

557: MATHEMATICAL STATISTICS II

THE EM ALGORITHM

The EM Algorithm is a method for producing the maximum likelihood estimates in **incomplete data** problems, that is, models formulated for data that are only partially observed.

Suppose that random variables to be modelled can be partitioned $(\underline{X}, \underline{Y})$ where

- $\underline{X} = (X_1, \dots, X_m)^\top$ are **unobserved**, termed the **augmented data**
- $\underline{Y} = (Y_1, \dots, Y_n)^\top$ are **observed**, termed the **incomplete data**
- $(\underline{X}, \underline{Y})$ are termed the **complete data**

where

$$f_{\underline{Y}|\underline{\theta}}(\underline{y}|\underline{\theta}) = \int f_{\underline{X}, \underline{Y}|\underline{\theta}}(\underline{x}, \underline{y}|\underline{\theta}) d\underline{x}$$

In this formulation,

$$L(\underline{\theta}|\underline{y}) = f_{\underline{Y}|\underline{\theta}}(\underline{y}|\underline{\theta}).$$

is the **incomplete data** likelihood, and

$$L(\underline{\theta}|\underline{x}, \underline{y}) = f_{\underline{X}, \underline{Y}|\underline{\theta}}(\underline{x}, \underline{y}|\underline{\theta})$$

is the **complete data** likelihood.

Algorithm

The EM Algorithm facilitates maximization of the incomplete data likelihood $L(\underline{\theta}|\underline{y})$ by working with the complete data likelihood $L(\underline{\theta}|\underline{x}, \underline{y})$ and the conditional distribution

$$f_{\underline{X}|\underline{Y}, \underline{\theta}}(\underline{x}|\underline{y}, \underline{\theta}) = \frac{f_{\underline{X}, \underline{Y}|\underline{\theta}}(\underline{x}, \underline{y}|\underline{\theta})}{f_{\underline{Y}|\underline{\theta}}(\underline{y}|\underline{\theta})} = \frac{L(\underline{\theta}|\underline{x}, \underline{y})}{L(\underline{\theta}|\underline{y})} = K(\underline{x}|\underline{y}, \underline{\theta}) \quad (1)$$

$$\therefore \log L(\underline{\theta}|\underline{y}) = \log L(\underline{\theta}|\underline{x}, \underline{y}) - \log K(\underline{x}|\underline{y}, \underline{\theta}) \quad (2)$$

However, the data $\underline{x}$ are not observed, so consider replacing the right-hand side of equation (2) by the expectations with respect to the conditional density $f_{\underline{X}|\underline{Y}, \underline{\theta}'}(\underline{x}|\underline{y}, \underline{\theta}')$, for some $\underline{\theta}' \in \Theta$. This yields

$$\log L(\underline{\theta}|\underline{y}) = E_{f_{\underline{X}|\underline{Y}, \underline{\theta}'}}[\log L(\underline{\theta}|\underline{X}, \underline{Y})|\underline{y}, \underline{\theta}'] - E_{f_{\underline{X}|\underline{Y}, \underline{\theta}'}}[\log K(\underline{X}|\underline{Y}, \underline{\theta})|\underline{y}, \underline{\theta}']. \quad (3)$$

Note that the notation indicates that we condition on a specific (but as yet unspecified) value of $\underline{\theta}'$ when computing the expectations of $\log L(\underline{\theta}|\underline{X}, \underline{Y})$ and $\log K(\underline{X}|\underline{y}, \underline{\theta})$ at the $\underline{\theta}$ at which the likelihood on the left-hand side of equation (3) is being computed.

The EM algorithm is an iterative algorithm that produces a sequence of estimates that converges to the (incomplete data) maximum likelihood estimate. Generically, starting from an initial value $\underline{\theta} = \underline{\theta}^{(0)}$, the $(r + 1)$ st value in the sequence, $\underline{\theta}^{(r+1)}$, is constructed given the r th value, $\underline{\theta}^{(r)}$,

$$\underline{\theta}^{(r+1)} = \operatorname{argmax}_{\underline{\theta} \in \Theta} E_{f_{\underline{X}|\underline{Y}, \underline{\theta}^{(r)}}}[\log L(\underline{\theta}|\underline{X}, \underline{Y})|\underline{y}, \underline{\theta}^{(r)}]$$

Two components of this calculation are

- **E-step** : compute the expected conditional log-likelihood
- **M-step** : carry out the maximization of the expectation.

In the traditional notation, we write

$$Q(\underline{\theta}|\underline{\theta}') = E_{f_{\underline{X}|\underline{Y},\underline{\theta}}}[\log L(\underline{\theta}|\underline{X}, \underline{Y})|\underline{y}, \underline{\theta}']$$

We wish to show that the sequence of estimates produced by

$$\underline{\theta}^{(r+1)} = \underset{\underline{\theta} \in \Theta}{\operatorname{argmax}} Q(\underline{\theta}|\underline{\theta}^{(r)}) \quad r = 1, 2, \dots$$

converges to the maximum likelihood estimate. First, note that for two pdfs f_1 and f_2 for random variable Z , we have

$$\begin{aligned} E_{f_1}[\log f_1(Z)] - E_{f_1}[\log f_2(Z)] &= -E_{f_1}[\log\{f_2(Z)/f_1(Z)\}] \geq -\log E_{f_1}[\{f_2(Z)/f_1(Z)\}] \\ &= -\log \int_{\mathcal{Z}} \{f_2(z)/f_1(z)\} f_1(z) dz \\ &= -\log \int_{\mathcal{Z}} f_2(z) dz = 0 \end{aligned}$$

so therefore $E_{f_1}[\log f_1(Z)] \geq E_{f_1}[\log f_2(Z)]$, with equality if and only if $f_1 \equiv f_2$. For $\underline{\theta} \in \Theta$, as

$$K(\underline{\theta}|\underline{X}, \underline{Y}) = \frac{L(\underline{\theta}|\underline{X}, \underline{Y})}{L(\underline{\theta}|\underline{Y})} = f_{\underline{X}|\underline{Y},\underline{\theta}}(\underline{x}|\underline{y}, \underline{\theta})$$

is itself a (conditional) pdf for all $\underline{\theta} \in \Theta$, and hence we have, for any $\underline{\theta}$,

$$\begin{aligned} Q(\underline{\theta}|\underline{\theta}^{(r)}) - \log L(\underline{\theta}|\underline{y}) &= E_{f_{\underline{X}|\underline{Y},\underline{\theta}}}[\log L(\underline{\theta}|\underline{X}, \underline{Y})|\underline{y}, \underline{\theta}^{(r)}] - \log L(\underline{\theta}|\underline{y}) = E_{f_{\underline{X}|\underline{Y},\underline{\theta}}}[\log K(\underline{\theta}|\underline{X}, \underline{Y})|\underline{y}, \underline{\theta}^{(r)}] \\ &\leq E_{f_{\underline{X}|\underline{Y},\underline{\theta}}}[\log K(\underline{\theta}^{(r)}|\underline{X}, \underline{Y})|\underline{y}, \underline{\theta}^{(r)}] \\ &= Q(\underline{\theta}^{(r)}|\underline{\theta}^{(r)}) - \log L(\underline{\theta}^{(r)}|\underline{y}). \end{aligned}$$

Thus $\log L(\underline{\theta}|\underline{y}) - Q(\underline{\theta}|\underline{\theta}^{(r)})$ achieves its **minimum** value when $\underline{\theta} = \underline{\theta}^{(r)}$. Now suppose that $\underline{\theta}^{(r+1)}$ is the value that maximizes $Q(\underline{\theta}|\underline{\theta}^{(r)})$ over Θ ; we have that

$$\begin{aligned} \log L(\underline{\theta}^{(r+1)}|\underline{y}) &\equiv Q(\underline{\theta}^{(r+1)}|\underline{\theta}^{(r)}) + (\log L(\underline{\theta}^{(r+1)}|\underline{y}) - Q(\underline{\theta}^{(r+1)}|\underline{\theta}^{(r)})) \\ &\geq Q(\underline{\theta}^{(r)}|\underline{\theta}^{(r)}) + (\log L(\underline{\theta}^{(r)}|\underline{y}) - Q(\underline{\theta}^{(r)}|\underline{\theta}^{(r)})) \\ &= \log L(\underline{\theta}^{(r)}|\underline{y}) \end{aligned}$$

and the likelihood attained is **increasing** with the sequence $\underline{\theta}^{(0)}, \underline{\theta}^{(1)}, \underline{\theta}^{(2)}, \dots$

Thus, provided $L(\underline{\theta}|\underline{y})$ is bounded, we have that the sequence of $\log L(\underline{\theta}^{(r)}|\underline{y})$ converges to a (local) maximum of $L(\underline{\theta}|\underline{y})$.

Example 1 Finite Mixture Model

Suppose that $Y_1, \dots, Y_n$ are a random sample from the K component finite mixture model

$$f_{Y|\underline{\theta}}(y|\underline{\theta}) = \sum_{k=1}^K \omega_k f_k(y|\theta_k) \quad y \in \mathbb{R}$$

where $f_1, \dots, f_K$ are component densities, and $0 < \omega_k < 1$ and $\sum_{k=1}^K \omega_k = 1$.

Estimation of $\underline{\theta} = (\theta_1, \dots, \theta_K)^\top$ from the likelihood $L(\underline{\theta}|y)$ is in general difficult. However, consider the augmented data $X_1, \dots, X_n$, where

$$\Pr[X_i = k] = \omega_k \quad i = 1, \dots, K$$

are independent random variables so that

$$L(\underline{\theta}|\underline{X}, \underline{Y}) = \prod_{i=1}^n \prod_{k=1}^K \{\omega_k f_k(y_i|\theta_k)\}^{\mathbb{I}_{\{k\}}(X_i)}$$

and

$$\log L(\underline{\theta}|\underline{X}, \underline{Y}) = \sum_{i=1}^n \sum_{k=1}^K \mathbb{I}_{\{k\}}(X_i) (\log \omega_k + \log f_k(y_i|\theta_k)).$$

The conditional distribution $f_{X|Y, \underline{\theta}}(x|y, \underline{\theta})$ is a discrete distribution on the set $\{1, 2, \dots, K\}$ where for each $i = 1, \dots, n$

$$\Pr[X_i = k|\underline{Y}, \underline{\omega}, \underline{\theta}] = \frac{\omega_k f_k(y_i|\theta_k)}{\sum_{j=1}^K \omega_j f_j(y_i|\theta_j)} = \varpi_k(y_i, \underline{\theta}) \quad k = 1, \dots, K$$

where $X_1, \dots, X_n$ are conditionally independent. Thus

$$\mathbb{E}_{f_{X_i|Y_i, \underline{\omega}, \underline{\theta}}}[\mathbb{I}_{\{k\}}(X_i)|y_i, \underline{\theta}, \underline{\omega}] = \varpi_k(y_i, \underline{\theta})$$

and hence

$$\begin{aligned} Q(\underline{\theta}, \underline{\omega}|\underline{\theta}^{(r)}, \underline{\omega}^{(r)}) &= \mathbb{E}_{f_{\underline{X}|\underline{Y}, \underline{\theta}, \underline{\omega}}}[\log L(\underline{\theta}|\underline{X}, \underline{Y})|\underline{y}, \underline{\theta}^{(r)}, \underline{\omega}^{(r)}] \\ &= \sum_{i=1}^n \sum_{k=1}^K \varpi_k^{(r)}(y_i, \underline{\theta}^{(r)}) (\log \omega_k + \log f_k(y_i|\theta_k)) \\ &= \sum_{k=1}^K \left\{ \sum_{i=1}^n \varpi_k^{(r)}(y_i, \underline{\theta}^{(r)}) \right\} \log \omega_k + \sum_{k=1}^K \sum_{i=1}^n \varpi_k^{(r)}(y_i, \underline{\theta}^{(r)}) \log f_k(y_i|\theta_k) \end{aligned} \quad (4)$$

We seek to maximize over $(\underline{\theta}, \underline{\omega})$ to obtain $(\underline{\theta}^{(r+1)}, \underline{\omega}^{(r+1)})$ presuming that the values $\varpi_k^{(r)}(y_i, \underline{\theta}^{(r)})$ are fixed. From the form of equation (4) it is evident that the function is sum of two parts, the first only depending on $\underline{\omega}$, the second only dependent on $\underline{\theta}$. We can therefore maximize the two parts separately to obtain $(\underline{\theta}^{(r+1)}, \underline{\omega}^{(r+1)})$.

The first part of equation (4) is of the form of a multinomial likelihood in ϖ , therefore, by previous results, it follows that

$$\omega_k^{(r+1)} = \frac{\sum_{i=1}^n \varpi_k^{(r)}(y_i, \underline{\theta}^{(r)})}{\sum_{j=1}^K \sum_{i=1}^n \varpi_j^{(r)}(y_i, \underline{\theta}^{(r)})} \quad k = 1, \dots, K$$

The second part of equation (4) is the sum of K log-likelihoods for the K mixture components which can be maximized separately

$$\theta_k^{(r+1)} = \operatorname{argmax}_{\theta_k} \sum_{i=1}^n \varpi_k^{(r)}(y_i, \underline{\theta}^{(r)}) \log f_k(y_i | \theta_k) \quad (5)$$

For certain choices of the component densities, this maximization can be carried out analytically. For example, if $f_k(y | \theta_k)$ is the normal density with expectation μ_k and variance σ_k^2 , it follows that the new maximizing value equals $\theta_k^{(r+1)} = (\mu_k^{(r+1)}, \sigma_k^{(r+1)})$ where

$$\mu_k^{(r+1)} = \frac{\sum_{i=1}^n \varpi_k^{(r)}(y_i, \underline{\theta}^{(r)}) y_i}{\sum_{i=1}^n \varpi_k^{(r)}(y_i, \underline{\theta}^{(r)})}$$

and

$$\sigma_k^{(r+1)} = \sqrt{\frac{\sum_{i=1}^n \varpi_k^{(r)}(y_i, \underline{\theta}^{(r)}) (y_i - \mu_k^{(r+1)})^2}{\sum_{i=1}^n \varpi_k^{(r)}(y_i, \underline{\theta}^{(r)})}}$$

Note that in the normal model the terms in (5) correspond to likelihood components of the form

$$\{f_k(y_i | \theta_k)\}^{\varpi_k^{(r)}} = \left(\frac{1}{2\pi\sigma_k^2}\right)^{\varpi_k^{(r)}/2} \exp \left\{ -\frac{\varpi_k^{(r)}}{2\sigma_k^2} (y_i - \mu_k^{(r)})^2 \right\}$$

so the terms $\varpi_k^{(r)} \equiv \varpi_k^{(r)}(y_i, \underline{\theta}^{(r)})$ are acting as weighting factors.

Example 2 The EM Algorithm: Genetics of Human Blood Groups

In human genetics, the **genotype** at a genomic locus is a pair of **alleles** corresponding to small segments of DNA lying on the two chromosomal strands. The **phenotype** is the physical presentation or trait arising from the genotype. At a certain locus that determines the phenotype of blood group, the relationship between genotype and phenotype is somewhat complex; there are

- three alleles (A, B and O) yielding **six** possible genotypes (ordering is not important)
- only **four** phenotypes (A,B,AB and O).

The relationship between phenotype and genotype in this case is determined by the following table. The third column, headed X , denotes a label for the genotype class. However, only the phenotype may be observed; let $Y_1, \dots, Y_n$ denote the recorded phenotype for each of the n data.

Genotype	Phenotype	X	Y
AA	A	1	1
AB	AB	2	3
AO	A	3	1
BB	B	4	2
BO	B	5	2
OO	O	6	4

Suppose that inference about the proportions of the three alleles A,B and O, denoted $\theta_A, \theta_B, \theta_O$ is required from a sample of size n of phenotype data. We formulate a data augmentation approach, and use the EM algorithm to perform maximum likelihood estimation. One independence assumption (based on so-called *Hardy-Weinberg equilibrium*) is needed; we assume that the probability of observing a genotype is the product of the individual allele probabilities. For example

$$P(AA) = \theta_A \times \theta_A \quad P(AB) = \theta_A \times \theta_B$$

and so on.

Define the augmented data $X_1, \dots, X_n$, where for $i = 1 \dots, n$,

$$\Pr[X_i = j] = \Pr[i\text{th genotype is in class } j] \quad j = 1, \dots, 6$$

that is

$$\Pr[X_i = j] = \begin{cases} \theta_A^2 & j = 1 \\ \theta_A \theta_B & j = 2 \\ \theta_A \theta_O & j = 3 \\ \theta_B^2 & j = 4 \\ \theta_B \theta_O & j = 5 \\ \theta_O^2 & j = 6 \end{cases} \quad \text{for } i = 1, \dots, n$$

with $X_1, \dots, X_n$ a random sample. This simplification yields a complete data likelihood

$$L(\underline{\theta}|\underline{x}, \underline{y}) \equiv L(\underline{\theta}|\underline{x}) = \prod_{i=1}^n \left\{ \theta_A^{2\mathbb{I}_{\{1\}}(x_i) + \mathbb{I}_{\{2\}}(x_i) + \mathbb{I}_{\{3\}}(x_i)} \theta_B^{\mathbb{I}_{\{2\}}(x_i) + 2\mathbb{I}_{\{4\}}(x_i) + \mathbb{I}_{\{5\}}(x_i)} \theta_O^{\mathbb{I}_{\{3\}}(x_i) + \mathbb{I}_{\{5\}}(x_i) + 2\mathbb{I}_{\{6\}}(x_i)} \right\}$$

say, where

$$n_j = \sum_{i=1}^n \mathbb{I}_{\{j\}}(x_i) \quad j = 1, \dots, 6.$$

The complete data likelihood is a multinomial-type likelihood in $\underline{\theta}$.

In the standard notation, for the EM steps, we have to

- E-step: compute

$$Q(\underline{\theta}|\underline{\theta}^{(r)}) = E_{f_{\underline{X}|\underline{Y},\underline{\theta}}}[\log L(\underline{\theta}|\underline{X}, \underline{Y})|\underline{y}, \underline{\theta}^{(r)}]$$

taking the expectation over $X_1, \dots, X_n$ etc.

- M-step: maximize $Q(\underline{\theta}|\underline{\theta}^{(r)})$ to get $\underline{\theta}^{(r+1)}$.

Here the M-step is straightforward due to the multinomial likelihood. The E-step is also quite straightforward, but some steps need clarification.

The log complete data likelihood takes the form

$$\begin{aligned} \log L(\underline{\theta}|\underline{x}, \underline{y}) &= \sum_{i=1}^n (2\mathbb{I}_{\{1\}}(x_i) + \mathbb{I}_{\{2\}}(x_i) + \mathbb{I}_{\{3\}}(x_i)) \log \theta_A \\ &\quad + \sum_{i=1}^n (\mathbb{I}_{\{2\}}(x_i) + 2\mathbb{I}_{\{4\}}(x_i) + \mathbb{I}_{\{5\}}(x_i)) \log \theta_B \\ &\quad + \sum_{i=1}^n (\mathbb{I}_{\{3\}}(x_i) + \mathbb{I}_{\{5\}}(x_i) + 2\mathbb{I}_{\{6\}}(x_i)) \log \theta_O \end{aligned}$$

which is linear and additive in the indicator functions.

Conditional on $\underline{Y}$ and $\underline{\theta}$, some expectations can be written down automatically. For example

$$\begin{aligned} E_{f_{X_i|Y_i,\underline{\theta}}}[\mathbb{I}_{\{j\}}(X_i)|Y_i = 3, \underline{\theta}] &= \begin{cases} 1 & j = 2 \\ 0 & j \neq 2 \end{cases} \\ E_{f_{X_i|Y_i,\underline{\theta}}}[\mathbb{I}_{\{j\}}(X_i)|Y_i = 4, \underline{\theta}] &= \begin{cases} 1 & j = 6 \\ 0 & j \neq 6 \end{cases} \end{aligned}$$

as by definition $Y = 3 \implies X = 2$ and $Y = 4 \implies X = 6$. For the remaining conditional expectations, we have by Bayes theorem

$$E_{f_{X_i|Y_i,\underline{\theta}}}[\mathbb{I}_{\{j\}}(X_i)|Y_i = 1, \underline{\theta}] = \begin{cases} \frac{\theta_A^2}{\theta_A^2 + 2\theta_A\theta_O} & j = 1 \\ \frac{2\theta_A\theta_O}{\theta_A^2 + 2\theta_A\theta_O} & j = 3 \\ 0 & \text{otherwise} \end{cases}$$

as if $Y = 1$, then either $X = 1$ or $X = 3$, with conditional probability for each determined by noting that

$$\Pr[X = 1|Y = 1] = \frac{\Pr[X = 1, Y = 1]}{\Pr[Y = 1]} = \frac{\Pr[X = 1, Y = 1]}{\Pr[X = 1, Y = 1] + \Pr[X = 3, Y = 1]} = \frac{P(AA)}{P(AA) + P(AO)}$$

Similarly,

$$\mathbb{E}_{f_{X_i|Y_i,\underline{\theta}}}[\mathbb{I}_{\{j\}}(X_i)|Y_i = 2, \underline{\theta}] = \begin{cases} \frac{\theta_B^2}{\theta_B^2 + 2\theta_B\theta_O} & j = 4 \\ \frac{2\theta_B\theta_O}{\theta_B^2 + 2\theta_B\theta_O} & j = 5 \\ 0 & \text{otherwise} \end{cases}$$

Thus $Q(\underline{\theta}|\underline{\theta}^{(r)})$ takes the form

$$Q(\underline{\theta}|\underline{\theta}^{(r)}) = \alpha_A^{(r)} \log \theta_A + \alpha_B^{(r)} \log \theta_B + \alpha_O^{(r)} \log \theta_O$$

where

$$\begin{aligned} \alpha_A^{(r)} &= \frac{2n_1\theta_A^{(r)2}}{\theta_A^{(r)2} + 2\theta_A^{(r)}\theta_O^{(r)}} + n_3 + \frac{2n_1\theta_A^{(r)}\theta_O^{(r)}}{\theta_A^{(r)2} + 2\theta_A^{(r)}\theta_O^{(r)}} \\ \alpha_B^{(r)} &= n_3 + \frac{2n_2\theta_B^{(r)2}}{\theta_B^{(r)2} + 2\theta_B^{(r)}\theta_O^{(r)}} + \frac{2n_2\theta_B^{(r)}\theta_O^{(r)}}{\theta_B^{(r)2} + 2\theta_B^{(r)}\theta_O^{(r)}} \\ \alpha_O^{(r)} &= \frac{2n_1\theta_A^{(r)}\theta_O^{(r)}}{\theta_A^{(r)2} + 2\theta_A^{(r)}\theta_O^{(r)}} + \frac{2n_2\theta_B^{(r)}\theta_O^{(r)}}{\theta_B^{(r)2} + 2\theta_B^{(r)}\theta_O^{(r)}} + 2n_4. \end{aligned}$$

and $n_1, \dots, n_4$ are the observed counts for phenotypes A,B,AB and O. By the results for the multinomial likelihood, we can maximize $Q(\underline{\theta}|\underline{\theta}^{(r)})$ analytically to get

$$\theta_A^{(r+1)} = \frac{\alpha_A^{(r)}}{\alpha_A^{(r)} + \alpha_B^{(r)} + \alpha_O^{(r)}} \quad \theta_B^{(r+1)} = \frac{\alpha_B^{(r)}}{\alpha_A^{(r)} + \alpha_B^{(r)} + \alpha_O^{(r)}} \quad \theta_O^{(r+1)} = \frac{\alpha_O^{(r)}}{\alpha_A^{(r)} + \alpha_B^{(r)} + \alpha_O^{(r)}}$$

Real Data Example: Data from Clarke *et. al.* (1959)

We have $n_1 = 186$, $n_2 = 38$, $n_3 = 13$ and $n_4 = 284$ for the numbers of A,B,AB and O phenotypes in a sample of $n = 521$. Starting the iterative procedure at $\underline{\theta}^{(0)} = (1/3, 1/3, 1/3)^T$ yields the following first ten iterations:

r	$\theta_A^{(r)}$	$\theta_B^{(r)}$	$\theta_O^{(r)}$
1	0.25047985	0.06110045	0.68841971
2	0.21845436	0.05049394	0.73105170
3	0.21418233	0.05016173	0.73565593
4	0.21366195	0.05014667	0.73619139
5	0.21359944	0.05014547	0.73625508
6	0.21359196	0.05014535	0.73626270
7	0.21359106	0.05014533	0.73626361
8	0.21359095	0.05014533	0.73626372
9	0.21359094	0.05014533	0.73626373
10	0.21359094	0.05014533	0.73626373

indicating that convergence to the maximum value is fairly rapid.

Example 3 The EM Algorithm: Censored Data

Suppose that $Y_1, \dots, Y_n$ are the realized failure times of electronic components, and that in addition there are m additional components that are censored at times $t_{n+1}, \dots, t_{n+m}$. Denote by $X_{n+1}, \dots, X_{n+m}$ the unobserved failure times of these m components (so that we observe only that $X_{n+j} > t_{n+j}$ for $j = 1, \dots, m$).

Under the assumption that the data are $Exponential(\theta)$ distributed, we may carry out inference about θ using the EM algorithm. We have the complete data likelihood as

$$L(\theta|\underline{x}, \underline{y}) = \prod_{i=1}^n \theta e^{-\theta y_i} \times \prod_{i=n+1}^{n+m} \theta e^{-\theta x_i} = \theta^{n+m} \exp \left\{ -\theta \left[\sum_{i=1}^n y_i + \sum_{i=n+1}^{n+m} x_i \right] \right\}$$

so that

$$\log L(\theta|\underline{x}, \underline{y}) = (n+m) \log \theta - \theta \left[\sum_{i=1}^n y_i + \sum_{i=n+1}^{n+m} x_i \right].$$

Bearing in mind the constraint that $X_{n+j} > t_{n+j}$, we note that for $i = n+1, \dots, n+m$, in the Exponential model that exhibits the lack of memory property

$$E_{f_{X_i|\underline{y}, \theta}}[X_i|\underline{y}, \theta] = t_i + \frac{1}{\theta}$$

Thus

$$Q(\theta|\theta^{(r)}) = (n+m) \log \theta - \theta \left[\sum_{i=1}^n y_i + \sum_{i=n+1}^{n+m} t_i + \frac{m}{\theta^{(r)}} \right]$$

which is readily maximized to yield

$$\theta^{(r+1)} = \frac{n+m}{\sum_{i=1}^n y_i + \sum_{i=n+1}^{n+m} t_i + \frac{m}{\theta^{(r)}}}$$

For the following data

$$3.479 \quad 0.57 \quad 1.067^* \quad 1.736^* \quad 0.156^* \quad 0.265 \quad 0.044^* \quad 0.595 \quad 4.515^* \quad 1.617$$

where the $\star$ superscript indicates censored values, we have $n = m = 5$. If $\theta^{(0)} = 1$, we have

r	$\theta^{(r)}$	r	$\theta^{(r)}$
1	0.525137	11	0.356170
2	0.424376	12	0.356114
3	0.387227	13	0.356086
4	0.370989	14	0.356072
5	0.363370	15	0.356065
6	0.359677	16	0.356061
7	0.357858	17	0.356060
8	0.356956	18	0.356059
9	0.356506	19	0.356058
10	0.356282	20	0.356058

indicating that convergence to the maximum value is slower than in earlier examples. Note that in the exponential model, the maximum likelihood estimate is available directly as

$$L(\theta|\underline{y}, \underline{t}) = \theta^n \exp \left\{ -\theta \left[\sum_{i=1}^n y_i + \sum_{i=n+1}^{n+m} t_i \right] \right\} \quad \therefore \quad \hat{\theta}(\underline{y}, \underline{t}) = \frac{n}{\sum_{i=1}^n y_i + \sum_{i=n+1}^{n+m} t_i} = \frac{5}{6.525 + 7.518} = 0.356058.$$