The Annals of Statistics
1984, Vol. 12, No. 1, 241–260

INTEGRATED SQUARE ERROR PROPERTIES OF KERNEL ESTIMATORS OF REGRESSION FUNCTIONS

BY PETER HALL

Australian National University

Weak laws of large numbers and central limit theorems are proved for integrated square error of kernel estimators of regression functions. The regressor is assumed to take values in $\mathbb{R}^p$, and the regressand, X, to be real valued. It is shown that in many cases, integrated square error is asymptotically normally distributed and independent of the X-sample. As an application, a test for the regression function (such as that proposed by Konakov) is seen to be asymptotically independent of an arbitrary test based on the X-sample. The proofs involve martingale methods.

1. Introduction. Kernel estimators of a regression function were introduced by Nadaraya (1964) and Watson (1964). They take the following form when the regressand is a multivariate random variable. Suppose (X_j, Y_j), $1 \le j \le n$, is a random sample from the $(p+1)$-variate distribution of (X, Y), where X is a p-vector and Y is a scalar. Let K be a density function on $\mathbb{R}^p$, and let h be a small positive constant. Then

$$\hat{\mu}_n(x) = \hat{\mu}_n(x \mid h) \equiv [\textstyle\sum_{i=1}^n Y_i K\{(x - X_i)/h\}]/[\sum_{i=1}^n K\{(x - X_i)/h\}]$$

is a nonparametric estimator of $\mu(x) \equiv E(Y \mid X = x)$. (The ratio 0/0 is interpreted as unity.)

This definition is obviously closely related to that of a nonparametric density estimator. Indeed, the denominator in our expression for $\hat{\mu}_n(x)$ is proportional to

$$\hat{f}_n(x) = (nh^p)^{-1} \textstyle\sum_{i=1}^n K\{(x - X_i)/h\}, \tag{1.1}$$

which is a nonparametric estimator of the marginal density of X at $x \in \mathbb{R}^p$. The similarity between these two concepts suggests that we might transfer techniques from the better-known field of nonparametric density estimation, to regression function estimation. In particular, recalling that mean integrated square error is the most commonly accepted measure of the performance of a density estimator, we might decide to construct a regression estimator on a set $A \subseteq \mathbb{R}^p$ so as to minimise

$$\int_A E\{\hat{\mu}_n(x) - \mu(x)\}^2 \, dx. \tag{1.2}$$

However, at this point we immediately encounter difficulties. The estimator $\hat{\mu}_n(x)$ is defined as the *ratio* of two random variables, and so computation of the

Received January 1983; revised September 1983.
AMS 1980 *subject classifications*. Primary 62G20, 62E20; secondary 60F05.
Key words and phrases. Central limit theorem, integrated square error, kernel estimator, law of large numbers, multivariate, nonparametric, regression, test for regression.

expectation in (1.2) is not a straightforward matter. Collomb (1977) has described an asymptotic formula for $E\{\hat{\mu}_n(x) - \mu(x)\}^2$ at a fixed point x, although under somewhat restrictive conditions on the "window size", h. He has remarked (Collomb, 1981) that similar results could be obtained, under appropriate conditions, for weighted mean integrated square error. However, it is known that $\hat{\mu}_n$ is consistent for μ under the minimal conditions, $h = h(n) \to 0$ and $nh^p \to \infty$, which are necessary for the consistency of $\hat{f}_n$; see for example Devroye and Wagner (1979). Therefore it is very desirable to describe integrated square error under such basic constraints. Furthermore, a sufficiently detailed description of integrated square error could be used to construct hypothesis tests for the unknown regression function, μ; see Nadaraya (1974) and Konakov (1977).

Our aim in the present paper is to prove laws of large numbers and central limit theorems for a very general version of weighted integrated square error,

$$I_n = \int_A \{\hat{\mu}(x) - \mu(x)\}^2 w_n(x)\, dx,$$

where $w_n(x)$ is a weight function which may be *either random or deterministic.* Among the conclusions of our study are the following points.

(i) Laws of large numbers for I_n, of the form $I_n/c_n \to 1$ in probability for constants c_n converging to zero, hold under minimal conditions on $h(n)$. We suggest that these results could play the role of mean integrated square error for regression function estimators. The constant c_n breaks up cleanly into two parts, one corresponding to the variance of $\hat{\mu}_n$ and the other to the squared bias. The rate of convergence of integrated square error is maximised when these two terms are of the same order of magnitude.

(ii) Central limit theorems for I_n hold under very general conditions. Konakov (1977) used approximations to empirical processes to derive the first central limit theorems of this type, but his techniques forced him to confine attention to the case $p = 1$ and to impose very restrictive conditions on the window size $h(n)$. Indeed, the conditions require that the estimator be constructed suboptimally with respect to mean square error. In this paper we use different methods, including martingale theory, to substantially generalise Konakov's results.

(iii) For the most part, central limit theorems proved for integrated square error are valid *conditional on the X-sample,* as well as unconditionally. The centering constants are not affected by conditioning. This means that hypothesis tests for $\mu(x)$ which are based on integrated square error, such as that proposed by Nadaraya (1974) and Konakov (1977), are asymptotically independent of arbitrary hypothesis tests based on the X-sample. Under various conditions on h and n, "Studentized" versions of centered integrated square error could be constructed, and used to test hypotheses about the regression function μ. Assessment of the practicality of such a procedure would probably rely heavily on simulation studies. See Liero (1982) for an example of "Studentization" in the context of nonparametric regression.

Our proofs of the results summarized above are founded on a single basic lemma, which is stated as Theorem 1 in the next section. All the proofs are relegated to Section 3.

The history of integrated square error theory for density estimators is longer and more detailed than that for nonparametric regression function estimators. Bickel and Rosenblatt (1973), Rosenblatt (1975), Csörgö and Révész (1981, Chapter 6) and Hall (1981) have proved central limit theorems for integrated square error of nonparametric density estimators. However, these results are of a very different nature from their counterparts for regression function estimators, since in the latter case the dominant source of variability in limit theorems is the Y-sample, not the X-sample. (See point (iii) above.) Indeed, limit theory in the regression case is perhaps closer to that for sums of weighted independent random variables, than to limit theory for density estimators.

All our results describe L^2 properties of the error in nonparametric regression. By way of comparison, we mention the recent L^∞ results of Johnston (1982) and Liero (1982).

2. Results. Throughout this section we impose the following conditions. Let $A \subseteq \mathbb{R}^p$ be a rectangle (not necessarily bounded; $A = \mathbb{R}^p$ is permissable), and let A^ε be the set of all points in $\mathbb{R}^p$ which are distant less than ε from A, where ε is an arbitrary positive number. We take K to be a bounded density function on $\mathbb{R}^p$ with compact support and satisfying

$$\int z^{(i)}K(z)\,dz = 0, \quad \int z^{(i)}z^{(j)}K(z)\,dz = k\delta_{ij}$$

for $1 \le i, j \le p$, where k does not depend on i and δ_{ij} is the Kronecker delta. Assume that X has a uniformly continuous density, f, on A^ε, whose first derivatives exist, are bounded and uniformly continuous on A^ε; that the conditional moments

$$\mu(x) \equiv E(Y \mid X = x), \quad \sigma^2(x) \equiv E(Y^2 \mid X = x) - \mu^2(x),$$
$$\mu_4(x) = E[\{Y - \mu(X)\}^4 \mid X = x]$$

are well defined for $x \in A^\varepsilon$; that μ is uniformly continuous on A^ε; that the first and second derivatives of μ exist, are bounded and uniformly continuous on A^ε; that $\sigma^2(x)$ is bounded on A^ε; and that

$$\int_{A^\varepsilon} \mu_4(x)f(x)\,dx < \infty. \tag{2.1}$$

To simplify our notation we write $w_n(x) = \hat{f}_n^2(x)v_n(x)$, where v_n is a random function measurable in $\mathscr{F}\{X_1, \ldots, X_n\}$ (the σ-field generated by $X_1, \ldots, X_n$) and defined on $x \in A^\varepsilon$, and where $\hat{f}_n$ is defined by (1.1). We suppose that there exists a bounded, nonnegative, deterministic function v such that

$$\sup\nolimits_{x\in A^\varepsilon} | v_n(x)/v(x) - 1 | \to 0 \tag{2.2}$$

in probability as $n \to \infty$.

Note that if A is a bounded rectangle then none of our conditions, in particular the constraint (2.1), require the assumption that Y have a finite moment. This is made possible by our restriction that K have compact support, and it seems unlikely that the latter condition can be relaxed without more stringent assumptions about the tail behaviour of Y. We do not require that μ be bounded.

Our first theorem is used in the proofs of all subsequent results in this section. Define

$$\gamma(x) = k \sum_{j=1}^{p} [\{\partial f(x)/\partial x^{(j)}\}\{\partial\mu(x)/\partial x^{(j)}\} + \tfrac{1}{2} f(x)\{\partial^2\mu(x)/\partial x^{(j)}\partial x^{(j)}\}],$$

$$\alpha_1 = 2\left[\int_A \sigma^4(x)f^2(x)v^2(x)\,dx\right]\left[\int \left\{\int K(x)K(x+y)\,dx\right\}^2 dy\right]$$

and

$$\alpha_2 \equiv 4\int_A \sigma^2(x)\gamma^2(x)v^2(x)\,dx.$$

Here and below, an unqualified integral is taken over all of $\mathbb{R}^p$. To avoid trivialities we assume that $\alpha_1\alpha_2 > 0$.

THEOREM 1. *Under the above conditions, and if $h \to 0$ and $nh^p \to \infty$ as $n \to \infty$,*

$$\int_A \{\hat\mu_n(x) - \mu(x)\}^2\hat f_n^2(x)v_n(x)\,dx$$

$$= (nh^p)^{-2}\sum_{i=1}^{n}\sigma^2(X_i)\int_A K^2\{(x-X_i)/h\}v_n(x)\,dx \tag{2.3}$$

$$+ (nh^p)^{-2}\int_A \left[\sum_{i=1}^{n}\{\mu(X_i)-\mu(x)\}K\{(x-X_i)/h\}\right]^2 v_n(x)\,dx$$

$$+ (n^{-2}h^{-p}\alpha_1 + n^{-1}h^4\alpha_2)^{1/2}N_n,$$

where N_n is asymptotically normal $N(0, 1)$ and asymptotically independent of $X_1, \ldots, X_n$.

The statement that N_n is asymptotically independent of $X_1, \ldots, X_n$ means that for any sequence of events $\{A_n\}$ for which A_n is measurable in $\mathscr{F}\{X_1, \ldots, X_n\}$ for each n, we have

$$\sup_{-\infty<x<\infty} |\,P(A_n;\, N_n \le x) - P(A_n)\Phi(x)\,| \to 0$$

as $n \to \infty$, Φ being the standard normal distribution function. It is easily seen from the proof that the variable N_n has conditional mean (given $X_1, \ldots, X_n$) exactly equal to zero, and so the sum of the first two terms on the right hand side of (2.3) equals the conditional mean of integrated square error.

Our next result describes a weak law of large numbers for integrated square error.

THEOREM 2. *Assume the conditions of Theorem* 1, *and in addition that* v *is uniformly continuous on* A^ε *and satisfies*

$$\int_A v(x)\,dx < \infty. \tag{2.4}$$

Then

$$\begin{aligned}\int_A \{\hat{\mu}_n(x) - \mu(x)\}^2 \hat{f}_n^2(x) v_n(x)\,dx \\ = (nh^p)^{-1} \int_A \sigma^2(x) f(x) v(x)\,dx \int K^2(u)\,du + h^4 \int_A \gamma^2(x) v(x)\,dx \\ + o_p\{(nh^p)^{-1} + h^4\}\end{aligned} \tag{2.5}$$

as $n \to \infty$.

The first term on the right hand side in (2.5) is derived from the error about the mean of $\hat{\mu}$, while the second term corresponds to the bias. The minimum order of magnitude of the sum of these terms is $n^{-4/(p+4)}$, and is achieved when $h \approx$ const. $n^{-1/(p+4)}$ (cf. Stone, 1982). Note too that $n^{-1}h^4 >> n^{-2}h^{-p}$ if and only if $h >> n^{-1/(p+4)}$. (A term of order $(n^{-1}h^4 + n^{-2}h^{-p})^{1/2}$ appears on the right hand side in (2.3).)

Let us choose A to be a bounded rectangle such that $f(x)$ is bounded away from zero on A^ε, and select $v_n(x) = w(x)/\hat{f}_n^2(x)$, where w is an arbitrary bounded, continuous, nonnegative weight function. Condition (2.2) is satisfied if we take $v = w/f^2$, and we may deduce from Theorem 2 that

$$\begin{aligned}\int_A \{\hat{\mu}_n(x) - \mu(x)\}^2 w(x)\,dx \\ = (nh^p)^{-1} \int_A \sigma^2(x)\{f(x)\}^{-1} w(x)\,dx \cdot \int K^2(u)\,du \\ + h^4 \int_A \gamma^2(x)\{f(x)\}^{-2} w(x)\,dx + o_p\{(nh^p)^{-1} + h^4\}.\end{aligned} \tag{2.6}$$

The sum of the first two terms on the right in (2.6) is the same as would be obtained by formally integrating the asymptotic formula for $E\{\hat{\mu}_n(x) - \mu(x)\}^2 w(x)$, given in Collomb (1977) for a fixed point x, over the set A.

Konakov (1977) used empirical process methods to prove central limit theorems for the weighted integrated square error given in (2.5), in the special case where $v_n(x) \equiv v(x)$ for all n. He confined attention to the case $p = 1$, and considered only values of h satisfying $nh^2 \to \infty$ and $nh^{5/2} \to 0$. The methods used in this paper permit us to remove these restrictions, and to prove in addition that the normal limit is asymptotically independent of the X-process.

THEOREM 3. *Assume the conditions of Theorem* 1, *and take* $v_n(x) \equiv v(x)$ *for* *all x, where v is uniformly continuous on A^ε and satisfies* (2.4). *Then*

$$(n^{-2}h^{-p}\alpha_1 + n^{-1}h^4\alpha_2)^{-1/2}\Big(\int_A \{\hat{\mu}_n(x) - \mu(x)\}^2\hat{f}_n^2(x)v(x)\,dx \tag{2.7}$$
$$-E\Big[\int_A \{\hat{\mu}_n(x) - \mu(x)\}^2\hat{f}_n^2(x)v(x)\,dx\Big]\Big) \to N(0, 1)$$

in distribution as $n \to \infty$, and the variable on the left hand side is asymptotically independent of $X_1, \ldots, X_n$.

Our last result provides a central limit theorem for integrated square error when the weight function w is purely deterministic. More general results may be proved under weaker conditions on h and p. It is necessary only to use an arbitrarily long Taylor expansion of $\{\hat{f}_n(x)\}^{-2} = \{E\hat{f}_n(x) + \hat{f}_n(x) - E\hat{f}_n(x)\}^{-2}$, extending (3.28) below.

THEOREM 4. *Assume the conditions of Theorem* 1, *and take A to be a bounded rectangle such that $f(x)$ is bounded away from zero on A^ε. Set $v_n(x) = w(x)/\hat{f}_n^2(x)$, where w is a bounded, continuous, nonnegative weight function. If in addition $1 \le p \le 3$ and $n^2h^{3p} \to \infty$ as $n \to \infty$,*

$$(n^{-2}h^{-p}\alpha_1 + n^{-1}h^4\alpha_2)^{-1/2}\Big(\int_A \{\hat{\mu}_n(x) - \mu(x)\}^2 w(x)\,dx$$
$$-E\Big[\int_A \{\hat{\mu}_n(x) - \mu(x)\}^2\hat{f}_n^2(x)\{E\hat{f}_n(x)\}^{-2}w(x)\,dx\Big]\Big)$$
$$\to N(0, 1)$$

in distribution as $n \to \infty$, and the variable on the left hand side is asymptotically independent of $X_1, \ldots, X_n$.

Note that the conditions on h and p imposed in Theorem 4 include the "optimal" h, $h \approx n^{-1/(p+4)}$. The constraint $1 \le p \le 3$ is a "technical necessity", and is not directly connected with the problem.

3. Proofs. We let E' denote expectation conditional on $X_1, \ldots, X_n$, and $C, C_1, C_2, \ldots$ denote generic positive constants.

PROOF OF THEROEM 1. Observe that

$$\mu_n(x) - \mu(x) = (nh^p)^{-1}\textstyle\sum_{i=1}^n \{Y_i - \mu(X_i)\}K\{(x - X_i)/h\}/\hat{f}_n(x) + g_n(x)/\hat{f}_n(x),$$

where

$$g_n(x) = (nh^p)^{-1}\textstyle\sum_{i=1}^n \{\mu(X_i) - \mu(x)\}K\{(x - X_i)/h\}.$$

Therefore

$$(3.1)\qquad \int_A [\{\hat{\mu}_n - \mu(x)\}^2 \hat{f}_n^2(x) v_n(x)\, dx = I_{n1} + 2I_{n2} + 2I_{n3} + I_{n4},$$

where

$$I_{n1} = (nh^p)^{-2} \sum_{i=1}^n \{Y_i - \mu(X_i)\}^2 \int_A K^2\{(x - X_i)/h\} v_n(x)\, dx,$$

$$I_{n2} = (nh^p)^{-2} \sum_{1 \le i < j \le n}\sum \{Y_i - \mu(X_i)\}\{Y_j - \mu(X_j)\}$$

$$\times \int_A K\{(x - X_i)/h\} K\{(x - X_j)/h\} v_n(x)\, dx,$$

$$I_{n3} = (nh^p)^{-1} \sum_{i=1}^n \{Y_i - \mu(X_i)\} \int_A K\{(x - X_i)/h\} g_n(x) v_n(x)\, dx$$

and

$$I_{n4} = \int_A g_n^2(x) v_n(x)\, dx.$$

We shall handle these terms one by one. Our major task is to separate out behaviour which depends on the X's and Y's together, from behaviour which depends solely on the X's. This is achieved in steps (i), (ii) and (iii) below.

(i) I_{n1}: Note that

$$(nh^p)^4 E'\{I_{n1} - E'(I_{n1})\}^2$$

$$= \sum_{i=1}^n E'[\{Y_i - \mu(X_i)\}^2 - \sigma^2(X_i)]^2 \left[\int_A K^2\{(x - X_i)/h\} v_n(x)\, dx\right]^2$$

$$= O_p(1) \sum_{i=1}^n \mu_4(X_i) \left[\int_A K^2\{(x - X_i)/h\}\, dx\right]^2.$$

If $K(z)$ vanishes for $|z| > \lambda$ then the integral within square brackets is zero unless $X_i \in A^{\lambda h}$. In any event,

$$\int_A K^2\{(x - X_i)/h\}\, dx \le \int K^2\{(x - X_i)/h\}\, dx = h^p \int K^2(x)\, dx,$$

and so

$$(nh^p)^4 E'\{I_{n1} - E'(I_{n1})\}^2 = O_p(h^{2p}) \sum_{X_i \in A^\epsilon} \mu_4(X_i) = O_p(nh^{2p}),$$

the last equality following from (2.1) and the weak law of large numbers. It now follows via Chebyshev's inequality that if $\{\lambda_n, n \ge 1\}$ is any sequence of constants diverging to $+\infty$,

$$P\{|I_{n1} - E'(I_{n1})| > \lambda_n n^{-3/2}h^{-p} | X_1, \ldots, X_n\} \to 0$$

in probability as $n \to \infty$. The unconditional probability must also converge to zero, by dominated convergence. Therefore

$$\begin{aligned} I_{n1} &= E'(I_{n1}) + O_p(n^{-3/2}h^{-p}) \\ &= (nh^p)^{-2} \textstyle\sum_{i=1}^n \sigma^2(X_i) \displaystyle\int_A K^2\{(x - X_i)/h\}v_n(x)\, dx + O_p(n^{-3/2}h^{-p}) \end{aligned} \tag{3.2}$$

as $n \to \infty$.

(ii) I_{n2}: Let

$$W_{nij} = \int_A K\{(x - X_i)/h\}K\{(x - X_j)/h\}v_n(x)\, dx,$$

and define $\tilde{W}_{nij}$ in the same way but with v_n replaced by v. Observe that

$$(nh^p)^2 I_{n2} = \sum\sum_{1 \le i < j \le n} \{Y_i - \mu(X_i)\}\{Y_j - \mu(X_j)\}W_{nij} = \textstyle\sum_{i=2}^n Y_{ni}, \tag{3.3}$$

where

$$Y_{ni} = \{Y_i - \mu(X_i)\} \textstyle\sum_{j=1}^{i-1} \{Y_j - \mu(X_j)\}W_{nij}, \quad 2 \le i \le n.$$

If $\mathscr{F}_{ni}$ denotes the σ-field generated by $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_i$ for $0 \le i \le n$ (so that $\mathscr{F}_{n0} = \mathscr{F}\{X_1, \ldots, X_n\}$), then $E(Y_{ni} | \mathscr{F}_{n,i-1}) = 0$ almost surely for all i. Therefore the sequence

$$\{(S_{ni} = \textstyle\sum_{j=2}^i Y_{nj}, \mathscr{F}_{ni}), \quad 2 \le i \le n < \infty\}$$

is a martingale triangular array. (See Hall and Heyde, 1980, pages 52–53.) The conditional variance of S_{nn} is given by

$$\begin{aligned} V_n^2 &= \textstyle\sum_{i=2}^n E(Y_{ni}^2 | \mathscr{F}_{n,i-1}) = \sum_{i=2}^n \sigma^2(X_i)[\sum_{j=1}^{i-1} \{Y_j - \mu(X_j)\}W_{nij}]^2 \\ &= \textstyle\sum_{i=2}^n \sigma^2(X_i) \sum_{j=1}^{i-1} \{Y_j - \mu(X_j)\}^2 W_{nij}^2 \\ &\quad + 2 \textstyle\sum_{i=2}^n \sigma^2(X_i) \displaystyle\sum\sum_{1 \le j < k \le i-1} \{Y_j - \mu(X_j)\}\{Y_k - \mu(X_k)\}W_{nij}W_{nik} \\ &= V_{n1} + V_{n2}, \end{aligned} \tag{3.4}$$

say.

Our aim is to prove a central limit theorem for S_{nn}. This entails checking two conditions, which is done in Lemmas 1 and 2 below.

LEMMA 1.

$$\begin{aligned} n^{-2}h^{-3p}V_{n1} &\to \frac{1}{4}\alpha_1 \\ &\equiv \frac{1}{2}\left[\int_A \sigma^4(x)f^2(x)v^2(x)\, dx\right]\left[\int \left\{\int K(u)K(u+v)\, du\right\}^2 dv\right] \end{aligned} \tag{3.5}$$

and

$$n^{-2}h^{-3p}V_{n2} \to 0 \tag{3.6}$$

in probability as $n \to \infty$.

PROOF. Since $V_{n1} = \{1 + o_p(1)\}\tilde{V}_{n1}$, where

$$\tilde{V}_{n1} = \sum_{i=2}^{n} \sigma^2(X_i) \sum_{j=1}^{i-1} \{Y_j - \mu(X_j)\}^2 \tilde{W}_{nij}^2,$$

then it suffices to prove (3.5) with $\tilde{V}_{n1}$ replacing V_{n1}. Now,

$$\begin{aligned} E'\{\tilde{V}_{n1} - E'(\tilde{V}_{n1})\}^2 &= \sum_{j=1}^{n-1} E'[\{Y_j - \mu(X_j)\}^2 - \sigma^2(X_j)]^2 \\ &\quad \times \{\textstyle\sum_{i=j+1}^{n} \sigma^2(X_i)\tilde{W}_{nij}^2\}^2 \\ &\le C_1 \sum_{j=1}^{n-1} \mu_4(X_j)\{\textstyle\sum_{i=j+1}^{n} \sigma^2(X_i)\tilde{W}_{nij}^2\}^2 \\ &\le C_2 \sum_{j=1}^{n-1} \mu_4(X_j)(\textstyle\sum_{i=j+1}^{n} \tilde{W}_{nij}^2)^2, \end{aligned} \tag{3.7}$$

using the fact that $\sigma^2(x)$ is bounded on A^ϵ, and $\tilde{W}_{nij} = 0$ unless both X_i and X_j are inside $A^{\lambda h}$, where λ is such that $K(z)$ vanishes for $|z| > \lambda$. Furthermore,

$$\tilde{W}_{nij} \le C_3 \int_A K\{(x - X_i)/h\}\, dx \le C_3 h^p, \tag{3.8}$$

and so by (3.7),

$$E'\{\tilde{V}_{n1} - E'(\tilde{V}_{n1})\}^2 \le C_4 n h^{3p} \sum_{\substack{j \le n-1,\\ X_j \in A^\epsilon}} \mu_4(X_i) \sum_{\substack{i \ge j+1,\\ X_i \in A^\epsilon}} \tilde{W}_{nij}.$$

But the mean of the double series on the right is dominated by a constant multiple of

$$\begin{aligned} n^2 &\int_{A^\epsilon} \mu_4(x) f(x)\, dx \int_{A^\epsilon} f(y)\, dy \int_A K\{(z - x)/h\} K\{(z - y)/h\}\, dz \\ &\le C\, n^2 h^{2p} \int_{A^\epsilon} \mu_4(x) f(x)\, dx' \int dv \int K(u) K(u + v)\, du \\ &= O(n^2 h^{2p}), \end{aligned}$$

whence $E'\{\tilde{V}_{n1} - E'(\tilde{V}_{n1})\}^2 = O(n^3 h^{5p})$ and

$$\tilde{V}_{n1} = E'(\tilde{V}_{n1}) + O_p(n^{3/2} h^{5p/2}). \tag{3.9}$$

(Note that for positive random variables Z_n, $E(Z_n) = O(\delta_n)$ implies $Z_n = O_p(\delta_n)$.)

Let $A_n = E'(\tilde{V}_{n1})$ and $a_n = E(A_n)$, and observe that

$$
\begin{aligned}
a_n &= \frac{1}{2} n(n-1)E\{\sigma^2(X_1)\sigma^2(X_2)\tilde{W}_{n12}^2\} \\
&= \frac{1}{2} n(n-1)h^{3p} \int\int \sigma^2(x)\sigma^2(x-wh)f(x)f(x-wh)\,dx\,dw \\
&\quad \times \left\{\int_{(A-x)/h} K(z)K(z+w)v(x-zh)\,dz\right\}^2 \\
&\sim \frac{1}{4} n^2h^{3p}\alpha_1.
\end{aligned}
\tag{3.10}
$$

We may write $A_n - a_n$ as a U-statistic,

$$U_n \equiv A_n - a_n = \sum_{1\le i<j\le n}\sum H_n(X_i, X_j),$$

where $H_n(X_i, X_j) = \sigma^2(X_i)\sigma^2(X_j)\tilde{W}_{nij}^2 - E\{\sigma^2(X_1)\sigma^2(X_2)\tilde{W}_{n12}^2\}$. Let $G_n(x) = E\{H_n(X_1, x)\}$. The U-statistic's projection is given by

$$S_n = (n-1)\textstyle\sum_{i=1}^n G_n(X_i),$$

and

$$
\begin{aligned}
&E(S_n^2) + E(U_n - S_n)^2 \\
&\qquad \le n^3E\{G_n^2(X_1)\} + n^2E\{H_n(X_1, X_2) - G_n(X_1) - G_n(X_2)\}^2 \\
&\qquad \le C_1n^3E\{\sigma^4(X_1)\sigma^4(X_2)\tilde{W}_{n12}^4\} \\
&\qquad \le C_2n^3E(\tilde{W}_{n12}^4).
\end{aligned}
$$

Now,

$$
\begin{aligned}
E(\tilde{W}_{n12}^4) &= h^{5p}\int\int f(x)f(x-wh)\,dx\,dw \\
&\qquad \times \left[\int_{(A-x)/h} K(z)K(z+w)v(x-zh)\,dz\right]^4 \\
&= O(h^{5p}),
\end{aligned}
$$

and so $E(A_n - a_n)^4 = O(n^3h^{5p})$, whence $A_n = a_n + O_p(n^{3/2}h^{5p/2})$. Since $nh^p \to \infty$ as $n \to \infty$, we may now deduce from (3.9) and (3.10) that $n^{-2}h^{-3p}\tilde{V}_{n1} \to \frac{1}{4}\alpha_1$ in probability, from which follows (3.5).

To prove (3.6), observe that

$$
\begin{aligned}
E'(V_{n2}^2) &= 4 \sum_{1\le j<k\le n-1}\sum \sigma^2(X_j)\sigma^2(X_k)\{\textstyle\sum_{i=k+1}^n \sigma^2(X_i)W_{nij}W_{nik}\}^2 \\
&= O_p(1) \sum_{1\le j<k\le n-1}\sum (\textstyle\sum_{i=k+1}^n \tilde{W}_{nij}\tilde{W}_{nik})^2.
\end{aligned}
\tag{3.11}
$$

If $1 \le j < k \le n - 1$ then

$$E\Big(\sum_{i=k+1}^{n} \tilde W_{nij} \tilde W_{nik}\Big)^2 = (n-k)E(\tilde W_{n12}^2 \tilde W_{n13}) + (n-k)(n-k-1)E(\tilde W_{n13}\tilde W_{n14}\tilde W_{n23}\tilde W_{n24}), \tag{3.12}$$

and

$$E(\tilde W_{n12}^2 \tilde W_{n13}^2) = O(h^{4p}), \tag{3.13}$$

using (3.8) and the fact that $E(\tilde W_{n12}) = O(h^{2p})$. Furthermore,

$$\begin{aligned} CE(\tilde W_{n13}\tilde W_{n14}\tilde W_{n23}\tilde W_{n24}) &\le \int_A \int_A \int_A \int_A B_n(u_1, u_2, v_1, v_2)\, du_1\, du_2\, dv_1\, dv_2 \\ &= h^{3p} \int_A \int_{A_1} \int_{A_2} \int_{A_3} \\ &\quad \cdot B_n(u_1, u_1 + a_1 h, u_1 + a_2 h - a_3 h, u_1 - a_3 h)\, du_1\, da_1\, da_2\, da_3, \end{aligned} \tag{3.14}$$

where the A_j are transformed versions of the set A, and

$$\begin{aligned} &B_n(u_1, u_2, v_1, v_2) \\ &\quad = E[K\{(u_1 - X_1)/h\}K\{(u_2 - X_1)/h\}]E[K\{(v_1 - X_1)/h\}K\{(v_2 - X_1)/h\}] \\ &\qquad \times E[K\{(u_1 - X_1)/h\}K\{(v_1 - X_1)/h\}]E[K\{(u_2 - X_1)/h\}K\{(v_2 - X_1)/h\}] \\ &\quad = h^{4p}\left[\int K(w)K\{w + (u_2 - u_1)/h\}f(u_1 - wh)\, dw\right] \\ &\qquad \times \left[\int K(w)K\{w + (v_1 - v_2)/h\}f(v_2 - wh)\, dw\right] \\ &\qquad \times \left[\int K(w)K\{w + (v_1 - u_1)/h\}f(u_1 - wh)\, dw\right] \\ &\qquad \times \left[\int K(w)K\{w + (u_2 - v_2)/h\}f(v_2 - wh)\, dw\right]. \end{aligned}$$

On substituting this formula into (3.14) we may deduce that

$$E(\tilde W_{n13}\tilde W_{n14}\tilde W_{n23}\tilde W_{n24}) = O(h^{7p}). \tag{3.15}$$

It follows from (3.12), (3.13) and (3.15) that

$$E\Big(\sum_{i=k+1}^{n} \tilde W_{nij}\tilde W_{nik}\Big)^2 = O(nh^{5p} + n^2 h^{7p})$$

uniformly in $1 \le j < k \le n - 1$, and so by (3.11) and the fact that $nh^p \to \infty$,

$$E'(n^{-2}h^{-3p}V_{n2})^2 = O_p(n^{-1}h^{-p} + h^p) = o_p(1).$$

The result (3.6) is a consequence of this estimate.

LEMMA 2. *For each $\varepsilon > 0$,*

$$n^{-2}h^{-3p}\sum_{i=2}^{n} E'\{Y_{ni}^2 I(|Y_{ni}| > \varepsilon n h^{3p/2})\} \to 0$$

in probability as $n \to \infty$.

PROOF. Define $Z_{ni} = \sum_{j=1}^{i-1}\{Y_j - \mu(X_j)\}W_{nij}$, so that $Y_{ni} = \{Y_i - \mu(X_i)\}Z_{ni}$. Fix $k > 0$, and observe that $|Y_{ni}| > \varepsilon n h^{3p/2}$ entails either $|Y_i - \mu(X_i)| > k$, or $k|Z_{ni}| > \varepsilon n h^{3p/2}$. Therefore

$$\begin{aligned} E\{Y_{ni}^2 I(|Y_{ni}| > \varepsilon n h^{3p/2}) \mid \mathscr{F}_{n,i-1}\} \\ \le Z_{ni}^2[E'\{|Y_i - \mu(X_i)|^2 I(|Y_i - \mu(X_i)| > k)\} \\ + \sigma^2(X_i) I(|Z_{ni}| > k^{-1}\varepsilon n h^{3p/2})]. \end{aligned}$$

Now, $E'(Z_{ni}^2) = \sum_{j=1}^{i-1}\sigma^2(X_j)W_{nij}^2$. Thus, since $\sigma^2(x)$ is bounded on A^ε, and $W_{nij} = 0$ for large n unless both $X_i, X_j \in A^\varepsilon$,

$$\begin{aligned} E'\{Y_{ni}^2 I(|Y_{ni}| > \varepsilon n h^{3p/2})\} \\ \le C_1\Big(\sum_{j=1}^{i-1} W_{nij}^2\Big)E'\{|Y_i - \mu(X_i)|^2 I(|Y_i - \mu(X_i)| > k)\} \\ + C_1 E'\{Z_{ni}^2 I(|Z_{ni}| > k^{-1}\varepsilon n h^{3p/2})\}. \end{aligned}$$

But

$$\sum_{j=1}^{i-1} W_{nij}^2 = O_p(1)\sum_{j=1}^{i-1}\tilde{W}_{nij}^2 = O_p(h^p)\sum_{j=1}^{i-1}\tilde{W}_{nij}$$

uniformly in i, using (3.8), and so Lemma 2 will follow if we prove that

$$n^{-2}h^{-3p}\sum_{i=2}^{n} E'\{Z_{ni}^2 I(|Z_{ni}| > \varepsilon n h^{3p/2})\} \to 0 \tag{3.16}$$

in probability as $n \to \infty$ for all $\varepsilon > 0$, and

$$\begin{aligned} \limsup_{n\to\infty} n^{-2}h^{-2p} \\ \times E\Big[\sum_{i=2}^{n}\Big(\sum_{j=1}^{i-1}\tilde{W}_{nij}\Big)E\{|Y_i - \mu(X_i)|^2 I(|Y_i - \mu(X_i)| > k) \mid X_i\}\Big] \to 0 \end{aligned} \tag{3.17}$$

as $k \to \infty$.

To prove (3.16), observe that

$$\begin{aligned} E'(Z_{ni}^4) &= \sum_{j=1}^{i-1} E'\{Y_j - \mu(X_j)\}^4 W_{nij}^4 + 6\sum\sum_{1\le j<k\le i-1}\sigma^2(X_i)\sigma^2(X_j)W_{nij}^2W_{nik}^2 \\ &= O_p(h^{3p})\sum_{j=1}^{i-1}\mu_4(X_j)\tilde{W}_{nij} + O_p(nh^{3p})\sum_{j=1}^{i-1}\tilde{W}_{nij} \end{aligned}$$

uniformly in i, and

$$E\Big\{\sum_{i=2}^{n-1}\sum_{j=1}^{i-1}\mu_4(X_j)\tilde{W}_{nij}\Big\} + E\Big(\sum_{i=2}^{n-1}\sum_{j=1}^{i-1}\tilde{W}_{nij}\Big) = O(n^2h^{2p}),$$

using the argument leading to (3.9). Therefore $\sum_{i=2}^{n} E'(Z_{ni}^4) = O(n^3h^{5p})$, from which follows (3.16). The result (3.17) may be proved by slightly modifying the argument leading to (3.9). This proves Lemma 2.

It follows from Lemma 1 and the identity (3.4) that $n^{-2}h^{-3p}V_n^2 \to \frac{1}{4}\alpha_1$ in probability as $n \to \infty$. From this result, Lemma 2 and a central limit theorem due to Brown (1971) (see Corollary 3.1, page 58 of Hall and Heyde, 1980), we may deduce that $n^{-1}h^{-3p/2}S_{nn}$ is asymptotically normal $N(0, \frac{1}{4}\alpha_1)$. It is not difficult to prove even more than this. The σ-fields $\mathscr{F}_{n0} = \mathscr{F}\{X_1, \ldots, X_n\}$ have the properties $E(Y_{ni} | \mathscr{F}_{n0}) = 0$ for all i, and $\mathscr{F}_{n0} \subseteq \mathscr{F}_{ni}$ for all i. By modifying arguments of Eagleson (1975) we may prove that for any sequence of events $A_n \in \mathscr{F}_{n0}$, and all real x,

$$| P(A_n;\ n^{-1}h^{-3p/2}S_{nn} \le \tfrac{1}{2}\alpha_1^{1/2}x) - P(A_n)\Phi(x) | \to 0$$

as $n \to \infty$. It now follows from (3.3) that

$$| P(A_n;\ 2nh^{p/2}I_{n2} \le \alpha_1^{1/2}x) - P(A_n)\Phi(x) | \to 0. \tag{3.18}$$

(iii) I_{n3}: Let $\gamma_n(x) = E\{g_n(x)\}$, and define

$$J_{nj} = (nh^p)^{-1} \textstyle\sum_{i=1}^n \{Y_i - \mu(X_i)\} \displaystyle\int_A K\{(x - X_i)/h\} a_j(x) v_n(x)\, dx,$$

for $j = 1, 2$, where $a_1(x) = \gamma_n(x)$ and $a_2(x) = g_n(x) - \gamma_n(x)$. Then

$$I_{n3} = J_{n1} + J_{n2}. \tag{3.19}$$

We shall handle these terms individually.

LEMMA 3. *$J_{n2} = o_p(n^{-1}h^{-p/2})$ as $n \to \infty$.*

PROOF. It suffices to prove that

$$E'(J_{n2}^2) = o_p(n^{-2}h^{-p}). \tag{3.20}$$

Since

$$\begin{aligned} E'(J_{n2}^2) &= (nh^p)^{-2} \textstyle\sum_{i=1}^n \sigma^2(X_i) \\ &\quad \times \left[\displaystyle\int_A K\{(x - X_i)/h\}\{g_n(x) - \gamma_n(x)\} v_n(x)\, dx\right]^2 \\ &\le (nh^p)^{-2}\{\sup_{x \in A} v_n^2(x)\} \left\{\displaystyle\int K(x/h)\, dx\right\} \\ &\quad \times \textstyle\sum_{i=1}^n \sigma^2(X_i) \displaystyle\int_A K\{(x - X_i)/h\}\{g_n(x) - \gamma_n(x)\}^2\, dx, \end{aligned}$$

then (3.20) will follow if we show that

$$\textstyle\sum_{i=1}^n \displaystyle\int_A K\{(x - X_i)/h\}\{g_n(x) - \gamma_n(x)\}^2\, dx = o_p(1). \tag{3.21}$$

Let $g_{ni}(x) = (nh^p)^{-1} \sum_{j \ne i}\{\mu(X_j) - \mu(x)\}K\{(x - X_j)/h\}$, and observe that

$$\begin{aligned} &\{g_n(x) - \gamma_n(x)\}^2 \\ &\quad \le 2\{g_{ni}(x) - Eg_{ni}(x)\}^2 + 4(nh^p)^{-2}\{\mu(X_i) - \mu(x)\}^2 K^2\{(x - X_i)/h\} \\ &\qquad + 4(nh^p)^{-2}[E\{\mu(X_i) - \mu(x)\}K\{(x - X_i)/h\}]^2. \end{aligned}$$

Consequently if $x \in A$,

$$\begin{aligned} b_n(x) &\equiv E[K\{(x - X_i)/h\}\{g_n(x) - \gamma_n(x)\}^2] \\ &\le C_1((nh^{2p})^{-1}E[K\{(x - X)/h\}]\ E[\{\mu(X) - \mu(x)\}^2K\{(x - X)/h\}] \\ &\quad + (nh^p)^{-2}E[\{\mu(X) - \mu(x)\}^2K\{(x - X)/h\}]) \\ &\le C_2(nh^p)^{-1}E[\{\mu(X) - \mu(x)\}^2K\{(x - X)/h\}], \end{aligned}$$

using the fact that f is bounded on A^ε. Now, if $K(z) = 0$ for $|z| > \lambda$ then

$$\begin{aligned} \int_A E[\{\mu(X) - \mu(x)\}^2K\{(x - X)/h\}]\ dx \\ &= h^p \int_A dx \int \{\mu(x - zh) - \mu(x)\}^2K(z)f(x - zh)\ dz \\ &\le h^p\{\sup_{x\in A,|z|\le\lambda h} |\mu(x + z) - \mu(x)|^2\} = o(h^p), \end{aligned}$$

using the uniform continuity of μ. This proves (3.21).

The next step is to derive a central limit theorm for J_{n1}. Conditional on $X_1, \ldots, X_n$, J_{n1} is just a sum of independent random variables with zero means, and so we shall prove the theorem conditional on $X_1, \ldots, X_n$. This requires us to check the two conditions in the following lemma. Let $Y_{ni} = \{Y_i - \mu(X_i)\}\tilde{Z}_{ni}$, where

$$\tilde{Z}_{ni} = \int_A K\{(x - X_i)/h\}\gamma_n(x)v_n(x)\ dx.$$

LEMMA 4.

$$n^{-1}h^{-2p-4}E'(\textstyle\sum_{i=1}^n Y_{ni})^2 \to \frac{1}{4}\alpha_2 \equiv \displaystyle\int_A \sigma^2(x)\gamma^2(x)v^2(x)f(x)\ dx \tag{3.22}$$

in probability as $n \to \infty$, and for each $\varepsilon > 0$,

$$n^{-1}h^{-2p-4} \textstyle\sum_{i=1}^n E'\{Y_{ni}^2I(|Y_{ni}| > \varepsilon n^{1/2}h^{p+2})\} \to 0 \tag{3.23}$$

in probability as $n \to \infty$.

PROOF. Since

$$\gamma_n(x) = \int \{\mu(x - zh) - \mu(x)\}K(z)f(x - zh)\ dz = h^2\gamma(x) + o(h^2) \tag{3.24}$$

uniformly in $x \in A$, and $v_n(x) = v(x) + o_p(1)$ uniformly in $x \in A$, then

$$\begin{aligned} T_{n1} &\equiv E'(\textstyle\sum_{i=1}^n Y_{ni})^2 = \sum_{i=1}^n \sigma^2(X_i)\tilde{Z}_{ni}^2 \\ &= h^4 \textstyle\sum_{i=1}^n \sigma^2(X_i)\left[\displaystyle\int_A K\{(x - X_i)/h\}\gamma(x)v(x)\ dx\right]^2 + r_n, \end{aligned} \tag{3.25}$$

where

$$|r_n| = o_p(h^4) \sum_{i=1}^n \left[\int_A K\{(x - X_i)/h\} \, dx\right]^4 = o_p(nh^{4p+4}),$$

using (3.8). Let T_{n2} denote the last-written series in (3.25). Then

$$E(T_{n2}) = nh^{2p} \int \sigma^2(z)f(z) \left[\int_{(A-z)/h} K(u)\gamma(z + uh)v(z + uh) \, du\right]^2 dz$$

$$\sim nh^{2p} \int \sigma^2(z)\gamma^2(z)v^2(z)f(z) \, dz,$$

while

$$\operatorname{var}(T_{n2}) \le CnE\left[\int_A K\{(x - X)/h\} \, dx\right]^4 \le Cnh^{4p}.$$

The result (3.22) follows on combining the estimates from (3.25) down.

To prove (3.23) it suffices to show that

$$n^{-2}h^{-4p-8} \sum_{i=1}^n E'(Y_{ni}^4) \to 0 \tag{3.26}$$

in probability as $n \to \infty$. Now,

$$\sup_i |\tilde{Z}_{ni}| = O_p(h^2) \sup_i \int_A K\{(x - X_i)/h\} \, dx = O_p(h^{p+2}),$$

using (3.24), and so

$$\sum_{i=1}^n E'(Y_{ni}^4) = \sum_{i=1}^n \tilde{Z}_{ni}^4 E'\{Y_i - \mu(X_i)\}^4 = O_p(h^{4p+8}) \sum_{\substack{i \le n, \\ X_i \in A^\epsilon}} \mu_4(X_i)$$

$$= O_p(nh^{4p+8}).$$

This proves (3.26).

Let F_n denote the distribution function of

$$T_{n3} \equiv 2\alpha_2^{-1/2} n^{-1/2} h^{-p-2} \sum_{i=1}^n Y_{ni},$$

conditional on $X_1, \ldots, X_n$. We may deduce from Lemma 4 and the Lindeberg-Feller theorem that $F_n(x) \to \Phi(x)$ in probability as $n \to \infty$. Therefore for any sequence of events $A_n \in \mathscr{F}\{X_1, \ldots, X_n\}$, $n \ge 1$, and all real x,

$$|P(A_n; T_{n3} \le x) - P(A_n)\Phi(x)| \to 0$$

as $n \to \infty$. Consequently

$$|P(A_n; n^{1/2}h^{-2}J_{n1} \le \tfrac{1}{2} \alpha_2^{1/2} x) - P(A_n)\Phi(x)| \to 0. \tag{3.27}$$

The expansion (2.3), with the term $(n^{-2}h^{-p}\alpha_1 + n^{-1}h^2\alpha_2)^{1/2}N_n$ replaced by $(nh^{p/2})^{-1}\alpha_1^{1/2}N_{n1} + n^{-1/2}h^2\alpha_2^{1/2}N_{n2}$ for asymptotically $N(0, 1)$ variables N_{n1} and N_{n2}, follows immediately on combining the results (3.1), (3.2), (3.18), (3.19), (3.27) and Lemma 2. If one of the terms $(nh^{p/2})^{-1}$ and $n^{-1/2}h^2$ is asymptotically negligible in comparison with the other as $n \to \infty$, then the theorem is proved. It remains only to derive the last term in (2.3) in the special case where $(nh^{p/2})^{-1}n^{-1/2}h^2 \to \lambda$, $0 < \lambda < \infty$, as $n \to \infty$ along a subsequence of n-values. This in turn entails proving the asymptotic independence of N_{n1} and N_{n2} as $n \to \infty$ along this subsequence. Now, N_{n1} and N_{n2} are principally derived from the terms I_{n2} and J_{n1} (see (3.18) and (3.27)), and I_{n2} and J_{n1} both have zero conditional mean and satisfy $E(I_{n2}J_{n1} \mid X_1, \ldots, X_n) = 0$. For all real constants a, b, the quantity $aI_{n2} + bJ_{n1}$ can be expressed as a martingale, very much as was done early in step (ii). The proof in steps (ii) and (iii) can now be reworked to show that $2(aI_{n2} + bJ_{n1})$ is asymptotically normal with variance $n^{-2}h^{-p}\alpha_1 a^2 + n^{-1}h^4\alpha_2 b^2$. The desired independence now follows via the Cramér-Wold device. (See Billingsley, 1968, page 48.)

PROOF OF THEOREM 2. In view of condition (2.2),

$$\sum_{i=1}^{n} \sigma^2(X_i) \int_A K^2\{(x - X_i)/h\}v_n(x)\,dx$$
$$= \{1 + o_p(1)\} \sum_{i=1}^{n} \sigma^2(X_i) \int_A K^2\{(x - X_i)/h\}v(x)\,dx.$$

The last-written series has mean equal to

$$nh^p \int \sigma^2(x)f(x)\,dx \int_{(A-x)/h} K^2(u)v(x - uh)\,du \sim nh^p \int_A \sigma^2(x)f(x)v(x)\,dx \cdot \int K^2(u)\,du,$$

and variance of order nh^{2p}. Therefore the first term on the right in (2.3) is asymptotically equivalent in probability to the first term on the right in (2.5).

Defining $g_n(x)$ and $\gamma_n(x)$ as in the previous proof, we see that the second term on the right in (2.3) equals $\{1 + o_p(1)\} \int_A g_n^2(x)v(x)\,dx$, while

$$\int_A g_n^2(x)v(x)\,dx = \int_A \{g_n(x) - \gamma_n(x)\}^2 v(x)\,dx$$
$$+ 2\int_A \{g_n(x) - \gamma_n(x)\}\gamma_n(x)v(x)\,dx + \int_A \gamma_n^2(x)v(x)\,dx,$$

$$\int_A E\{g_n(x) - \gamma(x)\}^2 v(x)\,dx$$
$$\leq (nh^p)^{-1} \int_A v(x)\,dx \int \{\mu(x - zh) - \mu(x)\}^2 K^2(z)f(x - zh)\,dz = o(n^{-1}h^{-p}),$$

$$\int_A E|g_n(x) - \gamma_n(x)\|\gamma_n(x)|v(x)\,dx$$

$$\leq \int_A |\gamma_n(x)|v(x)[(nh^p)^{-1}\int \{\mu(x-zh) - \mu(x)\}^2K^2(z)f(x-zh)\,dz]^{1/2}dx$$

$$= o\{h^2(nh^p)^{-1/2}\},$$

and

$$\int_A \gamma_n^2(x)v(x)\,dx = h^4\int_A \gamma^2(x)v(x)\,dx + o(h^4),$$

using (3.24). Therefore the second term on the right in (2.3) equals the second term on the right in (2.5), plus terms of $o_p(n^{-1}h^{-p} + h^4)$. The last-written term in (2.3) equals $o_p(n^{-1}h^{-p} + h^4)$.

PROOF OF THEOREM 3. The variance of the first term on the right in (2.3) is dominated by a constant multiple of

$$(n^3h^{4p})^{-1}E\left[\int_A K^2\{(x-X)/h\}\,dx\right]^2 = O(n^{-3}h^{-2p}) = o(n^{-2}h^{-p}).$$

Therefore the error about the mean of the first term is asymptotically negligible in comparison with the last term on the right in (2.3). The difference between the second term on the right in (2.3), and its mean, equals $L_{n1} + 2L_{n2}$, where

$$L_{n1} = (nh^p)^{-2}\textstyle\sum_{i=1}^n \displaystyle\int_A \xi_n^{(2)}(X_i, x)v(x)\,dx,$$

$$L_{n2} = (nh^p)^{-2}\sum\sum_{1\leq i<j\leq n}\int_A \xi_n^{(1)}(X_i, x)\xi_n^{(1)}(X_j, x)v(x)\,dx$$

and

$$\xi_n^{(l)}(X_i, x) = \{\mu(X_i) - \mu(x)\}^lK^l\{(x-X_i)/h\}$$
$$- E[\{\mu(X_i) - \mu(x)\}^lK^l\{(x-X_i)/h\}].$$

Therefore

$$\operatorname{var}(L_{n1}) = (n^3h^{4p})^{-1}\int_A\int_A E\{\xi_n^{(2)}(X, x)\xi_n^{(2)}(X, y)\}v(x)v(y)\,dx\,dy$$

and

$$\operatorname{var}(L_{n2}) = \frac{1}{2}(nh^p)^{-4}n(n-1)\int_A\int_A \{E\xi_n^{(1)}(X, x)\xi_n^{(1)}(X, y)\}^2v(x)v(y)\,dx\,dy.$$

Each of these integrals may be expanded into several terms, and shown to be of smaller order than $n^{-2}h^{-p} + n^{-1}h^4$. Therefore the error about the mean of the second term on the right in (2.3) is asymptotically negligible in comparison with the last term in (2.3). We treat here only the first term in an expansion of $\mathrm{var}(L_{n2})$, which (if $K(z)$ vanishes for $|z| > \lambda$) is dominated by

$$(n^2h^{4p})^{-1}\int_A\int_A [E\{|\mu(X)-\mu(x)|\,|\mu(X)-\mu(y)| \times K((x-X)/h)K((y-X)/h)\}]^2 v(x)v(y)\,dx\,dy$$

$$\leq \{\sup_{x\in A,|x-y|\leq\lambda h}|\mu(x)-\mu(y)|\}^4(n^2h^p)^{-1}\int_A v(x)\,dx \times \int_{(A-x)/h} v(x+yh)\left\{\int K(y+z)K(z)f(x-zh)\,dz\right\}^2 dy$$

$$= o(n^{-2}h^{-p}).$$

PROOF OF THEOREM 4. Observe that

$$\{\hat f_n(x)\}^{-2} = \{E\hat f_n(x)\}^{-2} + O_p(1)\,|\hat f_n(x) - E\hat f_n(x)| \tag{3.28}$$

uniformly in $x \in A$, and so the term I_{n4} from the proof of Theorem 1 may be written as

$$I_{n4} = \int_A g_n^2(x)\{E\hat f_n(x)\}^{-2}w(x)\,dx + O_p(1)\int_A g_n^2(x)\,|\hat f_n(x) - E\hat f_n(x)|\,dx. \tag{3.29}$$

The argument we used to prove Theorem 3 may be employed to show that the first term on the right in (3.29) equals

$$\int_A E\{g_n^2(x)\}\{E\hat f_n(x)\}^{-2}w(x)\,dx + o_p(n^{-1}h^{p/2} + n^{-1/2}h^2),$$

while the second term equals

$$O_p(1)\left\{\int_A g_n^4(x)\,dx\right\}^{1/2}\left\{\int_A |\hat f_n(x) - E\hat f_n(x)|^2\,dx\right\}^{1/2}.$$

It can be proved that

$$\int_A E\{g_n^4(x)\}\,dx = O\{(nh^p)^{-2}h^4 + h^8\}$$

and

$$\int_A E\{\hat f_n(x) - E\hat f_n(x)\}^2\,dx = O(n^{-1}h^{-p}),$$

and so the last-written term in (3.29) equals

$$O_p\{(nh^p)^{-3/2}h^2 + (nh^p)^{-1/2}h^4\} = o_p(n^{-1}h^{p/2} + n^{-1/2}h^2)$$

under the conditions imposed on h and p. Consequently

$$I_{n4} = \int_A E\{g_n^2(x)\}\{E\hat{f}_n(x)\}^{-2}w(x)\,dx + o_p(n^{-1}h^{p/2} + n^{-1/2}h^2). \tag{3.30}$$

Using a longer expansion than (3.28) we may deduce that

$$\begin{aligned} E'(I_{n1}) = {} & (nh^p)^{-2}\sum_{i=1}^n \sigma^2(X_i)\int_A K^2\{(x-X_i)/h\}\{E\hat{f}_n(x)\}^{-2}w(x)\,dx \\ & - 2(nh^p)^{-2}\sum_{i=1}^n \sigma^2(X_i)\int_A K^2\{(x-X_i)/h\}\{E\hat{f}_n(x)\}^{-3} \\ & \{\hat{f}_n(x) - E\hat{f}_n(x)\}w(x)\,dx \\ & + O_p\{(nh^p)^{-1}\}\int_A [(nh^p)^{-1}\sum_{i=1}^n \\ & K^2\{(x-X_i)/h\}]\{\hat{f}_n(x) - E\hat{f}_n(x)\}^2\,dx. \end{aligned} \tag{3.31}$$

The argument leading to Theorem 3 may be used to show that the first term on the right hand side in (3.31) equals its mean plus $o_p(n^{-1}h^{-p/2} + n^{-1/2}h^2)$, while the last term equals $O_p\{(nh^p)^{-2}\}$. The second term multiplied by $-\frac{1}{2}$ equals $L_{n1} + L_{n2} + L_{n3}$, where

$$L_{n1} = (nh^p)^{-3}\sum_{i=1}^n \int_A E[\sigma^2(X_i)K^2\{(x-X_i)/h\}]\xi_n(X_i, x)w_n(x)\,dx,$$

$$L_{n2} = (nh^p)^{-3}\sum_{i=1}^n \int_A \eta_n(X_i, x)\xi_n(X_i, x)w_n(x)\,dx,$$

$$L_{n3} = (nh^p)^{-3}\sum_{i\neq j}\sum \int_A \eta_n(X_i, x)\xi_n(X_j, x)w_n(x)\,dx,$$

$$\xi_n(X_i, x) = K\{(x-X_i)/h\} - E[K\{(x-X_i)/h\}],$$

$$\eta_n(X_i, x) = \sigma^2(X_i)K^2\{(x-X_i)/h\} - E[\sigma^2(X_i)K^2\{(x-X_i)/h\}]$$

and $w_n(x) = w(x)/\{E\hat{f}_n(x)\}^3$. Now, L_{n1} and L_{n3} both have zero mean and variance of order $(nh^p)^{-4}$, while L_{n2} has mean of order $(nh^p)^{-2}$ and variance of order $(nh^p)^{-4}$. Therefore the second term on the right hand side in (3.31) equals $O_p\{(nh^p)^{-2}\} = o_p(n^{-1}h^{-p/2})$, since $n^2h^{3p} \to \infty$, whence

$$\begin{aligned} E'(I_{n1}) = {} & (nh^p)^{-2}\sum_{i=1}^n \int_A E[\sigma^2(X_i)K^2\{(x-X_i)/h\}]\{E\hat{f}_n(x)\}^{-2}w(x)\,dx \\ & + o_p(n^{-1}h^{-p/2} + n^{-1/2}h^2). \end{aligned}$$

Theorem 4 follows from this estimate and (3.30).

Acknowledgments. The helpful comments of two referees have been incorporated into this revised version.

REFERENCES

BICKEL, P. J. and ROSENBLATT, M. (1973). On some global measures of the deviations of density function estimates. *Ann. Statist.* **1** 1071–1095.

BILLINGSLEY, P. (1968). *Convergence of Probability Measures.* Wiley, New York.

BROWN, B. M. (1971). Martingale central limit theorems. *Ann. Math. Statist.* **42** 59–66.

COLLOMB, G. (1977). Quelques propriétés de la méthode du noyau pour l'estimation non-paramétrique de la régression en un point fixé. *Comptes Rendus Acad. Sci. Paris* **285 A** 289–292.

COLLOMB, G. (1981). Estimation non-paramétrique de la régression: revue bibliographique. *Int. Statist. Rev.* **49** 75–93.

CSÖRGÖ, M. and RÉVÉSZ, P. (1981). *Strong Approximations in Probability and Statistics.* Academic, New York and Budapest.

DEVROYE, L. P. and WAGNER, T. J. (1979). Distribution-free consistency results in nonparametric discrimination and regression function estimation. *Ann. Statist.* **8** 231–239.

EAGLESON, G. K. (1975). Martingale convergence to mixtures of infinitely divisible laws. *Ann. Probab.* **3** 557–562.

HALL, P. (1981). Central limit theorems for integrated square error of multivariate nonparametric density estimators. *J. Multivariate Anal.*, to appear.

HALL, P. and HEYDE, C. C. (1980). *Martingale Limit Theory and its Application.* Academic, New York.

JOHNSTON, G. J. (1982). Probabilities of maximum deviations for nonparametric regression function estimates. *J. Multivariate Anal.* **12** 402–414.

KONAKOV, V. D. (1977). On a global measure of deviation for an estimate of the regression line. *Theor. Probab. Appl.* **22** 858–868.

LIERO, H. (1982). On the maximal deviation of the kernel regression function estimate. *Math. Oper. Statist. ser. Statist.* **13** 171–182.

NADARAYA, E. A. (1964). On estimating regression. *Theor. Probab. Appl.* **9** 141–142.

NADARAYA, E. A. (1974). A limit distribution of the squared error deviation of nonparametric estimators of the regression function. (In Russian.) *Bull. Acad. Sci. Georgian S.S.R.* **74** 33–36.

ROSENBLATT, M. (1975). A quadratic measure of deviation of two-dimensional density estimates and a test of independence. *Ann. Statist.* **3** 1–14.

STONE, C. J. (1982). Optimal global rates of convergence for nonparametric regression. *Ann. Statist.* **10** 1040–1053.

WATSON, G. S. (1964). Smooth regression analysis. *Sankhyā* **A26** 359–372.

DEPARTMENT OF STATISTICS
FACULTY OF ECONOMICS
THE AUSTRALIAN NATIONAL UNIVERSITY
GPO BOX 4
CANBERRA A.C.T. 2601
AUSTRALIA