The Korean Communications
in Statistics Vol. 4, No. 3, 1997
pp. 833 - 845

A Study on Log-Fourier Deconvolution[1)]

Ja-Yong Koo[2)] and Hyun-Suk Park[3)]

Abstract

Fourier expansion is considered for the deconvolution problem of estimating a probability density function when the sample observations are contaminated with random noise. In the log-Fourier method of density estimation for data without noise, the logarithm of the unknown density function is approximated by a trigonometric function, the unknown parameters of which are estimated by maximum likelihood. The log-Fourier density estimation method, which has been considered theoretically by Koo and Chung (1997), is studied for the finite-sample case with noise. Numerical examples using simulated data are given to show the performance of the log-Fourier deconvolution.

1. Introduction

Suppose that Y is an observable random variable such that

$$Y=X+Z.$$

Here we assume that X and Z are independent random variables with probability density functions f_X and f_Z, respectively. Then Y has the density

$$f_Y(y)=\int f_Z(y-x)f_X(x)dx. \tag{1}$$

Assuming f_Z is known, we will consider a deconvolution problem of estimating f_X from a set of independent observations $Y_1, \dots, Y_n$ having the common density f_Y.

Integral equations such as (1) are known as Fredholm equations of the first kind. A mathematical inversion problem of solving equation (1) for f_X, when f_Y and f_Z are given, is often ill-posed, since a small change incurred in measuring f_Y would give rise to large perturbations in the solution. This fact makes our deconvolution problem somewhat more difficult.

Such a model of measurements being contaminated with error exists in many different fields and has been widely studied in both theoretical and applied settings. Recent related works

1) This paper was supported by a grant from Hallym University, 1996.
2) Associate professor, Department of Statistics, Hallym University, Chunchon, 200-702, Korea.
3) Candidate of Ph. D., Department of Statistics, Hallym University, Chunchon, 200-702, Korea.

include Mendelsohn and Rice (1982), Carroll and Hall (1988), Liu and Taylor (1989), Stefanski and Carroll (1990), Zhang (1990), Fan (1991), and Koo and Chung (1997). Koo (1993) considered inverse problems including deconvolution in a regression framework. Silverman *et al.* (1990), Vardi and Lee (1993), and Koo and Park (1996) used the EM algorithm for indirect estimation problems.

The approximation of log-densities in direct problem, where the main interest lies in the estimation of f_Y based on $Y_1, \ldots, Y_n$, has been considered by many people. Related works include Stone and Koo (1986), Stone (1990), Kooperberg and Stone (1991, 1992), Barron and Sheu (1991), Koo (1996), Koo and Kim (1996), and Koo, Lee and Park (1997). Estimates of density functions based on exponential families have advantages that they are automatically positive and integrate to one. For other traditional methods for nonparametric density estimation, such as kernel estimators and orthogonal series expansions of the density rather than the log-density, refer to Devroye and Gyofi (1985) and Silverman (1986).

In this paper we use numerical simulation to study the finite-sample performance of the method proposed by Koo and Chung (1997). The approach taken here is to seek a solution with f_X in an exponential family determined by trigonometric functions. In this way positivity and integrability (to one) of the estimate are ensured. The difference is that while direct maximum likelihood density estimation in exponential families [Koo, Lee and Park (1997)] chooses the parameters to match the moments of the trigonometric functions to the sample moments, in the deconvolution problem the moment of each trigonometric function is related to a corresponding moment of the direct problem by a factor given by a Fourier coefficient λ_ν of f_Z. Thus an appropriate analogue of the maximum likelihood estimate is obtained by matching moments with respect to f_X to $1/\lambda_\nu$ times the empirical moments associated with the sample from f_Y.

2. Log-Fourier Models

In this section we describe log-Fourier density estimation based on a Fourier expansion having a fixed number of trigonometric functions.

Consider the case when X takes values in the unit interval $I=[0,1]$. Let $B_1, \ldots, B_J$ be a subset of the set of trigonometric functions consisting of $\{\cos 2\pi jx, \sin 2\pi jx\}$ and let $\boldsymbol{B}=\{B_1, \ldots, B_J\}$. Let $\theta=(\theta_1, \ldots, \theta_J)'$ be a J-dimensional vector, and set $s(\cdot\,;\theta)=\sum_{j=1}^{J}\theta_j B_j$. Consider the log-Fourier model

$$f(x;\theta)=\exp(s(x;\theta)-C(\theta)), \quad x \in I,$$

where

$$C(\theta)=\log \int_I \exp(s(x;\theta))dx.$$

Let Θ be the set of θ such that $C(\theta) < \infty$. Then $f(\cdot;\theta)$, $\theta \in \Theta$ is a positive density function on I. A constant function on I will not used to make the log-Fourier models identifiable.

Let $X_1, \ldots, X_n$ be a random sample of size n from the unknown density f_X. The log-likelihood function corresponding to the logspline family is defined by

$$l_X(\theta) = \frac{1}{n} \sum_{m=1}^{n} \log f(X_m;\theta) = \sum_{j=1}^{J} \theta_j \bar{B}_j - C(\theta), \tag{2}$$

where $\bar{B}_j = n^{-1} \sum_{m=1}^{n} B_j(X_m)$. The maximizing value $\hat{\theta}$ of $l_X(\theta)$ over $\theta \in \Theta$ is referred to as the maximum likelihood estimator (MLE). Let $H(\theta)$, $\theta \in \Theta$ be the Hessian matrix of $C(\theta)$ with elements

$$\int_I B_j(x) B_k(x) f(x;\theta) dx - \int_I B_j(x) f(x;\theta) dx \int_I B_k(x) f(x;\theta) dx.$$

Since $H(\theta)$ is strictly concave, the MLE $\hat{\theta}$ is unique if it exists. The Newton-Raphson method can be used to compute $\hat{\theta}$ as in Koo, Lee and Park (1997).

3. Log-Fourier deconvolution

Consider the problem of finding unbiased estimators of $E\bar{B}_j = EB_j(X)$ where X_m's are not observable but only Y_m's are available. For a random variable U, let ϕ_U denote the characteristic function of U which is defined by $\phi_U(t) = Ee^{itU}$, and define C_U and S_U by $\phi_U(t) = C_U(t) + iS_U(t)$. From the assumption that X and Z are independent, we have the relation

$$\phi_X(t) = \phi_Y(t)/\phi_Z(t). \tag{3}$$

Here and after it is assumed that $\phi_Z(t) \neq 0$ for all t. Equivalently, we have

$$C_X = (C_Y C_Z + S_Y S_Z)/|\phi_Z|^2 \text{ and } S_X = (S_Y C_Z - C_Y S_Z)/|\phi_Z|^2,$$

where $|\phi_Z|^2 = C_Z^2 + S_Z^2$. Let $\hat{C}_Y(t) = n^{-1}\sum_m \cos(tY_m)$ and $\hat{S}_Y(t) = n^{-1}\sum_m \sin(tY_m)$. The relation (3) implies that $n^{-1}\sum_m e^{itY_m}/\phi_Z(t)$ has the same expectation as $n^{-1}\sum_m e^{itX_m}$ or equivalently that

$$\hat{C}_X = (\hat{C}_Y C_Z + \hat{S}_Y S_Z)/|\phi_Z|^2 \text{ and } \hat{S}_X = (\hat{S}_Y C_Z - \hat{C}_Y S_Z)/|\phi_Z|^2 \tag{4}$$

are unbiased estimators of C_X and S_X; see Lemma 1 in Appendix for the proof of this fact. From (4), we can find unbiased estimators $\hat{B}_j$ of $EB_j(X)$ based on the observable data Y_m's although X_m's are not observable; for example, $\hat{B}_j = \hat{C}_X(2\pi j)$ when $B_j(x) = \cos(2\pi jx)$.

Now we give a density estimator of f_X when only Y_m's are available. We define the indirect log-likelihood function by

$$l_Y(\theta)=\sum_{j=1}^{J}\theta_j\ \hat{B}_j-C(\theta),\ \ \theta\in\Theta. \tag{5}$$

Though the log-likelihood function l_X defined by (2) can be interpreted as a log-likelihood if X_m's are observable, the indirect log-likelihood function $l_Y(\theta)$ in (5) may not be necessarily interpretable as a log-likelihood; it has been introduced by Koo and Chung (1997) as an object function to be optimized for the definition of density estimator. Let $\hat{\theta}$ be the maximizer of the indirect log-likelihood $l_Y(\theta)$ over $\theta\in\Theta$ and refer to $\hat{\theta}$ as the maximum indirect likelihood estimator (MILE). We can note that $\hat{\theta}$ should satisfy the equation $S(\hat{\theta})=0$, where $S(\hat{\theta})$ is the J-dimensional vector of elements $\hat{B}_j-\partial C(\theta)/\partial\theta_j$. Since the Hessian matrix $H(\theta)$ of $C(\theta)$ is strictly convex for $\theta\in\Theta$, the MILE is unique if it exists. The MILE $\hat{f}$ of f_X is now defined by $\hat{f}=f(\cdot\ ;\hat{\theta})$.

Consider the case when X takes values in $[L,U]$ which may not be equal to I. First we define the MILE when $L=-\infty$ and $U=\infty$. Let $Y_{(1)}$ and $Y_{(n)}$ be the first and nth order statistic of $Y_1,\ldots,Y_n$. For suitable constants y_L and y_U to be defined below, let $W_m=aY_m+b,\ m=1,\ldots,n$, where

$$a=\frac{y_U-y_L}{Y_{(n)}-Y_{(1)}}\ \text{ and }\ b=\frac{y_LY_{(n)}-y_UY_{(1)}}{Y_{(n)}-Y_{(1)}}.$$

Then the rescaled data W_m's take values in $[y_L,y_U]$. Setting $0<y_L<y_U<1$, we find the MILE $\hat{f}_0$ for the density function f_0 of $aX+b$ based on $W_m=(aX_m+b)+aZ_m,\ m=1,\ldots,n$; see the Appendix for the computation of $\hat{f}_0$. By the change-of-variable, the MILE $\hat{f}$ of f_X is given by

$$\hat{f}(x)=\begin{cases}\frac{1}{a}\hat{f}_0\left(\frac{x-b}{a}\right) & \text{if } x\in[x_L,x_U]\\ 0 & \text{otherwise,}\end{cases}$$

where $x_L=(y_UY_{(1)}-y_LY_{(n)})/(y_U-y_L)$ and $x_U=((1-y_L)Y_{(n)}+(y_U-1)Y_{(1)})/(y_U-y_L)$. Secondly we choose $y_L=0$ when X is known to be positive, i.e. $L=0$ and $U=\infty$. In our simulation, $y_L=0.1$ and $y_U=0.9$.

4. Numerical Examples

4.1 Unimodal Examples

Figures 1 and 2 contain examples involving normal, exponential, and lognormal distributions. The random variable $X=\exp(U/2)$ has a lognormal distribution if U has a standard normal distribution. As the error distribution, normal or double exponential distribution was used. The random variable $D=\lambda(E_1-E_2)$ has a double exponential distribution if E_1 and E_2 are independent exponential random variables. As in Fuller (1987) we define the reliability ratio r by

$$r = \frac{Var(X)}{Var(X) + Var(Z)} \, .$$

We report results here for density estimates based on samples of various sample size n. The subroutines in Press *et al.* (1992) were used for the generation of random numbers. The solid line indicates MILEs and the dotted line is the true density from which we generated the unobserved X_i's. In the figure we also provide kernel density estimates, which is based on a very small window. It is included only as a descriptor of the observed data $Y_1, \ldots, Y_n$.

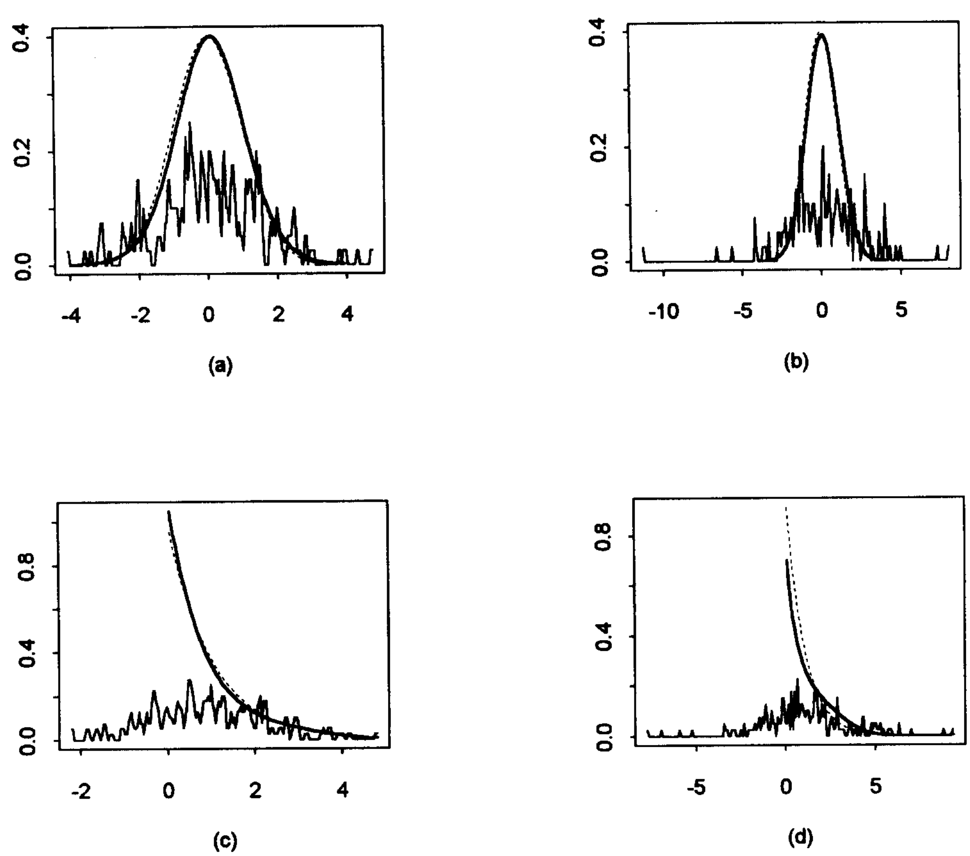

Figure 1 Normal and Exponential Density. (a) and (b) are MILEs for normal density with normal error, double exponential error based on $\boldsymbol{B} = \{\sin 2\pi x, \cos 2\pi x\}$, respectively; (c) and (d) are MILEs for exponential density with normal error, double exponential error based on $\boldsymbol{B} = \{\sin 2\pi x, \cos 2\pi x, x, x^2\}$, respectively. For each case, n=200. - Solid = MILE, dotted = Truth.

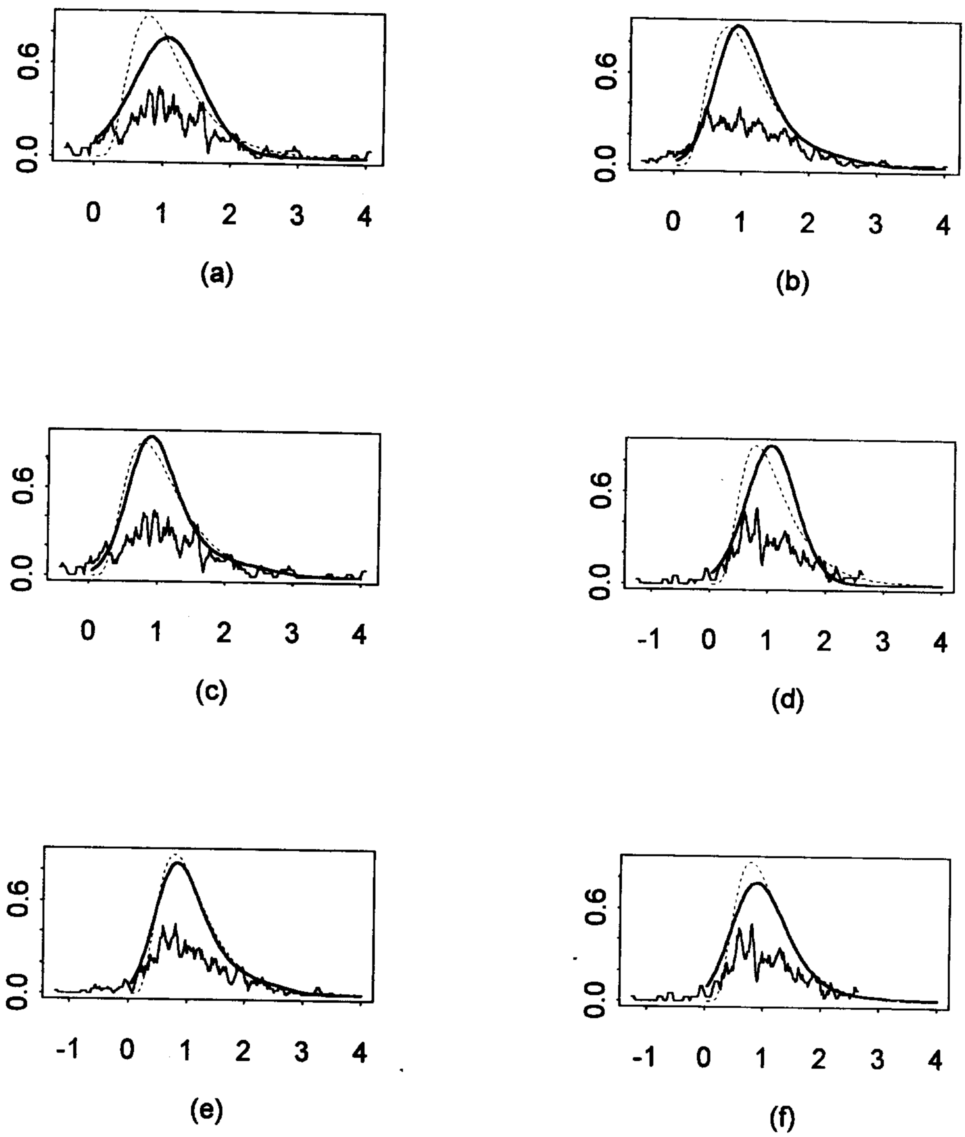

Figure 2 Lognormal Density. (a) is for normal error with $\underline{\mathbf{B}}=\{\sin 2\pi x, \cos 2\pi x\}$ and (b) and (c) are for normal error with $\underline{\mathbf{B}}=\{\sin 2\pi x, \cos 2\pi x, \sin 4\pi x, \cos 4\pi x\}$, respectively; (d) is for double exponential error with $\underline{\mathbf{B}}=\{\sin 2\pi x, \cos 2\pi x\}$ and (e) and (f) are double exponential error with $\underline{\mathbf{B}}=\{\sin 2\pi x, \cos 2\pi x, \sin 4\pi x, \cos 4\pi x\}$, respectively. Sample size: (a) 200, (b) 400, (c) 600, (d) 200, (e) 400, (f) 600. - Solid = MILE, dotted = Truth.

Suppose that X has a density $N(\mu, \sigma^2)$ and the noise Z has a density $N(0,1)$. Then the maximum likelihood estimates of μ and σ^2 are given by $\hat{\mu} = n^{-1}\sum_m Y_m$ and $\hat{\sigma}^2 = n^{-1}\sum_m (Y_m - \hat{\mu})^2 - 1$. This model was fitted by the kernel method of Liu and Taylor (1989) and the spline method of Koo and Park (1996); see Figure 1. We feel that our MILEs performed better than the estimates in Liu and Taylor (1989); the amount of smoothing near the mode of the density and the amount of smoothing in the tails seem to be correct. Our MILEs appear to be comparable with the B-spline fit. We also present in Figure 1(b) the result in which the noise distribution is double exponential with the unit variance. Figure 1(c) and 1(d) show MILEs when the target distribution is exponential. It is based on data of size $n=200$ and $r=0.5$. Since the logarithm of exponential density takes different values of derivative at tails, we have included linear and quadratic terms. One can find unbiased estimators of linear and quadratic functions as in trigonometric functions; for example, $\sum_m Y_m - EZ$ has the same expectation as $n^{-1}\sum_m X_m$.

Figures 1(c) and 2(d) show that MILEs give reasonable fits for this case.

In the case with lognormal X having normal or double exponential noise, we generated data of size $n=200$, $n=400$ and $n=600$ with $r=0.7$. As the sample size increases, MILEs perform better. One may want to compare Figure 2 with B-spline estimates in Koo and Park (1996); the performance looks similar. An advantage of MILE over the EM-algorithm of Koo and Park (1996) is the speed of computation; the whole computing time for an MILE is approximately the same as one M-step of the EM-algorithm. Comparing it with the result in Liu and Taylor (1989) for chi-squared density, we may conclude that the tail behavior of MILEs seems to be better than the kernel-type estimator of Liu and Taylor (1989).

4.2 Bimodal Examples

To investigate how MILEs behave if the true density is of a more complex nature, for example a bimodal density, we have generated X_i's from a density of the form

$$f_1(x) = 0.7N(x; 0.5, 0.5^2) + 0.3N(x; 2, 0.13^2)$$

and the density

$$f_2(x) = 0.5N(x; -(2/3)^{1/2}, 1/3) + 0.5N(x; (2/3)^{1/2}, 1/3).$$

In the case of f_1, both error distribution were used $r=0.7$. For the second case, normal measurement error with variance 1/3 was considered so that f_Y is unimodal. The bimodal density f_2 was used in Stefanski and Carroll (1990).

Figure 3 shows MILEs for samples of size $n=400$ and $n=2500$ when X has the density f_1, and Figure 4 displays MILEs when the target density is f_2. The computations and figures are organized in the same way as the examples in the previous section.

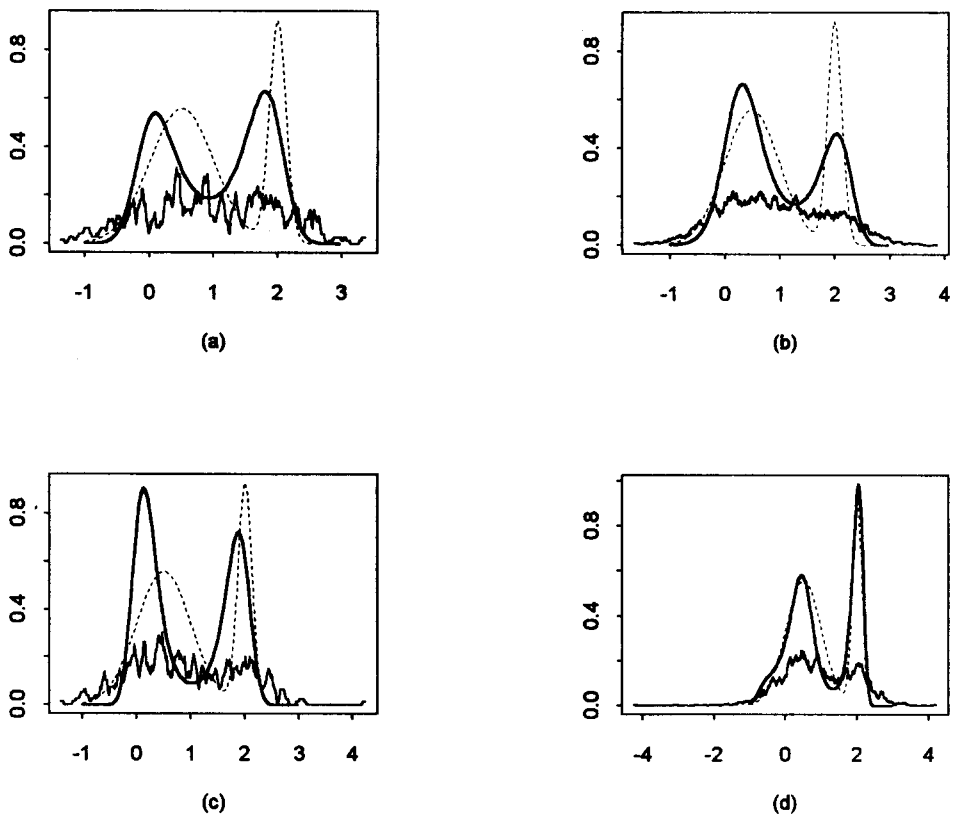

Figure 3 Bimodal Density 1. (a), (b) are MILEs for normal error with $\underline{\boldsymbol{B}}=\{\sin 2\pi x,\ \cos 2\pi x,\ \sin 4\pi x,\ \cos 4\pi x\}$ and $\underline{\boldsymbol{B}}=\{\sin 2\pi x, \cos 2\pi x, \sin 4\pi x, \cos 4\pi x, \sin 6\pi x, \cos 6\pi x\}$; (c) and (d) are MILEs for double exponential error with $\underline{\boldsymbol{B}}=\{\sin 2\pi x,\ \cos 2\pi x,\ \sin 4\pi x,\ \cos 4\pi x\}$ and $\underline{\boldsymbol{B}}=\{\sin 2\pi x,\ \cos 2\pi x,\ \sin 4\pi x,\ \cos 4\pi x,\ \sin 6\pi x,\ \cos 6\pi x, \sin 8\pi x, \cos 8\pi x\}$, respectively. Sample size: (a) 400, (b) 2500, (c) 400 and (d) 2500. - Solid = MILE, dotted = Truth.

In figure 3 are MILEs for f_1. We can observe the effect of error distribution and the size of random samples in this example. For $n=400$, MILEs for both error distributions underestimated the second sharp mode. When $n=2500$ and Z is normal, MILE underestimated the second sharp mode; whereas the second sharp mode was estimated quite well when $n=2500$ and Z was double exponential. This example seems consistent with existing theory in that the deconvolution problem is more difficult when the error is normal; see Fan (1991), Koo and Park (1996), and Koo and Chung (1997).

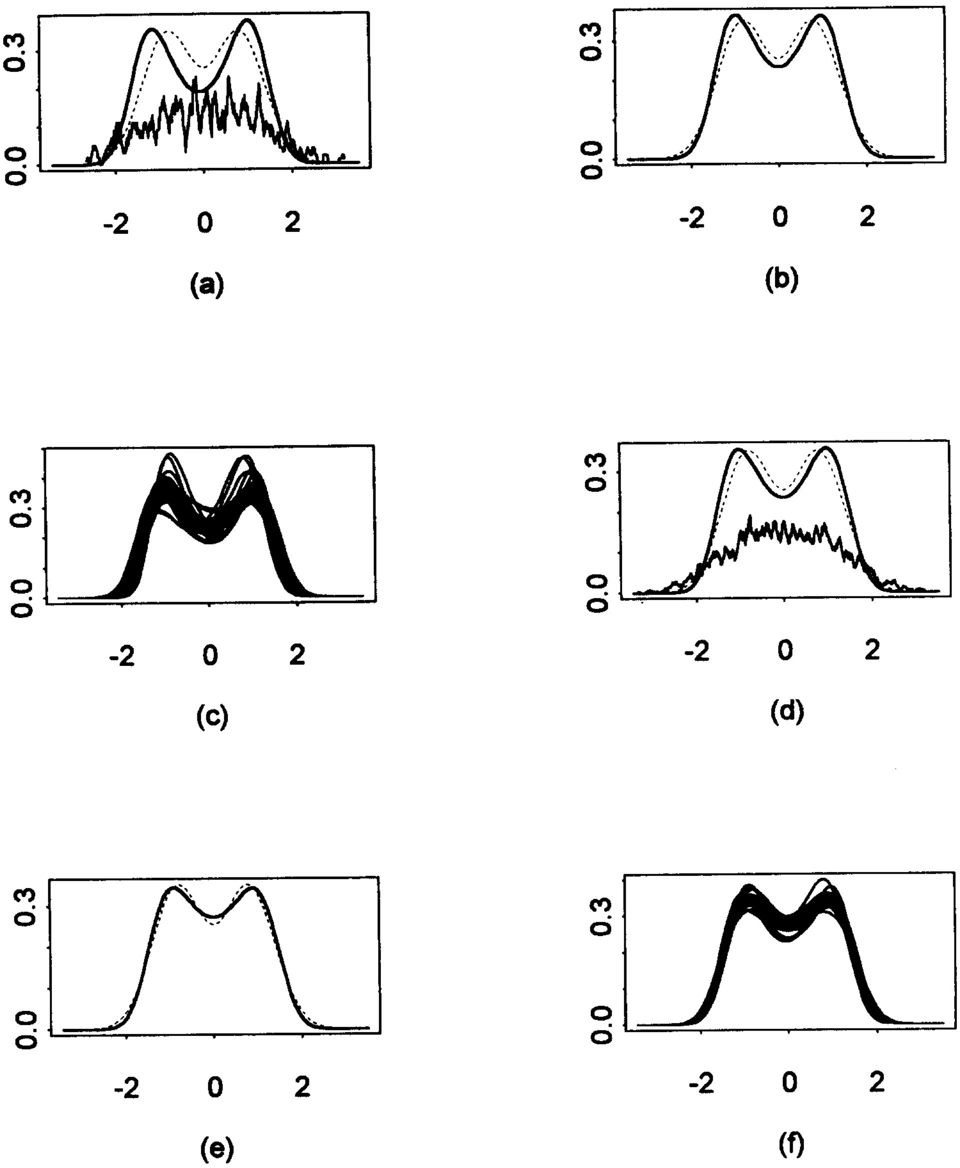

Figure 4 Bimodal Density 2. (a) and (d) denote MILEs for the Stefanski and Carroll's bimodal density with $\mathcal{B} = \{\sin 2\pi x,\ \cos 2\pi x,\ \sin 4\pi x,\ \cos 4\pi x\}$. The means of 25 repetitions are shown in (b) and (e). Overlap plots of 25 repetitions are given in (c) and (f). Sample size: 500 for (a), (b) and (c); 2500 for (d), (e) and (f). - Solid = MILE, dotted = Truth.

For the case with f_2, twenty-five repetitions were performed as in Stefanski and Carroll (1990) and Koo and Park (1996). The sample sizes in Figure 4 are $n=500$ and $n=2500$, where 500 is the smallest n among {500, 1000, 1500, 2000, 2500} for which the bimodal structure of f_2 was captured reasonably. Figure 4(a) and (d) show an estimate with sample size $n=500$, $n=2500$ respectively, Figure 4(b) and (e) are displayed the mean of the 25 MILEs respectively, and Figure 4(c) and (f) give a good idea of the variability inherent to the estimators, respectively. In Stefanski and Carroll (1990), some estimates seemed misleading: several estimates were unimodal and had negative values. However, in our 25 repetitions, we observed that our MILEs performed better in that every MILE is bimodal and the mean of MILEs is much more similar to their true density. Again MILEs appear to perform similarly to the B-spline estimates in Koo and Park (1996).

5. Concluding Remarks

An advantage of computer simulation over analytic study is that attractive and mathematically unwidely modifications can be studied. We have seen that MILEs perform reasonably well for a variety of densities. For unimodal densities, the results look impressive even for small n. When the density is of more complex nature, we need a larger sample size. The deconvolution problem with normal error is harder than the case with exponential error.

In order find an MILE, we should determine $B_1, \ldots, B_J$. It is worth extending the automatic way of choosing trigonometric functions of Koo, Lee and Park (1997) to the log-Fourier deconvolution.

References

[1] Barron, A. R. and Sheu, C. -H. (1991). Approximation of density functions by sequences of exponential families. *The Annals of Statistics*, 19, 1347-1369.

[2] de Boor, C. (1978). *A Practical Guide to Splines*. Springer-Verlag, New-York.

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

[4] Devroye, L. and Grorfi, L. (1985). *Nonparametric Density Estimation : The L_1 view*. Wiley, New York.

[5] Fan, J. (1991). On the optimal rate of convergence for nonparametric deconvolution problems. *The Annals of Statistics*, 19, 1257-1272.

[6] Fuller, W. A. (1987). *Measurement Error Models*. Wiley, New York.

[7] Koo, J. -Y. (1993). Optimal rates of convergence for nonparametric statistical inverse problems. *The Annals of Statistics*, 21, 590-599.

[8] Koo, J. -Y. (1996). Bivariate B-splines for tensor logspline density estimation. *Computational Statistics and Data Analysis*, 21, 31-42.

[9] Koo, J. -Y. and Chung, H. -Y. (1997). Log-density estimation in linear inverse problems. Being revised for *The Annals of Statistics*.

[10] Koo, J. -Y. and Kim, W. -C. (1996). Wavelet density estimation by approximation of log-densities. *Statistics and Probability Letters*, 26, 271-278.

[11] Koo, J. -Y. and Park, B. U. (1996). B-spline deconvolution based on the EM algorithm. *Journal of statistical Computation and Simulation*, 54, 275-288.

[12] Koo, J. -Y., Lee, K. -W. and Park, H. -S. (1997). Log-density estimation based on the Fourier expansion. *The Korean Journal of Applied statistics*. 10. 137-149.

[13] Kooperberg, C. and Stone, C. J. (1991). A study of logspline density estimation. *Computational Statistics and Data Analysis*, 12, 327-347.

[14] Kooperberg, C. and Stone, C. J. (1992). Logspline density estimation for censored data. *Journal of Computational and Graphical Statistics*, 1, 301-328.

[15] Liu, M. C. and Taylor, R. L. (1989). A consistent nonparametric density estimator for the deconvolution problem. *Canadian Journal of Statistics*, 17, 427-438.

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

[17] Press, W. H., Teukolsky, S. A., Vetterling, W. T. and Flannery, B. P.(1992). *Numerical Recipes in FORTRAN*. 2nd edition. Cambridge University Press, New York.

[18] Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, London.

[19] Silverman, B. W., Jones, M. C., Nychka, D. W. and Wilson, J. D. (1990). A smoothed EM algorithm to indirect estimation problems, with particular reference to stereology and emission tomography. *Journal of Royal Statistical Society. Series* B, 52, 271-324.

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

[21] Stone, C. J. (1990). Large sample inference for logspline model. *The Annals of Statistics,* 18, 717-741.

[22] Stone, C. J. and Koo, C. -Y. (1986). Logspline density estimation.. *Contemporary Mathematics*, 59, 1-15.

[23] Vardi, Y. and Lee, D. (1993). From image deblurring to optimal investments: Maximum likelihood solutions for positive linear inverse problems. *Journal of Royal Statistical Society. Series* B, 55, 569-612.

[24] Zhang, C. H. (1990). Fourier methods for estimating mixing densities and distributions. *The Annals of Statistics*, 18, 806-831.

Appendix

A-1 Lemma 1

Lemma 1 : $E\hat{C}_X(t) = C_X(t)$ and $E\hat{S}_X(t) = S_X(t)$.

Proof. Since X and Z are independent, $\phi_Y(t) = Ee^{itY} = Ee^{itX}Ee^{itZ} = \phi_X(t)\phi_Z(t)$, or equivalently

$$C_Y(t) = C_X(t)C_Z(t) + S_X(t)S_Z(t)$$
$$S_Y(t) = C_X(t)S_Z(t) - S_X(t)C_Z(t) \qquad (6)$$

If we solve the equation (6) for $C_X(t)$ and $S_X(t)$, we have

$$C_X(t) = (C_Y(t)C_Z(t) + S_Y(t)S_Z(t)) \,/\, |\phi_Z(t)|^2$$

and,

$$S_X(t) = (S_Y(t)C_Z(t) - C_Y(t)S_Z(t)) \,/\, |\phi_Z(t)|^2,$$

where $|\phi_Z(t)|^2 = C_X^2(t) + S_Z^2(t)$. Observe that $E\hat{C}_Y(t) = C_Y(t)$ and $E\hat{S}_Y(t) = S_Y(t)$. Hence,

$$E\hat{C}_X(t) = [\{E\hat{C}_Y(t)\}C_Z(t) + \{E\hat{S}_Y(t)\}S_Z(t)] \,/\, |\phi_Z(t)|^2 = C_X(t).$$

Similarly, we can show $E\hat{S}_X(t) = S_Y(t)$. This complete the proof of Lemma 1.

A-2 Computation of $\hat{f}_0$

For the fixed constants a and b determined in Section 3, let $X^* = aX + b$ and $Z^* = aX$. Then, the density f_W of $W = X^* + Z^*$ is also given by convolution equation

$$f_W(w) = \int f_{Z^*}(w - x) f_X(x) dx .$$

Based on a random sample $W_1, \ldots, W_n$ from the distribution of W, we compute $\hat{B}_j$ as before; for example, when $B_j(x) = \cos(2\pi jx)$,

$$\hat{B}_j = (\hat{C}_W(t)C_{Z^*}(t) + \hat{S}_W(t)S_{Z^*}(t))/|\phi_{Z^*}(t)|^2, \quad t = 2\pi j.$$

Let $\hat{\theta}$ be the maximizer of $l_W(\theta)$ over $\theta \in \Theta$, where $l_W(\theta)$ is defined as in (5). Then, $\hat{f}_0$ is given by $f(\cdot\,; \hat{\theta})$.

A-3 Numerical implementation

The program for log-Fourier deconvolution is written in C. The algorithm for the C program is described in the following.

Log-Fourier Deconvolution :

Choose $\underline{B}=\{B_1,\ldots,B_J\}$.

Initialize $\hat{\theta}^{(1)} \leftarrow 0$.

Compute $C(\hat{\theta}^{(1)})$ and $l_W(\hat{\theta}^{(1)})$.

Iterate for $r=1,\cdots,50$.

Compute $S(\hat{\theta}^{(r)})$ and Hessian $H(\hat{\theta}^{(r)})$,

Solve $H(\hat{\theta}^{(r)})\eta = S(\hat{\theta}^{(r)})$ for η

Find $m=\min\{k : l_W(\hat{\theta}^{(r)}+2^{-k}\eta) > l_W(\hat{\theta}^{(r)}), k=0,1,\ldots,30\}$.

If $l_W(\hat{\theta}^{(r+1)}) - l_W(\hat{\theta}^{(r)}) < 10^{-5}$, then exit iteration.

Otherwise, $\hat{\theta}^{(r+1)} \leftarrow \hat{\theta}^{(r)} + 2^{-m}\eta$.

$M \leftarrow r+1$

End iterate.

Return with $\hat{\theta} \leftarrow \hat{\theta}^{(r+1)}$.

By the change-of-variable, compute the MILE $\hat{f}$ from $\hat{f}_0 = f(\cdot\,;\hat{\theta})$.

Acknowledgements

The authors are grateful to a referee for helpful comments and suggestions.