The Annals of Statistics
1984, Vol. 12, No. 4, 1388–1399

ASYMPTOTIC PROPERTIES OF MAXIMUM LIKELIHOOD ESTIMATES IN THE MIXED POISSON MODEL

BY DIANE LAMBERT AND LUKE TIERNEY[1]

Carnegie-Mellon University and University of Minnesota

This paper considers the asymptotic behavior of the maximum likelihood estimators (mle's) of the probabilities of a mixed Poisson distribution with a nonparametric mixing distribution. The vector of estimated probabilities is shown to converge in probability to the vector of mixed probabilities at rate $n^{1/2-\varepsilon}$ for any $\varepsilon > 0$ under a generalized χ^2 distance function. It is then shown that any finite set of the mle's has the same joint limiting distribution as does the corresponding set of sample proportions when the support of the mixing distribution G_0 is an infinite set with a known upper bound and G_0 satisfies a certain condition at zero.

1. Introduction and summary. In many applications count data can be viewed as arising from a mixed Poisson distribution. That is, the counts can be viewed as a random sample $X_1, \cdots, X_n$ with

$$P\{X_k = i\} = \pi_{i,0} = \int f_i(\lambda)\, dG_0(\lambda),$$

where $f_i(\lambda)$ is the probability of a count of i under a Poisson distribution with mean λ. For applications where there is little information about the mixing distribution, Simar (1976) has proposed estimating G_0 nonparametrically. (It follows from results of Teicher (1961) that the mixing distribution of a Poisson mixture is identifiable.) Specifically, Simar proposes to estimate G_0 by the distribution $\hat{G}_n$ on $[0, \infty)$ that maximizes the log likelihood

$$\mathscr{L}(G) = n \textstyle\sum_{i=0}^{\infty} \hat{p}_{i,n} \log\left(\displaystyle\int f_i(\lambda)\, dG(\lambda)\right),$$

where $\hat{p}_{i,n}$ denotes the proportion of the n observations equaling i. One algorithm for computing $\hat{G}_n$ is described in Simar (1976); other computational approaches are described in Laird (1978). Simar also proves that for each n the maximum likelihood estimator (mle) $\hat{G}_n$ is a unique distribution on $[0, \infty)$ with finite support and that $\hat{G}_n$ is strongly consistent in the sense that, almost surely, $\hat{G}_n$ converges weakly to G_0.

Recently, the results obtained by Simar for mixtures of Poisson distributions have been extended to mixtures of other families of distributions. Jewell (1982)

Received October 1982; revised May 1984.

[1] The research for this article was done while both authors were at The Department of Statistics, Carnegie-Mellon University. The research was partially funded by NSF Grant MCS-8301692 at Carnegie-Mellon University.

AMS 1980 *subject classifications*. 62E20, 62G05.

Key words and phrases. Asymptotic normality, consistency, nonparametric maximum likelihood estimation.

considers mixtures of exponential and Weibull distributions and shows that, in this case also, $\hat{G}_n$ is unique, has finite support, and is strongly consistent. Lindsay (1983a, b) shows that the uniqueness and finite support properties of the mle are closely related to the geometry of the likelihood function and that these properties hold for many parametric families, including most one parameter exponential families.

In order for nonparametric maximum likelihood techniques to be useful in practice, the precision of the mle's should be known at least approximately. In this paper we consider the large sample precision of the mle's $\hat{\pi}_{i,n} = \int f_i(\lambda)\, d\hat{G}_n(\lambda)$ of the mixed Poisson probabilities. Our major result is that if the support of the mixing distribution G_0 is an infinite set with a known upper bound M and $G_0(x) - G_0(0)$ tends to zero no faster than some power of x as x tends to zero, then any finite set of the $\hat{\pi}_{i,n}$'s has the same limiting distribution as the corresponding set of sample proportions.

Our problem can roughly be described as follows. The vector $\hat{\pi}_n$ can be viewed as a projection of the vector $\hat{p}_n$ of sample proportions onto the space $\mathscr{M}$ of all mixed Poisson probability vectors. If the number of support points of G_0, say N_0, is finite and unknown, then the mixed probability vector π_0 is on the boundary of $\mathscr{M}$. Hence it is unreasonable to expect that $\hat{\pi}_{i,n}$ is asymptotically normal. On the other hand, if N_0 is large then $\hat{\pi}_n$ can come closer to the global maximum $\hat{p}_n$. The question to be answered is whether the difference between $\hat{\pi}_n$ and $\hat{p}_n$ is in fact asymptotically negligible when the support of G_0 is infinite. The somewhat unexpected answer is yes, at least if G_0 is well-behaved at the origin and has a support set with a known upper bound.

Two χ^2-type norms that are convenient for studying the limiting behavior of $n^{1/2}(\hat{\pi}_n - \pi_0)$ and $n^{1/2}(\hat{p}_n - \pi_0)$ are given in Section 2. Both $\hat{\pi}_n$ and $\hat{p}_n$ are shown to converge in these norms to π_0 at rate $n^{1/2-\varepsilon}$ for any $\varepsilon > 0$ in Section 3. The major result of the paper is developed in Section 4. There we show that if the support of the mixing distribution G_0 is an infinite set with a known upper bound M, then the linear combinations $\sum c_i \hat{\pi}_{i,n}/\pi_{i,0}$ and $\sum c_i \hat{p}_{i,n}/\pi_{i,0}$ have the same limiting normal distribution when c belongs to a certain class of vectors $\mathscr{L}_1$. We do not know whether the unit vectors $e_j = (e_{0,j}, e_{i,j}, \cdots)$, given by $e_{i,j} = 1$ if $i = j$ and 0 otherwise, belong to $\mathscr{L}_1$ in general. The unit vectors do, however, belong to $\mathscr{L}_1$ in the special case that $G_0(x) - G_0(0)$ tends to zero no faster than some power of x as x tends to zero (Section 4). Consequently, under these conditions $\hat{p}_n$ and $\hat{\pi}_n$ have the same limiting distributions. Further remarks about the limiting behavior of $\hat{\pi}_n - \pi_0$ are made in Section 5.

It might be argued that our asymptotic result is irrelevant or even misleading when G_0 is in fact supported on only a few points. This does not appear to be the case, however. To see this, note that if G_0 has N_0 support points and N_0 is known, then asymptotic standard errors can be computed by standard techniques. When N_0 is unknown, using these formulas with the number of points $\hat{N}_n$ of the mle in place of N_0 should underestimate the true standard errors of the nonparametric $\hat{\pi}_{i,n}$'s. Comparing these estimates with $\hat{N}_n$ in place of N_0 to our asymptotic standard error estimates $n^{-1/2}(\hat{\pi}_{i,n}(1 - \hat{\pi}_{i,n}))^{1/2}$ should indicate the conservatism of the asymptotics of this paper when the support of G_0 is finite. We have made such a comparison for the data given by Simar (1976) on the number of car accidents of 9461 policyholders of La Royale Belge Insurance Company. The number of accidents per policyholder ranged from 0 to 7. Simar fit a nonparametric Poisson mixture and obtained an estimated mixing distribution with four support points. For counts 0, 1 and 2, the estimated standard errors multiplied by $9461^{1/2}$ using $\hat{N}_n = 4$ in place of N_0 and the estimates $(\hat{\pi}_{i,n}(1 - \hat{\pi}_{i,n}))^{1/2}$ agreed to three digits. With a count of six, for which $\hat{\pi}_{6,n} = .00024$, the two standard error estimates multiplied by $9461^{1/2}$ agreed to two digits.

Our results can be described as giving conditions under which the sample proportions $\hat{p}_n$ are asymptotically as efficient as the mle's $\hat{\pi}_n$. Note, however, that we are not suggesting that the estimator $\hat{p}_n$ be used instead of $\hat{\pi}_n$, but rather that $(\hat{\pi}_{i,n}(1 - \hat{\pi}_{i,n})/n)^{1/2}$ provides a reasonable estimate of the standard error of $\hat{\pi}_{i,n}$. Although $\hat{\pi}_{i,n}$ need not be more efficient than $\hat{p}_{i,n}$ in large samples, $\hat{\pi}_{i,n}$ is a smoother estimator and it may be nonzero when $\hat{p}_{i,n}$ is zero. Smoothness may be more important than asymptotic efficiency when rare events are of concern. Neverthless, our results do raise the question of whether there is some estimator that is more efficient that $\hat{p}_n$. Tierney and Lambert (1984) show that the answer is no if only estimators that are "regular" in a sense analogous to that of Hajek (1970) and Beran (1977) are considered.

2. Two norms for comparing the mle and $\hat{p}_n$.

A comparison of the limiting behavior of the mle $\hat{\pi}_n$ and the nonparametric estimator $\hat{p}_n$ requires that a distance between $\hat{\pi}_n$ and π_0 and between $\hat{p}_n$ and π_0 be defined. We choose to work with the norm $\| \cdot \|_0$ on $\mathbb{R}^\infty$ defined by $\| x \|_0 = (\sum_{i=0}^\infty x_i^2/\pi_{i,0})^{1/2}$ and its empirical analogue $\| \cdot \|_n$ defined by $\| x \|_n = (\sum_{i=0}^\infty x_i^2/\hat{\pi}_{i,n})^{1/2}$. Here π_0 is an arbitrary but fixed member of the set $\mathscr{M}(M)$ of mixtures of Poisson distributions with rates in $[0, M]$ for some known finite bound M. To avoid trivialities, π_0 is assumed not to be the degenerate Poisson distribution with rate zero. Since $\| \pi_1 - \pi_0 \|_0^2 = \sum (\pi_{i,1} - \pi_{i,0})^2/\pi_{i,0}$, the norm $\| \cdot \|_0$ can be thought of as a kind of χ^2-distance.

Let $\mathscr{C} = \{c \in \mathbb{R}^\infty : \| c \|_0 < \infty\}$. Note that $\hat{p}_n \in \mathscr{C}$ for all n and $\pi \in \mathscr{C}$ for all $\pi \in \mathscr{M}(M)$. An inner product $\langle \cdot, \cdot \rangle_0$ is defined on $\mathscr{C}$ by $\langle x, y \rangle_0 = \sum x_i y_i/\pi_{i,0}$. An empirical inner product $\langle \cdot, \cdot \rangle_n$ is defined on the set of vectors in $\mathbb{R}^\infty$ for which $\| x \|_n$ is finite by $\langle x, y \rangle_n = \sum x_i y_i/\hat{\pi}_{i,n}$.

Some interesting properties of $\hat{\pi}_n$ and $\hat{p}_n$ can be described in terms of the inner products $\langle \cdot, \cdot \rangle_0$ and $\langle \cdot, \cdot \rangle_n$. For a mixed Poisson probability vector π, the log likelihood $\mathscr{L}$ at π is proportional to

$$\ell(\pi) = \textstyle\sum_{i=0}^\infty \hat{p}_{i,n} \log \pi_i.$$

The directional derivative of ℓ at a mixed Poisson probability vector π_1 in the direction of a second mixed Poisson probability vector π_2 is defined by

$$D(\pi_1; \pi_2) = \lim_{\varepsilon \downarrow 0} \varepsilon^{-1} \langle \ell((1 - \varepsilon)\pi_1 + \varepsilon\pi_2) - \ell(\pi_1) \rangle = \textstyle\sum \hat{p}_{i,n}((\pi_{i,2} - \pi_{i,1})/\pi_{i,1}).$$

The mle $\hat{\pi}_n$ satisfies $D(\hat{\pi}_n; \pi) \le 0$ for any other mixed Poisson vector π. In terms of the empirical inner product the directional derivative at $\hat{\pi}_n$ can be written as

$$(2.1)\quad D(\hat{\pi}_n; \pi) = \langle \hat{p}_n, \pi - \hat{\pi}_n \rangle_n = \langle \hat{p}_n - \hat{\pi}_n, \pi - \hat{\pi}_n \rangle_n = \langle \hat{p}_n - \hat{\pi}_n, \pi \rangle_n.$$

Therefore, for any n

$$\| \hat{p}_n - \pi_0 \|_n^2 = \| \hat{p}_n - \hat{\pi}_n \|_n^2 + \| \hat{\pi}_n - \pi_0 \|_n^2 + 2\langle \hat{p}_n - \hat{\pi}_n, \hat{\pi}_n - \pi_0 \rangle_n$$

$$(2.2) \qquad = \| \hat{p}_u - \hat{\pi}_n \|_n^2 + \| \hat{\pi}_n - \pi_0 \|_0^2 - 2D(\hat{\pi}_n, \pi_0)$$

$$\geq \| \hat{p}_n - \hat{\pi}_n \|_n^2 + \| \hat{\pi}_n - \pi_0 \|_n^2.$$

Thus, in finite samples the mle $\hat{\pi}_n$ is closer to the underlying mixed Poisson probability vector π_0 than is $\hat{p}_n$ when distance is measured according to the empirical norm $\| \cdot \|_n$.

Finally, note that $\mathscr{L}$, the set of $c \in \mathbb{R}^\infty$ for which $\| c \|_0 < \infty$, is identical to the set of vectors c for which the quantity $n^{1/2}\langle \hat{p}_n, c\rangle_0$ has finite variance for all n, and for $c \in \mathscr{L}$ the variance of $n^{1/2}\langle \hat{p}_n, c\rangle_0$ is $\sigma^2(c)$, where

$$(2.3) \qquad \sigma^2(c) = \| c \|_0^2 - \langle \pi_0, c \rangle_0^2.$$

To see this, note that if we write $h(i) = c_i/\pi_{i,0}$ then $\langle \hat{p}_n, c\rangle_0 = \sum_j h(X_j)/n$ where $\{h(X_j)\}$ are iid random variables with mean $\langle \pi_0, c\rangle_0$ and variance $\sigma^2(c)$. Thus for any $c \in \mathscr{L}$ the quantity $n^{1/2}\langle \hat{p}_n - \pi_0, c\rangle_0$ has a limiting normal distribution with mean zero and variance $\sigma^2(c)$.

3. Consistency of the mle and $\hat{p}_n$. One source of difficulty in analyzing the asymptotic behavior of the mle $\hat{\pi}_n$ is that $n^{1/2}\| \hat{p}_n - \pi_0 \|_0$ diverges as n increases. Nevertheless, both $\| \hat{p}_n - \pi_0 \|_0$ and $\| \hat{\pi}_n - \pi_0 \|_0$ tend to zero in probability faster than $n^{1/2-\varepsilon}$ for any $\varepsilon > 0$.

PROPOSITION 3.1. *For any $\varepsilon > 0$ the four quantities* (i) $n^{1/2-\varepsilon}\| \hat{p}_n - \pi_0 \|_0$, (ii) $n^{1/2-\varepsilon}\| \hat{p}_n - \pi_0 \|_n$, (iii) $n^{1/2-\varepsilon}\| \hat{\pi}_n - \pi_0 \|_0$ *and* (iv) $n^{1/2-\varepsilon}\| \hat{\pi}_n - \pi_0 \|_n$ *tend to zero in probability. More generally, the convergence of* (i) *holds for any infinite dimensional multinomial vector π_0 for which the moment generating function $\sum \pi_{j,0}e^{jt}$ is finite for some $t > 0$.*

PROOF. Since $G_0(0) < 1$, there are positive constants b, B and m such that $(*)$ $bf_i(m) < \pi_{i,0} < Bf_i(M)$ for all i. Furthermore, since $\hat{G}_n$ is a consistent estimator of G_0, $(**)$ almost surely $bf_i(m) < \hat{\pi}_{i,n} < Bf_i(M)$ for all i for n sufficiently large.

First consider (i). Fix an $\varepsilon > 0$, choose a $\gamma > 0$ in $(1 - 2\varepsilon, 1)$ and an $a > 1$. In view of $(*)$, $\sum \pi_{i,0}a^i < \infty$. Set $A = a^{(1-\gamma)/3}$. By Holder's Inequality,

$$n^{1-2\varepsilon}\| \hat{p}_n - \pi_0 \|_0^2 = n^{1-2\varepsilon} \sum (\hat{p}_{i,n} - \pi_{i,0})^2\pi_{i,0}^{-1}A^{-i}A^i$$

$$\leq n^{1-2\varepsilon}[\sum (\hat{p}_{i,n} - \pi_{i,0})^2\pi_{i,0}^{-1}A^{-i/\gamma}]^\gamma[\sum (\hat{p}_{i,n} - \pi_{i,0})^2\pi_{i,0}^{-1}A^{i/(1-\gamma)}]^{1-\gamma}.$$

Since $A > 1$, we have $nE[\sum (\hat{p}_{i,n} - \pi_{i,0})^2\pi_{i,0}^{-1}A^{-i/\gamma}] \leq \sum A^{-i/\gamma} < \infty$, and thus

$$\sum (\hat{p}_{i,n} - \pi_{i,0})^2\pi_{i,0}^{-1}A^{-i/\gamma} = O_p(n^{-1}).$$

So it suffices to show that $\sum (\hat{p}_{i,n} - \pi_{i,0})^2\pi_{i,0}^{-1}A^{i/(1-\gamma)} = \sum (\hat{p}_{i,n} - \pi_{i,0})^2\pi_{i,0}^{-1}a^{i/3}$ is bounded in probability.

Expanding the quadratic gives

$$(3.1)\qquad \sum (\hat{p}_{i,n} - \pi_{i,0})^2 \pi_{i,0}^{-1} a^{i/3} \le \sum \hat{p}_{i,n}^2 \pi_{i,0}^{-1} a^{i/3} + \sum \pi_{i,0}^2 a^{i/3}.$$

Since $a > 1$, the second term on the right is bounded by $\sum \pi_{i,0} a^i$ which is finite by $(*)$. For any fixed $k > 0$, define E_n to be the event $\{\hat{p}_{i,n} > k\pi_{i,0} a^{i/3}$ for some $i\}$. Then by Markov's inequality

$$P(E_n) \le \sum P\{\hat{p}_{i,n} > k\pi_{i,0} a^{i/3}\} \le k^{-1} \sum E(\hat{p}_{i,n}) \pi_{i,0}^{-1} a^{-i/3}$$
$$= k^{-1} \sum a^{-i/3} = k^{-1}(1 - a^{-1/3})^{-1}.$$

Thus $P(E_n)$ can be made arbitrarily small by choosing k sufficiently large. On the complement of E_n, $\sum \hat{p}_{i,n}^2 \pi_{i,0}^{-1} a^{i/3} \le k^2 \sum \pi_{i,0} a^i < \infty$. Thus the first term on the right of (3.1) is bounded in probability, and the convergence of (i) is proved.

Next, consider (ii). For ε and γ as above and all large n,

$$\begin{aligned} n^{1-2\varepsilon} \| \hat{p}_n - \pi_0 \|_n^2 &= n^{1-2\varepsilon} \sum (\hat{p}_{i,n} - \pi_{i,0})^2 \pi_{i,0}^{-1} (\pi_{i,0}/\hat{\pi}_{i,n}) \\ &\le n^{1-2\varepsilon} B b^{-1} \sum (\hat{p}_{i,n} - \pi_{i,0})^2 \pi_{i,0}^{-1} (M/m)^i \\ &\le n^{1-2\varepsilon} B b^{-1} \| \hat{p}_n - \pi_0 \|_0^{2\gamma} [\sum (\hat{p}_{i,n} - \pi_{i,0})^2 \pi_{i,0}^{-1} \langle (M/m) \rangle^{i/(1-\gamma)}]^{1-\gamma} \end{aligned}$$

where the first inequality follows from $(*)$ and $(**)$ and the second inequality follows from Holder's inequality. The same argument used to show (i) implies that the term in square brackets is bounded in probability. Consequently, (ii) follows from (i).

Now consider (iv). Inequality (2.2) implies that $\| \hat{\pi}_n - \pi_0 \|_n^2 \le \| \hat{p}_n - \pi_0 \|_n^2$, and thus (iv) follows from (ii).

Finally, with ε and γ as above, for large n

$$\begin{aligned} n^{1-2\varepsilon} \| \hat{\pi}_n - \pi_0 \|_0^2 &= n^{1-2\varepsilon} \sum (\hat{\pi}_{i,n} - \pi_{i,0})^2 \hat{\pi}_{i,n}^{-1} (\pi_{i,0}/\hat{\pi}_{i,n})^{-1} \\ &\le n^{1-2\varepsilon} B b^{-1} \sum (\hat{\pi}_{i,n} - \pi_{i,0})^2 \hat{\pi}_{i,n}^{-1} (M/m)^i \\ &\le n^{1-2\varepsilon} B b^{-1} \| \hat{\pi}_n - \pi_0 \|_n^{2\gamma} [\sum (\hat{\pi}_{i,n} - \pi_{i,0})^2 \hat{\pi}_{i,n}^{-1} (M/m)^{i/(1-\gamma)}]^{1-\gamma}. \end{aligned}$$

The term in square brackets is bounded by

$$\sum \hat{\pi}_{i,n} (M/m)^{i/(1-\gamma)} + \sum \pi_{i,0}^2 \hat{\pi}_{i,n}^{-1} (M/m)^{i/(1-\gamma)},$$

which by $(*)$ and $(**)$ is bounded for all large n by $2Bb^{-1} \sum \pi_{i/0} (M/m)^{2i/(1-\gamma)}$, which is finite. Thus, (iii) follows from (iv) and the proof is complete.

4. Asymptotic normality. In this section a subset $\mathscr{S}_1$ of $\mathscr{S}$ of vectors c such that $\langle n^{1/2}(\hat{\pi}_n - \hat{p}_n), c \rangle_0$ tends to zero in probability is given. For $c \in \mathscr{S}_1$ the quantities $\langle n^{1/2}(\hat{p}_n - \pi_0), c \rangle_0$ and $\langle n^{1/2}(\hat{\pi}_n - \pi_0), c \rangle_0$ thus have the same limiting normal distribution. The set $\mathscr{S}_1$ is shown to be dense in $\mathscr{S}$ when the support of G_0 is an infinite set, but the convergence result on $\mathscr{S}_1$ is not uniform and thus an immediate extension to all of $\mathscr{L}$ does not seem possible. Each unit vector e_i, however, can be shown to belong to $\mathscr{L}_1$ when G_0 has a certain behavior at zero. The required behavior is prescribed in Condition 4.1. Thus, $n^{1/2}(\hat{\pi}_n - \pi_0)$ and $n^{1/2}(\hat{p}_n - \pi_0)$ have the same finite dimensional limiting normal distributions when G_0 satisfies Condition 4.1.

An important step towards establishing the limiting behavior of $\hat{\pi}_n$ is provided by the following lemma.

LEMMA 4.1. *For any* $\varepsilon > 0$, $n^{1-\varepsilon}\langle \hat{\pi}_n - \hat{p}_n, \pi_0 \rangle_n$ *tends to zero in probability.*

PROOF. Recall from inequality (2.1) that $\langle \hat{\pi}_n - \hat{p}_n, \pi_0 \rangle_n = -D(\hat{\pi}_n; \pi_0) \geq 0$. Concavity of the log-likelihood ℓ along with the observations of Section 2 imply that

$$0 \leq -n^{1-\varepsilon} D(\hat{\pi}_n; \pi_0) \leq n^{1-\varepsilon} \sum \hat{p}_{i,n} \log(\hat{\pi}_{i,n}/\pi_{i,0}) \leq n^{1-\varepsilon} D(\pi_0; \hat{\pi}_n)$$

$$= n^{1-\varepsilon}\langle \hat{p}_n - \pi_0, \hat{\pi}_n - \pi_0 \rangle_0 \leq n^{1-\varepsilon} \| \hat{p}_n - \pi_0 \|_0 \| \hat{\pi}_n - \pi_0 \|_0,$$

which tends to zero by Proposition 3.1. □

Before proceeding, we give a brief outline of our approach. Our objective is to substitute ½ for ε, e_i for π_0 and $\langle \cdot, \cdot \rangle_0$ for $\langle \cdot, \cdot \rangle_n$ in the statement of Lemma 4.1. The major difficulty lies in substituting e_i for π_0. Our approach is to first consider a class $\mathscr{L}_0$ of vectors c for which $n^{1/2}\langle \hat{p}_n - \hat{\pi}_n, c \rangle_0$ tends to zero in probability. In particular, we consider the class $\mathscr{L}_0$ of all vectors in $\mathscr{L}$ for which there is a bounded function h such that the ith component of c can be written as $\int f_i(\lambda) h(\lambda)\, dG_0(\lambda)$, where $f_i(\lambda)$ is the probability of a count of i under the Poisson distribution with mean λ. Denote the vector in $\mathscr{L}_0$ corresponding to the function h by $c(h)$. $\mathscr{L}_0$ is dense in $\mathscr{L}$ in terms of $\langle \cdot, \cdot \rangle_0$, but the convergence of $\langle n^{1/2}(\hat{p}_n - \hat{\pi}_n), c(h) \rangle_0$ to zero in probability apparently depends on the function h (see the proof of Theorem 4.1). Therefore, the convergence may not be uniform and need not hold for all of C even though $\mathscr{L}_0$ is dense in $\mathscr{L}$. Since the unit vector e_i does not belong to $\mathscr{L}_0$, convergence in $\mathscr{L}_0$ thus does not yet imply convergence of $n^{1/2}\langle \hat{p}_n - \hat{\pi}_n, e_i \rangle_0$ to zero in probability.

The convergence of $n^{1/2}\langle \hat{p}_n - \hat{\pi}_n, c \rangle_0$ to zero in probability does hold, however, for a larger class than $\mathscr{L}_0$. In particular, this convergence holds for those $c \in \mathscr{L}$ that can be approximated by a sequence $c(h_k)$ of vectors in $\mathscr{L}_0$ that grow at most linearly in k (specifically, we assume that $|h_k(\lambda)| \leq k$ for all $\lambda \in [0, M]$) and approach c sufficiently fast (that is, $\| c - c(h_k) \|_0 = O(k^{-\beta})$ for some $\beta > 0$). Denote the set of vectors c that can be so approximated by $\mathscr{L}_1$. Theorem 4.1, which establishes this convergence result for $\mathscr{L}_1$, then implies that $n^{1/2}\langle \hat{p}_n - \pi_0, c \rangle_0$ and $n^{1/2}\langle \hat{\pi}_n - \pi_0, c \rangle_0$ have the same limiting distribution if $c \in \mathscr{L}_1$.

The final step is to show that e_i belongs to $\mathscr{L}_1$ by constructing a sequence of appropriate approximations $c(h_k)$ to e_i. First note that differentiation of each component of $f(\lambda)$ i times and evaluation of the derivatives at zero gives a vector that has a one in the ith position and zeros elsewhere, in other words, the vector

e_i. Consequently we have a linear mapping H such that $H(f(\cdot)) = e_i$. It suffices to show that this H can be suitably approximated by mappings $\hat{H}$ of the form $\hat{H}(f)_j = \int f_j(\lambda)h(x)\,dG_0(x)$ for some bounded function h. To ensure that such $\hat{H}$'s exist, we impose Condition 4.1 on G_0. The desired result is then proved as Theorem 4.2. Whether Condition 4.1 is necessary as well as sufficient is not known.

THEOREM 4.1. *For $c \in \mathscr{C}_1$, $\langle n^{1/2}(\hat{\pi}_n - \hat{p}_n), c\rangle_0$ tends to zero in probability and, thus, $\langle n^{1/2}(\hat{\pi}_n - \pi_0), c\rangle_0$ has a limiting normal distribution with mean zero and variance $\sigma^2(c)$. If the support of G_0 is an infinite subset of $[0, M]$, then $\mathscr{C}_0$ (and hence $\mathscr{C}_1$) is dense in $\mathscr{C}$ under $\langle \cdot, \cdot\rangle_0$.*

PROOF. Choose and fix $c \in \mathscr{C}_1$, its corresponding sequence $c_k \in \mathscr{C}_0$ and its exponent $\beta > 0$. Fix an $\varepsilon \in (0, \frac{1}{2})$ and let $k(n)$ be a sequence of integers such that $k(n) \sim n^{1/2-\varepsilon}$. Now write

$$\langle n^{1/2}(\hat{\pi}_n - \hat{p}_n), c\rangle_0 = \langle n^{1/2}(\hat{\pi}_n - \hat{p}_n), c - c_{k(n)}\rangle_0 + \langle n^{1/2}(\hat{\pi}_n - \hat{p}_n), c_{k(n)}\rangle_0. \tag{4.1}$$

The first term on the right is bounded by

$$|\langle n^{1/2}(\hat{\pi}_n - \hat{p}_n), c - c_{k(n)}\rangle_0| \le n^{1/2}(\|\hat{\pi}_n - \pi_0\|_0 + \|\hat{p}_n - \pi_0\|_0)\|c - c_{k(n)}\|_0.$$

Since $\|c - c_{k(n)}\|_0 = O(n^{-\beta(1/2-\varepsilon)})$, this tends to zero in probability by Proposition 3.1. The second term on the right of (4.1) can be written as

$$\begin{aligned} &\langle n^{1/2}(\hat{\pi}_n - \hat{p}_0), c_{k(n)}\rangle_0 \\ &\quad = \langle n^{1/2}(\hat{\pi}_n - \hat{p}_{i,n}), c_{k(n)}\rangle_n - n^{1/2}\sum(\hat{\pi}_{i,n} - \hat{p}_{i,n})(\pi_{i,0} - \hat{\pi}_{i,n})\frac{c_{i,k(n)}}{\pi_{i,0}\hat{\pi}_{i,n}}. \end{aligned} \tag{4.2}$$

Since $|c_{i,k(n)}| \le k(n)\pi_{i,0}$, the second term on the right of (4.2) is no larger in absolute value than

$$\begin{aligned} &n^{1/2}\sum|\hat{\pi}_{i,n} - \hat{p}_{i,n}|\,|\hat{\pi}_{i,n} - \pi_{i,0}|\,(k(n)\pi_{i,0}/\pi_{i,0}\hat{\pi}_{i,n}) \\ &\qquad \le n^{1/2}k(n)\|\hat{\pi}_n - \hat{p}_n\|_n\|\hat{\pi}_n - \pi_0\|_n - n^{1-\varepsilon}\|\hat{\pi}_n - \hat{p}_n\|_n\|\hat{\pi}_n - \pi_0\|_n, \end{aligned}$$

which tends to zero in probability by Proposition 3.1. Finally, for the first term on the right of (4.2)

$$\begin{aligned} |\langle n^{1/2}(\hat{\pi}_n - \hat{p}_n), c_{k(n)}\rangle_n| &= \left|\int \langle n^{1/2}(\hat{\pi}_n - \hat{p}_n), f(\lambda)\rangle_n h_{k(n)}(\lambda)\,dG_0(\lambda)\right| \\ &\le \int |\langle n^{1/2}(\hat{\pi}_n - \hat{p}_n), f(\lambda)\rangle_n|\,|h_{k(n)}(\lambda)|\,dG_0(\lambda) \\ &\le n^{1/2}k(n)\int |\langle \hat{\pi}_n - \hat{p}_n, f(\lambda)\rangle_n|\,dG_0(\lambda). \end{aligned}$$

Now $\langle \hat{\pi}_n - \hat{p}_n, f(\lambda)\rangle_n = -D(\hat{\pi}_n, f(\lambda)) \ge 0$ for all $\lambda \in [0, M]$ by (2.1) with $\pi = f(\lambda)$.

So

$$n^{1/2}k(n)\int |\langle \hat{\pi}_n - \hat{p}_n, f(\lambda)\rangle_n|\, dG_0(\lambda)$$
$$= n^{1/2}k(n)\int \langle \hat{\pi}_n - \hat{p}_n, f(\lambda)\rangle_n\, dG_0(\lambda)$$
$$= n^{1/2}k(n)\langle \hat{\pi} - \hat{p}_n, \pi_0\rangle_n \sim n^{1-\varepsilon}\langle \hat{\pi}_n - \hat{p}_n, \pi_0\rangle_n,$$

since $\langle \hat{\pi}_n - \hat{p}_n, f(\lambda)\rangle_n \geq 0$ for all $\lambda \in [0, M]$ by (2.1) with $\pi = f(\lambda)$. Lemma 4.1 implies that the final term tends to zero in probability.

To show that $\mathscr{L}_0$ is dense in $\mathscr{L}$ suppose that $z \in \mathscr{L}$ and that z is orthogonal to $\mathscr{L}_0$ in the inner product $\langle \cdot, \cdot \rangle_0$. Then for any $c = c(h) \in \mathscr{L}_0$, it holds that $\sum c_j z_j/\pi_{j,0} = \int \sum z_j f_j(\lambda)h(\lambda)\, dG_0(\lambda)/\pi_{j,0} = 0$. Hence, $\sum z_j f_j(\lambda)/\pi_{j,0} = 0$ for G_0-almost all λ. Therefore, if the support of G_0 is an infinite subset of $[0, M]$, $z_j = 0$ for all j. □

In order to show that $e_i \in \mathscr{L}_1$, we impose the following condition on G_0.

CONDITION 4.1. There exist positive constants d, β, ε such that $G_0(x + y) - G_0(x) \geq dy^\beta$ for all x, y in $(0, \varepsilon)$.

Note that Condition 4.1 implies that the support of G_0 is an infinite set. Condition 4.1 is satisfied if the derivative $G_0'(x)$ exists and is continuous for $x > 0$ in some neighborhood of the origin and $G_0'(x) \geq d$ in a neighborhood of the origin. More generally, Condition 4.1 is satisfied if for some k the derivatives $G_0^{(1)}(x) = dG_0(x)/dx, \cdots, G_0^k(x) = d^kG_0(x)/dx^k$ exist and are continuous in some neighborhood of the origin and $G_0^{(1)}(0) = \cdots = G_0^{(k-1)}(0) = 0$, but $G_0^{(k)}(0) > 0$. These requirements on G_0 are sufficient even if $G_0(0) > 0$.

Recall that the mapping $H(f)$ described above involves differentiation of f i times and evaluation of the derivatives at the origin. Also recall that

$$k^i \sum_{r=0}^{i} \binom{i}{r}(-1)^{i-r} g\left(\frac{r}{k}\right) \to \frac{d^i}{dx^i} g(x)\bigg|_{x=0}$$

as $k \to \infty$ if g is i times differentiable near zero. The latter fact is exploited to construct mappings $\hat{H}(f)$ into $\mathscr{L}_0$ that approximate H appropriately (cf. the proof of Theorem 4.2). The proof requires the following simple combinatorial lemma (cf. problem 16 on page 65 of Feller, 1968.)

LEMMA 4.2. *If* $r \leq n$ *then* $(1/n!)\sum_{k=0}^{n} (-1)^{n-k}\binom{n}{k}k^r$ *equals* 0 *if* $r < n$ *and* 1 *if* $r = n$.

THEOREM 4.2. *Suppose* G_0 *satisfies Condition* 4.1. *Then for any* i *the vector* e *belongs to* C_1. *Thus* $n^{1/2}(\hat{\pi}_{i,n} - \pi_{i,0})$ *and* $n^{1/2}(\hat{p}_{i,n} - \pi_{i,0})$ *have the same finite dimensional limiting normal distributions.*

PROOF. Fix i and define the constants

$$\lambda_{r,k} = \begin{cases} k^i\binom{i}{r}(-1)^{i-r} & r \le i \\ 0 & \text{otherwise} \end{cases}$$

for $k = i, i+1, \cdots$ and $r = 0, 1, \cdots, k$. Define a set of i intervals near the origin by $I(r, k) = [r/k, (r/k) + (1/k^{(i+1)})]$ and set

$$h_k(x) = \sum_{r=0}^{i} \lambda_{r,k} 1_{I(r,k)}(x)\left(\int_{I(r,k)} e^{-\lambda}\, dG_0(\lambda)\right)^{-1}.$$

Once k is large enough, h_k is finite and bounded by a constant times $k^{i+\beta(i+1)}$. Thus it is enough to show that $\| e_i - c(h_k) \|_0 = O(k^{-\gamma})$ for some $\gamma > 0$. To show this, set $c_k = c(h_k)$, let $c_{j,k}$ denote the jth component of c_k, and note that the mean value theorem implies

$$c_{j,k} = \int f_j(\lambda)h_k(\lambda)\, dG_0(\lambda) = \frac{1}{j!}\int \lambda^j h_k(\lambda)e^{-\lambda}\, dG_0(\lambda) = \frac{1}{j!}\sum_{r=0}^{i} \lambda_{r,k}\left(\frac{r}{k} + \xi_{r,k}\right)^j$$

for some $\xi_{r,k} \in [0, k^{-(i+1)}]$; the dependence of the ξ's on j has been suppressed. We now show that $c_{0,k}, \cdots, c_{i-1,k}$, $1 - c_{i,k}$, and $\sum_{j=i+1}^{\infty} c_{j,k}^2/\pi_{j,0}$ are all $O(k^{-1})$, which gives $\| e_i - c(h_k) \|_0 = O(k^{-1})$ as needed. Consider the values $j \le i$. For some $\eta_{r,k} \in [0, k^{-(i+1)}]$,

$$c_{j,k} = \frac{1}{j!}\sum_{r=0}^{i} \lambda_{r,k}\frac{r^j}{k^j} + \frac{j}{j!}\sum_{r=0}^{i} \lambda_{r,k}\xi_{r,k}\left(\frac{r}{k} + \eta_{r,k}\right)^{j-1}$$

(the coefficient of the second term is written as $j/j!$ to allow for the case $j = 0$). Since $\xi_{r,k} = O(k^{-(i+1)})$, the second term on the right is $O(k^{-1})$. By Lemma 4.2 the first term equals zero if $j < i$ and one if $j = i$. So $c_{0,k}, \cdots, c_{i-1,k}$, $1 - c_{i,k}$ are $O(k^{-1})$. For $j > i$, the $c_{j,k}$ are bounded by

$$| c_{j,k} | \le \frac{1}{j!}\frac{(i+1)^i 2^i}{k^{j-1}} \le \frac{1}{k}\frac{(i+1)^j 2^i}{j!}.$$

Hence,

$$\sum_{j=i+1}^{\infty} \frac{c_{j,k}^2}{\pi_{j,0}} \le \frac{2^{2i}}{k^2}\sum_{j=0}^{\infty} \frac{(i+1)^{2j}}{(j!)^2\pi_{j,0}}.$$

The sum on the right is finite since there are positive numbers b and m such that $\pi_{i,0} \ge bf_i(m)$ for all i. Thus $\sum_{j=i+1}^{\infty} c_{j,k}^2/\pi_{i,0} = O(k^{-1})$ as well, and therefore $\| e_i - c(h_k) \|_0 = O(k^{-1})$. □

5. Some remarks. The results in Section 4 can be extended to functionals T on $\mathscr{S}$ that are smooth at π_0 in the sense that there is a $c \in \mathscr{S}$ such that

$$T(x) - T(\pi_0) = \langle x - \pi_0, c\rangle_0 + O(\| x - \pi_0 \|_0^{\gamma}) \tag{5.1}$$

for some $\gamma > 1$. In particular, if T satisfies (5.1) for $c \in \mathscr{S}_1$, then $n^{1/2}(T(\hat{\pi}_n) - T(\pi_0))$ and $n^{1/2}\langle \hat{p}_n - \pi_0, c\rangle_0$ both have limiting normal distributions with mean zero and variance $\sigma^2(c)$ where $\sigma^2(c)$ is given by (2.3). If, as we conjecture, the more general result that $n^{1/2}\langle\hat{\pi}_n - \hat{p}_n, c\rangle_0$ has a limiting normal distribution for all $c \in \mathscr{S}$ when G_0 has infinite bounded support is true, then $n^{1/2}(T(\hat{\pi}_n) - T(\pi_0))$ is asymptotically normal $(0, \sigma^2(c))$ for any T satisfying (5.1) for some $c \in \mathscr{S}$.

For the case where G_0 has finite support consisting of N_0 points, say, certain linear combinations of the $\hat{\pi}_{i,n}$'s can be shown to be asymptotically normal. For example, Theorem 4.1 implies the following Corollary 5.1 which applies whenever λ is a support point in $(0, \infty)$ of G_0.

COROLLARY 5.1. *If* λ *is such that* (i) $[G_0(\lambda + \varepsilon) - G_0(\lambda - \varepsilon)]^{-1} = O(\varepsilon^\beta)$ *for some* $\beta > 0$, *then* $f(\lambda) \in \mathscr{S}_1$ *and thus* $\langle n^{1/2}(\hat{\pi}_n - \pi_0), f(\lambda)\rangle_0$ *has a limiting normal distribution with mean zero and variance* $\sigma^2(f(\lambda))$. *If, in addition,* (ii) $0 < \lambda < M$, *then* $\langle n^{1/2}(\hat{\pi}_n - \pi_0), f'(\lambda)\rangle$ *has a limiting normal distribution with mean zero and variance* $\sigma^2(f'(\lambda))$, *where* $f'(\lambda)$ *is the derivative of* $f(\lambda)$.

PROOF. Suppose λ satisfies (i) for a particular β. Define h_ε by

$$h_\varepsilon(x) = [G_0(\lambda + \varepsilon) - G_0(\lambda - \varepsilon)]^{-1} 1_{[\lambda-\varepsilon,\lambda+\varepsilon]}(x).$$

Then if $c_\varepsilon = c(h_\varepsilon)$, we have

$$\| f(\lambda) - c_\varepsilon \|_0^2 = \sum \left[f_i(\lambda) - \int f_i(x) h_\varepsilon(x)\, dG_0(x) \right]^2 \Big/ \pi_{i,0}$$

$$= \sum [f_i(\lambda) - f_i(\xi_i)]^2 / \pi_{i,0}$$

for some $\xi_i \in [\lambda - \varepsilon, \lambda + \varepsilon]$ by the mean value theorem. Applying the mean value theorem a second time gives

$$\sum [f_i(\lambda) - f_i(\xi_i)]^2 \pi_{i,0}^{-1} \le \varepsilon^2 \sum f_i'(\eta_i)^2 \pi_{i,0}^{-1} \tag{5.2}$$

for some $\eta_i \in [\lambda - \varepsilon, \lambda + \varepsilon]$. Since $f_i'(\lambda) = f_{i-1}(\lambda) - f_i(\lambda)$, the sum on the right of equation (5.2) is finite, and thus $\| f(\lambda) - c_\varepsilon \|_0 \to 0$ and

$$\| f(\lambda) - c_\varepsilon \|_0 = O((G_0(\lambda + \varepsilon) - G_0(\lambda - \varepsilon))^{1/\beta})$$

as $\varepsilon \to 0$. So $f(\lambda)$ is in $\mathscr{S}_1$.

To prove the second claim, assume λ satisfies (ii) in addition to (i). Then arguments along the lines of the proof of Proposition 3.1 show that

$$\langle n^{1/2}(\hat{\pi}_n - \hat{p}_n), f(\lambda)\rangle_n - \langle n^{1/2}(\hat{\pi}_n - \hat{p}_n), f(\lambda)\rangle_0$$

tends to zero in probability. Note that $f(\lambda) \in \mathscr{S}_1$. Hence, a variant of the proof of Theorem 4.1 shows that $\langle n^{1/2}(\hat{\pi}_n - \hat{p}_n), f(\lambda)\rangle_n = -D(\hat{\pi}_n, f(\lambda))$ tends to zero in probability. Now assume $\langle n^{1/2}(\hat{\pi}_n - \hat{p}_n), f'(\lambda)\rangle_n = -dD(\hat{\pi}_n, f(\lambda))/d\lambda$ does not tend to zero in probability for some $\lambda \in (0, M)$. Then there is a subsequence n_* for which almost surely $\sum n_*^{1/2}(\hat{\pi}_{i,n_*} - \hat{p}_{i,n_*}) f_i'(\lambda)/\hat{\pi}_{i,n_*}$ is bounded away from zero but $\sum n_*^{1/2}(\pi_{i,n_*} - \hat{p}_{i,n_*}) f_i(\lambda)/\hat{\pi}_{i,n_*}$ converges to zero. Then for any large enough n_* there must exist some δ near λ such that $\langle n_*^{1/2}(\hat{\pi}_{n_*} - \hat{p}_{n_*}), f(\delta)\rangle_{n_*} = -D(\hat{\pi}_{n_*}, f(\delta)) < 0$. This contradicts the fact that $D(\hat{\pi}_n, \pi) \le 0$ for all mixed Poisson vectors π since $\hat{\pi}_n$ is the mle.

For mixing distributions G_0 with finite support, Corollary 5.1 implies that $\langle n^{1/2}(\hat{\pi}_n - \pi_0), c\rangle_0$ has a limiting normal distribution if c belongs to the finite dimensional vector space $\mathscr{L}_* = \text{span}(\{f(\lambda): \lambda \in \text{support}(G_0)\} U \{f'(\lambda): 0 < \lambda < M, \lambda \in \text{support}(G_0)\})$. For other c's, $\langle n^{1/2}(\hat{\pi}_n - \pi_0), c\rangle_0$ may not have a nondegenerate limiting normal distribution. When c corresponds to a support hyperplane of $\mathscr{M}(M)$, $\langle f(\lambda) - \pi_0, c\rangle_0 \geq 0$. For example, if $N_0 = 1$ and G_0 is supported on λ_0 then $\langle f(\lambda) - \pi_0, f''(\lambda_0)\rangle_0 \geq 0$ for all λ, so that $\langle n^{1/2}(\hat{\pi}_n - \pi_0), f''(\lambda_0)\rangle_0$ being nonnegative cannot have a nondegenerate limiting normal distribution. This reasoning suggests that if G_0 has finite support then all the finite dimensional limiting distributions of $n^{1/2}(\hat{\pi}_n - \pi_0)$ cannot be nondegenerate normal. (Nevertheless, as the numerical calculation mentioned in Section 1 suggests, the formula $(\hat{\pi}_{i,n}(1 - \hat{\pi}_{i,n}))^{1/2}$ may provide a reasonable estimate of the standard deviation of $n^{1/2}(\hat{\pi}_{i,n} - \pi_{i,0})$ even if N_0 is finite.)

It should be possible to extend the results of the previous sections to mixtures of an exponential family $\{f(x, \theta)\}$ of densities with respect to a measure μ on the real line for a parameter set that is a subset of the extended real line. Lindsay (1983a, b) gives conditions under which the mle $\hat{G}_n$ of G_0 exists, is unique and has finite support. Let $\hat{f}_n(x) = \int f(x, \theta)\, d\hat{G}_n(\theta)$ be the mle of the mixed density f. Define a norm $\| g \|_0 = (\int g^2/f_0\, d\mu)^{1/2}$, an empirical analogue

$$\| g \|_n = \left(\int g^2/\hat{f}_n\, d\mu\right)^{1/2}$$

and the associated inner products. Also define $\mathscr{L}$ to be the set of functions c for which $\| c \|_0 < \infty$, define $\sigma^2(c) = \| c \|_0^2 - \langle c, f_0\rangle_0^2$, and define the linear functional $\langle\langle H, c\rangle\rangle_0 = \int (c/f_0)\, dH$ for any signed measure for which it is meaningful. The analogue of $\langle \hat{p}_n, c\rangle_0$ would then be $\langle\langle d\hat{F}_n, c\rangle\rangle_0$, where $\hat{F}_n$ is the empirical distribution function.

Analogues of Lemma 4.1 and Theorem 4.1 for suitably defined sets $\mathscr{L}_0$ and $\mathscr{L}_1$ are straightforward to establish, and Corollary 5.1 can be extended to this more general context. But producing an analogue of Theorem 4.2 is more difficult, since for most classes of c's of practical interest, such as indicator functions, there is no obvious way to establish membership in $\mathscr{L}_1$. It may be possible, however, to use a differentiation approach to determine conditions under which polynomials belong to $\mathscr{L}$. We have not pursued this in detail.

Acknowledgement. We would like to thank the associate editor and the referees for their careful reading of the original manuscript and for their suggestions that led to substantial improvements in the presentation. We are particularly grateful to the referee whose suggestions led to significant simplifications in the proofs of Proposition 3.1 and Lemma 4.1.

REFERENCES

BERAN, R. (1977). Robust location estimates. *Ann. Statist.* **5** 431–444.

FELLER, W. (1968). *An Introduction to Probability Theory and its Applications*, vol. 1, 3rd Ed. Wiley, New York.

HAJEK, J. (1970). A characterization of limiting distributions of regular estimates. *Z. Wahrsch. verw. Gebiete* **14** 323–330.

JEWELL, N. (1982). Mixtures of exponential distributions. *Ann. Statist.* **10** 479–484.

KARLIN, S. J. and STUDDEN, W. J. (1966). *Tchebycheff Systems: With Applications in Analysis and Statistics*. Wiley, New York.

LAIRD, N. (1978). Nonparametric maximum likelihood estimation of a mixing distribution. *J. Amer. Statist. Assoc.* **73** 805–808.

LINDSAY, B. G. (1983a). The geometry of mixture likelihoods: A general theory. *Ann. Statist.* **11** 86–94.

LINDSAY, B. G. (1983b). The geometry of mixture likelihoods, part II: The exponential family. *Ann. Statist.* **11** 783–792.

SIMAR, L. (1976). Maximum likelihood estimation of a compound Poisson process. *Ann. Statist.* **4** 1200–1209.

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

TIERNEY, L. and LAMBERT, D. (1984). The asymptotic efficiency of estimators of functionals of mixed distributions. *Ann. Statist.* **12** 1380–1387.

DEPARTMENT OF STATISTICS
CARNEGIE-MELLON UNIVERSITY
PITTSBURGH, PENNSYLVANIA 15213

DEPARTMENT OF THEORETICAL STATISTICS
UNIVERSITY OF MINNESOTA
206 CHURCH STREET, S.E.
MINNEAPOLIS, MINNESOTA 55455