

Central Limit Theorem for Integrated Square Error of Multivariate Nonparametric Density Estimators

PETER HALL

Australian National University, Canberra, Australia

Communicated by P. Révész

Martingale theory is used to obtain a central limit theorem for degenerate U -statistics with variable kernels, which is applied to derive central limit theorems for the integrated square error of multivariate nonparametric density estimators. Previous approaches to this problem have employed Komlós–Major–Tusnády type approximations to the empiric distribution function, and have required the following two restrictive assumptions which are not necessary using the present approach: (i) the data are in one or two dimensions, and (ii) the estimator is constructed suboptimally.

1. INTRODUCTION

Let f_n be a nonparametric estimator of the p -variate density f based on a sample of size n . The most widely accepted measure of the global performance of f_n is its integrated square error (ISE),

$$I_n \equiv \int \{f_n(x) - f(x)\}^2 dx.$$

Indeed, it is often suggested that f_n be constructed to minimize mean integrated square error (MISE), defined by

$$E(I_n) = \int E\{f_n(x) - f(x)\}^2 dx,$$

in an asymptotic sense. It has been shown [7] that in the case $p = 1$, and when f_n is a Rosenblatt–Parzen kernel estimator, ISE and MISE are asymptotically equivalent, in the sense that $I_n/E(I_n) \rightarrow 1$ in probability as $n \rightarrow \infty$. Several authors [1, 4, pp. 228–229, 13] have sharpened this result to

Received November 9, 1982; revised January 21, 1983.

AMS 1980 subject classifications: Primary 62G05, Secondary 60F05, 60G42.

Key words and phrases: central limit theorem, integrated square error, Martingale, nonparametric density estimator, U -statistic.

a central limit theorem in certain cases. However, the latter results are not sufficiently general to include the case of an optimally constructed density estimator, and the general multivariate case has not been treated. Our aim in the present paper is to develop a completely new method of proof which rectifies these deficiencies.

Integrated square error is often used in simulation studies to measure the performance of density estimators. It is also used implicitly in adaptive constructions of estimators, when the aim is to minimise mean integrated square error in some sense; see, for example, [2, 8]. Both these applications involve the assumption that integrated square error is somehow “close” to mean integrated square error. Our central limit theorems for ISE provide an explicit description of the order of this “closeness,” by showing that

$$d(n)\{I_n - E(I_n)\} \rightarrow N(0, 1),$$

in distribution as $n \rightarrow \infty$, where $d(n)$, $n \geq 1$, is a sequence of positive constants diverging to infinity. If the estimator is of the Rosenblatt–Parzen type, and is constructed to minimise MISE, then the constants $d(n)$ may be taken equal to $dn^{(p+8)/2(p+4)}$ for a positive constant d , where p equals the number of dimensions; see Theorem 2 in Section 4.

The techniques used in [1, 4, 7, 13] are similar, in that they employ Komlós–Major–Tusnády [10] type approximations to the empiric distribution function. This approach overcomes many of the analytic difficulties which are encountered in a more direct assault on the problem, but it has two major drawbacks: (i) It does not generalise easily to the multivariate case, since the best available multivariate versions of the Komlós–Major–Tusnády approximation are not sufficiently informative. (ii) In the case of the central-limit theorem, it forces rather severe restrictions on the method of construction of the density estimator (such as, that it be suboptimal) and on the unknown density (such as, that it vanish outside a compact interval, or be positive everywhere). The approach adopted in this paper is very different. We show that the problem can be set up in the context of degenerate U -statistics, and apply Martingale theory to derive a central limit theorem for degenerate U -statistics with variable kernels. The desired central limit theorem for I_n may be deduced from this result and the Lindeberg–Feller theorem for sums of independent random variables.

The necessary theory for degenerate U -statistics is developed in Section 2. In Section 3 we provide a decomposition of I_n into a U -statistic plus several other terms, and in Section 4 we apply the results of Section 2 to the U -statistic. Section 5 contains proofs of some of the results in Section 4. The case of a Rosenblatt–Parzen estimator is treated in greatest detail, although it is shown that the U -statistic approach may be used very widely.

2. CENTRAL LIMIT THEOREM FOR DEGENERATE U -STATISTICS

A simple one-sample U -statistic is a random variable of the form

$$U_n = \sum_{1 \leq i < j \leq n} H(X_i, X_j),$$

where H is a symmetric function and X_1, \dots, X_n are independent and identically distributed random variables (or vectors). We may assume without loss of generality that U_n has been centred, so that $E\{H(X_1, X_2)\} = 0$. In this case, the U -statistic is said to be *degenerate* if $E\{H(X_1, X_2)|X_1\} = 0$, almost surely.

Define

$$Y_i = \sum_{j=1}^{i-1} H(X_i, X_j), \quad 2 \leq i \leq n,$$

and note that in the case of a centred degenerate U -statistic, $E(Y_i|X_1, \dots, X_{i-1}) = 0$, almost surely. Therefore the sequence $\{S_i = \sum_{j=2}^i Y_j, 2 \leq i \leq n\}$ is a Martingale, in which $S_n = U_n$.

Limit theory for degenerate U -statistics when H is fixed has been worked out by Gregory [5], Neuhaus [12], Hall [6], and Weber [14]. In that case, the limit distribution is a linear combination of independent, centred χ_1^2 distributions, and cannot be derived using classical Martingale methods. However, in certain cases in which $H (=H_n)$ depends on n , a normal distribution can result. Roughly speaking, this occurs when the eigenvalues v_{nr} , $r \geq 1$, of the linear operator \mathcal{H}_n on the space of square integrable measurable functions $\alpha: \mathbb{R} \rightarrow \mathbb{R}$, defined by

$$(\mathcal{H}_n \alpha)(x) = E\{H_n(X_1, x) \alpha(X_1)\},$$

satisfy an ‘‘asymptotic negligibility’’ condition, such as

$$\lim_{n \rightarrow \infty} \left(\sum_{r=1}^{\infty} |v_{nr}|^t \right)^{2/t} \left/ \left(\sum_{r=1}^{\infty} v_{nr}^2 \right) \right. = 0,$$

for some $t > 2$. Unfortunately, this type of condition can be rather difficult to check in practice, since the eigenvalues are seldom known with any precision. We shall use Martingale theory to derive a central limit theorem under more practicable conditions. Define

$$G_n(x, y) = E\{H_n(X_1, x) H_n(X_1, y)\}.$$

THEOREM 1. *Assume H_n is symmetric, $E\{H_n(X_1, X_2)|X_1\} = 0$ almost surely and $E\{H_n^2(X_1, X_2)\} < \infty$ for each n . If*

$$[E\{G_n^2(X_1, X_2)\} + n^{-1}E\{H_n^4(X_1, X_2)\}]/[E\{H_n^2(X_1, X_2)\}]^2 \rightarrow 0, \quad (2.1)$$

as $n \rightarrow \infty$, then $U_n \equiv \sum_{i=1}^n \sum_{1 < j < n} H_n(X_i, X_j)$ is asymptotically normally distributed with zero mean and variance given by $\frac{1}{2}n^2 E\{H_n^2(X_1, X_2)\}$.

Our proof involves checking two conditions (2.2) and (2.3), which are sufficient for an invariance principle as well as a central limit theorem. We have chosen to state the simpler result since it is more closely related to the applications in the next section. However, it is possible to adapt the argument below and derive an invariance principle under even more general conditions than (2.1). We have settled on condition (2.1) as a compromise between generality and simplicity.

Note that the first part of condition (2.1),

$$E\{G_n^2(X_1, X_2)\} / [E\{H_n^2(X_1, X_2)\}]^2 \rightarrow 0,$$

is equivalent to

$$\left(\sum_{r=1}^{\infty} v_{nr}^4 \right)^{1/2} \left/ \left(\sum_{r=1}^{\infty} v_{nr}^2 \right) \right. \rightarrow 0.$$

Proof of Theorem 1. We shall apply Brown's [3] Martingale central limit theorem; see Hall and Heyde [9, Corollary 3.1, p. 58]. This requires us to check two conditions:

$$s_n^{-2} \sum_{i=2}^n E\{Y_{ni}^2 I(|Y_{ni}| > \varepsilon s_n)\} \rightarrow 0, \quad (2.2)$$

as $n \rightarrow \infty$ for each $\varepsilon > 0$, where $Y_{ni} = \sum_{j=1}^{i-1} H_n(X_i, X_j)$ and $s_n^2 = E(U_n^2)$, and

$$s_n^{-2} V_n^2 \rightarrow 1 \quad \text{in probability}, \quad (2.3)$$

as $n \rightarrow \infty$, where $V_n^2 = \sum_{i=2}^n E(Y_{ni}^2 | X_1, \dots, X_{i-1})$. From (2.2) and (2.3) it follows that $s_n^{-1} U_n$ is asymptotically normal $N(0, 1)$. Since

$$\begin{aligned} E(Y_{ni}^2) &= \sum_{j=1}^{i-1} \sum_{k=1}^{i-1} E\{H_n(X_i, X_j) H_n(X_i, X_k)\} \\ &= \sum_{j=1}^{i-1} E\{H_n^2(X_i, X_j)\} = (i-1) E\{H_n^2(X_1, X_2)\}, \end{aligned}$$

then $s_n^2 = \sum_{i=2}^n E(Y_{ni}^2) = \frac{1}{2}n(n-1) E\{H_n^2(X_1, X_2)\}$. Furthermore,

$$\begin{aligned} &E\{H_n(X_1, X_2) H_n(X_1, X_3) H_n(X_1, X_4) H_n(X_1, X_5)\} \\ &= E\{H_n(X_1, X_2) H_n^3(X_1, X_3)\} = 0, \end{aligned}$$

and so

$$\begin{aligned} E(Y_{ni}^4) &= \sum_{j=1}^{i-1} E\{H_n^4(X_i, X_j)\} + 3 \sum_{1 \leq j, k \leq i-1; j \neq k} E\{H_n^2(X_i, X_j) H_n^2(X_i, X_k)\} \\ &= (i-1) E\{H_n^4(X_1, X_2)\} + 3(i-1)(i-2) E\{H_n^2(X_1, X_2) H_n^2(X_1, X_3)\}, \end{aligned}$$

whence

$$\begin{aligned} \sum_{i=2}^n E(Y_{ni}^4) &\leq \text{const} [n^2 E\{H_n^4(X_1, X_2)\} + n^3 E\{H_n^2(X_1, X_2) H_n^2(X_1, X_3)\}] \\ &\leq \text{const. } n^3 E\{H_n^4(X_1, X_2)\}. \end{aligned}$$

It now follows from condition (2.1) that

$$s_n^{-4} \sum_{i=2}^n E(Y_{ni}^4) \rightarrow 0,$$

as $n \rightarrow \infty$, which implies (2.2). Observe that

$$\begin{aligned} v_{ni} &\equiv E(Y_{ni}^2 | X_1, \dots, X_{i-1}) = \sum_{j=1}^{i-1} \sum_{k=1}^{i-1} G_n(X_j, X_k) \\ &= 2 \sum_{1 \leq j < k \leq i-1} G_n(X_j, X_k) + \sum_{j=1}^{i-1} G_n(X_j, X_j). \end{aligned}$$

If $j_1 \leq k_1$ and $j_2 \leq k_2$ then

$$\begin{aligned} &E\{G_n(X_{j_1}, X_{k_1}) G_n(X_{j_2}, X_{k_2})\} \\ &= E\{G_n^2(X_1, X_1)\} \quad \text{if } j_1 = k_1 = j_2 = k_2, \\ &= [E\{G_n(X_1, X_1)\}]^2 \quad \text{if } j_1 = k_1 \neq j_2 = k_2, \\ &= E\{G_n^2(X_1, X_2)\} \quad \text{if } j_1 = j_2, k_1 = k_2, j_1 < k_1, \\ &= 0 \quad \text{otherwise.} \end{aligned}$$

Therefore if $i_1 \leq i_2$,

$$\begin{aligned} E(v_{ni_1} v_{ni_2}) &= 4 \sum_{1 \leq j < k \leq i_1-1} E\{G_n^2(X_1, X_2)\} + \sum_{j=1}^{i_1-1} \sum_{k=1}^{i_2-1} [E\{G_n(X_1, X_1)\}]^2 \\ &\quad + \sum_{j=1}^{i_1-1} (E\{G_n^2(X_1, X_1)\} - [E\{G_n(X_1, X_1)\}]^2) \\ &= 2(i_1-1)(i_1-2) E\{G_n^2(X_1, X_2)\} + (i_1-1)(i_2-1) [E\{G_n(X_1, X_1)\}]^2 \\ &\quad + (i_1-1) \text{var}\{G_n(X_1, X_1)\}. \end{aligned}$$

Consequently

$$\begin{aligned}
 E(V_n^4) &= 2 \sum_{2 \leq i < j \leq n} \sum_{i=2}^n E(v_{ni}v_{nj}) + \sum_{i=2}^n E(v_{ni}^2) \\
 &= 2 \left\{ \sum_{i=2}^n (i-1)(i-2)(2n-2i+1) \right\} E\{G_n^2(X_1, X_2)\} \\
 &\quad + \left\{ \sum_{i=2}^n (i-1)(2n-2i+1) \right\} \text{var}\{G_n(X_1, X_1)\} \\
 &\quad + \left[\frac{1}{2}n(n-1) E\{G_n(X_1, X_1)\} \right]^2,
 \end{aligned}$$

whence

$$\begin{aligned}
 E(V_n^2 - s_n^2)^2 &\leq \text{const}[n^4 E\{G_n^2(X_1, X_2)\} + n^3 E\{G_n^2(X_1, X_1)\}] \\
 &\leq \text{const}[n^4 E\{G_n^2(X_1, X_2)\} + n^3 E\{H_n^2(X_1, X_2)\}].
 \end{aligned}$$

It now follows from (2.1) that $s_n^{-4} E(V_n^2 - s_n^2)^2 \rightarrow 0$, which proves (2.3).

3. A *U*-STATISTIC ARISING IN DENSITY ESTIMATION

Let X_1, \dots, X_n be a random sample from a distribution with density f on \mathbb{R}^p . Most nonparametric estimators of f may be written in the form

$$f_n(x) = \sum_{i=1}^n K_n(x, X_i),$$

where K_n is a “kernel function.” For example, if we take

$$K_n(x, y) = (nh^p)^{-1} K\{(x-y)/h\}, \quad (3.1)$$

where K is a density in \mathbb{R}^p and $h = h(n)$ is a sequence of constants converging to zero as $n \rightarrow \infty$, we have the classical Rosenblatt–Parzen kernel estimator. The (weighted) integrated square error of the estimator f_n is given by

$$\begin{aligned}
 \int \{f_n(x) - f(x)\}^2 w(x) dx &= \int \{f_n(x) - Ef_n(x)\}^2 w(x) dx \\
 &\quad + 2 \int \{f_n(x) - Ef_n(x)\} \{Ef_n(x) - f(x)\} w(x) dx \\
 &\quad + \int \{Ef_n(x) - f(x)\}^2 w(x) dx, \quad (3.2)
 \end{aligned}$$

where w is the weight function and where an unqualified integral denotes integration over \mathbb{R}^p . The last term on the right-hand side in (3.2) is purely deterministic in character, and can be analysed by routine analytic methods. The second last term can be written as a sum of independent and identically distributed random variables, and so is readily described by a central limit theorem. The first term may be expressed as

$$\begin{aligned} & \int \{f_n(x) - Ef_n(x)\}^2 w(x) dx \\ &= 2 \sum_{1 \leq i < j \leq n} \int \{K_n(x, X_i) - EK_n(x, X_i)\} \{K_n(x, X_j) - EK_n(x, X_j)\} w(x) dx \\ &+ \sum_{i=1}^n \int \{K_n(x, X_i) - EK_n(x, X_i)\}^2 w(x) dx. \end{aligned} \quad (3.3)$$

The last term on the right-hand side of (3.3), being a sum of independent random variables, is very easily described by a central limit theorem, while the first term equals twice a centred, degenerate U -statistic whose variable kernel function is given by

$$\tilde{H}_n(x, y) = \int \{K_n(u, x) - EK_n(u, X_1)\} \{K_n(u, y) - EK_n(u, X_1)\} w(u) du.$$

Thus, central limit theorems for degenerate U -statistics with variable kernels, and for triangular arrays of sums of independent random variables, are basic tools for proving a central limit theorem for the difference between integrated square error and mean integrated square error. We shall illustrate this point by considering the most important case of a Rosenblatt–Parzen kernel estimator, taking $w(x) \equiv 1$. Many other cases may be treated similarly.

4. CENTRAL LIMIT THEOREM FOR INTEGRATED SQUARE ERROR

Henceforth we assume that the kernel function is given by (3.1), and that $h \rightarrow 0$ and $nh^p \rightarrow \infty$ as $n \rightarrow \infty$. (The latter condition is necessary for the mean square consistency of the estimator.) We assume throughout the following conditions on K and f , referred to below as the “stated conditions”:

K is a bounded, nonnegative function on \mathbb{R}^p satisfying

$$\int K(z) dz = 1, \quad \int z_i K(z) dz = 0, \quad \text{and} \quad \int z_i z_j K(z) dz = 2k\delta_{ij} < \infty,$$

for each i , where k does not depend on i ; and f and its second order partial derivatives are bounded and uniformly continuous on \mathbb{R}^p .

We shall consider the terms in the expansion (3.2) individually, via a sequence of lemmas which are proved in Section 5.

(i) $I_{n1} \equiv \int \{f_n(x) - Ef_n(x)\} \{Ef_n(x) - f(x)\} dx$. Observe that $I_{n1} = (nh^p)^{-1} \sum_{i=1}^n Z_{n1i}$, where

$$Z_{n1i} = \int [K\{(x - X_i)/h\} - EK\{(x - X_i)/h\}] \{Ef_n(x) - f(x)\} dx.$$

LEMMA 1. Under the stated conditions on K and f , $E(Z_{n1i}) = 0$,

$$E(Z_{n1i}^2) \sim h^{2p+4} k^2 \left[\int \{\nabla^2 f(x)\}^2 f(x) dx - \left\{ \int (\nabla^2 f(x)) f(x) dx \right\}^2 \right],$$

and

$$E(Z_{n1i}^4) = O(h^{4p+8}), \quad (4.1)$$

as $h \rightarrow 0$, where ∇^2 is the Laplacian.

Let $s_n^2 = \sum_{i=1}^n E(Z_{n1i}^2)$. It follows from (4.1) that

$$s_n^{-2} \sum_{i=1}^n E\{Z_{n1i}^2 I(|Z_{n1i}| > \varepsilon s_n)\} \leq \varepsilon^{-2} s_n^{-4} \sum_{i=1}^n E(Z_{n1i}^4) \rightarrow 0,$$

as $n \rightarrow \infty$, and so I_{n1} is asymptotically normally distributed with zero mean and variance given by $\sigma_{n1}^2 = n^{-1} h^4 k^2 \sigma_1^2$, where

$$\sigma_1^2 = \int \{\nabla^2 f(x)\}^2 f(x) dx - \left[\int \{\nabla^2 f(x)\} f(x) dx \right]^2.$$

(ii) $I_{n2} \equiv (nh^p)^{-2} \sum_{i=1}^n \int [K\{(x - X_i)/h\} - EK\{(x - X_i)/h\}]^2 dx$. Let Z_{n2i} denote the i th term in this series.

LEMMA 2. Under the stated conditions on K and f ,

$$E(Z_{n2i}) = h^p \int K^2(z) dz - h^{2p} \iint K(u) K(u+v) du dv \int f(x) f(x+hv) dx \quad (4.2)$$

and

$$E(Z_{n2i}^2) = O(h^{2p}), \quad (4.3)$$

as $h \rightarrow 0$.

Let $\sigma_{n2}^2 = (nh^{2p})^{-1} E(Z_{n2i})$. By (4.3), $\text{var}(I_{n2}) = O\{(n^3 h^{2p})^{-1}\}$, and so

$$I_{n2} = \sigma_{n2}^2 + O_p(n^{-3/2} h^{-p}),$$

as $n \rightarrow \infty$.

(iii) $I_{n3} \equiv 2(nh^p)^{-2} U_n$, where $U_n = \sum \sum_{1 \leq i < j \leq n} H_n(X_i, X_j)$, $H_n(x, y) = \int A_n(u, x) A_n(u, y) du$, and $A_n(u, x) = K\{(u-x)/h\} - E[K\{(u-X_1)/h\}]$. Define the function G_n in terms of H_n as in Section 2. The next lemma enables us to check the conditions of Theorem 1 for the U -statistic U_n .

LEMMA 3. *Under the stated conditions on K and f ,*

$$E\{H_n^2(X_1, X_2)\} \sim h^{3p} \left\{ \int f^2(x) dx \left[\int \int K(u) K(u+v) du \right]^2 dv \right\}, \quad (4.4)$$

$$E\{H_n^4(X_1, X_2)\} = O(h^{5p}), \quad (4.5)$$

and

$$E\{G_n^2(X_1, X_2)\} = O(h^{7p}), \quad (4.6)$$

as $h \rightarrow 0$.

Since $nh^p \rightarrow 0$ as $n \rightarrow \infty$, it follows from Theorem 1 and Lemma 3 that U_n is asymptotically normally distributed with zero mean and variance equal to $\frac{1}{2} n^2 E\{H_n^2(X_1, X_2)\}$. Therefore I_{n3} is asymptotically normal $N(0, \sigma_{n3}^2)$, where $\sigma_{n3}^2 = 2n^{-2} h^{-p} \sigma_3^2$ and

$$\sigma_3^2 = \left\{ \int f^2(x) dx \left[\int \int K(u) K(u+v) du \right]^2 dv \right\}.$$

(iv) In this step we combine the results from steps (i)–(iii). Observe that

$$\begin{aligned} A_n &\equiv \int \{f_n(x) - f(x)\}^2 dx - \int \{E f_n(x) - f(x)\}^2 dx \\ &= 2I_{n1} + I_{n2} + I_{n3}. \end{aligned} \quad (4.7)$$

From (ii), (iii), and the fact that $nh^p \rightarrow \infty$, we see that the error about the mean of I_{n2} , viz. $I_{n2} - \sigma_{n2}^2$, is asymptotically negligible in comparison with I_{n3} . Therefore (4.7) may be written as

$$A_n = 2k\sigma_1 n^{-1/2} h^2 N_{n1} + \sigma_{n2}^2 + 2^{1/2} \sigma_3 n^{-1} h^{-(1/2)p} N_{n3}, \quad (4.8)$$

where the random variables N_{n1} and N_{n3} are each asymptotically normal $N(0, 1)$.

If $nh^{p+4} \rightarrow \infty$ then the last term on the right-hand side in (4.8) is asymptotically negligible in comparison with the first, while if $nh^{p+4} \rightarrow 0$, the first term is negligible in comparison with the last. The case where $nh^{p+4} \rightarrow \lambda$, $0 < \lambda < \infty$, is also of interest, since this corresponds to choosing h to minimise mean integrated square error,

$$\text{MISE} = \int E\{f_n(x) - f(x)\}^2 dx.$$

See, for example, Mack and Rosenblatt [11]. In that situation,

$$\begin{aligned} \Delta_n = & n^{-(p+8)/2(p+4)}(2k\sigma_1\lambda^{2/(p+4)}N_{n_1} + 2^{1/2}\sigma_3\lambda^{-p/2(p+4)}N_{n_3}) + \sigma_{n_2}^2 \\ & + o_p(n^{-(p+8)/2(p+4)}). \end{aligned} \quad (4.9)$$

Now, the variables N_{n_1} and N_{n_3} are principally derived from the terms I_{n_1} and I_{n_3} , which are easily seen to be uncorrelated. Furthermore, for any real numbers a and b , the quantity

$$a(\text{var } I_{n_1})^{-1/2} I_{n_1} + b(\text{var } I_{n_2})^{-1/2} I_{n_3},$$

can be written as the n th partial sum of a Martingale difference array, and so (using techniques from the proof of Theorem 1) can be proved to be asymptotically normal $N(0, a^2 + b^2)$. It now follows via the Cramér–Wold device that the variables I_{n_1} and I_{n_3} are asymptotically independent and normally distributed. Therefore we may rewrite the expansion (4.9) as

$$\Delta_n = n^{-(p+8)/2(p+4)}(4k^2\sigma_1^2\lambda^{4/(p+4)} + 2\sigma_3^2\lambda^{-p/(p+4)})^{1/2}N_{n_4} + \sigma_{n_2}^2,$$

where N_{n_4} is asymptotically normal $N(0, 1)$.

The results we have just derived are collected together in the following theorem. In order to explain the appearance of the constant term $c(n)$ below, note that

$$\int E(f_n - f)^2 = \int (Ef_n - f)^2 + \sigma_{n_2}^2.$$

THEOREM 2. *Let*

$$\begin{aligned} c(n) = & \int E\{f_n(x) - f(x)\}^2 dx \\ = & \int \{Ef_n(x) - f(x)\}^2 dx + (nh^p)^{-1} \left[\int K^2(z) dz \right. \\ & \left. - h^p \iint K(u)K(u+v) du dv \int f(x)f(x+hv) dx \right], \end{aligned}$$

and define

$$\begin{aligned} d(n) &= n^{1/2} h^{-2} && \text{if } nh^{p+4} \rightarrow \infty, \\ &= nh^{(1/2)p} && \text{if } nh^{p+4} \rightarrow 0, \\ &= n^{(p+8)/2(p+4)} && \text{if } nh^{p+4} \rightarrow \lambda, 0 < \lambda < \infty. \end{aligned}$$

Under the stated conditions on K and f , and assuming that $h \rightarrow 0$ and $nh^p \rightarrow \infty$, we have

$$\begin{aligned} d(n) \left[\int \{f_n(x) - f(x)\}^2 dx - c(n) \right] &\rightarrow 2k\sigma_1 Z && \text{if } nh^{p+4} \rightarrow \infty, \\ &\rightarrow 2^{1/2}\sigma_3 Z && \text{if } nh^{p+4} \rightarrow 0, \\ &\rightarrow (4k^2\sigma_1^2\lambda^{4/(p+4)} + 2\sigma_3^2\lambda^{-p/(p+4)})^{1/2} Z, && \\ &&& \text{if } nh^{p+4} \rightarrow \lambda, 0 < \lambda < \infty, \end{aligned}$$

in distribution as $n \rightarrow \infty$, where Z has the standard normal distribution.

We may replace $c(n)$ by

$$c^*(n) = \int \{Ef_n(x) - f(x)\}^2 dx + (nh^p)^{-1} \int K^2(z) dz,$$

in Theorem 2, without affecting the asymptotics. In general, $c(n)$ (and $c^*(n)$) depend on the unknown f in a nontrivial way. This was not the case in some earlier results; see, for example, [4, 13].

If we assume in addition that $\nabla^2 f$ is square integrable then

$$\int \{Ef_n(x) - f(x)\}^2 dx = h^4 k^2 \int \{\nabla^2 f(x)\}^2 dx + o(h^4),$$

and so

$$\begin{aligned} \int \{f_n(x) - f(x)\}^2 dx &= (nh^p)^{-1} \int K^2(z) dz + h^4 k^2 \int \{\nabla^2 f(x)\}^2 dx \\ &\quad + o_p\{(nh^p)^{-1} + h^4\}. \end{aligned} \tag{4.10}$$

This result may be interpreted as a “weak law of large numbers” corresponding to the central limit theorem of Theorem 2. It extends earlier work due to Hall [7, Theorem 2] in two ways: it places only the minimal conditions $h \rightarrow 0$ and $nh^p \rightarrow \infty$ on the window size h , and more importantly, it treats the case of a general $p \geq 1$.

5. PROOFS OF LEMMAS

Proof of Lemma 1. Note that $Z_{n1i} = Y_{n1i} - EY_{n1i}$, where

$$Y_{n1i} = \int K\{(x - X_i)/h\} \{Ef_n(x) - f(x)\} dx.$$

Define $t_n^{(j)} = E(Y_{n1i}^j)$ for positive integers j . Since $E\{f_n(x)\} - f(x) = h^2 k \nabla^2 f(x) + o(h^2)$ uniformly in x , then

$$\begin{aligned} t_n^{(1)} &= h^p \int \{Ef_n(x) - f(x)\} dx \int K(z) f(x - zh) dz \\ &= h^{p+2} k \int \nabla^2 f(x) dx \int K(z) f(x - zh) dz + o(h^{p+2}) \\ &= h^{p+2} k \int \{\nabla^2 f(x)\} f(x) dx + o(h^{p+2}), \\ t_n^{(2)} &= h^{2p} \iint \{Ef_n(x) - f(x)\} \{Ef_n(x + uh) - f(x + uh)\} dx du \\ &\quad \times \int K(z) K(z + u) f(x - zh) dz \\ &= h^{2p+4} k^2 \iint \nabla^2 f(x) \nabla^2 f(x + uh) dx du \\ &\quad \times \int K(z) K(z + u) f(x - zh) dz + o(h^{2p+4}) \\ &= h^{2p+4} k^2 \int \{\nabla^2 f(x)\}^2 f(x) dx + o(h^{2p+4}), \end{aligned}$$

and for any $k \geq 1$,

$$\begin{aligned} |t_n^{(k)}| &\leq \int \dots \int \left\{ \prod_{j=1}^k |Ef_n(x^{(j)}) - f(x^{(j)})| dx^{(1)} \dots dx^{(k)} \right. \\ &\quad \times \int \left[\prod_{j=1}^k K\{(x^{(j)} - z)/h\} \right] f(z) dz \\ &\leq C_k h^{2k} \int \dots \int dx^{(1)} \dots dx^{(k)} \int \left[\prod_{j=1}^k K\{(x^{(j)} - z)/h\} \right] f(z) dz \\ &= C_k h^{k(p+2)}. \end{aligned}$$

The desired results follow immediately, on noting that $E(Z_{n1i}^2) = t_n^{(2)} - (t_n^{(1)})^2$ and $E(Z_{n1i}^4) = t_n^{(4)} - 4t_n^{(3)}t_n^{(1)} + 6t_n^{(2)}(t_n^{(1)})^2 - 3(t_n^{(1)})^4$.

Proof of Lemma 2. The result (4.2) follows on noting that

$$\int E[K^2\{(x - X_1)/h\}] dx = h^p \int K^2(z) dz, \quad (5.1)$$

and that $\int [EK\{(x - X_1)/h\}]^2 dx$ equals the second term on the right-hand side in (4.2). To prove (4.3), observe that for positive constants C_1 and C_2 ,

$$\begin{aligned} E(Z_{n2l}^2) &= \iint E\{[K\{(x - X_1)/h\} - EK\{(x - X_1)/h\}]^2 \\ &\quad \times [K\{(y - X_1)/h\} - EK\{(y - X_1)/h\}]^2\} dx dy \\ &\leq C_1 \iint (E[K^2\{(x - X_1)/h\}] K^2\{(y - X_1)/h\} \\ &\quad + [EK\{(x - X_1)/h\}]^2 E[K^2\{(y - X_1)/h\}] \\ &\quad + E[K\{(x - X_1)/h\}] E[K\{(x - X_1)/h\}] K^2\{(y - X_1)/h\}] \\ &\quad + [EK\{(x - X_1)/h\}]^2 [EK\{(y - X_1)/h\}]^2) dx dy \\ &\leq C_2 \iint (E[K^2\{(x - X_1)/h\}] K^2\{(y - X_1)/h\}] \\ &\quad + E[K\{(x - X_1)/h\}] K^2\{(y - X_1)/h\}) dx dy \\ &\quad + C_2 \left(\int E[K^2\{(x - X_1)/h\}] dx \right)^2. \end{aligned} \quad (5.2)$$

Now, for any $r \geq 1$ and $s \geq 1$,

$$\begin{aligned} &\iint E[K^r\{(x - X_1)/h\}] K^s\{(y - X_1)/h\}] dx dy \\ &= h^{2p} \iint K^r(u) K^s(u + v) du dv = O(h^{2p}). \end{aligned} \quad (5.3)$$

The result (4.3) follows on combining (5.1), (5.2), and (5.3).

Proof of Lemma 3. The result (4.4) follows from the identities

$$\begin{aligned} E\{H_n^2(X_1, X_2)\} &= \iint [E\{A_n(x, X_1) A_n(y, X_1)\}]^2 dx dy \\ &= h^{3p} \iint \left[\int K(z) K(z + u) f(x - zh) dz \right. \\ &\quad \left. - h^p \left\{ \int K(z) f(x - zh) dz \right\} \right. \\ &\quad \left. \times \left\{ \int K(z) f(x + uh - zh) dz \right\} \right]^2 dx du \end{aligned}$$

$$\begin{aligned} &\sim h^{3p} \iint \left[\int K(z) K(z+u) f(x-zh) dz \right]^2 dx du \\ &\sim h^{3p} \left\{ \int f^2(x) dx \right\} \left[\int \int K(z) K(z+u) dz \right]^2 du. \end{aligned}$$

Next observe that

$$E\{H_n^4(X_1, X_2)\} = \iiint \left[E \left\{ \prod_{i=1}^4 A_n(x^{(i)}, X_1) \right\} \right]^2 dx^{(1)} \dots dx^{(4)}.$$

The integrand may be expanded into several terms, and each of these shown to be of order h^{5p} . We shall illustrate the procedure in the case of the first of these terms,

$$\begin{aligned} T_{n1} &= \iiint \left[E \left\{ \prod_{i=1}^4 K((x^{(i)} - X_1)/h) \right\} \right]^4 dx^{(1)} \dots dx^{(4)} \\ &= h^{2p} \iiint \left[\int K(v) \left\{ \prod_{i=2}^4 K(v + (x^{(i)} - x^{(1)})/h) \right\} f(x^{(1)} - vh) dv \right]^2 \\ &\quad \times dx^{(1)} \dots dx^{(4)} \\ &= h^{5p} \iiint \left[\int K(v) \left\{ \prod_{i=2}^4 K(v + w^{(i)}) \right\} f(x^{(1)} - vh) dv \right]^2 \\ &\quad \times dx^{(1)} dw^{(2)} dw^{(3)} dw^{(4)} \\ &\leq h^{5p} \iiint dx^{(1)} dw^{(2)} dw^{(3)} dw^{(4)} \\ &\quad \times \int K(v) \left\{ \prod_{i=2}^4 K^2(v + w^{(i)}) \right\} f^2(x^{(1)} - vh) dv \\ &= h^{5p} \left\{ \int K^2(v) dv \right\}^3. \end{aligned}$$

This proves (4.5).

Let

$$\begin{aligned} B_n(x, y) &= E\{A_n(x, X_1) A_n(y, X_1)\} \\ &= E\{K\{(x - X_1)/h\} K\{(y - X_1)/h\}\} - E\{K\{(x - X_1)/h\}\} \\ &\quad \times E\{K\{(y - X_1)/h\}\}. \end{aligned} \tag{5.4}$$

Then

$$\begin{aligned} E\{G_n^2(X_1, X_2)\} &= \iiint B_n(u^{(1)}, u^{(2)}) B_n(v^{(1)}, v^{(2)}) B_n(u^{(1)}, v^{(1)}) \\ &\quad \times B_n(u^{(2)}, v^{(2)}) du^{(1)} du^{(2)} dv^{(1)} dv^{(2)}. \end{aligned}$$

Using the formula (5.4) for B_n , this quadruple integral may be expanded into several terms, each of which is of order h^{7p} . We treat only the first such term,

$$\begin{aligned} T_{n2} &= \iiint\!\!\!\int D_n(u^{(1)}, u^{(2)}, v^{(1)}, v^{(2)}) du^{(1)} du^{(2)} dv^{(1)} dv^{(2)} \\ &= h^{3p} \iiint\!\!\!\int D_n(u^{(1)}, u^{(1)} + a^{(1)}h, u^{(1)} + a^{(2)}h - a^{(3)}h, u^{(1)} - a^{(3)}h) \\ &\quad \times du^{(1)} da^{(1)} da^{(2)} da^{(3)}, \end{aligned}$$

where

$$\begin{aligned} D_n(u^{(1)}, u^{(2)}, v^{(1)}, v^{(2)}) &= E[K\{(u^{(1)} - X_1)/h\} K\{(u^{(2)} - X_1)/h\}] E[K\{(v^{(1)} - X_1)/h\} \\ &\quad \times K\{(v^{(2)} - X_1)/h\}] \\ &\quad \times E[K\{(u^{(1)} - X_1)/h\} K\{(v^{(1)} - X_1)/h\}] E[K\{(u^{(2)} - X_1)/h\} \\ &\quad \times K\{(v^{(2)} - X_1)/h\}] \\ &= h^{4p} \left[\int K(w) K\{w + (u^{(2)} - u^{(1)})/h\} f(u^{(1)} - wh) dw \right] \\ &\quad \times \left[\int K(w) K\{w + (v^{(1)} - v^{(2)})/h\} f(v^{(2)} - wh) dw \right] \\ &\quad \times \left[\int K(w) K\{w + (v^{(1)} - u^{(1)})/h\} f(u^{(1)} - wh) dw \right] \\ &\quad \times \left[\int K(w) K\{w + (u^{(2)} - v^{(2)})/h\} f(v^{(2)} - wh) dw \right]. \end{aligned}$$

Therefore

$$\begin{aligned} T_{n2} &= h^{7p} \iiint\!\!\!\int \left[\int K(w) K(w + a^{(1)}) f(u^{(1)} - wh) dw \right] \\ &\quad \times \left[\int K(w) K(w + a^{(2)}) f(u^{(1)} - a^{(3)}h - wh) dw \right] \\ &\quad \times \left[\int K(w) K(w + a^{(2)} - a^{(3)}) f(u^{(1)} - wh) dw \right] \\ &\quad \times \left[\int K(w) K(w + a^{(1)} + a^{(3)}) f(u^{(1)} - a^{(3)}h - wh) dw \right] \\ &\quad \times du^{(1)} da^{(1)} da^{(2)} da^{(3)} \\ &= O(h^{7p}), \end{aligned}$$

as $h \rightarrow 0$.

REFERENCES

- [1] BICKEL, P. J., AND ROSENBLATT, M. (1973). On some global measures of the deviations of density function estimates. *Ann. Statist.* **1** 1071–1095.
- [2] BOWMAN, A. W. (1982). A comparative study of some kernel-based nonparametric density estimates. Manchester–Sheffield School of Probability and Statistics Research Report 84/AWB/1.
- [3] BROWN, B. M. (1971). Martingale central limit theorems. *Ann. Math. Statist.* **42** 59–66.
- [4] CŠORGÖ, M., AND RÉVÉSZ, P. (1981). *Strong Approximations in Probability and Statistics*. Academic Press, New York/Budapest.
- [5] GREGORY, G. G. (1977). Large sample theory for U-statistics and tests of fit. *Ann. Statist.* **5** 110–123.
- [6] HALL, P. (1979). On the invariance principle for U-statistics. *Stochastic Process. Appl.* **9** 163–174.
- [7] HALL, P. (1982). Limit theorems for stochastic measures of the accuracy of density estimators. *Stochastic Process. Appl.* **13** 11–25.
- [8] HALL, P. (1982). Large sample optimality of least squares cross-validation in density estimation. To appear, *Ann. Statist.*
- [9] HALL, P., AND HEYDE, C. C. (1980). *Martingale Limit Theory and Its Application*. Academic Press, New York.
- [10] KOMLÓS, J., MAJOR, P., AND TUSNÁDY, G. (1975). An approximation of partial sums of independent r.v.'s and the sample d.f. I. *Z. Wahrsch. Verw. Gebiete* **32** 111–131.
- [11] MACK, Y. P., AND ROSENBLATT, M. (1979). Multivariate k-nearest neighbour density estimates. *J. Multivar. Anal.* **9** 1–15.
- [12] NEUHAUS, G. (1977). Functional limit theorems for U-statistics in the degenerate case. *J. Multivar. Anal.* **7** 424–439.
- [13] ROSENBLATT, M. (1975). A quadratic measure of deviation of two-dimensional density estimates and a test of independence. *Ann. Statist.* **3** 1–14.
- [14] WEBER, N. C. (1981). Incomplete degenerate U-statistics. *Scand. J. Statist.* **8** 120–123.