The Annals of Statistics
1990, Vol. 18, No. 2, 806–831

FOURIER METHODS FOR ESTIMATING MIXING DENSITIES AND DISTRIBUTIONS[1]

BY CUN-HUI ZHANG

State University of New York at Stony Brook and Rutgers University

Let $X_1, X_2, \ldots$ be iid observations from a mixture density $f(x) = \int f(x|\theta)\,dG(\theta)$, where $f(x|\theta)$ is a known parametric family of density functions and G is an unknown distribution function. This paper concerns estimating the mixing density $g = G'$ and the mixing distribution G. Fourier methods are used to derive kernel estimators, upper bounds for their rates of convergence and lower bounds for the optimal rate of convergence. Sufficient conditions are given under which the kernel estimators are asymptotically normal. Our estimators achieve the optimal rate of convergence $(\log n)^{-1/2}$ for the normal family and $(\log n)^{-1}$ for the Cauchy family.

1. Introduction. Let $\{f(x|\theta),\ \theta \in \Theta\}$ be a known parametric family of probability density functions with respect to a σ-finite measure μ. The density function $f(x)$ of a random variable X belongs to a mixture model if

$$f(x) = f(x;G) = \int f(x|\theta)\,dG(\theta), \tag{1}$$

where G is a probability distribution function. This paper concerns estimating the mixing density function $g = dG/d\theta$ and the mixing distribution function G based on iid observations $X_1, \ldots, X_n$ with common marginal density $f(x)$.

We consider kernel estimators for mixing densities and distributions. A sequence of functions $K_n(x)$ is a kernel for a functional $\lambda = \lambda_G$ of G if $\lim_n E_G K_n(X) = \lambda_G$ under suitable conditions. A kernel estimator for λ is basically the average of iid random variables $n^{-1}\sum_{j=1}^n K_n(X_j)$.

EXAMPLE 1 (Normal). Suppose that $f(x|\theta) = \varphi(x-\theta)$, $\varphi(x) = (2\pi)^{-1/2}e^{-x^2/2}$. Then $e^{t^2/2}Ee^{itX}$ is the characteristic function of G, so that a kernel for the mixing density $g = G'$ can be obtained by making use of the Fourier inversion formula.

In Section 2 Fourier methods are used to derive kernel estimators for mixing densities and upper bounds for their rates of convergence. In Section 3 kernels for mixing distributions are obtained by integrating kernels for mixing densities. Asymptotical normality of the kernel estimators is obtained under conditions on the growth of the second moment of the kernels.

Received January 1987; revised November 1988.
[1]Research supported by NSF Grant DMS-8857774.
AMS 1980 *subject classifications*. Primary 62G05; secondary 62G20, 62E20.
Key words and phrases. Mixing distribution, kernel estimation, contiguity, Fourier transformation.

Lower bounds for the optimal rate of convergence are given in Section 4. Let $g_n = G_n'$ be a sequence of mixing density functions such that the sequences $\{\prod_{j=1}^n f(x_j; G_n)\}$ and $\{\prod_{j=1}^n f(x_j; G)\}$ are contiguous. Then the order of $g_n - g$, $g = G'$, can serve as a lower bound for the rate of convergence for estimating the mixing density. Our problem is to find "bad" contiguous sequences and quantify the rate of $g_n - g$. To motivate our method, let us consider

EXAMPLE 1 (Normal, continuation). Let $g_0(\theta) = G_0'(\theta) = \varphi((\theta - \mu)/\sigma)/\sigma$ be a normal mixing density. Then $f(x|\theta)g_0(\theta)/f(x; G_0)$ is a normal density (in θ) with variance $\sigma^2/(1 + \sigma^2)$ for every fixed x, so that

$$\left|\int \exp[id_n\theta] f(x|\theta) g_0(\theta)\, d\theta\right| = \exp\left[-d_n^2\sigma^2/(2(1+\sigma^2))\right] f(x; G_0).$$

Let $g_n = g_0 + (w/d_n)(\cos(d_n\theta) - e_n)g_0$, where $0 < w < 1/2$, $d_n^2\sigma^2/(1+\sigma^2) = \log n$ and the constant e_n is chosen such that g_n is a probability density function. Then it can be shown that $|f(x; G_n)/f(x; G_0) - 1| \leq 1/\sqrt{n}$, so that the sequences $\{\prod_{j=1}^n f(x_j; G_n)\}$ and $\{\prod_{j=1}^n f(x_j; G_0)\}$ are contiguous. Therefore, $1/d_n \sim (\log n)^{-1/2}$, the order of $g_n - g_0$, is a lower bound for the optimal rate of convergence for the class of G with a bounded second derivative, if we are interested in estimating mixing densities. This idea of using Fourier methods to construct contiguous sequences will be developed in detail under a smoothness condition on the family $f(x|\theta)$.

We shall consider location families

$$\begin{aligned} &f(x|\theta) = f_0(x - \theta), \qquad \mu(dx) = dx, \qquad \theta \in (-\infty, \infty), \\ &f_0^*(t) = \int e^{itx} f_0(x)\, dx \neq 0, \qquad \forall\, t, \end{aligned} \tag{2}$$

throughout the paper. Some of our arguments can be extended to exponential families, but these extensions are incomplete and omitted to save space. For the location case, both the rate of convergence for our kernel estimators and the lower bound for the optimal rate of convergence are essentially given in terms of the tail of the characteristic function f_0^* of f_0. The rates are slower for f_0^* with thinner tails. Since smooth density functions have characteristic functions with thin tails, the optimal rates of convergence are related to the smoothness of the family $f(x|\theta)$ with slower rates corresponding to smoother families. See the discussions after Theorems 1 and 6. For the sake of space, we only consider the rates for the largest class of G with minimum smoothness conditions to which our methods are directly applicable. Three examples, the normal, Cauchy and double exponential cases, are considered in each section. The rates for these examples are summarized in Table 1. Our kernel estimators have the optimal rate of convergence $(\log n)^{-1/2}$ for the normal case and $(\log n)^{-1}$ for the Cauchy case.

Problems related to mixture models were proposed by Robbins (1951, 1955) in connection with the empirical Bayes approach to compound decision prob-

TABLE 1
Lower and upper bounds for the rates of convergence

		Estimating density g		Estimating distribution G	
Family	f_0^*	**Rate achieved**	**Lower bound**	**Rate achieved**	**Lower bound**
Normal	e^{-t^2}	$(\log n)^{-1/2}$	$(\log n)^{-1/2}$	$(\log n)^{-1/2}$	$(\log n)^{-1/2}$
Cauchy	$e^{-\|t\|}$	$(\log n)^{-1}$	$(\log n)^{-1}$	$(\log n)^{-1}$	$(\log n)^{-1}$
Double Exp.	$(1+t^2)^{-1}$	$n^{-1/7}$	$n^{-1/6}$	$n^{-1/5}$	$n^{-1/4}$

lems, by Kiefer and Wolfowitz (1956) in connection with estimating an unknown parameter in the presence of infinitely many nuisance parameters and by many others in various contexts. Estimating mixing densities and/or mixing distributions has been studied by Deely and Kruse (1968), Jewell (1982), Kiefer and Wolfowitz (1956), Laird (1978), Lindsay (1983a, b), Robbins (1950) and Rolph (1968) among others. Various consistent estimators have been proposed, but their rates of convergence and asymptotic distributions are unknown. Recently, Edelman (1988) proposed an estimator for the "empirical" mixing distribution function for the normal case and gave an upper bound $(\log n)^{-1/4}$ for its rate of convergence. The relationship between a part of this work and some of Edelman's results is discussed in Section 3. Estimating mixing densities and the optimal rate of convergence were also studied independently by Stefanski and Carroll (1988) and Carroll and Hall (1988).

Throughout this paper we denote by $P = P_G$ and $E = E_G$ the probability and expectation corresponding to G, respectively, by H^* the characteristic function of any distribution function H and by h^* the Fourier transformation of any integrable function h, so that $h^*(t) = \int e^{itx}h(x)\,dx$ for any $\int|h(x)|dx < \infty$.

2. Kernel estimation for mixing densities. In this section we give kernel estimators for mixing density functions and their rates of convergence in the L^2 norm $\|h\| = (\int|h(x)|^2\,dx)^{1/2}$ and in the mean square error at fixed $\theta = a$. The asymptotic normality at fixed $\theta = a$ is also given under a condition on the growth of the variance of the kernel.

A sequence of functions $K_n(x,a)$ is a kernel for $g(\theta) = G'(\theta) = dG/d\theta$ at $\theta = a$ if

$$(3) \qquad \int_{a-\delta}^{a+\delta}\int K_n(x,a)\,f(x|\theta)\mu(dx)\,d\theta \to 1 \quad \text{as } n\to\infty, \forall\, 0<\delta<\infty.$$

Under certain regularity conditions on G, (3) implies that $EK_n(X,a) \to g(a)$, so that the statistic

$$(4) \qquad \hat{g}_n(a) = \left[n^{-1}\sum_{j=1}^{n} K_n(X_j,a)\right]^+, \qquad x^+ = \max(x,0),$$

can be used to estimate $g(a)$. As $\hat{g}_n(a)$ is the nonnegative part of a sum of iid random variables, we can bound the mean square error $E[\hat{g}_n(a) - g(a)]^2$ (and therefore the rate of convergence) if we can find upper bounds for $\mathrm{Var}(K_n(X, a))$ and $|EK_n(X, a) - g(a)|$. It is clear by (3) that the choice of $K_n(x, a)$ depends on the family $f(x|\theta)$ and the measure μ.

Consider the location case (2). Similar to the normal case, Example 1 in Section 1, the expectation of the random variable $e^{itX}/f_0^*(t)$ is the characteristic function $g^*(t)$ of G. By the Fourier inversion formula,

$$(2\pi)^{-1}\int k^*(t/c)e^{-ita}g^*(t)\,dt \to g(a) \quad \text{as } c \to \infty$$

for any smooth integrable function k with $k^*(0) = 1$, so that we can define a kernel by

$$K_n(x, a) = (2\pi)^{-1}\int \mathbb{R}\{k^*(t/c_n)e^{it(x-a)}/f_0^*(t)\}\,dt, \qquad 0 < c_n \uparrow \infty, \tag{5}$$

where $\mathbb{R}\{z\}$ is the real part of z and $k(x)$ is a probability density function such that

$$k(x) = k(-x), \qquad \int x^2 k(x)\,dx < \infty, \qquad \int |xk'(x)|dx < \infty, \tag{6}$$
$$k^*(t) = 0, \qquad \forall\, |t| > 1.$$

Since $k^*(t/c_n)$ is the Fourier transformation of $c_n k(c_n x)$,

$$\begin{aligned}\int K_n(x, a) f(x|\theta)\mu(dx) &= \mathbb{R}\left\{\int\int (2\pi f_0^*(t))^{-1} k^*(t/c_n)e^{it(x-a)}f_0(x-\theta)\,dx\,dt\right\} \\ &= (2\pi)^{-1}\mathbb{R}\left\{\int k^*(t/c_n)e^{it(\theta-a)}\,dt\right\} = c_n k(c_n(a-\theta)),\end{aligned} \tag{7}$$

which implies (3). Hence, K_n is a kernel. There are many density functions satisfying (6). For example, $k(x) = (6/\pi)|(2/x)\sin(x/4)|^4$, whose characteristic function is proportional to the density function of the average of four independent unif$(-1, 1)$ random variables with $k^*(0) = 1$. In the sequel we shall always use the notation

$$C_1 = \int |x|k(x)\,dx, \qquad C_2 = \int x^2 k(x)\,dx, \qquad C_3 = \int |k^*(t)|^2\,dt,$$

$$\varepsilon(x) = \int_x^\infty yk(y)\,dy + \sup_{y \ge x} y^2 k(y).$$

The constants c_n in (5) should be chosen to "balance" the bias and the variance of the kernel estimators. In view of (7) we shall choose large c_n to ensure small bias $EK_n(X, a) - g(a)$. On the other hand, if c_n is too large, then by (5) the variance $\mathrm{Var}(K_n(X, a))$ is hard to control, since $f_0^*(t) \to 0$ as

$t \to \infty$. The properties of our kernel estimators are studied in Theorems 1 and 2, where the constants c_n and upper bounds for the risks are expressed in terms of the tail of $|f_0^*(t)|$. Let us first consider the L^2 norm. Define

$$\mathscr{G}_1 = \mathscr{G}_1(M_1) = \{G: G \text{ has a density } g \text{ with } \|tg^*(t)\| \le M_1\}. \tag{8}$$

THEOREM 1. *Let the estimator $\hat{g}_n$ for the mixing density $g = G'$ be defined by* (4) *with the kernel $K_n(x,a)$ given by* (5) *and constants c_n satisfying*

$$1/n \le c_0 |f_0^*(c_n)|^2 / c_n^3, \qquad c_0 < \infty, \qquad |f_0^*(c_n)| = \min_{|t| \le c_n} |f_0^*(t)|. \tag{9}$$

Then

$$\lim_{n\to\infty} E\|\hat{g}_n - g\|^2 = 0, \qquad \forall\, G \textit{ with a density } g \textit{ such that } \|g\| < \infty, \tag{10}$$

$$\sup_{G \in \mathscr{G}_1(M_1)} c_n^2 E_G \|\hat{g}_n - g\|^2 \le \left[c_0 C_3 + (M_1 C_1)^2\right]/(2\pi), \qquad \forall\, n \ge 1. \tag{11}$$

Clearly, the smoother the density f_0 is, the faster $|f_0^*(t)|$ vanishes as $t \to \infty$ and by (9) the slower the upper bound $c_n^{-2}O(1)$ in (11) for the risk $E_G\|\hat{g}_n - g\|^2$ converges to 0. By (9) we see that the constants c_n are of polynomial order in n (Example 3) if $f_0^*(t)$ is of polynomial order in t for large t and c_n are of logarithmic order (Examples 1 and 2) if $f_0^*(t)$ is of exponential order. As far as the rate of convergence is concerned, the choice of $k(\cdot)$ and c_0 does not matter. Of course, one can choose them to minimize the right-hand side of (11). But this requires the knowledge of M_1 or a good estimation of $\|tg^*(t)\|$. We shall not discuss the problem of estimating the "best" $k(\cdot)$ and/or c_0 here.

EXAMPLE 1 (Normal, continuation). Since $f(x|\theta) \sim N(\theta, 1)$, we have $f_0^*(t) = e^{-t^2/2}$, so that

$$K_n(x,a) = (2\pi)^{-1} \int \cos(t(x-a)) e^{t^2/2} k^*(t/c_n)\, dt.$$

Choose, $c_n = \sqrt{\alpha \log n}$, $0 < \alpha < 1$. It follows that

$$\sup_{G \in \mathscr{G}_1(M_1)} (\log n) E_G \|\hat{g}_n - g\|^2 \le \left[C_3 \sqrt{\alpha}\, n^{\alpha-1} (\log n)^{3/2} + C_1^2 M_1^2/\alpha\right]/(2\pi).$$

EXAMPLE 2 (Cauchy). Suppose $f_0(x) = (1+x^2)^{-1}/\pi$. Let $c_n = \alpha \log n$, $0 < \alpha < 1$. Then

$$\sup_{G \in \mathscr{G}_1(M_1)} (\log n)^2 E_G \|\hat{g}_n - g\|^2 \le \left[C_3 \alpha n^{\alpha-1} (\log n)^3 + C_1^2 M_1^2/\alpha^2\right]/(2\pi).$$

EXAMPLE 3 (Double exponential). Suppose $f_0(x) = e^{-|x|}/2$. Let $c_n = \alpha n^{1/7}$, $0 < \alpha$. Then

$$\limsup_{n\to\infty} \sup_{G \in \mathscr{G}_1(M_1)} n^{2/7} E_G \|\hat{g}_n - g\|^2 \le \left[\alpha^5 C_3 + C_1^2 M_1^2/\alpha^2\right]/(2\pi).$$

PROOF OF THEOREM 1. Since $K_n(x,\cdot)$ is the real part of the Fourier inversion of $k^*(t/c_n)e^{itx}/f_0^*(t)$ for fixed x, it follows from the Plancherel identity that

$$2\pi E\int K_n^2(X,\theta)\,d\theta/\le \int |k^*(t/c_n)/f_0^*(t)|^2\,dt/n$$
$$\le \left[c_n\int|k^*(t)|^2\,dt\right]/\left[n|f_0^*(c_n)|^2\right]$$
$$= c_nC_3/\left[n|f_0^*(c_n)|^2\right].$$

By (7) $EK_n(X,\cdot)$ has the Fourier transformation $k^*(t/c_n)g^*(t)$, so that for $\|g\|<\infty$,

$$2\pi\int|EK_n(X,\theta)-g(\theta)|^2\,d\theta=\int|k^*(t/c_n)g^*(t)-g^*(t)|^2\,dt\to 0.$$

Since by (4)

$$E\|\hat{g}_n-g\|^2\le E\|K_n(X,\cdot)\|^2/n+\|EK_n(X,\cdot)-g\|^2,$$

we have (10) by adding the two inequalities together. For (11) we have

$$2\pi c_n^2E\|\hat{g}_n-g\|^2\le c_0C_3+c_n^2\int|k^*(t/c_n)g^*(t)-g^*(t)|^2\,dt$$
$$\le c_0C_3+C_1^2\int|tg^*(t)|^2\,dt. \qquad \square$$

Now consider a fixed point $\theta=a$. We shall give the asymptotic normality for our estimators as well as upper bounds for their rate of convergence. A mixing density g satisfies the Lipschitz condition at $\theta=a$ if

$$(12)\qquad |g(a+\theta)-g(a)|\le M_2|\theta|,\qquad \forall\,|\theta|\le\delta.$$

Define

$$(13)\qquad \begin{aligned}\mathscr{G}_2&=\mathscr{G}_2(a,\delta,M_2)\\&=\{G\colon g=G' \text{ exists in } [a-\delta,a+\delta] \text{ and } (12) \text{ holds}\}.\end{aligned}$$

THEOREM 2. *Let the estimator* $\hat{g}_n$ *be as in Theorem* 1 *such that* (5) *and* (9) *hold.*

(i) *If* $g=G'$ *exists at* $\theta=a$, *the* $\lim_n E|\hat{g}_n(a)-g(a)|^2=0$.
(ii) *For any* $n\ge 1$,

$$(14)\qquad \begin{aligned}&\sup_{G\in\mathscr{G}_2(a,\delta,M_2)} c_n^2E_G|\hat{g}_n(a)-g(a)|^2\\&\le\frac{c_0}{2\pi^2}\int|f_0^*(t)|\,dt+\left[\varepsilon(c_n\delta)(\delta^{-2}+M_2)+M_2C_1\right]^2.\end{aligned}$$

(iii) *Suppose that* $\lim_{n\to\infty} EK_n^2(X,a)=\infty$,

$$(15)\qquad \lim_{n\to\infty}\sup_x |K_n(x,a)|/\sqrt{nEK_n^2(X,a)}=0,\qquad \liminf_{n\to\infty} c_n^2 EK_n^2(X,a)/n>0.$$

Let $\hat\sigma_n^2=\hat\sigma_n^2(a)=\sum_{j=1}^n K_n^2(X_j,a)/n$. *If* $g'(a)$ *exists, then*

$$\sqrt{n}\,(\hat g_n(a)-g(a))/\hat\sigma_n \to_{\mathscr{D}} N(0,1)\qquad \text{if } 0<g(a)<\infty,$$

$$\sqrt{n}\,\hat g_n(a)/\hat\sigma_n \to_{\mathscr{D}} [N(0,1)]^+\quad \text{if } g(a)=0.$$

(iv) *Suppose that* $g''(a)$ *exists and* $\lim_{x\to\infty} x^3k(x)=0$. *If*

$$\lim_{n\to\infty} c_n^5/\left[n|f_0^*(c_n)|^2\right]=0,$$

then

$$\lim_{n\to\infty} c_n^2(\hat g_n(a)-g(a))=g''(a)\int x^2k(x)\,dx/2,\quad \text{in mean square}.$$

REMARKS. (1) If $\int|f_0^*(t)|\,dt=\infty$, then the first term on the right-hand side of (14) can be replaced by $c_0\int_{-2c_n}^{2c_n}|f_0^*(t)|\,dt/[2\pi^2]$. (2) Since $x^2k(x)$ has a bounded Fourier transformation with compact support, $\lim_{x\to\infty}\varepsilon(x)=0$. (3) The inequality (14) is still meaningful when $\delta=\infty$.

Condition (15) simply ensures that the bias of the kernel estimator has a smaller order than the standard deviation. In this case part (iii) shall be used to obtain the asymptotic normality (Example 3). If the standard deviation is of smaller order than the bias, then the limiting distribution is degenerate and part (iv) shall be used (Examples 1 and 2).

EXAMPLE 1 (Normal, continuation). Since $\int|f_0^*(t)|\,dt=\sqrt{2\pi}$,

$$\sup_{G\in\mathscr{G}_2(a,\delta,M_2)} (\log n)E_G|\hat g_n(a)-g(a)|^2$$
$$\le \frac{\sqrt{\alpha}\,(\log n)^{3/2}}{\pi\sqrt{2\pi}\,n^{1-\alpha}}+\frac{1}{\alpha}\left[\varepsilon\left(\delta\sqrt{\alpha\log n}\right)\left(\delta^{-2}+M_2\right)+C_1M_2\right]^2.$$

If $g''(a)$ exists, then

$$\lim_{n\to\infty}(\log n)(\hat g_n(a)-g(a))=g''(a)C_2/(2\alpha),\quad \text{in mean square}.$$

EXAMPLE 2 (Cauchy, continuation). Since $\int|f_0^*(t)|\,dt=2$,

$$\sup_{G\in\mathscr{G}_2(a,\delta,M_2)} (\log n)^2E_G|\hat g_n(a)-g(a)|^2$$
$$\le \frac{\alpha(\log n)^3}{\pi^2 n^{1-\alpha}}+\frac{1}{\alpha^2}\left[\varepsilon(\delta\alpha\log n)\left(\delta^{-2}+M_2\right)+C_1M_2\right]^2.$$

If $g''(a)$ exists, then

$$\lim_{n\to\infty} (\log n)^2(\hat{g}_n(a) - g(a)) = g''(a)C_2/(2\alpha^2), \quad \text{in mean square.}$$

EXAMPLE 3 (Double exponential, continuation). Since $\int|f_0^*(t)|\,dt = \pi$,

$$\limsup_{n\to\infty} \sup_{G\in\mathscr{G}_2(a,\delta,M_2)} n^{2/7}E_G|\hat{g}_n(a) - g(a)|^2 \le \alpha^5/(2\pi) + C_1^2M_2^2/\alpha^2.$$

Let $\hat{\sigma}_n^2 = \sum_{j=1}^n K_n^2(X_j, a)/n$. If $g'(a)$ exists, then (15) holds by Lemma 1 in the Appendix, so that

$$\sqrt{n}\,(\hat{g}_n(a) - g(a))/\hat{\sigma}_n \to_{\mathscr{D}} N(0,1) \qquad \text{if } g(a) > 0,$$

$$\sqrt{n}\,\hat{g}_n(a)/\hat{\sigma}_n \to_{\mathscr{D}} [N(0,1)]^+ \quad \text{if } g(a) = 0.$$

PROOF OF THEOREM 2. (i) To bound the variance at $\theta = a$, we have by (5) and (9)

$$\begin{aligned}\operatorname{Var}(K_n(X,a)) &\le EK_n^2(X,a)\\ &\le (2\pi)^{-2}\int_{-c_n}^{c_n}\int_{-c_n}^{c_n}[k^*(t/c_n)/f_0^*(t)]\overline{[k^*(s/c_n)/f_0^*(s)]}\\ &\qquad\qquad \times E\exp[i(t-s)(X-a)]\,ds\,dt\\ &\le (2\pi)^{-2}\int_{-c_n}^{c_n}\int_{-c_n}^{c_n}|f_0^*(c_n)|^{-2}|f_0^*(t-s)|\,ds\,dt\\ &\le (2\pi)^{-2}2c_n|f_0^*(c_n)|^{-2}\int_{-2c_n}^{2c_n}|f_0^*(u)|\,du.\end{aligned} \tag{16}$$

It follows from (7) that

$$\begin{aligned}EK_n(X,a) &= -\int c_nk(x)\,dG(a - x/c_n)\\ &= \int c_n[G(a-x/c_n) - G(a)]k'(x)\,dx\\ &= -g(a)\int xk'(x)\,dx + O(1)\int_{\delta c_n}^{\infty}|xk'(x)|\,dx + o(1)\int|xk'(x)|\,dx,\end{aligned}$$

which implies the mean square consistency by (4) and (16), since $-\int xk'(x)\,dx = \int k(x)\,dx = 1$.

(ii) Let $\varepsilon_1(x) = \sup_{y\ge x} y^2k(y)$ and $\varepsilon_2(x) = \int_x^\infty yk(y)\,dy$. By (7) and (12)

$$\begin{aligned}c_n|EK_n(X,a) - g(a)| &\le c_n\left|\int_{a-\delta}^{a+\delta}c_nk(c_n(a-\theta))g(\theta)\,d\theta - g(a)\right| + \varepsilon_1(c_n\delta)/\delta^2\\ &\le c_n\left|\int_{-c_n\delta}^{c_n\delta}k(x)[g(a+x/c_n) - g(a)]\,dx\right|\\ &\qquad + \varepsilon_1(c_n\delta)/\delta^2 + 2g(a)\varepsilon_2(c_n\delta)/\delta\\ &\le \varepsilon_1(c_n\delta)/\delta^2 + 2g(a)\varepsilon_2(c_n\delta)/\delta + M_2C_1.\end{aligned} \tag{17}$$

By the Lipschitz condition,

$$1 \geq \int_{a-\delta}^{a+\delta} g(\theta)\, d\theta \geq \int_{a-\delta}^{a+\delta} \big[g(a) - M_2|\theta - a|\big]\, d\theta = 2\delta g(a) - M_2\delta^2.$$

Hence, we have part (ii) by (16) and (17).

(iii) Basically the asymptotic normality follows from the Lindeberg central limit theorem. Since $EK_n(X,a) \to g(a)$ and $EK_n^2(X,a) \to \infty$, $\operatorname{Var}(K_n(X,a)) = (1 + o(1))EK_n^2(X,a)$, so that by (15)

$$\begin{aligned} \varepsilon_n &= \max_x |K_n(x,a)| \Big/ \sqrt{n\sigma_n^2} \\ &= (1 + o(1)) \max_x |K_n(x,a)| \Big/ \sqrt{nEK_n^2(X,a)} = o(1), \end{aligned} \tag{18}$$

where $\sigma_n^2 = \operatorname{Var}(K_n(X,a))$. Let $Y_{nj} = [K_n(X_j,a) - EK_n(X,a)]/\sqrt{n\sigma_n^2}$. Then $P\{|Y_{nj}| \leq 2\varepsilon_n\} = 1$, which implies by the Lindeberg central limit theorem that

$$\sqrt{n}\left[n^{-1}\sum_{j=1}^{n} K_n(X_j,a) - EK_n(X,a)\right]\Big/\sigma_n = \sum_{j=1}^{n} Y_{nj} \to_{\mathscr{D}} N(0,1).$$

Taking the Taylor expansion of $g(a + x/c_n) - g(a)$ at the last step of (17), we have by (15)

$$\sqrt{n}\,[EK_n(X,a) - g(a)]/\sigma_n = O(1)c_n[EK_n(X,a) - g(a)] = o(1).$$

Since

$$P\{n^{-1}\Sigma_{j=1}^{n} K_n(X_j,a) = \hat{g}_n(a)\} \to 1 \quad \text{for } g(a) > 0$$

and $P\{n^{-1}\Sigma_{j=1}^{n} K_n(X_j,a) < 0\} \to 1/2$ for $g(a) = 0$,

$$\begin{aligned} \sqrt{n}\,(\hat{g}_n(a) - g(a))/\sigma_n &\to_{\mathscr{D}} N(0,1) \qquad \text{if } 0 < g(a) < \infty, \\ \sqrt{n}\,(\hat{g}_n(a) - g(a))/\sigma_n &\to_{\mathscr{D}} [N(0,1)]^{+} \quad \text{if } g(a) = 0. \end{aligned}$$

It follows from (18) that

$$\begin{aligned} E\big[\hat{\sigma}_n^2/\sigma_n^2 - 1\big]^2 &\leq (1 + o(1))EK_n^4(X,a)/(n\sigma_n^4) \\ &\leq (1 + o(1))\varepsilon_n^2 EK_n^2(X,a)/\sigma_n^2 = o(1). \end{aligned}$$

Hence, $\hat{\sigma}_n^2/\sigma_n^2 \to 1$ in probability and part (iii) follows the Slutsky theorem.

(iv) Taking further Taylor expansions of $g(a + x/c_n) - g(a)$ in (17), we have part (iv) by (16). □

3. Estimating mixing distributions. Let the kernel be the same as in Section 2,

$$K_n(x,a) = (2\pi)^{-1}\int \mathbb{R}\{k^*(t/c_n)e^{it(x-a)}/f_0^*(t)\}\, dt, \qquad 0 < c_n \uparrow \infty, \tag{19}$$

with a density function k such that (6) holds. We shall use the integration $\int_{-\infty}^{a} K_n(x,\theta)\, d\theta$ as our kernel for the mixing distribution and the statistic

$$\hat{G}_n(a) = \min\big((\overline{G}_n(a))^{+}, 1\big), \qquad \overline{G}_n(a) = \int_{-\infty}^{a} n^{-1}\sum_{j=1}^{n} K_n(X_j,\theta)\, d\theta, \tag{20}$$

as our estimation for $G(a)$. In this section we study the integrability of the kernel (19) and the asymptotics of $\hat{G}_n$ for the location case (2).

Let $H(\cdot)$ be a distribution function such that $\int_0^1 |1 - H^*(t)| t^{-1}\, dt + \int_1^\infty |H^*(t)| t^{-1}\, dt < \infty$. Assume that $\int_0^1 |1 - f_0^*(t)| t^{-1}\, dt < \infty$. Then

$$\int \left| \frac{k^*(t/c_n) e^{itx} - f_0^*(t) H^*(t)}{t f_0^*(t)} \right| dt < \infty,$$

so that we can define

$$(21) \qquad J_n(x, a) = H(a) - \frac{1}{2\pi} \int e^{-ita} \frac{k^*(t/c_n) e^{itx} - f_0^*(t) H^*(t)}{it f_0^*(t)}\, dt.$$

By the Riemann–Lebesgue lemma

$$\lim_{a \to -\infty} J_n(x, a) = 0 \quad \text{and} \quad \lim_{a \to \infty} J_n(x, a) = 1.$$

It follows from the Fourier inversion formula that

$$\int_a^b K_n(x, \theta)\, d\theta = \mathbb{R}\{ J_n(x, b) - J_n(x, a) \},$$

which implies that $K_n(x, \cdot\,)$ is Riemann integrable and $\int_{-\infty}^a K_n(x, \theta)\, d\theta = \mathbb{R}\{J_n(x, a)\}$. Hence, letting $F_n(x)$ be the ECDF of $X_1, \dots, X_n$, we have

$$(22) \qquad \begin{aligned} \overline{G}_n(a) &= \mathbb{R}\left\{ \frac{1}{n} \sum_{j=1}^n J_n(X_j, a) \right\} \\ &= H(a) - \frac{1}{2\pi} \int \mathbb{R}\left\{ e^{-ita} \frac{k^*(t/c_n) F_n^*(t) - f_0^*(t) H^*(t)}{it f_0^*(t)} \right\} dt. \end{aligned}$$

We first consider the L^2 loss. The following theorem is proved in the Appendix.

THEOREM 3. *Assume that* $\int |x| f_0(x)\, dx < \infty$. *Let* $\hat{G}_n$ *be defined by* (20) *with the kernel in* (19).

(i) *Let* c_n *be such that* $1/n \le c_0 |f_0^*(c_n)|^2 / c_n$, $|f_0^*(c_n)| = \min_{|t| \le c_n} |f_0^*(t)|$ *and* $c_n \uparrow \infty$. *Then*

$$(23) \qquad \begin{aligned} c_n E_G \left\| \hat{G}_n - G \right\|^2 \le c_0 \int F_0(x)(1 - F_0(x))\, dx + \frac{1 + C_1^2}{\pi} \\ + \frac{c_n}{n} \int G(\theta)(1 - G(\theta))\, d\theta, \end{aligned}$$

where $F_0(x) = \int_{-\infty}^x f_0(y)\, dy$.

(ii) *Let* c_n *be such that* $1/n \le c_0|f_0^*(c_n)|^2/c_n^2$, $|f_0^*(c_n)| = \min_{|t|\le c_n}|f_0^*(t)|$ *and* $c_n \uparrow \infty$. *Then*

$$(24)\qquad \begin{aligned} c_n^2 E_G\|\hat{G}_n - G\|^2 &\le c_0\int F_0(x)(1 - F_0(x))\,dx + \frac{C_1^2\|G^*(t)\|^2}{2\pi} \\ &\quad + \frac{c_n^2}{n}\int G(\theta)(1 - G(\theta))\,d\theta. \end{aligned}$$

REMARK. $\int F_0(x)(1 - F_0(x))\,dx < \infty$ if and only if $\int|x|f_0(x)\,dx < \infty$.

Let (X_i, θ_i) be iid random vectors such that $X_i|\theta_i \sim N(\theta_i, 1)$ and $\theta_i \sim G$. Choose $c_n = \sqrt{\alpha \log n}$. Then our argument in the proof shows that

$$(25)\qquad \begin{aligned} \sqrt{\log n}\,E_G\|\hat{G}_n - G_n\|^2 &\le n^{\alpha-1}\sqrt{\log n}\big/\sqrt{\tau} + (1 + C_1^2)/(\pi\sqrt{\alpha}), \\ (\log n)E_G\|\hat{G}_n - G_n\|^2 &\le n^{\alpha-1}\log n/\sqrt{\pi} + C_1^2\|G^*(t)\|^2/(2\pi\alpha), \end{aligned}$$

where G_n is the ECDF of $\theta_1, \dots, \theta_n$. The upper bound of Edelman (1988) for estimating G_n in the normal case is comparable to (25). The quantity $\int G(\theta) \times (1 - G(\theta))\,d\theta/n = E\|G_n - G\|^2$ in (23) and (24) is the risk of estimating G by the ECDF G_n based on the extra knowledge of $\theta_1, \dots, \theta_n$. Given the observations $X_1, \dots, X_n$, the estimator $\hat{G}_n$ is not necessarily a distribution function. However, it can be shown that we can always construct a distribution function $\hat{G}_n^0$ based on $\hat{G}_n$ such that $\|\hat{G}_n^0 - H\| \le \|\hat{G}_n - H\|$ for any distribution function H.

Let us consider a fixed point $\theta = a$. Under certain moment conditions on X,

$$EJ_n(X, a) = H(a) - \frac{1}{2\pi}\int e^{-ita}\frac{k^*(t/c_n)G^*(t) - H^*(t)}{it}\,dt.$$

Since H is arbitrary, we can take $H^* = (k^*(t/c_n)G^*(t)$, which implies that

$$(26)\qquad EJ_n(X, a) = \int G(\theta)c_n k(c_n(a - \theta))\,d\theta.$$

By (6) $k(x) = k(-x)$ and the right-hand side of (26) converges to

$$(27)\qquad \underline{G}(a) = [G(a+) + G(a-)]/2,$$

which equals to $G(a)$ if G is continuous at a. Theorem 4 is also proved in the Appendix.

THEOREM 4. *Assume that* $E(\log(1 + |X|))^2 < \infty$ *and* G *satisfies the Lipschitz condition*

$$(28)\qquad \left|\frac{G(a + \theta) + G(a - \theta)}{2} - \underline{G}(a)\right| \le |\theta|M_3 < \infty, \qquad \forall\, 0 < \theta < \delta.$$

Choose c_n *such that* $1/n \le c_0|f_0^*(c_n)|^2/[c_n \log^+ c_n]^2$, $|f_0^*(c_n)| = \min_{|t| \le c_n}|f_0^*(t)|$ *and* $c_n \uparrow \infty$. *Then*

$$\limsup_{n\to\infty} c_n^2 E_G\left|\hat{G}_n(a) - \underline{G}(a)\right|^2 \le c_0(2/\pi)^4 + M_3^2C_1^2.$$

EXAMPLE 1 (Normal, continuation). Let $H(\cdot) \sim N(a, 1)$. Then

$$\mathbb{R}\{J_n(x,a)\} = 1/2 - (2\pi)^{-1}\int t^{-1}\sin(t(x-a))k^*(t/c_n)e^{t^2/2}\,dt.$$

Since $\int F_0(x)(1 - F_0(x))\,dt = 1/\sqrt{\pi}$, by Theorems 3 and 4 we have

$$\sqrt{\log n}\, E_G\left\|\hat{G}_n - G\right\|^2 \le \frac{\sqrt{\log n}}{\sqrt{\pi}\, n^{1-\alpha}} + \frac{1 + C_1^2}{\pi\sqrt{\alpha}} + \frac{\sqrt{\log n}}{n}\int G(\theta)(1 - G(\theta))\,d\theta,$$

$$(\log n) E_G\left\|\hat{G}_n - G\right\|^2 \le \frac{\log n}{\sqrt{\pi}\, n^{1-\alpha}} + \frac{C_1^2\|G^*(t)\|^2}{2\pi\alpha} + \frac{\log n}{n}\int G(\theta)(1 - G(\theta))\,d\theta,$$

and if $\int \log^2(1 + |x|)\,dG(x) < \infty$ and (28) holds, then

$$\limsup_{n\to\infty}(\log n) E_G\left|\hat{G}_n(a) - \underline{G}(a)\right|^2 \le \frac{M_3^2C_1^2}{\alpha}.$$

EXAMPLE 2 (Cauchy, continuation). Since $E|X| = \infty$, Theorem 3 is not applicable. If $\int \log^2(1 + |x|)\,dG(x) < \infty$ and (28) holds, then by Theorem 4

$$\limsup_{n\to\infty}(\log n)^2 E_G\left|\hat{G}_n(a) - \underline{G}(a)\right|^2 \le M_3^2C_1^2/\alpha^2.$$

EXAMPLE 3 (Double exponential, continuation). Let $c_n = \alpha n^{1/5}$. We have $\int F_0(x)(1 - F_0(x))\,dt = 3/4$. If $\int G(\theta)(1 - G(\theta))\,d\theta < \infty$, then by Theorem 3

$$\limsup_{n\to\infty} n^{1/5} E_G\left\|\hat{G}_n - G\right\|^2 \le 3\alpha^4/4 + \left(1 + C_1^2\right)/(\pi\alpha).$$

Let $c_n = \alpha n^{1/6}$. If $\int G(\theta)(1 - G(\theta))\,d\theta < \infty$, then by Theorem 3

$$\limsup_{n\to\infty} n^{1/3} E_G\left\|\hat{G}_n - G\right\|^2 \le 3\alpha^4/4 + C_1^2\|G^*(t)\|^2/(2\pi\alpha^2).$$

Let $c_n = \alpha(n/\log^2 n)^{1/6}$. If $\int \log^2(1 + |x|)\,dG(x) < \infty$ and (28) holds, then by Theorem 4

$$\limsup_{n\to\infty} n^{1/3}(\log n)^{-2/3} E_G\left|\hat{G}_n(a) - \underline{G}(a)\right|^2 \le 16\alpha^4/\pi^2 + M_3^2C_1^2/\alpha^2.$$

Both Theorems 3 and 4 require moment conditions on F_0 as well as G. However, these moment conditions are not necessary if we estimate

$$(29)\quad \underline{G}(a,b) = \underline{G}(b) - \underline{G}(a) = \frac{G(b+) + G(b-)}{2} - \frac{G(a+) + G(a-)}{2}.$$

Define

$$(30)\quad I_n(x,a,b) = \int_a^b K_n(x,\theta)\,d\theta = \frac{1}{2\pi}\int \mathbb{R}\left\{ e^{itx}\frac{(e^{-ita}-e^{-itb})k^*(t/c_n)}{itf_0^*(t)}\right\} dt,$$

$$(31)\qquad \hat{G}_n(a,b) = \min\left(\left(\sum_{j=1}^{n} I_n(X_j,a,b)/n\right)^+, 1\right),$$

$$(32)\qquad \mathscr{G}_3 = \mathscr{G}_3(a,b,\delta,M_3) = \mathscr{G}_3(a,\delta,M_3)\cap\mathscr{G}_3(b,\delta,M_3),$$

where $\mathscr{G}_3(a,\delta,M_3) = \{G\colon (28) \text{ holds for } G\}$.

THEOREM 5. *Let the estimator* $\hat{G}_n(a,b)$ *be given by* (31). *Choose the constants* c_n *such that*

$$(33)\qquad \frac{1}{n} \le \frac{c_0}{c_n}|f_0^*(c_n)|^2, \qquad \min_{|t|\le 2/|b-a|}\frac{2}{b-a}|f_0^*(t)| \ge c_n|f_0^*(c_n)|,$$
$$\min_{2/|b-a|\le|t|\le c_n}|tf_0^*(t)| = c_n|f_0^*(c_n)|.$$

(i) *For any* $a\le b$, $\lim_{n\to\infty} E|\hat{G}_n(a,b) - \underline{G}(a,b)|^2 = 0$.

(ii) *Let* $\varepsilon_0(x) = \int_x^\infty yk(y)\,dy$. *Then*

$$\sup_{G\in\mathscr{G}_3(a,b,\delta,M_3)} c_n^2 E_G\left|\hat{G}_n(a,b) - \underline{G}(a,b)\right|^2$$
$$\le \frac{2c_0}{\pi^2}\int|f_0^*(t)|\,dt + 4[\varepsilon_0(c_n\delta)/\delta + M_3C_1]^2.$$

(iii) *Suppose that* $\lim_{n\to\infty} EI_n^2(X,a,b) = \infty$,

$$(34)\qquad \lim_{n\to\infty}\sup_x|I_n(x,a,b)|/\sqrt{nEI_n^2(X,a,b)} = 0,$$
$$\liminf_{n\to\infty} c_n^2EI_n^2(X,a,b)/n > 0.$$

Let $\hat{\sigma}_n^2 = \hat{\sigma}_n^2(a,b) = \sum_{j=1}^n I_n^2(X_j,a,b)/n$. *If* $g(a) = G'(a)$ *and* $g(b) = G'(b)$ *exist, then*

$$\sqrt{n}\,(\hat{G}_n(a,b) - G(b) + G(a))/\hat{\sigma}_n \to_{\mathscr{D}} N(0,1) \qquad \text{if } G(b) - G(a)\in(0,1),$$
$$\sqrt{n}\,\hat{G}_n(a,b)/\hat{\sigma}_n \to_{\mathscr{D}} [N(0,1)]^+ \qquad \text{if } G(b) = G(a),$$
$$\sqrt{n}\,(\hat{G}_n(a,b) - 1)/\hat{\sigma}_n \to_{\mathscr{D}} -[N(0,1)]^+ \qquad \text{if } G(b) - G(a) = 1.$$

(iv) *Suppose that* $g'(a) = G''(a)$ *and* $g'(b) = G''(b)$ *exist. If* $\lim_{n\to\infty} c_n^3/[n|f_0^*(c_n)|^2] = 0$, *then*

$$\lim_{n\to\infty} c_n^2(\hat{G}_n(a,b) - G(b) + G(a))$$
$$= [g'(b) - g'(a)]\int x^2k(x)\,dx/2, \quad \textit{in mean square}.$$

PROOF. By (33) for $|t| \le c_n$,

$$\max_{|t| \le c_n} \left| \frac{e^{-ita} - e^{-itb}}{t f_0^*(t)} \right| \le \min\left(\frac{|b-a|}{|f_0^*(t)|}, \frac{2}{|t f_0^*(t)|} \right) \le \frac{2}{c_n |f_0^*(c_n)|}.$$

The argument of (16) provides a bound for the variance term. For the bias term we have

$$\begin{aligned} &|EI_n(X, a, b) - \underline{G}(a, b)| \\ &\quad = \left| \int \{[G(b + x/c_n) - \underline{G}(b)] - [G(a + x/c_n) - \underline{G}(a)]\} k(x)\, dx \right| \\ &\quad \le 2M_3 \int |x| k(x)\, dx + 2\int_{\delta c_n}^{\infty} k(x)\, dx, \end{aligned}$$

by (8) and the Lipschitz condition. The rest of the proof is similar to that of Theorem 2 and omitted. □

EXAMPLE 1 (Normal, continuation). The estimator $\hat{G}_n(a, b)$ is given by (31) with

$$I_n(x, a, b) = \pi^{-1} \int t^{-1} \sin(t(b-a)/2) \cos(t(x - (b+a)/2)) \, k^*(t/c_n) e^{t^2/2}\, dt.$$

It follows from part (ii) of Theorem 5 that

$$\begin{aligned} &\sup_{G \in \mathscr{G}_3(a, b, \delta, M_3)} (\log n) E_G \left| \hat{G}_n(a, b) - \underline{G}(a, b) \right|^2 \\ &\quad \le \frac{\sqrt{8 \log n}}{\sqrt{\pi^3 \alpha}\, n^{1-\alpha}} + \frac{4}{\alpha} \left[\varepsilon\left(\delta \sqrt{\alpha \log n}\right)/\delta + M_3 C_1 \right]^2. \end{aligned}$$

Since the standard deviation is of smaller order than the bias, we do not apply part (iii). If $g'(a)$ and $g'(b)$ exist, then by part (iv)

$$\begin{aligned} &\lim_{n \to \infty} (\log n)\left(\hat{G}_n(a, b) - G(b) + G(a)\right) \\ &\quad = [g'(b) - g'(a)] C_2/(2\alpha), \quad \text{in mean square}. \end{aligned}$$

EXAMPLE 2 (Cauchy, continuation). Similar to the normal case, we have

$$\begin{aligned} &\sup_{G \in \mathscr{G}_3(a, b, \delta, M_3)} (\log n)^2 E_G \left| \hat{G}_n(a, b) - \underline{G}(a, b) \right|^2 \\ &\quad \le \frac{4 \log n}{\pi^2 \alpha n^{1-\alpha}} + \frac{4}{\alpha^2} [\varepsilon(\delta \alpha \log n)/\delta + M_3 C_1]^2, \end{aligned}$$

and if $g'(a)$ and $g'(b)$ exist, then

$$\lim_{n\to\infty} (\log n)\big(\hat{G}_n(a,b) - G(b) + G(a)\big)$$
$$= [g'(b) - g'(a)]C_2/(2\alpha^2), \quad \text{in mean square.}$$

EXAMPLE 3 (Double exponential, continuation). Let $c_n = \alpha n^{1/5}$, $\alpha > 0$. Then

$$\limsup_{n\to\infty} \sup_{G\in\mathscr{G}_3(a,b,\delta,M_3)} n^{2/5} E_G\big|\hat{G}_n(a,b) - \underline{G}(a,b)\big|^2 \le 2\alpha^3/\pi + 4M_3^2C_1^2/\alpha^2.$$

Let $\hat{\sigma}_n^2 = \sum_{j=1}^n I_n^2(X_j, a, b)/n$. If $g(a)$ and $g(b)$ exist, then by Lemma 1 in the Appendix

$$\sqrt{n}\big(\hat{G}_n(a,b) - G(b) + G(a)\big)/\hat{\sigma}_n \to_{\mathscr{D}} N(0,1) \qquad \text{if } G(b) - G(a) \in (0,1),$$
$$\sqrt{n}\,\hat{G}_n(a,b)/\hat{\sigma}_n \to_{\mathscr{D}} [N(0,1)]^+ \qquad \text{if } G(b) = G(a),$$
$$\sqrt{n}\big(\hat{G}_n(a,b) - 1\big)/\hat{\sigma}_n \to_{\mathscr{D}} -[N(0,1)]^+ \qquad \text{if } G(b) - G(a) = 1.$$

4. Lower bounds for the optimal rate of convergence. We study lower bounds for the optimal rate of convergence in this section. Our basic idea is borrowed from Le Cam (1972) and Hájek (1972) among other papers on asymptotic information and efficiency. Instead of directly using the LAN condition, we observe that a lower bound for the rate of convergence can be obtained for any pair of contiguous sequences in infinitesimal neighborhoods of $f(x;G)$. As demonstrated by Example 1 in Section 1, Fourier methods can be used to derive contiguous sequences. We consider estimating mixing densities and distributions in two separate sections.

4.1. *Lower bounds for estimating mixing densities.* We shall try to find density functions $g_j = G_j'$, $j = 1,2$, such that $f(\cdot;G_1)$ and $f(\cdot;G_2)$ are relatively close but not g_1 and g_2 for a given sample size n. Let $g_0 = G_0'$ be a density function and G be a distribution function. For constants $b_n, d_n > 0$ and e_n define

$$(35) \quad dG_1(\theta) = (1-w)\,dG(\theta) + wg_0(\theta)\,d\theta, \qquad 0 < w < 1,$$

$$(36) \quad dG_2(\theta) = dG_1(\theta) + [w/(4d_n)][\cos(d_n\theta - b_n) - e_n]g_0(\theta)\,d\theta.$$

It follows from the Riemann–Lebesgue lemma that $\lim_{d_n\to\infty} d_n[f(x;G_2) - f(x;G_1)] = 0$, so that $f(\cdot;G_1)$ and $f(\cdot;G_2)$ are at least as close as $o(d_n^{-1})$ and $|G_1' - G_2'|$ is bounded from below by some thing of order d_n^{-1} for large d_n. Note that we have contiguous sequences if the $o(d_n^{-1})$ is actually $O(n^{-1/2})$.

THEOREM 6. *Suppose there exist a density function $g_0 = G_0'$ and nonnegative functions $\rho(t)$ and $u(x)$ such that $\rho(t)/t \downarrow 0$ as $t \to \infty$ and*

$$(37) \quad \left|\int e^{it\theta} f(x|\theta)g_0(\theta)\,d\theta\right| \le \rho(t)u(x)f(x;G_0), \qquad \forall\, t \ge 1,\ \rho(1) = 1.$$

Let G be a distribution function with derivative g(a) at $\theta = a$. Let G_1 and G_2 be defined by (35) *and* (36) *with*

$$(38)\quad e_n = \int \cos(d_n\theta - b_n) g_0(\theta)\, d\theta \in (-1, 1), \qquad \rho(d_n)/d_n \le n^{-1/2}.$$

Then for any estimator $\tilde{g}_n(a)$ based on $X_1, \dots, X_n$,

$$(39)\quad \begin{aligned} &\max_{j=1,2} P_j\left\{ d_n |\tilde{g}_n(a) - g_j(a)| \ge \frac{w}{8} |\cos(d_n a - b_n) - e_n| g_0(a) \right\} \\ &\quad \ge \left[2(1 + e^{2M})\right]^{-1}, \end{aligned}$$

where $P_j = P_{G_j}$, $E_j = E_{G_j}$, $g_j = G_j'$, $j = 0, 1, 2$, and $M = 1 + 3E_1(u(X) + E_0 u(X))^2/16$. Furthermore, if G has a density g and $M < \infty$, then

$$(40)\quad \inf_{n \ge 1} \max_{j=1,2} d_n^2 \int E_j \left|\tilde{g}_n(\theta) - g_j(\theta)\right|^2 d\theta > 0.$$

REMARKS. By the Riemann–Lebesgue lemma $\lim e_n = 0$. If $g_0(a) > 0$ and $b_n = d_n a$, then $(w/8)|\cos(d_n a - b_n) - e_n| g_0(a)$ is bounded away from 0. (2) If $G_0 \in \mathscr{G}_j$, then such pairs (G_1, G_2) are dense in $\mathscr{G}_j$, $j = 1, 2$, where $\mathscr{G}_1$ and $\mathscr{G}_2$ are defined by (8) and (13), respectively. (3) The constant M does not depend on n, d_n, b_n or e_n.

For the location case the integration on the left-hand side of (37) is the convolution of $f_0(x)$ and $e^{itx} g_0(x)$ for fixed t. Since $e^{itx} g_0(x)$ has the Fourier transformation $g_0^*(t + s)$, we have

$$\int e^{it\theta} f(x|\theta) g_0(\theta)\, d\theta = (2\pi)^{-1} \int e^{-ixs} f_0^*(s) g_0^*(t + s)\, ds,$$

so that $\rho(t)$ is basically a mixture of the tail of f_0^* near t if g_0^* has very thin tails. The examples below suggest that $\rho(t)$ and $|f_0^*(t)|$ have similar behavior for large t. For smooth f_0, we have f_0^* with thin tails and slow lower bounds for the optimal rate of convergence.

COROLLARY 1. *Suppose that the conditions of Theorem* 6 *hold with $E_{G_0} u^2(X) < \infty$ and $G_0 \in \mathscr{G}_1(M_{10})$ for some $M_{10} < \infty$, where $\mathscr{G}_1$ is defined by* (8). *Then for any estimator $\tilde{g}_n$ and $M_1 > 0$,*

$$\inf_{n \ge 1} \sup_{G \in \mathscr{G}_1(M_1)} d_n^2 \int E_G (\tilde{g}_n(\theta) - g(\theta))^2 d\theta > 0.$$

PROOF. Let G_j, $j = 1, 2$, be defined by (35) and (36) with $G \in \mathscr{G}_1(M_1/2)$ and $E_G u^2(X) < \infty$. Since

$$\begin{aligned} \|t g_2^*(t)\| &\le \|t g_1^*(t)\| + (w/(8d_n))\big(2\|t g_0^*(t + d_n)\| + 2|e_n|\, \|t g_0^*(t)\|\big) \\ &= \|t g_1^*(t)\| + (w/4)\big(\|(t + d_n) g_0^*(t)/d_n\| + \|t g_0^*(t)\|/d_n\big), \end{aligned}$$

there exists a (small) $w > 0$ such that $\|tg_j^*(t)\| < M_1$ for all $j = 1, 2$ and $n \geq 1$. The corollary follows from (40) and the fact that $M < \infty$. □

COROLLARY 2. *Suppose that the conditions of Theorem 6 hold with $E_{G_0}u^2(X) < \infty$ and $G_0 \in \mathscr{G}_2(a, \delta_0, M_{20})$ for some $\delta_0 > 0$ and $M_{20} < \infty$, where $\mathscr{G}_2$ is defined by (13). Then for any estimator $\tilde{g}_n$ and constants $\delta < \infty$ and $M_2 > 0$ there exists an $\varepsilon > 0$ such that*

$$\inf_{n \geq 1} \sup_{G \in \mathscr{G}_2(a, \delta, M_2)} P_G\{d_n|\tilde{g}_n(a) - g(a)| > \varepsilon\} > 0.$$

EXAMPLE 1 (Normal, continuation). The condition (37) holds with

$$G_0 \sim N(0, \sigma^2), \qquad \rho(t) = \exp[-(t^2 - 1)\sigma^2/(2(1 + \sigma^2))]$$

and

$$u(x) = e^{-\sigma^2/(2(1+\sigma^2))},$$

so that $d_n = \sqrt{1 + (1 + \sigma^{-2})\log n}$, which implies by Corollaries 1 and 2 that there exists an $\varepsilon > 0$ such that for any statistics $\tilde{g}_n$ or $\tilde{g}_n(\cdot)$,

$$\inf_{n \geq 1} \sup_{G \in \mathscr{G}_1(M_1)} \log n E_G\|\tilde{g}_n - g\|^2 > \varepsilon,$$

$$\inf_{n \geq 1} \sup_{G \in \mathscr{G}_2(a, \delta, M_2)} P_G\{\sqrt{\log n}\,|\tilde{g}_n - g(a)| > \varepsilon\} > \varepsilon.$$

Hence, our estimators achieve the optimal rate of convergence $(\log n)^{-1/2}$ for the classes $\mathscr{G}_1$ and $\mathscr{G}_2$.

EXAMPLE 2 (Cauchy, continuation). Let $g_0(\theta) = \sigma/[\pi(\sigma^2 + \theta^2)]$. Then by the Cauchy residue theorem for $t \geq 0$,

$$\begin{aligned}
&\int e^{it\theta} f(x|\theta) g_0(\theta)\, d\theta \\
&\quad = (\sigma/\pi^2)\int e^{it\theta}\left(1 + (x - \theta)^2\right)^{-1}(\sigma^2 + \theta^2)^{-1}\, d\theta \\
&\quad = \frac{\sigma e^{-t} e^{ixt}(\sigma^2 + x^2 - 1 - 2xi) + e^{-\sigma t}(x^2 + 1 - \sigma^2 + 2\sigma xi)}{\pi\left[(\sigma + 1)^2 + x^2\right]\left[(\sigma - 1)^2 + x^2\right]} \\
&\quad = \exp[-\min(1, \sigma)t]\frac{(\sigma + 1)/\pi}{(\sigma + 1)^2 + x^2} h(x, t, \sigma), \quad \text{say.}
\end{aligned}$$

By algebra $\sup\{|h(x, t, \sigma)|: -\infty < x < \infty, t \geq 0\} = M(\sigma) < \infty$, so that

$$\left|\int e^{it\theta} f(x|\theta) g_0(\theta)\, d\theta\right| \leq \exp[-\min(1, \sigma)|t|]\, f(x; G_0) M(\sigma), \qquad \forall\, t,$$

which implies that (37) holds with $\rho(t) = \exp[-\min(1, \sigma)(|t| - 1)]$ and $u(x) =$

$e^{-\min(1,\sigma)}M(\sigma)$. It follows that $d_n = 1 + (\log n)/[2\min(1,\sigma)]$. By Corollaries 1 and 2 for some $\varepsilon > 0$,

$$\inf_{n\ge 1}\sup_{G\in\mathscr{G}_1}(\log n)^2 E_G\|\tilde{g}_n - g\|^2 > \varepsilon, \inf_{n\ge 1}\sup_{g\in\mathscr{G}_2} P_G\{(\log n)|\tilde{g}_n - g(a)| > \varepsilon\} > \varepsilon.$$

Hence, our estimators achieve the optimal rate of convergence $(\log n)^{-1}$ for the classes $\mathscr{G}_1$ and $\mathscr{G}_2$.

EXAMPLE 3 (Double exponential, continuation). To derive lower bounds for the optimal rate of convergence, we set $g_0(\theta) = e^{-|\theta|}/2$. Since

$$\int e^{it\theta}e^{-|x-\theta|-|\theta|}\,d\theta = e^{-x}\left(\frac{4}{4+t^2} + \frac{2(e^{itx}-1)}{t^2+2it}\right), \qquad \forall\, x \ge 0,$$

$$\begin{aligned}\left|\int e^{it\theta}f(x|\theta)g_0(\theta)\,d\theta/f(x;G_0)\right| &= \frac{1}{1+|x|}\left|\frac{4}{4+t^2} + \frac{2(e^{itx}-1)}{t^2+2it}\right| \\ &\le 8/t^2, \qquad \forall\, x \ge 0,\end{aligned}$$

so that $\rho(t) = t^{-2}$ and $u(x) = 8$. Since $\|g_0'\|^2 = 1/2$, by Corollaries 1 and 2 for some $\varepsilon > 0$,

$$\inf_{n\ge 1}\sup_{g\in\mathscr{G}_1} n^{1/3}E_G\|\tilde{g}_n - g\|^2 > \varepsilon, \qquad \inf_{n\ge 1}\sup_{G\in\mathscr{G}_2} P_G\{n^{1/6}|\tilde{g}_n - g(a)| > \varepsilon\} > \varepsilon.$$

4.2. *Lower bounds for estimating mixing distributions.* Now it is easy to study the case of estimating mixing distributions. Let $g_0 = G_0'$ be a density function and G be an arbitrary distribution function. Define $dG_1(\theta)$ by (35) and

$$(41) \qquad dG_2(\theta) = dG_1(\theta) + (w/4)(\cos(d_n\theta - b_n) - e_n)g_0(\theta)\,d\theta.$$

THEOREM 7. *Suppose* (37) *holds such that* $\rho(t)\downarrow 0$ *as* $t\to\infty$. *Define* G_1 *and* G_2 *by* (35) *and* (41) *with constants* e_n *in* (38) *and* $\rho(d_n) \le n^{-1/2}$. *Then for any Borel function* h *and estimator* $\tilde{G}_n(h)$ *for* $G(h) = \int h(\theta)\,dG(\theta)$ *based on* $X_1,\dots,X_n$,

$$\begin{aligned}&\max_{j=1,2} P_j\left\{|\tilde{G}_n(h) - G_j(h)| \ge \frac{w}{8}\left|\int h(\theta)(\cos(d_n\theta - b_n) - e_n)g_0(\theta)\,d\theta\right|\right\} \\ &\qquad \ge [2(1+e^{2M})]^{-1},\end{aligned}$$

where P_j, $j = 1,2$, *and* M *are as in Theorem* 6.

We omit the proof since it is almost identical to that of Theorem 6. Define

$$(42) \qquad \begin{aligned}\mathscr{G}_3' &= \mathscr{G}_3'(a,b,\delta,M_3) \\ &= \mathscr{G}_3(a,b,\delta,M_3) \cap \{G\colon G(a+) = G(a-), G(b+) = G(b-)\}.\end{aligned}$$

COROLLARY 3. *Suppose that the conditions of Theorem 7 hold with a differentiable density function* $g_0 = G_0'$ *such that* $E_{G_0}u^2(X) < \infty$, $\int |g_0'(\theta)| d\theta < \infty$ *and* $g_0(a) + g_0(b) > 0$. *Let* d_n *be such that* $\rho(d_n) = 1/\sqrt{n}$. *Then for any* $\delta \geq 0$ *and* $M_3 > 0$ *there exists an* $\varepsilon > 0$ *such that for any estimator* $\tilde{G}_n(a, b)$,

$$\inf_{n \geq 1} \sup_{G \in \mathscr{G}_3'(a, b, \delta, M_3)} P_G\{d_n |\tilde{G}_n(a, b) - (G(b) - G(a))| \geq \varepsilon\} > 0.$$

PROOF. Let the constants d_n' b_n' and e_n' be such that $d_n'|b - a|/2 = 2k_n\pi + \pi/2$ for some integer k_n, $d_n \leq d_n' < d_n + 4\pi/|b - a|$, $b_n' = d_n'(b + a)/2$ and $e_n' = \int \cos(d_n'\theta - b_n')g_0(\theta)\, d\theta$. Then $\rho(d_n') \leq 1/\sqrt{n}$ and by Theorem 7

$$\sup_{G \in \mathscr{G}_3'(a, b, \delta, M_3)} P_G\left\{d_n |\tilde{G}_n(a, b) - (G(b) - G(a))| \geq \frac{w}{8} d_n \left| \int_a^b (\cos(d_n'\theta - b_n') - e_n') g_0(\theta)\, d\theta \right| \right\}$$

$$\geq [2(1 + e^{2M})]^{-1},$$

which implies the corollary, provided that

$$\liminf \left| d_n' \int_a^b (\cos(d_n'\theta - b_n') - e_n') g_0(\theta)\, d\theta \right| > 0,$$

since $d_n/d_n' \to 1$. Integrating by parts, we have by the Riemann–Lebesgue lemma

$$\begin{aligned} d_n' \int_a^b &(\cos(d_n'\theta - b_n') - e_n') g_0(\theta)\, d\theta \\ &= g_0(\theta)\sin(d_n'\theta - b_n')\Big|_a^b - \int_a^b \sin(d_n'\theta - b_n') g_0'(\theta)\, d\theta \\ &\quad + \int_a^b g_0(\theta)\, d\theta \int \sin(d_n'\theta - b_n') g_0'(\theta)\, d\theta \\ &= g_0(b)\sin(2k_n\pi + \pi/2) - g_0(a)\sin(-2k_n\pi - \pi/2) + o(1) \\ &= g_0(a) + g_0(b) + o(1). \end{aligned}$$

□

EXAMPLE 1 (Normal, continuation). Similar to the case of estimating mixing density

$$\inf_{n \geq 1} \sup_{G \in \mathscr{G}_3'(a, b, \delta, M_3)} P_G\{\sqrt{\log n}\, |\tilde{G}_n - (G(b) - G(a))| > \varepsilon\} > \varepsilon, \quad \text{for some } \varepsilon > 0.$$

Again, our estimators achieve the optimal rate of convergence $(\log n)^{-1/2}$ for the class $\mathscr{G}_3$.

EXAMPLE 2 (Cauchy, continuation). Similar to the case of estimating mixing density

$$\inf_{n\ge 1}\sup_{G\in\mathscr{G}_3^*(a,b,\delta,M_3)} P_G\{(\log n)|\tilde{G}_n - (G(b) - G(a))| > \varepsilon\} > \varepsilon, \quad \text{for some } \varepsilon > 0.$$

Our estimators also achieve the optimal rate of convergence $(\log n)^{-1}$ for the class $\mathscr{G}_3$.

EXAMPLE 3 (Double exponential, continuation). For this case we have

$$\inf_{n\ge 1}\sup_{G\in\mathscr{G}_3^*(a,b,\delta,M_3)} P_G\{n^{1/4}|\tilde{G}_n - (G(b) - G(a))| > \varepsilon\} > \varepsilon, \quad \text{for some } \varepsilon > 0.$$

APPENDIX

PROOF OF THEOREM 3. Since $\int|x|f_0(x)\,dx < \infty$, $J_n(x,a)$ is well defined and $K_n(x,\cdot)$ is Riemann integrable. Since the Fourier transformation of $H - G$ is $(G^* - H^*)/(it)$, by (22) and the Plancherel identity

$$(2\pi)E\|\hat{G}_n - G\|^2 \le \int E\left|\frac{k^*(t/c_n)F_n^*(t) - f_0^*(t)G^*(t)}{itf_0^*(t)}\right|^2 dt.$$

Let $\theta_1 \sim G$ and $X_1|\theta_1 \sim f(\cdot|\theta_1)$. Since $E[e^{itX_1}|\theta_1] = f_0^*(t)e^{it\theta_1}$, we have

$$\begin{aligned}\operatorname{Var}(e^{itX_1}) &= E|e^{itX_1} - f_0^*(t)e^{it\theta_1}|^2 + \operatorname{Var}(f_0^*(t)e^{it\theta_1})\\ &= E|e^{it(X_1-\theta_1)} - F_0^*(t)|^2 + |f_0^*(t)|^2E|e^{it\theta_1} - G^*(t)|^2.\end{aligned}$$

It follows from the Plancherel identity that

$$\begin{aligned}(2\pi)^{-1}\int E|e^{it\theta_1} - G^*(t)|^2t^{-2}\,dt &= E\int|G_1(\theta) - G(\theta)|^2\,d\theta\\ &= \int G(\theta)(1 - G(\theta))\,d\theta,\end{aligned}$$

$$(2\pi)^{-1}\int E|e^{it(X_1-\theta_1)} - F_0^*(t)|^2t^{-2}\,dt = \int F_0(x)(1 - F_0(x))\,dx,$$

so that $E\|\hat{G}_n - G\|^2$ is bounded by

$$\begin{aligned}&\int\frac{|k^*(t/c_n)|^2\operatorname{Var}(F_n^*(t))}{2\pi|tf_0^*(t)|^2}dt + \int\frac{|k^*(t/c_n)EF_n^*(t) - f_0^*(t)G^*(t)|^2}{2\pi|tf_0^*(t)|^2}dt\\ &\le \int\frac{|k^*(t/c_n)|^2\operatorname{Var}(e^{itX_1})}{2\pi n|tf_0^*(t)|^2}dt + \int\frac{|(k^*(t/c_n) - 1)f_0^*(t)G^*(t)|^2}{2\pi|tf_0^*(t)|^2}dt\\ &\le \frac{\int F_0(x)(1 - F_0(x))\,dx}{n|f_0^*(c_n)|^2} + \int G(\theta)(1 - G(\theta))\,d\theta/n\\ &\quad + \int\frac{|G^*(t)(k^*(t/c_n) - 1)|^2}{2\pi t^2}dt.\end{aligned}$$

Hence, (23) and (24) follow from

$$\int |G^*(t)(k^*(t/c_n) - 1)/t|^2\,dt \le \int_{-c_n}^{c_n} + 2\int_{c_n}^{\infty} \le \frac{2}{c_n}\left[\left(\int |x|k(x)\,dx\right)^2 + 1\right],$$

$$\int |G^*(t)(k^*(t/c_n) - 1)/t|^2\,dt \le c_n^{-2}\left(\int |x|k(x)\,dx\right)^2 \|G^*(t)\|^2. \qquad \square$$

PROOF OF THEOREM 4. By (21)

$$\begin{aligned}(2\pi)^2 \operatorname{Var}(J_n(X,a)) &\le E\left|\int e^{-ita}k^*(t/c_n)[e^{itX} - 1]/(tf_0^*(t))\,dt\right|^2\\ &\le |f_0^*(c_n)|^{-2} E\left\{2\int_0^{c_n}|(e^{itX} - 1)/t|\,dt\right\}^2.\end{aligned}$$

Since $|e^{i2\pi t} - 1|^2 = 2(1 - \cos(2\pi t)) = 4\sin^2(\pi t)$ and

$$\begin{aligned}\int_0^x |\sin(\pi t)/t|\,dt &\le \int_0^1 \sin(\pi t)/t\,dt + \sum_{1\le k<x} k^{-1}\int_0^1 \sin(\pi t)\,dt\\ &\le \int_0^1 (\pi t - (\pi t)^3/3! + (\pi t)^5/5!)/t\,dt + (2/\pi)(1 + \log^+ x)\\ &\le \pi + (2/\pi)\log^+ x,\end{aligned}$$

$$\int_0^c |(e^{itx} - 1)/t|\,dt = 2\int_0^{cx/2\pi}|\sin(\pi s)/s|\,ds \le 2\pi + (4/\pi)\log^+(cx/2\pi), \qquad \forall\, x \ge 0.$$

It follows that

$$\begin{aligned}(2\pi)^2 \operatorname{Var}(J_n(X,a)) &\le |f_0^*(c_n)|^{-2} E\{4\pi + (8/\pi)\log^+(c_n|X|/(2\pi))\}^2\\ &\le |f_0^*(c_n)|^{-2}(8/\pi + o(1))^2(\log^+ c_n)^2.\end{aligned}$$

Hence, by (26)

$$\begin{aligned}&c_n^2 E_G|\hat{G}_n(a) - \underline{G}(a)|^2\\ &\quad\le c_n^2 \operatorname{Var}(J_n(X,a))/n + c_n^2|EJ_n(X,a) - \underline{G}(a)|^2\\ &\quad\le (2\pi)^{-2}c_n^2|f_0^*(c_n)|^{-2}(8/\pi + o(1))^2(\log^+ c_n)^2/n\\ &\qquad + \left|c_n\int G(\theta)c_n k(c_n(a - \theta))\,d\theta - \underline{G}(a)\right|^2\\ &\quad\le c_0(4/\pi^2)^2 + o(1)\\ &\qquad + \left|\int_0^\infty c_n[G(a + x/c_n) + G(a - x/c_n) - 2\underline{G}(a)]k(x)\,dx\right|^2\\ &\quad\le c_0(4/\pi^2)^2 + o(1) + \left(\int M_3|x|k(x)\,dx\right)^2. \qquad \square\end{aligned}$$

PROOF OF THEOREM 6. We shall first prove (39) under the additional assumption that $E_1|\Lambda_n| \le M$, where $\Lambda_n = \sum_{j=1}^n \log(f(X_j;G_2)/f(X_j;G_1))$ is the log-likelihood ratio. By symmetry we may assume $g_2(a) \ge g_1(a)$. Let p be the left-hand side of (39) and R be the event that $\tilde{g}_n(a) > [g_1(a) + g_2(a)]/2$. Since $d_n[g_2(a) - g_1(a)] = (w/4)(\cos(d_n a - b_n) - e_n)g_0(a)$, it follows from the definition of p that $P_1\{R\} \le p$ and $P_2\{R^c\} \le p$. Therefore,

$$\begin{aligned}1/2 &\le P_1\{|\Lambda_n| \le 2M\} \le P_1\{R\} + P_1\{R^c, |\Lambda_n| \le 2M\}\\ &\le p + E_1 e^{\Lambda_n + 2M} I\{R^c\} = p + P_2\{R^c\}e^{2M} \le p(1 + e^{2M}).\end{aligned}$$

Let us prove that $E_1|\Lambda_n| \le M$. Set $v(x) = f(x;G_2)/f(x;G_1) - 1$. Since $d_n \ge 1$ by (38), it follows from (35) and (36) that $|v(x)| \le wf(x,G_0)/(2f(x;G_1)) \le 1/2$, which implies that

$$\sum_{j=1}^n v(X_j) - 2\sum_{j=1}^n v^2(X_j) \le \Lambda_n = \sum_{j=1}^n \log\bigl(1 + v(X_j)\bigr) \le \sum_{j=1}^n v(X_j).$$

Since $f(x;G_2)/f(x;G_1)$ is a likelihood ratio, $E_1 v(X) = 0$, so that

$$\begin{aligned}E_1|\Lambda_n| &\le E_1 \sum_{j=1}^n v(X_j) + 2E_1\sum_{j=1}^n v^2(X_j)\\ &\le 2nE_1 v^2(X) + \sqrt{nE_1 v^2(X)}.\end{aligned} \tag{43}$$

It follows from (38) and (37) that

$$\begin{aligned}|e_n| &= \left|\int\int f(x|\theta)\mu(dx)\cos(d_n\theta - b_n)g_0(\theta)\,d\theta\right|\\ &\le \left|\int\int f(x|\theta)\mu(dx)e^{i(d_n\theta - b_n)}g_0(\theta)\,d\theta\right|\\ &\le \int \rho(d_n)u(x)f(x;G_0)\mu(dx) = \rho(d_n)E_0 u(X)\end{aligned}$$

and that

$$\begin{aligned}|v(x)| &= |f(x;G_2)/f(x;G_1) - 1|\\ &\le [w/(4d_n)][\rho(d_n)u(x)f(x;G_0) + |e_n|f(x;G_0)]/f(x;G_1)\\ &\le [\rho(d_n)u(x) + |e_n|]/(4d_n) \le n^{-1/2}[u(x) + E_0 u(X)]/4.\end{aligned} \tag{44}$$

Hence, $E_1|\Lambda_n| \le M$ by (43) and (44).

Let us prove (40). It follows from (39) that

$$\begin{aligned}&2\max_{j=1,2} d_n^2 E_j\|\tilde{g}_n - g_j\|^2\\ &\ge d_n^2\left(E_1\|\tilde{g}_n - g_1\|^2 + E_2\|\tilde{g}_n - g_2\|^2\right)\\ &\ge [2(1 + e^{2M})]^{-1}\int (w/8)^2|\cos(d_n\theta - b_n) - e_n|^2 g_0^2(\theta)\,d\theta.\end{aligned}$$

Let $h(\theta) = \min(g_0^2(\theta), C)$ for some $C > 0$. Then $\int h < \infty$. It follows from (38) and the Riemann–Lebesgue lemma that

$$\liminf \int |\cos(d_n\theta - b_n) - e_n|^2 g_0^2(\theta)\, d\theta$$

$$\geq \lim \int \cos^2(d_n\theta - b_n) h(\theta)\, d\theta$$

$$= \lim \int (1 + \cos(2d_n\theta - 2b_n)) h(\theta)\, d\theta/2 = \int h(\theta)\, d\theta/2 > 0.$$

The proof is complete since $\int |\cos(d_n\theta - b_n) - e_n|^2 g_0^2(\theta)\, d\theta > 0$, $\forall\ n$. □

We use the following lemma to check the conditions (15) and (34) for Example 3.

LEMMA 1. *Suppose* $f_0(x) = e^{-|x|}/2$. *Let* $K_n(x, a)$ *and* $I_n(x, a)$ *be given by* (19) *and* (30). *Then*

$$\lim_{n\to\infty} c_n^{-5} E_G |K_n(X, a)|^2 = (2\pi)^{-1} f(a; G) \|t^2 k^*(t)\|^2,$$

$$\lim_{n\to\infty} c_n^{-3} E_G |I_n(X, a)|^2 = (2\pi)^{-1} (f(a; G) + f(b; G)) \|t k^*(t)\|^2.$$

PROOF. We shall first prove the lemma based on the fact that for any real-valued function $h(x)$ satisfying $\int x^2 |h(x)|\, dx < \infty$ and $h^*(t) = 0$, $\forall |t| > 1$,

$$\lim_{c\to\infty} (2\pi c)^{-1} \int\int e^{-i(t-s)a} f^*(t-s) h^*(t/c) \overline{h^*(s/c)}\, dt\, ds = f(a) \|h^*(t)\|^2, \tag{45}$$

$$\lim_{c\to\infty} (2\pi c)^{-1} \int\int e^{-i(t-s)\xi} f^*(t-s) \sin(t\tau) \sin(s\tau) h^*(t/c) \overline{h^*(s/c)}\, dt\, ds = (f(a) + f(b)) \|h^*(t)\|^2/4, \tag{46}$$

where $\xi = (a+b)/2$ and $\tau = (b-a)/2$. Since $\int \sin(t(x-a)) k^*(t/c_n) \times (1+t^2)\, dt = 0$, by (45)

$$\begin{aligned} EK_n^2(X, a) &= (2\pi)^{-2} \int\int e^{-i(t-s)a} f^*(t-s)(1+t^2)(1+s^2) \\ &\quad \times k^*(t/c_n) k^*(s/c_n)\, dt\, ds \\ &= (2\pi)^{-2} c_n^4 \int\int e^{-i(t-s)a} f^*(t-s)(t/c_n)^2 k^*(t/c_n)(s/c_n)^2 \\ &\quad \times k^*(s/c_n)\, dt\, ds + o(c_n^5) \\ &= (2\pi)^{-1} c_n^5 f(a; G) \|t^2 k^*(t)\|^2 + o(c_n^5). \end{aligned}$$

Set $h(x) = k'(x)$. By a similar argument $EI_n^2(X, a, b)$ is equal to

$$\pi^{-2} \int\int e^{-i(t-s)\xi} f^*(t-s) \frac{\sin(\tau t)}{t} \frac{\sin(\tau s)}{s} (1+t^2)(1+s^2)$$

$$\times k^*(t/c_n) k^*(s/c_n)\, dt\, ds$$

$$= \frac{c_n^2}{\pi^2} \int\int e^{-i(t-s)a} f^*(t-s) \sin(\tau t) \sin(\tau s)$$

$$\times (t/c_n) k^*(t/c_n)(s/c_n) k^*(s/c_n)\, dt\, ds + o(c_n^3)$$

$$= \frac{2c_n^3}{\pi} (f(a;G) + f(b;G)) \|h^*(t)\|^2/4 + o(c_n^3)$$

$$= \frac{c_n^3}{2\pi} (f(a;G) + f(b;G)) \|tk^*(t)\|^2 + o(c_n^3).$$

Now let us prove (45) and (46). Let $u = t + s$ and $v = t - s$. Then

$$(2\pi)^{-1} \int\int e^{-i(t-s)a} f^*(t-s) h^*(t/c) \overline{h^*(s/c)}\, dt\, ds$$

$$= (2\pi)^{-1} \int\int e^{-iva} f^*(v) h^*((u+v)/(2c)) h^*((v-u)/(2c))\, du\, dv/2.$$

Set $h_1(x) = (2\pi) h^2(x)$. Then $h_1^*(t) = \int h^*(t-s) h^*(s)\, ds$ and

$$\int h^*((u+v)/(2c)) h^*((v-u)/(2c))\, du$$

$$= 2c \int h^*(w) h^*(v/c - w)\, dw = 2c h_1^*(v/c),$$

so that

$$(c2\pi)^{-1} \int\int e^{-i(t-s)a} f^*(t-s) h^*(t/c) \overline{h^*(s/c)}\, dt\, ds$$

$$= (2\pi)^{-1} \int e^{-iva} f^*(v) h_1^*(v/c)\, dv$$

$$= \int f(a - x/c) h_1(x)\, dx = f(a) \int h_1(x)\, dx + o(1)$$

$$= f(a) h_1^*(0) + o(1) = f(a) \|h^*\|^2 + o(1).$$

This proves (45). For (46) we have

$$(2\pi)^{-1}\int\int e^{-i(t-s)\xi}f^*(t-s)\sin(t\tau)\sin(s\tau)h^*(t/c)\overline{h^*(s/c)}\,dt\,ds$$

$$= (2\pi)^{-1}\int\int e^{-iv\xi}f^*(v)[\cos(v\tau)-\cos(u\tau)]h^*((u+v)/(2c)) \times h^*((v-u)/(2c))\,du\,dv/4.$$

Since $\{h^*(w)h^*(z-w),\ -\infty < z < \infty\}$ is uniformly continuous with support $[-1,1]$, by the Riemann–Lebesgue lemma

$$\sup_v\left|\int\cos(u\tau)h^*((u+v)/(2c))h^*((v-u)/(2c))\,du\right|$$

$$= 2c\sup_v\left|\int\cos(\tau(2cw-v))h^*(w)h^*(v/c-w)\,dw\right| = o(c),$$

which implies that

$$(c2\pi)^{-1}\int\int e^{-i(t-s)\xi}f^*(t-s)\sin(t\tau)\sin(s\tau)h^*(t/c)\overline{h^*(s/c)}\,dt\,ds$$

$$= (2\pi)^{-1}\int e^{-iv\xi}f^*(v)\cos(v\tau)h_1^*(v/c)\,dv/2 + o(1)$$

$$= (2\pi)^{-1}\int[e^{-iva}+e^{-ivb}]f^*(v)h_1^*(v/c)\,dv/4 + o(1)$$

$$= [f(a)+f(b)]\|h^*\|^2/4 + o(1). \qquad \square$$

Acknowledgments. I would like to thank Professor Herbert Robbins who introduced this problem to me. Also, I would like to thank the associate editor and a referee for many helpful suggestions.

REFERENCES

CARROLL, R. J. and HALL, P. (1988). Optimal rates of convergence for deconvolving a density. *J. Amer. Statist. Assoc.* **83** 1184–1186.

DEELY, J. J. and KRUSE, R. L. (1968). Construction of sequences estimating the mixing distribution. *Ann. Math. Statist.* **39** 286–288.

EDELMAN, D. (1988). Estimation of the mixing distribution for a normal mean with applications to the compound decision problem. *Ann. Statist.* **16** 1609–1622.

HÁJEK, J. (1972). Local asymptotic minimax and admissibility in estimation. *Proc. Sixth Berkeley Symp. Math. Statist. Probab.* **1** 175–194. Univ. California Press, Berkeley.

JEWELL, N. P. (1982). Mixtures of exponential distributions. *Ann. Statist.* **10** 479–484.

KIEFER, J. and WOLFOWITZ, J. (1956). Consistency of the maximum likelihood estimator in the presence of infinitely many nuisance parameters. *Ann. Math. Statist.* **27** 887–906.

LAIRD, N. (1978). Nonparametric maximum likelihood estimation of a mixing distribution. *J. Amer. Statist. Assoc.* **73** 805–811.

LE CAM, L. (1972). Limits of experiments. *Proc. Sixth Berkeley Symp. Math. Statist. Probab.* **1** 245–261. Univ. California Press, Berkeley.

LINDSAY, B. G. (1983a). The geometry of mixture likelihoods: A general theory. *Ann. Statist.* **11** 86–94.

LINDSAY, B. G. (1983b). The geometry of mixture likelihoods. II. The exponential family. *Ann. Statist.* **11** 783–792.

ROBBINS, H. (1950). A generalization of the method of maximal likelihood: Estimating a mixing distribution (Abstract). *Ann. Math. Statist.* **21** 314–315.

ROBBINS, H. (1951). Asymptotically subminimax solutions of compound statistical decision problems. *Proc. Second Berkeley Symp. Math. Statist. Probab.* **1** 131–148. Univ. California Press, Berkeley.

ROBBINS, H. (1955). An empirical Bayes approach to statistics. *Proc. Third Berkeley Symp. Math. Statist. Probab.* **1** 157–163. Univ. California Press, Berkeley.

ROLPH, J. E. (1968). Bayesian estimation of mixing distributions. *Ann. Math. Statist.* **39** 1289–1302.

STEFANSKI, L. A. and CARROLL, R. J. (1988). Deconvolving kernel density estimators. Preprint.

DEPARTMENT OF STATISTICS
HILL CENTER FOR THE MATHEMATICAL SCIENCES
BUSCH CAMPUS
RUTGERS UNIVERSITY
NEW BRUNSWICK, NEW JERSEY 08903