ELSEVIER Computational Statistics & Data Analysis 29 (1998) 81–103

COMPUTATIONAL STATISTICS & DATA ANALYSIS

Minimum Hellinger distance estimation for Poisson mixtures

Dimitris Karlis, Evdokia Xekalaki*

Department of Statistics, Athens University of Economics and Business, 76 Patission Str., 104 34 Athens, Greece

Received 1 March 1997; received in revised form 1 October 1997; accepted 2 April 1998

Abstract

Minimum Hellinger distance (MHD) Estimation is an appealing method of estimation for discrete data as it works well in cases where the assumed model provides a poor fit to observed data and the maximum-likelihood (ML) method fails. Often, spurious observations that may cause problems to the ML method do not seem to affect the MHD method which in general performs better with such data. In this paper we derive MHD estimates for finite Poisson mixtures. The properties of these estimators are examined and a comparison is made of their performance relative to that of the ML estimators. MHD estimators are both efficient and robust. A numerical example involving data sets on environmental complaints is presented. An iterative algorithm that facilitates computation is provided. The algorithm always converges to a minimum, but several initial values are needed to ensure that the global minimum is obtained.

Keywords: Algorithm; Efficiency; Environmental data; Minimum distance; Robustness

Introduction

Minimum-distance estimation methods are appealing in parametric inference especially in cases where the model is suspected to be inexact. In fact, the ML method is of minimum distance type, since the maximization of the likelihood is equivalent to the minimization of the Kullback–Leibler distance.

A variety of distances have been considered for mixture models, especially for normal mixtures. Choi and Bulgren (1968) and Henna (1983) examined the

* Corresponding author.

0167-9473/98/$19.00
PII S0167-9473(98)00047-4

averaged L_2-norm, Bartlett and MacDonald (1968) employed the weighted L_2-norm while MacDonald (1971) and Woodward et al. (1984) considered the Cramer–Von Mises distance. Deely and Kruse (1968) and Blum and Susarla (1977) used the sup-norm (or Kolmogorov distance), Phillips (1990) applied the Chebyshev norm. Further, Fryer and Robertson (1972) considered the chi-square distance and Edelman (1988) and Clarke and Heathcote (1994) employed the square distance. More recently Woodward et al. (1995) used the Hellinger distance for estimating the mixing proportion of 2-finite normal mixtures and Cutler and Cordero-Brana (1996) extended the same method for all the parameters of the mixtures. A comprehensive review of minimum distance methods for mixture problems can be found in Titterington, et al. (1985).

The main difficulty of all such methods is the computational complexity. Numerical techniques are required limiting the applicability of all the methods. On the other hand, the computer implementation of the MLE via the EM algorithm is easy and inexpensive (see for example, Hasselblad, 1969). This has made the ML method the widely used estimation technique for mixture models.

Most of the above-mentioned distances have been applied to normal mixtures but very few have been applied to more general mixture models. Among these, the Hellinger distance is an important member. The robustness of the MHD estimates makes MHD estimation appealing for normal models (see Beran, 1977). Simpson (1987) extended the method to count data and showed that the MHD method works well for data sets prone to outliers. Lindsay (1994) made a thorough comparison of the MHD method to the ML method.

Many robust estimators achieve robustness at some cost in the first-order efficiency. This is not true for the minimum Hellinger distance estimates. Lindsay (1994) shows that MLE and MHDE are members of a larger class of efficient estimators with various robustness and second order efficiency properties. This balancing between robustness and efficiency has caused difficulties since a researcher desires an efficient and robust method. This is not always possible to achieve, so a trade off between these two issues is necessary. The proposed MHD method is examined in relation to both of these issues.

Before getting into to the details of the method let us consider the following artificial example that motivated our research. A sample of size 25 is taken from a 2-finite Poisson mixture. Table 1 contains the observed frequencies.

In order to demonstrate how an outlier can adversely affect parameter estimates using this sample, we contaminated the data by adding a new observation far from the bulk of the data, namely, we added an observation with value $X_{26} = 12$. Table 2 contains the estimates for the uncontaminated (original) model (UN) and the contaminated model (CO) by both the ML and the MHD methods of estimation. Deviations from the true values are attributed to random sampling as well as the small sample size.

The new observation (which may be considered an outlier) influenced the MLE very much while its influence on the MHDE is almost negligible. This example gives an indication that the MHD estimates may work better than the ML estimates in situations when there are outliers.

Table 1
The observed frequencies for a simulated sample of size $n = 25$, from a 2-finite Poisson mixture with $p_1 = 0.5$, $\lambda_1 = 1$ and $\lambda_2 = 3$

x	0	1	2	3	4	5	6	
frequency	8	4	5	1	3	2	2	$n = 25$

Table 2
MLE and MHDE for the uncontaminated (UN) and the contaminated (CO) data

	Model	p_1	λ_1	λ_2
MLE	UN	0.470	0.480	3.425
	CO	0.591	0.768	4.813
MHDE	UN	0.409	0.354	2.992
	CO	0.399	0.368	3.043

If a three-component Poisson distribution is considered, then the ML estimates will be $p_1 = 0.4211$, $p_2 = 0.5363$, $p_3 = 0.0426$, $\lambda_1 = 0.4062$, $\lambda_2 = 3.3031$ and $\lambda_3 = 11.2976$. We can see that the third component has led to a mixing proportion of 0.0426, and a parameter estimate of 11.2976. The contamination was effected through one of the 26 observations ($x = 12$) representing almost 4% of our data. In other words, the outlier observation is treated as one more component of the model while the other two components are very close to the values obtained before the contamination. This fact has lead researchers to the strategy of fitting one more component for possible outliers when using mixture methods (e.g., Aitkin and Wilson, 1980, Harris and Basu, 1994). The above results support the use of MHD method as robust alternative to the ML method in the presence of outliers.

In Section 2, the MHD method is described for Poisson mixtures and properties of the estimators are examined in Section 3. In Section 4 an algorithm is given for the derivation of the estimates, which is easily programmed and provides an insight into the method. In Section 5, the new method is applied to a real data set. In Section 6 the MHD method of estimation is compared to the ML method. In Section 7, we summarize the behaviour of the new method and we discuss possible applications of the method.

2. Minimum Hellinger estimation for Poisson mixtures

Suppose that $f_n(x)$ is the observed proportion of the value x from a sample of size n and $f_\theta(x)$ is the probability under the assumed model that the random variable X takes the value x, $X \in \mathcal{X}$. Here $\mathcal{X}$ denotes the set of all possible discrete outcomes and θ denotes the vector of parameters of interest. The MHDEs for discrete data

can be derived as the vector $\theta_{\min}$ which minimizes the Hellinger distance D given by

$$D(f_n, f_\theta) = \sum_{x \in \mathscr{X}} [\{f_n(x)\}^{1/2} - \{f_\theta(x)\}^{1/2}]^2. \tag{1a}$$

Letting $\rho(f_n, f_\theta) = \sum_{x \in \mathscr{X}} \{f_n(x) f_\theta(x)\}^{1/2}$ the Hellinger distance can be written as

$$D(f_n, f_\theta) = 2 - 2\rho(f_n, f_\theta). \tag{1b}$$

Note that minimizing $D(f_n, f_\theta)$ is equivalent to maximizing $\rho(f_n, f_\theta)$.
For k-finite Poisson mixtures $f_\theta(x)$ has the form

$$f_\theta(x) = \sum_{i=1}^{k} p_i \frac{e^{-\lambda_i} \lambda_i^x}{x!}, \quad x = 0, 1, \ldots, \tag{2}$$

where $\theta = (p_1, p_2, \ldots, p_{k-1}, \lambda_1, \lambda_2, \ldots, \lambda_k)$ is the vector of the unknown parameters to be estimated, $\lambda_i > 0$, $i = 1, 2, \ldots, k$ and $p_i \in (0, 1)$ for $i = 1, 2, \ldots, k$ with $\sum_{i=1}^{k} p_i = 1$. Finite Poisson mixtures naturally arise as models when the whole population consists of k subpopulations each having a Poisson distribution with parameter λ_i, $i = 1, 2, \ldots, k$, and proportions p_i, $i = 1, 2, \ldots, k$,. Such models are used to describe data that are overdispersed and hence cannot be fitted by a simple Poisson distribution. The distance given in (1a) has to be minimized to obtain MHD estimates for the parameters of the k-finite Poisson distribution. The system of estimating equations can be written in the form

$$\sum_{x=0}^{\infty} \left(\frac{f_n(x)}{f_\theta(x)} \right)^{1/2} (f(x, \lambda_j) - f(x, \lambda_k) = 0, \quad j = 1, 2, \ldots, k-1, \tag{3}$$

$$\sum_{x=0}^{\infty} \left(\frac{f_n(x)}{f_\theta(x)} \right)^{1/2} p_j (f(x-1, \lambda_j) - f(x, \lambda_j) = 0, \quad j = 1, 2, \ldots, k-1, \tag{4}$$

where

$$f(x, \lambda) = \frac{\exp(-\lambda)\lambda^x}{x!},$$

i.e. the probability function of a Poisson distribution with parameter λ which we will denote as $P(\lambda)$ in the sequel.

An analytical solution of the above system of equations is not feasible. Numerical methods are required to solve it.

3. Some statistical aspects

3.1. Properties of the estimators

Some properties of the estimators will be derived. All these properties are consequences of the theorems given by Simpson (1987) for MHD estimators of discrete distributions.

3.1.1. Identifiability

The identifiability of Poisson mixtures (finite or not) has been proved by Teicher (1961). We say that mixtures of the probability function $g(x|\theta)$ are identifiable if and only if $\int g(x|\theta)h_1(\theta)\,d\theta = \int g(x|\theta)h_2(\theta)\,d\theta$ implies that $h_1(\theta) = h_2(\theta)$ for all the values of θ. We replace the integration by summation in the case of discrete mixtures. If $g(x|\theta)$ is the Poisson distribution the mixtures are identifiable. In finite mixtures, however, we restrict the parameters λ_i, $i = 1, 2, \ldots, k$ to be in ascending order as interchanging their values and their mixing proportions would lead to the same finite mixture. This identifiability assumption is necessary for the application of Simpson's theorems.

3.1.2. Existence

From Theorem 1 of Simpson (1987), the MHD estimators exist, since the k-finite Poisson mixtures are continuous in their parameters for each x. Their consistency is also a consequence of Simpson's Theorem 1. Finite Poisson mixtures are identifiable (see Teicher, 1961) and, thus, the estimates are consistent and asymptotically unbiased.

3.1.3. Asymptotic Normality

Simpson (1987) imposed some smoothness conditions on the derivatives of $f_\theta(x)$ to prove the asymptotic normality of the estimators. These conditions are satisfied by the k-finite Poisson mixtures since the derivatives of the probability function of a k-finite Poisson mixture with respect to the parameters are linear functions of the probability functions of the component Poisson distributions. This property holds for mixtures of any distribution for which Simpson's (1987) conditions are satisfied. From the above argument and Simpson's Theorem 2, the asymptotic normality of the estimators can be established. So, the MHD estimator follows asymptotically a multivariate normal distribution with mean vector θ and variance–covariance matrix V. The variance covariance matrix V is calculated as $V = H^{-1}\,\mathrm{I}(\theta)\,H^{-1}$, where H is the matrix with its ijth element equal to

$$H_{ij} = \frac{\partial^2 \rho(f_n, f_\theta)}{\partial\theta_i \partial\theta_j}$$

$$= -\frac{1}{4}\sum_{x=0}^{m} [f_n(x)]^{1/2} \frac{2\dfrac{\partial^2 f_\theta(x)}{\partial\theta_j \partial\theta_i} f_\theta(x) - \dfrac{\partial f_\theta(x) \partial f_\theta(x)}{\partial\theta_j \quad \partial\theta_i}}{(f_\theta(x))^{3/2}}, \quad i, j = 1, 2, \ldots, 2k-1$$

and $I(\theta)$ is the Fisher information matrix.

Our simulations showed that the convergence to normality is rather slow. Figs. 1 and 2 present normal plots for the estimators for simulated data from 2-finite mixtures with parameters $p_1 = 0.5$, $\lambda_1 = 1$ and $\lambda_2 = 3$ and $p_1 = 0.5$, $\lambda_1 = 1$ and $\lambda_2 = 10$, respectively. We can see that for well-separated data the convergence is faster. Moreover, the distribution of the MHD estimator compared to that of the ML estimator is quite closer to a normal distribution. Finally, the distributions of the ML estimators exhibit longer tails.

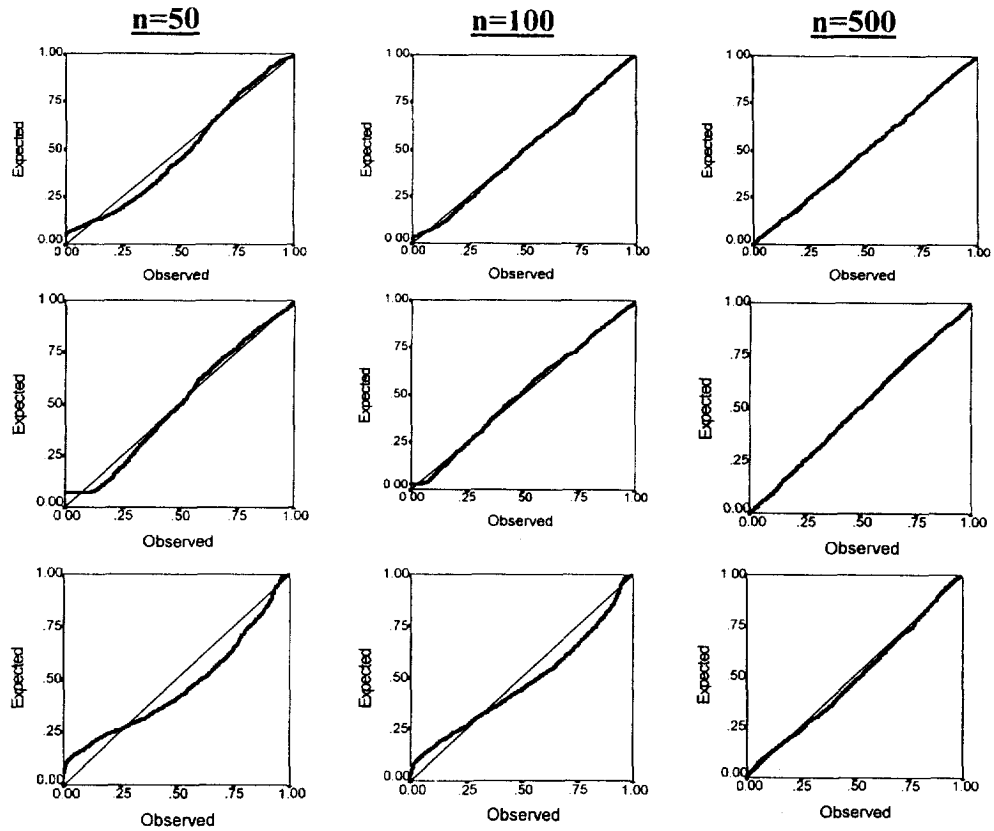

Fig. 1. Normal P–P plots for p_1, λ_1 and λ_2. 1000 samples of size n were drawn from a 2-finite mixture with parameters 0.5, 1, 3. The sample sizes used were $n = 50, 100, 500$. We can see that as the sample size increases the estimates tend to normality. The rows are the parameters and the columns the sample sizes.

3.2. *Measures of robustness and breakdown points*

To examine the robustness of estimators certain performance measures are needed. A common measure is the influence function (IF) defined by

$$\mathrm{IF}(x, T, F) = \lim_{t \downarrow 0} \frac{T((1-t)F + t\Delta_x) - T(F)}{t},$$

whenever this limit exists (see e.g. Hampel et al., 1985). $T(F)$ is a functional based on the distribution function F, which is usually the empirical distribution function of the data, and Δ_x is a degenerate distribution at x. For the MHD estimators the functional $T(F)$ is defined as $T(F) = \{\theta \in \Theta : D(f_n, f_\theta)$ is minimized $\}$ where Θ is the parameter space. The importance of the influence function lies in its heuristic interpretation: it describes the effect on the estimate of an infinitesimal contamination at the point x, standardized by the mass of contamination. The IF is often hard

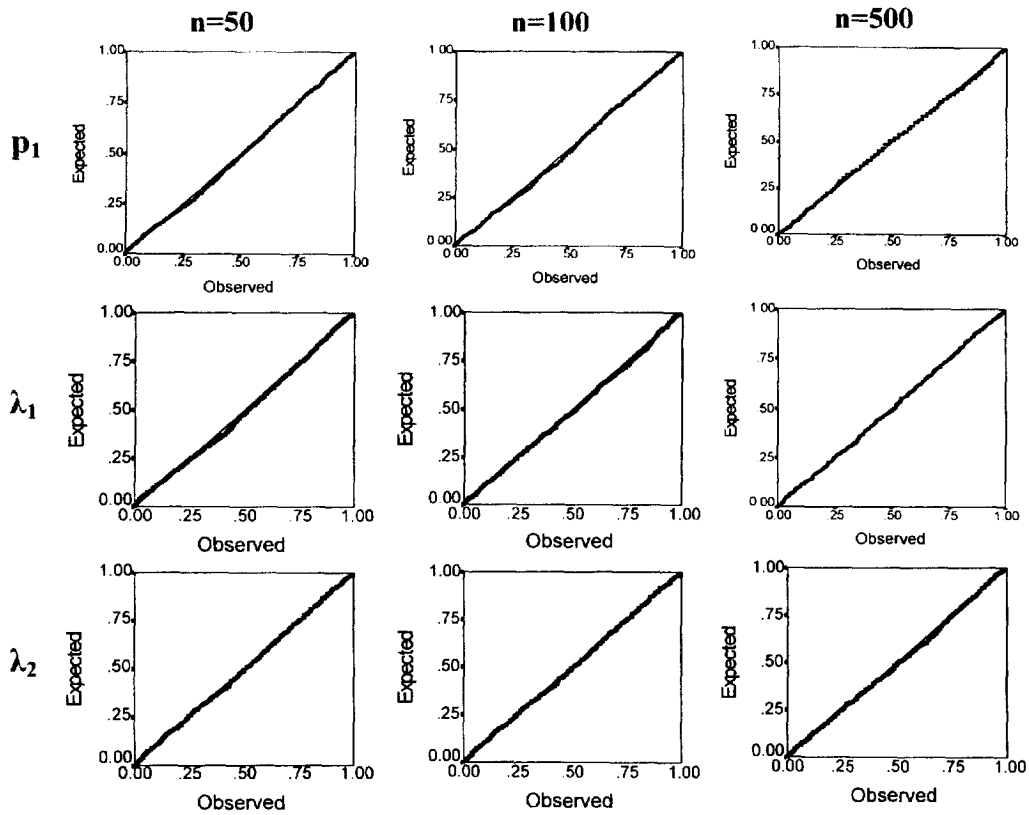

Fig. 2. Normal *P-P* plots for p_1, λ_1 and λ_2. 1000 samples of size n were drawn from a 2-finite mixture with parameters 0.5, 1, 10. The sample sizes used were $n = 50, 100, 500$. We can see that as the sample size increases the estimates tend to normality. The rows are the parameters and the columns the sample sizes.

to calculate and thus other versions are more appropriate. One alternative is to use the empirical counterpart of the IF, namely, the empirical influence function (EIF). According to Hampel et al. (1985, pp. 93), the EIF of the estimator based on any sample is a plot of the values of the estimator if one more observation (contaminant) is added at the point x. Fig. 3a–c provides the EIF of the sample in the example of Section 1 for the 2-finite Poisson mixture and its estimates, for both the ML and the MHD methods of estimation. Each time we contaminated our initial sample of size 25 (given in Table 1), by adding to it a 26th observation at the point x, ($x = 0, 1, 2, \ldots, 20$) and calculated both the MLE and the MHDE on the resulting contaminated sample of size 26.

From Fig. 3a–c we can see that the MHD estimates are not influenced much by the addition of one more observation, especially at points far from the main body of the sample. It is interesting how stable the MHD estimator remains for $x > 10$. Jorgensen (1990) proposed the use of the EIF as a diagnostic tool for the influence of an observation in finite mixture models. He also reported the influence of observations far from the main body of data to the MLE.

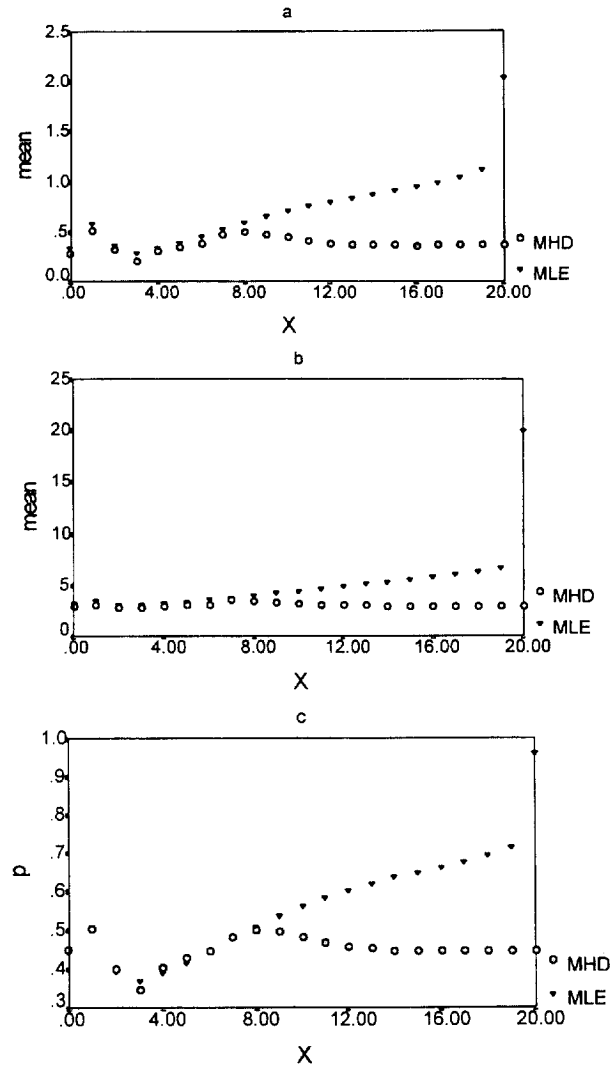

Fig. 3. The empirical influence function for the parameters of a 2-finite Poisson mixture fitted to the data in Table 1. Fig. 1a–c depicts the functions for λ_1, λ_2 and p_1 respectively. The MHD method seems to be more robust than the ML method when an outlier is present to the data.

An alternative measure of robustness is the α-influence function, (Beran, 1977). This measures the change in the estimators if we add one more component in the model and we assign to it a probability equal to α. In particular, the α-influence function (α-IF) is defined as

$$a - \mathrm{IF}(z, T, F) = \lim_{a \downarrow 0} \frac{T((1-a)F + ag_z) - T(F)}{a},$$

where g_z is one more component of the same distribution (Poisson in our case), with parameter z.

The difference from the simple IF is that the simple IF measures the influence of one more observation at the point x, while the α-IF measures the influence of one more component with mixing proportion α. To illustrate this, consider the model 0.5 $P(1) + 0.5\ P(3)$ and also consider a $P(12)$ distribution ($z = 12$) as a contaminant. Analytical evaluation of the α-IF is not possible, so we calculated the α-IF numerically for both the MHD and ML methods of estimation. These are depicted in Fig. 4a–c.

For small α (low contamination) the MHD estimates are far better than the ML estimates but with α increasing, the two methods work in the same manner. We

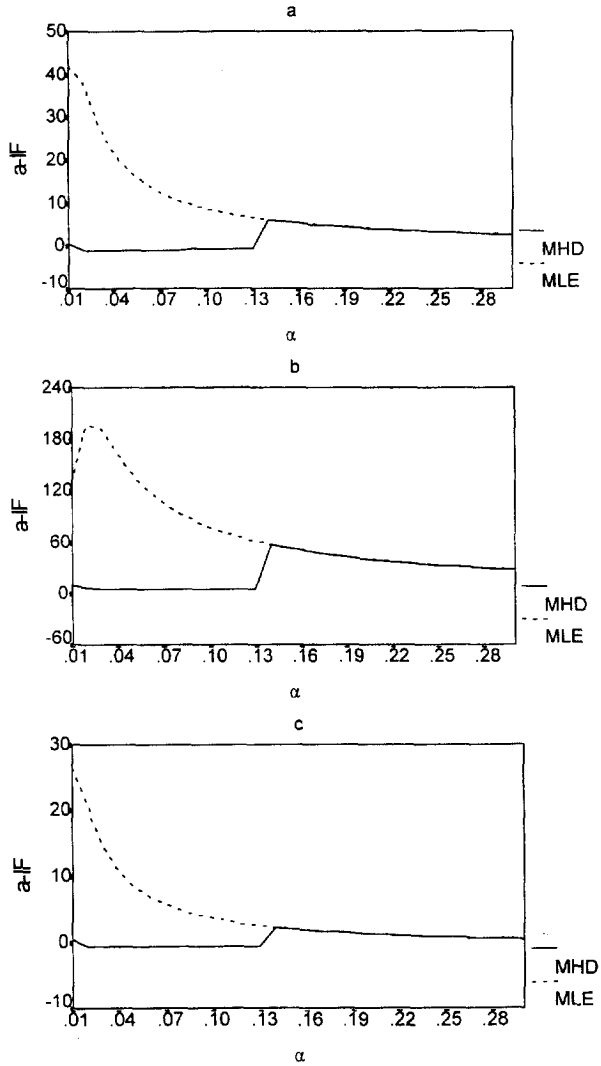

Fig. 4. The α-Influence function for the model [0.5 Poisson(1) + 0.5 Poisson(3)] with a Poisson (12) distribution as the contaminant. Fig. 2a–c depicts the function λ_1, λ_2 and p_1, respectively. If α is small the ML method is influenced very much. note also that the influence is larger for the parameter λ_2. The MHD method seems to be more robust to contamination.

often consider a few observations far from the main body of the data to be outliers. In other words, we may regard a small fraction α of the observations as outliers. If the proportion of spurious observations is large, then clearly they cannot be regarded as outliers. The jump of the α-IF of the MHD method at the point $\alpha = 0.13$ can be considered as the breakdown point in the sense of Simpson (1987).

Lindsay (1994) criticized the ability of the influence function to present the robustness properties of estimators and, in particular, the robustness of MHD estimators. He proposed the use of the residual adjustment function (RAF). More specifically, Lindsay (1994) showed that the Influence function underestimates the true robustness of a method, and he proposed the use of the RAF for comparing the robustness of different methods. It should be noted however that the RAF cannot be used for detecting the influence of one more component. This is possible with the use of the α-influence function which by its definition measures how the existence of one more component influences the any estimator. Moreover, the approach based on the RAF lacks simple interpretation which can be given using the Influence function (i.e. the relative change of the estimate when a new observation is added).

4. The HELMIX algorithm

In this section an algorithm is given to facilitate the estimation procedure. The algorithm is developed using the estimating equations given in Section 2. From (4), using the recurrence relation $f(x, \lambda) = f(x-1, \lambda)\,\lambda/x$ for the poisson probabilities we obtain

$$\sum_{x=0}^{m} (f_n(x))^{1/2} w_{xj}(x - \lambda_j) = 0, j = 1, 2, \ldots, k,$$

where m denotes the largest observed value and $w_{xj} = f(x, \lambda_j)\{f_\theta(x)\}^{-1/2}$. Solving these equations with respect to the parameters λ_j, $j = 1, 2, \ldots, k$ we obtain

$$\lambda_j = \frac{\sum_{x=0}^{m} w_{xj} x \{f_n(x)\}^{1/2}}{\sum_{x=0}^{m} w_{xj}\{f_n(x)\}^{1/2}}, \quad j = 1, 2, \ldots, k,$$

i.e., the MHD estimates are weighted versions of the sample mean. From (3) we obtain

$$\sum_{x=0}^{m} \left\{\frac{f_n(x)}{f_\theta(x)}\right\}^{1/2} f(x, \lambda_j) = \sum_{x=0}^{m} \left\{\frac{f_n(x)}{f_\theta(x)}\right\}^{1/2} f(x, \lambda_k) \quad \text{for } j = 1, 2, \ldots, k, \tag{5}$$

Also, multiplying the ith equation in (3) by p_j and adding the resulting equations yields

$$\sum_{x=0}^{m} \left\{\frac{f_n(x)}{f_\theta(x)}\right\}^{1/2} f(x, \lambda_k) = \sum_{x=0}^{m} \{f_n(x) f_\theta(x)\}^{1/2} \quad \text{for} \quad j = 1, 2, j = 1, 2, \ldots, k. \tag{6}$$

Eqs. (5) and (6) lead to

$$\sum_{x=0}^{m} \{f_n(x) f_\theta(x)\}^{1/2} = \sum_{x=0}^{m} \{f_n(x)\}^{1/2} w_{xj}, \quad j = 1, 2, \ldots, k. \tag{7}$$

Following Behboodian (1969), we may multiply them by p_j obtaining

$$p_j = \frac{\sum_{x=0}^{m} p_j w_{xj} \{f_n(x)\}^{1/2}}{\sum_{x=0}^{m} \{f_n(x) f_\theta(x)\}^{1/2}}, \quad j = 1, 2, \ldots, k-1, \text{ and } p_k = 1 - \sum_{i=1}^{k-1} p_i.$$

This provides the basis for the genesis of the following algorithm for MHD estimation of Poisson mixtures (HELMIX). Note that this derivation is very similar to that introduced by Behboodian (1969) for the derivation of the ML estimators in the case of finite normal mixtures.

Step 1: Given the values obtained from the ith iteration $\lambda_j^{(i)}, j = 1, 2, \ldots, k$, and $p_j^{(i)}, j = 1, 2, \ldots, k$, calculate the weights $w_{xj} = f(x, \lambda_j) \{f_\theta(x)\}^{-1/2}$, where $f(x, \lambda_j)$ and $f_\theta(x)$ are calculated using the estimates from the ith iteration.

Step 2. Calculate the new parameter estimates using

Step 2a: $$\lambda_j^{(i+1)} = \frac{\sum_{x=0}^{m} w_{xj} x [f_n(x)]^{1/2}}{\sum_{x=0}^{m} w_{xj} [f_n(x)]^{1/2}}, \quad j = 1, 2, \ldots, k.$$

Step 2b $$p_j^{(i+1)} = \frac{\sum_{x=0}^{m} p_j^{(i)} w_{xj} [f_n(x)]^{1/2}}{\sum_{x=0}^{m} [f_n(x) f_\theta(x)]^{1/2}}, \quad j = 1, \ldots, k-1$$

and

$$p_k^{(i+1)} = 1 - \sum_{j=1}^{k-1} p_j^{(i+1)},$$

where m denotes the largest observed value.

Step 3. Check if some convergence criterion is satisfied, otherwise go back to step 1, using the current estimates as initial values to make the next iteration.

The procedure requires initial values for the estimates. If the initial values are within the acceptable range for the parameters, the estimated values are also within this range. In particular, the λ_j's remain positive if their starting values are positive. For the mixing proportions p_j it suffices to observe at step 2b that $p_j^{(i+1)}$ is less than or equal to $\sum_{x=0}^{m} w_{xj} [f_n(x)]^{1/2} / \sum_{x=0}^{m} [f_n(x) f_\theta(x)]^{1/2}$ which by (7) equals 1.

Note the similarity of the HELMIX algorithm with the EM algorithm for mixture models, introduced by Hasselblad (1969) and examined more formally by Redner and Walker (1984). If we were to use $w_{xj} = f(x, \lambda_j^{(i)}) / f_\theta(x)$ as weights in step 1 and the observed frequencies themselves instead of their square roots in step 2, then HELMIX algorithm reduces to the EM algorithm for ML estimation for finite mixture models.

The HELMIX algorithm shares some weaknesses of the EM algorithm for ML estimation in the case of mixture models. These are slow convergence and dependence on the choice of the initial values. For all our simulations the algorithm converged to a minimum. However, the attained minimum might not be a global one. A good strategy is to start from several different initial values so as to ensure

that the global minimum is obtained. One may start with equidistant points from 0 to the maximum observed value given equal probabilities and stop iterating when the maximum relative difference between the estimated parameters of two successive iterations first becomes less than a small number, say tol (we used tol = 0.00001).

5. An application

To illustrate the MHD method of estimation for data sets prone to outliers in the case of k-finite Poisson mixtures, consider the data in Table 3. They concern the number of environmental complaints placed by phone in an environmental station in Nederlands for the year 1985. The high overdispersion of the data makes the use of a mixed Poisson model appropriate to model the number of environmental complaints. The mean is 22.11 while the variance is 324.08 (almost 15 times bigger than the mean). Moreover, the data are highly skewed, with a very long right tail. The ML estimators are expected to be influenced by the data at the tail and thus their estimates will not be a reasonable choice. The MHD method of estimation may be more appropriate as it seems not to be affected so much by the observations at the right tail.

To these data a 3-finite Poisson distribution was fitted using both the ML and MHD methods of estimation. Table 3 contains the expected frequencies using both methods while Table 4 provides the parameter estimates.

In Fig. 5a–c we can see the observed frequencies with the fitted frequencies for both the methods. From Table 3 it can be observed that the distribution fitted by the ML method of estimation has a heavier right tail than that of the observed

Table 3
Observed and expected frequencies of environmental complaints placed in an environmental station in 1985

x	Observed frequencies	Expected frequencies		x	Observed frequencies	Expected frequencies	
		MHDE	MLE			MHDE	MLE
0–4	37	22.95	4.70	45–49	11	7.20	2.19
5–9	67	96.71	85.71	50–54	3	1.85	0.30
10–14	69	60.75	114.47	55–59	3	0.30	0.13
15–19	56	70.20	26.19	60–64	7	0.03	0.49
20–24	28	37.76	8.68	65–69	2	0.01	1.47
25–29	23	12.17	25.19	70–79	3	0	7.64
30–34	21	16.08	37.42	80–89	1	0	9.05
35–39	13	22.19	26.90	90–99	2	0	3.65
40–44	13	16.78	10.20	≥ 100*	6	0	0.60

*The actual observations were (102,108,118,134,158,185).

Table 4
The parameter estimates for both the methods for data in Table 3

	p_1	p_2	λ_1	λ_2	λ_3
MHDE	0.390	0.418	7.136	17.331	37.676
MLE	0.635	0.302	10.559	32.587	81.423

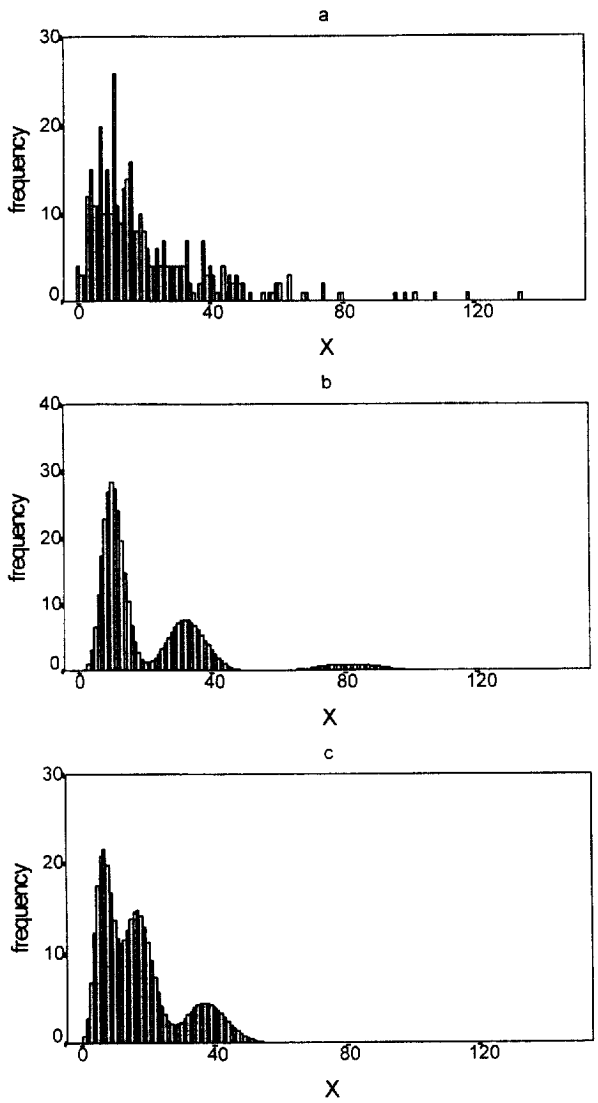

Fig. 5. Histograms of the observed frequencies (3a), the expected frequencies via the ML method (3b) and the expected frequencies via the MHD method (3c) for the number of environmental complaints placed by phone at an environmental station for the year 1985, are presented in Table 3. The ML method assigns one component at the tail of the distribution due to the presence of a few large observations.

distribution with a bump in the range 70–89. The distribution fitted by the MHD method, on the contrary, has a smoother right tail and it provides a relatively better fit to the data. The ML method tries to fit a component at the tail and hence the tail influences the estimation. The MHD method seems to ignore these observations. These high values may be outliers, for example, some days with unexpectedly high number of phone calls. The great difference in the two estimates demonstrates how the choice of an estimation method can affect the results. An interpretation of why the markedly different estimates occur is given in the next section.

For the data with a long right tail that can be attributed to "unexpected" or even "unreasonable" values (due to some mistakes in the collection of data) the ML method must be used with care while the MHD method offers an interesting alternative.

6. Comparison of the minimum Hellinger distance method to the maximum-likelihood method

6.1. General comparison

In parametric estimation two fundamental – but potentially competing – aspects are of interest: The aspect of efficiency when the model has been appropriately specified and the aspect of robustness when it has not. Unfortunately, satisfying both is very difficult and thus a trade-off between them is necessary. We will compare the MHD method for finite Poisson mixtures to the ML method with reference to both aspects.

For the parameter θ_i, the estimating equation in the case of the ML method is given by

$$\sum_{x=0}^{\infty} \frac{f_n(x)}{f_\theta(x)} \frac{\partial f_\theta(x)}{\partial \theta_i} = 0,$$

while in the case of the MHD method it is given by

$$\sum_{x=0}^{\infty} \left[\frac{f_n(x)}{f_\theta(x)}\right]^{1/2} \frac{\partial f_\theta(x)}{\partial \theta_i} = 0.$$

Clearly, if the model is well specified and the sample size is large ($n \to \infty$) the quantity under the square root must be close to 1 and, hence, the square root of this quantity is itself close to 1; we expect that the two methods will behave similarly. In the ideal case of exact specification of the model, the ratio $f_n(x)/f_\theta(x)$ equals 1 for every x, and the two methods thus coincide. On the other hand, for values of x for which the ratio $f_n(x)/f_\theta(x)$ is large (as in the case of outliers) the MHD method gives less weight to the estimation being thus not so sensitive to outliers. As a result, the MHD method works better with datasets prone to outliers. Simpson (1987) showed that for large values of x, an improbable count has little impact on the MHDE rather than the MLE. Our results referring to the example of Section 1 on how

a new observation far away from the bulk of data influences the two methods are not in disagreement with Simpson's findings.

Note also that the quantities

$$A_{\mathrm{ML}}(x)=\frac{f_n(x)}{f_\theta(x)} \text{ and } A_{\mathrm{HD}}(x)=\left[\frac{f_n(x)}{f_\theta(x)}\right]^{1/2}$$

are similar to the residuals adjustment functions introduced by Lindsay (1994) in investigating the robustness of estimators.

In robust analysis we consider a contaminated model of the form $(1-e)\,M_1+eM_2$, where M_1 is the underlying model, M_2 is a contaminant which causes the departure from model M_1 and e is the probability that an observation belongs to the contaminant. Note that the quantity e itself is of practical interest since as Simpson (1989) and Lindsay (1994) pointed out, for every model there is a value of e which gives an upper bound of possible contamination. Above this point there is a breakdown point for the model.

Clearly, the above model is a mixture and it is true that contaminated models are described as mixture models (e.g. Titterington et al., 1985, pp. 22). So, for example, a contaminated model for a 2-finite Poisson mixture can be considered as a 3-finite Poisson mixture. The ML method usually models the contamination with an additional component. In our case the ML method will work well for a three-component model, but not for a model with two components.

The remainder of this section is devoted to simulation comparisons of the two methods. The HELMIX algorithm was used for deriving the MHD estimates. It converged in all the cases. The iterations were stopped when the maximum relative difference between the estimated parameters of two successive iterations is less than 0.00001. For the ML method the EM algorithm described by Hasselblad (1969) was used. The same stopping rule was used, but now the relative increase in the distance between two successive iterations was considered. Two sets of initial values were used to increase the chance that the obtained maximum (minimum) was not a local extreme. For both methods, the true parameter values and equiprobable values $m \pm 0.5$, around the sample mean m were considered initial values as such. We examined both correctly specified models and contaminated models. We used a table look-up method for generating the variates. All the programs were written in Pascal.

6.2. *Simulation comparison for correctly specified models.*

First, we examined the case where the model is correctly hypothesized to be a 2-finite Poisson mixture. For each parameter vector four sample sizes were examined, namely, $n = 50, 100, 250, 500$ and each case for given parameters and sample size was replicated 1000 times. The values given to the parameters of the 2-finite Poisson mixtures used in the simulation are shown in Tables 5–8.

The usual method for comparing two methods in multiparameter model estimation utilizes the ratio of the generalized variances of the estimators. For our case,

Table 5
The ratio of generalized variances of the ML estimator divided by that of the MHD estimator for certain 2-finite mixtures and sample sizes. The reported values have been calculated using 1000 replication samples

n	50	100	250	500	50	100	250	500
p_1		$\lambda_1 = 1$	$\lambda_2 = 2$			$\lambda_1 = 1$	$\lambda_2 = 3$	
0.2	9.293	16.035	6.104	1.952	3.897	5.056	3.390	1.301
0.5	4.725	4.115	3.261	0.929	1.625	1.182	1.164	0.955
0.8	4.238	2.772	1.904	1.113	1.359	1.103	0.806	0.680
		$\lambda_1 = 1$	$\lambda_2 = 5$			$\lambda_1 = 1$	$\lambda_2 = 8$	
0.2	1.612	1.271	1.098	1.025	0.702	0.890	0.962	0.994
0.5	0.908	0.964	0.976	0.973	0.628	0.821	0.965	0.980
0.8	0.505	0.662	0.866	0.906	0.570	0.645	0.873	0.934
		$\lambda_1 = 2$	$\lambda_2 = 4$			$\lambda_1 = 2$	$\lambda_2 = 5$	
0.2	7.119	6.317	4.102	1.794	2.579	1.679	1.419	1.103
0.5	2.510	3.062	1.740	0.869	1.140	1.091	0.963	0.937
0.8	2.707	1.865	0.854	0.714	0.952	0.722	0.687	0.660
		$\lambda_1 = 2.8$	$\lambda_2 = 3.2$					
0.2	7.607	28.969	22.886	2.213				
0.5	4.771	14.784	5.547	0.954				
0.8	9.286	9.510	4.449	0.844				

the generalized variances were computed on the basis of the covariance matrices calculated from the simulation. Table 5 summarizes the results for several 2-finite Poisson mixture distributions. The entries are the values $|V_{\mathrm{ML}}|/|V_{\mathrm{MHD}}|$, where $|V|$ denotes the determinant of the covariance matrix, and the subscripts indicate the method used.

Table 5 confirms that the ML method works far better for models with well-separated components. Hasselblad (1969) warned that ML estimators have large standard errors when the components are close together. In this case the MHD method is far superior for all sample sizes. Generally, the MHD method performs better for small sample sizes and low mixing proportions.

Usually, the ML method works better for well-specified models as compared to the MHD method (Lindsay, 1994). However, our results constitute a case in which the MHD method performs better. To further examine this remarkable result we report in Table 6 the estimated relative efficiencies for the parameters. The entries are values of the relative efficiency of any parameter ϕ defined as $\mathrm{reff}(\phi) = \mathrm{Var}(\hat{\phi}_{\mathrm{ML}})/\mathrm{Var}(\hat{\phi}_{\mathrm{MHD}})$. The subscript denotes the method used for obtaining the estimator. The corresponding variances have been calculated using

$$\mathrm{Var}(\hat{\phi}) = \sum_{i=1}^{n} (\hat{\phi}_i - \bar{\hat{\phi}})^2/n,$$

Table 6
Relative efficiencies of the MHD estimators of the parameters of a 2-finite Poisson mixture for certain 2-finite mixtures and sample sizes. The reported values have been calculated using 1000 replication samples

	$p_1 = 0.2$			$p_1 = 0.5$			$p_1 = 0.8$		
	p_1	λ_1	λ_2	p_1	λ_1	λ_2	p_1	λ_1	λ_2
n				$\lambda_1 = 1$	$\lambda_2 = 2$				
50	1.78	1.00	2.33	1.29	1.00	2.07	1.06	1.05	1.98
100	1.60	0.95	2.72	1.19	0.93	1.93	0.96	0.95	1.87
250	1.32	0.86	2.03	1.00	0.83	1.76	0.82	0.79	1.89
500	1.05	0.76	1.51	0.87	0.75	1.29	0.76	0.73	1.68
				$\lambda_1 = 1$	$\lambda_2 = 3$				
50	1.38	0.96	1.81	1.12	0.95	1.53	0.93	0.93	1.41
100	1.44	0.96	2.09	1.03	0.95	1.30	0.86	0.87	1.41
250	1.23	1.00	1.58	0.99	0.97	1.16	0.81	0.82	1.19
500	1.10	0.96	1.13	0.95	0.95	1.03	0.78	0.84	1.06
				$\lambda_1 = 1$	$\lambda_2 = 5$				
50	1.27	1.07	1.36	0.96	1.04	1.03	0.72	0.88	0.99
100	1.14	1.10	1.07	0.97	1.02	1.01	0.82	0.95	0.97
250	1.04	1.03	1.02	0.98	1.01	0.99	0.93	0.99	1.00
500	1.01	1.00	1.00	0.99	1.00	0.99	0.96	0.99	0.99
				$\lambda_1 = 1$	$\lambda_2 = 8$				
50	0.92	1.00	0.94	0.88	1.01	0.90	0.87	0.99	0.83
100	0.96	1.02	0.97	0.95	1.00	0.95	0.90	0.99	0.89
250	0.99	1.01	0.98	0.98	1.01	0.98	0.97	1.00	0.96
500	0.99	1.01	0.99	0.99	1.00	0.99	0.98	1.00	0.98
				$\lambda_1 = 2$	$\lambda_2 = 4$				
50	1.67	0.95	2.17	1.23	0.94	1.89	1.05	1.01	1.77
100	1.58	0.91	2.34	1.12	0.90	1.72	0.94	0.89	1.70
250	1.29	0.87	1.69	1.02	0.91	1.51	0.84	0.79	1.51
500	1.07	0.86	1.32	0.93	0.85	1.13	0.73	0.71	1.17
				$\lambda_1 = 2$	$\lambda_2 = 5$				
50	1.50	0.87	1.84	1.06	0.92	1.42	0.92	0.92	1.29
100	1.36	0.93	1.51	1.00	0.94	1.27	0.84	0.84	1.25
250	1.21	1.01	1.19	0.98	0.97	1.08	0.80	0.85	1.10
500	1.07	0.96	1.07	0.96	0.96	1.02	0.80	0.89	1.01
				$\lambda_1 = 2.8$	$\lambda_2 = 3.2$				
50	2.20	1.14	2.64	1.62	1.18	2.63	1.27	1.21	2.21
100	2.24	1.05	3.29	1.42	1.11	2.88	1.21	1.21	2.30
250	1.78	1.01	3.00	1.22	0.98	2.17	0.87	0.85	2.70
500	0.96	0.77	1.91	0.75	0.67	1.74	0.63	0.60	1.93

Table 7
Relative MSEs for the ML and MHD methods (correctly specified models) for certain 2-finite mixtures and sample sizes. The reported values have been calculated using 1000 replication samples

	$p_1 = 0.2$			$p_1 = 0.5$			$p_1 = 0.8$		
	p_1	λ_1	λ_2	p_1	λ_1	λ_2	p_1	λ_1	λ_2
n				$\lambda_1 = 1$	$\lambda_2 = 2$				
50	3.62	0.98	5.58	1.56	1.00	3.54	1.15	1.13	2.17
100	2.85	0.89	7.57	1.26	0.88	3.33	0.87	0.90	2.13
250	1.96	0.76	4.44	0.88	0.65	3.04	0.55	0.59	2.87
500	1.14	0.58	2.34	0.67	0.50	1.69	0.47	0.48	2.35
				$\lambda_1 = 1$	$\lambda_2 = 3$				
50	2.14	0.94	3.41	1.09	0.85	1.91	0.81	0.83	1.28
100	2.28	0.92	4.33	0.91	0.81	1.43	0.62	0.67	1.39
250	1.58	0.92	2.48	0.86	0.80	1.21	0.53	0.58	1.08
500	1.23	0.85	1.25	0.81	0.81	0.91	0.50	0.61	0.92
				$\lambda_1 = 1$	$\lambda_2 = 5$				
50	1.70	1.16	1.60	0.91	1.01	0.78	0.52	0.70	0.71
100	1.33	1.18	0.97	0.93	0.98	0.79	0.66	0.83	0.67
250	1.09	1.06	0.94	0.96	0.99	0.78	0.85	0.94	0.75
500	1.02	0.99	0.93	0.96	0.98	0.87	0.89	0.94	0.80
				$\lambda_1 = 1$	$\lambda_2 = 8$				
50	0.84	1.01	0.74	0.69	0.98	0.60	0.55	0.90	0.59
100	0.93	1.04	0.81	0.84	0.99	0.70	0.69	0.95	0.57
250	0.98	1.02	0.85	0.93	1.01	0.77	0.90	0.98	0.67
500	0.99	1.01	0.88	0.98	1.01	0.86	0.93	0.99	0.77
				$\lambda_1 = 2$	$\lambda_2 = 4$				
50	3.23	0.89	4.52	1.37	0.88	2.64	1.12	1.06	1.50
100	2.81	0.84	5.40	1.07	0.77	2.57	0.78	0.76	1.70
250	1.80	0.76	2.94	0.84	0.69	2.05	0.55	0.55	1.78
500	1.20	0.69	1.76	0.77	0.63	1.19	0.41	0.42	1.08
				$\lambda_1 = 2$	$\lambda_2 = 5$				
50	2.49	0.76	3.16	1.02	0.79	1.61	0.83	0.82	1.02
100	1.97	0.86	2.21	0.85	0.76	1.23	0.59	0.64	1.08
250	1.53	0.95	1.37	0.86	0.83	0.89	0.54	0.62	0.91
500	1.16	0.88	1.07	0.86	0.85	0.90	0.53	0.69	0.76
				$\lambda_1 = 2.8$	$\lambda_2 = 3.2$				
50	5.32	1.37	7.38	2.52	1.44	6.93	1.69	1.50	4.37
100	5.40	1.16	1.38	1.93	1.30	8.02	1.50	1.51	4.75
250	3.26	1.09	9.21	1.45	0.99	4.62	0.73	0.73	6.94
500	0.92	0.63	3.72	0.54	0.45	2.78	0.38	0.37	3.74

where $\hat{\phi}_i$ is the estimate of the ith sample and $\bar{\hat{\phi}}$ is the mean of the estimated parameter over all samples.

The ML estimator of λ_1 is usually better than the MHD estimator. The difference in the performance of the two estimators is greater with respect to λ_2. An

explanation for this might be that for count data outliers most often occur only at the right tail. As a result, outliers influence this parameter more. In addition, the covariances of the parameter estimators were in general smaller in the case of the MHD method, resulting in the superiority of the MHD method as judged by the generalized variance ratios reported in Table 5.

The relative mean-squared–errors are also reported in Table 7. The relative mean squared–error RMSE(φ) of a parameter φ is defined as RMSE(ϕ) = MSE($\hat{\phi}_{ML}$)/MSE($\hat{\phi}_{MHD}$) where the mean-squared error (MSE) of an estimate $\hat{\phi}$ is given by

$$\mathrm{MSE}(\hat{\phi}) = \sum_{i=1}^{n} (\hat{\phi}_i - \phi)^2/n,$$

where φ is the true value of the parameter.

Examining the relative MSEs of the two methods reported in Table 7, we can see that, for the parameter λ_2 the performance of the ML estimator is inferior. It is the presence of outliers that contributes to this situation. On the other hand, the ML estimator is more accurate for λ_1 and in the case of models with well-separated components.

6.3. *Simulation comparison for contaminated models*

We now examine the robustness of the method when the model is not correctly specified. For this purpose, it was assumed that the data come from contaminated 2-finite Poisson mixtures. In particular, it was assumed that an additional component was present at the right tail of the distribution. The probability α associated with this component, (the level of contamination), was allowed to take three values so as to investigate whether the amount of contamination affects the plausibility of the method. Specifically, α was given the values 0.01, 0.05, 0.1.

The RMSE defined previously was used as a measure of robustness. For some of the models considered the notion of contamination is not well defined. The reason is that the level of contamination is very high relative to the mixing proportion of the smallest component. In this case it is not clear with respect to what parameter the mean-squared error should be calculated. For example, when the parameter $p_1 = 0.8$ and the contamination level α is high, the probability associated with the second component is very close to the probability assigned to the contaminant, namely $p_2 = 0.18$ and $\alpha = 0.10$.

The relative MSEs for several contaminated models are tabulated in Table 8a–d; λ_3 is the parameter of the Poisson variable which is assumed to contaminate the 2-finite Poisson model.

Tables 8a–d show that the MHD estimator is more robust than the ML estimator when the incorrect model is hypothesized, particularly when the sample size is small, the mixing proportion is small and the contaminant is far from the other components. We note again the same behaviour with respect to the parameters, namely that for λ_2 the MHD estimator is almost always more robust while for λ_1 it is less robust, depending on the mixing proportion. A dependence of the robustness of the method on the mixing proportion is again manifested. Careful

Table 8a
Relative MSEs based on 1000 replications from a 2-finite Poisson mixture distribution with $p_1 = p$, $\lambda_1 = 1$, $\lambda_2 = 3$, and contamination α from a distribution with $\lambda_3 = 7$

	$\alpha = 0.01$			$\alpha = 0.05$			$\alpha = 0.1$		
	p_1	λ_1	λ_2	p_1	λ_1	λ_2	p_1	λ_1	λ_2
n					$p = 0.2$				
50	2.84	0.98	9.12	2.91	1.26	8.86	2.05	1.39	4.56
100	2.91	1.09	9.36	2.38	1.55	6.08	1.80	1.61	3.03
250	2.57	1.34	10.73	2.00	1.72	3.65	1.37	1.36	1.81
500	1.84	1.25	5.87	1.69	1.58	2.45	1.22	1.21	1.46
					$p = 0.5$				
50	1.17	0.89	3.62	1.34	1.08	4.40	1.29	1.36	3.53
100	1.13	0.95	4.32	1.45	1.29	3.98	1.34	1.46	2.20
250	1.22	1.07	3.45	1.56	1.56	2.40	1.33	1.42	1.61
500	1.35	1.17	2.26	1.47	1.50	1.75	1.22	1.27	1.36
					$p = 0.8$				
50	0.69	0.74	2.55	0.55	0.70	3.08	0.46	0.87	2.17
100	0.55	0.65	3.31	0.45	0.80	2.68	0.52	1.12	1.78
250	0.54	0.71	3.34	0.92	1.29	1.89	0.95	1.29	1.41
500	0.80	1.03	2.54	1.26	1.40	1.55	1.05	1.23	1.24

Table 8b
Relative MSEs based on 1000 replications from a 2-finite Poisson mixture distribution with $p_1 = p$, $\lambda_1 = 1$, $\lambda_2 = 3$, and contamination α from a distribution with $\lambda_3 = 12$.

	$\alpha = 0.01$			$\alpha = 0.05$			$\alpha = 0.1$		
	p_1	λ_1	λ_2	p_1	λ_1	λ_2	p_1	λ_1	λ_2
n					$p = 0.2$				
50	5.58	1.31	44.50	4.96	2.12	16.04	2.76	2.00	5.50
100	6.71	1.85	74.63	3.06	2.48	7.07	1.70	1.70	2.72
250	7.04	3.32	51.94	1.84	1.87	3.01	1.06	1.12	1.32
500	7.98	5.33	48.85	1.29	1.32	1.70	0.99	1.02	1.12
					$p = 0.5$				
50	1.59	1.12	19.43	2.32	1.83	13.85	1.64	1.80	4.33
100	2.03	1.52	43.09	2.25	2.43	6.58	1.35	1.72	2.28
250	3.56	3.11	42.44	1.88	2.19	3.15	1.03	1.18	1.31
500	5.69	5.32	48.36	1.34	1.51	1.85	0.97	1.05	1.11
					$p = 0.8$				
50	0.65	0.76	11.59	0.32	0.73	6.69	0.22	0.96	2.74
100	0.49	0.63	14.63	0.39	0.93	3.28	0.32	1.08	1.61
250	0.67	1.18	18.28	0.86	1.40	1.73	0.70	1.18	1.15
500	1.62	2.88	16.43	0.93	1.23	1.27	0.82	1.12	1.08

Table 8c
Relative MSEs based on 1000 replications from a 2-finite Poisson mixture distribution with $p_1 = p$, $\lambda_1 = 1$, $\lambda_2 = 5$, and contamination α from a distribution with $\lambda_3 = 12$

	$\alpha = 0.01$			$\alpha = 0.05$			$\alpha = 0.1$		
	p_1	λ_1	λ_2	p_1	λ_1	λ_2	p_1	λ_1	λ_2
n					$p = 0.2$				
50	2.81	1.56	8.57	4.01	2.70	11.75	3.10	2.78	7.19
100	2.52	1.74	5.62	4.50	3.24	10.43	3.73	3.36	5.84
250	1.69	1.40	2.05	5.86	4.24	7.60	3.27	2.89	4.04
500	1.40	1.34	1.96	6.69	4.78	6.35	3.17	2.82	3.32
					$p = 0.5$				
50	1.00	1.22	1.94	1.33	2.20	5.48	1.35	2.49	4.45
100	1.03	1.20	1.54	1.62	2.35	4.70	1.66	2.50	3.43
250	1.09	1.19	1.86	1.93	2.58	3.77	1.87	2.45	2.57
500	1.21	1.26	2.24	2.20	2.64	3.08	1.91	2.26	2.24
					$p = 0.8$				
50	0.51	0.88	2.03	0.39	1.02	2.73	0.47	1.48	2.14
100	0.62	0.95	2.03	0.54	1.50	3.03	0.64	1.60	1.90
250	0.97	1.24	2.76	0.95	1.96	2.54	0.86	1.50	1.49
500	1.20	1.47	3.40	1.09	1.83	2.09	1.00	1.35	1.29

Table 8d
Relative MSEs based on 1000 replications from a 2-finite Poisson mixture distribution with $p_1 = p$, $\lambda_1 = 2$, $\lambda_2 = 5$, and contamination α from a distribution with $\lambda_3 = 10$.

	$\alpha = 0.01$			$\alpha = 0.05$			$\alpha = 0.1$		
	p_1	λ_1	λ_2	p_1	λ_1	λ_2	p_1	λ_1	λ_2
n					$p = 0.2$				
50	2.91	0.82	6.52	2.93	1.10	9.46	2.37	1.27	5.07
100	2.54	1.01	6.33	2.21	1.46	4.90	1.84	1.54	3.22
250	2.29	1.27	4.47	2.39	1.89	4.28	1.59	1.49	2.09
500	1.60	1.12	2.00	1.99	1.75	2.63	1.41	1.35	1.64
					$p = 0.5$				
50	1.07	0.84	3.45	1.16	1.00	4.37	1.17	1.24	3.20
100	1.03	0.91	3.22	1.47	1.34	3.80	1.41	1.54	2.52
250	1.19	1.04	2.06	1.68	1.66	2.35	1.45	1.52	1.74
500	1.25	1.19	1.95	1.62	1.62	1.89	1.30	1.35	1.44
					$p = 0.8$				
50	0.64	0.67	1.98	0.40	0.55	2.59	0.38	0.83	2.04
100	0.51	0.56	2.75	0.43	0.76	2.50	0.48	1.12	1.81
250	0.57	0.79	2.55	0.84	1.32	2.02	0.86	1.30	1.44
500	0.78	1.08	2.54	1.22	1.41	1.59	1.02	1.24	1.25

examination of the results reveals some points that might give some explanation for this fact. For some samples with large p_1 the other components are represented in the sample with a few observations not far from the origin. As a result, the other components were confounded with the first component yielding thus an MHD estimate of p_1 that was close to 1. An indication supporting this observation is the increased relative MSEs when the sample sizes were increased.

Concluding, we can say that the MHD method for finite Poisson mixtures is appealing compared to ML method with respect to both efficiency and robustness.

7. Discussion.

The MHD method for finite Poisson mixtures seems to be a very attractive method of estimation, because it can optimize the estimates with respect to both efficiency and robustness. The attainment of robust estimates with high efficiency is very important in mixture models because of the failure of the standard likelihood approach on the one hand, and of the fact that mixture models are very often not appropriately specified on the other. The latter include cases where the number of components cannot be assigned prior to the analysis, and thus the ML method with an incorrect number of components can cause some problems. The MHD estimation method seems to be more reliable in such situations.

In this paper the robustness of the MHD method was examined under contamination models. Examining the performance of the technique under models with at least one component distribution being not Poisson is another question of interest that needs to be and is currently being investigated by the present authors

Another very interesting aspect is that we may use similar devices to the well-known likelihood ratio test for hypothesis testing. The derivation and use of a Hellinger deviance test (see Simpson, 1989) for detecting the presence of a mixture is still an ongoing research by the authors.

The application of MHD estimates for clustering purposes seems also appealing. Mixture models are widely used in Cluster Analysis, for more details the reader is referred to the book by McLachlan and Basford (1989). Symons et al. (1983) used ML estimates for 2-finite Poisson mixtures in a clustering application. The robustness of the MHD method relative to the ML method makes this method very appealing for clustering applications.

Acknowledgements

The authors wish to thank Prof. Paul Eiler who kindly provided us with the data set used in the example. The authors would also like to thank the referees and the associate editor for their comments that helped improve the presentation of the results. Dimitris Karlis is being supported with a scholarship from the State Scholarships Foundation of Greece.

References

Aitkin, M., Wilson, T, 1980. Mixture models, outliers and the EM algorithm. Technometrics 22, 325–331.

Bartlett, M., Mc Donald, P., 1968. Least squares estimation of distribution mixtures. Nature E.S.I. London 217, 195–196.

Behboodian, J., 1969. On a mixture of normal distributions. Biometrika 56, 215–217.

Beran, R.J., 1977. Minimum Hellinger distance estimates for parametric models. Ann. Statist. 5, 445–463.

Blum, J.R., Susarla, V., 1977. Estimation of a mixing distribution function. Ann. Probab. 5, 200–209.

Choi, K., Bulgren, W., 1968. An estimation procedure for mixtures of distributions. J. Roy. Statist. Soc., B 30, 444–460.

Clarke, B. R., Heathcote, C.R., 1994. Robust estimation of k-components univariate mixtures. Ann. Inst. Statist. Math. 46, 83–93.

Cutler, A., Cordero-Brana, O., 1996. Minimum Hellinger distance estimation for finite mixture models. J. Amer. Statist. Assoc. 91, 1716–1724.

Deely, J., Kruse, R., 1968. Construction of sequences estimating the mixing distribution. Ann. Math. Statist., 39, 286–288.

Eddelman, D., 1988. Estimation of the mixing distribution for a normal mean with applications to the compound decision problem. Ann. Statist. 16, 1609–1622.

Fryer, J. G., Robertson, C.A., 1972. A comparison of some methods for estimating mixed normal distributions. Biometrika 59, 639–648.

Hampel, F., Ronchetti, E., Rousseeuw, P., Stahel, W., 1986. Robust Statistics. Wiley, New York.

Harris, I., Basu, A., 1994. Hellinger distance as a penalised loglikelihood.Comm. Statis. Simulation Computat. 23, 1097–1113.

Hasselblad, V., 1969. Estimation of finite mixtures from the exponential family. J. Amer. Statist. Assoc. 64, 1459–1471.

Henna, J., 1983. A note on a consistent estimator of a mixing distribution function. Ann. Inst. Statist. Math. 35, 229–234.

Jorgensen M., 1990. Influence-based diagnostics for finite mixture models. Biometrics 46, 1047–1058.

Lindsay, B., 1994. Efficiency versus robustness: the case of minimum Hellinger distance and related methods. Ann. Statist. 22, 1081–1114.

Mac Donald, P.D.M., 1971. Comment on "An estimation procedure for mixtures of distributions" by Choi and Bulgren. J. Roy. Statist. Soc. B 33, 326–329.

McLachlan, G., Basford, K., 1988. Mixture Models: Inference and Applications to Clustering. Marcel Dekker, New York.

Phillips, R., 1990. On constructing sequences estimating the mixing distribution with applications. Comm. Statist. Simulation Comput. 19, 705–720.

Redner, R., Walker, H., 1984. Mixture densities, maximum likelihood and the EM algorithm. SIAM Rev. 26, 195–230.

Simpson, D., 1987. Minimum Hellinger distance estimation for the analysis of count data. J. Amer. Statist. Assoc. 82, 802–807.

Simpson, D., 1989. Hellinger deviance tests: efficiency, breakdown points and examples. J. Amer. Statist. Assoc. 84, 107–113.

Symons, M., Grimson, R., Yuan, Y., 1983. Clustering of rare events. Biometrics 39, 193–205.

Teicher, H., 1961. Identifiability of mixtures. Ann. Math. Statist., 32, 244–248.

Titterington, D. M, Smith, A.F.M., Makov, U.E., 1985. Statist. Analysis of Finite Mixtures Distributions. Willey, New York.

Woodward, W., Parr, W., Schucany, R., Lindsey, H., 1984. A comparison of minimum distance and maximum likelihood estimation of a mixture proportion. J. Amer. Statist. Assoc. 79, 590–598.

Woodward, W., Whitney, P., Eslinger, P., 1995. Minimum Hellinger distance estimation of mixture proportions. J. Statist. Plann. Inference 48, 303–319.