The Annals of Statistics
1992, Vol. 20, No. 2, 797–806

USING STOPPING RULES TO BOUND THE MEAN INTEGRATED SQUARED ERROR IN DENSITY ESTIMATION[1]

BY ADAM T. MARTINSEK

University of Illinois

Suppose $X_1, X_2, \ldots, X_n$ are i.i.d. with unknown density f. There is a well-known expression for the asymptotic mean integrated squared error (MISE) in estimating f by a kernel estimate $\hat{f}_n$, under certain conditions on f, the kernel and the bandwidth. Suppose that one would like to choose a sample size so that the MISE is smaller than some preassigned positive number w. Based on the asymptotic expression for the MISE, one can identify an appropriate sample size to solve this problem. However, the appropriate sample size depends on a functional of the density that typically is unknown. In this paper, a stopping rule is proposed for the purpose of bounding the MISE, and this rule is shown to be asymptotically efficient in a certain sense as w approaches zero. These results are obtained for data-driven bandwidths that are asymptotically optimal as n goes to infinity.

1. Introduction and summary. Assume that $X_1, X_2, \ldots$ are i.i.d. random variables with unknown continuous-type density f on the real line. Based on the random sample $X_1, \ldots, X_n$ of size n, one can estimate f by the kernel estimate

$$\hat{f}_n(x|h_n) = \frac{1}{nh_n} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h_n}\right), \tag{1.1}$$

where K is the kernel function and h_n is the bandwidth. It will be assumed throughout that K is nonnegative and $h_n > 0$, $h_n \to 0$ as $n \to \infty$. A popular measure of the global performance of the estimate $\hat{f}_n(\cdot|h_n)$ is its mean integrated squared error (MISE), which is defined as

$$E\int_{-\infty}^{\infty} \left(\hat{f}_n(x|h_n) - f(x)\right)^2 dx = \int_{-\infty}^{\infty} E\left[\left(\hat{f}_n(x|h_n) - f(x)\right)^2\right] dx. \tag{1.2}$$

Under certain regularity conditions, the following asymptotic expansion for the MISE holds:

$$\begin{aligned} &E\int_{-\infty}^{\infty} \left(\hat{f}_n(x|h_n) - f(x)\right)^2 dx \\ &\quad = \frac{1}{nh_n}\int_{-\infty}^{\infty} K^2(t)\, dt + \frac{k_2^2}{4} h_n^4 \int_{-\infty}^{\infty} \left(f''(x)\right)^2 dx + o\left(\frac{1}{nh_n} + h_n^4\right), \end{aligned} \tag{1.3}$$

Received July 1990; revised June 1991.

[1]Research supported in part by NSF Grant DMS-88-02556 and by the Air Force under Grant AFOSR-87-0041.

AMS 1980 *subject classifications*. Primary 62G07; secondary 62L12, 62G20.

Key words and phrases. Density estimation, stopping rule, sequential estimation, asymptotic efficiency, mean integrated squared error.

where

$$k_2 = \int_{-\infty}^{\infty} t^2 K(t)\, dt$$

[see Rosenblatt (1956, 1971), Nadaraya (1974) and Prakasa Rao (1983), Theorem 2.1.7]. Put

$$\theta = \int_{-\infty}^{\infty} (f''(x))^2\, dx.$$

It follows from (1.3) that the best asymptotic rate for h_n is $n^{-1/5}$. If $h_n = cn^{-1/5}$, where $c > 0$ is fixed, then (1.3) becomes

$$\left\{\frac{1}{c}\int_{-\infty}^{\infty} K^2(t)\, dt + \frac{k_2^2 c^4}{4}\theta\right\} n^{-4/5} + o(n^{-4/5}). \tag{1.4}$$

A calculus argument shows that the best choice of c [i.e., the one that minimizes the leading term in (1.4)] is

$$c^* = k_2^{-2/5}\left\{\int_{-\infty}^{\infty} K^2(t)\, dt\right\}^{1/5} \theta^{-1/5}, \tag{1.5}$$

with corresponding MISE

$$\tfrac{5}{4} k_2^{2/5}\left\{\int_{-\infty}^{\infty} K^2(t)\, dt\right\}^{4/5} \theta^{1/5} n^{-4/5} + o(n^{-4/5}). \tag{1.6}$$

In practice, of course, c^* is unknown and must be estimated from the data.

Suppose that we wish to estimate the unknown density f with MISE no larger than w, where w is a preassigned positive bound. In other words, the goal is to achieve sufficiently good global fit, where lack of fit is measured by the MISE. If an asymptotically optimal bandwidth is used [i.e., one for which the MISE behaves as in (1.6)], a reasonable way to proceed is to require

$$\tfrac{5}{4} k_2^{2/5}\left\{\int_{-\infty}^{\infty} K^2(t)\, dt\right\}^{4/5} \theta^{1/5} n^{-4/5} \le w. \tag{1.7}$$

This yields the sample size n_w^*, which is the smallest integer greater than or equal to

$$\left[\frac{5}{4} k_2^{2/5}\left\{\int_{-\infty}^{\infty} K^2(t)\, dt\right\}^{4/5} \frac{\theta^{1/5}}{w}\right]^{5/4}. \tag{1.8}$$

However, if θ is unknown, this sample size is also unknown. The expression for n_w^* suggests the following sequential procedure. Define the stopping rule T_w = first n such that

$$n \ge \left[\frac{5}{4} k_2^{2/5}\left\{\int_{-\infty}^{\infty} K^2(t)\, dt\right\}^{4/5} \frac{\{\hat{\theta}_n + b_n\}^{1/5}}{w}\right]^{5/4}, \tag{1.9}$$

where $\hat{\theta}_n$ is an estimate of θ based on $X_1, \dots, X_n$, and $b_n > 0$, $b_n \to 0$. Once sampling is terminated, f will be estimated by

$$\hat{f}_{T_w}(x|\hat{h}_{T_w}) = \frac{1}{T_w \hat{h}_{T_w}} \sum_{i=1}^{T_w} K\left(\frac{x - X_i}{\hat{h}_{T_w}}\right), \tag{1.10}$$

where

$$\hat{h}_n = k_2^{-2/5}\left\{\int_{-\infty}^{\infty} K^2(t)\,dt\right\}^{1/5}\left\{\hat{\theta}_{BW,n} + b_n\right\}^{-1/5} n^{-1/5} \tag{1.11}$$

is an estimate of the optimal bandwidth that is in turn based on an estimate $\hat{\theta}_{BW,n}$ of θ.

The estimate $\hat{\theta}_n$ used in the stopping rule T_w will be either

$$\hat{\theta}_n^K(\tilde{h}_n) = \int_{-\infty}^{\infty}\left(\hat{f}_n''(x|\tilde{h}_n)\right)^2 dx, \tag{1.12}$$

where $\hat{f}_n(\cdot|\tilde{h}_n)$ is a kernel estimate of f using the kernel K but with a bandwidth $\tilde{h}_n$ that is wider than h_n, or else

$$\hat{\theta}_n^{\mathrm{BC}K}(\tilde{h}_n) = (n/(n-1))\left\{\hat{\theta}_n^K(\tilde{h}_n) - n^{-1}\tilde{h}_n^{-5}K^{(2)} * K^{(2)}(0)\right\}, \tag{1.13}$$

where $*$ denotes convolution. $\hat{\theta}_n^{\mathrm{BC}K}(\tilde{h}_n)$ is in effect a bias-corrected version of $\hat{\theta}_n^K$ based on the work of Hall and Marron (1987). Both estimates assume of course that K is at least twice differentiable. The estimate $\hat{\theta}_{BW,n}$ used to estimate the optimal bandwidth will be either $\hat{\theta}_n^K(\tilde{h}_{BW,n})$ or $\hat{\theta}_n^{\mathrm{BC}K}(\tilde{h}_{BW,n})$. The construction of the plug-in bandwidth $\hat{h}_n$ follows the ideas of Woodroofe (1970) and Nadaraya (1974); for more sophisticated versions and a comparison of them with least squares cross-validation, see Park and Marron (1990). The constant b_n is added to the estimate of θ for technical reasons that will be apparent in the proof of the following theorem.

THEOREM 1. *Assume f is bounded and twice differentiable, that f'' is uniformly continuous and square integrable and that K is a symmetric probability density function with compact support and continuous third derivative. Assume further that $\tilde{h}_n = c_1 n^{-\alpha}$ for $c_1 > 0$ and $0 < \alpha < 1/10$,*

$$b_n = c_2 n^{(10\alpha - 1)/2 + \beta} \tag{1.14}$$

for some $c_2 > 0$ and $0 < \beta < (1 - 10\alpha)/2$, and $\tilde{h}_{BW,n} = c_3 n^{-\alpha_{BW}}$ for some $c_3 > 0$ and $0 < \alpha_{BW} < 1/50$. Then as $w \to 0$,

$$T_w/n_w^* \to 1 \quad a.s., \tag{1.15}$$

$$E(T_w)/n_w^* \to 1, \tag{1.16}$$

$$\left[\int_{-\infty}^{\infty}\left(\hat{f}_{T_w}(x|\hat{h}_{T_w}) - f(x)\right)^2 dx\right]\frac{1}{w} \to 1 \quad a.s., \tag{1.17}$$

and

$$E\left[\int_{-\infty}^{\infty}\left(\hat{f}_{T_w}\left(x|\hat{h}_{T_w}\right)-f(x)\right)^2 dx\right]\frac{1}{w}\to 1, \tag{1.18}$$

where the estimates of θ are those given above.

REMARK. The assumption that K has compact support is not necessary, but it simplifies the conditions that must be placed on K. The conclusions of the theorem also hold for normal kernels, for example.

The proof of Theorem 1 is given in the next section. (1.15)–(1.18) assert that the sequential procedure with stopping rule T_w is asymptotically efficient, in the sense that its MISE (and also ISE) are asymptotically equivalent to the desired bound w, and this is accomplished with a sample size that is asymptotically equivalent to n_w^*. All asymptotics are for the case when the bound on the MISE approaches zero, that is, increasingly better fit (in the limit, infinitely good fit) is desired.

The sequential procedure given above requires a great deal of computation, in part because the estimate of θ must be completely recomputed after each new observation. This suggests the desirability of using a recursive kernel estimate of f'' [see, for instance, Wolverton and Wagner (1969) and Yamato (1971)]. Unfortunately, results parallel to (1.15)–(1.18) have yet to be obtained for sequential procedures based on such recursive estimates. An alternative method for cutting down the amount of computation is to resort to two-stage, rather than fully sequential, procedures. In view of (1.9) and the assumptions of Theorem 1, define $m_1(w)$ to be the smallest integer greater than or equal to

$$w^{-10/(9-10\alpha-2\beta)}$$

and $m_2(w)$ to be the smallest integer greater than or equal to

$$\left[\left[\frac{5}{4}k_2^{2/5}\left\{\int_{-\infty}^{\infty}K^2(t)\,dt\right\}^{4/5}\frac{\left\{\hat{\theta}_{m_1(w)}+b_{m_1(w)}\right\}^{1/5}}{w}\right]^{5/4}-m_1(w)\right]^+,$$

and estimate f by

$$\hat{f}_{m_1(w)+m_2(w)}\left(x|\hat{h}_{m_1(w)+m_2(w)}\right),$$

where $\hat{h}_n$ is given by (1.11). Results analogous to those in Theorem 1 may be obtained for this two-stage procedure. For details, see Martinsek (1990), which also contains theoretical and simulation results for bandwidths $h_n = cn^{-1/5}$, where $c > 0$ is fixed (nonrandom).

2. Proof of Theorem 1. We begin with a series of lemmas for nonrandom sample sizes. Assume throughout this section that the assumptions of Theorem 1 hold.

LEMMA 1. *As* $n \to \infty$,

$$\hat{\theta}_n^K(\tilde{h}_n) \to \theta \quad a.s., \tag{2.1}$$

and

$$\hat{\theta}_n^{BCK}(\tilde{h}_n) \to \theta \quad a.s. \tag{2.2}$$

PROOF. This follows from Silverman (1978) and Lemma 1 of Nadaraya (1974). □

Define

$$a_n(x|h_n) = \int_{-\infty}^{\infty} h_n^{-1} K((x-y)/h_n) f(y)\, dy = E(\hat{f}_n(x|h_n))$$

and

$$\tilde{a}_n(x|\tilde{h}_n) = E(\hat{f}_n''(x|\tilde{h}_n)).$$

LEMMA 2. *For every positive integer* m,

$$E\left\{\left|\hat{\theta}_n^K(\tilde{h}_n) - \int_{-\infty}^{\infty} \tilde{a}_n^2(x|\tilde{h}_n)\, dx\right|^{2m}\right\} = O(n^{-m(1-10\alpha)}) \tag{2.3}$$

and

$$E\left\{\left|\hat{\theta}_n^{BCK}(\tilde{h}_n) - \int_{-\infty}^{\infty} \tilde{a}_n^2(x|\tilde{h}_n)\, dx\right|^{2m}\right\} = O(n^{-m(1-10\alpha)}). \tag{2.4}$$

PROOF. This is immediate from the proof of Lemma 3 in Nadaraya (1974) and (1.13). □

LEMMA 3. *For every positive integer* m, *if* $h_n n^{1/5} \to c > 0$, *then*

$$\begin{aligned} E\Bigg\{\Bigg|&\int_{-\infty}^{\infty} (\hat{f}_n(x|h_n) - a_n(x|h_n))^2\, dx \\ &- \left\{(nh_n)^{-1} \int_{-\infty}^{\infty} K^2(t)\, dt\right\} n^{-4/5}\Bigg|^{2m}\Bigg\} = O(n^{-9m/5}). \end{aligned} \tag{2.5}$$

PROOF. This follows from Lemma 8 and the proof of Lemma 4 in Stone (1984). □

LEMMA 4. *For every positive integer* m, *if* $h_n n^{1/5} \to c > 0$, *then*

$$\begin{aligned} &E\left\{\left|\int_{-\infty}^{\infty} (\hat{f}_n(x|h_n) - a_n(x|h_n))(a_n(x|h_n) - f(x))\, dx\right|^{2m}\right\} \\ &\quad = O(n^{-9m/5}). \end{aligned} \tag{2.6}$$

PROOF. By Bernstein's inequality [cf. the proof of Lemma 2 of Stone (1984)], if $(*)$ denotes the absolute value of the integral in (2.6), for positive t we have

$$P\left[(*) \geq tn^{-4/5}\right] \leq 2\exp\left(-n^{1/5}t^2/(2c' + (2/3)c't)\right), \tag{2.7}$$

for some $c' > 0$. It follows that

$$\begin{aligned} E\left\{\left(n^{4/5}(*)\right)^{2m}\right\} &\leq 4m\int_0^1 t^{2m-1}\exp\left(-3n^{1/5}t^2/(8c')\right)dt \\ &\quad + 4m\int_1^\infty t^{2m-1}\exp\left(-3n^{1/5}t/(8c')\right)dt. \end{aligned} \tag{2.8}$$

The second term on the right side of (2.8) is bounded by

$$4m\Gamma(2m)\left(8c'n^{-1/5}/3\right)^{2m} = O(n^{-2m/5}).$$

The first term on the right side of (2.8) can be shown by a change of variable to be $O(n^{-m/5})$ and (2.6) follows immediately. □

PROOF OF THEOREM 1. We have the inequalities

$$T_w \geq \left[\frac{5}{4}k_2^{2/5}\left\{\int_{-\infty}^\infty K^2(t)\,dt\right\}^{4/5}\frac{\left\{\hat\theta_{T_w} + b_{T_w}\right\}^{1/5}}{w}\right]^{5/4} \tag{2.9}$$

and

$$T_w - 1 < \left[\frac{5}{4}k_2^{2/5}\left\{\int_{-\infty}^\infty K^2(t)\,dt\right\}^{4/5}\frac{\left\{\hat\theta_{T_w-1} + b_{T_w-1}\right\}^{1/5}}{w}\right]^{5/4}. \tag{2.10}$$

Moreover, from (2.9),

$$T_w \geq \left[\frac{5}{4}c_2^{1/5}k_2^{2/5}\left\{\int_{-\infty}^\infty K^2(t)\,dt\right\}^{4/5}\frac{1}{w}\right]^{10/(9-10\alpha-2\beta)} \to \infty \tag{2.11}$$

as $w \to 0$.

(1.15) for both estimates of θ is an immediate consequence of Lemma 1 of Chow and Robbins (1965), Lemma 1 above, (2.9) and (2.10). (1.16) will follow from (1.15) if

$$\left\{w^{5/4}T_w\colon w \leq 1\right\} \text{ is uniformly integrable,} \tag{2.12}$$

and in view of (2.10) it suffices to establish uniform integrability of $\hat\theta_{T_w-1}^{1/4}$ for both estimates of θ. By Lemma 2 and the Cauchy-Schwarz inequality, for positive integers $m > 1/2(1 - 10\alpha)$, if LB(w) denotes the lower bound (2.11) for T_w,

$$
\begin{aligned}
&E\left\{\left|\hat{\theta}_{T_w-1} - \int_{-\infty}^{\infty} \tilde{a}^2_{T_w-1}\left(x|\tilde{h}_{T_w-1}\right) dx\right|^{2m}\right\} \\
&\quad = \sum_{n \geq \mathrm{LB}(w)-1} E\left\{\left|\hat{\theta}_n - \int_{-\infty}^{\infty} \tilde{a}^2_n\left(x|\tilde{h}_n\right) dx\right|^{2m} I_{\{T_w-1=n\}}\right\} \\
&\quad \leq \sum_{n \geq \mathrm{LB}(w)-1} E^{1/2}\left\{\left|\hat{\theta}_n - \int_{-\infty}^{\infty} \tilde{a}^2_n\left(x|\tilde{h}_n\right) dx\right|^{4m}\right\} P^{1/2}(T_w - 1 = n) \\
&\quad = O(1) \sum_{n \geq \mathrm{LB}(w)-1} n^{-m(1-10\alpha)} P^{1/2}(T_w - 1 = n) \\
&\quad \leq O(1) \left(\sum_{n \geq \mathrm{LB}(w)-1} n^{-2m(1-10\alpha)}\right)^{1/2} \left(\sum_{n \geq \mathrm{LB}(w)-1} P(T_w - 1 = n)\right)^{1/2} \\
&\quad = O(1) \left(\sum_{n \geq \mathrm{LB}(w)-1} n^{-2m(1-10\alpha)}\right)^{1/2} = o(1)
\end{aligned} \tag{2.13}
$$

as $w \to 0$, since $\mathrm{LB}(w) \to \infty$. Moreover, by Lemma 1 of Nadaraya (1974),

$$
\sup_w E\left\{\left|\int_{-\infty}^{\infty} \tilde{a}^2_{T_w-1}\left(x|\tilde{h}_{T_w-1}\right) dx - \theta\right|^{2m}\right\} < \infty. \tag{2.14}
$$

It follows from (2.13) and (2.14) that all positive powers of $\hat{\theta}_{T_w-1}$ are uniformly integrable, proving (1.16).

Define

$$
\begin{aligned}
\tau_n = {} & k_2^{-2/5}\left\{\int_{-\infty}^{\infty} K^2(t)\, dt\right\}^{1/5} \\
& \times \left\{\int_{-\infty}^{\infty} E^2\left(\hat{f}''_n\left(x|\tilde{h}_{\mathrm{BW},n}\right)\right) dx + b_n\right\}^{-1/5} n^{-1/5}.
\end{aligned} \tag{2.15}
$$

To establish (1.17) and (1.18), we will first prove analogous results for $\hat{f}_n(\cdot|\tau_n)$.

Write

$$
\begin{aligned}
&\left[\int_{-\infty}^{\infty}\left(\hat{f}_{T_w}(x|\tau_{T_w}) - f(x)\right)^2 dx\right]\frac{1}{w} \\
&\quad = \left\{\frac{5}{4} k_2^{2/5}\left(\int_{-\infty}^{\infty} K^2(t)\, dt\right)^{4/5} \theta^{1/5}\right\}\frac{1}{wT_w^{4/5}} \\
&\qquad + \left[T_w^{4/5}\int_{-\infty}^{\infty}\left(\hat{f}_{T_w}(x|\tau_{T_w}) - f(x)\right)^2 dx\right. \\
&\qquad\qquad \left. - \frac{5}{4} k_2^{2/5}\left\{\int_{-\infty}^{\infty} K^2(t)\, dt\right\}^{4/5} \theta^{1/5}\right]\frac{1}{wT_w^{4/5}}.
\end{aligned} \tag{2.16}
$$

From (1.15), the first term on the right side of (2.16) converges almost surely to 1. From Lemmas 2 and 4 of Stone (1984) and Lemma 1 of Nadaraya (1974),

$$n^{4/5}\int_{-\infty}^{\infty}\left(\hat{f}_n(x|\tau_n) - f(x)\right)^2 dx - \left\{\frac{5}{4}k_2^{2/5}\left(\int_{-\infty}^{\infty}K^2(t)\,dt\right)^{4/5}\theta^{1/5}\right\} \to 0 \quad \text{a.s.},$$

and in view of (1.15), this implies that the second term on the right side of (2.16) vanishes almost surely, proving

$$\left[\int_{-\infty}^{\infty}\left(\hat{f}_{T_w}(x|\tau_{T_w}) - f(x)\right)^2 dx\right]\frac{1}{w} \to 1 \quad \text{a.s.} \tag{2.17}$$

An argument similar to (2.13) shows that all negative powers of $w^{5/4}T_w$ are uniformly integrable (the rate on b_n plays a crucial role here). Hence from (1.15), the expectation of the first term on the right side of (2.16) converges to 1. Using Lemmas 3 and 4 above, an argument similar to (2.13) and Lemma 1 of Nadaraya (1974), it can be shown that all moments of

$$\left|T_w^{4/5}\int_{-\infty}^{\infty}\left(\hat{f}_{T_w}(x|\tau_{T_w}) - a_{T_w}(x|\tau_{T_w})\right)^2 dx - k_2^{2/5}\left\{\int_{-\infty}^{\infty}K^2(t)\,dt\right\}^{4/5}\theta^{1/5}\right|$$

and

$$\left|T_w^{4/5}\int_{-\infty}^{\infty}\left(\hat{f}_{T_w}(x|\tau_{T_w}) - a_{T_w}(x|\tau_{T_w})\right)\left(a_{T_w}(x|\tau_{T_w}) - f(x)\right) dx\right|$$

vanish as $w \to 0$. Moreover, since

$$n^{4/5}\int_{-\infty}^{\infty}\left(a_n(x|\tau_n) - f(x)\right)^2 dx \to k_2^{2/5}\left\{\int_{-\infty}^{\infty}K^2(t)\,dt\right\}^{4/5}\frac{\theta^{1/5}}{4}$$

as $n \to \infty$ [this is part of the proof of (1.3)], an easy argument shows that all moments of

$$T_w^{4/5}\int_{-\infty}^{\infty}\left(a_{T_w}(x|\tau_{T_w}) - f(x)\right)^2 dx - k_2^{2/5}\left\{\int_{-\infty}^{\infty}K^2(t)\,dt\right\}^{4/5}\frac{\theta^{1/5}}{4}$$

vanish as $w \to 0$. Combining these observations with uniform integrability of all negative powers of $w^{5/4}T_w$ and the equality

$$\begin{aligned}
&\int_{-\infty}^{\infty}\left(\hat{f}_{T_w}(x|\tau_{T_w}) - f(x)\right)^2 dx \\
&\quad = \int_{-\infty}^{\infty}\left(\hat{f}_{T_w}(x|\tau_{T_w}) - a_{T_w}(x|\tau_{T_w})\right)^2 dx \\
&\qquad + 2\int_{-\infty}^{\infty}\left(\hat{f}_{T_w}(x|\tau_{T_w}) - a_{T_w}(x|\tau_{T_w})\right)\left(a_{T_w}(x|\tau_{T_w}) - f(x)\right) dx \\
&\qquad + \int_{-\infty}^{\infty}\left(a_{T_w}(x|\tau_{T_w}) - f(x)\right)^2 dx,
\end{aligned}$$

shows that the expectation of the second term on the right side of (2.16) vanishes as $w \to 0$, and this shows

$$(2.18) \qquad E\left[\int_{-\infty}^{\infty}\left(\hat{f}_{T_w}(x|\tau_{T_w}) - f(x)\right)^2 dx\right]\frac{1}{w} \to 1.$$

By a slight generalization of the proof of Theorem 2 in Nadaraya (1974), we have for all positive integers m,

$$(2.19) \quad E\left[\left\{\int_{-\infty}^{\infty}\left(\hat{f}_n(x|\hat{h}_n) - \hat{f}_n(x|\tau_n)\right)^2 dx\right\}^{2m}\right] = O(n^{-2m+20m\alpha_{BW}}).$$

Because $\alpha_{BW} < 1/50$, it follows from (2.19) that for all positive integers $m > 5/(4 - 200\alpha_{BW})$,

$$\begin{aligned}
&E\left[(T_w)^{8m/5}\left\{\int_{-\infty}^{\infty}\left(\hat{f}_{T_w}(x|\hat{h}_{T_w}) - \hat{f}_{T_w}(x|\tau_{T_w})\right)^2 dx\right\}^{2m}\right]\\
&\quad = \sum_{n \geq \mathrm{LB}(w)} E\left[n^{8m/5}\left\{\int_{-\infty}^{\infty}\left(\hat{f}_n(x|\hat{h}_n) - \hat{f}_n(x|\tau_n)\right)^2 dx\right\}^{2m} I_{\{T_w = n\}}\right]\\
(2.20)\quad &\quad \leq \sum_{n \geq \mathrm{LB}(w)} E^{1/2}\left[n^{16m/5}\left\{\int_{-\infty}^{\infty}\left(\hat{f}_n(x|\hat{h}_n) - \hat{f}_n(x|\tau_n)\right)^2 dx\right\}^{4m}\right] P^{1/2}(T_w = n)\\
&\quad = O(1)\sum_{n \geq \mathrm{LB}(w)} n^{8m/5 - 2m + 20m\alpha_{BW}} P^{1/2}(T_w = n)\\
&\quad \leq O(1)\left(\sum_{n \geq \mathrm{LB}(w)} n^{16m/5 - 4m + 40m\alpha_{BW}}\right)^{1/2} = o(1)
\end{aligned}$$

as $w \to 0$. Because all negative powers of $w^{5/4}T_w$ are uniformly integrable, by the Cauchy–Schwarz inequality and (2.20),

$$(2.21) \qquad \frac{1}{w}E\left[\left\{\int_{-\infty}^{\infty}\left(\hat{f}_{T_w}(x|\hat{h}_{T_w}) - \hat{f}_{T_w}(x|\tau_{T_w})\right)^2 dx\right\}\right] = o(1)$$

as $w \to 0$. (1.18) now follows from (2.18) and (2.21).

To prove (1.17), note that from Silverman (1978),

$$(2.22) \qquad n^{1/5}\left(\hat{h}_n - \tau_n\right) = o(n^{-41/100}) \quad \text{a.s.}$$

(recall the assumption on $\tilde{h}_{BW,n}$). Combining (2.22) and an almost sure version of the method used in the proof of Theorem 2 of Nadaraya (1974) yields

$$(2.23) \qquad n^{4/5}\int_{-\infty}^{\infty}\left(\hat{f}_n(x|\hat{h}_n) - \hat{f}_n(x|\tau_n)\right)^2 dx \to 0 \quad \text{a.s.},$$

as $n \to \infty$. (1.17) now follows from (1.15), (2.11), (2.17) and (2.23). □

REFERENCES

CARROLL, R. J. (1976). On sequential density estimation. *Z. Wahrsch. Verw. Gebiete* **36** 137–151.

CHOW, Y. S. and ROBBINS, H. (1965). On the asymptotic theory of fixed-width confidence intervals for the mean. *Ann. Math. Statist.* **36** 457–462.

EFROIMOVICH, S. Y. (1989). Sequential nonparametric estimation of a density. *Theory Probab. Appl.* **34** 228–239.

HALL, P. and MARRON, J. S. (1987). Estimation of integrated squared density derivatives. *Statist. Probab. Lett.* **6** 109–115.

ISOGAI, E. (1981). Stopping rules for sequential density estimation. *Bulletin of Mathematical Statistics* **19** 53–67.

ISOGAI, E. (1987). The convergence rate of fixed-width sequential confidence intervals for a probability density function. *Sequential Anal.* **6** 55–69.

ISOGAI, E. (1988). A note on sequential density estimation. *Sequential Anal.* **7** 11–21.

KORONACKI, J. and WERTZ, W. (1988). A global stopping rule for recursive density estimators. *J. Statist. Plann. Inference* **20** 23–39.

MARTINSEK, A. T. (1990). Using stopping rules to bound the mean integrated squared error in density estimation. Technical report, Univ. Illinois.

NADARAYA, E. A. (1974). On the integral mean square error of some nonparametric estimates for the density function. *Theory Probab. Appl.* **19** 133–141.

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

PRAKASA RAO, B. L. S. (1983). *Nonparametric Functional Estimation*. Academic, Orlando.

ROSENBLATT, M. (1956). Remarks on some nonparametric estimates of a density function. *Ann. Math. Statist.* **27** 832–837.

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

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.

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

STUTE, W. (1983). Sequential fixed-width confidence intervals for a nonparametric density function. *Z. Wahrsch. Verw. Gebiete* **62** 113–123.

WOLVERTON, C. T. and WAGNER, T. J. (1969). Recursive estimates of probability densities. *IEEE Transactions on Systems Science and Cybernetics* **5** 246–247.

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

YAMATO, H. (1971). Sequential estimation of a continuous probability density function and mode. *Bulletin of Mathematical Statistics* **14** 1–12.

DEPARTMENT OF STATISTICS
UNIVERSITY OF ILLINOIS
101 ILLINI HALL
725 SOUTH WRIGHT STREET
CHAMPAIGN, ILLINOIS 61820