Ann. Inst. Statist. Math.
Vol. 51, No. 1, 149–162 (1999)

ON TESTING FOR THE NUMBER OF COMPONENTS IN A MIXED POISSON MODEL

DIMITRIS KARLIS AND EVDOKIA XEKALAKI

Department of Statistics, Athens University of Economics and Business,
76 Patission St., 10434, Athens, Greece

(Received May 22, 1996; revised November 10, 1997)

Abstract. Poisson mixtures are usually used to describe overdispersed data. Finite Poisson mixtures are used in many practical situations where often it is of interest to determine the number of components in the mixture. Identifying how many components comprise a mixture remains a difficult problem. The likelihood ratio test (LRT) is a general statistical procedure to use. Unfortunately, a number of specific problems arise and the classical theory fails to hold. In this paper a new procedure is proposed that is based on testing whether a new component can be added to a finite Poisson mixture which eventually leads to the number of components in the mixture. It is a sequential testing procedure based on the well known LRT that utilises a resampling technique to construct the distribution of the test statistic. The application of the procedure to real data reveals some interesting features of the distribution of the test statistic.

Key words and phrases: Poisson mixture, k-finite mixture, number of components, likelihood ratio test, resampling.

1. Introduction

It is widely accepted that the Poisson distribution can describe adequately situations where only randomness is present. In a variety of applications, however, the data manifest some sort of overdispersion in the sense of having a variance that is larger than their mean. Such situations can be modelled by a mixed Poisson distribution. We will restrict our attention to finite mixtures, assuming that the parameter of the Poisson distribution takes only a finite number of distinct values.

In the sequel, we use the term k-finite mixture of Poisson distributions to refer to the distribution defined by the probability function

$$f_{\theta_k}(x) = \sum_{i=1}^{k} p_i \frac{\exp(-\lambda_i)\lambda_i^x}{x!}, \qquad x = 0, 1, 2, \dots \tag{1.1}$$

where k is the number of Poisson components, $\lambda_i \geq 0$, $\forall i = 1, 2, \dots, k$, p_i, $i = 1, \dots, k$, are the mixing proportions with $\sum_{i=1}^{k} p_i = 1$ and $\boldsymbol{\theta}_k$ represents the vector of parameters of the above mixture, namely $\boldsymbol{\theta}_k = (p_1, p_2, \ldots, p_{k-1}, \lambda_1, \lambda_2, \ldots, \lambda_k)$. We also assume that $0 \leq \lambda_1 < \lambda_2 < \cdots < \lambda_k$ so as to avoid identifiability problems (see Teicher (1961)).

Such a distribution may arise when the entire population can be thought of as consisting of k subpopulations, each following a Poisson distribution with some parameter. The value of the Poisson parameter is assumed to vary from subpopulation to subpopulation according to a distribution function P with jumps of size $p_j > 0$, at the points λ_j, $j = 1, 2, \ldots, k$. P is usually referred to as the mixing distribution. Obviously, the number of support points can give us information about the number of subpopulations comprising a finite mixture of Poisson distributions. It would, therefore, be interesting to determine this number.

A common method of testing for the number of components in a model is the likelihood ratio test (LRT). Unfortunately, the general theory of the test fails for mixture models as noted by several authors (e.g. Self and Liang (1987)). All the test procedures suggested by various research workers have been restricted to the simplest case of testing for a one-component model versus a 2-component model. The test procedure that is proposed in this paper generalises such tests so as to test for a model with k components against a model with $k+1$ components.

Following a brief review of results on the use of the likelihood ratio procedure for 2-finite mixtures provided in Section 2, a new method for determining the number of components in a mixture is introduced in Section 3. This is based on a sequential approach and can serve both as a method for general testing for k components against $k+1$ components in a finite Poisson mixture as well as a method for determining the optimal number of components. In Section 4, the performance of the new method is examined via simulation. The new procedure is illustrated on a real dataset in Section 5. Our findings are summarised in Section 6.

2. The likelihood ratio test for mixture models

The LRT is used for testing nested hypotheses. The test statistic is $-2\log\lambda$, where $\lambda = L_0/L_1$ is the ratio of the maximised likelihood L_0 under the model in H_0 to the maximised likelihood L_1 under the model in H_1. Under some regularity conditions, this statistic follows asymptotically a χ^2 distribution with a number of degrees of freedom equal to the difference in the numbers of parameters between the two models.

Suppose that we want to test the hypothesis H_0: The data come from a Poisson distribution, against the hypothesis H_1: The data come from a 2-finite mixture of Poisson distributions. From (1.1), for $k = 2$, we can see that we may rewrite the hypotheses to be tested as H_0: $p_1 = 1$ or $p_1 = 0$, against H_1: $0 < p_1 < 1$. A common testing procedure for such hypotheses would employ an LRT statistic. However, carrying out this test for mixture models presents some difficulties. The reason for this is that the value of p_1 under the null hypothesis is on the boundary of the parameter space and hence the regularity conditions fail. (See for example Self and Liang (1987).)

Many attempts have been made in the literature to determine the asymptotic distribution of the test statistic in the general framework of finite mixtures. A few of them were focused on Poisson mixtures (e.g. Symons *et al.* (1983) and Bohning *et al.* (1994)), while the majority of them were in relation to the case of normal mixtures. Titterington *et al.* (1985), Self and Liang (1987) and Bohning *et al.* (1994) showed that the asymptotic distribution of the test statistic is a mixture of a distribution degenerate at 0 and a χ^2 distribution with one degree of freedom in equal mixing proportions. Alternatively, the distribution of the test statistic can be derived via simulation as in Aitkin *et al.* (1981), Symons *et al.* (1983), McLachlan (1987), Thode *et al.* (1988), Mendell *et al.* (1991, 1993), and Feng and McCullogh (1994, 1996) among others. See also the work of Beran (1988) for a thorough justification of bootstrap tests. Feng and McCullogh (1996) showed that the loglikelihood is identifiable even when the parameters are not and hence the fact that the ML estimates are not consistent under the alternative hypothesis does not affect the procedure.

3. Determining the optimal number of components—The new approach

The unknown form of the null distribution causes problems in the application of the LRT. The test procedure that we propose in this section aims at overcoming these problems. The procedure makes a sequential use of the LRT adopting the bootstrap approach for constructing the null distribution of the test statistic at every stage. The proposed test procedure constitutes the first attempt to use bootstrap tests for such kinds of hypothesis testing. It can also serve as a method of determining the number of components in a mixture. The utility of this possibility is obvious, as it gives us an insight into the structure of the population under investigation.

Consider the hypothesis H_0: the number of components in a Poisson mixture is k versus the hypothesis H_1: the number of components in the mixture is $k+1$. The proposed procedure tests H_0 against H_1 sequentially for $k = 1, 2, \ldots$ using the LRT statistic until H_0 is accepted for the first time at the chosen significance level. The value $k_{\max}$ of k which does not lead to the rejection of H_0 represents the optimal number of components in the mixture. Due to the fact that the standard asymptotic result is not applicable, we adopt a resampling approach for the construction of the null distribution of the LRT statistic. The steps for carrying out the proposed test are:

Set $k = 1$

Step 1. Find the ML estimates of the parameters of the finite Poisson mixture for k and $k+1$, say $\boldsymbol{\theta}_k$ and $\boldsymbol{\theta}_{k+1}$ respectively and calculate the LRT statistic, say L_{obs}. Note that for $k = 1$ the MLE is the sample mean.

Step 2. Simulate B bootstrap samples of size n, (n is the sample size from the data set) from the k-finite Poisson mixture with parameter vector $\boldsymbol{\theta}_k$, and for each bootstrap sample calculate the value of the LRT, say w_j, $j = 1, \ldots, B$.

Step 3. Estimate the a-percentile of the distribution of the test statistic by the $(100a)$-th order statistic from the bootstrap values w_j, $j = 1, \ldots, B$. Let this percentile be denoted by C_a.

Step 4. **If** $L_{\text{obs}} > C_a$ **then** set $k = k + 1$ and go to Step 1, **else** deduce that the optimal number of components is k and stop.

The procedure terminates when H_0 cannot be rejected for the first time, i.e. when there is no sufficient evidence that adding one more component will significantly improve the likelihood.

As will be illustrated in Section 5, this procedure achieves a dual goal: it reveals the number of components in the assumed mixture model while at the same time it provides the appropriate goodness-of-fit test. Approaches alternative to the one proposed for determining the number of components in a mixture that already exist in the literature have been based on penalized estimation methods (Henna (1985), Leroux and Putterman (1992), Leroux (1992) and Chen and Kaldbfleisch (1996)), moment based methods (Lindsay (1989) and Fruman and Lindsay (1994)), graphical method (Lindsay and Roeder (1992)), Bayesian methods (Richardson and Green (1997)). Wolfe (1970) proposed the use of a chi-square distribution as an asymptotic approximation to the distribution of the LRT statistic. Izenmann and Sommer (1988) adopted this method. Other methods include the SEM algorithm described by Celeux and Diebolt (1985), the method based on Fisher information matrices proposed by Windham and Culter (1992) and the posterior Bayes factors method used by Aitkin *et al.* (1996).

The proposed procedure is based on a forward search technique aiming mainly at reducing the computational effort. In their concluding remarks Bohning *et al.* (1994) proposed a backward search type. In fact, we expect that both backward and forward elimination techniques will produce the same results. The reason is that the improvement of the likelihood between two successive models decreases as the number of the already fitted components increases (Lindsay (1983)). The results of our simulation show that the critical values decrease also, thought slower than the observed value of the LRT statistic. Based on this, one would expect that if the k-component model hypothesis is not rejected when tested versus the $(k+1)$-component model hypothesis, then neither will the $(k+1)$-component hypothesis when tested versus a $(k+2)$-component hypothesis and so on. So, we expect that a backward technique will produce almost the same results as the forward technique. On the other hand, for small values of k, the ML estimates are derived easier and faster. Hence, the forward technique can save a lot of computational time. Note also that with a backward technique the starting value of k is not known. Thus, we have either to choose it arbitrarily or to determine it using sophisticated and computationally demanding methods (for such methods see Bohning (1995)).

4. Simulation results

In this section an extensive simulation study of the newly proposed method is made. Two issues are of special interest. The first is the ability of the procedure to determine the correct number of components and the second is the power of the sequential tests used for obtaining the optimal number of components.

In order to use the LRT we need the ML estimates of the mixing distribution under each of the two hypotheses. The EM algorithm for the ML estimation in finite mixture models was used (e.g. Hasselblad (1969), see also Dempster *et* *al.* (1977)). The details can be found in the Appendix. We also employed the conditions introduced by Karlis and Xekalaki (1996*a*) for checking if the LRT statistic is equal to 0.

In the sequel, some k-finite Poisson mixtures were considered for selected values of k ($k = 2, 3, 4$) so as to allow the representation of models with well separated components, models with components close together and models that result in skew distributions. For each distribution three sample sizes were used ($n = 50, 100, 500$) and 500 samples were generated from each distribution, for each sample size. The sequential method proposed was then applied using 500 bootstrap samples ($B = 500$) for constructing the null distribution of each test statistic. For all the tests we used $\alpha = 5\%$. Table 1 presents the relative frequencies of the numbers of components which the new method detected.

Table 1 reveals that the method is quite successful in determining the number of components when the latter are not close and the sample size is large enough. How large the sample size must be depends on the value of k. So, for $k = 3$ a sample size of 500 is sufficient for the accurate determination of the number of components. For $k = 4$ larger sample sizes are needed. Clearly, when the components are very close the method cannot distinguish between them. On the other hand, components with small mixing probabilities are usually ignored especially in the case of small sample sizes. It is interesting that the method seldom overestimates the number of components. For small sample sizes it performs better for models with small numbers of components. This may be connected with the high variances of the ML estimates for finite Poisson mixtures with not well separated components and with a small sample size first reported by Hasselblad (1969). For our simulation purposes, only cases plausible in practice and small sample sizes were considered. Selecting cases with extraordinarily large separation between the components would only lead to more impressive results but of little practical interest as most often count data consist of small positive integer values.

The sequential nature of the tests employed makes the calculation of the power of the method (in the sense of the term used in hypothesis testing) very difficult. The simulation results reported in Table 1, however, can also be regarded as revealing the power of the proposed method.

The power of each separate test proposed is also examined. Thode *et al.* (1988) and Mendell *et al.* (1991, 1993) have examined the power of the LRT for testing a one-component model versus a 2-component model in normal mixtures with equal variances via simulation. Recently, Berdai and Garrel (1996) examined the power of the LRT deriving an asymptotic distribution. All the authors agree in that the power of the test is susceptible to the sample size and to the closeness of the components. The power for the LRT in the case of finite Poisson mixtures has not been examined, up to now. So the results given in this section on the power of such tests are new and they are reported for the first time in the present paper.

In order to investigate the power of the proposed method the empirical power of the test for k components versus $k+1$ components was examined for $k = 1, 2, 3$. We define as the empirical power of the test the proportion of times we rejected the null hypothesis when the data actually were generated from the alternative

Table 1. The relative frequencies of the estimated number of components among 500 simulated samples from k-finite Poisson mixtures ($\alpha = 5\%$).

sample size	$n = 50$					$n = 100$					$n = 500$				
						$k = 2$									
						estimated number of components									
parameter vector θ_2	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
$(0.5, 1, 9)$	—	0.95	0.05	—	—	—	0.95	0.05	—	—	—	0.96	0.04	—	—
$(0.8, 1, 9)$	—	0.92	0.08	—	—	—	0.95	0.05	—	—	—	0.96	0.04	—	—
$(0.5, 1, 1.1)$	0.96	0.04	—	—	—	0.93	0.07	—	—	—	0.94	0.05	0.01	—	—
$(0.95, 1, 10)$	0.11	0.83	0.06	—	—	—	0.93	0.07	—	—	—	0.95	0.05	—	—
						$k = 3$									
						estimated number of components									
parameter vector θ_3	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
$(0.45, 0.45, 1, 5, 10)$	—	0.62	0.36	0.01	—	—	0.39	0.58	0.02	—	—	—	0.94	0.06	—
$(0.4, 0.4, 1, 3, 3.1)$	0.42	0.56	0.01	—	—	0.14	0.82	0.03	—	—	—	0.96	0.04	—	—
$(0.33, 0.33, 1, 5, 10)$	—	0.54	0.44	0.01	—	—	0.30	0.66	0.03	—	—	—	0.94	0.06	—
						$k = 4$									
						estimated number of components									
parameter vector θ_4	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
$(0.3, 0.4, 0.25, 1, 5, 9, 15)$	—	0.31	0.61	0.08	—	—	0.09	0.78	0.13	—	—	—	0.59	0.38	0.03
$(0.3, 0.3, 0.2, 1, 1.2, 5, 9)$	—	0.78	0.21	0.01	—	—	0.68	0.31	0.01	—	—	0.17	0.78	0.03	0.02
$(0.25, 0.25, 0.25, 1, 5, 10, 15)$	—	0.17	0.76	0.07	—	—	0.02	0.86	0.12	—	—	—	0.59	0.40	0.01

Table 2a. The empirical power of the LRT for testing $k = 1$ versus $k = 2$ ($\alpha = 5\%$).

Null distribution		sample size $n = 50$	$n = 100$	$n = 500$	$n = 1000$	$n = 2000$
	alternative					
$\lambda = 1$	1A	0.063	0.068	0.064	0.050	0.048
	1B	0.340	0.526	0.974	0.999	1.000
	1C	0.225	0.327	0.753	0.936	0.997
	1D	0.081	0.089	0.087	0.080	0.095
$\lambda = 3$	1A	0.041	0.048	0.039	0.031	0.035
	1B	0.868	0.989	1.000	1.000	1.000
	1C	0.538	0.793	1.000	1.000	1.000
	1D	0.061	0.070	0.094	0.100	0.147
$\lambda = 5$	1A	0.033	0.039	0.035	0.027	0.034
	1B	0.992	1.000	1.000	1.000	1.000
	1C	0.802	0.972	1.000	1.000	1.000
	1D	0.062	0.074	0.132	0.171	0.294
$\lambda = 10$	1A	0.027	0.034	0.042	0.035	0.060
	1B	1.000	1.000	1.000	1.000	1.000
	1C	0.988	1.000	1.000	1.000	1.000
	1D	0.073	0.107	0.320	0.511	0.813

distribution. The level of significance was $\alpha = 5\%$. The critical value of each test was calculated via simulation of 50000 samples of given size from the null distribution. For each value of k several distributions were chosen so as to represent various cases. For every distribution considered, several alternatives were considered. As Karlis and Xekalaki (1996*b*) have shown, the alternative distributions have to be chosen so that they have the same mean with the distribution under the null hypothesis. The reason is that when applying the test to real data sets the MLE for k-finite mixtures must satisfy the first moment equation whatever the value of k.

For the case where $k = 1$, the null distributions used were Poisson distributions with parameters $\lambda = 1, 3, 5, 10$ respectively. Then the vectors of parameters for the 2-finite mixture alternatives were: (1A) $(0.5, 0.95\lambda, 1.05\lambda)$, (1B) $(0.5, 0.5\lambda, 1.5\lambda)$, (1C) $(0.8, 0.8\lambda, 1.8\lambda)$ and (1D) $(0.2, 0.8\lambda, 1.05\lambda)$. Alternative (1A) is very close to the null distribution while (1C) and (1D) result in distributions more skew to the left and to the right respectively.

For $k = 2$ the distributions considered in the null hypothesis had vectors of parameters of the form $(p_1, \lambda_1, \lambda_2)$: (2a) $(0.5, 1, 5)$, (2b) $(0.8, 3, 11)$, (2c) $(0.5, 2, 2.2)$, (2d) $(0.5, 5, 15)$. The four alternatives to each null hypothesis considered were of the form: (2A) $(0.5p_1, 0.5p_1, 0.95\lambda_1, 1.05\lambda_1, \lambda_2)$, (2B) $(0.5p_1, 0.5p_1, 0.5\lambda_1, 1.5\lambda_1,$

Table 2b. The empirical power of the LRT for testing $k = 2$ versus $k = 3$ ($\alpha = 5\%$).

		sample size				
Null distribution		$n = 50$	$n = 100$	$n = 500$	$n = 1000$	$n = 2000$
	alternative					
	2A	0.046	0.059	0.065	0.052	0.044
	2B	0.052	0.090	0.222	0.314	0.514
2a	2C	0.027	0.035	0.026	0.006	0.000
	2D	0.066	0.080	0.078	0.063	0.049
	2E	0.104	0.197	0.544	0.764	0.949
	2A	0.049	0.067	0.051	0.039	0.040
	2B	0.393	0.697	0.999	1.000	1.000
2b	2C	0.060	0.074	0.084	0.080	0.075
	2D	0.108	0.123	0.166	0.174	0.190
	2E	0.152	0.192	0.347	0.502	0.755
	2A	0.034	0.038	0.077	0.085	0.077
	2B	0.097	0.119	0.190	0.287	0.460
2c	2C	0.002	0.001	0.000	0.000	0.000
	2D	0.011	0.011	0.003	0.002	0.002
	2E	0.012	0.020	0.045	0.033	0.018
	2A	0.051	0.065	0.050	0.040	0.037
	2B	0.422	0.723	1.000	1.000	1.000
2d	2C	0.048	0.056	0.061	0.044	0.044
	2D	0.084	0.100	0.104	0.113	0.119
	2E	0.105	0.130	0.198	0.272	0.447

λ_2), (2C) $(\pi p_1, \pi, 0.95\lambda_1, (\lambda_1 + \lambda_2)/2, \lambda_2)$, (2D) $(\pi p_1, \pi p_2, 0.95\lambda_1, \lambda_2, 1.5\lambda_2)$ and (2E) $(0.33, 0.33, 1, 1 + a, \lambda_2 + 1)$, where π is the probability assigned to the third component so that the mean does not change and a is chosen so that the alternative distribution can have the same mean as the null distribution. Again (2A) differs very little from the null distribution, (2B) differs more, while (2C) and (2D) add the new component in the left and the right tail respectively.

Similarly, for $k = 3$ the distributions used under the null hypothesis had vectors of parameters of the form $(p_1, p_2, \lambda_1, \lambda_2, \lambda_3)$: (3a) $(0.33, 0.33, 1, 5, 12)$, (3b) $(0.8, 0.1, 1, 5, 12)$, (3c) $(0.1, 0.4, 1, 5, 12)$, (3d) $(0.5, 0.25, 1, 8, 8.5)$ and (3e) $(0.33, 0.33, 1, 10, 20)$. The alternatives to each of them considered were of the form: (3A) $(0.5p_1, 0.5p_1, p_2, 0.95\lambda_1, 1.05\lambda_1, \lambda_2, \lambda_3)$, (3B) $(0.5p_1, 0.5p_1, p_2, 0.5\lambda_1, 1.5\lambda_1, \lambda_2, \lambda_3)$, (3C) $(p_1, p_2, 0.5p_3, \lambda_1, \lambda_2, 0.5\lambda_3, 1.5\lambda_3)$, (3D) $(\pi p_1, \pi p_2, \pi p_3, \lambda_1, \lambda_2, \lambda_3, 1.5\lambda_3)$ and (3E) $(0.25, 0.25, 0.25, 1, 1{+}a, 1{+}2a, \lambda_3{+}1)$, where a and π are defined as previously. Again (3A) differs very little from the null distribution, (3B) differs more, while (3C) and (3D) add the new component between the 2nd and the 3rd component

Table 2c. The empirical power of the LRT for testing $k = 3$ versus $k = 4$ ($\alpha = 5\%$).

Null distribution		sample size				
		$n = 50$	$n = 100$	$n = 500$	$n = 1000$	$n = 2000$
	alternative					
	3A	0.032	0.068	0.097	0.085	0.047
	3B	0.025	0.064	0.183	0.232	0.355
3a	3C	0.015	0.030	0.058	0.053	0.032
	3D	0.092	0.180	0.551	0.764	0.941
	3E	0.029	0.068	0.225	0.302	0.462
	3A	0.050	0.067	0.112	0.121	0.109
	3B	0.063	0.123	0.424	0.632	0.831
3b	3C	0.013	0.027	0.104	0.126	0.126
	3D	0.050	0.123	0.528	0.766	0.937
	3E	—	—	—	—	—
	3A	0.021	0.050	0.107	0.082	0.036
	3B	0.017	0.040	0.114	0.138	0.142
3c	3C	0.016	0.031	0.008	0.001	0.000
	3D	0.032	0.066	0.092	0.087	0.115
	3E	0.025	0.043	0.081	0.070	0.063
	3A	0.049	0.060	0.071	0.063	0.074
	3B	0.080	0.125	0.368	0.575	0.808
3d	3C	0.001	0.003	0.001	0.000	0.000
	3D	0.008	0.017	0.064	0.140	0.303
	3E	0.041	0.058	0.113	0.122	0.129
	3A	0.046	0.057	0.064	0.047	0.041
	3B	0.082	0.152	0.431	0.644	0.893
3e	3C	0.036	0.045	0.051	0.036	0.013
	3D	0.146	0.252	0.297	0.537	0.761
	3E	0.244	0.471	0.962	0.999	1.000

and at the right tail respectively. Tables 2a–2c contain the values of the empirical power for all the cases.

For testing a one component mixture versus a two component mixture the power of the test increases with the distance of the components. This result is similar to the one obtained for normal mixtures by Mendell *et al.* (1991) and it was expected. For testing $k = 2$ versus $k = 3$ the power is increased only when the sample size is large and the components are well separated. Adding a well separated new component, but with a small probability, does not improve the power of the test. This is also the case when testing for $k = 3$ versus $k = 4$. Concluding, we can say that the LRT applied to the general case $k = m$ versus

$k = m + 1$ has low power when the components are not well separated and when one of the components has a small mixing probability. As the value of m increases, the sample size required for obtaining a specific power increases very much. This result verifies the behaviour of the method for the simulated cases of Table 1. As far as the asymptotic distribution of the test statistic is concerned, the χ^2 form does not seem plausible. On the other hand, the null distribution depends highly on the value of k and the sample size used.

5. An application

In the present section the proposed procedure for determining the number of components in the case of a finite Poisson mixture is illustrated by a real dataset example.

Table 3. Number of accidents incurred by 414 machinists over a period of three months (Greenwood and Yule (1920)).

x	0	1	2	3	4	5	6	7	8
frequency	296	74	26	8	4	4	1	0	1

The data refer to the number of accidents incurred by 414 machinists over a period of three months, taken from the classical paper of Greenwood and Yule (1920), and analysed by several authors. The fit provided by the simple Poisson distribution is very poor ($X^2 = 57.81$ with 2 d.f) a fact noted also by Greenwood and Yule. As can be seen from Table 4, a notable improvement can be achieved by mixtures of Poisson models.

Table 4. Sequential testing results for the data of Table 3.

k	LRT statistic	p-value
1	88.068	0
2	3.122	0.033
3	0.094	0.216

Column 1 of Table 4 contains the values of k, the number of the components in the mixture. Column 2 contains the values of the test statistic for testing $m = k$ against $m = k + 1$ and the last column contains the associated p-values calculated via simulation. Using the bootstrap approach described previously, we constructed the null distribution of the test statistic for various values of k, using 50000 bootstrap samples. Our procedure leads to the selection of the model with

3 components. Note, however that, had we erroneously used the χ^2 approximation we would have been led to select the 2-component model.

Figure 1 depicts the cumulative distribution function of the test statistic under the hypotheses tested. Clearly, the distributions differ markedly from the χ^2 distribution with 2 degrees of freedom and there is a difference between the distributions corresponding to different values of k in the null hypothesis. We may deduce therefore that the use of the χ^2 can lead to invalid conclusions and hence it should be avoided. Bohning *et al.* (1994) have come to the same conclusion for a variety of models in the case $k = 1$.

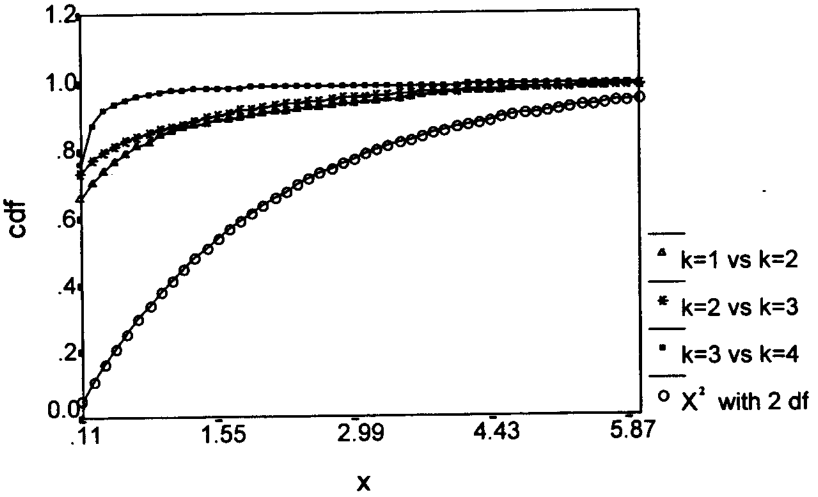

Fig. 1. The cumulative distribution function (cdf) for the test statistic for testing H_0 : $k = 1$ vs H_1 : $k = 2$, H_0 : $k = 2$ vs H_1 : $k = 3$ and H_0 : $k = 3$ vs H_1 : $k = 4$ for the data of Table 3 and the cdf of a χ^2 distribution with 2 degrees of freedom. The form of the distribution clearly depends on the value of k.

In order to assess the performance of the newly proposed method we calculated the empirical power of the test procedure involved. As mentioned before, this is defined as the proportion of times we rejected the null hypothesis when the data actually came from the alternative distribution. So, for our example, we used as critical values for given α the corresponding α-percentiles of the null distribution constructed via simulation. 50000 samples were generated from the distribution in H_1, namely the distribution with parameters the ML estimates for a model with the number of components specified in H_1. For each sample, the LRT statistic was calculated so as to construct the distribution of the test statistic under the alternative hypothesis. The proportions of times H_0 was rejected for a given level of significance α are reported in Table 5. As can easily be seen, the LRT performs well only for the case $k = 1$ vs $k = 2$. It can be noted that the test has a lower performance when it is used to detect components that are very close. This is usually true for models with a large number of components as the new added

Table 5. Simulated power calculation for the data of Table 3 (α denotes the significance level).

	α			
Test	0.10	0.05	0.025	0.01
$k = 1$ vs $k = 2$	1.000	1.000	1.000	1.000
$k = 2$ vs $k = 3$	0.560	0.430	0.318	0.207
$k = 3$ vs $k = 4$	0.055	0.023	0.008	0.003

point is usually very close to the previously estimated points. Note that the null distribution of all the test statistics is highly skewed to the right.

Based on the above results on both real and simulated data, the method presented in this paper does not seem to overestimate the number of components in the mixture. This is the consequence of the fact that in a model with too many components, two or more components are very close together, and thus the improvement of the loglikelihood is negligible.

6. Discussion

In the present paper a new technique for testing hypotheses on finite Poisson mixtures was introduced. Testing such hypotheses proceeds in a sequential manner which lends particularity to the procedure and distinguishes it from just being another technique for merely testing Poisson mixture hypotheses with fixed numbers of components. In parallel to that, it allows for determining the optimal number of components for which a Poisson mixture provides the most satisfactory fit to the data, without an excess of support points. Thus, the innovation brought by this procedure lies in its dual character that permits testing a hypothesis of a Poisson mixture with k components against one model with $k + 1$ components, parallel to determining the optimal Poisson mixture. The implications of this technique in cluster analysis and other fields of application are obvious.

The application of the procedure to both real and simulated data led to the interesting conclusions that the asymptotic distribution of the test statistic has a form that can in no way be approximated by the χ^2 distribution which is the standard choice in problems of this type and that the value of k in H_0 affects the distribution of the test statistic. The latter needs further investigation as it differs from results already existing in the literature that fail to examine how the value of k affects the procedure. The examination of the power of the procedure reveals that for well separated components the proposed testing procedure may work well, otherwise the LRT has very low power, and a more powerful test should be sought.

Finally, it is clear that our method which we applied to Poisson mixtures is applicable to normal mixtures or to mixtures from the exponential distribution as well and in general can be considered as a method for determining the number of components in finite mixtures from any family of distributions.

Acknowledgements

The authors would like to thank Prof. Dankmar Bohning for helpful comments on an earlier version of this work. The authors would also like to thank the referees for their comments that helped improve the presentation of the results. Dimitris Karlis is being supported by a scholarship from the State Scholarships Foundation of Greece.

Appendix. About the simulations

All the calculations were performed using the EM algorithm. As a consequence, the stopping rule may lead to slightly different values of the LRT. The stopping rule considered for all the simulations in the present paper stopped iterating when the absolute value of the ratio $(L^{(m-1)}/L^{(m)})$ first exceeded 0.9999995. ($L^{(m)}$ denotes the maximised loglikelihood after m iterations.) The same criterion was used for obtaining the ML estimates from the data.

The convergence of the EM depends on the choice of the initial values. Three different initial guesses were used for finding the ML estimates of a k-component model: a) The MLE for a k-component model calculated from the data b) Equally spaced points between the minimum and the maximum observed value with equal initial probabilities and c) Equally spaced points around the mean of the sample. For even k the mean itself was used as a starting point. Again the initial probabilities were assumed equal.

The algorithm run for all the different initial values and the "best" solution was selected. A good strategy for avoiding local maxima is to start from well separated values. We note that the computational error in the sense of not finding the global maximum might have had a slight downward bias effect on the percentiles. The acceleration scheme described in Karlis and Xekalaki (1996*b*) was used.

References

Aitkin, M., Anderson, D. and Hinde, J. (1981). Statistical modelling of data on teaching styles, *J. Roy. Statist. Soc. Ser. A*, **144**, 419–461.

Aitkin, M., Finch, S., Mendell, N. and Thode, H. (1996). A new test for the presence of a normal mixture distribution based on the posterior Bayes factor, *Statistics and Computing*, **6**, 121–125.

Beran, R. (1988). Prepivoting test statistics: a bootstrap review of asymptotic refinements, *J. Amer. Statist. Assoc.*, **83**, 687–697.

Berdai, A. and Garrel, B. (1996). Detecting a univariate normal mixture with two components, *Statist. Decisions*, **14**, 35–51.

Bohning, D. (1995). A review of reliable maximum likelihood algorithms for semiparametric mixture models, *J. Statist. Plann. Inference*, **47**, 5–28.

Bohning, D., Dietz, Ek., Schaub, R., Schlattman, P. and Lindsay, B. (1994). The distribution of the likelihood ratio for mixtures of densities from the one-parameter exponential family, *Ann. Inst. Statist. Math.*, **46**, 373–388.

Celeux, G. and Diebolt, J. (1985). The SEM algorithm: a probabilistic teacher algorithm derived from the EM algorithm for the mixture problem, *Computational Statistics Quarterly*, **2**, 73–92.

Chen, J. and Kalbfleisch, J. D. (1996). Penalised minimum-distance estimates in finite mixture models, *Canad. J. Statist.*, **24**, 167–175.

Dempster, A. P., Laird N. M. and Rubin, D. (1977). Maximum likelihood from incomplete data via the EM aglgorithm, *J. Roy. Statist. Soc. Ser. B*, **39**, 1–38.

Feng, Z. and McCulloch, C. E. (1994). On the likelihood ratio test statistic for the number of components in a normal mixture with unequal variances, *Biometrics*, **50**, 1158–1162.

Feng, Z. and McCulloch, C. E. (1996). Using bootstrap likelihood ratios in finite mixture models, *J. Roy. Statist. Soc. Ser. B*, **58**, 609–617.

Fruman, W. D. and Lindsay, B. (1994). Testing for the number of components in a mixture of normal distributions using moment estimators, *Comput. Statist. Data Anal.*, **17**, 473–492.

Greenwood, M. and Yule, G. (1920). An inquiry into the nature of frequency distributions representative of multiple happenings with particular reference to the occurrence of multiple attacks of disease or of repeated accidents, *J. Roy. Statist. Soc. Ser. A*, **83**, 255–279.

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

Henna, J. (1985). On estimating the number of constituents of a finite mixture of continuous distributions, *Ann. Inst. Statist. Math.*, **37**, 235–240.

Izenmann, A. J. and Sommer, C. (1988). Philatelic mixtures and multimodal densities, *J. Amer. Statist. Assoc.*, **83**, 941–953.

Karlis, D. and Xekalaki, E. (1996*a*). Testing for finite mixtures via the likelihood ratio test, Tech. Report, No. 28, Department of Statistics, Athens University of Economics and Business.

Karlis, D. and Xekalaki, E. (1996*b*). A note on the maximum likelihood estimation of the parameters of finite Poisson mixtures, Tech. Report, No. 24, Department of Statistics, Athens University of Economics and Business.

Leroux, B. (1992). Consistent estimation of a mixing distribution, *Ann. Statist.*, **20**, 1350–1360.

Leroux, B. and Puterman, M. (1992). Maximum-penalised-likelihood for independent and Markov-dependent mixture models, *Biometrics*, **48**, 545–558.

Lindsay, B. (1983). The geometry of mixture likelihood: A general theory, *Ann. Statist.*, **11**, 86–94.

Lindsay, B. (1989). Moment matrices: Application in mixtures, *Ann. Statist.*, **17**, 722–740.

Lindsay, B. and Roeder, K. (1992). Residuals diagnostics for mixture models, *J. Amer. Statist. Assoc.*, **87**, 785–794.

McLachlan, G. (1987). On bootstraping the likelihood ratio test statistic for the number of components in a normal mixture, *Applied Statistics*, **36**, 318–324.

Mendell, N., Thode, H. and Finch, S. J. (1991). The likelihood ratio test for the 2-component normal mixture problem: Power and sample size analysis, *Biometrics*, **47**, 1143–1148.

Mendell, N., Finch, S. J. and Thode, H. C. (1993). Where is the likelihood ratio test powerful for detecting two components normal mixture? (The consultant's forum), *Biometrics*, **49**, 907–915.

Richardson, S. and Green, P. (1997). On Bayesian analysis of mixtures with an unknown number of components, *J. Roy. Statist. Soc. Ser. B*, **59**, 751–793.

Self, S. and Liang, K. (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions, *J. Amer. Statist. Assoc.*, **82**, 605–610.

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

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

Titterington, M., Markov, G. and Smith, A. F. M. (1985). *Statistical Analysis of Finite Mixtures*, Willey, London.

Thode, H., Finch, S. and Mendell, N. (1988). Simulated percentage points for the null distribution of the likelihood ratio test for a mixture of two normals, *Biometrics*, **44**, 1195–1201.

Windham, M. and Cutler, A. (1992). Information ratios for validating mixture analyses, *J. Amer. Statist. Assoc.*, **87**, 1188–1192.

Wolfe, J. H. (1970). Pattern clustering by multivariate mixture analysis, *Multivariate Behavioral Research*, **5**, 329–350.