Wavelet Linear Estimation for Different Distributed Random Variables

Jinru Wang  
Department of Applied Science  
Beijing University of Technology  
Beijing, China, 100124  
wangjinru@bjut.edu.cn

Meng Wang  
Department of Applied Science  
Beijing University of Technology  
Beijing, China, 100124  
330870283@qq.com

Abstract—In this paper, we construct a wavelet linear estimator for the component of a finite mixture under independent identically distributed biased observations. We evaluate its performance by determining an upper bound of $L_p$ risk of $\hat{f}_i^* (A, L)$ function classes.

Keywords— Density function; Besov space; Wavelet estimation; $L_p$ risk;

I. INTRODUCTION

The problem of analysis of mixtures with varying mixing proportions occurs in the study of medical, biological, social and other types of data. We observe $n$ random variables $X_1, X_2, \ldots, X_n$ such that for any $i \in \{1, \ldots, n\}$, the density of $X_i$ is the finite mixture:

$$p_i(x) = \sum_{i=1}^{\infty} a_i f_i(x),$$

where $m \in N^+$, and

- $(a_i)_{i,j \in \{1, \ldots, n\}}$ are known positive weights such that for $i \in \{1, \ldots, n\}$,

$$\sum_{i=1}^{\infty} a_i = 1;$$

- $f_1, \ldots, f_n$ are unknown densities.

For a fixed $l \in \{1, \ldots, m\}$, we will to estimate $f_j$ from i.i.d. random variables $X_1, X_2, \ldots, X_n$. Let us now present a brief survey related to this problem under various configurations. When $X_1, X_2, \ldots, X_n$ are i.i.d., i.e. $m = 1$, the estimation of $f_j$ can be considered in e.g. Doukhan (1990), Kerkyacharian and Picard (1992), Donoho (1996). On the other hand, when $X_1, X_2, \ldots, X_n$ are independent different distribution, i.e. $m > 1$, the estimation of $f_j$ has been investigated in e.g. Maiboroda (1996), Hall and Zhou (2003), Pokhylko (2005), Prakasa Rao (2010).Tribouilloy (1995) studied estimation of multivariate densities using wavelet methods. Donoho et al. (1996) investigated density estimation by wavelet thresholding. For a discussion on statistical modeling by wavelets, see Vidakovic (1999). The advantages and disadvantages of the use of wavelet based probability density estimators are discussed in Walter and Ghorai (1992) 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 (1996) that one can obtain precise limits on the asymptotic mean squared error for a wavelet based linear estimator for the density function.

II. PRELIMINARIES ON WAVELETS

The definitions of the scaling function $\phi$, wavelet function $\psi$, weak differentiability and weak derivative can be found in[1].

Definition 2.1. The scaling function $\phi$ is called $r$ - regular for a given integer $r \geq 1$, if for any nonnegative constant $t \leq r$, and for any integer $k \geq 1$,

$$|\phi^{(i)}(x)| \leq c_k (1 + |x|)^{-r-i}, -\infty < x < \infty,$$

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

Definition 2.2. Let $1 \leq p < \infty$ and $n \geq 0$ be an integer. A function $f \in L_p(R)$ belongs to Sobolve space $W^n_p(R)$, if it is $n$-times weakly differentiable and $f^{(n)} \in L_p(R)$.

In particular, $W^n_p(R) = L_p(R)$. The space $W^n_p(R)$ is equipped with the norm

$$\|f\|_{W^n_p} = \|f\|_p + \|f^{(n)}\|_p,$$

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

Let $\hat{W}^n_p(R) = W^n_p(R)$, if $1 \leq p < \infty$, and $\hat{W}^n_p(R) = \{f : f \in W^n_p(R), f^{(n)}$ uniformly continuous $\}.$

Note that $\hat{W}^n_p(R) = L_p(R), 1 \leq p < \infty.$
Let $f$ be a function in $L^p(\mathbb{R}), 1 \leq p < \infty$. Let $\tau_h f(x) = f(x-h), \Delta_s f = \tau_s f - f$. We also define $\Delta_s^l f = \Delta_s \Delta_s f$.

For $t \geq 0$, the moduli of continuity are defined by 
$$\omega^t(f, t) = \sup_{|h| \leq t} \|f(x+h)-f(x)\|_p.$$ 
Let $1 \leq p < \infty$, suppose there exists a function $\varepsilon(t)$ on $[0, \infty)$ such that $\|f\|_p < \infty$, where 
$$\|f\|_p = \left(\int_0^\infty \varepsilon(t)^p \, dt\right)^{1/p},$$ 
if $1 \leq p < \infty$.

**Definition 2.4** Let $1 \leq p, q \leq \infty$ and $n = n + \alpha$, with $n \in \{0, 1, \cdots \}$. The Besov space $B^p_q (\mathbb{R})$ is the space of all function $f$ such that 
$$f \in W^p_q (\mathbb{R}), \omega^t(f^\alpha, t) = \varepsilon(t)t^\alpha,$$ 
where $\|f\|_{B^p_q} < \infty$.

The space $B^p_q (\mathbb{R})$ is equipped with the norm 
$$\|f\|_{B^p_q} = \|f\|_p + \omega^t(f^\alpha, t).$$

**Definition 2.5** Let $\widetilde{B}^p_q (\mathbb{R})$ be the set defined by 
$$\widetilde{B}^p_q (\mathbb{R}) := \{f \in B^p_q (\mathbb{R}), \|f\|_{B^p_q} \leq L, f \text{ is probability density, } \text{supp} f \leq A\}.$$ 

Here are some necessary conditions posed on $\phi$.

**Condition (\(\theta\))** Let $\theta_\phi (x) := \sum_{k \in Z} |\phi(x-k)|$, then 
$$\text{ess sup}_x \theta_\phi (x) < \infty.$$ 

**Condition H.** There exists an integral function $F(x)$, such that 
$$|K(x,y)| \leq F(x-y), \forall x,y \in \mathbb{R}.$$ 

**Condition H(N).** Condition H holds and 
$$\int |F(x)| \, dx < \infty.$$ 

**Condition M(N).** Condition H(N) is satisfied and 
$$\int K(x,y)(y-x)^s \, dy = \delta_{nk}, \forall k = 0, \ldots, N.$$ 

**Definition 2.6.** Let $\phi$ be a scaling function that satisfies condition (\(\theta\)). A kernel $K(x,y) = \sum_{k \in Z} \phi(x-k)\overline{\phi(y-k)}$ is called orthogonal projective kernel associated with $\phi$.

**Proposition 2.7.** Let $\phi$ be a scaling function. If condition (\(\theta\)) holds for $\phi$, then 
$$P(f) = \int K(x,y)f(y) \, dy,$$ 
where $K(x,y) = 2^j K(2^j x - 2^j y)$.

The following lemmas are all from the references [1].

**Lemma 2.8.** Let $s > 0, 1 \leq r, q \leq \infty$, then 
(1) $B^r_q (\mathbb{R}) \subset B^s_q (\mathbb{R})$, $s' > s$ or $s' = s, q' \leq q$;
(2) $B^r_q (\mathbb{R}) \subset B^r_q (\mathbb{R})$, $r' > r$, $s' = s-1/r + 1/r'$;
(3) $B^r_q (\mathbb{R}) \subset B^r_{q''} (\mathbb{R})$, $s > 1/r$.

**Lemma 2.9.** If a function $\phi$ satisfies Condition (\(\theta\)), then for any sequence $\{\lambda_k, k \in Z\}$ satisfying 
$$\|A\|_2 = \left(\sum_k |\lambda_k|^2\right)^{1/2} < \infty,$$ 
and $p$ and $q$ such that $1 \leq p \leq \infty, 1/p + 1/q = 1$, we have 
$$C_1 |A|_2^{2 - \frac{1}{p}} \leq \|\sum_k \lambda_k \phi_{\lambda k}\|_2 \leq C_2 |A|_2^{2 - \frac{1}{q}}.$$ 

**Lemma 2.10.** (Rosenthal inequality) Assume that $X_1, X_2, \ldots, X_n$ are independent random variables such that $E(X_i) = 0, E(|X_i|^r) < \infty$, then there is a constant $C(p) > 0$, such that 
(1) $p \geq 2,$ 
$$E\left(\sum_{i=1}^n |X_i|^r\right) \leq C(p) \sum_{i=1}^n E|X_i|^r + (\sum_{i=1}^n E(X_i^r))^{1/2};$$
(2) $0 < p \leq 2,$ 
$$E\left(\sum_{i=1}^n |X_i|^r\right) \leq (\sum_{i=1}^n E(X_i^r))^{1/2}.$$ 

**Lemma 2.11.** (Approximation theorem in Besov space) Let the $K$ satisfy the Condition $M(N)$, and Condition $H(N)$ for some integer $N \geq 0$. Let $1 \leq p, q \leq \infty$, and $0 < s < N+1$, if $f \in B^p_q (\mathbb{R})$, then 
$$\|K_{f,j}f - f\|_p = 2^{-\delta_{nk}} E_{\delta_{nk}}.$$
where \( \{e_j\} \in l_q \).

### III. ESTIMATOR AND RESULTS

For \( x, y \in R^n \), we define the scalar product as follows:

\[
\langle x, y \rangle := \sum_{i=1}^{n} x_i y_i.
\]

Let \( a_k = (a_k(1), \cdots, a_k(n)) \), suppose that the vector \( \{a_k\}, 1 \leq k \leq n \) are linearly independent in \( R^n \), then it follows that the matrix \( \Gamma_n = \langle a_k a_l \rangle \) is nonsingular and its determinant \( (\Gamma_n) > 0 \).

Let \( a_i = (a_i(1), \cdots, a_i(n)) \) be a vector such that

1. \( \langle a_i, a_l \rangle = \delta_{i,l}, 1 \leq k, l \leq n; \)
2. \( \langle a_i, a_l \rangle = \frac{1}{\det(\Gamma_n)} \sum_{j=1}^{n} (-1)^{j+i} \gamma^\Gamma_{a_i}(i), \)

where \( \gamma^\Gamma_{a_i}(i) \) denotes the determinant of the minor \( (l,k) \) of the matrix \( \Gamma_n \).

Assuming that \( f_j \in \tilde{B}_a^n (A, L) \), we define the linear estimator \( \hat{f}_j(x) \) by

\[
\hat{f}_j(x) = \sum_{a_i} \hat{a}_i \phi_{a_i}(x),
\]

where

\[
\hat{a}_i = \frac{1}{n} \sum_{i=1}^{n} a_i(\hat{\phi}_a(X_i)).
\]

It is easy to see that \( E(\hat{a}_{a,j}) = \alpha_{a,j} \).

**Theorem 3.1.** (Upper bound for \( \hat{f}_j \))

Let the scaling function \( \phi \) be compactly supported and \( r + 1 \)-regular ( \( r + 1 > s \) ). Suppose that \( f_j \in \tilde{B}_a^n (A, L) \) and for all \( 1 \leq l \leq m, 1 \leq r, p < \infty, s > r/p \), then we have

\[
\sup_{f_j \in B_n^s(A, L)} E \left\| f_j(x) - f_j(x) \right\|_p < \frac{1}{n} \left\| \sum_{i=1}^{n} a_i \phi_{a_i}(X_i) \right\|_p \left\| \phi \right\|_{L^p} \left\| \phi \right\|_{L^\infty}.
\]

Proof: First of all, we divide the risk into two parts: the stochastic error and the bias error.

\[
E \left\| \hat{f}_j(x) - f_j(x) \right\|_p \leq E \left\| \hat{f}_j(x) - P_j f_j(x) \right\|_p + E \left\| P_j f_j(x) - f_j(x) \right\|_p.
\]

Now we estimate the bias error \( \left\| P_j f_j(x) - f_j(x) \right\|_p \).

1. For \( r = p \), using the Approximation theorem in Besov space, we have

\[
\left\| P_j f_j(x) - f_j(x) \right\|_p < 2^{-bp}.
\]

2. For \( r < p \), from the Approximation in Besov space and Corollary 2.1, we have

\[
\sup_{f_j \in B_n^s(A, L)} \left\| P_j f_j(x) - f_j(x) \right\|_p < 2^{-bp},
\]

where \( s' = s - (1/r + 1/p) \).

3. For \( r > p \), using Hölder inequality, we have

\[
\left\| P_j f_j(x) - f_j(x) \right\|_p < 2^{-bp}.
\]

Hence for \( \forall r \in [1, \infty) \), we get

\[
\sup_{f_j \in B_n^s(A, L)} \left\| f_j(x) - f_j(x) \right\|_p < 2^{-bp},
\]

where \( s' = s - (1/r + 1/p) \).

Now we estimate a stochastic error

\[
E \left\| \hat{f}_j(x) - P_j f_j(x) \right\|_p.
\]

Applying Lemma 2.2, we get

\[
E \left\| \hat{f}_j(x) - P_j f_j(x) \right\|_p = E \left\| \sum_{i=1}^{n} (\hat{a}_{a,j} - \alpha_{a,j}) \phi_{a_j}(x) \right\|_p \leq 2^{bp(1/2 - 1/p)} E \left\| \sum_{i=1}^{n} (\hat{a}_{a,j} - \alpha_{a,j}) \right\|_p \leq 2^{bp(1/2 - 1/p)} E \left\| \hat{a}_{a,j} - \alpha_{a,j} \right\|_p.
\]

To estimate the term \( E \left\| \hat{a}_{a,j} - \alpha_{a,j} \right\|_p \), using triangle inequality, we get

\[
E \left\| \hat{a}_{a,j} - \alpha_{a,j} \right\|_p = E \left\| \sum_{i=1}^{n} a_i(\hat{\phi}_a(X_i)) - \sum_{i=1}^{n} a_i(\phi_a(X_i)) \right\|_p = E \left\| \sum_{i=1}^{n} (\hat{a}_i(i) - a_i) \phi_a(X_i) \right\|_p = E \left\| \sum_{i=1}^{n} \gamma_i \right\|,
\]

where \( \gamma_i \) denotes the determinant of the minor \( (i, l) \) of the matrix \( \Gamma_n \).

Now we estimate the bias error \( \left\| P_j f_j(x) - f_j(x) \right\|_p \).

1. For \( r = p \), using the Approximation theorem in Besov space, we have

\[
\left\| P_j f_j(x) - f_j(x) \right\|_p < 2^{-bp}.
\]

2. For \( r < p \), from the Approximation theorem in Besov space and Corollary 2.1, we have

\[
\sup_{f_j \in B_n^s(A, L)} \left\| P_j f_j(x) - f_j(x) \right\|_p < 2^{-bp},
\]

Hence for \( \forall r \in [1, \infty) \), we get

\[
\sup_{f_j \in B_n^s(A, L)} \left\| f_j(x) - f_j(x) \right\|_p < 2^{-bp},
\]

where \( s' = s - (1/r + 1/p) \).

Now we estimate a stochastic error

\[
E \left\| \hat{f}_j(x) - P_j f_j(x) \right\|_p.
\]

Applying Lemma 2.2, we get

\[
E \left\| \hat{f}_j(x) - P_j f_j(x) \right\|_p = E \left\| \sum_{i=1}^{n} (\hat{a}_{a,j} - \alpha_{a,j}) \phi_{a_j}(x) \right\|_p \leq 2^{bp(1/2 - 1/p)} E \left\| \sum_{i=1}^{n} (\hat{a}_{a,j} - \alpha_{a,j}) \right\|_p \leq 2^{bp(1/2 - 1/p)} E \left\| \hat{a}_{a,j} - \alpha_{a,j} \right\|_p.
\]

To estimate the term \( E \left\| \hat{a}_{a,j} - \alpha_{a,j} \right\|_p \), using triangle inequality, we get

\[
E \left\| \hat{a}_{a,j} - \alpha_{a,j} \right\|_p = E \left\| \sum_{i=1}^{n} a_i(\hat{\phi}_a(X_i)) - \sum_{i=1}^{n} a_i(\phi_a(X_i)) \right\|_p = E \left\| \sum_{i=1}^{n} (\hat{a}_i(i) - a_i) \phi_a(X_i) \right\|_p = E \left\| \sum_{i=1}^{n} \gamma_i \right\|,
\]

where \( \gamma_i \) denotes the determinant of the minor \( (i, l) \) of the matrix \( \Gamma_n \).
where \( Y_i = a_i(t)\phi_n(X_i) - E(\phi_n(X_i)) \) are i.i.d. centered random variables.

Note that \( Y_i \) are also uniformly bounded by \( M \gg 2^{\frac{p}{2}} \| E \|_{\infty} < \infty \), it is clear that
\[
\sum_{i=1}^{n} E(Y_i^2) \leq \sum_{i=1}^{n} (a_i(t))^2 E(\phi_n(X_i))^2 \leq \sum_{i=1}^{n} (a_i(t))^2 \int p(x)dx \leq \sum_{i=1}^{n} (a_i(t))^2 \| p \|_{\infty} \leq L \sum_{i=1}^{n} (a_i(t))^2.
\]

From the Rosenthal inequality, for \( p > 2 \), i.e. \( 1 - p < -p/2 \), and \( 2^j \leq n \), we can get
\[
T_i = E\left( \frac{1}{n} \sum_{i=1}^{n} Y_i^2 \right) \leq C(p) \left( \frac{1}{n^p} \sum_{i=1}^{n} E(Y_i^2) + \left( \sum_{i=1}^{n} E(Y_i^2) \right)^{\frac{p}{2}} \right) \leq C\left( 2^{\frac{p}{2} - 1} \sum_{i=1}^{n} (a_i(t))^2 + \left( \sum_{i=1}^{n} (a_i(t))^2 \right)^{\frac{p}{2}} \right) \leq C\left( 2^{\frac{p}{2} - 1} n^{-p+1} \left\{ a_i, a_j \right\}_{n} + n^{\frac{p}{2}} \left\{ a_i, a_j \right\}_{n} \right) \leq C\left( 2^{\frac{p}{2} - 1} n^{-p+1} \left\{ a_i, a_j \right\}_{n} + n^{\frac{p}{2}} \left\{ a_i, a_j \right\}_{n} \right) \leq C\left( n^{-p+1} \left\{ a_i, a_j \right\}_{n} + n^{\frac{p}{2}} \left\{ a_i, a_j \right\}_{n} \right) < n^{\frac{p}{2}} \left\{ a_i, a_j \right\}_{n}.
\]

On the other hand, for \( 1 \leq p < 2 \), we get
\[
T_i = E\left( \frac{1}{n} \sum_{i=1}^{n} Y_i^2 \right) \leq \frac{1}{n^p} \left( \sum_{i=1}^{n} E(Y_i^2) \right)^{\frac{p}{2}} = \frac{1}{n^p} \left( \sum_{i=1}^{n} (a_i(t))^2 \right)^{\frac{p}{2}} \leq \frac{1}{n^p} \left( \sum_{i=1}^{n} (a_i(t))^2 \right)^{\frac{p}{2}} = n^{\frac{p}{2}} \left\{ a_i, a_j \right\}_{n}.
\]

Put \( T_i, T_j \) together, we have
\[
E\left( \hat{x}_{j,k} - \alpha_{j,k} \right)^2 \leq C\left( T_i + T_j \right) < n^{\frac{p}{2}} \left\{ a_i, a_j \right\}_{n}.
\]

Hence
\[
E\left\| \hat{f}_i(x) - P_z f_i(x) \right\|^p \leq 2^{p\left(1/2 - 1/p \right)} \sum_{k} E\left( \hat{x}_{j,k} - \alpha_{j,k} \right)^p \leq 2^{p\left(1/2 - 1/p \right)} 2^{1/p} n^{p/2} \left\{ a_i, a_j \right\}_{n}.
\]

Finally, through the stochastic error and bias error estimate, we get
\[
\sup_{f \in \mathcal{B}^{(1)}(\mu, 1)} E\left\| \hat{f}_i(x) - f_i(x) \right\|^p < \left( \frac{2^{1/p}}{n} \right)^p \left\{ a_i, a_j \right\}_{n}^2 + 2^{-2p}.
\]

The above expression has a minimum when the two antagonistic terms are balanced, i.e. for \( 2^{j} \equiv n^{\frac{1}{2p-1}} \). In this case, we obtain
\[
\sup_{f \in \mathcal{B}^{(1)}(\mu, 1)} E\left\| \hat{f}_i(x) - f_i(x) \right\|^p < n^{-\frac{2p}{2p-1}} \left\{ a_i, a_j \right\}_{n}^2 + 1\right]
\]

where \( s' = s - \left( \frac{1}{p} - \frac{1}{p} \right) \).

ACKNOWLEDGMENT

This paper is supported by Beijing Educational Committee Foundation (No. PHR201008022) and Fundamental Research Foundation of Beijing University of Technology.

REFERENCES