NTNU
Innovation and Creativity

Bandwith selection based on a special choice of the kernel

Thomas Oksavik

Master of Science in Physics and Mathematics
Submission date: June 2007
Supervisor: Nikolai Ushakov, MATH

Norwegian University of Science and Technology
Department of Mathematical Sciences

Problem Description

In the frames of the project it is supposed to study influence of the choice of the kernel in kernel density estimation on deviation of the asymptotic bandwith from the optimal bandwith. On the basis on this investigation, it is supposed to develop methods of the kernel choice.

Assignment given: 15. January 2007
Supervisor: Nikolai Ushakov, MATH

Preface

This report is the result of my work in the course TMA4905 Statistics. It is written in my 10th and final semester of my Master of Science degree at the Department of Mathematical Sciences, NTNU. This course corresponds to a work load of 30 points, or one semester.

I would like to thank my supervisor Professor Nikolai Ushakov for help and guidance through the semester.

Abstract

The bandwidth in kernel density estimation can be selected in different ways. This paper focus on investigation of the bandwidth selected on basis of the mean integrated squared error, and its asymptotic case, for different choices of the kernel. Also this paper look at properties of the sinc-kernel, and investigates if this non-conventional kernel is better than standard kernels in some cases. Use of empirical data for bandwidth selection is also an important topic which this paper deals with.

Contents

Chapter 1

Introduction

The main goal of this report is to analyze and investigate the bandwidth selection in kernel density estimation. We have studied already existing theories, and we have developed some results based on these theories. We have done this for a wide range of the kernel choice to get more substance behind our investigation.
In chapter 2 we give a short introduction to the theory used in this work. Some basics about kernel density estimation, an introduction to characteristic functions and some calculations of important expressions are given here. In chapter 3 we pick out a large number of kernels, and use these kernels to investigate a conjecture given in Marron and Wand about bandwidth selection. In chapter 4 we use derivatives to investigate the same problem. We also try to construct our own method of bandwidth selection.
In chapter 5 we try to prove the same conjecture for some special, simple case and in the last chapter we consider a non-conventional kernel and investigate if its properties are different from conventional kernels. We also compare the bandwidth selected from a simulated sample with the theoretical bandwidth, and comment on the result.

In our work we have used the mathematical software MAPLE 9.5 to make the difficult calculations and to draw the graphs. The programming code and some comments are given in the appendix.

Chapter 2

Theory

2.1 Kernel Density Estimate

A lot of the theory in this chapter is obtained from Wand and Jones [4, chapter 2]. Kernel estimation is a non-parametric estimation method, and a very effective tool to structure data before analysis. It is very useful when parametric methods are inappropriate. We consider a sample $X_1, ..., X_n$, which are independent and identically distributed. We will deal with univariate kernel density estimator on the form

$$\widehat{f}(x;h) = (nh)^{-1} \sum_{i=1}^{n} K(\frac{x - X_i}{h}) \tag{2.1}$$

In this formula K(x) satisfies $\int K(x)dx = 1$, and h is the bandwidth, which is a positive number. K(x) is usually chosen to be a density function, but it is also possible to use kernels that are not a density. There are several types of kernels. The most common is the Gaussian kernel or the Epanechnikov (quadratic) kernel. Figure 2.1 shows four different kernels.

The choice of kernel type is however not that important, when we work with conventional kernels. Conventional kernels are kernels which are probability density functions. The most important choice, is the choice of the bandwidth h. If h is too small, the kernel estimate will be undersmoothed. This means that there will be many local maxima, and the estimator is very dependent on the data. If h is too big, the graph will be very smooth, and all structure will be gone. This is called oversmoothing. So it is important to choose h correct. Figure 2.2 shows how the choice of h affects the kernel density estimate, when the kernel is chosen to be the skewed bimodal density $f(x) = \frac{3}{4}N(0,1) + \frac{1}{4}N(\frac{3}{2}, (\frac{1}{3})^2)$, and the sample size is $n = 5000$.

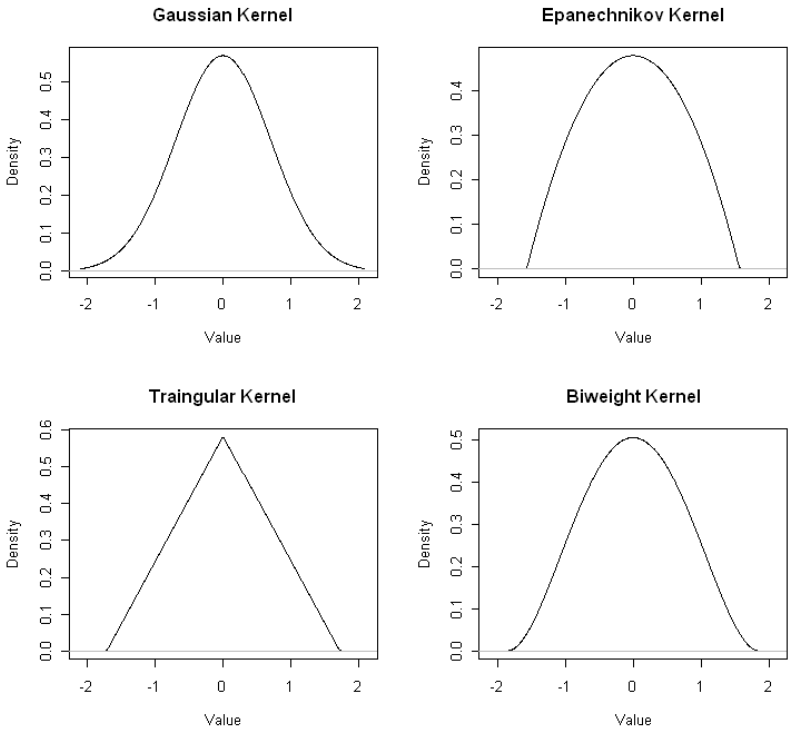
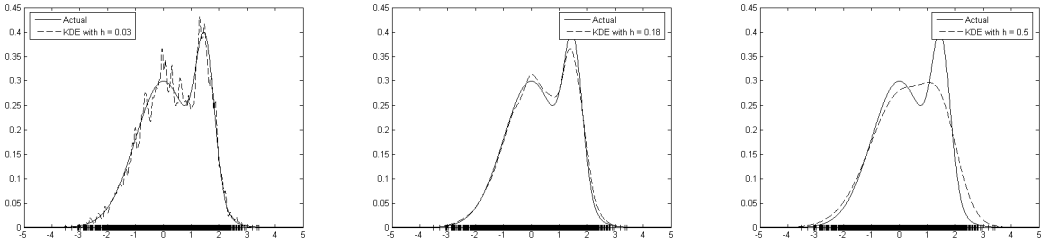


Figure 2.1: Four different kernels



(a) h=0.03 (b) h=0.18 (c) h=0.5

Figure 2.2: Kernel density estimate with different bandwidth.

2.2 MSE/MISE/AMISE

The mean squared error (MSE) is a useful tool for measuring the error in the density estimate. It is defined as

$$MSE(\widehat{f}(x;h)) = E(\widehat{f}(x;h) - f(x))^2 = Var\widehat{f}(x;h) + (Bias(\widehat{f}(x;h)))^2 \tag{2.2}$$

where $\widehat{f}(x;h)$ is the density estimate of the density function $f(x)$. MSE computes the error in a single point on the density estimate. To compute the error over the entire real line, we use another error criterion, that is the mean integrated squared error (MISE). This is simply the integral over the whole real line of MSE

$$MISE(\widehat{f}(x;h)) = E\int(\widehat{f}(x;h) - f(x))^2dx = \int MSE(\widehat{f}(x;h))dx \tag{2.3}$$

As the formula says, MISE measures the distance between an estimate of f, and f itself. A problem with MSE/MISE is that it depends on the bandwidth in a rather complicated way. One way to deal with this problem, is to make asymptotic approximations of MSE/MISE. We will use the following assumptions:
1) f'' is continuous, square integrable and ultimately monotone (A function f is ultimately monotone if there exists y such that for $x > y$, $f(x)$ is monotone).
2) $h = h_n$ is a non-random sequence of positive numbers, satisfying
$\lim_{n\to\infty} h = 0$ and $\lim_{n\to\infty} nh = \infty$
3) K is a bounded probability density function having finite fourth moment and symmetry about the origin.
The asymptotic MISE (AMISE) is a large sample approximation for the variance and bias. These approximations have easier expressions, and depends on the bandwidth in a much simpler way.

$$AMISE(\widehat{f}(\cdot;h)) = (nh)^{-1}R(K) + \frac{1}{4}h^4\mu_2(K)^2R(f'') \tag{2.4}$$

Here $R(K) = \int(K(x))^2dx$, $\mu_2(K) = \int x^2K(x)dx$ and $R(f'') = \int(f''(x))^2dx$. The h that minimizes MISE, is called h_{MISE} and the h that minimizes AMISE, is called h_{AMISE}

2.3 Characteristic functions

The characteristic function ϕ_X of a random variable X is a complex function that uniquely determines the distribution function. It is defined as

$$\phi_X(t) = E[e^{itX}] = \int_{-\infty}^{\infty} e^{itx} f(x)dx \tag{2.5}$$

When $f(x)$ is symmetric about the origin, we can write the characteristic function like

$$\phi_X(t) = E[e^{itX}] = \int_{-\infty}^{\infty} \cos(tx) f(x)dx \tag{2.6}$$

The characteristic function has many advantages. It makes it easier to find moments of a random variable, since the moments are derivatives at zero of its characteristic functions. There are two important theorems for characteristic functions. They are the inversion theorem and the uniqueness theorem. For proofs look in [2].

The inversion theorem

$$P(a < X < b) = \lim_{T\to\infty} \frac{1}{2\pi} \int_{-T}^{T} \frac{e^{-ita} - e^{-itb}}{it} \phi_X(t)dt \tag{2.7}$$

The uniqueness theorem
If $\phi_X(t) = \phi_Y(t)$ for all t, then $X = Y$ in distribution.

Another important formula, which have been used later in this report is the Parseval identity.
If $\phi(t)$ is the characteristic function of $f(x)$, and $\psi(t)$ is the characteristic function of $g(x)$, then

$$\int (f(x) - g(x))^2 dx = \frac{1}{2\pi} \int |\phi(t) - \psi(t)|^2 dt \tag{2.8}$$

Proof of Parseval identity:
If X and Y are iid from a distribution $f(x) = \frac{1}{2\pi} \int e^{-itx} \phi(t)dt$, where $\phi(t)$ is the corresponding characteristic function $\phi(t) = \int e^{itx} f(x)dx$, we have

$$
\begin{aligned}
f_{X-Y}(x) &= \int f(x+t)f(t)dt \\
\phi_{X-Y}(t) &= Ee^{it(X-Y)} = Ee^{itX}Ee^{-itY} = \phi(t)\phi(-t) = |\phi(t)|^2 \\
f_{X-Y}(x) &= \frac{1}{2\pi}\int e^{-itx}\,|\phi(t)|^2\,dt \\
\int (f(x))^2 dx &= f_{X-Y}(0) = \frac{1}{2\pi}\int |\phi(t)|^2\,dt
\end{aligned}
$$

It is easy to derive that this also holds for the sum of two densities $f(x)$ and $g(x)$ with characteristic functions $\phi(t)$ and $\psi(t)$.
Let X and Y be two independent random variables from distribution $f(x)$ and $g(x)$ respectively, and let $\phi(t)$ and $\psi(t)$ be the characteristic functions to $f(x)$ and $g(x)$ respectively. Then

$$
\begin{aligned}
f_{X-Y} &= \int f(x+t)g(t)dt \\
\phi_{X-Y}(t) &= Ee^{it(X-Y)} = Ee^{itX}Ee^{-itY} \\
&= \phi(t)\psi(-t) = \phi(t)\overline{\psi(t)} \\
\int f(x)g(x)dx &= f_{X-Y}(0) = \frac{1}{2\pi}\int \phi(t)\overline{\psi(t)}dt
\end{aligned}
$$

Parseval identity for two densities becomes

$$
\begin{aligned}
\int (f(x)-g(x))^2 dx &= \int (f(x))^2 dx - 2\int f(x)g(x)dx + \int (g(x))^2 dx \\
&= \frac{1}{2\pi}\int |\phi(t)|^2\,dt - 2\frac{1}{2\pi}\int \phi(t)\overline{\psi(t)}dt + \frac{1}{2\pi}\int |\psi(t)|^2\,dt \\
&= \frac{1}{2\pi}\int |\phi(t)|^2\,dt - \frac{1}{2\pi}\int \phi(t)\overline{\psi(t)}dt - \\
&\quad \frac{1}{2\pi}\int \overline{\phi(t)}\psi(t)dt + \frac{1}{2\pi}\int |\psi(t)|^2\,dt \\
&= \frac{1}{2\pi}\int |\phi(t)-\psi(t)|^2\,dt
\end{aligned}
\tag{2.9}
$$

Parseval identity for derivatives becomes

$$
\int (f^n(x))^2 dx = \frac{1}{2\pi}\int t^{2n}\,|\phi(t)|^2\,dt \tag{2.10}
$$

2.4 MISE in terms of characteristic functions

To be able to work more easily, we had to find an expression for MISE in terms of characteristic functions.
Let

$$\phi(t) = Ee^{itX_j} = \int e^{itx} f(x)dx$$

be the c.f of a random variable X_j,
and let

$$\phi_n(t) = \frac{1}{n}\sum_{j=1}^{n} e^{itX_j}$$

be the empirical c.f associated with sample X_1, ... , X_n.
Let also

$$\psi(t) = \int e^{itx} K(x)dx$$

be the c.f of the kernel $K(x)$.
From this it follows that the c.f of $K_h(x) = \frac{1}{h}K(\frac{x}{h})$ is $\psi(ht)$.
The kernel density estimate can be written as

$$\frac{1}{nh}\sum_{i=1}^{n} K(\frac{x - X_i}{h}) = \frac{1}{n}\sum_{i=1}^{n} K_h(x - X_i) = \int K_h(x - y)dF_n(y)$$

where

$$F_n(x) = \frac{1}{n}\sum_{k=1}^{n} I_{(-\infty,x]}(X_k) = \begin{cases} 0 & \text{if } x < X_{(1)} \\ \frac{k}{h} & \text{if } X_{(k)} \leq x < X_{(k+1)} \\ 1 & \text{if } x \geq X_{(k)} \end{cases}$$

is the empirical distribution function.
This is the convolution between two functions, the first one with c.f $\psi(ht)$ and the other one with c.f $\phi_n(t)$. So the characteristic function of the kernel density estimate is

$$\chi(t) = \phi_n(t)\psi(ht). \tag{2.11}$$

Then we can express MISE the following way

$$
\begin{aligned}
MISE(f_n(x;h)) &= E\int (f_n(x;h)-f(x))^2 dx \\
&= \frac{1}{2\pi} E\int |\phi_n(t)\psi(ht)-\phi(t)|^2\, dt \\
&= \frac{1}{2\pi}\int E(\phi_n(t)\psi(ht)-\phi(t))(\overline{\phi_n(t)\psi(ht)}-\overline{\phi(t)})dt \\
&= \frac{1}{2\pi}\int |\psi(ht)|^2 E|\phi_n(t)|^2 - \phi(t)\overline{\psi(ht)}E\overline{\phi_n(t)} - \\
&\quad \overline{\phi(t)}\psi(ht)E\phi_n(t) + |\phi(t)|^2\, dt \\
&= \frac{1}{2\pi}\int |\psi(ht)|^2 E|\phi_n(t)|^2 - \overline{\psi(ht)}\,|\phi(t)|^2 - \\
&\quad \psi(ht)\,|\phi(t)|^2 + |\phi(t)|^2\, dt \\
&= \frac{1}{2\pi}\int |\psi(ht)|^2 E|\phi_n(t)-\phi(t)|^2 + |\psi(ht)|^2\,|\phi(t)|^2 - \\
&\quad \overline{\psi(ht)}\,|\phi(t)|^2 - \psi(ht)\,|\phi(t)|^2 + |\phi(t)|^2\, dt \\
&= \frac{1}{2\pi}\int |\psi(ht)|^2 \frac{1}{n}(1-|\phi(t)|^2) + \\
&\quad |\phi(t)|^2\,(|\psi(ht)|^2 - \overline{\psi(ht)} - \psi(ht) + 1)dt \\
&= \frac{1}{2n\pi}\int (1-|\phi(t)|^2)\,|\psi(ht)|^2\, dt + \\
&\quad \frac{1}{2\pi}\int |1-\psi(ht)|^2\,|\phi(t)|^2\, dt
\end{aligned}
\tag{2.12}
$$

In this proof we have used Parsevals identity (2.8),
$E|\phi_n(t)-\phi(t)|^2 = \frac{1}{n}(1-|\phi(t)|^2)$, and that $E\phi_n(t)=\phi(t)$.

We know from (2.2) that MSE can be written as $Var + (bias)^2$. In the same way MISE can be written as

$$
MISE = \int Var + \int (bias)^2 \tag{2.13}
$$

We calculate explicit formulas for the integrated variance (IV) and the in-

tegrated squared bias (ISB), to show that this corresponds to (2.12).

$$\begin{aligned}
f_n(x;h) &= \frac{1}{2\pi}\int e^{-itx}\psi(ht)\phi_n(t)dt \\
Ef_n(x;h) &= \frac{1}{2\pi}\int e^{-itx}\psi(ht)\phi(t)dt \\
Ef_n(x;h) - f(x) &= \frac{1}{2\pi}\int e^{-itx}\phi(t)(\psi(ht)-1)dt \\
\int (Ef_n(x;h) - f(x))^2 &= \frac{1}{2\pi}\int |\phi(t)|^2 |1-\psi(ht)|^2 dt = ISB
\end{aligned}$$

$$\begin{aligned}
\int Ef_n(x;h)^2 - \int (Ef_n(x;h))^2 &= \frac{1}{2\pi}\int |\psi(ht)|^2 E|\phi_n(t)|^2 dt - \\
&\quad \frac{1}{2\pi}\int |\psi(ht)|^2 |\phi(t)|^2 dt \\
&= \frac{1}{2\pi n}\int |\psi(ht)|^2 (1-|\phi(t)|^2)dt + \\
&\quad \frac{1}{2\pi}\int |\psi(ht)|^2 |\phi(t)|^2 dt - \\
&\quad \frac{1}{2\pi}\int |\psi(ht)|^2 |\phi(t)|^2 dt \\
&= \frac{1}{2\pi n}\int |\psi(ht)|^2 (1-|\phi(t)|^2)dt = IV \\
ISB + IV &= MISE
\end{aligned} \tag{2.14}$$

Chapter 3

Exact MISE and AMISE for a number of kernels

In the article by Marron and Wand [5], the writers compare MISE and AMISE for several different kernels. These kernels were all combinations of normal densities. Even though all of them have different form, they also have some common properties, like existence of all moments.
The conjecture they made based on their results was that $h_{AMISE} < h_{MISE}$, at least for nonnegative kernels. In this chapter we want to investigate this conjecture to see if it holds for a wider range of kernels. Instead of using combinations of normal densities, we will use a different set of kernels. All the kernels we use have finite second moment and their corresponding characteristic function symmetric about zero. These kernels are selected from a book written by Nikolai Ushakov [6], and the numbers used are the same numbers as in the book. The kernels we used in this investigation are shown in table 3.1.

A note to table 3.1 is that for all kernels except the uniform (B27), x operates on the entire real line. For the uniform kernel the range of x is $-\sqrt{3} \leq x \leq \sqrt{3}$.
We see that all the expressions both for the kernels and their corresponding characteristic function are quite nice, and easy to interpret on the computer. This makes it possible to get exact results. To check the conjecture that $h_{AMISE} < h_{MISE}$ we used the standard normal distribution as the density, and $n = 150$.

Table 3.1: Kernels used in this project

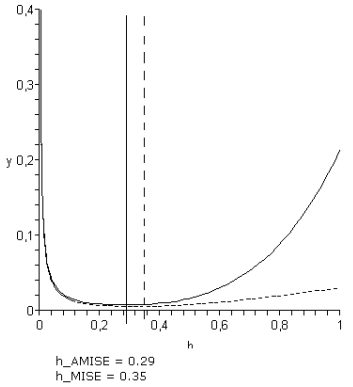
kernel number	*formula*	*characteristic function*
B5	$\frac{1}{2}e^{-\lvert x\rvert}$	$\frac{1}{1+t^2}$
B13	$\frac{1}{\pi cosh(x)}$	$\frac{1}{cosh(\frac{\pi t}{2})}$
B19	$\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$	$e^{-\frac{t^2}{2}}$
B27	$\frac{1}{2\sqrt{3}}$	$\frac{sin(\sqrt{3}t)}{\sqrt{3}t}$
B31	$\frac{x^2}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$	$(1-t^2)e^{-\frac{t^2}{2}}$
B33	$\frac{2}{\pi(1+x^2)^2}$	$(1+\lvert t\rvert)e^{-\lvert t\rvert}$
B35	$\frac{1}{4}e^{-\lvert x\rvert}(1+\lvert x\rvert)$	$\frac{1}{(1+t^2)^2}$
B38	$\frac{\pi}{4cosh^2(\frac{\pi x}{2})}$	$\frac{t}{sinh(t)}$
B39	$\frac{x}{2sinh(\frac{\pi x}{2})}$	$\frac{1}{cosh^2(t)}$

We calculated MISE and AMISE using the formulas we obtained in chapter 2.

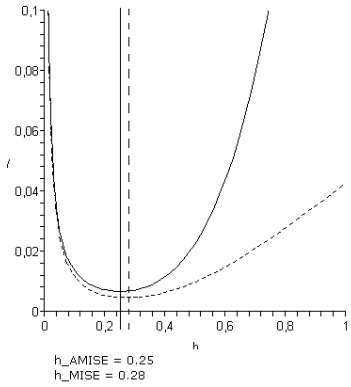
$$MISE(f_n(x;h)) = \frac{1}{2n\pi}\int (1-|\phi(t)|^2)\,|\psi(ht)|^2\,dt + \frac{1}{2\pi}\int |1-\psi(ht)|^2\,|\phi(t)|^2\,dt \tag{3.1}$$

$$AMISE(\widehat{f}(\cdot;h)) = (nh)^{-1}R(K) + \frac{1}{4}h^4\mu_2(K)^2R(f'') \tag{3.2}$$

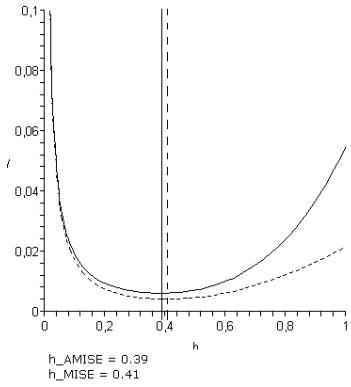
We plotted these functions against each other for all nine kernels, to see if the conjecture was correct for these examples. The result is shown in figure 3.1. Here the dashed line is AMISE, and the solid line is MISE.



(a) kernel B5

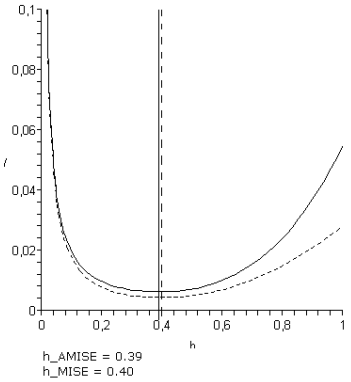


(b) kernel B13

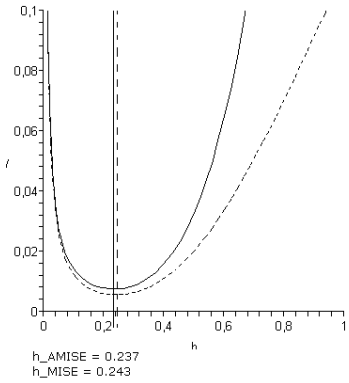


(c) kernel B19

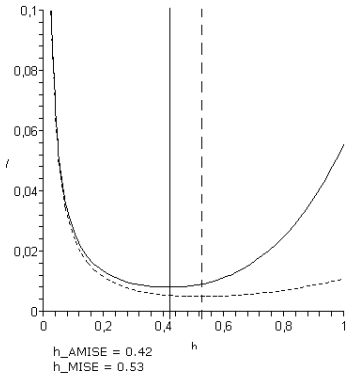
Figure 3.1: h_{AMISE} and h_{MISE}



(d) kernel B27

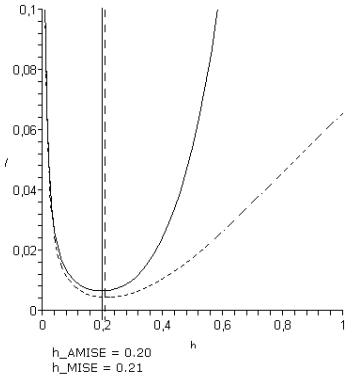


(e) kernel B31

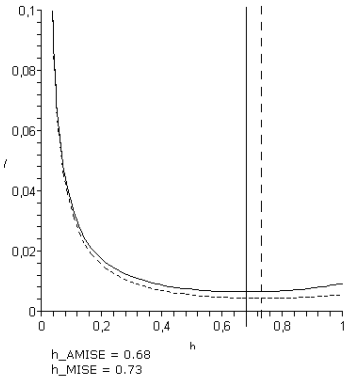


(f) kernel B33

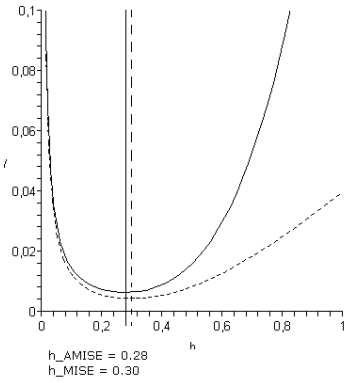
Figure 3.1: h_{AMISE} and h_{MISE}



(g) kernel B35



(h) kernel B38



(i) kernel B39

Figure 3.1: h_{AMISE} and h_{MISE}

As we can see from figure 3.1 the conjecture from Marron and Wand holds in all cases. The difference between h_{AMISE} and h_{MISE} varies a lot from kernel to kernel, and when using kernels B27, B31 and B35 they are almost equal. This means that for these kernels h_{AMISE} is a very good choice of the bandwidth. On the other hand we have kernels B33 and B38, where there is a larger difference between h_{AMISE} and h_{MISE}, and for these kernels h_{AMISE} may not be the best choice.
We also tried to change the sample size n. When n got smaller, both h_{AMISE} and h_{MISE} got bigger, and also the difference between them got bigger. When n got bigger than 150, h_{AMISE} and h_{MISE} got smaller, and the difference got smaller, but always $h_{AMISE} < h_{MISE}$. When we let n go to infinity, AMISE was almost equal to MISE as expected, and h_{AMISE} and h_{MISE} both approached zero.

Chapter 4

Comparison of derivatives of MISE and AMISE

The main idea for this chapter is to use derivatives of MISE and AMISE to investigate the conjecture of Marron and Wand. We want to compare these two derivatives, and try to find situations where $h_{AMISE} > h_{MISE}$. We also want to construct a method to select the bandwidth, by using these derivatives. We want to compare our method to the already existing method

$$h_{AMISE} = \left[\frac{R(K)}{\mu_2(K)^2 R(f'')n}\right]^{1/5} \tag{4.1}$$

and see if our method can be better, at least in some cases.
We will also in this chapter use the kernels listed in table 3.1, the density will be the standard normal distribution and the sample size is $n = 150$.

4.1 Comparison of derivatives of MISE and AMISE

The expressions for MISE and AMISE are as before

$$\begin{aligned} MISE = ISB + IV = \frac{1}{2\pi}\int |\phi(t)|^2 \, |1 - \psi(ht)|^2 \, dt + \\ \frac{1}{2\pi n}\int |\psi(ht)|^2 \, (1 - |\phi(t)|^2) dt \end{aligned} \tag{4.2}$$

$$AMISE = AIV + AISB = (nh)^{-1}R(K) + \frac{1}{4}h^4\mu_2(K)^2 R(f'') \tag{4.3}$$

We can rewrite (4.2) like

$$\begin{aligned} MISE &= \frac{1}{2\pi}\int |\phi(t)|^2 \, |1-\psi(ht)|^2 \, dt + \\ &\quad \frac{1}{2\pi n}\int |\psi(ht)|^2 - \\ &\quad \frac{1}{2\pi n}\int |\psi(ht)|^2 \, |\phi(t)|^2 \, dt \\ &= (nh)^{-1}R(K) + \frac{1}{2\pi}\int |1-\psi(ht)|^2 \, |\phi(t)|^2 \, dt - \\ &\quad \frac{1}{2\pi n}\int |\phi(t)|^2 \, |\psi(ht)|^2 \, dt \\ &= AIV + g(h) \end{aligned} \tag{4.4}$$

where

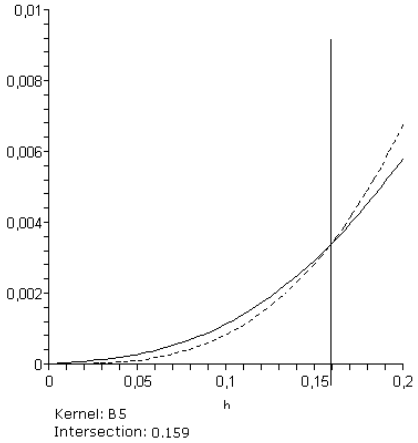
$$\begin{aligned} g(h) = \frac{1}{2\pi}\int |1-\psi(ht)|^2 \, |\phi(t)|^2 \, dt - \\ \frac{1}{2n\pi}\int |\phi(t)|^2 \, |\psi(ht)|^2 \, dt \end{aligned} \tag{4.5}$$

Here we use Parsevals identity (2.8). The advantage by writing MISE as (4.4) is that MISE and AMISE now has a common term. Comparing MISE and AMISE is therefore the same as comparing AISB and $g(h)$.
Since h_{MISE} is the minimum of MISE, it is a root of the derivative of MISE. The same thing holds for h_{AMISE}. Therefore the conjecture that $h_{MISE} > h_{AMISE}$ is correct if $(MISE)' < (AMISE)'$. This is equivalent to $g'(h) < (AISB)'$. We will try to check this second theory for our set of kernels.
We calculate the derivative for $g(h)$ and $(AISB)$.

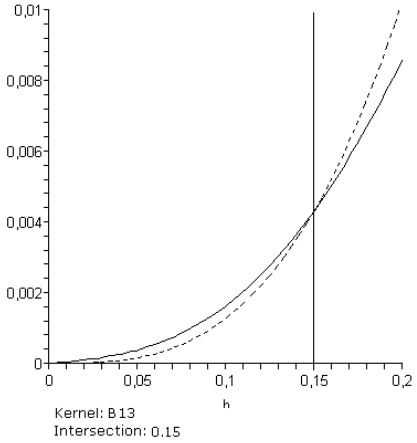
$$
\begin{aligned}
g'(h) = & \frac{1}{2\pi}\int 2t\left|1-\psi(ht)\right|^2\left(-\psi'(ht)\left|\phi(t)\right|^2 dt-\right.\\
& \frac{1}{2n\pi}\int 2t\left|\phi(t)\right|^2\psi'(ht)\left|\psi(ht)\right|^2 dt\\
= & \frac{1}{\pi}\int t|\phi(t)|^2(-\psi'(ht))dt-\\
& \frac{1}{\pi}\int t|\phi(t)|^2\psi(ht)(-\psi'(ht))dt-\\
& \frac{1}{n\pi}\int t|\phi(t)|^2\psi(ht)(\psi'(ht))dt\\
= & \frac{1}{\pi}\Bigg[-\int t|\phi(t)|^2\psi'(ht)dt+\\
& (1-\frac{1}{n})\int t|\phi(t)|^2\psi(ht)\psi'(ht)dt\Bigg]
\end{aligned}
\tag{4.6}
$$

$$
(AISB)' = h^3\mu_2(K)^2R(f'') \tag{4.7}
$$

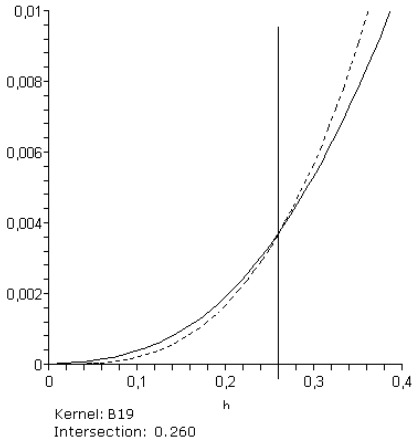
From here on we will call $(AISB)'$ for $g_0'(h)$.
We plotted $g'(h)$ and $g_0'(h)$ for all kernels (see figure 4.1). For all graphs the solid line is the graph of $g'(h)$, and the dashed line is the graph of $g_0'(h)$.



(a) kernel B5

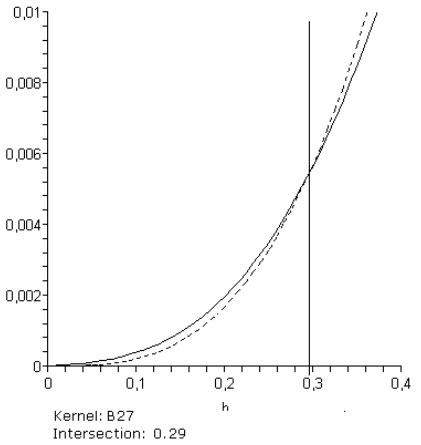


(b) kernel B13

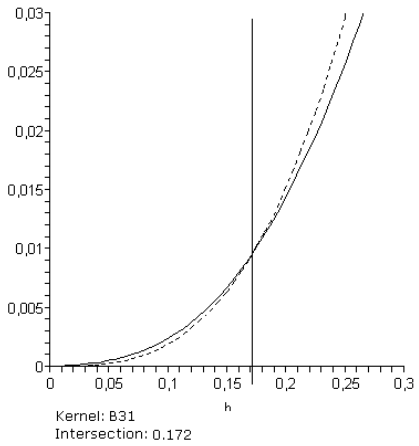


(c) kernel B19

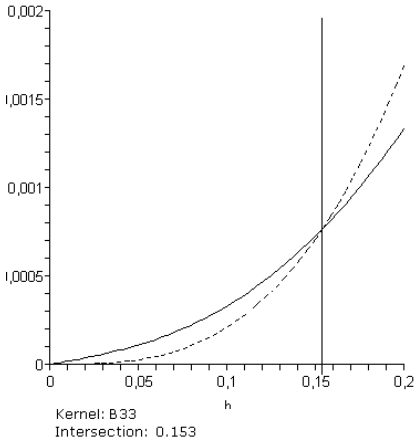
Figure 4.1: $g'(h)$(solid line) and $g'_0(h)$(dashed line)



(d) kernel B27

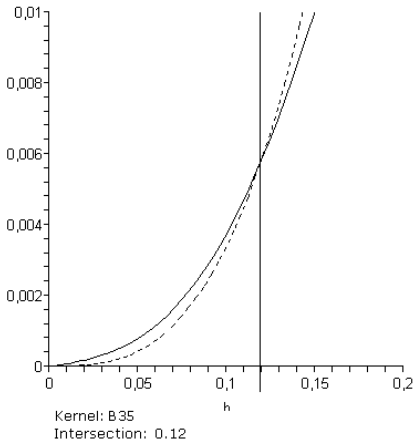


(e) kernel B31

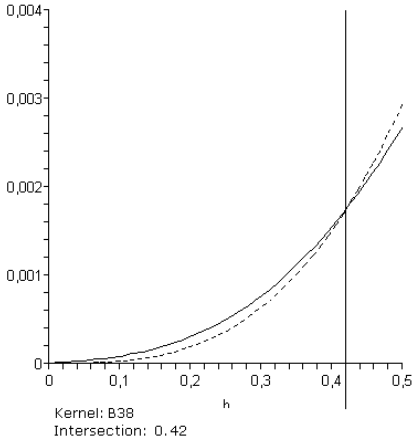


(f) kernel B33

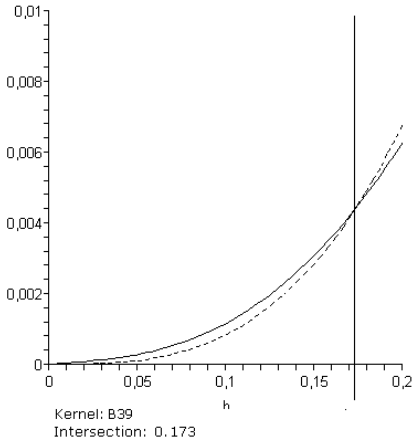
Figure 4.1: $g'(h)$(solid line) and $g_0'(h)$(dashed line)



(g) kernel B35



(h) kernel B38



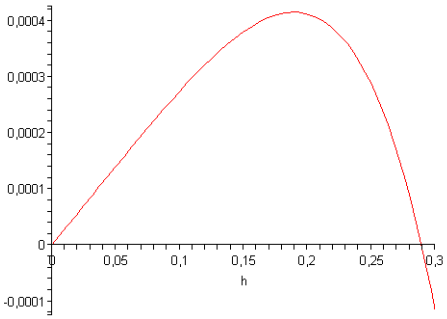
(i) kernel B39

Figure 4.1: $g'(h)$(solid line) and $g'_0(h)$(dashed line)

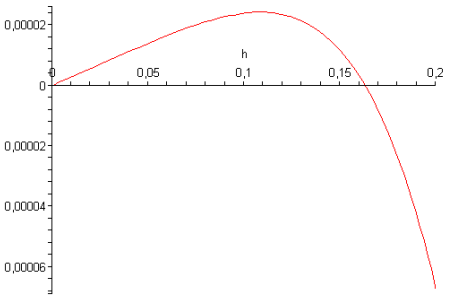
All figures show the same pattern. For small h, $g'(h)$ is greater than $g_0'(h)$, but from a certain h, which depends on the kernel, $g_0'(h)$ is bigger. This h that intersects the two graphs is always smaller than the respective h_{AMISE} (see figure 3.1 for values of h_{AMISE}). This means that $(MISE)' < (AMISE)'$ holds, which again means that the conjecture of Marron and Wand is true in these examples.
We have already shown that h_{AMISE} and h_{MISE} tends to zero as n tends to infinity. If we could show that the intersection between $g'(h)$ and $g_0'(h)$ did not tend to zero as n tends to infinity, this would mean that for some big n, h_{AMISE} and h_{MISE} will be smaller than this intersection, and h_{MISE} would be smaller than h_{AMISE}, which would have proved the conjecture wrong. We calculated this intersection and checked what happened when we increased n very much. The kernel and the density is both the standard normal distribution. Figure 4.2 draws $g'(h) - g_0'(h)$, so the intersection is were the graph is zero.

As we can see from figure 4.2 it seems like the intersection tends to zero, as n tends to infinity, and therefore this is not a contradiction to the conjecture. This is also the case when using the other kernels. When n got very big, the computer got some problems with the calculation, and the graphs were not smooth.

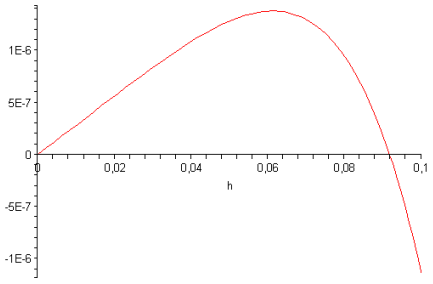
Based on our investigation it seems like the conjecture that $h_{AMISE} < h_{MISE}$ might be correct. Even though it is not proved, we have made many examples using different kernels and different methods, and always the conjecture is true.



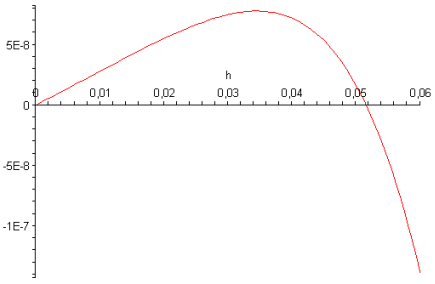
(a) n=10E1



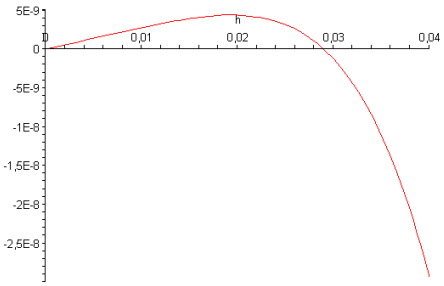
(b) n=10E2



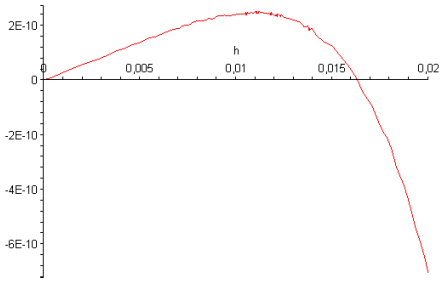
(c) n=10E3



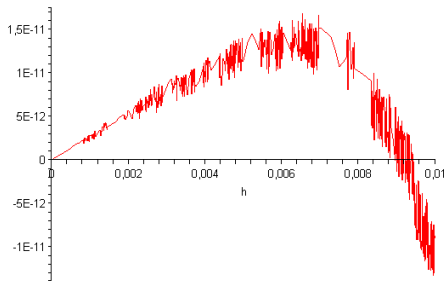
(d) n=10E4



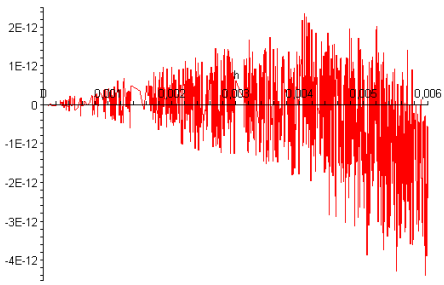
(e) n=10E5



(f) n=10E6



(g) n=10E7



(h) n=10E8

Figure 4.2: $g'(h) - g'_0(h)$

4.2 A new method of bandwidth selection?

The idea of this chapter is to try to find a new method of bandwidth selection by using the derivatives of MISE and AMISE. By computing an upper bound for $g'(h)$, and finding the root of this upper bound, which we call h_{UB}, we have another way of selecting the bandwidth. Then we compare h_{UB} with the already existing method h_{AMISE} and we check if our method is better.
To develop such a method of bandwidth selection, we first find an upper bound for $g'(h)$ and compare this upper bound with $g_0'(h)$. Then we will find if h_{UB} is closer to h_{MISE} than h_{AMISE}. If h_{UB} is closer, it will surely be a better choice of the bandwidth.
To find this upper bound we need to prove the following two inequalities.

$$\psi(t) \geq 1 - \frac{\mu_2(K)t^2}{2}, \forall t \tag{4.8}$$

$$-\psi'(t) \leq t\mu_2(K), \forall t \geq 0 \tag{4.9}$$

To prove these two inequalities we need to prove some properties of $\cos x$ and $\sin x$ first.

$$\cos x \geq 1 - (x^2/2), \forall x$$

$$\sin x \leq x, \forall x \geq 0$$

Proof:
We start with the second one. If $g(x) = x - sin(x)$ we see that $g'(x) = 1 - cos(x) \geq 0$.
We also see that $g(0) = 0$.
This means that $g(x)$ is always increasing and for $x \geq 0$ $g(x) \geq 0$, which implies that

$$\sin x \leq x, \forall x \geq 0$$

Let us then prove the first inequality. Let $f(x) = cos(x) - 1 - \frac{x^2}{2}$.
We get that $f'(x) = x - sin(x) \geq 0$ $\forall x \geq 0$ and
$f'(x) = x - sin(x) \leq 0$ $\forall x \leq 0$.
We also see that $f(0) = 0$.
This means that $f(x)$ is increasing for $x \geq 0$ and decreasing for $x \leq 0$ and therefore is always positive. This again implies that

$$\cos x \geq 1 - (x^2/2), \forall x$$

Now we can prove (4.8) and (4.9).

$$\begin{aligned}
\psi(t) &= \int_{-\infty}^{\infty} \cos(tx)K(X)dx \\
&\geq \int_{-\infty}^{\infty} (1 - \frac{t^2x^2}{2}K(x))dx \\
&= \int_{-\infty}^{\infty} K(x)dx - \frac{t^2}{2}\int_{-\infty}^{\infty} x^2K(x)dx \\
&= 1 - \frac{t^2}{2}\mu_2(K) \\
-\psi'(t) &= \int_{-\infty}^{\infty} \sin(tx)xK(x)dx \\
&\leq t\int_{-\infty}^{\infty} x^2 * K(x))dx \\
&= t\mu_2(K)
\end{aligned}$$

Now we can use (4.8) and (4.9) to find an upper bound for $g'(h)$, by inserting these two bounds into the expression for $g'(h)$.

$$
\begin{aligned}
g'(h) &= \frac{1}{\pi}\Bigg[- \int t \left|\phi(t)\right|^2 \psi'(ht)dt + \\
&\quad \left(1-\frac{1}{n}\right) \int t \left|\phi(t)\right|^2 \psi(ht)\psi'(ht)dt \Bigg] \\
&\leq \frac{1}{\pi}\Bigg[\int ht^2 \left|\phi(t)\right|^2 \mu_2(K)dt + \\
&\quad \left(1-\frac{1}{n}\right) \int t \left|\phi(t)\right|^2 \psi(ht)\psi'(ht)dt \Bigg] \\
&\leq \frac{1}{\pi}\Bigg[\int ht^2 \left|\phi(t)\right|^2 \mu_2(K)dt + \\
&\quad \left(1-\frac{1}{n}\right) \int ht^2 \left|\phi(t)\right|^2 (-\psi(ht))\mu_2(K)dt \Bigg] \\
&\leq \frac{1}{\pi}\Bigg[\int ht^2 \left|\phi(t)\right|^2 \mu_2(K)dt + \\
&\quad \left(1-\frac{1}{n}\right) \int ht^2 \left|\phi(t)\right|^2 \left(-\left(1-\frac{\mu_2(K)h^2t^2}{2}\right)\right) \mu_2(K)dt \Bigg] \\
&= \frac{1}{\pi}\Bigg[\frac{1}{n} \int ht^2 \left|\phi(t)\right|^2 \mu_2(K)dt + \\
&\quad \frac{1}{2}\left(1-\frac{1}{n}\right) \int h^3t^4\mu_2(K)^2 \left|\phi(t)\right|^2 dt \Bigg] \\
&= \frac{1}{\pi}\Bigg[\frac{1}{n} h\mu_2(K) \int t^2 \left|\phi(t)\right|^2 dt + \\
&\quad \left(1-\frac{1}{n}\right) h^3\mu_2(K)^2 \int t^4 \left|\phi(t)\right|^2 dt \Bigg] \\
&= \frac{1}{n} 2h\mu_2(K)R(f') + \left(1-\frac{1}{n}\right) h^3\mu_2(K)^2 R(f'') \\
&= \frac{1}{n} 2h\mu_2(K)R(f') + \left(1-\frac{1}{n}\right) g_0'(h)
\end{aligned}
\tag{4.10}
$$

We have used Parsevals identity for derivatives (2.10).

We compared this upper bound with the $g_0'(h)$ which is

$$g_0'(h) = h^3 \mu_2(K)^2 R(f'')$$

The result is that these two functions are very similar, and the graph of the two functions looks like figure 3.3.

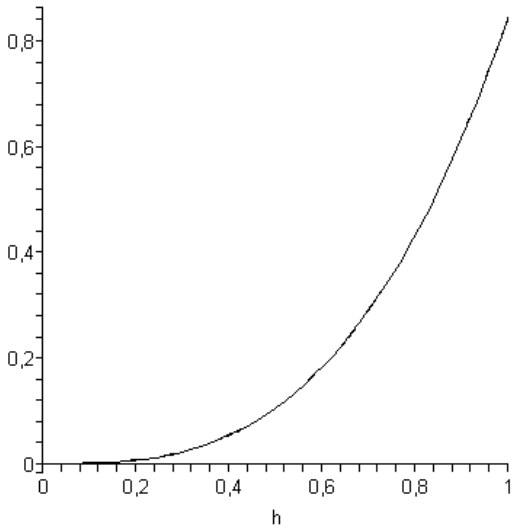


Figure 4.3: $g'(h)$ and $g_0'(h)$

As we can see it is impossible to separate the graphs by looking at the figure. A more detailed analysis of the function shows that the upper bound of $g'(h)$ is bigger than $g_0'(h)$ when h is small, and vice versa when h is big. The number where the intersection is, depends only on the second moment of the kernel, since $g'(h)$ is independent on n. The bigger second moment, the smaller h at the intersection. Table 4.1 shows the value of h at the intersection.

Table 4.1: Values of intersection at different second moments

kernel	*second moment*	*intersection*
B5	2	0.816497
B13	2.46	0,73621017
B19	1	1.1547005
B27	1	1.1547005
B31	3	0.6666667
B33	1	1.1547005
B35	4	0.5773503
B38	0.3333	1.99999997
B39	2	0.816497

Now we know that the derivative of the upper bound of $g'(h)$ is greater than $g'_0(h)$ when h is less than some value (see table 4.1). Unfortunately this value is greater than h_{AMISE} for all n, so the bandwidth selected from this derivative h_{UB}, is not a better choice than h_{AMISE}. If we could make h_{AMISE} greater than this intersection, it would mean that the intersection is below the zero line, and the bandwidth selected by this derivative, would have been better. The situation is explained in figure 4.4

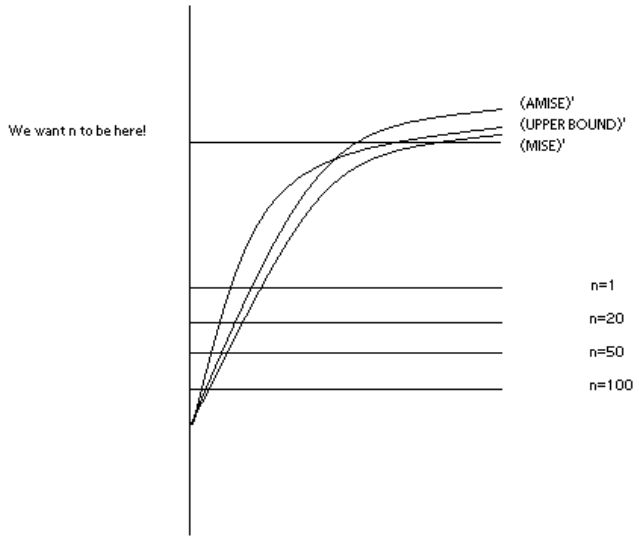


Figure 4.4: Our situation

Figure 4.4 is just an explanation of the situation, and not a real graph.
We have found that in the case when both the kernel and the density is standard normal, the intersection lies at 1.15. h_{AMISE} is 1.05 when $n = 1$, and 0.42 at $n = 100$.
The same situation exists for the kernel B5. The intersection is at 0.816, and h_{AMISE} is 0.784 when $n = 1$, and 0.312 when $n = 100$. The same situation also holds for the rest of the kernels. The intersection is higher than h_{AMISE} for all n. This means that our new method of choosing the bandwidth is not as good as h_{AMISE}, and also it has a more complicated form.
The conclusion is therefore that selecting the bandwidth by using this method gives a result which is close, but not as good as h_{AMISE}. To use this method it is necessary to have access to a computer, since the equations for $(MISE)'$ and $(AMISE)'$ are quite difficult.

Chapter 5

A special case: $n = 1$

We have considered the conjecture of Marron and Wand [5], that $h_{MISE} > h_{AMISE}$ in several different ways, but we have not managed to find examples when it is not true. Therefore we will in this chapter try to prove the conjecture, instead of finding counter examples. We will try to prove this conjecture in the simplest case when $n = 1$. This has no practical interest since the sample size always is greater than 1, but it has some theoretical interest.
We assume that

$$\int |\phi(t)|^2 \, dt = 1 \tag{5.1}$$

We can make this assumption without loss of generality, since this is only a change of a scaling factor.
When $n = 1$ we have the following expressions for $g_0'(h)$ and $g'(h)$

$$g_0'(h) = h^3 \mu_2(K)^2 R(f'') \tag{5.2}$$

$$g'(h) = \frac{1}{\pi} \int t \, |\phi(t)|^2 \, (-\psi'(ht)) dt \tag{5.3}$$

Let h_1 be the solution of the equation $g'(h) = g_0'(h)$.
Let h_2 be the solution of the equation $g_0'(h) = \frac{R(K)}{h^2}$.
This means that $h_2 = h_{AMISE}$. Now we see from figure 5.1 that the conjecture is correct if we can prove that $h_1 < h_2$.

We have to find expressions for h_1 and h_2. This is easy for h_2. It becomes

$$h_2 = \left(\frac{R(K)}{\mu_2(K)^2 R(f'')} \right)^{\frac{1}{5}} \tag{5.4}$$

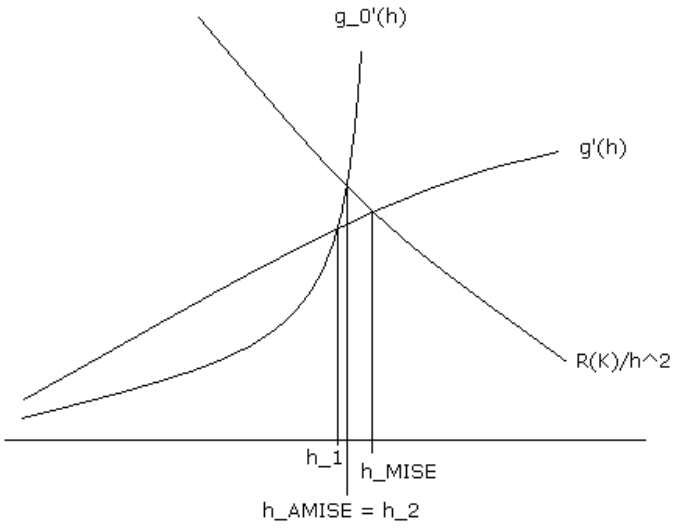


Figure 5.1: Our situation

To find an explicit form of h_1 is not possible, so we need to find a simple $g_1'(h)$ such that

$$g'(h) \leq g_1'(h)$$

and

$$g_1'(h) \geq g_0'(h) \forall h \leq h_3$$
$$g_1'(h) \leq g_0'(h) \forall h \geq h_3$$

Here, h_3 is the solution of $g_1'(h) = g_0'(h)$. Evidently we have that $h_1 \leq h_3$. Therefore if

$$h_3 \leq h_2 \tag{5.5}$$

holds, than the conjecture is correct.
To find this $g_1'(h)$ we need a special form of Jensen's inequality. It states that if $f(x)$ is a non-negative function such that

$$\int_{-\infty}^{\infty} f(x)dx = 1 \tag{5.6}$$

g is any real-valued measurable function and τ is convex over the range of g, then

$$\tau\left(\int_{-\infty}^{\infty} g(x)f(x)dx\right) \leq \int_{-\infty}^{\infty} \tau(g(x))f(x)dx \tag{5.7}$$

For proof of the general form of this inequality look in Casella and Berger [1]
We have

$$\begin{aligned} g'(h) &= \frac{1}{\pi}\int t\left|\phi(t)\right|^2\left(-\psi'(ht)\right)dt \\ &\leq \frac{1}{\pi}h\mu_2(K)\int t^2\left|\phi(t)\right|^2 dt \\ &= \frac{1}{\pi}h\mu_2(K)\left(\left(\int t^2\left|\phi(t)\right|^2 dt\right)^2\right)^{\frac{1}{2}} \\ &\leq \frac{1}{\pi}h\mu_2(K)\left(\int t^4\left|\phi(t)\right|^2\right)^{\frac{1}{2}} \\ &= \frac{1}{\pi}h\mu_2(K)\left(2\pi R(f'')\right)^{\frac{1}{2}} \\ &= \sqrt{\frac{2}{\pi}}h\mu_2(K)\sqrt{R(f'')} = g_1'(h) \end{aligned} \tag{5.8}$$

Here we have used (4.9), Jensen's inequality (5.6) and Parseval identity (2.10).
Now we can derive a nice expression for h_3, since h_3 is the solution of $g_1'(h) = g_0'(h)$. It is

$$h_3 = \left(\frac{2}{\pi R(f'')\mu_2(K)^2}\right)^{\frac{1}{4}} \tag{5.9}$$

Now we have expressions for both h_2 and h_3, and to prove the conjecture, we must prove (5.5)

$$\begin{aligned} h_3 &\leq h_2 \\ \left(\frac{2}{\pi R(f'')\mu_2(K)^2}\right)^{\frac{1}{4}} &\leq \left(\frac{R(K)}{\mu_2(K)^2 R(f'')}\right)^{\frac{1}{5}} \\ \left(\frac{2}{\pi}\right)^{\frac{1}{4}} &\leq \left(R(K)^2\mu_2(K)\right)^{\frac{1}{10}} R(f'')^{\frac{1}{20}} \end{aligned} \tag{5.10}$$

We need lower bounds for $\left(R(K)^2\mu_2(K)\right)^{\frac{1}{10}}$ and $R(f'')^{\frac{1}{20}}$.
According to [4], chapter 2.7, the Epanechnikov kernel is the solution that minimizes $\left(R(K)^4\mu_2(K)^2\right)^{\frac{1}{5}}$ and therefore also it is the solution that minimizes $\left(R(K)^2\mu_2(K)\right)^{\frac{1}{10}}$, so the lower bound we will use for this expression

is $\left(R(K^*)^2\mu_2(K^*)\right)^{\frac{1}{10}}$, where $K^* = \frac{3}{4}(1-x^2)$ which is the Epanechnikov kernel.
To obtain a lower bound of $R(f'')$ we use Parseval identity (2.10) to obtain

$$R(f'') = \frac{1}{2\pi}\int t^4 \left|\phi(t)\right|^2 dt \tag{5.11}$$

To find a lower bound for this integral we use the following function for $\phi(t)$

$$\phi(t) = \begin{cases} 1 & \text{if } |t| \leq \frac{1}{2} \\ 0 & \text{if } |t| \geq \frac{1}{2} \end{cases}$$

This function is not a characteristic function, but the bound we get from this function is even lower than any characteristic function can give us, so the function will be good to use in our example.
Proof that $\phi(t)$ gives the lower bound for $R(f'')$:
Let $f(t) \geq 0$ and $f_0(t) \geq 0$. Let also

$$\begin{aligned} f(t) &\leq f_0(t) \forall |t| \leq \frac{1}{2} \\ f(t) &\geq f_0(t) \forall |t| > \frac{1}{2} \end{aligned} \tag{5.12}$$

Finally let $\int f(t)dt = \int f_0(t)dt = 1$.
We want to prove that

$$\int t^4 f(t)dt \geq \int t^4 f_0(t)dt \tag{5.13}$$

Let us look at the difference between the to integrals

$$\int t^4 f(t)dt - \int t^4 f_0(t)dt = \int t^4(f(t) - f_0(t))dt$$
$$= \int_{|t|\leq\frac{1}{2}} t^4(f(t) - f_0(t))dt + \int_{|t|>\frac{1}{2}} t^4(f(t) - f_0(t))dt$$

The first part of this sum is always negative because of (5.12), and the second part is always positive, also because of (5.12). This means that our proof is correct if

$$\left|\int_{|t|\leq\frac{1}{2}} t^4(f(t)-f_0(t))dt\right| \leq \int_{|t|>\frac{1}{2}} t^4(f(t)-f_0(t))dt$$
$$\int_{|t|\leq\frac{1}{2}} t^4(f_0(t)-f(t))dt \leq \int_{|t|>\frac{1}{2}} t^4(f(t)-f_0(t))dt$$

Now we use the following two obvious inequalities

$$\int_{|t|\leq\frac{1}{2}} t^4(f_0(t)-f(t))dt \leq \left(\frac{1}{2}\right)^4 \int_{|t|\leq\frac{1}{2}} (f_0(t)-f(t))dt$$
$$\int_{|t|>\frac{1}{2}} t^4(f(t)-f_0(t))dt \geq \left(\frac{1}{2}\right)^4 \int_{|t|>\frac{1}{2}} (f(t)-f_0(t))dt$$

The proof will now be correct if

$$\left(\frac{1}{2}\right)^4 \int_{|t|\leq\frac{1}{2}} (f_0(t)-f(t))dt = \left(\frac{1}{2}\right)^4 \int_{|t|>\frac{1}{2}} (f(t)-f_0(t))dt$$

We use the fact that $\int f(t)dt = \int f_0(t)dt = 1$, and see that

$$\int_{|t|\leq\frac{1}{2}} (f_0(t)-f(t))dt + \int_{|t|>\frac{1}{2}} (f_0(t)-f(t))dt = 0$$
$$\int_{|t|\leq\frac{1}{2}} (f_0(t)-f(t))dt - \int_{|t|>\frac{1}{2}} (f(t)-f_0(t))dt = 0$$

So the proof is correct. The only function $f_0(t)$ that fulfills this property for any other function $f(t)$ is $\phi(t)$.
This gives the following lower bound for $R(f'')^{\frac{1}{20}}$

$$R(f'')^{\frac{1}{20}} \geq \left(\frac{1}{2\pi}\int_{-\frac{1}{2}}^{\frac{1}{2}} t^4\right)^{\frac{1}{20}} \tag{5.14}$$

When we insert numbers we get the following result

$$h_3 \leq h_2$$
$$\left(\frac{2}{\pi}\right)^{\frac{1}{4}} \leq \left(R(K^*)^2\mu_2(K^*)\right)^{\frac{1}{10}} \left(\frac{1}{2\pi}\int_{-\frac{1}{2}}^{\frac{1}{2}} t^4\right)^{\frac{1}{20}} \tag{5.15}$$
$$0.8932438417 \leq 0.7686581736 \cdot 0.7327174038$$
$$0.8932438417 \leq 0.5632092214$$

We see that the inequality is incorrect, and the proof is wrong.

We will try the same procedure to establish this proof once more, but now we will use a different inequality for $\psi'(ht)$. We will use that

$$\psi'(ht) \leq \mu_1(K) \tag{5.16}$$

Proof:

$$\begin{aligned} \psi(ht) &= \int cos(htx)K(x)dx \\ \psi'(ht) &= -\int |x|sin(htx)K(x)dx \\ -\psi'(ht) &\leq \int |x|K(x) \\ &= \mu_1(K) \end{aligned} \tag{5.17}$$

We find a new expression for $g_1'(h)$ using this inequality and the fact that $\mu_1(K) \leq (\mu_2(K))^{\frac{1}{2}}$.

$$\begin{aligned} g'(h) &= \frac{1}{\pi}\int t\left|\phi(t)\right|^2(-\psi'(ht))\,dt \\ &\leq \frac{1}{\pi}\mu_1(K)\int t\left|\phi(t)\right|^2 dt \\ &= \frac{1}{\pi}\mu_1(K)\left(\left(\int t\left|\phi(t)\right|^2 dt\right)^4\right)^{\frac{1}{4}} \\ &\leq \frac{1}{\pi}\mu_2(K)^{\frac{1}{2}}\left(\int t^4\left|\phi(t)\right|^2\right)^{\frac{1}{4}} \\ &= \frac{1}{\pi}\mu_2(K)^{\frac{1}{2}}\left(2\pi R(f'')\right)^{\frac{1}{4}} \\ &= \frac{1}{\pi}\mu_2(K)^{\frac{1}{2}}\left(2\pi R(f'')\right)^{\frac{1}{4}} = g_1'(h) \end{aligned} \tag{5.18}$$

Here we have used (5.17), Jensen's inequality (5.6) and Parseval identity (2.10).
The expression for h_2 is still

$$h_2 = \left(\frac{R(K)}{\mu_2(K)R(f'')}\right)^{\frac{1}{5}}$$

Now we can derive a new expression for h_3 from the equation $g_1'(h) = g_0'(h)$. It is

$$h_3 = \left[\left(\frac{2}{\pi^3} \right)^{\frac{1}{12}} (\mu_2(K)^2 R(f''))^{-\frac{1}{4}} \right] \tag{5.19}$$

Now we again insert the expressions for h_2 and h_3 in (5.5)

$$\begin{aligned} h_3 &\leq h_2 \\ \left[\left(\frac{2}{\pi^3} \right)^{\frac{1}{12}} \left(\mu_2(K)^2 R(f'')\right)^{-\frac{1}{4}} \right] &\leq \left(\frac{R(K)}{\mu_2(K)^2 R(f'')} \right)^{\frac{1}{5}} \\ \left(\frac{2}{\pi^3} \right)^{\frac{1}{12}} &\leq \left(R(K)^2 \mu_2(K)\right)^{\frac{1}{10}} R(f'')^{\frac{1}{20}} \end{aligned} \tag{5.20}$$

We need lower bounds for the expression on the right side, and we use, as we did earlier in this chapter, the Epanechnikov kernel to find a lower bound for $\left(R(K)^2 \mu_2(K)\right)^{\frac{1}{10}}$, and the fact that $R(f'')^{\frac{1}{20}} \geq 0.7327174038$. When we insert numbers we get

$$0.7957897933 \leq 0.5632092213 \tag{5.21}$$

As we can see, the inequality is again incorrect, and the proof is wrong.
The conclusion to this chapter is that despite a good effort, we were not able to prove this conjecture, even in the case when $n = 1$. This does not mean that the conjecture is wrong, but since we have used quite sensitive inequalities, this means that, even if the conjecture is correct, there is no simple proof of it. Our results also mean that if the conjecture is not correct, it is not correct only for very special, irregular kernels and densities.

Chapter 6

Sinc kernel

6.1 Bandwidth Selection

So far in this report we have dealt with standard choices of the kernel. Kernels that are symmetric, and have finite second moment. In this chapter we shall use another kernel, which is not standard because it takes negative values. This kernel is the sinc kernel.

$$K(x) = \frac{\sin x}{\pi x} \tag{6.1}$$

The characteristic function of the sinc kernel is

$$\psi(t) = \begin{cases} 1 & \text{for } |t| \leq 1 \\ 0 & \text{for } |t| > 1 \end{cases}$$

According to [3] the expressions for the sinc estimator and its characteristic function are

$$f_n(x;h) = \frac{1}{\pi n} \sum_{j=1}^{n} \frac{\sin\left[(x - X_j)/h\right]}{x - X_j} \tag{6.2}$$

$$cf = \phi_n(t)\psi(ht) = \phi_n(t) I_{[-1/h,1/h]}(t) \tag{6.3}$$

where $\phi_n(t)$ is the empirical characteristic function associated with the sample $X_1, ..., X_n$, and $I_A(t)$ is the indicator function of the set A.

Also according to [3] we have the following expression for MISE of the sinc

estimator.

$$MISE(f_n(x;h)) = \frac{1}{\pi nh} + R(f) - \left(1 + \frac{1}{n}\right)\frac{1}{\pi}\int_0^{1/h} |\phi(t)|^2 dt \qquad (6.4)$$

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

We want to investigate the bandwidth selection based on this estimator. We want to compare the result we get from the expression of MISE, with the result we get when we select the bandwidth using the empirical characteristic function. We know from [5] that the difference between h_{MISE} and h_{AMISE} when the kernel is gaussian can be quite big for some densities. Sometimes h_{MISE} is 5 times as big as h_{AMISE}. This makes h_{AMISE} a poor estimator for the bandwidth. We want to investigate if the empirical characteristic function based on the sinc estimator gives a better estimation for the bandwidth.
The optimal bandwidth based on the sinc estimator is given as (for proof see [3])

$$|\phi(1/h)| = \frac{1}{\sqrt{n+1}} \qquad (6.5)$$

The method based on the empirical characteristic function is (for proof see [3])

$$|\phi_n(1/h)| = \frac{1}{\sqrt{n+1}} \qquad (6.6)$$

We have calculated these two equalities for large, medium and small sample sizes and using six of the densities from Marron and Wand [5].

The last density, number 11, is difficult to simulate with small sample sizes, since its expression contains very small fractions which requires big sample sizes to simulate. Therefore this is only considered in the case when n is large. Table 6.2 - 6.4 shows the first solution of these two equations, and the sample size. Figure 6.1 shows the graphs for the characteristic function and the empirical characteristic function. The horizontal line shows the value of $\frac{1}{\sqrt{n+1}}$.

Table 6.1: Densities from Marron & Wand

Density	*Formula*
#1	N(0,1)
#2	$\frac{1}{5}N(0,1) + \frac{1}{5}N(\frac{1}{2}, \frac{2}{3}) + \frac{3}{5}N(\frac{13}{12}, \frac{5}{9})$
#4	$\frac{2}{3}N(0,1) + \frac{1}{3}N(0, \frac{1}{10})$
#6	$\frac{1}{2}N(-1, \frac{2}{3}) + \frac{1}{2}N(1, \frac{2}{3})$
#10	$\frac{1}{2}N(0,1) + \sum_{l=0}^{4} \frac{1}{10}N(\frac{l}{2} - 1, \frac{1}{10})$
#11	$\frac{49}{100}N(-1, \frac{2}{3}) + \frac{49}{100}N(1, \frac{2}{3}) + \sum_{l=0}^{6} \frac{1}{350}N(\frac{l-3}{2}, \frac{1}{100})$

Table 6.2: Solutions based on the c.f and the e.c.f for different densities with large n

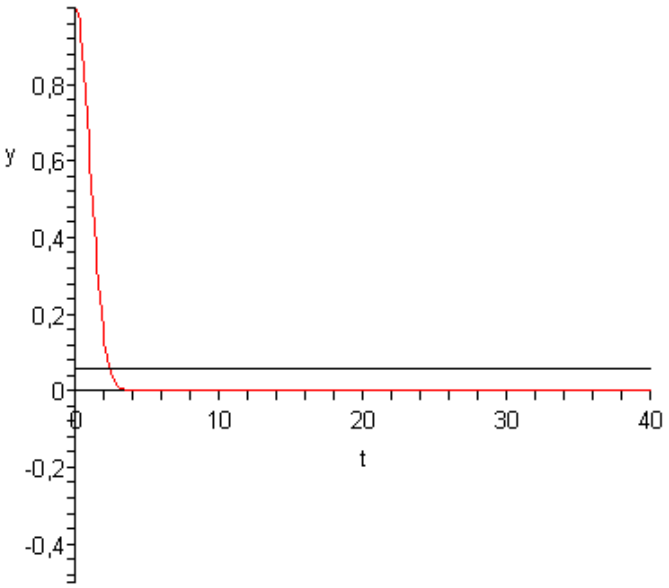
Density	$\lvert\phi\left(\frac{1}{h}\right)\rvert = \frac{1}{\sqrt{(n+1)}}$	$\lvert\phi_n\left(\frac{1}{h}\right)\rvert = \frac{1}{\sqrt{(n+1)}}$	n
#1	0.4202	0.4545	300
#2	0.6173	0.6250	300
#4	0.0532	0.0518	300
#6	0.6757	0.6993	300
#10	0.4202	0.4545	300
#11	0.6623	0.6410	700

Table 6.3: Solutions based on the c.f and the e.c.f for different densities with $n = 100$

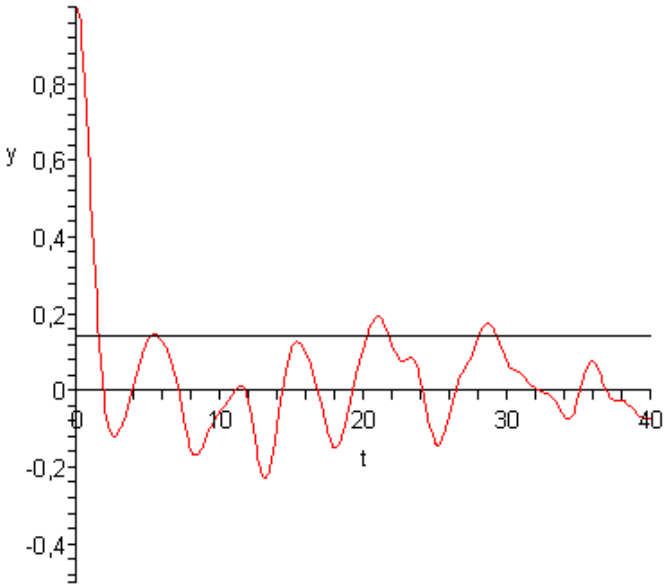
Density	$\lvert\phi\left(\frac{1}{h}\right)\rvert = \frac{1}{\sqrt{(n+1)}}$	$\lvert\phi_n\left(\frac{1}{h}\right)\rvert = \frac{1}{\sqrt{(n+1)}}$	n
#1	0.4651	0.4878	100
#2	0.6410	0.6757	100
#4	0.0641	0.0617	100
#6	0.7042	0.7299	100
#10	0.4444	0.4651	100

Table 6.4: Solutions based on the c.f and the e.c.f for different densities with $n = 50$

Density	$\lvert\phi\left(\frac{1}{h}\right)\rvert = \frac{1}{\sqrt{(n+1)}}$	$\lvert\phi_n\left(\frac{1}{h}\right)\rvert = \frac{1}{\sqrt{(n+1)}}$	n
#1	0.5051	0.6410	50
#2	0.6667	0.5988	50
#4	0.0758	0.0917	50
#6	0.7246	0.7042	50
#10	0.4695	0.4545	50

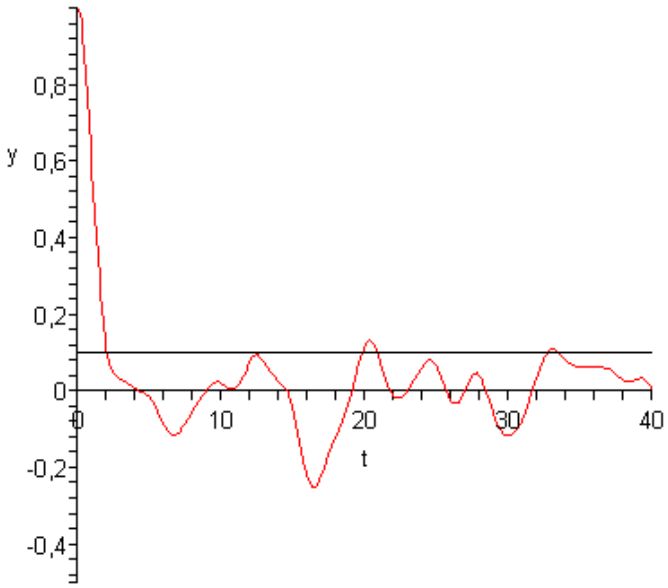


(a) ch.function of density 1

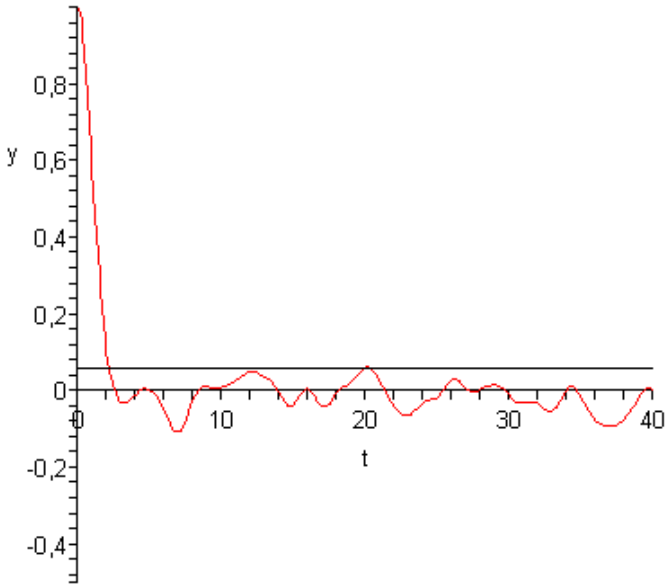


(b) emp.ch.function of density 1, n=50

Figure 6.1: ch.function and emp.ch.function for density nr.1

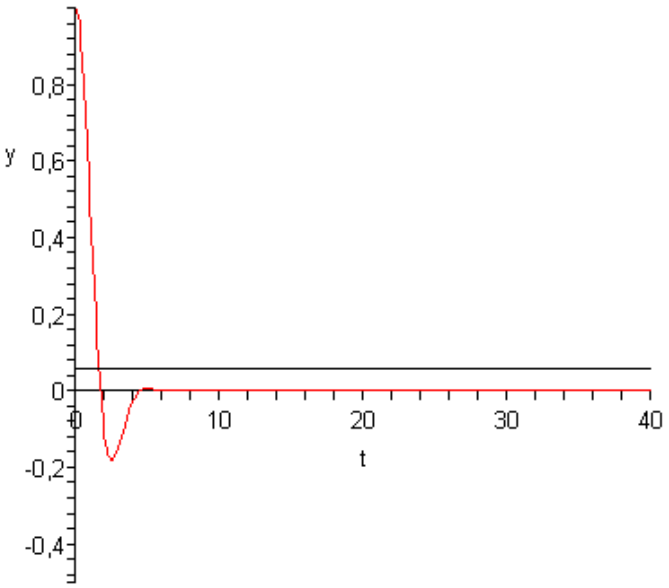


(c) emp.ch.function of density 1, n=100

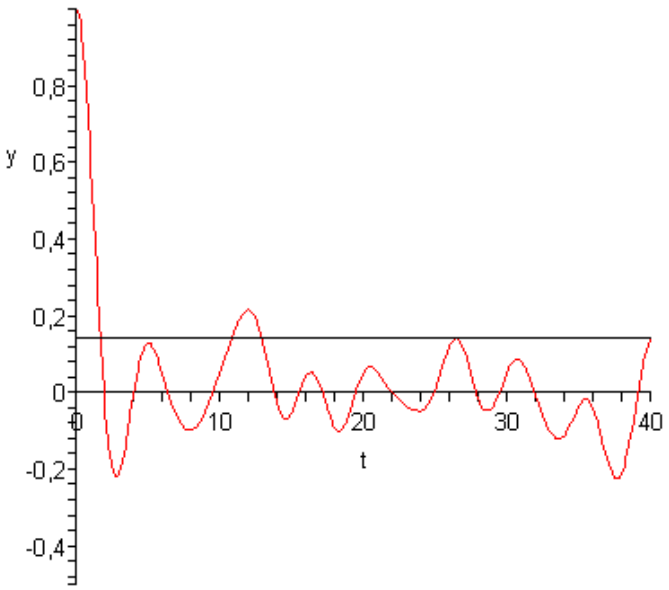


(d) emp.ch.function of density 1, n=300

Figure 6.1: ch.function and emp.ch.function for density nr.1

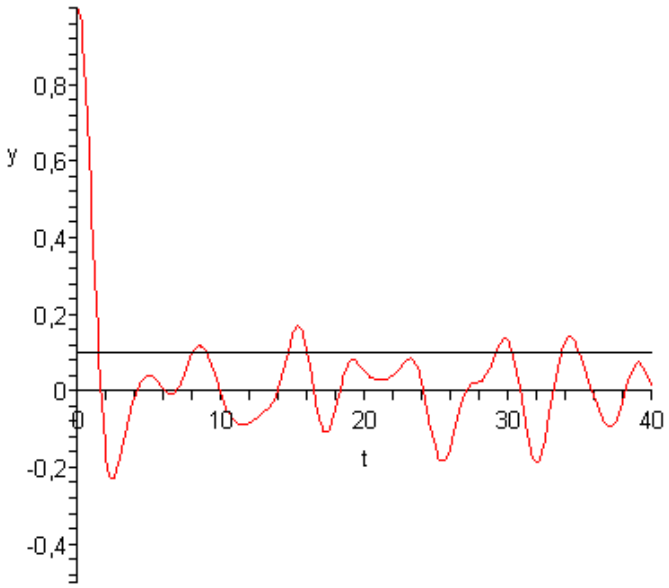


(e) ch.function of density 2

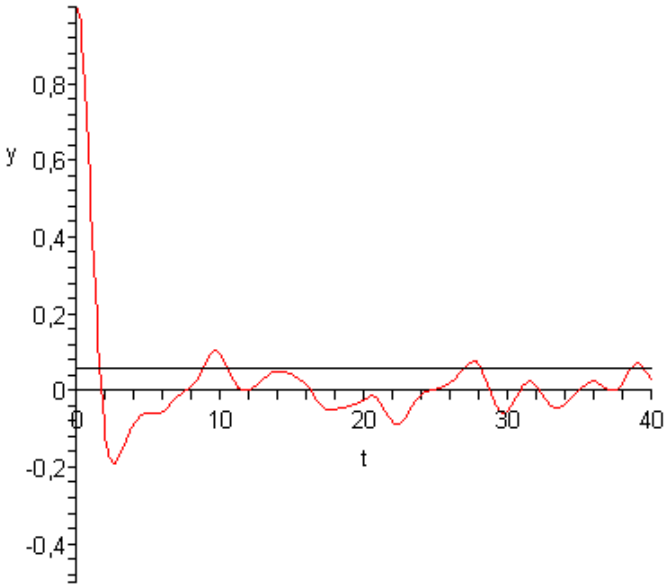


(f) emp.ch.function of density 2, n=50

Figure 6.1: ch.function and emp.ch.function for density nr.2

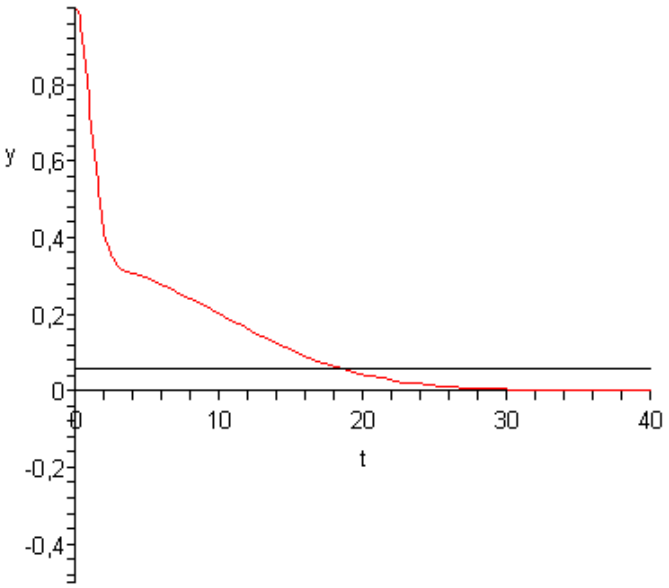


(g) emp.ch.function of density 2, n=100

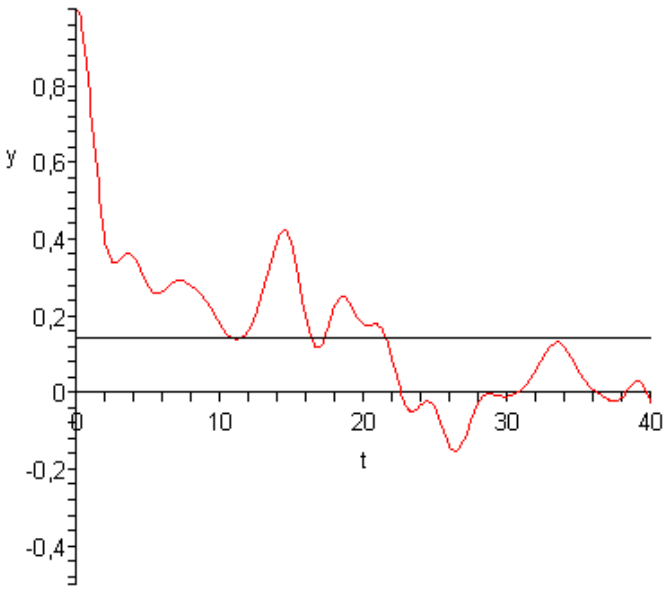


(h) emp.ch.function of density 2, n=300

Figure 6.1: ch.function and emp.ch.function for density nr.2

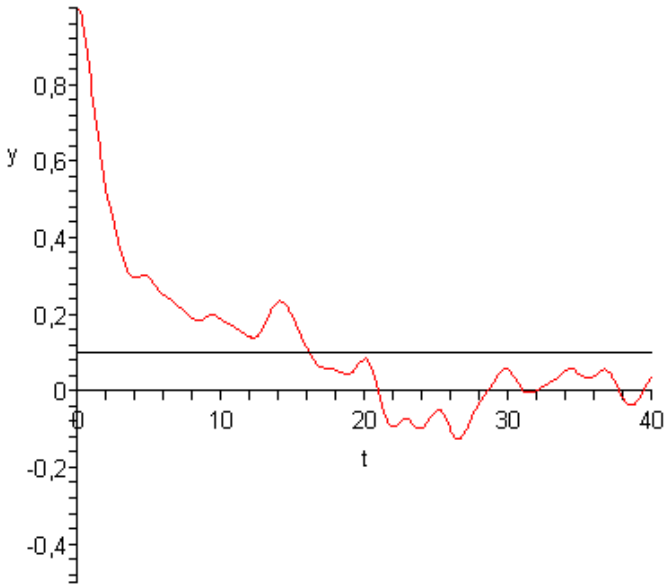


(i) ch.function of density 4

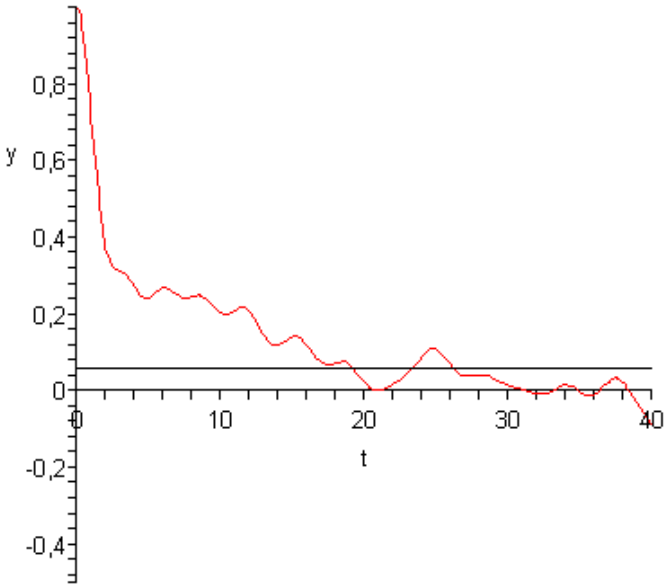


(j) emp.ch.function of density 4, n=50

Figure 6.1: ch.function and emp.ch.function for density nr.4

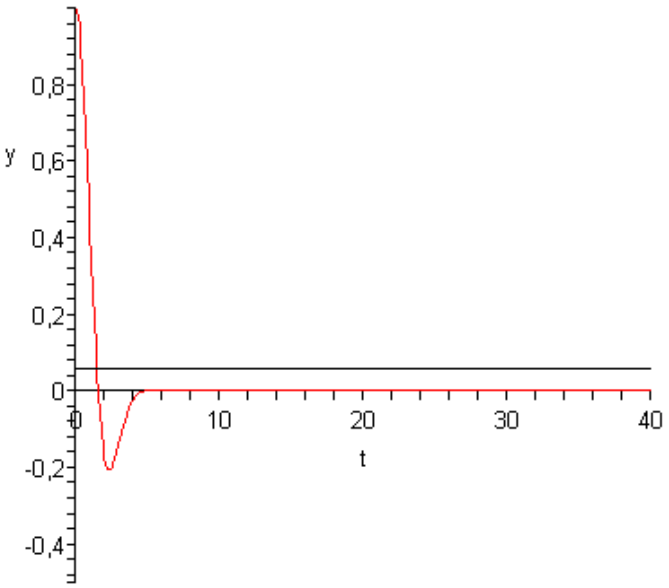


(k) emp.ch.function of density 4, n=100

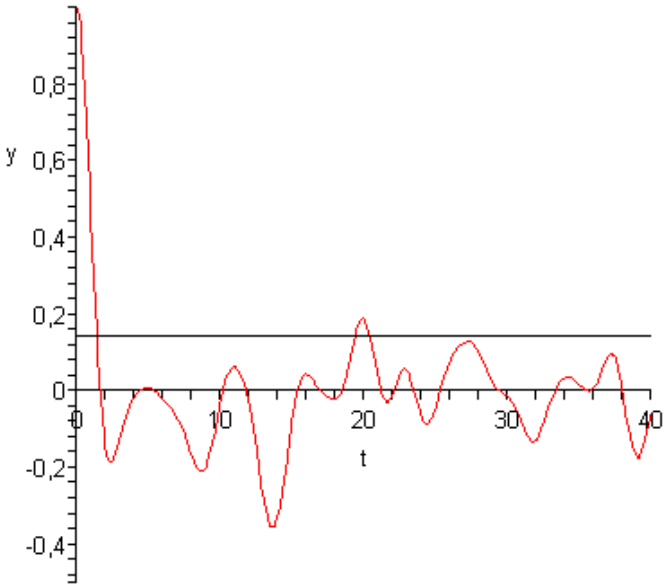


(l) emp.ch.function of density 4, n=300

Figure 6.1: ch.function and emp.ch.function for density nr.4

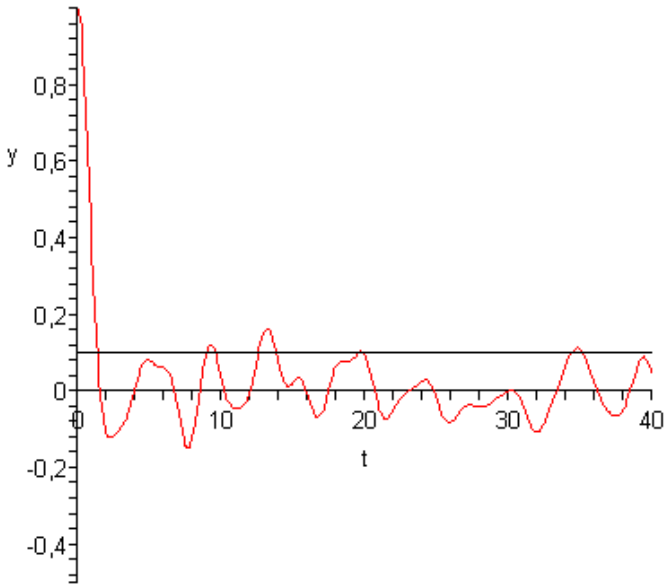


(m) ch.function of density 6

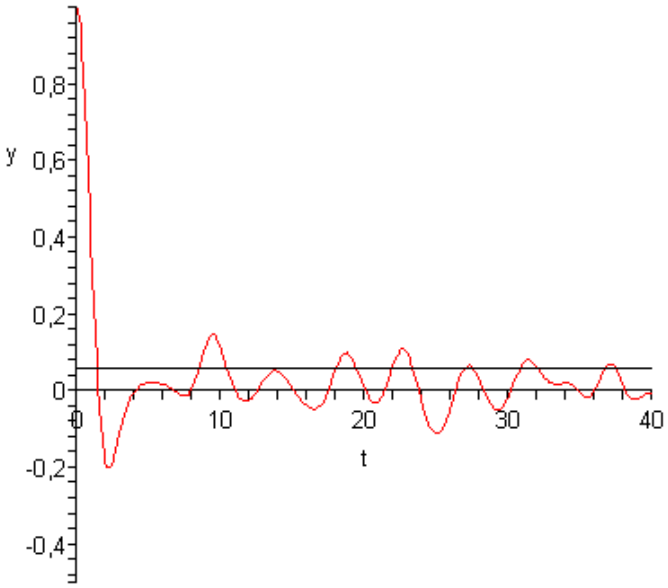


(n) emp.ch.function of density 6, n=50

Figure 6.1: ch.function and emp.ch.function for density nr.6

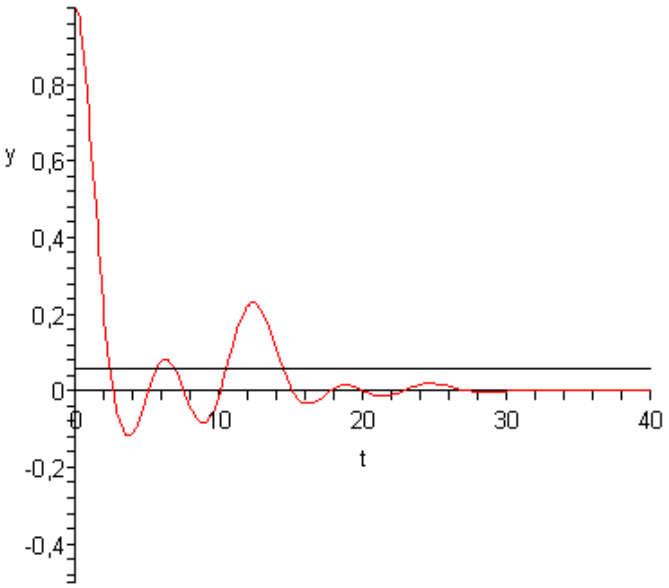


(o) emp.ch.function of density 6, n=100

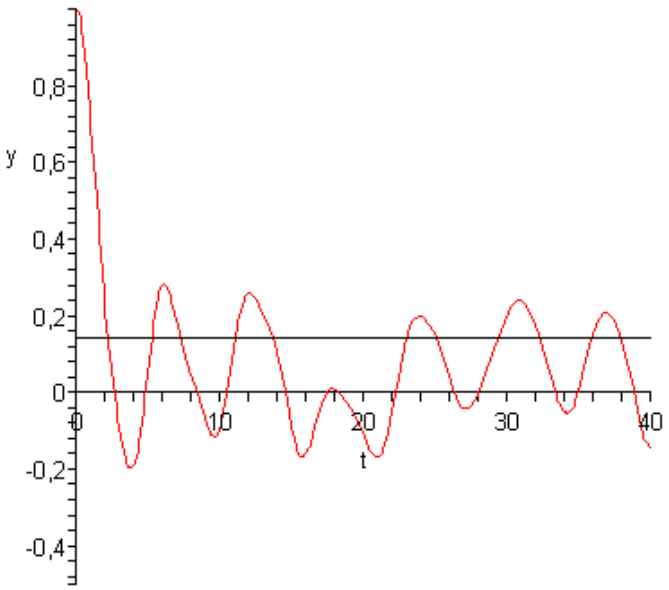


(p) emp.ch.function of density 6, n=300

Figure 6.1: ch.function and emp.ch.function for density nr.6

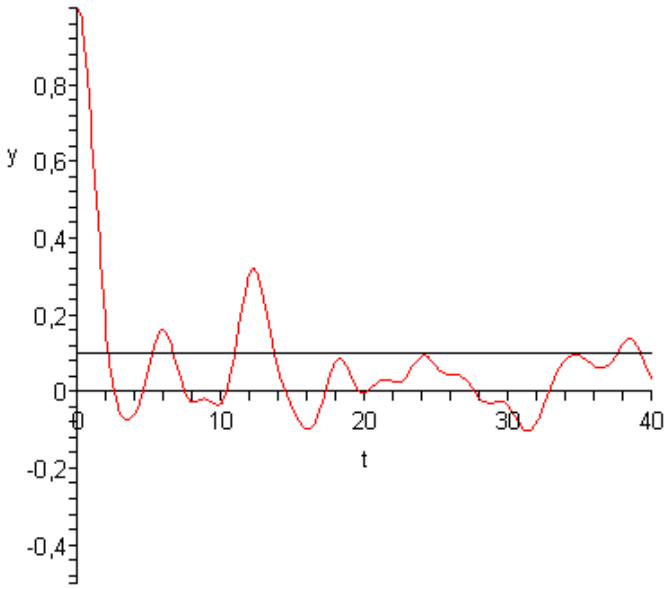


(q) ch.function of density 10

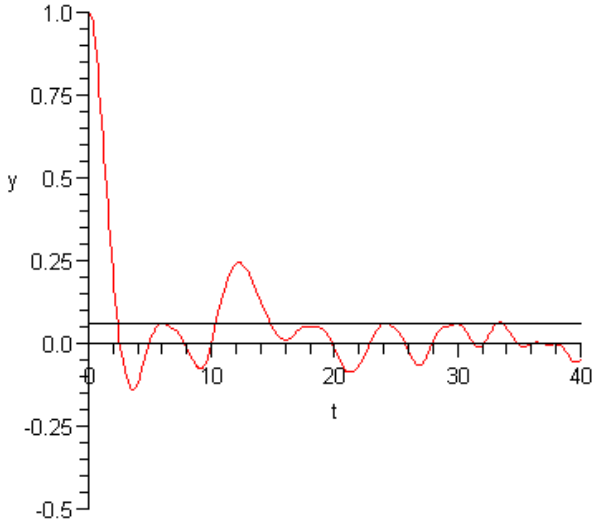


(r) emp.ch.function of density 10, n=50

Figure 6.1: ch.function and emp.ch.function for density nr.10

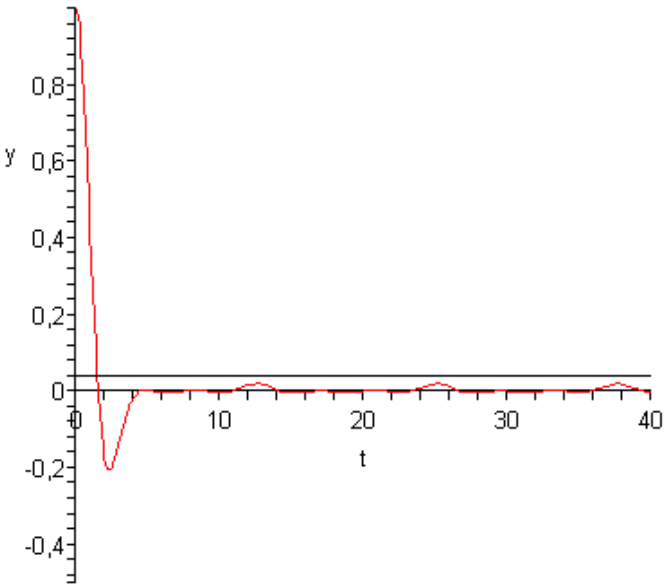


(s) emp.ch.function of density 10, n=100

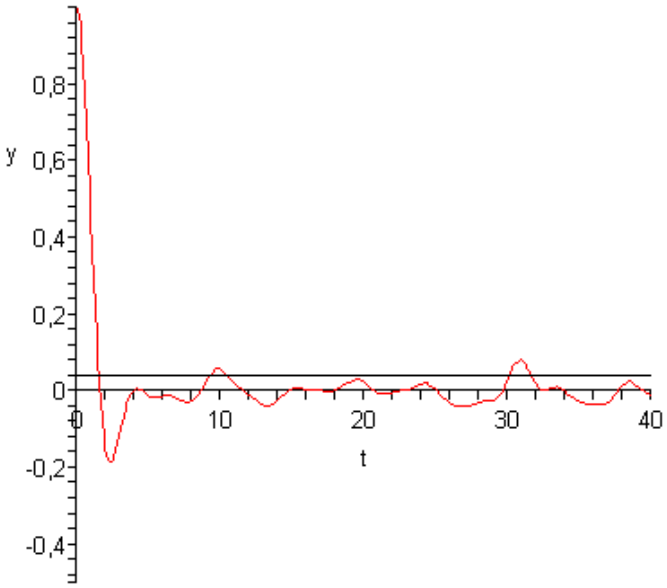


(t) emp.ch.function of density 10, n=300

Figure 6.1: ch.function and emp.ch.function for density nr.10



(u) ch.function of density 11



(v) emp.ch.function of density 11, n=700

Figure 6.1: ch.function and emp.ch.function for density nr.11

To calculate the bandwidth from these graphs we use $\phi(1/h)$ instead of $\phi(t)$. This means that the intersection between the two functions in the figures is the value for $1/h$. To find the value for h we had to calculate $h = 1/value$.

As we can see from the tables and figures, the result we get using the empirical characteristic function is a very good approximation for the bandwidth even for small sample sizes. Compared with the result Marron and Wand [5] got from using the gaussian kernel, our results are much better. The difference between the value we get from calculating the optimal bandwidth, and the value we get from the method based on the empirical characteristic function have a very small difference. We can see that when n is small, the difference is a little bigger, but compared to Marron and Wand, it is still very small.
This makes the sinc kernel a very good choice of the kernel, when it comes to estimating the bandwidth.

6.2 Performance

Asymptotically, the sinc estimator beats any finite-order kernel estimator, in particular any conventional estimator, provided that the density to be estimated is smooth enough. But in the finite sample case, a conventional estimator can have smaller error, the situation is very similar to that for higher-order estimators studied by Jones and Signorini (1997). In this section we compare the sinc estimator with a conventional estimator for several different densities. We use a method, developed by Marron and Wand (1992), which is based on the exact MISE calculation for normal mixture densities. The following densities (normal mixtures) are used.

#1. Normal

$$N(0,1).$$

#2. Bimodal

$$\frac{1}{2}(N(-1.4,1)+N(1.4,1)).$$

#3. Bimodal

$$\frac{1}{2}(N(-1.8,1)+N(1.8,1)).$$

#4. Plateau

$$\frac{1}{2}(N(-1,1)+N(1,1)).$$

#5. Separated bimodal

$$\frac{1}{2}(N(-2.5,1)+N(2.5,1)).$$

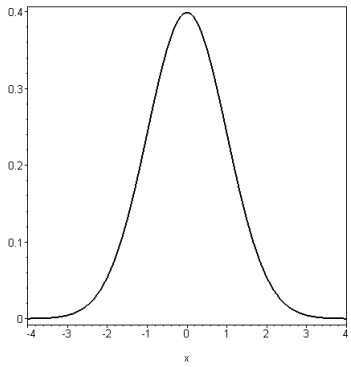
#6. Kurtotic

$$\frac{1}{2}(N(0,1/16)+N(0,3)).$$

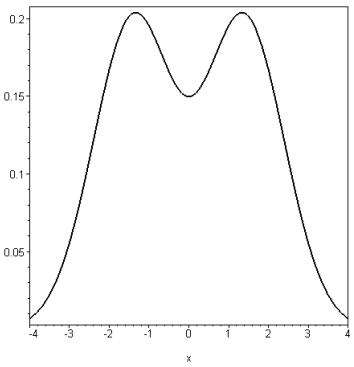
#7. Trimodal

$$0.3N(-2.7,1/2)+0.4N(0,1/2)+0.3N(2.7,1/2).$$

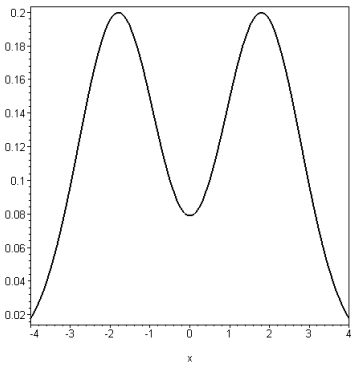
These seven densities are presented in Figure 1.



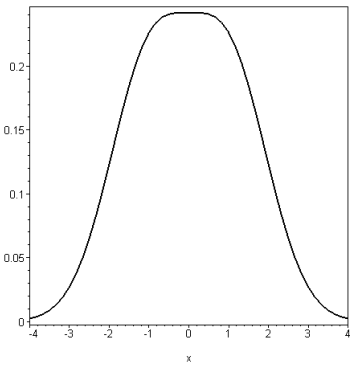
(a) Normal Density



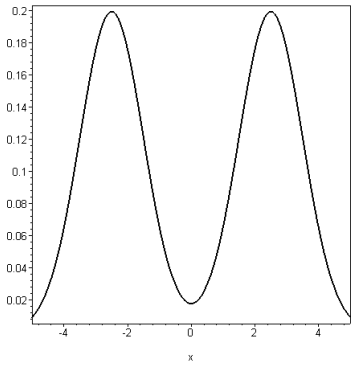
(b) Bimodal Density



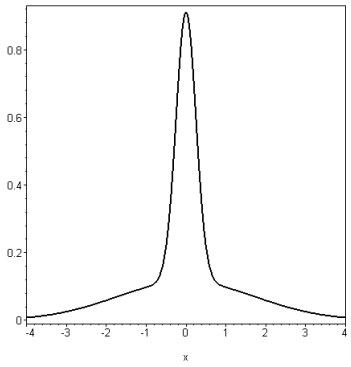
(c) Bimodal Density



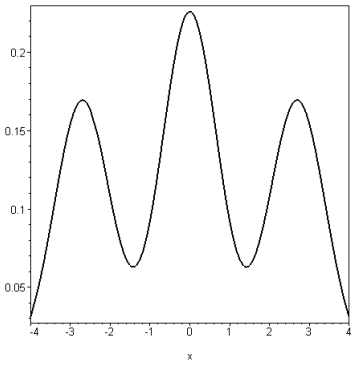
(d) Plateau Density



(e) Separated Bilmodal Density



(f) Kurtotic Density



(g) Trimodal Density

Figure 6.2: Densities

The comparison is made between the sinc estimator and conventional estimator with normal kernel

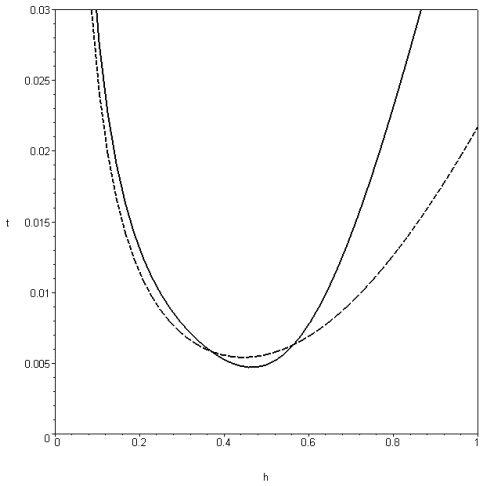
$$K(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}.$$

Results of the comparison are presented in Table 6.5 and in Figures 6.3 and 6.3. Table 6.5 contains the minimized MISE of the sinc estimator and normal based estimator ("norm") for two sample sizes, $n = 100$ and $n = 1000$. For $n = 100$, the Fourier integral estimator is better than the normal based estimator for densities #1, #3, #4, #5, #7 and is worse for densities #2, #6. For $n = 1000$, the sinc estimator is better for all seven densities, usually essentially (more than 50%).

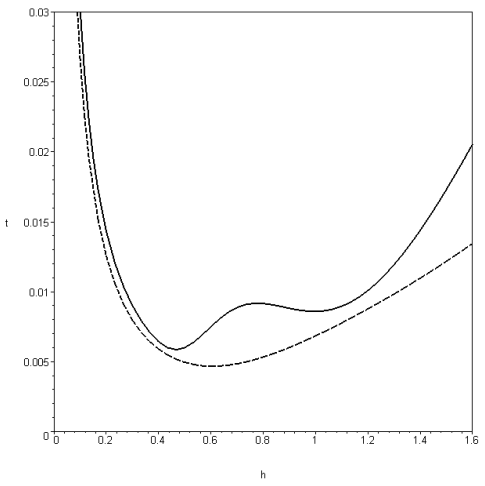
Table 6.5: Minimal MISE of the sinc estimator and the normal based estimator

	$n = 100$		$n = 1000$	
	sinc	norm	sinc	norm
#1	0.00470	0.00541	0.00061	0.00103
#2	0.00586	0.00467	0.00070	0.00089
#3	0.00551	0.00553	0.00063	0.00100
#4	0.00323	0.00374	0.00069	0.00073
#5	0.00481	0.00570	0.00073	0.00101
#6	0.02166	0.01941	0.00283	0.00344
#7	0.00771	0.00777	0.00085	0.00137

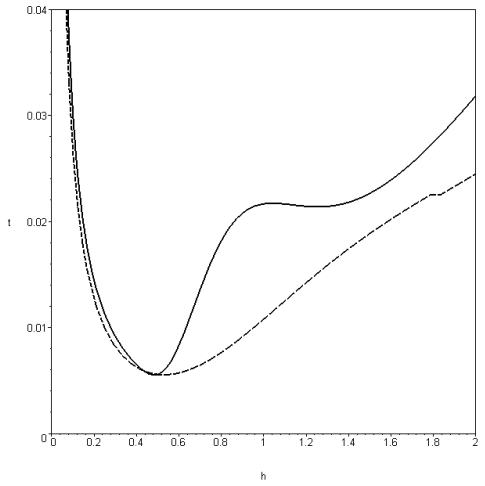
More detailed information is presented in Figures 2 and 3. Here, the MISE of the estimators are presented as a function of h, for $n = 100$ in Figure 6.3 and for $n = 1000$ in Figure 6.4. The dashed line is MISE of the normal based estimator and the solid line is MISE of the sinc estimator.



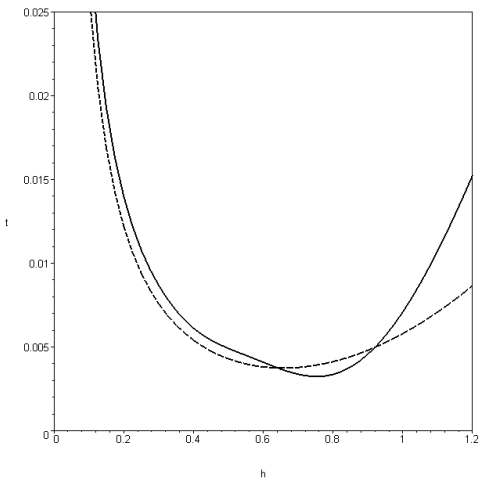
(a) Normal Density



(b) Bimodal Density

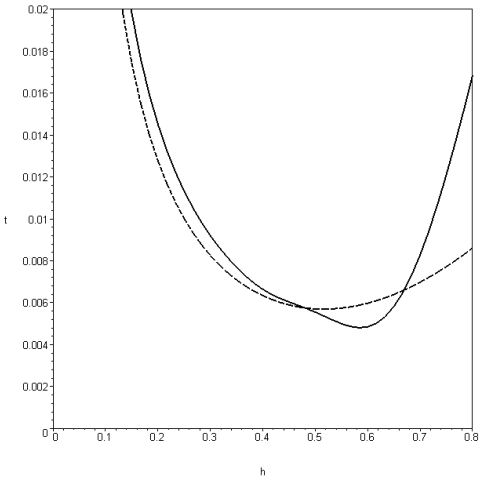


(c) Bimodal Density

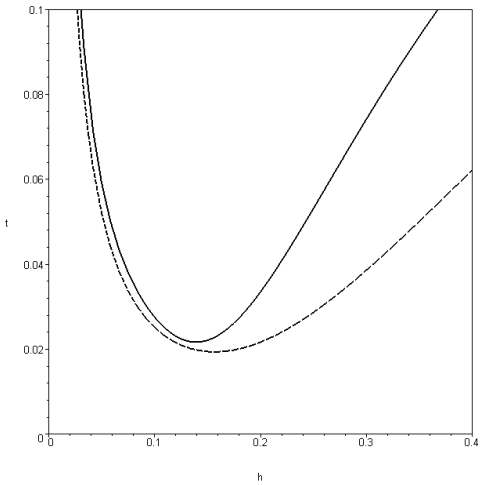


(d) Plateau Density

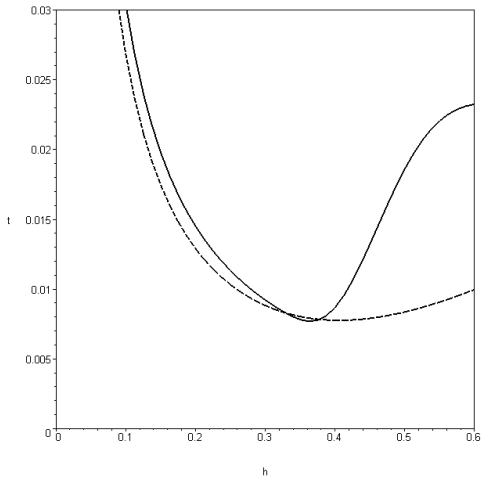
Figure 6.3: MISE, n=100



(e) Separated Bilmodal Density

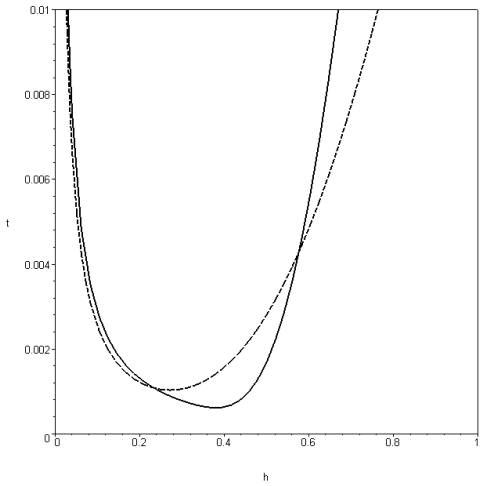


(f) Kurtotic Density

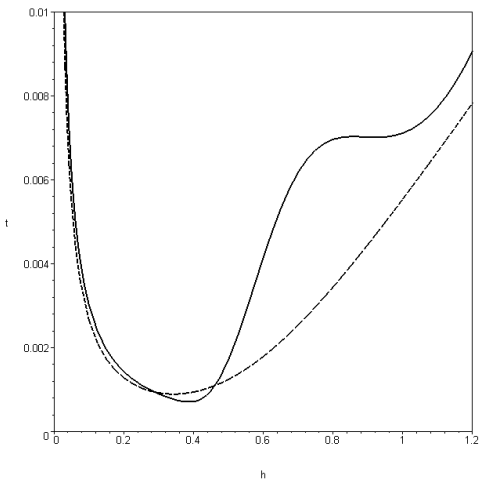


(g) Trimodal Density

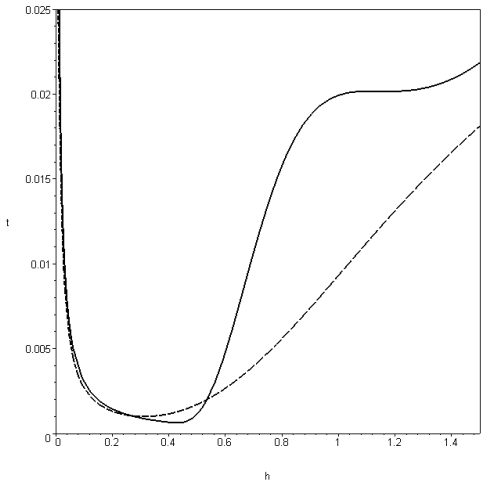
Figure 6.3: MISE, n=100



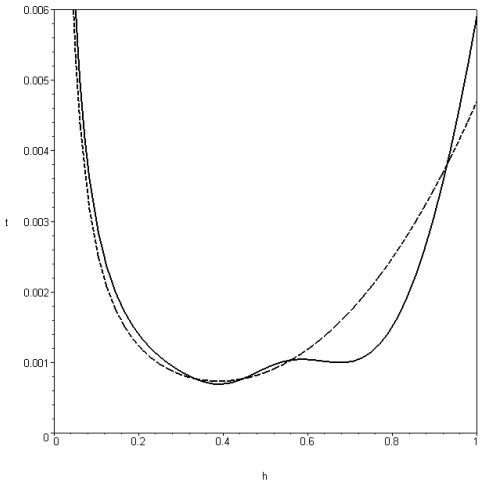
(a) Normal Density



(b) Bimodal Density

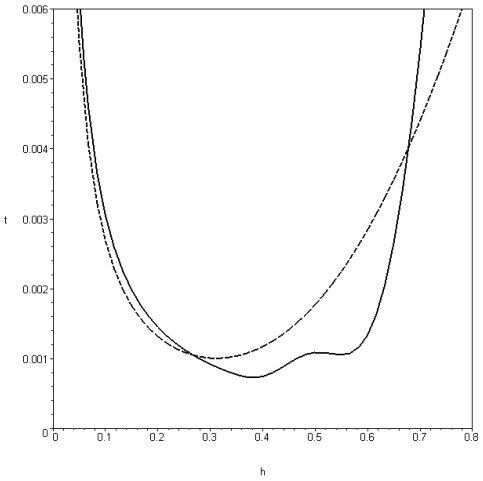


(c) Bimodal Density

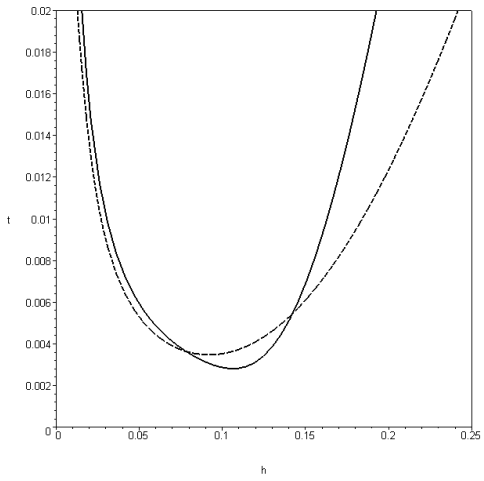


(d) Plateau Density

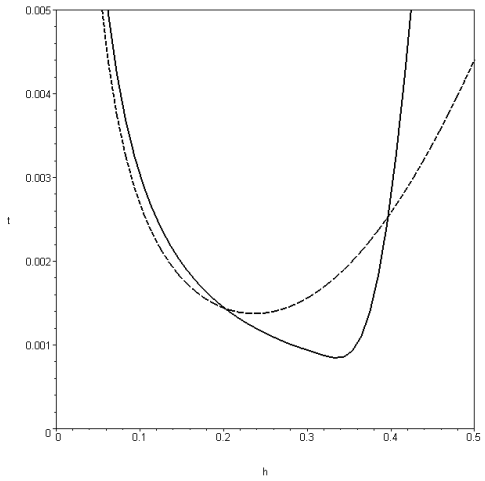
Figure 6.4: MISE, n=1000



(e) Separated Bilmodal Density



(f) Kurtotic Density



(g) Trimodal Density

Figure 6.4: MISE, n=1000

Bibliography

[1] George Casella and Roger L. Berger. *Statistical Inference*. Duxbury Advanced Series, 2002.

[2] Alan F.Karr. *Probability*. Springer Verlag, 1993.

[3] Ingrid K. Glad, Nils Lid Hjort, and Nikolai Ushakov. Density estimation using the sinc kernel. Preprint, Statistics, NTNU, 2007.

[4] M.P.Wand and M.C.Jones. *Kernel Smoothing*. Chapman and Hall, 1995.

[5] J.S.Marron; M.P.Wand. Exact Mean Integrated Squared Error. *The Annals of Statistics*, 20(712-736), 1992.

[6] Nikolai G. Ushakov. *Selected Topics in Characteristic Functions*. VSP, Utrecht, 1999.

Appendix A

Appendix

Programming-code used in MAPLE to achieve the results in this paper.

```
#The kernels I have used

B5: 0.5*exp(-abs(x));
B13: 1/(Pi*(cosh(x)));
B19: N;
B27: 1/(2*sqrt(3));
B31: (x^2/(sqrt(2*Pi)))*exp(-(x^2/2));
B33: 2/(Pi*(1+x^2)^2);
B35: 0.25*(1+abs(x))*exp(-abs(x));
B38: Pi/(4*(cosh((Pi*x)/2))^2);
B39: x/(2*sinh((Pi*x)/2));

#Setting the density and kernel functions
N := (1/(sqrt(2*Pi)))*exp(-(x^2)/2);  #normal density
K := 1/(Pi*(cosh(x)));  #kernel

#Calculating the characteristic functions
nt := int((cos(t*x))*N,x=-infinity..infinity);
kht := int((cos(h*t*x))*K,x=-infinity..infinity);

#Plotting the characteristic functions
plot(nt,t=-3..3);
plot(kht,t=-1..1);
```

```
#Setting parameters
n:=1;
mom:=2.46;
phi := nt;
psi := kht;

#Calculating R(K), R(F) and R(f'')
RK := int(K^2,x=-infinity..infinity);
RF := int((diff(N,x))^2,x=-infinity..infinity);
RFF := int((diff(diff(N,x),x))^2,x=-infinity..infinity);

#Evaluating values for different expressions
hhh := evalf((2/(mom^2*RFF*Pi))^0.25);
hh := evalf((RK/((mom^2)*RFF))^0.2);

#Calculating IV, ISB and MISE
IV := ((1/(2*n*Pi))*int((1-abs(phi)^2)*(abs(psi))^2,
t=-infinity..infinity));
ISB := ((1/(2*Pi))*int((abs(1-psi))^2*(abs(phi))^2,
t=-infinity..infinity));
MISE := IV + ISB;

#Plots of IV, ISB and MISE
plot(evalf(ISB),h=0..1);
plot(evalf(IV),h=0..1);
plot(evalf(MISE),h=0..1);

#Calculating asymptotic expressions
AIV := (1/(n*h))*int((K)^2,x=-infinity..infinity);
AISB := (1/4)*h^4*mom^2*int((diff(diff(N,x),x))^2,x=-infinity..infinity)
AMISE := AIV + AISB;

#Calculating h_AMISE
hamise := solve(diff(AMISE,h));

#Plots MISE and AMISE in the same picture
```

```
plot([evalf(MISE),AMISE],h=0..1, y=0..0.4,color=[black,black]);

#Finding values for h_MISE and h_AMISE
for h from 0.2 by 0.01 to 0.32 do
tot[h] := simplify(AMISE)  end do;
for h from 0.24 by 0.01 to 0.36 do
tot[h] := simplify(evalf(MISE))  end do;

#Calculating derivatives
AISBderiv := h^3*mom^2*int((diff(diff(N,x),x))^2,
x=-infinity..infinity);
AIVderiv := -(1/(n*h^2))*int((K)^2,x=-infinity..infinity);
gderiv := (-(2/Pi)*int((abs(phi))^2*diff(psi,h),t=0..infinity))
+((2/Pi)*(1-(1/n))*int((abs(phi))^2*psi*diff(psi,h),t=0..infinity))
aaa := gderiv - AISBderiv;

#Plotting derivatives
plot([AISBderiv,gderiv],h=0..0.2,y=0..0.01,color=[black,black]);
plot(aaa,h=0..0.3);

#Calculating derivatives for MISE and AMISE, and plotting them
AMISEderiv := AIVderiv + AISBderiv;
MISEderiv := AIVderiv + gderiv;
plot([AMISEderiv,MISEderiv],h=0.2..0.4,y=-0.1..0.1,
color=[blue,black]);

#Finding the function g(h), and its derivative
g := ((1/(2*Pi))*int((abs(1-psi))^2*(abs(phi))^2,t=0..infinity))
-((1/(2*n*Pi))*int((abs(phi)^2)*(abs(psi))^2,t=0..infinity));
gder := diff(g,h);

#Finding an upper bound and compare with the derivative of AISB
ub := ((1/n)*2*h*mom*int((diff(N,x))^2,x=-infinity..infinity))
+((1-(1/n))*AISBderiv);
plot([ub,AISBderiv],h=0..1,color=[black,blue]);
a := ub - AISBderiv;
solve(a=0,h);
```

```
#Calculating the normal mixture densities fromMarron and Wand

n01 := (1/sqrt(2*Pi))*exp((-1/2)*x^2);
n1b22b3 := (1/sqrt(2*Pi))*(1/(2/3))*
exp(((-1)*(x-0.5)^2)/(2*(2/3)^2));
n13b125b9 := (1/sqrt(2*Pi))*(1/(5/9))*
exp(((-1)*(x-(13/12))^2)/(2*(5/9)^2));
n01b10 := (1/sqrt(2*Pi))*(1/(1/10))*
exp(((-1)*(x-0)^2)/(2*(1/10)^2));
nm12b3 := (1/sqrt(2*Pi))*(1/(2/3))*
exp(((-1)*(x-(-1))^2)/(2*(2/3)^2));
n12b3 := (1/sqrt(2*Pi))*(1/(2/3))*
exp(((-1)*(x-(1))^2)/(2*(2/3)^2));
nm11b10 := (1/sqrt(2*Pi))*(1/(1/10))*
exp(((-1)*(x-(-1))^2)/(2*(1/10)^2));
nm1b21b10 := (1/sqrt(2*Pi))*(1/(1/10))*
exp(((-1)*(x-(-1*(1/2)))^2)/(2*(1/10)^2));
n1b21b10 := (1/sqrt(2*Pi))*(1/(1/10))*
exp(((-1)*(x-(1/2))^2)/(2*(1/10)^2));
n11b10 := (1/sqrt(2*Pi))*(1/(1/10))*
exp(((-1)*(x-1)^2)/(2*(1/10)^2));
nm3b21b100 :=(1/sqrt(2*Pi))*(1/(1/100))*
exp(((-1)*(x-(-1*(3/2)))^2)/(2*(1/100)^2));
nm11b100 :=(1/sqrt(2*Pi))*(1/(1/100))*
exp(((-1)*(x-(-1*(2/2)))^2)/(2*(1/100)^2));
nm1b21b100 :=(1/sqrt(2*Pi))*(1/(1/100))*
exp(((-1)*(x-(-1*(1/2)))^2)/(2*(1/100)^2));
n01b100 :=(1/sqrt(2*Pi))*(1/(1/100))*
exp(((-1)*(x-(-1*(0/2)))^2)/(2*(1/100)^2));
n1b21b100 :=(1/sqrt(2*Pi))*(1/(1/100))*
exp(((-1)*(x-(1*(1/2)))^2)/(2*(1/100)^2));
n11b100 :=(1/sqrt(2*Pi))*(1/(1/100))*
exp(((-1)*(x-(1*(2/2)))^2)/(2*(1/100)^2));
n3b21b100 :=(1/sqrt(2*Pi))*(1/(1/100))*
exp(((-1)*(x-(1*(3/2)))^2)/(2*(1/100)^2));
```

```
n1 := n01;
n2 := (0.2*n01) + (0.2*n1b22b3) + (0.6*n13b125b9);
n4 := ((2/3)*n01) + ((1/3)*n01b10);
n6 := (0.5*nm12b3) + (0.5*n12b3);
n10:= (0.5*n01)+((1/10)*nm11b10)+((1/10)*nm1b21b10)+
((1/10)*n01b10)+((1/10)*n1b21b10)+((1/10)*n11b10);
n11:= ((49/100)*nm12b3)+((49/100)*n12b3)+((1/350)*nm3b21b100)+
((1/350)*nm11b100)+((1/350)*nm1b21b100)+
((1/350)*n01b100)+((1/350)*n1b21b100)+
((1/350)*n11b100)+((1/350)*n3b21b100);

#Calculating the ch.function and the empirical ch.function
and makes plots for all the six densities

cf1 := int((cos(t*x))*n1,x=-infinity..infinity);
plot([cf1,(1/sqrt(301))],t=0..40,y=-0.5..1);
plot(cf1-(1/sqrt(301)),t=1.0..2.5);
with(stats);
sample1 := array(1..300,[stats[random, normald[0,1]](300)] );
emp1 := (1/300)*sum(cos(t*sample1[k]),k=1..300);
plot([emp1,(1/sqrt(301))],t=0..40,y=-0.5..1);
plot(emp1-(1/sqrt(301)),t=1.0..2.5);

cf2 := int((cos(t*x))*n2,x=-infinity..infinity);
plot([cf2,(1/sqrt(301))],t=0..40,y=-0.5..1);
plot(cf2-(1/sqrt(301)),t=1.0..2.5);
with(stats);
sample2 := array(1..300,[stats[random, normald[0,1]](60),
stats[random, normald[0.5,(2/3)]](60),
stats[random, normald[(13/12),(5/9)]](180)] );
emp2 := (1/300)*sum(cos(t*sample2[k]),k=1..300);
plot([emp2,(1/sqrt(301))],t=0..40,y=-0.5..1);
plot(emp2-(1/sqrt(301)),t=1.0..2.5);

cf4 := int((cos(t*x))*n4,x=-infinity..infinity);
plot([cf4,(1/sqrt(301))],t=0..40,y=-0.5..1);
```

```
plot(cf4-(1/sqrt(301)),t=15..20);
with(stats);
sample4 := array(1..300,[stats[random, normald[0,1]](200),
stats[random, normald[0,(1/10)]](100)]);
emp4 := (1/300)*sum(cos(t*sample4[k]),k=1..300);
plot([emp4,(1/sqrt(301))],t=0..40,y=-0.5..1);
plot(emp4-(1/sqrt(301)),t=15..20);

cf6 := int((cos(t*x))*n6,x=-infinity..infinity);
plot([cf6,(1/sqrt(301))],t=0..40,y=-0.5..1);
plot(cf6-(1/sqrt(301)),t=1.0..2.5);
with(stats);
sample6 := array(1..300,[stats[random, normald[-1,(2/3)]](150),
stats[random, normald[1,(2/3)]](150)]);
emp6 := (1/300)*sum(cos(t*sample6[k]),k=1..300);
plot([emp6,(1/sqrt(301))],t=0..40,y=-0.5..1);
plot(emp6-(1/sqrt(301)),t=1.0..2.5);

cf10 := int((cos(t*x))*n10,x=-infinity..infinity);
plot([cf10,(1/sqrt(301))],t=0..40,y=-0.5..1);
plot(cf10-(1/sqrt(301)),t=1.5..2.5);
with(stats);
sample10 := array(1..300,[stats[random, normald[0,1]](150),
stats[random, normald[-1,(1/10)]](30) ,
stats[random, normald[-0.5,(1/10)]](30) ,
stats[random, normald[0,(1/10)]](30) ,
stats[random, normald[0.5,(1/10)]](30) ,
stats[random, normald[1,(1/10)]](30)]);
emp10 := (1/300)*sum(cos(t*sample10[k]),k=1..300);
plot([emp10,(1/sqrt(301))],t=0..40,y=-0.5..1);
plot(emp10-(1/sqrt(301)),t=1.5..2.5);

cf11 := int((cos(t*x))*n11,x=-infinity..infinity);
plot([cf11,(1/sqrt(701))],t=0..40,y=-0.5..1);
plot(cf11-(1/sqrt(701)),t=1.5..2.5);
with(stats);
sample11 := array(1..700,[stats[random, normald[-1,(2/3)]](343) ,
```

```
stats[random, normald[1,(2/3)]](343) ,
stats[random, normald[-1.5,(1/100)]](2) ,
stats[random, normald[-1,(1/100)]](2) ,
stats[random, normald[-0.5,(1/100)]](2) ,
stats[random, normald[0,(1/100)]](2) ,
stats[random, normald[0.5,(1/100)]](2) ,
stats[random, normald[1,(1/100)]](2) ,
stats[random, normald[1.5,(1/100)]](2)]);
emp11 := (1/700)*sum(cos(t*sample11[k]),k=1..700);
plot([emp11,(1/sqrt(701))],t=0..40,y=-0.5..1);
plot(emp11-(1/sqrt(701)),t=1.5..2.5);

#Setting the density and calculates its ch.function
F := (1/(sqrt(2*Pi)))*exp(-(x^2)/2);
f := int((cos(h*t*x))*F,x=-infinity..infinity);

#Setting new parameters
n:=100;
phi := cf1;
psi := f;

#Calculating MISE, using the gaussian kernel and the sinc-kernel
and compares these
IV := ((1/(2*n*Pi))*int((1-abs(phi)^2)*(abs(psi))^2,t=-100..100));
ISB := ((1/(2*Pi))*int((abs(1-psi))^2*(abs(phi))^2,t=-100..100));
MISE := IV + ISB;
sincMISE := (1/(Pi*n*h))+int(n01^2,x=-infinity..infinity)-
((1+(1/n))*(1/Pi)*int(phi^2,t=0..(1/h)));
plot([MISE,sincMISE],h=0..1,t=0..0.1,color=[red,blue]);

#Comparing MISE, using the gaussian kernel and the sinc-kernel
for a special density DE
a := 1;
F := (1/(sqrt(2*Pi)))*exp(-(x^2)/2);
f := int((cos(h*t*x))*F,x=-infinity..infinity);
DE := 0.5*((1/(sqrt(2*Pi)))*exp(-((x-a)^2)/2) +
(1/(sqrt(2*Pi)))*exp(-((x+a)^2)/2));
```

```
d := int((cos(h*t*x))*DE,x=-infinity..infinity);
dd := (cos(a*t))*exp(-(t^2/2));
n:=100;
phi := dd;
psi := f;
IV := ((1/(2*n*Pi))*int((1-abs(phi)^2)*(abs(psi))^2,t=-100..100));
ISB := ((1/(2*Pi))*int((abs(1-psi))^2*(abs(phi))^2,t=-100..100));
MISE := IV + ISB;
sincMISE := (1/(Pi*n*h))+int(DE^2,x=-infinity..infinity)-
((1+(1/n))*(1/Pi)*int(phi^2,t=0..(1/h)));
plot([MISE,sincMISE],h=0..1,t=0..0.1,color=[red,blue]);
```