ACCEPTED MANUSCRIPT

Pointwise and uniform convergence of multivariate kernel density estimators using random bandwidths

Santanu Dutta * Koushik Saha †

Abstract

We obtain the rates of pointwise and uniform convergence of multivariate kernel density estimators using a random bandwidth vector obtained by some data based algorithm. We are able to obtain faster rate for pointwise convergence. The uniform convergence rate is obtained under some moment condition on the marginal distribution. The rates are obtained under i.i.d. and strongly mixing type dependence assumptions.

keywords. multivariate kernel density estimator, random bandwidth.

*Mathematical Science Dept., Tezpur University, Napaam, Tezpur, Assam 784028, India, e-mail: tezpur1976@gmail.com.

†Department of Mathematics, IIT Bombay, Mumbai 400076, India, e-mail:koushiksaha877@gmail.com

ACCEPTED MANUSCRIPT

1 Introduction

Let $X_1, \ldots, X_n$ be $\mathbb{R}^d$ valued random variables with density f. A kernel density estimator (KDE) with bandwidth vector $h = (h_1, \ldots, h_d)$ is defined as

$$\hat{f}_{n,h}(x) = \frac{1}{n\prod_{j=1}^{d} h_j} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right) = \frac{1}{n}\sum_{i=1}^{n} g_{x-X_i}(h), \tag{1.1}$$

where $x = (x_1, \ldots, x_d) \in \mathbb{R}^d$ and K is a density function on $\mathbb{R}^d$, $X_i = (X_{i1}, X_{i2}, \ldots, X_{id})$ and

$$g_z(h) = \frac{1}{\prod_{i=1}^{d} h_i} K\left(\frac{z}{h}\right) = \frac{1}{\prod_{i=1}^{d} h_i} K\left(\frac{z_1}{h_1}, \ \ldots, \frac{z_d}{h_d}\right), \text{ for } z = (z_1, z_2, \ldots, z_d) \in \mathbb{R}^d.$$

From Scott (1992) we see that under i.i.d. assumption the optimal values of $h_1, \ldots, h_d$ in terms of minimizing the MISE are of the order $O(n^{-1/(4+d)})$.

The asymptotic properties of the KDE has been widely studied for nonrandom h. See for instance Rao (1983) for a detailed review. However, a practical problem in density estimation is to determine the appropriate values of $h_1, h_2, \ldots, h_d$ based on $X_1, X_2, \ldots, X_n$. Sain et al. (1994) discussed a number of algorithms, such as the least squares or the unbiased cross validation (UCV), the biased cross validation (BCV) and the smooth bootstrap bandwidth selector (SCV) for data based choice of h. Duong and Hazelton (2005) and Dutta (2011) have proposed other methods of choosing the bandwidth vector h based on the given data. Let $\hat{h} = (\hat{h}_1, \ldots, \hat{h}_d)$ be a bandwidth vector obtained by any one of these bandwidth selectors. Such a $\hat{h}$ is a random vector, i.e. a $\mathbb{R}^d$ valued function of $X_1, \ldots, X_n$. A natural problem is to investigate the asymptotic properties of the resulting estimator $\hat{f}_{n,\hat{h}}$. Obtaining properties of such an estimator can be much more challenging than the estimator based on non random h. For instance, it is very difficult to compute the expectation of $E[\hat{f}_{n,\hat{h}}]$ and hence the bias.

For univariate data, Dutta and Goswami (2013) have studied the convergence of the KDE using a random bandwidth. Far less seems to be known for multivariate KDE using a random bandwidth vector. Under the assumption that $h_1 = h_2 = \cdots = h_d$, Tenreiro (2001) established the convergence

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

rates of the L_2 error of these estimators for a wide range of random bandwidths (we are thankful to the reviewer for this reference). Dutta and Saha (2013) studied the L_1 consistency of the multivariate KDE for a broad class of random bandwidth vectors. Under similar assumptions on the bandwidth vector $\hat{h}$, in this paper we study the pointwise and sup-norm convergence of such estimators with increase in sample size. In particular we obtain the rates at which $P(|\hat{f}_{n,\hat{h}}(x) - f(x)| > \epsilon)$ and $P(||\hat{f}_{n,\hat{h}} - f|| > \epsilon)$ converges to zero, where $\epsilon > 0$. These rates extend the results in Dutta and Goswami (2013) for multivariate KDE and hold for a wide variety of bandwidth selectors discussed in Sain et al. (1994), Duong and Hazelton (2005), Dutta and Saha (2013).

1.1 Notation and Assumptions.

Let us define some notation which are used in the sequel. For any function g of d variables, let $\partial_j g$ denote the partial derivative of g with respect to the jth variable, $j = 1, \ldots, d$. For $x \in \mathbb{R}^d$, $||x||$ denotes its Eucledian norm and for any real valued function f, $||f||$ denotes its sup-norm. For any event A, $I(A)$ denotes a binary valued random variable such that $I(A)(w) = 1$ for $w \in A$ and $I(A)(w) = 0$ otherwise.

Let

$$I_n = \prod_{j=1}^{d} [a_j n^{-1/(4+d)}, b_j n^{-1/(4+d)}], \tag{1.2}$$

where $a_j < b_j$, $j = 1, \ldots, d$, are arbitrary positive constants. Let us state the assumptions on the kernel K and density f.

Assumption 1: The kernel K is a differentiable density in $\mathbb{R}^d$ satisfying the following conditions

(i)

$$||K||,\ ||\partial_j K||,\ \sup_{x \in \mathbb{R}^d} \left| x_j \partial_j K(x) \right| < \infty \text{ for } j = 1, \ldots, d,$$

ACCEPTED MANUSCRIPT

(ii)

$$\sup_{x\in\mathbb{R}^d} ||x||^d K(x) < \infty \text{ where } x = (x_1, \ldots, x_d).$$

Assumption 2: $||f|| < \infty$ and $||\partial_i f|| < \infty$ for $i = 1, \ldots, d$.

Assumption 3: There exists $\gamma > 0$ such that $\int_{\mathbb{R}^d} ||x||^\gamma f(x)dx < \infty$.

Assumption 4: f_j, jth marginal density, is strictly positive on some open neighbourhood of Q_{1j} and Q_{3j}, $j = 1, \ldots, d$.

2 Main results

2.1 I.I.D. case

Theorem 1. *Let $\{X_n\}_{n\geq 1}$ be i.i.d. $\mathbb{R}^d$ valued random variables with density f, satisfying Assumption 2. Let K be a kernel satisfying Assumption 1(i) and let $\hat{h} \in I_n$. Then for every $\epsilon > 0$ there exists $C > 0$ such that, as $n \to \infty$*

$$P(|\hat{f}_{n,\hat{h}}(x) - f(x)| > \epsilon) = O\left(n^{d^2/(d+4)} \exp(-Cn^{4/(4+d)})\right).$$

Theorem 2. *Let $\{X_n\}_{n\geq 1}$ be i.i.d. $\mathbb{R}^d$ valued random variables with density f. Let $\hat{h} \in I_n$. Then for every $\epsilon > 0$ there exists $C' > 0$ such that, under Assumptions $1 - 3$, as $n \to \infty$*

$$P(||\hat{f}_{n,\hat{h}} - f|| > \epsilon) = O\left(n^{[d+d^2(2+1/\gamma)]/(d+4)} \exp(-C'n^{4/(4+d)})\right),$$

where $C' > 0$ and γ is a positive constant as defined in Assumption 3.

The proof of Theorems 1 and 2 are direct consequences of Lemma 4 which is stated and proved in Section 3.1. Using Borel-Cantelli Lemma and Theorem 2 we get the following Corollary.

Corollary 1. *Let $\{X_n\}_{n\geq 1}$ be i.i.d. $\mathbb{R}^d$ valued random variables with density f. Under Assumptions $1 - 3$, $||\hat{f}_{n,\hat{h}} - f|| = o(1)$ completely as $n \to \infty$.*

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

The coefficients $a_j,\ b_j,\ j = 1, \ldots, d$, in I_n can also be specified as functions of $X_1, X_2, \ldots, X_n$. For example, Dutta and Saha (2013) suggested to search for $\hat{h}$ in the interval

$$\prod_{j=1}^{d}[c_{1j}\eta_j^* n^{-1/(4+d)},\ c_{2j}\eta_j^* n^{-1/(4+d)}], \tag{2.1}$$

where $0 < c_{1j} < c_{2j}$ and $\eta_j^* = Q_{3j}^* - Q_{1j}^*,\ j = 1, \ldots, d$, and $Q_{1j}^*,\ Q_{3j}^*$ be the 1st and the 3rd sample quartiles based on $X_{1j}, \ldots, X_{nj}$. X_{ij} denotes the jth component of the vector X_i, $j = 1, \cdots, d$. The role and importance of η_j^* are discussed in detail in Dutta and Goswami (2013). See for instance the introduction and final remark number 5 in Dutta and Goswami (2013). From Dutta and Goswami (2013) we see that if the coefficients $a_j,\ b_j,\ j = 1, \ldots, d$, are random, their convergence rates effect the rate of point wise and uniform convergence of the density estimator.

In the next Theorem we extend the results in Theorems 1 and 2 to the case where $a_j,\ b_j,\ j = 1, \ldots, d$ are functions of $X_1, \ldots, X_n$ satisfying certain conditions.

Theorem 3. *Let $\{X_n\}_{n\geq 1}$ be i.i.d. $\mathbb{R}^d$ valued random variables with density f. Let*

$$H_n = \prod_{j=1}^{d}[a_{nj} n^{-1/(4+d)}, b_{nj} n^{-1/(4+d)}],$$

where $0 < a_{nj} < b_{nj},\ j = 1, \ldots, d$, are functions of $X_{1j}, \ldots, X_{nj}$ such that $P(|a_{nj} - a_j| > \epsilon) = O(r_{n,\epsilon})$ and $P(|b_{nj} - b_j| > \epsilon) = O(r_{n,\epsilon})$, where $a_j,\ b_j,\ j = 1, \ldots, d$, are positive constants and $r_{n,\epsilon} = o(1)$ as $n \to \infty$. Let $\hat{h} \in H_n$. Then under Assumptions $1 - 3$, as $n \to \infty$

$$(i) P(|\hat{f}_{n,\hat{h}}(x) - f(x)| > \epsilon) = O\left(r_{n,\epsilon} + n^{d^2/(d+4)} \exp(-C n^{4/(4+d)})\right),$$
$$(ii) P(\|\hat{f}_{n,\hat{h}} - f\| > \epsilon) = O\left(r_{n,\epsilon} + n^{[d+d^2(2+1/\gamma)]/(d+4)} \exp(-C' n^{4/(4+d)})\right),$$

where $C,\ C' > 0$.

In particular if we use $a_j = c_{1j}\eta_j^*,\ b_j = c_{2j}\eta_j^*$, as suggested by Dutta and Saha (2013), we have the following result (at the expense of one more assumption, viz. Assumption 4).

ACCEPTED MANUSCRIPT

Corollary 2. *Under the Assumptions* $1-4$*, and assuming* $\hat{h} \in H_n$ *we see that as as* $n \to \infty$

$$(i) P(|\hat{f}_{n,\hat{h}}(x) - f(x)| > \epsilon) = O\left(n^{d^2/(d+4)} \exp(-Cn^{4/(4+d)})\right),$$
$$(ii) P(\|\hat{f}_{n,\hat{h}} - f\| > \epsilon) = O\left(n^{[d+d^2(2+1/\gamma)]/(d+4)} \exp(-C' n^{4/(4+d)})\right).$$

The above corollary follows from the following inequality in page 74 of Serfling (1980)

$$P(|Q_{1j} - Q^*_{1j}| > \eta_j/4),\ P(|Q_{3j} - Q^*_{3j}| > \eta_j/4) \le \exp(-2n\delta^2),\ j = 1, \ldots, d,$$

where $\delta = \min\{\delta_1, \ldots, \delta_d\}$, $\delta_i^2 = \min\{\delta_{1i}^2, \delta_{2i}^2\}$, and $\delta_{1i} = \min\{F_i(Q^*_{1i} + \eta_i/4) - 1/4,\ 1/4 - F_i(Q^*_{1i} - \eta_i/4)\}$ and $\delta_{2i} = \min\{F_i(Q^*_{3i} + \eta_i/4) - 3/4,\ 3/4 - F_i(Q^*_{3i} - \eta_i/4)\}$, where F_i is the i-th marginal cumulative distribution function. Under Assumption 4, δ is a positive constant. Clearly δ is free of ϵ. This ensures that $r_{n,\epsilon}$ goes to zero at at exponential rate (which does not depend on ϵ).

Remark 1. *(i) The Theorems 1 and 2 are obtained under the assumption* $\hat{h} \in I_n$*. Therefore it is natural to question whether such an assumption is very artificial and restrictive? A number of well known bandwidth selectors such as the unbiased, the biased, and the smoothed cross validation and the different versions of the smooth bootstrap based bandwidth selectors in Sain et al. (1994) and Dutta (2011) involve optimization of some criterion (say* F_n*) which is a continuous function of* $h_1, \ldots, h_d$*. Even the multivariate plug-in bandwidth selection rule proposed by Wand and Jones (1994) involve optimization of the asymptotic mean integrated squared error(AMISE), where the unknowns are replaced by kernel based estimates, with respect to the bandwidths* $h_1, \ldots, h_d$*.*

Let $\hat{h}$ *denote the bandwidth vector that optimizes the bandwidth selection criterion* $F_n(h_1, \cdots, h_d)$*. If we optimize the bandwidth selection criterion* $F_n(h_1, \ldots, h_d)$ *on* $(0, \infty)^d$*, the optimum* $\hat{h}$ *may not be attained. Therefore it is natural to optimize the bandwidth selection criterion on some compact set, for example* I_n *or* H_n*. In that case, the resulting random bandwidth vector* $\hat{h}$ *will automatically satisfy the condition* $\hat{h} \in I_n$ *or* $\hat{h} \in H_n$*. Therefore one can ensure that a wide variety of the above mentioned random bandwidth vectors satisfy the assumption* $\hat{h} \in I_n$ *or* H_n *simply by optimizing the*

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

bandwidth selection criterion on I_n or H_n.

Moreover the assumption that $\hat{h} \in I_n$ or H_n is not very restrictive. For $d = 1$, I_n covers a wide range of bandwidths (see Dutta and Goswami (2013)). The simulation and real data analysis in Dutta and Saha (2013) also suggest that even for $d \geq 2$, the cross validation bandwidth restricted to H_n compares well with the usual unconstrained cross validation bandwidth vector.

(ii) The esteemed reviewer pointed out that the data-based cross validation selectors are known to be of the order $n^{-1/(4+d)}$. But their coefficients are unknown. Therefore is it possible to ensure that the unconstrained cross validation bandwidth satisfies the condition $\hat{h} \in I_n$ at least for large sample size?

Under the constraint $h_1 = h_2 = \cdots = h_d = h$, Sain et al. (1994) showed that as $n \to \infty$, $n^{(d+2)/(2d+8)}(\hat{h} - h_0)$ converges in law to Z, which follows normal distribution with mean zero and finite variance (see their proposition in Page 810). Where $\hat{h}$ is the cross validation bandwidth obtained under the assumption $h_1 = h_2 = \cdots = h_d = h$. h_0 is the optimal bandwidth minimizing the asymptotic mean integrated squared error, and h_0 equals some constant times $n^{-1/(d+4)}$. This result implies that for any open interval (a, b), $b > a > 0$,

$$\lim_{n\to\infty} P\left(a < n^{(d+2)/(2d+8)}(\hat{h} - h_0) < b\right) = P\left(a < Z < b\right).$$

Since h_0 is a multiple of $n^{-1/(d+4)}$, we can assume $\frac{c}{n^{1/(d+4)}} < h_0 < \frac{d}{n^{1/(d+4)}}$ for some $0 < c < d$. Then

$$\lim_{n\to\infty} P\left(\frac{c}{n^{1/(d+4)}} + \frac{a}{n^{d/(2d+8)}} < \hat{h} < \frac{d}{n^{1/(d+4)}} + \frac{b}{n^{d/(2d+8)}}\right) \geq P\left(a < Z < b\right).$$

The above result ensures that for any $0 < \eta < 1$, there exists $0 < c' < d'$ and an positive integer N (depending on η) such that

$$P\left(\hat{h} \in \left[\frac{c'}{n^{1/(d+4)}}, \frac{d'}{n^{1/(d+4)}}\right]\right) \geq \eta, \forall\, n > N.$$

Consequently for large n there is high probability that the cross validation bandwidth $\hat{h} \in \left[\frac{c'}{n^{1/(d+4)}}, \frac{d'}{n^{1/(d+4)}}\right]$.

ACCEPTED MANUSCRIPT

(iii) The range of bandwidths I_n can be further widened to $\prod_{j=1}^{d}\left[\frac{a_j}{n^{1/(4+d)+\delta}},\ \frac{b_j}{n^{1/(4+d)-\delta}}\right]$, where $0 < \delta < 1/(4+d)$, to accommodate more values of h especially the multiples of $n^{-1/(d+4)}$. In that case, under Assumptions $1-4$, we get that for $\hat{h} \in I_n$

$$P(|\hat{f}_{n,\hat{h}}(x) - f(x)| > \epsilon) = O\left(n^{(2d^2+2d)/(d+4)} \exp\left(-Cn^{4/(4+d)-d\delta}\epsilon^2\right)\right)$$

$$\text{and} \quad P(\|\hat{f}_{n,\hat{h}} - f\| > \epsilon) = O\left(n^{[4d+d^2(4+2/\gamma)]/(d+4)} \exp\left(-C'n^{4/(4+d)-d\delta}\epsilon^2\right)\right),$$

where $C,\ C'$ is positive constant. For $d = 1$, the above rates are same as those obtained by Dutta and Goswami (2013) in their final remark 3.

The reviewer rightly pointed out that a random bandwidth obtained by optimizing the selection criterion on I_n or H_n can be different from a bandwidth obtained by unconstrained optimization. If the range of bandwidths is widened to $(0,\ \infty)^d$, the extension of our theoretical results appear to be quite challenging. We leave this problem for further research.

2.2 strongly mixing case

A stationary process $\{X_t\}_{t\in\mathbb{Z}}$ is a strong mixing process, with mixing coefficient α, if

$$\alpha(n) = \sup_t \sup\{|P(A \cap B) - P(A)P(B)| : A \in M_{t+n}^{\infty}, B \in M_{-\infty}^{t}\} \downarrow 0, \text{ as } n \to \infty,$$

where $M_{-\infty}^{t}$ and M_{t+n}^{∞} denote σ-fields generated by $\{X_l,\ l \leq t\}$ and by $\{X_l, l \geq t+n\}$ respectively. Under very general dependence assumptions (that includes strong mixing condition), Lardjane (2007) has shown that the MSE of a KDE goes to zero at the rate similar (up to a logarithm) to the rate of convergence of the MSE under i.i.d.assumptions for $h_1 = h_2 = \cdots = h_d$ equal to a multiple of $(\log n/n)^{1/(d+4)}$ (see page 213, Lardjane (2007)). Therefore under strongly mixing dependence assumption, we search for the appropriate value of the random bandwidth vector $\hat{h}$ in the following

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

d-dimensional rectangle.

$$I'_n = \prod_{j=1}^{d} \left[a_j \{\log n/n\}^{1/(4+d)},\ b_j \{\log n/n\}^{1/(4+d)} \right],$$

where $0 < a_j < b_j,\ j = 1, \ldots, d$, are arbitrary constants.

Theorem 4. *Let $\{X_n\}_{n\geq 1}$ be a d-dimensional strongly mixing process with marginal density f and Assumptions 1(i) and 2 hold. Also let $\hat{h} \in I'_n$ and the mixing coefficient $\alpha(n)$ satisfies $\alpha(n) \leq D\rho^n$, where $0 < \rho < 1$ and $D > 0$. Then for every $\epsilon > 0$ there exists $C > 0$, such that as $n \to \infty$*

$$P\left(|\hat{f}_{n,\hat{h}}(x) - f(x)| > \epsilon\right) = O\left((n/\log n)^{d^2/(d+4)} \exp\left(-Cn^{(4-d)/(4+d)}(\log n)^{2d/(4+d)}\right)\right).$$

The right side of the above equation goes to zero for $d \leq 3$.

Theorem 5. *Let $\{X_n\}_{n\geq 1}$ be a d-dimensional strongly mixing process with marginal density f and Assumptions $1-3$ hold. Also let $\hat{h} \in I'_n$ and the mixing coefficient $\alpha(n)$ satisfies $\alpha(n) \leq D\rho^n$ where $0 < \rho < 1$ and $D > 0$. Then for every $\epsilon > 0$ there exists $C' > 0$, such that as $n \to \infty$*

$$P(\|\hat{f}_{n,\hat{h}} - f\| > \epsilon) = O\left(\left(\frac{n}{\log n}\right)^{[d+d^2(2+1/\gamma)]/(d+4)} \exp\left\{-C'n^{(4-d)/(4+d)}(\log n)^{2d/(4+d)}\right\}\right),$$

where γ is a positive constant as defined in Assumption 2. The right side of the above equation goes to zero for $d \leq 3$.

The proof of Theorem 4 and Theorem 5 are direct consequences of Lemma 6, stated and proved in Section 3.1.

Remark 2. *(i) Like in i.i.d. case, here also the range of bandwidths I'_n can be further widened to*

$$\prod_{j=1}^{d} \left[a_j \{\log n/n\}^{1/(4+d)+\delta}, b_j \{\log n/n\}^{1/(4+d)-\delta} \right],$$

where $0 < \delta < 1/(4+d)$, to accommodate more values of h. In that case, under Assumptions $1-3$,

ACCEPTED MANUSCRIPT

we get that for $\hat{h} \in I'_n$

$$P(|\hat{f}_{n,\hat{h}}(x) - f(x)| > \epsilon) = O\left(\left(\frac{n}{\log n}\right)^{\frac{2d^2+2d}{d+4}} \exp\left(-Cn^{\frac{4-d}{4+d}-2d\delta}(\log n)^{\frac{2d}{4+d}+2d\delta}\right)\right)$$

$$\text{and} \quad P(\|\hat{f}_{n,\hat{h}} - f\| > \epsilon) = O\left(\left(\frac{n}{\log n}\right)^{\frac{4d+d^2(4+2/\gamma)}{d+4}} \exp\left(-Cn^{\frac{4-d}{4+d}-2d\delta}(\log n)^{\frac{2d}{4+d}+2d\delta}\right)\right),$$

where $C,\ C'$ *is positive constant. Observe that if* $\delta < \frac{4-d}{(4+d)2d}$, *then right hand side of the above equations go to zero.*

(ii) Like in i.i.d. case, one may use data based coefficients $a_{nj},\ b_{nj},\ j = 1, \ldots, d$, *in* I'_n. *In that case the convergence rates in Theorems 4 and 5 are replaced by*

$$O\left(r_{n\epsilon} + (n/\log n)^{d^2/(d+4)} \exp\left(-Cn^{(4-d)/(4+d)}(\log n)^{2d/(4+d)}\right)\right)$$

and

$$O\left(r_{n\epsilon} + \left(\frac{n}{\log n}\right)^{[d+d^2(2+1/\gamma)]/(d+4)} \exp\left\{-C'n^{(4-d)/(4+d)}(\log n)^{2d/(4+d)}\right\}\right)$$

respectively, where $P(|a_{nj} - a_j| > \epsilon) = O(r_{n,\epsilon})$ *and* $P(|b_{nj} - b_j| > \epsilon) = O(r_{n,\epsilon})$ *and* $r_{n\epsilon} = o(1)$ *as* $n \to \infty$, *under the stated dependence assumption.* $0 < \epsilon,\ a_j,\ b_j,\ j = 1, \ldots, d$ *are arbitrary.*

In particular, one may use $a_{nj} = c_{1j}\eta^*_j,\ b_{nj} = c_{2j}\eta^*_j$, *where* $0 < c_{1j} < c_{2j}$ *and* $\eta^*_j = Q^*_{3j} - Q^*_{1j},\ j = 1, \ldots, d$, *and* $Q^*_{1j},\ Q^*_{3j}$ *be the 1st and the 3rd sample quartiles based on* $X_{1j}, \ldots, X_{nj}$. *In that case the rate of convergence of* $P\left(|a_{nj} - a_j| > \epsilon\right),\ P\left(|b_{nj} - b_j| > \epsilon\right)$ *are determined by the rate of convergence in probability of sample first and third quartiles under the stated dependence assumptions. Therefore, under Assumptions* $1 - 4$ *and strong mixing dependence assumption stated in Theorem 4, following similar argument given in the proof of Theorem 3 in Section 3.2 and using* (3.15) *from*

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Lemma 7, it is easy to see that

$$P\left(|\hat{f}_{n,\hat{h}}(x) - f(x)| > \epsilon\right) = O\left((n/\log n)^{d^2/(d+4)} \exp\left(-Cn^{(4-d)/(4+d)}(\log n)^{2d/(4+d)}\right)\right),$$
$$P(\|\hat{f}_{n,\hat{h}} - f\| > \epsilon) = O\left(\left(\frac{n}{\log n}\right)^{[d+d^2(2+1/\gamma)]/(d+4)} \exp\left\{-C' n^{(4-d)/(4+d)}(\log n)^{2d/(4+d)}\right\}\right).$$

3 Proofs

3.1 Lemmas and their proofs:

Here we state and prove a few lemmas. Lemma 1 and Lemma 2 are used to prove all the theorems stated in last section. Lemma 3 and Lemma 4 are used to prove Theorems 1, 2 and 3. Lemma 5 and Lemma 6 are used in the proof of Theorems 4 and 5. In Lemma 7, we derive the rate of convergence sample quantiles to population quantiles under strongly mixing dependence assumption. This result is used in Remark 2 (ii).

Lemma 1. *Let K be a density in $\mathbb{R}^d$ such that $\|K\|, \|z_j \partial_j K(z)\| < \infty$ for $j = 1, \ldots, d$ and $z = (z_1, \ldots, z_d) \in \mathbb{R}^d$. Then given $\epsilon > 0$, there exists $C > 0$ such that for any two bandwidth vectors $h = (h_1, \ldots, h_d)$ and $h' = (h'_1, \ldots, h'_d)$ in $I_n = \prod_{j=1}^d [a_j n^{-1/(4+d)},\ b_j n^{-1/(4+d)}]$, satisfying $|h_i - h'_i| < \delta_n = \frac{\epsilon}{4dCn^{(d+1)/(d+4)}},\ i = 1, \ldots, d$, we have*

$$\left|\hat{f}_{n,h}(x) - \hat{f}_{n,h'}(x)\right| < \epsilon/2.$$

Proof. Under given condition on K, the partial derivatives $\frac{\partial g_z(h)}{\partial h_i},\ i = 1, \ldots, d$, exist. Therefore, for fixed z and any two bandwidth vectors h and h' we have

$$g_z(h) - g_z(h') = \sum_{i=1}^{d} (h_i - h'_i) \int_0^1 \left[\frac{\partial g_z(h)}{\partial h_i}\right]_{h=h'+t(h-h')} dt.$$

ACCEPTED MANUSCRIPT

For any bandwidth vector $h = (h_1, \ldots, h_d) \in I_n$, it is easy to verify that

$$\left|\frac{\partial g_z(h)}{\partial h_i}\right| \leq \frac{1}{h_i^2 \prod_{l=1,\ l\neq i}^{d} h_l}\left\{\|K\| + \sup_{-\infty<z<\infty} |z_i \partial_i K(z)|\right\} \leq Cn^{(d+1)/(d+4)},\ i = 1, \ldots, d.$$

Therefore, under given conditions on K, for any two bandwidth vectors $h,\ h' \in I_n$ we see that

$$|g_z(h) - g_z(h')| \leq Cn^{(d+1)/(d+4)} \sum_{i=1}^{d} \left|h_i - h_i'\right|. \tag{3.1}$$

The right side of the above inequality is free of z. A direct consequence of (1.1) and (3.1) is that for any two bandwidth vectors $h,\ h' \in I_n$,

$$\left|\hat{f}_{n,h}(x) - \hat{f}_{n,h'}(x)\right| \leq \frac{1}{n}\sum_{i=1}^{n} \left|g_{x-X_i}(h) - g_{x-X_i}(h')\right| \leq Cn^{(d+1)/(d+4)} \sum_{i=1}^{d} \left|h_i - h_i'\right|. \tag{3.2}$$

The above inequality (3.2) implies that for any two random bandwidth vectors $h,\ h' \in I_n$, satisfying $|h_i - h_i'| < \delta_n,\ i = 1, \ldots, d$,

$$\left|\hat{f}_{n,h}(x) - \hat{f}_{n,h'}(x)\right| \leq dCn^{(d+1)/(d+4)}\delta_n.$$

For given $\epsilon > 0$, using $\delta_n = \frac{\epsilon}{4dCn^{(d+1)/(d+4)}}$ in the right side of the above inequality we get

$$\left|\hat{f}_{n,h}(x) - \hat{f}_{n,h'}(x)\right| < \epsilon/2.$$

This completes the proof the Lemma. □

Lemma 2. *Let K be a density in $\mathbb{R}^d$ such that $\|\partial_j K\| < \infty$ for $j = 1, 2, \ldots, d$. Then given $\epsilon > 0$ and a bandwidth vector $h = (h_1, \ldots, h_d)$, there exists $b_n > 0$ such that for any two vectors $x = (x_1, \ldots, x_d)$ and $x' = (x_1', \ldots, x_d')$ satisfying $|x_j - x_j'| < b_n,\ j = 1, \ldots, d$,*

$$\left|\hat{f}_{n,h}(x) - \hat{f}_{n,h}(x')\right| < \epsilon/2.$$

ACCEPTED MANUSCRIPT

Proof. It is easy to see that

$$\begin{aligned} \left|\hat{f}_{n,h}(x) - \hat{f}_{n,h}(x')\right| &\leq \frac{1}{n\prod_{j=1}^{d} h_j} \sum_{i=1}^{n} \left| K\left(\frac{x - X_i}{h}\right) - K\left(\frac{x' - X_i}{h}\right)\right| \\ &\leq \frac{1}{\prod_{j=1}^{d} h_j} \sum_{j=1}^{d} \frac{|x_j - x'_j|}{h_j} \|\partial_j K\| < \frac{d b_n}{h_*^{d+1}} \sum_{j=1}^{d} \|\partial_j K\|, \end{aligned}$$

where $h_* = \min\{h_1, \ldots, h_d\}$. Using $b_n = \frac{\epsilon h_*^{d+1}}{4d\sum_{j=1}^{d} \|\partial_j K\|}$ in the right side of the above inequality we get

$$\left|\hat{f}_{n,h}(x) - \hat{f}_{n,h}(x')\right| < \epsilon/2.$$

This completes the proof the Lemma. □

Lemma 3. *Let $X_1, \ldots, X_n$ be i.i.d. random variables with density f, satisfying Assumptions 2 and 3. Further let $h = (h_1, \ldots, h_d)$ be a bandwidth vector such that $h_i = o(1),\ i = 1, \ldots, d$ and $nh_*^d \to \infty$ as $n \to \infty$, where $h_* = \min\{h_1, \ldots, h_d\}$. Then under Assumption 1, as $n \to \infty$*

$$P\left(\|\hat{f}_{n,h} - f\| > \epsilon\right) = O\left(\frac{1}{(h_*)^{d+d^2(1+1/\gamma)}} \exp(-Cnh_*^d)\right),\ C > 0,$$

where γ is a positive constant as stated in Assumption 3.

Proof. Let us introduce some notation

$$K_1 = \|K\|,\ K_2 = \sup_{x\in\mathbb{R}^d} \|x\|^d K(x),\ K_4 = 2^\gamma \int \|x\|^\gamma f(x)dx.$$

Given $\epsilon > 0$, let $R_n = [-a_n,\ a_n]^d$ where $a_n = \left(16K_1K_4/\epsilon h_*^d\right)^{1/\gamma}$ and $b_n = \frac{\epsilon {h_*}^{d+1}}{4d\sum_{j=1}^{d} \|\partial_j K\|}$. It is easy to see that

$$P\left(\|\hat{f}_{n,h} - f\| > \epsilon\right) \leq P\left(\sup_{x\in R_n} |\hat{f}_{n,h}(x) - f(x)| > \epsilon\right) + P\left(\sup_{x\in R_n^C} |\hat{f}_{n,h}(x) - f(x)| > \epsilon\right), \tag{3.3}$$

where R_n contains a d-dimensional sphere S_n with center $0 = (0, \ldots, 0)$ and radius a_n. If x is not in R_n, it is not in S_n. Therefore

$$
\begin{aligned}
P\left(\sup_{x\in R_n^C}|\hat{f}_{n,h}(x)-f(x)|>\epsilon\right) &\leq P\left(\sup_{x\in S_n^C}|\hat{f}_{n,h}(x)-f(x)|>\epsilon\right)\\
&= P\left(\sup_{||x||>a_n}|\hat{f}_{n,h}(x)-f(x)|>\epsilon\right) \qquad (3.4)
\end{aligned}
$$

Under Assumptions 1 and 2, and the condition $h_i = o(1),\ i = 1, \ldots, d$ as $n \to \infty$, we get

$$
||E(\hat{f}_{n,h})-f|| \leq \sum_{i=1}^{d}|h_i|||\partial_i f||\int|z_i|K(z_1,\ldots,z_d)dz_1\ldots dz_d = o(1). \qquad (3.5)
$$

Now $$P\left(\sup_{\{x:||x||>a_n\}}|\hat{f}_{n,h}(x)-f(x)|>\epsilon\right)$$

$$
\leq P\left(\sup_{\{x:||x||>a_n\}}|\hat{f}_{n,h}(x)|\geq\epsilon/2\right)+P\left(\sup_{\{x:||x||>a_n\}}f(x)\geq\epsilon/4\right)+P\left(||E(\hat{f}_{n,h})-f||\geq\epsilon/4\right)
$$

Under Assumption 2, $f(x) \to 0$ as $||x|| \to \infty$ and under the stated conditions on h, $a_n \to \infty$ as $n \to \infty$. Therefore for any $\epsilon > 0$, $P\left(\sup_{\{x:||x||>a_n\}} f(x) \geq \epsilon/4\right)$ is equal to zero for sufficiently large value of n.

Moreover from (3.5) we see that as $n \to \infty$, $P\left(||E(\hat{f}_{n,h}) - f|| \geq \epsilon/4\right)$ eventually equals zero. Therefore we see that for every $\epsilon > 0$, there exists $N_\epsilon > 1$ such that for every $n > N_\epsilon$

$$
\begin{aligned}
P\left(\sup_{||x||>a_n}|\hat{f}_{n,h}(x)-f(x)|>\epsilon\right) &\leq P\left(\sup_{||x||>a_n}|\hat{f}_{n,h}(x)|\geq\epsilon/2\right)\\
&\leq P\left(\sup_{||x||>a_n}\frac{1}{n\prod h_j}\sum_{\{i:||X_i-x||>a_n/2\}}K(\frac{x-X_i}{h})\geq\epsilon/4\right)\\
&\quad+P\left(\sup_{||x||>a_n}\frac{1}{n\prod h_j}\sum_{\{i:||X_i-x||\leq a_n/2\}}K(\frac{x-X_i}{h})\geq\epsilon/4\right).
\end{aligned}
$$

Now $P\left(\sup_{||x||>a_n}\frac{1}{n\prod h_j}\sum_{\{i:||X_i-x||>a_n/2\}}K(\frac{x-X_i}{h})\geq\epsilon/4\right)\leq P\left(\frac{(\max h_j)^d}{\prod h_j}(a_n/2)^{-d}K_2\geq\epsilon/4\right)\to 0$

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

as $n \to \infty$. This follows from the fact that $a_n \to \infty$ as $n \to \infty$. Also

$$P\left(\sup_{\|x\|>a_n} \frac{1}{n\prod h_j} \sum_{\{i:\|X_i - x\|\le a_n/2\}} K(\frac{x-X_i}{h}) \ge \epsilon/4\right) \le P\left(\frac{1}{n}\sum_{\{i:\|X_i\|\ge a_n/2\}} K_1 \ge \frac{\epsilon}{4}h_*^d\right)$$

Moreover, using Markov inequality and recalling the definition of a_n, we see that

$$P(\|X_1\| > a_n/2) \le E\|X_1\|^\gamma/(a_n/2)^\gamma = \frac{K_4}{a_n^\gamma} = \epsilon h_*^d/16K_1.$$

Therefore every $\epsilon > 0$, there exists $N'_\epsilon > N_\epsilon$ such that for every $n > N'_\epsilon$

$$\begin{aligned} P\left(\sup_{\|x\|>a_n} |\hat{f}_{n,h}(x) - f(x)| > \epsilon\right) &\le P\left(\frac{1}{n}\sum_{\{i:\|X_i\|\ge a_n/2\}} K_1 \ge \frac{\epsilon}{4}h_*^d\right) \\ &\le P\left(\frac{1}{n}\sum_{i=1}^n I_{\|X_i\|>a_n/2} \ge \frac{\epsilon h_*^d}{4K_1}\right) \\ &\le P\left(\frac{1}{n}\sum_{i=1}^n \left[I_{\|X_i\|>a_n/2} - P(\|X_1\| > a_n/2)\right] \ge \frac{\epsilon h_*^d}{K_1}(\frac{1}{4} - \frac{1}{16})\right) \end{aligned}$$

Using inequality (48) in page 187 in Rao (1983), in the right side of the above inequality we get the following equation

$$P\left(\sup_{\|x\|>a_n} |\hat{f}_{n,h}(x) - f(x)| > \epsilon\right) = O\left(\exp(-C_1 n(h_*)^d)\right), \text{ as } n \to \infty. \tag{3.6}$$

From (3.3), (3.4) and the above equation we get as $n \to \infty$,

$$P\left(\|\hat{f}_{n,h} - f\| > \epsilon\right) \le P\left(\sup_{x\in R_n} |\hat{f}_{n,h}(x) - f(x)| > \epsilon\right) + O\left(\exp(-Cn(h_*)^d)\right), \; C > 0. \tag{3.7}$$

We partition R_n into $k'(n) = (1 + [2a_n/b_n])^d$ (where $[y]$ is the largest integer not exceeding y) subintervals $\{J_{ni} : i = 1, \ldots, k'(n)\}$ each of volume b_n^d. $\{J_{ni} : i = 1, \ldots, k'(n)\}$ are d-fold product of the non-overlapping intervals of a partition of $[-a_n, \; a_n]$ into sub-intervals of length b_n. We see that

$$P\left(\sup_{x\in R_n} |\hat{f}_{n,h}(x) - f(x)| > \epsilon\right) \le P\left(\max_{1\le i\le k'(n)} \left\{|\hat{f}_{n,h}(y_i) - f(y_i)| + \sup_{x\in J_{ni}} |\hat{f}_{n,h}(y_i) - \hat{f}_{n,h}(x)|\right\} > \epsilon\right),$$

ACCEPTED MANUSCRIPT

where $y_i \in J_{ni},\ i = 1, \ldots, k'(n)$. For $x \in J_{ni},\ i = 1, \ldots, k'(n)$, x and y_i satisfy the conditions of Lemma 2. Therefore by Lemma 2

$$|f_{n,h}(y_i) - \hat{f}_{n,h}(x)| < \epsilon/2, \forall x \in J_{ni}, i = 1, \ldots, k'(n).$$

Therefore

$$\begin{aligned} P\left(\sup_{x \in R_n} |\hat{f}_{n,h}(x) - f(x)| > \epsilon\right) &\leq P\left(\max_{1 \leq i \leq k'(n)} \left\{|\hat{f}_{n,h}(y_i) - f(y_i)|\right\} > \epsilon/2\right) \\ &\leq \sum_{i=1}^{k'(n)} P\left(|\hat{f}_{n,h}(y_i) - f(y_i)| > \epsilon/2\right). \end{aligned} \tag{3.8}$$

Using (3.5), we see that for $\epsilon > 0$ there exists N_1 such that for

$$\sum_{i=1}^{k'(n)} P\left(|\hat{f}_{n,h}(y_i) - f(y_i)| > \epsilon/2\right) \leq \sum_{i=1}^{k'(n)} P\left(|\hat{f}_{n,h}(y_i) - E\hat{f}_{n,h}(y_i)| > \epsilon/4\right).$$

Now using inequality (27) in the proof of Theorem 3.1.5. in Rao (1983), we get that as $n \to \infty$

$$P\left(|\hat{f}_{n,h}(x) - E[\hat{f}_{n,h}(x)]| > \epsilon/4\right) = O(\exp(-Cnh_*^d \epsilon^2). \tag{3.9}$$

Therefore as $n \to \infty$,

$$P\left(\sup_{x \in R_n} |\hat{f}_{n,h}(x) - f(x)| > \epsilon\right) = O\left(k'(n) \exp(-C_2 n h_*^d)\right),\ C_2 > 0.$$

Substituting the above equation in the right side of (3.7) we get as $n \to \infty$

$$P\left(\|\hat{f}_{n,h} - f\| > \epsilon\right) = O\left(k'(n) \exp(-Cnh_*^d)\right),\ C > 0.$$

Recalling the notions $a_n,\ b_n$ we see that $1 + [2a_n/b_n]$ is a multiple of $\frac{1}{(h_*)^{d+1+d/\gamma}}$. $k'(n) = (1 + [2a_n/b_n])^d$ is a multiple of $\frac{1}{(h_*)^{d^2(1+1/\gamma)+d}}$. This completes the proof. □

Lemma 4. *Let $X_1, \ldots, X_n$ be i.i.d. random variables with density f, satisfying Assumptions 2 and*

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

3. Under Assumption 1, we see that as $n \to \infty$

$$(i) P(\sup_{h\in I_n} |\hat{f}_{n,h}(x) - f(x)| > \epsilon) = O\left(n^{d^2/(d+4)} \exp(-C n^{4/(4+d)})\right),$$
$$(ii) P(\sup_{h\in I_n} \|\hat{f}_{n,h} - f\| > \epsilon) = O\left(n^{[d+d^2(2+1/\gamma)]/(d+4)} \exp(-C' n^{4/(4+d)})\right),$$

where $I_n = \prod_{j=1}^{d}[a_j n^{-1/(4+d)},\ b_j n^{-1/(4+d)}]$, $C, C' > 0$ and γ is a positive constant as defined in Assumption 2.

Proof. Given $\epsilon > 0$, we partition $I_n = \prod_{j=1}^{d}[a_j n^{-1/(4+d)},\ b_j n^{-1/(4+d)}]$ into $k(n)$ non-overlapping d-dimensional rectangles $\{I_{ni},\ i = 1, \ldots, k(n)\}$. To achieve this, each $[a_j n^{-1/(4+d)},\ b_j n^{-1/(4+d)}]$, $j = 1, \ldots, d$, is partitioned into non-overlapping sub-intervals of length proportional to $\delta_n = \frac{\epsilon}{4dCn^{(d+1)/(d+4)}}$. $\{I_{ni},\ i = 1, \ldots, k(n)\}$ are the d-fold product of these sub-intervals of length δ_n. Consequently $k(n)$ is proportional to $\frac{1}{n^{d/(4+d)}\delta_n^d} = n^{d^2/(d+4)}$. Now using Lemma 1, we get that

$$\begin{aligned}
P(\sup_{h\in I_n} |\hat{f}_{n,h}(x) - f(x)| > \epsilon) &\leq P(\max_{1\leq i\leq k(n)} \{|\hat{f}_{n,h^i}(x) - f(x)| + \sup_{h\in I_{ni}} |\hat{f}_{n,h}(x) - f_{n,h^i}(x)|\} > \epsilon) \\
&\leq P(\max_{1\leq i\leq k(n)} |\hat{f}_{n,h^i}(x) - f(x)| > \epsilon/2) \\
&\leq \sum_{i=1}^{k(n)} P\left(|\hat{f}_{n,h^i}(x) - f(x)| > \epsilon/2\right) \qquad (3.10)
\end{aligned}$$

and similarly

$$P(\sup_{h\in I_n} \|\hat{f}_{n,h} - f\| > \epsilon) \leq \sum_{i=1}^{k(n)} P\left(\|\hat{f}_{n,h^i} - f\| > \epsilon/2\right) \qquad (3.11)$$

where $h^i \in I_{ni}, i = 1, \ldots,\ k(n)$ and each h^i is a non random bandwidth vector.

Using (3.5) we see that for $\epsilon > 0$ there exists N_1 such that for $1 \leq i \leq k(n)$,

$$\|E(\hat{f}_{n,h^i}) - f\| < \epsilon/4,\ \forall\ n > N_1.$$

ACCEPTED MANUSCRIPT

Hence for $n > N_1$, we get from (3.10) that

$$P(\sup_{h\in I_n} |\hat{f}_{n,h}(x) - f(x)| > \epsilon) \leq \sum_{i=1}^{k(n)} P\left(|\hat{f}_{n,h^i}(x) - E[\hat{f}_{n,h^i}(x)]| > \epsilon/4\right) \tag{3.12}$$

Now using inequality (27) in the proof of Theorem 3.1.5. in Rao (1983), we get that as $n \to \infty$

$$P\left(|\hat{f}_{n,h^i}(x) - E[\hat{f}_{n,h^i}(x)]| > \epsilon/4\right) = O(\exp(-Cn(h_*^i)^d\epsilon^2) = O\left(\exp(-Cn^{4/(d+4)}\epsilon^2)\right),$$

where $h_*^i = \min\{h_1^i, \ldots, h_d^i\},\ i = 1, \ldots, k(n)$. Substituting the above result in the right side of (3.12) and recalling that $k(n)$ is a multiple of $n^{d^2/(d+4)}$, we get the result (i) of Lemma 4.

To prove (ii), we use Lemma 3 and see that as $n \to \infty$

$$P\left(||\hat{f}_{n,h^i} - f|| > \epsilon\right) = O\left(\frac{1}{(h_*^i)^{d+d^2(1+1/\gamma)}} \exp(-Cn(h_*^i)^d)\right) \text{ where } C > 0,$$

γ is a positive constant as stated in Assumption 3. Substituting the above equation in the right side of (3.11) and using the fact that $k(n)$ is a multiple of $n^{d^2/(d+4)}$, we get the following equation

$$P(\sup_{h\in I_n} ||\hat{f}_{n,h} - f|| > \epsilon) = O\left(n^{[d+d^2(2+1/\gamma)]/(d+4)} \exp(-C' n^{4/(4+d)})\right).$$

This completes the proof of (ii). □

Lemma 5. *Let $\{X_n\}_{n\geq 1}$ be a d-dimensional strongly mixing process with marginal density f, satisfying Assumptions 2 and 3. Further let the mixing coefficient $\alpha(n)$ satisfies $\alpha(n) \leq D\rho^n$ where $0 < \rho < 1$ and $D > 0$ and $h = (h_1, \ldots, h_d)$ be a bandwidth vector such that $h_i = o(1),\ i = 1, \ldots, d$ and $nh_*^d \to \infty$, $h_*^d \log n \log\log n \to 0$ as $n \to \infty$, where $h_* = \min\{h_1, \ldots, h_d\}$. Then under Assumption 1, as $n \to \infty$*

$$P\left(||\hat{f}_{n,h} - f|| > \epsilon\right) = O\left(\frac{1}{(h_*)^{d+d^2(1+1/\gamma)}} \exp\left(-Cnh_*^{2d}\right)\right),\ C > 0,$$

where γ is a positive constant as stated in Assumption 3.

Proof. Proof of this lemma is exactly similar to the proof of Lemma 3. Observe that most of the

ACCEPTED MANUSCRIPT

arguments used in the proof of Lemma 3 do not depend on independence of $\{X_n\}$. Only in (3.6) and (3.9), we used independence. In particular, we used Bernstein type inequality for i.i.d. entries to find the convergence rates of $P\left(\frac{1}{n}\sum_{i=1}^{n} I_{\|X_i\|>a_n/2} \geq \frac{\epsilon h_*^d}{4K_1}\right)$ and $P\left(\left|\hat{f}_{n,h}(y_l) - E\left[\hat{f}_{n,h}(y_l)\right]\right| > \epsilon/4\right)$, respectively.

Here we use Bernstein type inequality from Merleveede et al. (2009) for strongly mixing process to prove (3.6) and (3.9). Observe that

$$P\left(\left|\hat{f}_{n,h}(y_l) - E\left[\hat{f}_{n,h}(y_l)\right]\right| > \epsilon/4\right) \leq P(|\sum_{j=1}^{n} Y_{nj}| > nh_*^d\epsilon/4), \tag{3.13}$$

where $Y_{nj} = K\left(\frac{x-X_j}{h}\right) - E[K\left(\frac{x-X_j}{h}\right)]$. If $\{X_n\}$ is a stationary strongly mixing process with mixing coefficient $\alpha(n)$, then each row of the triangular array $\{Y_{nj}, j = 1, \ldots, n\}$ represents a strongly mixing stationary sequence of mean zero bounded random variables with a sequence of mixing coefficients $\{\alpha_j(n)\}$ bounded above by $\{\alpha(n)\}$ (see Dutta and Goswami (2013)). Under the stated condition on $\alpha(n)$, the sequence $\{\alpha_j(n)\}$, $j = 1, \ldots, n$ also satisfies the condition $\alpha_j(n) \leq exp(-2cn)$ for some $c > 0$. Under this, using Theorem 1 of Merleveede et al. (2009) we get that for $\epsilon > 0$ and $n \geq 4$,

$$P(|\sum_{j=1}^{n} Y_{nj}| > nh_*^d\epsilon/4) \leq \exp\left(-\frac{Cnh_*^{2d}\epsilon^2}{4K_1^2 + 2K_1\epsilon h_*^d \log n \log\log n}\right)$$

where $C,\ K_1 > 0$. Under the stated assumption on the bandwidth vector h, the right side of the above inequality equals $O\left(\exp(-C'nh_*^{2d})\right)$ as $n \to \infty$. Hence under the stated dependence assumptions and the other conditions, we have as $n \to \infty$

$$P\left(\left|\hat{f}_{n,h}(y_j) - E\left[\hat{f}_{n,h}(y_j)\right]\right| > \epsilon/4\right) = O\left(\exp(-C'nh_*^{2d})\right), \tag{3.14}$$

for $j = 1, \ldots, k'(n)$ and this is analogue of (3.9) of Lemma 3.

Similarly, using Theorem 1 of Merleveede et al. (2009), we see that

$$P\left(\frac{1}{n}\sum_{i=1}^{n} I_{\|X_i\|>a_n/2} \geq \frac{\epsilon h_*^d}{4K_1}\right) \leq \exp\left(-\frac{Cnh_*^{2d}\epsilon^2}{4K_1^2 + 2K_1\epsilon h_*^d \log n \log\log n}\right) = O\left(\exp(-C_1 nh_*^{2d})\right)$$

ACCEPTED MANUSCRIPT

and we get (3.6) for strongly mixing case. This completes the proof of this lemma. □

Lemma 6. *Let* $\{X_n\}_{n\geq 1}$ *be a d-dimensional strongly mixing process with marginal density* f*, satisfying Assumptions 2 and 3. Also let the mixing coefficient* $\alpha(n)$ *satisfies* $\alpha(n) \leq D\rho^n$ *where* $0 < \rho < 1$ *and* $D > 0$*. Under Assumption 1, we see that as* $n \to \infty$

$$(i) P(\sup_{h\in I_n} |\hat{f}_{n,h}(x) - f(x)| > \epsilon) = O\left(\left(\frac{n}{\log n}\right)^{d^2/(d+4)} \exp\left(-Cn^{(4-d)/(4+d)}(\log n)^{2d/(4+d)}\right)\right),$$
$$(ii) P(\sup_{h\in I_n} \|\hat{f}_{n,h} - f\| > \epsilon) = O\left(\left(\frac{n}{\log n}\right)^{[d+d^2(2+1/\gamma)]/(d+4)} \exp\left\{-C' n^{(4-d)/(4+d)}(\log n)^{2d/(4+d)}\right\}\right),$$

where $I_n = \prod_{j=1}^{d}[a_j(\log n/n)^{1/(4+d)},\ b_j(\log n/n)^{1/(4+d)}]$*,* $C, C' > 0$ *and* γ *is a positive constant as defined in Assumption 3.*

Proof. We follow the similar arguments given in the proof of Lemma 4. From the proof of the Lemma 4, we see that the arguments up to the inequality (3.12) do not depend on independence of $X_1, X_2, \ldots, X_n$. Therefore under strong mixing dependence assumption, we repeat these arguments, replacing $n^{1/(4+d)}$ by $(\frac{\log n}{n})^{1/(4+d)}$. Consequently the inequalities (3.11) and (3.12) hold, with $k(n)$ proportional to $(\frac{n}{\log n})^{d^2/(d+4)}$, under the stated dependence assumption. Now we complete the proof of (i) by using (3.14) in the right side of (3.12) and of (ii) by using Lemma 5 in the right side of (3.11). Most of the arguments used in the proof of Lemma 3 do not depend on any dependence assumption. For instance, under the stated conditions, the inequalities (3.5) and (3.8) do not require the independence assumption to be true. These results can be repeated as it is in the presence of mixing type dependence.

Only the convergence rate of $P\left(\left|\hat{f}_{n,h}(y_l) - E\left[\hat{f}_{n,h}(y_l)\right]\right| > \epsilon/4\right)$, $l = 1, \ldots, k'(n)$, is obtained under i.i.d. assumption in Lemma 3. In Lemma 5, we obtain the same convergence rate under strongly mixing dependence assumption, where $\alpha(n) \leq D\rho^n$, $0 < D$, $\rho < 1$. From (3.13) we see that

$$P\left(\left|\hat{f}_{n,h}(y_l) - E\left[\hat{f}_{n,h}(y_l)\right]\right| > \epsilon/4\right) \leq P(|\sum_{j=1}^{n} Y_{nj}| > nh_*^d\epsilon/4),$$

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

where $Y_{nj} = K\left(\frac{x-X_j}{h}\right) - E[K\left(\frac{x-X_j}{h}\right)],\ j = 1, \ldots, n$. If $\{X_n\}$ is a stationary strongly mixing process with mixing coefficient α, each row of the triangular array $\{Y_{nj}, j = 1, ..., n, n \in N\}$ represents a strongly mixing stationary sequence of mean zero bounded random variables with a sequence of mixing coefficients bounded above by $\{\alpha(n)\}$ (see Dutta and Goswami (2013)). Under the stated condition on α, the sequence of mixing coefficients of $\{Y_{nj}, j = 1, ..., n, n \in N\}$ also satisfies the condition $\alpha(n) \leq exp(-2cn),\ c > 0$. Under this using the Theorem 1 in page 3 of Merleveede et al. (2009) we get the following inequality. For $n \geq 4$ and $\epsilon > 0$,

$$P(|\sum_{j=1}^{n} Y_{nj}| > nh_*^d \epsilon/4) \leq \exp\left(-\frac{Cnh_*^{2d}\epsilon^2}{4K_1^2 + 2K_1\epsilon h_*^d \log n \log\log n}\right),$$

where $C,\ K_1 > 0$. Under the stated assumption on the bandwidth vector h, the right side of the above inequality equals $O\left(\exp(-C'nh_*^{2d})\right)$ as $n \to \infty$. Hence under the stated dependence assumptions and the other conditions we have as $n \to \infty$

$$P\left(\left|\hat{f}_{n,h}(y_l) - E\left[\hat{f}_{n,h}(y_l)\right]\right| > \epsilon/4\right) = O\left(\exp(-C'nh_*^{2d})\right),\ l = 1, \ldots, k'(n),$$

Now using t the condition that $h_i = o(1),\ i = 1, \ldots, d$, and the inequalities (3.5) and (3.8) we get the stated rate of convergence $P\left(\|\hat{f}_{n,h} - f\| > \epsilon\right)$. This completes the proof. □

In the following lemma we prove the rate of convergence of sample quantiles to population quantiles for strongly mixing case.

Lemma 7. *Let $\{X_n\}_{n\geq 1}$ be a d-dimensional strongly mixing process with marginal density f satisfying Assumption 4. Also let the mixing coefficient $\alpha(n)$ satisfies $\alpha(n) \leq D\rho^n$ where $0 < \rho < 1$ and $D > 0$. Then for $i = 1, 3$ and $j = 1, 2, \ldots, d$,*

$$P\left(|Q_{ij} - Q_{ij}^*| > \epsilon\right) = O\left(\sqrt{n}\exp(-\sqrt{n}s)\right) \tag{3.15}$$

where s is a positive constant.

Proof. If $\{X_n = (X_{n1}, X_{n2}, \ldots, X_{nd})\}_{n\geq 1}$ is d-dimensional strongly mixing process with mixing coef-

ACCEPTED MANUSCRIPT

ficient $\alpha(n)$ then for $1 \le j \le d$, $\{X_{nj}\}_{n\ge1}$ is one dimensional strongly mixing process with mixing coefficient $\alpha_j(n)$ and $\alpha_j(n) \le \alpha(n)$. With this observation, this lemma follows from Lemma 4 of Dutta and Goswami (2013). □

3.2 Proof of Theorems

Theorems 1 and 2 follow from Lemma 4 and Theorems 4 and 5 follow from Lemma 6. It remains to prove only Theorem 3.

Proof of Theorem 3. Let $I_n = \prod_{j=1}^{d}\left[\frac{a_j}{2}n^{-1/(4+d)},\ \frac{3b_j}{2}n^{-1/(4+d)}\right]$. We note that

$$\begin{aligned} P(|\hat{f}_{n,\hat{h}}(x) - f(x)| > \epsilon) &\le P\left(\hat{h} \notin I_n\right) + P\left(|\hat{f}_{n,\hat{h}}(x) - f(x)| > \epsilon,\ \hat{h} \in I_n\right) \\ &\le P\left(\hat{h} \notin I_n\right) + P\left(\sup_{h\in I_n} |\hat{f}_{n,h}(x) - f(x)| > \epsilon\right). \end{aligned} \tag{3.16}$$

Similarly

$$P(\|\hat{f}_{n,\hat{h}} - f\| > \epsilon) \le P\left(\hat{h} \notin I_n\right) + P\left(\sup_{h\in I_n} \|\hat{f}_{n,h} - f\| > \epsilon\right) \tag{3.17}$$

We assume that $\hat{h} \in H_n = \prod_{j=1}^{d}[a_{nj}n^{-1/(4+d)},\ b_{nj}n^{-1/(4+d)}]$. Therefore it is easy to verify that

$$\begin{aligned} P\left(\hat{h} \notin I_n\right) &\le \sum_{j=1}^{d}[P(|a_{nj} - a_j| > a_j/2) + P(|b_{nj} - b_j| > b_j/2)] \\ \Rightarrow P\left(\hat{h} \notin I_n\right) &= O\left(r_{n,\epsilon}\right) \end{aligned} \tag{3.18}$$

The convergence rates of $P\left(\sup_{h\in I_n} |\hat{f}_{n,h}(x) - f(x)| > \epsilon\right)$ and $P\left(\sup_{h\in I_n} \|\hat{f}_{n,h} - f\| > \epsilon\right)$ are obtained in Lemma 4 (*i*) and (*ii*). Hence Theorem 3 is a consequence of (3.18) and Lemma 4. □

Acknowledgement. We are thankful to the esteemed reviewer for the detailed report and constructive suggestions which helped us to improve this manuscript.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

References

[1] Dutta, S. (2011). Estimation of the MISE and the optimal bandwidth vector of a product kernel density estimate. *Journal of Statistical Planning and Inference*, 141, 1817-1831.

[2] Dutta, S. and Goswami, A. (2013). Pointwise and uniform convergence of kernel density estimators using random bandwidths. *Statistics and Probability Letters* 83, 2711-2720.

[3] Dutta, S. and Saha, K. (2013). Consistency of multivariate density estimators using random bandwidths. accepted for publication in *Communications in Statistics-Theory and Methods.*

[4] Duong, T. and Hazelton M. L. (2005). Cross-validation bandwidth matrices for multivariate kernel density estimation. *Scandinavian Journal of Statistics*, 32, 485-506.

[5] Lardjane, S. (2007). Nonparametric density estimation for nonmixing approximable Stochastic Processes. Stat Infer Stoch Process, 10, 209–221.

[6] Merlev‘ede, F., Peligrad, M., and Rio, E. (2009). Bernstein inequality and moderate deviations under strong mixing conditions. IMS Collections, High Dimensional Probability V, 273-292.

[7] Park, B. U. and Marron , J. S. (1990). Comparison of Data-Driven Bandwidth Selectors *The Journal of the American Statistical Association,* Theory and Methods, 85, No. 409 , 66-72.

[8] Rao, B.L.S.P. (1983). *Nonparametric Functional Estimation*. New York: Academic Press.

[9] Sain, S. R., Baggerly, K. A. and Scott, D. W. (1994). Cross-Validation of Multivariate Densities, *The Journal of the American Statistical Association,* Theory and Methods, 89, No. 427, 807-817.

[10] Scott D. W. (1992). *Multivariate Density Estimation* Theory, Practice and Visualization, John Wiley and Sons, Inc.

[11] Serfling, R.J. (1980). *Aproximation Theorems of Mathematical Statistics*, New York: John Wiley & Sons.

[12] Tenreiro, C. (2001). On the asymptotic behaviour of the integrated square error of kernel density estimators with data-dependent bandwidth. *Statistics* & *Probability Letters*, 53, 283-292.

[13] Wand, M. P. and Jones, M.C. (1994). Multivariate plug-in bandwidth selection, *Computational Statistics*, 9, 97-116.

ACCEPTED MANUSCRIPT