1 Paired factor analysis (PFA) model

Let D_{nj} be the data corresponding to n-th sample and j-th feature, where n runs from 1 to N and j runs from 1 to J. Suppose these data come from a graph with K nodes (factors) and E edges. In the PFA set up $E = \frac{K(K-1)}{2}$.

Let us define latent variables Z and Λ. Z_n is a $(E+K)\times 1$ binary vector. We have a prior on $Z_{n,k1,k2,l}$, with the constraint that if $l=0$, then k_1 and k_2 are both 0, and if k_1 and k_2 are non-zero, then l is 0.

$$Pr\left[Z_{n,k_1,k_2,0}=1\right] = \pi_{k_1,k_2} \qquad k_1 < k_2$$
$$Pr\left[Z_{n,0,0,l}=1\right] = \pi_l \qquad l = 1,2,\cdots,K$$

So, we assume that

$$\sum_{k_1<k_2} \pi_{k_1,k_2} + \sum_{l=1}^{K} \pi_l = 1$$
$$Pr(Z|\pi) = \prod_{n=1}^{N} \prod_{k_1<k_2}^{E} \pi_{k1,k2}^{z_{n,k1,k2,0}} \prod_{l=1}^{K} \pi_l^{z_{n,0,0,l}}$$

Λ_n is a $Q\times 1$ binary vector, where Q is the cardinality of the set of values that q can take. For computational convenience we assume that q can take a finite set of values between 0 and 1, say $1/100, 2/100, \cdots, 99/100, 1$.

We have a prior on Λ,

$$Pr\left[\Lambda_{n,q}=1\right] = \frac{1}{Q}$$

For data on the graph given edge (k_1,k_2) and position on edge q we assume the model

$$E\left[D_{nj}|Z_{n,k_1,k_2,0}=1, \Lambda_{n,q}=1, F\right] = qF_{k1,j} + (1-q)F_{k2,j} \tag{1}$$

$$E\left[D_{nj}|Z_{n,0,0,l}=1, F\right] = F_{l,j} \tag{2}$$

Where F is a $K\times J$ matrix of factors. Then we can marginalize over Z and Λ, assuming as a first pass that the data is Gaussian in its distribution,

$$Pr\left[D_n|Z_{n,k_1,k_2,0}=1, F, s^2_{j=1,2,\cdots,J}\right] = \sum_q \frac{1}{Q} Pr\left[D_n|Z_{n,k_1,k_2}=1, \Lambda_{n,q}=1, F, s^2_{j=1,2,\cdots,J}\right]$$

$$\begin{aligned} Pr\left[D_n|\pi, F, s^2_{j=1,2,\cdots,J}\right] = \textstyle\sum_{k_1<k_2} \pi_{k_1,k_2} Pr\left[D_n|Z_{n,k_1,k_2,0}=1, F, s^2_{j=1,2,\cdots,J}\right] \\ + \sum_l \pi_l Pr\left[D_n|Z_{n,0,0,l}=1, F, s^2_{j=1,2,\cdots,J}\right] \end{aligned} \tag{3}$$

where s_j^2 is the residual variance of the jth feature.

We define the joint prior over the edges and the fraction of the edge represented as

$$\pi_{k_1,k_2,q} = \pi_{k_1,k_2} \times \frac{1}{Q} \qquad k_1 < k_2$$

The overall likelihood

$$L(\pi, F) = \prod_{n=1}^{N} Pr\left[D_n|\pi, F, s^2_{j=1,2,\cdots,J}\right]$$

or we can write it as

$$L(\pi, F) = \textstyle\prod_{n=1}^{N}\left[\sum_{k_1<k_2}\sum_{q=1}^{Q}\left[\pi_{k_1,k_2,q} \times \prod_{j=1}^{J} N\left(D_{nj}; qF_{k_1,g} + (1-q)F_{k_2,g}, s_j^2\right)\right] + \right.$$
$$\left.\sum_{l}\left[\pi_l \times \prod_{j=1}^{J} N\left(D_{nj}; F_{l,g}, s_j^2\right)\right]\right] \quad (4)$$

And the log likelihood

$$\ln L(\pi, F) = \sum_{n=1}^{N} \ln\left(\sum_{k_1<k_2}\sum_{q}\left[\pi_{k_1,k_2,q} \times \prod_{j=1}^{J} N\left(D_{nj}; qF_{k_1,g} + (1-q)F_{k_2,g}, s_j^2\right)\right] + \right.$$
$$\left.\sum_{l=1}^{K}\pi_l \times \prod_{j=1}^{J} N\left(D_{nj}; F_{l,g}, s_j^2\right)\right) \quad (5)$$

This is the quantity we want to maximize.

2 EM algorithm

2.1 E step

We assume that q can take a finite set of values between 0 and 1, say $1/100, 2/100, \cdots, 90/100, 1$.

Suppose we have run upto m iterations. For the $(m+1)$th iteration, we have

$$\begin{aligned}
\delta^{(m+1)}_{n,k_1,k_2,0} &= Pr\left[Z_{n,k1,k2,0} = 1|\pi^{(m)}, F^{(m)}, s^{(m)}_{j=1,2,\cdots,J}, D_n\right] \\
&\propto Pr\left[Z_{n,k1,k2,0} = 1\right]\sum_{q=1}^{Q} Pr\left[\lambda_{n,q} = 1\right] Pr\left[D_n|\pi^{(m)}, F^{(m)}, s^{(m)}_{j=1,2,\cdots,J}, Z_{n,k1,k2,0} = 1, \lambda_{n,q} = 1\right] \\
&\propto \pi^{(m)}_{k_1,k_2}\sum_{q=1}^{Q}\left[\prod_{j} N\left(D_{nj}|qF^{(m)}_{k_1,j} + (1-q)F^{(m)}_{k_2,j}, {s_j^{(m)}}^2\right)\right]
\end{aligned}$$

$$
\begin{aligned}
\delta_{n,0,0,l}^{(m+1)} &= Pr\left[Z_{n,0,0,l}=1|\pi^{(m)},F^{(m)},s_{j=1,2,\cdots,J}^{(m)},D_n\right] \\
&\propto Pr\left[Z_{n,0,0,l}=1\right]Pr\left[D_n|\pi^{(m)},F^{(m)},s_{j=1,2,\cdots,J}^{(m)},Z_{n,0,0,l}=1\right] \\
&\propto \pi_l^{(m)}\prod_j N\left(D_{nj}|F_{l,j}^{(m)},{s_j^{(m)}}^2\right)
\end{aligned}
$$

where ${s_j^{(m)}}^2$ is the residual variance of feature j.
We normalize δ so that

$$
\sum_{k_1<k_2}\delta_{n,k_1,k_2,0}^{(m+1)}+\sum_{l=1}^{K}\delta_{n,0,0,l}^{(m+1)}=1 \qquad \forall n
$$

We define

$$
\pi_{k_1,k_2}^{(m+1)}=\frac{1}{N}\sum_{n=1}^{N}\delta_{n,k_1,k_2,0}^{(m+1)}
$$

$$
\pi_l^{(m+1)}=\frac{1}{N}\sum_{n=1}^{N}\delta_{n,0,0,l}^{(m+1)}
$$

We have therefore updated $\pi_{k_1,k_2}^{(m)}$ to $\pi_{k_1,k_2}^{(m+1)}$.
In the M step, we define

$$
\delta_{n,k_1,k_2,0,q}^{(m+1)}=\delta_{n,k_1,k_2,0}^{(m+1)}\times\frac{1}{Q}
$$

2.1.1 Variational EM - Model 2

In model 2, we do not assume independence of the latent variables Z and Λ and instead estimate their joint variational distribution.

$$
q(Z,\Lambda,\pi)=q(Z)q(\pi)q(\Lambda)
$$

We assume up front that $q*(\Lambda)=\prod_{n=1}^{N}\left[\frac{1}{100}\right]^{\Lambda_{nq}}$.
In this set up, we assume prior distributions of π as follows

$$
Pr(\pi|\alpha_0)=C(\alpha_0)\prod_{k_1<k_2}\pi_{k1,k2}^{\alpha_0-1}\prod_{l=1}^{L}\pi_l^{\alpha_0-1}
$$

$$
\begin{aligned}
\ln q^{\star}(Z) &= E_{\pi,\Lambda}\left[\ln p(\pi|\alpha_0) + \ln p(\Lambda) + \ln p(Z|\pi) + \ln p(D|Z,\Lambda,F,s_{j=1,2,\cdots,J})\right] \\
&= E_{\pi,\Lambda}\left[\ln p(Z|\pi) + \ln p(D|Z,\Lambda,F,s_{j=1,2,\cdots,J})\right] + constant \\
&= \sum_{n=1}^{N}\sum_{k_1<k_2} z_{n,k1,k2,0} E_{\pi}\left[\ln(\pi_{k1,k2})\right] \\
&\quad + \sum_{n=1}^{N}\sum_{k_1<k_2} z_{n,k1,k2,0} \sum_{q=1}^{Q}\frac{1}{Q}\left[-\sum_{j=1}^{J}\ln(s_j) - \frac{J}{2}\ln(2\pi) - \sum_{j=1}^{J}\frac{(D_{nj}-qF_{k1,j}-(1-q)F_{k2,j})^2}{2s_j^2}\right] \\
&\quad + \sum_{n=1}^{N}\sum_{l} z_{n,0,0,l}\left[-\sum_{j=1}^{J}\ln(s_j) - \frac{J}{2}\ln(2\pi) - \sum_{j=1}^{J}\frac{(D_{nj}-F_{l,j})^2}{2s_j^2}\right] \\
&\quad + \sum_{n=1}^{N}\sum_{l=1}^{K} z_{n,0,0,l} E_{\pi}\left[\ln(\pi_l)\right]
\end{aligned} \tag{6}
$$

$$
\begin{aligned}
\ln q^{\star}(Z,\Lambda) &= E_{\pi}\left[\ln p(\pi|\alpha_0) + \ln p(\Lambda) + \ln p(Z|\pi) + \ln p(D|Z,\Lambda,F,s_{j=1,2,\cdots,J})\right] \\
&= E_{\pi,\nu}\left[\ln p(Z|\pi) + \ln p(\Lambda) + \ln p(D|Z,\Lambda,F,s_{j=1,2,\cdots,J})\right] + constant \\
&= \sum_{n=1}^{N}\sum_{k_1<k_2} z_{n,k1,k2,0} E_{\pi}\left[\ln(\pi_{k1,k2})\right] + + \sum_{n=1}^{N}\sum_{l=1}^{K} z_{n,0,0,l} E_{\pi}\left[\ln(\pi_l)\right] \\
&\quad + \sum_{n=1}^{N}\sum_{k_1<k_2} z_{n,k1,k2,0}\left[-\sum_{j=1}^{J}\ln(s_j) - \frac{J}{2}\ln(2\pi) - \sum_{j=1}^{J}\frac{(D_{nj}-qF_{k1,j}-(1-q)F_{k2,j})^2}{2s_j^2}\right] \\
&\quad + \sum_{n=1}^{N}\sum_{l} z_{n,0,0,l}\left[-\sum_{j=1}^{J}\ln(s_j) - \frac{J}{2}\ln(2\pi) - \sum_{j=1}^{J}\frac{(D_{nj}-F_{l,j})^2}{2s_j^2}\right]
\end{aligned} \tag{7}
$$

From here one can get

$$
q^{\star}(Z) \propto \prod_{n=1}^{N}\left[\prod_{k_1<k_2}\delta_{n,k1,k2,0}^{Z_{n,k1,k2,0}}\prod_{l=1}^{K}\delta_{n,0,0,l}^{Z_{n,0,0,l}}\right]
$$

$$
q^{\star}(\pi) \propto \prod_{k_1<k_2}\pi_{k1,k2}^{a_{k1,k2,0}-1}\prod_{l=1}^{L}\pi_l^{a_{0,0,l}-1}
$$

then

$$
\begin{aligned}
\delta_{n,k1,k2,0} &\propto exp\left(E_{\pi}\left[\ln(\pi_{k1,k2})\right]\right. \\
&\quad \left. + \sum_{q=1}^{Q}\frac{1}{Q}\left[-\sum_{j=1}^{J}\ln(s_j) - \frac{J}{2}\ln(2\pi) - \sum_{j=1}^{J}\frac{(D_{nj}-qF_{k1,j}-(1-q)F_{k2,j})^2}{2s_j^2}\right]\right)
\end{aligned} \tag{8}
$$

$$\delta_{n,k1,k2,0} \propto exp\Bigg(\psi(a_{k1,k2,0}) - \psi(\sum_l a_{0,0,l} + \sum_{k_1<k_2} a_{k_1,k_2,0}) + \sum_{q=1}^{Q}\frac{1}{Q}\left[-\sum_{j=1}^{J}\ln(s_j) - \frac{J}{2}\ln(2\pi) - \sum_{j=1}^{J}\frac{(D_{nj} - qF_{k1,j} - (1-q)F_{k2,j})^2}{2s_j^2}\right]\Bigg) \quad (9)$$

$$\delta_{n,0,0,l} \quad \propto \quad exp\left(E_\pi\left[\ln(\pi_l)\right] + \left[-\sum_{j=1}^{J}\ln(s_j) - \frac{J}{2}\ln(2\pi) - \sum_{j=1}^{J}\frac{(D_{nj} - F_{l,j})^2}{2s_j^2}\right]\right) \quad (10)$$

$$\delta_{n,0,0,l} \propto exp\left(\psi(a_{0,0,l}) - \psi(\sum_l a_{0,0,l} + \sum_{k_1<k_2} a_{k_1,k_2,0}) + \left[-\sum_{j=1}^{J}\ln(s_j) - \frac{J}{2}\ln(2\pi) - \sum_{j=1}^{J}\frac{(D_{nj} - F_{l,j})^2}{2s_j^2}\right]\right) \quad (11)$$

We can also derive variational distributions similarly for π.

$$\begin{aligned}
\ln q^\star(\pi) &= E_{\Lambda,Z}\left[\ln p(\pi|\alpha_0) + \ln p(Z|\pi) + \ln p(D|Z,\Lambda,F,s_{j=1,2,\cdots,J})\right] \\
&= E_Z\left[\ln p(Z|\pi)\right] + \ln p(\pi|\alpha_0) + constant \\
&= \sum_{n=1}^{N}\sum_{k1<k2} E(z_{n,k1,k2,0})\ln \pi_{k1,k2} + \sum_{n=1}^{N}\sum_{l=1}^{K} E(z_{n,0,0,l})\ln \pi_l + (\alpha_0 - 1)\sum_{k1<k2}\ln \pi_{k1,k2} \\
&= \sum_{k1<k2}\left[\sum_{n=1}^{N}\delta_{n,k1,k2,0} + (\alpha_0 - 1)\right]\ln \pi_{k1,k2} + \sum_{l=1}^{K}\left[\sum_{n=1}^{N}\delta_{n,0,0,l} + (\alpha_0 - 1)\right]\ln \pi_l
\end{aligned}$$

We define

$$a_{k1,k2,0} = \alpha_0 + \sum_{n=1}^{N}\delta_{n,k1,k2,0}$$

$$a_{0,0,l} = \alpha_0 + \sum_{n=1}^{N}\delta_{n,0,0,l}$$

$$q^\star(\pi) = Dir(\pi|a)$$

We initialize π first, then use $a_{k1,k2} = \alpha_0$ to begin with and estimate $\delta_{n,k1,k2,0}$. Then use the $\delta_{n,k1,k2,0}$ to update $a_{k1,k2}$ and proceed in this way. In this case, we do not assume independence of the Λ and Z variational distributions, so this model is more generalized.

Again just as before, set

$$\delta_{n,k1,k2,0,q} = \delta_{n,k1,k2,0} \times \frac{1}{Q}$$

2.2 M step

We take the expectation of this quantity with respect to $\left[Z, \lambda | D, \theta^{(m)}\right]$.

$$Q(\theta|\theta^{(m)}) \propto -\sum_{n=1}^{N} \sum_{k_1<k_2} \sum_{q=1}^{Q} \delta^{(m+1)}_{n,k_1,k_2,0,q} \sum_{j} \left[log s_j^{(m+1)} + \frac{(D_{nj} - qF_{k_1,j} - (1-q)F_{k_2,j})^2}{2{s_j^{(m+1)}}^2} \right] \tag{12}$$

$$-\sum_{n=1}^{N} \sum_{l=1}^{K} \sum_{q=1}^{Q} \delta^{(m+1)}_{n,0,0,l,q} \sum_{j} \left[log s_j^{(m+1)} + \frac{(D_{nj} - F_{l,j})^2}{2{s_j^{(m+1)}}^2} \right] \tag{13}$$

We try to maximize this quantity with respect to F, So, we can take derivative with respect to F and try to solve the resulting normal equation.

This equation, conditional on $\left[Z, \lambda | D, \theta^{(m)}\right]$, can be written as

$$D_{N\times J} = L_{N\times K} F_{K\times J} + E_{N\times J} \tag{14}$$

where

$$e_{nj} \sim N(0, s_j^2)$$

We define

$$D'_{nj} := \frac{D_{nj}}{s_j}$$

If we consider finding the factors on a feature-by-feature basis, we do not need to worry about s_j.

$$L_{nk} = \begin{cases} q \text{ or } (1-q) & \lambda_n = q \quad Z_{n,k,*,0} = 1 \text{ or } \quad Z_{n,*,k,0} = 1 \\ 1 & Z_{n,0,0,k} = 1 \\ 0 & \text{o.w.} \end{cases}$$

We have

$$E_{Z,\lambda|D,\theta^{(m)}}\left[L_{nk}\right] = \sum_{q} \sum_{k_2>k} q\delta^{(m+1)}_{n,k,k_2,0,q} + \sum_{q} \sum_{k_1<k} (1-q)\delta^{(m+1)}_{n,k1,k,0,q} + \delta^{(m+1)}_{n,0,0,k}$$

$$E_{Z,\lambda|D,\theta^{(m)}}\left[L^2_{nk}\right] = \sum_{q} \sum_{k_2>k} q^2\delta^{(m+1)}_{n,k,k_2,q} + \sum_{q} \sum_{k_1<k} (1-q)^2\delta^{(m+1)}_{n,k1,k,q} + \delta^{(m+1)}_{n,0,0,k}$$

Also for any $k \neq l$,

$$E_{Z,\lambda|D,\theta^{(m)}}\left[L_{nk}L_{nl}\right] = \sum_{q} q(1-q)\delta^{(m+1)}_{n,k,l,q}$$

We use these to solve for the equation

$$\left[E_{Z,\lambda|D,\theta^{(m)}}\left(L^T L\right)\right] F \approx \left[E_{Z,\lambda|D,\theta^{(m)}}(L)\right]^T D$$

The solution therefore is

$$F \approx \left[E_{Z,\lambda|D,\theta^{(m)}}\left(L^T L\right)\right]^{-1} \left[E_{Z,\lambda|D,\theta^{(m)}}(L)\right]^T D$$

For $W = L^T L$

$$W_{kl} = \sum_n L_{kn} L_{nl}$$

$$E_{Z,\lambda|D,\theta^{(m)}}\left(W_{kl}\right) = \sum_n E_{Z,\lambda|D,\theta^{(m)}}\left(L_{nk} L_{nl}\right)$$

We use the definition of $E_{Z,\lambda|D,\theta^{(m)}}\left[L_{nk}^2\right]$ and $E_{Z,\lambda|D,\theta^{(m)}}\left[L_{nk} L_{nl}\right]$ from above to solve F.

In the same way as we computed F by solving for the normal equation obtained from taking derivative of the function $Q(\theta|\theta^{(m)})$, we take derivative of the latter with respect to s_j^2 to obtain EM updates of the residual variance terms. Taking the derivative, we obtain the estimate as

$$\widehat{s_j^{(m+1)}}^2 = \frac{1}{N} \sum_{n=1}^{N} \sum_{k_1 < k_2} \sum_{q} \delta_{n,k_1,k_2,0,q}^{(m+1)} (D_{nj} - qF_{k_1,j} - (1-q)F_{k_2,j})^2 + \frac{1}{N} \sum_{n=1}^{N} \sum_{l=1}^{K} \delta_{n,0,0,l,q}^{(m+1)} (D_{nj} - F_{l,j})^2 \quad (15)$$

where the F are the estimated values of the factors from the previous step.

We then continue this procedure described above for multiple iterations.