Rates of Convergence of Orthogonal Polynomial Estimators for a Mixing Density

by

Wei-Liem Loh and Cun-Hui Zhang
Purdue University Rutgers University

Technical Report # 93–46

Department of Statistics
Purdue University

August 1993

△

RATES OF CONVERGENCE OF ORTHOGONAL POLYNOMIAL ESTIMATORS FOR A MIXING DENSITY

BY WEI-LIEM LOH[1] AND CUN-HUI ZHANG[2]

Purdue University and Rutgers University

Walter and Hamedani (1989) proposed a class of orthogonal polynomial estimators for the mixing density of a mixture of discrete exponential families. In this paper the convergence rates of these estimators are studied with respect to weighted L^2 loss over various smoothness classes of mixing density functions. In particular, sufficient conditions are obtained in which the estimators attain the optimal convergence rate.

1 Introduction

Let $X_1, \ldots, X_n$ be independent observations from a mixture distribution with probability law

$$f(x;g) = \int_0^{\theta^*} f(x|\theta)g(\theta)d\theta, \tag{1}$$

where g is a mixing probability density function on $(0, \theta^*)$ and $f(x|\theta)$ is a known parametric family of probabilty density functions with respect to a σ-finite measure ν. In particular we assume that

$$f(x|\theta) = C(\theta)q(x)\theta^x, \quad \forall x = 0, 1, 2, \cdots, \tag{2}$$

where $0 < \theta < \theta^* \leq \infty$, $q(x) > 0$ whenever $x = 0, 1, 2, \ldots$ and ν is the counting measure on the set of nonnegative integers. This class is quite broad and includes the following mixture distributions.

EXAMPLE 1. The random variable X is said to have a Poisson mixture distribution with mixing density g if

$$P(X = x) = \int_0^{\theta^*} e^{-\theta}(\theta^x/x!)g(\theta)d\theta, \quad \forall x = 0, 1, 2, \cdots.$$

[1]Research supported in part by NSF Grant DMS 89-23071 and NSA Grant MDA 904-93-H-3011.

[2]Research supported in part by NSF Grant DMS 89-16180 and ARO Grant DAAL03-91-G-0045.

AMS 1980 *subject classifications.* Primary 62G05; secondary 62E20, 62G20.

Key words and phrases. Mixing density, orthogonal polynomials, discrete exponential family, rate of convergence.

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

$$P(X = x) = \int_0^1 \binom{x + \nu - 1}{x} \theta^x (1 - \theta)^\nu g(\theta) d\theta, \quad \forall x = 0, 1, 2, \cdots.$$

In 1989 Walter and Hamedani proposed a class of orthogonal polynomial estimators for the mixing density g of the mixture distribution as described in (1). In this paper, we shall investigate the convergence rates of these orthogonal polynomial estimators with respect to weighted L^2 loss. In particular Section 2 gives upper bounds for the rates of convergence of the estimators over various smoothness classes of mixing density functions. Section 3 supplies lower bounds for the optimal convergence rates for a subset of these smoothness classes. As a consequence of the results of Sections 2 and 3, sufficient conditions are obtained in which the estimators attain (or almost attain) the optimal convergence rate. Finally the Appendix contains somewhat technical lemmas that are needed in previous sections.

Over the last few years, there has been a great deal of interest in mixture problems. Important advances have been made on the deconvolution problem by Devroye and Wise (1979), Carroll and Hall (1988), Zhang (1990), Fan (1991a), (1991b) (1991c) and many others using Fourier techniques. In particular kernel estimators have been obtained which achieve the optimal convergence rate.

In the context of mixtures of discrete exponential families, Tucker (1963) considered the estimation of the mixing distribution of a Poisson mixture via the method of moments and Simar (1976) approached the same problem using maximum likelihood. More recently, Zhang (1988), (1992) and Loh and Zhang (1993) considered estimating the mixing density of a mixture of discrete exponential families via Fourier methods and obtained kernel estimators which attain (or almost attain) the optimal convergence rate under suitable conditions. As mentioned earlier, Walter and Hamedani (1989), (1991) and Walter (1985) considered similar problems and have proposed alternative mixing density estimators using orthogonal polynomials.

Other mixture problems were studied by Robbins (1964), Deely and Kruse (1968), Jewell (1982) and Lindsay (1983), (1989) among others. Rolph (1968), Meeden (1972) and Datta (1991) used Bayesian methods to construct consistent estimators for the mixing distribution.

Throughout this paper, we shall denote by $P = P_g$ and $E = E_g$ the probability and expectation corresponding to g respectively, by $h^{(j)}$ the jth derivative (if it exists) of any function h with $h^{(0)} = h$, and the weighted L^p-norm of any measurable function h by

$$||h||_{w,p} = \begin{cases} (\int |h(y)|^p w(y)dy)^{1/p} & \text{if } 1 \le p < \infty, \\ \text{ess sup}_y |h(y)| & \text{if } p = \infty, \end{cases}$$

where $w(y)$ is the weight function and ess sup is with respect to the measure $w(y)dy$. If $w(y) \equiv 1$, we denote $||.||_{w,p}$ by $||.||_p$.

2 Mixtures of discrete exponential families

Let $C : (0, \theta^*) \to R^+$ be as in (2) and $w : (0, \theta^*) \to R^+$ be a measurable function such that $||C^2/w||_1 < \infty$. Let $\{p_{w_0,j}\}_{j=0}^{\infty}$ be a sequence of orthogonal polynomials on $(0, \theta^*)$ with weight function

$$w_0(\theta) = C^2(\theta)/w(\theta). \tag{3}$$

In particular, these polynomials are normalized so that

$$p_{w_0,j}(\theta) = \sum_{x=0}^{j} k_{w_0,j,x}\theta^x, \tag{4}$$

with $k_{w_0,j,j} > 0$ for all $j \ge 0$, and

$$\int_0^{\theta^*} p_{w_0,i}(\theta)p_{w_0,j}(\theta)w_0(\theta)d\theta = \delta_{ij},$$

where δ_{ij} denotes the Kronecker delta. We further assume that $\{p_{w_0,j}\}_{j=0}^{\infty}$ is complete with respect to $||.||_{w_0,2}$. Note that this is always true if $\theta^* < \infty$ [see for example Szegö (1975) page 40]. Next define

$$\lambda_{w_0,j}(x) = \begin{cases} k_{w_0,j,x}/q(x) & \text{if } 0 \le x \le j, \\ 0 & \text{otherwise.} \end{cases}$$

We write

$$h(\theta) = w(\theta)g(\theta)/C(\theta), \quad \forall 0 < \theta < \theta^*, \tag{5}$$

and assume that the mixing density g satisfies $||g||_{w,2} = ||h||_{w_0,2} < \infty$. Then h has the formal orthogonal polynomial series expansion

$$h(\theta) \sim \sum_{j=0}^{\infty} h_{w_0,j}p_{w_0,j}(\theta),$$

where

$$h_{w_0,j} = \int_0^{\theta^*} h(\theta)p_{w_0,j}(\theta)w_0(\theta)d\theta, \quad \forall j = 0,1,2,\cdots. \tag{6}$$

Observing that

$$E\lambda_{w_0,j}(X_1) = \sum_{x=0}^{\infty} f(x;g)\lambda_{w_0,j}(x) = h_{w_0,j}, \quad \forall j = 0,1,2,\cdots,$$

we estimate $h_{w_0,j}$ by $\hat{h}_{w_0,j} = n^{-1}\sum_{i=1}^{n}\lambda_{w_0,j}(X_i)$ and $g(\theta)$ by

$$\hat{g}_n(\theta) = [C(\theta)/w(\theta)]\sum_{j=0}^{c_n}\hat{h}_{w_0,j}p_{w_0,j}(\theta), \quad \forall 0 < \theta < \theta^*, \tag{7}$$

where c_n is a positive constant tending to ∞.

Proposition 1 *Suppose* $||C^2/w||_1 < \infty$ *and* $||g||_{w,2} < \infty$. *Let* $\hat{g}_n$ *be as in (7). Then*

$$E_g||\hat{g}_n - g||_{w,2} \le \{n^{-1}\sum_{j=0}^{c_n}\max_{0\le x\le j}[k_{w_0,j,x}/q(x)]^2 + \sum_{j=c_n+1}^{\infty}h_{w_0,j}^2\}^{1/2},$$

with $k_{w_0,j,x}$ *and* $h_{w_0,j}$ *as in (4) and (6) respectively.*

PROOF. We observe that

$$\begin{aligned} & E_g\{\int_0^{\theta^*}[\hat{g}_n(\theta) - g(\theta)]^2 w(\theta)d\theta\}^{1/2} \\ = \ & E_g\{\int_0^{\theta^*}[\sum_{j=0}^{c_n}\hat{h}_{w_0,j}p_{w_0,j}(\theta) - h(\theta)]^2 w_0(\theta)d\theta\}^{1/2} \\ \le \ & \{\sum_{j=0}^{c_n}E_g(\hat{h}_{w_0,j} - h_{w_0,j})^2 + \sum_{j=c_n+1}^{\infty}h_{w_0,j}^2\}^{1/2}. \end{aligned} \tag{8}$$

The last inequality follows from Jensen's inequality and the completeness of $\{p_{w_0,j}\}_{j=0}^{\infty}$. Since $\hat{h}_{w_0,j} = n^{-1}\sum_{i=1}^{n}\lambda_{w_0,j}(X_i)$, the r.h.s. of (8) is bounded by

$$\begin{aligned} & \{n^{-1}\sum_{j=0}^{c_n}E_g[\lambda_{w_0,j}^2(X_1)] + \sum_{j=c_n+1}^{\infty}h_{w_0,j}^2\}^{1/2} \\ \le \ & \{n^{-1}\sum_{j=0}^{c_n}\max_{0\le x\le j}[k_{w_0,j,x}/q(x)]^2 + \sum_{j=c_n+1}^{\infty}h_{w_0,j}^2\}^{1/2}. \end{aligned}$$

This proves the proposition. □

REMARK. The motivation for (7) originates from Walter and Hamedani (1989) who proposed a similar class of estimators. They also obtained a result analogous to Proposition 1.

We now study the performance of the estimators $\hat{g}_n$ with respect to the following class of mixing density functions. For positive constants α, M and $m = 1, 2, \ldots$, we define $\mathcal{G}(\alpha, m, M, w_0)$ to be the set of all probability density functions g on $(0, \theta^*)$ such that $||g||_{w,2} < \infty$ and $\sum_{j=m}^{\infty} j^{2\alpha} h_{w_0,j}^2 < M$ with $h_{w_0,j}$ as in (6). We note that this class implicitly depends on the discrete exponential family of interest, in particular on $C(\theta)$. This ellipsoidal class is chosen mainly for reasons of mathematical tractability. However ellipsoid conditions can amount to the imposition of smoothness and integrability requirements, see for example Johnstone and Silverman (1990) page 258. In our case, we have the following characterization.

Proposition 2 *Let $m \geq 1$ and $\{p_{w_0,j}\}_{j=0}^{\infty}$ be as in (4). Suppose there exist constants $\nu_{j,m}$, $j \geq m$ and another sequence of (normalized) complete orthogonal polynomials $\{p_{w_1,j}\}_{j=0}^{\infty}$ with weight function w_1 such that*

$$(9) \qquad [p_{w_1,j}(\theta)w_1(\theta)]^{(m)} = (-1)^m \nu_{j+m,m} p_{w_0,j+m}(\theta) w_0(\theta), \qquad \forall j \geq 0,$$

and

$$(10) \qquad \alpha_1 < \inf_{j \geq m} |\nu_{j,m}|/j^{\alpha} \leq \sup_{j \geq m} |\nu_{j,m}|/j^{\alpha} < \alpha_2,$$

where α, α_1 and α_2 are positive constants. Then if h is a measurable function on $(0, \theta^*)$ such that $h^{(m)}$ exists,

$$\begin{aligned} 0 &= \lim_{\theta \to 0+} h^{(m-i)}(\theta)[p_{w_1,j}(\theta)w_1(\theta)]^{(i-1)} \\ &= \lim_{\theta \to \theta^*-} h^{(m-i)}(\theta)[p_{w_1,j}(\theta)w_1(\theta)]^{(i-1)} \end{aligned}$$

whenever $0 < i < m$, $j \geq 0$, and $||h^{(m)}||_{w_1,2} < \infty$, we have

$$\alpha_1 (\sum_{j=m}^{\infty} j^{2\alpha} h_{w_0,j}^2)^{1/2} \leq ||h^{(m)}||_{w_1,2} \leq \alpha_2 (\sum_{j=m}^{\infty} j^{2\alpha} h_{w_0,j}^2)^{1/2},$$

where $h_{w_0,j}$ is defined as in (6).

We defer the proof of Proposition 2 to the Appendix. The following argument shows that (9) and (10) are satisfied by the classical orthogonal polynomials of Laguerre and Jacobi.

LAGUERRE POLYNOMIALS. Suppose $w_0(\theta) = \theta^\beta e^{-\theta}$, with $\theta > 0$ and $\beta > -1$, is the weight function of the normalized Laguerre polynomials

$$p_{w_0,j}(\theta) = [\Gamma(\beta+1)\binom{j+\beta}{j}]^{-1/2} \sum_{x=0}^{j} \binom{j+\beta}{j-x} \frac{(-\theta)^x}{x!}, \quad \forall j \geq 0.$$

For $j \geq 0$ and $m \geq 1$, we write

$$\begin{aligned} w_1(\theta) &= \theta^{\beta+m} e^{-\theta}, \\ p_{w_1,j}(\theta) &= [\Gamma(\beta+m+1)\binom{j+\beta+m}{j}]^{-1/2} \sum_{x=0}^{j} \binom{j+\beta+m}{j-x} \frac{(-\theta)^x}{x!}, \end{aligned}$$

and

$$\begin{aligned} \nu_{j+m,m} &= (-1)^m \frac{(j+m)!}{j!} [\Gamma(\beta+1)\binom{j+\beta+m}{j+m}]^{1/2} \\ &\quad \times [\Gamma(\beta+m+1)\binom{j+\beta+m}{j}]^{-1/2}. \end{aligned}$$

Then (9) follows from the Rodrigues' formula for Laguerre polynomials and (10) holds for $\alpha = m/2$. □

JACOBI POLYNOMIALS. Suppose $w_0(\theta) = \theta^{\beta_1}(\theta^* - \theta)^{\beta_2}$, with $\beta_1 > -1$, $\beta_2 > -1$ and $0 < \theta < \theta^* < \infty$. Then the orthogonal polynomials with w_0 as the weight function correspond to the normalized Jacobi polynomials

$$\begin{aligned} p_{w_0,j}(\theta) &= C_{j,\beta_1,\beta_2} \binom{j+\beta_2}{j} (\theta^*)^{-j} \\ &\quad \times \sum_{x=0}^{j} \frac{j(j-1)\cdots(j-x+1)}{(\beta_2+1)(\beta_2+2)\cdots(\beta_2+x)} \binom{j+\beta_1}{x} \theta^{j-x} (\theta-\theta^*)^x, \end{aligned}$$

where

$$C_{j,\beta_1,\beta_2} = [\frac{(2j+\beta_1+\beta_2+1)\Gamma(j+1)\Gamma(j+\beta_1+\beta_2+1)}{(\theta^*)^{\beta_1+\beta_2+1}\Gamma(j+\beta_1+1)\Gamma(j+\beta_2+1)}]^{1/2} \quad \text{if } j \geq 1,$$

and is equal to

$$[\frac{\Gamma(\beta_1+\beta_2+2)}{(\theta^*)^{\beta_1+\beta_2+1}\Gamma(\beta_1+1)\Gamma(\beta_2+1)}]^{1/2} \quad \text{if } j = 0.$$

For $m \geq 1$, let $p_{w_1,j}$, $j \geq 0$, denote the set of normalized Jacobi polynomials with weight function

$$w_1(\theta) = \theta^{\beta_1+m}(\theta^* - \theta)^{\beta_2+m}, \quad \forall 0 < \theta < \theta^*,$$

and

$$\nu_{j+m,m} = (\theta^*)^m (j+m)! C_{j,\beta_1+m,\beta_2+m} / [j! C_{j+m,\beta_1,\beta_2}].$$

Then (9) follows from the Rodrigues' formula for Jacobi polynomials and (10) holds for $\alpha = m$. □

For the rest of this paper, we shall assume that M is sufficiently large so that $\mathcal{G}(\alpha, m, M, w_0)$ is nonempty.

Theorem 1 *Suppose* $||C^2/w||_1 < \infty$. *Let* $\hat{g}_n$ *be as in (7) and*

$$\max_{0 \leq x \leq j \leq c_n} \log(|k_{w_0,j,x}|/q(x)) \leq \beta_0 \log n, \tag{11}$$

for some constant $0 < \beta_0 < 1/2$. *Then*

$$\sup\{E_g ||\hat{g}_n - g||_{w,2} : g \in \mathcal{G}(\alpha, m, M, w_0)\} = O(1)(c_n^{-\alpha} + c_n^{1/2} n^{(2\beta_0 - 1)/2}).$$

Proof. We first observe from (11) that

$$n^{-1} \sum_{j=0}^{c_n} \max_{0 \leq x \leq j} [k_{w_0,j,x}/q(x)]^2 = O(c_n n^{2\beta_0 - 1}). \tag{12}$$

We also observe that

$$\sup\{\sum_{j=c_n+1}^{\infty} h_{w_0,j}^2 : g \in \mathcal{G}(\alpha, m, M, w_0)\} = O(c_n^{-2\alpha}). \tag{13}$$

Now the theorem follows from (12), (13) and Proposition 1. □

Corollary 1 *Suppose* $\theta^* = \infty$, $w(\theta) = \theta^{-\beta} C^2(\theta) e^{\theta}$ *and* $w_0(\theta) = \theta^{\beta} e^{-\theta}$ *with* $\beta > -1$. *Let* $\{p_{w_0,j}\}_{j=0}^{\infty}$ *be the sequence of (normalized) Laguerre polynomials on* $(0, \infty)$ *with weight function* w_0, $\hat{g}_n$ *as in (7) and*

$$q(x)\gamma_0 \gamma_1^x (x!)^\gamma > 1, \quad \forall x \geq 0,$$

for constants γ_0, $\gamma_1 \geq 1$ *and* $0 < \gamma \leq 1$. *Then by choosing* $c_n = \delta \log n$ *with* $0 < \delta \leq \beta_0 / \log(2\gamma_1)$ *and* $0 < \beta_0 < 1/2$, *we have*

$$\sup\{E_g ||\hat{g}_n - g||_{w,2} : g \in \mathcal{G}(\alpha, m, M, w_0)\} = O(1)(1/\log n)^{\alpha}.$$

PROOF. From the properties of Laguerre polynomials, we have

$$\begin{aligned}
&|k_{w_0,j,x}/q(x)| \\
\leq\ & \gamma_0\gamma_1^x(x!)^{\gamma-1}\binom{j+\beta}{j-x}[\Gamma(\beta+1)\binom{j+\beta}{j}]^{-1/2} \\
=\ & \gamma_0\gamma_1^x(x!)^{\gamma-1}\binom{j}{x}[\prod_{i=x+1}^{j}(1+\beta i^{-1})]^{1/2}[\Gamma(\beta+1)\prod_{i=1}^{x}(1+\beta i^{-1})]^{-1/2} \\
(14) \leq\ & \gamma_0\gamma_1^j 2^j[\prod_{i=x+1}^{j}(1+\beta i^{-1})]^{1/2}[\Gamma(\beta+1)\prod_{i=1}^{x}(1+\beta i^{-1})]^{-1/2}.
\end{aligned}$$

Here we follow the convention that $\prod_{i=x_1}^{x_2}(1+\beta i^{-1}) = 1$ if $x_1 > x_2$. We further observe that there exist positive constants c_1^* and c_2^* such that

$$c_1^* j^{-1} \leq \prod_{i=1}^{j}(1+\beta i^{-1}) \leq c_2^* j, \quad \forall j \geq 1.$$

Thus it follows from (14) that

$$\begin{aligned}
\max_{0\leq x\leq j\leq c_n} \log(|k_{w_0,j,x}|/q(x)) &= c_n(1+o(1))\log(2\gamma_1) \\
&\leq \beta_0(1+o(1))\log n.
\end{aligned}$$

This proves (11) and the corollary follows from Theorem 1. □

The next theorem is a specialization of Theorem 1 which proves to be useful when $\theta^* < \infty$. The proof is immediate and is omitted.

Theorem 2 *Let $\hat{g}_n$ be as in (7) and that for some constant $\zeta > 1$,*

$$\max_{0\leq x\leq j} k^2_{w_0,j,x} < \zeta^{2j}, \quad \forall j \geq 0. \tag{15}$$

Suppose further that

$$\max_{0\leq x\leq c_n} \log(1/q(x)) + c_n \log\zeta \leq \beta_0 \log n, \tag{16}$$

with constant $0 < \beta_0 < 1/2$. Then

$$\sup\{E_g||\hat{g}_n - g||_{w,2} : g \in \mathcal{G}(\alpha, m, M, w_0)\} = O(c_n^{-\alpha}).$$

Corollary 2 *Let $\hat{g}_n$ be as in (7) and that (15) holds for some constant $\zeta > 1$, Suppose*

$$q(x)\gamma_0\gamma_1^x(x!)^\gamma > 1, \quad \forall x \geq 0,$$

for nonnegative constants γ_0, $\gamma_1 \geq 1$ and γ. Then

(a) if $\gamma = 0$, by choosing $c_n = \delta \log n$ with $0 < \delta \leq \beta_0/\log(\gamma_1\zeta)$ and $0 < \beta_0 < 1/2$, we have

$$\sup\{E_g||\hat{g}_n - g||_{w,2} : g \in \mathcal{G}(\alpha, m, M, w_0)\} = O(1)(1/\log n)^\alpha,$$

(b) if $0 < \gamma < \infty$, by choosing $c_n = \delta \log n/\log\log n$ with $0 < \delta \leq \beta_0/\gamma$ and $0 < \beta_0 < 1/2$, we have

$$\begin{aligned} & \sup\{E_g||\hat{g}_n - g||_{w,2} : g \in \mathcal{G}(\alpha, m, M, w_0)\} \\ = \; & O(1)(\log\log n/\log n)^\alpha. \end{aligned}$$

Proof. If $\gamma = 0$, we observe that

$$\begin{aligned} \max_{0 \leq x \leq c_n} \log(1/q(x)) + c_n \log\zeta &\leq c_n(1 + o(1))\log(\gamma_1\zeta) \\ &\leq \beta_0(1 + o(1))\log n. \end{aligned}$$

This proves (16) and (a) follows from Theorem 2. The case of $0 < \gamma < \infty$ is similar and is omitted. □

Remark. Corollary 2 applies to the negative binomial and Poisson mixtures as the value of γ is 0 for the negative binomial mixture and 1 for the Poisson mixture.

As the classical orthogonal polynomials of Jacobi satisfy (15), we have the following corollary.

Corollary 3 *Suppose θ^* is finite,*

$$w(\theta) = C^2(\theta)\theta^{-\beta_1}(\theta^* - \theta)^{-\beta_2}, \quad \forall 0 < \theta < \theta^*,$$

for constants $\beta_1 > -1$ and $\beta_2 > -1$ and w_0 be as in (3). Let $\{p_{w_0,j}\}_{j=0}^\infty$ denote the set of (normalized) Jacobi polynomials on $(0, \theta^*)$ with respect to the weight function w_0, $\hat{g}_n$ be defined by (7) and

$$q(x)\gamma_0\gamma_1^x(x!)^\gamma > 1, \quad \forall x \geq 0,$$

for nonnegative constants γ_0, $\gamma_1 \geq 1$ and γ. Then conclusions (a) and (b) of Corollary 2 hold.

3 Lower bounds on the optimal convergence rate

In this section we derive lower bounds for the optimal convergence rate over the class of mixing densities $\mathcal{G}(\alpha, m, M, w_0)$ with α satisfying (9) and (10), i.e. the rate at which

$$\inf_{\tilde{g}_n} \sup\{E_g||\tilde{g}_n - g||_{w,2} : g \in \mathcal{G}(\alpha, m, M, w_0)\}$$

tends to 0 as $n \to \infty$, where the infimum runs over all statistics $\tilde{g}_n$ based on $X_1, \cdots, X_n$.

Theorem 3 *Let $w : (0, \theta^*) \to R^+$ be a measurable function such that $||w||_1 < \infty$ and $||w_0||_1 < \infty$ with w_0 as in (3) and $\{p_{w_0,j}\}_{j=0}^{\infty}$ be a sequence of (normalized) orthogonal polynomials with weight function w_0 such that (9) and (10) are satisfied. Suppose there exists a nonempty interval $(\theta_0, \theta_1) \subset (0, \theta^*)$ such that w is strictly positive and m times continuously differentiable on (θ_0, θ_1) and w_1 is bounded from above on the same interval. Then for sufficiently large M,*

$$\lim_{n\to\infty} (\log n)^m \inf_{\tilde{g}_n} \sup\{E_g||\tilde{g}_n - g||_{w,2} : g \in \mathcal{G}(\alpha, m, M, w_0)\} > 0.$$

Proof. Since w is bounded away from 0 on (θ_0, θ_1), let a be an interior point of that interval and δ, θ_2, θ_3, θ_4 and θ_5 be fixed constants satisfying

$$0 < \theta_0 < \theta_2 < \theta_3 < a < \theta_4 < \theta_5 < \theta_1 < \theta^*,$$

with

$$\theta_4 - a < a, \quad w(\theta) > \delta > 0, \quad \forall \theta_0 < \theta < \theta_1.$$

Next define

$$l_{u,v}(\theta) = v^u \theta^{u-1} e^{-v\theta} / \Gamma(u), \tag{17}$$

with $u/v = a$ and $g_{u,v} : (0, \theta^*) \to R$ be an m times continuously differentiable function such that

$$g_{u,v}(\theta) = \begin{cases} 0 & \text{if } 0 < \theta < \theta_2, \\ l_{1,u,v}(\theta)/C(\theta) & \text{if } \theta_2 \le \theta < \theta_3, \\ l_{u,v}(\theta)/C(\theta) = \sum_{x=0}^{\infty} q(x)\theta^x l_{u,v}(\theta) & \text{if } \theta_3 \le \theta \le \theta_4, \\ l_{2,u,v}/C(\theta) & \text{if } \theta_4 < \theta \le \theta_5, \\ 0 & \text{if } \theta_5 < \theta < \theta^*, \end{cases}$$

where $l_{1,u,v}$ and $l_{2,u,v}$ are $(2m+1)$th degree polynomials on (θ_2, θ_3) and (θ_4, θ_5) respectively. Furthermore we write

$$t(u,v) = \int_{\theta_2}^{\theta_5} \cos(u(\theta - a)/\theta_4) g_{u,v}(\theta) d\theta / \int_{\theta_2}^{\theta_5} g_{u,v}(\theta) d\theta. \tag{18}$$

Now choose constants u_n and $v_n = u_n/a$ such that

$$\begin{aligned} \frac{u_n}{\log n} &= \max\{\frac{\theta_4/(\theta_5 - \theta_4)}{\log(\theta_5/\theta_4)}, \frac{2}{\log(1 + a^2/\theta_4^2)}, \\ &\quad \frac{1}{\theta_4/a - 1 - \log(\theta_4/a)}, \frac{1}{\theta_3/a - 1 - \log(\theta_3/a)}\}. \end{aligned} \tag{19}$$

For constants $\varepsilon > 0$ and $0 < \varepsilon_0 < M$, let $g_0 \in \mathcal{G}(\alpha, m, M - \varepsilon_0, w_0)$ such that for sufficiently large n,

$$\begin{aligned} g_0(\theta) &\geq \frac{\varepsilon}{u_n^{m-1/4} \theta C(\theta)} e^{-u_n(\theta/a - 1 - \log(\theta/a))} I\{\theta_3 \leq \theta \leq \theta_4\} \\ &\quad + \frac{\varepsilon}{u_n^{m-1/4} \theta_3 C(\theta_3)} e^{-u_n(\theta_3/a - 1 - \log(\theta_3/a))} I\{\theta_2 \leq \theta < \theta_3\} \\ &\quad + \frac{\varepsilon}{u_n^{m-1/4} \theta_4 C(\theta_4)} e^{-u_n(\theta_4/a - 1 - \log(\theta_4/a))} I\{\theta_4 < \theta \leq \theta_5\}. \end{aligned} \tag{20}$$

For $0 < \theta < \theta^*$, define $g_{1n}(\theta) = g_0(\theta)$ and

$$g_{2n}(\theta) = g_0(\theta) + \varepsilon u_n^{-m-1/4} [\cos(u_n(\theta - a)/\theta_4) - t(u_n, v_n)] g_{u_n, v_n}(\theta). \tag{21}$$

For the sake of clarity we shall now divide the rest of the proof into 3 separate steps and we shall drop the subscript in u_n and v_n.

STEP 1. Verify the membership of g_{1n} and g_{2n} in $\mathcal{G}(\alpha, m, M, w_0)$. From the definition of g_{1n}, it is immediate that $g_{1n} \in \mathcal{G}(\alpha, m, M, w_0)$. We observe from (18), (20) and (29) that g_{2n} is a density function on $(0, \theta^*)$. Define for $0 < \theta < \theta^*$,

$$h(\theta) = \varepsilon u^{-m-1/4} w(\theta) [\cos(u(\theta - a)/\theta_4) - t(u,v)] g_{u,v}(\theta)/C(\theta).$$

Then using Leibniz rule and observing that w_1 is bounded on (θ_2, θ_5), it follows from Lemma 1 (see Appendix) that

$$||h^{(m)}||_{w_1,2} = \varepsilon O(1), \tag{22}$$

where the $O(1)$ term does not depend on ε. Since (9) and (10) hold, we observe from (22) and Proposition 2 that

$$(\sum_{j=m}^{\infty} j^{2\alpha} h_{w_0,j}^2)^{1/2} = \varepsilon O(1),$$

where $h_{w_0,j} = \int_0^{\theta^*} h(\theta) p_{w_0,j}(\theta) w_0(\theta) d\theta$. Writing

$$g_{2n,w_0,j} = \int_0^{\theta^*} C(\theta) g_{2n}(\theta) p_{w_0,j}(\theta) d\theta, \quad \forall j \geq m,$$

it follows from Minkowski's inequality and (21) that

$$(\sum_{j=m}^{\infty} j^{2\alpha} g_{2n,w_0,j}^2)^{1/2} \leq M - \varepsilon_0 + \varepsilon O(1).$$

Thus we conclude that $g_{2n} \in \mathcal{G}(\alpha, m, M, w_0)$ for sufficiently small ε.

STEP 2. To show that

$$\sum_{x=0}^{\infty} |f(x; g_{1n}) - f(x; g_{2n})| = o(1/n). \tag{23}$$

Since $l_{1,u,v}$ is a $(2m+1)$th degree polynomial, we write

$$l_{1,u,v}(\theta) = \sum_{i=0}^{2m+1} \beta_i \theta^i, \quad \forall \theta_2 < \theta < \theta_3,$$

where β_i, $0 \leq i \leq 2m+1$, are constants satisfying

$$\begin{aligned} \frac{d^j}{d\theta^j}(\sum_{i=0}^{2m+1} \beta_i \theta^i)|_{\theta=\theta_2} &= 0, \\ \frac{d^j}{d\theta^j}(\sum_{i=0}^{2m+1} \beta_i \theta^i)|_{\theta=\theta_3} &= O(u^{j+1/2}/n), \end{aligned}$$

uniformly in $\theta_2 < \theta < \theta_3$ and $0 \leq j \leq m$. Thus it follows that $\beta_i = O(u^{m+1/2}/n)$ uniformly in i. Now let $x^* = (\log\log n)/[2\log(\theta_4/\theta_3)]$. Then

$$\varepsilon u^{-m-1/4} \sum_{x=0}^{\infty} q(x) \left| \int_{\theta_2}^{\theta_3} \theta^x \cos(u(\theta - a)/\theta_4) l_{1,u,v}(\theta) d\theta \right|$$

$$\begin{aligned}
&\leq \varepsilon u^{-m-1/4}\{\sum_{x\leq x^*} q(x)|\int_{\theta_2}^{\theta_3} \theta^x e^{iu\theta/\theta_4}(\sum_{j=0}^{2m+1} \beta_j\theta^j)d\theta| \\
&\quad + \sum_{x>x^*} q(x)\theta_3^x O(u^{m+1/2}/n)\} \\
&\leq \varepsilon u^{-m-1/4}\{\theta_4 \sum_{x\leq x^*} q(x) \sum_{j=0}^{2m+1} |\beta_j[\frac{\theta_3^{x+j}e^{iu\theta_3/\theta_4}}{iu} - \frac{\theta_2^{x+j}e^{iu\theta_2/\theta_4}}{iu} \\
&\quad -I\{x+j>0\}\int_{\theta_2}^{\theta_3}(x+j)\theta^{x+j-1}\frac{e^{iu\theta/\theta_4}}{iu}d\theta]| \\
&\quad +(\theta_3/\theta_4)^{x^*}O(u^{m+1/2}/n)\sum_{x>x^*} q(x)\theta_4^x\} \\
(24)\quad &= o(1/n),
\end{aligned}$$

where $I\{x+j>0\}$ denotes the indicator function of $\{x+j>0\}$. Similarly it can be shown that

$$(25)\quad \varepsilon u^{-m-1/4}\sum_{x=0}^{\infty} q(x)|\int_{\theta_4}^{\theta_5} \theta^x \cos(u(\theta-a)/\theta_4) l_{2,u,v}(\theta)d\theta| = o(1/n).$$

Finally as in Zhang (1992), it can be seen using (19) that

$$(26)\quad \varepsilon u^{-m-1/4}\sum_{x=0}^{\infty} q(x)|\int_{\theta_3}^{\theta_4} \theta^x \cos(u(\theta-a)/\theta_4) l_{u,v}(\theta)d\theta| = o(1/n).$$

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

$$(27)\, \varepsilon u^{-m-1/4}\sum_{x=0}^{\infty} q(x)|\int_{\theta_2}^{\theta_5} \theta^x \cos(u(\theta-a)/\theta_4) C(\theta) g_{u,v}(\theta)d\theta| = o(1/n).$$

Since

$$\begin{aligned}
&\varepsilon u^{-m-1/4} t(u,v)\int_{\theta_2}^{\theta_5} g_{u,v}(\theta)d\theta \\
\leq\ &\varepsilon u^{-m-1/4}\sum_{x=0}^{\infty} q(x)|\int_{\theta_2}^{\theta_5} \theta^x \cos(u(\theta-a)/\theta_4) C(\theta) g_{u,v}(\theta)d\theta| \\
=\ &o(1/n),
\end{aligned}$$

(23) follows from (21).

STEP 3. Also as in Loh and Zhang (1993), we have

$$(28)\qquad \liminf_{n\to\infty}(\log n)^m ||g_{1n} - g_{2n}||_{w,2} > 2\varepsilon_1$$

where ε_1 is a suitably small positive constant. To conclude the proof, we observe that for $Z = (X_1, \cdots, X_n)$ and $\lambda > 1$,

$$\begin{aligned} P_{g_{1n}}(f(Z; g_{1n}) > \lambda f(Z; g_{2n})) &\leq \frac{\lambda}{\lambda - 1} \sum_Z |f(Z; g_{1n}) - f(Z; g_{2n})| \\ &\leq \frac{n\lambda}{\lambda - 1} \sum_x |f(x; g_{1n}) - f(x; g_{2n})| \\ &\rightarrow 0. \end{aligned}$$

Now it follows from (28) and Lemma 2 (see Appendix) that

$$\inf_{\tilde{g}_n} \max\{P_g[||\tilde{g}_n - g||_{w,2} > \varepsilon_1/(\log n)^m] : g = g_{1n} \text{ or } g_{2n}\} \geq 1/4,$$

and hence

$$\lim_{n\to\infty} (\log n)^m \inf_{\tilde{g}_n} \max\{E_g||\tilde{g}_n - g||_{w,2} : g = g_{1n} \text{ or } g_{2n}\} \geq \varepsilon_1/4 > 0.$$

This proves Theorem 3. □

We close this section with the following consequence of Corollary 3 and Theorem 3. Suppose $\theta^* < \infty$ and that there exist constants $\beta_1 > -1$, $\beta_2 > -1$, $\gamma_0 > 0$, $\gamma_1 \geq 1$ and $\gamma \geq 0$ such that

$$w(\theta) = C^2(\theta)\theta^{-\beta_1}(\theta^* - \theta)^{-\beta_2}, \quad \forall 0 < \theta < \theta^*,$$

and

$$q(x)\gamma_0\gamma_1^x(x!)^\gamma > 1, \quad \forall x \geq 0.$$

Then

(a) if $\gamma = 0$, the optimal convergence rate with respect to $||.||_{w,2}$ loss is $(1/\log n)^m$ for mixing densities g in the class $\mathcal{G}(m, m, M, w_0)$ where w_0 is as in (3). This rate is attained by the mixing density estimators $\hat{g}_n$ of Corollary 3.

(b) if $0 < \gamma < \infty$, the convergence rate [namely $(\log\log n/\log n)^m$] of the estimators of Corollary 3 almost achieve the lower bound of $(1/\log n)^m$ obtained in Theorem 3 for mixing densities within the class $\mathcal{G}(m, m, M, w_0)$.

4 Appendix

Proof of Proposition 2.

Since

$$\begin{aligned} 0 &= \lim_{\theta\to 0+} h^{(m-i)}(\theta)[p_{w_1,j}(\theta)w_1(\theta)]^{(i-1)} \\ &= \lim_{\theta\to\theta^*-} h^{(m-i)}(\theta)[p_{w_1,j}(\theta)w_1(\theta)]^{(i-1)} \end{aligned}$$

whenever $0 < i < m$, we observe from (9) and repeated integration by parts that

$$\begin{aligned} \int_0^{\theta^*} h^{(m)}(\theta)p_{w_1,j}(\theta)w_1(\theta)d\theta &= (-1)^m \int_0^{\theta^*} h(\theta)[p_{w_1,j}(\theta)w_1(\theta)]^{(m)}d\theta \\ &= \nu_{j+m,m}\int_0^{\theta^*} h(\theta)p_{w_0,j+m}(\theta)w_0(\theta)d\theta \\ &= \nu_{j+m,m}h_{w_0,j+m}, \quad \forall j \geq 0. \end{aligned}$$

From the completeness of $\{p_{w_1,j}\}_{j=0}^{\infty}$, we obtain

$$||h^{(m)}||_{w_1,2}^2 = \sum_{j=m}^{\infty} \nu_{j,m}^2 h_{w_0,j}^2.$$

Now the proposition follows immediately from (10). □

Lemma 1 *Let $l_{u,v}$ be given by (17). Then,*

(29) $$l_{u,v}(\theta) = \theta^{-1}\sqrt{u/(2\pi)}\exp[-u(\theta/a - 1 - \log(\theta/a)) - \varepsilon_u],$$

where $(12u+1)^{-1} \leq \varepsilon_u \leq (12u)^{-1}$. Also as $u \to \infty$, we have for $1 \leq p < \infty$

$$\begin{aligned} ||l_{u,v}||_p &= (v^u/\Gamma(u))[(pv)^{pu-p+1}\Gamma(pu-p+1)]^{1/p} \\ &\approx (v^2/(2\pi u))^{(p-1)/(2p)}p^{-1/(2p)}. \end{aligned}$$

In addition, there exists constants c_j^* such that

$$|l_{u,v}^{(j)}(\theta)|/l_{u,v}(\theta) \leq c_j^*\theta^{-j}[1 + |u-1-v\theta|^j + (v\theta)^{j/2}], \quad \forall j \geq 0,$$

and as $u \to \infty$

$$||l_{u,v}^{(j)}||_p/||l_{u,v}||_p \approx (u/v)^{-j}u^{j/2}[E|Q_j(p^{-1/2}Z)|^p]^{1/p}, \quad \forall j \geq 0,$$

where Z is a N(0,1) random variable and $Q_j(x)$ are polynomials such that

$$Q_{j+1}(x) = xQ_j(x) - \frac{d}{dx}Q_j(x), \quad Q_0(x) = 1.$$

REMARK. The Burkholder-Davis-Gundy inequality can be used to obtain an upper bound for $||l_{u,v}^{(j)}||_p$ and it can be proved by mathematical induction that

$$Q_j(x) = \sum_{0 \le i \le j/2} \frac{j!(-1)^i x^{j-2i}}{(j-2i)!i!2^i}.$$

It is also clear that $0 < E|Q_j(p^{-1/2}Z)|^p < \infty$.

PROOF OF LEMMA 1. (29) follows from the Stirling formula $\Gamma(u+1) > u^{u+1/2}e^{-u}\sqrt{2\pi}$. Next since $l_{u,v}$ is a gamma density,

$$l_{u,v}^p(\theta) = ||l_{u,v}||_p^p l_{p(u-1)+1,pv}(\theta),$$

so that $||l_{u,v}||_p^p = (v^u/\Gamma(u))^p\Gamma(pu-p+1)/(pv)^{pu-p+1}$. By the Stirling formula,

$$\begin{aligned} & (v^u/\Gamma(u))^p\Gamma(pu-p+1)/(pv)^{pu-p+1} \\ \approx\ & \frac{v^{up}(pu)^{pu-p+1/2}e^{-pu}\sqrt{2\pi}}{(u^{u-1/2}e^{-u}\sqrt{2\pi})^p(pv)^{pu-p+1}} \\ =\ & v^{up-pu+p-1}u^{pu-p+1/2-pu+p/2}p^{pu-p+1/2-pu+p-1}(2\pi)^{(1-p)/2} \\ =\ & v^{p-1}u^{(1-p)/2}p^{-1/2}(2\pi)^{(1-p)/2}, \end{aligned}$$

which implies the approximation for $||l_{u,v}||_p$.

Let $Q_j^*(x,y)$ be a function such that

$$l_{u,v}^{(j)}(\theta) = l_{u,v}(\theta)\theta^{-j}Q_j^*(u-1-v\theta, \sqrt{v\theta}).$$

Clearly $Q_0^* = 1$. Since $[\log l_{u,v}(\theta)]^{(1)} = (u-1)/\theta - v$,

$$\begin{aligned} & \theta^{-j-1}Q_{j+1}^*(u-1-v\theta, \sqrt{v\theta}) \\ =\ & [(u-1)/\theta - v]\theta^{-j}Q_j^* - j\theta^{-j-1}Q_j^* + \theta^{-j}\left[-vQ_{j,1}^* + \sqrt{v\theta}Q_{j,2}^*/(2\theta)\right], \end{aligned}$$

where $Q_{j,1}^*(x,y) = (\partial/\partial x)Q_j^*(x,y)$ and $Q_{j,2}^*(x,y) = (\partial/\partial y)Q_j^*(x,y)$. It follows that

$$Q_{j+1}^*(x,y) = xQ_j^*(x,y) - jQ_j^*(x,y) - y^2Q_{j,1}^*(x,y) + (y/2)Q_{j,2}^*(x,y),$$

so that $Q_j^*(x,y)$ is a polynomial of degree j. This gives the inequality for $|l_{u,v}^{(j)}(\theta)|$.

For the $\|\cdot\|_p$ norm, we have $|l_{u,v}^{(j)}|^p/\|l_{u,v}\|_p^p = |\theta^{-j}Q_j^*|^p l_{p(u-1)+1,pv}$. Since $l_{u,v}$ has mean u/v and variance u/v^2, by the central limit theorem

$$(v/\sqrt{u})(\theta - u/v) = (v\theta - u)/\sqrt{u} \to Z$$

in distribution under the density $l_{u,v}$ as $u \to \infty$. Since gamma distributions have finite moments,

$$\int_0^\infty |Q\left((u - v\theta)/\sqrt{u}, \sqrt{v\theta/u}\right)|^p l_{u,v}(\theta)d\theta \to E|Q(Z,1)|^p$$

for all polynomials Q and $p \geq 0$. Therefore, as $u \to \infty$

$$\begin{aligned} & \|l_{u,v}^{(j)}\|_p^p/\|l_{u,v}\|_p^p \\ = & \int |\theta^{-j}Q_j^*(u - 1 - v\theta, \sqrt{v\theta})|^p l_{p(u-1)+1,pv}(\theta)d\theta \\ \approx & E|(u/v)^{-j}Q_j^*(Z\sqrt{pu}/p, \sqrt{u})|^p, \\ \approx & (u/v)^{-j}u^{j/2}E|Q_j(Z/\sqrt{p})|^p, \end{aligned}$$

where $Q_j(x) = Q_j(x,1)$ and $Q_j(x,y)$ is the sum of all terms of degree j in $Q_j^*(x,y)$. The recursion of Q_j follows from that of Q_j^*. □

Lemma 2 *Let $\mathcal{P}$ be a class of probability measures and T be a mapping from $\mathcal{P}$ to a metric space with distance function $d(.,.)$. Let f_j be the joint densities of observations $X_1, \cdots, X_n$ under $P_j \in \mathcal{P}$, $j = 1,2$. If $\rho \leq P_1(f_1 \leq \lambda f_2)$, then for any estimator $\tilde{T}_n$ based on $X_1, \cdots, X_n$,*

$$\max_{j=1,2} P_j\{d(\tilde{T}_n, T(P_j)) \geq d(T(P_1), T(P_2))/2\} \geq \rho/(1+\lambda).$$

PROOF. Lemma 2 follows directly from the argumant in Zhang (1990, top of page 827). □

References

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

[2] DATTA, S. (1991). On the consistency of posterior mixtures and its applications. *Ann. Statist.* **19** 338-353.

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

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

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

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

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

[9] JOHNSTONE, I. M. and SILVERMAN, B. W. (1990). Speed of estimation in positron emission tomography and related inverse problems. *Ann. Statist.* **18** 251-280.

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

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

[12] LOH, W. L. and ZHANG, C. H. (1993). Global properties of kernel estimators for mixing densities in exponential family models for discrete variables. Preprint.

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

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

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

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

[17] SZEGÖ, G. (1975). *Orthogonal Polynomials.* Amer. Math. Soc., Providence, Rhode Island.

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

[19] WALTER, G. G. (1985). Orthogonal polynomial estimators of the prior distribution of a compound Poisson distribution. *Sankhya* **47** A 222-230.

[20] WALTER, G. G. and HAMEDANI, G. G. (1989). Bayes empirical Bayes estimation for discrete exponential families. *Ann. Inst. Statist. Math.* **41** 101-119.

[21] WALTER, G. G. and HAMEDANI, G. G. (1991). Bayes empirical Bayes estimation for natural exponential families with quadratic variance functions. *Ann. Statist.* **19** 1191-1224.

[22] ZHANG, C. H. (1988). Fourier methods for estimating mixing densities and distributions. Manuscript.

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

[24] ZHANG, C. H. (1992). On estimating mixing densities in exponential family models for discrete variables. Preprint.