Practical bandwidth selection in deconvolution kernel density estimation

A. Delaigle, I. Gijbels

Institut de Statistique, Université catholique de Louvain, 20 Voie du Roman Pays, Louvain-la-Neuve B1348, Belgium

Received 25 October 2002

Abstract

Kernel estimation of a density based on contaminated data is considered and the important issue of how to choose the bandwidth parameter in practice is discussed. Some plug-in (PI) type of bandwidth selectors, which are based on non-parametric estimation of an approximation of the mean integrated squared error, are proposed. 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. A simulation study compares these PI bandwidth selectors with a bootstrap (BT) and a cross-validated (CV) bandwidth selector. It is concluded that in finite samples, an appropriately chosen PI bandwidth selector and the BT bandwidth selector perform comparably and both outperform the CV bandwidth. The use of the various practical bandwidth selectors is illustrated on a real data example.

Keywords: Bandwidth selection; Bootstrap; Cross-validation; Deconvolution; Errors-in-variables; Kernel density estimation; Plug-in methods

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 those of Stefanski and Carroll (1990) and Hesse (1999) who propose a cross-validation (CV) procedure, and Delaigle and Gijbels (2001) who investigate a (BT) method. In this paper we introduce plug-in (PI) 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 (NR) bandwidth selector. The use of more sophisticated (non-parametric) estimators for the unknown quantities in the approximated mean integrated squared error leads to the more elaborated PI bandwidth selectors.

After having introduced the PI 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 NR method, the more elaborated PI method, the CV method and the BT method. We will see that in some cases the PI method brings considerable improvement to the NR method and that it competes with the BT procedure. Both, the PI and the BT bandwidth selectors, outperform the CV 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 NR, a PI and a solve-the-equation (SEQ) bandwidth selector. Section 4 briefly discusses the other available data-driven bandwidth selectors: the CV method of Stefanski and Carroll (1990) and the BT 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

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

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

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

where \( K^Z(u; h) = (2\pi)^{-1} \int e^{-iu} \varphi_L(t)/\varphi_Z(t/h) dt, \) with \( \varphi_L \) 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 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–c, 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 non-parametric 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, 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 estimator (2.1) requires some restrictions to be imposed on \( \varphi_K \), the Fourier transform of \( K \). Throughout this paper we will, for simplicity, use kernels with \( \varphi_K \) compactly supported although in the ordinary smooth error case, \( K \) could be taken with \( \varphi_K \) supported on \( \mathbb{R} \) and such that \( \varphi_K \) 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 \( \hat{f}_X \) 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 \( \hat{f}_X(\cdot;h) \). In order to select an ‘optimal’ bandwidth, we need to choose a way to measure the distance between the estimator \( \hat{f}_X(\cdot;h) \) and its target density \( f_X \). A commonly used criterion is the mean integrated squared error (MISE):

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

The optimal bandwidth, \( h_{\text{MISE}} \), 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 discuss practical bandwidth selectors from there on. In order to highlight the dependency on \( h \), we will in what follows write \( \text{MISE}\{\hat{f}_X(\cdot;h)\} \) as \( \text{MISE}(h) \).

3. PI type of bandwidth selection

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

\[
\text{AMISE}(h) = (2\pi nh)^{-1} \int |\phi_K(t)|^2 |\phi_Z(t/h)|^{-2} \, dt + \frac{h^4}{4} \mu_{k,2}^2 R(f_X'''),
\]

(3.1)

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, (3.1) involves the unknown quantity \( R(f_X''') \), 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 \( \text{AMISE}(h) \). Application of our bandwidth selection procedures in practice requires minimization of the resulting estimator of the asymptotic mean integrated squared error in (3.1). In general, we can only approximate the solution numerically: on a discrete grid of \( h \)-values we evaluate \( \text{AMISE}(h) \) by numerical integration and select the \( h \) that minimizes \( \text{AMISE}(h) \) 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 more easily (for example analytically). See Delaigle (1999) for more details.

Section 3.3 discusses a SEQ type of bandwidth selection method.

3.1. NR bandwidth selection

A first elementary estimator of \( R(f''_X) \) is provided by making reference to a normal distribution: one assumes that \( f_X \) is a normal \( N(\mu_X; \sigma_X^2) \) density which implies that \( R(f''_X) = 0.375\sigma_X^{-5}\pi^{-1/2} \). The estimator \( R(\hat{f}''_X) \) is then defined by \( R(\hat{f}''_X) = 0.375\hat{\sigma}_X^{-5}\pi^{-1/2} \), with \( \hat{\sigma}_X^2 \) being 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, \( \hat{\sigma}_X^2 = \hat{\sigma}_Y^2 - \hat{\sigma}_Z^2 \), where \( \hat{\sigma}_Y \) is the sample standard deviation of the observations \( Y_i \).

In general, when \( f_X \) is not a normal density, this estimator \( R(\hat{f}''_X) \) 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 is 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 non-parametric estimator for \( R(f''_X) \) is discussed in the next section.

3.2. PI bandwidth selection

Non-parametric estimation of \( \theta_r = R(f''_X) \), with \( r \) being 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 \( \theta_r \) by \( \hat{\theta}_r = R(\hat{f}^{(r)}_X(\cdot; h_r)) \), where \( \hat{f}^{(r)}_X(\cdot; h_r) \) is the deconvolving kernel density estimator of the \( r \)th derivative of \( f_X \), defined by

\[
\hat{f}^{(r)}_X(x; h_r) = \left(\frac{n}{nh_r^r} + 1\right)^{-1} \sum_{i=1}^{n} K^{(r)}_Z \left( \frac{x - Y_i}{h_r} ; h_r \right)
\]

with \( K^{(r)}_Z(x; h_r) = (2\pi)^{-1} \int (-it)^r e^{-itx} \phi_X(t) \phi_Z(t/h_r) \, dt \), and where \( h_r \) is the mean squared error (MSE) theoretical optimal bandwidth for the estimation of \( \theta_r \) (possibly different from \( h \)).

Delaigle and Gijbels (2002) prove that, under sufficient regularity conditions, the asymptotic MSE of the estimator \( \hat{\theta}_r \) 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,

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

(3.2)
In practice, the computation of \( \hat{\theta}_r \) necessitates a numerical integration of \( \{ \hat{f}_X^{(r)}(\cdot; h_r) \}^2 \) over the whole real line. The computations involved can be reduced considerably as we will explain now. Let \( \hat{\phi}_{X,h_r}(t) \) denote the Fourier transform of \( \hat{f}_X^{(r)}(\cdot; h_r) \). One can easily prove that \( \hat{\phi}_{X,h_r}(t) = (-it)^r \hat{\phi}_{X,h_r}(t) \), where \( \hat{\phi}_Y \) being the empirical characteristic function of \( Y \) (see Delaigle and Gijbels (2001)). An application of Parseval’s identity leads to

\[
\hat{\theta}_r = \frac{1}{2\pi} \int |\hat{\phi}_{X,h_r}(t)|^2 \, dt = \frac{1}{2\pi h_r^{2r+1}} \int t^{2r} |\hat{\phi}_Y(t/h_r)|^2 |\varphi_K(t)^2| |\varphi_Z(t/h_r)|^{-2} \, dt,
\]

which is easier to compute since this integral has finite integration bounds, due to the factor \( \varphi_K \) in the integrand.

From expression (3.2) we see that selecting \( h_2 \), the optimal bandwidth for the estimation of \( \theta_2 = R(f_X^{(2)}) \) will necessitate the estimation of \( \theta_{2+k/2} \) which itself requires the estimation of \( \theta_{2+k} \) and so on. After \( \ell \) 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 \( \ell \)-stage procedure. As described in Delaigle and Gijbels (2002), a trade-off between bias and variance has to be made in order to choose \( \ell \). A theoretical 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-stage) 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-stage procedure.

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

**Step 0:** Estimate \( \hat{\theta}_4 \) via the NR method, i.e. \( \hat{\theta}_4 = 8!/(\hat{\sigma}_X^4 2^34!\sqrt{\pi}) \).

**Step 1:** Use \( \hat{\theta}_4 \) to select a bandwidth \( h_1 \) for estimating \( \theta_3 \).

**Step 2:** Use \( \hat{\theta}_3 \) to select a bandwidth \( h_2 \) for estimating \( \theta_2 \), the quantity of interest.

Here \( h_i \) is the bandwidth found by minimization of the squared asymptotic bias.

### 3.3. SEQ bandwidth selection

In the error-free case, Park and Marron (1990) and Sheather and Jones (1991), among others, study a SEQ rule to estimate \( h_{\text{AMISE}} \), 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 \( R(f_X^{(2)}) \), as a function of \( h_{\text{AMISE}} \), say \( h_2 = F(h_{\text{AMISE}}) \), estimate the unknown quantities in \( F \) and obtain \( \hat{F} \), then plug \( \hat{F}(h_{\text{AMISE}}) \) into the expression of \( h_{\text{AMISE}} \) and solve the resulting fixed point equation in \( h_{\text{AMISE}} \):

\[
h_{\text{AMISE}} = \left( \frac{R(K)}{[\mu_{K,2} R(\hat{f}_X^{(2)}(\cdot; \hat{F}(h_{\text{AMISE}})))]} \right)^{1/5} n^{-1/5}.
\]
It is quite complicated to think of a similar method in the error case. Note that both quantities $h_{\text{AMISE}}$ and $h_2$ necessitate the evaluation of a rather involved integral (see expressions (3.1) and (3.2)). 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 $h_{\text{AMISE}}$ 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 (3.1) and (3.2). In the ordinary smooth error case, one can compute these integrals analytically and obtain an explicit form for (3.1) and (3.2). However, this does not automatically result in an expression for $h_{\text{AMISE}}$ and $h_2$ and, in general, one has to drop some lower order terms in (3.1) and (3.2) in order to be able to express $h_2$ as a function of $h_{\text{AMISE}}$. In conclusion, when applicable, a SEQ rule demands a lot of analytic computations (the formulas have to 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 illustrated 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 CV and BT 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. CV bandwidth selection

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

$$CV(h) = \int |\varphi_K(ht)|^2 |\hat\phi_Y(t)|^2 + 2(n-1)^{-1} \varphi_K(-ht) [-n|\hat\phi_Y(t)|^2 + 1] \frac{dt}{2\pi |\varphi_Z(t)|^2}, \quad (4.1)$$

with $\hat\phi_Y(t)$ being the empirical characteristic function of $Y$, is an unbiased estimator of $\text{MISE}(h) = \int f_X^2(x) \, dx$. This motivates the CV bandwidth selector, obtained via minimization of (4.1). 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 CV bandwidth selection procedure has been carried out by Hesse (1999) in the particular case of ordinary smooth error densities.

An advantage of CV 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 et al. (1994) and Jones et al. (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 et al. (1996) in the error-free case.

4.2. BT bandwidth selection

The BT method of Delaigle and Gijbels (2001) selects the bandwidth through minimization of the following consistent BT estimator of the MISE

\[ \text{MISE}_2^*(h) = (2\pi nh)^{-1} \int |\varphi_K(t)|^2 |\varphi_Z(t/h)|^2 dt - \pi^{-1} \int |\hat{\varphi}_{X,g}(t)|^2 |\phi_K(ht)| dt \]

\[ + (1 - n^{-1})(2\pi)^{-1} \int |\hat{\varphi}_{X,g}(t)|^2 |\phi_K(ht)|^2 dt, \]

(4.2)

where \( g \) is a pilot bandwidth and \( \hat{\varphi}_{X,g}(t) \) is the Fourier transform of \( \hat{f}_X(:, g) \) given by \( \hat{\varphi}_X(t) = \hat{\phi}_Y(t) \varphi_K(gt) / \varphi_Z(t) \), with \( \hat{\phi}_Y \) being 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 PI 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-stage bandwidth \( h_2 \) of Section 3.2.

From a computational point of view, this BT 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, (4.2) 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 \( \text{MISE}_2^*(h) \).

5. Simulation results

Simulations were performed for all the 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 (\( N(0; \sigma^2_Z) \) 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 these 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: $X \sim 0.4 \text{ Gamma}(5)+0.6 \text{ Gamma}(13)$, which is bimodal, but with close and different sized modes;

and are represented in Fig. 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 $\text{Var}(Z)$ was controlled by the noise-to-signal ratio $\text{Var}(Z)/\text{Var}(X)$. For densities #1, #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.
Table 1
Simulation results for the N(0;1) density: median(ISE) with empirical interquartile range in brackets

<table>
<thead>
<tr>
<th>n = 50</th>
<th>n = 100</th>
<th>n = 250</th>
</tr>
</thead>
<tbody>
<tr>
<td>Z ~ Laplace</td>
<td>Z ~ Gaussian</td>
<td>Z ~ Laplace</td>
</tr>
<tr>
<td>NR 0.011 (0.011)</td>
<td>0.011 (0.012)</td>
<td>0.0071 (0.0080)</td>
</tr>
<tr>
<td>PI 0.013 (0.013)</td>
<td>0.014 (0.017)</td>
<td>0.0084 (0.0095)</td>
</tr>
<tr>
<td>CV 0.016 (0.018)</td>
<td>0.018 (0.020)</td>
<td>0.011 (0.012)</td>
</tr>
<tr>
<td>BT 0.012 (0.012)</td>
<td>0.013 (0.013)</td>
<td>0.0079 (0.0087)</td>
</tr>
</tbody>
</table>

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

\[ \text{ISE}(h) = \int \left( \hat{f}_X(x; h) - f_X(x) \right)^2 dx, \]

and in each reported case, we computed ISE(\(\hat{h}\)) where \(\hat{h}\) 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 CV bandwidth selection method).

In each table we compare the results obtained from four bandwidth selection methods: the NR bandwidth selector, the two-stage PI bandwidth selector, the CV bandwidth selector and the BT bandwidth selector. We present in each case a kernel density estimate of the bandwidth selector based on the 500 values of \(\hat{h}\), using a NR bandwidth and a standard normal kernel. We also calculated the exact value of \(h_{\text{MISE}}\), 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 ISEs.

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

Table 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 BT and the PI method perform almost as well. By contrast the poorest method seems to be the CV method, which even for a sample of size 250 remains quite variable. Fig. 2 presents the kernel density estimates for the 500 \(\hat{h}\) values together with the exact value of \(h_{\text{MISE}}\). From this figure we see that the NR and PI methods select a bandwidth which is slightly underestimated whereas CV and BT led in general to a slightly overestimated bandwidth.
Table 2
Simulation results for the $\chi^2(3)$ density with $n = 100$

<table>
<thead>
<tr>
<th>Method</th>
<th>Laplace error</th>
<th>Gaussian error</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>median(ISE) (IQR)</td>
<td>median(ISE) (IQR)</td>
</tr>
<tr>
<td>NR</td>
<td>0.018 (0.0063)</td>
<td>0.021 (0.0073)</td>
</tr>
<tr>
<td>PI</td>
<td>0.015 (0.0063)</td>
<td>0.018 (0.0084)</td>
</tr>
<tr>
<td>CV</td>
<td>0.018 (0.010)</td>
<td>0.022 (0.011)</td>
</tr>
<tr>
<td>BT</td>
<td>0.016 (0.0062)</td>
<td>0.020 (0.0083)</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Method</th>
<th>$n = 100$</th>
<th>$n = 250$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$Z \sim \text{Laplace}$</td>
<td>$Z \sim \text{Gaussian}$</td>
</tr>
<tr>
<td></td>
<td>med(ISE) (IQR)</td>
<td>med(ISE) (IQR)</td>
</tr>
<tr>
<td>NR</td>
<td>0.031 (0.0067)</td>
<td>0.034 (0.0072)</td>
</tr>
<tr>
<td>PI</td>
<td>0.018 (0.012)</td>
<td>0.027 (0.013)</td>
</tr>
<tr>
<td>CV</td>
<td>0.022 (0.016)</td>
<td>0.032 (0.018)</td>
</tr>
<tr>
<td>BT</td>
<td>0.022 (0.013)</td>
<td>0.032 (0.013)</td>
</tr>
</tbody>
</table>

From Table 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 PI method and the BT method seem to give the best results, although the NR bandwidth selector also performs reasonably well. The CV method again gives the poorest results. See also Figs. 3 and 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 3 and Figs. 5 and 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 PI method brings considerable improvements over the NR method. The BT method works well and the CV technique remains very variable.
Table 4 and Figs. 7 and 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. Fig. 9 will confirm that the results obtained by the practical methods compete with those using the optimal global bandwidth ($h_{MISE}$).
Fig. 5. Estimation of the mixed normal density (#3) for a sample of size $n = 100$ and a Laplace error by the NR method (top left), the PI method (bottom left), the CV method (top right), and the BT method (bottom right).

Fig. 6. Kernel density estimates of the four bandwidth selectors for estimation of the mixed normal density (#3), for sample size $n = 100$ and Laplace errors (left) and Gaussian errors (right).

Figs. 9 and 10 compare, for densities #1–#10, the results of the four practical bandwidth selection methods as well as of the SEQ method (discussed in Section 3.3), relatively to the results obtained by using the MISE optimal bandwidth ($h_{\text{MISE}}$). These figures present boxplots of the ratio $\text{ISE}(h_{\text{MISE}})/\text{ISE}(\hat{h})$ 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, since $h_{\text{MISE}}$ is the optimal bandwidth on the average, but for a given sample, $\hat{h}$ can give a smaller ISE. The bigger this ratio, the better the method. For densities #1,
Table 4
Simulation results for the mixed gamma density (#6), n = 250

<table>
<thead>
<tr>
<th>Method</th>
<th>Laplace error median(ISE) (IQR)</th>
<th>Gaussian error median(ISE) (IQR)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NR</td>
<td>0.0024 (0.00078)</td>
<td>0.0026 (0.00085)</td>
</tr>
<tr>
<td>PI</td>
<td>0.0021 (0.0011)</td>
<td>0.0023 (0.0011)</td>
</tr>
<tr>
<td>CV</td>
<td>0.0024 (0.0017)</td>
<td>0.0025 (0.0014)</td>
</tr>
<tr>
<td>BT</td>
<td>0.0024 (0.0011)</td>
<td>0.0026 (0.0010)</td>
</tr>
</tbody>
</table>

Fig. 7. Kernel density estimates of the four bandwidth selectors for estimation of the mixed gamma density (#6), for sample size n = 250 and Laplace errors (left) and Gaussian errors (right).

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 across 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. This together with the remarks of Section 3.3 makes that we cannot really recommend the SEQ bandwidth selection method. From Figs. 9 and 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 ratios are relatively close to 1).

From Figs. 2, 4, 6 and 7 we see that in general the four bandwidth selectors are biased to the right, meaning that they give slightly too big bandwidth values. The PI bandwidth was in general smaller than the BT bandwidth, which is apparently also
Fig. 8. Estimation of the mixed gamma density (#6) for a sample of size $n = 250$ and a Laplace error by the NR method (top left), the PI method (bottom left), the CV method (top right), and the BT method (bottom right).

Fig. 9. Boxplots of the relative values $\text{ISE}(\hat{h})/\text{ISE}(\hat{h}_{\text{MISE}})$ for the bandwidth 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 $\text{Var}(Z)/\text{Var}(X) = 0.25$.

the case in the error-free case, looking at the results of Cao et al. (1994). This is related to a remark of Marron and Wand (1992) who noticed that $h_{\text{AMISE}}$ was often smaller than $h_{\text{MISE}}$ in the error-free case. We obtained 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. Figs. 9 and 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 et al. (1996) or Cao et al. (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 PI and BT 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 analysed by, for example, Stefanski and Carroll (1990) and Carroll et al. (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 et al. (1995)).

We applied the PI method, the NR method, the BT bandwidth and the CV 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 case $\sigma_Z^2=(1/5)\sigma_X^2$ or $\sigma_Z^2=(1/3)\sigma_X^2$. Each time we compare the results with the case $\sigma_Z^2 = 0$ (the error-free case).

The resulting estimators are depicted in Fig. 11. The curves indicated by the * characters in each picture represent the results for the error-free (EF) case.
a classical PI bandwidth selector. We see that the PI, NR and BT 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 1). By contrast the CV 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 \( (1.5\sigma^2) \). 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 the variance, the smoother the estimator.

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

Financial support from the contract ‘Projet d’Actions de Recherche Concertées’ Nr 98/03-217 from the Belgian government, and from the IAP research network Nr P5/24 of the Belgian State (Federal Office for Scientific, Technical and Cultural Affairs) is gratefully acknowledged.

The authors gratefully acknowledge the Editor, an Associate Editor and two referees for their valuable remarks on a first version of the paper, which led to a considerable improvement of the original manuscript.

References

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,