1 EM model incorporating sample size error

The E-step log likelihood under the EM algorithm has the following form

$$E_{Z,M}|X, f_k^{(n)}, Q^{(n)}, F_u^{(n)}, F_k^{(n)} \left[logP(X, Z, M, f_k^{(n)}|Q^{(n)}, F_u^{(n)}, F_k^{(n)}) \right]$$

We first determine the conditional distribution of Z and M given X, Q and F as in Skotte et al 2013.

$$P(Z_{ij} = t, M_{ij} = 1|X_{ij}, f_k^{(n)}, Q^{(n)}, F_u^{(n)}, F_k^{(n)}) = c_n^{(ijt)} X_{ij}/2$$

$$P(Z_{ij} = t, M_{ij} = 0|X_{ij}, f_k^{(n)}, Q^{(n)}, F_u^{(n)}, F_k^{(n)}) = d_n^{(ijt)} (2 - X_{ij})/2$$

When the updates do not distinguish between F_k and F_u, we will use the notation F for the sake of readability.

$$c_n^{(ijt)} = \frac{Q_{it}^n F_{j,k}^{(n)}}{\sum_l Q_{il}^n F_{j,l}^{(n)}}$$

$$d_n^{(ijt)} = \frac{Q_{it}^n (1 - F_{j,k}^{(n)})}{\sum_l Q_{il}^n (1 - F_{j,l}^{(n)})}$$

In addition,

$$P(Z_{ij} = t, M_{ij} = 1|Q^{(n)}, F^{(n)}) = Q_{it}^n F_{j,t}^{(n)}$$

$$P(Z_{ij} = t, M_{ij} = 0|Q^{(n)}, F^{(n)}) = Q_{it}^n (1 - F_{j,t}^{(n)})$$

We can write down the E step log likelihood in that case as

$$\begin{aligned}
&E_{Z,M}|X, f_k^{(n)}, Q^{(n)}, F_u^{(n)}, F_k^{(n)} \left[logP(X, Z, M, f_k|Q, F_u, F_k) \right] \\
= \quad & E_{Z,M}|X, f_k^{(n)}, Q^{(n)}, F_u^{(n)}, F_k^{(n)} \left[logP(X, Z, M|Q, F_u, F_k) + logP(f_k|F_k) \right] \\
= \quad & \sum_{i,j} E_{Z_{ij}, M_{ij}|X_{ij}, f_k^{(n)}, Q^{(n)}, F_u^{(n)}, F_k^{(n)}} \left[logP(X_{ij}, Z_{ij}, M_{ij}|Q, F_u, F_k) \right] + logP(f_k|F_k) \\
\propto \quad & \sum_{ij} \sum_{t=1}^{T_u} \left[log(q_{it} F_u^{jt}) a_n^{ijt} + log(q_{it}(1 - F_u^{(jt)})) b_n^{ijt} \right] + \sum_{ij} \sum_{t=1}^{T_k} \left[log(q_{it} F_k^{jt}) a_n^{ijt} + log(q_{it}(1 - F_k^{(jt)})) b_n^{ijt} \right] + \\
& \sum_{j} \sum_{t=1}^{T_k} \left[f_{jt} n_t log(F_{jt}) + n_t (1 - f_{jt}) log(1 - F_{jt}) \right]
\end{aligned}$$

here

$$a_n^{ijt} = c_n^{ijt} X_{ij}/2$$

$$b_n^{ijt} = d_n^{ijt} (2 - X_{ij})/2$$

Take the derivative with respect to F_u and solve for roots to get the $(n+1)$ th update of F_u to be

$$F_{u,(n+1)}^{jt} = \frac{\sum_i a_n^{ijt}}{\sum_i a_n^{ijt} + \sum_i b_n^{ijt}}$$

Take the derivative with respect to F_k and solve for roots to get the $(n+1)$ th update of F_u to be

$$\begin{aligned} F_{k,(n+1)}^{jt} &= \frac{\sum_i a_n^{ijt} + n_t f_{jt}}{\sum_i a_n^{ijt} + \sum_i b_n^{ijt} + n_t} \\ &= \left(\frac{\sum_i a_n^{ijt} + \sum_i b_n^{ijt}}{\sum_i a_n^{ijt} + \sum_i b_n^{ijt} + n_t} \right) \frac{\sum_i a_n^{ijt}}{\sum_i a_n^{ijt} + \sum_i b_n^{ijt}} + \left(\frac{n_t}{\sum_i a_n^{ijt} + \sum_i b_n^{ijt} + n_t} \right) f_{jt} \end{aligned}$$

Take the derivative with respect to Q with constraints and solve for roots to get the $(n+1)$ th update of Q to be

$$q_{it}^{(n+1)} = \frac{1}{M} \sum_{j=1}^{M} \left(a_n^{ijt} + b_n^{ijt} \right)$$