# Comparison of Mixture and Classification Maximum Likelihood approaches in Poisson Regression Models

Susana Faria[1] and Gilda Soromenho[2]

[1] Department of Mathematics for Science and Technology, Research Centre Officina Mathematica, University of Minho, 4800-058 Guimarães, Portugal *sfaria@mct.uminho.pt*

[2] Faculty of Psychology and Sciences of Education, Research LEAP, University of Lisbon. Portugal *soromenhop@sapo.pt*

**Abstract.** In this work, we propose to compare two algorithms to compute maximum likelihood estimators of the parameters of a mixture Poisson regression models. To estimate these parameters, we may use the EM algorithm in a mixture approach or the CEM algorithm in a classification approach. The comparison of the two procedures was done through a simulation study of the performance of these approaches on simulated data sets in a target number of iterations. Simulation results show that the CEM algorithm is a good alternative to the EM algorithm for fitting Poisson mixture regression models, having the advantage of converging more quickly.

**Keywords:** Maximum Likelihood Estimation, EM algorithm, Classification EM Algorithm, Mixture Poisson Regression Models, Simulation Study

## 1 Introduction

Finite mixture models are a well-known method for modelling unobserved heterogeneity (see e.g. McLachlan et al. (2000) and Fruhwirth-Schnatter (2006) for a review). The study of these models is a well-established and active area of statistical research and mixtures of regressions have also been studied fairly extensively.

In this work, we study the procedure for fitting Poisson mixture regression models, which are commonly used to analyze heterogeneous count data (see Wedel et al. (1993)), by means of maximum likelihood. We apply two maximization algorithms to obtain the maximum likelihood estimates: the Expectation Maximization (EM) algorithm (see Dempster et al.(1977)) and the Classification Expectation Maximization (CEM) algorithm (see Celeux et al. (1992)).

The comparison of these two different approaches in a cluster analysis is well known in the mixture models literature (see Celeux et al. (1993) and Govaert et al. (1996))

Our goal is to compare the performance of these two approaches using samples drawn from mixtures of Poisson regression model. The paper is organized as follows: in Section 2, we present the Poisson mixture regression models and the two maximization algorithms to obtain the maximum likelihood estimates. Section 3 provides a simulation study investigating the performance of the algorithms for fitting two and three component mixtures of Poisson regression models. In Section 4 the conclusions of our study are drawn.

## 2 Poisson Mixture Regression Models

Let $Y_i$ denote the $i-$th response variable, observed in reaction to an independent variable $\mathrm{x}_i^T = (1, x_{i1}, \ldots, x_{ip})$. It is assumed that the marginal distribution of $Y_i$ follows a mixture of Poisson distributions,

$$Y_i \sim \sum_{j=1}^{J} \pi_j f_j(y_i|\lambda_{i|j}) \tag{1}$$

where

$$f_j(y_i|\lambda_{i|j}) = \frac{\exp^{-\lambda_{i|j}}(\lambda_{i|j})^{y_i}}{y_i!}, \quad i = 1, \ldots, n, j = 1, \ldots, J \tag{2}$$

and $\lambda_{i|j} = exp(\mathrm{x}_i\beta_j)$, $\beta_j = (\beta_{0j}, \beta_{1j}, \ldots, \beta_{pj})$ $(j = 1, \ldots, J)$ denotes the (p+1)-dimensional vector of regressor variables for the $j$th component and $\pi_j$ are the mixing probabilities ($0 < \pi_j < 1$, for all $j = 1, \ldots, J$ and $\sum_j \pi_j = 1$).

### 2.1 The EM Algorithm

The standard tool for finding maximum likelihood solution is the EM algorithm (Dempster et al. (1977) and McLachlan et al. (2000)).

### 2.2 The CEM Algorithm

## 3 Simulation Study of Algorithm Performance

In order to compare the performance of the two algorithms in fitting Poisson mixture regression models, a simulation study was performed. The scope was limited to the study of two and three components. We used the freeware $R$ to develop the simulation program.

### 3.1 Design of the Study

In this study, the simulated data sets were generated according to the following factors:

*Initial Conditions.* In our simulation study, two different strategies of choosing initial values were considered. In the first strategy, the true values were used as the starting values. In the other strategy we ran the algorithm 20 times from random initial position and we selected the solution out of 20 runs which provided the best value of the optimized criterion (see Celeux et al. (1993)).

*Stopping Rules.* A rather strict stopping criterion for the two algorithms was used: iterations were stopped when the relative change in log-likelihood was smaller than $10^{-40}$.

*Number of Samples.* For each type of simulated data set, we generated 100 samples of given sample size $n = 100$ and $n = 500$.

*Data set.* Each datum $(x_i, y_i)$ was generated by the following scheme. First, a uniform $[0, 1]$ random number $c_i$ was generated and its value is used to select a particular component $j$ from mixture of regressions model. Next, $x_i$ was randomly generated from a uniform $[x_L, x_U]$ distribution and then we have $\lambda = exp(\beta_0 + \beta_1\, x)$. Finally, we simulate the value $y_i$ from the Poisson distribution $P(\lambda)$.

*Measures of Algorithm Performance*: In order to examine the performance of two algorithms, the following criteria was used:

- the mean number of iterations required for convergence;
- the mean square error (MSE) of the parameter estimates over the 100 replications which is given by:

$$MSE(\hat{\theta}_j) = \frac{1}{100} \sum_{m=1}^{100} \left( \hat{\theta}_j^{(m)} - \theta_j \right)^2 \tag{3}$$

The simulation process consists of the following steps:

1. Create a data set of size $n$.

2. Fit a mixture of Poisson regression model to the data using the EM and the CEM algorithms. Save the number of iterations required for convergence and the estimated parameters.

3. Repeat steps 1-2, for a total of 100 trials. Compute the mean number of iterations required for convergence and the mean square error (MSE) of the parameter estimates.

### 3.2 Simulation Results: Two Component Mixture of Linear Regressions

In our numerical experiments, for two component models ($J = 2$), we considered eight groups of different parameters $\theta = (\pi, \beta_{10}, \beta_{11}, \beta_{02}, \beta_{21})$ studied in Yang et al. (2005).

Samples of two different sizes $n$ ($n = 100, 500$) were generated for each set of true parameter values $(\pi, \beta)$ shown on Table 1.

**Table 1.** True parameter values for the essays with a two component mixture of Poisson regressions

| Cases | $\beta_{10}$ | $\beta_{11}$ | $\beta_{20}$ | $\beta_{21}$ | $\pi$ |
|---|---|---|---|---|---|
| $A1$ | 3 | $-1$ | 4 | 1 | 0.5 |
| $A2$ | 4 | $-1$ | 4 | 1 | 0.5 |
| $A3$ | 5 | $-1$ | 4 | 1 | 0.5 |
| $B1$ | 4 | $-0.5$ | 4 | 0.5 | 0.3 |
| $B2$ | 4 | $-0.5$ | 4 | 0.5 | 0.5 |
| $B3$ | 4 | $-0.5$ | 4 | 0.5 | 0.7 |
| $C1$ | 3 | $-0.5$ | 4 | 0.5 | 0.7 |
| $C2$ | 4 | $-0.5$ | 2 | 0.5 | 0.7 |

Table 2 and 3 provide the mean square error of estimators and the mean number of iterations required for convergence using the EM and CEM algorithm for fitting two component mixtures of Poisson regression models. In all cases, the mean number of iterations for convergence is smaller using the CEM algorithm rather than using the EM algorithm. It is evident that the EM and CEM estimators of the regression coefficients and the mixture proportion have relatively small MSE and, in generality, the MSE of EM estimates are smaller than the MSE of CEM estimates. However, when the algorithms are initiated with the true parameter values and for $n = 100$, the CEM algorithm performs generally better. It also seems that the EM and CEM algorithm have practically the same behavior in situations where the overlap is small (A1, C1). The MSE of both estimates tend to approach zero when the sample size increases.

### 3.3 Simulation Results: Three Component Mixture of Linear Regressions

For three component models (J=3), samples of size $n = 100$ and $n = 500$ were generated for the five sets of parameter values $(\pi, \beta)$ shown in Table 4. For illustration we show scatter plots of random samples of 500 points in Figure 1.

Table 4 and 5 report the mean square error of estimators and the mean number of iterations required for convergence using the EM and CEM algorithm for fitting three component mixtures of Poisson regression models. Also in all cases, the mean number of iterations for convergence is smaller using the CEM algorithm rather than using the EM algorithm. It is evident that the EM and CEM estimators of the regression coefficients and the mixture proportion have relatively small MSE, especially when the algorithms

**Table 2.** The mean number of iterations required for convergence and mean square error (MSE) of estimators based on 100 replications of the two component mixtures of Poisson regression models when the true values were used as the starting values

| Cases | $n$ | Algorithm | MSE | | | | | ITER |
|---|---|---|---|---|---|---|---|---|
| | | | $\pi_1$ | $\beta_{10}$ | $\beta_{11}$ | $\beta_{20}$ | $\beta_{21}$ | |
| A1 | | EM | 2,4059E-03 | 1,2489E-02 | 8,2789E-03 | 1,5521E-04 | 5,7553E-06 | 2,84 |
| | | CEM | 2,4060E-03 | 1,2477E-02 | 8,2752E-03 | 1,5520E-04 | 5,7549E-06 | 2,35 |
| A2 | | EM | 2,9046E-03 | 2,0510E-03 | 8,9126E-04 | 1,8048E-04 | 7,1487E-06 | 9,17 |
| | | CEM | 3,3000E-03 | 1,8923E-03 | 8,4791E-04 | 1,7764E-04 | 7,0212E-06 | 5,09 |
| A3 | | EM | 2,9497E-03 | 2,5048E-03 | 1,2430E-03 | 1,6964E-04 | 6,9760E-06 | 9,45 |
| | | CEM | 3,3360E-03 | 2,4072E-03 | 1,2156E-03 | 1,7325E-04 | 7,1209E-06 | 6,04 |
| B1 | 100 | EM | 2,5834E-03 | 6,6455E-03 | 1,1248E-03 | 4,7483E-04 | 2,2572E-05 | 13,04 |
| | | CEM | 2,4710E-03 | 7,6146E-03 | 1,2494E-03 | 4,6372E-04 | 2,1893E-05 | 5,45 |
| B2 | | EM | 2,7556E-03 | 4,2026E-03 | 7,7901E-04 | 7,5701E-04 | 3,6127E-05 | 11,98 |
| | | CEM | 3,3220E-03 | 4,3697E-03 | 7,2379E-04 | 8,3557E-04 | 4,0198E-05 | 6,66 |
| B3 | | EM | 2,6979E-03 | 1,8411E-03 | 4,4683E-04 | 1,1903E-03 | 6,2489E-05 | 10,75 |
| | | CEM | 2,8690E-03 | 1,8867E-03 | 4,2485E-04 | 1,2404E-03 | 6,5341E-05 | 6,86 |
| C1 | | EM | 2,1660E-03 | 4,7207E-03 | 8,6553E-04 | 1,0338E-03 | 5,2448E-05 | 3,18 |
| | | CEM | 2,1650E-03 | 4,7264E-03 | 8,6504E-04 | 1,0346E-03 | 5,2493E-05 | 2,78 |
| C2 | | EM | 2,4164E-03 | 1,9191E-03 | 3,3287E-04 | 9,0757E-03 | 4,6045E-04 | 12,31 |
| | | CEM | 3,1910E-03 | 1,9654E-03 | 3,4092E-04 | 8,6398E-03 | 4,4700E-04 | 6,52 |
| A1 | | EM | 4,8406E-04 | 2,8154E-03 | 1,2703E-03 | 3,2709E-05 | 1,2815E-06 | 3,40 |
| | | CEM | 4,8380E-04 | 2,8111E-03 | 1,2710E-03 | 3,2715E-05 | 1,2813E-06 | 2,70 |
| A2 | | EM | 4,3286E-04 | 1,1347E-03 | 5,8600E-04 | 3,3374E-05 | 1,3708E-06 | 9,64 |
| | | CEM | 4,8832E-04 | 1,4985E-03 | 7,1637E-04 | 3,3479E-05 | 1,3719E-06 | 5,60 |
| A3 | | EM | 5,6578E-04 | 3,7905E-04 | 2,1872E-04 | 3,1860E-05 | 1,1669E-06 | 8,12 |
| | | CEM | 6,5812E-04 | 4,0076E-04 | 2,2725E-04 | 3,2058E-05 | 1,1739E-06 | 5,87 |
| B1 | 500 | EM | 4,4439E-04 | 1,3048E-03 | 1,9581E-04 | 8,7125E-05 | 3,9960E-06 | 12,04 |
| | | CEM | 4,6296E-04 | 2,0424E-03 | 2,8605E-04 | 9,4530E-05 | 4,2836E-06 | 6,40 |
| B2 | | EM | 4,6078E-04 | 6,2492E-04 | 1,1839E-04 | 1,4969E-04 | 6,9001E-06 | 10,82 |
| | | CEM | 6,0076E-04 | 1,4175E-03 | 1,8849E-04 | 1,6580E-04 | 7,5461E-06 | 6,45 |
| B3 | | EM | 4,7933E-04 | 4,0734E-04 | 9,1096E-05 | 2,4983E-04 | 1,1926E-05 | 9,63 |
| | | CEM | 5,1424E-04 | 5,6083E-04 | 1,1413E-04 | 2,2744E-04 | 1,1173E-05 | 6,10 |
| C1 | | EM | 4,6506E-04 | 1,1051E-03 | 2,2268E-04 | 1,7817E-04 | 9,3458E-06 | 3,45 |
| | | CEM | 4,6512E-04 | 1,1080E-03 | 2,2302E-04 | 1,7804E-04 | 9,3416E-06 | 2,90 |
| C2 | | EM | 5,8938E-04 | 4,4807E-04 | 8,8556E-05 | 2,0458E-03 | 9,5420E-05 | 11,35 |
| | | CEM | 1,4763E-03 | 4,5191E-04 | 9,1394E-05 | 2,1511E-03 | 9,9098E-05 | 6,57 |

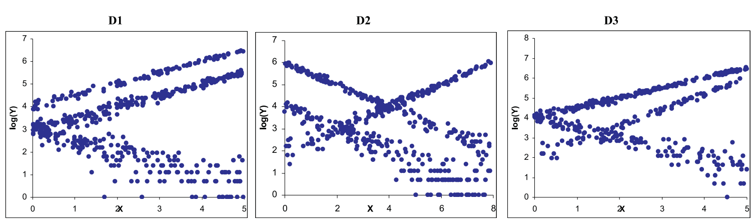


**Fig. 1.** Scatter plot of samples from 3 component models with $n = 500$

are initiated with the true parameter values. In generality, EM outperforms CEM by producing estimates of the parameters that have smaller MSE, but its performance decreases when the sample size decreases. Also, the MSE of both estimates tend to approach zero when the sample size increases.

**Table 3.** The mean number of iterations required for convergence and mean square error (MSE) of estimators based on 100 replications of the two component mixtures of Poisson regression models when the second strategy was used as the starting values

| Cases | $n$ | Algorithm | MSE | | | | | ITER |
|---|---|---|---|---|---|---|---|---|
| | | | $\pi_1$ | $\beta_{10}$ | $\beta_{11}$ | $\beta_{20}$ | $\beta_{21}$ | |
| A1 | | EM | 2,9225E-03 | 9,0650E-03 | 2,5830E-03 | 1,2813E-04 | 4,8572E-06 | 3,81 |
| | | CEM | 2,9226E-03 | 9,0650E-03 | 2,5830E-03 | 1,2812E-04 | 4,8570E-06 | 3,73 |
| A2 | | EM | 2,7007E-03 | 8,3178E-03 | 3,6713E-03 | 2,1419E-04 | 7,9847E-06 | 10,74 |
| | | CEM | 2,8140E-03 | 7,7843E-03 | 3,4395E-03 | 2,1553E-04 | 8,0382E-06 | 7,65 |
| A3 | | EM | 2,5467E-03 | 1,4764E-03 | 8,0824E-04 | 1,3363E-04 | 5,1782E-06 | 11,38 |
| | | CEM | 2,9055E-03 | 2,1145E-03 | 9,3221E-04 | 1,3391E-04 | 5,2184E-06 | 8,65 |
| B1 | 100 | EM | 1,9315E-03 | 5,5026E-03 | 8,5110E-04 | 3,7692E-04 | 1,8047E-05 | 15,64 |
| | | CEM | 2,0240E-03 | 7,5523E-03 | 9,5542E-04 | 3,9024E-04 | 1,8376E-05 | 8,36 |
| B2 | | EM | 3,0347E-03 | 3,7530E-03 | 6,1041E-04 | 6,5411E-04 | 3,0978E-05 | 12,54 |
| | | CEM | 3,4798E-03 | 3,9230E-03 | 6,4672E-04 | 7,5143E-04 | 3,5387E-05 | 7,90 |
| B3 | | EM | 2,0122E-03 | 1,6752E-03 | 6,3327E-04 | 8,7551E-04 | 4,3932E-05 | 12,03 |
| | | CEM | 2,0445E-03 | 1,7366E-03 | 6,8045E-04 | 8,9967E-04 | 4,6028E-05 | 10,34 |
| C1 | | EM | 1,8954E-03 | 4,8617E-03 | 9,3506E-04 | 9,6653E-04 | 4,7294E-05 | 4,42 |
| | | CEM | 1,8955E-03 | 4,8619E-03 | 9,3503E-04 | 9,6652E-04 | 4,7297E-05 | 4,38 |
| C2 | | EM | 2,7182E-03 | 2,3314E-03 | 4,8860E-04 | 8,2764E-03 | 3,9897E-04 | 17,94 |
| | | CEM | 3,8127E-03 | 2,2748E-03 | 4,8116E-04 | 9,5038E-03 | 4,4581E-04 | 11,35 |
| A1 | | EM | 5,4356E-04 | 2,9842E-03 | 1,4281E-03 | 3,3599E-05 | 1,3061E-06 | 4,48 |
| | | CEM | 5,4363E-04 | 2,9856E-03 | 1,4264E-03 | 3,3606E-05 | 1,3064E-06 | 4,18 |
| A2 | | EM | 4,3297E-04 | 8,5663E-04 | 5,2794E-04 | 2,6460E-04 | 1,1582E-06 | 10,76 |
| | | CEM | 4,6288E-04 | 1,5729E-03 | 6,0889E-04 | 3,3372E-05 | 1,2899E-06 | 9,94 |
| A3 | | EM | 5,6275E-04 | 3,2312E-04 | 1,5401E-04 | 3,8856E-05 | 1,4417E-06 | 10,32 |
| | | CEM | 6,9108E-04 | 3,3758E-04 | 1,5378E-04 | 3,8108E-05 | 1,4274E-06 | 9,51 |
| B1 | 500 | EM | 6,2822E-03 | 8,8400E-04 | 1,8212E-04 | 5,6435E-04 | 4,6851E-06 | 16,19 |
| | | CEM | 3,8152E-04 | 2,5997E-03 | 3,7517E-04 | 1,0030E-04 | 4,8691E-06 | 8,75 |
| B2 | | EM | 5,0787E-04 | 7,4878E-04 | 1,6089E-04 | 2,3008E-04 | 4,6002E-06 | 12,11 |
| | | CEM | 6,2093E-04 | 1,3715E-03 | 1,9079E-04 | 1,1952E-04 | 5,0020E-06 | 10,42 |
| B3 | | EM | 3,5098E-04 | 4,2143E-04 | 8,3468E-05 | 2,1760E-04 | 1,0176E-05 | 11,73 |
| | | CEM | 4,3180E-04 | 6,5103E-04 | 1,0461E-04 | 2,8667E-04 | 1,1222E-05 | 11,08 |
| C1 | | EM | 3,6296E-04 | 1,5168E-03 | 2,7209E-04 | 1,9389E-04 | 8,8194E-06 | 5,18 |
| | | CEM | 3,6357E-04 | 1,4960E-03 | 2,6903E-04 | 1,9340E-04 | 8,7814E-06 | 5,00 |
| C2 | | EM | 5,6718E-04 | 4,6880E-04 | 6,6114E-05 | 1,0945E-03 | 2,9860E-05 | 13,08 |
| | | CEM | 1,0871E-03 | 5,1625E-04 | 7,1189E-05 | 1,1732E-03 | 3,1553E-05 | 11,31 |

**Table 4.** True parameter values for the essays with a three component mixture of Poisson regressions

| Cases | $\beta_{10}$ | $\beta_{11}$ | $\beta_{20}$ | $\beta_{21}$ | $\beta_{30}$ | $\beta_{31}$ | $\pi_1$ | $\pi_2$ |
|---|---|---|---|---|---|---|---|---|
| $D1$ | 3 | $-0.5$ | 4 | 0.5 | 3 | 0.5 | 0.4 | 0.4 |
| $D2$ | 4 | $-0.5$ | 2 | 0.5 | 6 | $-0.5$ | 0.4 | 0.3 |
| $D3$ | 4 | $-0.5$ | 4 | 0.5 | 2 | 0.8 | 0.4 | 0.4 |
| $D4$ | 4 | $-0.5$ | 4 | 0.5 | 2 | 0.8 | 0.4 | 0.3 |
| $D5$ | 4 | $-0.5$ | 4 | 0.5 | 2 | 0.8 | 0.3 | 0.5 |

**Table 5.** Mean square error (MSE) and the mean number of iterations required for convergence when the true values were used as the starting values for three component mixtures of Poisson regression models

| | $n$ | Alg | MSE | | | | | | | | | IT |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | $\pi_1$ | $\pi_2$ | $\pi_3$ | $\beta_{10}$ | $\beta_{11}$ | $\beta_{20}$ | $\beta_{21}$ | $\beta_{30}$ | $\beta_{31}$ | |
| D1 | | EM | 5,9029E-05 | 6,5323E-03 | 6,0397E-04 | 1,3167E-02 | 2,4942E-03 | 2,7613E-03 | 2,6532E-03 | 7,6084E-04 | 1,7330E-03 | 15,28 |
| | | CEM | 5,8975E-05 | 6,9648E-03 | 6,1803E-04 | 1,4987E-02 | 2,7730E-03 | 2,8892E-03 | 2,6540E-03 | 7,6024E-04 | 2,1250E-03 | 7,02 |
| D2 | | EM | 1,6465E-04 | 1,0145E-03 | 1,8595E-04 | 6,1498E-03 | 2,8287E-03 | 7,2314E-04 | 2,9125E-03 | 6,5752E-03 | 2,7855E-03 | 13,02 |
| | 100 | CEM | 1,7441E-04 | 9,8840E-04 | 1,7521E-04 | 6,0942E-03 | 3,3920E-03 | 7,3820E-04 | 3,8560E-03 | 6,9452E-03 | 3,4120E-03 | 7,22 |
| D3 | | EM | 4,4075E-05 | 6,9382E-03 | 2,9730E-04 | 6,2091E-03 | 2,2859E-03 | 1,1027E-03 | 2,0705E-03 | 8,6231E-04 | 1,4406E-03 | 14,49 |
| | | CEM | 4,6454E-05 | 6,9113E-03 | 2,9292E-04 | 6,4598E-03 | 2,5470E-03 | 9,8237E-04 | 2,3890E-03 | 9,0597E-04 | 1,4640E-03 | 7,10 |
| D4 | | EM | 4,2643E-05 | 5,9664E-03 | 2,4007E-04 | 5,0188E-03 | 2,7088E-03 | 7,6235E-04 | 1,2762E-02 | 7,8814E-04 | 1,2735E-02 | 14,24 |
| | | CEM | 4,2653E-05 | 5,8225E-03 | 2,3715E-04 | 5,8173E-03 | 3,4970E-03 | 8,3082E-04 | 1,3129E-02 | 8,0515E-04 | 1,4804E-02 | 8,20 |
| D5 | | EM | 3,7884E-05 | 5,7084E-03 | 2,7355E-04 | 6,7253E-03 | 2,3963E-03 | 1,1168E-03 | 2,8565E-03 | 6,8101E-04 | 1,5833E-03 | 17,45 |
| | | CEM | 3,6196E-05 | 5,8297E-03 | 2,7760E-04 | 7,6287E-03 | 2,5270E-03 | 1,1825E-03 | 3,0060E-03 | 6,5483E-04 | 1,7390E-03 | 6,93 |
| D1 | | EM | 5,2864E-04 | 4,5448E-04 | 3,5210E-04 | 9,6984E-04 | 1,6294E-04 | 1,4355E-04 | 7,1127E-06 | 1,0983E-03 | 4,3782E-05 | 12,59 |
| | | CEM | 7,9596E-04 | 5,3060E-04 | 5,1960E-04 | 1,1632E-03 | 1,7715E-04 | 1,5796E-04 | 7,3823E-06 | 1,0747E-03 | 4,5091E-05 | 7,69 |
| D2 | | EM | 5,9451E-04 | 4,4723E-04 | 5,0924E-04 | 8,6938E-04 | 1,5701E-04 | 1,9070E-04 | 9,6561E-06 | 9,6937E-04 | 3,8801E-05 | 12,71 |
| | 500 | CEM | 8,9240E-04 | 4,7860E-04 | 6,9164E-04 | 1,3378E-03 | 1,9477E-04 | 2,2205E-04 | 1,0436E-05 | 9,7401E-04 | 3,9105E-05 | 7,95 |
| D3 | | EM | 4,1384E-04 | 5,8108E-04 | 4,4181E-04 | 1,5409E-03 | 2,4640E-04 | 8,7286E-05 | 4,9136E-06 | 1,2775E-03 | 5,0717E-05 | 13,80 |
| | | CEM | 4,8456E-04 | 6,2888E-04 | 5,1448E-04 | 2,1753E-03 | 3,2236E-04 | 9,9044E-05 | 5,4319E-06 | 1,3675E-03 | 5,3688E-05 | 7,82 |
| D4 | | EM | 4,2075E-04 | 5,3687E-04 | 2,9318E-04 | 1,8425E-03 | 4,0906E-04 | 1,1235E-04 | 8,7797E-06 | 1,1642E-03 | 7,9796E-05 | 12,06 |
| | | CEM | 5,4108E-04 | 5,3716E-04 | 3,6632E-04 | 3,2361E-03 | 6,1755E-04 | 1,1265E-04 | 8,7931E-06 | 1,7814E-03 | 1,2193E-04 | 6,48 |
| D5 | | EM | 4,0161E-04 | 4,6683E-04 | 4,5871E-04 | 9,7668E-04 | 1,2755E-04 | 1,3316E-03 | 3,2107E-05 | 1,6169E-04 | 2,2801E-05 | 11,15 |
| | | CEM | 6,7548E-04 | 9,2964E-04 | 5,5992E-04 | 1,0561E-03 | 1,4072E-04 | 1,3853E-03 | 3,3031E-05 | 1,5906E-04 | 2,2383E-05 | 7,27 |

**Table 6.** Mean square error (MSE) and the mean number of iterations required for convergence when the second strategy was used as the starting values for three component mixtures of Poisson regression models

| | $n$ | Alg | MSE | | | | | | | | | IT |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | $\pi_1$ | $\pi_2$ | $\pi_3$ | $\beta_{10}$ | $\beta_{11}$ | $\beta_{20}$ | $\beta_{21}$ | $\beta_{30}$ | $\beta_{31}$ | |
| D1 | | EM | 2,3345E-03 | 2,4518E-03 | 1,8433E-03 | 8,2859E-03 | 1,7057E-03 | 1,0728E-03 | 9,0160E-05 | 5,4683E-03 | 3,8075E-04 | 17,81 |
| | | CEM | 2,5734E-03 | 2,4181E-03 | 2,1191E-03 | 1,1987E-02 | 1,9442E-03 | 1,6930E-03 | 3,8338E-04 | 9,5458E-03 | 4,4256E-04 | 10,72 |
| D2 | | EM | 3,1526E-03 | 3,5667E-03 | 1,5101E-03 | 8,5315E-03 | 1,1428E-03 | 2,5975E-03 | 9,3229E-03 | 1,2574E-03 | 5,2940E-03 | 21,56 |
| | 100 | CEM | 3,2516E-03 | 5,1452E-03 | 2,5548E-03 | 4,3336E-03 | 1,1973E-03 | 9,5463E-03 | 8,2628E-03 | 7,0493E-03 | 4,8767E-03 | 13,55 |
| D3 | | EM | 2,8015E-03 | 2,3836E-03 | 1,5625E-03 | 7,1209E-04 | 1,3200E-04 | 5,4053E-04 | 3,9797E-03 | 1,6472E-03 | 9,0171E-03 | 19,54 |
| | | CEM | 3,5129E-03 | 2,7121E-03 | 1,7356E-03 | 8,1833E-04 | 5,5074E-05 | 2,8697E-04 | 5,5681E-03 | 1,0886E-03 | 6,8317E-03 | 11,18 |
| D4 | | EM | 2,6741E-03 | 1,1034E-03 | 1,0219E-03 | 8,4975E-04 | 1,9258E-04 | 7,3966E-04 | 4,4754E-03 | 1,8592E-03 | 1,1680E-03 | 19,73 |
| | | CEM | 3,3587E-03 | 1,1614E-03 | 1,2767E-03 | 9,4387E-04 | 1,8490E-04 | 1,0496E-03 | 6,5185E-03 | 1,9667E-03 | 1,4852E-03 | 11,63 |
| D5 | | EM | 1,0552E-03 | 1,2708E-03 | 1,3621E-03 | 9,4138E-04 | 1,0813E-04 | 6,7421E-04 | 4,3799E-03 | 1,4003E-03 | 1,0709E-02 | 19,04 |
| | | CEM | 1,1355E-03 | 1,2603E-03 | 1,2613E-03 | 9,5659E-04 | 5,7220E-05 | 5,1211E-04 | 4,9318E-03 | 8,7235E-04 | 8,3325E-03 | 11,88 |
| D1 | | EM | 4,0157E-04 | 5,0423E-04 | 1,9974E-03 | 5,9170E-04 | 9,9258E-05 | 2,6945E-04 | 1,2116E-05 | 2,8963E-03 | 4,7819E-05 | 17,10 |
| | | CEM | 1,2796E-03 | 1,2310E-03 | 1,5460E-02 | 2,8506E-03 | 3,6450E-04 | 3,5328E-03 | 1,6895E-05 | 2,0301E-03 | 3,3105E-04 | 11,65 |
| D2 | | EM | 4,9674E-04 | 4,1812E-04 | 4,7946E-04 | 1,0872E-03 | 1,9308E-04 | 2,4648E-04 | 1,3329E-05 | 7,7652E-04 | 3,2616E-05 | 18,15 |
| | 500 | CEM | 9,4343E-04 | 8,0388E-04 | 9,9794E-04 | 1,5631E-03 | 2,2947E-04 | 7,9862E-04 | 1,2169E-05 | 8,0165E-04 | 3,3786E-05 | 14,00 |
| D3 | | EM | 6,5126E-04 | 4,7191E-04 | 1,1913E-03 | 1,4969E-03 | 4,5723E-04 | 2,0237E-04 | 5,5819E-06 | 8,8861E-03 | 4,8776E-05 | 21,23 |
| | | CEM | 7,3477E-04 | 1,7864E-03 | 3,1636E-03 | 2,1111E-03 | 5,2864E-04 | 4,1630E-03 | 6,0537E-05 | 1,1667E-03 | 1,3157E-05 | 10,49 |
| D4 | | EM | 4,3870E-04 | 3,6082E-04 | 3,3218E-04 | 3,4986E-03 | 5,6718E-04 | 2,0644E-04 | 1,3989E-05 | 1,3689E-03 | 9,5668E-05 | 18,68 |
| | | CEM | 4,8877E-04 | 3,6129E-04 | 4,0586E-04 | 5,0061E-03 | 9,0093E-04 | 2,0568E-04 | 1,3921E-05 | 1,9598E-03 | 1,2887E-04 | 13,42 |
| D5 | | EM | 5,1023E-04 | 4,3449E-04 | 4,7250E-04 | 1,2076E-03 | 1,5268E-04 | 1,2660E-03 | 2,8645E-05 | 1,9520E-04 | 2,9281E-05 | 22,65 |
| | | CEM | 7,7904E-04 | 9,6812E-04 | 5,8524E-04 | 1,2528E-03 | 1,7133E-04 | 1,2887E-03 | 2,8470E-05 | 1,9529E-04 | 3,0116E-05 | 15,60 |

### 3.4 Conclusion

In this work, we compared the performance of two algorithms to compute maximum likelihood estimators of the parameters of a mixture Poisson regression models, the EM and the CEM algorithm.

We run a number of simulations, and in all of them the CEM algorithm converged in fewer iterations than the EM algorithm, which implies a reduction in the computational time to reach the parameter estimates.

Simulation results show that the CEM algorithm is a good alternative to the EM algorithm for fitting Poisson mixture regression models, however, the choice of the approach seems depend on the size of the samples.

## References

CELEUX, G. and GOVAERT, G. (1992): A classification EM Algorithm and two stochastic versions, *Computational Statistics & Data Analysis, 14, 315-332*.

CELEUX, G. and GOVAERT, G.(1993): Comparison of the mixture and the classification maximum likelihood in cluster analysis. *J. Statistical Computation and Simulation, 47, 127-146*.

DEMPSTER, A.P. and LAIRD, N.M. and RUBIN, D.B.(1977): Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society B, 39, 1-38*.

FRUHWIRTH-SCHNATTER(2006): *Finite Mixture and Markov Switching Models*, Springer, Heidelberg.

GOVAERT, G. and NADIF, M. (1993): Comparison of the mixture and the classification maximum likelihood in cluster analysis with binary data. *Computational Statistics & Data Analysis, 23, 65-81*.

MCLACHLAN, G.J. and PEEL, D., (2000): *Finite Mixture Models*, Wiley, New York.

WEDEL, M., DESARBO, W.S., BULT, J.R. and RAMASWAMY, V.(1993): A Latent Class Poisson Regression Model for Heterogeneous Count Data. *Journal of Applied Econometrics, 8, 397 - 411*.

YANG, MS and LAI, CY. (2005): Mixture Poisson regression models for heterogenous count data based on latent and fuzzy class analysis. *Soft Computing, 9, 519-524*