\right) dy\right\} dx, \end{aligned}$$

using the independence of X_1 and X_2.

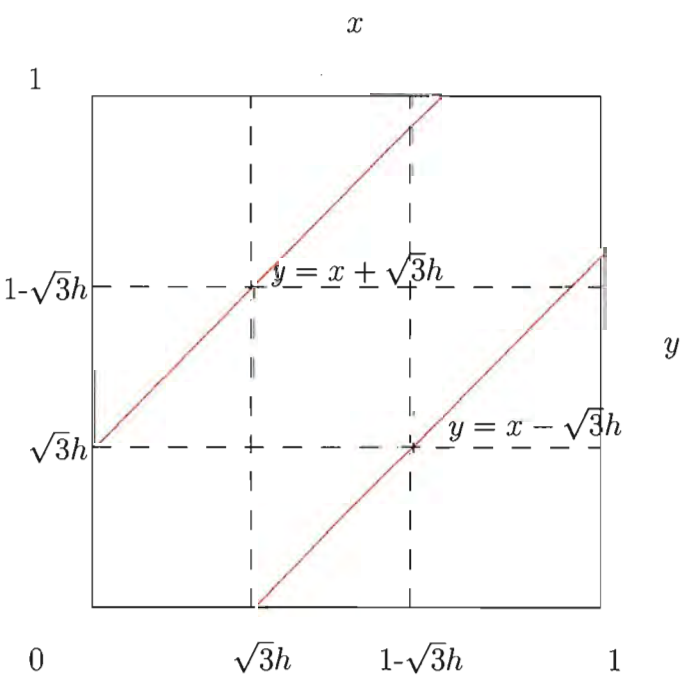


Figure 4.1: Case 1, $\sqrt{3}h < \frac{1}{2}$

Considering Figure 4.1 it follows that for case 1 we have

i. $0 < x < \sqrt{3}h$

$$\begin{aligned}
\mathrm{E}\left[K\left(\frac{X_1 - X_2}{h}\right)\right] &= \int_0^{\sqrt{3}h}\left\{\int_0^{x+\sqrt{3}h}\frac{1}{2\sqrt{3}h}\left[\frac{x-y}{h}+\sqrt{3}\right]dy\right\}dx \\
&= \int_0^{\sqrt{3}h}\left[\frac{x^2}{4\sqrt{3}h}+\frac{x}{2}+\frac{\sqrt{3}h}{4}\right]dx \\
&= \frac{7h^2}{4}.
\end{aligned}$$

ii. $\sqrt{3}h < x < 1 - \sqrt{3}h$

$$\begin{aligned}
\mathrm{E}\left[K\left(\frac{X_1 - X_2}{h}\right)\right] &= \int_{\sqrt{3}h}^{1-\sqrt{3}h}\left\{\int_0^{x-\sqrt{3}h}1dy+\int_{x-\sqrt{3}h}^{x+\sqrt{3}h}\frac{1}{2\sqrt{3}h}\left[\frac{x-y}{h}+\sqrt{3}\right]dy\right\}dx \\
&= \int_{\sqrt{3}h}^{1-\sqrt{3}h}xdx \\
&= \frac{1}{2}-\sqrt{3}h.
\end{aligned}$$

iii. $1 - \sqrt{3}h < x < 1$

$$\begin{aligned}
&\mathrm{E}\left[K\left(\frac{X_1 - X_2}{h}\right)\right] \\
&= \int_{1-\sqrt{3}h}^{1}\left\{\int_0^{x-\sqrt{3}h}1dy+\int_{x-\sqrt{3}h}^{1}\frac{1}{2\sqrt{3}h}\left[\frac{x-y}{h}+\sqrt{3}\right]dy\right\}dx \\
&= \int_{1-\sqrt{3}h}^{1}\left[-\frac{x^2}{4\sqrt{3}h}+\frac{x}{2\sqrt{3}h}+\frac{x}{2}-\frac{1}{4\sqrt{3}h}+\frac{1}{2}-\frac{\sqrt{3}h}{4}\right]dx \\
&= -\frac{7h^2}{4}+\sqrt{3}h.
\end{aligned}$$

Compiling the results it follows that

$$\mathrm{E}\left[K\left(\frac{X_1 - X_2}{h}\right)\right] = \frac{1}{2}. \tag{4.3}$$

For case 2, following the steps outlined in (i), (ii) and (iii) and using Figure 4.2 we have

$$\begin{aligned}
&\mathrm{E}\left[K\left(\frac{X_1 - X_2}{h}\right)\right] \\
&= \int_0^{1-\sqrt{3}h}\int_0^{x+\sqrt{3}h}\frac{1}{2\sqrt{3}h}\left[\frac{x-y}{h}+\sqrt{3}\right]dydx \\
&+ \int_{1-\sqrt{3}h}^{\sqrt{3}h}\frac{1}{2\sqrt{3}h}\int_0^{1}\left[\frac{x-y}{h}+\sqrt{3}\right]dydx \\
&+ \int_{\sqrt{3}h}^{1}\left[\frac{1}{2\sqrt{3}h}\int_{x-\sqrt{3}h}^{1}\left\{\frac{x-y}{h}+\sqrt{3}\right\}dy+\int_0^{x-\sqrt{3}h}1dy\right]dx
\end{aligned}$$

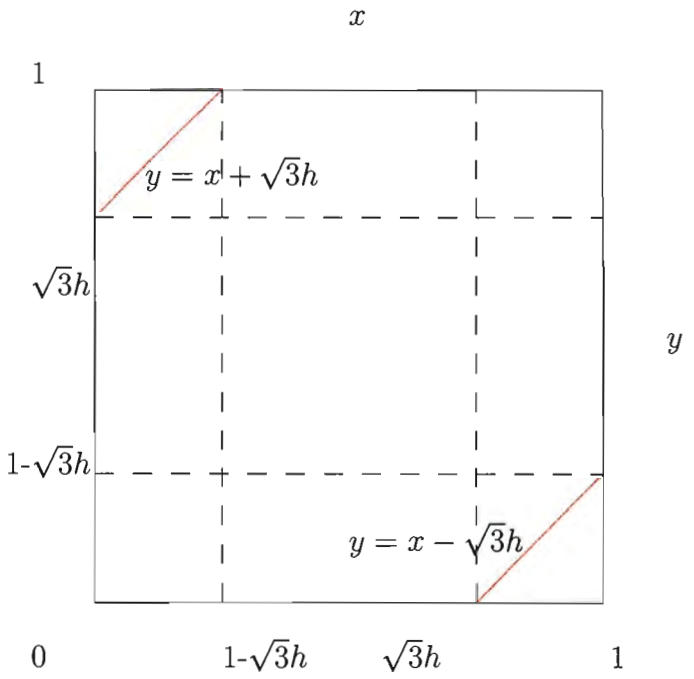


Figure 4.2: Case 2, $\frac{1}{2} < \sqrt{3}h < 1$

$$
\begin{aligned}
&= \int_0^{\sqrt{3}h} \left[\frac{x^2}{4\sqrt{3}h} + \frac{x}{2} + \frac{\sqrt{3}h}{4} \right] dx + \int_{\sqrt{3}h}^{1-\sqrt{3}h} \left[\frac{x}{2\sqrt{3}h} - \frac{1}{4\sqrt{3}h} + \frac{1}{2} \right] dx \\
&+ \int_{1-\sqrt{3}h}^{1} \left[-\frac{x^2}{4\sqrt{3}h} + \frac{x}{2\sqrt{3}h} + \frac{x}{2} - \frac{1}{4\sqrt{3}h} + \frac{1}{2} - \frac{\sqrt{3}h}{4} \right] dx \\
&= -\frac{h^2}{4} + \frac{1}{12\sqrt{3}h} + \sqrt{3}h - \frac{1}{2} - \frac{7}{4}h^2 + \sqrt{3}h + \frac{h^2}{4} - \frac{1}{12\sqrt{3}h} - \sqrt{3}h + 1 \\
&= \frac{1}{2}.
\end{aligned}
$$

For case 3, in a similar way as for cases 1 and 2, using Figure 4.3 leads to

$$
\begin{aligned}
&\mathrm{E}\left[K\left(\frac{X_1 - X_2}{h} \right) \right] \\
&= \int_0^1 \left\{ \int_0^1 \frac{1}{2\sqrt{3}h} \left[\frac{x-y}{h} + \sqrt{3} \right] dy \right\} dx \\
&= \int_0^1 \left[\frac{x}{2\sqrt{3}h} - \frac{1}{4\sqrt{3}h} + \frac{1}{2} \right] dx \\
&= \frac{1}{4\sqrt{3}h} - \frac{1}{4\sqrt{3}h} + \frac{1}{2} \\
&= \frac{1}{2}.
\end{aligned}
$$

Next, consider evaluating

$$
\begin{aligned}
\mathrm{E}\left[K\left(\frac{X_1 - X_2}{h} \right) F(X_1) \right] &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} K\left(\frac{x-y}{h} \right) f(x,y) F(x) dy dx \\
&= \int_0^1 \left\{ \int_0^1 K\left(\frac{x-y}{h} \right) dy \right\} F(x) dx,
\end{aligned}
$$

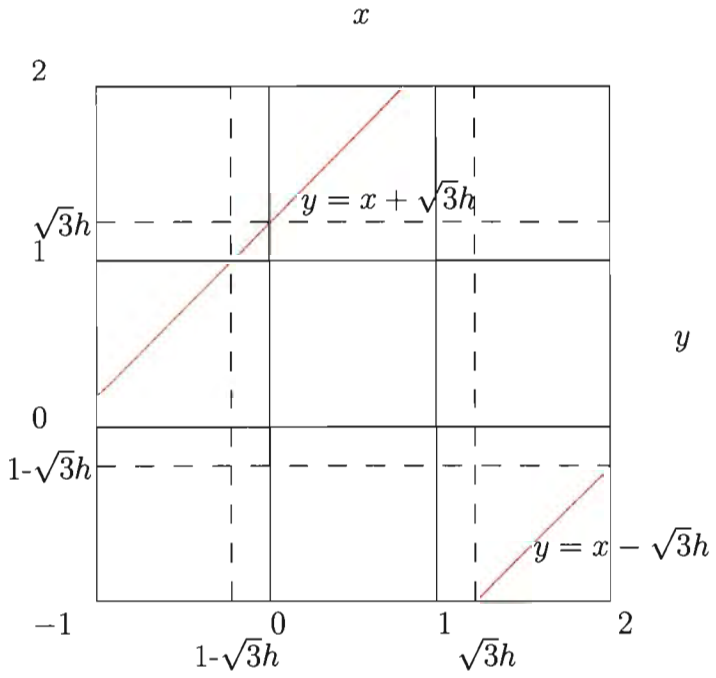


Figure 4.3: Case 3, $\sqrt{3}h > 1$

using the independence of X_1 and X_2. Considering Figure 4.1 on page 32 it follows that for case 1 we have

i. $0 < x < \sqrt{3}h$

$$\begin{aligned} & \mathrm{E}\left[K\left(\frac{X_1 - X_2}{h}\right) F(X_1)\right] \\ = & \int_0^{\sqrt{3}h} \frac{x}{2\sqrt{3}h} \int_0^{x+\sqrt{3}h} \left[\frac{x-y}{h} + \sqrt{3}\right] dydx \\ = & \int_0^{\sqrt{3}h} \left[\frac{x^3}{4\sqrt{3}h} + \frac{x^2}{2} + \frac{\sqrt{3}hx}{4}\right] dx \\ = & \frac{17\sqrt{3}h^3}{16}. \end{aligned}$$

ii. $\sqrt{3}h < x < 1 - \sqrt{3}h$

$$\begin{aligned} & \mathrm{E}\left[K\left(\frac{X_1 - X_2}{h}\right) F(X_1)\right] \\ = & \int_{\sqrt{3}h}^{1-\sqrt{3}h} \left[\frac{x}{2\sqrt{3}h} \int_{x-\sqrt{3}h}^{x+\sqrt{3}h} \left(\frac{x-y}{h} + \sqrt{3}\right) dy + \int_0^{x-\sqrt{3}h} 1 dy\right] dx \\ = & \int_{\sqrt{3}h}^{1-\sqrt{3}h} x^2 dx \\ = & -2\sqrt{3}h^3 + 3h^2 - \sqrt{3}h + \frac{1}{3}. \end{aligned}$$

iii. $1-\sqrt{3}h < x < 1$

$$
\begin{aligned}
&\mathrm{E}\left[K\left(\frac{X_1-X_2}{h}\right)F(X_1)\right] \\
&= \int_{1-\sqrt{3}h}^{1}\left[\frac{x}{2\sqrt{3}h}\int_{x-\sqrt{3}h}^{1}\left(\frac{x-y}{h}+\sqrt{3}\right)dy + \int_{0}^{x-\sqrt{3}h} 1dy\right]dx \\
&= \int_{1-\sqrt{3}h}^{1}\left[-\frac{x^3}{4\sqrt{3}h}+\frac{x^2}{2\sqrt{3}h}+\frac{x^2}{2}-\frac{x}{4\sqrt{3}h}+\frac{x}{2}-\frac{\sqrt{3}hx}{4}\right]dx \\
&= \frac{17\sqrt{3}h^3}{16}-\frac{13h^2}{4}+\sqrt{3}h.
\end{aligned}
$$

Collecting the results it follows that for case 1

$$
\mathrm{E}\left[K\left(\frac{X_1-X_2}{h}\right)F(X_1)\right] = \frac{\sqrt{3}h^3}{8}-\frac{h^2}{4}+\frac{1}{3}.
$$

For case 2, following the steps outlined in (i), (ii) and (iii) and using Figure 4.2 on page 34 we have

$$
\begin{aligned}
&\mathrm{E}\left[K\left(\frac{X_1-X_2}{h}\right)F(X_1)\right] \\
&= \int_{0}^{1-\sqrt{3}h}\frac{x}{2\sqrt{3}h}\int_{0}^{x+\sqrt{3}h}\left[x-y+\sqrt{3}h\right]dydx \\
&+ \int_{1-\sqrt{3}h}^{\sqrt{3}h}\frac{x}{2\sqrt{3}h}\int_{0}^{1}\left[x-y+\sqrt{3}h\right]dydx \\
&+ \int_{\sqrt{3}h}^{1}\left[\frac{x}{2\sqrt{3}h}\int_{x-\sqrt{3}h}^{1}\left[x-y+\sqrt{3}h\right]dy + \int_{0}^{x-\sqrt{3}h}1dy\right]dx \\
&= \int_{0}^{\sqrt{3}h}\left[\frac{x^3}{4\sqrt{3}h}+\frac{x^2}{2}+\frac{\sqrt{3}hx}{4}\right]dx + \int_{\sqrt{3}h}^{1-\sqrt{3}h}\left[\frac{x^2}{2\sqrt{3}h}-\frac{x}{4\sqrt{3}h}+\frac{x}{2}\right]dx \\
&+ \int_{1-\sqrt{3}h}^{1}\left[-\frac{x^3}{4\sqrt{3}h}+\frac{x^2}{2\sqrt{3}h}+\frac{x^2}{2}-\frac{x}{4\sqrt{3}h}+\frac{x}{2}-\frac{\sqrt{3}hx}{4}\right]dx \\
&= \frac{\sqrt{3}h^3}{16}-\frac{1}{12}+\frac{1}{16\sqrt{3}h}+h^2-\frac{1}{24\sqrt{3}h}+\frac{\sqrt{3}h^3}{16}-\frac{5h^2}{4}+\frac{5}{12}-\frac{1}{48\sqrt{3}h} \\
&= \frac{\sqrt{3}h^3}{8}-\frac{h^2}{4}+\frac{1}{3}.
\end{aligned}
$$

For case 3, in a similar way as for cases 1 and 2, using Figure 4.3 on page 35 leads to

$$
\begin{aligned}
&\mathrm{E}\left[K\left(\frac{X_1-X_2}{h}\right)F(X_1)\right] \\
&= \int_{0}^{1}\frac{x}{2\sqrt{3}h}\int_{0}^{1}\left[x-y+\sqrt{3}h\right]dydx \\
&= \int_{0}^{1}\frac{x^2}{2\sqrt{3}h}-\frac{x}{4\sqrt{3}h}+\frac{x}{2}dx
\end{aligned}
$$

$$= \frac{1}{4\sqrt{3}h} - \frac{1}{4\sqrt{3}h} + \frac{1}{2}$$
$$= \frac{1}{4} + \frac{1}{24\sqrt{3}h}.$$

Next, consider evaluating

$$\mathrm{E}\left[K^2\left(\frac{X_1 - X_2}{h}\right)\right] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} K^2\left(\frac{x-y}{h}\right) f(x,y)dydx$$
$$= \int_0^1 \left\{\int_0^1 K^2\left(\frac{x-y}{h}\right)dy\right\}dx,$$

using the independence of X_1 and X_2. Considering Figure 4.1 on page 32 it follows that for case 1 we have

i. $0 < x < \sqrt{3}h$

$$\mathrm{E}\left[K^2\left(\frac{X_1 - X_2}{h}\right)\right] = \int_0^{\sqrt{3}h}\int_0^{x+\sqrt{3}h}\left[\frac{1}{2\sqrt{3}h}\left\{\frac{x-y}{h} + \sqrt{3}\right\}\right]^2 dydx$$
$$= \int_0^{\sqrt{3}h}\left[\frac{x^3}{36h^2} + \frac{x^2}{4\sqrt{3}h} + \frac{x}{4} + \frac{h}{4\sqrt{3}}\right]dx$$
$$= \frac{15h^2}{16}.$$

ii. $\sqrt{3}h < x < 1 - \sqrt{3}h$

$$\mathrm{E}\left[K^2\left(\frac{X_1 - X_2}{h}\right)\right]$$
$$= \int_{\sqrt{3}h}^{1-\sqrt{3}h}\left[\int_{x-\sqrt{3}h}^{x+\sqrt{3}h}\left\{\frac{1}{2\sqrt{3}h}\left(\frac{x-y}{h} + \sqrt{3}\right)\right\}^2 dy + \int_0^{x-\sqrt{3}h} 1dy\right]dx$$
$$= \int_{\sqrt{3}h}^{1-\sqrt{3}h}\left[x - \frac{h}{\sqrt{3}}\right]dx$$
$$= 2h^2 - \frac{4\sqrt{3}h}{3} + \frac{1}{2}.$$

iii. $1 - \sqrt{3}h < x < 1$

$$\mathrm{E}\left[K^2\left(\frac{X_1 - X_2}{h}\right)\right]$$
$$= \int_{1-\sqrt{3}h}^{1}\left[\int_{x-\sqrt{3}h}^{1}\left\{\frac{x-y+\sqrt{3}h}{2\sqrt{3}h}\right\}^2 dy + \int_0^{x-\sqrt{3}h} 1dy\right]dx$$
$$= \int_{1-\sqrt{3}h}^{1}\Bigg[-\frac{x^3}{36h^2} - \frac{x^2}{4\sqrt{3}h} + \frac{x^2}{12h^2} + \frac{x}{2\sqrt{3}h} + \frac{3x}{4}$$
$$- \frac{x}{12h^2} + \frac{1}{36h^2} - \frac{1}{4\sqrt{3}h} - \frac{5h}{4\sqrt{3}} + \frac{1}{4}\Bigg]dx$$
$$= -\frac{41h^2}{16} + \sqrt{3}h.$$

Collecting the results it follows that for case 1

$$\mathrm{E}\left[K^2\left(\frac{X_1-X_2}{h}\right)\right] = \frac{3h^2}{8}-\frac{h}{\sqrt{3}}+\frac{1}{2}.$$

For case 2, following the steps outlined in (i), (ii) and (iii) and using Figure 4.2 on page 34 we have

$$\begin{aligned}
\mathrm{E}\left[K^2\left(\frac{X_1-X_2}{h}\right)\right] &= \int_0^{1-\sqrt{3}h}\int_0^{x+\sqrt{3}h}\left[\frac{x-y+\sqrt{3}h}{2\sqrt{3}h}\right]^2 dydx \\
&\quad + \int_{1-\sqrt{3}h}^{\sqrt{3}h}\int_0^{1}\left[\frac{x-y+\sqrt{3}h}{2\sqrt{3}h}\right]^2 dydx \\
&\quad + \int_{\sqrt{3}h}^{1}\left[\int_{x-\sqrt{3}h}^{1}\left[\frac{x-y+\sqrt{3}h}{2\sqrt{3}h}\right]^2 dy + \int_0^{x-\sqrt{3}h} 1dy\right] dx \\
&= \int_0^{\sqrt{3}h}\left[\frac{x^3}{36h^2}+\frac{x^2}{4\sqrt{3}h}+\frac{x}{4}+\frac{\sqrt{3}h}{12}\right] dx \\
&\quad + \int_{\sqrt{3}h}^{1-\sqrt{3}h}\left[\frac{x^2}{12h^2}-\frac{x}{12h^2}+\frac{x}{2\sqrt{3}h}+\frac{1}{36h^2}-\frac{1}{4\sqrt{3}h}+\frac{1}{4}\right] dx \\
&\quad + \int_{1-\sqrt{3}h}^{1}\Bigg[-\frac{x^3}{36h^2}+\frac{x^2}{12h^2}-\frac{x^2}{4\sqrt{3}h}+\frac{x}{2\sqrt{3}h} \\
&\qquad +\frac{3x}{4}-\frac{x}{12h^2}+\frac{1}{36h^2}-\frac{1}{4\sqrt{3}h}-\frac{5\sqrt{3}h}{12}+\frac{1}{4}\Bigg] dx \\
&= -\frac{h^2}{16}+\frac{1}{144h^2}+\frac{2h}{\sqrt{3}}-\frac{1}{2}+\frac{1}{6\sqrt{3}h}-\frac{1}{72h^2} \\
&\quad +\frac{7h^2}{16}-\sqrt{3}h+1-\frac{1}{6\sqrt{3}h}+\frac{1}{144h^2} \\
&= \frac{3h^2}{8}-\frac{h}{\sqrt{3}}+\frac{1}{2}.
\end{aligned}$$

For case 3, in a similar way as for cases 1 and 2, using Figure 4.3 on page 35 leads to

$$\begin{aligned}
&\mathrm{E}\left[K^2\left(\frac{X_1-X_2}{h}\right)\right] \\
&= \int_0^1\int_0^1\left[\frac{x-y+\sqrt{3}h}{2\sqrt{3}h}\right]^2 dydx \\
&= \int_0^1\left[\frac{x^2}{12h^2}-\frac{x}{12h^2}+\frac{x}{2\sqrt{3}h}+\frac{1}{36h^2}-\frac{1}{4\sqrt{3}h}+\frac{1}{4}\right] dx \\
&= \frac{1}{4}+\frac{1}{72h^2}.
\end{aligned}$$

Next, consider evaluating

$$\mathrm{E}\left[K\left(\frac{X_1-X_2}{h}\right)K\left(\frac{X_1-X_3}{h}\right)\right]$$

$$
\begin{aligned}
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} K\left(\frac{x-y}{h}\right) K\left(\frac{x-z}{h}\right) f(x,y,z)dzdydx \\
&= \int_0^1\int_0^1\int_0^1 K\left(\frac{x-y}{h}\right) K\left(\frac{x-z}{h}\right) dzdydx \\
&= \int_0^1 \left\{\int_0^1 K\left(\frac{x-y}{h}\right) dy \int_0^1 K\left(\frac{x-z}{h}\right) dz\right\} dx \\
&= \int_0^1 \left\{\int_0^1 K\left(\frac{x-y}{h}\right) dy\right\}^2 dx,
\end{aligned}
$$

using the independence of X_1, X_2 and X_3. Considering Figure 4.1 on page 32 and using the results for $\mathrm{E}\left[K\left(\frac{X_1-X_2}{h}\right)\right]$ it follows that for case 1 we have

i. $0 < x < \sqrt{3}h$

$$
\begin{aligned}
&\mathrm{E}\left[K\left(\frac{X_1-X_2}{h}\right) K\left(\frac{X_1-X_3}{h}\right)\right] \\
&= \int_0^{\sqrt{3}h}\left[\frac{x^2}{4\sqrt{3}h} + \frac{x}{2} + \frac{\sqrt{3}h}{4}\right]^2 dx \\
&= \int_0^{\sqrt{3}h}\left[\frac{x^4}{48h^2} + \frac{x^3}{4\sqrt{3}h} + \frac{3x^2}{8} + \frac{\sqrt{3}hx}{4} + \frac{3h^2}{16}\right] dx \\
&= \frac{93\sqrt{3}h^3}{80}.
\end{aligned}
$$

ii. $\sqrt{3}h < x < 1-\sqrt{3}h$

$$
\begin{aligned}
\mathrm{E}\left[K\left(\frac{X_1-X_2}{h}\right) K\left(\frac{X_1-X_3}{h}\right)\right] &= \int_{\sqrt{3}h}^{1-\sqrt{3}h} [x]^2\, dx \\
&= \int_{\sqrt{3}h}^{1-\sqrt{3}h} \left[x^2\right] dx \\
&= -2\sqrt{3}h^3 + 3h^2 - \sqrt{3}h + \frac{1}{3}.
\end{aligned}
$$

iii. $1-\sqrt{3}h < x < 1$

$$
\begin{aligned}
&\mathrm{E}\left[K\left(\frac{X_1-X_2}{h}\right) K\left(\frac{X_1-X_3}{h}\right)\right] \\
&= + \int_{1-\sqrt{3}h}^{1}\left[-\frac{x^2}{4\sqrt{3}h} + \frac{x}{2\sqrt{3}h} + \frac{x}{2} - \frac{1}{4\sqrt{3}h} + \frac{1}{2} - \frac{\sqrt{3}h}{4}\right]^2 dx \\
&= \int_{1-\sqrt{3}h}^{1}\Bigg[\frac{x^4}{48h^2} - \frac{x^3}{4\sqrt{3}h} - \frac{x^3}{12h^2} + \frac{x^2}{4\sqrt{3}h} + \frac{3x^2}{8} + \frac{x^2}{8h^2} + \frac{x}{4\sqrt{3}h} \\
&\qquad + \frac{x}{4} - \frac{x}{12h^2} - \frac{\sqrt{3}hx}{4} + \frac{3h^2}{16} - \frac{\sqrt{3}h}{4} - \frac{1}{4\sqrt{3}h} + \frac{3}{8} + \frac{1}{48h^2}\Bigg] dx \\
&= \frac{93\sqrt{3}h^3}{80} - \frac{7h^2}{2} + \sqrt{3}h.
\end{aligned}
$$

Collecting the results it follows that for case 1

$$\mathrm{E}\left[K^2\left(\frac{X_1-X_2}{h}\right)\right] \;=\; \frac{3h^2}{8}-\frac{h}{\sqrt{3}}+\frac{1}{2}.$$

For case 2, following the steps outlined in (i), (ii) and (iii) and using Figure 4.2 on page 34 we have

$$\begin{aligned}
&\mathrm{E}\left[K\left(\frac{X_1-X_2}{h}\right)K\left(\frac{X_1-X_3}{h}\right)\right]\\
&=\int_0^{\sqrt{3}h}\left[\frac{x^2}{4\sqrt{3}h}+\frac{x}{2}+\frac{\sqrt{3}h}{4}\right]^2dx+\int_{\sqrt{3}h}^{1-\sqrt{3}h}\left[\frac{x}{2\sqrt{3}h}-\frac{1}{4\sqrt{3}h}+\frac{1}{2}\right]^2dx\\
&\quad+\int_{1-\sqrt{3}h}^{1}\left[-\frac{x^2}{4\sqrt{3}h}+\frac{x}{2\sqrt{3}h}+\frac{x}{2}-\frac{1}{4\sqrt{3}h}+\frac{1}{2}-\frac{\sqrt{3}h}{4}\right]^2dx\\
&=\int_0^{\sqrt{3}h}\left[\frac{x^4}{48h^2}+\frac{x^3}{4\sqrt{3}h}+\frac{3x^2}{8}+\frac{\sqrt{3}hx}{4}+\frac{3h^2}{16}\right]dx\\
&\quad+\int_{\sqrt{3}h}^{1-\sqrt{3}h}\left[\frac{x^2}{12h^2}-\frac{x}{12h^2}+\frac{x}{\sqrt{3}h}+\frac{1}{4}-\frac{1}{4\sqrt{3}h}+\frac{1}{48h^2}\right]dx\\
&\quad+\int_{1-\sqrt{3}h}^{1}\left[\frac{x^4}{48h^2}-\frac{x^3}{4\sqrt{3}h}-\frac{x^3}{12h^2}+\frac{x^2}{4\sqrt{3}h}+\frac{3x^2}{8}+\frac{x^2}{8h^2}+\frac{x}{4\sqrt{3}h}\right.\\
&\qquad\left.+\frac{x}{4}-\frac{x}{12h^2}-\frac{\sqrt{3}hx}{4}+\frac{3h^2}{16}-\frac{\sqrt{3}h}{4}-\frac{1}{4\sqrt{3}h}+\frac{3}{8}+\frac{1}{48h^2}\right]dx\\
&=-\frac{3\sqrt{3}h^3}{80}+\frac{1}{240h^2}+\frac{2h}{\sqrt{3}}-\frac{1}{2}+\frac{1}{8\sqrt{3}h}-\frac{1}{144h^2}\\
&\quad-\frac{3\sqrt{3}h^3}{80}+\frac{h^2}{2}-\sqrt{3}h+1-\frac{1}{6\sqrt{3}h}+\frac{1}{240h^2}\\
&=-\frac{3\sqrt{3}h^3}{40}+\frac{1}{720h^2}-\frac{h}{\sqrt{3}}+\frac{1}{2}-\frac{1}{24\sqrt{3}h}+\frac{h^2}{2}.
\end{aligned}$$

For case 3, in a similar way as for cases 1 and 2, using Figure 4.3 on page 35 leads to

$$\begin{aligned}
&\mathrm{E}\left[K\left(\frac{X_1-X_2}{h}\right)K\left(\frac{X_1-X_3}{h}\right)\right]\\
&=\int_0^1\left[\frac{x}{2\sqrt{3}h}-\frac{1}{4\sqrt{3}h}+\frac{1}{2}\right]^2dx\\
&=\int_0^1\left[\frac{x^2}{12h^2}-\frac{x}{12h^2}+\frac{x}{\sqrt{3}h}+\frac{1}{4}-\frac{1}{4\sqrt{3}h}+\frac{1}{48h^2}\right]dx\\
&=\frac{1}{4}+\frac{1}{144h^2}.
\end{aligned}$$

Summarising,

$$\mathrm{E}\left[F(X_1)=\frac{1}{2},\right] \tag{4.4a}$$

$$\mathrm{E}\left[F^2(X_1)=\frac{1}{3},\right] \tag{4.4b}$$

$$\mathrm{E}\left[K\left(\frac{X_1-X_2}{h}\right)=\frac{1}{2}\right], \tag{4.4c}$$

$$\mathrm{E}\left[K\left(\frac{X_1-X_2}{h}\right)F(X_1)\right]=\begin{cases}\frac{\sqrt{3}h^3}{8}-\frac{h^2}{4}+\frac{1}{3} & \text{if } 0\le\sqrt{3}h<\frac{1}{2}\\ \frac{\sqrt{3}h^3}{8}-\frac{h^2}{4}+\frac{1}{3} & \text{if } \frac{1}{2}\le\sqrt{3}h<1\\ \frac{1}{4}+\frac{1}{24\sqrt{3}h} & \text{if } \sqrt{3}h>1,\end{cases} \tag{4.4d}$$

$$\mathrm{E}\left[K^2\left(\frac{X_1-X_2}{h}\right)\right]=\begin{cases}\frac{3h^2}{8}-\frac{\sqrt{3}h}{3}+\frac{1}{2} & \text{if } 0\le\sqrt{3}h<\frac{1}{2}\\ \frac{3h^2}{8}-\frac{\sqrt{3}h}{3}+\frac{1}{2} & \text{if } \frac{1}{2}\le\sqrt{3}h<1\\ \frac{1}{4}+\frac{1}{72h^2} & \text{if } \sqrt{3}h>1,\end{cases} \tag{4.4e}$$

$$\begin{aligned}&\mathrm{E}\left[K\left(\frac{X_1-X_2}{h}\right)K\left(\frac{X_1-X_3}{h}\right)\right]\\ &=\begin{cases}\frac{13\sqrt{3}h^3}{40}-\frac{h^2}{2}+\frac{1}{3} & \text{if } 0\le\sqrt{3}h<\frac{1}{2}\\ -\frac{3\sqrt{3}h^3}{40}+\frac{1}{720h^2}-\frac{h}{\sqrt{3}}+\frac{1}{2}-\frac{1}{24\sqrt{3}h}+\frac{h^2}{2} & \text{if } \frac{1}{2}\le\sqrt{3}h<1\\ \frac{1}{4}+\frac{1}{144h^2} & \text{if } \sqrt{3}h>1.\end{cases}\end{aligned} \tag{4.4f}$$

Using (4.4a) to (4.4f), (1.24) for the uniform population/uniform kernel in case 1 can be written as

$$\begin{aligned}\mathrm{MASE}(h) &= \frac{1}{4n^2}+\frac{(n-1)}{n^2}\mathrm{E}\left[K\left(\frac{X_1-X_2}{h}\right)\right]-\frac{1}{n}\mathrm{E}\left[F(X_1)\right]+\\ &\quad\frac{n-1}{n^2}\mathrm{E}\left[K^2\left(\frac{X_1-X_2}{h}\right)\right]-\frac{2(n-1)}{n}\mathrm{E}\left[K\left(\frac{X_1-X_2}{h}\right)F(X_1)\right]\\ &\quad+\mathrm{E}\left[F^2(X_1)\right]+\frac{(n-1)(n-2)}{n^2}\mathrm{E}\left[K\left(\frac{X_1-X_2}{h}\right)K\left(\frac{X_1-X_3}{h}\right)\right]\\ &= \frac{1}{4n^2}+\frac{(n-1)}{n^2}\left[\frac{1}{2}\right]-\frac{1}{n}\left[\frac{1}{2}\right]\\ &\quad+\frac{n-1}{n^2}\left[\frac{1}{2}+\frac{3h^2}{8}-\frac{\sqrt{3}h}{3}\right]-\frac{2(n-1)}{n}\left[\frac{\sqrt{3}h^3}{8}+\frac{1}{3}-\frac{h^2}{4}\right]\\ &\quad+\left[\frac{1}{3}\right]+\frac{(n-1)(n-2)}{n^2}\left[\frac{1}{3}-\frac{h^2}{2}+\frac{13\sqrt{3}h^3}{40}\right]\\ &= \frac{3\sqrt{3}h^3}{40}+\frac{1}{n}\left[\frac{1}{6}-\frac{29\sqrt{3}h^3}{40}+\frac{11h^2}{8}-\frac{h}{\sqrt{3}}\right]\\ &\quad+\frac{1}{n^2}\left[-\frac{1}{12}-\frac{11h^2}{8}+\frac{h}{\sqrt{3}}+\frac{13\sqrt{3}h^3}{20}\right].\end{aligned}$$

Similarly for case 2

$$
\begin{aligned}
\mathrm{MASE}(h) &= \frac{1}{4n^2} + \frac{(n-1)}{n^2}\,\mathrm{E}\left[K\left(\frac{X_1-X_2}{h}\right)\right] - \frac{1}{n}\,\mathrm{E}\left[F(X_1)\right] + \\
&\quad \frac{n-1}{n^2}\,\mathrm{E}\left[K^2\left(\frac{X_1-X_2}{h}\right)\right] - \frac{2(n-1)}{n}\,\mathrm{E}\left[\left(\frac{X_1-X_2}{h}\right)F(X_1)\right] + \mathrm{E}\left[F^2(X_1)\right] + \\
&\quad \frac{(n-1)(n-2)}{n^2}\,\mathrm{E}\left[K\left(\frac{X_1-X_2}{h}\right)K\left(\frac{X_1-X_3}{h}\right)\right] \\
&= \frac{1}{4n^2} + \frac{(n-1)}{n^2}\left[\frac{1}{2}\right] - \frac{1}{n}\left[\frac{1}{2}\right] \\
&\quad + \frac{n-1}{n^2}\left[\frac{1}{2} - \frac{3h^2}{8} - \frac{\sqrt{3}h}{3}\right] - \frac{2(n-1)}{n}\left[\frac{1}{3} - \frac{h^2}{4} + \frac{\sqrt{3}h^3}{8}\right] \\
&\quad + \left[\frac{1}{3}\right] + \frac{(n-1)(n-2)}{n^2}\left[-\frac{3\sqrt{3}h^3}{40} + \frac{1}{720h^2} - \frac{h}{\sqrt{3}} + \frac{1}{2} - \frac{1}{24\sqrt{3}h} + \frac{h^2}{2}\right] \\
&= -\frac{13\sqrt{3}h^3}{40} + h^2 + \frac{1}{6} + \frac{1}{720h^2} - \frac{h}{\sqrt{3}} - \frac{1}{24\sqrt{3}h} \\
&\quad + \frac{1}{n}\left[-\frac{1}{3} + \frac{19\sqrt{3}h^3}{40} - \frac{13h^2}{8} + \frac{2h}{\sqrt{3}} - \frac{1}{240h^2} + \frac{1}{8\sqrt{3}h}\right] \\
&\quad + \frac{1}{n^2}\left[\frac{1}{4} + \frac{5h^2}{8} - \frac{h}{\sqrt{3}} - \frac{3\sqrt{3}h^3}{20} + \frac{1}{360h^2} - \frac{1}{12\sqrt{3}h}\right].
\end{aligned}
$$

For case 3

$$
\begin{aligned}
\mathrm{MASE}(h) &= \frac{1}{4n^2} + \frac{(n-1)}{n^2}\,\mathrm{E}\left[K\left(\frac{X_1-X_2}{h}\right)\right] - \frac{1}{n}\,\mathrm{E}\left[F(X_1)\right] + \\
&\quad \frac{n-1}{n^2}\,\mathrm{E}\left[K^2\left(\frac{X_1-X_2}{h}\right)\right] - \frac{2(n-1)}{n}\,\mathrm{E}\left[K\left(\frac{X_1-X_2}{h}\right)F(X_1)\right] \\
&\quad + \mathrm{E}\left[F^2(X_1)\right] + \frac{(n-1)(n-2)}{n^2}\,\mathrm{E}\left[K\left(\frac{X_1-X_2}{h}\right)K\left(\frac{X_1-X_3}{h}\right)\right] \\
&= \frac{1}{4n^2} + \frac{(n-1)}{n^2}\left[\frac{1}{2}\right] - \frac{1}{n}\left[\frac{1}{2}\right] \\
&\quad + \frac{n-1}{n^2}\left[\frac{1}{4} + \frac{1}{72h^2}\right] - \frac{2(n-1)}{n}\left[\frac{1}{4} + \frac{1}{24\sqrt{3}h}\right] \\
&\quad + \left[\frac{1}{3}\right] + \frac{(n-1)(n-2)}{n^2}\left[\frac{1}{4} + \frac{1}{144h^2}\right] \\
&= \frac{1}{12} - \frac{n-1}{12\sqrt{3}hn} + \frac{n-1}{144h^2n}.
\end{aligned}
$$

Summarising:

$$\text{MASE}(h) = \begin{cases} \frac{3\sqrt{3}h^3}{40} + \frac{1}{n}\left[\frac{1}{6} - \frac{29\sqrt{3}h^3}{40} + \frac{11h^2}{8} - \frac{h}{\sqrt{3}}\right] \\ +\frac{1}{n^2}\left[-\frac{1}{12} - \frac{11h^2}{8} + \frac{h}{\sqrt{3}} + \frac{13\sqrt{3}h^3}{20}\right] & \text{if } 0 \le \sqrt{3}h < \frac{1}{2} \\ -\frac{13\sqrt{3}h^3}{40} + h^2 + \frac{1}{6} + \frac{1}{720h^2} - \frac{h}{\sqrt{3}} - \frac{1}{24\sqrt{3}h} \\ +\frac{1}{n}\left[-\frac{1}{3} + \frac{19\sqrt{3}h^3}{40} - \frac{13h^2}{8} + \frac{2h}{\sqrt{3}} - \frac{1}{240h^2} + \frac{1}{8\sqrt{3}h}\right] \\ +\frac{1}{n^2}\left[\frac{1}{4} + \frac{5h^2}{8} - \frac{h}{\sqrt{3}} - \frac{3\sqrt{3}h^3}{20} + \frac{1}{360h^2} - \frac{1}{12\sqrt{3}h}\right] & \text{if } \frac{1}{2} \le \sqrt{3}h < 1 \\ \frac{1}{12} - \frac{n-1}{n}\cdot\frac{1}{12\sqrt{3}h} + \frac{n-1}{n}\cdot\frac{1}{144h^2} & \text{if } \sqrt{3}h > 1. \end{cases} \tag{4.5}$$

From Van Graan (1983) it is known that

$$\text{MISE}(h) = \begin{cases} \frac{3\sqrt{3}h^3}{40} + \frac{1}{n}\left[-\frac{13\sqrt{3}h^3}{40} + \frac{7h^2}{8} - \frac{h}{\sqrt{3}} + \frac{1}{6}\right] & \text{if } 0 \le \sqrt{3}h < \frac{1}{2} \\ -\frac{13\sqrt{3}h^3}{40} + h^2 + \frac{1}{6} + \frac{1}{720h^2} - \frac{h}{\sqrt{3}} - \frac{1}{24\sqrt{3}h} \\ +\frac{1}{n}\left[\frac{3\sqrt{3}h^3}{40} - \frac{h^2}{8} - \frac{1}{720h^2} + \frac{1}{24\sqrt{3}h}\right] & \text{if } \frac{1}{2} \le \sqrt{3}h < 1 \\ \frac{1}{12} - \frac{1}{12\sqrt{3}h} + \frac{n+1}{n}\cdot\frac{1}{144h^2} & \text{if } \sqrt{3}h > 1. \end{cases} \tag{4.6}$$

It can be seen in Figures 4.4 — 4.8 which represents MASE(h) and MISE(h) for $n = 10$, $n = 20$, $n = 50$, $n = 100$ and $n = 500$ that MASE(h) tends to MISE(h) as n increases.

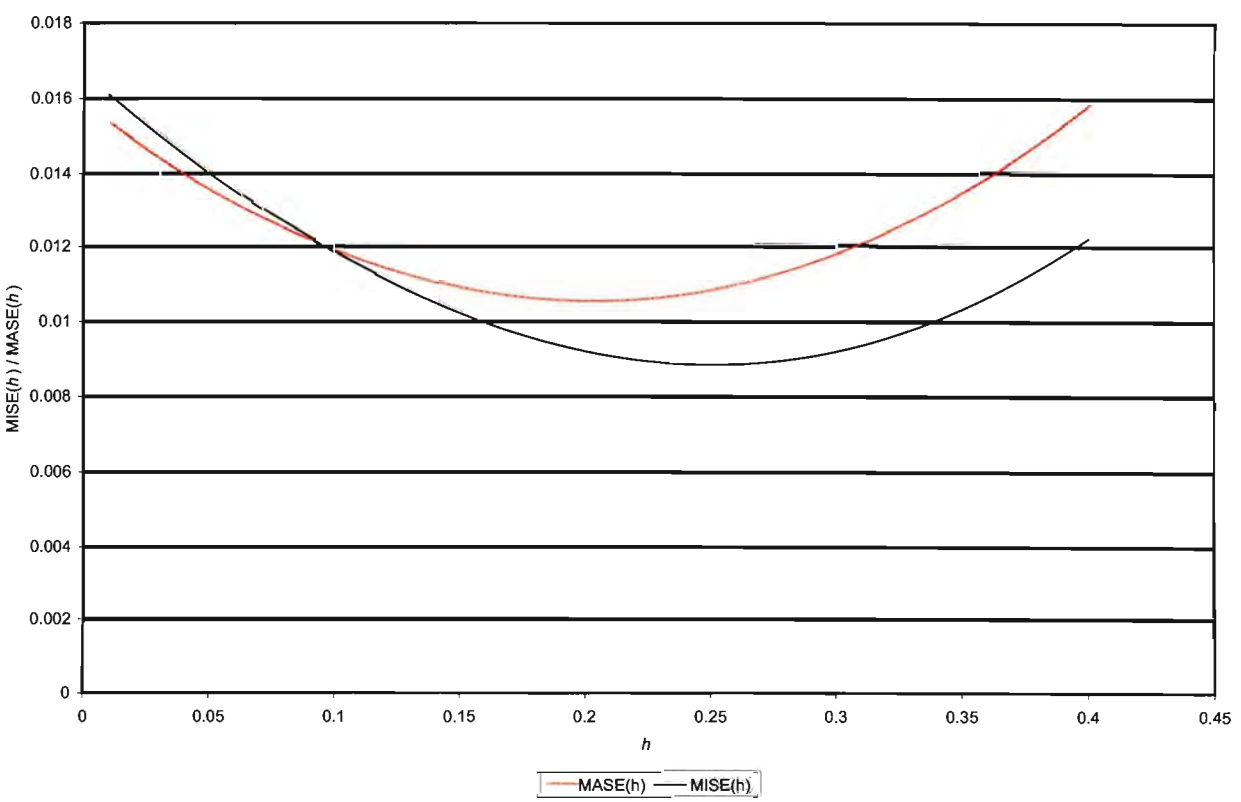


Figure 4.4: MASE(h) and MISE(h) with n=10

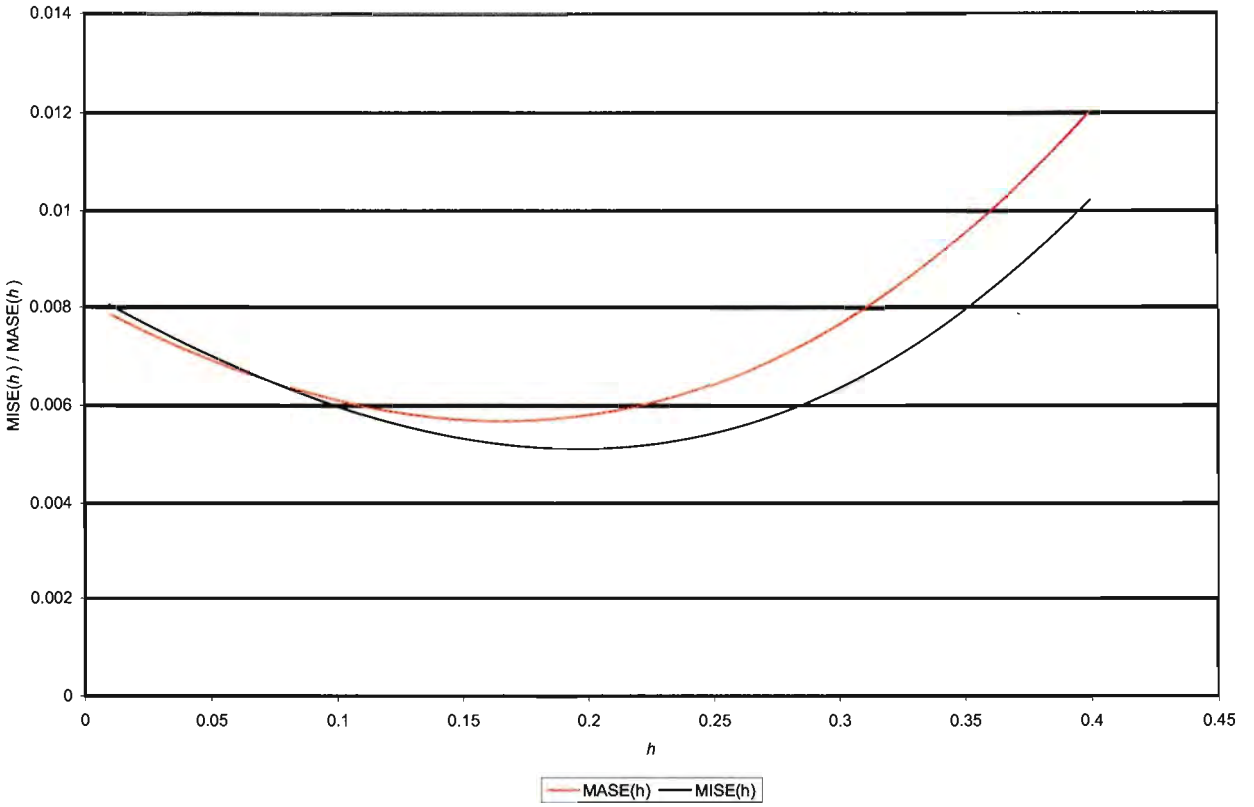


Figure 4.5: MASE(h) and MISE(h) with n=20

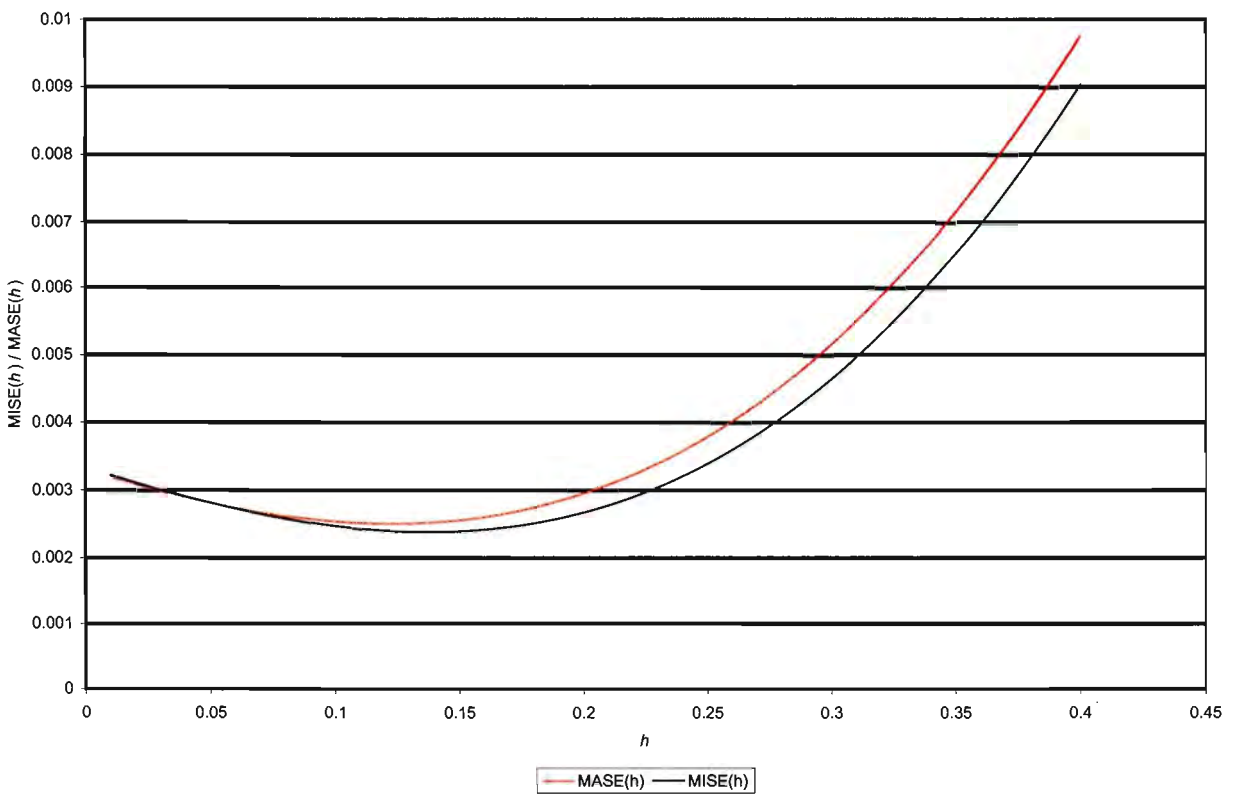


Figure 4.6: MASE(h) and MISE(h) with n=50

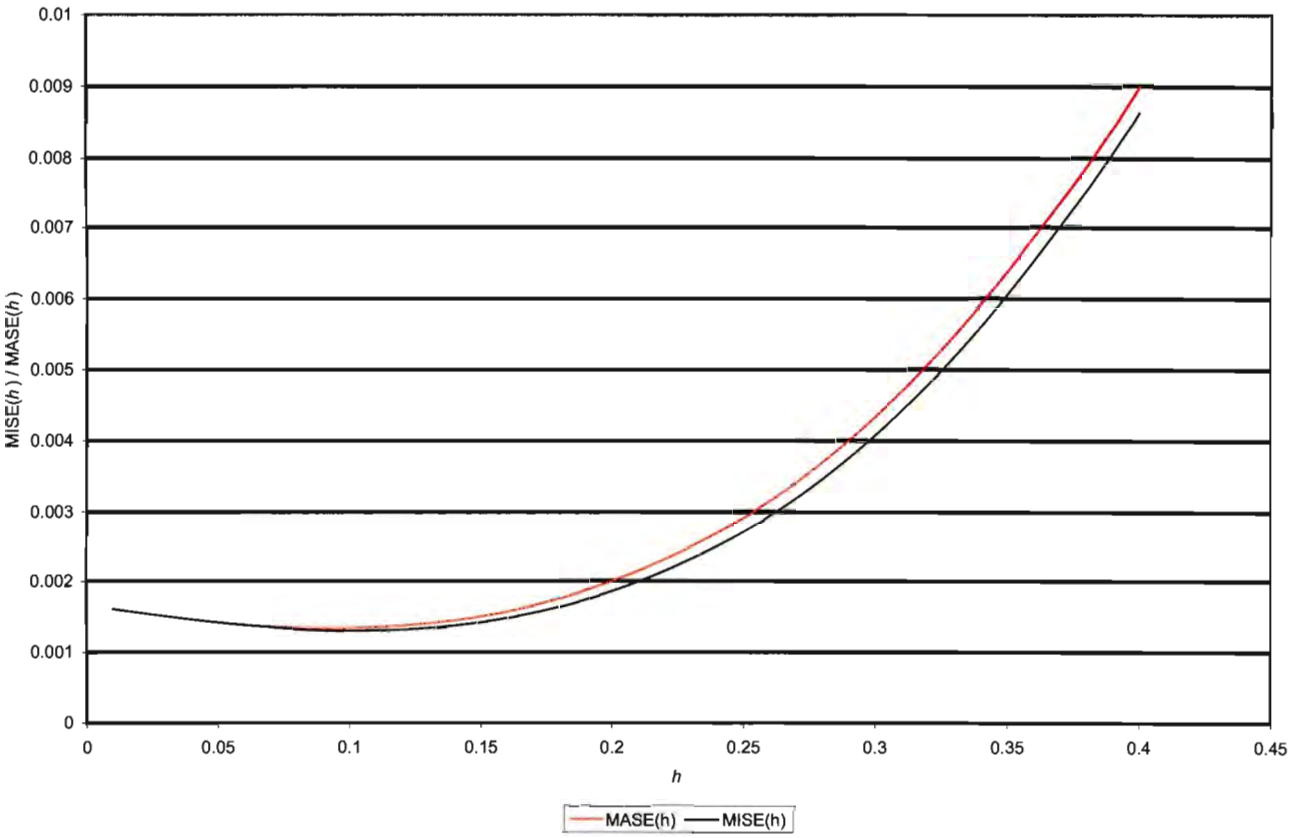


Figure 4.7: MASE(h) and MISE(h) with n=100

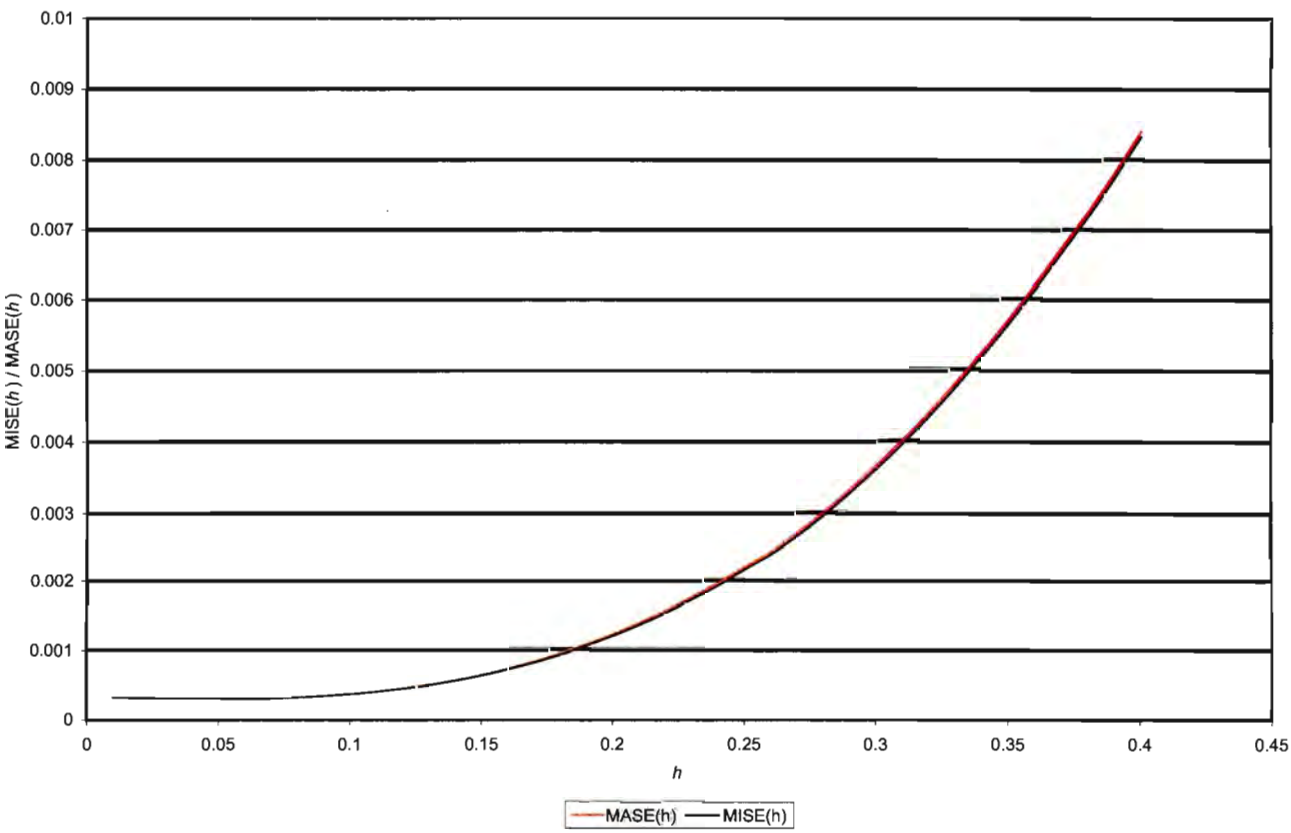


Figure 4.8: MASE(h) and MISE(h) with n=500

Now that we have expressions for MASE(h) and MISE(h) we can obtain the optimal values h_{MASE} and h_{MISE} that minimises the functions MASE(h) and MISE(h). Differentiating the expressions in (4.5) and (4.6), equating to 0 and solving we find that for $0 \leq \sqrt{3}h < \frac{1}{2}$

$$h_{\text{MASE}} = \frac{55-55n+\sqrt{360n^3-815n^2+550n-95}}{3\sqrt{3}\cdot(3n^2-29n+26)},$$

and

$$h_{\text{MISE}} = \frac{-35+\sqrt{360n-335}}{3\sqrt{3}\cdot(3n-13)}.$$

Optimal values h_{MASE} and h_{MISE} for different values of n are shown in Table 4.1.

n	h_{MISE}	h_{MASE}
10	0.251	0.202
15	0.218	0.181
25	0.180	0.155
50	0.138	0.122
100	0.103	0.094
500	0.050	0.048
1000	0.036	0.035

Table 4.1: Optimal values of the bandwidth

Remarks

1. Inspection of the expressions for MISE(h) and MASE(h) for $\sqrt{3}h > 1$, shows clearly that the measures MISE(h) and MASE(h) tends to each other for increasing values of n.

2. Table 4.1 shows that $h_{\text{MASE}} \leq h_{\text{MISE}}$ for typical values of n. Furthermore, larger values of n produce smaller values of $h_{\text{MASE}}/h_{\text{MISE}}$. This is in agreement with the observation that as $n \to \infty$ then $h \to 0$.

4.3 Asymptotic Theory

In this section the asymptotic properties of MASE will be investigated. To evaluate the expression for MASE(h) in (1.24), each of the terms, i.e., $\text{E}\left[K\left(\frac{X_1-X_2}{h}\right)\right]$, $\text{E}\left[K^2\left(\frac{X_1-X_2}{h}\right)\right]$, $\text{E}\left[K\left(\frac{X_1-X_2}{h}\right)K\left(\frac{X_1-X_3}{h}\right)\right]$ and $\text{E}\left[K\left(\frac{X_1-X_2}{h}\right)F(X_1)\right]$, will be treated separately.

The following notation will be used in the derivation of the results. Let

$$\mu_j(k) \text{ as defined in (1.5)}, \tag{4.7a}$$

$$D_j(K) \text{ as defined in (1.7)}, \tag{4.7b}$$

$$A_j(F) \text{ as defined in (3.9)}, \tag{4.7c}$$

$$E_j(F) \text{ as defined in (3.10)}. \tag{4.7d}$$

Assume that the kernel K has a symmetric density with respect to 0 and that its fifth moment exists. Assume also that F has five derivatives and that the fifth derivative is bounded.

In addition, we will need the following lemma:

Lemma 1. Let $\mu_2(k)$ and $D_2(K)$ be defined as in (1.5) and (1.7). Then

$$D_2(K) = \frac{1}{2}\mu_2(k)$$

Proof:

$D_2(K) = \int_{-\infty}^{\infty} z^2 k(z) K(z) dz$. Using the substitution $t = -z$ and the fact that $k(\cdot)$ is the symmetric around zero kernel density it follows that

$$\begin{aligned} D_2(K) &= \int_{-\infty}^{\infty} t^2 k(t) K(-t) dt \\ &= \int_{-\infty}^{\infty} t^2 k(t) \left[1 - K(t)\right] dt \\ &= \int_{-\infty}^{\infty} t^2 k(t) dt - \int_{-\infty}^{\infty} t^2 k(t) K(t) dt \\ &= \mu_2(k) - D_2(K), \end{aligned}$$

from which the result can be obtained.

Using the method of integration by parts, we find that

$$\begin{aligned} D_0(K) &= \int_{-\infty}^{\infty} K(z) k(z) dz \\ &= \frac{1}{2}. \end{aligned} \tag{4.8}$$

Furthermore, using the same method as above it follows that

$$\int_{-\infty}^{\infty} F(x) dF(x) = \frac{1}{2}, \tag{4.9}$$

and

$$\int_{-\infty}^{\infty} F^2(x) dF(x) = \frac{1}{3}. \tag{4.10}$$

The function $F(x-hz)$ can be written as follows using a Taylor series expansion

$$\begin{aligned} & F(x-hz) \\ &= F(x) - hzf(x) + \frac{1}{2}h^2z^2f'(x) - \frac{1}{6}h^3z^3f''(x) + \frac{1}{24}h^4z^4f'''(x) + h^5R(x,z), \end{aligned} \tag{4.11}$$

where $R(x,z)$ is a remainder term depending on both x and z.

The following result will be used several times in this section. Using integration by parts, the substitution $\frac{x-y}{h} = z$ and the Taylor series expansion in (4.11)

$$\begin{aligned} & \mathrm{E}\left[K\left(\frac{x-X}{h}\right)\right] \\ &= \int_{-\infty}^{\infty} K\left(\frac{x-y}{h}\right) f(y)dy \\ &= \frac{1}{h}\int_{-\infty}^{\infty} F(y)k\left(\frac{x-y}{h}\right) dy \\ &= \int_{-\infty}^{\infty} F(x-hz)k\left(z\right) dz \\ &= \int_{-\infty}^{\infty} \left[F(x) - hzf(x) + \frac{1}{2}h^2z^2f'(x) - \frac{1}{6}h^3z^3f''(x) + \frac{1}{24}h^4z^4f'''(x) + h^5R(x,z)\right] k(z)dz \\ &= F(x) + \frac{1}{2}h^2f'(x)\int_{-\infty}^{\infty} z^2k(z)dz + \frac{1}{24}h^4f'''(x)\int_{-\infty}^{\infty} z^4k(z)dz + O(h^5) \\ &= F(x) + \frac{1}{2}h^2f'(x)\mu_2(k) + \frac{1}{24}h^4f'''(x)\mu_4(k) + O(h^5), \end{aligned} \tag{4.12}$$

where $R(x,z)$ denotes a remainder term. Using the results in (4.9) and (4.12), as well as the independence of X_1 and X_2, we find that

$$\begin{aligned} & \mathrm{E}\left[K\left(\frac{X_1-X_2}{h}\right)\right] \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} K\left(\frac{x-y}{h}\right) f(x,y)dxdy \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left[K\left(\frac{x-y}{h}\right) f(y)dy\right] f(x)dx \\ &= \int_{-\infty}^{\infty} \left[F(x) + \frac{1}{2}h^2f'(x)\mu_2(k) + \frac{1}{24}h^4f'''(x)\mu_4(k) + O(h^5)\right] f(x)dx \\ &= \int_{-\infty}^{\infty} F(x)f(x)dx + \frac{1}{2}h^2\mu_2(k)\int_{-\infty}^{\infty} f'(x)f(x)dx + \frac{1}{24}h^4\mu_4(k)\int_{-\infty}^{\infty} f'''(x)f(x)dx + O(h^5) \\ &= \frac{1}{2} + \frac{1}{2}h^2\mu_2(k)A_1(F) + \frac{1}{24}h^4A_3(F) + O(h^5). \end{aligned} \tag{4.13}$$

Using integration by parts, the substitution $\frac{x-y}{h} = z$, (4.8) and the Taylor series expansion

in (4.11) it can be shown that

$$
\begin{aligned}
&\mathrm{E}\left[K^2\left(\frac{x-X}{h}\right)\right] \\
&= \int_{-\infty}^{\infty} K^2\left(\frac{x-y}{h}\right) f(y)dy \\
&= \frac{2}{h}\int_{-\infty}^{\infty} F(y)K\left(\frac{x-y}{h}\right) k\left(\frac{x-y}{h}\right) dy \\
&= 2\int_{-\infty}^{\infty} F(x-hz)K(z)k(z)dz \\
&= 2\int_{-\infty}^{\infty} \Big[F(x) - hzf(x) + \frac{1}{2}h^2z^2f'(x) - \frac{1}{6}h^3z^3f''(x) \\
&\quad + \frac{1}{24}h^4z^4f'''(x) + h^5R(x,z)\Big] K(z)k(z)dz \\
&= 2F(x)\int_{-\infty}^{\infty} K(z)k(z)dz - 2hf(x)\int_{-\infty}^{\infty} zK(z)k(z)dz + h^2f'(x)\int_{-\infty}^{\infty} z^2K(z)k(z)dz \\
&\quad - \frac{1}{3}h^3f''(x)\int_{-\infty}^{\infty} z^3K(z)k(z)dz + \frac{1}{12}h^4f'''(x)\int_{-\infty}^{\infty} z^4K(z)k(z)dz + O(h^5) \\
&= 2F(x)D_0(K) - 2hf(x)D_1(K) + h^2f'(x)D_2(K) - \frac{1}{3}h^3f''(x)D_3(K) \\
&\quad + \frac{1}{12}h^4f'''(x)D_4(K) + O(h^5) \\
&= F(x) - 2hf(x)D_1(K) + h^2f'(x)D_2(K) \\
&\quad - \frac{1}{3}h^3f''(x)D_3(K) + \frac{1}{12}h^4f'''(x)D_4(K) + O(h^5).
\end{aligned} \tag{4.14}
$$

Using the results in (4.9) and (4.14), as well as the independence of X_1 and X_2, we find that

$$
\begin{aligned}
&\mathrm{E}\left[K^2\left(\frac{X_1-X_2}{h}\right)\right] \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} K^2\left(\frac{x-y}{h}\right) f(x,y)dxdy \\
&= \int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty} K^2\left(\frac{x-y}{h}\right) f(y)dy\right] f(x)dx \\
&= \int_{-\infty}^{\infty} F(x)dF(x) - 2hD_1(K)\int_{-\infty}^{\infty} f(x)dF(x) \\
&\quad + h^2D_2(K)\int_{-\infty}^{\infty} f'(x)dF(x) - \frac{1}{3}h^3D_3(K)\int_{-\infty}^{\infty} f''(x)dF(x) \\
&\quad + \frac{1}{12}h^4D_4(K)\int_{-\infty}^{\infty} f'''(x)dF(x) + O(h^5) \\
&= \frac{1}{2} - 2hD_1(K)A_0(F) + h^2D_2(K)A_1(F) \\
&\quad - \frac{1}{3}h^3D_3(K)A_2(F) + \frac{1}{12}h^4D_4(K)A_3(F) + O(h^5).
\end{aligned} \tag{4.15}
$$

Using (4.10), (4.12) and the independence of X_1, X_2 and X_3, we find that

$$
\mathrm{E}\left[K\left(\frac{X_1-X_2}{h}\right) K\left(\frac{X_1-X_3}{h}\right)\right]
$$

$$
\begin{aligned}
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} K\left(\frac{x-y}{h}\right) K\left(\frac{x-z}{h}\right) f(x,y,z)dxdydz \\
&= \int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty} K\left(\frac{x-y}{h}\right) f(y)dy\right]\left[\int_{-\infty}^{\infty} K\left(\frac{x-z}{h}\right) f(z)dz\right] f(x)dx \\
&= \int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty} K\left(\frac{x-y}{h}\right) f(y)dy\right]^2 f(x)dx \\
&= \int_{-\infty}^{\infty}\left[F(x)+\frac{1}{2}h^2 f'(x)\mu_2(k)+\frac{1}{24}h^4 f'''(x)\mu_4(k)+O(h^5)x)\right]^2 f(x)dx \\
&= \int_{-\infty}^{\infty}\Big[F^2(x)+h^2F(x)f'(x)\mu_2(k)+\frac{1}{12}h^4F(x)f'''(x)\mu_4(k) \\
&\quad +\frac{1}{4}h^4\mu_2^2(k)\left\{f'(x)\right\}^2+O(h^5)\Big] f(x)dx \\
&= \int_{-\infty}^{\infty} F^2(x)dF(x)+h^2\mu_2(k)\int_{-\infty}^{\infty} F(x)f'(x)dF(x)+\frac{1}{4}h^4\mu_2^2(k)\int_{-\infty}^{\infty}\left\{f'(x)\right\}^2 dF(x) \\
&\quad +\frac{1}{12}h^4\mu_4(k)\int_{-\infty}^{\infty} F(x)f'''(x)dF(x)+O(h^5) \\
&= \frac{1}{3}+h^2\mu_2(k)E_1(F)+\frac{1}{4}h^4\mu_2^2(k)\int_{-\infty}^{\infty}\left\{f'(x)\right\}^2 dF(x) \\
&\quad +\frac{1}{12}h^4\mu_4(k)E_3(F)+O(h^5). \qquad (4.16)
\end{aligned}
$$

Finally, using (4.10), (4.12) and the independence of X_1 and X_2 it follows that

$$
\begin{aligned}
&\mathrm{E}\left[K\left(\frac{X_1-X_2}{h}\right)F(X_1)\right] \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} K\left(\frac{x-y}{h}\right) F(x)f(x,y)dxdy \\
&= \int_{-\infty}^{\infty} F(x)\left[\int_{-\infty}^{\infty} K\left(\frac{x-y}{h}\right) f(y)dy\right] f(x)dx \\
&= \int_{-\infty}^{\infty} F(x)\left[F(x)+\frac{1}{2}h^2f'(x)\mu_2(k)+\frac{1}{24}h^4f'''(x)\mu_4(k)+O(h^5)\right] f(x)dx \\
&= \int_{-\infty}^{\infty} F^2(x)dF(x)+\frac{1}{2}h^2\mu_2(k)\int_{-\infty}^{\infty} F(x)f'(x)dF(x) \\
&\quad +\frac{1}{24}h^4\mu_4(k)\int_{-\infty}^{\infty} F(x)f'''(x)dF(x)+O(h^5) \\
&= \frac{1}{3}+\frac{1}{2}h^2\mu_2(k)E_1(F)+\frac{1}{24}h^4\mu_4(k)E_3(F)+O(h^5). \qquad (4.17)
\end{aligned}
$$

Substituting the results obtained from (4.13), (4.15), (4.16) and (4.17) into (1.24) we obtain

$$
\begin{aligned}
&\mathrm{MASE}(h) \\
&= \frac{n-1}{n^2}\left\{\mathrm{E}\left[K\left(\frac{X_1-X_2}{h}\right)\right]+\mathrm{E}\left[K^2\left(\frac{X_1-X_2}{h}\right)\right]\right\} \\
&\quad +\frac{(n-1)(n-2)}{n^2}\mathrm{E}\left[K\left(\frac{X_1-X_2}{h}\right)K\left(\frac{X_1-X_3}{h}\right)\right] \\
&\quad -\frac{2(n-1)}{n}\mathrm{E}\left[F(X_1)K\left(\frac{X_1-X_2}{h}\right)\right]+\frac{1}{3}-\frac{1}{2n}+\frac{1}{4n^2}
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{(n-1)}{n^2}\left[\frac{1}{2} + \frac{1}{2}h^2\mu_2(k)A_1(F) + \frac{1}{24}h^4A_3(F) + O(h^5)\right] \\
&+ \frac{n-1}{n^2}\left[\frac{1}{2} - 2hD_1(K)A_0(F) + h^2D_2(K)A_1(F) - \frac{1}{3}h^3D_3(K)A_2(F)\right. \\
&\quad \left. + \frac{1}{12}h^4D_4(K)A_3(F) + O(h^5)\right] + \frac{1}{3} - \frac{1}{2n} + \frac{1}{4n^2} \\
&- \frac{2(n-1)}{n}\left[\frac{1}{3} + \frac{1}{2}h^2\mu_2(k)E_1(F) + \frac{1}{24}h^4\mu_4(k)E_3(F) + O(h^5)\right] \\
&+ \frac{(n-1)(n-2)}{n^2}\left[\frac{1}{3} + h^2\mu_2(k)E_1(F) + \frac{1}{4}h^4\mu_2^2(k)\int_{-\infty}^{\infty}\{f'(x)\}^2\,dF(x)\right. \\
&\quad \left. + \frac{1}{12}h^4\mu_4(k)E_3(F) + O(h^5)\right]. \qquad (4.18)
\end{aligned}
$$

Using the result in lemma 1, (4.18) simplifies to

$$
\begin{aligned}
&\text{MASE}(h) \\
&= \frac{1}{6n} - \frac{1}{12n^2} - 2\frac{h}{n}D_1(K)A_0(F) + \frac{1}{4}h^4\mu_2^2(k)\int_{-\infty}^{\infty}[f'(x)]^2\,dF(x) \\
&+ \frac{2h^2}{n}\left[\frac{1}{2}\mu_2(k)A_1(F) - \mu_2(k)E_1(F)\right] + O\left(\frac{h^3}{n}\right). \qquad (4.19)
\end{aligned}
$$

Remarks

1. Comparison of (3.4) to (4.19) leads to the conclusion that

$$\text{MASE}(h) = \text{MISE}(h) + O\left(\frac{1}{n^2} + \frac{h^2}{n}\right).$$

   The MASE(h) measure picks up an extra O($\frac{1}{n^2}$) term and the constant before the O($\frac{h^2}{n}$) term becomes twice as large.

2. Under the usual assumptions, it follows that the asymptotic optimal bandwidth determined from MASE(h) will be as in (3.11).

3. From the analysis above it can be concluded that the MASE(h) and MISE(h) measures are asymptotically equivalent. In the next section, a bootstrap *estimator for MASE(h)* will be introduced.

4.4 A Bootstrap Estimator for MASE

Consider an unknown distribution function F that we wish to estimate with $F_{n,h}$. One way to measure the discrepancy between F and $F_{n,h}$ is the average squared error (ASE) or

$$\mathrm{ASE}(h) = \frac{1}{n}\sum_{i=1}^{n}\left[F_{n,h}(X_i) - F(X_i)\right]^2.$$

However, this is still a random variable. Taking the expected value of this function the mean average squared error (MASE) or

$$\mathrm{MASE}(h) = \mathrm{E}\left\{\frac{1}{n}\sum_{i=1}^{n}\left[F_{n,h}(X_i) - F(X_i)\right]^2\right\} \tag{4.20}$$

is obtained. For more details on ASE and MASE, see Section 1.2.

Applying the plug-in principle from (2.1) to (4.20), using the definition of $F_{n,h}$ in (1.2) the bootstrap estimator of $\mathrm{MASE}(h)$ becomes:

$$\begin{aligned}
&\widehat{\mathrm{MASE}}(h)\\
&= \mathrm{E}_*\left[\frac{1}{n}\sum_{i=1}^{n}\left\{F_{n,h}(X_i^*) - F_n(X_i^*)\right\}^2\right]\\
&= \mathrm{E}_*\left[\frac{1}{n}\sum_{i=1}^{n}\left\{\frac{1}{n}\sum_{j=1}^{n}K\left(\frac{X_i^* - X_j^*}{h}\right) - F_n(X_i^*)\right\}^2\right]\\
&= \mathrm{E}_*\left[\left\{\frac{1}{n}\sum_{j=1}^{n}K\left(\frac{X_i^* - X_j^*}{h}\right) - F_n(X_i^*)\right\}^2\right]\\
&= \mathrm{E}_*\left[\left\{\frac{1}{n}\sum_{i\neq j}^{n}K\left(\frac{X_i^* - X_j^*}{h}\right) + \frac{1}{n}K(0) - F_n(X_i^*)\right\}^2\right]\\
&= \mathrm{E}_*\left[\left\{\frac{1}{n}\sum_{i\neq j}^{n}\left[K\left(\frac{X_i^* - X_j^*}{h}\right) + \frac{1}{n-1}K(0) - \frac{n}{n-1}F_n(X_i^*)\right]\right\}^2\right]\\
&= \mathrm{E}_*\left[\mathrm{E}_*\left\{\frac{1}{n}\sum_{i\neq j}^{n}\left[K\left(\frac{X_i^* - X_j^*}{h}\right) + \frac{1}{n-1}K(0) - \frac{n}{n-1}F_n(X_i)\right]\right\}^2\middle| X_i^*\right].
\end{aligned} \tag{4.21}$$

Now let $X_i^* = x$ and define

$$\Psi_x(X_j^*) = K\left(\frac{x - X_j^*}{h}\right) + \frac{K(0)}{n-1} - \frac{n}{n-1}F_n(x). \tag{4.22}$$

Using $\mathrm{E}(X^2) = \mathrm{Var}(X) + [\mathrm{E}(X)]^2$, where X is any random variable, the inner expected

value in (4.21) can be written as follows

$$
\begin{aligned}
&\mathrm{E}_*\left[\left\{\frac{1}{n}\sum_{i\neq j}^{n}\left[K\left(\frac{x-X_j^*}{h}\right)+\frac{1}{n-1}K(0)-\frac{n}{n-1}F_n(x)\right]\right\}^2\right]\\
&=\mathrm{E}_*\left[\left\{\frac{1}{n}\sum_{i\neq j}^{n}\Psi_x(X_j^*)\right\}^2\right]\\
&=\mathrm{E}_*\left[\left\{\frac{n-1}{n}\cdot\frac{1}{n-1}\sum_{i\neq j}^{n}\Psi_x(X_j^*)\right\}^2\right]\\
&=\frac{(n-1)^2}{n^2}\,\mathrm{E}_*\left[\left\{\frac{1}{n-1}\sum_{i\neq j}^{n}\Psi_x(X_j^*)\right\}^2\right]\\
&=\frac{(n-1)^2}{n^2}\left\{\frac{\mathrm{Var}_*\left[\Psi_x(X_1^*)\right]}{n-1}+\left\{\mathrm{E}_*\left[\Psi_x(X_1^*)\right]\right\}^2\right\}\\
&=\frac{(n-1)^2}{n^2}\left\{\frac{\mathrm{E}_*\left[\left\{\Psi_x(X_1^*)\right\}^2\right]}{n-1}-\frac{\left\{\mathrm{E}_*\left[\Psi_x(X_1^*)\right]\right\}^2}{n-1}+\left\{\mathrm{E}_*\left[\Psi_x(X_1^*)\right]\right\}^2\right\}\\
&=\frac{n-1}{n^2}\,\mathrm{E}_*\left[\left\{\Psi_x(X_1^*)\right\}^2\right]-\frac{n-1}{n^2}\left\{\mathrm{E}_*\left[\Psi_x(X_1^*)\right]\right\}^2+\frac{(n-1)^2}{n^2}\left\{\mathrm{E}_*\left[\Psi_x(X_1^*)\right]\right\}^2\\
&=\frac{n-1}{n^2}\,\mathrm{E}_*\left[\left\{\Psi_x(X_1^*)\right\}^2\right]+\frac{(n-1)(n-2)}{n^2}\left\{\mathrm{E}_*\left[\Psi_x(X_1^*)\right]\right\}^2\\
&=\frac{n-1}{n^2}\cdot\frac{1}{n}\sum_{j=1}^{n}\Psi_x^2(X_j)+\frac{(n-1)(n-2)}{n^2}\left[\frac{1}{n}\sum_{j=1}^{n}\Psi_x(X_j)\right]^2. \qquad (4.23)
\end{aligned}
$$

Substituting (4.22) and (4.23) into (4.21)

$$
\begin{aligned}
&\widehat{\mathrm{MASE}}(h)\\
&=\mathrm{E}_*\left[\frac{1}{n}\sum_{i=1}^{n}\left\{\frac{1}{n}\sum_{j=1}^{n}K\left(\frac{X_i^*-X_j^*}{h}\right)-F_n(X_i^*)\right\}^2\right]\\
&=\frac{n-1}{n^3}\sum_{j=1}^{n}\mathrm{E}_*\left[\Psi_{X_i^*}^2(X_j)\right]+\frac{(n-1)(n-2)}{n^2}\,\mathrm{E}_*\left[\left\{\frac{1}{n}\sum_{j=1}^{n}\Psi_{X_i^*}(X_j)\right\}^2\right]\\
&=\frac{n-1}{n^3}\cdot\frac{1}{n}\sum_{j=1}^{n}\sum_{i=1}^{n}\Psi_{X_i}^2(X_j)+\frac{(n-1)(n-2)}{n^2}\cdot\frac{1}{n}\sum_{i=1}^{n}\left\{\frac{1}{n}\sum_{j=1}^{n}\Psi_{X_i}(X_j)\right\}^2\\
&=\frac{n-1}{n^4}\sum_{j=1}^{n}\sum_{i=1}^{n}\Psi_{X_i}^2(X_j)+\frac{(n-1)(n-2)}{n^5}\sum_{i=1}^{n}\left\{\sum_{j=1}^{n}\Psi_{X_i}(X_j)\right\}^2\\
&=\frac{n-1}{n^4}\sum_{j=1}^{n}\sum_{i=1}^{n}\left[K\left(\frac{X_i-X_j}{h}\right)+\frac{K(0)}{n-1}-\frac{n}{n-1}F_n(X_i)\right]^2\\
&+\frac{(n-1)(n-2)}{n^5}\sum_{i=1}^{n}\left\{\sum_{j=1}^{n}\left[K\left(\frac{X_i-X_j}{h}\right)+\frac{K(0)}{n-1}-\frac{n}{n-1}F_n(X_i)\right]\right\}^2. \qquad (4.24)
\end{aligned}
$$

Replacing $X_1, X_2, \ldots, X_n$ with the order statistics $X_{(1)}, X_{(2)} \ldots, X_{(n)}$ and using the property

that $F_n\left(X_{(i)}\right) = \frac{i}{n}$ the bootstrap estimator of MASE(h) becomes:

$$
\begin{aligned}
&\widehat{\text{MASE}}(h) \\
&= \frac{n-1}{n^4}\sum_{j=1}^{n}\sum_{i=1}^{n}\left[K\left(\frac{X_{(i)}-X_{(j)}}{h}\right)+\frac{K(0)}{n-1}-\frac{n}{n-1}F_n\left(X_{(i)}\right)\right]^2 \\
&+\frac{(n-1)(n-2)}{n^5}\sum_{i=1}^{n}\left\{\sum_{j=1}^{n}\left[K\left(\frac{X_{(i)}-X_{(j)}}{h}\right)+\frac{K(0)}{n-1}-\frac{n}{n-1}F_n\left(X_{(i)}\right)\right]\right\}^2 \\
&= \frac{n-1}{n^4}\sum_{j=1}^{n}\sum_{i=1}^{n}\left[K\left(\frac{X_{(i)}-X_{(j)}}{h}\right)+\frac{K(0)}{n-1}-\frac{i}{n-1}\right]^2 \\
&+\frac{(n-1)(n-2)}{n^5}\sum_{j=1}^{n}\left\{\sum_{i=1}^{n}\left[K\left(\frac{X_{(i)}-X_{(j)}}{h}\right)+\frac{1}{n-1}K(0)-\frac{j}{n-1}\right]\right\}^2 \\
&= \frac{n-1}{n^4}\sum_{j=1}^{n}\sum_{i=1}^{n}\left[1-K\left(\frac{X_{(j)}-X_{(i)}}{h}\right)-\frac{1}{n-1}K(0)+\frac{i}{n-1}\right]^2 \\
&+\frac{(n-1)(n-2)}{n^5}\sum_{j=1}^{n}\left\{\sum_{i=1}^{n}\left[K\left(\frac{X_{(i)}-X_{(j)}}{h}\right)+\frac{1}{n-1}K(0)-\frac{j}{n-1}\right]\right\}^2 \\
&= \frac{n-1}{n^4}\sum_{j=1}^{n}\sum_{i=1}^{n}\left[K\left(\frac{X_{(j)}-X_{(i)}}{h}\right)-\frac{K(0)}{n-1}+\frac{i-n+1}{n-1}\right]^2 \\
&+\frac{(n-1)(n-2)}{n^5}\sum_{j=1}^{n}\left\{\sum_{i=1}^{n}\left[K\left(\frac{X_{(i)}-X_{(j)}}{h}\right)+\frac{K(0)}{n-1}-\frac{j}{n-1}\right]\right\}^2 \\
&= C'_n\sum_{j=1}^{n}\sum_{i=1}^{n}\left[K\left(\frac{X_{(j)}-X_{(i)}}{h}\right)-\left(1-\frac{\left(i-\frac{1}{2}\right)}{n-1}\right)\right]^2 \\
&+D'_n\sum_{j=1}^{n}\left\{\sum_{i=1}^{n}\left[K\left(\frac{X_{(i)}-X_{(j)}}{h}\right)-\frac{\left(j-\frac{1}{2}\right)}{n-1}\right]\right\}^2, \qquad (4.25)
\end{aligned}
$$

since $K(x)$ is symmetrical about 0 and where

$$C'_n = \frac{n-1}{n^4} \text{ and } D'_n = \frac{(n-1)(n-2)}{n^5}.$$

The expression in (4.25) can now be used to estimate MASE(h) and when minimised with respect to h, should yield an estimate of the optimal bandwidth.

Remark

Van Graan (1983) proposed an estimator for MISE(h) which was first given in (3.2), i.e.,

$$
\begin{aligned}
J(F_n,h) &= C_n\sum_{j=1}^{n}\sum_{i=1}^{n}\left[K(\frac{X_{(j)}-X_{(i)}}{h})-\frac{j}{n}\right]^2 \\
&+D_n\sum_{j=1}^{n}\left[\sum_{i=1}^{n}\left\{K(\frac{X_{(j)}-X_{(i)}}{h})-\frac{j}{n}\right\}\right]^2,
\end{aligned}
$$

Where $C_n = \frac{1}{n^3}$ and $D_n = \frac{n-1}{n^4}$. It can be seen that (3.2) and (4.25) are similar in form. Noting the results in the last section, it can be expected that the large sample behaviour of (3.2) and (4.25) will be the same.

Chapter 5

Simulation Study

5.1 Introduction

A simulation study was performed to study the small to medium sample behaviour of the three bandwidth selectors:

- The bandwidth selector defined in (4.25) was derived by proposing a bootstrap estimator for the MASE discrepancy measure.
- The bandwidth selector defined in (3.2) that was derived by proposing a bootstrap estimator for the MISE discrepancy measure.
- The plug-in bandwidth selector obtained by estimating the asymptotical optimal bandwidth defined in (3.11). The procedure of Altman and Léger (1995) in its improved form (see Section 3.3.2) will be used in the simulation study.

Unless mentioned otherwise, the bandwidth selectors defined in (3.2) and (4.25), were calculated as follows: the normal distribution function $K(x) = \Phi(x),\ -\infty < x < \infty$ was used to compute the kernel distribution function estimator and the plug-in bandwidth selector was implemented as discussed in section 3.3.2.

Samples from four distributions were used in the simulation study. These were: standard normal distribution, exponential distribution, skewed unimodal distribution #2 and the asymmetric bimodal distribution #8. The last two distributions are from the examples of normal mixtures given in Marron and Wand (1992). The cumulative distribution functions are listed in Table 5.1

Name	Distribution Function
Standard normal distribution	$N(0,1)$
Standard exponential distribution	$\text{Exp}(1)$
Skewed unimodal distribution #2	$\frac{1}{5}N(0,1)+\frac{1}{5}N(\frac{1}{2},\frac{4}{9})+\frac{3}{5}N(\frac{13}{12},\frac{25}{81})$
Skewed bimodal distribution #8	$\frac{3}{4}N(0,1)+\frac{1}{4}N(\frac{3}{2},\frac{1}{9})$

Table 5.1: Distribution functions used in the simulation study. Plots of normal mixtures densities are in Marron and Wand (1992).

To generate observations from a normal mixture of the form $(1-\epsilon)N(\mu_1,\sigma_1^2)+\epsilon N(\mu_2,\sigma_2^2)$, the following procedure was followed: Generate an uniformly distributed random number between 0 and 1. If the number is less than $(1-\epsilon)$ generate a random variable from $N(\mu_1,\sigma_1^2)$, otherwise generate a random variable from $N(\mu_2,\sigma_2^2)$. This procedure can also be extended to include more complicated normal mixtures like the skewed unimodal distribution #2.

In Section 5.2 it will be discussed how the optimal bandwidths for the distributions studied in this chapter were obtained.

Section 5.3 describes how the asymptotical optimal bandwidth was obtained and gives the values of the asymptotical optimal bandwidth for the distributions and sample sizes studied.

Since two of the procedure are based on estimating MISE(h) or MASE(h), the efficiency of the chosen bandwidth will depend on how well the procedures estimates MISE(h) or MASE(h). In Section 5.4 this aspect will be considered.

All that remains is to compare the different procedures to each other. This will be done in Section 5.5 where the bandwidths obtained from the simulation study, that will be described later in this chapter, and the discrepancy measures associated with the bandwidths will be compared for the different procedures, distribution functions and sample sizes.

5.2 Computation of the optimal bandwidth

The optimal bandwidth h_{MISE} is defined as the bandwidth minimising MISE(h) and h_{MASE} as the bandwidth that minimises MASE(h), i.e.,

$$h_{\text{MISE}} = \arg\min_h \text{MISE}(h), \tag{5.1}$$

and

$$h_{\text{MASE}} = \arg\min_h \text{MASE}(h). \tag{5.2}$$

The bandwidth selectors defined in (3.2) and (4.25) are aimed at finding estimators for (5.1) and (5.2). In Section 4.2 an exact theoretical calculation was performed to evaluate MASE(h) when both F and K are known. However, it is difficult to perform this calculation for more involved choices of F and K. Therefore, to obtain values for h_{MISE} and h_{MASE} for different choices of F, K and n a Monte Carlo simulation procedure is proposed to obtain approximations to h_{MISE} and h_{MASE}. The following algorithm was used to obtain approximations to the values h_{MISE} and h_{MASE}:

1. Let h_0 be a fixed value of the bandwidth parameter and $X_1, X_2, \ldots, X_n$ i.i.d. random variables from a certain distribution F. Let $F_{n,h_0,i}(x)$ be the non-parametric kernel distribution function estimator calculated for the i^{th} Monte Carlo sample, $i = 1, \ldots, \text{MC}$. Define

$$\widehat{\text{ISE}}_{\text{i}}(h_0) = \int_{-\infty}^{\infty} [F_{n,h_0,i}(x) - F(x)]^2 \, dF(x),$$

and

$$\widehat{\text{ASE}}_{\text{i}}(h_0) = \sum_{j=1}^{n} [F_{n,h_0,i}(X_j) - F(X_j)]^2 .$$

2. The Monte Carlo estimates of MISE and MASE at h_0 can be calculated as

$$\widehat{\text{MISE}}(h_0) = \frac{1}{\text{MC}} \sum_{i=1}^{MC} \widehat{\text{ISE}}_i(h_0),$$

and

$$\widehat{\text{MASE}}(h_0) = \frac{1}{\text{MC}} \sum_{i=1}^{MC} \widehat{\text{ASE}}_i(h_0).$$

3. The calculations in the previous steps were performed for a fixed value of the bandwidth parameter, i.e., $h = h_0$. The calculations can be performed for a range of values of h, say $h \in [0.01, 0.02, \ldots, r]$ where r depends on the sample size. Monte Carlo estimates of h_{MISE} and h_{MASE} can be obtained as

$$\tilde{h}_{\text{MISE}} = \arg\min_h \widehat{\text{MISE}}(h),$$

and

$$\tilde{h}_{\text{MASE}} = \arg\min_h \widehat{\text{MASE}}(h).$$

The bandwidths contained in Table 5.2 are the Monte Carlo estimated bandwidths that minimises MISE(h) and MASE(h) for the distributions in Table 5.1 and the normal kernel. The number of Monte Carlo trials were set at $MC = 500$.

	n	Normal	Exponential	Skewed Unimodal	Skewed Bimodal
$\tilde{h}_{\text{MISE}}$	20	0.63	0.33	0.42	0.80
	100	0.34	0.13	0.25	0.40
	200	0.28	0.09	0.20	0.28
	500	0.20	0.06	0.14	0.20
$\tilde{h}_{\text{MASE}}$	20	0.54	0.27	0.36	0.61
	100	0.32	0.12	0.24	0.32
	200	0.27	0.09	0.18	0.23
	500	0.20	0.06	0.14	0.17

Table 5.2: Optimal bandwidths

Remarks

1. From Table 5.2 it follows that $\tilde{h}_{\text{MASE}} \leq \tilde{h}_{\text{MISE}}$ for the combinations of sample sizes and distribution functions considered.

2. Furthermore, from Table 5.2 it is clear that both of $\tilde{h}_{\text{MISE}}$ and $\tilde{h}_{\text{MASE}}$ tends to zero as n tends to infinity. Also, $\tilde{h}_{\text{MASE}} \to \tilde{h}_{\text{MISE}}$ as $n \to \infty$.

5.3 Asymptotical optimal bandwidth

The direct plug-in estimate in (3.30) estimates the asymptotical optimal bandwidth and for comparison purposes the asymptotical optimal bandwidth needs to be calculated for the distributions contained in the simulation study.

It was shown in (3.11) that the asymptotical optimal bandwidth h_{opt} can be written as

$$h_{\text{opt}} = \left[\frac{V_2}{4B_3}\right]^{1/3} n^{-1/3},$$

where B_3 and V_2 were defined in (3.5) and (3.6) respectively. Note that since $f'(x) = 0$ for the uniform distribution, the asymptotical optimal bandwidth cannot be computed using (3.11).

In the case of the normal distribution, it was shown by Van Graan (1983) that

$$V_2 = \frac{1}{2\pi},$$

and

$$B_3 = \frac{1}{24\sqrt{3}\pi},$$

which results in the asymptotical optimal bandwidth

$$h_{\text{opt}} = 1.7321 n^{-\frac{1}{3}}.$$

In the case of the exponential distribution analytical calculations yield

$$V_2 = \frac{1}{2\sqrt{\pi}},$$

and

$$B_3 = \frac{1}{12},$$

which results in the asymptotical optimal bandwidth

$$h_{\text{opt}} = 0.9459 n^{-\frac{1}{3}}.$$

Values of V_2 and B_3 for the skewed unimodal and skewed bimodal distributions were calculated using numerical integration techniques. Values of V_2 and B_3 for the distributions considered in the simulation study are summarised in Table 5.3. The asymptotical optimal bandwidths calculated with (3.11) using the results in Table 5.3 are summarised in Table 5.4.

Distribution	V_2	B_3
Normal	0.1592	0.0077
Exponential	0.2821	0.0833
Skewed Unimodal	0.0867	0.0025
Skewed Bimodal	0.1025	0.0033

Table 5.3: The values of V_2 and B_3

Remarks

1. In the case of a normal population distribution, there are noticeable similarities between the optimal bandwidth $\tilde{h}_{\text{MISE}}$ in Table 5.2 and the asymptotical bandwidth h_{opt} in Table 5.4 for the different sample sizes.

n	Normal	Exponential	Skewed unimodal	Skewed Bimodal
20	0.64	0.35	0.76	0.73
100	0.37	0.20	0.44	0.43
200	0.30	0.16	0.35	0.34
500	0.22	0.12	0.26	0.25

Table 5.4: Asymptotical optimal bandwidths

2. It is reasonable to expect that the asymptotical optimal bandwidth, h_{opt}, and the optimal bandwidth determined from the MISE criterion, h_{MISE}, should be relatively close to each other for large sample sizes. Inspection of Tables 5.2 and 5.4 reveals that sample sizes larger than 500 will be necessary to observe this property in the case of the skewed distributions.

5.4 Estimation of the measures MISE and MASE

In section 5.2 computation of the optimal bandwidth was considered. If the true distribution F is unknown, then the values of h_{MISE} and h_{MASE} are also unknown.

The measures MISE(h) and MASE(h) cannot be computed since F is unknown. Therefore (3.2) and (4.25) can be used to estimate the unknown MISE(h) and MASE(h). If the bandwidth selection procedures are to be good procedures in the sense that the selected bandwidths should be in the region of (5.1) or (5.2), then the expected values of (3.2) and (4.25) should also be similar in form to MISE(h) and MASE(h) in the region of (5.1) and (5.2).

In the case of an uniform F and K, exact expressions for MASE(h) and MISE(h) are available (see (4.5) and (4.6) respectively). Note that analytical expressions for MASE(h) and MISE(h) are available only for the case mentioned previously. Approximations for other combinations of F and K can be obtained using the algorithm described in section 5.2. Note that the optimal bandwidths in section 5.2 were obtained from approximations of MASE(h) and MISE(h) for the distributions considered there and a normal kernel. Denote (4.25) by $J'(F_n, h)$, then the following algorithm can be employed to obtain an approximation to the values of $\mathrm{E}\left[J(F_n, h)\right]$ and $\mathrm{E}\left[J'(F_n, h)\right]$:

1. Let h_0 be a fixed value of the bandwidth parameter and $X_1, X_2, \ldots, X_n$ i.i.d. random

variables from a certain distribution F, with associated order statistics $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$. Define $J_i(F_n, h_0)$ and $J_i'(F_n, h_0)$ to be (3.2) and (4.25) calculated for the i^{th} Monte Carlo sample, $i = 1, ..., \text{MC}$ and $h = h_0$.

2. The Monte Carlo estimates of $\text{E}\,[J(F_n, h_0)]$ and $\text{E}\,[J'(F_n, h_0)]$ are then given by

$$\text{E}\,[\widehat{J(F_n, h_0)}] = \frac{1}{\text{MC}} \sum_{i=1}^{\text{MC}} J_i(F_n, h_0),$$

and

$$\text{E}\,[\widehat{J'(F_n, h_0)}] = \frac{1}{\text{MC}} \sum_{i=1}^{\text{MC}} J_i'(F_n, h_0).$$

3. It follows that $\text{E}\,[\widehat{J(F_n, h)}]$ and $\text{E}\,[\widehat{J'(F_n, h)}]$ can be calculated for a range of values $h = h_0$ (see section 5.2). Now define

$$h'_{\text{MISE}} = \arg\min_h \text{E}\,[\widehat{J(F_n, h)}],$$

and

$$h'_{\text{MASE}} = \arg\min_h \text{E}\,[\widehat{J'(F_n, h)}].$$

For the case of a uniform population and a uniform kernel, estimates of $\text{E}\,[J(F_n, h)]$ and $\text{E}\,[J'(F_n, h)]$ are shown in Figures 5.1 to 5.4 for different choices of the sample size. The bandwidths minimising these functions are summarised in Table 5.5. For the case of a normal population and a normal kernel, the same estimates were calculated for the same choices of sample size. In both of the cases the number of Monte Carlo trials were set at $\text{MC} = 500$. The results are shown in Figures 5.5 to 5.8. The bandwidths minimising these functions are summarised in Tables 5.6 and 5.7.

Note that in Figures 5.1 to 5.8 the red line represents the MISE discrepancy measure (as a function of h), the blue line the MASE discrepancy measure (as a function of h), the green line the estimate of $\text{E}\,[J'(F_n, h)]$ and the black line the estimate of $\text{E}\,[J(F_n, h)]$.

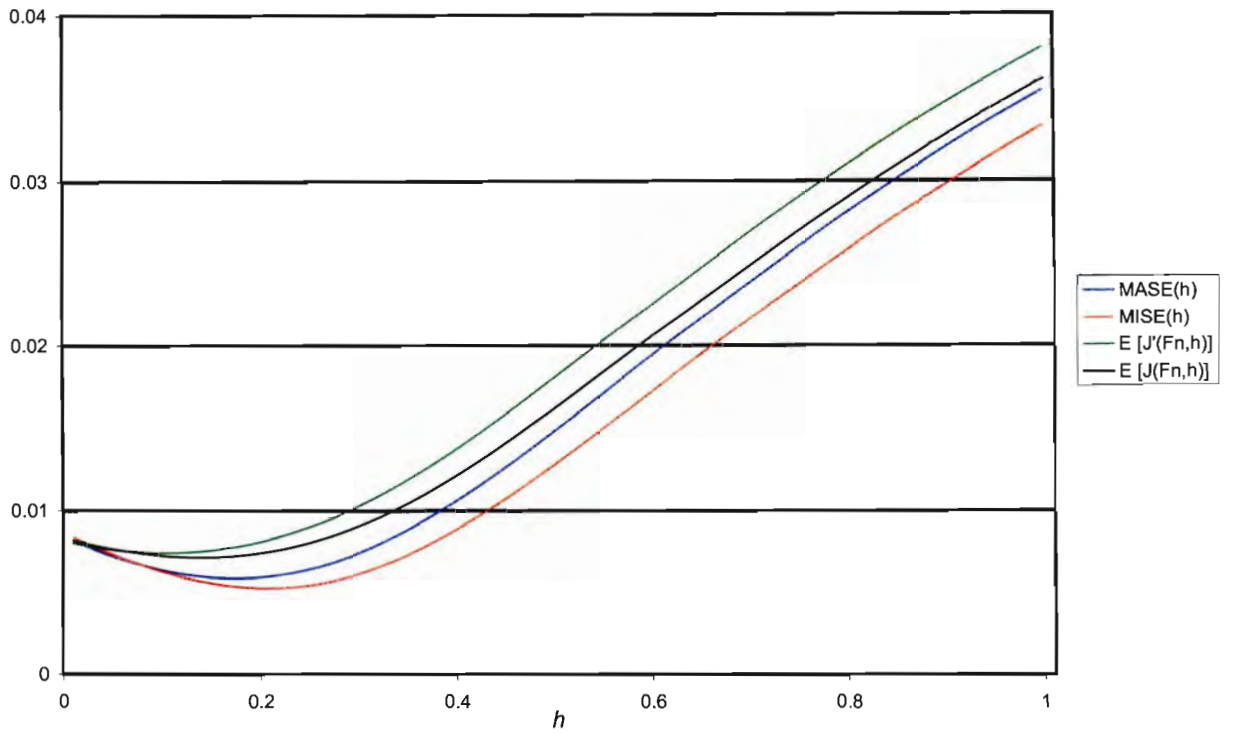


Figure 5.1: Comparison for n=20, uniform population, uniform kernel

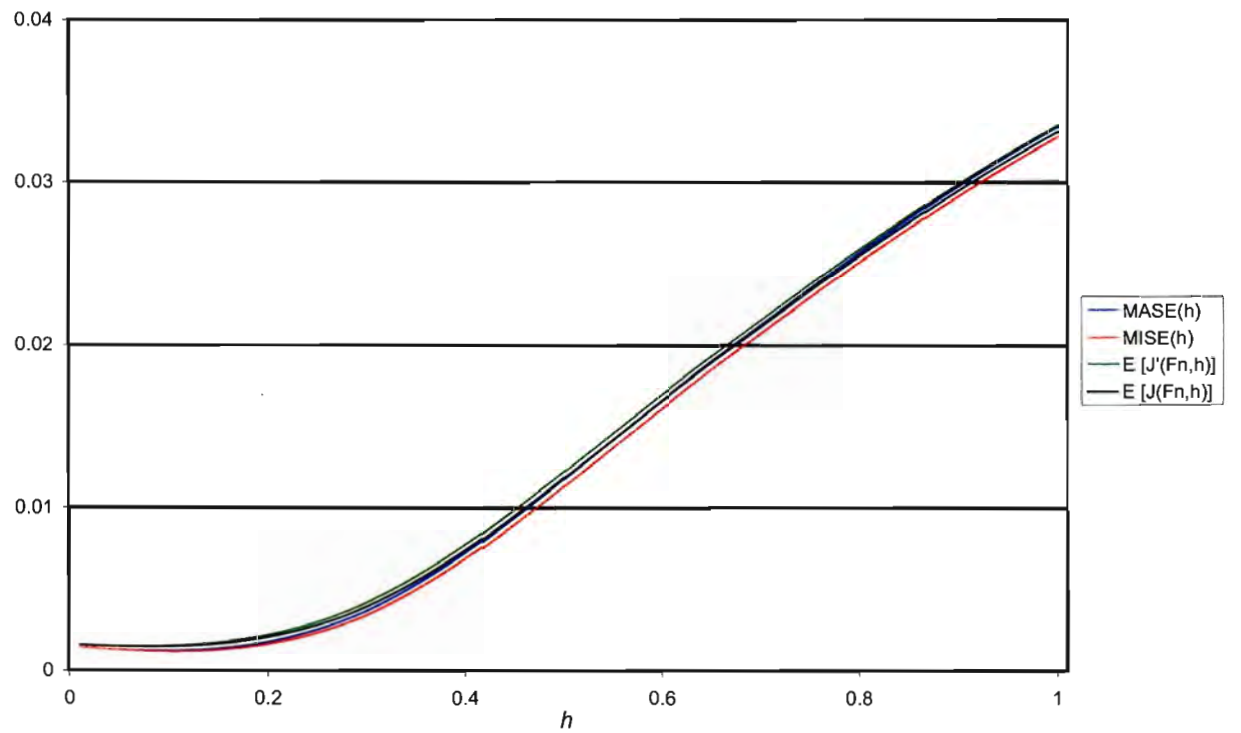


Figure 5.2: Comparison for n=100, uniform population, uniform kernel

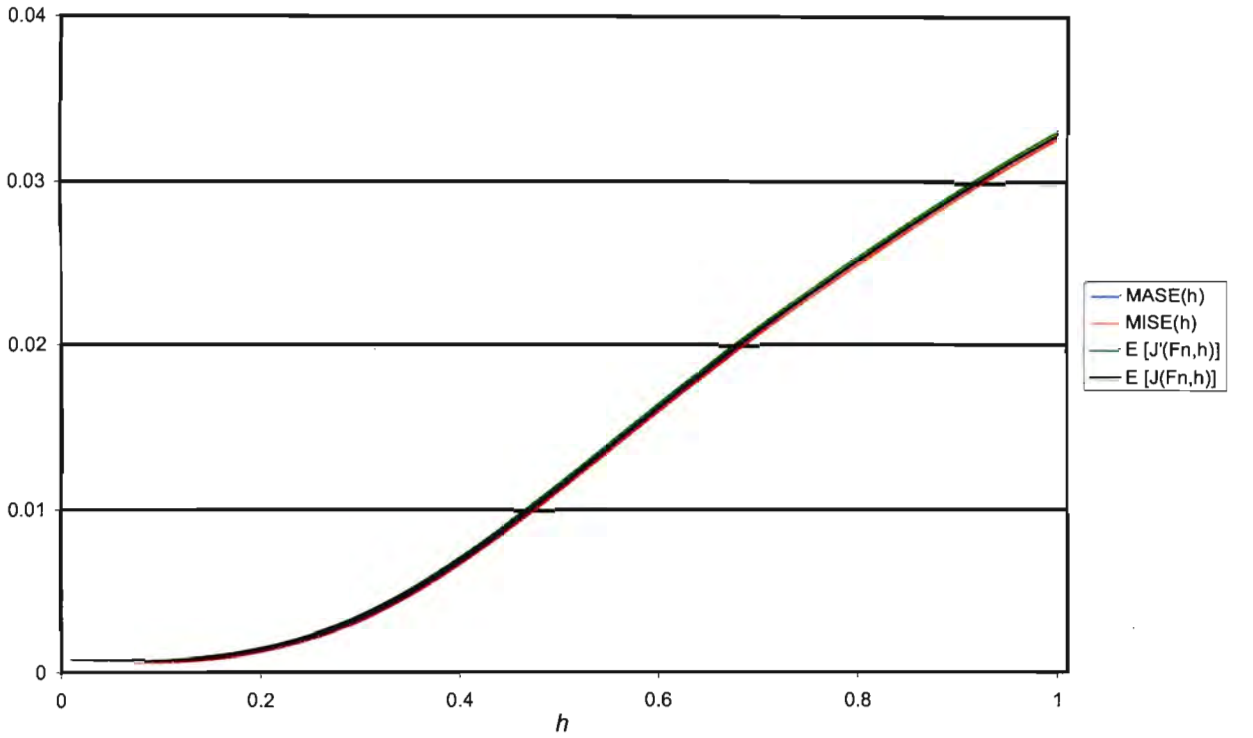


Figure 5.3: Comparison for n=200, uniform population, uniform kernel

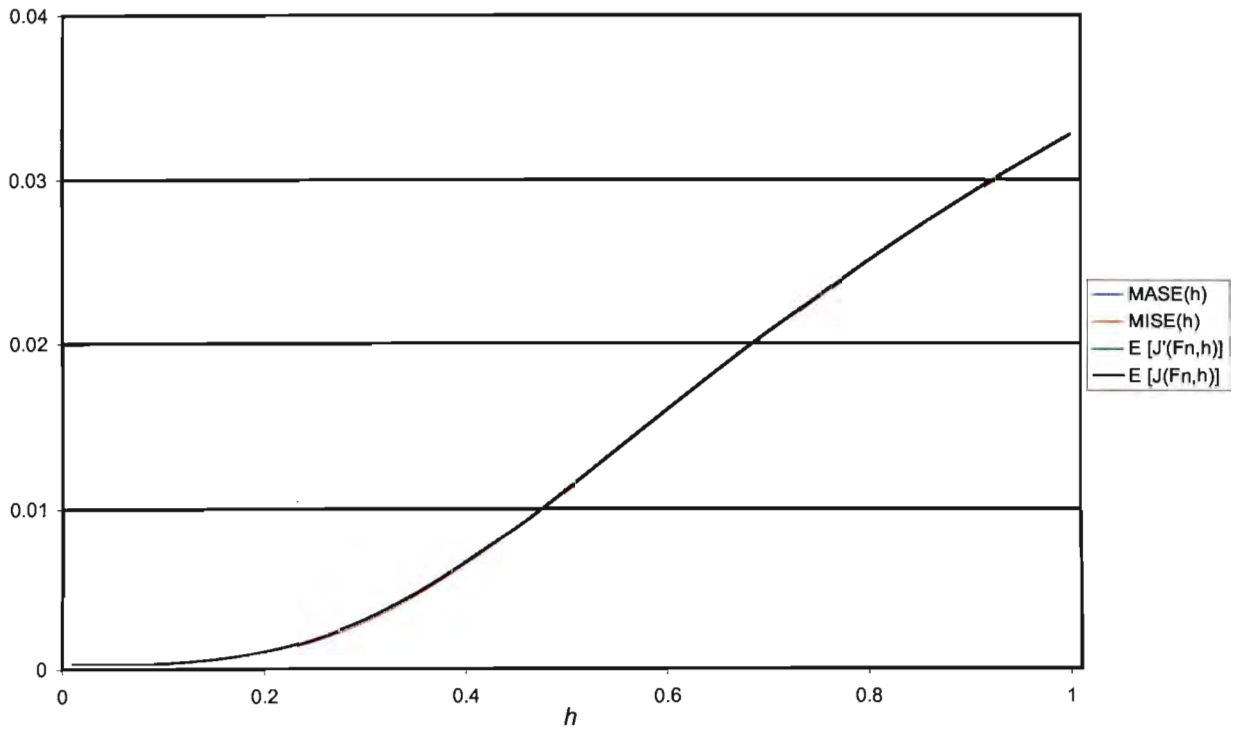


Figure 5.4: Comparison for n=500, uniform population, uniform kernel

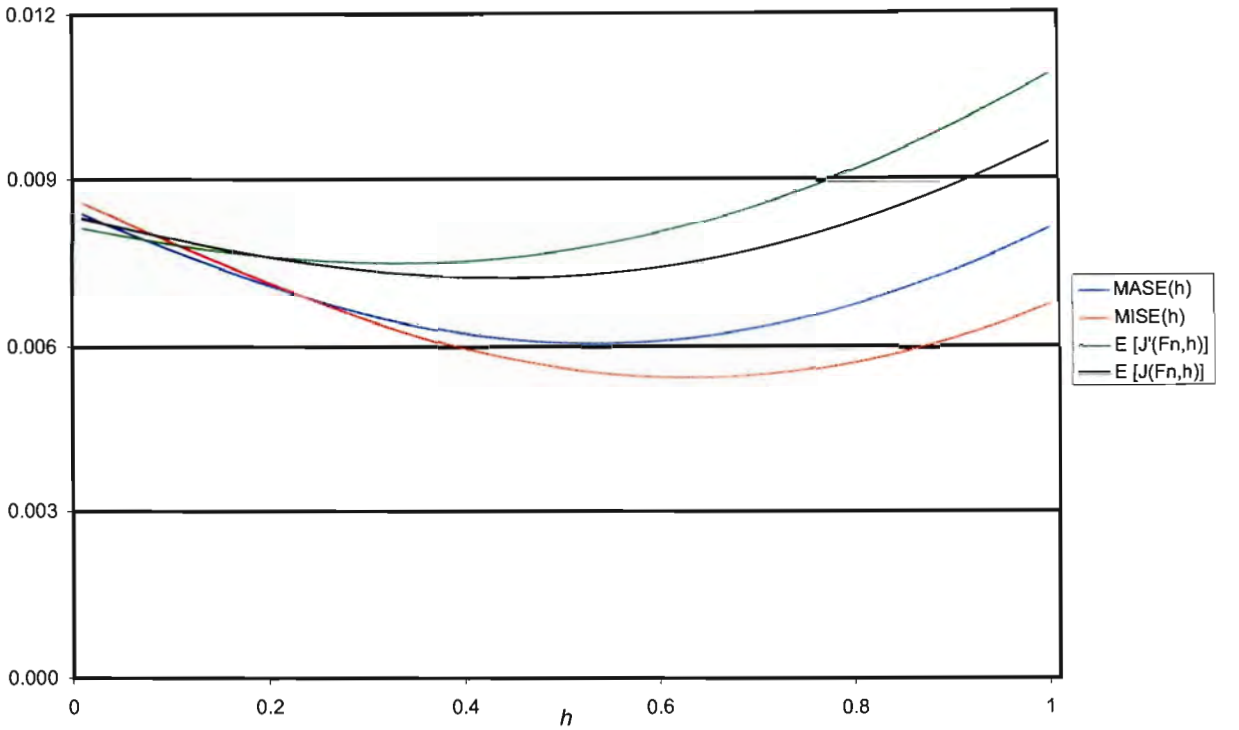


Figure 5.5: Comparison for n=20, normal population, normal kernel

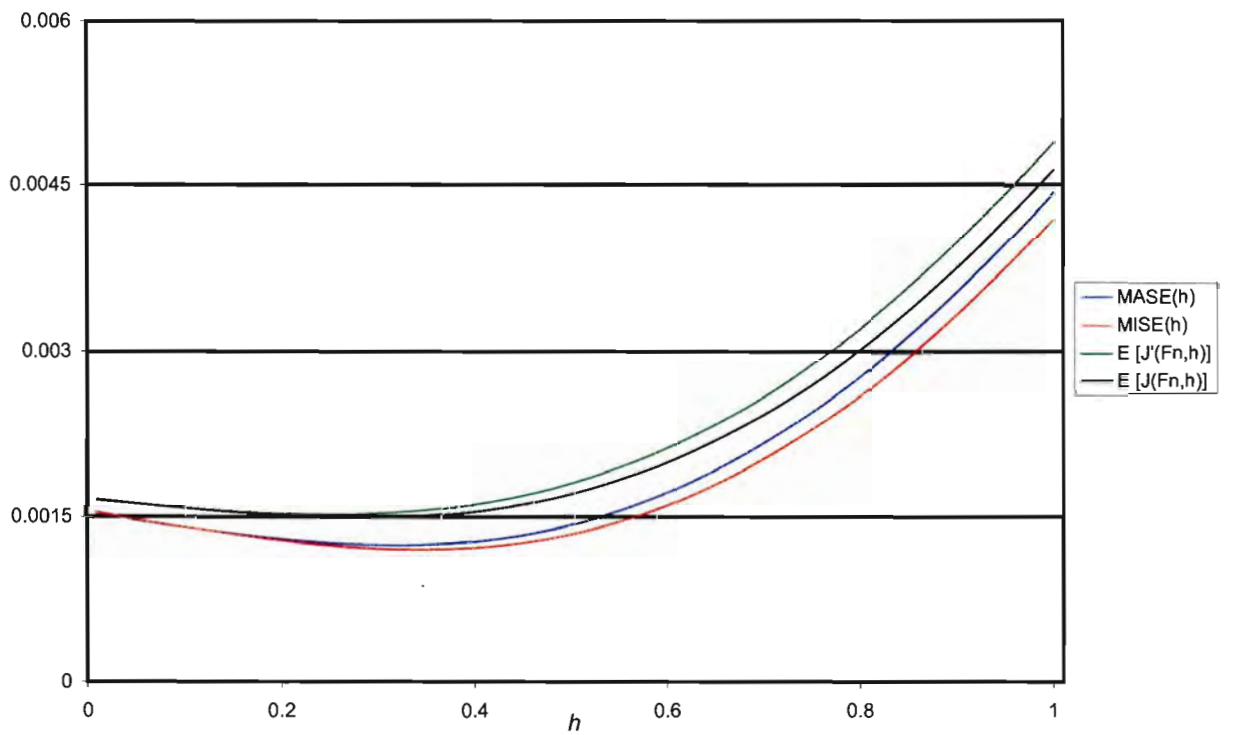


Figure 5.6: Comparison for n=100, normal population, normal kernel

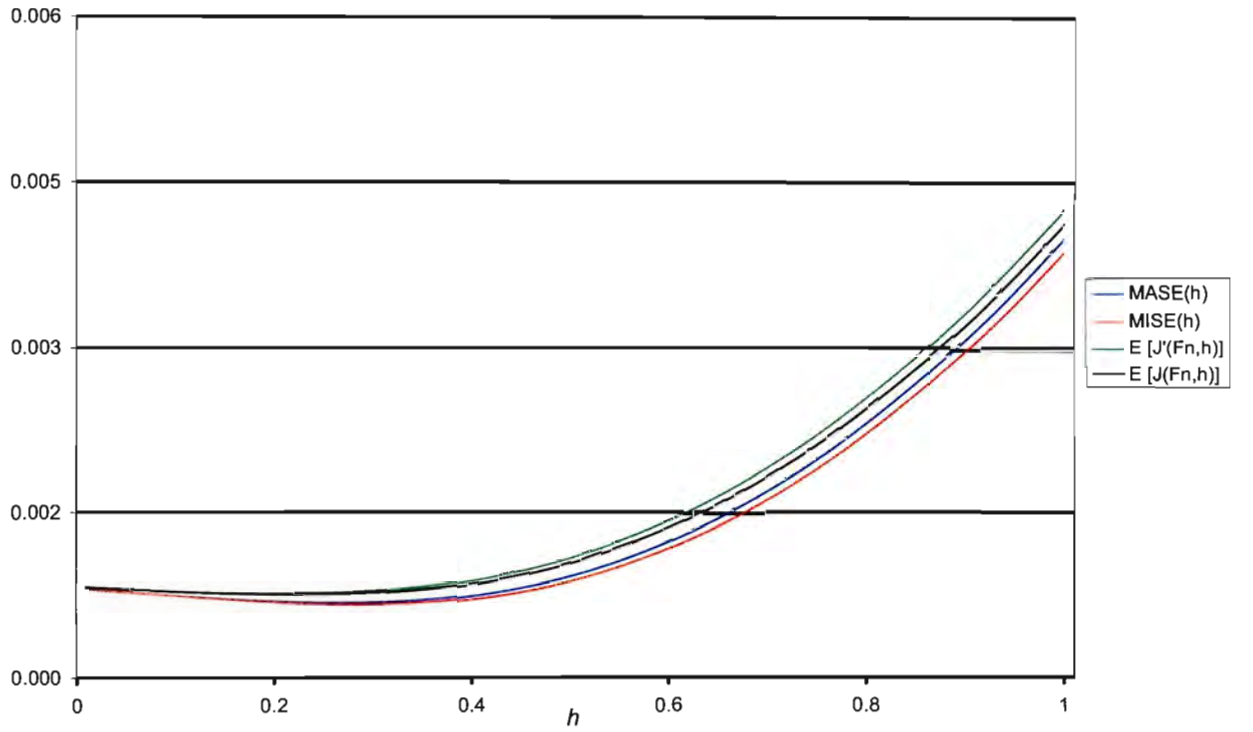


Figure 5.7: Comparison for n=200, normal population, normal kernel

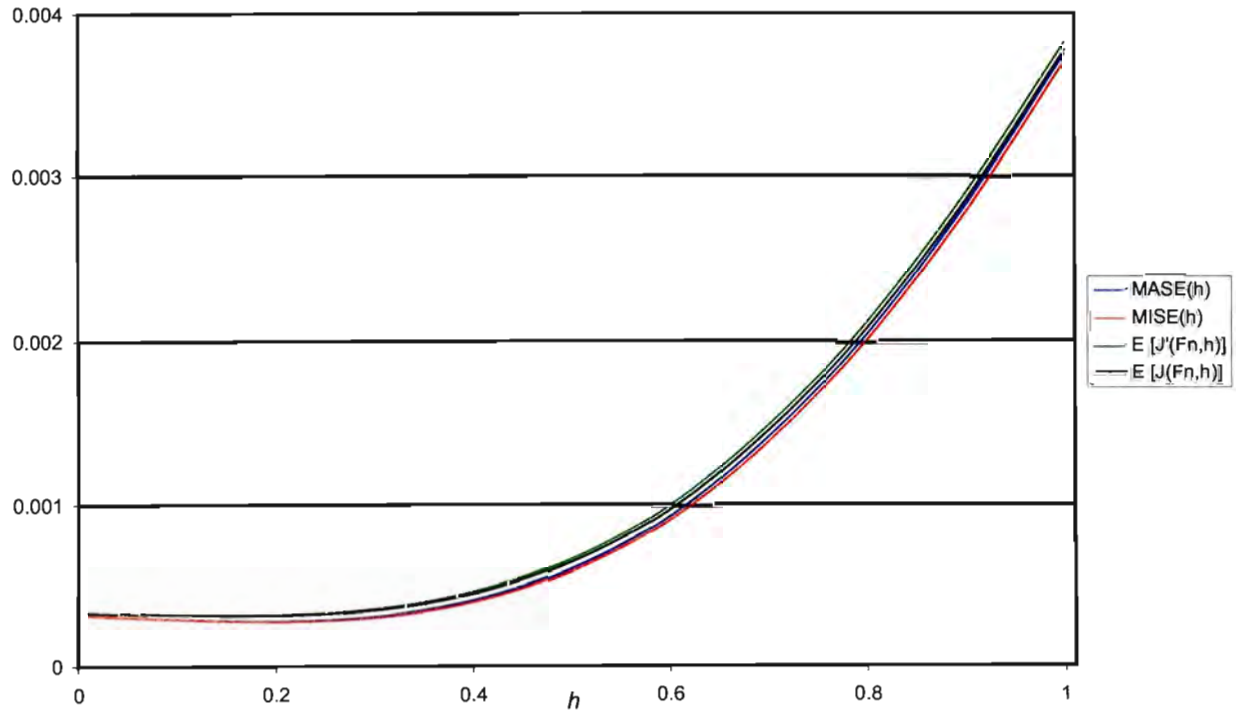


Figure 5.8: Comparison for n=500, normal population, normal kernel

n	h_{MISE}	h'_{MISE}	h_{MASE}	h'_{MASE}
20	0.20	0.14	0.17	0.10
100	0.10	0.08	0.09	0.07
200	0.08	0.06	0.07	0.05
500	0.05	0.04	0.05	0.04

Table 5.5: Uniform population, uniform kernel

	Normal		Exponential		Skewed Unimodal		Skewed Bimodal	
n	$\tilde{h}_{\text{MISE}}$	h'_{MISE}	$\tilde{h}_{\text{MISE}}$	h'_{MISE}	$\tilde{h}_{\text{MISE}}$	h'_{MISE}	$\tilde{h}_{\text{MISE}}$	h'_{MISE}
20	0.63	0.44	0.33	0.20	0.42	0.33	0.80	0.50
100	0.34	0.29	0.13	0.09	0.25	0.20	0.40	0.26
200	0.28	0.23	0.09	0.07	0.20	0.16	0.28	0.21
500	0.20	0.17	0.06	0.04	0.14	0.11	0.20	0.14

Table 5.6: $\tilde{h}_{\text{MISE}}$, h'_{MISE} with a normal kernel

	Normal		Exponential		Skewed Unimodal		Skewed Bimodal	
n	$\tilde{h}_{\text{MASE}}$	h'_{MASE}	$\tilde{h}_{\text{MASE}}$	h'_{MASE}	$\tilde{h}_{\text{MASE}}$	h'_{MASE}	$\tilde{h}_{\text{MASE}}$	h'_{MASE}
20	0.54	0.33	0.27	0.15	0.36	0.25	0.61	0.36
100	0.32	0.25	0.12	0.08	0.24	0.17	0.32	0.23
200	0.27	0.21	0.09	0.06	0.18	0.14	0.23	0.19
500	0.19	0.16	0.06	0.04	0.14	0.12	0.17	0.14

Table 5.7: $\tilde{h}_{\text{MASE}}$, h'_{MASE} with a normal kernel

Remarks

1. In the case of a uniform population and uniform kernel, inspection of Figures 5.1 to 5.4 reveal that $\mathrm{E}\left[J(F_n, h)\right]$ and $\mathrm{E}\left[J'(F_n, h)\right]$ tend to MISE(h) and MASE(h) respectively as n increases. It can be seen from Figure 5.4 the different lines can no longer be distinguished. This behaviour is also evident in Table 5.5 where $\tilde{h}_{\text{MISE}} \to h'_{\text{MISE}}$ and $\tilde{h}_{\text{MASE}} \to h'_{\text{MASE}}$ as n takes on larger values. Also, there seems to be a bias present in that $h'_{\text{MISE}} \leq \tilde{h}_{\text{MISE}}$ and $h'_{\text{MASE}} \leq \tilde{h}_{\text{MASE}}$ for the values of n considered there.

2. Inspection of Figures 5.5 to 5.8 and Tables 5.6 and 5.7 shows that the same behaviour continues to hold for the normal kernel and the other distributions and sample sizes considered there.

5.5 Comparison of methods

The final part of the simulation study is to compare the different procedures which were discussed in this study with each other. In order to compare the different procedures to each other a relevant criterion must be used. It was decided that a measure based on the average squared error

$$\text{ASE}(h) = \frac{1}{n}\sum_{i=1}^{n}\left[F_{n,h}(X_i) - F(X_i)\right]^2,$$

will be used (see also (1.15)). For any competing procedure, let $\hat{h}$ be the data-driven bandwidth obtained from the procedure for a random sample $X_1, X_2, \ldots, X_n$ from a distribution F. Let

$$\text{ASE}(\hat{h}) = \frac{1}{n}\sum_{i=1}^{n}\left[F_{n,\hat{h}}(X_i) - F(X_i)\right]^2. \tag{5.3}$$

The criterion that will be used is based on the mean value of the random variable defined in (5.3), i.e.,

$$\text{E}\left[\text{ASE}\left(\hat{h}\right)\right] = d_{\text{ASE}}. \tag{5.4}$$

The distributions considered here are those stated in Table 5.1 and the sample sizes are the same as that in Tables 5.2 and 5.4. The following algorithm, for which the Fortran 90 source code for the program can be found on the CD attached to the back cover of this dissertation, was employed to calculate the criterion in (5.4) and the related quantities:

1. Let $X_1, X_2, \ldots, X_n$ be an i.i.d. random sample from a certain distribution F, with associated order statistics $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$. Define $J_i(F_{n,h})$ and $J_i'(F_{n,h})$ to be (3.2) and (4.25) calculated for the i^{th} Monte Carlo sample, $i = 1, \ldots, \text{MC}$.

2. For the i^{th} Monte Carlo sample, let $\hat{h}_i$ denote that value of h which minimises the functions $J_i(F_{n,h})$ or $J_i'(F_{n,h})$ or let $\hat{h}_i$ denote the direct plug-in bandwidth obtained by using the algorithm in Section 3.3.2. The functions $J_i(F_{n,h})$ and $J_i'(F_{n,h})$ are calculated over a range of values of h, where $h \in [0.01, 0.02, \ldots, r]$ where r depends on the sample

size. The method of minimisation consisted of both a traditional as well as the golden section method.

3. For $i = 1, \ldots, \text{MC}$ and for each of the competing procedures, calculate

$$\text{ASE}(\hat{h}_i) = \frac{1}{n}\sum_{j=1}^{n}\left[F_{n,\hat{h}_i}(X_j) - F(X_j)\right]^2,$$

and

$$\hat{d}_{\text{ASE}} = \frac{1}{\text{MC}}\sum_{i=1}^{\text{MC}}\text{ASE}\left(\hat{h}_i\right), \tag{5.5}$$

where $\hat{d}_{\text{ASE}}$ represent the Monte Carlo estimate of the mean average squared error d_{ASE}. Also calculate

$$\overline{h} = \frac{1}{\text{MC}}\sum_{i=1}^{\text{MC}}\hat{h}_i, \tag{5.6}$$

the average bandwidth for each of the procedures. The Monte Carlo estimates of SE (standard error) for (5.5) and (5.6) are respectively

$$\widehat{\text{SE}\left(\hat{d}_{\text{ASE}}\right)} = \frac{1}{\sqrt{\text{MC}}}\sqrt{\frac{1}{\text{MC}-1}\sum_{i=1}^{MC}\left[\text{ASE}\left(\hat{h}_i\right) - \hat{d}_{\text{ASE}}\right]^2}, \tag{5.7}$$

and

$$\widehat{\text{SE}\left(\overline{h}\right)} = \frac{1}{\sqrt{\text{MC}}}\sqrt{\frac{1}{\text{MC}-1}\sum_{i=1}^{MC}\left[\hat{h}_i - \overline{h}\right]^2}. \tag{5.8}$$

In all cases the number of Monte Carlo simulations was chosen to be 1000. The following abbreviations will be used for the different procedures:

- New to indicate the procedure in (4.25)
- VG to indicate the procedure in (3.2)
- AL to indicate the direct plug-in bandwidth procedure in Section 3.3.2.

The results for the simulation study in this section will be shown in Appendixes A and B. As an example of the tables that will appear in Appedix A and the figures that will appear in Appendix B consider Table 5.8 and Figure 5.9.

In Table 5.8 the optimal bandwidth column refers to the values of $\tilde{h}_{\text{MASE}}$, $\tilde{h}_{\text{MISE}}$ and h_{opt} for the normal distribution when $n = 20$. The column containing $\overline{h}$ refers to (5.6), and the next

Procedure	Optimal bandwidth	$\overline{h}$	SE$\times 10^{-2}$	$\hat{d}_{\mathrm{ASE}} \times 10^{-2}$	SE$\times 10^{-3}$
New	0.54	0.377	0.457	0.686	0.2438
VG	0.63	0.491	0.532	0.678	0.2387
AL	0.64	0.650	0.446	0.676	0.2305

Table 5.8: Normal distribution, n=20

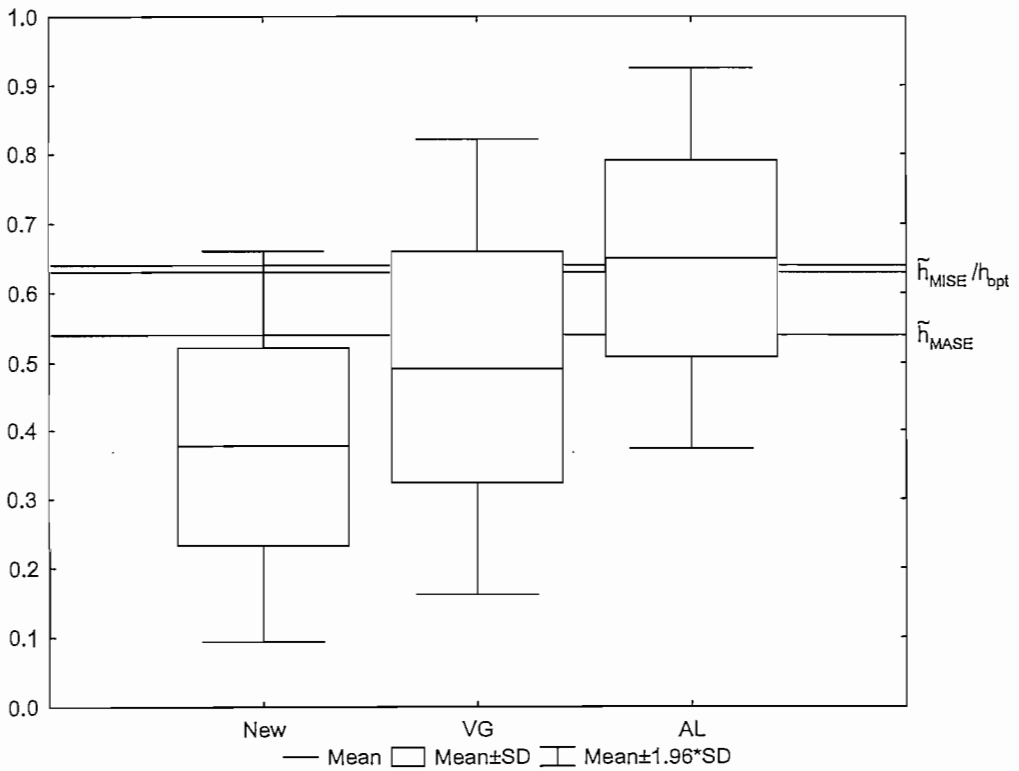


Figure 5.9: The bandwidths obtained from the New, VG and AL procedures for the normal distribution and n=20

column to the estimated standard error in (5.8). The last two columns refers respectively to $\hat{d}_{\mathrm{ASE}}$, defined in (5.5), and its estimated standard error defined in (5.7). Figure 5.9 shows the distribution of the bandwidths obtained from the different procedures. Note that the horizontal lines indicate the values of the optimal bandwidths $\tilde{h}_{\mathrm{MASE}}$, $\tilde{h}_{\mathrm{MISE}}$ and h_{opt}.

5.6 Conclusions

1. The criterion in (5.4) was formulated in order to measure the average squared discrepancy of the non-parametric distribution function estimation with an estimated

bandwidth from the theoretical distribution function, both evaluated in the sample observations. Other criteria to measure the performance of the bandwidth selectors can be formulated. In the simulation study a criterion based on the "integrated squared error concept" was also used, but similar results was obtained. Thus it will not be reported.

2. Inspection of the tables in Appendix A shows that the estimated measure in (5.5) do not differ much for the procedures New, VG and AL in the cases where the normal, skewed unimodal and skewed bimodal distributions are considered.

3. In the case of the exponential distribution, the New and VG procedures outperforms the AL procedure even for sample sizes as large as $n = 200$

4. It has been mentioned in Section 4.3 that the measures MASE(h) and MISE(h) are asymptotically equivalent measures. It is to be expected that the bandwidths determined from the procedures New and VG should approximately be the same for larger sample sizes. This behaviour can be seen in the tables of Appendix A by noting that the values of $\overline{h}$ tends to each other for larger sample sizes.

5. Furthermore, it should be remembered that bandwidths determined from the AL procedure are estimates of the asymptotical optimal bandwidth. The same behaviour that was noted in the previous conclusion can also be seen in the case of the AL procedure.

6. Inspecting the boxplots in Appendix B which reflects the distribution of the estimated bandwidths determined from the different procedures, shows that most of the distributions are skewed with respect to the optimal values $\tilde{h}_{\text{MISE}}$, $\tilde{h}_{\text{MASE}}$ and h_{opt}.

7. In most cases the optimal values $\tilde{h}_{\text{MISE}}$ and $\tilde{h}_{\text{MASE}}$ are located in the upper 25% of the distribution. This might be an indication that there is a bias present in the procedures New and VG when determining the values of the respective bandwidths.

8. In the case of the skewed unimodal distribution the value of h_{opt} is more than at least two standard deviations away from $\overline{h}$ for all sample sizes considered in the simulation study. This observation indicates that convergence of the sample bandwidth to the optimal bandwidth might be very slow.

5.7 Recommendations

In view of the results obtained from the simulation study, the following recommendations can be made:

1. As was noted in point 3 of the previous section, the AL procedure does not perform as well as the other procedures for the exponential distribution. It is recommended that one of the other procedures be used in cases where the data seems to come from an skewed distribution.

2. The difference between the New and VG procedures are negligible, especially for larger sample sizes.

3. The procedures in this study are not dependent on assumptions about the distributional form of the underlying distribution function F. However, the simulation results suggest that the form of the underlying distribution or density function has an impact on the performance of the proposed procedures (see point 1). The results suggest that the procedures tend to differ not much for larger sample sizes in the case of non-skewed symmetrical distributions. The interested reader is referred to Koekemoer (2004) for more information on transformation to normality, which might help to improve the performance of the procedures.

Appendix A

Tables of simulation study

Procedure	Optimal bandwidth	$\overline{h}$	SE$\times 10^{-2}$	$\hat{d}_{\mathrm{ASE}} \times 10^{-2}$	SE$\times 10^{-3}$
New	0.54	0.377	0.457	0.686	0.2438
VG	0.63	0.491	0.532	0.678	0.2387
AL	0.64	0.650	0.446	0.676	0.2305

A.1 Normal distribution, n=20

Procedure	Optimal bandwidth	$\overline{h}$	SE$\times 10^{-2}$	$\hat{d}_{\mathrm{ASE}} \times 10^{-2}$	SE$\times 10^{-4}$
New	0.32	0.255	0.185	0.138	0.4545
VG	0.34	0.290	0.195	0.137	0.4526
AL	0.37	0.369	0.131	0.136	0.4453

A.2 Normal distribution, n=100

Procedure	Optimal bandwidth	$\overline{h}$	SE$\times 10^{-2}$	$\hat{d}_{\mathrm{ASE}} \times 10^{-3}$	SE$\times 10^{-4}$
New	0.27	0.210	0.127	0.688	0.2004
VG	0.28	0.232	0.128	0.685	0.2000
AL	0.30	0.292	0.076	0.683	0.1998

A.3 Normal distribution, n=200

Procedure	Optimal bandwidth	$\overline{h}$	SE$\times 10^{-3}$	$\hat{d}_{\mathrm{ASE}} \times 10^{-3}$	SE$\times 10^{-5}$
New	0.20	0.162	0.726	0.296	0.9109
VG	0.20	0.174	0.734	0.295	0.9105
AL	0.22	0.215	0.381	0.295	0.9097

A.4 Normal distribution, n=500

Procedure	Optimal bandwidth	$\overline{h}$	SE$\times 10^{-2}$	$\hat{d}_{\text{ASE}} \times 10^{-2}$	SE$\times 10^{-3}$
New	0.27	0.188	0.304	0.725	0.2481
VG	0.33	0.247	0.392	0.721	0.2453
AL	0.35	0.478	0.445	0.811	0.2369

A.5 Exponential distribution, n=20

Procedure	Optimal bandwidth	$\overline{h}$	SE$\times 10^{-2}$	$\hat{d}_{\text{ASE}} \times 10^{-2}$	SE$\times 10^{-4}$
New	0.12	0.093	0.101	0.148	0.4646
VG	0.13	0.104	0.112	0.148	0.4640
AL	0.20	0.228	0.095	0.168	0.4554

A.6 Exponential distribution, n=100

Procedure	Optimal bandwidth	$\overline{h}$	SE$\times 10^{-3}$	$\hat{d}_{\text{ASE}} \times 10^{-3}$	SE$\times 10^{-4}$
New	0.09	0.067	0.589	0.740	0.2024
VG	0.09	0.073	0.637	0.740	0.2024
AL	0.16	0.167	0.497	0.831	0.2057

A.7 Exponential distribution, n=200

Procedure	Optimal bandwidth	$\overline{h}$	SE$\times 10^{-3}$	$\hat{d}_{\text{ASE}} \times 10^{-3}$	SE$\times 10^{-5}$
New	0.06	0.043	0.301	0.315	0.9154
VG	0.06	0.045	0.318	0.314	0.9150
AL	0.12	0.110	0.212	0.343	0.9075

A.8 Exponential distribution, n=500

Procedure	Optimal bandwidth	$\overline{h}$	SE$\times 10^{-2}$	$\hat{d}_{\mathrm{ASE}} \times 10^{-2}$	SE$\times 10^{-3}$
New	0.36	0.278	0.354	0.629	0.2080
VG	0.42	0.361	0.406	0.625	0.2059
AL	0.76	0.491	0.359	0.639	0.2033

A.9 Skewed unimodal distribution, n=20

Procedure	Optimal bandwidth	$\overline{h}$	SE$\times 10^{-2}$	$\hat{d}_{\mathrm{ASE}} \times 10^{-2}$	SE$\times 10^{-4}$
New	0.24	0.177	0.140	0.141	0.4682
VG	0.25	0.200	0.148	0.140	0.4665
AL	0.44	0.265	0.103	0.140	0.4623

A.10 Skewed unimodal distribution, n=100

Procedure	Optimal bandwidth	$\overline{h}$	SE$\times 10^{-3}$	$\hat{d}_{\mathrm{ASE}} \times 10^{-3}$	SE$\times 10^{-4}$
New	0.18	0.148	0.889	0.733	0.2261
VG	0.20	0.162	0.920	0.732	0.2263
AL	0.35	0.209	0.620	0.734	0.2276

A.11 Skewed unimodal distribution, n=200

Procedure	Optimal bandwidth	$\overline{h}$	SE$\times 10^{-3}$	$\hat{d}_{\mathrm{ASE}} \times 10^{-3}$	SE$\times 10^{-4}$
New	0.14	0.111	0.541	0.317	0.1013
VG	0.14	0.119	0.549	0.317	0.1012
AL	0.26	0.151	0.295	0.316	0.1007

A.12 Skewed unimodal distribution, n=500

Procedure	Optimal bandwidth	$\overline{h}$	SE$\times 10^{-2}$	$\hat{d}_{\text{ASE}} \times 10^{-2}$	SE$\times 10^{-3}$
New	0.61	0.410	0.522	0.664	0.2210
VG	0.80	0.542	0.608	0.652	0.2155
AL	0.73	0.743	0.420	0.645	0.2034

A.13 Skewed Bimodal distribution, n=20

Procedure	Optimal bandwidth	$\overline{h}$	SE$\times 10^{-2}$	$\hat{d}_{\text{ASE}} \times 10^{-2}$	SE$\times 10^{-4}$
New	0.32	0.245	0.210	0.135	0.4171
VG	0.40	0.278	0.233	0.135	0.4137
AL	0.43	0.404	0.114	0.135	0.3989

A.14 Skewed Bimodal distribution, n=100

Procedure	Optimal bandwidth	$\overline{h}$	SE$\times 10^{-2}$	$\hat{d}_{\text{ASE}} \times 10^{-3}$	SE$\times 10^{-4}$
New	0.23	0.195	0.137	0.721	0.2280
VG	0.28	0.213	0.149	0.719	0.2273
AL	0.34	0.309	0.075	0.726	0.2227

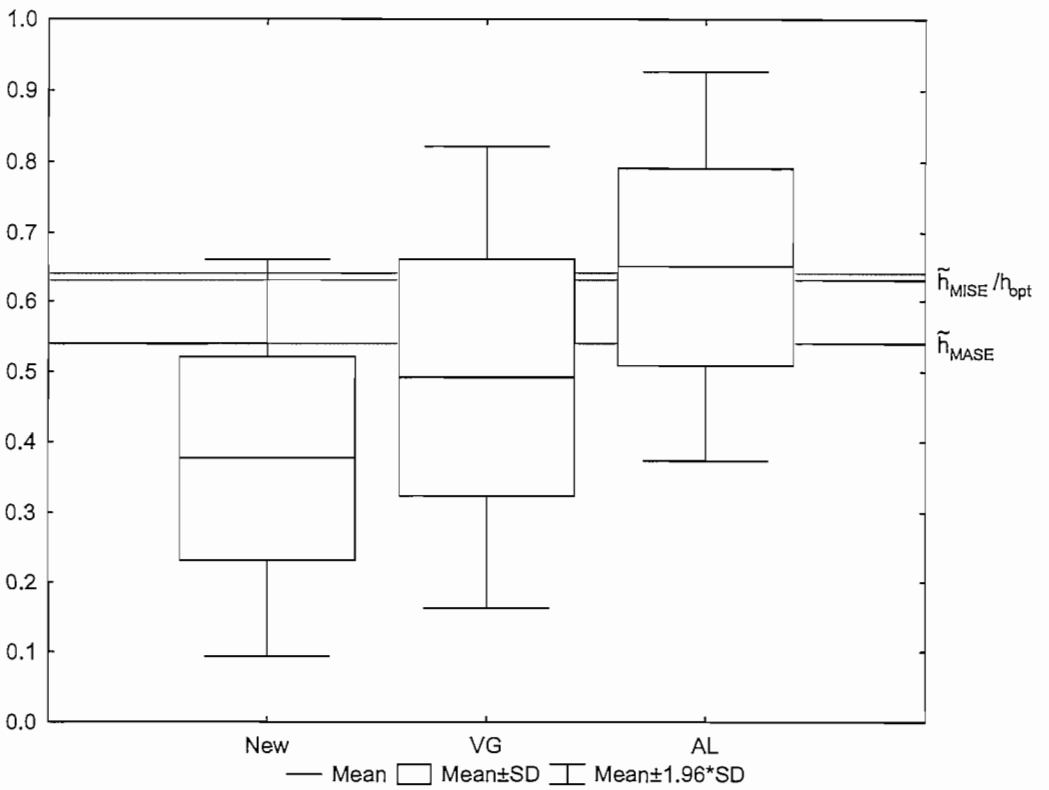
A.15 Skewed Bimodal distribution, n=200

Procedure	Optimal bandwidth	$\overline{h}$	SE$\times 10^{-3}$	$\hat{d}_{\text{ASE}} \times 10^{-3}$	SE$\times 10^{-5}$
New	0.17	0.140	0.766	0.300	0.9251
VG	0.20	0.148	0.806	0.300	0.9243
AL	0.25	0.213	0.448	0.302	0.9185

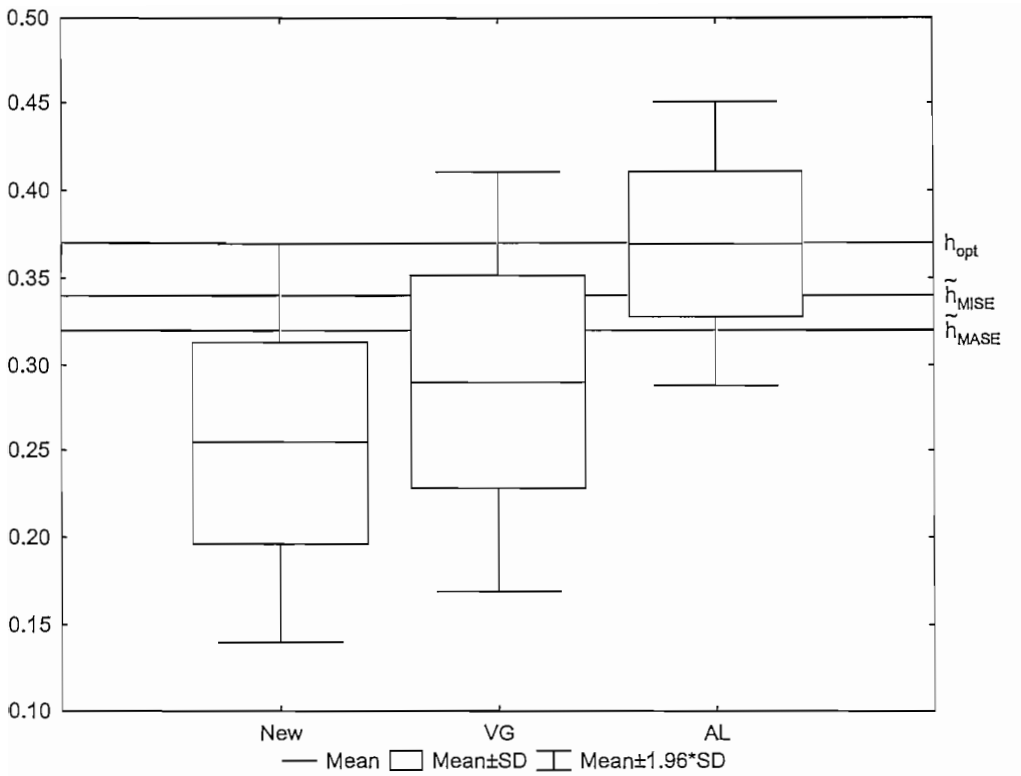
A.16 Skewed Bimodal distribution, n=500

Appendix B

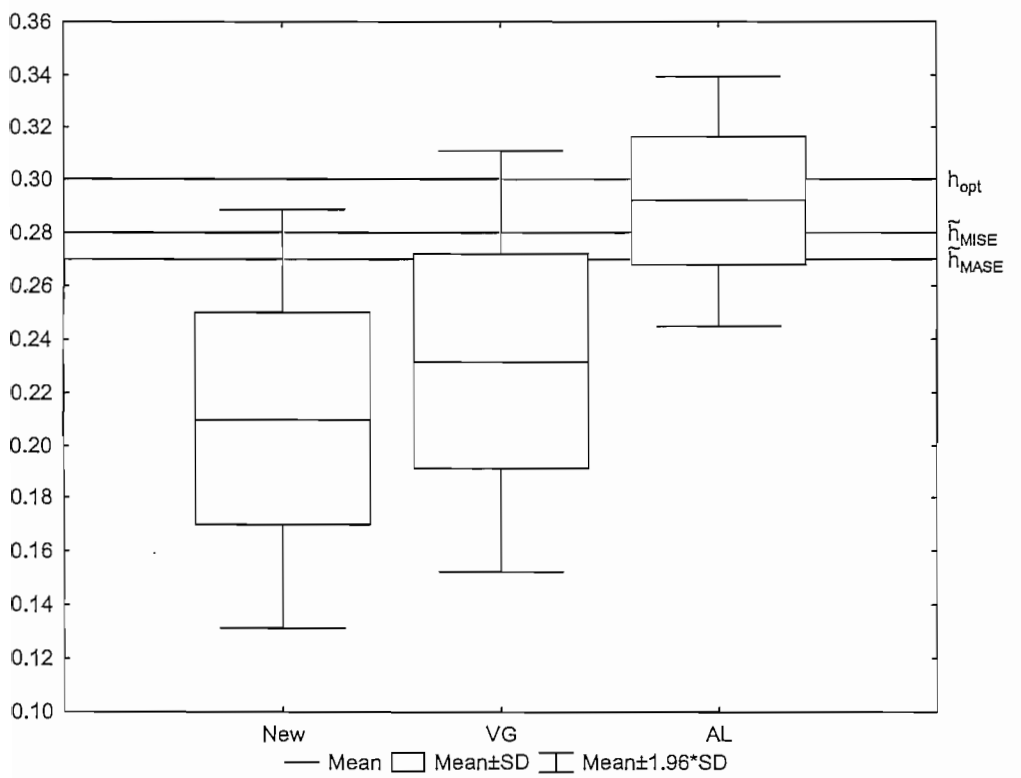
Figures of simulation study



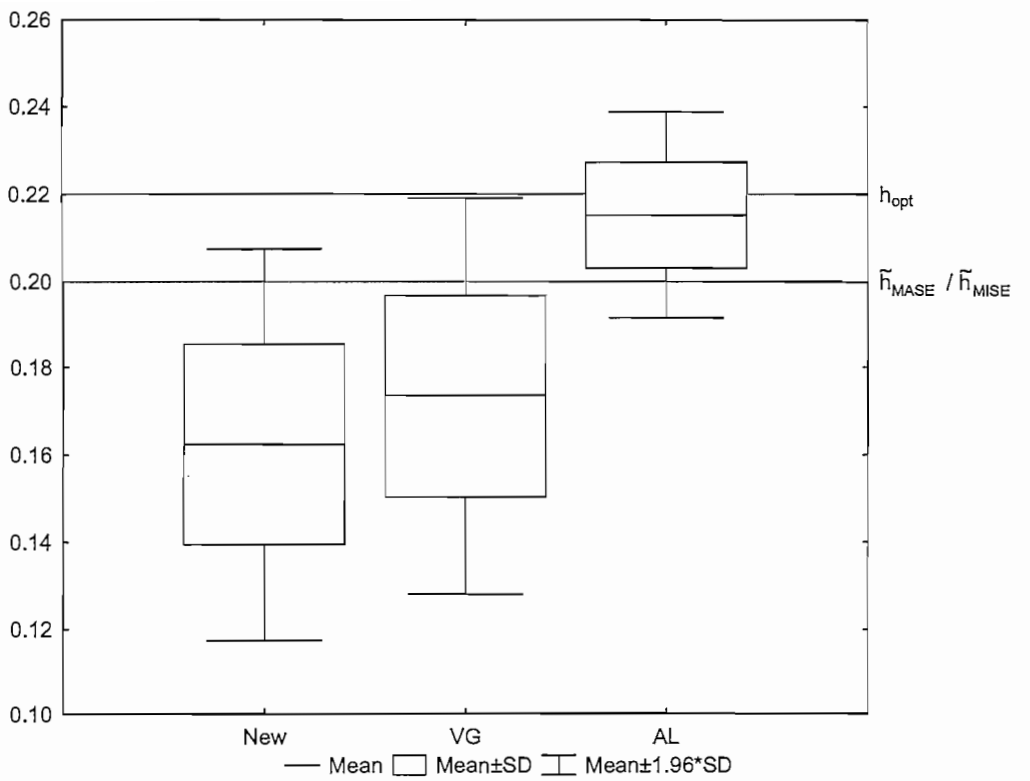
B.1 The bandwidths obtained from the New, VG and AL procedures for the normal distribution and n=20



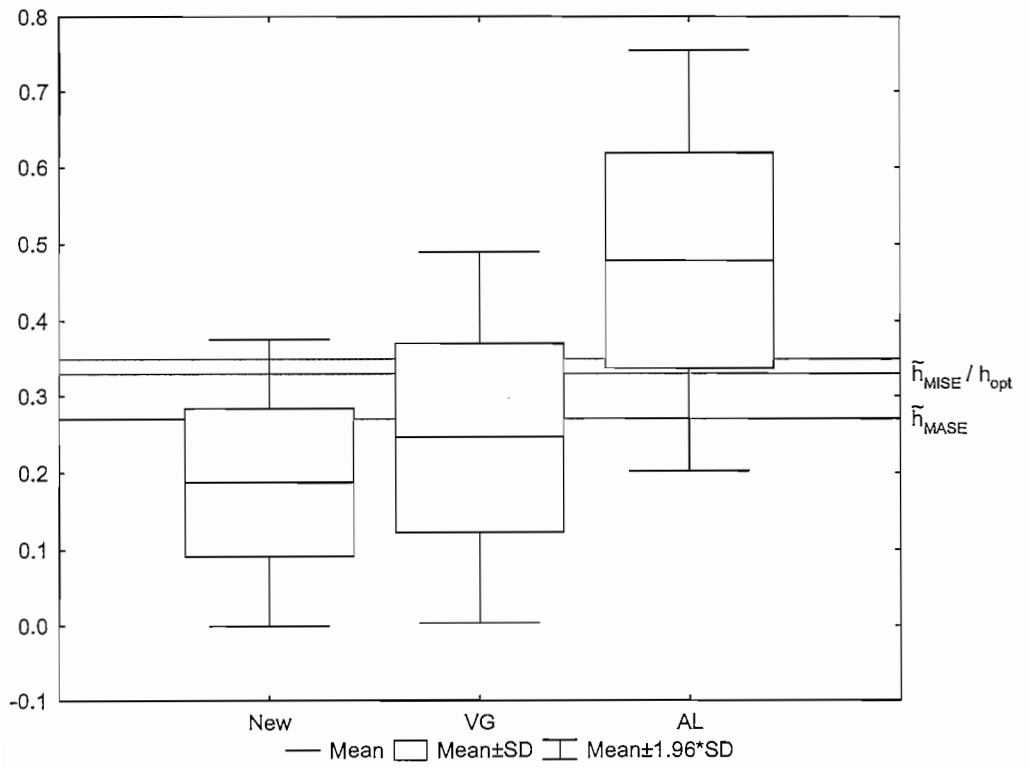
B.2 The bandwidths obtained from the New, VG and AL procedures for the normal distribution and n=100



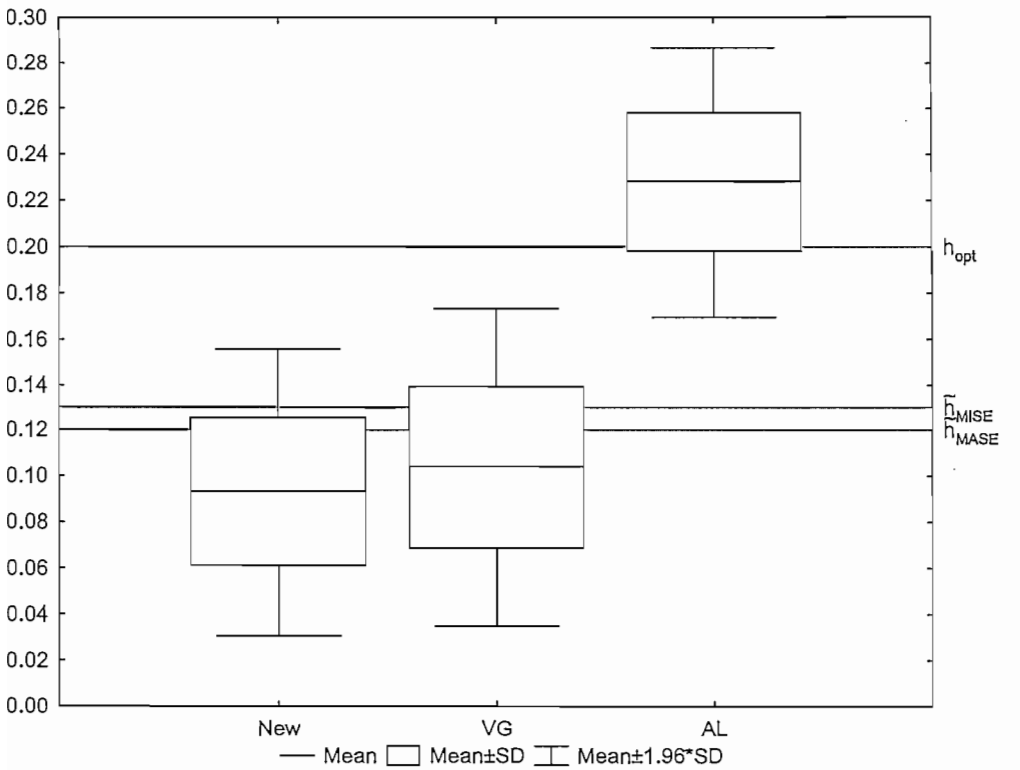
B.3 The bandwidths obtained from the New, VG and AL procedures for the normal distribution and n=200



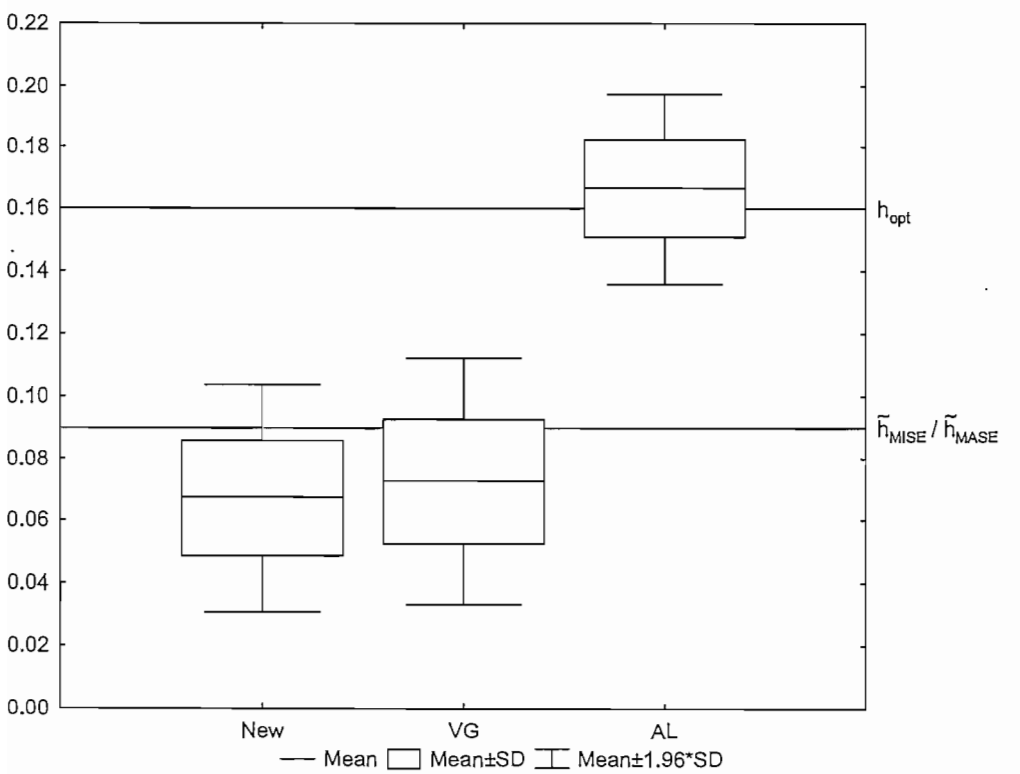
B.4 The bandwidths obtained from the New, VG and AL procedures for the normal distribution and n=500



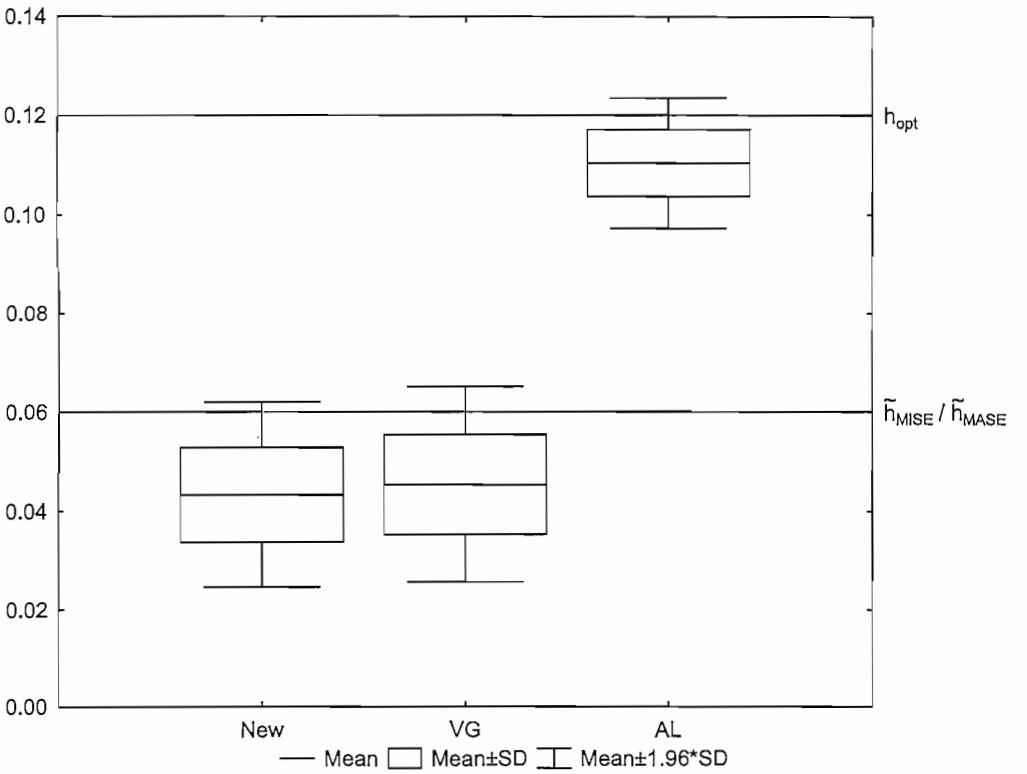
B.5 The bandwidths obtained from the New, VG and AL procedures for the exponential distribution and n=20



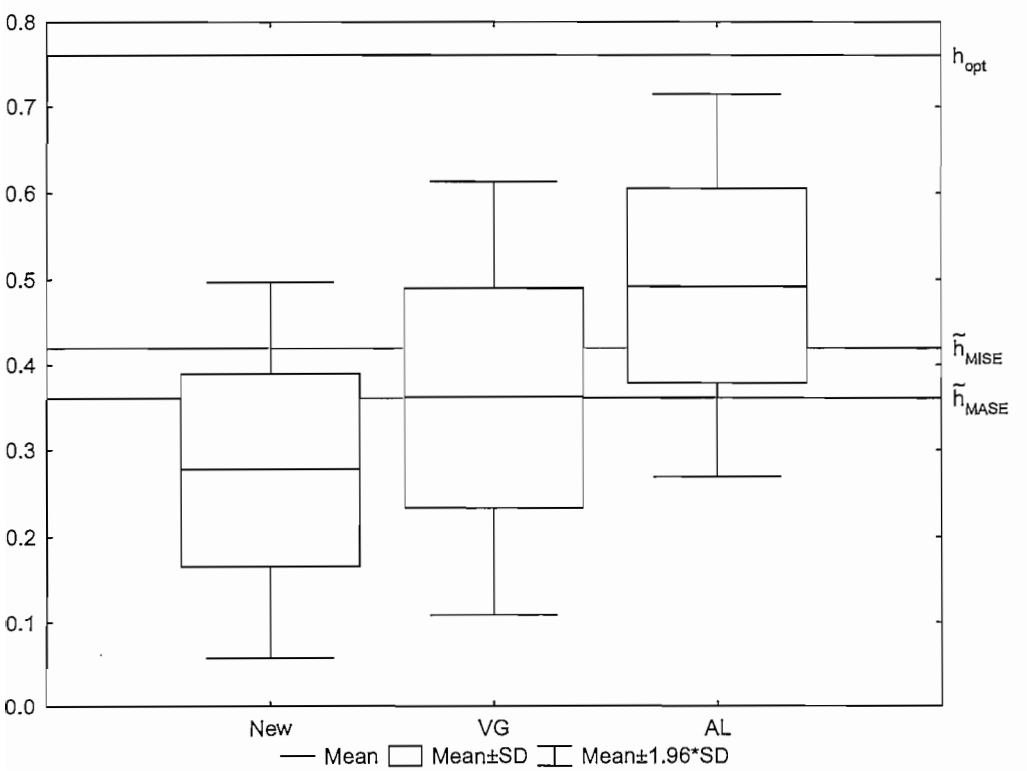
B.6 The bandwidths obtained from the New, VG and AL procedures for the exponential distribution and n=100



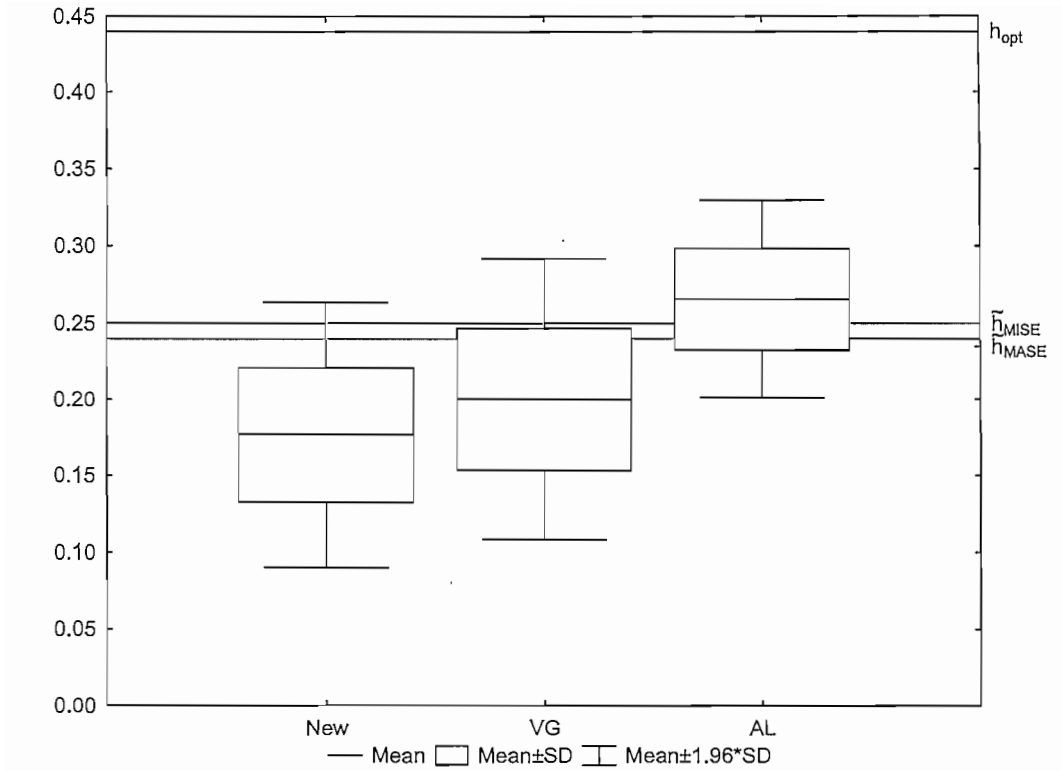
B.7 The bandwidths obtained from the New, VG and AL procedures for the exponential distribution and n=200



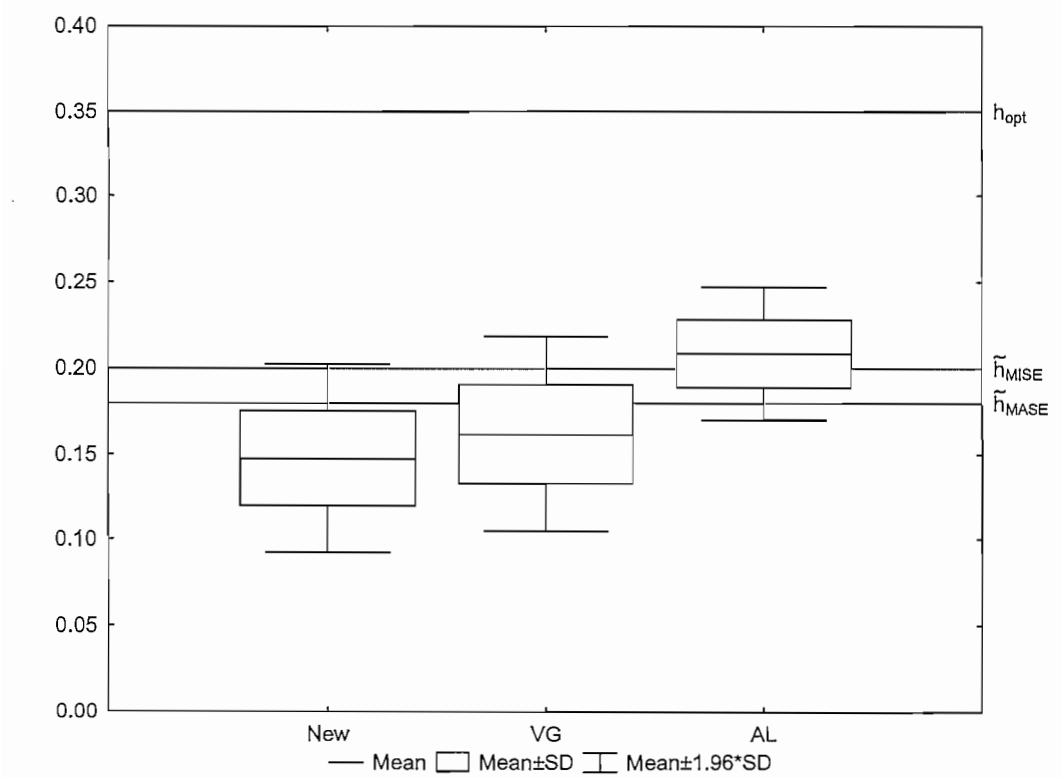
B.8 The bandwidths obtained from the New, VG and AL procedures for the exponential distribution and n=500



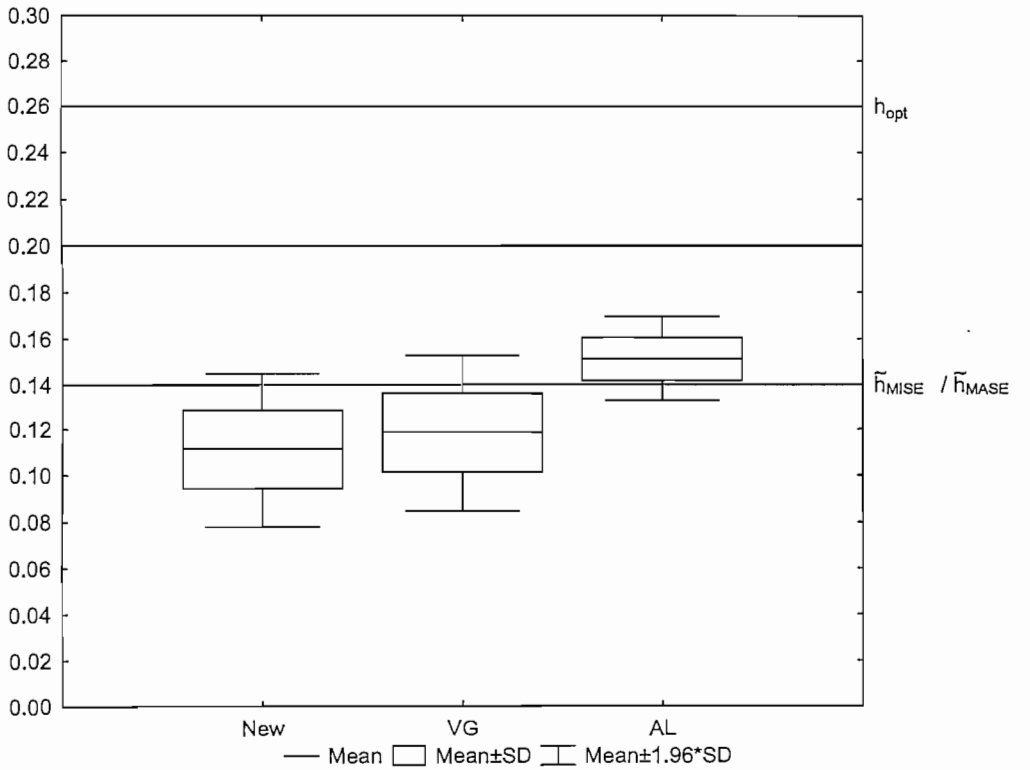
B.9 The bandwidths obtained from the New, VG and AL procedures for the skewed unimodal distribution and n=20



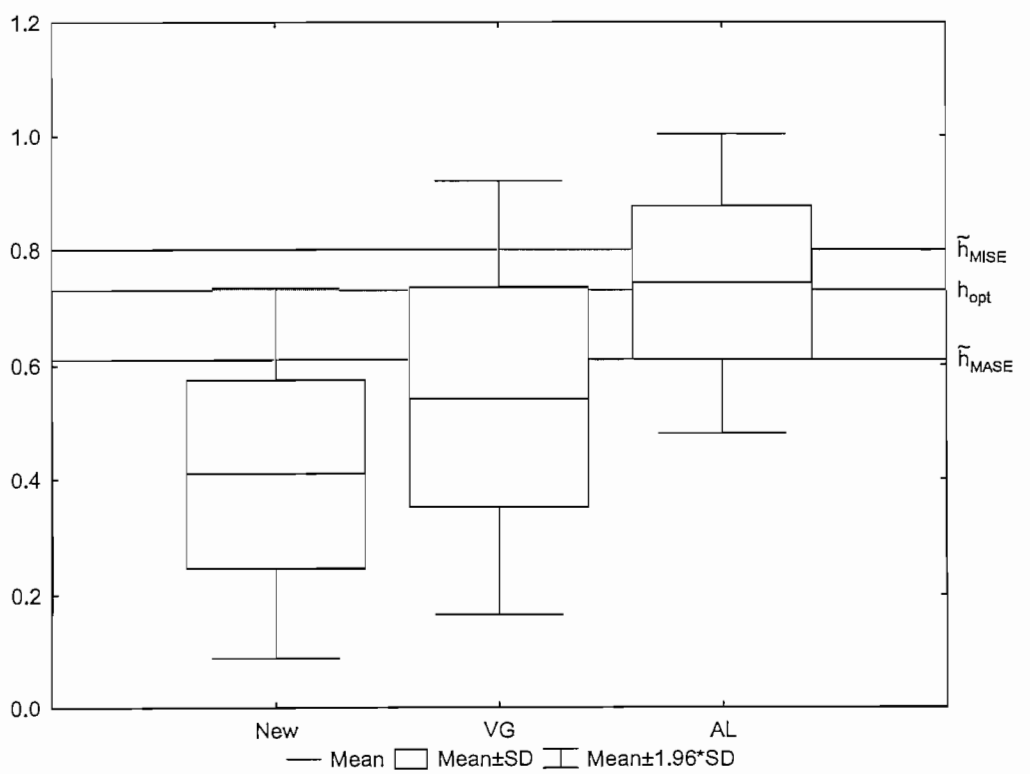
B.10 The bandwidths obtained from the New, VG and AL procedures for the skewed unimodal distribution and n=100



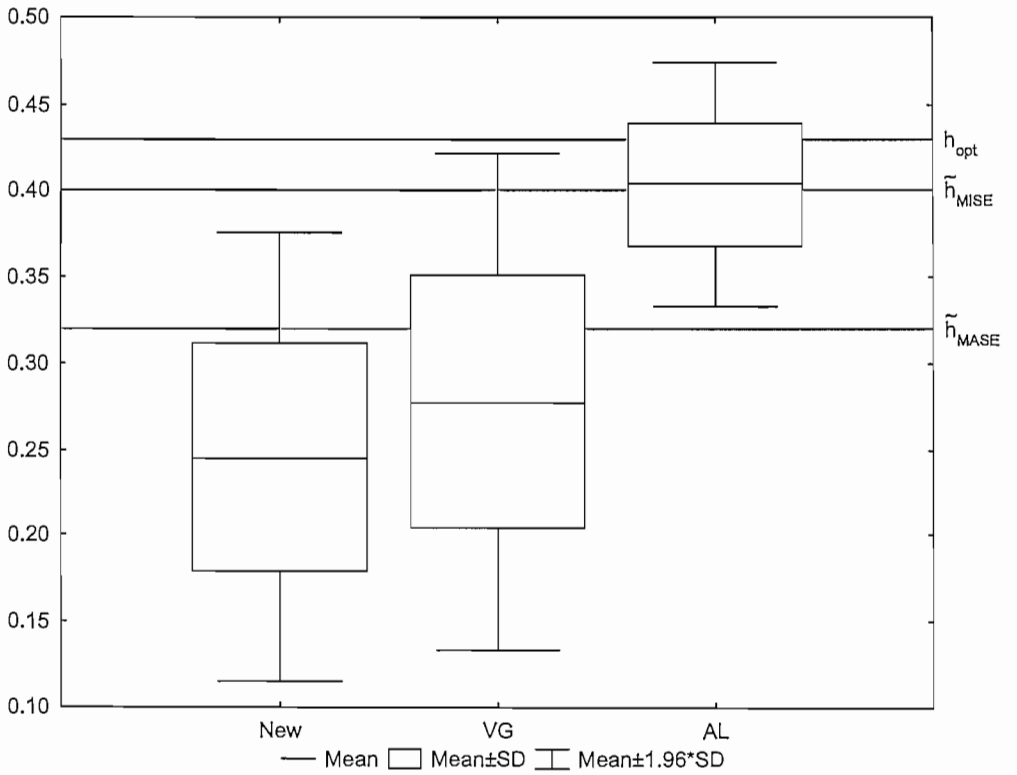
B.11 The bandwidths obtained from the New, VG and AL procedures for the skewed unimodal distribution and n=200



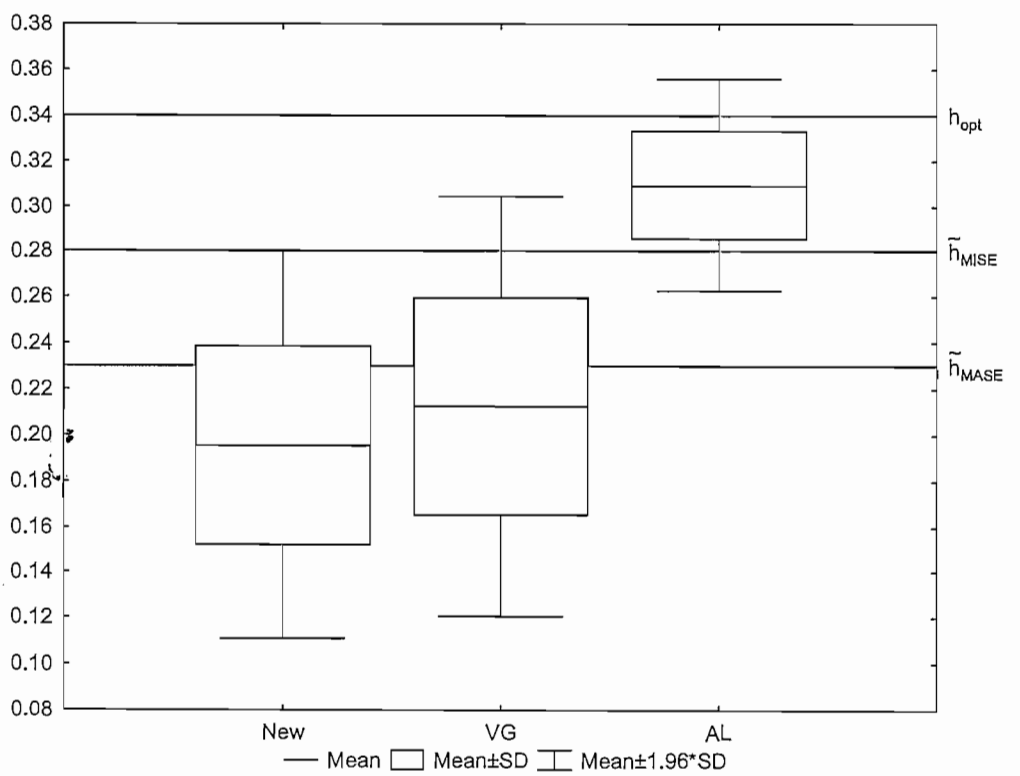
B.12 The bandwidths obtained from the New, VG and AL procedures for the skewed unimodal distribution and n=500



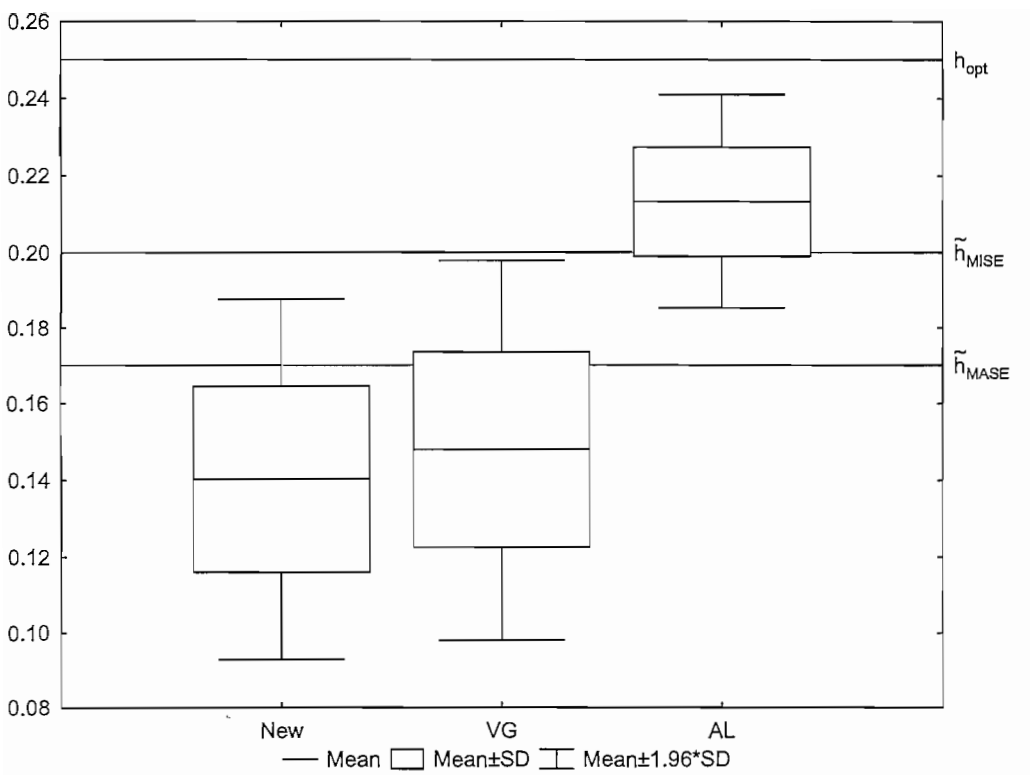
B.13 The bandwidths obtained from the New, VG and AL procedures for the skewed bimodal distribution and n=20



B.14 The bandwidths obtained from the New, VG and AL procedures for the skewed bimodal distribution and n=100



B.15 The bandwidths obtained from the New, VG and AL procedures for the skewed bimodal distribution and n=200



B.16 The bandwidths obtained from the New, VG and AL procedures for the skewed bimodal distribution and n=500

List of Figures

List of Tables

Bibliography

Altman, N. and Léger, C. (1995). Bandwidth selection for kernel distribution function estimators, *Journal of Statistical Planning and Inference* **46**: 195–214.

Bowman, A. (1984). An alternative method of cross-validation for the smoothing of density estimates, *Biometrika* **71**: 353–360.

Bowman, A., Hall, P. and Prvan, T. (1998). Bandwidth selection for the smoothing of distribution functions, *Biometrika* **85**(4): 799–808.

Efron, B. and Tibshirani, R. (1993). *An introduction to the bootstrap*, Chapman and Hall, New York.

Falk, M. (1983). Relative efficiency and deficiency of kernel type estimators of smooth distribution functions, *Statist. Neer.* **37**: 73–83.

Hall, P. and Marron, J. S. (1987). Estimation of integrated squared density derivatives, *Statistics and Probability Letters* **6**: 109–115.

Jacod, J. and Protter, P. (2000). Heidelberg: Springer Verlag, Berlin. 250 p.

Jones, M. C. (1990). The performance of kernel density functions in kernel distribution function estimation, *Statistics and Probability Letters* **9**: 129–132.

Koekemoer, G. (2004). *A new method for transforming data to normality with application to density estimation*, PhD thesis, North-West University, Potchefstroom.

Lejuene, M. and Sarda, P. (1992). Smooth estimators od distribution and density estimators, *Computational Statistics and Data Analysis* **14**(4): 457–471.

Marron, J. (1987). A comparison of cross-validation techniques in density estimation, *Annals of Statistics* **15**: 152–162.

Marron, J. S. and Wand, M. P. (1992). Exact mean integrated squared error, *Annals of Statistics* **20**: 1919–1932.

Nadaraya, E. A. (1964). Some new estimates for distribution functions, *Theory of probability and its Applications* **15**: 497–500.

Parzen, E. (1962). On estimation of a probability density and mode, *Annals of Mathematical Statistics* **33**: 1065–1076.

Polansky, A. M. (1997). Bandwidth selection for kernel distribution functions, Unpublished Manuscript.

Reiss, R. D. (1981). Nonparametric estimation of smooth distribution functions, *Scandinavian Journal of Statistics* (8): 116–119.

Rosenblat, M. (1956). Remarks on some nonparametric estimates of a density function, *Annals of Mathematical Statistics* **27**: 832–837.

Sarda, P. (1993). Smoothing parameter selection for smooth distribution functions, *Journal of Statistical Planning and Inference* **35**: 65–75.

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, Series B* **53**: 683–690.

Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*, Chapman and Hall, London.

Swanepoel, J. W. H. (1986). A note on proving that the (modified) bootstrap works, *Communications in Statistics - Theory and Methods* **15**(11): 3193–3203.

Van Graan, F. C. (1983). *Nie-parametriese beraming van verdelingsfunksies*, Master's thesis, P.U. for C.H.E, Potchefstroom. 65 p.

Watson, G. S. and Leadbetter, M. R. (1964). Hazard analysis ii, *Sankhyā Series A* **26**: 101–116.

Winter, B. B. (1973). Strong uniform consistency of integrals of density estimators, *Canadian Journal of Statistics* **1**: 247–253.

Winter, B. B. (1979). Convergence rate of perturbed empirical distribution functions, *Journal of Applied. Probability* **16**: 163–176.

Yamato, H. (1973). Uniform convergence of a distribution function, *Bulletin of Mathematical Statistics* **15**: 69–78.