The EM algorithm applied to diversification modeling

F. Richter Mendoza

October 28, 2016

The EM algorithm

The EM algorithm is a broadly applicable algorithm that provides an iterative procedure for computing MLE in situations where, but for the absence of some aditional data, ML estimation would be straightfoward.

The EM algorithm

The situations where the EM algorithm can be applies include not only evidently incomplete-data situations, where there are missing data, truncated distributions, censored or grouped observations, but also a whole variety of situations where the incompleteness of data is not all that natural or evident. These includes statistical models such as random effect, mixtures, convolutions, log linear models, and latent variable structures. The Em algorithm has thus found applications in almost all fields where statistical techniques have been applied– medical imaging, diary science, correcting census undercount, and AIDS apidemiology, to mention a few.

Formulation of the EM algorithm

The EM algorithm approaches the problem of solving the incomplete-data likelihood equation indirecly by proceeding iteratively in terms of the complete-data log likelihood function $L_c(\theta)$. As it is unobserved, it is replaced by its conditional expectation given y, using the current fit for θ.

Monotonicity of the EM algorithm

Theorem

The EM algorithm increases the likelihood at every iteration.

Proof:

Let

$$k(x|y;\theta) = g_c(x;\theta)/g(y;\theta)$$

be the conditional density of X given $Y = y$.

Monotonicity of the EM algorithm

Theorem

The EM algorithm increases the likelihood at every iteration.

Proof:

Let

$$k(x|y;\theta) = g_c(x;\theta)/g(y;\theta)$$

be the conditional density of X given $Y = y$.Then the log likelihood is given by

$$\begin{aligned} logL(\theta) &= \log g(y;\theta) \\ &= \log g_c(x;\theta) - \log k(x|y;\theta) \\ &= \log L_c(\theta) - \log k(x|y;\theta) \end{aligned} \tag{1}$$

Monotonicity of the EM algorithm

On taking the expectations of both sides with respect to the conditional distribution of X given $Y = y$, using the fit $\theta^{(k)}$ fot θ, we have that

$$\begin{aligned} \log L(\theta) &= E_{\theta^{(k)}}\{\log L_c(\theta)|y\} - E_{\theta^{(k)}}\{\log k(X|y;\theta)|y\} \\ &= Q(\theta;\theta^{(k)}) - H(\theta;\theta^{(k)}) \end{aligned} \tag{2}$$

Monotonicity of the EM algorithm

On taking the expectations of both sides with respect to the conditional distribution of X given $Y = y$, using the fit $\theta^{(k)}$ fot θ, we have that

$$\begin{aligned} \log L(\theta) &= E_{\theta^{(k)}}\{\log L_c(\theta)|y\} - E_{\theta^{(k)}}\{\log k(X|y;\theta)|y\} \\ &= Q(\theta;\theta^{(k)}) - H(\theta;\theta^{(k)}) \end{aligned} \tag{2}$$

where

$$H(\theta;\theta^{(k)}) = E_{\theta^{(k)}}\{\log k(X|y;\theta)|y\}$$

Monotonicity of the EM algorithm

With that, we have

$$\log L(\theta^{(k+1)}) - \log L(\theta^{(k)}) \\ = \{Q(\theta^{(k+1)};\theta^{(k)}) - Q(\theta^{(k)};\theta^{(k)})\} - \{H(\theta^{(k+1)};\theta^{(k)}) - H(\theta^{(k)};\theta^{(k)})\} \quad (3)$$

Monotonicity of the EM algorithm

With that, we have

$$\log L(\theta^{(k+1)}) - \log L(\theta^{(k)})$$
$$= \{Q(\theta^{(k+1)}; \theta^{(k)}) - Q(\theta^{(k)}; \theta^{(k)})\} - \{H(\theta^{(k+1)}; \theta^{(k)}) - H(\theta^{(k)}; \theta^{(k)})\} \tag{3}$$

the first difference on the right hand is not negative since $\theta^{(k+1)}$ is chosen so that

$$Q(\theta^{(k+1)}; \theta^{(k)}) \geq Q(\theta; \theta^{(k)}), \qquad \forall \theta \in \Theta$$

Monotonicity of the EM algorithm

$$
\begin{aligned}
H(\theta^{(k+1)};\theta^{(k)}) - H(\theta^{(k)};\theta^{(k)}) &= E_{\theta^{(k)}}[\log\{k(X|y;\theta)/k(X|y;\theta^{(k)})\}|y] \\
&\leq \log E_{\theta^{(k)}}[\{k(X|y;\theta)/k(X|y;\theta^{(k)})\}|y] \\
&= \log \int_{\mathcal{X}(y)} k(x|y;\theta)dx \\
&= 0
\end{aligned} \tag{4}
$$

Monotonicity of the EM algorithm

$$\begin{aligned}
H(\theta^{(k+1)};\theta^{(k)}) - H(\theta^{(k)};\theta^{(k)}) &= E_{\theta^{(k)}}[\log\{k(X|y;\theta)/k(X|y;\theta^{(k)})\}|y] \\
&\leq \log E_{\theta^{(k)}}[\{k(X|y;\theta)/k(X|y;\theta^{(k)})\}|y] \\
&= \log \int_{\mathcal{X}(y)} k(x|y;\theta)dx \\
&= 0
\end{aligned} \tag{4}$$

Finally, we can conclude that the EM algorithm increases the Likelihood function every iteracion.

Convergence of the EM algorithm

Theorem

For unimodal likelihood functions the EM algorithm converges to its global maximum.

MCEM

$$Q(\theta|\theta_{(i)}) = \int_{\{\ ,\ ,\ ,\dots\}} \log L(\theta|\ )\mathrm{d}\ \longrightarrow \theta_{(i+1)} = \arg\max_{\theta} Q\left(\theta|\theta_{(i)}\right)$$