Nonparametric Estimation of Partial Derivatives of a Multivariate Probability Density by the Method of Wavelets

B. L. S. Prakasa Rao
Indian Statistical Institute, New Delhi
and
Australian National University, Canberra

July 1, 1999

Abstract

A method of estimation of the partial derivatives of a multivariate probability density using wavelet systems is proposed. Rates for the almost sure convergence of these estimators are investigated.

Keywords and Phrases: Nonparametric estimation of Partial Derivatives; Multivariate Density; Wavelets; Rates of almost sure Convergence.

AMS (1991) Subject Classification: Primary 62G05.

1 Introduction

Methods of nonparametric estimation of a multivariate probability density function and regression function are extensively discussed in the literature (cf. Prakasa Rao (1983,1999)). The problem of estimation of partial derivatives of a multivariate probability density is of interest (cf. Singh (1981), Prakasa Rao (1983)) especially to detect concavity or convexity properties of the regression function. Asymptotic properties of the kernel type estimators for the partial derivatives of a multivariate probability density have been investigated earlier (cf. Prakasa Rao (1983)). Recent work in this area is due to Delecroix (1996) and Koshkin and Vasiliev (1998).

We now discuss the estimation of the partial derivatives of a multivariate probability density using the method of wavelets. Prakasa Rao (1996) studied the problem in the univariate case. We obtain rates for the almost sure convergence of wavelet based estimators for the partial derivatives of a multivariate probability density.

2 Preliminaries

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

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

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

(iii) there exists a scaling function $\phi \in V_0$ such that $\int_{\mathbb{R}^d} \phi(\mathbf{x})\, d\mathbf{x} = 1$ and

$$\{\phi(\mathbf{x}-\mathbf{k}), \mathbf{k} \in \mathbb{Z}^d\}$$

is an orthonormal basis for V_0; and for all $h \in L^2(\mathbb{R}^d)$,

(iv) for all $\mathbf{k} \in \mathbb{Z}^d$, $h(\mathbf{x}) \in V_0 \Longrightarrow h(\mathbf{x}-\mathbf{k}) \in V_0$, and

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

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

The multiresolution is said to be r-regular if $\phi \in C^{(r)}$ 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

$$\left|(D^{\boldsymbol{\beta}}\phi)(\mathbf{x})\right| \leq \frac{c_m}{(1+\|\mathbf{x}\|)^m} \text{ for all } |\boldsymbol{\beta}| \leq r \tag{2.1}$$

where

$$(D^{\boldsymbol{\beta}}\phi)(\mathbf{x}) = \frac{\partial^{\boldsymbol{\beta}}\phi(\mathbf{x})}{\partial x_1^{\beta_1}\cdots\partial x_d^{\beta_d}} \tag{2.2}$$

and

$$\boldsymbol{\beta} = (\beta_1, \dots, \beta_d), |\boldsymbol{\beta}| = \sum_{i=1}^{d} \beta_i. \tag{2.3}$$

Define W_j by the relation $V_{j+1} = V_j \oplus W_j$. The space W_j is called the detail space at level j. Let $(P_{V_j}f)(\mathbf{x})$ be the orthogonal projection of f on V_j. Then

$$(P_{V_j}f)(\mathbf{x}) = \sum_{\mathbf{k}\in\mathbb{Z}^d} a_{j\mathbf{k}}\phi_{j,\mathbf{k}}(\mathbf{x}). \tag{2.4}$$

3 Main Result

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

$$(D^{\boldsymbol{\beta}}f)(\mathbf{x}) = \frac{\partial^{\boldsymbol{\beta}}f(\mathbf{x})}{\partial x_1^{\beta_1}\cdots\partial x_d^{\beta_d}}$$

where

$$\boldsymbol{\beta} = (\beta_1, \dots, \beta_d), |\boldsymbol{\beta}| = \sum_{i=1}^{d} \beta_i \text{ and } |\boldsymbol{\beta}| \leq r.$$

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

$$(D^{\boldsymbol{\beta}}f)(\mathbf{x}) = \sum_{\mathbf{k}\in\mathbb{Z}^d} a_{l\mathbf{k}}\phi_{l,\mathbf{k}}(\mathbf{x}) \tag{3.1}$$

where

$$a_{l\mathbf{k}} = \int_{\mathbb{R}^d} (D^{\boldsymbol{\beta}}f)(\mathbf{u})\phi_{l,\boldsymbol{\beta}}(\mathbf{u})\, d\mathbf{u} \tag{3.2}$$

$$= (-1)^{\sum_{i=1}^{d}\beta_i} \int_{\mathbb{R}^d} f(\mathbf{u})(D^{\boldsymbol{\beta}}\phi_{l,\mathbf{k}})(\mathbf{u})\, d\mathbf{u}. \tag{3.3}$$

The last relation follows from (2.1). Hence

$$a_{l\mathbf{k}} = (-1)^{\sum_{i=1}^{d} \beta_i} E\left[(D^{\boldsymbol{\beta}} \phi_{l,\mathbf{k}})(\mathbf{X}) \right]. \tag{3.4}$$

Let

$$\hat{a}_{l\mathbf{k}} = \frac{(-1)^{\sum_{i=1}^{d} \beta_i}}{n} \sum_{i=1}^{n} (D^{\boldsymbol{\beta}} \phi_{l,\mathbf{k}})(\mathbf{X}_i). \tag{3.5}$$

Let us estimate $D^{\boldsymbol{\beta}} f$ by

$$(D^{\boldsymbol{\beta}} f)\hat{}_n(\mathbf{x}) = \sum_{\mathbf{k} \in \mathbb{Z}^d} \hat{a}_{l\mathbf{k}} \phi_{l,\mathbf{k}}(\mathbf{x}). \tag{3.6}$$

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

$$K(\mathbf{u}, \mathbf{v}) = \sum_{\mathbf{k} \in \mathbb{Z}^d} \phi(\mathbf{u} - \mathbf{k}) (D^{\boldsymbol{\beta}} \phi)(\mathbf{v} - \mathbf{k}). \tag{3.7}$$

Let

$$E(\mathbf{u}, \mathbf{v}) = \sum_{\mathbf{k} \in \mathbb{Z}^d} \phi(\mathbf{u} - \mathbf{k}) \phi(\mathbf{v} - \mathbf{k}). \tag{3.8}$$

Note that

$$\begin{aligned} \partial_{\mathbf{v}}^{\boldsymbol{\beta}} E(\mathbf{u}, \mathbf{v}) &= \sum_{\mathbf{k} \in \mathbb{Z}^d} \phi(\mathbf{u} - \mathbf{k}) (D^{\boldsymbol{\beta}} \phi)(\mathbf{v} - \mathbf{k}) \\ &= K(\mathbf{u}, \mathbf{v}). \end{aligned} \tag{3.9}$$

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

$$|\partial_{\mathbf{u}}^{\mathbf{v}} \partial_{\mathbf{v}}^{\mathbf{x}} K(\mathbf{u}, \mathbf{v})| \leq C_m (1 + \|\mathbf{u} - \mathbf{v}\|)^{-m} \tag{3.10}$$

for $|\boldsymbol{\alpha}| \leq r$ and $|\boldsymbol{\alpha} + \boldsymbol{\beta}| \leq r$(cf. Meyer (1992), p. 33). Choosing $\boldsymbol{\alpha} = \mathbf{0} = \boldsymbol{\gamma}$, we obtain that

$$|K(\mathbf{u}, \mathbf{v})| \leq C_m (1 + \|\mathbf{u} - \mathbf{v}\|)^{-m}, \qquad m \geq 1. \tag{3.11}$$

In particular

$$|K(\mathbf{u}, \mathbf{v})| \leq C_{d+1} (1 + \|\mathbf{u} - \mathbf{v}\|)^{-(d+1)} \tag{3.12}$$

and

$$\int_{\mathbb{R}^d} |K(\mathbf{u}, \mathbf{v})|^j \, d\mathbf{v} \leq G_j(d) \tag{3.13}$$

where

$$G_j(d) = 2\pi^{\frac{d}{2}} \frac{\Gamma(d)\Gamma(j + d(j-1))}{\Gamma(d/2)\Gamma((d+1)j)} C_{d+1}^{j} \tag{3.14}$$

Note that

$$\begin{aligned} (D^{\boldsymbol{\beta}} f_n)\hat{}(\mathbf{x}) &= \sum_{\mathbf{k} \in \mathbb{Z}^d} \phi_{l,\mathbf{k}}(\mathbf{x}) \left\{ \frac{(-1)^{\sum_{i=1}^{d} \beta_i}}{n} \sum_{i=1}^{n} (D^{\boldsymbol{\beta}} \phi_{l,\mathbf{k}})(\mathbf{X}_i) \right\} \\ &= \frac{(-1)^{|\boldsymbol{\beta}|}}{n} \sum_{i=1}^{n} \sum_{\mathbf{k} \in \mathbb{Z}^d} \phi_{l,\mathbf{k}}(\mathbf{x}) (D^{\boldsymbol{\beta}} \phi_{l,\mathbf{k}})(\mathbf{X}_i) \\ &= \frac{(-1)^{|\boldsymbol{\beta}|}}{n} \sum_{i=1}^{n} \sum_{\mathbf{k} \in \mathbb{Z}^d} 2^{ld/2} \phi(2^l \mathbf{x} - \mathbf{k}) 2^{(ld/2) + l|\boldsymbol{\beta}|} (D^{\boldsymbol{\beta}} \phi)(2^l \mathbf{X}_i - \mathbf{k}) \\ &= \frac{(-1)^{|\boldsymbol{\beta}|} 2^{ld + l|\boldsymbol{\beta}|}}{n} \sum_{i=1}^{n} K(2^l \mathbf{x}, 2^l \mathbf{X}_i). \end{aligned} \tag{3.15}$$

Hence

$$\begin{aligned} Var\left[(D^{\boldsymbol{\beta}} f_n)\hat{}(\mathbf{x})\right] &= \frac{2^{2ld+2l|\boldsymbol{\beta}|}}{n^2} n Var\left(K(2^l\mathbf{x}, 2^l\mathbf{X}_1)\right) \\ &= \frac{2^{2ld+2l|\boldsymbol{\beta}|}}{n} Var\left(K(2^l\mathbf{x}, 2^l\mathbf{X}_1)\right) \end{aligned} \tag{3.16}$$

and

$$\begin{aligned} Var\big(K(2^l\mathbf{x}, 2^l\mathbf{X}_1)\big) &= \int_{\mathbb{R}^d} K^2(2^l\mathbf{x}, 2^l\mathbf{u}) f(\mathbf{u})\, d\mathbf{u} - \left\{\int_{\mathbb{R}^d} K(2^l\mathbf{x}, 2^l\mathbf{u}) f(\mathbf{u})\, d\mathbf{u}\right\}^2 \\ &= \int_{\mathbb{R}^d} K^2(2^l\mathbf{x}, \mathbf{v}) f(2^{-l}\mathbf{v}) 2^{-ld}\, d\mathbf{v} - \left\{\int_{\mathbb{R}^d} K(2^l\mathbf{x}, \mathbf{v}) f(2^{-l}\mathbf{v}) 2^{-ld}\, d\mathbf{v}\right\}^2 \\ &\leq 2^{-ld} M_1 G_2(d) + 2^{-2ld} M_1^2 G_1^2(d) \\ &= 2^{-ld}\big(M_1 G_2(d) + 2^{-ld} M_1^2 G_1^2(d)\big). \end{aligned} \tag{3.17}$$

Relation (3.16) and (3.17) imply that

$$Var\left[(D^{\boldsymbol{\beta}} f_n)\hat{}(\mathbf{x})\right] \leq \frac{2^{ld+2l|\boldsymbol{\beta}|}}{n} M_1 G_2(d)\big(1 + o(1)\big).$$

Therefore

$$n 2^{-l(d+2|\boldsymbol{\beta}|)} Var\left[(D^{\boldsymbol{\beta}} f_n)\hat{}(\mathbf{x})\right] \leq M_1 G_2(d)\big(1 + o(1)\big).$$

Hence

$$\sup_{\mathbf{x}\in\mathbb{R}^d} Var\left[(D^{\boldsymbol{\beta}} f_n)\hat{}(\mathbf{x})\right] \leq M \frac{2^{l(d+2|\boldsymbol{\beta}|)}}{n}. \tag{3.18}$$

Let D be a compact set in $\mathbb{R}^d$ 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 centres $\mathbf{x}_j = \mathbf{x}_{nj}$ with sides of length m_n for $j = 1, \ldots, L(n)$, $m_n = \text{const}/L^{1/d}(n)$. This is possible since D is compact. Then

$$\begin{aligned} \sup_{\mathbf{x}\in D}\left|(D^{\boldsymbol{\beta}} f_n)\hat{}(\mathbf{x}) - E\left[(D^{\boldsymbol{\beta}} f_n)\hat{}(\mathbf{x})\right]\right| &= \max_{1\leq j\leq L(n)} \sup_{\mathbf{x}\in D\cap I_j} \left|(D^{\boldsymbol{\beta}} f_n)\hat{}(\mathbf{x}) - E\left[(D^{\boldsymbol{\beta}} f_n)\hat{}(\mathbf{x})\right]\right| \\ &\leq \max_{1\leq j\leq L(n)} \sup_{\mathbf{x}\in D\cap I_j} \left|(D^{\boldsymbol{\beta}} f_n)\hat{}(\mathbf{x}) - (D^{\boldsymbol{\beta}} f_n)\hat{}(\mathbf{x}_j)\right| \\ &\quad + \max_{1\leq j\leq L(n)} \left|(D^{\boldsymbol{\beta}} f_n)\hat{}(\mathbf{x}_j) - E\left[(D^{\boldsymbol{\beta}} f_n)\hat{}(\mathbf{x}_j)\right]\right| \\ &\quad + \max_{1\leq j\leq L(n)} \sup_{\mathbf{x}\in D\cap I_j} \left|E\left[(D^{\boldsymbol{\beta}} f_n)\hat{}(\mathbf{x})\right] - E\left[(D^{\boldsymbol{\beta}} f_n)\hat{}(\mathbf{x}_j)\right]\right| \\ &= T_1 + T_2 + T_3 \qquad \text{(say)}. \end{aligned} \tag{3.19}$$

Note that

$$\left|\frac{\partial K(\mathbf{u}, \mathbf{y})}{\partial u_i}\right| \leq \frac{C_2}{1 + \|\mathbf{u} - \mathbf{y}\|^2}. \tag{3.20}$$

In view of (3.10), it follows that

$$|K(\mathbf{u}, \mathbf{y}) - K(\mathbf{v}, \mathbf{y})| \leq C_2 \|\mathbf{u} - \mathbf{v}\| d^{1/2}. \tag{3.21}$$

Hence

$$\begin{aligned} \left|(D^{\boldsymbol{\beta}} f_n)\hat{}(\mathbf{x}) - (D^{\boldsymbol{\beta}} f_n)\hat{}(\mathbf{x}_j)\right| &\leq \frac{2^{ld+l|\boldsymbol{\beta}|}}{n} n C_2 \|2^l\mathbf{x} - 2^l\mathbf{x}_j\| d^{1/2} \\ &= 2^{l(d+1+|\boldsymbol{\beta}|)} C_2 d^{1/2} \|\mathbf{x} - \mathbf{x}_j\|. \end{aligned} \tag{3.22}$$

Therefore

$$\begin{aligned} T_1 &= \max_{1\le j\le L(n)} \sup_{\mathbf{x}\in D\cap I_j} \left|(D^{\boldsymbol{\beta}} f_n)\hat{}(\mathbf{x}) - (D^{\boldsymbol{\beta}} f_n)\hat{}(\mathbf{x}_j)\right| \\ &\le C_2 d^{1/2} m_n 2^{l(d+1+|\boldsymbol{\beta}|)} \\ &= O\left(\frac{2^{l(d+1+|\boldsymbol{\beta}|)}}{L^{1/d}(n)}\right) \qquad \text{a.s.} \end{aligned} \tag{3.23}$$

Relation (3.23) implies that

$$\begin{aligned} T_3 &= \max_{1\le j\le L(n)} \sup_{\mathbf{x}\in D\cap I_j} \left|E\left[(D^{\boldsymbol{\beta}} f_n)\hat{}(\mathbf{x})\right] - E\left[(D^{\boldsymbol{\beta}} f_n)\hat{}(\mathbf{x}_j)\right]\right| \\ &= O\left(\frac{2^{l(d+1+|\boldsymbol{\beta}|)}}{L^{1/d}(n)}\right). \end{aligned} \tag{3.24}$$

Now, for any $\mathbf{x}$,

$$\begin{aligned} Z_n(\mathbf{x}) &\equiv (D^{\boldsymbol{\beta}} f_n)\hat{}(\mathbf{x}) - E\left[(D^{\boldsymbol{\beta}} f_n)\hat{}(\mathbf{x})\right] \\ &= \frac{(-1)^{|\boldsymbol{\beta}|} 2^{ld+l|\boldsymbol{\beta}|}}{n} \sum_{i=1}^{n} \left\{K(2^l\mathbf{x}, 2^l\mathbf{X}_i) - E\left[K(2^l\mathbf{x}, 2^l\mathbf{X}_i)\right]\right\} \\ &= \frac{1}{n}\sum_{i=1}^{n} Y_{ni} \qquad \text{(say)} \end{aligned} \tag{3.25}$$

where

$$Y_{ni} = (-1)^{|\boldsymbol{\beta}|} 2^{ld+l|\boldsymbol{\beta}|} \left\{K(2^l\mathbf{x}, 2^l\mathbf{X}_i) - E\left[K(2^l\mathbf{x}, 2^l\mathbf{X}_i)\right]\right\}. \tag{3.26}$$

Note that relation (3.11) implies that

$$\left|K(\mathbf{u}, \mathbf{v})\right| \le C_2 \tag{3.27}$$

and hence

$$\left|Y_{ni}\right| \le 2^{ld+l|\boldsymbol{\beta}|} 2C_2, \qquad 1 \le i \le n. \tag{3.28}$$

Furthermore Y_{ni}, $1 \le i \le n$, are i.i.d. random variables for any fixed n. Note that, for any $\eta_n > 0$ and $t_n > 0$,

$$\begin{aligned} P\left(|Z_n(\mathbf{x})| > \eta_n\right) &= P\left(\left|\sum_{i=1}^{n} Y_{ni}\right| > n\eta_n\right) \\ &\le e^{-nt_n\eta_n} E\left[e^{t_n\left|\sum_{i=1}^{n} Y_{ni}\right|}\right] \\ &= e^{-nt_n\eta_n} e^{t_n n 2^{ld+l|\boldsymbol{\beta}|} 2C_2}. \end{aligned} \tag{3.29}$$

Therefore

$$P\left(\max_{1\le j\le L(n)} |Z_n(\mathbf{x}_j)| > \eta_n\right) \le L(n) e^{-nt_n\eta_n} e^{t_n n 2^{ld+l|\boldsymbol{\beta}|} 2C_2}. \tag{3.30}$$

Let

$$\eta_n = \frac{2^{l(d+1+|\boldsymbol{\beta}|)}}{L^{1/d}(n)}. \tag{3.31}$$

Then

$$P\left(\max_{1\le j\le L(n)} |Z_n(\mathbf{x}_j)| > \eta_n\right) \le L(n) e^{-\frac{nt_n 2^{l(d+1+|\boldsymbol{\beta}|)}}{L^{1/d}(n)}} e^{t_n n 2^{l(d+|\boldsymbol{\beta}|)} 2C_2}. \tag{3.32}$$

Let

$$L(n) = \left(2^{(d+2)l} n / \log n\right)^{d/2}. \tag{3.33}$$

Then

$$\begin{aligned} T_1 + T_3 &= O\left(\frac{2^{l(d+1+|\boldsymbol{\beta}|)} (\log n)^{1/2}}{2^{\frac{(d+2)}{2} l} n^{1/2}}\right) \\ &= O\left(2^{l((d/2)+|\boldsymbol{\beta}|)} \frac{(\log n)^{1/2}}{n^{1/2}}\right) \end{aligned} \tag{3.34}$$

Suppose that $l_n \to \infty$ and

$$\frac{2^{l(d+2|\boldsymbol{\beta}|)} \log n}{n} \to 0 \tag{3.35}$$

as $n \to \infty$. Following the relation (3.32), it can be checked that

$$\sum_{n=1}^{\infty} P\left(\max_{1 \le j \le L(n)} |Z_n(\mathbf{x}_j)| > \eta_n\right) \le \sum_{n=1}^{\infty} \frac{L(n)}{n^a} < \infty \tag{3.36}$$

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

$$T_2 = \max_{1 \le j \le L(n)} |Z_n(\mathbf{x}_j)| \le \eta_n \qquad \text{a.s.} \tag{3.37}$$

Combining the earlier estimates, it follows from (3.19) that

$$\begin{aligned} T_1 + T_2 + T_3 &= O\left(2^{l((d/2)+|\boldsymbol{\beta}|)} \left(\frac{\log n}{n}\right)^{1/2}\right) \\ &= O\left(\left(\frac{2^{l(d+2|\boldsymbol{\beta}|)} \log n}{n}\right)^{1/2}\right). \end{aligned} \tag{3.38}$$

Note that

$$E\left[(D^{\boldsymbol{\beta}} f_n)\hat{}(\mathbf{x})\right] = \sum_{\mathbf{k} \in \mathbb{Z}^d} a_{l\mathbf{k}} \phi_{l,\mathbf{k}}(\mathbf{x}) = P_{V_l}(D^{\boldsymbol{\beta}} f)(\mathbf{x}). \tag{3.39}$$

Hence

$$\begin{aligned} (D^{\boldsymbol{\beta}} f)(\mathbf{x}) - E\left[(D^{\boldsymbol{\beta}} f_n)\hat{}(\mathbf{x})\right] &= (D^{\boldsymbol{\beta}} f)(\mathbf{x}) - P_{V_l}(D^{\boldsymbol{\beta}} f)(\mathbf{x}) \\ &= \sum_{j \ge l} P_{W_j}(D^{\boldsymbol{\beta}} f)(\mathbf{x}). \end{aligned} \tag{3.40}$$

If $(D^{\boldsymbol{\beta}} f)(\mathbf{x}) \in \mathcal{B}_{spq}$, where $\mathcal{B}_{spq}$ is a Besov space, for some $0 < s < r$, $1 \le p, q < \infty$ with $s > d/p$ and if the multiresolution analysis is r-regular, then it follows by arguments given in Masry(1997) and Kerkyacharian and Picard(1992) that

$$\begin{aligned} \sup_{\mathbf{x} \in \mathbb{R}^d} \left| E\left[(D^{\boldsymbol{\beta}} f_n)\hat{}(\mathbf{x})\right] - (D^{\boldsymbol{\beta}} f)(\mathbf{x}) \right| &= \sup_{\mathbf{x} \in \mathbb{R}^d} \left| (D^{\boldsymbol{\beta}} f)(\mathbf{x}) - (P_{V_l} f)(\mathbf{x}) \right| \\ &\le (\text{constant}) 2^{-(s-d/p)l} J_{spq}(D^{\boldsymbol{\beta}} f) \end{aligned} \tag{3.41}$$

where

$$J_{spq}(g) = \|P_{V_0} g\|_{L^p} + \left(\sum_{j \ge 0} \left(2^{js} \|P_{W_j} g\|_{L^p}\right)^q\right)^{1/q} \tag{3.42}$$

for $g \in \mathcal{B}_{spq}$. Combining (3.38) and (3.41), we have

$$\sup_{\mathbf{x} \in D} \left| (D^{\boldsymbol{\beta}} f_n)\hat{}(\mathbf{x}) - (D^{\boldsymbol{\beta}} f)(\mathbf{x}) \right| = O\left(\left(\frac{2^{l(d+2|\boldsymbol{\beta}|)} \log n}{n}\right)^{1/2}\right) + O\left(2^{-(s-d/p)l}\right) \qquad \text{a.s.} \tag{3.43}$$

We now have the following main result of the paper.

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

$$\frac{2^{l(d+2|\boldsymbol{\beta}|)}\log n}{n} \to 0 \text{ as } n \to \infty. \tag{3.44}$$

Then there exists a constant $M > 0$ *such that*

$$\sup_{\mathbf{x}\in\mathbb{R}^d} Var\left[(D^{\boldsymbol{\beta}} f_n)\hat{}(\mathbf{x})\right] \leq M\frac{2^{l(d+2|\boldsymbol{\beta}|}\big)}{n}. \tag{3.45}$$

Furthermore, for any compact set $D \subset \mathbb{R}^d$,

$$\sup_{\mathbf{x}\in D}\left|(D^{\beta} f_n)\hat{}(\mathbf{x}) - E\left[(D^{\boldsymbol{\beta}} f_n)\hat{}(\mathbf{x})\right]\right| = O\left(\left(\frac{2^{l(d+2|\boldsymbol{\beta}|)}\log n}{n}\right)^{1/2}\right) \qquad a.s. \tag{3.46}$$

If, in addition, $(D^{\boldsymbol{\beta}} f)(\mathbf{x}) \in \mathcal{B}_{spq}$ *for some* $0 < s < r$, $1 \leq p$, $q < \infty$ *where* $s > d/p$, *then*

$$\sup_{\mathbf{x}\in D}\left|(D^{\boldsymbol{\beta}} f_n)\hat{}(\mathbf{x}) - (D^{\beta} f)(\mathbf{x})\right| = O\left(\left(\frac{2^{l(d+2|\boldsymbol{\beta}|)}\log n}{n}\right)^{1/2}\right) + O\left(2^{-(s-d/p)l}\right) \qquad a.s. \tag{3.47}$$

and if $(D^{\beta} f)(\mathbf{x}) \in \mathcal{B}_{s\infty\infty}$, $s > 0$, *then*

$$\sup_{\mathbf{x}\in D}\left|(D^{\boldsymbol{\beta}} f_n)\hat{}(\mathbf{x}) - (D^{\beta} f)(\mathbf{x})\right| = O\left(\left(\frac{2^{l(d+2|\boldsymbol{\beta}|)}\log n}{n}\right)^{1/2}\right) + O\left(2^{-sl}\right) \qquad a.s. \tag{3.48}$$

Remarks:

(1). If $\{l_n\}$ is chosen so that

$$2^{l_n} \backsimeq \left(\frac{\log n}{n}\right)^{\frac{1}{(d+2|\boldsymbol{\beta}|+2(s-d/p))}},$$

then, for every compact set $D \subset \mathbb{R}^d$,

$$\sup_{\mathbf{x}\in D}\left|(D^{\boldsymbol{\beta}} f_n)\hat{}(\mathbf{x}) - (D^{\beta} f)(\mathbf{x})\right| = O\left(\left(\frac{\log n}{n}\right)^{\frac{(s-d/p)}{(d+2|\boldsymbol{\beta}|+2(s-d/p))}}\right) \qquad \text{a.s.}$$

for $f \in \mathcal{B}_{spq}$ with $1 \leq p, q < \infty$. If $f \in \mathcal{B}_{s\infty\infty}$, then

$$\sup_{\mathbf{x}\in D}\left|(D^{\boldsymbol{\beta}} f_n)\hat{}(\mathbf{x}) - (D^{\beta} f)(\mathbf{x})\right| = O\left(\left(\frac{\log n}{n}\right)^{\frac{s}{(d+2|\boldsymbol{\beta}|+2s)}}\right) \qquad \text{a.s.}$$

(2). Masry(1997) has pointed out that, for multivariate density estimators, if the uniform strong convergence of the estimators over compact sets is chosen as the performance measure, then there is no advantage in using nonlinear threshold wavelet estimates. These comments continue to hold for the estimation of partial derivatives of multivariate densities by wavelet estimators.

Acknowledgement: The author wishes to thank Professor Peter Hall, Centre for Mathematics and its Applications, Australian National University, for his invitation to visit during which time this work was completed.

References

[1] Delecroix, M. (1996) L^2-consistency of functional parameters estimated under ergodicity assumptions, *Publ. Inst. Stat. Univ. de. Paris* **40**, 33–56.

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

[3] Koshkin, G. A. and Vasiliev, V. A. (1998) Nonparametric estimation of a multivariate density from dependent observations, *Mathematical Methods of Statistics* **7**, 361–400.

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

[5] Meyer, Y. (1992) *Wavelets and operators*, Cambridge Univ. Press, Cambridge, UK.

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

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

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

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