The Annals of Statistics
1993, Vol. 21, No. 2, 694–710

ON PLUG-IN RULES FOR LOCAL SMOOTHING OF DENSITY ESTIMATORS

By Peter Hall

Mathematical Sciences Research Institute, Berkeley

Optimal local smoothing of a curve estimator requires knowledge of various derivatives of the curve in the neighbourhood of the point at which estimation is being conducted. One empirical approach to selecting the amount of smoothing is to employ pilot estimators to approximate those derivatives, and substitute the approximate values into an analytical formula for the desired local bandwidth. In the present paper we study how bandwidth choice for the pilot estimators affects the performance of the final estimator. Our conclusions are rather curious. Depending on circumstance, the pilot estimators should be substantially oversmoothed or undersmoothed, relative to the amount of smoothing that would be optimal if they were to be employed themselves for point estimation. Occasionally, the optimal amount of undersmoothing is so extreme as to render the pilot estimators inconsistent. Here, the resulting local bandwidth is asymptotically random; it is not asymptotic to a sequence of constants.

1. Introduction. In problems of nonparametric curve estimation, the optimal amount of smoothing depends on unknown characteristics of the curve, such as derivatives of the curve at the point of estimation. One way of estimating those characteristics is to construct one or more preliminary curve estimators, compute derivative estimators from those, and substitute back into the formula for the optimal smoothing parameter. This so-called plug-in approach to local, adaptive bandwidth selection is not new; see, for example, Woodroofe (1970), Krieger and Pickands (1981), Park and Marron (1990) and the references therein. However, very little advice is available on how the smoothing parameters should be chosen for the pilot estimators. In the present paper we remedy this deficiency. Our conclusions suggest that the pilot estimators should be either substantially oversmoothed or substantially undersmoothed, depending on circumstance, and relative to the amount of smoothing that would be appropriate if the primary purpose of the pilot estimators was point estimation. Sometimes the amount of undersmoothing may be so extreme as to render one of the point estimators inconsistent.

To describe our conclusions in more detail, let $\hat{f}(x|h)$ denote a kernel estimator of a density f at the point x, based on the kernel K (a known symmetric density) and bandwidth h. Details of the construction of $\hat{f}$ will be given in Section 2. In the sense of minimising mean squared error, the

Received April 1992; revised July 1992.
AMS 1991 *subject classifications*. Primary 62G07; secondary 62G20.
Key words and phrases. Adaptive estimation, bandwidth, density estimator, kernel estimator, local smoothing, nonparametric density estimator, plug-in rule, smoothing, variable bandwidth.

asymptotically optimal bandwidth for estimating f at x is given by

$$h = c_K f(x)^{1/5} |f''(x)|^{-2/5} n^{-1/5}, \tag{1.1}$$

where the constant c_K depends only on K:

$$c_K = \left(\int K^2\right)^{1/5} \left\{\int y^2 K(y)\, dy\right\}^{-2/5}.$$

See, for example, Rosenblatt (1971) and Silverman [(1986), page 103]. Replacing f and f'' in (1.1) by estimators $\tilde{f}$ and $\check{f}''$ we obtain a plug-in rule for computing an empirical bandwidth $\hat{h}$.

If $\tilde{f}$ and $\check{f}$ are kernel density estimators, then they depend on bandwidths, say h_1 for $\tilde{f}$ and h_2 for $\check{f}$. We shall derive an expansion of the mean squared error of $\hat{f}(x|\hat{h})$ which shows how choice of h_1 and h_2 influences over-all performance. In particular, we shall show that the estimator $\tilde{f}$ should be substantially oversmoothed if $f''(x) > 0$, or substantially undersmoothed if $f''(x) < 0$, relative to the optimal amount of smoothing for point estimation of f. In practice, since estimation of $f''(x)$ is a prerequisite for implementing the plug-in rule, it would often be possible to make a qualitative decision to oversmooth $\tilde{f}$ somewhat in the event that $\check{f}'' > 0$, and to undersmooth when $\check{f}'' < 0$. However, specific quantitative rules are not really possible, owing to the sheer complexity of this multiparameter smoothing problem.

The same principles apply to smoothing of $\check{f}''(x)$, except that here the appropriate amount of undersmoothing can occasionally be so extreme as to render the resulting estimator inconsistent for $f''(x)$. In this case the plug-in bandwidth $\hat{h}$ is asymptotically random; it is not asymptotic to a nonrandom sequence. In appropriate circumstances, employing a bandwidth defined in this way can improve on the size of mean squared error by a constant factor, although not by an order of magnitude.

Alternative, asymptotically random bandwidth selection procedures have been suggested by, for example, Abramson (1982). They are simpler than undersmoothing f'', and more effective in the sense that they do reduce the order of magnitude of mean squared error. However, the random bandwidth constructions given here are quite different from those considered by Abramson, in that they inherently depend significantly on a large number of data values and take the same value in each summand of the kernel estimator. They form an unexpected link between the two kinds of variable bandwidth discussed by Jones (1990). Our principal purpose in this paper is to describe the effect of bandwidth choice for pilot estimators when those estimators are consistent. The fact that inconsistent estimators can sometimes give better performance emerges as a somewhat pathological feature.

Section 2 states our main technical result on second-order expansion of mean squared error, and Section 3 discusses issues concerning asymptotically random bandwidths. A proof of the theorem in Section 2 is given in Section 4. Related results may be established for curve estimation by nonparametric regression, and for a variety of alternative approaches to nonparametric density estimation, such as histogram and histospline methods. Thus, the results in this paper might be seen as examples of more general phenomena, although it appears to be impossible to derive a result which, in a useful and meaningful way, embraces a wide range of contexts.

2. Main results. Let $X_1, \dots, X_n$ denote independent and identically distributed random variables from the population with density f, and let K, L and M be bounded, compactly supported kernel funtions. If N represents one of K, L or M, we assume that

$$\int y^j N(y)\, dy = \begin{cases} 1, & \text{if } j = 0, \\ 0, & \text{if } 1 \le j \le r-1, \\ \nu \neq 0, & \text{if } j = r, \end{cases}$$

where $r = 2$, l or m and $\nu = \kappa$, λ or μ according as $N = K$, L or M, respectively. In common parlance, K, L and M are kernels of orders 2, l and m. It is usual to take K to be a symmetric probability density with $K(0) > 0$, and l and m to be even integers, so we make those assumptions here. They allow us to simplify notation a little, but are otherwise inessential.

Assume that $l, m \ge 2$; that f has $\min(l, m+2)$ continuous derivatives in a neighbourhood of x; that $f(x) > 0$ and $f''(x) \neq 0$; and that K''' and M'' exist and are bounded. Two estimators of f, $\hat{f}$ and $\tilde{f}$, and one estimator of f'', are given by

$$\hat{f}(x|h) = (nh)^{-1} \sum_{i=1}^{n} K\{(x - X_i)/h\}, \quad \tilde{f}(x) = (nh_1)^{-1} \sum_{i=1}^{n} L\{(x - X_i)/h_1\},$$

$$\check{f}''(x) = (nh_2^3)^{-1} \sum_{i=1}^{n} M''\{(x - X_i)/h_2\}.$$

The quantities $\tilde{f}$ and $\check{f}''$ will be used to determine the bandwidth h for $\hat{f}(\cdot|h)$. We wish to ascertain how best to select the subsidiary bandwidths, h_1 and h_2, for this purpose.

Recall from Section 1 that the asymptotically optimal nonrandom bandwidth h_0 is given by $h_0 = c_K f(x)^{1/5} |f''(x)|^{-2/5} n^{-1/5}$, where c_K depends only on K and so is known. Our plug-in version of h_0 is defined by, essentially,

$$\hat{h} = c_K |\tilde{f}(x)|^{1/5} |\check{f}''(x)|^{-2/5} n^{-1/5}.$$

However, since L is compactly supported then for each $n \ge 1$, $\tilde{f}(x)$ takes the value zero with positive probability. This means that $P(\hat{h} = 0) > 0$ for all $n \ge 1$, and of course $\hat{f}(x|0)$ is not generally well defined. [On the other hand, the possibility that $\check{f}''(x)$ might vanish causes no difficulties, since $\hat{f}(x|\infty)$ is properly defined.] Therefore, we consider the following modified definition of

$\hat{h}$. Let $q > 0$ be arbitrary, and put

$$\hat{h} = \begin{cases} c_K \tilde{f}(x)^{1/5} |\check{f}''(x)|^{-2/5} n^{-1/5}, & \text{if } \tilde{f}(x) > n^{-q}, \\ n^{-1/5}, & \text{otherwise}. \end{cases}$$

Consistency of $\tilde{f}$ and $\check{f}''$ for f and f'', respectively, demands that $h_1, h_2 \to 0$ and $nh_1, nh_2^5 \to \infty$. We ask only a little more: that for some $\varepsilon > 0$,

$$h_1 + h_2 + (nh_1)^{-1} + \left(nh_2^5\right)^{-1} = O(n^{-\varepsilon}) \tag{2.1}$$

as $n \to \infty$.

Since mean squared error was used for the optimality criterion that produced h_0 as the "asymptotically best" nonrandom bandwidth, then that yardstick should also be used to assess the performance of $\hat{h}$. Thus, we seek the mean squared error of $\hat{f}(x|\hat{h})$, which is described by the following result.

THEOREM . *Assume the conditions above. Then as $n \to \infty$,*

$$E\left\{\hat{f}(x|\hat{h}) - f(x)\right\}^2 = E\left\{\hat{f}(x|h_0) - f(x)\right\}^2 + \alpha_n + o(\xi_n), \tag{2.2}$$

where, dropping the argument x in f, f'', $f^{(l)}$ and $f^{(m+2)}$,

$$\begin{aligned} \alpha_n = {} & \frac{1}{10} h_0^4 \left(\frac{1}{l!} h_1^l \lambda f^{-1} f'' f^{(l)} - \frac{2}{m!} h_2^m \mu f^{(m+2)} \right)^2 \kappa^2 \\ & + \frac{1}{10} h_0^4 (nh_1)^{-1} \kappa^2 f^{-1} f''^2 \int L^2 \\ & + \frac{1}{5} h_0^2 \Bigg[(nh_1)^{-1} f'' \int \{K(y) - yK'(y)\} L(h_0 y / h_1)\, dy \\ & \qquad - 4\left(nh_2^3\right)^{-1} \kappa M''(0) f \Bigg] \kappa, \end{aligned} \tag{2.3}$$

$$\xi_n = h_0^4 \{ h_1^{2l} + h_2^{2m} + (nh_1)^{-1} \} + h_0^2 \left\{ (nh_1)^{-1} \min(1, h_1/h_0) + \left(nh_2^3\right)^{-1} \right\}.$$

REMARK 2.1. If we neglect pathological cancellation, then the term α_n is of precise size ξ_n as $n \to \infty$. Thus (2.2) accurately describes asymptotic properties of the mean squared error of $\hat{f}(x|\hat{h})$ up to terms of smaller order than ξ_n.

REMARK 2.2. The quantity $E\{\hat{f}(x|h_0) - f(x)\}^2$, of size $n^{-4/5}$, is the first-order term in formula (2.2). Of course, it does not depend on h_1 or h_2. The second-order term α_n depends critically on the smoothing parameters h_1 and h_2. If those quantities are chosen appropriately then it is of smaller order than $n^{-4/5}$. Third- and higher-order terms are collected together in the remainder $o(\xi_n)$.

REMARK 2.3. Minimisation of α_n with respect to the first bandwidth h_1 is relatively straightforward. It takes the following form. If $f''(x) > 0$, then the optimal h_1 is asymptotic to a constant multiple of $n^{-3/\{5(2l+1)\}}$, which is an order of magnitude larger than the size $n^{-1/(2l+1)}$ that is suitable for point estimation of f using the estimator $\tilde{f}$. With $h_1 \sim \text{const.}\, n^{-3/\{5(2l+1)\}}$, the quantity α_n is of size $n^{-2(7l+2)/\{5(2l+1)\}} = o(n^{-4/5})$, provided m is sufficiently large and h_2 is chosen appropriately. ["Appropriate" choice of h_2 means, if $M''(0) < 0$, that $h_2 \sim \text{const.}\, n^{-3/\{5(2m+3)\}}$. See Remark 2.4.] In the case where $f''(x) < 0$, define

$$J(p) = \int \{K(py) - pyK'(py)\} L(y)\, dy$$

and choose $p_0 \in [0, \infty)$ to maximise $J(p)$. For appropriate choice of h_2, any h_1 that satisfies $h_1/h_0 \to p_0$ (and also, in the case $p_0 = 0$, $n^{3/5}h_1 \to \infty$) will produce $\alpha_n \sim (1/5)n^{-1}h_0 J(p_0) f''(x) < 0$, which is of size $n^{-6/5}$ and is asymptotically the "smallest" (i.e., largest negative, on this occasion) value that α_n can assume. Thus, when $f''(x) > 0$, oversmoothing of $\tilde{f}$ is optimal, whereas undersmoothing is required when $f''(x) < 0$. All these results may be derived by routine analytical methods, outlined in Minimisation Problem I.

REMARK 2.4. Minimisation of α_n with respect to the second bandwidth h_2 is straightforward when $M''(0) < 0$. The latter inequality typically holds in practice, since M is usually concave in a neighbourhood of the origin; consider, for example, a kernel of the form $M(y) = \text{const.}(1 - y^2)^r$, $r > 0$, for $|y| \le 1$, and $M(y) = 0$ for $|y| > 1$. In this circumstance, the optimal h_2 is asymptotic to a constant multiple of $n^{-3/\{5(2m+3)\}}$, which is an order of magnitude larger than the size $n^{-1/(2m+5)}$ that is appropriate for point estimation of f'' using the estimator $\tilde{f}''$. However, should $M''(0)$ be positive, h_2 should be chosen as small as possible subject to conditions required for consistency; see condition (2.1), and also Minimisation Problem II. An example where $M''(0) > 0$ is afforded by

$$M(x) = \tfrac{15}{16}\{1 - (x + \theta)^2\}^2 I(|x + \theta| \le 1) + \tfrac{15}{16}\{1 - (x - \theta)^2\}^2 I(|x - \theta| \le 1),$$

where $3^{-1/2} < \theta < 1/2$.

Next we treat the unusual case where $M''(0) = 0$. Should $f''f^{(l)}$ and $f^{(m+2)}$ be of opposite signs then the contributions from $h_1^{(l)}$ and h_2^m to the first term of α_n tend to reinforce one another. Thus, choosing $h_2 = o(h_1^{l/m})$ leads to minimisation of α_n. If $f''f^{(l)}$ and $f^{(m+2)}$ are of the same sign, then it is theoretically possible to choose h_2 so that the first term in α_n vanishes. However, this is hardly a practical suggestion, and so once again it would be advisable to select h_2 so that $h_2 = o(h_1^{l/m})$.

Consistency of $\tilde{f}''$ for f'' is crucial to the methods that we use to prove the theorem. However, it does not always produce a bandwidth estimator that is optimal in the mean squared error sense. This issue will be discussed in greater detail in the next section.

We conclude this discussion by describing two minimisation problems that elucidate the conclusions of Remarks 2.3 and 2.4.

MINIMISATION PROBLEM I. Minimise

$$(2.4)\qquad C_1 h_0^4 h_1^{2l} + C_2 h_0^4 (nh_1)^{-1} + C_3 h_0^2 (nh_1)^{-1} \min(1, h_1/h_0)$$

with respect to $h_1 > 0$, when $C_1, C_2 > 0$ and C_3 may be either positive or negative. We claim that if $C_3 > 0$, then the optimal h_1 satisfies $h_1 \sim C_4 n^{-3/\{5(2l+1)\}} \gg h_0$; the minimum is asymptotic to $C_5 n^{-2(7l+2)/\{5(2l+1)\}}$; and the second term in (2.4) plays a negligible role. If $C_3 \leq 0$, then the minimum of (2.4) is asymptotic to $C_3 h_0 n^{-1}$, and this asymptotic minimum is attained by any h_1 sequence satisfying $n^{-3/5} \ll h_1 \leq h_0$.

The claim in the case $C_3 < 0$ is clear, on reflection. To elucidate the claim when $C_3 > 0$, observe that the sum of the first two terms in (2.4) is minimised with h_1 of size $h_1' = n^{-1/(2l+1)}$, which, since $l \geq 2$, is of the same order as h_0 or larger. The minimum of this two-term sum is of size $n^{-2(9l+2)/\{5(2l+1)\}}$, but is dominated by the third term in (2.4), which is of size $n^{-2(7l+1)/\{5(2l+1)\}}$. When $C_3 > 0$, the sum of the first and third terms in (2.4) is minimised with h_1 of size $h_1'' = n^{-3/\{5(2l+1)\}}$, which is of larger order than h_0. The resulting two-term sum is of size $n^{-2(7l+2)/\{5(2l+1)\}}$, which dominates the second term in (2.4) (evaluated at $h_1 = h_1''$), and is of smaller order than $n^{-2(7l+1)/\{5(2l+1)\}}$.

MINIMISATION PROBLEM II. Minimise

$$(2.5)\qquad C_1 h_0^4 h_2^{2m} + C_2 h_0^2 (nh_2^3)^{-1}$$

with respect to $h_2 > 0$, where $C_1 > 0$ and C_2 may be either positive or negative. If $C_2 > 0$, then the optimal h_2 satisfies $h_2 \sim C_3 n^{-3/\{5(2m+3)\}}$, and the minimum is asymptotic to $C_4 n^{-2(7m+6)/\{5(2m+3)\}}$. If $C_2 < 0$ and $h_2 \geq h_2' = n^{-(1/5)+\varepsilon}$, where $0 < \varepsilon < 2m/\{5(2m+3)\}$, then the minimum of (2.5) is achieved with $h_2 = h_2'$ and is asymptotic to $C_2 h_0^2 (nh_2^3)^{-1}$.

3. Asymptotically random bandwidths.

3.1. *Introduction and summary.* In the case where our theory predicts that the estimator $\check{f}''$ should be substantially undersmoothed, we ask that h_2 be taken as small as possible subject to constraints required by consistency of $\check{f}''$ for f''. Thus, we ask that h_2 be of larger order than $n^{-1/5}$. Taking h_2, in formula (2.3), to be of size $n^{-1/5}$, and ignoring the fact that this extreme choice of h_2 invalidates the assumptions required for that result, we see that if $M''(0) > 0$ then the crucial term

$$-\tfrac{4}{5} h_0^2 (nh_2^3)^{-1} \kappa^2 M''(0) f(x)$$

in (2.3) is of the same size (viz., $n^{-4/5}$) as $E\{\hat{f}(x|h_0) - f(x)\}^2$. This suggests that such an extreme choice of h_2 might produce a significant reduction in the size of mean squared error, by a constant factor if not by an order of magnitude.

Generally, reductions in the order of magnitude of mean squared error are possible if one employs an asymptotically random bandwidth. This is clear from work of Abramson (1982), for example. However, Abramson's method employs a different bandwidth for each summand in the kernel density estimator, and so is very different from the undersmoothed plug-in rule being considered here. An alternative approach, using the same bandwidth for each summand but still reducing the order of magnitude of mean squared error, will be outlined in subsection 3.2. Variants of the plug-in rule suggested in Section 2, in the case where h_2 is chosen so that $\check{f}''$ is not consistent for f'', can reduce mean squared error by a constant factor although not, usually, by an order of magnitude. This will be shown in subsection 3.3.

There should be no misconception concerning the relative practical virtues of Abramson's method and our own; Abramson's is clearly more practical. In particular it tends to be more numerically stable. The point of the work in the present section is to elucidate the curious results about undersmoothing $\check{f}''$, encountered in Section 2, and to show that variable bandwidth methods quite different from Abramson's can, like Abramson's, reduce mean squared error.

To better appreciate what is occurring when $\check{f}''$ is substantially undersmoothed, using a bandwidth of size $h_2 = cn^{-1/5}$ for a constant $c > 0$, observe that in this case, $\check{f}''$ has an asymptotically normal $N\{f''(x), \sigma^2\}$ distribution, where $\sigma^2 = c^{-5}f(x)\int M''^2$. Therefore, the empirical bandwidth selector $\hat{h}$ admits the representation $\hat{h} = c_K n^{-1/5} f(x)^{1/5} |f''(x) + \zeta_n|^{-2/5}$, where the random variable ζ_n is asymptotically normal $N(0, \sigma^2)$. More generally, we might define $\hat{h} = hg(Z_1)$, where $h = c_K n^{-1/5}$, g is an appropriate function, and Z_1 is a random variable obtained by centring a kernel estimator. In the case discussed just above we have, essentially,

$$g(z) = f(x)^{1/5} |f''(x) + c^{-2}z|,$$

$$Z_1 = (cn^{4/5})^{-1/2} \sum_{i=1}^{n} \left[N\{(x - X_i)/cn^{-1/5}\} - EN\{(x - X_i)/cn^{-1/5}\} \right].$$

This is the context that we shall study in subsection 3.3.

3.2. *An asymptotically random bandwidth selector: first approach.* We begin by describing the decomposition of $\hat{f}$ into stochastic and deterministic terms. This expansion will form the basis of our first bandwidth selector.

The stochastic process U_n, defined by

$$U_n(t) = n^{-2/5} \sum_{i=1}^{n} \left[K\{(x - X_i)n^{1/5}t^{-1}\} - EK\{(x - X_i)n^{1/5}t^{-1}\} \right],$$

represents the standardised error about the mean for $\hat{f}(x|h)$, when the bandwidth is $h = n^{-1/5}t$. If K is continuous, then so are the sample paths of U_n, and for any $0 < a < b < \infty$, U_n converges weakly on the space $C[a, b]$ of continuous functions on $[a, b]$ to a Gaussian process U with continuous sample paths. See, for example, Silverman (1976, 1978) and Krieger and Pickands (1981). Now, the mean of $\hat{f}$ admits the usual expansion,

$$E\hat{f}(x|h) = f(x) + \tfrac{1}{2}h^2\kappa f''(x) + O(h^4)$$

provided $f^{(4)}$ exists and is bounded in a neighbourhood of x. Therefore,

$$
\begin{aligned}
\hat{f}(x|n^{-1/5}t) &= n^{-2/5}t^{-1}U_n(t) + E\{\hat{f}(x|n^{-1/5}t)\} \\
&= f(x) + n^{-2/5}\{t^{-1}U_n(t) + \tfrac{1}{2}t^2\kappa f''(x)\} + O(n^{-4/5}),
\end{aligned} \tag{3.1}
$$

the remainder here being nonstochastic and of order $n^{-4/5}$ uniformly in $a \le t \le b$, for any $0 < a < b < \infty$.

Next we use formula (3.1) to inspire an asymptotically random bandwidth selector $\tilde{h}$. Let $\pi_n(a, b), \pi(a, b)$ denote the respective probabilities that the equations

$$t^{-1}U_n(t) + \tfrac{1}{2}t^2\kappa f''(x) = 0, \qquad t^{-1}U(t) + \tfrac{1}{2}t^2\kappa f''(x) = 0$$

have solutions in $[a, b]$. If solutions exist, let T_n, T denote the respective solutions that are nearest to unity. The covariance of U is given by

$$\operatorname{cov}\{U(s), U(t)\} = f(x)\left\{\int K(z/s)K(z/t)\,dz - st\right\}.$$

It may be shown that $\pi_n(a, b) \to \pi(a, b)$ as $n \to \infty$; and, noting the covariance structure of U, that $\pi(a, b) \to 1$ as $a \to 0$ and $b \to \infty$. Hence, given $\varepsilon > 0$ we may choose a small and b large such that, for all sufficiently large n, $\pi_n(a, b) > 1 - \varepsilon$. Thus, T_n is well defined with probability at least $1 - \varepsilon$. Let us agree to take T_n, T equal to 1 if the variable is not otherwise defined. (Then $T_n \to T$ in distribution.) Put $\tilde{h} = n^{-1/5}T_n$.

Next we outline mean squared error properties when the bandwidth is taken to be $\tilde{h}$. In view of (3.1),

$$E\{\hat{f}(x|\hat{h}) - f(x)\}^2 \le n^{-4/5}(ES_n^4)^{1/2}\{1 - \pi_n(a, b)\}^{1/2} + o(n^{-4/5}), \tag{3.2}$$

where $S_n = U_n(1) + (1/2)\kappa f''(x)$ and has all moments bounded. Letting $a = a_n \to 0$ and $b = b_n \to \infty$ at a sufficiently slow rate, we may deduce from (3.2) that

$$E\{\hat{f}(x|\tilde{h}) - f(x)\}^2 = o(n^{-4/5}). \tag{3.3}$$

Thus, the convergence rate has been reduced from $n^{-4/5}$ to $o(n^{-4/5})$ by employing the asymptotically random bandwidth $\tilde{h}$ rather than an asymptotically constant bandwidth.

Of course, the bandwidth selector $\hat{h}$ is not a practical choice, since it requires the unknown density. We show next how to modify this approach to make it feasible. Note that the centring term,

$$EK\{(x - X_i)n^{1/5}t^{-1}\} = O(n^{-1/5}),$$

may be estimated with an error of only $O_p(n^{-1/5}n^{-(1/2)+\delta}) = O_p(n^{-(7/10)+\delta})$ for any $\delta > 0$, assuming enough smoothness of f in a neighbourhood of x. Similarly, $f''(x)$ may be estimated with error $O_p(n^{-(1/2)+\delta})$. Arguing in this way we may derive computable, empirical approximations $\tilde{U}_n$, $\tilde{f}''$ to U_n, f'', respectively, such that the following analogue of (3.1) holds:

$$\hat{f}(x|n^{-1/5}t) = f(x) + n^{-2/5}\{t^{-1}\tilde{U}_n(t) + \tfrac{1}{2}t^2\kappa\tilde{f}''(x)\} + R_n(t),$$

where for each $0 < a < b < \infty$, any integer $p \geq 1$, and $\delta > 0$,

$$E\Big\{\sup_{a \leq t \leq b} |R_n(t)|^p\Big\} = O(n^{-p\{(1/2)-\delta\}}).$$

The approximations may be constructed so that $\tilde{U}_n$ is continuous, the probability $\tilde{\pi}_n(a, b)$ that the equation

$$t^{-1}\tilde{U}_n(t) + \tfrac{1}{2}t^2\kappa\tilde{f}''(x) = 0 \tag{3.4}$$

admits a solution $t \in [a, b]$ converges to $\pi(a, b)$ as $n \to \infty$, and

$$E\{\tilde{U}_n(1) + \tfrac{1}{2}\kappa\tilde{f}''(x)\}^4 = O(1).$$

Defining $\tilde{T}_n$ to be the solution of (3.4) within $[a, b]$ that is nearest to 1, if a solution exists, and to equal 1 otherwise; putting $\hat{h} = n^{-1/5}\tilde{T}_n$; and letting $a_n \to 0$, $b_n \to \infty$ sufficiently slowly; we may deduce that this feasible bandwidth selector has property (3.3).

3.3. *An asymptotically random bandwidth selector: second approach.* Our second approach is more in keeping with the kernel-based plug-in rule considered in Section 2, although like our first approach, it produces a bandwidth selector that does not vary among the summands that comprise the kernel estimator. It does not generally allow the convergence rate of mean squared error (MSE) to be reduced from $n^{-4/5}$ to $o(n^{-4/5})$, although it can reduce the constant C in the formula MSE $\sim Cn^{-4/5}$. For the sake of brevity our argument will be heuristic, although it can be made rigorous under certain conditions on the function g, which contributes the stochastic component to the asymptotically random bandwidth. Appropriate conditions are that g be continuous, bounded away from zero, and equal to a nonzero constant outside an interval $[a, b]$, where $0 < a < b < \infty$. In these circumstances, a rigorous proof may be given via an Edgeworth expansion of the joint distribution of Z_1 and $Z_3(z)$, defined below.

We begin by introducing functions and random variables on which our bandwidth selector will be based. Let $g > 0$ and N be functions, with N bounded, symmetric and compactly supported. Let $h = cn^{-1/5}$ for a constant $c > 0$, and put

$$Z_1 = (nh)^{-1/2} \sum_{i=1}^{n} [N\{(x - X_i)/h\} - EN\{(x - X_i)/h\}],$$

$$Z_2 = \{nhg(Z_1)\}^{-1} \sum_{i=1}^{n} K[(x - X_i)/\{hg(Z_1)\}],$$

$$Z_3(z) = \{nhg(z)\}^{-1} \sum_{i=1}^{n} K[(x - X_i)/\{hg(z)\}].$$

Our bandwidth selector is $\hat{h} = hg(Z_1)$, and our density estimator is $\hat{f}(x|\hat{h}) = Z_2$.

Next we outline the mean and covariance properties of Z_1 and $Z_3(z)$, which lead to mean squared error properties of $\hat{f}(x|\hat{h})$. Define

$$a(z) = \int K(y)N\{g(z)y\}\,dy,$$

and observe that

$$EZ_3(z) = f(x) + \tfrac{1}{2}h^2 g(z)^2 \kappa f''(x) + O(h^4),$$

$$\sigma_1^2 = \mathrm{var}(Z_1) = f(x)\int N^2 + O(h^2),$$

$$\sigma_3(z)^2 = \mathrm{var}\{Z_3(z)\} = \{nhg(z)\}^{-1} f(x)\int K^2 + O(n^{-1}h),$$

$$\mathrm{cov}\{Z_1, Z_3(z)\} = (nh)^{-1/2} f(x)a(z) + O(n^{-1/2}h^{1/2}),$$

$$\rho(z) = \mathrm{corrln}\{Z_1, Z_3(z)\} = g(z)^{1/2}a(z)\left(\int K^2 \int N^2\right)^{-1/2} + O(h).$$

If Z_1, and Z_2 conditional on Z_1, had joint normal distributions then the following formulae would be valid:

$$\begin{aligned} E(Z_2|Z_1 = z) &= f(x) + \tfrac{1}{2}h^2 g(z)^2 \kappa f''(x) + \rho(z)\sigma_3(z)\sigma_1^{-1}z + O(h^4) \\ &= f(x) + \tfrac{1}{2}h^2 g(z)^2 \kappa f''(x) \\ &\quad + (nh)^{-1/2} a(z) z \left(\int N^2\right)^{-1} + O(n^{-1/2}), \end{aligned}$$

$$\begin{aligned} \mathrm{var}(Z_2|Z_1 = z) &= \{1 - \rho(z)^2\}\sigma_3(z)^2 \\ &= (nh)^{-1} f(x)\left\{g(z)^{-1}\int K^2 - a(z)^2\left(\int N^2\right)^{-1}\right\} + O(n^{-1}). \end{aligned}$$

Therefore,

$$\begin{aligned} E\left[\{Z_2 - f(x)\}^2 | Z_1 = z\right] &= \{E(Z_2|Z_1 = z) - f(x)\}^2 + \mathrm{var}(Z_2|Z_1 = z) \\ &= \left\{\tfrac{1}{2}h^2 g(z)^2 \kappa f''(x) + (nh)^{-1/2} a(z) z \left(\int N^2\right)^{-1}\right\}^2 \\ &\quad + (nh)^{-1} f(x)\left\{g(z)^{-1}\int K^2 - a(z)^2\left(\int N^2\right)^{-1}\right\} \\ &\quad + O(n^{-1}). \end{aligned}$$

Thus,

$$
\begin{aligned}
E\{\hat{f}(x|\hat{h}) - f(x)\}^2
&= E\{Z_2 - f(x)\}^2 \\
&\sim E\left\{\tfrac{1}{2}h^2 g(Z)^2 \kappa f''(x) + (nh)^{-1/2} a(Z) Z\left(\int N^2\right)^{-1}\right\}^2 \\
&\quad + (nh)^{-1} f(x)\left[\left(\int K^2\right) E\{g(Z)^{-1}\} - \left(\int N^2\right)^{-1} E\{a(Z)^2\}\right],
\end{aligned}
$$

where Z is normal $N\{0, f(x)\int N^2\}$.

Although this formula has been derived under the assumption that Z_1, and Z_2 given Z_1, have normal distributions, it is available more generally, as noted in the first paragraph of this section.

4. Proof of theorem. Let d_1, d_2 be constants such that $0 < d_1 < f(x)$ and either $0 < d_2 < f''(x)$ or $f''(x) < d_2 < 0$. Consider redefining $\hat{h}$ as

$$
\hat{h}' = \begin{cases} c_K \tilde{f}(x)^{1/5} |\check{f}''(x)|^{-2/5} n^{-1/5}, & \text{if } \tilde{f}(x) > d_1 \text{ and either} \\ & \quad 0 < d_2 < \check{f}''(x) \text{ or } \check{f}''(x) < d_2 < 0, \\ n^{-1/5}, & \text{otherwise}. \end{cases}
$$

Then $\hat{h} = \hat{h}'$ unless the event $\mathscr{E} = \{\tilde{f}(x) < d_1$, or $\check{f}''(x) < d_2$ and $d_2 > 0$, or $\check{f}''(x) > d_2$ and $d_2 < 0\}$ obtains. Bernstein's, or Bennett's, or Hoeffding's or even Markov's inequality may be used to prove that $\mathscr{P}(\mathscr{E}) = O(n^{-C})$ for each $C > 0$. Observe that $\hat{h}^{-1}$ and $\hat{h}'^{-1}$ are both dominated by n^{C_1}, for some $C_1 = C_1(q) \geq 1/5$ sufficiently large. It follows that for each $C_2, C_3 > 0$,

$$
E\left\{\hat{f}(x|\hat{h})^{C_2} I(\mathscr{E})\right\} = O(n^{-C_3}),
$$

$$
E\left\{\hat{f}(x|\hat{h}')^{C_2} I(\mathscr{E})\right\} = O(n^{-C_3}).
$$

If we choose C_3 sufficiently large, then the right-hand sides of these identities are negligible relative to $\xi_n^{C_2/2}$. Hence, there is no essential loss of generality in proving the theorem for $\hat{h}'$ instead of $\hat{h}$. This we shall do, although we shall drop the dash from our notation.

For the sake of clarity, the remainder of our proof will be given in a sequence of seven steps.

STEP (i). *Taylor expansion of* $\hat{f}(x|\hat{h})$. Define $\Delta(h) = \hat{f}(x|h) - f(x)$, $\Delta_0 = \Delta(h_0)$, $\hat{\Delta} = \Delta(\hat{h})$, $\delta = (\hat{h} - h_0)/h_0$, $K_1(y) = K(y) + yK'(y)$, $K_2(y) = 2K(y) + 4yK'(y) + y^2K''(y)$, $S_j(h) = h^j(\partial/\partial h)^2\hat{f}(x|h)$ and $T_j = S_j(h_0)$. In this notation, $S_j(h) = (-1)^j(nh)^{-1}\sum_i K_j\{(x - X_i)/h\}$ for $j = 1, 2$, and $ES_j(h) = O(h^2)$ for $j \geq 1$. Since K''' is bounded, and since $P(C_1 n^{-1/5} \leq \hat{h} \leq C_2 n^{-1/5}) = 1$ for some $0 < C_1 < C_2$, then $\hat{\Delta} = \Delta_0 + \delta T_1 + (1/2)\delta^2 T_2 + |\delta|^3 R_1$, where

$$|R_1| \le C_3 \sup{}^*|S_3(h)| \le C_3 \sup{}^*\{|ES_3(h)| + |S_3(h) - ES_3(h)|\}$$

and $\sup^*$ denotes the supremum over $C_1 n^{-1/5} \le h \le C_2 n^{-1/5}$. For each integer $p \ge 1$, and each $\varepsilon > 0$, $E\{\sup^*|S_3(h) - ES_3(h)|^p\} = O(n^{-(2p/5)+\varepsilon})$. Furthermore,

$$(E|\Delta_0|)^{1/p} + (E|T_1|^p)^{1/p} + (E|T_2|^p)^{1/p} = O(n^{-2/5}). \tag{4.1}$$

Hence, for each $\varepsilon > 0$,

$$E(\hat{\Delta}^2) = E\{\Delta_0^2 + \delta^2(T_1^2 + \Delta_0 T_2) + 2\delta\Delta_0 T_1\} + O\left\{n^{-(4/5)+\varepsilon}(E|\delta|^6)^{1/2}\right\}. \tag{4.2}$$

STEP (ii). *Taylor expansion of* δ. By Taylor-expanding $(\tilde{f}/\tilde{f}''^2)^{1/5}$ we may show that

$$\delta = \delta_1 + \delta_2 + R_2, \tag{4.3}$$

where

$$\begin{aligned} \delta &= \tfrac{1}{5}\{f^{-1}(\tilde{f} - f) - 2f''^{-1}(\tilde{f}'' - f'')\}, \\ \delta_2 &= \tfrac{1}{25}\Big\{-f^{-2}(\tilde{f} - f)^2 + 7f''^{-2}(\tilde{f}'' - f'')^2 \\ &\qquad -2(ff'')^{-1}(\tilde{f} - f)(\tilde{f}'' - f'')\Big\}, \\ |R_2| &\le C_4(|\tilde{f} - f|^3 + |\tilde{f}'' - f''|^3). \end{aligned} \tag{4.4}$$

STEP (iii). *Simplified formula for* $E(\hat{\Delta}^2)$. Since $E|\tilde{f} - f|^p = O\{(nh_1)^{-1/2} + h_1^l\}^p$ and $E|\tilde{f}'' - f''|^p = O\{(nh_2^5)^{-1/2} + h_2^m\}^p$ for $p \ge 2$, then by (4.1), and substituting (4.3) and (4.4) into (4.2), we obtain for any $\varepsilon > 0$,

$$\begin{aligned} E(\hat{\Delta}^2) &= E\{\Delta_0^2 + \delta_1^2(T_1^2 + \Delta_0 T_2) + 2(\delta_1 + \delta_2)\Delta_0 T_1\} \\ &\quad + O\left[n^{-(4/5)+\varepsilon}\left\{(nh_1)^{-1/2} + h_1^l + (nh_2^5)^{-1/2} + h_2^m\right\}^3\right]. \end{aligned} \tag{4.5}$$

STEP (iv). *Expansion of* $E(\delta_1\Delta_0 T_1)$. We first state a simple formula. Let (U_i, V_i, W_i) be independent and identicaly distributed as (U, V, W), with zero mean and finite third moment. Let u, v, w be constants. Then

$$\begin{aligned} &E\left\{\left(u + \sum U_i\right)\left(v + \sum V_i\right)\left(w + \sum W_i\right)\right\} \\ &\quad = uvw + n\{uE(VW) + vE(UW) + wE(UV)\} + nE(UVW). \end{aligned} \tag{4.6}$$

Let N denote either L or M''. When $N = L$, put $(\alpha, \beta) = (1, 1)$, $r = l$ and $D = \tilde{f} - f$. When $N = M''$, put $(\alpha, \beta) = (2, 3)$, $r = m$ and $D = \tilde{f}'' - f''$. Define $u = ED$, $v = E\hat{f}_0 - f$, $w = ET_1$, $\rho_\alpha = h_0/h_\alpha$, $U_i' = (nh_\alpha^\beta)^{-1}N\{(x - X_i)/h_\alpha\}$, $V_i' = (nh_0)^{-1}K\{(x - X_i)/h_0\}$, $W_i' = -(nh_0)^{-1}K_1\{(x - X_i)/h_0\}$ and $(U_i, V_i, W_i) = (U_i', V_i', W_i') - E(U_i', V_i', W_i')$. Let (U, V, W) have the distribution of (U_i, V_i, W_i), and observe that $\int y^2 K_1(y)\,dy = -2\kappa$, $\int KK_1 = (1/2)\int K^2$, $u = O(h_\alpha^r)$, $v = \frac{1}{2}h_0^2\kappa f'' + O(h_0^4)$, $w = h_0^2\kappa f'' + O(h_0^4)$,

$$E(UV) = \left(n^2h_\alpha^\beta\right)^{-1} f\int K(y)N(p_\alpha y)\,dy + o\left\{\left(n^2h_\alpha^\beta\right)^{-1}\min(1, h_\alpha/h_0)\right\},$$

$$E(UW) = -\left(n^2h_\alpha^\beta\right)^{-1} f\int K_1(y)N(\rho_\alpha y)\,dy + o\left\{\left(n^2h_\alpha^\beta\right)^{-1}\min(1, h_\alpha/h_0)\right\},$$

$$E(VW) = -\left(n^2h_0\right)^{-1}\tfrac{1}{2}f\int K^2 + O\left(n^{-2}h_0\right),$$

$$E(UVW) = O\left\{\left(n^3h_0h_\alpha^\beta\right)^{-1}\min(1, h_\alpha/h_0)\right\}.$$

Now, $D = u + \Sigma U_i$, $\Delta_0 = v + \Sigma V_i$, $T_1 = w + \Sigma W_i$. Hence, by (4.6),

$$\begin{aligned} E(D\Delta_0 T_1) &= u\{vw + nE(VW)\} \\ &\quad + n\{vE(UW) + wE(UV)\} + nE(UVW) \\ &= I_1 + I_2 + o\left\{h_0^2\left(nh_\alpha^\beta\right)^{-1}\min(1, h_\alpha/h_0)\right\}, \end{aligned} \tag{4.7}$$

where

$$\begin{aligned} I_1 &= u\{vw + nE(VW)\} = O\left(h_0^6 h_\alpha^r\right), \\ I_2 &= n\{vE(UW) + wE(UV)\} \\ &= \tfrac{1}{2}h_0^2\left(nh_\alpha^\beta\right)^{-1}\kappa ff''\int\{K(y) - yK'(y)\}N(\rho_\alpha y)\,dy \\ &\quad + o\left\{\left(n^2h_\alpha^\beta\right)^{-1}\min(1, h_\alpha/h_0)\right\}. \end{aligned}$$

Remembering that $\delta_1 = (1/5)(f^{-1}D_1 - 2f''^{-1}D_2)$, where $D_1 = \tilde{f} - f$ and $D_2 = \tilde{f}'' - f''$, we may deduce from (4.7) that

$$\begin{aligned} E(\delta_1\Delta_0T_1) &= \tfrac{1}{10}h_0^2\Big[(nh_1)^{-1}f''\int\{K(y) - yK'(y)\}N(\rho_1 y)\,dy \\ &\qquad - 2\left(nh_2^3\right)^{-1}f\int\{K(y) - yK'(y)\}M''(\rho_2 y)\,dy\Big]\kappa \\ &\quad + O\left\{h_0^6\left(h_1^l + h_2^m\right)\right\} \\ &\quad + o\left[h_0^2\left\{(nh_1)^{-1}\min(1, h_1/h_0) + \left(nh_2^3\right)^{-1}\min(1, h_2/h_0)\right\}\right]. \end{aligned} \tag{4.8}$$

Note that

$$\begin{aligned} h_0^6\left(h_1^l + h_2^m\right) &= o\Big[h_0^4\left(h_1^{2l} + h_2^{2m}\right) \\ &\qquad + h_0^2\left\{(nh_1)^{-1}\min(1, h_1/h_0) + \left(nh_2^3\right)^{-1}\right\}\Big], \end{aligned} \tag{4.9}$$

put $\xi_n = h_0^4\{h_1^{2l} + h_2^{2m} + (nh_1)^{-1}\} + h_0^2\{(nh_1)^{-1}\min(1, h_1/h_0) + (nh_2^3)^{-1}\}$, and observe from (4.8) and (4.9) that, since $h_2/h_0 \to \infty$,

$$\begin{aligned} E(\delta_1\Delta_0 T_1) = \tfrac{1}{10}h_0^2\Big[&(nh_1)^{-1}f''\int\{K(y) - yK'(y)\}L(\rho_1 y)\,dy \\ &-2(nh_2^3)^{-1}f\int\{K(y) - yK'(y)\}M''(\rho_2 y)\,dy\Big]\kappa + o(\xi_n). \end{aligned} \tag{4.10}$$

STEP (v). *Expansion of* $E(\delta_2\Delta_0 T_1)$. We first state a simple formula. Let $Z_0, \ldots, Z_4$ be jointly normally distributed with zero means and covariances $\gamma_{ij} = E(Z_i Z_j)$. Let $z_0, \ldots, z_4$ denote constants. Then

$$\begin{aligned} E\left\{\prod_{i=1}^{4}(z_i + Z_i)\right\} &= (z_1 z_2 + \gamma_{12})(z_3 z_4 + \gamma_{34}) \\ &\quad+(z_1 z_3\gamma_{24} + z_1 z_4\gamma_{23} + z_2 z_3\gamma_{14} + z_2 z_4\gamma_{13}) \\ &\quad+(\gamma_{13}\gamma_{24} + \gamma_{14}\gamma_{23}). \end{aligned} \tag{4.11}$$

In the event that $Z_1 = Z_2 = Z_0$ and $z_1 = z_2 = z_0$, the right-hand side reduces to

$$(z_0^2 + \gamma_{00})(z_3 z_4 + \gamma_{34}) + 2z_0(z_3\gamma_{04} + z_4\gamma_{03}) + 2\gamma_{03}\gamma_{04}. \tag{4.12}$$

Consider applying (4.11) in the case where, for $i = 1, 2$, $(z_i, Z_i) = (E\tilde{f} - f, \tilde{f} - E\tilde{f})$ or $(E\breve{f}'' - f'', \breve{f}'' - E\breve{f}'')$, and where $(z_3, Z_3) = (E\hat{f}_0 - f, \hat{f}_0 - E\hat{f}_0)$, $(z_4, Z_4) = (ET_1, T_1 - ET_1)$. Even though $Z_1, \ldots, Z_4$ are not normally distributed, the fact that they are asymptotically normal may be used to prove that (4.11) continues to hold, provided that a correction term of size $o(\eta_n)$, where $\eta_n = \{\prod_i E(z_i + Z_i)^4\}^{1/4}$, is added to the right-hand side. Now,

$$\begin{aligned} \eta_n &= O\left\{\prod_{i=1}^{n}(z_i^2 + EZ_i^2)^{1/2}\right\} \\ &= O\left[h_0^4\left\{h_1^{2l} + h_2^{2m} + (nh_1)^{-1} + (nh_2^5)^{-1}\right\}\right]. \end{aligned} \tag{4.13}$$

Therefore,

$$\begin{aligned} \left|E\left\{\prod_{i=1}^{4}(z_i + Z_i)\right\}\right| = O\big(&|z_1 z_2 + \gamma_{12}|\,|z_3 z_4 + \gamma_{34}| + |z_1 z_3\gamma_{24}| \\ &+|z_1 z_4\gamma_{23}| + |z_2 z_3\gamma_{14}| + |z_2 z_4\gamma_{13}| + |\gamma_{13}\gamma_{24}| + |\gamma_{14}\gamma_{23}|\big) \\ &+ o\left[h_0^4\left\{h_1^{2l} + h_2^{2m} + (nh_1)^{-1} + (nh_2^5)^{-1}\right\}\right]. \end{aligned}$$

Observe that $|z_1 z_2 + \gamma_{12}| = O\{h_1^{2l} + h_2^{2m} + (nh_1)^{-1} + (nh_2^5)^{-1}\}$; that in the notation of Step (iv), $|z_3 z_4 + \gamma_{34}| = |z_3 z_4 + nE(VW)| = O(h_0^6)$, using the estimates developed in Step (iv); and that $|z_1| + |z_2| = O(h_1^l + h_2^m)$, $|z_3| + |z_4| = O(h_0^2)$, $|\gamma_{13}| + |\gamma_{14}| + |\gamma_{23}| + |\gamma_{24}| = O\{(nh_1)^{-1}\min(1, h_1/h_0) + (nh_2^3)^{-1}\}$. Hence,

$$|E(\delta_2\Delta_0T_1)| = O\Big[h_0^2(h_1^l + h_2^m)\Big\{(nh_1)^{-1}\min(1, h_1/h_0) + \big(nh_2^3\big)^{-1}\Big\} + (nh_1)^{-2}\min\big(1, h_1^2/h_0^2\big) + \big(nh_2^3\big)^{-2}\Big] + o\Big[h_0^4\Big\{h_1^{2l} + h_2^{2m} + (nh_1)^{-1} + \big(nh_2^5\big)^{-1}\Big\}\Big].$$

From this result and the fact that $(nh_1)^{-2}\min(1, h_1^2/h_0^2) + (nh_2^3)^{-2} + h_0^4(nh_2^5)^{-1} = o(\xi_n)$ we may deduce that

$$|E(\delta_2\Delta_0T_1)| = o(\xi_n). \tag{4.14}$$

STEP (vi). *Expansion of* $E(\delta_1^2T_1^2)$ *and* $E(\delta_1^2\Delta_0T_2)$. The argument employed here is similar to that used in Step (v), but is based on the simpler formula (4.12).

Let $(z_0, Z_0) = (E\delta_1, \delta_1 - E\delta_1)$. When calculating $E(\delta_1^2T_1^2)$, put

$$(z_3, Z_3) = (z_4, Z_4) = (ET_1, T_1 - ET_1), \tag{4.15}$$

and when computing $E(\delta_1^2\Delta_0T_1)$, let

$$(z_3, Z_3) = (E\Delta_0, \Delta_0 - E\Delta_0), \qquad (z_4, Z_4) = (ET_2, T_2 - ET_2). \tag{4.16}$$

Then, $|\gamma_{03}| + |\gamma_{04}| = O\{(nh_1)^{-1}\min(1, h_1/h_0) + (nh_2^3)^{-1}\}$, $|z_0| = O(h_1^l + h_2^m)$ and $|z_3| + |z_4| = O(h_0^2)$. Therefore,

$$\begin{aligned} &|2z_0(z_3\gamma_{04} + z_4\gamma_{03}) + 2\gamma_{03}\gamma_{04}| \\ &= O\Big[(h_1^l + h_2^m)h_0^2\Big\{(nh_1)^{-1}\min(1, h_1/h_0) + \big(nh_2^3\big)^{-1}\Big\} + (nh_1)^{-2}\min\big(1, h_1^2/h_0^2\big) + \big(nh_2^3\big)^{-2}\Big] = o(\xi_n). \end{aligned} \tag{4.17}$$

(Note that $nh_2^5 \to \infty$). As in Step (v), the error in assuming that $E\{\prod(z_i + Z_i)\}$ equals the quantity in (4.12) equals $o(\eta_n)$, where a bound for η_n is given by (4.13). Therefore, by (4.12), (4.17) and the fact that $\eta_n = O(\xi_n)$,

$$E\left\{\prod_{i=1}^{4}(z_i + Z_i)\right\} = \big(z_0^2 + \gamma_{00}\big)(z_3z_4 + \gamma_{34}) + o(\xi_n). \tag{4.18}$$

The left-hand side of (4.18) equals either $E(\delta_1^2T_1^2)$ or $E(\delta_1^2\Delta_0T_2)$, depending on whether (z_3, Z_3) and (z_4, Z_4) are defined by (4.15) or (4.16), respectively. Since $\int y^2K_1 = -2\kappa$, $\int y^2K_2 = 2\kappa$, $\int K_1^2 = \int y^2K'^2$, $\int KK_2 = \int(K^2 - y^2K'^2)$, then the sum of the respective versions of $z_3z_4 + \gamma_{34}$ equals

$$\big\{(ET_1)^2 + \mathrm{var}(T_1)\big\} + \{(E\Delta_0)(ET_2) + \mathrm{cov}(\Delta_0, T_2)\} = \tfrac{5}{2}h_0^4\kappa^2f''^2 + o\big(h_0^4\big).$$

Therefore,

$$E\big(\delta_1^2T_1^2\big) + E\big(\delta_1^2\Delta_0T_2\big) = \tfrac{5}{2}h_0^4E\big(\delta_1^2\big)\kappa^2f''^2 + o(\xi_n). \tag{4.19}$$

Let $\rho = h_1/h_2$, and observe that

$$25E(\delta_1^2) = \left(\frac{1}{l!}h_1^l \lambda f^{-1} f^{(l)} - \frac{2}{m!} h_2^m \mu f''^{-1} f^{(m+2)}\right)^2$$
$$+ \left\{(nh_1)^{-1} f^{-2} \int L^2 + 4(nh_2^5)^{-1} f''^{-2} \int M''^2\right.$$
$$\left. -4(nh_2^3)^{-1} (ff'')^{-1} \int L(y) M''(\rho y)\, dy\right\} f$$
$$+ o\left\{h_1^{2l} + h_2^{2m} + (nh_1)^{-1} + (nh_2^5)^{-1}\right\}.$$

Combining this formula with (4.19), and remembering that $nh_2^5 \to \infty$, we deduce that

$$E\{\delta_1^2(T_1^2 + \Delta_0 T_2)\}$$
$$= \frac{1}{10} h_0^4 \left(\frac{1}{l!} h_1^l \lambda f^{-1} f'' f^{(l)} - \frac{2}{m!} h_2^m \mu f^{(m+2)}\right)^2 \kappa^2 \tag{4.20}$$
$$+ \frac{1}{10} h_0^4 (nh_1)^{-1} f^{-1} f''^2 \kappa^2 \int L^2 + o(\xi_n).$$

STEP (vii). *Completion*. In view of (4.5),

$$E(\hat{\Delta}^2) = E(\Delta_0^2) + E\{\delta_1^2(T_1^2 + \Delta_0 T_2)\} + 2E(\delta_1 \Delta_1 T_1) + 2E(\delta_2 \Delta_0 T_1) + o(\xi_n).$$

Using (4.20), (4.10) and (4.14) to expand the second, third and fourth terms, respectively, on the right-hand side, we see that

$$E(\hat{\Delta}^2) = E(\Delta_0^2) + \frac{1}{10} h_0^4 \left(\frac{1}{l!} h_1^l \lambda f^{-1} f'' f^{(l)} - \frac{2}{m!} h_2^m \mu f^{(m+2)}\right)^2 \kappa^2$$
$$+ \frac{1}{10} h_0^4 (nh_1)^{-1} \kappa^2 f^{-1} f''^2 \int L^2$$
$$+ \frac{1}{5} h_0^2 \left[(nh_1)^{-1} f'' \int \{K(y) - yK'(y)\} L(\rho_1 y)\, dy \right.$$
$$\left. -4(nh_2^3)^{-1} \kappa M''(0) f\right] \kappa + o(\xi_n),$$

as had to be shown.

Acknowledgments. This paper has benefited significantly from the helpful comments of two referees and an Associate Editor.

REFERENCES

ABRAMSON, I. S. (1982). On bandwidth variation in kernel estimates—a square root law. *Ann. Statist.* **10** 1217–1223.

JONES, M. C. (1990). Variable kernel density estimates and variable kernel density estimates. *Austral. J. Statist.* **32** 361–371.

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

PARK, B. U. and MARRON, J. S. (1990). Comparison of data-driven bandwidth selectors. *J. Amer. Statist. Assoc.* **85** 66–72.

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

SILVERMAN, B. W. (1976). On a Gaussian process related to multivariate probability density estimation. *Math. Proc. Cambridge Philos. Soc.* **80** 135–144.

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

SILVERMAN, B. W. (1986). *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, London.

WOODROOFE, M. (1970). On choosing a delta-sequence. *Ann. Math. Statist.* **41** 1665–1671.

CENTRE FOR MATHEMATICS
AND ITS APPLICATIONS
AUSTRALIAN NATIONAL UNIVERSITY
G.P.O. BOX 4
CANBERRA ACT 2601
AUSTRALIA