Practical bandwidth selection in deconvolution kernel density estimation

A. Delaigle and I. Gijbels [*]

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

Abstract

We consider kernel estimation of a density based on contaminated data and discuss the important issue of how to choose the bandwidth parameter in practice. We propose some plug-in type of bandwidth selectors, which are based on nonparametric estimation of an approximation of the mean integrated squared error. The selectors are a refinement of the simple normal reference bandwidth selector, which is obtained by parametrically estimating the approximated mean integrated squared error by referring to a normal density. In a simulation study we compare these plug-in bandwidth selectors with a bootstrap and a cross-validated bandwidth selector. We conclude that in finite samples, an appropriately chosen plug-in bandwidth selector and the bootstrap bandwidth selector perform comparably and both outperform the cross-validated bandwidth. We also illustrate the use of the various practical bandwidth selectors on a real data example.

Key words: bandwidth selection, bootstrap, cross-validation, deconvolution, errors-in-variables, kernel density estimation, plug-in methods.

[*] Corresponding author
  Email addresses: `delaigle@stat.ucl.ac.be` (A. Delaigle),

1 Introduction

In this paper we focus on the practical choice of the bandwidth in the context of contaminated data. Although the ideas behind this selection problem are close to those in the error-free case, its theoretical and practical study is quite different and far more complex than for the non-contaminated case, and has been studied by very few authors. For an overview of bandwidth selection procedures in the error-free case see for example Silverman (1986) and Wand and Jones (1995), among others, and the references therein. Papers dealing with practical bandwidth selection in the deconvolution problem are Stefanski and Carroll (1990) and Hesse (1999) who propose a cross-validation procedure, and Delaigle and Gijbels (2001) who investigate a bootstrap method. In this paper we introduce plug-in type of bandwidth selectors for contaminated data. We define the theoretical optimal bandwidth as the minimizer of the mean integrated squared error, and then consider an (asymptotic) approximation of this mean integrated squared error. The unknown quantities in this asymptotic expression can simply be estimated by making reference to a normal density (parametric estimation). This then leads to the simple normal reference bandwidth selector. The use of more sophisticated (nonparametric) estimators for the unknown quantities in the approximated mean integrated squared error leads to the more elaborated plug-in bandwidth selectors.

After having introduced the plug-in selection method we study the finite sample performances of various bandwidth selectors and their corresponding density estimates via a simulation study. We compare four procedures: the normal reference method, the more elaborated plug-in method, the cross-validation

gijbels@stat.ucl.ac.be (I. Gijbels).

method and the bootstrap method. We will see that in some cases the plug-in method brings considerable improvement to the normal reference method and that it competes with the bootstrap procedure. Both, the plug-in and the bootstrap bandwidth selectors, outperform the cross-validation method which, as in the error-free case, suffers from a large variability and multiplicity of solution.

This paper is organized as follows. In Section 2 we recall the definition of the deconvolving kernel density estimator as well as the distinction between ordinary smooth and supersmooth error distributions. The choice of the theoretical optimal bandwidth is also discussed. In Section 3 we propose three practical bandwidth selectors based on estimation of an (asymptotic) approximation of the mean integrated squared error: a normal reference, a plug-in and a solve-the-equation bandwidth selector. Section 4 briefly discusses the other available data-driven bandwidth selectors: the cross-validation method of Stefanski and Carroll (1990) and the bootstrap procedure of Delaigle and Gijbels (2001). The finite sample performances of the bandwidth selectors are explored and compared to each other in Section 5 via a simulation study. In Section 6 we illustrate the use of the bandwidth selection methods in a real data example.

2 Deconvolving kernel density estimator

2.1 The estimator

Suppose we want to estimate a continuous density f_X but we observe an i.i.d. sample $Y_1, \ldots, Y_n$ from the density f_Y, where

$$Y_i = X_i + Z_i \qquad i = 1, \ldots, n$$

and where for all i, Z_i is a random variable independent of X_i, of known continuous density f_Z, representing the error in the data, and X_i is a random variable of density f_X. In the case where f_Z is known up to some parameters, one may estimate these parameters through repeated measurements on several individuals. For a real data example see for example Delaigle and Gijbels (2001). The case where f_Z is totally unknown may also be considered. Such a problem necessitates further observations such as for example a sample from f_Z itself, and will not be studied here. See Barry and Diggle (1995) and Neumann (1997).

Consider a kernel function K and a smoothing parameter $h > 0$, depending on n, called the bandwidth. The deconvolving kernel estimator of f_X at $x \in \mathbb{R}$ is then 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 $K^Z(u; h) = (2\pi)^{-1} \int e^{-itu} \varphi_K(t)/\varphi_Z(t/h) \, dt$, with φ_L the Fourier transform of a function or a random variable L. See Carroll and Hall (1988) and Stefanski and Carroll (1990) for an introduction to this estimator. In order to guarantee that this estimator is well-defined we need to impose

4

that $\varphi_Z(t) \neq 0$ for all t, $\int |\varphi_X(t)|\, dt < \infty$, $\sup_t |\varphi_K(t)/\varphi_Z(t/h)| < \infty$, and $\int |\varphi_K(t)/\varphi_Z(t/h)|\, dt < \infty$.

Asymptotic properties of the deconvolving kernel estimator, such as consistency or rates of convergence to the target density have been studied in Carroll and Hall (1988), Devroye (1989), Stefanski (1990), Stefanski and Carroll (1990), Fan (1991a,b,c) and Fan (1992) among others. These properties depend strongly on the error distribution. Fan (1991a) distinguishes two categories of errors: the ordinary smooth and the supersmooth distributions. These two classes of error distributions differ in the way that their characteristic function $\varphi_Z(t)$ behaves in the tails. An ordinary smooth error distribution is characterized by a characteristic function that decays at a polynomial rate when $|t|$ tends to infinity. The characteristic function of a supersmooth error distribution on the other hand tends to zero at an exponentially fast rate when $|t|$ tends to infinity. Examples of supersmooth error distributions are the normal and Cauchy distributions and examples of ordinary smooth error distributions are the gamma and Laplace distributions. The rates of convergence of a nonparametric density estimator for f_X will differ strongly according to this type of smoothness of the error density. See also, for example, Fan and Koo (2002). For supersmooth errors this rate cannot be faster than logarithmic, whereas for ordinary smooth errors the rate of convergence of the estimator to f_X is of a much better polynomial rate. The deconvolving kernel estimator reaches those optimal rates. See Carroll and Hall (1988), Stefanski (1990), Stefanski and Carroll (1990), Fan (1991b,c) and Fan (1992).

In kernel density estimation for error-free data, the choice of the kernel function K has not a big influence on the quality of the estimator. In the error case however the particular structure of the estimator (2.1) requires some restric-

tions to be imposed on φ_K, the Fourier transform of K. Throughout this paper we will, for simplicity, use kernels with φ_K compactly supported although in the ordinary smooth error case, K could be taken with φ_K supported on $I\!\!R$ and such that φ_K decreases rapidly enough when $|t| \to \infty$ (see Fan (1991c)). This then immediately ensures the existence of all integrals, but also rules out kernels in the so-called Beta family such as Uniform, Epanechnikov, Biweight or Triweight kernels, which are yet the most commonly-used kernels in error-free kernel density estimation. We also impose that the kernel integrates to one (since then $\widehat{f}_X$ also integrates to one), but it may take negative values. Note that this does not say anything about the positiveness of the kernel density estimate, since this depends on the behaviour of K^Z. See (2.1). Some examples of kernels satisfying all these conditions can be found in, for example, Fan (1992) and Wand (1998).

2.2 *Theoretical optimal bandwidth*

As in usual kernel density estimation, the choice of the bandwidth h will strongly influence the shape of the estimator $\widehat{f}_X(.;h)$. In order to select an 'optimal' bandwidth, we need to choose a way to measure the distance between the estimator $\widehat{f}_X(.;h)$ and its target density f_X. A commonly-used criterion is the Mean Integrated Squared Error (MISE):

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

The optimal bandwidth, h_{MISE}, is then defined as the minimizer of this MISE quantity with respect to h.

This criterion however depends on unknown quantities involving f_X, and thus

is of no direct practical use for selecting the bandwidth. In the next section
we propose two estimators of an (asymptotic) approximated MISE and dis-
cuss practical bandwidth selectors from there on. In order to highlight the
dependency on h, we will in what follows write $\mathrm{MISE}\{\widehat{f}_X(\cdot; h)\}$ as $\mathrm{MISE}(h)$.

3 Plug-in type of bandwidth selection

Stefanski and Carroll (1990) provide an asymptotic expansion for the MISE.
Under some regularity assumptions, they prove that the asymptotic dominat-
ing term of the MISE is given by

$$\mathrm{AMISE}(h) = (2\pi nh)^{-1} \int |\varphi_K(t)|^2 . |\varphi_Z(t/h)|^{-2}\, dt + \frac{h^4}{4}\mu_{K,2}^2 R\left(f_X''\right) , \quad (3.2)$$

where, for any positive integer k, $\mu_{K,k} = \int x^k K(x)\, dx$, and for any square
integrable function g, $R(g) = \int g^2(x)\, dx$.

The main advantage of this asymptotic approximation over the exact MISE is
that it provides a rather simple expression, and the role of h can be evaluated
more easily. However, (3.2) involves the unknown quantity $R\left(f_X''\right)$, which has
to be estimated in order to propose practical bandwidth selectors. In the next
two sections we discuss two possible estimators for $R(f_X'')$, and obtain as such
an estimator $\widehat{\mathrm{AMISE}}(h)$. Application of our bandwidth selection procedures
in practice requires minimization of the resulting estimator of the asymptotic
mean integrated squared error in (3.2). In general we can only approximate
the solution numerically: on a discrete grid of h-values we evaluate $\widehat{\mathrm{AMISE}}(h)$
by numerical integration and select the h that minimizes $\widehat{\mathrm{AMISE}}(h)$ on that
grid. In some cases however (for example for most of the ordinary smooth
error densities), this expression simplifies and the minimizer can be found

7

more easily (for example analytically). See Delaigle (1999) for more details.

3.1 Normal reference bandwidth selection

A first elementary estimator of $R\left(f_X''\right)$ is provided by making reference to a normal distribution: one assumes that f_X is a normal $\mathrm{N}(\mu_X; \sigma_X^2)$ density which implies that $R(f_X'') = 0.375\,\sigma_X^{-5}\,\pi^{-1/2}$. The estimator $R(\widehat{f}_X'')$ is then defined by $R(\widehat{f}_X'') = 0.375\,\widehat{\sigma}_X^{-5}\,\pi^{-1/2}$, with $\widehat{\sigma}_X^2$ a consistent estimator of the variance of X. See for example Silverman (1986) for the error-free case. In the error case this variance can be estimated by, for example, $\widehat{\sigma}_X^2 = \widehat{\sigma}_Y^2 - \sigma_Z^2$, where $\widehat{\sigma}_Y$ is the sample standard deviation of the observations Y_i.

In general, when f_X is not a normal density, this estimator $R(\widehat{f}_X'')$ is not a consistent estimator of $R(f_X'')$, and the resulting bandwidth selector is not a consistent estimator of h_{MISE} either. Hence, although the order (rate) of the bandwidth selector and of the MISE estimator are not affected by this estimation step, a poor estimation of $R(f_X'')$ sometimes accounts for a bad selection of the bandwidth, resulting itself in a poor estimation of the density. This happens with densities f_X that possess particular non-normal features such as, for example, strong multimodality and/or asymmetry. See Section 5. Therefore we need a more elaborated procedure to estimate $R(f_X'')$. An appropriate nonparametric estimator for $R(f_X'')$ is discussed in the next section.

3.2 Plug-in bandwidth selection

Nonparametric estimation of $\theta_r = R(f_X^{(r)})$, with r any positive integer, when data are measured with errors, has been studied in detail by Delaigle and Gijbels (2002). They propose to estimate the quantity θ_r by $\widehat{\theta}_r = R(\widehat{f}_X^{(r)}(\cdot; h_r))$, where $\widehat{f}_X^{(r)}(\cdot; h_r)$ is the deconvolving kernel density estimator of the r-th derivative of f_X, defined by

$$\widehat{f}_X^{(r)}(x; h_r) = (nh_r^{r+1})^{-1} \sum_{i=1}^{n} K_Z^{(r)} \left(\frac{x - Y_i}{h_r}; h_r \right)$$

with $K_Z^{(r)}(x; h_r) = (2\pi)^{-1} \int (-it)^r e^{-itx} \varphi_K(t)/\varphi_Z(t/h_r)\, dt$, and where h_r is the Mean Squared Error (MSE) theoretical optimal bandwidth for the estimation of θ_r (possibly different from h).

Delaigle and Gijbels (2002) prove that, under sufficient regularity conditions, the asymptotic MSE of the estimator $\widehat{\theta}_r$ is dominated by its squared bias part, and thus one may choose the bandwidth h_r as the minimizer of the absolute value of the asymptotic bias (ABias), where for a k-th order kernel with k even

$$\text{ABias}[\widehat{\theta}_r] = (-1)^{k/2} \frac{2h_r^k}{k!} \mu_{K,k} \theta_{r+k/2} + (2\pi nh_r^{2r+1})^{-1} \int t^{2r} |\varphi_K(t)|^2 |\varphi_Z(t/h_r)|^{-2}\, dt.$$

(3.3)

In practice the computation of $\widehat{\theta}_r$ necessitates a numerical integration of $\{\widehat{f}_X^{(r)}(.; h_r)\}^2$ over the whole real line. The computations involved can be reduced considerably as we will explain now. Let $\widehat{\varphi}_{X,h_r}^r(t)$ denote the Fourier transform of $\widehat{f}_X^{(r)}(.; h_r)$. One can easily prove that $\widehat{\varphi}_{X,h_r}^r(t) = (-it)^r \widehat{\varphi}_{X,h_r}(t)$, where $\widehat{\varphi}_{X,h_r}(t)$ represents the Fourier transform of the density estimate $\widehat{f}_X(.; h_r)$, given by $\widehat{\varphi}_{X,h_r}(t) = \widehat{\varphi}_Y(t).\varphi_K(h_r t)/\varphi_Z(t)$, with $\widehat{\varphi}_Y$ the empirical characteristic

function of Y (see Delaigle and Gijbels (2001)). An application of Parseval's idendity leads to

$$\widehat{\theta}_r = \frac{1}{2\pi} \int |\widehat{\varphi}_{X,h_r}^r(t)|^2 \, dt = \frac{1}{2\pi h_r^{2r+1}} \int t^{2r} |\widehat{\varphi}_Y(t/h_r)|^2 . |\varphi_K(t)|^2 |\varphi_Z(t/h_r)|^{-2} \, dt,$$

which is easier to compute since this integral has finite integration bounds, due to the factor φ_K in the integrand.

From expression (3.3) we see that selecting h_2, the optimal bandwidth for the estimation of $\theta_2 = R(f_X'')$, will necessitate the estimation of $\theta_{2+k/2}$ which itself requires the estimation of θ_{2+k} and so on. After ℓ steps of iteration one still needs to deal with a pilot estimation of $\theta_{2+\ell \times k/2}$, by referring to a normal density for example. We refer to this as an ℓ-stages procedure. As described in Delaigle and Gijbels (2002), a trade-off between bias and variance has to be made in order to choose ℓ. A theorectical study establishing this trade-off is lacking so far and would be very technical in the deconvolution problem context. From a simulation study which is not reported here, it appears that when the density to be estimated does not present any strong features, a low order (for example one-stage) procedure is preferable, but a higher order (a two- or three-stages) procedure remains quite acceptable. In other cases, such as for example in case of a strong multimodal density, a higher order procedure considerably improves the results, since the decrease of the bias is much more important than the increase of the variance. Hence in general we would advise to use a two- or three-stages procedure.

In Section 5 we report on results using the two-stages selection procedure of Delaigle and Gijbels (2002), which for a second order kernel ($k = 2$) reads as follows

Step 0. Estimate θ_4 via the normal reference method, i.e. $\widehat{\theta}_4 = 8!/(\widehat{\sigma}_X^9 2^9 4! \sqrt{\pi})$;

Step 1. Use $\widehat{\theta}_4$ to select a bandwidth h_3 for estimating θ_3;

Step 2. Use $\widehat{\theta}_3$ to select a bandwidth h_2 for estimating θ_2, the quantity of interest,

where h_i is the bandwidth found by minimization of the squared asymptotic bias.

3.3 Solve-the-equation bandwidth selection

In the error-free case, Park and Marron (1990) and Sheather and Jones (1991), among others, study a solve-the-equation rule to estimate h_{AMISE}, the minimizer of the AMISE. The idea behind the method is to express h_2, the optimal asymptotic MSE bandwidth for the kernel estimation of $R(f_X'')$, as a function of h_{AMISE}, say $h_2 = F(h_{\mathrm{AMISE}})$, estimate the unknown quantities in F and obtain $\widehat{F}$, then plug $\widehat{F}(h_{\mathrm{AMISE}})$ into the expression of h_{AMISE} and solve the resulting fixed point equation in h_{AMISE} :

$$h_{\mathrm{AMISE}} = \left(R(K)/[\mu_{K,2}^2 R(\widehat{f}_X''(.; \widehat{F}(h_{\mathrm{AMISE}})))] \right)^{1/5} n^{-1/5}.$$

It is quite complicated to think of a similar method in the error case. Note that both quantities h_{AMISE} and h_2 necessitate the evaluation of a rather involved integral (see expressions (3.2) and (3.3)). The main difficulty is in the fact that the bandwidths appear in a very implicit way in the integrals. It is not possible to give a general formula for h_{AMISE} and h_2 as in the error-free case, and for a supersmooth error density, it is even not possible to find an analytic expression for the integrals in (3.2) and (3.3). In the ordinary smooth error case, one can compute those integrals analytically and obtain an explicit form for (3.2) and

11

(3.3). However this does not automatically result in an expression for h_{AMISE} and h_2 and in general, one has to drop some lower order terms in (3.2) and (3.3) in order to be able to express h_2 as a function of h_{AMISE}. In conclusion, when applicable, a solve-the-equation rule demands a lot of analytic computations (the formulas have the be recomputed for each error density and each kernel) and implies a lot of asymptotic approximations. For all these reasons, we do not really consider this method as a potential competitor of other practical bandwidth selection methods. We made the necessary computations for the Laplace error case and the kernel used in the simulations and illustated the performance of the method. Some results are presented in Section 5.

4   Other bandwidth selection procedures

The bandwidth selection procedures introduced in Section 3 are based on an asymptotic expression for the mean integrated squared error. Another approach is to directly try to estimate the mean integrated squared error and to minimize this estimator with respect to h. This is the general approach behind the cross-validation and bootstrap bandwidth selectors studied in the literature, which we briefly discuss in the next two sections. In Section 5, we then provide a finite sample comparison of the performances of all available practical bandwidth selectors.

4.1   Cross-validation bandwidth selection

The cross-validation bandwidth selection method of Stefanski and Carroll (1990) relies on the fact that, under sufficient conditions, the cross-validation

quantity

$$\text{CV}(h) = \int \frac{|\varphi_K(ht)|^2|\widehat{\varphi}_Y(t)|^2 + 2(n-1)^{-1}\varphi_K(-ht)\left[-n|\widehat{\varphi}_Y(t)|^2 + 1\right]}{2\pi|\varphi_Z(t)|^2}\,dt,$$

(4.4)

with $\widehat{\varphi}_Y(t)$ the empirical characteristic function of Y, is an unbiased estimator of $\text{MISE}(h) - \int f_X^2(x)\,dx$. This motivates the *cross-validated bandwidth selector*, obtained via minimization of (4.4). Stefanski and Carroll (1990) applied this method in the context of Gaussian errors and using the particular sinc kernel (see Davis (1975) and Tapia and Thompson (1978)). A theoretical study of the cross-validation bandwidth selection procedure has been carried out by Hesse (1999) in the particular case of ordinary smooth error densities.

An advantage of cross-validation is that it requires few assumptions, but in practice it suffers from some drawbacks such as non-uniqueness of the solution or sometimes even very poor performance of the estimator. See Section 5 and Delaigle (1999) for a more complete description. Cao, Cuevas and González-Manteiga (1994) and Jones, Marron and Sheather (1996) among others, encountered the same problems when applying the method in the error-free case.

In the case of non-uniqueness of the solution we would, in a real data example, recommend to select the smallest solution for which the estimator of the density 'appears' smooth enough. Since in a simulation study, a visual inspection of each density estimate is not feasible, in our simulation study we have kept as solution the largest bandwidth found in the search interval, as was suggested by Jones, Marron and Sheather (1996) in the error-free case.

13

The bootstrap method of Delaigle and Gijbels (2001) selects the bandwidth through minimization of the following consistent bootstrap estimator of the $\mathrm{MISE} - \int f_X^2(x)\,dx$:

$$
\begin{aligned}
\mathrm{MISE}_2^*(h) = (2\pi n h)^{-1} \int |\varphi_K(t)|^2 |\varphi_Z(t/h)|^{-2}\,dt - \pi^{-1} \int |\widehat{\varphi}_{X,g}(t)|^2 \varphi_K(ht)\,dt \\
+ (1 - n^{-1})(2\pi)^{-1} \int |\widehat{\varphi}_{X,g}(t)|^2 |\varphi_K(ht)|^2\,dt \quad,
\end{aligned}
\tag{4.5}
$$

where g is a pilot bandwidth and $\widehat{\varphi}_{X,g}(t)$ is the Fourier transform of $\widehat{f}_X(.;g)$ given by $\widehat{\varphi}_{X,g}(t) = \widehat{\varphi}_Y(t).\varphi_K(gt)/\varphi_Z(t)$, with $\widehat{\varphi}_Y$ the empirical characteristic function of Y.

Delaigle and Gijbels (2001) proved that, under sufficient regularity assumptions, this leads to a consistent bandwidth selector. Their theoretical results also establish conditions on the pilot bandwidth g. In their paper it is shown that a good choice for the pilot bandwidth g is the plug-in bandwidth which is optimal (in the MSE sense) for estimation of $R(f_X'')$. We follow Delaigle and Gijbels (2001) and choose g to be the two-stages bandwidth h_2 of Section 3.2.

From a computational point of view, this bootstrap procedure competes with other more classical procedures since no bootstrap sample needs to be generated. As a matter of fact, once a pilot bandwidth g has been chosen, (4.5) can be computed rather easily directly from the original sample. In practice all integrals involved will be computed by numerical integration and the bandwidth will be that value on a grid which minimizes $\mathrm{MISE}_2^*(h)$.

5 Simulation results

Simulations were performed for all above methods using kernel K with characteristic function $\varphi_K(t) = (1-t^2)^3.1_{[-1,1]}(t)$ (see Fan (1992) for an expression of K itself). We consider two different error densities ($N(0;\sigma_Z^2)$ and $Laplace(\sigma_Z)$) and three sample sizes (50, 100 and 250). We present detailed results for four different target densities f_X (#1, #2, #3 and #6 below) that have been chosen because they present each a particular feature that can be encountered in practice, while still being quite standard densities. For a better comparison of the methods we also provide results using densities #4, #5, #7, #8 and #9 of Marron and Wand (1992), which are all a mixture of two (#4, #5, #7, #8) or three (#9) normal densities. See Marron and Wand (1992) for the definition and a graphical representation of those densities. We also consider the $\chi^2(8)$ density (#10). Densities #1, #2, #3 and #6 are, in increasing order of estimation difficulty, defined by

(1) Density #1: $X \sim N(0;1)$, the standard normal density.

(2) Density #2: $X \sim \chi^2(3)$, 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 #6: $X \sim 0.4\,Gamma(5)+0.6\,Gamma(13)$, which is bimodal, but with close and different sized modes;

and are represented in Figure 5.1.

In each case, we generated 500 samples from f_X and f_Z, which, after the addition $X + Z$ resulted in 500 samples from f_Y. The error variance $Var(Z)$ was controlled by the noise to signal ratio $Var(Z)/Var(X)$. For densities #1,
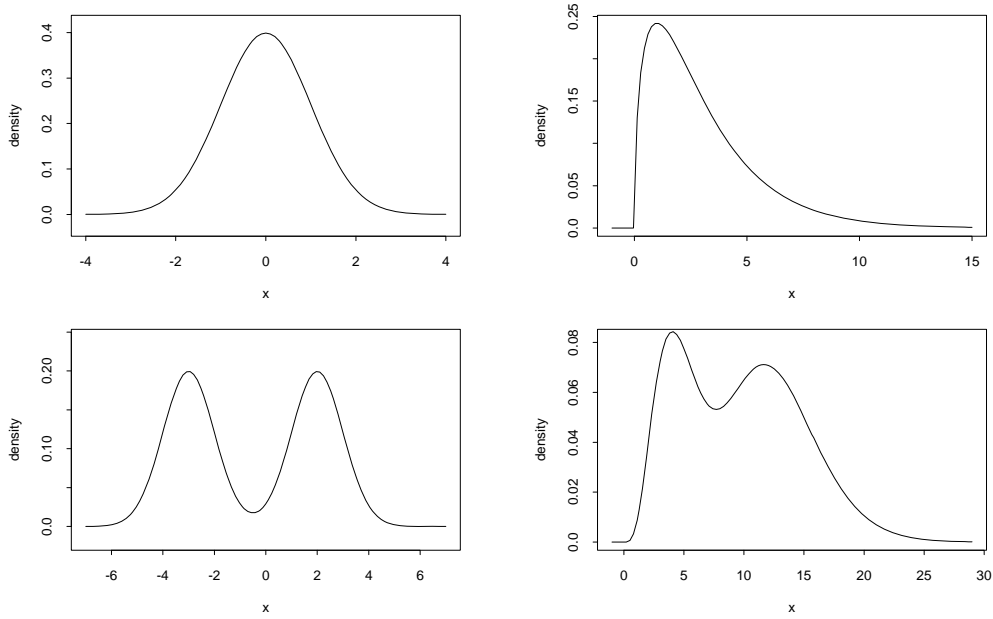
Fig. 5.1. Four target densities: density #1 (top left), density #2 (top right), density #3 (bottom left), and density #6 (bottom right).

#2, #3 and #10 this ratio was set to 25%, but for all the other densities this ratio was set to 10%, since for those densities there are more particular features to recover.

We carried out a more extensive simulation study by considering other values of the error variance and by using other kernels. We chose to report on only a part of this study, since the other simulations led to similar conclusions. Full results of the simulation study can be obtained from the authors upon request.

To assess the quality of the density estimator, we used the criterion

$$\mathrm{ISE}(h) = \int \left\{ \widehat{f}_X(x; h) - f_X(x) \right\}^2 dx,$$

and in each reported case, we computed $\mathrm{ISE}(\widehat{h})$ where $\widehat{h}$ is the bandwidth selected by the method considered.

We first present detailed results (tables and graphs) for the estimation of densities #1, #2, #3 and #6. In the tables we report on the empirical median and interquartile range of the 500 values of the ISE. These robust estimators were preferred to the mean and the standard deviation since they deal with the problem of extreme values encountered in practice (mainly with the cross-validation bandwidth selection method).

In each table we compare the results obtained from four bandwidth selection methods: the normal reference bandwidth selector (NR), the 2-stages plug-in bandwidth selector (PI), the cross-validation bandwidth selector (CV) and the bootstrap bandwidth selector (BT). We present in each case a kernel density estimate of the bandwidth selector based on the 500 values of $\widehat{h}$, using a normal reference bandwidth and a standard normal kernel. We also calculated the exact value of h_{MISE}, which we indicate on those graphs by a vertical line for comparison purpose. For most cases we also provide a picture which shows the target density f_X along with three out of the 500 replicated estimates, corresponding to the first quartile (1st quart), the second quartile (median) and the third quartile (3rd quart) of the 500 calculated ISE's.

At the end of the section we provide a further comparison of the performances of the bandwidth selectors by giving boxplots of the ratio $\mathrm{ISE}(h_{\mathrm{MISE}})/\mathrm{ISE}(\widehat{h})$ for densities #1 to #10.

Table 5.1 compares, for the N(0;1) target density and Laplace or Gaussian error, the results for three different sample sizes (50, 100 and 250). First note that all four methods performed quite well. As one could have expected, the best method in this case was the one referring to a normal density. Nevertheless we see that as n increases, the bootstrap and the plug-in method perform al-

Table 5.1

Simulation results for the N(0; 1) density: median(ISE) with empirical interquartile range between brackets.

	$n = 50$		$n = 100$		$n = 250$	
	$Z \sim$ Laplace	$Z \sim$ Gaussian	$Z \sim$ Laplace	$Z \sim$ Gaussian	$Z \sim$ Laplace	$Z \sim$ Gaussian
NR	0.011 (0.011)	0.011 (0.012)	0.0071 (0.0080)	0.0080 (0.0075)	0.0041 (0.0035)	0.0051 (0.0042)
PI	0.013 (0.013)	0.014 (0.017)	0.0084 (0.0095)	0.0094 (0.0090)	0.0047 (0.0043)	0.0058 (0.0053)
CV	0.016 (0.018)	0.018 (0.020)	0.011 (0.012)	0.012 (0.013)	0.0059 (0.0061)	0.0072 (0.0078)
BT	0.012 (0.012)	0.013 (0.013)	0.0079 (0.0087)	0.0087 (0.0084)	0.0045 (0.0038)	0.0053 (0.0046)
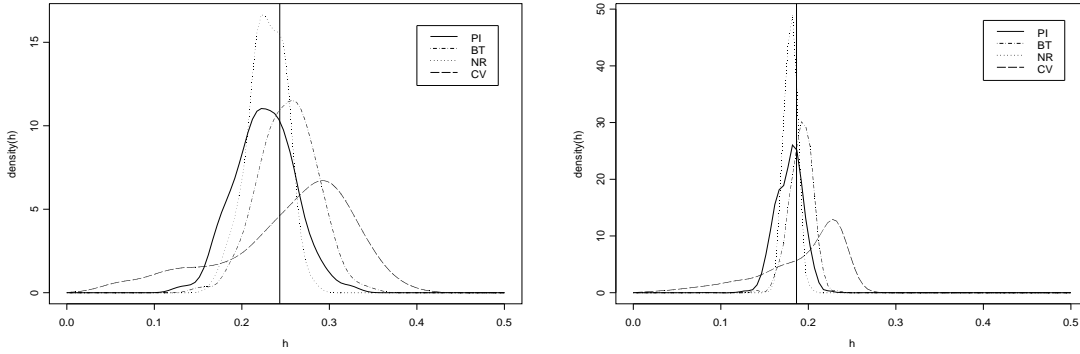
Fig. 5.2. Kernel density estimates of the four bandwidths selectors for estimation of the N(0; 1) density (#1), for sample sizes $n = 50$ (left) and $n = 250$ (right).

most as well. By contrast the poorest method seems to be the cross-validation method, which even for a sample of size 250 remains quite variable. Figure 5.2 presents the kernel density estimates for the 500 $\widehat{h}$ values together with the exact value of h_{MISE}. From this figure we see that the normal reference and plug-in methods select a bandwidth which is slightly underestimated whereas cross-validation and bootstrap led in general to a slightly overestimated bandwidth.

18

Table 5.2

Simulation results for the $\chi^2(3)$ density with $n = 100$.

Method	Laplace error median(ISE) (IQR)	Gaussian error median(ISE) (IQR)
NR	0.018 (0.0063)	0.021 (0.0073)
PI	0.015 (0.0063)	0.018 (0.0084)
CV	0.018 (0.010)	0.022 (0.011)
BT	0.016 (0.0062)	0.020 (0.0083)

From Table 5.2 we can see that when estimating the $\chi^2(3)$ density from a sample of size 100 and a Laplace or Gaussian error, the plug-in method and the bootstrap method seem to give the best results, although the normal reference bandwidth selector also performs reasonably well. The cross-validation method again gives the poorest results. See also Figures 5.3 and 5.4, for the estimates of the $\chi^2(3)$ density and a kernel density estimate of the bandwidth selectors.

Similar conclusions can be drawn from Table 5.3 and Figures 5.5 and 5.6, where, for $n = 100$ or 250, we tried to recover the mixed normal density (#3) contaminated by a Laplace or a Gaussian error. In this case the density is far from a normal, and the plug-in method brings considerable improvements over the NR method. The bootstrap method works well and the cross-validation technique remains very variable.
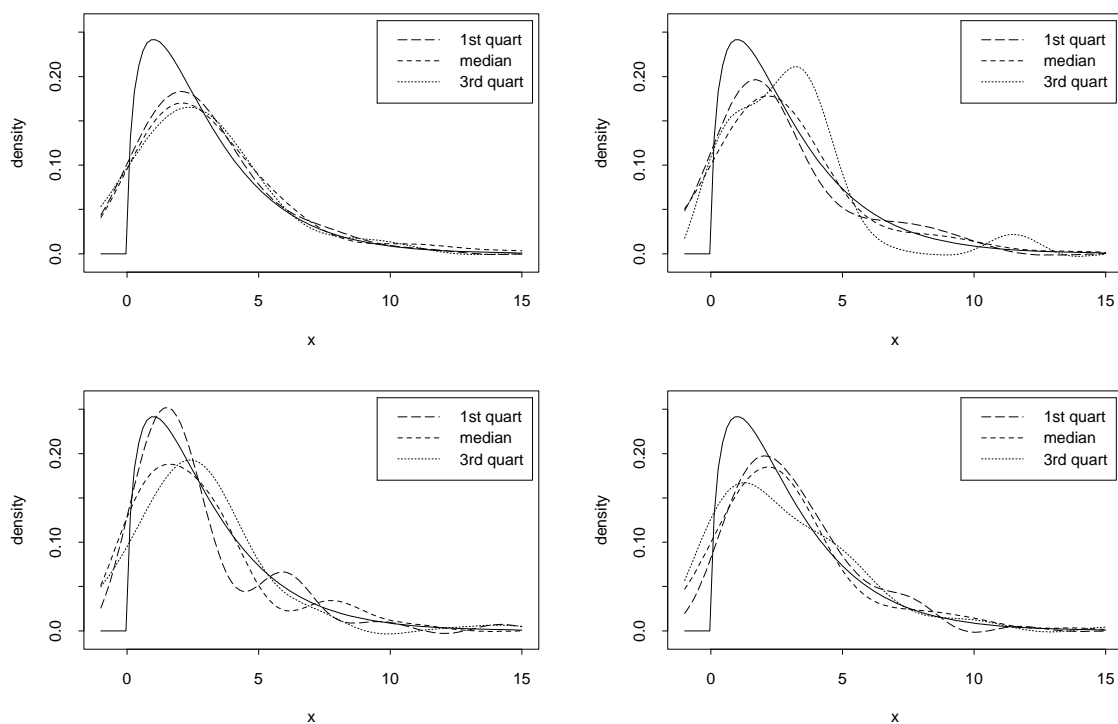
Fig. 5.3. Estimation of the $\chi^2(3)$ density (#2) for a sample of size $n = 100$ and a Laplace error by the normal reference method (top left), the plug-in method (bottom left), the cross-validation method (top right), and the bootstrap method (bottom right).
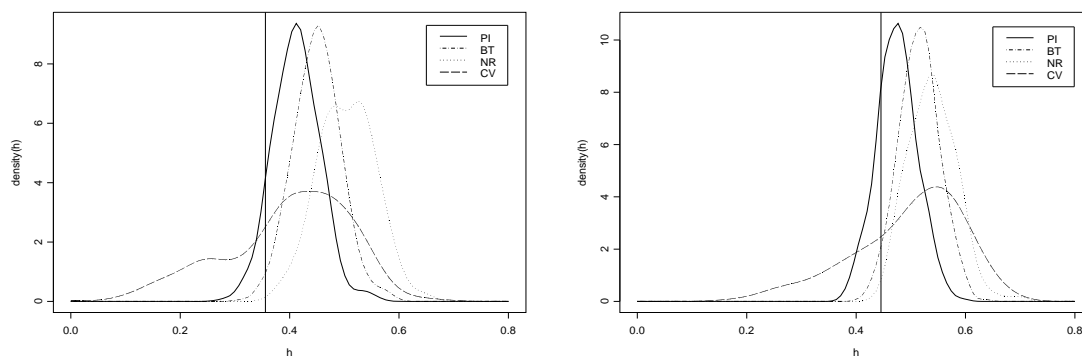

Fig. 5.4. Kernel density estimates of the four bandwidths selectors for estimation of the $\chi^2(3)$ density (#2), for sample size $n = 100$ and Laplace errors (left) and Gaussian errors (right).
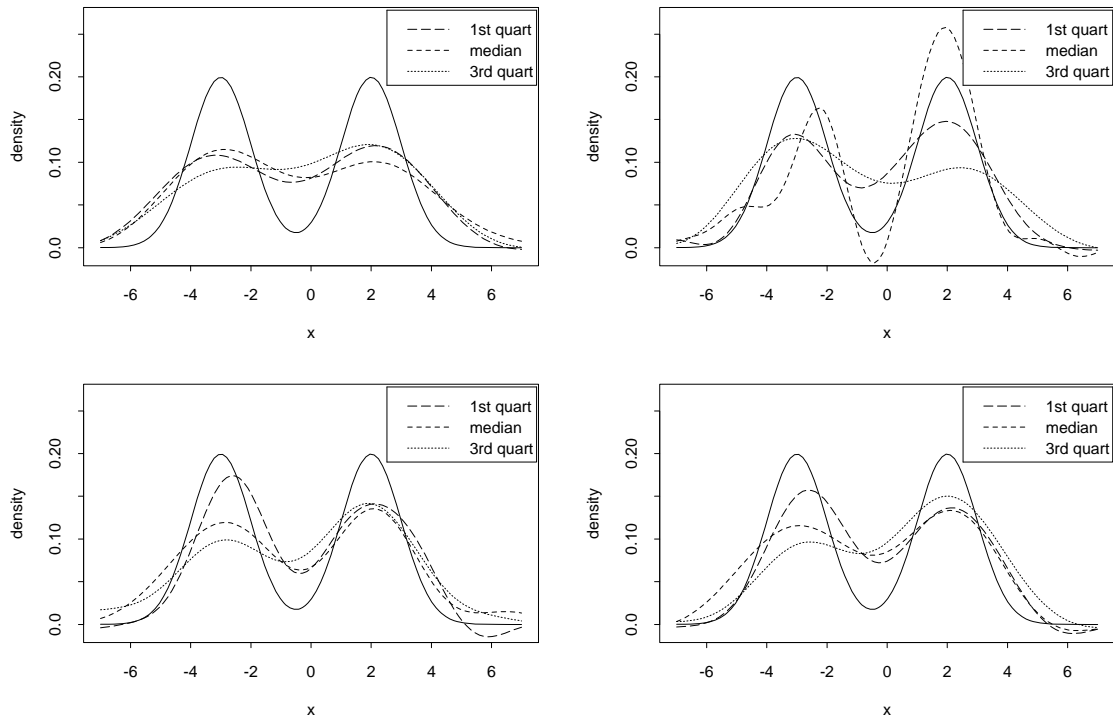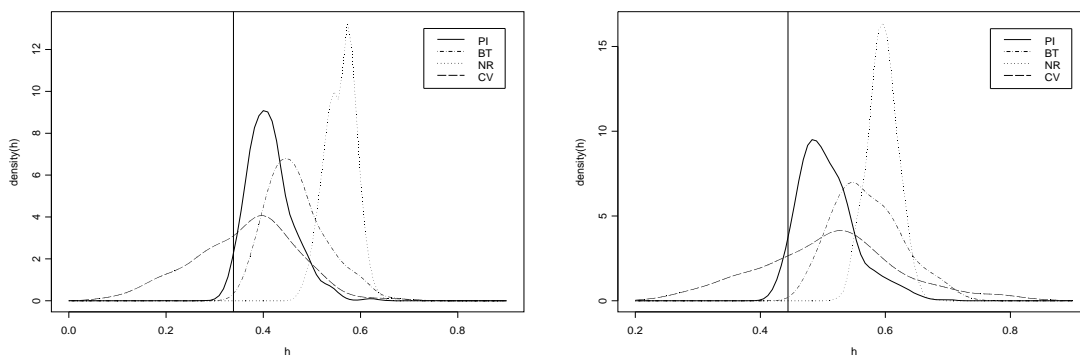
Fig. 5.5. Estimation of the mixed normal density (#3) for a sample of size $n = 100$ and a Laplace error by the normal reference method (top left), the plug-in method (bottom left), the cross-validation method (top right), and the bootstrap method (bottom right).

Fig. 5.6. Kernel density estimates of the four bandwidths selectors for estimation of the mixed normal density (#3), for sample size $n = 100$ and Laplace errors (left) and Gaussian errors (right).

Table 5.3

Simulation results for mixed normal density (#3) with $n = 100$ or 250.

Method	$n = 100$				$n = 250$			
	$Z \sim$ Laplace		$Z \sim$ Gaussian		$Z \sim$ Laplace		$Z \sim$ Gaussian	
	med(ISE)	(IQR)	med(ISE)	(IQR)	med(ISE)	(IQR)	med(ISE)	(IQR)
NR	0.031	(0.0067)	0.034	(0.0072)	0.023	(0.0051)	0.028	(0.0058)
PI	0.018	(0.012)	0.027	(0.013)	0.011	(0.0074)	0.020	(0.011)
CV	0.022	(0.016)	0.032	(0.018)	0.013	(0.010)	0.025	(0.012)
BT	0.022	(0.013)	0.032	(0.013)	0.012	(0.0082)	0.023	(0.010)

Table 5.4 and Figures 5.7 and 5.8 show that even in the more involved mixed gamma density case (#6), the deconvolving kernel density estimator with an appropriate bandwidth still performs reasonably well, taking into account that the estimates are all using a global bandwidth parameter. In fact we see that even if it is hard to recover the two modes properly, globally, the estimator has a good behaviour. Figure 5.9 will confirm that the results obtained by the practical methods compete with those using the optimal global bandwidth (h_{MISE}).

Figures 5.9 and 5.10 compare, for densities #1 to #10, the results of the four practical bandwidth selection methods as well as of the solve-the-equation (SEQ) method (discussed in Section 3.3), relatively to the results obtained by using the MISE optimal bandwidth (h_{MISE}). These figures present boxplots of the ratio $\text{ISE}(h_{\text{MISE}})/\text{ISE}(\widehat{h})$ for each target density and in the Laplace error case. In general this ratio is smaller than one, but it can be larger than one
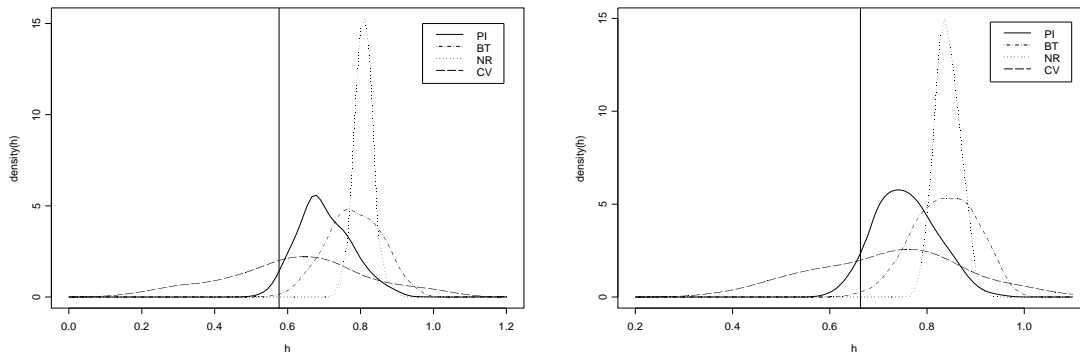
22

Fig. 5.7. Kernel density estimates of the four bandwidths selectors for estimation of the mixed gamma density (#6), for sample size $n = 250$ and Laplace errors (left) and Gaussian errors (right).
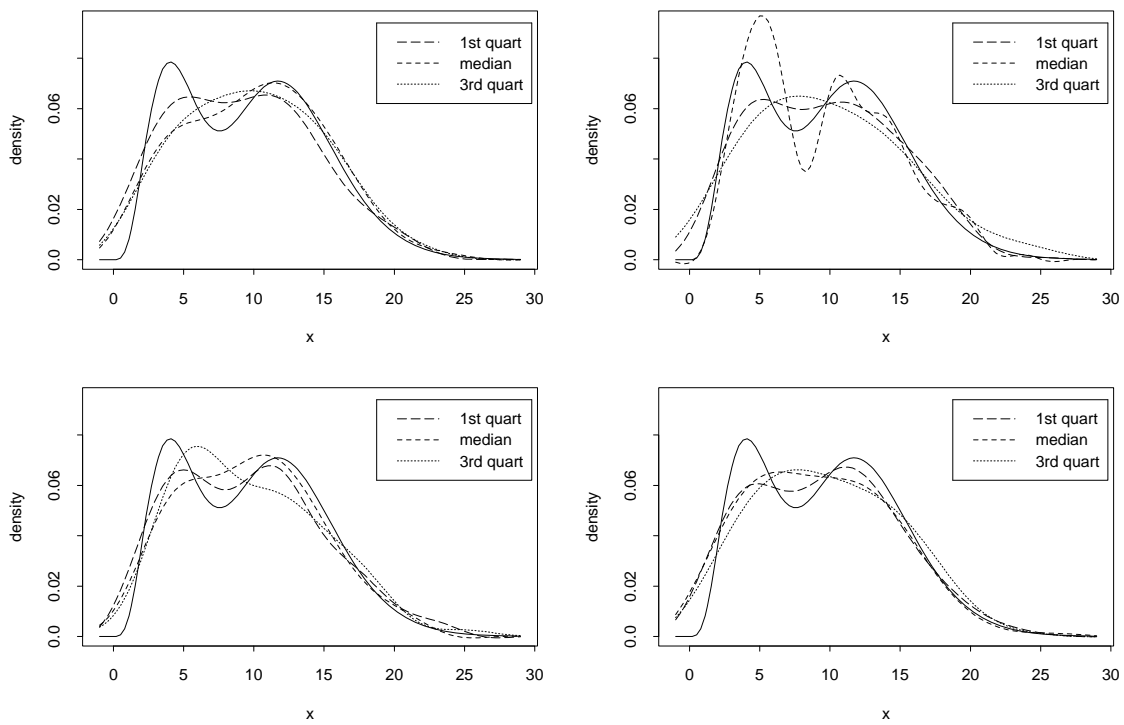

Fig. 5.8. Estimation of the mixed gamma density (#6) for a sample of size $n = 250$ and a Laplace error by the normal reference method (top left), the plug-in method (bottom left), the cross-validation method (top right), and the bootstrap method (bottom right).

Table 5.4

Simulation results for the mixed gamma density (#6), $n = 250$.

Method	Laplace error median(ISE) (IQR)		Gaussian error median(ISE) (IQR)	
NR	0.0024	(0.00078)	0.0026	(0.00085)
PI	0.0021	(0.0011)	0.0023	(0.0011)
CV	0.0024	(0.0017)	0.0025	(0.0014)
BT	0.0024	(0.0011)	0.0026	(0.0010)

since h_{MISE} is the optimal bandwidth on the average, but for a given sample, $\widehat{h}$ can give a smaller ISE. The bigger this ratio the better the method. For densities #1, #2, #3, #6 and #10, the boxplots are given for a sample of size $n = 100$ and 25% of error. For all the other densities the boxplots correspond to a sample of size 250 and 10% of error.

We see that except for density #4 (which has a very particular structure), the PI method gave overall the best ratio accross all simulations, and that this ratio was rather large in general. In most cases the BT method gave similar (but slightly less good) results. The NR method failed with special densities such as densities #3, #4, #5 and #7 but worked reasonably well for the other densities. The CV method gave generally a smaller ratio than the other methods, except for density #4, but the major drawback is that it is always very variable, with a percentage of totally unacceptable results (relative performance close to 0). Moreover, in practice this method does not always have a unique solution. See Section 4. In most cases the SEQ method did not improve the results of the PI method, and suffers from a rather large variability.
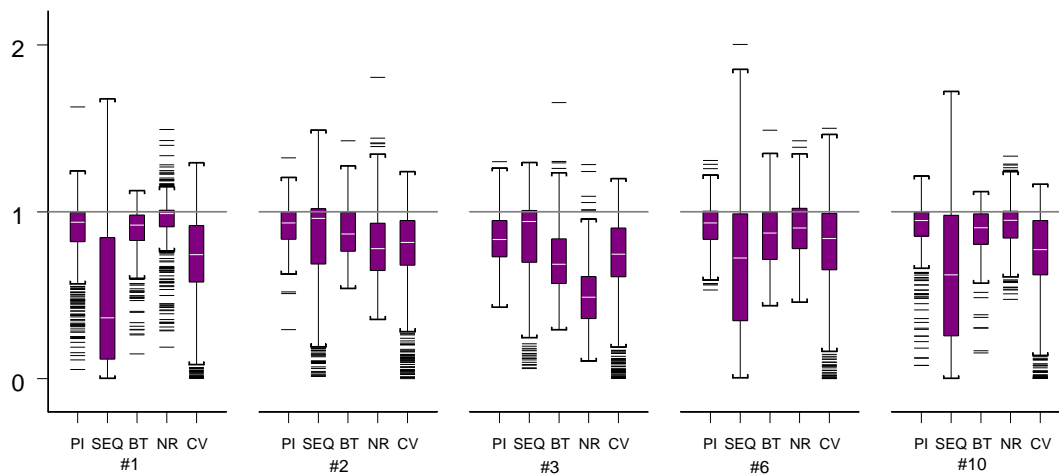
Fig. 5.9. Boxplots of the relative values $\mathrm{ISE}(h_{\mathrm{MISE}})/\mathrm{ISE}(\widehat{h})$ for the bandwidths selectors for estimation of the densities #1, #2, #3, #6 and #10 across the 500 simulations, for a Laplace error. Sample size $n = 100$ and $\mathrm{Var}(Z)/\mathrm{Var}(X) = 0.25$.
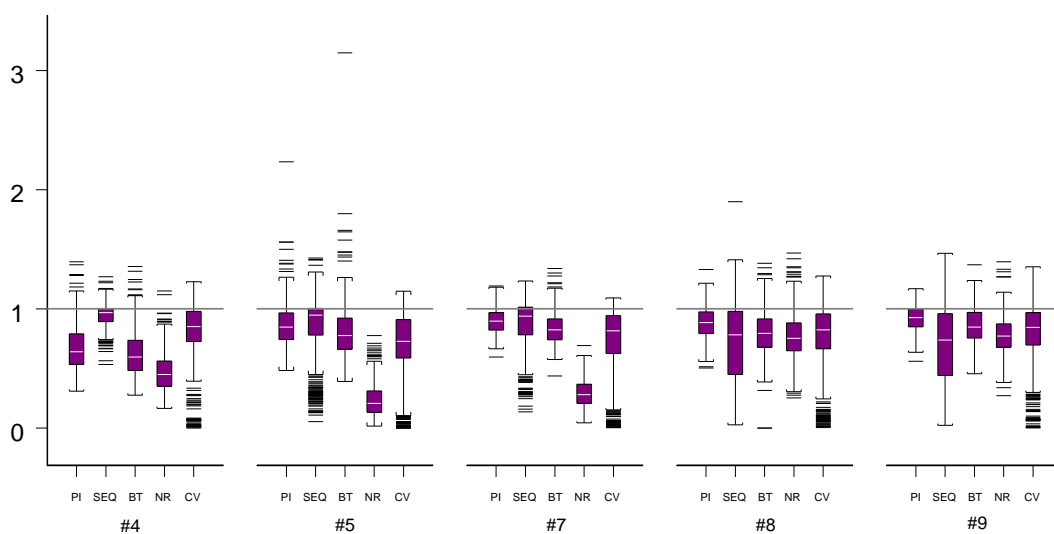

Fig. 5.10. Boxplots of the relative values $\mathrm{ISE}(h_{\mathrm{MISE}})/\mathrm{ISE}(\widehat{h})$ for the bandwidths selectors for estimation of the densities #4, #5, #7, #8 and #9 across the 500 simulations, for a Laplace error. Sample size $n = 250$ and $\mathrm{Var}(Z)/\mathrm{Var}(X) = 0.10$.

This together with the remarks of Section 3.3 makes that we can not really recommend the solve-the-equation bandwidth selection method. From Figures 5.9 and 5.10 we can conclude that some of our practical selection procedures have a performance close to the one for the theoretical MISE-based bandwidth (since the ratio's are relatively close to 1).

From Figures 5.2, 5.4, 5.6 and 5.7 we see that in general the four bandwidth selectors are biased to the right, meaning that they give slighlty too big bandwidth values. The plug-in bandwidth was in general smaller than the bootstrap bandwidth, which is apparently also the case in the error-free case, looking at the results of Cao, Cuevas and González-Manteiga (1994). This is related to a remark of Marron and Wand (1992) who noticed that h_{AMISE} was often smaller than h_{MISE} in the error-free case. We got to a similar conclusion from our exact computations in the error case. As a consequence, the PI bandwidth is generally less biased than the other methods. The CV bandwidth is less biased in the mixed densities cases, but to the extent of a large variance. As mentioned in Section 3, the NR bandwidth is quite biased for densities far from a normal. Figures 5.9 and 5.10 show that for the PI bandwidth (mainly), this bias does not have a serious effect on the efficiency of the method, since the ratio is relatively close to 1.

In general our conclusions are similar to those in the error-free case, but the derivation of the results and the application of the methods in practice are much more involved in the error case. See for example Jones, Marron and Sheather (1996) or Cao, Cuevas and González-Manteiga (1994) for detailed results in the error-free case. Further, the simulation results with a Laplace or Gaussian error are very similar, but as one could expect, the Laplace case always gives better results than the Gaussian case.

Finally no method proved to be the best in all cases, but, nevertheless, the plug-in and bootstrap methods seem to be reliable techniques, whatever the density to estimate.

6   A real data example: The NHANES study

The 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). From previous nutrition studies it is reasonable to assume that for this type of data, more than 50% of the variance is due to the noise. See for example Stefanski and Carroll (1990) or Carroll, Ruppert and Stefanski (1995).

We applied the plug-in method, the normal reference method, the bootstrap bandwidth and the cross-validation method to these data, for two different error densities (Gaussian and Laplace), and an error variance of approximately $1.5\,\sigma_X^2$, which corresponds to 60% of the variance of the data due to the noise.

Since this variance comes from the knowledge of other nutrition studies, and thus is probably just a rough estimate of the actual variance, we follow Stefanski and Carroll (1990) and also consider the cases $\sigma_Z^2 = (1/5)\,\sigma_X^2$ or $\sigma_Z^2 = (1/3)\,\sigma_X^2$. Each time we compare the results with the case $\sigma_Z^2 = 0$ (the error-
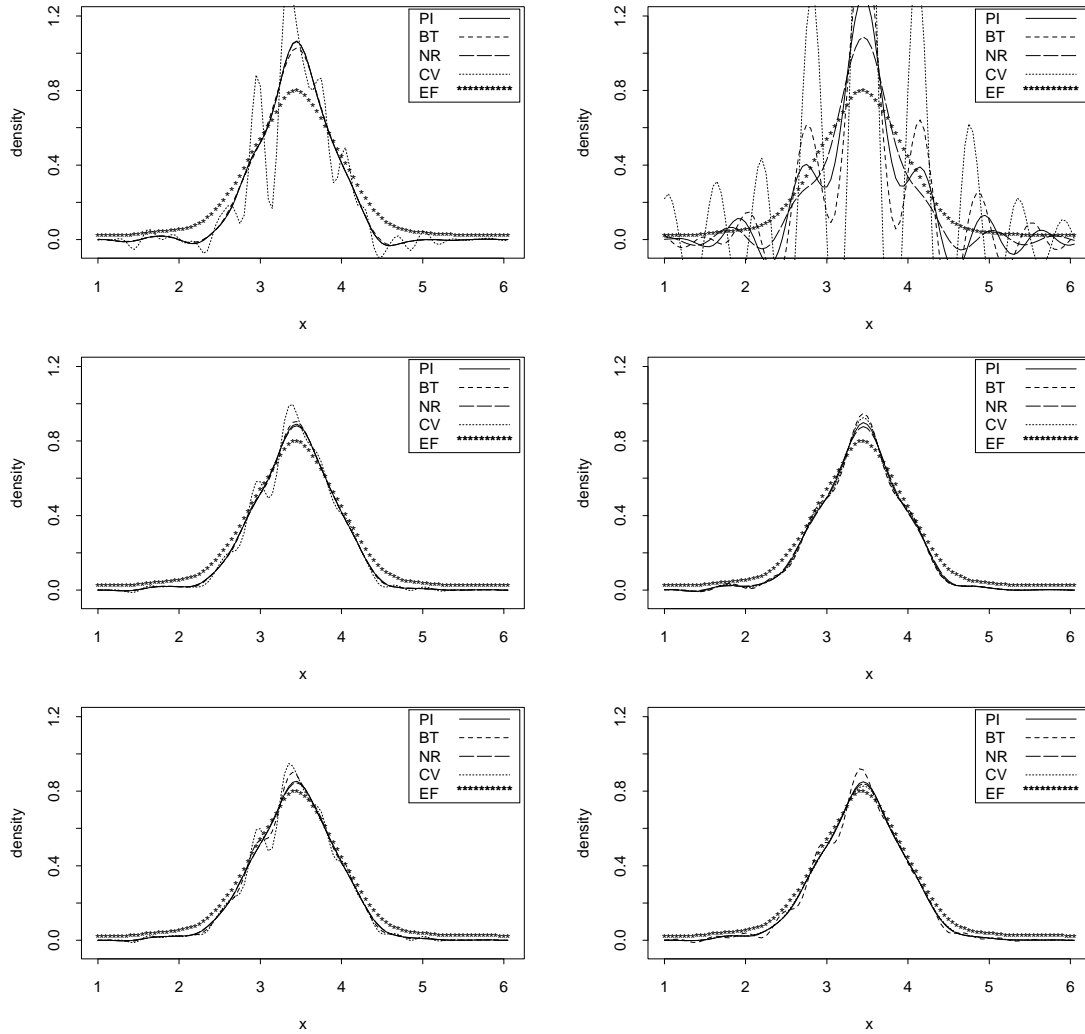
Fig. 6.11. Estimation of the log saturated fat density for a Laplace error (left panels) with $\sigma_Z^2 = 1.5\,\sigma_X^2$ (top left), $\sigma_Z^2 = (1/3)\,\sigma_X^2$ (center left), $\sigma_Z^2 = (1/5)\,\sigma_X^2$ (bottom left), and for a normal error (right panels) with $\sigma_Z^2 = 1.5\,\sigma_X^2$ (top right), $\sigma_Z^2 = (1/3)\,\sigma_X^2$ (center right), $\sigma_Z^2 = (1/5)\,\sigma_X^2$ (bottom right). The curves indicated by the $*$-characters represent the results for the error-free (EF) case.

free case).

The resulting estimators are depicted in Figure 6.11. The curves indicated by the $*$-characters in each picture represent the results for the error-free case, obtained by using a classical plug-in bandwidth selector. We see that the

plug-in, normal reference and bootstrap methods are almost indistinguishable. This is not surprising since the density appears as almost normal and our simulations already revealed that for such a density all three methods perform comparably (see Table 5.1). By contrast the cross-validation method seems to give, in general, a too small bandwidth. This small bandwidth has a dramatic effect on the resulting density estimator, mainly if we assume a large error variance ($1.5\ \sigma_X^2$). Note that Stefanski and Carroll (1990) encountered similar problems when deconvolving with this much noise.

Except in this case of large error variance, the estimators do not differ much when considering different error densities. It seems that the important point is to specify an error variance, whatever the distribution of the error. This goes along with results of Hesse (1999) who studied the robustness of the deconvolving kernel density estimator to error misspecification. The error variance plays an important role in the estimation process: the smaller this variance, the smoother the estimator will be.

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).

Acknowledgements

References

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

Cao, R., Cuevas, A. and González-Manteiga, W. (1994). A comparative study of several smoothing methods in density estimation. Computational Statistics and Data Analysis, 17, 153–176.

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.

Davis , K.B., 1975. Mean square error properties of density estimators. Annals of Statistics, 3, 1025–1030.

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.
http://www.stat.ucl.ac.be/ISpersonnel/delaigle/liens_en.htm#publications

Delaigle, A. and Gijbels, I., 2001. Bootstrap bandwidth selection in kernel density estimation from a contaminated sample. Discussion Paper #0116, Institut de Statistique, Université catholique de Louvain, Belgium.
http://www.stat.ucl.ac.be

Delaigle, A. and Gijbels, I., 2002. Estimation of integrated squared density

30

derivatives from a contaminated sample. Journal of the Royal Statistical Society, Series B, to appear.

Devroye, L., 1989. Consistent deconvolution in density estimation. The Canadian Journal of Statistics, 7, 235–239.

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

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

Fan, J. ,1991c. Global behaviour of deconvolution kernel estimates. Statistica Sinica, 1, 541–551.

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

Fan, J. and Koo, J.Y., 2002. Wavelet deconvolution. IEEE Transactions on Information Theory, 48, 734–747.

Hesse, C.H., 1999. Data-driven deconvolution. Nonparametric Statistics, 10, 343–373.

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

Marron, J.S. and Wand, M.P., 1992. Exact mean integrated squared error. The Annals of Statistics, 20, 712–736.

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

Park, B. and Marron, J.S., 1990. Comparison of data-driven bandwidth selectors. Journal of the American Statistical Association, 85, 66–72.

Sheather, S.J. and Jones, M.C., 1991. A reliable data-based bandwidth selection method for kernel density estimation. Journal of the Royal Statistical

Society, Series B, 53, 683–690.

Silverman, B.W., 1986. Density estimation for statistics and data analysis. Chapman and Hall, London and New York.

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.

Tapia, R.A. and Thompson, J.R., 1978. Nonparametric density estimation. John Hopkins University Press, Baltimore.

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 and New York.