Ann. Inst. Statist. Math.
Vol. 48, No. 3, 451–467 (1996)

MINIMAX KERNELS FOR DENSITY ESTIMATION WITH BIASED DATA

COLIN O. WU AND ANDREW Q. MAO

Department of Mathematical Sciences, The Johns Hopkins University, Baltimore, MD 21218, U.S.A.

(Received August 5, 1994; revised September 25, 1995)

Abstract. This paper considers the asymptotic properties of two kernel estimates $\tilde{f}_n$ and $\hat{f}_n$, which have been proposed by Bhattacharyya *et al.* (1988, *Comm. Statist. Theory Methods*, **A17**, 3629–3644) and Jones (1991, *Biometrika*, **78**, 511–519), respectively, for estimating the underlying density f at a point under a general selection biased model. The asymptotic optimality of $\hat{f}_n$ and $\tilde{f}_n$ is measured by the corresponding asymptotic minimax mean squared errors under a compactly supported Lipschitz continuous family of the underlying densities. It is shown that, in general, $\hat{f}_n$ is a superior local estimate than $\tilde{f}_n$ in the sense that the asymptotic minimax risk of $\hat{f}_n$ is lower than that of $\tilde{f}_n$. The minimax kernels and bandwidths of $\hat{f}_n$ are computed explicitly and shown to have simple forms and depend on the weight functions of the model.

Key words and phrases: Kernel density estimate, minimax mean squared error, minimax kernel, bandwidth, weighted distribution, selection biased data.

1. Introduction

Let $X_1, X_2, \ldots, X_n$ be an i.i.d. sample with distribution function $G(x) = P(X_i \leq x)$, $x > 0$, $i = 1, \ldots, n$, and density g with respect to Lebesgue measure on the line. Then G is a weighted distribution or selection biased distribution if there is an underlying density f and a non-negative weight function w on $(0, \infty)$ such that

$$g(x) = w(x)f(x)/\mu, \qquad x > 0 \tag{1.1}$$

where $\mu = \int_0^\infty w(t)f(t)dt$. To avoid any identifiability problem for (1.1), we assume throughout that the weight function $w(x)$ is known. Theory and applications of (1.1) have been studied by Cox (1969), Patil *et al.* (1988), Patil and Taillie (1989), Vardi (1985), Ahmad (1995), among others.

In density estimation, two types of kernel estimates of $f(x)$,

$$\tilde{f}_n(x) = \frac{w^{-1}(x)}{nh_n\hat{\nu}_n} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h_n}\right) \tag{1.2}$$

and

$$\hat{f}_n(x) = \frac{1}{nh_n\hat{\nu}_n}\sum_{i=1}^{n} w^{-1}(X_i)K\left(\frac{x - X_i}{h_n}\right), \tag{1.3}$$

have been proposed by Bhattacharyya *et al.* (1988) and Jones (1991), respectively, where $h_n > 0$ is a bandwidth, $K(\cdot)$ a kernel function and

$$\hat{\nu}_n = \frac{1}{n}\sum_{i=1}^{n} w^{-1}(X_i). \tag{1.4}$$

In statistical literature, the asymptotic efficiency of a density estimate can be measured either by a Bayes criterion or by a minimax criterion. Under the minimax criteria, the efficiency of a density estimate can be described by its asymptotic minimax mean squared errors (MSE), or its global analogue, the asymptotic minimax integrated mean squared errors (IMSE). Studies of asymptotic minimaxity for kernel type estimates in the classical i.i.d. direct sample case can be found in Sacks and Ylvisaker (1981), Donoho and Liu (1991*b*), among others. For the special case of length bias model where $w(x) = x$, Wu (1995) has shown that, under a compactly supported Lipschitz continuous family with order one, that is,

$$\begin{aligned}\mathcal{F}(C, 1; x, a, b) = \{f : f(y) \equiv 0 \text{ if } y \notin [a, b], 0 \le a < x < b < \infty \text{ and}\\ |f(x) - f(y)| \le C|x - y| \text{ for all } y \in [a, b],\ \text{some } C > 0\},\end{aligned}$$

$\hat{f}_n(x)$ and $\tilde{f}_n(x)$ are locally equivalent in the sense that they have the same asymptotically minimax mean squared errors, and furthermore, the minimax kernel is a non-negative triangular kernel and the minimax bandwidth is a deterministic bandwidth which depends on $w(x) = x$ and is of order $n^{-1/3}$.

The aim of this paper is to generalize the approach of Wu (1995) to the general selection biased model (1.1) and a more general family of the underlying density functions. Specifically, we consider some compactly supported Lipschitz continuous families defined by

$$\begin{aligned}\mathcal{F}(C, \alpha; x, a, b) = \{f : f(y) \equiv 0 \text{ if } y \notin [a, b], 0 \le a < x < b < \infty \text{ and}\\ |f(x) - f(y)| \le C|x - y|^{\alpha}\\ \text{for all } y \in [a, b] \text{ and some } C > 0\}\end{aligned} \tag{1.5}$$

where $0 < \alpha \le 1$ is assumed to be a known constant, and study the asymptotic minimax mean squared errors of $\hat{f}_n(x)$ and $\tilde{f}_n(x)$ under $\mathcal{F}(C, \alpha; x, a, b)$. Denote $R(\hat{f}, f(x))$ to be the risk (MSE) of any estimate $\hat{f}(x)$ such that

$$R(\hat{f}, f(x)) = E_g[\hat{f}(x) - f(x)]^2. \tag{1.6}$$

The asymptotic properties of $\hat{f}_n(x)$ and $\tilde{f}_n(x)$ are measured by the large sample behavior of their minimax risks

$$\rho_n(\hat{f}; x) = \inf_{K, h_n} \sup_{f \in \mathcal{F}(C, \alpha; x, a, b)} R(\hat{f}_n, f(x))$$

and

$$\rho_n(\tilde{f};x) = \inf_{K,h_n} \sup_{f\in\mathcal{F}(C,\alpha;x,a,b)} R(\tilde{f}_n, f(x)),$$

respectively. The families defined in (1.5) include a large class of interesting densities which are less smooth than the twice differentiable family $\mathcal{F}$ considered by Jones (1991) and Ahmad (1995). Unlike the direct sample case, the compactness of the support of f ensures that the minimax mean squared errors of $\hat{f}_n$ actually exist for those $w(x)$ which might be unbounded when $x \to \infty$, such as the length biased model $w(x) = x$. Without the compactness on the support of f, one has to make additional assumptions such as $w(x)$ is bounded away from 0 and ∞ on $(0,\infty)$, or $\int_0^\infty w(x)f(x)dx < \infty$ and $\int_0^\infty w^{-1}(x)f(x)dx < \infty$ uniformly for all $f \in \mathcal{F}(C,\alpha;x,0,\infty)$, etc. Since most real life situations have compact supports, such compactness assumption should be mostly realistic.

The main results of this paper have several interesting features. First, the asymptotically minimax MSE of $\hat{f}_n$ is always lower than that of $\tilde{f}_n$. This theoretically justifies the heuristic impression that $\hat{f}_n$ is a locally better estimate than $\tilde{f}_n$ (see Jones (1991) or Ahmad (1995)). Second, since the optimality of $\hat{f}_n$ (or $\tilde{f}_n$) is measured by their mean squared errors, we assume here only the minimal conditions that the expectations and variances of $\hat{f}_n$ (or $\tilde{f}_n$) exist for all $f \in \mathcal{F}(C,\alpha;x,a,b)$ and $n \geq 1$. These conditions guarantee the existence of $R(\hat{f}_n, f(x))$ (or $R(\tilde{f}_n, f(x))$) for all $f \in \mathcal{F}(C,\alpha;x,a,b)$ and $n \geq 1$. Even though there are no prior restrictions on bandwidths h_n and kernels K, it is shown in Section 2 that the natural conditions on h_n and K as assumed in Jones (1991) and Ahmad (1995), such as $h_n \to 0$ and $nh_n \to \infty$ as $n \to \infty$ and $\int K(u)du = 1$, are necessary to ensure the consistency of $\hat{f}_n$ in the sense that $\lim_{n\to\infty} R(\hat{f}_n, f(x)) = 0$ for all $f \in \mathcal{F}(C,\alpha;x,a,b)$. However, the consistency of $\tilde{f}_n$ may require the value of $\int K(u)du$ to be other than 1. Third, unlike the usual differentiable families, we can not apply Taylor expansions to the less smooth family $\mathcal{F}(C,\alpha;x,a,b)$. Thus our derivations of the minimax kernels and bandwidths are based on the straightforward approach of step by step analysis of the least favorable risks of $\hat{f}_n$ and $\tilde{f}_n$. Fourth, the asymptotic minimax risks of $\hat{f}_n$ and $\tilde{f}_n$, and the choices of optimal kernels and bandwidths depend on the smoothness of $w(x)$. When $w(x)$ is continuous at x, our optimal kernels are symmetric about zero. When $w(x)$ is discontinuous at x, optimal kernels may be asymmetric about zero. Finally, the minimax kernels obtained in this paper have simple forms, and the optimal bandwidths are deterministic and computationally tractable. For practical purposes, it is also worthwhile to develop some data driven bandwidth selection procedures. However, to maintain our focus on asymptotic minimaxity, we omit the discussion of such procedures in this paper.

In Section 2, we give some useful notation and preliminary results. In Section 3, we formulate the minimax estimation procedures of $\hat{f}_n$ and $\tilde{f}_n$, compute their asymptotic minimax risks, and derive the corresponding minimax kernels and bandwidths. In Section 4, we present some examples and compare the relative efficiency of different kernels. Finally, the proofs of the main results of Section 3 are deferred to Section 5.

2. Preliminary results

Throughout we make the following assumptions on $w(\cdot)$:

ASSUMPTION A. The weight functions $w(\cdot)$ are strictly positive and uniformly bounded away from 0 and ∞ on $[a,b]$, i.e. for each $[a,b]$, there exist universal constants $0 < m(a,b) < M(a,b) < \infty$ such that

$$m(a,b) \le \inf_{t\in[a,b]} w(t) \le \sup_{t\in[a,b]} w(t) \le M(a,b).$$

ASSUMPTION B. For each $w(\cdot)$, there exists a neighborhood $(x-\delta, x+\delta)$ of x such that $w(t)$, $t \in (x-\delta, x+\delta)$, has right and left limits $w(x_+)$ and $w(x_-)$ at x, i.e.

$$\lim_{t\downarrow x} w(t) = w(x_+) \qquad \text{and} \qquad \lim_{t\uparrow x} w(t) = w(x_-).$$

Suppose that $\nu = \int w^{-1}(t)g(t)dt$ is known. Then the natural analogues of $\hat{f}_n$ and $\tilde{f}_n$ are given by, respectively,

$$\hat{f}_n^*(x) = \frac{1}{nh_n\nu}\sum_{i=1}^{n} w^{-1}(X_i)K\left(\frac{x-X_i}{h_n}\right) \tag{2.1}$$

and

$$\tilde{f}_n^*(x) = \frac{w^{-1}(x)}{nh_n\nu}\sum_{i=1}^{n} K\left(\frac{x-X_i}{h_n}\right). \tag{2.2}$$

The main effort of this section is to show that $\hat{f}_n$ and $\tilde{f}_n$ are asymptotically equivalent to $\hat{f}_n^*$ and $\tilde{f}_n^*$, respectively. Thus it suffices only to study the asymptotic behaviors of $R(\hat{f}_n^*, f(x))$ and $R(\tilde{f}_n^*, f(x))$.

Let $\mathrm{B}(\hat{f}, f(x))$ and $\mathrm{V}(\hat{f}, f(x))$ be the bias and variance of any estimate $\hat{f}(x)$, respectively. By routine computation, $\nu = \mu^{-1}$ and the change of variables, the risk of $\hat{f}_n^*(x)$ is given by

$$R(\hat{f}_n^*, f(x)) = \mathrm{B}^2(\hat{f}_n^*, f(x)) + \mathrm{V}(\hat{f}_n^*, f(x)) \tag{2.3}$$

where

$$\begin{aligned}\mathrm{B}(\hat{f}_n^*, f(x)) &= \int K(u)f(x-h_nu)du - f(x),\\ \mathrm{V}(\hat{f}_n^*, f(x)) &= \frac{\mu}{nh_n}\int w(x-h_nu)^{-1}K^2(u)f(x-h_nu)du\\ &\quad - \frac{1}{n}\left(\int K(u)f(x-h_nu)du\right)^2.\end{aligned}$$

Similarly, the risk of $\tilde{f}_n^*(x)$ is given by

$$R(\tilde{f}_n^*, f(x)) = \mathrm{B}^2(\tilde{f}_n^*, f(x)) + \mathrm{V}(\tilde{f}_n^*, f(x)) \tag{2.4}$$

where

$$\mathrm{B}(\tilde{f}_n^*, f(x)) = w^{-1}(x) \int K(u)w(x - h_n u)f(x - h_n u)du - f(x),$$

$$\mathrm{V}(\tilde{f}_n^*, f(x)) = \frac{w^{-2}(x)\mu}{nh_n} \int K^2(u)w(x - h_n u)f(x - h_n u)du - \frac{w^{-2}(x)}{n} \left(\int K(u)w(x - h_n u)f(x - h_n u)du \right)^2 .$$

The next theorem summarizes the necessary conditions on K and h_n for the consistency of $\hat{f}_n(x)$ and $\tilde{f}_n(x)$, and the asymptotically equivalence between $\hat{f}_n(x)$, $\tilde{f}_n(x)$ and their corresponding $\hat{f}_n^*(x)$, $\tilde{f}_n^*(x)$. Its proof can be obtained by modifying the method in the proofs of Theorem 2.1 and Lemma 3.1 of Wu (1995), hence is omitted for brevity. The details of the proof can be found in Wu and Mao (1994) and Wu (1994).

THEOREM 2.1. (A) *Suppose that* $\mathrm{B}(\hat{f}_n, f(x))$ *and* $\mathrm{V}(\hat{f}_n, f(x))$ *exist for all* $n \geq 1$, *and*

$$\lim_{n\to\infty} R(\hat{f}_n, f(x)) = 0 \quad \textit{for all} \quad f \in \mathcal{F}(C, \alpha; x, a, b), \tag{2.5}$$

then the following hold:

(a) $\lim_{n\to\infty} h_n = 0$ *and* $\lim_{n\to\infty} nh_n = \infty$,
(b) $K(\cdot)$ *is integrable on* $(-\infty, \infty)$ *and* $\int K(u)du = 1$,
(c) *when* n *is sufficiently large*,

$$\sup_{f\in\mathcal{F}(C,\alpha;x,a,b)} R(\hat{f}_n, f(x)) = \sup_{f\in\mathcal{F}(C,\alpha;x,a,b)} R(\hat{f}_n^*, f(x)) + o\left(\sup_{f\in\mathcal{F}(C,\alpha;x,a,b)} R(\hat{f}_n^*, f(x)) \right).$$

Furthermore, if

$$\lim_{n\to\infty} \sup_{f\in\mathcal{F}(C,\alpha;x,a,b)} n^{2\alpha/(2\alpha+1)} R(\hat{f}_n^*, f(x)) = \lambda \quad \textit{for some} \quad 0 \leq \lambda < \infty, \tag{2.6}$$

then

(d) $h_n = O(n^{-1/(2\alpha+1)})$ *for sufficiently large* n,
(e) $\int |K(u)||u|^\alpha du < \infty$ *and* $\int K^2(u)du < \infty$.

(B) *If* $\hat{f}_n$ *and* $\hat{f}_n^*$ *in* (2.5) *and* (2.6) *were replaced by* $\tilde{f}_n$ *and* $\tilde{f}_n^*$, *then* (a), (c), (d) *and* (e) *still hold, but* (b) *is replaced by*

(b′) $\lim_{n\to\infty} \int_{(x-b)/h_n}^{(x-a)/h_n} K(u)w(x - h_n u)du = w(x)$.

Remark 2.1. It is well-known that $n^{-2\alpha/(2\alpha+1)}$ is the optimal convergence rate for kernel density estimates with i.i.d. direct samples (cf. Stone (1980), Donoho and Liu (1991*a*)). We will see later in Theorem 3.1 that $n^{-2\alpha/(2\alpha+1)}$ is also the best attainable convergence rate for kernel density estimates with biased samples. A clear implication of Theorem 2.1 is that, for the selection of optimal kernels and bandwidths, it suffices to eliminate any $K(\cdot)$ and h_n which do not satisfy (a), (b) (or (b$'$)), (d) and (e) of Theorem 2.1.

Remark 2.2. Contrary to the usual condition of $\int_{-\infty}^{\infty} K(u)du = 1$ in kernel density estimation, Theorem 2.1(B) shows that, in certain cases where $w(\cdot)$ is discontinuous at x, the consistency of $\tilde{f}_n(x)$ may require that $\int_{-\infty}^{\infty} K(u)du \neq 1$. For example, suppose that $w(x) = 2$ and $w(t) = 1$ if $t \neq x$, then Theorem 2.1(a) and (b$'$) imply that

$$\lim_{n\to\infty} \int_{(x-b)/h_n}^{(x-a)/h_n} K(u)du = \int_{-\infty}^{\infty} K(u)du = 2.$$

3. Minimax risks

This section is devoted to deriving the minimax risks of $\hat{f}_n$ and $\tilde{f}_n$. From Remark 2.1, it suffices to consider (K, h_n) which satisfies (a), (b) (or (b$'$)), (d) and (e) of Theorem 2.1. Here, we first compute the asymptotic form of the least favorable risk $\sup_{f\in\mathcal{F}(C,\alpha;x,a,b)} R(\hat{f}_n^*, f(x))$, and then derive the minimax kernels and bandwidths for $\hat{f}_n$. The asymptotic minimax risk of $\hat{f}_n$ is shown to be a lower bound of the minimax risk $\inf_{K,h_n} \sup_{f\in\mathcal{F}(C,\alpha;x,a,b)} R(\tilde{f}_n^*, f(x))$.

LEMMA 3.1. (A) *If* (2.5) *and* (2.6) *are satisfied, then there exist constants* p_0 *and* μ_0 *such that*

$$0 < p_0\mu_0 = \sup_{f\in\mathcal{F}(C,\alpha;x,a,b)} f(x)\mu < \infty$$

and, for sufficiently large n,

$$\begin{aligned}(3.1)\quad & \sup_{f\in\mathcal{F}(C,\alpha;x,a,b)} R(\hat{f}_n^*, f(x)) \\ &= h_n^{2\alpha}C^2\left[\int |K(u)||u|^\alpha du\right]^2 + o(h_n^{2\alpha}) + o(n^{-1}h_n^{-1}) \\ &\quad + \frac{p_0\mu_0}{nh_n}\left[w^{-1}(x_+)\int_{-\infty}^0 K^2(u)du + w^{-1}(x_-)\int_0^\infty K^2(u)du\right].\end{aligned}$$

(B) *If* (2.5) *and* (2.6) *are satisfied with* $\hat{f}_n$ *and* $\hat{f}_n^*$ *replaced by* $\tilde{f}_n$ *and* $\tilde{f}_n^*$, *then, for sufficiently large* n,

$$(3.2)\quad \sup_{f\in\mathcal{F}(C,\alpha;x,a,b)} R(\tilde{f}_n^*, f(x)) \geq \Gamma(h_n, K) + o(h_n^{2\alpha}) + o(n^{-1}h_n^{-1}),$$

where

$$\Gamma(h_n, K) = h_n^{2\alpha} C^2 \left[\frac{w(x_+)}{w(x)} \int_{-\infty}^{0} |K(u)||u|^\alpha du + \frac{w(x_-)}{w(x)} \int_0^\infty |K(u)||u|^\alpha du \right] + \frac{p_0\mu_0}{nh_n} \left[\frac{w(x_+)}{w^2(x)} \int_{-\infty}^{0} K^2(u) du + \frac{w(x_-)}{w^2(x)} \int_0^\infty K^2(u) du \right].$$

Remark 3.1. For some special cases, the value of $p_0\mu_0$ can be computed explicitly. However, when there are more complicated weight functions involved, explicit solutions of $p_0\mu_0$ may be difficult to obtain. Thus we have to compute $p_0\mu_0$ numerically. A special density sequence which may be used to compute the value of $p_0\mu_0$ is given in (5.1) of Section 5.

It is now clear from Theorem 2.1 and Lemma 3.1 that the asymptotically minimax risks of $\hat{f}_n$ can be derived by finding a sequence of bandwidths and kernels which minimize the dominating terms of the right hand side of (3.1). Similarly, minimizing the leading term $\Gamma(h_n, K)$ at the right hand side of (3.2) with respect to some (h_n, K), we can obtain a lower bound of the minimax risks of $\tilde{f}_n(x)$.

For the minimax risks of $\hat{f}_n(x)$, we define a kernel K_{opt} and bandwidth h_{opt} such that

$$K_{opt}(u) = \begin{cases} \dfrac{2w(x_+)}{w(x_+) + w(x_-)} \left[1 - \left(\dfrac{2\alpha}{\alpha+1} \right)^\alpha |u|^\alpha \right] & \text{if } -\dfrac{\alpha+1}{2\alpha} \le u < 0, \\ \dfrac{2w(x_-)}{w(x_+) + w(x_-)} \left[1 - \left(\dfrac{2\alpha}{\alpha+1} \right)^\alpha |u|^\alpha \right] & \text{if } 0 \le u \le \dfrac{\alpha+1}{2\alpha}, \\ 0 & \text{otherwise}, \end{cases} \tag{3.3}$$

and

$$h_{opt} = n^{-1/(2\alpha+1)} (p_0\mu_0)^{1/(2\alpha+1)} C^{-2/(2\alpha+1)} \left[\frac{2}{w(x_+) + w(x_-)} \right]^{1/(2\alpha+1)} \times (\alpha+1)^{-2\alpha/(2\alpha+1)} (2\alpha)^{2\alpha/(2\alpha+1)} (2\alpha+1)^{1/(2\alpha+1)}. \tag{3.4}$$

For the lower bound of the minimax risks of $\tilde{f}_n(x)$, we consider the same bandwidth h_{opt} as defined in (3.4) and a new kernel K_{opt*} such that

$$K_{opt*}(u) = \begin{cases} \dfrac{2w(x)}{w(x_-) + w(x_+)} \left[1 - \left(\dfrac{2\alpha}{\alpha+1} \right)^\alpha |u|^\alpha \right] & \text{if } |u| \le \dfrac{\alpha+1}{2\alpha}, \\ 0 & \text{otherwise}. \end{cases} \tag{3.5}$$

Since K_{opt} and K_{opt*} have the same support on a compact interval, it is easy to verify that they satisfy the necessary conditions of Theorem 2.1. The conclusions of Lemma 3.1 lead to the following main results of this section.

THEOREM 3.1. *The kernel K_{opt} and the bandwidth h_{opt} as defined in (3.3) and (3.4) are asymptotically minimax for $\hat{f}_n(x)$. Furthermore, the asymptotic minimax risks of $\hat{f}_n(x)$ and $\tilde{f}_n(x)$ satisfy*

$$\begin{aligned}
(3.6) \qquad & \lim_{n\to\infty} n^{2\alpha/(2\alpha+1)} \rho_n(\tilde{f}; x) \\
& \geq \lim_{n\to\infty} n^{2\alpha/(2\alpha+1)} \Gamma(h_{opt}, K_{opt*}) \\
& = \lim_{n\to\infty} n^{2\alpha/(2\alpha+1)} \rho_n(\hat{f}; x) \\
& = \lim_{n\to\infty} n^{2\alpha/(2\alpha+1)} \sup_{f\in\mathcal{F}(C,\alpha;x,a,b)} R(\hat{f}_{K_{opt},h_{opt}}, f(x)) \\
& = C^{2/(2\alpha+1)} (p_0\mu_0)^{2\alpha/(2\alpha+1)} [(2\alpha)^{-2\alpha/(2\alpha+1)} + (2\alpha)^{1/(2\alpha+1)}] \\
& \quad \times \left[\frac{2}{w(x_+) + w(x_-)}\right]^{2\alpha/(2\alpha+1)} \\
& \quad \times (2\alpha+1)^{-1-(1/(2\alpha+1))} (\alpha+1)^{2\alpha/(2\alpha+1)}.
\end{aligned}$$

Remark 3.2. The minimax kernels and bandwidths of $\hat{f}_n(x)$ are not unique. In fact, the asymptotic least favorable risk

$$\lim_{n\to\infty} n^{2\alpha/(2\alpha+1)} \sup_{f\in\mathcal{F}(C,\alpha;x,a,b)} R(\hat{f}_n, f(x))$$

attains the right hand side of (3.6) if and only if for any constant $\tau > 0$ the kernel and bandwidth of $\hat{f}_n(x)$ are given by

$$K_\tau(u) = \begin{cases} w(x_+)(\tau - \tau_*|u|^\alpha) & \text{if } -\dfrac{\alpha+1}{\alpha\tau(w(x_+)+w(x_-))} \leq u < 0, \\ w(x_-)(\tau - \tau_*|u|^\alpha) & \text{if } 0 \leq u \leq \dfrac{\alpha+1}{\alpha\tau(w(x_+)+w(x_-))}, \\ 0 & \text{otherwise,} \end{cases}$$

and

$$\begin{aligned}
h(\tau) &= n^{-1/(2\alpha+1)} \left(\frac{p_0\mu_0}{2\alpha C^2}\right)^{1/(2\alpha+1)} \\
&\quad \times \left[\frac{w^{-1}(x_+)\int_{-\infty}^0 K_\tau^2(u)du + w^{-1}(x_-)\int_0^\infty K_\tau^2(u)du}{(\int K_\tau(u)|u|^\alpha du)^2}\right]^{1/(2\alpha+1)}
\end{aligned}$$

where $\tau_* = \tau^{\alpha+1}[\alpha(w(x_+)+w(x_-))]^\alpha(\alpha+1)^{-\alpha}$. In particular, $(K_\tau, h(\tau))$ reduces to (K_{opt}, h_{opt}) if $\tau = 2[w(x_+)+w(x_-)]^{-1}$. The above minimax kernels and bandwidths depend on the weight function $w(\cdot)$. For most of the interesting cases where $w(x)$ is continuous at x, these minimax kernels are symmetric about zero. However, contrary to the frequently used kernels in the i.i.d. direct sample case, when $w(x)$ is discontinuous at x, the minimax kernels depend on the values of $w(x_+)$ and $w(x_-)$ and should be asymmetric about zero.

Remark 3.3. A direct consequence of Theorem 3.1 is that $\hat{f}_n(x)$ is generally a better local estimate than $\tilde{f}_n(x)$ under the minimax MSE criterion. It is also straightforward to conclude that the best possible convergence rate for $\hat{f}_n(x)$ is $n^{-2\alpha/(2\alpha+1)}$, and this rate is attained by (K_{opt}, h_{opt}). However, Theorem 3.1 only gives a pointwise result in the sense that the risk is taken to be the MSE. For global properties, further study is needed to investigate the global minimax risks of $\hat{f}_n$ and $\tilde{f}_n$ such as the minimax integrated mean squared errors, minimax L_1-norms or minimax L_∞-norms.

Remark 3.4. It is interesting to know whether the lower bound of (3.6) can be always attained by some proper choices of kernels and bandwidths for $\tilde{f}_n(x)$. For the special case of length biased model with family $\mathcal{F}(C, 1; x, a, b)$, Wu (1995) showed that the lower bound is attained by using a triangular kernel of the form $K(u) = (1 - |u|)1_{[|u| \le 1]}$ and a bandwidth sequence $h_n = n^{-1/3}x^{-1/3}C^{-2/3} \cdot (3p_0\mu_0)^{1/3}$. However, such kind of attainability does not hold in general. In some cases, $\hat{f}_n(x)$ is strictly superior than $\tilde{f}_n(x)$. To see this, consider the problem of estimating $f_0(0)$ where f_0 is a uniform density on $[-1/2, 1/2]$. Let $w(t) = \max\{0, 1 - |t|^{\alpha/2}\}$ be the weight function. Then, by Theorem 3.1, we know that the asymptotic lower bound of (3.6) can only be attained by kernels which are non-negative and have compact supports. Now for any compactly supported non-negative kernel K satisfying (2.5), direct computation based on (2.4) shows that, when n is sufficiently large, $\int_{-1/2h_n}^{1/2h_n} K(u)du = 1$ and $\inf_{h_n} R(\tilde{f}_n^*, f_0(x)) = O(n^{-\alpha/(1+\alpha)})$. Thus the asymptotic minimax risk of $\tilde{f}_n(x)$ must be strictly larger than that of $\hat{f}_n(x)$.

4. Comparison of kernels

It is well-known in the i.i.d. direct sample case that most commonly used kernels such as the uniform kernel, the Gaussian kernel and the Epanechnikov kernel are nearly optimal for kernel density estimates (cf. Silverman (1986)). Thus it is natural to ask that whether this phenomenon still holds in the biased sampling case. The numerical results of this section show that these popular kernels are asymptotically suboptimal when $w(\cdot)$ is discontinuous at x. Because of Remark 3.3, the discussion here is limited to $\hat{f}_n$ only.

Let $\hat{f}_{K,h_n}(x)$ be the kernel estimate $\hat{f}_n(x)$ with any particular choice of (K, h_n). Define

$$e_{(K,h_n),(K_{opt},h_{opt})} = \frac{\lim_{n\to\infty} \sup_{f\in\mathcal{F}(C,\alpha;x,a,b)} R(\hat{f}_{K_{opt},h_{opt}}, f(x))}{\lim_{n\to\infty} \sup_{f\in\mathcal{F}(C,\alpha;x,a,b)} R(\hat{f}_{K,h_n}, f(x))} \tag{4.1}$$

to be the the asymptotic relative efficiency of $\hat{f}_{K,h_n}(x)$ with respect to the minimax estimate $\hat{f}_{K_{opt},h_{opt}}(x)$. If the pair (K, h_n) does not satisfy (a), (b), (d) and (e) of Theorem 2.1, then by Theorem 2.1(c) and Theorem 3.1, it automatically follows that

$$e_{(K,h_n),(K_{opt},h_{opt})} = 0.$$

If (K, h_n) satisfies (a), (b), (d) and (e) of Theorem 2.1, then the optimal bandwidth h_{opt} as defined in (3.4) should be used. Substituting h_{opt} back to the dominating term at the right hand side of (3.1), we have

$$e_{K,K_{opt}} = \left[\frac{q(K_{opt})}{q(K)}\right]^{2/(2\alpha+1)}. \tag{4.2}$$

where $e_{K,K_{opt}} = e_{(K,h_{opt}(K)),(K_{opt},h_{opt})}$ and

$$q(K) = \left[w^{-1}(x_+)\int_{-\infty}^{0} K^2(u)du + w^{-1}(x_-)\int_{0}^{\infty} K^2(u)du\right]^{\alpha} \times \left[\int K(u)|u|^{\alpha}du\right]. \tag{4.3}$$

Let $\alpha = 1$, $\beta = w(x_-)/w(x_+)$ and K_1 through K_3 be the following kernels:

1. Uniform: $K_1(u) = (1/2)1_{[-1\le u\le 1]}$,
2. Epanechnikov: $K_2(u) = (3/4\sqrt{5})(1-u^2/5)1_{[|u|\le\sqrt{5}]}$,
3. Gaussian: $K_3(u) = (2\pi)^{-1/2}\exp\{-u^2/2\}$.

The relative efficiency $e_{K,K_{opt}}$ for various β values are computed in Table 1 which shows that all three kernels considered here have very high $e_{K,K_{opt}}$ values when β is close to one, but low $e_{K,K_{opt}}$ values when β is away from one.

Table 1. Relative efficiency $e_{K,K_{opt}}$ when $\alpha = 1$.

β	0.1	0.5	1.0	5.0
K_{opt}	1.000	1.000	1.000	1.000
K_1	0.442	0.855	0.925	0.625
K_2	0.474	0.917	0.992	0.670
K_3	0.474	0.917	0.991	0.670

5. Proofs

This section gives the proofs of Lemma 3.1.(A) and Theorem 3.1. Since the proofs for $\tilde{f}_n$ can be obtained by modifying the calculation for $\hat{f}_n$, the details for $\tilde{f}_n$ can be found in Wu (1994), and are omitted here for brevity.

First, we give a technical lemma which is useful for the proofs of Lemma 3.1 and Theorem 3.1. Define a density $f_p \in \mathcal{F}(C, \alpha; x, a, b)$ such that

$$f_p(t) = \begin{cases} \max\{0, p - C|x-t|^{\alpha}\} & \text{if } t \in A_1[W(p)] \cup A_{01}[W(p), T(p)], \\ p + C|x-t|^{\alpha} & \text{if } t \in A_2[W(p)] \cup A_{02}[W(p), T(p)], \\ 0 & \text{otherwise,} \end{cases} \tag{5.1}$$

where

$$\begin{aligned}
A_1[W(p)] &= \{t : w(t) < W(p), t \in [a,b]\},\\
A_2[W(p)] &= \{t : w(t) > W(p), t \in [a,b]\},\\
A_0[W(p)] &= \{t : w(t) = W(p), t \in [a,b]\},\\
A_{01}[W(p),T(p)] &= \{t : t \in A_0[W(p)] \text{ and } t \le T(p)\},\\
A_{02}[W(p),T(p)] &= \{t : t \in A_0[W(p)] \text{ and } t > T(p)\},
\end{aligned}$$

and $(p, W(p), T(p))$ satisfies the technical conditions $0 \le p$, $a \le T(p) \le b$, $m(a,b) \le W(p) \le M(a,b)$, $\int_a^b f_p(t)dt = 1$. To ensure that f_p is a density, p has to be bounded above by a constant. Let

$$p_1 = \begin{cases} 0 & \text{if } \displaystyle\int_a^b C|x-t|^\alpha dt \ge 1,\\ p_* & \text{if } \displaystyle\int_a^b C|x-t|^\alpha dt < 1, \end{cases}$$

where $p_* = \max\{0, (b-a)^{-1}(1 - \int_a^b C|t-x|^\alpha dt)\}$, and p_2 be the unique solution of

$$\int_a^b \max\{0, p - C|t-x|^\alpha\}dt = 1.$$

It is easy to see that the constraint of $f_p \in \mathcal{F}(C,\alpha;x,a,b)$ implies $p_1 \le p \le p_2$. For any $f_p(t)$, define $\mu_p = \int w(t) f_p(t)dt$.

LEMMA 5.1. *If $w(\cdot)$ satisfies Assumptions* A *and* B, *then*
(a) *μ_p is a continuous function of p,*
(b) *there exist a constant $c > 0$ and a point p_0 such that*

$$\max[p_1, c] \le p_0 \le p_2 \quad \text{and} \quad p_0\mu_0 = \sup_{p_1 \le p \le p_2} p\mu_p = \sup_{f \in \mathcal{F}(C,\alpha;x,a,b)} f(x)\mu, \tag{5.2}$$

where $\mu_0 = \int w(t) f_{p_0}(t)dt$.

PROOF. (a) For any $p_0 \in (p_1, p_2)$ and $\epsilon > 0$, we may assume that without loss of generality $p = p_0 + \epsilon$. The analysis for $p = p_0 - \epsilon$ is similar. It suffices to show that, for any $\delta > 0$, we can select ϵ sufficiently small such that $|\mu_p - \mu_{p_0}| \le \delta$.

Since $p \ge p_0$, the technical condition of $\int f_p(t)dt = \int f_{p_0}(t)dt = 1$ implies that $W(p) \ge W(p_0)$. Let

$$\begin{aligned}
\mathcal{A}_1(p) &= A_1[W(p)] \cup A_{01}[W(p),T(p)] \quad \text{and}\\
\mathcal{A}_2(p) &= A_2[W(p)] \cup A_{02}[W(p),T(p)].
\end{aligned}$$

Consequently, we have that

$$\mathcal{A}_1(p_0) \subseteq \mathcal{A}_1(p) \quad \text{and} \quad \mathcal{A}_2(p) \subseteq \mathcal{A}_2(p_0). \tag{5.3}$$

If $t \in \mathcal{A}_1(p_0)$, then by (5.1) and (5.3)

$$0 \leq f_p(t) - f_{p_0}(t) = \max\{0, p - C|x-t|^\alpha\} - \max\{0, p_0 - C|x-t|^\alpha\} \leq \epsilon;$$

and if $t \in \mathcal{A}_2(p)$, then $0 \leq f_p(t) - f_{p_0}(t) = p - p_0 = \epsilon$. Thus, it follows that

$$0 \leq f_p(t) - f_{p_0}(t) \leq \epsilon \quad \text{if} \quad t \in \mathcal{A} \tag{5.4}$$

where $\mathcal{A} = \mathcal{A}_1(p_0) \cup \mathcal{A}_2(p)$.

If $t \in [a,b] \setminus \mathcal{A}$ and $f_p(t) - f_{p_0}(t) > 0$, then (5.1) implies that $f_{p_0}(t) = p_0 + C|x-t|^\alpha$ and $f_p(t) = \max\{0, p - C|x-t|^\alpha\}$, hence,

$$|f_p(t) - f_{p_0}(t)| \leq \epsilon. \tag{5.5}$$

Thus, there exists some constant $c_1 > 0$, such that

$$\left| \int_{[a,b]\setminus\mathcal{A}} (f_p(t) - f_{p_0}(t)) 1_{[f_p(t) - f_{p_0}(t) > 0]} dt \right| \leq c_1 \epsilon. \tag{5.6}$$

Similarly, by (5.4), (5.6) and the technical condition of $\int f_p(t)dt = \int f_{p_0}(t)dt = 1$, we can verify that

$$\int_{[a,b]\setminus\mathcal{A}} |f_p(t) - f_{p_0}(t)| 1_{[f_p(t) - f_{p_0}(t) < 0]} dt \leq c_2 \epsilon \quad \text{for some} \quad c_2 > 0. \tag{5.7}$$

Combining (5.4) through (5.7), we have

$$|\mu_p - \mu_{p_0}| = \left| \int w(t)(f_p(t) - f_{p_0}(t))dt \right| \leq M(a,b)c_3\epsilon$$

for some $c_3 > 0$. Thus, (a) holds.

(b) By the definition of p_1 and p_2, we have $p_1 \leq f(x) \leq p_2$ for all $f \in \mathcal{F}(C,\alpha;x,a,b)$. Thus, it suffices to only consider those densities with $p_1 \leq f(x) \leq p_2$.

Now, for all $f \in \mathcal{F}(C,\alpha;x,a,b)$ such that $f(x) = p \in [p_1,p_2]$, the problem of maximizing $f(x)\mu$ is equivalent to maximizing μ subject to $f(x) = p$ and $f \in \mathcal{F}(C,\alpha;x,a,b)$. We first show that $\mu_p = \sup_{f\in\mathcal{F}(C,\alpha;x,a,b), f(x)=p} \mu$. Let $f_{1p}(t) = \max\{0, p - C|x-t|^\alpha\}$ and $f_{2p}(t) = p + C|x-t|^\alpha$. Then, (5.1) shows that $f_p(t) = f_{1p}(t)1_{[t\in\mathcal{A}_1(p)]} + f_{2p}(t)1_{[t\in\mathcal{A}_2(p)]}$. Since $f_{1p}(t) \leq f(t) \leq f_{2p}(t)$ holds for all $f \in \mathcal{F}(C,\alpha;x,a,b)$, it can be seen from direct integration that $\mu_p \geq \mu$ for all $f \in \mathcal{F}(C,\alpha;x,a,b)$ with $f(x) = p$.

Next, for each $p_1 \leq p \leq p_2$, (a) implies that $f_p(x)\mu_p = p\mu_p$ is a continuous function of p. Furthermore, it is obvious that $\sup_{p_1\leq p\leq p_2} p\mu_p > 0$. Thus p_0 must be bounded below from 0, that is, $p_0 \geq c$ for some $c > 0$. Now, since $p\mu_p$ is a continuous function in p on the closed interval $[\max[p_1,c],p_2]$, there exists a point p_0 such that $p_0\mu_0 = \sup_{p\in[\max[p_1,c],p_2]} p\mu_p$. This completes the proof. □

PROOF OF LEMMA 3.1(A). The first assertion is proved in Lemma 5.1, so it is only necessary to prove (3.1). We first establish that the right hand side of (3.1) is an upper bound. By Assumptions A and B, and Theorem 2.1, we have the following inequalities

$$\sup_{f\in\mathcal{F}(C,\alpha;x,a,b)}\left|\int K(u)f(x-h_nu)du-f(x)\right|\le Ch_n^\alpha\int|K(u)||u|^\alpha du, \tag{5.8}$$

and, for any $s>0$,

$$\begin{aligned}
&\sup_{f\in\mathcal{F}(C,\alpha;x,a,b)}\frac{\mu}{nh_n}\int w^{-1}(x-h_nu)K^2(u)f(x-h_nu)du \\
&\quad=\sup_{f\in\mathcal{F}(C,\alpha;x,a,b)}\frac{\mu}{nh_n}\left[\int_{-s}^{s}w^{-1}(x-h_nu)K^2(u)f(x-h_nu)du\right.\\
&\qquad+\int_{-\infty}^{-s}w^{-1}(x-h_nu)K^2(u)f(x-h_nu)du\\
&\qquad\left.+\int_{s}^{\infty}w^{-1}(x-h_nu)K^2(u)f(x-h_nu)du\right]\\
&\quad\le\sup_{f\in\mathcal{F}(C,\alpha;x,a,b)}\frac{\mu}{nh_n}\left[w^{-1}(x_+)f(x)\int_{-s}^{0}K^2(u)du\right.\\
&\qquad+w^{-1}(x_-)f(x)\int_0^sK^2(u)du\\
&\qquad+o(1)+m^{-1}(a,b)\sup_{t\in[a,b]}\{f(t)\}\\
&\qquad\left.\times\left(\int_{-\infty}^{-s}K^2(u)du+\int_s^\infty K^2(u)du\right)\right].
\end{aligned}\tag{5.9}$$

Since $s>0$ can be selected arbitrarily large, (5.9), Theorem 2.1 and Lemma 5.1 imply that $\int_{-\infty}^{-s}K^2(u)du$ and $\int_s^\infty K^2(u)du$ can be arbitrarily small. Thus,

$$\begin{aligned}
&\sup_{f\in\mathcal{F}(C,\alpha;x,a,b)}\frac{\mu}{nh_n}\int w^{-1}(x-h_nu)K^2(u)f(x-h_nu)du\\
&\quad\le\frac{p_0\mu_0}{nh_n}\left[w^{-1}(x_+)\int_{-\infty}^0K^2(u)du+w^{-1}(x_-)\int_0^\infty K^2(u)\right]\\
&\qquad+o(n^{-1}h_n^{-1}).
\end{aligned}\tag{5.10}$$

Then (5.8) and (5.10) imply that the right hand side of (3.1) is an upper bound.

To see that the right hand side of (3.1) is also a lower bound, we need to show that it is attained asymptotically by a particular density sequence. By Lemma 5.1, there exists $p_0>0$ such that $W(p_0)>0$, $T(p_0)>0$, and f_{p_0} satisfies (5.1). Thus, for any $s>0$ and sufficiently large n, we can define a density sequence $\{f_{*n}\}$ such that

$$f_{*n}(t)=\eta_n(t)\left(\int\eta_n(t)dt\right)^{-1}\tag{5.11}$$

where

$$\eta_n(t) = \begin{cases} p_n + C|x-t|^\alpha & \text{if } t \in [x - h_n s, x + h_n s] \text{ and } K[(x-t)/h_n] > 0, \\ p_n - C|x-t|^\alpha & \text{if } t \in [x - h_n s, x + h_n s] \text{ and } K[(x-t)/h_n] \leq 0, \\ p_n + C|x-t|^\alpha & \text{if } t \in [a,b] \setminus [x - h_n s, x + h_n x] \\ & \text{and } t \in A_2[W(p_0)] \cup A_{02}[W(p_0), T(p_0)], \\ \gamma_n(t) & \text{if } t \in [a,b] \setminus [x - h_n s, x + h_n s] \\ & \text{and } t \in A_1[W(p_0)] \cup A_{01}[W(p_0), T(p_0)], \\ 0 & \text{otherwise,} \end{cases}$$

$$\gamma_n(t) = \max\{2Ch_n^\alpha s^\alpha, p_n - C|x-t|^\alpha\} \quad \text{and} \quad p_n = p_0 + 2Ch_n^\alpha s^\alpha.$$

Comparing $\eta_n(t)$ with f_{p_0} defined in (5.1), it is straightforward to verify that $\int \eta_n(t)dt \geq 1$ for all sufficiently large n, and $\lim_{n\to\infty} \int \eta_n(t)dt = 1$. Hence, f_{*n} is a density in $\mathcal{F}(C, \alpha; x, a, b)$ when n is sufficiently large. Moreover, we can also verify that $\lim_{n\to\infty} f_{*n}(x) = p_0$ and $\lim_{n\to\infty} \mu_{f_{*n}} = \mu_0$.

By Assumption A and Theorem 2.1, we can take $s > 0$ sufficiently large so that the following inequalities hold for any $\epsilon > 0$ and $f \in \mathcal{F}(C, \alpha; x, a, b)$,

$$\left| \int_{|u|>s} K(u)(f(x - h_n u) - f(x))du \right| \leq Ch_n^\alpha \epsilon,$$

and

$$\int_{|u|>s} w^{-1}(x - h_n u) K^2(u) f(x - h_n u) du \leq \epsilon.$$

Thus, when n is sufficiently large, (5.11) and the definition of η_n imply that

$$\begin{aligned} \left| \int K(u) f_{*n}(x - h_n u)du - f_{*n}(x) \right| &= \left| \int K(u)(f_{*n}(x - h_n u) - f_{*n}(x))du \right| \\ &\geq Ch_n^\alpha \left[\int_{-s}^{s} |K(u)||u|^\alpha du - \epsilon \right] \\ &\geq Ch_n^\alpha \left[\int |K(u)||u|^\alpha du - 2\epsilon \right], \end{aligned} \tag{5.12}$$

and

$$\begin{aligned} &\frac{\mu_{f_{*n}}}{nh_n} \int w^{-1}(x - h_n u) K^2(u) f_{*n}(x - h_n u) du \\ &\geq \frac{\mu_{f_{*n}}}{nh_n} \left[\int_{-s}^{s} w^{-1}(x - h_n u) K^2(u) f_{*n}(x - h_n u) du - \epsilon \right] \\ &\geq \frac{p_0 \mu_0}{nh_n} \left[w^{-1}(x_+) \int_{-\infty}^{0} K^2(u)du + w^{-1}(x_-) \int_{0}^{\infty} K^2(u)du - \epsilon \right] \\ &\quad + o(n^{-1}h_n^{-1}). \end{aligned} \tag{5.13}$$

Hence, by (5.12) and (5.13), the right hand side of (3.1) is also a lower bound. □

PROOF OF THEOREM 3.1. By Lemma 3.1, it suffices to find K and h_n which satisfy (a), (b), (d) and (e) of Theorem 2.1, to minimize the following dominating term of the right hand side of (3.1)

$$Q(K,h_n) = h_n^{2\alpha}C^2\left[\int |K(u)||u|^\alpha du\right]^2 + \frac{p_0\mu_0}{nh_n}\left[w^{-1}(x_+)\int_{-\infty}^0 K^2(u)du + w^{-1}(x_-)\int_0^\infty K^2(u)du\right].$$

Let $K(\cdot)$ be any kernel such that $\int K(u)du = 1$. There exists a set $\mathcal{A}_+ \subset (-\infty,\infty)$ such that $K(u) \geq 0$ if $u \in \mathcal{A}_+$ and $\int_{\mathcal{A}_+} K(u)du = 1$. Let $K_+(u) = K(u)1_{[u\in\mathcal{A}_+]}$ to be a non-negative kernel. Then it is easy to see that $K_+(\cdot)$ satisfies (a), (b), (d) and (e) of Theorem 2.1, and $Q(K,h_n) \geq Q(K_+,h_n)$. Thus in order to minimize $Q(K,h_n)$ it is only necessary to consider non-negative kernels. For the rest of the proof, we assume that K is non-negative.

For any fixed K, the corresponding optimal bandwidth $h_{opt}(K)$ can be obtained through the solution of $\partial Q(K,h_n)/\partial h_n = 0$. Routine computation then shows that

$$\text{(5.14)} \quad h_{opt}(K) = n^{-1/(2\alpha+1)}\left(\frac{p_0\mu_0}{2\alpha C^2}\right)^{1/(2\alpha+1)} \times \left[\frac{w^{-1}(x_+)\int_{-\infty}^0 K^2(u)du + w^{-1}(x_-)\int_0^\infty K^2(u)du}{(\int K(u)|u|^\alpha du)^2}\right]^{1/(2\alpha+1)}.$$

Substituting (5.14) back to $Q(K,h_n)$, we have that

$$\text{(5.15)} \quad Q(K,h_{opt}(K)) = n^{-2\alpha/(2\alpha+1)}C^{2/(2\alpha+1)}(p_0\mu_0)^{2\alpha/(2\alpha+1)} \times [(2\alpha)^{-2\alpha/(2\alpha+1)} + (2\alpha)^{1/(2\alpha+1)}]q^{2/(2\alpha+1)}(K),$$

where

$$q(K) = \left(w^{-1}(x_+)\int_{-\infty}^0 K^2(u)du + w^{-1}(x_-)\int_0^\infty K^2(u)du\right)^\alpha \times \left(\int K(u)|u|^\alpha du\right).$$

Let $K_\delta(u) = K(u) + \delta\phi(u)$, $\delta \geq 0$, be another non-negative kernel such that $\int K_\delta(u)du = 1$. Thus $\int \phi(u)du = 0$ and $\phi(u) \geq 0$ when u is outside the support of $K(u)$. The Lagrangian of $q(K_\delta)$ is given by

$$L(K_\delta) = q(K_\delta) + d\left(\int K_\delta(u)du - 1\right)$$

for some constant d. Since $q(K_\delta)$ is a convex function of δ, there is a minimizer K of $q(K)$ such that it is a solution of

$$\lim_{\delta\downarrow 0}\delta^{-1}(L(K_\delta)-L(K))\geq 0 \tag{5.16}$$

for all $\phi(u)$ such that $\int\phi(u)du=0$ and $K_\delta(\cdot)$ is non-negative.

Direct computation shows that

$$\begin{aligned}&\lim_{\delta\downarrow 0}\delta^{-1}(L(K_\delta)-L(K))\\&\quad=\int_{-\infty}^{0}[(2w^{-1}(x_+)\lambda_1\lambda_2)K(u)+\lambda_3|u|^\alpha+d]\phi(u)du\\&\qquad+\int_0^\infty[(2w^{-1}(x_-)\lambda_1\lambda_2)K(u)+\lambda_3|u|^\alpha+d]\phi(u)du\end{aligned}$$

where

$$\lambda_1=\alpha\left[w^{-1}(x_+)\int_{-\infty}^{0}K^2(u)du+w^{-1}(x_-)\int_0^\infty K^2(u)du\right]^{\alpha-1},$$

$$\lambda_2=\int K(u)|u|^\alpha du$$

and

$$\lambda_3=\left[w^{-1}(x_+)\int_{-\infty}^{0}K^2(u)du+w^{-1}(x_-)\int_0^\infty K^2(u)du\right]^{\alpha}.$$

Hence any solution K of (5.16) has the form

$$\begin{aligned}K(u)&=w(x_+)[a_1+a_2|u|^\alpha]1_{[a_1+a_2|u|^\alpha>0,u\leq 0]}\\&\quad+w(x_-)[a_1+a_2|u|^\alpha]1_{[a_1+a_2|u|^\alpha>0,u>0]}\end{aligned}$$

where $a_1=-d/(2\lambda_1\lambda_2)$ and $a_2=-\lambda_3/(2\lambda_1\lambda_2)$. The constraint of $\int K(u)du=1$ then implies that

$$2(w(x_+)+w(x_-))\int(a_1+a_2|u|^\alpha)1_{[a_1+a_2|u|^\alpha>0,u\leq 0]}du=1$$

and

$$a_2=a_1^{\alpha+1}\left(\frac{\alpha(w(x_+)+w(x_-))}{\alpha+1}\right)^\alpha.$$

Let $a_1=\tau$. It is easy to see that any solution of (5.16) is given by K_τ as defined in Remark 3.2 for some τ. It is easy to verify that, for all $\tau>0$, K_τ satisfies (a), (b), (d) and (e) of Theorem 2.1.

Finally, substituting K_τ back to $q(K)$, we get

$$\left(w^{-1}(x_+)\int_{-\infty}^{0}K_\tau^2(u)du+w^{-1}(x_-)\int_0^\infty K_\tau^2(u)du\right)^\alpha=\frac{2^\alpha\tau^\alpha\alpha^\alpha}{(2\alpha+1)^\alpha}$$

and

$$\int K_\tau(u)|u|^\alpha du = \frac{(\alpha+1)^\alpha}{\alpha^\alpha(w(x_+)+w(x_-))^\alpha \tau^\alpha(2\alpha+1)}.$$

Substituting the above equations back to (5.15), we see that

$$q(K_\tau) = \left(\frac{2}{w(x_+)+w(x_-)}\right)^\alpha (\alpha+1)^\alpha (2\alpha+1)^{-\alpha-1}$$

which is independent of τ, and that $n^{2\alpha/(2\alpha+1)}Q(K_{opt}, h_{opt})$ has the desired form given in the right hand side of (3.6). The proof for $\tilde{f}_n$ can be obtained by the same method. □

Acknowledgements

The authors would like to thank Professor Ibrahim A. Ahmad for sending us copies of his work, and two referees for many insightful comments.

REFERENCES

Ahmad I. A. (1995). On multivariate kernel estimation for samples from weighted distributions, *Statist. Probab. Lett.*, **22**, 121–129.

Bhattacharyya, B. B., Franklin, L. A. and Richardson, G. D. (1988). A comparison of nonparametric unweighted and length-biased density estimation of Fibres, *Comm. Statist. Theory Methods*, **A 17**, 3629–3644.

Cox, D. R. (1969). Some sampling problems in technology, *New Developments in Survey Sampling* (eds. N. L. Johnson and H. Smith, Jr.), 506–527, Wiley-Interscience, New York.

Donoho, D. L. and Liu, R. C. (1991*a*). Geometrizing rates of convergence II, *Ann. Statist.*, **19**, 633–667.

Donoho, D. L. and Liu, R. C. (1991*b*). Geometrizing rates of convergence III, *Ann. Statist.*, **19**, 668–701.

Jones, M. C. (1991). Kernel density estimation for length biased data, *Biometrika*, **78**, 511–519.

Patil, G. P. and Taillie, C. (1989). Probing encountered data, meta analysis and weighted distribution methods, *Statistical Data Analysis and Inference* (ed. Y. Dodge), 317–346. North-Holland, Amsterdam.

Patil, G. P., Rao, C. R. and Zelen, M. (1988). Weighted distributions, *Encyclopedia of Statistical Sciences, 9* (eds. S. Kotz and N. L. Johnson), 565–571, Wiley, New York.

Sacks, J. and Ylvisaker, D. (1981). Asymptotically optimum kernels for density estimation at a point, *Ann. Statist.*, **9**, 334–346.

Silverman, B. W. (1986). Density estimation for statistics and data analysis, *Monographs Statist. Appl. Probab.*, **26**, Chapman and Hall.

Stone, C. J. (1980). Optimal convergence rates for nonparametric estimators, *Ann. Statist.*, **8**, 1348–1360.

Vardi, Y. (1985). Empirical distributions in selection bias models (with discussions), *Ann. Statist.*, **13**, 178–205.

Wu, C. O. (1994). On comparison of kernel estimates of density at a point with selection biased data, Tech. Report, No. 536, Department of Mathematical Sciences, The Johns Hopkins University, Maryland.

Wu, C. O. (1995). Minimax kernel density estimators with length biased data, *Mathematical Methods of Statistics*, **4**, 56–80.

Wu, C. O. and Mao, A. Q. (1994). Minimax kernels for Jones type density estimators with selection biased data, Tech. Report, No. 535, Department of Mathematical Sciences, The Johns Hopkins University, Maryland.