Ann. Inst. Statist. Math.
Vol. 44, No. 3, 563–577 (1992)

OPTIMUM BANDWIDTHS AND KERNELS FOR ESTIMATING CERTAIN DISCONTINUOUS DENSITIES

B. K. GHOSH AND WEI-MIN HUANG

Department of Mathematics, Lehigh University, Bethlehem, PA 18015, U.S.A.

(Received August 22, 1990; revised June 21, 1991)

Abstract. Rosenblatt and Parzen proposed a well-known estimator f_n for an unknown density function f, and later Schuster suggested a modification $\hat{f}_n$ to rectify certain drawbacks of f_n. This paper gives the asymptotically optimum bandwidth and kernel for $\hat{f}_n$ under the standard measure of IMSE when f is discontinuous at one or both endpoints of its support. We also consider an alternative definition of the IMSE under which the optimum bandwidths and kernels for f_n and $\hat{f}_n$ are derived. The latter supplement van Eeden's results.

Key words and phrases: Estimation of discontinuous densities, alternative notions of IMSE, modified kernel density estimates, optimal bandwidths, optimal kernels.

1. Introduction

Let $X_1, \dots, X_n$ be independent and identically distributed random variables having a common density function $f \in \mathcal{F}$ with known support $[\alpha, \beta]$, $[\alpha, \infty)$ or $(-\infty, \beta]$. Consider the problem of estimating $f(x)$ at a given point x using the sample $(X_1, \dots, X_n)$. A common estimator, proposed by Rosenblatt (1956) and Parzen (1962), is given by

$$f_n(x) = (nb_n)^{-1} \sum_{i=1}^{n} w(b_n^{-1}[x - X_i]). \tag{1.1}$$

In (1.1), b_n is a predetermined *bandwidth* satisfying

$$b_n \to 0 \quad \text{and} \quad nb_n \to \infty \quad \text{as} \quad n \to \infty \tag{1.2}$$

and $w \in \mathcal{W}$ is a suitably chosen *kernel*, where the class $\mathcal{W}$ is defined by (all integrals without limits are over $\mathbb{R}$)

$$\begin{gathered} w \geq 0, \quad \int w(z)dz = 1, \quad \int z^2 w(z)dz < \infty, \\ \int w^2(z)dz < \infty, \quad \int |z| w^2(z)dz < \infty. \end{gathered} \tag{1.3}$$

The class $\mathcal{W}$ defined by (1.3) is slightly more general than what one usually encounters in the literature; w need not be bounded or symmetric (e.g., $w(z) = (16)^{-1}z^{-1/4}$ for $0 \leq z \leq 16$ and $w(z) = 4^{-1}|z|^{-1/4}$ for $-1 \leq z \leq 0$ is permissible in the developments below). The varying nature of the family $\mathcal{F}$ will be described in the third paragraph. The properties of f_n are generally quite difficult to investigate for arbitrary n. Consequently, the decision on the best choice of the pair (b_n, w) is usually made from the asymptotic behavior of f_n. We stipulate that the best choice of (b_n, w) should, for large n, minimize the *integrated mean square error* (IMSE), which is defined by

$$(1.4)\qquad \begin{aligned} R_n(f, b_n, w) &= \int E[f_n(x) - f(x)]^2 dx \\ &= \int [Ef_n(x) - f(x)]^2 dx + \int [\operatorname{Var} f_n(x)] dx. \end{aligned}$$

Suppose one can show that, for each $f \in \mathcal{F}$ and large n, R_n in (1.4) reduces to

$$(1.5)\qquad \begin{aligned} &R_n(f, b_n, w) \\ &= \begin{cases} (nb_n)^{-1}A(f,w) + b_n^{\delta}B(f,w) & \\ \quad +o(n^{-1}b_n^{-1}) + o(b_n^{\delta}) & \text{for } w \in \mathcal{W}_o \\ (nb_n)^{-1}A'(f,w) + b_n^{\delta'}B'(f,w) & \\ \quad +o(n^{-1}b_n^{-1}) + o(b_n^{\delta'}) & \text{for } w \in \mathcal{W} - \mathcal{W}_o, \end{cases} \end{aligned}$$

where $A > 0$, $B > 0$, $A' > 0$, $B' > 0$, $\delta > \delta' > 0$, and $\mathcal{W}_o$ is nonempty. Then, for each fixed $w \in \mathcal{W}_o$ and $n \geq 1$, the choice (see Parzen (1962), Lemma 4a)

$$(1.6)\qquad b_n(f, w) = [A(f,w)/\delta B(f,w)]^{1/(\delta+1)} n^{-1/(\delta+1)}$$

minimizes the dominant part of (1.5) with respect to b_n under $w \in \mathcal{W}_o$. In fact, for large n,

$$\min_{b_n} R_n(f, b_n, w) = R_n(f, b_n(f,w), w) + o(n^{-\delta/(\delta+1)}),$$

where

$$(1.7)\qquad \begin{aligned} &R_n(f, b_n(f,w), w) \\ &\quad = (\delta+1)[\delta^{-\delta}A(f,w)^{\delta}B(f,w)]^{1/(\delta+1)} n^{-\delta/(\delta+1)} + o(n^{-\delta/(\delta+1)}). \end{aligned}$$

It is easily verified from (1.5) and (1.7) that, for sufficiently large n, (1.7) is less than $R_n(f, b_n, w')$ for each $w \in \mathcal{W}_o$ and $w' \in \mathcal{W} - \mathcal{W}_o$ so that the class $\mathcal{W} - \mathcal{W}_o$ can be ignored for asymptotic purposes. We call $b_n(f, w)$ in (1.6) the (*asymptotically*) *optimum bandwidth* for f_n *under* $w \in \mathcal{W}_o$ (note that (1.6) need not be the optimum bandwidth under $w \in \mathcal{W} - \mathcal{W}_o$). Moreover, if there exists a $w^* \in \mathcal{W}_o$ which minimizes the leading term in $R_n(f, b_n(f,w), w)$ or, equivalently, the functional $A(f,w)^{\delta}B(f,w)$ then we call w^* and $b_n(f, w^*)$ the (*asymptotically*) *optimum kernel* and *bandwidth* for f_n *under* $\mathcal{W}$ (w^* may not minimize $A(f,w)^\delta B(f,w)$ in $\mathcal{W}$ but clearly $R_n(f,b_n,w')/R_n(f,b_n(f,w^*),w^*) \to \infty$ for any $w' \in \mathcal{W} - \mathcal{W}_o$).

Consider now three families of densities,

$\mathcal{F}(-\infty,\infty)$: *f is continuous and bounded in $\mathbb{R}$, $f \geq 0$, $\int f(x)dx = 1, f''$ is continuous and square-integrable;*

$\mathcal{F}(\alpha,\infty)$: *f is continuous and bounded on $[\alpha,\infty), f \geq 0$ on (α,∞), $f = 0$ on $(-\infty,\alpha)$, $f(\alpha) > 0$, $\int f(x)dx = 1, f'$ is continuous and square-integrable on $[\alpha,\infty)$;*

$\mathcal{F}(\alpha,\beta)$: *f is continuous on $[\alpha,\beta]$, $f(\alpha) > 0$, $f(\beta) > 0$, $f \geq 0$ on (α,β) and zero elsewhere, $\int f(x)dx = 1$, f' is bounded on $[\alpha,\beta]$.*

Observe that we are demanding from $\mathcal{F}(\alpha,\infty)$ that f is right-continuous at α and the right-derivative $f'(\alpha)$ exists. It is, of course, understood that $\mathcal{F}(-\infty,\infty)$ and $\mathcal{F}(\alpha,\infty)$ can contain densities with finite support $[\alpha,\beta]$, which implies $f(\alpha) = 0 = f(\beta)$ in the former and $f(\alpha) > 0 = f(\beta)$ in the latter. The family $\mathcal{F}(\alpha,\infty)$ has a dual $\mathcal{F}(-\infty,\beta)$ and the results for the latter follow from those of $\mathcal{F}(\alpha,\infty)$ by taking $X_i' = -X_i$. The bulk of the literature on kernel density estimation involves $\mathcal{F}(-\infty,\infty)$. On the other hand, the families $\mathcal{F}(\alpha,\infty)$, $\mathcal{F}(-\infty,\beta)$ and $\mathcal{F}(\alpha,\beta)$ arise in practice when the observed data follow the uniform, U-shaped, J-shaped and certain Pearsonian distributions as well as singly and doubly truncated versions of all continuous densities (see Johnson and Kotz (1970)). In fact, one would seldom expect an f with a finite or semi-infinite support to belong in $\mathcal{F}(-\infty,\infty)$.

If $f \in \mathcal{F}(-\infty,\infty)$, considerations of the second paragraph lead to the following conclusions under assumptions (1.2) and (1.3) (see Epanechnikov (1969), Rosenblatt (1971) and Silverman (1986)). Let

$$\xi_w = \int zw(z)dz, \qquad \tau_w^2 = \int (z-\xi_w)^2 w(z)dz \tag{1.8}$$

and define $\mathcal{W}_o$ as the subclass of $\mathcal{W}$ having $\xi_w = 0$. Then the optimum bandwidth for f_n under $w \in \mathcal{W}_o$ is

$$b_n(f,w) = \left[\tau_w^{-4}\int w^2(z)dz\right]^{1/5}\left[\int f''(x)^2dx\right]^{-1/5} n^{-1/5}, \tag{1.9}$$

$$\begin{aligned} &R_n(f,b_n(f,w),w) \\ &= \frac{5}{4}\left[\tau_w\int w^2(z)dz\right]^{4/5}\left[\int f''(x)^2dx\right]^{1/5} n^{-4/5} + o(n^{-4/5}), \end{aligned} \tag{1.10}$$

and the kernel minimizing the leading term in (1.10), subject to $\tau_w = \tau$, is

$$w_o^*(z) = \begin{cases} 3(4\tau\sqrt{5}\,)^{-1}(1-z^2/5\tau^2) & \text{for } |z| \leq \tau\sqrt{5} \\ 0 & \text{for } |z| > \tau\sqrt{5}. \end{cases} \tag{1.11}$$

Thus, if our criterion is to minimize R_n for large n, w_o^* and $b_n(f,w_o^*)$ are the optimum kernel and bandwidth for f_n when $f \in \mathcal{F}(-\infty,\infty)$. Similarly, if $f \in \mathcal{F}(\alpha,\beta)$ or $\mathcal{F}(\alpha,\infty)$, then van Eeden (1985) shows that the optimum kernel and bandwidth for f_n are, respectively,

$$(1.12)\qquad w_1^*(z) = (\tau\sqrt{2}\,)^{-1}\exp[-\sqrt{2}\,|z|/\tau] \quad \text{for} \quad -\infty < z < \infty,$$

$$(1.13)\qquad b_n(f, w_1^*) = \sqrt{2}[\tau\{f^2(\alpha) + f^2(\beta)\}^{1/2}]^{-1}n^{-1/2},$$

and

$$(1.14)\qquad R_n(f, b_n(f, w_1^*), w_1^*) = \frac{1}{2}[f^2(\alpha) + f^2(\beta)]^{1/2}n^{-1/2} + o(n^{-1/2}),$$

where τ is an arbitrary positive number (one takes $f(\beta) = 0$ above when $f \in \mathcal{F}(\alpha, \infty)$). van Eeden actually assumes symmetry and boundedness for w but her proof can easily be extended to $\mathcal{W}_o$ (see Cline and Hart (1990)). It is customary to take $\tau = 1$ in w_o^* and w_1^* because any scale change in w can be absorbed in b_n.

The purpose of the present paper is to explore two new aspects of estimation for the families $\mathcal{F}(\alpha, \infty)$ and $\mathcal{F}(\alpha, \beta)$. First, it is easily shown by Parzen's (1962) method that, if x_o is an arbitrary point of discontinuity in f, then (1.2) implies that with probability one we have

$$(1.15)\qquad f_n(x_o) \to f(x_o) + [f(x_o^-) - f(x_o)][1 - W(0)] + [f(x_o^+) - f(x_o)]W(0),$$

where $W(z) = \int_{-\infty}^{z} w(x)dx$. It follows from (1.15) that f_n under any $w \in \mathcal{W}$ cannot be uniformly consistent for $f \in \mathcal{F}(\alpha, \beta)$ while the only uniformly consistent f_n for $f \in \mathcal{F}(\alpha, \infty)$ are those which use a w with support in $\mathbb{R}^-$. In particular, the lack of uniform consistency renders (1.12) and (1.13) unsatisfactory. In Sections 2 and 3, we investigate Schuster's (1985) estimator $\hat{f}_n$ to remedy the situation. Our main result (Theorem 2.1) obtains the optimum kernel and bandwidth for $\hat{f}_n$. Apart from being uniformly consistent, the IMSE of $\hat{f}_n$ is of order $n^{-3/4}$, which tends to zero faster than (1.14). Second, in Section 4, we introduce an alternative definition $\tilde{R}_n$ for the IMSE of f_n, which is sometimes more meaningful than R_n. It is shown (Theorems 4.1 and 4.2) that $\tilde{R}_n$ leads to optimum kernel and bandwidth for f_n that are different from (1.12)–(1.13) for families $\mathcal{F}(\alpha, \infty)$ and $\mathcal{F}(\alpha, \beta)$ but the same as (1.9)–(1.11) for the family $\mathcal{F}(-\infty, \infty)$.

2. A modified estimator and the main result

Schuster (1985) proposed a simple modification $\hat{f}_n$ of f_n. For $f \in \mathcal{F}(\alpha, \infty)$, define

$$(2.1)\qquad \hat{f}_n(x) = \begin{cases} f_n(x) + f_n(2\alpha - x), & x \geq \alpha \\ 0, & \text{elsewhere.} \end{cases}$$

For $f \in \mathcal{F}(\alpha, \beta)$, define

$$(2.2)\qquad \hat{f}_n(x) = \begin{cases} f_n(x) + f_n(2\alpha - x) + f_n(2\beta - x), & \alpha \leq x \leq \beta \\ 0, & \text{elsewhere.} \end{cases}$$

The idea of the reflection principle (i.e., *folding* f_n at α in (2.1) or at both α and β in (2.2)) involved in $\hat{f}_n$ seems to have been well known to time series analysts, but Schuster (1985) and Silverman ((1986), Section 2.10) provide formal, albeit intuitive, justifications for using $\hat{f}_n$ in the present context. Schuster proves the uniform consistency of $\hat{f}_n$ when $f \in \mathcal{F}(\alpha, \infty)$ or $\mathcal{F}(\alpha, \beta)$ and w is symmetric; one can show more generally that the uniform consistency holds if and only if w satisfies $\int_0^\infty w(z)dz = 1/2$ (w need not be symmetric or nonnegative).

Define the IMSE $\hat{R}_n(f, b_n, w)$ of $\hat{f}_n$ as the right-hand side of (1.4) with f_n replaced by $\hat{f}_n$. Our main result is

THEOREM 2.1. *Assume that* (1.2) *and* (1.3) *hold. If* $f \in \mathcal{F}(\alpha, \beta)$ *and* f' *is continuous on* $[\alpha, \beta]$ *with either* $f'(\alpha) \neq 0$ *or* $f'(\beta) \neq 0$, *then the optimum kernel and bandwidth for* $\hat{f}_n$ *in* (2.2) *are, respectively,*

$$w_2^*(z) = \begin{cases} \frac{1}{2}m[(C-1)e^{m|z|}\{\cos(mz) - \sin(m|z|)\} \\ \qquad + Ce^{-m|z|}\{\cos(mz) + \sin(m|z|)\}] & \text{for } |z| \leq \pi/(2m) \\ 0 & \text{for } |z| > \pi/(2m), \end{cases} \tag{2.3}$$

$$b_n(f, w_2^*) = 2^{1/4} m [f'(\alpha)^2 + f'(\beta)^2]^{-1/4} n^{-1/4}, \tag{2.4}$$

where

$$C = (1 - e^{-\pi})^{-1}, \qquad m = 2^{1/2}\tau^{-1}(e^{\pi/2} - e^{-\pi/2})^{-1/2}$$

and τ *is an arbitrary positive number. If* $f \in \mathcal{F}(\alpha, \infty)$ *and* $f'(\alpha) \neq 0$, *then* (2.3) *and* (2.4) *are the optimum kernel and bandwidth for* $\hat{f}_n$ *in* (2.1) *with* $f'(\beta)$ *in* (2.4) *replaced by zero. Moreover, in either case* $\hat{R}_n(f, b_n(f, w_2^*), w_2^*)$ *is given by the right-hand side of* (2.27) *with* $L(w)$ *given by* (2.28).

We make a few remarks before proving the theorem. One consequence of Theorem 2.1 is that $\hat{R}_n(f, b_n(f, w_2^*), w_2^*) = O(n^{-3/4})$, which formally shows why $\hat{f}_n$ should be a better estimate than f_n. In Section 3, we investigate more precisely the efficiency of $\hat{f}_n$ based on (2.3) and (2.4). For the case of $f \in \mathcal{F}(\alpha, \infty)$, Cline and Hart (1990) derive expressions for $\hat{R}_n$ (take their x_o as our α), but not for the optimum w, under various conditions on f and w. However, their expression for $\hat{R}_n$ is incorrect (e.g., their (4.6)), they use stronger conditions (e.g., $\int |z|^3 w(z)dz < \infty$), and there are certain gaps in their proof (e.g., they treated $\hat{p}$ as a constant). Cline and Hart (1990) do not attempt to derive $\hat{R}_n$ when $f \in \mathcal{F}(\alpha, \beta)$.

The kernel w_2^* appears somewhat unusual but the uniform and Epanechnikov kernels can, in fact, be thought of as first and second order approximations to w_2^*. This follows from the power series expansion of (2.3),

$$w_2^*(z) = m\left(C - \frac{1}{2}\right)(1 - m^2 z^2) + O(|z|^3) \quad \text{for} \quad |z| \leq \pi/(2m).$$

Finally, note that (2.1) and (2.2) require the actual values of α and β, which must come either from a knowledge of the population that is being sampled or from some adaptive estimation method (e.g., start with $\hat{\alpha} = \min[X_1, \ldots, X_n]$ and $\hat{\beta} = \max[X_1, \ldots, X_n]$); in the adaptive case the resulting IMSE must be reanalyzed, which is not done in this paper. On the other hand, the occurrence of $f'(\alpha)$ and $f'(\beta)$ in $b_n(f, w_2^*)$ seems to be less restrictive than the occurrence of $\int f''(x)^2 dx$ in $b_n(f, w_o^*)$.

The proof of Theorem 2.1 requires two lemmas.

LEMMA 2.1. *If $f \in \mathcal{F}(\alpha,\beta)$ or $\mathcal{F}(\alpha,\infty)$, (1.2) and (1.3) hold, f' is continuous in the support $[\alpha,\beta]$ of f and w is symmetric about zero, then (defining W as in (1.15))*

$$\begin{aligned}(2.5)\qquad &\hat{R}_n(f, b_n, w)\\ &= (nb_n)^{-1}\int w^2(z)dz\\ &\quad + 4b_n^3\{f'(\alpha)^2 + f'(\beta)^2\}\int_0^\infty\left\{\int_y^\infty [1-W(z)]dz\right\}^2 dy\\ &\quad + O(n^{-1}b_n^{-1/2}) + o(b_n^3).\end{aligned}$$

PROOF. We will drop the subscript from b_n and prove the result (2.5) for the case $\beta = \infty$. The proof for the case $\beta < \infty$ is analogous but more tedious. By definition

$$(2.6)\qquad \hat{R}_n(f, b_n, w) = \int_\alpha^\infty [E\hat{f}_n(x) - f(x)]^2 dx + \int_\alpha^\infty [\text{Var}\,\hat{f}_n(x)]dx.$$

For the first integral, note that the Taylor expansion of (2.1) yields, for all $x \geq \alpha$,

$$\begin{aligned}(2.7)\quad E\hat{f}_n(x) &= f(x)\left[W\left(\frac{x-\alpha}{b}\right) + W\left(-\frac{x-\alpha}{b}\right)\right]\\ &\quad - b\int_{-\infty}^{(x-\alpha)/b} zw(z)f'(x-\theta_1 bz)dz\\ &\quad + \int_{-\infty}^{-(x-\alpha)/b} (2\alpha - 2x - bz)w(z)f'(x+\theta_2\{2\alpha - 2x - bz\})dz,\end{aligned}$$

where $0 < \theta_1, \theta_2 < 1$. Using the symmetry of w twice one gets

$$\begin{aligned}(2.8)\quad &\int_\alpha^\infty [E\hat{f}_n(x) - f(x)]^2 dx\\ &= \int_\alpha^\infty\left[b\int_{-\infty}^{(x-\alpha)/b} zw(z)f'(x-\theta_1 bz)dz\right.\\ &\quad \left. + \int_{-\infty}^{-(x-\alpha)/b}(bz + 2x - 2\alpha)w(z)f'(x+\theta_2\{2\alpha - 2x - bz\})dz\right]^2 dx\\ &= b^3\int_0^\infty\left[\int_y^\infty zw(z)f'(\alpha + by + \theta_1 bz)dz\right.\\ &\quad \left. + \int_y^\infty (z-2y)w(z)f'(\alpha + by - 2\theta_2 by + \theta_2 bz)dz\right]^2 dy.\end{aligned}$$

By the right-continuity of f' at α (see Rosenblatt (1971), p. 1819 or Apostol (1957), p. 441) one finally gets

$$(2.9)\qquad \int_\alpha^\infty [E\hat{f}_n(x) - f(x)]^2 dx = 4b^3 f'(\alpha)^2 \int_0^\infty \left[\int_y^\infty (z-y)w(z)dz\right]^2 dy + o(b^3).$$

Consider next the second integral in (2.6). It follows from (2.1) that

$$(2.10)\qquad \int_\alpha^\infty [\operatorname{Var} \hat{f}_n(x)]dx = \int_\alpha^\infty [\operatorname{Var} f_n(x)]dx + \int_\alpha^\infty [\operatorname{Var} f_n(2\alpha - x)]dx + 2\int_\alpha^\infty [\operatorname{Cov}(f_n(x), f_n(2\alpha - x))]dx.$$

The first term on the right-hand side of (2.10) is given by (see Rosenblatt (1971))

$$(2.11)\qquad \int_\alpha^\infty [\operatorname{Var} f_n(x)]dx = (nb)^{-1}\int w^2(z)dz + O(n^{-1}),$$

and the second term is

$$(2.12)\qquad \int_\alpha^\infty [\operatorname{Var} f_n(2\alpha - x)]dx = (nb)^{-1}\int_\alpha^\infty \int_{-\infty}^{-(x-\alpha)/b} w^2(z)f(2\alpha - x - bz)dzdx - n^{-1}\int_\alpha^\infty [Ef_n(2\alpha - x)]^2 dx.$$

It is easily shown that the first term in (2.12) is $O(n^{-1})$ and the second term is $O(bn^{-1})$, so that (2.12) is $O(n^{-1})$. Since

$$|\operatorname{Cov}(f_n(x), f_n(2\alpha - x))| \le [\operatorname{Var} f_n(x)]^{1/2}[\operatorname{Var} f_n(2\alpha - x)]^{1/2}$$

for all $x \ge \alpha$, it also follows that the third term in (2.10) is $O(n^{-1}b^{-1/2})$. Thus

$$(2.13)\qquad \int_\alpha^\infty [\operatorname{Var} \hat{f}_n(x)]dx = (nb)^{-1}\int w^2(z)dz + O(n^{-1}b^{-1/2}).$$

Relations (2.6), (2.9) and (2.13) lead to (2.5). □

If w is nonsymmetric, the first term on the right-hand side of (2.7) does not simplify to $f(x)$ for all x. One can easily show that this, in turn, causes (2.8) to be of order $O(b)$. Note that the symmetry of w has not been used in (2.13). If w is bounded then one can replace $O(n^{-1}b^{-1/2})$ in (2.13) and (2.5) by $O(n^{-1})$.

The reason is that the third term in (2.10) can then be written as $2n^{-1}(A_n - B_n)$, where

$$A_n = b^{-1} \int_{\alpha}^{\infty} \int_{-\infty}^{(x-\alpha)/b} w(z)w(z - 2(x-\alpha)b^{-1})f(x - bz)dzdx$$
$$\leq 4[\sup w(z)][\sup f(x)] \int |z|w(z)dz$$

and

$$B_n = \int_{\alpha}^{\infty} [Ef_n(x)][Ef_n(2\alpha - x)]dx \leq b[\sup f(x)]^2 \int |z|w(z)dz.$$

LEMMA 2.2. *Let $\mathcal{W}_o$ be the class of all functions w on $\mathbb{R}$ satisfying*

$$w(z) = w(-z) \geq 0, \qquad \int w(z)dz = 2\gamma, \qquad \int z^2 w(z)dz = 2\delta, \tag{2.14}$$

where γ and δ are given positive numbers. Then the functional

$$L(w) = \left[\int_0^{\infty} w^2(z)dz\right]^3 \left[\int_0^{\infty} \left\{\int_y^{\infty} [1 - W(z)]dz\right\}^2 dy\right]. \tag{2.15}$$

is minimized on $\mathcal{W}_o$ uniquely by

$$w_2^*(z) = \begin{cases} m\gamma(C-1)e^{m|z|}\{\cos(mz) - \sin(m|z|)\} & \\ \quad +m\gamma Ce^{-m|z|}\{\cos(mz) + \sin(m|z|)\} & \text{for } |z| \leq \pi/(2m) \\ 0 & \text{for } |z| > \pi/(2m), \end{cases} \tag{2.16}$$

where

$$C = (1 - e^{-\pi})^{-1}, \qquad m = (2\gamma/\delta)^{1/2}(e^{\pi/2} - e^{-\pi/2})^{-1/2}. \tag{2.17}$$

PROOF. For each $w \in \mathcal{W}_o$, define a function g on $\mathbb{R}^+$ by

$$g(z) = \int_z^{\infty} w(x)dx \quad \text{for} \quad z \geq 0, \tag{2.18}$$

and express (2.15) under (2.14) as

$$L(g) = \left[\int_0^{\infty} g'(z)^2 dz\right]^3 \left[\int_0^{\infty} \left\{\int_y^{\infty} g(z)dz\right\}^2 dy\right], \tag{2.19}$$

$$g'(z) \leq 0 \quad \text{for } z \geq 0, \qquad g(0) = \gamma, \qquad \int_0^{\infty} zg(z)dz = \delta/2, \tag{2.20}$$

where g' denotes the first derivative of g. Denote by $\mathcal{G}$ the class of all g on $\mathbb{R}^+$ satisfying (2.20). Since there is a one-to-one correspondence between $w \in \mathcal{W}_o$ and $g \in \mathcal{G}$, it suffices to show that the g corresponding to w_2^*, i.e.,

$$(2.21)\quad g_*(z) = \int_z^{\pi/(2m)} w_2^*(x)dx = \begin{cases} \gamma[Ce^{-mz} - (C-1)e^{mz}]\cos(mz) & \text{for } 0 \le z \le \pi/(2m) \\ 0 & \text{for } z > \pi/(2m) \end{cases}$$

minimizes L of (2.19) uniquely on $\mathcal{G}$. The calculus of variations involved in proving this is quite similar to that in the proof of Theorem 1 in Ghosh and Huang (1991) and we mention here the essential steps. It is easily shown that L is strictly convex on $\mathcal{G}$. Simple computations show that g_* indeed satisfies the constraints in (2.20) and $g_*'(\pi/2m) = 0$. Let $h \in (0,1)$ be a real number and define $R(z) = g(z) - g_*(z)$ for an arbitrary $g \in \mathcal{G}$. One gets from (2.19), by integrating by parts, the Gateaux differential as

$$(2.22)\quad \partial L(g_* + hR)/\partial h\,|_{h=0} = \int_0^\infty \left[-g_*''(z) + (A/3B)\int_0^z \int_y^\infty g_*(x)dxdy\right] R(z)dz,$$

where

$$(2.23)\quad A = \int_0^\infty g_*'(x)^2 dx, \qquad B = \int_0^\infty \left\{\int_y^\infty g_*(x)dx\right\}^2 dy.$$

It is easily verified that

$$A/(3B) = 4m^4, \qquad \int_0^{\pi/(2m)} \int_y^\infty g_*(x)dxdy = 1/4,$$

where m is defined in (2.17). Consequently,

$$(2.24)\quad -g_*''(z) + (A/3B)\int_0^z \int_y^\infty g_*(x)dxdy = \begin{cases} 0 & \text{for } 0 \le z \le \pi/(2m) \\ m^4 & \text{for } z > \pi/(2m). \end{cases}$$

It follows from (2.22) and (2.24) that

$$(2.25)\quad \begin{aligned} \partial L(g_* + hR)/\partial h\,|_{h=0} &= \int_{\pi/(2m)}^\infty \left[-g_*''(z) + (A/3B)\int_0^z \int_y^\infty g_*(x)dxdy\right] R(z)dz \\ &= m^4 \int_{\pi/(2m)}^\infty g(z)dz \ge 0 \quad \text{for every } g \in G. \end{aligned}$$

The first equation in (2.24) and the convexity of L show that g_* minimizes L uniquely among all $g \in \mathcal{G}$ with support in $[0, \pi/(2m)]$. The result in (2.25) globalizes the conclusion to $\mathcal{G}$. □

PROOF OF THEOREM 2.1. It follows from (1.6) and (2.5) that the bandwidth which asymptotically minimizes $\hat{R}_n$ is given by

$$b_n(f,w) = 6^{-1/4}[\{f'(\alpha)\}^2 + \{f'(\beta)\}^2]^{-1/4} \cdot \left[\int_0^\infty w^2(z)dz\right] [L(w)]^{-1/4} n^{-1/4}, \tag{2.26}$$

where $L(w)$ is defined in (2.15). Substituting (2.26) in (2.5) one gets

$$\min_{b_n} \hat{R}_n(f,b_n,w) = 2^{13/4}3^{-3/4}[f'(\alpha)^2 + f'(\beta)^2]^{1/4}[L(w)]^{1/4}n^{-3/4} + o(n^{-3/4}). \tag{2.27}$$

It follows from the remark after Lemma 2.1 that, for any nonsymmetric w, $\hat{R}_n(f,b_n,w)$ is $an^{-1/2} + o(n^{-1/2})$ for some $a > 0$, which is larger than (2.27) for sufficiently large n. Consequently, it suffices to restrict attention to symmetric w for the purpose of determining the optimal kernel. The latter problem is equivalent to minimizing $L(w)$ of (2.15) with respect to symmetric $w \in \mathcal{W}$. It follows from Lemma 2.2 with $\gamma = 1/2$ and $\delta = \tau^2/2$ that $L(w)$ is minimized, subject to $\tau_w = \tau$, uniquely by w_2^* of (2.3). Finally, substituting w_2^* in (2.26) one gets (2.4). □

It is easily verified that (2.15) is invariant in τ_w for any symmetric w. Substitution of (2.16) in (2.15) yields

$$L(w_2^*) = (27/262144)(e^\pi + 1)^4(e^\pi - 1)^{-4}. \tag{2.28}$$

3. Relative efficiency of $\hat{f}_n$ and f_n

Suppose $f \in \mathcal{F}(\alpha,\infty)$ or $\mathcal{F}(\alpha,\beta)$. Consider first the efficiency of $\hat{f}_n$ based on an arbitrary w relative to $\hat{f}_n$ based on w_2^*. An asymptotic measure of this relative efficiency is biven by

$$\lim_{n\to\infty} [\hat{R}_n(f,b_n(f,w_2^*),w_2^*)/\hat{R}_n(f,b_n(f,w),w)] = \begin{cases} [L(w_2^*)/L(w)]^{1/4} & \text{if } w \text{ is symmetric} \\ 0 & \text{if } w \text{ is nonsymmetric,} \end{cases} \tag{3.1}$$

where $L(w)$ is defined in (2.15) and $L(w_2^*)$ is given by (2.28). Theorem 2.1 shows that $L(w_2^*) < L(w)$ for any symmetric $w \neq w_2^*$. Table 1 shows values of $L(w)$ and (3.1) for some popular symmetric kernels (Epanechnikov (1969), Rosenblatt (1971)). It can be seen that, in practical terms, there is very little difference between w_2^* and the quadratic kernel in the table. This quadratic kernel happens to be the best choice w_o^* for f_n when $f \in \mathcal{F}(-\infty,\infty)$. Conversely, it is interesting to note here that there is very little difference between R_n under the kernels w_o^* and w_2^* when $f \in \mathcal{F}(-\infty,\infty)$ (e.g., in Epanechnikov's (1969) Table 1, $r = 1.004$

under w_2^*). Consider next the efficiency of f_n based on its optimum kernel w_1^* relative to $\hat{f}_n$ based on any symmetric w. It follows from (1.14) and (2.27) that the relative efficiency of f_n is asymptotically zero, in addition to the fact that $\hat{f}_n$ is globally consistent but f_n is not.

Table 1. Relative efficiency of some symmetric kernels for $\hat{f}_n$.

Kernel	$L(w)$	$[L(w_2^*)/L(w)]^{1/4}$
$w_2^*(z)$ of (2.3)	see (2.28)	
$w(z) = (\sqrt{6} - \lvert z\rvert)/6$ for $\lvert z\rvert \le \sqrt{6}$ †	1/6804	.9976
$w(z) = 3(4\sqrt{5})^{-1}(1 - z^2/5)$ for $\lvert z\rvert \le \sqrt{5}$ †	33/224000	.9970
$w(z) = (2\pi)^{-1/2}\exp(-z^2/2)$	$(\sqrt{2}-1)/192\sqrt{2}\pi^2$	.9851
$w(z) = (2\sqrt{3})^{-1}$ for $\lvert z\rvert \le \sqrt{3}$ †	1/5120	.9291
$w(z) = (\sqrt{2})^{-1}\exp(-\sqrt{2}\lvert z\rvert)$	1/4096	.8787

† $w(z) = 0$ otherwise.

In principle, one may think of using $\hat{f}_n$ of (2.2) to estimate an f with compact support $[\alpha, \beta]$ even when the latter is continuous at α and β. However, folding seems to be undesirable at a continuity point for two reasons. Suppose first that $f \in \mathcal{F}(-\infty, \infty)$, w is symmetric, and both f_n and $\hat{f}_n$ use the same w. The conditions underlying $\mathcal{F}(-\infty,\infty)$ imply $f(\alpha) = f'(\alpha) = f''(\alpha) = 0$ and $f(\beta) = f'(\beta) = f''(\beta) = 0$, using which it is easy to show that $|R_n - \hat{R}_n|$ is $o(b_n^4) + O(n^{-1})$. This means that folding will not affect the asymptotic adequacy of f_n for the family $\mathcal{F}(-\infty, \infty)$. Suppose next that $f(\alpha) = 0 = f(\beta)$ but $f'(\alpha) \ne 0$ or $f'(\beta) \ne 0$; such an f does not belong to $\mathcal{F}(-\infty,\infty)$, $\mathcal{F}(\alpha, \infty)$ or $\mathcal{F}(\alpha, \beta)$. One can then show that $\hat{R}_n$ is as in (2.5) and R_n is given by (2.5) with the factor 4 replaced by 2. This implies

$$\min_{b_n} \hat{R}_n(f, b_n, w) = 2^{1/4} \min_{b_n} R_n(f, b_n, w) + o(n^{-3/4}),$$

and therefore f_n is an asymptotically better estimate than $\hat{f}_n$. It follows from the results for $\hat{f}_n$ that, under the present conditions on f, the optimal kernel for f_n is w_2^* (and not w_o^* nor w_1^*).

4. An alternative definition of IMSE of f_n

Suppose that the support of f is $[\alpha, \beta]$, $-\infty \le \alpha < \beta \le \infty$, and consider

$$(4.1) \qquad \tilde{R}_n(f, b_n, w) = \int_\alpha^\beta E[f_n(x) - f(x)]^2 dx = \int_\alpha^\beta [Ef_n(x) - f(x)]^2 dx + \int_\alpha^\beta [\operatorname{Var} f_n(x)] dx.$$

Clearly, $R_n \ge \tilde{R}_n$ for each n because f_n need not be zero outside $[\alpha, \beta]$. In some situations, $\tilde{R}_n$ may be a more meaningful indicator of the IMSE of f_n than R_n.

One reason is that, if α and β are known, one would not really use f_n to estimate $f = 0$ outside $[\alpha,\beta]$ and, therefore, no loss is incurred by $f_n(x)$ for $x \notin [\alpha,\beta]$ (see Lehmann (1983), pp. 55–56, for a discussion). A second reason is that, for estimating $f \in \mathcal{F}(\alpha,\infty)$, f_n is uniformly consistent for f if and only if the support of w is in $\mathbb{R}^-$. This feature is not revealed by the minimization process of R_n (see (1.12) and (1.13)) whereas $\tilde{R}_n$ does exhibit this feature as noted below. Finally, in the context of using f_n for hypothesis testing, Bickel and Rosenblatt (1973) also ignored the loss in $f_n(x)$ for $x \notin [\alpha,\beta]$.

It is of some interest to know the asymptotically optimum kernel and bandwidth for f_n that minimize $\tilde{R}_n$. The relevant results follow from those under R_n by noting that

$$R_n(f,b_n,w) - \tilde{R}_n(f,b_n,w) = \int_{-\infty}^{\alpha} [Ef_n(x)]^2 dx + \int_{\beta}^{\infty} [Ef_n(x)]^2 dx + O(n^{-1})$$

and then analyzing the two integrals above under the appropriate assumptions of f. We summarize them without proofs which are similar to those in Rosenblatt (1971) and van Eeden (1985). First, if $f \in \mathcal{F}(-\infty,\infty)$, then w_o^* and $b_n(f,w_o^*)$ are also the optimum kernel and bandwidth under $\tilde{R}_n$. Second, if $f \in \mathcal{F}(\alpha,\infty)$ or $\mathcal{F}(\alpha,\beta)$, then we have

THEOREM 4.1. *Let $f \in \mathcal{F}(\alpha,\infty)$ and assume that* (1.2) *and* (1.3) *hold. Then, under the IMSE $\tilde{R}_n$, the optimum kernel and bandwidth for f_n are, repectively,*

$$(4.2) \qquad w_3^*(z) = \begin{cases} (z+3\tau\sqrt{2})/9\tau^2 & \text{for } -3\tau\sqrt{2} \le z \le 0 \\ 0 & \text{elsewhere,} \end{cases}$$

$$(4.3) \qquad b_n(f,w_3^*) = \tau^{-1}\left[9\sqrt{2}\int f'(x)^2 dx\right]^{-1/3} n^{-1/3},$$

where τ is an arbitrary positive number. Moreover,

$$(4.4) \qquad \tilde{R}_n(f,b_n(f,w_3^*),w_3^*) = \left[\frac{4}{3}\int f'(x)^2 dx\right]^{1/3} n^{-2/3} + o(n^{-2/3}).$$

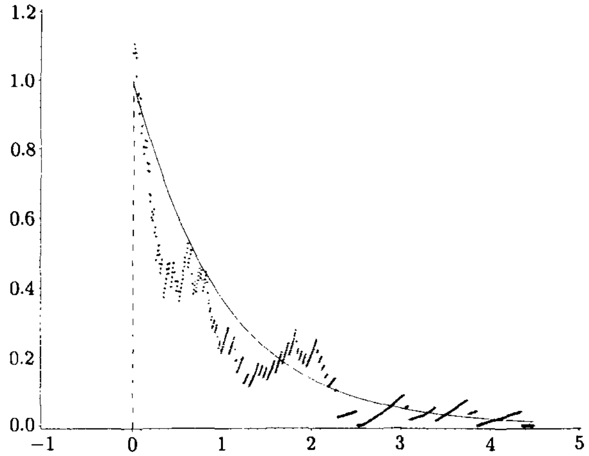
THEOREM 4.2. *Let $f \in \mathcal{F}(\alpha,\beta)$ and assume that* (1.2) *and* (1.3) *hold. Then, under the IMSE $\tilde{R}_n$, the optimum kernel and bandwidth for f_n are, respectively,*

$$(4.5) \quad w_4^*(z) = \begin{cases} [\tau(1+\rho)]^{-1}(1+\rho^2)^{1/2}\exp[-(1+\rho^2)^{1/2}z/\tau] & \text{for } z \ge 0 \\ [\tau(1+\rho)]^{-1}(1+\rho^2)^{1/2}\exp[(1+\rho^2)^{1/2}z/\rho\tau] & \text{for } z \le 0, \end{cases}$$

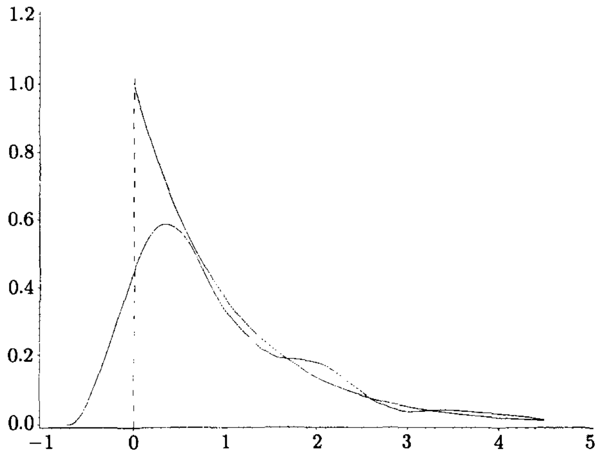
$$(4.6) \quad b_n(f,w_4^*) = (1+\rho^2)^{1/2}[\tau f(\alpha)]^{-1}n^{-1/2},$$

where $\rho = f(\alpha)/f(\beta)$ and τ is an arbitrary positive number. Moreover,

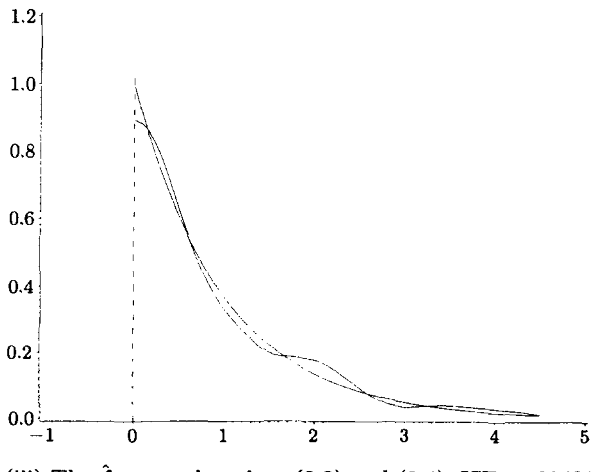
$$(4.7) \qquad \tilde{R}_n(f,b_n(f,w_4^*),w_4^*) = [f^{-1}(\alpha) + f^{-1}(\beta)]^{-1}n^{-1/2} + O(n^{-1}).$$



(i) The f_n-curve based on (4.2) and (4.3), ISE = .03492.



(ii) The f_n-curve based on (1.9) and (1.11), ISE = .06952.



(iii) The $\hat{f}_n$-curve based on (2.3) and (2.4), ISE = .00423.

Fig. 1. Three estimates of (5.1) using $n = 100$ observations.

If $f \in \mathcal{F}(\alpha, \infty)$, then (1.15) shows that f_n under every w with support in $\mathbb{R}^-$ is uniformly consistent for f; (4.2) effectively gives the best kernel among these w's under the criterion $\tilde{R}_n$. On the other hand, (1.12) is not related to the consistency property because it is constrained by the very choice of R_n as the IMSE. The best kernel among uniformly consistent f_n under the criterion R_n is, in fact, $w(z) = \tau^{-1} \exp(z/\tau)$ for $z \leq 0$.

In practical terms, Theorems 4.1 and 4.2 are moot because, when $f \in \mathcal{F}(\alpha, \infty)$ or $\mathcal{F}(\alpha, \beta)$, $\hat{f}_n$ is a superior estimator also under (4.1). If $f_n(x)$ in (4.1) is replaced by $\hat{f}_n(x)$ of (2.1) or (2.2), it is easily shown that Theorem 2.1 remains valid so that the right-hand side of (4.1) is of order $n^{-3/4}$.

5. Numerical results

The efficacy of f_n and $\hat{f}_n$ as global estimators of a discontinuous f can be best judged by estimating a known f over a range of x-values using actual samples from f. We consider the exponential density

$$f(x) = \begin{cases} \exp(-x) & \text{for } x \geq 0 \\ 0 & \text{elsewhere} \end{cases} \tag{5.1}$$

as an example of $\mathcal{F}(\alpha, \infty)$ to illustrate the essential features. A pseudo-random sample of size $n = 100$ was obtained by computer from the exponential distribution. Figure 1 shows the target curve (5.1) and three estimates: (i) the f_n-curve based on (4.2)–(4.3), (ii) the f_n-curve based on (1.9) and (1.11) and (iii) the $\hat{f}_n$-curve based on (2.3)–(2.4). Also shown are the observed integrated squared errors (ISE) $\int [f_n(x) - f(x)]^2 dx$ and $\int [\hat{f}_n(x) - f(x)]^2 dx$. In Fig. 1(i), the "local kinks" are caused by the discontinuity of w_3^* at 0 while the "global waves" are due to the one-sided nature of w_3^* (i.e., only observations in $[x, x + b_n]$ are contributing to $f_n(x)$). At the expense of inconsistency of $f_n(0)$, the local kinks are corrected in Fig. 1(ii) by the continuity of w_o^*. Figure 1(iii) shows, as a confirmation of the formal results of Section 2, that the folded estimator rectifies the drawbacks of f_n. Actual computations show that the f_n-curve based on (1.12) and (1.13) is indistinguishable from Fig. 1(ii), and the $\hat{f}_n$-curve based on (1.11) and (2.26) is indistinguishable from Fig. 1(iii).

Acknowledgements

The authors are grateful to the referees for several comments which helped improve the paper.

References

Apostol, T. M. (1957). *Mathematical Analysis*, Addison-Wesley, Reading, Massachusetts.

Bickel, P. J. and Rosenblatt, M. (1973). On some global measures of the deviations of density function estimates, *Ann. Statist.*, **1**, 1071–1095 (correction: ibid, (1975) **3**, 1370).

Cline, D. B. H. and Hart, J. D. (1990). Kernel estimates of densities with discontinuities or discontinuous derivatives, *Statistics*, **22**, 69–84.

Epanechnikov, V. A. (1969). Nonparametric estimates of a multivariate probability density, *Theory Probab. Appl.*, **14**, 153–158.

Ghosh, B. K. and Huang, W.-M. (1991). The power and optimal kernel of the Bickel-Rosenblatt test for goodness of fit, *Ann. Statist.*, **19**, 999–1009.

Johnson, N. L. and Kotz, S. (1970). *Distributions in Statistics: Continuous Univariate Distributions*, Houghton-Mifflin, Boston.

Lehmann, E. L. (1983). *Theory of Point Estimation*, Wiley, New York.

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

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.

Schuster, E. F. (1985). Incorporating support constraints into nonparametric estimates of densities, *Comm. Statist. Theory Methods*, **14**, 1123–1136.

Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*, Chapman, New York.

van Eeden, C. (1985). Mean integrated squared error of kernel estimators when the density and its derivative are not necessarily continuous, *Ann. Inst. Statist. Math.*, **37**, 461–472.