ELSEVIER

Statistics & Probability Letters 37 (1998) 131–139

STATISTICS & PROBABILITY LETTERS

Finite sample performance of deconvolving density estimators

M.P. Wand

Australian Graduate School of Management, University of New South Wales, Sydney, Australia

Received 1 November 1996; received in revised form 1 July 1997

Abstract

Recent studies have shown that the asymptotic performance of nonparametric curve estimators in the presence of measurement error will often be very much inferior to that when the observations are error-free. For example, deconvolution of Gaussian measurement error worsens the usual algebraic convergence rates of kernel estimators to very slow logarithmic rates. However, the slow convergence rates mean that very large sample sizes may be required for the asymptotics to take effect, so the finite sample properties of the estimator may not be very well described by the asymptotics. In this article finite sample calculations are performed for the important cases of Gaussian and Laplacian measurement error which provide insight into the feasibility of deconvolving density estimators for practical sample sizes. Our results indicate that for lower levels of measurement error deconvolving density estimators can perform well for reasonable sample sizes.

Keywords: Errors-in-variables; Kernel estimator; Measurement error; Mean integrated squared error; Nonparametric regression

1. Introduction

The problem of nonparametric estimation of curves such as probability densities and regression functions in the presence of measurement error has recently received considerable treatment in the literature. In the density estimation setting the problem can be stated as follows. Let X and Z be continuous independent random variables having densities f_X, which is unknown, and f_Z, which is known. The random variable $Y = X + Z$ has density $f_Y = f_X * f_Z$ where $*$ is the convolution operator. Given a sample $Y_1, \ldots, Y_n$ of independent observations having density f_Y the problem is to estimate f_X. This is often referred to as the *deconvolution* density estimation problem. Throughout we will call f_X the *target density* and f_Z the *error density*.

An important question is that concerning how well f_X can be estimated nonparametrically for a particular error density f_Z and sample size n. This problem also has implications for nonparametric regression in the presence of measurement error. Valuable insight into the answer has been gained through asymptotic analysis where it has been shown that convergence rates are typically considerably worse than for error-free data. In the important case of Gaussian measurement error it has been demonstrated by Carroll and Hall (1988) that, in a minimax sense, no estimator has a better than logarithmic rate of convergence. On the other hand, it may shown that for Laplacian measurement error reasonably good algebraic rates are obtainable. Stefanski and Carroll (1990) proposed a deconvolving density estimator, which can be viewed as the natural extension of the

0167-7152/98/$19.00
PII S0167-7152(97)00110-7

ordinary kernel density estimator, that achieves the optimal rates. For other asymptotic results on deconvolving density estimators see Devroye (1989), Liu and Taylor (1989), Stefanski (1990) and Fan (1991, 1992).

While asymptotic theory gives important insight into the difficulties of curve estimation in the presence of measurement error, the fact that the convergence rates are often slow means that very large samples may be required before the asymptotics take effect. For example, if f_X has two continuous square integrable derivatives and f_Z is the $N(0, \sigma^2)$ density then, for each $\sigma > 0$, the best possible rate of convergence of the mean integrated squared error (MISE) of the deconvolving kernel estimator of Stefanski and Carroll (1990) is of order $(\log n)^{-1}$. However, when there is no measurement error ($\sigma = 0$) the best possible MISE has order $n^{-4/5}$ which is considerably faster. Obviously, for small positive values of σ the asymptotics will not accurately describe the actual situation unless the sample size is very large.

One way of understanding the finite sample properties of deconvolving estimators is through simulation studies such as those carried out by Liu and Taylor (1990), Stefanski and Carroll (1990) and Fan (1992). While these have allowed better understanding of the deconvolution problem, they suffer from the disadvantage of being computationally expensive and subject to Monte Carlo error.

In this article we use an alternative technique for evaluating the feasibility of deconvolving kernel estimators in practice. This involves exact calculations of the MISE criterion for Gaussian and Laplacian error densities and a collection of important target density shapes. It is possible to do the same for several other error densities (Wand, 1991), however because of space restrictions we will only deal with these two important cases. Exact MISE calculations in density estimation were first considered by Fryer (1976) and Deheuvels (1977) and, more recently, by Marron and Wand (1992) and Wand (1992) where they have been demonstrated to be a useful tool for evaluating the finite sample performance of nonparametric curve estimators.

Our exact calculations indicate that it is often possible to deconvolve low levels of Gaussian measurement error with practical sample sizes. However, the poor relative asymptotic performance of the kernel estimator for Gaussian error compared to Laplacian error is evident for larger levels of error as well as larger sample sizes. As described by the asymptotic theory, practical deconvolution of Laplacian error is shown to be quite feasible.

The important problem of nonparametric regression when the response variables are contaminated by error has recently been studied by Fan and Truong (1991). These authors applied the deconvolving kernel ideas to kernel regression estimators and obtained asymptotic results similar to the density estimation setting. It is envisaged that the ideas presented in this study could be adapted to allow finite sample analysis in the more complicated nonparametric regression context.

In Section 2 we derive formulae for exact computation of MISE for a collection of example target densities. Section 3 gives some examples of these calculations and an appreciation of the difficulty of deconvolution in practice is obtained through exact MISE calculations. We conclude with a discussion in Section 4.

2. Exact mise for deconvolving density estimators

An estimator for f_X based on the $Y_1, \ldots, Y_n$ can be motivated by noting that

$$f_X(x) = (2\pi)^{-1} \int e^{-itx} \{\psi_Y(t)/\psi_Z(t)\}\, dt,$$

where ψ_Y and ψ_Z are the characteristic functions of Y and Z, respectively. Replacing the $f_Y(y)$ in $\psi_Y(t) = \int e^{ity} f_Y(y)\, dy$ by its kernel estimator $(nh)^{-1} \sum_{j=1}^{n} K\{(y - Y_j)/h\}$, for some kernel function K, we obtain the deconvolving kernel density estimator of Stefanski and Carroll (1990),

$$\hat{f}_X(x; h) = (nh)^{-1} \sum_{j=1}^{n} K^Z\{(x - Y_i)/h; h\}, \tag{2.1}$$

where

$$K^Z(x;h) = (2\pi)^{-1} \int e^{itx} \{\kappa(t)/\psi_Z(t/h)\}\, dt$$

and $\kappa(t) = \int e^{itx} K(x)\, dx$ is the Fourier transform of K. The smoothing parameter h is usually called the window width or bandwidth. Rescalings of K by h are denoted by $K_h(u) = K(u/h)/h$.

For a given value of h the MISE is given by

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

From the calculations in Stefanski and Carroll (1990) it is straightforward to derive

$$\text{MISE}(h) = (nh)^{-1} \int K^Z(\cdot;h)^2 + (1 - n^{-1}) \int (K_h * f_X)^2 - 2 \int (K_h * f_X) f_X + \int f_X^2. \tag{2.2}$$

This expression differs from that for error-free observations only in the first term. In the error-free case this term is $(nh)^{-1} \int K^2$. Observe that application of Parseval's identity leads to the Fourier domain version of (2.2),

$$\begin{aligned} \text{MISE}(h) = {} & (2\pi nh)^{-1} \int \kappa(t)^2 |\psi_Z(t/h)|^{-2}\, dt \\ & + (2\pi)^{-1} \int \{(1 - n^{-1})\kappa(ht)^2 - 2\kappa(ht) + 1\} |\psi_X(t)|^2\, dt, \end{aligned} \tag{2.3}$$

where ψ_X is the characteristic function of X. Respectively, the terms on the right hand side of (2.3) essentially represent the integrated variance, which depends only on the error density, and the integrated squared bias, which depends only on the target density.

If we set Z to be a $N(0, \sigma^2)$ random variable then the effect of having to deconvolve Gaussian error on the MISE depends on the size of the first integral in (2.3) which becomes $\int \kappa(t)^2 e^{\sigma^2 t^2/h^2}\, dt$. For many common kernels this integral does not converge, including the normal kernel if $h < \sigma$. For Gaussian error a finite integral can be obtained by taking κ to have compact support. Examples are

$$\kappa_1(t) = (1 - t^2)^3, \qquad |t| < 1,$$

$$\kappa_2(t) = \begin{cases} 1 - 6t^2 + 6|t|^3, & |t| \leqslant \frac{1}{2}, \\ 2(1 - |t|)^3, & \frac{1}{2} \leqslant |t| < 1. \end{cases}$$

The corresponding kernels are

$$K_1(x) = \frac{48x(x^2 - 15)\cos x - 144(2x^2 - 5)\sin x}{\pi x^7},$$

$$K_2(x) = \frac{3}{8\pi} \left(\frac{\sin(x/4)}{x/4} \right)^4,$$

respectively. Note that K_1 is used in simulations by Fan (1992). Apart from the Gaussian error density we will also consider Z having a Laplace density with scale parameter σ, given by $f_Z(x) = (2\sigma)^{-1} e^{-|x|/\sigma}$. In our

calculations we consider densities which are either of the form

$$f_X(x) = w(2\pi\sigma_1^2)^{-1/2}e^{-x^2/(2\sigma_1^2)} + (1-w)(2\pi\sigma_2^2)^{-1/2}e^{-x^2/(2\sigma_2^2)} \tag{2.5}$$

or

$$f_X(x) = wx^{p_1-1}e^{-x}/(p_1-1)! + (1-w)x^{p_2-1}e^{-x}/(p_2-1)!, \qquad x > 0, \tag{2.6}$$

where $0 \leqslant w \leqslant 1$, σ_1, $\sigma_2 > 0$ and $p_2 \geqslant p_1$ are positive integers. These are, respectively, two-component normal scale and gamma mixture densities and provide a wide variety of shapes.

It is useful to note that for the kernels K_1 and K_2 and the target densities (2.5) and (2.6) with $p_1 > 3$, the best possible rate of convergence of the MISE of (2.1) is $n^{-4/5}$ for error-free data, $(\log n)^{-1}$ for Gaussian error and $n^{-4/9}$ for Laplacian error.

3. Examples and equivalent sample size calculations

Using the results from the previous section, exact MISE calculations for Gaussian and Laplace error have been performed for several target densities. Integrals were computed using trapezoidal quadrature with successive doubling of the integration mesh until convergence was obtained. Because of space restrictions we will present our results for four target densities which represent some important density shapes which one would typically hope to resolve well in practice. These target densities are (i) the standard normal density, (ii) the two-component normal mixture with $w = \frac{2}{3}$, $\sigma_1 = 1$ and $\sigma_2 = \frac{1}{5}$, (iii) the gamma density with shape parameter 4 and (iv) the two-component gamma mixture density with $w = \frac{2}{5}$, $p_1 = 5$ and $p_2 = 13$. Density (ii) is symmetric but with a kurtosis coefficient about 2.23 times that of the normal, density (iii) is skewed and density (iv) is bimodal. Fig. 1 shows the graphs of each of these densities.

To compare the two different types of error we will use the percentage $100p\%$, $0 \leqslant p \leqslant 1$, of the variance of the observed data which is due to measurement error. Then $p = \mathrm{var}(Z)/\mathrm{var}(Y)$ where Z and Y are generic random variables having densities f_Z and f_Y, respectively. While p seems to be satisfactory for comparing Gaussian and Laplacian error for a fixed target density it should be noted that p is not very useful for comparing across target densities since the meaning of variance can differ considerably for different density shapes. It is easy to show from (2.3) that, for fixed values of p, the MISE for Gaussian error dominates that for Laplacian error uniformly in n and h which indicates that deconvolution of Gaussian error will always be the harder of the two.

Fig. 2 shows graphs of the MISE for Gaussian error (long dashed curves) and Laplacian error (short dashed curves) with p taking values of 10% and 30% and kernel K_1. The target density is (iii) the Gamma (4) and the sample size is $n = 100$. The MISE for error-free data is also graphed for comparison (solid curve). The most striking feature of the MISE curves corresponding to measurement error is the dramatic increase in error incurred by using too small a window width. This is a manifestation of the high sensitivity that the integrated variance has to undersmoothing and is particularly acute for Gaussian error. However, at the minima, the difference in MISE is less dramatic, especially for $p = 10\%$. For $p = 30\%$ it is seen that the minimum MISE can be several factors higher than for error-free data.

To better understand how the effect of measurement error changes for increasing sample size Fig. 3 plots values of $\inf_{h>0}\mathrm{MISE}(h; p)$ versus n on a $\log_{10}$–$\log_{10}$ scale for n ranging from 10 to 100 000 and the same density and values of p as in Fig. 2 and for kernel K_1. Here the argument p in $\mathrm{MISE}(h; p)$ indicates the level of error. The asymptotic results for these cases dictate that the curve for error-free data is asymptotic to a line having slope $-\frac{4}{5}$ while those for Laplacian error approach a line having slope $-\frac{4}{9}$. The fact that the curves for Gaussian error do not decrease linearly is a consequence of the logarithmic convergence rate in this case. For $p = 10\%$ the closeness of the lower pair of broken curves for n up to about 200 indicates

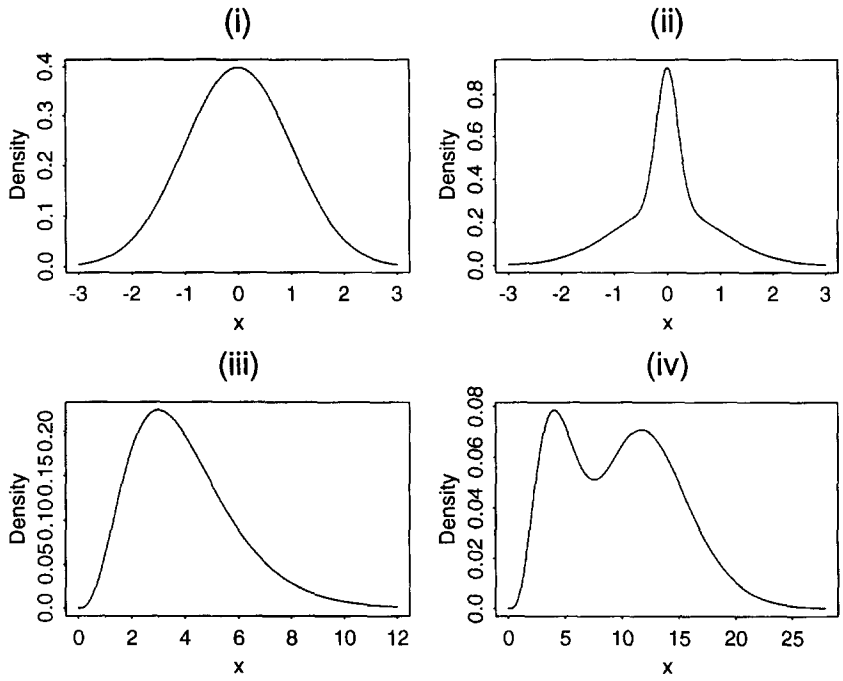

Fig. 1. The four example target densities: (i) standard normal, (ii) normal mixture, (iii) gamma (4) and (iv) gamma mixture.

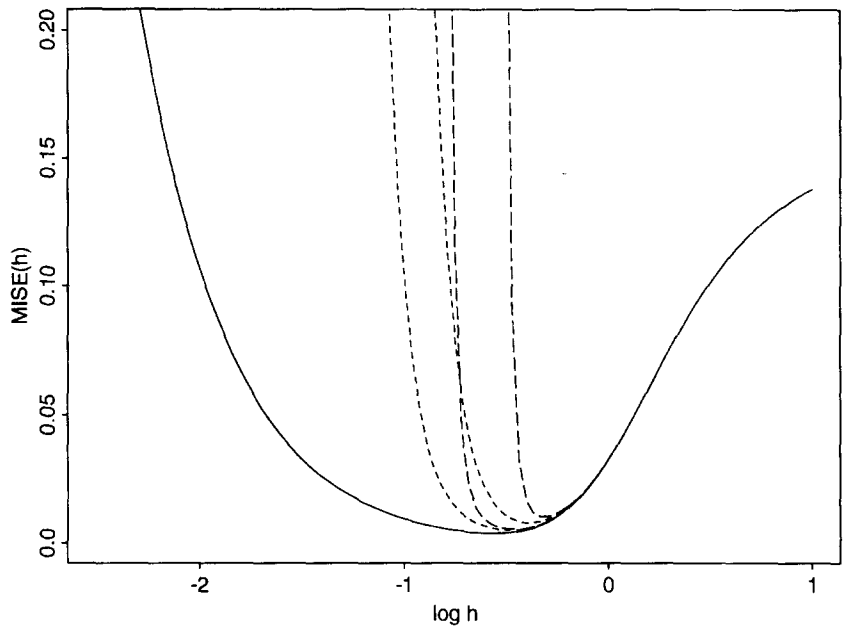

Fig. 2. Plots of MISE(h; p) versus $\log_{10} h$ for the gamma (4) target density. The solid curve is for the error-free data, the long dashed curves are for Gaussian error with $p = 10\%$ (lower curve) and $p = 30\%$ (upper curve) and the short dashed curves are for Laplacian error with $p = 10\%$ (lower curve) and $p = 30\%$ (upper curve).

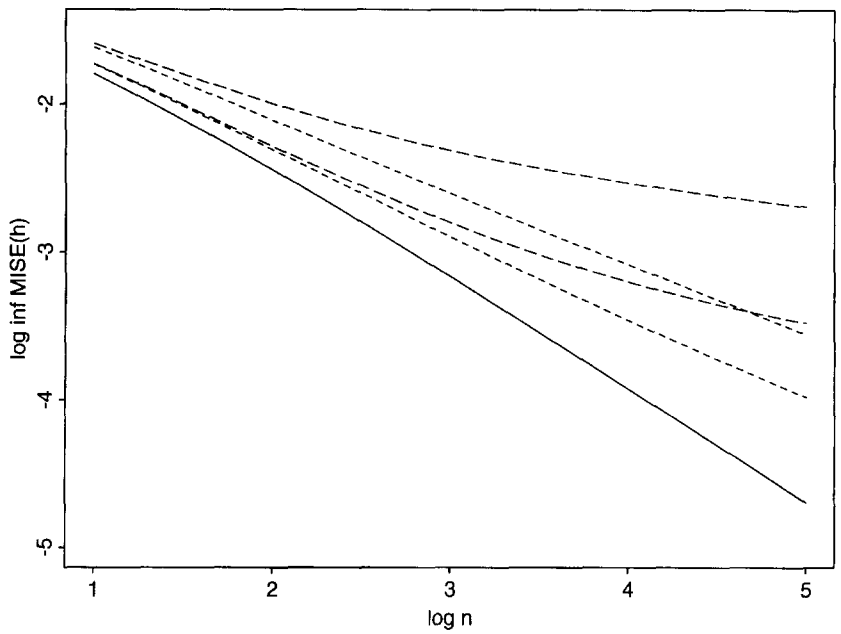

Fig. 3. Plots of $\log_{10} \inf_{h>0} \mathrm{MISE}(h; p)$ versus $\log_{10} n$ for the gamma (4) target density. The solid curve is for the error-free data, the long dashed curves are for Gaussian error with $p = 10\%$ (lower curve) and $p = 30\%$ (upper curve) and the short dashed curves are for Laplacian error with $p = 10\%$ (lower curve) and $p = 30\%$ (upper curve).

that, for this target density, there is little difference between deconvolution of Gaussian and Laplacian error. However, the asymptotics begin to take effect for larger values of n and the minimum MISE for Gaussian error becomes considerably worse. For the higher error level, $p = 30\%$, the detrimental effect of Gaussian measurement error is evident for sample sizes as low as about 50.

A more precise appreciation of how much error one can expect to deconvolve in practice can be realised through equivalent sample size calculations. For varying values of p, Tables 1 and 2 present, for each kernel K_1 and K_2, sample sizes required for the deconvolving density estimator to achieve the same value of $\inf_{h>0} \mathrm{MISE}(h; p)$ as that obtained for samples of size $n = 100$ and $n = 1000$ of error-free data. If the sample size exceeded 10^6 then we did not report this very high value. For Laplacian error it is seen that the kernels perform about the same. However, K_1 does a much better job at deconvolving Gaussian error, particularly for larger p. For $n = 100$ and values of p less than about 25% we see that the equivalent sample sizes can be reasonable for both types of error but for the higher levels they can become prohibitively large for Gaussian error. As mentioned above, p is not useful for comparison across target densities, as indicated by the much higher equivalent sample sizes obtained for density (ii). A different measure of error level would be appropriate for this case. For $n = 1\,000$ the leading terms tend to be more dominant and the equivalent sample sizes are relatively larger for small p. Nevertheless, it is clear that there are many situations where one can estimate f fairly satisfactorily if the data is contaminated by measurement error.

4. Discussion

It comes as no surprise that Gaussian deconvolution is difficult when the error level is high. However, our results in Section 3 show that deconvolution can be viable for detecting distributional shape in the variable

Table 1
Minimum sample sizes required to achieve $\inf_{h>0} \mathrm{MISE}(h; p) \leqslant \inf_{h>0} \mathrm{MISE}(h; 0)$ for Gaussian and Laplacian error and kernel K_1. The quantity $\inf_{h>0} \mathrm{MISE}(h; 0)$ is the minimum MISE for error-free samples of $n = 100$ and $n = 1000$

	Sample size $n = 100$		$n = 1000$	
p	Gaussian error density	Laplacian error density	Gaussian error density	Laplacian error density
(i) Standard normal density				
10%	156	146	2931	2064
20%	293	215	24795	3986
30%	788	325	$> 10^6$	7447
40%	4342	506	$> 10^6$	13814
50%	103093	828	$> 10^6$	26109
(ii) Normal mixture density				
10%	1318	362	$> 10^6$	12561
20%	625595	924	$> 10^6$	46873
30%	$> 10^6$	2045	$> 10^6$	123733
40%	$> 10^6$	4237	$> 10^6$	283467
50%	$> 10^6$	8650	$> 10^6$	617080
(iii) Gamma (4) density				
10%	197	171	7841	2948
20%	565	284	$> 10^6$	6975
30%	3507	474	$> 10^6$	14822
40%	99217	804	$> 10^6$	29989
50%	$> 10^6$	1415	$> 10^6$	60276
(iv) Gamma mixture density				
10%	273	204	149182	5186
20%	1373	382	$> 10^6$	15477
30%	24450	691	$> 10^6$	37152
40%	$> 10^6$	1249	$> 10^6$	80876
50%	$> 10^6$	2311	$> 10^6$	170577

of interest for lower levels of error. Because the number of examples which one can study in practice is always limited, there is always room for further study. Nevertheless, we have demonstrated that exact MISE calculations are a valuable tool for understanding problems as important as how much measurement error one may expect to deconvolve in practice.

As pointed out by a referee, another possibility for understanding the effects of measurement error is to perform asymptotics as $p \to 0$ for fixed n. Fan (1992) performs a variation of this idea by letting the error variance approach zero.

5. Acknowledgements

I am grateful to Professor Jianqing Fan and a referee for their very helpful comments on this project. This research was supported by a grant from the Office of Naval Research.

Table 2
Minimum sample sizes required to achieve $\inf_{h>0} \mathrm{MISE}(h; p) \leqslant \inf_{h>0} \mathrm{MISE}(h; 0)$ for Gaussian and Laplacian error and kernel K_2. The quantity $\inf_{h>0} \mathrm{MISE}(h; 0)$ is the minimum MISE for error-free samples of $n = 100$ and $n = 1000$

	Sample size $n = 100$		$n = 1000$	
p	Gaussian error density	Laplacian error density	Gaussian error density	Laplacian error density
(i) Standard normal density				
10%	166	152	4260	2286
20%	376	232	271 217	4735
30%	1730	361	$> 10^6$	9294
40%	43 306	579	$> 10^6$	17 869
50%	$> 10^6$	974	$> 10^6$	34 681
(ii) Normal mixture density				
10%	4819	401	$> 10^6$	15 382
20%	$> 10^6$	1082	$> 10^6$	59 720
30%	$> 10^6$	2466	$> 10^6$	160 027
40%	$> 10^6$	5210	$> 10^6$	369 755
50%	$> 10^6$	10 777	$> 10^6$	809 267
(iii) Gamma (4) density				
10%	222	180	22 485	3354
20%	1010	313	$> 10^6$	8499
30%	29 323	540	$> 10^6$	18 805
40%	$> 10^6$	945	$> 10^6$	39 041
50%	$> 10^6$	1705	$> 10^6$	79 865
(iv) Gamma mixture density				
10%	345	219	$> 10^6$	5993
20%	5387	431	$> 10^6$	18 857
30%	$> 10^6$	814	$> 10^6$	46 425
40%	$> 10^6$	1520	$> 10^6$	102 537
50%	$> 10^6$	2886	$> 10^6$	218 278

References

Carroll, R.J., Hall, P., 1988. Optimal rates of convergence for deconvolving a density. J. Amer. Statist. Assoc. 83, 1184–1186.
Deheuvels, P., 1977. Estimation nonparametrique de la densité par histogrammes generalisés. Rev. Statist. Appl. 25, 5–42.
Devroye, L., 1989. Consistent deconvolution in density estimation. Can. J. Statist. 17, 235–239.
Fan, J., 1991. On the optimal rates of convergence for nonparametric deconvolution problem. Ann. of Statist. 19, 1257–1272.
Fan, J., 1992. Deconvolution with supersmooth distributions. Can. J. Statist. 20, 155–169.
Fan, J., Truong, Y., 1991. Nonparametric regression with errors-in-variables. Ann. Statist. 21, 1900–1925.
Fryer, M.J., 1976. Some errors associated with the nonparametric estimation of density functions. J. Instit. Math. Appl. 18, 371–380.
Gradshteyn, I.S., Ryzhik, I.M., 1980. Tables of Integrals, Series and Products, Academic Press, San Diego.
Marron, J.S., Wand, M.P., 1992. Exact mean integrated squared error. Ann. Statist. 20, 712–736.
Liu, M.C., Taylor, R.L., 1989. A consistent nonparametric density estimator for the deconvolution problem. Can. J. Statist. 17, 427–438.
Liu, M.C., Taylor, R.L., 1990. Simulations and computations of nonparametric estimates for the deconvolution problem. J. Statist. Comput. Simul. 35, 145–167.
Stefanski, L.A., 1990. Rates of convergence of some estimators in a class of deconvolution problems. Statist. Probab. Lett. 9, 229–235.

Stefanski, L., Carroll, R.J., 1990. Deconvoluting kernel density estimators. Statistics 2, 169-1-84.

Wand, M.P., 1991. Exact mean integrated squared error for some non-Gaussian deconvolution problems. Department of Statistics, Technical Report No. 91–8, Rice University, Houston.

Wand, M.P., 1992. Finite sample performance of density estimators under moving average dependence. Statist. Probab. Lett. 13, 109 – 115.