Multivariate probability density estimation by wavelet methods: Strong consistency and rates for stationary time series¹

Elias Masry

Department of Electrical and Computer Engineering, University of California at San Diego, La Jolla, CA 92093, USA

Received 19 June 1996; revised 6 January 1997

Abstract

The estimation of the multivariate probability density functions $f(x_1, \ldots, x_d)$, $d \geq 1$, of a stationary random process $\{X_t\}$ using wavelet methods is considered. Uniform rates of almost sure convergence over compact subsets of $\mathbb{R}^d$ for densities in the Besov space $B_{spq}$ are established for strongly mixing processes.

Keywords: Probability density estimation; Wavelet method; Besov spaces; Rates of strong convergence; Strongly mixing processes

AMS classification: primary 62G07

1. Introduction

Probability density estimation plays an important role in statistical data analysis with applications in pattern recognition and classification (Fukunaga, 1972; Fukunaga and Hostetler, 1975; Krzyzak and Pawlak, 1984; Krzyzak, 1986). There is a vast literature on non-parametric density estimation; see the books by Silverman (1986), and Scott (1992) and the references therein for available methods and results. Among the classical approaches we briefly mention the kernel estimators introduced by Rosenblatt (1956a) and Parzen (1962); the orthogonal series estimators of Cencov (1962), Schwartz (1967), Walter (1977), and Kronmal and Tartar (1968); the spline estimators of Boneva et al. (1971), Lii and Rosenblatt (1975) and Wahba (1975). It is not the purpose of this paper to survey the extensive literature on density estimation under i.i.d. or time-series setting. We only mention that for the classical kernel density

¹This work was supported by the Office of Naval Research under Grant N00014-90-J-1175.

Recently, wavelet methods were introduced for the estimation of probability density functions. The first papers by Doukhan and Leon (1990), Kerkyacharian and Picard (1992), and Walter (1992) dealt with linear wavelet estimators in i.i.d. setting and established upper bounds on convergence in the mean $L_p$ norm. Leblanc (1996) extended the work of Kerkyacharian and Picard (1992) to dependent observations again establishing upper bounds in the mean $L_p$ norm estimation error. Masry (1994) established precise mean $L_2$ norm results (rate of convergence and the value of the asymptotic constant) for densities in the Sobolev space $H^s$, $s > 0$, for dependent observations using linear wavelet density estimates. The work of Donoho et al. (1996) is the most comprehensive to date in this area: they were the first to introduce nonlinear wavelet estimates using thresholded empirical wavelet coefficients and established minimax results over a large range of densities belonging to the Besov space $B_{spq}$, $s > 0$, $1 \leq p \leq \infty$, $1 \leq q \leq \infty$ and a range of global mean $L_{p'}$ error measures with $1 \leq p' \leq \infty$. For certain values of $p$ and $p'$, linear estimates have a suboptimal rate of convergence. An important result is given in their Theorem 2 which provides a lower bound on the performance measure of any estimator of $f \in B_{spq}$. The basic assumptions made in their paper are that the density function $f \in B_{spq}$ has a compact support and that the observations are i.i.d. Hall and Patil (1995) also considered nonlinear thresholded univariate density estimates in i.i.d. setting and developed mean $L_2$ estimation’s error formulas for density functions having piecewise differentiability properties.

This paper considers the estimation of the multivariate probability density functions $f(x) = f(x_1, \ldots, x_d)$, $d \geq 1$, of the random process $\{X_i\}$ using wavelet methods. The goal is to establish the strong consistency of such estimates along with rates of convergence which are uniform over compact subsets of $\mathbb{R}^d$. The multivariate density function $f$ is assumed to belong to the Besov space $B_{spq}$; the support of $f$ is not necessarily compact and the underlying random process $\{X_i\}$ is strongly mixing. Almost sure rates of convergence are important in practice since one is normally given a single realization on the basis of which an estimate is formed. It will be seen that the uniform almost sure rates established in this paper are of the same order as the best attainable rates (in the mean) given in Theorem 2 of Donoho et al. (1996) (after adjustment for $d \geq 1$). Moreover, when $f \in B_{spq}$, $s > 0$, the a.s. rates of convergence given in this paper coincide with the best attainable rates (in probability) given by Stone (1982) for density/regression estimates of functions with $s$ derivatives and i.i.d. observations.

The organization of the paper is as follows: Basic properties of wavelets and multiresolution analysis and Besov spaces on $\mathbb{R}^d$, needed in this paper, are presented in Section 2. The main results are presented and discussed in Section 3. Derivations are given in Section 4.
2. Wavelets and Besov spaces

Following Meyer (1992) a multiresolution analysis on the Euclidean space $\mathbb{R}^d$ is a decomposition of the space $L_2(\mathbb{R}^d)$ into an increasing sequence of closed subspaces

$$\cdots V_{-2} \subset V_{-1} \subset V_0 \subset V_1 \subset V_2 \cdots$$

such that

$$f(2x) \in V_{j+1} \Leftrightarrow f(x) \in V_j \quad \text{for all } j$$

$$\bigcap_j V_j = \emptyset, \quad \bigcup_j V_j = L_2(\mathbb{R}^d).$$

$V_0$ is closed under integer translation. Finally, there exists a scale function $\phi \in L_2(\mathbb{R}^d)$ with $\int_{\mathbb{R}^d} \phi(x) \, dx = 1$ such that $\{\phi_k(x) = \phi(x - k), k \in \mathbb{Z}^d\}$ is an orthonormal basis for $V_0$. It follows that $\{\phi_j,k(x) = 2^{jd/2}\phi(2^jx - k), k \in \mathbb{Z}^d\}$ is an orthonormal basis for $V_j$.

**Definition.** The multiresolution analysis is called $r$-regular if $\phi \in C^r$ and all its partial derivatives up to total order $r$ are rapidly decreasing, i.e., for every integer $i \geq 0$, there exists a constant $A_i$ such that

$$|(D^\beta \phi)(x)| \leq \frac{A_i}{(1 + \|x\|)^i} \quad \text{for all } |\beta| \leq r, \quad (2.1)$$

where

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

and

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

Define the detail space $W_j$ by

$$V_{j+1} = V_j \oplus W_j.$$ 

Then there exist $N = 2^d - 1$ wavelet functions $\{\psi_i(x), i = 1, \ldots, N\}$ such that

(i) $\{\psi_i(x - k), k \in \mathbb{Z}^d, i = 1, \ldots, N\}$ is an orthonormal basis for $W_0$.

(ii) With $\{\psi_{i,j,k}(x) = 2^{j/2}\psi_i(2^jx - k), \psi_i(x), i = 1, \ldots, N, k \in \mathbb{Z}^d, j \in \mathbb{Z}\}$ constitute an orthonormal basis for $L_2(\mathbb{R}^d)$.

(iii) $\psi_i$ has the same regularity as $\phi$.

For any $f \in L_2(\mathbb{R}^d)$ we have the orthonormal representation

$$f(x) = \sum_{k \in \mathbb{Z}^d} a_{mk} \phi_{m,k}(x) + \sum_{j \geq m} \sum_{i=1}^N \sum_{k \in \mathbb{Z}^d} b_{ijk} \psi_{i,j,k}(x) \quad (2.4a)$$
for any integer $m$ where

$$a_{mk} = \int_{\mathbb{R}^d} f(u) \phi_{m,k}(u) \, du, \quad b_{ijk} = \int_{\mathbb{R}^d} f(u) \psi_{i,j,k}(u) \, du. \quad (2.4b)$$

Note that the orthogonal projection of $f$ on $V_l$ can be written in two equivalent ways:

$$(P_{V_l} f)(x) = \sum_{k \in \mathbb{Z}^d} a_{lk} \phi_{l,k}(x)$$

$$= \sum_{k \in \mathbb{Z}^d} a_{mk} \phi_{m,k}(x) + \sum_{j=-m}^{l} \sum_{k \in \mathbb{Z}^d} b_{ijk} \psi_{i,j,k}(x) \quad (2.5)$$

for any $m \leq l$.

There are many equivalent definitions for Besov spaces. The most usual and natural definition is in terms of the modulus of continuity (Bergh and Lofstrom, 1976; Triebel, 1983). Assume that $1 \leq p \leq \infty$, $1 \leq q \leq \infty$. Let $(S_{\tau} f)(x) = f(x - \tau)$. For $0 < s < 1$ set

$$\gamma_{spq}(f) = \left( \int_{\mathbb{R}^d} \left( \frac{\|S_{\tau} f - f\|_{L_p}}{\|\tau\|^s} \right)^q \, d\tau \right)^{1/q},$$

$$\gamma_{spq}(f) = \sup_{\tau \in \mathbb{R}^d} \|S_{\tau} f - f\|_{L_p}. \quad (2.6)$$

For $s = 1$, set

$$\gamma_{1pq}(f) = \left( \int_{\mathbb{R}^d} \left( \frac{\|S_{\tau} f + S_{-\tau} f - 2 f\|_{L_p}}{\|\tau\|^s} \right)^q \, d\tau \right)^{1/q},$$

$$\gamma_{1pq}(f) = \sup_{\tau \in \mathbb{R}^d} \|S_{\tau} f + S_{-\tau} f - 2 f\|_{L_p}. \quad (2.7)$$

For $0 < s < 1$ and $1 \leq p, q \leq \infty$, define $B_{spq} = \{f \in L_p : \gamma_{spq} < \infty\}$. For $s > 1$, put $s = \lfloor s \rfloor + \{s\}$ with $\lfloor s \rfloor$ an integer and $0 < \{s\} \leq 1$. Define $B_{spq}$ to be the space of functions in $L_p(\mathbb{R}^d)$ such that $D^j f \in B_{\lfloor s \rfloor, \{s\}, pq}$ for all $|j| \leq \lfloor s \rfloor$. The norm is defined by

$$\|f\|_{B_{spq}} = \|f\|_{L_p} + \sum_{|j| \leq \lfloor s \rfloor} \gamma_{\lfloor s \rfloor, \{s\}, pq}(D^j f).$$

Meyer (1992) provided a characterization of the space $B_{spq}$ in terms of the wavelets coefficients: Assume the multiresolution analysis is $r$-regular and $s < r$. Then $f \in B_{spq}$ if and only if

$$J_{spq}(f) = \|P_{V_0} f\|_{L_p} + \left( \sum_{j \geq 0} \left( 2^{js} \|P_{V_j} f\|_{L_p} \right)^q \right)^{1/q} < \infty$$

with the usual sup-norm modification for $q = \infty$. Using the wavelet representation (2.3) we also have that $f \in B_{spq}$ if and only if

$$J'_{spq}(f) = \|a_0\|_{L_p} + \left( \sum_{j \geq 0} \left( 2^{js+\frac{d(1/2-1/p)}{2}} \|b_j\|_{L_p} \right)^q \right)^{1/q} < \infty.$$
where \( \|b_j\|_{p} = \left( \sum_{|l|=1}^{N} \sum_{k \in \mathbb{Z}^d} |b_{j,l,k}|^p \right)^{1/p} \). The following Sobolev embedding (Triebel, 1983, p. 129) is useful in the sequel:

\[
B_{s,pq} \subset B_{s',p'q} \quad \text{for } p' \geq p, \ s' \leq s, \ s - \frac{d}{p} = s' - \frac{d}{p'}.
\]  

(2.6)

Special cases of Besov spaces include the \( L_2(\mathbb{R}^d) \) Sobolev space \( H_2 = B_{s,22} \), and the set of bounded \( s \)-Lipschitz functions \( B_{s,\infty,\infty} \).

3. Multivariate density estimation

Let \( \{X_i\} \) be a stationary random process. Set

\[
X_j = (X_{j+1}, ..., X_{j+d})
\]

(3.1)

and let \( f(x) \) be the joint probability density of the vector \( X_j \). Assume that \( f \in L_2(\mathbb{R}^d) \). Then \( f \) admits the wavelet representation (2.4a). Given \( n \) observations \( \{X_i\}_{i=1}^{n} \) we estimate the coefficients \( \{a_{mk}\} \) and \( \{b_{ijk}\} \) by

\[
\hat{a}_{mk} = \frac{1}{n} \sum_{i=1}^{n} \phi_{m,k}(X_i), \quad \hat{b}_{ijk} = \frac{1}{n} \sum_{i=1}^{n} \psi_{i,j,k}(X_i),
\]

(3.2)

and note that these estimates are unbiased,

\[
E[\hat{a}_{mk}] = a_{mk}, \quad E[\hat{b}_{ijk}] = b_{ijk}.
\]

A linear estimate of \( f \) can be obtained from (2.4a) by

\[
\hat{f}_n(x) = \sum_{k \in \mathbb{Z}^d} \hat{a}_{mk} \phi_{m,k}(x)
\]

(3.3a)

or, equivalently, as

\[
\hat{f}_n(x) = \sum_{k \in \mathbb{Z}^d} \hat{a}_{j,k} \phi_{j,k}(x) + \sum_{j=j_0}^{m} \sum_{i=1}^{N} \sum_{k \in \mathbb{Z}^d} \hat{b}_{ijk} \psi_{i,j,k}(x)
\]

(3.3b)

for any \( j_0 \leq m \). Here the resolution level \( m = m(n) \to \infty \) at a rate specified below. We assume that \( \phi \) and \( \psi_i \) have a compact support so that the summations above are finite for each fixed \( x \) (note that in this case the support of \( \phi \) and \( \psi_i \) is a monotonically increasing function of their degree of differentiability (Daubechies, 1992)). A nonlinear thresholded estimate of \( f \) (in one dimension) was introduced by Donoho et al. (1996); for arbitrary dimension \( d \geq 1 \) it takes the form

\[
\hat{f}_n(x) = \sum_{k \in \mathbb{Z}^d} \hat{a}_{j,k} \phi_{j,k}(x) + \sum_{j=j_0}^{m} \sum_{i=1}^{N} \sum_{k \in \mathbb{Z}^d} \hat{b}_{ijk} I\{ |\hat{b}_{ijk}| > \delta_{j,n} \} \psi_{i,j,k}(x)
\]

(3.4)

with properly chosen threshold \( \delta_{j,n} \).

Our goal in this paper is to establish rates of strong convergence which are uniform over compact subsets of \( \mathbb{R}^d \). From Theorem 2 of Donoho et al. (1996) it is seen in the one-dimensional case \( d = 1 \) that for any estimator \( \hat{f}_n \), we have with \( s > 1/p \) and
\[ p' \geq (1 + 2s)p, \]
\[ \left( \inf_{f_n} \sup_{f \in F_{spq}(M)} E\left( \| f_n(x) - f(x) \|_{L_p'} \right)^{1/p'} \right) \geq \text{const.} \left( \frac{\log n}{n} \right)^{\frac{1}{2}} \left( \frac{s-1/p + 1/p'}{1 + 2(s-1/p)} \right), \]  
(3.5)

where \( F_{spq}(M) \) is the space of densities with \( J_{spq}(f) \leq M \) and the infimum is taken over all estimators \( f_n \) taking values in a space containing \( F_{spq}(M) \). As \( p' \to \infty \) this rate is attained by linear estimators and thus nonlinear thresholded estimators do not improve the rate of convergence in this case. Motivated by this result, we focus our attention on multivariate linear estimators (3.3) which will be shown to have uniform almost sure convergence rates over compact sets attaining the lower bound (3.5) with \( p' = \infty \) (and adjusted for dimension \( d \geq 1 \)). Nonlinear estimates provide no improvement in this case.

Let \( \mathcal{F}^k \) be the \( \sigma \)-algebra of events generated by the random variables \( \{X_j, i \leq j \leq k\} \). The stationary process \( \{X_j\} \) is called strongly mixing (Rosenblatt, 1956b) if

\[ \sup_{A \in \mathcal{F}^0} \sup_{B \in \mathcal{F}^k} \left| P[AB] - P[A]P[B]\right| = \varepsilon(k) \to 0 \text{ as } k \to \infty. \]

Condition 1. Assume that

(i) \( f(x) < M_1 \leq \infty. \)

(ii) \( |f(x,y;j) - f(x)f(y)| \leq M_2 < \infty \) for all \( j \geq 1 \), where \( f(x,y;j) \) is the joint probability density of the vectors \( (X_0, X_j) \) which is assumed to exist.

(iii) The strongly mixing coefficient \( \varepsilon(j) \) satisfies

\[ \sum_{j=1}^{\infty} \varepsilon(j)^{1-2/v} < \infty \]  
(3.6)

for some \( v > 2 \) and \( 0 < a < 1 - 2/v. \)

Note that (3.6) is equivalent to \( \varepsilon(j) = O(1/j^a) \) for some \( c > 2. \) Also note that for \( 1 \leq j \leq d - 1, \) the components of \( X_0 \) and of \( X_j \) overlap; the joint density \( f(x,y;j) \) in Condition 1(ii) is then the density of the random variables \( (X_1, \ldots, X_{d+j}). \)

Lemma 1. Under Condition 1, there exists a constant \( M \) (which depends only on \( f \) and \( \phi \)) such that

\[ \sup_{x \in \mathbb{R}^d} \text{var}[f_n(x)] \leq M \frac{2^d m(n)}{n}. \]  
(3.7)

Theorem 1. Let Condition 1 hold and let \( m(n) \to \infty \) such that \( (2^d m(n) \log n)/n \to 0 \) as \( n \to \infty. \) Define

\[ r(n) = \left( n/(2^d m(n) \log n) \right)^{1/2}, \]  
(3.8)

and

\[ \theta(n) = \left( \frac{n^3 + 2^d 2^d (7 + 2d) m(n)}{(\log n)^{2d-1}} \right)^{1/4} \varepsilon[r(n)], \]  
(3.9a)
where $\log = \log_n$. If the mixing coefficient $\alpha(j)$ satisfies

$$
\sum_{j=1}^{\infty} \theta(j) < \infty
$$

(3.9b)

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

$$
\sup_{x \in D} |\hat{f}_n(x) - E[\hat{f}_n(x)]| = O \left( \frac{\left( \log n \right)^{2m(n)} n^{1/2}}{n} \right)
$$

almost surely.

**Remark 1.** Theorem 1 requires that the mixing coefficient $\alpha(j)$ satisfy (3.6) and (3.9). These two conditions are identical to the one assumed for establishing uniform a.s. rates of convergence, like Theorem 1, for multivariate kernel density estimators by setting $h_n = 2^{-m(n)}$ so that $h_n$ is the bandwidth of the corresponding kernel density estimator (see, for example, Tran, 1990; Masry, 1993b). We provide an explicit condition on the decay rate of $\alpha(j)$. Of the two conditions (3.6) and (3.9), the latter is the more stringent one. Assume that $h_n = (\log n)^a$ for some $0 < a < 1/d$. Consider first the case of algebraic decay for the mixing coefficient $\alpha(j) = O(1/j^\delta)$ for some $c > 0$ to be specified. Then $\theta(j)$ is of the form

$$
\theta(j) = \frac{(\log j)^{a_1}}{j^{a_2}}, \quad j \geq 2
$$

with

$$
a_1 = c(1 - ad)/2 - \frac{1}{4} [2d - 1 + ad(7 + 2d)],
$$

$$
a_2 = c(1 - ad)/2 - \frac{1}{4} [3 + 2d + ad(7 + 2d)]
$$

and $\sum_{j=2}^{\infty} \theta(j) < \infty$ provided $a_2 > 1$, i.e.

$$
\alpha(j) = O(1/j^\delta) \quad \text{for some } c > \frac{(3.5 + d)(1 + ad)}{1 - ad} \quad \text{with } 0 < a < 1/d.
$$

(3.10)

Next, consider the geometric decay case whereby $\alpha(j) = \exp(-\beta j)$ for some $\beta > 0$. Clearly, this decay is faster than any algebraic decay so that condition (3.9) is automatically satisfied. Specifically, $\theta(j)$ is now of the form

$$
\theta(j) = \frac{j^{a_1} \exp(-\beta(j/\log j)^{a_2})}{(\log j)^{a_3}}, \quad j \geq 2,
$$

(3.11)

with

$$
a_1 = \frac{1}{4} [3 + 2d + ad(7 + 2d)],
$$

$$
a_2 = (1 - ad)/2 \in (0, 0.5),
$$

$$
a_3 = \frac{1}{4} [2d - 1 + ad(7 + 2d)]
$$

and the exponential term in (3.11) guarantees the summability of $\theta(j)$ over $j \geq 2$. Thus, in the case of geometric decay of the mixing coefficient, no additional constrains are needed for the validity of Theorem 1.

We remark that Lemma 1 and Theorem 1 do not require the density $f$ to belong to the Besov space $B_{spq}$ nor do they require the multiresolution analysis to be $r$-regular.
(it is only required that \( f \) is bounded and that \( \phi \) satisfies (2.1) for \( r = 1 \)). On the other hand, the bias of \( \hat{f}_n \) requires Condition 2.

**Condition 2.** (i) Assume that the multiresolution analysis is \( r \)-regular.

(ii) The density \( f \in B_{spq} \) for some \( 0 < s < r, 1 \leq p, q \leq \infty \).

**Lemma 2.** Under Condition 2 we have for \( s > d/p \),

\[
\sup_{x \in \mathbb{R}^d} |E[\hat{f}_n(x)] - f(x)| = \sup_{x \in \mathbb{R}^d} |f(x) - (P_{x-}\hat{f})(x)| \\
\leq \text{const.} \, 2^{-s(d/p)\min(n)} J_{spq}(f).
\]

**Remark 2.** For the special case of \( f \in H^s_2 = B_{s,2} \) the uniform rate of approximation given in Lemma 2 coincides with that given by Kon and Raphael (1993).

Combining Theorem 1 and Lemma 2 we obtain our principal result.

**Theorem 2.** Let Conditions 1 and 2 hold and the mixing coefficient \( \alpha(j) \) satisfies (3.9). If \( s > d/p \) and the resolution level \( 2^{-m(n)} \) satisfies

\[
2^{-m(n)} \sim \left( \frac{\log n}{n} \right)^{1/(d+2(s-d/p))},
\]

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

\[
\sup_{x \in D} |\hat{f}_n(x) - f(x)| = O \left( \left( \frac{\log n}{n} \right)^{(s-d/p)/(d+2(s-d/p))} \right) \text{ almost surely.}
\]

**Corollary 1.** Under the assumptions of Theorem 2 with \( f \in B_{s,\infty,\infty}, s > 0 \), we have on every compact subset \( D \subset \mathbb{R}^d \),

\[
\sup_{x \in D} |\hat{f}_n(x) - f(x)| = O \left( \left( \frac{\log n}{n} \right)^{s(d+2s)} \right) \text{ almost surely.}
\]

**Remark 3.** For one-dimensional densities \( (d = 1) \) the uniform almost sure rate of convergence furnished by Theorem 2 for linear wavelet density estimators attains the lower bound (3.5) (in the mean) for any density estimator (with \( p' = \infty \)). Moreover, the rate given in Corollary 1 coincides with the best attainable rate of convergence (in probability) established by Stone (1982) for density/regression functions with integer-valued \( s \) bounded and continuous derivatives in i.i.d. setting. Thus when the performance measure is uniform strong convergence over compact sets, linear wavelet estimates (3.3) achieve the best rate possible for densities in the Besov space \( B_{spq} \) and there is no advantage in using nonlinear thresholded estimates (3.4).

**Remark 4.** In view of Remark 1, it is seen that when the mixing coefficient \( \alpha(j) \) has geometric decay \( \alpha(j) = \exp(-\beta j) \) for some \( \beta > 0 \), no additional conditions on the mixing coefficient are needed for the validity of Theorem 2. On the other hand, when
the mixing coefficient decay algebraically, \( \alpha(j) = O(1/j^c) \), then Theorem 2 holds provided that

\[
\alpha(j) = O(1/j^{c'}) \quad \text{for some} \quad c > (3.5 + d) \left(1 + \frac{d}{s'}\right)
\]

(3.13)

with \( s' = s - d/p > 0 \). This is obtained by substituting \( a = \frac{1}{(d + 2(s - d/p))} \) in (3.10).

Note that in this case there is a tradeoff between the rate of decay of the mixing coefficient and the smoothness parameter \( s \) of the density \( f \in B_{s,p,q} \): for a smoother density \( s \) large the required rate of decay of the mixing coefficient is slower. We remark that the condition (3.13) is compatible with that given in Tran (1990) for kernel density estimators under a standard smoothness assumption. At the same time (3.13) is very close to the condition stated in Theorem 1 of Ango Nze and Rios (1995) for kernel density estimators under the smoothness assumption that \( f \in C^s \) and its partial derivatives of order \( s \) satisfy a Hölder’s condition of degree \( \rho - s \):

\[
\alpha(j) = O(1/j^{c'}) \quad \text{for some} \quad c > (3 + d) \left(1 + \frac{d}{1 + 5}\right)
\]

(3.14)

It is seen that (3.13) and (3.14) differ only in a minor way with corresponding factors \( (3.5 + d) \) and \( (3 + d) \). The cause for this slight discrepancy is difficult to determine since the proof of Theorem 1 in Ango Nze and Rios (1995) is only sketched in few lines.

4. Derivations

Proof of Lemma 1. Define the kernel \( K(u, v) \) by

\[
K(u, v) \triangleq \sum_{k \in \mathbb{Z}^d} \phi(u - k) \phi(v - k).
\]

(4.1)

Since

\[
\left| \phi(x) \right| \leq \frac{A_{d+1}}{(1 + \|x\|^{d+1})}
\]

the kernel \( K \) defined in (4.1) converges uniformly and satisfies (Meyer, 1992, p. 33)

\[
\left| K(u, v) \right| \leq \frac{C_{d+1}}{(1 + \|u - v\|)^{d+1}}
\]

(4.2)

for some constant \( C_{d+1} \). It follows from (4.2) that for any \( j \geq 1 \)

\[
\int_{\mathbb{R}^d} |K(u, v)|^j \, dv \leq G_j(d).
\]

(4.3)

where

\[
G_j(d) = 2\pi^{d/2} \frac{\Gamma(d)\Gamma(j + d(j - 1))}{\Gamma(d/2)\Gamma((d + 1)j)} C_{d+1}^j
\]
and \( f(x) \) is the Gamma function. Then by (3.2), (3.3), and (4.1) we can write \( \hat{f}_n(x) \) as an extended kernel estimator

\[
\hat{f}_n(x) = \frac{1}{nh_n^d} \sum_{i=1}^{n} K\left(\frac{x}{h_n}, h_n \right), \quad h_n = 2^{-m(n)}.
\]

(4.4)

By stationarity,

\[
\begin{align*}
\frac{1}{h_n^d} \text{var}[\hat{f}_n(x)] & = \frac{1}{h_n^d} \left\{ \text{var} \left[ K\left(\frac{x}{h_n}, h_n \right) \right] + 2 \sum_{j=1}^{n-1} \left( 1 - \frac{j}{n} \right) \right. \\
& \quad \times \text{cov} \left\{ K\left(\frac{x}{h_n}, h_n \right), K\left(\frac{x}{h_n}, h_n \right) \right\} \\
& = I_1(x) + I_2(x).
\end{align*}
\]

(4.5)

Now,

\[
I_1(x) = \int_{\mathbb{R}^d} K^2\left(\frac{x}{h_n}, u\right)f(h_n u) \, du - h_n^d \left( \int_{\mathbb{R}^d} K\left(\frac{x}{h_n}, u\right)f(h_n u) \, du \right)^2
\]

and by Condition 1(i) and (4.3)

\[
I_1(x) \leq M_1 G_2(d) + M_1^2 G_2^2(d) h_n^d.
\]

Thus,

\[
\sup_{x \in \mathbb{R}^d} I_1(x) \leq M \left( 1 + O(h_n^d) \right); \quad M \triangleq M_1 G_2(d).
\]

(4.6)

For \( I_2(x) \), let \( \pi_n \to \infty \) such that \( \pi_n h_n^d \to 0 \) as \( n \to \infty \). Write

\[
|I_2(x)| \leq \frac{1}{h_n^d} \sum_{j=1}^{d-1} \left| \text{cov} \left\{ K\left(\frac{x}{h_n}, h_n \right), K\left(\frac{x}{h_n}, h_n \right) \right\} \right|
\]

\[
+ \frac{1}{h_n^d} \sum_{j=d}^{\pi_n} \left| \text{cov} \left\{ K\left(\frac{x}{h_n}, h_n \right), K\left(\frac{x}{h_n}, h_n \right) \right\} \right|
\]

\[
+ \frac{1}{h_n^d} \sum_{j=\pi_n+1}^{n-1} \left| \text{cov} \left\{ K\left(\frac{x}{h_n}, h_n \right), K\left(\frac{x}{h_n}, h_n \right) \right\} \right|
\]

\[
= I_{21}(x) + I_{22}(x) + I_{23}(x).
\]

(4.7)

For \( I_{21} \) there is an overlap between the components of \( X_0 \) and \( X_j \). Let \( f(u', u'', u''') \) be the joint density of the \( d + j \) distinct random variables in \( (X_0, X_j) \), where \( u', u'' \), and \( u''' \) are of dimensions \( j, d - j, \) and \( j \), respectively. Then

\[
\int_{\mathbb{R}^{d+j}} \left| \text{cov} \left\{ K\left(\frac{x}{h_n}, h_n \right), K\left(\frac{x}{h_n}, h_n \right) \right\} \right|
\]

\[
= h_n^d \int_{\mathbb{R}^{d+j}} \left| K\left(\frac{x}{h_n}, h_n \right)K\left(\frac{x}{h_n}, h_n \right) \right| \left[ f(h_n u', h_n u'', h_n u'''); j \right]
\]

\[
- f(h_n u', h_n u'')f(h_n u'', h_n u''') \left( u' \, du'' \, du''' \right)
\]

\[
\leq M_2 h_n^d \int_{\mathbb{R}^{d+j}} \left| K\left(\frac{x}{h_n}, h_n \right)K\left(\frac{x}{h_n}, h_n \right) \right| \left( u' \, du'' \, du''' \right)
\]

\[
= M_2 h_n^d \int_{\mathbb{R}^{d+j}} \left| K\left(\frac{x}{h_n}, h_n \right)K\left(\frac{x}{h_n}, h_n \right) \right| \left( u' \, du'' \, du''' \right)
\]
by Condition 1(ii). Using (4.2) and changing variables of integration, we find
\[ \leq M_2 C_{d+1}^2 h_n^{d+1} \int_{\mathbb{R}^{d+1}} \frac{du' \, du'' \, du'''}{(1 + \|u' + u''\|^2)^{d+1}} \leq \text{const}. h_n^d. \]

Thus,
\[ \sup_{x \in \mathbb{R}^d} I_{21}(x) \leq \text{const.} \sum_{j=1}^{d-1} h_n^j = O(h_n) \rightarrow 0 \quad \text{as } n \rightarrow \infty. \quad (4.8) \]

For \( I_{22} \), there is no overlap between the components of \( X_0 \) and \( X_j \) so that by Condition 2(ii) and (4.5) we have for \( 1 \leq j \leq \pi_n \)
\[ \frac{1}{h_n^d} \left| \text{cov} \left\{ \frac{x}{h_n^d}, \frac{x}{h_n^d} \right\} \right| \leq M_2 h_n^d \left( \int_{\mathbb{R}^d} \left| K \left( \frac{x}{h_n^d}, u \right) \right| du \right)^2 \leq M_2 G_1^2(d) h_n^d. \]

Hence,
\[ \sup_{x \in \mathbb{R}^d} I_{21}(x) \leq M_2 G_1^2(d) \pi_n h_n^d = o(1). \quad (4.9) \]

For \( I_{23}(x) \) we use Davydov’s lemma [Hall and Heyde (1980), Corollary A.2] to conclude that for \( j > \pi_n \)
\[ \frac{1}{h_n^d} \left| \text{cov} \left\{ \frac{x}{h_n^d}, \frac{x}{h_n^d} \right\} \right| \leq \frac{8}{h_n^d} [x(j - d + 1)]^{1 - 2/v} \left( E \left| K \left( \frac{x}{h_n^d}, \frac{x}{h_n^d} \right) \right| \right)^{2/v}. \]

Now by Condition 1(i) and (4.3)
\[ E \left| K \left( \frac{x}{h_n^d}, \frac{x}{h_n^d} \right) \right| \leq M_1 h_n^d \left( \int_{\mathbb{R}^d} \left| K \left( \frac{x}{h_n^d}, u \right) \right| du \right)^{v} \leq M_1 G_1(d) h_n^d. \]

Hence,
\[ \sup_{x \in \mathbb{R}^d} I_{23}(x) \leq \frac{\text{const.}}{h_n^d(1 - 2/v)} \sum_{j=\pi_n + 1}^{\infty} [x(j)]^{1 - 2/v} \leq \frac{\text{const.}}{h_n^d(1 - 2/v)} \pi_n^d \sum_{j=\pi_n + 1}^{\infty} j^a [x(j)]^{1 - 2/v}. \]

Select \( \pi_n = h_n^{(1 - 2/v)/a} \). Then \( \pi_n h_n^d \rightarrow 0 \) as required and by Condition 1(iii)
\[ \sup_{x \in \mathbb{R}^d} I_{23}(x) = o(1). \quad (4.10) \]

It follows by (4.7)–(4.10) that
\[ \sup_{x \in \mathbb{R}^d} I_2(x) = o(1). \quad (4.11) \]

and by (4.5), (4.6), and (4.11),
\[ nh_n^d \text{var} \left[ \hat{f}_n(x) \right] \leq M(1 + o(1)). \]

Proof of Theorem 1. Set
\[ L(n) = (2^{(d + 2)m(n)} n/\log n)^{d/2}. \]
Since $D$ is compact, it can be covered by a finite number $L = L(n)$ of cubes $I_j = I_{n,j}$ with centers $x_j = x_{n,j}$ having sides of length $\ell_n$ for $j = 1, \ldots, L(n)$. Clearly, $\ell_n = \text{const.}/(L^{1/4}(n))$ since $D$ is compact. Write
\[
\sup_{x \in D} |\hat{f}_n(x) - E[\hat{f}_n(x)]| = \max_{1 \leq j \leq L(n)} \sup_{x \in D \setminus I_j} |\bar{f}_n(x) - E[\bar{f}_n(x)]|
\]
\[
\leq \max_{1 \leq j \leq L(n)} \sup_{x \in D \setminus I_j} |\hat{f}_n(x) - \hat{f}_n(x_j)|
+ \max_{1 \leq j \leq L(n)} |\hat{f}_n(x_j) - E[\hat{f}_n(x_j)]|
+ \max_{1 \leq j \leq L(n)} \sup_{x \in D \setminus I_j} |E[\hat{f}_n(x)] - E[\hat{f}_n(x)]|
\]
\[
= Q_1 + Q_2 + Q_3. \tag{4.12}
\]

Since $\phi$ satisfies $(2.1)$ for $|\beta| = 1$ we have (Meyer, 1992, p. 33)
\[
\frac{|\partial K(u, y)|}{\partial u_i} \leq \frac{C_2}{(1 + \|u - y\|)^2} \leq C_2, \quad i = 1, \ldots, d. \tag{4.13a}
\]

It follows that
\[
|K(u, y) - K(v, y)| \leq C_2 \sum_{i=1}^{d} |u_i - v_i| \leq d^{1/2}C_2 \|u - v\|. \tag{4.13b}
\]

Thus by $(4.2)$ and $(4.13b)$
\[
|\bar{f}_n(x) - \bar{f}_n(x_j)| \leq \frac{d^{1/2}C_2}{h_n^{d+1}} \|x - x_j\|,
\]
so that
\[
Q_1 \leq \frac{d^{1/2}C_2\ell_n}{h_n^{d+1}} = \frac{\text{const.}}{L^{1/4}(n)h_n^{d+1}} = O \left( \frac{\log n}{n h_n^d} \right)^{1/2} \text{ almost surely.} \tag{4.14}
\]

From (4.14) we find immediately that
\[
Q_3 = O \left[ \left( \frac{\log n}{n h_n^d} \right)^{1/2} \right]. \tag{4.15}
\]

The main task is to show that
\[
Q_2 = O \left[ \left( \frac{\log n}{n h_n^d} \right)^{1/2} \right] \text{ almost surely.}
\]

Write
\[
W_n \equiv \hat{f}_n(x) - E[\hat{f}_n(x)] = \frac{1}{n} \sum_{i=1}^{n} Y_{n,i}. \tag{4.16}
\]

where
\[
Y_{n,i} = \frac{1}{h_n^d} \left[ K \left( \frac{x_i}{h_n} \right) - K \left( \frac{x_i}{h_n} \right) \right]. \tag{4.17}
\]
Partition the set \{1, \ldots, n\} into \(2k = 2k(n)\) consecutive blocks of size \(r(n)\); \(n = 2k(n)r(n) + v(n)\) with \(0 \leq v(n) < r(n)\). Write

\[ V_n(j) = \frac{1}{n} \sum_{i=(j-1)r+1}^{jr} Y_{n,i}, \quad j = 1, \ldots, 2k \]  

and

\[ W'_n = \sum_{j=1}^{k} V_n(2j-1), \quad W''_n = \sum_{j=1}^{k} V_n(2j), \quad W'''_n = \sum_{i=2kr+1}^{k} Y_{n,i}, \]  

so that \(W'_n\) and \(W''_n\) are the sums of the odd-numbered and even-numbered blocks, respectively. The contribution of the remainder term \(W'''_n\) is negligible (and is subsequently ignored) since it consists of at most \(r(n)\) terms whereas \(W'_n\) and \(W''_n\) each consists of \(k(n)r(n)\) terms and \(k(n) \to \infty\) as \(n \to \infty\). Then for each \(\eta > 0\),

\[ P[Q_2 > \eta] \leq P \left[ \max_{1 \leq j \leq L(n)} |W'_n(x_j)| > \eta/2 \right] + P \left[ \max_{1 \leq j \leq L(n)} |W''_n(x_j)| > \eta/2 \right] \]

\[ \leq 2L(n) \sup_{x \in [0,1]} P \left[ |W'_n(x)| > \eta/2 \right]. \]  

We proceed to bound the expression \(P[|W'_n(x)| > \eta/2]\). We use the strong approximation theorem of Bradley (1983) to approximate the random variables \(V_n(1), V_n(3), \ldots, V_n(2k-1)\) by independent random variables. By enlarging the probability space if necessary, introduce a sequence \((U_1, U_2, \ldots)\) of independent uniform \([0, 1]\) random variables that is independent of \(\{V_n(2j-1)\}_{j=1}^{k}\). Define \(V_n^*(0) = 0, V_n^*(1) = V_n(1)\). Then for each \(j \geq 2\), there exists a random variable \(V_n^*(2j-1)\) which is a measurable function of \(V_n(1), V_n(3), \ldots, V_n(2j-3), U_j\) such that \(V_n^*(2j-1)\) is independent of \(V_n(1), V_n(3), \ldots, V_n(2j-3)\), has the same distribution as \(V_n(2j-1)\), and satisfies

\[ P[|V_n^*(2j-1) - V_n(2j-1)| > \mu] \]

\[ \leq 18(\|V_n(2j-1)\|_\infty/\mu)^{1/2} \sup |P[AB] - P[A]P[B]|. \]  

where the supremum is over all sets \(A, B\) with \(A, B\) in the \(\sigma\)-algebras of events generated by \(\{V_n(1), V_n(3), \ldots, V_n(2j-3)\}\) and \(V_n(2j-1)\), respectively. Here \(\mu\) is any positive numbers such that \(0 < \mu \leq \|V_n(2j-1)\|_\infty < \infty\). Now,

\[ P \left[ |W'_n(x)| > \frac{\eta}{2} \right] \leq P \left[ \sum_{j=1}^{k} V_n^*(2j-1) > \frac{\eta}{4} \right] \]

\[ + P \left[ \sum_{j=1}^{k} V_n(2j-1) - V_n^*(2j-1) > \frac{\eta}{4} \right] = J_1(x) + J_2(x). \]  

We bound \(J_1\) as follows. By (4.13a), \(|K(u, v)| \leq C_2\) so that by (4.17),

\[ |Y_{n,i}| \leq \frac{2C_2}{h^2}, \quad i = 1, \ldots, n. \]
Moreover by (4.18),
\[ |V_n(j)| \leq \frac{2C_2 r(n)}{n h_n^d}. \] (4.25)

Define
\[ \lambda_n = \frac{1}{4C_2} [n h_n^d \log n]^{1/2}, \] (4.26)

then
\[ \lambda_n |V_n(j)| \leq \frac{1}{2}, \quad j = 1, \ldots, 2k. \] (4.27)

Because \( e^x \leq 1 + x + x^2 \) for \( |x| \leq 1/2 \) and \( V_n^*(2j - 1) \) has the same distribution as \( V_n(2j - 1) \), it follows by (4.27) that \( \lambda_n |V_n^*(2j - 1)| \leq 1/2 \) so that
\[ e^{\lambda_n V_n^*(2j - 1)} \leq 1 + \lambda_n V_n^*(2j - 1) + \lambda_n^2 [V_n^*(2j - 1)]^2. \]

Hence,
\[ E \left[ e^{\lambda_n V_n^*(2j - 1)} \right] \leq 1 + \lambda_n^2 E[V_n^*(2j - 1)]^2 \leq e^{\lambda_n^2 E[V_n^*(2j - 1)]^2}. \] (4.28)

Now by (4.23), Markov inequality, and the independence of the \( \{V_n^*(2j - 1)\}_{j=1}^k \),
\[ J_1 \leq \frac{E \left[ e^{\lambda_n \sum_{j=1}^k V_n^*(2j - 1)} \right] + E \left[ e^{-\lambda_n \sum_{j=1}^k V_n^*(2j - 1)} \right]}{e^{\lambda_n n/4}} \]
\[ \leq 2e^{-\lambda_n n/4} \left\{ e^{\lambda_n^2 \sum_{j=1}^k E[V_n^*(2j - 1)]^2} \right\} \] (4.29)
by (4.28). We obtain an upper bound on \( E[V_n^*(2j - 1)]^2 \):
\[ \sum_{j=1}^k E[V_n^*(2j - 1)]^2 = \sum_{j=1}^k E[V_n(2j - 1)]^2 = \frac{1}{n^2} \sum_{j=1}^k \left\{ \sum_{i=2(j-1)+1}^{2j-1} Y_{n,i} \right\}^2 \]
\[ \leq \frac{1}{n^2} \left\{ \sum_{i=1}^n \text{var}(Y_{n,i}) + \sum_{i \neq j} \sum_{i=1}^n \text{cov}(Y_{n,i}, Y_{n,j}) \right\} \leq \frac{M}{nh_n^2} (1 + o(1)) \] (4.30)
by Lemma 1. Thus by (4.29) and (4.30),
\[ \sup_{x \in \mathbb{R}_+} J_1(x) \leq 2 \exp \left\{ -\lambda_n \eta + \frac{\lambda_n^2 M}{n h_n^d} \right\}. \] (4.31)

We now bound the term \( J_2 \) on the right-hand side of (4.23):
\[ J_2 \leq \sum_{j=1}^k P \left[ \left| V_n(2j - 1) - V_n^*(2j - 1) \right| > \frac{\eta}{4k} \right] \] (4.32)
We make use of (4.22): (i) If \( \|V_n(2j - 1)\|_\infty > \eta/(4k) \), then by (4.22)
\[ J_2 \leq 18k \left\{ \left( \frac{\|V_n(2j - 1)\|_\infty}{\eta/4k} \right)^{1/2} \right\} \sup_{A} |P[A B] - P[A] P[B]|. \] (4.33)
where \( A \in \sigma(V_n(1), V_n(3), \ldots, V_n(2j - 3)) \), \( B \in \sigma(V_n(2j - 1)) \). By (4.17) and (4.18) we have
\[ |P[AB] - P[A] P[B]| \leq \alpha [r(n) - d + 1]. \]
By (4.25) we then have
\[
\sup_{x \in \mathbb{R}^d} J_2(x) \leq \text{const. } k(n) (k(n)/\eta)^{1/2} \left\{ \frac{r(n)}{n h_n^d} \right\}^{1/2} \left[ r(n) - d + 1 \right].
\] (4.34i)

(ii) If \( \| V_n(2j - 1) \|_\infty < \eta/(4k) \) then by (4.32) and (4.22) with \( \mu \equiv \| V_n(2j - 1) \|_\infty \) we have
\[
\sup_{x \in \mathbb{R}^d} J_2(x) \leq \text{const. } k(n) \left[ \eta(r(n) - d + 1) \right],
\] (4.34ii)
which is of smaller order than (4.34i). Thus by (4.23), (4.31), and (4.34),
\[
\sup_{x \in \mathbb{R}^d} P\left[ |W_n(x)| > \eta/2 \right] \leq 2 \exp \left\{ - \frac{\lambda_n \eta}{4} + \frac{\lambda_n^2 M}{n h_n^d} \right\}
\]
and similarly for \( P[|W_n'| > \eta/2] \). Then by (4.21) we have with \( \eta = \eta_n = M_3 \left[(\log n)/(n h_n^d)\right]^{1/2} \) that
\[
P\left[ \max_{1 \leq j \leq L(n)} |\tilde{f}_n(x_j) - E[\tilde{f}_n(x_j)]| > \eta_n \right] \leq 4L(n)
\]
\[
\times \exp \left\{ - \frac{\lambda_n \eta_n}{4} + \frac{M \lambda_n^2}{n h_n^d} \right\} + \text{const. } \theta(n).
\] (4.35)

Now by (4.26)
\[
\frac{\lambda_n \eta_n}{4} = \frac{M}{16C_2} \log n, \quad \frac{\lambda_n^2 M}{n h_n^d} = \frac{M}{16C_2} \log n.
\]
The first term on the right-hand side of (4.35) is therefore of the form
\[
\frac{4L(n)}{n^a}; \quad a = \frac{1}{16C_2} \left( M_3 - \frac{M}{C_2} \right),
\]
and by selecting \( M_3 \) large enough we can ensure that \( L(n)/n^a \) is summable. Because \( \{\theta(n)\} \) is summable by assumption, it follows by (4.35) and the Borel–Cantelli lemma that
\[
Q_2 = O(\eta_n) = O\left[ \left( \frac{\log n}{n h_n^d} \right)^{1/2} \right] \text{ almost surely}
\] (4.36)
and Theorem 1 now follows from (4.12), (4.14), (4.15), and (4.36).

**Proof of Lemma 2.** We have
\[
E[\tilde{f}_n(x)] = \sum_{k \in \mathbb{Z}^d} a_{mk} \phi_{mk}(x) = (P_{V_n} f)(x).
\]
Thus,

$$f(x) - E [\hat{f}_n(x)] = f(x) - (P_{V_n} f)(x) = \sum_{j \geq m} (P_{W_j} f)(x).$$

As in Kerkyacharian and Picard (1992) we have

$$\| f - E [\hat{f}_n] \|_{L_p} \leq \text{const.} \left( \sum_{j \geq m} (\| P_{W_j} f \|_{L_p})^q \right)^{1/q}$$

$$\leq \text{const.} \ 2^{-m(n)s'} \left( \sum_{j \geq m} (2^{js'} \| P_{W_j} f \|_{L_p})^q \right)^{1/q}$$

and since $f \in B_{spq}$ it follows by the embedding (2.6)

$$\| f - E [\hat{f}_n] \|_{L_p} \leq \text{const.} \ 2^{-m(n)s'} \left( \sum_{j \geq m} (2^{js'} \| P_{W_j} f \|_{L_p})^q \right)^{1/q}$$

$$\leq \text{const.} \ 2^{-m(n)s'} J_{spq}(f),$$

where $s' = s - d(1/p - 1/p')$ and the result follows with $p' = \infty$. □

References


