The Annals of Statistics
1983, Vol. 11, No. 4, 1156–1174

LARGE SAMPLE OPTIMALITY OF LEAST SQUARES CROSS-VALIDATION IN DENSITY ESTIMATION

BY PETER HALL

Australian National University

We prove that the method of cross-validation suggested by A. W. Bowman and M. Rudemo achieves its goal of minimising integrated square error, in an asymptotic sense. The tail conditions we impose are only slightly more severe than the hypothesis of finite variance, and so least squares cross-validation does not exhibit the pathological behaviour which has been observed for Kullback-Leibler cross-validation. This is apparently the first time that a cross-validatory procedure for density estimation has been shown to be asymptotically optimal, rather then simply consistent.

1. Introduction. Cross-validatory methods in density estimation have generated considerable interest in recent years. They were introduced on an *ad hoc* basis by Habbema, Hermans and van den Broek (1974) and Duin (1976), and shown to be cross-validatory by Titterington (1978, 1980). Their strong intuitive appeal and relative simplicity have caused them to be adopted by several workers. However, support for cross-validation is not universal. For example, it is rather disturbing to learn that the common cross-validatory methods produce inconsistent estimators unless the distribution tails are very small; see Schuster and Gregory (1981) and Chow, Geman and Wu (1982). In particular, the cross-validatory estimators are inconsistent when used with a distribution which has regularly varying tails, such as Student's t. They are even inconsistent in the case of the exponential distribution. Chow, Geman and Wu (1982) have proved consistency in the case of a distribution with bounded support. Bowman (1980) has established consistency in the categorical case. Stone (1974a, 1977) has discussed general properties of cross-validatory techniques.

Given the popularity of cross-validatory methods, and the problems encountered with them for moderately long-tailed distributions, it is important to consider modifications which make them more robust. The most obvious approach is to truncate the sampling distribution, although Hall (1982a) has shown that this procedure can lead to poor performance. (But see Marron, 1983.) An alternative, delightfully simple method has been suggested by Bowman (1982, 1983) and Rudemo (1982). Following Titterington (1978, 1980), Bowman points out that the commonly used cross-validation can be derived by minimising the Kullback-Leibler loss function,

$$I(p, q) = \int p(x)\log\{p(x)/q(x)\}\,dx, \tag{1.1}$$

averaged over the sample, in which $p(x)$ is taken as the Dirac delta function at the sample point X_i, and $q(x)$ is a kernel density estimate based on the sample excluding X_i. If we have reservations about the estimator derived using this loss function, we might consider employing a different measure of loss. The most widely accepted means of determining the performance of an estimator f_n of a density f is in terms of its mean integrated square error,

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

Received November 1982; revised June 1983.
AMS 1980 *subject classifications.* 62G05, 62E20.
Key words and phrases. Asymptotically optimal, Bowman and Rudemo's method, cross-validation, integrated square error, least squares, nonparametric density estimation.

This suggests that we might change I to

$$I(p, q) = \int \{p(x) - q(x)\}^2\, dx. \tag{1.2}$$

In order to describe the estimator resulting from the loss function (1.2), we shall introduce some notation. Given a random sample $X_1, \cdots, X_n$ from a common distribution with density f, let

$$f_n(x) = (nh)^{-1} \sum_{j=1}^n K\{(x - X_j)/h\}$$

denote a kernel estimator of f, and let

$$f_{ni}(x) = \{(n-1)h\}^{-1} \sum_{j \neq i} K\{(x - X_j)/h\}, \quad 1 \le i \le n,$$

denote the estimator computed with the ith observation deleted from the sample. Of course, the quantities f_n and f_{ni} depend on the window size, h, and when it is necessary to stress this dependence we shall write them as $f_n(x \mid h)$ and $f_{ni}(x \mid h)$, respectively. Bowman (1982) showed that the cross-validatory estimator based on the loss function (1.2), is obtained by using the value of h, say $\hat{h}$, which minimises

$$\alpha_n(h) = n^{-1} \sum_{i=1}^n \int f_{ni}^2(x \mid h)\, dx - 2n^{-1} \sum_{i=1}^n f_{ni}(X_i \mid h).$$

Rudemo (1982) introduced the same technique from a slightly different viewpoint. In many practical cases the integral in the definition of α_n can be calculated analytically, and does not require numerical computation.

A simple analytic argument shows that

$$\begin{aligned} n^{-1}\sum_{i=1}^n \int f_{ni}^2(x)\, dx &= (1 - n^{-1})^{-2}(1 - 2n^{-1}) \\ &\quad \cdot \int f_n^2(x)\, dx + (n-1)^{-2}h^{-1} \int K^2(x)\, dx \\ &= \int f_n^2(x)\, dx + O(1/n^2h) \end{aligned} \tag{1.3}$$

in probability, as $n \to \infty$. It follows from our proofs in Section 3 that one of the terms in an expansion of $\alpha_n(h)$ is of order $1/nh$, and so the remainder term in (1.3) is negligible in comparison. Therefore the cross-validatory criterion could be changed from $\alpha_n(h)$ to

$$\beta_n(h) = \int f_n^2(x)\, dx - 2n^{-1} \sum_{i=1}^n f_{ni}(X_i \mid h),$$

which is slightly simpler to compute, without affecting the asymptotics. However, as Bowman (1982) and Rudemo (1982) have pointed out,

$$E\{\alpha_n(h)\} = \int E\{f_{n-1}(x) - f(x)\}^2\, dx - \int f^2(x)\, dx,$$

and so $\alpha_n(h)$ is perhaps more closely related to the concept of mean integrated square error. Our aim in the present paper is to prove that the Bowman-Rudemo method of least squares cross-validation achieves its goal of minimising integrated square error, in an asymptotic sense.

2. Results. Let K be a bounded, symmetric density function with finite second moment. Define $k = \frac{1}{2} \int z^2 K(z)\, dz$. Under quite mild conditions on f, the mean integrated square error admits the expansion

$$\int E\{f_n(x) - f(x)\}^2\,dx \tag{2.1}$$
$$= (nh)^{-1}\int K^2(z)\,dz + h^4k^2\int\{f''(x)\}^2\,dx + o\{(nh)^{-1} + h^4\}$$

as $n \to \infty$ and $h \to 0$; see for example Rosenblatt (1971). (Here and below, an unqualified integral denotes integration over the whole real line.) The "asymptotically optimal" window size with respect to mean integrated square error is that value h^* which minimises the sum of the first and second terms on the right hand side of (2.1). Thus, $h^* = \gamma n^{-1/5}$, where the constant γ depends on f as well as K. Hall (1982b) has shown that integrated square error is asymptotically equivalent to mean integrated square error. In particular,

$$\left[\int\{f_n(x \mid h^*) - f(x)\}^2\,dx\right] \Big/ \left[\int E\{f_n(x \mid h^*) - f(x)\}^2\,dx\right] \to 1$$

in probability as $n \to \infty$. Thus, an estimated window $\hat{h}$ will be asymptotically as good as the "best" window h^*, if

$$\left[\int\{f_n(x \mid \hat{h}) - f(x)\}^2\,dx\right] \Big/ \left[\int E\{f_n(x \mid h^*) - f(x)\}^2\,dx\right] \to 1 \tag{2.2}$$

in probability.

Our aim is to prove the result (2.2), in the case where $\hat{h}$ is the least squares cross-validated window size. In practice, an experimenter implementing cross-validation would be aware that h^* is of order $n^{-1/5}$, and would seek a window of this order of magnitude. Therefore we shall prove that with probability tending to one as $n \to \infty$, choosing a value of h of order $n^{-1/5}$ to minimise $\alpha_n(h)$, is asymptotically equivalent to choosing h of order $n^{-1/5}$ to minimise integrated square error. For this reason we confine attention to h within the interval $[\varepsilon n^{-1/5}, \lambda n^{-1/5}]$, for arbitrarily small ε and large λ. At the end of this section we shall prove that $\hat{h}/h^* \to 1$ in probability.

The proof of (2.2) involves three steps, which are described by Theorems 1, 2 and 3, respectively. The first step consists of deriving an asymptotic expansion for $\alpha_n(h)$. We assume throughout that f has two bounded derivatives on $(-\infty, \infty)$, and in Theorem 1 we suppose in addition that

(2.3) f'' exists, is bounded and is uniformly continuous on $(-\infty, \infty)$,

and

$$\int |f'(x)|\,dx < \infty \quad \text{and} \quad \int [F(x)\{1 - F(x)\}]^{1/2}\,dx < \infty, \tag{2.4}$$

where F is the distribution function corresponding to f. The second part of condition (2.4) holds if $E(X^2 |\log|X||^{2+\varepsilon}) < \infty$ for some $\varepsilon > 0$. Our assumption that K be symmetric can be avoided by using a slightly longer proof of Theorem 1, but since symmetric kernels are the rule in practice, it seems pointless to use a weaker condition. In Theorem 1 we assume that K has two bounded derivations, and that

$$\int z^2\{K(z) + |K'(z)| + |K''(z)|\}\,dz < \infty. \tag{2.5}$$

Note that $z^2K(z) = -\int_z^\infty \{2uK(u) + u^2K'(u)\}\,du$, and so (2.5) implies that $z^2K(z) \to 0$ as $z \to \infty$. The smoothness condition on K does not appear to be intrinsically important.

THEOREM 1. *Under conditions* (2.3) – (2.5), *the criterion* $\alpha_n(h)$ *admits the expansion*

$$\alpha_n(h) = \int \{f_n(x \mid h) - f(x)\}^2\, dx - 2n^{-1} \textstyle\sum_1^n f(X_j) + \int f^2(x)\, dx + o_p(n^{-4/5}) \tag{2.6}$$

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

The conclusion of uniform convergence in this theorem means that the remainder term $r_n(h)$ denoted by $o_p(n^{-4/5})$ in (2.6), satisfies

$$n^{4/5} \sup_{\varepsilon n^{-1/5} \le h \le \lambda n^{-1/5}} | r_n(h) | \to 0$$

in probability as $n \to \infty$.

Let $\hat{h}$ be the value of h in the range $\varepsilon n^{-1/5} \le h \le \lambda n^{-1/5}$ which minimises $\alpha_n(h)$. The second and third terms on the right hand side in (2.6) do not depend on h, and so an immediate corollary of Theorem is that

$$\int \{f_n(x \mid \hat{h}) - f(x)\}^2\, dx = \inf_{\varepsilon n^{-1/5} \le h \le \lambda n^{-1/5}} \int \{f_n(x \mid h) - f(x)\}^2\, dx + o_p(n^{-4/5}). \tag{2.7}$$

Our next theorem provides us with a lower bound to the infimum on the right hand side of (2.7).

Assume that K is of bounded variation on $(-\infty, \infty)$, and that

$$\int | z |^{5/2} K(z)\, dz < \infty \quad \text{and} \quad \int | z \| \, dK(z) | < \infty. \tag{2.8}$$

THEOREM 2. *Under conditions* (2.3) *and* (2.8),

$$\begin{aligned}\int_{-d}^{d} \{f_n(x \mid h) - f(x)\}^2\, dx &= (nh)^{-1} \left\{\int_{-d}^{d} f(x)\, dx\right\} \left\{\int K^2(z)\, dz\right\} \\ &\quad + h^4 k^2 \int_{-d}^{d} \{f''(x)\}^2\, dx + o_p(n^{-4/5})\end{aligned} \tag{2.9}$$

uniformly in $\varepsilon n^{-1/5} \le h \le \lambda n^{-1/5}$, *for any* $0 < \varepsilon < \lambda < \infty$ *and* $0 < d < \infty$.

Choose ε so small and λ so large that $\varepsilon < \gamma < \lambda$. (Indeed, it is not difficult to see that Theorems 1 and 2 remain true if $\varepsilon = \varepsilon(n) \to 0$ and $\lambda = \lambda(n) \to \infty$ sufficiently slowly.) Then if f'' is square integrable, it follows from (2.7) and (2.9) that

$$\begin{aligned}&\int \{f_n(x \mid \hat{h}) - f(x)\}^2\, dx \\ &\quad \ge (nh^*)^{-1} \int K^2(z)\, dz + (h^*)^4 k^2 \int \{f''(x)\}^2\, dx + o_p(n^{-4/5}) \\ &\quad = \int E\{f_n(x \mid h^*) - f(x)\}^2\, dx + o_p(n^{-4/5}).\end{aligned} \tag{2.10}$$

In order to prove the reverse inequality, we note the following special case of Theorem 2 of Hall (1982b).

THEOREM 3. *Assume condition* (2.8), *and that* f'' *is square integrable. If* $h = h(n)$ *denotes a deterministic sequence converging to zero and such that* $nh(\log n)^{-2} \to \infty$, *then*

$$\int \{f_n(x \mid h) - f(x)\}^2\, dx = \int E\{f_n(x \mid h) - f(x)\}^2\, dx + o_p\{(nh)^{-1} + h^4\} \tag{2.11}$$

as $n \to \infty$.

Taking $h = h^*$ in (2.11), and combining that result with (2.7), we obtain

$$\int \{f_n(x \mid \hat{h}) - f(x)\}^2\, dx \le \int E\{f_n(x \mid h^*) - f(x)\}^2\, dx + o_p(n^{-4/5}).$$

The limit theorem (2.2) follows from this inequality and (2.10).

We shall prove finally that under the conditions of Theorems 1–3,

$$\hat{h}/h^* \to 1 \tag{2.12}$$

in probability. Using Helly's extraction principle, we may choose a subsequence $\{n_k\}$ such that $\hat{h}/h^*$ has a limiting distribution as $n \to \infty$ along the subsequence. Since $\hat{h}$ lies within the interval $[\varepsilon n^{-1/5}, \lambda n^{-1/5}]$, the limiting distribution is proper and confined to $[\varepsilon/\gamma, \lambda/\gamma]$. Let $l > 0$ be a point of support of the distribution. If condition (2.12) fails then we may choose the subsequence $\{n_k\}$ in such a way that $l \neq 1$. Let us suppose that $0 < l < 1$; the case $l > 1$ may be treated similarly. Define $\varepsilon = \frac{1}{2}\min(l, 1 - l)$, and choose d so large that for some $\rho > 1$,

$$\frac{(nh^*z)^{-1}\left\{\int_{-d}^{d} f(x)\, dx\right\}\left\{\int K^2(y)\, dy\right\} + (h^*z)^4 k^2 \int_{-d}^{d} \{f''(x)\}^2\, dx}{(nh^*)^{-1}\int K^2(y)\, dy + (h^*)^4 k^2 \int \{f''(x)\}^2\, dx} > \rho$$

whenever $z \in [l - \varepsilon, l + \varepsilon]$. Note that the ratio on the left hand side does not depend on n, and that the existence of such a d is ensured by the fact that h^* minimises the denominator. From Theorem 2 we may deduce that

$$\int \{f_n(x \mid \hat{h}) - f(x)\}^2\, dx \ge (n\hat{h})^{-1}\left\{\int_{-d}^{d} f(x)\, dx\right\}\left\{\int K^2(z)\, dz\right\} + \hat{h}^4 k^2 \int_{-d}^{d} \{f''(x)\}^2\, dx + o_p(n^{-4/5}). \tag{2.13}$$

When $\hat{h}/h^* \in [l - \varepsilon, l + \varepsilon]$, the right hand side of (2.13) exceeds

$$\rho \int E\{f_n(x \mid h^*) - f(x)\}^2\, dx + o_p(n^{-4/5}),$$

and since l is a point of support of the subsequence limit of $\hat{h}/h^*$,

$$\limsup_{n\to\infty} P\{\hat{h}/h^* \in [l - \varepsilon, l + \varepsilon]\} > 0.$$

These results contradict the limit theorem (2.2), and so the result (2.12) follows by contradiction.

Note that we have proved (2.12) as a corollary of the asymptotic optimality of $\hat{h}$, which is quite the reverse of the usual approach. Theorem 2 may be extended to the case $d = \infty$ under slightly more restrictive conditions, and thus it may be proved that the result (2.2) holds for any sequence of random window sizes $\hat{h}$ satisfying $\hat{h}/h^* \to 1$ in probability. Having derived the relation (2.12), we may apply the results of Krieger and Pickands (1981) to prove that $f_n(x \mid \hat{h})$ satisfies the same central limit theorem as $f_n(x \mid h^*)$.

3. Proofs.

PROOF OF THEOREM 1. Observe that with $m = n - 1$,

$$\textstyle\sum_{i=1}^{n} f_{ni}(X_i) = (n/m) \sum_{i=1}^{n} f_n(X_i) - (n/mh)K(0), \tag{3.1}$$

and that

$$
\begin{aligned}
n^{-1} \textstyle\sum_{i=1}^{n} f_n(X_i) &= \int f_n(x)\, dF_n(x) \\
&= \int f_n(x) f(x)\, dx - \int \{F_n(x) - F(x)\} f'(x)\, dx \\
&\quad - \int \{F_n(x) - F(x)\}\{f'_n(x) - f'(x)\}\, dx,
\end{aligned} \tag{3.2}
$$

where F_n is the empiric distribution function. Therefore the proof of Theorem 1 consists largely of deriving information about the term

$$I_{n1}(h) \equiv \int \{F_n(x) - F(x)\}\{f'_n(x) - f'(x)\}\, dx.$$

Write $I_{n1}(h) = I_{n2}(h) + I_{n3}(h)$, where

$$I_{n2}(h) = \int \{F_n(x) - F(x)\}\{f'_n(x) - \mu(x, h)\}\, dx,$$

$$I_{n3}(h) = \int \{F_n(x) - F(x)\}\{\mu(x, h) - f'(x)\}\, dx$$

and

$$\mu(x, h) = \int f'(x - hz) K(z)\, dz.$$

By Theorem 3 of Komlós, Major and Tusnády (1975), there exists a sequence of Brownian bridges W_n^0, $n \geq 2$, such that the variables

$$Z_n = n(\log n)^{-1} \sup_{-\infty<x<\infty} |F_n(x) - F(x) - n^{-1/2} W_n^0\{F(x)\}|, \quad n \geq 2,$$

from a tight sequence. (Henceforth we shall drop the subscript, n, on W_n^0.) Therefore

$$
\begin{aligned}
I_{n3}(h) &= n^{-1/2} \int W^0\{F(x)\}\{\mu(x, h) - f'(x)\}\, dx \\
&\quad + O_p(n^{-1} \log n) \int \{|\mu(x, h)| + |f'(x)|\}\, dx.
\end{aligned}
$$

It follows from (2.4) that $\int \{|\mu(x, h)| + |f'(x)|\}\, dx$ is bounded uniformly in h, and so

$$\sup_{\varepsilon n^{-1/5} \leq h \leq \lambda n^{-1/5}} |I_{n3}(h)| = o_p(n^{4/5}), \tag{3.3}$$

provided we can prove that with

$$\zeta(h) = \int W^0\{F(x)\}\{\mu(x, h) - f'(x)\}\, dx,$$

we have

$$\sup_{\varepsilon n^{-1/5} \leq h \leq \lambda n^{-1/5}} |\zeta(h)| = o_p(n^{-3/10}).$$

The Gaussian process $\zeta(h)$ is continuous on $I_n = [\varepsilon n^{-1/5}, \lambda n^{-1/5}]$, and may be shown to have variance $E\{\zeta^2(h)\} \leq C_1 h^4$ and quadratic variation $E\{\zeta(h_1) - \zeta(h_2)\}^2 - C_2 h_2^3(h_2 - h_1)$ for $0 < h_1 < h_2$. Therefore by Fernique's inequality (see the Corollary, page 307 of Marcus, 1970), there exist constants C_3 and C_4 depending only on C_1 and C_2 such that

$$P\{\sup_{h \in I_n} |\zeta(h)| > n^{-2/5} C_3 t\} \leq C_4 \exp(-t^2/2)$$

for all $t \geq 1$. This completes the proof of (3.3).

Observe that

$$f'_n(x) - \mu(x, h) = h^{-2} \int K'\{(x - y)/h\} d\{F_n(y) - F(y)\}$$
$$= -h^{-2} \int \{F_n(x - zh) - F(x - zh)\}\, dK'(z)$$
$$= -n^{-1/2}h^{-2} \int W^0\{F(x - zh)\}\, dK'(z) + r_n(x),$$

where

$$\sup_{-\infty<x<\infty} | r_n(x) | \le C_0 n^{-1}(\log n) h^{-2} Z_n$$

and $C_o = \int | dK'(z) |$. Therefore we may write $I_{n2}(h) = I_{n4}(h) + I_{n5}(h)$, where

$$I_{n4}(h) = -n^{-1/2}h^{-2} \int \{F_n(x) - F(x)\}\, dx \int W^0\{F(x - zh)\}\, dK'(z)$$

and

$$| I_{n5}(h) | \le C_0 n^{-1}(\log n) h^{-2} Z_n \int | F_n(x) - F(x) |\, dx.$$

The argument leading to (3.3) now produces the estimate

(3.4) $$\sup_{\epsilon n^{-1/5} \le h \le \lambda n^{-1/5}} | I_{n5}(h) | = o_p(n^{-4/5}).$$

Applying the Komlós, Major and Tusnády expansion directly to the term $F_n(x) - F(x)$ in the expression for $I_{n4}(h)$, we find that

$$I_{n4}(h) = -n^{-1}h^{-2} \int W^0\{F(x)\}\, dx \int W^0\{F(x - zh)\}\, dK'(z) + I_{n6}(h),$$

where

$$| I_{n6}(h) | \le n^{-3/2}(\log n) h^{-2} Z_n \int | dK'(z) | \int | W^0\{F(x - zh)\} |\, dx$$
$$= C_0 n^{-3/2}(\log n) h^{-2} Z_n \int | W^0\{F(x)\} |\, dx.$$

But

$$\int E| W^0\{F(x)\} |\, dx \le \int [E| W^0\{F(x)\} |^2]^{1/2}\, dx = \int [F(x)\{1 - F(x)\}]^{1/2}\, dx,$$

and so

(3.5) $$\sup_{\epsilon n^{-1/5} \le h \le \lambda n^{-1/5}} | I_{n6}(h) | = o_p(n^{-4/5}).$$

Define

$$Y_1(h) \equiv \frac{1}{2} \int dK'(z) \int W^0\{F(x)\} W^0\{F(x - zh)\}\, dx$$
$$= \frac{1}{2} \int dK'(z) \int W^0\left\{F\left(x + \frac{1}{2} zh\right)\right\} W^0\left\{F\left(x - \frac{1}{2} zh\right)\right\} dx$$
$$= \int_0^\infty dK'(z) \int W^0\{F(x)\} W^0\{F(x + zh)\}\, dx,$$

using the symmetry of K. From the estimates (3.3), (3.4) and (3.5) we may deduce that

$$I_{n1}(h) = -2n^{-1}h^{-2}Y_1(h) + I_{n7}(h), \tag{3.6}$$

where

$$\sup_{\varepsilon n^{-1/5} \le h \le \lambda n^{-1/5}} | I_{n7}(h) | = o_p(n^{-4/5}). \tag{3.7}$$

Thus, the properties of $I_{n1}(h)$ are determined in a large part by those of the stochastic process $Y_1(h)$, which may be shown to equal the limit of quadratic forms in zero-mean normal random variables. Shortly we shall prove the following lemmas.

LEMMA 1. *Under the conditions of Theorem* 1,

$$\sup_{\varepsilon \le t \le \lambda, t \le s \le t + n^{-1/5-\Delta}} | Z(t) - Z(s) | = O_p\{n^{-1/5-\Delta/2}(\log n)^{1/2}\}$$

as $n \to \infty$.

LEMMA 2. *Let* $Y(t) = Y_1(th_0)$ *for* $\varepsilon \le t \le \lambda$, *where* $h_0 = n^{-1/5}$, *and define* $Z(t) = Y(t) - E\{Y(t)\}$. *Suppose* $0 \le \Delta \le 1/100$. *Then under the conditions of Theorem* 1,

$$E\{Y(t)\} = -\frac{1}{2} h_0 t K(0) + \frac{1}{2} (h_0 t)^2 \int f^2(x)\, dx + O(h_0^3), \tag{3.8}$$

$$\operatorname{var}\{Y(t)\} = O(h_0^3) \tag{3.9}$$

and

$$E\{Z(t + n^{-1/5-\Delta}) - Z(t)\}^2 = O(h_0^4) \tag{3.10}$$

uniformly in $\varepsilon \le t \le \lambda$, *as* $n \to \infty$.

We utilise Lemmas 1 and 2 in the following way. Let $0 < \Delta \le 1/100$, and divide the interval $[\varepsilon, \lambda]$ into N or $N + 1$ intervals (t_{j-1}, t_j), each of length $n^{-1/5-\Delta}$, where $\varepsilon = t_0 < t_1 < \cdots < t_N \le \lambda < t_{N+1}$. Then

$$\begin{aligned}\sup_{\varepsilon \le t \le \lambda} | Z(t) | &\le \sup_{j \ge 0} | Z(t_j) | + \sup_{\varepsilon \le t \le \lambda, t \le s \le t + n^{-1/5-\Delta}} | Z(s) - Z(t) | \\ &= \sup_{j \ge 0} | Z(t_j) | + o_p(n^{-1/5}),\end{aligned}$$

using Lemma 2. Now,

$$\sup_{j \ge 0} | Z(t_j) |^2 \le | Z(t_0) |^2 + \textstyle\sum_{j \ge 1} | Z^2(t_j) - Z^2(t_{j-1}) |,$$

and so

$$\begin{aligned}E\{\sup_{j \ge 0} | Z(t_j) |\} &\le [E\{Z^2(t_0)\} + \textstyle\sum_{j \ge 1} \{E | Z(t_j) - Z(t_{j-1}) |^2 E | Z(t_j) + Z(t_{j-1}) |^2\}^{1/2}]^{1/2} \\ &\le C_1\{h_0^3 + (N + 1)(h_0^4 h_0^3)^{1/2}\}^{1/2} \le C_2 n^{-1/4+\Delta/2}.\end{aligned}$$

Therefore

$$\sup_{\varepsilon \le t \le \lambda} | Z(t) | = o_p(n^{-1/5}),$$

and so by (3.6) and (3.7),

$$I_{n1}(h) = -2n^{-1}h^{-2}E\{Y_1(h)\} + I_{n8}(h)$$

where

$$\sup_{\varepsilon n^{-1/5} \le h \le \lambda n^{-1/5}} | I_{n8}(h) | = o_p(n^{-4/5}).$$

We may now deduce from (3.1), (3.2) and (3.8) that

$$n^{-1} \sum_{i=1}^{n} f_{ni}(X_i) = \int f_n(x) f(x)\, dx - \int \{F_n(x) - F(x)\} f'(x)\, dx + o_p(n^{-4/5})$$

uniformly in $\varepsilon n^{-1/5} \le h \le \lambda n^{-1/5}$. Theorem 1 follows easily from this expansion and result (1.3).

The remainder of the proof consists of deriving Lemmas 1 and 2. Let us temporarily assume Lemma 2. To prove Lemma 1, observe that

$$| Y_1(sh_0) - Y_1(th_0) | \le \int_0^\infty | dK'(z) | \int | W^0\{F(x)\} | \, w^0(| F(x - zsh_0) - F(x - zth_0) |)\, dx,$$

where $w^0(u) = \sup_{0<s,t<1;\, |s-t| \le u} | W^0(s) - W^0(t) |$ denotes the modulus of continuity of W^0. Since $w^0(u) \le Z^0\{u(1 - \log u)\}^{1/2}$ for an almost surely finite variable Z^0 (see for example Proposition 3 of Silverman, 1978), and since

$$\{u(1 - \log u)\}^{1/2} \le C_3 n^{-1/5-\Delta/2} (\log n)^{1/2} \{1 + | z | | \log | z | | \}^{1/2}$$

when $u = | F(x - zsh_0) - F(x - zth_0) |$ and $| s - t | \le n^{-1/5-\Delta}$, then

$$| Y_1(sh_0) - Y_1(th_0) | \le C_3 n^{-1/5-\Delta/2} (\log n)^{1/2} Z^0 \times \left[\int_0^\infty \{1 + | z | | \log | z | | \}^{1/2} | dK'(z) | \right] \left[\int | W^0\{F(x)\} | \, dx \right].$$

Therefore

$$\sup_{\varepsilon \le t \le \lambda,\, t \le s \le t + n^{-1/5-\Delta}} | Y_1(sh_0) - Y_1(th_0) | = O_p\{n^{-1/5-\Delta/2} (\log n)^{1/2}\}.$$

Lemma 1 follows on combining this estimate with (3.8).

PROOF OF LEMMA 2. To prove (3.8), note that

$$E\{Y_1(h)\} = \int_0^\infty dK'(z) \int F(x)\{1 - F(x + zh)\}\, dx$$
$$= -h \int F(x) \left[f(x) K(0) + \frac{1}{2} h f'(x) + h \int_0^\infty \{f'(x + zh) - f'(x)\} K(z)\, dz \right] dx.$$

Now,

$$\int F(x)\, dx \int_0^\infty \{f'(x + zh) - f'(x)\} K(z)\, dz = - \int_0^\infty K(z)\, dz \int f(x) \{f(x + zh) - f(x)\}\, dx,$$

which in absolute value is dominated by

$$\{\sup_{-\infty < x < \infty} | f'(x) | \} h \int_0^\infty z K(z)\, dz \int f(x)\, dx.$$

Therefore (3.8) holds.

In the remainder of the proof we write h for h_0. The next step is to calculate

$$E\{Y_1(hs)Y_1(ht)\} = \int\int dx\, dy \int_0^\infty \int_0^\infty G_1\, dK'(u)\, dK'(v), \tag{3.11}$$

where $G_1 = E[W^0\{F(x)\}W^0\{F(y)\}W^0\{F(x + uhs)\}W^0\{F(x + vht)\}]$. If N_1, N_2, N_3 and N_4 denote normal random variables with zero means and covariance matrix (σ_{ij}), then

$$E(N_1N_2N_3N_4) = \sigma_{12}\sigma_{34} + \sigma_{13}\sigma_{24} + \sigma_{14}\sigma_{23}.$$

Therefore if $0 \le a \le b \le c \le d \le 1$,

$$E\{W^0(a)W^0(b)W^0(c)W^0(d)\} = a(1 - d)(2b + c - 3\, bc). \tag{3.12}$$

Let $a_1 = F(x)$, $b_1 = F(y)$, $a_2 = F(x + uhs)$ and $b_2 = F(y + vht)$, and assume for the time being that $x < y$, $u > 0$ and $v > 0$. Then $a_1 < a_2$, $b_1 < b_2$ and $a_1 < b_1$. The relationships among the a_i's and b_i's take three possible forms: $a_1 < a_2 < b_1 < b_2$, $a_1 < b_1 < a_2 < b_2$ or $a_1 < b_1 < b_2 < a_2$. These inequalities determine the regions $0 < us < (y - x)/h$, $(y - x)/h < us < vt + (y - x)/h$ and $us > vt + (y - x)/h$, respectively. Using the formula (3.12) to evaluate G_1 in these regions, we may deduce that

$$\int_0^\infty \int_0^\infty G_1\, dK'(u)\, dK'(v) = F(x) \int_0^\infty G_2\, dK'(v),$$

where

$$\begin{aligned} G_2 = &\int_0^{(y-x)/hs} \{1 - F(y + vht)\} \\ &\cdot\{2F(x + uhs) + F(y) - 3F(x + uhs)F(y)\}\, dK'(u) \\ &+ \int_{(y-x)/hs}^{vt/s+(y-x)/hs} \{1 - F(y + vht)\} \\ &\cdot\{2F(y) + F(x + uhs) - 3F(y)F(x + uhs)\}\, dK'(u) \\ &+ \int_{vt/s+(y-x)/hs}^{\infty} \{1 - F(x + uhs)\} \\ &\cdot\{2F(y) + F(y + vht) - 3F(y)F(y + vht)\}\, dK'(u) \\ = &\{1 - F(y + vht)\} \\ &\cdot\left[\int_0^{vt/s+(y-x)/hs} \{F(x + uhs) + F(y) - 3F(x + uhs)F(y)\}\, dK'(u)\right. \\ &\left. + \int_0^{(y-x)/hs} F(x + uhs)\, dK'(u) + \int_{(y-x)/hs}^{vt/s+(y-x)/hs} F(y)\, dK'(u)\right] \\ &+ \{2F(y) + F(y + vht) - 3F(y)F(y + vht)\} \\ &\cdot\int_{vt/s+(y-x)/hs}^{\infty} \{1 - F(x + uhs)\}\, dK'(u). \end{aligned} \tag{3.13}$$

It may be proved after some algebra that

$$\int_0^{vt/s+(y-x)/hs} \{F(x+uhs)+F(y)-3F(x+uhs)F(y)\}\,dK'(u)$$

$$= \{F(y+vht)+F(y)-3F(y+vht)F(y)\}K'\left(\frac{vt}{s}+\frac{y-x}{hs}\right)$$

$$+ hs\{1-3F(y)\}$$

$$\cdot\left\{f(x)K(0)-f(y+vht)K\left(\frac{vt}{s}+\frac{y-x}{hs}\right)\right.$$

$$\left.+ hs\int_0^{vt/s+(y-x)/hs} f'(x+uhs)K(u)\,du\right\},$$

$$\int_0^{(y-x)/hs} F(x+uhs)\,dK'(u) + \int_{(y-x)/hs}^{vt/s+(y-x)/hs} F(y)\,dK'(u)$$

$$= F(y)K'\left(\frac{vt}{s}+\frac{y-x}{hs}\right)$$

$$+ hs\left\{f(x)K(0)-f(y)K\left(\frac{y-x}{hs}\right)+hs\int_0^{(y-x)/hs} f'(x+uhs)K(u)\,du\right\}$$

and

$$\int_{vt/s+(y-x)/hs}^{\infty} \{1-F(x+uhs)\}\,dK'(u)$$

$$= -\{1-F(y+vht)\}K'\left(\frac{vt}{s}+\frac{y-x}{hs}\right)$$

$$- hs\left\{f(y+vht)K\left(\frac{vt}{s}+\frac{y-x}{hs}\right)+hs\int_{vt/s+(y-x)/hs}^{\infty} f'(x+uhs)K(u)\,du\right\}.$$

Substituting these results into (3.13) and simplifying, we obtain

$$G_2 = hs\{1-3F(y)\}\{1-F(y+vht)\}\left\{f(x)K(0)+hs\int_0^{\infty} f'(x+uhs)K(u)\,du\right\}$$

$$+ hs\{1-F(y+vht)\}$$

$$\cdot\left\{f(x)K(0)-f(y)K\left(\frac{y-x}{hs}\right)+hs\int_0^{(y-x)/hs} f'(x+uhs)K(u)\,du\right\} \tag{3.14}$$

$$- hs\{1-F(y)\}$$

$$\cdot\left\{f(y+vht)K\left(\frac{vt}{s}+\frac{y-x}{hs}\right)+hs\int_{vt/s+(y-s)/hs}^{\infty} f'(x+uhs)K(u)\,du\right\}.$$

Next observe that

$$\int_0^{\infty} \{1-F(y+vht)\}\,dK'(v) = -\left\{htf(y)K(0)+(ht)^2\int_0^{\infty} f'(y+vht)K(v)\,dv\right\}$$

and

$$\int_0^\infty \left\{f(y+vht)K\left(\frac{vt}{s}+\frac{y-x}{hs}\right)+hs\int_{vt/s+(y-x)/hs}^\infty f'(x+uhs)K(u)\,du\right\}dK'(v)$$

$$=-(t/s)\int_0^\infty f(y+vht)K'\left(\frac{vt}{s}+\frac{y-x}{hs}\right)dK(v).$$

Therefore by (3.14),

$$\begin{aligned}\int_0^\infty G_2\,dK'(v) &= -(hs)(ht)\left\{f(y)K(0)+ht\int_0^\infty f'(y+vht)K(v)\,dv\right\}\\ &\quad\times\left[\{1-3F(y)\}\left\{f(x)K(0)+hs\int_0^\infty f'(x+uhs)K(u)\,du\right\}\right.\\ &\quad+f(x)K(0)-f(y)K\left(\frac{y-x}{hs}\right)\\ &\quad\left.+hs\int_0^{(y-x)/hs} f'(x+uhs)K(u)\,du\right]\\ &\quad+ht\{1-F(y)\}\int_0^\infty f(y+vht)K'\left(\frac{vt}{s}+\frac{y-x}{hs}\right)dK(v).\end{aligned}\tag{3.15}$$

Now,

$$\begin{aligned}\int_{-\infty}^y F(x)\{f'(x+uhs)-f'(x)\}\,dx &= F(y)\{f(y+uhs)-f(y)\}\\ &\quad-\int_{-\infty}^y f(x)\{f(x+uhs)-f(x)\}\,dx\\ &= hur_1,\end{aligned}$$

$$\begin{aligned}\int_{-\infty}^y F(x)K\left(\frac{y-x}{hs}\right)dx &= \frac{1}{2}hsF(y)+hs\int_0^\infty\{F(y-zhs)-F(y)\}K(z)\,dz\\ &= \frac{1}{2}hsF(y)+h^2r_2\end{aligned}$$

and

$$\begin{aligned}&\int_{-\infty}^y F(x)\,dx\int_0^{(y-x)/hs} f'(x+uhs)K(u)\,du\\ &\quad=\int_0^\infty K(u)\left\{F(y-uhs)f(y)-\int_{-\infty}^y f(x)f(x-uhs)\,dx\right\}du\\ &\quad=\frac{1}{2}f(y)F(y)-\frac{1}{2}\int_{-\infty}^y f^2(x)\,dx+hr_3,\end{aligned}$$

where r_j stands for a function which is uniformly bounded as $h\to 0$. Combining these estimates with (3.15) and simplifying we may deduce that

$$
\begin{aligned}
\int_{-\infty}^{y} F(x)\,dx \int_0^\infty G_2\,dK'(v)
&= -\frac{1}{2}(hs)(ht)\left\{f(y)K(0) + ht\int_0^\infty f'(y+vht)K(v)\,dv\right\} \\
&\quad \times \Big[\{1-3F(y)\}\left\{F^2(y)K(0) + hs\int_{-\infty}^{y} f'(x)F(x)\,dx\right\} \\
&\qquad + F^2(y)K(0) - hsf(y)F(y) + hs\int_{-\infty}^{y} f'(x)F(x)\,dx + h^2 r_4\Big] \\
&\quad + ht\{1-F(y)\}\int_{-\infty}^{y} F(x)\,dx \\
&\quad \cdot \int_0^\infty f(y+vht)K'\left(\frac{vt}{s}+\frac{y-x}{hs}\right)dK(v).
\end{aligned} \tag{3.16}
$$

Next observe that

$$
\begin{aligned}
\int_{-\infty}^{y} F(x)\,dx \int_0^\infty f(y+vht)K'\left(\frac{vt}{s}+\frac{y-x}{hs}\right)dK(v)
&= hs\int_{-\infty}^{y} f(x)\,dx\int_0^\infty f(y+vht)K\left(\frac{vt}{s}+\frac{y-x}{hs}\right)dK(v) \\
&\quad - hsF(y)\int_0^\infty f(y+vht)K(vt/s)\,dK(v),
\end{aligned}
$$

$$
\begin{aligned}
\int_{-\infty}^{y} f(x)\,dx \int_0^\infty f(y+vht)K\left(\frac{vt}{s}+\frac{y-x}{hs}\right)dK(v)
&= hs\int_0^\infty f(y-zhs)\,dz\int_0^\infty f(y+vht)K(vt/s+z)\,dK(v) \\
&= hs\,f(y)\int_0^\infty f(y+vht)\,dK(v)\int_0^\infty K(vt/s+z)\,dz \\
&\quad + h^2 r_5\int_0^\infty f(y+vht)\,|dK(v)|\int_0^\infty zK(vt/s+z)\,dz
\end{aligned}
$$

and

$$
\begin{aligned}
\int_0^\infty f(y+vht)K(vt/s)\,dK(v) &= f(y)\int_0^\infty K(vt/s)\,dK(v) \\
&\quad + ht\,f'(y)\int_0^\infty vK(vt/s)\,dK(v) + h^2 r_6.
\end{aligned}
$$

Using these estimates to simplify the last term in (3.16), we may deduce that

$$\int_{-\infty}^{y} F(x)\,dx \int_{0}^{\infty} G_2\,dK'(v)$$

$$= -\frac{1}{2}(hs)(ht)\Big[f(y)F^2(y)\{2 - 3F(y)\}K^2(0)$$

$$+ hs\,f(y)K(0)\Big(\{2 - 3F(y)\}\int_{-\infty}^{y} f'(x)\,F(x)\,dx - f(y)F(y)\Big)$$

$$+ ht\,F^2(y)\{2 - 3F(y)\}K(0)\int_{0}^{\infty} f'(y + vht)K(v)\,dv\Big]$$

$$- (hs)(ht)\{1 - F(y)\}$$

$$\cdot\Big[f(y)F(y)\int_{0}^{\infty} K(vt/s)\,dK(v) + ht\,f'(y)F(y)\int_{0}^{\infty} vK(vt/s)dK(v)$$

$$- hs\,f(y)\int_{0}^{\infty} f(y + vht)\,dK(v)\int_{0}^{\infty} K(vt/s + z)\,dz\Big]$$

$$+ h^4 r_7 f(y) + h^4 r_8 \int_{0}^{\infty} f'(y + vht)K(v)\,dv + h^4 r_9 F(y)\{1 - F(y)\}$$

$$+ h^4 r_{10}\int_{0}^{\infty} f(y + vht)\,|\,dK(v)\,|\int_{0}^{\infty} zK(vt/s + z)\,dz.$$

Consequently

$$\int_{-\infty}^{\infty} dy \int_{-\infty}^{y} F(x)\,dx \int_{0}^{\infty} G_2 dK'(v)$$

$$= -\frac{1}{2}(hs)(ht)\{(-1/12)K^2(0) + c_1 hs + c_2 ht\} - (hs)(ht)$$

$$\cdot\Big\{(1/6)\int_{0}^{\infty} K(vt/s)\,dK(v) + c_3 ht\int_{0}^{\infty} vK(vt/s)\,dK(v)$$

$$+ c_4 hs\int_{0}^{\infty} dK(v)\int_{0}^{\infty} K(vt/s + z)\,dz\Big\} + O(h^4)$$

uniformly in $\varepsilon \le s, t \le \lambda$, where c_j stands for a constant depending only on f and K. Since $|K(vt/s + z) - K(v + z)| \le |v||1 - t/s|\{\sup_x |K'(x)|\}$ uniformly in z, then

$$\int_{-\infty}^{\infty} dy \int_{-\infty}^{y} F(x)\,dx \int_{0}^{\infty} G_2\,dK'(v)$$

$$= (1/24)h^2 st\Big\{K^2(0) - 4\int_{0}^{\infty} K(vt/s)\,dK(v)\Big\} \tag{3.17}$$

$$+ c_5 h^3 s^2 t + c_6 h^3 s t^2 + O(h^4)$$

uniformly in $\varepsilon \le t \le \lambda$ and $|s - t| \le h$.

The expression in (3.17) denotes the quadruple integral of (3.11) in the special case where (x, y) is constrained by $x < y$. To treat the case $y < x$ it is necessary only to interchange the roles of s and t. Thus,

$$E\{Y_1(hs)Y_1(ht)\} = (1/12)h^2st\left[K^2(0) - 2\int_0^\infty \{K(vt/s) + K(vs/t)\}\,dK(v)\right]$$

$$+ (c_5 + c_6)h^3st(s+t) + O(h^4)$$

uniformly in $\varepsilon \le t \le \lambda$ and $|s-t| \le h$. From this formula and (3.8) it follows that

$$\mathrm{cov}\{Y_1(hs),\ Y_1(ht)\} = -(1/6)h^2st\left[K^2(0) + \int_0^\infty \{K(vt/s) + K(vs/t)\}\,dK(v)\right]$$

$$+ c_7h^3st(s+t) + O(h^4).$$

The result (3.9) follows immediately. Furthermore,

$$E\{Z(s) - Z(t)\}^2 = -(1/6)h^2(s-t)^2K^2(0)$$

$$- (1/3)h^2\int_0^\infty [(s^2+t^2)K(v) - st\{K(vt/s) + K(vs/t)\}]\,dK(v)$$

$$+ c_7h^3(s+t)(s-t)^2 + O(h^4).$$

Writing $\delta = (s-t)/t$, and expanding

$$K(vt/s) + K(vs/t) = K\{v/(1+\delta)\} + K\{v(1+\delta)\}$$

as a Taylor series in δ, we may deduce that

$$E\{Z(s) - Z(t)\}^2 = O(h^4)$$

uniformly in $\varepsilon \le t \le \lambda$ and $|s-t| \le h$. This proves (3.10).

PROOF OF THEOREM 2. Let $\nu(x, h) = \int f(x-hz)K(z)\,dz$, and observe that

$$|\nu(x,h) - f(x) - h^2kf''(x)|$$

$$= \left|\int \{f(x-hz) - f(x) + hzf(x) - \frac{1}{2}(hz)^2f''(x)\}K(z)\,dz\right|$$

$$\le h^2\int |f''(x-\theta hz) - f''(x)|\,z^2K(z)\,dz = o(h^2)$$

uniformly in x, using condition (2.3), where $\theta = \theta(x, z, h) \in [0, 1]$. By Theorem 3 of Komlós, Major and Tusnády (1975),

$$f_n(x) - \nu(x,h) = n^{-1/2}h^{-1}\int W^0\{F(x-zh)\}\,dK(z) + r_n(x,h),$$

where

$$\sup_{-\infty<x<\infty}\sup_{\varepsilon n^{-1/5}\le h\le\lambda n^{-1/5}}|r_n(x,h)| = O_p(n^{-4/5}\log n).$$

Combining these estimates we may deduce that

$$\begin{aligned}\int_{-d}^{d} \{f_n(x) - f(x)\}^2\,dx &= n^{-1}h^{-2}\int_{-d}^{d}\left[\int W^0\{F(x-zh)\}\,dK(z)\right]^2 dx \\ &\quad + h^4k^2\int_{-d}^{d}\{f''(x)\}^2\,dx \\ &\quad + 2n^{-1/2}h^{-1}\int_{-d}^{d}\{\nu(x,h) - f(x)\}\,dx \\ &\quad \cdot \int W^0\{F(x-zh)\}\,dK(z) + o_p(n^{-4/5})\end{aligned} \tag{3.18}$$

uniformly in $\varepsilon n^{-1/5} \le h \le \lambda n^{-1/5}$.

Our next task is to prove that the Gaussian process

$$\zeta(h) = \int_{-d}^{d}\{\nu(x,h) - f(x) - h^2kf''(x)\}\,dx\int W^0\{F(x-zh)\}\,dK(z)$$

satisfies

$$\sup_{\varepsilon n^{-1/5}\le h\le\lambda n^{-1/5}}|\zeta(h)| = o_p(n^{-1/2}). \tag{3.19}$$

However, this is readily accomplished using Fernique's inequality, since $E\{\zeta^2(h)\} \le C_1h^5$ and $E\{\zeta(h_1) - \zeta(h_2)\}^2 \le C_2h_2^4(h_2 - h_1)$ for $0 < h_1 < h_2$. Similarly it may be proved that

$$\sup_{\varepsilon n^{-1/5}\le h\le\lambda n^{-1/5}}\left|\int_{-d}^{d} f''(x)\,dx\int W^0\{F(x-zh)\}\,dK(z)\right| = o_p(n^{-1/10}). \tag{3.20}$$

Combining (3.18) − (3.20) we see that

$$\begin{aligned}\int_{-d}^{d}\{f_n(x) - f(x)\}^2\,dx &= n^{-1}h^{-2}\int_{-d}^{d}\left[\int W^0\{F(x-zh)\}\,dK(z)\right]^2 dx \\ &\quad + h^4k^2\int_{-d}^{d}\{f''(x)\}^2\,dx + o_p(n^{-4/5})\end{aligned} \tag{3.21}$$

uniformly in $\varepsilon n^{-1/5} \le h \le \lambda n^{-1/5}$.

We shall prove next that

$$\begin{aligned}&\int_{-d}^{d}\left[\int W^0\{F(x-zh)\}\,dK(z)\right]^2 dx \\ &\qquad = h\left\{\int_{-d}^{d} f(x)\,dx\right\}\left\{\int_{-\infty}^{\infty} K^2(z)\,dz\right\} + o_p(n^{-1/5})\end{aligned} \tag{3.22}$$

uniformly in $\varepsilon n^{-1/5} \le h \le \lambda n^{-1/5}$. Theorem 2 follows from (3.21) and (3.22). Now,

$$\begin{aligned}&\int_{-d}^{d}\left[\int W^0\{F(x-zh)\}\,dK(z)\right]^2 dz \\ &= \int_{-d}^{d}\left[\int W\{F(x-zh)\}\,dK(z)\right]^2 dx + W^2(1)\int_{-d}^{d}\left[\int F(x-zh)\,dK(z)\right]^2 dx \\ &\quad - 2W(1)\int_{-d}^{d}\left[\int W\{F(x-zh)\}\,dK(z)\right]\left\{\int F(x-zh)\,dK(z)\right\} dx.\end{aligned}$$

The second term on the right hand side is easily seen to be of order $n^{-2/5}$ in probability, uniformly in $\varepsilon n^{-1/5} \le h \le \lambda n^{-1/5}$, while the third term equals

$$- 2hW(1) \int_{-d}^{d} \left[\int W\{F(x - zh)\} \, dK(z) \right] \left\{ \int f(x - zh) K(z) \, dz \right\} dx.$$

Fernique's inequality may be used to show that this quantity equals $o_p(n^{-1/5})$ uniformly in $\varepsilon n^{-1/5} \le h \le \lambda n^{-1/5}$. Therefore (3.22) will follow if we prove that

$$\begin{aligned} Y_3(h) &\equiv \int_{-d}^{d} \left[\int W\{F(x - zh)\} \, dK(z) \right]^2 dx \\ &= h \left\{ \int_{-d}^{d} f(x) \, dx \right\} \left\{ \int_{-\infty}^{\infty} K^2(z) \, dz \right\} + o_p(n^{-1/5}) \end{aligned} \tag{3.23}$$

uniformly in $\varepsilon n^{-1/5} \le h \le \lambda n^{-1/5}$.

Observe that $\int W\{F(x - zh)\} \, dK(z) = \int [W\{F(x - zh)\} - W\{F(x)\}] \, dK(z)$, and so

$$\begin{aligned} &| Y_3(h_0 s) - Y_3(h_0 t) | \\ &\quad \le \int_{-d}^{d} \left\{ \int w(| F(x - zh_0 s) - F(x - zh_0 t) |) \, | dK(z) | \right\} dx \\ &\quad \times \left[\int \{ w(| F(x - zh_0 s) - F(x) |) + w(| F(x - zh_0 t) - F(x) |) \} \, | dK(z) | \right] \\ &\quad = O_p(n^{-1/5 - \Delta/2} \log n) \end{aligned} \tag{3.24}$$

uniformly in $\varepsilon \le t \le \lambda$ and $| s - t | \le n^{-\Delta}$. Furthermore,

$$\begin{aligned} E\{ Y_3(h) \} &= h \int_{-d}^{d} dx \int f(x - uh) K^2(u) \, du \\ &= h \left\{ \int_{-d}^{d} f(x) \, dx \right\} \left\{ \int K^2(u) \, du \right\} + O(h^2). \end{aligned} \tag{3.25}$$

Writing

$$\begin{aligned} Y_3^2(h) = \int_{-d}^{d} \int_{-d}^{d} &\left(\int [W\{F(x - zh)\} - W\{F(x)\}] \, dK(z) \right)^2 \\ &\times \left(\int [W\{F(y - zh)\} - W\{F(y)\}] \, dK(z) \right)^2 dx \, dy, \end{aligned}$$

and breaking each of the squared integrals into two parts, one over $| zh | < | x - y | /2$ and the other over $| zh | > | x - y | /2$, we see that

$$\begin{aligned} Y_3^2(h) = \int_{-d}^{d} \int_{-d}^{d} &\left(\int_{| zh | < | x - y | /2} [W\{F(x - zh)\} - W\{F(x)\}] \, dK(z) \right)^2 \\ &\times \left(\int_{| zh | < | x - y | /2} [W\{F(y - zh)\} - W\{F(y)\}] \, dK(z) \right)^2 dx \, dy + R_n, \end{aligned}$$

where $| R_n |$ satisfies an inequality similar to that given for $| R_{n3} |$ on page 17 of Hall (1982b). (In that inequality, a $\times$ sign is misprinted as a $+$ sign.) Arguing as in Hall (1982b) we may deduce that $E(|R_n|) = O(h^{5/2-\eta})$ for any $\eta > 0$, and thence that

$$E\{Y_3^2(h)\} = \int_{-d}^{d}\int_{-d}^{d} E\left(\int_{|zh|<|x-y|/2}[W\{F(x-zh)\} - W\{F(x)\}]\,dK(z)\right)^2$$

$$\times\left(\int_{|zh|<|x-y|/2}[W\{F(y-zh)\} - W\{F(y)\}]\,dK(z)\right)^2 dx\,dy + O(h^{5/2-\eta})$$

$$= \{EY_3(h)\}^2 + O(h^{5/2-\eta}).$$

(The referee has suggested an even simpler proof, based on formula (2.2), page 427 of Doob, 1953.) Therefore

$$\text{var}\{Y_3(h_0t)\} = O(n^{-1/2+\eta/5}) \tag{3.26}$$

uniformly in $\varepsilon \le t \le \lambda$.

Let $Z(t) = Y_3(h_0t) - E\{Y_3(h_0t)\}$, and divide the interval $[\varepsilon, \lambda]$ up into N or $N+1$ subintervals (t_{j-1}, t_j) of length $n^{-\Delta}$, where $0 < \Delta < 1/10$. Then

$$\sup_{\varepsilon\le t\le\lambda}|Z(t)| \le \sup_{j\ge0}|Z(t_j)| + \sup_{\varepsilon\le t\le\lambda, t\le s\le t+n^{-\Delta}}|Z(s) - Z(t)|$$

$$= \sup_{j\ge0}|Z(t_j)| + O_p\{n^{-1/5-\Delta/2}\log n\},$$

using (3.24) and (3.25). Furthermore,

$$E\{\sup_{j\ge0}|Z(t_j)|\} \le C_1\{(N+1)n^{-1/2+\eta/5}\}^{1/2} \le C_2 n^{-1/4+\Delta/2+\eta/10},$$

using (3.26), and so

$$\sup_{\varepsilon\le t\le\lambda}|Z(t)| = o_p(n^{-1/5}).$$

The result (3.23) follows from this estimate and (3.25).

Acknowledgment. The motivation behind this research was supplied by correspondence with Dr. A. W. Bowman and conversations with Professor D. M. Titterington. The referee's detailed comments have resulted in many clarifications.

Note added in proof. A treatment of the multivariate case by the author will appear in *Proceedings of the Sixth International Symposium on Multivariate Analysis*, Pittsburgh, 25–29 July 1983.

REFERENCES

BOWMAN, A. W. (1980). A note on consistency of the kernel method for the analysis of categorical data. *Biometrika* **67** 682–684.

BOWMAN, A. W. (1982). A comparative study of some kernel-based non-parametric density estimators. Manchester-Sheffield School of Probability and Statistics Research Report No. 84/AWB/1.

BOWMAN, A. W. (1983). An alternative method of cross-validation for the smoothing of density estimates. *Biometrika*, to appear.

CHOW, Y. S., GEMAN, S. and WU, L.-D. (1983). Consistent cross-validated density estimation. *Ann. Statist.* **11** 25–28.

DOOB, J. L. (1953). *Stochastic Processes.* Wiley, New York.

DUIN, R. P. W. (1976). On the choice of smoothing parameters for Parzen estimators of probability density functions. *IEEE Trans. on Computers* **C-25** 1175–1179.

HABBEMA, J. D. F., HERMANS, J. and VAN DEN BROEK, K. (1974). A stepwise discriminant analysis program using density estimation. In *Compstat* 1974, ed. G. Bruckman, 101–110. Physica Verlag, Vienna.

HALL, P. (1982a). Cross-validation in density estimation. *Biometrika* **69** 383–390.

HALL, P. (1982b). Limit theorems for stochastic measures of the accuracy of density estimators. *Stochastic Process Appl.* **13** 11–25.

KOMLÓS, J., MAJOR, P. and TUSNÁDY, G. (1975). An approximation of partial sums of independent random variables, and the sample distribution function. I. *Z. Wahrsch. verw. Gebiete* **32** 111–131.

KRIEGER, A. M. and PICKANDS, J. III. (1981). Weak convergence and efficient density estimation at a point. *Ann. Statist.* **9** 1066–1078.

MARCUS, M. (1970). A bound for the distribution of the maximum of a continuous Gaussian process. *Ann. Math. Statist.* **41** 305–309.

MARRON, S. (1983). An asymptotically efficient solution to the bandwidth problem of kernel density estimation. Abstract 83t–8. *IMS Bulletin* **12** 80.

ROSENBLATT, M. (1971). Curve estimates. *Ann. Math. Statist.* **42** 1815–1842.

RUDEMO, M. (1982). Empirical choice of histograms and kernel density estimators. *Scand. J. Statist.* **9** 65–78.

SCHUSTER, E. F. and GREGORY, C. G. (1981). On the nonconsistency of maximum likelihood nonparametric density estimators. In: *Computer Science and Statistics: Proceeding of the 13th Symposium on the Interface.* Ed. W. F. Eddy, 295–298. Springer-Verlag, New York.

SILVERMAN, B. W. (1978). Weak and strong uniform consistency of the kernel estimate of a density and its derivations. *Ann. Statist.* **6** 177–184.

STONE, M. (1974). Cross-validation and multinomial prediction. *Biometrika* **61** 509–515.

STONE, M. (1977). Asymptotics for and against cross-validation. *Biometrika* **64** 29–35.

TITTERINGTON, D. M. (1978). Contribution to discussion of paper by T. Leonard. *J. Roy. Statist. Soc.* **B40** 139–140.

TITTERINGTON, D. M. (1980). A comparative study of kernel-based density estimates for categorical data. *Technometrics* **22** 259–268.

DEPARTMENT OF STATISTICS
THE FACULTY OF ECONOMICS
THE AUSTRALIAN NATIONAL UNIVERSITY
G.P.O. BOX 4, CANBERRA
A.C.T. 2601, AUSTRALIA