

CONVERGENCE PROPERTIES OF AN EMPIRICAL ERROR CRITERION
FOR MULTIVARIATE DENSITY ESTIMATION

James Stephen Marron
University of North Carolina at Chapel Hill

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

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

ABSTRACT

For the purpose of comparing different nonparametric density estimators, Wegman (1972) introduced an empirical error criterion. In a recent paper by Hall (1982) 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 (1975) 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, \dots, X_n from f . If $\hat{f}(x) = \hat{f}(x, X_1, \dots, X_n)$ denotes some estimator of $f(x)$, a very popular means of measuring the accuracy of \hat{f} is Mean Integrated Square Error. Given a nonnegative function $w(x)$, this error criterion is defined by

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

Writing the weight function in the form $w(x)f(x)$ is for notational convenience in the proofs of this paper. There is no loss of generality in this device because if a weight function $w^*(x)$ is desired, simply take $w(x) = w^*(x)f(x)^{-1}$.

In a survey paper, Wegman (1972) 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

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

and gave some heuristic justification.

Recently there has been some controversy regarding this practice. The difficulties have been essentially settled by Hall (1982), 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 $n \rightarrow \infty$,

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

In this paper, Hall's theorem 1 (dealing with kernel estimation) is extended in several ways. First, the assumptions made on f , K and w 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 (1983) 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

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

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

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

$$(K.2) \quad K \text{ is bounded,}$$

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

$$(K.4) \quad \overline{\lim}_{\|x\| \rightarrow \infty} \|x\| |K(x)| < \infty,$$

where the integrals are taken over \mathbb{R}^d and $\|\cdot\|$ denotes the usual Euclidean norm.

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 (1982).

It will also be assumed that the underlying density satisfies the following assumptions:

(f.1) f is bounded,

(f.2) f is uniformly continuous.

Assumption (f.1) is a consequence of (f.2) but it is required often enough in the upcoming proofs to be listed separately. Assumption (f.2) is much weaker than the bounded variation and differentiability assumptions used in theorem 1 of Hall.

Finally, it is assumed that,

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

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

Hall deals only with the special case $w(x) \equiv 1$.

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

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

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 $n \rightarrow \infty$,

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

It is worth noting that in the case $d = 1$ the range of allowable h is much larger than that of Hall (1982). This fact is vital to the results of Marron (1983a).

3. PROOF OF THEOREM

First note that by the familiar variance and bias² decomposition (see Rosenblatt (1971) or (3.6) of Marron (1983)),

$$(3.1) \quad \begin{aligned} \text{MISE} = n^{-1}h^{-d} & (\int f(x)^2 w(x) dx) (\int K(u)^2 du) + o(n^{-1}h^{-d}) + \\ & + \int [\int K(u) f(x-hu) du - f(x)]^2 w(x) f(x) dx . \end{aligned}$$

It will be convenient to let the bias² part of this expansion be denoted by

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

Note that by (f.2)

$$(3.3) \quad \lim_{h \rightarrow 0} s_f(h) = 0 .$$

It is seen in Marron (1983) that the rate of convergence of $s_f(h)$ to 0 provides a measure of the "smoothness" of f .

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

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

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

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

Note that for $j = 1, \dots, n$

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

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

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

Theorem 1 is now a consequence of the following.

Theorem 2: Under the conditions of theorem 1,

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

4. PROOF OF THEOREM 2

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

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

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

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

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

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

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

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

Hence,

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

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.)

<u># of Terms</u>	<u>General Terms</u>
$O(n)$	$\text{var}(V_{ij}^2)$
$O(n^2)$	$\text{var}(V_{ij} V_{i'j'})$
$O(n^2)$	$\text{cov}(V_{ij}^2, V_{i'j'}^2)$
$O(n^2)$	$\text{cov}(V_{ij}^2, V_{ij} V_{i'j'})$
$O(n^3)$	$\text{cov}(V_{ij}^2, V_{i'j'} V_{i''j''})$
$O(n^3)$	$\text{cov}(V_{ij} V_{i'j'}, V_{ij} V_{i''j''})$
$O(n^4)$	$\text{cov}(V_{ij} V_{i'j'}, V_{i''j''} V_{i^*j^*})$

Bounds will now be computed for each general term.

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

$$\begin{aligned}
 \text{var}(V_{ij}^2) &\leq EV_{ij}^4 = \iint [h^{-d} K(\frac{x-y}{h}) - f(x)]^4 w(x)^2 f(x) f(y) dx dy = \\
 (4.5) \quad &= \iint [h^{-3d} K(u)^4 - 4h^{-2d} K(u)^3 f(x) + 6h^{-d} K(u)^2 f(x) - \\
 &\quad - 4K(u) f(x)^3 + h^d f(x)^4] w(x)^2 f(x) f(x-hu) du dx \\
 &= O(h^{-3d}) .
 \end{aligned}$$

Similarly,

$$\begin{aligned}
 \text{var}(V_{ij} V_{i'j'}) &\leq EV_{ij}^2 V_{i'j'}^2 = \\
 (4.6) \quad &= \iiint [h^{-d} K(\frac{x-y}{h}) - f(x)]^2 [h^{-d} K(\frac{x-z}{h}) - f(x)]^2 w(x)^2 f(x) f(y) f(z) dx dy dz = \\
 &= \int [\int [h^{-d} K(u)^2 - 2K(u) f(x) + h^d f(x)^2] f(x-hu) du]^2 f(x) dx = \\
 &= O(h^{-2d}) .
 \end{aligned}$$

Very similar computations show that

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

It follows from the above that

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

In the same way it is easily seen that

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

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

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

To bound the final term, note that

$$(4.8) \quad E[V_{ij} | X_j] = \int [h^{-d} K(\frac{X_j - y}{h}) - f(X_j)] w(X_j)^{1/2} f(y) dy$$

Hence, by (3.2)

$$(4.9) \quad \begin{aligned} E[V_{ij} V_{i',j}] &= E(E[V_{ij} | X_j]^2) = \\ &= \int [\int K(u) f(x+hu) du - f(x)]^2 w(x) f(x) dx = \\ &= s_f(h) . \end{aligned}$$

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

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

It follows from the above that

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

Hence,

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

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

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

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

$$\begin{aligned} (4.11) \quad n^{-1}\text{var}(U_j) &= \frac{O(n^{-1}h^{-d}) + O(s_f(h))}{n[Cn^{-1}h^{-d} + s_f(h) + o(n^{-1}h^{-d})]^2} \leq \frac{O(n^{-1}h^{-d})}{C^2 n^{-1}h^{-2d}} + \frac{O(s_f(h))}{2Ch^{-d}s_f(h)} \\ &= O(h^d) \rightarrow 0 . \end{aligned}$$

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

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

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):

<u># of Terms</u>	<u>General Term</u>
$O(1)$	$\text{cov}(V_{j,j}^2, V_{jj}^2)$
$O(n)$	$\text{cov}(V_{j,j}^2, V_{ij}^2)$
$O(n)$	$\text{cov}(V_{ij}^2, V_{ij}^2)$
$O(n^2)$	$\text{cov}(V_{ij}^2, V_{i'j}^2)$
$O(n)$	$\text{cov}(V_{j,j}^2, V_{jj}, V_{ij})$
$O(n^2)$	$\text{cov}(V_{j,j}^2, V_{ij}, V_{i'j})$
$O(n)$	$\text{cov}(V_{ij}^2, V_{ij}, V_{jj})$
$O(n^2)$	$\text{cov}(V_{ij}^2, V_{ij}, V_{i'j})$
$O(n^2)$	$\text{cov}(V_{ij}^2, V_{jj}, V_{i'j})$
$O(n^3)$	$\text{cov}(V_{ij}^2, V_{i'j}, V_{i''j})$
$O(n)$	$\text{cov}(V_{j,j}, V_{ij}, V_{jj}, V_{ij})$
$O(n^2)$	$\text{cov}(V_{j,j}, V_{ij}, V_{jj}, V_{i'j})$
$O(n^2)$	$\text{cov}(V_{j,j}, V_{ij}, V_{ij}, V_{i'j})$
$O(n^3)$	$\text{cov}(V_{j,j}, V_{ij}, V_{i'j}, V_{i''j})$
$O(n^2)$	$\text{cov}(V_{ij}, V_{i'j}, V_{ij}, V_{i'j})$
$O(n^3)$	$\text{cov}(V_{ij}, V_{i'j}, V_{ij}, V_{i''j})$
$O(n^4)$	$\text{cov}(V_{ij}, V_{i'j}, V_{i''j}, V_{i^*j})$

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

By the independence of X_1, \dots, X_n ,

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

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)

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

Thus, by (3.3) and (4.9),

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

Similarly, again using the Schwartz Inequality,

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

from which it follows that

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

By the above techniques it is easy to see that

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

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

$$\begin{aligned} E[V_{ij}^2 V_{ij}, V_{i',j},] &= E(V_{ij}^2 V_{ij}, E[V_{i',j}, |X_i, X_j, X_j,]) = \\ &= \iiint [h^{-d} K(\frac{y-x}{h}) - f(y)]^2 [h^{-d} K(\frac{z-x}{h}) - f(z)] \cdot \\ &\quad \cdot w(y)w(z)^{\frac{1}{2}} E[V_{ij} |X_j=z] f(x) f(y) f(z) dx dy dz = \\ &= \iiint h^{-d} [K(u) - h^d f(x+hu)]^2 [K(v) - h^d f(x+hv)] \cdot \\ &\quad \cdot w(x+hu)w(x+hv)^{\frac{1}{2}} E[V_{ij} |X_j = x+hv] f(x) f(x+hu) f(x+hv) dx dudv \\ &= O(h^{-d}), \end{aligned}$$

and hence,

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

Similarly,

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

and so,

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

By the same technique,

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

and hence,

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

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

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

Note that

$$\begin{aligned} (4.13) \quad E(V_{ij} V_{i,j} V_{ij}, V_{i,j'}) &= E(E[V_{ij} V_{i,j} | X_j, X_{j'}]^2) = \\ &= \iint \{ \int [h^{-d} K(\frac{x-z}{h}) - f(x)] [h^{-d} K(\frac{y-z}{h}) - f(y)] f(z) dz \}^2 w(x)w(y) f(x) f(y) dx dy = \\ &= I + I' , \end{aligned}$$

where

$$\begin{aligned} I &= \iint_A \{ \int [h^{-d} K(\frac{x-z}{h}) - f(x)] [h^{-d} K(\frac{y-z}{h}) - f(y)] f(z) dz \}^2 w(x)w(y) f(x) f(y) dx dy, \\ I' &= \iint_{A^c} \{ \int [h^{-d} K(\frac{x-z}{h}) - f(x)] [h^{-d} K(\frac{y-z}{h}) - f(y)] f(z) dz \}^2 w(x)w(y) f(x) f(y) dx dy . \end{aligned}$$

Note that, uniformly over x, y ,

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

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

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

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 $A_3 = \mathbb{R}^d \setminus (A_1 \cup A_2)$. Define I_1, I_2 , and I_3 by

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

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,

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

Similarly,

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

It follows from the above that

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

Now from (4.13) and (4.14)

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

The above computations will prove useful in bounding the last term.

Using the Schwartz Inequality and the independence of X_j and $X_{j'}$,

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

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

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

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

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

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

$$\text{cov}(U_j, U_j,) \rightarrow 0 .$$

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

$$\text{var}(n^{-1} \sum_{j=1}^n U_j) \rightarrow 0 ,$$

and hence, by the Chebychev Inequality,

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

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

ACKNOWLEDGEMENT

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

REFERENCES

- Hall, P. (1982). Limit theorems for stochastic measures of the accuracy of nonparametric density estimators. Stoch. Processes Appln. 13, 11-25.
- Komlós, J., Major, P. and Tusnády (1975). An approximation of partial sums of independent random variables, and the sample distribution function. Z. Wahrsch. Verw. Geb. 32, 111-131 .
- Marron, J.S. (1983). An asymptotically efficient solution to the bandwidth problem of kernel density estimation. North Carolina Institute of Statistics, Mimeo Series #1518.
- Wegman, E.J. (1972). Nonparametric density estimation. II. A comparison of density estimation methods. J. Statist. Comp. Simul., 1, 225-245.