USING ORTHONORMAL POLYNOMIALS TO ESTIMATE MIXING DISTRIBUTIONS

Bo E. Honoré[1]
Northwestern University

August 1989

(front page recreated 2016. The rest is from the 1989 version)

[1] This paper extends results from my University of Chicago Ph. D. dissertation (Honoré (1987)) and from Honoré (1986). Support from NSF Grant No SES-8809352 is greatfully acknowledged.

ABSTRACT: This paper illustrates how orthonormal polynomials can be used to estimate the densities of the mixing distributions in mixtures of Poissons, exponentials and normals. For the models considered, there is a simple relationship between the distribution of the data and inner products of the mixing density and a sequence of orthonormal polynomials.

ABBREVIATED TITLE: Mixing Distributions.

Estimation of the mixing distribution in a mixture model has long been a topic of interest in statistics. The contribution of this paper is to observe that in some mixtures models, the implied distribution reveals the inner products of the mixing density with a sequence of orthonormal polynomials. This observation is used to construct estimators of the mixing density. Assuming that the mixing distribution does have a density, the estimators considered in this paper are consistent in the metric defined by the inner product, and the implied estimators of the mixing distributions are consistent in the sense of weak convergence.

Orthogonal series estimators have previously been used in density estimation (for references see Prakasa–Rao (1980)), but to our knowledge this is the first time they have been used for estimation of mixing distributions. Previous estimators of the mixing distribution in mixture models include maximum likelihood estimators (Simar (1976), Lambert and Tierney (1984), and Jewell (1982)) and method of moments estimators (Tucker (1963), Heckman and Walker (1988), and Lindsay (1989)). There is a tradeoff between these estimators and the estimators described in this paper. The former are consistent (in the sense of weak convergence) for all mixing distributions, whereas we have only been able to prove consistency of the latter if the true mixing distribution has a density. On the other hand, assuming that the mixing densities do exist, it might be of interest

[1] This paper extends results from my University of Chicago Ph. D. dissertation (Honoré (1987)) and from Honoré (1986). Support from NSF Grant No SES-8809352 is greatfully acknowledged.
AMS 1980 SUBJECT CLASSIFICATION: Primary 62G05; secondary 62G25.
KEY WORDS: Mixing distributions; projections; orthonormal polynomials; consistency.

to estimate them, just as it might be of interest to estimate densities in general. The maximum likelihood and method of moments estimators do not give consistent estimators of the density. The estimators in this paper do.

The necessary notation is defined in Section 2. In Section 3, it is demonstrated that the probabilities in a mixture of Poissons can be interpreted as inner products of the mixing density and polynomials. This fact is used to construct an estimator which is consistent in the norm defined by the inner product. The problem with this estimator is that it can be negative and might not integrate to 1. However, it can easily be modified to a proper density by changing the negative values to zero and rescaling the resulting density. Assuming that the true mixing distribution has a density, the resulting density estimator gives an estimator of the mixing distribution which converges (in probability) weakly to the true mixing.

Section 4 demonstrates that the idea behind the estimator for the mixing density in a mixture of Poissons, can be used to construct estimators in a mixture of exponentials and in a mixture of normals (with known variance). In Section 5, we discuss the use of alternative inner products in the estimation of the mixing density in mixtures of Poissons and normals.

Before defining and motivating the estimators, we need to briefly review Laguerre, Legendre and Hermite Polynomials.

Let $L_2(0,\infty)$ be the space consisting of all real valued measurable functions f on $(0,\infty)$ for which $\int_0^\infty e^{-t} f(t)^2 \, dt < \infty$. Define an inner product by $\langle f, g\rangle = \int_0^\infty e^{-t} g(t) f(t) \, dt$. $L_2(0,\infty)$ is then a Hilbert space with norm $\|f\| = \langle f, f\rangle^{\frac{1}{2}}$. Let $L_i(t)$, $i = 0, 1, \ldots$, be the Laguerre polynomials defined by

$$L_i(t) = \sum_{j=0}^{i} c_{ji}^L \, t^j,$$

where

$$c_{ji}^L = (-1)^j \binom{i}{i-j} \frac{1}{j!}$$

(see, for example, Abramowitz and Stegun (1970)). The Laguerre polynomials are a complete orthonormal sequence in $L_2(0,\infty)$. It follows that $\sum_{i=0}^\infty \langle f, L_i\rangle L_i$ converges to f for $f \in L_2(0,\infty)$ (Luenberger (1969), Theorem 2, page 60).

Likewise, $L_2[-1,1]$ is the space consisting of all the real valued measurable functions f on $[-1,1]$ for which f^2 is Lebesque integrable. $L_2[-1,1]$ is a Hilbert space with inner product $\langle f,g\rangle = \int_{-1}^{1} f(t)g(t)\,dt$. Let $P_i(t)$ $i=0,1,\ldots$ be the Legendre polynomials,

$$P_i(t) = \sum_{j=0}^{i} c_{ji}^{P}\, t^j,$$

where

$$c_{ji}^{P} = \begin{cases} 0 & \text{for } i-j \text{ odd,} \\ \dfrac{1}{2^i}(-1)^{\frac{i-j}{2}}\dbinom{i}{\frac{i-j}{2}}\dbinom{i+j}{i} & \text{for } i-j \text{ even.} \end{cases}$$

Let $e_i(t) = \sqrt{\frac{2i+1}{2}}P_i(t)$. The sequence $\{e_i : i=0,1,\ldots\}$ is then a complete orthonormal sequence in $L_2[-1,1]$. It therefore follows that $\sum_{i=0}^{\infty}\langle f, e_i\rangle e_i$ converges to f for $f \in L_2(-1,1)$.

Finally, let $L_2(-\infty,\infty)$ be the space consisting of all real valued measurable functions f on $(-\infty,\infty)$ for which $\int_{-\infty}^{\infty} e^{-t^2/2} f(t)^2\ dt < \infty$. Define an inner product by $\langle f,g\rangle = \int_{-\infty}^{\infty} e^{-t^2/2} g(t)f(t)\ dt$. $L_2(-\infty,\infty)$ is then a Hilbert space with norm $\|f\| = \langle f,f\rangle^{\frac{1}{2}}$. Let $H_i(t)$, $i=0,1,\ldots$, be the Hermite polynomials defined by

$$H_i(t) = \sum_{j=0}^{i} c_{ji}^{H}\, t^j,$$

where

$$c_{ji}^{H} = \begin{cases} 0 & \text{for } i-j \text{ odd,} \\ \left(-\dfrac{1}{2}\right)^{(i-j)/2} \dfrac{i!}{(\frac{i-j}{2})!\, j!} & \text{for } i-j \text{ even.} \end{cases}$$

(see, for example, Abramowitz and Stegun (1970)). The Hermite polynomials are a complete orthonormal sequence in $L_2(-\infty,\infty)$. Hence $\sum_{i=0}^{\infty}\langle f, H_i\rangle H_i$ converges to f for $f \in L_2(-\infty,\infty)$.

In this paper, we use $\langle\cdot,\cdot\rangle$ and $\|\cdot\|$ to denote the inner products and norms on all three spaces.

A mixture of Poissons has support on the non–negative integers, and the probabilities $P(X = i) = \pi_i$ are given by

$$\pi_i = \int_0^{\infty} \frac{1}{i!} e^{-\theta}\theta^i\, dG(\theta) \tag{1}$$

where G is the mixing distribution. In this section, it will be assumed that G has a bounded density, g. This implies that $g \in L_2(0,\infty)$. The aim is to estimate g and/or G from a sample of n i.i.d. observations of X.

With the notation of Section 2, note that

$$i!\,\pi_i = \int_0^\infty e^{-\theta}\theta^i g(\theta)\,d\theta = \langle g, \theta^i\rangle, \tag{2}$$

and therefore

$$\langle g, L_i\rangle = \sum_{j=0}^{i} c_{ij}^L \langle g, \theta^j\rangle = \sum_{j=0}^{i} c_{ij}^L\, j!\,\pi_j.$$

This implies that $\gamma_i \stackrel{\text{def}}{=} \langle g, \theta^i\rangle$ can be consistently estimated by $\hat{\gamma}_i \stackrel{\text{def}}{=} \sum_{j=0}^{i} c_{ji}^L\, j!\,\hat{\pi}_j$, where $\hat{\pi}_i$ is the fraction of the observation for which $X = i$.

The idea behind the estimator defined in this section is to approximate g by $\sum_{i=0}^{I(n)} \hat{\gamma} L_i$, where the order of the approximation, $I(n)$, depends on the sample size in such a way that the estimator is consistent in the norm $\|\cdot\|$.

To specify the rate at which the degree of the polynomial increases with sample size $I(n)$, notice that for a given degree, I,

$$\begin{aligned}\Big\|g - \sum_{i=0}^{I}\hat{\gamma}_i L_i\Big\| &\le \Big\|g - \sum_{i=0}^{I}\gamma_i L_i\Big\| + \Big\|\sum_{i=0}^{I}(\gamma_i - \hat{\gamma}_i)L_i\Big\| \\ &= \Big\|g - \sum_{i=0}^{I}\gamma_i L_i\Big\| + \Big(\sum_{i=0}^{I}(\gamma_i - \hat{\gamma}_i)^2\Big)^{\frac{1}{2}}\end{aligned}$$

The first term goes to 0 as $I \to \infty$, so we must find an $I(n)$ such that $I(n) \to \infty$ as $n \to \infty$ and $\Big(\sum_{i=0}^{I(n)}(\gamma_i - \hat{\gamma}_i)^2\Big)^{\frac{1}{2}} \xrightarrow{p} 0$ as $n \to \infty$. To do this, it suffices to find a function $I(n)$ such that $I(n) \to \infty$ and $E\Big[\sum_{i=0}^{I(n)}(\hat{\gamma}_i - \gamma_i)^2\Big] \to 0$ as $n \to \infty$.

From

$$\gamma_i - \hat{\gamma}_i = \sum_{j=0}^{i} c_{ji}^L\, j!\,(\hat{\pi}_j - \pi_j)$$

it is seen that

$$E\big[(\gamma_i - \hat{\gamma}_i)^2\big] = \sum_{j=0}^{i}\sum_{k=0}^{i} c_{ji}c_{ki}\,k!\,j!\;\mathrm{cov}(\hat{\pi}_j, \hat{\pi}_k) \le \sum_{j=0}^{i}\sum_{k=0}^{i} |c_{ji}|\,|c_{ki}|\;j!\,k!\,n^{-1}.$$

To simplify notation let $d_I = \sum_{i=0}^{I} \sum_{j=0}^{i} \sum_{k=0}^{i} |c_{ji}|\,|c_{ki}|\; j!\,k!$. Then

$$E\Big[\sum_{i=0}^{I} (\gamma_i - \hat{\gamma}_i)^2\Big] \le \tilde{d}_I n^{-1}$$

So any rule $I(n)$ for which $\tilde{d}_{I(n)} n^{-1} \to 0$ and $I(n) \to \infty$ as $n \to \infty$ will suffice (such rules obviously exist).

We have therefore proved

THEOREM. *Assume that G in (2) has a bounded density g. Let $\hat{\gamma}_i = \sum_{j=0}^{i} c_{ji}^L j!\,\hat{\pi}_j$ and let $I(n)$ be any sequence of integers such that $I(n) \to \infty$ and $I(n)n^{-1} \to 0$ as $n \to \infty$, then $\hat{g}_n(t) = \sum_{i=0}^{I(n)} \hat{\gamma}_i L_i(t)$ is a consistent (in the metric, $\|\cdot\|$, defined above) estimator of g:*

$$\big\|\hat{g}_n - g\big\| \xrightarrow{p} 0 \qquad \text{as} \qquad n \to \infty.$$

In kernel estimation of densities, it is of interest to let the bandwidth depend on the data. Similarly, it might be of interest to let the order of the polynomial for the estimator depend on that data. It is clear from the discussion above that the consistency result would still hold if $I(n)$ is random and $\tilde{d}_{I(n)} n^{-1} \xrightarrow{a.s.} 0$ and $I(n) \xrightarrow{a.s.} \infty$ as $n \to \infty$.

There is no guarantee that $\hat{g}_n$ will be positive and integrate to 1. To get a non–negative estimator of g, define $\tilde{g}_n \stackrel{\text{def}}{=} 1_{\{\hat{g}_n \ge 0\}} \hat{g}_n$. As $g \ge 0$, $\big\|\tilde{g}_n - g\big\| \le \big\|\hat{g}_n - g\big\|$ and hence $\big\|\tilde{g}_n - g\big\| \xrightarrow{p} 0$ as $n \to \infty$. So $\tilde{g}_n$ is consistent in the norm $\|\cdot\|$.

If the support of g is known to be bounded from above by $M < \infty$, then $\tilde{g}_n$ can be normalized to integrate to 1. Define $\breve{g}_n(\theta) = 1_{\{0 \le \theta \le M\}} a_n^{-1} \tilde{g}_n(\theta)$, where $a_n = \int_0^M \tilde{g}_n(\theta)\, d\theta$.

First notice that $\big\|\tilde{g}_n - g\big\| \xrightarrow{p} 0$ implies that $\int_0^M |\tilde{g}_n(\theta) - g(\theta)|\, e^{-\theta}\, d\theta \xrightarrow{p} 0$, and hence

$$\big|a_n - 1\big| = \left|\int_0^M \big(\breve{g}_n(\theta) - g(\theta)\big)\, d\theta\right| \le \int_0^M \big|\breve{g}_n(\theta) - g(\theta)\big|\, d\theta \le e^M \int_0^M \big|\tilde{g}_n(\theta) - g(\theta)\big| e^{-\theta}\, d\theta \xrightarrow{p} 0$$

so $a_n \xrightarrow{p} 1$. This implies that

$$\begin{aligned}\big\|\breve{g}_n - g\big\| &= \left(\int_0^M \big(\breve{g}_n(\theta) - g(\theta)\big)^2 e^{-\theta}\, d\theta\right)^{\frac{1}{2}} \\ &\le \left(\int_0^M \big(\breve{g}_n(\theta) - \tilde{g}_n(\theta)\big)^2 e^{-\theta}\, d\theta\right)^{\frac{1}{2}} + \left(\int_0^M \big(\tilde{g}_n(\theta) - g(\theta)\big)^2 e^{-\theta}\, d\theta\right)^{\frac{1}{2}}\end{aligned}$$

$$
\begin{aligned}
&\leq \left(\int_0^M \big(\breve{g}_n(\theta) - \tilde{g}_n(\theta)\big)^2 e^{-\theta}\, d\theta \right)^{\frac{1}{2}} + \left(\int_0^M \big(\tilde{g}_n(\theta) - g(\theta)\big)^2 e^{-\theta}\, d\theta \right)^{\frac{1}{2}} \\
&\leq \big|a_n - 1\big| \big\|\tilde{g}_n\big\| + \big\|\tilde{g}_n - g\big\| \\
&\xrightarrow{\ p\ } 0 \qquad \text{as} \qquad n \to \infty,
\end{aligned}
$$

So $\breve{g}_n$ is consistent in the norm $\|\cdot\|$.

If $\breve{G}_n$ is defined by $\breve{G}_n(\theta) \stackrel{\text{def}}{=} \int_0^\theta \breve{g}_n(\eta)\, d\eta$, then $\breve{G}_n$ is a distribution function and $\breve{G}_n$ converges weakly to G (in probability):

$$
\sup_{0 \leq \theta \leq M} \big|\breve{G}_n(\theta) - G(\theta)\big| \leq \sup_{0 \leq \theta \leq M} \int_0^\theta \big|\breve{g}_n(\eta) - g(\eta)\big|\, d\eta = \int_0^M \big|\breve{g}_n(\eta) - g(\eta)\big|\, d\eta \xrightarrow{\ p\ } 0.
$$

To summarize, we have proved

THEOREM. *Assume that G in (2) has a bounded density g and support bounded by some known M. Then the estimator $\breve{g}_n$ defined above is a proper density and it is consistent in the metric $\|\cdot\|$: $\big\|\breve{g}_n - g\big\| \xrightarrow{\ p\ } 0$ as $n \to \infty$. Also, the implied distribution function, $\breve{G}_n$ converges weakly to G in probability.*

In Section 5, it is demonstrated that the same idea that was used to construct $\breve{g}_n$ can be used to construct an alternative estimator for which it is not necessary to assume that M is finite and known.

Finally, observe that for $j \geq 1$

$$
\begin{aligned}
j!\,\pi_j = \int_0^\infty e^{-\theta}\theta^j\, dG(\theta) = e^{-\theta}\theta^j G(\theta)\Big|_0^\infty - \int_0^\infty \big(j\theta^{j-1}e^{-\theta} - e^{-\theta}\theta^j\big) G(\theta)\, d\theta \\
= 0 - j\gamma_{j-1} + \gamma_j
\end{aligned}
$$

and

$$
\pi_0 = \int_0^\infty e^{-\theta}\, dG(\theta) = e^{-\theta} G(\theta)\Big|_0^\infty - \int_0^\infty -e^{-\theta} G(\theta)\, d\theta = \gamma_0
$$

so $\gamma_j = j! \sum_{k=0}^j \pi_k$ for $j = 0, 1, \ldots$, and therefore

$$
\langle G, L_i \rangle = \sum_{j=0}^{i} c_{ji}^L j! \sum_{k=0}^{j} \pi_k = \sum_{j=0}^{i} c_{ji}^L j!\, P(X \leq j).
$$

This could be used to construct an estimator for G which is consistent in $\|\cdot\|$.

The idea behind the estimator defined in Section 3 can also be used to construct estimators of the mixing distribution in mixtures of exponentials and normals (with known variance).

Let X be a random variable with survivor function

$$S(x) = P(X > x) = \int_0^\infty e^{-\theta x}\, dG(\theta). \tag{3}$$

Define

$$f(\eta) = \begin{cases} g(-\log \eta) & \text{for } 0 < \eta \leq 1 \\ 0 & \text{for } -1 \leq \eta \leq 0 \end{cases}$$

Assume that g is bounded. It then follows that $f \in L_2[-1, 1]$. As in Section 3, the basis for the estimator is that it is very easy to consistently estimate the projections of f on a set of orthonormal polynomials. For any integer $i \geq 0$,

$$S(i) = \int_0^\infty e^{-\theta i}\, dG(\theta) = \int_0^\infty \left(e^{-}\theta\right)^i g(\theta)\, d\theta = \int_0^1 \eta^{i-1} f(\eta)\, d\eta$$

which implies

$$\langle f, x^i \rangle = S(i+1) - \frac{(-1)^{i+1}}{i+1}$$

This means that we can estimate the projection of f on all polynomials by simply substituting the empirical survivor function S_n for S in the expression above, and hence the projection on the Legendre polynomials can be estimated.

Letting

$$\tilde{c}_{ji} = \sqrt{\frac{2i+1}{2}}\, c^P_{ji}, \qquad d_i = \sum_{j=0}^{i} \sum_{k=0}^{i} |\tilde{c}_{ji}|\, |\tilde{c}_{ki}|\,, \qquad \text{and} \qquad \tilde{d}_I = \sum_{i=0}^{I} d_i,$$

it is straightforward to show that if $I(n)$ is a function such that $I(n) \to \infty$ and $\tilde{d}_{I(n)} n^{-1} \to 0$ as $n \to \infty$, then the estimator of f described above is consistent in the $L_2[-1, 1]$ metric. As was the case for the estimator of the compound distribution in a mixture of Poissons, it is clear that we can allow $I(n)$ to be random as long as $I(n) \xrightarrow{a.s.} \infty$ and $\tilde{d}_{I(n)} n^{-1} \xrightarrow{a.s.} 0$ as $n \to \infty$.

Two points should be made about the estimator for f defined in this section. First, notice that it is not necessary to have exact observations on T in order to estimate f. We only need to estimate S at the integers, so it is sufficient to have grouped observations on T. The other thing to note is that an estimator for f could be constructed by estimating the projections of f on the shifted Legendre polynomials which span the space $L_2[0,1]$.

From the estimator $\hat{f}$ of f, an estimator of g can be defined by $\hat{g}(\theta) = \hat{f}\big(e^{-\theta}\big)$. The consistency of $\hat{f}$ in the metric $\|\cdot\|$ implies that

$$\int_0^1 \big(\hat{f}(\eta) - f(\eta)\big)^2 \, d\eta = \int_0^1 \big(\hat{g}(-\log\eta) - g(-\log\eta)\big)^2 \, d\eta = \int_0^\infty \big(\hat{g}(\theta) - g(\theta)\big)^2 e^{-\theta} \, d\eta \xrightarrow{p} 0.$$

So the estimator $\hat{g}$ is consistent in the same metric as the estimator of the mixing density in a mixture of Poissons. Hence, if the support of g is known to be bounded from above by some M, then the logic of Section 3 can be mimicked to construct an estimator of g. The resulting estimator is positive and integrates to 1 and the integral of which converges weakly to G (in probability).

Notice that orthonormal polynomials can be used to estimate G in (3) directly. Define

$$\tilde{G}(\eta) = \begin{cases} 1 - G(-\log\eta) & \text{for } 0 < \eta \le 1 \\ 0 & \text{for } -1 \le \eta \le 0 \end{cases}$$

It follows that $\tilde{G} \in L_2[-1,1]$. Then for any integer $i \ge 0$,

$$S(i) = \int_0^\infty e^{-\theta i} \, dG(\theta) = \int_0^1 \eta^i \, d\tilde{G}(\eta) = 1 - i \int_0^1 \eta^{i-1}(1 - F(\eta)) \, d\eta,$$

so we have

$$\langle \tilde{G}, \eta^i \rangle = \int_{-1}^1 \eta^i \tilde{G}(\eta) \, d\eta = \int_0^1 \eta^i \tilde{G}(\eta) \, d\eta = \frac{1 - S(i+1)}{i+1}$$

which means that we can easily estimate the projection of $\tilde{G}$ on all polynomials.

As a final example of how orthonormal polynomials can be used to estimate mixing distributions, we consider a mixture of normals with known variance (which without loss of generality will be assumed to equal 1).

The mixture of normals is defined by

$$\text{(4)} \qquad \mathrm{P}(X \le x) = \int_{-\infty}^\infty \Phi(x - \theta) \, dG(\theta)$$

where Φ is the normal c.d.f.

It is helpful to notice that for $i \ge 0$ and $\kappa < \frac{1}{2}$

$$
\begin{aligned}
E\Big[Y^i e^{\kappa Y^2}\big|Y\sim N(\mu,1)\Big] &= \int_{-\infty}^{\infty}\frac{y^i}{\sqrt{2\pi}}\ \exp\{\kappa y^2\}\exp\Big\{\frac{-(y-\mu)^2}{2}\Big\}\,dy\\
&= \int_{-\infty}^{\infty}\frac{y^i}{\sqrt{2\pi}}\ \exp\Big\{-\frac{y^2(1-2\kappa)-2\mu y+\mu^2}{2}\Big\}\,dy\\
&= \int_{-\infty}^{\infty}\frac{y^i}{\sqrt{2\pi}}\ \exp\Big\{-\frac{y^2-2\mu y(1-2\kappa)^{-1}+\mu^2(1-2\kappa)^{-1}}{2(1-2\kappa)^{-1}}\Big\}\,dy\\
&= \int_{-\infty}^{\infty}\frac{y^i}{\sqrt{2\pi}}\ \exp\Big\{-\frac{(y-\mu(1-2\kappa)^{-1})^2}{2(1-2\kappa)^{-1}}-\frac{\mu^2}{2}+\frac{\mu^2}{2(1-2\kappa)}\Big\}\,dy\\
&= \frac{\exp\Big\{\frac{\kappa\mu^2}{1-2\kappa}\Big\}}{\sqrt{1-2\kappa}}\int_{-\infty}^{\infty}\frac{y^i}{\sqrt{2\pi(1-2\kappa)^{-1}}}\ \exp\Big\{-\frac{(y-\mu(1-2\kappa)^{-1})^2}{2(1-2\kappa)^{-1}}\Big\}\,dy\\
&= \frac{\exp\Big\{\frac{\kappa\mu^2}{1-2\kappa}\Big\}}{\sqrt{1-2\kappa}}E\Big[Y^i\mid Y\sim N(\frac{\mu}{1-2\kappa},\frac{1}{1-2\kappa})\Big]\\
&= \frac{\exp\Big\{\frac{\kappa\mu^2}{1-2\kappa}\Big\}}{\big(\sqrt{1-2\kappa}\big)^{i+1}}E\Big[\big(Y+\frac{\mu}{\sqrt{1-2\kappa}}\big)^i\mid Y\sim N(0,1)\Big]
\end{aligned}
$$

For $\kappa=-\frac{1}{2}$ we have

$$
\begin{aligned}
E\Big[Y^i e^{-Y^2/2}\big|Y\sim N(\mu,1)\Big] &= e^{-\mu^2/4}\sum_{j=0}^{i}\binom{i}{j}m_{i-j}\left(\frac{1}{2}\right)^{\frac{i+j+1}{2}}\mu^j\\
&= e^{-\mu^2/4}\sum_{j=0}^{i}k_{ij}\mu^j
\end{aligned}
$$

where m_j is the j^{th} moment of a standard normal and

$$k_{ij}=\binom{i}{j}m_{i-j}\left(\frac{1}{2}\right)^{\frac{i+j+1}{2}}.$$

So if X is distributed as in (4) and G has bounded density g, then with $f(\eta)=g(\sqrt{2}\eta)$ and $\tilde{k}_{ij}=k_{ij}2^{(i+1)/2}$,

$$E\big[X^i e^{-X^2/2}\big]=\sum_{j=0}^{i}k_{ij}\int_{-\infty}^{\infty}e^{-\theta^2/4}\theta^j g(\theta)\,d\theta=\sum_{j=0}^{i}\tilde{k}_{ij}\int_{-\infty}^{\infty}e^{-\eta^2/2}\eta^j f(\eta)\,d\eta=\sum_{j=0}^{i}\tilde{k}_{ij}\langle\eta^j,f\rangle$$

which can be rewritten as

$$\langle\eta^i,f\rangle=\begin{cases}E[e^{-X^2/2}]/\tilde{k}_{00} & \text{for } i=0\\ \Big(E[X^i e^{-X^2/2}]-\sum_{j=0}^{i-1}\tilde{k}_{ij}\langle\eta^j,f\rangle\Big)/\tilde{k}_{ii} & \text{for } i=1,2,\ldots\end{cases}$$

where we have used that $\tilde{k}_{ii} \neq 0$ for $i = 0, 1, 2, \ldots$.

This implies that the inner products of f with polynomials can be consistently estimated. Furthermore, observe that

$$E\Big[\big(X^i e^{-X^2/2}\big)^2\Big] = \sum_{j=0}^{2i} \binom{2i}{j} m_{2i-j} \left(\frac{1}{3}\right)^{\frac{2i+j+1}{3}} \int_{-\infty}^{\infty} e^{-\theta^2/3}\theta^j g(\theta)\, d\theta$$

$$\leq \sum_{j=0}^{2i} \binom{2i}{j} m_{2i-j} \left(\frac{1}{3}\right)^{\frac{2i+j+1}{3}} r_j$$

where $r_j = \sup_{0 \leq \theta < \infty} e^{-\theta^2/3}\theta^j < \infty$. This implies that (except for algebra) it is simple to find a function $I(n)$ such that the estimator for f based on orthonormal polynomials is consistent.

If it is assumed that the support of g is included in the interval $(-M, M)$ for some known M, then this estimator can be used to construct an estimator for G which is consistent (in the sense of weak convergence) by exactly the same argument as in Section 3.

As mentioned in the previous sections, the density estimators based on orthonormal polynomials can be integrated to give estimators of the mixing distributions which converge weakly to the true mixing distribution. In order to obtain that result, it was assumed that the support of the mixing distribution is bounded by known constants. We now demonstrate that the orthonormal polynomial estimators of the mixing distribution in Poissons and normals can be modified in such a way that the weak convergence of the estimated distribution function to the true mixing distribution holds even if the mixing distribution has unbounded support.

The modification of the estimators will be based on a slight modification of the definitions in Section 2. Let $L_2(0,\infty)$ be the space consisting of all real valued measurable functions, f, on $(0,\infty)$ for which $\int_0^\infty f(t)^2\, dt < \infty$. Define an inner product by $\langle f, g\rangle = \int_0^\infty g(t)f(t)\, dt$. $L_2(0,\infty)$ is then a Hilbert space with norm $\|f\| = \langle f, f\rangle^{\frac{1}{2}}$. Let $L_i^\circ(t) = e^{-t/2}L_i(i)$, $i = 0, 1, \ldots$, where L_i are the Laguerre polynomials defined in Section 2. The functions L_i° are a complete orthonormal sequence in $L_2(0,\infty)$. It follows that $\sum_{i=0}^\infty \langle f, L_i^\circ\rangle L_i^\circ$ converges to f for $f \in L_2(0,\infty)$. Likewise, let $L_2(-\infty,\infty)$ be the space consisting of all real valued measurable functions f on $(-\infty,\infty)$ for which $\int_{-\infty}^\infty f(t)^2\, dt < \infty$. Define an inner product by $\langle f, g\rangle = \int_{-\infty}^\infty g(t)f(t)\, dt$. $L_2(-\infty,\infty)$ is then a Hilbert space with norm $\|f\| = \langle f, f\rangle^{\frac{1}{2}}$. Let $H_i^\circ(t) = e^{-t^2/4}H_i(t)$, $i = 0, 1, \ldots$, where H_i are the

Hermite polynomials defined in Section 2. The functions H_i° are a complete orthonormal sequence in $L_2(-\infty,\infty)$. Hence $\sum_{i=0}^{\infty}\langle f, H_i^\circ\rangle H_i^\circ$ converges to f for $f \in L_2(-\infty,\infty)$.

Consider first the mixture of Poissons given by (1). Assume that G has density g satisfying $\int_0^\infty g(\theta)^2 d\theta < \infty$, so $g \in L_2(0,\infty)$. Then

$$i!\,\pi_i = \int_0^\infty e^{-\theta}\theta^i g(\theta)\,d\theta = \left(\frac{1}{2}\right)^i \int_0^\infty e^{-\eta/2}\eta^i f(\eta)\,d\eta = \left(\frac{1}{2}\right)^i \langle f, e^{-\eta/2}\eta^i\rangle$$

where $f(\theta) = \frac{1}{2}g(\theta/2)$, i.e. f is the density of two times a random variable with density g. So the inner product of f and L_i° can be expressed as a linear combination of the probabilities π_j, $j = 0, 1, \ldots, i$. This implies that if $\int_0^\infty g(\theta)^2\,d\theta < \infty$ (and hence $\int_0^\infty f(\theta)^2\,d\theta < \infty$), then the logic of Section 3 can be used to construct an estimator for f, $\hat{f}_n$, such that $\int_0^\infty \big(\hat{f}_n(\theta) - f(\theta)\big)^2\,d\theta \xrightarrow{p} 0$, and therefore $\int_0^\infty \big(\hat{g}_n(\theta) - g(\theta)\big)^2\,d\theta \xrightarrow{p} 0$ where $\hat{g}_n(\theta) = 2\hat{f}_n(2\theta)$. Mimicking the discussion in Section 3, this implies that if $\hat{g}_n$ is first modified by changing the negative values to 0, and then normalized to integrate to 1, then the resulting estimator, $\check{g}_n$, is consistent in the norm defined above, and $\sup_{0\le\theta<\infty}\big|\check{G}_n(\theta) - G(\theta)\big| \xrightarrow{p} 0$, where $\check{G}_n(\theta) = \int_0^\theta \check{g}_n(\eta)\,d\eta$.

Next consider the mixture of normals (4). From the expression for $E\big[X^i e^{-X^2/4}\big]$, it is clear that the inner products of g with $e^{-\theta^2/4}\theta^i$ (and the inner products of g with H_i° can be consistently estimated. This can be used to construct an estimator for g, $\hat{g}_n$, such that $\int_{-\infty}^\infty \big(\hat{g}_n(\theta) - g(\theta)\big)^2\,d\theta \xrightarrow{p} 0$. If this estimator is modified to be non–negative and integrated to 1 as above, then the resulting estimator $\check{g}_n$, is consistent in the same norm as $\hat{g}_n$, and $\sup_{-\infty\le\theta<\infty}\big|\check{G}_n(\theta) - G(\theta)\big| \xrightarrow{p} 0$, where $\check{G}_n(\theta) = \int_{-\infty}^\theta \check{g}_n(\eta)\,d\eta$.

Abramowitz, M., and I. A. Stegun: *Handbook of Mathematical Functions.* New York: Dover Press.

Heckman, J. J. and J. R. Walker (1987a): "Testing the Mixtures of Exponentials Hypothesis and Estimating the Mixing Distribution by the Method of Moments." Unpublished manuscript.

Honoré, B. E. (1986): "Estimation of the Mixing Distribution in a Mixture of Poissons Using Orthogonal Polynomials: A Note." Unpublished manuscript.

Honoré, B. E. (1987): "Identification and Estimation of Econometric Duration Models." Unpublished Ph. D. dissertation, University of Chicago.

Jewell, N. (1982): "Mixtures of Exponential Distributions." *Annals of Statistics*, **10**, pp. 479–484.

Lambert, D., and L. Tierney (1984) "Asymptotic Properties of Maximum Likelihood Estimates in the Mixed Poisson Model," *Annals of Statistics*, **12**, pp. 1388–1399.

Lindsay, B. G. (1989): "Moment Matrices: Applications in Mixtures," *Annals of Statistics*, **17**, pp. 722–740.

Luenberger, D. G. (1970): *Optimation by Vector Space Methods.* New York: John Wiley and Sons.

Prakasa–Rao, B. L. S. (1983): *Nonparametric Functional Estimation.* Orlando: Academic Press.

Simar, L. (1976): "Maximum Likelihood Estimation of a Compound Poisson Process," *Annals of Statistics*, **4**, pp. 1200–1209.

Tucker, H. G. (1963): "An Estimate of the Compounding Distribution of a Compound Poisson Distribution," *Theory of Probability and Its Applications*, **8**, pp. 195–200.

Bo E. Honoré
Department of Economics
Northwestern University
2003 Sheridan Road
Evanston, IL 60208–2400