ELSEVIER

Journal of Statistical Planning and Inference 61 (1997) 141–154

journal of statistical planning and inference

Kernel density estimation under weak dependence with sampled data

Berlin Wu*

Department of Mathematical Sciences, National Chengchi University, Taipei, Taiwan

Received 15 March 1994

Abstract

Kernel type estimators of the density of continuous time $\mathbb{R}^d$-valued stochastic processes are studied. Uniform strong consistency on $\mathbb{R}^d$ of the estimators and their rates of convergence are obtained. The stochastic processes are assumed to satisfy the strong mixing condition and the sampling instants are random. It is shown that the estimators can attain the optimal L' rates of convergence.

AMS classifications: Primary: 62G05; Secondary: 60J25

Keywords: Strong mixing; Density estimation; Kernel; Bandwidth

1. Introduction

This paper is concerned with the estimation of the marginal probability density function $f(x)$ of a stationary continuous-time $\mathbb{R}^d$-valued process $X = \{X(t), -\infty < t < \infty\}$ on the basis of the discrete time samples $\{X(t_k)\}$, $1 \leqslant k \leqslant n$, where the sampling instants $\{t_k\}$ are random. As an estimator of $f(x)$ we shall consider the kernel estimate defined by

$$f_n(x) = (nb_n^d)^{-1} \sum_{j=1}^{n} K((x - X(t_j))/b_n),$$

where K is a kernel function and $\{b_n\}$ is a sequence of bandwidths tending to zero as n tends to infinity.

Density estimation has been studied extensively since the works of Rosenblatt (1956) and Parzen (1962). Under dependent situations, kernel type density estimators

* Tel.: + 886 2 9387387; fax: + 886 2 93 90064; e-mail: berlin@math.necu.edu.tw.

PII S0378-3758(96)00151-6

have been investigated by Masry (1983, 1986), Robinson (1983), Roussas (1988) and Tran (1989a, b, 1993) for various weakly dependent processes. Györfi et al. (1988) studied the uniform convergence and the L_1 convergence of f_n under different mixing conditions. Masry (1983) investigated the random sampling case considered in the present paper.

Assume that X_t satisfies the strong mixing condition defined below.

Definition 1.1. Let $\mathscr{F}_{-\infty}^{0}$ and $\mathscr{F}_{\tau}^{\infty}$ denote, respectively, the σ-fields generated by $X(t)$, $t \leqslant 0$ and by X_t, $t \geqslant \tau$. Then X_t is strong mixing if

$$\alpha(\tau) = \sup\{|P(A \cap B) - P(A)P(B)|\colon A \in \mathscr{F}_{-\infty}^{0},\ B \in \mathscr{F}_{\tau}^{\infty}\} \downarrow 0.$$

The strong mixing condition is well known to be weaker than many dependence conditions, for example, the absolutely regular condition or the ϕ-mixing condition. For more information on strong mixing processes, see Rosenblatt (1956), or Roussas (1988).

The purpose of this paper is to establish weak conditions under which f_n converges to f a.s. and uniformly on $\mathbb{R}^d$, and also to obtain sharp rates of convergence of f_n to f. The rates of convergence obtained are sharp. From Theorem 3.1, it follows that f_n can achieve the uniform rate of convergence on $\mathbb{R}^d$ of order $(n^{-1}\log n)^{1/(d+2)}$, which is the optimal uniform rate of convergence on compacts for nonparametric estimators of a density function (see Stone, 1983). The letter C will be used to denote a generic constant. All limits are taken as $n \to \infty$ unless indicated otherwise.

2. Assumptions and preliminaries

Assumption 1. The density $f(x)$ is bounded.

Assumption 2. The kernel K is a probability density function defined on $\mathbb{R}^d$. In addition K satisfies a Lipschitz condition $|K(x) - K(y)| < C\|x - y\|$, where $\|\cdot\|$ is the usual norm on $\mathbb{R}^d$.

Assumption 3. The process X_t is strong mixing and the strong mixing coefficient satisfies $\int_0^\infty (1+\tau)[\alpha(\tau)]^\rho \, d\tau < \infty$ for some $0 < \rho < 1$.

We assume that the sampling instants $\{t_k\}$ are random, constituting a renewal process on $[0, \infty)$. Let $\{\tau_k\}$, $1 \leqslant k < \infty$, be a sequence of i.i.d. random variables with a common distribution $G(x)$ on $[0, \infty)$ with $G(0) = 0$ and a finite mean $\int_0^\infty x \, dG(x) = 1/\beta < \infty$. The sampling instants are defined by $t_k = \sum_{i=1}^{k} \tau_i$, $k = 1, 2, \ldots$.

Assumption 4. The renewal-type sampling instants $\{t_k\}$ have an intensity density $g(x)$ on $[0, \infty)$ and the second-order differential $m(x)$ satisfies $m(x) \leqslant C(1 + x)$ on $[0, \infty)$ where C is a constant.

Let $G_k(x)$ be the cumulative distribution function of t_k. If $G(x)$ is absolutely continuous with density $g(x)$ then G_k has a derivative, say, $g_k(x)$, which is the probability density function of t_k. Define then $m(t) = 2\sum_{k=1}^{\infty} kg_k(t)$, $t > 0$. The quantity m is often referred to in the renewal theory literature as the second-order factorial density.

Assumption 5. The joint probability density $f(x, y; \tau)$ of (X_0, X_τ) exists and satisfies

$$p(x, y; s) \equiv \int_0^\infty f(x, y; \tau + s)\, g(\tau)\, \mathrm{d}\tau \leqslant C < \infty,$$

for all x, y and $s \geqslant 0$.

Denote

$$\psi(n, 1) = (\log n)^{1/2}/(nb_n^d)^{1/2}.$$

Assumption 6. For some $\ell > 0$,

$$(\psi(n, 1))^{-1} \sup_{\|x\| \geqslant n^\ell} |f(x)| = \mathrm{O}(1), \tag{2.1}$$

$$\sum_{n=1}^{\infty} n\left(1 - \int_{\|x\| > n^\ell} f(x)\, \mathrm{d}x\right) < \infty. \tag{2.2}$$

Assumption 7. Assume

$$(\psi(n, 1))^{-1} b_n^{-d} \sup_{\|x\| \geqslant n^\ell} K(x/b_n) = \mathrm{O}(1). \tag{2.3}$$

Note that the constant ℓ in Assumptions 6 and 7 are identical.

Assumption 8. Let $0 < \tau < 1/2$. Denote

$$\gamma = -(d/2)(d + 3 + \tau), \qquad a(n, 1) \equiv n^{(1/2)(2/d + d + \tau + 1)} b_n^{\gamma} (\log n)^{(1 - d - \tau)/2}.$$

Define

$$\alpha(2\tau, p) \equiv \int (\alpha(s))^{2\tau}\, \mathrm{d}G_p(s),$$

and

$$\psi(n, 2) \equiv a(n, 1)\, \alpha(2\tau, p), \tag{2.4}$$

with

$$p = [(\psi(n, 1))^{-1}], \tag{2.5}$$

where $[x]$ denotes the integer part of x. Suppose b_n tends to zero in a manner that

$$\psi(n, 2)\, h(n) \to 0 \tag{2.6}$$

for some function $h(n) > 0$ with $\sum_{n=1}^{\infty} (1/h(n)) < \infty$ and for some $0 < \tau < \frac{1}{2}$.

Lemma 2.1. *Let X_t satisfy the strong mixing condition. Let X and Y be random variables measurable with respect to $\mathscr{F}_{-\infty}^{0}$ and $\mathscr{F}_{\tau}^{\infty}$, respectively. Assume that $\|X\|_{2+\delta} < \infty$ and $\|Y\|_{2+\delta} < \infty$, where $\|X\|_{2+\delta} = (E|X|^{2+\delta})^{1/(2+\delta)}$. Then*

$$|EXY - EXEY| \leqslant 10 \|X\|_{2+\delta} \|Y\|_{2+\delta} \{\alpha(\tau)\}^{\delta/(2+\delta)}.$$

For a proof, see Deo (1973, Lemma 1, p. 871).

Let

$$K_n(x) = (1/b_n^d)\, K(x/b_n). \tag{2.7}$$

Then

$$f_n(x) = (1/n) \sum_{j=1}^{n} K_n(x - X(t_j)). \tag{2.8}$$

Define

$$I_{n,0}(x) \equiv (1/n) \operatorname{var} K_n[x - X(0)],$$

$$I_{n,k}(x) \equiv (1/n)\left(1 - \frac{|k|}{n}\right) \operatorname{cov}\{K_n[x - X(t_{|k|})],\, K_n[x - X(0)]\}.$$

Then

$$\operatorname{cov}\{f_n(x), f_n(y)\} = I_{n,0}(x) + \sum_{\substack{k=-(n-1) \\ k \neq 0}}^{n-1} I_{n,k}(x). \tag{2.9}$$

Denote

$$J_n(x) = \sum_{\substack{k=-(n-1) \\ k \neq 0}}^{n-1} |I_{n,k}(x)|. \tag{2.10}$$

Lemma 2.2. *If Assumptions 1–5 are satisfied, then*

(i) $n b_n^d \sup_{x \in \mathbb{R}^d} I_{n,0}(x) \leqslant C$,

(ii) $n b_n^d \sup_{x \in \mathbb{R}^d} J_n(x) \leqslant C$.

We will need the following result from Bradley (1983):

Lemma 2.3. *Suppose X and Y are random variables taking their values on $\mathscr{S}$ and $\mathbb{R}$, respectively, where $\mathscr{S}$ is a Borel space; suppose U is a uniform $-[0,1]$ r.v. independent of (X, Y); and suppose ξ and v are positive numbers such that $\xi \leqslant \|Y\|_v < \infty$, where $\|Y\|_v$ is as defined in Lemma 2.1. Then there exists a real-valued r.v. $Y^* = f(X, Y, U)$, where f is a measurable function from $S \times \mathbb{R} \times [0,1]$ into $\mathbb{R}$, such that*

(i) *Y^* is independent of X,*
(ii) *the probability distributions of Y and Y^* are identical, and*
(iii) $P(|Y^* - Y) \geqslant \xi) \leqslant 18(\|Y\|_v/\xi)^{v/(2v+1)}[\sup|P(A \cap B) - P(A)P(B)|]^{2v/2v+1}$,

where the supremum is taken over all sets A, B with $a \in \mathscr{F}(X)$, $B \in \mathscr{F}(Y)$. Here, $\mathscr{F}(X)$, $\mathscr{F}(Y)$ are the σ-fields generated by X and Y, respectively.

Lemma 2.4. *Suppose Assumptions 1–8 hold and b_n tends to zero slowly enough that $nb_n^d(\log n)^{-1} \to \infty$. Then*

$$\sup_{\|x\| \leqslant 2n^\zeta} |f_n(x) - Ef_n(x)| = \mathrm{O}(\psi(n, 1)) \text{ a.s.} \quad \text{as } n \to \infty. \tag{2.11}$$

Lemma 2.5. *Suppose the conditions of Lemma 2.4 hold. Then*

$$\sup_{\|x\| > 2n^\zeta} |f_n(x) - f(x)| = \mathrm{O}(\psi(n, 1)) \text{ a.s.} \tag{2.12}$$

3. Uniform convergence of f_n

Theorem 3.1. *Suppose the conditions of Lemma 2.4 hold. In addition assume*

$$|f(x) - f(y)| \leqslant C\|x - y\|, \qquad \int \|x\| |K(x)| \,\mathrm{d}x < \infty$$

and

$$(\psi(n, 1))^{-1} b_n = \mathrm{O}(1). \tag{3.1}$$

Then $\sup_{x \in \mathbb{R}^d} |f_n(x) - f(x)| = \mathrm{O}(\psi(n, 1))$ a.s.

Proof. By the Lipschitz condition of f and since $\int K(x)\,\mathrm{d}x = 1$ and $\int \|x\| |K(x)| \,\mathrm{d}x < \infty$, following Roussas (1988, p. 141), we have

$$\sup_{x \in \mathbb{R}^d} |Ef_n(x) - f(x)| \leqslant Cb_n. \tag{3.2}$$

But $b_n = \mathrm{O}(\psi(n, 1))$ by the condition stated in (3.1). The rest of the proof follows from Lemmas 2.4, 2.5 and (3.2).

Example 3.1. We will consider the important case that $\{\tau_k\}$ constitutes an ordinary renewal process with τ_k having a probability density function $\theta \exp\{-\theta x\}$ where θ is a positive number. In this case t_k has the probability density function of a Gamma distribution, namely

$$\theta(\theta x)^{k-1}\exp(-\theta x)/\Gamma(k), \tag{3.3}$$

where

$$\Gamma(k) = \int_0^{\infty} x^{k-1}\,\mathrm{e}^{-x}\,\mathrm{d}x. \tag{3.4}$$

We will also assume $\alpha(s) = \mathrm{O}(s^{-\kappa})$ for some $\kappa > 2$. Then Assumption 3 holds since $\kappa > 2$. For sequences of positive integers $\{a_n\}$ and $\{b_n\}$, we write $a_n \sim b_n$ if $a_n/b_n \to 1$. The sampling intervals $\{\tau_k\}$ have mean value $1/\theta$. Let σ^2 be the variance of τ_1. Following Masry (1983, p. 704), we have

$$m(t) \sim 2\theta(t_\theta^2\sigma^2) = 2\theta(t+1),$$

which shows that Assumption 4 holds.

Using (3.3),

$$\hat{\alpha}(2\hat{\tau}, p) = (\theta^p/\Gamma(p))\int s^{-2\kappa\hat{\tau}+p-1}\exp(-\theta s)\,\mathrm{d}s, \tag{3.5}$$

where $\hat{\alpha}(2\hat{\tau}, p)$ is defined in Assumption 8. Let $t = \theta s$. Then

$$\hat{\alpha}(2\hat{\tau}, p) = \theta^{2\kappa\hat{\tau}}(1/\Gamma(p))\int t^{-2\kappa\hat{\tau}+p-1}\exp(-t)\,\mathrm{d}t = \theta^{2\kappa\hat{\tau}}\Gamma(p-2k\hat{\tau})/\Gamma(p). \tag{3.6}$$

By Stirling's formula

$$\Gamma(p) \sim (2\pi/p)^{1/2}(p/\mathrm{e})^p,$$

which implies that

$$\hat{\alpha}(2\hat{\tau}, p) \sim \theta^{2\kappa\hat{\tau}}p^{-2\kappa\tau}. \tag{3.7}$$

By (2.5) and (3.7),

$$\hat{\alpha}(2\hat{\tau}, p) \sim \theta^{2\kappa\hat{\tau}}\psi(n,1)^{2\kappa\hat{\tau}} \sim \theta^{2\kappa\hat{\tau}}(nb_n^d/\log n)^{-\kappa\tau} \equiv a(n,2). \tag{3.8}$$

Using (2.4), (3.8) and the value of γ given in Assumption 8,

$$\psi(n,2) \sim a(n,1)\,a(n,2). \tag{3.9}$$

Choose

$$h(n) = \hat{h}(n) \equiv n\log n(\log\log n)^{1+\varepsilon},$$

where ε is an arbitrary positive integer. Then $\sum_{n=1}^{\infty}(1/h(n)) < \infty$ and using (3.9), it is clear that (2.6) is satisfied if

$$a(n,1)\,a(n,2)\,h(n) \to 0. \tag{3.10}$$

After a simple computation using the definitions of $a(n, 1)$, $a(n, 2)$ and $h(n)$, we obtain that (3.10) holds if

$$n^{(1/2)(2/d+d+\hat{\tau}+3)-\kappa\hat{\tau}}\, b^{(d/2)(-d-3-\hat{\tau}-2\kappa\hat{\tau})}(\log n)^{((3-d-\hat{\tau})/2)+\kappa\hat{\tau}}(\log\log n)^{1+\varepsilon} \to 0. \quad (3.11)$$

For definiteness, let us assume further that $b_n = n^{-\beta}$ for some $\alpha > \beta > 0$. Then (3.11) is satisfied if

$$\beta(d/2)(-d-3-\hat{\tau}-2\kappa\hat{\tau}) > (1/2)(2/d+d+\hat{\tau}+3) - \kappa\hat{\tau}, \quad (3.12)$$

for some $0 < \hat{\tau} < \frac{1}{2}$. Both the left-hand side and the right-hand side of (3.12) are continuous functions of $\hat{\tau}$. Thus (3.12) holds for some $0 < \hat{\tau} < \frac{1}{2}$ if it holds for $\hat{\tau} = \frac{1}{2}$, that is,

$$\tfrac{1}{2}\beta d[-d-\tfrac{7}{2}-\kappa] > \tfrac{1}{2}[2/d+d+\tfrac{7}{2}] - \tfrac{1}{2}\kappa, \quad (3.13)$$

or equivalently,

$$\beta < \frac{2\kappa - 4/d - 2d - 7}{d(2\kappa + 2d + 7)}. \quad (3.14)$$

Suppose that $|f(x)| = \mathrm{O}(\|x\|^{-r})$ as $\|x\| \to \infty$ for some $r > 0$. Then

$$\left(1 - \int_{\|x\| < n^{\ell}} f(x)\,\mathrm{d}x\right) = \int_{\|x\| > n^{\ell}} f(x)\,\mathrm{d}x \leqslant C\int_{\|x\| > n^{\ell}} \|x\|^{-r}\,\mathrm{d}x \leqslant Cn^{\ell(-r+d)}. \quad (3.15)$$

Thus (2.2) is finite if

$$\sum_{n=1}^{\infty} n^{1+\ell(-r+d)} < \infty, \quad (3.16)$$

which is in turn satisfied if $\ell > 2/(r-d)$ for some r greater than d.

Now

$$\sup_{\|x\| \geqslant n^{\ell}} |f(x)| \leqslant Cn^{-r\ell}. \quad (3.17)$$

Note that $\psi(n, 1) = (\log n)^{1/2} n^{(\beta d-1)/2}$. Using (3.17), it is seen that (2.1) holds since

$$(\log n)^{-1/2} n^{(1-\beta d)/2} n^{-r\ell} = \mathrm{O}(1). \quad (3.18)$$

Assume $K(x) = \mathrm{O}(\|x\|^{-s})$ for some $s > \frac{1}{2}d$ as $\|x\| \to \infty$. Then (2.3) is valid if

$$(\log n)^{-1} n^{1-\beta d+2\beta d-2\ell s-2\beta s} = \mathrm{O}(1),$$

which follows if $\beta > (1-2\ell s)/(2s-d)$. Finally, condition (3.1) is satisfied if $nb_n^{d+2}\log n^{-1} = \mathrm{O}(1)$, which holds for $b_n = n^{-\beta}$ as long as $nn^{-\beta(d+2)}\log n^{-1} = \mathrm{O}(1)$, or $\beta > 1/(d+2)$.

Summarizing, we have the following:

Proposition 3.1. *Suppose that $\{\tau_k\}$ constitutes an ordinary renewal process with τ_k having a probability density function $g(x) = \theta\exp\{-\theta x\}$ where θ is a positive number.*

Suppose that Assumptions 1, 2, 5 *hold and* $\alpha(s) = O(s^{-\kappa})$ *for some* $\kappa > 0$. *In addition, assume that*

$$|f(x)| = O(\|x\|^{-r}) \quad \text{for some } r > d,$$

and

$$K(x) = O(\|x\|^{-s}) \quad \text{for some } s > \tfrac{1}{2}d$$

as $\|x\| \to \infty$. *Let* $b_n = n^{-\beta}$ *where* $\beta > 0$ *satisfies*

$$(2\kappa - 4\ell d - 2d - 7)/(d(2\kappa + 2d + 7)) > \beta > \max\{(1 - 2\ell s)/(2s - d),\ 1/(d+2)\}$$

for some $\ell > 2/(r - d)$. *Then* $\sup_{x \in \mathbb{R}^d} |f_n(x) - f(x)| = O(\psi(n, 1))$ *a.s.*

Acknowledgements

I would like to thank the referee and the Associate Editor for many useful comments.

Appendix A

Proof of Lemma 2.2. (i) Clearly

$$nb_n^d I_{n,0}(x) \leqslant \int (1/b_n^d) K^*((x-u)/b_n) f(u)\,du, \tag{A.1}$$

with $K^* = K^2$. Set $v = (x-u)/b_n$. Since f is bounded by Assumption 1 and K^* is integrable by Assumption 2,

$$\int (1/b_n^d) K^*((x-u)/b_n) f(u)\,du \leqslant C \int K^*(v)\,dv \leqslant C.$$

(ii) Write

$$J_n(x) = 2\sum_{k=1}^{c(n)} |I_{n,k}(x)| + 2\sum_{k=1}^{c(n)} |I_{n,k}(x)| = J_{1n}(x) + J_{2n}(x), \tag{A.2}$$

where $c(n)$ is a positive integer satisfying $c(n) \to \infty$ and $c(n) b_n^d = O(1)$. Then

$$J_{1n}(x) \leqslant (2/n) \sum_{k=1}^{c(n)} \iint_{\mathbb{R}^d \times \mathbb{R}^d} |K_n(x-u) K_n(x-v)| |q_k(u,v) - f(u)f(v)|\,du\,dv, \tag{A.3}$$

where

$$q_k(u,v) = \int f(u,v;t)\,dG_k(t). \tag{A.4}$$

Following Masry (1983) and using Assumption 5, we have $|q_k(u, v)| < C$ for all u, v and $k \geqslant 1$. Therefore,

$$\sup_{x \in \mathbb{R}^d} J^{1n}(x) \leqslant C/(nb_n^d). \tag{A.5}$$

Clearly,

$$J_{2n}(x) = 2 \sum_{k=c(n)+1}^{n-1} n^{-1}\left(1 - \frac{k}{n}\right)|\mathrm{cov}\{K_n[x - X(t_k)],\ K_n[x - X(0)]\}|. \tag{A.6}$$

By Assumptions 3 and 4

$$\hat{\alpha}(\rho, k) < \infty. \tag{A.7}$$

Employing Lemma 2.1 with $\delta - 2\rho/(1-\rho)$,

$$|\mathrm{cov}\{K_n[x - X(t_k)],\ K_n[x - X(0)]\}| \leqslant C(1/b_n^d)^{1+\rho}\hat{\alpha}(\rho, k). \tag{A.8}$$

Note that the constant C in (A.8) is independent of x. Using Assumptions 3 and 4 and following the argument of Theorem 7 of Masry (1983), we obtain

$$\sup_{x \in \mathbb{R}^d} J_{2n}(x) \leqslant C/(nb_n^d). \tag{A.9}$$

The proof of (ii) follows by combining (A.2), (A.5) and (A.9).

Proof of Lemma 2.4. The proof of this lemma uses similar techniques employed in Tran (1989a, b). Let $\hat{\ell} = b_n^{(d+1)}\psi(n, 1)$. Then the sphere $\{x: \|x\| \leqslant 2n^{\ell}\}$ can be covered with, say v, cubes B_k's having sides of length $\hat{\ell}$ and center at x_k, where $v \leqslant Cn^{\ell d}[b_n^{(d+1)}\psi(n, 1)]^{-d}$. Now

$$\sup_{\|x\| \leqslant n^{\ell}} |f_n(x) - Ef_n(x)| \leqslant \max_{1 \leqslant k \leqslant v} \sup_{x \in B_k} |f_n(x) - f_n(x_k)| + \leqslant \max_{1 \leqslant k \leqslant v} |f_n(x_k) - Ef_n(x_k)|$$
$$+ \max_{1 \leqslant k \leqslant v} \sup_{x \in B_k} |Ef_n(x_k) - Ef_n(x)|.$$

For $x \in B_k$, by the Lipschitz condition of K in Assumption 2,

$$|f_n(x) - f_n(x_k)| \leqslant Cb_n^{-(d+1)}\|x - x_k\| \leqslant Cb_n^{-(d+1)}\hat{\ell} = \mathrm{O}(\psi(n, 1)) \text{ a.s.}$$

Therefore

$$\max_{1 \leqslant k \leqslant v} \sup_{x \in B_k} |f_n(x) - f_n(x_k)| = \mathrm{O}(\psi(n, 1)) \text{ a.s.},$$

$$\max_{1 \leqslant k \leqslant v} \sup_{x \in B_k} |Ef_n(x_k) - Ef_n(x)| = \mathrm{O}(\psi(n, 1)) \text{ a.s.}$$

It remains to show that

$$\max_{1 \leqslant k \leqslant v} |f_n(x_k) - Ef_n(x_k)| = \mathrm{O}(\psi(n, 1)). \tag{A.10}$$

Assume $n = 2pq$ for some increasing integer valued function $q = q(n)$, where p is defined in (2.5). Note that $p \to \infty$ since $nb_n^d(\log n)^{-1} \to \infty$ by assumption. Define

$$\Delta_j(x) \equiv K_n(x - X(t_j)) - EK_n(x - X(t_j)).$$

Then the random variables Δ_j's can be grouped successively into $2q$ blocks of size p. Write $S(n, x)$ as $S(n, x) = S(n, x, 1) + S(n, x, 2)$, where

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

with

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

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

$$\begin{aligned} P\left(\max_{1 \leqslant k \leqslant v} |f_n(x_k) - Ef_n(x_k)| > \varepsilon_n\right) &= P\left(\max_{1 \leqslant k \leqslant v} |S(n, x_k, 1) + S(n, x_k, 2)| > \varepsilon_n\right) \\ &\leqslant P\left(\max_{1 \leqslant k \leqslant v} |S(n, x_k, 1)| > \varepsilon_n/2\right) \\ &\quad + P\left(\max_{1 \leqslant k \leqslant v} |S(n, x_k, 2)| > \varepsilon_n/2\right). \end{aligned} \tag{A.11}$$

Since K is Lipschitz and absolutely integrable, $|K|$ is bounded by some constant $\tilde{K}$. Thus

$$|V(n, x, j)| \leqslant 2\tilde{K}p/(nb_n^d).$$

Let

$$\lambda_n = (4\tilde{K})^{-1}(nb_n^d \log n)^{1/2}. \tag{A.12}$$

Then $\lambda_n|V(n, x, j)| \leqslant 2\tilde{K}p\lambda_n/(nb_n^d) \leqslant \frac{1}{2}$. Refer to $V(n, x, 2(j-1))$ simply as $V_j(x)$. Define $W_0(x) = 0$, $W_1(x) = V_1(x)$. By Lemma 2.3, for each $j > 2$, there exists a r.v. $W_j(x)$ such that $W_j(x)$ is independent of $V_1(x), \ldots, V_{j-1}(x)$ has the same distribution as $V_j(x)$ and satisfies

$$\begin{aligned} &P[|V_j(x) - W_j(x)| > \xi| \leqslant 18(\|V_j(x)\|_\gamma/\xi)^{\nu/(2\nu+1)} \\ &\quad [\sup|P(A \cap B) - P(A)P(B)]^{2\nu/(2\nu+1)}, \end{aligned} \tag{A.13}$$

where the supremum is taken over all sets A, B with A and B in the σ-fields generated, respectively, by $V_1(x), \ldots, V_{j-1}(x)$ and $V_j(x)$, respectively. Here ξ and ν are any positive numbers such that $0 < \xi \leqslant \| V_j(x) \|_v < \infty$. Now,

$$P[|S(n,x,1)| > \varepsilon_n/2] \leqslant P\left[\left|\sum_{j=1}^{q} W_j(x)\right| > \varepsilon_n/4\right] + P\left[\left|\sum_{j=1}^{q} V_j(x) - W_j(x))\right| > \varepsilon_n/4\right]. \tag{A.14}$$

Now $\lambda_n |W_j(x)| \leqslant 1/2$. Following the same line of argument in Tran (1989a),

$$P\left[\left|\sum_{j=1}^{q} W_j(x)\right| > \varepsilon_n\right] \leqslant 2\exp\left(-\lambda_n\varepsilon_n + \lambda_n^2 \sum_{j=1}^{q} E(W_j(x))^2\right). \tag{A.15}$$

Clearly,

$$\sum_{j=1}^{q} E(W_j(x))^2 \leqslant n^{-2}\left(\sum_{k=1}^{n} E\Delta_k^2 + 2\sum_{1\leqslant k<l\leqslant n} |E\Delta_k(x)\Delta_l(x)|\right)$$
$$= I_{n,0}(x) + J_n(x), \tag{A.16}$$

where $I_{n,0}(x)$ and $J_n(x)$ are defined in (2.9) and (2.10). By Lemma 2.2 and (A.16),

$$\sup_{x\in\mathbb{R}^d} \sum_{j=1}^{q} E(W_j(x))^2 \leqslant C/nb_n^d). \tag{A.17}$$

Employing (A.15) and (A.17)

$$\sup_{x\in\mathbb{R}^d} P\left[\left|\sum_{j=1}^{q} W_j(x)\right| > \varepsilon_n\right] \leqslant C\exp[-\lambda_n\varepsilon_n + C\lambda_n^2/(nb_n^d)]. \tag{A.18}$$

Choose $v = \hat{\tau}/(1-2\hat{\tau})$. Then $\hat{\tau} = v/(2v+1)$. Clearly,

$$\sup_{x\in\mathbb{R}^d} \max_{1\leqslant j\leqslant q} \| V_j(x) \|_v^{\hat{\tau}} \leqslant C(p/(nb_n^d))^{\hat{\tau}}.$$

Consider the last term of (A.14). If $\varepsilon_n/4q \leqslant \| V_j(x) \|_v$, we have by using (A.13),

$$\sup_{x\in\mathbb{R}^d} P\left[\left|\sum_{j=1}^{q} V_j(x) - W_j(x)\right| > \varepsilon_n/4\right] \leqslant Cq(q\varepsilon_n^{-1})^{\hat{\tau}}(p/(nb_n^d))^{\hat{\tau}}\hat{\alpha}(2\hat{\tau},p) \equiv \psi(n,3), \text{ say.} \tag{A.19}$$

If $\varepsilon_n/4q > \| V_j(x) \|_v$, then

$$\sup_{x\in\mathbb{R}^d} P\left[\left|\sum_{j=1}^{q} V_j(x) - W_j(x)\right| > \varepsilon_n/4\right] \leqslant 18q\hat{\alpha}(2\hat{\tau},p) \leqslant \psi(n,3), \tag{A.20}$$

since $q(q\varepsilon_n^{-1}))^{\hat{\tau}}(p/(nb_n^d))^{\hat{\tau}} \to \infty$ by a simple computation. Replacing ε_n by $\varepsilon_n/4$ in (A.18), then using (A.14), (A.18), (A.19) and (A.20),

$$P\left(\max_{1\leqslant k\leqslant r} |S(n,x_k,1)| > \varepsilon_n/2\right) \leqslant Cr\exp[-\lambda_n\varepsilon_n/2 + C\lambda_n^2/(nb_n^d)] + Cr\psi(n,3). \tag{A.21}$$

Similarly, $P(\max_{1 \leqslant j \leqslant v} |S(n, x_k, 2)| > \frac{1}{2}\varepsilon_n)$ is bounded by the right-hand side of (A.21). A simple computation shows that $\lambda_n \varepsilon_n = \eta \log n$ and $\lambda_n^2/(nb_n^d) = (4\tilde{K})^{-2} \log n$. From (A.11) and (A.21)

$$P\left(\max_{1 \leqslant k \leqslant v} |f_n(x_k) - Ef_n(x_k)| > \varepsilon_n \right) \leqslant Cvn^{-((\eta/2)+C)} + Cv\psi(n, 3). \tag{A.22}$$

Note that

$$v \leqslant Cn^{d(\ell + 1/2)}(b_n^d)^{-(d/2)-1}(\log n)^{-d/2}. \tag{A.23}$$

Since $\psi(n, 1) \to \infty$ by assumption,

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

Using (A.23),

$$v \leqslant Cn^{d(\ell + 1)+1}(\log n)^{-d-1}.$$

Hence, for sufficiently large η,

$$\sum_{n=1}^{\infty} vn^{((\eta/2)+C)} < \infty. \tag{A.24}$$

Let

$$\xi(n, 1) \equiv q(q\varepsilon_n^{-1})^{\hat{\tau}}(p/(nb_n^d))^{\hat{\tau}}, \qquad \xi(n, 2) \equiv n^{\ell d}(b_n^{d+1}\psi(n, 1))^{-d}.$$

Since $2pq \leqslant n$,

$$\xi(n, 1) \leqslant Cn(\psi(n, 1))^{1-\hat{\tau}} b_n^{-d\hat{\tau}}.$$

Therefore

$$\xi(n, 1)\xi(n, 2) \leqslant Cn^{\ell d+1} b^{-d(d+1+\hat{\tau})} \{(nb_n^d)^{1/2}(\log n)^{1/12}\}^{d-1+\hat{\tau}}. \tag{A.25}$$

Using (2.19), (A.25), (2.6) and note that $v \leqslant C\xi(n, 2)$,

$$v\psi(n, 3) \leqslant C\xi(n, 1)\xi(n, 2)\hat{\alpha}(2\hat{\tau}, p) \leqslant C\psi(n, 2) = o(1/h(n)).$$

Since $\sum_{n=1}^{\infty} (1/h(n)) < \infty$,

$$\sum_{n=1}^{\infty} Cv\psi(n, 3) < \infty. \tag{A.26}$$

The proof of the lemma follows by (A.22), (A.24), (A.26) and the Borel–Cantelli lemma.

Proof of Lemma 2.5. Obviously,

$$\sup_{\|x\| > n^{\ell}} |f_n(x) - f(x)| \leqslant \sup_{\|x\| \geqslant 2n^{\ell}} f_n(x) + \sup_{\|x\| \geqslant 2n^{\ell}} f(x). \tag{A.27}$$

From (2.1) of Assumption 6

$$(\psi(n, 1))^{-1} \sup_{\|x\| > 2n^{\ell}} f(x) \leqslant (\psi(n, 1))^{-1} \sup_{\|x\| \geqslant n^{\ell}} |f(x)| = O(1). \tag{A.28}$$

By (A.27) and (A.28), it is clear that the proof of the lemma is completed if we can show that

$$\sup_{\|x\| > 2n'} |f_n(x)| = O(\psi(n, 1)), \tag{A.29}$$

which is equivalent to

$$(\psi(n, 1))^{-1} \sup_{\|x\| > n'} |f_n(x)| = (\psi(n, 1))^{-1}(nb_n^d)^{-1} \sup_{\|x\| > n'} \sum_{j=1}^{n} K((x - X(t_j))/b_n) = O(1). \tag{A.30}$$

Define

$$G_n = \{\omega: \|X(t_i, \omega)\| \leqslant n' \text{ for all } 1 \leqslant i \leqslant n\}. \tag{A.31}$$

Clearly,

$$P[G_n^c] = P[\|X(t_i)\| > n' \text{ for some } 1 \leqslant i \leqslant n] \leqslant nP[\|X(t_1)\| > n']. \tag{A.32}$$

By (2.2) of Assumption (A.4)

$$\sum_{n=1}^{\infty} P[G_n^c] \leqslant \sum_{n=1}^{\infty} n\left[1 - \inf_{\|x\| < n'} f(x)\,dx\right] < \infty. \tag{A.33}$$

Thus

$$P[G_n^c \text{ infinitely often}] = 0 \tag{A.34}$$

by the Borel–Cantelli lemma.

For $\omega \in G_n$, and $\|x\| > 2n'$, we have $\|X(t_j, \omega) - x\| \geqslant n'$. Thus,

$$(nb_n^d)^{-1} \sup_{\|x\| > 2n'} \sum_{j=1}^{n} K((x - X(t_j, \omega))/b_n) \leqslant b_n^{-d} \sup_{\|x\| > n'} K(x/b_n). \tag{A.35}$$

The proof of (A.30) follows by Assumption 5, (A.34) and (A.35).

References

Bradley, R.C. (1983). Approximation theorems for strongly mixing random variables. *Michigan Math. J.* **30**, 69–81.

Deo, C.M. (1973). A note on empirical processes of strong mixing sequences. *Ann. Probab.* **1**, 870–875.

Gyorfi, L., W. Hardle, P. Sarda and P. Vieu (1989) *Nonparametric Curve Estimation From time Series*. Springer-Verlag, New York.

Masry, E. (1983). Probability density estimation from sampled data. *IEEE Trans. Inform. Theory*, **IT-29**, 699–709.

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

Parzen, E. (1962). On estimation of a probability density function and mode. *Ann Math. Statist.* **33**, 1065–1076.

Robinson, P.M. (1983). Nonparametric estimators for time series. *J. Time Series Anal.* **4**, 185–197.

Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function. *Ann. Math. Statist.* **27**, 832–837.

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

Stein, E.M. (1970). *Singular Integrals and Differentiability Properties of Functions.* Princeton University Press, Princeton, New Jersey.

Stone, C.J. (1983). Optimal uniform rate of convergence for nonparametric estimators of a density function and its derivatives. In: m. H. REzvi, J.S. Rustagi and D. Siegmund, Eds., Recent Advances in Statistics: Papers in honor of Herman Chernoff, 293–306.

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

Tran, L.T. (1989b). The L_1 convergence of kernel density estimators under dependence. *Can. J. Statist.* **17**, 197–208.

Tran, L.T. (1983). Nonparametric function estimation for time series by local average estimators. *Ann. Statist.* **40**, 1040–1057.