Z. Wahrscheinlichkeitstheorie verw. Gebiete
67, 175–196 (1984)

Zeitschrift für
Wahrscheinlichkeitstheorie
und verwandte Gebiete

Asymptotic Properties of Integrated Square Error and Cross-Validation for Kernel Estimation of a Regression Function

Peter Hall

The Australian National University, P.O. Box 4, Canberra ACT 2600, Australia

Summary. We derive an asymptotic expansion of integrated square error in kernel-type nonparametric regression. A similar result is obtained for a cross-validatory estimate of integrated square error. Together these expansions show that cross-validation is asymptotically optimal in a certain sense.

1. Introduction

Let (X, Y), $(X_1, Y_1), \ldots, (X_n, Y_n)$ be independent observations from a bivariate distribution, and let $\mu(x) = E(Y \mid X = x)$ denote the regression function. Nadaraya [18] and Watson [26] introduced kernel estimators of $\mu(x)$, which are defined in the following way. Let K be a density function on the real line, and h be a small positive constant. Set

$$\hat{\mu}_n(x) = \hat{\mu}_n(x \mid h) \equiv \left[\sum_{j=1}^{n} Y_j K\{(x - X_j)/h\}\right] \Bigg/ \left[\sum_{j=1}^{n} K\{(x - X_j)/h\}\right].$$

If $h = h(n)$ is chosen so that $h \to 0$ and $nh \to \infty$ as $n \to \infty$, then $\mu_n(x) \to \mu(x)$ in probability. Detailed accounts of the consistency of such estimators have been given by Collomb [3], Devroye [8, 9], Devroye and Wagner [10, 11], Mack and Silverman [17] and Spiegelman and Sacks [21]. Some of these results describe the *rate* of convergence for different choices of the 'window size', h, and show that the order of consistency depends crucially on the selection of h. Cross-validation has been suggested by Wahba and Wold [25] as a practical method of determining h in real statistical problems of this type. See also Diaz [12]. Our aim in the present paper is to describe large sample properties of integrated square error (ISE) in non-parametric regression. We derive asymptotic expansions of ISE and of a cross-validatory estimate of ISE. These lead to a proof that the cross-validatory estimator is asymptotically optimal in the sense of minimising a version of ISE, in the case where K is the rectangular kernel.

Collomb [4, 5] has derived several results which shed light on the problem of selecting window size. In particular, he has shown that under appropriate conditions on h, K and the underlying distribution of (X, Y),

$$E\{\hat{\mu}_n(x)-\mu(x)\}^2=(nh)^{-1}[\sigma^2(x)\{f(x)\}^{-1}\int K^2(u)\,du] \\ +h^4[\{f'(x)\,\mu'(x)+\tfrac{1}{2}f(x)\,\mu''(x)\}^2\{f(x)\}^{-2}\{\int u^2 K(u)\,du\}^2] \\ +o\{(nh)^{-1}+h^4\}$$

as $h\to 0$ and $n\to\infty$. (We have assumed here that X has marginal density f, and written $\sigma^2(x)=\operatorname{var}(Y\,|\,X=x)$.) Under additional constraints it is permissible to formally integrate this expression [5, p. 82]. Thus, if A is a bounded interval on which f is bounded away from zero, and if w is a bounded weight function, then

$$\begin{aligned} I_n(h)&\equiv\int_A E\{\hat{\mu}_n(x)-\mu(x)\}^2\,w(x)\,dx \\ &=(nh)^{-1}[\int_A \sigma^2(x)\{f(x)\}^{-1}\,w(x)\,dx\cdot\int K^2(u)\,du] \\ &\quad+h^4[\int_A\{f'(x)\,\mu'(x)+\tfrac{1}{2}f(x)\,\mu''(x)\}^2\{f(x)\}^{-2}\,w(x)\,dx\cdot\{\int u^2 K(u)\,du\}^2] \\ &\quad+o\{(nh)^{-1}+h^4\}=(nh)^{-1}c_1+h^4 c_2+o\{(nh)^{-1}+h^4\}, \end{aligned} \tag{1.1}$$

say. The sum of the first two terms on the right hand side of (1.1) is minimised by taking

$$h=h_0\equiv(c_1/4c_2)^{1/5}\,n^{-1/5}, \tag{1.2}$$

which is the "asymptotically optimal" window size in the sense of minimising $I(h)$.

A natural choice for w is $w\equiv f$. In this case we might conjecture that $I_n(h)$, which can be written as

$$I_n(h)=\int_A E\{\hat{\mu}_n(x\,|\,h)-\mu(x)\}^2\,dF(x)$$

where F is the marginal distribution function of X, is closely approximated by mean summed square error, given by

$$\begin{aligned} \beta_n(h)&=\int_A\{\hat{\mu}_n(x\,|\,h)-\mu(x)\}^2\,dF_n(x) \\ &=n^{-1}\sum_{X_i\in A}\{\hat{\mu}_n(X_i\,|\,h)-\mu(X_i)\}^2 \end{aligned} \tag{1.3}$$

where F_n is the empiric distribution function of the X-sample. In Sect. 2 we derive an asymptotic expansion for $\beta_n(h)$ which shows that $I_n(h)$ and $\beta_n(h)$ *are* asymptotically equivalent. It will follow that $\beta_n(h_0)/I_n(h_0)\to 1$ in probability. Thus, an adaptive, "data-driven" estimate of window size, $\hat{h}$, will be asymptotically as good as the "best" window, h_0, if

$$\beta_n(\hat{h})/I(h_0)\to 1 \tag{1.4}$$

in probability. We shall prove that the cross-validatory window size attains this optimal performance, and also that $\hat{h}/h_0 \to 1$ in probability.

Craven and Wahba [6, Theorem 4.3] and Golub, Heath and Wahba [13] have described certain asymptotic properties of cross-validation, in the context of fitting smooth splines to noisy data. However, their assumptions differ from ours in a major respect: they assume that the values of the regressand are regularly spaced on a bounded interval, being determined by a precise integral formula. This obviously excludes the case where the design variables, X_i, are random. The classical theory of nonparametric regression (as distinct from curve-fitting) deals with the case of a random regressand; see for example Stone [22]. Our results are different in nature from those of [6, 13], which were effectively concerned with showing that the *expectation* of the left hand side in (1.4) converges to the *expectation* of the right hand side.

More recent work by Härdle and Marron [14] uses an ingenuous argument which allows mean integrated square error to be calculated "conditional" on a set on which the denominator of $\hat{\mu}_n$ behaves well. These results are in the spirit of Stone [23], in that the order of magnitude of the bias contribution is left undetermined. This allows the authors to impose only mild smoothness conditions on f.

2. Results

Our early attempts at solving this problem were founded on a two-dimensional version of the Komlós-Major-Tusnády [15] approximation to the empiric distribution function, due to Tusnády [24]. However, this required very restrictive conditions on the regressor variable, Y, such as that it have high-order moments. It seems to us that if cross-validation is to be recommended for practical purposes, such restrictions should be avoided at almost all costs, since doubt has been cast upon cross-validatory methods for density estimation when the distribution is unbounded [2, 20]. Therefore we shall present an alternative solution, based in part upon techniques we learned from Révész [19] and Csörgő and Révész [7, Sect. 6.3]. This requires us to restrict attention to the rectangular kernel,

$$K(u)=\begin{cases}1 & \text{for } |u| \leqq \frac{1}{2} \\ 0 & \text{otherwise,}\end{cases} \tag{2.1}$$

but permits comparatively generous conditions on the underlying distribution of (X, Y). (Our argument via Tusnády's [24] approximation did not include the rectangular kernel.)

Let $A=[a, b]$ denote a compact interval, and set $A^\delta=(a-\delta, b+\delta)$ for $\delta>0$. We assume that X has a twice differentiable density, f, on A^ε for some $\varepsilon>0$, and that f is bounded away from zero on A and satisfies a Lipschitz condition of order $\frac{1}{2}$:

$$\sup_{x \in A^{\frac{1}{2}\varepsilon};\, 0<y<\delta} |f''(x+y)-f''(x)|=O(\delta^{\frac{1}{2}}) \quad \text{as } \delta \to 0. \tag{2.2}$$

Suppose $\mu(x)\equiv E(Y|X=x)$, $\sigma^2(x)\equiv \mathrm{var}(Y|X=x)$ and $\mu_4(x)\equiv E[\{Y-\mu(x)\}^4|X=x]$ are all well-defined and bounded on A^ε, and μ has two continuous derivatives on A^ε. Set

$$\gamma(x)=f'(x)\,\mu'(x)+\tfrac{1}{2}f(x)\,\mu''(x).$$

The cross-validatory criterion for estimation on A is defined by

$$\alpha_n(h)=n^{-1}\sum_{X_i\in A}\{Y_i-\hat{\mu}_{ni}(X_i|h)\}^2,$$

where $\hat{\mu}_{ni}(x|h)=\hat{\mu}_{ni}(x)\equiv[\sum_{j\neq i} Y_j K\{(x-X_j)/h\}]/[\sum_{j\neq i} K\{(x-X_j)/h\}]$ is a kernel estimator of μ based on the sample excluding X_i. We choose $\hat{h}$ to minimise $\alpha_n(\hat{h})$. In view of the results of Collomb [4, 5] and the expansion (1.1), we know in advance that $\hat{h}$ should be selected in the vicinity of $n^{-1/5}$. Therefore we may suppose that $\alpha_n(h)$ is minimised over $h\in[\eta n^{-1/5}, \lambda n^{-1/5}]$, for η arbitrarily small and λ arbitrarily large. In the proof below we shall assume that η and λ are fixed as $n\to\infty$, although it is easily seen that we may choose $\eta\to 0$ and $\lambda\to\infty$ at a sufficiently slow rate.

Our first result shows that minimising $\alpha_n(h)$ is asymptotically equivalent to minimising mean summed square error, $\beta_n(h)$, defined at (1.3). Of course, the value of μ is unknown, and so β_n cannot be minimised directly.

Theorem 1. *Under the conditions stated above,*

$$\alpha_n(h)=\beta_n(h)+n^{-1}\sum_{X_i\in A}\{Y_i-\mu(X_i)\}^2+o_p(n^{-4/5}) \tag{2.3}$$

uniformly in $\eta n^{-1/5}\leq h\leq \lambda n^{-1/5}$, *for any* $0<\eta<\lambda<\infty$.

Our next theorem provides an expansion for $\beta_n(h)$, in which the first two terms are deterministic functions of h, and the remainder is negligibly small. Note that the kernel in (2.1) satisfies

$$\int K^2(u)\,du=1 \quad \text{and} \quad \int u^2 K(u)\,du=\frac{1}{12},$$

and so the first two terms of our expansion coincide with the first two terms on the right hand side in (1.1), with $w\equiv f$.

Theorem 2. *Under the conditions stated above,*

$$\beta_n(h)=(nh)^{-1}\int_A \sigma^2(x)\,dx+\frac{1}{144}h^4\int_A \gamma^2(x)\{f(x)\}^{-1}\,dx+o_p(n^{-4/5})$$

uniformly in $\eta n^{-1/5}\leq h\leq \lambda n^{-1/5}$, *for any* $0<\eta<\lambda<\infty$.

Let h_0 denote the "asymptotically optimal" window size defined at (1.2), and suppose $c_1 c_2\neq 0$ and η is so small and λ so large that $\eta<(c_1/4c_2)^{1/5}<\lambda$. Then it follows from the expansion (1.1) and Theorems 1 and 2 that $\hat{h}/h_0\to 1$, $\beta_n(h_0)/I_n(h_0)\to 1$ and $\beta_n(\hat{h})/f_n(h_0)\to 1$ in probability. Therefore the cross-validatory window size, $\hat{h}$, is asymptotically equivalent to h_0.

3. Proofs

We begin by introducing some notation. Let H_n denote the interval $[\eta n^{-1/5}, \lambda n^{-1/5}]$, and define $\Delta_i = Y_i - \mu(X_i)$,

$$v(x) = v(x \mid h) = E[\{\mu(x) - \mu(X)\} K\{(x-X)/h\}],$$

$$f_n(x \mid h) = (nh)^{-1} \sum_{i=1}^{n} K\{(x-X_i)/h\}$$

and

$$F_n(x) = n^{-1} \sum_{i=1}^{n} I(X_i \leqq x)$$

(the empiric distribution function of the X-sample). Let $F(x) = E\{F_n(X)\}$ denote the true distribution function, and $\mathscr{F}_n = \mathscr{F}\{X_1, \ldots, X_n\}$ denote the σ-field generated by the X-sample. Write E' for expectation conditional on $\mathscr{F}_n$, $\sum_i^*$ for summation over i such that $X_i \in A$ and $1 \leqq i \leqq n$, and $\sum_j^{(\delta)}$ for summation over j such that $X_j \in A^\delta$ and $1 \leqq j \leqq n$. Let $C, C_1, C_2, \ldots$ denote positive generic constants. Arrange the variables $X_1, \ldots, X_n$ in order of increasing magnitude, obtaining $X_{(1)} \leqq X_{(2)} \leqq \ldots \leqq X_{(n)}$, and set $Y_{(i)} = Y_j$ and $\Delta_{(i)} = \Delta_j$ if $X_{(i)} = X_j$.

Observe that

$$n^{-1} \sum_i^* \{Y_i - \hat{\mu}_{ni}(X_i)\}^2 = 2n^{-1} \sum_i^* \{Y_i - \mu(X_i)\}\{\mu(X_i) - \hat{\mu}_{ni}(X_i)\} + n^{-1} \sum_i^* \{\mu(X_i) - \hat{\mu}_{ni}(X_i)\}^2 + n^{-1} \sum_i^* \{Y_i - \mu(X_i)\}^2. \tag{3.1}$$

The proof is divided into two main sections, the first showing that the first term on the right hand side in (3.1) equals $o(n^{-4/5})$ in probability, and the second showing that the second term equals mean integrated square error plus a remainder of $o(n^{-4/5})$ in probability. Each section is divided into subsections, which control individual components in expansions of these two terms.

(I) Note that

$$n^{-1} \sum_i^* \{Y_i - \mu(X_i)\}\{\mu(X_i) - \hat{\mu}_{ni}(X_i)\} = -T_1 + T_2 + T_3, \tag{3.2}$$

where

$$T_1 = n^{-1} \sum_i^* \Delta_i [\sum_{j \neq i} \Delta_j K\{(X_i - X_j)/h\}][\sum_{j \neq i} K\{(X_i - X_j)/h\}]^{-1},$$

$$T_2 = n^{-1} \sum_i^* \Delta_i [\sum_{j \neq i} \{\mu(X_i) - \mu(X_j)\} K\{(X_i - X_j)/h\} - (n-1) v(X_i)] \times [\sum_{j \neq i} K\{(X_i - X_j)/h\}]^{-1}$$

and

$$T_3 = (1 - n^{-1}) \sum_i^* \Delta_i v(X_i) [\sum_{j \neq i} K\{(X_i - X_j)/h\}]^{-1}.$$

We shall handle these terms individually, in Sects. (I.i)–(I.iii) below. Our proof is prefaced by three lemmas, the first of which is easily derived. Let $\phi(x \mid h) = E\{f_n(x \mid h)\}$.

Lemma 1. *Under condition* (2.2),

$$\sup_{x\in A} |\phi(x\,|\,h) - f(x) - (1/24)\,h^2 f''(x)| = O(h^{5/2})$$

as $h \to 0$.

Lemma 2.

$$|f_n(x\,|\,h) - f(x)| = O_p(n^{-2/5} \log^{\frac{1}{2}} n) \tag{3.3}$$

uniformly in $x \in A$ *and* $h \in H_n$. *Furthermore, if* $\alpha > 2/5$ *then*

$$|f_n(x\,|\,h_1) - f_n(x\,|\,h_2)| = O_p(n^{-\frac{1}{2}}) \tag{3.4}$$

uniformly in $x \in A$ *and values* $h_1, h_2 \in H_n$ *satisfying* $|h_1 - h_2| \leqq n^{-\alpha}$.

Proof. Applying the Komlós-Major-Tusnády approximation [15] to the empiric distribution function F_n, we see that there exist Brownian bridges W_n^0, $n \geqq 1$, such that

$$f_n(x\,|\,h) - \phi(x\,|\,h) = n^{-\frac{1}{2}} h^{-1} [W_n^0\{F(x + \tfrac{1}{2}h)\} - W_n^0\{F(x - \tfrac{1}{2}h)\}] + O_p(n^{-4/5} \log n) \tag{3.5}$$

uniformly in $x \in A$ and $h \in H_n$. The modulus of continuity of W^0,

$$w^0(u) \equiv \sup_{0<s,\,t<1;\;|s-t|\leqq u} |W^0(s) - W^0(t)|,$$

is dominated by $Z\{u(1 - \log u^{-1})\}^{\frac{1}{2}}$ for an almost surely finite variable Z. Therefore $|f_n(x\,|\,h) - \phi(x\,|\,h)| = O_p(n^{-2/5} \log^{\frac{1}{2}} n)$, uniformly in $x \in A$ and $h \in H_n$. The result (3.3) follows from this estimate and Lemma 1.

By Lemma 1,

$$\sup_{x\in A} |\phi(x\,|\,h_1) - \phi(x\,|\,h_2)| \leqq C|h_1^2 - h_2^2| + O(n^{-\frac{1}{2}}) \tag{3.6}$$

uniformly in $h_1, h_2 \in H_n$ satisfying $|h_1 - h_2| \leqq n^{-\alpha}$. Furthermore,

$$\sup_{x\in A} |W_n^0\{F(x \pm h_1)\} - W_n^0\{F(x \pm \tfrac{1}{2}h_2)\}| = O_p\{(n^{-\alpha} \log n)^{\frac{1}{2}}\} = O_p(n^{-1/5})$$

uniformly in $h_1, h_2 \in H_n$ satisfying $|h_1 - h_2| \leqq n^{-\alpha}$. The result (3.4) follows from this estimate, (3.5) and (3.6).

Lemma 3.

$$U_n \equiv \sup_{x\in A;\, h\in H_n} \left| \sum_{j=1}^{n} \Delta_{(j)} K\{(x - X_{(j)})/h\} \right| = O_p(n^{9/20}), \tag{3.7}$$

$$\sup_{1\leqq i\leqq n;\, x\in A;\, h\in H_n} |\sum_{j\neq i} \Delta_{(j)} K\{(x - X_{(j)})/h\}| = O_p(n^{9/20}) \tag{3.8}$$

and

$$\sup_{1\leqq r\leqq n} \left| \sum_{j=1}^{r}{}^{(\frac{1}{2}\varepsilon)} \Delta_j \right| + \sup_{1\leqq r\leqq n} \left| \sum_{j=1}^{r}{}^{(\frac{1}{2}\varepsilon)} \Delta_{(j)} \right| = O_p(n^{\frac{1}{2}}). \tag{3.9}$$

Proof. For large n,

$$U_n \leqq \sup_{0 \leqq r < s \leqq n;\, s-r \leqq m} \left| \sum_{j=r+1}^{s}{}^{(\frac{1}{2}\varepsilon)} \Delta_{(j)} \right|,$$

$$\leqq \sup_{0 \leqq i \leqq n/m+1;\, 1 \leqq r \leqq m;\, im+r \leqq n} 4 \left| \sum_{j=im+1}^{im+r}{}^{(\frac{1}{2}\varepsilon)} \Delta_{(j)} \right|, \tag{3.10}$$

where $m = n \sup_{x \in A} \{F_n(x+\frac{1}{2}\lambda n^{-1/5}) - F_n(x-\frac{1}{2}\lambda n^{-1/5})\} + 1$. The first inequality follows on noting that the term $K\{(x-x_{(j)})/h\}$ appearing within modulus signs in (3.7), equals unity for values of j in the range $r+1 \leqq j \leqq s$, for some r and s, and zero for all other values of j. Consequently

$$4^{-4} E'(U_n^4) \leqq \sum_{i=0}^{n/m+1} E' \left\{ \sup_{1 \leqq r \leqq m} \left| \sum_{j=im+1}^{im+r}{}^{(\frac{1}{2}\varepsilon)} \Delta_{(j)} \right|^4 \right\}$$
$$= C_1 \sum_{i=0}^{n/m+1} \left[\left\{ \sum_{j=im+1}^{im+m}{}^{(\frac{1}{2}\varepsilon)} \sigma^2(X_{(j)}) \right\}^2 + \sum_{j=im+1}^{im+m}{}^{(\frac{1}{2}\varepsilon)} E'(\Delta_{(j)}^4) \right]$$
$$\leqq C_2 (n/m)\, m^2 = C_2\, nm,$$

using a version of Rosenthal's inequality [1, (2.1.5), p. 40]. Therefore $E'(U_n) = O_p\{(nm)^{1/4}\}$, and since $m = O_p(n^{4/5})$, the result (3.7) follows. We may deduce from (3.10) that whenever $m \geqq 1$,

$$\sup_{X_{(i)} \in A^{\frac{1}{2}\varepsilon}} |\Delta_{(i)}| \leqq \sup_{0 \leqq r < s \leqq n;\, s-r \leqq m} 2 \left| \sum_{j=r+1}^{s}{}^{(\frac{1}{2}\varepsilon)} \Delta_{(j)} \right| = O_p(n^{9/20}),$$

and so (3.7) implies (3.8). The result (3.9) is proved similarly.

(I.i) T_1: We may deduce from (3.3) of Lemma 2 that

$$[(nh)^{-1} \sum_{j \neq i} K\{(x-X_{(j)})/h\}]^{-1}$$
$$= \{f(x)\}^{-1} - \{f(x)\}^{-2} \{f_n(x \mid h) - f(x)\} + O_p(n^{-4/5} \log n) \tag{3.11}$$

uniformly in $1 \leqq i \leqq n$, $x \in A$ and $h \in H_n$. Combining (3.8) and (3.11) we see that

$$[\sum_{j \neq i} \Delta_{(j)} K\{(x-X_{(j)})/h\}] [\sum_{j \neq i} K\{(x-X_{(j)})/h\}]^{-1}$$
$$= (nh)^{-1} [\sum_{j \neq i} \Delta_{(j)} K\{(x-X_{(j)})/h\}] [\{f(x)\}^{-1} - \{f(x)\}^{-2} \{f_n(x \mid h) - f(x)\}]$$
$$+ O_p(n^{-23/20} \log n) \tag{3.12}$$

uniformly in $1 \leqq i \leqq n$, $x \in A$ and $h \in H_n$. Since $n^{-1} \sum_i^{(\frac{1}{2}\varepsilon)} |\Delta_i| = O_p(1)$, we may deduce from (3.12) and the definition of T_1 that

$$T_1 = 2T_{11} - T_{12} + o_p(n^{-4/5}) \tag{3.13}$$

uniformly in $h \in H_n$, where

$$T_{11} = n^{-2} h^{-1} \sum_i{}^{*} \Delta_{(i)} \{f(X_{(i)})\}^{-1} \sum_{j \neq i} \Delta_{(j)} K\{(X_{(i)} - X_{(j)})/h\}$$

and

$$T_{12}=n^{-2}h^{-1}\sum_i^* \Delta_{(i)}\{f(X_{(i)})\}^{-2}f_n(X_{(i)}\,|\,h)\sum_{j\neq i}\Delta_{(j)}K\{(X_{(i)}-X_{(j)})/h\}.$$

We shall prove that

$$\sup_{h\in H_n}|T_{12}|=o_p(n^{-4/5}) \tag{3.14}$$

as $n\to\infty$. A similar (but shorter) argument may be used to show that

$$\sup_{h\in H_n}|T_{11}|=o_p(n^{-4/5}).$$

Let $\alpha>2/5$, and break the interval H_n up into N or $N+1$ disjoint subintervals $[t_{i-1}, t_i)$, each of precise length $n^{-\alpha}$, where N equals the integer part of $(\lambda-\eta)\,n^{\alpha-1/5}$ and $\eta n^{-1/5}=t_0<t_1<\ldots<t_N\leqq\lambda n^{-1/5}<t_{N+1}$. Given $h\in H_n$, let $k=k(h)$ equal the positive integer such that $h\in[t_k, t_{k+1})$. Write $T_{12}=T_{13}+T_{14}+T_{15}(k(h))$, where

$$T_{13}=n^{-2}h^{-1}\sum_i^* \Delta_{(i)}\{f(X_{(i)})\}^{-2}\{f_n(X_{(i)}\,|\,h)-f_n(X_{(i)}\,|\,t_k)\}$$
$$\times\sum_{j\neq i}\Delta_{(j)}K\{(X_{(i)}-X_{(j)})/h\},$$
$$T_{14}=n^{-2}h^{-1}\sum_i^* \Delta_{(i)}\{f(X_{(i)})\}^{-2}f_n(X_{(i)}\,|\,t_0)\sum_{j\neq i}\Delta_{(j)}K\{(X_{(i)}-X_{(j)})/h\},$$

and for a general l,

$$T_{15}(l)=n^{-2}h^{-1}\sum_i^* \Delta_{(i)}\{f(X_{(i)})\}^{-2}\{f_n(X_{(i)}\,|\,t_l)-f_n(X_{(i)}\,|\,t_0)\}$$
$$\times\sum_{j\neq i}\Delta_{(j)}K\{(X_{(i)}-X_{(j)})/h\}.$$

Then

$$|T_{13}|\leqq n^{-1}h^{-1}\{\inf_{x\in A}f(x)\}^{-2}\sup_{\substack{x\in A;\,h_1,\,h_2\in A\\ \text{s.t. }|h_1-h_2|\leqq n^{-\alpha}}}|f_n(x\,|\,h_1)-f_n(x\,|\,h_2)|$$
$$\times\sup_{1\leqq i\leqq n;\,x\in A;\,h\in H_n}|\sum_{j\neq i}\Delta_{(j)}K\{(X_{(i)}-X_{(j)})/h\}|\,(n^{-1}\sum_i^*|\Delta_i|).$$

It now follows from (3.4) and (3.8) that

$$|T_{13}|=O_p(n^{-4/5}\,n^{-1/2}\,n^{9/20})=o_p(n^{-4/5}) \tag{3.15}$$

uniformly in $h\in H_n$.

Let w_{ni} stand for either $\{f(X_{(i)})\}^{-2}f_n(X_{(i)}\,|\,t_0)$ or

$$\{f(X_{(i)})\}^{-2}\{f_n(X_{(i)}\,|\,t_l)-f_n(X_{(i)}\,|\,t_0)\}.$$

Then we may write both T_{14} and $T_{15}(l)$ in the form

$$T_{16}=n^{-2}h^{-1}\sum_i^* \Delta_{(i)}w_{ni}\sum_{j\neq i}\Delta_{(j)}K\{(X_{(i)}-X_{(j)})/h\},$$

where w_{ni} does not depend on h and is measurable in $\mathscr{F}_n$. We may further subdivide T_{16} into two parts, in which the inner sum is taken first over $j \geqq i+1$ and then over $j \leqq i-1$. Both series may be handled similarly, and so we shall treat only the first, equal to $n^{-2} h^{-1} T_{17}$, where

$$T_{17} = \sum_i^* \Delta_{(i)} w_{ni} \sum_{j=i+1}^{r(i,h)} \Delta_{(j)}$$

and $r(i, h)$ is the largest j such that $|X_{(i)} - X_{(j)}| \leqq \frac{1}{2}h$ and $j \leqq n$. For each i, $r(i, \cdot)$ is an $\mathscr{F}_n$-measurable function and right continuous with left hand limits. Let $u_1 < \ldots < u_M$ be the points of discontinuity of all functions $r(i, \cdot)$, $1 \leqq i \leqq n$, in the interval $(\eta n^{-1/5}, \lambda n^{-1/5}]$. Set $u_0 = \eta n^{-1/5}$, $s(i, p) = r(i, u_p)$ and

$$S_p = \sum_i^* \Delta_{(i)} w_{ni} \sum_{j=i+1}^{s(i,p)} \Delta_{(j)}.$$

(The inner series is taken as zero if $s(i, p) \leqq i$.) Let $Y_0 = S_0$ and $Y_p = S_p - S_{p-1}$, $1 \leqq p \leqq M$, and observe that $E'(Y_p) = 0$ for each p. It may be shown after some algebra that $E'(Y_p Y_q) = 0$ unless $p = q$, and so the variables Y_p are conditionally orthogonal. We may now deduce from the Rademacher-Mensov inequality [16, p. 457] that

$$\begin{aligned} E'(\sup_{h\in H_n} T_{17}^2) &= E'(\sup_{0\leqq p\leqq M} S_p^2) \\ &\leqq (\log 4M/\log 2)^2 E'(S_M^2) \\ &= (\log 4M/\log 2)^2 \sum_i^* \sigma^2(X_{(i)}) w_{ni}^2 \sum_{j=i+1}^{r(i,u_M)} \sigma^2(X_{(j)}). \end{aligned}$$

The integer M does not exceed n^2, and if $X_{(i)} \in A$ then $r(i, u_M)$ is dominated by

$$\sup_{x\in A} n\{F_n(x + \tfrac{1}{2}\lambda n^{-1/5}) - F_n(x - \tfrac{1}{2}\lambda n^{-1/5})\} + 1 = O_p(n^{4/5}), \tag{3.16}$$

using Lemma 2. Therefore

$$E'(\sup_{h\in H_n} T_{17}^2) = O_p(n^{9/5} \log^2 n)(\sup_{X_{(i)}\in A} w_{ni}^2)$$

and

$$E'(\sup_{h\in H_n} T_{16}^2) = O_p(n^{-9/5} \log^2 n)(\sup_{X_{(i)}\in A} w_{ni}^2). \tag{3.17}$$

Since the formula (3.16) does not depend on the weights w_{ni}, then in the special case where $T_{16} \equiv T_{15}(l)$, the formula (3.17) holds uniformly in l:

$$\sup_{0\leqq l\leqq N} E'\{\sup_{h\in H_n} T_{15}^2(l)\} = O_p(n^{-9/5} \log^2 n)\{\sup_{0\leqq l\leqq N} \sup_{X_{(i)}\in A} w_{ni}^2(l)\}. \tag{3.18}$$

When $T_{16} \equiv T_{14}$, $\sup w_{ni}^2 = O_p(1)$, and so by (3.17),

$$\sup_{h\in H_n} |T_{14}| = o_p(n^{-4/5}). \tag{3.19}$$

In the case $T_{16} \equiv T_{15}(l)$,

$$\sup_{0 \leqq l \leqq N} \sup_{X_{(i)} \in A} |w_{ni}(l)| \leqq C \sup_{x \in A;\, h_1, h_2 \in H_n} |f_n(x \mid h_1) - f_n(x \mid h_2)| = O_p(n^{-2/5} \log^{\frac{1}{2}} n),$$

by Lemma 2. Hence by (3.18), and remembering that $N \leqq C n^{\alpha - 1/5}$,

$$E'\left[\sum_{l=0}^{N} \{\sup_{h \in H_n} T_{15}^2(l)\}\right] = O_p(n^{\alpha - 1/5} \cdot n^{-13/5} \log^3 n).$$

Therefore if $2/5 < \alpha < 1$,

$$\sup_{0 \leqq l \leqq N;\, h \in H_n} |T_{15}(l)| \leqq \left[\sum_{l=0}^{N} \{\sup_{h \in H_n} T_{15}^2(l)\}\right]^{\frac{1}{2}} = o_p(n^{-4/5}). \tag{3.20}$$

Combining (3.19) and (3.20) we see that (3.14) holds, and so by (3.13),

$$\sup_{h \in H_n} |T_1| = o_p(n^{-4/5}). \tag{3.21}$$

(I.ii) T_2: We begin with a lemma. Define

$$S(x \mid h) = \sum_j [\{\mu(x) - \mu(X_j)\} K\{(x - X_j)/h\} - v(x)].$$

Lemma 4.

$$\sup_{x \in A;\, h \in H_n} |S(x \mid h)| = O_p(n^{2/5}). \tag{3.22}$$

Furthermore, if $3/5 < \alpha < 1$ *then*

$$|S(x \mid h_1) - S(x \mid h_2)| = O_p(n^{3/10 - \alpha/2} \log^{\frac{1}{2}} n) \tag{3.23}$$

uniformly in $x \in A$ *and* $h_1, h_2 \in H_n$ *satisfying* $|h_1 - h_2| \leqq n^{-\alpha}$.

Proof. Since $|\mu(x+y) - \mu(x) - y\mu'(x)| \leqq C y^2$ uniformly in $x \in A$ and $|y| \leqq \frac{1}{2}\varepsilon$, then if $x \in A$ and $0 \leqq h_1 < h_2 \leqq \lambda n^{-1/5} < \varepsilon$,

$$\left|\sum_j \{\mu(x) - \mu(X_j)\} I(\tfrac{1}{2}h_1 < |x - X_j| \leqq \tfrac{1}{2}h_2) - \mu'(x) \sum_j (x - X_j) I(\tfrac{1}{2}h_1 < |x - X_j| \leqq \tfrac{1}{2}h_2)\right| \leqq \tfrac{1}{4} C h_2^2 \sum_j I(\tfrac{1}{2}h_1 < |x - X_j| \leqq \tfrac{1}{2}h_2). \tag{3.24}$$

First we shall prove (3.22), taking $h_1 = 0$ and $h_2 = h$ in (3.24). Now,

$$\sup_{x \in A;\, h \in H_n} \sum_j I(|x - X_j| \leqq \tfrac{1}{2}h) \leqq \lambda n^{4/5} \sup_{x \in A} f_n(x \mid \lambda n^{4/5}) = O_p(n^{4/5}),$$

and a similar formula holds for the supremum of expectations. Therefore

$$\sup_{x \in A;\, h \in H_n} \left|S(x \mid h) - \mu'(x) \sum_j [(x - X_j) I(|x - X_j| \leqq \tfrac{1}{2}h) - E\{(x - X_j) I(|x - X_j| \leqq \tfrac{1}{2}h)\}]\right| = O_p(n^{2/5}). \tag{3.25}$$

Let $\tilde{S}(x\,|\,h)$ denote the sum of the terms within square brackets in (3.25). Integrating by parts in the identity

$$n^{-1}\tilde{S}(x\,|\,h)=\int_{x-\frac{1}{2}h\leqq y\leqq x+\frac{1}{2}h}(x-y)\,d\{F_n(y)-F(y)\},$$

and using the fact that $\sup|F_n(y)-F(y)|=O_p(n^{-\frac{1}{2}})$, we may deduce that

$$\sup_{x\in A;\,h\in H_n}|\tilde{S}(x\,|\,h)|=O_p(n\cdot n^{-\frac{1}{2}}\cdot n^{-1/5})=O_p(n^{3/10}).$$

The result (3.22) follows from this estimate and (3.25).

To prove (3.23), let $\sup^\dagger$ denote the supremum over $x\in A$ and $h_1, h_2\in H_n$ satisfying $0\leqq h_2-h_1\leqq n^{-\alpha}$. Note that

$$\begin{aligned}\sup{}^\dagger\sum_j I(\tfrac{1}{2}h_1<x-X_j\leqq\tfrac{1}{2}h_2)&=n\sup{}^\dagger\{F_n(x-\tfrac{1}{2}h_1)-F_n(x-\tfrac{1}{2}h_2)\}\\&=O_p(n^{1-\alpha}),\end{aligned}$$

using the Komlós-Major-Tusnády [15] approximation. A similar formula holds for the supremum of expectations, and also for the series

$$\sum_j I(\tfrac{1}{2}h_1<X_j-x\leqq\tfrac{1}{2}h_2).$$

Therefore by (3.24),

$$\begin{aligned}\sup{}^\dagger|\{S(x\,|\,h_2)-S(x\,|\,h_1)\}-\mu'(x)\sum_j[(x-X_j)\,I(\tfrac{1}{2}h_1<|x-X_j|\leqq\tfrac{1}{2}h_2)\\-E\{(x-X_j)\,I(\tfrac{1}{2}h_1<|x-X_j|\leqq\tfrac{1}{2}h_2)\}]|=O_p(n^{3/5-\alpha}).\end{aligned}\tag{3.26}$$

Integrating by parts in the identity

$$\begin{aligned}&n^{-1}\sum_j[(x-X_j)\,I(\tfrac{1}{2}h_1<X_j-x\leqq\tfrac{1}{2}h_2)-E\{(x-X_j)\,I(\tfrac{1}{2}h_1<X_j-x\leqq\tfrac{1}{2}h_2)\}]\\&=\int_{x+\frac{1}{2}h_1<y\leqq x+\frac{1}{2}h_2}(x-y)\,d\{F_n(y)-F(y)\},\end{aligned}$$

we may deduce that the left hand side is dominated by

$$\begin{aligned}&h_2\,|\{F_n(x+\tfrac{1}{2}h_2)-F(x+\tfrac{1}{2}h_2)\}-\{F_n(x+\tfrac{1}{2}h_1)-F(x+\tfrac{1}{2}h_1)\}|\\&+|h_1-h_2|\sup_{y\in A^{\frac{1}{2}\varepsilon}}|F_n(y)-F(y)|\end{aligned}$$

uniformly in $x\in A$ and $h_1, h_2\in H_n$ with $0\leqq h_2-h_1\leqq n^{-\alpha}$, and so equals

$$O_p(n^{-1/5}\,n^{-\frac{1}{2}}\,n^{-\alpha/2}\log^{\frac{1}{2}} n)$$

uniformly, using the Komlós-Major-Tusnády approximation [15]. A similar estimate holds for the series in which $I(\frac{1}{2}h_1<X_j-x\leqq\frac{1}{2}h_2)$ is replaced by $I(\frac{1}{2}h_1<x-X_j\leqq\frac{1}{2}h_2)$. The result (3.23) now follows via (3.26). This proves Lemma 4.

Since $|v(x\,|\,h)|=O(h^3)$ uniformly in $x\in A$ as $h\to 0$, then if we replace $S(x\,|\,h)$ by $R(x\,|\,h)=S(x\,|\,h)+v(x\,|\,h)$, the results (3.22) and (3.23) continue to hold. Let (3.22)′ and (3.23)′ denote these alternative results.

We may deduce from (3.22)′ and the expansion (3.11) that

$$T_2 = 2T_{21} - T_{22} + O_p(n^{-6/5}\log n) \tag{3.27}$$

uniformly in $h \in H_n$, where

$$T_{21} = n^{-2}h^{-1}\sum_i{}^* \Delta_i\{f(X_i)\}^{-1}R(X_i \mid h)$$

and

$$T_{22} = n^{-2}h^{-1}\sum_i{}^* \Delta_i\{f(X_i)\}^{-2}f_n(X_i \mid h)\,R(X_i \mid h).$$

Our aim is to show that $\sup_{h\in H_n}|T_{21}| = o_p(n^{-2/5})$ and

$$\sup_{h\in H_n}|T_{22}| = o_p(n^{-4/5}). \tag{3.28}$$

We shall prove only the latter result, since the former is simpler.

Divide H_n into intervals $[t_{i-1}, t_i)$ of length $n^{-\alpha}$, as in (I.i). Recall that $k = k(h)$ is defined by $h \in [t_k, t_{k-1})$. Set $T_{22} = T_{23} + T_{24} + T_{25}(k(h))$, where

$$T_{23} = n^{-2}h^{-1}\sum_i{}^* \Delta_i\{f(X_i)\}^{-2}\{f_n(X_i \mid h) - f_n(X_i \mid t_k)\}\,R(X_i \mid h),$$

$$T_{24} = n^{-2}h^{-1}\sum_i{}^* \Delta_i\{f(X_i)\}^{-2}f_n(X_i \mid t_k)\{R(X_i \mid h) - R(X_i \mid t_k)\}$$

and

$$T_{25}(l) = n^{-2}h^{-1}\sum_i{}^* \Delta_i\{f(X_i)\}^{-2}f_n(X_i \mid t_l)\,R(X_i \mid t_l).$$

Using (3.4), (3.22)′ and the fact that $n^{-1}\sum_i{}^* |\Delta_i| = O_p(1)$, we obtain

$$\sup_{h\in H_n}|T_{23}| = O_p(n^{-1}\cdot n^{1/5}\cdot n^{-\frac{1}{2}}\cdot n^{2/5}) = o_p(n^{-4/5}). \tag{3.29}$$

From (3.23)′ we may deduce that if $3/5 < \alpha < 1$,

$$\sup_{h\in H_n}|T_{24}| = O_p(n^{-1}\cdot n^{1/5}\cdot n^{3/10-\alpha/2}\log^{\frac{1}{2}} n) = o_p(n^{-4/5}). \tag{3.30}$$

For each l,

$$E'\{n^2hT_{25}(l)\}^2 = \sum_i{}^* \sigma^2(X_i)\{f(X_i)\}^{-4}f_n^2(X_i \mid t_l)\,R^2(X_i \mid t_l),$$

and so

$$\sup_{h\in H_n} T_{25}^2(k(h)) = O_p(n^{-18/5})\sum_i{}^*\sum_{l=0}^{N} R^2(X_i \mid t_l).$$

But if $x \in A$,

$$\begin{aligned} E\{R^2(X_i \mid t_l) \mid X_i = x\} &= (n-1)\,\mathrm{var}\,[\{\mu(x)-\mu(X)\}\,K\{(x-X)/t_l\}] \\ &\le n\sup_{x\in A}\int_{-\frac{1}{2}t_l}^{\frac{1}{2}t_l}\{\mu(x)-\mu(x+u)\}^2 f(x+u)\,du \\ &\le Cn(\lambda n^{-1/5})^3 \end{aligned}$$

uniformly in l. Consequently, since $N \leqq Cn^{\alpha-1/5}$,

$$\sup_{h \in H_n} T_{25}^2(k(h)) = O_p(n^{-18/5} \cdot n \cdot n^{\alpha-1/5} \cdot n^{2/5}) = o_p(n^{-8/5}),$$

provided $3/5 < \alpha < 4/5$. The result (3.28) follows from this estimate, (3.29) and (3.30). We may now deduce from (3.27) that

$$\sup_{h \in H_n} |T_2| = o_p(n^{-4/5}). \tag{3.31}$$

(I.iii) T_3: Since $|v(x)| = O(h^3)$ uniformly in $x \in A$ as $h \to 0$, then by (3.11),

$$T_3 = 2T_{31} - T_{32} + O_p(n^{-6/5} \log n) \tag{3.32}$$

uniformly in $h \in H_n$, where

$$T_{31} = (1-n^{-1})\, n^{-1} h^{-1} \sum_i{}^* \Delta_i\, v(X_i) \{f(X_i)\}^{-1}$$

and

$$T_{32} = (1-n^{-1})\, n^{-1} h^{-1} \sum_i{}^* \Delta_i\, v(X_i) \{f(X_i)\}^{-2} f_n(X_i \mid h).$$

We shall prove that

$$\sup_{h \in H_n} |T_{32}| = o_p(n^{-4/5}). \tag{3.33}$$

Similarly, it may be shown that $|T_{31}| = o_p(n^{-4/5})$ uniformly in $h \in H_n$.

Divide H_n into N or $N+1$ intervals $[t_{i-1}, t_i)$ of length $n^{-\alpha}$, as before, and define $k(h)$ by $h \in [t_k, t_{k+1})$. For definiteness we shall take $\alpha = 7/10$, so that $3/5 < \alpha < 4/5$. Let $\beta = 13/40$ and $m = [n^\beta]$. Given an integer l in the interval $[0, N]$, define $t_l' = t_{m[l/m]}$. Observe that

$$T_{32} = (1-n^{-1})\, h^{-1} \left\{ \sum_{j=3}^{4} T_{3j} + \sum_{j=5}^{7} T_{3j}(k(h)) \right\}, \tag{3.34}$$

where

$$T_{33} = n^{-1} \sum_i{}^* \Delta_i \{f(X_i)\}^{-2} \{f_n(X_i \mid h) - f_n(X_i \mid t_k)\}\, v(X_i \mid h),$$

$$T_{34} = n^{-1} \sum_i{}^* \Delta_i \{f(X_i)\}^{-2} f_n(X_i \mid t_k) \{v(X_i \mid h) - v(X_i \mid t_k)\},$$

$$T_{35}(l) = n^{-1} \sum_i{}^* \Delta_i \{f(X_i)\}^{-2} \{f_n(X_i \mid t_l) - f_n(X_i \mid t_0)\}\, v(X_i \mid t_l),$$

$$T_{36}(l) = n^{-1} \sum_i{}^* \Delta_i \{f(X_i)\}^{-2} f_n(X_i \mid t_0) \{v(X_i \mid t_l) - v(X_i \mid t_l')\}$$

and

$$T_{37}(l) = n^{-1} \sum_i{}^* \Delta_i \{f(X_i)\}^{-2} f_n(X_i \mid t_0)\, v(X_i \mid t_l').$$

In view of (3.4),

$$\sup_{h \in H_n} |T_{33}| = O_p(n^{-\frac{1}{2}} \cdot n^{-3/5}) = o_p(n^{-1}). \tag{3.35}$$

Under the conditions imposed on μ and f,

$$|\{\mu(x)-\mu(x+u)\}f(x+u)+u\mu'(x)f(x)|\leqq Cu^2$$

uniformly in $x\in A$ and $|u|\leqq\frac{1}{2}\varepsilon$. Therefore

$$\begin{aligned}|v(x\,|\,h_2)-v(x\,|\,h_1)| &= \left|\int_{\frac{1}{2}h_1}^{\frac{1}{2}h_2}\{\mu(x)-\mu(x+u)\}f(x+u)\,du\right.\\ &\qquad \left.-\int_{-\frac{1}{2}h_2}^{-\frac{1}{2}h_1}\{\mu(x)-\mu(x+u)\}f(x+u)\,du\right|\\ &\leqq Cn^{-2/5}|h_1-h_2|\end{aligned}$$

uniformly in $x\in A$ and $h_1, h_2\in H_n$. Hence

$$\sup_{h\in H_n}|T_{34}|=O_p(n^{-2/5-\alpha})=o_p(n^{-1}). \tag{3.36}$$

It follows from (3.3) that

$$\begin{aligned}&\sup_{0\leqq l\leqq N} E'\{T_{35}^2(l)\}\\ &\quad=n^{-2}\sup_{0\leqq l\leqq N}\sum_i{}^{*}\sigma^2(X_i)\{f(X_i)\}^{-4}\{f_n(X_i\,|\,t_l)-f_n(X_i\,|\,t_0)\}^2v^2(X_i\,|\,t_0)\\ &\quad=O_p\{n^{-1}\cdot n^{-4/5}(\log n)\cdot n^{-6/5}\},\end{aligned}$$

and so

$$\begin{aligned}\sup_{0\leqq l\leqq N}|T_{35}(l)| &\leqq\left\{\sum_{l=0}^{N}T_{35}^2(l)\right\}^{\frac{1}{2}}\\ &=O_p\{(n^{\alpha-1/5}\cdot n^{-3}\log n)^{\frac{1}{2}}\}=o_p(n^{-1}).\end{aligned} \tag{3.37}$$

Similarly, since

$$\sup_{x\in A;\,0\leqq l\leqq N}|v(x\,|\,t_l)-v(x\,|\,t_l')|\leqq Cn^{-2/5}\cdot n^{-(\alpha-\beta)},$$

then

$$\sup_{0\leqq l\leqq N}|T_{36}(l)|=O_p\{(n^{\alpha-1/5}\cdot n^{-1}\cdot n^{-4/5-2(\alpha-\beta)})^{\frac{1}{2}}\}=o_p(n^{-1}). \tag{3.38}$$

As l ranges over the integers $0,\dots,N$, t_l' takes no more than $[N/M]+1=O(n^{\alpha-\beta-1/5})$ values. Since $\sup_{0\leqq l\leqq N}E'\{T_{37}^2(l)\}=O_p(n^{-11/5})$ then

$$\sup_{0\leqq l\leqq N}|T_{37}(l)|=O_p\{(n^{\alpha-\beta-1/5}\cdot n^{-11/5})^{\frac{1}{2}}\}=o_p(n^{-1}). \tag{3.39}$$

The result (3.33) follows on combining (3.25)–(3.39). We may now deduce from (3.32) that

$$\sup_{h\in H_n}|T_3|=o_p(n^{-4/5}). \tag{3.40}$$

It follows from (3.2), (3.21), (3.31) and (3.40) that

$$\sup_{h\in H_n}|n^{-1}\sum_i{}^{*}\{Y_i-\mu(X_i)\}\{\mu(X_i)-\hat{\mu}_{ni}(X_i)\}|=o_p(n^{-4/5}). \tag{3.41}$$

This takes care of the first term on the right hand side of (3.1). We now examine the second term.

(II). We may deduce from (3.3) that

$$\sup_{1 \leq i \leq n; x \in A} |[\sum_{j \neq i} K\{(x-X_j)/h\}]/[\sum_j K\{(x-X_j)/h\}]-1| \to 0$$

is probability as $n \to \infty$. Consequently

$$\begin{aligned}\{1+o_p(1)\} n^{-1} &\sum_i^* \{\mu(X_i)-\mu_{ni}(X_i)\}^2 \\ &= n^{-1} \sum_i^* ([\sum_{j \neq i} \{Y_j-\mu(X_i)\} K\{(X_i-X_j)/h\}][\sum_j K\{(X_i-X_j)/h\}]^{-1})^2 \\ &= n^{-1} \sum_i^* \{\mu(X_i)-\hat{\mu}_n(X_i)\}^2 - 2T_4 + T_5, \end{aligned} \tag{3.42}$$

where

$$T_4 = n^{-1} \sum_i^* \Delta_i [\sum_j \{Y_j-\mu(X_i)\} K\{(X_i-X_j)/h\}][\sum_j K\{(X_i-X_j)/h\}]^{-2}$$

and

$$T_5 = n^{-1} \sum_i^* \Delta_i^2 [\sum_j K\{(X_i-X_j)/h\}]^{-2}.$$

Now,

$$\begin{aligned}\sum_j &\{Y_j-\mu(X_i)\} K\{(X_i-X_j)/h\} \\ &= \sum_j \Delta_j K\{(X_i-X_j)/h\} + \sum_j \{\mu(X_j)-\mu(X_i)\} K\{(X_i-X_j)/h\} \\ &= O_p(n^{9/20})\end{aligned}$$

uniformly in $h \in H_n$, by (3.7) and (3.22). Therefore $T_4 = O_p(n^{-23/20})$ uniformly in $h \in H_n$, and it is easily proved that $T_5 = O_p(n^{-8/5})$ uniformly in $h \in H_n$. Hence by (3.42),

$$n^{-1} \sum_i^* \{\mu(X_i)-\hat{\mu}_{ni}(X_i)\}^2 = \{1+o_p(1)\} n^{-1} \sum_i^* \{\mu(X_i)-\hat{\mu}_n(X_i)\}^2 + o_p(n^{-4/5}), \tag{3.43}$$

uniformly in $h \in H_n$.

Next we examine the first term on the right hand side in (3.43), which we write asymptotically as

$$\begin{aligned}\{1+o_p(1)\} n^{-1} &\sum_i^* \{\mu(X_i)-\hat{\mu}_n(X_i)\}^2 \\ &= n^{-3} h^{-2} \sum_i^* \{f(X_i)\}^{-2} [\sum_j \{Y_j-\mu(X_i)\} K\{(X_i-X_j)/h\}]^2 \\ &= n^{-3} h^{-2} (\sum_i^* \{f(X_i)\}^{-2} [\sum_j \Delta_j K\{(X_i-X_j)/h\}]^2 \\ &\quad + 2 \sum_i^* \{f(X_i)\}^{-2} [\sum_j \Delta_j K\{(X_i-X_j)/h\}] \\ &\quad \times [\sum_j \{\mu(X_j)-\mu(X_j)\} K\{(X_i-X_j)/h\}] \\ &\quad + \sum_i^* \{f(X_i)\}^{-2} [\sum_j \{\mu(X_j)-\mu(X_i)\} K\{(X_i-X_j)/h\}]^2) \\ &= n^{-3} h^{-2} (T_6 + 2T_7 + T_8), \end{aligned} \tag{3.44}$$

say. We shall treat these terms individually.

(II.i) T_6: Divide H_n into N or $N+1$ intervals $[t_{i-1}, t_i)$ of length $n^{-\alpha}$, as before, and define $k(h)$ by $h \in [t_k, t_{k+1})$. We assume $3/5 < \alpha < 1$. Now,

$$\begin{aligned}
&|T_6(h) - T_6(t_k)| \\
&\quad \leqq C \sum_i{}^* |\sum_j{}' \Delta_{(j)} \{I(|X_{(i)} - X_{(j)}| \leqq \tfrac{1}{2}h) - I(|X_{(i)} - X_{(j)}| \leqq \tfrac{1}{2}t_k)\}| \\
&\qquad \times |\sum_j{}' \Delta_{(j)} \{I(|X_{(i)} - X_{(j)}| \leqq \tfrac{1}{2}h) + I(|X_{(i)} - X_{(j)}| \leqq \tfrac{1}{2}t_k)\}| \\
&\quad \leqq 4C \left\{ \sup_{1 \leqq r \leqq n} \left| \sum_{j=1}^{r}{}' \Delta_{(j)} \right| \right\} \sum_i{}^* \{\sup{}^\dagger |\sum_j{}' \Delta_{(j)} I(X_{(i)} - \tfrac{1}{2}\delta \leqq X_{(j)} < X_{(i)} - \tfrac{1}{2}t_l)| \\
&\qquad + \sup{}^\dagger |\sum_j{}' \Delta_{(j)} I(X_{(i)} + \tfrac{1}{2}t_l < X_{(j)} \leqq X_{(i)} + \tfrac{1}{2}\delta)|\},
\end{aligned} \tag{3.45}$$

where on this occasion, $\sup^\dagger$ denotes the supremum over $0 \leqq l \leqq N$ and values δ satisfying $t_l \leqq \delta \leqq t_l + n^{-\alpha}$. We shall prove next that if $\alpha > 3/5$,

$$\sum_i{}^* \sup{}^\dagger |\sum_j{}' \Delta_{(j)} I(X_{(i)} + \tfrac{1}{2}t_l < X_{(j)} \leqq X_{(i)} + \tfrac{1}{2}\delta)| = o_p(n^{13/10}). \tag{3.46}$$

From this estimate, a similar result in which the indicator function is replaced by $I(X_{(i)} - \frac{1}{2}\delta \leqq X_{(j)} < X_{(i)} - \frac{1}{2}t_l)$, and (3.9) and (3.45), it follows that

$$\sup_{h \in H_n} |T_6(h) - T_6(t_k)| = o_p(n^{9/5}). \tag{3.47}$$

Define

$$\begin{aligned}
M(x) &\equiv \sup{}^\dagger |\sum_j{}' \Delta_{(j)} I(x + \tfrac{1}{2}t_l < X_{(j)} \leqq x + \tfrac{1}{2}\delta)|^4 \\
&\leqq \sum_{l=0}^{N} \sup_{t_l \leqq \delta < t_l + n^{-\alpha}} |\sum_j{}' \Delta_{(j)} I(x + \tfrac{1}{2}t_l < X_{(j)} \leqq x + \tfrac{1}{2}\delta)|^4,
\end{aligned}$$

where $x \in A$. Then

$$\begin{aligned}
E'\{M(x)\} &\leqq C_1 \sum_{l=0}^{N} [\{\sum_j{}' \sigma^2(X_{(j)}) I(x + \tfrac{1}{2}t_l < X_{(j)} \leqq x + \tfrac{1}{2}t_l + \tfrac{1}{2}n^{-\alpha})\}^2 \\
&\qquad + \sum_{j: x + \frac{1}{2}t_l < X_{(j)} \leqq x + \frac{1}{2}t_l + \frac{1}{2}n^{-\alpha}} E'(\Delta_{(j)}^4)] \\
&\leqq C_2 n^{\alpha - 1/5} [n \sup_{y \in A^{\frac{1}{2}\varepsilon}} \{F_n(y + \tfrac{1}{2}n^{-\alpha}) - F_n(y)\}]^2,
\end{aligned}$$

where neither C_1 nor C_2 depends on x or l. (The first inequality follows from [1, (2.1.5), p. 40].) Therefore the left hand side of (3.46), which equals $\sum_i^* \{M(X_{(i)})\}^{1/4}$, has conditional mean (given $X_1, \ldots, X_n$) dominated by

$$C n^{29/20 + \alpha/4} [\sup_{y \in A^{\frac{1}{2}\varepsilon}} \{F_n(y + \tfrac{1}{2}n^{-\alpha}) - F_n(y)\}]^2 = O_p(n^{29/20 - \alpha/4}) = o_p(n^{13/10}),$$

provided $\alpha > 3/5$. This proves (3.46).

Next observe that $T_6(t_l) = 2T_{61}(l) + T_{62}(l)$, where

$$T_{61}(l) = \sum_i{}^* \{f(X_{(i)})\}^{-2} \sum_{j<k} \sum \Delta_{(j)} \Delta_{(k)} I(X_{(k)} - \tfrac{1}{2}t_l \leqq X_{(i)} \leqq X_{(j)} + \tfrac{1}{2}t_l)$$

and

$$T_{62}(l)=\sum_i^* \{f(X_i)\}^{-2}\sum_j \Delta_j^2 I(X_j-\tfrac{1}{2}t_l \leqq X_i \leqq X_j+\tfrac{1}{2}t_l).$$

Let $m=n^\beta$, where $0<\beta<\alpha-1/5$, and set $l'=m[l/m]$. Then

$$\begin{aligned} E'\{T_{61}(l)-T_{61}(l')\}^2 = \sum_{j<k}\sum \sigma^2(X_{(j)})\,\sigma^2(X_{(k)})\Big[\sum_i^* \{f(X_{(i)})\}^{-2} \\ \times\{I(X_{(k)}-\tfrac{1}{2}t_l \leqq X_{(i)} \leqq X_{(j)}+\tfrac{1}{2}t_l) \\ -I(X_{(k)}-\tfrac{1}{2}t_{l'} \leqq X_{(i)} \leqq X_{(j)}+\tfrac{1}{2}t_{l'})\}\Big]^2, \end{aligned}$$

and the absolute value of the term within square brackets is dominated by

$$\begin{aligned} C_1 \sum_i^* \{I(X_{(k)}-\tfrac{1}{2}t_l \leqq X_{(i)} < X_{(k)}-\tfrac{1}{2}t_{l'}) + I(X_{(j)}+\tfrac{1}{2}t_{l'} < X_{(i)} \leqq X_{(j)}+\tfrac{1}{2}t_l)\} \\ \leqq C_2 n \sup_{x\in A^{\frac{1}{2}\varepsilon}} \{F_n(x+t_l-t_{l'})-F_n(x-)\} = O_p(n^{1-(\alpha-\beta)}) \end{aligned}$$

uniformly in j,k and l. Therefore

$$\sup_{0\leqq l\leqq N} E'\{T_{61}(l)-T_{61}(l')\}^2 = O_p(n^{4-2(\alpha-\beta)}),$$

whence

$$\begin{aligned} E'\{\sup_{0\leqq l\leqq N} |T_{61}(l)-T_{61}(l')|\} &\leqq \Big[\sum_{l=0}^{N} E'\{T_{61}(l)-T_{61}(l')\}^2\Big]^{\frac{1}{2}} \\ &= O_p(n^{19/10+\beta-\alpha/2}) = o_p(n^{9/5}), \end{aligned} \tag{3.48}$$

provided we choose $\beta<\alpha/2-1/10$. Repeating this argument but with α, β, m and l' replaced by $\alpha-\beta$, γ, $p=[n^\gamma]$ and $l''=p[l'/p]$, respectively, where $0<\gamma<\alpha-\beta-1/5$, we see that if $\gamma<(\alpha-\beta)/2-1/10$,

$$E'\{\sup_{0\leqq l\leqq N} |T_{61}(l')-T_{61}(l'')|\} = o_p(n^{9/5}). \tag{3.49}$$

Let us take $\alpha=13/20$, $\beta=1/5$ and $\gamma=1/10$ for definiteness. As l ranges over $0\leqq l\leqq N$, the total number of values taken by l'' is of order $O(n^{\alpha-\beta-\gamma-1/5}) = O(n^{3/20})$. Therefore if we prove that

$$\sup_{0\leqq l\leqq N} E'\{T_{61}^2(l)\} = O_p(n^{17/5}), \tag{3.50}$$

it will follow that

$$\sup_{0\leqq l\leqq N} |T_{61}(l'')| = O_p\{(n^{3/20}\cdot n^{17/5})^{\frac{1}{2}}\} = o_p(n^{9/5}). \tag{3.51}$$

The left hand side of (3.50) is dominated by a constant multiple of

$$S \equiv \sum_{j<k}\sum \{\sum_i^* I(X_{(k)}-\tfrac{1}{2}\lambda n^{-1/5} \leqq X_{(i)} \leqq X_{(j)}+\tfrac{1}{2}\lambda n^{-1/5})\}^2.$$

The series within parentheses equals zero unless $X_{(k)}-X_{(j)}\leqq n^{-1/5}$, and for large n the series is always dominated by

$$n \sup_{x\in A^{\frac{1}{2}\varepsilon}} \{F_n(x+\lambda n^{-1/5})-F_n(x)\} = O_p(n^{4/5}).$$

Therefore

$$S=O_p(n^{8/5})\sum_{j=1}^{n-1}{}^{(\frac{1}{2}\varepsilon)}\sum_{k=j+1}^{n}{}^{(\frac{1}{2}\varepsilon)}I(X_{(k)}-X_{(j)}\leqq\lambda n^{-1/5})=O_p(n^{17/5}),$$

which proves (3.50).

Combining (3.48), (3.49) and (3.51), we see that

$$\sup_{0\leqq l\leqq N}|T_{61}(l)|=o_p(n^{9/5}). \tag{3.52}$$

Next we examine T_{62}. Set

$$T_{63}(l)\equiv T_{62}(l)-E'\{T_{62}(l)\}$$
$$=\sum_j\{\Delta_j^2-\sigma^2(X_j)\}\sum_i{}^*\{f(X_i)\}^{-2}I(X_j-\tfrac{1}{2}t_l\leqq X_i\leqq X_j+\tfrac{1}{2}t_l),$$

and observe that

$$\sup_{0\leqq l\leqq N}\{E'|T_{63}^2(l)|\}\leqq C\sum_j\{\sum_i{}^*I(X_j-\tfrac{1}{2}\lambda n^{-1/5}\leqq X_i\leqq X_j+\tfrac{1}{2}\lambda n^{-1/5})\}^2$$
$$=O_p(n^{13/5}).$$

Therefore

$$E'\{\sup_{0\leqq l\leqq N}|T_{63}(l)|\}=O_p\{(n^{\alpha-1/5}\cdot n^{13/5})^{\frac{1}{2}}\}=o_p(n^{9/5}) \tag{3.53}$$

when $\alpha<4/5$.

We may write $E'\{T_{62}(l)\}=\sum_{j=4}^{6}T_{6j}(l)$, where

$$T_{64}(l)=\sum_i{}^*\{f(X_i)\}^{-2}\sigma^2(X_i)=O_p(n), \tag{3.54}$$

$$T_{65}(l)=\sum_i{}^*\{f(X_i)\}^{-2}\sum_{j\neq i}[\sigma^2(X_j)I(X_i-\tfrac{1}{2}t_l\leqq X_j\leqq X_i+\tfrac{1}{2}t_l)$$
$$-E\{\sigma^2(X_j)I(X_i-\tfrac{1}{2}t_l\leqq X_j\leqq X_i+\tfrac{1}{2}t_l)|X_i\}]$$

and

$$T_{66}(l)=(n-1)\sum_i{}^*\{f(X_i)\}^{-2}E\{\sigma^2(X_j)I(X_i-\tfrac{1}{2}t_l\leqq X_j\leqq X_i+\tfrac{1}{2}t_l)|X_i\},$$

where $j\neq i$. Let $U(i,j,l)$ denote the random variable within square brackets in the expression for $T_{65}(l)$. Then

$$\sup_{0\leqq l\leqq N}|T_{65}(l)|\leqq C\sum_i{}^*\sup_{0\leqq l\leqq N}|\sum_{j\neq i}U(i,j,l)|$$

and

$$E[\sup_{0\leqq l\leqq N}|\sum_{j\neq i}U(i,j,l)|^2|X_i]\leqq C_1\sum_{l=0}^{N}n\leqq C_2n^{1+\alpha-1/5}.$$

Consequently

$$E\{\sup_{0\leqq l\leqq N}|T_{65}(l)|\}=O\{n\cdot(n^{1+\alpha-1/5})^{\frac{1}{2}}\}=o(n^{9/5}) \tag{3.55}$$

if $\alpha<4/5$. Let M denote the number of X_i's in A. Then

$$E[\sup_{0\leqq l\leqq N}|T_{65}(l)-E\{T_{66}(l)|M\}|^2|M]\leqq\sum_{l=0}^{N}\operatorname{var}\{T_{66}(l)|M\}$$
$$=O_p(n^{\alpha-1/5}\cdot n^{3-1/5})=o_p(n^{18/5}).$$

Furthermore,

$$E\{T_{66}(l)|M\}=(n-1)M\int_A\{f(x)\}^{-1}dx\int_{x-\frac{1}{2}t_l}^{x+\frac{1}{2}t_l}\sigma^2(y)f(y)\,dy/P(X\in A)$$
$$=n^2t_l\int_A\sigma^2(x)\,dx+o_p(n^{9/5}) \tag{3.56}$$

uniformly in l. Combining (3.53)-(3.56) we see that

$$T_{62}(l)=n^2t_l\int_A\sigma^2(x)\,dx+o_p(n^{9/5})$$

uniformly in l, whence by (3.47) and (3.52),

$$T_6=n^2h\int_A\sigma^2(x)\,dx+o_p(n^{9/5}) \tag{3.57}$$

uniformly in $h\in H_n$.

(II.ii) T_7: Write

$$T_{71}(x|h)\equiv n^{-1}\sum_j\{\mu(X_j)-\mu(x)\}\,K\{(x-X_j)/h\}$$

and

$$\tau_{71}(x|h)\equiv E\{T_{71}(x|h)\}$$
$$=\int_{-\frac{1}{2}h}^{\frac{1}{2}h}\{\mu(x+u)-\mu(x)\}\,f(x+u)\,du$$
$$=\tfrac{1}{12}h^3\,\gamma(x)+o(n^{-3/5}) \tag{3.58}$$

uniformly in $x\in A$ and $h\in H_n$. Then

$$T_{71}(x|h)-\tau_{71}(x|h)=\int_{x-\frac{1}{2}h\leqq u\leqq x+\frac{1}{2}h}\{\mu(y)-\mu(x)\}\,d\{F_n(y)-F(y)\}$$
$$=O_p(n^{-7/10}) \tag{3.59}$$

uniformly in $x\in A$ and $h\in H_n$, on integrating by parts. We may now deduce from (3.7) that

$$T_7=n\sum_i{}^*\{f(X_i)\}^{-2}\tau_{71}(X_i|h)\sum_j\Delta_jK\{(X_i-X_j)/h\}$$
$$+O_p(n^2\cdot n^{-7/10}\cdot n^{9/20}). \tag{3.60}$$

Define

$$T_{72}(x|h)=\sum_j\Delta_jK\{(x-X_j)/h\},$$

and note that if $0<h_1\leqq h_2$,

$$T_{72}(x|h_2)-T_{72}(x|h_1)=\sum_j\Delta_{(j)}\{I(x-\tfrac{1}{2}h_2\leqq X_{(j)}<x-\tfrac{1}{2}h_1)+I(x+\tfrac{1}{2}h_1<X_{(j)}$$
$$\leqq x+\tfrac{1}{2}h_2)\}.$$

The argument leading to (3.7) may now be used to prove that if $2/5<\alpha<1$,

$$\sup_{x\in A;\, h_1,h_2\in H_n \text{s.t.} |h_1-h_2|\leqq n^{-\alpha}} |T_{72}(x|h_1)-T_{72}(x|h_2)| = O_p\{(n\cdot n^{1-\alpha})^{1/4}\}=o_p(n^{2/5}). \quad (3.61)$$

Therefore if we divide H_n into intervals $[t_{i-1},t_i)$ of length $n^{-\alpha}$, and define $k(h)$ by $h\in[t_k,t_{k+1})$, we see from (3.58) and (3.60) that

$$T_7=n\sum_i{}^* \{f(X_i)\}^{-2}\tau_{71}(X_i|h)\sum_j \Delta_j K\{(X_i-X_j)/t_k\}+o_p(n^{9/5}) \quad (3.62)$$

uniformly in $h\in H_n$. The argument leading to (3.58) may be modified to prove that

$$|\tau_{71}(x|h_1)-\tau_{71}(x|h_2)\leqq Cn^{-2/5}|h_1-h_2|$$

uniformly in $h_1,h_2\in H_n$. It now follows from (3.7) and (3.62) that

$$\begin{aligned} T_7&=T_{73}(k(h))+O_p(n^2\cdot n^{-2/5-\alpha}\cdot n^{9/20})+o_p(n^{9/5})\\ &=T_{73}(k(h))+o_p(n^{9/5}) \end{aligned} \quad (3.63)$$

uniformly in $h\in H_n$, where

$$T_{73}(l)=n\sum_i{}^* \{f(X_i)\}^{-2}\tau_{71}(X_i|t_l)\sum_j{}' \Delta_j K\{(X_i-X_j)/t_l\}.$$

Now,

$$\begin{aligned} n^{-4}E'\{T_{73}^4(l)\}&\leqq C_1([\sum_j{}' \sigma^2(X_j)\{\sum_i{}^* (f(X_i))^{-2}\tau_{71}(X_i|t_l)I(|X_i-X_j|\leqq\tfrac{1}{2}t_l)\}^2]^2\\ &\quad+\sum_j{}' E'(\Delta_j^4)|\sum_i{}^* \{f(X_i)\}^{-2}\tau_{71}(X_i|t_l)I(|X_i-X_j|\leqq\tfrac{1}{2}t_l)|^4)\\ &\leqq C_2\{n\cdot(n^{4/5}\cdot n^{-3/5})^2\}^2=C_2n^{14/5}, \end{aligned}$$

and so

$$E'\{\sup_{0\leqq l\leqq N}|T_{73}(l)|\}=O_p\{(n^{\alpha-1/5}\cdot n^{34}/^5)^{1/4}\}=o_p(n^{9/5}), \quad (3.64)$$

provided $2/5<\alpha<3/5$. Combining (3.63) and (3.64) we see that

$$\sup_{h\in H_n}|T_7|=o_p(n^{9/5}). \quad (3.65)$$

(II.iii) T_8: The results (3.58) and (3.59) imply that

$$T_{71}(x|h)=\tfrac{1}{12}nh^3\gamma(x)+o_p(n^{2/5})$$

uniformly in $x\in A$ and $h\in H_n$. Therefore

$$\begin{aligned} T_8&=\tfrac{1}{144}n^2h^6\sum_i{}^* \{f(X_i)\}^{-2}\gamma^2(X_i)+o_p(n^{9/5})\\ &=\tfrac{1}{144}n^3h^6\int_A \{f(x)\}^{-1}\gamma^2(x)\,dx+o_p(n^{9/5}) \end{aligned} \quad (3.66)$$

uniformly in $h\in H_n$.

Combining the estimates (3.44), (3.57), (3.65) and (3.66), we see that

$$n^{-1}\sum_i^* \{\mu(X_i)-\hat{\mu}_n(X_i)\}^2 = (nh)^{-1}\int_A \sigma^2(x)\,dx+\tfrac{1}{144}h^4\int_A \{f(x)\}^{-1}\gamma^2(x)\,dx+o_p(n^{-4/5}) \quad (3.67)$$

uniformly in $h\in H_n$. This estimate, (3.1), (3.41) and 3.43) imply that

$$n^{-1}\sum_i^* \{Y_i-\hat{\mu}_{ni}(X_i)\}^2 = n^{-1}\sum_i^* \{Y_i-\mu(X_i)\}^2 + (nh)^{-1}\int_A \sigma^2(x)\,dx+\tfrac{1}{144}h^4\int_A \{f(x)\}^{-1}\gamma^2(x)\,dx+o_p(n^{-4/5})$$

uniformly in $h\in H_n$. Theorem 1 follows from (3.67) and (3.68), and Theorem 2 from (3.67).

References

1. Burkholder, D.L.: Distribution function inequalities for martingales. Ann. Probability **1**, 19–42
2. Chow, Y.S., Geman, S., Wu, L.-D.: Consistent cross-validated density estimation. Ann. Statist **11**, 25–38 (1983)
3. Collomb, G.: Estimation non-parametrique de la régression par la méthode du noyau. Thesis, Université Paul Sabatier, Toulouse (1976)
4. Collomb, G.: Quelques proprietes de la méthode du noyau pour l'estimation non-paramétrique de la régression en un point fixé. C. R. Acad. Sci. Paris **285A**, 289–292 (1977)
5. Collomb, G.: Estimation non-paramétrique de la régression: revue bibliographique. Internat. Statist. Rev. **49**, 75–93 (1981)
6. Craven, P., Wahba, G.: Smoothing noisy data with spline functions. Numer. Math. **31**, 377–403 (1979)
7. Csörgő, M., Révész, P.: Strong Approximations in Probability and Statistics. New York and Budapest: Academic Press 1981
8. Devroye, L.P.: The uniform convergence of nearest neighbour regression function estimators and their application in optimization. IEEE Trans. Information Theory **IT-24**, 142–151 (1978)
9. Devroye, L.P.: The uniform convergence of the Nadaraya-Watson regression function estimate. Canad. J. Statist. **6**, 179–191 (1979)
10. Devroye, L.P., Wagner, T.J.: Distribution free consistency results in nonparametric discrimination and regression function estimation. Ann. Statist. **8**, 231–239 (1980)
11. Devroye, L.P., Wagner, T.J.: On the L_1 convergence of kernel estimators of regression functions with applications in discrimination. Z. Wahrscheinlichkeitstheorie verw. Geb. **51**, 15–25 (1980)
12. Diaz, F.U.: Sur le choix du paramètre d'ajustement dans le lissage par functions spline. Numer. Math. **34**, 15–28 (1980)
13. Golub, G.H., Heath, M., Wahba, G.: Generalized cross-validation as a method for choosing a good ridge parameter. Technometrics **21**, 215–223 (1979)
14. Härdle, W., Marron, J.S.: Optimal bandwidth selection in nonparametric regression function estimation.
15. Komlós, J., Major, P., Tusnády, G.: An approximation of partial sums of independent random variables, and the sample distribution function. I. Z. Wahrscheinlichkeitstheorie verw. Geb. **32**, 111–131 (1975)
16. Loève, M.: Probability Theory. Princeton, N.J.: Van Nostrand 1963
17. Mack, Y.P., Silverman, B.W.: Weak and strong uniform consistency of kernel regression estimates. Z. Wahrscheinlichkeitstheorie verw. Geb. **61**, 405–415 (1982)
18. Nadaraya, E.A.: On estimating regression. Theor. Probability Appl. **9**, 141–142 (1964)

19. Révész, P.: On the nonparametric estimation of the regression function. Problems of Control and Information Theory **8**, 297–302
20. Schuster, E.F., Gregory, C.G.: On the nonconsistency of maximum likelihood nonparametric density estimators. In: Computer Science and Statistics: Proceedings of the 13th Symposium on the Interface. Ed. W.F. Eddy, 295–298. New York: Springer 1981
21. Spiegelman, C., Sacks, J.: Consistent window estimation in nonparametric regression. Ann. Statist. **8**, 240–246
22. Stone, C.J.: Consistent nonparametric regression. Ann. Statist. **5**, 595–620 (1977)
23. Stone, C.J.: Optimal global rates of convergence for nonparametric regression. Ann. Statist. **10**, 1040–1053 (1982)
24. Tusnády, G.: A remark on the approximation of the sample distribution function in the multidimensional case. Period. Math. Hungar. **8**, 53–55 (1977)
25. Wahba, G., Wold, S.: A completely automatic French curve: fitting spline functions by cross-validation. Comm. Statist. **4**, 1–17 (1975)
26. Watson, G.S.: Smooth regression analysis. Sankhyā **A26**, 359–372 (1964)

Received January 13, 1983; in final form April 6, 1984