



On Estimating Mixing Densities in Discrete Exponential Family Models

Cun-Hui Zhang

The Annals of Statistics, Vol. 23, No. 3. (Jun., 1995), pp. 929-945.

Stable URL:

<http://links.jstor.org/sici?sici=0090-5364%28199506%2923%3A3%3C929%3AOEMDID%3E2.0.CO%3B2-N>

The Annals of Statistics is currently published by Institute of Mathematical Statistics.

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/about/terms.html>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at <http://www.jstor.org/journals/ims.html>.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is an independent not-for-profit organization dedicated to creating and preserving a digital archive of scholarly journals. For more information regarding JSTOR, please contact support@jstor.org.

ON ESTIMATING MIXING DENSITIES IN DISCRETE EXPONENTIAL FAMILY MODELS¹

BY CUN-HUI ZHANG

Rutgers University

This paper concerns estimating a mixing density function g and its derivatives based on iid observations from $f(x) = \int f(x | \theta)g(\theta) d\theta$, where $f(x | \theta)$ is a known exponential family of density functions with respect to the counting measure on the set of nonnegative integers. Fourier methods are used to derive kernel estimators, upper bounds for their rate of convergence and lower bounds for the optimal rate of convergence. If $f(x | \theta_0) \geq \varepsilon^{x+1} \forall x$, for some positive numbers θ_0 and ε , then our estimators achieve the optimal rate of convergence $(\log n)^{-\alpha+m}$ for estimating the m th derivative of g under a Lipschitz condition of order $\alpha > m$. The optimal rate of convergence is almost achieved when $(x!)^\beta f(x | \theta_0) \geq \varepsilon^{x+1}$. Estimation of the mixing distribution function is also considered.

1. Introduction. Let $f(x | \theta)$ 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

$$(1) \quad f(x) = f(x; g) = \int f(x | \theta)g(\theta) d\theta.$$

We are interested in estimating the mixing density function g and its derivatives based on independent identically distributed observations $X_1, \dots, X_n$ from $f(x)$. When $f(x | \theta)$ is a location family and μ is the Lebesgue measure, this is also called the deconvolution problem, which was investigated recently by Carroll and Hall (1988), Devroye (1989), Edelman (1988), Fan (1991a, b), Stefanski (1990) and Zhang (1990) among others. The purpose of this paper is to study the case where

$$(2) \quad f(x | \theta) = C(\theta)q(x)\theta^x, \quad x = 0, 1, 2, \dots, \quad 0 \leq \theta \leq (\text{or } <) \theta^* \leq \infty,$$

and μ is the counting measure on the set of nonnegative integers. This model includes the Poisson, binomial and negative binomial distributions.

In Section 2 we use Fourier methods to construct kernel estimators for the mixing density g and its derivatives, and to obtain upper bounds for their mean squared error at a fixed point under the condition that the value of θ^*

Received June 1992; revised July 1994.

¹ Supported in part by the NSF and Army Research Office.

AMS 1991 subject classifications. Primary 62G05; secondary 62G20.

Key words and phrases. Mixture, mixing distribution, kernel estimator, rate of convergence, Fourier transformation.

is finite and known. Let $g^{(m)}$ be the m th derivative of g , $g^{(0)} = g$. For positive numbers α , $\alpha < \theta^*$, δ , M and M_0 , define

$$\begin{aligned} \mathcal{G}_{\alpha, \theta^*} &= \mathcal{G}_{\alpha, \theta^*}(\alpha, \delta, M, M_0) \\ (3) \quad &= \left\{ g : |g^{(\alpha')}(\theta) - g^{(\alpha')}(a)| \leq M|\theta - a|^{\alpha''}, \forall |a - \theta| \leq \delta, \right. \\ &\quad \left. \sum_{j=0}^{\alpha'} |g^{(j)}(a)| \leq M_0, \int_0^{\theta^*} g(\theta) d\theta = 1 \right\}. \end{aligned}$$

Here and in the sequel, α' is always the largest integer less than α and $\alpha'' = \alpha - \alpha' \in (0, 1]$. If there exist nonnegative numbers β , B_0 and B such that $q(x)B_0 B^x (x!)^\beta \geq 1$, $\forall x$, then the mean squared error of our kernel estimator for $g^{(m)}$, $0 \leq m < \alpha$, is bounded by $O((\log n)^{-\alpha+m})$ uniformly over the class $\mathcal{G}_{\alpha, \theta^*}$ when $\beta = 0$ and by $O((\log n / \log \log n)^{-\alpha+m})$ when $\beta > 0$. In Section 3 we obtain $(\log n)^{-\alpha+m}$ as a lower bound for the minimax mean squared error for estimating $g^{(m)}$ over $\mathcal{G}_{\alpha, \theta^*}$. Consequently, our kernel estimators achieve the optimal rate of convergence when $\beta = 0$ (e.g., the negative binomial family) and almost achieve the optimal rate of convergence when $0 < \beta < \infty$, (e.g., the Poisson family). In Section 4 we derive kernel estimators for the case $\theta^* = \infty$ under the condition $\beta < 2$, which achieve the rate of $(\log n)^{-(1-\beta/2)(\alpha-m)}$ for the mean squared error. Section 5 concerns estimating the mixing distribution function.

Nonparametric maximum likelihood estimation of mixing densities for families satisfying (2) has been studied by Kiefer and Wolfowitz (1956), Laird (1978), Lambert and Tierney (1984), Lindsay (1983a, b) and Simar (1976) among others. Deely and Kruse (1968) considered minimum distance estimators; Rolph (1968) and Walter and Hamedani (1991) considered Bayes estimators, whereas Tucker (1963) used the method of moments. Although the nonparametric maximum likelihood and other estimators are consistent under quite general conditions (e.g., $\theta^* < \infty$), their rate of convergence is still unknown. Mixture models and their applications were proposed by Robbins (1951, 1956), Kiefer and Wolfowitz (1956) and many others in various contexts. The identifiability of mixing distributions was studied by Teicher (1961).

Throughout this paper we shall use the following: $P = P_g$ and $E = E_g$ denote the probability and expectation corresponding to g , respectively; $\|\cdot\|_p$ denotes the L^p -norm, $p = 1, \infty$; $h^{(j)}$ denotes the j th derivative (if it exists) of any function h ($h^{(0)} = h$); and h^* denotes the Fourier transformation of any integrable function h , so that $h^*(t) = \int \exp(itx)h(x) dx$.

2. Kernel estimators. In this section, we shall give kernel estimators for $g^{(m)}$ and upper bounds for their rate of convergence under the assumption that the value of θ^* in (2) is finite and known. The case of $\theta^* = \infty$ is considered in Section 4.

Define an even function

$$(4) \quad h(\theta) = C(|\theta|)g(|\theta|), \quad -\infty < \theta < \infty.$$

It follows from (1) and (2) that, for $q(x) > 0$,

$$(5) \quad \frac{P_g\{X=x\}}{q(x)} = \frac{f(x;g)}{q(x)} = \int_0^{\theta^*} \theta^x C(\theta)g(\theta) d\theta = \int_0^{\theta^*} \theta^x h(\theta) d\theta,$$

so that $f(x)/q(x)$ is the x th moment of h . Since $q(\cdot)$ and $C(\cdot)$ are known functions, our problem is essentially to estimate $h(\cdot)$ based on the observed frequency distribution from f . This is related to the moment problem [Feller (1971), pages 227–228 and 514–515].

We shall first construct kernels for $h^{(j)}(a)$ by finding functions $K_n^{(j)}(x, a)$ such that $E K_n^{(j)}(X, a) = \int_{-\infty}^{\infty} h^{(j)}(a - \theta/c_n)k(\theta) d\theta + o(1)$ for some suitable $c_n \rightarrow \infty$ and $k(\cdot)$. Let $k(x)$ be a function satisfying

$$(6) \quad \int x^j k(x) dx = I\{j=0\}, \quad \|x^{\alpha_0+1}k^{(j)}(x)\|_{\infty} < \infty, \quad 0 \leq j < \alpha_0, \\ \|x^{\alpha_0}k(x)\|_1 < \infty,$$

for a positive number α_0 , and

$$(7) \quad k(x) = k(-x) \quad \forall x, \quad k^*(t) = 0 \quad \forall |t| > 1.$$

Note that $\|x^{\alpha_0+1}k^{(j)}(x)\|_{\infty} < \infty$ for all integers $j \geq 0$ if (7) holds and $\|x^{\alpha_0+1}k(x)\|_1 < \infty$. Since $h(\cdot)$ and $k(\cdot)$ are both even functions, for $c > 0$,

$$\int_{-\infty}^{\infty} h\left(a - \frac{\theta}{c}\right)k(\theta) d\theta = \int_0^{\infty} \{ck(c(a - \theta)) + ck(c(a + \theta))\}h(\theta) d\theta.$$

By (7), $k^*(t)$ is a real even function with a compact support, so that, by the Fourier inversion formula,

$$\begin{aligned} & ck(c(a - \theta)) + ck(c(a + \theta)) \\ &= \left(\frac{c}{2\pi}\right) \int \{\exp[-itc(a - \theta)] + \exp[-itc(a + \theta)]\}k^*(t) dt \\ &= \frac{\int \cos(t\theta)\cos(ta)k^*(t/c) dt}{\pi}. \end{aligned}$$

Taking the infinite series expansion $\cos(t\theta) = \sum_{x=0}^{\infty} (t\theta)^x \cos(x\pi/2)/x!$, we find by (5) that

$$\int_{-\infty}^{\infty} h\left(a - \frac{\theta}{c}\right)k(\theta) d\theta = \sum_{x=0}^{\infty} \frac{f(x;g)}{q(x)x!\pi} \int t^x \cos\left(\frac{x\pi}{2}\right)\cos(ta)k^*\left(\frac{t}{c}\right) dt.$$

Differentiating with respect to a under the integrations, we obtain, for integers $j \geq 0$,

$$\begin{aligned} & \int_{-\infty}^{\infty} h^{(j)}\left(a - \frac{\theta}{c}\right)k(\theta) d\theta \\ &= \sum_{x=0}^{\infty} \frac{f(x;g)}{\pi q(x)x!} \int \cos\left(ta - \frac{j\pi}{2}\right)t^{x+j} \cos\left(\frac{x\pi}{2}\right)k^*\left(\frac{t}{c}\right) dt. \end{aligned}$$

This motivates the kernel for $h^{(j)}(a)$,

$$(8) \quad K_n^{(j)}(x, a) = \frac{1}{\pi q(x) x!} \int \cos\left(ta - \frac{j\pi}{2}\right) t^{x+j} \cos\left(\frac{x\pi}{2}\right) \\ \times k^*\left(\frac{t}{c_n}\right) dt I\{0 \leq x \leq m_n\},$$

provided that $q(x) > 0$ for all nonnegative *even* integers $x \leq m_n$. Since $\cos(x\pi/2) = 0$ when x is odd, $K_n^{(j)}(x, a)$ is defined to be 0 unless x is a nonnegative even integer. For $m_n = \infty$, $E_g K_n^{(j)}(X, a) = \int_{-\infty}^{\infty} h^{(j)}(a - \theta/c_n) k(\theta) d\theta \rightarrow h^{(j)}(a)$ as $c_n \rightarrow \infty$ when $h^{(j)}(\cdot)$ is continuous at a . In general, we can reduce the variance of $K_n^{(j)}(X, a)$ by choosing suitable finite m_n , which may depend on the unknown g . For $\theta^* < \infty$, Theorem 1 gives proper choices of m_n and c_n , which depend on g only through θ^* .

The bias of our kernel (8) is

$$(9) \quad EK_n^{(j)}(X, a) - h^{(j)}(a) = b_{1n}^{(j)}(a) + b_{2n}^{(j)}(a),$$

where $b_{1n}^{(j)}(a) = \int \{h^{(j)}(a - \theta/c_n) - h^{(j)}(a)\} k(\theta) d\theta$ and

$$b_{2n}^{(j)}(a) = \sum_{x > m_n} \{\pi x!\}^{-1} \int_0^{\theta^*} \theta^x h(\theta) d\theta \int \cos\left(ta - \frac{j\pi}{2}\right) t^{x+j} \cos\left(\frac{x\pi}{2}\right) k^*\left(\frac{t}{c_n}\right) dt.$$

The first term $b_{1n}^{(j)}(a)$ in (9) is well understood by the standard theory of kernel density estimation [cf. Prakasa Rao (1983)]. Theorem 1 provides upper bounds for the second term $b_{2n}^{(j)}(a)$ and the variance which do not depend on g .

THEOREM 1. *Let the kernel $K_n^{(j)}(x, a)$ be given by (8). Suppose the value of θ^* in (2) is finite. Let $\alpha \geq 0$ and $0 < \beta_0 < \frac{1}{2}$, and choose m_n and c_n such that*

$$(10) \quad \max\left\{\log\left(\frac{1}{q(2x)}\right), x = 0, 1, \dots, \left\lceil \frac{m_n}{2} \right\rceil\right\} + c_n \leq \beta_0 \log n,$$

$$(11) \quad (\theta^* e) c_n + \alpha \log c_n \leq m_n + 1,$$

where $[c]$ is the integer part of c . Then,

$$(12) \quad \sqrt{E|K_n^{(j)}(X, a)|^2} \leq \sup_{x, a} |K_n^{(j)}(x, a)| \leq B_1 c_n^j n^{\beta_0}, \quad |b_{2n}^{(j)}(a)| \leq B_2 c_n^{-\alpha+j},$$

where $B_1 = 2\|k\|_1/\pi$ and $B_2 = 2C(0)\|k\|_1/(\pi\theta^*)$.

REMARK. It is worthwhile noting that the above derivation of the kernel and the results in Theorem 1 are valid as long as $f(x | \theta) = C(\theta)q(x)\theta^x > 0$ on the set of even integers $0 \leq x \leq m_n$ (possibly after reparametrization and data transformation), even when $P\{X = \text{integers} | \theta\} < 1$. If $q(x) > 0$ for all nonnegative integers x , we may consider the family $f(y | \eta) = C(\eta^2)q(y/2)\eta^y$ with $Y = 2X$ and $\eta = \sqrt{\theta}$, so that the kernel will utilize the observed fre-

quencies at all nonnegative integers. This reparametrization technique improves the rate of mean squared error when $\theta^* = \infty$, considered in Section 4. Define $c_n(y) = \beta_0 \log n + \min_{0 \leq x \leq y/2} \log q(2x)$. Then $\theta^* e c_n(y) + \alpha \log c_n(y)$ is decreasing in y and crosses the straight line y at a unique point y_n . Clearly, both (10) and (11) hold for

$$(13) \quad c_n = c_n(y_n), \quad m_n \leq y_n < m_n + 1.$$

PROOF OF THEOREM 1. By (8), for even $x \leq m_n$,

$$\begin{aligned} |K_n^{(j)}(x, a)| &\leq \{\pi q(x) x!\}^{-1} 2 \|k\|_1 \frac{c_n^{x+j+1}}{x+j+1} \\ &\leq \left(2 \frac{\|k\|_1}{\pi}\right) c_n^j \exp[-\log q(x) + c_n] \\ &\leq B_1 c_n^j n^{\beta_0}. \end{aligned}$$

For the bias $b_{2n}^{(j)}(a)$ in (9) we have

$$\begin{aligned} |b_{2n}^{(j)}(a)| &\leq \sum_{x > m_n} (\pi x!)^{-1} (\theta^*)^x C(0) c_n^{x+j+1} \frac{2 \|k\|_1}{x+j+1} \\ &\leq C(0) 2 \|k\|_1 (\pi \theta^*)^{-1} c_n^j \sum_{x > m_n} \frac{(\theta^* c_n)^{x+1}}{(x+1)!}. \end{aligned}$$

The proof is complete, since

$$\begin{aligned} \sum_{x > m_n} \frac{(\theta^* c_n)^{x+1}}{(x+1)!} &\leq \left\{ \frac{\theta^* c_n}{m_n + 1} \right\}^{m_n+1} \sum_{x > m_n} \frac{(m_n + 1)^{x+1}}{(x+1)!} \\ &\leq \left\{ \frac{e \theta^* c_n}{m_n + 1} \right\}^{m_n+1} \leq c_n^{-\alpha}. \quad \square \end{aligned}$$

Consider the estimation of $g^{(m)}(a)$. By (4) and (2), $g(\theta) = h(\theta) \sum_{x=0}^{\infty} q(x) \theta^x$, so that

$$(14) \quad g^{(m)}(a) = \sum_{j=0}^m C_{j,m}(a) h^{(j)}(a), \quad C_{j,m}(\theta) = \binom{m}{j} \sum_{x=m-j}^{\infty} \frac{q(x) x! \theta^{x-m+j}}{(x-m+j)!}.$$

Since $C_{j,m}(a)$ are known constants, we can estimate $g^{(m)}(a)$ by

$$(15) \quad \hat{g}_n^{(m)}(a) = \sum_{j=0}^m C_{j,m}(a) \hat{h}_n^{(j)}(a), \quad \hat{h}_n^{(j)}(a) = n^{-1} \sum_{i=0}^n K_n^{(j)}(X_i, a).$$

THEOREM 2. Let $\mathcal{G}_{\alpha, \theta^*}$ be given by (3), and $\hat{g}_n^{(m)}(a)$ by (15) with the kernel $K_n^{(j)}(x, a)$ in (8) satisfying (6) and (7), $\alpha_0 \geq \alpha$. Choose m_n and c_n by (13). Then, for $0 \leq m < \alpha$ and $0 < a < \theta^* < \infty$,

$$\sup \{ |E_g[\hat{g}_n^{(m)}(a) - g^{(m)}(a)]|^2 : g \in \mathcal{G}_{\alpha, \theta^*} \} = O(c_n^{-2(\alpha-m)}).$$

REMARK. By (13), the rate of convergence $O(c_n^{-\alpha+m})$ is determined by the rate at which $q(x) \rightarrow 0$. The faster $q(x) \rightarrow 0$, the slower $E_g|\hat{g}_n^{(m)}(a) - g^{(m)}(a)|^2 \rightarrow 0$. This is expected, as $q(x)$ is the denominator of $\int \theta^x h(\theta) d\theta = f(x; g)/q(x)$ through which $g(\cdot)$ is estimated. The estimator $\hat{g}_n^{(m)}$ depends on α only through c_n in (13). Since $\delta > 0$ can be arbitrarily small with $\mathcal{G}_{\alpha, \theta^*}(a, \delta, M, M_0)$, the rate of convergence is characterized by local properties of g near a . If $\delta \geq a$ and $g^{(j)}(0) = 0$ for $0 \leq j < \alpha$, then the Lipschitz condition in (3) implies $\sum_{j=0}^{\alpha'} |g^{(j)}(a)| \leq M_0 = 2Ma^{\alpha'} e^a$.

PROOF OF THEOREM 2. Due to the smoothness of $C(\cdot)$, $g \rightarrow h$ maps $\mathcal{G}_{\alpha, \theta^*}(a, \delta, M, M_0)$ into $\mathcal{G}_{\alpha, \theta^*}(a, \delta, M', M'_0)$ for some (large) M' and M'_0 . Since the kernel is bounded for each n by Theorem 1, $E_g\{\hat{g}_n^{(m)}(a)\}^j$, $j = 1, 2$, are continuous under the convergence in distribution as functionals of $G(\theta) = \int_0^\theta g(y) dy$, so that the supremum can be taken over all $g \in \mathcal{G}_{\alpha, \theta^*}$ such that (4) is α' times continuously differentiable on $(-\infty, \infty)$. The bias term $b_{1n}^{(j)}(a) = \int \{h^{(j)}(a - \theta/c_n) - h^{(j)}(a)\}k(\theta) d\theta$ in (9) can be split into two integrals: $\int_{|\theta| > \delta c_n}$ and $\int_{|\theta| \leq \delta c_n}$. Integration by parts is used toward $h^{(0)}(a - \theta/c)$ for the first one with the boundary values of $O(c_n^{-\alpha+j})$ due to (6) and the uniform boundedness of $h^{(j)}(a \pm \delta)$ and $h^{(j)}(a)$, $0 \leq j \leq \alpha'$. It follows from (6) and standard methods in density estimation that

$$\begin{aligned} b_{1n}^{(j)}(a) &= O(c_n^{-\alpha+j}) + (-1)^j c_n^j \int_{|\theta| > \delta c_n} h\left(a - \frac{\theta}{c_n}\right) k^{(j)}(\theta) d\theta + \int_{|\theta| \leq \delta c_n} \\ &= O(c_n^{-\alpha+j}), \end{aligned}$$

$0 \leq j < \alpha$. Since $c_n = O(\log n)$, equation (15) and Theorem 1 imply

$$\begin{aligned} E_g|\hat{g}_n^{(m)}(a) - g^{(m)}(a)|^2 &\leq O(B_1^2 c_n^{2m} n^{2\beta_0-1}) + O(c_n^{-2(\alpha-m)}) \\ &= O(c_n^{-2(\alpha-m)}). \end{aligned} \quad \square$$

COROLLARY 1. Suppose there exist nonnegative constants B_0 , B and β such that, for all even $x \geq 0$,

$$(16) \quad q(x) B_0 B^x (x!)^\beta \geq 1$$

and the conditions of Theorem 2 hold. Then, for $0 < a < \theta^*$,

$$\begin{aligned} &\sup\{E_g|\hat{g}_n^{(m)}(a) - g^{(m)}(a)|^2 : g \in \mathcal{G}_{\alpha, \theta^*}\} \\ &= \begin{cases} O(1)(\log n)^{-2(\alpha-m)}, & \text{if } \beta = 0, \\ O(1)\left(\frac{\log n}{\log \log n}\right)^{-2(\alpha-m)}, & \text{if } 0 < \beta < \infty. \end{cases} \end{aligned}$$

The value of β is 0 for the negative binomial family and 1 for the Poisson family.

PROOF OF COROLLARY 1. By (16) and the Stirling formula, $-\log q(x) \leq x \log B + \beta\{(x + \frac{1}{2})\log x - x\} + O(1)$, so that the order of c_n and m_n is $(\log \log n)^{-I(\beta > 0)} \log n$ by (13). $\square$

3. Lower bounds for the optimal rate of convergence. Let us characterize the degree of difficulty for estimating $g^{(m)}(a)$ by the minimax mean squared error

$$(17) \quad r_{n,\alpha,\theta^*} = \inf_{\tilde{g}_n^{(m)}} \sup \left\{ E_g |\tilde{g}_n^{(m)}(a) - g^{(m)}(a)|^2 : g \in \mathcal{G}_{\alpha,\theta^*} \right\},$$

where $\mathcal{G}_{\alpha,\theta^*}$ is given by (3), and the infimum runs over all statistics $\tilde{g}_n^{(m)}$ based on $X_1, \dots, X_n$. In this section we derive lower bounds for the rate at which r_{n,α,θ^*} tends to 0, which is called the optimal rate of convergence over the class $\mathcal{G}_{\alpha,\theta^*}$. The basic idea is to find pairs of densities g_{1n} and g_{2n} in $\mathcal{G}_{\alpha,\theta^*}$ such that $g_{1n}^{(m)}(a) - g_{2n}^{(m)}(a)$ tends to 0 at a much slower rate than $f(\cdot; g_{1n}) - f(\cdot; g_{2n})$ does.

THEOREM 3. *Let $\mathcal{G}_{\alpha,\theta^*}$ and r_{n,α,θ^*} be given by (3) and (17), respectively, $0 < a < \theta^*$. Then there exists $\varepsilon_0 > 0$ such that*

$$\liminf_{n \rightarrow \infty} \inf_{\tilde{g}_n^{(m)}} \sup_{g \in \mathcal{G}_{\alpha,\theta^*}} P_g \{ |\tilde{g}_n^{(m)}(a) - g^{(m)}(a)| > \varepsilon_0 (\log n)^{-\alpha+m} \} > 0.$$

Consequently, $\liminf_{n \rightarrow \infty} (\log n)^{2(\alpha-m)} r_{n,\alpha,\theta^*} > 0$.

It follows from this theorem and Corollary 1 that the optimal rate of convergence is achieved by our kernel estimators for the negative binomial family and is almost achieved for the Poisson family.

The densities g_{1n} and g_{2n} are constructed in the following manner. Let a , θ_0 and θ_1 be fixed constants satisfying $0 < a < \theta_0 < \theta_1 < \theta^*$. Set

$$(18) \quad h_{u,v} = h_{u,v}(\theta) = \frac{v^u \theta^{u-1} \exp(-v\theta)}{\Gamma(u)}, \quad \frac{u}{v} = a,$$

$$(19) \quad g_{u,v} = g_{u,v}(\theta) = \frac{h_{u,v}(\theta)}{C(\theta)} = \sum_{x=0}^{\infty} q(x) \theta^x h_{u,v}(\theta),$$

$$(20) \quad w = w(u, v) = \int_0^{\theta_0} \cos\left(\frac{u(\theta - a)}{\theta_0} - \frac{m\pi}{2}\right) g_{u,v}(\theta) d\theta.$$

Let $\varepsilon_0 > 0$, and let g_0 be a density in $\mathcal{G}_{\alpha,\theta^*}(a, \delta, M - \varepsilon_0, M_0 - \varepsilon_0)$ such that

$$(21) \quad g_0(\theta) \geq \min(\theta^{1/\varepsilon_0}, \varepsilon_0) I\{0 \leq \theta \leq \theta_0\}.$$

Choose constants u_n and $v_n = u_n/a$ such that

$$(22) \quad \frac{u_n}{\log n} = \delta_0 \geq \max \left\{ \frac{\theta_0/(\theta_1 - \theta_0)}{\log(\theta_1/\theta_0)}, \frac{1}{\theta_0/a - 1 - \log(\theta_0/a)}, \frac{2}{\log(1 + a^2/\theta_0^2)} \right\}.$$

Define $g_{1n} = g_0$ and

$$(23) \quad g_{2n}(\theta) = g_0(\theta) + \frac{w_0}{\sqrt{u_n}} \left(\frac{\theta_0}{u_n} \right)^\alpha \left[\cos \left(u_n \frac{\theta - a}{\theta_0} - \frac{m\pi}{2} \right) - w(u_n, v_n) \right] \\ \times g_{u_n, v_n}(\theta) I\{0 \leq \theta \leq \theta_0\}.$$

We shall show in the proof of Theorem 3 that g_{1n} and g_{2n} are members of $\mathcal{G}_{\alpha, \theta^*}$ for small w_0 , and

$$(24) \quad p_n = \inf_{\tilde{g}_n^{(m)}} \max \left\{ P_{\tilde{g}} \{ |\tilde{g}_n^{(m)}(a) - g^{(m)}(a)| > \varepsilon_n (\log n)^{-\alpha+m} \} : \right. \\ \left. g = g_{1n} \text{ or } g_{2n} \right\} \rightarrow p_0,$$

for some $\varepsilon_n \rightarrow \varepsilon > 0$ and $p_0 > 0$. This will prove the theorem since $r_{n, \alpha, \theta^*} \geq (\varepsilon_n (\log n)^{-\alpha+m})^2 p_n$.

LEMMA 1. Let $\mathcal{P}$ be a class of probability measures, and let $T(P)$ be a mapping from $\mathcal{P}$ to a metric space with distance function $d(\cdot, \cdot)$. Let f_j be the joint densities of observations $X_1, \dots, X_n$ 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 $X_1, \dots, X_n$,

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

LEMMA 2. Let $h_{u,v}$ be given by (18). Then, for $j \geq 0$, $\alpha > 0$ and $x \geq 0$, the following hold:

$$(25) \quad h_{u,v}(\theta) = \theta^{-1} \sqrt{\frac{u}{2\pi}} \exp \left[-u \left(\frac{\theta}{a} - 1 - \log \left(\frac{\theta}{a} \right) \right) - \varepsilon_u \right], \\ \frac{1}{12u + 1} \leq \varepsilon_u \leq \frac{1}{12u};$$

$$(26) \quad |h_{u,v}^{(j)}(\theta)| \leq \frac{v^{j+1}}{2\pi} \frac{\Gamma((j+1)/2) \Gamma((u-j-1)/2)}{\Gamma(u/2)} = O(u^{(j+1)/2});$$

$$(27) \quad \begin{aligned} & |h_{u,v}^{(\alpha')}(\theta) - h_{u,v}^{(\alpha')}(a)| \\ & \leq |\theta - a|^{\alpha''} \frac{v^{\alpha+1}}{\pi} \frac{\Gamma((\alpha+1)/2)\Gamma((u-\alpha-1)/2)}{\Gamma(u/2)} \end{aligned}$$

$$(28) \quad \begin{aligned} & = O(|\theta - a|^{\alpha''} u^{(\alpha+1)/2}); \\ & \left| \int_0^\infty \cos\left(\frac{u(\theta - a)}{\theta_0} - \frac{m\pi}{2}\right) \theta^x h_{u,v}(\theta) d\theta \right| \\ & \leq \left(1 + \frac{a^2}{\theta_0^2}\right)^{-u/2} \left\{ \theta_0 \left(1 + \frac{x}{u}\right) \right\}^x; \end{aligned}$$

and

$$(29) \quad \int_\theta^\infty \theta^x h_{u,v}(\theta) d\theta \leq \exp\left[-u\left\{\frac{\theta_0}{a} - 1 - \log\left(\frac{\theta_0}{a}\right)\right\}\right] \left\{ \theta_0 \left(1 + \frac{x}{u}\right) \right\}^x.$$

Lemma 1 follows directly from the argument in Zhang [(1990), top of page 827]. Lemma 2 is proved at the end of the section.

PROOF OF THEOREM 3. We shall drop the subscript n in u_n and v_n throughout the proof.

Step 1. Verify $g_{2n} \in \mathcal{G}_{\alpha,\theta^*}$ ($g_{1n} = g_0 \in \mathcal{G}_{\alpha,\theta}$). It follows from (21), (25) and (20) that

$$\begin{aligned} g_0(\theta) & \geq \frac{w_0(1 + 1/C(\theta_0))}{C(\theta)\theta\sqrt{2\pi}} \left(\frac{\theta_0}{u}\right)^\alpha \exp\left[-u\left(\frac{\theta}{a} - 1 - \log\left(\frac{\theta}{a}\right)\right)\right] \\ & \geq \frac{w_0(1 + |w|)}{C(\theta)\sqrt{u}} \left(\frac{\theta_0}{u}\right)^\alpha h_{u,v}, \end{aligned}$$

for large u uniformly on $0 \leq \theta \leq \theta_0$, so that g_{2n} is a density on $[0, \theta^*]$ by (23). It follows from (23) and (14) that, for $0 \leq \theta \leq \theta_0$,

$$\begin{aligned} & g_{2n}^{(m)}(\theta) - g_0^{(m)}(\theta) \\ & = g_{2n}^{(m)}(\theta) - g_{1n}^{(m)}(\theta) \\ & = \frac{w_0}{\sqrt{u}} \left(\frac{\theta_0}{u}\right)^\alpha \left[-w(u, v) g_{u,v}^{(m)}(\theta) \right. \\ & \quad \left. + \sum_{j=0}^m \binom{m}{j} \left(\frac{u}{\theta_0}\right)^{m-j} \cos\left(u \frac{\theta - a}{\theta_0} - \frac{j\pi}{2}\right) g_{u,v}^{(j)}(\theta) \right] \end{aligned}$$

and $g_{u,v}^{(m)}(\theta) = \sum_{j=0}^m C_{j,m}(\theta) h_{u,v}^{(j)}(\theta)$. Since $C_{j,m}(\cdot)$ and $\cos(\cdot)$ are continuously differentiable functions, by (26), (27) and the fact that $|\cos(uy) - 1| \leq 2|uy|^{\alpha''}$ we have

$$\begin{aligned} & |g_{2n}^{(\alpha')}(\theta) - g_{2n}^{(\alpha')}(a)| \leq |\theta - a|^{\alpha''} (M - \varepsilon_0 + w_0 O(1)), \quad |\theta - a| \leq \delta, \\ & \sum_{j=0}^{\alpha'} |g_{2n}^{(j)}(a)| \leq M_0 - \varepsilon_0 + w_0 O(1), \end{aligned}$$

where the $O(1)$ is uniform in θ and does not depend on w_0 . Therefore, $g_{2n} \in \mathcal{G}_{\alpha, \theta^*}(a, \delta, M, M_0)$ for small w_0 . Also, since $u/\log n = \delta_0$ by (22), it follows from (23), (25) and (26) that

$$\begin{aligned} \varepsilon_n &= (\log n)^{\alpha-m} \frac{|g_{2n}^{(m)}(a) - g_{1n}^{(m)}(a)|}{2} \rightarrow \varepsilon \\ (30) \quad &= \frac{w_0}{2\alpha C(a)\sqrt{2\pi}} \left(\frac{\theta_0}{\delta_0}\right)^{\alpha-m} > 0. \end{aligned}$$

Step 2. Verify (24) for the ε_n in (30). We shall first find a good bound for

$$\begin{aligned} &\sum_{x=0}^{\infty} |f(x; g_{1n}) - f(x; g_{2n})| \\ (31) \quad &= \frac{w_0}{\sqrt{u}} \left(\frac{\theta_0}{u}\right)^{\alpha} \sum_{x=0}^{\infty} q(x) \left| \int_0^{\theta_0} \cos\left(u \frac{\theta-a}{\theta_0} - \frac{m\pi}{2}\right) - w \right| \theta^x h_{u,v}(\theta) d\theta \Big|. \end{aligned}$$

Set $x^* = \log n / \log(\theta_1/\theta_0)$. Then, by (22), $\theta_0(1+x/u) \leq \theta_1 \forall x \leq x^*$, and $(\theta_0/\theta_1)^{x^*} \leq 1/n$. Also, (22) implies $\exp[-u\{\theta_0/a - 1 - \log(\theta_0/a)\}] \leq 1/n$ and $(1+a^2/\theta_0^2)^{-u/2} \leq 1/n$. By (28) and (29),

$$\begin{aligned} &\sum_{x=0}^{\infty} q(x) \left| \int_0^{\theta_0} \cos\left(\frac{u(\theta-a)}{\theta_0} - \frac{m\pi}{2}\right) \theta^x h_{u,v}(\theta) d\theta \right| \\ &\leq \sum_{x > x^*} q(x) \theta_0^x + \sum_{x \leq x^*} q(x) \int_{\theta_0}^{\infty} \theta^x h_{u,v}(\theta) d\theta \\ &\quad + \sum_{x \leq x^*} q(x) \left| \int_0^{\infty} \cos\left(\frac{u(\theta-a)}{\theta_0} - \frac{m\pi}{2}\right) \theta^x h_{u,v}(\theta) d\theta \right| \\ &\leq \frac{(\theta_0/\theta_1)^{x^*}}{C(\theta_1)} + \sum_{x \leq x^*} q(x) \exp\left[-u\left\{\frac{\theta_0}{a} - 1 - \log\left(\frac{\theta_0}{a}\right)\right\}\right] \left\{\theta_0\left(1 + \frac{x}{u}\right)\right\}^x \\ &\quad + \sum_{x \leq x^*} q(x) \left(1 + \frac{a^2}{\theta_0^2}\right)^{-u/2} \left\{\theta_0\left(1 + \frac{x}{u}\right)\right\}^x \\ &\leq \frac{3}{nC(\theta_1)}, \end{aligned}$$

which implies by (20) that

$$(32) \quad |w| = \left| \int_0^{\theta_0} \cos\left(\frac{u(\theta-a)}{\theta_0} - \frac{m\pi}{2}\right) \sum_{x=0}^{\infty} q(x) \theta^x h_{u,v}(\theta) d\theta \right| \leq \frac{3}{nC(\theta_1)}.$$

Since $\sum_{x=0}^{\infty} q(x) |\int_0^{\theta_0} w \theta^x h_{u,v}(\theta) d\theta| = |w| \int_0^{\theta_0} h_{u,v}(\theta)/C(\theta) d\theta \leq |w|/C(\theta_0)$, we have, by (31),

$$\sum_{x=0}^{\infty} |f(x; g_{1n}) - f(x; g_{2n})| \leq \frac{w_0}{\sqrt{u}} \left(\frac{\theta_0}{u} \right)^{\alpha} \left(1 + \frac{1}{C(\theta_0)} \right) \frac{3}{nC(\theta_1)} = o\left(\frac{1}{n}\right).$$

Now, for $Z = (X_1, \dots, X_n)$ and $\lambda > 1$, we have

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

which implies by Lemma 1 that (24) holds with $p_n \rightarrow p_0 = \frac{1}{2}$ and $\varepsilon_n \rightarrow \varepsilon$ in (30). $\square$

PROOF OF LEMMA 2. Equation (25) follows from the Stirling formula. Inequalities (26) and (27) are based on the fact that

$$\int_0^{\infty} t^{\alpha} (1 + t^2)^{-u/2} dt = \frac{\Gamma((\alpha + 1)/2) \Gamma((u - \alpha - 1)/2)}{2\Gamma(u/2)}$$

for all $u > \alpha + 1 > 0$. Since $h_{u,v}$ is a gamma density, $h_{u,v}^*(t) = \int_0^{\infty} \exp(it\theta) h_{u,v}(\theta) d\theta = (1 - it/v)^{-u}$. By the Fourier inversion formula

$$h_{u,v}^{(j)}(\theta) = (2\pi)^{-1} \int_{-\infty}^{\infty} (-it)^j \exp(-it\theta) \left(1 - \frac{it}{v}\right)^{-u} dt.$$

It follows that

$$\begin{aligned} |h_{u,v}^{(j)}(\theta)| &\leq \frac{2}{2\pi} \int_0^{\infty} t^j \left(1 + \left(\frac{t}{v}\right)^2\right)^{-u/2} dt \\ &= \frac{v^{j+1}}{2\pi} \frac{\Gamma((j + 1)/2) \Gamma((u - j - 1)/2)}{\Gamma(u/2)} \end{aligned}$$

and

$$\begin{aligned} |h_{u,v}^{(\alpha')}(\theta) - h_{u,v}^{(\alpha')}(a)| &\leq \max_t \left\{ \frac{|e^{it} - 1|}{t^{\alpha''}} \right\} \frac{|\theta - a|^{\alpha''}}{2\pi} \int_{-\infty}^{\infty} |t|^{\alpha} \left(1 + \left(\frac{t}{v}\right)^2\right)^{-u/2} dt \\ &\leq 2|\theta - a|^{\alpha''} \frac{v^{\alpha+1}}{2\pi} \frac{\Gamma((\alpha + 1)/2) \Gamma((u - \alpha - 1)/2)}{\Gamma(u/2)}. \end{aligned}$$

To prove (28) we have

$$\begin{aligned}
 & \left| \int_0^\infty \cos\left(\frac{u(\theta - a)}{\theta_0} - \frac{m\pi}{2}\right) \theta^x h_{u,v}(\theta) d\theta \right| \\
 & \leq \left| \int_0^\infty \frac{\exp(iu\theta/\theta_0) v^u \theta^{u+x-1} \exp(-v\theta)}{\Gamma(u)} d\theta \right| \\
 & = \frac{v^u}{\Gamma(u)} \frac{\Gamma(u+x)}{|v - iu/\theta_0|^{u+x}} \leq \left(1 + \frac{a^2}{\theta_0^2}\right)^{-u/2} \left\{\theta_0\left(1 + \frac{x}{u}\right)\right\}^x.
 \end{aligned}$$

To prove (29) we have

$$\begin{aligned}
 \int_{\theta_0}^\infty \theta^x h_{u,v}(\theta) d\theta &= \int_{\theta_0}^\infty v^u \theta^{u+x-1} \frac{\exp(-v\theta)}{\Gamma(u)} d\theta \\
 &= \left(\frac{v\theta_0}{u}\right)^u \left(\frac{u}{\theta_0}\right)^{-x} \int_{\theta_0}^\infty \left(\frac{u}{\theta_0}\right)^{u+x} \\
 &\quad \times \theta^{u+x-1} \frac{\exp(-(u/a - u/\theta_0 + u/\theta_0)\theta)}{\Gamma(u)} d\theta \\
 &\leq \left(\frac{\theta_0}{a}\right)^u \left(\frac{u}{\theta_0}\right)^{-x} \frac{\exp(-(u/a - u/\theta_0)\theta_0) \Gamma(u+x)}{\Gamma(u)} \\
 &\leq \exp\left[-u\left(-\log\left(\frac{\theta_0}{a}\right) + \frac{\theta_0}{a} - 1\right)\right] \theta_0^x \left(1 + \frac{x}{u}\right)^x. \quad \square
 \end{aligned}$$

4. The case of infinite θ^* . The natural value of θ^* in (2) is $\theta_0^* = \sup\{\theta: \sum_x q(x)\theta^x < \infty\}$, which is always known. If θ_0^* is finite [e.g., (16) with $\beta = 0$], then we can set $\theta^* = \theta_0^*$ and use the results in Section 2. If $\theta_0^* = \infty$, then the condition $\theta^* < \infty$ becomes an assumption in addition to the knowledge of $q(\cdot)$. In this section we study the case $\theta^* = \infty$, where Theorem 1 does not apply. The kernel in (8) is used with $m_n = \infty$. An alternative bound for the variance of (8) is provided with a square-root reparametrization. If (16) holds for $0 \leq \beta < 2$ and all nonnegative integers x , then our kernel estimators for $g^{(m)}$ have the rate $(\log n)^{-(1-\beta/2)(\alpha-m)}$ of convergence over the class $\mathcal{G}_{\alpha,\infty}$ in (3).

Let $Y = 2X$ and $\eta = \sqrt{\theta}$. The conditional density $f(y | \eta) = C(\eta^2)q(y/2)\eta^y$ is of the form (2). If $q(y/2) > 0$ for all nonnegative even integers y , then we can estimate the density g_η of η via the kernel in (8). The effect of this reparametrization is essentially halving the value of β under (16), which dictates the rate of our alternative bound of variance in Theorem 4 below. For example, the upper bounds are not useful for the Poisson family without this reparametrization. Since $g_\eta(\sqrt{a}) = 2\sqrt{a}g(a)$, (8) provides a kernel (with $j = 0$

and $m_n = \infty$ for $g_\eta(\sqrt{a})C(a) = 2\sqrt{a}C(a)g(a)$,

$$(33) \quad \begin{aligned} K_n^{(0)}(y, \sqrt{a}) &= \bar{K}_n(x, a) \\ &= \{\pi q(x)(2x)!\}^{-1}(-1)^x \int \cos(t\sqrt{a})t^{2x}k^*\left(\frac{t}{c_n}\right)dt \end{aligned}$$

with expectation $E\bar{K}_n(x, a) = \int_{-\infty}^{\infty} ck(c(\sqrt{a} - \theta/\sqrt{|\theta|}))h(\theta)d\theta$ and estimators

$$(34) \quad \hat{g}_n(a) = n^{-1} \sum_{i=1}^n \frac{\bar{K}_n(X_i, a)}{2\sqrt{a}C(a)}, \quad \hat{g}_n^{(m)}(a) = \left(\frac{d}{d\theta}\right)^m \hat{g}_n(\theta) \Big|_{\theta=a}.$$

Since $\bar{K}_n(x, a)$ utilizes all observed frequencies at nonnegative integers, a truncated version (up to $x \leq m_n$) should probably be used even when $\theta^* < \infty$ [provided $q(x) > 0$, $x = 0, 1, 2, \dots$], although this does not improve the rate in Section 2.

THEOREM 4. *Let the kernel $\bar{K}_n(x, a)$ be given by (33). Suppose*

$$(35) \quad 1 \leq q(x)B_0B^{2x}((2x)!)^{\beta/2}, \quad x = 0, 1, \dots,$$

for some $0 \leq \beta < 2$. Choose $0 < \beta_0 < \frac{1}{2}$ and

$$(36) \quad c_n = B^{-1} \left(\frac{\beta_0 \log n}{1 - \beta/2} \right)^{1 - \beta/2}, \quad n \geq 1.$$

Then

$$\sqrt{E_g |\bar{K}_n(x, a)|^2} \leq \sup_{x, a} |\bar{K}_n(x, a)| \leq B_3 n^{\beta_0},$$

$$\sqrt{E_g |\bar{K}_n^{(j)}(x, a)|^2} \leq \sup_{x, a} |\bar{K}_n^{(j)}(x, a)| \leq (1 + o(1)) \left(\frac{c_n}{2\sqrt{a}} \right)^j B_3 n^{\beta_0},$$

where $B_3 = 2B_0\|k\|_1/(\pi B)$.

REMARK. The value of β in (35) is the same as in (16), although B_0 and B could be different.

PROOF OF THEOREM 4. By (33), (35) and (36),

$$\begin{aligned} |\bar{K}_n(x, a)| &\leq \{\pi q(x)(2x)!\}^{-1} 2\|k(y)\|_1 \frac{c_n^{2x+1}}{2x+1} \\ &\leq \frac{2\|k(y)\|_1 B_0/B}{\pi q(x)B_0B^{2x}((2x)!)^{\beta/2}} \frac{(Bc_n)^{2x+1}}{((2x+1)!)^{1-\beta/2}} \\ &\leq B_3 n^{\beta_0}, \end{aligned}$$

since $c^x/(x!)^{1-\beta/2} < \exp[(1 - \beta/2)c^{1/(1-\beta/2)}]$. The rest follows from $|\bar{K}_n^{(j)}(x, a)| \leq c_{n,j} B_3 n^{\beta_0}$ with

$$c_{n,j} = \max_{|t| \leq c_n} \left| \left(\frac{d}{da} \right)^j \cos(t\sqrt{a}) \right| \leq (1 + o(1)) \left(\frac{c_n}{2\sqrt{a}} \right)^j \quad \text{as } c_n \rightarrow \infty. \quad \square$$

COROLLARY 2. Let $\hat{g}_n^{(m)}(a)$ be given by (34) and let $a > 0$. Under the conditions of Theorem 4,

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

5. Estimating a mixing distribution. A function $G(\cdot)$ is the mixing distribution for the population X if

$$(37) \quad f(x) = f(x; G) = \int f(x | \theta) dG(\theta),$$

where $f(x | \theta)$ is given by (2). If $dG/d\theta = g$ exists, then (37) is the same as (1). Since

$$\int_{-\infty}^{\infty} G(a - \theta/c) k(\theta) d\theta \rightarrow \{G(a+) + G(a-)\}/2$$

for $k(\cdot)$ satisfying (6) and (7), we shall take the version $G(\theta) = \{G(\theta+) + G(\theta-)\}/2$ of G .

We shall assume $\theta^* < \infty$ and consider the kernel estimator

$$(38) \quad \hat{G}_n(a) = \int_0^a \hat{g}_n(\theta) d\theta,$$

where $\hat{g}_n = \hat{g}_n^{(0)}$ is given by (15). The performance of estimators will be measured over the class

$$\begin{aligned} \mathcal{G}_{\alpha,\theta^*}^{cdf} &= \mathcal{G}_{\alpha,\theta^*}^{cdf}(a, \delta, M, M_0, M_1) \\ &= \left\{ G: |G_c^{(\alpha'+1)}(\theta) - G_c^{(\alpha'+1)}(a)| \leq M|\theta - a|^{\alpha''}, \forall 0 \leq |\theta - a| \leq \delta, \right. \\ (39) \quad &\sum_{j=0}^{\alpha'+1} |G_c^{(j)}(a)| \leq M_0, G_d([a - \delta, a] \cup (a, a + \delta]) = 0, \\ &\left. G(\theta) \leq M_1\theta^\alpha, \forall 0 \leq \theta \leq \delta, G(\theta^*) = 1 \right\}, \end{aligned}$$

where G_c and G_d are the continuous and discrete components of G ; α' is the greatest integer less than α ; and $\alpha'' = \alpha - \alpha' \in (0, 1]$. Note that, for a given discrete distribution function G with finite number of atoms in $(0, \theta^*)$, there exists a $\delta > 0$ such that $G \in \mathcal{G}_{\alpha,\theta^*}^{cdf}(a, \delta, M, M_0, M_1)$ for all $\alpha > -1$ and positive numbers M, M_0 and M_1 .

THEOREM 5. Let $\hat{G}_n(a)$ be given by (38) and let $0 \leq a < \theta^*$. Suppose the conditions of Theorem 2 hold and (6) is satisfied with $\alpha_0 \geq \alpha + 1$. Then, for $\alpha > -1$,

$$\sup\{E_G|\hat{G}_n(a) - G(a)|^2: G \in \mathcal{G}_{\alpha,\theta^*}^{cdf}\} = O(c_n^{-2(\alpha+1)}).$$

COROLLARY 3. Suppose (16) and the conditions of Theorem 5 hold. Then, for $0 \leq \alpha < \theta^*$,

$$\sup\{E_G|\hat{G}_n(a) - G(a)|^2 : G \in \mathcal{G}_{\alpha, \theta^*}^{cdf}\} = \begin{cases} O(1)(\log n)^{-2(\alpha+1)}, & \text{if } \beta = 0, \\ O(1)\left(\frac{\log n}{\log \log n}\right)^{-2(\alpha+1)}, & \text{if } 0 < \beta < \infty. \end{cases}$$

PROOF OF THEOREM 5. The argument is the same as the proof of Theorem 1 for the variance and the “second” term of the bias except for the use of upper bound

$$\left| \int_0^a \frac{\cos(t\theta)}{C(\theta)} d\theta \right| \leq \begin{cases} |t|^{-1} \left| \int_0^a (C(\theta))^{-1} d \sin(t\theta) \right| \leq \frac{2|t|^{-1}}{C(a)}, & x > 0, \\ \frac{a}{C(a)}, & x = 0. \end{cases}$$

Define $H(\theta) = \int_0^\theta C(y) dG(y)$. Then, the “first” term of the bias can be written as

$$\begin{aligned} B_{1n}(a) &= \int \left[\int_0^a \left(\frac{1}{C(y)} \right) H \left(dy - \frac{\theta}{c_n} \right) - G(a) \right] k(\theta) d\theta \\ &= \int \left[\int \frac{C(y)}{C(y + \theta/c_n)} I \left\{ 0 \leq y + \frac{\theta}{c_n} \leq a \right\} dG(y) - G(a) \right] k(\theta) d\theta. \end{aligned}$$

Since $C(y)$ is an analytic function and $C(y + \theta/c_n)$ is bounded away from zero for $0 \leq y + \theta/c_n \leq a$, we can take the Taylor expansion

$$\frac{C(y)}{C(y + \theta/c_n)} = \sum_{j=0}^{\alpha'+1} d_j(y) \left(\frac{\theta}{c_n} \right)^j + R_y \left(\frac{\theta}{c_n} \right),$$

where $d_j(y)$, $d_0(y) = 1$, are analytic functions on $[0, a + \delta_0]$; $|R_y(\theta/c)| \leq M^*|\theta/c|^{\alpha'+1}$ for $0 \leq y + \theta/c_n \leq a$ and $|\theta/c_n| \leq \delta_0$; and δ_0 and M^* are positive numbers. Hence, by (6),

$$\begin{aligned} B_{1n}(a) &\leq O(c_n^{-\alpha-1}) + \int_{|\theta/c_n| \leq \delta_0} \sum_{j=0}^{\alpha'+1} \left[\int_{-\theta/c_n}^{a-\theta/c_n} d_j(y) dG(y) - \int_0^a d_j(y) dG(y) \right] \\ &\quad \times \left(\frac{\theta}{c_n} \right)^j k(\theta) d\theta \\ &= O(c_n^{-\alpha-1}), \end{aligned}$$

where the $O(1)$ is uniform in the class $G \in \mathcal{G}_{\alpha, \theta^*}^{cdf}$. $\square$

Lower bounds for the optimal rate of convergence are given below.

THEOREM 6. For $0 \leq a < \theta^*$,

$$\liminf_{n \rightarrow \infty} (\log n)^{2(\alpha+1)} \inf_{\tilde{G}_n} \sup \{E_G |\bar{G}_n(a) - G(a)|^2 : G \in \mathcal{G}_{\alpha, \theta^*}^{cdf}\} > 0.$$

Again, the optimal rate of convergence is achieved for the negative binomial family and is almost achieved for the Poisson family.

PROOF OF THEOREM 6. We shall use the same pairs (23) as in Section 3 with $m = -1$ and $u_n/\log n = \delta_0$ in (22) replaced by $u_n/\log(u_n n) = \delta_0$. Our statement holds, since by (32), (25) and (26) we have, via integration by parts,

$$\begin{aligned} & G_{2n}(a) - G_{2n}(a) \\ &= w_0 \theta_0^\alpha u^{-\alpha-1/2} \int_0^a \left[\cos\left(u \frac{\theta-a}{\theta_0} + \frac{\pi}{2}\right) - w \right] g_{u,v}(\theta) d\theta \\ &= w_0 \theta_0^{\alpha+1} u^{-\alpha-1/2-1} \int_0^a g_{u,v}(\theta) d \sin\left(u \frac{\theta-a}{\theta_0} + \frac{\pi}{2}\right) + O\left(\frac{u^{-\alpha-1/2}}{n}\right) \\ &= -w_0 \theta_0^{\alpha+1} u^{-\alpha-1/2-1} \left\{ g_{u,v}(a) + \int_0^a \cos\left(u \frac{\theta-a}{\theta_0}\right) dg_{u,v}(\theta) \right\} \\ &\quad + O\left(\frac{u^{-\alpha-1/2}}{n}\right) \\ &= -\frac{(w_0 \theta_0^{\alpha+1} + o(1)) u^{-\alpha-1}}{aC(a)\sqrt{2\pi}}. \end{aligned}$$

□

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.
- DEVROYE, L. P. (1989). Consistent deconvolution in density estimation. *Canad. J. Statist.* **17** 235–239.
- EDELMAN, D. (1988). Estimation of the mixing distribution for a normal mean with applications to the compound decision problem. *Ann. Statist.* **16** 1609–1622.
- FAN, J. (1991a). On the optimal rates of convergence for nonparametric deconvolution problems. *Ann. Math. Statist.* **19** 1257–1272.
- FAN, J. (1991b). Global behavior of deconvolution kernel estimates. *Statist. Sinica* **1** 541–551.
- FELLER, W. (1971). *An Introduction to Probability Theory and Its Applications, II*. Wiley, New York.
- 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.
- LAMBERT, D. and TIERNEY, L. (1984). Asymptotic properties of maximum likelihood estimates in the mixed Poisson model. *Ann. Statist.* **12** 1388–1399.
- 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, part II: the exponential family. *Ann. Statist.* **11** 783–792.
- PRAKASA RAO, B. L. S. (1983). *Nonparametric Functional Estimation*. Academic Press, New York.
- 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. (1956). 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.
- SIMAR, L. (1976). Maximum likelihood estimation of a compound Poisson process. *Ann. Statist.* **4** 1200–1209.
- STEFANSKI, L. A. (1990). Rates of convergence of some estimators in a class of deconvolution problems. *Statist. Probab. Lett.* **9** 229–235.
- TEICHER, H. (1961). Identifiability of mixtures. *Ann. Math. Statist.* **32** 244–248.
- TUCKER, H. G. (1963). An estimate of the compounding distribution of a compound Poisson distribution. *Theory Probab. Appl.* **8** 195–200.
- 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.
- ZHANG, C.-H. (1990). Fourier methods for estimating mixing densities and distributions. *Ann. Statist.* **18** 806–831.

DEPARTMENT OF STATISTICS
HILL CENTER, BUSCH CAMPUS
RUTGERS UNIVERSITY
NEW BRUNSWICK, NEW JERSEY 08903