NONPARAMETRIC ESTIMATION OF A MULTIVARIATE PROBABILITY DENSITY FOR MIXING SEQUENCES BY THE METHOD OF WAVELETS

Narges Hosseinioun

Statistics Departement
Payame Noor University
19395-4697 Tehran
Iran

Hassan Doosti

Department of Mathematics
Tarbiat Moallem University
Tehran
Iran
e-mail: h.doosti@tmu.ac.ir

Hossein Ali Niroumand

Department of Statistics
Ferdowsi University of Mashhad
Mashhad
Iran

Abstract. The mathematical theory of wavelet and their applications in statistics have become a well-known technique for non-parametric curve estimation: see e.g. Meyer (1990), Daubachies (1992), Chui (1992), Donoho and Johnstone (1995) and Vidakovic (1999). We Consider the problem of estimation of the partial derivatives of a multivariate probability density f of mixing sequences, using wavelet-based method. Many stochastic processes and time series are known to be mixing. Under certain weak assumptions autoregressive and more generally bilinear time series models are strongly mixing with exponential mixing coefficients. The problem of density estimation from dependent samples is often considered. For instance quadratic losses were considered by Ango Nze and Doukhan (1993). Bosq (1995) and Doukhan and Loen (1990). We investigate the variance and the rate of the almost convergence of wavelet-based estimators. Rate of convergence of estimators when f belongs to the Besov space is also established.

Keywords and phrases: nonparametric estimation of partial derivatives, multivariate density, wavelet, mixing process.

1. Introduction

Methods of nonparametric estimation of a multivariate probability density function and regression function are discussed in Prakasa Rao (1983,1999). The problem of estimation of partial derivatives of multivariate probability density is of interest of Singh (1981), Prakasa Rao (1983), especially to detect concavity or convexity properties of the regression function.

Kernel-type estimation the functional $I_2(f)$ has been investigated by Hall and Marron (1987). and Bickel and Ritov (1988) among others. Prakasa Rao (1996) studied nonparametric estimation of the derivative of a density by wavelets and obtained a precise asymptotic expression for the mean integrated squared error following techniques of Masry (1994). Estimation of the integrated squared density derivatives was discussed in Prakasa Rao (1999) by the method of wavelets and a precise asymptotic expression for the mean squared error had been obtained.

Prakasa Rao (2000) also obtained the almost sure convergence of estimation of the partial derivatives of a probability density. We now extend the result to the case of strongly mixing process . We show that the L_p error of the proposed estimator attains the same rate when the observations are independent. Certain week dependence conditions are imposed to the $\{x_i\}$ defined in $\{\Omega, N, P\}$.

Let N_k^m denote the σ-algebra generated by events $X_k \in A_k, ..., X_m \in A_m$. We consider the following classical mixing conditions:

1. Strong mixing (s.m) also called α-mixing:

$$\sup\sup |p(AB) - p(A)p(B)| = \alpha(s) \to 0 \qquad as \quad s \to \infty$$

2. Complete regularity (c.r.), also called β-mixing:

$$\sup E\{var|p(B|N_1^m) - p(B)|\} = \beta(s) \to 0 \qquad as \quad s \to \infty$$

3. Uniformely strong mixing (u.s.m.), also called $\phi - mixing$:

$$\sup\sup \frac{|p(AB) - p(A)p(B)|}{p(A)} = \phi(s) \to 0 \qquad as \quad s \to \infty$$

4. ρ-mixing:

$$\sup\sup |corr(X, Y)| = \rho(s) \to 0 \qquad as \quad s \to \infty$$

Among various mixing conditions used in the literature, α-mixing is reasonably weak, and has many practical applications. Many stochastic processes and time series are known to be mixing. Under certain weak assumptions autoregressive and more generally bilinear time series models are strongly mixing with exponential mixing coefficients. The problem of density estimation from dependent samples is often considered. For instance quadratic losses were considered by

Ango Nze and Doukhan (1993), Bosq (1995) and Doukhan and Loen (1990). Linear wavelet estimators were also used in context: Doukhan (1988) and Doukhan and Loen (1990). Leblance (1996) also established that the $L_{\acute{p}}$-loss ($2 \leq \acute{p} < \infty$) of the linear wavelet density estimators for a stochastic process converges at the rate $N^{\frac{-s}{(2s+1)}}$ ($s = 1/p + 1/\acute{p}$), when the density of f belongs to the Besov space $B^s_{p,\acute{p}}$. Dooti, Niroumand and Afshari (2006) extended the above result for derivative of a density.

2. Discussion of Theorem's Assumptions

Consider the following conditions:

C_1: The process is ρ-mixing and $\sum_{t=1}^{\infty} \rho(t) \leq R < \infty$.

C_2: The process is ϕ-mixing and $\sum_{t=1}^{\infty} \phi^{1/2}(t) \leq \phi < \infty$.

Since the inequality $\rho(t) \leq 2\varphi^{1/2}(t)$ holds (see Doukhan 1994), C_2 implies C_1. Also note that if X and Y are random variables , then the following covariance inequalities hold.(see Doukhan, 1994, section 1.2.2)

$$cov(X,Y) \leq 2\rho(j-i)\|X\|_2.\|Y\|_2 \tag{2.1}$$

$$cov(X,Y) \leq 2\phi^{1/p}(j-i)\|X\|_p.\|Y\|_q \tag{2.2}$$

for any $p, q \geq 1$ and $1/p + 1/q = 1$.

3. Preliminaries

A multiresolution in R^d is a decomposition of the space $L^2(R^d)$ in to an interesting sequence of closed subspaces $V_j, -\infty < j < \infty$ such that

(i) $\bigcap_{j=-\infty}^{\infty} V_j = 0$,

(ii) $\overline{\bigcup_{j=-\infty}^{\infty}} V_j = L^2(R^d)$,

(iii) there exists a scaling function $\varphi \in V_0$ such that

$$\int_{R^d} \varphi(x)dx = 1$$

and $\{\varphi(x-k), k \in Z^d\}$ is an orthogonal basis for V_0 and for all $h \in L^2(R^d)$,

(iv) for all $k \in z^d$, $h(X) \in V_0 \to h(x-k) \in V_0$ and

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

In fact, the family $\{\varphi_{j,k} = 2^{\frac{jd}{2}}(2^j x - k), k \in Z^d\}$ is an orthonormal basis for V_j.

Definition 3.1. The multiresolution analysis is said to be *r-regular* if $\varphi \in C^{(r)}$ and all its partial derivatives up to total order r are rapidly decreasing, that is for any integer $m \geq 1$ there exists a constant c_m such that

$$|(D^\beta \varphi)(x)| \leq \frac{c_m}{(1+\|x\|)^m} \tag{3.1}$$

for all $|\beta| \leq r$, where

$$(D^\beta \varphi)(x) = \frac{\partial \varphi(x)}{\partial x_1^{\beta_1} \dots \partial x_n^{\beta_n}} \tag{3.2}$$

and

$$\beta = (\beta_1, \dots, \beta_d), |\beta| = \sum_{i=1}^{d} \beta_i. \tag{3.3}$$

Define $\{X_n, n \geq 1\}$ be i.i.d d-dimensional random vectors with density $f(x)$. Suppose f is partially differentiable up to total order r. The problem is to estimate

$$(D^\beta \varphi)(x) = \frac{\partial \varphi(x)}{\partial x_1^{\beta_1} \dots \partial x_n^{\beta_n}}$$

We assume that $D^\beta f = L^2(R^d)$. Let $(D^\beta f)_l$ be the orthogonal projection of $D^\beta f$ on V_l. Then

$$(D^\beta f)(x) = \sum_{k \in z^d} a_{lk} \varphi_{l,k}(x) \tag{3.4}$$

where

$$a_{lk} = \int_{R^d} (D^\beta f)(u) \varphi_{l,k}(u) du \tag{3.5}$$

$$= (-1)^{\sum_{i=1}^{d} \beta_i} \int_{R^d} (D^\beta \varphi_{l,k}) f(u) du \tag{3.6}$$

Hence

$$\hat{a}_{lk} = \frac{(-1)^{\sum_{i=1}^{d} \beta_i}}{n} \sum_{i=1}^{n} (D^\beta \varphi_{l,k})(x_i) \tag{3.7}$$

Let us estimate $D^\beta f$ by

$$(D^\beta f)_n(x) = \sum_{k\in z^d} \hat{a}_{lk}\varphi_{l,k}(x) \tag{3.8}$$

Assume that the support of φ is compact. then the series above has a finite number of nonzero terms for any fixed x. Define the kernel $K(u,v)$ by

$$K(u,v) = \sum_{k\in z^d} \varphi(u-k)(D^\beta\varphi)(v-k)$$

Let

$$E(u,v) = \sum_{k\in z^d} \varphi(u-k)\varphi(v-k) \tag{3.9}$$

Note that

$$\partial_v^\beta E(u,v) = \sum_{k\in z^d} \varphi(u-k)(D^\beta\varphi)(v-k) = K(u,v) \tag{3.10}$$

Then there exists constants $C_m > 0$, for any $m \geq 1$, such that

$$|\partial_u^v \partial_v^x K(u,v)| \leq C_m(1+\|u-v\|)^{-m}, m \geq 1 \tag{3.11}$$

for $|\alpha| \leq r$ and $|\alpha+\beta| \leq r$ (cf. Meyer (1992)). Choosing $\alpha = 0 = \gamma$, we obtain that in particular

$$|K(u,v)| \leq C_m(1+\|u-v\|)^{-m} \tag{3.12}$$

In particular

$$|K(u,v)| \leq C_{d+1}(1+\|u-v\|)^{-(d+1)} \tag{3.13}$$

and

$$\int_{R^d} |K(u,v)|^j dv \leq G_j(d) \tag{3.14}$$

where

$$G_j(d) = 2\pi^{d/2}\frac{\Gamma(d)\Gamma(j+d(j-1))}{\Gamma(d/2)\Gamma((d+1)j))}C_{d+1}^j \tag{3.15}$$

Note that

$$(D^\beta f)_n(x) = \sum_{k\in z^d} \varphi_{l,k}(x)\frac{(-1)^{\sum_{i=1}^d \beta_i}}{n}\sum_{i=1}^n (D^\beta\varphi_{l,k})(x_i)$$

$$= \frac{(-1)^{|\beta_i|}}{n}\sum_{i=1}^n\sum_{k\in z^d}\varphi_{l,k}(x)(D^\beta\varphi_{l,k})(x_i) = \frac{(-1)^{|\beta|}}{n}2^{ld+l|\beta|}\sum_{i=1}^n K(2^l x, 2^l X_i) \tag{3.16}$$

Hence

$$Var[(D^\beta f)_n(x)] = \frac{2^{2ld+2l|\beta|}}{n^2}\left[\sum_{i=1}^{n} Var(K(2^l x, 2^l X_i)) + 2\sum_{i>j} Cov(K(2^l x, 2^l X_i), K(2^l x, 2^l X_j))\right].$$

Following along the lines of Rao (1999), one may easily get

$$n2^{-l(d+2|\beta|)} Var[K(2^l x, 2^l X_i)] \leq M_1 G_2(d)(1+o(1)). \tag{3.17}$$

Applying (2.2), it is straightforward to have

$$\begin{aligned} &Cov(K(2^l x, 2^l X_i), K(2^l x, 2^l X_j)) \\ &\quad = 2\phi^{1/2}(j-i)\|K(2^l x, 2^l x_i)\|_2 \|K(2^l x, 2^l x_j)\|_2 \\ &\quad d = 2\phi^{1/2}(j-i)\left(\int 2^{-2ld}K^2(2^l x, u)du\right)^{1/2}\left(\int 2^{-2ld}K^2(2^l x, v)dv\right)^{1/2} \\ &\quad = 2^{-2ld+1}\phi^{1/2}(j-i)G_2(d). \end{aligned} \tag{3.18}$$

Having C_2 in mind with (3.17) and (3.18), it follows that

$$\sup_{x\in R^d} Var[(D^\beta f)_n(x)] \leq M_1\frac{2^{l(d+2\beta)}}{n} + M_2\frac{2^{l(d+2\beta)}}{n}. \tag{3.19}$$

Let D be a compact set in R^d and $L(n); n \geq 1$ be a non-negative sequence to be chosen later. Suppose that the set D can be covered by a finite number $L(n)$ of cubes $I_j = I_{nj}$ with centers $x_j = x_{nj}$ with sides of length m_n for $j = 1, ..., L(n)$, $m_n = const/L^{1/d}(n)$. This is possible since D is compact. Then

$$\begin{aligned} &\sup_{x\in D}|(D^\beta f_n)(x) - E((D^\beta f_n)(X))| \\ &\quad = \max_{1\leq j\leq L(n)}\sup_{x\in D\cap I_j}|(D^\beta f_n)(x) - E((D^\beta f_n)(X))| \\ &\quad \leq \max_{1\leq j\leq L(n)}\sup_{x\in D\cap I_j}|(D^\beta f_n)(x) - (D^\beta f_n)(X_j)| \\ &\quad + \max_{1\leq j\leq L(n)}|(D^\beta f_n)(x_j) - E((D^\beta f_n)(X))| \\ &\quad + \max_{1\leq j\leq L(n)}\sup_{x\in D\cap I_j}|E((D^\beta f_n)(X)) - E((D^\beta f_n)(X_j))| \\ &\quad = T_1 + T_2 + T_3. \end{aligned} \tag{3.20}$$

Applying (3.23),(3.24) in Rao (1999), one may have

$$T_1 = T_2 = O\left(\frac{2^{l(d+|\beta|+1)}}{L^{1/d}(n)}\right). \tag{3.21}$$

Now, for any x,

$$\begin{aligned} Z_n(x) &= (D^\beta f_n)(x) - E[(D^\beta f_n)(X)] \\ &= = \frac{(-1)^{|\beta|}2^{ld+l|\beta|}}{n}\sum_{i=1}^n K(2^l x, 2^l X_i) - E(K(2^l x, 2^l X_i)) = 1/n\sum_{i=1}^n Y_{ni}, \end{aligned}$$

where

$$Y_{ni} = (-1)^{|\beta|}2^{ld+l|\beta|}\{K(2^l x, 2^l X_i) - E(K(2^l x, 2^l X_i))\}.$$

Note that relation (3.12) implies that

$$|K((u,v)| \le C_2 \tag{3.22}$$

Lemma 3.1. (Doukhan (1994)) *Assume the sequence to satisfy the ϕ-mixing condition and $n\phi(n)$ is bounded and*

(i) $\forall t \in N, EX_t = 0$

(ii) $\exists \sigma^2 \in R^+, \forall n, m \in N : 1/mE(X_n, ..., X_m)^2 \le \sigma^2$

(iii) $\forall t \in N, |x_t| \le M$.

Then, there exists constants a and $b > 0$ such that

$$P\left(\sum_{i=1}^n |X_t| \ge x\sqrt{n}\right) \le a\exp(-bx^2).$$

It is easy to check (i), (ii) and (iii) in the above Lemma applying (3.19),(3.23) as well as the definition of Y_{ni}. Using (3.19) and (3.22), one may easily apply the above result to conclude that for any $\eta_n > 0$ and $t_n > 0$,

$$P(|Z_n(x)| \ge \eta_n) \le P\left(|\sum_{i=1}^n Y_{ni}| \ge n\eta_n\right) \le a\exp(-nb\eta_n^2) \tag{3.23}$$

therefore

$$P\left(\max_{1\le j\le L(n)} |Z_n(x)| \ge \eta_n\right) \le L(n)\exp(-nb\eta_n^2). \tag{3.24}$$

Let

$$\eta_n = \frac{2^{l(d+1+|\beta|)}}{L^{1/d}(n)}.$$

Then

$$P\left(\max_{1\le j\le L(n)} |Z_n(x)| \ge \eta_n\right) \le L(n)\exp\left(-nb\frac{2^{l(d+1+\beta)}}{L^{2/d}(n)}\right). \tag{3.25}$$

Let

$$L(n) = (2^{l(d+1+\beta)}n/\log n)^{d/2}.$$

Then

$$T_1 + T_3 = O\left(2^{l(d/2)+\beta}\frac{(\log)^{1/2}}{n^{1/2}}\right). \tag{3.26}$$

Suppose that $l_n \to \infty$ and

$$\frac{2^{l(d+1+\beta)}\log n}{n} \to 0$$

as $n \to \infty$. Following the relation (3.25), it can be checked that

$$\sum_{n=1}^{\infty} P(|Z_n(x)| \geq \eta_n) \leq \sum_{n=1}^{\infty} \frac{L(n)}{n^a} < \infty$$

for a suitable constant $a > 1$. Hence, by Borel-Cantelli Lemma, it follows that

$$T_2 = \max_{1\leq j\leq L(n)} |Z_n(x_j)| \leq \eta_n \tag{3.27}$$

Combining (3.26) and (3.27), it follows that

$$T_1 + T_2 + T_3 = O\left(2^{l(d/2)+\beta}\frac{(\log)^{1/2}}{n^{1/2}}\right) = O\left(\left(\frac{2^{l(d)+2\beta}\log}{n}\right)^{1/2}\right). \tag{3.28}$$

Note that

$$E[(D^\beta f_n)(x)] = \sum_{k\in z^d} a_{lk}\varphi_{l,k}(x) = P_{V_i}(D^\beta f_n)(x).$$

Hence

$$(D^\beta f_n)(x) - E[(D^\beta f_n)(x)] = (D^\beta f_n)(x) - P_{V_i}(D^\beta f_n)(x) = \sum_{j\geq 1} P_{W_j}(D^\beta f_n)(x).$$

If $(D^\beta f_n)(x) \in B_{spq}$, where B_{spq} is a Besov space, for some $0 < s < r, 1 \leq p, q < \infty$ with $s > d/p$ and if the multiresolution analysis is r-regular, then it follows by arguments in Masry and Kerkyacharian and Picard (1992) that

$$\begin{aligned}
&\sup_{x\in R^d}(D^\beta f_n)(x) - E[(D^\beta f_n)(x)]\\
&\qquad\sup_{x\in R^d}(D^\beta f_n)(x) - P_{V_i}(D^\beta f_n)(x)\\
&\qquad\leq (constant)2^{-(s-d/p)l}J_{spq}(D^\beta f),
\end{aligned} \tag{3.29}$$

where

$$J_{spq}(D^\beta f) = \|P_{v_0 g}\|_{L_P} + \left(\sum_{j\geq 0}(2^{js}\|P_{w_j g}\|_{L_P})^q\right)^{1/q}$$

for $g \in B_{spq}$. Combining (3.28) and (3.29), we have

$$\begin{aligned}
&\sup_{x\in R^d}(D^\beta f_n)(x) - Et[(D^\beta f_n)(x)]\\
&= O\left(\left(\frac{2^{l(d)+2\beta}\log n}{n}\right)^{1/2}\right) + O(2^{-(s-d/p)l})\ a.s.
\end{aligned} \tag{3.30}$$

Now we are ready to have the main result of the paper.

5. Main Results

Let the sequence $\{X_n\}_{n\geq 1}$ be the ϕ-mixing sequence of random variables with the property C_2 such that $n\phi(n)$ is bounded.

Theorem 5.1. *Suppose the multivariate probability density $f(x)$ is bounded with partial derivative up to order r. Further, suppose that $D^\beta f \in L^2(R^d)$ and $l = l_n \to \infty$ such that*

$$\frac{2^{l(d+1+\beta)}\log n}{n} \to 0$$

Then, there exists constants $M_1, M_2 > 0$ such that

$$\sup_{x\in R^d} Var[(D^\beta f)_n(x)] \leq M_1 \frac{2^{l(d+2\beta)}}{n} + M_2 \frac{2^{l(d+2\beta)}}{n}$$

Furthermore, for any compact set $D \in R^d$,

$$\sup_{x\in D} |(D^\beta f_n)(x) - E((D^\beta f_n)(X))| = O\left(\left(\frac{2^{l(d)+2\beta}\log}{n}\right)^{1/2}\right) \quad a.s.$$

In addition if $D \in B_{spq}$ for some $0 < s < r$, $1 \leq p$, $q < \infty$ with $s > d/p$, then

$$\sup_{x\in R^d} (D^\beta f_n)(x) - E[(D^\beta f_n)(x)] = O\left(\left(\frac{2^{l(d)+2\beta}\log n}{n}\right)^{1/2}\right) + O\left(2^{-(s-d/p)l}\right) \quad a.s.$$

and if $D \in B_{s\infty\infty}$, then

$$\sup_{x\in R^d} (D^\beta f_n)(x) - E[(D^\beta f_n)(x)] = O\left(\left(\frac{2^{l(d)+2\beta}\log n}{n}\right)^{1/2}\right) + O(2^{-sl}) \quad a.s.$$

References

[1] Ango Nze, P. and Doukhan, P., *Functional estimation for time series: a general approach*, Prépublication de I'Université Paris-Sun, 1993, 93-43.

[2] Bickel, P. and Ritrov, Y., *Estimation of integrated squared density derivatives; sharp best order of convergence estimate*, Sankhya A, 50 (1988), 381-393.

[3] Bosq, D., *Optimal asymptotic quadratic error of density estimators for strong mixing or chaotic data*, Statist. Probab. Lett., to appear.

[4] Dooti, Niroumand and Afshari, *Wavelet-based estimation of derivative of a density for discrete-time stochastic process: L_p-Losses*, Journal of Science, 17 (1) (2006), 75-81.

[5] Doukhan, P., *Forme de Toeplitz associée a une analyse multi-échelle*, C.R. Acad Sci. Paris, Série 1, t. 306 (1988), 663-666.

[6] Doukhan, P. and J.R. Loen, *Une note sur la déviation quadratique d'estimators de densités par projections orthogonales*, C.R. Acad Sci. Paris, série 1, t. 310 (1990), 425-430.

[7] Doukhan, P., *Mixing Properties and Example*, Springer-Verlag, 1994, 35-39.

[8] Hall, P. and Marron, J.S., *Estimation of integrated squared density derivatives*, Statist. Prob. Lett., 6 (1987), 109-115.

[9] Kerkyacharian, G. and and Picard, D., *Density estimation in Besov spaces*, Statist. Probab. Lett., 13 (1992), 15-24.

[10] Koshkin, G.A. and Vasiliev, V.A., *Nonparametric estimation of multivariate density from dependent observation*, Mathematical Methods of Statistics, 7 (1998), 361-400.

[11] Leblance, F., *Wavelet linear density estimator for a discrete-time stochastic process: L_P-losses*, Statist. Probab. Lett., 27 (1996), 71-84.

[12] Masry, E., *Probability density estimation from dependent observation using wavelet orthonormal bases*, Statist. Probab. Lett., 21 (1994), 181-194.

[13] Masry, E., *Multivariate probability estimation by wavelet methods. Strong consistency and rates for stationary time series*, Stochastic Process. Appl., 67 (1997), 177-193.

[14] Meyer, Y., *Wavelet and operators*, Combridge Univ. Press, Cambridge, UK, 1992.

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

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

[17] Prakasa Rao. B.L.S., *Estimation of the integrated squared density derivatives by wavelets*, Bulletin of Information and Cybernetics, vol. 31, no. 1 (1999).

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

[19] Prakasa Roa. B.L.S., *Nonparametric estimation of partial derivatives of a multivariate probability density by the method of wavelets*, Asymptotic in Statistics and Probability, Festschrift for G.G. Roussas (Ed. M.L. Puri), Vsp, The Netherlands, 2000, 321-330.

[20] Singh, R.S., *Speed of convergence in nonparametric estimation of a multivariate density and its mixed partial derivatives,* J. Stat. Plan. Inf., 5 (1981), 287-298.

Accepted: 16.02.2009