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 $\tilde{B}^s_{rq}(A,L)$ 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 $X_1, X_2, \cdots, X_n$ such that for any $i \in \{1, \cdots, n\}$, the density of X_i is the finite mixture:

$$p_i(x) = \sum_{l=1}^{m} \omega_l(i) f_l(x),$$

where $m \in N^+$, and

- $(\omega_l(i))_{(i,l)\in\{1,\cdots n\}\times\{1,\cdots m\}}$ are known positive weights such that for $i \in \{1, \cdots, n\}$,

$$\sum_{l=1}^{m} \omega_l(i) = 1;$$

- $f_1, \cdots f_m$ are unknown densities.

For a fixed $l \in \{1, \cdots m\}$, we will to estimate f_l from i.i.d . random variables $X_1, X_2, \cdots, X_n$.

Let us now present a brief survey related to this problem under various configurations. When $X_1, X_2, \cdots, X_n$ are i.i.d., i.e. $m = 1$, the estimation of f_l can be considered in e.g. Doukhan (1990), Kerkyacharian and Picard (1992), Donoho (1996). On the other hand, when $X_1, X_2, \cdots, X_n$ are independent different distribution, i.e. $m > 1$, the estimation of f_l has been investigated in e.g. Maiboroda (1996), Hall and Zhou (2003), Pokhylko (2005) , Prakasa Rao (2010).Tribouliy (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.

To estimate f_l, several methods such as kernel, spine, wavelet are possible (see e.g. Prakasa Rao (1983, 1999), hardle et al. (1998) and Tsybakov (2004)). In this paper, our aim is to discuss wavelet linear estimators for probability density function when the samples of observation come from a mixture of several components with varying mixing proportions. We propose an estimator for the density based on wavelets and obtain upper bounds on the L_p risk of $\tilde{B}^s_{rq}(A,L)$ function classes.

The paper is organized as follows: Assumptions on the model and some notations are introduced in section I. Section II briefly describes the preliminaries on wavelet and space $\tilde{B}^s_{rq}(A,L)$. The estimators and the main result are presented in section III.

II. PRELIMINARIES ON WAVELETS

The definitions of the scaling function ϕ, wavelet function ψ, weak differentiability and weak derivative can be found in[1].

Definition 2.1. The scaling function ϕ is called r-regular for a given integer $r \ge 1$, if for any nonnegative constant $l \le r$, and for any integer $k \ge 1$,

$$\left|\phi^{(l)}(x)\right| \le c_k(1+|x|)^{-k}, -\infty < x < \infty,$$

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

Definition 2.2. Let $1 \le p < \infty$ and $n \ge 0$ 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, $W^0_p(R) = L_p(R)$. The space $W^n_p(R)$ is equipped with the norm

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

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

Let $\tilde{W}^n_p(R) = W^n_p(R)$, if $1 \le p < \infty$, and

$$\tilde{W}^n_\infty(R) = \{f : f \in W^n_\infty(R), f^{(n)} \text{ uniformly continuous}\}.$$

Note that $\tilde{W}^0_p(R) = L_p(R), 1 \le p < \infty$.

Definition 2.3. Let f be a function in $L_p(R), 1\le p<\infty$.

Let $\tau_h f(x)=f(x-h), \Delta_h f=\tau_h f-f$. We also define

$$\Delta_h^2 f=\Delta_h\Delta_h f.$$

For $t\ge 0$, the moduli of continuity are defined by

$$\omega_p^1(f,t)=\sup_{|h|\le t}\|\Delta_h f\|_p, \omega_p^2(f,t)=\sup_{|h|\le t}\|\Delta_h^2 f\|_p.$$

Let $1\le p<\infty$. suppose there exists a function $\varepsilon(t)$ on $[0,\infty)$ such that $\|\varepsilon\|_q^*<\infty$, where

$$\|\varepsilon\|_q^*=(\int_0^\infty t^{-1}|\varepsilon(t)|^q dt)^{\frac{1}{q}}, \text{ if } 1\le p<\infty,$$

Definition 2.4 Let $1\le p,q\le\infty$ and $s=n+\alpha$, with $n\in\{0,1,\cdots\}$, The Besov space $B_{pq}^s(R)$ is the space of all function f such that

$$f\in W_p^n(R), \ \omega_p^2(f^{(n)},t)=\varepsilon(t)t^\alpha,$$

where $\|\varepsilon\|_q^*<\infty$.

The space $B_{pq}^s(R)$ is equipped with the norm

$$\|f\|_{spq}=\|f\|_{W_p^n}+\left\|\frac{\omega_p^2(f^{(n)},t)}{t^\alpha}\right\|_q^*.$$

Definition 2.5. Let $\tilde{B}_{rq}^s(A,L)$ be the set defined by

$\tilde{B}_{rq}^s(A,L):=\{f\,|\, f\in B_{rq}^s(R), \|f\|_{B_{rq}^s}\le L, f$ is probability density, $|\text{supp}f|\le A\}$.

Here are some necessary conditions posed on ϕ.

Condition (θ) Let $\theta_\phi(x):=\sum_{k\in z}|\phi(x-k)|$, then

$$ess\sup_x \theta_\varphi(x)<\infty.$$

Condition H. There exists an integrel function $F(x)$, such that

$$|K(x,y)|\le F(x-y), \forall x,y\in R.$$

Condition $H(N)$. Condition H holds and

$$\int|x|^N F(x)dx<\infty.$$

Condition $M(N)$. Condition $H(N)$ is satisfied and

$$\int K(x,y)(y-x)^k dy=\delta_{0k}, \forall k=0,\ldots,N.$$

Definition 2.6. Let ϕ be a scaling function that satisfies condition (θ). A kernel

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

is called orthogonal projective kernel associated with ϕ.

Proposition 2.7. Let ϕ be a scaling function. If condition (θ) holds for ϕ, then

$$P_J f(x)=K_J f(x)=\int_R K_J(x,y)f(y)dy,$$

where $K_J(x,y)=2^J K(2^J x-2^J y)$.

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

lemma 2.8. Let $s>0, 1\le r,q\le\infty$, then

(1) $B_{rq'}^{s'}\subset B_{rq}^s$, $s'>s$ or $s'=s$, $q'\le q$;

(2) $B_{rq}^s\subset B_{r'q}^{s'}$, $r'>r$, $s'=s-1/r+1/r'$;

(3) $B_{r\infty}^s\subset B_{\infty\infty}^{s-1/r}$, $s>1/r$.

lemma 2.9. If a function ϕ satisfies Condition (θ), then for any sequence $\{\lambda_k, k\in Z\}$ satisfying

$$\|\lambda\|_{l_p}=(\sum_k|\lambda_k|^p)^{\frac{1}{p}}<\infty,$$

and p and q such that $1\le p\le\infty$, $\frac{1}{p}+\frac{1}{q}=1$, we have

$$C_1\|\lambda\|_{l_p}2^{(\frac{J}{2}-\frac{J}{p})}\le\left\|\sum_k\lambda_k\phi_{J,k}\right\|_p\le C_2\|\lambda\|_{l_p}2^{(\frac{J}{2}-\frac{J}{p})}.$$

lemma 2.10. (Rosenthal inequality)

Assume that $X_1, X_2,\ldots, X_n$ are independent random variables such that $E(X_i)=0$, and $E(|X_i|^p)<\infty$, then there is a constant $C(p)>0$, such that

(1) $p\ge 2$,

$$E(\left|\sum_{i=1}^n X_i\right|^p)\le C(p)\{\sum_{i=1}^n E|X_i|^p+(\sum_{i=1}^n E(X_i^2))^{p/2}\};$$

(2) $0<p\le 2$,

$$E(\left|\sum_{i=1}^n X_i\right|^p)\le(\sum_{i=1}^n E(X_i^2))^{p/2}.$$

lemma 2.11. (Approximation theorem in Besov space)

Let the K satisfy the Condition $M(N)$, and Condition $H(N)$ for some integer $N\ge 0$. Let $1\le p,q\le\infty$, and $0<s<N+1$, if $f\in B_{pq}^s(R)$, then

$$\|K_J f-f\|_p=2^{-Js}\varepsilon_J,$$

where $\{\varepsilon_J\} \in l_q$.

III. ESTIMATOR AND RESULTS

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

$$\langle x, y \rangle_n := \frac{1}{n} \sum_{k=1}^{n} x_k y_k .$$

Let $\omega_k = (\omega_k(1), \cdots \omega_k(n))$, suppose that the vector $\{\omega_k, 1 \le k \le n\}$ are linearly independent in R^n, then it follows that the matrix $\Gamma_m = (\langle \omega_k\ \omega_l \rangle_n)$ is nonsingular and it's determinant $(\Gamma_m) > 0$.

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

(1) $\langle a_l, \omega_k \rangle_n = \delta_{kl}, 1 \le k, l \le n$;

(2) $\langle a_l, a_l \rangle = \frac{1}{n} \sum_{i=1}^{n} (a_l(i))^2$ is minimum,

here δ_{kl} is the Kronecker delta function. By using Lagrange multipliers, it can be checked that

$$a_l(i) = \frac{1}{\det(\Gamma_m)} \sum_{k=1}^{n} (-1)^{(l+k)} \gamma_{lk}^n \omega_k(i),$$

where γ_{lk}^n denotes the determinant of the minor (l,k) of the matrix Γ_m.

Assuming that $f_l \in \tilde{B}_{rq}^s(A, L)$, we define the linear estimator $\hat{f}_l(x)$ by

$$\hat{f}_l(x) = \sum_{k \in Z} \hat{\alpha}_{Jk} \phi_{Jk}(x),$$

where

$$\hat{\alpha}_{Jk} = \frac{1}{n} \sum_{i=1}^{n} a_l(i) \phi_{Jk}(X_i).$$

It is easy to see that $E(\hat{\alpha}_{J,k}) = \alpha_{J,k}$.

Theorem 3.1. (Upper bound for $\hat{f}_l$)

Let the scaling function ϕ is compactly supported and $r+1$-regular ($r+1 > s$). Suppose that $f_l \in \tilde{B}_{rq}^s(A, L)$ and for all $1 \le l \le m$, $1 \le r, p < \infty$, $s > 1/r$, then we have

$$\sup_{f_l \in \tilde{B}_{rq}^s(A,L)} E\left\| \hat{f}_l(x) - f_l(x) \right\|_p^p \prec n^{-\frac{s'p}{2s'+1}} [\langle a_l, a_l \rangle_n^{\frac{1}{2}} + 1],$$

where $s' = s - (\frac{1}{r} - \frac{1}{p})_+$.

Remark 3.12 $a \prec b$ means that exists a positive constant C such that $a \le Cb$.

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

$$E\left\| \hat{f}_l(x) - f_l(x) \right\|_p^p$$
$$\le E(\left\| \hat{f}_l(x) - P_J f^X(x) \right\|_p + \left\| P_J f_l(x) - f^X(x) \right\|_p)^p$$
$$\le 2^{p-1}(E\left\| \hat{f}_l(x) - P_J f_l(x) \right\|_p^p + \left\| P_J f_l(x) - f_l(x) \right\|_p^p).$$

Now we estimate the bias error $\| P_J f_l(x) - f_l(x) \|_p^p$.

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

$$\| P_J f_l(x) - f_l(x) \|_p^p \prec 2^{-Jsp}.$$

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

$$\sup_{f_l \in \tilde{B}_{rq}^s(A,L)} \| P_J f_l(x) - f_l(x) \|_p^p$$
$$\le \sup_{f_l \in \tilde{B}_{pq}^{s'}(A,L)} \| P_J f_l(x) - f_l(x) \|_p^p$$
$$\prec 2^{-Js'p},$$

where $s' = s - 1/r + 1/p$.

(3) For $r > p$, using *Hölder* inequality， we have

$$\| P_J f_l(x) - f_l(x) \|_p^p$$
$$= \int | P_J f_l(x) - f_l(x) |^p dx$$
$$\le (\int | P_J f_l(x) - f_l(x) |^{p \cdot \frac{r}{p}} dx)^{\frac{p}{r}} (\int 1 dx)^{1 - \frac{p}{r}}$$
$$\le C \| P_J f_l(x) - f_l(x) \|_r^p$$
$$\prec 2^{-sJp}.$$

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

$$\sup_{f_l \in \tilde{B}_{rq}^s(A,L)} \| P_J f_l(x) - f_l(x) \|_p^p \prec 2^{-s'Jp},$$

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

Now we estimate a stochastic error

$$E\left\| \hat{f}_l(x) - P_J f_l(x) \right\|_p^p.$$

Applying lemma 2.2 , we get

$$E\left\| \hat{f}_l(x) - P_J f_l(x) \right\|_p^p$$
$$= E\left\| \sum_k (\hat{\alpha}_{J,k} - \alpha_{J,k}) \phi_{J,k}(x) \right\|_p^p$$
$$\le 2^{Jp(1/2 - 1/p)} E\left\| \sum_k (\hat{\alpha}_{J,k} - \alpha_{J,k}) \right\|_p^p$$
$$\le 2^{Jp(1/2 - 1/p)} \sum_k E\left| \hat{\alpha}_{J,k} - \alpha_{J,k} \right|^p.$$

To estimate the term $E\left| \hat{\alpha}_{J,k} - \alpha_{J,k} \right|^p$, using triangle inequality, we get

$$E\left| \hat{\alpha}_{J,k} - \alpha_{J,k} \right|^p$$
$$= E\left| \frac{1}{n} \sum_{i=1}^{n} a_l(i) \phi_{Jk}(X_i) - E(\frac{1}{n} \sum_{i=1}^{n} a_l(i) \phi_{Jk}(X_i)) \right|^p$$
$$= E\left| \frac{1}{n} \sum_{i=1}^{n} [a_l(i)(\phi_{Jk}(X_i) - E(\phi_{Jk}(X_i))] \right|^p$$
$$= E\left| \frac{1}{n} \sum_{i=1}^{n} Y_i \right|^p,$$

where $Y_i = a_l(i)[\phi_{Jk}(X_i) - E(\phi_{Jk}(X_i))]$ are i.i.d. centered random variables.

Note that Y_i are also uniformly bounded by $M := 2^{\frac{j}{2}} \|\phi\|_\infty < \infty$, it is clear that

$$\sum_{i=1}^{n} E(Y_i^2) \le \sum_{i=1}^{n} (a_l(i))^2 E(\phi_{Jk}(X_i))^2$$
$$\le \sum_{i=1}^{n} (a_l(i))^2 \int (\phi_{Jk}(x))^2 p_i(x)dx$$
$$\le \sum_{i=1}^{n} (a_l(i))^2 \|p_i\|_\infty$$
$$\le L\sum_{i=1}^{n} (a_l(i))^2 \cdot$$

From the Rosenthal inequality, for $p > 2$, i.e. $1-p < -p/2$, and $2^j \le n$, we can get

$$T_1 = E\left|\frac{1}{n}\sum_{i=1}^{n} Y_i\right|^p$$
$$\le C(p)\frac{1}{n^p}\{M^{p-2}\sum_{i=1}^{n} E(Y_i^2) + (\sum_{i=1}^{n} E(Y_i^2))^{\frac{p}{2}}\}$$
$$\le C\{2^{\frac{j}{2}(p-2)}\frac{1}{n^p}\sum_{i=1}^{n}(a_l(i))^2 + \frac{1}{n^p}(\sum_{i=1}^{n}(a_l(i))^2)^{\frac{p}{2}}\}$$
$$\le C\{2^{\frac{j}{2}(p-2)} n^{-p+1}\langle a_l, a_l\rangle_n + n^{-\frac{p}{2}}\langle a_l, a_l\rangle_n^{\frac{p}{2}}\}$$
$$\le C\{2^{\frac{j}{2}(p-2)} n^{-p+1}\langle a_l, a_l\rangle_n + n^{-\frac{p}{2}}\langle a_l, a_l\rangle_n^{\frac{p}{2}}\}$$
$$\le C\{n^{\frac{p}{2}-1} n^{-p+1}\langle a_l, a_l\rangle_n + n^{-\frac{p}{2}}\langle a_l, a_l\rangle_n^{\frac{p}{2}}\}$$
$$\prec n^{-\frac{p}{2}}\langle a_l, a_l\rangle_n^{\frac{p}{2}}.$$

On the other hand, for $1 \le p < 2$, we get

$$T_2 = E\left|\frac{1}{n}\sum_{i=1}^{n} Y_i\right|^p$$
$$\prec \frac{1}{n^p}(\sum_{i=1}^{n} E(Y_i^2))^{\frac{p}{2}}$$
$$\prec \frac{1}{n^p}(\sum_{i=1}^{n}(a_l(i))^2)^{\frac{p}{2}}$$
$$\prec n^{-\frac{p}{2}}\langle a_l, a_l\rangle_n^{\frac{p}{2}}.$$

Put T_1, T_2 together, we have

$$E\left|\hat{\alpha}_{J,k} - \alpha_{J,k}\right|^p \le C\{T_1 + T_2\} \prec n^{-\frac{p}{2}}\langle a_l, a_l\rangle_n^{\frac{p}{2}}.$$

Hence

$$E\left\|\hat{f}_l(x) - P_J f_l(x)\right\|_p^p$$
$$\le 2^{Jp(1/2-1/p)}\sum_k E\left|\hat{\alpha}_{J,k} - \alpha_{J,k}\right|^p$$
$$\prec 2^{Jp(1/2-1/p)} 2^J n^{-p/2}\langle a_l, a_l\rangle_n^{\frac{p}{2}}$$
$$= (\frac{2^J}{n})^{p/2}\langle a_l, a_l\rangle_n^{\frac{p}{2}}.$$

Finally, through the stochastic error and bias error estimate, we get

$$\sup_{f_l \in \tilde{B}_{rq}^s(A,L)} E\left\|\hat{f}_l(x) - f_l(x)\right\|_p^p \prec (\frac{2^J}{n})^{p/2}\langle a_l, a_l\rangle_n^{\frac{p}{2}} + 2^{-s'Jp}.$$

The above expression has a minimum when the two antagonistic terms are balanced, i.e. for $2^J \cong n^{\frac{1}{2s'+1}}$. In this case, we obtain

$$\sup_{f_l \in \tilde{B}_{rq}^s(A,L)} E\left\|\hat{f}_l(x) - f_l(x)\right\|_p^p \prec n^{-\frac{s'p}{2s'+1}}[\langle a_l, a_l\rangle_n^{\frac{1}{2}} + 1]\cdot$$

where $s' = s - (\frac{1}{r} - \frac{1}{p})_+$.

ACKNOWLEDGMENT

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

REFERENCES

[1] W. hardle, G. Kerkyacharian, D. Picard, A. Tsybakow, Wavelets, Approximation and Statistical Applications. Springer-Verlag, 1997.

[2] B. L. S. Prakasa Rao, "Wavelet linear density estimation for associated sequences," J. Indian Statist. Assoc., pp. 369-379, 2003.

[3] D. Pokhylko, "Wavelet estimations of a density constructed from observations of a mixture," Theor. Probability and Math, no. 70, pp. 135-145, 2005.

[4] P. Chaubey and H. Doosti, "Wavelet based estimation of the derivatives of a density with associated variables," Int. J. Pure and Appl. Math, no. 27, pp. 97-106, 2006.

[5] B. L. S. Prakasa Rao, "Wavelet linear estimation for derivatives of a density from observations of a mixture with varying mixing proportions," Indian J. Pure Appl. Math, no. 41, pp. 275-291, 2010.

[6] C. Chesneau, "Wavelet linear estimation of a density and its derivatives from observations of a mixtures under quadrant dependence," hal-00510690, version 4-3, 2011.

[7] S. Efromovich, "Density estimation for biased data," The Annals of Statistics, no. 32, pp. 1137-1161, 2004.

[8] D. Donoho, I. Johnstone, G. Kerkyacharian, and D. Picard, "Density estimation by wavelet thresholding ," The Annals of Statistics, no. 24, pp. 508-539, 1996.

[9] H. Tribouley, "Practical estimation of multivariate densities using wavelet methods ," Statistics Neerlandica, no. 49, pp.41-62, 1995.

[10] B. Vidakovic, Statistical Moseling by Wavelets,Wiley, New York, 1999.