

# Frequent problems in calculating integrals and optimizing objective functions: a case study in density deconvolution

A. Delaigle<sup>1,2</sup> and I. Gijbels<sup>3</sup>

<sup>1</sup>Department of Mathematics, University of Bristol, Bristol, BS8 4JS, UK;

<sup>2</sup>Department of Mathematics and Statistics, University of Melbourne, Parkville, VIC, 3010, Australia; and <sup>3</sup>Department of Mathematics and University Center for Statistics, Katholieke Universiteit Leuven, Belgium.

## Abstract

Many statistical procedures involve calculation of integrals or optimization (minimization or maximization) of some objective function. In practical implementation of these, the user often has to face specific problems such as seemingly numerical instability of the integral calculation, choices of grid points, appearance of several local minima or maxima, etc. In this paper we provide insights into these problems (why and when are they happening?), and give some guidelines of how to deal with them. Such problems are not new, neither are the ways to deal with them, but it is worthwhile to devote serious considerations to them. For a transparent and clear discussion of these issues, we focus on a particular statistical problem: nonparametric estimation of a density from a sample that contains measurement errors. The discussions and guidelines remain valid though in other contexts. In the density deconvolution setting, a kernel density estimator has been studied in detail in the literature. The estimator is consistent and fully data-driven procedures have been proposed. When implemented in practice however, the estimator can turn out to be very inaccurate if no adequate numerical procedures are used. We review the steps leading to the calculation of the estimator and in selecting parameters of the method, and discuss the various problems encountered in doing so.

*Key words and phrases:* Bandwidth selection, dramatic cancellation, fast Fourier transform, numerical approximations, optimization.

## 1 Introduction

We consider kernel estimation of a density from a sample  $Y_1, \dots, Y_n$  of independent and identically distributed (i.i.d.) random variables corrupted by measurement errors. More

precisely, for  $i = 1, \dots, n$ , we assume that  $Y_i = X_i + Z_i$ , where  $X_i \sim f_X$  is the variable of interest, and  $Z_i \sim f_Z$  ( $f_Z$  known) represents a measurement error independent of  $X_i$ . In this context, Carroll and Hall (1988) and Stefanski and Carroll (1990) proposed a so-called deconvolution kernel estimator of the density  $f_X$ . Let  $K$  be a kernel function that integrates to 1 and  $0 < h = h_n \rightarrow 0$ ,  $n \rightarrow \infty$  be a smoothing parameter called the bandwidth. Under sufficient regularity conditions, the deconvolution kernel density estimator of  $f_X$  is defined by

$$\widehat{f}_X(x; h) = \frac{1}{nh} \sum_{j=1}^n K_Z\left(\frac{x - Y_j}{h}; h\right), \quad (1.1)$$

where

$$K_Z(x; h) = \frac{1}{2\pi} \int e^{-itx} \varphi_K(t) / \varphi_Z(t/h) dt, \quad (1.2)$$

and  $\varphi_g$  represents the Fourier transform or characteristic function of a function or density  $g$ . When  $\widehat{f}_X(x; h)$  is not real, the estimator is defined as the real part of (1.1). Asymptotically this does not make any difference.

Theoretical properties of the estimator have been studied in many papers including Carroll and Hall (1988), Stefanski and Carroll (1990), Fan (1991a,b,c), Fan (1992) and Masry (1993a,b). Recent papers include Zhang and Karunamuni (2000), Meister (2004, 2006) and van Es and Uh (2004, 2005). It is well known that both the asymptotic and finite sample performance of the estimator depend crucially on the value of the bandwidth. Fully data-driven bandwidth selectors have been proposed by Stefanski and Carroll (1990), Hesse (1999) and Delaigle and Gijbels (2002, 2004a,b). An extensive simulation study conducted by Delaigle and Gijbels (2004b) shows that, when combined with these bandwidth selectors, the deconvolution kernel density estimator has good finite sample performance.

Although, in theory, the estimator and its bandwidth selectors work well, the success of their practical implementations very much relies on adequate use of appropriate techniques for calculating integrals and for optimizing objective functions. In particular, direct implementation of some formulas available for the estimator and its bandwidth can produce inaccurate results and we show how to deal with this. We will see that calculating the estimator generally requires numerical approximations because, even when analytic expressions are available, they may lead to erroneous results in practice. Next, we tackle the minimization problem associated to bandwidth selection. In this deconvolution context, one often has to minimize a function approximated by numerical means, and this can be done by performing a grid search. We discuss an automatic choice of the grid based on a normal reference procedure and illustrate the impact of the grid for different bandwidth selectors. In particular, we show that the cross-validation procedure is extremely

sensitive to the grid choice and that caution is needed here. Obviously, the problems investigated here might also appear in other contexts, and the chosen context serves as a working platform. The aim of this paper is to report on these practical implementation problems and discuss some approaches to deal with them. In addition, this will be a very useful guideline for people who want to use the deconvolution kernel density estimator in practice.

This paper is organized as follows. In Section 2, we discuss practical implementation of the estimator. We show why analytic expressions cannot always be used and discuss the choice of appropriate methods to evaluate the estimator numerically. In Section 3, we deal with issues on the calculation of data-driven bandwidths in practice, referring to optimization of objective functions in general.

## 2 Calculation of the estimator in practice

Calculation of the estimator in practice requires evaluation of the integral in (1.2), which is usually not easy to calculate. In order to avoid problems of integrability, it is rather common to work with a kernel  $K$  that has a compactly supported Fourier transform  $\varphi_K$ . Two popular kernels in this context are the sinus cardinal (sinc) kernel  $K_1(x) = \sin(x)/x$  and the kernel  $K_2(x) = 48 \cos x (1 - 15x^{-2})/(\pi x^4) - 144 \sin x (2 - 5x^{-2})/(\pi x^5)$ , with characteristic functions  $\varphi_{K_1}(t) = 1_{[-1,1]}(t)$  and  $\varphi_{K_2}(t) = (1 - t^2)^3 1_{[-1,1]}(t)$ , where  $1_A$  denotes the indicator function on a set  $A$ , i.e.  $1_A(x) = 1$  for  $x \in A$  and 0 otherwise. Delaigle and Hall (2006) show that, despite its unconventional shape, the kernel  $K_2$  is very appropriate for deconvolution problems, and we therefore focus our attention on this type of kernels.

### 2.1 Numerical approximations

A first, frequent scenario is that no closed form expression can be found for the integral in (1.2). In cases where  $\varphi_Z$  behaves like a negative exponential in the tails (the so-called supersmooth case), the integral involves at least a sine/cosine function, an exponential function and the kernel  $K$ . A simple example is for a  $N(0; \sigma^2)$  error density combined with the kernel  $K_2$ , where  $K_Z(x; h)$  is proportional to  $\int_{-1}^1 \cos(tx)(1 - t^2)^3 \exp(t^2\sigma^2/(2h^2)) dt$ .

Obviously, the function to integrate is oscillating and cannot be approximated accurately by usual fast iterative numerical methods, such as Romberg's method for example, especially if successive iterations evaluate the integrand at grids of equidistant points  $x_i$  where successive grids are simply augmented by the middle points  $(x_i + x_{i+1})/2$ . For  $x$  a multiple of  $2\pi$ , if the initial grid is not large enough, the points of the first few successive grids are located at maxima of  $\cos(tx) \cdot \varphi_K(t)/\varphi_Z(t/h)$  and the integrand is mistakenly

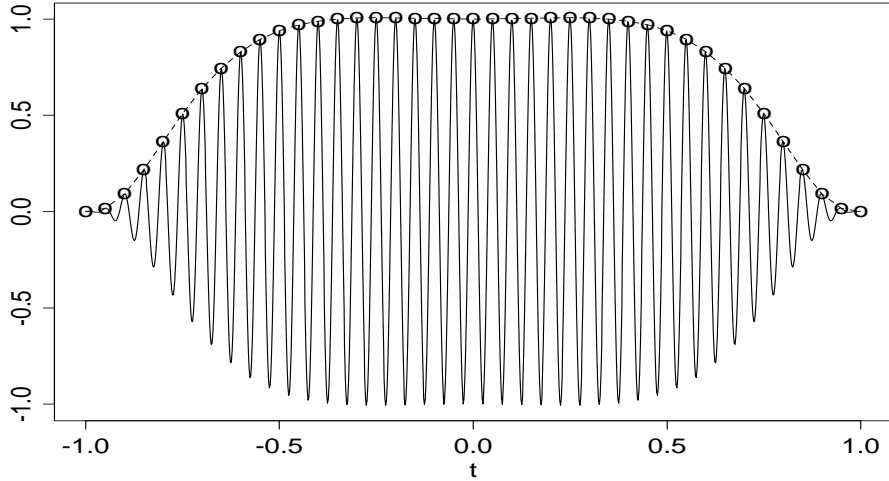


Figure 2.1: Plot of  $\cos(tx)(1 - t^2)^3 \exp(t^2\sigma^2/(2h^2))$  for  $x = 40\pi$  with  $\sigma^2/h^2 = 6.4$ ; the dashed curve represents the function used by the numerical method of integration.

confused with a slowly varying function passing through these points, giving a completely wrong approximation of the integral.

In Figure 2.1 for example, we present the function  $\cos(tx)(1 - t^2)^3 \exp(t^2\sigma^2/(2h^2))$  for  $x = 40\pi$  (the solid curve), with  $\sigma^2/h^2 = 6.4$ . Figure 2.1 also depicts, as circles, the corresponding equidistant grid points of successive iterations for an initial 11 points grid, and the wrong approximated function used by the numerical method (dashed curve). In order to avoid such dramatic numerical problems,  $K_Z$  must be calculated using numerical methods devoted to approximations of Fourier transforms, such as, for example the Fast Fourier Transform (FFT). See for example Press et al. (1992), Chapter 13.

## 2.2 Analytic expressions

Although the simplest case seems to be when analytic expressions of  $K_Z$  are available, we show here that these cannot always be used in practice. We illustrate the potential numerical problems on a simple example, namely in the Laplace ( $\sigma$ ) error case with  $f_Z(z) = (2\sigma)^{-1}e^{-|z|/\sigma}$ , for all  $z$ . For the sinc kernel ( $K_1$ ), straightforward calculations lead to

$$K_Z(x; h) = \left\{ \frac{1}{\pi} \frac{\sin x}{x} + \frac{\sigma^2}{\pi h^2} \left( \frac{\sin x}{x} + 2 \frac{\cos x}{x^2} - 2 \frac{\sin x}{x^3} \right) \right\} \cdot 1_{\{x \neq 0\}} + \left\{ \frac{1}{\pi} + \frac{\sigma^2}{3\pi h^2} \right\} \cdot 1_{\{x=0\}} \quad (2.3)$$

whereas for the kernel  $K_2$ , we have

$$K_Z(x; h) = \frac{48}{\pi} \left\{ \frac{\sigma^2}{h^2} \left[ \frac{\cos x}{x^8} (x^4 - 95x^2 + 840) + \frac{\sin x}{x^9} (-14x^4 + 375x^2 - 840) \right] \right\}$$

$$+ \left[ \frac{\cos x}{x^4} - 15 \frac{\cos x}{x^6} - 6 \frac{\sin x}{x^5} + 15 \frac{\sin x}{x^7} \right] \cdot 1_{\{x \neq 0\}} + \left\{ \frac{16\sigma^2}{315\pi h^2} + \frac{16}{35\pi} \right\} \cdot 1_{\{x=0\}}. \quad (2.4)$$

These expressions seem very simple but, in practice, they cannot be used for small  $x$ , as explained below. Let  $T_n(g(x))$  denote the  $n$ th order Taylor expansion of a function  $g(x)$  around zero. The above expressions for  $K_Z$  are coming from linear combinations of integrals of the form  $\int_0^1 t^{2j} \cos(tx) dt$ , with  $j$  a positive integer. Let's look first at the simple example of the integral  $\int_0^1 t^2 \cos(tx) dt$ . By integration by parts, it is easy to see that

$$\int_0^1 t^2 \cos(tx) dt = \frac{2}{x^3} \left[ \left( \frac{x^2}{2} - 1 \right) \sin x + x \cos x \right] = \frac{2}{x^3} [-\sin x T_2(\cos x) + \cos x T_1(\sin x)].$$

More generally, for any positive integer  $j$ , we have, by integration by parts,

$$\int_0^1 t^{2j} \cos(tx) dt = \frac{(-1)^{j+1} (2j)!}{x^{2j+1}} [-\sin x \cdot T_{2j}(\cos x) + \cos x \cdot T_{2j-1}(\sin x)], \quad (2.5)$$

where, for  $x$  close to zero, (2.5) subtracts two close values, and multiplies the difference by a very large value ( $x^{-2j-1}$ ). Evaluation of the difference by a computer can only be approximated by a number with a finite number of digits which, when multiplied by  $x^{-2j-1}$ , can lead to huge errors of approximations called dramatic cancellation. As a toy example, a computer that would work with only 6 digits would approximate  $(1.0000011 - 1.000001) \times 10^{20} = 10^{13}$  by  $(1.000001 - 1.000001) \times 10^{20} = 0$ .

This problem has a huge impact on the calculation of  $K_Z$  for the kernels  $K_1$  and  $K_2$ , and, more generally for kernels whose characteristic function is a combination of powers of  $t^2$  on a finite support. See our discussion about Figure 2.2 below. Hence, in the neighborhood of  $x = 0$ , we need to use an appropriate numerical approximation such as e.g. the FFT.

Alternatively, the functions  $\sin x$  and  $\cos x$  in (2.5) could be approximated by their Taylor expansion of order, respectively,  $2j + 3$  and  $2j + 2$ . The resulting approximations for  $K_Z$  for the kernels  $K_1$  and  $K_2$  are respectively  $K_Z(x; h) = \pi^{-1}[1 - x/2 + \sigma^2/h^2(1/3 - x^2/10)] + O(x^3/h^2)$ , and  $K_Z(x; h) = \pi^{-1}[16/35 - 8/315x^2 - \sigma^2/h^2(16/315 + 8/1155x^2)] + O(x^3/h^2)$ . Of course, better approximations of  $K_Z$  can be obtained by using higher order Taylor expansions for  $\sin x$  and  $\cos x$ , but, in any case, these will always only be valid in a neighborhood of  $x = 0$ . Figure 2.2 (left panel) compares the curves obtained by calculating  $K_Z$  for the kernel  $K_2$  using the three methods. The analytic curve and the

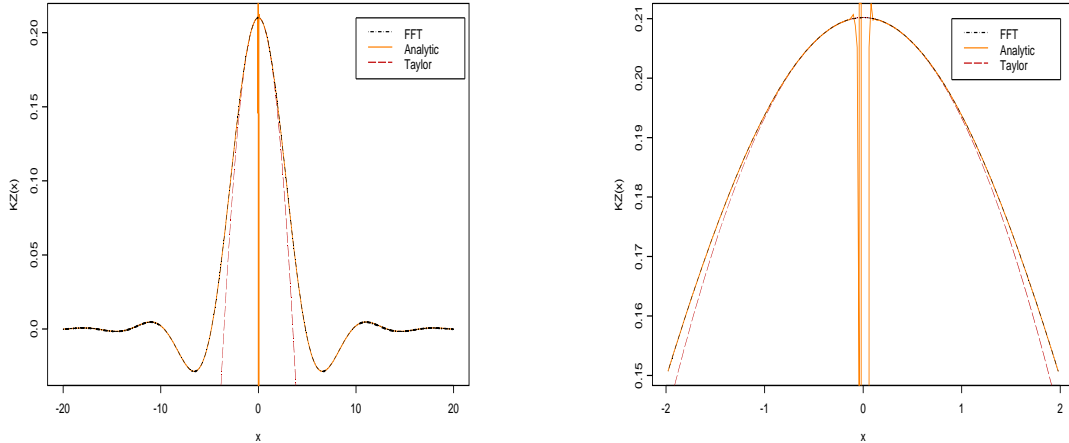


Figure 2.2: Comparison of  $K_Z$  in the Laplace error case for the kernel  $K_2$ , using the analytic expression (2.4), the FFT approximation or a Taylor expansion. The right panel shows a zoom-in of the middle top part of the figure on the left-hand side.

FFT curve are confounded almost everywhere, but in the neighborhood of  $x = 0$ , the analytic expression provides completely erroneous values. This is clearly seen from the right panel in Figure 2.2 which shows a zoom-in of the middle top part of the picture in the left panel. Further, without any surprise, the Taylor approximation is only good in a neighborhood of  $x = 0$ .

Our discussion illustrates the care with which analytic expressions sometimes need to be handled. The calculation of  $K_Z$  in this particular study is just an example. In the same context, for other more general errors/kernels, similar problems might occur. Careful analysis has to be performed before using any analytic expression. In case of potential problems and/or doubts, the FFT method seems to be an appropriate alternative, which is quite efficient and can be calculated reasonably fast, and this for any value of  $x$ . It would also be possible to combine the analytic expression for certain values of  $x$  with a Taylor approximation for other values of  $x$ . Here one has to determine the values  $x$  where each method can be applied.

### 3 Bandwidth selection

#### 3.1 Calculation of objective functions to be optimized

Delaigle and Gijbels (2004b) propose several ways to select the bandwidth in practice, based on minimization of a consistent estimator of the Mean Integrated Squared Error (MISE) of the estimator, where  $\text{MISE}(h) = E \int [\widehat{f}_X(x; h) - f_X(x)]^2 dx$ . They propose a normal reference method, a bootstrap method and a plug-in method, and compare

the procedures with the cross-validation method of Stefanski and Carroll (1990). Their extensive simulation results show that the plug-in bandwidth seems to perform uniformly better. In this section, we show how the quantities involved in the plug-in bandwidth selection procedure can be handled more easily when written in their Fourier domain.

The plug-in (PI) bandwidth is obtained by minimization of the following estimator of the asymptotic MISE

$$\widehat{\text{AMISE}}(h) = (nh)^{-1} \int \{K_Z(x; h)\}^2 dx + \frac{h^4}{4} \mu_{K,2}^2 \widehat{\theta}_2, \quad (3.6)$$

where  $\mu_{K,2} = \int x^2 K(x) dx$ ,  $\widehat{\theta}_r = \int \{\widehat{f}_X^{(r)}(x; h_r)\}^2 dx$ , with  $\widehat{f}_X^{(r)}$  representing the  $r$ th derivative of the density estimator (1.1) and where  $h_r$  is the bandwidth optimal for estimation of  $\theta_r = \int \{f_X^{(r)}(x)\}^2 dx$  by  $\widehat{\theta}_r$ . Calculation of a well-performing PI bandwidth in practice requires evaluation of  $\widehat{\theta}_r$  for different values of  $r$ . For a second order kernel  $K$  for example, we need to calculate  $\widehat{\theta}_r$  for  $r = 2$  and  $3$ , where in each case, the bandwidth  $h_r$  is selected by minimization of the quantity

$$D(h_r) = -h_r^2 \mu_{K,2} \widehat{\theta}_{r+1} + (nh_r^{2r+1})^{-1} \int \{K_Z^{(r)}(x; h)\}^2 dx. \quad (3.7)$$

The complete PI bandwidth selection procedure requires the calculation of  $\int \{K_Z^{(r)}(\cdot; h)\}^2$ , for  $r = 0, 2$  and  $3$  and  $\widehat{\theta}_r$  for  $r = 2$  and  $3$ . Direct calculation necessitates numerical integration of  $\{\widehat{f}_X^{(r)}(\cdot; h_r)\}^2$  and  $\{K_Z^{(r)}(\cdot; h)\}^2$  over the whole real line, which, in general, proves to be a difficult and inaccurate task, partly because the integration bounds are infinite and  $\widehat{f}_X^{(r)}$  and  $K_Z^{(r)}$  are themselves defined by an integral which can often only be approximated numerically.

The PI bandwidth can be calculated accurately and rather easily when all integrals in (3.6), (3.7) and  $\widehat{\theta}_r$  are transformed in their Fourier domain. Application of Parseval's identity leads to

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

where  $\widehat{\varphi}_Y = n^{-1} \sum_{j=1}^n e^{itY_j}$  denotes the empirical characteristic function of  $Y$ , and

$$\int \{K_Z^{(r)}(x; h)\}^2 dx = (2\pi n h_r^{2r+1})^{-1} \int t^{2r} |\varphi_K(t)|^2 |\varphi_Z(t/h_r)|^{-2} dt.$$

These integrals are considerably easier to calculate since now the integration bounds are finite when  $\varphi_K$  is compactly supported (as is the case for e.g. the popular kernels  $K_1$  and  $K_2$ ).

## 3.2 Minimization procedures

The PI bandwidth, as well as the other practical bandwidth selectors, require minimization of an approximation of the MISE, say  $\widehat{\text{MISE}}$  for convenience. In the PI case, an analytic expression can sometimes be found for the  $\widehat{\text{MISE}}$ . For example, for the kernel  $K_2$  and a Laplace( $\sigma$ ) error, simple calculations show that  $h$  is the solution of  $-256/153153 \sigma^4 - 512/45045 \sigma^2 h^2 - 256/9009 h^4 + 3\widehat{\theta}_2 \pi n h^9 = 0$ . In this case, the bandwidth  $h$  is thus the solution of a 9th order polynomial and can be very easily approximated numerically by most mathematical softwares. In many cases however, like for example a normal error and a kernel such as  $K_2$ , the integral  $\int |\varphi_K(t)|^2 |\varphi_Z(t/h)|^{-2} dt$  cannot be calculated analytically and has to be approximated by numerical integration. If the integrand is very smooth and does not oscillate much, a fast method such as Romberg's method can be used. In other cases, the slow but more robust trapezoid rule is more appropriate.

More generally, for many bandwidth selection methods, the objective function  $\widehat{\text{MISE}}$  can only be approximated numerically and, as a result, its dependence in  $h$  is unclear. In order to find the minimum, one has to perform a grid search, i.e. on a grid of values of  $h$ , search for the bandwidth that minimizes  $\widehat{\text{MISE}}$ . A first task is thus to fix an interval where to search for the solution. For the PI, bootstrap and normal reference approximations of the MISE, this step is not crucial because  $\widehat{\text{MISE}}$  is generally a convex function (if no solution is found in the selected interval, the solution will be found after enlarging the interval). Whereas the calculations involved with the normal reference bandwidth  $\widehat{h}_{\text{NR}}$  of Delaigle and Gijbels (2004b) are simple and very fast, a good choice of an initial grid can reduce considerably the computational time involved in the PI and bootstrap procedures. For these reasons, we propose to

- 1). Search for  $\widehat{h}_{\text{NR}}$  on an overlarge grid – for example  $[m, M]$ , where  $m$  (resp.  $M$ ) denotes the minimum (resp. maximum) distance between two observations;
- 2). Since  $\widehat{h}_{\text{NR}}$  is known to usually oversmooth the data, search for the PI/bootstrap bandwidth on the grid  $[0.1\widehat{h}_{\text{NR}}, 2\widehat{h}_{\text{NR}}]$ ;
- 3). If the solution is not found, widen the grid until a solution is found.

Obviously, these guidelines apply for many bandwidth selection procedures of similar type.

For some methods of bandwidth selection, once the grid has been fixed, one can still face another type of problem: non uniqueness of the solution. In the case of the sinus cardinal kernel  $K_1$ , for example, it is easy to check that the optimal bandwidth (the one that minimizes the MISE) is a solution of the equation  $\text{MISE}'(h) = 0 \iff -1 + (n+1)|\varphi_Y(1/h)|^2 = 0$ . A data-driven bandwidth can be found by replacing  $\varphi_Y$  by its empirical estimator, which leads to an equation that has multiple roots. A similar

phenomenon occurs even with more general kernels with the cross-validation (CV) method of Stefanski and Carroll (1990). Interestingly, among the multiple solutions, the global minimiser of the  $\widehat{\text{MISE}}$  is not necessarily the most appropriate bandwidth, neither are we guaranteed to even find an acceptable bandwidth.

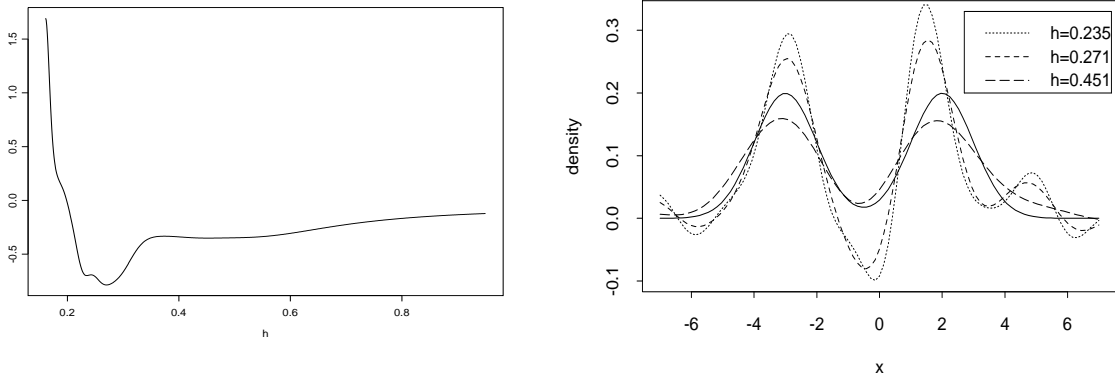


Figure 3.3: Typical shape of  $\widehat{\text{MISE}} - \int f_X^2$  for a sinc kernel (left). Estimators of the mixture density (the solid curve) from one sample using three local bandwidth solutions (right).

In Figure 3.3 for example, we show, on the interval  $[0.15, 1.0]$ , the typical shape of  $\widehat{\text{MISE}} - \int f_X^2$  for the sinc kernel (in this case when the sample comes from a mixture of two Gaussian densities contaminated by Laplace error). On this restricted interval, we already have 3 local minimisers (0.235, 0.271, 0.451). Figure 3.3 also shows the target mixture density (solid curve) along with the estimators using these three solutions. It is clear that  $h = 0.451$  is the best of all the solutions, but yet this bandwidth is not the global minimiser of  $\widehat{\text{MISE}}$ . In practice one cannot guess which of the multiple roots is the best one. Moreover, the performance of this and the CV method depends heavily on the interval chosen. It is possible to define rules such as, restrict to a small interval  $[a \cdot \hat{h}_{\text{NR}}, b \cdot \hat{h}_{\text{NR}}]$  with  $a < b$  (for example  $b = 1.2$  and  $a = 0.2$ ), but, in this case, these choices have a strong impact on the bandwidth selected. An illustration of this is provided in Figure 3.4 in which we report on results from 500 simulated samples of size  $n = 100$  from the same mixed normal density contaminated with Laplace error. For the left figure, the bandwidth chosen was the largest solution in the interval  $[0.15, 1]$ . For the right figure, the bandwidth was searched for in the interval  $[0.15, 0.6]$ . In the cases no solution was found on  $[0.15, 0.6]$ , a solution was searched outside that interval. Based on all simulated samples we ordered the obtained estimators according to their performance in terms of Integrated Squared Error (ISE) and in Figure 3.4 we present the estimated curves for which this performance corresponds to the first quantile, the median and the third quartile of the

sequence of ordered ISE-values.

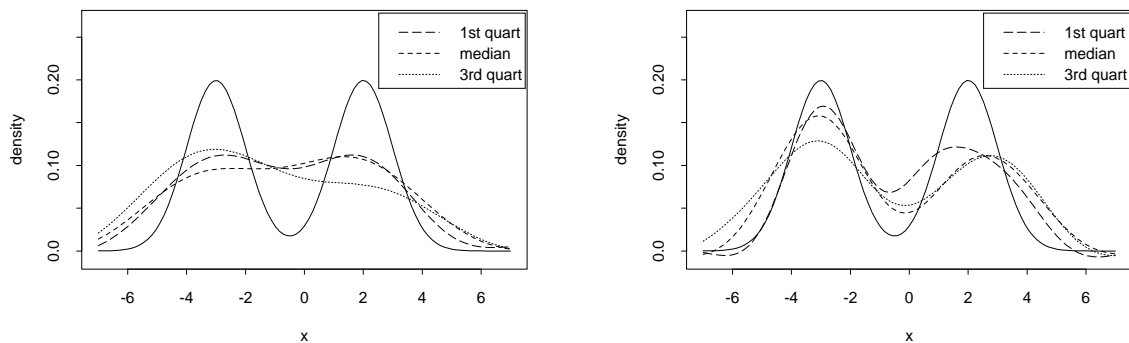


Figure 3.4: *Estimators of the mixed normal density, from samples of size  $n = 100$  contaminated with a Laplace error, using the sinc kernel, and taking grids  $[0.15, 1]$  (left) and  $[0.15, 0.6]$  (right).*

In summary, methods such as cross-validation turn out to be quite unreliable and should be used with the necessary cautions towards difficulties mentioned above. We advise to use, when possible, other methods such as plug-in or bootstrap methods, which are more stable. See also Delaigle and Gijbels (2004b).

The above reported items related to calculation of objective functions and to optimization of such functions are just illustrations of practical problems one needs to be aware of, and needs to take care of when using statistical procedures.

### Acknowledgement

The authors are grateful to the Editor, an Associate Editor and the reviewers for valuable comments which led to an improvement of the presentation. This paper was finished while the second author was visiting the Department of Mathematics of the University of California, San Diego. The author thanks the Department for its hospitality and support, and also gratefully acknowledges financial support from the Research Fund K.U.Leuven (GOA/2007/4).

### References

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

- Delaigle, A. and Gijbels, I. (2002). Estimation of integrated squared density derivatives from a contaminated sample. *Journal of the Royal Statistical Society, B*, **64**, 869 – 886.
- Delaigle, A. and Gijbels, I. (2004a). Bootstrap bandwidth selection in kernel density estimation from a contaminated sample. *Annals of the Institute of Statistical Mathematics*, **56**, 19 – 47.
- Delaigle, A. and Gijbels, I. (2004b). Practical bandwidth selection in deconvolution kernel density estimation. *Computational Statistics and Data Analysis*, **45**, 249–267.
- Delaigle, A. and Hall, P. (2006). On the optimal kernel choice for deconvolution. *Statistics & Probability Letters*, **76**, 1594–1602.
- Van Es, A.J and H.-W. Uh (2004). Asymptotic normality of kernel type deconvolution estimators: Crossing the Cauchy boundary. *Nonparametric Statistics*, **16**, 261–277.
- Van Es, A.J and H.-W. Uh (2005). Asymptotic normality of kernel type deconvolution estimators . *Scandinavian Journal of Statistics*, **32**, 467–483.
- Fan, J. (1991a). Asymptotic normality for deconvolution kernel density estimators. *Sankhya A*, **53**, 97–110.
- Fan, J. (1991b). Global behaviour of deconvolution kernel estimates. *Statistica Sinica*, **1**, 541–551.
- Fan, J. (1991c). On the optimal rates of convergence for nonparametric deconvolution problems. *The Annals of Statistics*, **19**, 1257–1272.
- Fan, J. (1992). Deconvolution with supersmooth distributions. *The Canadian Journal of Statistics*, **20**, 155–169.
- Hesse, C.H. (1999). Data-driven deconvolution. *Nonparametric Statistics*, **10**, 343–373.
- Masry, E. (1993a). Asymptotic normality for deconvolution estimators of multivariate densities of stationary processes. *Journal of Multivariate Analysis*, **44**, 47–68.
- Masry, E. (1993b). Strong consistency and rates for deconvolution of multivariate densities of stationary processes. *Stochastic Processes and their Applications*, **47**, 53–74.
- Meister, A. (2004). On the effect of misspecifying the error density in a deconvolution problem. *Canadian Journal of Statistics*, **32**, 439–449.

- Meister, A. (2006). Density estimation with normal measurement error with unknown variance. *Statistica Sinica*, **16**, 195–211.
- Press, W.H. , Flannery, B.P., Teukolsky, S.A., Vetterling, W.T. (1992). *Numerical Recipes in C, The Art of Scientific Computing (Second Edition)*. Cambridge University Press.
- Stefanski, L. and Carroll, R.J. (1990). Deconvoluting kernel density estimators. *Statistics*, **2**, 169–184.
- Zhang, S. and Karunamuni, R. (2000). Boundary bias correction for nonparametric deconvolution. *Ann. Inst. Statist. Math.*, **52**, 612–629.