JOURNAL OF MULTIVARIATE ANALYSIS 19, 1–13 (1986)

Convergence Properties of an Empirical Error Criterion for Multivariate Density Estimation

JAMES STEPHEN MARRON

University of North Carolina at Chapel Hill

Communicated by G. Kallianpur

For the purpose of comparing different nonparametric density estimators, Wegman (*J. Statist. Comput. Simulation* **1** 225–245) introduced an empirical error criterion. In a recent paper by Hall (*Stochastic Process. Appl.* **13** 11–25) it is shown that this empirical error criterion converges to the mean integrated square error. Here, in the case of kernel estimation, the results of Hall are improved in several ways, most notably multivariate densities are treated and the range of allowable bandwidths is extended. The techniques used here are quite different from those of Hall, which demonstrates that the elegant Brownian Bridge approximation of Komlós, Major, and Tusnády (*Z. Warsch. Verw. Gebrete* **32** 111–131) does not always give the strongest results possible.

1. INTRODUCTION

Consider the problem of estimating a probability density $f(x)$, defined on $\mathbb{R}^d$, using a sample of random vectors $X_1, ..., X_n$ from f. If $\hat{f}(x) = \hat{f}(x, X_1, ..., X_n)$ denotes some estimator of $f(x)$, a very popular means of measuring the accuracy of $\hat{f}$ is Mean Integrated Square Error (MISE). Given a nonnegative function $w(x)$, this error criterion is defined by

$$\text{MISE} = E \int [\hat{f}(x) - f(x)]^2 \, w(x) \, f(x) \, dx. \tag{1.1}$$

Writing the weight function in the form $w(x) f(x)$ is for notational convenience in the proof of this paper. There is little loss of generality in this device because if a weight function $w^*(x)$ is desired, simply take $w(x) = w^*(x) f(x)^{-1}$.

Received April 1983; revised October 1983.

AMS 1980 Subject Classifications: Primary 62G20; Secondary 62H12.

Key words and Phrases: Nonparametric density estimation, kernel estimator, empirical error criterion, stochastic measures of accuracy.

0047-259X/86 $3.00

In a survey paper, Wegman [6] was interested in comparing the MISE of several different density estimators, in the case $d = 1$. Unfortunately, as Wegman pointed out, the computation of MISE can be quite tedious. Hence Wegman used the empirical error criterion

$$\widehat{\text{MISE}} = n^{-1} \sum_{j=1}^{n} [\hat{f}(X_j) - f(X_j)]^2 \, w(X_j),$$

and gave some heuristic justification.

Recently there has been some controversy regarding this practice. The difficulties have been essentially settled by Hall [1], who has shown that Wegman's heuristics are valid if $\hat{f}(x)$ is either a kernel estimator or an orthogonal series estimator. Hall has shown that, in the case $d = 1$, under relatively mild conditions on f and w, as $n \to \infty$,

$$\text{MISE} = \text{MISE} + o_p(\text{MISE}).$$

In this paper, Hall's Theorem 1 (dealing with kernel estimation) is extended in several ways. First, the assumptions made on f and K are substantially weaker here. Of more interest, the case of general dimension d is treated here. Also of importance is the fact that the bandwidth of the kernel estimator satisfies much weaker restrictions than in Hall's theorem. This is vital to the results of Marron [3] where an asymptotically efficient means of choosing the bandwidth is proposed.

It is interesting to note that the crude, "brute force" method of proof used in this paper gives stronger results than the elegant techniques employed by Hall.

2. Assumptions and Statement of Theorem

Given a "bandwidth," $h > 0$, and a "kernel function," K, defined on $\mathbb{R}^d$, the usual kernel estimator of $f(x)$ is given by

$$\hat{f}(x, h) = \frac{1}{nh^d} \sum_{j=1}^{n} K\left(\frac{x - X_j}{h}\right). \tag{2.1}$$

For the rest of this paper it is assumed that K satisfies the following assumptions.

(K.1) $\int K(x)\, dx = 1$.

(K.2) K is bounded,

(K.3) $\int K(x)^2 \, dx < \infty$,

(K.4) $\overline{\lim}_{\|x\| \to \infty} \|x\| \, |K(x)| < \infty$, where the integrals are taken over $\mathbb{R}^d$ and $\|\cdot\|$ denotes the usual Euclidean norm. It will also be assumed that the underlying density, f, and the kernel, K, satisfy the assumptions

(f.1) f is bounded,

(f.2) f is uniformly continuous,

(w.1) $w(x)$ is nonnegative,

(w.2) $w(x)$ is bounded.

Assumption (K.3) and (f.1) are redundant, but are included because they seem to lend some insight to the upcoming proofs. Assumption (K.1) is required to make $\hat{f}(x, h)$ integrate to 1. Assumptions (K.2), (K.3), and (K.4) are milder than those used in Theorem 1 of Hall [1]. Assumptions (f.1) and (f.2) are much weaker than the bounded variation and differentiability assumptions used in Theorem 1 of Hall. It should be noted that the assumptions (w.1) and (w.2) are more restrictive than they appear at first glance, in particular the popular choice $w(x) = 1/f(x)$ (i.e., $w^*(x) \equiv 1$) is eliminated.

Given sequences $\{a_n\}$ and $\{b_n\}$ it will be convenient to let the phrase "h is between a_n and b_n" mean the sequence $h = h(n)$ satisfies

$$\lim_{n \to \infty} a_n h^{-1} = 0, \qquad \lim_{n \to \infty} b_n h^{-1} = \infty.$$

The main theorem of this paper can now be stated.

THEOREM 1. *Under the assumptions* (K.1)–(K.4), (f.1), (f.2), (w.1), *and* (w.2), *if $\hat{f}$ is the kernel estimator of* (2.1) *and if h is between $n^{-1/d}$ and* 1, *then as $n \to \infty$,*

$$\widehat{\text{MISE}} = \text{MISE} + o_p(\text{MISE}).$$

Using the expansion (3.1) in the next section, together with Taylor's Theorem, it is easy to see that Theorem 1 of Hall (1982) follows easily from the above. The above theorem is in fact stronger because the case that f has fewer derivatives than K has vanishing moments is included here (this case has been treated, e.g., by Stone [5]).

3. PROOF OF THEOREM 1

First, note that by the familiar variance and squared bias decomposition (see Rosenblatt [4]),

$$\text{MISE} = n^{-1}h^{-d}\left(\int f(x)^2\, w(x)\, dx\right)\left(\int K(u)^2\, du\right) + o(n^{-1}h^{-d})$$
$$+ \int \left[\int K(u)\, f(x-hu)\, du - f(x)\right]^2 w(x)\, f(x)\, dx. \qquad (3.1)$$

It will be convenient to denote the squared bias part of this expansion by

$$s_f(h) = \int \left[\int K(u)\, f(x-hu)\, du - f(x)\right]^2 w(x)\, f(x)\, dx. \qquad (3.2)$$

Note that by (f.2)

$$\lim_{h \to 0} s_f(h) = 0. \qquad (3.3)$$

It is seen in Marron [3] that the rate of convergence of $s_f(h)$ to 0 provides a measure of the "smoothness" of f.

Next, for $j = 1, ..., n$, it will be useful to define the "leave one out" kernel estimator

$$\hat{f}_j(x, h) = \frac{1}{(n-1)\, h^d} \sum_{i \neq j} K\left(\frac{x - X_i}{h}\right). \qquad (3.4)$$

A quantity which is more tractable than $\widehat{\text{MISE}}$ is

$$\widehat{\widehat{\text{MISE}}} = \frac{1}{n} \sum_{j=1}^{n} [\hat{f}_j(X_j, h) - f(X_j)]^2\, w(X_j). \qquad (3.5)$$

Note that for $j = 1, ..., n$

$$\hat{f}_j(x, h) - \hat{f}(x, h) = \frac{1}{n(n-1)\, h^d} \sum_{i \neq j} K\left(\frac{x - X_i}{h}\right) - \frac{1}{nh^d} K\left(\frac{x - X_j}{h}\right).$$

Hence, by (K.2), for h between $n^{-1/d}$ and 1,

$$\sup_x \sup_{j=1,...,n} |\hat{f}_j(x, h) - \hat{f}(x, h)| = O(n^{-1}h^{-d}). \qquad (3.6)$$

Theorem 1 is now a consequence of

THEOREM 2. *Under the conditions of Theorem* 1,

$$\widehat{\widehat{\text{MISE}}} = \text{MISE} + o_p(\text{MISE}).$$

4. Proof of Theorem 2

First, for $j = 1,\dots, n$ define

$$U_j = \frac{[\hat{f}_j(X_j, h) - f(X_j)]^2\, w(X_j)}{\text{MISE}}. \tag{4.1}$$

Note that by (1.1), (3.1), and (3.6), for h between $n^{-1/d}$ and 1,

$$\begin{aligned} EU_j &= \text{MISE}^{-1} \int E[\hat{f}_j(x, h) - f(x)]^2\, w(x)\, f(x)\, dx \\ &= \text{MISE}^{-1} \int E[\hat{f}(x, h) - f(x)]^2\, w(x)\, f(x)\, dx + O(n^{-1}h^{-d}) \\ &= 1 + o(1). \end{aligned} \tag{4.2}$$

The proof of Theorem 2 will be complete when a weak law of large numbers is established for the U_j.

For $i \neq j = 1,\dots, n$ define

$$V_{ij} = \left[h^{-d} K\left(\frac{X_j - X_i}{h}\right) - f(X_j)\right] w(X_j)^{1/2}. \tag{4.3}$$

It follows from (3.4) that, for $j = 1,\dots, n$,

$$[\hat{f}_j(X_j, h) - f(X_j)]^2\, w(X_j) = \left[(n-1)^{-1} \sum_{i \neq j} V_{ij}\right]^2.$$

Hence,

$$\text{var}(U_j) = (n-1)^{-4}\, \text{MISE}^{-2}\, \text{var}\left[\sum_{i \neq j} V_{ij}\right]^2. \tag{4.4}$$

Now if this last variance is expanded in terms of variances and covariances, there will be $(n-1)^4$ terms of the following types (where i, i', i'', i^*, j are all different).

No. of terms	General terms
$O(n)$	$\text{var}(V_{ij}^2)$
$O(n^2)$	$\text{var}(V_{ij} V_{i'j})$
$O(n^2)$	$\text{cov}(V_{ij}^2, V_{i'j}^2)$
$O(n^2)$	$\text{cov}(V_{ij}^2, V_{ij} V_{i'j})$
$O(n^3)$	$\text{cov}(V_{ij}^2, V_{i'j} V_{i''j})$
$O(n^3)$	$\text{cov}(V_{ij} V_{i'j}, V_{ij} V_{i''j})$
$O(n^4)$	$\text{cov}(V_{ij} V_{i'j}, V_{i''j} V_{i^*j})$

Bounds will now be computed for each general term.

Using (K.2), (K.3), (f.1), (w.2), and (4.3), as $h \to 0$,

$$\operatorname{var}(V_{ij}^2) \leqslant EV_{ij}^4 = \iint \left[h^{-d}K\left(\frac{x-y}{h}\right) - f(x) \right]^4 w(x)^2 f(x) f(y)\, dxdy$$

$$= \iint [h^{-3d}K(u)^4 - 4h^{-2d}K(u)^3 f(x) + 6h^{-d}K(u)^2 f(x)^2$$

$$- 4K(u) f(x)^3 + h^d f(x)^4]\, w(x)^2 f(x) f(x-hu)\, dudx$$

$$= O(h^{-3d}). \tag{4.5}$$

Similarly,

$$\operatorname{var}(V_{ij}V_{i'j}) \leqslant EV_{ij}^2 V_{i'j}^2$$

$$= \iiint \left[h^{-d}K\left(\frac{x-y}{h}\right) - f(x) \right]^2 \left[h^{-d}K\left(\frac{x-z}{h}\right) - f(x) \right]^2$$

$$\times w(x)^2 f(x) f(y) f(z)\, dxdydz$$

$$= \int \left[\int [h^{-d}K(u)^2 - 2K(u) f(x) + h^d f(x)^2] f(x-hu)\, du \right]^2 f(x)\, dx$$

$$= O(h^{-2d}). \tag{4.6}$$

Very similar computations show that

$$EV_{ij}^2 = O(h^{-d}). \tag{4.7}$$

It follows from the above that

$$\operatorname{cov}(V_{ij}^2, V_{i'j}^2) = O(h^{-2d}).$$

In the same way it is easily seen that

$$\operatorname{cov}(V_{ij}^2, V_{ij}V_{i'j}) = O(h^{-2d}),$$

$$\operatorname{cov}(V_{ij}^2, V_{i'j}V_{i''j}) = O(h^{-d}),$$

$$\operatorname{cov}(V_{ij}V_{i'j}, V_{ij}V_{i''j}) = O(h^{-d}).$$

To bound the final term, note that

$$E[V_{ij} \mid X_j] = \int \left[h^{-d}K\left(\frac{X_j - y}{h}\right) - f(X_j) \right] w(X_j)^{1/2} f(y)\, dy. \tag{4.8}$$

Hence, by (3.2)

$$E[V_{ij}V_{i'j}] = E(E[V_{ij}|X_j]^2)$$
$$= \int \left[\int K(u) f(x-hu)\,du - f(x) \right]^2 w(x) f(x)\,dx$$
$$= s_f(h). \tag{4.9}$$

As another consequence of (4.8), note that by (K.2), (f.1) and (w.2),

$$\sup_x |E[V_{ij}|X_j = x]| = \sup_x \left| w(x)^{1/2} \left[\int K(u) f(x+hu)\,du - f(x) \right] \right| < \infty. \tag{4.10}$$

It follows from the above that

$$E[V_{ij}V_{i'j}V_{i''j}V_{i*j}] = E(E[V_{ij}|X_j]^4)$$
$$\leqslant \sup_x E[V_{ij}|X_j = x]^2\, E(E[V_{ij}|X_j]^2)$$
$$= O(s_f(h)).$$

Hence,

$$\operatorname{cov}(V_{ij}V_{i'j},\, V_{i''j}V_{i*j}) = O(s_f(h)).$$

Looking back to (4.4), it is now apparent that for h between $n^{-1/d}$ and 1,

$$\operatorname{var}(U_j) = \frac{O(n^{-1}h^{-d}) + O(s_f(h))}{\text{MISE}^2}.$$

Hence, by (3.1) and (3.2), there is a constant C so that, for h between $n^{-1/d}$ and 1,

$$n^{-1}\operatorname{var}(U_j) = \frac{O(n^{-1}h^{-d}) + O(s_f(h))}{n[Cn^{-1}h^{-d} + s_f(h) + o(n^{-1}h^{-d})]^2}$$
$$\leqslant \frac{O(n^{-1}h^{-d})}{C^2 n^{-1} h^{-2d}} + \frac{O(s_f(h))}{2Ch^{-d}s_f(h)}$$
$$= O(h^d) \to 0. \tag{4.11}$$

Next, for $j \neq j'$, a similar bound will be computed for

$$\operatorname{cov}(U_j, U_{j'}) = (n-1)^{-4}\,\text{MISE}^{-2} \operatorname{cov}\left(\left[\sum_{i \neq j} V_{ij} \right]^2, \left[\sum_{i \neq j'} V_{ij'} \right]^2 \right). \tag{4.12}$$

Once again, by the linearity of covariance, the right-hand covariance in (4.12) may be written as the sum of $(n-1)^4$ terms of the form (again assume i, i', i'', i^*, j, and j' are all different):

No. of terms	General term
$O(1)$	$\operatorname{cov}(V_{j'j}^2, V_{jj'}^2)$
$O(n)$	$\operatorname{cov}(V_{j'j}^2, V_{ij'}^2)$
$O(n)$	$\operatorname{cov}(V_{ij}^2, V_{ij'}^2)$
$O(n^2)$	$\operatorname{cov}(V_{ij}^2, V_{i'j'}^2)$
$O(n)$	$\operatorname{cov}(V_{j'j}^2, V_{jj'} V_{ij'})$
$O(n^2)$	$\operatorname{cov}(V_{j'j}^2, V_{ij'} V_{i'j'})$
$O(n)$	$\operatorname{cov}(V_{ij}^2, V_{ij'} V_{jj'})$
$O(n^2)$	$\operatorname{cov}(V_{ij}^2, V_{ij'} V_{i'j'})$
$O(n^2)$	$\operatorname{cov}(V_{ij}^2, V_{jj'} V_{i'j'})$
$O(n^3)$	$\operatorname{cov}(V_{ij}^2, V_{i'j'} V_{i''j'})$
$O(n)$	$\operatorname{cov}(V_{j'j} V_{ij}, V_{jj'} V_{ij'})$
$O(n^2)$	$\operatorname{cov}(V_{j'j} V_{ij}, V_{jj'} V_{i'j'})$
$O(n^2)$	$\operatorname{cov}(V_{j'j} V_{ij}, V_{ij'} V_{i'j'})$
$O(n^3)$	$\operatorname{cov}(V_{j'j} V_{ij}, V_{i'j'} V_{i''j'})$
$O(n^2)$	$\operatorname{cov}(V_{ij} V_{i'j}, V_{ij'} V_{i'j'})$
$O(n^3)$	$\operatorname{cov}(V_{ij} V_{i'j}, V_{ij'} V_{i''j'})$
$O(n^4)$	$\operatorname{cov}(V_{ij} V_{i'j}, V_{i''j'} V_{i^*j'})$

These general terms will now be bounded in order of increasing difficulty.

By the independence of $X_1, \ldots, X_n$,

$$\operatorname{cov}(V_{ij}^2, V_{i'j'}^2) = \operatorname{cov}(V_{ij}^2, V_{i'j'} V_{i''j'}) = \operatorname{cov}(V_{ij} V_{i'j}, V_{i''j'} V_{i^*j'}) = 0,$$

Using the Schwartz Inequality, (4.5) and (4.6), each term which appears $O(n)$ times may be bounded by $O(h^{-3d})$.

Using (4.7) and (4.10)

$$\begin{aligned} E[V_{j'j}^2 V_{ij'} V_{i'j'}] &= E(V_{j'j}^2 E[V_{ij'} \mid X_j, X_{j'}]\, E[V_{i'j'} \mid X_j, X_{j'}]) \\ &\leqslant \sup_x E[V_{ij} \mid X_j = x]^2\, E(V_{j'j}^2) = O(h^{-d}). \end{aligned}$$

Thus, by (3.3) and (4.9),

$$\operatorname{cov}(V_{j'j}^2, V_{ij'} V_{i'j'}) = O(h^{-d}).$$

Similarly, again using the Schwartz Inequality,

$$\begin{aligned} E(V_{j'j} V_{ij} V_{jj'} V_{i'j'}) &= E(V_{j'j} V_{jj'} E[V_{ij} \mid X_j, X_{j'}]\, E[V_{i'j'} \mid X_j, X_{j'}]) \\ &\leqslant \sup E[V_{ij} \mid X_j = x]^2\, E(V_{j'j} V_{jj'}) = O(h^{-d}). \end{aligned}$$

from which it follows that

$$\operatorname{cov}(V_{j'j}V_{ij},\, V_{jj'}V_{i'j'}) = O(h^{-d}).$$

By the above techniques it is easy to see that

$$\operatorname{cov}(V_{j'j}V_{ij},\, V_{i'j'}V_{i''j'}) = O(s_f(h)).$$

The remaining terms will be a little more difficult. By (4.3),

$$\begin{aligned}
E[V_{ij}^2V_{ij'}V_{i'j'}] &= E(V_{ij}^2V_{ij'}E[V_{i'j'}\,|\,X_i, X_j, X_{j'}]) \\
&= \iiint \left[h^{-d}K\left(\frac{y-x}{h}\right) - f(y)\right]^2 \left[h^{-d}K\left(\frac{z-x}{h}\right) - f(z)\right] \\
&\quad \times w(y)\, w(z)^{1/2}\, E[V_{ij}\,|\,X_j = z]\, f(x)\, f(y)\, f(z))\, dxdydz \\
&= \iiint h^{-d}[K(u) - h^d f(x+hu)]^2[K(v) - h^d f(x+hv)]. \\
&\quad \times w(x+hu)\, w(x+hv)^{1/2}\, E[V_{ij}\,|\,X_j = x+hv] \\
&\quad \times f(x)\, f(x+hu)\, f(x+hv)\, dxdudv \\
&= O(h^{-d}),
\end{aligned}$$

and hence,

$$\operatorname{cov}(V_{ij}^2,\, V_{ij'}V_{i'j'}) = O(h^{-d}).$$

Similarly,

$$\begin{aligned}
E[V_{ij}^2V_{jj'}V_{i'j'}] &= E(V_{ij}^2V_{jj'}E[V_{i'j'}\,|\,X_i, X_j, X_{j'}]) \\
&= \iiint \left[h^{-d}K\left(\frac{y-x}{h}\right) - f(y)\right]^2 \left[h^{-d}K\left(\frac{z-y}{h}\right) - f(z)\right] \\
&\quad \times w(y)\, w(z)^{1/2}\, E[V_{ij}\,|\,X_j = z]\, f(x)\, f(y)\, f(z)\, dxdydz \\
&= O(h^{-d}),
\end{aligned}$$

and so,

$$\operatorname{cov}(V_{ij}^2,\, V_{jj'}V_{i'j'}) = O(h^{-d}).$$

By the same technique,

$$
\begin{aligned}
&E[V_{j'j}V_{ij}V_{ij'}V_{i'j'}] \\
&\quad = E(V_{j'j}V_{ij}V_{ij'}E[V_{i'j'} \mid X_j, X_{j'}, X_i]) \\
&\quad = \iiint \left[h^{-d}K\left(\frac{y-x}{h}\right) - f(y)\right]\left[h^{-d}K\left(\frac{y-z}{h}\right) - f(y)\right] \\
&\qquad \times \left[h^{-d}K\left(\frac{x-z}{h}\right) - f(x)\right] \\
&\qquad \times w(y)\, w(x)^{1/2}\, E[V_{ij} \mid X_j = x]\, f(x)\, f(y)\, f(z)\, dxdydz \\
&\quad = O(h^{-d}),
\end{aligned}
$$

and hence,

$$\operatorname{cov}(V_{j'j}V_{ij}, V_{ij'}V_{i'j'}) = O(h^{-d}).$$

The most difficult terms have been saved for last. It will be convenient to define the sets

$$
\begin{aligned}
A &= \{(x, y) \colon \|x-y\| > 2h^{1/2}\}, \\
A^c &= \mathbb{R}^d \times \mathbb{R}^d \backslash A.
\end{aligned}
$$

Note that

$$
\begin{aligned}
E(V_{ij}V_{i'j}V_{ij'}V_{i'j'}) &= E(E[V_{ij}V_{ij'} \mid X_j, X_{j'}]^2) \\
&= \iint \left\{ \int \left[h^{-d}K\left(\frac{x-z}{h}\right) - f(x)\right] \right. \\
&\qquad \left. \times \left[h^{-d}K\left(\frac{y-z}{h}\right) - f(y)\right] f(z)\, dz \right\}^2 w(x) \\
&\qquad w(y)\, f(x)\, f(y)\, dxdy \\
&= I + I', \qquad (4.13)
\end{aligned}
$$

where

$$I = \iint_A \{\ \}^2\, w(x)\, w(y)\, f(x)\, f(y)\, dxdy,$$

$$I' = \iint_{A^c} \{\ \}^2\, w(x)\, w(y)\, f(x)\, f(y)\, dxdy.$$

Note that, uniformly over x, y,

$$\int \left[h^{-d}K\left(\frac{x-z}{h}\right)-f(x)\right]\left[h^{-d}K\left(\frac{y-z}{h}\right)-f(y)\right]f(z)\,dz=O(h^{-d}).$$

Thus, since for each $x\in\mathbb{R}^d$, the set $\{y: (x, y)\in A^c\}$ has Lebesgue measure $O(h^{d/2})$,

$$I'=O(h^{-3d/2}). \tag{4.14}$$

Now given $(x, y)\in A$, note that the sets $A_1=\{z: \|x-z\|<h^{1/2}\}$ and $A_2=\{z: \|y-z\|<h^{1/2}\}$ are disjoint. Let $A_3=\mathbb{R}^d\backslash(A_1\cup A_2)$. Define I_1, I_2, and I_3 by

$$\begin{aligned} I&=\iint_A\left[\int_{A_1} dz+\int_{A_2} dz+\int_{A_3} dz\right]^2 w(x)\,w(y)\,f(x)\,f(y)\,dxdy\\ &=\iint_A [I_1+I_2+I_3]^2\,w(x)\,w(y)\,f(x)\,f(y)\,dxdy. \end{aligned}$$

Using (K.4), uniformly over $(x, y)\in A$ and $z\in A_1$,

$$\left|K\left(\frac{y-z}{h}\right)\right|=O\left(\left\|\frac{y-z}{h}\right\|^{-1}\right)=O(h^{1/2}).$$

Hence,

$$\begin{aligned}\sup_{(x,y)\in A} I_1&=\sup_A\int_{A_1}\left[h^{-d}K\left(\frac{x-z}{h}\right)-f(x)\right]\left[h^{-d}K\left(\frac{y-z}{h}\right)-f(y)\right]f(z)\,dz\\ &=O(h^{(1/2)-d}).\end{aligned}$$

Similarly,

$$\sup_A I_2=O(h^{(1/2)-d}),$$
$$\sup_A I_3=O(h^{(1/2)-d}).$$

It follows from the above that

$$I=O(h^{1-2d}).$$

Now from (4.13) and (4.14),

$$\operatorname{cov}(V_{ij}V_{i'j},\, V_{ij'}V_{i'j'})=O(h^{-3d/2})+O(h^{1-2d})=o(h^{-2d}).$$

The above computations will prove useful in bounding the last term. Using the Schwartz Inequality and the independence of X_j and $X_{j'}$,

$$\begin{aligned} &|E(V_{ij} V_{i'j} V_{ij'} V_{i''j'})| \\ &\quad = |E(E[V_{ij} V_{ij'} \mid X_j, X_{j'}]\, E[V_{i'j} \mid X_j]\, E[V_{i''j'} \mid X_{j'}])| \\ &\quad \leqslant [E(E[V_{ij} V_{ij'} \mid X_j, X_{j'}]^2)]^{1/2} [E(E[V_{i'j} \mid X_j]^2)\, E(E[V_{i''j'} \mid X_{j'}]^2)]^{1/2}. \end{aligned}$$

Note that the first factor on the right side appears in (4.13). Thus the computations following (4.13) together with (4.9) imply

$$\operatorname{cov}(V_{ij} V_{i'j}, V_{ij'} V_{i''j'}) = o(h^{-d} s_f(h)).$$

Now looking back to (4.12) it is apparent that, for h between $n^{-1/d}$ and 1,

$$\operatorname{cov}(U_j, U_{j'}) = \frac{o(n^{-2} h^{-2d}) + o(n^{-1} h^{-d} s_f(h))}{\text{MISE}^2}.$$

Thus by computations similar to (4.11), for h between $n^{-1/d}$ and 1,

$$\operatorname{cov}(U_j, U_{j'}) \to 0.$$

It follows from this together with (4.11) that for h between $n^{-1/d}$ and 1,

$$\operatorname{var}\left(n^{-1} \sum_{j=1}^{n} U_j\right) \to 0,$$

and hence, by the Chebychev Inequality,

$$n^{-1} \sum_{j=1}^{n} U_j \to E(U_j) \qquad \text{in probability.}$$

Theorem 2 is an easy consequence of this together with (3.5), (4.1), and (4.2).

Acknowledgment

The author is grateful to David Ruppert for recommending the techniques employed in this paper.

References

[1] Hall, P. (1982). Limit theorems for stochastic measures of the accuracy of nonparametric density estimators. *Stochastic Process. Appl.* **13** 11–25.

[2] Komlós, J., Major, P., and Tusnády (1975). An approximation of partial sums of independent random variables, and the sample distribution function. *Z. Wahrsch. Verw. Gebiete* **32** 111–131.

[3] Marron, J. S. (1983). An asymptotically efficient solution to the bandwidth problem of kernel density estimation. North Carolina Institute of Statistics, Mimeo Series, No. #1518.

[4] Rosenblatt, M. (1971). Curve estimates. *Ann. Math. Statist.* **42** 1815–1842.

[5] Stone, C. J. (1980). Optimal rates of convergence for nonparametric estimators. *Ann. Statist.* **8** 1348–1360.

[6] Wegman, E. J. (1972). Nonparametric density estimation. II. A comparison of density estimation methods. *J. Statist. Comput. Simulation* **1** 225–245.