I N S T I T U T D E
S T A T I S T I Q U E

UNIVERSITÉ CATHOLIQUE DE LOUVAIN

D I S C U S S I O N
P A P E R

0116

BOOTSTRAP BANDWIDTH SELECTION IN KERNEL DENSITY ESTIMATION FROM A CONTAMINATED SAMPLE

A. DELAIGLE and I. GIJBELS

http://www.stat.ucl.ac.be

Bootstrap bandwidth selection in kernel density estimation from a contaminated sample*

A. Delaigle and I. Gijbels

Institut de Statistique, Université catholique de Louvain, 20 Voie du Roman Pays, Louvain-la-Neuve, Belgium.

April 2001

Abstract

In this paper we consider kernel estimation of a density when the data are contaminated by random noise. More specifically we deal with the problem of how to choose the bandwidth parameter in practice. A theoretical optimal bandwidth is defined as the minimizer of the mean integrated squared error. We propose a bootstrap procedure to estimate this optimal bandwidth, and show its consistency. These results remain valid for the case of no measurement error, and hence also summarize part of the theory of bootstrap bandwidth selection in ordinary kernel density estimation. The finite sample performance of the proposed bootstrap selection procedure is demonstrated with a simulation study. Some applications to real data examples illustrate the use of the method.

Key words and phrases: bandwidth selection, bootstrap, consistency, deconvolution, errors-in-variables, kernel density estimation.

1 Introduction

In this paper we consider the problem of estimating a density from a sample of size n that has been contaminated by random noise. This problem is usually referred to as a *deconvolution problem*, and has applications in many different fields such as biostatistics, chemistry and public health. See for example Stefanski and Carroll (1990) or Carroll, Ruppert and Stefanski (1995) who deal with analyzing such data. Here we consider the so-called *deconvolving kernel estimator* introduced by Carroll and Hall (1988) and investigated by Stefanski and Carroll (1990), among others.

* This research was supported by 'Projet d'Actions de Recherche Concertées' (No. 98/03–217) from the Belgian government.

This method of estimation has already received considerable attention in the literature. Recent related works include Neumann (1997), Wand (1998), Hesse (1999) and Rachdi and Sabre (2000). See also Stefanski and Carroll (1990), Fan (1991a,b,c), Fan (1992) and Wand and Jones (1995), among others, for earlier contributions. Most of those papers however deal only with theoretical aspects of the estimation, and very few focus attention on the yet important issue of choosing the bandwidth in practice. Stefanski and Carroll (1990) and Hesse (1999) investigate a cross-validation choice of the bandwidth. However in practice this method suffers from the same drawbacks as those known in the non-contaminated case, such as large variability or multiplicity of the solution. See Delaigle (1999) or Delaigle and Gijbels (2001b). The latter paper proposes plug-in type of bandwidth selectors, based on asymptotic approximations of the mean integrated squared error. These plug-in procedures already improve upon the cross-validation technique. In the present paper we propose a bandwidth selection procedure based on bootstrap techniques, which can be considered as a competitor of the asymptotic techniques. We establish the asymptotic theory which justifies the use of this bootstrap bandwidth selector.

A detailed comparison, via simulation studies, of the three practical bandwidth selection procedures, the plug-in procedure, the bootstrap bandwidth selection and the cross-validation bandwidth, will be provided in a forthcoming manuscript. See Delaigle and Gijbels (2001b). This finite sample comparison revealed that the plug-in method and the bootstrap bandwidth selection method outperform the cross-validation method, and that they are competitive procedures in the sense that none of the two can be claimed to be best in all cases. A very interesting point for the bootstrap procedure is that it does not require the generation of bootstrap samples. Due to its special structure, the bootstrap estimate of the mean integrated squared error can be computed directly from the original sample. In fact, it simplifies even a lot further since the only quantity that needs to be computed from the original sample is an empirical characteristic function.

Bootstrap bandwidth selection in case of non-contaminated data has already been studied in several papers, including Faraway and Jhun (1990), Marron (1992), Grund and Polzehl (1997) and Hazelton (1999). In the present paper we consider the case of data that are contaminated by random noise, and develop a bootstrap procedure to choose the bandwidth parameter. This bootstrap-based bandwidth selection method requires the choice of a pilot bandwidth. We will see that a good choice for this pilot bandwidth is a bandwidth which is optimal for estimating the integrated squared second derivative of the target density. Such a bandwidth has been proposed by Delaigle and Gijbels (2001a), and is easy to use in practice.

We prove the consistency of the proposed bootstrap bandwidth selector, defined as the minimizer of the bootstrap estimator of the mean integrated squared error. We first prove that the bootstrap estimator of the mean integrated squared error of the density estimator converges (in L^2) to the exact mean integrated squared error. We then show that the ratio of the bootstrap bandwidth selector and the minimizer of the mean integrated squared error converges to one in probability. The finite sample performance of the bandwidth selection procedure is illustrated via a simulation study from which we can conclude that the method performs quite well.

The paper is organized as follows. In Section 2 we recall the definition of the deconvolving kernel estimator and some of its theoretical properties. We also briefly discuss the choice of the kernel function. In Section 3 we define the bootstrap estimator of the bandwidth, and discuss how to choose the pilot bandwidth. In Section 4 we establish theoretical properties of the bootstrap bandwidth selector. Finally in Section 5 we illustrate our method on a few simulated examples, and also apply it to some real data. All proofs are deferred to the Appendix.

2 Deconvolving kernel estimator

Let $Y_1, \ldots, Y_n$ be an i.i.d. sample of size n from a random variable Y of unknown density f_Y, verifying $Y = X + Z$, where X is a random variable distributed according to f_X, and Z is a random variable representing the measurement error, distributed according to f_Z. Suppose that X is independent of Z and f_X and f_Z are continuous. We assume here that the distribution of the error Z is fully known, which is the usual assumption in this context. This assumption might seem to be quite restrictive, but reflects the reality that one often has not sufficient information to estimate the distribution of Z and hence one needs to assume full knowledge of Z. However, in some applications it might happen that one has additional sampling information which permits to either estimate the variance of the error term, or even the entire distribution. An estimation of the variance for example can be carried out, under certain assumptions, in case of repeated measurements. See the application section. Occasionally one also might have to deal with the comfortable situation that an i.i.d sample $Z_1, \cdots, Z_n$ from Z has been observed separately. See for example Barry and Diggle (1995) and Neumann (1997).

The *deconvolving kernel estimator* of f_X at $x \in \mathbb{R}$ is defined by

$$\widehat{f}_X(x;h) \;=\; \frac{1}{nh}\sum_{j=1}^{n} K^Z\left(\frac{x-Y_j}{h};h\right), \tag{2.1}$$

where

$$K^Z(u;h) = (2\pi)^{-1}\int e^{-itu}\varphi_K(t)/\varphi_Z(t/h)\,dt\;, \tag{2.2}$$

with K a kernel function, supported on $\mathbb{R}$, real-valued and such that $\int K(x)\,dx = 1$, and where $h > 0$ is a smoothing parameter depending on n, called the bandwidth. Further, we denote the Fourier transform of K (or its characteristic function if K is a density) by $\varphi_K(t) = \int e^{itx}K(x)dx$, and denote by φ_Z the characteristic function of Z. Throughout this paper we suppose that for all $t \in I\!R$, $\varphi_Z(t) \neq 0$. See for example Carroll and Hall (1988) and Stefanski and Carroll (1990). See also Delaigle (1997, 1999) for computational details on how to calculate the estimator.

Furthermore we make the assumption that $K^Z(\cdot;h)$ is supported on $I\!R$, is real-valued, and $|K_Z(\cdot;h)|$ is integrable.

From studies about the deconvolving kernel density estimator it is already known that the kernel should rather be chosen among densities whose characteristic function has a compact and symmetric support. See for example Stefanski (1990), Stefanski and Carroll (1990), Fan (1992), Wand and Jones (1995), Wand (1998), Delaigle (1999) and Hesse (1999). Although this assumption can be relaxed in the case of ordinary smooth error densities (see Definition 1 below), we will in this paper work exclusively with compactly supported φ_K. The use of such kernels guarantees the existence of the density estimator in (2.1). An example of such a kernel is given by

$$K(x) = \frac{48\cos x}{\pi x^4}\left(1-\frac{15}{x^2}\right) - \frac{144\sin x}{\pi x^5}\left(2-\frac{5}{x^2}\right), \tag{2.3}$$

with $\varphi_K(t) = (1-t^2)^3 1_{[-1,1]}(t)$.

A common way to measure the closeness of the density estimator $\widehat{f}_X(\cdot;h)$ to its target $f_X(\cdot)$ is to compute the Mean Integrated Squared Error (MISE) of $\widehat{f}_X(\cdot;h)$, defined by

$$\begin{aligned}\text{MISE}\{\widehat{f}_X(\cdot;h)\} &= \text{E}\int \{\widehat{f}_X(x;h) - f_X(x)\}^2\,dx \\ &= \int \left[\text{Bias}\{\widehat{f}_X(x;h)\}\right]^2 dx + \int \text{Var}\{\widehat{f}_X(x;h)\}\,dx,\end{aligned}$$

which, after some algebra, can be written as

$$\begin{aligned}\text{MISE}\{\widehat{f}_X(\cdot;h)\} &= (2\pi nh)^{-1}\int |\varphi_K(t)|^2.|\varphi_Z(t/h)|^{-2}\,dt + \left(1-n^{-1}\right)R(K_h * f_X) + R(f_X) \\ &\quad -2\int (K_h * f_X)(x) f_X(x)\,dx, \end{aligned} \tag{2.4}$$

where we introduced the notations $K_h(x) = h^{-1}K(x/h)$, for any function K, $R(g) = \int g^2(x)\,dx$, for any square integrable function g, and where $f * g(\cdot) = \int f(\cdot - y)g(y)dy$ denotes the convolution of the functions f and g. See for example Wand and Jones (1995). The optimal bandwidth h_n for the estimation of f_X is then the bandwidth which minimizes $\text{MISE}\{\widehat{f}_X(\cdot;h)\}$. For simplicity of notation and to highlight the dependency in h, we will rather write $\text{MISE}\{\widehat{f}_X(\cdot;h)\}$ as $\text{MISE}(h)$ in what follows.

The MISE can also be written in the following way

$$\text{MISE}(h) = (2\pi nh)^{-1}\int |\varphi_K(t)|^2|\varphi_Z(t/h)|^{-2}\,dt + (1-n^{-1})(2\pi)^{-1}\int |\varphi_X(t)|^2|\varphi_K(ht)|^2\,dt$$

$$+ R(f_X) - \pi^{-1} \int |\varphi_X(t)|^2 \varphi_K(ht)\, dt. \tag{2.5}$$

See Wand (1998). This expression is easier to manipulate in practice (see for example Delaigle (1999)).

The asymptotic properties of the deconvolving kernel estimator have been studied in several papers, among which Stefanski and Carroll (1990), Fan (1991a,b,c) and Fan (1992). These properties depend strongly on the error distribution. As in Fan (1991a) we classify the errors in two categories, the ordinary smooth distributions and the supersmooth distributions.

Definition 1. *The distribution of a random variable Z is said to be*

(i) supersmooth of order β if its characteristic function $\varphi_Z(t)$ satisfies:

$$d_0|t|^{\beta_0} \exp(-|t|^{\beta}/\gamma) \leq |\varphi_Z(t)| \leq d_1|t|^{\beta_1} \exp(-|t|^{\beta}/\gamma) \quad \text{as } t \to \infty,$$

for some positive constants d_0, d_1, β, γ and constants β_0 and β_1;

(ii) ordinary smooth of order β if its characteristic function $\varphi_Z(t)$ satisfies:

$$d_0|t|^{-\beta} \leq |\varphi_Z(t)| \leq d_1|t|^{-\beta} \quad \text{as } t \to \infty,$$

for some positive constants d_0, d_1 and β.

It has been proved in Fan (1991c) that for supersmooth error densities (e.g. normal and Cauchy densities) the optimal rate of convergence of the density estimator to f_X is logarithmic, and hence very slow. By contrast this rate is much faster (algebraic) for ordinary smooth error densities (e.g. gamma and Laplace densities). Recall that this rate is $n^{-2/5}$ in the error free case. This difference in convergence rate between the two classes of error densities also shows up in the simulation results in Section 5.

3 Bootstrap selection of the bandwidth

3.1 Bootstrap estimator of the mean integrated squared error

The optimal bandwidth h_n defined as the minimizer of MISE(h) cannot be found in practice since the MISE involves unknown f_X related quantities. In this section we define an estimator of the optimal bandwidth, based on smoothed bootstrap techniques.

Bootstrap procedures for selection of the bandwidth in kernel density estimation from non-contaminated data have been studied in previous papers including Taylor (1989), Faraway and Jhun (1990), Jones, Marron and Park (1991), Falk (1992), Hall, Marron and Park (1992), Marron (1992), Grund and Polzehl (1997) and Hazelton (1999).

Let $\widehat{f}_X(\cdot;g)$ be the deconvolving kernel estimator of f_X obtained from $Y_1,\dots,Y_n$, with kernel L and pilot bandwidth g. Draw a bootstrap sample $X_1^*,\dots,X_n^*$ from $\widehat{f}_X(\cdot;g)$, and after having added noise Z, use the contaminated bootstrap sample to construct a deconvolving estimator $\widehat{f}_X^*(\cdot;h)$ of $\widehat{f}_X(\cdot;g)$, with kernel K and bandwidth h. Then the bootstrap estimator of the MISE is given by

$$\begin{aligned}\mathrm{MISE}^*(h) &= \mathrm{E}^*\int\{\widehat{f}_X^*(x;h)-\widehat{f}_X(x;g)\}^2\,dx\\ &= \int\big[\mathrm{Bias}^*\{\widehat{f}_X^*(x;h)\}\big]^2\,dx+\int \mathrm{Var}^*\{\widehat{f}_X^*(x;h)\}\,dx,\end{aligned}$$

where $\mathrm{E}^*, \mathrm{Bias}^*$ and Var^* all involve expectations conditionally upon $Y_1,\cdots,Y_n$, and are taken with respect to the pseudo density $\widehat{f}_X(\cdot;g)$, and the estimator of the optimal bandwidth is defined as the minimizer of $\mathrm{MISE}^*(h)$. From (2.4), it is immediate to see that the bootstrap MISE may also be written as

$$\begin{aligned}\mathrm{MISE}^*(h) &= (2\pi nh)^{-1}\int|\varphi_K(t)|^2.|\varphi_Z(t/h)|^{-2}\,dt+\big(1-n^{-1}\big)\,R\big(K_h*\widehat{f}_X(\cdot;g)\big)\\ &\quad +R(\widehat{f}_X(\cdot;g))-2\int\big(K_h*\widehat{f}_X(\cdot;g)\big)(x)\widehat{f}_X(x;g)\,dx. \qquad (3.1)\end{aligned}$$

Note that (3.1) is nothing but an approximation of the exact MISE, where the unknown f_X has been replaced by a deconvolving kernel estimator $\widehat{f}_X(\cdot;g)$. Although it may then seem at first that g should be set equal to h, the bandwidth needed to estimate f_X, a closer look at expression (3.1) reveals that g need not be optimal for the estimation of f_X, but rather for the estimation of quantities involving f_X. See Section 3.2. Similarly, the kernel L need not be equal to K, but this choice is less important and will not be discussed here.

In the case of non-contaminated data Taylor (1989) and Marron (1992), among others, remarked that this bootstrap bandwidth selection procedure, unlike many other bootstrap estimation procedures, does not require the generation of any bootstrap sample in practice. This also holds in case of contaminated data. As a matter of fact, expression (3.1) can be computed entirely from the original sample. Similarly to (2.5), the bootstrap MISE estimator is easier manipulated when written as

$$\begin{aligned}\mathrm{MISE}^*(h) =&(2\pi nh)^{-1}\int|\varphi_K(t)|^2|\varphi_Z(t/h)|^{-2}\,dt+(1-n^{-1})(2\pi)^{-1}\int|\widehat{\varphi}_{X,g}(t)|^2|\varphi_K(ht)|^2\,dt\\ &+R(\widehat{f}_X(\cdot;g))-\pi^{-1}\int|\widehat{\varphi}_{X,g}(t)|^2\varphi_K(ht)\,dt, \qquad (3.2)\end{aligned}$$

where $\widehat{\varphi}_{X,g}(t)$ is the Fourier transform of $\widehat{f}_X(\cdot;g)$. Note that minimizing $\mathrm{MISE}^*(h)$ with respect to h is equivalent to minimizing

$$\mathrm{MISE}_2^*(h) =(2\pi nh)^{-1}\int|\varphi_K(t)|^2|\varphi_Z(t/h)|^{-2}\,dt+(1-n^{-1})(2\pi)^{-1}\int|\widehat{\varphi}_{X,g}(t)|^2|\varphi_K(ht)|^2\,dt$$

$$- \pi^{-1} \int |\widehat{\varphi}_{X,g}(t)|^2 \varphi_K(ht)\, dt,$$

with respect to h, and this only requires calculation of $\widehat{\varphi}_{X,g}(\cdot)$ from the original sample. Furthermore we have that $\widehat{\varphi}_{X,g}(t) = \widehat{\varphi}_{Y,n}(t).\varphi_L(gt)/\varphi_Z(t)$, with $\widehat{\varphi}_{Y,n}$ the empirical characteristic function of Y. See Remark 2 below. So, in practice the only quantity that needs to be calculated in $\text{MISE}_2^*(h)$ is this empirical characteristic function. This simplifies considerably the computations involved in this bootstrap bandwidth selection procedure.

Remark 1. In this method we used a smoothed estimator of f_X since resampling with replacement from the data would lead to a bootstrap estimator of the bias equal to zero, resulting in a bootstrap bandwidth chosen on the basis of the variance only (instead of the whole MISE). See Faraway and Jhun (1990) for the non-contaminated case and Delaigle (1999) for the contaminated case.

Remark 2. A useful expression for calculation of $\widehat{\varphi}_{X,g}(t)$ in (3.2) can be derived as follows. Let $\widehat{f}_Y(.;g)$ be the 'usual kernel estimator' of f_Y based on the sample $Y_1, \ldots, Y_n$, and using the kernel L, i.e. $\widehat{f}_Y(x;g) = n^{-1} \sum_{1 \le i \le n} L_g(x - Y_i)$. By Lemma A.6 of the Appendix, we see that $\widehat{\varphi}_{X,g}(t) = \widehat{\varphi}_{Y,g}(t)/\varphi_Z(t)$, where $\widehat{\varphi}_{Y,g}(t)$ denotes the characteristic function of $\widehat{f}_Y(x;g)$, for which it is easily proved that it equals $\widehat{\varphi}_{Y,n}(t).\varphi_L(gt)$, with $\widehat{\varphi}_{Y,n}$ the empirical characteristic function of Y. We conclude that $\widehat{\varphi}_{X,g}(t) = \widehat{\varphi}_{Y,n}(t).\varphi_L(gt)/\varphi_Z(t)$.

3.2 Choice of the pilot bandwidth

The quality of the estimator of the density f_X depends strongly on the choice of the bandwidth h, but it also depends on the pilot bandwidth g. In case of error free data, it is already known that, to give good estimates, the bandwidth g should preferably be of an order of magnitude larger than h, at least when using second order kernels. See Faraway and Jhun (1990), Falk (1992), Hall, Marron and Park (1992), Marron (1992), Hazelton (1999) and Jones (2000). In this context, efficient procedures for choosing g have been proposed. See for example Jones, Marron and Park (1991), Hall, Marron and Park (1992) or Jones, Marron and Sheather (1996) for a survey. Throughout this paper we assume that K is a second order kernel.

The simple choice of g that we propose here is based on the fact that, under sufficient smoothness conditions, an asymptotic representation of the MISE is

$$\text{AMISE}(h) = \frac{h^4}{4}\mu_2^2(K) R\left(f_X''\right) + (2\pi nh)^{-1} \int |\varphi_K(t)|^2 . |\varphi_Z(t/h)|^{-2}\, dt,$$

where $\mu_2(K) = \int u^2 K(u) du$ denotes the second moment of the kernel K. See Proposition

1 in Section 4. The bootstrap estimator of this quantity is

$$\text{AMISE}^*(h) = \frac{h^4}{4}\mu_2^2(K)R\big(\widehat{f}_X''(\cdot; g)\big) + (2\pi nh)^{-1}\int |\varphi_K(t)|^2.|\varphi_Z(t/h)|^{-2}\,dt\,. \tag{3.3}$$

From (3.3) we see that asymptotically, the only g related quantity is $R\big(\widehat{f}_X''(\cdot; g)\big)$, which is best estimated by choosing g as the optimal bandwidth for estimating $R(f_X'')$. See Marron (1992) for a similar remark in the error free case. Delaigle and Gijbels (2001a) provide a practical optimal bandwidth for estimating $R(f_X'')$, which is based on minimization of the mean squared error of $R\big(\widehat{f}_X''(\cdot; g)\big)$. In our simulation study this choice of initial bandwidth g proved to work well, and hence we propose to use this pilot bandwidth in practice. Of course, any bandwidth g which satisfies the conditions of the theorems of Section 4 would lead to a consistent bootstrap procedure, and therefore any appropriate choice of the initial g would do the job.

4 Consistency of the bootstrap method

Throughout this paper we will assume that h and g tend to zero and nh and ng tend to infinity as n tends to infinity. These are indeed classical conditions necessary to insure the convergence of a density estimator to the target density. The following conditions will be necessary.

Condition A:

(A1) $\sup_{x\in\mathbb{R}} |f_X^{(j)}(x)| < \infty \qquad$ for $j = 1, \ldots, 4$;

(A2) $\int |f_X''(x)|\,dx < \infty$ and $\int |f_X^{(3)}(x)|dx < \infty$;

(A3) f_X is square integrable.

Condition B:

(B1) K and L are symmetric and bounded continuous functions such that $\int K(x)\,dx = 1$, $\int L(x)\,dx = 1$, $\int y^{12}|K(y)|\,dy < \infty$ and $\int y^2|L(y)|\,dy < \infty$;

(B2) $\varphi_K(t)$ and $\varphi_L(t)$ are supported on respectively $[-B_K, B_K]$ and $[-B_L, B_L]$, with $0 < B_K, B_L < \infty$.

Condition C:

(C1) $\left\{\int |L^Z * L^Z(u; g)|\,du\right\}\left\{\int |t|^j|\varphi_L(t)|^2|\varphi_Z(t/g)|^{-2}\,dt\right\} = o(n^2g^{j+2})$ for $j = 0, \ldots, 8$;

(C2) $\int |L * L^Z(u; g)|\,du = o(\sqrt{n})$;

(C3) $\lim_{n\to\infty} h^{-1}\int |\varphi_K(t)|^2|\varphi_Z(t/h)|^{-2}\,dt = \infty \quad$ and $\quad \int |\varphi_K(t)|^2|\varphi_Z(t/h)|^{-2}\,dt = o(nh)$;

(C4) $\int t^4 |\varphi_L(t)|^2 |\varphi_Z(t/g)|^{-2}\, dt = o(ng^5)$.

The above set of conditions might look quite overwhelming at first sight, especially Condition C. However, for a given density function f_Z and given kernels K and L, this condition can be shown to result into conditions on the bandwidths h and g. In Section 4.2 we will discuss to which conditions on h and g it leads in case of a certain class of ordinary smooth error densities. Moreover, we will explain in that section that the practical pilot bandwidth g as proposed in the previous section, and the optimal bandwidth h of for example Fan (1991b) satisfy the above assumptions.

Concerning Condition B, an example of a kernel with 12 finite moments is given by

$$K(x) = c\left(\frac{\sin(x/14)}{x/14}\right)^{14},$$

with c a normalizing constant. Its characteristic function is supported on [-1,1].

4.1 Consistency results

In what follows we prove that the bootstrap MISE estimator is a consistent estimator of the MISE. Therefore we first establish the asymptotic orders of the bias and variance of the bootstrap integrated squared bias and the bootstrap integrated variance separately. Then we gather those properties in a theorem concerning the whole bootstrap estimated MISE. To simplify the proofs, we will suppose that φ_Z is symmetric. Note however that with a little more effort, the proofs can be adapted to a non symmetric φ_Z.

Remark 3. Note that the symmetry of φ_Z together with the symmetry of a kernel K implies that the function $K^Z(\cdot; h)$ as defined in (2.2) is symmetric. This fact will be used in the proofs.

Proposition 1 below describes in detail the behaviour of the integrated squared bias, the integrated variance, and mean integrated squared error of the deconvolving kernel estimator. See Stefanski and Carroll (1990). In the following, the integrated squared bias and its bootstrap counterpart will be denoted by respectively $\mathrm{ISB}(h)$ and $\mathrm{ISB}^*(h)$.

Proposition 1.

(i). Under Conditions (A1), (A2), (B2), and if K is a second order kernel with $\int |u^3 K(u)|\, du < \infty$, we have

$$\mathrm{ISB}(h) = \frac{h^4}{4}\mu_2^2(K) R(f_X'') + o(h^4).$$

(ii). Under Conditions (A3), (B2) and (C3) we have

$$\int \mathrm{Var}\{\widehat{f}_X(t;h)\}\,dt = (2\pi nh)^{-1}\int |\varphi_K(t)|^2|\varphi_Z(t/h)|^{-2}\,dt + O(n^{-1}).$$

(iii). Under Conditions (A1) to (A3), Conditions (B2) and (C3) and if K is a second order kernel with $\int |u^3K(u)|\,du < \infty$, and $(nh^6)^{-1}\int |\varphi_K(t)|^2|\varphi_Z(t/h)|^{-2}\,dt \to \infty$ as $n\to\infty$ and $nh^4 \to \infty$, as $n \to \infty$, we have

$$\mathrm{MISE}(h) = \frac{h^4}{4}\mu_2^2(K)R(f_X'') + (2\pi nh)^{-1}\int |\varphi_K(t)|^2|\varphi_Z(t/h)|^{-2}\,dt + O(n^{-1}) + o(h^4).$$

The next two propositions provide the asymptotic order of the bias and the variance of the bootstrap integrated variance (Proposition 2) and of the bootstrap integrated squared bias (Proposition 3).

Proposition 2. *Under Conditions (A3), B and (C3), we have*

(i) $\mathrm{Bias}\left[\int \mathrm{Var}^*\{\widehat{f}_X^*(t;h)\}\,dt\right] = O(n^{-1})$.

(ii) $\mathrm{Var}\left[\int \mathrm{Var}^*\{\widehat{f}_X^*(t;h)\}\,dt\right] = o(n^{-2})$.

Proposition 3. *Under Conditions (A1), (A2) and B*

(i) If (C4) is satisfied, we have $\mathrm{Bias}\,[\mathrm{ISB}^*(h)] = o(h^4)$.

(ii) If (C1) and (C2) are satisfied, we have $\mathrm{Var}[\mathrm{ISB}^*(h)] = o(h^8)$.

Since $\mathrm{MISE}^*(h) = \mathrm{ISB}^*(h) + \int \mathrm{Var}^*\{\widehat{f}_X^*(t;h)\}\,dt$, and hence

$$\mathrm{E}\,[\mathrm{MISE}^*(h)] = \mathrm{E}\,[\mathrm{ISB}^*(h)] + \mathrm{E}\left[\int \mathrm{Var}^*\{\widehat{f}_X^*(t;h)\}\,dt\right]$$

and

$$\begin{aligned}\mathrm{Var}\,[\mathrm{MISE}^*(h)] \;\le\;& \mathrm{Var}\,[\mathrm{ISB}^*(h)] + \mathrm{Var}\left[\int \mathrm{Var}^*\{\widehat{f}_X^*(t;h)\}\,dt\right]\\ &+2\sqrt{\mathrm{Var}\,[\mathrm{ISB}^*(h)]\cdot \mathrm{Var}\left[\int \mathrm{Var}^*\{\widehat{f}_X^*(t;h)\}\,dt\right]}\,,\end{aligned}$$

combining Propositions 2 and 3 leads to the next theorem, which establishes the convergence of the bootstrap MISE estimator to the exact MISE.

Theorem 1. *Under Conditions A, B and C, we have*

$$\begin{aligned}
\mathrm{E}\left[\mathrm{MISE}^*(h)\right] &= \mathrm{MISE}(h) + o(h^4) + O(n^{-1}) = \mathrm{MISE}(h) + o(\mathrm{MISE}(h)),\\
\mathrm{Var}\left[\mathrm{MISE}^*(h)\right] &= o(n^{-2}) + o\left(h^8\right) + o(h^4 n^{-1}) = o(\mathrm{MISE}^2(h)),
\end{aligned}$$

or also $\mathrm{E}\left[\dfrac{\mathrm{MISE}^*(h)}{\mathrm{MISE}(h)}\right] = 1 + o(1)$ *and* $\mathrm{Var}\left[\dfrac{\mathrm{MISE}^*(h)}{\mathrm{MISE}(h)}\right] = o(1)$, *i.e.*

$$\frac{\mathrm{MISE}^*(h)}{\mathrm{MISE}(h)} \xrightarrow{L_2} 1, \quad \text{as } n \to \infty.$$

We conclude that under certain conditions, the MISE and its bootstrap counterpart are asymptotically equivalent. From Theorem 1, we deduce the following theorem, which shows that if we restrict our search of the bootstrap bandwidth to an interval around the real optimal bandwidth, the MISE and the bootstrap MISE used either with the real optimal bandwidth or with its bootstrap estimator are also asymptotically equivalent.

Theorem 2. *Let* $h_n = \mathrm{argmin}_h\, \mathrm{MISE}(h)$ *and* $h_n^* = \mathrm{argmin}_{h\in[A_1 h_n, A_2 h_n]}\, \mathrm{MISE}^*(h)$, *with constants* $0 < A_1 < 1$ *and* $A_2 > 1$. *Under Conditions A, B and C, we have*

$$\frac{\mathrm{MISE}^*(h_n^*)}{\mathrm{MISE}(h_n^*)} \xrightarrow{P} 1 \quad \text{and} \quad \frac{\mathrm{MISE}^*(h_n)}{\mathrm{MISE}(h_n)} \xrightarrow{P} 1, \ \text{as } n \to \infty.$$

Finally the next theorem proves the consistency of our bootstrap bandwidth selector.

Theorem 3. *Let* h_n *and* h_n^* *be as in Theorem 2. Under Conditions A, B and C, we have*

$$\frac{h_n^*}{h_n} \xrightarrow{P} 1, \quad \text{as } n \to \infty.$$

From this theorem we learn that under sufficient conditions, as the sample size increases, the bootstrap bandwidth tends to the real optimal bandwidth.

4.2 Verifying Condition C

The aim of this section is to get a closer look to Condition C, in order to demonstrate that for a given error density and kernels K and L, this set of conditions can be translated into conditions on the bandwidths h and g.

Let f_Z be such that $\{\varphi_Z(t)\}^{-1} = \sum_{j=0}^{p} c_{2j} t^{2j}$, where $c_0 = 1, c_2, c_4, \dots, c_{2p-2}$ are constants and c_{2p} is a constant different from zero. Note that f_Z is an ordinary smooth density of order $\beta = 2p$. An example of such an error density is a Laplace density for which $\{\varphi_Z(t)\}^{-1} = 1 + \sigma_Z^2 t^2$, corresponding to a case with $p = 1$ in the above class of densities.

Suppose now that the kernel function K satisfies (B2) and the following condition:

Condition D. K is $2p$ times differentiable and $\int |K^{(2j)}(x)\,dx| < \infty$, for $j = 0, \ldots, p$;

In order to check Condition C, we investigate first the behaviour of the quantities appearing in this condition. The following results can be shown:

Result 1 : Under Condition D, we have

$$\int |t|^j |\varphi_K(t)|^2 |\varphi_Z(t/h)|^{-2}\,dt = O(h^{-4p})\ , \qquad \text{for } j = 0, 1, 2, \cdots$$

Result 2 : Under Condition D, $\int |K * K^Z(x; h)|\,dx = O(h^{-2p})$.

Result 3 : Under Condition D, $\int |K^Z * K^Z(x; h)|\,dx = O(h^{-4p})$.

The proofs of these results are reasonably straightforward and are omitted. Details can be obtained from the authors upon request.

The above results will allow us to verify Condition C for this class of error densities and kernels. Assuming that the kernels K and L satisfy Condition D, it is easily seen that the quantities involved in Condition C behave as follows:

$$\left\{\int |L^Z * L^Z(u; g)|\,du\right\}\left\{\int |t|^j |\varphi_L(t)|^2 |\varphi_Z(t/g)|^{-2}\,dt\right\}$$
$$= O(g^{-8p}) = o(n^2 g^{j+2}) \qquad \text{for } j = 0, \cdots, 8; \tag{4.1}$$

$$\int |L * L^Z(u; g)|\,du = O(g^{-2p}) = o(\sqrt{n}); \tag{4.2}$$

$$h^{-1} \int |\varphi_K(t)|^2 |\varphi_Z(t/h)|^{-2}\,dt = O(h^{-1-4p}) \to \infty \text{ as } n \to \infty. \tag{4.3}$$

$$\int |\varphi_K(t)|^2 |\varphi_Z(t/h)|^{-2}\,dt = O(h^{-4p}) = o(nh). \tag{4.4}$$

$$\int t^4 |\varphi_L(t)|^2 |\varphi_Z(t/g)|^{-2}\,dt = O(g^{-4p}) = o(ng^5), \tag{4.5}$$

where the expression following the second equality sign or the arrow indicates what is required by Condition C.

Putting $h = cn^{-\alpha}$ and $g = dn^{-\gamma}$, with c and d positive constants, we can translate the requirements in (4.1)—(4.5) into the following very simple requirements on h and g:

$$0 < \alpha < \frac{1}{4p+1} \qquad \text{and} \qquad 0 < \gamma < \frac{1}{4p+5}.$$

The optimal bandwidths h and g satisfy these requirements. Indeed, for ordinary smooth error densities of order $\beta = 2p$, the MISE-optimal bandwidth h for estimating f_X in case of contaminated error is of order

$$h \sim n^{-\frac{1}{2k+4p+1}},$$

where $k > 0$ denotes the order of the kernel function. This result can be found in Fan (1991c). The MSE-optimal bandwidth g for estimating $R(f_X'')$ is, for ordinary smooth error densities order of $\beta = 2p$, of order

$$g \sim n^{-\frac{1}{k+4p+5}},$$

as has been shown by Delaigle and Gijbels (2001a).

One can also easily check that when h and g satisfy Condition C, then also do $c.h$ and $d.g$ where c and d are positive constants. As a consequence, all bandwidths h and g which behave (in rate) as the optimal bandwidths also satisfy the conditions.

5 Simulations and real data examples

5.1 Simulation study

We now investigate the finite sample performance of the bootstrap bandwidth selection procedure via a simulation study. Our study involves four f_X densities, chosen because they show some typical features that can be encountered in practice. These densities, in increasing order of estimation difficulty, are:

1. Density #1: $X \sim$ N(0,1), the standard normal density.
2. Density #2: $X \sim \chi^2(8)$, chosen because it is skewed.
3. Density #3: $X \sim$ 0.5 N(-3,1)+0.5 N(2,1), chosen for its two clearly separated modes.
4. Density #4: $X \sim$ 0.4 Gamma(5)+0.6 Gamma(13), which is bimodal, but with close and different sized modes.

Figure 5.1 shows the four target densities. From each of these densities, 500 samples of size $n = 50$, 100 and 250 were generated, each of which was then contaminated by a sample from either a N(0;σ_Z^2) or a Laplace($\sigma_Z/\sqrt{2}$) error density. For each configuration, the parameter σ_Z was chosen such that the ratio σ_Z^2/σ_X^2 equals 0.25, except for density #4 where we took $\sigma_Z^2/\sigma_X^2 = 0.1$ (the latter density is more difficult to recover). As Wand (1998) remarked, this ratio does not allow to compare, in an absolute sense, the difficulty of the estimation across densities since its meaning can be very different according to the density considered. In our simulation study we use the kernels $K = L$ as defined in (2.3).

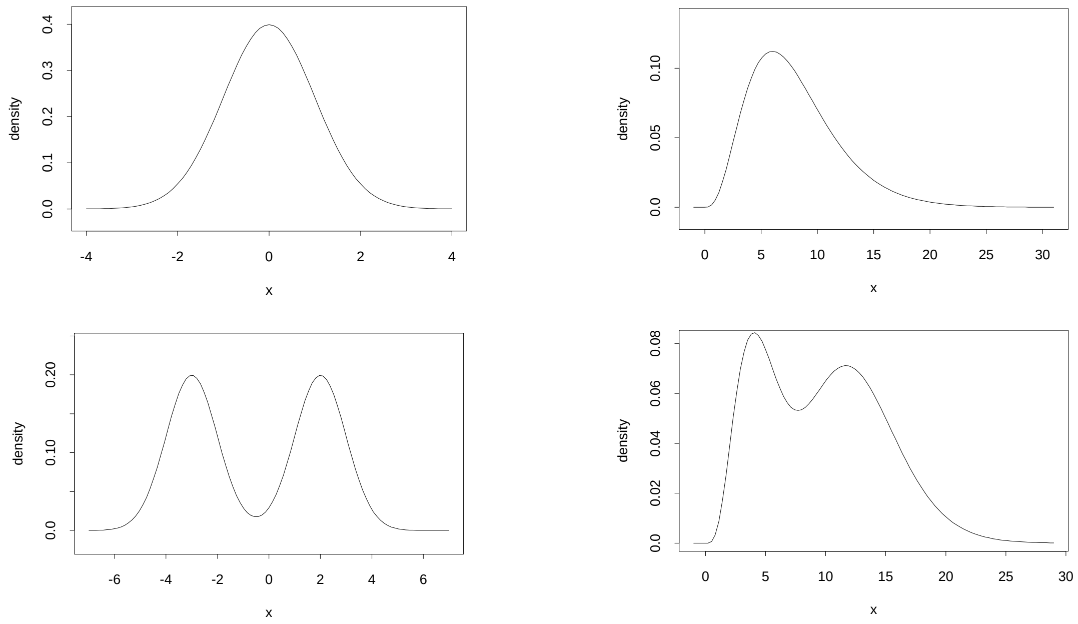

Figure 5.1: The four target densities: a normal density (top left panel), a chi-squared density (top right panel), a mixture of two normal densities (bottom left panel), and a mixture of two gamma densities (bottom right panel).

To apply our method in practice, we need to

1. Choose an initial bandwidth g. We used the 3-stage plug-in bandwidth selector to estimate $R(f_X'')$ discussed in Delaigle and Gijbels (2001a);

2. On a grid of h, search the value which minimizes $\mathrm{MISE}_2^*(h)$ and obtain $\widehat{h}_{n,\mathrm{boot}}$;

3. Estimate $f_X(x)$ by $\widehat{f}_X(x; \widehat{h}_{n,\mathrm{boot}})$.

In order to evaluate the performance of the estimation procedure, we compute the Integrated Squared Error (ISE) of $\widehat{f}_X(\cdot; \widehat{h}_{n,\mathrm{boot}})$ for each calculated estimate. This allows us to classify the 500 estimates from the best one (i.e. the one with the smallest ISE) till the worse one (i.e. the one with the largest ISE). In all figures reported below, the estimates represented correspond to the first, second and third quartile of these 500 ordered ISE's. The target density is always represented as a continuous curve.

Figure 5.2 shows the results of the estimation of a $\chi^2(8)$ density (density #2) with different error densities (Laplace error and gaussian error) and sample sizes. We see that even for a sample of size as small as 50, the method performs quite well. We observed similar results when trying to recover a standard normal density, and hence we do not report them here.

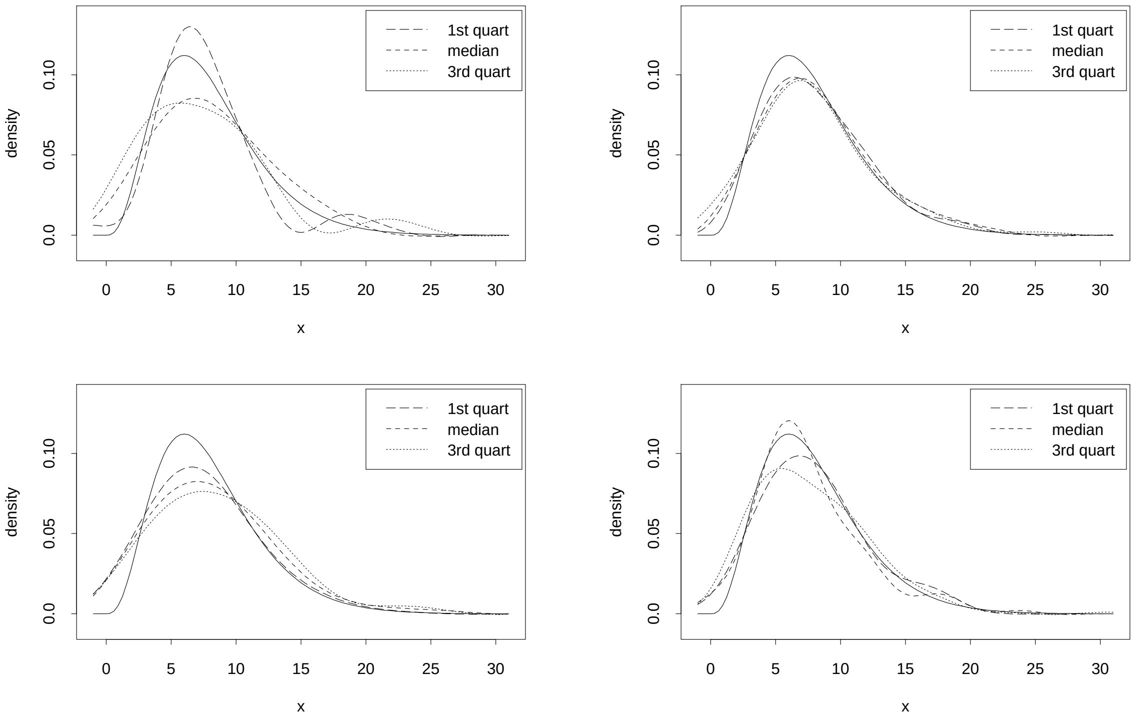

Figure 5.2: Estimation of the $\chi^2(8)$ density for a Laplace error with $n = 50$ (top left panel), and $n = 250$ (top right panel), or for a gaussian error with $n = 50$ (bottom left panel), and $n = 250$ (bottom right panel).

Figure 5.3 is similar to Figure 5.2, but for the configuration with the mixed normal density (density #3). We see that the estimation was a bit more difficult in this case. The estimate managed to recover the two modes, but the peaks were underestimated. In this case, since the density presents quite different features at different places, the estimator would rather require use of a local bandwidth: a small bandwidth at peaks and a larger bandwidth elsewhere. Despite this difficulty, we see that the bootstrap selection method performs quite good in recovering the density, except with a gaussian error for too small sample sizes.

Recovering the mixed gamma density (density #4) turned out to be more difficult. Figure 5.4 shows the results obtained with a Laplace noise for sample sizes 50 and 250. We see that even if from a MISE point of view the density was well estimated, the estimate had yet difficulties to recover the two modes. For a larger sample size such as 250, the two modes were more often detected, but then the estimate had difficulties to recover their shapes. However as in the case of a mixed normal density, this problem is inherent to the choice of a global parameter, and is not really due to a failure of the method itself. It would not be difficult to adapt our technique to the search of a local bandwidth, and it is obvious that such a bandwidth would perform better (but would require more computations).

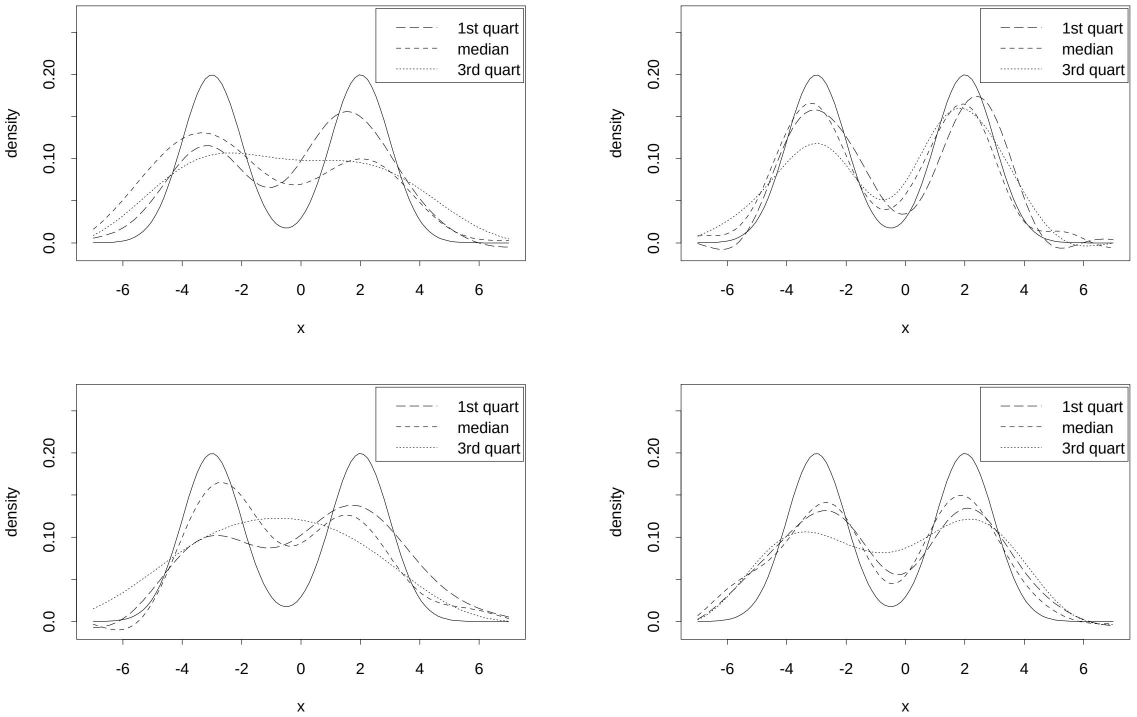

Figure 5.3: Estimation of the mixed normal density for a Laplace error with $n = 50$ (top left panel), and $n = 250$ (top right panel), or for a gaussian error with $n = 50$ (bottom left panel), and $n = 250$ (bottom right panel).

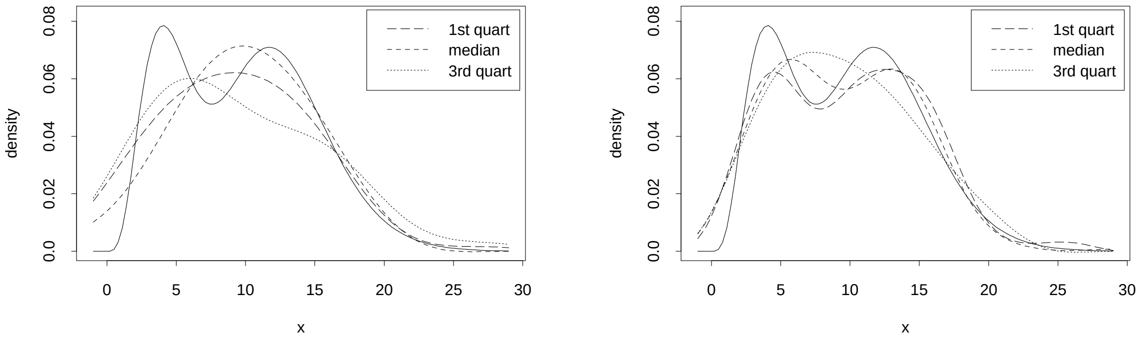

Figure 5.4: Estimation of the mixed gamma density for a Laplace error with $n = 50$ (left panel) and $n = 250$ (right panel).

The conclusions from the simulations remain unchanged if we replace the Laplace error by any ordinary smooth density, or if we replace the gaussian error by any supersmooth density. We also believe that for most regular densities f_X (i.e. densities without any strong feature) the estimation method with bootstrap bandwidth selector will perform quite well, at least for a reasonable value of the ratio σ_Z^2/σ_X^2. As a matter of fact one should not expect to get a very good estimate of f_X if the sample is too contaminated by noise.

5.2 Real data applications

5.2.1 The NHANES study

The first data come from the Second National Health And Nutrition Examination Survey (1976–1980), abbreviated as NHANES II study. The interest is to estimate the density of the long-term log daily saturated fat intake based on a sample of size 4708, consisting of women aged between 25 and 50. NHANES II is the second phase of a previous study NHANES I, which has been analyzed by, for example, Stefanski and Carroll (1990) and Carroll, Ruppert and Stefanski (1995). We use the same transformation as in the latter paper, and work with the variable log(5+saturated fat). These data are known to be contaminated by a considerable noise: it is reasonable to assume that more than 50% of the variance of the data is due to the noise. See for example Stefanski and Carroll (1990) or Carroll, Ruppert and Stefanski (1995).

We applied the bootstrap bandwidth estimation method to these data, for two different error densities (gaussian and Laplace), and an error variance of approximately 1.5 σ_X^2, which corresponds to 60% of the variance of the data due to the noise. Note that Stefanski and Carroll (1990) encountered problems in deconvolving this much noise with their cross-validation method. By contrast we will see that the bootstrap method performs quite well. For the sake of illustration, and since this variance is estimated through knowledge of previous studies and may not be the actual variance, we follow Stefanski and Carroll (1990) and also try to estimate the density of X assuming that $\sigma_Z^2 = (1/5)\,\sigma_X^2$ or $\sigma_Z^2 = (1/3)\,\sigma_X^2$. We also consider the case $\sigma_Z^2 = 0$ (the error free case).

The resulting estimators are depicted in Figure 5.5. We see that the error variance plays an important role in the estimation process: the smaller this variance, the smoother the estimator will be. For a large error variance (top left panel), the cases of normal and Laplace errors differ only a little, but we see that decreasing the value of σ_Z^2 makes the estimators based on the gaussian and the Laplace error assumption almost indistinguishable, as is to be expected. In all cases, the estimators are almost symmetric, with a small tendency to be skewed to the left, which could be due to underreport of the high saturated fat intake from the patients, as already remarked by Stefanski and Carroll (1990).

This example illustrates the importance of taking the noise into account, since if we compare the top left panel with the bottom right panel, we see that assuming no noise in the data results in a too smooth estimator of the density. This could hide some important features of the data for example in case of multimodality.

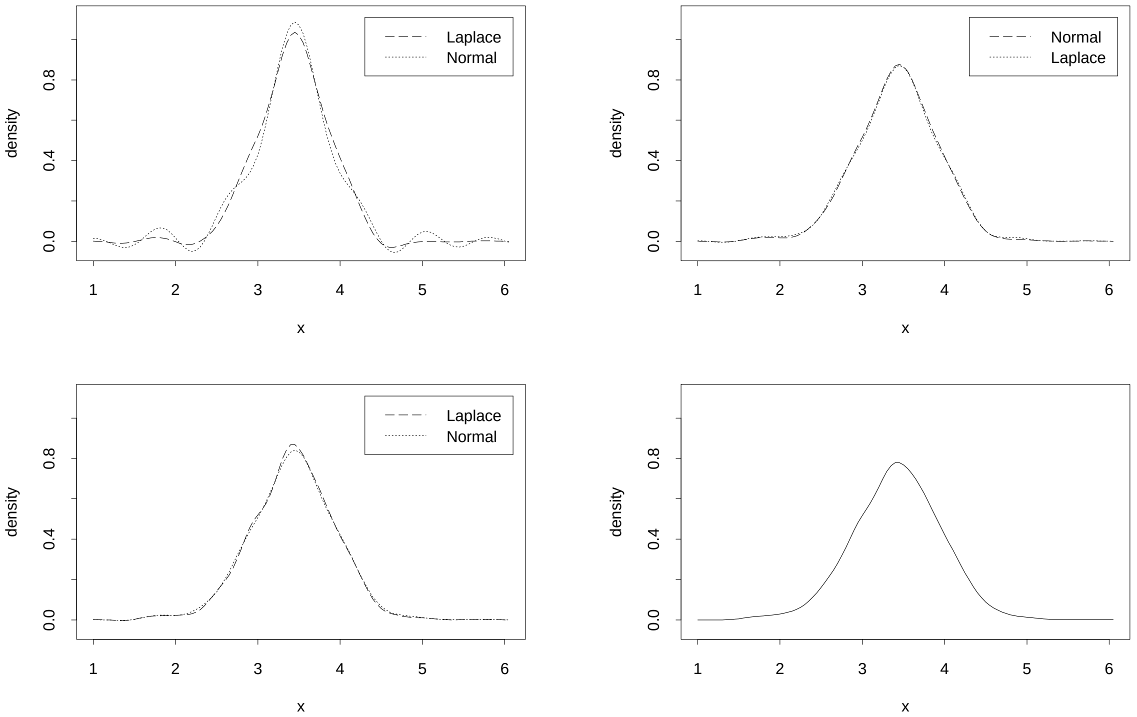

Figure 5.5: Estimation of the log saturated fat density for a Laplace error or a normal error with $\sigma_Z^2 = 1.5\,\sigma_X^2$ (top left panel), $\sigma_Z^2 = (1/3)\,\sigma_X^2$ (top right panel), $\sigma_Z^2 = (1/5)\,\sigma_X^2$ (bottom left panel), and no error (bottom right panel).

5.2.2 Diet and heart data

The second data come from a pilot study on coronary heart disease, reported by Morris, Marr and Clayton (1977) and analyzed in, for example, Clayton (1992) and Cordy and Thomas (1997). The data consist of measurements of the ratio of poly-unsaturated fat to saturated fat intake for 336 men in a one-week full weighted dietary survey. Among them, 61 individuals completed two such surveys, reporting on their diet during two weeks, separated by 6 months in time. We suppose that the actual ratio is invariant within an individual (see also Clayton (1992)). Since the error distribution is unknown, we will compute the density estimator of the actual ratio in case of a normal error distribution and a Laplace error distribution.

The error variance may be estimated through the two repeated measurements on 61 individuals. More precisely we will estimate σ_Z^2 by half the empirical variance of the 61 differences between the repeated measurements on each individual, which corresponds approximately to the situation that $\sigma_Z^2 = (2/3)\,\sigma_X^2$. This estimation of the error variance may be not very accurate but can at least give us some insight into it. Since we cannot guarantee that this error variance is close to the exact one, we also consider $\sigma_Z^2 = (4/3)\,\sigma_X^2$, $\sigma_Z^2 = (1/3)\,\sigma_X^2$ and even $\sigma_Z^2 = 0$.

The results are depicted in Figure 5.6. The figure suggests that the actual density is skewed to the right, perhaps meaning that people have a tendency to underreport their saturated fat intake. The estimators assuming different error distributions do not differ much. Note that the variance estimated through repeated measurements seems to be plausible (the other variances seem to be either too large or too small).

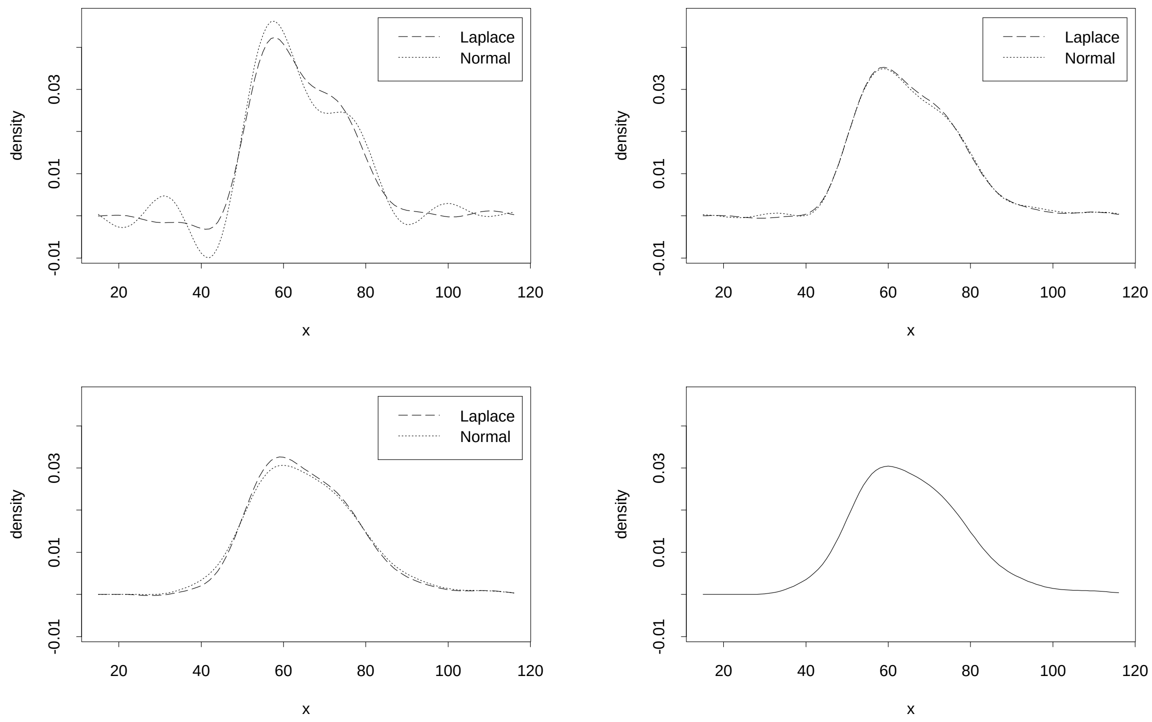

Figure 5.6: Estimation of the ratio of poly-unsaturated fat to saturated fat intake for a Laplace error or a normal error with with $\sigma_Z^2 = 4/3\,\sigma_X^2$ (top left panel), $\sigma_Z^2 = (2/3)\,\sigma_X^2$ (top right panel), $\sigma_Z^2 = (1/3)\,\sigma_X^2$ (bottom left panel), and no error (bottom right panel).

A Appendix

In this section we provide the proofs of the main results as stated in Section 4. These proofs rely on some useful facts about various functionals involving the kernels K and L and the 'kernels' K^Z and L^Z, defined in (2.2), and on some facts about the empirical characteristic function. These facts are summarized in the next section.

We first introduce some notations which will be used throughout the appendix:

$$\begin{aligned}
&D(x) = K(x) - \delta_0(x) && \text{with } \delta_0 \text{ the Dirac delta function}\\
&a(x) = D * D(x) && \\
&b(x) = L^Z * L^Z(x; g) && \\
&B(x) = L * L^Z(x; g) && \\
&\mu_i(K) = \textstyle\int u^i K(u) du && \text{the } i\text{th moment of the kernel } K\\
&\mu_i(K * K) = \textstyle\int u^i (K * K)(u) du && \text{the } i\text{th moment of the convolution } K * K,
\end{aligned} \tag{A.1}$$

and similar moments for the kernel L and the convolution $L * L$.

From Condition B, we know that the functions $\varphi_K(\cdot)$ and $\varphi_L(\cdot)$ are bounded by a finite constant. Without loss of generality we will suppose that this bound is equal to 1.

A.1 Preliminary lemmas

We start by recalling a form of Taylor's theorem with integral form of the remainder term.

Theorem 4 (Integral form of the remainder term in a Taylor expansion). *Let $m \in \mathbb{N}$ and f be a real-valued function which is $(m+1)$ times differentiable on an interval I. Then $\forall a \in I$ and $\forall h \neq 0 \in I$ such that $a + h \in I$, we can write*

$$f(a+h) = \sum_{j=0}^{m} f^{(j)}(a) \frac{h^j}{j!} + \frac{h^{m+1}}{m!} \int_0^1 f^{(m+1)}(a+hu)(1-u)^m \, du. \tag{A.2}$$

The next lemma is essentially due to Stefanski and Carroll (1990). See also Delaigle and Gijbels (2001a) for a proof.

Lemma *A.1.* For a kernel K satisfying Condition (B2) we have

$$\mathrm{E}\Big[K^Z\Big(\frac{x-Y}{h}; h\Big)\Big] = \mathrm{E}\Big[K\Big(\frac{x-X}{h}\Big)\Big]. \tag{A.3}$$

Lemma *A.2.* Under Condition (B1), we have

$$\int |D * D(x)| |x|^j \, dx < \infty \qquad \text{for } j = 0, \ldots, 12. \tag{A.4}$$

Proof of Lemma A.2. Under Condition (B1), we have

$$
\begin{aligned}
D*D(x) &= \int D(x-y)D(y)\,dy \\
&= \int [K(x-y)-\delta_0(x-y)][K(y)-\delta_0(y)]\,dy \\
&= \int K(x-y)K(y)\,dy - \int K(x-y)\delta_0(y)\,dy - \int \delta_0(x-y)K(y)\,dy + \int \delta_0(x-y)\delta_0(y)\,dy \\
&= \int K(x-y)K(y)\,dy - 2K(x) + \delta_0(x),
\end{aligned}
$$

and thus $|D*D(x)| \le \int |K(x-y)K(y)|\,dy + 2|K(x)| + \delta_0(x)$, and hence

$$
\begin{aligned}
\int |D*D(x)||x|^j\,dx &\le \int_{\mathcal{R}^2} |x|^j |K(x-y)K(y)|\,dy\,dx + 2\int |x|^j|K(x)|\,dx + \int |x|^j\delta_0(x)\,dx \\
&= \int_{\mathcal{R}^2} |x|^j |K(x-y)K(y)|\,dy\,dx + 2\int |x|^j|K(x)|\,dx.
\end{aligned}
$$

Using Condition (B1) and the change of variable $u = x-y$, we find

$$
\begin{aligned}
\int_{\mathcal{R}^2} |x|^j|K(x-y)K(y)|\,dy\,dx &= \int_{\mathcal{R}^2} |u+y|^j|K(u)||K(y)|\,dy\,du \\
&\le \sum_{k=0}^{j} C_j^k \int_{\mathcal{R}^2} |u|^k|y|^{j-k}|K(u)||K(y)|\,dy\,du < \infty,
\end{aligned}
$$

where C_j^k denotes the binomial coefficient $\binom{j}{k}$. This proves the lemma. □

Lemma *A.3.* Under Condition (B1), we have

$$
\mu_i(K) = \mu_i(L) = \mu_i(L*L) = \mu_j(D*D) = 0 \quad \text{for } i = 1,3,5 \text{ and } j = 0,1,2,3. \tag{A.5}
$$

$$
\mu_0(L*L) = 1, \qquad \mu_2(L*L) = 2\mu_2(L), \qquad \mu_4(D*D) = 6\mu_2^2(K). \tag{A.6}
$$

Proof of Lemma A.3. The proof is straightforward, but rather technical and will be omitted here. □

Lemma *A.4.* Under Condition (B2), for all $x \in I\!R$ and $j \in I\!N$, we have

$$
|b^{(j)}(x)| \le (2\pi)^{-1} \int |t|^j |\varphi_L(t)|^2 |\varphi_Z(t/g)|^{-2}\,dt.
$$

Proof of Lemma A.4. Under (B2), we have

$$\begin{aligned}
b^{(j)}(x) &= \int \{L^Z\}^{(j)}(x-y;g)L^Z(y;g)\,dy \\
&= (2\pi)^{-2}\int\Big(\int(-it)^j e^{-it(x-y)}\varphi_L(t)/\varphi_Z(t/g)\,dt\Big)\Big(\int e^{-isy}\varphi_L(s)/\varphi_Z(s/g)\,ds\Big)\,dy \\
&= (2\pi)^{-2}\int\Big(\int e^{ity}g_1(t)\,dt\Big)\Big(\int e^{-isy}\bar{g}_2(s)\,ds\Big)\,dy \\
&= (2\pi)^{-2}\int \varphi_{g_1}(y)\bar{\varphi}_{g_2}(y)\,dy \\
&= (2\pi)^{-1}\int g_1(t)\bar{g}_2(t)\,dt \\
&= (2\pi)^{-1}\int(-it)^j e^{-itx}\varphi_L^2(t)\varphi_Z^{-2}(t/g)\,dt,
\end{aligned}$$

where $g_1(t)=(-it)^j e^{-itx}\varphi_L(t)/\varphi_Z(t/g)$, $\bar{g}_2(s)=\varphi_L(s)/\varphi_Z(s/g)$, and where we used Parseval's identity. This proves the result. $\square$

Lemma *A.5.* Under Conditions (B2) and (C1) we have

$$\int |b(u)b^{(j)}(u+x)|\,du = o(n^2g^{j+2}) \qquad \forall x\in I\!R, \qquad \text{for } j=0,\dots,8. \tag{A.7}$$

Proof of Lemma A.5. Under Condition (B2), we have

$$\begin{aligned}
\int |b(u)b^{(j)}(u+x)|\,du \le& \sup_u |b^{(j)}(u)|\int |b(u)|\,du \\
\le&(2\pi)^{-1}\int |t|^j|\varphi_L(t)|^2|\varphi_Z(t/g)|^{-2}\,dt\int|b(u)|\,du \\
=&(2\pi)^{-1}\Big\{\int |t|^j|\varphi_L(t)|^2|\varphi_Z(t/g)|^{-2}\,dt\Big\}\Big\{\int |L^Z * L^Z(u;g)|\,du\Big\} \\
=&o(n^2g^{j+2}),
\end{aligned}$$

where we used Lemma A.4 and Condition (C1). $\square$

Let $\widehat{f}_Y(\cdot;h)$ denote the usual kernel density estimator of f_Y based on the kernel K, and let $\widehat{f}_X(\cdot;h)$ be the deconvolving kernel density estimator of f_X based on the kernel K, and defined in (2.1).

Lemma *A.6.* Under Condition B, we have

$$\widehat{f}_Y(x;h) = \left(\widehat{f}_X(\cdot;h) * f_Z\right)(x).$$

Proof of Lemma A.6. If we use (B1), (B2) and the change of variable $v = x - u$, we find

$$\begin{aligned}
\left(\widehat{f}_X(\cdot;h) * f_Z\right)(x) =& \int \widehat{f}_X(u;h) f_Z(x-u)\,du \\
=& n^{-1} \sum_{1\le i\le n} \int K_h^Z(u - Y_i; h) f_Z(x-u)\,du \\
=& (2\pi nh)^{-1} \sum_{1\le i\le n} \int_{\mathcal{R}^2} e^{-it\frac{u-Y_i}{h}} \varphi_K(t) f_Z(x-u)/\varphi_Z(t/h)\,du\,dt \\
=& (2\pi nh)^{-1} \sum_{1\le i\le n} \int \left(\int e^{itv/h} f_Z(v)\,dv\right) e^{-it(x-Y_i)/h} \varphi_K(t)/\varphi_Z(t/h)\,dt \\
=& (2\pi nh)^{-1} \sum_{1\le i\le n} \int e^{-it(x-Y_i)/h} \varphi_K(t)\,dt \\
=& n^{-1} \sum_{1\le i\le n} K_h(x - Y_i) \\
=& \widehat{f}_Y(x;h),
\end{aligned}$$

where we used the Fourier inversion Theorem. $\square$

Let $\widehat{\varphi}_{Y,n}(t) = n^{-1}\sum_{j=1}^{n} e^{itY_j}$, denote the empirical characteristic function of Y. The following lemma summarizes some useful facts about this quantity.

Lemma *A.7.* For the empirical characteristic function we have

$$|\widehat{\varphi}_{Y,n}(t)|^2 = \frac{1}{n} + \frac{1}{n^2}\sum_{\substack{j,k=1\\ j\neq k}}^{n} e^{it(Y_j - Y_k)} \qquad \text{and} \qquad E|\widehat{\varphi}_{Y,n}(t)|^2 = |\varphi_Y(t)|^2 + O(n^{-1}).$$

Proof of Lemma A.7. For the squared modulus of the empirical characteristic function $\widehat{\varphi}_{Y,n}(t)$ we find

$$|\widehat{\varphi}_{Y,n}(t)|^2 = \widehat{\varphi}_{Y,n}(t)\overline{\widehat{\varphi}}_{Y,n}(t) = \frac{1}{n^2}\sum_{j,k=1}^{n} e^{it(Y_j - Y_k)} = \frac{1}{n} + \frac{1}{n^2}\sum_{\substack{j,k=1\\ j\neq k}}^{n} e^{it(Y_j - Y_k)}.$$

Note however that this quantity is of course real-valued and hence equals $n^{-1}+n^{-2}\sum_{j\neq k}\cos(t(Y_j-Y_k))$. We prefer to keep its complex writing since it will help us in the following. Taking the expectation we find

$$\begin{aligned}
E|\widehat{\varphi}_{Y,n}(t)|^2 =&\frac{1}{n}+\frac{1}{n^2}\sum_{\substack{j,k=1\\ j\neq k}}^{n} E[e^{it(Y_j-Y_k)}]\\
=&\frac{1}{n}+\frac{1}{n^2}\sum_{\substack{j,k=1\\ j\neq k}}^{n} \mathrm{E}[e^{itY_j}]E[e^{-itY_k}]\\
=&\frac{1}{n}+\frac{n(n-1)}{n^2}\varphi_Y(t)\overline{\varphi_Y(t)}\\
=&|\varphi_Y(t)|^2+n^{-1}(1-|\varphi_Y(t)|^2)=|\varphi_Y(t)|^2+O(n^{-1}).
\end{aligned}$$

□

A.2 Proofs of Propositions 2 and 3

A.2.1 Proof of Proposition 2

To deal with the integrated variance, first note that under (A3) and (B2) and after basic manipulations, the integrated variance of the deconvolving kernel density estimator may be written as follows (see for example Stefanski and Carroll (1990))

$$\int \mathrm{Var}\{\widehat{f}_X(t;h)\}\,dt=\frac{1}{2\pi nh}\int|\varphi_K(t)|^2|\varphi_Z(t/h)|^{-2}\,dt+\frac{1}{2\pi n}\int|\varphi_K(ht)|^2|\varphi_X(t)|^2\,dt, \tag{A.8}$$

and its bootstrap estimator as

$$\int \mathrm{Var}^*\{\widehat{f}^*_X(t;h)\}\,dt=\frac{1}{2\pi nh}\int|\varphi_K(t)|^2|\varphi_Z(t/h)|^{-2}\,dt+\frac{1}{2\pi n}\int|\varphi_K(ht)|^2|\widehat{\varphi}_{X,g}(t)|^2\,dt. \tag{A.9}$$

In what follows we will calculate the expectation and the variance of this bootstrap estimator. For simplicity we introduce the shortcut notation $\mathrm{Var}^*(t;h)$ for $\mathrm{Var}^*\{\widehat{f}^*_X(t;h)\}$.

Proof of statement (i):
From (A.9) we find

$$\mathrm{E}\Big[\int\mathrm{Var}^*(t;h)\,dt\Big]=\frac{1}{2\pi nh}\int|\varphi_K(t)|^2|\varphi_Z(t/h)|^{-2}\,dt+\frac{1}{2\pi n}\int|\varphi_K(ht)|^2\,\mathrm{E}\left\{|\widehat{\varphi}_{X,g}(t)|^2\right\}\,dt.$$

Recall from Remark 2 that $\widehat{\varphi}_{X,g}(t)=\widehat{\varphi}_{Y,n}(t).\varphi_L(gt)/\varphi_Z(t)$, and hence

$$\mathrm{E}\left[|\widehat{\varphi}_{X,g}(t)|^2\right]=\mathrm{E}\left[|\widehat{\varphi}_{Y,n}(t)|^2\right].|\varphi_L(gt)|^2|\varphi_Z(t)|^{-2}$$

$$=|\varphi_X(t)|^2|\varphi_L(gt)|^2 + |\varphi_L(gt)|^2|\varphi_Z(t)|^{-2}O\left(n^{-1}\right),$$

where we used Lemma A.7 and the fact that $\varphi_Y(t) = \varphi_X(t).\varphi_Z(t)$.
Therefore we can write

$$\begin{aligned}\mathrm{E}\Big[\int \mathrm{Var}^*(t;h)\,dt\Big] =& \frac{1}{2\pi nh}\int |\varphi_K(t)|^2|\varphi_Z(t/h)|^{-2}\,dt + \frac{1}{2\pi nh}\int |\varphi_K(t)|^2|\varphi_X(t/h)|^2|\varphi_L(gt/h)|^2\,dt\\ &+ O\left(n^{-2}h^{-1}\right)\int |\varphi_K(t)|^2|\varphi_L(gt/h)|^2|\varphi_Z(t/h)|^{-2}\,dt,\end{aligned}$$

where, by (A3), the behaviour of the second term is provided by

$$n^{-1}\int |\varphi_K(ht)|^2|\varphi_X(t)|^2|\varphi_L(gt)|^2\,dt \le n^{-1}\int |\varphi_X(t)|^2\,dt = O(n^{-1}),$$

and, by (C3), the behaviour of the third term is described by

$$\begin{aligned}n^{-2}h^{-1}\int |\varphi_K(t)|^2|\varphi_L(gt/h)|^2|\varphi_Z(t/h)|^{-2}\,dt \le& (n^2h)^{-1}\int |\varphi_K(t)|^2|\varphi_Z(t/h)|^{-2}\,dt\\ =& o(n^{-1}).\end{aligned}$$

Using (A.8) completes the proof of statement (i).

Proof of statement (ii):
Note first that from (A.9), we get $\mathrm{Var}\left[\int \mathrm{Var}^*\{\widehat{f}^*_X(t;h)\}\,dt\right] = (4\pi^2n^2)^{-1}\,\mathrm{Var}\left[\int |\varphi_K(ht)|^2|\widehat{\varphi}_{X,g}(t)|^2\,dt\right]$, and hence the task is to compute the latter variance term.

Let $\varphi_K^Z(t)$ denote $|\varphi_L(gt)|^2|\varphi_Z(t)|^{-2}|\varphi_K(ht)|^2$, and note that this is an even function. Then by Remark 2 and Lemma A.7, we have

$$\begin{aligned}\mathrm{Var}\Big[\int |\varphi_K(ht)|^2|\widehat{\varphi}_{X,g}(t)|^2|\,dt\Big] =& \mathrm{Var}\Big[n^{-2}\sum_{\substack{j,k=1\\ j\neq k}}^{n}\int e^{it(Y_j-Y_k)}\varphi_K^Z(t)\,dt\Big]\\ =& n^{-4}\sum_{\substack{j,k=1\\ j\neq k}}^{n}\ \sum_{\substack{j',k'=1\\ j'\neq k'}}^{n}\mathrm{Cov}\Big[\int e^{it(Y_j-Y_k)}\varphi_K^Z(t)\,dt, \int e^{it(Y_{j'}-Y_{k'})}\varphi_K^Z(t)\,dt\Big]\\ =& 2n^{-3}(n-1)\,\mathrm{Var}\Big[\int e^{it(Y_1-Y_2)}\varphi_K^Z(t)dt\Big]\\ &+ 4n^{-3}(n-1)(n-2)\,\mathrm{Cov}\Big[\int e^{it(Y_1-Y_2)}\varphi_K^Z(t)\,dt, \int e^{it(Y_1-Y_3)}\varphi_K^Z(t)\,dt\Big].\end{aligned} \tag{A.10}$$

We will now compute separately the two terms appearing on the right-hand side of (A.10).

For the first term, let $T = \int e^{it(Y_1-Y_2)}\varphi_K^Z(t)\,dt$. Then $\mathrm{Var}(T) \le \mathrm{E}(T^2)$, where under (C3)

$$\mathrm{E}(T^2) = \mathrm{E}\int_{\mathcal{R}^2} e^{i(t+u)Y_1}e^{-i(t+u)Y_2}|\varphi_L(gt)|^2|\varphi_Z(t)|^{-2}|\varphi_K(ht)|^2|\varphi_L(gu)|^2|\varphi_Z(u)|^{-2}|\varphi_K(hu)|^2\,dt\,du$$

$$
\begin{aligned}
&= \int_{\mathcal{R}^2} |\varphi_Y(t+u)|^2 |\varphi_Z(t)|^{-2} |\varphi_Z(u)|^{-2} |\varphi_L(gt)|^2 |\varphi_K(ht)|^2 |\varphi_L(gu)|^2 |\varphi_K(hu)|^2 \, dt\, du \\
&\leq \int_{\mathcal{R}^2} |\varphi_Z(t)|^{-2} |\varphi_Z(u)|^{-2} |\varphi_K(ht)|^2 |\varphi_K(hu)|^2 \, dt\, du \\
&= o(n^2). \qquad \text{(A.11)}
\end{aligned}
$$

For the second term, let T be defined as above and put $U = \int e^{iu(Y_1 - Y_3)} \varphi_K^Z(u)\, du$. We then have

$$
\begin{aligned}
|\,\mathrm{E}(TU)| &= \Big| \int_{\mathcal{R}^2} \varphi_Y(t+u) \bar{\varphi}_Y(t) \bar{\varphi}_Y(u) \varphi_K^Z(t) \varphi_K^Z(u) \, dt\, du \Big| \\
&\leq \int_{\mathcal{R}^2} |\varphi_Y(t+u)| |\varphi_Y(t)| |\varphi_Y(u)| |\varphi_Z(t)|^{-2} |\varphi_K(ht)|^2 |\varphi_Z(u)|^{-2} |\varphi_K(hu)|^2 \, dt\, du \\
&= \int_{\mathcal{R}^2} |\varphi_Y(t+u)| |\varphi_X(t)| |\varphi_X(u)| |\varphi_Z(u)|^{-1} |\varphi_Z(t)|^{-1} |\varphi_K(ht)|^2 |\varphi_K(hu)|^2 \, dt\, du \\
&\leq \int_{\mathcal{R}^2} |\varphi_X(t)| |\varphi_X(u)| |\varphi_Z(u)|^{-1} |\varphi_Z(t)|^{-1} |\varphi_K(ht)| |\varphi_K(hu)| \, dt\, du
\end{aligned}
$$

and this term is of order $o(n)$ since by Cauchy-Schwartz and applying Conditions (A3) and (C3) we have

$$
\int |\varphi_X(t)| |\varphi_Z(t)|^{-1} |\varphi_K(ht)| \, dt \leq \Big[\int |\varphi_X(t)|^2 \, dt \Big]^{1/2} \Big[\int |\varphi_Z(t)|^{-2} |\varphi_K(ht)|^2 \, dt \Big]^{1/2} = o(\sqrt{n}).
$$

For the expectation of the random quantity T we get

$$
\begin{aligned}
E(T) &= \int_{\mathcal{R}^2} e^{it(y-z)} \varphi_K^Z(t) f_Y(y) f_Y(z) dt\, dy\, dz \\
&= \int \varphi_Y(t) \varphi_Y(-t) \varphi_K^Z(t) \, dt \\
&= \int |\varphi_Y(t)|^2 \varphi_K^Z(t) \, dt \\
&= \int |\varphi_X(t)|^2 |\varphi_L(gt)|^2 |\varphi_K(ht)|^2 \, dt \\
&\leq \int |\varphi_X(t)|^2 \, dt,
\end{aligned}
$$

which is of order $O(1)$ by Condition (A3). Similarly for $E(U)$.
Therefore the covariance term $\mathrm{Cov}(T, U) = E(TU) - E(T)E(U)$ is of order $o(n)$. Substituting this finding and (A.11) into (A.10) we get the result. □

A.2.2 Proof of Proposition 3

Some arguments of this proof are similar to those used by Jones, Marron and Park (1991) for proving their results. The main technical difficulties here come from the fact that we have to deal with contaminated data.

Recall the definition of $D(\cdot)$ in (A.1), and denote by $D_h(\cdot) = \frac{1}{h}D(\cdot/h)$ its usual rescaled version. It is easy to verify that $D_h(x)$ is equal to $K_h(x) - \delta_0(x)$.

Using Lemma A.1 and Condition B, we can rewrite the integrated squared bias as follows

$$\begin{aligned}
\mathrm{ISB}(h) &= \int \left\{ \mathrm{E}\left[\widehat{f}_X(x;h)\right] - f_X(x)\right\}^2 dx \\
&= \int \left\{ \mathrm{E}\left[K_h^Z(x-Y;h)\right] - f_X(x)\right\}^2 dx \\
&= \int \left\{ \int K_h(x-y) f_X(y)\, dy - f_X(x)\right\}^2 dx \\
&= \int \left\{ \int D_h(x-y) f_X(y)\, dy \right\}^2 dx \\
&= \int_{\mathcal{R}^3} D_h(x-y) f_X(y) D_h(x-z) f_X(z)\, dx\, dy\, dz \\
&= \int_{\mathcal{R}^3} D_h(u) f_X(y) D_h(u-(z-y)) f_X(z)\, du\, dy\, dz \\
&= \int_{\mathcal{R}^2} D_h * D_h(z-y) f_X(y) f_X(z)\, dy\, dz \\
&= \int \left(D_h * D_h * f_X\right)(z) f_X(z)\, dz, \qquad \text{(A.12)}
\end{aligned}$$

where we applied the change of variable $u = x - y$.
In what follows, we use the shortcut notation $D_h * D_h * f_X(z)$ to denote $\left(D_h * D_h * f_X\right)(z)$ and similar expressions. Recall that, under Condition B, $L^Z(\cdot; g)$ is symmetric (see Remark 3). The bootstrap estimator of the integrated squared bias can then be written as

$$\begin{aligned}
\mathrm{ISB}^*(h) &= \int \left(D_h * D_h * \widehat{f}_X(\cdot;g)\right)(z) \widehat{f}_X(z;g)\, dz \\
&= n^{-2} \sum_{i,j=1}^{n} \int D_h * D_h * L_g^Z(z-Y_j;g) L_g^Z(z-Y_i;g)\, dz \\
&= n^{-2} \sum_{i,j=1}^{n} \int D_h * D_h * L_g^Z(u;g) L_g^Z(u-Y_i+Y_j;g)\, du \\
&= n^{-2} \sum_{i,j=1}^{n} \int D_h * D_h * L_g^Z(u;g) L_g^Z(Y_i-Y_j-u;g)\, du \\
&= n^{-2} \sum_{i,j=1}^{n} D_h * D_h * L_g^Z * L_g^Z(Y_i-Y_j;g) \\
&= n^{-1} D_h * D_h * L_g^Z * L_g^Z(0;g) + n^{-2} \sum_{\substack{i,j=1 \\ j\neq i}}^{n} T_{ij},
\end{aligned}$$

where we used a change of variable $u = z - Y_j$, and introduced the notation

$$T_{ij} = \int D_h * D_h * L_g^Z(z - Y_j; g) L_g^Z(z - Y_i; g)\, dz = (D_h * D_h * L_g^Z * L_g^Z)(Y_i - Y_j; g). \tag{A.13}$$

Statement (i) of Proposition 3 is an immediate consequence of the following lemma, the proof of which is given below.

Lemma 1. *Under Condition B,*

(i) *if (A1) and (A2) are satisfied, we have* $\mathrm{E}(T_{12}) = \mathrm{ISB}(h) + O(g^2 h^4)$.

(ii) *if (C4) is satisfied, we have* $D_h * D_h * L_g^Z * L_g^Z(0; g) = o(nh^4)$.

To compute the variance of the bootstrap ISB, note first that

$$\begin{aligned}
\mathrm{Var}[\mathrm{ISB}^*(h)] =& n^{-4}\,\mathrm{Var}\Big[\sum_{\substack{i,j=1\\ j\neq i}}^{n} T_{ij}\Big]\\
=& n^{-4}\,\mathrm{Cov}\Big[\sum_{\substack{i,j=1\\ j\neq i}}^{n} T_{ij}, \sum_{\substack{i',j'=1\\ j'\neq i'}}^{n} T_{i'j'}\Big]\\
=& n^{-4}\big[2n(n-1)\,\mathrm{Var}(T_{12}) + 4n(n-1)(n-2)\,\mathrm{Cov}\,(T_{12}, T_{13})\big].
\end{aligned} \tag{A.14}$$

The behaviour of the variance and covariance term in this expression is established in Lemma 2 below, and hence the proof of statement (ii) of Proposition 3 is completed with the proof of that lemma.

Lemma 2. *Under Conditions (A1) and B,*

(i) *if (C1) is satisfied we have*

$$\mathrm{Var}(T_{12}) \leq \mathrm{E}(T_{12}^2) = o(n^2 h^8 g)$$

(ii) *if (A2) and (C2) are satisfied we have*

$$\mathrm{Cov}(T_{12}, T_{13}) = o(nh^8).$$

Proof of Lemma 1. *Proof of statement* (i):
Using Lemma A.1, the symmetry of L, and introducing the notation $(L*L)_g(\cdot)$ for the usual rescaled version of $L*L$ we can write

$$\begin{aligned}
\mathrm{E}(T_{12}) =& \mathrm{E}\int D_h * D_h * L_g^Z(z-Y_2;g)L_g^Z(z-Y_1;g)\,dz \\
=& \mathrm{E}\int D_h * D_h * L_g(z-X_2)L_g(z-X_1)\,dz \\
=& \mathrm{E}\left\{D_h * D_h * L_g * L_g(X_1-X_2)\right\} \\
=& \int_{\mathcal{R}^2} D_h * D_h * L_g * L_g(x-y)f_X(x)f_X(y)\,dx\,dy \\
=& \int_{\mathcal{R}^3} D_h * D_h(x-y-u)L_g * L_g(u)f_X(x)f_X(y)\,dx\,dy\,du \\
=& \int_{\mathcal{R}^3} D_h * D_h(x-v)L_g * L_g(v-y)f_X(x)f_X(y)\,dx\,dy\,dv \\
=& \int_{\mathcal{R}^3} D_h * D_h(x-v)f_X(y)(L*L)_g(v-y)f_X(x)\,dx\,dy\,dv \\
=& \int_{\mathcal{R}^3} D_h * D_h(x-v)f_X(v-gw)L*L(w)f_X(x)\,dx\,dw\,dv \\
=& \int_{\mathcal{R}^2} D_h * D_h(x-v)f_X(v)f_X(x)\,dx\,dv \\
&+ g^2\int_{\mathcal{R}^3}\int_0^1 D_h * D_h(x-v)L*L(w)w^2(1-t)f_X''(v-gtw)f_X(x)\,dt\,dw\,dx\,dv \\
=& \mathrm{ISB}(h) + g^2\int_{\mathcal{R}^3}\int_0^1 (D*D)_h(x-v)L*L(w)w^2(1-t)f_X''(v-gtw)f_X(x)\,dt\,dw\,dx\,dv \\
\equiv& \mathrm{ISB}(h) + g^2.(I), \qquad\qquad (A.15)
\end{aligned}$$

where we used a first change of variable $v = y + u$ and a second change of variable $w = (v-y)/g$, followed by a first order Taylor expansion of f_X around v and an application of Lemma A.3.

We will now show that the term (I) in expression (A.15) is of order $O(h^4)$. To see this, use the change of variable $z = (x-v)/h$, a third order Taylor expansion of f_X around v, and get

$$\begin{aligned}
(I) =& \Big|\int_{\mathcal{R}^3}\int_0^1 D*D(z)L*L(w)w^2(1-t)f_X''(v-gtw)f_X(v+hz)\,dt\,dw\,dz\,dv\Big| \\
=& \frac{h^4}{3!}\Big|\int_{\mathcal{R}^3}\int_0^1\int_0^1 D*D(z)z^4L*L(w)w^2(1-t)f_X''(v-gtw)(1-s)^3f_X^{(4)}(v+zsh)\,ds\,dt\,dw\,dz\,dv\Big| \\
\leq& cst.h^4\int_{\mathcal{R}^3}\int_0^1\int_0^1 |D*D(z)|\,z^4\,|L*L(w)|w^2(1-t)(1-s)^3|f_X''(v-gtw)|\,ds\,dt\,dw\,dz\,dv \\
=& O(h^4),
\end{aligned}$$

where we used subsequently Lemma A.3, Condition (A1), Conditions (A2) and (B1) and Lemma A.2.

Proof of statement (*ii*):
First note that using the notations in (A.1) and applying Condition (B2) we have

$$D_h * D_h * L_g^Z * L_g^Z(0; g) = (D * D)_h * (L^Z * L^Z)_g(0; g) = a_h * b_g(0).$$

Now we can write

$$\begin{aligned}
a_h * b_g(x) &= \int a_h(x-y) b_g(y)\, dy \\
&= \int a(u) b_g(x - hu)\, du \\
&= \frac{h^4}{3!} \iint_0^1 a(u) u^4 (1-t)^3 b_g^{(4)}(x - thu)\, dt\, du \\
&= \frac{h^4 g^{-5}}{3!} \iint_0^1 a(u) u^4 (1-t)^3 b^{(4)} \left(\frac{x - thu}{g} \right) dt\, du,
\end{aligned}$$

where we used subsequently the change of variable $u = (x - y)/h$, a third order Taylor expansion of b_g around x and Lemma A.3.

Finally from Lemmas A.4 and A.2 and using (C4), we obtain

$$|a_h * b_g(0)| \leq \frac{h^4 g^{-5}}{3! 2\pi} \int |a(u)| u^4\, du \int_0^1 |1-t|^3\, dt \int v^4 |\varphi_L(v)|^2 |\varphi_Z(v/g)|^{-2}\, dv = o(nh^4),$$

which proves the statement. □

Proof of Lemma 2. *Proof of statement* (*i*):
Note first that, by Definition (A.13) and the notations in (A.1), $E(T_{12}^2) = \mathrm{E}[a_h * b_g(Y_1 - Y_2).a_h * b_g(Y_1 - Y_2)]$. Now, by the changes of variables $s = (x - y - z)/h$, $t = z/g$ and $w = (hs + gt - u)/h$, we have

$$\begin{aligned}
&\mathrm{E}[a_h * b_g(Y_1 - Y_2).a_h * b_g(Y_1 - Y_2)] \\
&\quad = \int_{\mathcal{R}^4} a_h(x - y - z) b_g(z) a_h(x - y - u) b_g(u) f_Y(x) f_Y(y)\, dx\, dy\, dz\, du \\
&\quad = \int_{\mathcal{R}^4} a(s) b(t) a_h(hs + gt - u) b_g(u) f_Y(sh + y + gt) f_Y(y)\, ds\, dy\, dt\, du \\
&\quad = \int_{\mathcal{R}^4} a(s) b(t) a(w) b_g(hs + gt - wh) f_Y(sh + y + gt) f_Y(y)\, ds\, dy\, dt\, dw \\
&\quad = g^{-1} \int_{\mathcal{R}^4} a(s) b(t) a(w) b(t + h(s - w)/g) f_Y(y + sh + gt) f_Y(y)\, ds\, dy\, dt\, dw. \quad \text{(A.16)}
\end{aligned}$$

Using a seventh order Taylor expansion of the function $b(\cdot)$ around t, we can write

$$b(t+h(s-w)/g)=\sum_{k=0}^{7}\frac{h^k g^{-k}}{k!}(s-w)^k b^{(k)}(t)+\frac{h^8 g^{-8}}{7!}(s-w)^8\int_0^1(1-\theta)^7 b^{(8)}(t+h\theta(s-w)/g)\,d\theta,$$

where by the binomial expansion, $(s-w)^k=\sum_{j=0}^{k}C_k^j(-1)^{k-j}s^j w^{k-j}$. Combining this with expression (A.16) and using Lemma A.3, we find

$$\begin{aligned}
&\mathrm{E}[a_h * b_g(Y_1-Y_2).a_h * b_g(Y_1-Y_2)]\\
&=\sum_{k=4}^{7}\sum_{j=0}^{k-4}\frac{h^k g^{-k-1}}{k!}C_k^j(-1)^{k-j}\mu_{k-j}(a)\int_{\mathcal{R}^3}a(s)b(t)b^{(k)}(t)s^j f_Y(y+sh+gt)f_Y(y)\,ds\,dy\,dt\\
&+\frac{h^8 g^{-9}}{7!}\int_{\mathcal{R}^4}\int_0^1 a(s)b(t)a(w)(s-w)^8(1-\theta)^7 b^{(8)}(t+\theta h(s-w)/g)f_Y(y+sh+gt)f_Y(y)\,d\theta ds\,dy\,dt\,dw\\
&=\sum_{k=5}^{7}\sum_{\substack{1\le j\le k-4\\4-j\le\ell\le3}}h^{k+\ell}g^{-k-1}\frac{C_k^j(-1)^{k-j}}{k!\ell!}\mu_{k-j}(a)\mu_{\ell+j}(a)\int_{\mathcal{R}^2}b(t)b^{(k)}(t)f_Y(y)f_Y^{(\ell)}(y+gt)\,dy\,dt\\
&+\sum_{\substack{4\le k\le7\\0\le j\le k-4}}\frac{h^{k+4}C_k^j(-1)^{k-j}}{g^{k+1}k!3!}\mu_{k-j}(a)\int_{\mathcal{R}^3}\int_0^1 a(s)b(t)b^{(k)}(t)s^j s^4(1-\xi)^3 f_Y^{(4)}(y+gt+\xi hs)f_Y(y)\,d\xi ds\,dy\,dt\\
&+\sum_{0\le\ell\le3}\frac{h^{8+\ell}g^{-9}}{7!\ell!}\int_{\mathcal{R}^4}\int_0^1 a(s)b(t)a(w)(s-w)^8 s^\ell(1-\theta)^7 b^{(8)}(t+\theta h(s-w)/g)f_Y^{(\ell)}(y+gt)f_Y(y)\,d\theta dw\,ds\,dy\,dt\\
&+\frac{h^{12}g^{-9}}{7!3!}\int_{\mathcal{R}^4}\int_0^1\int_0^1 a(s)b(t)a(w)s^4(s-w)^8(1-\theta)^7 b^{(8)}(t+\theta h(s-w)/g)(1-\xi)^3 f_Y^{(4)}(y+gt+\xi hs)\\
&\qquad\times f_Y(y)\,d\xi d\theta dw\,ds\,dy\,dt,
\end{aligned}\tag{A.17}$$

where we used a third order Taylor expansion of f_Y around $y+gt$ and applied Lemma A.3.

By using Lemmas A.5 and A.2 and Condition (A1), we can now conclude that

$$\begin{aligned}
&|\,\mathrm{E}[a_h * b_g(Y_1-Y_2).a_h * b_g(Y_1-Y_2)]|\\
&\le\sum_{k=5}^{7}\sum_{\substack{1\le j\le k-4\\4-j\le\ell\le3}}o(n^2h^{k+\ell}g^{k+2-k-1})+\sum_{\substack{4\le k\le7\\0\le j\le k-4}}o(n^2h^{k+4}g^{k+2-k-1})\iint_0^1|a(s)||s|^{j+4}\,ds\\
&\quad+o(n^2h^8g)\int_{\mathcal{R}^2}|a(s)||a(w)|(s-w)^8\,dw\,ds+o(n^2h^{12}g)\int_{\mathcal{R}^2}|a(s)||a(w)|s^4(s-w)^8\,dw\,ds\\
&=o(n^2h^8g),
\end{aligned}$$

which proves the statement.

Proof of statement (ii):
In order to evaluate $\mathrm{Cov}(T_{12},T_{13})=E(T_{12}T_{13})-E(T_{12})E(T_{13})$ we first investigate the

expectation of the product term. It should be understood that all operations below are carried out after having written the expectations as integrals, so the various steps should be understood as such.

Using Condition B, the notation from (A.1), a first change of variables $s' = x + Y_2$, $t' = z + Y_3$, Lemma A.1, a second set of changes of variables $u = (x - s')/h$, $v = (x - t')/h$, $s = (s' - y)/g$ and $t = (t' - z)/g$, we find

$$
\begin{aligned}
\mathrm{E}[T_{12}T_{13}] &= \mathrm{E}[a_h * b_g(Y_1 - Y_2).a_h * b_g(Y_1 - Y_3)] \\
&= \mathrm{E}\Big[\int_{\mathcal{R}^2} a_h(Y_1 - Y_2 - x)b_g(x)a_h(Y_1 - Y_3 - z)b_g(z)\,dx\,dz\Big] \\
&= \int_{\mathcal{R}^2} \mathrm{E}[a_h(Y_1 - s')a_h(Y_1 - t')]\,\mathrm{E}[b_g(s' - Y_2)]\,\mathrm{E}[b_g(t' - Y_3)]\,ds'\,dt' \\
&= \int_{\mathcal{R}^2} \mathrm{E}[a_h(Y_1 - s')a_h(Y_1 - t')]\,\mathrm{E}[B_g(s' - X_2)]\,\mathrm{E}[B_g(t' - X_3)]\,ds'\,dt' \\
&= \int_{\mathcal{R}^5} a_h(x - s')a_h(x - t')B_g(s' - y)B_g(t' - z)f_Y(x)f_X(y)f_X(z)\,ds'\,dt'\,dx\,dy\,dz \\
&= \int_{\mathcal{R}^5} a(u)B(s)a(v)B(t)f_Y(x)f_X(x - hu - gs)f_X(x - hv - gt)\,ds\,dt\,dx\,du\,dv \\
&= \frac{h^4}{3!}\int_{\mathcal{R}^5}\int_0^1 u^4a(u)B(s)a(v)B(t)f_Y(x)(1-\theta)^3 f_X^{(4)}(x - gs - \theta hu)f_X(x - hv - gt)\,d\theta\,ds\,dt\,dx\,du\,dv \\
&= \frac{h^8}{3!3!}\int_{\mathcal{R}^5}\int_0^1\int_0^1 u^4a(u)B(s)a(v)B(t)f_Y(x)v^4(1-\theta)^3 f_X^{(4)}(x - gs - \theta hu) \\
&\qquad\qquad \times (1-\xi)^3 f_X^{(4)}(x - gt - \xi hv)\,d\xi\,d\theta\,ds\,dt\,dx\,du\,dv,
\end{aligned}
$$

where we applied subsequently a third order Taylor expansion of f_X around $x - gs$, and a third order Taylor expansion of f_X around $x - gt$ and used Condition (A1) and Lemma A.3.

Using Condition (A1) and Lemma A.2 we then find

$$
|\,\mathrm{E}[T_{12}T_{13}]| \le M\frac{h^8}{3!3!}\Big\{\int |u^4a(u)|\,du\Big\}^2\Big\{\int |B(s)|\,ds\Big\}^2\int f_Y(x)\,dx = o(nh^8),
$$

where we also used Condition (C2), and where M is a positive constant. Using Lemma A.1 and Proposition 1 (i) it is clear that $\mathrm{E}(T_{ij}) = \mathrm{E}(T_{ik}) = O(h^4)$. This then proves that $\mathrm{Cov}(T_{12}, T_{13}) = o(nh^8)$. □

A.3 Proof of Theorem 3

The proof of this result needs to be done with the necessary mathematical rigour.
Denote by Ω the set on which all random variables are defined. Then $\forall\omega \in \Omega$ we define $h_n^*(\omega) = \mathrm{argmin}_{h\in[A_1h_n, A_2h_n]}\,\mathrm{MISE}^*(h,\omega)$. If we suppose that, $\forall\omega$, $\mathrm{MISE}^*(\cdot,\omega)$ has a single global minimum, we have for all $h \neq h_n^*(\omega)$, $\mathrm{MISE}^*(h_n^*(\omega),\omega) < \mathrm{MISE}^*(h,\omega)$.

Similarly if we suppose that MISE has a single global minimum we have for all $h \neq h_n$, $\mathrm{MISE}(h_n) < \mathrm{MISE}(h)$. Let $A_n = \{\omega : h_n^*(\omega) \neq h_n\}$. Then $\forall n \in I\!N$ and $\forall \omega \in A_n$, we have

$$\mathrm{MISE}^*(h_n^*(\omega), \omega)/\mathrm{MISE}^*(h_n, \omega) < 1 \quad \text{and} \quad \mathrm{MISE}(h_n^*(\omega))/\mathrm{MISE}(h_n) > 1, \qquad (A.18)$$

where we suppose that the second inequality remains strict when $n \to \infty$. Thus $\exists \delta_1 \geq 0$ and $\delta_2 > 0$ such that for all $n \in I\!N$ and $\omega \in A_n$,

$$\mathrm{MISE}^*(h_n^*(\omega), \omega)/\mathrm{MISE}^*(h_n, \omega) \leq 1 - \delta_1 \quad \text{and} \quad \mathrm{MISE}(h_n^*(\omega))/\mathrm{MISE}(h_n) \geq 1 + \delta_2.$$

Now for $\varepsilon_1 > 0$ define $B_n = \{\omega : |\mathrm{MISE}(h_n)/\mathrm{MISE}^*(h_n, \omega) - 1| > \varepsilon_1\}$, and for $\varepsilon_2 > 0$, let $D_n = \{\omega : |\mathrm{MISE}^*(h_n^*(\omega), \omega)/\mathrm{MISE}(h_n^*(\omega)) - 1| > \varepsilon_2\}$.
Since by Theorem 2 $\mathrm{MISE}^*(h_n)/\mathrm{MISE}(h_n) \xrightarrow{P} 1$ and $\mathrm{MISE}^*(h_n^*)/\mathrm{MISE}(h_n^*) \xrightarrow{P} 1$, we have

$$\forall \varepsilon_1, \varepsilon_2 > 0, P(B_n) \to 0 \text{ and } P(D_n) \to 0, \quad \text{as } n \to \infty, \text{ i.e.}$$

$$\begin{aligned} &\forall \varepsilon_1 > 0, \eta_1 > 0, \exists N_1 \in I\!N, \text{ such that } \forall n \geq N_1, P(B_n^C) > 1 - \eta_1 \\ \text{and} \quad &\forall \varepsilon_2 > 0, \eta_2 > 0, \exists N_2 \in I\!N, \text{ such that } \forall n \geq N_2, P(D_n^C) > 1 - \eta_2. \end{aligned}$$

The proof now is proceeded by contradiction. Suppose that $h_n^*/h_n \overset{P}{\nrightarrow} 1$, as $n \to \infty$. This then would mean that

$$\exists \varepsilon > 0, \exists \eta > 0 \text{ such that } \forall n \in I\!N, \exists N_3 \geq n \ : \ P\left(\left|\frac{h_{N_3}^*}{h_{N_3}} - 1\right| > \varepsilon\right) \geq \eta. \qquad (A.19)$$

Without loss of generality we can assume that $\eta < 1$.

From (A.19) it follows that,

$$\exists \varepsilon > 0, \exists \eta_3 > 0 \text{ such that } \forall n \in I\!N, \exists N_3 \geq n \ : \ P\left(\left|\frac{h_{N_3}^*}{h_{N_3}} - 1\right| > \varepsilon\right) > \eta_3,$$

which implies that

$$\exists \eta_3 > 0 \text{ such that } \forall n \in I\!N, \exists N_3 \in I\!N \ : \ P\left(A_{N_3}\right) > \eta_3,$$

with A_n the set as defined before.

For this η_3 given, take $\eta_1 > 0$ and $\eta_2 > 0$ such that $2\eta_1 + \eta_2 < \eta_3$ (see below). Take also $\varepsilon_1 = \varepsilon_2 > 1 + \sqrt{\frac{1}{1+\delta_2}}$. Then, for $N = \max(N_1, N_2)$, consider the $N_3 \geq N$. For this N_3 we have

$$P(B_{N_3}^C) > 1 - \eta_1 \ , \qquad P(D_{N_3}^C) > 1 - \eta_2 \ , \qquad P(A_{N_3}) > \eta_3,$$

and hence

$$P\left(A_{N_3} \cap B_{N_3}^C \cap D_{N_3}^C\right)$$

$$
\begin{aligned}
&= P\left(A_{N_3} \cup B_{N_3}^C \cup D_{N_3}^C\right) + P(A_{N_3}) + P(B_{N_3}^C) + P(D_{N_3}^C) \\
&\qquad - P(A_{N_3} \cup B_{N_3}^C) - P(A_{N_3} \cup D_{N_3}^C) - P(B_{N_3}^C \cup D_{N_3}^C) \\
&\geq P(A_{N_3}) + 2P(B_{N_3}^C) + P(D_{N_3}^C) - 1 - 1 - 1 \\
&\geq \eta_3 - 2\eta_1 - \eta_2 \\
&> 0,
\end{aligned}
$$

if one has chosen η_1 and η_2 small enough such that $\eta_3 > 2\eta_1 + \eta_2$. Now for all $\omega \in A_{N_3} \cap B_{N_3}^C \cap D_{N_3}^C$ we have

$$
\begin{aligned}
1 - \delta_1 \geq \frac{\mathrm{MISE}^*(h_n^*(\omega), \omega)}{\mathrm{MISE}^*(h_n, \omega)} &= \frac{\mathrm{MISE}^*(h_n^*(\omega), \omega)}{\mathrm{MISE}(h_n^*(\omega))} \cdot \frac{\mathrm{MISE}(h_n^*(\omega))}{\mathrm{MISE}(h_n)} \cdot \frac{\mathrm{MISE}(h_n)}{\mathrm{MISE}^*(h_n, \omega)} \\
&\geq (1 - \varepsilon_2)(1 + \delta_2)(1 - \varepsilon_1) \\
&= (1 - \varepsilon_1)^2(1 + \delta_2) \\
&> 1,
\end{aligned}
$$

which leads to a contradiction. □

References

Barry, J. and Diggle, P. (1995). Choosing the smoothing parameter in a Fourier approach to nonparametric deconvolution of a density function. *Journal of Nonparametric Statistics*, **4**, 223–232.

Carroll, R.J. and Hall, P. (1988). Optimal rates of convergence for deconvolving a density. *Journal of the American Statistical Association*, **83**, 1184–1186.

Carroll, R.J., Ruppert, D. and Stefanski, L. (1995). *Measurement error in nonlinear models.* Chapman and Hall, London.

Clayton, D.G. (1992). Models for the analysis of cohort and case control studies with inaccurately measured exposures. In Dwyer, Feinleib, Lippert and Hoffmeister (Eds.), *Statistical Models for Longitudinal Studies of Health*, 301–331. Oxford University Press, New York.

Cordy, C.B. and Thomas, D.R. (1997). Deconvolution of a distribution function. *Journal of the American Statistical Association*, **92**, 1459–1465.

Delaigle, A. (1997). L'estimateur à noyau d'une densité de probabilité sur base d'échantillons avec erreurs de mesure. *Mémoire de licence (Diploma Thesis)*, Département de Mathématiques, Université catholique de Louvain, Belgium.

Delaigle, A. (1999). Bandwidth selection in kernel estimation of a density when the data are contaminated by errors. *Mémoire de DEA (Master thesis)*, Institut de Statistique, Université catholique de Louvain, Belgium.

Delaigle, A. and Gijbels, I. (2001a).Estimation of integrated squared density derivatives from a contaminated sample. *Discussion Paper* #0101, Institut de Statistique, Université catholique de Louvain, Belgium.

Delaigle, A. and Gijbels, I. (2001b). Comparison of several data-driven bandwidth selection procedures in kernel estimation of a density when the data are contaminated by random noise. *Manuscript*.

Falk, M. (1992). Bootstrap optimal bandwidth selection for kernel density estimates. *Journal of Statistical Planning and Inference*, **30**, 13–22.

Fan, J. (1991a). Asymptotic normality for deconvolution kernel density estimators. *Sankhya A*, **53**, 97–110.

Fan, J. (1991b). Global behaviour of deconvolution kernel estimates. *Statistica Sinica*, **1**, 541–551.

Fan, J. (1991c). On the optimal rates of convergence for nonparametric deconvolution problems. *The Annals of Statistics*, **19**, 1257–1272.

Fan, J. (1992). Deconvolution with supersmooth distributions. *The Canadian Journal of Statistics*, **20**, 155–169.

Faraway, J. and Jhun, M. (1990). Bootstrap choice of bandwidth for density estimation. *Journal of the American Statistical Association*, **85**, 1119–1122.

Grund, B. and Polzehl, J. (1997). Bias corrected bootstrap bandwidth selection. *Journal of Nonparametric Statistics*, **8**, 97–126.

Hall, P., Marron, J. and Park, B. (1992). Smoothed cross-validation. *Probability Theory and Related Fields*, **92**, 1–20.

Hazelton, M.L. (1999). An optimal local bandwidth selector for kernel density estimation. *Journal of Statistical Planning and Inference*, **77**, 37–50.

Hesse, C. (1999). Data-driven deconvolution. *Journal of Nonparametric Statistics*, **10**, 343–373.

Jones, M.C. (2000). Rough-and-ready assessment of the degree and importance of smoothing in functional estimation. *Statistica Neerlandica*, **54**, 37–46.

Jones, M.C., Marron, J. and Park, B. (1991). A simple root n bandwidth selector. derivatives. *The Annals of Statistics*, **19**, 1919–1932.

Jones, M.C., Marron, J. and Sheather, S.J. (1996). Progress in data-based bandwidth selection for kernel density estimation. *Computational Statistics*, **11**, 337–381.

Marron, J. (1992). Bootstrap bandwidth selection. In LePage and Billard (Eds.), *Exploring the limits of bootstrap*, 249–262. John Wiley and Sons.

Morris, J.N., Marr, J.W. and Clayton, D.G. (1977) Diet and heart: a postscript. *British Medical Journal*, **2**, 1307–1314.

Neumann, M.H. (1997). On the effect of estimating the error density in nonparametric deconvolution.*Journal of Nonparametric Statistics*, **7**, 307–330.

Rachdi, M. and Sabre, R. (2000). Consistent estimates of the mode of the probability density function in nonparametric deconvolution problems. *Statistics & Probability Letters*, **47**, 105–114.

Stefanski, L.A. (1990). Rates of convergence of some estimators in a class of deconvolution problems. *Statistics & Probability Letters*, **9**, 229–235.

Stefanski, L. and Carroll, R.J. (1990). Deconvoluting kernel density estimators. *Statistics*, **2**, 169–184.

Taylor, C. (1989). Bootstrap choice of the tuning parameter in kernel density estimation. *Biometrika*, **76**, 705–712.

Wand, M.P. (1998). Finite sample performance of deconvolving density estimators. *Statistics & Probability Letters*, **37**, 131–139.

Wand, M.P. and Jones, M.C. (1995). *Kernel Smoothing*. Chapman and Hall, London.