Ann. Inst. Statist. Math.
Vol. 42, No. 2, 305–329 (1990)

RECURSIVE KERNEL DENSITY ESTIMATORS UNDER A WEAK DEPENDENCE CONDITION

LANH TAT TRAN

Department of Mathematics, Indiana University, Bloomington, IN 47405, U.S.A.

(Received March 8, 1989; revised June 19, 1989)

Abstract. Let X_t, $t = \ldots, -1, 0, 1, \ldots$ be a strictly stationary sequence of random variables (r.v.'s) defined on a probability space $(\Omega, \mathscr{F}, P)$ and taking values in R^d. Let $X_1, \ldots, X_n$ be n consecutive observations of X_t. Let f be the density of X_1. As an estimator of $f(x)$, we shall consider $\hat{f}_n(x) = n^{-1} \sum_{j=1}^{n} b_j^{-d} K((x - X_j)/b_j)$. Here K is a kernel function and b_n is a sequence of bandwidths tending to zero as $n \to \infty$. The asymptotic distribution and uniform convergence of $\hat{f}_n$ are obtained under general conditions. Appropriate bandwidths are given explicitly. The process X_t is assumed to satisfy a weak dependence condition defined in terms of joint densities. The results are applicable to a large class of time series models.

Key words and phrases: Asymptotic normality, uniform convergence, absolute regularity, density estimation, kernel, bandwidth.

1. Introduction

The nonparametric estimation of a probability density f is an interesting problem in statistical inference and plays an important role in communication theory and pattern recognition. An account of this information can be found in Fukunaga (1972) and Fukunaga and Hostetler (1975). The purpose of this paper is to investigate recursive density estimators when the observations are dependent. Let X_t, $t = \ldots, -1, 0, 1, \ldots$ be a strictly stationary sequence of random variables (r.v.'s) defined on a probability space $(\Omega, \mathscr{F}, P)$ and taking values in R^d. Let $X_1, \ldots, X_n$ be n consecutive observations of X_t. Let f be the density of X_1. As an estimator of $f(x)$ we shall consider

$$(1.1) \qquad \hat{f}_n(x) = n^{-1} \sum_{j=1}^{n} b_j^{-d} K((x - X_j)/b_j) ,$$

which was introduced by Wolverton and Wagner (1969*a*, 1969*b*) and Yamato (1971). Note that $\hat{f}_n$ can be computed recursively by

$$\hat{f}_n(x) = \frac{n-1}{n}\hat{f}_{n-1}(x) + nb_n^{-d}K((x - X_n)/b_n)\,. \tag{1.2}$$

This property is particularly useful in large sample sizes since $\hat{f}_n$ can be easily updated with each additional observation. Here K is a kernel function and b_n is a sequence of bandwidths tending to zero as $n \to \infty$.

Assume the joint densities of $(X_1,\ldots, X_k)$ exist for all $k > 1$ and that X_t satisfies the dependence condition defined below:

DEFINITION 1.1. Let m, k and l be arbitrary positive integers. Let $\boldsymbol{X} = (X_1,\ldots, X_k)$ and $\boldsymbol{Y} = (X_{k+l+1},\ldots, X_{k+l+m})$. Let $f_{X,Y}, f_X, f_Y$ be the densities of $(\boldsymbol{X}, \boldsymbol{Y})$, $\boldsymbol{X}$ and $\boldsymbol{Y}$, respectively. The process X_t is said to satisfy the absolute regularity condition in the locally transitive sense (ARLT) if for some function h

$$\iint_{R^{dk}\times R^{dm}} |f_{X,Y}(\boldsymbol{x},\boldsymbol{y}) - f_X(\boldsymbol{x})f_Y(\boldsymbol{y})|\, d\boldsymbol{x}d\boldsymbol{y} \le h(k,m)\varphi(l) \tag{1.3}$$

where $\varphi(l) \downarrow 0$ as $l \to \infty$. Here $\boldsymbol{x}, \boldsymbol{y}$ denote values of $\boldsymbol{X}$, $\boldsymbol{Y}$.

The letter C will be used to denote constants whose values are unimportant and may vary. We assume throughout the paper that

$$h(k,m) = C(k+m)^\theta$$

for some $\theta \ge 0$. Let

$$\hat{\beta}(k,l,m) = E\{\sup |P(B|\mathscr{F}(X_t\colon 1 \le t \le k)) - P(B)|\}\,, \tag{1.4}$$

where the supremum is taken over all sets $B \in \mathscr{F}(X_t\colon k + l + 1 \le t \le k + l + m)$; and where $\mathscr{F}(X_t\colon 1 \le t \le k)$ and $\mathscr{F}(X_t\colon k + l + 1 \le t \le k + l + m)$ are the σ-fields generated, respectively, by $\{X_t\colon 1 \le t \le k\}$ and $\{X_t\colon k + l + 1 \le t \le k + l + m\}$. Employing Lemma 2 of Ibragimov and Rozanov ((1978), p. 118),

$$\iint_{R^{dk}\times R^{dm}} |f_{X,Y}(\boldsymbol{x},\boldsymbol{y}) - f_X(\boldsymbol{x})f_Y(\boldsymbol{y})|\, d\boldsymbol{x}d\boldsymbol{y} = 2\hat{\beta}(k,l,m)\,. \tag{1.5}$$

It is now easy to see that the ARLT condition is weaker than the absolute regularity condition defined below when the joint densities of $(X_1,\ldots, X_k)$ exist for all $k > 1$.

DEFINITION 1.2. Let $\mathscr{F}_{-\infty}^{0}$ and $\mathscr{F}_{n}^{\infty}$ be the σ-fields generated by $\{X_t: t \leq 0\}$ and $\{X_t: t \geq n\}$ respectively. Then $X(t)$ is absolutely regular if

$$\beta(n) = E\{\sup |P(A|M_{-\infty}^{0}) - P(A)|: A \in \mathscr{F}_{n}^{\infty}\} \downarrow 0 \tag{1.6}$$

as $n \to \infty$.

From Definitions 1.1 and 1.2, it is clear that $\hat{\beta}(k,l,m) \leq \beta(l)$. For relevant literature on absolutely regular processes, the reader is referred to Yoshihara (1976, 1978 and 1984). Pham and Tran (1985) and Pham (1986) have shown that a large class of time series is absolutely regular. In particular, autoregressive moving average time series models and bilinear time series models are absolutely regular with $\beta(n)$ decaying to zero exponentially fast under weak conditions. Thus, autoregressive processes and bilinear time series models satisfy the ARLT condition with $h(k,m) \equiv 1$.

The absolute regularity condition is weaker than the ϕ-mixing condition but is stronger than the strong mixing condition. The definitions of these dependence conditions can be found in Ibragimov and Rozanov (1978).

In the independent case, $\hat{f}_n$ has been thoroughly examined in Wegman and Davies (1979). In the dependent case, quadratic mean convergence and asymptotic normality of $\hat{f}_n$ have been obtained by Masry (1986) under various assumptions on the dependence of X_t. Strong pointwise consistency of $\hat{f}_n$ has been proved by Gyorfi (1981). Takahata (1980) and Masry and Gyorfi (1987) obtained sharp a.s. rates of $\hat{f}_n$ to f for the class of asymptotically uncorrelated processes, the definition of which can be found in Takahata (1977*a*). Masry (1987) established sharp rates of almost sure convergence of $\hat{f}_n$ to f for vector-valued stationary strong mixing processes under weak assumptions on the strong mixing condition. These rates were recently improved by Tran (1989*a*).

The role of the smoothing parameter b_n is crucial in kernel density estimation. The book of Devroye and Gyorfi (1985) points out the prominent role of b_n in the behaviour of kernel type estimators. Our main effort is devoted to finding the appropriate bandwidths and determining the trade-off between the rates at which b_n and $\varphi(n)$ tend to zero. The paper is organized as follows: in Section 2, weak and explicit conditions under which $\hat{f}_n$ is asymptotically normal are found. Section 3 deals with the almost sure convergence of $\hat{f}_n$ to f.

The methods of proof are closely related to those of Yoshihara (1976, 1978 and 1984), Takahata (1980), Masry (1986) and Tran (1989*a*). The conditions under which $\hat{f}_n$ converges to f uniformly on compacts are weaker than those assumed by Tran (1989*a*) and the rates of convergence obtained are sharp. As an example, choose $b_n = C(n^{-1} \log n)^{1/(d+2)}$ where $C > 0$; then under the conditions of Theorem 3.1, the uniform rate of convergence of $\hat{f}_n$

to f on compact sets is of order $(n^{-1}\log n)^{1/(d+2)}$. This is the optimal uniform rate of convergence for nonparametric estimators of a density function (see Stone (1983)). An interesting open problem is to find the best constant C to minimize the L_∞ distance. In the independent case, Silverman (1978) has developed a practical method to determine b_n to give the best possible rate of uniform consistency for nonrecursive kernel density estimators. The reader is referred to Silverman (1978) for further statistical motivation.

Let $J_n = \int |\hat{f}_n(x) - f(x)|\,dx$ be the L_1 distance, where integration is over the entire space. Assume that $K \in L_2$ and X_t is stationary and ergodic, and that there is an integer $m > 0$ such that the conditional distribution of X_m given $\mathscr{F}_{-\infty}^{0}$ is absolutely continuous a.s. Under some additional conditions requiring no information on the dependence structure of X_t, Gyorfi and Masry (1988) have shown that J_n converges to zero a.s. This result is also presented in Gyorfi *et al.* ((1988), Theorem 4.3.1). The conditions on the dependence structure of X_t in Gyorfi and Masry (1988) are quite weak in comparison with those of the present paper. Here, appropriate rates convergence of $\varphi(n)$ to zero are assumed; also, the assumption made earlier that the joint densities of $(X_1,\ldots, X_k)$ exist for all $k > 1$ is itself a dependence condition. Recently, for ARLT processes, Tran (1989*b*) has shown that J_n tends to zero completely under the assumption that $K \in L_1$ with $\varphi(n)$ tending to zero sufficiently fast.

The literature on density estimation under dependence is extensive. The book of Gyorfi *et al.* (1988) gives an in-depth treatment of the subject. For a bibliography and additional background material, the reader is referred to Roussas (1969, 1988), Rosenblatt (1970), Takahata (1977*a*, 1977*b*, 1979 and 1980), Robinson (1983), Yoshihara (1984), Masry (1986), Masry and Gyorfi (1987), Yakowitz (1987), Ioannides and Roussas (1987), Hart and Vieu (1988).

2. Asymptotic normality of $\hat{f}_n$

Let $K_n(x)$ be the averaging kernel defined by $K_n(x) = (1/b_n^d)K(x/b_n)$. Then $\hat{f}_n(x) = (1/n)\sum_{j=1}^{n} K_j(x - X_j)$. Let $\mu_j = E[K_j(x - X_j)]$. By (1.2)

$$\hat{f}_n(x) - E\hat{f}_n(x) = \frac{1}{n}\sum_{j=1}^{n}[K_j(x - X_j) - \mu_j]\,. \tag{2.1}$$

Let $\Delta_j(x) = K_j(x - X_j) - \mu_j$. Then

$$\hat{f}_n(x) - E\hat{f}_n(x) = \frac{1}{n}\sum_{j=1}^{n}\Delta_j(x)\,.$$

LEMMA 2.1. *Let U and V be random variables measurable with respect to $\mathscr{F}(X_t\colon 1 \le t \le k)$ and $\mathscr{F}(X_t\colon k + l + 1 \le t \le k + l + m)$, respectively. Let r, s be positive numbers. Assume that $\| U \|_r < \infty$ and $\| V \|_s < \infty$ where $\| U \|_r = \{E| U |^r\}^{1/r}$.* If $r^{-1} + s^{-1} + h^{-1} = 1$, *then*

$$|EUV - EUEV| \le C\| U \|_r \| V \|_s \{\hat{\beta}(k, l, m)\}^{1/h} .$$

Lemma 2.1 can be found in Yoshihara (1984). Let $\delta > 0$. Assume $E| U |^{2+\delta} < \infty$ and $E| V |^{2+\delta} < \infty$. Then by Lemma 2.1,

$$|EUV - EUEV| \le CE\{| U |^{2+\delta} E| V |^{2+\delta}\}^{1/(2+\delta)} \{\hat{\beta}(k, l, m)\}^{\delta/(2+\delta)} .$$

LEMMA 2.2. *Let $N \ge 1$ be a positive integer. Assume $n = (p + q)r$ for some positive integers p, q* and *r. Let $\{\boldsymbol{\eta}_j, 1 \le j \le r\}$ be a family of N-dimensional random vectors such that for each j $(1 \le j \le r)$, $\boldsymbol{\eta}_j$ is measurable with respect to the σ-field generated by*

$$X(k), \quad (j - 1)(p + q) + 1 \le k \le (j - 1)(p + q) + p .$$

Let $g(\boldsymbol{x}_1,\dots, \boldsymbol{x}_r)$ be a Borel function such that $|g(\boldsymbol{x}_1,\dots, \boldsymbol{x}_r)| \le M$ for some constant M, where $\boldsymbol{x}_1,\dots, \boldsymbol{x}_r$ are N-dimensional vectors. Let $F^{(1)}$ and $F^{(2)}$ be distribution functions of random vectors $(\boldsymbol{\eta}_1,\dots, \boldsymbol{\eta}_j)$ and $(\boldsymbol{\eta}_{j+1},\dots, \boldsymbol{\eta}_r)$, with $1 < j < r$. Then

$$\begin{aligned}(2.2)\qquad &\Big| Eg(\boldsymbol{\eta}_1,\dots, \boldsymbol{\eta}_r) - \int\cdots\int_{R^{Nr}} g(\boldsymbol{x}_1,\dots, \boldsymbol{x}_j, \boldsymbol{x}_{j+1},\dots, \boldsymbol{x}_r)\, dF^{(1)} \\ &\qquad \cdot (\boldsymbol{x}_1,\dots, \boldsymbol{x}_j)\, dF^{(2)}(\boldsymbol{x}_{j+1},\dots, \boldsymbol{x}_r) \Big| \\ &\le 2M\hat{\beta}(n, q, n) \le Cn^{\theta}\varphi(q) .\end{aligned}$$

PROOF. By assumption, $|g(\boldsymbol{x}_1,\dots, \boldsymbol{x}_r)| \le M$. Using Lemma 1 in Yoshihara (1978), the left-hand side of (2.2) is bounded by

$$2ME\{\sup| P(B|\mathscr{F}(\eta_k\colon 1 \le k \le j)) - P(B)|\} ,$$

where the supremum is taken over all sets $B \in \mathscr{F}(\eta_k\colon j + 1 \le k \le r)$; and where $\mathscr{F}(\eta_k\colon 1 \le k \le j)$ and $\mathscr{F}(\eta_k\colon j + 1 \le k \le r)$ are the σ-fields generated, respectively, by $\{\eta_k\colon 1 \le k \le j\}$ and $\{\eta_k\colon j + 1 \le k \le r\}$.

Note that $(\boldsymbol{\eta}_1,\dots, \boldsymbol{\eta}_j)$ is measurable with respect to the σ-field generated by $X_1, X_2,\dots, X_{(j-1)(p+q)+p}$ and $\boldsymbol{\eta}_{j+1},\dots, \boldsymbol{\eta}_r$ is measurable with respect to the σ-field generated by $X_{j(p+q)+1},\dots, X_{(r-1)(p+q)+p}$. Each group contains no more than n r.v.'s variables and the indexes of the r.v.'s in the two groups are far apart by a distance of at least q. Thus clearly,

$$E\{\sup|P(B|\mathscr{F}(\eta_k\colon 1\le k\le j)) - P(B)|\} \le \hat{\beta}(n,q,n)\,,$$

from which the lemma follows.

LEMMA 2.3. *Assume $n=(p+q)r$ for some positive integers p, q and r. Let $\{\eta_j\colon 1\le j\le r\}$ be a family of real valued r.v.'s such that for each j $(1\le j\le r)$, η_j is measurable with respect to the σ-field generated by*

$$X(k),\quad (j-1)(p+q)+1\le k\le (j-1)(p+q)+p\,.$$

Let $\varepsilon>0$. Then

$$P\left[\sum_{i=1}^{r}\eta_i<\varepsilon\right]\le P\left[\sum_{i=1}^{r}Z_i<\varepsilon\right]+2r\hat{\beta}(n,q,n)\,, \tag{2.3}$$

$$P\left[\sum_{i=1}^{r}\eta_i<\varepsilon\right]\ge P\left[\sum_{i=1}^{r}Z_i<\varepsilon\right]-2r\hat{\beta}(n,q,n) \tag{2.4}$$

where $\{Z_i\colon 1\le i\le r\}$ are independent r.v.'s such that for each i $(1\le i\le r)$, Z_i has the same distribution function as that of the random vector η_i.

PROOF. We will prove (2.3) since the proof of (2.4) is similar. Let

$$A=\{(x_1,\dots,x_r)\colon x_1+\cdots+x_r<\varepsilon\}\,.$$

Set

$$g(x_1,\dots,x_r)=\begin{cases}1 & \text{if}\quad (x_1,\dots,x_r)\in A\\ 0 & \text{otherwise}\,.\end{cases}$$

Then $g(x_1,\dots,x_r)$ is bounded by $M=1$. Let $F^{(1)},\dots,F^{(r)}$ be, respectively, the distribution functions of the random vectors $\eta_1,\dots,\eta_r$. Using Lemma 2.2, we get

$$\begin{aligned}P\left[\sum_{i=1}^{r}\eta_i<\varepsilon\right]&=Eg(\eta_1,\dots,\eta_r)\\ &\le\int\cdots\int g(x_1,\dots,x_r)\,dF^{(1)}(x_1)\cdots dF^{(r)}(x_r)+2r\hat{\beta}(n,q,n)\\ &=P\left[\sum_{i=1}^{r}Z_i<\varepsilon\right]+2r\hat{\beta}(n,q,n)\,.\end{aligned} \tag{2.5}$$

ASSUMPTION 1. The kernel function $K \in L_1$, with $\int K(x)dx = 1$ and K has an integrable radial majorant $Q(x)$, that is, $Q(x) \equiv \sup\{|K(y)|: \|y\| \geq \|x\|\}$ is integrable, where $\|x\|$ denotes the Euclidean norm of x and the integration is over the entire space. Assume in addition that K satisfies the following Lipschitz condition:

$$|K(x) - K(y)| \leq C\|x - y\| .$$

Remark 2.1. Note that Assumption 1 implies that $|K|$ is bounded since K is Lipschitz and absolutely integrable.

ASSUMPTION 2. The bandwidth parameter $\{b_n\}$ satisfies

$$(1/n) \sum_{j=1}^{n} (b_n/b_j)^{dr} \to \hat{\theta}_{dr}$$

as $n \to \infty$ for $1 \leq r < 2$.

ASSUMPTION 3. (i) The joint probability density $f(x,y,k)$ of the r.v. X_1 and X_{1+k} exists and satisfies $|f(x,y,k) - f(x)f(y)| \leq M < \infty$ for all x, y and $k \geq 1$.

(ii) Assume X_t satisfies the ARLT condition with $\varphi(n) = O(n^{-\nu})$ for some $\nu > 2$.

The following is an immediate consequence of the Lebesgue Density Theorem (see Devroye and Gyorfi (1985)).

LEMMA 2.4. *Assume that Assumption* 1 *holds. Then*

$$\lim_{n\to\infty} \int_{R^d} K_n(x-u)f(u)du = f(x)\int K(u)du = 1 .$$

LEMMA 2.5. *Assume X_t is a strictly stationary ARLT process satisfying Assumptions* 1–3. *Let x be a point of continuity of f. Suppose that $b_n \to 0$ and $nb_n^d \to \infty$ as $n \to \infty$. Then*

$$\lim_{n\to\infty} nb_n^d \operatorname{var} \hat{f}_n(x) = \hat{\theta}_d f(x) \int K^2(u)du , \tag{2.6}$$

where $\operatorname{var} \hat{f}_n(x)$ *denotes the variance of* $\hat{f}_n(x)$.

PROOF. The proof of Lemma 2.5 is essentially the same as the proof of Theorem 3 of Masry (1986) except here the theorem is proved under the ARLT condition and X_t takes value in R^d instead of R^1. We will therefore just sketch the proof. Note that $\operatorname{var} \hat{f}_n(x) = I_n(x) + R_n(x)$, where

$$(2.7)\qquad \begin{aligned} I_n(x) &= n^{-2} \sum_{i=1}^{n} \operatorname{var} K_i(x - X_i) , \\ R_n(x) &= n^{-2} \sum_{\substack{i=1\\ i\neq j}}^{n} \sum_{j=1}^{n} \operatorname{cov} \{K_i(x - X_i), K_j(x - X_j)\} . \end{aligned}$$

Using Assumptions 1 and 2

$$(2.8)\qquad \lim_{n\to\infty} nb_n^d I_n(x) = \hat{\theta}_d f(x) \int K^2(u)\, du .$$

From (2.7)

$$(2.9)\qquad |R_n(x)| \le 2n^{-2} \sum_{\substack{i=1\\ i>j}}^{n} \sum_{j=1}^{n} |\operatorname{cov} \{K_i(x - X_i), K_j(x - X_j)\}| .$$

Choose $c_n = b_n^{-d(1-\gamma)/\eta}$ where $\eta = -1 - \varepsilon + (1-\gamma)a^{-1}$ where $a = 1/v$ with γ and ε being small positive numbers such that $a^{-1} - (1+\varepsilon)(1-\gamma)^{-1} > 1$. This can be done since $0 < a < 1/2$. Note that

$$\eta > -1 - \varepsilon + (1-\gamma)[1 + (1+\varepsilon)(1-\gamma)^{-1}] = (1-\gamma) .$$

Split the sum above into two regions, S_1 and S_2,

$$(2.10)\qquad \begin{aligned} S_1 &= \{i, j \in \{1,\dots,n\} \colon 1 \le i - j \le c_n\} , \\ S_2 &= \{i, j \in \{1,\dots,n\} \colon c_n + 1 \le i - j \le n - 1\} , \end{aligned}$$

and write the bound as

$$(2.11)\qquad |R_n(x)| \le J_1 + J_2 .$$

Note that $c_n b_n^d = b_n^{-d[((1-\gamma)/\eta)-1]}$, which tends to zero as $n \to \infty$ since $((1-\gamma)/\eta) - 1 < 0$ and $b_n \to 0$ as $n \to \infty$. Using Assumption 3(i)

$$(2.12)\qquad \begin{aligned} J_1 &\le 2Mn^{-2} \left[\int |K(v)|\, dv \right]^2 \sum_{S_1} 1 \\ &= O(c_n/n) = O(c_n b_n^d / nb_n^d) = o(1/(nb_n^d)) . \end{aligned}$$

Let $\delta = 2(1-\gamma)/\gamma$ or $\gamma = 2/(2+\delta)$. Define

$$(2.13)\qquad q_i(x) = \int (1/b_i^d) |K((x-u)/b_i)|^{2+\delta} f(u)\, du ,$$

which tends to $f(x)\int |K(u)|^{2+\delta}du < \infty$ by Lemma 2.4 since $\int |K(u)|^{2+\delta}du$ is finite by Assumption 1.

Using Lemma 2.1 with $r = s = 2 + \delta$ and $h = (2 + \delta)/\delta$ and following a computation similar to (3.25) of Masry (1986), we have

$$J_2 \le Cn^{-2} \sum_{l=c_n+1}^{n-1} [\hat{\beta}(1, l, 1)]^{1-\gamma} \sum_{j=1}^{n-l} \left[\frac{q_{l+j}(x)}{b_{l+j}^{d(1+\delta)}} \right]^{\gamma/2} \left[\frac{q_j(y)}{b_j^{d(1+\delta)}} \right]^{\gamma/2}.$$

Using the Cauchy-Schwarz inequality and Assumption 2, we obtain

$$(2.14)\qquad J_2 \le Cn^{-1}b_n^{-d(2-\gamma)} \left(\sum_{l=c_n}^{\infty} [\varphi(l)]^{1-\gamma} \right) \hat{\theta}_{d(2-\gamma)} \left[\int |K(u)|^{2/\gamma} du \right]^{\gamma} (f(x))^{\gamma}.$$

Thus

$$(2.15)\qquad \begin{aligned} nb_n^d J_2 &\le Cb_n^{-d(1-\gamma)} \sum_{l=c_n}^{\infty} [\varphi(l)]^{1-\gamma} \\ &\le Cb_n^{-d(1-\gamma)} c_n^{-\eta} \sum_{l=c_n}^{\infty} l^{\eta}[\varphi(l)]^{1-\gamma}. \end{aligned}$$

All we need to show is that $\sum_{l=1}^{\infty} l^{\eta}[\varphi(l)]^{1-\gamma} < \infty$. However,

$$(2.16)\qquad \sum_{l=1}^{\infty} l^{\eta}[\varphi(l)]^{1-\gamma} \le C \sum_{l=1}^{\infty} l^{-1-\varepsilon+(1-\gamma)\nu} l^{-(1-\gamma)\nu} = C \sum_{l=1}^{\infty} l^{-1-\varepsilon} < \infty.$$

Finally by (2.12), (2.15) and (2.16), $nb_n^d J_2 \to 0$ and $nb_n^d |R_n(x)| \to 0$ as $n \to \infty$ since $b_n^{-d(1-\gamma)} c_n^{-\eta} = 1$ and $c_n \to \infty$.

ASSUMPTION 4. Suppose $\sup_x |f_j'(x)| < \infty$ for each $1 \le j \le d$, where f_j' denotes the partial derivative of f with respect to x_j, the j-th coordinate of x. Suppose also that $\sum_{i=1}^{n} (b_i/b_n) = O(n)$ as $n \to \infty$.

THEOREM 2.1. (i) *Suppose Assumptions* 1–3 *are satisfied and* $b_n \downarrow 0$ *as* $n \to \infty$ *slowly enough that*

$$(2.17)\qquad nb_n^{d(3-2\gamma)} \to \infty$$

for some $\gamma < 1 - \nu^{-1}$. *Assume also*

$$(2.18)\qquad n^{(\nu-\gamma\nu-1)/2} b_n^{d\{(3-2\gamma)\nu(1-\gamma)-1\}/2} \to \infty$$

and

$$n^{(\nu-2\theta-1)/2} b_n^{d(3\nu-2\gamma\nu+1)/2} \to \infty . \tag{2.19}$$

Then $(nb_n^d)^{1/2}(\hat{f}_n(x) - E\hat{f}_n(x))/\sigma$ *has a standard normal distribution as* $n \to \infty$, *where*

$$\sigma^2 = \hat{\theta}_d f(x) \int K^2(u)\,du .$$

(ii) *If in addition, Assumption* 4 *is satisfied and* $nb_n^{d+2} \to 0$, *then* $(nb_n^d)^{1/2}(\hat{f}_n(x) - f(x))/\sigma$ *has a standard normal distribution as* $n \to \infty$.

PROOF OF THEOREM 2.1(i). Define $Y_j = b_j^{d/2}\Delta_j(x)$. Then

$$S_n = nb_n^{d/2}(\hat{f}_n(x) - E\hat{f}_n(x)) = \sum_{j=1}^{n} b_j^{d/2}\Delta_j(x) = \sum_{j=1}^{n} Y_j . \tag{2.20}$$

Let $\omega(n)$ be a positive function increasing to infinity sufficiently slowly that

$$(nb_n^{d(3-2\gamma)})^{1/2}(\omega(n))^{-2} \to \infty ,$$

$$n^{(-\nu+\gamma\nu+1)/2} b_n^{d\{(3-2\gamma)\nu(-1+\gamma)+1\}/2}(\omega(n))^{2\{(1-\gamma)\nu-1\}} \to 0 , \tag{2.21}$$

$$n^{(2\theta+1-\nu)/2} b_n^{-d(3\nu-2\gamma\nu+1)/2}(\omega(n))^{1+2\nu} \to 0 . \tag{2.22}$$

Choose $q = q_n = [(nb_n^{d(3-2\gamma)})^{1/2}(\omega(n))^{-2}]$ and $p = p(n) = [(nb_n^d)^{1/2}/\omega(n)]$, where $[a]$ denotes the integer part of a. Then

$$q/p = b_n^{d(1-\gamma)}(\omega(n))^{-1} = o(1) . \tag{2.23}$$

Assume for now that $n/(p+q) = r$ where r is a positive integer. If n is not a multiple of $p+q$ then the proof can be altered, but the result remains valid as will be pointed out later. We now set the r.v.'s Y_j's in alternate blocks of size p and q. Let

$$U(n,x,j) = \sum_{i=(j-1)(p+q)+1}^{(j-1)(p+q)+p} Y_i(x) ,$$

$$V(n,x,j) = \sum_{i=(j-1)(p+q)+p+1}^{j(p+q)} Y_i(x) , \tag{2.24}$$

where $j = 1, \ldots, r$. Let

$$S_n' = \sum_{j=1}^{r} U(n,x,j) \quad \text{and} \quad S_n'' = \sum_{j=1}^{r} V(n,x,j) , \tag{2.25}$$

be the sum of r.v.'s Y_j in large blocks and small blocks, respectively. Let $\varepsilon > 0$ and let $\sigma^2(S_n)$ denote the variance of S_n. It is easy to see that,

$$(2.26)\quad \begin{aligned} P[S_n/\sigma(S_n) \le x] &\le P[|S_n''|/\sigma(S_n) \le \varepsilon,\ S_n/\sigma(S_n) < x] \\ &\quad + P[|S_n''|/\sigma(S_n) > \varepsilon] \\ &\le P[S_n - S_n''/\sigma(S_n) \le \varepsilon + x] + P[|S_n''|/\sigma(S_n) > \varepsilon] \\ &= P[S_n'/\sigma(S_n) \le \varepsilon + x] + P[|S_n''|/\sigma(S_n) > \varepsilon]\,. \end{aligned}$$

By (2.3) of Lemma 2.3,

$$(2.27)\quad P[S_n'/\sigma(S_n) \le \varepsilon + x] \le P\left[\sum_{i=1}^{r} Z_{ni} \le x + \varepsilon\right] + 2r\hat{\beta}(n, q, n)$$

where Z_{ni} $(1 \le i \le r)$ are independent r.v.'s such that for each i, Z_{ni} has the same distribution as that of $U(n, x, i)/\sigma(S_n)$. By (2.26) and (2.27),

$$(2.28)\quad \begin{aligned} &P[S_n/\sigma(S_n) \le x] \\ &\quad \le P\left[\sum_{i=1}^{r} Z_{ni} \le x + \varepsilon\right] + P[|S_n''|/\sigma(S_n) \ge \varepsilon] + 2r\hat{\beta}(n, q, n)\,. \end{aligned}$$

Similarly, using (2.4), it can be shown that

$$(2.29)\quad \begin{aligned} &P[S_n/\sigma(S_n) \le x] \\ &\quad \ge P\left[\sum_{i=1}^{r} Z_{ni} \le x - \varepsilon\right] - P[|S_n''|/\sigma(S_n) \ge \varepsilon] - 2r\hat{\beta}(n, q, n)\,. \end{aligned}$$

By a simple computation, we have

$$(2.30)\quad \begin{aligned} r\hat{\beta}(n, q, n) &\le Cn(p+q)^{-1} n^{\theta} q^{-\nu} \\ &\le Cn(p+q)^{-1} n^{\theta} \{(n b_n^{d(3-2\gamma)})^{1/2} (\omega(n))^{-2}\}^{-\nu} \\ &\le C n^{1+\theta-(\nu/2)} (p+q)^{-1} (\omega(n))^{2\nu} b_n^{-d\nu(3-2\gamma)/2} \\ &\le C n^{1+\theta-(\nu/2)} (n b_n^{d})^{-1/2} (\omega(n))^{1+2\nu} b_n^{-d\nu(3-2\gamma)/2} \\ &\le C n^{(2\theta+1-\nu)/2} (\omega(n))^{1+2\nu} b_n^{-d(3\nu-2\gamma\nu+1)/2} = o(1)\,, \end{aligned}$$

by (2.22).

Using (2.25)

(2.31) $$(1/n)E|S_n''|^2 = (1/n)\sum_{i=1}^{r} E[V(n,x,i)]^2 + (2/n)\sum_{\substack{i=1\\ i>j}}^{r}\sum_{j=1}^{r} \operatorname{cov}\{V(n,x,i), V(n,x,j)\}.$$

Employing (2.24)

(2.32) $$E[V(n,x,j)]^2 = \sum_{i=1}^{q} E[Y_{(j-1)(p+q)+p+i}]^2 + \sum_{\substack{i=1\\ i\neq l}}^{q}\sum_{l=1}^{q} \operatorname{cov}\{Y_{(j-1)(p+q)+p+i}, Y_{(j-1)(p+q)+p+l}\}.$$

By a simple computation using Lemma 2.4,

(2.33) $$E[Y_{(j-1)(p+q)+p+i}]^2 \le M_1(x),$$

for some constant $M_1(x)$ independent of i. Note that $\sum_{l=1}^{\infty}[\varphi(l)]^{1-\gamma} < \infty$ since $\gamma < 1 - \nu^{-1}$. By Lemma 2.4

(2.34) $$\operatorname{cov}\{Y_{(j-1)(p+q)+p+k}, Y_{(j-1)(p+q)+p+l}\} \le C[\varphi(|k-l|)^{1-\gamma}]\|Y_{(j-1)(p+q)+p+k}\|_{2/\gamma}\|Y_{(j-1)(p+q)+p+l}\|_{2/\gamma}.$$

Again, by Lemma 2.4,

(2.35) $$\max \|Y_i\|_{2/\gamma} \le [M_2(x)b_n^{d(1-(1/\gamma))}]^{\gamma/2},$$

for some constant $M_2(x)$ independent of i. Therefore

(2.36) $$\begin{aligned}(1/n)\sum_{i=1}^{r} E[V(n,x,i)]^2 &\le (1/n)\sum_{i=1}^{r}\sum_{k=1}^{q} M_1(x) \\ &\quad + C(1/n)[M_2(x)b_n^{d(1-(1/\gamma))}]^{\gamma}\sum_{i=1}^{r}\sum_{k=1}^{q}\sum_{l=1}^{q}[\varphi(|k-l|)]^{1-\gamma} \\ &\le M_1(x)n^{-1}rq + CM_2(x)n^{-1}b_n^{d(\gamma-1)}rq\sum_{l=1}^{\infty}[\varphi(l)]^{1-\gamma} \\ &\le M_1(x)n^{-1}rq + CM_2(x)n^{-1}b_n^{d(\gamma-1)}rq,\end{aligned}$$

since

$$\sum_{l=1}^{\infty} [\varphi(l)]^{1-\gamma} \leq C \sum_{l=1}^{\infty} l^{-\nu} < \infty .$$

Clearly

$$(2.37) \qquad 2 \sum_{\substack{i=1\\ i>j}}^{r} \sum_{j=1}^{r} \operatorname{cov}\{V(n,x,i), V(n,x,j)\}$$
$$= 2 \sum_{\substack{i=1\\ i>j}}^{r} \sum_{j=1}^{r} \sum_{k=(j-1)(p+q)+p+1}^{j(p+q)} \sum_{l=(i-1)(p+q)+p+1}^{i(p+q)} \operatorname{cov}\{Y_k, Y_l\} .$$

For $i>j$, the indices k, l differ by at least p, so by (2.35) and Lemma 2.1

$$(2.38) \qquad 2 \sum_{\substack{i=1\\ i>j}}^{r} \sum_{j=1}^{r} \operatorname{cov}\{V(n,x,i), V(n,x,j)\}$$
$$\leq 4 \sum_{k=1}^{n-p} \sum_{l=k+p}^{n} |\operatorname{cov}\{Y_k, Y_l\}|$$
$$\leq C \sum_{k=1}^{n-p} \sum_{l=k+p}^{n} [\varphi(l-k)]^{1-\gamma} \| Y_k \|_{2/\gamma} \| Y_l \|_{2/\gamma}$$
$$\leq C M_2^{\gamma}(x) b_n^{d(\gamma-1)} \sum_{k=1}^{n-p} \sum_{l=k+p}^{n} [\varphi(l-k)]^{1-\gamma}$$
$$\leq C M_2^{\gamma}(x) b_n^{d(\gamma-1)} n \sum_{k=p}^{\infty} [\varphi(k)]^{1-\gamma} .$$

By (2.31), (2.36) and (2.38),

$$(2.39) \quad (1/n)E|S_n''|^2 \leq Cn^{-1}rq + Cn^{-1}b_n^{d(\gamma-1)}rg + Cb_n^{d(\gamma-1)} \sum_{k=p}^{\infty} [\varphi(k)]^{1-\gamma} .$$

Since $q/p \to 0$ as $n \to \infty$,

$$(2.40) \qquad rq/n = nq/((p+q)n) = q/(p+q) = o(1) .$$

By (2.23)

$$(2.41) \qquad n^{-1}b_n^{d(\gamma-1)}rq = n^{-1}b_n^{d(\gamma-1)}n(p+q)^{-1}q = b_n^{d(\gamma-1)}(p+q)^{-1}q$$
$$\leq b_n^{d(\gamma-1)}p^{-1}q = 1/\omega(n) \to 0$$

as $n \to \infty$.

Since $q \leq p$, by (2.21)

$$(2.42)\quad b_n^{d(\gamma-1)} \sum_{k=p}^{\infty} [\varphi(k)]^{1-\gamma}$$
$$\le C b_n^{d(\gamma-1)} q^{-(1-\gamma)\nu+1} = C b_n^{d(\gamma-1)} [(n b_n^{d(3-2\gamma)})^{1/2} (\omega(n))^{-2}]^{-(1-\gamma)\nu+1}$$
$$\le C b_n^{d(\gamma-1)} \{(n b_n^{d(3-2\gamma)})^{1/2}\}^{-(1-\gamma)\nu+1} (\omega(n))^{2\{(1-\gamma)\nu-1\}}$$
$$\le C b_n^{d(\gamma-1)} n^{(-\nu+\gamma\nu+1)/2} b_n^{d(3-2\gamma)(-\nu+\gamma\nu+1)/2} (\omega(n))^{2\{(1-\gamma)\nu-1\}}$$
$$\le C n^{(-\nu+\gamma\nu+1)/2} b_n^{d\{(3-2\gamma)\nu(-1+\gamma)+1\}/2} (\omega(n))^{2\{(1-\gamma)\nu-1\}} = o(1)\,.$$

Finally, from (2.39)–(2.42)

$$(2.43)\qquad (1/n) E|S_n''|^2 \to 0\,,$$

which entails

$$(2.44)\qquad P[|S_n''| > \varepsilon n^{1/2}] = o(1)\,.$$

Lemma 2.3 implies

$$(2.45)\qquad \lim_{n\to\infty} (\sigma^2(S_n)/n) = n b_n^d \operatorname{var}(\hat f_n(x)) = \sigma^2\,.$$

It follows from (2.44) and (2.45) that

$$(2.46)\qquad P[|S_n''| > \varepsilon\sigma(S_n)] = o(1)\,.$$

Since $S_n = S_n' + S_n''$, by (2.43) and (2.45),

$$(2.47)\qquad (1/n) E S_n'^2 \to \sigma^2\,.$$

Note that

$$(2.48)\qquad (1/n) E|S_n'|^2 = (1/n) \sum_{j=1}^{r} E U^2(n,x,j) + (2/n) \sum_{\substack{i=1\\ i>j}}^{r} \sum_{j=1}^{r} \operatorname{cov}\{U(n,x,i) U(n,x,j)\}\,.$$

Similar to the proof of (2.41),

$$(2.49)\qquad (2/n) \sum_{\substack{i=1\\ i>j}}^{r} \sum_{j=1}^{r} \operatorname{cov}\{U(n,x,i) U(n,x,j)\} \le C b_n^{d(\gamma-1)} \sum_{k=q}^{\infty} [\varphi(k)]^{1-\gamma} = o(1)$$

by (2.42).

Employing (2.47), (2.48) and (2.49),

$$(2.50)\qquad (1/n)\sum_{j=1}^{r} EU^2(n,x,j) \to \sigma^2 ,$$

which together with (2.45) gives

$$(2.51)\qquad (1/\sigma^2(S_n))\sum_{i=1}^{r} EU^2(n,x,i) \to 1 .$$

Note that

$$(2.52)\qquad \begin{aligned} |Y_j| &= b_j^{d/2}\Delta_j(x) \\ &= b_j^{d/2}\left| b_j^{-d}K((x-X_j)/b_j) - \int K_j(x-u)f(u)du \right| \\ &\le C(b_j^{-d/2} + b_j^{d/2}) \end{aligned}$$

since $\int K_j(x-u)f(u)du \to f(x)$ as $j \to \infty$ and $|K|$ is bounded by Remark 2.1. Hence $\max_{1\le j\le n} |Y_j| \le Cb_n^{-d/2}$ and

$$(2.53)\qquad \max_{1\le j\le r} |U(n,x,j)| \le Cpb_n^{-d/2} ,$$

which entails

$$(2.54)\qquad \begin{aligned} &(n\sigma^2)^{-1}\sum_{i=1}^{r} E[U^2(n,x,i)I\{|U(n,x,i)| > \eta n^{1/2}\sigma\}] \\ &\le (n\sigma^2)^{-1}p^2 b_n^{-d} r \max_{1\le i\le r} P[|U(n,x,i)| > \eta n^{1/2}\sigma] \to 0 , \end{aligned}$$

since

$$pb_n^{-d/2}(\eta n^{1/2}\sigma)^{-1} \le Cp(nb_n^d)^{-1/2} \to 0$$

by the choice of p. Here $I\{A\}$ denotes the indicator of the set A.

By (2.45) and (2.54),

$$(2.55)\qquad (1/\sigma^2(S_n))\sum_{i=1}^{r} E[U^2(n,x,i)I\{|U(n,x,i)| \ge \eta\sigma(S_n)\}] \to 0 ,$$

or equivalently

$$(2.56)\qquad \sum_{i=1}^{r} E[Z_{ni}^2 I\{|Z_{ni}| > \eta\}] \to 0 ,$$

where the Z_{ni}'s are defined in (2.27). Let

$$s_{nr} = \sigma(Z_{n1} + \cdots + Z_{nr}) . \tag{2.57}$$

By (2.45) and (2.51)

$$\lim_{r\to\infty} s_{nr} = 1 . \tag{2.58}$$

Employing (2.56) and (2.58)

$$\sum_{i=1}^{r} E[(Z_{ni}/s_{nr})^2 I[|Z_{ni}| > \eta s_{nr}]] \to 0 . \tag{2.59}$$

Then by the Lindeberg-Feller theorem

$$P\left[\sum_{i=1}^{r} (Z_{ni}/s_{nr}) \le x + \varepsilon\right] \to \Phi(x + \varepsilon) . \tag{2.60}$$

Since ε can be chosen arbitrarily small, by (2.28), (2.29), (2.30), (2.44) and (2.60),

$$P[S_n/\sigma(S_n) \le x] \to \Phi(x) . \tag{2.61}$$

The proof of (i) is completed by (2.20) and Lemma 2.5.

Finally, suppose n is not an integer multiple of $p + q$. Let $r = [n/(p + q)]$. Define

$$S_n''' = Y_{r(p+q)+1} + \cdots + Y_n .$$

It is not hard to show that $P[S_n'''/\sigma(S_n) > \varepsilon] \to 0$. The proof of the theorem can be completed in the same manner as in the case $n = (p + q)r$ since $S_n = S_n' + S_n'' + S_n'''$.

PROOF OF THEOREM 2.1(ii). We have

$$\begin{aligned} E\hat{f}_n(x) &= n^{-1} \sum_{i=1}^{n} b_i^{-d} \int K((x-y)/b_i) f(y)\,dy \\ &= n^{-1} \sum_{i=1}^{n} \int K(z) f(x - b_i z)\,dz \\ &= n^{-1} \sum_{i=1}^{n} \int K(z)[f(x - b_i z) - f(x) + f(x)]\,dz \\ &= f(x) + n^{-1} \sum_{i=1}^{n} \int K(z)[f(x - b_i z) - f(x)]\,dz . \end{aligned} \tag{2.62}$$

By Taylor's theorem

$$(2.63)\quad |E\hat{f}_n(x) - f(x)| \le n^{-1} \sum_{i=1}^{n} \int |K(z)|\,|f(x - b_i z) - f(x)|\,dz$$
$$\le \sup_{1\le j\le d} \sup_{x} |f_j'(x)| n^{-1} \sum_{i=1}^{n} b_i \int \left(\sum_{j=1}^{d} |z_j| \right) |K(z)|\,dz$$
$$\le C \sup_{1\le j\le d} \sup_{x} |f_j'(x)| n^{-1} \sum_{i=1}^{n} b_i \int \|z\|\,|K(z)|\,dz$$
$$\le C n^{-1} \sum_{i=1}^{n} b_i = C b_n n^{-1} \sum_{i=1}^{n} (b_i/b_n) \le C b_n$$

by Assumption 4. Therefore

$$(nb_n^d)^{1/2} |E\hat{f}_n(x) - f(x)|/\sigma \le C(nb_n^d)^{1/2} b_n = Cn^{1/2} b_n^{(d/2)+1},$$

which tends to zero since $nb_n^{d+2} \to 0$ by assumption.

Example 2.1. Let X_t be a stationary autoregressive process of order 1, that is, $X_t = aX_{t-1} + e_t$ where $|a| < 1$. Assume the e_t's are i.i.d. and each e_t has a Cauchy density with density symmetric about zero. Then X_t satisfies the ARLT condition with $\theta = 0$ and $\varphi(n) = O(e^{-sn})$ for some $s > 0$. Hence $\varphi(n) = O(n^{-\nu})$ for all $\nu > 0$. It is not hard to show that Assumption 3 is satisfied. Consider the case where $b_n = n^{-\rho}$ for some $\rho > 0$. Then (2.17), (2.18) and (2.19) are satisfied if for some arbitrarily large ν and some $\gamma < 1 - \nu^{-1}$,

$$(2.64)\qquad \rho < \frac{1}{d(3 - 2\gamma)},$$

$$(2.65)\qquad \rho < \frac{\nu - \gamma\nu - 1}{d\{(3 - 2\gamma)\nu(1 - \gamma) - 1\}}$$

and

$$(2.66)\qquad \rho < \frac{\nu - 1}{d(3\nu - 2\gamma\nu + 1)}.$$

By letting $\nu \to \infty$ and $\gamma \to 1$, it is seen that (2.64)–(2.66) are satisfied if $\rho < 1/d$. The condition $nb_n^{d+2} \to 0$ is satisfied if $\rho > 1/(d + 2)$.

3. Uniform convergence of $\hat{f}_n$

LEMMA 3.1. *Assume the conditions of Lemma* 2.5 *hold and in addition f is continuous on R^d. Let D be a compact subset of R^d. Let $I_n(x)$ be defined as in Lemma* 2.4 *and let*

$$R_n^*(x) = n^{-2} \sum_{i=1}^{n} \sum_{\substack{j=1 \\ i \neq j}}^{n} |\text{cov}\,\{K_i(x - X_i),\, K_j(x - X_j)\}|\,.$$

Then for some constant $C > 0$,

$$nb_n^d I_n(x) \leq C \quad \textit{and} \quad nb_n^d |R_n^*(x)| \leq C \quad \textit{for all} \quad x \in D\,.$$

Lemma 3.1 can be obtained by the same argument as of Lemma 2.5 by noting that f is uniformly continuous on any compact subset of R^d.

LEMMA 3.2. *Assume* $n = 2pq$ *for some positive integers* p, q. *Let* $\{(\eta_{j1},\ldots,\eta_{jN}),\ 1 \leq j \leq q\}$ *be a family of N-dimensional random vectors such that for each* j $(1 \leq j \leq q)$, $\boldsymbol{\eta}_j = (\eta_{j1},\ldots,\eta_{jN})$ *is measurable with respect to the* σ*-field generated by*

$$X(k), \quad 2(j-1)p - p + 1 \leq k \leq 2(j-1)p\,.$$

Let $\varepsilon > 0$. *Then*

$$P\left[\max_{1 \leq j \leq N} \left|\sum_{i=1}^{q} \eta_{ij}\right| > \varepsilon\right] \leq P\left[\max_{1 \leq j \leq N} \left|\sum_{i=1}^{q} Z_{ij}\right| > \varepsilon\right] + 2q\hat{\beta}(n,p,n)$$

where $\{(Z_{j1},\ldots,Z_{jN}),\ 1 \leq j \leq q\}$ *are independent random vectors such that for each* j $(1 \leq j \leq q)$, $\boldsymbol{Z}_j = (Z_{j1},\ldots,Z_{jN})$ *has the same distribution function as that of the random vector* $\boldsymbol{\eta}_j$.

Lemma 3.2 is essentially the same as Lemma 3.1 of Yoshihara (1984). The proof can be easily obtained from Lemma 2.1.

LEMMA 3.3. *Suppose Assumptions* 1 *and* 2 *hold and in addition* f *is continuous on* R^d. *Suppose* $b_n \downarrow 0$ *slowly enough that*

$$nb_n^d(\log n)^{-1} \to \infty\,. \tag{3.1}$$

Let $\Psi(n) = (\log n)^{1/2}(nb_n^d)^{-1/2}$ *and* $\lambda_n = (nb_n^d \log n)^{1/2}$. *Then for any compact subset D of* R^d, $\sup_{x \in D} |\hat{f}_n(x) - E\hat{f}_n(x)| = O(\Psi(n))$ *a.s. as* $n \to \infty$ *if* b_n *tends to zero in a manner that*

$$(\lambda_n g(n)/b_n^d) n^\theta \varphi([(nb_n^d/\lambda(n)g(n))])h(n) \to 0\,, \tag{3.2}$$

for some function $g(n)$ *increasing to infinity arbitrarily slowly, and some function* $h(n) > 0$ *with* $\sum_{n=1}^{\infty} 1/h(n) < \infty$.

PROOF. Let

$$l = n^{-(1/d)-(3/(2\delta))}(\log n)^{3/(2\delta)+(1/d)} b_n^{-d/(2\delta)} . \tag{3.3}$$

Since D is compact, it can be covered by, say, υ cubes I_k with sides parallel to the coordinate axes and with center at x_k. Now

$$\begin{aligned} \sup_{x \in D} |\hat{f}_n(x) - E\hat{f}_n(x)| \le & \max_{1 \le k \le \upsilon} \sup_{x \in I_k} |\hat{f}_n(x) - \hat{f}_n(x_k)| \\ & + \max_{1 \le k \le \upsilon} |\hat{f}_n(x_k) - E\hat{f}_n(x_k)| \\ & + \max_{1 \le k \le \upsilon} \sup_{x \in I_k} |E\hat{f}_n(x_k) - E\hat{f}_n(x)| . \end{aligned} \tag{3.4}$$

For $x \in I_k$, using the Lipschitz condition of K and noting that $b_j^d > C(\log j)/j$ by (3.1),

$$\begin{aligned} |\hat{f}_n(x) - \hat{f}_n(x_k)| & \le (C/n) \sum_{j=1}^{n} b_j^{-d}(\|x - y\|/b_j)^{\delta} \\ & \le Cl^{\delta} n^{-1} \sum_{j=2}^{n} (j/\log j)^{1+(\delta/d)} \le Cl^{\delta} n^{1+(\delta/d)}(\log n)^{-1-(\delta/d)} \\ & = O(\Psi(n)) \end{aligned} \tag{3.5}$$

a.s. as $n \to \infty$. Therefore

$$\begin{aligned} & \max_{1 \le k \le \upsilon} \sup_{x \in I_k} |\hat{f}_n(x) - \hat{f}_n(x_k)| = O(\Psi(n)) \quad \text{a.s. as } n \to \infty , \\ & \max_{1 \le k \le \upsilon} \sup_{x \in I_k} |E\hat{f}_n(x_k) - E\hat{f}_n(x)| = O(\Psi(n)) \quad \text{a.s. as } n \to \infty . \end{aligned} \tag{3.6}$$

It remains to show that

$$\max_{1 \le k \le \upsilon} |\hat{f}_n(x_k) - E\hat{f}_n(x_k)| = O(\Psi(n)) . \tag{3.7}$$

Assume $n = 2pq$ for some increasing integer valued functions $p = p(n)$, $q = q(n)$. Then the r.v.'s Δ_j's defined in (2.1) can be grouped successively into $2q$ blocks of size p. Let $S(n, x) = \hat{f}_n(x) - E\hat{f}_n(x) = (1/n) \sum_{i=1}^{n} \Delta_i(x)$. Write $S(n, x)$ as

$$S(n, x) = S(n, x, 1) + S(n, x, 2) ,$$

where

$$S(n,x,1)=\sum_{j=1}^{q} V(n,x,2(j-1)), \qquad S(n,x,2)=\sum_{j=1}^{q} V(n,x,2j-1)$$

with

$$V(n,x,j)=(1/n)\sum_{i=(j-1)p+1}^{jp} \Delta_i(x) \qquad (j=1,\dots,q)\,.$$

Note that $S(n,x,1)$ and $S(n,x,2)$ are, respectively, the sum of the even-numbered and odd-numbered groups. If it is not the case that $n=2pq$, then the last blocks of $S(n,x,1)$ and $S(n,x,2)$ can be shorter than p but this does not effect the proofs of the results, as will be seen. Let $\varepsilon_n=\eta\Psi(n)$, where η is a large number to be specified later. Observe that

$$\text{(3.8)} \qquad \begin{aligned} &P\Big[\max_{1\le k\le v}|\hat{f}_n(x_k)-E\hat{f}_n(x_k)|>\varepsilon_n\Big] \\ &\quad = P\Big[\max_{1\le k\le v}|S(n,x_k,1)+S(n,x_k,2)|>\varepsilon_n\Big] \\ &\quad \le P\Big[\max_{1\le k\le v}|S(n,x_k,1)|>\varepsilon_n/2\Big] \\ &\qquad + P\Big[\max_{1\le k\le v}|S(n,x_k,2)|>\varepsilon_n/2\Big]. \end{aligned}$$

Since $|K|$ is bounded by Remark 2.1, we have

$$\text{(3.9)} \qquad \begin{aligned} V(n,x,j) &\le (1/n)\sum_{i=(j-1)p+1}^{jp} |b_i^{-d}[K((x-X_j)/b_j)-EK((x-X_j)/b_j)]| \\ &\le Cn^{-1}\sum_{i=(j-1)p+1}^{jp} b_i^{-d}\,. \end{aligned}$$

By (3.1) and (3.2), there exists a function $g^*(n)$ increasing to infinity such that

$$nb_n^d(\log n)^{-1}/g^*(n)\to\infty\,,$$

and in addition (3.2) is satisfied with $g(n)\equiv g^*(n)$. Choose

$$\text{(3.10)} \qquad p=[nb_n^d/(\lambda_n g^*(n))]\ge C(nb_n^d)^{1/2}(\log n)^{-1/2}/g^*(n)\,.$$

Then $p\to\infty$. By Assumption 2 and (3.9),

$$\text{(3.11)} \quad \lambda_n|V(n,x,j)|\le C\lambda_n n^{-1}b_n^{-d}\sum_{i=(j-1)p+1}^{jp}(b_n/b_i)^d\le C\lambda_n pn^{-1}b_n^{-d}=1/g^*(n)\,,$$

which tends to zero. We now approximate ARLT r.v.'s by independent ones. By Lemma 3.2

$$P\Big[\max_{1\le k\le \upsilon} |S(n,x_k,1)| > \varepsilon_n/2\Big] \le P\Big[\max_{1\le k\le \upsilon}\Big|\sum_{j=1}^{q} V_j^*(x_k)\Big| > \varepsilon_n/2\Big] + 4q\hat{\beta}(n,p,n)\,, \tag{3.12}$$

where the $\{(V_j^*(x_1),\ldots, V_j^*(x_\upsilon), j=1,\ldots,q)\}$ is a family of independent υ-dimensional random vectors such that for each j, the random vector $(V_j^*(x_1),\ldots, V_j^*(x_\upsilon))$ has the same distribution as that of $\{V(n,x_1,2(j-1)), \ldots, V(n,x_\upsilon,2(j-1))\}$. Note that $|\lambda_n V_j^*(x)| \le 1/2$ for large n. Hence there exists an N such that

$$\begin{aligned} &\exp \lambda_n V_j^*(x) \le 1 + \lambda_n V_j^*(x) + (V_j^*(x))^2\lambda_n^2\,, \\ &\exp(-\lambda_n V_j^*(x)) \le 1 - \lambda_n V_j^*(x) + (V_j^*(x))^2\lambda_n^2 \end{aligned} \tag{3.13}$$

for $n > N$. Using the independence of the $V_j^*(x)$'s and applying Markov's inequality separately to the summands $\sum_{j=1}^{q} V_j^*(x)$ and $\sum_{j=1}^{q} - V_j(x)$, we obtain

$$\begin{aligned} P\Big[\Big|\sum_{j=1}^{q} V_j^*(x)\Big| > \varepsilon_n\Big] &\le P\Big[\sum_{j=1}^{q} V_j^*(x) > \varepsilon_n\Big] + P\Big[\sum_{j=1}^{q} - V_j^*(x) > \varepsilon_n\Big] \\ &\le e^{-\lambda_n\varepsilon_n}\Big(\prod_{j=1}^{q} E[\exp \lambda_n V_j^*(x)] \\ &\qquad + \prod_{j=1}^{q} E[\exp(-\lambda_n V_j^*(x))]\Big) \\ &\le 2\exp\Big(-\lambda_n\varepsilon_n + \lambda_n^2 \sum_{j=1}^{q} E(V_j^*(x))^2\Big)\,. \end{aligned} \tag{3.14}$$

Clearly

$$\sum_{j=1}^{q} E(V_j^*(x))^2 \le n^{-2}\Big[\sum_{k=1}^{n} E\Delta_k^2 + 2\sum\sum_{1\le k<l\le n} |E\Delta_k\Delta_l|\Big] = I_n + R_n^*\,. \tag{3.15}$$

From (3.15) and Lemma 3.1

$$\sum_{j=1}^{q} E(V_j^*(x))^2 \le C/(nb_n^d)\,. \tag{3.16}$$

By (3.12), (3.14) and (3.16)

$$P\left[\max_{1\le k\le v} |S(n, x_k, 1)| > \varepsilon_n/2\right] \le Cv \exp\left[-\lambda_n\varepsilon_n/2 + C\lambda_n^2/(nb_n^d)\right] + 4q\hat{\beta}(n,p,n)\,. \tag{3.17}$$

Similarly, $P\left[\max_{1\le k\le v} |S(n, x_k, 2)| > \varepsilon_n/2\right]$ is bounded by the right-hand side of (3.17). Now $\lambda_n\varepsilon_n = \eta \log n$ and $\lambda_n^2/(nb_n^d) = \log n$. From (3.8) and (3.17)

$$\begin{aligned} P\left[\max_{1\le k\le v} |\hat{f}_n(x_k) - E\hat{f}_n(x_k)| > \varepsilon_n\right] &\le Cv \exp\left[-\lambda_n\varepsilon_n/2 + C\lambda_n^2/(nb_n^d)\right] + 4q\hat{\beta}(n,p,n) \\ &= Cv \exp\left[-(\eta/2)\log n + C\log n\right] + 4q\hat{\beta}(n,p,n) \\ &= Cvn^{-((\eta/2)+C)} + 4q\hat{\beta}(n,p,n)\,. \end{aligned} \tag{3.18}$$

From (3.3)

$$v \le Cn^{1+(3d/(2\delta))}(\log n)^{-(3d/(2\delta))-1}b_n^{d/(2\delta)}\,. \tag{3.19}$$

By (3.1)

$$b_n > C(n^{-1}\log n)^{1/d}\,. \tag{3.20}$$

Using (3.19) and (3.20), it is easy to show that for sufficiently large η

$$\sum_{n=1}^{\infty} vn^{-((\eta/2)+C)} < \infty\,. \tag{3.21}$$

Note that (3.2) implies that $\sum_{n=1}^{\infty} q\hat{\beta}(n,p,n)$ is finite. The proof of the lemma follows by the Borel-Cantelli Lemma from (3.2), (3.18) and (3.21).

THEOREM 3.1. (i) *Assume that* $\varphi(n) = O(n^{-\nu})$ *for some* $\nu > 3$. *Suppose Assumptions* 1 *and* 2 *hold and* $b_n \downarrow 0$ *in a manner that*

$$n^{((\nu-3)/2)-\theta}(\log n)^{-(3+\nu)/2}(b_n)^{(1+\nu)d/2}(\operatorname{loglog} n)^{-(1+\varepsilon)} \to \infty \tag{3.22}$$

for some $\varepsilon > 0$. *Then* $\sup_{x\in D} |\hat{f}_n(x) - E\hat{f}_n(x)| = O(\Psi(n))$ *a.s. for any compact subset* D *of* R^d.

(ii) *If in addition, Assumption* 4 *holds and* $n(\log n)^{-1}b_n^{d+2} = O(1)$, *then*

$$\sup_{x \in D} |\hat{f}_n(x) - f(x)| = O(\Psi(n)) \qquad a.s.$$

Remark 3.1. Since $b_n > 0$, for (3.22) to be satisfied, it is necessary that $((\nu - 3)/2) - \theta > 0$ or $\nu > 3 + 2\theta$.

PROOF. (i) By (3.22), there exists a function $g(n)$ increasing to infinity such that

$$n^{(3-\nu)/2+\theta}(\log n)^{(3+\nu)/2}(b_n)^{-(1+\nu)d/2}(\log\log n)^{(1+\varepsilon)}(g(n))^{1+\nu} \to 0 \,. \tag{3.23}$$

By (3.23)

$$([nb_n^d \log n]^{1/2} g(n)/b_n^d) n^\theta \cdot [nb_n^d/((nb_n^d \log n)^{1/2} g(n))]^{-\nu} n \log n (\log\log n)^{(1+\varepsilon)} \to 0 \,,$$

which implies (3.2) of Lemma 3.2 with $h(n) = n \log n(\log\log n)^{1+\varepsilon}$. Note that (3.23) implies (3.1). The theorem thus follows from Lemma 3.2.

$$E\hat{f}_n(x) = n^{-1} \sum_{i=1}^{n} b_i^{-d} \int K((x-y)/b_i) f(y) dy \,. \tag{3.24}$$

(ii) By the proof of Theorem 2.1(ii), it follows that $\sup_x |E\hat{f}_n(x) - f(x)| \le Cb_n = O(\Psi(n))$ since $n(\log n)^{-1} b_n^{d+2} = O(1)$. Part (ii) follows from (i) and the fact that

$$\sup_{x \in D} |\hat{f}_n(x) - f(x)| \le \sup_{x \in D} |\hat{f}_n(x) - E\hat{f}_n(x)| + \sup_{x \in D} |E\hat{f}_n(x) - f(x)| \,.$$

Acknowledgements

The author would like to thank the referees for many valuable comments.

REFERENCES

Devroye, L. and Gyorfi, L. (1985). *Nonparametric Density Estimation: The L_1 View*, Wiley, New York.

Fukunaga, K. (1972). *Introduction to Statistical Pattern Recognition*, Academic Press, New York.

Fukunaga, K. and Hostetler, L. D. (1975). The estimation of the gradient of a density function, with applications in pattern recognition, *IEEE Trans. Inform. Theory*, **21**, 32–40.

Gyorfi, L. (1981). Strong consistent density estimate from ergodic sample, *J. Multivariate Anal.*, **11**, 81–84.

Gyorfi, L. and Masry, E. (1988). The L_1 and L_2 strong consistency of recursive kernel density estimation from dependent samples, Report, TU Budapest.

Gyorfi, L., Hardle, W., Sarda, P. and Vieu, P. (1988). Nonparametric curve estimation from time series, Preprint of book, 1988.

Hart, J. D. and Vieu, P. (1988). Data driven bandwidth choice for density estimation based on dependent data, Tech. Report, Department of Statistics, Texas A&M University, College Station, Texas.

Ibragimov, I. A. and Rozanov, Yu. V. (1978). *Gaussian Random Processes*, Springer, New York.

Ioannides, D. A. and Roussas, G. G. (1987). Note on the uniform convergence of density estimates for mixing random variables, *Statist. Probab. Lett.*, **5**, 279–285.

Masry, E. (1986). Recursive probability density estimation for weakly dependent stationary processes, *IEEE Trans. Inform. Theory*, **32**, 254–267.

Masry, E. (1987). Almost sure convergence of recursive density estimators for stationary mixing processes, *Statist. Probab. Lett.*, **5**, 249–254.

Masry, E. and Gyorfi, L. (1987). Strong consistency and rates for recursive probability density estimators of stationary processes, *J. Multivariate Anal.*, **22**, 79–93.

Pham, D. T. (1986). The mixing property of bilinear and generalized random coefficient autoregressive models, *Stochastic Process. Appl.*, **23**, 291–300.

Pham, D. T. and Tran L. T. (1985). Some mixing properties of time series models, *Stochastic Process. Appl.*, **19**, 297–303.

Robinson, P. M. (1983). Nonparametric estimators for time series, *J. Time Ser. Anal.*, **4**, 185–207.

Rosenblatt, M. (1970). Density estimates and Markov sequences, *Nonparametric Techniques in Statistical Inference*, (ed. M. Puri), 199–210, Cambridge Univ. Press, London.

Roussas, G. G. (1969). Nonparametric estimation of the transition distribution function of a Markov Process, *Ann. Math. Statist.*, **40**, 1386–1400.

Roussas, G. G. (1988). Nonparametric estimation in mixing sequences of random variables, *J. Statist. Plann. Inference*, **18**, 135–149.

Silverman, B. W. (1978). Choosing a window width when estimating a density, *Biometrika*, **65**, 1–11.

Stone, C. J. (1983). Optimal uniform rate of convergence for nonparametric estimators of a density function and its derivatives, *Recent Advances in Statistics*: *Papers in Honor of H. Chernoff*, (eds. M. H. Rezvi, J. S. Rustagi and D. Siegmund), 393–306, Academic Press, New York.

Takahata, H. (1977*a*). Limiting behavior of density estimates for stationary asymptotically uncorrelated processes, *Bull. Tokyo Gakugei Univ.* (*4*), **29**, 1–9.

Takahata, H. (1977*b*). On recursive density estimators for a class of stationary processes, *Bull. Tokyo Gakugei Univ.* (*4*), **29**, 10–18.

Takahata, H. (1979). An application of the Strassen law of iterated logarithm to density estimations, *Bull. Tokyo Gakugei Univ.* (*4*), **31**, 45–50.

Takahata, H. (1980). Almost sure convergence of density estimators for weakly dependent stationary processes, *Bull. Tokyo Gakugei Univ.* (*4*), **32**, 11–32.

Tran, L. T. (1989*a*). Recursive density estimation under dependence, *IEEE Trans. Inform. Theory*, **35**, 1103–1108.

Tran, L. T. (1989*b*). The L_1 complete convergence of recursive density estimators under weak dependence (manuscript).

Wegman, E. J. and Davies, H. I. (1979). Remarks on some recursive estimators of a probability density function, *Ann. Statist.*, **7**, 316–327.

Wolverton, C. T. and Wagner, T. J. (1969*a*). Asymptotically optimal discriminant functions for pattern classification, *IEEE Trans. Inform. Theory*, **15**, 258–265.

Wolverton, C. T. and Wagner, T. J. (1969*b*). Recursive estimates of probability densities, *IEEE Trans. Syst. Sci. Cybern.*, **5**, 307.

Yakowitz, S. (1987). Nearest-neighbor methods for time series analysis, *J. Time Ser. Anal.*, **8**, 235–247.

Yamato, H. (1971). Sequential estimation of a continuous probability density function and mode, *Bull. Math. Statist.*, **14**, 1–12.

Yoshihara, K. (1976). Limiting behavior of *U*-statistics for stationary absolutely regular processes, *Z. Wahrsch. Verw. Gebiete*, **35**, 237–252.

Yoshihara, K. (1978). Probability inequalities for sums of absolutely regular processes and their applications, *Z. Wahrsch. Verw. Gebiete*, **43**, 319–330.

Yoshihara, K. (1984). Density estimation for samples satisfying a certain absolute regularity condition, *J. Statist. Plann. Inference*, **9**, 19–32.