Stat 543 EM Algorithm Example

Suppose that n iid trials each produce one of k possible outcomes. If

$$p_j = \text{the probability that any trial produces outcome } j$$

and

$$X_{ij} = I\left[\text{trial } i \text{ produces outcome } j\right]$$

and

$$n_j = \sum_{i=1}^{n} X_{ij} = \text{the number of trials producing outcome } j$$

a joint pmf for all nk of the variables X_{ij} is

$$f\left(\mathbf{x} \mid \mathbf{p}\right) = \prod_{j=1}^{k} p_j^{n_j} = \left(\prod_{j=1}^{k-1} p_j^{n_j}\right)\left(1 - \sum_{j=1}^{k-1} p_j\right)^{n_k}$$

A loglikelihood is

$$l_{\mathbf{x}}\left(\mathbf{p}\right) = \sum_{j=1}^{k} n_j \ln p_j$$

and it is possible to argue that an MLE of $\mathbf{p}$ is $\left(\frac{n_1}{n}, \frac{n_2}{n}, \ldots, \frac{n_k}{n}\right)$.

For sake of example, suppose that what is observed in $n = 10$ trials in a problem where $k = 4$ is as below

trial 1 outcome is 1 $\quad X_{11} = 1$
trial 2 outcome is 3 $\quad X_{23} = 1$
trial 3 outcome is 2 or 4 $\quad X_{32} + X_{34} = 1$
trial 4 outcome is 2 $\quad X_{42} = 1$
trial 5 outcome is 3 $\quad X_{53} = 1$
trial 6 outcome is 2 or 3 $\quad X_{62} + X_{63} = 1$
trial 7 outcome is 1 $\quad X_{71} = 1$
trial 8 outcome is 1 or 2 $\quad X_{81} + X_{82} = 1$
trial 9 outcome is 2 $\quad X_{92} = 1$
trial 10 outcome is 4 $\quad X_{104} = 1$

The likelihood function based on this information is

$$\begin{aligned} L_{\mathbf{Y}}(\mathbf{p}) &= p_1^2 p_2^2 p_3^2 (1-p_1-p_2-p_3)(1-p_1-p_3)(p_2+p_3)(p_1+p_2) \\ &= p_1^2 p_2^2 p_3^2 p_4 (p_2+p_4)(p_2+p_3)(p_1+p_2) \end{aligned}$$

and one might potentially simply try to optimize this directly. Another possibility is consider an EM algorithm.

If I knew all X_{ij} I would have the log-likelihood

$$l_X(\mathbf{p}) = n_1 \ln p_1 + n_2 \ln p_2 + n_3 \ln p_3 + n_4 \ln(1-p_1-p_2-p_3)$$

that I know how to optimize. I don't have all the n_j's. But I do know that

$$\begin{aligned} n_1 &= 2 + X_{81} \\ n_2 &= 2 + X_{32} + X_{62} + X_{82} \\ n_3 &= 2 + X_{63} \\ n_4 &= 1 + X_{34} \end{aligned}$$

and for a particular $\mathbf{p}_0 = (p_{01}, p_{02}, p_{03}, p_{04})$ conditioned on the data in hand (call them $\mathbf{Y} = \mathbf{y}$)

$$\begin{aligned} \mathrm{E}_{\mathbf{p}_0}[X_{81} \mid \mathbf{Y}=\mathbf{y}] &= \frac{p_{01}}{p_{01}+p_{02}} \\ \mathrm{E}_{\mathbf{p}_0}[X_{32} \mid \mathbf{Y}=\mathbf{y}] &= \frac{p_{02}}{p_{02}+p_{04}} \\ \mathrm{E}_{\mathbf{p}_0}[X_{62} \mid \mathbf{Y}=\mathbf{y}] &= \frac{p_{02}}{p_{02}+p_{03}} \\ \mathrm{E}_{\mathbf{p}_0}[X_{82} \mid \mathbf{Y}=\mathbf{y}] &= \frac{p_{02}}{p_{01}+p_{02}} \\ \mathrm{E}_{\mathbf{p}_0}[X_{63} \mid \mathbf{Y}=\mathbf{y}] &= \frac{p_{03}}{p_{02}+p_{03}} \\ \mathrm{E}_{\mathbf{p}_0}[X_{34} \mid \mathbf{Y}=\mathbf{y}] &= \frac{p_{04}}{p_{02}+p_{04}} \end{aligned}$$

So $\mathrm{E}_{\mathbf{p}_0}\left[l_{\mathbf{X}}(\mathbf{p}) \mid \mathbf{Y}=\mathbf{y}\right]$ is easy enough to compute. For

$$
\begin{aligned}
a(\mathbf{p}_0) &= 2 + \frac{p_{01}}{p_{01} + p_{02}} \\
b(\mathbf{p}_0) &= 2 + p_{02}\left(\frac{1}{p_{01} + p_{02}} + \frac{1}{p_{02} + p_{03}} + \frac{1}{p_{02} + p_{04}}\right) \\
c(\mathbf{p}_0) &= 2 + \frac{p_{03}}{p_{02} + p_{03}} \\
d(\mathbf{p}_0) &= 1 + \frac{p_{04}}{p_{02} + p_{04}}
\end{aligned}
$$

note that $a(\mathbf{p}_0) + b(\mathbf{p}_0) + c(\mathbf{p}_0) + d(\mathbf{p}_0) = 10 = n$ and then that

$$\mathrm{E}_{\mathbf{p}_0}\left[l_{\mathbf{X}}(\mathbf{p}) \mid \mathbf{Y} = \mathbf{y}\right] = a(\mathbf{p}_0)\ln p_1 + b(\mathbf{p}_0)\ln p_2 + c(\mathbf{p}_0)\ln p_3 + d(\mathbf{p}_0)\ln p_4$$

This is maximized at

$$\mathbf{p}^*(\mathbf{p}_0) = \left(\frac{a(\mathbf{p}_0)}{10}, \frac{b(\mathbf{p}_0)}{10}, \frac{c(\mathbf{p}_0)}{10}, \frac{d(\mathbf{p}_0)}{10}\right)$$

So, with a starting value $\mathbf{p}^{(0)}$, define iterates by

$$\mathbf{p}^{(l+1)} = \mathbf{p}^*\left(\mathbf{p}^{(l)}\right)$$

and hope to iterate to a fixed point optimizing the loglikelihood.