NORGES TEKNISK-NATURVITENSKAPELIGE UNIVERSITET

Bandwidth Selection for the Fourier Integral Estimator

by

Thomas Oksavik and Nikolai Ushakov

PREPRINT
STATISTICS NO. 8/2007

NORWEGIAN UNIVERSITY OF SCIENCE AND TECHNOLOGY
TRONDHEIM, NORWAY

This report has URL http://www.math.ntnu.no/preprint/statistics/2007/S8-2007.ps
Nikolai Ushakov has homepage: http://www.math.ntnu.no/∼ushakov
E-mail: ushakov@stat.ntnu.no
Address: Department of Mathematical Sciences, Norwegian University of Science and Technology, N-7491 Trondheim, Norway.

Bandwidth Selection for the Fourier Integral Estimator

Thomas Oksavik and Nikolai Ushakov
Department of Mathematical Sciences
Norwegian University of Science and Technology
Trondheim, Norway

Abstract

In this work, we consider the problem of bandwidth selection for the Fourier integral estimator. Using a simulation study, we investigate the performance of two bandwidth selectors, find out their merits and defects.

Key words: Bandwidth selection, Kernel estimation, Sinc kernel, Fourier integral kernel, Mean integrated squared error, Empirical characteristic function, Finite samples

1. Introduction

One of main problems in nonparametric density estimation is the smoothing parameter selection. In this work, we consider this problem for the kernel density estimator based on so-called sinc or Fourier integral kernel. Let $X_1, ..., X_n$ be a random sample (iid random variables) from an absolutely continuous distribution with the density function $f(x)$. Consider the problem of estimating $f(x)$. Suppose that a functional form of $f(x)$ is unknown, so it has to be estimated nonparametrically. The kernel estimator, based on the sample $X_1, ..., X_n$, is defined as

$$f_n(x; h) = \frac{1}{nh} \sum_{j=1}^{n} K\left(\frac{x - X_j}{h}\right),$$

where $K(x)$ is the kernel and h is the smoothing parameter — bandwidth. In this paper, we use the mean integrated squared error (MISE) as the performance criterion:

$$\text{MISE}(f_n) = \text{E} \int (f_n(x; h) - f(x))^2 dx.$$

The kernel usually belongs to one of the following families: conventional kernels, higher-order kernels, superkernels, sinc kernel. We say that a kernel is conventional if it is a probability density function i.e. is nonnegative and integrates to one. Higher order kernels are defined as follows. Let k be an integer and $k \geq 2$. $K(x)$ is called a k-order kernel if $K(x)$ is symmetric,

$$\int K(x)dx = 1, \quad \int x^j K(x)dx = 0 \text{ for } j = 1, ..., k-1,$$

and

$$\int x^k K(x)dx \neq 0.$$

Conventional kernels are second-order kernels. If $k > 2$, then we say that the kernel is higher-order. Higher-order kernel estimators have better asymptotic MISE than conventional estimators, provided that the density to be estimated is smooth enough.

Let $K(x)$ be a kernel and $\psi(t)$ be its characteristic function

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

We say that $K(x)$ is a superkernel if $\psi(t)$ has form

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

where $g(t)$ is a real-valued, even function, satisfying the inequality $|g(t)| \leq 1$ and chosen in such a way that $\psi(t)$ is continuous, $\Delta > 0$, $c > 1$. Provided the density to be estimated is smooth enough, superkernel estimators have better asymptotic MISE than higher-order estimators.

The Fourier integral kernel (sinc kernel) is the function

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

Asymptotic behaviour of the MISE of this estimator is similar to that of superkernel estimators.

Thus, if the density to be estimated is smooth enough, then conventional estimators are asymptotically inferior to both higher-order, superkernel and Fourier integral estimators. Situation, however, is essentially different in the finite sample case, i.e. when the sample size is small or moderate. Extensive simulations display that there are neither unquestionable leaders nor unquestionable outsiders in this case, see for example Jones and Signorini (1997). While a conventional estimator is inferior to a higher-order estimator for some densities to be estimated, it is superior for others.

Therefore, making a choice of the kernel, one should take into account other properties of kernels, first of all how the problem of the optimal bandwidth approximation is solved for the given kernel. From this point of view, the Fourier integral estimator is very attractive because there is a very simple connection between the optimal bandwidth and the underlying distribution, see Glad et al. (2007).

The rest of the paper is as follows. In Section 2, we make a comparison of the MISE of the Fourier integral estimator with a conventional kernel estimator for a number of different densities. The comparison shows that the sinc estimator is quite competetive in the finite sample case. Two methods of bandwidth selection, suggested by Glad et al. (2007), are considered in Section 2. Performance of these methods is studied in Section 3.

2. Fourier integral and conventional estimators: finite sample comparison

The Fourier integral estimator was studied by Davis (1975), Davis (1977) and Glad et al. (2007). This estimator can produce estimates which take negative values and do not integrate to one. This defect however is corrected without loss of performance, see Glad et al. (2003).

Asymptotically, the Fourier integral 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 Fourier integral 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. Skewed unimodal

$$\frac{1}{5}N(0, 1) + \frac{1}{5}N(1/2, 4/9) + \frac{3}{5}N(13/12, 25/81).$$

#8. Trimodal

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

#9. Asymmetric unimodal

$$\frac{1}{3}(N(-0.6, 1/16) + N(0, 1) + N(2, 1)).$$

#10. Asymmetric bimodal

$$0.7N(-1.8, 1) + 0.3N(1.8, 1).$$

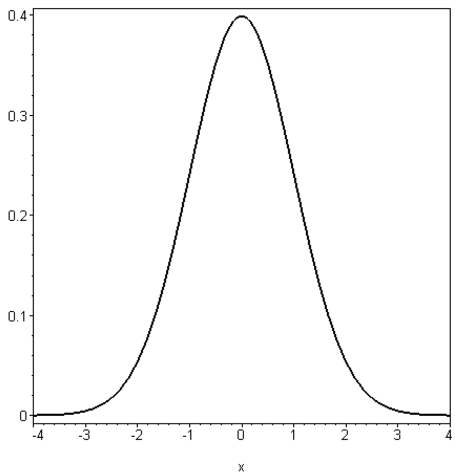
These ten densities are presented in Figure 1.

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

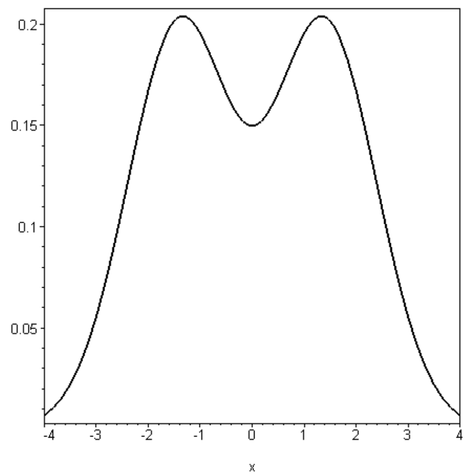
$$K(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}.$$

Results of the comparison are presented in Table 1 and in Figures 2 and 3. Table 1 contains the minimized MISE of the Fourier integral estimator ("sinc") 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, #8, #10 and is worse for densities #2, #6, #9. For $n = 1000$, the sinc estimator is better for all ten densities, usually essentially (more than 50%).

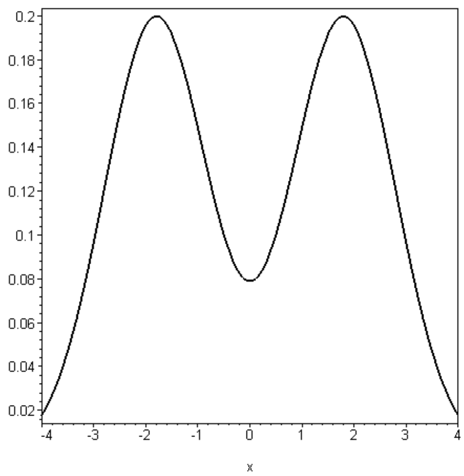
The reason that we include two very similar densities #2 and #3 is that we would like to show the behaviour of the ratio between the minimal MISE of the sinc and conventional estimators (for $n = 100$) for such bimodal symmetric densities when peaks go away from each other: first the sinc estimator is better, then, from some distance between the peaks, the conventional estimator becomes better (for example #2), but starting approximately from #3 the sinc becomes better again.



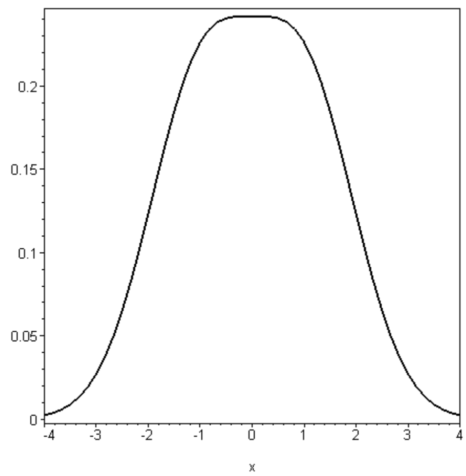
(a) Normal Density



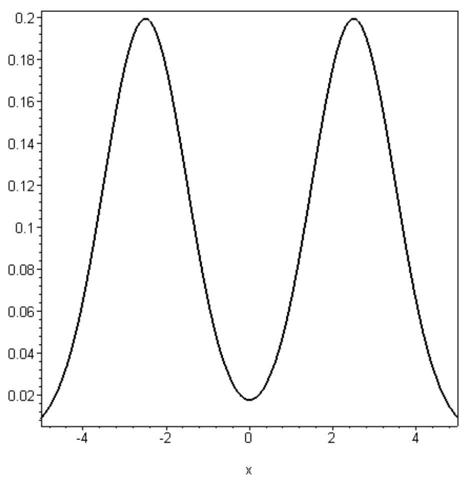
(b) Bimodal Density



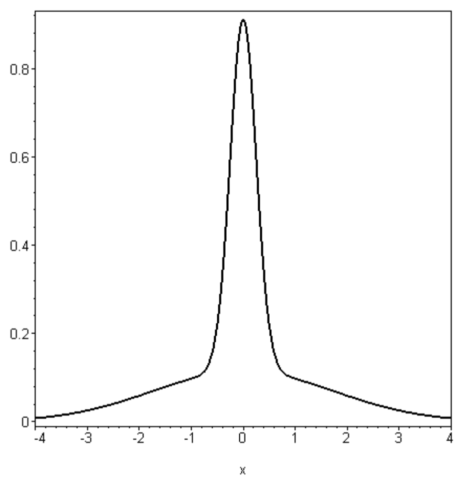
(c) Bimodal Density



(d) Plateau Density

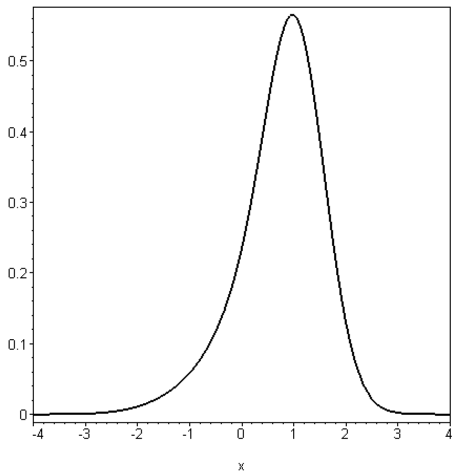


(e) Separated Bilmodal Density

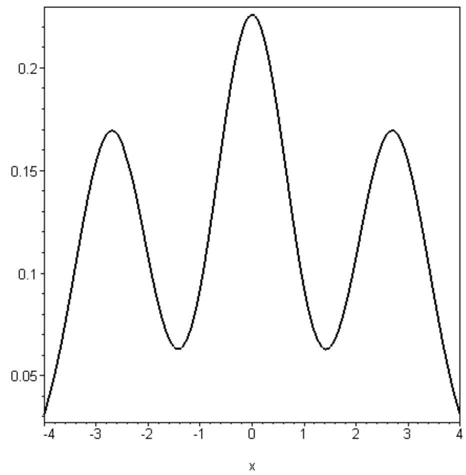


(f) Kurtotic Density

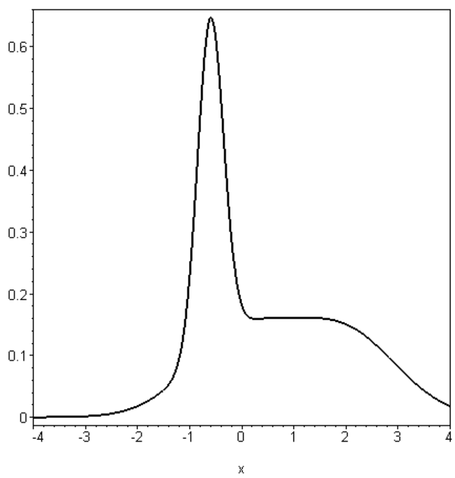
Figure 1: Densities



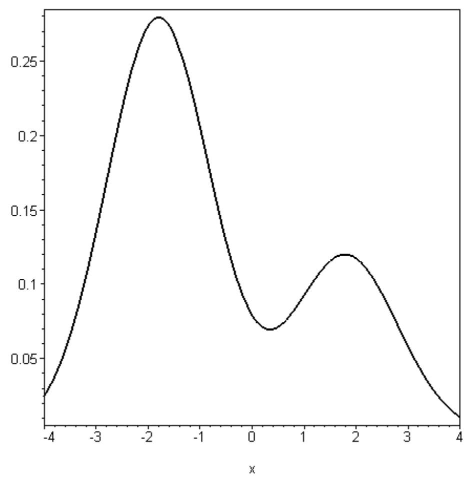
(g) Skewed Unimodal Density



(h) Trimodal Density



(i) Asymmetric Unimodal Density



(j) Assymetric Bimodal Density

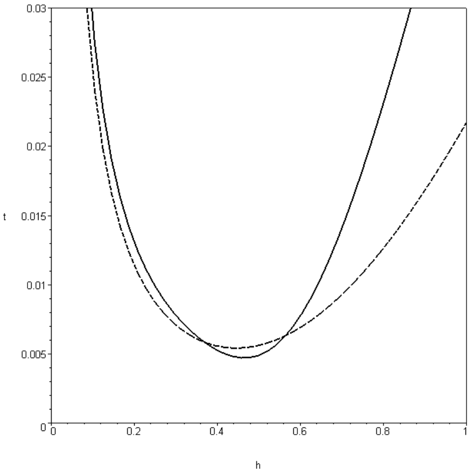
Figure 1: Densities

Table 1

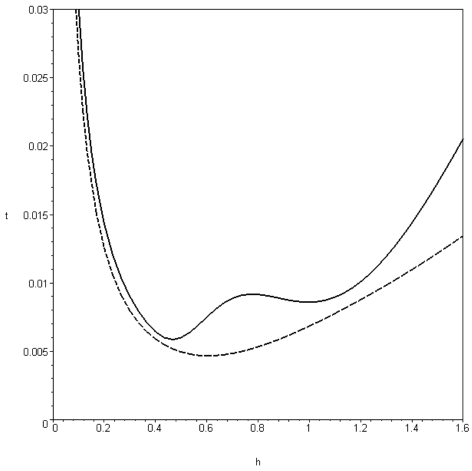
Minimal MISE of the sinc estimator and the normal based estimator

	$n = 100$		$n = 1000$	
	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	0.00827	0.00831	0.00111	0.00157
#8	0.00771	0.00777	0.00085	0.00137
#9	0.02046	0.01659	0.00277	0.00299
#10	0.00539	0.00557	0.00064	0.00101

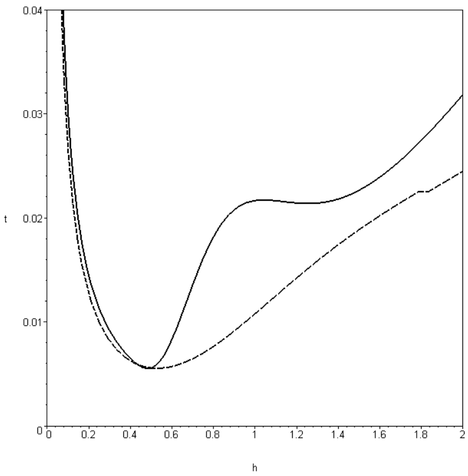
More detailed information is presented in Figures 2 and 3. Here, the MISE of the estimators are presented as a function of h, for $n = 100$ in Figure 2 and for $n = 1000$ in Figure 3. Solid lines correspond to the Fourier integral estimator and dotted lines — to the normal estimator.



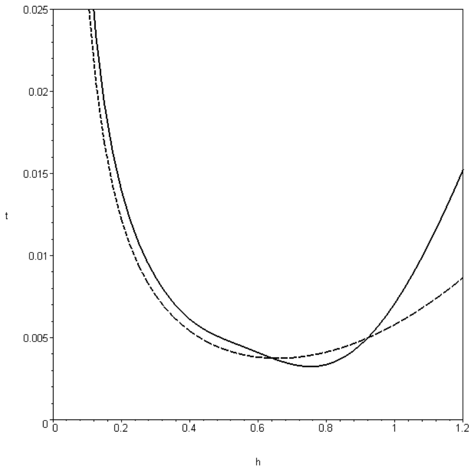
(a) Normal Density



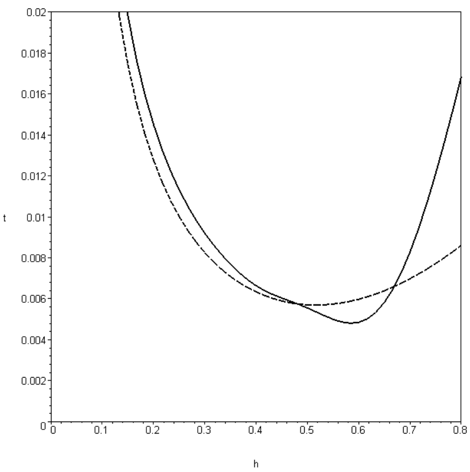
(b) Bimodal Density



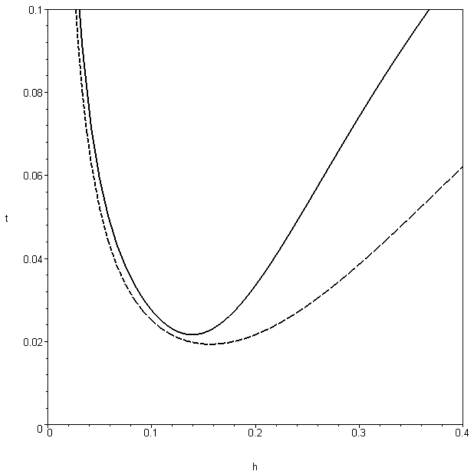
(c) Bimodal Density



(d) Plateau Density

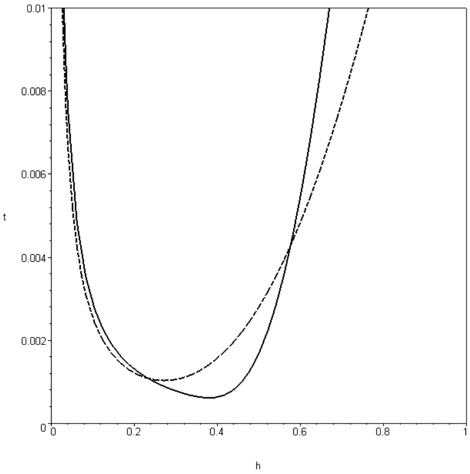


(e) Separated Bilmodal Density

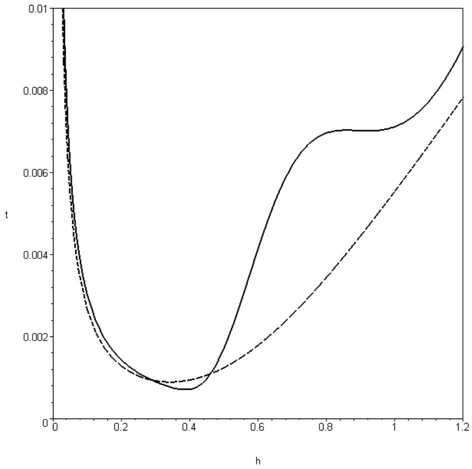


(f) Kurtotic Density

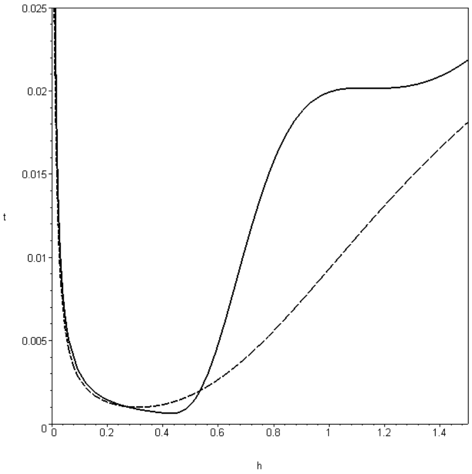
Figure 2: MISE, n=100



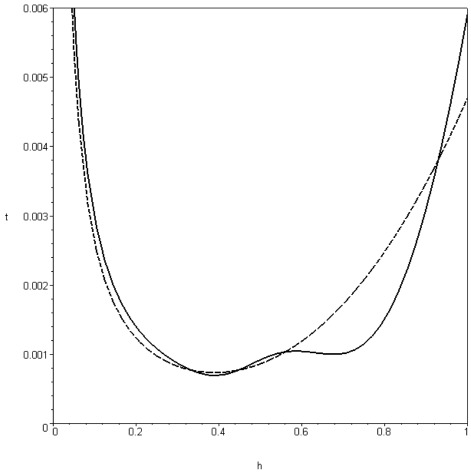
(a) Normal Density



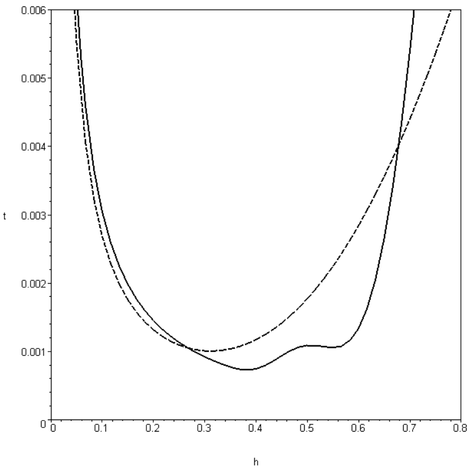
(b) Bimodal Density



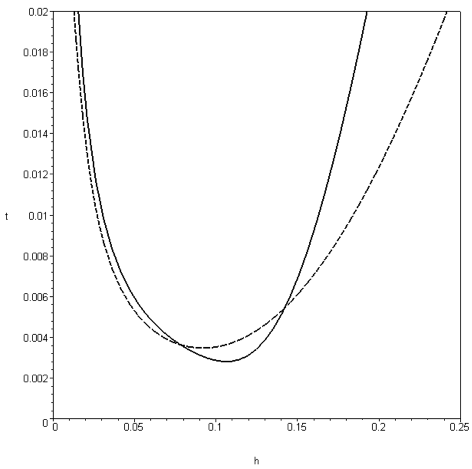
(c) Bimodal Density



(d) Plateau Density

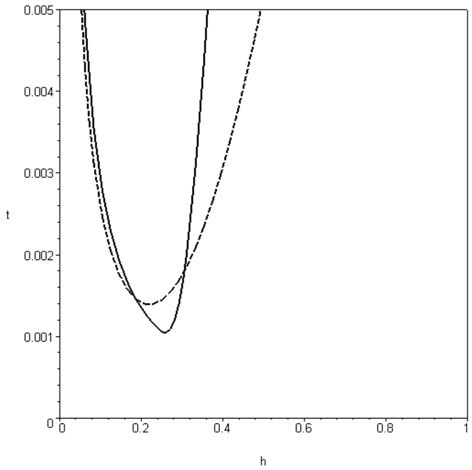


(e) Separated Bilmodal Density

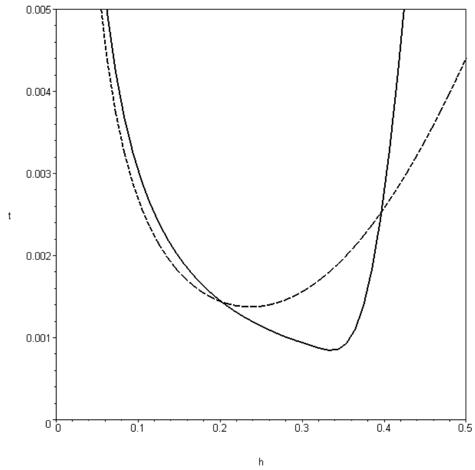


(f) Kurtotic Density

Figure 3: MISE, n=1000



(g) Skewed Unimodal Density



(h) Trimodal Density

Figure 3: MISE, n=1000

3. Optimal bandwidth of the Fourier integral estimator and its estimation

Data-driven bandwidth selection is a large and difficult topic for conventional estimators and even more for many higher-order estimators, see Jones et al. (1996), Jones and Signorini (1997). At the same time, a simple representation of the MISE of the Fourier integral estimator suggests simple methods of bandwidth selection.

The characteristic function of the sinc estimator is simply the indicator

$$\psi(t) = I_{[-1,1]}(t)$$

(so the sinc kernel can be considered as a limit case of superkernels as $c \to 1$). Let $\varphi(t)$ be the characteristic function of the density to be estimated. Then the MISE of the Fourier integral estimator is represented as

$$\text{MISE} = \frac{1}{2\pi}\int_{|t|>1/h} |\varphi(t)|^2 dt + \frac{1}{n}\cdot\frac{1}{2\pi}\int_{-1/h}^{1/h}(1-|\varphi(t)|^2)dt =$$

$$= \frac{1}{\pi n h} + R(f) - \left(1+\frac{1}{n}\right)\frac{1}{\pi}\int_0^{1/h}|\varphi(t)|^2 dt, \tag{1}$$

where

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

It follows from this representation that local minimums of the MISE are solutions of the equation

$$|\varphi(1/h)| = \frac{1}{\sqrt{n+1}} \tag{2}$$

for which $|\varphi(t)|$ decreases in some neighbourhood. Fortunately, for moderate n ($n < 1000$) and for more or less regular densities, there are only one or two such solutions. For example, for normal mixtures, considered in Section 2, the number of local minimums of the MISE of the Fourier integral estimator is given in Table 2. For smaller n situation is even better because the number of solutions of equation (2) decreases as n decreases.

Table 2

Number of local minimums of the MISE

	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10
$n = 100$	1	2	2	1	2	1	1	2	1	1
$n = 1000$	1	2	2	2	2	1	1	2	1	1

Glad et al. (2007) suggested the following methods of bandwidth selection. One method consists in replacing the characteristic function φ in (2) by the empirical characteristic function. Let $\varphi_n(t)$ be the empirical characteristic function associated with the sample $X_1, ..., X_n$, i.e.

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

It is known that $\varphi_n(t)$ converges to $\varphi(t)$ uniformly on each interval, therefore roots of equation (2) can be approximated by roots of the equation

$$|\varphi_n(1/h)| = \frac{1}{\sqrt{n+1}}. \tag{3}$$

So first, roots of (3) are found such that $|\varphi_n(t)|$ decreases in a neighbourhood of each root. Then the approximate MISE is calculated using (1) with $\varphi(t)$ replaced by $\varphi_n(t)$ (and with an appropriate truncation for calculation the integrals). Bandwidth is selected as a root having minimal approximate MISE.

Now suppose that the MISE has one local minimum. Then the suggested bandwidth selector becomes

$$\hat{h}_{\rm ecf} = 1/\delta_0 \tag{4}$$

where

$$\delta_0 = \min\{\delta : \ |\varphi_n(\delta)| = (n+1)^{-1/2}\}. \tag{5}$$

Another method from Glad et al. (2007) is based on the idea of the normal rule: if we suppose that X_i have the standard normal distribution, then the unique solution of (2) is

$$h_{\rm opt} = \frac{\sigma}{\sqrt{\ln(n+1)}},$$

where σ^2 is the variance of X_i. Thus the normal rule selector is

$$\hat{h}_{\rm norm} = \frac{\hat{\sigma}}{\sqrt{\ln(n+1)}}, \tag{6}$$

where $\hat{\sigma}^2$ is a reasonable estimator of σ^2, for example

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2, \quad \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

In the next section, we study the performance of these two selectors using simulation.

4. Performance of the bandwidth selectors

The first bandwidth selector, suggested by Glad et al. (2007), which we briefly described in the previous section, consists of two stages: approximation of solutions of equation (2) and choice of the solution with minimal approximate MISE. Since in this work the first stage is our main target, we simplify the problem and consider those densities for which the MISE has only one local minimum, namely densities #1, #4, #6, #7, #9, #10 for $n = 100$, and #1, #6, #7, #9, #10 for $n = 1000$. For each of this densities we find $h_{\rm opt}$ — the optimal bandwidth. Then we simulate 100 independent samples from the corresponding underlying distribution and find 100 values of $\hat{h}_{\rm ecf}$ using (4) and (5) and 100 values of $\hat{h}_{\rm norm}$ using (6). On the basis of these values we estimate expectations $\mathrm{E}\,\hat{h}_{\rm ecf}$ and $\mathrm{E}\,\hat{h}_{\rm norm}$ and standard errors $(\mathrm{Var}\,\hat{h}_{\rm ecf})^{1/2}$ and $(\mathrm{Var}\,\hat{h}_{\rm norm})^{1/2}$ of these selectors. Results are presented in Tables 3 and 4. In addition, the number of $\hat{h}_{\rm ecf}$ and $\hat{h}_{\rm norm}$, for which the absolute value of the relative error is less than 0.2, i.e. $|\hat{h}_{\rm ecf}/h_{\rm opt}-1|<0.2$ and $|\hat{h}_{\rm norm}/h_{\rm opt}-1|<0.2$, are presented in the tables ($N_{\rm ecf}$ and $N_{\rm norm}$ respectively), so that one can see in how many cases a method under consideration gives a good approximation.

The simulation shows that the normal rule selector is good only when the density to be estimated is normal or very close to normal (but in this case it is very good), otherwise it is very little effective. But this must be so not only for the sinc estimator but for all other kernels as well, because the selector depends only on one parameter of the density — the variance, which can be the same for densities with very different shapes and, therefore, for very different optimal bandwidths.

As to the empirical characteristic function based selector, it seems to be very promising. Comparison parameters of $\hat{h}_{\rm ecf}$ with parameters of some hi-tech selectors, for the latter see for example Park and Marron (1990), displays that $\hat{h}_{\rm ecf}$ usually has substantially better performance.

Analysis of simulated $\hat{h}_{\rm ecf}$ shows that values, which are not close to $h_{\rm opt}$, often appear as outliers. An example is in Table 5. Here those values, for which $|\hat{h}_{\rm ecf}/h_{\rm opt}-1|>0.2$, are marked. This is also an advantage of the selector. First, outliers are easier disclosed. Secondly, the rest of the sample (without outliers) has smaller scatter than the whole sample.

Some modifications of $\hat{h}_{\rm ecf}$ seem to be reasonable. For example, the empirical characteristic function can be replased by the uniformly consistent estimator suggested by Lebedeva and Ushakov (2007). There are grounds to expect that this replacement can reduce both the bias and the variance of the selector.

Table 3

Performance of the selectors, $n = 100$

	$h_{\rm opt}$	$\mathrm{E}\,\hat{h}_{\rm ecf}$	$(\mathrm{Var}\,\hat{h}_{\rm ecf})^{1/2}$	$N_{\rm ecf}$	$\mathrm{E}\,\hat{h}_{\rm norm}$	$(\mathrm{Var}\,\hat{h}_{\rm norm})^{1/2}$	$N_{\rm norm}$
#1	0.465	0.405	0.106	59	0.460	0.034	80
#4	0.752	0.688	0.142	75	0.652	0.043	99
#6	0.139	0.133	0.027	63	0.570	0.064	0
#7	0.279	0.258	0.071	50	0.379	0.036	4
#9	0.161	0.181	0.059	55	0.646	0.041	0
#10	0.482	0.451	0.116	77	0.887	0.052	0

Table 4

Performance of the selectors, $n = 1000$

	$h_{\rm opt}$	$\mathrm{E}\,\hat{h}_{\rm ecf}$	$(\mathrm{Var}\,\hat{h}_{\rm ecf})^{1/2}$	$N_{\rm ecf}$	$\mathrm{E}\,\hat{h}_{\rm norm}$	$(\mathrm{Var}\,\hat{h}_{\rm norm})^{1/2}$	$N_{\rm norm}$
#1	0.380	0.328	0.082	57	0.382	0.010	90
#6	0.106	0.102	0.016	84	0.473	0.017	0
#7	0.232	0.202	0.047	53	0.310	0.010	0
#9	0.115	0.115	0.013	88	0.527	0.011	0
#10	0.419	0.382	0.063	82	0.731	0.014	0

Table 5

A sample of $\hat{h}_{\text{ecf}}$, density #10, $n = 100$, $h_{\text{opt}} = 0.482$

0.498	$\boxed{\mathbf{0.349}}$	0.467	0.457	$\boxed{\mathbf{0.312}}$	0.496	0.485	0.481	0.464	0.462
0.391	0.462	0.450	0.467	0.436	$\boxed{\mathbf{0.269}}$	0.530	0.491	$\boxed{\mathbf{0.319}}$	0.448
0.424	0.442	$\boxed{\mathbf{0.356}}$	$\boxed{\mathbf{0.302}}$	0.491	0.462	$\boxed{\mathbf{0.276}}$	$\boxed{\mathbf{1.241}}$	0.519	0.496
0.490	$\boxed{\mathbf{0.296}}$	0.512	0.402	0.509	0.492	0.487	0.477	0.467	0.543
0.424	0.508	0.472	0.423	0.419	0.427	0.444	0.547	0.379	0.515
0.472	0.473	$\boxed{\mathbf{0.310}}$	0.545	0.457	0.466	0.521	0.457	0.484	0.427
0.428	0.424	0.522	0.405	0.495	0.478	0.479	$\boxed{\mathbf{0.252}}$	$\boxed{\mathbf{0.565}}$	0.536
0.517	0.496	0.439	$\boxed{\mathbf{0.233}}$	0.433	0.509	$\boxed{\mathbf{0.325}}$	0.450	0.382	$\boxed{\mathbf{0.318}}$
0.428	0.481	0.529	0.452	$\boxed{\mathbf{0.256}}$	0.544	$\boxed{\mathbf{0.578}}$	0.471	0.455	0.484
0.472	0.504	$\boxed{\mathbf{0.339}}$	0.540	$\boxed{\mathbf{0.326}}$	0.510	$\boxed{\mathbf{1.130}}$	0.370	0.468	0.481

References

Davis, K.B. (1975). Mean square error properties of density estimates. *Ann. Statist.*, **3**, no. 4, 1025-1030.

Davis, K.B. (1977). Mean integrated square error properties of density estimates. *Ann. Statist.*, **5**, no. 3, 530-535.

Glad, I.K., Hjort, N.L. and Ushakov, N.G. (2003). Correction of density estimators that are not densities. *Scand. J. Statist.*, **30**, no. 2, 415-427.

Glad I.K., Hjort N.L. and Ushakov N.G. (2007) Density estimation using the sinc kernel. Statistics Preprint No. 2. Department of Mathematical Sciences, NTNU, Trondheim, Norway (http://www.math.ntnu.no/preprint/statistics).

Jones, M.C., Marron, J.S. and Sheather, S.J. (1996). A brief survey of bandwidth selection for density estimation. *J. Amer. Stat. Assoc.*, **91**, 401-407.

Jones, M.C. and Signorini, D.F. (1997). A comparison of higher-order bias kernel density estimators. *J. Amer. Stat. Assoc.*, **92**, 1063-1073.

Lebedeva T.V. and Ushakov N.G. (2007). Uniformly consistent estimators of the characteristic function. Statistics Preprint No. 7. Department of Mathematical Sciences, NTNU, Trondheim, Norway (http://www.math.ntnu.no/preprint/statistics).

Marron, J.S. and Wand, M.P. (1992). Exact mean integrated squared error. *Ann. Statist.*, **20**, 712-736.

Park B.U. and Marron J.S. (1990). Comparison of data-driven bandwidth selectors. *J. Amer. Stat. Assoc.*, **85**, 66-72.