The Annals of Statistics
1992, Vol. 20, No. 3, 1647–1657

ASYMPTOTICS FOR LEAST SQUARES CROSS-VALIDATION BANDWIDTHS IN NONSMOOTH CASES

BY BERT VAN ES

University of Amsterdam

We consider the problem of bandwidth selection for kernel density estimators. Let H_n denote the bandwidth computed by the least squares cross-validation method. Furthermore, let H_n^* and h_n^* denote the minimizers of the integrated squared error and the mean integrated squared error, respectively. The main theorem establishes asymptotic normality of $H_n - H_n^*$ and $H_n - h_n^*$, for three classes of densities with comparable smoothness properties. Apart from densities satisfying the standard smoothness conditions, we also consider densities with a finite number of jumps or kinks. We confirm the $n^{-1/10}$ rate of convergence to 0 of the relative distances $(H_n - H_n^*)/H_n^*$ and $(H_n - h_n^*)/h_n^*$ derived by Hall and Marron in the smooth case. Unexpectedly, in turns out that these relative rates of convergence are faster in the nonsmooth cases.

1. Introduction and results. Let $X_1, \dots, X_n$ be a random sample from a distribution on the real line with an unknown density f. We consider the problem of selecting a bandwidth h for a kernel estimator

$$f_{nh}(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right) \tag{1.1}$$

of the density f. Here the bandwidth h is a parameter which determines the amount of smoothness of the estimator. For small h we get a very rough estimate and for large h a smooth estimate. The kernel K is supposed to be a probability density function. The major problem in bandwidth selection is the fact that bandwidths which are asymptotically optimal in some sense tend to depend on the unknown density f. For rough densities a relatively small bandwidth is required and for smooth densities a relatively large one.

In recent years several *automatic* or *data-driven* bandwidth selection methods have been introduced. These methods compute, a hopefully good, bandwidth from the sample itself. Thus the resulting bandwidth is random. The methods we shall discuss here are based on the cross-validation principle. In particular, we shall discuss the least squares cross-validation method introduced by Rudemo (1982) and Bowman (1984). They proposed to compute the bandwidth H_n which minimizes an unbiased estimate, $\text{LS}_n(h)$, of the mean integrated squared error $\text{MISE}_n(h) = E\,\text{ISE}_n(h)$, minus a quantity not depending on the bandwidth h. Here the integrated squared error ISE_n is

Received November 1990; revised September 1991.
AMS 1980 *subject classifications*. Primary 62G05; secondary 62E20.
Key words and phrases. Density estimation, kernel estimators, bandwidth selection, cross-validation, rates of convergence.

defined by

$$\mathrm{ISE}_n(h) = \int_{-\infty}^{\infty} (f_{nh}(x) - f(x))^2\, dx, \tag{1.2}$$

and the least squares cross-validation criterion function $\mathrm{LS}_n(h)$ by

$$\mathrm{LS}_n(h) = \int_{-\infty}^{\infty} f_{nh}^2(x)\, dx - 2\frac{1}{n}\sum_{i=1}^{n} f_{nhi}(X_i), \tag{1.3}$$

where f_{nhi}, $i = 1, \dots, n$, are the "leave one out estimators,"

$$f_{nhi}(x) = \frac{1}{(n-1)h}\sum_{j\neq i}^{n} K\left(\frac{x - X_j}{h}\right), \qquad i = 1, \dots, n. \tag{1.4}$$

It is readily shown that

$$E\,\mathrm{LS}_n(h) = E\,\mathrm{ISE}_n(h) - \int_{-\infty}^{\infty} f^2(x)\, dx, \tag{1.5}$$

so the value of h which minimizes $\mathrm{LS}_n(h)$ indeed also minimizes an unbiased estimate of the mean integrated squared error of f_{nh}.

Stone (1984) has shown that, up to first order, asymptotically the least squares cross-validation bandwidths perform just as well as the best possible deterministic ones, under the sole assumption that f be bounded. Later Hall and Marron (1987a, b), under the assumption that f is sufficiently smooth, established a disappointingly slow rate of convergence to 0 of $H_n - H_n^*$ and $H_n - h_n^*$, where H_n^* is the minimizer of the integrated squared error (1.2) and h_n^* is the minimizer of its expected value, the mean integrated squared error. Our main result, stated below, establishes the asymptotic normality of both $H_n - H_n^*$ and $H_n - h_n^*$ for densities f classified into three groups with different smoothness properties. Thus we extend the results of Hall and Marron (1987a) to classes of nonsmooth densities. Doing so we shall encounter a remarkable paradox which shall be further discussed after we have stated our main theorem.

To introduce the three classes of densities mentioned above we first impose the following conditions.

CONDITION F. The density f has a bounded support. The first and second derivatives of f, denoted by f' and f'', exist at every point of the real line except at a finite set of points which we denote by D. In these points we give f' and f'' arbitrary values. The functions f, f' and f'' have finite left and right limits at the points in D. The function f has finite left and right first and second derivatives at the points in D. The second derivative f'' is continuous on the complement of D.

Now, writing $\delta^{(0)}(d) = f(d+) - f(d-)$ and $\delta^{(1)}(d) = f'(d+) - f'(d-)$, we introduce the quantities

$$\Delta^{(0)} = \sum_{d \in D} \delta^{(0)}(d)^2 \quad \text{and} \quad \Delta^{(1)} = \sum_{d \in D} \delta^{(1)}(d)^2.$$

We consider densities satisfying Condition F and one of the following three conditions:

$$\Delta^{(0)} > 0, \tag{1.6}$$

$$\Delta^{(0)} = 0 \quad \text{and} \quad \Delta^{(1)} > 0, \tag{1.7}$$

$$\Delta^{(0)} = \Delta^{(1)} = 0. \tag{1.8}$$

Roughly speaking, under (1.6) we allow the densities to have a finite number of jumps, while under (1.7) we at the most allow kinks. Condition (1.8) corresponds to the usual classical smoothness conditions.

Since it is generally recognized that the choice of the kernel is of less importance for the performance of a kernel estimator than the choice of bandwidth, we feel free to impose the following conditions on the kernel. Note that in this paper we do not consider higher-order kernels.

CONDITION K. The kernel function K is a symmetric, differentiable probability density function with support $[-1, 1]$.

Before we can state our main theorem we have to give some definitions.

DEFINITION 1.1. For G a bounded symmetric measurable function with bounded support, the functions b^G, b_0^G, b_1^G and b_2^G are defined by

$$b^G(x, h) = \frac{1}{h} \int_{-\infty}^{\infty} G\left(\frac{x-u}{h}\right) f(u)\, du - f(x) \int_{-\infty}^{\infty} G(u)\, du,$$

$$b_m^G(t) = \begin{cases} \int_{-\infty}^{t} (t-u)^m G(u)\, du, & \text{for } t < 0, \\ -\int_{t}^{\infty} (t-u)^m G(u)\, du, & \text{for } t \geq 0, \end{cases}$$

for $m = 0$, 1 or 2.

The functions introduced in Definition 1.1 play an important role in kernel estimation theory in nonsmooth cases. For G equal to the kernel K of a kernel estimator, the function b^G is equal to the bias of the estimator, that is, we have $b^K(x, h) = Ef_{nh}(x) - f(x)$. The following constants appear in the asymptotic variances in Theorem 1.3.

DEFINITION 1.2.

$$\alpha_0 = \int_{-1}^{1} K^2(u)\,du,$$

$$\alpha_1 = \Delta^{(0)}\int_{-1}^{1} b_0^K(t)^2\,dt,$$

$$\alpha_2 = 3\Delta^{(1)}\int_{-1}^{1} b_1^K(t)^2\,dt,$$

$$\alpha_3 = \left(\int_{-1}^{1} u^2 K(u)\,du\right)^2 \int_{-\infty}^{\infty} f''(x)^2\,dx,$$

$$\sigma^2(G) = \int_{-\infty}^{\infty} G^2(u)\,du \int_{-\infty}^{\infty} f^2(x)\,dx,$$

$$\sigma_1^2(G) = \frac{\alpha_0}{\alpha_1}\int_0^{\infty} b_0^G(t)^2\,dt \sum_{d\in D} \delta^{(0)}(d)^2(f(d-)+f(d+)),$$

$$\sigma_2^2(G) = \frac{\alpha_0}{\alpha_2}\int_0^{\infty} b_1^G(t)^2\,dt \sum_{d\in D} \delta^{(1)}(d)^2(f(d-)+f(d+)),$$

$$\sigma_3^2(g) = \frac{1}{4}\frac{\alpha_0}{\alpha_3}\left(\int_{-\infty}^{\infty} u^2 G(u)\,du\right)^2\left(\int_{-\infty}^{\infty} f''(x)^2 f(x)\,dx - \left(\int_{-\infty}^{\infty} f''(x) f(x)\,dx\right)^2\right).$$

We shall need these constants for G equal to special functions L and M given by

$$L(x) = K(x) + xK'(x) \quad \text{and} \quad M(x) = -K * L(x) + L(x), \tag{1.9}$$

with $*$ denoting convolution. Note that L and M are symmetric about 0 and that their integrals vanish.

Our main theorem establishes the asymptotic normality of $H_n - H_n^*$ and $H_n - h_n^*$ for the three classes of densities specified by assumptions (1.6), (1.7) and (1.8) above.

THEOREM 1.3. *Assume that Conditions* F *and* K *are satisfied. Under* (1.6), (1.7) *and* (1.8) *we have, respectively,*

$$n^{3/4}(H_n - H_n^*) \to_{\mathscr{D}} N\big(0, \alpha_0^{-5/2}\alpha_1^{5/2}(2\sigma^2(L) + \sigma_1^2(L))\big),$$

$$n^{3/8}(H_n - H_n^*) \to_{\mathscr{D}} N\big(0, \tfrac{1}{4}\alpha_0^{-5/4}\alpha_2^{-3/4}(2\sigma^2(L) + \sigma_2^2(L))\big),$$

$$n^{3/10}(H_n - H_n^*) \to_{\mathscr{D}} N\big(0, \tfrac{4}{25}\alpha_0^{-7/5}\alpha_3^{-3/5}(2\sigma^2(L) + \sigma_3^2(L))\big).$$

The same results hold for $H_n - h_n^*$, *provided we replace* $\sigma^2(L)$ *by* $\sigma^2(M)$ *and* $\sigma_1^2(L)$, $\sigma_2^2(L)$ *and* $\sigma_3^2(L)$ *by* $\sigma_1^2(2M)$, $\sigma_2^2(2M)$ *and* $\sigma_3^2(2M)$.

Notice that the third part of the theorem, which deals with densities satisfying the classical smoothness conditions, corresponds to Theorem 2.1 in Hall and Marron (1987a).

COROLLARY 1.4. *Assume that Conditions* F *and* K *are satisfied. Under* (1.6), (1.7) *and* (1.8) *we have, respectively,*

$$(H_n - H_n^*)/H_n^* = O_p(n^{-1/4}),$$

$$(H_n - H_n^*)/H_n^* = O_p(n^{-1/8}),$$

$$(H_n - H_n^*)/H_n^* = O_p(n^{-1/10}).$$

The same results hold for the relative distances $(H_n - h_n^*)/h_n^*$.

Now let us compare these rates of convergence with the minimax results in Hall and Marron (1991). Assuming that the second derivative of f exists, they show that the rate of convergence of the relative of a bandwidth selector with respect to H_n^* can never be faster than $n^{-1/10}$. Hall, Marron and Park (1989) show that the deterministic bandwidth h_n^*, the minimizer of the mean integrated squared error, can be estimated at a faster rate, but only when even more smoothness is assumed.

Clearly, with less smoothness, the relative rates of convergence of the cross-validation bandwidths given by the corollary get better, which is a rather uncommon phenomenon in functional estimation. It should be noted, however, that, in spite of the better properties of the selected bandwidth in the less smooth cases, the resulting density estimate will have a larger error compared to estimates of smooth densities (see Remark 1.6).

REMARK 1.5. By the same techniques as employed in the proof of Theorem 1.3, together with the Cramér–Wold device, joint asymptotic normality of $(H_n - H_n^*, H_n - h_n^*)$ can be derived. In the smooth case this corresponds to Remark 2.3 in Hall and Marron (1987a).

REMARK 1.6. The performance of kernel estimators of nonsmooth densities has been studied by Van Eeden (1985), Cline and Hart (1991), Swanepoel (1987) and van Es (1991a, b). It turns out that the existence of jumps or kinks of the density f causes the bias of a kernel estimator to be of different order compared to the well-known order h^4 in the smooth case. This leads to the following expansion of the mean squared error. We have

$$\mathrm{MISE}_n(h_n) = \frac{1}{nh_n}\int_{-1}^{1} K^2(u)\,du + O\left(\frac{1}{n}\right)$$

$$+\begin{cases} h_n \Delta^{(0)} \int_{-1}^{1} b_0^K(t)^2\,dt + o(h_n), & \text{under (1.6)},\\ h_n^3 \Delta^{(1)} \int_{-1}^{1} b_1^K(t)^2\,dt + o(h_n^3), & \text{under (1.7)},\\ \frac{1}{4} h_n^4 \left(\int_{-1}^{1} u^2 K(u)\,du\right)^2 \int_{-\infty}^{\infty} f''(x)^2\,dx + o(h_n^4), & \\ \quad \text{under (1.8)}. & \end{cases}$$

These expansions in turn yield the following asymptotically optimal bandwidths h_n^*:

$$h_n^* = \begin{cases} (\alpha_0/\alpha_1)^{1/2} n^{-1/2}, & \text{under (1.6)},\\ (\alpha_0/\alpha_2)^{1/4} n^{-1/4}, & \text{under (1.7)},\\ (\alpha_0/\alpha_3)^{1/5} n^{-1/5}, & \text{under (1.8)}. \end{cases} \tag{1.10}$$

Furthermore, the minimizer of the integrated squared error, H_n^*, is almost surely asymptotically equivalent to the minimizer of the mean integrated squared error. So, for the densities we consider here, H_n^* is almost surely asymptotically equivalent to the optimal bandwidth, a fact which we shall need in the next section.

Plugging in these bandwidths in the expansion of the mean integrated squared error yields the orders $n^{-1/2}$, $n^{-3/4}$ and $n^{-4/5}$, under (1.6), (1.7) and (1.8), respectively. An important feature is that these orders decrease as the smoothness of f increases. So, although the optimal bandwidths can be estimated better in the nonsmooth cases, the error of a kernel estimator with such an estimated bandwidth will be larger.

REMARK 1.7. The phenomenon of a faster relative rate of convergence for less smooth densities also occurs in the case of likelihood cross-validation. It is shown in van Es (1991a, b) that under (1.7) and (1.8) the relative rates with respect to the minimizer of a suitably weighted integrated squared error are of order $O_p(n^{-1/8})$ and $O_p(n^{-1/10})$, respectively, just as in Corollary 1.4. Under (1.6), when the density f has jumps, the likelihood cross-validation bandwidth does no longer asymptotically minimize the integrated squared error, nor the mean integrated squared error.

2. Proofs. To prove Theorem 1.3, we follow the arguments of Hall and Marron (1987a), adjusted for the possible nonsmoothness of the density f. We shall give the proof in outline and focus on the points where the essential differences occur with the smooth case. Let $D_n^{(1)}(h)$ and $D_n^{(2)}(h)$ denote the first and second derivative of the least squares cross-validation criterion function LS_n with respect to the bandwidth h. By the mean value theorem we have

$$D_n^{(1)}(H_n) - D_n^{(1)}(H_n^*) = -D_n^{(1)}(H_n^*) = D_n^{(2)}(\tilde{H}_n)(H_n - H_n^*) \tag{2.1}$$

for some random variable $\tilde{H}_n$ between H_n and H_n^*. Introducing a similar random variable $\overline{H}_n$, the distance $H_n - h_n^*$ can be treated analogously. We get

$$H_n - H_n^* = -\frac{D_n^{(1)}(H_n^*)}{D_n^{(2)}(\tilde{H}_n)} \tag{2.2}$$

and

$$H_n - h_n^* = -\frac{D_n^{(1)}(h_n^*)}{D_n^{(2)}(\overline{H}_n)}.$$

The two main ingredients of the proof are almost sure expansions of the denominators in (2.2) and asymptotic normality of the numerators. Theorem 1.3 follows from the next two lemmas.

LEMMA 2.1. *Assume that Conditions* F *and* K *are satisfied. Under* (1.6), (1.7) *and* (1.8) *we have, respectively,*

$$D_n^{(2)}(\tilde{H}_n) \sim 2\alpha_0^{1/2}\alpha_1^{-1/2}n^{1/2} \quad \text{almost surely},$$

$$D_n^{(2)}(\tilde{H}_n) \sim 4\alpha_0^{1/4}\alpha_2^{3/4}n^{-1/4} \quad \text{almost surely},$$

$$D_n^{(2)}(\tilde{H}_n) \sim 5\alpha_0^{2/5}\alpha_3^{3/5}n^{-2/5} \quad \text{almost surely}.$$

The same results hold for $D_n^{(2)}(\tilde{H}_n)$.

LEMMA 2.2. *Assume that Conditions* F *and* K *are satisfied. Under* (1.6), (1.7) *and* (1.8) *we have, respectively,*

$$n^{1/4}D_n^{(1)}(H_n^*) \to_{\mathscr{D}} N\Big(0, 4(\alpha_0/\alpha_1)^{-3/2}\big(2\sigma^2(L) + \sigma_1^2(L)\big)\Big),$$

$$n^{5/8}D_n^{(1)}(H_n^*) \to_{\mathscr{D}} N\Big(0, 4(\alpha_0/\alpha_2)^{-3/4}\big(2\sigma^2(L) + \sigma_2^2(L)\big)\Big),$$

$$n^{7/10}D_n^{(1)}(H_n^*) \to_{\mathscr{D}} N\Big(0, 4(\alpha_0/\alpha_3)^{-3/5}\big(2\sigma^2(L) + \sigma_3^2(L)\big)\Big).$$

The same results hold for $D_n^{(1)}(h_n^*)$, *provided we replace* $\sigma^2(L)$ *by* $\sigma^2(M)$ *and* $\sigma_1^2(L)$, $\sigma_2^2(L)$ *and* $\sigma_3^2(L)$ *by* $\sigma_1^2(2M)$, $\sigma_2^2(2M)$ *and* $\sigma_3^2(2M)$.

The almost sure rates presented in Lemma 2.1 can be obtained by plugging in the optimal bandwidths (1.11) in the second derivative of the mean integrated squared error expansions in Remark 1.6. Since a rigorous proof only contains standard arguments, it is omitted. We proceed with the proof of Lemma 2.2.

PROOF OF LEMMA 2.2. Note that, using

$$\frac{d}{dh}\frac{1}{h}K\left(\frac{x}{h}\right) = -\frac{1}{h^2}K\left(\frac{x}{h}\right) - \frac{x}{h^3}K'\left(\frac{x}{h}\right) = -\frac{1}{h^2}L\left(\frac{x}{h}\right),$$

the derivative of $\mathrm{LS}_n(h)$ can be written as

$$\begin{aligned} D_n^{(1)}(h) = &-\frac{2}{n^2h^3}\sum_{i=1}^{n}\sum_{j=1}^{n}\int_{-\infty}^{\infty} L\left(\frac{u-X_i}{h}\right)K\left(\frac{u-X_j}{h}\right)du \\ &+\frac{2}{n(n-1)h^2}\sum_{i\neq j} L\left(\frac{X_i-X_j}{h}\right). \end{aligned} \tag{2.3}$$

The fact that H_n^* minimizes $\mathrm{ISE}_n(h)$ implies that at the point $h = H_n^*$ the derivative of $\mathrm{ISE}_n(h)$ is equal to 0. A straightforward computation shows that, for $h = H_n^*$, we have

$$\begin{aligned} &\frac{2}{n^2h^3}\sum_{i=1}^{n}\sum_{j=1}^{n}\int_{-\infty}^{\infty} L\left(\frac{u-X_i}{h}\right)K\left(\frac{u-X_j}{h}\right)du \\ &= \frac{2}{nh^2}\sum_{i=1}^{n}\int_{-\infty}^{\infty} L\left(\frac{x-X_i}{h}\right)f(x)\,dx. \end{aligned} \tag{2.4}$$

Using (2.4) to rewrite (2.3), we get $D_n^{(1)}(H_n^*) = U_n(H_n^*)$, where $U_n(h)$ is equal to

$$\frac{2}{n(n-1)h^2}\sum_{i\neq j} L\left(\frac{X_i-X_j}{h}\right) - \frac{2}{nh^2}\sum_{i=1}^{n}\int_{-\infty}^{\infty} L\left(\frac{x-X_i}{h}\right)f(x)\,dx. \tag{2.5}$$

Next we rewrite (2.3) in the more convenient form $D_n^{(1)}(h) = V_n(h) + \tilde{V}_n(h)$, where $V_n(h)$ and $\tilde{V}_n(h)$ are statistics of the type

$$T_n(h) = \sum_{i\neq j} G\left(\frac{X_i-X_j}{h}\right) + \sum_{i=1}^{n} g_n(X_i). \tag{2.6}$$

For $V_n(h)$ the function G is equal to $2n^{-2}h^{-2}M$ and $g_n = 0$, while for $\tilde{V}_n(h)$ the function G equals $2n^{-2}(n-1)^{-1}h^{-2}L$ and g_n is equal to the constant $-2n^{-2}h^{-2}\int KL$. The statistic $U_n(h)$ is also of the type (2.6).

The next proposition establishes asymptotic normality of such statistics. The proof is a direct consequence of Theorem 2.2 in Jammalamadaka and Janson (1986). For the details of the proof, see van Es [(1991a), Appendix C]. Alternative approaches are provided by Hall (1984), de Jong (1987, 1990) and Nolan and Pollard (1987, 1988).

PROPOSITION 2.3. *Let f be a bounded almost everywhere continuous density and let the functions G and g_n also be bounded. Furthermore, assume that G is symmetric and integrable. Let the statistic $T_n(h)$ be defined by* (2.6) *and let (h_n) be a sequence of positive bandwidths converging to* 0, *such that $nh_n \to \infty$.*

Let the function g_n^* be defined by

$$g_n^*(x) = 2(n-1)\int_{-\infty}^{\infty} G\left(\frac{x-y}{h}\right) f(y)\,dy + g_n(x), \tag{2.7}$$

and suppose that this function satisfies

$$\frac{1}{nh_n^{1/2}} \sup_x \left|g_n^*(x) - Eg_n^*(X_1)\right| \to 0, \tag{2.8}$$

$$\frac{1}{nh_n} \operatorname{var}(g_n^*(X_1)) \to \alpha^2, \qquad 0 < \alpha^2 < \infty. \tag{2.9}$$

Then

$$\frac{1}{nh_n^{1/2}}(T_n(h_n) - ET_n(h_n)) \to_{\mathscr{D}} N(0, 2\sigma^2(G) + \alpha^2),$$

with

$$\sigma^2(G) = \int_{-\infty}^{\infty} G^2(v)\,dv \int_{-\infty}^{\infty} f^2(x)\,dx.$$

Recall that we have $D_n^{(1)}(H_n^*) = U_n(H_n^*)$, with $U_n(h)$ as in (2.5). In Remark 1.6 we have seen that H_n^* is almost surely asymptotically equivalent to the deterministic asymptotic minimizer of the mean integrated squared error, that is, $H_n^* \sim h_n^*$, with h_n^* as in (1.10). In order to derive the asymptotic distribution of $U_n(h_n^*)$, we rescale U_n in a form suitable for Proposition 2.3. Consider the statistic $T_n(h) = n(n-1)h^2U_n(h)/2$. Now $T_n(h)$ is of the form (2.6) with G equal to L and the function g_n^*, defined in (2.7), equal to

$$g_n^*(x) = (n-1)\int_{-\infty}^{\infty} L\left(\frac{y-X_i}{h}\right) f(y)\,dy = (n-1)hb^L(x,h).$$

We first derive the asymptotic distribution of $T_n(h_n^*)$ under (1.6). Recall that in that case $h_n^* \sim (\alpha_0/\alpha_1)^{1/2}n^{-1/2}$. Condition (2.8) of Proposition 2.3 is readily verified. To compute the constant α^2 in (2.9), under (1.6), write, with $h = h_n^*$,

$$Eg_n^*(X_1)^2 = (n-1)^2h^2Eb^L(X_1,h)^2 \sim (nh)^2\int_{-\infty}^{\infty} b^L(x,h)^2 f(x)\,dx. \tag{2.10}$$

Now let D be equal to the set of singular points of f in Condition F. Using Lemma 2.2 in van Es (1991b), the integral in (2.10) can be expanded as follows:

$$\int_{-\infty}^{\infty} b^L(x,h)^2 f(x)\,dx \sim h\int_0^1 b_0^L(t)^2\,dt \sum_{d\in D} \delta^{(0)}(d)^2(f(d-) + f(d+)).$$

Since the squared expectation of $g_n^*(X_1)$ turns out to be negligible, we see that under (1.6) we have

$$\frac{1}{nh_n^*}\operatorname{var}(g_n^*(X_1)) \to \frac{\alpha_0}{\alpha_1}\int_0^1 b_0^L(t)^2\,dt \times \sum_{d\in D}\delta^{(0)}(d)^2(f(d-)+f(d+)), \tag{2.11}$$

so the asymptotic variance α^2 in (2.9) is equal to the right-hand side of (2.11), which we shall denote by $\sigma_1^2(L)$. Proposition 2.3 now gives

$$\frac{1}{n(h_n^*)^{1/2}}T_n(h_n^*) \to_{\mathscr{D}} N\big(0, 2\sigma^2(L) + \sigma_1^2(L)\big).$$

For $U_n(h_n^*)$, using $h_n^* \sim (\alpha_0/\alpha_1)^{1/2}n^{-1/2}$, we now get

$$n^{1/4}U_n(h_n^*) \to_{\mathscr{D}} N\Big(0, 4(\alpha_0/\alpha_1)^{-3/2}\big(2\sigma^2(L) + \sigma_1^2(L)\big)\Big). \tag{2.12}$$

The asymptotic distribution of $U_n(h_n^*)$ under (1.7) can be derived similarly. Under (1.8) standard arguments for the smooth case can be used. Both cases are not treated in detail here.

To prove the statements concerning $D_n^{(1)}(H_n^*)$ in Lemma 2.2, it remains to show that h_n^* can be replaced by H_n^*, without disturbing the asymptotics. For the smooth case, that is, under (1.8), this is shown in Hall and Marron (1987a). By the same techniques it can also be shown to be true under assumptions (1.6) and (1.7).

Following the same steps in the derivation above, one can prove asymptotic normality of $D_n^{(1)}(h_n^*)$. First, consider the statistic $n^2h^2V_n(h)/2$, which is of the form (2.6), with G equal to M and g_n equal to 0. Applying Proposition 2.3 in this case, we have

$$g_n^*(x) = 2(n-1)hb^M(x,h) = (n-1)hb^{2M}(x,h),$$

so we get the same asymptotics for $n^2h^2V_n(h)/2$ as for $n^2h^2U_n(h)/2$ above, provided we replace $\sigma^2(L)$ by $\sigma^2(M)$ and $\sigma_1^2(L)$, $\sigma_2^2(L)$ and $\sigma_3^2(L)$ by $\sigma_1^2(2M)$, $\sigma_2^2(2M)$ and $\sigma_3^2(2M)$. The proof is completed by noting the fact that the contribution of $\tilde{V}_n(h_n^*)$ to the variation of $D_n^{(1)}(h_n^*)$ is asymptotically negligible. □

Acknowledgment. I would like to thank Steve Marron for suggesting the problem and pointing out its significance.

REFERENCES

BOWMAN, A. W. (1984). An alternative method of cross-validation for the smoothing of density estimates. *Biometrika* **71** 353–360.

CLINE, D. B. H. and HART, J. D. (1991). Kernel estimation of densities with discontinuities or discontinuous derivatives. *Statistics* **22** 69–84.

DE JONG, P. (1987). A central limit theorem for generalized quadratic forms. *Probab. Theory Related Fields* **75** 261–277.

DE JONG, P. (1990). *Central Limit Theorems for Generalized Quadratic Forms*. *CWI Tract* **61**. CWI, Amsterdam.

HALL, P. (1984). Central limit theorem for integrated square error for multivariate density estimation. *J. Multivariate Anal.* **14** 1–16.

HALL, P. and MARRON, J. S. (1987a). Extent to which least squares cross-validation minimizes integrated squared error in nonparametric density estimation. *Probab. Theory Related Fields* **74** 567–581.

HALL, P. and MARRON, J. S. (1987b). On the amount of noise inherent in bandwidth selection for a kernel density estimator. *Ann. Statist.* **15** 163–181.

HALL, P. and MARRON, J. S. (1991). Lower bounds for bandwidth selection in density estimation. *Probab. Theory Related Fields* **90** 149–173.

HALL, P., MARRON, J. S. and PARK, B. (1989). Smoothed cross-validation. Preprint.

JAMMALAMADAKA, R. S. and JANSON, S. (1986). Limit theorems for a triangular scheme of U-statistics with applications to interpoint distances. *Ann. Probab.* **14** 1347–1358.

NOLAN, D. and POLLARD, D. (1987). U-processes: Rates of convergence. *Ann. Statist.* **15** 780–800.

NOLAN, D. and POLLARD, D. (1988). Functional limit theorems for U-processes. *Ann. Probab.* **16** 1291–1298.

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

STONE, C. J. (1984). An asymptotically optimal window selection rule for kernel density estimators. *Ann. Statist.* **12** 1285–1297.

SWANEPOEL, J. H. (1987). Optimal kernels when estimating non-smooth densities. *Comm. Statist. Theory Methods* **16** 1835–1848.

VAN EEDEN, C. (1985). Mean integrated squared error of kernel estimators when the density and its derivatives are not necessarily continuous. *Ann. Inst. Statist. Math.* **37** 461–472.

VAN ES, A. J. (1991a). *Aspects of Nonparametric Density Estimation. CWI Tract* **77**. CWI, Amsterdam.

VAN ES, A. J. (1991b). Likelihood cross-validation bandwidth selection for nonparametric kernel density estimators. *Nonparametric Statist.* **1** 83–110.

DEPARTMENT OF MATHEMATICS AND
COMPUTER SCIENCE
UNIVERSITY OF AMSTERDAM
PLANTAGE MUIDERGRACHT 24
1018TV AMSTERDAM
THE NETHERLANDS