ON ESTIMATING MIXING DENSITIES IN DISCRETE
EXPONENTIAL FAMILY MODELS

BY CUN-HUI ZHANG

Rutgers University

This paper concerns estimating a mixing density function $g$ and its
derivatives based on iid observations from $f(x) = f(x \mid \theta)g(\theta)\,d\theta$, where
$f(x \mid \theta)$ 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
$f(x \mid \theta_0) \geq \epsilon^{x+1} \forall x$, for some positive numbers $\theta_0$ and $\epsilon$, then our esti-
mators achieve the optimal rate of convergence $(\log n)^{-\alpha - m}$ for estimat-
ing the $m$th derivative of $g$ under a Lipschitz condition of order $\alpha > m$. The
optimal rate of convergence is almost achieved when $(x!)^\alpha f(x \mid \theta_0) \geq \epsilon^{x+1}$. Estimation of the mixing distribution function is also considered.

1. Introduction. Let $f(x \mid \theta)$ be a known parametric family of probabil-
ity density functions with respect to a $\sigma$-finite measure $\mu$. The density
function $f(x)$ of a random variable $X$ belongs to a mixture model if

$$f(x) = f(x \mid \theta) = \int f(x \mid \theta)g(\theta)\,d\theta. \quad (1)$$

We are interested in estimating the mixing density function $g$ and its
derivatives based on independent identically distributed observations
$X_1, \ldots, X_n$ from $f(x)$. When $f(x \mid \theta)$ is a location family and $\mu$ 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

$$f(x \mid \theta) = C(\theta)q(x)\theta^x, \quad x = 0, 1, 2, \ldots, 0 \leq \theta \leq (or <) \theta^* \leq \infty, \quad (2)$$

and $\mu$ 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 $\theta^*$

---

Received June 1992; revised July 1994.


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 \( \alpha, \alpha < \theta^* \), \( \delta, M \) and \( M_0 \), define

\[
\mathcal{G}_{\alpha, \theta^*} = \mathcal{G}_{\alpha, \theta^*}(a, \delta, M, M_0)
\]

\[
= \left\{ g : |g^{(\alpha)}(\theta) - g^{(\alpha)}(a)| \leq M|\theta - a|^\alpha, \forall |a - \theta| \leq \delta, \sum_{j=0}^{\alpha'} |g^{(j)}(a)| \leq M_0, \int_0^{\theta^*} g(\theta) d\theta = 1 \right\}.
\]

(3)

Here and in the sequel, \( \alpha' \) is always the largest integer less than \( \alpha \) and \( \alpha'' = \alpha - \alpha' \in (0, 1] \). If there exist nonnegative numbers \( \beta, B_0 \) and \( B \) such that \( q(x)B_0B^2(x!)^\beta \geq 1, \forall x \), 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 \( \theta^* \) 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 \( q(x) > 0, \)
\( (5) \quad \frac{P_{\theta}(X = x)}{q(x)} = \frac{f(x; g)}{q(x)} = \int_{0}^{\theta^*} \theta^* C(\theta) g(\theta) d\theta = \int_{0}^{\theta^*} \theta^* h(\theta) d\theta, \)
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 \( K^{(j)}_n(x, a) \) such that \( E K^{(j)}_n(X, a) = \int_{-\infty}^{\infty} h^{(j)}(a - \theta/c_n)k(\theta) d\theta + o(1) \) for some suitable \( c_n \to \infty \) and \( k(\cdot) \). Let \( k(x) \) be a function satisfying
\( (6) \quad \int x^j k(x) dx = I[j = 0], \quad \|x^{\alpha_0 + 1}h^{(j)}(x)\|_\infty < \infty, \quad 0 \leq j < \alpha_0, \)
\( \|x^{\alpha_0}k(x)\|_1 < \infty, \)
for a positive number \( \alpha_0 \), and
\( (7) \quad k(x) = k(-x) \quad \forall \ x, \quad k^*(t) = 0 \quad \forall \ |t| > 1. \)

Note that \( \|x^{\alpha_0 + 1}h^{(j)}(x)\|_\infty < \infty \) 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 \( c > 0, \)
\[ \int_{-\infty}^{\infty} h\left(a - \frac{\theta}{c}\right)k(\theta) d\theta = \int_{0}^{\infty} \left\{ ck(c(a - \theta)) + ck(c(a + \theta)) \right\} h(\theta) d\theta. \]

By (7), \( k^*(t) \) is a real even function with a compact support, so that, by the Fourier inversion formula,
\[ ck(c(a - \theta)) + ck(c(a + \theta)) \]
\[ = \left(\frac{c}{2\pi}\right) \int \left\{ \exp[-itc(a - \theta)] + \exp[-itc(a + \theta)] \right\} k^*(t) dt \]
\[ = \frac{\int \cos(t\theta) \cos(ta) k^*(t/c) dt}{\pi}. \]
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
\[ \int_{-\infty}^{\infty} h\left(a - \frac{\theta}{c}\right)k(\theta) d\theta = \sum_{x=0}^{\infty} \frac{f(x; g)}{q(x) x!} \int t^x \cos\left(\frac{x\pi}{2}\right) \cos\left(\frac{ta}{2}\right) k^*\left(\frac{t}{c}\right) dt. \]

Differentiating with respect to \( a \) under the integrations, we obtain, for integers \( j \geq 0, \)
\[ \int_{-\infty}^{\infty} h^{(j)}\left(a - \frac{\theta}{c}\right) k(\theta) d\theta \]
\[ = \sum_{x=0}^{\infty} \frac{f(x; g)}{\pi q(x) x!} \int \cos\left(\frac{ta - j\pi}{2}\right) t^{x+j} \cos\left(\frac{x\pi}{2}\right) k^*\left(\frac{t}{c}\right) dt. \]
This motivates the kernel for $h^{(j)}(a)$,

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

(8)

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 $m_n = \infty$, $E_{\theta} K_n^{(j)}(X, a) = \int_0^\infty h^{(j)}(a - \theta/c_n) k(\theta) d\theta \rightarrow h^{(j)}(a)$ 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 $\theta^*$.

The bias of our kernel (8) is

$$
E K_n^{(j)}(X, a) - h^{(j)}(a) = b_1^{(j)}(a) + b_2^{(j)}(a),
$$

(9)

where $b_1^{(j)}(a) = \int h^{(j)}(a - \theta/c_n) - h^{(j)}(a) k(\theta) d\theta$ and

$$
b_2^{(j)}(a) = \sum \{ x! \}^{-1} \int_0^{\theta^*} \theta^* h(\theta) d\theta \int \cos\left(\frac{ta - j\pi}{2}\right) t^{x+j} \cos\left(\frac{x\pi}{2}\right) k^*\left(\frac{t}{c_n}\right) dt.
$$

The first term $b_1^{(j)}(a)$ 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 $b_2^{(j)}(a)$ 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 $\theta^*$ in (2) is finite. Let $\alpha \geq 0$ and $0 < \beta_0 < \frac{1}{2}$, and choose $m_n$ and $c_n$ such that

$$
\max\left\{ \log\left(\frac{1}{q(2x)}\right), x = 0, 1, \ldots, \left\lfloor \frac{m_n}{2} \right\rfloor \right\} + c_n \leq \beta_0 \log n,
$$

(10)

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

(11)

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

$$
\sqrt{E|K_n^{(j)}(X, a)|^2} \leq \sup_{x, a} |K_n^{(j)}(x, a)| \leq B_1 c_n^{1/2} n^{\beta_0}, \quad |b_2^{(j)}(a)| \leq B_2 c_n^{-a+j},
$$

(12)

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 $f(x \mid \theta) = C(\theta) q(x) \theta^x > 0$ on the set of even integers $0 \leq x \leq m_n$ (possibly after reparametrization and data transformation), even when $P(X = \text{integers} \mid \theta) < 1$. If $q(x) > 0$ for all nonnegative integers $x$, we may consider the family $f(y \mid \eta) = C(\eta^2) q(y/2) \eta^y$ with $Y = 2X$ and $\eta = \sqrt{\theta}$, so that the kernel will utilize the observed fre-
quences 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 \( \theta^* c_n(y) + \alpha \log c_n(y) \) 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 \( x \leq m_n \),
\[
|K_n^{(j)}(x, a)| \leq \left\{ \pi q(x) x! \right\}^{-1} 2\|k\|_1 c_n^{x+j+1} \frac{c_n^{x+j+1}}{x+j+1}
\leq \left( 2\frac{\|k\|_1}{\pi} \right) c_n^j \exp\left[ -\log q(x) + c_n \right]
\leq B_1 c_n^j n^{\beta_0}.
\]
For the bias \( b_{2n}(\alpha) \) in (9) we have
\[
|b_{2n}^{(j)}(\alpha)| \leq \sum_{x > m_n} (\pi x!)^{-1} (\theta^*)^x C(0) c_n^{x+j+1} \frac{2\|k\|_1}{x+j+1}
\leq C(0) 2\|k\|_1 (\pi \theta^*)^{-1} c_n^j \sum_{x > m_n} \frac{(\theta^* c_n)^{x+1}}{(x+1)!}.
\]
The proof is complete, since
\[
\sum_{x > m_n} \frac{(\theta^* c_n)^{x+1}}{(x+1)!} \leq \left( \frac{\theta^* c_n}{m_n + 1} \right)^{m_n+1} \sum_{x > m_n} \frac{(m_n + 1)^{x+1}}{(x+1)!}
\leq \left( \frac{e^{\theta^* c_n}}{m_n + 1} \right)^{m_n+1} \leq c_n^{-\alpha}. \quad \square
\]

Consider the estimation of \( g^{(m)}(\alpha) \). By (4) and (2), \( g(\theta) = h(\theta) \sum_{x=0}^{\infty} q(x) \theta^x \), so that
\[
(14) \quad g^{(m)}(\alpha) = \sum_{j=0}^{m} C_{j,m}(\alpha) h_{\alpha}(j), \quad C_{j,m}(\alpha) = \binom{m}{j} \sum_{x=m-j}^{\infty} q(x) x! \theta^{x-m+j}.
\]
Since \( C_{j,m}(\alpha) \) are known constants, we can estimate \( g^{(m)}(\alpha) \) by
\[
(15) \quad \hat{g}^{(m)}(\alpha) = \sum_{j=0}^{m} C_{j,m}(\alpha) \hat{h}_{\alpha}(j), \quad \hat{h}_{\alpha}(j) = n^{-1} \sum_{i=0}^{n} K_{\alpha}(X_i, \alpha).
\]

**Theorem 2.** Let \( \mathcal{G}_{a, \theta^*} \) be given by (3), and \( \hat{g}^{(m)}(\alpha) \) by (15) with the kernel \( K_{\alpha}(x, a) \) in (8) satisfying (6) and (7), \( a \geq \alpha \). Choose \( m_n \) and \( c_n \) by (13).

Then, for \( 0 \leq m < \alpha \) and \( 0 < \alpha < \theta^* < \infty \),
\[
\sup \left\{ E_\Theta \left[ \hat{g}^{(m)}(\alpha) - g^{(m)}(\alpha) \right]^2 : g \in \mathcal{G}_{a, \theta^*} \right\} = O(c_n^{-2(\alpha-m)}).
\]
REMARK. By (13), the rate of convergence $O(c_n^{-\alpha + m})$ is determined by the rate at which $q(x) \to 0$. The faster $q(x) \to 0$, the slower $E_g |\hat{g}(m)(a) - g(m)(a)|^2 \to 0$. 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 $\hat{g}(m)$ depends on $\alpha$ only through $c_n$ in (13). Since $\delta > 0$ can be arbitrarily small with $\mathcal{F}_{a, \theta*}(a, \delta, M, M_0)$, 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 $\sum_{j=0}^a |g^{(j)}(a)| \leq M_0 = 2Ma^\delta e^\alpha$.

PROOF OF THEOREM 2. Due to the smoothness of $C(\cdot)$, $g \to h$ maps $\mathcal{F}_{a, \theta*}(a, \delta, M, M_0)$ into $\mathcal{F}_{a, \theta*}(a, \delta, M', M'_0)$ for some (large) $M'$ and $M'_0$. Since the kernel is bounded for each $n$ by Theorem 1, $E_g |\hat{g}(m)(a)|^j$, $j = 1, 2$, are continuous under the convergence as functionals of $G(\theta) = \int_0^\theta g(y) dy$, so that the supremum can be taken over all $g \in \mathcal{F}_{a, \theta*}$ such that (4) is $a'$ times continuously differentiable on $(-\infty, \infty)$. The bias term $b_{1n}(a) = \int (h^{(j)}(a - \theta/c_n) - h^{(j)}(a)) k(\theta) d\theta$ 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 $h^{(j)}(a - \theta/c)$ 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 $h^{(j)}(a)$, $0 \leq j \leq a'$. It follows from (6) and standard methods in density estimation that

$$b_{1n}(a) = O(c_n^{-\alpha + j}) + (-1)^j c_n \int_{|\theta| > \delta c_n} h^{(j)}(\theta) d\theta + \int_{|\theta| \leq \delta c_n}$$

$$b_{1n}(a) = O(c_n^{-\alpha + j}),$$

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

$$E_g |\hat{g}(m)(a) - g(m)(a)|^2 \leq O(B_n^2 c_n^{2m n^{2\alpha_0 - 1}}) + O(c_n^{-2(\alpha - m)})$$

$$= O(c_n^{-2(\alpha - m)}).$$

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

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

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

$$\sup \{E_g |\hat{g}(m)(a) - g(m)(a)|^2 : g \in \mathcal{F}_{a, \theta*} \}$$

$$= \begin{cases} O(1)(\log n)^{-2(\alpha - m)}, & \text{if } \beta = 0, \\ O(1) \left( \frac{\log n}{\log \log n} \right)^{-2(\alpha - m)}, & \text{if } 0 < \beta < \infty. \end{cases}$$

The value of $\beta$ 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 \((\log n)^{-1}(\beta > 0)\log n\) by (13). \(\Box\)

**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

\[
(17) \quad r_{n, \alpha, \theta^*} = \inf_{\hat{g}_{n, \alpha}} \sup_{g \in \mathcal{G}_{\alpha, \theta^*}} \left\{ \mathbb{E}_g | \hat{g}_{n, \alpha}^{(m)}(a) - g^{(m)}(a)|^2 : g \in \mathcal{G}_{\alpha, \theta^*} \right\},
\]

where \(\mathcal{G}_{\alpha, \theta^*}\) is given by (3), and the infimum runs over all statistics \(\hat{g}_{n, \alpha}\) based on \(X_1, \ldots, X_n\). In this section we derive lower bounds for the rate at which \(r_{n, \alpha, \theta^*}\) 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 \(\hat{g}_{1n}^{(m)}(a) - \hat{g}_{2n}^{(m)}(a)\) 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 \(r_{n, \alpha, \theta^*}\) be given by (3) and (17), respectively, \(0 < \alpha < \theta^*\). Then there exists \(\varepsilon_0 > 0\) such that

\[
\liminf_{n \to \infty} \inf_{\hat{g}_{n, \alpha}^{(m)}} \sup_{g \in \mathcal{G}_{\alpha, \theta^*}} P_g \left\{ | \hat{g}_{n, \alpha}^{(m)}(a) - g^{(m)}(a) | > \varepsilon_0 (\log n)^{-\alpha + m} \right\} > 0.
\]

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

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 \(a, \theta_0\) and \(\theta_1\) be fixed constants satisfying \(0 < a < \theta_0 < \theta_1 < \theta^*\). Set

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

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

\[
(20) \quad w = w(u, v) = \int_0^{\theta_1} \cos \left( \frac{u(\theta - a)}{\theta_0} - \frac{m\pi}{2} \right) g_{u, v}(\theta) \, d\theta.
\]

Let \(\varepsilon_0 > 0\), and let \(g_0\) be a density in \(\mathcal{G}_{\alpha, \theta^*}(\alpha, \delta, M - \varepsilon_0, M_0 - \varepsilon_0)\) 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 \( v_n = u_n/\alpha \) such that
\[
\frac{u_n}{\log n} = \delta_0 \geq \max \left\{ \frac{\theta_0/(\theta_1 - \theta_0)}{\log(\theta_1/\theta_0)}, \frac{1}{\theta_0/\alpha - 1 - \log(\theta_0/\alpha)}, \frac{2}{\log(1 + \alpha^2/\theta_0^2)} \right\}.
\]

Define \( g_{1n} = g_0 \) and
\[
g_{2n}(\theta) = g_0(\theta) + \frac{w_0}{\sqrt{u_n}} \left( \frac{\theta_0}{u_n} \right)^\alpha \left[ \cos \left( \frac{u_n}{\theta_0} \left( \frac{\theta - a}{\theta_0} - \frac{m\pi}{2} \right) \right) - w(u_n, v_n) \right] \times g_{u_n, v_n}(\theta) I(0 \leq \theta \leq \theta_0).
\]

We shall show in the proof of Theorem 3 that \( g_{1n} \) and \( g_{2n} \) are members of \( \mathcal{H}_{\alpha, \theta} \), for small \( w_0 \), and
\[
p_n = \inf_{\bar{g}_{n}^{(m)}} \max \left\{ P_{\bar{g}} \left( \left| \bar{g}_{n}^{(m)}(a) - g_{n}^{(m)}(a) \right| > \varepsilon_n (\log n)^{-\alpha + m} \right) \right\};
\]
\[
\bar{g} = g_{1n} \text{ or } g_{2n} \rightarrow p_0,
\]
for some \( \varepsilon_n \to \varepsilon > 0 \) 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 \( X_1, \ldots, X_n \) under \( P_j \in \mathcal{P} \), \( j = 1, 2 \). If \( \rho \leq P_1 f_1 \leq \lambda f_2 \), then, for any estimator \( T_n \) based on \( X_1, \ldots, X_n \),
\[
\max_{j=1,2} P_j \left( d(T_n, T(P_j)) \geq \frac{d(T(P_1), T(P_2))}{2} \right) \geq \frac{\rho}{1 + \lambda}.
\]

**Lemma 2.** Let \( h_{u,v} \) be given by (18). Then, for \( j \geq 0 \), \( \alpha > 0 \) and \( x \geq 0 \), the following hold:
\[
h_{u,v}(\theta) = \theta^{-1} \frac{\sqrt{u}}{2\pi} \exp \left[ -u \left( \frac{\theta}{\alpha} - 1 - \log \left( \frac{\theta}{\alpha} \right) \right) - \varepsilon_u \right],
\]
\[
\frac{1}{12u + 1} \leq \varepsilon_u \leq \frac{1}{12u};
\]
\[
|h_{u,v}(\theta)| \leq \frac{u^{j+1}}{2\pi} \frac{\Gamma((j+1)/2)\Gamma((u-j-1)/2)}{\Gamma(u/2)} = O(u^{(j+1)/2});
\]
\[ |h_{u,v}^{(\alpha')} (\theta) - h_{u,v}^{(\alpha')} (a)| \]
\[ \leq |\theta - a|^{\alpha} \frac{\sqrt{\alpha + 1}}{\pi} \frac{\Gamma((\alpha + 1)/2) \Gamma((u - \alpha - 1)/2)}{\Gamma(u/2)} \]
\[ = O(|\theta - a|^{\alpha} u^{(\alpha + 1)/2}); \]
\[ \left| \int_0^\infty \cos \left( \frac{u(\theta - a)}{\theta_0} - \frac{m\pi}{2} \right) \theta^x h_{u,v}(\theta) \, d\theta \right| \]
\[ \leq \left( 1 + \frac{a^2}{\theta_0^2} \right)^{-u/2} \left\{ \theta_0 \left( 1 + \frac{u}{x} \right) \right\}^x; \]
and
\[ \int_0^x \theta^x h_{u,v}(\theta) \, d\theta \leq \exp \left[ -u \left( \frac{\theta_0}{a} - 1 - \log \left( \frac{\theta_0}{a} \right) \right) \right] \left( \theta_0 \left( 1 + \frac{x}{u} \right) \right)^x. \]

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
\[ g_0(\theta) \leq \frac{w_0(1 + 1/C(\theta_0)) \left( \frac{\theta_0}{u} \right)^{\alpha} \exp \left[ -u \left( \frac{\theta}{a} - 1 - \log \left( \frac{\theta}{a} \right) \right) \right]}{C(\theta)\sqrt{2\pi}} \]
\[ \geq \frac{w_0(1 + |w|) \left( \frac{\theta_0}{u} \right)^{\alpha} \exp \left[ -u \left( \frac{\theta}{a} - 1 - \log \left( \frac{\theta}{a} \right) \right) \right]}{C(\theta)\sqrt{2\pi}} \]
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 \( 0 \leq \theta \leq \theta_0 \),
\[ g_{2n}^{(m)}(\theta) - g_0^{(m)}(\theta) \]
\[ = g_{2n}^{(m)}(\theta) - g_{1n}^{(m)}(\theta) \]
\[ = \frac{w_0}{\sqrt{u}} \left( \frac{\theta_0}{u} \right)^{\alpha} \left[ -w(u,v)g_{u,v}^{(m)}(\theta) \right. \]
\[ + \sum_{j=0}^m \binom{m}{j} \left( \frac{u}{\theta_0} \right)^{m-j} \cos \left( \frac{u}{\theta_0} - \frac{j\pi}{2} \right) g_{u,v}^{(j)}(\theta) \]
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
\[ |g_{2n}^{(\alpha')} (\theta) - g_{2n}^{(\alpha')} (a)| \leq |\theta - a|^{\alpha'} (M - \varepsilon_0 + w_0 O(1)), \quad |\theta - a| \leq \delta, \]
\[ \sum_{j=0}^{\alpha'} |g_{2n}^{(j)}(a)| \leq M_0 - \varepsilon_0 + w_0 O(1), \]
where the $O(1)$ is uniform in $\theta$ and does not depend on $w_0$. Therefore, $g_{2n} \in \mathscr{G}_{\theta}(\alpha, \delta, M, M_0)$ for small $w_0$. Also, since $u/\log n = \delta_0$ by (22), it follows from (23), (25) and (26) that

$$
\varepsilon_n = (\log n)^{\alpha - m} \frac{|g_{2n}(a) - g_{1n}(a)|}{2} \to \varepsilon
$$

$$
= \frac{w_0}{2aC(a)\sqrt{2\pi}} \left( \frac{\theta_0}{\delta_0} \right)^{\alpha - m} > 0.
$$

**Step 2.** Verify (24) for the $\varepsilon_n$ in (30). We shall first find a good bound for

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

$$
= \frac{w_0}{\sqrt{u}} \left( \frac{\theta_0}{u} \right)^{\alpha} \sum_{x=0}^{\infty} q(x) \left| \int_0^{\theta_0} \left[ \cos \left( \frac{u(\theta - a)}{\theta_0} - \frac{m\pi}{2} \right) - u \right] \theta^x u_v(\theta) \, d\theta \right|.
$$

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),

$$
\sum_{x=0}^{\infty} q(x) \left| \int_0^{\theta_0} \cos \left( \frac{u(\theta - a)}{\theta_0} - \frac{m\pi}{2} \right) \theta^x u_v(\theta) \, d\theta \right|
$$

$$
\leq \sum_{x > x^*} q(x) \theta_0^x + \sum_{x \leq x^*} q(x) \int_{\theta_0}^{\infty} \theta^x u_v(\theta) \, d\theta
$$

$$
+ \sum_{x \leq x^*} q(x) \int_0^{\theta_0} \cos \left( \frac{u(\theta - a)}{\theta_0} - \frac{m\pi}{2} \right) \theta^x u_v(\theta) \, d\theta
$$

$$
\leq \frac{(\theta_0/\theta_1)^{x^*}}{C(\theta_1)} + \sum_{x \leq x^*} q(x) \exp\left[-u\left(\frac{\theta_0}{a} - 1 - \log\left(\frac{\theta_0}{a}\right)\right)\right] \left(\theta_0\left(1 + \frac{x}{u}\right)\right)^x
$$

$$
+ \sum_{x \leq x^*} q(x) \left(1 + \frac{a^2}{\theta_0^2}\right)^{-u/2} \left(\theta_0\left(1 + \frac{x}{u}\right)\right)^x
$$

$$
\leq \frac{3}{nC(\theta_1)},
$$

which implies by (20) that

$$
|w| = \left| \int_0^{\theta_0} \cos \left( \frac{u(\theta - a)}{\theta_0} - \frac{m\pi}{2} \right) \sum_{x=0}^{\infty} q(x) \theta^x u_v(\theta) \, d\theta \right| \leq \frac{3}{nC(\theta_1)}.
$$
Since \( \sum_{x=0}^{\infty} q(x) \int_0^\theta w \theta^x h_{u,v}(\theta) \, d\theta = |w| \int_0^\theta h_{u,v}(\theta) / C(\theta) \, d\theta \leq |w| / C(\theta_0) \), 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 \( Z = (X_1, \ldots, X_n) \) and \( \lambda > 1 \), we have

\[
P_{g_{1n}}\{f(Z; g_{1n}) > \lambda f(Z; g_{2n})\} \leq \frac{\lambda}{\lambda - 1} \sum_{Z} |f(Z; g_{1n}) - f(Z; g_{2n})|
\leq \frac{n\lambda}{\lambda - 1} \sum_{x} |f(x; g_{1n}) - f(x; g_{2n})| \to 0,
\]

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

**Proof of Lemma 2.** Equation (25) follows from the Stirling formula. Inequalities (26) and (27) are based on the fact that

\[
\int_0^\infty t^a \left( 1 + t^2 \right)^{-u/2} dt = \frac{\Gamma((\alpha + 1)/2)\Gamma((u - \alpha - 1)/2)}{2\Gamma(u/2)}
\]

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

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

and

\[
|h_{u,v}^{(\alpha)}(\theta) - h_{u,v}^{(\alpha)}(a)| \leq \max_t \left\{ \frac{|e^{it} - 1|}{t^{\alpha}} \frac{\left| \theta - a \right|^{\alpha}}{2\pi} \int_{-\infty}^{\infty} |t|^{\alpha} \left( 1 + \left( \frac{t}{v} \right)^2 \right)^{-u/2} dt \right\}
\leq 2\left| \theta - a \right|^{\alpha} \frac{v^{\alpha+1}}{2\pi} \frac{\Gamma((\alpha + 1)/2)\Gamma((u - \alpha - 1)/2)}{\Gamma(u/2)}.
\]
To prove (28) we have

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

To prove (29) we have

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

4. The case of infinite $\theta^*$. The natural value of $\theta^*$ in (2) is $\theta_0^* = \sup \{ \theta : \Sigma_x q(x) \theta^x < \infty \}$, which is always known. If $\theta_0^*$ 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 $(\log n)^{-(1 - \beta/2k_m - m)}$ of convergence over the class $G_0$ in (3).

Let $Y = 2X$ and $\eta = \sqrt{\theta}$. The conditional density $f(y \mid \eta) = C(\eta^2)q(y/2)\eta^y$ is of the form (2). If $q(y/2) > 0$ for all nonnegative even integers $y$, then we can estimate the density $g_\eta$ of $\eta$ via the kernel in (8). The effect of this reparametrization is essentially halving the value of $\beta$ 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 $g_\eta(\sqrt{\alpha}) = 2\sqrt{\alpha} g(\alpha)$, (8) provides a kernel (with $j = 0$
and \( m_n = \infty \) for \( g_n(\sqrt{a}) C(a) = 2\sqrt{a} C(a) g(a) \),

\[
K_n^{(0)}(y, \sqrt{a}) = \bar{K}_n(x, a)
\]

\[
= \pi q(x)(2x)!^{-1}(-1)^x \int \cos(t\sqrt{a}) t^{2x} k_* \left( \frac{t}{c_n} \right) dt
\]

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

\[
\hat{g}_n(a) = n^{-1} \sum \frac{\bar{K}_n(X_i, a)}{2\sqrt{a} C(a)}, \quad \hat{g}_n^{(m)}(a) = \left( \frac{d}{d\theta} \right)^m \hat{g}_n(\theta) \bigg|_{\theta = a}.
\]

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 \( q(x) > 0, x = 0, 1, 2, \ldots \)], 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

\[
1 \leq q(x) B_0 B^{2x}((2x)!)^{\beta/2}, \quad x = 0, 1, \ldots,
\]

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

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

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 \( \beta \) 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),

\[
|\bar{K}_n(x, a)| \leq \{ \pi q(x)(2x)! \}^{-1/2} \|k(y)\|_1 \frac{c_n^{2x+1}}{2x+1}
\]

\[
\leq \frac{2\|k(y)\|_1 B_0/B}{\pi q(x) B_0 B^{2x}((2x)!)^{\beta/2}} \left( Bc_n \right)^{2x+1} (2x+1)!^{1-\beta/2}
\]

\[
\leq B_3 n^{\beta_0},
\]

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

\[
c_{n,j} = \max_{|d| \leq c_n} \left| \frac{d}{da} \right|^j \cos(t\sqrt{a}) \leq (1 + o(1)) \left( \frac{c_n}{2\sqrt{a}} \right)^j
\]

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

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

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

$$f(x) = f(x; G) = \int f(x \mid \theta) \, dG(\theta),$$

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

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

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

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

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

$$\mathcal{F}_{\alpha, \theta^*}^{cdef} = \mathcal{F}_{\alpha, \theta^*}^{cdef}(a, \delta, M, M_0, M_1)$$

$$= \left\{ G : |G_c^{(\alpha' + 1)}(\theta) - G_c^{(\alpha' + 1)}(a)| \leq M|\theta - a|^\alpha, \forall 0 \leq |\theta - a| \leq \delta, \right\},$$

$$\sum_{j=0}^{\alpha' + 1} |G_c^{(j)}(a)| \leq M_0, G_d([a - \delta, a) \cup (a, a + \delta)] = 0,$$

$$G(\theta) \leq M_1\theta^\alpha, \forall 0 \leq \theta \leq \delta, G(\theta^*) = 1 \right\},$$

where $G_c$ and $G_d$ are the continuous and discrete components of $G$; $\alpha'$ is the greatest integer less than $\alpha$; 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 $G \in \mathcal{F}_{\alpha, \theta^*}^{cdef}(a, \delta, M, M_0, M_1)$ 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 $\alpha > -1$,

$$\sup \{ E_G |\hat{G}_n(a) - G(a)|^2 : G \in \mathcal{F}_{\alpha, \theta^*}^{cdef} \} = O(n^{2(\alpha + 1)}).$$
COROLLARY 3. Suppose (16) and the conditions of Theorem 5 hold. Then, for \(0 \leq a < \theta^*\),

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

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

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

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

\[
B_{1n}(a) = \int \left[ \int_0^\theta \left( \frac{1}{C(y)} \right) H\left( dy - \frac{\theta}{c_n} \right) - G(a) \right] k(\theta) \, d\theta
\]

\[
= \int \left[ \int \frac{C(y)}{C(y + \theta/c_n)} I\left\{ 0 \leq y + \frac{\theta}{c_n} \leq a \right\} \, dG(y) - G(a) \right] k(\theta) \, d\theta.
\]

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 \(\delta_0\) and \(M^*\) are positive numbers. Hence, by (6),

\[
B_{1n}(a) \leq O(c_n^{-\alpha-1}) + \int \sum_{j=0}^{\alpha+1} \left[ \int_{\theta/c_n \leq \delta_0} d_j(y) \, dG(y) - \int_0^\theta d_j(y) \, dG(y) \right] \left( \frac{\theta}{c_n} \right)^j k(\theta) \, d\theta
\]

\[
= O(c_n^{-\alpha-1}),
\]

where the \(O(1)\) is uniform in the class \(G \in \mathcal{F}_{a, \theta}^c\). \(\square\)

Lower bounds for the optimal rate of convergence are given below.
THEOREM 6. For $0 \leq \alpha < \theta^*$,
\[
\liminf_{n \to \infty} (\log n)^{2(\alpha+1)} \inf_{\theta_n} \sup \{|\tilde{G}_n(a) - G(a)|^2 : G \in \mathcal{G}_{\alpha, \delta_0}^{cdf}\} > 0.
\]

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,

\[
G_{2n}(a) - G_{2n}(a) = w \theta^a u^{-a-1} \int_0^a \left[ \cos \left( u \frac{\theta - a}{\theta_0} + \frac{\pi}{2} \right) - w \right] g_{u,v}(\theta) d\theta
\]

\[
= w \theta^a u^{-a-1} \int_0^a g_{u,v}(\theta) d\sin \left( u \frac{\theta - a}{\theta_0} + \frac{\pi}{2} \right) + O \left( \frac{u^{-a-1/2}}{n} \right)
\]

\[
= -w \theta^a u^{-a-1} \left[ g_{u,v}(a) + \int_0^a \cos \left( u \frac{\theta - a}{\theta_0} \right) dg_{u,v}(\theta) \right]
\]

\[
+ O \left( \frac{u^{-a-1/2}}{n} \right)
\]

\[
= - \frac{(w \theta^a u^{-a-1} + o(1)) u^{-a-1}}{aC(a)\sqrt{2\pi}}.
\]

REFERENCES


LINDSAY, B. G. (1983b). The geometry of mixture likelihoods, part II: the exponential family. 
Ann. Statist. 11 783–792.
ROBBINS, H. (1951). Asymptotically subminimax solutions of compound statistical decision prob-
Statist. 18 806–831.

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