Ann. Inst. Statist. Math.
Vol. 49, No. 1, 57–78 (1997)

BOUNDING THE L_1 DISTANCE IN NONPARAMETRIC DENSITY ESTIMATION

SUBRATA KUNDU[1] AND ADAM T. MARTINSEK[2]

[1] *Department of Statistics and Applied Probability, University of California, Santa Barbara, CA 93106-3110, U.S.A.*
[2] *Department of Statistics, University of Illinois, 101 Illini Hall, 725 South Wright Street, Champaign, IL 61820, U.S.A.*

(Received October 23, 1995; revised May 13, 1996)

Abstract. Let $X_1, X_2, \ldots, X_n$ be i.i.d. random variables with common unknown density function f. We are interested in estimating the unknown density f with bounded Mean Integrated Absolute Error (MIAE). Devroye and Győrfi (1985, *Nonparametric Density Estimation: The L_1 View*, Wiley, New York) obtained asymptotic bounds for the MIAE in estimating f by a kernel estimate $\hat{f}_n$. Using these bounds one can identify an appropriate sample size such that an asymptotic upper bound for the MIAE is smaller than some pre-assigned quantity $w > 0$. But this sample size depends on the unknown density f. Hence there is no fixed sample size that can be used to solve the problem of bounding the MIAE. In this work we propose stopping rules and two-stage procedures for bounding the L_1 distance. We show that these procedures are asymptotically optimal in a certain sense as $w \to 0$, i.e., as one requires increasingly better fit.

Key words and phrases: Density estimation, mean integrated absolute error, stopping rule, sequential estimation.

1. Introduction

Let $X_1, X_2, X_3, \ldots$ be independent random variables with common unknown density f on the real line. Consider a random sample $X_1, X_2, X_3, \ldots, X_n$ of size n. A kernel estimate $\hat{f}_n$ of f is given by

$$\hat{f}_n(x) = h_n^{-1} \int K((x-y)/h_n) dF_n(y), \tag{1.1}$$

where F_n is the sample distribution function, K is the kernel function and h_n is the bandwidth. There are several measures of the global performance of the estimate $\hat{f}_n$. For $0 < p < \infty$ the L_p distance between $\hat{f}_n$ and f is defined by

$$\|\hat{f}_n - f\|_p = \left(E \int |\hat{f}_n(x) - f(x)|^p dx\right)^{1/p} \tag{1.2}$$

The most widely used L_p distance is the Mean Integrated Squared Error (MISE) or the L_2 distance. Devroye and Győrfi (1985) argued that the L_1 distance (Mean Integrated Absolute Error or MIAE) EJ_n is the most natural choice, where $J_n = \int |\hat{f}_n - f|$. They showed that it is invariant under monotone transformations and it is always well defined. L_1 distance also has the advantage of corresponding more closely than L_2 distance to the visual comparison of densities.

The problem of sequential density estimation was first considered by Yamato (1971). Wegman and Davies (1975) developed some sequential procedures which satisfy certain error control. Carroll (1976) considered the problem of estimating the unknown density at a particular point which may be known or unknown. He proposed stopping rules to construct a fixed width confidence interval for the value of the unknown density. Stute (1983) considered similar problems. Isogai (1987) investigated the problem of sequential estimation of f at a given point x_0 in p—dimensional Euclidean space. He considered a class of recursive kernel estimators $f_n(x)$ and proposed a class of stopping rules based on the idea of fixed width interval estimation. In his 1988 paper he investigated the asymptotic behavior of the moments of these stopping rules when the length of the interval goes to zero. Koronacki and Wertz (1988) obtained some results for sequential recursive density estimators. Efroimovich (1989) derived an asymptotically exact minimax lower bound for the risk in a sequential nonparametric estimation plan.

Martinsek (1992) studied the problem of estimating the density such that the most commonly used global measure of error, Mean Integrated Squared Error (MISE) is smaller than some pre-assigned positive quantity w. Using the well known asymptotic expansion (see Rosenblatt (1956, 1971), Nadaraya (1974), and Prakasa Rao ((1983), Theorem 2.1.7))

$$(1.3)\qquad E\int_{-\infty}^{\infty}(\hat{f}_n(x)-f(x))^2dx$$
$$= (nh_n)^{-1}\int_{-\infty}^{\infty}K^2(t)dt + (\beta^2h_n^4/4)\int_{-\infty}^{\infty}(f''(x))^2dx$$
$$+\, o((nh_n)^{-1}+h_n^4),$$

where

$$(1.4)\qquad \beta = \int_{-\infty}^{\infty}x^2K(x)dx,$$

Martinsek (1992) identified an appropriate sample size. Since the appropriate sample size depends on a functional of the unknown density, the problem cannot be solved using a fixed sample size. Martinsek (1992) considered some fully sequential and two stage procedures to solve the problem. He showed that those procedures are asymptotically efficient.

Here we are concerned with the problem of estimating the unknown density f such that the L_1 distance between the true density and the estimated density is bounded by some pre-assigned positive quantity w. The organization of this paper is as follows. Section 2 proposes sequential and two-stage procedures for bounding the L_1 error and contains the statements and proofs of the main results. Section 3 deals with data-driven, asymptotically optimal bandwidths. Finally, some simulation results for the two-stage procedure are presented in the appendix.

2. Bounding the L_1 distance

Recently much progress has been made on the "L_1 view" of nonparametric density estimation. Among others this was largely due to the works of Devroye, Győrfi, Hall, and Wand. In their monograph Devroye and Győrfi (1985) developed a smooth L_1 theory. They did not obtain an asymptotic expansion for the L_1 error, instead they worked with upper and lower bounds for MIAE.

We want to estimate f with MIAE no larger than w, i.e., with sufficiently good global fit. Unlike the MISE case there is no explicit asymptotic expansion for the MIAE (for precise but somewhat implicit expansions, see Hall and Wand (1988), Devroye and Wand (1993) and Wand and Devroye (1993)). But there are some upper bounds for the MIAE. Under the assumptions that K is a symmetric, bounded probability density function with compact support, f is twice differentiable with bounded continuous second derivative, $\int \sqrt{f} < \infty$, $\int |f''| < \infty$, and $h_n \to 0$, $nh_n \to \infty$, the kernel estimate satisfies

$$EJ_n \leq \left(\alpha \int \sqrt{f}\right) / \sqrt{nh_n} + (\beta/2)h_n^2 \int |f''| + o(h_n^2 + (nh_n)^{-1/2}), \tag{2.1}$$

where

$$\alpha = \sqrt{\int K^2}, \qquad \beta = \int x^2 K(x)dx, \tag{2.2}$$

see Devroye and Győrfi (1985). Put $\theta_1 = \int \sqrt{f}$, $\theta_2 = \int |f''|$ and let H_n be the R.H.S of the expression (2.1), excluding the higher order term. Since the exact asymptotic behavior of EJ_n is unknown, we deal with the bound for EJ_n. From (2.1) it is clear that the best asymptotic rate for h_n is $n^{-1/5}$. Let $h_n = cn^{-1/5}$ where c is a positive constant. Then we get

$$H_n = \{\alpha\theta_1/\sqrt{c} + (\beta/2)\theta_2 c^2\} n^{-2/5}. \tag{2.3}$$

The best choice for c is

$$c^* = [(\alpha\theta_1)/(2\beta\theta_2)]^{2/5}. \tag{2.4}$$

For this choice of c we get

$$H_n^* = \tau^* A(K) B(f) n^{-2/5}, \tag{2.5}$$

where $\tau^* = 5/2^{(8/5)}$,

$$A(K) = \left(\int K^2\right)^{2/5} \left(\int x^2 K\right)^{1/5} \tag{2.6}$$

and

$$B(f) = \left(\left(\int \sqrt{f} \right)^4 \int |f''|/2 \right)^{1/5}. \tag{2.7}$$

We want to estimate the unknown density f with both IAE and MIAE no larger than w ($w > 0$), i.e., with sufficiently good L_1 fit. We first consider the case $h_n = cn^{-1/5}$ for some fixed $c > 0$. From (2.3) it suffices (at least asymptotically) to achieve

$$(\alpha\theta_1/\sqrt{c} + (\beta/2)c^2\theta_2)n^{-2/5} \leq w. \tag{2.8}$$

This yields the sample size n_w^*, where

$$n_w^* = \text{smallest integer} \geq ([\alpha\theta_1/\sqrt{c} + (\beta/2)c^2\theta_2]w^{-1})^{5/2}. \tag{2.9}$$

But this choice of n_w^* involves the unknown quantities θ_1 and θ_2 and hence n_w^* cannot be used in practice.

2.1 *Sequential procedure*

One way of circumventing this problem is to replace the unknown parameters θ_1 and θ_2 in the expression for n_w^* with estimates, which suggests the following sequential procedure. Define the stopping rule

$$T_w = \text{first } n \geq 1 \text{ such that } n^{-2/5}[\alpha\hat{\theta}_{1n}/\sqrt{c} + (\beta/2)c^2\hat{\theta}_{2n} + n^{-\xi}] \leq w, \tag{2.10}$$

where $\hat{\theta}_{1n}$ and $\hat{\theta}_{2n}$ are estimators of θ_1 and θ_2 respectively, based on $X_1, X_2, \ldots, X_n$, and $0 < \xi < 1/16$ is a positive number that will be determined later.

The term $n^{-\xi}$ is added to the left side of (2.10) for technical reasons: see the proof of Theorem 2.1. Estimate the unknown density f by

$$\hat{f}_{T_w}(x) = (T_w h_{T_w})^{-1} \sum_1^{T_w} K((x - X_i)/h_{T_w}), \tag{2.11}$$

where $h_{T_w} = cT_w^{-1/5}$. We hope that this procedure will yield the desired results, i.e.,

$$\text{MIAE} = E\left(\int_{-\infty}^{\infty} |\hat{f}_{T_w}(x) - f(x)| dx \right) \leq w \tag{2.12}$$

and

$$T_w \approx n_w^*. \tag{2.13}$$

Choice of $\hat{\theta}_{1n}$ and $\hat{\theta}_{2n}$: In order to construct the estimates of θ_1 and θ_2 we need to assume the following conditions on f and K.

CONDITIONS.

$$\|f''\|_{BL} = \sup_x |f''(x)| + \sup_{x \neq y}(|f''(x) - f''(y)|/|x - y|) < \infty. \tag{2.14}$$

(2.15) Both K and K'' have finite total variation.

Remark. It is natural to wonder about the behavior of the sequential procedure without the smoothness condition (2.14). Because of the term $n^{-\xi}$ in (2.10), the random sample size $T_w \to \infty$ a.s. as $w \to 0$. It follows that the procedure will be strongly consistent (i.e., $J_{T_w} \to 0$ a.s.) provided that $J_n \to 0$ a.s. as $n \to \infty$. In other words, the sequential estimator is strongly consistent whenever the usual nonsequential estimator is. But the nonsequential estimator is strongly consistent for *all* f: see Devroye (1983). The sequential procedure therefore inherits strong consistency from the nonsequential version, without any smoothness conditions on f. By bounded convergence one immediately has the same result for the MIAE as well. This remark applies also to the two stage and data-driven methods to be discussed later in the paper.

We propose the following estimates of θ_1 and θ_2. Let

$$\hat{\theta}_{1n} = \hat{\theta}_{1n}(\tilde{h}_n) = \int_{-\tilde{d}_n}^{\tilde{d}_n} \sqrt{\tilde{f}_n}, \tag{2.16}$$

where $\tilde{d}_n \to \infty$ as $n \to \infty$ and $\tilde{f}_n$ is a kernel estimate of f based on $X_1, X_2, \ldots, X_n$ with kernel K and bandwidth $\tilde{h}_n$. The quantities $\tilde{d}_n$ and $\tilde{h}_n$ will be determined later. Similarly we define

$$\hat{\theta}_{2n} = \hat{\theta}_{2n}(\check{h}_n) = \int_{-\check{d}_n}^{\check{d}_n} |\check{f}_n''|, \tag{2.17}$$

where $\check{f}_n$ is a kernel estimate of f with kernel K and bandwidth $\check{h}_n$, and $\check{d}_n$ and $\check{h}_n$ will be determined later.

The following theorem gives some asymptotic optimality results for the sequential procedure.

THEOREM 2.1. *Assume* (2.14) *and* (2.15). *Put*

$$\tilde{d}_n = \check{d}_n = n^{1/8-\delta},$$
$$\tilde{h}_n = n^{-1/4}(\log\log n)^{1/4},$$

and

$$\check{h}_n = n^{-1/8}(\log\log n)^{1/8},$$

where $0 < \delta < 1/8$. *Then as* $w \to 0$

$$T_w/n_w^* \to 1 \quad a.s. \tag{2.18}$$

$$ET_w/n_w^* \to 1 \tag{2.19}$$

$$\limsup_{w \to 0} J_{T_w}/w \le 1 \quad a.s. \tag{2.20}$$

$$\limsup_{w \to 0} EJ_{T_w}/w \le 1. \tag{2.21}$$

Remark. Equations (2.18) and (2.19) state that the random sample size T_w is, for small w, almost surely equivalent to n_w^* and similarly that the expected sample size is equivalent to n_w^*. Equations (2.20) and (2.21) say that the IAE and MIAE are asymptotically bounded above by w.

Note that this procedure requires a lot of computation, because the estimates $\hat{\theta}_{1n}$ and $\hat{\theta}_{2n}$ must be recomputed after each new observation. Another possible method is a two stage procedure which reduces computational cost and addresses the main problem.

2.2 *Two stage procedure*

We propose the following two stage procedure. First draw a sample of size $n_0(w)$, where $n_0(w) \to \infty$ and $w^{5/2} n_0(w) \to 0$ as $w \to 0$. One possible choice for $n_0(w)$ is $n_0(w) = w^{-\gamma}$ for some $0 < \gamma < 5/2$. This is done to ensure that we have enough observations at the first stage but not more than we need. Based on the first sample estimate θ_1 and θ_2. Next we draw a second sample of size $n_1(w)$, where $n_1(w)$ is the smallest integer greater than or equal to

$$[\{[\alpha\hat{\theta}_{1n_0}/\sqrt{c} + (\beta/2)c^2\hat{\theta}_{2n_0} + n_0^{-\xi}]/w\}^{5/2} - n_0(w)]^+, \tag{2.22}$$

where ξ is as in (2.10).

Estimate f by

$$\hat{f}_{n_2(w)}(x) = [h_{n_2(w)} n_2(w)]^{-1} \sum_1^{n_2(w)} K((x - X_i)/h_{n_2(w)}), \tag{2.23}$$

where $h_n = cn^{-1/5}$ and $n_2(w) = n_0(w) + n_1(w)$. We hope that this procedure will yield the desired results, i.e., results analogous to Theorem 2.1. We get the following theorem which summarizes the performance of the two stage procedure.

THEOREM 2.2. *Under the conditions of Theorem* 2.1, *as* $w \to 0$,

$$n_2(w)/n_w^* \to 1 \quad a.s. \tag{2.24}$$

$$En_2(w)/n_w^* \to 1 \tag{2.25}$$

$$\limsup_{w \to 0} J_{n_2(w)}/w \le 1 \quad a.s. \tag{2.26}$$

$$\limsup_{w \to 0} E[J_{n_2(w)}]/w \le 1. \tag{2.27}$$

Remark. From Theorem 2.2 we can see that the two stage procedure performs satisfactorily. In other words, (2.24) and (2.25) state that $n_2(w)$ is asymptotically equivalent to the optimal sample size n_w^*. Equations (2.26) and (2.27) say that the IAE and MIAE are asymptotically bounded above by w.

2.3 *Proofs of Theorems* 2.1 *and* 2.2

We assume (2.14) and (2.15), as well as the choices of $\tilde{h}_n$, $\check{h}_n$, $\tilde{d}_n$ and $\check{d}_n$ given in Theorem 2.1, throughout.

LEMMA 2.1.

$$\sup_x |\sqrt{\tilde{f}_n(x)} - \sqrt{f(x)}| = O(n^{-1/8}(\log\log n)^{1/8}) \qquad a.s. \tag{2.28}$$

PROOF. From Theorem 2.1.10 of Prakasa Rao (1983), it follows that

$$\sup_x |\tilde{f}_n(x) - f(x)| = O(n^{-1/4}(\log\log n)^{1/4}) \qquad \text{a.s.} \tag{2.29}$$

and since

$$\begin{aligned} |\sqrt{\tilde{f}_n(x)} - \sqrt{f(x)}| &= |\tilde{f}_n(x) - f(x)|/(\sqrt{\tilde{f}_n(x)} + \sqrt{f(x)}) \\ &\le |\tilde{f}_n(x) - f(x)|/\sqrt{|\tilde{f}_n(x) - f(x)|} \\ &= |\tilde{f}_n(x) - f(x)|^{1/2}, \end{aligned} \tag{2.30}$$

(2.28) follows from (2.29) and (2.30).

LEMMA 2.2.

$$\hat{\theta}_{1n} \to \theta_1 \qquad a.s.$$

PROOF. By Lemma 2.1,

$$\begin{aligned} |\hat{\theta}_{1n} - \theta_1| &\le \left| \int_{-\tilde{d}_n}^{\tilde{d}_n} \sqrt{\tilde{f}_n} - \int_{-\tilde{d}_n}^{\tilde{d}_n} \sqrt{f} \right| + \left| \int_{-\tilde{d}_n}^{\tilde{d}_n} \sqrt{f} - \int \sqrt{f} \right| \\ &\le \int_{-\tilde{d}_n}^{\tilde{d}_n} \left| \sqrt{\tilde{f}_n} - \sqrt{f} \right| + \left| \int_{-\tilde{d}_n}^{\tilde{d}_n} \sqrt{f} - \int \sqrt{f} \right| \\ &= O(\tilde{d}_n n^{-1/8}(\log\log n)^{1/8}) + \left| \int_{-\tilde{d}_n}^{\tilde{d}_n} \sqrt{f} - \int \sqrt{f} \right|. \end{aligned} \tag{2.31}$$

The first term in this equation goes to zero a.s and since $\tilde{d}_n \to \infty$, the second term also goes to zero.

LEMMA 2.3.

$$\sup_x |\check{f}''_n(x) - f''(x)| = O(n^{-1/8}(\log\log n)^{1/8}) \qquad a.s. \tag{2.32}$$

PROOF. By Silverman (1978)

$$\sup_x |\check{f}_n''(x) - E\check{f}_n''(x)| = O(n^{-1/2}\check{h}_n^{-5/2}[\log(\check{h}_n^{-1})]^{1/2}) \quad \text{a.s.} \tag{2.33}$$

Integrating by parts it can be shown that

$$E\check{f}_n''(x) = (K_h * f'')(x) = h_n^{-1}\int K((x-y)/h_n)f''(y)dy.$$

From (2.14) we have $\|f''\|_{BL} < \infty$. Therefore,

$$\begin{aligned} \sup_x |E\check{f}_n''(x) - f''(x)| & \\ &= \sup_x \left| \int_{\infty}^{\infty} K(u)[f''(x - \check{h}_n u) - f''(x)]du \right| \\ &\leq \|f''\|_{BL}\check{h}_n \int_{\infty}^{\infty} |u|K(u)du \\ &= O(\check{h}_n). \end{aligned} \tag{2.34}$$

(2.32) now follows from (2.33) and (2.34).

Remark. Note that here we are not choosing the optimal $\check{h}_n$. That is because we shall need this choice of $\check{h}_n$ for another result which does not hold for the optimal choice of $\check{h}_n$.

LEMMA 2.4.

$$\hat{\theta}_{2n} \to \theta_2 \qquad a.s. \tag{2.35}$$

PROOF. Lemma 2.4 follows immediately from Lemma 2.3 and the fact that $\check{d}_n = n^{1/8-\delta}$.

LEMMA 2.5.

$$E\left[\left(\sup_n \left| \int_{-\check{d}_n}^{\check{d}_n} \sqrt{\tilde{f}_n} - \int_{-\check{d}_n}^{\check{d}_n} \sqrt{f} \right|\right)^p\right] < \infty \qquad \forall p > 0. \tag{2.36}$$

PROOF. From Theorem 2.1.11 of Prakasa Rao (1983) we get

$$E(\exp(\beta M^2)) < \infty \tag{2.37}$$

for all $\beta > 0$, where

$$M = \sup_n \left[\sup_x |\tilde{f}_n(x) - E\tilde{f}_n(x)|\tilde{h}_n(n/\log\log n)^{1/2}\right]. \tag{2.38}$$

Again, as in (2.34),

$$\sup_x |E[\tilde{f}_n(x) - f(x)]| \le \|f\|_{BL}\tilde{h}_n \int |u|K(u)du. \tag{2.39}$$

Therefore,

$$\left[\tilde{h}_n^{-1} \sup_n \sup_x |E[\tilde{f}_n(x)] - f(x)|\right]^p < \infty \tag{2.40}$$

and from (2.38) we have $EM^p < \infty$ for all $p > 0$, i.e.,

$$E\left[\sup_n \sup_x |\tilde{f}_n(x) - E\tilde{f}_n(x)|\tilde{h}_n(n/\log\log n)^{1/2}\right]^p < \infty. \tag{2.41}$$

Substituting $\tilde{h}_n$ and combining (2.40) and (2.41) we get

$$E\left[\sup_n \sup_x |\tilde{f}_n(x) - f(x)|(n/\log\log n)^{1/4}\right]^p < \infty. \tag{2.42}$$

From (2.30) we get

$$\left|\sqrt{\tilde{f}_n(x)} - \sqrt{f(x)}\right| \le |\tilde{f}_n(x) - f(x)|^{1/2}.$$

It follows that

$$E\left[\sup_n \sup_x \left|\sqrt{\tilde{f}_n(x)} - \sqrt{f(x)}\right| (n/\log\log n)^{1/8}\right]^p < \infty. \tag{2.43}$$

Now,

$$\tilde{d}_n(n/\log\log n)^{\ \ 1/8} = O(n^{\ \ \delta}(\log\log n)^{1/8}).$$

Therefore,

$$E\left[\left(\sup_n \left|\int_{-\tilde{d}_n}^{\tilde{d}_n} \sqrt{\tilde{f}_n} - \int_{-\tilde{d}_n}^{\tilde{d}_n} \sqrt{f}\right| n^{\delta}(\log\log n)^{-1/8}\right)^p\right] < \infty \qquad \forall p > 0, \tag{2.44}$$

which implies (2.36).

LEMMA 2.6.

$$E\left[\left(\sup_n \left|\int_{-\check{d}_n}^{\check{d}_n} (|\check{f}_n''(x)| - |f''(x)|)dx\right|\right)^p\right] < \infty \qquad \forall p > 0. \tag{2.45}$$

PROOF. Let

$$W = \sup_n \left[\sup_x |\check{f}_n''(x) - E\check{f}_n''(x)|(n/\log\log n)^{1/2}\check{h}_n^3\right]. \tag{2.46}$$

Now,

$$
\begin{aligned}
(2.47)\qquad & \sup_x |\check{f}_n''(x) - E\check{f}_n''(x)| \\
&= \sup_x \left| (n\check{h}_n^3)^{-1} \sum_1^n K''((x - X_i)/\check{h}_n) \right. \\
&\qquad \left. - \int \check{h}_n^{-3} K''((x-y)/\check{h}_n) dF(y) \right| \\
&= \sup_x \check{h}_n^{-3} \left| \int K''((x-y)/\check{h}_n)(dF_n(y) - dF(y)) \right| \\
&= \sup_x \check{h}_n^{-3} \left| \int [F_n(y) - F(y)] dK''((x-y)/\check{h}_n) \right| \\
&\le \check{h}_n^{-3} \sup_x |F_n(y) - F(y)| \mu,
\end{aligned}
$$

where $\mu < \infty$ is the total variation of K''. Following Prakasa Rao ((1983), Corollary 2.1.1) we have for all $\beta > 0$, $E[\exp(\beta W^2)] < \infty$. In particular $E[W^p] < \infty$, for all $p > 0$. From (2.34) we get

$$
(2.48)\qquad \left[\sup_n \sup_x |E\check{f}_n''(x) - f''(x)| \check{h}_n^{-1} \right]^p < \infty.
$$

We have $\check{h}_n = (n/\log\log n)^{-1/8}$. From here it follows that

$$
(2.49)\qquad E\left[\sup_n \sup_x |\check{f}_n''(x) - f''(x)| (n/\log\log n)^{1/8} \right]^p < \infty.
$$

In particular,

$$
(2.50)\qquad E\left[\left(\sup_n \left| \int_{-\check{d}_n}^{\check{d}_n} (|\check{f}_n''(x)| - |f''(x)|) dx \right| n^{\delta} (\log\log n)^{-1/8} \right)^p \right] < \infty \qquad \forall p > 0,
$$

which is slightly stronger than Lemma 2.6.

PROOF OF THEOREM 2.1. From our definition of T_w it is obvious that as $w \to 0$, $T_w \to \infty$ a.s. and we get the following inequalities

$$
(2.51)\qquad T_w \ge \{[\alpha\hat{\theta}_{1T_w}/\sqrt{c} + (\beta/2)c^2\hat{\theta}_{2T_w} + T_w^{-\xi}]w^{-1}\}^{5/2}
$$

and

$$
(2.52)\qquad T_w - 1 < \{[\alpha\hat{\theta}_{1T_w-1}/\sqrt{c} + (\beta/2)c^2\hat{\theta}_{2T_w-1} + (T_w - 1)^{-\xi}]w^{-1}\}^{5/2}.
$$

(2.18) follows immediately from Lemma 2.2, Lemma 2.4 and inequalities (2.51) and (2.52).

In view of (2.18), to prove (2.19) it suffices to show $\{w^{5/2}T_w : 0 < w < 1\}$ is uniformly integrable. Therefore, it is enough to show that

$$\{[\alpha\hat{\theta}_{1T_w-1}/\sqrt{c} + (\beta/2)c^2\hat{\theta}_{2T_w-1} + (T_w - 1)^{\ \xi}]^{5/2} : 0 < w < 1\}$$

is uniformly integrable. This is equivalent to showing

$$\{\hat{\theta}^{5/2}_{1T_w-1} : 0 < w < 1\} \quad \text{is u.i.} \tag{2.53}$$

and

$$\{\hat{\theta}^{5/2}_{2T_w-1} : 0 < w < 1\} \quad \text{is u.i.} \tag{2.54}$$

From Lemma 2.5,

$$E\left[\left(\sup_n \left|\int_{-\tilde{d}_n}^{\tilde{d}_n} \sqrt{\tilde{f}_n} - \int_{-\tilde{d}_n}^{\tilde{d}_n} \sqrt{f}\right|\right)^p\right] < \infty \qquad \forall p > 0$$

and

$$\left|\int_{-\tilde{d}_n}^{\tilde{d}_n} \sqrt{f} - \int \sqrt{f}\right|^p \to 0 \quad \text{as} \quad n \to \infty.$$

Hence,

$$E\left[\left(\sup_n \left|\int_{-\tilde{d}_n}^{\tilde{d}_n} \sqrt{\tilde{f}_n} - \int \sqrt{f}\right|\right)^p\right] < \infty \qquad \forall p > 0$$

i.e.,

$$E\left[\sup_n |\hat{\theta}_{1n} - \theta_1|\right]^p < \infty \qquad \forall p > 0. \tag{2.55}$$

This implies $\{\hat{\theta}^p_{1T_w-1} : 0 < w < 1\}$ is u.i. for all $p > 0$. Similarly, it can be shown using Lemma 2.6 that $\{\hat{\theta}^p_{2T_w-1} : 0 < w < 1\}$ is u.i. for all $p > 0$, and this finishes the proof of (2.19).

To prove (2.20), first write

$$J_{T_w}/w = (J_{T_w}/H_{T_w}) \cdot (H_{T_w}/w). \tag{2.56}$$

To show that the left hand side is less than or equal to 1 asymptotically, we shall show that the first term in this decomposition is asymptotically no larger than 1 and the second term approaches 1 as w goes to zero. First consider the second term. From the definition of n^*_w (2.9) we have

$$H_{T_w}/w = \{\alpha\theta_1/\sqrt{c} + (\beta/2)\theta_2 c^2\}T_w^{-2/5}w^{-1} \approx (n^*_w)^{2/5}T_w^{-2/5}. \tag{2.57}$$

From (2.18) it follows that

$$H_{T_w}/w \to 1 \quad \text{a.s.} \tag{2.58}$$

Next we consider the first term of (2.56). From Pinelis (1990), we obtain

$$P(|J_n - EJ_n| > \epsilon) \leq 2\exp(-n\epsilon^2/2\|K\|^2) \quad \forall \epsilon > 0. \tag{2.59}$$

Therefore, for n sufficiently large,

$$\begin{aligned} P(|J_n/EJ_n - 1| > \epsilon) &= P(|J_n - EJ_n| > \epsilon EJ_n) \\ &\leq P(|J_n - EJ_n| > cD_2(f)n^{-2/5}\epsilon), \end{aligned} \tag{2.60}$$

from Devroye and Győrfi ((1985), p. 37), where $D_2(f)$ is a constant depending on f.

Therefore, by (2.59) and (2.60),

$$P(|J_n/EJ_n - 1| > \epsilon) \leq 2\exp(-n^{1/5}c^2(D_2(f))^2\epsilon^2/(2\|K\|^2)). \tag{2.61}$$

Using the Borel-Cantelli Lemma it is easy to see that

$$J_n/EJ_n \to 1 \quad \text{a.s.} \tag{2.62}$$

From (2.1) we have $EJ_n \leq H_n + o(n^{-2/5})$. Hence,

$$\limsup_{n\to\infty} J_n/H_n \leq 1 \quad \text{a.s.} \tag{2.63}$$

Since $T_w \to \infty$ a.s. as $w \to 0$,

$$\limsup_{w\to 0} J_{T_w}/H_{T_w} \leq 1 \quad \text{a.s.} \tag{2.64}$$

Combining we get (2.20). To prove (2.21) we first prove the following lemma.

LEMMA 2.7.

$$\{(T_w^{2/5}J_{T_w})^p : 0 < w < 1\} \quad \textit{is u.i.} \quad \forall p. \tag{2.65}$$

PROOF. By (2.59) we get

$$\begin{aligned} E(n^{2/5}|J_n - EJ_n|)^p &\leq 2p\int_0^\infty t^{p-1}\exp(-n^{1/5}t^2/\|K\|^2)dt \\ &= O(n^{-p/10}). \end{aligned} \tag{2.66}$$

This implies that

$$\sup_{n\geq m} E(n^{2/5}|J_n - EJ_n|)^p \to 0 \quad \text{as} \quad m \to \infty. \tag{2.67}$$

Write $m(w) = w^{-5/(2+5\zeta)}$ and note that from (2.10), $T_w \geq m(w)$. Define $a(n) = EJ_n$. Then we have

$$
\begin{aligned}
(2.68)\quad & E(T_w^{2/5}|J_{T_w} - a(T_w)|)^p \\
&= \sum_{n\geq m(w)} E\{(n^{2/5}|J_n - EJ_n|)^p 1_{(T_w=n)}\} \\
&\leq \sum_{n\geq m(w)} E^{1/2}(n^{2/5}|J_n - EJ_n|)^{2p} \cdot \sqrt{P(T_w = n)} \\
&\leq \left[\sum_{n\geq m(w)} E(n^{2/5}|J_n - EJ_n|)^{2p}\right]^{1/2} \cdot \left[\sum_{n\geq m(w)} P(T_w = n)\right]^{1/2} \\
&= \left[\sum_{n\geq m(w)} E(n^{2/5}|J_n - EJ_n|)^{2p}\right]^{1/2} \\
&= \left[O(1) \sum_{n\geq m(w)} n^{-p/5}\right]^{1/2} \longrightarrow 0
\end{aligned}
$$

for $p > 5$ as $w \to 0$, which implies (2.65).

Returning to the proof of (2.21), by Hölder's inequality

$$E(J_{T_w}/w) \leq \{E(J_{T_w}/H_{T_w})^p\}^{1/p}\{E(H_{T_w}/w)^q\}^{1/q}, \tag{2.69}$$

where $1/p + 1/q = 1$. Write

$$\theta = \{\alpha\theta_1/\sqrt{c} + (\beta/2)\theta_2 c^2\}.$$

Then $H_n = \theta n^{-2/5} = (l_1\theta_1 + l_2\theta_2)n^{-2/5}$, say. By (2.64) and Fatou's Lemma,

$$\limsup_{w\to 0} E(J_{T_w}/H_{T_w})^p \leq 1. \tag{2.70}$$

It remains to show that

$$\limsup_{w\to 0} E(H_{T_w}/w)^q \leq 1. \tag{2.71}$$

Write

$$H_{T_w}/w = \theta T_w^{-2/5} w^{-1} \leq \theta/\hat{\theta}_{T_w}$$

where θ is as before and

$$\hat{\theta}_n = \alpha\hat{\theta}_{1n}/\sqrt{c} + (\beta/2)c^2\hat{\theta}_{2n} + n^{-\xi} = l_1\hat{\theta}_{1n} + l_2\hat{\theta}_{2n} + n^{-\xi}. \tag{2.72}$$

Define $\theta_{1n} = \int_{-\check{d}_n}^{\check{d}_n} \sqrt{f}$, $\theta_{2n} = \int_{-\check{d}_n}^{\check{d}_n} |f''|$, and $\theta_n = l_1\theta_{1n} + l_2\theta_{2n} + n^{-\xi}$. It is obvious that $\theta_n \to \theta$.

Because $0 < \xi < \delta/2$,

$$
\begin{aligned}
(2.73)\quad E\left(\sup_{n\geq m(w)} |\hat{\theta}_n^{-1} - \theta_n^{-1}|^q\right)
&= E\left(\sup_{n\geq m(w)} |(l_1\hat{\theta}_{1n} + l_2\hat{\theta}_{2n} + n^{-\xi})^{-1} - (l_1\theta_{1n} + l_2\theta_{2n} + n^{-\zeta})^{-1}|^q\right)\\
&\leq E\left(\sup_{n\geq m(w)} |[l_1(\theta_{1n} - \hat{\theta}_{1n}) + l_2(\theta_{2n} - \hat{\theta}_{2n})]n^{2\xi}|^q\right)\\
&= O(m(w)^{-(\delta-2\xi)q}),
\end{aligned}
$$

by (2.44) and (2.50). (2.73) yields

$$
(2.74)\qquad E|\hat{\theta}_{T_w}^{-1} - \theta_{T_w}^{-1}|^q \to 0.
$$

This completes the proof of Theorem 2.1.

The proof of Theorem 2.2 is analogous to the proof of Theorem 2.1 and is omitted.

2.4 *Estimating densities with bounded support*

We can improve the results obtained in the previous sections if we further assume that the unknown density f has a bounded support. Under the assumptions that K is a symmetric, bounded probability density function with compact support, f is twice differentiable with bounded continuous second derivative and bounded support, $\int \sqrt{f} < \infty$, $\int |f''| < \infty$, and $h_n \to 0$, $nh_n \to \infty$, we can improve upon the bounds in (2.1). The new bound is given by

$$
\begin{aligned}
(2.75)\qquad EJ_n \leq (\sqrt{2/\pi})\left(\alpha\int\sqrt{f}\right)/\sqrt{nh_n} + (\beta/2)h_n^2\int|f''|\\
+ o(h_n^2 + (nh_n)^{-1/2}),
\end{aligned}
$$

see Devroye and Győrfi (1985).

As in Subsection 2.3 we can get stopping times based on this bound. We define the stopping rule T_w = first $n \geq 1$ such that

$$
(2.76)\qquad n^{-2/5}[(\sqrt{2/\pi})(\alpha\hat{\theta}_{1n})/\sqrt{c} + (\beta/2)c^2\hat{\theta}_{2n} + n^{-\xi}] \leq w,
$$

where $\hat{\theta}_{1n}$ and $\hat{\theta}_{2n}$ are estimates of θ_1 and θ_2 respectively, based on $X_1, X_2, \ldots, X_n$, and $0 < \xi < 1/16$. Here the estimates of θ_1 and θ_2 are

$$
(2.77)\qquad \hat{\theta}_{1n} = \hat{\theta}_{1n}(\tilde{h}_n) = \int_{-\infty}^{\infty}\sqrt{\tilde{f}_n},
$$

$$
(2.78)\qquad \hat{\theta}_{2n} = \hat{\theta}_{2n}(\check{h}_n) = \int_{-\infty}^{\infty}|\check{f}_n''|,
$$

where $\tilde{f}_n$ and $\check{f}_n$ are as before. Note that no truncation of the integral is necessary.

Estimate the unknown density f by

$$(2.79)\qquad \hat{f}_{T_w}(x) = (T_w h_{T_w})^{-1} \sum_1^{T_w} K((x - X_i)/h_{T_w}),$$

where $h_{T_w} = cT_w^{-1/5}$. It is quite easy to show that these stopping rules are asymptotically optimal and we get the following theorem.

THEOREM 2.3. *Assume* (2.14) *and* (2.15). *Then as* $w \to 0$

i) $T_w/n_w^* \to 1$ *a.s.*

ii) $ET_w/n_w^* \to 1$

iii) $\limsup_{w\to 0} J_{T_w}/w \leq 1$ *a.s.*

iv) $\limsup_{w\to 0} EJ_{T_w}/w \leq 1.$

PROOF. The proof is analogous to the proof of Theorem 2.1 and is omitted.

We can also obtain analogous results for a two-stage version of the above sequential procedure.

3. Data driven bandwidth selection for the two-stage procedure

Recently data driven bandwidth selection has received considerable attention. Many researchers have observed great potential for automatic bandwidth choice in kernel density estimation: see Wand and Jones (1995), Cao *et al.* (1994), Sheather (1992), and Park and Turlach (1992), and the references they cite.

In this paper so far we have dealt with bandwidths that go to zero at the optimal rate. Equation (2.4) gives the optimal choice among such bandwidths in the sense that the resulting bound (H_n) for the MIAE is asymptotically minimized. Suppose now we want to achieve the bound w for MIAE such that the resulting H_n behaves as in (2.5). The appropriate nonrandom sample size is given by

$$(3.1)\qquad N_w^{OPT} = [\tau^* A(K)B(f)/w]^{5/2}.$$

This sample size would work if we used the optimal bandwidth given by $h_n^{OPT} = c^* n^{-1/5}$ where

$$c^* = [(\alpha\theta_1)/(2\beta\theta_2)]^{2/5}.$$

Once again the choices of N_w^{OPT} and bandwidth h_n^{OPT} depend on the unknown density f which in turn suggests the following stopping rule. Let T_w^{OPT} = first $n \geq 1$ such that

$$(3.2)\qquad n^{-2/5}[\tau^* A(K)\widehat{B(f)} + n^{-\xi}] \leq w,$$

where $\widehat{B(f)}$ is an estimate of $B(f)$ and the estimate of the density is given by

$$\hat{f}_{T_w^{OPT}}(x \mid \hat{h}_{T_w^{OPT}}) = (T_w^{OPT}\hat{h}_{T_w^{OPT}})^{-1} \sum_{1}^{T_w^{OPT}} K((x - X_i)/\hat{h}_{T_w^{OPT}}) \tag{3.3}$$

where $\hat{h}_{T_w^{OPT}}$ is a data driven bandwidth.

Under certain conditions we would like to obtain results similar to those in Theorem 2.1. The analogues of (2.18) and (2.19), namely

$$T_w^{OPT}/N_w^{OPT} \to 1 \quad \text{a.s.}$$

and

$$ET_w^{OPT}/N_w^{OPT} \to 1,$$

as $w \to 0$, follow by essentially the same arguments as for (2.18) and (2.19), provided that the optimal bandwidth is estimated as indicated in (3.6) and (3.7) below, with T_w replacing both $n_0(w)$ and $n_1^{OPT}(w)$. However, proving the analogues of (2.20) and (2.21) is much more difficult and so far we have been unable to obtain the desired results.

We propose a two stage procedure which achieves more of the desired asymptotic goals. (3.2) suggests the following two stage procedure. First draw a sample of size $n_0 = n_0(w) \sim m(w)$ where $m(w)$ is as before. Then based on this sample estimate $B(f)$ and N_w^{OPT}. Now we draw a second sample of size $n_1^{OPT}(w)$, where

$$n_1^{OPT}(w) = \text{smallest integer} \geq [\tau^* A(K)\widehat{B(f)}/w]^{5/2} \vee n_0(w). \tag{3.4}$$

Estimate f by

$$\hat{f}_{n_1^{OPT}(w)}(x) = (\hat{h}_{n_1^{OPT}(w)} n_1^{OPT}(w))^{-1} \cdot \sum_{n_0(w)+1}^{n_0(w)+n_1^{OPT}(w)} K((x - X_i)/\hat{h}_{n_1^{OPT}(w)}), \tag{3.5}$$

where $\hat{h}_{n_1^{OPT}(w)}$ is a data driven bandwidth depending only on $X_1, X_2, \ldots, X_{n_0}$. Note that we are using only the second sample for estimating the density. The optimal choice for h_n that minimizes H_n is given by $h_n^{OPT} = c^* n^{-1/5}$. Once again since c^* is unknown we use the "plug-in" bandwidth

$$\hat{h}^{OPT}_{n_1^{OPT}(w)} = \hat{c}^*_{n_0}(n_1^{OPT}(w))^{-1/5}, \tag{3.6}$$

where

$$\hat{c}^*_{n_0} = [(\alpha\hat{\theta}_{1n_0})/(2\beta\hat{\theta}_{2n_0})]^{2/5}, \tag{3.7}$$

and $\hat{\theta}_{1n_0}$ and $\hat{\theta}_{2n_0}$ are estimates of θ_1 and θ_2 respectively, based on $X_1, X_2, X_3, \ldots, X_{n_0}$. For estimating $B(f)$ and $\hat{c}^*_{n_0}$ we modify our estimates of θ_1 and θ_2 slightly. Take

$$\hat{\theta}_{1n} = \int_{-\tilde{d}_n}^{\tilde{d}_n} \sqrt{\tilde{f}_n + n^{-\gamma}}$$

and

$$\hat{\theta}_{2n} = \int_{-\check{d}_n}^{\check{d}_n} |\check{f}_n''| + n^{-\gamma}.$$

Define

$$\widehat{B(f)} = (\hat{\theta}_{1n_0}^4 \hat{\theta}_{2n_0}/2)^{1/5}.$$

Then we get the following results.

THEOREM 3.1. *Let $n_2^{OPT}(w) = n_0(w) + n_1^{OPT}(w)$ be the total sample size. Assume* (2.14), (2.15) *and $0 < \gamma < \delta/2$. Then, as $w \to 0$*

(3.8) $$n_2^{OPT}(w)/N_w^{OPT} \to 1 \qquad a.s.$$

(3.9) $$En_2^{OPT}(w)/N_w^{OPT} \to 1$$

(3.10) $$\limsup_{w\to 0} E[J_{n_1^{OPT}(w)}|\hat{h}^*_{n_1^{OPT}(w)}]/w \le 1.$$

PROOF. We have $\hat{\theta}_{1n} \to \theta_1$ a.s., $\hat{\theta}_{2n} \to \theta_2$ a.s. as $n \to \infty$, and as $w \to 0$, $N_w^{OPT} \to \infty$. Therefore, $\widehat{B(f)} \to B(f)$ a.s.,

$$n_2^{OPT}(w)/n_w^{OPT} \ge \{\widehat{B(f)}/B(f)\}^{5/2}$$

and

$$[n_2^{OPT}(w) - 1]/n_w^{OPT} \le \{\widehat{B(f)}/B(f)\}^{5/2}.$$

Moreover, by our choice of $n_0(w)$ we have $n_0(w)/n_w^{OPT} \to 0$. This proves (3.8).

To show

$$En_2^{OPT}(w)/n_w^{OPT} \to 1$$

it is enough to show that $\{w^{5/2}n_1^{OPT}(w) : 0 < w < 1\}$ is u.i., which is equivalent to showing that

(3.11) $$\{(\tau^* A(K)\widehat{B(f)})^{5/2} : 0 < w < 1\} \quad \text{is u.i.}$$

So, it is enough to prove that $\{\widehat{B(f)}^p : 0 < w < 1\}$ is u.i. for all $p \ge 1$. This follows from the fact that all powers of $\hat{\theta}_{1n}$ and $\hat{\theta}_{2n}$ are uniformly integrable, so the proof of (3.9) is complete.

Next we want to show (3.10). For convenience, abbreviate $n_1^{OPT}(w)$ by n_1. By independence, for every $\epsilon > 0$, if w is sufficiently small (so that $n_0(w)$ is sufficiently large),

(3.12) $$\begin{aligned}\int E(|\hat{f}_{n_1} - f|) &= \int \sum_m E[(|\hat{f}_m - f|)I_{\{n_1=m\}}] \\ &= \int \sum_m E\{E[(|\hat{f}_m - f|I_{\{n_1=m\}}) \mid X_1, X_2, \ldots, X_{n_0}]\}\end{aligned}$$

$$
\begin{aligned}
&= \int \sum_m E\{E[|\hat{f}_m - f| \mid X_1, X_2, \ldots, X_{n_0}] I_{\{n_1=m\}}\} \\
&\leq \sum_m E\Big\{\Big(\alpha \int \sqrt{f}/\sqrt{m\hat{h}_m^*} \\
&\qquad + (\beta/2)(\hat{h}_m^*)^2 \int |f''| + o(m^{-2/5})\Big) I_{\{n_1=m\}}\Big\} \\
&= E[(\hat{c}_{n_0}^*)^{-1/2}(\alpha\theta_1 + (\beta/2)\theta_2(\hat{c}_{n_0}^*)^{5/2}) m^{-2/5} I_{\{n_1=m\}}] \\
&\quad + \epsilon E[(n_1^{OPT})^{-2/5}] \\
&\leq E[((\alpha\hat{\theta}_{1n_0})/(2\beta\hat{\theta}_{2n_0}))^{-1/5} \\
&\qquad \cdot \{\alpha\theta_2(\theta_1/\theta_2 + \hat{\theta}_{1n_0}/(4\hat{\theta}_{2n_0}))\} w/(\tau^* A(K)\tilde{B}(\tilde{f}))] \\
&\quad + O(1)\epsilon w E[\hat{\theta}_{1n_0}^{-4/5}\hat{\theta}_{2n_0}^{-1/5}] \\
&\leq E[\theta_1\hat{\theta}_{1n_0}^{-1} + \theta_2\hat{\theta}_{2n_0}^{-1}/4](5w/4) + O(1)\epsilon w E^{4/5}[\hat{\theta}_{1n_0}^{-1}] E^{1/5}[\hat{\theta}_{2n_0}^{-1}].
\end{aligned}
$$

Now with the modified estimates it remains to show that

$$E(\theta_1\hat{\theta}_{1n_0}^{-1}) \to 1 \tag{3.13}$$

and

$$E(\theta_2\hat{\theta}_{2n_0}^{-1}) \to 1. \tag{3.14}$$

The proof of (3.13) and (3.14) is very similar to that for (2.73).

Now let ϵ go to 0 to get (3.10).

Acknowledgements

We would like to thank the Associate Editor and referees for a careful reading of the manuscript and for comments that improved it.

Appendix

We did some Monte Carlo simulations for empirical verification. These simulations are related to the results stated in Section 2. We considered i.i.d. observations from a mixture of normal distributions, $f \sim 0.3N(-4,4)+0.4N(0,1)+0.3N(2,1)$. A standard normal kernel was used. Although the standard normal kernel does not satisfy the required conditions still it gives quite good results. We had to use numerical integration for computing $\hat{\theta}_1$ and $\hat{\theta}_2$ as there are no closed form expressions available for them. To avoid huge computations we restricted ourselves to the two stage procedure. We used $n_0(w) = w^{-5/2}$ as the initial sample size and took $\xi = 1/32$. Fixed bandwidths were chosen as $h_n = cn^{-1/5}$ for different values of c. The results show that the choice of c influences the sample size greatly. Four different values of c were used: 0.5, 0.8, 1.0 and 2.0. Three values of w were used: 0.03, 0.05 and 0.1 . Thirty repetitions were conducted for each combination of h_n and w.

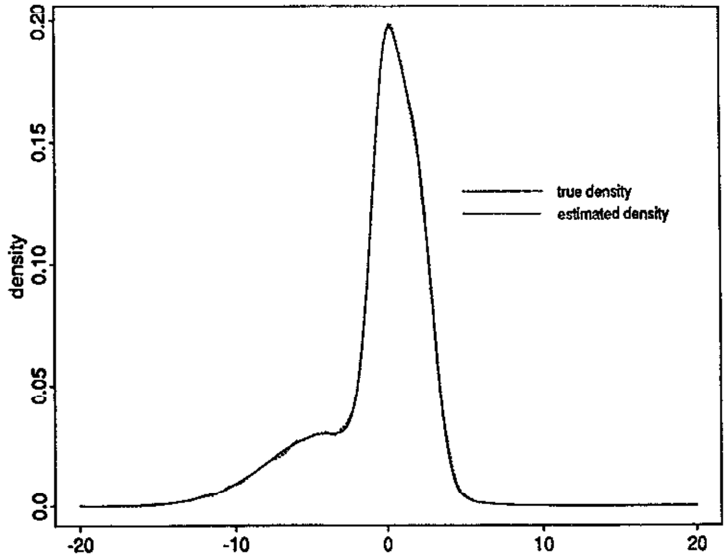

Fig. A1. Estimated density and true density when $w = 0.03$, $h_n = 2.0n^{-1/5}$.

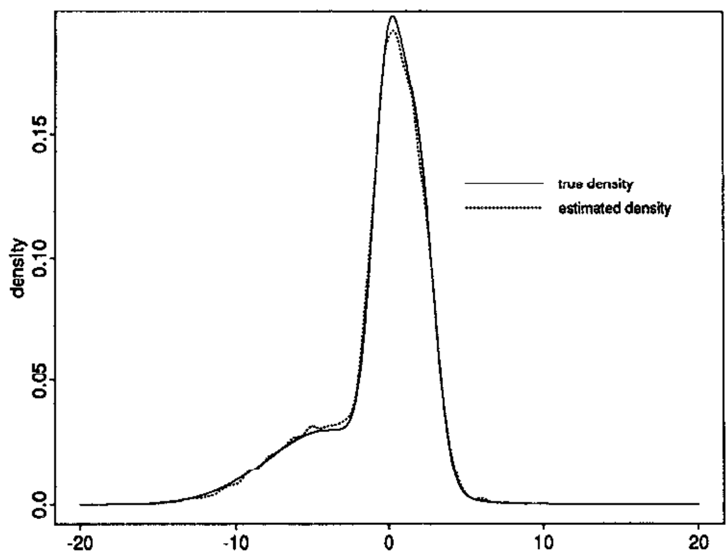

Fig. A2. Estimated density and true density when $w = 0.05$, $h_n = 2.0n^{-1/5}$.

We used Splus for carrying out the computations. Plots for estimated densities and the actual densities are shown for the following cases: (i) $w = 0.03$, $h_n = 2.0n^{-1/5}$; (ii) $w = 0.05$, $h_n = 2.0n^{-1/5}$; (iii) $w = 0.10$, $h_n = 2.0n^{-1/5}$; (iv) $w = 0.10$, $h_n = 0.5n^{-1/5}$. The plots for the first three cases are rather impressive. The fit of the estimated density to the true density, as judged by the human eye,

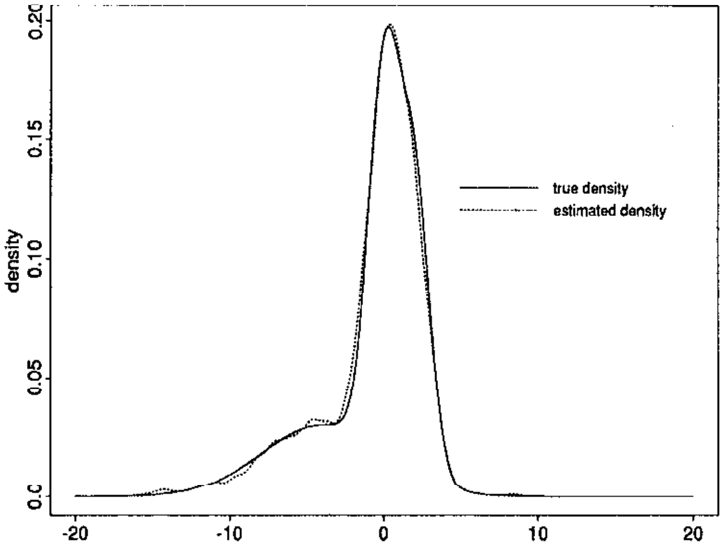

Fig. A3. Estimated density and true density when $w = 0.10$, $h_n = 2.0n^{-1/5}$.

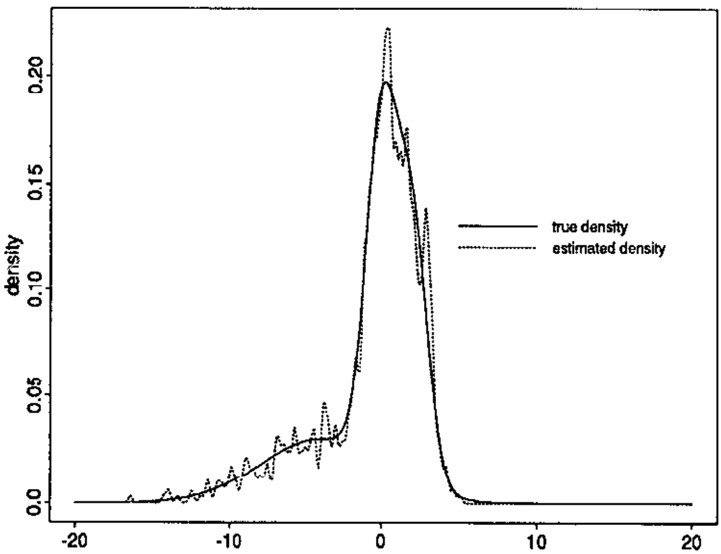

Fig. A4. Estimated density and true density when $w = 0.10$, $h_n = 0.5n^{-1/5}$.

is quite good. Moreover, as one would expect, the fit gets better as w decreases. The fourth case illustrates the sensitivity of the results to the choice of bandwidth. Because the bandwidth in this case is relatively small the estimated density is choppier, although it still gives a pretty good idea of the overall shape. As is typical in nonsequential density estimation, when the underlying density is very

Table A1. Summary of the simulation study for the two-stage procedure.

w	c	n_*	n_2	$\hat{\theta}_1$	$\hat{\theta}_2$	IAE
0.03	0.5	22302	21369.3	3.8632	0.6156	0.04271
0.03	0.8	16388	15392.0	3.8548	0.6349	0.03948
0.03	1.0	16250	14767.3	3.8545	0.6164	0.03531
0.03	2.0	44544	37127.2	3.8579	0.6234	0.01989
0.05	0.5	6219	5569.2	3.7934	0.4921	0.07101
0.05	0.8	4570	3921.7	3.8200	0.5117	0.06489
0.05	1.0	4532	3645.4	3.7972	0.5156	0.05870
0.05	2.0	12422	13188.6	3.8084	0.7265	0.03078
0.10	0.5	1100	1019.1	3.6971	0.9656	0.14351
0.10	0.8	808	813.9	3.6987	0.9114	0.12085
0.10	1.0	802	876.3	3.7514	0.8852	0.11084
0.10	2.0	2196	3485.1	3.7171	0.9210	0.05374

smooth larger bandwidths tend to outperform smaller ones.

Numerical integrations were used to find observed IAE and it was compared with w. For narrower bandwidths the values of IAE generally are not less than w as expected, even though they are quite close to w. One possible reason is that our method does not estimate θ_2 very well, causing $n_2(w)$ to be much less than n_w^*. On the other hand when we take $c = 2$ the estimates improve significantly, and in that case the IAE is less than w as expected. This shows the importance of the bandwidth selection. Overall the two stage procedure performs satisfactorily. In particular, the parameter choices examined here work reasonably well for this density. It would be of interest to see the extent to which they work well in other situations, or whether there are better (universal) choices.

In most cases the number of observations required to achieve (or at least come close to) the desired L_1 error bound is large but not ridiculously so. Silverman (1986) gives an example involving 15,000 observations of the height of a steel surface, taken from Bowyer (1980), and another example with 4763 observed time intervals between successive micro-earthquakes in an area in California, taken from Rice (1975). These sample sizes are of the same order of magnitude as those found in the simulation study.

Neutra *et al.* (1978) analyzed data on the effect of fetal monitoring on neonatal death rates. The study covered 15,846 babies born at Boston's Beth Israel Hospital, an illustration that our methods could be applied in a large scale medical study (in this particular instance the densities of certain covariates would be of interest).

References

Bowyer, A. (1980). Experiments and computer modelling in stick-slip, Ph.D. Thesis, Department of Actuarial Science and Statistics, University of London.

Cao, R., Cuevas, A. and Manteiga, W. G. (1994). A comparative study of several smoothing methods in density estimation, *Comput. Statist. Data Anal.*, **17**, 153–176.

Carroll, R. J. (1976). On sequential density estimation, *Zeit. Wahrscheinlichkeitsth.*, **36**, 137–151.

Devroye, L. (1983). The equivalence of weak, strong and complete convergence in L_1 for kernel density estimates, *Ann. Statist.*, **11**, 896–904.

Devroye, L. (1988). Asymptotic performance bounds for the kernel estimate, *Ann. Statist.*, **16**, 1162–1179.

Devroye, L. (1991). Exponential inequalities in nonparametric estimation, *Nonparametric Functional Estimation and Related Topics* (ed. G. Roussas), 31–44, Kluwer.

Devroye, L. and Györfi, L. (1985). *Nonparametric Density Estimation: The L_1 View*, Wiley, New York.

Devroye, L. and Wand, M. P. (1993). On the effect of density shape on the performance of its kernel estimate, *Statistics*, **24**, 215–233.

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

Hall, P. and Wand, M. P. (1988). Minimizing L_1 distance in nonparametric density estimation, *J. Multivar. Anal.*, **26**, 59–88.

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, D. (1988). A global stopping rule for recursive density estimators, *J. Statist. Plann. Inference*, **20**, 946–954.

Martinsek, A. T. (1992). Using stopping rules to bound the mean integrated squared error in density estimation, *Ann. Statist.*, **20**, 797–806.

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

Neutra, R. R., Fienberg, S. E., Greenland, S. and Friedman, E. A. (1978). The effect of fetal monitoring on neonatal death rates, *New England Journal of Medicine*, **299**, 324–326.

Park, B. U. and Turlach, B. A. (1992). Practical performance of several data driven bandwidth selectors, *Comput. Statist.*, **7**, 251–270.

Pinelis, I. (1990). To Devroye's estimates for the distribution of density estimators, Tech. Rep. (personal communication).

Prakasa Rao, B. L. S. (1983). *Nonparametric Functional Estimation*, Academic Press, Orlando.

Rice, J. (1975). Statistical methods of use in analyzing sequences of earthquakes, *Geophysical Journal of the Royal Astronomical Society*, **42**, 671–683.

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.

Sheather, S. J. (1992). The performance of six popular bandwidth selection methods on some real data sets, *Comput. Statist.*, **7**, 225–250.

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.

Stute, W. (1983). Sequential fixed-width confidence intervals for a nonparametric density function, *Zeit. Wahrscheinlichkeitsth.*, **62**, 113–123.

Wand, M. P. and Devroye, L. (1993). How easy is a given density to estimate?, *Comput. Statist. Data Anal.*, **16**, 311–323.

Wand, M. P. and Jones, M. C. (1995). *Kernel Smoothing*, Chapman and Hall, London.

Wegman, E. J. and Davies, H. I. (1975). Sequential nonparametric density estimation, *IEEE Transactions on Information Theory*, **21**, 619–628.

Yamato, H. (1971). Sequential estimate of a continuous probability density function and mode, *Bull. Math. Statist.*, **14**, 1–12.