Indian J. Pure Appl. Math., **41**(1): 275-291, February 2010

WAVELET LINEAR ESTIMATION FOR DERIVATIVES OF A DENSITY FROM OBSERVATIONS OF MIXTURES WITH VARYING MIXING PROPORTIONS

B. L. S. Prakasa Rao

Department of Mathematics and Statistics, University of Hyderabad,
Hyderabad 500 046, India
e-mail: blsprao@gmail.com

Abstract A wavelet based linear estimator is proposed for the derivatives of a probability density function based on a sample from a finite mixture of components with varying mixing proportions. It extends the linear estimator of a probability density function proposed by Pokhyl'ko (*Theor. Probability and Math. Statist*, **70** (2005) 135-145). Upper bounds on L_2 and L_∞ losses are obtained for such estimators.

Key words Estimation of derivatives of a density function, wavelets, mixtures of components, varying mixing proportions.

1. Introduction

The problem of analysis of mixtures with varying mixing proportions occur in the study of medical, biological, social and other types of data. The objects of observation $J_1, \ldots, J_N$ may belong to any one of M populations. Let $I(J_j)$ denote the indicator of the population that contains the object J_j. For every object J_j, we observe a random variable X_j based on the object J_j. Note that the distribution function of the random variable X_j depends on the indicator $I(J_j)$. Suppose that

$$P(X_j \leq x | I(J_j) = k) = H_k(x), 1 \leq j \leq N, 1 \leq k \leq M.$$

Suppose the distribution functions $H_k(x), 1 \leq k \leq M$ are unknown and the sequence $I(J_j), 1 \leq j \leq N$ are also unknown but we know the probability $w_k(j)$ that an object J_j belongs to the k-th population, that is,

$$w_k(j) = P(I(J_j) = k), 1 \leq j \leq N, 1 \leq k \leq M.$$

Note that $w_k(j) \geq 0, 1 \leq k \leq M$ and $\sum_{k=1}^{M} w_k(j) = 1, 1 \leq j \leq N.$

Observe that the probability $w_k(j)$ indicates the mixing proportion of the objects of k-th population in the mixture from which the object J_j is chosen. It is easy to check that

$$P(X_j \leq x) = \sum_{\ell=1}^{M} w_\ell(j) H_\ell(x), 1 \leq j \leq N. \tag{1.1}$$

The problem of estimation of the distribution function $H_\ell(x)$ was studied in Maiboroda [10] using a weighted empirical distribution function. Maiboroda [11] obtained a generalized version of the Kolmogorov-Smirnov test for testimg the hypothesis for the homogenity of mixtures with varying mixing proportions. Assuming that the distribution functions $H_\ell(x)$ are absolutely continuous with density functions $h_\ell(x),$ Pokhyl'ko [14] constructed linear and adaptive wavelet estimators for the density function $h_\ell(x).$

Methods of nonparametric estimation of a density function and regression function are widely discussed in the literature (cf. Prakasa Rao [15, 17]). It is known that the estimation of derivatives of a density as well as that of regression function are also of importance and interest to detect possible bumps in the case of a density and to detect concavity or convexity properties in the case of regression function. Asymptotic properties of the kernel type estimators for the derivatives of density have been investigated earlier (cf. Prakasa Rao [15], p.237).

Our aim in this paper is to discuss wavelet linear estimators for the derivatives of a probability density function when the sample of observations come from a mixture of several components with varying mixing proportions. We propose an estimator for the derivative of the density based on wavelets and obtain upper bounds on the L_2 and L_∞ losses for the proposed estimator. Estimators of density using wavelets was studied for independent and identically distributed random variables in Antoniadis *et al.* [1], for some stationary dependent random variables in Leblanc [9] and for stationary associated sequnces in Prakasa Rao [19]. Chaubey *et al.* [2, 3] extended these results to derivatives of density estimators for associated sequences and for negatively associated processes.

The advantages and disadvantages of the use of wavelet based probability density estimators are discussed in Walter and Ghorai [24] in the case of independent and identically distributed observations. The same comments continue to hold in this case. However it was shown in Prakasa Rao [16, 18] that one can obtain precise limits on the asymptotic mean squared error for a wavelet based linear estimator for the density function and its derivatives as well as some other functionals of the density. Tribouley [21] studied estimation of multivariate densities using wavelet methods. Donoho *et al.* [6] investigated density estimation by wavelet thresholding. For a discussion on statistical modeling by wavelets, see Vidakovic [23].

2. Preliminaries on Wavelets

A wavelet system is an infinite collection of translated and scaled versions of functions $\phi(\cdot)$ and $\psi(\cdot)$ called the *scaling function* and the *primary wavelet function* respectively. In the following discussion, we assume that $\phi(\cdot)$ is real-valued. The function $\phi(x)$ is a solution of the equation

$$\phi(x) = \sum_{k=-\infty}^{\infty} C_k \phi(2x-k), \tag{2.1}$$

with

$$\int_{-\infty}^{\infty} \phi(x)dx = 1, \tag{2.2}$$

and the function $\psi(x)$ is defined by

$$\psi(x) = \sum_{k=-\infty}^{\infty} (-1)^k C_{-k+1} \phi(2x-k). \tag{2.3}$$

The choice of the sequence $\{C_k\}$ determines the wavelet system. It is easy to see that

$$\sum_{k=-\infty}^{\infty} C_k = 2. \tag{2.4}$$

Define

$$\phi_{jk}(x) = 2^{j/2}\phi(2^j x - k), -\infty < j, k < \infty \tag{2.5}$$

and

$$\psi_{jk}(x) = 2^{j/2}\psi(2^j x - k), -\infty < j, k < \infty. \tag{2.6}$$

Suppose the coefficients $\{C_k\}$ satisfy the condition

$$\begin{aligned} \sum_{k=-\infty}^{\infty} C_k C_{k+2\ell} &= 2 \text{ if } \ell = 0 \\ &= 0 \text{ if } \ell \neq 0. \end{aligned} \tag{2.7}$$

It is known that, under some additional conditions on $\phi(\cdot)$, the collection $\{\psi_{j,k}, -\infty < j, k < \infty\}$ is an orthonormal basis for $L^2(R)$, and $\{\phi_{j,k}, -\infty < k < \infty\}$ is an orthonormal system in $L^2(R)$, for each $-\infty < j < \infty$ (cf. Daubechies [4]).

Definition 2.1. The scaling function ϕ is said to be r-regular for an integer $r \geq 1$, if for every nonnegative integer $\ell \leq r$, and for any integer $k \geq 1$,

$$|\phi^{(\ell)}(x)| \leq c_k(1+|x|)^{-k}, -\infty < x < \infty \tag{2.8}$$

for some $c_k \geq 0$ depending only on k. Here $\phi^{(\ell)}(\cdot)$ denotes the ℓ-th derivative of $\phi(\cdot)$.

Definition 2.2. A *multiresolution analysis* of $L^2(R)$ consists of an increasing sequence of closed subspaces $\{V_j\}$ of $L^2(R)$ such that

(i) $\cap_{j=-\infty}^{\infty} V_j = \{0\}$;

(ii) $\bar{\cup}_{j=-\infty}^{\infty} V_j = L^2(R)$;

(iii) there is a scaling function $\phi \in V_0$ such that $\{\phi(x-k), -\infty < k < \infty\}$ is an orthonormal basis for V_0;

(iv) for all $h(\cdot) \in L^2(R), -\infty < k < \infty, h(x) \in V_0 \Rightarrow h(x-k) \in V_0$; and

(v) $h(\cdot) \in V_j \Rightarrow h(2x) \in V_{j+1}$.

Mallat [12] has shown that, given any multiresolution analysis, it is possible to find a function $\psi(\cdot)$ (called the primary wavelet function) such that, for any fixed $j, -\infty < j < \infty$, the family $\{\psi_{j,k}, -\infty < k < \infty\}$ is an orthonormal basis of the orthogonal complement W_j of V_j in V_{j+1} so that $\{\psi_{j,k}, -\infty < j, k < \infty\}$ is an orthonormal basis of $L^2(R)$ (cf. Daubechies [4]). When the scaling function $\phi(\cdot)$ is r-regular, the corresponding multiresolution analysis is said to be r-regular.

Let $f \in L_2(R)$. The function f can be expanded in the form (cf. Daubechies [5]):

$$\begin{aligned} f &= \sum_{k=-\infty}^{\infty} a_{s,k}\phi_{s,k} + \sum_{j=s}^{\infty}\sum_{k=-\infty}^{\infty} b_{j,k}\psi_{j,k} \qquad (2.9) \\ &= P_s f + \sum_{j=s}^{\infty} D_j f \end{aligned}$$

for any integer $-\infty < s < \infty$. Observe that the wavelet coefficients are given by

$$a_{s,k} = \int_{-\infty}^{\infty} f(x)\phi_{s,k}(x)dx \qquad (2.10)$$

and

$$b_{j,k} = \int_{-\infty}^{\infty} f(x)\psi_{j,k}(x)dx. \qquad (2.11)$$

Suppose that the functions ϕ and ψ belong to C^r, the space of functions with r continuous derivatives for some $r \geq 1$, and have compact support contained in an interval $[-\delta, \delta]$ for some $\delta > 0$. It follows, from the Corollary 5.5.2 in Daubechies [4], that the function $\psi(\cdot)$ is orthogonal to polynomials of degree less than or equal to r. In particular

$$\int_{-\infty}^{\infty} \psi(x)x^{\ell}dx = 0, \ell = 0, 1, \ldots, r.$$

The above introduction to wavelets is based on Antoniades et al. [1]. For a detailed discussion, see Daubechies [5]. For a brief survey on wavelets, see Strang [20].

3. Introduction to Sobolev Spaces and Besov Spaces

Let f be a function defined on the real line which is integrable on every bounded interval. It is said to be weakly differentiable if there exists a function g defined on the real line which is integrable on every bounded interval such that

$$\int_x^y g(u)du = f(y) - f(x).$$

The function g is defined almost everywhere and is called the *weak derivative* of f (cf. Hardle et al. [8]). It is known that, if f is weakly differentiable with weak derivative g, then

$$\int_{-\infty}^{\infty} f(u)\phi'(u)du = -\int_{-\infty}^{\infty} g(u)\phi(u)du$$

for any $\phi \in D(R)$ where $D(R)$ denotes the space of infinitely differentiable functions, on the real line R, with compact support.

Definition 3.1. Let $1 \leq p \leq \infty$ and $m \geq 0$ be an integer. A function $f \in L_p(R)$ belongs to the Sobolev space $W_p^m(R)$, if it is m-times weakly differentiable and $f^{(m)} \in L_p(R)$. In particular $W_p^0(R) = L_p(R)$. The space $W_p^m(R)$ is equipped with the norm

$$||f||_{W_p^m} = ||f||_p + ||f^{(m)}||_p$$

where $||f||_p$ denotes the norm for $L_p(R)$.

Let $\tilde{W}_p^m(R) = W_p^m(R)$ if $1 \leq p < \infty$ and $\tilde{W}_\infty^m(R) = \{f : f \in W_\infty^m(R) : f^{(m)}$ uniformly continuous $\}$. Note that $\tilde{W}_p^0(R) = L_p(R), 1 \leq p < \infty$.

Let $f \in L_p(R)$ for some $1 \leq p \leq \infty$. Let $(\Delta_h f)(x) = f(x-h) - f(x)$ and define $\Delta_h^2 f = \Delta_h \Delta_h f$. For $t \geq 0$, define

$$\omega_p^1(f,t) = \sup_{|h|\leq t} ||\Delta_h f||_p$$

and

$$\omega_p^2(f,t) = \sup_{|h|\leq t} ||\Delta_h^2 f||_p.$$

Let $1 \leq q \leq \infty$. Suppose there exists a function $\epsilon(t)$ on $[0,\infty)$ such that $||\epsilon||_q^* < \infty$ where

$$\begin{aligned} ||\epsilon||_q^* &= (\int_0^\infty t^{-1}|\epsilon(t)|^q dt)^{1/q}, \text{ if } 1 \leq q < \infty \\ &= ess\sup_t |\epsilon(t)|, \text{ if } q = \infty. \end{aligned} \tag{3.1}$$

Definition 3.2. Let $1 \leq p, q \leq \infty$ and $s = n + \alpha$ where $n \geq 0$ is an integer and $0 < \alpha \leq 1$. The *Besov space* $B_{p,q}^s$ is the space of all functions f such that $f \in W_p^n(R)$ and $\omega_p^2(f^{(n)},t) = \epsilon(t)t^\alpha$ where $||\epsilon||_q^* < \infty$.

For properties of Besov spaces, see Meyer [13] and Triebel [22] (cf. Leblanc [9], Hardle et al. [8]).

Suppose that the function f belongs to the Besov class

$$F_{s,p,q}(L) = \{f \in B^s_{p,q}, ||f||_{B^s_{p,q}} \leq L\}$$

for some $0 < s < r+1, p \geq 1$ and $q \geq 1$, where

$$||f||_{B^s_{p,q}} = ||P_0 f||_p + [\sum_{j\geq 0}(||D_j f||_p 2^{js})^q]^{1/q}.$$

Given a double indexed sequence $\{\gamma_{j,k}\}$ define the norm

$$||\gamma_{j,.}||_{\ell_p} = (\sum_k \gamma^p_{j,k})^{1/p}. \tag{3.2}$$

In view of the representation (2.9), it be can shown that the function $f \in B^s_{p,q}$ if and only if

$$||a_{s,.}||_{\ell_p} < \infty, \text{ and } (\sum_{j\geq s}[||b_{s,.}||_{\ell_p} 2^{j(s+(1/2)-(1/p))}]^q)^{1/q} < \infty. \tag{3.3}$$

Let $\phi(\cdot)$ be a scaling function as defined earlier. Define

$$\theta_\phi(x) = \sum_{k=-\infty}^{\infty} |\phi(x-k)|.$$

Suppose the following conditions hold:

(C1) The $ess\sup_x \theta_\phi(x) < \infty$ where

$$ess\sup_x g(x) = \inf\{y : \lambda([x : g(x) > y]) = 0\}$$

and λ is the Lebesgue measure on the real line.

(C2) There exists a bounded nondecreasing function $\Phi(\cdot)$ such that $|\phi(u)| \leq \Phi(|u|)$ almost every where and

$$\int_0^\infty |u|^r \Phi(|u|) du < \infty.$$

for some integer $r \geq 0$.

Lemma 3.1. *Suppose that the scale function $\phi(\cdot)$ is such that the collection $\{\phi(x - k), -\infty < k < \infty\}$ is an orthonormal system in $L_2(R)$ and the spaces $V_j, -\infty < j < \infty$ are nested. Further suppose that the function ϕ satisfies the condition (C2) and it is $r+1$ times weakly differentiable. If $\phi^{(r+1)}$ satisfies the condition (C1), then the norm $||.||_{B^s_{p,q}}$ is equivalent to the norm $||.||'_{B^s_{p,q}}$ in the space of the wavelet coefficients for all s, p, q such that $0 < s < r+1$ and $1 \leq p, q \leq \infty$ where*

$$||f||'_{B^s_{p,q}} = ||a_0||_p + (\sum_{j=0}^{\infty}(2^{j(s+(1/2)-(1/p))}||b_j||_p)^q)^{1/q}.$$

(*Here* $||a_0||_p$ *denotes* $[\sum_{k=-\infty}^{\infty} |a_{0,k}|^p]^{1/p}$ *and* $||b_j||_p$ *denotes* $[\sum_{k=-\infty}^{\infty} |b_{j,k}|^p]^{1/p}$).

For a proof of Lemma 3.1, see Theorem 9.6 in Hardle *et al.* [8], p.123.

4. Estimation of the d-th Derivative of a Probability Density Function

Let $\{Y_i, 1 \leq i \leq n\}$ be independent and identically distributed random variables with probability density function f which is d-times differentiable. Suppose that $f^{(d)}$ is bounded and has compact support. Suppose that $f^{(d)} \in L_2(R)$. Let us first consider the case $d = 0$. The problem now is the estimation of the probability density function f. A wavelet based density estimator of the density function f can be motivated in the following way from the expansion given in (2.9) (cf. Prakasa Rao [19]). We can estimate $f(x)$ by $\hat{f}(x)$ where

$$\hat{f}(x) = \sum_{k \in N_s} \alpha_{s,k} \phi_{s,k}(x), \tag{4.1}$$

where

$$\alpha_{s,k} = \frac{1}{n} \sum_{i=1}^{n} \phi_{s,k}(Y_i). \tag{4.2}$$

Here N_s is the set of integers k such that $supp(f) \cap supp(\phi_{s,k})$ is nonempty. Since the functions f and ϕ have compact supports, the cardinality of the set N_s is finite and it is of the order $O(2^s)$.

Let us now consider the problem of estimation of the derivative $f^{(d)}$ of f. As in Prakasa Rao [16], we assume that the scaling function $\phi(\cdot)$ generates a r-regular multiresolution analysis for some $r \geq (d+1)$ and that there exists $C_m \geq 0$ and $\beta_m \geq 0$ such that

$$|f^{(m)}(x)| \leq C_m(1+|x|)^{-\beta_m}, \ \ 0 \leq m \leq r. \tag{4.3}$$

This assumption implies that that the derivative $\phi^{(d)}$ is bounded for every $d \geq 0$ (cf. Prakasa Rao [16]). Furthermore the projection of $f^{(d)}$ on V_s is

$$f_s^{(d)}(x) = \sum_{k \in N_s} a_{s,k} \phi_{s,k}(x), \tag{4.4}$$

where

$$a_{s,k} = (-1)^d \int_{-\infty}^{\infty} f(x) \phi_{s,k}^{(d)}(x) dx.$$

The equation given above can be justified by using integration by parts since the function $\phi(\cdot)$ is r-regular (cf. Prakasa Rao [16]). This expression motivates the following estimator for $f^{(d)}(x)$:

$$\hat{f}_s^{(d)}(x) = \sum_{k \in N_s} \hat{a}_{s,k} \phi_{s,k}(x), \tag{4.5}$$

where

$$\hat{a}_{s,k} = \frac{(-1)^d}{n} \sum_{i=1}^{n} \phi_{s,k}^{(d)}(Y_i).$$

Note that the estimator defined above reduces to the density estimator given in (4.1) for $d = 0$. We now rewrite the expression for the estimator $\hat{f}_s^{(d)}(x)$ in a slightly different form.

Note that

$$\begin{aligned}
\hat{f}_s^{(d)}(x) &= \sum_{k \in N_s} \hat{a}_{s,k}\phi_{s,k}(x) \\
&= \sum_{k \in N_s} [\frac{(-1)^d}{n} \sum_{i=1}^{n} \phi_{s,k}^{(d)}(Y_i)]\phi_{s,k}(x) \\
&= \frac{(-1)^d}{n} \sum_{i=1}^{n} \sum_{k \in N_s} \phi_{s,k}^{(d)}(Y_i)\phi_{s,k}(x) \\
&= \frac{(-1)^d}{n} \sum_{i=1}^{n} \sum_{k \in N_s} 2^{(s/2)+ds}\phi^{(d)}(2^s Y_i - k)2^{s/2}\phi(2^s x - k) \\
&= \frac{(-1)^d}{n} \sum_{i=1}^{n} [\sum_{k \in N_s} \phi^{(d)}(2^s Y_i - k)\phi(2^s x - k)]2^{s+ds} \\
&= \frac{(-1)^d}{n} \sum_{i=1}^{n} K^{(d)}(2^s Y_i, 2^s x)2^{s+ds} \\
&= \frac{(-1)^d}{n} \sum_{i=1}^{n} K_s^{(d)}(Y_i, x), \qquad (4.6)
\end{aligned}$$

where

$$K_s(x,y) = 2^s \ K(2^s x, 2^s y)$$

and

$$K(x,y) = \sum_{k \in N_s} \phi(x-k)\phi(y-k).$$

Here $K_s^{(d)}(x,y)$ denotes the d-th partial derivative of $K_s(x,y)$ with respect to x.

5. Estimation of the d-th Derivative of a Component Probability Density Function from a Mixture with Varying Mixing Proportions

Let $\{X_i, 1 \le i \le N\}$ be random variables as described Section 1. The problem is to estimate the d-th derivative of the component probability density function $h_\ell(x)$ corresponding to the mixing proportion $w_\ell(\cdot)$ based on the observations $X_1, \ldots, X_N$. Note that the probability density function of X_j is given by

$$p_j(x) = \sum_{\ell=1}^{M} w_\ell(j)h_\ell(x)$$

for $1 \leq j \leq N$. For $\mathbf{x}, \mathbf{y} \in R^N$, define the inner product $< \mathbf{x}, \mathbf{y} >_N$ by the relation

$$< \mathbf{x}, \mathbf{y} >_N = \frac{1}{N} \sum_{k=1}^{N} x_k y_k,$$

whenever $\mathbf{x} = (x_1, \ldots, x_N)$ and $\mathbf{y} = (y_1, \ldots, y_N)$. Let $\mathbf{w}_k = (w_k(1), \ldots, w_k(N))$. Suppose that the vectors $\mathbf{w}_k, 1 \leq k \leq M$ are linearly independent in R^N. Then it follows that the matrix $\Gamma_N = ((< \mathbf{w}_k, \mathbf{w}_\ell >_N))$ is nonsingular and $det(\Gamma_N) > 0$. Let $\mathbf{a}_\ell = (a_\ell(1), \ldots, a_\ell(N))$ be a vector such that

(i) $< \mathbf{a}_\ell, \mathbf{w}_k >= \delta_{k\ell}, 1 \leq k, \ell \leq N$; and

(ii) $< \mathbf{a}_\ell, \mathbf{a}_\ell >= \frac{1}{N} \sum_{j=1}^{N} a_\ell(j)^2$ is minimum.

Here $\delta_{k\ell}$ is the Kronecker delta function. By using Lagrange multipliers, it can be checked that

$$a_\ell(j) = \frac{1}{det(\Gamma_N)} \sum_{k=1}^{N} (-1)^{\ell+k} \gamma_{\ell k}^N \, w_k(j), \tag{5.1}$$

where $\gamma_{\ell k}^N$ denotes the determinant of the minor (ℓ, k) of the matrix Γ_N. We now construct the wavelet linear estimator, for the d-th derivative of the density $h_\ell(x)$ of the ℓ-th component, at resolution level s. It is defined by

$$[h_\ell^{\hat{(d)}}]_s(x) = \frac{(-1)^d}{N} \sum_{j=1}^{N} a_\ell(j) K_s^{(d)}(X_j, x). \tag{5.2}$$

We now study the properties of the estimator $[h_\ell^{\hat{(d)}}]_s(x)$. For $d = 0$, it can be shown that this estimator is essentially the same as the density estimator studied in Pokhyl'ko [14].

6. Properties of the Estimator $[h_\ell^{\hat{(d)}}]_s(x)$

Lemma 6.1. *Let $Pr_{V_s} g \equiv g_s$ denote the projection of a function $g \in L_2(R)$ on the space V_s. The estimator $[h_\ell^{\hat{(d)}}]_s(x)$ is an unbiased estimator of $[h_\ell^{(d)}]_s(x)$.*

Since X_j is a random variable with the density function $p_j(x) = \sum_{\ell=1}^{M} w_\ell(j) h_\ell(x)$, it follows that

$$E([h_\ell^{\hat{(d)}}]_s(x)) = \frac{(-1)^d}{N} \sum_{j=1}^{N} a_\ell(j) E[K_s^{(d)}(X_j, x)]. \tag{6.1}$$

Note that

$$E[(-1)^d K_s^{(d)}(X_j, x)] \;\; = \;\; \int_{-\infty}^{\infty} (-1)^d K_s^{(d)}(u, x) p_j(u) du$$

$$
\begin{aligned}
&= \int_{-\infty}^{\infty} (-1)^d K_s^{(d)}(u,x)[\sum_{\ell=1}^{M} w_\ell(j) h_\ell(u)]du \\
&= \sum_{\ell=1}^{M} w_\ell(j)[\int_{-\infty}^{\infty} (-1)^d K_s^{(d)}(u,x) h_\ell(u)du] \\
&= \sum_{\ell=1}^{M} w_\ell(j)[\int_{-\infty}^{\infty} K_s(u,x) h_\ell^{(d)}(u)du] \\
&= \sum_{\ell=1}^{M} w_\ell(j) Pr_{V_s} h_\ell^{(d)}(x) \\
&= Pr_{V_s}[\sum_{\ell=1}^{M} w_\ell(j) h_\ell^{(d)}](x) \\
&= Pr_{V_s} p_j^{(d)}(x) \\
&= [p_j^{(d)}]_s(x). \qquad (6.2)
\end{aligned}
$$

The second equality follows by integration by parts and the last three equalities use the fact that projection and derivative are linear operators. Therefore

$$
\begin{aligned}
E([\hat{h_\ell^{(d)}}]_s(x)) &= \frac{1}{N}\sum_{j=1}^{N} a_\ell(j)[p_j^{(d)}]_s(x) \\
&= \frac{1}{N}\sum_{j=1}^{N} a_\ell(j)(\sum_{k=1}^{M} w_k(j)[h_k^{(d)}]_s(x)) \\
&= \sum_{k=1}^{M} [h_k^{(d)}]_s(x) < \mathbf{a}_\ell, \mathbf{w}_k >_N \\
&= [h_\ell^{(d)}]_s(x). \qquad (6.3)
\end{aligned}
$$

□

Lemma 6.2. *Suppose the scaling function $\phi(\cdot)$ has the property that the function $K(x,y)$ is d-times differentiable with respect to x and there exists $F \in L_2(R)$ such that*

$$|K(x,y)| \leq F(x-y).$$

Suppose that Z is a random variable with density function $h \in L_2(R)$ which is d-times differentiable and $h^{(d)} \in L_2(R)$. Then

$$E||(-1)^d K_s^{(d)}(Z,\cdot) - [h^{(d)}]_s(\cdot)||_2^2 \leq 2^{2ds+2s} \int_{-\infty}^{\infty} F^2(u)du. \qquad (6.4)$$

Proof : Observe that

$$E[(-1)^d K_s^{(d)}(Z,x)] = [h^{(d)}]_s(x)$$

and

$$\begin{aligned} &E||(-1)^d K_s^{(d)}(Z,.) - [h^{(d)}]_s(\cdot)||_2^2 \\ &= E(\int_{-\infty}^{\infty} |(-1)^d K_s^{(d)}(Z,x) - E[(-1)^d K_s^{(d)}(Z,x)]|^2 dx) \\ &= \int_{-\infty}^{\infty} E[Y^2(x)]dx, \end{aligned} \tag{6.5}$$

where

$$Y(x) = (-1)^d K_s^{(d)}(Z,x) - E[(-1)^d K_s^{(d)}(Z,x)].$$

Since

$$K_s(x,y) = 2^s K(2^s x, 2^s y),$$

it is easy to see that

$$K_s^{(d)}(x,y) = 2^{ds+s} K(2^s x, 2^s y)$$

which implies that

$$|K_s^{(d)}(x,y)| \leq 2^{ds+s} F(2^s x - 2^s y).$$

Therefore

$$\begin{aligned} E[Y^2(x)] &\leq E([(-1)^d K_s^{(d)}(Z,x)]]^2) \\ &\leq 2^{2ds+2s} \int_{-\infty}^{\infty} F^2(2^s x - 2^s y)h(y)dy. \end{aligned} \tag{6.6}$$

Hence

$$\begin{aligned} E||(-1)^d K_s^{(d)}(Z,.) &- [h^{(d)}]_s(\cdot)||_2^2 \\ &\leq 2^{2ds+2s} \int_{-\infty}^{\infty} [\int_{-\infty}^{\infty} F^2(2^s x - 2^s y)h(y)dy]dx \\ &\leq 2^{2ds+2s} \int_{-\infty}^{\infty} [\int_{-\infty}^{\infty} F^2(2^s x - 2^s y)dx]h(y)dy \\ &= 2^{2ds+s} \int_{-\infty}^{\infty} F^2(v)dv. \end{aligned} \tag{6.7}$$

□

Lemma 6.3. *Suppose the scalar function* ϕ *satisfies the conditions stated in Lemma 6.2. In addition, suppose that the functions* $h_k^{(d)} \in L_2(R), 1 \leq k \leq M$. *Define the estimator* $[h_\ell^{\hat{(d)}}]_s(x)$ *of* $[h_\ell^{(d)}]_s(x)$ *as given in (5.2) for* $1 \leq \ell \leq M$. *Then*

$$E\int_{-\infty}^{\infty} ([h_\ell^{\hat{(d)}}]_s(x) - [h_\ell^{(d)}]_s(x))^2 dx \leq \frac{1}{N} < \mathbf{a}_\ell, \mathbf{a}_\ell)_N 2^{2ds+s} \int_{-\infty}^{\infty} F^2(v)dv \tag{6.8}$$

for $1 \leq \ell \leq M$ *and for all* $s \geq 0$.

Proof : Let

$$Y_j(x) = (-1)^d K_s^{(d)}(X_j, x) - [p_j^{(d)}]_s(x), 1 \leq j \leq N.$$

Then the random variables $Y_j(x), 1 \leq j \leq N$ are independent with mean zero for every x. Applying Lemmas 6.1 and 6.2, we get that

$$\begin{aligned} E\int_{-\infty}^{\infty}([\hat{h_\ell^{(d)}}]_s(x) - [h_\ell^{(d)}]_s(x))^2 dx &= E[\int_{-\infty}^{\infty}\frac{1}{N^2}\sum_{j=1}^{N} a_\ell^2(j) Y_j^2(x)dx] \\ &\leq \frac{1}{N} < \mathbf{a}_\ell, \mathbf{a}_\ell >_N 2^{2ds+s}\int_{-\infty}^{\infty} F^2(v)dv. \end{aligned} \tag{6.9}$$

We now obtain bounds on the mean integrated squared error

$$E(||[\hat{h_\ell^{(d)}}]_j - [h_\ell^{(d)}]]||_2^2).$$

Theorem 6.1. *Suppose the conditions stated in Lemma 6.2 and Lemma 6.3 hold for some $r \geq 0$. . Suppose that the functions $h_k^{(d)} \in F_{s,p,q}(L), 1 \leq k \leq M$ for some $L > 0$. Further suppose that $s \in (0, r+1)$ and $1 \leq q \leq 2$. Then there exists a constant $C = C(q, s, L) > 0$ such that for all $\ell = 1, \ldots, M$,*

$$E(||[\hat{h_\ell^{(d)}}]_j - h_\ell^{(d)}||_2^2) \leq C[(< \mathbf{a}_\ell, \mathbf{a}_\ell >_N \frac{2^{2dj+j}}{N} + 2^{-2js}]. \tag{6.10}$$

Proof. For any function $g \in L_2(R)$, let $Pr_{V_j} g$ denote the projection of g onto the subspace V_j. Since

$$E([\hat{h_\ell^{(d)}}]_j(x)) = [h_\ell^{(d)}]_j(x) = Pr_{V_j} h_\ell^{(d)}(x),$$

and $Pr_{V_j} h_\ell^{(d)} \in V_j$, it follows that

$$E(||[\hat{h_\ell^{(d)}}]_j - h_\ell^{(d)}||_2^2) = E(||[\hat{h_\ell^{(d)}}]_j - E([\hat{h_\ell^{(d)}}]_j)||_2^2) + ||E([\hat{h_\ell^{(d)}}]_j) - h_\ell^{(d)}||_2^2.$$

We now obtain a bound on the second term on the right side of the above equation by applying the Lemmas 3.1 and 6.1 to 6.3. Note that the second term on the right side of the above equation is given by

$$\begin{aligned} ||E([\hat{h_\ell^{(d)}}]_j) - h_\ell^{(d)}||_2^2 &= ||[h_\ell^{(d)}]_j - h_\ell^{(d)}||_2^2 \\ &= \sum_{i=j}^{\infty}\sum_{k=-\infty}^{\infty}|b_{i,k}|^2 \\ &\leq 2^{-2js}\sum_{i=j}^{\infty} 2^{2is}||b_i||_2^2 \\ &\leq 2^{-2js}(||h_\ell^{(d)}||'_{B_{2,2}^s})^2 \end{aligned} \tag{6.11}$$

since $B^s_{2,q} \subset B^s_{2,2}$ (cf. Hardle et al. [8], p.124). Here $b_{i,k}$ denote the wavelet coefficients of the function $h_\ell^{(d)}$ in the space orthogonal to V_j. Furthermore, for all $L > 0$ and all q such that $1 \le q \le 2$, there exists a constant $C_1 = C_1(s, q, L)$ such that for all $h_\ell^{(d)} \in F_{s,2,q}(L)$,

$$||h_\ell^{(d)}||'_{B^s_{2,2}} \le C_1.$$

Hence

$$||E[\hat{h_\ell^{(d)}}]_j - h_\ell^{(d)}||_2^2 \le C_1 2^{-2js}. \tag{6.12}$$

We now obtain an upper bound on the first term

$$E(||[\hat{h_\ell^{(d)}}]_j - E[\hat{h_\ell^{(d)}}]_j||_2^2).$$

Applying Lemma 6.3, we get that

$$\begin{aligned} E(||[\hat{h_\ell^{(d)}}]_j - E[\hat{h_\ell^{(d)}}]_j]||_2^2) &\le \frac{1}{N} < \mathbf{a}_\ell, \mathbf{a}_\ell)_N 2^{2dj+j} \int_{-\infty}^{\infty} F^2(v) dv \\ &= C_2 < \mathbf{a}_\ell, \mathbf{a}_\ell)_N \frac{2^{2dj+j}}{N}. \end{aligned} \tag{6.13}$$

Combining the relations (6.12) and (6.13), we get that there exists a constant $C > 0$ such that

$$E(||[\hat{h_\ell^{(d)}}]_j - h_\ell^{(d)}||_2^2) \le C[< \mathbf{a}_\ell, \mathbf{a}_\ell >_N \frac{2^{2dj+j}}{N} + 2^{-2js}]. \tag{6.14}$$

□

Remark 6.1. If the integer j is chosen so that $2^j \simeq N^{1/(2s+2d+1)}$, then the bound on the right side will be minimum and is of the order $N^{-2s/(2d+2s+1)}[< \mathbf{a}_\ell, \mathbf{a}_\ell >_N +1]$. This results extend Theorem 1 in Pokhyl'ko [14] on the L_2-loss in the problem of density estimation to the estimation of the d-th derivative of a density function constructed from observations of a mixture with varying mixing proportions.

The next result deals with L_∞-loss of the estimator $[\hat{h_\ell^{(d)}}]_j$ as an estimator of $h_\ell^{(d)}$. We first state a lemma due to Pokhyl'ko [14].

Lemma 6.4. *Suppose that a function $g \in C^1(R)$ with $||g||_\infty < \infty, ||g||_2 < \infty$ and $||g'||_\infty < \infty$, where g' denotes the derivative of the function g. Then*

$$||g||_\infty \le (4||g'||_\infty ||g||_2^2)^{1/3}.$$

Theorem 6.2. *Suppose the conditions stated in Lemma 3.1 and Lemma 6.3 hold for some $r \ge 0$. Suppose that the functions $h_k^{(d)} \in W_\infty^{r+1}, 1 \le k \le M$ for some $r \ge d$ and there exists $\gamma > 0$ such that $||h_k^{(d)}||_{W_\infty^{r+1}} \le \gamma$. Further suppose that* *the scaling function $\phi \in C^r$ with compact support. Then there exists a constant $C = C(r,\gamma) > 0$ such that for all $\ell = 1,\dots,M$,*

$$E(||[\hat{h_\ell^{(d)}}]_j - [h_\ell^{(d)}]]||_\infty) \le C[(< \mathbf{a}_\ell, \mathbf{a}_\ell >_N)^{1/2}\frac{2^{dj+j}}{N^{1/3}} + 2^{-j(r+1)}]. \quad (6.15)$$

Proof : Since

$$E([\hat{h_\ell^{(d)}}]_j(x)) = [h_\ell^{(d)}]_j(x) = Pr_{V_j} h_\ell^{(d)}(x),$$

it follows that

$$\begin{aligned} E(||[\hat{h_\ell^{(d)}}]_j - h_\ell^{(d)}||_\infty) &\le E(||[\hat{h_\ell^{(d)}}]_j - E([\hat{h_\ell^{(d)}}]_j)||_\infty) + ||E([\hat{h_\ell^{(d)}}]_j) - h_\ell^{(d)}||_\infty \\ &= E||[\hat{h_\ell^{(d)}}]_j - [h_\ell^{(d)}]_j||_\infty + ||[h_\ell^{(d)}]_j - h_\ell^{(d)}||_\infty. \quad (6.16) \end{aligned}$$

Since $\phi \in C^r(R)$ for some $r \ge 1$ with a compact support, the condition (C3) holds for all $r \ge 1$. Furthermore $\phi \in \tilde{W}^r_\infty(R)$. Hence there exists a constant $C = C(\gamma) > 0$ depending on γ such that

$$||[h_\ell^{(d)}]_j - h_\ell^{(d)}|| \le C(\gamma)2^{-j(r+1)}, \quad (6.17)$$

whenever $h_\ell^{(d)} \in W^{r+1}_\infty(R)$ with $||h_\ell^{(d)}|| \le \gamma$. This follows from Theorem 8.1(ii), Lemma 8.6 and Theorem 8.2 in Hardle et al. [8].

Since $\phi \in C^r(R), r \ge d$ with a compact support, the function $K(x,y)$ is a uniformly bounded and continuously d-times differentiable function with respect to x and there exists a constant $C > 0$ such that $|K_j^{(d)}(x,y)| \le C2^{dj+j}$. Therefore

$$|E((-1)^d K_j^{(d)}(X_i,x))| \le C2^{dj+j}.$$

Let

$$Y_{d,j}(x) = [\hat{h_\ell^{(d)}}]_j(x) - E([\hat{h_\ell^{(d)}}]_j(x)).$$

Then

$$\begin{aligned} |Y_{d,j}(x)| &= |\frac{1}{N}\sum_{i=1}^N a_\ell(i)((-1)^d K_j^{(d)}(X_i,x) - [p_i^{(d)}]_j(x))| \\ &\le \frac{2}{N}C2^{dj+j}\sum_{j=1}^N |a_\ell(j)| \\ &\le 2C2^{dj+j}[< \mathbf{a}_\ell, \mathbf{a}_\ell >_N]^{1/2}. \quad (6.18) \end{aligned}$$

The same argument shows that

$$||Y_{d-1,j}||_\infty \le C2^{dj}\sum_{j=1}^N |a_\ell(j)| < \infty.$$

Applying Lemma 6.4, we get that

$$||Y_{d,j}||_\infty \leq [C2^{dj+j}(<\mathbf{a}_\ell, \mathbf{a}_\ell>_N)^{1/2}||Y_{d,j}||_2^2]^{1/3}$$

for all $X_i, 1 \leq i \leq N$. Observe that $|K_j^{(d)}(x,y)| \leq C2^{dj+j}$ for some constant $C > 0$ and $K_j^{(d)}(x,y) = 0$ if $|x-y| > L = 2$ diam(supp ϕ). Let $F(x) = C2^{dj+j} I_{[0,L]}(|x|)$ where $I(A)$ denotes the indicator function of the set A. Then $F(\cdot) \in L_2(R)$. Applying arguments similar to those in Lemma 6.3, we get that

$$\begin{aligned} E||[\hat{h_\ell^{(d)}}]_j - E[\hat{h_\ell^{(d)}}]_j||_\infty &\leq [C2^{dj+j}(<\mathbf{a}_\ell, \mathbf{a}_\ell>_N>)^{1/2} 2^{2dj+2j} \\ &< \mathbf{a}_\ell, \mathbf{a}_\ell >_N> \int_{-\infty}^{\infty} F^2(x)dx]^{1/3} \\ &\leq C\frac{2^{dj+j}}{N^{1/3}}(<\mathbf{a}_\ell, \mathbf{a}_\ell>_N)^{1/2} \end{aligned} \tag{6.19}$$

for some constant $C > 0$ independent of j and d. Combining the inequalities (6.17) and (6.19), we get that there exists a constant $C > 0$ such that

$$E[||[\hat{h_\ell^{(d)}}]_j - h_\ell^{(d)}||_\infty] \leq C[2^{-j(r+1)} + \frac{2^{dj+j}}{N^{1/3}}(<\mathbf{a}_\ell, \mathbf{a}_\ell>_N)^{1/2}]. \tag{6.20}$$

Remark 6.2. If the integer j is chosen so that $2^j \simeq N^{1/[3(r+d+2)]}$, then the bound on the right side will be minimum and is of the order $N^{-2(r+1)/[3(r+d+2)]}$ $\{[<\mathbf{a}_\ell, \mathbf{a}_\ell>_N]^{1/2} + 1\}$. This result extends Theorem 2 in Pokhyl'ko [14] on the L_∞-loss in the problem of density estimation to the estimation of the d-th derivative of a density function constructed from observations of a mixture with varying mixing proportions.

Remark 6.3. Let $p' \geq \max(p, 2)$. Following the techniques in Leblanc [9] and Prakasa Rao [19], one can get bounds on the $L_{p'}$-loss

$$E(||[\hat{h_\ell^{(d)}}]_j - h_\ell^{(d)}||_{p'}^2),$$

whenever $h_k \in F_{s,p,q}(L), 1 \leq k \leq M$ by noting that

$$E[||[\hat{h_\ell^{(d)}}]_j - h_\ell^{(d)}||_{p'}^2] \leq 2[E(||[\hat{h_\ell^{(d)}}]_j - E[\hat{h_\ell^{(d)}}]_j||_{p'}^2) + ||E([\hat{h_\ell^{(d)}}]_j) - h_\ell^{(d)}||_{p'}^2]$$

and

$$||E([\hat{h_\ell^{(d)}}]_j) - h_\ell^{(d)}||_{p'}^2 \leq C_1 2^{-2js'}$$

for some positive constant C_1 whenever $s \geq \frac{1}{p}$ and $s' = s + \frac{1}{p'} - \frac{1}{p}$ (cf. Leblanc [9], p.83).

If $1 \leq p' \leq 2$, then a bound on the $L_{p'}$-loss

$$E(||[\hat{h_\ell^{(d)}}]_j - h_\ell^{(d)}||_{p'}^{p'})$$

can be obtained by noting that

$$E(||[\hat{h_\ell^{(d)}}]_j - h_\ell^{(d)}||_{p\prime}^{p'}) \leq 2^{p'-1}(E||[\hat{h_\ell^{(d)}}]_j - E[\hat{h_\ell^{(d)}}]_j]||_{p'}^{p'} + ||E([\hat{h_\ell^{(d)}}]_j) - h_\ell^{(d)}||_{p'}^{p'})$$

and

$$||E([\hat{h_\ell^{(d)}}]_j) - h_\ell^{(d)}||_{p'}^{p'} \leq C_2 2^{-2js'p'}$$

for some positive constant C_2. We do not discuss the details here.

References

1. A. Antoniadis, G. Gregoire and I. W. McKeague, Wavelet methods for curve estimation, *J. Amer. Statist. Assoc.*, **89** (1994), 1340-1353.

2. Y. P. Chaubey, H. Doosti and B. L. S. Prakasa Rao, Wavelet based estimation of the derivatives of a density with associated variables, *Int. J. Pure and Appl. Math.*, **27** (2006), 97-106.

3. Y. P. Chaubey, H. Doosti and B. L. S. Prakasa Rao, Wavelet based estimation of the derivatives of a density for a negatively associated process, *J. Statistical Theory and Practice*, **2** (2008), 453-463.

4. I. Daubecheis, Orthogonal bases of compactly supported wavelets, *Comm. Pure Appl. Math.*, **41** (1988), 909-996.

5. I. Daubechies, *Ten Lectures on Wavelets*, SIAM, Philadelphia, (1992).

6. D. Donoho, I. Johnstone, G. Kerkyacharian and D. Picard, Density estimation by wavelet thresholding, *Ann. Statistics*, **24** (1996), 508-539.

7. I. Dewan and B. L. S. Prakasa Rao, A general method of density estimation for associated random variables, *J. Nonparametric Statistics*, **10** (1999), 405-420.

8. W. Hardle, G. Kerkycharian, D. Picard and A. Tsybakov, *Wavelets, Approximations, and Statistical Applications*, Lecture Notes in Statistics, Vol. 129, Springer, New York, (1998).

9. F. Leblanc, Wavelet linear density estimator for a discrete-time stochastic process: L_p-losses, *Statist. Probab. Lett.*, **27** (1996), 71-84.

10. R. E. Maiboroda, Estimators of components of a mixture with varying concentrations, *Ukranian Math. J.*, **48** (1997), 618-622.

11. R. E. Maiboroda, A test for the homogenity of mixtures with varying concentrations, *Ukranian Math. J.* , **52** (2000), 1256-1263.

12. S. G. Mallat, A theory for multiresolution signal decompocition: the wavelet representation, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **11** (1989), 674-693.

13. Y. Meyer, *Ondolettes et Operateurs*, Hermann, Paris, 1990.

14. D. Pokhyl'ko, Wavelet estimators of a density constructed from observations of a mixture, *Theor. Probability and Math. Statist.*, **70** (2005), 135-145.

15. B. L. S. Prakasa Rao, *Nonparametric Functional Estimation*, Academic Press, Orlando, 1983.

16. B. L. S. Prakasa Rao, Nonparametric estimation of the derivatives of a density by the method of wavelets, *Bull. Inform. Cyb.*, **28** (1996), 91-100.

17. B. L. S. Prakasa Rao, Nonparametric functional estimation: an overview, in *Asymptotics, Nonparametrics and Time Series*, Ed. Subir Ghosh, pp. 461-509, Marcel Dekker Inc. New York, 1999.

18. B. L. S. Prakasa Rao, Estimation of the integrated squared density derivative by wavelets, *Bull. Inform. Cyb*., **31** (1999b), 47-65.

19. B. L. S. Prakasa Rao, Wavelet linear density estimation for associated sequences, *J. Indian Statist. Assoc*., **41** (2003), 369-379.

20. G. Strang, Wavelets and dilation equations: a brief introduction, *SIAM Review*, **31** (1989), 614-627.

21. K. Tribouley, Practical estimation of multivariate densities using wavelet methods, *Statistica Neerlandica*, **49** (1995), 41-62.

22. H. Triebel, *Theory of Function Spaces II*, Birkhäuser, Berlin, 1992.

23. B. Vidakovic, *Statistical Modeling by Wavelets*, Wiley, New York, 1999.

24. G. Walter and J. Ghorai, Advantages and disadvantages of density estimation with wavelets, In *Proceedings of the 24th Symp. on the Interface*, Ed. H. Joseph Newton, Interface FNA, VA, **24** (1992), 234-243.