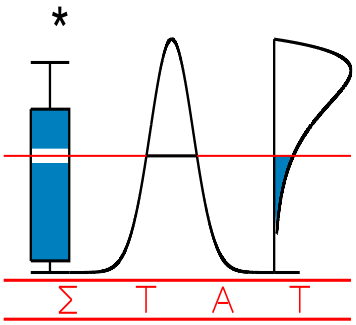
T E C H N I C A L

R E P O R T

0212

COMPARISON OF DATA-DRIVEN BANDWIDTH SELECTION PROCEDURES IN DECONVOLUTION KERNEL DENSITY ESTIMATION

A. DELAIGLE and I. GIJBELS



I A P S T A T I S T I C S

N E T W O R K

INTERUNIVERSITY ATTRACTION POLE

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

Comparison of data-driven bandwidth selection procedures 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.

March 2002

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 non-parametric 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 and phrases: bandwidth selection, bootstrap, cross-validation, deconvolution, errors-in-variables, kernel density estimation, plug-in methods.

Running title: deconvolution density estimation.

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.

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

Papers dealing with this practical bandwidth selection problem are Stefanski & Carroll (1990) and Hesse (1999) who propose a cross-validation procedure, and Delaigle & Gijbels (2001b) 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 the 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 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 the solution.

This paper is organized as follows. In Section 2 we recall the definition of the deconvolving kernel density estimator and of ordinary smooth and supersmooth error distributions. The choice of the kernel and the theoretical optimal bandwidth are also discussed. In Section 3 we propose two practical bandwidth selectors based on estimation of an (asymptotic) approximation of the mean integrated squared error: a normal reference and a plug-in bandwidth selector. Section 4 briefly discusses the other available data-driven bandwidth selectors: the cross-validation method of Stefanski & Carroll (1990) and the bootstrap procedure of Delaigle & Gijbels (2001b). The finite sample performances of the four 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

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

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 (2001b). 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 & Diggle (1995) and Neumann (1997).

Consider a kernel function K and a smoothing parameter h, depending on n, called the bandwidth. The deconvolving kernel estimator of f_X at $x \in I\!R$ 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 & Hall (1988) and Stefanski & Carroll (1990) for an introduction to this estimator. In order to guarantee that this estimator is well-defined we need the following conditions:

(C1). $\varphi_Z(t) \neq 0$ for all t;

(C2). $\displaystyle\int |\varphi_X(t)|\,dt < \infty$;

(C3). $\displaystyle\sup_t |\varphi_K(t)/\varphi_Z(t/h)| < \infty$, and $\displaystyle\int |\varphi_K(t)/\varphi_Z(t/h)|\,dt < \infty$.

Asymptotic properties of the deconvolving kernel estimator, such as consistency or rates of convergence to the target density have been studied in Carroll & Hall (1988), Devroye (1989), Stefanski (1990), Stefanski & Carroll (1990), Fan, Truong & Wang (1991), Fan (1991a,b,c) and Fan (1992). 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 & Hall (1988), Stefanski (1990), Stefanski & Carroll (1990), Fan (1991b,c) and Fan (1992).

2.2 Choice of the kernel

In kernel density estimation for error-free data, it has been shown that the choice of the kernel function K has not a big influence on the quality of the estimator. When we have errors in the data however, the particular structure of the estimator (2.1) requires some restrictions to be imposed on K. More precisely, we will impose conditions on φ_K, the Fourier transform of K.

For a supersmooth error distribution, condition (C3) will be satisfied as soon as we use a kernel with a compactly supported Fourier transform. This automatically 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. When the error distribution is ordinary smooth, K can be taken such that its Fourier transform is supported on $\mathbb{R}$ provided that the latter decreases rapidly enough when $t \to \infty$ (see Fan (1991c)).

Throughout this paper we will, for simplicity, use kernels with a compactly supported Fourier transform. This then immediately ensures the existence of all integrals. 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 are listed in Table 2.1.

Table 2.1: *Examples of kernels with compactly supported Fourier transform.*

kernel	characteristic function (Fourier transform)
$K_1(x) = \dfrac{\sin x}{\pi x}$	$\varphi_{K_1}(t) = 1_{[-1,1]}(t)$
$K_2(x) = \dfrac{48\cos x}{\pi x^4}\left(1 - \dfrac{15}{x^2}\right) - \dfrac{144\sin x}{\pi x^5}\left(2 - \dfrac{5}{x^2}\right)$	$\varphi_{K_2}(t) = (1-t^2)^3 1_{[-1,1]}(t)$
$K_3(x) = \dfrac{3}{8\pi}\left(\dfrac{\sin(x/4)}{x/4}\right)^4$	$\varphi_{K_3}(t) = \begin{cases} 2(\lvert t\rvert - 1)^3 & \text{for } 1/2 \le \lvert t\rvert \le 1 \\ 6\lvert t\rvert^3 - 6t^2 + 1 & \text{for } 0 \le \lvert t\rvert < 1/2 \end{cases}$

The kernel K_1 is the so-called sinc kernel, which is popularly-used in kernel density deconvolution. See for example Davis (1975), Tapia & Thompson (1978), Fan (1992) or Wand (1998).

2.3 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.

If conditions (C1) to (C3) hold and if K is integrable, then it has been shown that the MISE can be re-expressed as

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

where K_h is a rescaled version of K defined by $K_h(x) = h^{-1}K(x/h)$, and where $f * g$ denotes the convolution between two functions f and g. See Stefanski & Carroll (1990) and Wand & Jones (1995), among others.

The above expression 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 $\text{MISE}\{\widehat{f}_X(\cdot;h)\}$ as $\text{MISE}(h)$.

3 Plug-in type of bandwidth selection

3.1 Asymptotic approximation of the MISE

Stefanski & Carroll (1990) provide an asymptotic expansion for the MISE. See Theorem 1 below, which requires the following assumptions:

(C4). K is such that $\int yK(y)\,dy = 0$ and $0 \neq \int y^2K(y)\,dy < \infty$;

(C5). $h \to 0$ and $nh \to \infty$ as $n \to \infty$;

(C6). f_X'' exists and is bounded and square integrable.

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

Theorem 1. *(Stefanski & Carroll (1990))*

Under (C1) to (C6) and if K is such that $\int K(y)\,dy = 1$, we have

$$\mathrm{MISE}(h) = \mathrm{AMISE}(h) + o\left\{(nh)^{-1}\right\} + o\left(h^4\right)$$

where

$$\mathrm{AMISE}(h) = (2\pi nh)^{-1} \int |\varphi_K(t)|^2 . |\varphi_Z(t/h)|^{-2}\,dt + \frac{h^4}{4}\mu_{K,2}^2 R\left(f_X''\right) \ . \tag{3.1}$$

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 $R\left(f_X''\right)$, 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 $\widehat{\mathrm{AMISE}}(h)$. We then select the bandwidth which minimizes $\widehat{\mathrm{AMISE}}(h)$ with respect to h.

3.2 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. Based on the sample $Y_1, \cdots, Y_n$ 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 is not a consistent estimator for $R(f_X'')$. Although the order (rate) of the resulting bandwidth selector will not be affected by this estimation step (estimation of $R(f_X'')$ only influences the estimation of the constants appearing in the theoretical optimal bandwidth), 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 features such as, for example, strong multimodality and/or asymmetry. See also 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.

3.3 Plug-in bandwidth selection

Nonparametric estimation of $\theta_r = R(f_X^{(r)})$, with r any integer, when data are measured with errors, has been studied in detail by Delaigle & Gijbels (2001a). 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 & Gijbels (2001a) 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 this squared asymptotic bias (squared ABias), where for a k-th order kernel with k even

$$\text{ABias}[\widehat{\theta}_r] = (-1)^{k/2} \frac{2h_r^k}{k!} \mu_{K,k} \theta_{r+k/2} + (2\pi n h_r^{2r+1})^{-1} \int t^{2r} |\varphi_K(t)|^2 |\varphi_Z(t/h_r)|^{-2}\, dt. \quad (3.2)$$

Remark 1. Delaigle & Gijbels (2001a) show that in the ordinary smooth error case, the above mentioned regularity conditions are satisfied. The situation is more involved in the supersmooth error case, but from their simulation study it seems that even in that case, minimization of the squared asymptotic bias or of the asymptotic MSE lead to comparable results. Since taking into account the variance is computationally much more involved and requires some extra estimations, Delaigle & Gijbels (2001a) recommend to minimize the squared asymptotic bias.

Remark 2. 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 (2001b)). An application of Parseval's idendity leads to

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

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

From expression (3.2) we see that selecting h_2, the optimal bandwidth for the estimation of $\theta_2 = R(f_X'')$, 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 $\theta_{2+\ell\times k/2}$. We refer to this as an *ℓ-stages procedure*. A possible simple pilot estimator for $\theta_{2+\ell\times k/2}$ is obtained by referring to a normal density. One has to decide how many steps will be included in this iterative procedure to estimate θ_2. Typically, the more steps are included the smaller will be the bias, but the more variable will be the procedure. A trade-off between bias and variance has to be made. 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 & Gijbels (2001a), which for a second order kernel ($k = 2$) reads as follows

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

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.

Application of our bandwidth selection procedures in practice requires the estimation of $R(f_X'')$, followed by 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 $\widehat{\text{AMISE}}(h)$ by numerical integration and select the h that minimizes $\widehat{\text{AMISE}}(h)$ 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.

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 & Carroll (1990) relies on the fact that, under conditions (C1) to (C3), the cross-validation quantity

$$\mathrm{CV}(h) = \int \frac{|\varphi_K(ht)|^2 |\widehat{\varphi}_Y(t)|^2 + 2(n-1)^{-1}\varphi_K(-ht)\left[-n|\widehat{\varphi}_Y(t)|^2 + 1\right]}{2\pi|\varphi_Z(t)|^2}\, dt, \tag{4.1}$$

with $\widehat{\varphi}_Y(t)$ the empirical characteristic function of Y, is an unbiased estimator of $\mathrm{MISE}(h) - \int f_X^2(x)\, dx$. This motivates the *cross-validated bandwidth selector*, obtained via minimization of (4.1). Stefanski & Carroll (1990) applied this method in the context of Gaussian errors and using the particular sinc kernel (K_1 in Table 2.1). 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.

In our extensive simulation study we applied the procedure for a variety of cases, and encountered problems such as non-uniqueness of the solution or sometimes even very poor performance of the estimator. Moreover the resulting density estimator was generally quite variable. See Section 5. Note that these problems also arise when applying the method in the error-free case. See Delaigle (1999) for a more complete description of the numerical problems encountered in practice. 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. Such a recommendation is of course of no use in a simulation study, since 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.

4.2 Bootstrap bandwidth selection

The bootstrap method of Delaigle & Gijbels (2001b) selects the bandwidth through minimization of the following consistent bootstrap estimator of the MISE:

$$\begin{aligned}\mathrm{MISE}^*(h) =& (2\pi nh)^{-1}\int |\varphi_K(t)|^2 |\varphi_Z(t/h)|^{-2}\, dt + (1-n^{-1})(2\pi)^{-1}\int |\widehat{\varphi}_{X,g}(t)|^2 |\varphi_K(ht)|^2\, dt \\ &+ R(\widehat{f}_X(.;g)) - \pi^{-1}\int |\widehat{\varphi}_{X,g}(t)|^2 \varphi_K(ht)\, dt,\end{aligned} \tag{4.2}$$

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 & Gijbels (2001b) 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 & Gijbels (2001b) and choose g to be the two-stages bandwidth h_2 of Section 3.3.

Note that 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, (4.2) can be computed rather easily directly from the original sample. Note that the minimizer of $\text{MISE}^*(h)$ with respect to h is also the minimizer of

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

and thus in practice we do not even need to compute the estimator $\widehat{f}_X(.;g)$ of f_X.

In practice all integrals involved will be computed by numerical integration and the bandwidth will be that value on a grid which minimizes $\text{MISE}_2^*(h)$. The finite sample performance of this procedure is compared with the performances of the other practical bandwidth selectors in the next section.

5 Simulation results

Simulations were performed for all above methods with kernel K_2, two different error densities (N(0,σ_Z^2) and Laplace(σ_Z)) and three sample sizes (50, 100 and 250). We consider four different target densities f_X. They have been chosen because they present each a particular feature that can be encountered in practice. These densities are, in increasing order of estimation difficulty

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 #4: $X \sim 0.4$ Gamma(5)+0.6 Gamma(13), which is bimodal, but with close and different sized modes.

These densities are represented in Figure 5.1, and for simplicity, we will often refer to them as densities (1), (2), (3) and (4) in the following.

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 $\mathrm{Var}(Z)/\mathrm{Var}(X)$. For densities (1)—(3) this ratio was set to 25%, but for density (4) this ratio was set to 10%, since for this density there are more particular features to recover.

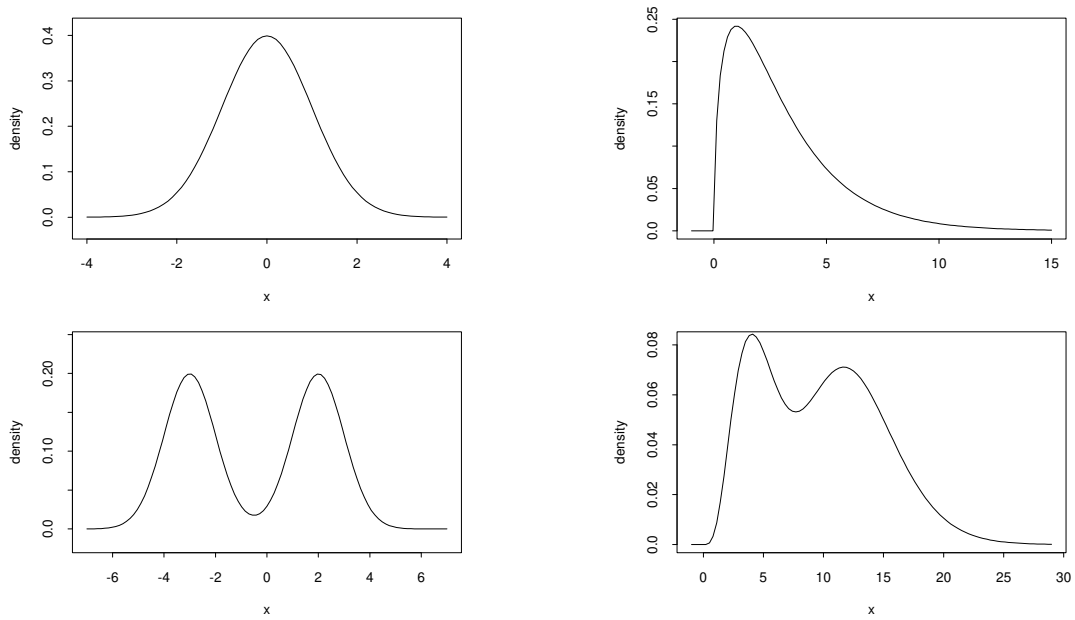


Figure 5.1: *The four target densities: a normal density (top left), a chi-squared density (top right), a mixture of two normal densities (bottom left), and a mixture of two gamma densities (bottom right).*

We carried out a more extensive simulation study by considering other target densities and 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. In the tables we report on the empirical median and interquartile range of the 500 estimated bandwidths. 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). For all cases we also calculated the exact value of h_{MISE}, which we indicate in the tables for comparison purpose.

In each table we compare the results obtained from the 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). In order to show the behaviour of the ISE-quantity across the simulations,

we present in each case a kernel density estimate based on the 500 values of ISE($\widehat{h}$), using a normal reference bandwidth and a standard normal kernel. 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.

Table 5.1: *Simulation results for the $N(0;1)$ density, with Laplace error density.*

	$n = 50$		$n = 100$		$n = 250$	
	$h_{\text{MISE}} = 0.2433$		$h_{\text{MISE}} = 0.2161$		$h_{\text{MISE}} = 0.1865$	
Method	med(h)	IQR	med(h)	IQR	med(h)	IQR
NR	0.23	0.031	0.21	0.020	0.18	0.011
PI	0.23	0.045	0.21	0.034	0.18	0.021
CV	0.26	0.072	0.24	0.065	0.21	0.058
BT	0.25	0.043	0.22	0.029	0.19	0.018

Put Figure 5.2 here

Table 5.1 compares, for the N(0;1) target density and Laplace 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 almost 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 ISE($\widehat{h}$) values, showing that these errors are close to zero, and that the variance of their distribution decreases when n increases.

Table 5.2: *Simulation results for the $\chi^2(3)$ density with $n = 100$.*

	Laplace error			Gaussian error		
Method	h_{MISE}	med(h)	IQR	h_{MISE}	med(h)	IQR
NR	0.3555	0.50	0.073	0.4455	0.54	0.060
PI		0.41	0.058		0.47	0.048
CV		0.41	0.14		0.51	0.13
BT		0.45	0.057		0.52	0.048

Put Figures 5.3 and 5.4 here

From Table 5.2 we can see that when estimating the $\chi^2(3)$ density from a sample of size 100 and a Laplace 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 still gives the poorest results. Similar conclusions can be drawn for the Gaussian error case, but as to be expected the resulting density estimates are less performant. See also Figures 5.3 and 5.4, for an evaluation of the ISE-values and the resulting estimates of the $\chi^2(3)$ density.

Table 5.3: *Simulation results for mixed normal density with* $n = 100$.

	Laplace error			Gaussian error		
Method	h_{MISE}	med(h)	IQR	h_{MISE}	med(h)	IQR
NR	0.3386	0.57	0.042	0.4154	0.60	0.031
PI		0.41	0.057		0.50	0.056
CV		0.37	0.13		0.51	0.12
BT		0.46	0.081		0.56	0.078

Put Figures 5.5 and 5.6 here

A similar conclusion can be drawn from Table 5.3 and Figures 5.5 and 5.6, where, for $n = 100$, we tried to recover the mixed normal density contaminated by a Laplace or a Guassian error. In this case the density is far from a normal, and the plug-in method brings improvements over the NR method. The bootstrap method works well and the cross-validation technique remains very variable.

Table 5.4: *Simulation results for the mixed gamma density,* $n = 250$*, with Laplace error.*

Method	h_{MISE}	med(h)	IQR
NR	0.5766	0.81	0.033
PI		0.69	0.094
CV		0.62	0.23
BT		0.78	0.10

Put Figures 5.7 and 5.8 here

Table 5.4 and Figures 5.7 and 5.8 show that even in the more involved mixed gamma density case, 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.

From our simulation results we conclude that, in general, all presented methods seem to work reasonably well, even in the mixed density cases. It is not clear which method performs the best, but yet we can make a few remarks. First the plug-in procedure using the method of Delaigle & Gijbels (2001a) for estimating $R(f_X'')$ brings improvements over the simple normal reference rule, since in the cases where the NR method is not very effective (the strongly mixed or strongly skewed densities cases), the plug-in method works much better, but, moreover, this plug-in method works also in the cases where the NR method would be the most appropriate one. Second, whatever the case, the bootstrap bandwidth selector always performed quite well. Note that the plug-in bandwidth is in general smaller than the bootstrap bandwidth. We noted that this also seems to be the case for non-contaminated data as can be seen from the simulation results presented by Cao, Cuevas & González-Manteiga (1994). By contrast, the cross-validation technique seems to be far more variable than all other techniques, and sometimes tends to select too small bandwidths, giving rise to rough estimates. In practice, this method also suffers from drawbacks, such as non-uniqueness of the solution or a percentage of totally unacceptable results.

Finally, we conclude that no method proved to be the best in all cases, but, nevertheless, the bootstrap and plug-in 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 & Carroll (1990) and Carroll, Ruppert & 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 & Carroll (1990) or Carroll, Ruppert & 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 σ_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 & 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-free case).

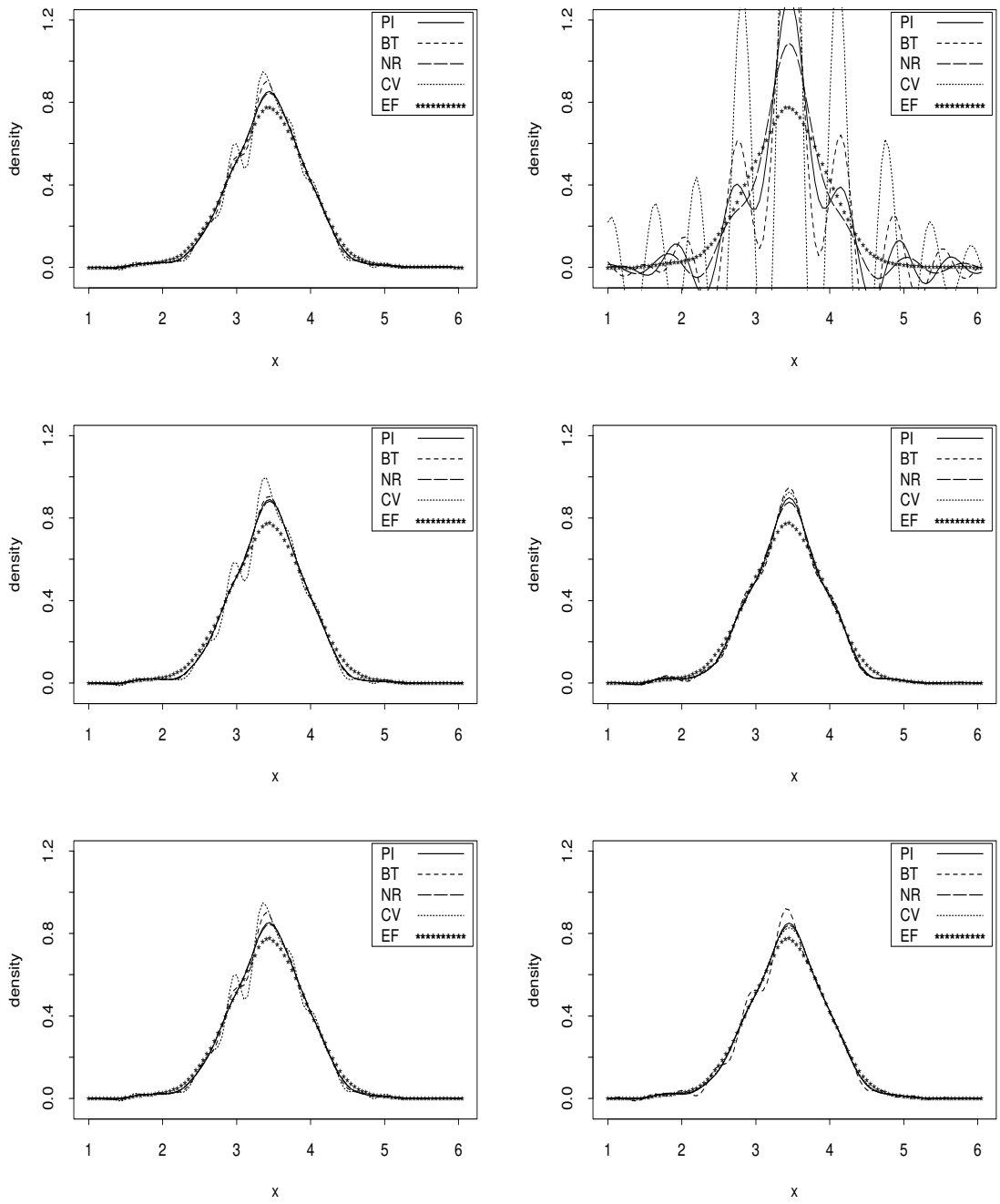


Figure 6.1: *Estimation of the log saturated fat density for a Laplace error (left panels) with* $\sigma_Z^2 = 1.5\,\sigma_X^2$ *(top left),* $\sigma_Z^2 = (1/3)\,\sigma_X^2$ *(center left),* $\sigma_Z^2 = (1/5)\,\sigma_X^2$ *(bottom left), and for a normal error (right panels) with* $\sigma_Z^2 = 1.5\,\sigma_X^2$ *(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.*

The resulting estimators are depicted in Figure 6.1. 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 in the noise ($1.5\ \sigma_X^2$). Note that Stefanski & Carroll (1990) encountered similar problems in deconvolution 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 & Carroll (1990).

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.

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.

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. (2001a). Estimation of integrated squared density derivatives from a contaminated sample. *Discussion Paper* #0101, Institut de Statistique, Université catholique de Louvain, Belgium. http://www.stat.ucl.ac.be

Delaigle, A. and Gijbels, I. (2001b). 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

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 tractions on Information Theory*, to appear.

Fan, J., Truong, Y.K. and Wang (1991). Nonparametric function estimation involving errors in variables. In *Nonparametric estimation and related topics* (G. Roussas, ed.), 613–327.

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

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. *Nonparametric Statistics*, **7**, 307–330.

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.

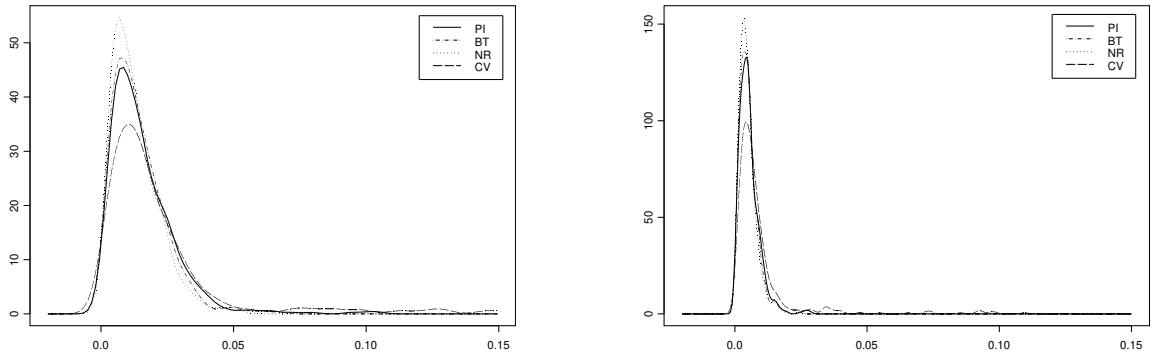


Figure 5.2: *Performance of the four bandwidth selectors for estimation of the* $N(0;1)$ *density. Kernel density estimates of the values ISE*$(\widehat{h})$ *across the 500 simulations, for sample sizes* $n = 50$ *(left) and* $n = 250$ *(right).*

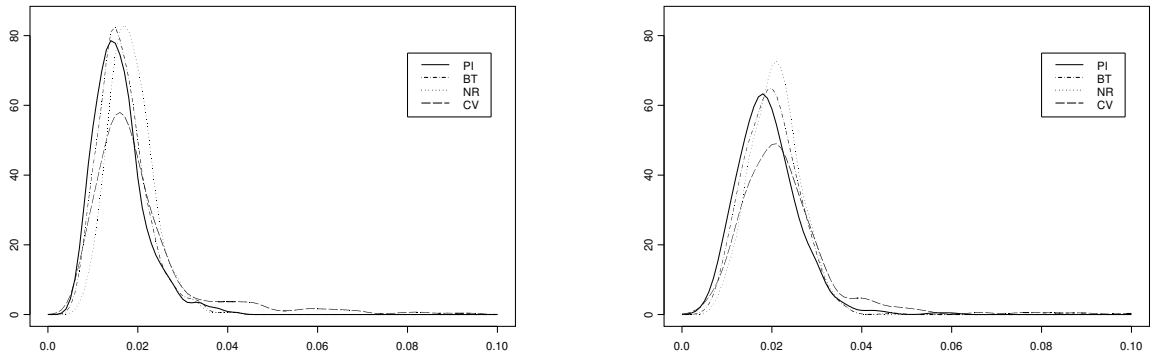


Figure 5.3: *Performance of the four bandwidth selectors for estimation of the* $\chi^2(3)$ *density. Kernel density estimates of the values ISE*$(\widehat{h})$ *across the 500 simulations, for sample size* $n = 100$ *and Laplace errors (left) and Gaussian errors (right).*

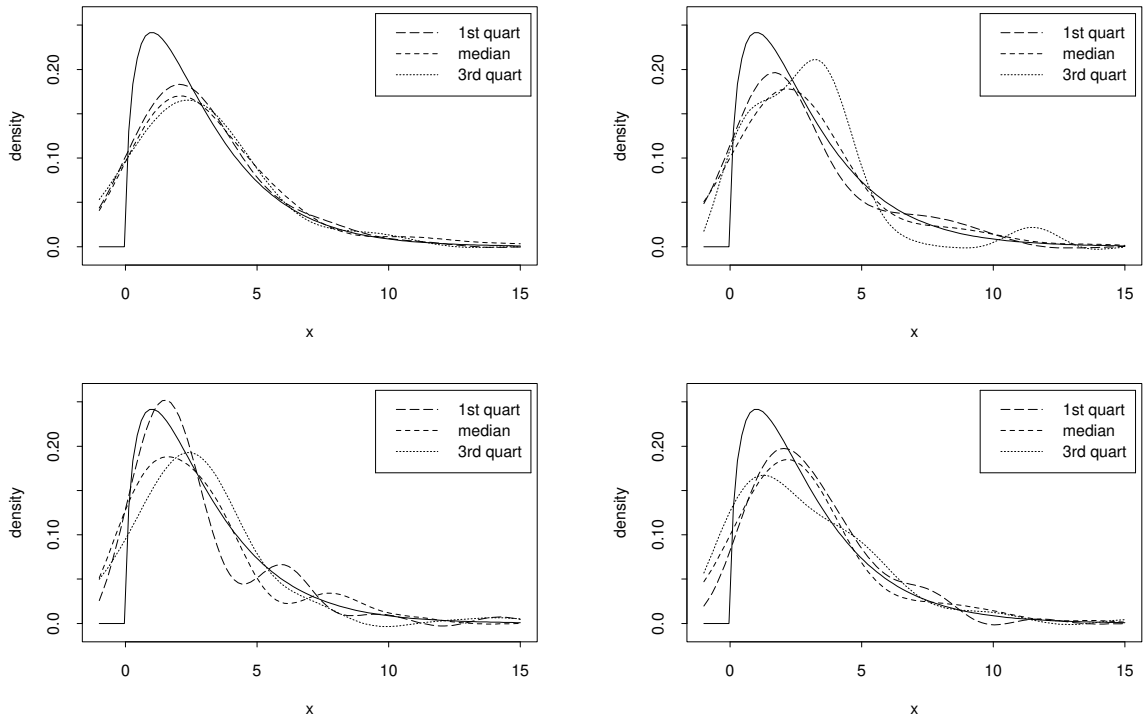


Figure 5.4: *Estimation of the* $\chi^2(3)$ *density for a sample of size* $n = 100$ *and a Laplace error by the normal reference method (top left), the plug-in method (bottom left) ,the cross-validation method (top right), and the bootstrap method (bottom right).*

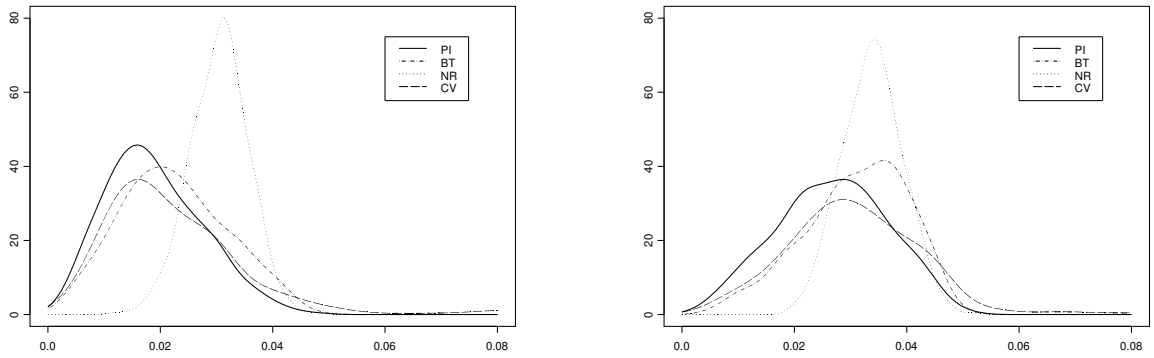


Figure 5.5: *Performance of the four bandwidth selectors for estimation of the mixed normal density. Kernel density estimates of the values ISE*$(\widehat{h})$ *across the 500 simulations, for sample size* $n = 100$ *and Laplace errors (left) and Gaussian errors (right).*

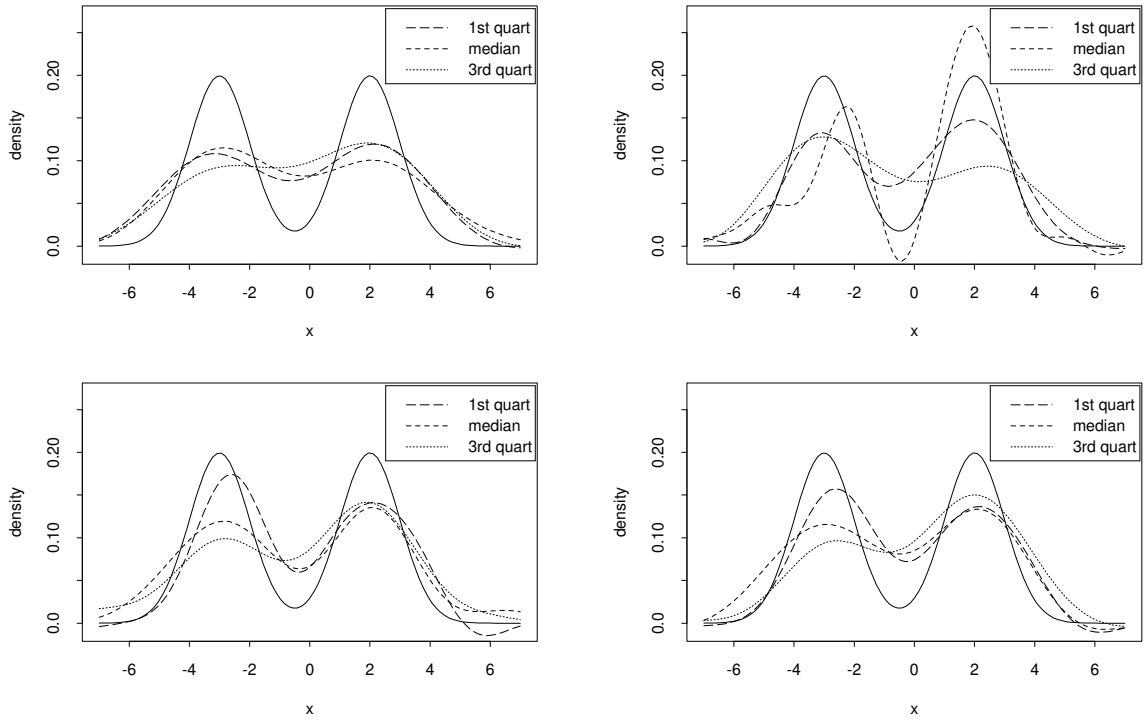


Figure 5.6: *Estimation of the mixed normal density for a sample of size* $n = 100$ *and a Laplace error by the normal reference method (top left), the plug-in method (bottom left), the cross-validation method (top right), and the bootstrap method (bottom right).*

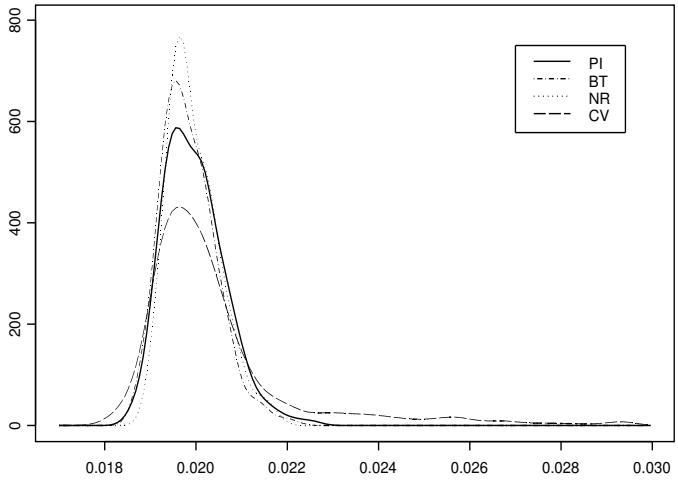


Figure 5.7: *Performance of the four bandwidth selectors for estimation of the mixed gamma density. Kernel density estimates of the values ISE*$(\widehat{h})$ *across the 500 simulations, for sample size* $n = 250$ *and Laplace errors.*

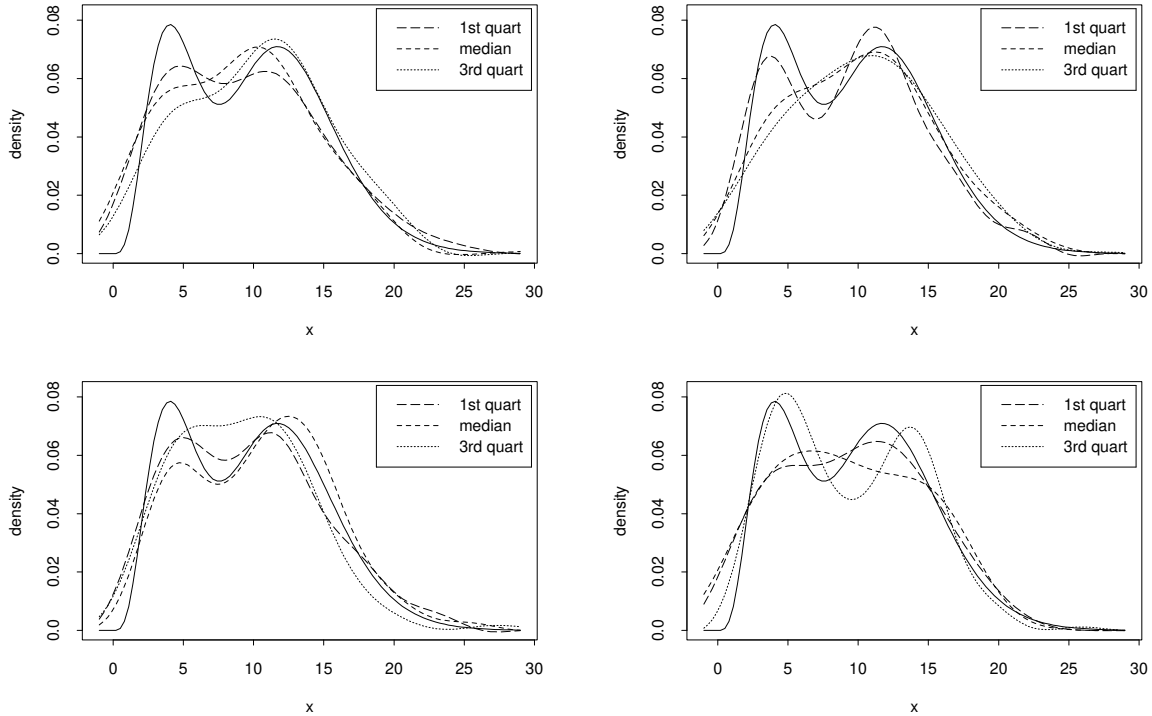


Figure 5.8: *Estimation of the mixed gamma density for a sample of size* $n = 250$ *and a Laplace error by the normal reference method (top left), the plug-in method (bottom left), the cross-validation method (top right), and the bootstrap method (bottom right).*