ESTIMATING THE MIXING DENSITY OF A MIXTURE OF POWER SERIES DISTRIBUTIONS

by

Wei-Liem Loh
Purdue University

Technical Report #92-14

Department of Statistics
Purdue University

April 1992

ESTIMATING THE MIXING DENSITY OF A MIXTURE OF POWER SERIES DISTRIBUTIONS[1]

BY WEI-LIEM LOH

Purdue University

Let $X_1, X_2, \cdots, X_n$ be independent and identically distributed observations from a mixture of power series distributions. Based on this random sample, we consider the problem of estimating the mixing density of the mixture distribution. A mixing density kernel estimator is proposed which, under mild assumptions, has $1/\log n$ as an upper bound for its rate of convergence to the true density under squared error loss. It is also shown that the optimal rate of convergence cannot exceed $1/n^r$ for any constant r.

1 Introduction

Let X be an integer-valued random variable with the following mixture distribution.

$$(1) \qquad P(X=j) = \int_0^\alpha x^j \phi(x)\psi(j)dG(x), \quad \forall j = 0,1,2,\cdots,$$

where G is a probability distribution function on $[0,\alpha]$ with α known, and $\phi : [0,\alpha] \to [0,\infty)$, $\psi : \{0,1,2,\cdots\} \to (0,\infty)$ are known functions. We assume that $x^j\phi(x)\psi(j)$ is a probability mass function in j for each $x \in [0,\alpha]\backslash\{y : \phi(y) = 0\}$, and the restriction of G to the set $\{y : \phi(y) = 0\}$ is a known measure (without loss of generality we assume that $G(\{y : \phi(y) = 0\}) < 1$ for otherwise G is completely known) if the cardinality of $\{y : \phi(y) = 0\}$ is greater than or equal to two. Here we adopt the convention that $0/0 = 0^0 = 1$. Under these assumptions, Lemma 3 (see Appendix) shows that this mixture distribution is identifiable in G.

We observe that (1) is a mixture of power series distributions [see, for example, Johnson and Kotz (1969) page 33]. This class is quite broad and includes the common mixture distributions listed below.

[1]Research supported in part by NSF Grant DMS 89-23071.

AMS 1980 *subject classifications.* Primary 62G05; Secondary 62E20.

Key words and phrases. Fourier analysis, mixing density, mixture of power series distributions, method of kernels, rate of convergence.

EXAMPLE 1 (Compound Poisson distribution). The random variable X is said to have a compound Poisson distribution with mixing distribution G having support on $[0, \alpha]$ if

$$P(X = j) = \int_0^\alpha e^{-\lambda}(\lambda^j / j!) dG(\lambda), \quad \forall j = 0, 1, 2, \cdots.$$

Here we take $\phi(x) = e^{-x}$ and $\psi(j) = 1/j!$. We remark that Lambert and Tierney (1984) also made a similar assumption that the support of the mixing distribution has a known finite upper bound.

EXAMPLE 2 (Negative binomial mixture). X is said to have a negative binomial mixture distribution with parameter $\nu \in \{1, 2, \cdots\}$ and mixing distribution G if

$$P(X = j) = \int_0^1 C_j^{j+\nu-1} p^j (1-p)^\nu dG(p), \quad \forall j = 0, 1, 2, \cdots,$$

where $C_j^{j+\nu-1} = (j+\nu-1)!/[j!(\nu-1)!]$. Here $\alpha = 1$, $\phi(x) = (1-x)^\nu$ and $\psi(j) = C_j^{j+\nu-1}$. In the case where $\nu = 1$, we have a mixture of geometric distributions.

EXAMPLE 3 (Mixture of logarithmic series distributions). X has a logarithmic series mixture distribution with mixing distribution G if

$$P(X = j) = -\int_0^1 x^{j+1}[(j+1)\log(1-x)]^{-1} dG(x), \quad \forall j = 0, 1, 2, \cdots.$$

In this instance, we take $\alpha = 1$, $\phi(x) = -x/\log(1-x)$ and $\psi(j) = 1/(j+1)$.

COUNTEREXAMPLE (Binomial mixture). X has a binomial mixture distribution with N fixed and mixing distribution G if

$$P(X = j) = \int_0^1 C_j^N p^j (1-p)^{N-j} dG(p), \quad \forall j = 0, 1, \cdots, N.$$

However the binomial mixture does not belong to (1). Indeed Teicher (1961), page 247, showed that it is not even identifiable.

Now let $X_1, X_2, \ldots, X_n$ be independent and identically distributed observations having the same distribution as that of X in (1). Assuming that $g(a) = G'(a)$ exists for some $0 < a < \alpha$, the main purpose of this paper is to discuss how well $g(a)$ can be estimated nonparametrically on the basis of $X_1, X_2, \ldots, X_n$ under squared error loss. In addition, the difficulty of estimating the whole mixing density (assuming that it exists) is also investigated with respect to a rather natural weighted squared error loss.

Among mixture problems, the deconvolution problem appears to be the best understood. Recent and important advances to the solution were made by Devroye and Wise (1979), Carroll and Hall (1988), Stefanski (1990), Zhang (1990), Fan (1991a), (1991b), (1991c) and many others using the techniques of Fourier analysis. In particular, kernel estimators for the mixing density (distribution) have been obtained which achieve the optimal rate of convergence to the true mixing density (distribution) respectively. Indeed, as the reader will probably observe, the present work is significantly influenced by the papers of Zhang and Fan.

Another problem that has been of much interest is the estimation of the mixing distribution of a compound Poisson distribution. Tucker (1963) approached this problem via the method of moments and Simar (1976) proposed a nonparametric maximum likelihood estimator for the mixing distribution. However, although these estimators have been shown to be consistent, their convergence rates are not known.

Other mixture problems were studied by Robbins (1964), Deely and Kruse (1968), Blum and Susarla (1977), Jewell (1982) and Lindsay (1983a), (1983b), (1989) among others. Rolph (1968) and Meeden (1972) used Bayesian methods to construct consistent estimators for the mixing distribution. Again the convergence rates of these estimators have not been worked out.

The rest of this paper is organized as follows. In Section 2, a kernel estimator $\hat{g}_n$ for the mixing density g is proposed. In addition, local and global upper bounds for the rate of convergence of $\hat{g}_n$ to g are obtained via Fourier analysis. Section 3 gives complementary local and global lower bounds for the optimal rate of convergence. Finally the Appendix contains proofs of a few somewhat technical lemmas that are needed in previous sections.

2 A mixing density kernel estimator

Let $0 < a < \alpha$ such that $g(a) = G'(a)$ exists and $\phi(a) > 0$. In this section, a kernel estimator for $g(a)$ will be constructed and studied using techniques of Fourier analysis. Let $k(x)$ be a probability density function, with characteristic function $k^*(t)$, such that

$$\begin{aligned} k^*(t) &= 0, \quad \forall |t| > 1, \\ k(x) &= k(-x), \quad \forall x \in R, \\ \int_{-\infty}^{\infty} x^2 k(x)dx &< \infty, \end{aligned} \tag{2}$$

$$\int_{-\infty}^{\infty} |xk'(x)|dx \quad < \quad \infty.$$

Zhang (1990) has observed the existence of such a $k(x)$; an example being $k(x) = (6/\pi)[(2/x)\sin(x/4)]^4$. Now define

$$K_n(j,x) = (2\pi)^{-1}\mathcal{R}\{\int_{-\sigma_n}^{\sigma_n} k^*(\sigma_n^{-1}t)e^{-itx}(it)^j[j!\psi(j)]^{-1}dt\},$$

where $\mathcal{R}\{z\}$ denotes the real part of the complex number z, and $\{\sigma_n\}$ is a suitably chosen sequence of constants. Based on a random sample $X_1, X_2, \cdots, X_n$ of independent and identically distributed observations distributed as in (1), the mixing density kernel estimator for $g(a)$ is given by

$$\hat{g}_n(a) = [n\phi(a)]^{-1}\sum_{j=1}^{n} K_n(X_j, a). \tag{3}$$

The motivation for (3) is as follows. We observe that with X as in (1), we have

$$\sum_{j=0}^{\infty} \frac{(it)^j}{j!\psi(j)} P(X=j) = \sum_{j=0}^{\infty} \int_0^{\alpha} \frac{(itx)^j}{j!}\phi(x)dG(x), \qquad \forall t \in R.$$

Using Fubini's theorem, we have

$$E(it)^X[X!\psi(X)]^{-1} = \int_0^{\alpha} e^{itx}\phi(x)dG(x), \qquad \forall t \in R.$$

We observe that $\psi^{-1}(0)P(X=0) = \int_0^{\alpha}\phi(x)dG(x) > 0$ since $0^0 = 1$ and $G(\{x : \phi(x) = 0\}) < 1$. Let

$$\begin{aligned} C_g &= [\int_0^{\alpha} \phi(x)dG(x)]^{-1}, \\ h^*(t) &= C_g \int_0^{\alpha} e^{itx}\phi(x)dG(x), \qquad \forall t \in R, \end{aligned} \tag{4}$$

and $H(x)$ be the distribution function of the measure $C_g\phi(x)dG(x)$.

Then under mild regularity conditions, it follows from the Fourier inversion theorem that

$$(2\pi)^{-1}\int_{-\sigma_n}^{\sigma_n} k^*(\sigma_n^{-1}t)e^{-ita}h^*(t)dt \to C_g\phi(a)g(a),$$

as $\sigma_n \to \infty$ and hence

$$(2\pi)^{-1}\int_{-\sigma_n}^{\sigma_n} k^*(\sigma_n^{-1}t)e^{-ita}E(it)^X[X!\psi(X)]^{-1}dt$$
$$\to \quad \phi(a)g(a), \quad \text{as } \sigma_n \to \infty. \tag{5}$$

$\hat{g}_n(a)$ is obtained from the l.h.s. of (5) by replacing $E(it)^X[X!\psi(X)]^{-1}$ with $n^{-1}\sum_{j=1}^n (it)^{X_j}[X_j!\psi(X_j)]^{-1}$.

Theorem 1 *Let* $0 < a < \alpha$, $\sigma_n = (\beta\log n)^{1/2}$ *with* $0 < \beta < 1/\alpha$ *and* $\hat{g}_n(a)$ *be as in (3). Suppose that* $g(a) = G'(a)$ *exists,* $\inf_{j\geq 0} j!\psi(j) > 0$ *and* $\phi(a) > 0$. *Then*

$$E[\hat{g}_n(a) - g(a)]^2 = o(1), \quad \text{as } n \to \infty.$$

PROOF. With X as in (1), we observe from Fubini's theorem that

$$\begin{aligned} C_gEK_n(X,a) &= (2\pi)^{-1}\int_{-\sigma_n}^{\sigma_n} k^*(\sigma_n^{-1}t)e^{-ita}h^*(t)dt \\ &= -\int_{-\infty}^{\infty} \sigma_n k(x)dH(a-\sigma_n^{-1}x) \\ &= \int_{-\infty}^{\infty} \sigma_n[H(a-\sigma_n^{-1}x) - H(a)]k'(x)dx. \end{aligned} \tag{6}$$

Letting ε denote a sufficiently small positive constant, it follows from Taylor's expansion that the r.h.s. of (6) is equal to

$$\begin{aligned} &-C_g\phi(a)g(a)\int_{-\infty}^{\infty} xk'(x)dx + \varepsilon^{-1}O(1)\int_{\varepsilon\sigma_n}^{\infty} |xk'(x)|dx \\ &+o(1)\int_{-\infty}^{\infty} |xk'(x)|dx \\ \to \quad &-C_g\phi(a)g(a)\int_{-\infty}^{\infty} xk'(x)dx + o(1)\int_{-\infty}^{\infty} |xk'(x)|dx, \quad \text{as } n \to \infty. \end{aligned}$$

Here $O(1)$ pertains to the limit as $n \to \infty$ uniformly over small ε and $o(1)$ pertains to the limit as $\varepsilon \to 0$ uniformly over large n. Since ε can be arbitrarily small, we have $\lim_n EK_n(X,a) = \phi(a)g(a)$. Next we observe that

$$\begin{aligned} \mathrm{Var}[K_n(X,a)] &\leq EK_n^2(X,a) \\ &\leq (2\pi)^{-2}\int_{-\sigma_n}^{\sigma_n}\int_{-\sigma_n}^{\sigma_n} E|st|^X[X!\psi(X)]^{-2}ds\, dt \\ &\leq \{\sup_{j\geq 0}[j!\psi(j)]^{-1}\}\pi^{-2}\psi^{-1}(0)\sigma_n^2 e^{\alpha\sigma_n^2}. \end{aligned} \tag{7}$$

Consequently we conclude that

$$
\begin{aligned}
& E[\hat{g}_n(a) - g(a)]^2 \\
(8) \qquad = \;& \phi^{-2}(a)\{n^{-1}\mathrm{Var}[K_n(X,a)] + [EK_n(X,a) - \phi(a)g(a)]^2\} \\
\to \;& 0, \quad \text{as } n \to \infty.
\end{aligned}
$$

This completes the proof. □

Next let δ be a positive constant such that $[a-\delta, a+\delta] \subset (0,\alpha)$. We define $\mathcal{G}(a,\delta,M)$ to be the set of probability distributions G with support on $[0,\alpha]$ such that $g = G'$ exists on $[a-\delta, a+\delta]$ and that g satisfies the following Lipschitz condition at $x = a$: there exists a positive constant M such that

$$
(9) \qquad |g(a+\varepsilon) - g(a)| < M|\varepsilon|, \quad \forall |\varepsilon| \le \delta.
$$

We shall also assume that ϕ satisfies a Lipschitz condition at $x = a$, i.e. there exists a positive constant M_1 such that

$$
(10) \qquad |\phi(a+\varepsilon) - \phi(a)| < M_1|\varepsilon|, \quad \forall |\varepsilon| \le \delta.
$$

Theorem 2 *Let $\sigma_n = (\beta \log n)^{1/2}$ with $0 < \beta < 1/\alpha$ and $\hat{g}_n$ be as in (3). Suppose that (10) holds, $\inf_{j\ge 0} j!\psi(j) > 0$ and $\phi(a) > 0$. Then*

$$
\sup_{G \in \mathcal{G}(a,\delta,M)} E[\hat{g}_n(a) - g(a)]^2 = O(1/\log n), \quad \text{as } n \to \infty.
$$

PROOF. Let $\varepsilon_1(y) = \sup_{x \ge y} x^2 k(x)$ and $\varepsilon_2(y) = \int_y^\infty x k(x) dx$, for all $y > 0$. It follows from (2) that $\varepsilon_1(y)$ and $\varepsilon_2(y)$ both tend to 0 as $y \to \infty$. Let $G \in \mathcal{G}(a,\delta,M)$ such that $g(x) = G'(x)$ whenever $x \in [a-\delta, a+\delta]$. Then with X as in (1), we observe that

$$
\begin{aligned}
& \sigma_n|EK_n(X,a) - \phi(a)g(a)| \\
\le \;& \sigma_n\left|\int_{a-\delta}^{a+\delta} \sigma_n k[\sigma_n(a-x)]\phi(x)g(x)dx - \phi(a)g(a)\right| + \varepsilon_1(\sigma_n\delta)\delta^{-2}C_g^{-1} \\
\le \;& \sigma_n\left|\int_{-\sigma_n\delta}^{\sigma_n\delta} k(x)[\phi(a+\sigma_n^{-1}x)g(a+\sigma_n^{-1}x) - \phi(a)g(a)]dx\right| \\
& +\varepsilon_1(\sigma_n\delta)\delta^{-2}C_g^{-1} + 2\phi(a)g(a)\delta^{-1}\varepsilon_2(\sigma_n\delta) \\
\le \;& \sigma_n \int_{-\sigma_n\delta}^{\sigma_n\delta} k(x)[\phi(a)|g(a+\sigma_n^{-1}x) - g(a)| + g(a+\sigma_n^{-1}x)|\phi(a+\sigma_n^{-1}x) \\
& -\phi(a)|]dx + \varepsilon_1(\sigma_n\delta)\delta^{-2}C_g^{-1} + 2\phi(a)g(a)\delta^{-1}\varepsilon_2(\sigma_n\delta) \\
\le \;& \{M\phi(a) + M_1[g(a) + M\delta]\}\int_{-\infty}^{\infty} |x|k(x)dx \\
& +\varepsilon_1(\sigma_n\delta)\delta^{-2}\psi^{-1}(0) + 2\phi(a)g(a)\delta^{-1}\varepsilon_2(\sigma_n\delta).
\end{aligned}
$$

The last inequality uses (9), (10) and $C_g \geq \psi(0)$. However from the Lipschitz condition on g at $x = a$ [see Zhang (1990) page 814], we have $g(a) \leq (1 + M\delta^2)/(2\delta)$. Thus we conclude that

$$\sigma_n |EK_n(X, a) - \phi(a)g(a)| = O(1), \tag{11}$$

as $n \to \infty$ uniformly over $G \in \mathcal{G}(a, \delta, M)$. Also as in (7), we have

$$\mathrm{Var}[K_n(X, a)] = O(\sigma_n^2 e^{\alpha\sigma_n^2}), \tag{12}$$

as $n \to \infty$ uniformly over $G \in \mathcal{G}(a, \delta, M)$. The theorem now follows from (8), (11) and (12). □

REMARK. We observe that the conditions $\phi(a) > 0$ and $\inf_{j \geq 0} j!\psi(j) > 0$ are satisfied by Examples 1, 2 and 3 of the first section. Consequently Theorems 1 and 2 apply to these cases.

Now assume that $g = G'$ exists on $[0, \alpha]$. We shall conclude this section with an upper bound for the global rate of convergence of $\hat{g}_n$ to the true mixing density g under the following weighted squared error loss:

$$L(g, \hat{g}_n) = \int_0^\alpha [\hat{g}_n(x) - g(x)]^2 \phi^2(x) dx.$$

It appears that this is a more natural loss function to consider than the usual unweighted L_2 loss as in the computation of mixture probabilities, see (1), the distribution G is weighted by the function ϕ. For simplicity we define $h : R \to R$ by

$$h(x) = \begin{cases} C_g \phi(x) g(x) & \text{if } 0 \leq x \leq \alpha, \\ 0 & \text{otherwise}, \end{cases}$$

where C_g is as in (4) and for $M > 0$,

$$\mathcal{F}(M) = \{G : g = G' \text{ and } h' \text{ exist on } [0, \alpha] \text{ and } R \text{ respectively with } \int_0^\alpha [h'(x)]^2 dx < M\}.$$

Theorem 3 *Let $\sigma_n = (\beta \log n)^{1/2}$ with $0 < \beta < 1/\alpha$ and $\hat{g}_n$ be as in (3). Suppose that $\inf_{j \geq 0} j!\psi(j) > 0$ and $\mathcal{F}(M)$ is nonempty. Then*

$$\sup_{G \in \mathcal{F}(M)} E \int_0^\alpha [\hat{g}_n(x) - g(x)]^2 \phi^2(x) dx = O(1/\log n), \quad \text{as } n \to \infty.$$

PROOF. Let $G \in \mathcal{F}(M)$ and $g = G'$. With X as in (1), we observe from Plancherel's identity and Fubini's theorem that

$$\begin{aligned} E\int_0^\alpha K_n^2(X,x)dx &\leq (2\pi)^{-1}E\int_{-\sigma_n}^{\sigma_n}\{k^*(\sigma_n^{-1}t)t^X[X!\psi(X)]^{-1}\}^2dt \\ &= O(1)\int_{-\sigma_n}^{\sigma_n}Et^{2X}[X!\psi(X)]^{-1}dt \\ &= O(\sigma_n e^{\alpha\sigma_n^2}), \qquad (13) \end{aligned}$$

as $n \to \infty$ uniformly over $G \in \mathcal{F}(M)$. From the Fourier inversion theorem and Taylor's expansion, we get

$$\begin{aligned} EK_n(X,x) &= C_g^{-1}(2\pi)^{-1}\int_{-\sigma_n}^{\sigma_n}k^*(\sigma_n^{-1}t)e^{-itx}h^*(t)dt \\ &= C_g^{-1}\int_{-\infty}^{\infty}k(y)h(x-\sigma_n^{-1}y)dy \\ &= \phi(x)g(x) - C_g^{-1}\sigma_n^{-1}\int_{-\infty}^{\infty}\int_0^1 yk(y)h'(x-\sigma_n^{-1}ty)dtdy. \end{aligned}$$

Hence it follows from Jensen's inequality and $C_g \geq \psi(0)$ that

$$\begin{aligned} &\int_0^\alpha[EK_n(X,x)-\phi(x)g(x)]^2dx \\ &= O(\sigma_n^{-2})\int_{-\infty}^{\infty}\int_0^1|y|k(y)\{\int_0^\alpha[h'(x-\sigma_n^{-1}ty)]^2dx\}dtdy \\ &= O(\sigma_n^{-2})\int_{-\infty}^{\infty}|y|k(y)dy, \qquad (14) \end{aligned}$$

as $n \to \infty$ uniformly over $G \in \mathcal{F}(M)$. Consequently we conclude from (13) and (14) that

$$\begin{aligned} &\sup_{G\in\mathcal{F}(M)} E\int_0^\alpha[\hat{g}_n(x)-g(x)]^2\phi(x)^2dx \\ \leq\ &\sup_{G\in\mathcal{F}(M)}\{n^{-1}E\int_0^\alpha K_n^2(X,x)dx + \int_0^\alpha[EK_n(X,x)-\phi(x)g(x)]^2dx\} \\ =\ &O(1/\log n), \quad \text{as } n\to\infty. \end{aligned}$$

This completes the proof. □

REMARK. The hypotheses of Theorem 3 are satisfied by Examples 1, 2 and 3.

REMARK. We note that the condition G has support on $[0, \alpha]$ can be weakened to the condition that G has a sufficiently rapidly decreasing tail, for example a normal tail will suffice. However the convergence rate of $\hat{g}_n$ will then be significantly slower and a further disadvantage of using the latter condition is that this condition will most likely be unverifiable in practice.

3 Lower Bounds

In this section, we shall complement the results of the previous section by establishing local and global lower bounds for the optimal rate of convergence. To obtain a local lower bound, we shall use the techniques developed by Donoho and Liu (1987), (1991a) and (1991b) in a remarkable series of papers.

Let X be as in (1) and

$$P(Y = j) = \int_0^\alpha x^j \phi(x)\psi(j)dF(x), \qquad \forall j = 0, 1, 2 \cdots,$$

where F is a probability distribution on $[0, \alpha]$. We define the L_1 distance between the law of X and Y by

$$L_1(F, G) = \sum_{j=0}^{\infty} |P(Y = j) - P(X = j)|.$$

If F and G possess densities f and g respectively, we write $L_1(f, g) = L_1(F, G)$. Now let $r > 0$ and m be a positive integer. Let $G_0 \in \mathcal{G}(a, \delta, M)$ such that G_0 has a continuously differentiable probability density g_0 on $[0, \alpha]$ with $g_0(a) > 0$. Define

$$(15) \qquad p_N(x) = \begin{cases} (x^2 - 1)^2(1 + c_2x^2 + \cdots + c_{2N}x^{2N}) & \text{if } |x| \le 1, \\ 0 & \text{if } |x| > 1. \end{cases}$$

Then for sufficiently large N, there exist constants $c_2, \ldots, c_{2N}$ such that

$$(16) \qquad \int_{-1}^{1} x^j p_N(x)dx = 0, \qquad \forall j = 0, \ldots, m - 1.$$

(This involves solving a system of linear equations.) For such a p_N and for each $n \ge 1$, we write

$$g_n(x) = g_0(x) + n^{-r/2}p_N(n^{r/8}(x - a)), \qquad \forall 0 \le x \le \alpha.$$

Let $G_n(x) = \int_0^x g_n(y)dy$. Then for sufficiently large n, we observe that $G_n \in \mathcal{G}(a, \delta, M)$.

Lemma 1 *Let $r > 0$ and m be a positive integer satisfying $mr > 8$. Suppose that ϕ is m times continuously differentiable on an open interval containing a and that there exists $\varepsilon > 0$ such that*

$$\sum_{j=m}^{\infty} j^m (a+\varepsilon)^j \psi(j) < \infty. \tag{17}$$

Then for sufficiently large n, we have

$$L_1(G_0, G_n) \leq 1/n. \tag{18}$$

The proof of Lemma 1 is somewhat technical and is deferred to the Appendix. Let $\hat{T}_n$ be an estimator for $g(a) = G'(a)$ based on a sample of n independent and identically distributed observations having the same law as X. If (18) is satisfied, Donoho and Liu (1987) observed that

$$\inf_{\hat{T}_n} \max_{G \in \{G_0, G_n\}} P_G\{|\hat{T}_n - g(a)| \geq |g_0(a) - g_n(a)|/2\} > c,$$

for some positive constant c, and hence

$$\begin{aligned} \inf_{\hat{T}_n} \sup_{G \in \mathcal{G}(a,\delta,M)} E[\hat{T}_n - g(a)]^2 &> c[g_0(a) - g_n(a)]^2/4 \\ &= c/(4n^r), \end{aligned} \tag{19}$$

for any constant r.

Theorem 4 *Suppose that ϕ is infinitely differentiable on an open interval containing a and that for each $m \geq 1$, there exists $\varepsilon_m > 0$ such that*

$$\sum_{j=m}^{\infty} j^m (a+\varepsilon_m)^j \psi(j) < \infty. \tag{20}$$

Then with $\hat{T}_n$ as in (19), we have

$$\liminf_{n\to\infty} \inf_{\hat{T}_n} \sup_{G \in \mathcal{G}(a,\delta,M)} n^r E[\hat{T}_n - g(a)]^2 = \infty,$$

for any positive constant r.

PROOF. This is immediate from (19) and Lemma 1. □

REMARK. Donoho and Liu (1987) has also observed a result similar to that of Theorem 4 for the case of estimating the mixing distribution of a scale mixture of exponentials. However an explicit proof has not been provided by them.

REMARK. We observe that (20) and the infinite differentiability of ϕ on $(0, \alpha)$ are satisfied by Examples 1, 2 and 3 of the first section. Hence Theorem 4 is applicable to them.

We end this section with a corresponding global lower bound for the optimal rate of convergence. Let $r > 0$, m be an integer satisfying $mr > 8$ and p_N be as in (15) satisfying (16). Furthermore let $\tau_n = \lceil n^{r/8} \rceil$, where $\lceil x \rceil$ denotes the smallest integer greater than or equal to x.

We assume that ϕ' is continuous on $[0, \alpha]$. This implies that ϕ' is right continuous at 0 and left continuous at α. Then for M sufficiently large, there exists $G_\vartheta \in \mathcal{F}(M)$ such that G_ϑ has a continuously differentiable density g_ϑ on $[0, \alpha]$, $g_\vartheta(x) > 0$ on $(0, \alpha)$, $g_\vartheta(x) = b_1 x^2(1 + o(1))$ as $x \downarrow 0$ and $g_\vartheta(x) = b_2(\alpha - x)^2(1 + o(1))$ as $x \uparrow \alpha$ where b_1 and b_2 are nonzero constants. Following Fan (1991c), define for $0 \le \rho < \gamma \le \alpha$,

$$
\begin{aligned}
x_{n,l} &= \rho + \frac{\gamma - \rho}{\tau_n}\Big(l - \frac{1}{2}\Big), \quad \forall 1 \le l \le \tau_n, \\
\vartheta_n &= (\theta_1, \cdots, \theta_{\tau_n}) \in \{0,1\}^{\tau_n}, \\
g_{\vartheta_n}(x) &= g_\vartheta(x) + n^{-r/2} \sum_{l=1}^{\tau_n} \theta_l p_N\Big(4\tau_n \frac{x - x_{n,l}}{\gamma - \rho}\Big), \quad \forall 0 \le x \le \alpha, \qquad (21)
\end{aligned}
$$

and given ϑ_n, we write

$$\vartheta_{n,l,q} = (\theta_1, \cdots, \theta_{l-1}, q, \theta_{l+1}, \cdots, \theta_{\tau_n}),$$

whenever $q = 0, 1$. Writing $G_{\vartheta_n}(x) = \int_0^x g_{\vartheta_n}(y) dy$, we observe that $G_{\vartheta_n} \in \mathcal{F}(M)$ for sufficiently large n uniformly over $\vartheta_n \in \{0,1\}^{\tau_n}$.

Lemma 2 *Let $r > 0$, $mr > 8$ and g_{ϑ_n} be as in (21). Suppose that ϕ is m times continuously differentiable on $[\rho, \gamma]$ and*

$$\sum_{j=m}^{\infty} j^m \gamma^j \psi(j) < \infty.$$

Then for sufficiently large n,

$$\max_{1 \le l \le \tau_n} \max_{\vartheta_n} L_1(g_{\vartheta_{n,l,0}}, g_{\vartheta_{n,l,1}}) \le 1/n. \qquad (22)$$

The proof of Lemma 2 is deferred to the Appendix.

Proposition 1 *Let g_{ϑ_n} be as in (21). Suppose that ϕ is continuous on $[\rho, \gamma]$ and (22) is satisfied. Then for any estimator $\hat{T}_n$ for g based on a random sample $X_1, \cdots, X_n$ having the same law as X in (1), we have*

$$\inf_{\hat{T}_n} \max_{g \in \{g_{\vartheta_n} : \vartheta_n\}} E \int_\rho^\gamma [\hat{T}_n(x) - g(x)]^2 \phi^2(x) dx \geq c/n^r,$$

for some positive constant c.

The proof of the above proposition can essentially be found in Theorem 1 of Fan (1991c). In that paper, the χ^2 distance is used instead of the L_1 distance. However we observe that the proof can be easily adapted to accommodate the L_1 distance.

Theorem 5 *Suppose that ϕ' is continuous on $[0, \alpha]$, ϕ is infinitely differentiable on $[\rho, \gamma]$ and that for each $m \geq 1$,*

$$\sum_{j=m}^{\infty} j^m \gamma^j \psi(j) < \infty,$$

where $0 \leq \rho < \gamma \leq \alpha$. Then with $\hat{T}_n$ as in Proposition 1, we have for sufficiently large M,

$$\lim_{n \to \infty} \inf_{\hat{T}_n} \sup_{G \in \mathcal{F}(M)} n^r E \int_\rho^\gamma [\hat{T}_n(x) - g(x)]^2 \phi^2(x) dx = \infty,$$

for any positive constant r.

PROOF. This follows directly from Lemma 2 and Proposition 1. □

REMARK. We observe that the hypotheses of Theorem 5 are satisfied by choosing $0 \leq \rho < \gamma \leq \alpha$ in Example 1. In the case of Examples 2 and 3, the hypotheses of the above theorem hold if we take $0 \leq \rho < \gamma < \alpha$.

4 Acknowledgments

I would like to thank Professors Herman Rubin and Bill Studden for kindling my interest in mixture problems, which eventually resulted in this work.

5 Appendix

Lemma 3 *Under the assumptions of the first paragraph of Section 1, the mixture distribution as defined by (1) is identifiable in G.*

PROOF. Let G_1 and G_2 be two probability distributions on $[0, \alpha]$ such that

$$\int_0^\alpha x^j \phi(x) dG_1(x) = \int_0^\alpha x^j \phi(x) dG_2(x), \quad \forall j = 0, 1, 2, \cdots.$$

Furthermore we assume that the restriction of $G_1 - G_2$ to $\{x : \phi(x) = 0\}$ is identically zero if the cardinality of $\{x : \phi(x) = 0\}$ is greater than or equal to two. As the Hausdorff moment problem is determinate [see, for example, Shohat and Tamarkin (1943) page 9], we conclude that

$$\phi(x) dG_1(x) = \phi(x) dG_2(x). \tag{23}$$

Now we consider two cases.

CASE I. Suppose that the cardinality of $\{x : \phi(x) = 0\}$ is greater than or equal to two. Then it follows from (23) that $G_1 = G_2$ on $[0, \alpha] \backslash \{x : \phi(x) = 0\}$ and hence $G_1 = G_2$.

CASE II. Suppose that the cardinality of $\{x : \phi(x) = 0\}$ is less than two. Again from (23) we have $G_1 = G_2$ on $[0, \alpha] \backslash \{x : \phi(x) = 0\}$. Hence $G_1 = G_2$ as both are probability measures. This proves Lemma 3. □

PROOF OF LEMMA 1. We observe that

$$\begin{aligned} L_1(G_0, G_n) &= \sum_{j=0}^{\infty} \left| \int_0^\alpha x^j \phi(x) \psi(j) n^{-r/2} p_N(n^{r/8}(x - a)) dx \right| \\ &= n^{-5r/8} \sum_{j=0}^{\infty} \left| \int_{-1}^{1} (a + x n^{-r/8})^j \phi(a + x n^{-r/8}) \psi(j) p_N(x) dx \right|. \end{aligned} \tag{24}$$

It follows from Taylor's expansion that for $j \geq m$ and n sufficiently large,

$$\begin{aligned} (a + x n^{-r/8})^j &= a^j + C_1^j a^{j-1}(x n^{-r/8}) + \cdots + C_{m-1}^j a^{j-m+1}(x n^{-r/8})^{m-1} \\ &\quad + C_m^j (a + \xi)^{j-m} (x n^{-r/8})^m, \end{aligned}$$

and for $1 \leq l \leq m$,

$$\phi(a + x n^{-r/8}) = \phi(a) + \cdots + \frac{\phi^{(l-1)}(a)}{(l-1)!}(x n^{-r/8})^{l-1} + \frac{\phi^{(l)}(a + \eta_l)}{l!}(x n^{-r/8})^l,$$

where $\max\{|\xi|, |\eta_l|\} \leq |xn^{-r/8}|$. Hence using (16), we get

$$
\begin{aligned}
& |\int_{-1}^{1}(a+xn^{-r/8})^j\phi(a+xn^{-r/8})p_N(x)dx| \\
= & |\int_{-1}^{1}[\phi(a+xn^{-r/8})C_m^j(a+\xi)^{j-m} \\
& +\sum_{l=1}^{m}\frac{\phi^{(l)}(a+\eta_l)}{l!}C_{m-l}^j a^{j-m+l}](xn^{-r/8})^m p_N(x)dx| \\
= & n^{-mr/8}O(j^m(a+\varepsilon)^j), \qquad (25)
\end{aligned}
$$

as $n \to \infty$ uniformly over $j \geq m$. Similarly we have for $0 \leq j < m$,

$$
\begin{aligned}
& |\int_{-1}^{1}(a+xn^{-r/8})^j\phi(a+xn^{-r/8})p_N(x)dx| \\
= & O(n^{-mr/8}), \quad \text{as } n \to \infty. \qquad (26)
\end{aligned}
$$

Thus we conclude from (17), (24), (25) and (26) that

$$L_1(G_0, G_n) = O(n^{-(5+m)r/8}) \leq 1/n,$$

for sufficiently large n. This completes the proof of Lemma 1. □

PROOF OF LEMMA 2. As in the proof of Lemma 1, we have

$$
\begin{aligned}
& \max_{1\leq l\leq \tau_n}\max_{\vartheta_n} L_1(g_{\vartheta_{n,l,0}}, g_{\vartheta_{n,l,1}}) \\
= & \max_{1\leq l\leq \tau_n}\frac{\gamma-\rho}{4\tau_n n^{r/2}}\sum_{j=0}^{\infty}|\int_{-1}^{1}[x_{n,l}+\frac{x(\gamma-\rho)}{4\tau_n}]^j \\
& \times\phi(x_{n,l}+\frac{x(\gamma-\rho)}{4\tau_n})\psi(j)p_N(x)dx| \\
= & O(n^{-r/2}\tau_n^{-m-1})\sum_{j=0}^{\infty}j^m\gamma^j\psi(j),
\end{aligned}
$$

as $n \to \infty$. Hence we conclude that

$$\max_{1\leq l\leq \tau_n}\max_{\vartheta_n} L_1(g_{\vartheta_{n,l,0}}, g_{\vartheta_{n,l,1}}) \leq 1/n,$$

for sufficiently large n. This proves Lemma 2. □

References

[1] BLUM, J. R. and SUSARLA, V. (1977). Estimation of a mixing distribution function. *Ann. Probab.* **5** 200-209.

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

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

[4] DEVROYE, L. P. (1989). Consistent deconvolution in density estimation. *Canad. J. Statist.* **17** 235-239.

[5] DEVROYE, L. P. and WISE, G. L. (1979). On the recovery of discrete probability densities from imperfect measurements. *J. Franklin Inst.* **307** 1-20.

[6] DONOHO, D. L. and Liu, R. C. (1987). Geometrizing rates of convergence, I. Technical Report, Dept. Statist., Univ. California, Berkeley.

[7] DONOHO, D. L. and LIU, R. C. (1991a). Geometrizing rates of convergence, II. *Ann. Statist.* **19** 633-667.

[8] DONOHO, D. L. and LIU, R. C. (1991b). Geometrizing rates of convergence, III. *Ann. Statist.* **19** 668-701.

[9] FAN, J. (1991a). On the optimal rates of convergence for nonparametric deconvolution problems. *Ann. Statist.* **19** 1257-1272.

[10] FAN, J. (1991b). Global behavior of deconvolution kernel estimates. *Statist. Sinica* **1** 541-551.

[11] FAN, J. (1991c). Adaptively local 1-dimensional subproblems. Preprint.

[12] HARRIS, I. R. (1991). The estimated frequency of zero for a mixed Poisson distribution. *Statist. Probab. Lett.* **12** 371-372.

[13] Jewell, N. (1982). Mixtures of exponential distributions. *Ann. Statist.* **10** 479-484.

[14] JOHNSON, N. L. and KOTZ, S. (1969). *Discrete Distributions.* Wiley, New York.

[15] LAMBERT, D. and TIERNEY, L. (1984). Asymptotic properties of maximum likelihood estimates in the mixed Poisson model. *Ann. Statist.* **12** 1388-1399.

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

[17] LINDSAY, B. G. (1983b). The geometry of mixture likelihoods, part II: The exponential family. *Ann. Statist.* **11** 783-792.

[18] LINDSAY, B. G. (1989). Moment matrices: Applications in mixtures. *Ann. Statist.* **17** 722-740.

[19] MEEDEN, G. (1972). Bayes estimation of the mixing distribution, the discrete case. *Ann. Math. Statist.* **43** 1993-1999.

[20] ROBBINS, H. (1964). The empirical Bayes approach to statistical decision problems. *Ann. Math. Statist.* **35** 1-20.

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

[22] SHOHAT, J. and TAMARKIN, J. (1943). *The Problem of Moments.* Amer. Math. Soc. Waverly Press, Baltimore.

[23] SIMAR, L. (1976). Maximum likelihood estimation of a compound Poisson process. *Ann. Statist.* **4** 1200-1209.

[24] STEFANSKI, L. A. (1990). Rates of convergence of some estimators in a class of deconvolution problems. *Statist. Probab. Lett.* **9** 229-235.

[25] STEFANSKI, L. A. and CARROLL, R. J. (1990). Deconvolving kernel density estimators. *Statist.* **21** 169-184.

[26] TEICHER, H. (1961). Identifiability of mixtures. *Ann. Math. Statist.* **32** 244-248.

[27] TUCKER, H. G. (1963). An estimate of the compounding distribution of a compound Poisson distribution. *Theor. Probab. Appl.* **8** 195-200.

[28] WISE, G. L., TRAGANITIS, A. P. and THOMAS, J. B. (1977). The estimation of a probability density function from measurements corrupted by Poisson noise. *IEEE Trans. Inform. Theory* **23** 764-766.

[29] ZHANG, C. H. (1990). Fourier methods for estimating mixing densities and distributions. *Ann. Statist.* **18** 806-831.

Department of Statistics
Mathematical Sciences Building
Purdue University
West Lafayette, Indiana 47907