Comparison of EM and SEM Algorithms in Poisson Regression Models: a simulation study

Susana Faria, Gilda Soromenho

Comparison of EM and SEM Algorithms in Poisson Regression Models: a simulation study

Journal:	*Communications in Statistics - Simulation and Computation*
Manuscript ID:	LSSP-2010-0452.R1
Manuscript Type:	Original Paper
Date Submitted by the Author:	23-May-2011
Complete List of Authors:	Faria, Susana; University of Minho, Department of Mathematics and Applications Soromenho, Gilda; University of Lisbon
Keywords:	Maximum likelihood estimation, EM algorithm, Stochastic EM algorithm, Mixture Poisson regression models, Simulation study
Abstract:	In this work, we propose to compare two algorithms to compute maximum likelihood estimates for the parameters of a mixture Poisson regression model: the EM algorithm and the Stochastic EM algorithm. The comparison of the two procedures was done through a simulation study of the performance of these approaches on simulated data sets and real data sets. Simulation results show that the choice of the approach depends essentially on the overlap of the regression lines. In the real data case, we show that the Stochastic EM algorithm resulted in model estimates that best fit the regression model.

Note: The following files were submitted by the author for peer review, but cannot be converted to PDF. You must view these files (e.g. movies) online.

sfariasoromenho.zip

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

SCHOLARONE™
Manuscripts

Comparison of EM and SEM Algorithms in Poisson Regression Models: a simulation study

Susana Faria[1] ,Gilda Soromenho[2]

[1] *Department of Mathematics and Applications , University of Minho, 4800-058 Guimarães, Portugal, sfaria@math.uminho.pt*

[2] *Institute of Education, University of Lisbon, Portugal , gspereira@ie.ul.pt*

Abstract: In this work, we propose to compare two algorithms to compute maximum likelihood estimates for the parameters of a mixture Poisson regression model: the EM algorithm and the Stochastic EM algorithm. The comparison of the two procedures was done through a simulation study of the performance of these approaches on simulated data sets and real data sets. Simulation results show that the choice of the approach depends essentially on the overlap of the regression lines. In the real data case, we show that the Stochastic EM algorithm resulted in model estimates that best fit the regression model.

Keywords: Maximum likelihood estimation, EM algorithm, Stochastic EM algorithm, Mixture Poisson regression models, Simulation study

1   Introduction

Finite mixture models are a well-known method for modelling data that arise from a heterogeneous population (see e.g. McLachlan et al., 2000 and Fruhwirth-Schnatter, 2006 for a review). The

1

study of these models is a well-established and active area of statistical research and mixtures of re-gressions have also been studied fairly extensively. In particular, Poisson mixture regression models are commonly used to analyze heterogeneous count data. Wedel et al. (1993) proposed a class Pois-son regression model and an EM algorithm for estimation was described. Wang et al. (1996) studied mixed Poisson regression models and maximum likelihood estimates of the parameters were ob-tained by combining EM and quasi-Newton algorithms.

In this work, we study the procedure for fitting Poisson mixture regression models by means of maximum likelihood (ML). We apply two maximization algorithms to obtain the maximum likeli-hood estimates: the Expectation Maximization (EM) algorithm (Dempster et al., 1977) and the Sto-chastic Expectation Maximization (SEM) algorithm (Celeux and Diebolt, 1985).

The comparison of EM and SEM approaches in a mixture of distributions is well known. Celeux et al. (1996) have investigated the practical behaviour of these algorithms through intensive Monte Carlo numerical simulations and a real data study. Dias and Wedel (2004) have compared EM and SEM algorithms to estimate the parameters of Gaussian mixture model. Faria and Soromenho (2010) have performed a simulation study to compare the performance of these two approaches on Gaussian mixtures of linear regressions.

This paper is organized as follows: Section 2 describes the model. Parameter estimation based on the EM algorithm and the Stochastic EM algorithm is discussed in Section 3. Section 4 provides a simulation study investigating the performance of these algorithms for fitting two and three compo-nent mixtures of Poisson regression models. We also study the performance of algorithms in real data sets in section 5. In Section 6 the conclusions of our study are drawn.

2

2   Poisson mixture regression models

Let the random variable Y_i denote the ith response variable, and let $(y_i, x_i), i = 1, \ldots, n$ denote observations where y_i is the observed value of Y_i and x_i is a $(p+1)$-dimensional covariate vector. It is assumed that the marginal distribution of Y_i follows a mixture of Poisson distributions,

$$h(y_i \mid x_i, \theta) = \sum_{j=1}^{J} \pi_j f_j(y_i \mid x_i) \qquad (1)$$

where

$$f_j(y_i \mid x_i) = \frac{\exp(-\lambda_{ij})(\lambda_{ij})^{y_i}}{y_i!}, \qquad i = 1, \ldots, n, j = 1, \ldots, J \qquad (2)$$

and $\lambda_{ij} = \exp(\beta_j^T x_i)$, with $\beta_j = (\beta_{j0}, \beta_{j1}, \ldots, \beta_{jp})^T$ denoting the (p+1)-dimensional vector of regression coefficients for jth component and $\theta = (\pi_1, \ldots, \pi_J, \beta_1, \ldots, \beta_J)$ denotes the vector of all parameters. The proportions π_j are the mixing probabilities ($0 < \pi_j < 1$, $j = 1, \ldots, J$ and $\sum_{j=1}^{J} \pi_j = 1$) and can be interpreted as the unconditional probabilities that an individual belongs to component j of the mixture.

To be able to reliably estimate the parameters of mixture models we require identifiability. That is, two sets of parameters do not yield the same mixture distribution. Finite mixtures of Poisson distributions are identifiable (see Teicher, 1960 for details). Fruhwirth-Schnatter (2006) shows that if the covariate matrix is of full rank and the mixing proportions, π_j, are all different, then the Poisson mixture regression model is identifiable.

3

3   Parameter Estimation

Among the various estimation methods considered in the literature for finite mixture models, the maximum likelihood (ML) has dominated the field.

For a given number of J components, the task is to estimate the vector of parameters $\theta = (\pi_1, ..., \pi_J, \beta_1, ..., \beta_J)$ that maximizes the log-likelihood

$$L(\theta \mid x_1, ..., x_n, y_1, ..., y_n) = \sum_{i=1}^{n} \log\left(\sum_{j=1}^{J} \pi_j f_j(y_i \mid x_i) \right) \qquad (3)$$

The standard tool for finding maximum likelihood solution is the Expectation Maximization (EM) algorithm. However, it suffers from slow convergence and may converge to local maxima or saddle points. The Stochastic Expectation Maximization (SEM) algorithm is a viable alternative to find the ML estimates of the parameters of a mixture model. The SEM algorithm by using random drawing at each iteration, prevents from being trapped in local optima. It has some advantages over the EM algorithm: it does not get stuck; it often provides more information about the data (see Diebolt and Ip, 1996), for instance when parameters cannot be estimated; and in certain conditions behaves better than EM algorithm (see Celeux et al., 1996).

3.1 The EM algorithm

The EM algorithm is a broadly applicable approach to the iterative computation of maximum-likelihood estimates when the observations can be viewed as incomplete data. The idea here is to think of the data as consisting of triples $(y_i, x_i, z_i), i = 1, ..., n$ where $z_i = (z_{i1}, ..., z_{ij})^T, i = 1, ..., n, j = 1, ... J$ is the unobserved indicator that specifies the mixture compo-

nent from which the observation (y_i, x_i) is drawn, i.e., z_{ij} equals 1 if observation i comes from component j and 0 otherwise.

The log likelihood for the complete data is

$$L(\theta \mid x_1,...,x_n, y_1,...,y_n) = \sum_{i=1}^{n} \sum_{j=1}^{J} z_{ij} \log(\pi_j) + \sum_{i=j=1}^{n} \sum_{j=1}^{J} z_{ij} \log(f_j(y_i \mid x_i)) \quad (4)$$

The EM algorithm is easy to program and proceeds iteratively in two steps, E (for expectation) and M (for maximization). At the E-step, it replaces the missing data by its expectation conditional on the observed data. At the M-step, it finds the parameter estimates which maximize the expected log likelihood for the complete data, conditional on the expected values of the missing data.

This procedure can be stated as follows.

E-step: Given the current parameter estimates $\theta^{(r)}$ in the rth iteration, replace the missing data z_{ij} by the estimated probabilities that the i observation belongs to the jth component of the mixture,

$$w_{ij}^{(r)} = \frac{\pi_j^{(r)} f_j(y_i \mid x_i, \beta_{ij}^{(r)})}{\sum\limits_{j=1}^{J} \pi_j^{(r)} f_j(y_i \mid x_i, \beta_{ij}^{(r)})} \quad (5)$$

M-step: Given the estimates for the probabilities $w_{ij}^{(r)}$ (which are functions of $\theta^{(r)}$), obtain new estimates $\theta^{(r+1)}$ of the parameters by maximizing

$$Q(\theta^{(r+1)} \mid \theta^{(r)}) = Q_1 + Q_2 \quad (6)$$

under the restriction for the component weights and where

$$Q_1 = \sum_{i=1}^{n} \sum_{j=1}^{J} w_{ij} \log(\pi_j^{(r+1)}) \quad (7)$$

5

and

$$Q_2 = \sum_{i=}^{n} \sum_{j=1}^{J} w_{ij} \log\left(f_j\left(y_i \mid x_i, \beta_j^{(r+1)} \right) \right) \quad (8)$$

The maximization of Q_1 under the restriction for the component weights, $\sum_{j=1}^{J} \pi_j = 1$, is equivalent

to maximizing the function

$$Q_1^* = \sum_{i=1}^{n} \sum_{j=1}^{J} w_{ij} \log\left(\pi_j^{(r+1)} \right) - \mu\left(\sum_{j=1}^{J} \pi_j^{(r+1)} - 1 \right)$$

where μ is the Lagrangian multiplier. Setting the derivative of function Q_1^* with respect to $\pi_j^{(r+1)}$

equal to zero yields

$$\hat{\pi}_j^{(r+1)} = \frac{\sum_{i=1}^{n} w_{ij}^{(r)}}{n}, j = 1, \ldots, J \quad (9)$$

and Q_2 is maximized separately for each $j = 1, \ldots, J$ using weighted ML estimation of generalized

linear models (GLM).

3.2 The Stochastic EM algorithm

We also apply a procedure for fitting Poisson mixture regression models using a stochastic ver-

sion of the EM algorithm, the so-called SEM algorithm. The SEM algorithm is an improvement of

the EM algorithm that incorporates a stochastic step (S-step) between the E- and M-steps of EM.

Starting from an initial parameter θ_0, an iteration of SEM algorithm consists of three steps.

6

E-step: The estimated probabilities that the i observation belongs to the jth component of the mixture, $w_{ij}^{(r)}, i = 1,\ldots,n, j = 1,\ldots,J$, are computed for the current value of θ as done in the standard EM.

S-step: A partition $P^{(r+1)} = (P_1^{(r+1)}, \cdots, P_J^{(r+1)})$ of $(y_1,x_1),\ldots,(y_n,x_n)$ is designed by assigning each observation at random to one of the mixture components according to the multinomial distribution with parameter $w_{ij}^{(r)}, i = 1,\ldots,n, j = 1,\ldots,J$, given by (5). If one of the P(r+1) is empty or has only one observation, it must be considered that the mixture has J−1 components instead of J and the estimation process begins with J − 1 components. Yet, in this case, it provides a bias towards uniform π_j parameters.

M-step: The ML estimate of θ is updated using the sub-samples $P_J^{(r+1)}$. It follows that on the M-step of the $(r+1)$th iteration, the parameter estimates $\pi_j^{(r+1)}$ are given by

$$\hat{\pi}_j^{(r+1)} = \frac{n_j}{n}, \qquad j = 1,\ldots,J \qquad (10)$$

where n_j is the total number of observations arising from component j and the maximization of

$$Q_2^* = \sum_{j=1}^{J} \sum_{\{i|z_{ij}=1\}} \log\left(f_j\left(y_i \mid x_i, \beta_j^{(r+1)} \right) \right) \qquad (11)$$

where $\{i \mid z_{ij} = 1\}$ is the set of observations arising from the jth mixture component, gives $\beta_j^{(r+1)}$.

7

4   Simulation study of algorithm performance

4.1 Design of the study

To investigate the statistical behaviour of the proposed methods in fitting Poisson mixture regression models, a simulation study was performed. The simulation is designed to evaluate the model performance considering the effects of sample sizes and the initialization of the algorithms as well as the configuration of the regression lines. The scope was limited to the study of two and three components. We used the freeware R to develop the simulation program.

Initial Conditions

Two different approaches for choosing initial values are compared in the study. In the first strategy, we use the true parameter values of the model by generating the observations as initial values in order to determine the performance of the algorithm in the best case. 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 (Celeux et al., 1993).

Stopping Rules

For the EM algorithm, iterations were stopped when the relative change in log-likelihood between two successive iterations were less than 10^{-20}. However, since SEM does not converge pointwise and it generates a Markov chain whose stationary distribution is more or less concentrated around the ML parameter estimate, we used as stopping rule for the SEM algorithm the total number of iterations required for convergence by the EM algorithm.

Number of Samples

For each type of simulated data set, we generated 500 samples of size *n*.

Data set

8

Each datum (y_i, x_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 $\lambda_{ij} = \exp(\beta_{j0} + \beta_{j1} x_i)$. Finally, we simulate the value $y_i \sim P(\lambda_{ij})$.

Measure of Algorithm Performance

In order to examine the performance of two algorithms, we report the Euclidean distance between estimated parameters and true parameter values.

Quality of the fit

In order to compare the quality of the fit of two algorithms, we report the root mean squared error of prediction (MRSEP):

$$MRSEP = \frac{1}{200} \sum_{m=1}^{200} RMSEP^{(m)}$$

where $RMSEP^{(m)}$ is the root mean squared error of prediction of the mth replication based on K-fold cross validation, which is given by:

$$RMSEP^{(m)} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} \left(\frac{y_i - \hat{\mu}_i}{\sqrt{V(\hat{\mu}_i)}} \right)^2}$$

where

$$\hat{\mu}_i = \sum_{j=1}^{J} \hat{\pi}_j \hat{\lambda}_{ij} \text{ and } V(\hat{\mu}_i) = \sum_{j=1}^{J} \hat{\pi}_j \hat{\lambda}_{ij} + \sum_{j=1}^{J} \hat{\pi}_j \hat{\lambda}^2{}_{ij} - \left(\sum_{j=1}^{J} \hat{\pi}_j \hat{\lambda}_{ij} \right)^2$$

For the K-fold cross validation, we have chosen $K = 5$ and $K = 10$ (Hastie el al., 2001, p.214).

9

4.2 Simulation results: two component mixture of Poisson regressions

For two-component models, samples of four different sizes $n(n = 50, 100, 500, 1000)$ were generated for each set of true parameter values (π, β) shown on Table 1 (Yang and Lai, 2005 and Leisch, 2004). For instance, we present in Figure 1 typical scatter plots for samples with size 500. Note that the cases considered correspond to varying degrees of overlapping. Case A3 has the highest overlapping and data from A1 show the lowest overlapping.

Figure 2 shows boxplots of the Euclidean distance between estimated and true parameters over the 500 replications using the EM and SEM algorithm for fitting two component mixtures of Poisson regression models.

Figure 2 shows that the three algorithms have practically the same behaviour. However, when the overlap is high (A3) EM outperforms SEM by producing estimates of the parameters that have smaller estimation error. As expected, estimation error decreases when the sample size increases.

The resulting values of *MRSEP* based on 10-fold cross validation, for each of the configurations of the true regression lines are plotted in Figure 3 and Figure 4. Similar results were obtained calculating *MRSEP* based on 5-fold cross validation. Figure 3 and Figure 4 show that, in generality, the SEM algorithm performs better than the EM algorithm.

4.3 Simulation results: three component mixture of Poisson regressions

For three-component models, samples of three different sizes $n(n = 100, 500, 1000)$ were generated for each set of true parameter values (π, β) shown on Table 2. Also, the cases considered correspond to varying degrees of overlapping. Cases B2, B3 and B4 have the highest overlapping and data from B1 show the lowest overlapping.

10

Figure 5 shows boxplots of the Euclidean distance between estimated and true parameters over the 500 replications using the EM and SEM algorithm for fitting three component mixtures of Poisson regression models. Figure 5 shows that EM outperforms SEM by producing estimates of the parameters that have lower estimation error, especially when the overlap is higher (B4). Also, as expected, estimation error tends to decrease as the sample size increases.

The resulting values of MRSEP based on 10-fold cross validation, for each of the configurations of the true regression lines are shown in Table 3 and Table 4. Similar results were obtained calculating MRSEP based on 5-fold cross validation. Table 3 and Table 4 show that, in generality, the SEM algorithm performs better than the EM algorithm.

5   Real Data Sets

We now compare the performance of the EM algorithm and the SEM algorithm for fitting Poisson mixture regression models in two real data sets.

5.1   Fabric faults

The "Fabric Faults" data set consists of 32 observations of number of faults in rolls of fabric of different length. The dataset is analysed using a finite mixture of Poisson regression models in Aitkin (1996). The response variable is the number of faults and the covariate is the length of role in meters. The data set can be loaded into R with the command data ("fabricfault", package="flexmix").

We fitted a component Poisson mixture regression model using EM algorithm and SEM algorithm where the logarithm of lengths is used as independent variable. The algorithms were initiated by random numbers (second strategy) and the stopping criterion was the same used in the simulation

11

study. For each algorithm, the optimal number of components was selected using the proposed procedure:

Step 1: Set j=2 and calculate the value of the MRSEP based on k-fold cross validation for a two component model. Let this value be denoted by MIN.

Step 2: Set j=j+1 and calculate the value of the MRSEP based on k-fold cross validation for a j component model.

Step 3: If the new value of the MRSEP is lower than MIN then set MIN equal to the new value of the MRSEP and go to Step 2, else deduce that the optimal number of components is j-1 and stop.

Table 5 presents the MRSEP based on 10-fold cross validation computed for each algorithm and the results show that the mixture with 2 components is selected. We can also observe that the SEM algorithm performs always better in fitting Poisson mixture regression model to the "Fabric Faults" data.

5.2 Patent

The patent data given in Wang et. al (1996) consist of 70 observations on patent applications, research-and-development (R&D) spending and sales in millions of dollar from pharmaceutical and biomedical companies in 1976 taken from the National Bureau of Economic Research R&D Master file. To model this data, Wang et. al (1996) used several covariates including logarithm of R&D spending and/or squared logarithm of R&D spending for different models. The data set can be loaded into R with the command data ("patent", package="flexmix").

We fitted a Poisson mixture regression model using EM algorithm and SEM algorithm where the logarithm of R&D spending is used as independent variable. The algorithms were initiated by

12

random numbers (second strategy), the stopping criterion was the same used in the simulation study and the optimal number of components was selected using the proposed procedure described in section 5.1.

Table 6 presents the MRSEP based on 10- fold cross validation computed for each algorithm and the results show that the mixture with 3 components is selected. We can also observe that the SEM algorithm performs always better in fitting Poisson mixture regression model to the patent data.

6   Conclusion

In this paper, we compare the performance of two algorithms to compute maximum likelihood estimates of a mixture Poisson regression models, the EM algorithm and the Stochastic EM algorithm (SEM).

The results of simulation show that the choice of approach depends essentially on the overlap of the regression lines. For some severely overlapping mixtures, the EM algorithm outperforms the SEM algorithm by producing estimates of the parameters that have smaller estimation error. However, simulation results indicated that the Stochastic EM Algorithm provides in general best estimates for those parameters in the sense of the best fit for the regression model.

In the real data case, we also show that the SEM algorithm resulted in model estimates that best fit the regression model. As we expected, the SEM algorithm and the EM algorithm can converge to the different estimates. EM convergence is very dependent upon the type of starting values and the stopping rule used, so the EM algorithm may converge to local maxima or saddle points. The SEM

13

algorithm exhibits more reliable convergence because the stochastic step enables this algorithm to escape from saddle points in the likelihood.

References

Aitkin, M. (1996). A general maximum likelihood analysis of overdispersion in generalized linear models. *Statistics and Computing* 6:251-262.

Celeux, G., Diebolt, J. (1985). The SEM algorithm: a probabilistic teacher algorithm derived from the EM algorithm for the mixture problem. *Computational Statistics Quarterly* 2:73-82.

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

Celeux, G., Chauveau, D., Diebolt, J. (1996). Stochastic versions of the EM algorithm: an experimental study in the mixture case. *J. Statistical Computation and Simulation* 55:287-314.

Dempster, A.P., Laird, N.M., Rubin, D.B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *J. of the Royal Statistical Soc. B* 39:1-38.

Dias, J., Wedel, M. (2004). An empirical comparison of EM, SEM and MCMC performance for problematic Gaussian mixture likelihoods. *Statistics and Computing* 14:323-332.

Diebolt, J. and Ip, E.H.S. (1996). *Stochastic EM: method and application.* In W.R. Gilks, S. Richardson, D.J. Speigelhalter (eds), Markov Chain Monte Carlo in Practice. London: Chapman & Hall.

Faria, S., Soromenho, G. (2010). Fitting mixtures of linear regressions. *J. Statistical Computation and Simulation* 80(2):201-225.

Fruhwirth-Schnatter, S. (2006). *Finite Mixture and Markov Switching Models*, Springer, Heidelberg.

14

Hastie, T., Tibshirani, R., Friedman, J.(2001). *Elements of Statistical Learning*, Springer, New York.

Leisch, F. (2004). FlexMix: A general framework for finite mixture models and latent class regression in R. *Journal of Statistical Software* 11(8):1-18.

McLachlan, G.J., Peel, D. (2000). *Finite Mixture Models*, Wiley, New York.

R Development Core Team (2008). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org

Teicher, H. (1960). On the mixture of distributions. *Annals of Mathematical Statistics* 31:55-73.

Wang, P., Puterman, M.L., Cockburn, I.M., Le, N. (1996). Mixed poisson regression models with covariate dependent rates. *Biometrics* 52(2):381-400.

Wedel, M., Desarbo, W.S., Bult, J.R., Ramaswamy, V. (1993). A latent class poisson regression model for heterogeneous count data . *Journal of Applied Econometrics* 8(4):397-411.

Yang, M.S., Lai, C.Y. (2005). Mixture poisson regression models for heterogeneous count data based on latent and fuzzy class analysis. *Soft Computing* 9:519-524.

15

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Table 1. True parameter values for the essays with a 2 component mixtures of Poisson regression

Cases	β_{10}	β_{11}	β_{20}	β_{21}	π_1
A1	4	-1	4	1	0.5
A2	4	-0.5	2	0.5	0.7
A3	2	-0.2	1	0.1	0.5
A4	3	-0.5	4	0.5	0.2

Table 2. True parameter values for the essays with a 3 component mixtures of Poisson regression

Cases	β_{10}	β_{11}	β_{20}	β_{21}	β_{30}	β_{31}	π_1	π_2
B1	3	-0.5	4	0.5	3	0.5	0.4	0.4
B2	6	-0.5	4	-0.5	2	0.5	0.3	0.3
B3	4	-0.5	4	0.5	2	0.8	0.3	0.2
B4	2.8	-2.9	2.6	0.4	3.6	0.2	0.3	0.2

Table 3. MRSEP by 10-fold cross-validation for 3 component models when the algorithms were

initiated by random numbers

		B 1						B 2					
		n= 100		n= 500		n= 1000		n= 100		n= 500		n= 1000	
π_1	π_2	EM	SEM	EM	SEM	EM	SEM	EM	SEM	EM	SEM	EM	SEM
0.2	0.2	1.091	1.057	1.050	1.027	0.903	0.889	1.379	1.307	1.121	1.134	1.047	1.036
0.2	0.3	1.067	1.052	1.024	1.015	0.900	0.889	1.554	1.450	1.160	1.127	0.996	0.989
0.2	0.4	1.050	1.040	1.000	0.999	0.912	0.896	1.211	1.179	1.081	1.042	0.933	0.940
0.2	0.5	1.045	1.020	1.012	0.997	0.899	0.894	1.150	1.117	1.151	1.064	0.920	0.930
0.2	0.6	1.052	1.030	1.032	0.995	0.904	0.907	1.123	1.064	1.060	0.971	0.922	0.868
0.3	0.2	1.045	1.027	1.022	1.010	0.905	0.893	1.631	1.583	1.228	1.229	0.960	0.951
0.3	0.3	1.040	1.026	1.020	1.003	0.902	0.899	1.566	1.513	1.187	1.185	0.990	0.988
0.3	0.4	1.039	1.038	0.990	0.995	0.910	0.902	1.208	1.188	1.119	1.062	0.977	0.950
0.3	0.5	1.047	1.040	1.026	0.997	0.896	0.900	1.089	1.073	1.018	1.005	0.920	0.959
0.4	0.2	1.058	1.061	1.012	1.010	0.914	0.900	1.219	1.217	1.176	1.214	0.960	0.950
0.4	0.3	1.045	1.033	1.020	0.990	0.899	0.899	1.294	1.296	1.208	1.120	0.964	0.976
0.4	0.4	1.002	1.014	0.989	0.980	0.896	0.898	1.107	1.124	1.210	1.152	0.975	1.002
0.5	0.2	1.025	1.057	1.017	0.992	0.899	0.883	1.547	1.235	1.269	1.190	1.026	1.040
0.5	0.3	1.001	1.005	1.002	1.000	0.907	0.911	1.257	1.426	1.194	1.151	0.959	0.953
0.6	0.2	1.149	1.065	1.017	1.010	0.911	0.913	1.461	1.365	1.304	1.269	0.961	0.980

Table 4. MRSEP by 10-fold cross-validation for 3 component models when the algorithms were

initiated by random numbers

|		B 3						B 4					
		n= 100		n= 500		n= 1000		n= 100		n= 500		n= 1000	
π_1	π_2	EM	SEM	EM	SEM	EM	SEM	EM	SEM	EM	SEM	EM	SEM
0.2	0.2	1.077	1.140	1.003	0.984	0.970	1.026	1.091	1.057	1.019	1.000	0.982	0.951
0.2	0.3	1.055	1.028	1.059	1.017	0.949	0.925	1.067	1.052	1.013	0.985	0.960	0.947
0.2	0.4	1.040	1.009	1.023	1.003	0.936	0.908	0.995	0.962	1.008	0.967	0.896	0.880
0.2	0.5	0.965	1.016	1.056	1.027	0.920	0.914	1.045	0.966	1.006	0.939	0.940	0.869
0.2	0.6	1.243	1.230	1.040	1.007	1.119	1.107	1.009	0.900	1.019	0.921	0.908	0.860
0.3	0.2	1.068	1.064	1.030	1.007	0.961	0.958	1.084	1.067	1.019	1.018	0.976	0.961
0.3	0.3	1.051	1.010	1.022	1.019	0.946	0.909	1.037	1.016	1.006	1.005	0.933	0.914
0.3	0.4	1.103	1.107	1.027	1.015	0.993	0.996	1.042	1.045	1.002	1.016	0.938	0.940
0.3	0.5	1.192	1.169	1.031	1.011	1.073	1.052	1.080	1.066	1.032	0.979	0.972	0.959
0.4	0.2	1.030	0.980	1.011	1.010	0.927	0.882	1.074	1.107	1.032	1.036	1.000	0.996
0.4	0.3	1.040	1.031	0.991	0.985	0.936	0.928	1.135	1.105	1.049	1.035	1.022	0.994
0.4	0.4	1.290	1.227	1.030	1.021	1.161	1.105	1.135	1.085	1.062	1.051	0.980	0.977
0.5	0.2	1.052	0.985	1.017	0.994	0.947	0.887	1.172	1.134	1.075	1.064	0.970	1.021
0.5	0.3	1.176	1.173	1.115	1.072	1.059	1.056	1.190	1.151	1.100	1.090	1.050	1.036
0.6	0.2	1.287	1.173	1.173	1.089	1.159	1.056	1.067	1.116	1.066	1.088	1.070	1.034

Table 5. MRSEP based on 10-fold cross validation for "fabric faults" dataset

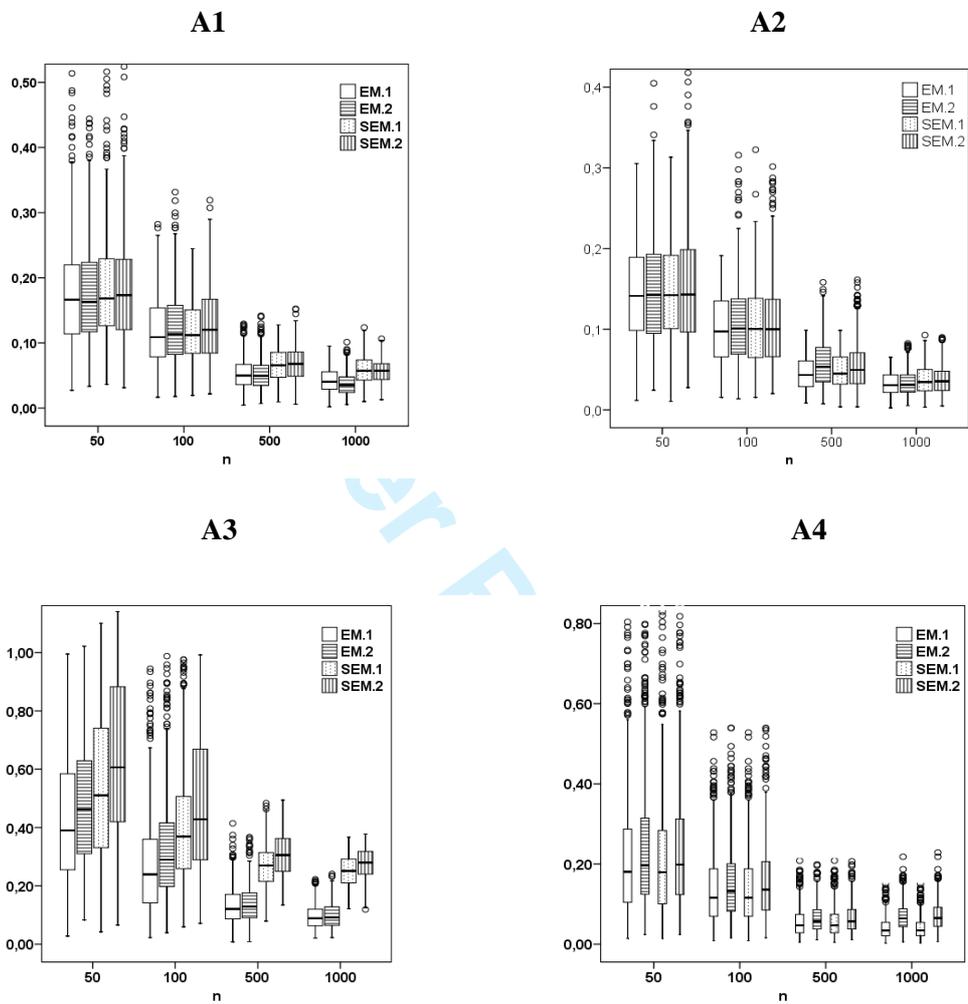|	EM Algorithm	SEM Algorithm
2 components	1.4503	1.020
3 components	1.4660	1.271

2

Table 6. MRSEP based on 10-fold cross validation for "Patent'' dataset

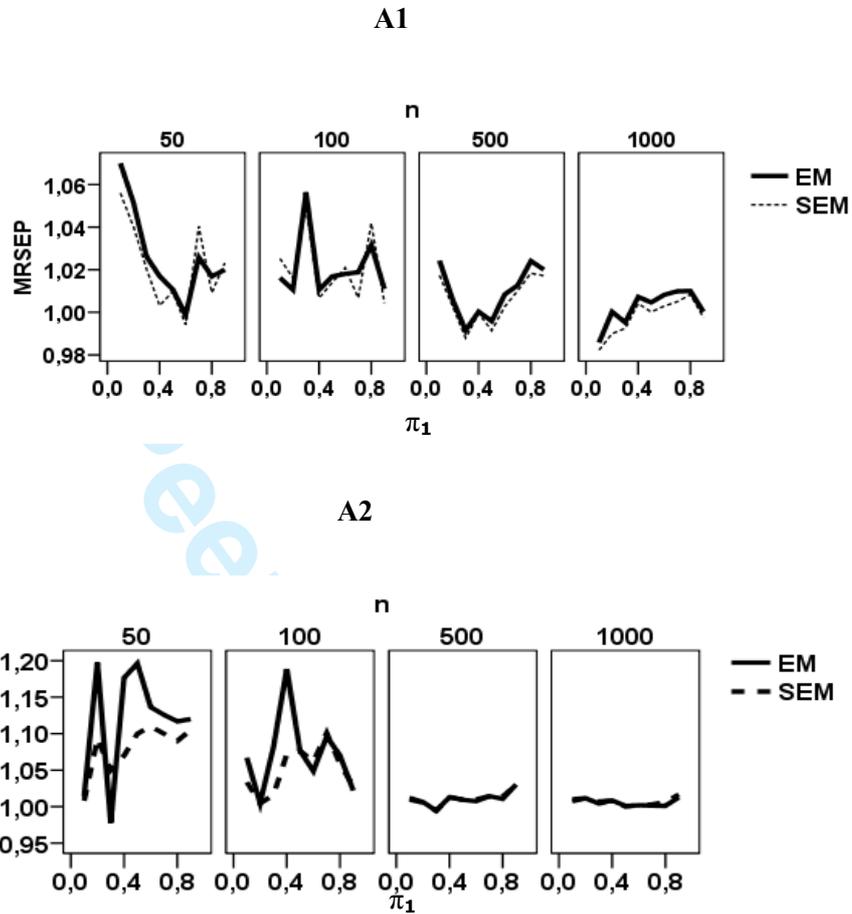|	EM Algorithm	SEM Algorithm
2 components	1.1072	0.9453
3components	1.0213	0.9156
4 components	1.0558	0.9203

3
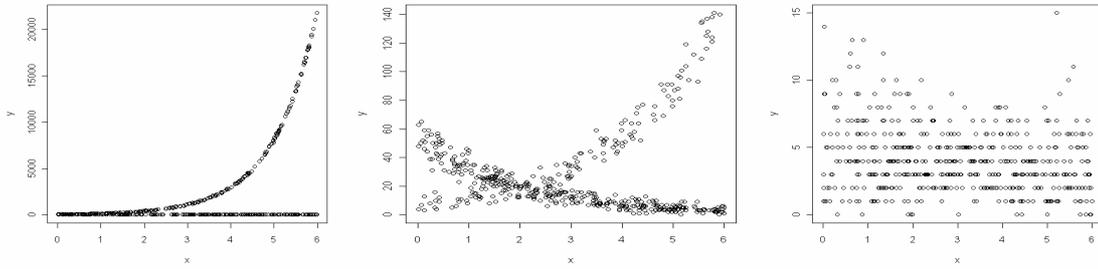
Figure 1. Scatter plot of samples from 2 component models with n = 500 (A1, A2 and A3)
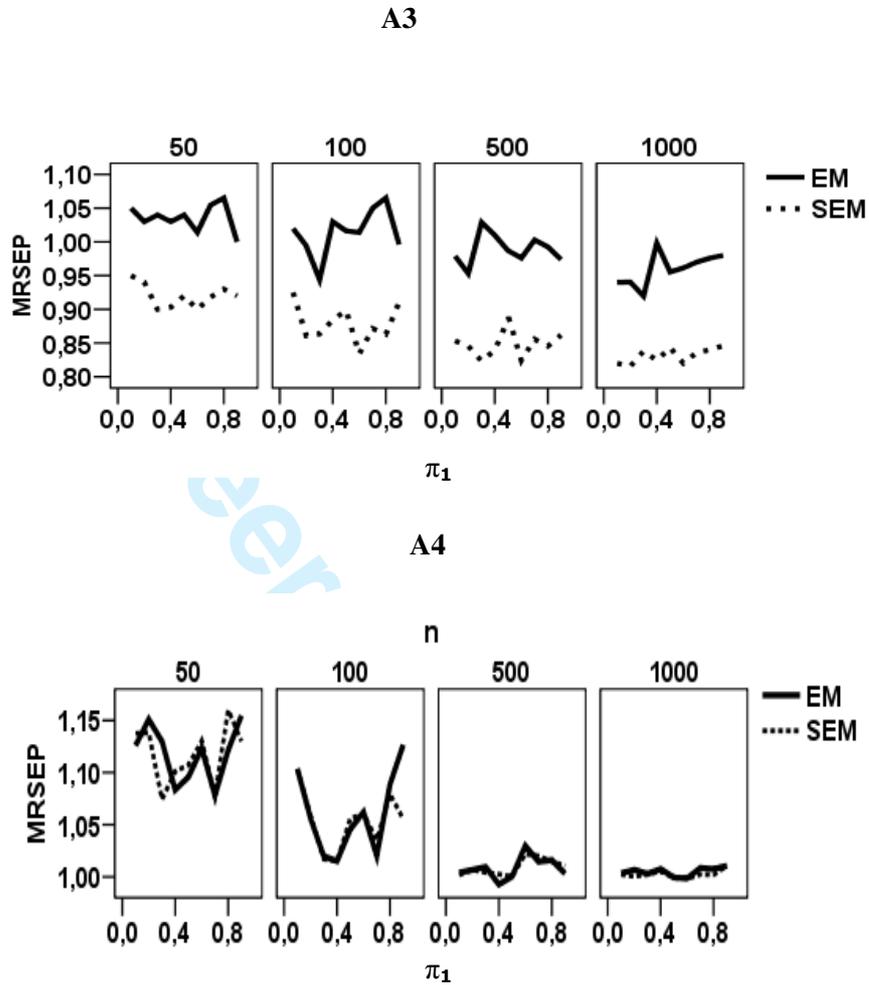
Figure 2. Distance between estimated and true parameter values for two-component Poisson mixture

regression models. (EM.1 and SEM.1 - the algorithms are initiated with the true parameter values;

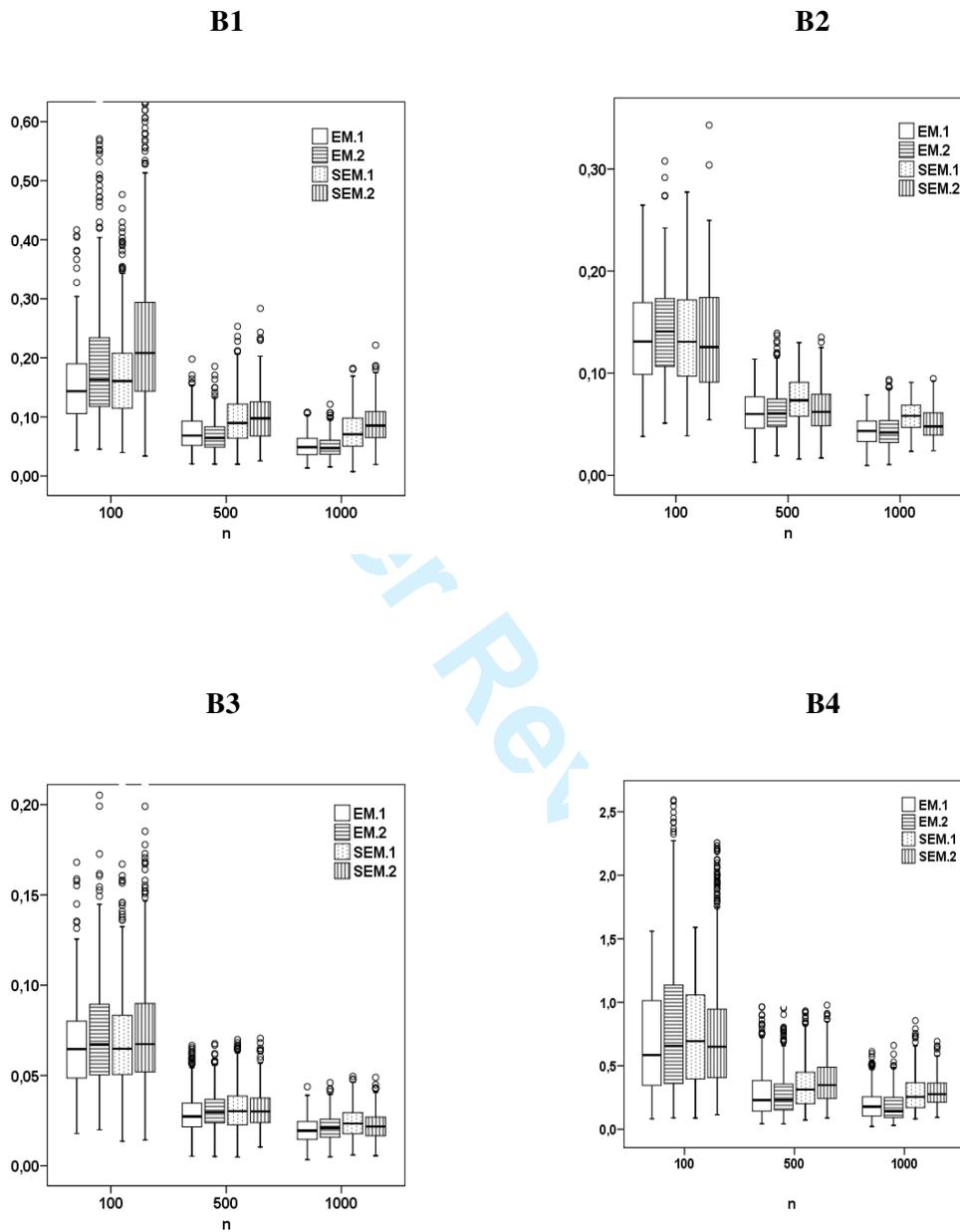EM.2 and SEM.2 – the algorithms are initiated by random numbers)

1

A1

A2

Figure 3. MRSEP by 10-fold cross-validation for 2 component models when the algorithms were initiated by random numbers.

1

A3

A4

Figure 4. MRSEP by 10-fold cross-validation for 2 component models when the algorithms were initiated by random numbers.

1

Figure 5. Distance between estimated and true parameter values for three-component Poisson mixture regression models. (EM.1 and SEM.1 - the algorithms are initiated with the true parameter values; EM.2 and SEM.2 – the algorithms are initiated by random numbers)

1

Communication in Statistics – Simulation and Computation

LSSP 2010 - 0452

Comparison of EM and SEM Algorithms in Poisson Regression Models: a
simulation study

Susana Faria and Gilda Soromenho

Comments to the issues raised by the referees:

1. Page 3, 5 lines from below: Text should read "… identifiability.
   That is, two sets of parameters do not yield the same…"

We rewrite the text in page 3. The changes we make in the manuscript
are in coloured text.

2. Page 4: First paragraph of Section 3: This material needs to be
very carefully rewritten. Most of the statements made here are just
not true. It is not clear what is intended. Does the SEM algorithm
always converge? Does it converge if MLE's and/or EM estimates can not
be   obtained?

Page 4. Line 3, sentence beginning "Given a set of independent…" I
don't believe that this statement is true at all. The exact conditions
for existence of maximum likelihood estimates is a very difficult
statement to make. Perhaps you can say that MLE's can sometimes be
estimated from this likelihood function provided such estimates exist.

We rewrite the text in Section 3. The changes we make in the
manuscript are in coloured text.

3. Page 4, Equation (4): Conditional on the lambda parameters, only
the pi's are estimated. The real condition is y given x, but not
lambda   here.

We rewrite the equation (4) and also equation (1), (2) and (3).

4. Page 5, equation (8): This is confusing. I thought that the
parameters we were estimating were the beta's, not the lambda's.

We rewrite the equation (8) and also equation (5) and (11).

5. Page 7, top paragraph: Please make a comment that this provides a
bias   towards   uniform   pi   parameters.

We make a comment. The changes we make in the manuscript are in
coloured text.

6. Page 9: Refer to page number in books such as Hastie, 2001.

We refer the page number.

7. Page 12: End Section 5.1 by explaining which method is better. What
do you conclude from this example? The two methods converge to
different   estimates.   Which   is   do   you   prefer?

We eliminate the end of Section 5.1 and 5.2 (and Figure 6 and Figure 7). The results in Table 5 and Table 6 show that the SEM algorithm perform always better in fitting Poisson mixture regression model to the data.

8. References: These are in different styles. Some are all capitals in titles, others are not. Please refer to the style requirements for this journal.

We rewrite some references. The changes we make in the manuscript are in coloured text.