Sankhyā: The Indian Journal of Statistics
1999, Volume 61, Series A, Pt. 1, pp. 101-112

UNBIASED ESTIMATORS OF A LATTICE MIXING DISTRIBUTION AND THE CHARACTERISTIC FUNCTION OF A GENERAL MIXING DISTRIBUTION*

By CUN-HUI ZHANG
Rutgers University, Piscataway

SUMMARY. 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 $f(x) = \int f(x|\theta)dG(\theta)$. We derive unbiased estimators of the characteristic functions of the mixing distribution G under some integrability conditions on G and the probability mass function of G when G is a lattice distribution. Upper bounds for the variances of these unbiased estimators are provided. Three types of exponential families and a location-type model are considered, including the Poisson and gamma families.

1. **Introduction**

Let (X, θ) be a random vector such that

$$X|\theta \sim f(x|\theta), \quad \theta \sim G, \qquad \ldots (1)$$

where the conditional density $f(x|\theta)$ belongs to a known parametric family of probability density functions with respect to a σ-finite measure μ, and G is an unknown mixing distribution. Let (X_j, θ_j), $1 \le j \le n$, be independent identically distributed random vectors from (X, θ). Suppose $\theta_1, \ldots, \theta_n$ are latent variables. We are interested in the demixing problem of estimating functionals of G based on observations $X_1, \ldots, X_n$ from the mixture density

$$f(x) = \int f(x|\theta)dG(\theta). \qquad \ldots (2)$$

When $f(x|\theta)$ is a location family and μ is the Lebesgue measure, this is also called the deconvolution problem.

Paper received. April 1994; revised July 1998.
AMS (1991) *subject classification.* Primary 62G05; secondary 62G20.
Key words and phrases. Mixture, mixing distribution, unbiasedness, Fourier transformation.
* This research was partially supported by the National Science Foundation, the Army Research Office and the National Security Agency.

Problems related to mixture models were proposed by Robbins (1951, 1955) in connection with the empirical Bayes approach to compound decision problems, by Kiefer-Wolfowitz (1956) in connection with estimating an unknown parameter in the presence of infinitely many nuisance parameters, and by many others in various contexts.

Maximum likelihood estimation (MLE) of mixing distributions were studied by Kiefer and Wolfowitz (1956), Laird (1978), Lambert and Tierney (1984), Lindsay (1983a,b), and Simar (1976) among others. Although the MLEs are consistent under quite general conditions, their rate of convergence is still unknown, unless we assume that G belongs to a finite-dimensional parametric family. Recently, there were a spate of papers on the subject which provided rates of convergence under smoothness assumptions on G. Minimum distance estimation for the normal case was considered by Edelman (1988). Kernel estimation in location models was considered by Carroll and Hall (1988), Fan (1991a,b), Stefanski (1990), Stefanski and Carroll (1990), and Zhang (1990) among others. Kernel estimation in the case of discrete exponential families was considered by Loh and Zhang (1996, 1997) and Zhang (1995), while Walter and Hamedani (1991) considered estimation based on orthogonal polynomial expansions. Most of these rates of convergence are extremely slow. In fact, except for a few location families such as the double exponential distribution, these estimators converge to the true G at logarithmic rates. Mixture models were also considered by Deely and Kruse (1968), Devroye (1989), Jewell (1982), Robbins (1950), Rolph (1968), Teicher (1961) and Wise, Traganitis and Thomas (1977) among others.

We shall consider four types models: (A) gamma-type model $f(x|\theta) = C(\theta)q(x)f_\alpha(x|\theta)$ with $f_\alpha(x|\theta)$ being the gamma$(\alpha+1, 1/\theta)$ density function; (B) Poisson-type model $f(x|\theta) = C(\theta)q(x)p(x|\theta)$ with $p(x|\theta)$ being the Poisson probability mass function; (C) discrete exponential model $f(x|\theta) = C(\theta)q(x)\theta^x$, $x = 0, 1, 2, \ldots$; and (D) location-type model $f(x|\theta) = C(\theta)q(x)f_0(x-\theta)$ for some known f_0. Here μ is the Lebesgue measure in models (A) and (D) and the counting measure in models (B) and (C). In Section 2, we derive unbiased estimators of the Fourier transformation

$$h^*_C(t) = \int e^{it\theta} C(\theta) dG(\theta) \qquad \ldots (3)$$

for general G and unbiased estimators of the mixing probability mass function when G is a lattice distribution. Clearly, $h^*_C(t)$ becomes the characteristic function of G when $C(\theta) \equiv q(x) \equiv 1$. Upper bounds for the variance of the unbiased estimators are provided under certain integrability conditions on G, so that the unbiased estimators are asymptotically normal with the usual $n^{-1/2}$ rate of convergence.

Estimation of the characteristic function of the mixing distribution G is often an important step in demixing problems. In the location model $f(x|\theta) = f_0(x-\theta)$, unbiased estimates of (3) with $C(\theta) \equiv 1$ was used in Stefanski and

Carroll (1990) and Zhang (1990) to derive estimates of the mixing density, while Zhang (1995) used asymptotically unbiased kernels for (3) in model (C) described above. In both instances, the resulting estimates of the mixing density possess optimal rates of convergence under smoothness conditions. The unbiased estimators here for the gamma- and Poisson-type models are new. Due to the Fourier inversion formula, estimates of (3) can be used in many ways to derive estimators of G or its density. Here we consider specifically lattice mixing distributions. Due to computational difficulties with the MLEs, the mixing distribution is often discretized in practice. The MLE can be easily computed using the EM and other algorithms for lattice G, but convergence rates for the MLE are still unknown when the support of G is unbounded. The methods here provide simple estimates with tractable properties.

2. **Unbiased Estimators**

In this section, we derive unbiased estimators for the Fourier transformation $h^*_C(t)$ in (3). From these estimators for $h^*_C(t)$, we derive unbiased estimators for the mixing probability mass function $p(a) = G(a) - G(a-)$ under the additional condition

$$\sum_{j=-\infty}^{\infty} p(a_j) = 1, \quad a_j = a_0 + j\delta, \qquad \ldots (4)$$

for some known $\delta > 0$ and a_0. Upper bounds for the variance of these unbiased estimators are also provided.

Let (X, θ) be as in (1). A function $K^*_C(x,t)$ is an unbiased kernel for $h^*_C(t)$ if

$$EK^*_C(X,t) = h^*_C(t) = \int e^{it\theta} C(\theta) dG(\theta) \qquad \ldots (5)$$

under certain integrability conditions. This provides an unbiased estimator $\widehat{h^*}_{C,n}(t) = \sum_{k=1}^n K^*_C(X_k,t)/n$ for $h^*_C(t)$ based on observations $X_1, \ldots, X_n$. If $E|K^*_C(X,t)|^2 < \infty$ in a set T, then the finite dimensional distributions of $\sqrt{n}\big(\widehat{h^*}_{C,n} - h^*_C\big)(t), t \in T$, converge to multivariate normal distributions with the covariance kernel

$$nE\big(\widehat{h^*}_{C,n} - h^*_C\big)(s)\overline{\big(\widehat{h^*}_{C,n} - h^*_C\big)(t)} = V_C(s,t) - h^*_C(s)\overline{h^*_C(t)}, \qquad \ldots (6)$$

where $V_C(s,t) = EK^*_C(X,s)\overline{K^*_C(X,t)}$ and $\bar{z}$ is the complex conjugate of z. Since $h^*_C(t) = \overline{h^*_C(-t)}$, $\big\{K^*_C(x,t) + \overline{K^*_C(x,-t)}\big\}/2$ is an unbiased kernel for $h^*_C(t)$ with smaller variance than $K^*_C(x,t)$. Thus, we shall consider kernels $K^*_C(x,t)$ satisfying $K^*_C(x,t) = \overline{K^*_C(x,-t)}$ in the rest of the paper.

Under condition (4), the Fourier inversion formula provides

$$C(a_j)p(a_j) = \frac{\delta}{2\pi} \int_{-\pi/\delta}^{\pi/\delta} e^{-ita_j} h^*_C(t) dt, \qquad \ldots (7)$$

where h_C^* is as in (3). Thus, if $K_C^*(x,t)$ satisfies (5) for $|t| \le \pi/\delta$,

$$\widehat{p}_n(a_j) = \sum_{k=1}^n \frac{K(X_k, a_j)}{n}, \quad K(x,a) = \frac{\delta}{2\pi C(a)} \int_{-\pi/\delta}^{\pi/\delta} e^{-ita} K_C^*(x,t)dt, \quad \ldots (8)$$

is an unbiased estimator for $p(a_j)$. It follows from (6) that

$$\begin{aligned} &nE\big(\widehat{p}_n(a_j) - p(a_j)\big)^2 \\ &\quad = \Big(\frac{\delta}{2\pi C(a_j)}\Big)^2 \int_{-\pi/\delta}^{\pi/\delta} \int_{-\pi/\delta}^{\pi/\delta} e^{-i(s-t)a_j} V_C(s,t) ds dt - p^2(a_j). \end{aligned} \quad \ldots (9)$$

Note that $K(x,a)$ is real by $K_C^*(x,t) = \overline{K_C^*(x,-t)}$. Furthermore, since $C(a_j)$ $\big\{K(X_k,a_j) - p(a_j)\big\}$ are the Fourier coefficients of $\sqrt{\delta/2\pi}\,\{K_C^*(X_j,t) - h_C^*(t)\}$ with the orthonormal basis $\{\sqrt{\delta/2\pi} e^{-ita_j}\}$ in $L^2(-\pi/\delta, \pi/\delta)$, the Parseval identity provides, with the $V_C(s,t)$ in (6),

$$\begin{aligned} &n \sum_{j=-\infty}^{\infty} C^2(a_j) E\big(\widehat{p}_n(a_j) - p(a_j)\big)^2 \\ &\quad = \frac{\delta}{2\pi} \int_{-\pi/\delta}^{\pi/\delta} E\left|K_C^*(X,t) - h_C^*(t)\right|^2 dt \\ &\quad = \frac{\delta}{\pi} \int_0^{\pi/\delta} V_C(t,t)dt - \frac{\delta}{\pi} \int_0^{\pi/\delta} \left|h_C^*(X,t)\right|^2 dt. \end{aligned} \quad \ldots (10)$$

In the following four subsections, we consider the four types of mixture models described in the introduction. We provide unbiased kernels and upper bounds for their variances. In models (A), (B) and (C), our kernels for (3) are derived from unbiased estimates of the moments $E\{\theta^k C(\theta)\}$, since

$$h_C^*(t) = \int e^{it\theta} C(\theta) dG(\theta) = \sum_{k=0}^{\infty} \frac{(it)^k}{k!} E\{\theta^k C(\theta)\} \quad \ldots (11)$$

when the infinite series converges under the expectation. This is related to the moment problem (Feller, 1971, pages 227-228 and 514-515). Upper bounds for the covariance kernel (6) will be provided. In this paper, $V_1(s,t) \le V_2(s,t)$ if $V_2(s,t) - V_1(s,t)$ is nonnegative definite.

We shall use the following notation: $\beta! = \Gamma(\beta+1)$, $\beta^{[k]} = \beta(\beta-1)\cdots(\beta-k+1)$ with $\beta^{[0]} = 1$, and $(\beta)_k = \beta(\beta+1)\cdots(\beta+k-1)$ with $(\beta)_0 = 1$. Note that $\beta^{[k]} = (-1)^k(-\beta)_k$ for $k \le \beta+1$ and $\binom{n}{k} = (-1)^k(-n)_k/k!$. Also, let $x^+ = \max(x,0)$.

A. Gamma-type models. Consider the gamma-type mixture model with

$$f(x|\theta) = C(\theta) q(x) f_\alpha(x|\theta), \ f_\alpha(x|\theta) = \frac{x^\alpha e^{-x/\theta}}{\theta^{\alpha+1}\Gamma(\alpha+1)} I_{\{x>0\}}, \quad \ldots (12)$$

and $d\mu = dx$ in (1) and (2), where $q(x) > 0$, $C(\cdot)$ and $\alpha > -1$ are known, and $\theta > 0$.

Let $E_{\alpha,\theta}$ be the expectation with respect to the density $f_\alpha(x|\theta)$. Since $E(X^k/q(X)|\theta) = C(\theta)E_{\alpha,\theta}X^k = C(\theta)\theta^k(1+\alpha)_k$, by (11) an unbiased kernel for $h_C^*(t)$ is

$$K_C^*(x,t) = \widetilde{K}_\alpha(x,t)/q(x), \quad \widetilde{K}_\alpha(x,t) = \sum_{k=0}^\infty \frac{(it)^k x^k}{k!(1+\alpha)_k}. \qquad \ldots(13)$$

Let $\psi_\alpha(\lambda,p) = E_\lambda\{(N_1+1+\alpha)/(1+\alpha)\}^p I\{N_1 = N_2\}$, where N_1 and N_2 are independent Poisson variables with $E_\lambda N_1 = E_\lambda N_2 = \lambda$. Let $F(a,b,c;z) = \sum_{m=0}^\infty \{(a)_m(b)_m/(c)_m\}z^m/m!$ be the hypergeometric function (Abramowitz and Stegun, 1964, 15.1.1), and set

$$V_{\alpha,r}(s,t|\theta) = \sum_{\ell=0}^\infty \frac{(is\theta)^\ell(1+\alpha+r)_\ell}{\ell!(1+\alpha)_\ell} F(-\ell,-\ell-\alpha,1+\alpha;-t/s).$$

THEOREM 1. *Let $K_C^*(x,t)$ be given by* (13) *and $\widehat{p}_n(a_j)$ by* (8).

(i) If $EC(\theta)e^{|t|\theta} < \infty$, then (5) *holds. If $EC(\theta)e^{\pi\theta/\delta} < \infty$, then $E\widehat{p}_n(a_j) = p(a_j)$, $\forall j$.*

(ii) If $1/q(x) \le \int \frac{x^r\alpha!}{(\alpha+r)!}Q(dr)$ for some measure Q on $(-1-\alpha,\infty)$, then (6) *holds with*

$$V_C(s,t) = EK_C^*(X,s)\overline{K_C^*(X,t)} \le \int EC(\theta)\theta^r V_{\alpha,r}(s,t|\theta)Q(dr). \qquad \ldots(14)$$

Moreover, with $c_\alpha(p) = \max\{1, \prod_{j=1}^\infty[\{1+p/(j+\alpha)\}/\{1+1/(j+\alpha)\}^p]\}$,

$$\begin{aligned} V_{\alpha,r}(t,t|\theta) &= \sum_{\ell=0}^\infty \frac{(t\theta)^{2\ell}(1+\alpha+r)_{2\ell}}{\ell!(1+\alpha)_{2\ell}(1+\alpha)_\ell} \\ &\le c_\alpha(-\alpha)c_\alpha(r)2^{r^+}\psi_\alpha(|t|\theta, r-\alpha)e^{2|t|\theta}. \end{aligned} \qquad \ldots(15)$$

Consequently, (10) *holds with*

$$\begin{aligned} &\frac{\delta}{\pi}\int_0^{\pi/\delta} V_C(t,t)dt \\ &\le \int \frac{c_\alpha(-\alpha)c_\alpha(r)2^{r^+}}{\min\{1,2(1+\alpha)\}} EC(\theta)\theta^r\psi_\alpha(\pi\theta/\delta, r-\alpha-1)e^{2\pi\theta/\delta}Q(dr). \end{aligned} \qquad \ldots(16)$$

REMARK 1. For the gamma mixture $q(x) \equiv C(\theta) \equiv 1$, (14) becomes equality with $r=0$, $V_C(s,t) = EV_{\alpha,0}(s,t|\theta)$, and $\{(-it\theta)^\alpha/\Gamma(\alpha+1)\}V_{\alpha,0}(t,t|\theta) = J_\alpha(-2it\theta)$, where J_α is the Bessel function. The kernel $\widetilde{K}_\alpha(x,t)$ is related to J_α in a similar manner.

REMARK 2. It will be shown in the proof that $V_{\alpha,r}(s,t|\theta) = E_{\alpha+r,\theta} \widetilde{K}_\alpha(X,s)\overline{\widetilde{K}_\alpha(X,t)}$ with the $\widetilde{K}_\alpha(x,t)$ in (13). For fixed $\alpha > -1$ and p, $\psi_\alpha(\lambda,p)$ is of the order $\lambda^{p-1/2}$ for large λ. For $p \notin (0,1)$, $c_\alpha(p) = 1$.

REMARK 3. If $1/q(x) \le \sum_{j=0}^k c_j x^j$ for some $c_j \ge 0$, we may take $Q(j) - Q(j-) = c_j$. For example, $\alpha = 1$, $q(x) = (1+x)/x$ and $1/q(x) \le (1+x)/4$ when $X = (Y-1)$ conditionally on $Y > 1$ with $Y|\theta \sim f_1(x|\theta)$.

PROOF. Let $\widetilde{V}_{\alpha,r}(s,t|\theta) = E_{\alpha+r,\theta}\widetilde{K}_\alpha(X,s)\overline{\widetilde{K}_\alpha(X,t)}$. It follows from (12) and (13) that

$$V_C(s,t) = E\int \widetilde{K}_\alpha(x,s)\overline{\widetilde{K}_\alpha(x,t)}\frac{f(x|\theta)}{q^2(x)}dx \le E\int C(\theta)\theta^r \widetilde{V}_{\alpha,r}(s,t|\theta)Q(dr),$$

as $f(x|\theta)/q^2(x) \le \int C(\theta)\theta^r f_{\alpha+r}(x|\theta)Q(dr)$ by the upper bound on $1/q(x)$. Let $z = -t/s$. Since $E_{\alpha+r,\theta}(X^\ell|\theta) = \theta^\ell(1+\alpha+r)_\ell$, we have, with $\ell = k+m$,

$$\begin{aligned}\widetilde{V}_{\alpha,r}(s,t|\theta) &= \sum_{k=0}^\infty\sum_{m=0}^\infty E_{\alpha+r,\theta}\left\{\frac{(isX)^k}{k!(1+\alpha)_k}\right\}\left\{\frac{(-itX)^m}{m!(1+\alpha)_m}\right\}\\ &= \sum_{\ell=0}^\infty \frac{(is\theta)^\ell(1+\alpha+r)_\ell}{\ell!(1+\alpha)_\ell}\sum_{m=0}^\ell z^m\binom{\ell}{m}\frac{(1+\alpha)_\ell}{(1+\alpha)_m(1+\alpha)_{\ell-m}}\\ &= V_{\alpha,r}(s,t|\theta),\end{aligned}$$

due to $(1+c)_\ell/(1+c)_{\ell-m} = (-1)^m(-\ell-c)_m$ for $c=0$ and $c=\alpha$. This proves (14).

Since $\sqrt{\pi}/\Gamma(1/2-\ell) = (-1/2)^\ell(2\ell-1)!! = (-1/4)^\ell(2\ell)!/\ell!$,

$$F(-2\ell,-2\ell-\alpha,1+\alpha;-1) = \frac{2^{2\ell}\sqrt{\pi}\Gamma(1+\alpha)}{\Gamma(1+\alpha+\ell)\Gamma(1/2-\ell)} = \frac{(-1)^\ell(2\ell)!}{\ell!(1+\alpha)_\ell}$$

by Abramowitz and Stegun (1964, 15.1.21). Since $V_{\alpha,r}(t,t|\theta)$ is real, the identity in (15) is obtained by summing over the even ℓ in its infinite series representation and using the above formula. The inequality in (15) follows from

$$\begin{aligned}\frac{(1+\alpha+r)_{2\ell}}{(1+\alpha)_{2\ell}}\cdot\frac{\ell!}{(1+\alpha)_\ell} &= \prod_{j=1}^{2\ell}\left(1+\frac{r}{\alpha+j}\right)\prod_{j=1}^{\ell}\left(1+\frac{-\alpha}{\alpha+j}\right)\\ &\le c_\alpha(-\alpha)c_\alpha(r)\prod_{j=\ell+1}^{2\ell}\left(1+\frac{1}{\alpha+j}\right)^r\prod_{j=1}^{\ell}\left(1+\frac{1}{\alpha+j}\right)^{r-\alpha}\\ &\le c_\alpha(-\alpha)c_\alpha(r)2^{r^+}\left(\frac{\ell+\alpha+1}{1+\alpha}\right)^{r-\alpha}.\end{aligned}$$

Inequality (16) follows from (14) and (15), as

$$\frac{\delta}{\pi}\int_0^{\pi/\delta} t^{2\ell}dt\binom{\ell+\alpha+1}{1+\alpha} = \frac{(\pi/\delta)^{2\ell}}{2\ell+1}\binom{\ell+\alpha+1}{1+\alpha} \le \frac{(\pi/\delta)^{2\ell}}{\min\{1,2(1+\alpha)\}}.$$

B. Poisson-type models. Consider the Poisson-type model

$$f(x|\theta) = C(\theta)q(x)p(x|\theta), \quad p(x|\theta) = e^{-\theta}\theta^x/x!, \quad q(x) > 0, \qquad \ldots (17)$$

with μ being the counting measure on $\{0, 1, 2, \ldots\}$ and known $q(x) > 0$ and $C(\cdot)$. Let E_θ be the expectation with respect to the Poisson density $p(x|\theta)$. Since $E(X^{[k]}/q(X)|\theta) = C(\theta)E_\theta X^{[k]} = C(\theta)\theta^k$, by (11) an unbiased kernel for $h_C^*(t) = \int e^{it\theta}C(\theta)dG(\theta)$ is

$$K_C^*(x,t) = \widetilde{K}(x,t)/q(x), \quad \widetilde{K}(x,t) = \sum_{k=0}^{\infty}(it)^k x^{[k]}/k!. \qquad \ldots (18)$$

THEOREM 2. *Let K_C^* be given by* (18) *and $\widehat{p}_n(a_j)$ by* (8).

(i) If $EC(\theta)e^{|t|\theta} < \infty$, then (5) *holds. If $EC(\theta)e^{\pi\theta/\delta} < \infty$, then $E\widehat{p}_n(a_j) = p(a_j)$, $\forall j$.*

(ii) If $1/q(x) \le \int x^{[r]}Q(dr)$, $x \ge 0$, for some measure Q on the set of non-negative integers $\{r = 0, 1, 2, \ldots\}$, then (6) *holds with*

$$V_C(s,t) \le \int EC(\theta)e^{st\theta}e^{i\theta(s-t)}\theta^r(1+is)^r(1-it)^r Q(dr). \qquad \ldots (19)$$

Consequently, (10) *holds with*

$$\frac{\delta}{\pi}\int_0^{\pi/\delta} V_C(t,t)dt \le \int EC(\theta)\theta^r\left\{\frac{\delta}{\pi}\int_0^{\pi/\delta} e^{t^2\theta}(1+t^2)^r dt\right\}Q(dr). \qquad \ldots (20)$$

REMARK 1. For the Poisson mixture $q(x) \equiv C(\theta) \equiv 1$, (19) and (20) become equality with $r = 0$; $V_C(s,t) = Ee^{st\theta + i(s-t)\theta}$.

REMARK 2. For any complex measure ξ with bounded support

$$E\left|\int K_C^*(X,t)d\xi(t)\right|^2 \le E\int e^{-\theta}\sum_{k=0}^{\infty}\frac{\theta^{k+r}}{k!}\left|\int(1+it)^{k+r}d\xi(t)\right|^2 dQ(r).$$

This applies to (9) with $d\xi(t) = I\{|t| \le \pi/\delta\}(2\pi C(a_j))^{-1}\delta e^{-ita_j}dt$.

REMARK 3. For example, the upper bound on $1/q(x)$ holds for suitable Q when $1/q(x) = (1+c)^x = \sum_{j=0}^{\infty} x^{[j]}c^j/j!$, or $1/q(x) \le (x!)^\beta \le \sum_{j=0}^{\infty} x^{[j]}/(j!)^{1-\beta}$, $0 < \beta < 1$.

PROOF. With the upper bound on $1/q(x)$, (17) and (18) imply

$$V_C(s,t) = EC(\theta)E_\theta\frac{\widetilde{K}(X,s)\overline{\widetilde{K}(X,t)}}{q(X)} \le \int EC(\theta)E_\theta X^{[r]}\widetilde{K}(X,s)\overline{\widetilde{K}(X,t)}dQ(r).$$

Thus, (19) is a consequence of

$$E_\theta X^{[r]}\widetilde{K}(X,s)\widetilde{K}(X,-t)=\{\theta(1+is)(1-it)\}^r\exp\big\{st\theta+i(s-t)\theta\big\}. \quad \ldots(21)$$

Let us prove (21). By the mathematical induction for $k=0,\ldots,m$ we obtain

$$\beta^{[k]}\beta^{[m]}=\sum_{j=0}^{\min(k,m)}\beta^{[k+m-j]}j!\binom{k}{j}\binom{m}{j}=\sum_{j=0}^{k}\beta^{[k+m-j]}m^{[j]}\binom{k}{j},$$

as it holds for $k=0$ and implies for $k<m$

$$\begin{aligned}\beta^{[k+1]}\beta^{[m]} &= \sum_{j=0}^{k}\big\{\beta^{[k+1+m-j]}+(m-j)\beta^{[k+m-j]}\big\}m^{[j]}\binom{k}{j}\\ &= \sum_{j=0}^{k}\beta^{[k+1+m-j]}m^{[j]}\binom{k}{j}+\sum_{j=1}^{k+1}\beta^{[k+1+m-j]}m^{[j]}\binom{k}{j-1}\\ &= \sum_{j=0}^{k+1}\beta^{[k+1+m-j]}m^{[j]}\binom{k+1}{j}.\end{aligned}$$

Two applications of this identity provide that for all integers $r\ge 0$,

$$\beta^{[r]}\beta^{[k]}\beta^{[m]}=\sum_{j=0}^{\min(k,m)}j!\binom{k}{j}\binom{m}{j}\sum_{u=0}^{r}\beta^{[r+k+m-j-u]}(k+m-j)^{[u]}\binom{r}{u}. \quad \ldots(22)$$

Since $E_\theta X^{[\ell]}=\theta^\ell$, we obtain with $v=k-j$, $w=m-j$ and $\ell=v+w+j=k+m-j$,

$$\begin{aligned}&E_\theta X^{[r]}\widetilde{K}(X,s)\widetilde{K}(X,-t)\\ &=\sum_{k=0}^{\infty}\sum_{m=0}^{\infty}\big\{(is)^k/k!\big\}\{(-it)^m/m!\}\,E_\theta X^{[r]}X^{[k]}X^{[m]}\\ &=\sum_{k=0}^{\infty}\sum_{m=0}^{\infty}\sum_{j=0}^{\min(k,m)}\sum_{u=0}^{r}\frac{(is)^k(-it)^m\theta^{r+k+m-j-u}(k+m-j)^{[u]}}{j!(k-j)!(m-j)!}\binom{r}{u}, \text{ by (22)}\\ &=\sum_{u=0}^{r}\binom{r}{u}\theta^{r-u}\sum_{\ell=0}^{\infty}\ell^{[u]}\theta^\ell\sum_{v+w+j=\ell}\frac{(st)^j(is)^v(-it)^w}{v!w!j!}I_{\{v\ge0,w\ge0,j\ge0\}}\\ &=\sum_{u=0}^{r}\binom{r}{u}\theta^{r-u}\sum_{\ell=0}^{\infty}\ell^{[u]}\theta^\ell\frac{(st+is-it)^\ell}{\ell!}\\ &=\sum_{u=0}^{r}\binom{r}{u}\theta^{r-u}e^{st\theta+i\theta(s-t)}\{st\theta+i\theta(s-t)\}^u\\ &=e^{st\theta+i\theta(s-t)}\{\theta+st\theta+i\theta(s-t)\}^r.\end{aligned}$$

This proves (21) and completes the proof.

C. Exponential families for nonnegative integer-valued variables. Let

$$f(x|\theta) = C(\theta)q(x)\theta^x, \quad x = 0, 1, 2, \ldots, \qquad \ldots (23)$$

with known $q(x) > 0$ and $C(\cdot)$. This is the model considered in Zhang (1995). It is identical to (17) but with a different formulation. For example, if $f(x|\theta)$ is the Poisson distribution, then $C(\theta) \equiv q(x) \equiv 1$ in (17) but $C(\theta) = e^{-\theta}$ and $q(x) = 1/x!$ in (23). Since $P(X = k|\theta) = C(\theta)\theta^k/q(x)$, by (11) we define

$$K^*_C(x, t) = (it)^x/\{x!q(x)\}. \qquad \ldots (24)$$

THEOREM 3. *Let K^*_C be given by* (24) *and $\widehat{p}_n(a_j)$ by* (8).

(i) If $EC(\theta)e^{|t|\theta} < \infty$, then (5) *holds. If $EC(\theta)e^{\pi\theta/\delta} < \infty$, then $E\widehat{p}_n(a_j) = p(a_j)$, $\forall j$.*

(ii) If $1/q(x) \leq x! \int x^{[r]}Q(dr)$, $x \geq 0$, for some measure Q on $\{r = 0, 1, \ldots\}$, then (6) *holds with $V_C(s, t) \leq \int EC(\theta)e^{st\theta}(st\theta)^r Q(dr)$. Consequently,* (10) *holds with*

$$\frac{\delta}{\pi}\int_0^{\pi/\delta} V_C(t, t)dt \leq \int EC(\theta)\theta^r \left\{\frac{\delta}{\pi}\int_0^{\pi/\delta} e^{t^2\theta}t^{2r}dt\right\}Q(dr).$$

PROOF. The proof is simpler than that of Theorem 2;

$$V_C(s, t) = EC(\theta)\sum_{m=0}^{\infty}\frac{(st)^m\theta^m q(m)}{\{q(m)m!\}^2} \leq \int EC(\theta)\sum_{m=0}^{\infty} m^{[r]}\frac{(st\theta)^m}{m!}Q(dr).$$

EXAMPLE 1. (Geometric distribution). For this case, $q(x) \equiv 1$ and $C(\theta) \equiv 1 - \theta$, (6) holds with $V_C(s, t) = E(1 - \theta)e^{st\theta}$, and (10) holds with

$$\frac{\delta}{\pi}\int_0^{\pi/\delta} V_C(t, t)dt = E(1 - \theta)\frac{\delta}{\pi}\int_0^{\pi/\delta} e^{t^2\theta}dt.$$

D. Location-type models. Consider the case where

$$f(x|\theta) = C(\theta)q(x)f_0(x - \theta), \ \mu(dx) = dx, \ \theta \in (-\infty, \infty), \qquad \ldots (25)$$

with known $f_0(\cdot)$, $q(x) > 0$ and $C(\cdot)$. This is a location family when $q(x) \equiv C(\theta) \equiv 1$. The general $q(x)$ is useful in applications such as biased sampling. Let $f^*_0(z) = \int e^{izx}f_0(x)dx$. Since the Fourier transformation of $f_0(x - \theta)$ is the product $e^{it\theta}f^*_0(t)$, we define

$$K^*_C(x, t) = e^{itx}/\{f^*_0(t)q(x)\}. \qquad \ldots (26)$$

THEOREM 4. *Let K_C^* be given by* (26) *and* $\widehat{p}_n(a_j)$ *by* (8).

(i) If $|f_0^*(t)| > 0$, *then* (5) *holds. If* $\int_{|t|\le\pi/\delta} |f_0^*(t)|^{-1}dt < \infty$, *then* $E\widehat{p}_n(a_j) = p(a_j)$.

(ii) If $1/q(x) \le \int e^{rx}dQ(r)$ *for some measure* Q, *then* (6) *holds with*

$$V_C(s,t) \le \int EC(\theta)e^{r\theta}\frac{f_0^*(s-t-ir)}{f_0^*(s)f_0^*(-t)}Q(dr).$$

Consequently, (10) *holds with*

$$\frac{\delta}{\pi}\int_0^{\pi/\delta} V_C(t,t)dt \le \int EC(\theta)e^{r\theta}\Big\{\frac{\delta}{\pi}\int_0^{\pi/\delta}\frac{f_0^*(-ir)}{|f_0^*(t)|^2}dt\Big\}Q(dr).$$

The proof is straightforward and omitted.

EXAMPLE 2. Normal case: $f_0(x) = (2\pi)^{-1/2}\exp[-x^2/2]$. Since $f_0^*(z) = e^{-z^2/2}$,

$$K(x,a) = \frac{\delta}{\pi C(a)q(x)}\int_0^{\pi/\delta}\cos(t(x-a))e^{t^2/2}dt$$

in (8), and

$$\frac{\delta}{\pi}\int_0^{\pi/\delta} V_C(t,t)dt \le \int EC(\theta)e^{r\theta+r^2/2}\Big\{\frac{\delta}{\pi}\int_0^{\pi/\delta} e^{t^2}dt\Big\}Q(dr).$$

3. Remarks

Suppose (4) does not hold and G is smooth. With the $K(x,a)$ in (8) define

$$K(x,a;\delta) = K(x,a)/\delta = \frac{1}{2\pi C(a)}\int_{-\pi/\delta}^{\pi/\delta} e^{-ita}K_C^*(x,t)dt.$$

If (5) holds, $EK(X,a;\delta) \to g(a) = G'(a)$ as $\delta \to 0+$, so that $K(x,a;\delta_n)$, $\delta_n \to 0+$, can be used as kernels for the estimation of the mixing density. These kernels can be further improved through truncation and smoothing methods. For these estimators, upper bounds for $V_C(s,t)$ and its integrations in Theorems 1-4 can be translated into upper bounds for rates of convergence over classes of smooth mixing densities. See Stefanski and Carroll (1990) and Zhang (1990) for location models and Zhang (1995) for model (C).

Due to the completeness of the exponential families, the kernels in (13), (18) and (24) are the unique unbiased ones.

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. (1987). 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. Statist.* **19** 1257-1272.

−−−− (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.

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

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.

−−−− (1983b). The geometry of mixture likelihoods, part II: the exponential family. *Ann. Statist.* **11** 783-792.

Loh, W. L. and Zhang, C.-H. (1996). Global properties of kernel estimators for mixing densities in discrete exponential family models. *Statistica Sinica* **6** 561-578.

−−−− (1997). Estimating mixing densities in exponential family models for discrete variables. *Scand. J. Statist.* **24** 15-32.

Robbins, H. (1950). A generalization of the method of maximal likelihood: estimating a mixing distribution (abstract). *Ann. Math. Statist.* **21** 314-315.

−−−− (1951). Asymptotically subminimax solutions of compound statistical decision problems. *Proc. Second Berkeley Symp. Math. Statist. Probab.* **1** 131-148. Univ. of California Press, Berkeley.

−−−− (1955). An empirical Bayes approach to statistics. *Proc. Third Berkeley Symp. Math. Statist. Probab.* **1** 157-163. Univ. of 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.

Stefanski, L. A. and Carroll, R. J. (1990). Deconvolving kernel density estimators. *Ann. Statist.* **21** 169-184.

Teicher, H. (1961). Identifiability of mixtures. *Ann. Math. Statist.* **32** 244-248

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.

Wise, G. L., Traganitis, A. P. and Thomas, J. B. (1977). The estimation of a probability density function from measurements corrupted by Poisson noise. *IEEE, Trans. Information Theory* **23** 764-765.

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

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

Cun-Hui Zhang
Department of Statistics
Hill Center for the Mathematical Sciences
Busch Campus, Rutgers University
Piscataway, New Jersey 08854
e-mail : czhang@stat.rutgers.edu