A Simple Deconvolving Kernel Density Estimator when Noise is Gaussian

Isabel Proença*
Instituto Superior de Economia e Gestão
Universidade Técnica de Lisboa

October 13, 2003

Abstract

Deconvolving kernel estimators when noise is Gaussian entail heavy calculations. In order to obtain the density estimates numerical evaluation of a specific integral is needed. This work proposes an approximation to the deconvolving kernel which simplifies considerably calculations by avoiding the typical numerical integration. Simulations included indicate that the lost in performance relatively to the true deconvolving kernel, is almost negligible in finite samples.

Keywords: deconvolution, density estimation, errors-in-variables, kernel, simulations.

1 Introduction

The estimation of a density by deconvolution consists in the estimation of the density of a random variable that is observed with an added unknown random noise. A typical example is the estimation of a density of a variable observed with measurement error. Another example is the estimation of the mixing distribution in a duration model. In what concerns applications, Fan and Truong (1993) introduce deconvolution techniques in the context of nonparametric regression with errors in variables. Calvet and Comon (2000) perform the deconvolution estimation of the joint density of spendig and tastes in presence of measurement error, and Horowitz and Markatou (1996) analyze earnings mobility using nonparametric deconvolution estimation of a density in the context of a random effects model for panel data.

To describe the problem, suppose a random variable Y such that $Y = X+U$, where X and U are independent random variables. Suppose more that Y is observable while X (the target) and U (the noise) are non-observable, and the aim here is to estimate the density of X, $f(x)$, also called the *target*

*Address for correspondence: R. do Quelhas 2, 1200-781 Lisboa, Portugal.E-mail: isabelp@iseg.utl.pt. Fax: 351 213922781.

density when $g(y)$, the density of Y is completely unknown. Usually the *noise* density, $q(u)$, is assumed to belong to a given family, most frequently the Normal, with zero mean. Observe that $g(y)$ is equal to the convolution of the densities of X and U, verifying,

$$g(y) = \int_{-\infty}^{\infty} f(y-u)q(u)du. \tag{1}$$

Then the density $f(x)$ may be obtained by deconvolution. The usual procedure is to obtain first the characteristic function of X, $\dot{f}$, using $\dot{g}$ and $\dot{q}$ (the characteristic functions of Y and U respectively) and then inverse Fourier transform leads to $f(x)$ according to,

$$f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itx}\frac{\dot{g}(t)}{\dot{q}(t)}dt \tag{2}$$

An estimator of $f(x)$ can be obtained by substituting the unknown quantities in (2) by consistent estimators. However, in practice the corresponding calculations can entail problems leading to high fluctuations in the aimed estimate. To avoid this a suitable damping factor is incorporated in the corresponding integral leading to the following deconvolving kernel density estimator introduced by Stefanski and Carroll (1990),

$$\hat{f}(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itx}\dot{k}(th)\frac{\widehat{\dot{g}}(t)}{\dot{q}(t)}dt \tag{3}$$

with $\dot{k}(t)$ the Fourier transform of a kernel function $K(x)$ (such that $\dot{k}(0) = 1$), h the bandwidth which tends to 0 (so that the damping factor tends to 1), and $\widehat{\dot{g}}(t)$ the empirical characteristic function of Y. Fan (1991) obtains convergence rates of this estimator for several *noise* distributions.

Stefanski and Carrol (1990) show that in case the function $\dot{k}(t)/\dot{q}(t/h)$ is integrable then expression (3) can be rewritten as,

$$\hat{f}(x) = \frac{1}{nh}\sum_{1}^{n} K_h^*\left(\frac{x-Y_j}{h}\right) \tag{4}$$

where

$$K_h^*(z) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itz}\frac{\dot{k}(t)}{\dot{q}(t/h)}dt \tag{5}$$

Observe that by equation (4) the deconvolving kernel estimate is just an ordinary kernel estimate but with specific kernel function equal to (5) where the shape of this kernel function depends on the bandwidth. For certain types of distributions for the noise variable, (5) has a closed-form expression and calculations are as hard as the ordinary kernel estimation. Unfortunately this is not the case when the noise is normally distributed as it is often assumed namely in econometric applications. When $q(u)$ belongs to the normal family the integral (5) has to be evaluated and calculations are much harder. Plus, the damping kernel K has to be carefully chosen, to guarantee that the integral exists. Consequently, deconvolving kernel density estimation for Guassian noise suffers from the drawback of being subject to Monte Carlo error and computationally very burdensome.

In this paper, an approximation of (5) is proposed to estimate $f(x)$ by kernel deconvolution when noise is Gaussian. It avoids the typical numerical integration making calculations incredibly easier, being as difficult as an ordinary kernel density estimator. The next section introduces the simple deconvolving Kernel. Section 3 presents a simulation study that analyzes the performance of the new estimator compared to the exact one for several sample sizes and target distributions. Section 4 concludes.

2 The simple Deconvolving Kernel density estimator

The main idea behind the simple deconvolving kernel estimator is to substitute in (5) the inverse of the true characteristic function of the normal density (which is an exponential function) by an approximation given by the first-order term of the respective Taylor series expansion around $\sigma_u^2 = 0$. Therefore, considering as damping kernel the standard normal the approximate deconvolving kernel is equal to,

$$K_h^{a*}(z) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itz}e^{-0.5t^2}\left(1 + \frac{t^2\sigma_u^2}{2h^2}\right)dt \tag{6}$$

with σ_u^2 the variance of the noise variable.

Elementary calculations simplify (6) to the following expression,

$$K_h^{a*}(z) = \phi(z) - \frac{\sigma_u^2}{2h^2}\phi''(z) \tag{7}$$

where $\phi(z)$ is the standard normal density function and $\phi''(z)$ is its second derivative. Finally, the density estimate is obtained with,

$$\hat{f}^a(x) = \frac{1}{nh}\sum_{1}^{n} K_h^{a*}\left(\frac{x - Y_j}{h}\right) \tag{8}$$

where $K_h^{a*}(\bullet)$ is given in (7).

Using the same arguments as Stefanski and Carroll (1990) is easy to show that $K_h^{a*}(\bullet)$ is symmetric and $\int \hat{f}^a(x)dx = 1$.

3 Simulation Results

In this section the performance of the approximated deconvolving kernel in finite samples is examined by a detailed simulation study. The main goal is to evaluate the deterioration in the accuracy of the deconvolving kernel density estimates due to the use of the much simpler approximated deconvolving kernel introduced in this paper instead of the exact one of Stefanski and Carroll (1990). With this aim the average integrated squared error (AISE) calculated for the optimal bandwidth (optimal in the sense that minimizes the AISE) is compared for the two procedures. The optimal bandwidth was found by grid search (with increment equal to 0.02). Depending on the particular design, the grid intervals were chosen wide enough in order to assure that they contain within the optimal bandwidth. It is also an aim to analyze whether the performance depends on the shape of the target density or on the importance of the disturbing noise measured by the reliability ratio equal to,

$$r = \frac{Var(X)}{Var(Y)} = \frac{\sigma_X^2}{\sigma_X^2 + \sigma_U^2}. \tag{9}$$

For the last issue, two situations were considered. One, with a relatively small noise with corresponding reliability ratio equal to 0.89 and another where r is 0.62 which expresses a severe perturbation of X (these particular quantities where chosen in order that σ_U^2 and σ_X^2 have suitable values). Observe that in this last situation the exact deconvolving kernel density estimate looses accuracy (in the sense that the corresponding AISE has tendency to be considerably bigger).

For $r = 0.89$ three different designs were chosen in order to illustrate three of the most important types of shapes of the densities of the target variable that are more frequently found in practice resulting respectively in a symmetric, a skewed, and a bimodal densities. The designs are,

DESIGN 1 - $X \sim N(0, 16)$, $U \sim N(0, 2)$ and $Y = X + U$.

DESIGN 2 - $X \sim \chi^2(8)$, $U \sim N(0, 2)$ and $Y = X + U$.

DESIGN 3 - X is a mixture of a $N(\sqrt{44}, 20)$ with a $N(-\sqrt{44}, 20)$, being $U \sim N(0, 8)$ and $Y = X + U$.

For each design 1000 replications were calculated for samples with size respectively of 100, 250 and 500 observations. The results can be seen in table 1. It is clear that the simpler approximated deconvolving kernel has a good performance given that the deterioration in the AISE is almost negligible even for small samples and for all designs tried. The worst case refers to an increase in the AISE of 7% for sample size of 100 obtained with Design 1. There are not remarkable differences among all the different density shapes tried. The fact that the rates are slightly less favorable for Design 1 may be due to the general better performance (in AISE) of the exact Kernel for this type of *target* densities (relatively to the skewed and bimodal) as is analyzed in Wand 1998.

The good performance of the approximated deconvolving kernel can be seen also in figures 1 to 3. These figures represent for each design the true density together with the exact and approximated deconvolving kernel density estimates (calculated each for the respective optimal bandwidth) for one sample randomly selected with 500 observations. Both estimates are remarkable close in all the graphics.

Av. Best AISE $\times 10^6$, $r = 0.89$			
	Exact	Approx.	App/Exa
Normal density			
n = 100	68.25	73.07	1.0706
n = 250	39.64	42.19	1.0643
n = 500	26.71	28.14	1.0535
Chi-square density			
n = 100	99.18	104.44	1.0530
n = 250	62.01	64.99	1.0481
n = 500	44.40	45.89	1.0336
Bimodal density			
n = 100	32.63	33.92	1.0395
n = 250	20.86	21.93	1.0513
n = 500	15.88	16.65	1.0485

Table 1: Average Best AISE in 1000 simulated samples of 100, 250 and 500 observations.

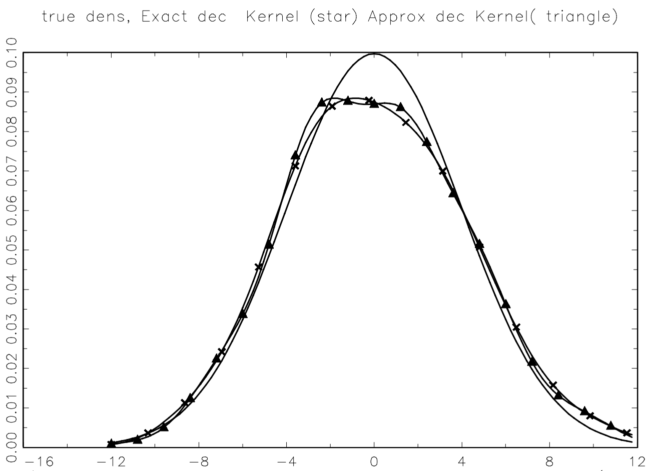

Figure 1: True density, Exact deconvolving kernel (star) Approximated deconvolving kernel (triangle). One random sample for Design 1, n=500.

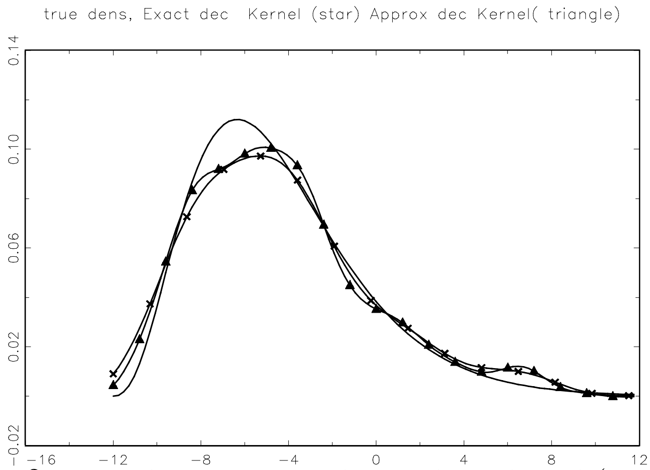

Figure 2: True density, Exact deconvolving kernel (star) Approximated deconvolving kernel (triangle). One random sample for Design 2, n=500.

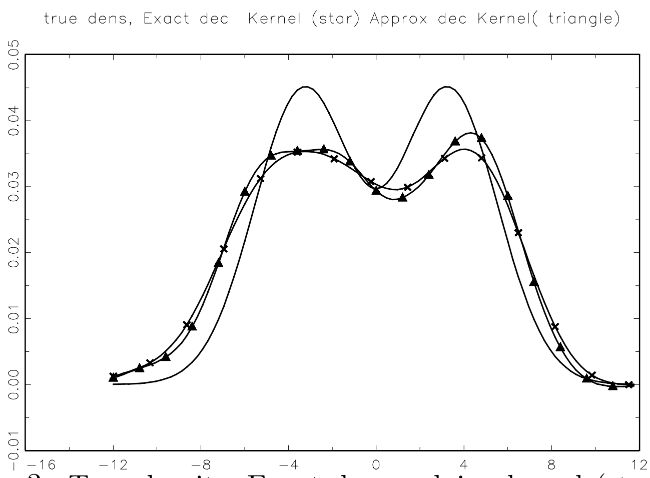

Figure 3: True density, Exact deconvolving kernel (star) Approximated deconvolving kernel (triangle). One random sample for Design 3,n=500.

Av. Best AISE $\times 10^6$ for Normal density			
	Exact	Approx.	App/Exa
	$r = 0.62$		
n = 100	150.16	157.75	1.0505
n = 250	108.51	110.79	1.0210
n = 500	85.70	85.06	0.9925

Table 2: Average Best AISE for normal *target* density in 1000 simulated samples of 100, 250 and 500 observations.

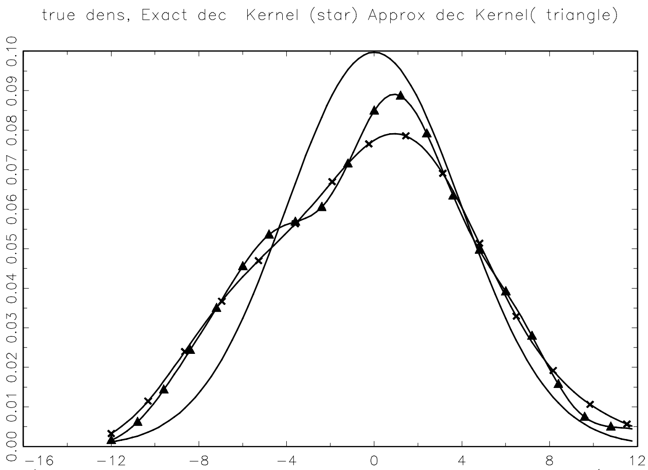

Figure 4: True density, Exact deconvolving kernel (plus) Approximated deconvolving kernel (triangle). One random sample for Design 4,n=500.

Given the heavy computations required by the exact density kernel estimator, and the fact that when $r = 0.89$ the performance of the approximated procedure for the different designs was very similar, only the symmetric shape for target variable was analyzed when $r = 0.62$, leading to

DESIGN 4 - $X \sim N(0, 16)$, $U \sim N(0, 10)$ and $Y = X + U$.

The results are included in table 2. The performance of the simple deconvolving kernel is better in approximating the exact deconvolving kernel density estimate when the variance of the error is bigger. So that, it seems that the deterioration in quality of the exact deconvolving density estimate caused by the increase of the importance of the disturbing noise is less significant in the simple approximated estimator.

Table 3 shows the ratio in the average computation time for each h between the exact deconvolving kernel density estimate and the simple approximated one. The gain in computation time of the simple deconvolving kernel

Ratio of calculation times Exact over Approximated	
Design 1	150.73
Design 2	295.68
Design 3	128.55
Design 4	27.78

Table 3: Ratio of the average calculation time for each h in one sample randomly selected.

over the exact can be impressively large. For instance, with data from Design 2 the exact calculations take more 296 times the time spent with the approximated, while for Design 4 the ratio is 28 which is already considerable, specially if one needs to replicate calculations for several bandwidths.

4 Concluding Remarks

This work introduces a simple deconvolving kernel to estimate a density by deconvolution when the noise variable is Normally distributed. It avoids the typical numerical integration necessary to obtain the ordinary deconvolving kernel density estimate in this situation, making calculations noticeably faster and less subject to numerical error. It has a simple and direct application being as difficult as an ordinary kernel density estimator.

A simulation study shows that the lost in performance of this simpler estimator in finite samples is reasonable low. Moreover, in situations where the accuracy of the exact ordinary deconvolving kernel tends to deteriorate (because of a more complex shape of the true density or a low reliability ratio) the simpler deconvolving kernel has a relatively better performance. Therefore, the use of this procedure seems to be beneficial when calculations of the deconvolving kernel density estimate have to be replicated several times, or numerical integration is a problem. On the other hand, it could be even relatively more beneficial in situations where the exact kernel is less accurate because of the shape of the density or an important noise with a low reliability ratio.

5 Acknowledgments

Thanks are due to Hidehiko Ichimura and João Santos Silva for valuable comments. The usual disclaimer applies. This work was partially done while the author was visiting the Economics Department of University College, London, due to the support of *Centre for Microdata Methods and Practice (CEMMAP)* which is gratefully appreciated. Financial support from Fundação para a Ciência e Tecnologia/MCT under FCT/POCTI and BFAB-212/2000 is also acknowledge.

6 References

Calvet, L. E. and Comon, E. (2000). "Behavioral Heterogeneity and the Income Effect". *Harvard Institute of Economic Research Working Papers*, No 1892.

Diggle, P.J. and Hall, P. (1993). "A Fourier Approach t Nonparametric Deconvolution of a Density Estimate". *J.R. Statist. Soc.*, B, **55**, 2, 523-531.

Fan, J. (1991). "On the Optimal Rates of Convergence for Nonparametric Deconvolution Problems". *The Annals of Statistics*, **19**, 3, 1257-1272.

Fan,J. and Truong, Y. K. (1993). "Nonparametric Regression with Errors in Variables". *The Annals of Statistics*, **21**, No 4, 1900-1925.

Horowitz, J.L. and Markatou, M. (1996). "Semiparametric Estimation of Regression Models for Panel Data". *Review of Economic Studies*, **63**, 145-168.

Stefanski, L. and Carrol,R.J. (1990). "Deconvoluting Kernel Density Estimators". *Statistics*, **21**, 2, 169-184.

Wand, M.P. (1998). "Finite Sample Performance of Deconvolving Density Estimators". *Statistics and Probability Letters*, **37**, 131-139.