Communications of the Korean Statistical Society
2010, Vol. 17, No. 3, 451–457

A Support Vector Method for the Deconvolution Problem

Sungho Lee[1,a]

[a]Department of Statistics, Daegu University

Abstract

This paper considers the problem of nonparametric deconvolution density estimation when sample observations are contaminated by double exponentially distributed errors. Three different deconvolution density estimators are introduced: a weighted kernel density estimator, a kernel density estimator based on the support vector regression method in a RKHS, and a classical kernel density estimator. The performance of these deconvolution density estimators is compared by means of a simulation study.

Keywords: Kernel density estimator, deconvolution, reproducing kernel Hilbert space(RKHS), support vector method.

1. Introduction

The problem of measurements being contaminated with noise exists in many different fields (*e.g.* Stefanski and Carroll, 1990; Louis, 1991; Zhang, 1992). This deconvolution problem of interest can be stated as follows. Let X and Z be independent random variables with density functions $f(x)$ and $q(z)$, respectively, where $f(x)$ is unknown and $q(z)$ is known. One observes a sample of random variables $Y_i = X_i + Z_i$, $i = 1, 2, \ldots, n$. The objective is to estimate the density function $f(x)$ where $g(y)$ is the convolution of $f(x)$ and $q(z)$, $g(y) = (f * q)(y) = \int_{-\infty}^{\infty} f(y-z)q(z)dz$. The most popular approach to this deconvolution problem has been to estimate $f(x)$ by a kernel estimator and Fourier transform (*e.g.* Carroll and Hall, 1988; Liu and Taylor, 1989; Fan, 1991). While kernel density estimation is widely considered as the most popular approach to density deconvolution, other alternatives have been proposed (*e.g.* Mendelsohn and Rice, 1982; Pensky and Vidakovic, 1999; Hall and Qiu, 2005; Lee and Taylor, 2008). Following the work of Fan (1991), two types of error distributions can be considered: ordinary smooth and super smooth distributions. Gamma or double exponential distribution functions are ordinary smooth, that is, the Fourier transform $\tilde{q}(\xi)$ $(= \int_{-\infty}^{\infty} e^{-i\xi z}q(z)dz)$ of $q(z)$ has a polynomial descent. Normal or Cauchy distribution functions are super smooth, that is, the Fourier transform $\tilde{q}(\xi)$ of $q(z)$ has an exponential descent.

In this paper three different deconvolution density estimators are introduced when the error distribution is double exponential: a weighted kernel density estimator proposed by Hazelton and Turlach (2009), a kernel density estimator based on the support vector regression method in a reproducing kernel Hilbert space(RKHS), and a classical(Parzen's) kernel density estimator. Finally, it will be shown through a simulation study that the kernel density estimator based on the support vector regression method in a RKHS is not as strong as the classical kernel density estimator.

This research was supported by the Daegu University Research Grant, 2009.
[1] Professor, Department of Statistics, Daegu University, Kyungbuk 712-714, Korea. E-mail: shlee1@daegu.ac.kr

2. Deconvolution when Measurement Errors are Double Exponential

This section will introduce three different deconvolution density estimators. First, we will evaluate a weighted kernel density estimator when the sample observations are contaminated by double exponentially distributed errors. When Hazelton and Turlach proposed this estimator, they showed that it is non-negative and that if the optimal weighting scheme $\omega_i\,(= f(Y_i)/g(Y_i))$ was known, then the estimator would have MISE(mean integrated squared error) of asymptotic order $n^{-4/5}$. They applied the estimator in cases with the Gaussian kernel and normal measurement error, and showed that the estimator can be evaluated without recourse to numerical integration techniques. Now, we can evaluate the weighted kernel estimator based on the Gaussian kernel in case of a double exponentially distributed error.

Let

$$\hat{f}_\omega(x) = \frac{1}{n}\sum_{i=1}^{n}\omega_i K_h(x - Y_i), \quad \omega_i \ge 0, \; \sum_{i=1}^{n}\omega_i = n$$

and

$$Q(\omega) = \int_{-\infty}^{\infty}\left(\hat{f}_\omega * q(y) - \hat{g}(y)\right)^2 dy, \quad q(z) = \frac{1}{2\sigma_z}e^{-|z|/\sigma_z}, \; \hat{g}(y) = \frac{1}{n}\sum_{i=1}^{n}K_h(y - Y_i).$$

Then

$$\begin{aligned}
Q(\omega) &= \int_{-\infty}^{\infty}\left(\hat{f}_\omega * q(y) - \hat{g}(y)\right)^2 dy \\
&= \int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty}\frac{1}{n}\sum_{i=1}^{n}\frac{\omega_i}{2\sqrt{2\pi}\sigma_h\sigma_z}e^{-(x-Y_i)^2/2\sigma_h^2-|y-x|/\sigma_z}dx - \frac{1}{n}\sum_{i=1}^{n}\frac{1}{\sqrt{2\pi}\sigma_h}e^{-(y-Y_i)^2/2\sigma_h^2}\right)^2 dy \\
&= \int_{-\infty}^{\infty}\left[\frac{e^{\sigma_h^2/2\sigma_z^2}}{2\sigma_z n}\sum_{i=1}^{n}\omega_i\left\{e^{-(y-Y_i)/\sigma_z}\Phi\left(\frac{y-Y_i-\sigma_h^2/\sigma_z}{\sigma_h}\right) + e^{(y-Y_i)/\sigma_z}\left(1-\Phi\left(\frac{y-Y_i+\sigma_h^2/\sigma_z}{\sigma_h}\right)\right)\right\}\right. \\
&\quad \left. -\frac{1}{n}\sum_{i=1}^{n}\frac{1}{\sqrt{2\pi}\sigma_h}e^{-(y-Y_i)^2/2\sigma_h^2}\right]^2 dy \\
&= \sum_{i=1}^{n}\omega_i^2\frac{e^{\sigma_h^2/\sigma_z^2}}{4\sigma_z^2 n^2}\int_{-\infty}^{\infty}\left[e^{-2(y-Y_i)/\sigma_z}\Phi^2\left(\frac{y-Y_i-\sigma_h^2/\sigma_z}{\sigma_h}\right) + e^{2(y-Y_i)/\sigma_z}\left\{1-\Phi\left(\frac{y-Y_i+\sigma_h^2/\sigma_z}{\sigma_h}\right)\right\}^2\right. \\
&\quad \left. + 2\Phi\left(\frac{y-Y_i-\sigma_h^2/\sigma_z}{\sigma_h}\right)\left\{1-\Phi\left(\frac{y-Y_i+\sigma_h^2/\sigma_z}{\sigma_h}\right)\right\}\right]dy \\
&\quad + \sum\sum_{i<j}\omega_i\omega_j\frac{e^{\sigma_h^2/\sigma_z^2}}{2\sigma_z^2 n^2}\int_{-\infty}^{\infty}\left[e^{-(y-Y_i)/\sigma_z}\Phi\left(\frac{y-Y_i-\sigma_h^2/\sigma_z}{\sigma_h}\right) + e^{(y-Y_i)/\sigma_z}\left\{1-\Phi\left(\frac{y-Y_i+\sigma_h^2/\sigma_z}{\sigma_h}\right)\right\}\right] \\
&\quad \times\left[e^{-(y-Y_j)/\sigma_z}\Phi\left(\frac{y-Y_j-\sigma_h^2/\sigma_z}{\sigma_h}\right) + e^{(y-Y_j)/\sigma_z}\left\{1-\Phi\left(\frac{y-Y_j+\sigma_h^2/\sigma_z}{\sigma_h}\right)\right\}\right]dy \\
&\quad - \sum_{i=1}^{n}\omega_i\frac{e^{\sigma_h^2/2\sigma_z^2}}{\sigma_z n^2}\int_{-\infty}^{\infty}\left[\left\{e^{-(y-Y_i)/\sigma_z}\Phi\left(\frac{y-Y_i-\sigma_h^2/\sigma_z}{\sigma_h}\right) + e^{(y-Y_i)/\sigma_z}\left(1-\Phi\left(\frac{y-Y_i+\sigma_h^2/\sigma_z}{\sigma_h}\right)\right)\right\}\right.
\end{aligned}$$

$$\times\left(\sum_{i=1}^{n}\frac{1}{\sqrt{2\pi}\sigma_h}e^{-(y-Y_i)^2/2\sigma_h^2}\right)\Bigg]dy+\int_{-\infty}^{\infty}\left(\frac{1}{n}\sum_{i=1}^{n}\frac{1}{\sqrt{2\pi}\sigma_h}e^{-(y-Y_i)^2/2\sigma_h^2}\right)^2 dy, \qquad (2.1)$$

where $K_h(x) = 1/(\sqrt{2\pi}\sigma_h)e^{-x^2/2\sigma_h^2}$ and $\Phi(x) = \int_{-\infty}^{x} 1/\sqrt{2\pi}e^{-t^2/2}dt$.

Thus optimizing $Q(\omega)$ in (2.1) under the constraints that the weights are non-negative and sum to n leads to a quadratic programming problem:

$$\begin{aligned}&\text{minimize}_{\omega}\ \frac{1}{2}\omega' H\omega + f'\omega\\ &\text{subject to}\ \sum_{i=1}^{n}\omega_i = n,\quad \omega_i \geq 0,\ i = 1,2,\dots,n.\end{aligned}$$

Let $\hat{\omega} = \arg\min_{\omega} Q(\omega)$. Then, we obtain $\hat{f}_{\hat{\omega}}(x) = 1/n\sum_{i=1}^{n}\hat{\omega}_i K_h(x-Y_i)$.

As Equation (2.1) indicates, the coefficients are not available in closed form and hence a numerical method is needed. We speculate that they will be computed relatively quickly through numerical integration method even though we could not obtain a solution to this problem here. Hazelton and Turlach (2009) recommended a penalized form of this criterion in order to regularize the optimization problem.

Next, we will introduce a method of deconvolution density estimation using the support vector regression method in a reproducing kernel Hilbert space(RKHS) with the Gaussian kernel. The following support vector method based on Phillips' residual method (Phillips, 1962) was proposed by Lee (2008) and Mukherjee and Vapnik (1999). In Lee (2008) the following estimator (2.2) was excluded in the simulation owing to the singular problem in K^{-1} and computing difficulties. In this paper the estimator (2.2) will be executed and compared in the simulation.

$$\begin{aligned}&\text{minimize}\ \ \Omega(g) = (g,g)_H = \sum_{i,j=1}^{n}\omega_i\omega_j k(y_i,y_j),\quad g(y,\omega) = \sum_{i=1}^{n}\omega_i k(y_i,y)\\ &\text{subject to}\ \ \max_i \left|G_n(y) - \int_{-\infty}^{y}\sum_{j=1}^{n}\omega_j k\left(y_j,y'\right)dy'\right|_{y=y_i} = \epsilon,\quad \omega_i \geq 0,\ \sum_{i=1}^{n}\omega_i = 1.\end{aligned}$$

Then, the coefficients ω_i's can be found by solving the following quadratic programming problem and applying the equation $\omega = K^{-1}R(\alpha-\alpha^*)$:

$$\begin{aligned}&\text{minimize}\ \ \frac{1}{2}(\alpha-\alpha*)^t R^t K^{-1}R(\alpha-\alpha^*) - \sum_{i=1}^{n}y_i(\alpha_i-\alpha_i^*) + \epsilon\sum_{i=1}^{n}(\alpha_i+\alpha_i^*)\\ &\text{subject to}\ \ 0 \leq \alpha_i^*,\quad \alpha_i \leq C,\ \ i = 1,\dots,n,\end{aligned}$$

where $K = [k_{ij}]_{n\times n}$, $R = [r_{ij}]_{n\times n}$, $r_{ij} = \int_{-\infty}^{y_i}k(y_j,y)dy$.

Then, applying the Fourier inversion formula,

$$\hat{f}(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{\tilde{\hat{g}}(\xi)}{\tilde{q}(\xi)}e^{i\xi x}\,d\xi = \frac{1}{2\pi}\int_{-\infty}^{\infty}\sum_{j=1}^{n}\omega_j\tilde{k}(y_j,\xi)\frac{e^{i\xi x}}{\tilde{q}(\xi)}\,d\xi.$$

Thus in the case of double exponential measurement error $q(z) = 1/(2\sigma_z)e^{-|z|/\sigma_z}$,

$$\begin{aligned}\hat{f}(x) &= \frac{1}{2\pi}\int_{-\infty}^{\infty}\sum_{j=1}^{n}\omega_j\,\tilde{k}(y_j,\xi)\frac{e^{i\xi x}}{\tilde{q}(\xi)}\,d\xi\\ &= \frac{1}{2\pi}\sum_{j=1}^{n}\omega_j\int_{-\infty}^{\infty}e^{-i\xi y_j-0.5\sigma_h^2\xi^2}\left(1+\sigma_z^2\xi^2\right)e^{i\xi x}d\xi\\ &= \sum_{j=1}^{n}\frac{\omega_j}{\sqrt{2\pi}\sigma_h}e^{-(x-y_j)^2/2\sigma_h^2}+\sigma_z^2\sum_{j=1}^{n}\frac{\omega_j}{\sqrt{2\pi}\sigma_h^3}e^{-(x-y_j)^2/2\sigma_h^2}\\ &\quad -\sigma_z^2\sum_{j=1}^{n}\frac{\omega_j}{\sqrt{2\pi}\sigma_h^5}(x-y_j)^2e^{-(x-y_j)^2/2\sigma_h^2},\end{aligned}\tag{2.2}$$

where $k(x,y) = (\sqrt{2\pi}\sigma_h)^{-1}e^{-(x-y)^2/2\sigma_h^2}$ and $\tilde{q}(\xi) = (1+\sigma_z\xi^2)^{-1}$.

Finally, the most popular approach to the deconvolution problem is to estimate $f(x)$ by a kernel estimator and Fourier transform. The deconvolution density estimator (*e.g.* Liu and Taylor, 1989; Fan, 1991) is given by

$$\hat{f}(x) = \frac{1}{2\pi n}\sum_{j=1}^{n}\int_{-\infty}^{\infty}e^{i\xi(x-y_j)}\frac{\tilde{K}(\sigma_h\xi)}{\tilde{q}(\xi)}\,d\xi.$$

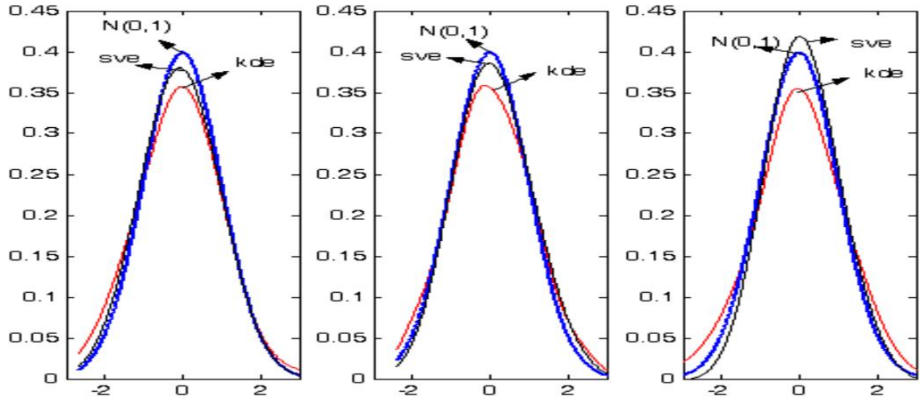
Then using the normalized Gaussian kernel, $\tilde{K}(\sigma_h\xi) = e^{-0.5\sigma_h^2\xi^2}$, and double exponential measurement error $q(z) = 1/(2\sigma_z)e^{-|z|/\sigma_z}$, the classical kernel density estimator $\hat{f}(x)$ (Pensky and Vidakovic, 1999) is evaluated as

$$\begin{aligned}\hat{f}(x) &= \frac{1}{2\pi n}\sum_{j=1}^{n}\int_{-\infty}^{\infty}e^{i\xi(x-y_j)}\frac{\tilde{K}(\sigma_h\xi)}{\tilde{q}(\xi)}\,d\xi\\ &= \frac{1}{\sqrt{2\pi}\sigma_h n}\sum_{j=1}^{n}e^{-0.5\left(\frac{x-y_j}{\sigma_h}\right)^2}\left[1-\frac{\sigma_Z^2}{\sigma_h^2}\left\{\left(\frac{x-y_j}{\sigma_h}\right)^2-1\right\}\right].\end{aligned}$$

3. Simulation and Discussion

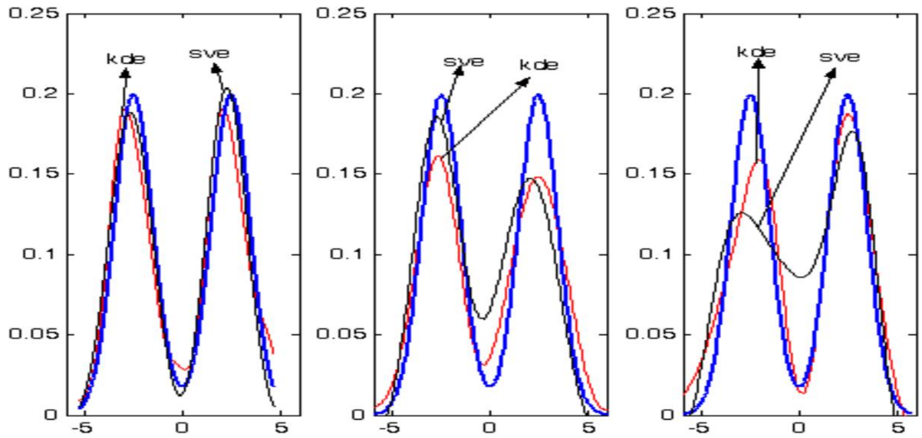
In this section we compare the performance of deconvolution density estimators when measurement errors are double exponential. The empirical distribution function, $G_n(y) = 1/n\sum_{i=1}^{n}I(Y_i \le y)$, is used as an estimator of $G(y)$. Target distributions are selected from distribution functions used in Hazelton and Turlach (2009). The weighted kernel density estimator is excluded here due to computing difficulties in Equation (2.1) and it will be the ongoing research project of the author. In this section the support vector kernel density estimator and the classsical kernel density estimator are compared by a simulation.

The following figures show plots of classical kernel density estimates and support vector kernel density estimates using Phillips' method when 100 points are randomly generated respectively from a target distribution $f(x)$ and a noise distribution, double exponential distribution $q(z)$ with mean zero. Each figure presents the case that the support vector kernel density estimator is best fitted to the exact probability density function $f(x)$ among 30 randomly generated data sets and the classical kernel density estimator is fitted to the data with the best possible parameter($=\sigma_h$). The measurement error



(a) var(Z)/var$(X) = 0.1$ (b) var(Z)/var$(X) = 0.25$ (c) var(Z)/var$(X) = 0.5$

Figure 1: *The simulation study when target density* $f(x)$ *is* $N(0, 1)$



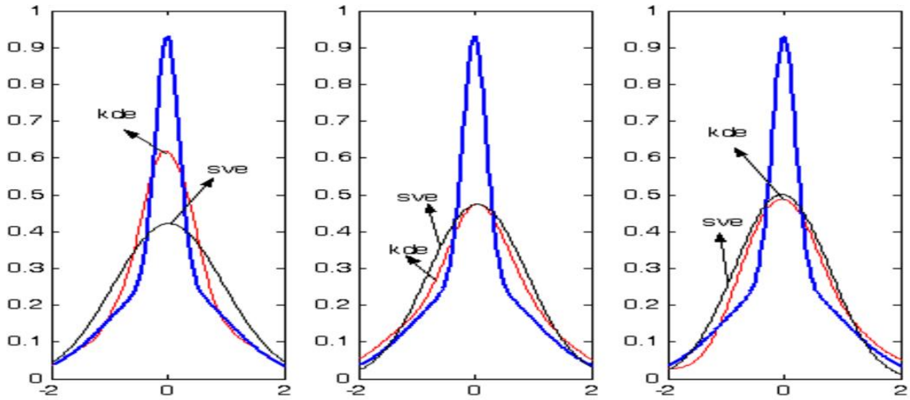
(a) var(Z)/var$(X) = 0.1$ (b) var(Z)/var$(X) = 0.25$ (c) var(Z)/var$(X) = 0.5$

Figure 2: *The simulation study when target density* $f(x)$ *is* $0.5N(-2.5, 1) + 0.5N(2.5, 1)$

variance is set at low (= var(Z)/var$(X) = 0.1$), moderate (= var(Z)/var$(X) = 0.25$), and high levels (= var(Z)/var$(X) = 0.5$) as shown in Hazelton and Turlach (2009). The exact probability density function $f(x)$ is shown in bold lines and the support vector kernel density estimate is shown in dashed lines. For the support vector kernel density estimates, Gunn's program (Gunn, 1998) and MATLAB 6.5 were used.

Figure 1 presents a simulation study when the target distribution is the standard normal probability distribution $f(x)$. The parameters (= σ_h) of classical kernel density estimates corresponding to variance ratios of 0.1, 0.25 and 0.5 are 0.6, 0.6 and 0.75 respectively. The parameters (= σ_h) of support vector kernel density estimates corresponding to variance ratios of 0.1, 0.25 and 0.5 are 0.95, 0.95 and 0.95 respectively and $\epsilon = 0.05$, $C = \infty$ are used.

Figure 2 presents a simulation study when the target distribution is the symmetric bimodal density $0.5N(-2.5, 1) + 0.5N(2.5, 1)$. The parameters (= σ_h) of classical kernel density estimates corresponding to variance ratios of 0.1, 0.25 and 0.5 are 0.75, 1.1 and 1.1 respectively. The parameters (= σ_h) of



(a) var(Z)/var(X) = 0.1 (b) var(Z)/var(X) = 0.25 (c) var(Z)/var(X) = 0.5

Figure 3: *The simulation study when target density* $f(x)$ *is* $2/3N(0, 1) + 1/3N(0, 0.04)$

support vector kernel density estimates corresponding to variance ratios of 0.1, 0.25 and 0.5 are 1.05, 1.1 and 1.1 respectively and $\epsilon = 0.05$, $C = \infty$ are used.

Figure 3 presents a simulation study when the target distribution is the kurtotic density $2/3N(0, 1)+ 1/3N(0, 0.04)$. The parameters ($= \sigma_h$) of classical kernel density estimates corresponding to variance ratios of 0.1, 0.25 and 0.5 are 0.3, 0.5 and 0.5 respectively. The parameters ($= \sigma_h$) of support vector kernel density estimates corresponding to variance ratios of 0.1, 0.25 and 0.5 are 0.95, 0.9 and 0.9 respectively and $\epsilon = 0.05$, $C = \infty$ are used.

As the illustrated figures suggest, most of Figures in 30 data sets did not show that the support vector kernel density estimator using Phillips' method is as good as the classical kernel density estimator. However, the estimator is attractive in the sense that some coefficients in $\omega = K^{-1}R(\alpha - \alpha^*)$ are very close to zero.

4. Concluding Remarks

In this paper three different deconvolution density estimators were introduced when the sample observations are contaminated by double exponentially distributed errors. It was shown that the coefficients of the weighted kernel density estimator based on the Gaussian kernel are not available in closed form. The weighted kernel density estimator was excluded in the simulation owing to the computing difficulties and it will be the ongoing research project of the author. Even though the simulation in this paper is limited, it appears to indicate that the classical(Parzen's) kernel density estimator is better than the support vector kernel density estimator using Phillilps' method. However, the support vector kernel density estimator is attractive in the sense that some coefficients in $\omega = K^{-1}R(\alpha - \alpha^*)$ are very close to zero. But its implementation seems to be more expensive than that of the classical kernel density estimator when the sample size is large.

References

Carroll, R. J. and Hall, P. (1998). Optimal rates of convergence for deconvoluting a density, *Journal of the American Statistical Association*, **83**, 1184–1886.

Fan, J. (1991). On the optimal rates of convergence for nonparametric deconvolution problem, *Annals of Statistics,* **19**, 1257–1272.

Gunn, S. R. (1998). Support Vector Machines for Classification and Regression, Technical report, University of Southampton.

Hall, P. and Qiu, P. (2005). Discrete-transform approach to deconvolution problems, *Biometrika*, **92**, 135–148.

Hazelton, M. L. and Turlach, B. A. (2009). Nonparametric density deconvolution by weighted kernel estimators, *Statistics and Computing*, **19**, 217–228.

Lee, S. (2008). A note on nonparametric density estimation for the deconvolution problem, *Communications of the Korean Statistical Society*, **15**, 939–946.

Lee, S. and Taylor, R. L. (2008). A note on support vector density estimation for the deconvolution problem, *Communications in Statistics: Theory and Methods*, **37**, 328–336.

Liu, M. C. and Taylor, R. L. (1989). A Consistent nonparametric density estimator for the deconvolution problem, *The Canadian Journal of Statistics*, **17**, 427–438.

Louis, T. A. (1991). Using empirical Bayes methods in biopharmaceutical research, *Statistics in Medicine*, **10**, 811–827.

Mendelsohn, J. and Rice, R. (1982). Deconvolution of microfluorometric histograms with B splines, *Journal of the American Statistical Association*, **77**, 748–753.

Mukherjee, S. and Vapnik, V. (1999). Support vector method for multivariate density estimation, *Proceedings in Neural Information Processing Systems*, 659–665.

Pensky, M. and Vidakovic, B. (1999). Adaptive wavelet estimator for nonparametric density deconvolutoin, *Annals of Statistics*, **27**, 2033–2053.

Phillips, D. L. (1962). A technique for the numerical solution of integral equations of the first kind, *Journal of the Association for Computing Machinery*, **9**, 84–97.

Stefanski, L. and Carroll, R. J. (1990). Deconvoluting kernel density estimators, *Statistics*, **21**, 169–184.

Zhang, H. P. (1992). On deconvolution using time of flight information in positron emission tomography, *Statistica Sinica*, **2**, 553–575

Received February 2010; Accepted March 2010