The Annals of Statistics
1995, Vol. 23, No. 6, 2218–2240

SEQUENTIAL CONFIDENCE BANDS FOR DENSITIES[1]

BY YI XU AND ADAM T. MARTINSEK

University of Illinois

This paper proposes a fully sequential procedure for constructing a fixed-width confidence band for an unknown density on a finite interval and shows the procedure has the desired coverage probability asymptotically as the width of the band approaches zero. The procedure is based on a result of Bickel and Rosenblatt. Its implementation in the sequential setting cannot be obtained using Anscombe's theorem, because the normalized maximal deviations between the kernel estimate and the true density are not uniformly continuous in probability (u.c.i.p.). Instead, we obtain a slightly weaker version of the u.c.i.p. property and a correspondingly stronger convergence property of the stopping rule. These together yield the desired results.

1. Introduction. Let $X_1, X_2, \ldots, X_n$ be independent and identically distributed random variables with continuous density function $f(x)$. A familiar method for estimating f is the kernel estimate due to Rosenblatt (1956) and Parzen (1962), which is given by

$$\hat{f}_n(x) = (nh_n)^{-1} \sum_1^n K[(x - X_i)/h_n], \tag{1.1}$$

where K is a bounded density function called the kernel and h_n is the bandwidth. There have been several papers on sequential density estimates. Carroll (1976) considered the problem of sequential estimation of the density f at a point x_0 which may be known or unknown and obtained sequential fixed-width confidence intervals. A similar problem was approached also by Stute (1983), where a different method was proposed. Isogai (1981, 1987, 1988) also considered aspects of this problem. However, all of them considered only estimation of f at a fixed point so that asymptotic normality can be applied. As mentioned in Stute (1983), it seems more interesting to construct sequential fixed-width confidence bands for f on a bounded interval $[a, b]$.

Our goal is to construct a confidence band for f on $[a, b]$, with approximate coverage probability $1 - \alpha$, that determines each value of the density to

Received February 1994; revised April 1995.

[1]This paper is based on the Ph.D. thesis of the first author, written under the direction of the second author.

AMS 1991 *subject classifications*. Primary 62L12; secondary 62G07, 62G20.

Key words and phrases. Density estimation, confidence band, sequential estimation, stopping rule, uniform continuity in probability.

within $\pm\varepsilon$, where ε is the desired precision. One motivation for wanting a fixed-width confidence band is that density estimates are often used for classification purposes in situations where misclassification can have serious consequences. One example is the heart rate data for heart disease patients and "normals" discussed in Izenman (1991). Without sufficiently precise estimates of the densities, accurate classification may be problematic.

We assume that f and K satisfy the following assumptions:

(A.1) K is a bounded, symmetric probability density function with finite second moment such that either (i) K has compact support and is absolutely continuous on its support or (ii) K is absolutely continuous on the real line with integrable and square integrable derivative;
(A.2) f is continuous, bounded and bounded away from zero on $[a, b]$;
(A.3) f is twice differentiable with bounded second derivative on $[a, b]$;
(A.4) $\int_{|x| \geq 3} (|x|)^{3/2} (\log\log|x|)^{1/2} (|K'(x)| + |K(x)|)\, dx < \infty$;
(A.5) the interval $[a, b]$ contains no boundary points of f.

Without loss of generality, take $a = 0$ and $b = 1$. Define

$$M_n = \sup_{0 \leq x \leq 1} \left[\frac{nh_n}{f(x)} \right]^{1/2} \left| \hat{f}_n(x) - f(x) \right|.$$

Then Bickel and Rosenblatt (1973) have shown that, for $h_n = n^{-\delta}$, $\frac{1}{5} < \delta < \frac{1}{2}$, as $n \to \infty$,

$$P\left[(2\delta \log n)^{1/2} \left\{ \left(\int K^2(t)\, dt \right)^{-1/2} M_n - d_n \right\} < z \right] \to \exp(-2\exp(-z)), \tag{1.2}$$

where

$$d_n = (2\delta \log n)^{1/2} + (2\delta \log n)^{-1/2} \left\{ \log \left(\frac{K_1(A)}{\pi^{1/2} \int K^2(t)\, dt} \right) + \frac{1}{2} [\log \delta + \log\log n] \right\}$$

if (A.1)(i) holds and the support of K is $[-A, A]$, $K_1(A) = \{K^2(A) + K^2(-A)\}/2$; otherwise,

$$d_n = (2\delta \log n)^{1/2} + (2\delta \log n)^{-1/2} \left\{ \log \left[(2\pi)^{-1} \left(\frac{\int (K'(t))^2\, dt}{\int K^2(t)\, dt} \right)^{1/2} \right] \right\}.$$

In view of the uniform almost-sure convergence of $\hat{f}_n$ to f, (1.2) implies

$$P\left[f(x) \in \hat{f}_n(x) \pm \left\{ \frac{\hat{f}_n(x) \int K^2(t)\,dt}{nh_n} \right\}^{1/2} \times \left\{ \frac{z}{(2\delta \log n)^{1/2}} + d_n \right\} \forall\, 0 \le x \le 1 \right] \to \exp(-2\exp(-z)). \tag{1.3}$$

It follows from (1.3) that if $z = z_\alpha$, where $0 < \alpha < 1$ and

$$\exp(-2\exp(-z_\alpha)) = 1 - \alpha,$$

then

$$P\left[f(x) \in \hat{f}_n(x) \pm \left\{ \frac{\hat{f}_n(x) \int K^2(t)\,dt}{nh_n} \right\}^{1/2} \left\{ \frac{z_\alpha}{(2\delta \log n)^{1/2}} + d_n \right\} \forall\, 0 \le x \le 1 \right] \to 1 - \alpha.$$

We wish to construct a confidence band for f that determines each of the values $f(x)$ to within $\pm\varepsilon$ in such a way that the probability that the band covers the entire function f is approximately $1 - \alpha$. It is natural to use the stopping rule

$$T_\varepsilon = \inf\left\{ n \ge 1, \left[\sup_{0 \le x \le 1} \frac{\hat{f}_n(x) \int K^2(t)\,dt}{nh_n} \right]^{1/2} \times \left[\frac{z_\alpha}{(2\delta \log n)^{1/2}} + d_n \right] \le \varepsilon \right\}. \tag{1.4}$$

Once sampling has terminated, one can use either of two confidence bands. The simpler one is the fixed-width band

$$\hat{f}_{T_\varepsilon}(x) \pm \varepsilon, \qquad 0 \le x \le 1.$$

The other possibility is the variable-width band

$$\hat{f}_{T_\varepsilon}(x) \pm \left\{ \frac{\hat{f}_{T_\varepsilon}(x) \int K^2(t)\,dt}{T_\varepsilon^{1-\delta}} \right\}^{1/2} \left\{ \frac{z_\alpha}{(2\delta \log T_\varepsilon)^{1/2}} + d_{T_\varepsilon} \right\}, \qquad 0 \le x \le 1.$$

Note that the variable-width band has half-width everywhere smaller than ε. The main purpose of this paper is to prove

$$P\left\{ f(x) \in \hat{f}_{T_\varepsilon}(x) \pm \left\{ \frac{\hat{f}_{T_\varepsilon}(x) \int K^2(t)\,dt}{T_\varepsilon^{1-\delta}} \right\}^{1/2} \left\{ \frac{z_\alpha}{(2\delta \log T_\varepsilon)^{1/2}} + d_{T_\varepsilon} \right\} \forall\, 0 \le x \le 1 \right\} \to 1 - \alpha$$

as $\varepsilon \to 0$ and

$$\liminf_{\varepsilon \to 0} P\Big\{ f(x) \in \hat{f}_{T_\varepsilon}(x) \pm \varepsilon, \forall\, 0 \le x \le 1 \Big\} \ge 1 - \alpha.$$

The first result of course implies the second. A natural way to approach this is by using Anscombe's theorem, which we now state.

ANSCOMBE'S THEOREM [Anscombe (1952)]. *Assume that* $Z_n \to_d Z$ *as* $n \to \infty$ *and the sequence* $\{Z_n\colon n \ge 1\}$ *is uniformly continuous in probability* (*u.c.i.p.*); *that is*, *for every* $\rho_1 > 0$ *there is a* $\rho_2 > 0$ *for which*

$$P\Big\{ \max_{0 \le k \le \rho_2 n} |Z_{n+k} - Z_n| \ge \rho_1 \Big\} < \rho_1 \quad \forall\, n \ge 1.$$

Assume further that $\{N_b\colon b > 0\}$ *is a sequence of integer-valued random variables for which* $N_b/b \to \lambda > 0$ *in probability as* $b \to \infty$, *where* λ *is a constant*. *Then*

$$Z_{N_b} \to_d Z \qquad \text{as } b \to \infty.$$

However, Martinsek (1993) presents an argument that suggests strongly that

$$(2\delta \log n)^{1/2} \Big\{ \Big(\int K^2(t)\, dt \Big)^{-1/2} M_n - d_n \Big\}$$

is not u.c.i.p. Careful examination of the arguments in Section 2 below provides a rigorous proof that this sequence is not u.c.i.p. [note that the order in (2.4) is exact]. Therefore Anscombe's theorem cannot be applied directly. Martinsek (1993) proposes a two-stage procedure to overcome this problem. Unfortunately, his procedure has the disadvantage that the first sample is used only to determine the size of the second sample and not in the final estimate of f. Hence the data are not fully utilized. Also, two-stage procedures typically are less efficient than fully sequential ones.

Our approach here is to get a slightly weaker u.c.i.p. property of the normalized maximum deviations and a correspondingly stronger convergence result for T_ε, so that the conclusion of Anscombe's theorem will still hold. We will use the Komlós–Major–Tusnády representation [Komlós, Major and Tusnády (1975)] for the empirical distribution function and a result on the modulus of continuity of Brownian motion. The main result is the following.

THEOREM 1. *Assume that conditions* (A.1)–(A.5) *hold and that if* (A.1)(i) *holds*, *then*

$$\int |K'(t)|\, dt < \infty.$$

Assume further that the bandwidth $h_n = n^{-\delta}$ *is used in the stopping rule as well as in the final estimate. Then*

$$P\left\{ f(x) \in \hat{f}_{T_\varepsilon}(x) \pm \left\{ \frac{\hat{f}_{T_\varepsilon}(x) \int K^2(t)\, dt}{T_\varepsilon^{1-\delta}} \right\}^{1/2} \times \left\{ \frac{z_\alpha}{(2\delta \log T_\varepsilon)^{1/2}} + d_{T_\varepsilon} \right\} \forall\, 0 \le x \le 1 \right\} \to 1 - \alpha \tag{1.5}$$

as $\varepsilon \to 0$ *and*

$$\liminf_{\varepsilon \to 0} P\Big[f(x) \in \hat{f}_{T_\varepsilon} \pm \varepsilon, \forall\, 0 \le x \le 1 \Big] \ge 1 - \alpha, \tag{1.6}$$

provided $\frac{1}{5} < \delta < \frac{1}{2}$.

The proof of Theorem 1 is given in Sections 2 and 3. Section 4 contains some simulation results. Results for censored data will appear elsewhere.

2. Some lemmas. The proof of Theorem 1 relies on two main results: the modified uniform continuity in probability result in Corollary 1 and the approximation of T_ε in Lemma 7. Corollary 1 follows easily from Lemma 3, which in turn relies on the Komlós–Major–Tusnády approximation of the empirical distribution function (Lemma 1) and the modulus of continuity of Brownian motion (2.4). Lemma 2, on the uniform rate of convergence of the density estimate $\hat{f}_n$, plays a key role in the arguments leading to the approximation in Lemma 7. Lemmas 4, 5 and 6 all lead to Lemma 7. Lemma 4 is technical. Lemma 5 gives an initial approximation of T_ε by a quantity n_ε^*. The order of approximation obtained there is used in Lemma 6 to refine n_ε^* and to produce a better approximation n_0. Finally, n_0 is refined in Lemma 7 to produce a still better approximation n_1. This final approximation will, together with Corollary 1, suffice to prove Theorem 1.

LEMMA 1. *Let*

$$Z_n^0(F(s)) = n^{1/2}(F_n(s) - F(s)),$$

where $F_n(s) = (1/n)\sum_1^n \mathscr{I}_{(X_i \le s)}$ *is the empirical distribution function. Then there exist a version of the sequence* Z_n^0 *and independent Brownian bridges* W_j^0, $j = 1, 2, \dots,$ *that is,* $W_j^0(t) = W_j(t) - tW_j(1)$, *where* $W_1, \dots$ *are independent Brownian motions, such that*

$$\sup_s \left| Z_n^0(F(s)) - \frac{1}{\sqrt{n}} \sum_1^n W_j^0(F(s)) \right| = O\left(\frac{(\log n)^2}{\sqrt{n}} \right) \quad a.s. \tag{2.1}$$

PROOF. This is immediate from Theorem 4 of Komlós, Major and Tusnády (1975). □

We will prove Theorem 1 for the version of Z_n^0 for which (2.1) holds. The general case will then follow immediately. The following lemmas apply to the version for which (2.1) holds, but we do not bother to introduce new notation.

LEMMA 2. *Under the assumptions of Theorem* 1,

$$\sup_{0 \le x \le 1} \left| \hat{f}_n(x) - E\hat{f}_n(x) \right| = O\left(n^{-(1-\delta)/2} (\log n)^{1/2}\right) \quad a.s. \tag{2.2}$$

PROOF. Lemma 2 follows from Lemma 1 and results on the modulus of continuity of Brownian motion [see (2.4)]. See Zheng (1988) for a related argument, and see Stute (1982) for exact rates. □

LEMMA 3. *Under the assumptions of Theorem* 1, *let* $b(n) > 0$ *s.t.*

$$\frac{(\log n)^2}{b(n)} \to 0$$

and

$$Q_n = (\log n)^{1/2} \sup_{0 \le x \le 1} \left\{ \left(\frac{n h_n}{f(x)} \right)^{1/2} \left| \hat{f}_n(x) - E\hat{f}_n(x) \right| \right\}.$$

Then

$$\max_{0 \le k \le n/b(n)} |Q_n - Q_{n+k}| \to 0 \quad a.s. \tag{2.3}$$

PROOF. We will give the proof for the case when (A.1)(i) holds and $\int |K'(t)|\, dt < \infty$. The proof under assumption (A.1)(ii) may be found in Xu and Martinsek (1994). Write

$$\begin{aligned} Q_n &= n^{\delta/2} (\log n)^{1/2} \\ &\quad \times \sup_{0 \le x \le 1} \left\{ \frac{1}{(f(x))^{1/2}} \left| K(A)\left[Z_n^0(F(x - h_n A)) - Z_n^0(F(x + h_n A))\right] \right.\right. \\ &\qquad \left.\left. - \int_{-A}^{A} Z_n^0(F(x - h_n y)) K'(y)\, dy \right| \right\} \\ &= o(1) + n^{\delta/2} (\log n)^{1/2} \\ &\quad \times \sup_{0 \le x \le 1} \left\{ \frac{1}{(f(x))^{1/2}} \left| K(A)\left[B_n(F(x - h_n A)) - B_n(F(x + h_n A))\right] \right.\right. \\ &\qquad \left.\left. - \int_{0}^{A} \left[B_n(F(x - h_n y)) - B_n(F(x + h_n y))\right] K'(y)\, dy \right| \right\} \quad \text{a.s.}, \end{aligned}$$

where $B_n(t)$ is the Brownian motion $n^{-1/2} \sum_1^n W_j(t)$. It therefore suffices to show (2.3) with Q_n replaced by each of

$$\begin{aligned} P_{n1}^* &= n^{\delta/2} (\log n)^{1/2} \sup_{0 \le x \le 1} \frac{1}{(f(x))^{1/2}} \\ &\qquad \times \left| K(A)\left[B_n(F(x - h_n A)) - B_n(F(x + h_n A))\right] \right| \end{aligned}$$

and

$$P_{n2}^* = n^{\delta/2}(\log n)^{1/2} \sup_{0 \le x \le 1} \frac{1}{(f(x))^{1/2}} \times \left| \int_0^A [B_n(F(x - h_n y)) - B_n(F(x + h_n y))] K'(y)\, dy \right|.$$

We have

$$\begin{aligned}
\frac{|P_{n+k,1}^* - P_{n1}^*|}{K(A)} &= \Bigg| (\log(n+k))^{1/2}(n+k)^{\delta/2} \\
&\quad \times \sup_{0 \le x \le 1} \left\{ \frac{1}{\sqrt{f(x)}} \frac{1}{\sqrt{n+k}} \left| \sum_1^{n+k} W_j\Big(F\big(x + (n+k)^{-\delta}A\big)\Big) - \sum_1^{n+k} W_j\Big(F\big(x - (n+k)^{-\delta}A\big)\Big) \right| \right\} \\
&\quad - (\log n)^{1/2} n^{\delta/2} \sup_{0 \le x \le 1} \left\{ \frac{1}{\sqrt{f(x)}} \left| \frac{1}{\sqrt{n}} \sum_1^n W_j(F(x + n^{-\delta}A)) - \frac{1}{\sqrt{n}} \sum_1^n W_j(F(x - n^{-\delta}A)) \right| \right\} \Bigg| \\
&\le I_1 + I_2 + I_3 + I_4,
\end{aligned}$$

where

$$\begin{aligned}
I_1 &= (\log(n+k))^{1/2}(n+k)^{\delta/2} \left| \frac{1}{\sqrt{n+k}} - \frac{1}{\sqrt{n}} \right| \\
&\quad \times \sup_{0 \le x \le 1} \frac{1}{\sqrt{f(x)}} \left| \sum_1^n W_j\Big(F\big(x + (n+k)^{-\delta}A\big)\Big) - \sum_1^n W_j\Big(F\big(x - (n+k)^{-\delta}A\big)\Big) \right|,
\end{aligned}$$

$$\begin{aligned}
I_2 &= (\log(n+k))^{1/2}(n+k)^{\delta/2} \\
&\quad \times \left| \sup_{0 \le x \le 1} \left[\frac{1}{\sqrt{f(x)}} \frac{1}{\sqrt{n}} \left| \sum_1^n W_j\Big(F\big(x + (n+k)^{-\delta}A\big)\Big) - \sum_1^n W_j\Big(F\big(x - (n+k)^{-\delta}A\big)\Big) \right| \right] \right.
\end{aligned}$$

$$- \sup_{0 \le x \le 1} \left[\frac{1}{\sqrt{f(x)}} \frac{1}{\sqrt{n}} \left| \sum_1^n W_j(F(x + n^{-\delta}A)) - \sum_1^n W_j(F(x - n^{-\delta}A)) \right| \right| \Bigg|,$$

$$I_3 = (\log(n+k))^{1/2}(n+k)^{\delta/2} \times \sup_{0 \le x \le 1} \left\{ \frac{1}{\sqrt{f(x)}} \frac{1}{\sqrt{n+k}} \left| \sum_{n+1}^{n+k} \left\{ W_j\left(F\left(x + (n+k)^{-\delta} A\right)\right) - W_j\left(F\left(x - (n+k)^{-\delta} A\right)\right) \right\} \right| \right\},$$

$$I_4 = \left| (\log(n+k))^{1/2}(n+k)^{\delta/2} - (\log n)^{1/2} n^{\delta/2} \right| \times \sup_{0 \le x \le 1} \frac{1}{\sqrt{f(x)}} \frac{1}{\sqrt{n}} \left| \sum_1^n \left(W_j(F(x + n^{-\delta}A)) - W_j(F(x - n^{-\delta}A)) \right) \right|.$$

By Lemma 1 of Csörgő and Révész (1979) (let $\varepsilon = 1$) and the Borel–Cantelli lemma,

$$\sup_{0 \le t \le 1-h} \sup_{0 \le s \le h} |B_n(t+s) - B_n(t)| = O\left(\sqrt{h \log h^{-1}}\right) \quad \text{a.s.,} \tag{2.4}$$

when $h \to 0$. Note that h can depend on n. Using (2.4) with $h = c(n + n/b(n))^{-\delta}$,

$$\max_{0 \le k \le n/b(n)} I_1 = O\left(\frac{\log n}{b(n)}\right) \quad \text{a.s.}$$

Similarly,

$$\max_{0 \le k \le n/b(n)} I_3 = O\left(\frac{\log n}{\sqrt{b(n)}}\right) \to 0 \quad \text{a.s.}$$

Since

$$\max_{0 \le k \le n/b(n)} \left| (\log(n+k))^{1/2}(n+k)^{\delta/2} - (\log n)^{1/2} n^{\delta/2} \right| = O\left(\frac{n^{\delta/2-1}(\log n)^{1/2} n}{b(n)}\right) = O\left(\frac{n^{\delta/2}(\log n)^{1/2}}{b(n)}\right),$$

by (2.4),

$$\max_{0 \le k \le n/b(n)} I_4 = O\left(\frac{\log n}{b(n)}\right) \to 0 \quad \text{a.s.}$$

Finally,

$$
\begin{aligned}
I_2 \le{} & (\log(n+k))^{1/2}(n+k)^{\delta/2} \\
& \times \Bigg\{ \sup_{0 \le x \le 1} \frac{1}{\sqrt{f(x)}} \Bigg| \frac{1}{\sqrt{n}} \sum_1^n W_j\big(F\big(x + (n+k)^{-\delta} A\big)\big) \\
& \qquad\qquad - \frac{1}{\sqrt{n}} \sum_1^n W_j(F(x + n^{-\delta} A)) \Bigg| \\
& \quad + \sup_{0 \le x \le 1} \frac{1}{\sqrt{f(x)}} \Bigg| \frac{1}{\sqrt{n}} \sum_1^n W_j\big(F\big(x - (n+k)^{-\delta} A\big)\big) \\
& \qquad\qquad - \frac{1}{\sqrt{n}} \sum_1^n W_j(F(x - n^{-\delta} A)) \Bigg| \Bigg\}.
\end{aligned}
$$

Since

$$
\big| F\big(x + (n+k)^{-\delta} A\big) - F(x + n^{-\delta} A) \big| \le O\left(\frac{n^{-\delta}}{b(n)} \right)
$$

and

$$
\big| F\big(x - (n+k)^{-\delta} A\big) - F(x - n^{-\delta} A) \big| \le O\left(\frac{n^{-\delta}}{b(n)} \right),
$$

again we can use (2.4),

$$
\begin{aligned}
\max_{0 \le k \le n/b(n)} I_2 &= O\left((\log n)^{1/2} n^{\delta/2} \sqrt{\frac{n^{-\delta} \log n}{b(n)}} \right) \\
&= O\left(\frac{\log n}{b(n)^{1/2}} \right) \to 0 \quad \text{a.s.}
\end{aligned}
$$

Hence,

$$
\max_{0 \le k \le n/b(n)} |P^*_{n+k,1} - P^*_{n1}| \to 0 \quad \text{a.s.}
$$

Turning to the sequence P^*_{n2}, we have

$$
\begin{aligned}
& |P^*_{n+k,2} - P^*_{n2}| \\
& \quad \le (n+k)^{\delta/2} \sqrt{\log(n+k)} \left| \frac{1}{\sqrt{n+k}} - \frac{1}{\sqrt{n}} \right| \sup_{0 \le x \le 1} \frac{1}{(f(x))^{1/2}} \Bigg|
\end{aligned}
$$

$$\times \int_0^A \left[\sum_1^n W_j(F(x - h_n y)) - \sum_1^n W_j(F(x + h_n y)) \right] K'(y)\, dy \Bigg|$$

$$+ \Bigg| (n+k)^{\delta/2} \sqrt{\log(n+k)}$$

$$\times \Bigg\{ \sup_{0 \le x \le 1} \frac{1}{(f(x))^{1/2}} \int_0^A \left[\frac{1}{\sqrt{n}} \sum_1^n W_j(F(x - h_{n+k} y)) - \frac{1}{\sqrt{n}} \sum_1^n W_j(F(x + h_{n+k} y)) \right] K'(y)\, dy$$

$$- \sup_{0 \le x \le 1} \frac{1}{(f(x))^{1/2}} \int_0^A \left[\frac{1}{\sqrt{n}} \sum_1^n W_j(F(x - h_n y)) - \frac{1}{\sqrt{n}} \sum_1^n W_j(F(x + h_n y)) \right] K'(y)\, dy \Bigg\} \Bigg|$$

$$+ (n+k)^{\delta/2} \sqrt{\log(n+k)}$$

$$\times \sup_{0 \le x \le 1} \frac{1}{(f(x))^{1/2}} \left| \int_0^A \frac{1}{\sqrt{n+k}} \sum_{n+1}^{n+k} \left[W_j(F(x - h_{n+k} y)) - W_j(F(x + h_{n+k} y)) \right] K'(y)\, dy \right|$$

$$+ \left| (n+k)^{\delta/2} \sqrt{\log(n+k)} - n^{\delta/2} \sqrt{\log n} \right|$$

$$\times \sup_{0 \le x \le 1} \frac{1}{(f(x))^{1/2}} \left| \frac{1}{\sqrt{n}} \int_0^A \sum_1^n \left[W_j(F(x - h_n y)) - W_j(F(x + h_n y)) \right] K'(y)\, dy \right|.$$

The second term is bounded by a constant times

$$(n+k)^{\delta/2} \sqrt{\log(n+k)}$$

$$\times \left| \sup_{0 \le x \le 1} \int_{-A}^A \left[B_n(F(x - h_{n+k} y)) - B_n(F(x - h_n y)) \right] K'(y)\, dy \right|$$

$$+ (n+k)^{\delta/2} \sqrt{\log(n+k)}$$

$$\times \left| \sup_{0 \le x \le 1} \int_{-A}^A \left[B_n(F(x + h_{n+k} y)) - B_n(F(x + h_n y)) \right] K'(y)\, dy \right|$$

$$
\begin{aligned}
&\le (n+k)^{\delta/2}\sqrt{\log(n+k)} \\
&\quad \times \sup_{0\le s,t\le 1}\ \sup_{|s-t|\le(\mathrm{const})(h_n-h_{n+k})} |B_n(s)-B_n(t)|\left|\int_{-A}^{A}|K'(y)|\,dy\right. \\
&= O\Big(n^{\delta/2}\sqrt{\log n}\sqrt{(h_n-h_{n+k})\log\big((h_n-h_{n+k})^{-1}\big)}\Big) \\
&= O\left(\frac{\log n}{\sqrt{b(n)}}\right)\to 0 \quad \text{a.s.}
\end{aligned}
$$

The other terms can be dealt with similarly, so we have

$$
\max_{0\le k\le n/b(n)} |P^*_{n+k,2}-P^*_{n2}| \to 0 \quad \text{a.s., } n\to\infty.
$$

This proves (2.3) □

COROLLARY 1. *Under the assumptions of Theorem* 1, *for* $\frac{1}{5}<\delta<\frac{1}{2}$,

$$
\begin{aligned}
\max_{0\le k\le n/b(n)} \Big|&(2\delta\log(n+k))^{1/2}[M_{n+k}-d_{n+k}] \\
&-(2\delta\log n)^{1/2}(M_n-d_n)\Big| \to 0 \quad a.s.,
\end{aligned}
\tag{2.5}
$$

with $b(n)$ *as in Lemma* 3.

PROOF.

$$
\begin{aligned}
&\Big|(2\delta\log(n+k))^{1/2}[M_{n+k}-d_{n+k}]-(2\delta\log n)^{1/2}(M_n-d_n)\Big| \\
&\quad\le \sqrt{2\delta}|Q_{n+k}-Q_n| \\
&\qquad+\sup_{0\le x\le 1}\frac{1}{\sqrt{f(x)}}\Big|E\hat{f}_{n+k}(x)-f(x)\Big|(\log(n+k))^{1/2}(n+k)^{(1-\delta)/2} \\
&\qquad+\sup_{0\le x\le 1}\left(\frac{1}{\sqrt{f(x)}}\Big|E\hat{f}_n(x)-f(x)\Big|\right)(\log n)^{1/2}n^{(1-\delta)/2} \\
&\qquad+\Big|\sqrt{2\delta\log(n+k)}\,d_{n+k}-\sqrt{2\delta\log n}\,d_n\Big|.
\end{aligned}
$$

By a standard expansion using (A.3), we have

$$
\sup_{0\le x\le 1}\Big|E\hat{f}_n(x)-f(x)\Big| = O(h_n^2), \tag{2.6}
$$

so the left-hand side of (2.5) is of order

$$
O\Big(n^{(1-5\delta)/2}\sqrt{\log n}\Big)+O\left(\frac{1}{b(n)}\right)+o(1)\to 0 \quad \text{a.s.} \qquad \square
$$

LEMMA 4. *Define* $G(n)$ *by*

$$G(n) = \left(\frac{z_\alpha}{\sqrt{2\delta \log n}} + d_n\right)^{2/(1-\delta)} - \left(\frac{4\delta}{1-\delta}\log \varepsilon^{-1}\right)^{1/(1-\delta)}.$$

Then

$$G(n) = (2\delta \log n)^{1/(1-\delta)} - \left(\frac{4\delta}{1-\delta}\log \varepsilon^{-1}\right)^{1/(1-\delta)} + H(n) + O\big((\log n)^{1/(1-\delta)-3}(\log\log n)^3\big),$$

where

$$\begin{aligned} H(n) &= a_1(\log n)^{1/(1-\delta)-1}\log\log n + a_2(\log n)^{1/(1-\delta)-1} \\ &\quad + a_3(\log n)^{1/(1-\delta)-2}(\log\log n)^2 + a_4(\log n)^{1/(1-\delta)-2}\log\log n \\ &\quad + a_5(\log n)^{1/(1-\delta)-2}, \end{aligned} \tag{2.7}$$

for certain constants a_1, a_2, a_3, a_4 *and* a_5.

PROOF. The proof is obvious by the Taylor expansion. □

LEMMA 5. *Let*

$$n_\varepsilon^* = \left(\frac{4\delta}{1-\delta}\sup_{0\le x\le 1} f(x)\int K^2(t)\,dt\,\frac{\log \varepsilon^{-1}}{\varepsilon^2}\right)^{1/(1-\delta)}.$$

Under the conditions of Theorem 1,

$$\left|\frac{T_\varepsilon - n_\varepsilon^*}{n_\varepsilon^*}\right| = o\left(\frac{(\log\log \varepsilon^{-1})^2}{\log \varepsilon^{-1}}\right) \quad a.s.\ as\ \varepsilon \to 0. \tag{2.8}$$

PROOF. By the definition of T_ε,

$$\begin{aligned} 1 &+ \left[\frac{1}{\varepsilon}\left(\frac{z_\alpha}{\sqrt{2\delta\log(T_\varepsilon - 1)}} + d_{T_\varepsilon - 1}\right)\right. \\ &\qquad \left.\times\left(\sup_{0\le x\le 1}\hat f_{T_\varepsilon-1}(x)\int K^2(t)\,dt\right)^{1/2}\right]^{2/(1-\delta)} \\ &\ge T_\varepsilon \ge \left[\frac{1}{\varepsilon}\left(\frac{z_\alpha}{\sqrt{2\delta\log T_\varepsilon}} + d_{T_\varepsilon}\right)\right. \\ &\qquad \left.\times\left(\sup_{0\le x\le 1}\hat f_{T_\varepsilon}(x)\int K^2(t)\,dt\right)^{1/2}\right]^{2/(1-\delta)} \end{aligned} \tag{2.9}$$

and

$$\frac{n_\varepsilon^*(\log\log \varepsilon^{-1})^2}{\log \varepsilon^{-1}} = O\Big(\varepsilon^{-2/(1-\delta)}(\log \varepsilon^{-1})^{1/(1-\delta)-1}(\log\log \varepsilon^{-1})^2\Big). \tag{2.10}$$

Then

$$\begin{aligned}
&\frac{T_\varepsilon - n_\varepsilon^*}{n_\varepsilon^*(\log\log \varepsilon^{-1})^2/\log \varepsilon^{-1}}\\
&\quad\le \mathrm{const}(\log \varepsilon^{-1})^{1-1/(1-\delta)}(\log\log \varepsilon^{-1})^{-2}\\
&\quad\quad\times\Bigg[\Big(\sup_{0\le x\le 1}\hat f_{T_\varepsilon-1}(x)\Big)^{1/(1-\delta)}\Bigg(\frac{z_\alpha}{\sqrt{2\delta\log(T_\varepsilon-1)}}+d_{T_\varepsilon-1}\Bigg)^{2/(1-\delta)}\\
&\qquad\qquad -\Big(\sup_{0\le x\le 1} f(x)\Big)^{1/(1-\delta)}\Big(\frac{4\delta}{1-\delta}\log \varepsilon^{-1}\Big)^{1/(1-\delta)}+1\Bigg]\\
&\quad\le \mathrm{const}(\log \varepsilon^{-1})^{1-1/(1-\delta)}(\log\log \varepsilon^{-1})^{-2}\\
&\quad\quad\times\Big(\sup_{0\le x\le 1}\hat f_{T_\varepsilon-1}(x)\Big)^{1/(1-\delta)}G(T_\varepsilon-1)\\
&\quad\quad+\mathrm{const}(\log \varepsilon^{-1})^{1-1/(1-\delta)}(\log\log \varepsilon^{-1})^{-2}\Big(\frac{4\delta}{1-\delta}\log \varepsilon^{-1}\Big)^{1/(1-\delta)}\\
&\quad\quad\times\Bigg[\Big(\sup_{0\le x\le 1} f_{T_\varepsilon-1}(x)\Big)^{1/(1-\delta)}-\Big(\sup_{0\le x\le 1} f(x)\Big)^{1/(1-\delta)}\Bigg]\\
&\quad\quad+o(1).
\end{aligned}$$

By Lemma 4 and

$$\begin{aligned}
\log n_\varepsilon^* &= \frac{1}{1-\delta}\log\Big(\frac{4\delta}{1-\delta}\sup_{0\le x\le 1} f(x)\int K^2(t)\,dt\Big)\\
&\quad+\frac{1}{1-\delta}\log\log \varepsilon^{-1}+\frac{2}{1-\delta}\log \varepsilon^{-1},
\end{aligned} \tag{2.11}$$

we have

$$\begin{aligned}
G(T_\varepsilon-1) &= (2\delta\log(T_\varepsilon-1))^{1/(1-\delta)}-\Big(\frac{4\delta}{1-\delta}\log \varepsilon^{-1}\Big)^{1/(1-\delta)}\\
&\quad+O\Big((\log(T_\varepsilon-1))^{1/(1-\delta)-1}\log\log(T_\varepsilon-1)\Big)\\
&=\frac{1}{1-\delta}\Big(\frac{4\delta}{1-\delta}\log \varepsilon^{-1}\Big)^{1/(1-\delta)-1}\\
&\quad\times\Big(2\delta\log(T_\varepsilon-1)-\frac{4\delta}{1-\delta}\log \varepsilon^{-1}\Big)
\end{aligned}$$

$$+ O\Big((\log \varepsilon^{-1})^{1/(1-\delta)-1} \log\log \varepsilon^{-1} \Big)$$

$$= \text{const}(\log \varepsilon^{-1})^{1/(1-\delta)-1} (\log(T_\varepsilon - 1) - \log n_\varepsilon^*)$$

$$+ O\Big((\log \varepsilon^{-1})^{1/(1-\delta)-1} \log\log \varepsilon^{-1} \Big).$$

Since

$$\sup_{0 \le x \le 1} \hat{f}_n(x) \to \sup_{0 \le x \le 1} f(x) \quad \text{a.s. } n \to \infty, \tag{2.12}$$

for

$$n_\varepsilon = \inf \left\{ n \colon n \ge 1, \left[\sup_{0 \le x \le 1} \frac{f(x) \int K^2(t)\, dt}{n h_n} \right]^{1/2} \left[\frac{z_\alpha}{\sqrt{2\delta \log n}} + d_n \right] \le \varepsilon \right\}$$

we have

$$\frac{T_\varepsilon}{n_\varepsilon} \to 1 \quad \text{a.s.}$$

Also,

$$\left(\sup_{0 \le x \le 1} \frac{f(x)}{n h_n} \right)^{1/2} \left(\frac{z_\alpha}{\sqrt{2\delta \log n}} + d_n \right)$$

$$= \Big(2\delta \sup_{0 \le x \le 1} f(x) \Big)^{1/2} (\log n)^{1/2} n^{-(1-\delta)/2} + o\Big(n^{-(1-\delta)/2} (\log n)^{1/2} \Big).$$

So we have $n_\varepsilon / n_\varepsilon^* \to 1$, which implies

$$\frac{T_\varepsilon}{n_\varepsilon^*} \to 1 \quad \text{a.s.,} \tag{2.13}$$

and, hence,

$$\log(T_\varepsilon - 1) - \log n_\varepsilon^* \to 0 \quad \text{a.s.}$$

By assumption, we also have

$$\sup_{0 \le x \le 1} \hat{f}_n(x) \to \sup_{0 \le x \le 1} f(x) \in [m, M], \qquad m > 0, M < \infty.$$

Hence

$$(\log \varepsilon^{-1})^{1 - 1/(1-\delta)} (\log\log \varepsilon^{-1})^{-2} \Big(\sup_{0 \le x \le 1} \hat{f}_{T_\varepsilon - 1}(x) \Big)^{1/(1-\delta)} G(T_\varepsilon - 1)$$

$$\to 0 \quad \text{a.s., } \varepsilon \to 0.$$

Also

$$\left| \Big(\sup_{0 \le x \le 1} \hat{f}_{T_\varepsilon - 1}(x) \Big)^{1/(1-\delta)} - \Big(\sup_{0 \le x \le 1} f(x) \Big)^{1/(1-\delta)} \right|$$

$$\le \text{const} \cdot \sup_{0 \le x \le 1} \left| \hat{f}_{T_\varepsilon - 1}(x) - f(x) \right|$$

$$(2.14)\qquad = O\left(T_\varepsilon^{-(1-\delta)/2}(\log T_\varepsilon)^{1/2}\right)$$

$$= O\left(\left(\frac{\log \varepsilon^{-1}}{\varepsilon^2}\right)^{-(1-\delta)/2}(\log \varepsilon^{-1})^{1/2}\right)$$

$$= O\left(\varepsilon^{1-\delta}(\log \varepsilon^{-1})^{\delta/2}\right) \quad \text{a.s.},$$

by Lemma 2 and (2.6).

It follows that

$$(\log \varepsilon^{-1})^{1-1/(1-\delta)}(\log\log \varepsilon^{-1})^{-2}\left(\frac{4\delta}{1-\delta}\log \varepsilon^{-1}\right)^{1/(1-\delta)}$$

$$\times\left|\left(\sup_{0\le x\le 1} \hat{f}_{T_\varepsilon-1}(x)\right)^{1/(1-\delta)} - \left(\sup_{0\le x\le 1} f(x)\right)^{1/(1-\delta)}\right|$$

$$\to 0 \quad \text{a.s.}, \ \varepsilon \to 0.$$

Similarly, by (2.9),

$$\frac{T_\varepsilon - n_\varepsilon^*}{n_\varepsilon^*(\log\log \varepsilon^{-1})^2/\log \varepsilon^{-1}}$$

$$\ge \text{const}(\log \varepsilon^{-1})^{1-1/(1-\delta)}(\log\log \varepsilon^{-1})^{-2}\left(\sup_{0\le x\le 1} \hat{f}_{T_\varepsilon}(x)\right)^{1/(1-\delta)} G(T_\varepsilon)$$

$$+\text{const}(\log \varepsilon^{-1})^{1-1/(1-\delta)}(\log\log \varepsilon^{-1})^{-2}\left(\frac{4\delta}{1-\delta}\log \varepsilon^{-1}\right)^{1/(1-\delta)}$$

$$\times\left[\left(\sup_{0\le x\le 1} \hat{f}_{T_\varepsilon}(x)\right)^{1/(1-\delta)} - \left(\sup_{0\le x\le 1} f(x)\right)^{1/(1-\delta)}\right]$$

$$\to 0 \quad \text{a.s.}$$

So (2.8) holds. □

LEMMA 6. *Let*

$$(2.15)\qquad n_0 = n_\varepsilon^* + \varepsilon^{-2/(1-\delta)}\left(\sup_{0\le x\le 1} f(x)\int K^2(t)\,dt\right)^{1/(1-\delta)} g(\varepsilon),$$

for

$$g(\varepsilon) = \frac{(4\delta)^{1/(1-\delta)}}{2(1-\delta)^{1/(1-\delta)+1}}(\log \varepsilon^{-1})^{1/(1-\delta)-1}$$

$$(2.16)\qquad \times\left(\log\log \varepsilon^{-1} + \log\left(\frac{4\delta}{1-\delta}\sup_{0\le x\le 1} f(x)\int K^2(t)\,dt\right) + H(n_\varepsilon^*)\right),$$

with $H(n)$ *defined by* (2.7). *Then, for all* $0 < \eta < 1$, *under the assumptions of Theorem* 1,

$$\frac{T_\varepsilon - n_0}{n_0/(\log n_0)^{1+\eta}} \to 0 \quad a.s., \ \varepsilon \to 0. \tag{2.17}$$

PROOF. By (2.9),

$$\begin{aligned}\frac{T_\varepsilon - n_0}{n_0/(\log n_0)^{1+\eta}} \le{} & \mathrm{const}(\log \varepsilon^{-1})^{1+\eta-1/(1-\delta)} \\ & \times [G(T_\varepsilon - 1) - g(\varepsilon)]\Big(\sup_{0\le x\le 1} \hat{f}_{T_\varepsilon - 1}(x)\Big)^{1/(1-\delta)} \\ & + \mathrm{const}(\log \varepsilon^{-1})^{1+\eta} \\ & \times \Big[\Big(\sup_{0\le x\le 1} f(x)\Big)^{1/(1-\delta)} - \Big(\sup_{0\le x\le 1} \hat{f}_{T_\varepsilon - 1}(x)\Big)^{1/(1-\delta)}\Big].\end{aligned}$$

By (2.14), it is easy to see that

$$(\log \varepsilon^{-1})^{1+\eta}\Big[\Big(\sup_{0\le x\le 1} f(x)\Big)^{1/(1-\delta)} - \Big(\sup_{0\le x\le 1} \hat{f}_{T_\varepsilon - 1}(x)\Big)^{1/(1-\delta)}\Big] \to 0 \quad \text{a.s.}, \varepsilon \to 0.$$

By (2.11),

$$\begin{aligned}G(T_\varepsilon - 1) - g(\varepsilon) ={} & (2\delta \log(T_\varepsilon - 1))^{1/(1-\delta)} - \Big(\frac{4\delta}{1-\delta}\log \varepsilon^{-1}\Big)^{1/(1-\delta)} \\ & + H(T_\varepsilon - 1) - g(\varepsilon) \\ & + O\Big((\log \varepsilon^{-1})^{1/(1-\delta)-3}(\log\log \varepsilon^{-1})^3\Big) \\ ={} & \frac{1}{1-\delta}\Big(\frac{4\delta}{1-\delta}\log \varepsilon^{-1}\Big)^{1/(1-\delta)-1} \\ & \times \Big[2\delta \log(T_\varepsilon - 1) - \frac{4\delta}{1-\delta}\log \varepsilon^{-1} - \frac{2\delta}{1-\delta}\log\log \varepsilon^{-1} \\ & \qquad - \frac{2\delta}{1-\delta}\log\Big(\frac{4\delta}{1-\delta}\sup_{0\le x\le 1} f(x)\int K^2(t)\,dt\Big)\Big] \\ & + H(T_\varepsilon - 1) - H(n_\varepsilon^*) \\ & + O\Big((\log \varepsilon^{-1})^{1/(1-\delta)-3}(\log\log \varepsilon^{-1})^3\Big) \\ ={} & \frac{1}{1-\delta}\Big(\frac{4\delta}{1-\delta}\log \varepsilon^{-1}\Big)^{1/(1-\delta)-1}\Big(2\delta \log\Big(\frac{T_\varepsilon - 1}{n_\varepsilon^*}\Big)\Big) \\ & + H(T_\varepsilon - 1) - H(n_\varepsilon^*) \\ & + O\Big((\log \varepsilon^{-1})^{1/(1-\delta)-3}(\log\log \varepsilon^{-1})^3\Big).\end{aligned}$$

Then,

$$
\begin{aligned}
&(\log \varepsilon^{-1})^{1+\eta-1/(1-\delta)}(G(T_\varepsilon - 1) - g(\varepsilon)) \\
&\quad = \text{const}(\log \varepsilon^{-1})^{\eta} \log\left(1 + \frac{T_\varepsilon - 1 - n_\varepsilon^*}{n_\varepsilon^*}\right) \\
&\qquad + (\log \varepsilon^{-1})^{1+\eta-1/(1-\delta)}(H(T_\varepsilon - 1) - H(n_\varepsilon^*)) \\
&\qquad + O\left((\log \varepsilon^{-1})^{\eta-2}(\log\log \varepsilon^{-1})^3\right).
\end{aligned}
$$

By Lemma 5,

$$
\left|(\log \varepsilon^{-1})^{\eta} \log\left(1 + \frac{T_\varepsilon - 1 - n_\varepsilon^*}{n_\varepsilon^*}\right)\right| = o\left((\log \varepsilon^{-1})^{\eta-1}(\log\log \varepsilon^{-1})^2\right) \to 0 \quad \text{a.s.}
$$

Using Taylor expansion for every term of $H(n)$, we have

$$
(\log \varepsilon^{-1})^{1+\eta-1/(1-\delta)}(H(T_\varepsilon - 1) - H(n_\varepsilon^*)) \to 0 \quad \text{a.s.}, \varepsilon \to 0.
$$

So

$$
\frac{T_\varepsilon - n_0}{n_0/(\log n_0)^{1+\eta}} \le o(1) \quad \text{a.s.} \ \forall\, 0 < \eta < 1.
$$

By (2.9) and the same reasoning, we have

$$
\frac{T_\varepsilon - n_0}{n_0/(\log \varepsilon^{-1})^{1+\eta}} \ge o(1) \quad \text{a.s.} \ \forall\, 0 < \eta < 1,
$$

so (2.17) holds. □

LEMMA 7. *Let*

$$
\begin{aligned}
n_1 &= n_0 + \tfrac{1}{2}\varepsilon^{-2/(1-\delta)}(\log \varepsilon^{-1})^{-1} \\
&\qquad \times \left(\sup_{0\le x\le 1} f(x) \int K^2(t)\,dt\right)^{1/(1-\delta)} g(\varepsilon) \\
&= n_\varepsilon^* + \varepsilon^{-2/(1-\delta)}\left(\sup_{0\le x\le 1} f(x) \int K^2(t)\,dt\right)^{1/(1-\delta)} \\
&\qquad \times g(\varepsilon)\left(1 + \tfrac{1}{2}(\log \varepsilon^{-1})^{-1}\right),
\end{aligned}
\tag{2.18}
$$

with n_0 *and* g *as defined in* (2.15) *and* (2.16). *Then there exists some* $\eta_0 > 0$, *such that*

$$
\frac{T_\varepsilon - n_1}{n_1/(\log \varepsilon^{-1})^{2+\eta_0}} \to 0 \quad a.s., \varepsilon \to 0. \tag{2.19}
$$

PROOF. Since

$$n_0 = n_\varepsilon^* + \varepsilon^{-2/(1-\delta)}\left(\sup_{0\le x\le 1} f(x)\int K^2(t)\,dt\right)^{1/(1-\delta)} g(\varepsilon)$$
$$= \varepsilon^{-2/(1-\delta)}\left(\frac{4\delta}{1-\delta}\sup_{0\le x\le 1} f(x)\int K^2(t)\,dt\right)^{1/(1-\delta)}$$
$$\times\left[1 + \frac{g(\varepsilon)}{\left([4\delta/(1-\delta)]\log\varepsilon^{-1}\right)^{1/(1-\delta)}}\right],$$

$$\log n_0 = \frac{2}{1-\delta}\log\varepsilon^{-1} + \frac{1}{1-\delta}\log\left(\frac{4\delta}{1-\delta}\sup_{0\le x\le 1} f(x)\int K^2(t)\,dt\right)$$
$$+ \frac{1}{1-\delta}(\log\log\varepsilon^{-1}) + \frac{g(\varepsilon)}{\left([4\delta/(1-\delta)]\log\varepsilon^{-1}\right)^{1/(1-\delta)}}$$
$$+ O\left(\left(\frac{\log\log\varepsilon^{-1}}{\log\varepsilon^{-1}}\right)^2\right);$$

$$\frac{4\delta}{1-\delta}\log\varepsilon^{-1} = 2\delta\log n_0$$
$$- \frac{2\delta}{1-\delta}\log\left(\frac{4\delta}{1-\delta}\sup_{0\le x\le 1} f(x)\int K^2(t)\,dt\right)$$
$$- \frac{2\delta}{1-\delta}\log\log\varepsilon^{-1} - \frac{2\delta g(\varepsilon)}{\left([4\delta/(1-\delta)]\log\varepsilon^{-1}\right)^{1/(1-\delta)}} \tag{2.20}$$
$$+ O\left(\left(\frac{\log\log\varepsilon^{-1}}{\log\varepsilon^{-1}}\right)^2\right).$$

By (2.20),

$$\frac{T_\varepsilon - n_1}{n_1/(\log n_1)^{2+\eta_0}}$$
$$\le \mathrm{const}_1(\log\varepsilon^{-1})^{2+\eta_0-1/(1-\delta)}$$
$$\times\left[G(T_\varepsilon - 1) - g(\varepsilon) - \frac{1}{2}(\log\varepsilon^{-1})^{-1}g(\varepsilon)\right]\left(\sup_{0\le x\le 1}\hat{f}_{T_\varepsilon-1}(x)\right)^{1/(1-\delta)}$$
$$+ \mathrm{const}_2(\log\varepsilon^{-1})^{2+\eta_0}\left[\left(\sup_{0\le x\le 1} f(x)\right)^{1/(1-\delta)} - \left(\sup_{0\le x\le 1}\hat{f}_{T_\varepsilon-1}(x)\right)^{1/(1-\delta)}\right]$$

$$
\begin{aligned}
&= \mathrm{const}_1 (\log \varepsilon^{-1})^{2+\eta_0 - 1/(1-\delta)} \Big(\sup_{0 \le x \le 1} \hat{f}_{T_\varepsilon - 1}(x) \Big)^{1/(1-\delta)} \\
&\quad \times \Bigg[(2\delta \log(T_\varepsilon - 1))^{1/(1-\delta)} - \left(\frac{4\delta}{1-\delta} \log \varepsilon^{-1} \right)^{1/(1-\delta)} + H(T_\varepsilon - 1) - g(\varepsilon) \\
&\qquad - \frac{1}{2} (\log \varepsilon^{-1})^{-1} g(\varepsilon) + O\Big((\log \varepsilon^{-1})^{1/(1-\delta)-3} (\log\log \varepsilon^{-1})^3 \Big) \Bigg] + o(1) \\
&= \mathrm{const}_1 (\log \varepsilon^{-1})^{2+\eta_0 - 1/(1-\delta)} \\
&\quad \times \Bigg[\frac{1}{1-\delta} \left(\frac{4\delta}{1-\delta} \log \varepsilon^{-1} \right)^{1/(1-\delta)-1} \\
&\qquad \times \Bigg\{ 2\delta \log(T_\varepsilon - 1) - 2\delta \log n_0 + \frac{2\delta}{1-\delta} \log\left(\frac{4\delta}{1-\delta} \sup_{0 \le x \le 1} f(x) \right) \\
&\qquad\quad + \frac{2\delta}{1-\delta} \log\log \varepsilon^{-1} + \frac{2\delta g(\varepsilon)}{([4\delta/(1-\delta)] \log \varepsilon^{-1})^{1/(1-\delta)}} \\
&\qquad\quad + O\left(\left(\frac{\log\log \varepsilon^{-1}}{\log \varepsilon^{-1}} \right)^2 \right) \Bigg\} \\
&\qquad + H(T_\varepsilon - 1) - g(\varepsilon) - \frac{1}{2} (\log \varepsilon^{-1})^{-1} g(\varepsilon) \Bigg] + o(1) \\
&= \mathrm{const}_1 (\log \varepsilon^{-1})^{2+\eta_0 + 1/(1-\delta)} \\
&\quad \times \Bigg[\frac{1}{1-\delta} \left(\frac{4\delta}{1-\delta} \log \varepsilon^{-1} \right)^{1/(1-\delta)-1} 2\delta \log\left(\frac{T_\varepsilon - 1}{n_0} \right) \\
&\qquad + H(T_\varepsilon - 1) - H(n_\varepsilon^*) \Bigg] + o(1) \\
&= \mathrm{const}_1 (\log \varepsilon^{-1})^{1+\eta_0} \log\left(\frac{T_\varepsilon - 1}{n_0} \right) \\
&\quad + \mathrm{const}_2 (\log \varepsilon^{-1})^{2+\eta_0 - 1/(1-\delta)} (H(T_\varepsilon - 1) - H(n_\varepsilon^*)) + o(1).
\end{aligned}
$$

We have

$$
\begin{aligned}
&\left| (\log \varepsilon^{-1})^{2+\eta_0 - 1/(1-\delta)} (H(T_\varepsilon - 1) - H(n_\varepsilon^*)) \right| \\
&\quad \le |a_1| (\log \varepsilon^{-1})^{2+\eta_0 - 1/(1-\delta)}
\end{aligned}
$$

$$
\begin{aligned}
&\times\Big|(\log(T_\varepsilon-1))^{1/(1-\delta)-1}\log\log(T_\varepsilon-1)\\
&\qquad -(\log(n_\varepsilon^*))^{1/(1-\delta)-1}\log\log n_\varepsilon^*\Big|\\
&+|a_2|(\log\varepsilon^{-1})^{2+\eta_0-1/(1-\delta)}\Big|(\log(T_\varepsilon-1))^{1/(1-\delta)-1}-(\log n_\varepsilon^*)^{1/(1-\delta)-1}\Big|\\
&+|a_3|(\log\varepsilon^{-1})^{2+\eta_0-1/(1-\delta)}\\
&\times\Big|(\log(T_\varepsilon-1))^{1/(1-\delta)-2}(\log\log(T_\varepsilon-1))^2\\
&\qquad -(\log n_\varepsilon^*)^{1/(1-\delta)-2}(\log\log n_\varepsilon^*)^2\Big|\\
&+|a_4|(\log\varepsilon^{-1})^{2+\eta_0-1/(1-\delta)}\\
&\times\Big|(\log(T_\varepsilon-1))^{1/(1-\delta)-2}\log\log(T_\varepsilon-1)\\
&\qquad -(\log n_\varepsilon^*)^{1/(1-\delta)-2}(\log\log n_\varepsilon^*)\Big|\\
&+|a_5|(\log\varepsilon^{-1})^{2+\eta_0-1/(1-\delta)}\Big|(\log(T_\varepsilon-1))^{1/(1-\delta)-2}-(\log n_\varepsilon^*)^{1/(1-\delta)-2}\Big|\\
&\to 0\quad \text{a.s., } \varepsilon\to 0.
\end{aligned}
$$

By Lemma 6,

$$
\begin{aligned}
\left|(\log\varepsilon^{-1})^{1+\eta_0}\log\left(\frac{T_\varepsilon-1}{n_0}\right)\right| &\le O\left((\log\varepsilon^{-1})^{1+\eta_0}\left|\frac{T_\varepsilon-n_0}{n_0}\right|\right)\\
&= O\left(\frac{(\log\varepsilon^{-1})^{1+\eta_0}}{(\log\varepsilon^{-1})^{1+\eta}}\right)\\
&= O\left((\log\varepsilon^{-1})^{\eta_0-\eta}\right)\\
&\to 0\quad\text{when } 1>\eta>\eta_0.
\end{aligned}
$$

So

$$
\frac{T_\varepsilon-n_1}{n_1/(\log n_1)^{2+\eta_0}}\le o(1).
$$

By using (2.9) and similar reasoning, we have

$$
\frac{T_\varepsilon-n_1}{n_1/(\log n_1)^{2+\eta_0}}\ge o(1).
$$

Hence, (2.19) holds. □

3. Proof of the theorem.

Proof of Theorem 1. By the result of Bickel and Rosenblatt (1973), under the assumptions of Theorem 1, for

$$
X_n=(2\delta\log n)^{1/2}\left\{\left(\int K^2(t)\,dt\right)^{-1/2}M_n-d_n\right\},
$$

there exists a r.v. X such that

$$X_n \to_d X,$$

where

$$P(X < z) = \exp(-2\exp(-z)).$$

This implies

$$X_{[n_1(\varepsilon)]} \to_d X,$$

with $n_1(\varepsilon) = n_1$ as defined in Lemma 7.

To prove the theorem, it suffices to show that

$$X_{T_\varepsilon} \to_d X,$$

so it suffices to show

$$X_{T_\varepsilon} - X_{[n_1(\varepsilon)]} \to_P 0.$$

Now, for all $b > 0$,

$$\begin{aligned} &P\{|X_{T_\varepsilon} - X_{[n_1(\varepsilon)]}| > b\} \\ &\quad \le P\left\{|X_{T_\varepsilon} - X_{[n_1(\varepsilon)]}| > b, |T_\varepsilon - [n_1(\varepsilon)]| \le \frac{n_1(\varepsilon)}{(\log(n_1(\varepsilon)))^{2+\eta_0}}\right\} \\ &\qquad + P\left\{|T_\varepsilon - [n_1(\varepsilon)]| > \frac{n_1(\varepsilon)}{(\log n_1(\varepsilon))^{2+\eta_0}}\right\}. \end{aligned}$$

Define $n_2(\varepsilon) = n_1(\varepsilon) - n_1(\varepsilon)/(\log(n_1(\varepsilon)))^{2+\eta_0}$. By Corollary 1,

$$\begin{aligned} &P\left\{|X_{T_\varepsilon} - X_{[n_1(\varepsilon)]}| > b, |T_\varepsilon - [n_1(\varepsilon)]| \le \frac{n_1(\varepsilon)}{(\log n_1(\varepsilon))^{2+\eta_0}}\right\} \\ &\quad \le P\left\{\max_{0 < k \le [n_1(\varepsilon)]/(\log[n_1(\varepsilon)])^{2+\eta_0}} |X_{[n_1(\varepsilon)]+k} - X_{[n_1(\varepsilon)]}| > b\right\} \\ &\qquad + P\left\{\max_{0 \le k \le [n_2(\varepsilon)]/(\log[n_2(\varepsilon)])^{2+\eta_0}} |X_{[n_2(\varepsilon)]+k} - X_{[n_2(\varepsilon)]}| > \frac{b}{2}\right\} \\ &\quad \to 0 \quad \text{as } \varepsilon \to 0. \end{aligned}$$

By Lemma 7,

$$\begin{aligned} &P\left\{|T_\varepsilon - [n_1(\varepsilon)]| > \frac{n_1(\varepsilon)}{(\log n_1(\varepsilon))^{2+\eta_0}}\right\} \\ &\quad \le P\left\{\left|\frac{T_\varepsilon - n_1}{n_1/(\log n_1)^{2+\eta_0}}\right| + \frac{1}{n_1/(\log n_1)^{2+\eta_0}} > 1\right\} \\ &\quad \to 0 \quad \text{as } \varepsilon \to 0. \end{aligned}$$

This proves (1.5). Relation (1.6) follows immediately. □

TABLE 1
Coverage frequencies and average sample sizes for $N(0, 1)$ *data on* $[0, 1]$ *with* $\delta = 0.21$*

α	ε	**Coverage frequency**	**Average sample size**
0.1	0.3	0.86	142.30 (4.07)
0.1	0.2	0.90	401.62 (6.49)
0.2	0.3	0.72	86.96 (2.82)
0.2	0.2	0.82	257.34 (5.43)

*Coverage frequency is frequency (in each case out of 50 trials) of the event that all values were covered. Numbers in parentheses are standard errors.

4. Simulation results. Simulations were carried out for the stopping rule that results from replacing the supremum in (1.4) with the maximum over a uniform grid of points in the interval under consideration. Two situations were considered: (1) standard normal on the interval $[0, 1]$, with a grid of 99 points; (2) exponential with scale 1 on the interval $[0.5, 1.5]$, with a grid of 100 points. Two choices of α were used: 0.1 and 0.2. There were also two choices of ε: 0.2 and 0.3; $\delta = 0.21$ was chosen because, according to the asymptotics, the closer δ is to $\frac{1}{5}$ the smaller the sample size will be. The uniform kernel with support $[-0.5, 0.5]$ was used throughout. All simulated observations X_i, including those outside the interval of interest, were used by the procedure, although of course the density estimates were computed only for points inside the interval [see assumption (A.5)]. Fifty simulations were conducted for each of the eight combinations of distribution, α and ε. The results are shown in Tables 1 and 2, where the coverage frequency [i.e., frequency with which all density values on the grid were covered by the band in (1.5)] and the average sample size are given.

Overall, the coverage frequencies are reasonably good. As one would expect, they are better at the smaller value of ε. When $\varepsilon = 0.3$, the observed frequencies are below the desired coverage probabilities, although not by an alarming amount. When $\varepsilon = 0.2$, the coverage frequencies in the normal case are either exactly at or slightly above the desired values, while in the exponential case they are noticeably above the desired values. We do not know why the procedure appears to be conservative for exponential data, but at least the discrepancy means increased rather than decreased confidence.

TABLE 2
Coverage frequencies and average sample sizes for exponential data on $[0.5, 1.5]$ *with* $\delta = 0.21$

α	ε	**Coverage frequency**	**Average sample size**
0.1	0.3	0.84	209.38 (3.61)
0.1	0.2	0.96	589.34 (9.75)
0.2	0.3	0.78	128.76 (2.87)
0.2	0.2	0.90	386.40 (7.27)

Of course one must keep in mind that the observed frequencies themselves contain random variation.

Acknowledgments. We would like to thank the referees and an Associate Editor for their careful reading of the original manuscript and for suggestions that improved the clarity of the paper.

REFERENCES

ANSCOMBE, F. (1952). Large sample theory of sequential estimation. *Proc. Cambridge Philos. Soc.* **48** 600–607.

BICKEL, P. J. and ROSENBLATT, M. (1973). On some global measures of the deviations of the density function estimates. *Ann. Statist.* **1** 1071–1095.

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

CHOW, Y. S., HSIUNG, C. A. and YU, K. F. (1980). Limit theorems for a positively drifting process and its related first passage times. *Bull. Inst. Math. Acad. Sinica* **8** 141–172.

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

CSÖRGŐ, M. and RÉVÉSZ, P. (1979). How big are the increments of a Wiener process? *Ann. Probab.* **7** 731–737.

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.

IZENMAN, A. (1991). Recent developments in nonparametric density estimation. *J. Amer. Statist. Assoc.* **86** 205–224.

KOMLÓS, J., MAJOR, P. and TUSNÁDY, G. (1975). An approximation of partial sums of independent RV's and the sample DF. I. *Z. Wahrsch. Verw. Gebiete* **32** 111–131.

MARTINSEK, A. T. (1983). Second order approximation to the risk of a sequential procedure. *Ann. Statist.* **11** 827–836.

MARTINSEK, A. T. (1993). Fixed width confidence bands for density functions. *Sequential Anal.* **12** 169–177.

PARZEN, E. (1962). On estimation of a probability density function and mode. *Ann. Math. Statist.* **33** 1065–1076.

ROBBINS, H. (1959). Sequential estimation of the mean of a normal population. In *Probability and Statistics. The Harald Cramér Volume* (U. Grenander, ed.) 235–245. Almquist and Wiksell, Stockholm.

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

STUTE, W. (1982). The oscillation behavior of empirical processes. *Ann. Probab.* **10** 86–107.

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

WOODROOFE, M. (1982). *Nonlinear Renewal Theory in Sequential Analysis*. SIAM, Philadelphia.

XU, Y. and MARTINSEK, A. T. (1994). Sequential confidence bands for densities. Technical report, Dept. Statistics, Univ. Illinois.

ZHENG, Z. (1988). Strong uniform consistency for density estimator from randomly censored data. *Chinese Ann. Math. Ser. B* **9** 167–175.

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