Implementation of EM-Algorithm in Normal Mixture Model

Submitted by Sagnika Chakraborty and Shwetab Sinha

November 21, 2010

Abstract

We use EM-Algorithm to give the estimate of the parameters of a Normal Mixture Model. We have generated observation from different Normal Mixture Model and observe the performance. After that we test the number of populations in the mixtute model by Likelihood Ratio Test.

1 Introduction:

One of the classical formulation of the discriminant analysis or the statistical pattern recognisation problem involves a mixture of two p- dimensional normal distributions with a common covariance matrix. Here our objective is to find the Maxmum Likelihood Estimate of the parmaters of k Normal Mixture model. Our methodology will be based on EM Algorithm.

2 Basic Set-up:

Our data consists of n observations $x_1, x_2, ...x_n$, from the mixed density $f(x)$. Suppose we have K normal populations with density $f_k(x)$, where $f_k(x) \sim N(\mu_k, \sigma_k^2)$, $k = 1(1)K$. Then the mixed density $f(x)$ is defined as $f(x) = \sum_{k=1}^{K} \alpha_k f_k(x)$, where α_k is the probability that an observation came from the k-th population. $k = 1(1)K$. We are interested to find the MLE's of α_k's, μ_k's and σ_k's with the help of EM Algorithm.

2.1 Why EM Algorithm?

An expectation-maximization (EM) algorithm is a method for finding maximum likelihood estimates of parameters in statistical models, where the model depends on unobserved latent variables. EM is an iterative method which alternates between performing an expectation (E) step, which computes the expectation of the log-likelihood evaluated using the current estimate for the latent variables, and a maximization (M) step, which computes parameters maximizing the expected log-likelihood found on the E step.

The EM Algorithm is not really an algorithm. Rather it is a recipe to create algorithm for specific MLE problems. EM algorithm can be applicable when we can view the problem as a missing data problem.

Here the complete data is $(x_1, \pi_1), (x_2, \pi_2), ..., (x_n, \pi_n)$, where $\pi_i = k$ if x_i came from f_k. $k = 1(1)K$, $i = 1(1)n$. π_i's are the missing data which is not available to us. Hence the EM Algorithm can be applied.

2.2 EM-Algorithm to This Problem:

Here we build the loglikelihood of the complete data and then compute the expectation of that. For that we need to introduce some new variables and notations. Let $Z_i \sim mult(1, (\alpha_1, \alpha_2, ..., \alpha_k))$, $i = 1(1)n$. Now, $\pi_i = k \Leftrightarrow Z_i = e_k$, which indicates that the i-th obsarvation came from the k-th population. One estimate of α_k is given by $\widehat{\alpha_k} = \frac{\#\{\pi_i = k\}}{n}$, $k = 1(1)K$, but we can't use this as we don't know which observation belons to which population. So, here the parameter to be estimated is given by $\phi = \{\underline{\alpha}, \underline{\mu}, \underline{\sigma}\}$.

- *Expectation-Step:*

  The loglikelihood function of $x|\phi$ ia as follows:

  $$logf(x|\phi) = \sum_{i=1}^{n} log\alpha_{\pi_i} + \sum_{i=1}^{n} logf(x_i|\mu_{\pi_i}, \sigma^2_{\pi_i})$$

  Eventually we need $Q(\phi'|\phi) = E[logf(x|\phi)]$

  $$\therefore Q(\phi'|\phi) = \sum_{i=1}^{n} \frac{f_k(x_i)\alpha_k}{\sum_{l=1}^{K} f_k(x_i)\alpha_l} + \sum_{i=1}^{n}\sum_{k=1}^{K} [\frac{\alpha_k f_k(x_i)}{\sum_l \alpha_l f_l(x_i)}] logf_k(x_i)$$

  $$= \widehat{N_k} + \sum_{i=1}^{n}\sum_{k=1}^{K} w_{ki} logf_k(x_i) \text{ (say) } ...(*)$$

- *Maximization-Step:*

  According to EM Algorithm we need to maximize $Q(\phi'|\phi)$ at every stage. Maximisation step gives $\phi^{(t+1)}$=argmax$_\phi$ $Q(\phi'|\phi^{(t)})$.
  At stage $(t+1)$,

  $$\widehat{\alpha_k}^{(t+1)} = \frac{\widehat{Nk}^{(t)}}{\sum_{k=1}^{K} \widehat{Nk}^{(t)}}, \widehat{\mu_k}^{(t+1)} = \frac{\sum_{i=1}^{n} w_{ki}^{(t)} x_i}{\sum_{i=1}^{n} w_{ki}^{(t)}} \text{ \& } \widehat{\sigma_k^2}^{(t+1)} = \frac{\sum_{i=1}^{n} w_{ki}^{(t)} (x_i - \mu_k^{(t)})^2}{\sum_{i=1}^{n} w_{ki}^{(t)}}$$

  We have to go on iterate this process untill convergence.

3 Analysis:

For our compuational purpose we simulate data from three univariate normal distributions. Also we fix different sample sizes and then try to observe the performance. We consider three cases viz

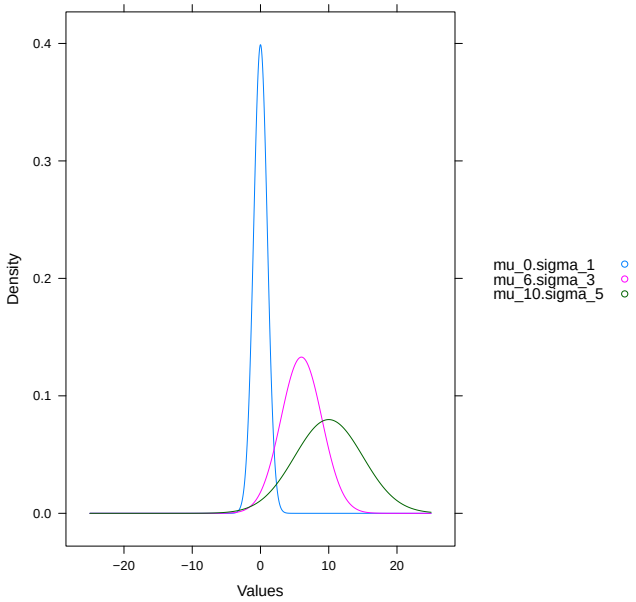
- Case I. All μ's and σ's are different.
- Case II. All μ's are same, but σ's are different.
- Case III. All μ's are different, but σ's are same.

  We vary sample sizes as $(20, 50, 30)'; (100, 250, 150)'; (200, 500, 300)'$.

3.1 Case I - All μ, σ Are Different:

We have generated obsrvation from $N(0, 1^2)$, $N(6, 3^2)$, $N(10, 5^2)$ with different sample sizes. The densities are as follows

Figure 1: The plot showing the three different densities having all μ and σ different



The estimated values are tabulated as follows

Table 1: Table showing the MLE's of α, μ, σ for different sample sizes

Parameter	$\alpha = (0.2, 0.5, 0.3)'$	$\mu = (0, 6, 10)'$	$\sigma = (1, 3, 5)'$
$n = \begin{pmatrix} 20 \\ 50 \\ 30 \end{pmatrix}$	$\begin{pmatrix} 0.235935 \\ 0.533862 \\ 0.230203 \end{pmatrix}$	$\begin{pmatrix} 0.124039 \\ 6.58757 \\ 10.9902 \end{pmatrix}$	$\begin{pmatrix} 1.42672 \\ 2.29191 \\ 4.64591 \end{pmatrix}$
$n = \begin{pmatrix} 100 \\ 250 \\ 150 \end{pmatrix}$	$\begin{pmatrix} 0.227059 \\ 0.418019 \\ 0.354922 \end{pmatrix}$	$\begin{pmatrix} 0.053341 \\ 6.20723 \\ 9.60652 \end{pmatrix}$	$\begin{pmatrix} 1.07785 \\ 2.49345 \\ 5.18161 \end{pmatrix}$
$n = \begin{pmatrix} 200 \\ 500 \\ 300 \end{pmatrix}$	$\begin{pmatrix} 0.20653 \\ 0.4686 \\ 0.32486 \end{pmatrix}$	$\begin{pmatrix} 0.2389 \\ 6.23578 \\ 9.6997 \end{pmatrix}$	$\begin{pmatrix} 1.11081 \\ 3.02592 \\ 4.9332 \end{pmatrix}$

We simulate the observation for 10 times for the sample size $n = (200, 500, 300)'$ and find the MLE's of the parameters each time. After that we plot them to see how the MLE's are deviated from the true parameter value.

Figure 2: Plot showing the MLE's for α for Case I

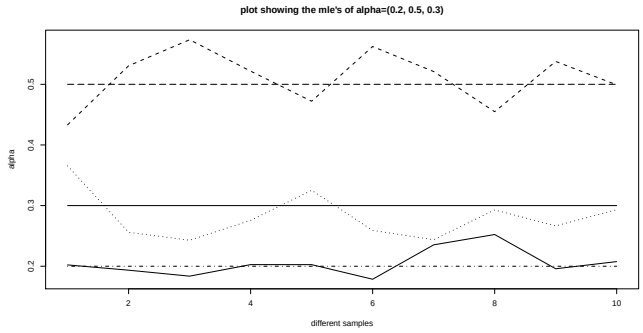


Figure 3: Plot showing the MLE's for μ for Case I

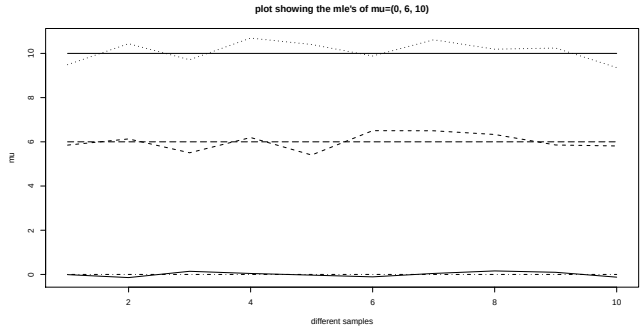
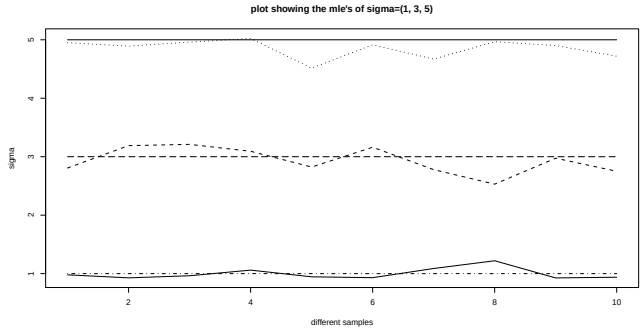


Figure 4: Plot showing the MLE's for σ for Case I



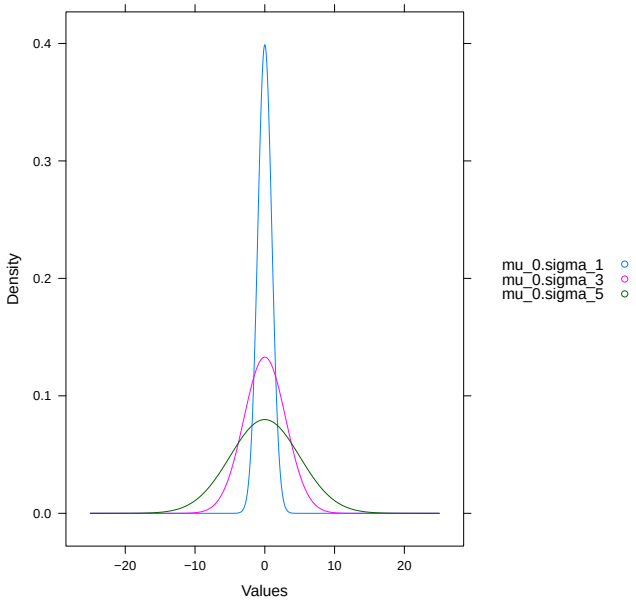
Remark:

- MLE's are found to be quite near to the original parameter values.

3.2 Case II - All μ's are same, σ's are different:

We have generated obsrvation from $N(0, 1^2)$, $N(0, 3^2)$, $N(0, 5^2)$ with different sample sizes. The densities are as follows

Figure 5: The plot showing the three different densities having same μ and different σ



The estimated values are tabulated as follows

Table2: Table showing the MLE's of α, μ, σ for different sample sizes

Parameter	$\alpha = (0.2, 0.5, 0.3)'$	$\mu = (0, 0, 0)'$	$\sigma = (1, 3, 5)'$
$n = \begin{pmatrix} 20 \\ 50 \\ 30 \end{pmatrix}$	$\begin{pmatrix} 0.204165 \\ 0.454283 \\ 0.341552 \end{pmatrix}$	$\begin{pmatrix} 0.671416 \\ -0.487204 \\ 0.770718 \end{pmatrix}$	$\begin{pmatrix} 0.993187 \\ 2.71302 \\ 5.50172 \end{pmatrix}$
$n = \begin{pmatrix} 100 \\ 250 \\ 150 \end{pmatrix}$	$\begin{pmatrix} 0.213757 \\ 0.366583 \\ 0.41966 \end{pmatrix}$	$\begin{pmatrix} 0.0259538 \\ 0.144978 \\ -0.103522 \end{pmatrix}$	$\begin{pmatrix} 1.66999 \\ 3.09095 \\ 4.65801 \end{pmatrix}$
$n = \begin{pmatrix} 200 \\ 500 \\ 300 \end{pmatrix}$	$\begin{pmatrix} 0.175921 \\ 0.373311 \\ 0.450768 \end{pmatrix}$	$\begin{pmatrix} 0.102794 \\ -0.609069 \\ 0.331768 \end{pmatrix}$	$\begin{pmatrix} 0.929219 \\ 2.88963 \\ 4.66718 \end{pmatrix}$

We simulate the observation for 10 times for the sample size $n = (200, 500, 300)'$ and find the MLE's of the parameters each time. After that we plot them to see how the MLE's are deviated from the true parameter value.

Figure 6: Plot showing the MLE's for α for Case II

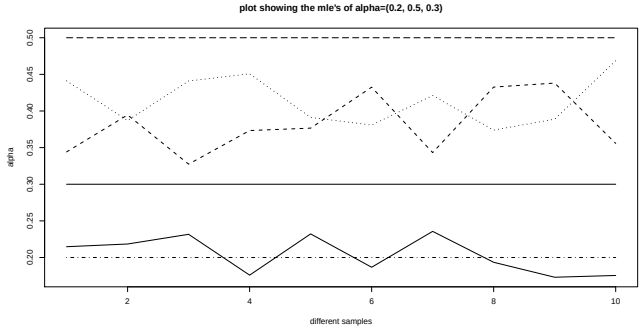


Figure 7: Plot showing the MLE's for μ for Case II

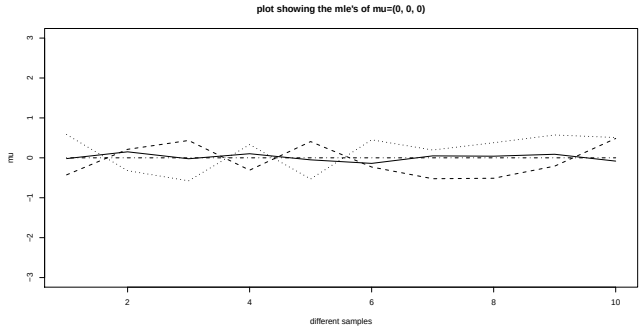
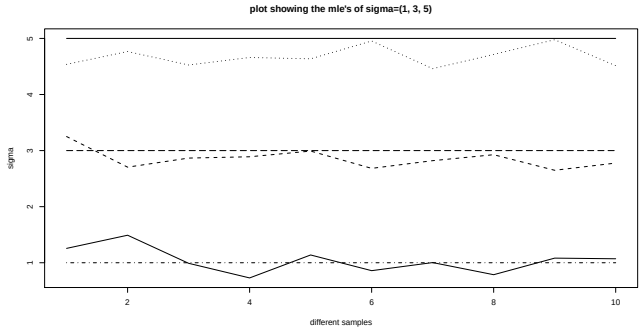


Figure 8: Plot showing the MLE's for σ for Case II



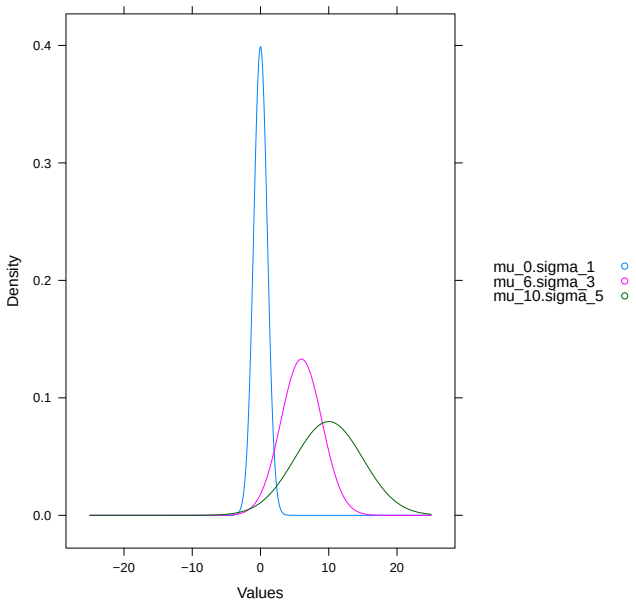
Remark:

- In this case as all the population has the same mean but different variances, the EM-Algoruthm fails to discriminant the proprtion of observations coming from a particular population properly. Infact, it over estimates α corresponding to the population with highest variance although in our data this population proprtion in the mixture model is not the highest. However, it provides the MLE's of μ's and σ's quite satisfactorily.

3.3 Case III - All μ's are different, σ are same.

We have generated obsrvation from $N(0, 3^2)$, $N(6, 3^2)$, $N(10, 3^2)$ with different sample sizes. The densities are as follows

Figure 9: The plot showing the three different densities having different μ and same σ



The estimated values are tabulated as follows

Table 3: Table showing the MLE's of α, μ, σ for different sample sizes for Case III

Parameter	$\alpha = (0.2, 0.5, 0.3)'$	$\mu = (0, 6, 10)'$	$\sigma = (3, 3, 3)'$
$n = \begin{pmatrix} 20 \\ 50 \\ 30 \end{pmatrix}$	$\begin{pmatrix} 0.176192 \\ 0.582639 \\ 0.241169 \end{pmatrix}$	$\begin{pmatrix} -0.477835 \\ 3.67089 \\ 8.18214 \end{pmatrix}$	$\begin{pmatrix} 3.62417 \\ 3.84471 \\ 3.17822 \end{pmatrix}$
$n = \begin{pmatrix} 100 \\ 250 \\ 150 \end{pmatrix}$	$\begin{pmatrix} 0.1655347 \\ 0.572136 \\ 0.262329 \end{pmatrix}$	$\begin{pmatrix} -1.354553 \\ 4.3705 \\ 8.37343 \end{pmatrix}$	$\begin{pmatrix} 2.49496 \\ 4.02304 \\ 3.81282 \end{pmatrix}$
$n = \begin{pmatrix} 200 \\ 500 \\ 300 \end{pmatrix}$	$\begin{pmatrix} 0.208965 \\ 0.515379 \\ 0.307747 \end{pmatrix}$	$\begin{pmatrix} 0.695545 \\ 4.86029 \\ 9.20074 \end{pmatrix}$	$\begin{pmatrix} 3.53093 \\ 2.69219 \\ 3.10883 \end{pmatrix}$

We simulate the observation for 10 times for the sample size $n = (200, 500, 300)'$ and find the MLE's of the parameters each time. After that we plot them to see how the MLE's are deviated from the true parameter values.

Figure 10: Plot showing the MLE's for α for Case III

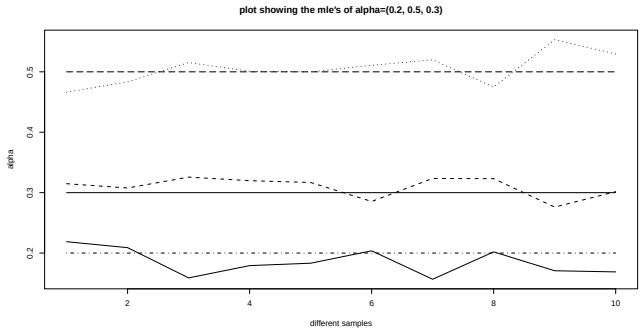


Figure 11: Plot showing the MLE's for μ for Case III

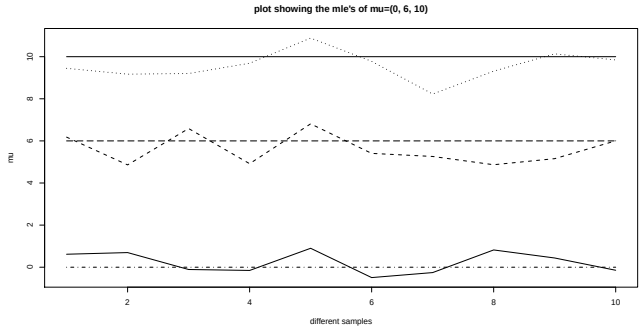
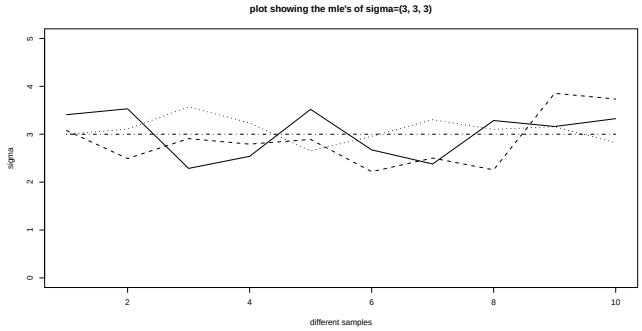


Figure 12: Plot showing the MLE's for σ for Case III



Remark:

- MLE's are found to be quite near to the original parameter values except for σ.

4 Test For Number of Population:

Now we want to test for the number of population. For our computation we have taken three normal populations having all μ's, σ's different [$20\%N(0, 1^2)$, $50\%N(6, 3^2)$, $30\%N(10, 5^2)$] and for the sample size $n = (20, 50, 30)'$ We the true number of population is 3. Here, our objective is to test H_{01} : K = 3 vs. H_{11} : K = 2 & H_{02} : K = 3 vs. H_{12} : K = 4. K : no of population.

4.1 Likelihood Ratio Test:

Neyman and Pearson have suggested a simple method of test construction which is closely related to the maximum likelihood method of estimation. Let $L(\phi|x_1, x_2, ..., x_n)$ be the likelihood. We want to test $H_0 : \phi \in \Theta_0$ vs $H_1 : \phi \in \Theta_1$. Define $\lambda(x_1, x_2, ..., x_n) = \frac{sup_{\theta\in\Theta_0} L(\phi|x_1,x_2,...,x_n)}{sup_{\theta\in\Theta_0\cup\Theta_1} L(\phi|x_1,x_2,...,x_n)}$ Clearly $0 \leq \lambda \leq 1$. It can be observed that we reject the null hypothesis for very small value of λ. Inparticular, for $\lambda = 1$ we surely accept the null hypotheis.

Here, $L(\phi|x_1, x_2, ..., x_n) = \prod_{i=1}^{n} \sum_{k=1}^{K} \alpha_k f_k(x_i)$.

For $\{K = 3\}$ and for $\{K = 3\}U\{K = 4\}$ the numerator and the denominator in λ are being maximised for the same value of ϕ. So $\lambda = 1$. And we get same result for H_{02} vs H_{12}. Hence, for both the cases we accept the null hypotheis. i.e. the number of population in the mixture model is 3.

5 Conclusion:

To find the Maximum Likelihood Estiamate of the parameters of an Normla mixture model using EM-Algorithm gives quite satisfactory result. When all the μ and σ are different and when all μ different σ same then we have better result than when all μ are same σ different.

6 Acknowledgement:

- Dr. Deepayan Sarkar
- Statistical Computing by Kundu and Das
- http://www.jstor.org/stable/2030064