

# A Note on Deconvolution Estimators when Measurement Errors are Normal

Sungho Lee<sup>1,a</sup>

<sup>a</sup>Department of Statistics, Daegu University

---

## Abstract

In this paper a support vector method is proposed for use when the sample observations are contaminated by a normally distributed measurement error. The performance of deconvolution density estimators based on the support vector method is explored and compared with kernel density estimators by means of a simulation study. An interesting result was that for the estimation of kurtotic density, the support vector deconvolution estimator with a Gaussian kernel showed a better performance than the classical deconvolution kernel estimator.

**Keywords:** Deconvolution, kernel estimator, support vector method, reproducing kernel Hilbert space(RKHS).

---

## 1. Introduction

The problem of measurements being contaminated with noise exists in many different fields (*e.g.* Mendelsohn and Rice, 1982; Stefanski and Carroll, 1990; 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 random sample  $Y_1, \dots, Y_n$  from  $Y = X + Z$ . The objective is to estimate the density function  $f(x)$  where  $g(y)$  is the convolution of  $f(x)$  and  $q(z)$ ,  $g(y) = (f * q)(y) = \int_{-\infty}^{\infty} f(y - z)q(z)dz$ . Following the work of Fan (1991), two types of error distributions have been considered: ordinary smooth and super smooth distributions. Gamma or double exponential distribution functions are ordinary smooth, that is, the Fourier transform  $\tilde{q}(\xi) (= \int_{-\infty}^{\infty} e^{-i\xi x} q(x)dx)$  of  $q$  has a polynomial descent. Normal or Cauchy distribution functions are super smooth, that is, the Fourier transform  $\tilde{q}$  of  $q$  has an exponential descent. The case of a normally distributed measurement error is generally more important in practice than that of a double exponential or gamma distributed measurement error. The most popular approach to this deconvolution problem has been to estimate  $f(x)$  using 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 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; Hazelton and Turlach, 2009).

Recently, the support vector method has drawn significant attention on classification and regression problems. The support vector method is a tool to solve multidimensional function estimation problems. It was developed in Russia in the sixties by Vapnik and co-workers (Vapnik and Lerner, 1963; Vapnik and Chervonenkis, 1964). It was initially designed to solve pattern recognition problems. Later the support vector method was extended to regression and real-valued function estimation.

---

This research was supported by the Daegu University Research Grant, 2011.

<sup>1</sup> Professor, Department of Statistics, Daegu University, Kyungbuk 712-714, Korea. E-mail: shlee1@daegu.ac.kr

The support vector regression algorithm (see, for example, Vapnik, 1995) computes a nonlinear function in the space of the input data  $R^m$  by using a linear function in high dimensional feature space  $\mathcal{F}$  with a dot product. The functions take the form  $f(x) = \omega \cdot \Phi(x) + b$  with  $\Phi : R^m \rightarrow \mathcal{F}$  and  $\omega \in \mathcal{F}$ . Weston *et al.* (1999) proposed the support vector method for the density function estimation using a support vector regression algorithm. The support vector regression method based on a reproducing kernel Hilbert space (RKHS) was discussed by Mukherjee and Vapnik (1999). Lee and Taylor (2008) and Lee (2010) applied the support vector method to the density deconvolution problem when the distribution of the error is double exponential.

In this paper two different deconvolution density estimators are briefly reviewed when the sample observations are contaminated by a normally distributed measurement error and then a support vector method is proposed for use in the case of a normally distributed measurement error. The performance of the deconvolution density estimator that uses a support vector method is compared with the classical deconvolution kernel density estimator via a simulation study.

## 2. Deconvolution Estimators when Measurement Errors are Normal

### 2.1. Classical deconvolution kernel estimators

The most popular approach to the deconvolution problem is to estimate  $f(x)$  using a kernel estimator and Fourier transform. The deconvolution kernel estimator (e.g. Carroll and Hall, 1988; Stefanski and Carroll, 1990; Fan, 1991) is

$$\hat{f}(x) = \frac{1}{2\pi n} \sum_{j=1}^n \int_{-\infty}^{\infty} e^{i\xi(x-Y_j)} \tilde{K}(h_n\xi) / \tilde{q}(\xi) d\xi, \quad (2.1)$$

where  $0 < h_n \rightarrow 0$ ,  $n \rightarrow \infty$  is a smoothing parameter called the bandwidth and  $\tilde{K}(\xi)$  is the Fourier transform of  $K$ . Carroll and Hall (1988) showed that if  $f(x)$  has  $m$  bounded derivatives and errors are normal, then the fastest attainable pointwise convergence rate of any nonparametric estimator of  $f(x)$  is only  $(\log n)^{-m/2}$ . The deconvolution kernel estimator (2.1) achieves the optimal rates (Stefanski and Carroll, 1990). The asymptotic theory gives that the optimal rates of convergence for a normal measurement error is very slow and hence very large samples may be required before the asymptotics take effect. Since the normal distribution is frequently used in application, Fan (1992) investigated how large a noise level is acceptable. To this goal, let  $Y_1, \dots, Y_n$  be a random sample from  $Y = X + Z$  and  $Z = \sigma_0\epsilon$ , where  $\sigma_0$  parametrizes the noise level and  $\epsilon \sim N(0, 1)$ . He showed that if  $\sigma_0 = \sqrt{\sigma_Z^2} = O(n^{-1/(1+2m)})$ , then deconvolution is just as difficult as the ordinary density estimation in terms of the rates of convergence (Theorem 4 in Fan, 1992). Thus, when  $\sigma_Z^2$  is small, the usual normalized Gaussian kernel can still be practical to calculate a deconvolution kernel estimator. He suggested a deconvolution kernel estimator (2.2) that uses the normalized Gaussian kernel,  $\tilde{K}(h_n\xi) = e^{-0.5h_n^2\xi^2}$ :

$$\begin{aligned} \hat{f}(x) &= \frac{1}{2\pi n} \sum_{j=1}^n \int_{-\infty}^{\infty} e^{i\xi(x-Y_j)} \tilde{K}(h_n\xi) / \tilde{q}(\xi) d\xi \\ &= \frac{1}{2\pi n} \sum_{j=1}^n \int_{-\infty}^{\infty} e^{i\xi(x-Y_j)} e^{-0.5\xi^2(h_n^2 - \sigma_Z^2)} d\xi \\ &= \frac{1}{n} \sum_{j=1}^n \frac{1}{\sqrt{2\pi(h_n^2 - \sigma_Z^2)}} e^{-(x-Y_j)^2/2(h_n^2 - \sigma_Z^2)}, \quad h_n^2 > \sigma_Z^2 \end{aligned} \quad (2.2)$$

and compared (via simulation studies) the performance of the estimator (2.2) with the following deconvolution kernel estimator (2.4) which is well-known in the presence of normal measurement error. In general, the estimator (2.2) is not appropriate for a deconvolution kernel estimator because the condition,  $h_n \rightarrow 0$  as  $n \rightarrow \infty$ , is not satisfied. Thus, when  $\sigma_Z^2$  is large, the estimator (2.2) cannot be used clearly. In this case, in order to avoid problems of integrability, a kernel  $K$  that has compactly supported Fourier transform  $\tilde{K}(t) = (1 - t^2)^3 I_{[-1,1]}(t)$  is used in common. The Fourier transform  $\tilde{K}(t)$  corresponds to the kernel function  $K$ ,

$$K(x) = \frac{48 \cos x}{\pi x^4} \left(1 - \frac{15}{x^2}\right) - \frac{144 \sin x}{\pi x^5} \left(2 - \frac{5}{x^2}\right). \tag{2.3}$$

Using this kernel  $\tilde{K}(t) = (1 - t^2)^3 I_{[-1,1]}(t)$ , a deconvolution kernel estimator  $\hat{f}(x)$  in the presence of normal measurement error can be calculated as follows:

$$\begin{aligned} \hat{f}(x) &= \frac{1}{2\pi n} \sum_{j=1}^n \int_{-\infty}^{\infty} e^{i\xi(x-Y_j)} \tilde{K}(h_n \xi) / \tilde{q}(\xi) d\xi \\ &= \frac{1}{\pi n h_n} \sum_{j=1}^n \int_0^1 \left( \cos \xi \left( \frac{x - Y_j}{h_n} \right) \right) (1 - \xi^2)^3 e^{\xi^2 \sigma_Z^2 / 2 h_n^2} d\xi. \end{aligned} \tag{2.4}$$

Fan (1992) gave an optimal rate of convergence(=  $O((\log n)^{-m/2})$ ) of this estimator (2.4) in terms of the mean of the weighted  $L^p$ -norms and presented several simulation studies that illustrates the difficulty of deconvolution. Wand (1998) derived formulae for an exact computation of the mean integrated squared error(MISE) for a collection of target densities and gave some examples of these calculations using the estimator (2.4). For example, if  $f(x)$  is a mixture normal distribution, then the best possible rate of convergence of MISE of the estimator (2.4) is of order  $(\log n)^{-1}$ . The results in Wand (1998) indicates that, for high levels of a normal measurement error, Gaussian deconvolution is difficult.

### 2.2. Weighted deconvolution kernel estimators

Recently Hazelton and Turlach (2009) proposed a weighted kernel density estimator for the deconvolution problem. In cases with the Gaussian kernel and normal measurement error  $Z \sim N(0, \sigma_Z^2)$ , they showed that  $\hat{f}_\omega(x)$  has the simple expression as follows:

$$\hat{f}_\omega(x) = \frac{1}{n} \sum_{i=1}^n \omega_i K_h(x - Y_i), \quad \omega_i \geq 0, \quad \sum_{i=1}^n \omega_i = n, \quad K_h(x) = (\sqrt{2\pi}\sigma_h)^{-1} e^{-x^2/2\sigma_h^2},$$

where unknown weight vector  $\omega$  will be estimated based on  $Q(\omega)$ ,

$$\begin{aligned} Q(\omega) &= \int_{-\infty}^{\infty} (\hat{f}_\omega * q(y) - \hat{g}(y))^2 dy \\ &= \frac{1}{n^2} \left( \sum_{i=1}^n \sum_{j=1}^n \omega_i \omega_j \phi_{\sqrt{2}\lambda}(Y_i - Y_j) + \sum_{i=1}^n \sum_{j=1}^n \phi_{h\sqrt{2}}(Y_i - Y_j) - 2 \sum_{i=1}^n \sum_{j=1}^n \omega_i \phi_\nu(Y_i - Y_j) \right), \end{aligned}$$

where  $\phi_\lambda(x) = 1/(\sqrt{2\pi}\sigma_\lambda)e^{-x^2/2\sigma_\lambda^2}$ ,  $\phi_h(x) = 1/(\sqrt{2\pi}\sigma_h)e^{-x^2/2\sigma_h^2}$ ,  $\sigma_\lambda^2 = \sigma_h^2 + \sigma_Z^2$ ,  $\sigma_\nu^2 = \sigma_Z^2 + 2\sigma_h^2$ , and  $\hat{g}(y) = 1/n \sum_{i=1}^n K_h(y - Y_i)$ .

One immediate attraction of  $\hat{f}_\omega(x)$  is that it will never take negative values. They also showed that if the optimal weighting scheme  $\omega_i (= f(Y_i)/g(Y_i))$  was known, then the estimator would have the MISE of an asymptotic order of  $n^{-4/5}$ . Through simulation they argued that  $\hat{f}_\omega(x)$  is a competitive estimator over the classical deconvolution kernel estimator.

### 2.3. Support vector deconvolution estimators

Now we will introduce the other method of estimation of a deconvolution density using the support vector regression method based on a reproducing kernel Hilbert space (RKHS). A (real) RKHS  $H$  is a Hilbert space of real-valued functions  $f$  on an interval  $\tau$  with the property that, for each  $t \in \tau$ , the evaluation functional  $L_t, L_t : f \rightarrow f(t)$ , is a bounded linear functional. Then, by Riesz representation theorem, for each  $t \in \tau$  there exists a unique element  $K_t \in H$  such that for each  $f \in H$ ,  $L_t(f) = f(t) = (K_t, f)$ . The function defined by  $K(u, v) = (K_u, K_v)$  for  $u, v \in \tau$  is the reproducing kernel. Then, by the Moore-Aronszajn Theorem (Aronszajn, 1950), to every positive definite function  $K$  on  $\tau \times \tau$  there corresponds a unique RKHS  $H_K$  of real valued functions on  $\tau$  with  $K$  as its reproducing kernel. Note that any positive definite function  $K(u, v)$  has an expansion  $K(u, v) = \sum_{i=1}^{\infty} \lambda_i \phi_i(u) \phi_i(v)$ . Let us consider the set of functions,  $f(x, \omega) = \sum_{i=1}^{\infty} \omega_i \phi_i(x)$ , and define the inner product as  $(f(x, \omega), f(x, \omega^*)) = \sum_{i=1}^{\infty} \omega_i \omega_i^* / \lambda_i$ . Then we have a RKHS  $H_K$  with its reproducing kernel  $K$  and will apply these properties of RKHS to the estimation of a deconvolution density. The following Bochner's theorem states that the Fourier transform of a positive measure constitutes a positive definite kernel.

**Theorem 1. (Bochner, 1959)** *A function  $K(s - t)$  is positive definite if and only if it is the Fourier transform of a symmetric, positive function  $\tilde{K}(\xi)$  decreasing to 0 at infinity.*

The Gaussian kernel represents a legitimate inner product in feature space  $\mathcal{F}$  and satisfies Mercer's condition. Let  $C_0(R)$  denote the set of continuous functions on  $R$  that vanish at infinity. Then the reproducing kernel Hilbert space  $H_\sigma$  (Vert and Vert, 2006) associated with the normalized Gaussian kernel  $K(x, y) = (\sigma \sqrt{2\pi})^{-1} e^{-(x-y)^2/2\sigma^2}$  is

$$H_\sigma = \left\{ f \in C_0(R) : f \in L_1(R) \text{ and } \int_R |\tilde{f}(\xi)|^2 e^{\sigma^2 \xi^2 / 2} d\xi < \infty \right\}$$

and the associated dot product is given by

$$\langle f, g \rangle_{H_\sigma} = \frac{1}{2\pi} \int_R \tilde{f}(\xi) \tilde{g}(\xi)^* e^{\sigma^2 \xi^2 / 2} d\xi.$$

Since  $\tilde{K}(\xi) (= (1 - \xi^2)^3 I_{[-1,1]}(\xi))$  is also a symmetric, positive function decreasing to at infinity, the reproducing kernel Hilbert space  $H$  associated with the kernel  $K$ ,

$$K(x) = \frac{48 \cos x}{\pi x^4} \left( 1 - \frac{15}{x^2} \right) - \frac{144 \sin x}{\pi x^5} \left( 2 - \frac{5}{x^2} \right)$$

is a subspace of  $L_2(R)$  with norm,  $\|f\|_H^2 = 1/(2\pi) \int_R |\tilde{f}(\xi)|^2 / \tilde{K}(\xi) d\xi < \infty$ .

The following support vector method based on Phillips' residual method (Phillips, 1962) was proposed by Mukherjee and Vapnik (1999) and Lee (2010). In order to estimate  $f(x)$ , first,  $g(y)$  will be estimated using the reproducing kernel of RKHS. Let

$$g(y, \omega) = \sum_{i=1}^n \omega_i K_h(y, Y_i), \quad \omega_i \geq 0, \quad \sum_{i=1}^n \omega_i = 1$$

and

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

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

$$\text{minimize } \frac{1}{2} (\alpha - \alpha^*)^t R^t \Gamma_h^{-1} R (\alpha - \alpha^*) - \sum_{i=1}^n Y_i (\alpha_i - \alpha_i^*) + \epsilon \sum_{i=1}^n (\alpha_i + \alpha_i^*), \quad 0 \leq \alpha_i^*, \alpha_i \leq C, i = 1, \dots, n$$

where  $\Gamma_h = [K_h(Y_i, Y_j)]_{n \times n}$ ,  $R^t = [r_{ij}]_{n \times n}$ ,  $r_{ij} = \int_{-\infty}^{Y_i} K_h(y, Y_j) dy = \int_{-\infty}^{Y_i} (1/h_n) K((y - Y_j)/h_n) dy$ .

Then, applying the Fourier inversion formula,

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

When  $Z \sim N(0, \sigma_Z^2)$  and  $\sigma_Z^2$  is small, Gaussian kernel can be used:

$$\begin{aligned} \hat{f}(x) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \sum_{j=1}^n \omega_j \tilde{K}_h(Y_j, \xi) e^{i\xi x} / \tilde{q}(\xi) d\xi \\ &= \frac{1}{2\pi} \sum_{j=1}^n \omega_j \int_{-\infty}^{\infty} e^{-i\xi Y_j - 0.5 h_n^2 \xi^2} e^{0.5 \sigma_Z^2 \xi^2} e^{i\xi x} d\xi \\ &= \sum_{j=1}^n \frac{\omega_j}{\sqrt{2\pi(h_n^2 - \sigma_Z^2)}} e^{-(x - Y_j)^2 / 2(h_n^2 - \sigma_Z^2)}, \quad h_n^2 > \sigma_Z^2 \end{aligned} \tag{2.5}$$

where

$$\begin{aligned} \tilde{K}_h(Y_j, \xi) &= \int_{-\infty}^{\infty} K_h(Y_j, y) e^{-i\xi y} dy \\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi} h_n} e^{-(y - Y_j)^2 / 2h_n^2} e^{-i\xi y} dy \\ &= e^{-i\xi Y_j - 0.5 h_n^2 \xi^2}. \end{aligned}$$

When  $Z \sim N(0, \sigma_Z^2)$  and  $\sigma_Z^2$  is large, using kernel (2.3)

$$\begin{aligned} \hat{f}(x) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \sum_{j=1}^n \omega_j \tilde{K}_h(Y_j, \xi) e^{i\xi x} / \tilde{q}(\xi) d\xi \\ &= \frac{1}{2\pi} \sum_{j=1}^n \omega_j \int_{-1}^1 (1 - h_n^2 \xi^2)^3 e^{-i\xi Y_j} e^{0.5 \sigma_Z^2 \xi^2} e^{i\xi x} d\xi \\ &= \frac{1}{\pi} \sum_{j=1}^n \omega_j \int_0^1 (1 - h_n^2 \xi^2)^3 \cos(\xi(x - Y_j)) e^{0.5 \sigma_Z^2 \xi^2} d\xi \end{aligned} \tag{2.6}$$

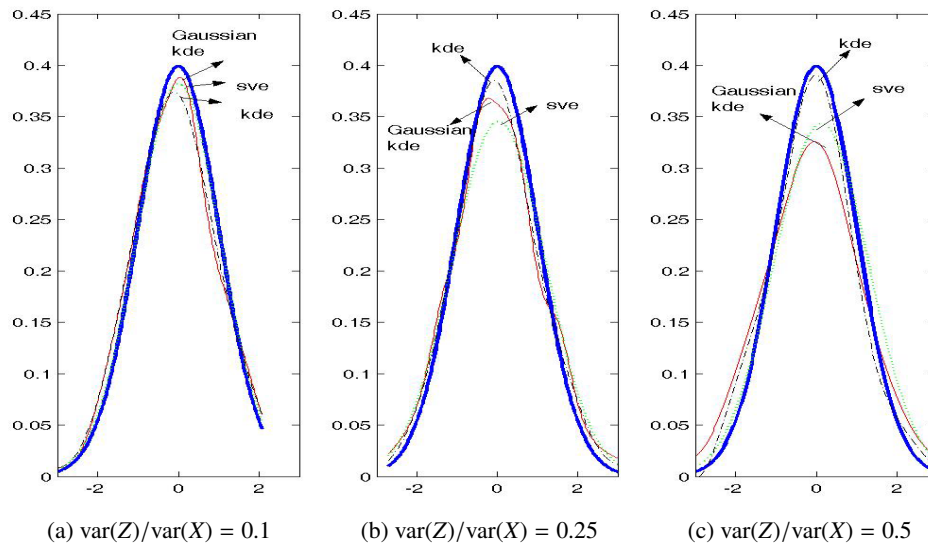


Figure 1: The simulation study when target density  $f(x)$  is  $N(0, 1)$

where

$$\begin{aligned} \tilde{K}(Y_j, \xi) &= \int_{-\infty}^{\infty} K(Y_j, y) e^{-i\xi y} dy \\ &= \int_{-\infty}^{\infty} \left\{ \frac{48 \cos(Y_j - y)}{\pi(Y_j - y)^4} \left( 1 - \frac{15}{(Y_j - y)^2} \right) - \frac{144 \sin(Y_j - y)}{\pi(Y_j - y)^5} \left( 2 - \frac{5}{(Y_j - y)^2} \right) \right\} e^{-i\xi y} dy \\ &= e^{-i\xi Y_j} (1 - \xi^2)^3 I_{[-1,1]}(\xi). \end{aligned}$$

### 3. Simulation and Discussion

In this section the support vector deconvolution estimator (2.5) and the classical deconvolution estimators (2.2) and (2.4) are compared via simulation studies when measurement errors are normal,  $Z \sim N(0, \sigma_Z^2)$ , and  $\sigma_Z^2$  is small or large. Target distributions are selected from distribution functions used in Hazelton and Turlach (2009). The empirical distribution function,  $G_n(y) = 1/n \sum_{i=1}^n I(Y_i \leq y)$ , is used as an estimator of  $G(y)$ . The support vector deconvolution estimator (2.6) was also tried to be compared, but unfortunately it showed the worst performance among four estimators. Thus only a part of simulations for the estimator (2.6) is introduced in Figure 4. The estimator (2.5) showed superior performance than the estimator (2.6) in almost all the simulations. Probably, the reason for these results seems to be due to computing difficulties in the estimation of  $\omega$ 's and numerical integration (Delaigle and Gijbels, 2007).

The Figure 1~Figure 3 show the plots of the classical deconvolution estimates (2.2) and (2.4) and the support vector deconvolution estimates (2.5) when 100 points are randomly generated respectively from a target distribution  $f(x)$  and a noise distribution, normal distribution  $q(z)$  with mean zero. The measurement error variance is set at low ( $= \text{var}(Z)/\text{var}(X) = 0.1$ ), moderate ( $= \text{var}(Z)/\text{var}(X) = 0.25$ ), and high levels ( $= \text{var}(Z)/\text{var}(X) = 0.5$ ) as shown in Hazelton and Turlach (2009). The exact probability density function  $f(x)$  is shown in a bold line and the support vector deconvolution estimate is

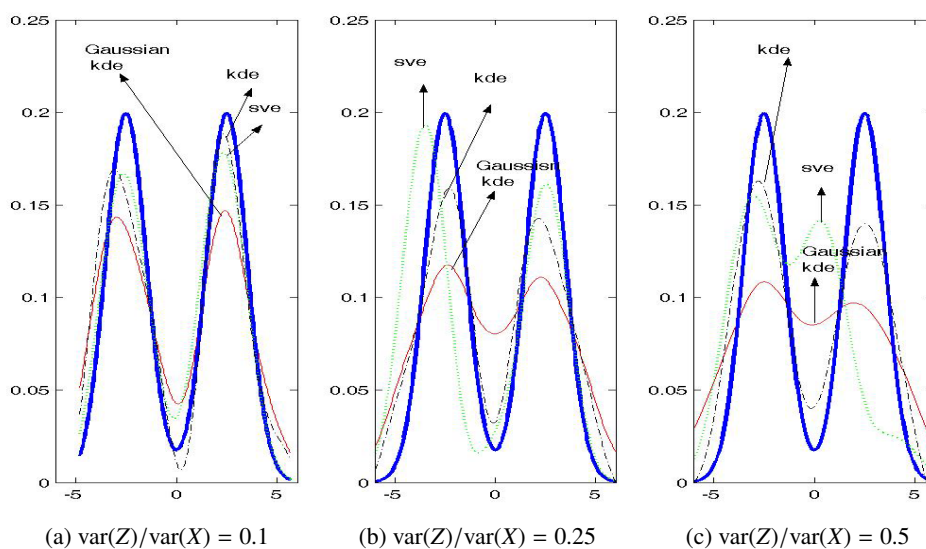


Figure 2: The simulation study when target density  $f(x)$  is  $0.5N(-2.5, 1) + 0.5N(2.5, 1)$

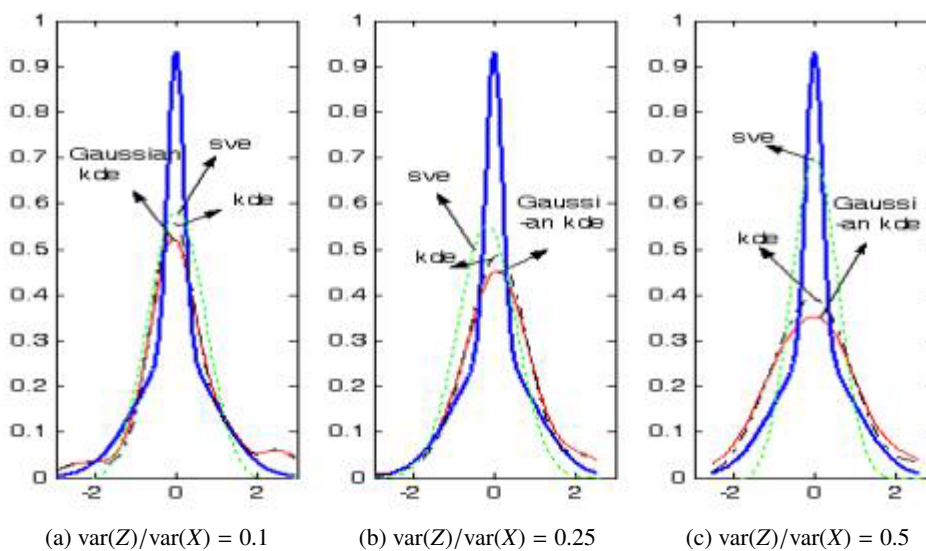


Figure 3: The simulation study when target density  $f(x)$  is  $2/3N(0, 1) + 1/3N(0, 0.04)$

shown in dashed lines. For the support vector deconvolution estimates, Gunn's program (Gunn, 1998) and MATLAB 6.5 were used. Each estimate was picked with the best possible value of parameters based on the exact probability density function  $f(x)$ .

Figure 1 presents the simulation study when a random sample of size 100 is generated from the standard normal probability distribution  $f(x)$ , and normal distribution  $q(z)$ . The parameters ( $= h_n$ ) of the kernel density estimator (2.2) and (2.4) corresponding to variance ratios of 0.1, 0.25 and 0.5 are 0.5, 0.6, 0.9 and 0.21, 0.21, 0.25 respectively.

The parameters ( $= h_n$ ) of the support vector density estimator corresponding to variance ratios of

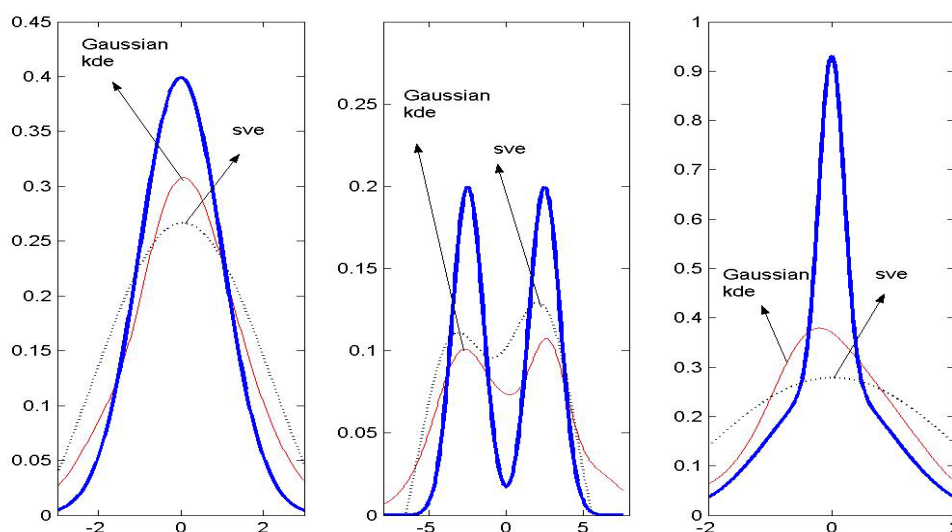


Figure 4: The simulation study for the estimator (2.6) when  $\text{var}(Z)/\text{var}(X) = 0.5$

0.1, 0.25 and 0.5 are 1.0, 1.1, 1.1 respectively and  $\epsilon = 0.05$ ,  $C = \infty$  are used.

Figure 2 presents the simulation study when a random sample of size 100 is generated from the symmetric bimodal density  $0.5N(-2.5, 1) + 0.5N(2.5, 1)$ . The parameters ( $= h_n$ ) of the kernel density estimator (2.2) and (2.4) corresponding to variance ratios of 0.1, 0.25 and 0.5 are 1.15, 1.6, 2.2 and 0.32, 0.40, 0.55 respectively. The parameters ( $= h_n$ ) of the support vector density estimator corresponding to variance ratios of 0.1, 0.25 and 0.5 are 1.1, 1.5, 2.1 respectively and  $\epsilon = 0.05$ ,  $C = \infty$  are used.

Figure 3 presents the simulation study when a random sample of size 100 is generated from the kurtotic density  $2/3N(0, 1) + 1/3N(0, 0.04)$ . The parameters ( $= h_n$ ) of the kernel density estimator (2.2) and (2.4) corresponding to variance ratios of 0.1, 0.25 and 0.5 are 0.45, 0.6, 0.8 and 0.15, 0.20, 0.25 respectively. The parameters ( $= h_n$ ) of the support vector density estimator corresponding to variance ratios of 0.1, 0.25 and 0.5 are 0.75, 0.8, 0.8 respectively and  $\epsilon = 0.05$ ,  $C = \infty$  are used.

Figure 4 presents the simulation study when a random sample of size 100 is generated from the standard normal probability distribution  $f(x)$ , and normal distribution  $q(z)$ . The parameters ( $= h_n$ ) of the kernel density estimator (2.2) corresponding to variance ratio of 0.5 are 0.9, 2.2, 0.8 and the parameters ( $= h_n$ ) of the support vector deconvolution estimator (2.6) corresponding to a variance ratio of 0.5 are 0.2, 0.2, 0.2 respectively and  $\epsilon = 0.05$ ,  $C = \infty$  are used.

Almost all the Figures show that the support vector deconvolution estimator with the Gaussian kernel (2.5) is as good as the classical deconvolution kernel estimator (2.2) and (2.4). Its implementation of the classical deconvolution kernel estimator (2.4) is more expensive than that of the classical kernel density estimator (2.2) and the support vector deconvolution estimator with the Gaussian kernel (2.5). Fan (1992) indicates, the classical deconvolution estimator with the Gaussian kernel is good for a small variance of  $Z$ ; however, not appropriate for a large variance of  $Z$ . An interesting result of the figures is that for the estimation of kurtotic density the support vector deconvolution estimator (2.5) shows good performance for the classical deconvolution kernel estimator (2.2) and (2.4) as Figure 3 indicates.



#### 4. Concluding Remarks

In this paper different deconvolution density estimators were introduced and explored when the sample observations are contaminated by a normally distributed error. Even though the simulation in this paper is limited, it appears to indicate that the support vector deconvolution estimator with a Gaussian kernel (2.5) shows a good performance for the classical deconvolution kernel estimators in  $\omega = \Gamma_h^{-1}R(\alpha - \alpha^*)$  the estimation of kurtotic density. Note that the support vector deconvolution estimator is attractive in the sense that some coefficients are very close to zero. Its implementation of the classical deconvolution kernel estimator (2.4) seems to be more expensive than that of the classical kernel density estimator (2.2) and the support vector deconvolution estimator with Gaussian kernel (2.5). A simulation study of the support vector deconvolution estimator (2.6) did not show the expected performance for other estimators. However, we speculate that the estimator (2.6) will show better performance through the improvement of a numerical integration method.

#### References

- Aronszajn, N. (1950). Theory of reproducing kernels, *Transactions of the American Mathematical Society*, **68**, 337–404.
- Bochner, S. (1959). *Lectures on Fourier Integral*, Princeton University Press, Princeton, New Jersey.
- Carroll, R. J. and Hall, P. (1988). Optimal rates of convergence for deconvoluting a density, *Journal of the American Statistical Association*, **83**, 1184–1886.
- Delaigle, A. and Gijbels, I. (2007). Frequent problems in calculating integrals and optimizing objective functions: A case study in density estimation, *Statistics and Computing*, **17**, 349–355.
- Fan, J. (1991). On the optimal rates of convergence for nonparametric deconvolution problem, *Annals of Statistics*, **19**, 1257–1272.
- Fan, J. (1992). Deconvolution with supersmooth distribution, *The Canadian Journal of Statistics*, **20**, 159–169.
- 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.
- Hazleton, M. L. and Turlach, B. A. (2009). Nonparametric density deconvolution by weighted kernel estimators, *Statistics and Computing*, **19**, 217–228.
- Lee, S. (2010). A support vector method for the deconvolution problem, *Communications of the Korean Statistical Society*, **17**, 451–457.
- 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.
- 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*, Technical Report A.I. Memo no. 1653, MIT AI Lab.
- Pensky, M. and Vidakovic, B. (1999). Adaptive wavelet estimator for nonparametric density deconvolution, *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.
- Vapnik, V. (1995). *The Nature of Statistical Learning Theory*, Springer Verlag, New York.
- Vapnik, V. and Chervonenkis, A. (1964). A note on one class of perceptrons, *Automation and Remote Control*, **25**, 103–109.
- Vapnik, V. and Lerner, L. (1963). Pattern Recognition using generalized portrait method, *Automation and Remote Control*, **24**.
- Vert, R. and Vert, J. (2006). Consistency and convergence rates of one-class svms and related algorithms, *Journal of Machine Learning Research*, **7**, 817–854.
- Wand, M. P. (1998). Finite sample performance of deconvolving density estimators, *Statistics and Probability Letters*, **37**, 131–139.
- Weston, J., Gammerman, A., Stitson, M., Vapnik, V., Vovk, V. and Watkins, C. (1999). Support vector density estimation. In Scholkopf, B. and Smola, A., editors, *Advances in Kernel Methods-Support Vector Learning*, 293–306, MIT Press, Cambridge, MA.
- Zhang, H. P. (1992). On deconvolution using time of flight information in positron emission tomography, *Statistica Sinica*, **2**, 553–575.

Received March 22, 2012; Revised June 19, 2012; Accepted June 20, 2012