International Journal of Pure and Applied Mathematics
Volume 27 No. 1 2006, 93-102

WAVELET BASED ESTIMATION OF THE DERIVATIVES OF A DENSITY WITH ASSOCIATED VARIABLES

Yogendra P. Chaubey[1 §], Hassan Doosti[2], B.L.S. Prakasa Rao[3]

[1]Department of Mathematics and Statistics
Concordia University
1455, de Maisonnekve Blvd. W.
Montreal, Quebec, H3G 1M8, CANADA
e-mail: chaubey@alcor.concordia.ca

[2]Department of Statistics
Ferdowsi University
Mashhad, 91755-1159, IRAN
e-mail: doosti@math.um.ac.ir

[3]Department of Mathematics and Statistics
University of Hyderabad
Hyderabad, 500 046, INDIA
e-mail: blsprsm@uohyd.ernet.in

Abstract: We propose a method of estimation of the derivatives of probability density based wavelets methods for a sequence of associated random variables with a common one-dimensional probability density function and obtain an upper bound on L_p-losses for the such estimators.

AMS Subject Classification: 62G05, 62G07
Key Words: associated dependence, multiresolution analysis, Besov space, wavelets, nonparametric estimation of derivative of density

1. Introduction

Estimation of the derivative of a density is of importance in detecting possible

Received: February 21, 2006

§Correspondence author

bumps in the associated density. As such there has been a lot of interest in nonparametric estimation of these (see Prakasa Rao [15] and Prakasa Rao [16]). The reader may be referred to Härdle et al [9] and Vidakovic [24] for a detailed coverage of wavelet theory in statistics and to Prakasa Rao [18] for a recent comprehensive review and application of these and other methods of nonparametric functional estimation. Antoniadis et al [1] and Masry [12] among others discuss the estimation of regression and density function using the wavelets. Prakasa Rao [16] considered the use of wavelets for estimating the derivatives of a density and investigated further their use for estimating the integrated squared density (see Prakasa Rao [17]). Walter and Ghorai [25] discuss the advantages and disadvantages of wavelet based methods of nonparametric estimation from i.i.d. sequences of random variables. Prakasa Rao [19] echoes the same advantages and disadvantages for the case of associated sequences. Furthermore, he points out that these methods allow one to obtain precise limits on the asymptotic mean squared error for the estimator of density and its derivatives as well as some other functionals of the density (see Prakasa Rao [16] and [17]). Recently, Doosti et al [6] have shown that the results in Prakasa Rao [19] can be extended to the case of negatively associated sequences. Here we study such properties for the derivatives of a density of positively associated sequences along the lines of Prakasa Rao [16]. We recall the definition of association for an arbitrary collection of random variables.

Definition. For a finite index set I, the random variables $\{X_i, i \in I\}$ are said to be positively associated (PA), or just associated, if for any real-valued coordinate-wise nondecreasing functions h_1 and Hh_2 defined on R^I,

$$\mathrm{Cov}\{h_1(X_i, i \in I), h_2(X_j, j \in I)\} \geq 0,$$

whenever $E[h_j^2(X_i, i \in I)] < \infty, j = 1, 2$.

A random process $\{X_i\}_{-\infty}^{\infty}$ is PA if every finite sub-collection has this property.

The concept of positive association (PA), here after called association was introduced by Esary et al [8], who studied it in detail. For a recent review of this concept along with many probabilistic and statistical results, one may refer to the paper by Prakasa Rao and Dewan [20].

The organization of the paper is as follows. In Section 2, we discuss the preliminaries of the wavelet based estimation of the derivatives of the density along with the necessary underlying setup considered in Prakasa Rao [16]. Then in Section 3, we extend his result to squared integrated error measured in $p-$norm for PA Case.

2. Preliminaries

Let $\{X_n, n \geq 1\}$ be a sequence of random variables on the probability space $(\Omega, \aleph, P)$. We suppose that X_i has a bounded and compactly supported marginal density $f(.)$, with respect to Lebesgue measure, which does not depend on i. We estimate this density from n observations $X_i, i = 1, ..., n$. For any function $f \in \mathbf{L}_2(\mathbf{R})$, we can write a formal expansion (see Daubechies [3]):

$$f = \sum_{k \in Z} \alpha_{j_0,k}\phi_{j_0,k} + \sum_{j \geq j_0} \sum_{k \in Z} \delta_{j,k}\psi_{j,k} = P_{j_0} f + \sum_{j \geq j_0} D_j f \, ,$$

where the functions

$$\phi_{j_0,k}(x) = 2^{j_0/2}\phi(2^{j_0}x - k)$$

and

$$\psi_{j,k}(x) = 2^{j/2}\psi(2^j x - k)$$

constitute an (inhomogeneous) orthonormal basis of $\mathbf{L}_2(\mathbf{R})$. Here $\phi(x)$ and $\psi(x)$ are the scale function and the orthogonal wavelet, respectively. Wavelet coefficients are given by the integrals

$$\alpha_{j_0,k} = \int f(x)\phi_{j_0,k}(x)dx, \delta_{j,k} = \int f(x)\psi_{j,k}dx \, .$$

We suppose that both ϕ and $\psi \in \mathbf{C}^{r+1}, r \in \mathbf{N}$, have compact supports included in $[-\delta, \delta]$. Note that, by Corollary 5.5.2 in Daubechies [2], ψ is orthogonal to polynomials of degree $\leq r$,, i.e.

$$\int \psi(x)x^l dx = 0, \quad \forall l = 0.1, ..., r \, .$$

We suppose that f belongs to the Besov class (see Meyer [13], Section VI.10), $F_{s,p,q} = \{f \in B^s_{p,q}, \|f\|_{B^s_{p,q}} \leq M\}$ for some $0 \leq s \leq r+1, p \geq 1$ and $q \geq 1$, where

$$\|f\|_{B^s_{p,q}} = \|P_{j_0}f\|_p + (\sum_{j \geq j_0} (\|D_j f\|_p 2^{js})^q)^{1/q} \, .$$

We may also say $f \in B^s_{p,q}$ if and only if

$$\|\alpha_{j_0,.}\|_{l_p} < \infty, \text{ and } (\sum_{j \geq j_0} (\|\delta_{j,.}\|_{l_p} 2^{j(s+1/2-1/p)})^q)^{1/q} < \infty \, , \tag{2.1}$$

where$\|\gamma_{j,.}\|_{l_p} = (\sum_{k\in Z}\gamma_{j,k}^p)^{1/p}$. We consider Besov spaces essentially because of their executional expressive power (see Triebel [22] and the discussion in Donoho et al [4]). We construct the density estimator (see Prakasa Rao [19])

$$\hat{f} = \sum_{k\in K_{j_0}} \hat{\alpha}_{j_0,k}\phi_{j_0,k}, \text{ with } \hat{\alpha}_{j_0,k} = \frac{1}{n}\sum_{i=1}^{n}\phi_{j_0,k}(X_i), \tag{2.2}$$

where K_{j_0} is the set of k such that $\mathrm{supp}(f)\cap\mathrm{supp}(\phi_{j_0,k}) \neq \emptyset$. The fact that ϕ has a compact support implies that K_{j_0} is finite and $\mathrm{card}(K_{j_0}) = O(2^{j_0})$. Wavelet density estimators aroused much interest in the recent literature, see Donoho et al [5] and Doukhan and Leon [7]. In the case of independent samples the properties of the linear estimator (2.2) have been studied for a variety of error measures and density classes (see Kerkyacharian and Picard [10], Leblanc [11] and Tribouley [23]).

In the setup considered by Prakasa Rao [16], we assume ϕ is a scaling function generating an $r-$regular multiresoulution analysis and $f^{(d)} \in \mathbf{L}_2(\mathbf{R})$. Furthermore, we assume that there exists $C_m \geq 0$ and $\beta_m \geq 0$ such that

$$|f^{(m)}(x)| \leq C_m(1+|x|)^{-\beta_m}, \quad 0 \leq m \leq d. \tag{2.3}$$

Prakasa Rao [16] showed that the projection of $f^{(d)}$ on V_{j_0} is

$$f_{n,d}^{(d)}(x) = \sum_k a_{j_0,k}\phi_{j_0,k}(x),$$

where

$$a_{j_0,k} = (-1)^d \int \phi_{j_0,k}^{(d)}(x) f_X(x) dx.$$

So its estimator is

$$\hat{f}_{n,d}^{(d)}(x) = \sum_k \hat{a}_{j_0,k}\phi_{j_0,k}(x), \tag{2.4}$$

where

$$\hat{a}_{j_0,k} = \frac{(-1)^d}{n}\sum_{i=1}^{n}\phi_{j_0,k}^{(d)}(X_i).$$

For the estimator defined by equation (2.4), the sum is considered over $k \in K_{j_0}$.

3. Main Results

We assume that the following conditions hold.

(A1) The sequence $\{X_n, n \geq 1\}$ is a sequence of associated random variables with

$$u(n) = \sup_{i \geq 1} \sum_{|j-i| \geq n} \operatorname{Cov}(X_i, X_j) \leq C n^{-\alpha}$$

for some $C > 0$ and $\alpha > 0$.

(A2) The function $f_{n,d}^{(d)}(x) \in F_{s,p,q}$ with $s \geq 1/p, p \geq 1$, and $q \geq 1$.

Theorem. *Suppose the conditions (A1) and (A2) hold. Let* $\max(2,p) \leq p' < h < \infty$. *Define the wavelet linear density estimator* $\hat{f}$ *by the relation (2.2). Then, for every* $\epsilon > 0$, *there corresponds a constant* $C > 0$ *such that*

$$\begin{aligned} \mathbf{E}\|\hat{f}_{n,d}^{(d)}(x) - f_{n,d}^{(d)}(x)\|_{p'}^2 &\leq C[2^{j_0} n^{-2}\{(n^{1+\epsilon} 2^{j_0(1/2+d)p'} \\ &+ (n[\max_{1 \leq i \leq n} \sum_{l=1}^{n} \operatorname{Cov}(X_i, X_l)] 2^{j_0(3+2d)})^{p'/2} \\ &+ n^{(h(p'-1)-p'+\alpha(p'-h))/(h-2) \vee (1+\epsilon)} \\ &\times 2^{j_0 h(1/2+d)(p'-2)/(h-2)} 2^{j_0(3+2d)})\}^{2/p'} + 2^{-2s' j_0}], \end{aligned}$$

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

Proof. First, we decompose $\mathbf{E}\|\hat{f}_{n,d}^{(d)}(x) - f_{n,d}^{(d)}(x)\|_{p'}^2$ into a bias term and a stochastic term:

$$\begin{aligned} &\mathbf{E}\|\hat{f}_{n,d}^{(d)}(x) - f_{n,d}^{(d)}(x)\|_{p'}^2 \\ &\qquad \leq 2(\|f_{n,d}^{(d)} - P_{j_0} f_{n,d}^{(d)}\|_{p'}^2 + \mathbf{E}\|\hat{f}_{n,d}^{(d)} - P_{j_0} f_{n,d}^{(d)}\|_{p'}^2) = 2(T_1 + T_2). \end{aligned} \tag{3.1}$$

Now, we want to find upper bounds for T_1 and T_2. Observe that

$$\begin{aligned} \sqrt{T_1} = \|\sum_{j \geq j_0} D_j f_{n,d}^{(d)}\|_{p'} &\leq \sum_{j \geq j_0} (\|D_j f_{n,d}^{(d)}\|_{p'} 2^{js'}) 2^{-js'} \\ &\leq \{\sum_{j \geq j_0} (\|D_j f_{n,d}^{(d)}\|_{p'} 2^{js'})^q\}^{1/q} \{\sum_{j \geq j_0} 2^{-js'\acute{q}}\}^{1/\acute{q}}, \end{aligned}$$

by Holder's inequality with $1/q + 1/\acute{q} = 1$. From the above inequality, we have

$$T_1 \leq C\|f_{n,d}^{(d)}\|_{B_{p',q}^{s'}} 2^{-s' j_0} \leq C\|f_{n,d}^{(d)}\|_{B_{p,q}^{s}} 2^{-s' j_0}.$$

The last inequality holds because of the continuous Sobolev injection (see Triebel [22] and the discussion in Donoho et al [5]). Further, $B^s_{p,q} \subset B^{s'}_{p',q}$ implies that $\|f^{(d)}_{n,d}\|_{B^{s'}_{p',q}} \leq \|f^{(d)}_{n,d}\|_{B^s_{p,q}}$. Hence

$$T_1 \leq K 2^{-2s' j_0} \tag{3.2}$$

for suitable constant $K > 0$. Next, we have

$$T_2 = \mathbf{E}\|\hat{f}^{(d)}_{n,d} - P_{j_0} f^{(d)}_{n,d}\|^2_{p'} = \mathbf{E}\| \sum_{k \in K_{j_0}} (\hat{a}_{j_0,k} - a_{j_0,k})\phi_{j_0,k}(x)\|^2_{p'}.$$

Now the use of Lemma 1 in Leblanc [11], p. 82 (using Meyer [13]) gives

$$T_2 \leq C\mathbf{E}\{\|\hat{a}_{j_0,k} - a_{j_0,k}\|^2_{l_{p'}}\} 2^{2j_0(1/2-1/p')}.$$

Further, by using Jensen's inequality the above equation implies,

$$T_2 \leq C 2^{2j_0(1/2-1/p')}\{\sum_{k \in K_{j_0}} \mathbf{E}|\hat{a}_{j_0,k} - a_{j_0,k}|^{p'}\}^{2/p'}. \tag{3.3}$$

To complete the proof, it is sufficient to estimate $\mathbf{E}_{|}\hat{a}_{j_0,k} - a_{j_0,k}|^{p'}$. We know that

$$\hat{a}_{j_0,k} - a_{j_0,k} = \frac{1}{n}\sum_{i=1}^{n}\{[\phi^{(d)}_{j_0,k}(X_i) - a_{j_0,k}]\}.$$

Denote $\xi_i = \eta(X_i) = [\phi^{(d)}_{j_0,k}(X_i) - a_{j_0,k}]$. Observe that the random variables $\xi_i, 1 \leq i \leq n$ are functions of associated random variables $X_i, 1 \leq i \leq n$. We will now estimate the term

$$\mathbf{E}\ |\hat{a}_{j_0,k} - a_{j_0,k}|^{p'}$$

by applying a Rosenthal type inequality for functions of associated random variables due to Shao and Yu [21], p. 210. Note that the sequence of random variables $\eta(X_i), 1 \leq i \leq n$ are identically distributed with mean zero. Furthermore the function $\eta(x)$ is differentiable with

$$\begin{aligned}\sup_{-\infty<x<\infty}|\acute{\eta}(x)| &= \sup_{-\infty<x<\infty}|\acute{\phi}^{(d+1)}_{j_0,k}(x)| \\ &\leq 2^{j_0(3/2+d)} \sup_{-\infty<x<\infty}|\acute{\phi}^{(d+1)}(x)| \leq B_0 2^{j_0(3/2+d)}\,,\end{aligned}$$

for some constant $B_0 > 0$ since $\phi \in \mathbf{C}^{r+1}$ for some $r \geq d+1$. In addition, for any $h \geq 0$,

$$\mathbf{E}|\eta(X_1)|^h \leq 2^h(\mathbf{E}|\phi^{(d)}_{j_0,k}(X_1)|^h + B_1^h) \\ \leq 2^h.B_1^h.2^{j_0(1/2+d)h} = B_2 2^{j_0(1/2+d)h}\,,$$

where B_1 is a bound on $\phi^{(d)}$ and B_2 is a positive constant independent of j_0.

Applying Theorem 4.2 in Shao and Yu [21], it follows that for any $\epsilon > 0$, there exists a constant D_0 depending only on ϵ, h, p' and α such that

$$\mathbf{E}|\sum_{i=1}^n \xi_i|^{p'} \leq D_0(n^{1+\epsilon}\mathbf{E}|\eta(X_1)|^{p'}) + (n \max_{1\leq i\leq n}\sum_{l=1}^n |\mathrm{Cov}(\eta(X_i), \eta(X_l))|)^{p'/2} \\ + n^{(h(p'-1)-p'+\alpha(p'-h))/(h-2)\vee(1+\epsilon)} \\ \times \|\eta(X_1)\|_h^{h(p'-2)/(h-2)}(B_0^2 2^{j_0(3+2d)}C)^{(h-p')/(h-2)})\,,$$

where B_0 is as defined before. Note that the constants D_0 and B_0 are independent of $k \in K_{j_0}$ and j_0. Applying the Newman's inequality (Newman [14]), we obtain

$$|\mathrm{Cov}(\eta(X_i), \eta(X_l))| \leq \{\sup_{-\infty<x<\infty} |\acute{\eta}(x)|\}^2 \mathrm{Cov}(X_i, X_l) \\ \leq B_0^2 2^{j_0(3+2d)}\mathrm{Cov}(X_i, X_l).$$

Combining the above estimates, we get

$$\begin{array}{rcl} \mathbf{E}|\sum_{i=1}^n \xi_i|^{p'} & \leq & D_0(n^{1+\epsilon}2^{j_0(1/2+d)p'}B_2 \\ & + & (n[\max_{1\leq i\leq n}\sum_{l=1}^n \mathrm{Cov}(X_i, X_l)]B_0^2 2^{j_0(3+2d)})^{p'/2} \\ & + & n^{(h(p'-1)-p'+\alpha(p'-h))/(h-2)\vee(1+\epsilon)} \\ & \times & (B_0 2^{j_0(1/2+d)})^{h(p'-2)/(h-2)}(B_0^2 2^{j_0(3+2d)}C)^{(h-p')/(h-2)}). \end{array} \tag{3.4}$$

Since the above estimator holds for all $k \in K_{j_0}$ and the cardinality of K is $O(2^{j_0})$, it follows that

$$\begin{array}{rcl} T_2 & \leq & C_2 2^{2j_0(1/2-1/p')}2^{2j_0/p'}n^{-\acute{P}}\{D_0(n^{1+\epsilon}2^{j_0(1/2+d)p'}B_2 \\ & + & (n[\max_{1\leq i\leq n}\sum_{l=1}^n \mathrm{Cov}(X_i, X_l)]B_0^2 2^{j_0(3+2d)})^{p'/2} \\ & + & n^{(h(p'-1)-p'+\alpha(p'-h))/(h-2)\vee(1+\epsilon)} \\ & \times & (B_0 2^{j_0(1/2+d)})^{h(p'-2)/(h-2)}(B_0^2 2^{j_0(3+2d)}C)^{(h-p')/(h-2)})\}^{2/p'} \end{array} \tag{3.5}$$

from (3.3) and (3.4). By substituting (3.2) and (3.5) in (3.1), we obtain the desired result. $\square$

Remark. Suppose $1 < p' \leq 2$. One can get upper bounds similar to those as in Theorem 3.1 for the expected loss $\mathbf{E}\|\hat{f} - f\|_{p'}^{p'}$ by observing that

$$\mathbf{E}\|\hat{f}_{n,d}^{(d)} - f_{n,d}^{(d)}\|_{p'}^{p'} \leq 2^{p'-1}(\|f_{n,d}^{(d)} - P_{j_0} f_{n,d}^{(d)}\|_{p'}^{p'} + \mathbf{E}\|\hat{f}_{n,d}^{(d)} - P_{j_0} f_{n,d}^{(d)}\|_{p'}^{p'}),$$

$$\|f_{n,d}^{(d)} - P_{j_0} f_{n,d}^{(d)}\|_{p'}^{p'} \leq C_1 2^{-p's'j_0},$$

and

$$\mathbf{E}\|\hat{f}_{n,d}^{(d)} - P_{j_0} f_{n,d}^{(d)}\|_{p'}^{p'} \leq C_2 2^{2j_0(p'/2-1)}\{\sum_{k \in K_{j_0}} \mathbf{E}|\hat{a}_{j_0,k} - a_{j_0,k}|^{p'}\}$$

for some positive constants C_1 and C_2.

References

[1] A. Antoniadis, G. Grégoire, I. McKeague, Wavelet methods for curve estimation, *Journal of the American Statistical Association*, **89** (1994), 1340-1353.

[2] I. Daubechies, Orthonormal bases of compactly supported wavelets, *Communication in Pure and Applied Mathematics*, **41** (1988), 909-996.

[3] I. Daubechies, *Ten Lectures on Wavelets*, CBMS-NSF Lecture Notes, No. 61, SIAM, Philadelphia (1992).

[4] D.L. Donoho, I.M. Johnstone, G. Kerkyacharian, D. Picard, Wavelet shrinkage: asymptopia (with discussion), *Journal of Royal Statistical Society,* Ser. B **57** (2) (1995), 301-370.

[5] D.L. Donoho, I.M. Johnstone, G. Kerkyacharian, D. Picard, Density estimation by wavelet thresholding, *The Annals of Statistics*, **2** (1996), 508-539.

[6] H. Doosti, V. Fakoor, Y.P. Chaubey, *Wavelet Linear Density Estimation for Negative Associated Sequences*, Preprint, Concordia University (2005).

[7] P. Doukhan, J.R. Leon, Une note sur la déviation quadratique déstimateurs de densités par projections orthogonales, *Comptes Rendus de l'Acadmie des Sciences*, **310**, Série 1 (1990), 425-430.

[8] J.D. Esary, F. Proschan, D.W. Walkup, Association of random variables with applications, *The Annals of Mathematical Statistics,* **38** (1967), 1466-1474.

[9] W. Härdle, G. Kerkyacharian, D. Picard, A. Tsybakov, *Wavelets, Approximations, and Statistical Applications,* Lecture Notes in Statistics, **129**, Springer-Verlag, Berlin (1998).

[10] G. Kerkyacharian, D. Picard, Density estimation in Besov spaces, *Statistics and Probability Letters*, **13** (1992), 15-24.

[11] F. Leblanc, Wavelet linear density estimator for a discrete-time stochastic process: L_p-losses, *Statistics and Probability Letters*, **27** (1996), 71-84.

[12] E. Masry, Probability density estimation from dependent observations using wavelets orthonormal bases, *Statistics and Probability Letters,* **21** (1994), 181–194.

[13] Y. Meyer, *Ondelettes et Operateurs*, Hermann, Paris (1990).

[14] C.M. Newman, Normal fluctuations and the FKG inequalities, *Communications in Mathematical Physics,* **74** (1980), 119-128.

[15] B.L.S. Prakasa Rao, *Nonparametric Functional Estimation*, Academic Press, New York (1983).

[16] B.L.S. Prakasa Rao, Nonparametric estimation of the derivatives of a density by the method of wavelets, *Bulletin of Informatics and Cybernetics*, **28** (1996), 91-100.

[17] B.L.S. Prakasa Rao, Estimation of the integrated squared density derivative by wavelets, *Bulletin of Informatics and Cybernetics*, **31** (1999), 47-65.

[18] B.L.S. Prakasa Rao, Nonparametric functional estimation: An overview, In: *Asymptotics, Nonparametrics and Time Series* (Ed. Subir Ghosh), Marcel Dekker Inc., New York (1999), 461-509.

[19] B.L.S. Prakasa Rao, Wavelet linear density estimation for associated sequences, *Journal of the Indian Statistical Association*, **41** (2003), 369-379

[20] B.L.S. Prakasa Rao, I. Dewan, Associated sequences and related inference problems, In: *Handbook of Statistics*, Volume **19** (Ed-s: D.N. Shanbhag, C.R. Rao), Elsevier Science B.V., Amsterdam (2001), 693-731.

[21] Q.M. Shao, H. Yu, Weak convergence for weighted empirical processes of dependent sequences, *Annals of Probability,* **24** (1996), 2098-2127.

[22] H. Triebel, *Theory of Function Spaces II*, Birkhaüser Verlag, Berlin (1992).

[23] K. Tribouley, Density estimation by cross-validation with wavelet method, *Statistica Neerlandica*, **45** (1995), 41-62.

[24] B. Vidakovic, *Statistical Modeling by Wavelets,* Wiley, New York (1999).

[25] G. Walter, J. Ghorai, Advantages and disadvantages of density estimation with wavelets, In: *Proceedings of the 24-th Symposium on the Interface* (Ed. H. Joseph Newton), Interface FNA, VA (1992), 234-343.