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 $(\mathbf{Y}, \mathbf{Z})$ where

- $\mathbf{Z} = (Z_1, \ldots, Z_m)^\top$ are **unobserved**, termed the **augmented data**;
- $\mathbf{Y} = (Y_1, \ldots, Y_n)^\top$ are **observed**, termed the **incomplete data**;
- $(\mathbf{Y}, \mathbf{Z})$ are termed the **complete data**.

where

$$f_{\mathbf{Y}}(\mathbf{y};\theta) = \int f_{\mathbf{Z},\mathbf{Y}|\theta}(\mathbf{z},\mathbf{y};\theta)\, d\mathbf{z}$$

In this formulation, we have

- the **incomplete data** likelihood, $f_{\mathbf{Y}}(\mathbf{y};\theta)$;
- the **complete data** likelihood, $f_{\mathbf{Z},\mathbf{Y}}(\mathbf{z},\mathbf{y};\theta)$.

The EM Algorithm facilitates maximization of the incomplete data likelihood by working with the complete data likelihood and the conditional distribution

$$f_{\mathbf{Z}|\mathbf{Y}}(\mathbf{z}|\mathbf{y};\theta) = \frac{f_{\mathbf{Z},\mathbf{Y}}(\mathbf{z},\mathbf{y};\theta)}{f_{\mathbf{Y}}(\mathbf{y};\theta)} = K(\mathbf{z}|\mathbf{y};\theta) \tag{1}$$

$$\therefore \quad \log f_{\mathbf{Y}}(\mathbf{y};\theta) = \log f_{\mathbf{Z},\mathbf{Y}}(\mathbf{z},\mathbf{y};\theta) - \log K(\mathbf{z}|\mathbf{y};\theta) \tag{2}$$

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

$$\log f_{\mathbf{Y}}(\mathbf{y};\theta) = \mathbb{E}_{\mathbf{Z}|\mathbf{Y}}[\log f_{\mathbf{Y},\mathbf{Z}}(\mathbf{Y},\mathbf{Z};\theta)|\mathbf{Y}=\mathbf{y};\theta'] - \mathbb{E}_{\mathbf{Z}|\mathbf{Y}}[\log K(\mathbf{Z}|\mathbf{Y};\theta)|\mathbf{Y}=\mathbf{y};\theta']. \tag{3}$$

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

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

$$\widehat{\theta}^{(t+1)} = \underset{\theta\,\in\,\Theta}{\text{argmax}}\ \mathbb{E}_{\mathbf{Z}|\mathbf{Y}}[\log f_{\mathbf{Y},\mathbf{Z}}(\mathbf{Y},\mathbf{Z};\theta)|\mathbf{Y}=\mathbf{y};\widehat{\theta}^{(t)}]$$

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

$$Q(\theta|\theta') = \mathbb{E}_{\mathbf{Z}|\mathbf{Y}}[\log f_{\mathbf{Y},\mathbf{Z}}(\mathbf{Y},\mathbf{Z};\theta)|\mathbf{Y}=\mathbf{y};\widehat{\theta}^{(t)}]$$

We wish to show that the sequence of estimates produced by

$$\widehat{\theta}^{(t+1)} = \underset{\theta\,\in\,\Theta}{\text{argmax}}\ Q(\theta|\widehat{\theta}^{(t)}) \qquad t = 1, 2, \ldots$$

converges to the maximum likelihood estimate. First, note that for two pdfs f_1 and f_2 for random variable Z, we have by the properties of the Kullback-Leibler divergence that

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

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

$$K(\mathbf{z}|\mathbf{y};\theta) = f_{\mathbf{Z}|\mathbf{Y}}(\mathbf{z}|\mathbf{y};\theta)$$

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

$$\begin{aligned}
Q(\theta|\widehat{\theta}^{(t)}) - \log f_{\mathbf{Y}}(\mathbf{y};\theta) &= \mathbb{E}_{\mathbf{Z}|\mathbf{Y}}[\log f_{\mathbf{Y},\mathbf{Z}}(\mathbf{Y},\mathbf{Z};\theta)|\mathbf{Y}=\mathbf{y};\widehat{\theta}^{(t)}] - \log f_{\mathbf{Y}}(\mathbf{y};\theta) \\
&= \mathbb{E}_{\mathbf{Z}|\mathbf{Y}}\left[\log K(\mathbf{Z}|\mathbf{Y};\theta)|\mathbf{Y}=\mathbf{y};\widehat{\theta}^{(t)}\right] \\
&\leq \mathbb{E}_{\mathbf{Z}|\mathbf{Y}}\left[\log K(\mathbf{Z}|\mathbf{Y};\widehat{\theta}^{(t)})|\mathbf{Y}=\mathbf{y};\widehat{\theta}^{(t)}\right] \\
&= Q(\widehat{\theta}^{(t)}|\widehat{\theta}^{(t)}) - \log f_{\mathbf{Y}}(\mathbf{y};\widehat{\theta}^{(t)}).
\end{aligned}$$

Thus the function

$$g(\theta) = \log f_{\mathbf{Y}}(\mathbf{y};\theta) - Q(\theta|\widehat{\theta}^{(t)})$$

achieves its **minimum** value when $\theta = \widehat{\theta}^{(t)}$.

Now suppose that $\widehat{\theta}^{(t+1)}$ is the value that maximizes $Q(\theta|\widehat{\theta}^{(t)})$ over Θ; we have that

$$\begin{aligned}
\log f_{\mathbf{Y}}(\mathbf{y};\widehat{\theta}^{(t+1)}) &\equiv Q(\widehat{\theta}^{(t+1)}|\widehat{\theta}^{(t)}) + \left(\log f_{\mathbf{Y}}(\mathbf{y};\widehat{\theta}^{(t+1)}) - Q(\widehat{\theta}^{(t+1)}|\widehat{\theta}^{(t)})\right) \\
&\geq Q(\widehat{\theta}^{(t)}|\widehat{\theta}^{(t)}) + \left(\log f_{\mathbf{Y}}(\mathbf{y};\widehat{\theta}^{(t)}) - Q(\widehat{\theta}^{(t)}|\widehat{\theta}^{(t)})\right) \\
&= \log f_{\mathbf{Y}}(\mathbf{y};\widehat{\theta}^{(t)})
\end{aligned}$$

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

Thus, provided $\log f_{\mathbf{Y}}(\mathbf{y};\theta)$ is bounded, we have that the sequence of $\log f_{\mathbf{Y}}(\mathbf{y};\widehat{\theta}^{(t)})$ converges to a (local) maximum of $\log f_{\mathbf{Y}}(\mathbf{y};\theta)$.

Example 1 Finite Mixture Model
Suppose that $Y_1 \ldots, Y_n$ are a random sample from the K component finite mixture model

$$f_{Y|\theta}(y;\theta) = \sum_{k=1}^{K} \pi_k f_k(y_i;\theta_k) \qquad y \in \mathbb{R}$$

where $f_1, \ldots, f_K$ are component densities, and $0 < \pi_k < 1$ and $\sum\limits_{k=1}^{K} \pi_k = 1$.

Estimation of parameters $\theta = (\theta_1, \ldots, \theta_K)^\top$ and $\pi = (\pi_1, \ldots, \pi_K)$ from the likelihood

$$f_{\mathbf{Y}}(\mathbf{y};\theta,\pi) = \prod_{i=1}^{n} \left\{ \sum_{k=1}^{K} \pi_k f_k(y_i;\theta_k) \right\}$$

is in general difficult. However, consider the augmented data $X_1, \ldots, X_n$, where

$$\Pr[Z_i = k] = \pi_k \qquad i = 1, \ldots, K$$

are independent random variables so that

$$f_{\mathbf{Y},\mathbf{Z}}(\mathbf{y},\mathbf{z};\theta,\pi) = \prod_{i=1}^{n}\prod_{k=1}^{K} \{\pi_k f_k(y_i;\theta_k)\}^{\mathbb{1}_{\{k\}}(Z_i)}$$

and

$$\log f_{\mathbf{Y},\mathbf{Z}}(\mathbf{y},\mathbf{Z};\theta,\pi) = \sum_{i=1}^{n}\sum_{k=1}^{K} \mathbb{1}_{\{k\}}(Z_i) \left(\log \pi_k + \log f_k(y_i;\theta_k)\right).$$

The conditional distribution $f_{Z|Y,\theta}(z|y;\theta)$ is a **discrete** distribution on the set $\{1, 2, \ldots, K\}$ where for each $i = 1, \ldots, n$

$$\Pr[Z_i = k|\mathbf{Y} = \mathbf{y}, \pi;\theta] = \frac{\pi_k f_k(y_i;\theta_k)}{\sum\limits_{j=1}^{K} \pi_j f_j(y_i;\theta_j)} = \omega_k(y_i;\theta,\pi) \qquad k = 1, \ldots, K$$

say, where $Z_1, \ldots, Z_n$ are conditionally independent. Thus

$$\mathbb{E}_{Z_i|Y_i}[\mathbb{1}_{\{k\}}(Z_i)|y_i;\theta,\pi] = \omega_k(y_i;\theta,\pi)$$

and hence, for the EM update, we have

$$\begin{aligned} Q(\theta,\pi|\widehat{\theta}^{(t)},\widehat{\pi}^{(t)}) &= \mathbb{E}_{\mathbf{Z}|\mathbf{Y}}[\log f_{\mathbf{Y},\mathbf{Z}}(\mathbf{y},\mathbf{Z};\theta,\pi)|\mathbf{Y} = \mathbf{y};\widehat{\theta}^{(t)},\pi^{(t)}] \\ &= \sum_{i=1}^{n}\sum_{k=1}^{K} \omega_k(y_i;\widehat{\theta}^{(t)},\widehat{\pi}^{(t)}) \left(\log \pi_k + \log f_k(y_i;\theta_k)\right) \\ &= \sum_{k=1}^{K} \left\{ \sum_{i=1}^{n} \omega_k(y_i;\widehat{\theta}^{(t)},\widehat{\pi}^{(t)}) \right\} \log \pi_k + \sum_{k=1}^{K}\sum_{i=1}^{n} \omega_k(y_i;\widehat{\theta}^{(t)},\widehat{\pi}^{(t)}) \log f_k(y_i;\theta_k) \qquad (4) \end{aligned}$$

We seek to maximize over (θ,π) to obtain $(\widehat{\theta}^{(t+1)},\widehat{\pi}^{(t+1)})$ presuming that the values $\omega_k(y_i;\widehat{\theta}^{(t)},\widehat{\pi}^{(t)})$ are fixed.

From the form of equation (4) it is evident that the function is sum of two parts, the first only depending on π, the second only dependent on θ. We can therefore maximize the two parts separately to obtain $(\widehat{\theta}^{(t+1)}, \widehat{\pi}^{(t+1)})$.

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

$$\widehat{\pi}_k^{(t+1)} = \frac{\sum_{i=1}^{n} \omega_k(y_i; \widehat{\theta}^{(t)}, \widehat{\pi}^{(t)})}{\sum_{j=1}^{K} \sum_{i=1}^{n} \omega_j(y_i; \widehat{\theta}^{(t)}, \widehat{\pi}^{(t)})} \qquad k = 1, \ldots, K$$

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

$$\widehat{\theta}_k^{(t+1)} = \underset{\theta_k}{\operatorname{argmax}} \sum_{i=1}^{n} \omega_k(y_i; \widehat{\theta}^{(t)}, \widehat{\pi}^{(t)}) \log f_k(y_i; \theta_k) \tag{5}$$

For certain choices of the component densities, this maximization can be carried out analytically. For example, if $f_k(y_i; \theta_k)$ is the Normal density with expectation μ_k and variance σ_k^2, it follows that the new maximizing value equals $\widehat{\theta}_k{}^{(t+1)} = (\widehat{\mu}_k^{(t+1)}, \widehat{\sigma}_k^{(t+1)})$ where

$$\widehat{\mu}_k^{(t+1)} = \frac{\sum_{i=1}^{n} \omega_k(y_i; \widehat{\theta}^{(t)}, \widehat{\pi}^{(t)}) y_i}{\sum_{i=1}^{n} \omega_k(y_i; \widehat{\theta}^{(t)}, \widehat{\pi}^{(t)})}$$

and

$$\widehat{\sigma}_k^{(t+1)} = \sqrt{\frac{\sum_{i=1}^{n} \omega_k(y_i; \widehat{\theta}^{(t)}, \widehat{\pi}^{(t)})(y_i - \widehat{\mu}_k^{(t+1)})^2}{\sum_{i=1}^{n} \omega_k(y_i; \widehat{\theta}^{(t)} \widehat{\pi}^{(t)})}}$$

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

$$\{f_k(y_i; \theta_k)\}^{\omega_k^{(t)}} = \left(\frac{1}{2\pi\sigma_k^2}\right)^{\omega_k^{(t)}/2} \exp\left\{-\frac{\omega_k^{(t)}}{2\sigma_k^2}(y_i - \mu_k^{(t)})^2\right\}$$

so the terms $\omega_k^{(t)} \equiv \omega_k(y_i; \widehat{\theta}^{(t)}, \widehat{\pi}^{(t)})$ 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 Z, denotes a label for the genotype class. However, only the phenotype may be observed; let $Y_1, \ldots, Y_n$ denote the recorded phenotype for each of the n data.

Genotype	Phenotype	Z	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. An 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

$$\mathrm{P}(\mathrm{AA}) = \theta_A \times \theta_A \qquad \mathrm{P}(\mathrm{AB}) = \theta_A \times \theta_B$$

and so on.

Define the augmented data

$$\Pr[Z_i = j] = \Pr[i\text{th genotype is in class j}] \qquad j = 1, \ldots, 6$$

that is

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

with $Z_1, \ldots, Z_n$ a random sample. This simplification yields a complete data likelihood

$$\begin{aligned} f_{\mathbf{Y},\mathbf{Z}}(\mathbf{y}, \mathbf{z}; \theta) &\equiv f_{\mathbf{Z}}(\mathbf{z}; \theta) \\ &= \prod_{i=1}^{n} \left\{ \theta_A^{2\mathbb{1}_{\{1\}}(z_i)+\mathbb{1}_{\{2\}}(z_i)+\mathbb{1}_{\{3\}}(z_i)} \theta_B^{\mathbb{1}_{\{2\}}(z_i)+2\mathbb{1}_{\{4\}}(z_i)+\mathbb{1}_{\{5\}}(z_i)} \theta_O^{\mathbb{1}_{\{3\}}(z_i)+\mathbb{1}_{\{5\}}(z_i)+2\mathbb{1}_{\{6\}}(z_i)} \right\} \end{aligned}$$

say, where

$$n_j = \sum_{i=1}^{n} \mathbb{1}_{\{j\}}(z_i) \qquad j = 1, \ldots, 6.$$

The complete data likelihood is a multinomial-type likelihood in θ. In the standard notation, for the EM steps, we have to

- E-step: compute
$$Q(\theta|\widehat{\theta}^{(t)}) = \mathbb{E}_{\mathbf{Z}|\mathbf{Y}}[\log f_{\mathbf{Y},\mathbf{Z}}(\mathbf{y},\mathbf{z};\theta)|\mathbf{Y}=\mathbf{y};\widehat{\theta}^{(t)}]$$
taking the expectation over $Z_1,\ldots,Z_n$ etc.
- M-step: maximize $Q(\theta|\widehat{\theta}^{(t)})$ to get $\widehat{\theta}^{(t+1)}$.

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

$$\begin{aligned}\log f_{\mathbf{Y},\mathbf{Z}}(\mathbf{y},\mathbf{z};\theta) \quad = \quad & \sum_{i=1}^{n}\left(2\mathbb{1}_{\{1\}}(z_i)+\mathbb{1}_{\{2\}}(z_i)+\mathbb{1}_{\{3\}}(z_i)\right)\log\theta_A \\ & +\sum_{i=1}^{n}\left(\mathbb{1}_{\{2\}}(z_i)+2\mathbb{1}_{\{4\}}(z_i)+\mathbb{1}_{\{5\}}(z_i)\right)\log\theta_B \\ & +\sum_{i=1}^{n}\left(\mathbb{1}_{\{3\}}(z_i)+\mathbb{1}_{\{5\}}(z_i)+2\mathbb{1}_{\{6\}}(z_i)\right)\log\theta_O\end{aligned}$$

which is linear and additive in the indicator functions.

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

$$\mathbb{E}_{Z_i|Y_i}[\mathbb{1}_{\{j\}}(Z_i)|Y_i=3;\theta] \quad = \quad \begin{cases} 1 & j=2 \\ 0 & j\neq 2\end{cases}$$

$$\mathbb{E}_{Z_i|Y_i}[\mathbb{1}_{\{j\}}(Z_i)|Y_i=4;\theta] \quad = \quad \begin{cases} 1 & j=6 \\ 0 & j\neq 6\end{cases}$$

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

$$\mathbb{E}_{Z_i|Y_i}[\mathbb{1}_{\{j\}}(Z_i)|Y_i=1;\theta] = \begin{cases} \dfrac{\theta_A^2}{\theta_A^2+2\theta_A\theta_O} & j=1 \\ \\ \dfrac{2\theta_A\theta_O}{\theta_A^2+2\theta_A\theta_O} & j=3 \\ \\ 0 & \text{otherwise}\end{cases}$$

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

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

Similarly,

$$\mathbb{E}_{Z_i|Y_i}[\mathbb{1}_{\{j\}}(Z_i)|Y_i=2;\theta] = \begin{cases} \dfrac{\theta_B^2}{\theta_B^2+2\theta_B\theta_O} & j=4 \\ \\ \dfrac{2\theta_B\theta_O}{\theta_B^2+2\theta_B\theta_O} & j=5 \\ \\ 0 & \text{otherwise}\end{cases}$$

Thus $Q(\theta|\widehat{\theta}^{(t)})$ takes the form

$$Q(\theta|\widehat{\theta}^{(t)}) = \widehat{\alpha}_A^{(t)} \log \theta_A + \widehat{\alpha}_B^{(t)} \log \theta_B + \widehat{\alpha}_O^{(t)} \log \theta_O$$

where

$$\begin{aligned}
\widehat{\alpha}_A^{(t)} &= \frac{2n_1 \widehat{\theta}_A^{(t)2}}{\widehat{\theta}_A^{(t)2} + 2\widehat{\theta}_A^{(t)}\widehat{\theta}_O^{(t)}} + n_3 + \frac{2n_1 \widehat{\theta}_A^{(t)}\widehat{\theta}_O^{(t)}}{\widehat{\theta}_A^{(t)2} + 2\widehat{\theta}_A^{(t)}\widehat{\theta}_O^{(t)}} \\
\widehat{\alpha}_B^{(t)} &= n_3 + \frac{2n_2 \widehat{\theta}_B^{(t)2}}{\widehat{\theta}_B^{(t)2} + 2\widehat{\theta}_B^{(t)}\widehat{\theta}_O^{(t)}} + \frac{2n_2 \widehat{\theta}_B^{(t)}\widehat{\theta}_O^{(t)}}{\widehat{\theta}_B^{(t)2} + 2\widehat{\theta}_B^{(t)}\widehat{\theta}_O^{(t)}} \\
\widehat{\alpha}_O^{(t)} &= \frac{2n_1 \widehat{\theta}_A^{(t)}\widehat{\theta}_O^{(t)}}{\widehat{\theta}_A^{(t)2} + 2\widehat{\theta}_A^{(t)}\widehat{\theta}_O^{(t)}} + \frac{2n_2 \widehat{\theta}_B^{(t)}\widehat{\theta}_O^{(t)}}{\widehat{\theta}_B^{(t)2} + 2\widehat{\theta}_B^{(t)}\widehat{\theta}_O^{(t)}} + 2n_4.
\end{aligned}$$

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

$$\theta_A^{(t+1)} = \frac{\widehat{\alpha}_A^{(t)}}{\widehat{\alpha}_A^{(t)} + \widehat{\alpha}_B^{(t)} + \widehat{\alpha}_O^{(t)}} \qquad \theta_B^{(t+1)} = \frac{\widehat{\alpha}_B^{(t)}}{\widehat{\alpha}_A^{(t)} + \widehat{\alpha}_B^{(t)} + \widehat{\alpha}_O^{(t)}} \qquad \theta_O^{(t+1)} = \frac{\widehat{\alpha}_O^{(t)}}{\widehat{\alpha}_A^{(t)} + \widehat{\alpha}_B^{(t)} + \widehat{\alpha}_O^{(t)}}$$

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 $\widehat{\theta}^{(0)} = (1/3, 1/3, 1/3)^\top$ yields the following first ten iterations:

r	$\widehat{\theta}_A^{(t)}$	$\widehat{\theta}_B^{(t)}$	$\widehat{\theta}_O^{(t)}$
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 Censored Data
Suppose that $Y_1, \ldots, Y_n$ are the realized failure times of electronic components, and that in addition there are m additional components that are censored at times $t_{n+1}, \ldots, t_{n+m}$. Let $Z_{n+1}, \ldots, Z_{n+m}$ be the **unobserved** failure times of these m components (so that we observe only that $Z_{n+j} > t_{n+j}$ for $j = 1, \ldots, m$). Under the assumption that the data are $Exponential(\theta)$ distributed, we have the complete data likelihood as

$$f_{\mathbf{Y},\mathbf{Z}}(\mathbf{y}, \mathbf{z}; \theta) = \prod_{i=1}^{n} \theta e^{-\theta y_i} \times \prod_{i=n+1}^{n+m} \theta e^{-\theta z_i} = \theta^{n+m} \exp\left\{-\theta \left[\sum_{i=1}^{n} y_i + \sum_{i=n+1}^{n+m} t_i\right]\right\}$$

so that

$$\log f_{\mathbf{Y},\mathbf{Z}}(\mathbf{y}, \mathbf{z}; \theta) = (n+m)\log\theta - \theta\left[\sum_{i=1}^{n} y_i + \sum_{i=n+1}^{n+m} t_i\right].$$

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

$$\mathbb{E}_{Z_i|Y_i}[Z_i | Y_i = y_i; \theta] = t_i + \frac{1}{\theta}$$

Thus

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

which is readily maximized to yield

$$\widehat{\theta}^{(t+1)} = \frac{n+m}{\sum\limits_{i=1}^{n} y_i + \sum\limits_{i=n+1}^{n+m} t_i + \dfrac{m}{\widehat{\theta}^{(t)}}}$$

For the following data

3.479 0.57 1.067* 1.736* 0.156* 0.265 0.044* 0.595 4.515* 1.617

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

r	$\widehat{\theta}^{(t)}$	r	$\widehat{\theta}^{(t)}$
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

$$f_{\mathbf{Y},\mathbf{T}}(\mathbf{y}, \mathbf{t}; \theta) = \theta^{n} \exp\left\{-\theta\left[\sum_{i=1}^{n} y_i + \sum_{i=n+1}^{n+m} t_i\right]\right\}$$

so

$$\widehat{\theta}(\mathbf{y}, \mathbf{t}) = \frac{n}{\sum\limits_{i=1}^{n} y_i + \sum\limits_{i=n+1}^{n+m} t_i} = \frac{5}{6.525 + 7.518} = 0.356058.$$