Department of Statistical Science

Southern Methodist University

Technical Report SMU-TR-326

December 2004

The Minimum L_2 Distance Estimator for Poisson Mixture Models

Shuyi Shen

Eli Lilly & Company

Indianapolis, IN 46256

(shensh@lilly.com)

Ian R. Harris

Department of Statistical Science, Southern Methodist University

Dallas, TX 75275

(iharris@smu.edu)

A robust estimator is developed for Poisson mixture models with known number of components. The proposed estimator minimizes the L_2 distance between the data and the model. When the component distributions are completely known, the estimators for the mixing proportions are in closed form. When the parameters for the component Poisson distributions are unknown, numerical methods are needed to calculate the estimators. The method offers a robust estimator for mixture models while retaining acceptable efficiency compared to the maximum likelihood estimator. Compared to the minimum Hellinger distance estimator, the minimum L_2 estimator is less robust to extreme outliers, and more robust to moderate outliers.

SOME KEY WORDS: Divergence; Influence function; L_2 distance; Maximum likelihood; Mixing proportion; Robustness.

1 Introduction

The random variable X has a k-component mixture distribution if its density can be represented in the form

$$p(x) = \pi_1 f_1(x) + \pi_2 f_2(x) + \ldots + \pi_k f_k(x), \quad 0 \le \pi_j \le 1, \quad j = 1, \ldots, k, \quad \sum_{j=1}^{k} \pi_j = 1. \qquad (1)$$

The parameters $\pi_1, \ldots, \pi_k$ are called the *mixing proportions* or *mixing weights*, and $f_1(\cdot), \ldots, f_k(\cdot)$ are the component densities of the mixture model. Theoretically, $f_1(\cdot), \ldots, f_k(\cdot)$ can be from different parametric families, but in practice they usually are from the same parametric family.

Several different estimation approaches have been applied to mixture problems. The method of moments was first applied by Pearson (1894) on the crab data, which was fitted by a two-component normal mixture model. Since the EM algorithm was introduced by Dempster et al. (1977), the maximum likelihood estimator (MLE) has been widely used in mixture problems. Like most of the maximum likelihood estimators, under some regularity conditions, the MLE is consistent for the parameters estimated, asymptotically efficient and normally distributed with the variance as the inverse of the Fisher information. The drawback of the MLE in mixture problems is that it has to be computed iteratively. When the mixing proportion is around the boundary of the parameter space or when the components are poorly separated, the convergence of iterative algorithms can be very slow. Another concern about the MLE is its sensitivity to the underlying assumptions.

When the model assumptions are satisfied, the MLE is asymptotically the most efficient estimator. However, the underlying model assumptions are often violated by real data. The existence of gross errors and even slight deviations from the parametric density can affect the performance of the MLE considerably.

If model assumptions are violated, minimum distance estimators can be more robust than the MLE. Choi and Bulgren (1968) proposed the minimum Wolfowitz distance estimator for mixing proportions with known component distributions. MacDonald (1971) and Woodward et al. (1984) examined a similar method of minimizing the Cramér-von Mises distance to estimate the mixing proportions in mixture of normal distributions. Clarke (1989) and Clarke and Heathcote (1994) developed explicit estimators for mixing proportions in mixture normal distributions by minimizing the L_2 distance between parametric and empirical distribution functions. Woodward et al. (1995) developed the minimum Hellinger distance estimator (MHDE) for two-component normal mixture models and Karlis and Xekalaki (1998) examined the MHDE of finite Poisson mixtures. The latter two papers showed that the MHDE is asymptotically normally distributed with *full efficiency* under model assumptions and more robust to departure from the underlying assumptions than the MLE. For a detailed history of finite mixture models, readers are referred to McLachlan and Peel (2001) and Titterington, Smith and Makov(1985).

Apart from the estimators of Clarke (1989) and Clarke and Heathcote (1994), the methods mentioned above give estimators for the mixing proportions that are not in explicit form. The present paper introduces a new estimator for mixture models, based on the minimum squared distance between parametric and empirical densities. When the component distributions are completely known, the proposed estimator for the mixing proportion exists in closed form. The estimation method also offers an estimator which is more robust to departure from the underlying assumptions while less efficient compared to the MLE. This is particularly appropriate for analyzing massive data sets where data cleaning is impractical and statistical efficiency is a secondary concern.

The rest of the paper is organized as follows. In section 2 the minimum L_2 distance estimator and the minimum density power divergence estimator are introduced. In section 3, the minimum L_2 distance estimator in k-component mixture models is developed, assuming only the mixing proportions are unknown. Some of its properties including robustness and asymptotic efficiency are discussed here and some asymptotic properties of the estimators are also established. Particular detail is given for the two-component case. In section 4 the minimum L_2 distance estimator is proposed for k-component Poisson mixtures with all parameters unknown. Some of the asymptotic properties are also investigated in this section. The results of some simulation studies are presented in section 5 to compare the performance of the L_2E, the MLE and the MHDE. In section 5.1 we present simulation results and a theoretical argument that suggest the L_2E is more robust to moderate outliers as compared to the MHDE. Concluding remarks and ideas for further research are presented in section 6.

2 Minimum L_2 Distance Estimator and MDPDE

Assume the data $X_1, \cdots, X_n$ are generated randomly from some distribution G with corresponding density g. The distribution G is unknown but one is willing to approximate it by an element from the parametric family $\mathcal{F}_\theta$. The L_2E for θ is obtained by minimizing the so-called L_2 distance $\int \{g(x) - f_\theta(x)\}^2 \, dx$, between the unknown density g and the parametric density f_θ. Note that the L_2 distance can be represented by $\int f_\theta^2(x) \, dx - 2\int f_\theta(x) \, dG(x) + C$, where the quantity C is independent of the parameter θ and does not affect the minimization procedure. Given a random sample $X_1, \ldots, X_n$ from the true distribution G, to obtain the L_2E of the best fitting parameter, one can actually minimize

$$\int f_\theta^2(x) \, dx - 2\int f_\theta(x) \, dG_n(x) dx = \int f_\theta^2(x) \, dx - 2n^{-1} \sum_{i=1}^n f_\theta(X_i)$$

with respect to θ, where G_n is the empirical distribution function. Under differentiability of the model and appropriate regularity conditions, the L_2E can be obtained by solving the estimating equation

$$n^{-1} \sum_{i=1}^n u_\theta(X_i) f_\theta(X_i) - \int u_\theta(x) f_\theta^2(x) \, dx = 0, \tag{2}$$

where $u_\theta(x) = \partial \log f_\theta(x)/\partial\theta$ is the maximum likelihood score function.

The minimum L_2 distance estimator (L_2E) in mixture problems was first proposed by Scott (1999), who investigated its application in normal mixture models. The L_2 distance is also an element of the family of the *density power divergence* (DPD) proposed by Basu et al. (1998). The DPD between two densities f and g is defined as

$$d_\alpha(g, f) = \int_{\mathcal{X}} \left\{ f^{1+\alpha}(x) - \left(1 + \frac{1}{\alpha}\right) g(x) f^\alpha(x) + \frac{1}{\alpha} g^{1+\alpha}(x) \right\} dx.$$

When $\alpha = 1$, the DPD is the L_2 *distance*. For $\alpha = 0$, the divergence is in the form

$$d_0(g, f) = \lim_{\alpha \to 0} d_\alpha(g, f) = \int_{\mathcal{X}} g(x) \log[g(x)/f(x)]\, dx,$$

which is also known as the *Kullback-Leibler divergence*. The minimum density power divergence estimator (MDPDE) for the parameter θ in f is obtained by choosing the value of θ in the parameter space such than the divergence d_α is minimized for a fixed value of α. With $\alpha = 0$, minimizing the *Kullback-Leibler divergence* is equivalent to maximizing $\sum_{i=1}^{n} \log f(X_i)$, which is the log-likelihood. So the MLE is the sample version of the MDPDE with tuning parameter $\alpha = 0$.

Basu et al. (1998) proposed that under mild regularity conditions, the MDPDE is consistent and asymptotically normally distributed with variance $J^{-1}KJ^{-1}$. The matrix K is defined as

$$K_\alpha(\theta) = \int u_\theta(x) u_\theta^T(x) f_\theta^{2\alpha}(x) g(x)\, dx - \xi_\theta \xi_\theta^T \quad \text{and} \quad \xi_\theta = \int u_\theta(x) f_\theta^{\alpha}(x) g(x)\, dx. \tag{3}$$

The matrix J is in the form

$$J_\alpha(\theta) = \int u_\theta(x) u_\theta^T(x) f_\theta^{1+\alpha}(x)\, dx + \int (i_\theta(x) - \alpha u_\theta(x) u_\theta^T(x))(g(x) - f_\theta(x)) f_\theta^{\alpha}(x)\, dx, \tag{4}$$

and $i_\theta(x) = -\partial\{u_\theta(x)\}/\partial\theta$, the so called information function of the model.

The single parameter α in the DPD controls the trade-off between robustness and asymptotic efficiency. As α increases, the robustness of the MDPDE increases but its asymptotic efficiency decreases. The influence function is a tool used to evaluate the local robustness of an estimator. Basu et al. (1998) showed that the influence function for the MDPDE has the form

$$I_\alpha(G, y) = J^{-1} \left\{ u_\theta(y) f_\theta^{\alpha}(y) - \xi_\theta \right\}, \tag{5}$$

and ξ_θ and J are as in (3) and (4). Besides the balance between robustness and efficiency, another appealing characteristic of the MDPDE is that it does not require a smoothing of the empirical density g_n. Thus, in contrast with other minimum distance estimators based on density, such as the MHDE, the MDPDE avoids the bandwidth selection involved with nonparametric density estimation.

3 L_2E for Mixing Proportion in k-component Mixture Models

In this section, the L_2E for the mixing proportions in k-component mixture models will be developed, assuming the component distributions are completely known. In section 3.1 to 3.3 we present the results for two-component mixture models.

3.1 L_2E for the mixing proportion in two-component models

In two-component mixture models, when only the mixing proportion is unknown, the parametric density is in the form $f(x;\pi) = \pi f_1(x) + (1-\pi)f_2(x)$, where $0 \le \pi \le 1$. By replacing the density f_θ with $f(x;\pi)$ in (2), the L_2E for π is obtained by solving the estimating equation

$$n^{-1}\sum_{i=1}^{n}[f_1(X_i) - f_2(X_i)] - \int (f_1(x) - f_2(x))f(x;\pi)\,dx = 0,$$

assuming $f(x;\pi)$ is an absolutely continuous density. For a discrete distribution the integral is replaced by a sum. By solving the estimating equation, the L_2E for π has the form

$$\hat{\pi}_n = \frac{n^{-1}\sum_{i=1}^{n}[f_1(X_i) - f_2(X_i)] - \int f_2(x)\{f_1(x) - f_2(x)\}\,dx}{\int [f_1(x) - f_2(x)]^2\,dx}. \tag{6}$$

When the model is correctly specified, the estimating equation is unbiased, and $\hat{\pi}_n$ is unbiased. The variance of $\sqrt{n}\hat{\pi}_n$ can be calculated as

$$\begin{aligned} \mathrm{Var}(\sqrt{n}\hat{\pi}_n) &= \frac{\mathrm{Var}(f_1(X) - f_2(X))}{\{\int (f_1(x) - f_2(x))^2\,dx\}^2} \\ &= \frac{\int (f_1(x) - f_2(x))^2 f(x;\pi)\,dx - \{\int (f_1(x) - f_2(x))f(x;\pi)\,dx\}^2}{\{\int [f_1(x) - f_2(x)]^2\,dx\}^2}. \end{aligned} \tag{7}$$

The variance of $(\sqrt{n}\hat{\pi}_n)$ can be consistently estimated in a sandwich fashion as

$$\hat{\mathrm{Var}}(\sqrt{n}\hat{\pi}_n) = \frac{(n-1)^{-1}\sum_{i=1}^{n}\left\{f_1(X_i) - f_2(X_i) - \left(n^{-1}\sum_{i=1}^{n}\{f_1(X_i) - f_2(X_i)\}\right)\right\}^2}{\{\int (f_1(x) - f_2(x))^2\,dx\}^2}.$$

It can be proved that when the tuning parameter $\alpha = 1$ and $0 < \pi < 1$, the regularity conditions A1-A5 in Basu et al. (1997) are satisfied in two-component mixture models, assuming only the mixing proportion π is unknown. Thus the L_2E is consistent and asymptotically normally distributed.

Proposition 3.1 In a mixture model with two known components, and $0 \le \pi \le 1$,

1. $\hat{\pi}_n$ is consistent for π, and
2. $n^{1/2}(\hat{\pi}_n - \pi)$ is asymptotically normally distributed with mean zero and variance as shown in (8).

Proof: See Appendix.

3.2 Influence Function and Asymptotic Relative Efficiency

From (5), by setting $\alpha = 1$, the influence function of the L_2E for the mixing proportion in two-component mixture models can be derived as

$$I_{L_2E}(y) = \frac{f_1(y) - f_2(y) - \int (f_1(x) - f_2(x))\, g(x)\, dx}{\int [f_1(x) - f_2(x)]^2\, dx}, \tag{8}$$

where y is the contaminating value. It is noticed that when $f_1(x) - f_2(x)$ is bounded, the influence function is also bounded. When the contaminating value $y \to \infty$, it follows that $f_1(y) - f_2(y)$ goes to zero for the Poisson and normal families and the influence function converges to

$$\lim_{y\to\infty} I_{L_2E}(y) = \frac{-\int (f_1(x) - f_2(x))\, g(x)\, dx}{\int (f_1(x) - f_2(x))^2\, dx}. \tag{9}$$

The influence function for the MLE can be obtained by setting $\alpha = 0$. From (5),

$$I_{\mathrm{MLE}}(y) = \frac{f_1(y) - f_2(y)}{g(y)} \times \left\{ \int (f_1(x) - f_2(x))^2 g(x)^{-1}\, dx \right\}^{-1}. \tag{10}$$

The second quantity in (10) is a constant with respect to y. When the contaminating value goes to infinity, both the denominator and the numerator of the first element tend to disappear. Methods for evaluating the quantity $\lim_{y\to\infty} I_{\mathrm{MLE}}(y)$ differ from density to density.

For two-component Poisson mixture models, $f_1(\cdot)$ and $f_2(\cdot)$ are Poisson distributions with parameters λ_1 and λ_2, where $\lambda_1 \neq \lambda_2$. Without loss of generality, let us assume $\lambda_1 < \lambda_2$. The limit of the influence function as $y \to \infty$ can be derived as

$$\lim_{y\to\infty} I_{\mathrm{MLE}}(y) = \frac{1}{\pi - 1} \times \left\{ \int (f_1(x) - f_2(x))^2\, f^{-1}(x;\pi)\, dx \right\}^{-1},$$

when the model is correctly specified in the sense $g(x) = \pi f_1(x) + (1-\pi) f_2(x)$. It is noticed that when $\pi = 1$, the influence function of the MLE can be unbounded.

To compare the robustness of the L_2E to that of the MLE, the influence functions of the L_2E and the MLE in certain two-component Poisson mixture models are plotted in Figure 1. To obtain a general view of the robustness of the estimators, the mixing proportion for the first component π is taken at .2 and .8. Also two different settings of the component distributions are chosen. In the first setting ($\lambda_1 = 1$, $\lambda_2 = 3$), the two components are poorly separated. The second setting ($\lambda_1 = 1$, $\lambda_2 = 8$) is designed to investigate the robustness of the estimators for models with well separated components. It is observed that the influence functions of the L_2E are closer to 0 than those of the MLE, which is particularly poor when the components are not well separated.

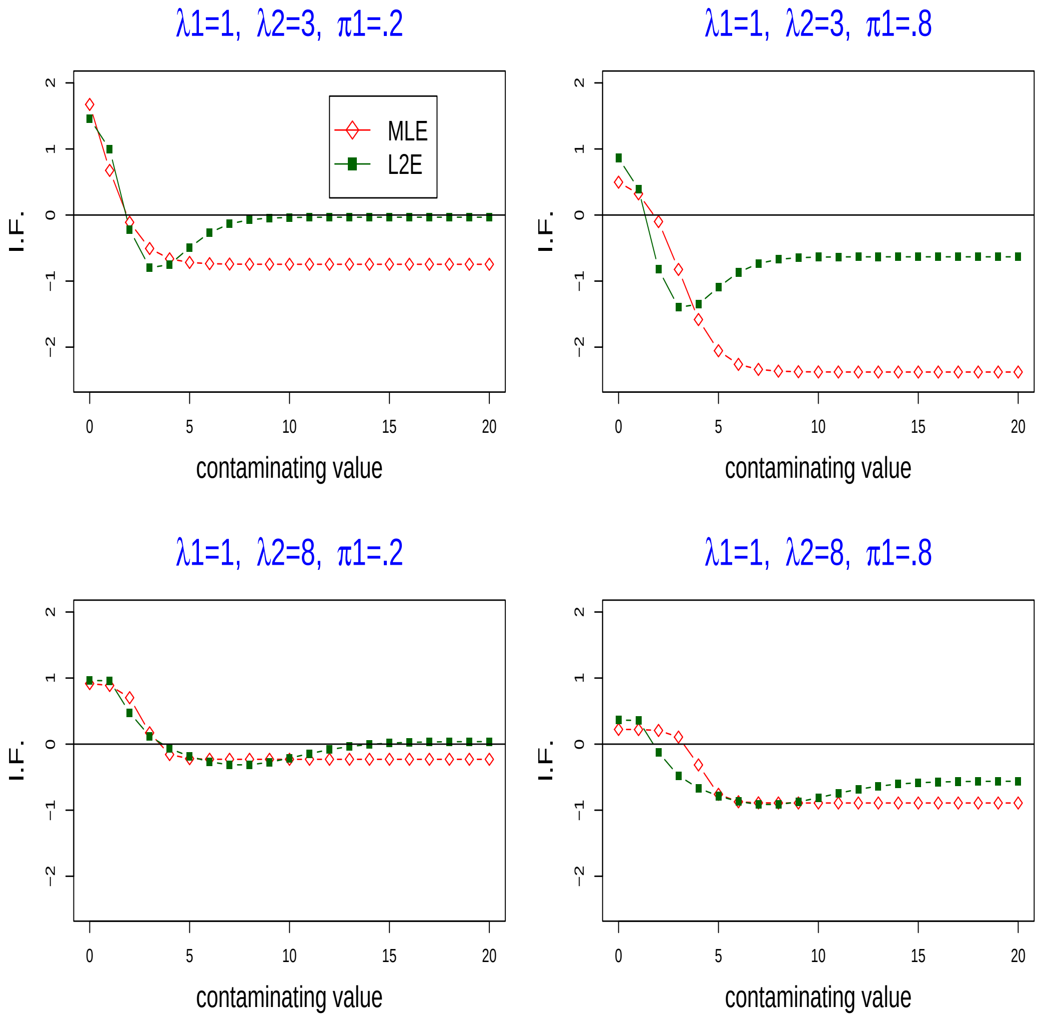

Figure 1: Influence Functions of the L_2E and the MLE for Certain 2-Component Poisson Mixture Models.

A desired estimator should be robust to the violations of the underlying assumptions and have acceptable efficiency. The Asymptotic Relative Efficiency (ARE) of the L_2E is defined as the ratio of the asymptotic variance of the MLE to that of the L_2E for the same parameter. From Basu et al. (1998), the asymptotic variance of $\sqrt{n}\hat{\theta}_{\text{MDPDE}}$ is equivalent to $J_\alpha^{-1}K_\alpha J_\alpha^{-1}$ where J_α and K_α are as shown in (3) and (4). In the two-component mixture model,

$$K_0 = J_0 = \int (f_1(x) - f_2(x))^2 f^{-1}(x;\pi)\, dx$$

and

$$\text{Var}(\sqrt{n}\hat{\pi}_{\text{MLE}}) = J_0^{-1} = \left\{ \int (f_1(x) - f_2(x))^2 f^{-1}(x;\pi)\, dx \right\}^{-1}.$$

The asymptotic variance of the L_2E for π is in the form as shown in (7).

Table 1 lists the asymptotic relative efficiencies of the L_2E in certain two-component Poisson mixture models. The mixing proportion is taken at different values from .1 to .9 to investigate its effect on the efficiency of the estimator. Several different combinations of the component distributions are selected.

Table 1: AREs of the L_2E for π in Certain 2-component Poisson Mixture Models

π	.1	.2	.3	.4	.5	.6	.7	.8	.9
$\lambda_1 = 1, \lambda_2 = 2$	.88	.88	.87	.84	.81	.77	.71	.64	.53
$\lambda_1 = 1, \lambda_2 = 3$	.85	.90	.91	.90	.88	.84	.77	.67	.50
$\lambda_1 = 1, \lambda_2 = 5$	.87	.95	.96	.95	.93	.90	.85	.75	.57
$\lambda_1 = 1, \lambda_2 = 8$	.93	.95	.95	.94	.92	.89	.85	.78	.62

In general, the L_2E performs better for models with well separated components and low mixing proportion of the first component. Also it is observed that for most of the models, the ARE of the L_2E is higher than .60.

3.3 The Truncated Estimator for the Mixing Proportion

The L_2E is obtained by minimizing the L_2 distance and the method does not guarantee that the estimator lies in the parameter space of [0,1]. Under such situations, some adjustment is necessary to keep the estimator in the appropriate parameter space. The gradient of the objective function is in the form

$$\frac{\partial^2}{\partial \pi^2} \int \{f(x;\pi) - g(x)\}^2 \, dx = \frac{\partial^2}{\partial \pi^2} \int f^2(x;\pi) \, dx - 2 \frac{\partial^2}{\partial \pi^2} \int f(x;\pi) \, dG(x).$$

When the unknown distribution $G(x)$ is estimated by its empirical distribution function $G_n(x)$,

$$\frac{\partial^2}{\partial \pi^2} \int \{f(x;\pi) - g_n(x)\}^2 \, dx = \int (f_1(x) - f_2(x))^2 \, dx.$$

It is obvious that the second derivative of the L_2 distance is positive assuming $\{x : f_1(x) \neq f_2(x)\}$ is a set of positive Lebesgue measure, and so the L_2 distance is concave and there exists a unique solution of the estimating equation. When the estimate is negative, it is clear that zero is the value with the minimum L_2 distance in the parameter space. Similarly, estimates larger than 1 can be truncated to 1.

From now on, we will denote the truncated L_2E for π as $\hat{\pi}_{\text{trunc}}$ to distinguish it from the non-truncated version $\hat{\pi}$. The truncated estimator can be expressed as

$$\begin{aligned}\hat{\pi}_{\text{trunc}} &= 0 \times I_{(-\infty,0)}(\hat{\pi}) + 1 \times I_{(1,\infty)}(\hat{\pi}) + \hat{\pi} \times I_{[0,1]} \\ &= I_{(1,\infty)}(\hat{\pi}) + \hat{\pi} \times I_{[0,1]}(\hat{\pi}),\end{aligned} \tag{11}$$

where I is an indicator function.

Theorem 3.1 *In two-component mixture model with component densities completely known, when $0 < \pi < 1$,*

1. *the truncated L_2E for π is consistent for π, and*

2. *$n^{1/2}(\hat{\pi}_{trunc} - \pi)$ is asymptotically normal with mean zero and variance as shown in (7).*

Proof: see appendix

3.4 L_2E in k-component Mixture Models

Since the sum of the mixing proportions is one, there are only $k-1$ unknown parameters.

Let $X_1, X_2, \ldots, X_n$ be a random sample from some absolutely continuous distribution G with density function (1). Assume the number of components k and all component densities $f_1(\cdot), \ldots, f_k(\cdot)$ are known. Let $\underline{\pi} = (\pi_1, \pi_2, \ldots, \pi_{k-1})^T$ be the vector of the $k-1$ unknown mixing proportions in the k-component mixture model. The results that follow also hold for discrete models as well. The L_2E for the mixing proportions $\underline{\pi}$ is obtained by minimizing

$$\int_{\mathcal{X}} f^2(x; \underline{\pi})\, dx - \frac{2}{n} \sum_{i=1}^{n} f(X_i; \underline{\pi})$$

over the parameter space

$$\Theta = \left\{ \pi_1, \pi_2, \ldots, \pi_{k-1} : 0 \leq \pi_j \leq 1, \quad j = 1, \ldots, k-1; \quad 0 \leq \sum_{j=1}^{k-1} \pi_j \leq 1 \right\}.$$

From (2), it can be easily checked that the estimating equations are in the form

$$\int_{\mathcal{X}} (f_j(x) - f_k(x)) f(x; \underline{\pi})\, dx - n^{-1} \sum_{i=1}^{n} \{f_j(X_i) - f_k(X_i)\} = 0, \quad j = 1, \ldots, k-1. \tag{12}$$

The estimating equations can also be expressed as $J\underline{\pi} = \underline{b}$, where J is a $k-1$ by $k-1$ matrix with elements in the form of

$$J_{i,j} = \int_{\mathcal{X}} (f_i(x) - f_k(x))(f_j(x) - f_k(x))\, dx; \quad i, j = 1, \ldots, k-1. \tag{13}$$

The j^{th} element of vector $\underline{b}$ is

$$b_j = \frac{1}{n} \sum_{i=1}^{n} \{f_j(X_i) - f_k(X_i)\} + \int_{\mathcal{X}} f_k(x)(f_k(x) - f_j(x))\, dx.$$

Assuming J is nonsingular, the L_2E can be derived as

$$\hat{\underline{\pi}} = J^{-1}\underline{b}. \tag{14}$$

Theoretically, the matrix J is singular only if some of the components are exactly the same as the others and hence the L_2E exists in closed form. In practice, the inverse of the matrix J cannot be obtained when some of the components are very poorly separated. When the model is correctly specified, the estimating equations and the L_2E are unbiased.

Because of the existence of the closed form of the estimator, the variance of the L_2E can be easily developed as

$$\begin{aligned} \text{Var}(\hat{\underline{\pi}}) &= \text{Var}(J^{-1}\underline{b}) \\ &= J^{-1}\text{Var}(\underline{b})(J^{-1})^T \\ &= J^{-1}\text{Var}(\underline{b})J^{-1}, \end{aligned}$$

since the matrix J is symmetric and it does not involve the random sample $X_1, \ldots, X_n$. The variance of $\underline{b}$ is a multiple of the variance-covariance matrix for $f_j(x) - f_k(x)$ and

$$K = n\text{Var}(\underline{b}) = n\text{Var}\begin{pmatrix} n^{-1}\sum_{i=1}^{n}(f_1(X_i) - f_k(X_i)) \\ n^{-1}\sum_{i=1}^{n}(f_2(X_i) - f_k(X_i)) \\ \cdot \\ \cdot \\ \cdot \\ n^{-1}\sum_{i=1}^{n}(f_{k-1}(X_i) - f_k(X_i)) \end{pmatrix} = \text{Var}\begin{pmatrix} f_1(X) - f_k(X) \\ f_2(X) - f_k(X) \\ \cdot \\ \cdot \\ \cdot \\ f_{k-1}(X) - f_k(X) \end{pmatrix}, \tag{15}$$

and the covariance matrix can be calculated under model conditions by replacing $g(x)$ with $f(x; \underline{\pi}) = \pi_1 f_1(x) + \pi_2 f_2(x) + \cdots + \left(1 - \sum_{j=1}^{k-1} \pi_j\right) f_k(x)$. The variance of ($\sqrt{n}$ times) the L_2E can be expressed as

$$\text{Var}(\sqrt{n}\hat{\pi}_{L_2E}) = J^{-1}KJ^{-1}, \tag{16}$$

where J and K are as in (13) and (15).

As in mixture models with two components, when $0 < \pi_j < 1$ for $j = 1, \ldots, k$, the L_2E for the mixing proportions can be proven to be consistent and asymptotically normally distributed.

Proposition 3.2 *Let* $f(x; \underline{\pi}) = \sum_{j=1}^{k} \pi_j f_j(x)$, $0 < \pi_j < 1$ *for* $j = 1, \ldots, k$ *and* $\sum_{j=1}^{k} \pi_j = 1$,

1. *the* L_2E *for the first* $k-1$ *mixing proportions,* $\hat{\underline{\pi}}_{L_2E,n}$ *is consistent for* $\underline{\pi} = (\pi_1, \ldots, \pi_{k-1})$, *and*

2. $n^{1/2}(\hat{\underline{\pi}}_{L_2E,n} - \underline{\pi})$ *is asymptotically multivariate normally distributed with mean vector of zero and covariance matrix* $J^{-1}KJ^{-1}$, *where* J *and* K *have the form as shown in (13) and (15).*

Proof: see Appendix.

Since the matrix K involves the unknown parameter vector $\underline{\pi}$, the variance of $n^{1/2}(\hat{\underline{\pi}}_{L_2E,n} - \underline{\pi})$ can be estimated by $J^{-1}\hat{K}J^{-1}$ where $\hat{K}$ is the matrix of the estimates for variances and covariances of $f_j(x) - f_k(x), j = 1, \cdots, k-1$ from the sample. On the diagonal of the matrix, there are estimates for variances of $f_j(x) - f_k(x)$ which have the form

$$\hat{\mathrm{Var}}[f_j(x) - f_k(x)] = (n-1)^{-1} \sum_{i=1}^{n} \left\{ f_j(X_i) - f_k(X_i) - n^{-1} \sum_{i=1}^{n} [f_j(X_i) - f_k(X_i)] \right\}^2 .$$

The $(j, j')^{th}$ element of $\hat{K}$ is the estimate for covariance of $f_j(x) - f_k(x)$ and $f_{j'}(x) - f_k(x)$ with the form

$$\frac{\sum_{i=1}^{n} \left\{ f_j(X_i) - f_k(X_i) - (\bar{f}_j - \bar{f}_k) \right\} \left\{ f_{j'}(X_i) - f_k(X_i) - (\bar{f}_{j'} - \bar{f}_k) \right\}}{n-1},$$

where $\bar{f}_j = n^{-1} \sum_{i=1}^{n} f_j(X_i)$. The covariance matrix of $n^{1/2}(\hat{\underline{\pi}}_{L_2E,n} - \underline{\pi})$ contains $k-1$ variances for the $k-1$ proportion estimators and $\frac{1}{2}(k-1)(k-2)$ potentially different covariances. The determinant of the covariance matrix can also be used to generalize the variability of the estimator. It is always useful to report both the variance-covariance matrix and the generalized variance.

Feng and McCulloch (1996) showed that the asymptotic variance for ($\sqrt{n}$ times) the MLE for the mixing proportions in mixture models can be approximated by the inverse of the Fisher information, which is J_0^{-1}, where J_0 is in the form shown in (4) with $\alpha = 0$. The variance for the L_2E has been developed in (16). For each of the mixing proportions, the asymptotic relative efficiency can be obtained through taking the ratio of the corresponding variance on the diagonal of the variance-covariance matrix of the MLE to that of the L_2E. In Table 2, the asymptotic relative efficiency of the L_2E for each of the mixing proportions in certain three-component Poisson mixture models are listed. In the last column, the ratio of the generalized asymptotic variance of the MLE to that of the L_2E is displayed, where the generalized asymptotic variance is defined as the determinant of the asymptotic covariance matrix of the estimator. It is observed that the L_2E performs better for the models with well separated components. The L_2E is less efficient when most of the data represent the first component, which has a smaller variance compared to other components.

Table 2: AREs of the L_2E in Certain 3-Component Poisson Mixture Models

λ	π	π_1	π_2	ratio of generalized variance
(1,2,3)	$(1/3, 1/3, 1/3)$	.83	.68	.59
	$(.10, .45, .45)$	.87	.75	.64
	$(.90, .05, .05)$	.49	.29	.16
(1,5,10)	$(1/3, 1/3, 1/3)$	.98	.92	.88
	$(.10, .45, .45)$	.96	.92	.86
	$(.90, .05, .05)$	.58	.42	.39

The influence function of the L_2E has the form as shown in (5). For the k-component mixture model, the matrix J involved in the formula is the same as shown in (13). Other quantities involved are calculated as

$$u(y; \underline{\pi}) f(y; \underline{\pi}) = [f_1(y) - f_k(y), \ldots, f_{k-1}(y) - f_k(y)]^T$$

and

$$\int_{\chi} u(x; \theta) f(x; \theta) g(x)\, dx = E_G \left[f_1(x) - f_k(x), \ldots, f_{k-1}(x) - f_k(x)\right]^T .$$

So the influence function of the L_2E for the mixing proportions has the form

$$I_{L_2E}(G, y) = J^{-1} \begin{pmatrix} f_1(y) - f_k(y) - E_G[f_1(x) - f_k(x)] \\ f_2(y) - f_k(y) - E_G[f_2(x) - f_k(x)] \\ \cdot \\ \cdot \\ \cdot \\ f_{k-1}(y) - f_k(y) - E_G[f_{k-1}(x) - f_k(x)] \end{pmatrix}, \tag{17}$$

where y is the contaminating value and G is the true distribution from which the random sample is generated. The influence function is bounded assuming the elements in J^{-1} are finite and $E_G[f_j(x) - f_k(x)]$ exists for $j = 1, 2, \ldots, k-1$.

The L_2E for the mixing proportions in k-component mixture models is not always in the parameter space Θ. If the L_2 estimate is outside the parameter space, some adjustment is necessary. A method of truncation is proposed by Shen (2004).

4 L_2E in Poisson Mixture Models with Unknown Parameters

In this section, the minimum L_2 distance estimation method is applied to k-component Poisson mixture models, assuming all the parameters $(\pi_1, \ldots, \pi_{k-1}, \lambda_1, \ldots, \lambda_k)$ are unknown. As in section 3, we will first consider $k = 2$ components and the results can be easily generalized to k-component models. The parametric distribution of a two-component Poisson mixture model is

$$f(x; \pi, \lambda_1, \lambda_2) = \pi e^{-\lambda_1}\lambda_1^x/x! + (1-\pi)e^{-\lambda_2}\lambda_2^x/x!, \quad x \geq 0, \quad (\pi, \lambda_1, \lambda_2) \in \Theta.$$

The parameter space is $\Theta = \{(\pi, \lambda_1, \lambda_2) \in \mathcal{R}^3 : 0 \leq \pi \leq 1, \lambda_1 > 0, \lambda_2 > 0\}$. The L_2E for the unknown parameters is the value $(\hat{\pi}, \hat{\lambda}_1, \hat{\lambda}_2)$ that minimizes

$$H_n(\pi, \lambda_1, \lambda_2) = \sum_{x=0}^{\infty} f^2(x; \pi, \lambda_1, \lambda_2) - \frac{2}{n}\sum_{i=1}^{n} f(X_i; \pi, \lambda_1, \lambda_2)$$

over the parameter space Θ.

Since the parameters λ_1 and λ_2 are not linear in the estimating equations, the L_2E for (λ_1, λ_2) does not have an explicit form. For computing the L_2E, numerical methods such as the Newton-Raphson algorithm are needed. The function 'nlm' (a Newton-Raphson algorithm) in R is used in this paper to perform the numerical minimization of the objective function. To increase the chance of finding the global minimum, different starting values are always suggested.

4.1 Asymptotic Properties for the L_2E

From (3) and (4), the matrices K_1 and J_1 of the L_2E in two-component Poisson mixture models are given by

$$K_1 = \sum_{x=0}^{\infty} \left\{ u(x; \pi, \lambda_1, \lambda_2) u^T(x; \pi, \lambda_1, \lambda_2) f^2(x; \pi, \lambda_1, \lambda_2) g(x) \right\} - \xi(\pi, \lambda_1, \lambda_2)\xi^T(\pi, \lambda_1, \lambda_2) \qquad (18)$$

where

$$\begin{aligned} u(x; \pi, \lambda_1, \lambda_2) &= \frac{\partial}{\partial(\pi, \lambda_1, \lambda_2)} \log f(x; \pi, \lambda_1, \lambda_2) \\ &= \frac{1}{f(x; \pi, \lambda_1, \lambda_2)} \left[f_1(x; \lambda_1) - f_2(x; \lambda_2), \pi \frac{\partial}{\partial \lambda_1} f_1(x; \lambda_1), (1-\pi) \frac{\partial}{\partial \lambda_2} f_2(x; \lambda_2) \right]^T \end{aligned}$$

and

$$\xi_1(\pi, \lambda_1, \lambda_2) = E_G\left[u(x;\pi,\lambda_1,\lambda_2)f(x;\pi,\lambda_1,\lambda_2)\right].$$

The matrix J_1 for the L_2E is defined as

$$\begin{aligned} J_1 &= \sum_{x=0}^{\infty} u(x;\pi,\lambda_1,\lambda_2)u^T(x;\pi,\lambda_1,\lambda_2)g(x)f(x;\pi,\lambda_1,\lambda_2) \\ &\quad -i(x;\pi,\lambda_1,\lambda_2)\left(g(x)f(x;\pi,\lambda_1,\lambda_2) - f^2(x;\pi,\lambda_1,\lambda_2)\right), \end{aligned} \tag{19}$$

where $i(x;\pi,\lambda_1,\lambda_2)$ is the information matrix of the two-component Poisson mixture distribution.

Theorem 4.1 *Let $(\pi, \lambda_1, \lambda_2)$ be the target parameter in two-component Poisson mixture models, where $0 < \pi < 1$, $\lambda_1 > 0$ and $\lambda_2 > 0$. There exists a sequence of minimum L_2 distance estimators $\{(\hat{\pi}_n, \hat{\lambda}_{1,n}, \hat{\lambda}_{2,n})\}$ such that*

1. *$(\hat{\pi}_n, \hat{\lambda}_{1,n}, \hat{\lambda}_{2,n})$ is consistent for $(\pi, \lambda_1, \lambda_2)$, and*

2. *$n^{1/2}\{(\hat{\pi}_n, \hat{\lambda}_{1,n}, \hat{\lambda}_{2,n})' - (\pi, \lambda_1, \lambda_2)'\}$ is asymptotically multivariate normally distributed with (vector) mean zero and covariance matrix $J_1^{-1}K_1J_1^{-1}$, where J_1 and K_1 have the forms as shown in* (14) *and* (15).

Proof: see Appendix.

Since both K_1 and J_1 involve the unknown density $g(x)$, the asymptotic variance of ($\sqrt{n}$ times) the L_2E can be consistently estimated by replacing K_1 and J_1 with their empirical versions from the random sample, $\hat{K}_1$ and $\hat{J}_1$, where

$$\begin{aligned} \hat{K}_1(\pi,\lambda_1,\lambda_2) &= \frac{1}{n-1}\sum_{i=1}^{n} u(X_i;\pi,\lambda_1,\lambda_2)u^T(X_i;\pi,\lambda_1,\lambda_2)f^2(X_i;\pi,\lambda_1,\lambda_2) \\ &\quad - \frac{1}{n^2(n-1)}\left\{\sum_{i=1}^{n} u(X_i;\pi,\lambda_1,\lambda_2)f(X_i;\pi,\lambda_1,\lambda_2)\right\}\left\{\sum_{i=1}^{n} u(X_i;\pi,\lambda_1,\lambda_2)f(X_i;\pi,\lambda_1,\lambda_2)\right\}^T, \end{aligned}$$

and

$$\begin{aligned} \hat{J}_1(\pi,\lambda_1,\lambda_2) &= \frac{1}{n}\sum_{i=1}^{n} u(X_i;\pi,\lambda_1,\lambda_2)u^T(X_i;\pi,\lambda_1,\lambda_2)f(X_i;\pi,\lambda_1,\lambda_2) \\ &\quad - \frac{1}{n}\sum_{i=1}^{n} i(X_i;\pi,\lambda_1,\lambda_2)f(X_i;\pi,\lambda_1,\lambda_2) + \sum_{x=0}^{\infty} i(x;\pi,\lambda_1,\lambda_2)f^2(x;\pi,\lambda_1,\lambda_2). \end{aligned}$$

Substituting estimates for π, λ_1 and λ_2 in these expressions yields the sandwich estimate of the variance-covariance matrix of the L_2E for $(\pi, \lambda_1, \lambda_2)$.

To measure the uncertainty of all the estimators, the generalized variance can be calculated.

4.2 Influence Function and Asymptotic Relative Efficiency

Using the results for the MDPDE from Basu et al. (1998), in two-component Poisson mixture models, the influence function of the L_2E for $(\pi, \lambda_1, \lambda_2)$ has the form

$$I_{L_2E}(G,y) = J_1^{-1}(\pi,\lambda_1,\lambda_2)\begin{pmatrix} f_1(y;\lambda_1) - f_2(y;\lambda_2) - E_G\left[f_1(x;\lambda_1) - f_2(x;\lambda_2)\right] \\ \pi\frac{\partial}{\partial\lambda_1}f_1(y;\lambda_1) - E_G\left[\pi\frac{\partial}{\partial\lambda_1}f_1(x;\lambda_1)\right] \\ (1-\pi)\frac{\partial}{\partial\lambda_2}f_2(y;\lambda_2) - E_G\left[(1-\pi)\frac{\partial}{\partial\lambda_2}f_2(x;\lambda_2)\right] \end{pmatrix}, \tag{20}$$

where y is the contaminating value, J_1 is as shown in (20) and G is the true distribution from which the random sample is generated.

The following argument shows the influence function for π is bounded. The first derivative of a Poisson density with respect to its parameter λ can be expressed in the form of $-f(x;\lambda)+f(x-1;\lambda)$. Since this is the difference of densities of two Poisson distributions, which are always between 0 and 1, the influence function for (λ_1, λ_2) is also bounded. When the contaminating value y goes to infinity, then $f_1(y;\lambda_1) - f_2(y;\lambda_2)$ vanishes for Poisson distributions, as do $\partial f_1(x;\lambda_1)/\partial\lambda_1$ and $\partial f_2(x;\lambda_2)/\partial\lambda_2$. The limit of the influence function as $y \to \infty$ can be expressed

$$\lim_{y\to\infty} I_{L_2E}(y) = -J^{-1}(\pi,\lambda_1,\lambda_2)\begin{pmatrix} \sum_{x=0}^{\infty}\{f_1(x;\lambda_1) - f_2(x;\lambda_2)\}g(x) \\ \sum_{x=0}^{\infty}\pi\{\frac{\partial}{\partial\lambda_1}f_1(x;\lambda_1)\}g(x) \\ \sum_{x=0}^{\infty}(1-\pi)\{\frac{\partial}{\partial\lambda_2}f_2(x;\lambda_2)\}g(x) \end{pmatrix}. \tag{21}$$

The influence function for the MLE can be derived as $I_{\text{MLE}}(G,y) = J_0^{-1}\left\{u(y;\pi,\lambda_1,\lambda_2) - \xi_0(\pi,\lambda_1,\lambda_2)\right\}$ where

$$\xi_0(\pi,\lambda_1,\lambda_2) = \sum_{x=0}^{\infty} u(x;\pi,\lambda_1,\lambda_2)g(x)$$

and

$$\begin{aligned} J_0 &= \sum_{x=0}^{\infty} u(x;\pi,\lambda_1,\lambda_2)u^T(x;\pi,\lambda_1,\lambda_2)f(x;\pi,\lambda_1,\lambda_2) \\ &\quad - \sum_{x=0}^{\infty} i(x;\pi,\lambda_1,\lambda_2)(g(x) - f(x;\pi,\lambda_1,\lambda_2)). \end{aligned}$$

Under model conditions, the unknown density $g(x)$ can be replaced by the parametric density $f(x;\pi,\lambda_1,\lambda_2)$. Figure 2 plots the influence functions of the L_2E and the MLE for the mixing proportion π in certain two-component Poisson mixture models, assuming all parameters are unknown. We observe that the influence functions of the L_2E stays closer to zero than those of the MLE.

And the change of the MLE can be dramatic when the contaminating value is very far away from the bulk of the data. Both the L_2E and the MLE are more robust for models with well separated components and low mixing proportion on the first component π.

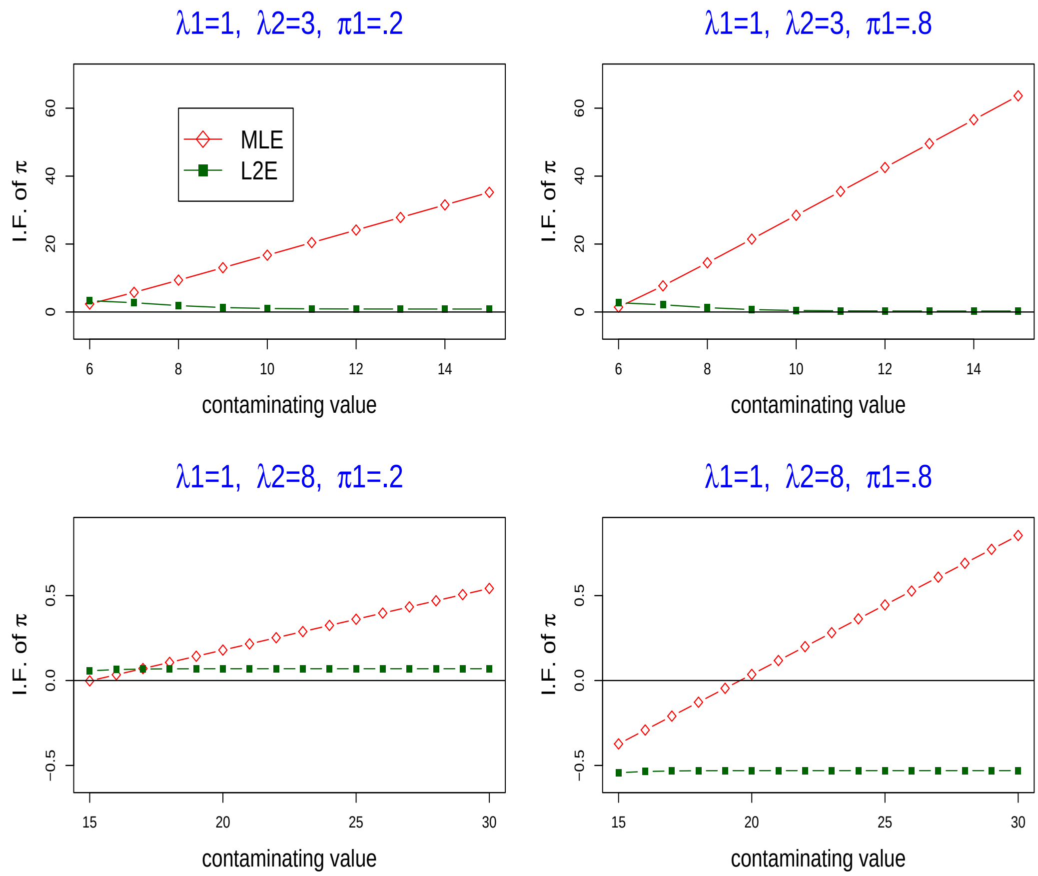

Figure 2: Plot of Influence Functions of the L_2E and MLE in 2-Component Poisson Mixture Models with $(\lambda_1 = 1,\ \lambda_2 = 3)$ and $(\lambda_1 = 1,\ \lambda_2 = 8)$.

The asymptotic relative efficiency of the L_2E is studied here to evaluate its asymptotic efficiency compared to the MLE. Under model conditions, the ARE of the L_2E can be calculated from the asymptotic variance-covariance matrix of the MLE

$$J_0^{-1} = \left\{ \sum_{x=0}^{\infty} u(x;\pi,\lambda_1,\lambda_2)u^T(x;\pi,\lambda_1,\lambda_2) f(x;\pi,\lambda_1,\lambda_2) \right\}^{-1}$$

and the asymptotic variance-covariance matrix of the L_2E, $J_1^{-1}K_1J_1^{-1}$, where K_1 and J_1 are as shown in (18) and (19). Table 3 displays the asymptotic relative efficiencies of the L_2E in some two-component Poisson mixture models. For each model, the ratio of the generalized variance of the MLE to that of the L_2E is displayed in the last column.

Table 3: AREs of the L_2E in Certain 2-Component Poisson Mixture Models with All Parameters Unknown

	π	$\hat{\pi}$	$\hat{\lambda}_1$	$\hat{\lambda}_2$	ratio of generalized variance
$\lambda_1 = 1, \lambda_2 = 2$	.2	.43	.49	.37	.33
	.5	.35	.42	.29	.25
	.8	.23	.31	.17	.13
$\lambda_1 = 1, \lambda_2 = 3$	.2	.68	.75	.55	.52
	.5	.56	.65	.43	.41
	.8	.35	.48	.24	.22
$\lambda_1 = 1, \lambda_2 = 5$	.2	.94	.93	.70	.67
	.5	.87	.87	.61	.59
	.8	.61	.70	.38	.37
$\lambda_1 = 1, \lambda_2 = 8$	.2	.95	.95	.73	.66
	.5	.92	.84	.73	.59
	.8	.77	.75	.64	.45

It is observed that the L_2E is more efficient for the model with well separated components and low mixing proportions on the first component. This pattern is consistent with the case with only the mixing proportion unknown. Also it is noticed that when the mixing proportion for the first component is high and the two components are poorly separated, the asymptotic relative efficiency of the L_2E can be even lower than 20%. The efficiency of the L_2E for λ_2 is the lowest compared to the other two parameters, possibly because the second component is more variable and the L_2 method tends to ignore some large values from the second component.

5 Simulation Study

The performances of the L_2E, the MLE and the MHDE in Poisson mixtures with finite sample size are compared in this section through some simulation studies. The relative mean squared error is utilized to measure the efficiencies of the estimators, assuming the model is correctly specified. To examine the robustness of the estimators in small samples, their sensitivity curves and relative mean squared errors in some contaminated models are compared. For brevity, we examine two-component cases only, but similar results can be obtained with k-component cases (Shen, 2004). In Section 5.1

the performance of the estimators will be evaluated in two-component Poisson mixtures with only the mixing proportion unknown. In Section 5.2 the two-component Poisson mixture model with all three parameters $(\pi, \lambda_1, \lambda_2)$ unknown is investigated. All computations were performed using R.

5.1 Two-component Poisson Mixture Model with π Unknown

In this section, it is assumed the data are from a two-component Poisson mixture model with density as (1) while $k = 2$ and only the mixing proportion π is unknown. To obtain the L_2E, the closed form estimator in (6) is used and the estimator is truncated when it is out of the parameter space. To calculate the MLE for π, the method of bisection is utilized. The HELMIX algorithm proposed by Karlis and Xekalaki (1998) is applied to get the MHDE for π. For both the ML method and the MHD method, two different initial values are used for π, $\pi = .5$ and the proportion in the model from which the data are generated.

5.1.1 Correctly Specified Models

In the case of a correctly specified model, the simulated samples are from a two-component Poisson mixture model with parameter $(\pi, \lambda_1, \lambda_2)$, where λ_1 and λ_2 are assumed known. To compare the efficiencies of estimators, for large samples, we compare their asymptotic relative efficiencies with respect to the MLE. For small samples, we compare their relative mean squared errors (RMSE) with respect to the MLE, that is,

$$\text{RMSE}(L_2E) = \frac{\text{MSE(MLE)}}{\text{MSE}(L_2E)}, \quad \text{RMSE(MHDE)} = \frac{\text{MSE(MLE)}}{\text{MSE(MHDE)}}. \tag{22}$$

The mean squared error (MSE) of an estimator $\hat{\theta}$ for a parameter θ is defined as

$$\text{MSE}(\hat{\theta}) = E[(\hat{\theta} - \theta)^2].$$

Via some simulation studies, the mean squared error can be estimated by its empirical mean squared error defined as

$$\hat{\text{MSE}}(\hat{\theta}) = \frac{1}{n_s} \sum_{i=1}^{n_s} (\hat{\theta}_i - \theta)^2, \tag{23}$$

where n_s is the number of replicates performed in the simulation study and $\hat{\theta}_i$ is the estimate obtained in i^{th} replicate.

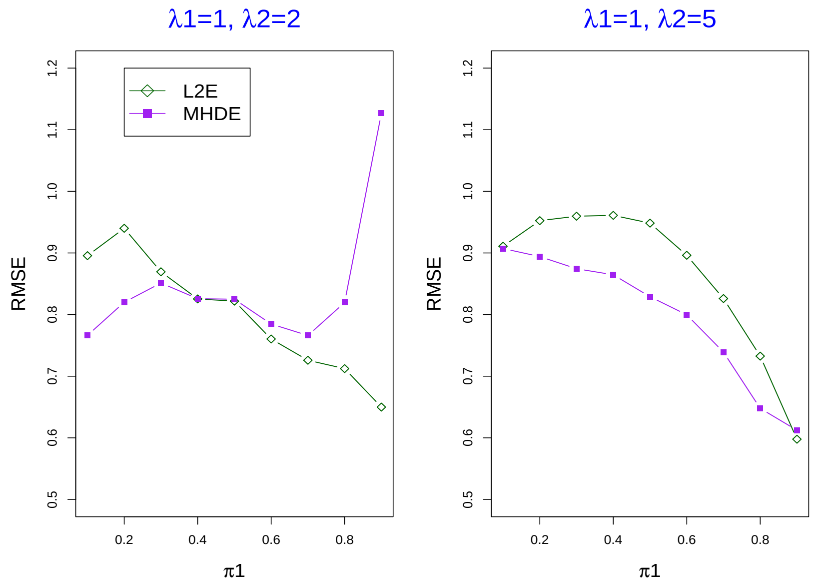

Figure 3: RMSE of L_2E and MHDE for π in Certain two-component Poisson Mixtures

Figure 3 plots the simulation results for certain two-component Poisson mixtures. The mixing proportion π is taken on different values from .1 to .9 to evaluate its effect on the efficiencies of the estimators. For each parameter vector 3000 samples with sample size $n = 100$ were generated independently from the corresponding mixture distribution and the empirical relative mean squared errors of the L_2E and the MHDE are calculated by taking the ratio of the empirical MSEs of the L_2E and the MHDE to that of the MLE. The estimator for the variance of the ratio of two empirical MSEs proposed by Juarez (2003) is calculated here to measure the accuracy of the estimates of RMSE. The magnitude of the estimates of these standard errors is much smaller than that of the empirical RMSEs.

With sample size 100, the MLE is more efficient than the L_2E and the MHDE if the model is correctly specified. The L_2E performs better for the models with well separated components and low mixing proportions of the first component. This observation is consistent with what we have seen for the asymptotic efficiencies of the L_2E. When the two components are well separated, the L_2E is more efficient than the MHDE. In models with poorly separated components, the performance of the L_2E is about the same as that of the MHDE except for the mixing proportion of the first component being very high. When $\pi = .9$, the MHDE is superior to the L_2E with respect to the RMSE. When the two components are poorly separated and most of the data are from the first component, the MHD method is more likely to ignore some relatively large values from the second component and treat all the observation as from the first component. Hence, most of the estimates are close to 1 and the MHDE has a smaller variance compared to the MLE and the L_2E. This can be confirmed with the positive bias of the MHDE. When the true value of π is .90, the bias of the MHDE does not increase the RMSE much compared to the decrease of the RMSE by its small variance.

5.1.2 Contaminated Models

This section evaluates the performance of the estimators when there are gross errors existing in the two-component Poisson mixture model. The data are generated from the contaminated distribution

$$(1-\epsilon)F(x;\pi,\lambda_1,\lambda_2)+\epsilon F(x;\lambda_3), \quad 0\leq\epsilon\leq .5. \tag{24}$$

In this model, the data are generated from two sources. The bulk of the data come from the specified two-component Poisson mixture distribution with the probability $1-\epsilon$. The rest of the data come from the contaminating Poisson distribution $F(x;\lambda_3)$ with probability ϵ. The ϵ is also known as the *contaminating rate*.

The relative mean squared error is also used here to measure the performance of the estimators. Table 4 lists the result of some simulation studies with data simulated from certain contaminated distributions. It is assumed that the data are from the contaminated distribution as shown in (24) with $\lambda_1 = 1$ and $\lambda_2 = 3$. The parameter of the contaminating component λ_3 is set at different values from 7 to 20 and two contaminating rates $\epsilon = .05$ and $\epsilon = .10$ are chosen. The mixing proportion for the first component π is taken from .2 to .8. The empirical mean squared errors are calculated based on 1000 replicates with sample size 100. The numbers in the parentheses are the estimates of the standard errors of the empirical RMSEs.

Table 4 shows that when the underlying distribution is contaminated, the L_2E and the MHDE are more robust than the MLE, particularly for a high contaminating rate and large proportion of the first component. In the simulation studies, the contaminating components are at the right tail of the underlying distribution, and the contaminants are more likely to affect the second component compared to the first component. When π is high, only a small portion of the data are from the second component and the estimators are more likely to be affected by the contaminants in this

Table 4: Empirical RMSEs of the L_2E and MHDE for π in Certain Contaminated 2-component Poisson Mixture Models with $\lambda_1 = 1$ and $\lambda_2 = 3$

		L_2E	MHDE	L_2E	MHDE
π	ϵ	$\lambda_3 = 7$		$\lambda_3 = 12$	
0.20	0.05	0.98 (0.03)	1.02 (0.01)	1.06 (0.01)	1.02 (0.03)
	0.10	1.38 (0.01)	1.21 (0.04)	1.52 (0.06)	1.35 (0.05)
0.50	0.05	1.13 (0.03)	1.17 (0.02)	1.39 (0.04)	1.34 (0.06)
	0.10	1.89 (0.05)	1.58 (0.03)	2.56 (0.08)	2.60 (0.11)
0.80	0.05	1.49 (0.04)	1.80 (0.04)	1.90 (0.06)	2.33 (0.11)
	0.10	2.39 (0.06)	2.10 (0.04)	3.50 (0.11)	5.53 (0.23)
		$\lambda_3 = 15$		$\lambda_3 = 20$	
0.20	0.05	1.05 (0.03)	0.97 (0.04)	1.05 (0.03)	0.91 (0.03)
	0.10	1.51 (0.06)	1.29 (0.06)	1.53 (0.07)	1.19 (0.06)
0.50	0.05	1.35 (0.04)	1.28 (0.06)	1.42 (0.05)	1.25 (0.07)
	0.10	2.58 (0.08)	2.49 (0.12)	2.74 (0.08)	2.64 (0.13)
0.80	0.05	1.94 (0.06)	2.26 (0.12)	1.90 (0.06)	2.29 (0.13)
	0.10	3.46 (0.10)	6.14 (0.28)	3.61 (0.10)	5.38 (0.27)

case.

The influence functions of the L_2E and the MLE for the mixing proportion in certain two-component Poisson mixture models have been shown in Chapter 3. One of the finite sample versions of the influence function is Tukey's (1970/1971) *Sensitivity Curve*. With a random sample $X_1, X_2, \ldots, X_n$, the sensitivity curve of an estimator T at the value x is defined as

$$SC_n(x) = n[T(X_1, \ldots, X_{n-1}, x) - T(X_1, \ldots, X_n)],$$

where T represents the dependence of the estimator on the underlying distribution. The average of sensitivity curves over the simulations is calculated to evaluate the local robustness of the estimators. Figure 4 plots the sensitivity curves of the MLE, the L_2E and the MHDE with different contaminating values, assuming the underlying distribution is a two-component Poisson mixture distribution with $\pi = .3, \lambda_1 = 1$ and $\lambda_2 = 3$.

It is observed that when the contaminating value is very extreme for the underlying parametric distribution, the L_2E and the MHDE are more robust than the MLE, and the MHDE performs the

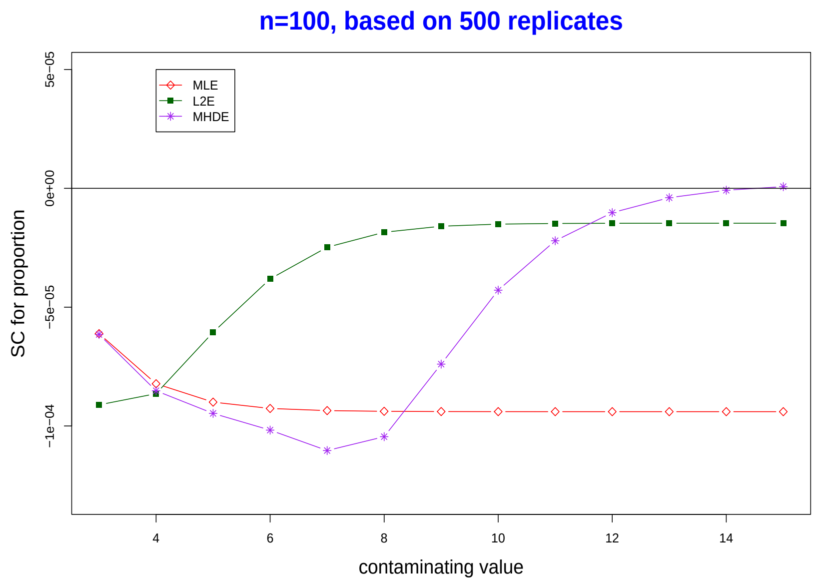

Figure 4: Sensitivity Curves of the L_2E, the MLE and the MHDE in 2-component Poisson Mixture Model with $\pi = .3$, $\lambda_1 = 1$ and $\lambda_2 = 3$.

best. When the contaminating value is moderately far away from the bulk of the data, the L_2E is less sensitive to the contaminants than the MLE and the MHDE.

Hjort (1994) shows that the log-likelihood is close to a weighted L_2 distance with weight given by the inverse of the empirical density of an observation. This procedure gives more weight to the contaminant, which has a lower frequency compared to the observations from the underlying distribution. The L_2 method gives equal weight to each observation. This explains why the MLE is more sensitive to the contaminants than the L_2E. The Hellinger distance can also be treated as a weighted L_2 distance with weight in the form

$$w = \frac{\sqrt{f_\theta(x)g(x)}}{(g(x) - f_\theta(x))^2}.$$

If x is an outlier to the parametric density $f_\theta(\cdot)$, the weight for x is close to

$$w \simeq \frac{\sqrt{f_\theta(x)g(x)}}{g^2(x)} = \frac{\sqrt{f_\theta(x)}}{g(x)\sqrt{g(x)}}.$$

Since the weight is proportional to the parametric density of the underlying distribution, if the observation is very extreme in the parametric distribution, it receives no weight in the procedure and the MHDE is the most robust estimator. In our sample shown in Figure 4, the MHDE is more robust than the L_2E only if the contaminant is more than 5 standard deviations away from the mean of the underlying distribution.

5.2 Two-component Poisson Mixture Models with $(\pi, \lambda_1, \lambda_2)$ Unknown

When all the parameters $(\pi, \lambda_1, \lambda_2)$ are unknown, the L_2E is calculated using the R function "nlm". The conventional EM algorithm is applied to calculate the MLE and the HELMIX algorithm given by Karlis and Xekalaki (1998) is the method for evaluating the MHDE.

5.2.1 Correctly Specified Models

This section compares the efficiencies of the estimators in correctly specified two-component Poisson mixture models for small samples. Table 5 summarizes the ratios of the estimated generalized variances of the MLE to those of the L_2E and the MHDE for several two-component Poisson mixture models from a simulation study with 1,000 replicates and sample size 100.

Table 5: The Ratios of the Estimated Generalized Variances of the MLE to Those of the L_2E and the MHDE for $(\pi, \lambda_1, \lambda_2)$ in Certain 2-component Poisson Mixture Models

	$\pi = 0.2$	$\pi = 0.5$	$\pi = 0.8$
Estimator		$\lambda_1 = 1, \quad \lambda_2 = 2$	
L_2E	0.84	0.82	0.91
MHDE	10.28	5.29	2.98
		$\lambda_1 = 1, \quad \lambda_2 = 5$	
L_2E	0.68	0.61	0.15
MHDE	1.43	0.93	0.61

It is observed that when the two components are well separated, the MLE performs the best. When the two components are poorly separated, the MHDE is more efficient than the L_2E and the MLE. For all the models studied, the MHDE is more efficient than the L_2E. The efficiencies of the L_2E and the MHDE depend on the mixing proportion π of the first component. Both estimators perform better when the majority of the data are representing the second component, which has a larger variance compared to the first component. A possible explanation for the superior performance of the MHDE can be that the MHD method tends to ignore large values and the estimator has much lower variability especially for λ_2. This is supported by the negative biases of the MHDE for π_1 and π_2.

5.2.2 Contaminated Models

In this section, the performance of the estimators is studied for certain contaminated two-component Poisson mixture models assuming all three parameters unknown. In addition, it is assumed that the additional contaminating component is located at the right tail of the specified two-component Poisson mixture model with the contaminating rate ϵ.

Figure 5 displays the plot of sensitivity curves of the estimators for the mixing proportion π with different contaminating values, assuming all three parameters are unknown. This figure is based on a simulation study with 500 replicates and sample size 100.

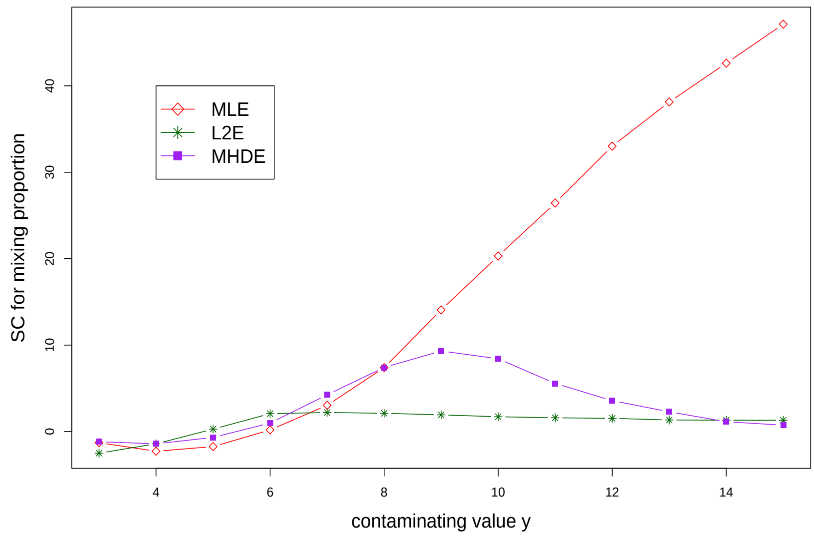

Figure 5: Sensitivity Curves of the Estimators for π in 2-component Poisson Mixture Model with $\pi = .3$, $\lambda_1 = 1$ and $\lambda_2 = 3$ (assuming all the parameters are unknown)

It is observed that when the contaminating value y is far away from the underlying distribution, the MLE can be changed dramatically. Both the L_2E and the MHDE are more robust to the contaminant than the MLE. If the contaminating value is very extreme to the underlying distribution, the sensitivity curves for both the L_2E and the MHDE are close to zero. When the contaminating value is moderately far away from the bulk of the data, the MHDE can be affected considerably while the L_2E does not change much. Based on the simulation study, the MHDE is less robust than the L_2E if the contaminant is within 6 standard deviation of the mean for the underlying distribution.

6 Conclusions

This paper introduces a minimum L_2 distance estimator (L_2E) for the mixing proportion in k-component mixture models with both known and unknown component distributions. One of the main advantages of this method over other estimation methods used currently in mixture estimation is that it offers a closed form estimator for the mixing proportion. The proposed estimator is more robust to the existence of gross errors than the MLE while still retaining acceptable efficiency. When the model is correctly specified, the ARE for the L_2E of the mixing proportion is over .60. If the weight for the component with larger variance is not too low, usually the ARE for the L_2E is higher than 80%. Compared to another minimum distance estimator, the minimum Hellinger distance estimator (MHDE), the L_2E is more robust when the contaminant is moderately far away from the underlying distribution. Arguably these more moderate outliers are of greater importance than extreme outliers, as they are harder to detect. The price to be paid for this improved robustness to moderate outliers is a reduced efficiency of the L_2E compared to the MHDE.

We have empirically demonstrated the improved robustness of the L_2E to moderate outliers in Poisson mixtures. However, the arguments in Section 5.1 imply the improved robustness should hold in general.

Developing the L_2E for Poisson mixture models when both the parameters for the Poisson distributions and also the number of components k are unknown are under investigation by the authors. For the Poisson mixture models, the fitted model can be obtained by starting with an initial model with all the possible component parameters and including only those components with positive weights. The model obtained by using this algorithm tends to have too many components than it is needed. Some model choice criteria can be applied to balance the complexity of the model and the goodness-of-fit with the data. The robustness of the L_2 method to outliers in such cases can be examined by comparing the number of components in the fitted model from the L_2 method with those from the ML and the MHD methods.

Appendix

ASYMPTOTIC BEHAVIOR OF THE L_2E FOR MIXING PROPORTION IN TWO-COMPONENT MIXTURE DISTRIBUTIONS

Theorem 2.2 in Basu et al. (1997) gives a set of regularity conditions that ensure consistence and asymptotic normality of the MDPDE. In this section, it is shown that this theorem can be applied to the L_2E for π in the mixture density by showing that the regularity conditions are satisfied in mixture density with 2 known components.

1. It is easily to see that the support of the mixture density doesn't depend on the interested parameter π. So the first condition is satisfied.

2. The density $f_\theta(x) = \pi f_1(x) + (1-\pi) f_2(x)$. The first derivative of the density with respect to π is $f_1(x) - f_2(x)$, which does not involve the parameter π. So the 2nd and 3rd derivatives are zero and hence the 3rd derivative is continuous with respect to the parameter π.

3. Assuming the mixture distribution is absolutely continuous, when the tuning parameter $\alpha = 1$,

$$\begin{aligned}
\int f_\theta^{1+\alpha}(x)\,dx &= \int [\pi f_1(x) + (1-\pi) f_2(x)]^2\,dx \\
&= \int \left\{\pi^2 f_1^2(x) + 2\pi(1-\pi) f_1(x) f_2(x) + (1-\pi)^2 f_2^2(x)\right\}\,dx \\
&= \pi^2 \int f_1^2(x)\,dx + 2\pi(1-\pi)\int f_1(x) f_2(x)\,dx + (1-\pi)^2 \int f_2^2(x)\,dx.
\end{aligned}$$

   Since both $f_1(\cdot)$ and $f_2(\cdot)$ do not involve π, the derivative of $\int f_\theta^2(x)\,dx$ can be taken under the integral sign, i.e.

$$\frac{\partial}{\partial \pi}\int f_\theta^2(x)\,dx = \int \left\{2\pi f_1^2(x) + 2(1-2\pi) f_1(x) f_2(x) - 2(1-\pi) f_2^2(x)\right\}\,dx$$

   and

$$\frac{\partial^2}{\partial \pi^2}\int \left\{f_\theta^2(x)\,dx = \int 2 f_1^2(x) - 4 f_1(x) f_2(x) + 2 f_2^2(x)\right\}\,dx.$$

   It is noticed that the 2^{nd} derivative does not involve π so the 3^{rd} derivative is zero, i.e. the integral $\int f_\theta^2(x)\,dx$ is three times differentiable and the differentiation can be taken under the integral sign. Similar result can be obtained for discrete mixture distribution.

4. It has been shown in Section 3.2 that the quantity $J(\theta) = \int [f_1(x) - f_2(x)]^2\,dx$ is positive definite.

5. When the tuning parameter $\alpha = 1$, then $V_{n,\theta}(x) = \int f_\theta^2(x)\,dx - 2f_\theta(x)$. We just showed above that the 3rd derivative of $\int f_\theta^2(x)\,dx$ is zero and the 3rd derivative of $f_\theta(x)$ is also zero. So the 3rd derivative of $V_{n,\theta}(x)$ is zero, and $M_{jkl}(x)$ can be set to any positive constant function and the expected value of $M_{jkl}(x)$ exists. Thus condition (5) is also satisfied.

ASYMPTOTIC BEHAVIOR OF THE TRUNCATED L_2E FOR MIXING PROPORTION IN TWO-COMPONENT MIXTURE DISTRIBUTIONS

From (11) and (12), when the mixing proportion π is not on the boundary of the parameter space, the indicator function $I_{(1,\infty)}(\hat{\pi})$ in (13) converges in probability to 0, as does the function $I_{[0,1]}(\hat{\pi})$. It is noticed that the L_2E , $\hat{\pi}_{L_2E}$, for the mixing proportion is a function of some sample average and is unbiased for π. Thus $\hat{\pi}_{L_2E}$ converges in probability to π by the weak law of large numbers. Theorem 2.1.3 in Lehmann (1998) indicates that if two sequences of random variables A_n and B_n converge in probability to constants a and b respectively, then the sum, difference and multiplication of A_n and B_n also converge in probability to the sum, difference and multiplication of a and b. With a simple application of this theorem, it can be proved that $\hat{\pi}_{\text{trunc}}$ converges in probability to π, i.e. it is also consistent for π. Also it is shown in Theorem 3.1 that the L_2E for π converges in distribution to a random variable with normal distribution. Lemma 2.3.2 in Lehmann (1998) states that if a sequence of random variables Y_n converge in law to some random variable Y, and a and b are constants with $b \neq 0$, then $bY_n + a$ converges in law to $bY + a$. By this lemma, the truncated L_2E for π converges in law to the same normal distribution.

ASYMPTOTIC BEHAVIOR OF L_2E FOR $(\pi_1, \ldots, \pi_{k-1})$

In this section, it is shown the regularity conditions in Theorem 2.2 in Basu et al. (1997) are satisfied in mixture models with k-component mixture models with only the mixing proportions unknown.

1. The support for the mixture density is the same as the union of supports for the density components and it does not involve the mixing proportions. So the first condition is satisfied.

2. The density $f(x;\theta) = \pi_1 f_1(x) + \pi_2 f_2(x) + \cdots + (1 - \pi_1 - \pi_2 - \cdots - \pi_{k-1}) f_k(x)$. The first derivative of the density with respect to π_i is $f_i(x) - f_k(x)$, which does not involve any parameter π_j. So the second and the third derivatives are equivalent to zero and hence the third derivative is continuous with respect to the parameter vector $\underline{\pi}$.

3. The integral $\int f^{1+\alpha}(x;\theta)\,dx = \int f^2(x;\theta)\,dx$ when the tuning parameter $\alpha = 1$.

$$\int f^2(x;\theta)\,dx = \int \left\{ \sum_{j=1}^{k} \pi_j f_j(x) \right\}^2 dx$$

Since all the component densities $f_1(\cdot), f_2(\cdot), \ldots, f_k(\cdot)$ do not involve the mixing proportions $\pi_1, \ldots, \pi_{k-1}$, the derivative of $\int f^2(x;\theta)\,dx$ with respect to $\underline{\pi}$ can be taken under the integral sign.

$$\frac{\partial}{\partial \pi_i} \int f^2(x;\theta)\,dx = \int 2 f(x;\underline{\pi})(f_i(x) - f_k(x))\,dx$$

$$\frac{\partial^2}{\partial \pi_i \pi_j} \int f^2(x;\theta)\,dx = \int 2(f_j(x) - f_k(x))(f_i(x) - f_k(x))\,dx$$

$$\frac{\partial^3}{\partial \pi_i \pi_j \pi_m} \int f^2(x;\theta)\,dx = 0$$

So the integral $\int f^2(x;\underline{\pi})\,dx$ is three times differentiable and the differentiation can be taken under the integral sign.

4. In section 3 it was shown that the matrix $J_1(\underline{\pi})$ is a $k-1$ by $k-1$ symmetric matrix with elements

$$J_{i,j} = \int (f_i(x) - f_k(x))(f_j(x) - f_k(x))\,dx \quad i,j = 1, \ldots, k-1.$$

Let $y = (y_1, \ldots, y_{k-1})'$ be any vector with real numbers, then $y'Jy$ can be expressed as

$$\begin{aligned} y'Jy &= \sum_{i=1}^{k-1} \sum_{j=1}^{k-1} y_i y_j \int (f_i(x) - f_k(x))(f_j(x) - f_k(x))\,dx \\ &= \int \sum_{i=1}^{k-1} \sum_{j=1}^{k-1} y_i y_j (f_i(x) - f_k(x))(f_j(x) - f_k(x))\,dx \\ &= \int \left(\sum_{i=1}^{k-1} y_i (f_i(x) - f_k(x)) \right)^2 dx. \end{aligned}$$

This quantity is positive assuming the density components are identifiable. So the matrix J is positive definite and $A4$ is satisfied.

5. When the tuning parameter $\alpha = 1$,

$$V_{n,\theta}(x) = \int f^2(x; \underline{\pi})\, dx - 2f(x; \underline{\pi}).$$

The third derivative of $\int f^2(x; \underline{\pi})\, dx$ is zero as shown in 3, and the third derivative of $f(x; \underline{\pi})$ is also zero as shown in 2. So the third derivative of $V_{n,\theta}(x)$ is zero. Thus $M_{jkl}(x)$ can be set to any positive constant function and the expected value of $M_{jkl}(x)$ exists.

The consistency and asymptotic normality are proven for the L_2E of $\underline{\pi}$.

ASYMPTOTIC BEHAVIOR OF THE L_2E FOR $(\pi, \lambda_1, \lambda_2)$ IN TWO-COMPONENT POISSON MIXTURE MODEL

In this section, it is proved that Theorem 2.2 in Basu et al. (1997) can be applied to the L_2E for $(\pi, \lambda_1, \lambda_2)$ in two-component Poisson mixtures. It is shown that the regularity conditions are satisfied in such models.

1. The support for the Poisson mixture model is $\chi = \{x : x = 0, 1, 2, \ldots\}$, which does not depend on a parameter. Thus the first condition is satisfied.

2. The density $f_\theta(x) = \pi_1 f_1(x) + (1-\pi) f_2(x)$, where $f_1(\cdot)$ and $f_2(\cdot)$ are Poisson densities with parameters λ_1 and λ_2. The third derivatives of the density are

$$\frac{\partial^3}{\partial \lambda_1^3} f_\theta(x) = \pi \frac{\partial^3}{\partial \lambda_1^3} f_1(x; \lambda_1)$$

$$\frac{\partial^3}{\partial \lambda_2^3} f_\theta(x) = (1-\pi) \frac{\partial^3}{\partial \lambda_2^3} f_2(x; \lambda_2)$$

$$\frac{\partial^3}{\partial \pi \partial \lambda_1^2} f_\theta(x) = \frac{\partial^2}{\partial \lambda_1^2} f_1(x; \lambda_1)$$

$$\frac{\partial^3}{\partial \pi \partial \lambda_2^2} f_\theta(x) = -\frac{\partial^2}{\partial \lambda_2^2} f_2(x; \lambda_2)$$

and all the other third derivatives are zero. It is known the Poisson density is three times differentiable with respect to the parameter λ and the third derivative is continuous with respect to λ. Thus it is obvious that the mixture Poisson density $f_\theta(x)$ is three times differentiable and the third partial derivatives are continuous with respect to the parameter $(\pi, \lambda_1, \lambda_2)$. Hence, the second condition is satisfied.

3. Now we prove that $\sum\limits_{x=0}^{\infty} f_\theta^2(x)$ is three times differentiable with respect to $(\pi, \lambda_1, \lambda_2)$.

$$\frac{\partial}{\partial\theta_1}\sum_{x=0}^{\infty} f_\theta^2(x) = \sum_{x=0}^{\infty} 2 f_\theta(x)\frac{\partial}{\partial\theta_1} f_\theta(x)$$

$$\frac{\partial^2}{\partial\theta_1\partial\theta_2}\sum_{x=0}^{\infty} f_\theta^2(x) = 2\sum_{x=0}^{\infty}\left\{\frac{\partial}{\partial\theta_1}f_\theta(x)\frac{\partial}{\partial\theta_2}f_\theta(x) + f_\theta(x)\frac{\partial^2}{\partial\theta_1\partial\theta_2}f_\theta(x)\right\}$$

$$\begin{aligned}\frac{\partial^3}{\partial\theta_1\partial\theta_2\partial\theta_3}\sum_{x=0}^{\infty} f_\theta^2(x) &= 2\sum_{x=0}^{\infty}\frac{\partial^2}{\partial\theta_1\partial\theta_3}f_\theta(x)\frac{\partial}{\partial\theta_2}f_\theta(x) + \frac{\partial}{\partial\theta_1}f_\theta(x)\frac{\partial^2}{\partial\theta_2\partial\theta_3}f_\theta(x)\\ &\quad + \frac{\partial}{\partial\theta_3}f_\theta(x)\frac{\partial^2}{\partial\theta_1\partial\theta_2}f_\theta(x) + f_\theta(x)\frac{\partial^3}{\partial\theta_1\partial\theta_2\partial\theta_3}f_\theta(x)\end{aligned}$$

Since the mixture density $f_\theta(x)$ is three times differentiable with respect to $\theta = (\pi, \lambda_1, \lambda_2)$, the quantity $\sum\limits_{x=0}^{\infty} f_\theta^2(x)$ is also three times differentiable and the derivative can be taken under the summation sign.

4. The symmetric matrix $\mathbf{J}(\pi, \lambda_1, \lambda_2)$ has the form

$$\sum_{x=0}^{\infty}\begin{pmatrix} (f_1-f_2)^2 & \pi f_1'(f_1-f_2) - f_1'(g-f_\theta) & (1-\pi)f_2'(f_1-f_2) + f_2'(g-f_\theta) \\ \cdots & \pi^2(f_1')^2 - \pi f_1''(g-f_\theta) & \pi(1-\pi)f_1'f_2' \\ \cdots & \cdots & (1-\pi)^2(f_2')^2 - (1-\pi)f_2''(g-f_\theta) \end{pmatrix}$$

where f_1 and f_2 are the Poisson component densities with parameters λ_1 and λ_2, and f_i' and f_i'' are the first and second derivatives of the Poisson density with respect to the parameter λ. Now let $\theta = (\pi, \lambda_1, \lambda_2)$ represents the best fitting value of the parameter such that

$$\begin{aligned}&\sum_{x=0}^{\infty} u_\theta(x) f_\theta(x)(g_n(x) - f_\theta(x)) = \\ &\qquad \sum_{x=0}^{\infty}\left(f_1(x) - f_2(x), \pi f_1'(x), (1-\pi)f_2'(x)\right)'(g(x) - f_\theta(x)) = \underline{0}.\end{aligned}$$

And let ω be an open subset of the parameter space Ω containing the best fitting parameter θ and

$$\omega = \left\{(\lambda_1,\lambda_2) : \left|\sum_{x=0}^{\infty} f_1'(g-f_\theta)\right| < \epsilon, \left|\sum_{x=0}^{\infty} f_1''(g-f_\theta)\right| < \epsilon, \left|\sum_{x=0}^{\infty} f_2'(g-f_\theta)\right| < \epsilon, \left|\sum_{x=0}^{\infty} f_2''(g-f_\theta)\right| < \epsilon\right\},$$

where ϵ is some infinitesimal number. The continuity of the first derivative of a Poisson density in the parameter guarantees the existence of such a subset ω. For any parameter value $\theta \in \omega$, the matrix J is in the form $J = \sum\limits_{x=0}^{\infty} u_\theta(x)u_\theta^T(x)f_\theta^2(x)$, which is positive definite, assuming $\lambda_1 \neq \lambda_2$.

5. When the tuning parameter $\alpha = 1$, then $V_{n,\theta}(x) = \sum\limits_{x=0}^{\infty} f_\theta^2(x) - 2f_\theta(x)$. The third derivative of $V_{n,\theta}$ is in the form of

$$\frac{\partial^3}{\partial\pi\partial\lambda_1\partial\lambda_2} V_{n,\pi,\lambda_1,\lambda_2} = 2(1-2\pi)\sum_{x=0}^{\infty}\left\{\frac{\partial}{\partial\lambda_1} f_1(x;\lambda_1)\frac{\partial}{\partial\lambda_2} f_2(x;\lambda_2)\right\},$$

and

$$\left|\frac{\partial^3}{\partial\pi\partial\lambda_1\partial\lambda_2} V_{n,\pi,\lambda_1,\lambda_2}\right| = 2\left|\sum_{x=0}^{\infty}\left\{(1-2\pi)\frac{\partial}{\partial\lambda_1} f_1(x;\lambda_1)\frac{\partial}{\partial\lambda_2} f_2(x;\lambda_2)\right\}\right|$$
$$\leq 2\sum_{x=0}^{\infty}|1-2\pi|\left|\frac{\partial}{\partial\lambda_1} f_1(x;\lambda_1)\right|\left|\frac{\partial}{\partial\lambda_2} f_2(x;\lambda_2)\right| = M_{jkl}(x) = E_G[M_{jkl}(x)].$$

Since $0 \leq \pi \leq 1$, $|1-2\pi| \leq 1$ and $M_{jkl}(x) \leq 2\sum\limits_{x=0}^{\infty}\left|\frac{\partial}{\partial\lambda_1} f_1(x;\lambda_1)\right|\left|\frac{\partial}{\partial\lambda_2} f_2(x;\lambda_2)\right|$. Also it has been shown in Section 4.1 that

$$\frac{\partial}{\partial\lambda} f(x;\lambda) = f(x-1;\lambda) - f(x;\lambda) \qquad x = 1,2,3,\ldots$$

and

$$\frac{\partial}{\partial\lambda} f(x;\lambda) = -f(x;\lambda) \qquad x = 0.$$

Both $\left|\frac{\partial}{\partial\lambda_1} f_1(x;\lambda_1)\right|$ and $\left|\frac{\partial}{\partial\lambda_2} f_2(x;\lambda_2)\right|$ are less than or equal to 1. We obtain

$$M_{jkl}(x) \leq 2\sum_{x=0}^{\infty}\left|\frac{\partial}{\partial\lambda_1} f_1(x;\lambda_1)\right| = 2\left(f_1(0;\lambda_1) + \sum_{x=1}^{\infty}|f_1(x-1;\lambda_1) - f_1(x;\lambda_1)|\right).$$

Also

$$|f_1(x-1;\lambda_1) - f_1(x;\lambda_1)| \leq |f_1(x-1;\lambda_1) + f_1(x;\lambda_1)| = f_1(x-1;\lambda_1) + f_1(x;\lambda_1),$$

and

$$M_{jkl}(x) \leq 2(f_1(0;\lambda_1) + 1 + 1 - f_1(0;\lambda_1)) = 4.$$

So $E_G[M_{jkl}(x)]$ is bounded for any $(\pi, \lambda_1, \lambda_2)$ in the parameter space and condition (5) in Theorem 2.2 in Basu et al. (1997) is satisfied.

The consistency and asymptotic normality are proven for $\hat{\pi}_{L2E}$.

7 REFERENCE

Basu, A, Harris, I.R., Hjort, N.L. and Jones, M.C. (1997) Robust and Efficient Estimation by Minimising a Density Power Divergence. *Technical Report; University of Oslo,Norway.*

Basu, A, Harris, I.R., Hjort, N.L. and Jones, M.C. (1998) Robust and Efficient Estimation by Minimising a Density Power Divergence. *Biometrika* 85, 549-559.

Choi, K and Bulgren, W.G.(1968) An estimation procedure for mixtures of distributions. *Journal of the Royal Statistical Society B* 30, 444-460.

Clarke, B.R. (1989). An unbiased minimum distance estimator of the proportion parameter in a mixture of two normal distributions. *Statistics & Probability Letters* 7, 275-281.

Clarke, B.R. and Heathcote, C.R. (1994). Robust estimation of k-component univariate normal mixtures. *Annals of the Institute of Statistical Mathematics* 46, 83-93.

Dempster, A.P., Laird, N.M., and Rubin, D.B. (1977). Maximum likelihood estimation from incomplete data via the EM algorithm (with discussion). *J.R.Statist.Soc.B*, 39, 1-38.

Feng, Z.D. and McCulloch, C.E. (1996). Using bootstrap likelihood ratio in finite mixture models. *Journal of the Royal Statistical Society B* 58, 609-617.

Gelfand, A.E. and Smith, A.F.M. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association* 85, 398-409.

Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J. and Stahel, W.A. (1985) *Robust Statistics: The Approach Based on Influence Functions.* New York: Wiley

Hjort, N.L. (1994) Minimum L_2 and robust Kullback-Leibler estimation. In *Proceedings of the 12th Prague Conference on Information Theory, Statistical Decision Functions and Random Processes,* eds. P. Lachout and J.À. Víšek, pp.102-105. Prague: Academy of Sciences of the Czech Republic.

Juarez, S. (2003). *Robust and Efficient Estimation for the Generalized Pareto Distribution.* Ph.D. thesis, Southern Methodist University.

Karlis, D and Xekalaki, E (1998). Minimum Hellinger distance estimation for Poisson mixtures. *Computational Statistics & Data Analysis* 29, 81-103.

Lehmann, E.L. (1998). *Elements of large-sample theory.* New York: Springer-Verlog

MacDonald, P.D.M. (1971). Comment on a paper by K.Choi and W.G.Bulgren. *Journal of the Royal Statistical Society B*, 33, 326-329.

McLachlan, G. and Peel, D. (2001) *Finite Mixture Models.* New York: Wiley

Pearson, K. (1894) Contributions to the theory of mathematical evolution. *Philosophical Transactions of the Royal Society of London A* 186, 343-414.

Pilla, R.S. and Lindsay, B.G. (2001) Alternative EM methods for nonparametric finite mixture models. *Biometrika* 88, 535-550.

Scott, D.W. (1999). Parametric modelling by minimum L_2 error. *Technical Report, Department of Statistics, Rice University.*

Shen, S. (2004). *The Minimum L_2 Distance Estimator in Poisson Mixtures.* Ph.D. thesis, Southern Methodist University.

Simpson, D.G. (1987). Minimum Hellinger Distance Estimation for the Analysis of Count Data. *Journal of the American Statistician Association* 82, 802-807.

Titterington, D.M., Smith, A.F.M., and Makov, U.E. (1985). *Statistical Analysis of Finite Mixture Distributions.* New York: Wiley

Tukey, J. W. (1970-1971). *Exploratory Data Analysis.* Addison-Wesley, Reading, Mass.

Wolfowitz, J.(1957). The minimum distance method. *Ann.Math.Statist.*, 28, 75-88.

Woodward, W.A., Parr, W.C., Schucany, W.R., and Lindsay, H. (1984). A comparison of minimum distance and maximum likelihood estimation of a mixture proportion. *Journal of the American Statistical Association* 79, 590-598.

Woodward, W.A., Whitney, P., and Eslinger, P.W. (1995). Minimum Hellinger distance estimation of mixture proportions. *Journal of Statistical Planning and Inference* 48, 303-319.

Yakowitz, S.J. (1969). A consistent estimator for the identification of finite mixtures. *Annals of Mathematical Statistics* 40, 1728-1735.