Statistica Sinica **6**(1996), 561-578

GLOBAL PROPERTIES OF KERNEL ESTIMATORS FOR MIXING DENSITIES IN DISCRETE EXPONENTIAL FAMILY MODELS

Wei-Liem Loh and Cun-Hui Zhang

Purdue University and Rutgers University

Abstract: This paper concerns the global performance of modifications of the kernel estimators considered in Zhang (1995) for a mixing density function g based on a sample from $f(x) = \int f(x|\theta)g(\theta)d\theta$ under weighted L^p-loss, $1 \le p \le \infty$, 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 upper bounds for the rate of convergence of the kernel estimators and lower bounds for the optimal convergence rate over various smoothness classes of mixing density functions. In particular under mild conditions, it is shown that these estimators achieve the optimal rate of convergence for the negative binomial mixture and are almost optimal for the Poisson mixture. Global estimation of the mixing distribution function under weighted L^p-loss is also considered.

Key words and phrases: Mixing density, kernel estimator, discrete exponential family, rate of convergence.

1. Introduction

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

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

where $f(x|\theta)$ is a known parametric family of probability density functions with respect to a σ-finite measure μ, and g is a mixing density function on $(0, \theta^*)$. Suppose

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

where $0 \le \theta \le$ (or $<$) $\theta^* \le \infty$, $q(x) > 0$ whenever $x = 0, 1, 2, \ldots$ and μ is the counting measure on the set of nonnegative integers.

Zhang (1995) considered a class of kernel estimators for the mixing density function g, its derivatives, and the mixing distribution function, and proved that the mean squared error at a fixed point achieves or almost achieves the optimal rate of convergence under mild conditions. However these estimators do not perform uniformly well asymptotically near the boundary of the domain of θ. In this paper we consider modifications of these estimators which improve their global performance (especially near the boundary) with respect to weighted L^p-loss, $1 \leq p \leq \infty$. In particular, Sections 2 and 3 give upper bounds for the convergence rates of the estimators for the mixing density and distribution respectively for the case where θ^* is finite and known. Section 4 supplies corresponding lower bounds for the optimal convergence rate. A consequence of Corollaries 1 and 2 and Theorem 3 in Sections 2, 3, and 4 is that, the kernel estimators achieve or almost achieve the optimal rate of convergence under mild assumptions. For the negative binomial distribution $f(x|\theta) = \binom{x+r-1}{r-1}(1-\theta)^r\theta^x$, the optimal rate for estimating $g(\cdot)$ is $O(\{1/\log n\}^\alpha)$ under a global weighted L^p-loss, where α denotes the degree of smoothness of g; whereas for the Poisson distribution, the optimal rate is at most as rapid as $O(\{1/\log n\}^\alpha)$ and the kernel estimates converge at the rate $O(\{\log\log n/\log n\}^\alpha)$ when θ^* is known and finite. Section 5 considers the case $\theta^* = \infty$, where the kernel estimators achieve certain rates of convergence which may not be optimal. For example, our kernel estimators converge at the rate $O(\{1/\log n\}^{\alpha/2})$ for the Poisson distribution under a global weighted L_p-loss when $\theta^* = \infty$. All the proofs are given in three appendices: upper bounds for finite θ^* in Appendix I, lower bounds in Appendix II, and upper bounds for $\theta^* = \infty$ in Appendix III.

A key point of this paper is that, in general, without further assumptions, global nonparametric estimation of a mixing density (or distribution) of a discrete exponential family is difficult in that the optimal rate of convergence is logarithmic (not polynomial).

Among related mixture problems, the deconvolution problem appears to be the best understood. Recent and important advances to the solution were made by Carroll and Hall (1988), Fan (1991a, b), Zhang (1990) and many others using Fourier analysis. In particular kernel estimators for the mixing density (or distribution) have been obtained which achieve the optimal convergence rate. However these methods are based on the convolution property of the mixture problem and are not directly applicable to (2).

Another problem that has been of much interest is the estimation of the mixing distribution of a Poisson mixture. Tucker (1963) approached this problem through the method of moments, and Lambert and Tierney (1984) and Simar (1976) considered the nonparametric maximum likelihood estimation for the mixing distribution. For mixtures of more general discrete exponential distributions, Loh (1993) and Zhang (1995) independently proposed estimates and obtained upper and lower bounds on the mean squared error at a fixed point via Fourier analysis, but they did not consider the global properties discussed here. Walter and Hamedani (1991) successfully applied orthogonal polynomial techniques to mixtures of exponential families. 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_g and E_g the probability and expectation, respectively, corresponding to g, by $I\{\cdot\}$ the indicator function, by $h^{(j)}$ the jth derivative (if it exists) of h with $h^{(0)} = h$, by h^* the Fourier transformation of any integrable h, $h^*(t) = \int e^{ity}h(y)dy$, and by $\|\cdot\|_p$ the L^p-norm with respect to the Lebesgue measure, $1 \le p \le \infty$. We shall use the notation κ' and κ'' to denote the decomposition $\kappa = \kappa' + \kappa''$ such that κ' is an integer and $0 < \kappa'' \le 1$ for all real numbers κ.

2. Kernel Estimators

In this and the next two sections, we assume that θ^* is finite and known. The case $\theta^* = \infty$ is considered in Section 5.

Let $k(\cdot)$ be a symmetric function satisfying $\int_{-\infty}^{\infty} k(y)dy = 1$, $k^*(t) = 0$, $\forall |t| > 1$,

$$\int_{-\infty}^{\infty} y^j k(y)dy = 0, \qquad \forall 1 \le j < \alpha_0, \qquad \int_{-\infty}^{\infty} |y^{\alpha_0} k(y)|\, dy < \infty. \tag{3}$$

For suitable positive constants m_n and c_n tending to ∞ and to be specified later, define

$$K_n(x,\theta) = \frac{I\{0 \le x \le m_n\}}{2\pi q(x) x!} \int_{-c_n}^{c_n} t^x \cos(x\pi/2 - t\theta) k^*(t/c_n)dt. \tag{4}$$

Given any probability density function g on $(0, \theta^*)$, we shall extend its domain to the whole real line by setting $g(y) = g(y)I\{0 < y < \theta^*\}$ for all $y \in R$. Let

$$h(y) = C(y)g(y), \qquad \forall -\infty < y < \infty. \tag{5}$$

It follows from (1) and (2) that $f(x; g)/q(x) = \int_0^{\theta^*} \theta^x h(\theta)d\theta$. Taking infinite series expansions in the Fourier inversion formula as in Zhang (1995), we obtain

$$\begin{aligned}\int_{-\infty}^{\infty} h(\theta - y/c_n)k(y)dy &= \int_0^{\theta^*} c_n k(c_n(\theta - y))h(y)dy \\ &= \sum_{x=0}^{\infty} (2\pi x!)^{-1} \int_0^{\theta^*} y^x h(y)dy \int_{-c_n}^{c_n} (it)^x e^{-it\theta} k^*(t/c_n)dt.\end{aligned}$$

Thus, $K_n(x,\theta)$ can be used as a kernel for h in the sense that for $-\infty < \theta < \infty$,

$$E_g K_n(X_1, \theta) - h(\theta) = b_{1n}(\theta) + b_{2n}(\theta) \to 0, \tag{6}$$

as $(m_n, c_n) \to (\infty, \infty)$ along a suitable path for continuous $h(\theta) = g(\theta)C(\theta)$, where

$$b_{1n}(\theta) = \int_{-\infty}^{\infty} [h(\theta - y/c_n) - h(\theta)]k(y)dy, \tag{7}$$

$$b_{2n}(\theta) = -\sum_{x>m_n} \frac{\int_0^{\theta^*} y^x h(y)dy}{2\pi x!} \int_{-c_n}^{c_n} t^x \cos(x\pi/2 - t\theta) k^*(t/c_n) dt. \tag{8}$$

With this as motivation, we estimate $g(\theta)$ by

$$\hat{g}_n(\theta) = n^{-1} \sum_{j=1}^{n} \{K_n(X_j, \theta)/C(\theta)\}\, I\{0 \le \theta \le a_n\}. \tag{9}$$

The positive integers m_n and positive constants c_n and a_n are chosen such that

$$m_n = \min\{m \ge 1 : (\theta^* e)c_n(m) + \beta_1 \log c_n(m) < m+1\}, \tag{10}$$

$$c_n = c_n(m_n), \quad c_n(m) = \beta_0 \log n - \max_{1\le x\le m} \log(1/q(x)), \tag{11}$$

$$a_n = \theta^* - a^*/c_n \quad \text{if} \quad C(\theta^*) = 0, \quad \text{and} \quad a_n = \theta^* \text{ if } C(\theta^*) > 0, \tag{12}$$

with some constants $0 < \beta_0 < 1/2$, $\beta_1 > 0$, and $0 < a^* < \infty$. It will be shown in the proof of Theorem 1 that (10) implies $\|b_{2n}\|_\infty = O(c_n^{-\beta_1})$ and that (11) implies $\|K_n\|_\infty = O(n^{\beta_0})$.

We shall investigate the global performance of (9) over the following classes of mixing densities. Let $1 \le p \le \infty$ and w be a member of L^p on $(0, \theta^*)$. For $\alpha > 0$ we define $\mathcal{G}_{\alpha,\theta^*} = \mathcal{G}_{\alpha,\theta^*}(p, w, M)$ to be the set of all densities g on $(0, \theta^*)$ such that

$$\|w(\theta)\{g^{(\alpha')}(\theta) - g^{(\alpha')}(\theta+\delta)\}\|_p < M|\delta|^{\alpha''}, \quad \forall \delta, \tag{13}$$

where α' is the integer with $0 < \alpha'' = \alpha - \alpha' \le 1$, and $0 < M < \infty$.

Assume that there exist constants $\gamma \ge 0$, C_1^*, C_2^*, and C_3^* such that

$$\sup_{0<\theta<\theta^*} (\theta^* - \theta)^\gamma / C(\theta) < C_1^*, \tag{14}$$

$$\sup_{0<\theta<\theta^*} (\theta^* - \theta)^j |C^{(j)}(\theta)| / \{C(\theta) j!\} < C_2^*, \quad \forall 0 \le j \le \rho', \tag{15}$$

$$|C^{(\rho')}(\theta+\delta) - C^{(\rho')}(\theta)| < C_3^* \delta^{\rho''}, \; 0 < \theta < \theta + \delta < \theta^*, \tag{16}$$

where ρ' is a nonnegative integer with $0 < \rho'' = \rho - \rho' \le 1$.

Theorem 1. *Suppose* $\alpha > 0$, $1 \le p \le \infty$, *and* (14)-(16) *hold with* $\gamma \ge 0$ *and* $\rho = \alpha + \gamma$. *Let* $\hat{g}_n$ *be given by* (9) *with the kernel* $K_n(x, \theta)$ *in* (4) *such that* $\alpha_0 \ge \alpha + \gamma$ *in* (3). *Let* (10)-(12) *hold with* $\beta_1 \ge \alpha + \gamma$. *Then,* $\sup\{E_g \|w(\hat{g}_n - g)\|_p : g \in \mathcal{G}_{\alpha,\theta^*}(p, w, M)\} = O(c_n^{-\alpha})$.

Remark. If $C(\theta^*) > 0$, we shall set $\gamma = 0$ although (14) holds for all $\gamma \ge 0$ and $\rho > 0$. Conditions (15) and (16) are satisfied for every $\rho > 0$ if $C(\theta)$ is an analytic function in a neighborhood of θ^* (e.g. $C(\theta) = (1-\theta)^\nu$ with $\theta^* = 1$ for the negative binomial family).

Corollary 1. *Suppose the conditions of Theorem* 1 *are satisfied and*

$$q(x)B_0B^x(x!)^\beta \geq 1, \quad \forall x \geq 0, \tag{17}$$

for some constants B_0*,* B*, and* β*. Then,* $\sup_{g\in\mathcal{G}_{\alpha,\theta^*}} E_g\|w(\hat{g}_n - g)\|_p = O(1)(1/\log n)^\alpha$ *if* $\beta = 0$*, and* $\sup_{g\in\mathcal{G}_{\alpha,\theta^*}} E_g\|w(\hat{g}_n - g)\|_p = O(1)(\log\log n/\log n)^\alpha$ *if* $0 < \beta < \infty$*.*

The conditions of Corollary 1 are satisfied in the following examples.

Example 1. (Negative Binomial) Since $C(\theta) = (1-\theta)^r$ and $q(x) = \binom{x+r-1}{r-1}$ for some known integer $r \geq 1$ and $\theta^* = 1$, (14)-(16) hold for $\gamma = r-1$ and all $\rho > 0$ and (17) holds for $\beta = 0$. By (10) and (11) $c_n = \beta_0 \log n + (r-1)\log\log n + O(1)$ and $m_n = e\beta_0 \log n + (e(r-1)+\beta_1)\log\log n + O(1)$.

Example 2. (Poisson with $\theta^* < \infty$) Here $C(\theta) = e^{-\theta}$ and $q(x) = 1/x!$, so that (14)-(16) hold for $\gamma = 0$ and all $\rho > 0$ and (17) holds for $\beta = 1$. By (10) and (11) $c_n = (\theta^* e)^{-1}(\beta_0 + o(1))\log n/\log\log n$ and $m_n = (\beta_0 + o(1))\log n/\log\log n$.

3. Estimating a Mixing Distribution

Let $f(x|\theta)$ be as in (2) with θ^* finite and known. Suppose the marginal density of X is

$$f(x;G) = \int_0^{\theta^*} f(x|\theta)dG(\theta), \tag{18}$$

where G is the mixing distribution. If the density $g = G^{(1)}$ exists, then $f(x;G) = f(x;g)$. In this section we consider the estimation of the mixing distribution G. Our results here are parallel to those in Section 2 for the estimation of the mixing density. We denote by E_G the expectation when G is the true mixing distribution.

Let $K_n(x,\theta)$ be as in (4) with the constants m_n and c_n in (10) and (11). Define

$$\hat{G}_n(\theta) = \begin{cases} n^{-1}\sum_{j=1}^n \int_{a_*}^\theta K_n(X_j,y)\{C(y)\}^{-1}dy, & \text{if } 0 < \theta \leq a_n, \\ 1, & \text{if } \theta > a_n, \end{cases} \tag{19}$$

where a_n is as in (12), and a_* is a negative constant such that $1/C(y) = \sum_{x=0}^\infty q(x)y^x$ is an increasing analytic function for $a_* \leq y < \theta^*$. Similar to (6)–(8), for continuous $G(\theta)$ we have

$$E_G\hat{G}_n(\theta) - G(\theta) = B_{1n}(\theta) + B_{2n}(\theta) \to 0 \tag{20}$$

for $0 < \theta < a_n$, where

$$B_{1n}(\theta) = \int_{-\infty}^\infty \int_{a_*}^\theta \left\{\frac{C(z - y/c_n)}{C(z)}\right\} d_z G(z - y/c_n)k(y)dy - G(\theta), \tag{21}$$

$$B_{2n}(\theta) = -\sum_{x>m_n} \frac{\int_0^{\theta^*} y^x h(y)dy}{2\pi x!} \int_{-c_n}^{c_n} t^x \int_{a_*}^{\theta} \frac{\cos(x\pi/2 - tz)}{C(z)} dz k^*(t/c_n)dt. \quad (22)$$

Let $\alpha > -1$, $1 \le p \le \infty$ and $w(\cdot)$ be a decreasing function on $[0, \theta^*]$. Define $\mathcal{G}^{cdf}_{\alpha,\theta^*} = \mathcal{G}^{cdf}_{\alpha,\theta^*}(p, w, M)$ to be the set of all distribution functions G on $(0, \theta^*)$ such that

$$\|w(\theta)\{G^{(\alpha'+1)}(\theta) - G^{(\alpha'+1)}(\theta + \delta)\}\|_p < M|\delta|^{\alpha''}, \quad \forall\delta. \quad (23)$$

Remark. If $0 < 1 + \alpha \le 1/p$, $w = 1$, and $M > 1$, then $\mathcal{G}^{cdf}_{\alpha,\theta^*}$ is the class of all distribution functions on $(0, \theta^*)$. If $\alpha > 0$, then $\mathcal{G}^{cdf}_{\alpha,\theta^*}(p, w, M) = \mathcal{G}_{\alpha,\theta^*}(p, w, M)$.

Theorem 2. *Let $\alpha > -1$ and $1 \le p \le \infty$. Suppose* (14)-(16) *hold with $\rho \ge \alpha + 1 + \max(\gamma, 1)$ if $\gamma \ne 1$ and $\rho > \alpha + 2$ if $\gamma = 1$. Let $\hat{G}_n$ be given by* (19) *with $\alpha_0 \ge \alpha + 1 + \gamma$ in* (3) *and $\beta_1 \ge \alpha + \gamma$ in* (10)*. Then,* $\sup\{E_G\|w(\hat{G}_n - G)\|_p : G \in \mathcal{G}^{cdf}_{\alpha,\theta^*}(p, w, M)\} = O(c_n^{-\alpha-1})$.

Corollary 2. *Suppose that* (17) *and the conditions of Theorem* 2 *hold. If $\beta = 0$ then* $\sup_{G\in\mathcal{G}^{cdf}_{\alpha,\theta^*}} E_G\|w(\hat{G}_n - G)\|_p = O(1)(1/\log n)^{\alpha+1}$. *If $0 < \beta < \infty$ then* $\sup_{G\in\mathcal{G}^{cdf}_{\alpha,\theta^*}} E_G\|w(\hat{G}_n - G)\|_p = O(1)(\log\log n/\log n)^{\alpha+1}$.

4. Optimal Rate of Convergence

In Sections 2 and 3, we obtained upper bounds for the maximum $\|\cdot\|_p$ risk of our kernel estimators over the classes $\mathcal{G}_{\alpha,\theta^*}$ and $\mathcal{G}^{cdf}_{\alpha,\theta^*}$. Here we derive corresponding lower bounds for the rate of the local minimax risk at interior points g_0 and G_0 of these classes

$$r_{n,\alpha,\theta^*}(g_0) = \inf_{\tilde{g}_n} \sup\{E_g\|\tilde{g}_n - g\|_p : g \in \mathcal{G}_{\alpha,\theta^*}, \|g - g_0\|_p \le M_1(\log n)^{-\alpha}\}, \quad (24)$$

$$r^{cdf}_{n,\alpha,\theta^*}(G_0) = \inf_{\tilde{G}_n} \sup\{E_G\|\tilde{G}_n - G\|_p : G \in \mathcal{G}^{cdf}_{\alpha,\theta^*}, \|G - G_0\|_p \le M_1(\log n)^{-\alpha-1}\}, \quad (25)$$

where the infimum runs over all statistics based on $X_1, \ldots, X_n$, and the classes are given by (13) and (23) respectively with $w(\theta) = I\{0 \le \theta \le \theta^*\}$. The rates of (24) and (25) can be regarded as characterizations of the degree of difficulty for estimating g and G respectively.

Theorem 3. *Let $M_1 > 0$ and g_0 and G_0 be interior points of $\mathcal{G}_{\alpha,\theta^*}(p, 1, M)$ and $\mathcal{G}^{cdf}_{\alpha,\theta^*}(p, 1, M)$.*
(i) *If $1 \le p \le \infty$ and $\alpha > 0$, then* $\liminf_{n\to\infty}(\log n)^{\alpha} r_{n,\alpha,\theta^*}(g_0) > 0$.
(ii) *If $1 \le p \le \infty$ and $\alpha \ge 0$, then* $\liminf_{n\to\infty}(\log n)^{\alpha+1} r^{cdf}_{n,\alpha,\theta^*}(G_0) > 0$.

The basic idea behind the proof of Theorem 3 is to find mixing densities g_{0n}, g_{1n}, and g_{2n}, close to g_0 and in $\mathcal{G}_{\alpha,\theta^*}(p,1,M)$, such that $\max_{j=1,2}\|g_{jn}-g_{0n}\|_p$ tends to 0 at a much slower rate than $f(\cdot;g_{jn})-f(\cdot;g_{0n})$, $j=1,2$.

We shall show in the proof of Theorem 3 that

$$\liminf_{n\to\infty} p_n = p_0 > 0, \quad p_n = \inf_{\tilde{g}_n}\max_{0\le j\le 2} P_{g_{jn}}\{\|\tilde{g}_n - g_{jn}\|_p > \varepsilon_0(\log n)^{-\alpha}\} \tag{26}$$

for some $\varepsilon_0 > 0$, and that g_{jn}, $0 \le j \le 2$, are members of $\mathcal{G}_{\alpha,\theta^*}(p,1,M)$ for small w_0 with $\|g_{jn}-g_0\|_p \le M_1(\log n)^{-\alpha}$. This will prove Theorem 3 (i), since $r_{n,\alpha,\theta^*}(g_0) \ge \varepsilon_0 p_n(\log n)^{-\alpha}$.

The densities g_{jn} are constructed in the following manner. Let $0<\theta_0<\theta_1<a<\theta_2<\theta_3<\theta^*$ be fixed constants. Define

$$h_{u,v}(\theta) = v^u\theta^{u-1}e^{-v\theta}/\Gamma(u), \tag{27}$$

$$g_{u,v}(\theta) = \{\chi_0(\theta)l_{1,u,v}(\theta)+\chi_1(\theta)h_{u,v}(\theta)+\chi_2(\theta)l_{2,u,v}(\theta)\}/C(\theta), \tag{28}$$

where $\chi_j=\chi_j(\theta)=I\{\theta_j\le\theta<\theta_{j+1}\}$ and $l_{j,u,v}$, $j=1,2$, are polynomials each of degree $(2\alpha'+1)$ such that $g_{u,v}$ is α' times continuously differentiable. Define

$$g_{0n}(\theta)=g_0(\theta)+\frac{3w_0}{u_n^{(p-1)/(2p)}}\left(\frac{\theta_2}{u_n}\right)^{\alpha}\{g_{u_n,v_n}(\theta)+g_{00}(\theta)-(w_{0n}+1)g_0(\theta)\}, \tag{29}$$

$$g_{1n}(\theta)=g_{0n}(\theta)+\frac{w_0}{u_n^{(p-1)/(2p)}}\left(\frac{\theta_2}{u_n}\right)^{\alpha}\left[\sin\left(u_n\frac{\theta-a}{\theta_2}\right)-\frac{w_{1n}}{w_{0n}}\right]g_{u_n,v_n}(\theta), \tag{30}$$

$$g_{2n}(\theta)=g_{0n}(\theta)+\frac{w_0}{u_n^{(p-1)/(2p)}}\left(\frac{\theta_2}{u_n}\right)^{\alpha}\left[\cos\left(u_n\frac{\theta-a}{\theta_2}\right)-\frac{w_{2n}}{w_{0n}}\right]g_{u_n,v_n}(\theta), \tag{31}$$

where g_{00} is a density in $\mathcal{G}_{\alpha,\theta^*}(p,1,1)$ bounded away from 0 in $[\theta_0,\theta_3]$, w_{jn} are constants given by $\int g_{jn}(\theta)d\theta=1$, $w_0>0$, $u_n=\delta_0\log n$, and $v_n=u_n/a$, with

$$\delta_0=\max\left\{\frac{\theta_2/(\theta_3-\theta_2)}{\log(\theta_3/\theta_2)},\frac{2}{\log(1+a^2/\theta_2^2)},\frac{1}{\theta_1/a-1-\log(\theta_1/a)},\frac{1}{\theta_2/a-1-\log(\theta_2/a)}\right\}. \tag{32}$$

5. The Case of Infinite θ^*

The natural value of θ^* in (2) is $\theta_0^*=\sup\{\theta:\sum_x q(x)\theta^x<\infty\}$. If (17) holds with $\beta=0$, then θ_0^* is finite and known, and we can set $\theta^*=\theta_0^*$ and the results of Sections 2 and 3 follow. However, when $\theta_0^*=\infty$ the condition $\theta^*<\infty$ becomes an assumption in addition to the knowledge of $q(\cdot)$. In this section, we consider the case of $\theta^*=\infty$. Upper bounds of the L^p risks of our kernel estimators are provided in Theorems 4 and 5 below. The lower bounds of Theorem 3 still apply here.

Let $\eta = \sqrt{\theta}$, $0 < c_n \to \infty$ and $k(\cdot)$ be as in (3). Define

$$K_{s,n}(x,\eta) = [\pi q(x)(2x)!]^{-1}(-1)^x \int_{-c_n}^{c_n} \cos(t\eta) t^{2x} k^*(t/c_n)dt, \tag{33}$$

$$g_s(\eta) = g(\eta^2)I\{\eta \geq 0\},\ C_s(\eta) = 2\eta C(\eta^2),\ h_s(\eta) = g_s(\eta)C_s(\eta). \tag{34}$$

Since $f(x;g) = \int_{-\infty}^{\infty} y^{2x} h_s(y)dy$ by (1) and (2), we have

$$\begin{aligned} & E_g K_{s,n}(X_1,\eta) - 2\eta C(\eta^2)g(\eta^2)I\{\eta \geq 0\} \\ = & \int_0^\infty c_n\{k(c_n(\eta - y)) + k(c_n(\eta + y))\}h_s(y)dy - h_s(\eta) = b_{3n}(\eta) + b_{4n}(\eta) \end{aligned}$$

by the Fourier inversion formula, where

$$b_{3n}(\eta) = \int_{-\infty}^{\infty} k(y)\{h_s(\eta - y/c_n) - h_s(\eta)\}dy, \tag{35}$$

$$b_{4n}(\eta) = \int_{-\infty}^{\infty} k(y)h_s(y/c_n - \eta)dy. \tag{36}$$

Zhang (1995) used $K_{s,n}(x,\sqrt{a})/[2\sqrt{a}C(a)]$ as a kernel for $g(a)$ in the case $\theta^* = \infty$. Define

$$\hat{g}_{n,\infty}(\theta) = \hat{g}_{s,n}(\sqrt{\theta})I\{a_{0n}^2 \leq \theta \leq a_{1n}^2\}, \quad \hat{g}_{s,n}(\eta) = n^{-1}\sum_{j=1}^n \frac{K_{s,n}(X_j,\eta)}{C_s(\eta)}, \tag{37}$$

where a_{0n} and a_{1n} are positive constants tending to 0 and ∞ respectively. We shall study the global performance of this estimator under weighted L^p-loss.

Let $1 \leq p \leq \infty$ and w be as in Section 2. For $\alpha > 0$ define $\mathcal{G}_{\alpha,\infty} = \mathcal{G}_{\alpha,\infty}(p,w,\alpha_1,M,M_1)$ to be the set of all probability density functions g on $(0,\infty)$ such that

$$\|w_s(|\eta|)\{g_s^{(\alpha')}(\eta+\delta) - g_s^{(\alpha')}(\eta)\}\|_p < M|\delta|^{\alpha''}, \quad \forall\delta, \tag{38}$$

$$\|w(\theta)g(\theta)I\{\theta > a\}\|_p < M_1[C(a)]^{\alpha_1}, \quad \forall a > 0, \tag{39}$$

where $w_s(\eta) = (2\eta)^{1/p}w(\eta^2)$, and α_1, M, and M_1 are given constants. Note that with $\eta = \sqrt{\theta}$, $\|w(\theta)h_0(\sqrt{\theta})\|_p = \|w_s(\eta)h_0(\eta)\|_p$ for all Borel functions h_0.

We assume that for every $0 < \delta < 1$ there exists a finite constant C_δ^* such that

$$|C^{(j)}(\theta)|\{C(\theta)\}^{\delta-1} < C_\delta^*, \quad \forall\theta \geq 0, \tag{40}$$

for all $0 \leq j < \rho$. This condition holds for $C(\theta) = e^{-\theta}$ of the Poisson mixture.

Theorem 4. *Let* $\alpha > 0$ *and* $1 \leq p \leq \infty$. *Suppose* (14)-(16) *and* (40) *hold with* $\rho > \alpha(1+1/\alpha_1)$ *and* $\rho \geq \alpha + 1$, *and* (17) *holds with* $0 < \beta < 2$. *Let* $\hat{g}_{n,\infty}$ *be given by* (37) *with* $a_{0n} = a_*/c_n$, $C(a_{1n}^2) = c_n^{-\alpha/\alpha_1}$, $c_n = B^{-1}\{(\beta_0 \log n)/(1-\beta/2)\}^{1-\beta/2}$, *and* $K_{s,n}(x,\eta)$ *as in* (33) *with* $\alpha_0 \ge \rho$ *in* (3), *where* a_* *and* $\beta_0 < 1/2$ *are positive constants. Then,*

$$\sup\{E_g\|w(\hat{g}_{n,\infty} - g)\|_p : g \in \mathcal{G}_{\alpha,\infty}(p, w, \alpha_1, M, M_1)\} = O(1)(\log n)^{-\alpha(1-\beta/2)}.$$

Conditions of Theorem 4 hold for the Poisson example.

Example 3. (Poisson with $\theta^* = \infty$) As in Example 2, (14)-(16) hold for $\gamma = 0$ and all $\rho > 0$, (40) holds for all $\rho > 0$ and $0 < \delta < 1$, and (17) holds for $B_0 = B = \beta = 1$. Taking $\beta_0 = 1/4$, $\alpha_1 = \alpha$ and $a^* = 1$, we have $c_n = \sqrt{(\log n)/2}$, $a_{0n} = 1/c_n$ and $a_{1n} = \sqrt{\log c_n}$.

Now consider the estimation of the mixing distribution G for $\theta^* = \infty$. Define

$$\hat{G}_{n,\infty}(\theta) = \begin{cases} n^{-1}\sum_{j=1}^n \int_0^{\sqrt{\theta}} K_{s,n}(X_j, y)\{C(y^2)\}^{-1}dy, & \text{if } 0 < \theta \le a_n^2, \\ 1, & \text{if } \theta > a_n^2, \end{cases} \tag{41}$$

where $K_{s,n}(x,\eta)$ is as in (33) and $0 < a_n \to \infty$. Similar to (20)-(22) and (35)-(36),

$$E_G\hat{G}_{n,\infty}(\theta) - G(\theta) = B_{3n}(\sqrt{\theta}) + B_{4n}(\sqrt{\theta}) \tag{42}$$

for $0 < \theta < a_n$, where, with $G_s(y) = G(y^2)I\{y \ge 0\}$ and $C_{s,0}(y) = C(y^2)$,

$$\begin{aligned} B_{3n}(\eta) &= \int_{-\infty}^{\infty}\int_0^{\eta}\Big\{\frac{C_{s,0}(z - y/c_n)}{C_{s,0}(z)}\Big\}d_zG_s(z - y/c_n)k(y)dy - G_s(\eta), \\ B_{4n}(\eta) &= -\int_{-\infty}^{\infty}\int_0^{\eta}\Big\{\frac{C_{s,0}(y/c_n - z)}{C_{s,0}(z)}\Big\}d_zG_s(y/c_n - z)k(y)dy. \end{aligned}$$

Let w be a decreasing function on $(0,\infty)$ with $\|w\|_p < \infty$. For $\alpha > 0$ define $\mathcal{G}_{\alpha,\infty} = \mathcal{G}_{\alpha,\infty}(p, w, \alpha_1, M, M_1, M_2)$ to be the set of all densities g on $(0,\infty)$ such that

$$\|w_s(|\eta|)\{G_s^{(\alpha'+1)}(\eta+\delta) - G_s^{(\alpha'+1)}(\eta)\}\|_p < M|\delta|^{\alpha''}, \quad \forall\delta, \tag{43}$$

$$\|w(\theta)(1 - G(\theta))I\{\theta > a\}\|_p < M_1[C(a)]^{\alpha_1}, \ \forall a > 0, \tag{44}$$

$$G(\theta) \le M_2\theta^{(\alpha+1)/2}, \ \ \forall\theta > 0, \tag{45}$$

where w_s is as in (38), and α_1, M, M_1, and M_2 are given constants.

Remark. Although $b_{3n}(\eta) \to 0$ for $-\infty < \eta < \infty$ as $c_n \to \infty$, we have $b_{4n}(\eta) \to h_s(-\eta)$. Thus, the bias of $\hat{G}_{n,\infty}$ will not tend to 0 if we integrate from a negative number in (41) as we did in (19). This caused us to add Condition (45).

Theorem 5. *Let* $\alpha > -1$ *and* $1 \le p \le \infty$. *Suppose* (14)-(16) *and* (40) *hold with* $\rho > (\alpha+1)(1+\alpha_1^{-1})$, *and that* (17) *holds for some* $0 < \beta < 2$. *Let* $\hat{G}_{n,\infty}$ *be given by* (41) *with* c_n *as in Theorem 4,* $C(a_n^2) = c_n^{-(\alpha+1)/\alpha_1}$ *and* $\alpha_0 \ge \rho$ *in* (3). *Then*

$$\begin{aligned} &\sup\{E_G\|w(\hat{G}_{n,\infty} - G)\|_p : G \in \mathcal{G}_{\alpha,\infty}^{cdf}(p, w, \alpha_1, M, M_1, M_2)\} \\ &= O(1)(\log n)^{-(\alpha+1)(1-\beta/2)}. \end{aligned}$$

Acknowledgement

Research of Wei-Liem Loh was supported in part by National Security Agency. Research of Cun-Hui Zhang was supported in part by National Science Foundation and Army Research Office.

Appendix I. Proofs of the Upper Bounds for Finite θ^*

Throughout Appendices I, II and III, we use $\mathrm{Rem}_h(x,\delta,m)$ to denote the remainder of the $(m+1)$-term Taylor expansion of h, which can be written as

$$\begin{aligned}\mathrm{Rem}_h(x,\delta,m) &= h(x+\delta)-\sum_{j=0}^{m} h^{(j)}(x)\delta^j/j!, \quad m\ge 0,\\ &= \int_0^\delta \frac{(\delta-y)^{m-1}}{(m-1)!}\{h^{(m)}(x+y)-h^{(m)}(x)\}dy,\ m\ge 1. \qquad (46)\end{aligned}$$

Proof of Theorem 1. Let $\chi_{[0,a]}(\theta)=I\{0\le\theta\le a\}$. By (46)

$$\begin{aligned}\|\chi_{[0,a]}wg^{(j)}\|_p &= \Big\|\chi_{[0,a]}(\theta)w(\theta)\mathrm{Rem}_{g^{(j)}}(\theta-a,a,\alpha'-j)\Big\|_p\\ &\le \int_0^a \frac{(a-y)^{\alpha'-j-1}}{(\alpha'-j-1)!}\Big\|w(\theta)\{g^{(\alpha')}(\theta-a+y)-g^{(\alpha')}(\theta-a)\}\Big\|_p dy\end{aligned}$$

due to $g^{(j)}(\theta-a)=0$ for $0\le\theta<a$, so that for all $a>0$ and $0\le j\le\alpha'$ (13) implies

$$\|\chi_{[0,a]}wg^{(j)}\|_p\le 2Ma^{\alpha-j}/(\alpha'-j)!\ . \qquad (47)$$

Taking the expansion at $\theta+\theta^*-a$, we obtain in the same manner

$$\|(1-\chi_{[0,a]})wg^{(j)}\|_p\le 2M(\theta^*-a)^{\alpha-j}/(\alpha'-j)!\ ,\ a\le\theta^*, 0\le j\le\alpha'. \qquad (48)$$

Let $\chi_n(\theta)=I\{0\le\theta\le a_n\}$ and $\delta_n=\theta^*-a_n$. By (5) and (6),

$$\begin{aligned}&E_g\|w(\hat g_n-g)\|_p\\ \le\ & E_g\|w(\hat g_n-E\hat g_n)\|_p+\|\chi_nwb_{1n}/C\|_p+\|\chi_nwb_{2n}/C\|_p+\|(1-\chi_n)wg\|_p\ .\end{aligned}$$

By (12) and (48) $\|(1-\chi_n)wg\|_p\le 2M\delta_n^\alpha/\alpha'!=O(c_n^{-\alpha})$. It follows from Zhang (1995), Proof of Theorem 1, that $\|b_{2n}\|_\infty=O(c_n^{-\beta_1})$ and $\|K_n(x,\cdot)\|_\infty=O(1)c_n^{x+1}/\{q(x)(x+1)!\}$, so that by (12) and (14) $\|\chi_nwb_{2n}/C\|_p=O(c_n^{-\beta_1})\|\chi_n/C\|_\infty=O(c_n^{-\alpha})$ and by (11)

$$\begin{aligned}E_g\|w(\hat g_n-E\hat g_n)\|_p &= O(1)\|w\|_p\sum_{x=0}^{m_n}E_g|f_n(x)-f(x,g)|c_n^{x+1}/\{q(x)(x+1)!\}\\ &= O(1)n^{-1/2}\exp\{c_n+\max_{o\le x\le m_n}\log(1/q(x))\}=O(n^{-1/2+\beta_0}),\end{aligned}$$

where $f_n(x), x \geq 0$, is the relative frequency function. These and Lemma 1 below imply $E_g\|w(\hat{g}_n - g)\|_p = \|\chi_n w b_{1n}/C\|_p + O(c_n^{-\alpha}) = O(c_n^{-\alpha})$.

Lemma 1. *Suppose the conditions of Theorem* 1 *hold. Then,* $\|\chi_n w b_{1n}/C\|_p \leq O(c_n^{-\alpha})$, *where the* $O(1)$ *is uniform over* $\mathcal{G}_{\alpha,\theta^*}(p, w, M)$.

Proof. We extend the domain of C such that (16) holds for all real numbers θ and δ. By (46) we have the expansion

$$C(\theta - y/c_n)g(\theta - y/c_n) = \sum_{j=1}^{4} \xi_j(\theta, -y/c_n), \tag{49}$$

where $\xi_j(\theta,\delta) = \xi_j(\theta,\delta; g, C, \alpha, \gamma, \rho)$ are given by $\xi_1(\theta,\delta) = \sum_{j=0}^{\alpha'} g^{(j)}(\theta)\{\delta^j/j!\}$ $\mathrm{Rem}_C(\theta, \delta, \rho' - j)$, $\xi_2(\theta,\delta) = \mathrm{Rem}_g(\theta,\delta,\alpha') \sum_{j=0}^{\gamma'} C^{(j)}(\theta)\{\delta^j/j!\}$, $\xi_3(\theta,\delta) = \mathrm{Rem}_g$ (θ,δ,α') $\mathrm{Rem}_C(\theta,\delta,\gamma')$ and $\xi_4(\theta,\delta) = \sum_{j=0}^{\alpha'} g^{(j)}(\theta)\{\delta^j/j!\} \sum_{l=0}^{\rho'-j} C^{(l)}(\theta)\{\delta^l/l!\}$.

Let $\rho = \alpha + \gamma$. As $\int \xi_4(\theta, -y/c_n)k(y)dy = h(\theta)$ by (3) and (5), (5) and (7) imply

$$b_{1n}(\theta) = \sum_{j=1}^{3} \int_{-\infty}^{\infty} \xi_j(\theta, -y/c_n)k(y)dy. \tag{50}$$

Treating $\xi_j = \xi_j(\theta,\delta)$ as functions of θ, we obtain $\|\chi_{[0,a]} w\xi_1/C\|_p = O(1)\delta^\rho/C(a)$ by (47), $\|\chi_{[0,a]} w\xi_2/C\|_p = O(1)\delta^\alpha \sum_{j=0}^{\gamma'} (\delta/(\theta^* - a))^j$ by (13) and (15), and $\|\chi_{[0,a]} w\xi_3/C\|_p = O(1)\delta^\rho/C(a)$ by (13), where the $O(1)$ is uniform in (a,δ). The details of the above calculation can be found in Loh and Zhang (1993). Hence, by (50), (14), (12), and (3)

$$\begin{aligned} \|\chi_n w b_{1n}/C\|_p &= O(1) \int_{-\infty}^{\infty} \Big\{ |y/c_n|^\rho / C(a_n) + |y/c_n|^\alpha \sum_{j=0}^{\gamma'} |y/a^*|^j \Big\} |k(y)| dy \\ &= O(c_n^{-\alpha}) \left\{ \|y^\rho k(y)\|_1 + \|y^\alpha k(y)\|_1 \right\} = O(c_n^{-\alpha}) \ . \end{aligned}$$

Proof of Theorem 2. By (20)-(22) and the proof of Theorem 1 $E_G\|w(\hat{G}_n - G)\|_p$ is bounded by

$$\begin{aligned} & E_G\|w(\hat{G}_n - E_G\hat{G}_n)\|_p + \|\chi_n w(B_{1n} + B_{2n})\|_p + \|(1-\chi_n)w(1-G)\|_p \\ \leq\ & O(c_n^{-\alpha-1}) + \|\chi_n w B_{1n}\|_p + \|(1-\chi_n)w(1-G)\|_p, \end{aligned}$$

since $c_n\|\chi_n B_{2n} C\|_\infty \leq 2(1 + 1/m_n) \sum_{x > m_n} (\theta^*)^x C(0) c_n^{x+1} \|k\|_1 / \{\pi(x+1)!\} = O(c_n^{-\alpha})$ by (10) and the fact that $|\int_{a_*}^{\theta} e^{-itz}\{C(z)\}^{-1}dz| \leq 2/\{|t|C(\theta)\}$. By the proof of (48), we have $\|(1-\chi_n)w(1-G)\|_p \leq 3M\delta_n^{\alpha+1}/(\alpha'+1)!$. The conclusion follows from

Lemma 2. *Let $\chi_n(\theta) = I\{0 \le \theta \le a_n\}$. Under the conditions of Theorem 2, $\|\chi_n w B_{1n}\|_p = O(c_n^{-\alpha-1})$, where the $O(1)$ is uniform over $\mathcal{G}^{cdf}_{\alpha,\theta^*}(p, w, M)$.*

Proof. Let $\xi_j(\theta, \delta; g, C, \alpha, \gamma, \rho)$ be as in (49). For $j = 0, 1$, define

$$\xi_{jn} = \xi_{jn}(\theta) = \int_{-\infty}^{\infty} \sum_{l=1}^{3} \xi_l(\theta, -y/c_n; G, C^{(j)}, \alpha + 1, \gamma_j, \rho_j) k(y) dy,$$

where $\rho_j = \alpha + 1 + \gamma_j$, $\min(\gamma_0, \gamma_1) \ge 0$, and $\max(\rho_0, \rho_1 + 1) \le \rho$. Note that the pair (g, C) in (49) is replaced by $(G, C^{(j)})$ here. Integrating by parts in (21), we find

$$B_{1n} = \frac{\xi_{0n}(\theta)}{C(\theta)} - \frac{\xi_{0n}(a_*)}{C(a_*)} + \int_{a_*}^{\theta} \Big\{ \frac{C^{(1)}(z)}{C(z)} \frac{\xi_{0n}(z)}{C(z)} - \frac{\xi_{1n}(z)}{C(z)} \Big\} dz. \tag{51}$$

Since the proof of Lemma 1 depends only on the smoothness and boundedness of $g^{(j)}$, $C^{(j)}$, and χ_n/C, it also applies to the components of ξ_{0n} and ξ_{1n}. It follows that $\|\chi_n w \xi_{0n}/C\|_p = O(c_n^{-\alpha-1})$ and that there exist functions ζ_{jn} with $\|\chi_n w \zeta_{jn}\|_p = O(c_n^{-\alpha-1})$ such that

$$\frac{C^{(1)}(z)}{C(z)} \frac{\xi_{0n}(z)}{C(z)} - \frac{\xi_{1n}(z)}{C(z)} = \sum_{j=0}^{2} \zeta_{jn}(z) h_{jn}(z),$$

where $h_{0n} = C^{(1)}/[c_n^{\gamma_0} C^2]$, $h_{1n} = 1/[c_n^{\gamma_1} C]$, and $h_{2n} = 1/[c_n(\theta^* - \theta)^2]$. Note the cancellation of the term $\{C^{(1)}/C\}\mathrm{Rem}_g$ here. By the Hölder inequality and the monotonicity of w,

$$\Big| w(\theta) \int_{a_*}^{\theta} \Big\{ \frac{C^{(1)}(z)}{C(z)} \frac{\xi_{0n}(z)}{C(z)} - \frac{\xi_{1n}(z)}{C(z)} \Big\} dz \Big| \le O(c_n^{-\alpha-1}) \sum_{j=0}^{2} \bar{h}_{jn}(\theta)$$

for fixed $\theta \le a_n$, where $\bar{h}_{jn}(\theta) = \|\chi_{[0,\theta]} h_{jn}\|_{p/(p-1)}$. This and (51) imply

$$\|\chi_n w B_{1n}\|_p \le \|w\|_p \Big| \frac{\xi_{0n}(a_*)}{C(a_*)} \Big| + O(c_n^{-\alpha-1}) \Big(1 + \sum_{j=0}^{2} \|\chi_n \bar{h}_{jn}\|_p \Big). \tag{52}$$

Set $\gamma_0 = \gamma$, $\gamma_1 = \max(\gamma - 1, 0)$ if $\gamma \ne 1$, and $1 \ge \gamma_1 > 0$ if $\gamma = 1$. Let $h_0(z) = (\theta^* - z)^{-\kappa-1}$ and $\bar{h}_0(\theta) = \|\chi_{[0,\theta]} h_0\|_{p/(p-1)}$. Then for $\kappa > 0$ and $1 \le p \le \infty$,

$$\|\chi_n(\theta) \bar{h}_0(\theta)\|_p \le (\theta^* - a_n)^{-\kappa} \{(p\kappa + 1)/(p-1)\}^{-(p-1)/p} \{p\kappa\}^{-1/p}.$$

This gives $\|\chi_n \bar{h}_{2n}\|_p = O(1)$ by (12). This also gives $\|\chi_n \bar{h}_{0n}\|_p = O(1)$ if $\gamma > 0$, while $\|\bar{h}_{0n}\|_\infty = O(1)$ by (16) if $\gamma = 0$. Since $C(\theta) = \sum_{x=0}^{\infty} q(x)\theta^x$, by (14) and the choice of γ_1, we have

$$\begin{aligned} &\|\chi_n \bar{h}_{1n}\|_p \\ \le\ & c_n^{-\gamma_1} \sum_{x=0}^{\infty} a_n^{x+1} (xp/(p-1) + 1)^{-(p-1)/p} (px + p)^{-1/p} \le c_n^{-\gamma_1} \|\chi_n/C\|_1 = O(1). \end{aligned}$$

So, $\|\chi_n \bar{h}_{jn}\|_p = O(1)$ in all the cases. The proof is completed by (52) and $|\xi_{0n}(a_*)/C(a_*)| \le \int G(a_* - y/c_n)|k(y)|dy \le \int_{-\infty}^{a_* c_n} |y/(c_n a_*)|^{\alpha+1}|k(y)|dy = O(c_n^{-\alpha-1})$.

Appendix II. Proofs of the Lower Bounds

The proofs of the following two lemmas are deferred to the end of Appendix II.

Lemma 3. *Let $h_{u,v}$ be given by (27) with $u/v = a$. Then as $u \to \infty$, we have $\|\theta^x h_{u,v}(\theta)\|_\infty \approx a^{x-1}\sqrt{u}/\sqrt{2\pi}$ for $x \ge 0$, $\|h_{u,v}\|_p \approx \{u/(2\pi a^2)\}^{(p-1)/(2p)} p^{-1/(2p)}$, and $\|h^{(j)}_{u,v}\|_p/\|h_{u,v}\|_p \approx a^{-j} u^{j/2}\{E|Q_j(p^{-1/2}Z)|^p\}^{1/p}, \forall j \ge 0$ for $1 \le p < \infty$, where Z is a $N(0,1)$ random variable and $Q_j(x)$ are polynomials such that $Q_0(x) = 1$ and $Q_{j+1}(x) = xQ_j(x) - (d/dx)Q_j(x)$. In addition, there exist constants c_j^* such that $|h^{(j)}_{u,v}(\theta)|/h_{u,v}(\theta) \le c_j^* \theta^{-j}\{1 + |u - 1 - v\theta|^j + (v\theta)^{j/2}\}$ for all $j \ge 0$. If $u = u_n$ and $v = v_n$ as in (30) and (31), then $|h^{(j)}_{u,v}(\theta_1)| + |h^{(j)}_{u,v}(\theta_2)| = O(n^{-1}u^{j+1/2})$ and $\|h_{u,v}(1 - \chi_1)\|_p = O(1)u^{(p-1)/(2p)}/n$, $1 \le p \le \infty$.*

Remark. $Q_j(x) = \sum_{0 \le l \le j/2}\{j!(-1)^l x^{j-2l}\}/\{(j-2l)!l!2^l\}$ by mathematical induction.

Lemma 4. *There exists a constant C^* such that $\|l^{(m)}_{1,u,v}\chi_0\|_p \le C^* \sum_{j=0}^{\alpha'} |h^{(j)}_{u,v}(\theta_1)|$ and $\|l^{(m)}_{2,u,v}\chi_2\|_p \le C^* \sum_{j=0}^{\alpha'} |h^{(j)}_{u,v}(\theta_2)|$ for all $u > 0$ and $v > 0$. If $u = u_n$ and $v = v_n$ as in (30) and (31), then $\|l^{(m)}_{1,u,v}\chi_0\|_p + \|l^{(m)}_{2,u,v}\chi_2\|_p = O(n^{-1}u_n^{\alpha'+1/2})$.*

Proof of Theorem 3. We drop the subscript n in u_n and v_n throughout the proof. Part (i) is proved in Steps 1-3, while Part (ii) is proved in Step 4.

Step 1. Verify the membership of g_{jn} in $\mathcal{G}_{\alpha,\theta^*}(p, 1, M)$. By Lemma 4, $\|(\chi_0 + \chi_2)g_{u,v}\|_\infty = O(n^{-1}u^{\alpha'+1/2})$. Also, we have by (29),

$$1/C(\theta_2) + o(1) \ge w_{0n} = \int_{\theta_0}^{\theta_3} g_{u,v}(y)dy \ge 1/C(\theta_1) + o(1). \tag{53}$$

By (29) and Lemma 4, $(\chi_0 + \chi_2)g_{0n} \ge 0$ for small w_0 and $(p-1)/(2p) + \alpha \ge 0$. It follows that g_{0n} is a density. In the same manner, we find $|w_{jn}| \le w_{0n} + o(1)$, so that by (30) and (31) g_{1n} and g_{2n} are all density functions.

It remains to verify (13). By the smoothness of $C(\theta)$ on $[\theta_0, \theta_3]$ and Lemmas 3 and 4, we have $u^{-(p-1)/(2p)}\|g^{(m)}_{u,v}\|_p = O(u^{m/2}) = O(u^m)$ for $m = \alpha', \alpha' + 1$, which implies

$$u^{-(p-1)/(2p)-\alpha}\|g^{(\alpha')}_{u,v}(\theta) - g^{(\alpha')}_{u,v}(\theta + \delta)\|_p = O(1)\min(u^{-\alpha''}, u^{1-\alpha''}\delta).$$

Since $g_0 \in \mathcal{G}_{\alpha,\theta^*}(p, 1, M - \epsilon_1)$ for some $\epsilon_1 > 0$ and $\min(u^{-\alpha''}, u^{1-\alpha''}\delta) \le \delta^{\alpha''}$,

$$\|g^{(\alpha')}_{0n}(\theta) - g^{(\alpha')}_{0n}(\theta + \delta)\|_p \le (M - \epsilon_1 + O(1)w_0)\delta^{\alpha''} \tag{54}$$

by (29). This implies (13) with $g = g_{0n}$ for small w_0. Since $\|(d/d\theta)^j h_0(u(\theta - a)/\theta_2)\|_\infty = (u/\theta_2)^j$ for $h_0(y) = \sin(y)$ and $h_0(y) = \cos(y)$, we also have for $m = \alpha', \alpha'+1$, and $j = 1, 2$,

$$\left\|g_{jn}^{(m)} - g_{0n}^{(m)}\right\|_p \le O(1)\frac{w_0\theta_2^\alpha}{u^{(p-1)/(2p)+\alpha}}\sum_{l=0}^m \|g_{u,v}^{(l)}\|_p\left\{1 + (u/\theta_2)^{(m-l)}\right\} = w_0 O(u^{m-\alpha})$$

so that $\|(g_{jn} - g_{0n})^{(\alpha')}(\theta) - (g_{jn} - g_{0n})^{(\alpha')}(\theta+\delta)\|_p = O(1)w_0\delta^{\alpha''}$. Therefore, (13) holds with $g = g_{1n}$ and g_{2n} for small w_0.

Step 2. Next we show that

$$\sum_{x=0}^\infty |f(x; g_{jn}) - f(x; g_{0n})| = \sum_{x=0}^\infty q(x)\Big|\int_0^{\theta^*}\theta^x C(\theta)\{g_{jn}(\theta) - g_{0n}(\theta)\}d\theta\Big| = o(n^{-1}). \tag{55}$$

We prove this only for g_{2n}. As in Zhang (1995) and by the definition of δ_0

$$\sum_{x=0}^\infty q(x)\Big|\int_{\theta_1}^{\theta_2}\cos(u(\theta - a)/\theta_2)\theta^x h_{u,v}(\theta)d\theta\Big| = O(1/n). \tag{56}$$

Set $l_{j,u,v,x}(\theta) = \theta^x l_{j,u,v}(\theta)$. By the definition of $l_{j,u,v}$ and Lemmas 3 and 4 we have

$$\begin{aligned}|l_{1,u,v,x}^{(m)}(\theta_1)| + |l_{2,u,v,x}^{(m)}(\theta_2)| &= O(1)\sum_{j=0}^m (x+1)^j[\theta_1^{x-j}|h_{u,v}^{(m-j)}(\theta_1)| + \theta_2^{x-j}|h_{u,v}^{(m-j)}(\theta_2)|] \\ &= O(n^{-1}u^{m+1/2})(x+1)^m\theta_2^x, \quad m \ge 0,\end{aligned}$$

and by Lemma 4 $\int[|l_{1,u,v,x}^{(\alpha'+1)}(\theta)|\chi_0(\theta) + |l_{2,u,v,x}^{(\alpha'+1)}(\theta)|\chi_2(\theta)]d\theta = O(n^{-1}u^{\alpha'+1/2})(x+1)^{\alpha'+1}\theta_3^x$. Integrating by parts $\alpha'+1$ times, we obtain

$$\begin{aligned}&\sum_{x=0}^\infty q(x)\Big|\int\cos(u(\theta - a)/\theta_2)\theta^x l_{2,u,v}(\theta)\chi_2(\theta)d\theta\Big| \\ \le\ &\sum_{x=0}^\infty q(x)(\theta_3/u)^{\alpha'+1}\int|l_{2,u,v,x}^{(\alpha'+1)}(\theta)|\chi_2(\theta)d\theta + \sum_{x=0}^\infty q(x)\sum_{j=0}^{\alpha'}(\theta_3/u)^{j+1}|l_{2,u,v,x}^{(j)}(\theta_2)| \\ \le\ &O(n^{-1}u^{-1/2})\sum_{x=0}^\infty q(x)(x+1)^{\alpha'+1}\theta_3^x = o(n^{-1})\end{aligned}$$

and the same with respect to $l_{2,u,v}(\theta)\chi_2(\theta)$, so that by (28) and (56)

$$\sum_{x=0}^\infty q(x)\Big|\int_0^{\theta_3}\cos(u(\theta - a)/\theta_2)\theta^x C(\theta)g_{u,v}(\theta)d\theta\Big| = O(n^{-1}). \tag{57}$$

Since $\sum_x q(x)\theta^x C(\theta) = 1$, (53) and (57) imply

$$\|g_{u,v}\|_1 \Big|\frac{w_{2n}}{w_{0n}}\Big| = \frac{\|g_{u,v}\|_1}{|w_{0n}|}\Big|\int_0^{\theta_3} \cos(u(\theta-a)/\theta_2) g_{u,v}(\theta) d\theta\Big| = O(n^{-1}). \qquad (58)$$

Thus, by (31) the left-hand side of (55) is $o(n^{-1})$ for $j = 2$, as it is bounded by the product of $w_0\theta_2^\alpha u^{-(p-1)/(2p)-\alpha}$ and the sum of (57) and (58).

Step 3. Verify (26) and prove Part (i). In view of (55) and the existence of $M^* < \infty$ such that $\|g_{jn} - g_{0n}\|_p \le w_0 M^* u_n^{-\alpha}$ for $j = 0, 1, 2$, we only need to show

$$\liminf_{n\to\infty} (\log n)^\alpha \max_{j=1,2} \|g_{jn} - g_{0n}\|_p > 2\varepsilon_0, \quad \text{for some } \varepsilon_0 > 0. \qquad (59)$$

By Lemma 3 there exists a positive constant δ_1 such that for large n, $C(0)\,\|g_{u,v}\|_p \ge \|h_{u,v}\|_p - \|(1-\chi_1)h_{u,v}\|_p \ge \delta_1 u^{(p-1)/(2p)}$. Since $\max\{|\sin(x)|, |\cos(x)|\} \ge 1/\sqrt{2}$, by (30) and (31)

$$w_0^{-1}(\frac{u}{\theta_2})^\alpha \{\|g_{1n} - g_{0n}\|_p^p + \|g_{2n} - g_{0n}\|_p^p\}^{1/p} \ge \{\delta_1/C(0)\}\Big(1/\sqrt{2} - \max_{j=1,2} |\frac{w_{jn}}{w_{0n}}|\Big),$$

which implies (59), as $w_{jn}/w_{0n} \to 0$ and $u = \delta_0 \log n$.

Step 4. Prove Part (ii). Let $G_{u,v}$ and G_{jn} be the integrals of $g_{u,v}$ and g_{jn} respectively. If $\alpha > 0$, then $g_{jn} \in \mathcal{G}_{\alpha,\theta^*} \subseteq \mathcal{G}^{cdf}_{\alpha,\theta^*}$ by Step 1. For $\alpha = 0$, we have $\|G_{u,v}(\theta) - G_{u,v}(\theta+\delta)\|_p \le \delta \|g_{u,v}\|_p$, so that $g_{jn} \in \mathcal{G}^{cdf}_{0,\theta^*}$ for small w_0 by Lemma 3. By Steps 2 and 3, Part (ii) holds if

$$\liminf_{n\to\infty} (\log n)^{\alpha+1} \max_{j=1,2} \|G_{jn} - G_{0n}\|_p > 2\varepsilon_0. \qquad (60)$$

The difference $G_{2n} - G_{0n}$ is proportional to $\int_0^\theta \cos(u(y-a)/\theta_2) g_{u,v}(y) dy$, which can be expressed via three times of integrating by parts by

$$\sum_{j=1}^{3} (-\theta_2/u)^j \cos(u(\theta-a)/\theta_2 - j\pi/2) g_{u,v}^{(j-1)}(\theta) + (\theta_2/u)^3 \int_0^\theta \sin(u(y-a)/\theta_2) g_{u,v}^{(3)}(y) dy.$$

It follows from (30), (31), (58), and Lemmas 3 and 4 that $\|G_{2n} - G_{0n}\|_p = (\theta/u)(1+O(u^{-1/2}))\|g_{1n} - g_{0n}\|_p$, and likewise $\|G_{1n} - G_{0n}\|_p = (\theta/u)(1+O(u^{-1/2}))$ $\|g_{2n} - g_{0n}\|_p$. Hence, (60) follows from (59).

Proof of Lemma 3. The approximations for $\|\theta^x h_{u,v}(\theta)\|_\infty$ and $\|h_{u,v}\|_p$ follow from the Stirling formula. Define $Q_j^*(x,y)$ by $h_{u,v}^{(j)}(\theta) = h_{u,v}(\theta)\theta^{-j} Q_j^*(u-1-v\theta, \sqrt{v\theta})$. Clearly $Q_0^* = 1$. Since $(\partial/\partial\theta)\log h_{u,v}(\theta) = (u-1)/\theta - v$, $\theta^{-j-1}Q_{j+1}^*(u-1-v\theta, \sqrt{v\theta})$ equals

$$[(u-1)/\theta - v]\theta^{-j}Q_j^* - j\theta^{-j-1}Q_j^* + \theta^{-j}\Big[-vQ_{j,1}^* + \sqrt{v\theta}Q_{j,2}^*/(2\theta)\Big],$$

where $Q^*_{j,1} = (\partial/\partial x)Q^*_j$ and $Q^*_{j,2} = (\partial/\partial y)Q^*_j$. It follows that $Q^*_{j+1} = xQ^*_j - jQ^*_j - y^2 Q^*_{j,1} + (y/2)Q^*_{j,2}$, so that $Q^*_j(x,y)$ is a polynomial of degree j. This gives the inequality for $|h^{(j)}_{u,v}(\theta)|$. For the $\|\cdot\|_p$ norm, we have $|h^{(j)}_{u,v}|^p/\|h_{u,v}\|_p^p = |\theta^{-j}Q^*_j|^p h_{p(u-1)+1,pv}$. Since $h_{u,v}$ has mean u/v and variance u/v^2, by the moment convergence in the central limit theorem and law of large numbers

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

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

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

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.

If $u = u_n$ and $v = v_n$ as in (30) and (31), then $\|h_{u,v}(1-\chi_1)\|_1 = O(1/n)$ by the standard large deviation formula. The rest follows, since by the expression for δ_0 and the Stirling formula we have $h_{u,v}(\theta)[1-\chi_1(\theta)] = O(1)u^{1/2}/n$.

Proof of Lemma 4. Define $\|Q\|_0 = \sum_{j=0}^{\alpha'}\{|Q^{(j)}(\theta_0)| + |Q^{(j)}(\theta_1)|\}$. Since $\|\cdot\|_0$ is a norm for the $(2\alpha'+2)$-dimensional space of all polynomials Q of degree $2\alpha'+1$ on $[\theta_0,\theta_1]$, it is equivalent to all other norms on this linear space. This implies $\|l^{(m)}_{1,u,v}\chi_0\|_p \leq C^*\sum_{j=0}^{\alpha'}|h^{(j)}_{u,v}(\theta_1)|$ as $l^{(j)}_{1,u,v}(\theta_0) = 0$ for $0 \leq j \leq \alpha'$. The proof of the inequality with respect to $\|l^{(m)}_{2,u,v}\chi_2\|_p$ is the same. The rest follows from Lemma 3.

Appendix III. Proofs of the Upper Bounds for $\theta^* = \infty$

Proof of Theorem 4. Clearly, $\|w(\theta)g(\theta)I\{\theta > a^2_{1n}\}\|_p = O(c_n^{-\alpha})$ by (39) and the choice of a_{1n}. Also, as in Zhang (1995) $\|K_{s,n}(x,\cdot)\|_\infty \leq 2n^{\beta_0}B_0\|k\|_1/(\pi B)$. It follows from (37) and the argument in the proof of Theorem 1 that

$$E_g\|w(\hat{g}_{n,\infty} - E_g\hat{g}_{n,\infty})\|_p = O(1)n^{\beta_0-1/2}\max_{a_{0n}\leq\eta\leq a_{1n}} 1/C_s(\eta) = O(1)n^{\beta_0-1/2}c_n^{\max(1,\alpha/\alpha_1)},$$

so that $E_g\|w(\hat{g}_{n,\infty} - g)\|_p \leq \|I\{0 \leq \theta \leq a^2_{1n}\}w(E_g\hat{g}_{n,\infty} - g)\|_p + O(c_n^{-\alpha})$.

For the rest of the proof, and unless otherwise specified, we write everything as functions of $\eta = \sqrt{\theta}$, for which w_s is the actual weight function. Set $\chi_{0n} = I\{0 \leq \eta < a_{0n}\}$, $\chi_{1n} = I\{a_{0n} \leq \eta < 1\}$, and $\chi_{2n} = I\{1 \leq \eta \leq a_{1n}\}$. By (47) and (38) we have $\|w_s g_s \chi_{0n}\|_p = O(c_n^{-\alpha})$, so that $E_g\|w(\theta)(\hat{g}_{n,\infty}(\theta) - g(\theta))\|_p \leq \|w_s(\chi_{1n}+\chi_{2n})(b_{3n}+b_{4n})/C_s\|_p + O(c_n^{-\alpha})$, where $b_{jn}(\eta)$ are given by (35) and (36).

For $j = 1, 2$, define $\xi_{jn}(\eta, y) = \sum_{l=1}^{3} \xi_l(\eta, -y/c_n; g_s, C_s, \alpha, \gamma_j, \rho_j)$, where ξ_j are as in (49), $\rho_j = \alpha + \gamma_j$, $\gamma_1 = 1$, and $\alpha/\alpha_1 < \gamma_2 \leq \rho - \alpha$. Then, by (35) $\chi_{jn} b_{3n}(\eta) = \chi_{jn} \int \xi_{jn}(\eta, y) k(y) dy$, $j = 1, 2$. As in the proof of Lemma 1, we have $\|w_s \chi_{jn} b_{3n}/C_s\|_p = O(c_n^{-\alpha})$, $j = 1, 2$. Note that on the set $[0, 1]$ (14) is replaced by $\eta/C_s(\eta) \leq 1/C(1)$ with $\gamma = \gamma_1 = 1$. Also note that $\|\chi_{2n} C_s^{(j)}/C_s\|_\infty = o(c_n^{\epsilon_1})$ and $a_{1n} = o(c_n^{\epsilon_1})$ for all small $\epsilon_1 > 0$, while $\gamma = \gamma_2 > \alpha/\alpha_1$.

The proof of $\|w_s \chi_{jn} b_{4n}/C_s\|_p = O(c_n^{-\alpha})$ is similar and is omitted. Note that (38) implies $\|I\{\eta > 0\} w_s(\eta) g_s^{(\alpha')}(\delta - \eta)\|_p < M|\delta|^{\alpha''}$ for all δ.

Proof of Theorem 5. We combine the methods in the proofs of Theorem 4 and Lemma 2 with the (g, C) in (49) replaced by $(G_s(y), C_{s,0}^{(j)}(y))$ in the case of B_{3n} and by $(G_s(-y), C_{s,0}^{(j)}(-y))$ in the case of B_{4n}, $j = 0, 1$. This gives

$$E_G\|w(\hat{G}_{n,\infty} - G)\|_p \leq O(c_n^{-\alpha}) + O(1)\Big|\int_0^\infty C((y/c_n)^2) G((y/c_n)^2) k(y) dy\Big|.$$

The proof is now complete, as the integration on the right-hand side is bounded in absolute value by $C(0) M_2 \int_0^\infty (y/c_n)^{\alpha+1} |k(y)| dy = O(c_n^{-\alpha-1})$ due to (45).

References

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

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

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

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

Lambert, D. and Tierney, L. (1984). Asymptotic properties of maximum likelihood estimates in the mixed Poisson model. *Ann. Statist.* **12**, 1388-1399.

Loh, W. L. (1993). Estimating the mixing density of a mixture of power series distributions. *Statist. Decision Theory and Related Topics V*, (Edited by S. S. Gupta and J. O. Berger), 87-98. Springer, New York.

Loh, W. L. and Zhang, C.-H. (1993). Global properties of kernel estimators for mixing densities in exponential family models for discrete variables. Technical Report, Statist. Dept., Purdue University.

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

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.

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.

Zhang, C.-H. (1995). On estimating mixing densities in discrete exponential family models. *Ann. Statist.* **23**, 929-945.

Department of Statistics, Purdue University, West Lafayette, IN 47907, U.S.A.

Department of Statistics, Hill Center, Busch Campus, Rutgers University, New Brunswick, NJ 08903, U.S.A.

(Received December 1993; accepted November 1995)