OPTIMAL BANDWIDTH SELECTION FOR DECONVOLUTED KERNEL

DENSITY ESTIMATION USING BOOTSTRAP METHOD

by

SOUAD SOSA

Presented to the Faculty of the Graduate School of

The University of Texas at Arlington in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

THE UNIVERSITY OF TEXAS AT ARLINGTON

August 2020

Dedicated to my husband Alfonso Sosa and my children Ilias and Ryan Sosa. I will remain grateful to their encouragement and support.

ACKNOWLEDGEMENTS

Firstly, I would like to express my sincere gratitude to my advisor Dr. Shan Sun-Mitchell for her continuous support of my Ph.D. study and related research, also for her patience, motivation, and immense knowledge. Her guidance has helped me during the time of research and writing this thesis.

Besides my advisor, I would like to thank the rest of my thesis committee members, Dr. Andrzej Korzeniowski , Dr. Ren-Cang Li, and Dr. Li Wang.

From Dr. Andrzej Korzeniowski, I have benefited by gaining an overview how different branches of mathematics are connected to each other, and solving problems he assigned taught me how to master understanding of theorems. I am very appreciative of Dr. Ren-Cang Li for giving me the opportunity to improve my proofs and allowing me to consult him for any questions in mathematics.

My sincere thanks also goes to Dr. Li Wang, for her help with programming and offering new ideas, and for going above and beyond in guiding me.

I wish to thank Dr. Ruth Gornet and Dr. Carole Krueger, for allowing me to acquire strong critical thinking skills. I also wish to thank Dr. Hristo Kojouharov, Dr. James Grover, Dr. Laura Mydlarz, and Dr. Christopher Kribs for their guidance throughout the UTTER (Undergraduate Training in Theoretical Ecology Research) program.

Many thanks to Dr. Mahmoud Jawad for challenging me when solving mathematical assignments. These challenges have contributed significantly to improve my reasoning skills. Thanks also to Dr. Wilber Ventura, Mr. Mohammad Latifi Jebelli,

iv

Dr. Kiran Kumar Mainali, Dr. Honguang Xi, Dr. Izzet Sozucok, and Mr. Hyung Wook Chun, as they have all inspired me and motivated me.

<div align="right">August 3, 2020</div>

ABSTRACT

OPTIMAL BANDWIDTH SELECTION FOR DECONVOLUTED KERNEL

DENSITY ESTIMATION USING BOOTSTRAP METHOD

SOUAD SOSA, Ph.D.

The University of Texas at Arlington, 2020

Supervising Professor: Dr. Shan Sun-Mitchell

To estimate an unknown density when observed measurements are from the convolution model contaminated by additive measurement errors, Stefanski and Carroll (1990) proposed using Fourier inversion on the product of Fourier transform of a kernel function and the characteristic function of the error variable. One important element in constructing such a density estimator is the bandwidth. The goal of this research is to establish an optimal bandwidth so that the mean integrated squared error of the estimator is minimized. The bootstrap method is used to accomplish this goal. The simulation results show that the estimated optimal bandwidths provide adequate estimation to the unknown densities.

<p align="center">TABLE OF CONTENTS</p>

LIST OF TABLES

CHAPTER 1

INTRODUCTION

The *deconvolving kernel density estimator* is a method which allows to estimate a population density function from a contaminated sample. Let us consider the estimation of an unknown continuous population density function from a sample containing random errors. This is a type of deconvolution problem which has many applications in the fields of engineering, econometrics, astronomy, public health, and others. Kernel density estimators have been used extensively since Rosenblatt (1956) [11]. They have proven to be useful when estimating an unknown density population f_X. The deconvolving kernel density estimator is a suitable estimator when random variables contain errors as contaminated samples are common in applications. Although many articles have described its mechanism and usage, many issues still arise such as estimating the optimal bandwidth given a contaminated random sample.

To introduce our model, we let X, Y, and Z be random variables which satisfy the general model

$$Y = X + Z,$$

where X is an unobservable random variable from an unknown p.d.f. (population density function) f_X, Y is an observable random variable with an unknown p.d.f. f_Y, and Z is a random variable representing the random error with the p.d.f. f_Z. We also assume that the random variables X and Z are independent.

Throughout this thesis, we assume f_Z is known and we choose the Laplace error distribution and discuss the reason for its selection in Chapter 2. Note that contrary to f_X and f_Y we know the distribution for f_Z. The assumption that we

1

know the distribution for f_z appears restrictive. Nevertheless, in reality it is common to obtain insufficient information regarding f_z, see Delaigle and Gijbels (2004) [3]. It is possible to consider the case where the distribution of Z is known. However, that case will not be considered in this thesis.

Our goal is to estimate the unknown density f_X based on a given observable random contaminated sample $Y_1, Y_2, ..., Y_n$. We achieve our goal by estimating the optimal bandwidth h for our deconvolving kernel density estimator. The bandwidth h is a scaling factor which influences the spread of the kernel density. When h is significantly small, it introduces too much variance, whereas when h is significantly large it introduces too much bias. The optimal bandwidth h is the one which establishes an equilibrium between the variance and the bias, see Wand and Jones (1995) [15]. We estimate the unknown density f_X via the convolution $f_Y = f_X * f_Z$ and

$$\varphi_{f_X}(t) = \frac{\varphi_{f_Y}(t)}{\varphi_{f_Z}(t)}, \tag{1.1}$$

where φ is the characteristic function of the corresponding density. The Fourier inversion theorem describes $f_X(x)$ as

$$\begin{aligned} f_X(x) &= \frac{1}{2\pi} \int e^{-itx} \, \varphi_{f_X}(t) \, dt \\ &= \frac{1}{2\pi} \int e^{-itx} \left(\frac{\varphi_{f_Y}(t)}{\varphi_{f_Z}(t)} \right) dt, \end{aligned}$$

and its estimator $\hat{f}_X(x)$ as

$$\hat{f}_X(x) = \frac{1}{2\pi} \int e^{-itx} \left(\frac{\varphi_{\hat{f}_Y}(t)}{\varphi_{f_Z}(t)} \right) dt. \tag{1.2}$$

Note that in the integrals above the t-values run over the common support of the respective characteristic functions φ_{f_X}, φ_{f_Y}, and φ_{f_Z}.

Given the observable i.i.d. sample $Y_1, Y_2, ..., Y_n$, the unknown density $f_Y(x)$ can be estimated by using its kernel estimator

$$\hat{f}_Y(x, h) = \frac{1}{nh} \sum_{j=1}^{n} K\left(\frac{x - Y_j}{h}\right), \tag{1.3}$$

where we use $\hat{f}_Y(x, h)$ instead of $\hat{f}_Y(x)$ to emphasize the dependence on the parameter h and we use K to denote the kernel function. Note that h actually depends on sample size n but in our notation we simply use h and suppress the dependence on n. We recall that in general the quantity h is actually a sequence of positive parameters called bandwidths. The kernel function K appearing in (1.3) is assumed to satisfy the following conditions:

(i) $\displaystyle\int_{-\infty}^{\infty} K(x)\, dx = 1.$

(ii) $\displaystyle\int_{-\infty}^{\infty} x\, K(x)\, dx = 0.$

(iii) $\displaystyle\int_{-\infty}^{\infty} x^2\, K(x)\, dx < \infty.$

Note that in (1.2) we have

$$\begin{aligned}
\varphi_{\hat{f}_Y}(t) &= E\left(e^{-ity}\right) \\
&= \int e^{-ity}\, \hat{f}_Y(x, h)\, dt \\
&= \frac{1}{nh} \sum_{j=1}^{n} \int e^{-ity}\, K\left(\frac{x - Y_j}{h}\right) dt.
\end{aligned}$$

Letting $u = (x - Y_j)/h$, we get

$$x = Y_j + hu, \quad du = \frac{dx}{h}.$$

Consequently, we obtain

$$\varphi_{\hat{f}_Y}(t) = \frac{1}{nh} \sum_{j=1}^{n} \int e^{-it(Y_j + hu)}\, K(u)\, h\, du$$

3

$$= \frac{1}{n} \sum_{j=1}^{n} \int e^{-itY_j} e^{-ithu} K(u) \, du \qquad (1.4)$$

$$= \frac{1}{n} \sum_{j=1}^{n} e^{-itY_j} \int e^{-ithu} K(u) \, du$$

$$= \frac{1}{n} \sum_{j=1}^{n} e^{-itY_j} \varphi_K(th).$$

Using (1.4) in (1.2) we have

$$\hat{f}_X(x, h) = \frac{1}{2\pi} \int e^{-itx} \left(\frac{\frac{1}{n} \sum_{j=1}^{n} e^{-itY_j} \varphi_K(th)}{\varphi_{f_Z}(t)} \right) dt$$

$$= \frac{1}{2\pi n} \sum_{j=1}^{n} \int e^{-it(x-Y_j)} \left(\frac{\varphi_K(th)}{\varphi_{f_Z}(t)} \right) dt.$$

We now use a particular choice of the bandwidth h. Letting $v = th$ we get

$$\hat{f}_X(x, h) = \frac{1}{2\pi n} \sum_{j=1}^{n} \int e^{-iv\left(\frac{x-Y_j}{h}\right)} \frac{\varphi_K(v)}{\varphi_{f_Z}\left(\frac{v}{h}\right)} \frac{dv}{h}$$

$$= \frac{1}{nh} \sum_{j=1}^{n} \frac{1}{2\pi} \int e^{-it\left(\frac{x-Y_j}{h}\right)} \frac{\varphi_K(v)}{\varphi_{f_Z}\left(\frac{v}{h}\right)} dv \qquad (1.5)$$

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

where the deconvoluted kernel density estimator K^Z is given by

$$K^Z(x, h) = \frac{1}{2\pi} \int e^{-ivx} \left(\frac{\varphi_K(v)}{\varphi_{f_Z}\left(\frac{v}{h}\right)} \right) dv.$$

We assume that the deconvoluted kernel density estimator $K^Z(\cdot, h)$ is a real-valued function. For the rest of this thesis we use the following assumption regarding the kernel density estimator $K(x)$:

Assumption 1 on the kernel density estimator: We assume that the kernel

4

density estimator $K(x)$ satisfies the property that it is a bounded p.d.f. with a finite fourth moment.

The triweight function is used as the kernel by Delaigle and Gijbels (2004) [3]. Here in our thesis we use the triweight function as well as two additional functions as our choice of kernels:

(i) The de la Vallée-Poussin kernel: $K(x) = \dfrac{1 - \cos x}{\pi x^2}$.

(ii) The Gaussian kernel: $K(x) = \dfrac{1}{\sqrt{2\pi}} \exp\left(\dfrac{-x^2}{2}\right)$.

(iii) The triweight kernel: $K(x) = \dfrac{48 \cos x}{\pi x^4}\left(1 - \dfrac{15}{x^2}\right) - \dfrac{144 \sin x}{\pi x^5}\left(2 - \dfrac{5}{x^2}\right)$.

Note that the de la Vallée-Poussin kernel does not meet the criteria of having a finite second moment. However, the reason for its selection is to emphasize the kernel involvement when constructing a pseudo density $\hat{f}_X(\cdot, g)$, which we will discuss in Chapter 2.

Our goal is to estimate the optimal bandwidth h by choosing the bandwidth h-value which minimizes $\mathrm{MISE}(\hat{f}_X(\cdot, h))$, which denotes the mean integrated squared error for the deconvoluted kernel density estimator $\hat{f}_X(x, h)$. The bootstrap method is a reliable method to achieve the approximation of such an optimal bandwidth.

To be able to use a bootstrap procedure, we need to first establish a bootstrap sampling from a pseudo density $\hat{f}_X(\cdot, g)$, which is itself obtained by constructing an initial bandwidth g called the pilot bandwidth. In Chapter 2, we show how to obtain such a pilot bandwidth g. The asymptotic representation of $\mathrm{MISE}(\hat{f}_X(\cdot, h))$ serves as the foundation to establish the two-step procedure, from which we will select the initial pilot bandwidth g. Our goal is to construct a pseudo density $\hat{f}_X(\cdot, g)$. Therefore, once g is selected, we focus our attention on the smoothness of the error

distribution $f_z(z)$, as the choice of the type of smoothness directly impacts the rate of convergence of the estimator. Finally, we test each of the three kernels described earlier to observe the overall role of the kernel function within the pseudo density $\hat{f}_X(\cdot, g)$.

CHAPTER 2

The construction of the pilot bandwidth g

In this chapter our goal is to determine a suitable pilot bandwidth g. To reach this goal, we must first visit the asymptotic representation of $\text{MISE}(\hat{f}_X(\cdot, h))$, as it serves as the foundation of a two-stage procedure. Note that we use g when referring to an initial pilot bandwidth, whereas h is the unique optimal bandwidth we aim to approximate.

A common way to approximate the optimal bandwidth h is to choose its approximate in such a manner that it minimizes the distance between the population density $f_X(x)$ and the deconvoluted kernel density estimator $\hat{f}_X(x, h)$. This is done by investigating the quantity

$$\text{MISE}\left(\hat{f}_X(\cdot, h)\right) = E \int \left(\hat{f}_X(x, h) - f_X(x)\right)^2 dx.$$

In applications it remains challenging to obtain such an optimal bandwidth h, as it acts on the kernel density in a complicated way. We start the estimation of the optimal bandwidth h by first constructing an initial pilot bandwidth g. The construction of such a bandwidth g relies on the asymptotic representation of $\text{MISE}\left(\hat{f}_X(\cdot, h)\right)$ as

$$\text{AMISE}\left(\hat{f}_X(\cdot, h)\right) = \frac{h^4}{4} \mu_2^2(K) R(f_X'') + \frac{1}{2\pi n h} \int \frac{|\varphi_K(t)|^2}{|\varphi_{f_Z}(\frac{t}{h})|^2} dt, \qquad (2.1)$$

where we have

$$\mu_2(K) = \int u^2 K(u) du,$$

and where for any square integrable function f we have

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

7

We remark that $\text{AMISE}\left(\hat{f}_X(\cdot, h)\right)$ denotes the asymptotic mean integrated square error of $\hat{f}_X(\cdot, h)$. The details for (2.1) are provided in Chapter 3.

Note that in (2.1) one can observe the behavior of h discussed in the introduction. In the first term on the right-hand side of (2.1) the bias becomes larger as h becomes larger, whereas in the second term there the variance becomes larger as h approaches 0. This is what Wand and Jones (1995) [15] refers to as the "variance-bias trade off." Although (2.1) is a simpler representation of MISE, its first term contains $R(f_X'')$ which is unknown. Note that f_X'' denotes the second derivative of $f_X(x)$ with respect to x. The reference by Delaigle and Gijbels (2004) [3] has shown that substituting the term $R(\hat{f}_X''(\cdot, g))$ for $R(f_X'')$ can be used to estimate (2.1). The term $R(\hat{f}_X''(\cdot, g))$ is then the only quantity relying on the bandwidth g, and hence g must be the optimal bandwidth to estimate $R(f_X'')$. The reference Delaigle and Gijbels (2002) [4] shows that such an optimal bandwidth g is obtained by minimizing the mean square error of $R(\hat{f}_X''(\cdot, g))$ denoted by $\text{MSE}[R(\hat{f}_X''(\cdot, g))]$. These two authors also show that asymptotically the optimal bandwidth g which reduces $\text{MSE}[R(\hat{f}_X''(\cdot, g))]$ is also the one which reduces the asymptotic bias for $R(\hat{f}_X''(\cdot, g))$. The asymptotic bias for $\hat{f}_X^{(r)}(\cdot, g_r)$ is given by

$$\text{ABias}\left(\hat{f}_X^{(r)}(\cdot, g_r)\right) = -g_r^2 \mu_2(K) R(f_X^{(r+1)}) + \frac{1}{2\pi n g_r^{(2r+1)}} \int t^{2r} \frac{|\varphi_K(t)|^2}{|\varphi_{f_Z}(\frac{t}{g_r})|^2} \, dt. \quad (2.2)$$

Note that the r-values in (2.2) are nonnegative integers. In our notation $f_X^{(r)}(x)$, the superscript in parentheses indicates the rth derivative of the density function $f_X(x)$ with respect to x. It is clear that $f_X^{(0)}(x)$ is the same as $f_X(x)$ itself. The expression in (2.2) is the foundation of the two-stage procedure which provides an initial pilot bandwidth g to construct pseudo density $\hat{f}_X(\cdot, g)$.

2.1 The selection of the pilot bandwidth g

In this section we are going to pick the initial pilot bandwidth g as described by the two-stage procedure. Once the pilot bandwidth g is constructed, the pseudo density $\hat{f}_X(\cdot, g)$ can be derived and provide us with a preliminary account of the underlying distribution $f_X(x)$ as well being the basis of our bootstrap method, which will be discussed in Chapter 3.

Recall that in (2.1) we have used the notation $R(f) = \int f(x)^2 dx$ and now we use a similar notation $R(f_X^{(r)})$ in case we use the rth instead of the second derivative. We refer to the quantity $R(f_X^{(r)})$ as the integrated square density derivative functional and we note that

$$R(f^{(r)}) = \int f^{(r)}(x)^2 \, dx.$$

Before visiting the two-stage procedure to obtain the optimal pilot bandwidth g, we analyze the order k of the kernel, where k is a nonnegative integer not exceeding a specific nonnegative integer value l. The references Wand and Jones (1995) [15] and Marron (1994) [10] show that a kernel $K(x)$ of order k satisfies

$$\int x^l K(x) \, dx = \begin{cases} 1, & l = 0, \\ 0, & l = 1, ..., k-1, \\ C > 0, & l = k, \end{cases} \tag{2.3}$$

where C is a positive quantity.

Obtaining the optimal pilot bandwidth g is achieved by implementing a two-stage procedure, as the number of stages corresponds to the order of the kernel and our kernel is a second order kernel. Our goal is to estimate $R(f_X'')$ which is related to the second derivative. However, we will start by estimating the related quantity involving the fourth derivative, and this will enable us to estimate the quantity related

9

to the third derivative. Finally, from the estimate for the third derivative we will be able to estimate the quantity related to the second derivative. We first estimate $R(f_X^{(4)})$ and obtain the integrated square density derivative with $r = 4$, which is denoted by $\widehat{R(f_X^{(4)})}$. The procedure is described as follows:

1. Step 0: The normal reference method indicates that

$$\widehat{R(f_X^{(4)})} = \frac{8! S_X^{-9}}{2^9 4! \sqrt{\pi}},$$

   see Delaigle and Gijbels (2004) [5] and Jones and Wand (1995) [15], for which we assume sample variance $S_X^2 = S_Y^2 - \mathrm{Var}(Z)$.

2. Step 1: Replace $R(f_X^{(4)})$ by $\widehat{R(f_X^{(4)})}$ in the asymptotic bias expression (2.2). Then, select a bandwidth g_3 such that g_3 is the optimal bandwidth to estimate $R(f_X^{(3)})$, meaning that g_3 reduces the asymptotic bias given to derive $R(\hat{f}_X^{(3)}(\cdot, g_3))$.

3. Step 2: $R(\hat{f}_X^{(3)}(\cdot, g_3))$ estimates $R(f_X^{(3)})$. Therefore, we substitute $R(\hat{f}_X^{(3)}(\cdot, g_3))$ for $R(f_X^{(3)})$ in (2.2). Then, we select the optimal bandwidth g_2 to derive $R(\hat{f}_X^{(2)}(\cdot, g_2))$ in order to estimate $R(f_X'')$, where g_2 is our pilot bandwidth g.

Remark: the two-stage procedure rests on the normal scaling reference. The normal scale requires the assumption that the density f_X is normal in $\widehat{R(f_X^{(4)})}$. This method is typically used as a preliminary analysis. As discussed by Delaigle and Gijbels (2004) [3], Delaigle and Gijbels (2004) [5], Chu, Henderson, and Parmeter (2015) [1], and Wand and Jones (1995) (page 72) [15], even if the random variable X does not follow a normal distribution, the use of g still provides an accurate representation of the target density, provided this target density does not contain a strong multimodality and/or an asymmetry. Simulation results show that for a target density such as Exponential(1) which lacks the symmetry property, we obtain good results only toward the tail of the distribution. Table 2.1 displays the values for g-values for the three types of kernel densities with the two target densities $N(0, 1)$

and the bimodal Gaussian((4,1),(7,1)) density for the three sample sizes $n = 200$, $n = 400$, and $n = 500$.

Table 2.1: The pilot bandwidth g estimations using the Laplace error distribution

Target Densities

Kernel	**Standard Normal**			**Bimodal Gaussian((4,1),(7,1))**		
	n = 200	**n=400**	**n=500**	**n = 200**	**n=400**	**n=500**
dlVP	0.1976	0.1575	0.1409	0.2058	0.1863	0.1762
Gaussian	0.4160	0.3457	0.3235	0.4339	0.4222	0.3708
Triweight	0.1808	0.1470	0.1347	0.1925	0.1863	0.1684

In spite of the choice of kernel K, we observe that as the sample sizes increase the value for the pilot bandwidth g decreases. This is also true when the target density is Exponential(1).

2.2   The choice of the error distribution

Having established a pilot bandwidth g using our two target densities, we are now ready to construct the pseudo density $\hat{f}_X(\cdot, g)$. Notice that this process involves the distribution for the error density $f_z(z)$. For the estimator to be consistent, we need to follow some assumptions regarding the error distribution. Jianqing Fan (1991) [6] has discussed the rate of convergence for the estimator $\hat{f}_X(x, h)$ and pointed out that it heavily depends on the tail of the characteristic of the error distribution. If we want to use any of the asymptotic properties for the estimator, then we must have the following conditions:

(i) $d_0|t|^{-\beta} \le \varphi_{f_z}(t) \le d_1|t|^{-\beta}$.

(ii) $d_0|t|^{\beta_0}e^{-\frac{|t|^\beta}{\gamma}} \le \varphi_{f_z}(t) \le d_1|t|^{\beta_1}e^{-\frac{|t|^\beta}{\gamma}}$,

11

where (i) is ordinary smooth, (ii) is supersmooth of order β as $t \to \infty$, the quantities d_0, d_1, β, and γ are positive constants and the parameters β_0 and β_1 are constants.

The rate of convergence for the kernel estimator is directly linked to the smoothness of the error distribution satisfying either condition (i) or (ii). In fact, since each condition for smoothness contains the term β, the value of β directly impacts the rate of convergence. Examples of ordinary smooth error distributions include the gamma distribution

$$f_z(z) = \frac{\alpha^m}{\Gamma(m)} z^{m-1} e^{-\alpha z}, \qquad \beta = m,$$

where m is a positive integer and the Laplace distribution

$$f_z(z) = \frac{1}{2} e^{|z|}, \qquad \beta = 2.$$

The distributions for $N(0,1)$ with $\beta = 2$ and for Cauchy$(0,1)$ with $\beta = 1$ are some examples of error densities satisfying super smooth conditions.

The reference Fan and Truong (1993) [7] explains how a super smooth error distribution displays a logarithmic convergence rate. Thus, selecting this kind of an error distribution will result in a very slow convergence while estimating $f_X(x)$ by $\hat{f}_X(x, h)$, whereas using an ordinary smooth error distribution allows the rate of convergence of $\hat{f}_X(x, h)$ to be fast. This specific reference describes the rate of convergence based on a different error distribution as shown in the following tables:

12

Table 2.2: The rate of convergence using the super smooth error distribution

Error Distribution	Rate of Convergence
$N(0,1)$	$\dfrac{1}{\sqrt{\log n}}$
Cauchy$(0,1)$	$\dfrac{1}{\log n}$

Table 2.3: The rate of convergence using the ordinary smooth error distribution

Error Distribution	Rate of Convergence
Gamma(α, m)	$\dfrac{1}{n^{\frac{1}{(2m+3)}}}$
Laplace	$\dfrac{1}{n^{1/7}}$

2.3   The kernel involvement

In Chapter 1, we defined the deconvolving kernel density estimator in (1.5). In this chapter we are showing the role of kernel $K(x)$ in $\hat{f}_X(x,h)$. The reference Jianqing Fan (1991) [6] shows the class of ordinary smooth error distribution will produce a faster convergence rate of the estimate than the class of supersmooth error distribution. Therefore, for the remainder of this thesis we select the ordinary smooth error characteristic distribution from the Laplace distribution, given by

$$\varphi_{f_Z}(t) = \frac{1}{1 + \dfrac{1}{2}\,\sigma^2\,t^2}.$$

13

Our simulations use the variance $\sigma^2 = 0.05$ as the contamination level is considered to be a reasonable representation for real life contaminated data.

We show that

$$K^Z(x, h) = K(x, h) - \frac{\sigma^2}{2h^2} K''(x, h), \tag{2.4}$$

provided that the kernel K'' for the second derivative of K exists on the support of K.

Note that we have

$$
\begin{aligned}
K^Z(x, h) &= \frac{1}{2\pi} \int e^{-itx} \left(\frac{\varphi_K(t)}{\varphi_{f_Z}\left(\frac{t}{h}\right)} \right) dt \\
&= \frac{1}{2\pi} \int e^{-itx} \frac{\varphi_K(t)}{\frac{1}{1 + \frac{1}{2h^2} \sigma^2 t^2}} dt \\
&= \frac{1}{2\pi} \int e^{-itx} \varphi_K(t) \left(1 + \frac{1}{2h^2} \sigma^2 t^2 \right) dt \tag{2.5} \\
&= \frac{1}{2\pi} \int e^{-itx} \varphi_K(t) \, dt + \frac{1}{2\pi} \int e^{-itx} \varphi_K(t) \frac{\sigma^2 t^2}{2h^2} dt \\
&= K(x, h) + \frac{\sigma^2}{2h^2} \frac{1}{2\pi} \int t^2 e^{-itx} \varphi_K(t) \, dt.
\end{aligned}
$$

We obtain

$$
\begin{aligned}
K(x, h) &= \frac{1}{2\pi} \int e^{-itx} \varphi_K(t) \, dt \\
&= \frac{1}{2\pi} \int (\cos tx - i \sin tx) \varphi_K(t) \, dt,
\end{aligned}
$$

and it follows that

$$K'(x, h) = \frac{1}{2\pi} \int (-t \sin tx - it \cos tx) \varphi_K(t) \, dt,$$

14

and

$$K''(x, h) = \frac{1}{2\pi} \int \left(-t^2 \cos tx + it^2 \sin tx \right) \varphi_K(t)\, dt$$
$$= \frac{1}{2\pi} \int -t^2 (\cos tx - i \sin tx)\, \varphi_K(t)\, dt$$
$$= \frac{-1}{2\pi} \int t^2\, e^{-itx}\, \varphi_K(t)\, dt.$$

Hence, (2.5) yields

$$K^Z(x, h) = K(x, h) - \frac{\sigma^2}{2h^2} K''(x, h).$$

Using an initial pilot bandwidth g we obtain

$$K^Z(x, g) = K(x, g) - \frac{\sigma^2}{2g^2} K''(x, g). \tag{2.6}$$

In Chapter 1 we have used three different kernel distributions in order to compare their performances against specific target densities. In this chapter we also use the de la Vallée-Poussin kernel, the Gaussian kernel, and the triweight kernel in (2.1), and we use the following assumptions which will be revisited in Chapter 3.

Assumption 2: As $n \to \infty$, we assume that

(i) $\left(\int |(K^Z * K^Z)(u, g)|\, du \right) \left(\int |t|^j \frac{|\varphi_K(t)|^2|}{|\varphi_{f_Z}\left(\frac{t}{g}\right)|^2}\, dt \right) = o(n^2\, g^{j+2}), \quad 0 \le j \le 8.$

(ii) $\int |(K * K^Z)(u, g)|\, du = o(\sqrt{n}).$

Next, we show that our pseudo density is a suitable estimator for any target density. To be able to do so, we draw a random sample from a chosen target density and add to it the noise from the Laplace error distribution in order to obtain $Y_1, Y_2, ..., Y_n$.

Note that the constructed g, the selected type of distribution, and the selected type of kernel function are the three necessary elements in the foundation of the pseudo density $\hat{f}_X(\cdot, g)$ given by

$$\hat{f}_X(\cdot, g) = \frac{1}{nh} \sum_{j=1}^{n} K^Z \left(\frac{x - Y_j}{g} \right).$$

2.4   The simulation results

From Table 2.1 we obtain the pilot bandwidth g from which we construct $\hat{f}_X(\cdot, g)$. We present our results when the underlying distribution satisfies the conditions stated earlier, namely that the underlying distribution density does not contain a strong multimodality and/or an asymmetry. Our simulations illustrate the pseudo densities $\hat{f}_X(\cdot, g)$ corresponding to the target population densities of $N(0,1)$ and the bimodal Gaussian$((4,1),(7,1))$. We have selected two sample sizes with $n = 200$ and $n = 500$. These results are the basis for our bootstrap procedure as described in Chapter 3. As Figure 2.1 and Figure 2.2 show, the effect of kernel selection on the estimation is insignificant for the estimation of the target density. These two figures also demonstrate that an increase in the sample size improves the performance of the estimation. We also demonstrate that when the underlying distribution does not satisfy all the properties mentioned above then the estimator does not perform well. This fact is represented in our simulations with the underlying distribution Exponential(1).
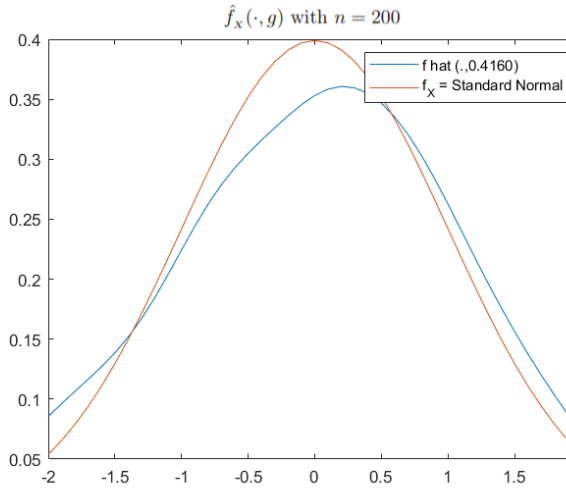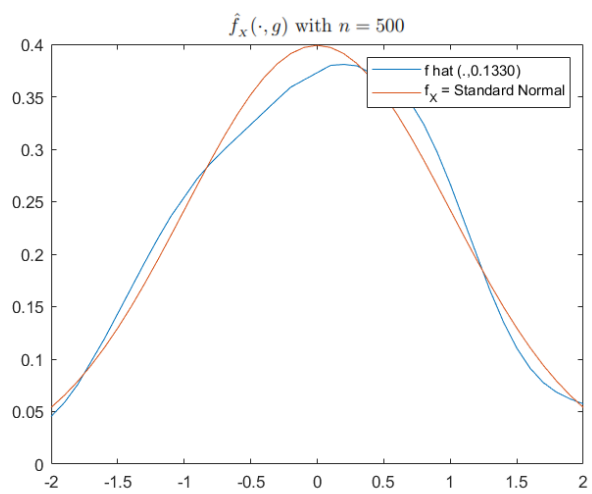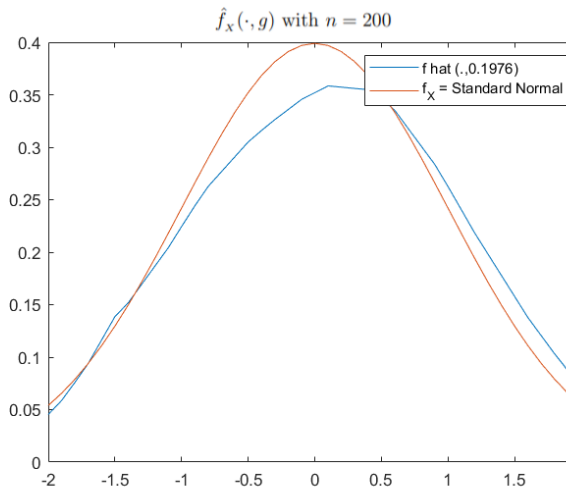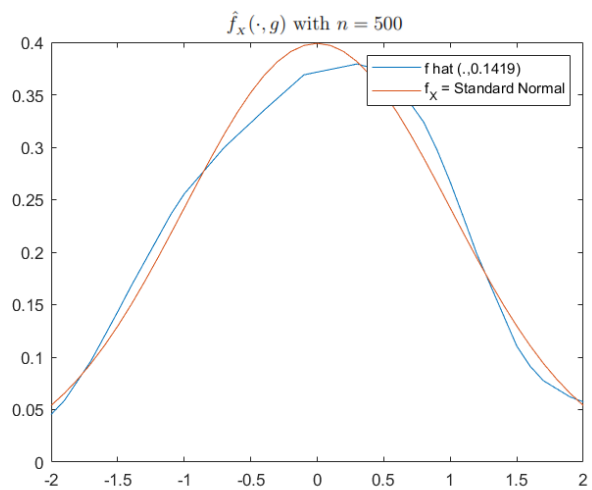
(a) n = 200 Gaussian

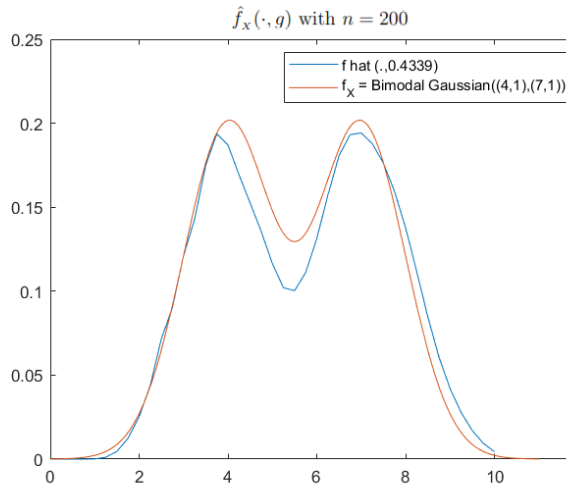(b) n = 500 Gaussian

(c) n = 200 Triweight

(d) n = 500 Triweight

(e) n = 200 dlVP
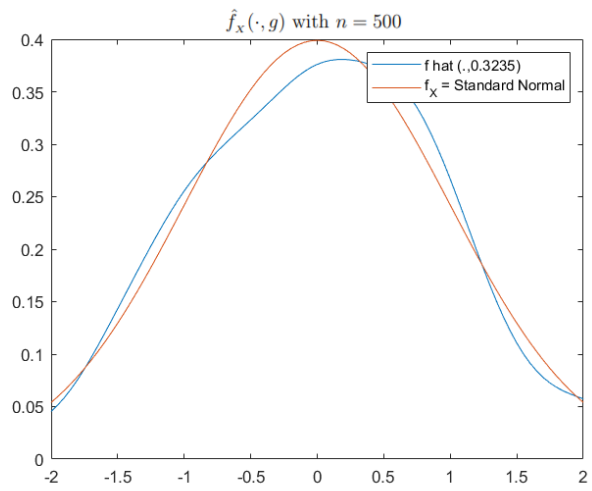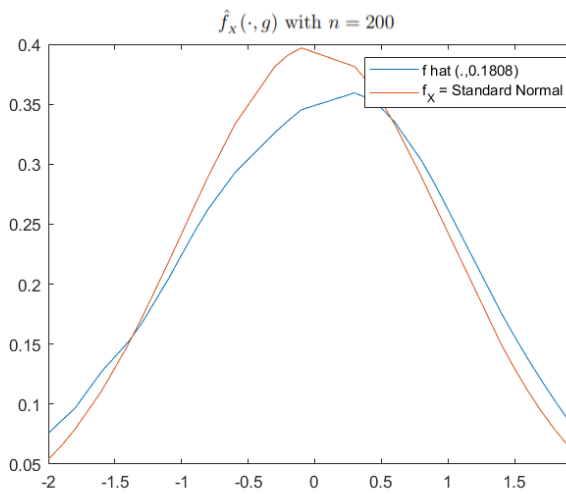
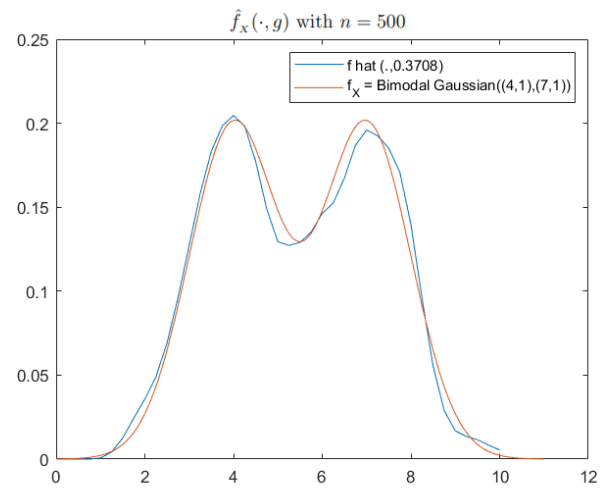(f) n = 500 dlVP
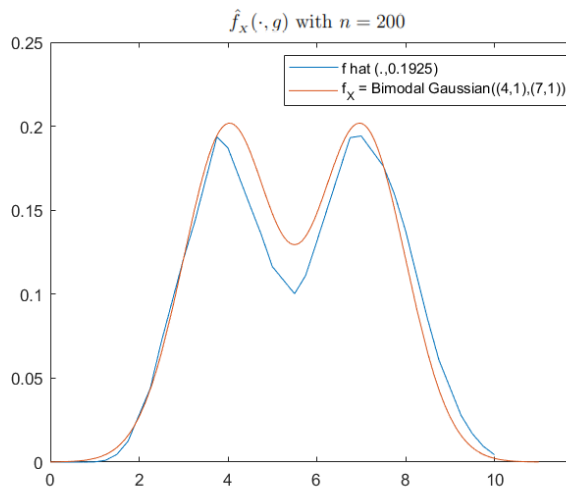
Figure 2.1: The pseudo densities $\hat{f}_X(\cdot, g)$ for $N(0,1)$ distribution with $n = 200$ and $n = 500$

(a) n = 200 Gaussian

(b) n = 500 Gaussian

(c) n = 200 Triweight

(d) n = 500 Triweight

(e) n = 200 dlVP

(f) n = 500 dlVP

Figure 2.2: The pseudo densities $\hat{f}_X(\cdot, g)$ for the bimodal Gaussian$((4,1),(7,1))$ distribution with $n = 200$ and $n = 500$

18

Again, when the target density is Exponential(1) it turns out that the performance of the estimation is poor. This is a direct consequence of the lack of symmetry for our target density, although the estimator improves toward the tail of that specific target density.
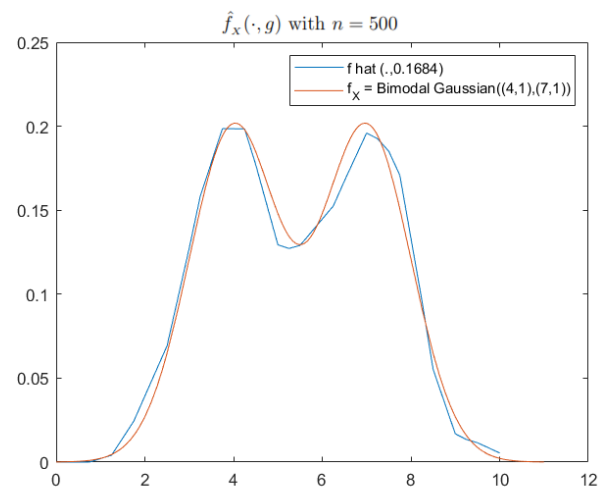

(a) n = 200 Gaussian
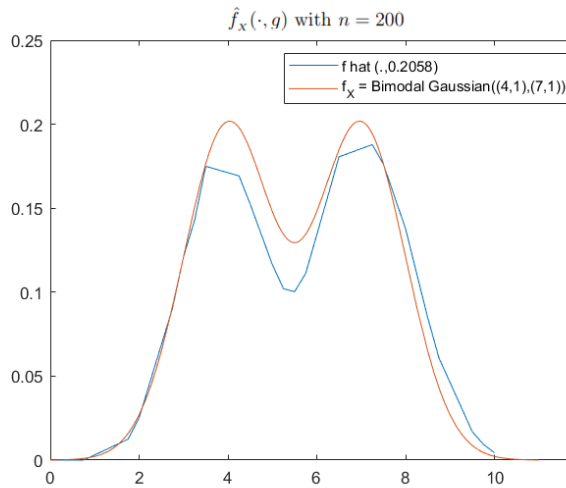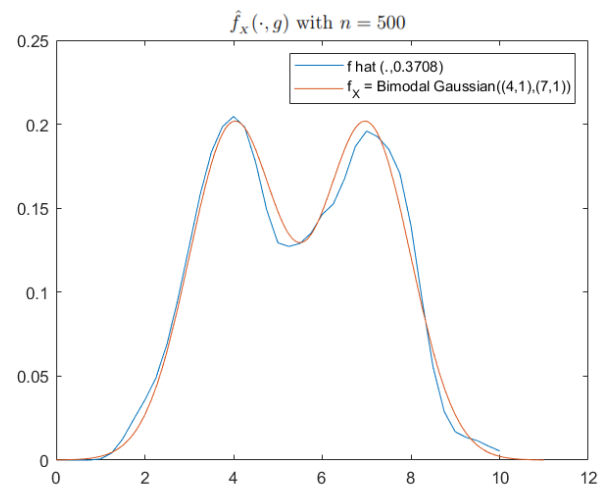
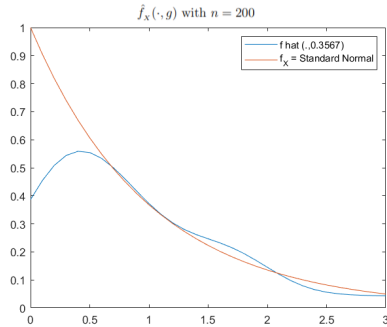(b) n = 500 Gaussian

(c) n = 200 Triweight

(d) n = 500 Triweight
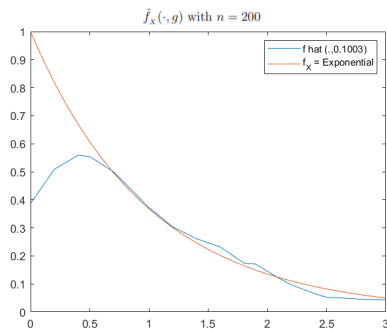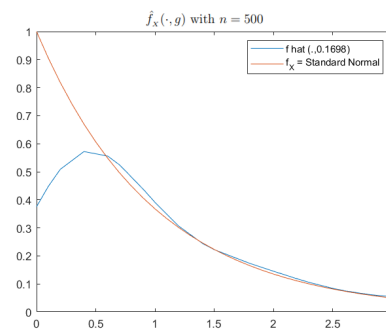
(e) n = 200 dlVP

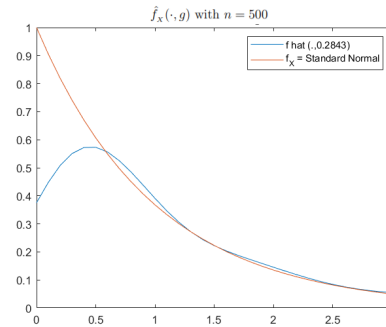(f) n = 500 dlVP

Figure 2.3: The pseudo densities $\hat{f}_X(\cdot, g)$ for the Exponential(1) distribution with $n = 200$ and $n = 500$

CHAPTER 3

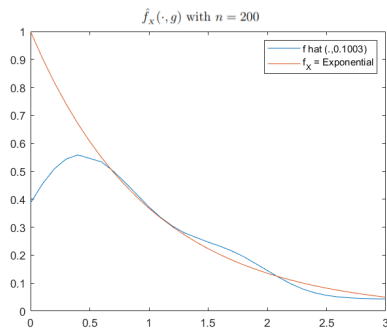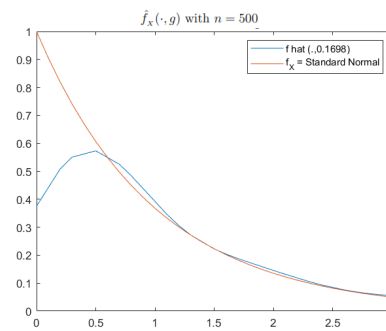The bootstrapping

In this chapter we first introduce the bootstrap procedure of Delaigle and Gijbels (2004) [3] used to approximate the optimal bandwidth h, and we illustrate the use of the bootstrap procedure we implement by using the pseudo density $\hat{f}_X(\cdot, g)$ constructed in Chapter 2.

3.1   An existing bootstrap method

The bootstrap method is an appealing method where one can approximate some relevant quantities from unknown densities. In the context of our study, the unknown quantity is the density itself. Before describing the bootstrap method of Delaigle and Gijbels (2004) [3], we present an analysis of $\mathrm{MISE}(\hat{f}_X(\cdot, h))$.

Recall that a common way to measure how close density estimator $\hat{f}_X(x, h)$ is to a population density $f_X(x)$ is to calculate $\mathrm{MISE}(\hat{f}_X(\cdot, h))$. We have

$$\mathrm{MISE}\left(\hat{f}_X(\cdot, h)\right) = \mathrm{E} \int \left(\hat{f}_X(x, h) - f_X(x)\right)^2 dx,$$

which can be written as

$$\mathrm{MISE}\left(\hat{f}_X(\cdot, h)\right) = \int \mathrm{Var}\left(\hat{f}_X(x, h)\right) dx + \int \left(\mathrm{Bias}\left(\hat{f}_X(x, h)\right)\right)^2 dx. \qquad (3.1)$$

We get

$$\mathrm{Var}\left(\hat{f}_X(x, h)\right) = \mathrm{Var}\left(\frac{1}{nh}\sum_{j=1}^{n} K^Z\left(\frac{x - Y_j}{h}\right)\right)$$

$$= \frac{1}{nh^2} E\left[K^Z\left(\frac{x - Y_1}{h}\right)^2\right] - \frac{1}{nh^2}\left[E\left(K^Z\left(\frac{x - Y_1}{h}\right)\right)\right]^2 dt. \qquad (3.2)$$

20

We first show that

$$\int \text{Var}\left(\hat{f}_X(\cdot, h)\right) dx = \frac{1}{2\pi h n} \int \frac{\varphi_K^2(t)}{|\varphi_{f_Z}(\frac{t}{h})|^2} dt - \frac{1}{2\pi n} \int \varphi_K^2(th)|\varphi_{f_X}(t)|^2 dt. \quad (3.3)$$

We now show that the second term on the right-hand side in (3.2) is given by

$$\frac{1}{nh^2} \int \left[E\left(K^Z\left(\frac{x-Y_1}{h}\right)\right)\right]^2 dx = \frac{1}{2\pi n} \int \varphi_K^2(ht)|\varphi_{f_X}(t)|^2 dt. \quad (3.4)$$

We will use the Plancherel theorem [16] which states that

$$\int |E(t)|^2 dt = \int |Ev|^2 dv,$$

where $E(t)$ and Ev are the Fourier transform pair. However, we first evaluate the left-hand side of (3.4). For notational simplicity, let us denote the left-hand side of (3.4) by Q1, i.e. we use

$$Q1 = \frac{1}{nh^2} \int \left[E\left(K^Z\left(\frac{x-Y_1}{h}\right)\right)\right]^2 dx.$$

We get

$$Q1 = \frac{1}{nh^2} \int \left[E\left(K^Z\left(\frac{x-(X+Z)}{h}\right)\right)\right]^2 dx$$

$$= \frac{1}{nh^2} \int \left(\int\int \left(K^Z\left(\frac{x-(r+s)}{h}\right)\right) f_{X,Z}(r,s) \, dr \, ds\right)^2 dx$$

$$= \frac{1}{nh^2} \int \left(\int\int \left(K^Z\left(\frac{x-(r+s)}{h}\right)\right) f_X(r) f_Z(s) \, dr \, ds\right)^2 dx$$

$$= \frac{1}{nh^2} \int \left(\int\int \frac{1}{2\pi}\int e^{-iv\left(\frac{x-(r+s)}{h}\right)} \frac{\varphi_K(v)}{\varphi_{f_Z}\left(\frac{v}{h}\right)} f_X(r) f_Z(s) \, dr \, ds \, dz\right)^2 dx$$

$$= \frac{1}{nh^2} \int \left(\int\int \frac{1}{2\pi}\int e^{-iv\left(\frac{x-(r+s)}{h}\right)} \frac{\varphi_K(v)}{\varphi_{f_Z}\left(\frac{v}{h}\right)} f_X(r) f_Z(s) \, dr \, ds \, dv\right)^2 dx$$

$$= \frac{1}{nh^2} \int \left(\int \frac{1}{2\pi} e^{iv\left(\frac{-x}{h}\right)} \frac{\varphi_K(v)}{\varphi_{f_Z}\left(\frac{v}{h}\right)} \int e^{iv\left(\frac{w}{h}\right)} f_X(r) \, dr \int e^{iv\left(\frac{z}{h}\right)} f_Z(s) \, ds \, dv\right)^2 dx$$

21

$$= \frac{1}{nh^2} \int \left(\int \frac{1}{2\pi} e^{iv\left(\frac{-x}{h}\right)} \frac{\varphi_K(v)}{\varphi_{f_Z}\left(\frac{v}{h}\right)} \varphi_{f_X}\left(\frac{v}{h}\right) \varphi_{f_Z}\left(\frac{v}{h}\right) dv \right)^2 dx$$

$$= \frac{1}{nh^2} \int \left(\int \frac{1}{2\pi} e^{iv\left(\frac{-x}{h}\right)} \varphi_K(v) \varphi_{f_X}\left(\frac{v}{h}\right) dv \right)^2 dx.$$

We will next apply the Plancherel theorem by letting $v = th$. We then get

$$t = \frac{v}{h}, \quad dt = \frac{dv}{h}, \quad m = \frac{x}{2\pi}, \quad dm = \frac{dx}{2\pi}.$$

We obtain

$$\begin{aligned}
\text{Q1} &= \frac{1}{nh^2} \int \left(\int \frac{1}{2\pi} e^{-2\pi m t i} \varphi_K(th) \varphi_{f_W}(t) h \, dt \right)^2 2\pi \, dm \\
&= \frac{2\pi h^2}{4\pi^2 nh^2} \int \left(\int e^{-2\pi m t i} \varphi_K(th) \varphi_{f_X}(t) h \, dt \right)^2 dm \\
&= \frac{1}{2\pi n} \int \left(\int \frac{1}{2\pi} e^{-2\pi m t i} \varphi_K(th) \varphi_{f_X}(t) \, dt \right)^2 dm \quad\quad (3.5) \\
&= \frac{1}{2\pi n} \int |\varphi_K(th) \varphi_{f_X}(t)|^2 \, dt \\
&= \frac{1}{2\pi n} \int \varphi_K^2(th) |\varphi_{f_X}(t)|^2 \, dt.
\end{aligned}$$

Let us now evaluate the first term on the right-hand side of (3.2) by denoting it with the symbol Q2, namely

$$\text{Q2} = \frac{1}{nh^2} \int E\left[K^Z \left(\frac{x - Y_1}{h} \right)^2 \right] dx.$$

Using the Plancherel theorem in a similar way with $t = \frac{x - Y_1}{h}$, we obtain

$$\text{Q2} = \frac{1}{2\pi hn} \int \frac{\varphi_K^2(t)}{|\varphi_{f_Z}\left(\frac{t}{h}\right)|^2} \, dt. \quad\quad (3.6)$$

As we see from (3.5) and (3.6), it follows that (3.3) holds.

22

Now, we evaluate the integral of the square of the bias, which is the second term in (3.1). Based on previous demonstrations of the first moment squared of the left hand-side in (3.4) we see that the integral of the bias square is given by

$$\int \left(\text{Bias}(\hat{f}_X(\cdot, h)) \right)^2 dt = \frac{1}{h^2} \int \left[E\left(K^Z \left(\frac{x - Y_1}{h} \right) \right) \right]^2 dx \tag{3.7}$$
$$- \frac{2}{h} \int E\left(K^Z \left(\frac{x - Y_1}{h} \right) \right) f_X(x)\, dx + R(f_X).$$

We will show that

$$\int \left(\text{Bias}(\hat{f}_X(\cdot, h)) \right)^2 dt = \frac{1}{2\pi} \int \varphi_K^2(th) |\varphi_{f_X}(t)|^2\, dt - \frac{1}{\pi} \int \varphi_K(th) |\varphi_{f_X}(t)|^2\, dt + R(f_X). \tag{3.8}$$

We let Q3 denote the first term of the right-hand side of (3.7), namely

$$\text{Q3} = \frac{1}{h^2} \int E\left[K^Z \left(\frac{x - Y_1}{h} \right)^2 \right] dx.$$

We then obtain

$$\text{Q3} = \frac{1}{h^2} \int \left[E\left(K^Z \left(\frac{x - (X + Z)}{h} \right) \right) \right]^2 dx$$
$$= \frac{1}{h^2} \int \left(\int \int \left(K^Z \left(\frac{x - (r + s)}{h} \right) \right) f_{X,Z}(r, s)\, dr\, ds \right)^2 dx$$
$$= \frac{1}{h^2} \int \left(\int \int \left(K^Z \left(\frac{x - (r + s)}{h} \right) \right) f_X(r)\, f_Z(s)\, dr\, ds \right)^2 dx$$
$$= \frac{1}{h^2} \int \left(\int \int \frac{1}{2\pi} \int e^{-iv\left(\frac{x - (r + s)}{h} \right)} \frac{\varphi_K(v)}{\varphi_{f_Z}\left(\frac{v}{h} \right)} f_X(r)\, f_Z(s)\, dr\, ds\, dz \right)^2 dx$$
$$= \frac{1}{h^2} \int \left(\int \int \frac{1}{2\pi} \int e^{-iv\left(\frac{x - (r + s)}{h} \right)} \frac{\varphi_K(v)}{\varphi_{f_Z}\left(\frac{v}{h} \right)} f_X(r)\, f_Z(s)\, dr\, ds\, dv \right)^2 dx$$
$$= \frac{1}{h^2} \int \left(\int \frac{1}{2\pi} e^{iv\left(\frac{-x}{h} \right)} \frac{\varphi_K(v)}{\varphi_{f_Z}\left(\frac{v}{h} \right)} \int e^{iv\left(\frac{w}{h} \right)} f_X(r)\, dr \int e^{iv\left(\frac{z}{h} \right)} f_Z(s)\, ds\, dv \right)^2 dx$$
$$= \frac{1}{h^2} \int \left(\int \frac{1}{2\pi} e^{iv\left(\frac{-x}{h} \right)} \frac{\varphi_K(v)}{\varphi_{f_Z}\left(\frac{v}{h} \right)} \varphi_{f_X}\left(\frac{v}{h} \right) \varphi_{f_Z}\left(\frac{v}{h} \right) dv \right)^2 dx$$
$$= \frac{1}{h^2} \int \left(\int \frac{1}{2\pi} e^{iv\left(\frac{-x}{h} \right)} \varphi_K(v) \varphi_{f_X}\left(\frac{v}{h} \right) dv \right)^2 dx.$$

23

We apply the Plancherel theorem by first letting $v = th$. As in the calculations for Q1, we get

$$t = \frac{v}{h}, \quad dt = \frac{dv}{h}, \quad m = \frac{x}{2\pi}, \quad dm = \frac{dx}{2\pi}.$$

We obtain

$$\text{Q3} = \frac{1}{h^2} \int \left(\int \frac{1}{2\pi} e^{-2\pi m t i} \varphi_K(th) \varphi_{f_W}(t) \, h \, dt \right)^2 2\pi \, dm$$

$$= \frac{2\pi h^2}{4\pi^2 h^2} \int \left(\int e^{-2\pi m t i} \varphi_K(th) \varphi_{f_X}(t) \, h \, dt \right)^2 dm$$

$$= \frac{1}{2\pi} \int \left(\int \frac{1}{2\pi} e^{-2\pi m t i} \varphi_K(th) \varphi_{f_X}(t) \, dt \right)^2 dm$$

$$= \frac{1}{2\pi} \int |\varphi_K(th) \varphi_{f_X}(t)|^2 \, dt$$

$$= \frac{1}{2\pi} \int \varphi_K^2(th) | \varphi_{f_X}(t)|^2 \, dt,$$

hence,

$$\text{Q3} = \frac{1}{2\pi} \int \varphi_K^2(th) | \varphi_{f_X}(t)|^2 \, dt. \tag{3.9}$$

Proceeding in a similar manner, we use Q4 to denote the second term of the right-hand side on (3.7), namely

$$\text{Q4} = \frac{2}{h} \int E \left(K^Z \left(\frac{x - Y_1}{h} \right) \right) f_X(x) \, dx.$$

We first show that

$$\text{Q4} = \frac{2}{h} \int E \left(K^Z \left(\frac{x - (X + Z)}{h} \right) \right) f_X(x) \, dx$$

$$= \frac{2}{h} \int \left(\int \int \left(K^Z \left(\frac{x - (r + s)}{h} \right) \right) f_{X,Z}(r, s) \, dr \, ds \right) f_X(x) \, dx$$

$$= \frac{2}{h} \int \left(\int \int \left(K^Z \left(\frac{x - (r + s)}{h} \right) \right) f_X(r) f_Z(s) \, dr \, ds \right) f_X(x) \, dx$$

$$= \frac{2}{h} \int \left(\int \int \frac{1}{2\pi} \int e^{-iv\left(\frac{x - (r+s)}{h} \right)} \frac{\varphi_K(v)}{\varphi_{f_Z}\left(\frac{v}{h} \right)} f_X(r) f_Z(s) \, dr \, ds \, dv \right) f_X(x) \, dx$$

24

$$= \frac{2}{h} \int \left(\int \frac{1}{2\pi} e^{iv\left(\frac{-x}{h}\right)} \frac{\varphi_K(v)}{\varphi_{f_Z}\left(\frac{v}{h}\right)} \int e^{iv\left(\frac{w}{h}\right)} f_X(r)\, dr \int e^{iv\left(\frac{z}{h}\right)} f_Z(s)\, ds\, dv \right) f_X(x)\, dx$$

$$= \frac{1}{\pi h} \int \int e^{iv\left(\frac{-x}{h}\right)} \frac{\varphi_K(v)}{\varphi_{f_Z}\left(\frac{v}{h}\right)} \varphi_{f_X}\left(\frac{v}{h}\right) \varphi_{f_Z}\left(\frac{v}{h}\right)\, dv\, f_X(x)\, dx$$

$$= \frac{1}{\pi h} \int \int e^{iv\left(\frac{-x}{h}\right)} \varphi_K(v)\, \varphi_{f_X}\left(\frac{v}{h}\right)\, dv\, f_X(x)\, dx$$

$$= \frac{1}{\pi h} \int \varphi_K(v)\, \varphi_{f_X}\left(\frac{v}{h}\right)\, dv \int e^{iv\left(\frac{-x}{h}\right)} f_X(x)\, dx$$

$$= \frac{1}{\pi h} \int \varphi_K(v)\, \varphi_{f_X}\left(\frac{v}{h}\right)\, \varphi_{f_X}\left(\frac{v}{h}\right)\, dv$$

$$= \frac{1}{\pi h} \int \varphi_K(v)\, |\varphi_{f_X}\left(\frac{v}{h}\right)|^2\, dv,$$

where we recall that

$$t = \frac{v}{h}, \qquad dt = \frac{dv}{h}.$$

Thus, we have

$$Q4 = \frac{1}{\pi} \int \varphi_K(th)\, |\varphi_{f_X}(t)|^2\, dt. \tag{3.10}$$

From (3.9), (3.10), and $R(f_X)$, we see that (3.8) holds.

Combining (3.3) and (3.8), we obtain

$$\mathrm{MISE}(\hat{f}(\cdot, h)) = Q2 - Q1 + Q3 - Q4 + R(f_X),$$

where the quantities Q1, Q2, Q3, and Q4 are given in (3.5), (3.6), (3.9), and (3.10), respectively. Therefore, we show

$$\mathrm{MISE}(\hat{f}(\cdot, h)) = \frac{1}{2\pi h n} \int \frac{\varphi_K^2(t)}{|\varphi_{f_Z}\left(\frac{t}{h}\right)|^2}\, dt - \frac{1}{2\pi n} \int \varphi_K^2(th)|\varphi_{f_X}(t)|^2\, dt + \frac{1}{2\pi} \int \varphi_K^2(th)|\varphi_{f_X}(t)|^2\, dt$$

$$- \frac{1}{\pi} \int \varphi_K(th)|\varphi_{f_X}(t)|^2\, dt + R(f_X)$$

$$= \frac{1}{2\pi h n} \int \frac{\varphi_K^2(t)}{|\varphi_{f_Z}\left(\frac{t}{h}\right)|^2}\, dt + \frac{1}{2\pi} \int \varphi_K^2(th)|\varphi_{f_X}(t)|^2\, dt - \frac{1}{2\pi n} \int \varphi_K^2(th)|\varphi_{f_X}(t)|^2\, dt$$

$$- \frac{1}{\pi} \int \varphi_K(th)|\varphi_{f_X}(t)|^2\, dt + R(f_X) \tag{3.11}$$

25

$$= \frac{1}{2\pi h n} \int \frac{\varphi_K^2(t)}{|\varphi_{f_Z}(\frac{t}{h})|^2} \, dt + \left(1 - \frac{1}{n}\right)\left(\frac{1}{2\pi}\right) \int \varphi_K^2(th)|\varphi_{f_X}(t)|^2 \, dt + R(f_X)$$

$$- \frac{1}{\pi} \int \varphi_K(th)|\varphi_{f_X}(t)|^2 \, dt.$$

Recall that in Chapter 2 we formed a pilot bandwidth g. Therefore, we were able to construct the pseudo density $\hat{f}_X(\cdot, g)$ from the observed random contaminated sample $Y_1, Y_2, ..., Y_n$. The reference Delaigle and Gijbels (2004) [3] proposes the following method. From the pseudo pilot density $\hat{f}_X(\cdot, g)$ we draw a bootstrap sample $X_1^*, X_2^*, ..., X_n^*$. Then, we add the error Z to the bootstrapped sample and obtain the contaminated bootstrapped sample $Y_1^*, Y_2^*, ..., Y_n^*$. We then use the contaminated bootstrapped sample to construct $\hat{f}_X^*(\cdot, h)$.

Our goal remains to obtain the optimal bandwidth h which is the minimizer of the quantity $\text{MISE}(\hat{f}_X(\cdot, h))$. While the pseudo density is constructed as an estimator of $f_X(x)$, the bootstrap method approximates $f_X(x)$ via the repetition of resampling with replacement to find an approximation of the optimal h value, see E. L. Lehmann (1995) [9].

The reference Delaigle and Gijbels (2004) [3] shows that the bootstrap approximation of (3.11) is given by

$$\text{MISE}_{(1)}^*(\hat{f}_X(\cdot, h)) = \frac{1}{2\pi h n} \int \frac{\varphi_K^2(t)}{|\varphi_{f_Z}(\frac{t}{h})|^2} \, dt + \left(1 - \frac{1}{n}\right)\left(\frac{1}{2\pi}\right) \int \varphi_K^2(th)|\hat{\varphi}_{X,g}(t)|^2 \, dt$$

$$+ R(\hat{f}_X^*(\cdot, g)) - \frac{1}{\pi} \int \varphi_K(th)|\hat{\varphi}_{X,g}(t)|^2 dt, \tag{3.12}$$

where $\hat{\varphi}_{X,g}(t)$ is the Fourier transform of $\hat{f}_X(\cdot, g)$. The calculation of the empirical characteristic function for $\hat{\varphi}_{X,g}(t)$ yields

$$\hat{\varphi}_{X,g}(t) = \hat{\varphi}_{Y,n}(t) \frac{\varphi_K(gt)}{\varphi_{f_Z}(t)}. \tag{3.13}$$

In (3.12) we want to estimate an optimal bandwidth g such that this optimal bandwidth g minimizes any term associated with it. We are not interested in obtaining

the optimal bandwidth g to minimize $f_X(x)$. If we want to minimize (3.12) depending on the bandwidth h, then this is equivalent to minimizing the expression

$$MISE^*_{(2)}(\hat{f}_X(\cdot, h)) = \frac{1}{2\pi h n} \int \frac{\varphi^2_K(t)}{|\varphi_{f_Z}(\frac{t}{h})|^2} dt + \left(1 - \frac{1}{n}\right) \left(\frac{1}{2\pi}\right) \int \varphi^2_K(th) |\hat{\varphi}_{X,g}(t)|^2 dt$$
$$- \frac{1}{\pi} \int \varphi_K(th) |\hat{\varphi}_{X,g}(t)|^2 dt, \qquad (3.14)$$

where the only quantity involving the pilot bandwidth g is the empirical characteristic function $\hat{\varphi}_{X,g}(t)$.

A simpler approach for estimating (3.11) is to use the asymptotic representation of $\mathrm{MISE}(\hat{f}_X(\cdot, h))$ introduced in Chapter 1 for which its bootstrap approximation is given by

$$\mathrm{AMISE}^*(\hat{f}_X(\cdot, h)) = \frac{h^4}{4} \mu_2^2(K) \, R(\hat{f}''_X(\cdot, g)) + \frac{1}{2\pi n h} \int \frac{|\varphi_K(t)|^2}{|\varphi_{f_Z}(\frac{t}{h})|^2} dt, \qquad (3.15)$$

where $R(\hat{f}''_X(\cdot, g))$ is an estimate of $R(f''_X)$. Note that the bootstrap method of Delaigle and Gijbels (2004) [3] involves the minimization of the $\mathrm{AMISE}(\hat{f}_X(\cdot, h))$ by defining a grid of possible values for the pilot bandwidth g candidates and by selecting the optimal pilot bandwidth g. Therefore, this bootstrap method relies only on the initial contaminated sample $Y_1, Y_2, ..., Y_n$. We remark that C. Léger and J. P Romano (1989) [8] and Marron (1992) [10] showed that if the observed sample is not contaminated, it is not necessary to obtain a bootstrap sample, contrary to other bootstrap procedures. The reference Delaigle and Gijbels (2004) [3] explains that the same is true for contaminated data.

Although we use a bootstrapping method involving resampling with replacement, our bootstrap method involves the bootstrap approximation of $\mathrm{MISE}(\hat{f}_X(\cdot, h))$ rather than its asymptotic representation. We obtain our bootstrap samples from the pseudo density $\hat{f}_X(\cdot, g)$ as our initial bootstrap sample. The reference Delaigle and Gijbels

(2004) [3] has shown that the optimal bandwidth h associated with the bootstrapped procedure in $\text{MISE}^*(\hat{f}_X^*(\cdot, h))$ is a consistent estimator of the $\text{MISE}(\hat{f}_X(\cdot, h))$.

3.2 The consistency of the bootstrap estimator

In this section, we evaluate the consistency of the bootstrap approximation. In order to observe the consistency of the bootstrapped representation of $\text{MISE}(\hat{f}_X(\cdot, h))$ we need to use the previous assumptions and also include a new set of assumptions stated below.

Assumption 3: We further assume that $f_X^{(j)}(x)$ satisfies

(i) $\sup_{x \in \mathbb{R}} |f_X^{(j)}(x)| < \infty, \qquad 1 \leq j \leq 4.$

(ii) $\int |f_X''(x)| \, dx < \infty$ and $\int |f_X^{(3)}(x)| \, dx < \infty.$

3.2.1 The consistency

To show the consistency we consider the case with $n \to \infty$, $h \to 0$, and where K is a second order kernel. The references Wand and Jones (1995) [15] and Stefanski and Carroll (1990) [13] describe the integrated bias squared as

$$\int \left(\text{Bias}(\hat{f}_X(x, h)) \right)^2 dx = \frac{h^4}{4} \mu_2^2(K) \, R(f_X'') + o(h^4), \qquad (3.16)$$

with a finite second moment and they show that

$$\int \text{Var}(\hat{f}_X(t, h)) \, dt = \frac{1}{2\pi n h} \int \frac{|\varphi_K(t)|^2}{|\varphi_{f_Z}\left(\frac{t}{h}\right)|^2} \, dt + O\left(\frac{1}{n}\right). \qquad (3.17)$$

Combining (3.16) and (3.17) we have

$$\text{MISE}(\hat{f}_X(\cdot, h)) = \frac{h^4}{4} \mu_2^2(K) \, R(f_X'') + \frac{1}{2\pi n h} \int \frac{|\varphi_K(t)|^2}{|\varphi_{f_Z}\left(\frac{t}{h}\right)|^2} \, dt + O\left(\frac{1}{n}\right) + o(h^4).$$

Analyzing the bootstrap version of the MISE via the asymptotic order of the bias of

28

the bootstrap integrated variance and the asymptotic order of the variance of the bootstrap integrated variance, we obtain

$$\text{Bias}\left[\int \text{Var}^* \hat{f}_X^*(t,h)\,dt\right] = O\left(\frac{1}{n}\right),$$

$$\text{Var}\left[\int \text{Var}^* \hat{f}_X^*(t,h)\,dt\right] = o\left(\frac{1}{n^2}\right).$$

The variance and the bias for the bootstrap integrated square bias are given by

$$\text{Bias}\int \left(\text{Bias}^* \hat{f}_X^*(x,h)\right)^2 dx = o(h^4),$$

$$\text{Var}\int \left(\text{Bias}^* \hat{f}_X^*(x,h)\right)^2 dx = o(h^8).$$

Compatible with our assumptions, we will now use the following theorem presented by Delaigle and Gijbels (2004) [3]:

Theorem [3] As $n \to \infty$ and $h \to 0$ we have

$$\text{E}\left(\text{MISE}^*(\hat{f}_X^*(\cdot,h))\right) = \text{MISE}(\hat{f}_X(\cdot,h)) + o(h^4) + O\left(\frac{1}{n}\right)$$

$$= \text{MISE}(\hat{f}_X(\cdot,h)) + o\left(\text{MISE}(\hat{f}_X(\cdot,h))\right),$$

and

$$\text{Var}\left[\text{MISE}^*(\hat{f}_X^*(\cdot,h))\right] \leq \text{Var}\left[\int \left(\text{Bias}^* \hat{f}_X^*(x,h)\right)^2 dx\right]$$

$$+ \text{Var}\left[\int \text{Var}^* \hat{f}_X^*(t,h)\,dt\right]$$

$$+ 2\sqrt{\text{Var}\left[\int \left(\text{Bias}^* \hat{f}_X^*(x,h)\right)^2 dx\right] + \text{Var}\left[\int \text{Var}^* \hat{f}_X^*(t,h)\,dt\right]},$$

$$\text{Var}\left(\text{MISE}^*(\hat{f}_X^*(\cdot,h))\right) = \text{MISE}(\hat{f}_X(\cdot,h)) + o(h^4) + O\left(\frac{1}{n}\right).$$

Therefore, $\text{MISE}^*(\hat{f}_X^*(\cdot,h)) \xrightarrow{L^2} \text{MISE}(\hat{f}_X(\cdot,h))$.

The asymptotical equivalence between $\mathrm{MISE}(\hat{f}_X(\cdot, h))$ and $\mathrm{MISE}^*(\hat{f}_X^*(\cdot, h))$ under our assumptions holds. The same reasoning is also valid for our bootstrap method, as we apply a bootstrapping method on $\mathrm{MISE}(\hat{f}_X(\cdot, h))$ rather than $\mathrm{AMISE}(\hat{f}_X(\cdot, h))$. We will discuss this point in the next chapter.

CHAPTER 4

The proposed bootstrap method

In this chapter we will describe our proposed bootstrap method which involves the minimization of $\mathrm{MISE}(\hat{f}_X(\cdot, h))$ using resampling with replacement. The foundation of our bootstrap rests on the selection of the pilot bandwidth g and the construction of pseudo density $\hat{f}_X(\cdot, g)$ as presented in Chapter 2. We also sample $X_1^*, X_2^*, ..., X_n^*$ from the pseudo density $\hat{f}_X(\cdot, g)$ as in 'Existing Method' in Chapter 3 before adding the error from the Laplace distribution to obtain the initial bootstrap sample $Y_1^*, Y_2^*, ..., Y_n^*$. We use the contaminated bootstrapped sample to construct the bootstrapped density $\hat{f}_X^*(\cdot, h)$.

We want to approximate the exact $\mathrm{MISE}(\hat{f}_X(\cdot, h))$ defined by

$$\mathrm{MISE}(\hat{f}_X(\cdot, h)) = \int \mathrm{Var}(\hat{f}_X(x, h)) \, dx + \int \left(\mathrm{Bias}(\hat{f}_X(x, h)) \right)^2 dx,$$

with its bootstrap approximation $\mathrm{MISE}^*(\hat{f}_X^*(\cdot, h))$ given by

$$\mathrm{MISE}^*(\hat{f}_X^*(\cdot, h)) = \int \mathrm{Var}^*(\hat{f}_X^*(x, h)) \, dx + \int \left(\mathrm{Bias}^*(\hat{f}_X^*(x, h)) \right)^2 dx$$

$$\approx \sum \mathrm{MSE}^*(\hat{f}_X^*(x, h)) \, \Delta(x),$$

and hence

$$\mathrm{MSE}^*(\hat{f}_X^*(x, h)) \approx \frac{1}{B-1} \sum_{i=1}^{B} \left(\bar{\hat{f}}_X^*(x, h) - \hat{f}_X^{*i}(x, h) \right)^2 + \left(\bar{\hat{f}}_X^*(x, h) - \hat{f}_X(x, g) \right)^2,$$

$$(4.1)$$

where

$$\bar{\hat{f}}_X^*(x, h) = \frac{1}{B} \sum_{i=1}^{B} \hat{f}_X^{*i}(x, h),$$

with $\bar{\hat{f}}_X^*(x, h)$ being calculated from the ith resampling with replacement for $i = 1, 2, ..., B$. For (4.1) see Sun, Sun, and Diao (2007) [14]. We select the optimal bandwidth h associated with the bootstrap procedure as the one which minimizes $\text{MSE}^*(\hat{f}_X^*(x, h))$.

4.1 The kernel selection

As previously discussed in Chapter 1 and Chapter 2, we use three different kernel density functions for $K(x)$. In other words, we have selected the de la Vallée-Poussin kernel, the Gaussian kernel, and the triweight kernel. For the de la Vallée-Poussin kernel and the triweight kernel the argument is never 0, i.e. we never have $x = Y_j$. Each bootstrapped sample $Y_1^*, Y_2^*, ..., Y_n^*$ becomes an input for $\hat{f}_X^*(x, h)$ constructed with each of the aforementioned three kernels. Moreover, the selection for the sequence of the bandwidths h associated with the bootstrap procedure consists of 20 values for which each one is used to construct $\hat{f}_X^*(x, h)$. Recall that the selection of the Laplace error distribution is also used to construct $\hat{f}_X^*(x, h)$. In order to construct $\hat{f}_X^*(x, h)$ the only remaining parameter needed is the number of resampling with replacement within the bootstrap.

4.2 The B Selection

In this section we discuss the importance of the number of resampling with replacement within the bootstrap. We use B to denote that relevant number. Since the optimal bandwidth h is unknown and we want to estimate it via the selection of bandwidth-values around the pilot bandwidth g. We must determine an optimal value for B such that B allows sufficient resampling to approximate the optimal bandwidth h. The reference B. W. Silverman (1998) [12] argues that it is reasonable to

simulate 1000 samples to find the optimal bandwidth h associated with our bootstrap method such that it estimates the optimal bandwidth h accurately. It is customary to regard $B = 1000$ as a reasonable lower bound, as the bootstrapping process is a time consuming technique which can also be numerically expensive. This drawback is balanced by the advantages offered by the bootstrapping method, which under certain conditions offers a consistent estimate when $B \geqslant 1000$. This value for B is commonly accepted to ensure a meaningful statistics. Whenever samples are drawn from the original sample with replacement, almost every sample contains repeated values. Of course, if n is large enough then most samples will contain repeated values several times. Therefore, we selected $B = 1000$ for each kernel we used. The next criteria we consider is the selection of the bandwidths h associated with the bootstrap procedure.

4.3 The selection of the bandwidths h associated with the bootstrap procedure

In this section, we discuss how to select the range of candidates for the optimal bandwidth h. The range of values for the bandwidths h used during the bootstrapping process must contain candidates which approximate the optimal bandwidth h. The optimal bandwidth h associated with the bootstrap process is the one which minimizes $\mathrm{MISE}^*(\hat{f}_X^*(\cdot, h))$. We select them around the pilot bandwidth g obtained in Chapter 2 by using increments of ± 0.01. This increment level produces values which do not violate the fact that a bandwidth h must be a positive value. This yields a total of 20 bandwidths and it provides a sufficiently fine grid of candidates. One of those h candidates approximates the optimal bandwidth h. We are, in essence, selecting the approximate of the optimal bandwidth h from a sample of bandwidths which contains a set of good performing bandwidths. As our sample size increases, our results show

that the range of the selected bandwidths h decreases. Selecting the range of values for these bandwidths is the last step to follow prior to initiating the bootstrapping.

4.4 The bootstrapping results

The bootstrapping results in Table 4.1 show that regardless of the choice of our target densities and regardless of the choice of our kernel densities, as the sample size increases, both $\text{MISE}^*(\hat{f}_X^*(\cdot, h))$ and the optimal selected bandwidth h decrease. Note the $\text{MISE}^*(\hat{f}_X^*(\cdot, h))$ results are to be considered $\times 10^{-3}$.

Table 4.1: The optimal selected bandwidth h and the corresponding $\text{MISE}^* \times 10^{-3}$

Target Densities

Kernel	Standard Normal			Bimodal Gaussian((4,1),(7,1))		
	n = 200	n=400	n=500	n = 200	n=400	n=500
dlVP	0.1678	0.1475	0.1309	0.1958	0.1763	0.1562
	(12.91)	(9.59)	(8.27)	(17.57)	(14.47)	(12.17)
Gaussian	0.4060	0.3257	0.3035	0.4339	0.4022	0.3508
	(41.42)	(35.44)	(33.17)	(45.05)	(37.49)	(31.22)
Triweight	0.1608	0.1270	0.1130	0.1825	0.1763	0.1584
	(4.82)	(3.47)	(2.98)	(4.97)	(4.02)	(3.86)

The following plots show the visual representation of the estimator using the optimal selected bandwidth h when the sample size is $n = 500$:
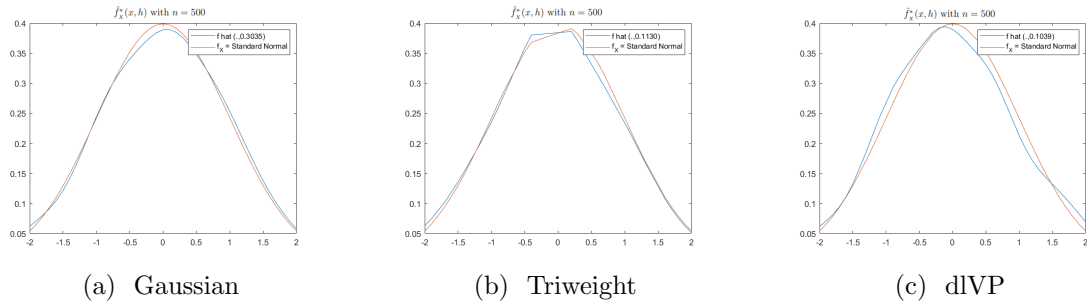
34

(a) Gaussian   (b) Triweight   (c) dlVP

Figure 4.1: The bootstrapped density $\hat{f}_X^*(x,h)$ estimating the target density $N(0,1)$ with $n = 500$

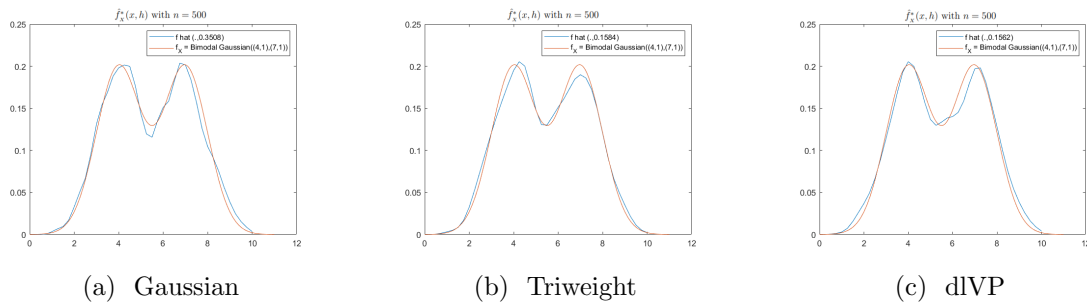(a) Gaussian   (b) Triweight   (c) dlVP

Figure 4.2: The bootstrapped density $\hat{f}_X^*(x,h)$ estimating the target density bimodal Gaussian$((4,1)(7,1))$ with $n = 500$

The results show that regardless of the kernel selection, the optimal selected bandwidth h has improved the estimation of the target densities. This is attributed in part to the sample size. Nevertheless, the role of such an optimal bandwidth h leads to a performing estimator as shown throughout Figure 4.1 and Figure 4.2.
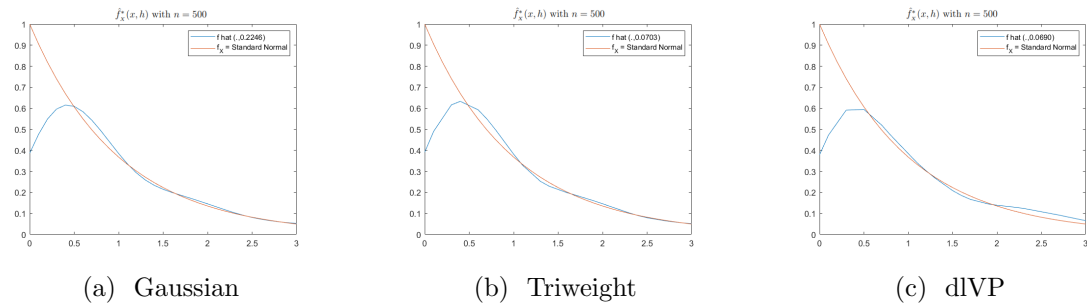

(a) Gaussian   (b) Triweight   (c) dlVP

Figure 4.3: The bootstrapped density $\hat{f}_X^*(x,h)$ estimating the target density Exponential(1) with $n = 500$

As in the cases of the $N(0, 1)$ target density and the bimodal Gaussian$((4,1),(7,1))$ target density, the kernel selection remains inconsequential to the performance of the estimator. When the target density is Exponential(1), Figure 4.3 displays how the lack of symmetry of this specific target density directly impacts the poor quality of estimation.

REFERENCES

[1] Chi-Yang Chu, D. J. Henderson, and C. Parmeter, *Plug-in bandwidth selection for kernel density estimation with discrete data*, Econometrics **3**, 199–214 (2015).

[2] A. Delaigle, Code for computing the deconvolution kernel density estimator and the pilot bandwidth,
`https://researchers.ms.unimelb.edu.au/aurored@unimelb//links.html#Code`.

[3] A. Delaigle and I. Gijbels, *Bootstrap bandwidth selection in kernel density estimation from a contaminated sample*, Ann. Inst. Stat. Math. **56**, 19–47 (2004).

[4] A. Delaigle and I. Gijbels, *Estimation of integrated squared density derivatives from a contaminated sample kernel density estimation*, Journal of the Royal Statistical Society **64**, 869–886 (2002).

[5] A. Delaigle and I. Gijbels, *Practical bandwidth selection in deconvolution kernel density estimation*, Computational Statistics Data Analysis **45**, 249–267 (2004).

[6] Jianqing Fan, *Asymptotic normality for deconvolving kernel density estimator*, Sankhyā: The Indian Journal of Statistics **53**, 97–110 (1991).

[7] Jianqing Fan and Young Truong, *Nonparametric regression with errors in variables*, Ann. Statist. **21**, 1900–1925 (1993).

[8] C. Léger and J. P Romano, *Bootstrap choice of the tuning parameter*, Ann. Instit. Stat. Math. **42**, 709–735 (1989).

[9] E. L. Lehmann, *Element of large-sample theory*, Springer-Verlag, New York, 1999.

[10] J. S. Marron, *Visual understanding of higher-order kernels*, Journal of Computational and Graphical Statistics **3**, 447–458 (1994).

[11] M. Rosenblatt, *Remark on some nonparametric estimates of a density function*, Ann. Math. Statist. **27**, 832–837 (1956).

[12] B. W. Silverman, *Density estimation for statistics and data analysis*, Chapman and Hall, London, 1994.

[13] L. A. Stefanski and R. J. Carroll, *Deconvolving kernel density estimators*, Statistics **21**, 169–184 (1990).

[14] Yanqing Sun, Shan Sun, and Yuanan Diao, *Smooth quantile processes from right censored data and construction of simultaneous confidence bands*, Communications in Statistics - Theory and Methods **30**, 707–727 (2007).

[15] M. P. Wand and M. C. Jones, *Kernel smoothing*, Springer, New York, 1995.

[16] `https://mathworld.wolfram.com/PlancherelsTheorem.html`