The Annals of Statistics
1995, Vol. 23, No. 1, 1–10

IMPROVED VARIABLE WINDOW KERNEL ESTIMATES OF PROBABILITY DENSITIES

By Peter Hall, Tien Chung Hu[1] and J. S. Marron[2]

Australian National University, National Tsing Hua University and University of North Carolina, Chapel Hill

Variable window width kernel density estimators, with the width varying proportionally to the square root of the density, have been thought to have superior asymptotic properties. The rate of convergence has been claimed to be as good as those typical for higher-order kernels, which makes the variable width estimators more attractive because no adjustment is needed to handle the negativity usually entailed by the latter. However, in a recent paper, Terrell and Scott show that these results can fail in important cases. In this paper, we characterize situations where the fast rate is valid, and also give rates for a variety of cases where they are slower. In addition, a modification of the usual variable window width estimator is proposed, which does have the earlier claimed rates of convergence.

1. Introduction. There are many curve estimation problems where it is intuitively clear that it would be desirable to use different amounts of smoothing at different locations. Since it provides a simple setting in which to study the main ideas, most of the research on this topic has been done in the context of kernel density estimation.

A useful mathematical structure for understanding this problem involves consideration of using independent, identically distributed observations $X_1, \dots, X_n$ from a probability density $f(x)$, in an attempt to estimate f. The (global bandwidth) kernel estimator of f is

$$\hat{f}(x \mid h) = n^{-1} \sum_{i=1}^{n} K_h(x - X_i), \tag{1}$$

where K is a kernel function, taken here to be a symmetric probability density, and the parameter h, often called bandwidth or window width, controls the scale in the sense $K_h(y) = K(y/h)/h$. See Silverman [(1986), Section 2.4] for good motivation and intuitive discussion of this estimator.

Bandwidth variation, that is, allowing the bandwidth to be different at different locations, is typically done in one of two ways, but see Wand, Marron and Ruppert (1991) for another approach based on transformations. The first of these usual methods involves allowing h to depend on x, and the most common choice of $h(x)$ involves nearest neighbor ideas [see Mack and Rosenblatt (1979) and Hall (1983)]. While this method can yield improvements and has a definite

Received June 1992; revised January 1994.
[1]Research done while on leave at the University of North Carolina, Chapel Hill.
[2]Research partially supported by NSF Grants DMS-89-02973 and DMS-92-03135.
AMS 1991 *subject classification*. 62G05.
Key words and phrases. Adaptive methods, bandwidth choice, curve estimation, nonparametric estimation, smoothing.

intuitive appeal, these improvements have been relatively minor compared to the dramatic improvements shown for allowing h to depend instead on X_i [see Abramson (1982a)].

When these ideas are studied asymptotically, as $n \to \infty$, $h \to 0$, it is often enough to consider the rate of convergence of the bias, because all these variations on the kernel estimator have variances of the same order $n^{-1}h^{-1}$. The bias rate of convergence for $h(x)$ bandwidth variation is the same as for the global bandwidth estimator (i.e., of order h^2), although the constant coefficient may sometimes be expected to be better. On the other hand, when the function $h(X_i)$ is inversely proportional to the square root of f, Silverman [(1986), see Section 5.3.3], Hall and Marron (1988), Hall (1990) and Jones (1990) have claimed a much faster bias rate, h^4, can be obtained. Hall and Marron (1988) go on to show that the same fast rate applies, even if a conventional global kernel estimator is used as an estimate of the best h-function.

However, more recently, Terrell and Scott (1992) have pointed out that the bias rate is much slower, in fact $(h/\log h)^2$, in the important special case of K uniform and f Normal. This example indicates a technical flaw with the earlier rate of convergence results. The h^4 rate given in the earlier papers is correct only under some rather stringent assumptions requiring f to have heavy tails.

In Section 2 of this paper, we establish conditions which ensure that the bias rate of h^4 is correct. It is seen that the tail of f needs to be nearly so heavy that the third moment does not exist. The bias rate of convergence is also quantified, in terms of tail conditions on f, in many cases where the rate is slower than h^4. In particular, an asymptotic representation is given for the bias, which has several terms, one of which is the same as found in the previous works. However, the other terms are dominant in many important cases. When f has roughly polynomial tails, these latter terms dominate when the tails of $f(x)$ go down more rapidly than x^{-4}. These latter terms always dominate when f has exponential tails.

In Section 3, we present an "ideal" means of fixing the location adaptive estimator, to give the previously advertised rates, under more reasonable conditions, based on a type of truncation of the kernel function. The estimator is "ideal" in the sense that it still makes use of the unknown density in the bandwidth function.

The ideal estimator is made practical in Section 4, where the idea of substituting a pilot kernel estimate into the bandwidth function is investigated. It is seen that this does have an effect on the limiting distribution, but that the data adaptive version still has the same fast rates of convergence.

All results in this paper are only for estimation at a single point. See Hall (1992) for analogs of many of these ideas in the rather different case of integrated squared error.

2. Asymptotic bias analysis. The modification of the estimator (1) discussed in this section allows the window width h to vary with the kernel center X_i according to the inverse of the square root of the density. In particular, define

$$\tilde{f}(x \mid h) = n^{-1} \sum_{i=1}^{n} K_{h_i}(x - X_i), \tag{2}$$

where $h_i = h\, f(X_i)^{-1/2}$. Of course this estimator is not useful in practice because f is unknown, but Hall and Marron (1988) have shown that when a pilot estimate is used, the asymptotic performance is essentially equivalent in a useful sense. As this idealized estimator $\tilde{f}$ exhibits the technical problems discovered by Terrell and Scott, it is the focus in this section.

We consider two types of underlying densities, those with regularly varying tails and those with exponentially decreasing tails. In the first case the defining criterion is

$$f(x) \sim x^{-\alpha_1} L_1(x), \qquad f(-x) \sim x^{-\alpha_2} L_2(x), \tag{3}$$

as $x \to \infty$, where $\alpha_1, \alpha_2 > 1$ and L_1 and L_2 are continuous, monotone, slowly varying functions. In the second case,

$$\begin{aligned} f(x) &\sim c_{11} \exp(-c_{12} x^{\alpha_1}), \\ f(-x) &\sim c_{21} \exp(-c_{22} x^{\alpha_2}) \end{aligned} \tag{4}$$

as $x \to \infty$, where c_{ij} and α_i are positive constants. Our proofs apply to more general distributions, for example to distributions which have one tail regularly varying and the other exponentially decreasing. However, from the viewpoint of stating theorems economically it is convenient to assume either (3) or (4).

In addition to (3) and (4) we suppose that, for fixed x, f has four bounded derivatives and that K is compactly supported and Lipschitz continuous, with

$$\int y^j K(y)\,dy = \begin{cases} 1, & \text{for } j = 0, \\ 0, & \text{for } j = 1, 3, \\ \kappa_j, & \text{for } j = 2, 4. \end{cases} \tag{5}$$

First we treat the context of (3) and let $\lambda_i = \lambda_i(h)$ denote a solution of the equation

$$L_i(h^{-2/(\alpha_i - 2)} \lambda_i) = \lambda_i^{\alpha_i - 2},$$

for $\alpha_i > 2$.

THEOREM 2.1. *Assume* (3) *and* (5). *Then, for constants* $\kappa_{\pm}(\alpha)$ *defined in Section* 5.1,

$$\begin{aligned} E\{\tilde{f}(x \mid h)\} - f(x) = {}& h^4 (\kappa_4/24)(d/dx)^4 f(x)^{-1} \\ & + h^{2\alpha_1/(\alpha_1 - 2)} \lambda_1^{-2} \kappa_-(\alpha_1) I(\alpha_1 \geq 4) \\ & + h^{2\alpha_2/(\alpha_2 - 2)} \lambda_2^{-2} \kappa_+(\alpha_2) I(\alpha_2 \geq 4) \\ & + o\left\{ h^4 + \sum_{i=1}^{2} h^{2\alpha_i/(\alpha_i - 2)} \lambda_i^{-2} I(\alpha_i \geq 4) \right\} \end{aligned}$$

REMARK 2.1. For each $\varepsilon > 0$, the quantity λ_i^{-2} is between h^{ε} and $h^{-\varepsilon}$ in order of magnitude. Thus, the term of order $h^{2\alpha_i/(\alpha_i-2)}\lambda_i^{-2}$ only becomes significant if $\alpha_i \ge 4$. When $\alpha_i > 4$, it dominates the term of order h^4. For $\alpha_i = 4$, the relative sizes of terms in $h^{2\alpha_i/(\alpha_i-2)}\lambda_i^{-2} = h^4\lambda_i^{-2}$ and h^4 are of course determined by λ_i, and hence by the slowly varying function L_i.

REMARK 2.2. If $\max(\alpha_1, \alpha_2) < 4$ or, equivalently, if, for some $\varepsilon > 0$,

$$\int_0^{\infty} y^{3-\varepsilon} f(y)\,dy = \int_0^{\infty} y^{3-\varepsilon} f(-y)\,dy = \infty,$$

then

$$E(\hat{f}(x)) - f(x) = h^4(\kappa_4/24)(d/dx)^4 f(x)^{-1} + o(h^4);$$

that is, the terms of order $h^{2\alpha_i/(\alpha_i-2)}\lambda_i^{-2}$ play no role, and the bias has the properties noted in Silverman [(1986), Section 5.3.3], in Hall and Marron (1988), in Hall (1990) and in Jones (1990). Thus, we see that, for f with tails that are either polynomial or exponential, bias is of the order stated in these earlier results "if and only if" moments of order 3 and above are infinite in each tail. It is easily checked that a different sufficient condition for the earlier results to hold is that f be compactly supported, and bounded above 0 on its support. This result is true as stated if we continue attention to the case where L_1 and L_2 are asymptotically constant. More generally, in the event that $\max(\alpha_1, \alpha_2) = 4$, a little additional qualification is required.

REMARK 2.3. Good heuristic discussion as to why tail behavior is so dominant in this "local" estimator can be found in Terrell and Scott [(1992), Section 4]. Essentially what happens is that when f has light enough tails, the kernel windows for extreme observations get such large bandwidths that they affect estimation even at interior points, to the level quantified here. A referee has made the nice comment that these tail effects cause the interchange of differentiation and integration assumed in Silverman [(1986), Section 5.3.3] to break down when $\min(\alpha_1, \alpha_2) \ge 4$.

Next we treat the case of distributions satisfying (4).

THEOREM 2.2. *Assume* (4) *and* (5). *Then, for constants* $\kappa_{\pm}(c_1, c_2, \alpha)$ *defined in Section* 5.1,

$$\begin{aligned} E\{\tilde{f}(x \mid h)\} - f(x) &= h^2(\log h^{-1})^{-1-(2/\alpha_1)}\kappa_-(c_{11}, c_{12}, \alpha_1) \\ &\quad + h^2(\log h^{-1})^{-1-(2/\alpha_2)}\kappa_+(c_{21}, c_{22}, \alpha_2) \\ &\quad + o\{h^2(\log h^{-1})^{-1-(2/\alpha_1)} + h^2(\log h^{-1})^{-1-(2/\alpha_2)}\}, \end{aligned}$$

as $h \to 0$.

REMARK 2.4. When tails are exponentially small the improvement in bias, from the use of variable window width estimates, is virtually negligible, being only logarithmic in character. However, bias is of smaller order than $h^2(\log\ h^{-1})^{-1}$, no matter how large the values of α_1 and α_2.

REMARK 2.5. A nice example in McKay (1993a) indicates there will be some interesting companion results to Theorems 2.1 and 2.2, in the case where f is compactly supported. In particular if f "comes down to the endpoints of its support sufficiently rapidly" (e.g., as a beta density with high enough order), then the bias will again be different, and will be dependent on the rate of decrease of f at these endpoints.

3. An improved variable window estimator. The extra, very troublesome, bias terms for the estimator $\tilde{f}$, made precise in the last section, may be viewed as coming from extreme tail observations. One would suspect that this problem could be avoided, if the effect of far away data points could be eliminated.

One means of eliminating this effect has already been proposed by Abramson (1982a), who replaced h_i in (2) by $h[f(X_i)\vee(f(x)/10)]$. Analysis as in this section shows that this modification does indeed yield a bias which is asymptotically equal to only the first term given in Theorem 2.1, without unnatural heavy tail conditions. Terrell and Scott (1992) discuss other modifications. See McKay (1993a, b) for a modification which is both nonnegative and integrates to 1.

To get the main ideas across as simply as possible, we work here with a modification which is chosen because it allows straightforward adaptation of the methods, and also the main lessons, of Hall and Marron (1988). In particular, define

$$\bar{f}(x \mid h) = n^{-1}\sum_{i=1}^{n} K_{h_i}(x - X_i)1_{\{|(x-X_i)/h|<C\}}, \tag{6}$$

where, as above, $K_h(\cdot) = K(\cdot/h)/h$ and $h_i = hf(X_i)^{-1/2}$.

Any sufficiently large constant C will give the desired asymptotic result; however, a sufficient condition is that $C \geq c_1c_2^{-1}$, where K vanishes outside $(-c_1, c_1)$, and $f(x)^{1/2} \geq 2c_2$. While specific choice of C makes no difference in the first-order asymptotic analysis presented here, it should make at least some difference in practice. Perhaps deeper asymptotic analysis could provide insight into possibilities for optimizing the choice of C, although this will clearly be of only second-order importance.

The beneficial effects of this modification are demonstrated by the following theorem.

THEOREM 3.1. *Assume* (5) *and* $C \geq c_1c_2^{-1}$. *Then*

$$E\{\bar{f}(x \mid h)\} - f(x) = h^4(\kappa_4/24)(d/dx)^4 f(x)^{-1} + o(h^4),$$

as $h \to 0$.

Explicit proof of Theorem 3.1 is not given, because it follows directly from arguments of Hall and Marron (1988). In particular, integration by substitution gives

$$E\bar{f}(x \mid h) = \int_{-C}^{C} K\{-z\, f(x+hz)^{1/2}\} f(x+hz)^{3/2}\, dz,$$

from which the assumption $C \geq c y_1 c_2^{-1}$ yields the desired result, as in that paper.

To analyze (6) further, define $\kappa = \int K^2$. For $C \geq c_1 c_2^{-1}$,

$$\begin{aligned} \operatorname{var}(\bar{f}(x \mid h)) &= (nh)^{-1} \int_{-C}^{C} K\{-z\, f(x+hz)^{1/2}\} f(x+hz)^2\, dz \\ &\quad - n^{-1} E\bar{f}(x \mid h) \\ &= (nh)^{-1} f(t)^{3/2} \kappa + o((nh)^{-1}). \end{aligned} \tag{7}$$

This, together with Theorem 3.1 gives

$$E[\bar{f}(x \mid h) - f(x)]^2 = c_4 (nh)^{-1} + c_5 h^8 + o((nh)^{-1} + h^8). \tag{8}$$

Therefore the minimizer in h of (8) is asymptotic to a constant multiple of $n^{-1/9}$, and the best mean squared error is of the order $n^{-8/9}$.

It is also straightforward to see that $\bar{f}$ satisfies a central limit theorem. In particular, when $n \to \infty$ with $h \to 0$ and $nh \to \infty$, $\bar{f}(x \mid h) - E\{\bar{f}(x \mid h)\}$ is asymptotically normal, with zero mean and variance satisfying (7).

The case where h is a random variable, denoted $\hat{h}$, as in data driven bandwidth selection, is also readily treated. If we add to our conditions on K the assumption that K is continuously differentiable, and if the bandwidth $\hat{h}$ satisfies $n^{1/9}\hat{h} \to c_6$ in probability [some $c_6 \in (0, \infty)$], then arguments of Abramson (1982b) may be adapted to show that

$$\bar{f}(x \mid \hat{h}) = \bar{f}(x \mid c_6 n^{-1/9}) + o_p(n^{-1/9}).$$

Note that in general the estimate $\bar{f}(x \mid h)$ will have an integral less than 1 and hence will not be a density. However, this effect will vanish asymptotically and is easily adjusted for in situations where this poses any difficulty.

4. An adaptive improved variable window estimator. The ideal estimator $\bar{f}(x \mid h)$ discussed in the last section has good asymptotic properties, but cannot be used in practice, since it still makes use of the unknown density $f(x)$. In this section we analyze the obvious idea of using a "pilot" kernel estimator in the place of f.

For this, let $\hat{f}(x \mid \hat{h}_1)$ denote the ordinary kernel estimator given in (1), using a data driven bandwidth $\hat{h}_1$. Assume that $\hat{h}_1$ is "good" in the sense that, for some $\alpha > 0$,

$$P[n^{-\alpha} < n^{1/5}\hat{h}_1 < n^{\alpha}] \to 1. \tag{9}$$

See Marron (1988) and Jones, Marron and Sheather (1993) for surveys of data driven bandwidths, including some that are good in this sense. Now consider replacing the bandwidth h in (6) with another data driven bandwidth $\hat{h}_2$, where

$$\lim_{\eta\to 0,\ \lambda\to\infty} \liminf_{n\to\infty} P[\eta < n^{1/9}\hat{h}_2 < \lambda] = 1. \tag{10}$$

It is straightforward to adapt, for example, the ideas of Park and Marron (1990) to find such data driven bandwidths. Now define the fully data adaptive, variable bandwidth, kernel density estimator by

$$\hat{f}(x \mid \hat{h}_1, \hat{h}_2) = n^{-1} \sum_{i=1}^{n} K_{\hat{h}_i}(x - X_i) 1_{\{|(x-X_i)/\hat{h}_2| < C\}}, \tag{11}$$

where, as above, $K_h(\cdot) = K(\cdot/h)/h$ and $\hat{h}_i = \hat{h}_2 \hat{f}(X_i \mid \hat{h}_1)^{-1/2}$.

In addition to (5) and the symmetry of K, we further assume that K has two bounded derivatives. Quantities that appear in the limiting behavior of $\hat{f}(x \mid \hat{h}_1, \hat{h}_2)$ include

$$\begin{aligned}\tau(u, x \mid h_1, h_2) = E\big[&f(X_1)^{-1/2}\{K((X_1 - u)/h_1) - h_1 \mu(X_1 \mid h_1)\} \\ &\times L\{f(X_1)^{1/2}(x - X_1)/h_2\} 1_{\{|(x-X_1)/h_2| < C\}}\big],\end{aligned}$$

where $\mu(x \mid h_1) = E[\hat{f}(x \mid h_1)]$ and $L(z) = K(z) + zK'(z)$, and also include

$$T(x \mid h_1, h_2) = (2nh_1h_2)^{-1} \sum_{i=1}^{n} \tau(X_i, x \mid h_1, h_2).$$

THEOREM 4.1. *Assume* (9), (10) *and that* $\hat{h}_2/h_2 \to 1$ *in probability. Then*

$$\hat{f}(x \mid \hat{h}_1, \hat{h}_2) - \bar{f}(x \mid h_2) - T(x \mid \hat{h}_1, \hat{h}_2) = o_p(n^{-4/9}).$$

The proof of Theorem 4.1 is omitted because it is essentially the same, although slightly easier because of the truncation represented by the indicator function, as that of Theorem 3.1 of Hall and Marron (1988).

Theorem 4.1 shows that, except for the term $T(x \mid \hat{h}_1, \hat{h}_2)$, the behavior of $\hat{f}(x \mid \hat{h}_1, \hat{h}_2)$ is the same as that described for $\bar{f}(x \mid h_2)$ in Section 3. To see how the term $T(x \mid \hat{h}_1, \hat{h}_2)$ affects this, define

$$S(x \mid h_2) = \bar{f}(x \mid h_2) - E\{\bar{f}(x \mid h_2)\}.$$

THEOREM 4.2. *Suppose that for nonrandom* h_1 *and* h_2 *satisfying* $n^{\varepsilon}(h_1 \vee h_2) \to 0$, $n^{1-\varepsilon}(h_1 \wedge h_2) \to \infty$ *for some* $\varepsilon > 0$, *and* $h_1 h_2^{-1} \to 0$. *Also suppose that* $\hat{h}_1/h_1 \to 1$ *and* $\hat{h}_2/h_2 \to 1$ *in probability. Then*

$$(nh_2)^{1/2}(S(x \mid \hat{h}_2), T(x \mid \hat{h}_1, \hat{h}_2)) \to (N_1, N_2)$$

in distribution, where (N_1, N_2) *are bivariate normal, with zero means and covariances given by*

$$\operatorname{var}(N_1) = f(x)^{3/2} \int K^2,$$

$$\operatorname{var}(N_2) = f(x)^{3/2} \int L^2/4,$$

$$\operatorname{cov}(N_1, N_2) = f(x)^{3/2} \int KL/2.$$

The proof of Theorem 4.2 is also omitted since it is so close to that of Theorem 3.2 of Hall and Marron (1988).

Theorem 4.2 shows that the limiting distribution of $\hat{f}(x \mid \hat{h}_1, \hat{h}_2)$ is indeed different from that of $\bar{f}(x \mid h_2)$. However, it also shows that $\hat{f}(x \mid \hat{h}_1, \hat{h}_2)$ still has the same fast $n^{-8/9}$ rate of mean square error convergence to $f(x)$.

5. Outline of proofs.

5.1. *Definition of constants.* The constants appearing in Theorem 2.1 are

$$\kappa_{\pm}(\alpha) = 2(\alpha - 2)^{-1} \int_0^{\infty} K(\pm u) u^{2\alpha/(\alpha-2)}\, du,$$

where $\alpha > 0$. The constants appearing in Theorem 2.2 are

$$\kappa_{\pm}(c_1, c_2, \alpha) = c_1^{3/2}\left(\tfrac{1}{2}c_2\right)^{1+(2/\alpha)} \int_{-\infty}^{\infty} K\left(\pm c_1^{1/2} \exp\left(-\tfrac{1}{2}\alpha c_2 u\right)\right) \exp\left(-\tfrac{3}{2}\alpha c_2 u\right) du$$

for $c_1, c_2, \alpha > 0$.

5.2. *Proof of Theorem* 2.1. Integration by substitution gives

$$E\{\hat{f}(X)\} = \int_{-\infty}^{\infty} K(-z f(x+hz)^{1/2}) f(x+hz)^{3/2}\, dz.$$

As in Section 3, for $c_3 \geq c_1 c_2^{-1}$,

$$\int_{-c_3}^{c_3} K\{-z f(x+hz)^{1/2}\} f(x+hz)^{3/2}\, dz = f(x) + h^4(\kappa_4/24)(d/dx)^4 f(x)^{-1} + o(h^4). \tag{12}$$

The extra terms in the expansion of the theorem come from integrals outside the range $(-c_3, c_3)$. We shall prove that, for $c_3 \geq c_1 c_2^{-1}$,

$$I \equiv \int_{c_3}^{\infty} K\{-z f(x+hz)^{1/2}\} f(x+hz)^{3/2}\, dz \tag{13}$$

$$= \kappa_-(\alpha_1) h^{2\alpha_1/(\alpha_1-2)} \lambda_1^{-2} + o(h^{2\alpha_1/(\alpha_1-2)} \lambda_1^{-2}) \tag{14}$$

if $\alpha_1 \geq 4$, and equals $o(h^4)$ if $\alpha_1 < 4$. The other tail, that is, the integral from $-\infty$ to $-c_3$, may be treated similarly and produces the analogous result.

For brevity of notation we shall delete the subscript "1" from L_1, α_1 and λ_1. The case $\alpha \le 2$ is relatively straightforward, and so we treat only $\alpha > 2$. There,

$$\begin{aligned} I &= h^{-\alpha/(\alpha-2)}\lambda \int_0^{c_3^{-1}\lambda h^{-\alpha/(\alpha-2)}} K\{-u^{-1}h^{-\alpha/(\alpha-2)}\lambda f(x+u^{-1}h^{-2/(\alpha-2)}\lambda)^{1/2}\} \\ &\qquad \times f(x+u^{-1}h^{-2/(\alpha-2)}\lambda)^{3/2}u^{-2}\,du \\ &\sim h^{-\alpha/(\alpha-2)}\lambda \int_0^{c_3^{-1}\lambda h^{-\alpha/(\alpha-2)}} K\{-u^{(\alpha-2)/2}\lambda^{-\alpha/(\alpha-2)}L(h^{-2/(\alpha-2)}\lambda)\} \\ &\qquad \times u^{(3\alpha/2)-2}h^{3\alpha/(\alpha-2)}\lambda^{-3\alpha/2}L(h^{-2/(\alpha-2)}\lambda)^{3/2}\,du \\ &= h^{2\alpha/(\alpha-2)}\lambda^{-2}\int_0^{\infty} K\{-u^{(\alpha-2)/2}\}u^{(3\alpha/2)-2}\,du \\ &= \kappa_-(\alpha)h^{2\alpha/(\alpha-2)}\lambda^{-2}. \end{aligned} \tag{15}$$

5.3. *Proof of Theorem* 2.2. Result (12) holds as before. With I defined as in (13), we must now prove the following analog of (14):

$$I = \kappa_-(c_{11}, c_{12}, \alpha_1)h^2(\log h^{-1})^{-1-(2/\alpha_1)} + o\{h^2(\log h^{-1})^{-1-(2/\alpha_1)}\}.$$

Drop the subscript "1" from α_1, and let λ denote the solution of the equation

$$\log(h^{-1}\lambda) + \tfrac{1}{2}c_{12}(\alpha\lambda^{\alpha-1}x - \lambda^{\alpha}) = 0.$$

Then $\lambda \sim (2c_{12}^{-1}\log h^{-1}\lambda)^{1/\alpha}$ as $h \to 0$. With $z = h^{-1}\lambda(1+\lambda^{-\alpha}v)$ and v fixed,

$$|x-hz|^{\alpha} = (hz)^{\alpha}\{1-(hz)^{-1}x\}^{\alpha} = \lambda^{\alpha} + \alpha v - \alpha\lambda^{\alpha-1}x + o(1),$$

and so

$$\begin{aligned} \log\{zc_{11}^{-1/2}f(x-hz)^{1/2}\} &= \log(h^{-1}\lambda) - \tfrac{1}{2}c_{12}(\lambda^{\alpha} + \alpha v - \alpha\lambda^{\alpha-1}x) + o(1) \\ &= -\tfrac{1}{2}\alpha c_{12}v + o(1). \end{aligned}$$

Therefore, arguing as in (15),

$$\begin{aligned} I &\sim h^{-1}\lambda^{1-\alpha}\int_{-\infty}^{\infty} K\{-c_{11}^{1/2}\exp(-\tfrac{1}{2}\alpha c_{12}v)\}c_{11}^{3/2}(h\lambda^{-1})^3\exp(-\tfrac{3}{2}\alpha c_{12}v)\,dv \\ &\sim \kappa_-(c_{11}, c_{12}, \alpha)h^2(\log h^{-1})^{-1-(2/\alpha)}, \end{aligned}$$

as had to be shown.

REFERENCES

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

ABRAMSON, I. S. (1982b). Arbitrariness of the pilot estimator in adaptive kernel methods. *J. Multivariate Anal.* **12** 562–567.

BREIMAN, L., MEISEL, W. and PURCELL, E. (1977). Variable kernel estimates of probability densities. *Technometrics* **19** 135–144.

HALL, P. (1983). On near neighbour estimates of a multivariate density. *J. Multivariate Anal.* **13** 24–39.

HALL, P. (1990). On the bias of variable bandwidth kernel estimators. *Biometrika* **77** 529–535.

HALL, P. (1992). On global properties of variable bandwidth density estimators. *Ann. Statist.* **20** 762–778.

HALL, P. and MARRON, J. S. (1988). On variable window width kernel estimates of probability densities. *Probab. Theory Related Fields* **80** 37–49.

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

JONES, M. C., MARRON, J. S. and SHEATHER, S. J. (1993). Progress in data based bandwidth selection for kernel density estimation. Unpublished manuscript.

MACK, Y. P. and ROSENBLATT, M. (1979). Multivariate k-nearest neighbor density estimates. *J. Multivariate Anal.* **9** 1–15.

MCKAY, I. (1993a). A note on bias reduction in variable kernel density estimates. *Canad. J. Statist.* **21** 367–375.

MCKAY, I. J. (1993b). Variable kernel methods in density estimation, Ph.D. dissertation, Queens Univ., Kingston, Ontario.

MARRON, J. S. (1988). Automatic smoothing parameter selection: a survey. *Empirical Economics* **13** 187–208.

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

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

TERRELL, G. R. and SCOTT, D. W. (1992). Variable kernel density estimation. *Ann. Statist.* **20** 1236–1265.

WAND, M. P., MARRON, J. S. and RUPPERT, D. (1991). Transformations in density estimation. *J. Amer. Statist. Assoc.* **86** 343–361.

PETER HALL
DEPARTMENT OF STATISTICS
AUSTRALIAN NATIONAL UNIVERSITY
G. P. O. BOX 4
CANBERRA, ACT 2601
AUSTRALIA

TIEN CHUNG HU
INSTITUTE OF MATHEMATICS
NATIONAL TSING HUA UNIVERSITY
HSINCHU, TAIWAN
REPUBLIC OF CHINA

J. S. MARRON
DEPARTMENT OF STATISTICS
UNIVERSITY OF NORTH CAROLINA
CHAPEL HILL, NORTH CAROLINA 27599-3260