THE EM ALGORITHM APPLIED TO THE SUM
OF EXPONENTIALS

Lucas Monteiro CHAVES
Devanil Jaques de SOUZA

ABSTRACT: The EM algorithm is used to determine the maximum likelihood estimates for the parameters of the sum of two exponentials with different scale factors. Simulations are developed using the R environment.

KEYWORDS: EM algorithm; expectation; maximization; exponential distribution; gamma distribution.

1 Introduction

The EM algorithm is a general method for finding the maximum-likelihood estimate of the parameters of an underlying distribution from a given data set when the data is incomplete or has missing values. One of its main applications is when the likelihood function is analytically intractable. It can be described as Silva (2002).

Let \( f_{XY}(x, y | \theta) \) be the distribution of the complete data and

\[
L^*(\theta | y) = \ln f_y(y | \theta) = \ln \int f_{X|Y}(x | y) f_Y(y) dx = \ln \int f_{XY}(x, y | \theta) dx
\]

the usual log-likelihood function of an observed sample. Verify that

\[
\ln \left( \frac{f_{XY}(x, y)}{f_{XY}(x | y)} \right) = \ln \left( \frac{f_{XY}(x, y)}{f_{XY}(x, y) / f_Y(y)} \right) = \ln [f_Y(y)]
\]

\[
= \ln [f(y | \theta)] = \ln \left[ \int f(x, y | \theta) dx \right]
\]

Plugging (2) into (1),
\[ L^*(\theta | y) = \ln\left[ f(x, y | \theta) \right] - \ln\left[ f(x | y, \theta) \right] \]  

(3)

Assuming a supposed \( \theta^{(0)} \) and an observation \( y \)

\[ L^*(\theta | y) = \int L^*(\theta | y) f(x | y, \theta^{(0)}) \, dx = E_{\{X,Y | \theta^{(0)}\}} [L^*(\theta | y)] \]  

(4)

Plugging (3) into (4) and denoting by

\[
Q(\theta, \theta^{(0)}) = E_{\{X,Y | \theta^{(0)}\}} [\ln f(x, y | \theta)]  \\
H(\theta, \theta^{(0)}) = E_{\{X,Y | \theta^{(0)}\}} [\ln f(x | y, \theta)]  
\]

we can write

\[ L^*(\theta | y) = E_{\{X,Y | \theta^{(0)}\}} [\ln f(x, y | \theta)] - E_{\{X,Y | \theta^{(0)}\}} [\ln f(x | y, \theta)] \\ = Q(\theta, \theta^{(0)}) - H(\theta, \theta^{(0)}) \]  

(5)

Now we want to maximize the log-likelihood of the incomplete data, \( L^*(\theta | y) \). Because it is supposed to be easier, we prefer to work with the log-likelihood of the complete data, \( L^*(\theta | x, y) = \ln f_{X,Y}(x, y | \theta) \). Equation (5) seems nice to help us on doing this, except for the second term in the right hand side. But, considering that \( \ln(.) \) is a concave function Jensen’s inequality may be used:

\[
H(\theta, \theta^{(k)}) - H(\theta^{(k)}, \theta^{(k)}) = \int \ln \left[ f(x | y; \theta) \right] f(x | y; \theta^{(k)}) \, dx \\ - \int \ln \left[ f(x | y; \theta^{(k)}) \right] f(x | y; \theta^{(k)}) \, dx \\ = \int \left[ \ln \left( \frac{f(x | y; \theta)}{f(x | y; \theta^{(k)})} \right) \right] f(x | y; \theta^{(k)}) \, dx 
\]

(6)

\[
= E_{\{X,Y | \theta^{(k)}\}} \left[ \ln \left( \frac{f(x | y; \theta)}{f(x | y; \theta^{(k)})} \right) \right] \leq \ln E_{\{X,Y | \theta^{(k)}\}} \left[ \frac{f(x | y; \theta)}{f(x | y; \theta^{(k)})} \right] \\ = \ln \int \frac{f(x | y; \theta)}{f(x | y; \theta^{(k)})} f(x | y; \theta^{(k)}) \, dx = \ln \int f(x | y; \theta) \, dx = 0
\]

Then \( H(\theta, \theta^{(k)}) \leq H(\theta^{(k)}, \theta^{(k)}) \) \( \forall (\theta, \theta^{(k)}) \).
This result is used in what follows.

The EM algorithm consists in constructing a sequenced \( \{ \theta^{(k)} \} \) such that

\[
Q(\theta^{(k+1)}, \theta^{(k)}) = \max_{\theta \in \Theta} \left[ Q(\theta, \theta^{(k)}) \right] = \max_{\theta \in \Theta} \left\{ E_{\theta^{(k)}} [f(x, y | \theta)] \right\}
\]

Dempster et al. (1977) show that this sequence converges to a critical point of \( L^*(\theta | y) \).

The most usual example of application for the EM algorithm is, probably, the one given by Rao (see Dempster, 1977). Although extremely instructive, it really doesn’t require using EM, since maximum likelihood estimates for the parameters can easily be computed. Another common example, in continuous case, is the mixture of normal variables. But this example is too much involved. One of the purposes of this work is to present a pedagogical example of application of the EM algorithm in a situation where the maximum likelihood procedures are not feasible. Besides this, as some references induces the reader the conclusion that the sum of exponential, with different parameters, has gamma distribution (Feller, 1970 - pg 47), another purpose is to clear that the sum of two independent gamma variables, with different scale parameters, has first form McKay distribution. This result may be found, without proof, in Johnson et al. 1994 (ch. 12 - 4.4). Holm et al. (2004) sketch the demonstration and comment that they couldn’t find the origin of it’s first presentation. McKay distributions involve Bessel functions and the corresponding likelihood functions are algebraically intractable.

In section two we compute the distribution of the sum of two exponentials with different parameters. In section three we develop the results needed for the use of EM algorithm. In section four we show some simulations and in section 5 we conclude.

2 The sum of two exponentials

Let \( X \sim \text{Exp}(\alpha) \) and \( Y \sim \text{Exp}(\lambda) \) be independent random variables.

The distribution of \( Z = X + Y \) is given by the convolution:

\[
f_z(z) = \int_0^z \alpha e^{-\alpha x} \lambda e^{-\lambda(z-x)} \, dx = \frac{\alpha \lambda}{\alpha - \lambda} \left( e^{-\lambda z} - e^{-\alpha z} \right)
\]

The maximum likelihood estimates for \( \alpha \) e \( \lambda \) are given by

\[
L^*(\alpha, \lambda | z) = \ln(\alpha) + \ln(\lambda) - \ln(\alpha - \lambda) + \ln\left( e^{-\lambda z} - e^{-\alpha z} \right)
\]
and this resulting nonlinear system is algebraically intractable.

3 The EM algorithm

Consider X and Y as the unobservable random variables and Z the observable one. Let \( \theta = (\alpha, \lambda) \) be the parameter vector and suppose a sample \( Z = (Z_1, \ldots, Z_n) \), where \( Z_i = X_i + Y_i \). The complete data are the pairs \((X_1, Y_1), \ldots, (X_n, Y_n)\) and the steps E (expectation) and M (maximization) of the algorithm are obtained with the following procedure:

\[
Q(\theta, \theta_0) = \int_0^z \cdots \int_0^z \left[ n \ln(\alpha) + n \ln(\lambda) - \alpha \sum t_i - \lambda \sum (z_i - t_i) \right] dt_1 \cdots dt_n
\]

\[
\frac{(\lambda - \alpha)^n e^{-\alpha \sum t_i} e^{-\lambda \sum (z_i - t_i)}}{(\sqrt{2})^n \prod (e^{-\alpha t_i} - e^{-\lambda t_i})} \left( \sqrt{2} \right)^n dt_1 \cdots dt_n
\]

\[
= \left[ n \ln(\alpha) + n \ln(\lambda) - \lambda \sum z_i \right] \int_0^z \cdots \int_0^z \frac{(\lambda - \alpha)^n e^{-\alpha \sum t_i} e^{-\lambda \sum (z_i - t_i)}}{\prod (e^{-\alpha t_i} - e^{-\lambda t_i})} dt_1 \cdots dt_n
\]

\[
+ \left[ \lambda - \alpha \right] \sum t_i \frac{(\lambda - \alpha)^n e^{-\alpha \sum t_i} e^{-\lambda \sum (z_i - t_i)}}{\prod (e^{-\alpha t_i} - e^{-\lambda t_i})} dt_1 \cdots dt_n
\]

As
\[
\int_0^z \frac{(\lambda_0 - \alpha_0) e^{-\alpha_1 t} e^{-\lambda t(z-t)}}{e^{\alpha_0 z} - e^{-\lambda z}} \, dt = \\
= \frac{1}{(e^{\alpha_0 z} - e^{-\lambda z})} e^{-\lambda z} \int_0^z (\alpha_0 - \lambda_0) e^{-(\alpha_0 - \lambda_0) t} \, dt \\
= \frac{e^{-\lambda z}}{(e^{\alpha_0 z} - e^{-\lambda z})} \left[ e^{-(\alpha_0 - \lambda_0) z} - 1 \right] = 1
\]

Using this result \(n\) times in (7),

\[
Q(\theta, \theta') = n Ln(\alpha) + n Ln(\lambda) - \lambda \sum z_i + \\
+ \int_0^z \int_0^z \cdots \int_0^z [(\lambda - \alpha) \sum \alpha_j e^{-\alpha_j t} \sum e^{-\lambda_j t} (z-t)] \left( \frac{\sqrt{2}}{\prod (e^{\alpha_j z} - e^{-\lambda_j z})} \right)^n dt_1 \cdots dt_n \\
= n Ln(\alpha) + n Ln(\lambda) - \lambda \sum z_i + \\
+ (\lambda - \alpha) \sum \int_0^z \int_0^z \cdots \int_0^z \left( \frac{\lambda_0 - \alpha_0 e^{-\alpha_0 t} \sum e^{-\lambda_j t} (z-t)}{\prod (e^{\alpha_0 z} - e^{-\lambda_j z})} \right)^n dt_1 \cdots dt_n
\]

Using the previous identity again:

\[
Q(\theta, \theta') = n Ln(\alpha) + n Ln(\lambda) - \lambda \sum z_i + \\
+ (\lambda - \alpha) \sum \int_0^z \left( \frac{\lambda_0 - \alpha_0 e^{-\alpha_1 t} e^{-\lambda t(z-t)}}{e^{\alpha_0 z} - e^{-\lambda z}} \right) dt_1 = n Ln(\alpha) + n Ln(\lambda) - \lambda \sum z_i + \\
+ (\lambda - \alpha) \sum e^{\lambda_0 z} \left( z e^{(\lambda_0 - \alpha_0) z} - e^{(\lambda_0 - \alpha_0) z} \frac{1}{(\lambda_0 - \alpha_0)} \right)
\]
Making

\[ k_i^0 = \frac{e^{-\lambda_i z_i}}{e^{-\alpha_0 z_i} - e^{-\lambda_i z_i}} \left[ z_i e^{(\lambda_0 - \alpha_0) z_i} - e^{(\lambda_0 - \alpha_0) z_i} + \frac{1}{(\lambda_0 - \alpha_0)} \right] \]  

(8)

\[ Q(\theta, \theta') = n \ln(\alpha) + n \ln(\lambda) - \lambda \sum_{i=1}^{n} (z_i - k_i^0) - \alpha \sum_{i=1}^{n} k_i^0 \]

The step M (maximization) is given by

\[
\frac{dQ(\theta, \theta')}{d\alpha} = 0 \quad \Rightarrow \quad \frac{n}{\alpha} - \sum_{i=1}^{n} k_i^0 = 0 \quad \Rightarrow \quad \alpha = \frac{n}{\sum_{i=1}^{n} k_i^0} = \frac{1}{n} \\
\frac{dQ(\theta, \theta')}{d\lambda} = 0 \quad \Rightarrow \quad \frac{n}{\lambda} - \sum_{i=1}^{n} (z_i - k_i^0) = 0 \quad \Rightarrow \quad \lambda = \frac{n}{\sum_{i=1}^{n} (z_i - k_i^0)} = \frac{1}{\tau - k^0} \\
\frac{dQ(\theta, \theta')}{d\lambda} = 0 \quad \Rightarrow \quad \frac{n}{\lambda} - \sum_{i=1}^{n} (z_i - k_i^0) = 0 \quad \Rightarrow \quad \lambda = \frac{n}{\sum_{i=1}^{n} (z_i - k_i^0)} = \frac{1}{\tau - k^0} \quad \Rightarrow \quad (9)
\]

So, the EM recurrence is: Start with any positive \((\lambda_0, \alpha_0)\), compute \(k^0\), and then use (9) and (8) until reaching some convergence criterion.

4 Simulations

To study the distribution of the estimates, we simulated (using R) samples of size 200, 500 e 1000 of the sum of exponentials with parameters \(\alpha = 1\) and \(\lambda = 5\). Initial values for the parameters were \(\alpha_0 = 2\) and \(\lambda_0 = 7\). Convergence criterion was the difference between consecutive estimates less than \(10^{-6}\). For the construction of the histograms the algorithm was run 200 times for each sample size.
Conclusions

The steps E (expectation) and M (maximization), necessary to implement the EM algorithm, are obtained with relatively trivial procedures. Therefore this can be a good example to be worked in classroom. The above histograms show that the algorithm is efficient to obtain estimates.


RESUMO: O algoritmo EM foi usado para obter estimativas de máxima verossimilhança dos parâmetros da soma de duas exponenciais com diferentes fatores de escala. Foram realizadas simulações usando o software R.

PALAVRAS-CHAVE: Algoritmo EM; esperança; maximização; distribuição exponencial; distribuição gama.
References


Received in 13.05.2007

Approved after revised in 12.09.2007