

# Paired factor analysis (PFA)

Gao Wang, Kushal Dey and Matthew Stephens

December 5, 2016

## 1 The PFA model

### 1.1 Data and latent variables

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  $\Lambda$ .  $Z_n$  is a  $(E + K) \times 1$  binary vector. We use  $Z_{n,k_1,k_2}$  to indicate samples in between nodes  $k_1$  and  $k_2$ , and  $Z_{n,l}$  to indicate samples near node  $l$ :

$$Pr [Z_{n,k_1,k_2} = 1] = \pi_{k_1,k_2} \quad k_1 < k_2$$

$$Pr [Z_{n,l} = 1] = \pi_l \quad l = 1, 2, \dots, K$$

with the constraint that

$$\sum_{k_1 < k_2}^E \pi_{k_1,k_2} + \sum_{l=1}^K \pi_l = 1$$

Likelihood of  $Z$  is

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

$\Lambda_n$  is a  $Q \times 1$  binary vector, where  $Q$  is the cardinality of the set of coordinates for positions in between nodes that samples are fitted to. For computational convenience we assume that these coordinates take a finite set of values between 0 and 1, say  $1/100, 2/100, \dots, 99/100, 1$ .

We have for density of  $\Lambda$ ,

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

### 1.2 The model likelihood

For data on the graph given a pair of nodes  $(k_1, k_2)$  and position  $q$  on the edge we assume a simple mixture model

$$\begin{aligned} E [D_{nj} | Z_{n,k_1,k_2} = 1, \Lambda_{n,q} = 1, F] &= qF_{k_1,j} + (1-q)F_{k_2,j} & k_1 < k_2 \\ E [D_{nj} | Z_{n,l} = 1, F] &= F_{l,j} & k_1 = k_2 = l \end{aligned} \tag{1}$$

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

$$Pr \left[ D_n | \pi, \delta, F, s_{j=1,2,\dots,J}^2 \right] = \sum_{k_1 < k_2} \pi_{k_1, k_2} Pr \left[ D_n | Z_{n, k_1, k_2} = 1, \delta, F, s_{j=1,2,\dots,J}^2 \right] + \sum_l \pi_l Pr \left[ D_n | Z_{n, l} = 1, F, s_{j=1,2,\dots,J}^2 \right] \quad (2)$$

where

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

and  $s_j^2$  is the residual variance of the  $j$ th feature. Let

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

The overall likelihood

$$\begin{aligned} L(\pi, F) &= \prod_{n=1}^N Pr \left[ D_n | \pi, F, s_{j=1,2,\dots,J}^2 \right] \\ &= \prod_{n=1}^N \left[ \sum_{k_1 < k_2} \sum_{q=1}^Q \left[ \pi_{k_1, k_2, q} \times \prod_{j=1}^J N \left( D_{nj}; qF_{k_1, j} + (1-q)F_{k_2, j}, s_j^2 \right) \right] + \right. \\ &\quad \left. \sum_l \left[ \pi_l \times \prod_{j=1}^J N \left( D_{nj}; F_{l, j}, s_j^2 \right) \right] \right] \end{aligned} \quad (3)$$

And the log likelihood

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

This is the quantity we want to maximize.

## 2 EM algorithm

### 2.1 E step

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

$$\begin{aligned}
\delta_{n,k_1,k_2}^{(m+1)} &= Pr \left[ Z_{n,k_1,k_2} = 1 | \pi^{(m)}, F^{(m)}, s_{j=1,2,\dots,J}^{(m)}, D_n \right] \\
&\propto Pr \left[ Z_{n,k_1,k_2} = 1 \right] \sum_{q=1}^Q Pr \left[ \lambda_{n,q} = 1 \right] Pr \left[ D_n | \pi^{(m)}, F^{(m)}, s_{j=1,2,\dots,J}^{(m)}, Z_{n,k_1,k_2} = 1, \lambda_{n,q} = 1 \right] \\
&\propto \pi_{k_1,k_2}^{(m)} \sum_{q=1}^Q \left[ \prod_j N \left( D_{nj} | qF_{k_1,j}^{(m)} + (1-q)F_{k_2,j}^{(m)}, s_j^{(m)2} \right) \right]
\end{aligned}$$

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

where  $s_j^{(m)2}$  is the residual variance of feature  $j$ .

We normalize  $\delta$  so that

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

We define

$$\begin{aligned}
\pi_{k_1,k_2}^{(m+1)} &= \frac{1}{N} \sum_{n=1}^N \delta_{n,k_1,k_2}^{(m+1)} \\
\pi_l^{(m+1)} &= \frac{1}{N} \sum_{n=1}^N \delta_{n,l}^{(m+1)}
\end{aligned}$$

We have therefore updated  $\pi_{k_1,k_2}^{(m)}$  to  $\pi_{k_1,k_2}^{(m+1)}$ .

#### 2.1.1 Variational inference

Now we introduce prior on parameter  $\pi$  for the latent variable  $Z$ . The motivation is to induce sparsity on factor pairs we identify from the model. Here we estimate their joint variational distribution:

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

We assume up front that  $q^*(\Lambda) = \prod_{n=1}^N \left[ \frac{1}{Q} \right]^{\Lambda_{nq}}$ . For  $\pi$  we use a Dirichlet prior as follows

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

$$\begin{aligned} \ln q^*(Z) &= E_{\pi, \Lambda} [\ln p(\pi|\alpha_0) + \ln p(\Lambda) + \ln p(Z|\pi) + \ln p(D|Z, \Lambda, F, s_{j=1,2,\dots,J})] \\ &= E_{\pi, \Lambda} [\ln p(Z|\pi) + \ln p(D|Z, \Lambda, F, s_{j=1,2,\dots,J})] + constant \\ &= \sum_{n=1}^N \sum_{k_1 < k_2} z_{n, k_1, k_2} E_{\pi} [\ln(\pi_{k_1, k_2})] \\ &\quad + \sum_{n=1}^N \sum_{k_1 < k_2} z_{n, k_1, k_2} \sum_{q=1}^Q \frac{1}{Q} \left[ - \sum_{j=1}^J \ln(s_j) - \frac{J}{2} \ln(2\pi) - \sum_{j=1}^J \frac{(D_{nj} - qF_{k_1, j} - (1-q)F_{k_2, j})^2}{2s_j^2} \right] \\ &\quad + \sum_{n=1}^N \sum_l z_{n, l} \left[ - \sum_{j=1}^J \ln(s_j) - \frac{J}{2} \ln(2\pi) - \sum_{j=1}^J \frac{(D_{nj} - F_{l, j})^2}{2s_j^2} \right] \\ &\quad + \sum_{n=1}^N \sum_{l=1}^K z_{n, l} E_{\pi} [\ln(\pi_l)] \end{aligned} \tag{5}$$

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

From here one can get

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

$$q^*(\pi) \propto \prod_{k_1 < k_2} \pi_{k_1, k_2}^{a_{k_1, k_2} - 1} \prod_{l=1}^L \pi_l^{a_l - 1}$$

then

$$\delta_{n, k_1, k_2} \propto \exp \left( E_\pi [\ln(\pi_{k_1, k_2})] + \sum_{q=1}^Q \frac{1}{Q} \left[ - \sum_{j=1}^J \ln(s_j) - \frac{J}{2} \ln(2\pi) - \sum_{j=1}^J \frac{(D_{nj} - qF_{k_1, j} - (1-q)F_{k_2, j})^2}{2s_j^2} \right] \right) \quad (7)$$

$$\delta_{n, k_1, k_2} \propto \exp \left( \psi(a_{k_1, k_2}) - \psi\left(\sum_l a_l + \sum_{k_1 < k_2} a_{k_1, k_2}\right) + \sum_{q=1}^Q \frac{1}{Q} \left[ - \sum_{j=1}^J \ln(s_j) - \frac{J}{2} \ln(2\pi) - \sum_{j=1}^J \frac{(D_{nj} - qF_{k_1, j} - (1-q)F_{k_2, j})^2}{2s_j^2} \right] \right) \quad (8)$$

$$\delta_{n, l} \propto \exp \left( E_\pi [\ln(\pi_l)] + \left[ - \sum_{j=1}^J \ln(s_j) - \frac{J}{2} \ln(2\pi) - \sum_{j=1}^J \frac{(D_{nj} - F_{l, j})^2}{2s_j^2} \right] \right) \quad (9)$$

$$\delta_{n, l} \propto \exp \left( \psi(a_l) - \psi\left(\sum_l a_l + \sum_{k_1 < k_2} a_{k_1, k_2}\right) + \left[ - \sum_{j=1}^J \ln(s_j) - \frac{J}{2} \ln(2\pi) - \sum_{j=1}^J \frac{(D_{nj} - F_{l, j})^2}{2s_j^2} \right] \right) \quad (10)$$

We can also derive variational distributions similarly for  $\pi$ .

$$\begin{aligned} \ln q^*(\pi) &= E_{\Lambda, Z} [\ln p(\pi|\alpha_0) + \ln p(Z|\pi) + \ln p(D|Z, \Lambda, F, s_{j=1,2,\dots,J})] \\ &= E_Z [\ln p(Z|\pi)] + \ln p(\pi|\alpha_0) + \text{constant} \\ &= \sum_{n=1}^N \sum_{k_1 < k_2} E(z_{n, k_1, k_2}) \ln \pi_{k_1, k_2} + \sum_{n=1}^N \sum_{l=1}^K E(z_{n, l}) \ln \pi_l + (\alpha_0 - 1) \sum_{k_1 < k_2} \ln \pi_{k_1, k_2} \\ &= \sum_{k_1 < k_2} \left[ \sum_{n=1}^N \delta_{n, k_1, k_2} + (\alpha_0 - 1) \right] \ln \pi_{k_1, k_2} + \sum_{l=1}^K \left[ \sum_{n=1}^N \delta_{n, l} + (\alpha_0 - 1) \right] \ln \pi_l \end{aligned}$$

We define

$$a_{k_1, k_2} = \alpha_0 + \sum_{n=1}^N \delta_{n, k_1, k_2}$$

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

$$q^*(\pi) = \text{Dir}(\pi|a)$$

We initialize  $\pi$  first, then use  $a_{k_1, k_2} = \alpha_0$  to begin with and estimate  $\delta_{n, k_1, k_2}$ . Then use the  $\delta_{n, k_1, k_2}$  to update  $a_{k_1, k_2}$  and proceed in this way. In this case, we do not assume independence of the  $\Lambda$  and  $Z$  variational distributions, so this model is more generalized.

## 2.2 M step

We define

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

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

$$Q(\theta|\theta^{(m)}) \propto - \sum_{n=1}^N \sum_{k_1 < k_2} \sum_{q=1}^Q \delta_{n, k_1, k_2, q}^{(m+1)} \sum_j \left[ \log s_j^{(m+1)} + \frac{(D_{nj} - qF_{k_1, j} - (1-q)F_{k_2, j})^2}{2s_j^{(m+1)^2}} \right] \quad (11)$$

$$- \sum_{n=1}^N \sum_{l=1}^K \sum_{q=1}^Q \delta_{n, l, q}^{(m+1)} \sum_j \left[ \log s_j^{(m+1)} + \frac{(D_{nj} - F_{l, j})^2}{2s_j^{(m+1)^2}} \right] \quad (12)$$

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

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

$$D_{N \times J} = L_{N \times K} F_{K \times J} + E_{N \times J} \quad (13)$$

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  $s_j$ .

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

We have

$$\begin{aligned} E_{Z,\lambda|D,\theta^{(m)}} [L_{nk}] &= \sum_q \sum_{k_2 > k} q \delta_{n,k,k_2,q}^{(m+1)} + \sum_q \sum_{k_1 < k} (1-q) \delta_{n,k_1,k,q}^{(m+1)} + \delta_{n,k}^{(m+1)} \\ E_{Z,\lambda|D,\theta^{(m)}} [L_{nk}^2] &= \sum_q \sum_{k_2 > k} q^2 \delta_{n,k,k_2,q}^{(m+1)} + \sum_q \sum_{k_1 < k} (1-q)^2 \delta_{n,k_1,k,q}^{(m+1)} + \delta_{n,0,0,k}^{(m+1)} \end{aligned}$$

Also for any  $k \neq l$ ,

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

We use these to solve for the equation

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

The solution therefore is

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

For  $W = L^T L$

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

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

We use the definition of  $E_{Z,\lambda|D,\theta^{(m)}} [L_{nk}^2]$  and  $E_{Z,\lambda|D,\theta^{(m)}} [L_{nk} L_{nl}]$  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

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

$$+ \frac{1}{N} \sum_{n=1}^N \sum_{l=1}^K \delta_{n,l,q}^{(m+1)} (D_{nj} - F_{l,j})^2 \quad (14)$$

where the  $F$  are the estimated values of the factors from the previous step. We then continue this procedure described above for multiple iterations.