Ann. Inst. Statist. Math.
Vol. 54, No. 3, 701–712 (2002)

DENSITY DECONVOLUTION OF DIFFERENT CONDITIONAL DISTRIBUTIONS

MARIANNA PENSKY* AND AHMED I. ZAYED

Department of Mathematics, University of Central Florida, Orlando, FL 32816, U.S.A.

(Received November 1, 2000; revised May 28, 2001)

Abstract. Recently, a new technique to circumvent the ill-posedness of the deconvolution problem has been suggested. This technique is based on what is known as multi-channel convolution system. In this paper, we modify and develop this technique in order to adapt it for statistical use. We then apply it to the problem of estimation of deconvolution density in the case of different conditional densities. This method enables us to combine equations efficiently for any set of conditional densities and to construct estimators in cases where the characteristic functions of the conditional distributions vanish at some points, as it happens in the case of uniform and triangular distributions.

Key words and phrases: Deconvolution, ill-posed problem, probability density.

1. Introduction

The convolution operation has been widely used in many statistical, mathematical, physical, and engineering applications for more than seventy five years. For example, in communication engineering the output signal of a linear filter can be represented by the convolution of the input signal and the system impulse response function. In statistics, the pdf of the sum of two independent random variables is the convolution of the pdf's of the random variables.

By a convolution here we mean the convolution operation, $*$, associated with the Fourier transform, which is defined by

$$h(x) = (f * g)(x) = \int_{-\infty}^{\infty} f(t)g(x-t)dt. \tag{1.1}$$

Define the Fourier transform of a function f as

$$\mathcal{F}[f](\omega) = \hat{f}(\omega) = \int_{-\infty}^{\infty} f(x)e^{i\omega x}dx, \tag{1.2}$$

so that the inverse transform is given by

$$\mathcal{F}^{-1}[\hat{f}](x) = f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \hat{f}(\omega)e^{-i\omega x}d\omega. \tag{1.3}$$

Then

$$\hat{h} = \widehat{f * g} = \hat{f}\hat{g}.$$

*Supported in part by National Science Foundation (NSF), grant DMS-0004173.

Equally important to convolution is deconvolution. The deconvolution problem, which arises in many applications, is the problem of reconstructing one of the convolved functions from the convolution, assuming that the other convolved function is known. In better words, find f, given that g and $f*g$ are known. Unfortunately, the deconvolution problem is ill-posed and cannot be always solved as seen from the last equation, which leads to

$$f(x)=\mathcal{F}^{-1}[\hat{h}/\hat{g}].$$

The right-hand side may not exist if $\hat{g}$ vanishes at some point.

The ill-posedness of the deconvolution can be phrased as follows: if f is a continuous function and g is a compactly supported finite Borel measure that is absolutely continuous with respect to Lebesgue measure on the real line, in particular, if g is a nonsingular probability measure with compact support, then there exists a continuous function μ, not identically zero, such that

$$\hat{g}(\omega)\hat{\mu}(\omega)=0.$$

This leads to

$$(f+\mu)*g=f*g,$$

which means that the solution of the deconvolution problem is not unique. Moreover, if we assume that $f\in L^2(\boldsymbol{R})$ and g as before, then it is known that the convolution transformation C_g

$$C_g: L^2(\boldsymbol{R})\to L^2(\boldsymbol{R})$$

defined by

$$C_g(f)=f*g$$

is a continuous injective transformation from $L^2(\boldsymbol{R})$ into the range of C_g, but its inverse, the deconvolution transformation, D_g, defined by $D_g(f*g)=f$ is discontinuous.

In 1994, Casey and Walnut proposed a new technique to circumvent the ill-posedness of the deconvolution problem by creating what they called a multidimensional channel system. Each channel in the system is represented by a convolution and the channels are grouped together in parallel so that the input (unknown) signal is fed to all the channels for processing. From the out of these channels, the input signal can be perfectly reconstructed. The underlying idea here is that perfect reconstruction of the input signal is possible because any information that is lost by one of the channels is retained by another. In other words, the system overdetermines the input signal f.

Symbolically, if f is the unknown function, q_i, $i=1,2,\dots,n$, is the system impulse response of the i-th channel, and $p_i=f*q_i$ is the output of the i-th channel, then f can be reconstructed if q_i and p_i are known, provided that the Fourier transforms, $\hat{q}_i$'s of the q_i's satisfy certain growth condition. This condition entails that the $\hat{q}_i$'s have no common zeros, which in turn implies that no information about f is lost by the system.

Functions that satisfy that condition are called coprimes. It is shown in Casey and Walnut (1994) that the indicator functions, $I_{[-r_1,r_1]}$, and $I_{[-r_2,r_2]}$ of the intervals $[-r_1,r_1]$ and $[-r_2,r_2]$ are coprimes if

$$\frac{r_1}{r_2}=\sqrt{p},\qquad p \text{ not perfect square.}$$

The work of Casey and Walnut (1994) has its roots in earlier work by Wiener (1949), Hörmander (1967), Berenstein and Yger (1989), and Berenstein and Gay (1991).

Although we shall follow a parallel path to that of Casey and Walnut (1994), we find that at some point we have to go back to the original work of Hörmander to extract some results and adjust them to our needs.

In this paper, we shall modify the work of Casey and Walnut (1994) in order to adapt it for statistical use. More precisely, we shall apply it to the problem of estimation of deconvolution density in the case of different conditional densities. Our approach, however, differs from theirs in two main points.

Firstly, the work of Casey and Walnut (1994) was focused mainly on the case where the q_i's were indicator functions of intervals and coprimes, while in this paper we deal with more general examples of coprime functions. Admittedly, we do not develop a general theory for coprime functions as Casey and Walnut (1994) did, but we focus our attention on some specific functions that are coprimes yet not necessarily indicator functions of intervals. To this end, we shall appeal to the original work of Hörmander (1967). Secondly, the work of Casey and Walnut was tailored to signal analysis applications, where the output signals $p_i = f * q_i$ are usually known. In contrast, in statistics applications, usually the functions p_i are also unknown, but fortunately can be estimated from observations. This adds more complexity to the solution of the deconvolution problem.

The problem we shall address in this paper can be formulated as follows. Suppose that n independent observations $Y_1, Y_2, \ldots, Y_n$ are available to estimate an unknown density $f(x)$ of i.i.d. random variables X_j, $j = 1, \ldots, n$, where

$$Y_j = X_j + \varepsilon_j, \qquad j = 1, \ldots, n, \tag{1.4}$$

with independent measurement errors ε_j having known probability density functions $q_j(\cdot)$. Mathematically, the problem reduces to searching for a common solution $f(x)$ of the n convolution equations

$$p_j(y) = \int_{-\infty}^{\infty} q_j(y - x) f(x) dx, \qquad j = 1, \ldots, n, \tag{1.5}$$

where $p_j(y)$ is the unknown density function of Y_j.

During the last two decades deconvolution problem has appeared in many contexts and has been studied by Carroll and Hall (1988), Devroye (1989), Diggle and Hall (1993), Fan (1991*a*,*b*,*c*,1992), Efromovich (1997), Liu and Taylor (1989), Stefanski and Carroll (1990), Taylor and Zhang (1990), Zhang (1990, 1995) among others. These authors were mainly concerned with deconvolution of Rosenblatt-Parzen type kernel estimates in the one-dimensional case, examination of MSE or MISE of these estimates, and the choice of the bandwidth parameter.

Masry (1991,1993*a*,1993*b*) constructed estimates for the stationary random process. Asymptotic normality of the estimator and asymptotic normality of its MISE were established, respectively, in the papers of Masry (1993*b*) and Piterbarg and Pensky (1993). Lower bounds for the errors of the estimators of deconvolution density were derived by Carroll and Hall (1988), Stefanski and Carroll (1990), Zhang (1990), Fan (1991*a*, 1991*b*, 1992, 1993) and Pensky and Singh (1994).

Abramovich and Silverman (1998) and Donoho (1995) constructed wavelet-based solutions of linear inverse problems treating density deconvolution as a special case. Pensky and Vidakovic (1999) and Walter (1994, 1999) studied density deconvolution based on wavelet expansions.

Yet, all these papers considered only the case when q_j are identical. However, in a variety of applications we cannot make this assumption. Let us consider, for instance, a simple example of measuring systolic blood pressure (see Carroll *et al.* (1995)). It is well known that a long-term systolic blood pressure X is an important predictor of the development of coronary heart diseases. However, during a clinical visit the immediate blood pressure Y is observed. The reason that the long-term X and the single-visit Y differ is that blood pressure has major daily, as well as seasonal variations. The other sources of error include simple machine recording error, administration error, etc. Therefore, $Y = X + \varepsilon$ where ε is the measurement error. It follows from the above discussion that it is unreasonable to assume that ε has the same distribution for various groups of patients and different types of equipment. In practice, for each chosen group we can estimate the pdf of ε by repeatedly measuring systolic blood pressure of patients in that group and then use the estimator as the true pdf of ε. Other examples where different conditional densities are required in medical and pharmaceutical research are given by Desouza (1991) and Louis (1991).

The present paper deals with the general situation when conditional densities are not identical. Pensky and Singh (1994) investigated density deconvolution with different conditional densities. However, estimators were constructed there under very restrictive assumptions that all conditional densities are uniformly bounded from below and their characteristic functions don't vanish on the real line.

The other shortcoming of the approach of Pensky and Singh (1994) is that each of the equations (1.5) was treated separately and then solutions were combined to obtain an estimator of $f(x)$. This technique does not allow one to derive estimators if the characteristic functions $\hat{q}_j$ vanish at some points as it happens when, for example, q_j are the pdf's of uniform or triangular distributions.

In what follows, we propose an approach to density deconvolution with different conditional distributions which enables one to combine equations (1.5) efficiently and construct an estimator of $f(x)$ even though $\hat{q}_j$, $j = 1, \ldots, n$, vanish at some points. As mentioned before, for this purpose we develop some ideas of Casey and Walnut (1994) and apply them in statistical environment.

The rest of the paper is organized as follows. In Section 2 we introduce the notation and terminology that will be used throughout the rest of the article and discuss some of the ideas introduced by Hörmander (1967) and Casey and Walnut (1994). In Section 3 we use some of these ideas to construct estimators of the deconvolution density. Section 4 is reserved for examples. Section 5 concludes the paper with the discussion.

2. Mathematical background

In what follows, the set of all real numbers is denoted by $\boldsymbol{R}$ and the set of all complex numbers by $\boldsymbol{C}$. Let Ω be an open set in the complex plane $\boldsymbol{C}$. If p is a nonnegative function defined on Ω, we shall denote by $Ap(\Omega)$ the set of all analytic functions $\varphi(z)$ in Ω such that for some constants C_1 and C_2

$$|\varphi(z)| \leq C_1 \exp(C_2 p(z)), \qquad z \in \Omega. \tag{2.1}$$

Let $\varphi_1, \varphi_2, \ldots, \varphi_n$ belong to $Ap(\Omega)$, so that

$$1 \leq \sum_{i=1}^{n} |\varphi_i(z)| c_1 \exp(c_2 p(z))$$

for some constants c_1, c_2. Thus,

$$d_1 \exp(-d_2 p(z)) \leq \sum_{i=1}^{n} |\varphi_i(z)|, \qquad z \in \Omega, \tag{2.2}$$

for some positive constants d_1, d_2.

Hörmander (1967) showed that condition (2.2) is both necessary and sufficient for $Ap(\Omega)$ to be finitely generated with generators $\varphi_1, \ldots, \varphi_n$. The latter implies that we can find $g_1, \ldots, g_n \in Ap(\Omega)$ such that $1 = \sum_{i=1}^{n} \varphi_i g_i$.

In the rest of this article, we will be interested in the case where $p(z) = |z|$. Since $|\mathrm{Im}\, z| \leq |z|$, and for any nonnegative integer N, we can find a and $0 < d$, such that $(1+|z|)^N \leq a e^{d|z|}$. it follows that

$$b_1 e^{-b_2|z|} \leq e^{-d|\mathrm{Im}\, z|} \left(1+|z|\right)^{-N},$$

for some appropriate constants b_1 and b_2. Moreover, since

$$\sqrt{\sum_{j=1}^{n} |\varphi_j(z)|^2} \leq \sum_{j=1}^{n} |\varphi_j(z)|,$$

we can now replace (2.2) by the stronger condition

$$a e^{-2\pi b|\mathrm{Im}\, z|} \left(1+|z|\right)^{-N} \leq \left(\sum_{j=1}^{n} |\varphi_j(z)|^2\right)^{1/2},$$

for some appropriate constants a and b.

Now let us consider the system of convolution equations (1.5) in which p_j and q_j are known. Writing the system of equations (1.5) in terms of the convolution operator, we arrive at

$$p_j = q_j * f, \qquad j = 1, \ldots, n. \tag{2.3}$$

The problem of solving the system for f is well-posed if the $\hat{q}_j(w)$ are entire functions on the complex ω-plane and

$$a e^{-2\pi b|\mathrm{Im}\, \omega|} \left(1+|\omega|\right)^{-N} \leq \left(\sum_{j=1}^{n} |\hat{q}_j(\omega)|^2\right)^{1/2}, \tag{2.4}$$

for some constants a and b and a positive integer N. For, by Hörmander's result, it follows that there exist entire functions $\hat{\nu}_j(\omega)$, $j = 1, 2, \ldots, n$, such that

$$\sum_{j=1}^{n} \hat{q}_j(\omega) \hat{\nu}_j(\omega) = 1. \tag{2.5}$$

Hence,

$$\sum_{j=1}^{n} (q_j * \nu_j)(x) = \delta(x),$$

and it follows that

$$\sum_{j=1}^{n} p_j * \nu_j = \sum_{j=1}^{n} (f * q_j) * \nu_j = f * \left(\sum_{j=1}^{n} q_j * \nu_j \right) = (f * \delta) = f. \tag{2.6}$$

It should be noted that Hörmander's result is an existence theorem, and as such it does not provide an algorithm for constructing the $\hat{\nu}_j(\omega)$'s. In fact, finding the deconvolvers ν_i explicitly is a very difficult problem, if not impossible in most cases. Yet, approximations thereof can be found. Consider nonnegative weights α_j, $j = 1, \ldots, n$. Then, as a first approximation of $\hat{\nu}_j$ one can take the function

$$\hat{\nu}_j(w) = \frac{\alpha_j \bar{\hat{q}}_j}{\sum_{i=1}^{n} \alpha_i |\hat{q}_i|^2},$$

which unfortunately is not analytic.

To devise an analytic approximation of the deconvolvers ν_j, $j = 1, \ldots, n$, consider a function $\psi(x)$ such that its Fourier transform $\hat{\psi}(\omega)$ satisfies the following conditions

C1. $\hat{\psi}(\omega)$ has a bounded support, with $\operatorname{supp} \hat{\psi} \subseteq [-A, A]$.

C2. $\hat{\psi}(0) = 1$.

C3. $\hat{\psi}(\omega)$ is s times continuously differentiable with $\hat{\psi}^{(j)}(0) = 0$ if $1 \le j \le s-1$ and $|\hat{\psi}^{(s)}(\omega)| \le C_\psi$.

Let $\psi_h(x) \equiv h^{-1}\psi(h^{-1}x)$, so that $\hat{\psi}_h(\omega) = \hat{\psi}(\omega h)$. From conditions C2 and C3 it follows that

$$|\hat{\psi}(\omega) - 1| \le C_\psi |\omega|^s. \tag{2.7}$$

Therefore, $|\hat{\psi}_h(\omega) - 1| \le C_\psi h^s |\omega|^s$ and $\hat{\psi}_h(\omega) \simeq 1$ for small h. Define

$$\hat{\nu}_{jh}(\omega) = \hat{\nu}_j(\omega)\hat{\psi}(\omega h) = \left[\sum_{i=1}^{n} \alpha_i |\hat{q}_i(\omega)|^2 \right]^{-1} \times \alpha_j \bar{\hat{q}}_j(\omega) \hat{\psi}(h\omega). \tag{2.8}$$

It is easy to see that $\nu_{jh} = \nu_j * \psi_h$ and $\sum_{j=1}^{n} \hat{\nu}_{jh}(\omega)\hat{q}_j(\omega) = \hat{\psi}(h\omega)$. Moreover, if f_h is defined by $f_h = f * \psi_h$, then

$$\hat{f}_h = \hat{f}\hat{\psi}_h = \frac{\sum_{j=1}^{n} \alpha_j \bar{\hat{q}}_j \hat{\psi}_h \hat{q}_j \hat{f}}{\sum_{j=1}^{n} \alpha_j |\hat{q}_j(\omega)|^2} = \sum_{j=1}^{n} \hat{\nu}_{jh} \hat{p}_j. \tag{2.9}$$

3. Statistical estimation

As the mathematical foundation has been developed, we apply the theory to estimation of deconvolution density. Suppose that n independent observations $Y_1, Y_2, \ldots, Y_n$ have been obtained with $Y_j \sim p_j(y)$ where $p_j(y)$ is defined by (1.5), $j = 1, \ldots, n$. Assume that q_j satisfy condition (2.4) and f belongs to Sobolev's space H_2^s, that is

$$\int_{-\infty}^{\infty} |\hat{f}(\omega)|^2 (\omega^2 + 1)^s d\omega < \infty. \tag{3.1}$$

To construct an estimator of $f(x)$ note that the function $\psi(x)$ can serve as a Parzen's kernel. Therefore, it is the standard technique to approximate f by

$$f_h(x) = h^{-1} \int_{-\infty}^{\infty} \psi(h^{-1}(y - z)) f(z) dz = f * \psi_h.$$

It follows from (2.9) that f_h can be written as

$$f_h = \sum_{j=1}^{n} p_j * \nu_{jh}.$$

Hence, $f_h(x) = \sum_{j=1}^{n} \int_{-\infty}^{\infty} \nu_{jh}(x-y)p_j(y)dy = \sum_{j=1}^{n} \boldsymbol{E}\nu_{jh}(x-Y_j)$ and the estimator of $f(x)$ has the form

$$\tilde{f}_h(x) = \sum_{j=1}^{n} \nu_{jh}(x - Y_j). \tag{3.2}$$

Let us calculate the mean integral squared error $\mathrm{MISE}(\tilde{f}_h)$ and choose the weights $\alpha_j, j = 1, \ldots, n$, minimizing it. It is well known that

$$\mathrm{MISE}(\tilde{f}_h(x)) = \boldsymbol{E}\int_{-\infty}^{\infty} (\tilde{f}_h(x) - f_h(x))^2 dx + \int_{-\infty}^{\infty} (f_h(x) - f(x))^2 dx \equiv \Delta_1 + \Delta_2. \tag{3.3}$$

Using Parseval's identity and (2.8) we can show that the variance component Δ_1 is bounded by

$$\begin{aligned} \Delta_1 &= \sum_{j=1}^{n} \int_{-\infty}^{\infty} \mathrm{Var}\, \nu_{jh}(x - Y_j)dx \\ &\le \sum_{j=1}^{n} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \nu_{jh}^2(x-y)p_j(y)dydx \\ &= (2\pi)^{-1} \sum_{j=1}^{n} \int_{-\infty}^{\infty} |\hat{\nu}_{jh}(\omega)|^2 d\omega \\ &= (2\pi h)^{-1} \sum_{k=1}^{n} \int_{-A}^{A} \left[\sum_{j=1}^{n} \alpha_j |\hat{q}_j(h^{-1}\omega)|^2\right]^{-2} \alpha_k^2 |\hat{q}_k(h^{-1}\omega)|^2 |\hat{\psi}(\omega)|^2 d\omega. \end{aligned} \tag{3.4}$$

The bias component does not depend on α_j:

$$\Delta_2 = \int_{-\infty}^{\infty} (f_h(x) - f(x))^2 dx \le (2\pi)^{-1} C_\psi h^{2s} \int_{-\infty}^{\infty} |\hat{f}(\omega)|^2 \omega^{2s} d\omega = O(h^{2s}) \tag{3.5}$$

due to (3.1), (2.7) and Parseval's identity. Therefore, to find optimal α_j, $j = 1, \ldots, n$, we need to minimize Δ_1 with respect to α_j, $j = 1, \ldots, n$. Applying Cauchy inequality to the denominator of (3.4), we obtain

$$\left[\sum_{j=1}^{n} \alpha_j |\hat{q}_j(h^{-1}\omega)|^2\right]^2 \le \sum_{j=1}^{n} \alpha_j^2 |\hat{q}_j(h^{-1}\omega)|^2 \sum_{j=1}^{n} |\hat{q}_j(h^{-1}\omega)|^2 \tag{3.6}$$

and the equality in (3.6) takes place when $\alpha_1 = \alpha_2 = \cdots = \alpha_n$. Hence, the estimator $\tilde{f}_h(x)$ has the form (3.2) with

$$\hat{\nu}_{kh}(\omega) = \left[\sum_{j=1}^{n} |\hat{q}_j(\omega)|^2\right]^{-1} \bar{\hat{q}}_k(\omega)\hat{\psi}(h\omega) \tag{3.7}$$

and

$$\text{MISE}(\tilde{f}_h) \le C_\psi h^{2s} \int_{-\infty}^{\infty} |\hat{f}(\omega)|^2 \omega^{2s} d\omega + (2\pi h)^{-1} \int_{-A}^{A} \frac{|\hat{\psi}(\omega)|^2}{\sum_{j=1}^{n} |\hat{q}_j(h^{-1}\omega)|^2} d\omega. \tag{3.8}$$

For $\text{MISE}(\tilde{f}_h)$ to converge to zero as $n \to \infty$ we need a somewhat stronger condition than (2.4). Namely, if

$$B\lambda(n)\,(1+|\omega|)^{-N} \le \left(\sum_{j=1}^{n} |\hat{q}_j(\omega)|^2 \right)^{1/2}, \tag{3.9}$$

for some positive B, N and $\lambda(n) \to \infty$ as $n \to \infty$, then

$$\begin{aligned} \text{MISE}(\tilde{f}_h) &\le C_\psi h^{2s} \int_{-\infty}^{\infty} |\hat{f}(\omega)|^2 \omega^{2s} d\omega \\ &\quad + [2\pi h \lambda^2(n)]^{-1} \int_{-A}^{A} |\hat{\psi}(\omega)|^2 \left(1+|\omega h^{-1}|\right)^{2N} d\omega \\ &= O\left(h^{2s}\right) + O\left(h^{-2N-1}\lambda^{-2}(n)\right). \end{aligned}$$

Choosing $h \sim [\lambda(n)]^{-2/(2s+2N+1)}$ we ensure that

$$\text{MISE}(\tilde{f}_h) = O([\lambda(n)]^{-4s/(2s+2N+1)}), \tag{3.10}$$

so that $\lim_{n\to\infty} \text{MISE}(\tilde{f}_h) = 0$.

4. Examples

Let us consider several examples of applications of the theory developed above.

Example 1. Uniform distributions. Let $q_j(x)$ be density functions of the uniform distributions

$$q_j(x) = (2a_j)^{-1} I(|x| \le a_j), \qquad a_j > 0,\ j = 1, \ldots, n.$$

Then $\hat{q}_j(\omega) = 2(a_j\omega)^{-1}\sin(a_j\omega)$ and $\hat{q}_j(\omega) = 0$ if $\omega = \pi k/a_j$ where k is an integer. Therefore,

$$\sum_{j=1}^{n} |\hat{q}_j(\omega)|^2 = \omega^{-2} \sum_{j=1}^{n} a_j^{-2} \sin^2(a_j\omega)$$

does not vanish if there are at least two different a_j's, say, $a_1 \ne a_2$, such that the ratio a_1/a_2 is irrational. Under this assumption,

$$\text{MISE}(\tilde{f}_h) \le C_\psi h^{2s} \int_{-\infty}^{\infty} |\hat{f}(\omega)|^2 \omega^{2s} d\omega + (2\pi)^{-1} h^{-3} C_a(n) \int_{-A}^{A} |\hat{\psi}(\omega)|^2 |\omega|^2 d\omega,$$

where

$$C_a(n) = \left[\inf_{\omega} \sum_{j=1}^{n} a_j^{-2} \sin^2(a_j\omega) \right]^{-1} > 0$$

is the constant depending on $a_1, a_2, \ldots, a_n$ and n only.

Consider, for example, the case when $a_j = a$ for k out of n cases and $a_j = b$ for the rest $(n-k)$. Assume that the ratio a/b is irrational. Denote $k/n = \varrho$ and $C_{ab} = \inf_\omega[\varrho a^{-2}\sin^2(\omega a) + (1-\varrho)b^{-2}\sin^2(\omega b)]$. Then condition (3.9) is valid with $\lambda(n) = \sqrt{n}$ and

$$\text{(4.1)} \qquad \text{MISE}(\tilde{f}_h) \le C_\psi h^{2s} \int_{-\infty}^{\infty} |\hat{f}(\omega)|^2 \omega^{2s} d\omega + (2\pi C_{ab} n)^{-1} h^{-3} \int_{-A}^{A} |\hat{\psi}(\omega)|^2 |\omega|^2 d\omega.$$

Note that since $\sin^2(\omega a)$ and $\sin^2(\omega b)$ both vanish at countable number of points, the estimators constructed for each group of conditional densities separately have infinite variances. However, according to (4.1), estimator (3.2) has

$$\text{MISE}(\tilde{f}_h) = O(n^{-2s/(2s+3)}).$$

Example 2. Triangular distributions. Assume that $q_j(x)$ are triangular distributions, i.e.

$$q_j(x) = a_j^{-1}(1 - a_j^{-1}|x|), \qquad a_j > 0, \quad j = 1, \ldots, n.$$

Then, $\hat{q}_j(\omega) = 4(a_j\omega)^{-2}\sin^2(0.5\, a_j\omega)$ and, similarly to the case of uniform distributions, $\sum_{j=1}^n |\hat{q}_j(\omega)|^2$ does not vanish if there are at least two different a_j's with irrational ratio. For instance, if ϱn of conditional densities have $a_j = a$ and $(1-\varrho)n$ of them have $a_j = b$, where $0 < \varrho < 1$ and a/b is irrational, then assumption (3.9) is valid with $\lambda(n) = \sqrt{n}$ and

$$\text{MISE}(\tilde{f}_h) \le C_\psi h^{2s} \int_{-\infty}^{\infty} |\hat{f}(\omega)|^2 \omega^{2s} d\omega + (8\pi C_{ab}^* n)^{-1} h^{-5} \int_{-A}^{A} |\hat{\psi}(\omega)|^2 |\omega|^2 d\omega$$

with $C_{ab}^* = \inf_\omega[\varrho a^{-4}\sin^4(0.5\,\omega a) + (1-\varrho)b^{-4}\sin^4(0.5\,\omega b)]$. Hence,

$$\text{MISE}(\tilde{f}_h) = O(n^{-2s/(2s+5)}).$$

Example 3. Scale-parameter family. Consider a sequence of positive integers σ_j, $j = 1, 2, 3, \ldots$ and let $q_j(x)$ be a scale parameter family

$$q_j(x) = \sigma_j q(\sigma_j x)$$

with $|\hat{q}(\omega)| \ge C_0(|\omega|^2 + 1)^{-d}$ for some positive d and C_0, so that $\hat{q}(\omega)$ does not vanish for real ω. Assume that σ_j's are bounded from above or from below.

If $\sigma_j \le \sigma_0$ for some positive σ_0, then for some $C_{\sigma 1} > 0$

$$\sum_{j=1}^n |\hat{q}_j(\omega)|^2 \ge C_{\sigma 1} \sum_{j=1}^n \sigma_j^{4d}(1 + |\omega|^2)^{-2d}$$

and condition (3.9) holds provided the series $\sum_{j=1}^\infty \sigma_j^{4d}$ is divergent. If $\sigma_j \le \sigma_0$ for some $\sigma_0 > 0$, then for some $C_{\sigma 2} > 0$

$$\sum_{j=1}^n |\hat{q}_j(\omega)|^2 \ge C_{\sigma 2} n(1 + |\omega|^2)^{-2d},$$

that is (3.9) is valid with $\lambda(n)=\sqrt{n}$.

To obtain a more precise asymptotic expression for $\mathrm{MISE}(\tilde{f}_h)$, observe that

$$\Delta_1 \le \frac{1}{2\pi C_0^2 h}\int_{-A}^{A}\frac{|\hat{\psi}(\omega)|^2}{\sum_{j=1}^{n}[\sigma_j^{4d}(h^{-2}|\omega|^2+\sigma_j^2)^{-2d}]}.$$

If we denote $\sigma_{n*}=\sup_{1\le j\le n}\sigma_j$, then from the previous inequality it follows that

$$\Delta_1 \le h^{-4d-1}\left(\sum_{j=1}^{n}\sigma_j^{4d}\right)^{-1}(2\pi)^{-1}C_0^{-2}\int_{-A}^{A}|\hat{\psi}(\omega)|^2(|\omega|^2+h^2\sigma_{n*}^2)^{2d}d\omega. \tag{4.2}$$

Thus, combining (3.3), (3.5) and (4.2), we arrive at

$$\mathrm{MISE}(\tilde{f}_h)=O\left(h^{2s}\right)+O\left(h^{-4d-1}\left[\sum_{j=1}^{n}\sigma_j^{4d}\right]^{-1}\right)+O\left(h^{-1}\sigma_{n*}^{4d}\left[\sum_{j=1}^{n}\sigma_j^{4d}\right]^{-1}\right). \tag{4.3}$$

It is interesting to examine the behavior of (4.3) in the situation when σ_j, $j=1,2,\dots$, are either increasing to infinity or decreasing to zero. We will consider the cases when $\sigma_j=\sigma_0 j^{\alpha}$ or $\sigma_j=\sigma_0 j^{-\alpha}$ with $\alpha>0$.

If $\sigma_j=\sigma_0 j^{\alpha}$, $\alpha>0$, then $\sigma_{n*}=\sigma_0 n^{\alpha}$ and $\sum_{j=1}^{n}\sigma_j^{4d}=O(n^{4\alpha d+1})$. Hence,

$$\mathrm{MISE}(\tilde{f}_h)=O(h^{2s})+O(h^{-4d-1}n^{-4\alpha d-1})+O(h^{-1}n^{-1}). \tag{4.4}$$

Formula (4.4) implies that in the case when $\alpha\ge(2s+1)^{-1}$ choosing $h=O(n^{-1/(2s+1)})$ we arrive at

$$\mathrm{MISE}(\tilde{f}_h)=O(n^{-2s/(2s+1)}).$$

This means that for $\alpha\ge(2s+1)^{-1}$ the presence of errors $\varepsilon_1,\dots,\varepsilon_n$ does not affect convergence rate of the estimator and we can construct the estimators with the same precision as if the observations $X_1,X_2,\dots,X_n$ were available. Nevertheless, if $\alpha<(2s+1)^{-1}$, the estimator converges slower than the estimator based on original observations, that is

$$\mathrm{MISE}(\tilde{f}_h)=O(n^{-2s(4\alpha d+1)/(2s+4d+1)}).$$

If $\sigma_j=\sigma_0 j^{-\alpha}$, $\alpha>0$, then $\sigma_{n*}=\sigma_1$ and $\sum_{j=1}^{n}j^{-4\alpha d}=O(n^{1-4\alpha d})$. Therefore,

$$\mathrm{MISE}(\tilde{f}_h)=O(h^{2s})+O(h^{-4d-1}n^{4\alpha d-1}). \tag{4.5}$$

It follows from (4.5) that $\lim_{n\to\infty}\mathrm{MISE}(\tilde{f}_h)\ne 0$ if $\alpha\ge d/4$, so that the estimator (3.2) is not mean square consistent in this case. However, if $\alpha<d/4$, we obtain $\sum_{j=1}^{n}j^{-4\alpha d}=O(n^{1-4\alpha d})$ and

$$\mathrm{MISE}(\tilde{f}_h)=O(n^{-2s(1-4\alpha d)/(2s+4d+1)}).$$

Note that the asymptotic expressions (4.4) and (4.5) are exact in the sense that if $\hat{q}(\omega)\le C_0(\omega^2+1)^{-d}$, then there exist absolute constants C_1 and C_2 such that $\mathrm{MISE}(\tilde{f}_h)\ge C_1(h^{2s}+h^{-4d-1}n^{-4\alpha d-1}+h^{-1}n^{-1})$ if $\sigma_j=\sigma_0 j^{\alpha}$, $\alpha>0$, and $\mathrm{MISE}(\tilde{f}_h)=C_2(h^{2s}+h^{-4d-1}n^{4\alpha d-1})$ if $\sigma_j=\sigma_0 j^{-\alpha}$, $\alpha>0$. This means that the analysis conducted above is adequate.

5. Discussion

In the present paper, we modified the work of Casey and Walnut (1994) in order to adapt it for statistical use. For this purpose, we appealed to the original work of Hörmander (1967) and developed the technique which allows to construct a common solution of convolution equations in a wider variety of situations than those on which Casey and Walnut (1994) concentrated.

This method is applied to the problem of estimation of deconvolution density in the case of different conditional densities. Although this problem has been studied earlier by Pensky and Singh (1994), the estimators were constructed under very restrictive assumptions that all conditional densities are uniformly bounded from above and their characteristic functions don't vanish on the real line. The other shortcoming of the approach of Pensky and Singh (1994) is that each of the equations (1.5) was treated separately and then solutions were combined to obtain an estimator of $f(x)$. This technique does not allow one to derive estimators if characteristic functions $\hat{q}_j$ vanish at some points as it happens when, for example, q_j are the pdf's of uniform or triangular distributions.

In the present paper we overcame both defects of the method of Pensky and Singh (1994) and proposed the method which allows one to construct the estimators even if all conditional characteristic functions vanish at some points. Moreover, this method allows one to combine equations efficiently and derive coefficients α_j, $j = 1, \ldots, n$, that are optimal for any set of conditional densities q_j. The MISE of the estimator is calculated and, under fairly non-restrictive condition, is shown to turn to zero as $n \to \infty$ for the appropriate choice of h.

The theory is illustrated by examples when the family of conditional distributions is uniform or triangular, or a scale parameter family. In the case of the scale-parameter family of conditional distributions, we consider the wide spectrum of situations including the ones when the scale parameters are increasing to infinity or decreasing to zero. The estimators are constructed and their mean integrates square errors are analyzed.

References

Abramovich, F. and Silverman, B. W. (1998). Wavelet decomposition approaches to statistical inverse problems, *Biometrika*, **85**, 115–129.

Berenstein, C. A. and Gay, R. (1991). *Complex Analysis: An Introduction*, Springer, New York.

Berenstein, C. A. and Yger, A. (1989). Analytic Bezout identities, *Adv. Math.*, **10**, 51–74.

Carroll, R. J. and Hall, P. (1988). Optimal rates of convergence for deconvolving a density, *J. Amer. Statist. Assoc.*, **83**, 1184–1186.

Carroll, R. J., Ruppert, D. and Stefanski, L. A. (1995). *Measurement Error in Nonlinear Models*, Chapman & Hall, London.

Casey, S. D. and Walnut, D. F. (1994). Systems of convolution equations, deconvolution, Shannon sampling, and the wavelet and Gabor transforms, *SIAM Rev.*, **36**, 537–577.

Desouza, C. M. (1991). An empirical Bayes formulation of cohort models in cancer epidemiology, *Statistics in Medicine*, **10**, 1241–1256.

Devroye, L. (1989). Consistent deconvolution in density estimation, *Canad. J. Statist.*, **17**, 235–239.

Diggle, P. J. and Hall, P. (1993). A Fourier approach to nonparametric deconvolution of a density estimate, *J. Roy. Statist. Soc. Ser. B*, **55**, 523–531.

Donoho, D. (1995). Nonlinear solution of linear inverse problems by wavelet-vaguelette decomposition, *Appl. Comput. Harmon. Anal.*, **2**, 101–126.

Efromovich, S. (1997). Density estimation for the case of supersmooth measurement error, *J. Amer. Statist. Assoc.*, **92**, 526–535.

Erdélyi, A. (1954). *Tables of Integral Transforms*, McGraw-Hill, New York.

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

Fan, J. (1991*b*). Asymptotic normality for deconvolution kernel density estimators, *Sankhyā Ser. A*, **53**, 97–110.

Fan, J. (1991*c*). Global behavior of deconvolution kernel estimates, *Statist. Sinica*, **1**, 541–551.

Fan, J. (1992). Deconvolution with supersmooth distributions, *Canad. J. Statist.*, **20**, 155–169.

Fan, J. (1993). Adaptively local one-dimensional subproblems with application to a deconvolution problem, *Ann. Statist.*, **21**, 600–610.

Hörmander, L. (1967). Generators for some rings of analytic functions, *Bull. Amer. Math. Soc.*, **73**, 943–949.

Liu, M. C. and Taylor, R. L. (1989). A consistent nonparametric density estimator for the deconvolution problem, *Canad. J. Statist.*, **17**, 427–438.

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

Masry, E. (1991). Multivariate probability density deconvolution for stationary random processes, *IEEE Trans. Inform. Theory*, **37**, 1105–1115.

Masry, E. (1993*a*). Strong consistency and rates for deconvolution of multivariate densities of stationary processes, *Stochastic Process. Appl.*, **47**, 53–74.

Masry, E. (1993*b*). Asymptotic normality for deconvolution estimators of multivariate densities of stationary processes, *J. Multivariate Anal.*, **44**, 47–68.

Pensky, M. and Singh, R. S. (1994). Optimal estimation of prior densities of multidimensional location and scale parameters, Tech. Report, Statistical Series 1994-275, Department of Mathematics and Statistics, University of Guelph, Canada.

Pensky, M. and Vidakovic, B. (1999). Adaptive wavelet estimator for nonparametric density deconvolution, *Ann. Statist.*, **27**, 2033–2053.

Piterbarg, V. and Pensky, M. (1993). On asymptotic distribution of integrated squared error of an estimate of a component of a convolution, *Math. Methods Statist.*, **2**, 30–41.

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

Taylor, R. L. and Zhang, H. M. (1990). On a strongly consistent non-parametric density estimator for deconvolution problem, *Comm. Statist. Theory Methods*, **19**, 3325–3342.

Walter, G. G. (1994). *Wavelets and Other Orthogonal Systems with Applications*, CRC Press, Boca Raton.

Walter, G. G. (1999). Density estimation in the presence of noise, *Statist. Probab. Lett.*, **41**, 237–246.

Wiener, N. (1949). *Extrapolation, Interpolation and Smoothing of Stationary Time Series*, MIT Technology Press, Cambridge, Massachusetts.

Zhang, C. H. (1990). Fourier methods for estimating mixing densities and distributions, *Ann. Statist.*, **18**, 806–831.

Zhang, C. H. (1995). On estimating mixing densities in discrete exponential family models, *Ann. Statist.*, **23**, 929–945.