Bayesian Inference for Sparse Factor Models

Yong See Foo

December 2019

1 Model

Let $\mathbf{Y} = \mathbf{L}\mathbf{F} + \mathbf{E}$, where $\mathbf{Y} \in \mathbb{R}^{G\times N}$ contains the observed data, $\mathbf{L} \in \mathbb{R}^{G\times K}$ are the factor loadings, $\mathbf{F} \in \mathbb{R}^{K\times N}$ are the factor weights, and $\mathbf{E} \in \mathbb{R}^{G\times N}$ indicates noise. The model is given by the following distributions:

$$
\begin{aligned}
p(\mathbf{y}_j \mid \mathbf{L}, \mathbf{F}, \boldsymbol{\tau}) &= \mathcal{N}\big(\mathbf{y}_i \bigm| \mathbf{L}\mathbf{f}_j, D_{\boldsymbol{\tau}}^{-1}\big) \\
p(l_{ik} \mid z_{ik}) &= \begin{cases} \delta(l_{ik}) & \text{if } z_{ik} = 0 \\ \mathcal{N}\big(l_{ik} \bigm| 0, \alpha_k^{-1}\big) & \text{if } z_{ik} = 1 \end{cases} \\
p(\mathbf{f}_j) &= \mathcal{N}(\mathbf{f}_j \mid \mathbf{0}, \mathbf{I}) \\
p(z_{ik}) &= \text{Bernoulli}(z_{ik} \mid \pi_k) \\
p(\tau_i) &= \Gamma(\tau_i \mid a_\tau, b_\tau) \\
p(\alpha_k) &= \Gamma(\alpha_k \mid a_\alpha, b_\alpha)
\end{aligned}
$$

where $D_{\mathbf{v}} = \text{diag}(\mathbf{v})$ for any vector $\mathbf{v}$. Note that vectors subscripted with i/j are rows/columns of matrices as column vectors. The conditional distribution of $\mathbf{Y}$ may also be expressed as $p(\mathbf{y}_i \mid \mathbf{L}, \mathbf{F}, \boldsymbol{\tau}) = \mathcal{N}\big(\mathbf{y}_i \bigm| \mathbf{F}^\top \mathbf{l}_i, \tau_i^{-1}\mathbf{I}\big)$.

2 MCMC

We need to ensure $l_{ik} = 0$ whenever $z_{ik} = 0$. We achieve this by introducing modifications to the conditional distribution of $\mathbf{Y}$. We have

$$
\begin{aligned}
p(\mathbf{l}_i, \mathbf{z}_i \mid \mathbf{Y}, \mathbf{F}, \boldsymbol{\tau}, \boldsymbol{\alpha}) &\propto \prod_{k\,:\,z_{ik}=1} \pi_k \sqrt{\frac{\alpha_k}{2\pi}} \times \prod_{k\,:\,z_{ik}=0} (1-\pi_k)\,\delta(l_{ik}) \\
&\quad \times \exp\left\{-\frac{\tau_i}{2}\Big(\mathbf{y}_i - [\mathbf{F}]_{\mathbf{z}_i}^\top [\mathbf{l}_i]_{\mathbf{z}_i}\Big)^\top \Big(\mathbf{y}_i - [\mathbf{F}]_{\mathbf{z}_i}^\top [\mathbf{l}_i]_{\mathbf{z}_i}\Big) - \frac{1}{2}[\mathbf{l}_i]_{\mathbf{z}_i}^\top [D_{\boldsymbol{\alpha}}]_{\mathbf{z}_i} [\mathbf{l}_i]_{\mathbf{z}_i}\right\} \\
&\propto \prod_{k\,:\,z_{ik}=1} \pi_k \sqrt{\frac{\alpha_k}{2\pi}} \times \prod_{k\,:\,z_{ik}=0} (1-\pi_k)\,\delta(l_{ik}) \\
&\quad \times \exp\left\{-\frac{1}{2}\big([\mathbf{l}]_{\mathbf{z}_i} - \boldsymbol{\mu}_{\mathbf{l}_i}\big)^\top \Sigma_{\mathbf{l}_i}^{-1}\big([\mathbf{l}]_{\mathbf{z}_i} - \boldsymbol{\mu}_{\mathbf{l}_i}\big) + \frac{1}{2}\boldsymbol{\mu}_{\mathbf{l}_i}^\top \Sigma_{\mathbf{l}_i}^{-1} \boldsymbol{\mu}_{\mathbf{l}_i}\right\}
\end{aligned} \tag{1}
$$

where

$$
\begin{aligned}
[\mathbf{F}]_{\mathbf{z}_i} &= \text{matrix consisting of rows of } \mathbf{F} \text{ whose corresponding entries of } \mathbf{z}_i \text{ are equal to 1} \\
[\mathbf{l}_i]_{\mathbf{z}_i} &= \text{vector consisting of entries of } \mathbf{l}_i \text{ whose corresponding entries of } \mathbf{z}_i \text{ are equal to 1} \\
[D_{\boldsymbol{\alpha}}]_{\mathbf{z}_i} &= \text{matrix consisting of rows of } D_{\boldsymbol{\alpha}} \text{ whose corresponding entries of } \mathbf{z}_i \text{ are equal to 1} \\
\Sigma_{\mathbf{l}_i} &= \left(\tau_i \, [\mathbf{F}]_{\mathbf{z}_i} [\mathbf{F}]_{\mathbf{z}_i}^\top + [D_{\boldsymbol{\alpha}}]_{\mathbf{z}_i}\right)^{-1} \\
\boldsymbol{\mu}_{\mathbf{l}_i} &= \tau_i \Sigma_{\mathbf{l}_i} [\mathbf{F}]_{\mathbf{z}_i} \mathbf{y}_i,
\end{aligned}
$$

and hence obtain the full conditional distribution of $\mathbf{l}_i$:

$$
p\big([\mathbf{l}_i]_{\mathbf{z}_i} \mid \mathbf{Y}, \mathbf{F}, \mathbf{Z}, \boldsymbol{\tau}, \boldsymbol{\alpha}\big) = \mathcal{N}\big([\mathbf{l}_i]_{\mathbf{z}_i} \mid \boldsymbol{\mu}_{\mathbf{l}_i}, \Sigma_{\mathbf{l}_i}\big) \times \prod_{k \,:\, z_{ik}=0} \delta(l_{ik}). \tag{2}
$$

Marginalising out $\mathbf{l}_i$ from Equation 1 gives a conditional distribution of z_{ik}:

$$
p(z_{ik} \mid \mathbf{Y}, \mathbf{F}, \mathbf{Z}_{-ik}, \boldsymbol{\tau}, \boldsymbol{\alpha}) \propto \left(\frac{\alpha_k}{2\pi}\right)^{\frac{z_{ik}}{2}} \det |\Sigma_{\mathbf{l}_i}|^{\frac{1}{2}} \exp\left\{\frac{1}{2} \boldsymbol{\mu}_{\mathbf{l}_i}^\top \Sigma_{\mathbf{l}_i}^{-1} \boldsymbol{\mu}_{\mathbf{l}_i}\right\} \pi_k^{z_{ik}} (1-\pi_k)^{1-z_{ik}}. \tag{3}
$$

We also have

$$
\begin{aligned}
p(\mathbf{f}_j \mid \mathbf{Y}, \mathbf{L}, \mathbf{Z}, \boldsymbol{\tau}, \boldsymbol{\alpha}) &\propto \exp\left\{-\frac{1}{2} (\mathbf{y}_j - \mathbf{L}\mathbf{f}_j)^\top D_{\boldsymbol{\tau}} (\mathbf{y}_j - \mathbf{L}\mathbf{f}_j) - \frac{1}{2} \mathbf{f}_j^\top \mathbf{f}_j\right\} \\
&\propto \exp\left\{-\frac{1}{2} \left(\mathbf{f}_j - \boldsymbol{\mu}_{\mathbf{f}_j}\right)^\top \Sigma_{\mathbf{f}_j}^{-1} \left(\mathbf{f}_j - \boldsymbol{\mu}_{\mathbf{f}_j}\right)\right\}
\end{aligned}
$$

where

$$
\begin{aligned}
\Sigma_{\mathbf{f}_j} &= \left(\mathbf{L}^\top D_{\boldsymbol{\tau}} \mathbf{L} + \mathbf{I}\right)^{-1} \\
\boldsymbol{\mu}_{\mathbf{f}_j} &= \Sigma_{\mathbf{f}_j} \mathbf{L}^\top D_{\boldsymbol{\tau}} \mathbf{y}_j,
\end{aligned}
$$

thus arriving at the full conditional distribution of $\mathbf{f}_j$:

$$
p(\mathbf{f}_j \mid \mathbf{Y}, \mathbf{L}, \mathbf{Z}, \boldsymbol{\tau}, \boldsymbol{\alpha}) = \mathcal{N}\left(\mathbf{f}_j \,\middle|\, \boldsymbol{\mu}_{\mathbf{f}_j}, \Sigma_{\mathbf{f}_j}\right). \tag{4}
$$

Lastly, we have the full conditional distribution of τ_i:

$$
p(\tau_i \mid \mathbf{Y}, \mathbf{L}, \mathbf{F}, \mathbf{Z}, \boldsymbol{\alpha}) = \Gamma\left(\tau_i \,\middle|\, a_\tau + \frac{N}{2}, b_\tau + \frac{1}{2}\left(\mathbf{y}_i - [\mathbf{F}]_{\mathbf{z}_i}^\top [\mathbf{l}_i]_{\mathbf{z}_i}\right)^\top \left(\mathbf{y}_i - [\mathbf{F}]_{\mathbf{z}_i}^\top [\mathbf{l}_i]_{\mathbf{z}_i}\right)\right), \tag{5}
$$

and the full conditional distribution of α_k:

$$
p(\alpha_k \mid \mathbf{Y}, \mathbf{L}, \mathbf{F}, \mathbf{Z}, \boldsymbol{\tau}) = \Gamma\left(\alpha_k \,\middle|\, a_\alpha + \frac{1}{2}\sum_{i=1}^{G} z_{ik}, b_\alpha + \frac{1}{2}\sum_{i \,:\, z_{ik}=1} l_{ik}^2\right). \tag{6}
$$

3 Variational Inference

3.1 Mean-field approximation

Use the variational factorisation

$$
q(\mathbf{L}, \mathbf{F}, \mathbf{Z}, \boldsymbol{\tau}, \boldsymbol{\alpha}) = \prod_{i=1}^{G}\left[\prod_{k=1}^{K} q(l_{ik} \mid z_{ik}) q(z_{ik})\right] q(\tau_i) \times \prod_{j=1}^{N} q(\mathbf{f}_j) \times \prod_{k=1}^{K} q(\alpha_k) \tag{7}
$$

as an approximation to the posterior distribution, where

$$
\begin{aligned}
q(l_{ik} \mid z_{ik}) &= \mathcal{N}\big(l_{ik} \mid \mu_{l_{ik}}, \sigma^2_{l_{ik}}\big)^{z_{ik}} \times \delta(l_{ik})^{1-z_{ik}} \\
q(z_{ik}) &= \text{Bernoulli}(z_{ik} \mid \eta_{ik}) \\
q(\mathbf{f}_j) &= \mathcal{N}\Big(\mathbf{f}_j \;\Big|\; \boldsymbol{\mu}_{\mathbf{f}_j}, \Sigma_{\mathbf{f}_j}\Big) \\
q(\tau_i) &= \Gamma\Big(\tau_i \;\Big|\; \hat{a}_{\tau_i}, \hat{b}_{\tau_i}\Big) \\
q(\alpha_k) &= \Gamma\Big(\alpha_k \;\Big|\; \hat{a}_{\alpha_k}, \hat{b}_{\alpha_k}\Big).
\end{aligned}
$$

Coordinate ascent for $\mathbf{l}_i$ and $\mathbf{z}_i$ gives

$$
\begin{aligned}
q^*(l_{ik}, z_{ik}) \propto\; & \exp\left\{\mathbb{E}_{\mathbf{l}_{-ik},\mathbf{F},\mathbf{z}_{-ik},,\tau_i,\boldsymbol{\alpha}}[\log p(l_{ik}, z_{ik} \mid \mathbf{Y}, \mathbf{l}_{-ik}, \mathbf{F}, \mathbf{z}_{-ik}, \boldsymbol{\tau}, \boldsymbol{\alpha})]\right\} \\
\propto\; & \exp\Bigg\{\frac{z_{ik}}{2}\mathbb{E}_{\alpha_k}\left[\log\frac{\alpha_k}{2\pi}\right] - \mathbb{E}_{\mathbf{l}_{-ik},\mathbf{F},\mathbf{z}_{-ik},\tau_i}\left[\frac{\tau_i}{2}\left(\mathbf{y}_i - [\mathbf{F}]^\top_{\mathbf{z}_i}[\mathbf{l}_i]_{\mathbf{z}_i}\right)^\top\left(\mathbf{y}_i - [\mathbf{F}]^\top_{\mathbf{z}_i}[\mathbf{l}_i]_{\mathbf{z}_i}\right)\right] \\
& \quad - \frac{z_{ik}}{2}\mathbb{E}_{\boldsymbol{\alpha}}\left[\alpha_k l^2_{ik}\right]\Bigg\} \times \prod_{k=1}^{K}\pi_k^{z_{ik}}(1-\pi_k)^{1-z_{ik}}, \\
\propto\; & \exp\Bigg\{-\frac{\hat{a}_{\tau_i}}{2\hat{b}_{\tau_i}}\mathbb{E}_{\mathbf{l}_{-ik},\mathbf{F},\mathbf{z}_{-ik}}\left[-2z_{ik}\mathbf{y}_i^\top\mathbf{f}_k l_{ik} + 2z_{ik}\mathbf{f}_k^\top\sum_{k'\neq k} z_{ik'}\mathbf{f}_{k'}l_{ik'}l_{ik} + z_{ik}\mathbf{f}_k^\top\mathbf{f}_k l^2_{ik}\right] \\
& \quad + \frac{z_{ik}}{2}\left(\psi(\hat{a}_{\alpha_k}) - \log 2\pi\hat{b}_{\alpha_k} - \frac{\hat{a}_{\alpha_k}}{\hat{b}_{\alpha_k}}l^2_{ik}\right)\Bigg\} \times \prod_{k=1}^{K}\pi_k^{z_{ik}}(1-\pi_k)^{1-z_{ik}} \\
\propto\; & \exp\Bigg\{-\frac{\hat{a}_{\tau_i}}{2\hat{b}_{\tau_i}}\Bigg[-2z_{ik}\mathbf{y}_i^\top\boldsymbol{\mu}_{\mathbf{f}_k}l_{ik} + 2z_{ik}\boldsymbol{\mu}^\top_{\mathbf{f}_k}\sum_{k'\neq k}\eta_{ik'}\boldsymbol{\mu}_{\mathbf{f}_{k'}}\mu_{\mathbf{l}_{ik'}}l_{ik} \\
& \quad + z_{ik}\sum_{j=1}^{N}\left(\left[\Sigma_{\mathbf{f}_j}\right]_{kk} + \left[\boldsymbol{\mu}_{\mathbf{f}_j}\right]^2_k\right)l^2_{ik}\Bigg] + \frac{z_{ik}}{2}\left(\psi(\hat{a}_{\alpha_k}) - \log 2\pi\hat{b}_{\alpha_k} - \frac{\hat{a}_{\alpha_k}}{\hat{b}_{\alpha_k}}l^2_{ik}\right)\Bigg\} \\
& \quad \times \pi_k^{z_{ik}}(1-\pi_k)^{1-z_{ik}},
\end{aligned}
$$

which corresponds to the updates

$$
\begin{aligned}
\sigma^{2*}_{l_{ik}} &= \left(\frac{\hat{a}_{\tau_i}}{\hat{b}_{\tau_i}}\sum_{j=1}^{N}\left(\left[\Sigma_{\mathbf{f}_j}\right]_{kk} + \left[\boldsymbol{\mu}_{\mathbf{f}_j}\right]^2_k\right) + \frac{\hat{a}_{\alpha_k}}{\hat{b}_{\alpha_k}}\right)^{-1} \\
\mu^*_{l_{ik}} &= \frac{\hat{a}_{\tau_i}}{\hat{b}_{\tau_i}}\sigma^{2*}_{l_{ik}}\boldsymbol{\mu}^\top_{\mathbf{f}_k}\left(\mathbf{y}_i - \sum_{k'\neq k}\eta_{ik'}\boldsymbol{\mu}_{\mathbf{f}_{k'}}\mu_{\mathbf{l}_{ik'}}\right)
\end{aligned} \tag{8}
$$

$$
q(z_{ik}) \propto \exp\left\{\frac{z_{ik}}{2}\left(\psi(\hat{a}_{\alpha_k}) - \log 2\pi\hat{b}_{\alpha_k} + \frac{\mu^{2*}_{l_{ik}}}{\sigma^{2*}_{l_{ik}}}\right)\right\}\left(\sqrt{\sigma^{2*}_{l_{ik}}}\pi_k\right)^{z_{ik}}(1-\pi_k)^{1-z_{ik}}. \tag{9}
$$

Coordinate ascent for $\mathbf{f}_j$ gives

$$
\begin{aligned}
q^*(\mathbf{f}_j) &\propto \exp\left\{\mathbb{E}_{,\mathbf{L},\mathbf{Z},\boldsymbol{\tau}}[\log p(\mathbf{f}_j \mid \mathbf{Y}, \mathbf{L}, \mathbf{Z}, \boldsymbol{\tau}, \boldsymbol{\alpha})]\right\} \\
&\propto \exp\left\{\mathbb{E}_{,\mathbf{L},\mathbf{Z},\boldsymbol{\tau}}\left[-\frac{1}{2}(\mathbf{y}_j - \mathbf{L}\mathbf{f}_j)^\top D_{\boldsymbol{\tau}}(\mathbf{y}_j - \mathbf{L}\mathbf{f}_j)\right] - \frac{1}{2}\mathbf{f}_j^\top\mathbf{f}_j\right\} \\
&\propto \exp\left\{\mathbf{y}_j^\top D_{\overline{\boldsymbol{\tau}}}\overline{\mathbf{L}}\mathbf{f}_j - \frac{1}{2}\mathbf{f}_j^\top\overline{\mathbf{L}^\top D_{\boldsymbol{\tau}}\mathbf{L}}\mathbf{f}_j - \frac{1}{2}\mathbf{f}_j^\top\mathbf{f}_j\right\}
\end{aligned}
$$

where

$$
\begin{aligned}
D_{\overline{\boldsymbol{\tau}}} &= \operatorname{diag}\left(\left\{\frac{\hat{a}_{\tau_i}}{\hat{b}_{\tau_i}}\right\}_{i=1}^{G}\right) \\
\left[\overline{\mathbf{L}}\right]_{ik} &= \eta_{ik}\mu_{l_{ik}} \\
\left[\overline{\mathbf{L}^\top D_{\boldsymbol{\tau}} \mathbf{L}}\right]_{kk'} &= \sum_{i=1}^{G} \frac{\hat{a}_{\tau_i}}{\hat{b}_{\tau_i}} \eta_{ik}\eta_{ik'}^{1-\delta_{kk'}} \left(\delta_{kk'}\sigma_{l_{ik}}^2 + \mu_{l_{ik}}\mu_{l_{ik'}}\right),
\end{aligned}
$$

which corresponds to the updates

$$
\begin{aligned}
\Sigma_{\mathbf{f}_j}^* &= \left(\overline{\mathbf{L}^\top D_{\boldsymbol{\tau}} \mathbf{L}} + \mathbf{I}\right)^{-1} \\
\boldsymbol{\mu}_{\mathbf{f}_j}^* &= \Sigma_{\mathbf{f}_j}^* \overline{\mathbf{L}}^T D_{\overline{\boldsymbol{\tau}}} \mathbf{y}_j.
\end{aligned}
\tag{10}
$$

Coordinate ascent for τ_i gives

$$
\begin{aligned}
q^*(\tau_i) &\propto \exp\left\{\mathbb{E}_{\mathbf{L},\mathbf{F},\mathbf{Z}}[\log p(\tau_i \mid \mathbf{Y}, \mathbf{L}, \mathbf{F}, \mathbf{Z}, \boldsymbol{\alpha})]\right\} \\
&\propto \exp\left\{\left(a_\tau + \frac{N}{2}\right)\log\tau_i - b_\tau\tau_i - \frac{\tau_i}{2}\mathbb{E}_{\mathbf{L},\mathbf{F},\mathbf{Z}}\left[\left(\mathbf{y}_i - \mathbf{F}^\top\mathbf{l}_i\right)^\top\left(\mathbf{y}_i - \mathbf{F}^\top\mathbf{l}_i\right)\right]\right\} \\
&\propto \exp\left\{\left(a_\tau + \frac{N}{2}\right)\log\tau_i - \left(b_\tau + \frac{1}{2}\left(\mathbf{y}_i^\top\mathbf{y}_i - 2\overline{\mathbf{l}_i}^\top\overline{\mathbf{F}}\mathbf{y}_i + \overline{\mathbf{l}_i^\top\mathbf{F}\mathbf{F}^\top\mathbf{l}_i}\right)\right)\tau_i\right\}
\end{aligned}
$$

where

$$
\begin{aligned}
\overline{\mathbf{l}_i} &= \{\eta_{ik}\mu_{l_{ik}}\}_{k=1}^{K} \\
\left[\overline{\mathbf{F}}\right]_{kj} &= \mu_{f_{kj}} \\
\overline{\mathbf{l}_i^\top\mathbf{F}\mathbf{F}^\top\mathbf{l}_i} &= \sum_{k=1}^{K}\sum_{k'=1}^{K}\left(\eta_{ik}\eta_{ik'}^{1-\delta_{kk'}}\left(\delta_{kk'}\sigma_{l_{ik}}^2 + \mu_{l_{ik}}\mu_{l_{ik'}}\right)\sum_{j=1}^{N}\left(\left[\Sigma_{\mathbf{f}_j}\right]_{kk'} + \left[\boldsymbol{\mu}_{\mathbf{f_j}}\right]_k\left[\boldsymbol{\mu}_{\mathbf{f_j}}\right]_{k'}\right)\right),
\end{aligned}
$$

which corresponds to the updates

$$
\begin{aligned}
\hat{a}_{\tau_i}^* &= a_\tau + \frac{N}{2} \\
\hat{b}_{\tau_i}^* &= b_\tau + \frac{1}{2}\left(\mathbf{y}_i^\top\mathbf{y}_i - 2\overline{\mathbf{l}_i}^\top\overline{\mathbf{F}}\mathbf{y}_i + \overline{\mathbf{l}_i^\top\mathbf{F}\mathbf{F}^\top\mathbf{l}_i}\right).
\end{aligned}
\tag{11}
$$

Coordinate ascent for α_k gives

$$
\begin{aligned}
q^*(\alpha_k) &\propto \exp\left\{\mathbb{E}_{\mathbf{L},\mathbf{Z}}[\log p(\alpha_k \mid \mathbf{Y}, \mathbf{L}, \mathbf{F}, \mathbf{Z}, \boldsymbol{\tau})]\right\} \\
&\propto \exp\left\{\left(a_\alpha + \frac{1}{2}\mathbb{E}_{\mathbf{Z}}\left[\sum_{i=1}^{G} z_{ik}\right]\right)\log\alpha_k - b_\alpha\alpha_k - \frac{\alpha_k}{2}\mathbb{E}_{\mathbf{L},\mathbf{Z}}\left[\sum_{i\,:\,z_{ik}=1} l_{ik}^2\right]\right\}
\end{aligned}
$$

which corresponds to the updates

$$
\begin{aligned}
\hat{a}_{\alpha_k}^* &= a_\alpha + \frac{1}{2}\sum_{i=1}^{G}\eta_{ik} \\
\hat{b}_{\alpha_k}^* &= b_\alpha + \frac{1}{2}\sum_{i=1}^{G}\eta_{ik}\left(\sigma_{l_{ik}}^2 + \mu_{l_{ik}}^2\right).
\end{aligned}
\tag{12}
$$

3.2 Capturing dependency within $\mathbf{l}_i$

To capture dependency within $\mathbf{l}_i$, the variational factorisation is modified:

$$q(\mathbf{L}, \mathbf{F}, \mathbf{Z}, \boldsymbol{\tau}, \boldsymbol{\alpha}) = \prod_{i=1}^{G} q(\mathbf{l}_i, \mathbf{z}_i) q(\tau_i) \times \prod_{j=1}^{N} q(\mathbf{f}_j) \times \prod_{k=1}^{K} q(\alpha_k) \tag{13}$$

where

$$\begin{aligned}
q(\mathbf{l}_i, \mathbf{z}_i) &= \mathcal{N}\big([\mathbf{l}_i]_{\mathbf{z}_i} \;\big|\; \boldsymbol{\mu}_{\mathbf{l}_i}, \Sigma_{\mathbf{l}_i}\big) \times \prod_{k\,:\,z_{ik}=0} \delta(l_{ik}) \times q(\mathbf{z}_i) \\
q(\mathbf{f}_j) &= \mathcal{N}\Big(\mathbf{f}_j \;\Big|\; \boldsymbol{\mu}_{\mathbf{f}_j}, \Sigma_{\mathbf{f}_j}\Big) \\
q(\tau_i) &= \Gamma\Big(\tau_i \;\Big|\; \hat{a}_{\tau_i}, \hat{b}_{\tau_i}\Big) \\
q(\alpha_k) &= \Gamma\Big(\alpha_k \;\Big|\; \hat{a}_{\alpha_k}, \hat{b}_{\alpha_k}\Big).
\end{aligned}$$

Note that $q(\mathbf{z}_i)$ is yet to be parameterised.

Coordinate ascent for $\mathbf{l}_i$ and $\mathbf{z}_i$ gives

$$\begin{aligned}
q^*(\mathbf{l}_i, \mathbf{z}_i) &\propto \exp\left\{\mathbb{E}_{\mathbf{F},\tau_i,\boldsymbol{\alpha}}[\log p(\mathbf{l}_i, \mathbf{z}_i \mid \mathbf{Y}, \mathbf{F}, \boldsymbol{\tau}, \boldsymbol{\alpha})]\right\} \\
&\propto \exp\Bigg\{\frac{1}{2}\sum_{k\,:\,z_{ik}=1} \mathbb{E}_{\alpha_k}\left[\log\frac{\alpha_k}{2\pi}\right] - \mathbb{E}_{\mathbf{F},\tau_i}\left[\frac{\tau_i}{2}\left(\mathbf{y}_i - [\mathbf{F}]_{\mathbf{z}_i}^{\mathsf{T}}[\mathbf{l}_i]_{\mathbf{z}_i}\right)^{\mathsf{T}}\left(\mathbf{y}_i - [\mathbf{F}]_{\mathbf{z}_i}^{\mathsf{T}}[\mathbf{l}_i]_{\mathbf{z}_i}\right)\right] \\
&\qquad - \frac{1}{2}\mathbb{E}_{\boldsymbol{\alpha}}\left[[\mathbf{l}_i]_{\mathbf{z}_i}^{\mathsf{T}}[D_{\boldsymbol{\alpha}}]_{\mathbf{z}_i}[\mathbf{l}_i]_{\mathbf{z}_i}\right]\Bigg\} \times \prod_{k=1}^{K} \pi_k^{z_{ik}}(1-\pi_k)^{1-z_{ik}} \\
&\propto \exp\Bigg\{\frac{1}{2}\sum_{k\,:\,z_{ik}=1}\left(\overline{\log\alpha_k} - \log 2\pi\right) - \frac{\hat{a}_{\tau_i}}{2\hat{b}_{\tau_i}}\mathbb{E}_{\mathbf{F}}\left[-2\mathbf{y}_i^{\mathsf{T}}[\mathbf{F}]_{\mathbf{z}_i}^{\mathsf{T}}[\mathbf{l}_i]_{\mathbf{z}_i} + \left[\mathbf{l}_i^{\mathsf{T}}\right]_{\mathbf{z}_i}[\mathbf{F}]_{\mathbf{z}_i}[\mathbf{F}]_{\mathbf{z}_i}^{\mathsf{T}}[\mathbf{l}_i]_{\mathbf{z}_i}\right] \\
&\qquad - \frac{1}{2}[\mathbf{l}_i]_{\mathbf{z}_i}^{\mathsf{T}}[D_{\overline{\boldsymbol{\alpha}}}]_{\mathbf{z}_i}[\mathbf{l}_i]_{\mathbf{z}_i}\Bigg\} \times \prod_{k=1}^{K} \pi_k^{z_{ik}}(1-\pi_k)^{1-z_{ik}} \\
&\propto \exp\Bigg\{\frac{1}{2}\sum_{k\,:\,z_{ik}=1}\left(\overline{\log\alpha_k} - \log 2\pi\right) + \frac{\hat{a}_{\tau_i}}{\hat{b}_{\tau_i}}\sum_{k\,:\,z_{ik}=1} l_{ik}\sum_{j=1}^{N} y_{ij}\overline{f_{kj}} \\
&\qquad - \frac{\hat{a}_{\tau_i}}{2\hat{b}_{\tau_i}}\sum_{k\,:\,z_{ik}=1}\sum_{k'\,:\,z_{ik'}=1} l_{ik}l_{ik'}\sum_{j=1}^{N}\overline{f_{kj}f_{k'j}} - \frac{1}{2}[\mathbf{l}_i]_{\mathbf{z}_i}^{\mathsf{T}}[D_{\overline{\boldsymbol{\alpha}}}]_{\mathbf{z}_i}[\mathbf{l}_i]_{\mathbf{z}_i}\Bigg\} \\
&\qquad \times \prod_{k=1}^{K} \pi_k^{z_{ik}}(1-\pi_k)^{1-z_{ik}},
\end{aligned}$$

where

$$\begin{aligned}
\overline{\log\alpha_k} &= \psi(\hat{a}_{\alpha_k}) - \log\hat{b}_{\alpha_k} \\
D_{\overline{\boldsymbol{\alpha}}} &= \operatorname{diag}\left(\left\{\frac{\hat{a}_{\alpha_k}}{\hat{b}_{\alpha_k}}\right\}_{k=1}^{K}\right) \\
\overline{f_{kj}} &= \left[\boldsymbol{\mu}_{\mathbf{f}_j}\right]_k \\
\overline{f_{kj}f_{k'j}} &= \left[\Sigma_{\mathbf{f}_j}\right]_{kk'} + \left[\boldsymbol{\mu}_{\mathbf{f}_j}\right]_k\left[\boldsymbol{\mu}_{\mathbf{f}_j}\right]_{k'}.
\end{aligned}$$

The variational parameters corresponding to $[\mathbf{l}_i]_{\mathbf{z}_i}$ are thus updated to

$$
\begin{aligned}
\Sigma^*_{\mathbf{l}_i} &= \left(\frac{\hat{a}_{\tau_i}}{\hat{b}_{\tau_i}} \sum_{j=1}^{N} \left(\left[\Sigma_{\mathbf{f}_j}\right]_{\mathbf{z}_i,\mathbf{z}_i} + \left[\boldsymbol{\mu}_{\mathbf{f}_j}\right]_{\mathbf{z}_i} \left[\boldsymbol{\mu}_{\mathbf{f}_j}\right]^{\top}_{\mathbf{z}_i} \right) + [D_{\overline{\boldsymbol{\alpha}}}]_{\mathbf{z}_i} \right)^{-1} \\
\boldsymbol{\mu}^*_{\mathbf{l}_i} &= \frac{\hat{a}_{\tau_i}}{\hat{b}_{\tau_i}} \Sigma^*_{\mathbf{l}_i} \sum_{j=1}^{N} y_{ij} \left[\boldsymbol{\mu}_{\mathbf{f}_j}\right]_{\mathbf{z}_i}
\end{aligned}
\tag{14}
$$

where $\left[\Sigma_{\mathbf{f}_j}\right]_{\mathbf{z}_i,\mathbf{z}_i}$ is a principal minor of $\Sigma_{\mathbf{f}_j}$ whose rows and columns indices correspond to the entries of 1s in $\mathbf{z}_i$, and $\left[\boldsymbol{\mu}_{\mathbf{f}_j}\right]_{\mathbf{z}_i}$ is a vector consisting of entries of $\boldsymbol{\mu}_{\mathbf{f}_j}$ whose corresponding entries of $\mathbf{z}_i$ are equal to 1.

Marginalising out $\mathbf{l}_i$ then gives

$$
q^*(\mathbf{z}_i) \propto \exp\left\{ \frac{1}{2} \sum_{k\,:\,z_{ik}=1} \left(\overline{\log \alpha_k} - \log 2\pi \right) + \frac{1}{2} \boldsymbol{\mu}^{\top}_{\mathbf{l}_i} \Sigma^{-1}_{\mathbf{l}_i} \boldsymbol{\mu}_{\mathbf{l}_i} \right\} \det |\Sigma_{\mathbf{l}_i}|^{\frac{1}{2}} \prod_{k=1}^{K} \pi_k^{z_{ik}} \left(1 - \pi_k\right)^{1-z_{ik}} . \tag{15}
$$

Unfortunately, this distribution is intractable to obtain. We instead approximate $q^*(\mathbf{z}_i)$ with

$$
\hat{q}(\mathbf{z}_i) = \prod_{k=1}^{K} \hat{q}(z_{ik}) = \prod_{k=1}^{K} \mathrm{Bernoulli}(z_{ik} \mid \gamma_{ik}). \tag{16}
$$

Methods for estimating γ_{ik} will be addressed in Section 3.3.

For the remaining variational parameters, computations similar to the previous section give the following updates:

$$
\begin{aligned}
\Sigma^*_{\mathbf{f}_j} &= \left(\overline{\mathbf{L}^{\top} D_{\boldsymbol{\tau}} \mathbf{L}} + \mathbf{I} \right)^{-1} \\
\boldsymbol{\mu}^*_{\mathbf{f}_j} &= \Sigma^*_{\mathbf{f}_j} \overline{\mathbf{L}}^{T} D_{\overline{\boldsymbol{\tau}}} \mathbf{y}_j
\end{aligned}
\tag{17}
$$

$$
\begin{aligned}
\hat{a}^*_{\tau_i} &= a_\tau + \frac{N}{2} \\
\hat{b}^*_{\tau_i} &= b_\tau + \frac{1}{2} \left(\mathbf{y}_i^{\top} \mathbf{y}_i - 2 \overline{\mathbf{l}_i}^{\top} \overline{\mathbf{F}} \mathbf{y}_i + \overline{\mathbf{l}_i^{\top} \mathbf{F} \mathbf{F}^{\top} \mathbf{l}_i} \right)
\end{aligned}
\tag{18}
$$

$$
\begin{aligned}
\hat{a}^*_{\alpha_k} &= a_\alpha + \frac{1}{2} \sum_{i=1}^{G} \eta_{ik} \\
\hat{b}^*_{\alpha_k} &= b_\alpha + \frac{1}{2} \sum_{i=1}^{G} \eta_{ik} \left(\sigma^2_{l_{ik}} + \mu^2_{l_{ik}} \right),
\end{aligned}
\tag{19}
$$

where

$$
\begin{aligned}
\left[\overline{\mathbf{L}^{\top} D_{\boldsymbol{\tau}} \mathbf{L}} \right]_{kk'} &= \sum_{i=1}^{G} \frac{\hat{a}_{\tau_i}}{\hat{b}_{\tau_i}} \mathbb{E}_{\mathbf{Z}}[z_{ik} z_{ik'}] \left([\Sigma_{\mathbf{l}_i}]_{kk'} + [\boldsymbol{\mu}_{\mathbf{l}_i}]_k [\boldsymbol{\mu}_{\mathbf{l}_i}]_{k'} \right) \\
\overline{\mathbf{L}} &= \left[\overline{\mathbf{l}_1} \quad \overline{\mathbf{l}_2} \quad \cdots \quad \overline{\mathbf{l}_G} \right]^{\top} \\
\left[\overline{\mathbf{l}_i} \right]_k &= \mathbb{E}_{\mathbf{Z}}[z_{ik}] \left[\mu_{\mathbf{l}_i} \right]_k \\
\overline{\mathbf{l}_i^{\top} \mathbf{F} \mathbf{F}^{\top} \mathbf{l}_i} &= \sum_{k=1}^{K} \sum_{k'=1}^{K} \left(\mathbb{E}_{\mathbf{Z}}[z_{ik} z_{ik'}] \left([\Sigma_{\mathbf{l}_i}]_{kk'} + [\boldsymbol{\mu}_{\mathbf{l}_i}]_k [\boldsymbol{\mu}_{\mathbf{l}_i}]_{k'} \right) \sum_{j=1}^{N} \left(\left[\Sigma_{\mathbf{f}_j}\right]_{kk'} + \left[\boldsymbol{\mu}_{\mathbf{f}_j}\right]_k \left[\boldsymbol{\mu}_{\mathbf{f}_j}\right]_{k'} \right) \right).
\end{aligned}
$$

The expectations taken over $\mathbf{Z}$ are approximated by

$$\mathbb{E}_{\mathbf{Z}}[z_{ik}] \approx \gamma_{ik}$$
$$\mathbb{E}_{\mathbf{Z}}[z_{ik}z_{ik'}] \approx \gamma_{ik}\gamma_{ik'}^{1-\delta_{kk'}}.$$

3.3 Estimating γ_{ik}

Let $B = \{0,1\}^K$, the set of binary vectors of size K. Through this section, q^* refers to the unnormalised density stated in Equation (15) for some fixed i. The direct approach of estimating γ_{ik} is to compute

$$\gamma_{ik}^* = \sum_{\substack{\boldsymbol{\zeta}\in B\\ \zeta_k=1}} q^*(\boldsymbol{\zeta}) \Bigg/ \sum_{\boldsymbol{\zeta}\in B} q^*(\boldsymbol{\zeta}) . \tag{20}$$

However, this would take 2^K matrix inversions, which is infeasible for large K. Instead, we seek to estimate γ_{ik} (for all k with some fixed i) using values of $q^*(\boldsymbol{\zeta})$ for $\boldsymbol{\zeta} \in \mathbf{B}_T$, where $\mathbf{B}_T$ is a random subset of B of size T. Define

$$g_{ik}(z) = \sum_{\substack{\boldsymbol{\zeta}\in \mathbf{B}_T\\ \zeta_k=z}} q^*(\boldsymbol{\zeta}) \Bigg/ |\{\boldsymbol{\zeta} \in \mathbf{B}_T : \zeta_k = z\}| \quad \text{for } z = 0, 1.$$

We then have an aggregation-based method for estimating γ_{ik}:

$$\gamma_{ik}^* = \frac{g_{ik}(1)}{g_{ik}(0) + g_{ik}(1)}. \tag{21}$$

Another *ad hoc* method is to use the independence assumption in Equation (16). Let

$$\gamma_{ik} = \frac{1}{1 + \exp(-u_{ik})},$$

the approximation

$$q^*(\boldsymbol{\zeta}) \approx \prod_{k=1}^{K} \gamma_{ik}^{\zeta_k} (1 - \gamma_{ik})^{1-\zeta_k}$$

is then equivalent to

$$q^*(\boldsymbol{\zeta}) \approx \prod_{k=1}^{K} \frac{\exp(u_{ik}\zeta_k)}{1 + \exp(u_{ik})}.$$

Taking logs of both sides then gives a regression problem:

$$\log q^*(\boldsymbol{\zeta}) \approx u_0 + \sum_{k=1}^{K} u_{ik}\zeta_k \tag{22}$$

where u_0 is some constant.

One last approach is to apply coordinate ascent to only a random subset of $\mathbf{l}_i$ and $\mathbf{z}_i$ during each iteration. If the subset is small enough, the direct approach found in Equation 20 can then be feasible. This idea is motivated by stochastic variational inference, but does not share the same justifications as values of $\mathbf{l}_i$ and $\mathbf{z}_i$ are assumed to be dependent in the variational space.